In silico discovery of metal-organic frameworks for precombustion CO\textsubscript{2} capture using a genetic algorithm


Discovery of new adsorbent materials with a high CO\textsubscript{2} working capacity could help reduce CO\textsubscript{2} emissions from newly commissioned power plants using precombustion carbon capture. High-throughput computational screening efforts can accelerate the discovery of new adsorbents but sometimes require significant computational resources to explore the large space of possible materials. We report the in silico discovery of high-performing adsorbents for precombustion CO\textsubscript{2} capture by applying a genetic algorithm to efficiently search a large database of metal-organic frameworks (MOFs) for top candidates. High-performing MOFs identified from the in silico search were synthesized and activated and show a high CO\textsubscript{2}/H\textsubscript{2} selectivity. One of the synthesized MOFs shows a higher CO\textsubscript{2} working capacity than any MOF reported in the literature under the operating conditions investigated here.

INTRODUCTION

Scientists, political leaders, and common citizens around the world are increasingly alarmed by the rapidly rising levels of CO\textsubscript{2} in the atmosphere (1–3). In the United States, nearly 40% of CO\textsubscript{2} emissions come from burning fossil fuels to generate electricity in power plants (4). Because renewable energy sources (such as wind and solar) are still far from replacing fossil fuels as the primary energy source to power the planet, almost all recent scenarios put forth to reduce CO\textsubscript{2} emissions include a significant midterm role for carbon capture and storage. Capturing CO\textsubscript{2} from existing power plants requires the separation of dilute amounts of CO\textsubscript{2} from N\textsubscript{2} in a low-pressure stream via a so-called “postcombustion” strategy. An easier strategy for newly commissioned power plants is to use “precombustion” CO\textsubscript{2} capture technology. An example of precombustion CO\textsubscript{2} capture is shown in Fig. 1 where natural gas is reformed to produce syngas (a mixture of CO and H\textsubscript{2}), which is run through a water-gas shift reaction (WGSR) to produce a high-pressure stream of CO\textsubscript{2} and H\textsubscript{2}. The CO\textsubscript{2} is then removed from this stream, and the resulting H\textsubscript{2} is burned for energy production, with water as a by-product (5). Currently available precombustion CO\textsubscript{2} capture technology using solvents, such as Selexol, methanol, or methylthanolamine, is estimated to cost around $60 per metric ton of captured CO\textsubscript{2} (6), which is 50% higher than the U.S. Department of Energy target. On the other hand, if pressure-swing adsorption (PSA) were used to capture the CO\textsubscript{2} from the high-pressure gas mixture obtained from the WGSR (Fig. 1), the cost of CO\textsubscript{2} capture could be reduced if a selective adsorbent material with a high CO\textsubscript{2} working capacity were available. Increasing the CO\textsubscript{2} working capacity means that, for instance, less adsorbent material is needed for the PSA unit, which in turn reduces the cost of CO\textsubscript{2} capture.

Metal-organic frameworks (MOFs) are a class of nanoporous materials that could potentially provide higher CO\textsubscript{2} working capacities for precombustion CO\textsubscript{2} capture than traditional sorbents, such as zeolites and activated carbons, because of their high pore volumes and surface areas. MOFs are synthesized by the self-assembly of organic and inorganic building blocks, and different combinations of the building blocks can produce MOFs with different physical and chemical properties, making this class of materials incredibly versatile and tunable for a wide range of applications (7–9). For instance, MOFs have been investigated for a wide range of gas storage and separation applications (10–13), but only a limited number of MOFs [for example, Cu-BTTrI (14, 15), Mg-MOF-74 (16), and Ni-4PyC (17)] have been tested for precombustion CO\textsubscript{2} capture. Given the large number of MOFs synthesized to date (18), experimental evaluation of all MOFs for this application would be impractical at best, and other approaches must be used to identify promising materials.

High-throughput computational screening has emerged in the past few years as a powerful approach to accelerating the evaluation of adsorbent materials for postcombustion CO\textsubscript{2} capture (19, 20), methane storage (21–24), Xe/Kr separation (25, 26), H\textsubscript{2} storage (27), and biofuel and hydrocarbon separations (28). In this approach, molecular simulations are carried out to evaluate the performance of existing or not-yetsynthesized adsorbent materials to find top-performing candidates and to reveal key structureproperty relationships between material performance and physical characteristics (for example, pore volume and enthalpy of adsorption). In one example, these computational screening efforts led to the synthesis of a new MOF for methane storage (24). In other cases, computational screening was used to find the performance limits of MOFs for applications such as methane and hydrogen storage (23, 29). To date, evaluation of the materials has been relatively fast, and hundreds of thousands of hypothetical materials could be screened because the guest species of interest were small molecules accurately described by classical force fields. However, there are millions of potential MOFs, and for many applications, calculating the performance of each material will be more time-consuming, for example, if quantum mechanical calculations are required (30). The current efforts have calculated the properties of all candidate materials in a given set of materials. However, many of these candidates have low performance, and much time is wasted in evaluating
performing materials was identified, synthesized, activated, and tested. Experimental pure-component CO2 and H2 isotherms and mixture isotherms on the activated MOF showed good agreement with the simulation predictions.

RESULTS

Validation of the GA
We implemented a GA search strategy as described in Materials and Methods and Fig. 2 (A and B) and applied it to a large database of 51,163 hMOFs (21). Before applying the GA to find high-performing hMOFs for precombustion CO2 capture, we tested the efficiency and robustness of our GA implementation by trying to find the hMOFs with the highest gravimetric and volumetric surface areas and methane working capacity. Note that these properties were previously calculated for all hMOFs; thus, we already knew the identity of the best materials for these test cases. Figure 2 (C to E) shows the histograms of methane working capacities and gravimetric and volumetric surface areas, respectively, for all hMOFs and for the initial population of 100 hMOFs used in all GA runs. These properties were used as different measures of hMOF “fitness” that the GA should attempt to improve. For each fitness measure, 100 independent GA runs were carried out for 100 generations each. The histograms of the performance of the best hMOF at the end of each GA run are shown in Fig. 2 (F to H). For each of the three performance measures, the GA always found a structure within the top 4% by the 10th generation.

Application of the GA
Starting with the same initial hMOF population that was used for the validation of the GA, the algorithm was applied to search for top hMOFs for precombustion CO2 capture. We note that, inherent to the GA formalism, it is not possible to determine whether the best hMOF in the database is identified through the GA search. However, as an objective measure of the success of our GA approach, we sought to identify MOFs with better performance metrics than those reported for MOFs to date. Three independent GA runs were performed to separately optimize three different fitness measures, namely, the CO2 working capacity (ΔNads), the CO2/H2 selectivity (αCO2H2), and an adsorbent performance score (APS), which is the product of the former two quantities, as defined in Materials and Methods. Each GA was run for 10 generations, and GCMC simulations were carried out for each new hMOF considered. To improve the computational efficiency, the GCMC results were saved, and if an hMOF in the nth generation was already evaluated in a previous generation (from any of the three runs), no new GCMC simulations were carried out for that structure.

Details of the GA runs are provided in section S3. Figure 3 summarizes the progress of the GA as it searched for top-performing hMOFs for each fitness measure. For each generation, the average fitness of the population and the fitness of the best-performing hMOF from the population (the elite) are plotted. As the GA evolved, for the top-performing MOF from each generation, the CO2 working capacity improved from ca. 7 mol/kg (1st generation) to ca. 8 mol/kg (10th generation), the CO2/H2 selectivity improved from ca. 700 to ca. 2600, and the APS improved from ca. 1000 mol/kg to ca. 1200 mol/kg.

The progression of genes during the GA run to optimize the APS is shown in Fig. 4. This figure shows that the building blocks that consistently result in top-performing hMOFs become dominant as the GA progresses. As the new generations of hMOFs are evolved, zinc paddlewheel nodes (gene “1”) from Fig. 4A and [1,1′,4′,1’”]terphenyl-3,3′,5,5′”-tetracarboxylic (TPTC) acid linkers (gene “38”) from Fig. 4, B and C) become dominant. On the other hand, no specific functional
group becomes dominant, which suggests that the choice of organic linkers and inorganic building blocks plays a more important role in optimizing the APS of MOFs for this application.

Table 1 shows that the number of GCMC simulations carried out during each GA run is significantly smaller than what would be required for an exhaustive search of all hMOFs. Each GA run required less than 1% of the computational time compared to a brute force search. Only 730 of 51,163 hMOFs were evaluated even if the three GA runs, together, were considered. Note that the GA is not guaranteed to find the very best solution (here, the best hMOF from the entire database). However, the GA validation tests illustrated in Fig. 2 (C to H) suggest that the top-performing hMOFs that were identified should be within the top 4% for each fitness measure (see section S3.)

Identification and synthesis of the top-performing MOFs

The three GA runs produced data for the CO₂ working capacity, the CO₂/H₂ selectivity, and the APS of 730 genetically unique hMOFs. Figure 5 summarizes the data obtained from all GA runs, where CO₂/H₂ selectivity is plotted as a function of CO₂ working capacity, with the color of the data points indicating the APS value. The top-performing hMOFs based on the CO₂ working capacity have void fractions between 0.6 and 0.8 and pore diameters between 8 and 10 Å. The MOFs with the highest CO₂/H₂ selectivity have lower void fractions...
fractions (<0.5) and smaller pore diameters (<5 Å), and consequently, they have a low CO2 working capacity because the pore space is filled before the pressure reaches 20 bar. Figure 5 shows that there is a clear trade-off between CO2 working capacity and CO2/H2 selectivity.

The APS aims to account for the effect of both the selectivity and the working capacity of the adsorbent on the purity and recovery of the PSA process. hMOFs with high APS values were considered for possible synthesis. In particular, we focused on hMOFs that were both located near the Pareto front in Fig. 5 and had better working capacities than the experimentally tested MOFs also shown in Fig. 5.

In this way, we obtained a “preliminary” list of nearly 50 high-performing hMOFs. From this list, we identified 12 MOFs based on the nbo topology, which combines metal paddlewheels and planar tetracarboxylate organic linkers, as materials that we anticipated we could successfully synthesize, on the basis of our previous experience in MOF synthesis. These 12 MOFs correspond to 6 zinc-based MOFs and their 6 copper-based counterparts. Before synthesis, the adsorption properties of these 12 hMOFs were recomputed using density functional theory (DFT)–derived partial atomic charges (34) for the hMOF atoms [approximate charges (35) were used for the GA screening]. We also arranged the functional groups in each of these hMOFs so that they were uniformly placed on each constituent organic linker in the material (section S5). From these simulations, we identified the TPTC acid linker [one of the linkers (gene 38) that was dominant in Fig. 4] functionalized with two ethoxy groups to optimize the APS value for both the copper- and zinc-based hMOFs (table S3).

For both the copper- and zinc-based hMOFs having this linker, we predicted high working capacities (5.7 mol/kg) and high selectivities (132 and 188 for copper and zinc cases, respectively). Because of the anticipated higher stability of copper paddlewheels upon activation, we selected the copper hMOF as the synthesis target. The selected hMOF is a functionalized version of a previously synthesized MOF, NOTT-101 (36), and our simulations predict it to have a CO2 working capacity that is higher than that reported for the few materials previously experimentally tested for precombustion CO2 capture (see Fig. 5).

The ethoxy-functionalized version of NOTT-101 (NOTT-101/OEt) was synthesized and activated. It should be noted that, unbeknown

---

**Table 1. Comparison of computational effort for brute force search versus GA.** $\Delta N_1$ is the CO$_2$ working capacity, $\alpha_{ads}$ is the CO$_2$/H$_2$ selectivity, and APS is the adsorbent performance score, as defined in Eqs. 1 to 3. The number of GCMC simulations for the GA search corresponds to the number of simulations carried out up to 10 generations.

<table>
<thead>
<tr>
<th>Method</th>
<th>Fitness measure</th>
<th>Number of GCMC simulations</th>
<th>Relative computational time (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brute force</td>
<td>—</td>
<td>51,163</td>
<td>100</td>
</tr>
<tr>
<td>GA</td>
<td>$\Delta N_1$</td>
<td>340</td>
<td>0.66</td>
</tr>
<tr>
<td></td>
<td>$\alpha_{ads}$</td>
<td>322</td>
<td>0.63</td>
</tr>
<tr>
<td></td>
<td>APS</td>
<td>268</td>
<td>0.52</td>
</tr>
</tbody>
</table>
The CO2 and H2 adsorption isotherms of NOTT-101/OEt were measured experimentally up to 16 bar at 313 K. Figure 6C shows the comparison between the experimental and simulated absolute adsorption isotherms for NOTT-101/OEt. In the figure, the experimental isotherms were multiplied by a factor of 1.09 to account for the 92% pore activation. There is good agreement between measured and simulated H2 isotherms, with only a slight (in absolute terms) underestimation by the simulation. There is fair agreement between measured and simulated CO2 isotherms, with simulations somewhat overpredicting the CO2 uptake, especially at the intermediate pressure range. Note that actual measured data instead of scaled-up data are presented and discussed in subsequent sections.

Application of structure-property relationships to find candidates in a separate database

High-throughput computational screening produces large volumes of data that can be used to find underlying structure-property relationships, such as how the performance for a given application depends on the MOF surface area, void fraction, etc. From the GA runs, structure-property relationships also emerged but with significantly fewer computations than with a brute force approach (Table 1). In principle, the structure-property relationships emerging from a GA screening could be used to find additional high-performing MOFs without the need for further simulations. To test the applicability of this approach, we searched for high-performing MOFs in a separate database: the computation-ready, experimental (CoRE) MOF database (18). Two key advantages of the 5109 structures in the CoRE MOF database are that all of them have already been synthesized and that their synthesis protocols are available in the literature, which can facilitate the synthesis and testing of any candidates identified from computational screening or other methods.

We identified ranges of optimal physical properties (pore-limiting diameter, largest cavity diameter, gravimetric surface area, and helium void fraction) for each performance measure on the basis of the properties of the hMOFs within the top 1% of the 730 hMOFs evaluated during the GA runs (table S4). These properties were then used to identify 75, 99, and 357 candidate CoRE MOFs for high CO2 working capacity, CO2/H2 selectivity, and APS, respectively. For each group of CoRE MOFs, GCMC simulations were carried out to evaluate their adsorption properties, and the results showed that 5 (of 75), 14 (of 99), and 13 (of 357) of these CoRE MOFs have a high working capacity, a high selectivity, and a high APS, respectively (see section S5). Note that although the “hit” rate was low (for example, only 5 of 75 candidates had a high working capacity), these properties were still useful in identifying high-performing CoRE MOFs without having to evaluate the full CoRE MOF database. One of the identified high-performing CoRE MOFs [Cambridge Structural Database REFCODE: VEXTUO (39)] with a predicted CO2 working capacity of 6.0 mol/kg and a CO2/H2 selectivity of 83 was selected for synthesis, activation, and testing. Note that although all CoRE MOFs have synthesis protocols available, successful activation is not guaranteed for all of these MOFs (18). Therefore, the comparison of reported BET areas and the geometrically calculated surface areas was also a factor in deciding which MOF to synthesize, because marked differences between the two values could indicate a tendency of the MOF to collapse upon activation or difficulty in removing trapped solvents or other impurities. VEXTUO was synthesized following the protocol in the literature (39). Experimental and simulated BET areas were 1977 and 2031 m²/g, respectively, and the measured and simulated pore volumes were 0.75 and 0.78 cm³/g, respectively. Single-component isotherms of CO2 and H2 were measured at 303 K. The simulated and experimental H2 isotherms were in good agreement, but the simulated CO2 isotherms were ~35% higher (at 15 bar) than the measured isotherms for VEXTUO (see section S9).

Comparison with other adsorbents

Table 2 summarizes a comparison among the two MOFs from this work and three high-performing MOFs (Mg-MOF-74, Cu-BTTr, and Ni-4PyC) known from the literature (all MOFs in Table 2 are illustrated in Fig. 6D). We used IAST to compute the mixture isotherms from high-pressure (up to 16 to 20 bar) experimental single-component CO2 and H2 isotherms of the MOFs in Table 2 (see
For NOTT-101/OEt and VEXTUO, experimental data were obtained in this work, whereas the experimental isotherms for Mg-MOF-74 and Cu-BTTri were obtained from the literature. Process simulations of an example PSA unit (see section S4) show that CO$_2$/H$_2$ selectivities higher than 30 are enough to achieve 99% H$_2$ purity, and higher working capacities reduce the amount of adsorbent required for the separation.

NOTT-101/OEt has the highest CO$_2$ working capacity among the five MOFs and a relatively high CO$_2$/H$_2$ selectivity of 60. Previously, Mg-MOF-74 has been noted for its high CO$_2$ working capacity (14). However, there is some question about how to define the working capacity for PSA processes. Here, we introduce a definition of the CO$_2$ working capacity (see Materials and Methods) based on the gas composition profiles from a PSA process simulation (40) under realistic operating conditions (see section S4). On the basis of this definition, the CO$_2$ working capacity of Mg-MOF-74 is the lowest among the listed MOFs, whereas its CO$_2$/H$_2$ selectivity is the highest. However, because our process modeling shows that selectivities higher than 30 are enough to reach >99% purity, MOFs such as NOTT-101/OEt (with an approximately 62% higher working capacity with respect to Mg-MOF-74) could be better suited for precombustion carbon capture. Similarly, both VEXTUO (which was identified from the CoRE MOF database) and Ni-4PyC (which was reported during the preparation of this manuscript) also have working capacities higher than that of Mg-MOF-74 (ca. 19% and ca. 31%, respectively), while having selectivities higher than 30. Note that our estimation of the Ni-4PyC working capacity is approximate, on the basis of the available experimental and simulation data reported by Nandi et al. (17). On the other hand, whereas Cu-BTTri has a high CO$_2$ working capacity (only slightly lower than NOTT-101/OEt), its selectivity is lower than 30.

**DISCUSSION**

Here, we successfully demonstrated that a GA could be used to efficiently identify top adsorbent materials for precombustion CO$_2$ capture among thousands of hMOFs. The GA reduced the computational time by at least two orders of magnitude relative to a brute force search. One of the top-performing MOFs, NOTT-101/OEt, was synthesized and tested, and the experimental pure-component CO$_2$ and H$_2$ isotherms agree well with the simulation predictions. IAST-predicted mixture isotherms show that the CO$_2$ working capacity of NOTT-101/OEt is 3.8 mol/kg, with a CO$_2$/H$_2$ selectivity of 60. We also showed that the structure-property relationships obtained from the GA-guided search could be used to discover top-performing
MOFs in different databases without the need for a large number of additional simulations. The methods demonstrated in this work (both the GA-guided and the structure-property-guided search) could be applied to search for high-performing MOFs for other applications and should be especially useful when the performance evaluation requires a large amount of computational time, such as simulations involving large, complex molecules, or when quantum mechanical calculations are required.

**MATERIALS AND METHODS**

**Calculation of adsorption properties**

GCMC simulations (41) were carried out as implemented in the RASPA simulation code (42, 43) to compute adsorption loadings at 313 K. Details of simulations and models are provided in section S1. Simulations were carried out to compute the CO₂ adsorption loadings for pure CO₂ at 1 bar and the CO₂ and H₂ adsorption loadings for a 20:80 CO₂/H₂ mixture at 20 bar. The following adsorbent evaluation criteria were used to measure the fitness of each MOF

\[
\Delta N_1 = N_{ads}^1 - N_{des}^1 \quad (1)
\]

\[
\alpha_{12}^{ads} = \frac{N_{ads}^1 / y_1}{N_{ads}^2 / y_2} \quad (2)
\]

\[
APS = \Delta N_1 \times \alpha_{12}^{ads} \quad (3)
\]

Here, \(\Delta N_1\) is the CO₂ working capacity, \(N_{ads}^1\) and \(N_{ads}^2\) are the CO₂ and H₂ adsorption loadings for the CO₂/H₂ mixture at 20 bar, and \(N_{des}^1\) is the CO₂ adsorption loading for pure CO₂ at 1 bar. Section S4 discusses the reasons for using this definition of the CO₂ working capacity, which is different from what is sometimes used, \(\alpha_{12}^{ads}\) is the CO₂/H₂ selectivity, and \(y_1\) and \(y_2\) are the mole fractions of CO₂ (0.2) and H₂ (0.8) in the gas phase, respectively. We also defined an APS in Eq. 3, similar to the performance measure defined by Bai et al. (28), as a way to account for the impact of both the CO₂ working capacity and the CO₂/H₂ selectivity on the performance of a PSA unit.

**Database of hMOFs**

The hMOFs explored in this work were obtained from the WLLFHS database of hMOFs (21). The structure of each MOF in this database can be characterized by a sequence of six integers (a chromosome). Genes 1 to 6 encode the interpenetration capacity, the actual interpenetration level, and the identities of the inorganic node, primary linker, secondary linker, and functional groups of a given hMOF (Fig. 2A). The 137,193 MOFs in the WLLFHS database can be described by 51,163 unique chromosomes due to conformational isomers and structures that differ only in the positioning of the functional groups (see section S2). The simulation of MOFs with identical genes resulted in very similar performance because they have similar structures. Therefore, the original WLLFHS database was reduced to 51,163 MOFs by selecting 1 MOF from each unique chromosome. This reduced database was subsequently explored using the GA developed in this work.

**Genetic algorithm**

GAs are a class of optimization methods that mimic natural selection. In a typical GA, a population of candidate solutions is evolved in the solution space toward higher values of some fitness function. Here, the solutions were hMOFs, and the GA evolved the genetic information of hMOFs to optimize one of the performance measures defined in Eqs. 1 to 3. We started with an initial population of 100 hMOFs (that is, the first generation) that was selected manually to ensure that each possible gene was carried by at least 1 hMOF. Each generation was evolved to create a subsequent generation. All generations had a population of 100 hMOFs. Elitism was implemented to ensure that the hMOF with the highest fitness in the nth generation appears in the (n + 1)th generation. All other hMOFs in the (n + 1)th generation were obtained by applying genetic operations on hMOF pairs selected from the nth generation. These hMOF pairs were selected using the tournament method (44). In a tournament, the hMOF with the higher fitness value between the two randomly selected hMOFs from the nth generation was selected with a 95% probability. Each hMOF pair was obtained from two independent tournament selections, and single-point crossover was subsequently carried out on the selected pair of hMOFs with a 65% probability. Each gene in the new chromosome had a 5% probability to undergo a mutation. Each hMOF pair in the nth generation produced one hMOF for the (n + 1)th generation. Figure 2B summarizes the workflow of the GA, and full details can be found in section S3.

**SUPPLEMENTARY MATERIALS**

Supplementary material for this article is available at http://advances.sciencemag.org/cgi/content/full/2/10/e1600909/DC1

section S1. Computational methods.

section S2. Genetic information for the WLLFHS hMOF database.

section S3. Genetic algorithm.

section S4. Discussion about a definition of CO₂ working capacity.

section S5. Identification of top-performing hMOFs for synthesis.

section S6. Synthesis of NOTT-101/OEt and VEXTUO.

section S7. Powder x-ray diffraction data.

section S8. N₂ sorption data.

section S9. CO₂ and H₂ simulated and measured isotherms for VEXTUO.

section S10. IAST calculations.

fig. S1. Nitrogen model.


In silico discovery of metal-organic frameworks for precombustion CO₂ capture using a genetic algorithm


Sci Adv 2 (10), e1600909.
DOI: 10.1126/sciadv.1600909