Journal of Materials Chemistry A

ROYAL SOCIETY OF CHEMISTRY

PAPER

View Article Online
View Journal | View Issue

Open Access Article. Published on 23 June 2023. Downloaded on 3/3/2026 11:13:59 AM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

Check for updates

Cite this: J. Mater. Chem. A, 2023, 11, 14788

Advancing CH_4/H_2 separation with covalent organic frameworks by combining molecular simulations and machine learning†‡

Gokhan Onder Aksu and Seda Keskin *

A high-throughput computational screening approach combined with machine learning (ML) was introduced to unlock the potential of both synthesized and hypothetical COFs (hypoCOFs) for adsorption-based CH_4/H_2 separation. We studied 597 synthesized COFs for adsorption of a CH_4/H_2 mixture using Grand Canonical Monte Carlo (GCMC) simulations under pressure-swing adsorption (PSA) and vacuum-swing adsorption (VSA) conditions. Based on the simulation results, the CH_4/H_2 selectivities, CH_4 working capacities, adsorbent performance scores, and regenerabilities of the synthesized COFs were assessed and the structural properties of the top-performing COFs were identified. The hypoCOF database composed of 69 840 materials was then filtered to identify 7737 hypothetical materials having similar structural properties to the top synthesized COFs. These hypothetical COFs were then examined for CH_4/H_2 separation using molecular simulations and the results showed that the top hypoCOFs have CH_4 selectivities and working capacities in the ranges of 21.9–28.7 (64.7–128.6) and 5.8–7.6 (1.3–3.1) mol kg^{-1} under PSA (VSA) conditions, respectively, outperforming the synthesized COFs and metal–organic frameworks (MOFs). ML models were then developed based on the hypoCOF simulation results to accurately predict the CH_4/H_2 mixture adsorption properties of all remaining hypothetical materials when their structural and chemical properties are fed into the models. These models accurately assessed the CH_4/H_2 mixture separation performances of any hypoCOF within seconds without performing computationally demanding molecular simulations. The computational approach that we have proposed in this study will provide an accurate and efficient assessment of COF materials for CH_4/H_2 separation and significantly accelerate the experimental efforts towards the design and discovery of new high-performing COF adsorbents.

Received 24th April 2023
Accepted 5th June 2023

DOI: 10.1039/d3ta02433d

rsc.li/materials-a

1. Introduction

Separating methane (CH_4) from hydrogen (H_2) is important in the refinery industry for pure H_2 recovery.[1] Pressure swing adsorption (PSA) and vacuum swing adsorption (VSA) are energy efficient CH_4/H_2 separation methods compared to traditional cryogenic separation and chemical absorption which suffer from high energy costs.[2,3] Porous adsorbents offering high selectivities and high working capacities are needed to achieve efficient and economical separation of the CH_4/H_2 mixture.[4] Traditional adsorbents, such as single-walled carbon nanotubes (SWNTs),[5] zeolites,[6] and activated carbons,[7] have been widely studied for CH_4/H_2 separation for PSA and VSA processes. Covalent organic frameworks (COFs) have recently emerged as one of the potential adsorbent candidates because of their high porosities, large surface areas, low densities, and high mechanical and chemical stabilities.[8,9] Thanks to these features and the wide variety in their chemistry, COFs have been examined for various applications including gas storage, gas separation, catalysis, and energy storage.[10]

The number of COFs that have been experimentally reported is rapidly increasing, and it is impossible to study all COFs using purely experimental trial-and-error methods to identify the best adsorbents among thousands of candidates. High-throughput computational screening (HTCS) of a large number of materials *via* Grand Canonical Monte Carlo (GCMC) simulations is highly useful to assess the gas adsorption

Department of Chemical and Biological Engineering, Koc University, Rumelifeneri Yolu, Sariyer, 34450, Istanbul, Turkey. E-mail: skeskin@ku.edu.tr; Tel: +90 212 338 1362

† ML scripts are available at https://github.com/gokhanonderaksu/COFS_CH4H2_ML.

‡ Electronic supplementary information (ESI) available: *R*%–APS relations of CoRE COFs; topology distributions among the top CoRE COFs and hypoCOFs; bond and linker type distributions of the top hypoCOFs; schematic representations of the most frequent linker types in the top hypoCOFs; a snapshot showing the adsorption of the gas mixture in the best hypoCOFs; correlation matrix between the structural and chemical properties of the trained hypoCOF set; comparisons of the predicted CH_4 and H_2 uptakes of trained hypoCOFs by ML models constructed with group A, B, and C descriptors and by simulations; feature importance distributions for group C models; comparisons of predicted and simulated CH_4 and H_2 uptakes, S_{CH_4/H_2}, and APSs of unseen hypoCOFs using original and extended ML models. See DOI: https://doi.org/10.1039/d3ta02433d

 This journal is © The Royal Society of Chemistry 2023

Open Access Article. Published on 23 June 2023. Downloaded on 3/3/2026 11:13:59 AM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

potentials of new materials in a time-efficient way and to guide the experimental efforts to the most promising of the many materials.[11–16] HTCS of COFs has accelerated after the introduction of two computation-ready COF databases which provide the simulation-ready crystal structures of synthesized COFs: The Computation-Ready Experimental COF (CoRE COF) database[17–19] consists of 613 different types of COFs, and the Clean, Uniform, and Refined with Automatic Tracking from Experimental Database (CURATED COFs)[20,21] is composed of 648 different types of structurally optimized COFs.

Synthesized COF databases have been computationally screened using GCMC for various separation applications. Tong *et al.*[17] evaluated 187 CoRE COFs for adsorption-based noble gas separations under PSA and VSA conditions and showed that COFs can achieve high Kr/Ar, Xe/Kr, and Rn/Xe adsorption selectivities. Lan *et al.*[22] screened the same number of CoRE COFs for iodine and methyl iodide capture, and revealed that COFs can outperform several traditional adsorbents such as activated carbons, alumina, and zeolites, showing very high iodine capacities. Yan *et al.*[19] screened 298 CoRE COFs for membrane-based CO_2/CH_4 separation at 10 bar and 298 K and showed that the presence of fluorine and chlorine groups improves the membrane selectivities of COFs. Ongari *et al.*[20] screened 296 CURATED COFs for CO_2/N_2 separation for a pressure–temperature swing adsorption (PTSA) process and concluded that the best COF adsorbents have the lowest parasitic energy. Our group studied CoRE COF and CURATED COF databases for CO_2/H_2,[23] CO_2/N_2,[24] CH_4/H_2, CH_4/N_2 and C_2H_6/CH_4[25] separations for both adsorption- and membrane-based processes combining GCMC and MD simulations. Results showed that COFs having narrow pores (<15 Å) achieve better performances for selective gas separation.

In addition to the synthesized COFs, a hypothetical COF (hypoCOF) database, composed of computer-generated but not yet synthesized materials, was established to expand the materials space. HTCS studies have been used to identify the hypothetical materials that can outperform the experimentally synthesized ones. Smit and co-workers constructed and screened a hypoCOF database consisting of 69 840 materials for CH_4 storage and revealed that 304 hypoCOFs achieve higher deliverable CH_4 capacities (>190 m^3 STP per m^3 adsorbent) at 65 bar compared to traditional adsorbents.[26] They further explored the same database for CO_2/N_2 separation under PTSA conditions and discovered that almost 400 hypoCOFs exhibit parasitic energies lower than that acquired for the traditional amine scrubbing process of CO_2 capture and 72 hypoCOFs achieve higher CO_2 working capacities than a well-known synthesized metal–organic framework (MOF), Mg-MOF-74 (0.05 kg CO_2 per kg adsorbent).[27] Our group explored the same hypoCOF database for adsorption-based CO_2/H_2 separation under PSA and VSA conditions, and for membrane-based H_2/CO_2 separation at 10 bar and 298 K, and revealed that hypoCOFs can achieve higher CO_2/H_2 adsorption selectivities (up to 954) and H_2/CO_2 membrane selectivities (up to 6.2) compared to the synthesized COF adsorbents and membranes.[28] We recently screened the same database together with CURATED COFs for adsorption-based removal of H_2S and CO_2 from a natural gas mixture and showed that many synthesized and hypothetical COFs achieve high selectivities up to 12.4 (8.5) under PSA (VSA) conditions, outperforming MOFs, zeolites, and SWNTs.[29] Very recently, Van Speybroeck and co-workers constructed a new hypothetical COF database consisting of 268 687 materials and showed that COFs can achieve similar deliverable volumetric CH_4 capacities to the best reported MOFs, such as MOF-5 (182 cm^3 STP per cm^3) and HKUST-1 (183 cm^3 STP per cm^3), in between 5.8 and 65 bar, at 298 K.[30]

As this literature search shows, HTCS studies using molecular simulations have unlocked the gas adsorption and separation potentials of synthesized COFs and some hypoCOFs. However, using the HTCS approach to study COFs is becoming challenging since the total number of experimentally reported and hypothetically constructed COFs is increasing very rapidly, almost daily. Machine learning (ML) methods have been useful to analyse the huge amount of materials' data obtained from HTCS for establishing the relations between the structural and chemical properties of materials and their performances in different applications.[31–34] ML methods have been adapted to MOFs,[35–37] and very recently to COFs for gas storage and separation.[38,39] For example, Pardakhti *et al.*[40] used ML algorithms to predict the CH_4 storage capacities of 69 839 hypoCOFs together with 17 846 porous polymer networks (PPNs), and showed that using chemical and structural properties as inputs of an ML algorithm leads to accurate CH_4 uptake predictions. Fanourgakis *et al.*[41] identified the best hypoCOFs for CH_4 storage among 69 840 hypoCOFs using a self-consistent ML approach to decrease the computational cost of molecular simulations. Cao *et al.*[42] combined ML algorithms and molecular simulations to predict the adsorption-based C_2H_6/C_2H_4 separation performance of CoRE COFs and hypoCOFs. They concluded that only two COFs have C_2H_6/C_2H_4 selectivities larger than 2, and the most C_2H_6 selective hypoCOF achieved a selectivity of ~45. The same group also constructed ML models to predict the i-C_4H_8 permeability and membrane selectivity of experimental COFs for i-C_4H_8/C_4H_6 mixtures, and showed that pore size and porosity are the key factors determining the separation performance of the membranes.[43]

There is a strong need for an accurate and efficient approach that can unlock the potential of both experimentally reported and hypothetically generated COFs for CH_4/H_2 separation. Motivated by this, we present a multi-level computational approach combining molecular simulations and ML algorithms to assess the CH_4/H_2 mixture adsorption and separation performance of all synthesized and hypothetical COFs under PSA and VSA conditions. We first computed CH_4/H_2 mixture adsorption for CoRE COFs using GCMC simulations and identified the top CoRE COFs by calculating various adsorbent performance metrics based on the simulation results. The structural properties of these top CoRE COFs were then used to filter 7737 potentially promising hypoCOFs, which were then studied by GCMC simulations for CH_4/H_2 separation. We then developed ML models that accurately predicted the CH_4/H_2 mixture adsorption data of 7737 hypoCOFs when their structural and chemical features were input into the models. These ML models were finally used to predict the CH_4/H_2 mixture

Open Access Article. Published on 23 June 2023. Downloaded on 3/3/2026 11:13:59 AM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

adsorption and separation performances of the whole hypoCOF database consisting of 69 840 materials. The top-performing hypoCOFs offering the highest adsorbent performance scores (product of selectivity and working capacity) together with high regenerabilities were identified. Our computational approach will be very useful (i) to evaluate the CH_4/H_2 mixture adsorption and separation potentials of any hypothetical COF within seconds without the need for performing computationally demanding molecular simulations, and (ii) to reveal the structural and chemical features of the best adsorbents which will accelerate the experimental efforts towards the design and development of new COF materials that can achieve high-performance gas separations.

2. Computational details

Our computational methodology combining molecular simulations and ML to examine the adsorption-based gas separation performances of COFs and hypoCOFs is summarized in Fig. 1. We focused on the latest version of CoRE COF[17] and the latest version of hypoCOF databases[26] which include 613 experimentally synthesized and 69 840 computer-generated structures, respectively. The structural features of all materials, pore limiting diameter (PLD), largest cavity diameter (LCD), accessible surface area (S_{acc}), density (ρ), and porosity (ϕ), were computed using Zeo++ software (version 0.3).[44] We eliminated the materials having PLDs less than 3.8 Å and zero S_{acc} so that both CH_4 and H_2 molecules can be adsorbed in the pores. After these eliminations, we had 597 CoRE COFs and 69 828 hypoCOFs.

We focused on CH_4/H_2 adsorption separation in the PSA and VSA processes, as the adsorption (desorption) pressure was configured at 10 (1) and 1 (0.1) bar, respectively, while maintaining a temperature of 298 K. GCMC simulations were performed to compute the adsorption of an equimolar CH_4/H_2 mixture at 0.1, 1, and 10 bar and 298 K using the RASPA software.[45] COF–gas and gas–gas dispersion interactions were described with the Lennard-Jones 12-6 (LJ) potentials. The DREIDING force field was used to define framework atoms.[46] CH_4 was defined by TraPPE[47] and H_2 was defined by Buch

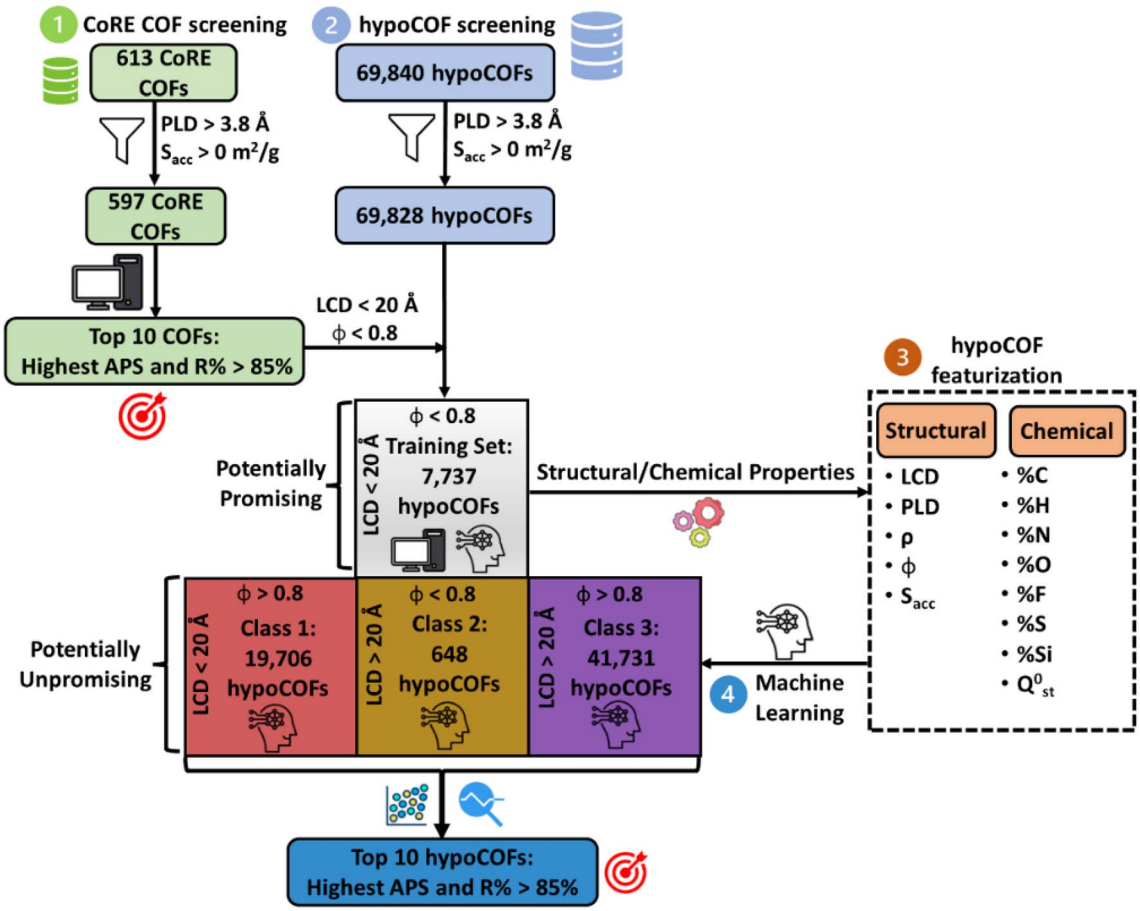

Fig. 1 Computational approach combining molecular simulations and ML to evaluate CoRE COFs and hypoCOFs for CH_4/H_2 separation: (1) screening of CoRE COFs using molecular simulations and identification of the top CoRE COFs based on simulation results, (2) screening of hypoCOFs based on the structural properties of the top CoRE COFs and molecular simulations of these potentially promising hypoCOFs, (3) featurization of structural and chemical properties of the potentially promising hypoCOFs, (4) development of ML models that use these features as input and transfer of the developed ML models to unseen hypoCOFs to identify the top hypoCOFs among 62 085 materials without the need for performing molecular simulations for every single material.

Open Access Article. Published on 23 June 2023. Downloaded on 3/3/2026 11:13:59 AM. This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

potentials.[48] Lorentz–Berthelot mixing rules were used to estimate the interactions between non-identical atoms. In GCMC simulations, we used 10 000 cycles for initialization and 20 000 cycles for taking the ensemble averages. We also computed the heats of adsorption of CH_4 and H_2 gases at infinite dilution by using the Widom insertion method.[49] Based on the mixture gas uptake results obtained from GCMC simulations, adsorbent performance evaluation metrics: adsorption selectivity (S_{CH_4/H_2}), working capacity (ΔN_{CH_4}), adsorbent performance score (APS), and percent regenerability (R%) were calculated as shown in Table S1 of the ESI.‡ APS is a metric combining both selectivity and working capacity, and it should be high for an efficient adsorbent. High R% is another requirement for cyclic usage of adsorbents to have an efficient separation. Thus, to find the most promising materials, all COFs having R% > 85% were ranked based on their calculated APSs and the top 10 COF adsorbents with the highest APSs were identified.

Due to the large materials space of the hypoCOF database consisting of 69 840 materials, performing brute-force molecular simulations for every single material would require very long computational time and sources. To tackle this problem, we first identified the structural properties of the top 10 CoRE COFs and then screened the hypoCOF database to find the materials having similar structural features to the top COFs. With this approach, the hypoCOFs having the potential to offer the highest separation performance were further explored by performing molecular simulations. The top CoRE COFs were found to have LCD < 20 Å and ϕ < 0.80. 7743 hypoCOFs with LCD < 20 Å and ϕ < 0.80 were identified among 69 828 hypoCOF materials and GCMC simulations were performed for these materials to compute their CH_4/H_2 mixture adsorption under the same conditions used in the simulations of CoRE COFs. Among these 7743 hypoCOFs, the top 10 adsorbents having R% > 85% and the highest APSs were also identified.

Following this computational strategy, we were able to unlock the CH_4/H_2 separation performances of 597 CoRE COFs and 7743 hypoCOFs. However, there are 62 085 hypoCOFs remaining to be explored in the database. Although we expected them to be potentially unpromising due to their structural properties, there can be outlier materials offering good (or even better) separation performance while exhibiting different structural and chemical features than the top CoRE COFs. To reveal the separation potentials of the remaining 62 085 hypoCOFs, we developed ML models that accurately predict CH_4/H_2 mixture adsorption for all hypoCOFs. We studied a total of 69 822 hypoCOFs consisting of 7737 hypoCOFs in the training set after data cleaning and 62 085 hypoCOFs in the unseen data set.

To establish the most accurate ML models, we first examined the relations between the descriptors of materials. We extracted a total of 13 different descriptors for 7737 hypoCOFs and divided them into three groups as shown in Table S2.‡ There are 5 structural (PLD, LCD, S_{acc}, ϕ, ρ) and 8 chemical features (elemental percentages (% C, % H, % N, % O, % F, % S, % Si) for COFs and isosteric heats of adsorption of CH_4 or H_2). Pearson correlation coefficients (r) were calculated to determine the feature correlations and the correlation matrix showing these values is provided in Fig. S1.‡ Group A includes only structural properties which were all calculated using Zeo++ for all materials, group B represents both structural and chemical properties, and group C includes all properties except PLD and ρ because these two parameters are highly correlated ($|r|$ > 0.8) with LCD and ϕ. To avoid overtraining, we eliminated these two parameters while constructing our models corresponding to group C descriptors.[50]

ML models were trained using the simulated gas adsorption results of 7737 hypoCOFs as the target data and three groups of descriptors as the input data, as shown in Table S2.‡ We used the tree-based pipeline optimization tool (TPOT) in auto-machine learning to identify the best ML algorithms and optimize the model parameters.[51] For the model selection in TPOT, the regression algorithms in the scikit-learn toolkit[52] were used. To keep the feature distribution in training and test data as uniform as feasible, a stratified sampling technique was used: 80% of the data served as a training set while 20% served as a test set. To prevent overfitting, we additionally performed a five-fold cross-validation. The accuracies of ML models were evaluated by using the coefficient of determination (R^2), mean absolute percentage error (MAPE), and root-mean square error (RMSE), which are all given in Table S3.‡ Several regressor models as shown in Tables S4–S6‡ such as the Extra Tree,[53] GradientBoost,[54] XG-Boost,[55] Random Forest,[42] and Lasso-LarsCV[56] were selected based on their accuracies to predict CH_4 and H_2 uptakes as will be discussed in the following sections.

To test the transferability of the ML models that we developed, three different hypoCOF subsets representing the remaining 62 085 hypoCOFs were generated as shown in Fig. 1 and used as the unseen data. We classified those hypoCOFs whether they have LCD > 20 Å and/or ϕ > 0.80 to isolate the effects of pore sizes and porosities on the separation performance of COFs. Class 1 has 19 706 hypoCOFs (LCD < 20 Å and ϕ > 0.80), Class 2 has 648 hypoCOFs (LCD > 20 Å and ϕ < 0.80), and Class 3 has 41 731 hypoCOFs (LCD > 20 Å and ϕ > 0.80). We randomly selected 1971, 648, and 4174 materials from Class 1, Class 2, and Class 3, respectively, used our ML models to predict their CH_4 and H_2 adsorption data, and compared these ML-predicted values with the simulated ones, to further validate the transferability of ML models. Finally, we used these ML models to unlock the separation potentials of all available 62 085 hypoCOFs, calculated their S_{CH_4/H_2}, ΔN_{CH_4}, APS, and R% using ML-predicted CH_4 and H_2 uptakes, and identified the top 10 hypoCOFs with the highest APSs and R% > 85%.

3. Results and discussion

3.1 Molecular simulations of CoRE COFs and hypoCOFs

We first evaluated the CH_4/H_2 mixture separation performances of 597 CoRE COFs based on the results of GCMC simulations. The optimal adsorbents for PSA and VSA processes should have both high selectivities and high working capacities. To account for this, the product of selectivity and working capacity, adsorbent performance score (APS), was computed and used as a performance metric to identify the promising candidates for CH_4/H_2 separation. Fig. 2(a and b) show the selectivity (S_{CH_4/H_2}) and working capacity (ΔN_{CH_4}) of CoRE COFs, which were

Open Access Article. Published on 23 June 2023. Downloaded on 3/3/2026 11:13:59 AM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

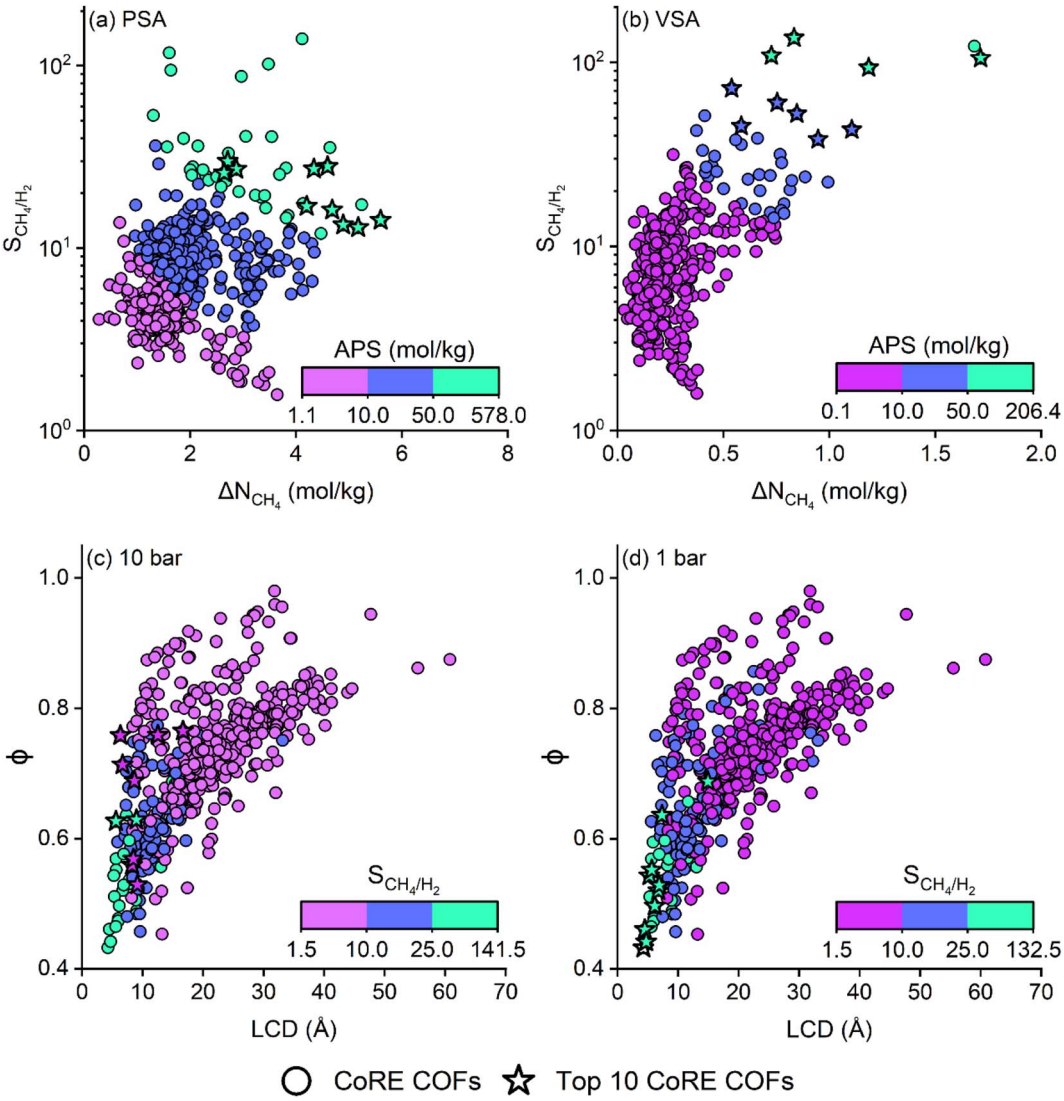

Fig. 2 Calculated S_{CH_4/H_2}, ΔN_{CH_4}, and APS of CoRE COFs for CH_4/H_2:50/50 separation under (a) PSA and (b) VSA conditions. Relation between LCD, ϕ, and S_{CH_4/H_2} of CoRE COFs at (c) 10 bar and (d) 1 bar. Stars represent the top 10 CoRE COFs showing $R\%$ > 85% and the highest APSs.

computed to be 1.6–141.2 (1.6–132.2) and 0.47–5.62 (0.05–1.72) mol kg^{-1} under PSA (VSA) conditions, respectively. COFs with low (<10 mol kg^{-1}), moderate (10–50 mol kg^{-1}), and high (>50 mol kg^{-1}) APSs are shown in Fig. 2(a and b) with pink, blue, and green points, respectively. There are 55 CoRE COFs having high APSs (>50 mol kg^{-1}) under PSA conditions, while there are only 7 COFs with high APSs (>50 mol kg^{-1}) under VSA conditions.

High regenerability ($R\%$) is one of the essential requirements in adsorption-based gas separation processes, but in general, materials having high APSs suffer from low $R\%$.[23,24,28] Fig. S2(a and b)‡ show the relation between $R\%$ and APS of 597 CoRE COFs under PSA and VSA conditions. 512 (561) CoRE COFs were computed to have $R\%$ > 85% under PSA (VSA) conditions. For each process, the top 10 CoRE COF adsorbents were selected among the ones having $R\%$ > 85% and the highest APSs. These top CoRE COFs are shown in Fig. 2(a and b) and listed in Tables S7 and S8‡ with their calculated structural properties and performance metrics. The APSs of the top 10 COFs were computed to be in the ranges of 65.6–129.1 and 26.4–180.6 mol kg^{-1} under PSA and VSA conditions, respectively. Although, COFs have higher APSs under PSA (1.1–578.0 mol kg^{-1}) than under VSA conditions (0.1–206.4 mol kg^{-1}), we observed that the top-performing COFs can achieve much higher APSs under VSA conditions compared to the ones identified under PSA conditions. This can be attributed to the fact that COFs with high APSs may suffer from low regenerabilities under PSA conditions. For example, NPN-2 was identified as the best material under VSA conditions, having an APS of 180.6 mol kg^{-1}. It was also computed to have a high APS of 355.2 mol kg^{-1} under PSA conditions, but it was not identified as a top material due to its low $R\%$ (63.6%).

We also investigated how structural properties affect the separation performance of CoRE COFs as shown in Fig. 2(c and d) where the top CoRE COF adsorbents are shown with stars. At adsorption pressures of 1 and 10 bar, COFs having small pore sizes (5 Å < LCD < 20 Å) and less porous structures (0.4 < ϕ < 0.8)

Open Access Article. Published on 23 June 2023. Downloaded on 3/3/2026 11:13:59 AM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

are high-performing materials as listed in Tables S7 and S8.‡ Narrow pores and low porosities favour the confinement of CH_4 molecules, resulting in high selectivities and APSs. Motivated by these results of CoRE COFs, we filtered the hypoCOF database to identify the potentially promising materials having LCD < 20 Å and ϕ < 0.8.

Fig. 3(a and b) show the calculated S_{CH_4/H_2} and ΔN_{CH_4} of 7737 hypoCOFs which were refined from the large hypoCOF database according to the structural properties of the top CoRE COFs (LCD < 20 Å and ϕ < 0.8). We observed that hypoCOFs can achieve very high APSs, 4.7–641.1 (0.5–473) mol kg^{-1}) in PSA (VSA) processes, outperforming the top 10 CoRE COFs. There are 173 (10) hypoCOFs achieving APS > 129.1 mol kg^{-1} (180.6 mol kg^{-1}), outperforming the best CoRE COF identified for the PSA (VSA) process in Fig. 2(a and b) with a corresponding APS of 129.1 (180.6) mol kg^{-1}. In a previous study of our group,[57] COF-5, COF-6, and COF-10 were studied for the adsorption-based separation of an equimolar CH_4/H_2 mixture and computed to have selectivities and working capacities in the range of 5–19 and 1.1–2.1 mol kg^{-1} under PSA conditions, respectively. Both top-performing CoRE COFs and hypoCOFs outperform these three COFs, suggesting that new materials offering higher separation potential have emerged. Fig. 3(a) also shows a comparison between the top CoRE COFs and the top hypoCOFs that we identified in this work together with the top MOFs identified in our group's previous study[58] for CH_4/H_2 separation. According to the results of our group's previous work,[58] the top 20 MOFs were computed to have S_{CH_4/H_2}, ΔN_{CH_4}, and APS of 22.7–31.2, 3.6–6.3 mol kg^{-1} and 102.9–189.2 mol kg^{-1} under PSA conditions, respectively. The top-performing hypoCOFs have similar selectivities (21.9–28.7) and higher APSs (146.8–205.4 mol kg^{-1}) as shown in Table S9.‡ When we compared the top materials identified under PSA conditions, hypoCOFs outperform both MOFs and CoRE COFs. We also note that MOFs mostly perform better than CoRE COFs, as they achieve higher APSs. The top hypoCOFs have generally higher selectivities (21.9–28.7) than those calculated for CoRE COFs (13–30), but comparable with those of MOFs (21–29). In terms of ΔN_{CH_4}, hypoCOFs were computed to have much higher values (5.82–7.60 mol kg^{-1}) compared to CoRE COFs and MOFs. Therefore, hypoCOFs can achieve much higher APSs, outperforming synthesized COFs and MOFs in CH_4/H_2 separation under PSA conditions. The lists of the top 10 hypoCOFs together with their calculated performance metrics are given in Tables S9 and S10‡ under PSA and VSA conditions. Fig. S3‡ shows snapshots of the CH_4/H_2 mixture adsorption in the best hypoCOFs identified under PSA and VSA conditions. The top hypoCOF adsorbents tend to have much smaller pore sizes (4.3–15.4 Å) and porosities (0.25–0.80) compared to the top CoRE COFs. This result shows that our proposed computational approach for identifying the hypoCOFs having narrow pores and low porosities based on the knowledge obtained from the structural analysis of the top CoRE COFs can be used to accurately find the most promising adsorbents.

We further investigated how structural and chemical variations in hypoCOFs affect their CH_4/H_2 separation performances. In the construction of a hypoCOF database, 111 different linker types and 839 topologies available in the Reticular Chemistry Structure Resource (RCSR)[59] were used.[26] Fig. S4 and S5‡ display the distributions of the topology and linker types that are most prevalent in the top 10 hypoCOFs identified among 7737 materials. The most frequent linker types are linker92 (benzene-based), linker91 (triazine-based), linker108 (pyrene-based), linker110 (adamantane-based) and linker100 (biphenyldiol-based). The corresponding names and structures of the linkers are given in Fig. S6.‡ The benzene and triazine-based linkers of linker92 and linker91 were also found to be among the top materials identified in our previous works

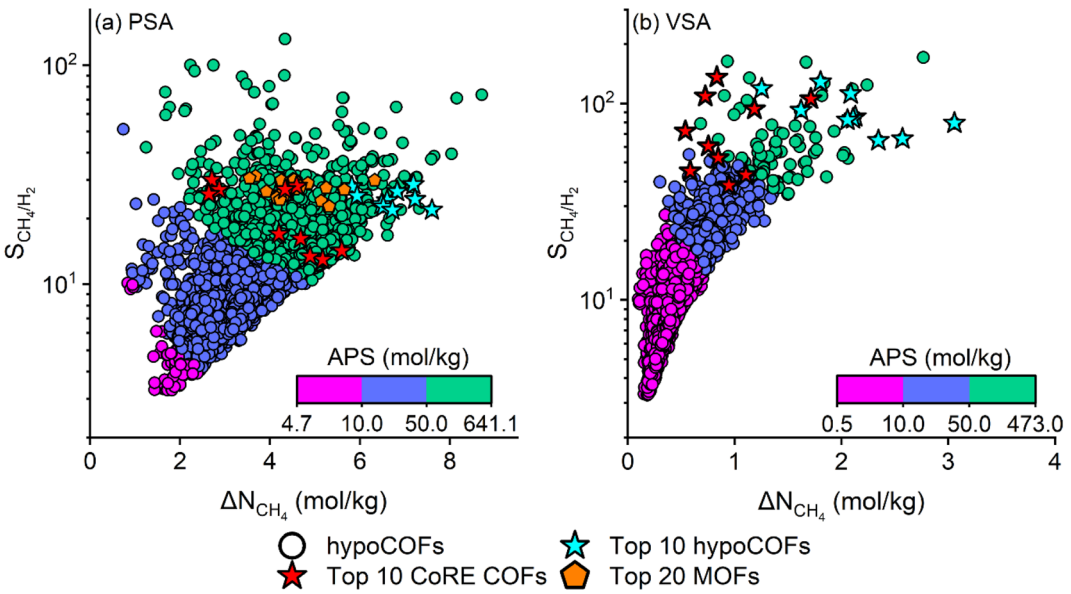

Fig. 3 Relations between S_{CH_4/H_2}, ΔN_{CH_4}, and APS of hypoCOFs for CH_4/H_2:50/50 separation under (a) PSA and (b) VSA conditions. Data of the top 20 MOFs identified in our previous work for CH_4/H_2 separation under PSA conditions was taken from ref. [58] and shown in (a).

Open Access Article. Published on 23 June 2023. Downloaded on 3/3/2026 11:13:59 AM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

related to natural gas purification[29] and adsorption-based CO_2/H_2 separation,[28] and Smit and co-workers' work focusing on flue gas separation under PTSA conditions.[27] In terms of topologies, tbo, lvt, and pts are the most dominant ones among the top hypoCOFs while dia, hcb and sql topologies are also found to be among the best-performing synthesized COFs. The emergence of the same linker types among the top hypoCOF materials identified for different gas separation applications, as well as the observation of the same topologies in the top synthesized and hypothetical COFs will be useful for the design of new COFs with high gas separation potential.

3.2 Development of ML models for hypoCOFs

The aim of developing ML models in this work is to predict the CH_4/H_2 mixture adsorption and separation performances of the whole hypothetical COF database (69 822 materials and even new ones which may be added into the database in the future) within seconds without performing computationally demanding molecular simulations. With this motivation, we developed 15 regression models to predict the adsorption of an equimolar CH_4/H_2 mixture at 0.1, 1, and 10 bar at 298 K. The models were trained using the structural and chemical properties of 7737 hypoCOFs as the input data and their simulated mixture gas adsorption as the target data. Detailed information about ML models is given in Tables S4–S6.‡ We compared the ML-predicted CH_4 and H_2 adsorption with the simulation results and calculated the coefficients of determination (R^2), mean absolute errors (MAE), and root mean square errors (RMSE) (Table S11 and Fig. S7–S9‡) to choose the best model.

First, we examined the models constructed by using group A, B, C descriptors to predict CH_4 uptakes at 0.1, 1, and 10 bar at 298 K. The models developed using group A descriptors were found to have the lowest accuracies and the R^2 values for test sets were calculated to be 0.464, 0.617, and 0.627 at 0.1, 1, and 10 bar, respectively, as shown in Table S11 and Fig. S7(a, c and e).‡ The models developed using group B descriptors were found to accurately predict simulated CH_4 uptakes and the R^2 values for test sets were calculated to be 0.841, 0.885, and 0.870 at 0.1, 1, and 10 bar, respectively, as shown in Fig. S8(a, c and e).‡ The ML models developed using group C descriptors also have good accuracies but they are slightly lower than those of models using group B descriptors, as R^2 for test sets were calculated to be 0.801, 0.856, and 0.829 at 0.1, 1, and 10 bar, respectively, as shown in Fig. S9(a, c and e).‡ Similar trends were also observed for the MAE and RMSE values calculated for each model as listed in Table S11.‡ Due to the existence of highly correlated features in group B descriptors (high correlations were observed between the PLD and LCD ($r = 0.83$), and porosity and density ($r = -0.81$) of hypoCOFs as shown in Fig. S1‡), we inferred that group B descriptors may be biased. Thus, we chose group C descriptors as the optimal ones for predicting CH_4 uptakes. We then compared the H_2 uptake predictions of the models constructed by using group A, B, and C descriptors at 1 and 10 bar, at 298 K. We observed that all models can accurately predict simulated H_2 uptakes leading to R^2 values larger than 0.9 for test sets at each adsorption condition as shown in Table S11.‡ To be consistent in our model selection, we used group C descriptors for H_2 uptake predictions as we did for CH_4 uptake predictions.

The feature importance distributions corresponding to each model constructed by using group C descriptors are given in Fig. S10.‡ As shown in Fig. S10(a–c),‡ the isosteric heat of adsorption for CH_4 is an important descriptor especially at low pressures. In contrast, structural properties such as porosity and surface area were observed to be the main descriptors of H_2 uptake in COFs as Fig. S10(d and e)‡ show. Since H_2 has weaker van der Waals interactions with the COFs than CH_4, the importance of chemical descriptors such as isosteric heat of adsorption may be less important for H_2. We also performed SHapley Additive exPlanations (SHAP) analysis[60] to gain more insights into the impact of the features on the ML predictions. The significance of the isosteric heats of adsorption for CH_4 in our models predicting CH_4 uptake was further confirmed in Fig. S11(a–c).‡ We observed that high CH_4 uptake predictions were associated with high isosteric heats of adsorption for CH_4, low LCDs, while low porosities can lower the CH_4 uptake predictions. Fig. S11(d and e)‡ demonstrate that surface area and porosity are the most important features playing a major role in predicting H_2 uptakes. High H_2 uptake predictions were associated with high values of these features as shown in Fig. S11(d and e).‡

After showing that our ML models can accurately predict CH_4 and H_2 uptakes for 7737 hypoCOFs, we utilized these models to calculate CH_4/H_2 selectivities and APSs. Fig. 4 shows the comparison of ML-predicted and simulated selectivity and APS of 7737 hypoCOFs under PSA and VSA conditions. ML-predicted S_{CH_4/H_2} and APS values are in good agreement with the simulated ones under both PSA and VSA conditions: for S_{CH_4/H_2} (APS) under PSA and VSA conditions, R^2 values for test sets were calculated to be 0.883 (0.847) and 0.868 (0.771), respectively. For example, S_{CH_4/H_2} and APS values were calculated to be in the range of 3.3–123 (3.3–132) and 5.1–567 (4.7–641) mol kg^{-1} by using ML-predicted (simulated) CH_4 and H_2 uptakes, respectively, under the PSA conditions. We demonstrated that our ML models accurately predict the separation performance of 7737 hypoCOFs, which are located within the region defined by the structural properties of the top-performing CoRE COFs. As we discussed, the S_{CH_4/H_2} and APS of the top 10 hypoCOFs were computed to be in the ranges of 21.9–28.7 (64.7–128.7) and 147–205 (149–243) mol kg^{-1} by using simulations under PSA (VSA) conditions. According to the ML-predicted results, the S_{CH_4/H_2} and APS of the same materials were calculated to be in the range of 17.4–27.1 (63.7–118) and 88.8–189 (112–249) mol kg^{-1} under PSA (VSA) conditions. The most prominent finding is that our ML models can find 6 (8) of the top 10 hypoCOFs identified based on the simulation results. Since ML predictions are obtained within seconds compared to molecular simulations which take several weeks, accurate identification of the most promising hypoCOF materials by ML is highly useful.

In our proposed computational approach, we chose the hypoCOFs based on the structural properties of the top CoRE COFs expecting that narrow-pored hypoCOFs can outperform

Open Access Article. Published on 23 June 2023. Downloaded on 3/3/2026 11:13:59 AM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

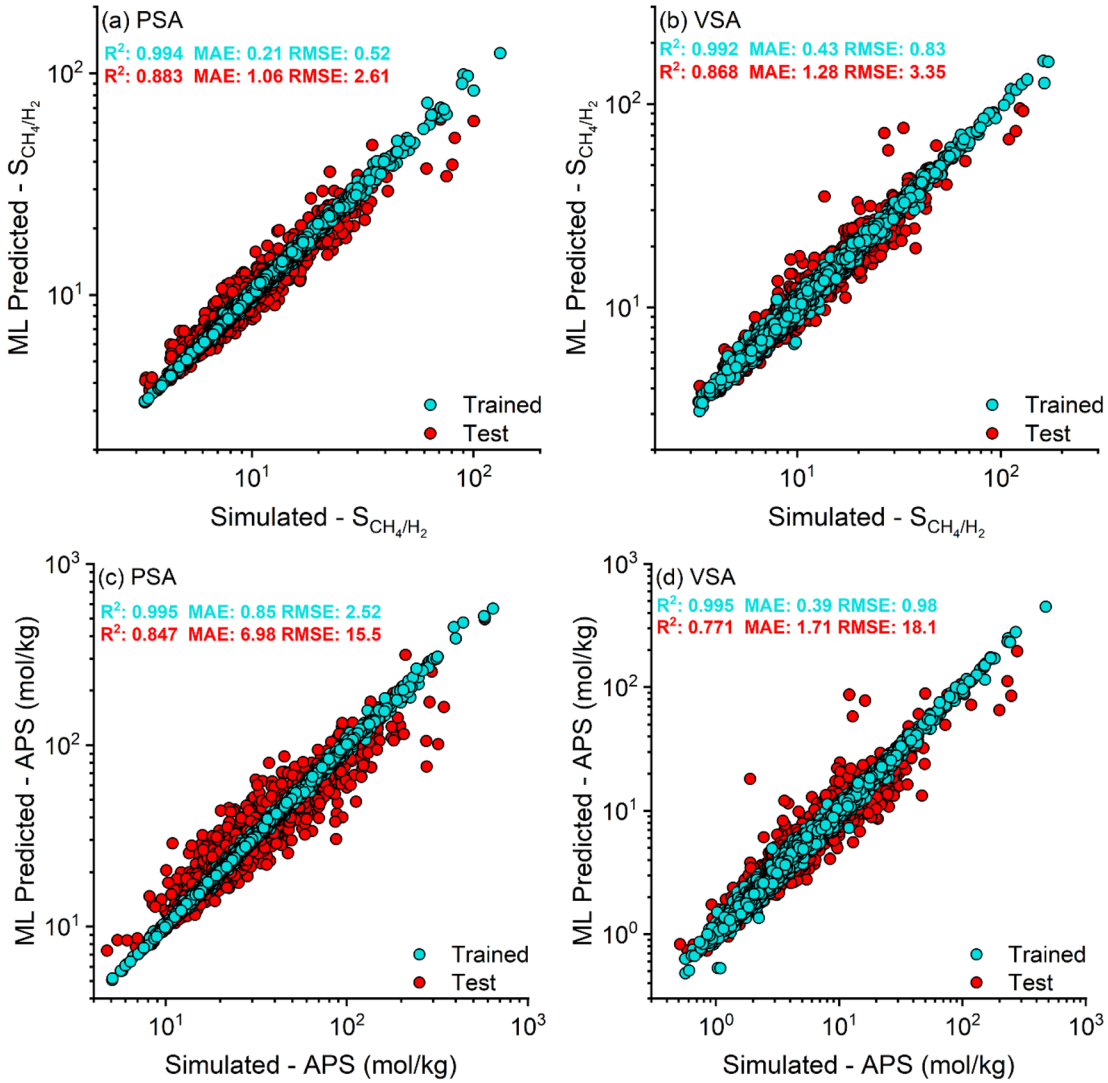

Fig. 4 Comparison of ML-predicted and simulated (a and b) selectivities and (c and d) APSs of 7737 hypoCOFs under PSA and VSA conditions. Blue (red) symbols represent training (test) data.

the materials having larger pores. As we discussed in Fig. 3, our computational approach targeting narrow-pored and low-porosity hypoCOFs was valid and hypoCOFs outperformed both synthesized COFs and MOFs. However, there may be exceptional structures since gas adsorption is a complex interplay between structural properties of the adsorbent and specific chemical interactions of gases with each other in the mixture and with the adsorbent material. To investigate this further, we focused on the hypoCOFs which do not satisfy the structural properties (LCD < 20 Å, ϕ < 0.80) identified for the top CoRE COFs. According to these limits, we specified three potentially unpromising hypoCOF classes: Class 1 (LCD < 20 Å, ϕ > 0.80), Class 2 (LCD > 20 Å, ϕ < 0.80) and Class 3 (LCD > 20 Å, ϕ > 0.80). Considering the computational costs, we sampled 10% of Class 1 and 3, representing 1971 and 4174 materials, respectively, and included all of Class 2 having 648 materials. We then further studied 6793 hypoCOFs as the unseen data, which were not used in the development of ML models.

A comparison of ML-predicted and simulated CH_4 uptakes of the 6793 unseen hypoCOFs is given in Fig. S12.‡ There is a good agreement between ML-predicted and simulated CH_4 uptakes of Class 1 hypoCOFs (Fig. S12(a and b)‡). For example, at 10 (1) bar, ML-predicted and simulated CH_4 uptakes are in the ranges of 2.09–7.79 (0.20–1.29) mol kg^{-1} and 2.13–9.50 (0.23–1.63) mol kg^{-1}, respectively. On the other hand, Fig. S12(d–i)‡ show that ML models that we developed for hypoCOFs having LCD < 20 Å cannot accurately predict CH_4 uptakes of Class 2 and Class 3 hypoCOFs, which both have LCD > 20 Å. For H_2, Fig. S13‡ shows the comparisons between the ML-predicted and simulated uptakes of unseen hypoCOFs. The ML models accurately predicted the H_2 uptakes of Class 2 hypoCOFs (Fig. S13(c and d)‡) but H_2 uptakes of Class 1 and Class 3 hypoCOFs cannot be predicted. Simulated H_2 uptakes of these unseen hypoCOFs have larger values compared to the H_2 uptake ranges used in training models. As shown in Tables S4–S6,‡ our models are based on supervised algorithms, which have limitations in

Open Access Article. Published on 23 June 2023. Downloaded on 3/3/2026 11:13:59 AM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

terms of extrapolation beyond the trained data set, resulting in inaccurate predictions of most of the unseen hypoCOFs. Thus, we decided to extend ML models by randomly selecting 1000 hypoCOFs specifically from Class 3, as it is the only material class having both different pore sizes and porosities than the original training set.

We developed new models by using 8737 hypoCOFs with their group C descriptors. These models with hyperparameters are listed in Table S12.‡ We observed that the extended models make much more accurate predictions for CH_4 uptakes of Class 2 and Class 3 materials under all conditions. With the use of extended models, R^2 values between ML-predicted and simulated CH_4 uptakes of Class 3 hypoCOFs increased from 0.022 (0.135) to 0.799 (0.836) at 10 (1) bar, as shown in Fig. S12(g and h), and S14(g and h),‡ respectively. The same trend is also valid for H_2 uptakes of Class 1 and Class 3 hypoCOFs. R^2 values between ML-predicted and simulated H_2 uptakes at 10 (1) bar for Class 1 were improved from 0.453 (0.320) to 0.984 (0.985), as shown in Fig. S13(a and b) and S15(a and b),‡ respectively. We note that our extended models are still not very good in making highly accurate predictions for two cases: (i) CH_4 uptakes of Class 2 materials at 10 bar, and (ii) H_2 uptakes of a small part of Class 3 materials at 1 and 10 bar. We inferred that our random sampling of unseen hypoCOFs added into the training set was not diverse enough to overcome the extrapolation limitations of regression models for these two cases. For case (i), our extended models can predict CH_4 uptakes of 551 hypoCOFs out of 648 with less than 20% error margin (defined as the difference between the ML-predicted and simulated values divided by the simulated one). For case (ii), H_2 uptakes of 3133 (3131) hypoCOFs out of 3174 Class 3 materials were predicted with less than 20% error at 10 (1) bar.

We then used these extended models to evaluate the CH_4/H_2 separation performance of the unseen hypoCOFs. Fig. 5 shows the APSs of all unseen hypoCOFs calculated from ML-predicted and simulated gas uptakes under VSA conditions. There is a strong agreement between ML and simulation results for APSs of Class 1 and Class 3 hypoCOFs, whereas the agreement is weaker for Class 2. Both Class 1 and Class 2 hypoCOFs perform on par, and they outperform Class 3 hypoCOFs. Narrow-pored Class 1 hypoCOFs and large-pored Class 2 hypoCOFs exhibit comparable performance under VSA conditions which can be attributed to the less pronounced impact of structural properties at low pressures. Fig. S16‡ shows the APSs of all unseen hypoCOFs calculated from ML-predicted and simulated gas uptakes under PSA conditions. Class 1 outperforms Class 2 and Class 3 hypoCOFs by achieving higher APSs. This can be explained by the increased impact of pore sizes in determining selectivities of materials at higher pressures as narrow pores (LCD < 20 Å) of Class 1 materials provide stronger confinement of CH_4 and lead to higher selectivities.

We finally calculated the CH_4/H_2 selectivities and CH_4 working capacities of all 69 822 hypoCOFs under PSA and VSA conditions by utilizing our ML models. To the best of our knowledge, this is the first representation of the CH_4/H_2 separation potential limits of the whole hypoCOF materials space in the literature. Fig. 6 shows that hypoCOFs used in the training of ML models which were specifically selected based on the structural properties of the top CoRE COFs are potentially promising adsorbent materials. They showed higher selectivities than the unseen hypoCOFs (Classes 1, 2, and 3), which were expected to be unpromising back in Fig. 1. For example, under PSA (VSA) conditions, trained hypoCOFs have selectivities between 3.3 and 123 (3.1 and 162.6) while the selectivities of the unseen hypoCOFs are between 1.4 and 36.3 (1.3 and 47.2). This shows the validity of our approach for targeting narrow-pored and low porosity hypoCOFs to find the best-performing adsorbents for CH_4/H_2 separation. We also identified the top materials with the highest APSs, which is an indicator for both high adsorption selectivities and working capacities. As we expected, all of the top 10 materials in hypoCOF materials space belong to our originally trained material set. The top 10 hypoCOFs in our training set were calculated to have ML-predicted APSs in the range of 156.7–268.2 (145.4–450.3) mol kg^{-1} under PSA (VSA) conditions, respectively. All in all, we were able to comprehensively map the CH_4/H_2 separation performance of all 69 822 hypoCOFs utilizing the ML models and our computational screening approach which focused on hypoCOFs having optimal pore sizes and porosities based on the results of experimentally synthesized COFs.

Fig. 6 can be considered as the key outcome of our work since it shows the selectivity and working capacity limits of all

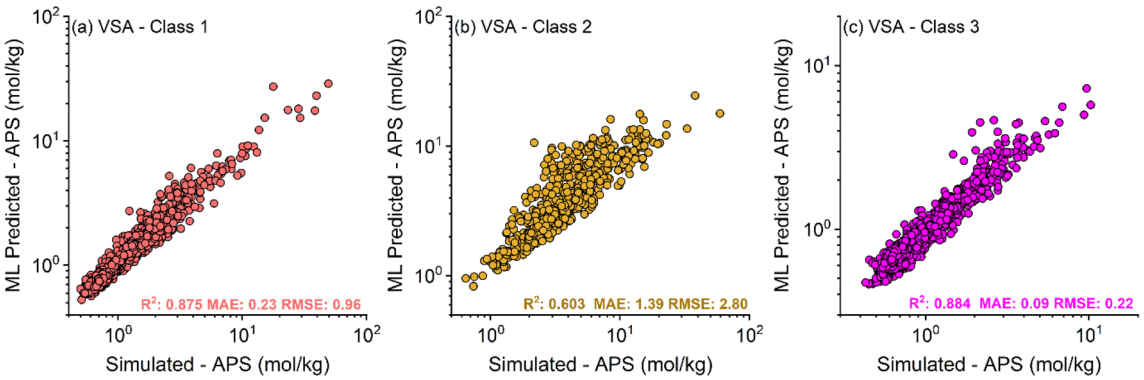

Fig. 5 ML-predicted and simulated APSs for unseen hypoCOFs under VSA conditions for (a) Class 1, (b) Class 2, and (c) Class 3 hypoCOFs.

Open Access Article. Published on 23 June 2023. Downloaded on 3/3/2026 11:13:59 AM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

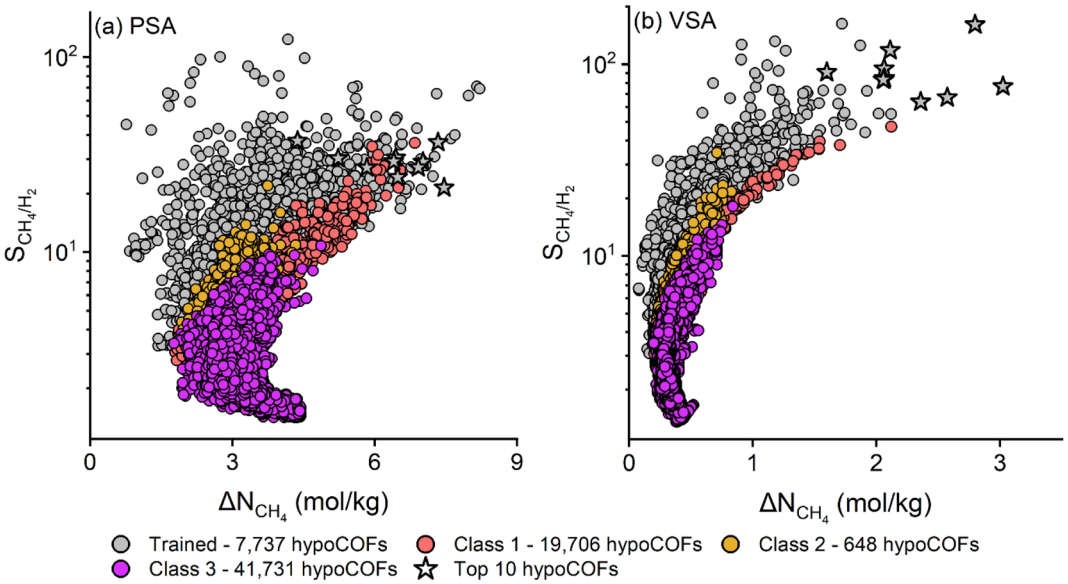

Fig. 6 CH_4/H_2 adsorption performance of the whole hypoCOF materials space predicted by ML models for (a) PSA and (b) VSA processes.

hypoCOFs which would not be feasible to compute by using solely molecular simulations due to the very large number of materials and large unit cell dimensions of COFs which make the computation very time demanding. Generating the CH_4/H_2 separation performance map of 69 822 hypoCOFs became possible when we combined molecular simulations with the ML models. At that point, it is important to note that ML models were developed based on the results of molecular simulations which were performed using classical force fields and rigid framework assumption. Thus, our ML models are as accurate as these assumptions and force fields, and we previously showed their validity by comparing the experimentally reported CH_4 and H_2 adsorption isotherms of several COFs with the simulations using the same force fields and assumptions.[23–25] Finally, it is important to discuss the synthesizability of hypothetical materials generated in the computer environment. The rationale behind constructing a hypoCOF database lies in the potential discovery of new COFs that can be synthesized. The distinguishing feature of the hypoCOF database is the validation of the framework construction approach against experimental structures, such as COF-300 and TAPB-PDA COF, by comparing their experimental powder X-ray diffraction spectra with those computationally generated.[26] Thus, we anticipate that with the recent advancements in the synthesis techniques, some of the promising hypoCOFs can be really synthesized in the future.

4. Conclusion

In this study, a novel approach combining HTCS with ML has been introduced to evaluate the CH_4/H_2 mixture separation performance of both synthesized and hypothetical COFs. We systematically screened nearly 70 000 COFs, which to the best of our knowledge is the largest set of COFs ever evaluated for a gas mixture separation application in the literature. After performing GCMC simulations to study the adsorption of equimolar CH_4/H_2 mixtures for 597 synthesized COFs under both PSA and VSA conditions, we identified the top adsorbents based on their CH_4/H_2 selectivities, CH_4 working capacities, and regenerabilities. The structural properties of the top-performing synthesized COFs were analysed in detail and a hypoCOF database composed of 69 840 materials was filtered to identify 7737 hypothetical materials having similar structural features to the top-performing synthesized COFs. GCMC simulations were performed to evaluate the CH_4/H_2 separation performance of these hypoCOFs and used as the target data to develop ML models that can accurately predict the CH_4/H_2 separation performance of all 62 085 unseen hypoCOF materials within seconds. ML models were shown to successfully identify the top adsorbent materials and many hypoCOFs were shown to outperform both synthesized COFs and MOFs in PSA and VSA-based CH_4/H_2 separation. Our results will be useful not only to completely reveal the separation potential of the whole hypoCOF materials space but also to accelerate the experimental efforts towards the design and discovery of new high-performing COF adsorbents.

Conflicts of interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

S. K. acknowledges the ERC-2017-Starting Grant. This research has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 Research and Innovation Programme (ERC-2017-Starting Grant, grant agreement no. 756489-COSMOS). This work is also supported by the

Open Access Article. Published on 23 June 2023. Downloaded on 3/3/2026 11:13:59 AM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

Scientific and Technological Research Council of Turkey (TUBITAK) under the 1001-Scientific and Technological Research Projects Funding Program (Project Number: 122Z536).

References

1 S. Sircar, W. Waldron, M. Rao and M. Anand, *Sep. Purif. Technol.*, 1999, **17**, 11–20.
2 S. Niaz, T. Manzoor and A. H. Pandith, *Renewable Sustainable Energy Rev.*, 2015, **50**, 457–469.
3 A. Malek and S. Farooq, *AIChE J.*, 1998, **44**, 1985–1992.
4 N. Kumar, S. Mukherjee, N. C. Harvey-Reid, A. A. Bezrukov, K. Tan, V. Martins, M. Vandichel, T. Pham, L. M. van Wyk and K. Oyekan, *Chem*, 2021, **7**, 3085–3098.
5 H. Chen and D. S. Sholl, *J. Membr. Sci.*, 2006, **269**, 152–160.
6 J. C. Poshusta, V. A. Tuan, E. A. Pape, R. D. Noble and J. L. Falconer, *AIChE J.*, 2000, **46**, 779–789.
7 A. M. Morales-Cas, C. Moya, B. Coto, L. F. Vega and G. Calleja, *J. Phys. Chem. C*, 2007, **111**, 6473–6480.
8 A. P. Cote, A. I. Benin, N. W. Ockwig, M. O'Keeffe, A. J. Matzger and O. M. Yaghi, *Science*, 2005, **310**, 1166–1170.
9 H. Furukawa and O. M. Yaghi, *J. Am. Chem. Soc.*, 2009, **131**, 8875–8883.
10 S. Mondal, B. Mohanty, M. Nurhuda, S. Dalapati, R. Jana, M. Addicoat, A. Datta, B. K. Jena and A. Bhaumik, *ACS Catal.*, 2020, **10**, 5623–5630.
11 G. Garberoglio, *Langmuir*, 2007, **23**, 12154–12158.
12 Q. Yang and C. Zhong, *Langmuir*, 2009, **25**, 2302–2308.
13 M. Tong, Q. Yang, Y. Xiao and C. Zhong, *Phys. Chem. Chem. Phys.*, 2014, **16**, 15189–15198.
14 M. Tong, Y. Lan, Q. Yang and C. Zhong, *Green Energy Environ.*, 2018, **3**, 107–119.
15 H. Daglar and S. Keskin, *Coord. Chem. Rev.*, 2020, **422**, 213470.
16 P. G. Boyd, Y. Lee and B. Smit, *Nat. Rev. Mater.*, 2017, **2**, 1–15.
17 M. Tong, Y. Lan, Q. Yang and C. Zhong, *Chem. Eng. Sci.*, 2017, **168**, 456–464.
18 M. Tong, Y. Lan, Z. Qin and C. Zhong, *J. Phys. Chem. C*, 2018, **122**, 13009–13016.
19 T. Yan, Y. Lan, M. Tong and C. Zhong, *ACS Sustainable Chem. Eng.*, 2018, **7**, 1220–1227.
20 D. Ongari, A. V. Yakutovich, L. Talirz and B. Smit, *ACS Cent. Sci.*, 2019, **5**, 1663–1675.
21 D. Ongari, L. Talirz and B. Smit, *ACS Cent. Sci.*, 2020, **6**, 1890–1900.
22 Y. Lan, M. Tong, Q. Yang and C. Zhong, *CrystEngComm*, 2017, **19**, 4920–4926.
23 G. O. Aksu, H. Daglar, C. Altintas and S. Keskin, *J. Phys. Chem. C*, 2020, **124**, 22577–22590.
24 O. F. Altundal, C. Altintas and S. Keskin, *J. Mater. Chem. A*, 2020, **8**, 14609–14623.
25 O. F. Altundal, Z. P. Haslak and S. Keskin, *Ind. Eng. Chem. Res.*, 2021, **60**, 12999–13012.
26 R. Mercado, R.-S. Fu, A. V. Yakutovich, L. Talirz, M. Haranczyk and B. Smit, *Chem. Mater.*, 2018, **30**, 5069–5086.
27 K. S. Deeg, D. Damasceno Borges, D. Ongari, N. Rampal, L. Talirz, A. V. Yakutovich, J. M. Huck and B. Smit, *ACS Appl. Mater. Interfaces*, 2020, **12**, 21559–21568.
28 G. O. Aksu, I. Erucar, Z. P. Haslak and S. Keskin, *Chem. Eng. J.*, 2022, **427**, 131574.
29 G. O. Aksu, I. Erucar, Z. P. Haslak and S. Keskin, *J. CO2 Util.*, 2022, **62**, 102077.
30 J. S. De Vos, S. Borgmans, P. Van Der Voort, S. M. Rogge and V. Van Speybroeck, *J. Mater. Chem. A*, 2023, **11**, 7468–7487.
31 C. Altintas, O. F. Altundal, S. Keskin and R. Yildirim, *J. Chem. Inf. Model.*, 2021, **61**, 2131–2146.
32 H. Demir, H. Daglar, H. C. Gulbalkan, G. O. Aksu and S. Keskin, *Coord. Chem. Rev.*, 2023, **484**, 215112.
33 K. M. Jablonka, D. Ongari, S. M. Moosavi and B. Smit, *Chem. Rev.*, 2020, **120**, 8066–8129.
34 G. H. Gu, J. Noh, I. Kim and Y. Jung, *J. Mater. Chem. A*, 2019, **7**, 17096–17117.
35 P. Krokidas, S. Karozis, S. Moncho, G. Giannakopoulos, E. N. Brothers, M. E. Kainourgiakis, I. G. Economou and T. A. Steriotis, *J. Mater. Chem. A*, 2022, **10**, 13697–13703.
36 Y. Lim, J. Park, S. Lee and J. Kim, *J. Mater. Chem. A*, 2021, **9**, 21175–21183.
37 Z. Shi, X. Yuan, Y. Yan, Y. Tang, J. Li, H. Liang, L. Tong and Z. Qiao, *J. Mater. Chem. A*, 2021, **9**, 7656–7666.
38 Z. Liu, W. Li, S. Cai, Z. Tu, X. Luo and S. Li, *J. Mater. Chem. A*, 2022, **10**, 9604–9611.
39 W. Li, X. Xia and S. Li, *J. Mater. Chem. A*, 2019, **7**, 25010–25019.
40 M. Pardakhti, P. Nanda and R. Srivastava, *J. Phys. Chem. C*, 2020, **124**, 4534–4544.
41 G. S. Fanourgakis, K. Gkagkas, E. Tylianakis and G. Froudakis, *J. Phys. Chem. C*, 2020, **124**, 19639–19648.
42 X. Cao, Z. Zhang, Y. He, W. Xue, H. Huang and C. Zhong, *Ind. Eng. Chem. Res.*, 2022, **61**, 11116–11123.
43 X. Cao, Y. He, Z. Zhang, Y. Sun, Q. Han, Y. Guo and C. Zhong, *Chem. Res. Chin. Univ.*, 2022, **38**, 421–427.
44 T. F. Willems, C. H. Rycroft, M. Kazi, J. C. Meza and M. Haranczyk, *Microporous Mesoporous Mater.*, 2012, **149**, 134–141.
45 D. Dubbeldam, S. Calero, D. E. Ellis and R. Q. Snurr, *Mol. Simul.*, 2016, **42**, 81–101.
46 S. L. Mayo, B. D. Olafson and W. A. Goddard, *J. Phys. Chem.*, 1990, **94**, 8897–8909.
47 M. G. Martin and J. I. Siepmann, *J. Phys. Chem. B*, 1998, **102**, 2569–2577.
48 V. Buch, *J. Chem. Phys.*, 1994, **100**, 7610–7629.
49 D. Frenkel and B. Smit, *Understanding Molecular Simulation: From Algorithms to Applications*, Elsevier, 2001.
50 P. Yang, H. Zhang, X. Lai, K. Wang, Q. Yang and D. Yu, *ACS Omega*, 2021, **6**, 17149–17161.
51 T. T. Le, W. Fu and J. H. Moore, *Bioinformatics*, 2020, **36**, 250–256.
52 P.-G. Martinsson, V. Rokhlin and M. Tygert, *Appl. Comput. Harmon. Anal.*, 2011, **30**, 47–68.
53 S. Meduri and J. Nandanavanam, *Energy and AI*, 2023, 100230.

 This journal is © The Royal Society of Chemistry 2023

Open Access Article. Published on 23 June 2023. Downloaded on 3/3/2026 11:13:59 AM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

54 H. Dureckova, M. Krykunov, M. Z. Aghaji and T. K. Woo, *J. Phys. Chem. C*, 2019, **123**, 4133–4139.
55 H. Liang, K. Jiang, T.-A. Yan and G.-H. Chen, *ACS Omega*, 2021, **6**, 9066–9076.
56 B. J. Bucior, N. S. Bobbitt, T. Islamoglu, S. Goswami, A. Gopalan, T. Yildirim, O. K. Farha, N. Bagheri and R. Q. Snurr, *Mol. Syst. Des. Eng.*, 2019, **4**, 162–174.
57 S. Keskin, *J. Phys. Chem. C*, 2012, **116**, 1772–1779.
58 C. Altintas, I. Erucar and S. Keskin, *ACS Appl. Mater. Interfaces*, 2018, **10**, 3668–3679.
59 M. O'Keeffe, M. A. Peskov, S. J. Ramsden and O. M. Yaghi, *Acc. Chem. Res.*, 2008, **41**, 1782–1789.
60 S. M. Lundberg, G. Erion, H. Chen, A. DeGrave, J. M. Prutkin, B. Nair, R. Katz, J. Himmelfarb, N. Bansal and S.-I. Lee, *Nat. Mach. Intell.*, 2020, **2**, 56–67.