JCTC Journal of Chemical Theory and Computation

pubs.acs.org/JCTC Article

Data Driven Discovery of MOFs for Hydrogen Gas Adsorption

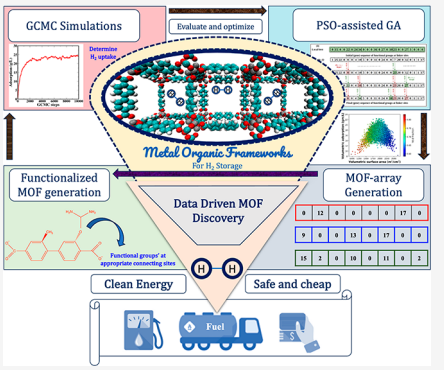
This paper was intended for Machine Learning for Molecular Simulation, published as the July 25, 2023, issue of J. Chem. Theory Comput. (Vol. 19, No. 14).

Samrendra K. Singh, Abhishek T. Sose, Fangxi Wang, Karteek K. Bejagam, and Sanket A. Deshmukh*

Cite This: *J. Chem. Theory Comput.* 2023, 19, 6686−6703 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Hydrogen gas (H_2) is a clean and renewable energy source, but the lack of efficient and cost-effective storage materials is a challenge to its widespread use. Metal−organic frameworks (MOFs), a class of porous materials, have been extensively studied for H_2 storage due to their tunable structural and chemical features. However, the large design space offered by MOFs makes it challenging to select or design appropriate MOFs with a high H_2 storage capacity. To overcome these challenges, we present a data-driven computational approach that systematically designs new functionalized MOFs for H_2 storage. In particular, we showcase the framework of a hybrid particle swarm optimization integrated genetic algorithm, grand canonical Monte Carlo (GCMC) simulations, and our in-house MOF structure generation code to design new MOFs with excellent H_2 uptake. This automated, data driven framework adds appropriate functional groups to IRMOF-10 to improve its H_2 adsorption capacity. A detailed analysis of the top selected MOFs, their adsorption isotherms, and MOF design rules to enhance H_2 adsorption are presented. We found a functionalized IRMOF-10 with an enhanced H_2 adsorption increased by ∼6 times compared to that of pure IRMOF-10 at 1 bar and 77 K. Furthermore, this study also utilizes machine learning and deep learning techniques to analyze a large data set of MOF structures and properties, in order to identify the key factors that influence hydrogen adsorption. The proof-of-concept that uses a machine learning/deep learning approach to predict hydrogen adsorption based on the identified structural and chemical properties of the MOF is demonstrated.



1. INTRODUCTION

In recent years, globally, energy consumption has reached unprecedented proportions, resulting in environmental hazards.[1] A variety of renewable energy sources, including solar, wind, and hydrogen gas (H_2), can assist us in not only meeting these demands of energy consumption but also in addressing environmental concerns.[2] Among all, H_2 is a more appealing fuel candidate, as its combustion produces water as a byproduct.[3] H_2 has potential applications in fuel cell vehicles, backup power for basic infrastructure, portable power devices, and so on.[4−7] Further, the limited availability of H_2 from the atmosphere has resulted in a need to address the increasing challenges associated with H_2 capture, storage, and dispensing. Traditionally, H_2 has been stored in commercial vehicles under high pressures up to 700 bar, which is an expensive and significant safety risk.[8] Thus, the need for an alternative, safe, and adequate solution for the storage and transportation of hydrogen is of utmost importance.

Recently, a number of porous and nonporous solid and liquid materials such as metal−organic frameworks (MOFs), zeolites, carbon nanotubes, complex hydrides, and ionic liquids have been used to achieve high-density H_2 storage to meet the required gravimetric capacity (55 g/kg for light-duty fuel cell vehicles according to the 2025 US Department of Energy (DOE) target) or volumetric capacity (40 g/L) for different applications.[9−15] Specifically, MOFs have attracted an impressive amount of attention due to their adjustable pore size and void fraction, large volumetric surface area, high gravimetric density, and tunable physical and chemical properties.[16−19] At cryogenic temperature (77 K), several MOFs have shown H_2 adsorption close to or better than the 2025 DOE target (5.5 wt %), such as zirconium-based NU-110x series,[20] zinc-based IRMOF-20,[21] DUT-32,[22] etc. Gomez-Gualdron et al. synthesized MOF NU-1101, NU-1102, and NU-1103 to study both the volumetric and gravimetric hydrogen adsorption. The hydrogen adsorptions at 5 bar and 160 K were 9.1 wt % (NU-1101), 9.6 wt % (NU-1102), and 12.6 wt % (NU-1103), which has been attributed to their surface area of 4340 m^2/g (NU-1101), 3720 m^2/g

Received: January 18, 2023
Published: September 27, 2023

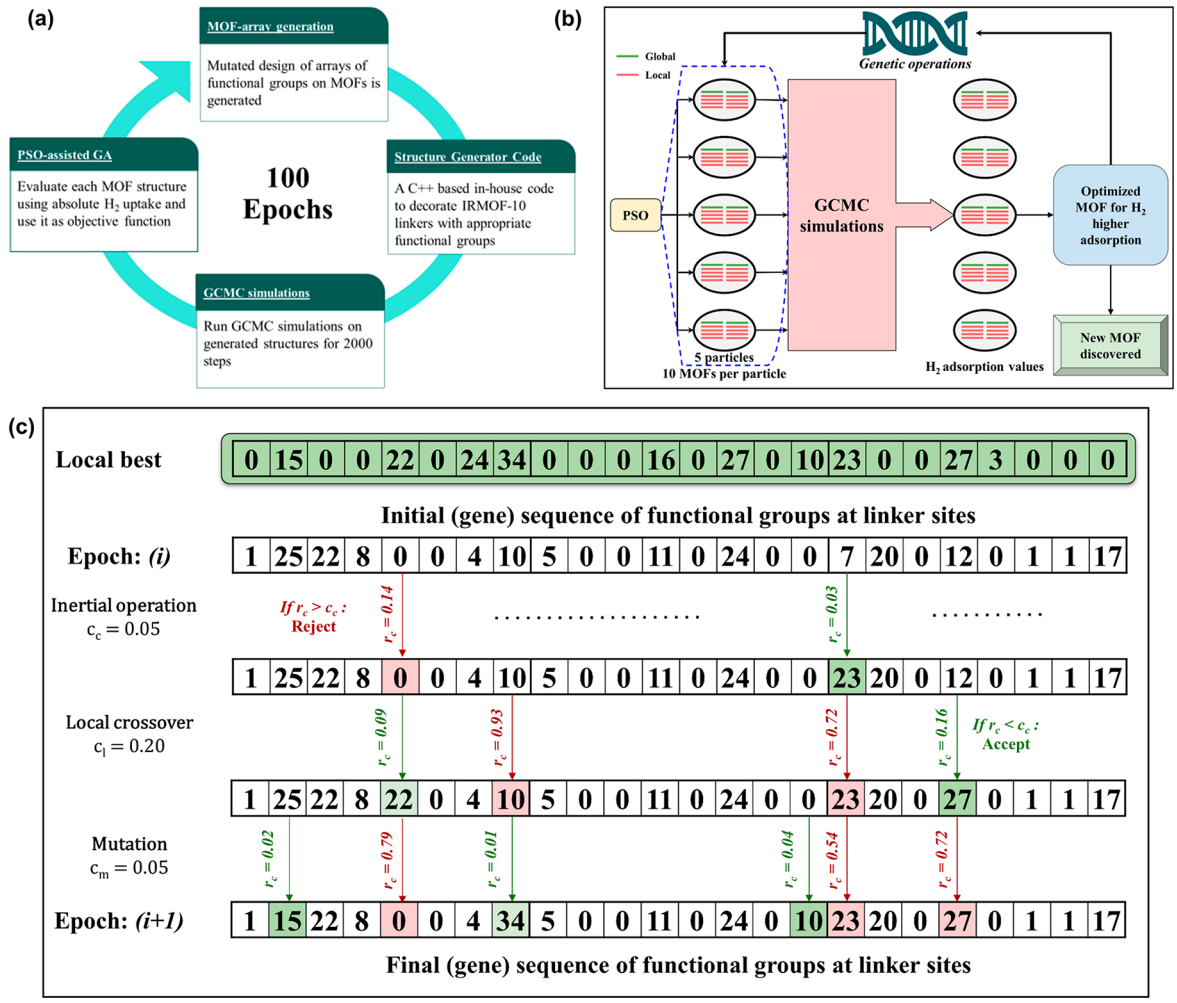


Figure 1. (a) Schematic of the overall framework of PSO-integrated GA combined with GCMC and the structure generator code. (b) Detailed schematic of PSO-integrated GA framework. (c) An example of various genetic operations performed on a functionalized MOF design going from the $(i)^{th}$ epoch to the $(i+1)^{th}$ epoch where functional groups change according to the local best and thresholds (c_c, c_l, and c_m).

(NU-1102), and 6245 m^2/g (NU-1103). Besides, they also showed that surface area and pore volume are correlated.[20] Ahmed et al. conducted the high-throughput computational screening using Grand Canonical Monte Carlo (GCMC) simulations and synthesized the best candidate, IRMOF-20 - that showed good H_2 adsorption in terms of both volumetric capacity (51.0 g/L) and gravimetric capacity (9.1 wt %) at 5 bar and 160 K. Compared to IRMOF-1, IRMOF-20 has a larger pore diameter (17.3 Å vs 15.1 Å), pore aperture (9.3 Å vs 7.9 Å), and geometric void fraction (0.84 vs 0.81). Different from IRMOF-5, the IRMOF-20 linker contains sulfur atoms, which may have a strong linker-H_2 interaction. All these structural and chemical factors resulted in the higher H_2 adsorption of IRMOF-20 compared with IRMOF-1 under the same conditions.[21]

Considering the pore size, void fraction, surface area, and interactions between gas molecules and organic linkers can alter the H_2 uptake, a powerful approach to control these MOF features is the functionalization of metal nodes or organic linkers of the existing MOFs.[23−28] For example, Chen et al. demonstrated that functionalizing the organic linkers of MOF-5 increased its H_2 uptake capacity by 84% compared to bare unfunctionalized MOF-5.[28] However, the three most critical factors that influence the adsorption performance of functionalized MOFs are the type of functional groups, the quantity of functional groups, and their locations on MOFs. By rearranging the locations of functional sites on MOFs, an extensive variety of hypothetical MOF structures (known as "hMOFs") can be generated through permutations and combinations. For example, with 39 functional groups and 24 functionalization sites on an MOF, it is possible to create $\sim 39^{24}/3! = 10^{37}$ distinct MOFs by arranging the functional groups in various combinations. However, exploration of such a vast design space using traditional computational and experimental approaches could be very challenging.

Efficient exploration of this "*infinite*" design space can be facilitated by optimization approaches including evolutionary algorithms like genetic algorithms (GAs) or the metaheuristic technique of particle swarm optimization (PSO).[29,30] For instance, Collins et al. successfully explored a search space of 1.65 trillion MOF structures for exceptional CO_2 uptake using a GA.[31] They developed an MOF functionalization GA (MOFF-GA). This MOFF-GA uses two mating schemas and two different mutation mechanisms along with 13 GA parameters to optimize the MOF functionality to achieve MOFs with good CO_2 uptake. They also showed that their MOFF-GA is powerful not only to find the best functionalized structure with large design space but also beneficial to search

https://doi.org/10.1021/acs.jctc.3c00081
J. Chem. Theory Comput. 2023, 19, 6686−6703

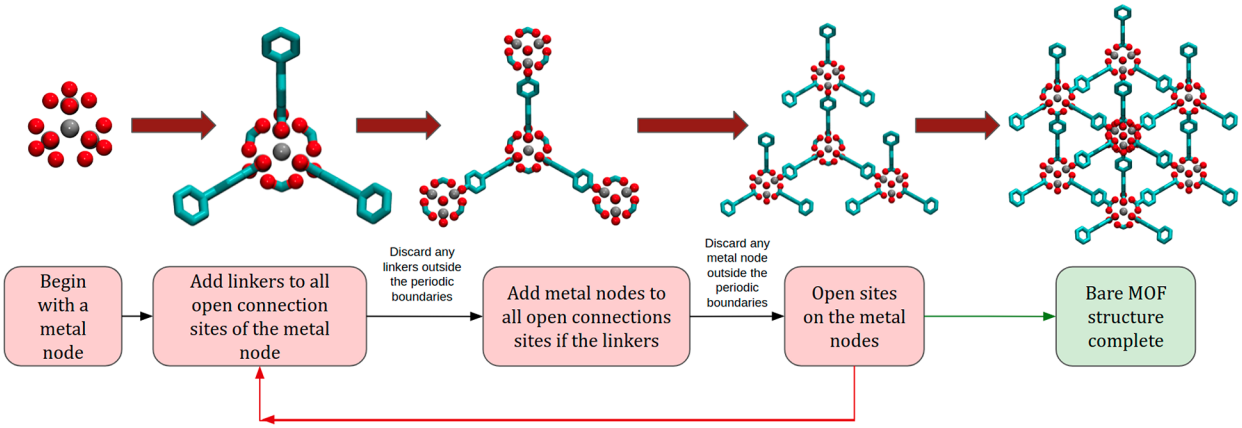


Figure 2. A schematic of the detailed outline of the workflow for generating structures.

functionalized MOF structures with good gas adsorption from a small space of less than 1000.[31] For materials design, PSO is often coupled with different machine learning (ML) methods.[32] Dashti et al. used a PSO-Adaptive Neuro-Fuzzy Inference System (PSO-ANFIS) to estimate CO_2 adsorption in MOFs.[33] Qiao et al. applied backpropagation neural network (BPNN) and decision tree (DT) ML methods along with PSO to investigate the xylene adsorption-based separation in MOFs, and it turned out that PSO performed well to accelerate the prediction from the ML methods.[34] Beauregard et al. carried out GCMC simulations for methane adsorption for over 130,000 *hMOFs*, and they applied the GA with random forest (GARF) model to evolve the structure and chemical properties of MOFs, which yielded high methane uptake successfully.[35]

In this study, we have developed a hybrid optimization algorithm by integrating GA with PSO, namely, the PSO-integrated GA method, which is used to optimize the combination of functional groups on MOF linkers to improve its H_2 adsorption capacity. Specifically, this framework identifies key functional groups and their appropriate location sites on a bare MOF with the goal of enhancing H_2 adsorption, as calculated by GCMC simulations. Furthermore, we develop machine learning and deep learning models using the data generated during this functionalized MOF design process to demonstrate the potential of artificial intelligence in order to replace expensive GCMC calculations.

2. MATERIALS AND METHODS

2.1. Computational Framework and Protocol. Figure 1(a) shows our novel computational framework that seamlessly integrates GCMC simulations with the PSO-integrated GA framework and the in-house MOF structure generation algorithm to efficiently explore the functionalization design space. The PSO-integrated GA framework is used to optimize the location and selection of available functional groups at appropriate sites. The H_2 adsorption in MOFs, calculated in GCMC simulations, acts as an objective function for optimization. Figure 1(b) and (c) summarizes the schematic representation of the PSO-integrated GA approach to generate MOF-arrays for optimization with GCMC. In the following sections, we will discuss different components of this framework in detail.

2.1.1. Computational Details for GCMC. We performed GCMC simulations using the LAMMPS simulations package.[36,37] The Monte Carlo runs consisted of translation, rotation, insertion, deletion, random reinsertion, and regrowth of an existing molecule on hydrogen with equal probabilities. In order to determine the potential parameters of MOFs and their functional groups, we applied the Universal Force Field (UFF).[38−40] Note, the MOF atoms were fixed to their crystallographic positions; however, the functional groups were subjected to MD simulations. The supercell used in this study was 33.35 × 33.35 × 33.35 $Å^3$. The cross-interactions between nonbonded atoms were defined using Lorentz−Berthelot (LB) combining rules.[41,42] Moreover, tail-corrections were implemented, and the interactions beyond 12.5 Å were neglected. Note, both GCMC and MD (on functional groups) simulations were performed at cryogenic temperature conditions *i.e.* 77 K. This temperature reduces thermal energy and mobility of H_2 gas, resulting in valid measurements of hydrogen uptake.[43−45] Additionally, at cryogenic temperatures, MOF-based systems are more stable due to slower degradation rates.[46] Along with most of the simulation studies, several experimental studies are aiming to maximize the hydrogen storage also performed under cryogenic temperature conditions *i.e.* 77 K and 1 bar.[47−49] Furthermore, cryogenic conditions have been shown to be effective in overcoming the challenges associated with physical adsorption and facilitating the feasibility of this method. In addition, BMW has introduced a groundbreaking vehicle concept that incorporates a cryo-compressed hydrogen (CcH2) storage mechanism.[50] A significant increase in storage capacity and greater thermal endurance has been achieved due to to this innovative technology, ultimately prolonging the life cycle of the automotive while ensuring a high level of safety.[50−52] Moreover, to match the set benchmark, several simulation studies have also studied H_2 adsorption in MOFs at these conditions, and therefore, for the sake of consistency, we have performed our simulations at these conditions. Gas−gas and gas−MOF interactions were defined using the Lennard−Jones (LJ) potential. A three-site rigid molecule with LJ 12−6 potential, with σ_{HH} = 0.296 nm, ε_{HH}/k = 34.2 K, and hydrogen bond length of 0.074 nm, was used to model H_2.[43,53] The partial charges were at the two end atoms (q = +0.4705) and at the center of the mass (2q = −0.941) of the H_2 molecule, which allowed reproduction of the quadrupole moment.[43,53] The total mass of the H_2 molecule was placed at the center of mass, while the two ends of H_2 were free of mass.

Extended Charge Equilibration Method (EQeq) charges were implemented in our code to assign partial point charges to MOF atoms to compute electrostatic interactions between

https://doi.org/10.1021/acs.jctc.3c00081
J. Chem. Theory Comput. 2023, 19, 6686−6703

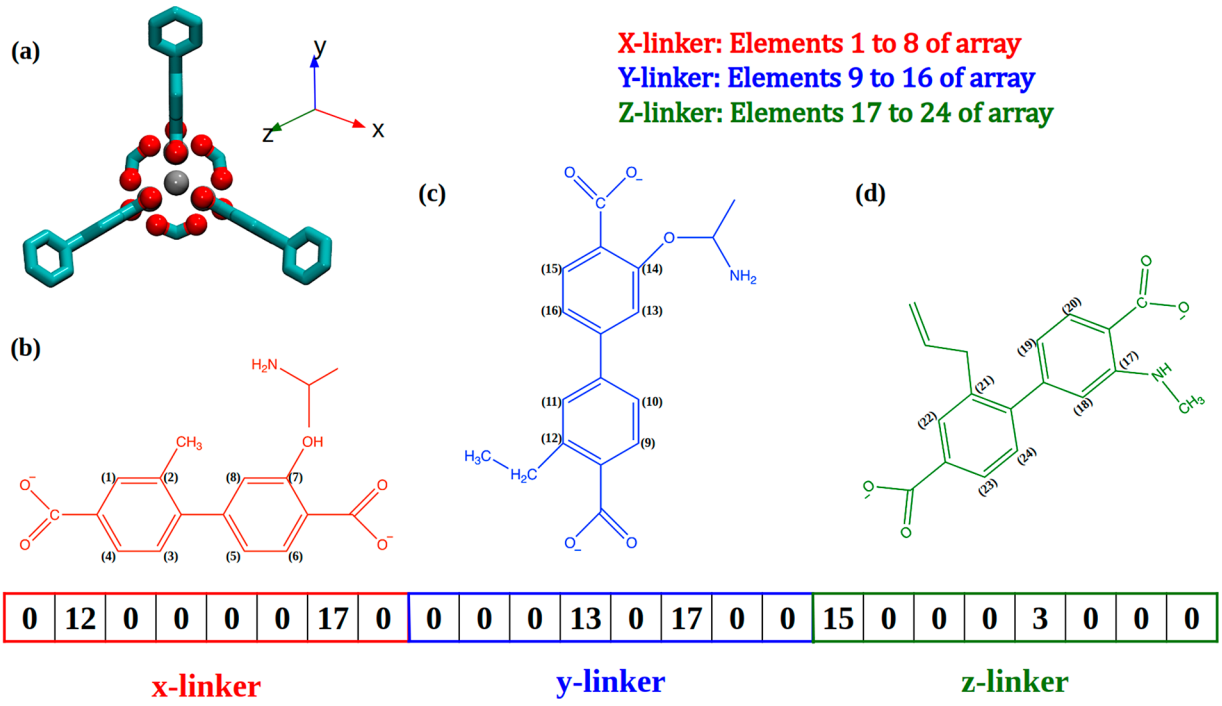


Figure 3. Systematic representation of functionalization of organic linkers of an MOF. (a) The elementary building block of IRMOF-10 is shown, which is repeated in *x*-, *y*-, and *z*-directions periodically to generate MOF structures. The indices of the array of size 24 represent the sites of functional groups specifically on the (b) x-linker (1−8), (c) y-linker (9−16), and (d) z-linker (17−24). Note, the elements of this array portray the functional groups, and the numbers in parentheses of each linker denote the indices of an array (or the location of functional groups on linkers).

H_2 and MOFs.[54] Wilmer et al. presented improved accuracy (over the traditional charge equilibration method) and faster computation of EQeq charge calculation, thus deeming it suitable for high-throughput screening of MOFs. In order to get a fair agreement in the qualitative trend, we plotted the adsorption relative to the number of GCMC steps to identify the plateau in the graph. As shown in Figure S5 of the SI, we determined that a plateau formed around ∼2000 GCMC steps, and as a result, 2000 steps were sufficient for broadly screening the MOFs. Accordingly, during the optimization cycle, we conducted 2000 GCMC steps and simultaneously MD simulations on the functional groups with a 0.5 fs time step in order to allow the functional groups to move. Besides, these 2000 GCMC steps, simulation results were used for the machine learning training shown in Section 2.2. To calculate the adsorption isotherms, we ran the selected top 10 structures for 500,000 GCMC steps.

To validate our simulation protocol and UFF parameters, we have performed GCMC simulations of IRMOF-10 to reproduce the H_2 adsorption isotherm using a different number of steps. As shown in Figure S2 in the SI, these results are in excellent agreement with experimental results reported by Ryan et al.[55] Note that all of the GCMC adsorption values in this work are absolute.

2.1.2. Bare MOF Assembly. In this research, a systematic structure generation algorithm was developed to construct an MOF (Metal−Organic Framework). The algorithm operates by iteratively attaching linkers to metal nodes and subsequently connecting metal nodes to the linkers, following a sequential process. Initially, a single metal node with a number of open sites (N_s), where linkers can be attached, is placed in a simulation cell (Figure 2). At each of these open sites of the metal node, a linker is attached in three directions *viz*. *x*-, *y*-, and *z*-directions, as shown in Figure 2 (second figure). The algorithm then searches every linker for any open sites and attaches a metal node at each site. In the next iteration, the metal nodes with the open sites for the connecting location for the linkers and this process continue iteratively until a complete MOF cell, as shown in Figure 2 (last figure), is formed. Note, only the principal atoms of MOFs are built in the bare MOF construction without the hydrogens on carbon linkers. Thus, in the latter part of the protocol, to ensure that MOFs are fully saturated (*i.e.* no missing hydrogen), we included hydrogens as one of the functional groups during the functionalization process.

2.1.3. Functionalization of Bare MOF. In addition, as the functional groups are attached to aromatic rings, they are arranged in the same plane as the aromatic rings themselves. The UFF force field is used to calculate the bond length between the atoms of the functional groups and the carbon atoms on the linker that they are attached to.[38] If the atoms of functional groups are overlapping (*i.e.* distance is < 0.1 Å), then they were rotated about the connecting bond axis until the overlap vanished to form a valid structure for GCMC simulations. These functional groups on these structures were then relaxed using a conjugate gradient (CG) algorithm implemented in the LAMMPS code and UFF parameters while holding IRMOF-10 atoms fixed to its crystallographic position.[36,56] This further eliminated any minor overlaps between any atoms of the functional groups. Following the minimization step, the structures were considered to be ready for GCMC + MD simulations. Table S1 of the SI describes the list of functional groups used in this study with their corresponding IDs and frequency of occurrence.

2.1.4. Functionalization of IRMOF-10. This study focuses on functionalizing the organic linker (*i.e.* 4,4′-biphenyldicarboxylate (BPDC)) of IRMOF-10, since it offers multiple potential sites to exploit, as shown in Figure 3. In addition to

https://doi.org/10.1021/acs.jctc.3c00081
J. Chem. Theory Comput. 2023, 19, 6686−6703

its simplistic structure and ease of synthesis, IRMOF-10 enables us to control and rotate the orientations of two phenyl rings in a linker. Particularly, it allowed us to model both natures of voids formed of two different sizes by the faces of the BPDC linker, as shown in Figure S1 of the SI. In the BPDC linker, there are 8 hydrogens/carbons on aromatic rings that serve as potential functional sites, as shown in Figure 3. To functionalize linkers of IRMOF-10, each linker's potential functional sites were numbered from 1 to 8 in three directions (*x*-, *y*-, and *z*-) generating an array of a total of 24 sites, as shown in Figure 3. Furthermore, any functional group can be chosen from a pool of 39 functional groups, including hydrogen, and was allowed to attach to the respective carbon site of the linker. Consequently, the MOF structure was defined by the array of functional groups at the appropriate locations.

Based on the adsorption results from the above designing method, we analyze the frequency of the 39 functional groups in the generated structures. The details for functionalized MOF generation are described in Sections 2.1.2 and 2.1.3 as well as in S1.1 of the SI. With an anticipated higher impact on enhancing H_2 adsorption, the top 20 highest occurring functional groups were shortlisted. The 19 functional groups (without considering H) were divided into four groups: 1) C≡C and benzene containing groups; 2) long-chain hydrocarbon groups with at least two C═C; 3) short-chain alkyl groups; and 4) short-chain nitrogen-containing groups. The functional groups and their frequencies are shown in Table S1 of the SI. Following the completion of the determination of an array of functional groups, our "in-house" structure generation code attaches the selected functional groups at appropriate locations. The X site of the functional group shown in Table S1 of the SI will be replaced by the ring carbon atom from the IRMOF-10 linker. Based on the van der Waals radius of each type of atom, the valid "*hMOF*" structure was generated with a minimal 0.1 between the functional groups atoms to avoid overlapping. Later, several structural and chemical features of the generated MOF, including void fraction (VF), surface area (SA), and mass in *"hMOFs"* were calculated. In order to facilitate the arrangement of the MOF computing cell, a uniformly sized grid of cubical tiles of 1 each was used. We examined every cell, in turn, to determine whether it is occupied by an atom based on its van der Waals radius. A void fraction was calculated by dividing the sum of the volumes of unoccupied grid cells by the total volume of the computational cell. Furthermore, if an occupied grid cell was adjacent to any unoccupied grid cell, it was defined as a surface cell and used to calculate the internal surface of the MOF. Moreover, the straightforward calculations of the total mass and number of carbons in a functional group were incorporated directly into the code for the generation of the structure. These features of the MOF accounted for the explicit prediction of H_2 adsorption through ML models. Consequently, it was deemed important to optimize an appropriate set of an array of relevant numbers representing the location and selection of functional groups on the linker of an MOF. Furthermore, enclosed herewith is the C++ source code responsible for the assembly of the family of IRMOFs and its subsequent decoration with functional groups. This code is included in the SI to foster its utilization by researchers in their investigations, while also requesting due acknowledgment of this study whenever it is employed.

2.1.5. Particle Swarm Optimization (PSO) Integrated with the Genetic Algorithm (GA). GAs are iterative processes designed to assess the fit of data over a period of several generations based on Darwin's theory of natural evolution.[57] In recent years, GAs have been successfully implemented to optimize large and complex design spaces pertaining in the design and development of new materials including metal−organic frameworks, metals, and their oxides and composites.[29,31] GAs are built through the application of rational design principles that lead to the creation of material structures that can be synthesized for a wide range of applications.[29,31,58] The use of GAs can improve MOF design processes by enabling iterative, data-informed decisions in order to optimize functionalization and their location.[59] The need for data-enhanced discovery methods of novel materials can be fulfilled by improving the ability of the GA to search a large design space by combining it with another evolutionary method, the PSO algorithm, which we refer to as a PSO-integrated GA.[60−62]

The GA and PSO have been integrated in different forms through a combination of mutation and crossover, combined with the mechanisms of PSO.[63,64] They are used for a variety of applications including data driven biology,[65] chemical kinetics,[66] cloud computing,[67] and so on for various constrained and unconstrained optimization problems. Ali and Tawhid proposed a hybrid Particle Swarm Optimization and Genetic Algorithm (HPSOGA) by improving the exploration through dimensionality reduction and the population partitioning process. Specifically, a genetic mutation operator in this algorithm prevents premature convergence. However, population partitioning, resembling a grid search in its approach, may encounter limitations in effectively exploring complex rough surfaces. Certainly, in a multidimensional space, the exploration process can face significant difficulties since the parameter space expands exponentially with the increase in number of dimensions.[68] Similarly, Singh and Singh presented hybrid PSO and gray wolf optimization (known as HPSOGWO) to improve the exploration of the parameter space.[69] However, the explicit local and global exploration of the parameter space still remains a grand challenge. In our current, novel framework, we borrow the concepts of particles, local best, and global best from PSO, and the GA's basic concepts of mutation and crossover. Conceptually, the use of particles enables us to run multiple GAs in parallel. The local and global bests are associated with each particle and all of the particles, respectively. Both local and global bests can participate in GA operations, which further enhances the possibilities of exploring larger design space compared to a traditional GA. Furthermore, in addition to hyperparameter tuning, the flexibility of starting from a specific point (array) provides a significant advantage for the algorithm.

During the PSO-integrated GA design of MOFs, a series of GA operations were performed upon an MOF array (of 24 size) representing the location of various functional groups on the MOF linker. In a PSO-integrated GA, rather than a common pool of candidates to be assessed through fitness function, separate groups (particles) were created so that GA operations would be performed according to both local (for each particle) and global (including all particles) best practices. Therefore, a total of 16 groups (N_p = 16 particles) were created in the optimizer, where each particle contained 8 children (N_c = 8). Each child represents a set of *"hMOFs"* parameters also known as *"genes"*. We performed GCMC

https://doi.org/10.1021/acs.jctc.3c00081
J. Chem. Theory Comput. 2023, 19, 6686−6703

Chart 1

Pseudo Code:

“

1: Set the initial values of the populations size P, acceleration constant c1 (local) and c2 (global) crossover probability w, mutation probability c3, initial randomization threshold (c_init: Higher value means more functional groups in initial population), number of particles (num_particles), number of local children (n_local) and global children (n_global) in each num_particle and the maximum number of iterations max_epochs. Define the range the genes could have *i.e.* min_P and max_P. We also define the velocity of each individual as the change for each gene in the given parameter space. Here, vel_inertial, vel_local, and vel_global are the change in gene space after crossover, and local and global mutations, respectively.

Start with assigning all genes as 0s for every individual.

2: Assign a starting array of features as first individual, and generate a random number (rn) for the rest of population randomly to generate P^0

If rn < c_init:

Change the 0s in the array of genes to = rn * (max_P-min_P) + min_P

End if

3: Evaluate the fitness function of all individuals in P^0 and determine the local (B_{local}) and global (B_{global}) bests.

4: Repeat

(i) Select all individuals from the population and randomly assign a number r1 and r2 in the range (0,1)

(ii) *(a)* **If r1 < w:**

vel_inertial = (max_P - min_P)*r2 + min_P

Iterate through each individual (local and global child in each particle separately)

End If

(b) Local Mutation:

Assign a random number r_c in the range (0,1)

If r_c < c1:

vel_local = B_{local}

End If

(c) Global Mutation:

Assign a random number r_c in the range (0,1)

If r_c < c1:

vel_global = B_{Global}

End If

(d) Local and Global Cross-breeding:

Add the updated velocity to initial set of genes and bound those genes in between (min_P, max_P)

Update P

Evaluate the fitness function of all individuals

Until Max_epochs”

https://doi.org/10.1021/acs.jctc.3c00081
J. Chem. Theory Comput. 2023, 19, 6686−6703

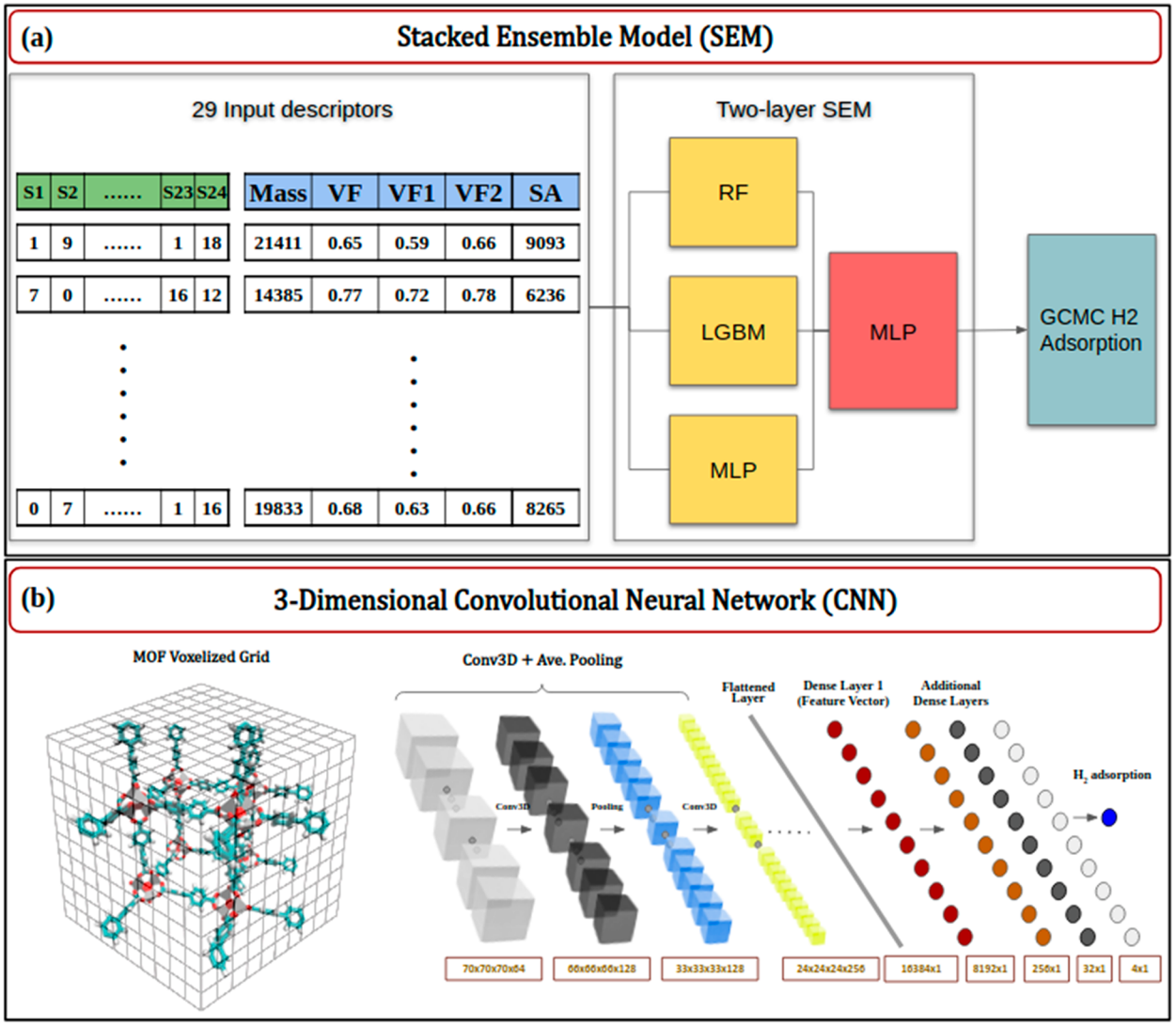


Figure 4. (a) A schematic representation of a stacked ensemble model (SEM) using RF, LGBM, and MLP as the hidden layers to output the H_2 adsorption. (b) A schematic representation of the 3-D CNN model.

simulations on each child, and the final H_2 adsorption for each MOF was considered to evaluate each MOF as a fitness. This PSO-integrated GA algorithm tracks the best design parameter (P_{best}) for each particle. A PSO particle has two generations of children: one for local evolution (N_{cl} = 4) and the other for global evolution (N_{cg} = 1) thereby adding new genes to the particle's gene pool. In this case, P_{best} and its design parameters are used at epoch (i) to determine the local children at epoch (i+1). Similar to the local children, the global children were created by crossing G_{best} with its design parameters at epoch i. Figure 1(b) illustrates how a child's genetic characteristics evolve over time through the process of crossing over and mutations. By combining both of these factors, this algorithm is capable of improving local genes, as well as bringing in improved genes from the global pool. In the next step, the local and global children were subjected to random mutations, in which design parameters could be randomly altered, as shown in Figure 1(c). The hyperparameters used for the PSO-integrated GA for local and global mutations are described in Table S2 of the SI. To shed light on the algorithm in a step-by-step protocol, the following pseudocode for this algorithm is provided, as shown in Chart 1.

To make a direct comparison of the PSO-integrated GA with a traditional GA and random baseline, as a test case, we ran the adsorption of H_2 simulations for 128 candidates for 100 optimization iterations for 200 GCMC steps. These results are discussed in Section S4.2 of the SI and suggest that the PSO-integrated GA is an effective approach to solve complex optimization problems, and it performs better than a traditional GA. The frequency of occurrences of a given functional group at the given specific location during evolution of a traditional GA and PSO-integrated GA are presented in Movies S5 and S6 of the SI, respectively. Movie S5 demonstrates that with a traditional GA algorithm only a single portion of the heatmap becomes darker in color (blue), indicating that only a limited section of the search space was thoroughly explored. In contrast, Movie S6 for the PSO-integrated GA shows a gradual increase in the number of darker spots across the expansive design space, indicating that as the algorithm evolves, more and more of the search space is explored efficiently.

2.2. Machine Learning and a Deep Learning Approach to Predict H_2 Adsorption. In principle, machine learning (ML) and deep learning (DL) models can be used to

https://doi.org/10.1021/acs.jctc.3c00081
J. Chem. Theory Comput. 2023, 19, 6686−6703

replace computationally inefficient GCMC calculations for H_2 gas adsorption. Here, five different ML and DL models were developed and compared for their ability to accurately predict H_2 gas adsorption. These constitute (*a*) a Random Forest (RF) model, (*b*) a Light Gradient Boosting Machine (LGBM) model, (*c*) a multilayer perceptron (MLP) model, (*d*) a stacked ensemble model (SEM), and (*e*) a 3-D convolutional neural network (CNN) model.

RF models combine several decision trees into an ensemble approach in order to create a predictive model. In general, the final prediction of RF flows either from voting by the members of the tree and picking the majority vote for a classifier or an averaged value for a regression.[70] As a result, RF models are simple to construct yet robust against overfitting, making reliable predictions with sufficiently large training data sets. As a result of its simplistic architecture, it is one of the fastest ML methods. The LGBM is one popular algorithm based on gradient boosting machine (GBM). It is a highly efficient decision tree, which shows good performance on the predicating CO_2[71] and arsenic[72] adsorption using an MOF. Compared with the RF model, it may perform better because the LGBM trains the gradient boosting trees in a sequential manner; each tree was trained to correct the errors of the previous tree. The MLP is a feed forward artificial neural network (ANN), which has been widely used in MOF adsorption studies.[73,74] It consists of an input layer, one or multiple hidden layers, and an output layer. Each layer is fully connected to the one before it. Anderson et al. have applied the MLP to study the adsorption isotherm for different small adsorbates at different pressures in an MOF. Their MLP model performs accurate prediction, on average, of adsorption isotherms in MOFs for adsorbates not in training data.[73]

One of the previous studies in our group demonstrated the feasibility of the SEM combined with MD for predicting contact angles and hydrogen bonding in a water droplet.[75] In the SEM, there are several layers of interconnected models, each layer containing a different ML model, as shown in Figure 4(a). More specifically, the first three ML models were trained using the features and properties as inputs and outputs. However, in the SEM, the outputs from the first three models were used as input, and subsequently, an MLP model was built to give adsorption as output in the final step of evaluation. A further enhancement of the prediction accuracy of the ML model is achieved by combining these predicted properties of the ML model in the first layer and feeding them into the ML model of the second layer as input and so on. It is through this process of stacking ML models in one or more layers that prediction accuracy is enhanced.[76] All four models were built and trained using the Scikit-Learn Python package.[77,78] The 12878 *"hMOFs"* data set from GCMC was split into two subdata sets, 80% for training and the rest 20% for testing. 10-fold cross-validation was performed on the training data for each model to ensure their validity. The input features have 29 descriptors, including the functional groups at 24 different functional sites, the total mass of the functionalized *"hMOFs"*, the surface area (in $Å^2$), and total void fraction (TVF) of the MOF as well as void fractions of two different pores of MOFs (VF1 and VF2). The volumetric adsorption was used as the label for training. In order to determine the best hyperparameters, a grid search algorithm was implemented to alter the following hyperparameters: (1) RF: n_estimators {50, 100, 200}, max_depth {None, 5,10}, and max_features {'auto', 'sqrt', 'log2'}, (2) LGBM: learning_rate {0.01, 0.05, 0.1}, n_estimators {100, 200, 300}, num_leaves {5, 10, 20}, max_depth {None, 5, 10}, (3) MLP: hidden layer sizes {(10,), (10, 30), (10, 30, 10)}, activation {'relu', 'tanh'}, solver {'adam', 'sgd'}, alpha {0.0001, 0.001, 0.01, 0.1, 1}, learning_rate {constant, adaptive}.

Further, deep learning techniques include convolutional neural networks (CNNs), which are typically used for image classification and recognition tasks, where the input data consists of 2D arrays of pixel values (e.g., a 32 × 32 × 3 image has 32 rows, 32 columns, and 3 color channels).[79] Nevertheless, CNNs are also capable of processing 3D data such as medical images or point clouds. It is the spatial characteristics of functional groups that remain an integral part of MOFs that determines their hydrogen adsorption capabilities. Thus, determining these features remains a crucial task in training an ML model. CNN through its three-dimensional grid approach maps the 3D shape and size of the functional groups through voxelization of the grid if an atom is present. In addition to incorporating these structural properties, we accounted for chemical characteristics by voxelizing the grids according to atom types and the origin of those atoms to the respective functional groups. For example, the labels of atom C1 (sp1 hybridized carbon) were considered as atom type 12, which was affiliated with F.G. number 9 and appended together resulting in label "912" for the grid. In this regard, we were able to emphasize the importance of both functional groups as well as atom type and hybridization. The voxelization and the overview of convolutional approaches are illustrated in Figure 4(b). We employed a 3D kernel of size 3 × 3 × 3 and exponential linear unit (ELU) activation function. A five-layered convolution neural network with 32, 64, 128, 256, and 512 channels was utilized, each followed by the MaxPooling layer. Since we have modeled F.G. as flexible, in order to incorporate the dynamical changes in the F.G. influenced by the H_2 adsorption, in 3D CNN, the last 5 frames spanning the last 500 steps of GCMC simulation were used for each structure for voxelizing 3D grid. Note, due to the restrictions on the computational resources, we have used only ∼9000 MOFs (∼45,000 frames) to train 3D CNN models.

To evaluate the performance of each ML model, we use R^2 score, mean absolute percentage error (MAPE), root mean squared error (RMSE), and mean absolute error (MAE), which their details are described in Section S3.1 of the SI.

3. RESULTS AND DISCUSSION

3.1. PSO-Integrated GA Optimization. In this study, the top 19 functional groups with the highest frequency of occurrences (see Table S1 of the SI) in addition to hydrogen were shortlisted and utilized in the optimization cycles. As a starting point, we used a known array that was anticipated to give higher adsorption, and for the remaining 127 arrays, random numbers were generated. This starting point was determined based on initial random runs performed with the 39 functional groups. Here, we performed two separate cycles of the PSO-integrated GA framework to discover new functionalized MOFs with enhanced H_2 adsorption. In Cycle-1, the optimization cycle ran for 39 epochs (iterations), with 128 functionalized MOFs in each epoch, totaling 4992 structures. Additionally, a restart Cycle-2 was implemented, with the best candidate from Cycle-1 used as the initial starting point. Here, we implemented a code to ensure that none of the structures from Cycle-1 were repeated by generating new random arrays in the case of any repetition. Thus, an additional

https://doi.org/10.1021/acs.jctc.3c00081
J. Chem. Theory Comput. 2023, 19, 6686−6703

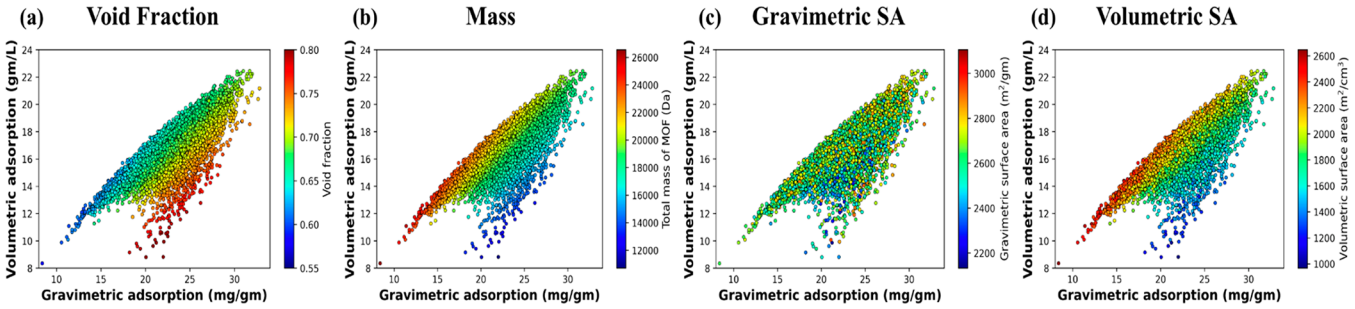


Figure 5. Dependence of volumetric and gravimetric adsorption with their correlation with (a) void fraction, (b) mass, (c) gravimetric surface area (m^2/g), and (d) volumetric surface area (m^2/cm^3).

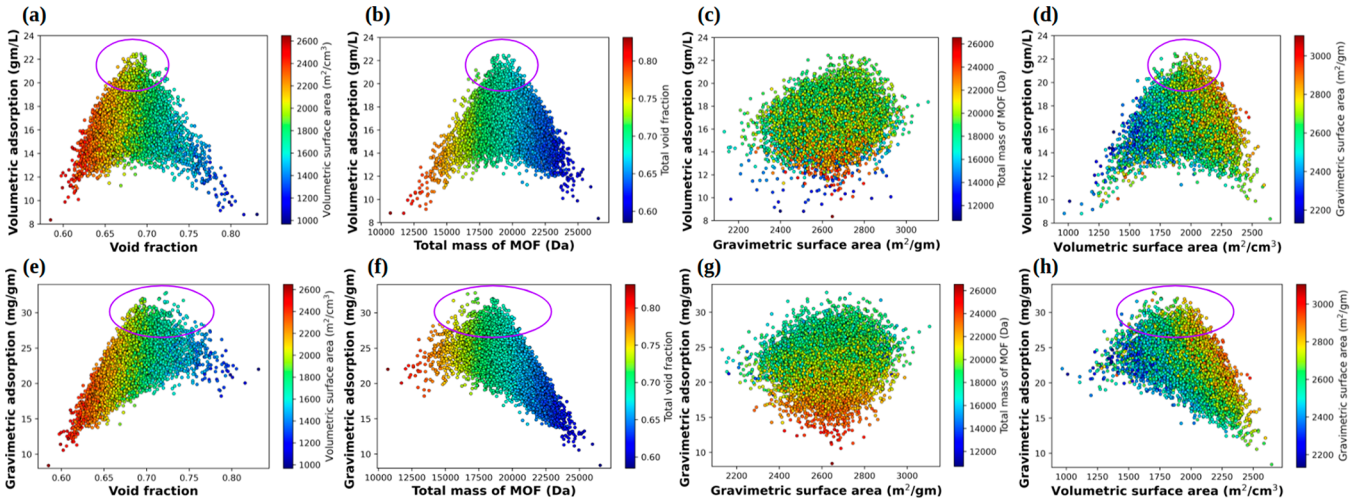


Figure 6. Dependence of volumetric and gravimetric adsorption with their correlation with **(a, e)** void fraction, **(b, f)** mass, **(c, g)** gravimetric surface area (m^2/g), and **(d, h)** volumetric surface area (m^2/cm^3).

8832 new structures were run to make a total of 13,824 structures in Cycle-1 and Cycle-2. In the optimization, all 128 *"hMOFs"* (from every epoch) were subjected to GCMC simulations to determine their absolute hydrogen adsorption in each epoch. GCMC simulations were used as the fitness function for PSO-integrated GA since absolute H_2 adsorption at 1 bar and 77 K was the only evaluating factor for each MOF. Particularly, in both cycles, our structure generation code generated these initial structures comprising the following features in the range: (a) total void fraction (TVF): ~0.55 to ~0.85, (b) total mass (M_{mof}): 10,000 to 26,000 Da, (c) gravimetric total surface area (GTSA): ~2,200 m^2/g to ~3,000 m^2/g, and (d) volumetric total surface area (VTSA): ~1,000 m^2/cm^3 to ~2,500 m^2/cm^3. Note, the range of these features is known to influence H_2 adsorption in MOFs.[9] While the design space spans across $20^{24}/3!$ MOFs, with the use of the PSO-integrated GA algorithm, the exploration of only 12,878 MOFs yielded us the best adsorption MOF. This highlights the effectiveness and efficiency of optimization algorithms as a tool for solving complex problems in the field of MOFs and other materials discovery.

Absolute hydrogen capacity in terms of the number of molecules was converted to volumetric and gravimetric adsorption capacity to compare with the DOE target.[80] According to their respective volumes and masses, unit cells of the MOF were used in the simulations. The best volumetric adsorption of each epoch starting from 18.86 g/L went on to increase until 22.44 g/L in the optimization, as plotted in Figure S4 of the SI. Further, the corresponding gravimetric adsorption was 31.72 mg/g. This MOF comprises 18 functional groups and six nonfunctionalized sites. In this structure, penta-1,3-diene, phenylacetylene, and unsaturated linear carbon chain functional groups appeared more frequently, while methyl, ethyl, pyrrole, and ethyl amine were present in fewer numbers on the bare MOF. Besides, when MOFs were ranked based on their gravimetric adsorption, an MOF structure with the gravimetric and volumetric adsorption of 32.82 and 21.17 g/L, respectively, was also found during optimization. This MOF includes 17 functional groups and seven nonfunctionalized sites. This MOF is notable for its higher occurrence of smaller functional groups like methyl, which decreases the overall weight of the MOF. However, there are fewer occurrences of smaller unsaturated carbon chains like acetylene, propyne, and styrene on the bare MOF. Note that these adsorption values were only after 2,000 steps of GCMC run, which we have further validated in the longer GCMC runs as discussed in Section 3.5 below.

Figure 5(a)-(d) broadly investigates the trend of the absolute volumetric adsorption with respect to gravimetric adsorption obtained from the GCMC calculations with the MOF features (on the color bar) across the set of all MOFs generated in this study. The highest volumetric and gravimetric adsorption was evident for a '*sweet-spot*' for TVF in the range ~0.65 to ~0.71, for mass (16,000 to 20,000 Da) and VTSA (1,600 to 2,000 m^2/cm^3). However, due to the lack of a discernible trend for GTSA, it can be concluded that a '*sweet-*

https://doi.org/10.1021/acs.jctc.3c00081
J. Chem. Theory Comput. 2023, 19, 6686−6703

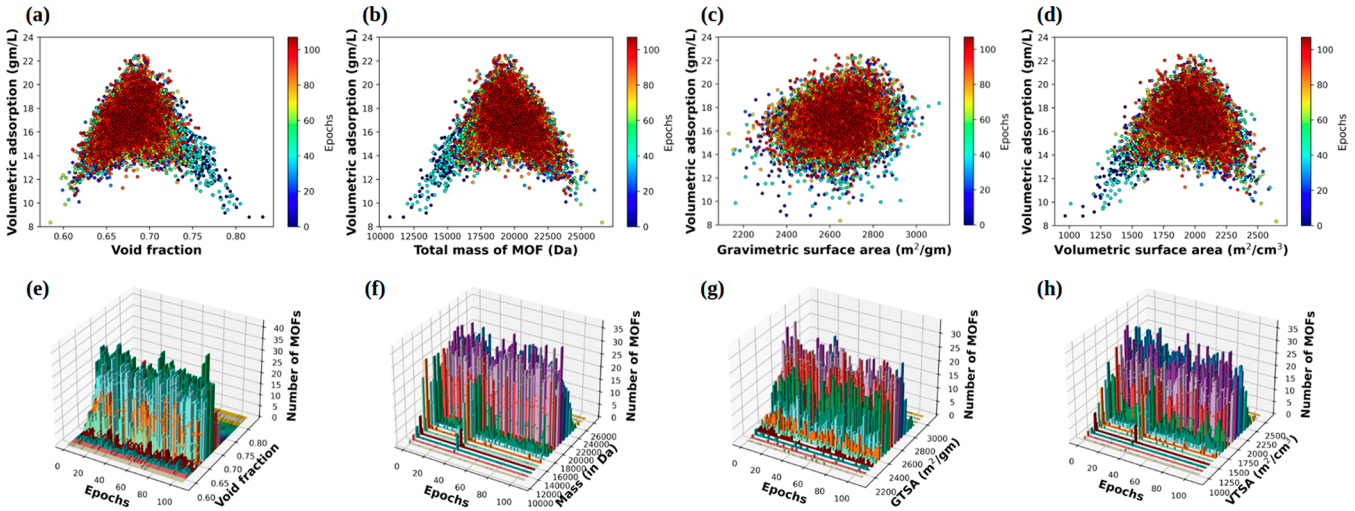


Figure 7. Dependence of volumetric adsorption on the features of MOF *i.e.* (a) TVF, (b) mass, (c) GTSA, and (d) VTSA correlating with the epochs of optimization cycles is shown. The evolution of occurrences of MOFs within specific ranges of the features mentioned above is shown in (e-h).

spot' for GTSA does not exist. As a result, attempting to optimize GTSA may not be a viable option since there is no specific target for which to aim for. Furthermore, a distinct color gradient, blue to red as we move from left to right in Figure 5(a), suggests that with an increase in void fraction, the gravimetric adsorption increases. This increased void fraction could be due to the decreased number of functional groups, thereby decreasing the mass of the MOF and resulting in increased gravimetric adsorption. Therefore, a similar color gradient was observed for the mass and VTSA in Figure 5(b) and (d). Finally, it can be concluded that the discernible trends in adsorption are greatly influenced by factors such as the void fraction, mass, and VTSA. These features play a significant role in determining the adsorption behavior.

Figure 6(a)-(h) demonstrates the dependence of H_2 adsorption of each MOF from the entire set of structures from PSO-integrated GA run in relation to their void fractions, mass, VTSA, and GTSA. With an increase in the void fraction (from ∼0.58 to ∼0.67), mass (from ∼10,000 Da to ∼19,000 Da), and VTSA (from ∼1,000 m^2/cm^3 to ∼2,000 m^2/cm^3), the absolute adsorption capacity of the MOFs increased from ∼8.2 g/L to ∼22 g/L, as shown in Figure 6(a), (b), and (d), respectively. However, further increases in the void fractions, mass, and VTSA led to the reduction of the absolute hydrogen adsorption capacities. Essentially, this is due to the fact that as the surface area increases, the exposure of unsaturated carbon atoms increases, resulting in more hydrogen adsorption sites.[81] The presence of double and triple bonds between carbon atoms allows unsaturated carbon chains, such as acetylene, propyne, penta-1,3-diene, and styrene, to have a greater electron density than saturated carbon chains. Consequently, the unsaturated carbon atoms and hydrogen molecules have stronger interactions because of the increased electron density, resulting in improved hydrogen sorption.[81] Eventually, further increase in the VTSA beyond ∼2,000 m^2/cm^3 leads to a decreased void fraction (to ∼0.6), leading to smaller pores and less volumetric space available for hydrogen adsorption. Thus, a "*sweet spot*" exists for optimal adsorption marked by the purple circle in Figure 6, suggesting that MOFs comprising only specific ranges in their void fraction, mass, and surface areas may yield higher adsorption. Similar behavior pertinent to the gravimetric adsorption is evident in Figure 6(e), (f), and (h). Moreover, as mentioned above, the increased void fraction, decreased mass, and VTSA, due to the decreased functional groups, resulted in a color trend (blue to red and vice versa) as we move from right to left. As anticipated, no evident trend can be observed in the graphs relating to the GTSA. In addition, we illustrate the effect of other parameters on the absolute uptake shown in Figure S6 and Figure S7 of the SI. These graphs suggest that the ranges of ∼0.67−0.71, ∼17,500 Da−20,000 Da, and ∼1,750−2,100 for the TVF, mass, and VTSA of a unit cell, respectively, may be optimal for high H_2 adsorption. Note, the overlap of points from over the epochs of optimization cycles suggests that the optimization may have reached the "point of saturation" by 108 epochs. The following section further sheds light on the evolution of functional groups and thus the MOF features during the MOF design using PSO-integrated GA framework.

3.2. Evolution of Functional Groups in the PSO-Integrated GA Algorithm. The evolution of the PSO-integrated GA algorithm can be quantified by analyzing the occurrences of the MOFs in the specific ranges of the features of MOF. For that purpose, we have plotted volumetric adsorption w.r.t. TVF, mass, GTSA, and VTSA corresponding to each epoch color-coded from blue (epoch number 1) to red (epoch number 108) in Figure 7(a)-(d). More details can be found in Movies S1−S4 in the SI. Further, the number of occurrences of MOFs in the ranges of TVF, mass, GTSA, and VTSA as the epochs progress is shown in Figure 7(e)-(h). As discussed in the previous sections, the '*sweet-spot*' for TVF in the range (∼0.67−0.70) was indeed achieved by the PSO-integrated GA algorithm as shown by appearances of dark red colored points in Figure 7(a). This can be confirmed by the peaks appearing at epochs > 100 marked by dark green colored bars in Figure 7(e). Further, for initial few epochs (<10), we observed that the majority (75%) of the structures were generated in the TVF range of ∼0.71 to ∼0.78, whereas as the PSO-integrated GA progressed, the structures generated in the range of ∼0.64 to ∼0.70 increased significantly. Therefore, in addition to the adsorption, this optimization cycle targeted the "center of percussion" for void fraction in this range. It was interesting to note that no trend was apparent in the

https://doi.org/10.1021/acs.jctc.3c00081
J. Chem. Theory Comput. 2023, 19, 6686−6703

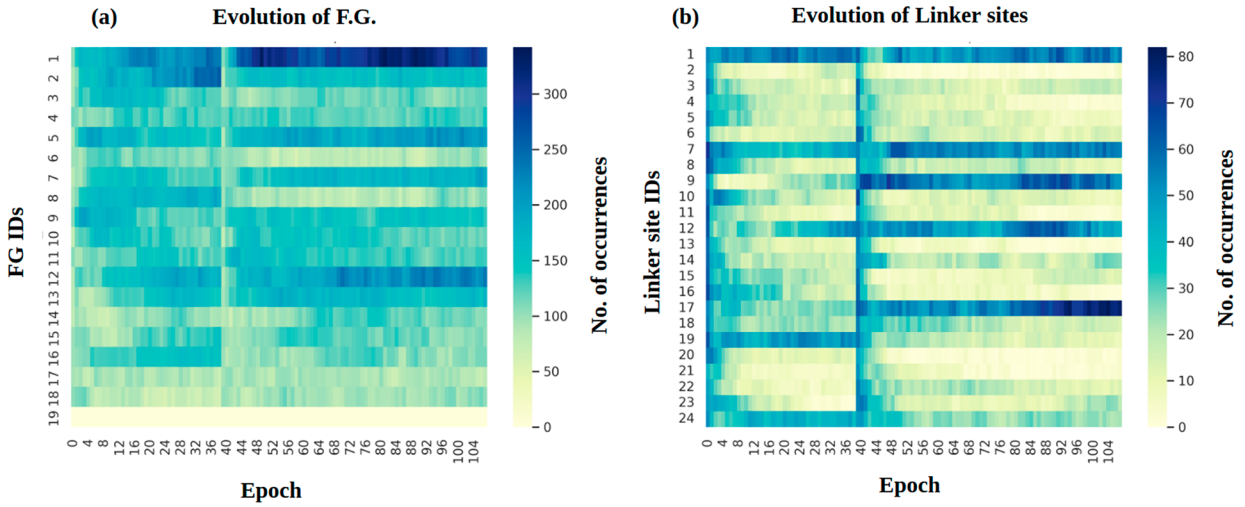


Figure 8. Evolution of occurrences of (a) functional groups and (b) the number of times linker sites were occupied in each iteration of optimization cycles.

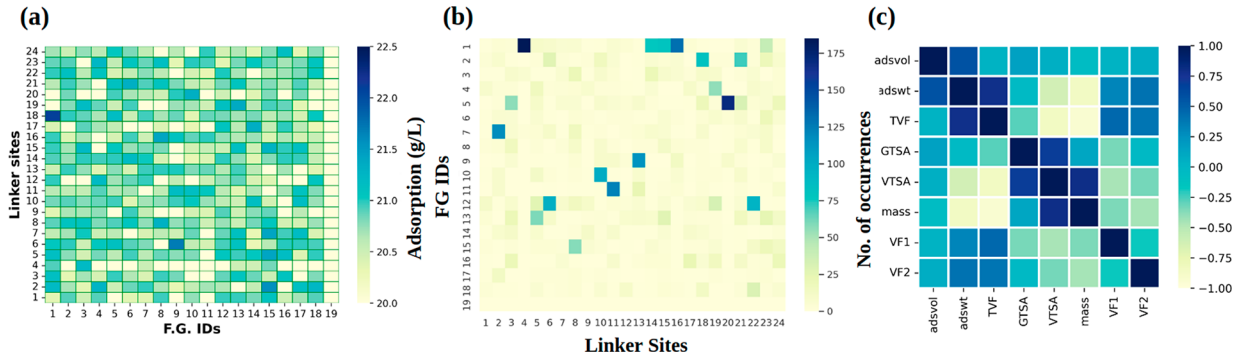


Figure 9. (a) Highest adsorption in MOFs recorded for the functional group-site pairs. (b) Occurrences of drugs with each combination of functional group and site IDs. (c) Heatmap of the correlation of features of top MOFs with the highest adsorption determined by the Pearson correlation coefficient. Refer to Table S1 in the Supporting Information for the names of the functional groups.

volumetric adsorption with respect to gravimetric total surface area, and consequently no '*sweet-spot*' could be attained. However, for the case of VTSA, the number of structures has gradually increased in the *"sweet-spot"* range of 1,750−2,100 m^2/g. Note that the peaks appearing at the ∼39th epoch for mass (10,000−16,000 Da) and VTSA (1000−1750 m^2/cm^3), shown in Figure 7 (f) and (h), correspond to the restart of the optimization cycle. The results of this analysis indicate that while optimizing H_2 adsorption as expected, the other structural parameters of MOFs such as total void fraction, mass, and VTSA were also changed.

To understand the evolution of functional groups and functionalization sites on the linker, we plotted the progression of genes during the PSO-integrated GA run. Figure 8(a) shows that functional groups phenylacetylene, Penta-1,3-diene, 1,3,5-Heptatriene, and methyl (IDs: 1, 5, 7, and 12) appeared more frequently compared to the other functional groups, as either they were smaller in size (like methyl) and/or they had unsaturated linear carbon chains, which assisted in enhancing the H_2 adsorption. Another interesting observation was that functional groups with IDs 1 to 11 appeared more often compared to those with IDs 12 to 19. This can be attributed to the higher adsorption driven by the larger exposure to the linear unsaturated carbon chains (for F.G. IDs: 1 to 11), as opposed to the saturated carbon chains (IDs: 12 to 14) and amine-based functional groups (IDs: 15 to 19). Moreover, in the case of the functionalization sites, the site numbers 1, 7, 9, and 17 were occupied more often in the optimization cycle, as shown in Figure 8(b). These linker sites correspond to the x-linker (sites 1 and 7), y-linker (site 9), and z-linker (site 17). This phenomenon can be attributed to the decreased likelihood of larger functional groups occupying adjacent sites, as they are prone to overlap. To avoid this, it is more common to have one site of a linker in each direction as this allows for larger functional groups to be attached without overlapping. The combination of top occurring functional groups and sites generated the MOFs within the '*sweet-spot*' of void fraction (even though it was not targeted), making this combination appear more frequently in the optimization. This outcome further demonstrates the effectiveness of this specific combination of functional groups in achieving the desired results.

3.3. Analysis of the Selected Top 200 MOFs. Figure 9(a) shows the importance of the location of the functional groups at each site in influencing H_2 adsorption. Specifically, it summarizes the dependence of average adsorption for the given functional groups at a particular site on the linker. The darker blue boxes, for example, for the location of functional group (or F.G.) number 1 at site number 18 primarily may drive the increased adsorption in the top selected MOFs. Figure 9(b) denotes the occurrence of the MOFs with the given functional groups at a particular site on the linker. The

https://doi.org/10.1021/acs.jctc.3c00081
J. Chem. Theory Comput. 2023, 19, 6686−6703

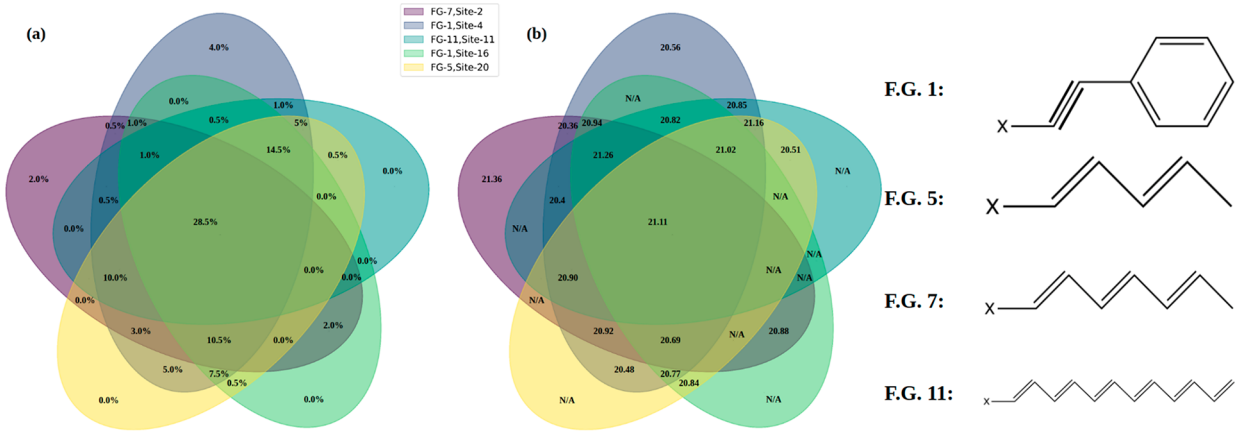


Figure 10. Multiple functional groups at various sites synergistically enhance H_2 adsorption. Therefore, a Venn diagram, which is the best representation of the overlap of the (a) number of MOFs and (b) their corresponding average adsorption values for the functionalization classified into 5 broad categories consisting of (i) Functional group 7 at site 2, (ii) Functional group 1 at site 4, (iii) Functional group 11 at site 11, (iv) Functional group 1 at site 16, and (v) Functional group 5 at site 20, is shown.

highest number of occurrences of 185 MOFs out of the top 200 consisted of functional group 1 (phenylacetylene) at site number 4, highlighted by the dark pearl-blue pixel in Figure 9(b). This was followed by F.G. 5 (Penta-1,3-diene) at site 20 (170 occurrences highlighted by the dark blue), F.G. 1 (phenylacetylene) occurring at site 16, and F.G. 11 (1,3,5,7,9,11-Dodecahexaene) occurring at site 11 for 132 and 123 MOFs (emerald green squares), respectively. However, it was interesting to note that their corresponding adsorption in Figure 9(a) represented by the lighter green shades was comparatively lower. Moreover, the darkest blue shade for FG-1 at site 18 mentioned earlier occurred for only 2 of 200. This clearly indicates that the adsorption within the MOFs is not significantly influenced by a single functional group at a specific site alone but rather by a combined effect of a set (or subsets) of functional groups at their appropriate locations. Therefore, we have carefully inspected and presented a Venn diagram to identify these subsets of functional groups, as shown in Section 3.4.

Indeed, the H_2 molecule is a nonpolar molecule, of which adsorption depends on their interactions with relatively weak polarizing centers.[82] Additionally, at cryogenic temperature (77 K), the high accessible adsorption surface and low bulk density result in a significant amount of H_2 adsorption. As a result of these nonpolar functional groups like phenylacetylene, Penta-1,3-diene, and 1,3,5,7,9,11-Dodecahexaene, which offer a high specific exposed area of unsaturated carbons, the H_2 adsorption in our MOFs is enhanced.[83] Similarly, the long-chain unsaturated alkenes, here particularly 1,3,5,7,9,11-Dodecahexaene (FG-11), could reach the central area of the pore enabling the higher adsorption of H_2 molecules in that region at low pressures.[9] At lower pressures, hydrogen adsorption is primarily driven by electrostatic interactions and is less affected by the presence of larger voids.[17] In fact, most of the top MOFs with the highest volumetric capacity were decorated with these functional groups.

Figure 9(c) outlines the correlation, determined by the Pearson correlation coefficient (PCC), among the features (TVF, mass, VTSA, GTSA, VF1, and VF2) of the top selected 200 MOFs and their corresponding H_2 adsorption (volumetric (adsvol) and gravimetric (adswt)). The correlation coefficient of +1 and −1 suggests that the two properties have positive or negative linear relationships between them. Furthermore, a negative relationship is evident for these properties with the void fractions with VTSA and mass ($PCC_{VTSA\text{-}TVF}$: −0.84, $PCC_{Mass\text{-}TVF}$: −0.94). This observation can be interpreted as the increase in the functional groups resulting in a decrease in the void fraction and an increase in the total surface area in the top 200 MOFs. It was interesting to note that the gravimetric adsorption is linearly correlated with a total void fraction ($PCC_{TVF\text{-}adswt}$: 0.79), while it is negatively correlated with volumetric total surface area ($PCC_{VTSA\text{-}adswt}$: −0.62) and mass ($PCC_{mass\text{-}adswt}$: −0.84), indicating that increasing functional groups contribute to MOF mass and thus decreasing gravimetric sorption. Moreover, the sky-blue shaded plots for volumetric adsorption dependence on the properties of MOFs ($PCC_{TVF\text{-}adsvol}$: 0.03, $PCC_{GTSA\text{-}adsvol}$: 0.16, $PCC_{VTSA\text{-}adsvol}$: 0.06, $PCC_{Mass\text{-}adsvol}$: −0.05) indicate that none of the properties are linearly correlated with the absolute adsorption. And the volumetric adsorption cannot be single-handedly determined by one factor (property).

3.4. Coupling of Functional Groups at Appropriate Sites.

Further, to investigate the occurrence of particular combinations of functional groups at appropriate sites, we plot the Venn diagram of the location of given functional groups, as shown in Figure 10. The top 200 MOFs with the highest adsorption in this study were selected, and the arrays (genes) were systematically inspected. Out of 200 MOFs, the occurrence of any functional at the one-particular site was determined, and the top 5 occurrences were selected. As mentioned in Section 3.3, the top 5 occurrences include phenylacetylene and penta-1,3-diene occurring in 185 and 170 MOFs out of 200 at sites 4 (here we term it as FG1-S4) and 20 (FG5-S20), respectively. Moreover, the next 3 combinations include 1,3,5-Heptatriene (FG7), 1,3,5,7,9,11-Dodecahexaene (FG11) and again phenylacetylene (FG1) at sites 2 (S2), 11 (S11), and 16 (S16, respectively, for 118 (FG7-S2), 123 (FG11-S11), and 132 (FG1-S16) MOFs of 200.

Particularly, we found that functional group numbers 7 (1,3,5-Heptatriene), 11 (1,3,5,7,9,11-Dodecahexaene), and 5 (Penta-1,3-diene) at linker sites 2, 11, and 20 occurred for 118, 123, and 170 MOFs, respectively, out of 200. In addition, functional group 1 (phenylacetylene) occurred at sites 4 and 16 both 185 and 132 times, respectively. To determine the coupling of functional groups in an MOF array, we inspected the overlap among them.

https://doi.org/10.1021/acs.jctc.3c00081
J. Chem. Theory Comput. 2023, 19, 6686−6703

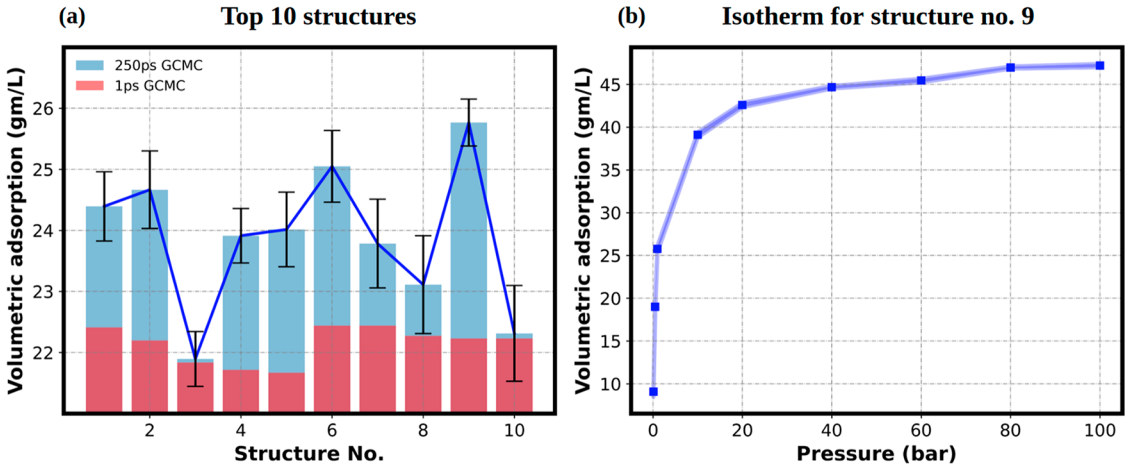


Figure 11. (a) The adsorption of the top 10 structures at 1 bar and 77 K is shown. (b) The adsorption isotherms for pressures ranging from 0.1 to 100 bar at 77 K for structure number 9 (which demonstrated the highest uptake at 1 and 77 K) are shown.

Figure 10(a) suggests that 57 MOFs of the top 200 have all of the aforementioned functional groups at their respective sites. This resulted in average adsorption of 21.1 g/L at 1 bar and 77 K, in 2000 GCMC steps. Evidently, the higher the surface area exposed by the carbons in functional groups, the higher the H_2 adsorption. Moreover, the average adsorption of each of the subsets is shown in Figure 10(b). These data indicate the adsorption values with their corresponding confidence in the subsets. For instance, 28.5% of the top MOFs (for all *i.e.* FG7-S2, FG1-S4, FG11-S11, FG1-S16, and FG5-S20) show an adsorption of 21.11 g/L; however, only 1% of the MOFs show 21.26 g/L (FG7-S2, FG1-S4, FG11-S11, and FG1-S16).

3.5. Adsorption Isotherms. The top 10 structures from the optimization cycle (top 5 from Cycle-1 and top 5 from Cycle-2) were identified, selected, and run for 500,000 GCMC steps. The final adsorption values are presented as bar plots with their standard deviations in Figure 11(a). We observed that 8 out of these top 10 structures had these functional groups connected at their appropriate sites (*i.e.* FG7-S2, FG1-S4, FG11-S11, FG1-S16, and FG5-S20). In this study, we designed an MOF with the highest volumetric adsorption, marking the first instance of such a design (25.76 g/L corresponding to 37.09 mg/g, which showed highest gravimetric adsorption too, at low pressure, *i.e.* at 1 bar and 77 K) for structure number 9 exhibiting these desired functionalization characteristics. It was interesting to note that VF1 (void fraction for pore 1) and VF2 (void fraction for pore 2) were fairly similar *i.e.* 0.695 and 0.68. Therefore, the similar range of void fractions of types 1 and 2 suggests (VF1 ≈ VF2) that higher adsorption may be established in an MOF where both the pores of MOF are uniformly functionalized. Moreover, the top 10 structures' volumetric adsorption ranged from 21.89 g/L to 25.76 g/L, and gravimetric adsorption ranged from 31.45 mg/g to 37.09 mg/g, which were considerably fair for 1 bar pressure DOE targets.[9] In particular, out of these 10 MOFs, 5 MOFs were generated with an absolute volumetric capacity greater than 24 g/L (and 7 MOFs with a gravimetric capacity greater than 32 mg/g) at 1 bar and 77 K. Compared to the unfunctionalized IRMOF-10, with H_2 absolute adsorption of 4.15 g/L and 1.8 mg/g, the functionalized MOFs generated in this study were ∼6 times higher.

The understanding of the effect of surrounding temperature and pressure may assist us to control the amount of deliverable H_2 capacity of MOFs by pressure swing adsorption (PSA) or temperature swing adsorption (TSA). Thus, to further probe the mechanism and validate the adsorption in these structures, the top 10 candidates from the optimization cycle were selected and run for 500,000 GCMC steps. We have employed a broad range of pressures to investigate adsorption isotherms, particularly, 0.1−100 bar at 77 K. The detailed arrays for these 10 structures are shown in Table S3 of the SI. As per results shown in Figure 11, adsorption increases linearly from 9.07 to 25.76 g/L in the range of 0.1 to 1 bar, whereas for the range of P ≈ 1 to 40 bar, *i.e.* at higher pressures, isotherms fluctuate around a constant value. Note that at low pressures, physisorption is the dominant mechanism of adsorption, where the H_2 molecules interact with the surface of the MOF through van der Waals and electrostatic forces,[17] since MOFs with smaller pores may exhibit stronger adsorption due to the increased surface area to volume ratio and the greater interaction between the H_2 molecules and the MOF surface. Additionally, the 3-site model incorporates the quadrupole moment of H_2 (which is relatively an accurate description of charge distribution) enabling stronger electrostatic interactions with MOF. The *hMOFs* studied here have increased surface area, and therefore, at lower pressures, we observe a linear trend of increase in adsorption. As the pressure continues to increase, the role of the surface area in the adsorption process diminishes. This results in the formation of a plateau, where the adsorption of H_2 reaches a constant and stable value (*i.e.* ∼45 g/L), fluctuating only slightly for pressures beyond 40 bar (*i.e.* P ≈ 40 to 100 bar).

According to PSA, the deliverable H_2 capacity for this structure was 18.902 g/L as we moved from 40 to 1 bar. However, the structure number 10 showed the highest deliverable capacity of 23.0 g/L. The deliverable capacities for all other MOFs range from 18.6 to 23.0 g/L and are shown in Table S4 of the SI. The adsorption isotherms for other 9 structures are shown in Figure S8 of the SI. Note, there are relatively large differences between the adsorption obtained from 500,000 and 2,000 step GCMC runs. However, on the basis of a qualitative trend obtained as shown in Figure 11, we performed a proof-of-concept study that showcases only 2000 GCMC steps are sufficient for high-throughput screening;

https://doi.org/10.1021/acs.jctc.3c00081
J. Chem. Theory Comput. 2023, 19, 6686−6703

however, a greater number of GCMC steps should be used to validate the uptake capacity of that MOF. Indeed, there are previous studies in the literature that have used ∼3,000−10,000 steps to perform high-throughput screening of MOFs for gas adsorption.[84−86]

3.6. ML Models to Predict H_2 Adsorption. As a preprocessing step, to ensure the elimination of symmetrically equivalent structures before ML training, all arrays of structures underwent preprocessing using a python script, and duplicates were removed. Specifically, we identified structures symmetric in *x*-, *y*-, and *z*-directions *i.e.* if the same set of numbers in columns 1 to 8, 9 to 16, and 17 to 24 is repeated in 9 to 16, 17 to 24, and 1 to 8, respectively, or in 17 to 24, and 1 to 8, and 9 to 16, respectively, then these structures were deemed as symmetrically equivalent. The performance of all ML and DL models developed in this study is shown in Table 1 and Figure 12. Considering the four ML models (RF, LGBM, MLP, and SEM) trained on the functional group's array and related structure properties, though the R^2 scores showed a best value of 0.68 from LGBM, the MAPE are less than 6%, which was good to compare with the MAPE (∼4.4% in average) predicted for different small molecules in MOFs using MLP from Anderson et al. work.[73] We also found that the LGBM model performed as the best of these four models. The 3D-CNN model has significantly increased the prediction performance with an R^2 score of 0.91 and MAPE of 0.017%. Instead of capturing only the functional group's type and site information, the 3D-CNN model also captures the location of each functional group related to the metal nodes and organic linkers, which significantly affects the adsorption results due to the flexible structure of functional groups during the GCMC simulation. From the parity plot shown in Figure 12(e), the majority of predicted adsorption values has a difference of less than 0.5 g/L from the GCMC simulated adsorption results. Though the total mass, void fraction, and surface area were not used in the 3D-CNN model for training, this information was indirectly captured by the 3D-CNN model. For example, as we use a 70 × 70 × 70 grid mesh, each grid was filled by no more than 2 atoms, usually 0 or 1 atom. The grids labeled with 0 could be used to estimate the void fraction of a functionalized MOF structure. The grid labeled with the functional group atom information gives the type of the certain element, which indirectly gives the total mass. Therefore, the 3D-CNN model actually captures both the chemical and structural information on the *"hMOFs"*, which enhanced the predictive performance of the gas uptake which is similar to the conclusion in Pardakhit et al. study that ML models based on structural and

Table 1. Comparison of Different Machine Learning Models on H_2 Adsorption

	Testing performance 10-fold cross-validation				
	MAPE (%)	RMSE (g/L)	MAE (g/L)	R^2	MSE (g/L)
RF	5.260	1.057	0.835	0.634	1.117
LGBM	4.790	0.989	0.789	0.680	0.979
MLP	5.980	1.220	0.980	0.540	1.490
SEM	4.890	1.011	0.809	0.670	1.034
3D-CNN	0.017	0.381	0.289	0.910	N/A

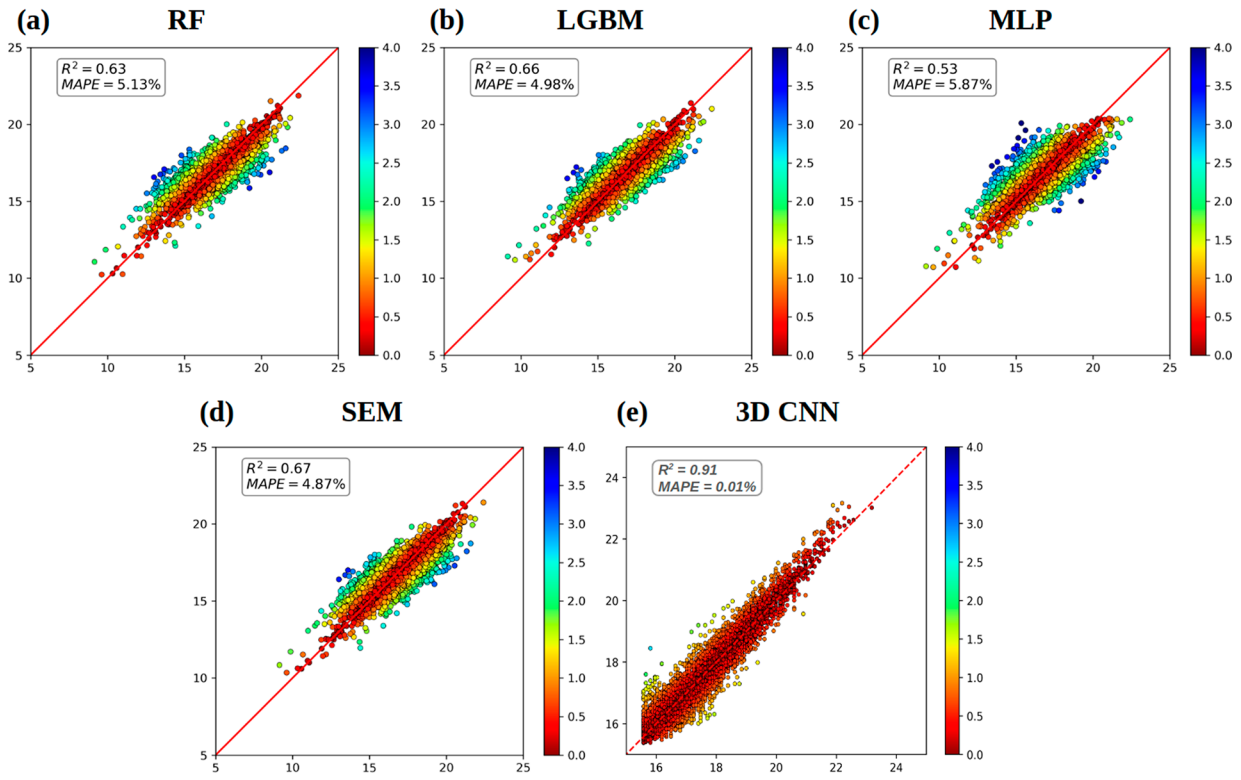


Figure 12. Testing accuracy of five models, *i.e.* a (a) Random forest (RF) model, (b) light gradient boosting machine (LGBM) model, (c) multilayer perceptron (MLP) model, (d) stacked ensemble model (SEM), and (e) three-dimensional convolutional neural network (3D-CNN) model, trained in this study.

https://doi.org/10.1021/acs.jctc.3c00081
J. Chem. Theory Comput. 2023, 19, 6686−6703

chemical descriptors achieve reliable predictive results.[87] This further demonstrates the capability of CNN models to achieve accurate predictions even with limited data sets (only ∼9000 MOFs (∼45,000 frames)).

3.7. Design Rules. According to the data generated via the PSO-integrated GA framework, we have formulated the following design rules and functionalization recommendations for the synthesis of these next-generation MOFs. These criteria may be applied to the design of experiments to test, rationalize, modify, and repurpose the current MOFs while providing new functional aspects specific to H_2 storage. A chemical rationale and the potentiality to modify functional groups based on the active site features specific to H_2 adsorption are described. Based on the data generated during the PSO-integrated GA framework, we provide the following design rules that may further assist the synthesis of these MOFs:

1 Functional group selection at low pressure: It has been reported that the high H_2 adsorption results from a higher specific surface area at low pressure and cryogenic temperature (77 K). Based on our data, the functional groups containing unsaturated carbon rings and carbon atoms (e.g., phenylacetylene, penta-1,3-diene, methyl, styrene, and ethyl), due to their high specific exposed surface area of carbons and unsaturated carbon bonds, may enhance the H_2 adsorption. Moreover, the longer linear chain alkenes, for example, decapentaene, could improve the H_2 adsorption by increasing the specific surface area by reaching the central region of voids at low pressures.

2 Selection of sites on MOFs: The top utilized functionalization sites were site numbers 24, 3, 14, 1, and 11. This was in accordance with the location of these sites being close to the metal node corner. This allows the enhanced entrapment of the H_2 molecules in the larger available pores.[88,89]

3 Selection of a combination of functional groups and site: the 4,4′-BPDC linker could be decorated with 1,3,5-Heptatriene and phenylacetylene at sites 2 and 4, respectively, in the *x*-direction. Subsequently, 1,3,5,7,9,11-Dodecahexaene on a linker in the *y*-direction at site 11 and phenylacetylene and Penta-1,3-diene on linkers in the *y*- and *z*-directions at sites 16 and 20 result in the higher volumetric adsorption. Approximately 28.5% of the top MOFs accounted for the above configuration in the top 200 MOFs selected on the basis of highest volumetric H_2 capacity.

4. CONCLUSIONS

We present a novel framework that seamlessly integrates molecular simulations (here, GCMC) with the PSO-integrated GA algorithm and in-house metal organic framework (MOF) structure generation code to design MOFs with enhanced H_2 uptake capacity. This framework reduced a large design space of $\sim 10^{37}$ MOFs created while functionalizing IRMOF-10 with 39 selected functional groups. Specifically, it was able to search for the combination of functionalization site and functional groups by exploring only 13,824 MOFs with the aid of the PSO-integrated GA algorithm. IRMOF-1 with the highest volumetric and gravimetric adsorption (*i.e.* 25.76 and 37.08 mg/g, respectively) was ∼6 times higher than those of bare/unfunctionalized IRMOF-10 (4.15 and 13.9 mg/g, respectively).

The in-depth analysis of the top 200 MOFs showed that unsaturated carbon atoms of long chain functional groups assisted in enhancing the H_2 adsorption because of their stronger interactions with H_2. Particularly, phenylacetylene, penta-1,3-diene, 1,3,5-Heptatriene, and 1,3,5,7,9,11-Dodecahexaene were the most occurring functional groups. Moreover, the adsorption isotherms revealed the highest volumetric and gravimetric deliverable capacity of functionalized IRMOF-10 to be of 23.0 g/L and 37.09 mg/g, respectively.

A proof-of-concept of training machine learning models to predict the H_2 adsorption, which in turn can be used as a surrogate for GCMC simulations, is also presented. ML models including random forests, light gradient boosting machine, and multilayer perceptron (artificial neural networks) were trained to achieve testing MAE values of 0.835 g/L (MAPE = 5.26%), 0.789 g/L (MAPE = 4.79%), and 0.980 g/L (MAPE = 5.98%), respectively. We also advanced our study further by training a stacked ensemble model, which improved the accuracy slightly by reducing the MAE value to 0.809 g/L (*i.e.* MAE reduced to 0.809 g/L and MAPE = 4.89%). However, 3-dimensional CNN, which incorporated both the spatial (or physical) features as well as chemical characteristics of functional groups attached to the MOFs performed the best and exhibited an MAE of 0.289 g/L (and MAPE = 0.017%). In the future, we envision integration of the ML methods with PSO-integrated GA models to discover new MOF candidates for hydrogen, carbon dioxide, ethane, and ethylene adsorption. Moreover, this novel computational framework and ML model development approach will be used to design new drug molecules and polysaccharides as well as polymer grafted nanoparticles for desired applications.

■ ASSOCIATED CONTENT

Data Availability Statement

The data sets generated and analyzed in the current article can be found at 10.5281/zenodo.7884470.

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jctc.3c00081.

Methodology; List of all functional groups; Comparison of PSO-integrated GA with traditional GA and randomly generated structures; Validation of convergence at 2000 GCMC steps; Results from optimization runs and insights on the structure of top performing MOFs; Deliverable H_2 capacity for top MOFs (PDF)

IRMOF structure generation code (ZIP)

PSO-integrated GA code (ZIP)

Movie S1, evolution of uptake capacity of MOFs w.r.t. MOF features (AVI)

Movie S2, evolution of uptake capacity of MOFs w.r.t. MOF features (AVI)

Movie S3, evolution of uptake capacity of MOFs w.r.t. MOF features (AVI)

Movie S4, evolution of uptake capacity of MOFs w.r.t. MOF features (AVI)

Movie S5, evolution of search space by PSO-integrated GA and traditional GA algorithm (MP4)

Movie S6, evolution of search space by PSO-integrated GA and traditional GA algorithm (MP4)

pdb files (ZIP)

https://doi.org/10.1021/acs.jctc.3c00081
J. Chem. Theory Comput. 2023, 19, 6686−6703

AUTHOR INFORMATION

Corresponding Author

Sanket A. Deshmukh − *Department of Chemical Engineering, Virginia Tech, Blacksburg, Virginia 24061, United States;* orcid.org/0000-0001-7573-0057; Phone: 1 540-231-8785; Email: sanketad@vt.edu

Authors

Samrendra K. Singh − *Department of Chemical Engineering, Virginia Tech, Blacksburg, Virginia 24061, United States*

Abhishek T. Sose − *Department of Chemical Engineering, Virginia Tech, Blacksburg, Virginia 24061, United States*

Fangxi Wang − *Department of Chemical Engineering, Virginia Tech, Blacksburg, Virginia 24061, United States*

Karteek K. Bejagam − *Department of Chemical Engineering, Virginia Tech, Blacksburg, Virginia 24061, United States;* orcid.org/0000-0002-8660-9946

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jctc.3c00081

Author Contributions

S.K.S. and A.T.S. contributed equally.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work was supported by GlycoMIP, a National Science Foundation Materials Innovation Platform funded through Cooperative Agreement DMR-1933525. This work was funded by the American Chemical Society Petroleum Research Fund [Grant No. 60488-DNI6]. The authors acknowledge Advanced Research Computing (ARC), Virginia Tech for providing the computational resources. The authors also acknowledge the resources from the National Energy Research Scientific Computing Center (NERSC) [Contract No. DE-AC02-05CH11231]. S.A.D. acknowledges the Institute for Critical Technology and Applied Science (ICTAS) Junior Faculty Award. S.A.D. acknowledges support from the NSF CAREER Award (Award No.: DMR-CMMT-2047743).

REFERENCES

(1) Zou, C.; Zhao, Q.; Zhang, G.; Xiong, B. Energy Revolution: From a Fossil Energy Era to a New Energy Era. *Natural Gas Industry B* **2016**, *3* (1), 1−11.

(2) Colbertaldo, P.; Agustin, S. B.; Campanari, S.; Brouwer, J. Impact of Hydrogen Energy Storage on California Electric Power System: Towards 100% Renewable Electricity. *Int. J. Hydrogen Energy* **2019**, *44* (19), 9558−9576.

(3) Yin, S. F.; Xu, B. Q.; Zhou, X. P.; Au, C. T. A Mini-Review on Ammonia Decomposition Catalysts for on-Site Generation of Hydrogen for Fuel Cell Applications. *Appl. Catal., A* **2004**, *277* (1), 1−9.

(4) Aminudin, M. A.; Kamarudin, S. K.; Lim, B. H.; Majilan, E. H.; Masdar, M. S.; Shaari, N. An Overview: Current Progress on Hydrogen Fuel Cell Vehicles. *Int. J. Hydrogen Energy* **2023**, *48*, 4371.

(5) Pramuanjaroenkij, A.; Kakaç, S. The Fuel Cell Electric Vehicles: The Highlight Review. *Int. J. Hydrogen Energy* **2023**, *48*, 9401.

(6) Raghavaiah, N. V.; Srinivasulu, G. N. Fuel Cells for Alternative and Sustainable Energy Systems. *Renewable Energy for Sustainable Growth Assessment*; Wiley: March 5, 2022; pp 363−388,.

(7) Nazir, H.; Muthuswamy, N.; Louis, C.; Jose, S.; Prakash, J.; Buan, M. E.; Flox, C.; Chavan, S.; Shi, X.; Kauranen, P.; Kallio, T.; Maia, G.; Tammeveski, K.; Lymperopoulos, N.; Carcadea, E.; Veziroglu, E.; Iranzo, A.; Kannan, A. M. Is the H2 Economy Realizable in the Foreseeable Future? Part II: H2 Storage, Transportation, and Distribution. *Int. J. Hydrogen Energy* **2020**, *45* (41), 20693−20708.

(8) Berro Ramirez, J. P.; Halm, D.; Grandidier, J.-C.; Villalonga, S.; Nony, F. 700 bar Type IV High Pressure Hydrogen Storage Vessel Burst - Simulation and Experimental Validation. *Int. J. Hydrogen Energy* **2015**, *40* (38), 13183−13192.

(9) Bobbitt, N. S.; Chen, J.; Snurr, R. Q. High-Throughput Screening of Metal-Organic Frameworks for Hydrogen Storage at Cryogenic Temperature. *J. Phys. Chem. C* **2016**, *120* (48), 27328−27341.

(10) Basdogan, Y.; Keskin, S. Simulation and Modelling of MOFs for Hydrogen Storage. *CrystEngComm* **2015**, *17* (2), 261−275.

(11) Suh, M. P.; Park, H. J.; Prasad, T. K.; Lim, D.-W. Hydrogen Storage in Metal-Organic Frameworks. *Chem. Rev.* **2012**, *112* (2), 782−835.

(12) Hu, Y. H.; Zhang, L. Hydrogen Storage in Metal-Organic Frameworks. *Adv. Mater.* **2010**, *22* (20), E117−E130.

(13) Yang, Q.; Zhong, C. Understanding Hydrogen Adsorption in Metal-Organic Frameworks with Open Metal Sites: A Computational Study. *J. Phys. Chem. B* **2006**, *110* (2), 655−658.

(14) Weitkamp, J.; Fritz, M.; Ernst, S. ZEOLITES AS MEDIA FOR HYDROGEN STORAGE. In *Proceedings from the Ninth International Zeolite Conference*; von Ballmoos, R., Higgins, J. B., Treacy, M. M. J., Eds.; Butterworth-Heinemann: 1993; pp 11−19,.

(15) Lim, K. L.; Kazemian, H.; Yaakob, Z.; Daud, W. R. W. Solid-State Materials and Methods for Hydrogen Storage: A Critical Review. *Chemical Engineering & Technology: Industrial Chemistry-Plant Equipment-Process Engineering-Biotechnology* **2010**, *33* (2), 213−226.

(16) Ren, J.; Langmi, H. W.; North, B. C.; Mathe, M. Review on Processing of Metal-Organic Framework (MOF) Materials towards System Integration for Hydrogen Storage. *Int. J. Energy Res.* **2015**, *39* (5), 607−620.

(17) Shet, S. P.; Shanmuga Priya, S.; Sudhakar, K.; Tahir, M. A Review on Current Trends in Potential Use of Metal-Organic Framework for Hydrogen Storage. *Int. J. Hydrogen Energy* **2021**, *46* (21), 11782−11803.

(18) Gangu, K. K.; Maddila, S.; Mukkamala, S. B.; Jonnalagadda, S. B. Characteristics of MOF, MWCNT and Graphene Containing Materials for Hydrogen Storage: A Review. *J. Mater. Chem. A Mater. Energy Sustain.* **2019**, *30*, 132−144.

(19) Purewal, J.; Veenstra, M.; Tamburello, D.; Ahmed, A.; Matzger, A. J.; Wong-Foy, A. G.; Seth, S.; Liu, Y.; Siegel, D. J. Estimation of System-Level Hydrogen Storage for Metal-Organic Frameworks with High Volumetric Storage Density. *Int. J. Hydrogen Energy* **2019**, *44* (29), 15135−15145.

(20) Gómez-Gualdrón, D. A.; Wang, T. C.; García-Holley, P.; Sawelewa, R. M.; Argueta, E.; Snurr, R. Q.; Hupp, J. T.; Yildirim, T.; Farha, O. K. Understanding Volumetric and Gravimetric Hydrogen Adsorption Trade-off in Metal-Organic Frameworks. *ACS Appl. Mater. Interfaces* **2017**, *9* (39), 33419−33428.

(21) Ahmed, A.; Liu, Y.; Purewal, J.; Tran, L. D.; Wong-Foy, A. G.; Veenstra, M.; Matzger, A. J.; Siegel, D. J. Balancing Gravimetric and Volumetric Hydrogen Density in MOFs. *Energy Environ. Sci.* **2017**, *10* (11), 2459−2471.

(22) Grünker, R.; Bon, V.; Müller, P.; Stoeck, U.; Krause, S.; Mueller, U.; Senkovska, I.; Kaskel, S. A New Metal-Organic Framework with Ultra-High Surface Area. *Chem. Commun.* **2014**, *50* (26), 3450−3452.

(23) Torrisi, A.; Bell, R. G.; Mellot-Draznieks, C. Functionalized MOFs for Enhanced CO2 Capture. *Cryst. Growth Des.* **2010**, *10* (7), 2839−2841.

(24) Yang, J.-M.; Liu, Q.; Sun, W.-Y. Co(II)-Doped MOF-5 Nano/microcrystals: Solvatochromic Behaviour, Sensing Solvent Molecules and Gas Sorption Property. *J. Solid State Chem.* **2014**, *218*, 50−55.

(25) Jia, X.-Y.; Song, Z.-Y.; Zhu, L.-J.; Xia, D.-H. Synthesis of Transition Metal-Doped MOF-5 and Their Application for Adsorptive Desulfurization. In *2nd Annual International Conference on Energy, Environmental & Sustainable Ecosystem Development*

(EESED 2016); Atlantis Press: Paris, France, 2017; DOI: 10.2991/eesed-16.2017.76.
(26) Sun, X.; Gu, X.; Xu, W.; Chen, W.-J.; Xia, Q.; Pan, X.; Zhao, X.; Li, Y.; Wu, Q.-H. Novel Hierarchical Fe(III)-Doped Cu-MOFs With Enhanced Adsorption of Benzene Vapor. *Front Chem.* **2019**, *7*, 652.
(27) Rowsell, J. L. C.; Millward, A. R.; Park, K. S.; Yaghi, O. M. Hydrogen Sorption in Functionalized Metal-Organic Frameworks. *J. Am. Chem. Soc.* **2004**, *126* (18), 5666−5667.
(28) Deng, H.; Doonan, C. J.; Furukawa, H.; Ferreira, R. B.; Towne, J.; Knobler, C. B.; Wang, B.; Yaghi, O. M. Multiple Functional Groups of Varying Ratios in Metal-Organic Frameworks. *Science* **2010**, *327* (5967), 846−850.
(29) Chung, Y. G.; Gómez-Gualdrón, D. A.; Li, P.; Leperi, K. T.; Deria, P.; Zhang, H.; Vermeulen, N. A.; Fraser Stoddart, J.; You, F.; Hupp, J. T.; Farha, O. K.; Snurr, R. Q. In Silico Discovery of Metal-Organic Frameworks for Precombustion CO2 Capture Using a Genetic Algorithm. *Science Advances* **2016**, *2*, e1600909.
(30) Sose, A. T.; Joshi, S. Y.; Kunche, L. K.; Wang, F.; Deshmukh, S. A. A Review of Recent Advances and Applications of Machine Learning in Tribology. *Phys. Chem. Chem. Phys.* **2023**, *25* (6), 4408−4443.
(31) Collins, S. P.; Daff, T. D.; Piotrkowski, S. S.; Woo, T. K. Materials Design by Evolutionary Optimization of Functional Groups in Metal-Organic Frameworks. *Sci. Adv.* **2016**, *2* (11), e1600954.
(32) Bejagam, K. K.; Singh, S.; An, Y.; Deshmukh, S. A. Machine-Learned Coarse-Grained Models. *J. Phys. Chem. Lett.* **2018**, *9*, 4667−4672.
(33) Dashti, A.; Bahrololoomi, A.; Amirkhani, F.; Mohammadi, A. H. Estimation of CO2 Adsorption in High Capacity Metal-organic Frameworks: Applications to Greenhouse Gas Control. *Journal of CO2 Utilization* **2020**, *41*, 101256.
(34) Qiao, Z.; Yan, Y.; Tang, Y.; Liang, H.; Jiang, J. Metal-Organic Frameworks for Xylene Separation: From Computational Screening to Machine Learning. *J. Phys. Chem. C* **2021**, *125* (14), 7839−7848.
(35) Beauregard, N.; Pardakhti, M.; Srivastava, R. In Silico Evolution of High-Performing Metal Organic Frameworks for Methane Adsorption. *J. Chem. Inf. Model.* **2021**, *61* (7), 3232−3239.
(36) Plimpton, S.; Crozier, P.; Thompson, A. *LAMMPS-Large-Scale Atomic/molecular Massively Parallel Simulator*; Sandia National Laboratories: 2007; Vol. *18*, p 43.
(37) Thompson, A. P.; Aktulga, H. M.; Berger, R.; Bolintineanu, D. S.; Brown, W. M.; Crozier, P. S.; in 't Veld, P. J.; Kohlmeyer, A.; Moore, S. G.; Nguyen, T. D.; Shan, R.; Stevens, M. J.; Tranchida, J.; Trott, C.; Plimpton, S. J. LAMMPS - a Flexible Simulation Tool for Particle-Based Materials Modeling at the Atomic, Meso, and Continuum Scales. *Comput. Phys. Commun.* **2022**, *271*, 108171.
(38) Rappe, A. K.; Casewit, C. J.; Colwell, K. S.; Goddard, W. A.; Skiff, W. M. UFF, a Full Periodic Table Force Field for Molecular Mechanics and Molecular Dynamics Simulations. *J. Am. Chem. Soc.* **1992**, *114*, 10024−10035.
(39) Sose, A. T.; Cornell, H. D.; Gibbons, B. J.; Burris, A. A.; Morris, A. J.; Deshmukh, S. A. Modelling Drug Adsorption in Metal-organic Frameworks: The Role of Solvent. *RSC Adv.* **2021**, *11*, 17064−17071.
(40) Wang, F.; Sose, A. T.; Singh, S. K.; Deshmukh, S. A. Dual-Force Zone Nonequilibrium Molecular Dynamics Simulations on Nanoporous Metal-Organic Framework Membranes for Separation of H2/CH4Mixtures. *ACS Appl. Nano Mater.* **2022**, *5* (3), 4048−4061.
(41) Lorentz, H. A. Ueber die Anwendung des Satzes vom Virial in der kinetischen Theorie der Gase. *Ann. Phys.* **1881**, *248* (1), 127−136.
(42) Berthelot, D.; Amoreux, G. *Acad. Sci.* **1898**, *126*, 1703−1855.
(43) Fischer, M.; Hoffmann, F.; Fröba, M. Preferred Hydrogen Adsorption Sites in Various MOFs-a Comparative Computational Study. *ChemPhysChem* **2009**, *10* (15), 2647−2657.
(44) Chen, Y.; Bai, X.; Liu, D.; Fu, X.; Yang, Q. High-Throughput Computational Exploration of MOFs with Open Cu Sites for Adsorptive Separation of Hydrogen Isotopes. *ACS Appl. Mater. Interfaces* **2022**, *14* (21), 24980−24991.
(45) Gómez-Gualdrón, D. A.; Colón, Y. J.; Zhang, X.; Wang, T. C.; Chen, Y.-S.; Hupp, J. T.; Yildirim, T.; Farha, O. K.; Zhang, J.; Snurr, R. Q. Evaluating Topologically Diverse Metal-organic Frameworks for Cryo-Adsorbed Hydrogen Storage. *Energy Environ. Sci.* **2016**, *9* (10), 3279−3289.
(46) Ming, Y.; Purewal, J.; Yang, J.; Xu, C.; Veenstra, M.; Gaab, M.; Müller, U.; Siegel, D. J. Stability of MOF-5 in a Hydrogen Gas Environment Containing Fueling Station Impurities. *Int. J. Hydrogen Energy* **2016**, *41* (22), 9374−9382.
(47) Eberle, U.; Felderhoff, M.; Schüth, F. Chemical and Physical Solutions for Hydrogen Storage. *Angew. Chem., Int. Ed. Engl.* **2009**, *48* (36), 6608−6630.
(48) Jaramillo, D. E.; Jiang, H. Z. H.; Evans, H. A.; Chakraborty, R.; Furukawa, H.; Brown, C. M.; Head-Gordon, M.; Long, J. R. Ambient-Temperature Hydrogen Storage via Vanadium(II)-Dihydrogen Complexation in a Metal-Organic Framework. *J. Am. Chem. Soc.* **2021**, *143* (16), 6248−6256.
(49) Hou, X. X.; Sulic, M.; Ortmann, J. P.; Cai, M.; Chakraborty, A. Experimental and Numerical Investigation of the Cryogenic Hydrogen Storage Processes over MOF-5. *Int. J. Hydrogen Energy* **2016**, *41* (6), 4026−4038.
(50) Brunner, T.; Kircher, O. Cryo-Compressed Hydrogen Storage. In *Hydrogen Science and Engineering: Materials, Processes, Systems and Technology*; Wiley-VCH Verlag GmbH & Co. KGaA: Weinheim, Germany, 2016; pp 711−732,.
(51) Aceves, S. M.; Espinosa-Loza, F.; Ledesma-Orozco, E.; Ross, T. O.; Weisberg, A. H.; Brunner, T. C.; Kircher, O. High-Density Automotive Hydrogen Storage with Cryogenic Capable Pressure Vessels. *Int. J. Hydrogen Energy* **2010**, *35* (3), 1219−1226.
(52) Barthelemy, H.; Weber, M.; Barbier, F. Hydrogen Storage: Recent Improvements and Industrial Perspectives. *Int. J. Hydrogen Energy* **2017**, *42* (11), 7254−7262.
(53) Ahmed, A.; Seth, S.; Purewal, J.; Wong-Foy, A. G.; Veenstra, M.; Matzger, A. J.; Siegel, D. J. Exceptional Hydrogen Storage Achieved by Screening Nearly Half a Million Metal-Organic Frameworks. *Nat. Commun.* **2019**, *10* (1), 1568.
(54) Manz, T. A.; Sholl, D. S. Chemically Meaningful Atomic Charges That Reproduce the Electrostatic Potential in Periodic and Nonperiodic Materials. *J. Chem. Theory Comput.* **2010**, *6* (8), 2455−2468.
(55) Ryan, P.; Broadbelt, L. J.; Snurr, R. Q. Is Catenation Beneficial for Hydrogen Storage in Metal-Organic Frameworks? *Chem. Commun.* **2008**, No. 35, 4132−4134.
(56) Berman, D.; Deshmukh, S. A.; Sankaranarayanan, S. K. R. S.; Erdemir, A.; Sumant, A. V. Friction. Macroscale Superlubricity Enabled by Graphene Nanoscroll Formation. *Science* **2015**, *348* (6239), 1118−1122.
(57) Forrest, S. Genetic Algorithms: Principles of Natural Selection Applied to Computation. *Science* **1993**, *261* (5123), 872−878.
(58) Patra, T. K.; Meenakshisundaram, V.; Hung, J.-H.; Simmons, D. S. Neural-Network-Biased Genetic Algorithms for Materials Design: Evolutionary Algorithms That Learn. *ACS Comb. Sci.* **2017**, *19* (2), 96−107.
(59) Lameijer, E.-W.; Bäck, T.; Kok, J. N.; Ijzerman, A. D. P. Evolutionary Algorithms in Drug Design. *Nat. Comput.* **2005**, *4* (3), 177−243.
(60) Bejagam, K. K.; Singh, S.; An, Y.; Berry, C.; Deshmukh, S. A. PSO-Assisted Development of New Transferable Coarse-Grained Water Models. *J. Phys. Chem. B* **2018**, *122* (6), 1958−1971.
(61) Kennedy, J.; Eberhart, R. Particle Swarm Optimization. In *Proceedings of ICNN'95 - International Conference on Neural Networks*; 1995; Vol. *4*, pp 1942−1948.
(62) Sose, A. T.; Mohammadi, E.; Achari, P. F.; Deshmukh, S. A. Determination of Accurate Interaction Parameters between the Molybdenum Disulfide and Water to Investigate Their Interfacial Properties. *J. Phys. Chem. C* **2022**, *126* (4), 2013−2022.
(63) Ru, N.; Jianhua, Y. A GA and Particle Swarm Optimization Based Hybrid Algorithm. In *2008 IEEE Congress on Evolutionary*

Computation (IEEE World Congress on Computational Intelligence); 2008; pp 1047−1050.
(64) Yang, B.; Chen, Y.; Zhao, Z. A Hybrid Evolutionary Algorithm by Combination of PSO and GA for Unconstrained and Constrained Optimization Problems. In *2007 IEEE International Conference on Control and Automation*; 2007; pp 166−170.
(65) El-Shafiey, M. G.; Hagag, A.; El-Dahshan, E.-S. A.; Ismail, M. A. A Hybrid GA and PSO Optimized Approach for Heart-Disease Prediction Based on Random Forest. *Multimed. Tools Appl.* **2022**, *81* (13), 18155−18179.
(66) Shi, L.; Gong, J.; Zhai, C. Application of a Hybrid PSO-GA Optimization Algorithm in Determining Pyrolysis Kinetics of Biomass. *Fuel* **2022**, *323*, 124344.
(67) Manasrah, A. M.; Ba Ali, H. Workflow Scheduling Using Hybrid GA-PSO Algorithm in Cloud Computing. *Wireless Communications and Mobile Computing* **2018**, *2018*, 1934784.
(68) Ali, A. F.; Tawhid, M. A. A Hybrid Particle Swarm Optimization and Genetic Algorithm with Population Partitioning for Large Scale Optimization Problems. *Ain Shams Engineering Journal* **2017**, *8* (2), 191−206.
(69) Singh, N.; Singh, S. B. Hybrid Algorithm of Particle Swarm Optimization and Grey Wolf Optimizer for Improving Convergence Performance. *J. Appl. Math.* **2017**, *2017*, 2030489.
(70) Breiman, L. *Classification and Regression Trees*; Routledge: 2017.
(71) Abdi, J.; Hadavimoghaddam, F.; Hadipoor, M.; Hemmati-Sarapardeh, A. Modeling of CO2 Adsorption Capacity by Porous Metal Organic Frameworks Using Advanced Decision Tree-Based Models. *Sci. Rep.* **2021**, *11* (1), 24468.
(72) Abdi, J.; Mazloom, G. Machine Learning Approaches for Predicting Arsenic Adsorption from Water Using Porous Metal-Organic Frameworks. *Sci. Rep.* **2022**, *12* (1), 16458.
(73) Anderson, R.; Biong, A.; Gómez-Gualdrón, D. A. Adsorption Isotherm Predictions for Multiple Molecules in MOFs Using the Same Deep Learning Model. *J. Chem. Theory Comput.* **2020**, *16* (2), 1271−1283.
(74) Guo, W.; Liu, J.; Dong, F.; Chen, R.; Das, J.; Ge, W.; Xu, X.; Hong, H. Deep Learning Models for Predicting Gas Adsorption Capacity of Nanomaterials. *Nanomaterials (Basel)* **2022**, *12* (19), 3376.
(75) Singh, S. K.; Bejagam, K. K.; An, Y.; Deshmukh, S. A. Machine-Learning Based Stacked Ensemble Model for Accurate Analysis of Molecular Dynamics Simulations. *J. Phys. Chem. A* **2019**, *123* (24), 5190−5198.
(76) Wang, G.; Hao, J.; Ma, J.; Jiang, H. A Comparative Assessment of Ensemble Learning for Credit Scoring. *Expert Syst. Appl.* **2011**, *38* (1), 223−230.
(77) Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; Others. Scikit-Learn: Machine Learning in Python. *Journal of Machine Learning Research* **2011**, *12*, 2825−2830.
(78) Ke, G.; Meng, Q.; Finley, T.; Wang, T.; Chen, W.; Ma, W.; Ye, Q.; Liu, T.-Y.; Research, M.; University, P.; Redmond, M. *LightGBM: A highly efficient gradient boosting decision tree*. https://proceedings.neurips.cc/paper/2017/file/6449f44a102fde848669bdd9eb6b76fa-Paper.pdf (accessed 2023-01-09).
(79) Joshi, S. Y.; Singh, S.; Deshmukh, S. A. Coarse-Grained Molecular Dynamics Integrated with Convolutional Neural Network for Comparing Shapes of Temperature Sensitive Bottlebrushes. *npj Computational Materials* **2022**, *8* (1), 45.
(80) *Materials-based hydrogen storage*. Energy.gov. https://www.energy.gov/eere/fuelcells/materials-based-hydrogen-storage (accessed 2023-01-18).
(81) Nguyen-Thuy, T.; Le-Hoang, P.; Hoang Vu, N.; Le, T. N.-M.; Le Hoang Doan, T.; Kuo, J.-L.; Nguyen, T. T.; Phan, T. B.; Nguyen-Manh, D. Hydrogen Adsorption Mechanism of MOF-74 Metal-Organic Frameworks: An Insight from First Principles Calculations. *RSC Adv.* **2020**, *10* (72), 43940−43949.
(82) Tsivion, E.; Long, J. R.; Head-Gordon, M. Hydrogen Physisorption on Metal-Organic Framework Linkers and Metalated Linkers: A Computational Study of the Factors That Control Binding Strength. *J. Am. Chem. Soc.* **2014**, *136* (51), 17827−17835.
(83) Xia, Y.; Yang, Z.; Zhu, Y. Porous Carbon-Based Materials for Hydrogen Storage: Advancement and Challenges. *J. Mater. Chem. A Mater. Energy Sustain.* **2013**, *1* (33), 9365−9381.
(84) Bobbitt, N. S.; Chen, J.; Snurr, R. Q. High-Throughput Screening of Metal-Organic Frameworks for Hydrogen Storage at Cryogenic Temperature. *J. Phys. Chem. C* **2016**, *120* (48), 27328−27341.
(85) Ren, J.; Wang, S.; Bi, K.; Cheng, M.; Liu, C.; Zhou, L.; Xue, X.; Ji, X. Machine Learning-Enabled Framework for High-Throughput Screening of MOFs: Application in Radon/Indoor Air Separation. *ACS Appl. Mater. Interfaces* **2023**, *15* (1), 1305−1316.
(86) Argueta, E.; Shaji, J.; Gopalan, A.; Liao, P.; Snurr, R. Q.; Gómez-Gualdrón, D. A. Molecular Building Block-Based Electronic Charges for High-Throughput Screening of Metal-Organic Frameworks for Adsorption Applications. *J. Chem. Theory Comput.* **2018**, *14* (1), 365−376.
(87) Pardakhti, M.; Moharreri, E.; Wanik, D.; Suib, S. L.; Srivastava, R. Machine Learning Using Combined Structural and Chemical Descriptors for Prediction of Methane Adsorption Performance of Metal Organic Frameworks (MOFs). *ACS Comb. Sci.* **2017**, *19* (10), 640−645.
(88) Stoddart, A. Predicting Perfect Pores. *Nature Reviews Materials* **2020**, *5* (5), 331−331.
(89) Zhang, X.; Lin, R.-B.; Wang, J.; Wang, B.; Liang, B.; Yildirim, T.; Zhang, J.; Zhou, W.; Chen, B. Optimization of the Pore Structures of MOFs for Record High Hydrogen Volumetric Working Capacity. *Adv. Mater.* **2020**, *32* (17), No. 1907995.

https://doi.org/10.1021/acs.jctc.3c00081
J. Chem. Theory Comput. 2023, 19, 6686−6703