Similarity-Driven Discovery of Zeolite Materials for Adsorption-Based Separations


Crystalline porous materials such as zeolites and metal-organic frameworks (MOFs) have applications as materials for separations, catalysis and gas storage.[1, 2] In particular, they are promising candidates for use as selective CO₂ adsorbents for carbon capture.[3] As their importance in energy-related applications grows, so does the need for new materials to meet engineering and economic constraints. Large databases of computationally predicted zeolites[4] and MOFs[5] are being generated, in principle enabling researchers to screen for outstanding materials. However, screening large databases of materials poses a serious challenge. Established computational techniques—including electronic structure calculations and molecular simulations—allow for accurate prediction of guest-related properties of a single material.[6] However, the computational cost of these calculations prohibits the characterization of a database of millions of structures, which would be required to perform brute-force screening. Moreover, as structures that exhibit desired properties constitute only a small fraction of predicted materials, a brute-force screening approach would involve many wasteful calculations.

Recent studies focused on screening porous materials for gas separations and storage have applied pre-screens to limit the number of structures to be characterized. These pre-screens have been based on either geometrical parameters[6] or approximated physical properties.[5] Furthermore, our group has developed tools to increase the number of structures that can be characterized in a brute-force approach to about 100000 materials by exploiting the parallel computing power of clusters of graphics processing units (GPUs),[7] incorporating automatic structure analysis algorithms.[8]

Still, these developments do not allow efficient screening of a database of millions of materials. Herein, we present a novel approach to materials discovery which further reduces the computational cost of identifying promising candidate materials by at least an order of magnitude. Our method is based on the similar property principle[9]—successfully employed in drug discovery[10]—which states that similar chemical structures have similar properties, a simple yet powerful concept that leads to efficient chemical space exploration. Our database screening approach has two steps. First, we characterize diverse materials to identify substructural features which dominate physical properties of interest; second, the database is analyzed using a computationally inexpensive substructure comparison method to identify materials which exhibit similar property-determining substructures. As a result, we can efficiently discover a majority of the highest-affinity adsorption structures by characterizing only a fraction of a database. We demonstrate this procedure by screening a previously unexamined database of millions of zeolites[9] for structures for the adsorption-based separation of CO₂/N₂, a critical separation for post-combustion carbon capture.

Adsorption-based separation processes are based on differences in adsorption of two or more to-be-separated species. Two physical properties that influence adsorption behavior of a material are the Henry coefficient (Kₐ), which determines the affinity of a guest for adsorption in the low-pressure regime, and the heat of adsorption (Hₐ), the latent heat given off by a guest adsorbing to the framework. These can be calculated in Monte Carlo simulations[11] (all simulations discussed herein were performed at 300 K using the force fields of ref. [12] and ref. [13]). In the case of CO₂/N₂ separation, since the Kₐ and Hₐ of N₂ are comparatively low, a good separation material will have a high Kₐ and Hₐ for CO₂.[14]

Kₐ and Hₐ are physical properties of an entire structure, but they can be dominated by the presence of strong binding sites for the guest molecule due to the exponential contribution of guest–host interaction energies.[11] A structure that exhibits strong binding sites—“sweet spots”—will have significant Kₐ and Hₐ (Figure 1). Moreover, structures with similar binding sites will be similar in their adsorption properties. We exploit this behavior in order to efficiently discover materials for optimum adsorption through geometry-based structure comparison.

The first step in our process is to characterize a small, diverse sample of structures. In a large dataset of porous materials there exist many different void space features: large and small cages; wide and narrow, straight and curved channels; junctions, and networks of different dimensionality. Through dissimilarity-based selection of structures we can efficiently sample this shape diversity with only a few tens of materials.[15]

The strength of our approach is the exploitation of the efficient sampling provided by diversity selection to obtain repre-

[9] DOI: 10.1002/cphc.201200554


3595
sentative geometric search queries from characterizing only a small training set of materials. By characterizing the diverse set we observe that the most energetically favorable binding sites for the examined guest molecules comprise framework atoms arranged so as to approximately enclose the molecule. Moreover, eight silicon atoms of the framework are positioned in close proximity to each of the two extremes of the linear molecule, for example, the oxygen atoms for CO₂. Thus, for each guest molecule, we can screen for materials which exhibit two such eight-silicon features in close proximity, indicating the potential existence of a guest-molecule-enveloping sweet spot (Figure 2).

Figure 1. Three hypothetical zeolites and their energy contours for CO₂: CO₂-accessible space is green, and low-energy (favorable) positions blue. a) PCOD8286959, with CO₂ 2.1 × 10⁻² mol kg⁻¹ Pa⁻¹. b) PCOD8290342, the 5th most similar by overall shape to (a); however without sweet spots the Kₜ is only 5.7 × 10⁻⁶. c) PCOD8278310, the 124647th most similar to (a); due to the presence of sweet spots, a similarly high Kₜ of 2.3 × 10⁻⁵, despite little shape similarity. Similarity of void space shape calculated using methodology of ref. [15].

Figure 2. a) Two features (black, purple) are detected in a substructure of zeolite PCOD8207615. A vector between features (blue) enables estimation of the CO₂ binding position. It can be seen that the minimum-energy position for CO₂ obtained from simulation lies approximately along this vector; b)–g) all six CO₂ binding site features identified in the diverse set of hypothetical zeolites.

To identify binding sites within a zeolite framework we first determine the guest-accessible regions of void space.¹⁶ By focusing on the surface of guest-accessible cavities, we can have confidence that the guest can indeed diffuse to identified binding sites. We then perform shape comparison following ref. [17] to detect geometric matches between substructures of the framework and query features. Finally, the interaction energy of the guest molecule in a binding site can be estimated by docking the guest between the two features, and estimating the molecule’s minimum-energy orientation and position (Figure 2a). The guest–host interaction energy can then be calculated using a force field.¹²,¹³ We note that for chemically heterogeneous materials such as MOFs, a simpler technique such as pharmacophore matching¹⁸ may be sufficient to identify binding sites.

We validated our screening approach on the PCOD set of about 140000 zeolites previously characterized using GPU-based molecular simulations.¹⁴ We rank each structure in the set by its predicted minimum guest–host interaction energy, our hypothesis being that a strongly favorable energy, correctly predicted, will dominate adsorption properties. The performance of our method is quantified by how this ranking positions the best adsorbing materials—those within the top 1% of Kₜ or Hₛ—near the top of the ranked list. We measure the enrichment in the first 500 structures in the ranked list, that is, the ratio of structures retrieved therein with the desired property, versus the number that would be retrieved at random.

Accurate retrieval of structures with outstanding adsorption properties can be achieved by our technique. At the cost of characterizing only 500 zeolites, we retrieve 327 materials within the top 1% of CO₂ Kₜ (i.e. > 1.301 × 10⁻⁷ mol kg⁻¹ Pa⁻¹) in the PCOD dataset, that is, a greater than 65-fold enrichment over random selection or brute-force characterization. A greater than 72-fold enrichment is similarly observed for materials within the top 1% of CO₂ Hₛ (i.e. < −49.730 kJ mol⁻¹).

Furthermore, to demonstrate the applicability of our method on a database one order of magnitude larger, we screened the PZEF set of about one million hypothetical zeolites that have not previously been characterized. Within this set we detected about 30000 new materials that exhibited the known features. This set of candidate materials represents a quantity that can be examined in detail with more sophisticated methods. The single best CO₂-adsorbing material discovered in the PZEF set is illustrated in Figure 3.

In conclusion, our method enables the efficient discovery of zeolite materials with optimum adsorption properties. As a result we can screen material databases at a fraction of the cost of brute-force characterization, enabling the analysis of datasets at least an order of magnitude larger than those examined by current state-of-the-art methods. Our technique is
based on material informatics: diversity selection, similarity searching and docking, and it selects promising candidate materials by identifying substructural features that dominate physical properties of interest and screening for them efficiently within a database. Our screening approach is novel in that it is based on local structural features, allowing us to identify outstanding adsorption materials based on their specific guest-binding sites. We have demonstrated this computationally inexpensive method by screening a database of about one million zeolites for structures for the adsorption-based separation of CO\textsubscript{2}/N\textsubscript{2} mixtures, and have shown that many of the strongest CO\textsubscript{2} adsorbing materials identified by material informatics are also appropriate for another gas separation, with results for the separation of C\textsubscript{2}H\textsubscript{4}/C\textsubscript{2}H\textsubscript{6} described in the Supporting Information.

Acknowledgements

This work was supported by the Director, Office of Science, Advanced Scientific Computing Research, of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231. In addition, it was supported jointly by DOE Office of Basic Energy Sciences and the Office of Advanced Scientific Computing Research through SciDAC project #CSNEW918 entitled “Knowledge-Guided Screening Tools for Identification of Porous Materials for CO\textsubscript{2} Separations”, and as part of the Center for Gas Separations Relevant to Clean Energy Technologies, an Energy Frontier Research Center funded by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences under Award Number DE-SC0001015. This research used resources of the National Energy Research Scientific Computing Center (NERSC), which is supported by the Office of Science of the U.S. Department of Energy under Contract No. DEAC02-05CH11231.

Keywords: adsorption • carbon dioxide • microporous materials • virtual screening • zeolites


Received: July 6, 2012 Published online on August 22, 2012