SUPPORTING INFORMATION

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, 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, Xiaoguwei Island, Panyu District, Guangzhou, 510006, P. R. China.

*Corresponding author

Email address: zbingj@mail.sysu.edu.cn
Descriptors selection

At the beginning, seven general 1D structural descriptors including LCD, PLD, ASA, AV, and $\rho$, as well as the size of the ring (SR) and void fraction of accessible volume ($\phi_{AV}$) were used to quantitatively describe the structure of the zeolite. As shown in Figure S1a, considering the discreteness of SR and the weak correlation between SR and the C$_3$H$_6$ adsorption capacity ($N_{C_3H_6}$), the SR was discarded. Moreover, the strong linear correlation between AV and $\phi_{AV}$ in Figure S1b suggests that only one of them could be selected. Ultimately, five 1D structural descriptors (LCD, PLD, ASA, AV, and $\rho$), a 2D structural descriptor (PSD) and an energy descriptor ($Q_{st}$) were adopted in our QSPR analysis.

![Figure S1](image_url)

**Figure S1.** Relationships between (a) $N_{C_3H_6} \sim$ SR, (b) AV $\sim \phi_{AV}$ at 300 K and 50.65 kPa.
Figure S2. Relationships between (a) \( N_{\text{C}_3\text{H}_6} \sim \text{LCD} \), (b) \( N_{\text{C}_3\text{H}_6} \sim \text{PLD} \), (c) \( N_{\text{C}_3\text{H}_6} \sim \text{ASA} \), (d) \( N_{\text{C}_3\text{H}_6} \sim \text{AV} \), (e) \( N_{\text{C}_3\text{H}_6} \sim \rho \), and (f) \( N_{\text{C}_3\text{H}_6} \sim Q_{\text{st}} \) at 300 K and 50.65 kPa.
Predicted $\text{N}_\text{C}_3\text{H}_6$ by multilinear regression models versus simulated results

Due to the limitations of univariate analysis in identifying synergies, five structural descriptors (LCD, PLD, ASA, AV, and $\rho$) with strong linear correlation with $\text{N}_\text{C}_3\text{H}_6$ in univariate analysis were used for multilinear regression analysis. The predicted results for $\text{N}_\text{C}_3\text{H}_6$ at 300 K, 5,065 and 50.65 kPa on the randomly selected training (80%) and test (20%) sets are illustrated in Figure S3. For high-pressure gas adsorption, guest molecules generally occupy almost the entire void space of materials. Therefore, at 300 K and 5,065 kPa, the fitting of the model with five structural descriptors capturing the global porosity characteristics of the zeolites is satisfactory on the training ($R^2 = 0.844$) and test ($R^2 = 0.800$) sets. In contrast, the role of these descriptors is weakened when only a small fraction of void space is filled at low pressure. It is not surprising that the generalization ability of the multilinear regression model with five structural descriptors is not acceptable with $R^2 = 0.558$ on the test set at 300 K and 50.65 kPa. Hence, some descriptors describing local porosity properties should be introduced for the adsorption at low pressure, which is diverse and difficult to obtain. Certainly, the limitation of the multilinear regression model may also be another factor.

**Figure S3.** Comparison of multilinear regression model predictions using five structural descriptors (LCD, PLD, ASA, AV, and $\rho$) with $\text{N}_\text{C}_3\text{H}_6$ obtained by GCMC simulation on the training (blue symbols) and test (orange symbols) sets at 300 K, 5,065 kPa (a) and
The prediction results of multilinear regression models at other pressures (1,013, 303.9, 202.6, and 101.3 kPa) are shown in Figure S4. Similar to univariate analysis, the interpretation of structural descriptors to $N_{C_3H_6}$ declines and the fitting accuracy of model deteriorates gradually with the decrease of adsorption pressure. Figure S5 shows moderate prediction results of multilinear regression model with the addition of pressure descriptor (P) by putting combined $N_{C_3H_6}$ data at six different pressures together.

**Figure S4.** Comparison of multilinear regression model predictions using five structural descriptors (LCD, PLD, ASA, AV, and $\rho$) with $N_{C_3H_6}$ obtained by GCMC simulation on the training (blue symbols) and test (orange symbols) sets at 300 K, 1,013 kPa (a), 303.9 kPa (b), 202.6 kPa (c), and 101.3 kPa (d).
Figure S5. Comparison of multilinear regression model predictions using five structural
descriptors (LCD, PLD, ASA, AV, and \( \rho \)) and P with combined \( N_{\text{C}_3\text{H}_6} \) obtained by GCMC
simulation on the training (blue symbols) and test (orange symbols) sets at six different
pressures.

Table S1 shows the coefficients of multilinear regression models at different pressures. It
can be seen from the table that the correlation between LCD, ASA, AV, \( \rho \), and \( N_{\text{C}_3\text{H}_6} \) is
consistent with the above univariate analysis, while the correlation between PLD and \( N_{\text{C}_3\text{H}_6} \) is
opposite, which indicates that the multilinear regression model cannot explain the influence
of PLD on \( N_{\text{C}_3\text{H}_6} \) well. At 300 K and 5,065 kPa, the importance of descriptors is AV > LCD >
\( \rho \approx \) PLD > ASA, which is different from that at 50.65 kPa (LCD > \( \rho \) > AV > ASA > PLD).
At high pressure, guest molecules are constantly squeezed into the zeolite window, and the
entire void space is almost completely filled. In this case, AV could reflect the global porosity
characteristics of the framework well among these descriptors. At low pressure, when there is
no strong external force pushing guest molecules into the window, LCD becomes the most
critical descriptor for determining the \( N_{\text{C}_3\text{H}_6} \). These outcomes are largely consistent with the
above univariate analysis, both of which confirm the significance of AV and LCD for high-
and low-pressure gas adsorptions, respectively. From the perspective of multilinear regression
analysis, P descriptor has little influence on $N_{C_3H_6}$ for combined pressures.

**Table S1**

Coefficients of multilinear regression models after input data normalization at different pressures.

| Pressure (kPa) | $\beta_1$-LCD | $\beta_2$-PLD | $\beta_3$-ASA | $\beta_4$-AV | $\beta_5$-$\rho$ | $\beta_6$-$P$ | $\beta_0$ |
|---------------|---------------|---------------|---------------|--------------|----------------|--------------|----------|
| 5.065         | 0.105         | -0.097        | 0.052         | 0.740        | -0.097         |              | 0.141    |
| 101.3         | 0.132         | -0.091        | 0.050         | 0.727        | -0.115         |              | 0.146    |
| 303.9         | 0.210         | -0.083        | 0.082         | 0.636        | -0.158         |              | 0.170    |
| 202.6         | 0.299         | -0.069        | 0.127         | 0.520        | -0.226         |              | 0.213    |
| 101.3         | 0.556         | -0.028        | 0.281         | 0.138        | -0.421         |              | 0.340    |
| 50.65         | 0.679         | -0.020        | 0.395         | -0.414       | -0.469         |              | 0.353    |
| combined      | 0.190         | -0.068        | 0.111         | 0.347        | -0.102         | 0.034        | 0.104    |
**Predicted $N_{C_3H_6}$ by quadratic regression models versus simulated results**

Since multilinear regression models could not clarify non-linear associations between descriptors and $N_{C_3H_6}$, quadratic regression models with the above five structural descriptors were further explored to correlate the relationships between them. Figure S6 demonstrates predicted $N_{C_3H_6}$ by quadratic regression at 300 K, 5,065 and 50.65 kPa, which is significantly superior to that of multilinear regression with $R^2 = 0.881$ at 5,065 kPa and $R^2 = 0.732$ at 50.65 kPa on the test set due to the involvement of complex binary interaction terms. Fortunately, the accuracy of the quadratic regression model is improved by 31.2% for 50.65 kPa on the test set, while the improvement is only 10.1% for 5,065 kPa, compared with the multilinear regression model. The prediction results of quadratic regression models at other pressures (1,013, 303.9, 202.6, and 101.3 kPa) are displayed in Figure S7. Resembling the previous analysis, the prediction performance of the model gets worse when the adsorption pressure is reduced, indicating that simple structural descriptors that cannot capture local porosity properties of the framework are unable to well reveal the adsorption behavior at low pressure. Given the combined pressures, Figure S8 exhibits improved prediction performance with the introduction of binary terms for combined pressures compared to multilinear regression models, but still fails to predict a high $N_{C_3H_6}$.

**Figure S6.** Comparison of quadratic regression model predictions using five structural descriptors (LCD, PLD, ASA, AV, and $\rho$) with $N_{C_3H_6}$ obtained by GCMC simulation on the
training (blue symbols) and test (orange symbols) sets at 300 K, 5,065 kPa (a) and 50.65 kPa (b).

**Figure S7.** Comparison of quadratic regression model predictions using five structural descriptors (LCD, PLD, ASA, AV, and ρ) with $N\text{C}_3\text{H}_6$ obtained by GCMC simulation on the training (blue symbols) and test (orange symbols) sets at 300 K, 1,013 kPa (a), 303.9 kPa (b), 202.6 kPa (c), and 101.3 kPa (d).
Figure S8. Comparison of quadratic regression model predictions using five structural descriptors (LCD, PLD, ASA, AV, and ρ) and P with combined $N_{C_3H_6}$ obtained by GCMC simulation on the training (blue symbols) and test (orange symbols) sets at six different pressures.

Table S2 illustrates a summary of coefficients of quadratic regression models at 300 K, 5,065 kPa, 50.65 kPa and combined pressures. It can be seen from the table that the absolute value of binary coefficient of AV and LCD or AV is large, implying that AV plays a crucial role in $N_{C_3H_6}$ at 5,065 kPa, which is consistent with the above outcomes of multilinear regression and univariate analysis. At 50.65 kPa, the influence of five structural descriptors to $N_{C_3H_6}$ is difficult to distinguish, which is somewhat different from the previous analysis. In view of combined pressures, the slight impact of P descriptor does not improve the prediction performance obviously. For simplicity, certain terms with small coefficients (less than 5% of the maximum at certain pressure) could be ignored, such as $\beta_{22}$ and $\beta_1$ at 5,065 kPa, $\beta_{11}$, $\beta_{12}$, $\beta_{14}$, $\beta_{22}$, and $\beta_1$ at 50.65 kPa, $\beta_{22}$, $\beta_{26}$, $\beta_{36}$, $\beta_{55}$, and $\beta_{56}$ and at combined pressures.
Table S2
Coefficients of quadratic regression models after input data normalization at 300 K, 5,065 kPa, 50.65 kPa and combined pressures.

| Pressure (kPa) | β₀-LCD | β₂-PLD | β₃-ASA | β₄-AV | β₅-ρ  |
|----------------|---------|---------|---------|-------|-------|
| 5,065          | -0.401  | -0.293  | -1.037  | 2.947 | 0.926 |
| 50.65          | 0.128   | 0.761   | 0.314   | 0.645 |       |
| Linear         | -2.244  | -0.69   |         |       |       |
| Constant       | 0.584   |         |         |       |       |

| Pressure (kPa) | β₀-LCD | β₂-PLD | β₃-ASA | β₄-AV | β₅-ρ  | β₆-P  |
|----------------|---------|---------|---------|-------|-------|-------|
| 50.65          | 0.029   | 0.469   | -1.124  | -0.393| 1.443 |
| 50.65 combined | 0.055   | 3.721   | -2.559  | 2.793 |       |       |
| Linear         | 2.716   | -6.22   | -1.616  | -9.721| 1.437 |
| Constant       | -0.26   | -2.676  | -4.032  | 10.373| -3.151|       |

| Pressure (kPa) | β₀-LCD | β₂-PLD | β₃-ASA | β₄-AV | β₅-ρ  | β₆-P  |
|----------------|---------|---------|---------|-------|-------|-------|
| 5,065          | -0.324  | -0.113  | -0.942  | 1.742 | 0.448 | -0.101|
| 50.65          | 0.077   | 1.208   | -0.282  | 0.847 | 0.977 | -0.083|
| Combined       | 0.825   | -0.99   | -1.351  | -1.991| 0.598 |       |
| Linear         | -1.351  | -1.991  | 0.076   | 0.006 |       |       |
| Constant       | 0.253   | -0.838  | 0.253   | -0.838| -0.791| 1.884 |

| Pressure (kPa) | β₀-LCD | β₂-PLD | β₃-ASA | β₄-AV | β₅-ρ  | β₆-P  |
|----------------|---------|---------|---------|-------|-------|-------|
| 5,065          | -0.324  | -0.113  | -0.942  | 1.742 | 0.448 | -0.101|
| 50.65          | 0.077   | 1.208   | -0.282  | 0.847 | 0.977 | -0.083|
| Combined       | 0.825   | -0.99   | -1.351  | -1.991| 0.598 |       |
| Linear         | -1.351  | -1.991  | 0.076   | 0.006 |       |       |
| Constant       | 0.253   | -0.838  | 0.253   | -0.838| -0.791| 1.884 |
$R^2$ on the training and test sets of multilinear and quadratic regression models at various pressures is summarized in Table S3.

**Table S3**

$R^2$ on the training and test sets of multilinear and quadratic regression models with five structural descriptors (LCD, PLD, ASA, AV, and $\rho$) at various pressures.

| Pressure (kPa) | Multilinear train | Multilinear test | Quadratic train | Quadratic test |
|---------------|-------------------|------------------|-----------------|----------------|
| 5.065         | 0.844             | 0.800            | 0.922           | 0.881          |
| 1.013         | 0.832             | 0.770            | 0.904           | 0.855          |
| 303.9         | 0.794             | 0.732            | 0.868           | 0.825          |
| 202.6         | 0.754             | 0.697            | 0.843           | 0.808          |
| 101.3         | 0.654             | 0.631            | 0.780           | 0.783          |
| 50.65         | 0.552             | 0.558            | 0.729           | 0.732          |
| combined      | 0.713             | 0.722            | 0.822           | 0.803          |
Predicted $N_{C_3H_6}$ by ANN models with six descriptors (LCD, PLD, ASA, AV, $\rho$, and PSD) versus simulated results

**Figure S9.** Comparison of ANN model predictions using six structural descriptors (LCD, PLD, ASA, AV, $\rho$, and PSD) with $N_{C_3H_6}$ obtained by GCMC simulation (fivefold CV approach) at 300 K, 5,065 kPa (a), 1,013 kPa (b), 303.9 kPa (c), 202.6 kPa (d), 101.3 kPa (e), and 50.65 kPa (f).
**Table S4**
LJ parameters $\sigma$ and $\varepsilon$ for C$_3$H$_6$ from the TraPPE-UA force field$^1$ and O atom for zeolites from a transferable force field$^2$. C$_3$H$_6$ molecule consists of three components CH$_3$, CH$_2$, and CH.

| Molecule | Group     | $\sigma$ (Å) | $\varepsilon/k$ (K) |
|----------|-----------|--------------|---------------------|
| C$_3$H$_6$ | CH$_3$ (sp$^3$) | 3.76         | 108                 |
|          | CH$_2$ (sp$^2$) | 3.68         | 85                  |
|          | CH (sp$^2$)   | 3.73         | 47                  |
| Zeolite  | O         | 3            | 93.53               |
Predicted $N_{C_3H_6}$ by ANN models with six descriptors (LCD, PLD, ASA, AV, $\rho$, and $Q_{st}$) versus simulated results

Figure S10. 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, 5,065 kPa (a), 1,013 kPa (b), 303.9 kPa (c), 202.6 kPa (d), and 101.3 kPa (e).
**Figure S11.** (a) Comparison of ANN model predictions using five structural descriptors (LCD, PLD, ASA, AV, and $\rho$) and an energy descriptor ($Q_{st}$), and P with combined $N_{C_{3}H_{6}}$ obtained by GCMC simulation (fivefold CV approach) at six different pressures; (b) residual statistical histogram for all sets.

**Table S5**

$R^2$ on the training and test sets of three ANN models (random validation) at various pressures with S representing the general structural descriptor.

| Pressure (kPa) | ANN (5S) train | ANN (5S) test | ANN (5S + PSD) train | ANN (5S + PSD) test | ANN (5S + $Q_{st}$) train | ANN (5S + $Q_{st}$) test |
|----------------|---------------|---------------|----------------------|---------------------|---------------------------|---------------------------|
| 5.065          | 0.950         | 0.907         | 0.963                | 0.909               | 0.938                      | 0.931                      |
| 1.013          | 0.927         | 0.878         | 0.958                | 0.893               | 0.927                      | 0.906                      |
| 303.9          | 0.865         | 0.840         | 0.923                | 0.855               | 0.881                      | 0.880                      |
| 202.6          | 0.826         | 0.807         | 0.896                | 0.817               | 0.840                      | 0.846                      |
| 101.3          | 0.820         | 0.789         | 0.870                | 0.779               | 0.832                      | 0.847                      |
| 50.65          | 0.809         | 0.752         | 0.862                | 0.726               | 0.804                      | 0.816                      |
| combined       | 0.920         | 0.877         | 0.987                | 0.974               | 0.927                      | 0.903                      |
References

(1) 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.

(2) Pascual, P.; Ungerer, P.; Tavitian, B.; Pernot, P.; Boutin, A. Development of a transferable guest-host force field for adsorption of hydrocarbons in zeolites I. Reinvestigation of alkane adsorption in silicalite by grand canonical Monte Carlo simulation. *Phys. Chem. Chem. Phys.* **2003**, *5*, 3684–3693.