Quantifying Gas Adsorption Variability and Optimal Si/Al Ratio for Rational Design of Aluminum Substituted Zeolite Frameworks

Akhilesh Gandhi[a], Silabrata Pahari[a], Joseph Sang-II Kwon[a], M. M. Faruque Hasan[a,b,*]

[a]Artie McFerrin Department of Chemical Engineering, Texas A&M University, College Station, TX 77843-3122, USA.

[b]Texas A&M Energy Institute, Texas A&M University, College Station, TX 77843, USA.

Abstract

Experimental measurements often show significant variations in gas adsorption on aluminosilicate zeolites, thereby inducing considerable uncertainty in gas separation and storage performance. These variations are largely attributed to the distribution of aluminum (Al) atoms within a zeolite framework. It is challenging to experimentally control the distribution of Al atoms during zeolite synthesis. The vast number of plausible Al-substituted configurations also makes it difficult to estimate the overall range of adsorption. To resolve this, we deploy a new representation of crystallographic frameworks using single repeating units (SRU). An SRU consists of the smallest network of tetrahedral atoms that can be repeated as single building block to represent an entire zeolite framework. SRUs enable a selective enumeration of unique Al-substituted configurations, thereby leading to an efficient computational framework for quantifying the variations in equilibrium gas adsorption on Al-substituted zeolites without exhaustive search. We apply this technique to analyze CO_2 adsorption on chabazite (CHA) zeolite. Using molecular simulations of gas adsorption on the unique Al-substituted configurations, we observe as much as 12% variation in CO_2 adsorption due to differences in the locations of Al atoms within the zeolite framework. Interestingly, our results indicate that variability in CO_2 adsorption in Al-substituted zeolites is significant only at moderate Si/Al ratios, primarily due to the non-uniform distribution of Al. At very high or very low Si/Al ratios, this variability appears to be negligible. Surprisingly, we also observe that the adsorption does not always increase with the number of Al sites, and there exists an inflection point beyond which additional Al substitution leads to a decrease in adsorption. This trade-off indicates an optimal Si/Al ratio that maximizes the equilibrium adsorption of CO_2 on Al-substituted CHA zeolites at some moderate values. We are able to systematically identify the optimal Si/Al ratio and the corresponding locations of Al sites in CHA framework that maximizes CO_2 adsorption. On further investigation using the Al-Al radial distribution function (RDF), we find the locations of Al sites that lead to high CO_2 adsorption. This demonstrates that the SRU-based selective enumeration combined with RDF-based structural screening is an enabling method towards the rational design of zeolites with optimal distribution of Al sites to achieve desired properties.

Keywords: Zeolites, Carbon capture, CO2 adsorption, Nanoporous Adsorbent, Aluminosilicates

*Correspondence concerning this article should be addressed to M.M. Faruque Hasan at hasan@tamu.edu, Tel.: 979-862-1449.

1 Introduction

Zeolites are aluminosilicate materials with a wide range of applications in gas separation, catalysis, ion exchange, and other fields.[1–5] More than 250 zeolites have been naturally found or chemically synthesized, and millions of hypothetical structures have been analyzed using computer-based techniques.[6] Like most aluminosilicates, zeolite frameworks are formed by aluminum (Al) and silicon (Si) atoms through forming Si-O-Al, Si-O-Si, and Al-O-Al oxide linkages. These linkages lead to a network of tetrahedral atoms (T-sites) to give rise to a three dimensional porous structure. The distribution and spatial arrangement of Al atoms within a zeolite framework particularly affect the adsorption of various gases within the framework. Experimental measurements have indicated considerable differences in equilibrium gas adsorption due to Al distribution within a zeolite framework. Up to two-fold differences in adsorption capacities of n-alkanes have been observed, for example, over Bronsted acid zeolites with the same Si/Al ratio.[7] Conventional wisdom is that Al substitutions generally increase gas adsorption.[8] However, this may not be true for all cases. Recent studies[9,10] on small-pore zeolite frameworks, such as Gismondine (GIS) and Merlinoite (MER), have revealed the existence of optimal Si/Al ratio beyond which adsorption decreases with additional Al substitution. Data have been also reported[11] where CO_2 adsorption increases with increasing Al substitution until a plateau or peak is reached, beyond which no significant increase in adsorption is observed.

The mechanisms underlying adsorption on Al substituted zeolites are complex and require further investigation. For fixed pore size and basicity, the electric field strength generated by cations through Al substitution influences the gas adsorption in microporous crystalline materials.[12] The siting and the orientation of Al atoms at important T-sites as well as the proximity of Al atoms to each other influence the force field within the lattice and determine the overall adsorption capacity. When Al atoms are strategically located at specific T-sites, they can synergistically enhance adsorption, leading to higher adsorption and selectivity for certain gases. Dispersion interactions also play an important role. For example, dispersion interactions account for about 50% of the overall adsorption enthalpy of CO_2 molecules in FAU zeolite with a Si/Al ratio of 2.55:1.[13] Guest CO_2 molecules predominantly adsorb on site II and tilt toward the zeolite wall due to stabilizing dispersion interactions, with minor heterogeneity in adsorption sites arising from differences in the number of Al atoms and the geometry affecting their spatial arrangement. Other factors, such as the cation type, size and charge, and the Si/Al ratio also affect the equilibrium and kinetic properties of gas adsorption on zeolites.[14–16]

It is important to accurately estimate the gas adsorption properties of zeolites to be able to design feasible chemical processes with desired separation and/or catalytic performance.[17–23] To this end, the following questions remain unanswered: How do the location and configuration of Al substitutions affect the adsorption capacity of a zeolite? What is the optimal Si/Al ratio that maximize gas adsorption, and how can we systematically identify this ratio?

The absence of literature regarding variation in adsorption on Al-substituted zeolites is primarily due to the challenges associated with comprehensive experimental and computational measurements

and observations. There are recent successes in experimentally biasing the arrangement of Al and acid sites,[24–27,27–29] but experimental synthesis of samples with precise control over each Al-siting and distribution is still difficult. Computational methods, on the other hand, typically rely on time-consuming and resource-intensive molecular simulations[30]. These calculations may range in the order of days to weeks to obtain a single adsorption estimate on a single zeolite. While AI/ML-based predictive models are being developed[31,32], high-throughput screening of Al-substituted zeolites remains computationally demanding, given the large number of possible framework structures. The complexity further increases when dopants, such as Al or other metals, are used to replace Si in zeolite frameworks.[33] Recent studies have demonstrated significant advancements in the development of quantitative structure-property relationships, driven by the emergence of advanced machine learning architectures such as Transformers.[34–37] These data-dependent approaches often lack a systematic understanding grounded in first principles. Moving forward, it is essential to establish effective screening rules while maintaining a deeper understanding of the underlying system.[38]

The most challenging issue in studying the variability in adsorption of Al-substituted zeolites is probably the vast number of possible Al-substituted configurations. To illustrate, consider the CHA zeolite framework, which has 36 T-sites in each unit cell. For a Si/Al ratio of 5.0, it would require substituting 6 of the 36 the sites with Al. Without considering additional constraints, such substitutions can be achieved in $\binom{36}{6}$ ways, leading to over 1.94 million possible structure configurations to check. As the number of Al atoms in the CHA framework increases, the number of possible Al-substituted configurations increases. For the Si/Al ratio of 1.0, which has 18 Al atoms and 18 Si atoms in a unit cell, this increases to over 9 million configurations. Even after limiting to only feasible structures via the Löwenstein's rule, which forbids Al-O-Al pairs in zeolites, the total number of plausible zeolite configurations is overwhelming. The enumeration of Al-substituted structures adds additional complexity to already time-intensive molecular simulations. No systematic approach currently exists to address this highly combinatorially complex problem, which seems to be daunting to address even for a fixed Si/Al ratio. To the best of our knowledge, a systematic study investigating the effect of Al-siting on gas adsorption has also not been performed yet.

In this work, we overcome these combinatorial challenges in assessing the unique Al-substituted structures through exploiting a new graph-theoretic representation of zeolite frameworks using Single Repeating Units (SRU). Through SRUs, we can place Al atoms in unique sites while complying with the Lowenstein's rule, thereby allowing us to systematically enumerate only the unique Al-substituted configurations for a given Si/Al ratio, while representing most of the variations in Al sites. This is a key contribution of this work, as it significantly reduces the number of configurations needed for analysis and leads to a very efficient computational framework, which is described in detail in Section 2, for quantifying the variability observed in equilibrium gas adsorption, storage, and other properties of Al-substituted zeolites.

The paper is structured as follows: Section 2 outlines the methodology, including the SRU-based

enumeration and generation of unique Al-substituted zeolite structure configurations, and molecular simulation to obtain gas adsorption. In Section 3.1, we quantify the variation in CO_2 adsorption on Al-substituted SOD and CHA zeolite frameworks for different Si/Al ratios, and systematically analyze the effects of proximity and location of distributed Al sites on CO_2 adsorption. In Section 3.2, we propose a new approach, based on the similarity in the radial distribution function (RDF) and the CO_2 adsorption, for high-throughput screening of optimal locations of Al sites in the lattice of Al-substituted CHA zeolite framework that maximize the CO_2 adsorption. We provide our concluding remarks in Section 4.

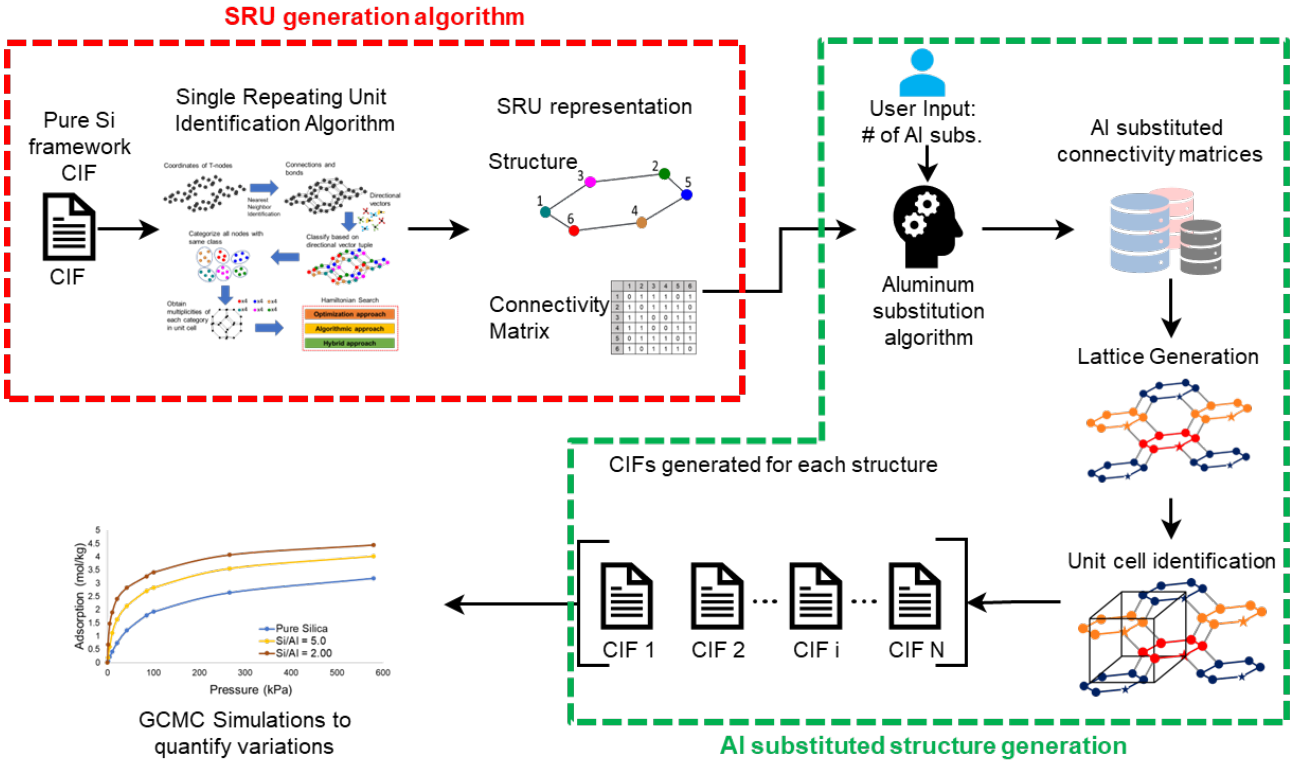


Figure 1: An overview of the workflow of the proposed computational framework, describing the enumeration and analysis of uniquely Al-substituted zeolite configurations. An efficient subset selection of unique distributions of Al for a given Si/Al ratio is enabled by the SRU (single repeating unit)-based representation of zeolite frameworks. An SRU is the smallest network of T-sites that can be repeated to generate an entire zeolite framework. It is used to generate the uniquely Al-substituted structures. GCMC simulations are then performed on select Al-distributed structures to quantify the adsorption capacities.

2 Methods

Figure 1 provides an overview of the workflow of the computational framework for the enumeration and analysis of uniquely Al-substituted zeolite configurations. It begins by taking the crystallographic information file (CIF) of the zeolite of interest. CIF is a standard way for describing a crystal framework. With the CIF available, the first major step is to obtain the SRU representation of the zeolite structure. An SRU is the single smallest network of T-sites with defined connectivity that can be repeated to represent an entire zeolite framework.[39] We have several efficient techniques/algorithms to identify the smallest set of T-nodes and their connectivity matrix that defines the SRU of a zeolite framework (see Section 2.1 on SRU identification). We then use the SRU representation to generate structures with the desired number of Al-substitutions, while maintaining a user-specified, fixed Si/Al ratio. As a check, the algorithm generates the post-substitution connectivity matrices and determines their feasibility according to the Lowenstein's rule. Once the algorithm determines the feasible Al-substituted connectivity matrices, we use them to create the lattices and the unit cells needed to generate the individual CIF files for Al-substituted zeolite frameworks (see Section 2.2 on generation of unique Al-substituted structures). We provide the new CIF files as inputs to a molecular simulation platform to predict the equilibrium adsorption of guest molecules, such as CO_2, on different Al-substituted zeolite configurations for the same Si/Al ratio. We analyze the simulation data and quantify the variations in gas adsorption due to variations in Al-substitution within a zeolite framework. We repeat the above procedure by parametrically changing the number of Al atoms within a zeolite framework. Overall, the SRU-based selective enumeration of Al-substituted structures enable a rational design of aluminosilicate zeolites with optimal Si/Al ratio to achieve desired gas adsorption and storage properties.

2.1 SRU Identification

The SRU representation of zeolite frameworks was first proposed in 2021.[39,40] An SRU consists of a structure that has fewer T-nodes than the unit cell and a connectivity matrix that governs the rules of connectivity between SRUs for re-generating the lattice. This graph-theoretic representation via the connectivity matrix is unique with the benefit of enumerating structural rules based on connectivity.

The graph-theoretic SRU representation enables the systematic enumeration of Al-substituted zeolite structures following Löwenstein's rule. For example, the CHA SRU has only 12 T-atoms and thus only 66 structures can be enumerated with a Si/Al ratio of 5.00 using the expression $\binom{12}{2}$. Using the SRU representation, a much larger number of structures can be efficiently generated and evaluated, enabling a better understanding of the impact of Si/Al substitution on zeolite properties and accelerating the discovery of new materials for various applications.

The benefit of the SRU representation lies in its reduced representation compared to the traditional unit cell representation. To demonstrate this, we consider the CHA unit cell as a reference, which requires 3 CHA-SRU units to represent. If we want to replicate all 1.9 million possibilities that the

unit cell offered, we can choose SRUs to get the desired Si/Al ratio in the unit cell and then identify different combinations of these SRUs. Once we generate these possibilities, we need to filter through all the 1.9 million structures based on Löwenstein's rule. Table 1 shows the calculations based on the smaller structure (12-member unit similar to SRU without the graph representation), without exploiting the benefits of the graph-theoretic representation. This table shows the possibilities for the reconstruction of unit cell enumeration where each column shows the possible number of structures for the selected Al substitution. In the table, we calculate the total number of ways of selecting which smaller units to use and unique placements of the same which lead to a total of 1.9 million, equal to $\binom{36}{6}$.

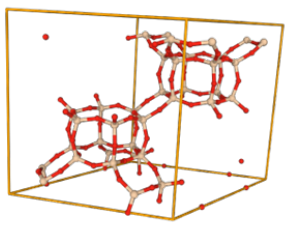
Table 1: Possibilities for reconstruction of unit cell enumeration with graph-theory not utilized for CHA with Si/Al ratio of 5.00

# subs. in SRU	0	1	2	3	4	5	6			
Possibilities	$\binom{12}{0}$	$\binom{12}{1}$	$\binom{12}{2}$	$\binom{12}{3}$	$\binom{12}{4}$	$\binom{12}{5}$	$\binom{12}{6}$	Select	Orientations	Total
	2						1	$\binom{12}{6}\binom{12}{0}\binom{12}{0}$	$\frac{3!}{2!1!}$	2,772
	1	1				1		$\binom{12}{5}\binom{12}{1}\binom{12}{0}$	$\frac{3!}{1!1!1!}$	57,024
	1		1		1			$\binom{12}{4}\binom{12}{2}\binom{12}{0}$	$\frac{3!}{1!1!1!}$	196,020
		2			1			$\binom{12}{4}\binom{12}{1}\binom{12}{1}$	$\frac{3!}{2!1!}$	213,840
	1			2				$\binom{12}{3}\binom{12}{3}\binom{12}{0}$	$\frac{3!}{2!1!}$	145,200
		1	1	1				$\binom{12}{3}\binom{12}{2}\binom{12}{1}$	$\frac{3!}{1!1!1!}$	1,045,440
			3					$\binom{12}{2}\binom{12}{2}\binom{12}{2}$	$\frac{3!}{3!}$	287,496
									Total	1,947,792

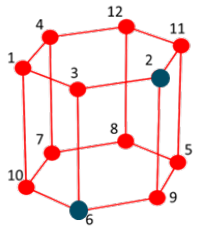
In Table 2, we demonstrate the reduced number of structures generated using the methodology proposed in this paper. The select column in this table refers to the possible selections of the selected substitutions and details on how these numbers are obtained are discussed partially in a previous work.[39] Note that the difference here is the number of possibilities for each level of substitution and details of how the number of possible structures is reduced will be covered in the methodology using the node-index method. There is one order of magnitude reduction in the total number of structures that need to be considered. However, given that the number is still significantly high on the computational requirements, we make the assumption on the periodicity of the lattice. We assume that the lattice is periodic over the SRU size. In the case of the unit cell, the periodicity is repeated over the unit cell, and thus this assumption is in line with assuming periodicity in the smallest representative unit considered. There are some consequences to this assumption that some high Si/Al ratio structures may not be covered however that is the case for any baseline structure that can be arbitrarily chosen and some structures will still be excluded. Figure 2 provides a visual demonstration of the reduction in possible structures due to the use of SRU in contrast to the unit cell.

Table 2: Possibilities for the reconstruction of unit cell from SRU with the benefit of graph-theoretic representation for CHA with Si/Al ratio of 5.00

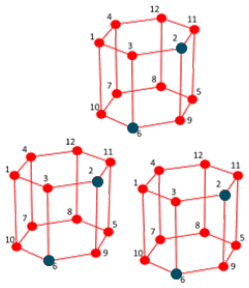
# subs. in SRU	0	1	2	3	4	5	6			
Possibilities	1	12	42	52	30	12	2	Select	Orientations	Total
	2						1	2	$\frac{3!}{2!1!}$	6
	1	1				1		12*12	$\frac{3!}{1!1!1!}$	864
	1		1		1			30*42	$\frac{3!}{1!1!1!}$	7,560
		2			1			30*12*12	$\frac{3!}{2!1!}$	12,960
	1			2				52*52	$\frac{3!}{2!1!}$	8,112
		1	1	1				52*42*12	$\frac{3!}{1!1!1!}$	157,248
			3					42*42*42	$\frac{3!}{3!}$	74,088
									Total	260,838



$C_6^{36} = 1,947,792$



$C_2^{12} = 66$



Different permutations of 3 SRUs with Al substitutions (total = 6 subs.) leads to 260,838 structures

Figure 2: Strategic reduction in total possible Al-substituted structure by using SRU and permutations and combinations. The total Al substitutions in the unit cell still remain 6.

A critical point to note is that both the SRU and traditional unit cell representations have inherent limitations in terms of capturing long-range lattice-scale effects. Attempting to construct exhaustive unit-cell-based configurations over multiple repeated units (e.g., 9-unit cells) leads to an astronomical number of possibilities—on the order of 10^{62} for Si/Al = 5.00 using the unit cell as the representation—making direct simulation impractical. Our SRU method addresses this bottleneck by focusing on a manageable yet chemically relevant subset of configurations, leveraging topological symmetry and

representative structural motifs. This approach allows for efficient exploration of configurational diversity without sacrificing the physical fidelity of adsorption predictions. The magnitude of the reduced enumerations are reported in the supporting information.

As shown in Figure 3, we have applied the SRU representation to several zeolite frameworks, and the resulting SRU structures are smaller than the unit cell, offering a 2-4 fold reduction in the number of T-atoms required in most zeolites. Figure 3 shows some examples of the SRU representation in contrast to the original lattice while the unit cell is shown in gray. Note that the SRU structure has each T-node numbered. The numbering of the nodes in the SRU is based on the connectivity rules defined by the connectivity matrix. Figure 4 demonstrates an example SRU connectivity matrix for the SOD framework. The size of the matrix is governed by the total number of T-nodes in the SRU structure, and the connectivity matrix and the structure together make up the SRU representation. Following the connectivity rules, the entire lattice can be regenerated.

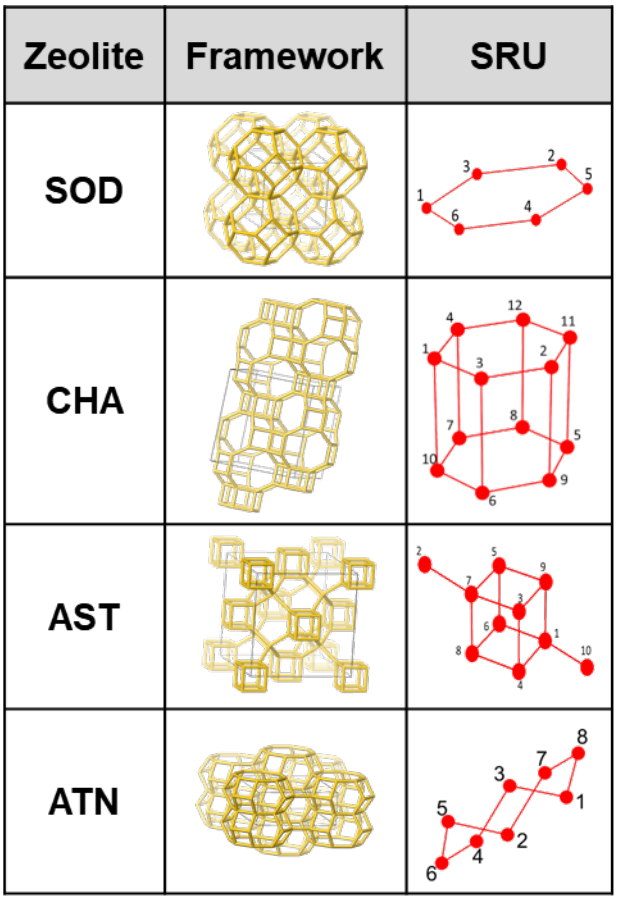


Figure 3: Examples of the SRU for a few zeolites. The unit cell has been highlighted in the framework and the SRU structure is shown in red with the numbering based on the connectivity matrix.

	1	2	3	4	5	6
1	0	1	1	1	0	1
2	1	0	1	1	1	0
3	1	1	0	0	1	1
4	1	1	0	0	1	1
5	0	1	1	1	0	1
6	1	0	1	1	1	0

Figure 4: The connectivity matrix for SOD. The SRU structure has 6 nodes leading to a connectivity matrix of size 6x6. Note that the sum of all elements in a row and column is equal to 4 due to the tetrahedral nature of the T-nodes.

2.2 Rational Generation of Al-substituted Zeolite Structures

With a more clear understanding of the SRU, we now focus on how this representation allows us to generate Al-substituted frameworks. The concept was briefly covered in our previous work however, we only discussed the number of enumerations and did not establish the complete unit cell generation of these enumerated structures.[39] To elaborate, we introduce the concept of a node index based on the neighboring 4 atoms in the lattice. We assign the node index a value of 4 for a Si atom connected to all four Si atoms. The node index value is reduced by 2 for every neighbor that is substituted as Al until the node index goes down to -4. For a central Al atom connected to 4 Si atoms, we define the node index as -6. Note that in this entire definition of the node index, there is no value assigned for an Al-Al connectivity. This lies within the scope of our problem due to Lowenstein's rule that forbids Al-O-Al linkage.

The node index is a theoretical concept that can be defined for each node in a structure. However, this concept can be mathematically programmed for matrices which allows for the systematic enumeration of structures. In a mathematical framework, the node index can be defined as the sum of all the row (or column since the matrix is symmetric) elements. With a pure Si framework, the node index for all the nodes is 4 based on the tetrahedral nature. For example, say we substitute site 1 with Al instead of Si. Mathematically, this means we multiply all elements in the row and column by -1 and update the diagonal element at (1,1) with a value of -2. Now when we recompute the node index for the entire matrix by summing the rows or columns we get the same value as we would get from a visual inspection of the structure. Additional substitutions can then be performed on nodes only where the node index has a value of 4. The benefit of the representation and the node index lies in the fact that the computation can be mathematically programmed thus making the process systematic.

To understand the framework in detail, let us first define the problem statement as: given the SRU matrix for a zeolite, determine feasible Al substitution locations while obeying Lowenstein's rule of no Al-O-Al linkage. We explain the solution procedure to obtain all feasible Al substitution locations via

two functions described in algorithms 1 and 2.

Algorithm 1 presents a function for adding a single Al atom to a given structure represented by a matrix M of dimension $m \times m$. The algorithm returns a set of all feasible matrices that result from the substitution of a single Al atom. The function AddOneAl starts by initializing an empty set $\mathcal{SSL}$ that will store the feasible matrices. Then, the diagonal elements of M are extracted into a vector de. The variable s is set to the highest index where de equals -2, which corresponds to the last Al substituted position site in the lattice. If no such site exists, s is set to 0, indicating that no Al substitution exists in the given matrix and all positions should be considered for Al substitution.

Further, a loop over the indices i ranging from s to m is performed. At each iteration, a copy of M is created and stored in the variable G. Then, the i^{th} row and column of G are updated to represent the addition of the Al atom at the empty site i. The update is only performed if the corresponding node index in the lattice is equal to 4, which ensures that the Al atom can be added without violating any constraints on the node index as defined previously. Finally, the updated matrix G is added to the set $\mathcal{SSL}$. After the loop, the function returns the set $\mathcal{SSL}$ containing all feasible matrices resulting from the addition of the Al atom. Note that the function does not modify the input matrix M, as it creates copies of it for each update.

Algorithm 1 Function for adding a single Al to given structure

function AddOneAl($M^{m \times m}$) ▷ Add single Al to M, return set of all feasible matrices
  $\mathcal{SSL} \leftarrow \{\}$
  $de \leftarrow diag(M)$
  $s \leftarrow$ highest index where $de == -2$, else 0
  for $i \in [s, ..., m]$ **do**
    $G \leftarrow M$
    update i^{th} row, col of G if Node Index == 4
    $\mathcal{SSL}.insert(G)$
  end for
  return $\mathcal{SSL}$ ▷ Set of single Al added matrices
end function

We now take a look at the most important function where we determine the addition of all possible Al that is described in Algorithm 2. Algorithm 2 takes a connectivity matrix C and a positive integer n as input, and generates a set of matrices A with Al substituted nodes, such that the number of Al substitutions is at most n.

The algorithm starts by initializing a copy of the input matrix C in Z as a temporary variable. The function GenMatrices is then called with Z, n, and an empty set $\mathcal{L}$ as input. If $\mathcal{L}$ is empty, the function checks if the number of nodes with value -2 (i.e., nodes that have not yet been substituted with Al) in Z is equal to n. If so, Z is added to the set $\mathcal{A}$ of matrices with Al substituted nodes. Otherwise, the function AddOneAl is called to generate a set $\mathcal{SSL}$ of all possible matrices that can be obtained by adding a single Al to Z. For each matrix I in $\mathcal{SSL}$, the function GenMatrices is called recursively with I, n, and $\mathcal{SSL}$ as input. This ensures that all possible combinations of Al substitutions are explored until the desired number of Al substitutions is reached.

If $\mathcal{L}$ is not empty, the function enters the recursive step. For each matrix O in $\mathcal{L}$ (which contains all possible matrices generated in previous recursive calls), the function AddOneAl is called to generate a set $\mathcal{T}$ of all possible matrices that can be obtained by adding a single Al to O. For each matrix, T in $\mathcal{T}$, T is added to the set $\mathcal{A}$ of matrices with Al substituted nodes. The function returns the set $\mathcal{A}$ of all matrices with Al substituted nodes that are generated during the execution of the algorithm.

Algorithm 2 An algorithm to generate Al substituted matrices from SRU connectivity matrix

Input: $C^{m\times m}, n \leq \frac{m}{2}+1$ ▷ n: number of Al substitutions
Output: $\{A^{m\times m} : A \in \mathcal{A}$ where $\mathcal{A}$ is the set of matrices with Al substituted nodes$\}$

$Z \leftarrow C$
function GenMatrices($Z, n, \mathcal{L} = \{\}$) ▷ Returns set of matrices with substitutions until n
  if $\mathcal{L}$ == {} **then**
    if no. of -2 in Z == n **then** ▷ Check if desired number of Al subs. acheived
      $\mathcal{A}.insert(Z)$
    else
      $\mathcal{SSL} \leftarrow$ AddOneAl(Z)
      for $I \in \mathcal{SSL}$ **do**
        GenMatrices(I,n, $\mathcal{SSL}$) ▷ Recursive call until desired Al subs. reached
      end for
    end if
  else ▷ Condition valid when in recursion
    for $O \in \mathcal{L}$ **do**
      $\mathcal{T} \leftarrow$ AddOneAl(O)
      for $T \in \mathcal{T}$ **do**
        $\mathcal{A}.insert(T)$
      end for
    end for
  end if
  return $\mathcal{A}$
end function

After generating the connectivity matrices we then use the structure to generate lattice coordinates. The next step is to convert the lattice coordinates to CIF file structures for molecular simulations. Since most molecular simulation software uses the CIF format, we expand the lattice sufficiently large to ensure at least one unit cell is contained within and then identify the atoms within this volume using the original parallelepiped dimensions of the unit cell. We then import these atoms as XYZ coordinates and convert the data to the CIF format. This process is critical since we want to exploit the periodic boundary conditions and mimic an infinitely repeating system to get accurate results. With the CIF structures generated, we can now perform molecular simulations to obtain adsorption isotherms for different Al substitutions. In the next section, we describe the molecular simulation methods we use for our analysis.

As we present the different structures considered in this work, we show some examples of the unit cells that we generated using the SRU framework and the proposed methodology. Since the SRU of CHA has 12 T-atoms, we generated structures from 1 substitution to 6 which led to Si/Al ratios ranging from 11.0 to 1.0, and these are shown in Figure 5. It is important to note that zeolites with

Si/Al ratios of 1.0 have not been observed or synthesized, but considering these structures allows us to study the effects of Al substitutions computationally.

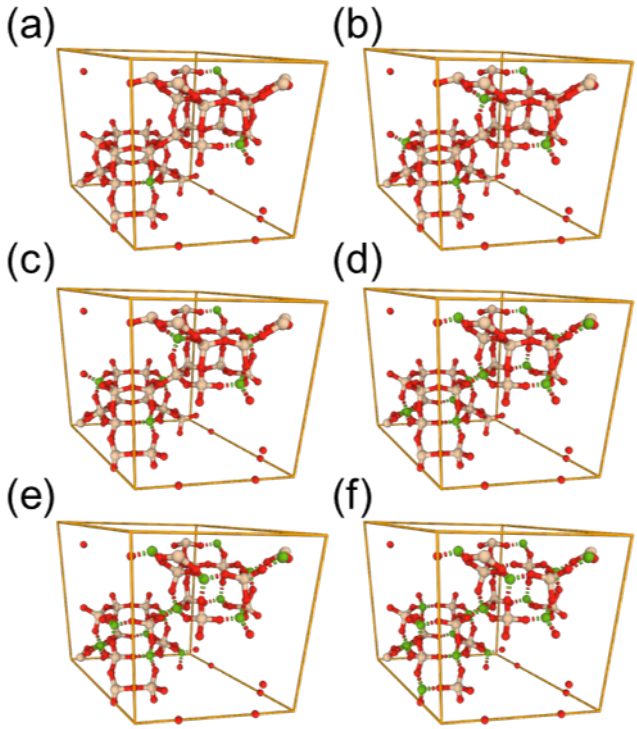


Figure 5: Examples of the unit cells of Al substituted zeolite frameworks. Each cell has a different Si/Al ratio. (a) has a ratio of 11.0, (b) has 5.0, (c) has 3.0, (d) has 2.0, (e) has 1.4, and (f) has 1.0. Note the color scheme where Al is highlighted in green while Si and O are in ivory and red respectively.

2.3 Molecular Simulation

To estimate the equilibrium gas adsorption capacity of the select Al-substituted structures, we use the Monte Carlo simulation technique for sampling. Molecular simulations were performed using the multi-purpose software RASPA 2,[41] which is developed for microporous materials, to perform the Grand Canonical Monte Carlo (GCMC) simulations. GCMC simulations are performed for CO_2 on Al-substituted zeolite frameworks at 298K under varying pressures. The simulations follow the conventional techniques reported in the literature[42,43]. The Lennard-Jones potential for the non-bonded interactions were taken from the Garcia Sanchez force field and are provided in the supporting information.[44] Additional details of the atoms mass, charge, and radii are also provided in the SI. The CO_2 molecules are based on the bond length and bond angle of 1.149Å and 180° respectively.

The simulations are initialized with 10,000 cycles and run for 25,000 production cycles. The thermodynamic properties are calculated every 1000 cycles. The simulation is performed in a $3 \times 3 \times 3$ unit cell. The $3 \times 3 \times 3$ simulation cells are generated using the VESTA software package, and

the GCMC simulations are performed under periodic boundary conditions[45]. Na ions are added to neutralize the system and allowed to float freely in the simulation box.[44] The atom labels are retained, and the Al atoms connected to the oxygen atoms are renamed to identify themselves as separate atoms with interaction parameters obtained from the force field. For the ions, only translational and insertion probabilities are used, with a value of 1 for both. For the adsorbate species, three moves are incorporated, which include translation, reinsertion, and swap, all with a probability value of 1.0. The cut-off radius used in GCMC simulations was 12Å.

After each simulation, the adsorbate's absolute loading on the adsorbent framework is calculated and recorded. The simulations were conducted for a total of 150 frameworks for the range of Si/Al ratios generated. Each framework is simulated at 9 different pressure values: 1, 5, 10, 20, 30, 40, 50, 75, and 100 kPa at 298K, resulting in a total of 1200 simulations. The chosen pressure values are selected based on their relevance in practical applications[46]. The molecular simulations are performed at the High-Performance Research Computing facility at Texas A&M University. The simulation time varies depending on the applied pressure. On average, each simulation takes about 60 hours to complete on a computer with 4 cores and 32GB RAM.

We also compare the molecular simulation results with those observed in literature. Due to their higher adsorption properties and stability, zeolites with some amount of Aluminum have been synthesized and studied in literature.[9] Impurities also skew the experimental results compared to the simulation results as shown in the study by Ghojavand et al.[46] The closest experimental zeolite with minor impurities that studies adsorption of CO_2 has been reported by Pourmahadi et al. where they report the adsorption of CO_2 and CH_4 on various CHA structures under different temperatures and pressures.[47] The closest to pure-silica CHA is their reported zeolite S5 and the adsorption isotherm reported in Figure 6 of their work where we compare the adsorption at 298K. Their experimental results on zeolite S5 are comparable to the simulation results reported in Figure 6 in this work with some margin for error for impurities, and computational vs experimental errors.

2.4 Radial Distribution Function in Al-Subsituted Zeolites

The effect of proximity of Al atoms on the adsorption can be analyzed using the radial distribution function (RDF). The RDF, often denoted as $g(r)$, is a fundamental tool in the study of materials that characterizes the distribution of particles around a tagged particle at a distance r. The RDF is a key indicator of the atomic structure of materials and varies significantly depending on the phase of the matter. It is therefore widely used for the characterization of solids, gases, and liquids. There are several possible RDFs, but we study only Al-Al RDFs in this work. In the case of materials with Al substitutions, such as zeolites, we focus on the Al atoms in the lattice and compute the RDF for these atoms. By analyzing the RDF, we can gain insights into the ordering and spatial arrangement of Al atoms in the material. The radial distribution function is defined as the ratio of $\langle \rho(r) \rangle$, the average local number density of particles at a distance r, to the bulk density of particles, ρ:

$$g(r) = \frac{\langle \rho(r) \rangle}{\rho} \tag{1}$$

The RDF is the computational method equivalent to the pair distribution function (PDF), which is obtained experimentally using X-ray or neutron scattering. The two methods are the same, in fact RDF is also called as the pair distribution function analogously. The PDF provides information on the positions of atoms in a crystal lattice and has soft peaks that correspond to high probabilities of atom presence. In contrast, the RDF has sharper peaks since the positions of atoms are already known. The significance of the RDF in material characterization lies in its ability to provide information on the local environment of atoms and molecules in a material. This information is crucial for understanding the physical and chemical properties of materials, such as their mechanical, electrical, and thermal properties. The RDF is therefore a valuable tool for materials science, chemistry, and condensed matter physics, and its continued development and application will be key in advancing our understanding of materials at the atomic scale.

Studying the RDF is similar to studying the PDF in that the focus is on the intensity and location of the peaks. The larger peaks in the RDF correspond to longer-range correlations in the lattice, while the first peak, typically the largest and most well-defined, corresponds to the nearest-neighbor distance between atoms. Subsequent peaks correspond to longer-range correlations between atoms, such as the second-nearest-neighbor distance and so on. The smaller peaks observed in the RDF correspond to interatomic distances between neighboring atoms in the lattice. These peaks are generally narrower and more well-defined than the broader peaks that correspond to longer-range correlations in the material. This is because atomic positions in a crystal lattice are highly ordered and regular, resulting in sharper peaks in the RDF.

The algorithm for computing the RDF is described in algorithms 3 and 4. The algorithm is given using a parallel computation process. Algorithm 3 does all the data manipulation and then splits the data for each parallel process which calls algorithm 4. The final result is compiled and reported in algorithm 3. In our case, the parameter dr is set to 0.25 since we have point locations of Al.

Algorithm 3 Function to compute the radial distribution function

Input: P has dimensions of $N \times d$ ▷ P: Array of coordinates of particles

function RDF($P, dr, \epsilon = 10^{-15}$)
  $r_{cutoff} \leftarrow 0.9$
  $min \leftarrow \min(P)$
  $max \leftarrow \max(P)$
  $P \leftarrow P - min$
  $dims \leftarrow max - min$
  $r_{max} \leftarrow (\min(dims)/2) * r_{cutoff}$
  $radii \leftarrow [dr, 2dr, ..., r_{max}]$
  $N, d \leftarrow$ **shape**(P)
  $\rho \leftarrow N/(\prod dims)$
  $tree \leftarrow$ **cKDTree**(P) ▷ Compute k-d tree for nearest neighbor lookup
  $N_{radii} \leftarrow$ **length**($radii$)
  $radii\_and\_indices \leftarrow$ **stack**([**range**(N_{radii}), $radii$])
  $radii_{splits} \leftarrow$ **split**($radii\_and\_indices$, **cpu cores**) ▷ Split data between number of CPU cores
  for $i \leftarrow 0$ **to** **length**($radii_{splits}$) $- 1$ **do**
    $values[i] \leftarrow$ **tuple**($radii_{splits}[i]$, $tree$, P, min, max, N_{radii}, dr, eps, ρ)
  end for
  $results \leftarrow$ **Pool.starmap**(PARALLELHISTLOOP(values)) ▷ Parallel process for RDF
  $g_r \leftarrow \sum(results)$
  return g_r, $radii$
end function

Algorithm 4 Parallel RDF Histogram Loop Process

Input: $radii\_and\_indices$, $kdtree$, P, min, max, N_{radii}, dr, ϵ, ρ

function PARALLELHISTLOOP($radii\_and\_indices$, $kdtree$, P, min, max, N_{radii}, dr, ϵ, ρ)
  $N, d \leftarrow$ **shape**(P) ▷ N: number of atoms, d: dimensions of problem (2D/3D)
  $g(r)_{partial} \leftarrow$ **zeros**(N_{radii})
  for (r_{idx}, r) **in** $radii\_and\_indices$ **do**
    $r_{idx} \leftarrow$ **int**(r_{idx})
    for $i \leftarrow 0$ **to** $d - 1$ **do**
      for $j \leftarrow 0$ **to** $N - 1$ **do**
        if $P[j, i] - (r + dr) \geq min[i]$ **and** $P[j, i] + (r + dr) \leq max[i]$ **then**
          $idxs_{valid}[j] \leftarrow True$
        end if
      end for
    end for
    $particles_{valid} \leftarrow P[idxs_{valid}]$
    for $particle$ **in** $particles_{valid}$ **do** ▷ Parallel process in code using k-d tree
      $n_1 \leftarrow$ Number of particles within a distance of $r + dr - \epsilon$ from $particle$
      $n_2 \leftarrow$ Number of particles within a distance of r from $particle$
      $n \leftarrow n_1 - n_2$
      $g(r_{idx})_{partial} \leftarrow g(r_{idx})_{partial} + n$
    end for
    $n_{valid} \leftarrow$ **len**($particles_{valid}$)
    $shell_{vol} \leftarrow (4/3) \times \pi \times ((r + dr)^3 - r^3)$ **if** $d ==$ 3 **else** $\pi \times ((r + dr)^2 - r^2)$
    $g(r_{idx})_{partial} \leftarrow g(r_{idx})_{partial}/(n_{valid} \times shell_{vol} \times \rho)$
  end for
  return $g(r)_{partial}$
end function

3 Results

3.1 Quantifying the Variation in CO_2 Adsorption

We have applied the proposed method to quantify the variation in CO_2 adsorption in Al-Substituted SOD and CHA zeolites. First, we have estimated the equilibrium adsorption of CO_2 on SOD zeolite for different Al contents or Si/Al ratios. The Grand Canonical Monte Carlo (GCMC) based simulation results, as shown in Figure 6 for Si/Al ratio of 2 and 5 as well for pure-silica zeolites for which the ratio is very very large due to the absence of any Al atom. We see a significant increase in CO_2 adsorption with increasing Al content in the SOD lattice, confirming the past observations[9,11] that the presence of Al sites enhances CO_2 adsorption. Among the designs, we achieve the highest adsorption for Si/Al ratio of 2. We could not decrease the ratio (or, increase the Al-content) further. For example, an Si/Al ratio of one for SOD zeolite is infeasible according to the Lowenstein rule. The Si/Al ratio of 2 is the lowest that can be achieved with the SRU representation for SOD. A unit cell level representation would give us a ratio of 1.4 but will generate 792 possibilities, which can be reduced by eliminating repeating structures, but the number will still be high.

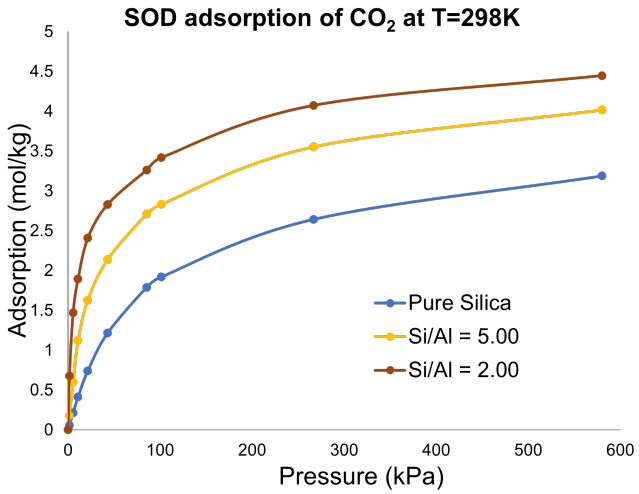


Figure 6: CO_2 adsorption on SOD framework with varying Si/Al ratios. The results are obtained using GCMC simulation.

We perform similar study of CO_2 adsorption on CHA zeolite framework for different Al substitutions. The variations in CO_2 adsorption on CHA zeolite framework for all selected Si/Al ratios are shown in Figure 7. For the same Si/Al ratio, the width of the band represents the variations in adsorption amount due to different enumerations of Al positions in the lattice. For large Si/Al ratio (e.g., Si/Al ratio equal to 11), we have narrow width, indicating small variations in adsorption. This is because the distributions of Al sites are sparse and, in most cases, they are located far from each other, thereby nullifying the effect of the geometric locations of Al sites within the zeolite framework. Most of the T-sites are Si atoms, and there is little variations in the overall composition of the framework in terms

of Al density. Similarly, we observe little variations in CO_2 adsorption when the Si/Al ratio is small (e.g., Si/Al = 1.0) due to the lack of compositional variation in terms of Si as most T-sites in this case consist of Al atoms.

A different result is obtained for moderate Si/Al ratios. The variations in CO_2 adsorption are most prominent at moderate Si/Al ratio. For example, we observe large variations in adsorption amount for Si/Al ratio of 2 and 5. We hypothesize that this can be due to the larger variations in Al density within the zeolite framework. For the same Si/Al ratio, the number of Al atoms per unit cell would be the same but the locations of the Al sites within the unit cell would be different for different Al distributions. Due to this, denser clusters of Al sites in specific locations may considerably enhance the adsorption of CO_2 compared to structures with more evenly distributed Al sites. This disparity in adsorption affinity is prominent when different clusters of T-atoms are present that give rise to non-uniform distribution of Al sites. Therefore, we predict that the variability in CO_2 adsorption amount would be large for moderate Si/Al ratios. We also observe that the variations are large at higher adsorption pressures. This is expected as higher system pressure allows higher adsorption for disparately located denser clusters of Al sites, thereby increasing the differences in local adsorption within the unit cells.

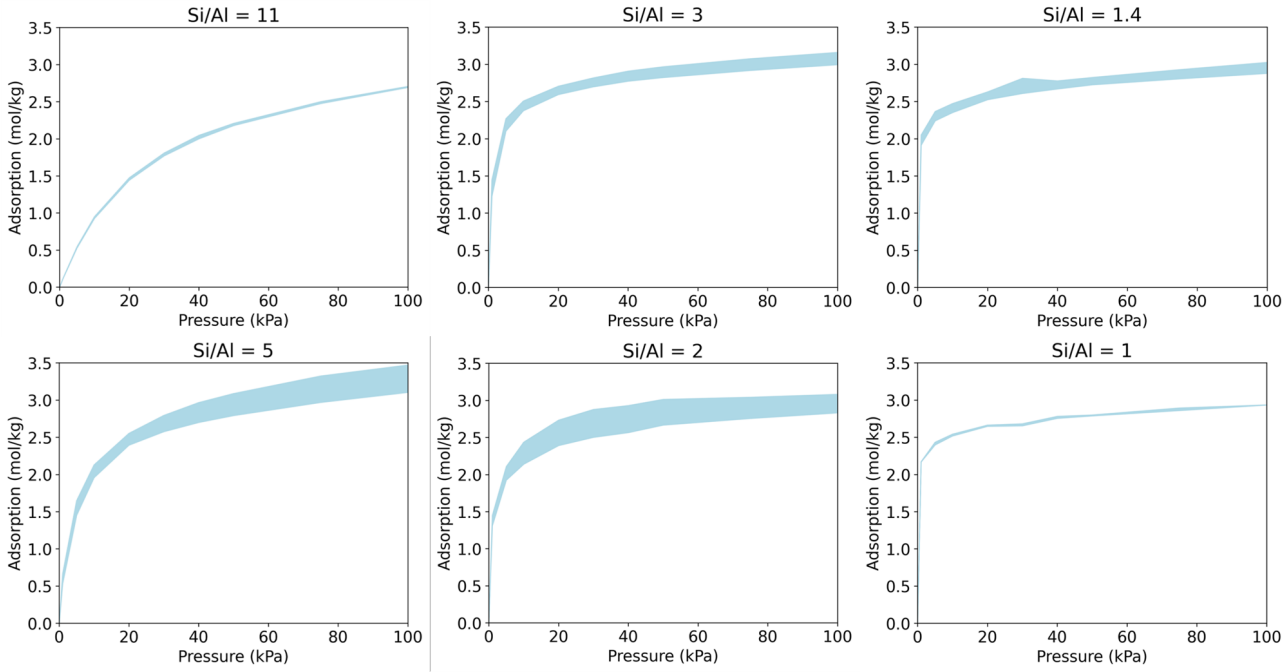


Figure 7: Variability in CO_2 adsorption on CHA framework with different Si/Al ratios.

We also observe that, the CO_2 adsorption does not always increase with the number of Al sites, and there exists an inflection point beyond which additional Al substitution leads to a decrease in adsorption. For CHA, CO_2 adsorption is maximum when Si/Al is 5.0 and drops on further Al addition. This result is consistent with previous studies in GIS-type frameworks.[9,10] The adsorption capacity for different Al substitutions per SRU cell (12-sized SRU for CHA) at 298 K, 100 kPa is shown in

figure 8. The maximum adsorption is observed at Si/Al ration of 5.0. Note that for this fixed ratio, there is a significant variation observed as well and we aim to quantify this variation and study the cause for it.

To understand the physio-chemical factors behind the optimal amount of Al that results in the highest adsorption, we need to examine the lattice structure. In a pure-silica lattice, the structure contains crevices for adsorption formed by the tetrahedral arrangement of Si and four oxygen atoms. When some of the Si atoms are replaced with Al, additional cations must be incorporated into the lattice to maintain charge balance, and these cations occupy certain lattice sites. The introduction of Al creates a charge field, enhancing adsorption. However, with further Al substitutions, the cations begin to occupy the CO_2 adsorption sites, leading to a decrease in adsorption. Therefore, there is an optimal amount of Al that maximizes adsorption before the charge-balancing cations occupy the adsorption sites.

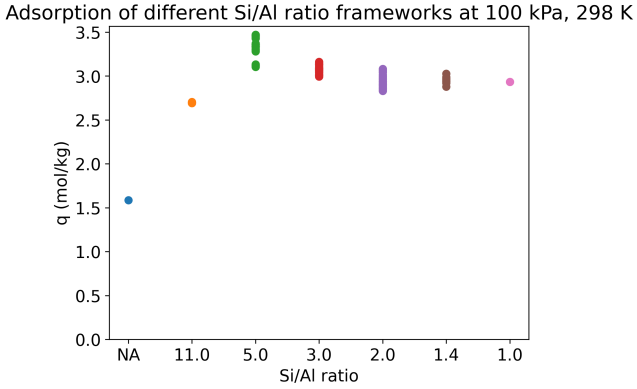


Figure 8: CO_2 adsorption isotherm for CHA frameworks with different Si/Al ratios.

We further investigate the effect of positions of Al on adsorption for the fixed Si/Al ratio of 5.0. There are 42 Al-substituted configurations of CHA zeolite frameworks with this ratio that are generated using the SRU approach. The variation observed in this case is significant ranging from 3.10 to 3.47 mol/kg at 100 kPa. This accounts for a 12% variation in adsorption. Figure 9 shows the variation in more detail and we can clearly see individual isotherms, some of which have higher adsorption than others. Few select unit cells of CHA zeolite with Si/Al ratio of 5.0 are shown in Figure 10. Each of the structures shown has exactly 6 Al atoms in the unit cell. However, due to difference in Al locations, they lead to different amounts of adsorption of CO_2.

While current CO_2 capture costs are estimated between USD 15 - 120 ton^{-1} CO_2,[48] a 12% improved adsorption capacity will have significant reduction in the costs. The implications on the process operability are even more critical as has been shown in the literature.[49] A process designed to operate with certain expected material performance in terms of purity and recovery will need significant modifications to operate at materials with properties differing by 12%. This is critical for not only simulation-based study, but also for optimization with processes that have narrow operability

domain. Even though a lot of hypothetical zeolites and crystalline materials have been proposed in literature previously, the high cost of experimental synthesis has hindered active research in this domain. Though the example selected in this study has focused on CO_2, zeolites are highly effective in selective adsorption, thus targeted adsorption of contaminants can also benefit from this approach. The low-density nature of zeolites results in a lighter weight in the overall set up facilitating ease of handling and potentially allow higher additive loading.

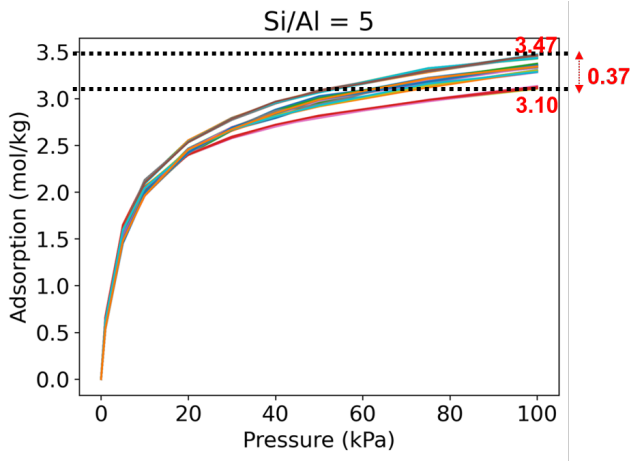


Figure 9: Variation observed in CO_2 adsorption on CHA frameworks with Si/Al=5.0 and different Al substitution locations.

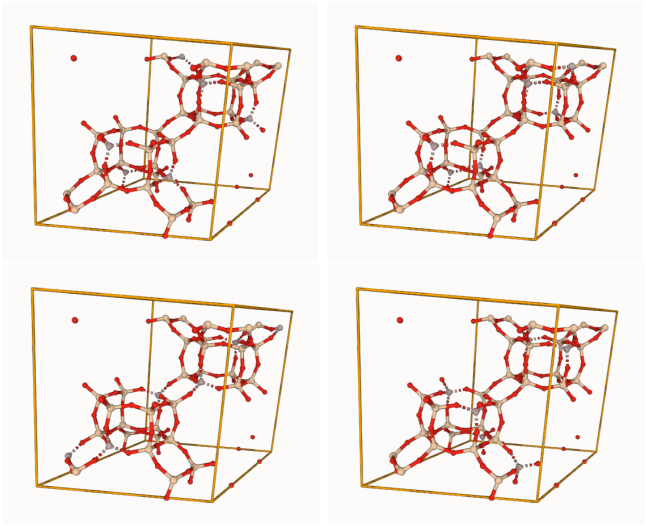


Figure 10: Examples of different Al substituted CHA framework configurations with Si/Al ration of 5.0. The colors for Al, Si, and O are grey, ivory, and red respectively. The bonds are depicted via dashed lines

We use the RDF to study how locations of Al affect adsorption. For the CHA framework with Si/Al=5.0, we have 42 different configurations of Al distributions, based on the SRU representation.

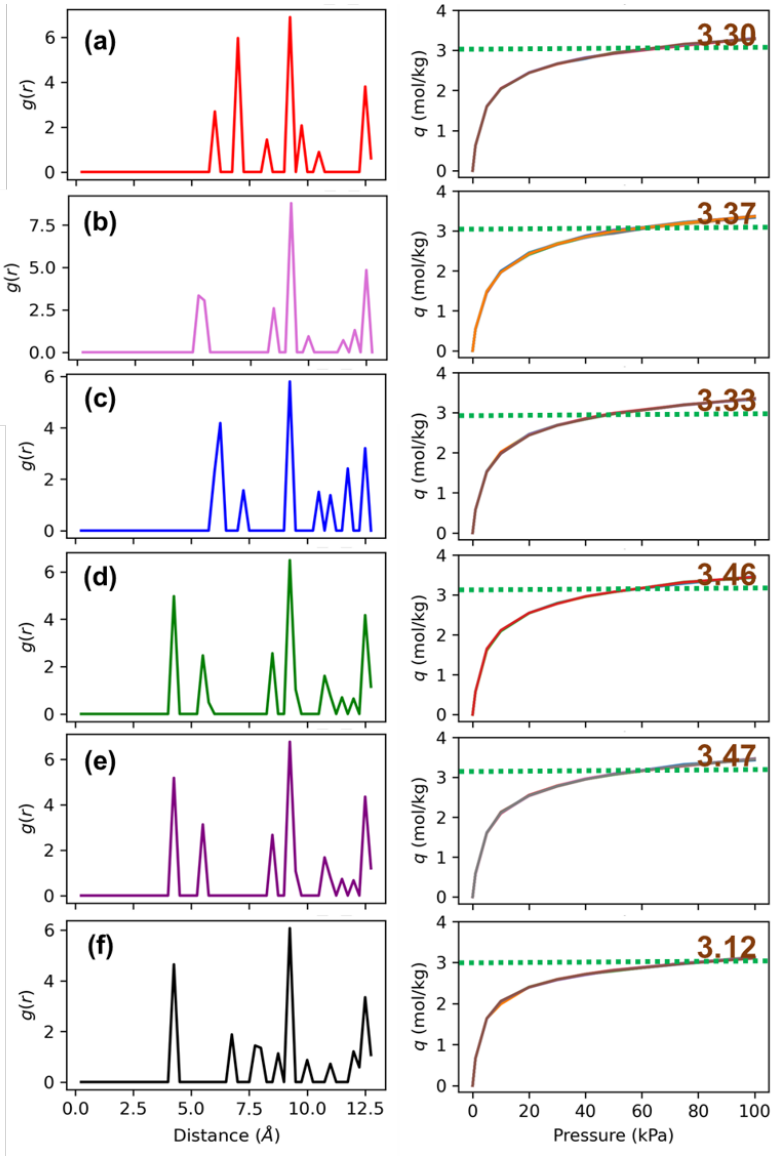


Figure 11: Radial distribution function $g(r)$ for: (a) group 1, (b) group 2, (c) group 3, (d) group 4, (e) group 5, and (f) group 6 shown on the left side and the corresponding adsorption isotherms shown for each group on the right side. all RDFs correspond to the same material class, CHA, with a Si/Al ratio of 5.

Of these, some of them are repeated due to the orientation of the 3-dimensional unit cell during the systematic enumeration. By observing the differences in RDF profiles of these structures, we classify them into six groups. These six different groups of RDF profiles are shown in Figure 11 along with the adsorption isotherms of the corresponding framework configurations. In all the RDFs, the highest peak is observed at 9.25Å which indicates the most frequently observed distances between two Al atoms in the lattice. Note the dimensions of the unit cell of CHA are $a = 13.675$Å $, b = 13.675$Å $, c = 14.767$Å $, \alpha = 90°, \beta = 90°, \gamma = 120°$, so the peak signifies the Al position due to repetition of the

SRU. The peak at 12.5Å also is observed in all RDF profiles. Since the unit cell has 6 Al substitutions (Si/Al=5.0), we are interested in the five tallest peaks since we are looking at other Al from one of the Al. Looking at the RDF of groups 4 and 5 (Fig. 11d and Fig. 11e) we observe that all the peaks are observed at the same values. However, due to the numerical precision floating decimals, there is some non-zero value at 5.75Å. Thus by visual inspection, we can still classify these groups together. Looking at the rest of the groups, we see significant differences in the locations of the peaks. Groups 4, 5, and 6 (Fig. 11d, Fig. 11e, and 11f) all have a peak at 4.25Å. Group 6 (11f) does not have the peak at 5.5Å. Groups 1, 2, and 3 (Fig. 11a, Fig. 11b, and 11c) all have their first peaks much further distances than 5Å.

The characterization of the positions of Al in the lattice is important to observe what factors lead to higher adsorption and contrasting factors that lead to lower adsorption. With that goal, we also plot adsorption isotherms from the GCMC data for each of these groups individually. These are shown on the right column in Figure 11. There is not much variation in the adsorption isotherms for the Al-substituted configurations with similar RDF. This justifies our classification of Al-substituted configurations in groups based on RDF. The acceptable variation for structures from the same group is 2% due to the stochastic nature of molecular simulations. Furthermore, groups 4 and 5 have very similar adsorption isotherms. This is expected due to the similarity in the RDFs for these groups. These groups (4 and 5) also show the highest saturation capacity or maximum equilibrium adsorption (3.46 and 3.47 mol/kg at 100 kPa). Both the extreme adsorption isotherms, i.e. the least and highest, are observed when the second peak in the RDF is at 4.25Å (groups 4, 5, and 6). On further inspection, we see that the highest adsorption is observed when a peak is present at 5.5Å. For groups 1, 2, and 3, since the adsorption isotherms are not extremes, we can comment on the features that stand out for all of them. Especially in group 2, which has the next best adsorption, shows a peak at 5.5Å. Thus we can infer that a peak at 5.5Å increases adsorption and when combined with a peak at 4.25Å, adsorption observed is highest (3.47 mol/kg).

3.2 High-throughput Screening of Al-Substituted Zeolite Configurations

The immediate next question is how do we use these insights to screen and select the structures with high adsorption from the large search space of over 2 million structures that exist for the same Si/Al ratio of 5.00. Note that insights generated from this analysis will also help experimental synthesis since only a few researchers are capable of synthesizing zeolites with Al located at selected sites.[50] With these insights, we screen among the 2 million structures and verify the validity of the filters that we identify, thus proposing novel material designs with desired properties. A quick preview of the screening is shown in figure 12. The first screening criterion is the Löwenstein's rule to the filter structures in which the Al-O-Al bond exists. This filter reduces the solution subset space to just 259,822 structures. Next, we introduce the criteria for the presence of a peak at 4.75Å and 5.5Å sequentially. Note that instead of hard constraints on the peaks, we allow for some flexibility by the

constraints that a peak should be less than 5Å and another peak between 5 - 5.75Å. The first filter reduces the search space by 13% to 226,469 and the second peak filter further reduces the subset size by 11% to 201,622.

While the peaks at 4.75Åand 5.5Åhave been observed for the CHA framework, the values at which peaks occur may vary for other zeolite frameworks. However, the methodology shown in this work can be applied to the zeolite framework of interest to identify the optimal Si/Al ratio, followed by the rest of the approach to identify the peaks. The contribution of this work highlights the specific case of CHA but expands beyond the single example considered. Using the SRU structure, one can reduce the potential structures and following this analysis, can identify zeolites with desired properties.

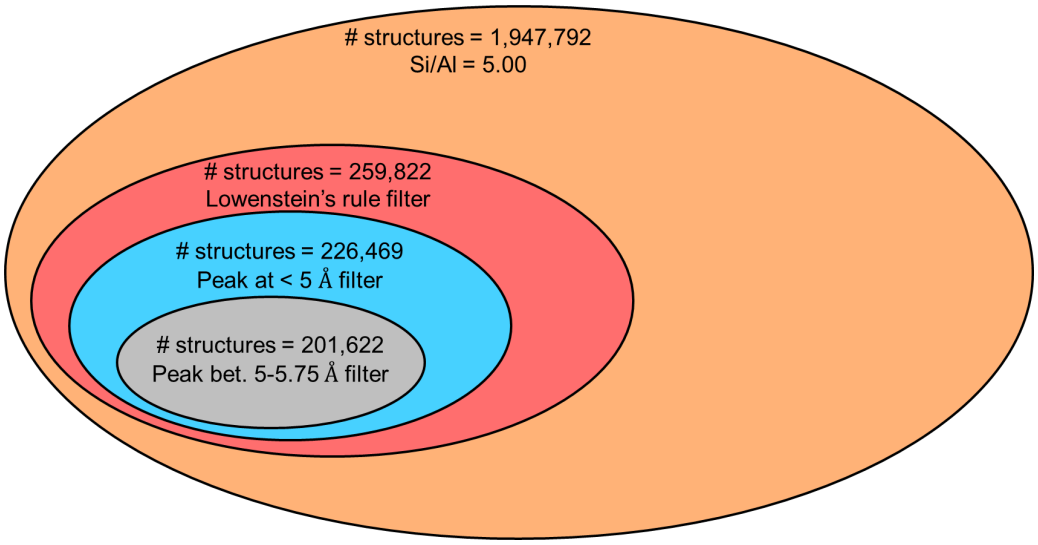


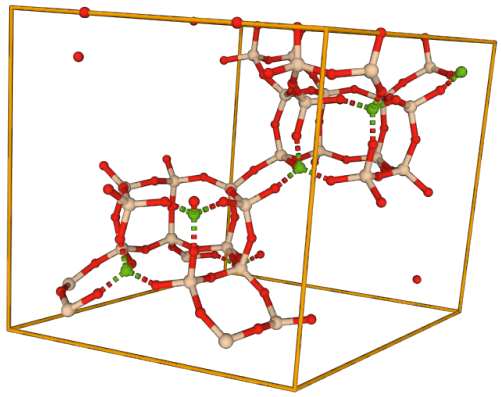
Figure 12: Subsets based on filters applied to the 2 million ways of generating Al-substituted structures.

To further screen from the selected subset based on the RDF profiles, we compare similar RDF profiles and observe the adsorption capacity of the corresponding structures. To compare RDF profiles, we use the cosine similarity and select the top 20 structures which have RDF similar to groups 4 and 5. The cosine similarity between two vectors, in this case, the RDF profiles, $\mathbf{A}$ and $\mathbf{B}$ can be computed using the following equation:

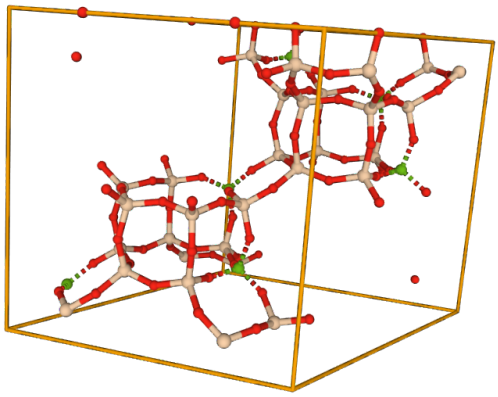
$$\text{cosine similarity}(\mathbf{A}, \mathbf{B}) = \frac{\mathbf{A} \cdot \mathbf{B}}{\|\mathbf{A}\|\|\mathbf{B}\|} \tag{2}$$

where, $\mathbf{A} \cdot \mathbf{B}$ represents the dot product of vectors $\mathbf{A}$ and $\mathbf{B}$, and $\|\mathbf{A}\|$ and $\|\mathbf{B}\|$ represent their respective Euclidean norms. With the top 20 structures, we run molecular simulations for all of these, of which the minimum adsorption is 3.17 mol/kg and four of the top 20 structures have adsorption between 3.45-3.48. Note that the cosine similarity is just one of the many metrics considered for vector comparisons. Further investigation into the top 20 structures shows that when we have similar RDF profiles (additional similarity check based on visual inspection for the intensity of the peaks), the adsorption is very high. The top four structures with high adsorption and similar

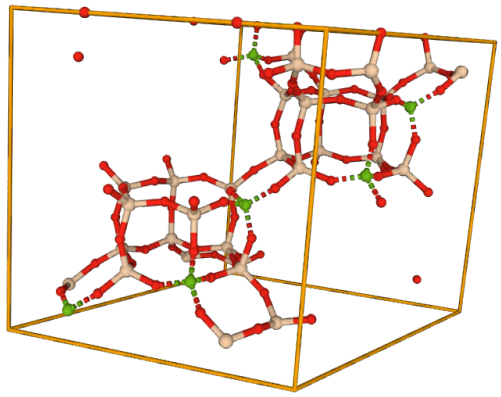
RDF profiles are shown in figure 13. The benefit of obtaining multiple structures that may lead to higher adsorption is of significant importance for experimental synthesis which allows for some flexibility in synthesizing amongst multiple structures. The RDF provides a unique mapping for the structure-property relationship.



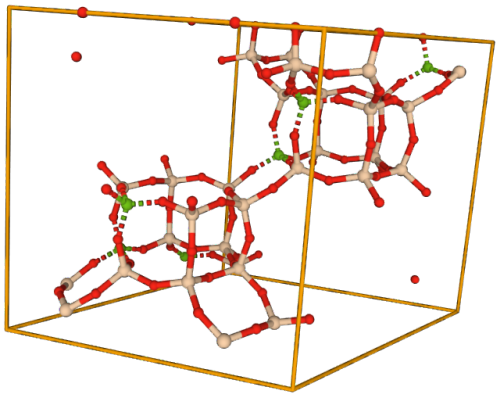
(a) CHA configuration 281759 with adsorption of 3.45 mol/kg



(b) CHA configuration 519226 with adsorption of 3.46 mol/kg



(c) CHA configuration 1043684 with adsorption of 3.46 mol/kg



(d) CHA configuration 1933297 with adsorption of 3.48 mol/kg

Figure 13: Top four identified configurations of Al-substituted CHA framework with high CO_2 adsorption at 298 K and 1 bar discovered via the proposed RDF-based high-throughput screening approach. all RDFs correspond to the same material class, CHA, with a Si/Al ratio of 5.

The physical inferences of the peaks at 4.75Å and 5.5Å are important to understand the underlying mechanisms of adsorption. On taking a closer look at the RDF profiles for SRU groups, we observe a high-intensity peak at 9Å. The repetition distance of the SRU structure is the cause of this peak. Specifically, the location of a specific node between two SRUs leads to the high-intensity peak. Given the two peaks closer to 5Å and another two peaks around 8.5Å, the physical interpretation is that Al are

locally close with some distance between them. This can be interpreted as some uniform distribution of Al in the lattice. To verify how Al-density affects adsorption, we also look at two structures with high local Al-density. The structures are shown in figure 14. The Al atoms are highlighted in green for contrast. Molecular simulation results for these structures validate our hypothesis of uniformity of distribution of Al. These structures show the adsorption of 3.15 and 2.7 mol/kg with lower adsorption observed for the densely packed Al. However, no comment can be made on the quantification of the uniformity and this requires large data of molecular simulations.

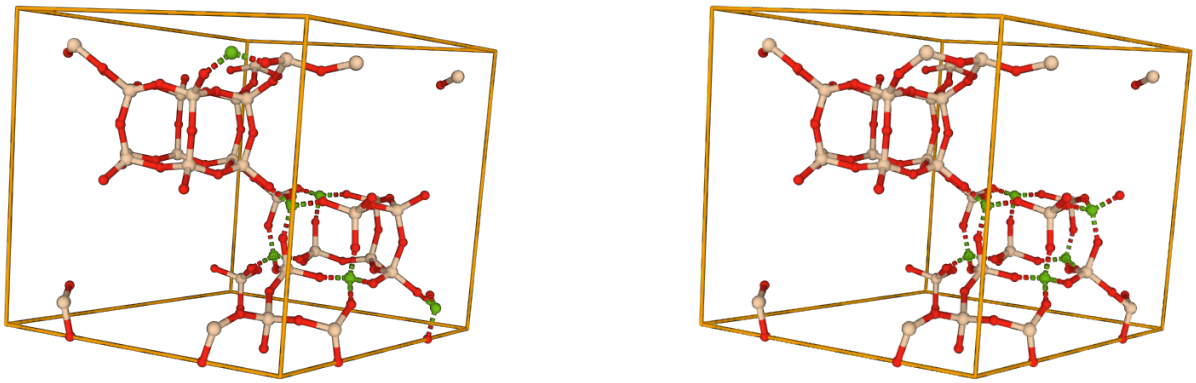


Figure 14: CHA structures with Si/Al=5.0 but local high density Al-substituted frameworks. Al atoms are shown in green.

4 Conclusions

We provide an efficient method for quantifying the variability in equilibrium gas adsorption on aluminosilicate zeolites due to the distribution of aluminum atoms within the framework. Such variability introduces considerable uncertainty in gas separation and storage applications, emphasizing the need for a systematic approach to address these challenges. We introduced a computational framework, based on a new representation of zeolite frameworks using single repeating units (SRUs), to perform efficient and selective enumeration of unique Al-substituted configurations. This computational approach avoids exhaustive searches while providing insights into equilibrium gas adsorption variability. Applying this methodology to CO_2 adsorption on SOD and CHA zeolites, we observed considerable variation in CO_2 adsorption based on Al atom locations, with the variability being most pronounced at moderate Si/Al ratios. At very high or low Si/Al ratios, the adsorption variability diminished. Our findings also reveal a non-monotonic relationship between the number of Al sites and CO_2 adsorption. Beyond a certain point, additional Al substitution reduces adsorption, highlighting the existence of an optimal Si/Al ratio for maximizing CO_2 uptake. Through systematic analysis, we can identify this optimal ratio and the corresponding Al site configurations that maximize adsorption performance.

As with any representation, the SRU has limitations that not all Si/Al ratios can be enumerated with a single unit. In those cases, multiple SRUs can be combined to make a bigger unit to represent these cases. With the enumerated structures, we performed GCMC simulations to obtain the adsorption isotherms for the same and study the variation that we initially hypothesized. Experimental data is not yet available for these cases and thus the GCMC parameters, specifically the interaction parameters may require further tuning to match with experimental data. However, the similar trend in variation in adsorption is expected to be observed in experimental results.

Our investigation further used radial distribution functions (RDFs) to pinpoint specific Al site arrangements that enhance CO_2 adsorption. These results underscore the utility of the SRU-based selective enumeration combined with RDF-based screening as a robust tool for rationally designing zeolites with tailored Al distributions. This approach paves the way for developing Al-substituted zeolite materials optimized for gas separation and storage applications. RDFs are found to be effective in describing the structural effects on the variation in gas adsorption. For example, upon studying the RDF of CHA zeolite we find that the distinguishing factor for high adsorption is the first two peaks being observed closely at 4.25Å and 5.5Å. Due to the unique mapping of the RDF, this representation can be used for modeling the structure-property relationship. The key contribution of this study is the methodology that allows identifying the optimal Si/Al ratio and the Aluminum RDF distribution which is demonstrated via the case study of CHA. Further, if zeolite synthesis can be done with targeted Al-substitution sites, then we can extend the current approach to discover new zeolite configurations with optimal desired properties.

Acknowledgment

M.M.F.H. gratefully acknowledge partial funding support from the NSF CDS&E grant 2245474. Portions of this research were conducted with the advanced computing resources provided by Texas A&M High Performance Research Computing.

Supporting Information

The supporting information consists of GCMC parameters used in this study.

Literature Cited

1. Mostapha Dakhchoune, Luis Francisco Villalobos, Rocio Semino, Lingmei Liu, Mojtaba Rezaei, Pascal Schouwink, Claudia Esther Avalos, Paul Baade, Vanessa Wood, Yu Han, et al. Gas-sieving zeolitic membranes fabricated by condensation of precursor nanosheets. *Nature Materials*, 20(3):362–369, 2021.

2. Jens Weitkamp. Zeolites and catalysis. *Solid State Ionics*, 131(1-2):175–188, 2000.

3. Javier Pérez-Ramírez, Claus H Christensen, Kresten Egeblad, Christina H Christensen, and Johan C Groen. Hierarchical zeolites: enhanced utilisation of microporous crystals in catalysis by advances in materials design. *Chem. Soc. Rev.*, 37(11):2530–2542, 2008.

4. MM Faruque Hasan, Eric L First, and Christodoulos A Floudas. Cost-effective CO_2 capture based on in silico screening of zeolites and process optimization. *Physical Chemistry Chemical Physics*, 15(40):17601–17618, 2013.

5. Eric L First, MM Faruque Hasan, and Christodoulos A Floudas. Discovery of novel zeolites for natural gas purification through combined material screening and process optimization. *AIChE Journal*, 60(5):1767–1785, 2014.

6. Atlas of prospective zeolite structures. http://www.hypotheticalzeolites.net. (accessed Jan 9, 2021).

7. Chi-Ta Yang, Amber Janda, Alexis T Bell, and Li-Chiang Lin. Atomistic investigations of the effects of Si/Al ratio and al distribution on the adsorption selectivity of n-alkanes in brønsted-acid zeolites. *The Journal of Physical Chemistry C*, 122(17):9397–9410, 2018.

8. Akhil Arora, Shachit S Iyer, and MM Faruque Hasan. Computational material screening using artificial neural networks for adsorption gas separation. *The Journal of Physical Chemistry C*, 124(39):21446–21460, 2020.

9. Hyun June Choi and Suk Bong Hong. Effect of framework Si/Al ratio on the mechanism of CO_2 adsorption on the small-pore zeolite gismondine. *Chemical Engineering Journal*, 433:133800, 2022.

10. Hyun June Choi, Donghui Jo, and Suk Bong Hong. Effect of framework Si/Al ratio on the adsorption mechanism of CO_2 on small-pore zeolites: Ii. merlinoite. *Chemical Engineering Journal*, 446:137100, 2022.

11. Sebastian Prodinger, Rama S Vemuri, Tamas Varga, B Peter McGrail, Radha Kishan Motkuri, and Miroslaw A Derewinski. Impact of chabazite ssz-13 textural properties and chemical composition on co 2 adsorption applications. *New Journal of Chemistry*, 40(5):4375–4385, 2016.

12. Danielle Bonenfant, Mourad Kharoune, Patrick Niquette, Murielle Mimeault, and Robert Hausler. Advances in principal factors influencing carbon dioxide adsorption on zeolites. *Science and technology of advanced materials*, 9(1):013007, 2008.

13. Ho Viet Thang, Lukáš Grajciar, Petr Nachtigall, Ota Bludskỳ, Carlos Otero Areán, Eva Frỳdová, and Roman Bulanek. Adsorption of CO_2 in FAU zeolites: Effect of zeolite composition. *Catalysis Today*, 227:50–56, 2014.

14. Tom Remy, Sunil A Peter, Leen Van Tendeloo, Stijn Van der Perre, Yannick Lorgouilloux, Christine EA Kirschhock, Gino V Baron, and Joeri FM Denayer. Adsorption and separation of CO_2 on KFI zeolites: effect of cation type and Si/Al ratio on equilibrium and kinetic properties. *Langmuir*, 29(16):4998–5012, 2013.

15. Firas N Ridha and Paul A Webley. Anomalous henry's law behavior of nitrogen and carbon dioxide adsorption on alkali-exchanged chabazite zeolites. *Separation and purification technology*, 67(3):336–343, 2009.

16. Firas N Ridha and Paul A Webley. Entropic effects and isosteric heats of nitrogen and carbon dioxide adsorption on chabazite zeolites. *Microporous and mesoporous materials*, 132(1-2):22–30, 2010.

17. MM Faruque Hasan, Richard C Baliban, Josephine A Elia, and Christodoulos A Floudas. Modeling, simulation, and optimization of postcombustion CO_2 capture for variable feed concentration and flow rate. 2. pressure swing adsorption and vacuum swing adsorption processes. *Industrial & engineering chemistry research*, 51(48):15665–15682, 2012.

18. MM Faruque Hasan, Eric L First, Fani Boukouvala, and Christodoulos A Floudas. A multi-scale framework for CO_2 capture, utilization, and sequestration: CCUS and CCU. *Computers & Chemical Engineering*, 81:2–21, 2015.

19. Akhil Arora and MM Faruque Hasan. Flexible oxygen concentrators for medical applications. *Scientific reports*, 11(1):14317, 2021.

20. Shachit S Iyer, Salih E Demirel, and MM Faruque Hasan. Combined natural gas separation and storage based on in silico material screening and process optimization. *Industrial & Engineering Chemistry Research*, 57(49):16727–16750, 2018.

21. MM Faruque Hasan, Eric L First, and Christodoulos A Floudas. Discovery of novel zeolites and multi-zeolite processes for p-xylene separation using simulated moving bed (smb) chromatography. *Chemical Engineering Science*, 159:3–17, 2017.

22. Tingting Liu, Eric L First, MM Faruque Hasan, and Christodoulos A Floudas. A multi-scale approach for the discovery of zeolites for hydrogen sulfide removal. *Computers & Chemical Engineering*, 91:206–218, 2016.

23. Shachit S Iyer and MM Faruque Hasan. Mapping the material-property space for feasible process operation: application to combined natural-gas separation and storage. *Industrial & Engineering Chemistry Research*, 58(24):10455–10465, 2019.

24. John R Di Iorio, Sichi Li, Casey B Jones, Claire T Nimlos, Yujia Wang, Eduard Kunkes, Vivek Vattipalli, Subramanian Prasad, Ahmad Moini, William F Schneider, et al. Cooperative

and competitive occlusion of organic and inorganic structure-directing agents within chabazite zeolites influences their aluminum arrangement. *Journal of the American Chemical Society*, 142(10):4807–4819, 2020.

25. Jialiang Li, Mingkun Gao, Wenfu Yan, and Jihong Yu. Regulation of the Si/Al ratios and al distributions of zeolites and their impact on properties. *Chemical Science*, 14(8):1935–1959, 2023.

26. Brandon C Knott, Claire T Nimlos, David J Robichaud, Mark R Nimlos, Seonah Kim, and Rajamani Gounder. Consideration of the aluminum distribution in zeolites in theoretical and experimental catalysis research. *ACS Catalysis*, 8(2):770–784, 2018.

27. John R Di Iorio and Rajamani Gounder. Controlling the isolation and pairing of aluminum in chabazite zeolites using mixtures of organic and inorganic structure-directing agents. *Chemistry of Materials*, 28(7):2236–2247, 2016.

28. Nao Tsunoji, Kazuyoshi Tsuchiya, Naoto Nakazawa, Satoshi Inagaki, Yoshihiro Kubota, Toshiki Nishitoba, Toshiyuki Yokoi, Takeshi Ohnishi, Masaru Ogura, Masahiro Sadakane, et al. Multiple templating strategy for the control of aluminum and phosphorus distributions in afx zeolite. *Microporous and Mesoporous Materials*, 321:111124, 2021.

29. Sen Wang, Yue He, Weiyong Jiao, Jianguo Wang, and Weibin Fan. Recent experimental and theoretical studies on al siting/acid site distribution in zeolite framework. *Current Opinion in Chemical Engineering*, 23:146–154, 2019.

30. Daan Frenkel and Berend Smit. *Understanding molecular simulation: from algorithms to applications*, volume 1. Elsevier, 2001.

31. Arijit Chakraborty, Akhilesh Gandhi, MM Faruque Hasan, and Venkat Venkatasubramanian. Discovering zeolite adsorption isotherms: a hybrid ai modeling approach. In *Computer Aided Chemical Engineering*, volume 53, pages 511–516. Elsevier, 2024.

32. Yuming Gu, Ziteng Liu, Changzhou Yu, Xu Gu, Lili Xu, Yang Gao, and Jing Ma. Zeolite adsorption isotherms predicted by pore channel and local environmental descriptors: feature learning on dft binding strength. *The Journal of Physical Chemistry C*, 124(17):9314–9328, 2020.

33. Yue Yu, Xu Li, Rajamani Krishna, Yuchuan Liu, Yuanzheng Cui, Jianfeng Du, Zhiqiang Liang, Xiaowei Song, and Jihong Yu. Enhancing CO_2 adsorption and separation properties of aluminophosphate zeolites by isomorphous heteroatom substitutions. *ACS applied materials & interfaces*, 10(50):43570–43577, 2018.

34. Zhonglin Cao, Rishikesh Magar, Yuyang Wang, and Amir Barati Farimani. Moformer: self-supervised transformer model for metal–organic framework property prediction. *Journal of the American Chemical Society*, 145(5):2958–2967, 2023.

35. Jingqi Wang, Jiapeng Liu, Hongshuai Wang, Musen Zhou, Guolin Ke, Linfeng Zhang, Jianzhong Wu, Zhifeng Gao, and Diannan Lu. A comprehensive transformer-based approach for high-accuracy gas adsorption predictions in metal-organic frameworks. *Nature Communications*, 15(1):1904, 2024.

36. Aahil Khambhawala, Chi Ho Lee, Silabrata Pahari, Paul Nancarrow, Nabil Abdel Jabbar, Mahmoud M El-Halwagi, and Joseph Sang-Il Kwon. Advanced transformer models for structure-property relationship predictions of ionic liquid melting points. *Chemical Engineering Journal*, page 158578, 2024.

37. Fang Wu, Dragomir Radev, and Stan Z Li. Molformer: Motif-based transformer on 3d heterogeneous molecular graphs. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pages 5312–5320, 2023.

38. Joseph Sang-Il Kwon. Adding big data into the equation. *Nature Chemical Engineering*, 1(11):724–724, 2024.

39. Akhilesh Gandhi and MM Faruque Hasan. A graph theoretic representation and analysis of zeolite frameworks. *Computers & Chemical Engineering*, 155:107548, 2021.

40. Akhilesh Gandhi and MM Faruque Hasan. Machine learning for the design and discovery of zeolites and porous crystalline materials. *Current Opinion in Chemical Engineering*, 35:100739, 2022.

41. David Dubbeldam, Sofía Calero, Donald E Ellis, and Randall Q Snurr. RASPA: molecular simulation software for adsorption and diffusion in flexible nanoporous materials. *Molecular Simulation*, 42(2):81–101, 2016.

42. E Garcia-Perez, D Dubbeldam, Th LM Maesen, and Sofia Calero. Influence of cation na/ca ratio on adsorption in lta 5a: a systematic molecular simulation study of alkane chain length. *The Journal of Physical Chemistry B*, 110(47):23968–23976, 2006.

43. David Dubbeldam, Houston Frost, Krista S Walton, and Randall Q Snurr. Molecular simulation of adsorption sites of light gases in the metal-organic framework irmof-1. *Fluid Phase Equilibria*, 261(1-2):152–161, 2007.

44. Almudena Garcia-Sanchez, Conchi O Ania, José B Parra, David Dubbeldam, Thijs JH Vlugt, Rajamani Krishna, and Sofia Calero. Transferable force field for carbon dioxide adsorption in zeolites. *The Journal of Physical Chemistry C*, 113(20):8814–8820, 2009.

45. Koichi Momma and Fujio Izumi. Vesta: a three-dimensional visualization system for electronic and structural analysis. *Journal of Applied crystallography*, 41(3):653–658, 2008.

46. Sajjad Ghojavand, Benoit Coasne, Edwin B Clatworthy, Rémy Guillet-Nicolas, Philippe Bazin, Marie Desmurs, Luis Jacobo Aguilera, Valérie Ruaux, and Svetlana Mintova. Alkali metal cations influence the CO_2 adsorption capacity of nanosized chabazite: modeling vs experiment. *ACS Applied Nano Materials*, 5(4):5578–5588, 2022.

47. Zahra Pourmahdi and Hafez Maghsoudi. Adsorption isotherms of carbon dioxide and methane on cha-type zeolite synthesized in fluoride medium. *Adsorption*, 23:799–807, 2017.

48. Xiang Yun Debbie Soo, Johnathan Joo Cheng Lee, Wen-Ya Wu, Longgang Tao, Cun Wang, Qiang Zhu, and Jie Bu. Advancements in co2 capture by absorption and adsorption: A comprehensive review. *Journal of CO2 Utilization*, 81:102727, 2024.

49. Reza Haghpanah, Arvind Rajendran, Shamsuzzaman Farooq, and Iftekhar A Karimi. Optimization of one-and two-staged kinetically controlled co2 capture processes from postcombustion flue gas on a carbon molecular sieve. *Industrial & Engineering Chemistry Research*, 53(22):9186–9198, 2014.

50. Siddarth H Krishna, Anshuman Goswami, Yujia Wang, Casey B Jones, David P Dean, Jeffrey T Miller, William F Schneider, and Rajamani Gounder. Influence of framework Al density in chabazite zeolites on copper ion mobility and reactivity during NOx selective catalytic reduction with NH3. *Nature Catalysis*, 6(3):276–285, 2023.