Uncovering Zeolitic Imidazolate Frameworks through Digitization of Chemical Insights

Soochan Lee[1], Hyein Jeong[1], Sungyeop Jung[1], Yeongjin Kim[1], Eunchan Cho[1], Joohan Nam[1], D. ChangMo Yang[1], Hyunchul Oh[1,2], and Wonyoung Choe*[,1,2]

[1]Department of Chemistry, Ulsan National Institute of Science and Technology, Ulsan 44919, Republic of Korea.

[2]Graduate School of Carbon Neutrality, Ulsan National Institute of Science and Technology, Ulsan 44919, Republic of Korea.

*Corresponding author: choe@unist.ac.kr

Abstract

The vast number of computational predictions presents challenges when transitioning from structural models to experimental confirmations. To address this challenge, we digitized chemical intuition into the discovery process, focusing on zeolitic imidazolate frameworks (ZIFs). Despite their potential, the limited topologies by the "zeolite conundrum" and an unclear synthetic roadmap have hindered ZIF discovery. We propose a data-driven approach for ZIF discovery using structural descriptors of known materials as a screening tool. From over 4 million zeolite structures, we identified potential ZIF candidates based on O−T−O angle differences, vertex symbols, and T−O−T angles. Energy calculations enabled the ranking of synthesizability of ZIFs, resulting in the successful synthesis of three ZIFs with two unprecedented topologies, UZIF-31 (*uft*1) and UZIF-32, -33 (*uft*2). Notably, UZIF-33 demonstrated remarkable selective adsorption of CO_2. This work underscores the synergistic potential of combining structure predictions with chemical intuition to advance the field of material discovery.

Introduction

Advanced computational methods, including materials informatics[1,2] and artificial intelligence[3-5], are now frequently used to accelerate materials discovery, addressing urgent sustainability needs[6-8]. Despite their successful advancements in materials discovery, the vast number of material candidates suggested by these methods often presents a significant hurdle for experimental verification[9-12]. To overcome this hurdle, an integrated approach is essential where computational predictions are closely aligned with targeted experimental efforts to efficiently realize the discovery of new materials[11-14]. For example, incorporating the chemical intuition of expert chemists into structure prediction can accelerate the materials discovery[15-19], similar to Lipinski's rule of five in drug discovery[20-22]. Digitizing the chemical intuition not only enhances the screening process[23] by reducing the computational costs[24] associated with large datasets but also lays out a blueprint for uncovering new functional materials[25-27]. In this context, we have focused our innovative approach on zeolitic imidazolate frameworks (ZIFs)[28-31], the metal-organic analogs of zeolites[32], to underscore the potential of this integrated strategy.

ZIFs mimic the tetrahedral units (TO_4, T = Si or Al) in zeolites through coordination bonds formed between four-coordinated metal cations and imidazolates (Im, hereafter). These materials hold the potential to supplement the zeolite ecosystems in applications such as gas separation[33-35], catalysis[36-38], and battery technologies[39-41], due to exceptional chemical stability and high porosity[29,30]. Despite strong attention to utilizing ZIFs, there are only 38 known ZIF topologies, significantly fewer than 255 realized zeolites[32]. Furthermore, the number of synthesized ZIF topologies is overwhelmingly less than over 4 million hypothetical zeolites[42]. This phenomenon is similar to the “zeolite conundrum”, a well-known challenge in the porous materials[43]. Another hurdle is the lack of clear synthetic strategies for ZIFs, in contrast to the case of MOFs, where reticular chemistry often serves as a blueprint for the discovery of MOF

phases[44-46]. The restricted coordination geometry of Im has further hindered the discovery of novel ZIF phases[47,48]. Even with extensive efforts in high-throughput synthesis, the chances of discovering new ZIF crystals remain extremely low, as reported by Yaghi *et al*[30]. Our objective is to integrate structural features identified through chemical intuition, together with energy calculations of ZIFs, allowing us to pinpoint synthesizable ZIFs from a vast pool of hypothetical zeolite databases[9,49].

Here, we present a new dataset for potential ZIF candidates extracted from over 4 million zeolite structures. This dataset was derived from intuitive structural descriptors such as O−T−O angle difference, vertex symbol, and T−O−T angle, all of which play a pivotal role in the construction of zeolitic networks[50-52]. Through energy calculations, we effectively ranked the synthesizability of affordable ZIFs for targeted synthesis. To our surprise, our approach has yielded the discovery of three previously unknown ZIFs, UZIF-31 (*uft*1), UZIF-32 (*uft*2), and UZIF-33 (*uft*2), featuring unprecedented topologies matched with our highly synthesizable dataset. Structural analysis unveils the internal hydrogen bonding network for framework stabilization. Notably, UZIF-33 stands out for exceptional adsorption selectivity in CO_2 over CH_4, positioning it as a model material for methane purification[53], contributing to carbon neutrality[54].

Result and discussion

Target system from the zeolite database. Given the structural interchangeability between zeolites and ZIFs, the data-driven discovery for feasible ZIFs was conducted by considering both the hypothetical zeolite database[42] and the International Zeolite Association (IZA) database[32]. In total, these databases have a sum of 4,450,797 zeolites (4,450,542 in the hypothetical database and 255 in IZA). The multitude of zeolitic networks, encompassing hypothetical zeolite databases, contains an extensive amount of structural data that complicates the pursuit of future targets for ZIF discovery (Supplementary Table 1). In comparison to the successful reticular design of MOFs[44-46], limiting the topological system could prove efficient for pinpointing a subset of zeolite-like materials. Initially, our focus was directed toward analyzing the number of nodes (the number of symmetrically identical nodes; *n*-nodal) present in synthesized ZIF topologies (Supplementary Figs. 1 and 2). Notably, a major portion of ZIFs were either uninodal or binodal nets (79% or 30 out of 38). Of particular interest were the trinodal topologies in ZIFs, which were notably scarce among the 38 ZIF topologies (merely 5%, 2 out of 38), even in tetranodal ZIFs accounting for 10% (4 out of 38). We hypothesized that these trinodal ZIFs might serve as missing links and hold high potential for introducing new ZIFs. Consequently, these trinodal topologies were identified as screening targets.

Screening process using structural descriptors. Employing efficient descriptors to screen through vast entries in databases could prove to be an effective method for identifying synthesizable materials. In earlier work, our group developed structural descriptors for feasible ZIFs, such as node number, vertex symbol, and T−O−T angle, which are associated with the topological frameworks within zeolitic networks[49]. In this study, we will utilize the intuition-based structural descriptors, specifically O−T−O angle difference, vertex symbol, and T−O−T angle, derived from common geometrical features in ZIFs.

In zeolites, the O−T−O angles are kept at the ideal angle for a 4-coordinated tetrahedral node (approximately 109.5°) to provide stability to the frameworks[52]. Nevertheless, the hypothetical zeolite database contains numerous impractical zeolite structures marked by highly deviated tetrahedral geometries from the ideal O−T−O angle, particularly encompassing energetically unstable zeolites (Fig. 1a). Avoiding the severely distorted geometry within the parent zeolite structures proves to be an effective approach when dealing with an extensive dataset. To manage this, we defined a cut-off parameter, ω (the angle difference between the maximum and minimum O−T−O angles), with a threshold of < 30° for ω in ZIFs. This selection was based on the observation that nearly all ZIFs (88%; 29 out of 33, the exclusion for those featuring partially 3-c nets; **moc** and **moz**, non-Zn ZIFs; **mog**, **ict**, and **dia-c**) exhibited ω values below 30° (Fig. 1b, Supplementary Figs. 3 and 4). The first descriptor is the limitation to the O−T−O angle difference (ω < 30°) in zeolitic topologies (Fig. 1c).

The second and third descriptors, namely the vertex symbol (4, 6, 8, or 12 combinations) and the T−O−T angle (θ > 130°) were adopted from our recent publication due to their proven efficacy in screening for reliable ZIF structures[49]. Specifically, 79% (30 out of 38) of ZIFs shared vertex symbols involving combinations of these ring sizes, 4, 6, 8, or 12 combinations (Fig. 1d and Supplementary Table 2). The distribution of T−O−T angles was notably narrow, with 97% (37 out of 38) of ZIF topologies displaying angles that did not exceed 130° (Fig. 1e and Supplementary Fig. 5). Thus, our second and final descriptors are defined by the limitation of vertex symbols to 4, 6, 8, or 12 combinations and the T−O−T angle exceeding 130°, respectively.

Following the approach guided by the structural descriptors, we initiated data-filtering in the following sequence: starting with O−T−O angle difference (ω < 30°), followed by vertex symbols (4, 6, 8, and 12), and then T−O−T angles (θ > 130°). The number of initial zeolites

with trinodal nets drastically reduced to 31,675 from the total 4,450,797 zeolites. The first descriptor, O−T−O angle difference ($\omega < 30°$) was applied to 31,637 hypothetical zeolites and resulted in only 10,219 structures. Subsequently, after the applying second (vertex symbol; 4, 6, 8, and 12) and third (T−O−T angle; $\theta > 130°$) descriptors, a total of 420 topologies, including hypothetical and IZA zeolites, remained in the pool for future ZIFs, marking a remarkable reduction from over 4 million zeolite structures (Fig. 1f). Out of 420 ZIF topologies, 23 topologies are found in IZA (BPH, AFS, ATS, SFW, AEI, AFT, SAV, PWN, OSI, MSO, UEI, AWO, AFV, OWE, AEL, THO, and AFG) and the reticular chemistry structure resource (RCSR)[55] database (**reo-t**, **ucb**, **tdc**, **gcd**, **cfd**, and **cmc**), including a realized ZIF topology[47] (**ucb**; ZIF-412, -413, and -414) (Supplementary Table 3).

Synthesizable future ZIFs. Based on the 420 topologies that resulted from our structural descriptors, we proceeded to carry out *in silico* construction of future ZIFs (f-ZIFs) using a topologically based crystal constructor (ToBaCCo)[56] (Supplementary Fig. 6). Once the initial structures were generated, we performed geometry optimization using the force field for ZIFs[57] (ab initio derived MOF-FF for ZIFs; hereafter FF) through the LAMMPS[58]. For the assessment of cell volume changes in known ZIFs, we applied both FF and density functional theory (DFT) calculations, as demonstrated by testing against experimental cell volumes (Supplementary Figs. 7 and 8). The resulting volume change in known ZIFs with FF was found to be a mere -1.41% compared to the experimental cell volume, displaying minimal divergence from the DFT outcome, which offers greater computational precision but also consumes more time, and registers at -0.93%. Several results have deliberated the synthesizability of porous crystalline materials[19,59,60] using low-level calculations to calculate the framework energy instead of the more computationally intensive periodic DFT method.

We also conducted energy calculations on 30 known ZIFs (converted to $Zn(Im)_2$ formula) after optimization using the FF and then plotted the energy against the framework density (nm^{-3}), *T/V* (Supplementary Fig. 9). Interestingly, ZIF topologies exhibited a similar trend to several DFT results[49,61,62], where the energy increased as the density decreased, with the most stable **zni** topology. We proceeded with the assumption that synthesizable ZIFs would likely align closely with real ZIFs and thus calculated a prediction distribution around the linear regression of known ZIF topologies. Intriguingly, the prediction interval within ±2σ encompassed 93 % of known ZIFs (two exceptions; **ana** and **zni** within ±3σ). This analysis was further extended to the newly generated 420 f-ZIFs, which were plotted alongside the known ZIF set (Fig. 2a). In crystal structure prediction, porous frameworks that possess lower energy at a given density are more likely to be achievable through synthesis[14,63-67]. Based on this approach, we established a categorization: Tier 1 f-ZIFs with high synthesizability were located under the regression line (21%, 90 out of 420), Tier 2 f-ZIFs with moderate synthesizability were positioned between the regression line to +1σ range (56%, 233 out of 420), and Tier 3 f-ZIFs with low synthesizability fell beyond the +1σ range (23%, 97 out of 420) (Fig. 2b and Supplementary Fig. 10). Remarkably, all Tier 1 f-ZIFs successfully passed the validation criteria based on the first descriptor, which involves assessing the O−T−O angle difference (ω < 30°), in contrast to several Tier 2 and 3 f-ZIFs that exceeded the established threshold for reliable O−T−O angle differences (Fig. 2c and Supplementary Fig. 12). This observation emphasizes the high efficiency of categorizing synthesizability by energy level in revealing chemically plausible structures. In essence, the ranking of synthesizability was established by the proximity to the -3σ line (Fig. 2d). Consequently, all 420 f-ZIFs were assigned names based on their synthesizability ranking, denoted as T-*n* (with T-1 being the most feasible ZIF and T-420 representing the least feasible). An overview of all 90 Tier 1 f-ZIFs (T-1 to T-90) is illustrated in Fig. 3. Significantly, we were able to experimentally validate

our rational strategy through the successful synthesis of two ZIF topologies, T-12 and T-50, both of which are in Tier 1 (Figs. 2d and 4a).

Synthetic strategy. In ZIF chemistry, the link-link interactions[47], mainly van der Waals forces, between Im linkers stand as a crucial factor in stabilizing distinct cages and topologies, and introducing hydrogen bonds within ZIF crystals presents a challenging work[48]. Our objective was to introduce weak hydrogen bonds[68] into ZIF construction using derivatives of azabenzimidazole (AzbIm). A unique combination of Im and three AzbIm derivatives, namely 4-azabenzimidazole (4-AzbIm), 5-azabenzimidazole (5-AzbIm), and purine (Pur), successfully facilitated the synthesis of new ZIF topologies (Fig. 4a). Following more than 1,600 synthesis attempts, UZIF-31 [$Zn(Im)_{0.88}(4\text{-}AzbIm)_{1.12}$], UZIF-32 [$Zn(Im)_{1.25}(5\text{-}AzbIm)_{0.75}$], and UZIF-33 [$Zn(Im)_{1.25}(Pur)_{0.75}$] were synthesized through solvothermal synthesis using a mixture of $Zn(NO_3)_2{\cdot}6H_2O$, Im, and the corresponding AzbIm derivative (4-AzbIm, 5-AzbIm, or Pur) in *N*,*N*-dimethylformamide (DMF) at 120 °C for 24−96 hours with no additives (See the detail in Methods section). The bulk purity of all crystals was verified by comparing their PXRD patterns to simulated ones (Supplementary Figs. 18-20).

Structural analysis. UZIF-31 (*uft*1) with Im and 4-AzbIm crystallized in the orthorhombic *Imma* (No. 74) space group (Supplementary Table 6). The whole crystal structure of UZIF-31 was determined by single-crystal X-ray diffraction (SCXRD) with help from the confirmed composition using ^{1}H-NMR spectroscopy (Supplementary Fig. 24). The topology of UZIF-31 exhibited an unprecedented 4,4,4-c topology with the point symbol of $(4^2.6^2.8^2)(4^3.6.8^2)(4^3.6^2.8)$ named *uft*1[69]. This novel topology, *uft*1, was documented in the personal topology repository[70] and not observed in any materials even zeolites. Notably, the topology of *uft*1 exhibits the same topological network as PCOD 8095196, also known as 4,4,4T-3872HZ, which is among hypothetical zeolites proposed by Deem *et al*.[71,72] Interestingly, the topology of UZIF-31 is

exactly matched with one of Tier 1, T-50 (Fig. 4b and Supplementary Fig. 14). Despite a slight distortion of Im, the difference in cell volume between UZIF-31 and the predicted structure was merely +6.14 %. During topological analysis of *uft*1, three types of tiles compose UZIF-31 (Fig. 5a). One of the tiles is a [4^6] cage, which is known as the *t-cub*, and the others ([$4^4.6^6.8^2$] and [$4^6.6^2.8^2.12^2$] cages) are not observed in any zeolites. The largest pore opening with a 12-membered ring (MR) was observed within the surrounding [$4^6.6^2.8^2.12^2$] cage.

UZIF-32 and -33 (*uft*2) crystallized within the orthorhombic *Cmmm* (No. 65) space group. The composition of crystals was confirmed by SCXRD and ^{1}H-NMR (Supplementary Tables 7, 8 and Supplementary Figs 25, 26). Thus far, due to the identical net, the topological analysis of UZIF-32 has been considered the same as that of UZIF-33 for clarity. UZIF-33 also presented a novel 4,4,4-c topology with the point symbol of ($4.6^4.8$)($4^2.6^3.8$)($4^3.6^2.8$) named *uft*2[69], not encountered in any other porous materials, including zeolite structures. Similar to *uft*1, the topology of *uft*2 shows the identical topological network with PCOD 8096930, also referred to as 4,4,4T-3884HZ, in hypothetical zeolites suggested by Deem *et al*.[71,72] The topology of UZIF-33 is completely matched to one of Tier 1, T-12, with a slight cell volume difference of +4.68 % (Fig. 4c and Supplementary Fig. 15). Eight types of tiles ([4^6] cage as *t-cub*, [$4^2.6^2$] cage as *t-lov*, [6^3] cage as *t-kah*, [$4^2.6^4$] cage as *t-lau*, [$6^2.8^2$] cage as *t-kaa*, [$4^2.8^4$] cage as *t-ste*, [$4^4.6^4.12^2$] cage as *t-sty*, [$8^4.12^2$] cage) compose UZIF-33 (Fig. 5b). To the best of our knowledge, UZIF-33 has the highest number of tiles (8 types) within known ZIF topologies. Among these tiles, [6^3] cage as *t-kah* and [$4^4.6^4.12^2$] cage as *t-sty* were first observed in ZIF structures, while [$8^4.12^2$] cage was novel and not even present in zeolites. 12-MRs as the largest pore opening were enclosed by alternating [$4^4.6^4.12^2$] cage (*t-sty*) and novel [$8^4.12^2$] cage.

The unique structural configuration of UZIF-31 and UZIF-33 emerges because of how the layers stack together, which are made up of different tiles. Specifically, UZIF-31 exhibits an A-A' stacking pattern (Fig. 5c), whereas UZIF-33 displays an A-B-A stacking pattern (Fig. 5d). Furthermore, there is a notable difference in the structural arrangement due to the ordering pattern of the *t-cub* ([4^6] cage) tile, illustrating the dissimilarity between UZIF-31 and UZIF-33 (Supplementary Figs 31 and 32). The framework density, denoted as *T/V* (the number of T sites per volume), for UZIF-31 (*T/V* = 2.33 nm^{-3}) and UZIF-33 (*T/V* = 2.42 nm^{-3}, identical to UZIF-32) is comparable to an iconic ZIF, ZIF-8 (*T/V* = 2.45 nm^{-3}).

The crystal structures of the new ZIFs confirm the link-link interactions attributed to hydrogen bonds. Notably, weak hydrogen bonds[68] between C(sp^2)−H as donors and N(C=N−C) as acceptors within 3.0 Å were monitored in coordinated Im linkers (Supplementary Fig. 33). In UZIF-31, an infinite hydrogen bond chain was evident along the [010] direction, contributing to a total of 60 hydrogen bonds in a unit cell (the number of hydrogen bonds per volume*; H/V* = 2.91 nm^{-3}) (Fig. 6a and Supplementary Fig. 34). The pair of hydrogen bonds between adjacent *t-cub* cages along the [010] direction was observed in UZIF-32 (*H/V* = 0.81 nm^{-3}) (Fig. 6b and Supplementary Fig. 35). In UZIF-33 (*H/V* = 1.82 nm^{-3}), a cluster of hydrogen bonds within a *t-lau* cage bolstered the crystal structure (Fig. 6c and Supplementary Fig. 36). The highest *H/V* value and infinite hydrogen bond chain among the three ZIFs were found in UZIF-31, influencing its chemical stability under varying aqueous conditions with pH levels (Supplementary Figs. 21-23).

Porosity and selective adsorption. The permanent porosity of new ZIFs, along with their microporosity attributed to well-defined 12-MR channel-type pores, was validated through N_2 sorption at 77 K (Fig. 6d). The Brunauer-Emmett-Teller (BET) surface areas were measured as 1291(2), 832(2), and 1371(6) $m^2 g^{-1}$ for UZIF-31, -32, and -33, respectively (Supplementary

Figs. 37-39). The pore size distribution indicated the presence of pores with diameters of 1.19 nm in UZIF-31, 1.27 nm in UZIF-32, and 1.36 nm in UZIF-33 (Supplementary Figs. 40 and 41). Notably, among the trio of new ZIFs, UZIF-33 stood out with the highest BET surface area of 1376 $m^2 g^{-1}$, surpassing all previously reported AzbIm-based ZIFs (Supplementary Table 15).

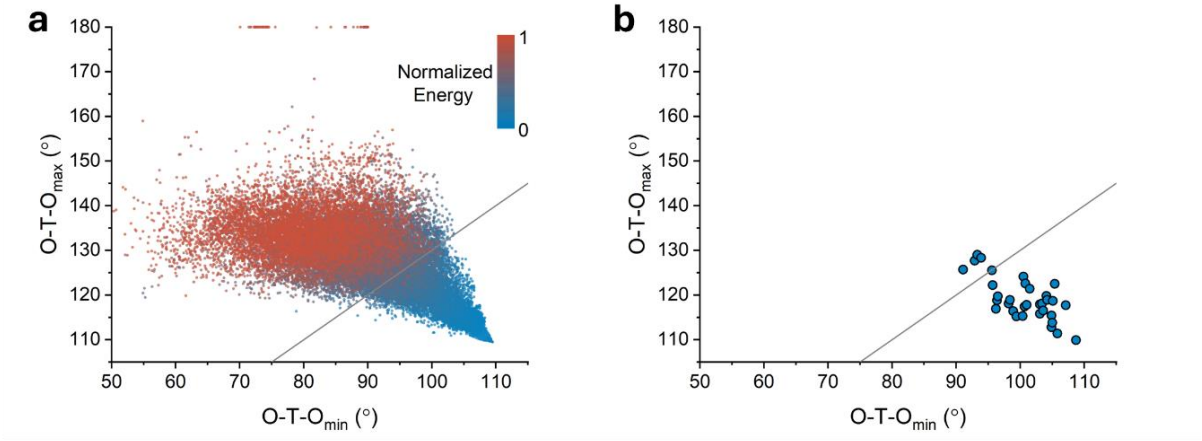
To explore the potential applications in separating harmful gases, we examined the adsorption behavior of CO_2 gas on ZIFs under ambient conditions. (Fig. 6e and Supplementary Figs. 43-45). Among these three ZIFs, UZIF-33 exhibited the highest capacity for CO_2 adsorption, followed by UZIF-31 and then UZIF-32. This sequence directly correlated with the number of exposed nitrogen atoms per unit volume, denoted as *N/V* (nm^{-3}): UZIF-32 (1.82 nm^{-3}), UZIF-31 (2.62 nm^{-3}), and UZIF-33 (3.63 nm^{-3}). The highest *N/V* value in UZIF-33 suggests a strong interaction between the frameworks and CO_2 gas molecules due to a more polar pore surface originating from a large number of uncoordinated nitrogen atoms. UZIF-33 also showed the highest isosteric heat of adsorption (Q_{st} = -35.1 kJ mol^{-1}) for CO_2 compared to the other two ZIFs (Fig. 6f). Further experiments were conducted to explore selective adsorption behaviors, particularly in separating CO_2 from CH_4, a vital industrial process for enhancing natural gas quality through carbon capture, contributing to a carbon-neutral society. As anticipated, UZIF-33 demonstrated significantly enhanced uptake capacities for CO_2 (1.73 mmol g^{-1}), C_2H_4 (1.71 mmol g^{-1}), and C_2H_6 (1.91 mmol g^{-1}), which were around five times higher than that for CH_4 (0.33 mmol g^{-1}) (Fig. 6g and Supplementary Fig. 46). Based on the ideal adsorbed solution theory (IAST)[73], the selectivity of equimolar (50/50) CO_2/CH_4, C_2H_4/CH_4, and C_2H_6/CH_4 mixtures at 298 K and 1 bar were determined as 11.2, 23.4, and 35.3, respectively (Fig. 6h). The outstanding separation efficiency of UZIF-33 in isolating C_2 gases and CO_2 from CH_4 was attributed to its highly polar pore environment. Intriguingly, the

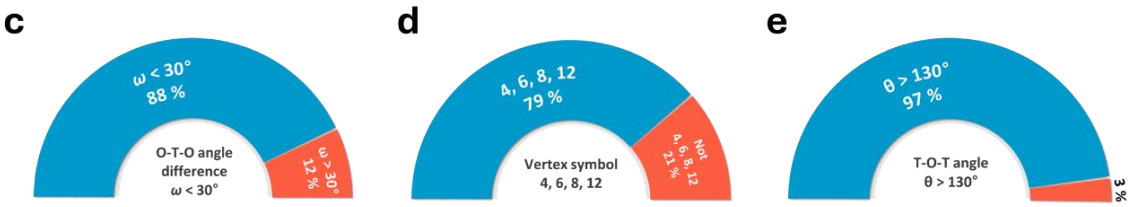
CO_2/CH_4 selectivity value of 11.9 is the highest among various reported ZIFs (Supplementary Table 22). To demonstrate the capturing CO_2 of UZIF-33 for practical applications, we conducted dynamic gas adsorption breakthrough experiments using CO_2/CH_4 gas mixtures at 298 K and 1 bar with a flow rate of 5 mL min^{-1} (See the details in Supplementary Information). The adsorption performance of UZIF-33 was assessed through breakthrough curve measurements conducted over three cycles. Notably, these breakthrough curves revealed no significant decline in adsorption capacity even after 120 sec g^{-1} of CO_2 retention time (Fig. 6i). Consequently, the dynamic breakthrough experiments demonstrated the remarkable ability of UZIF-33 to effectively separate CO_2 from CH_4.

Conclusions

In conclusion, we have developed an effective methodology for exploring new ZIFs by using a combination of structural descriptors, including O−T−O angle difference, vertex symbol, and T−O−T angle. Through this method, we have assessed the synthesizability of 420 trinodal ZIFs, establishing a collection of ZIFs with a high level of feasibility (90 ZIFs in Tier 1), from over 4 million zeolite structures. By introducing hydrogen bonds as link-link interactions between Im linkers, we successfully synthesized three novel ZIFs, UZIF-31 (*uft*1), UZIF-32 (*uft*2), and UZIF-33 (*uft*2), with novel topologies that classified within Tier 1 synthesizable ZIFs. Notably, UZIF-33 exhibited an exceptional ability to separate CO_2 from CH_4, making it a potential candidate for applications in carbon neutrality due to its highly polar pore surface. This data-driven discovery, which combines the digitization of chemical intuitions with targeted synthesis, has the potential to contribute significantly to the development of new functional materials and unravel challenges in the field of porous materials. In the forthcoming years, the

automation of synthesis processes[8] is expected to further accelerate the creation and evaluation of new ZIFs, merging computational predictions with real-world applications[74].





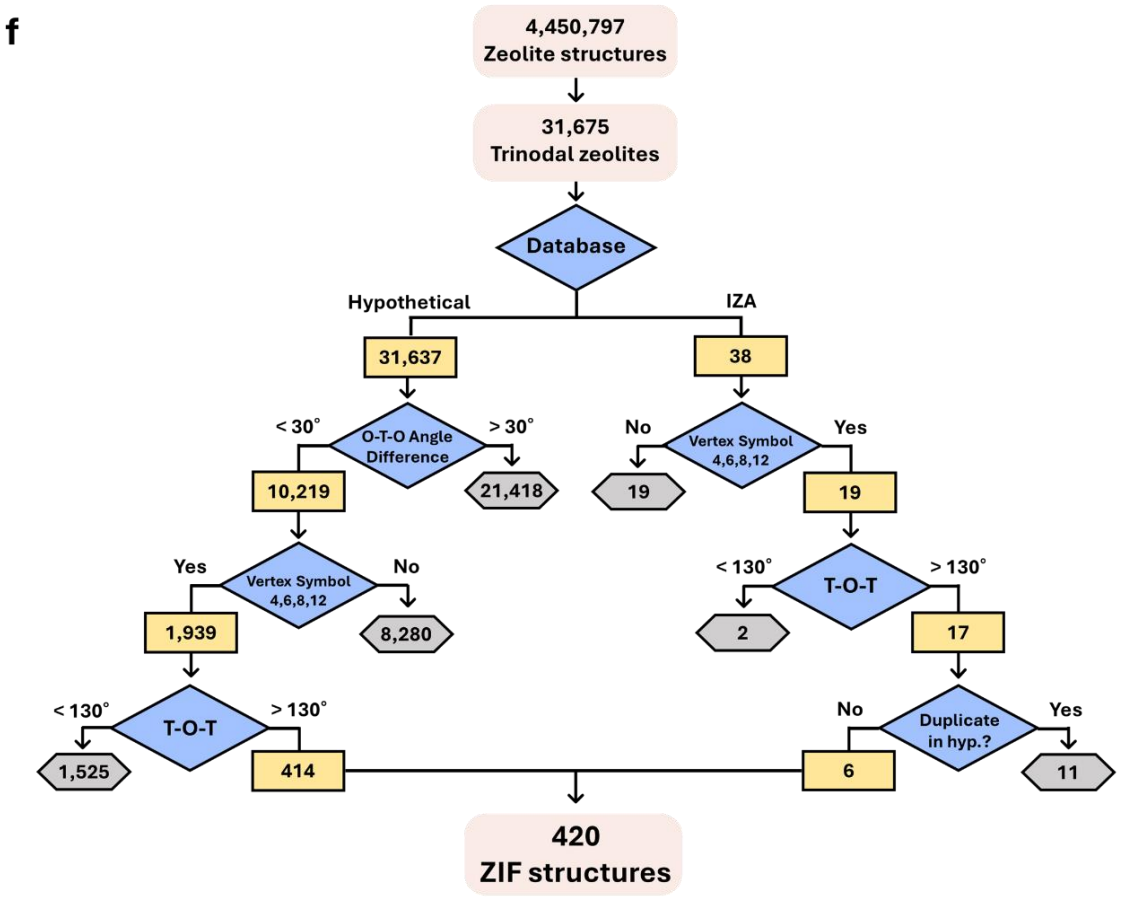


Fig. 1. Structural descriptors and flow chart for synthesizable ZIFs. a,**b** The O−T−O angle difference (ω) in (**a**) 31,637 trinodal hypothetical zeolites and (**b**) 33 known ZIFs. The grey line indicated ω = 30°; under the line (ω < 30°) and upper the line (ω > 30°). **c**,**d**,**e**, The percentage of known ZIFs with satisfying three structural descriptors. (**c**) First, O−T−O angle difference; ω < 30°. (**d**) Second, the vertex symbol; 4, 6, 8, and 12. (**e**) Third, T−O−T angle; > 130°. **f**, Flow chart for finding topology candidates for synthesizable ZIFs from zeolite database. After applying the descriptors, the number of zeolites reduced from an initial count of 4,450,797 to 420 ZIF topologies.

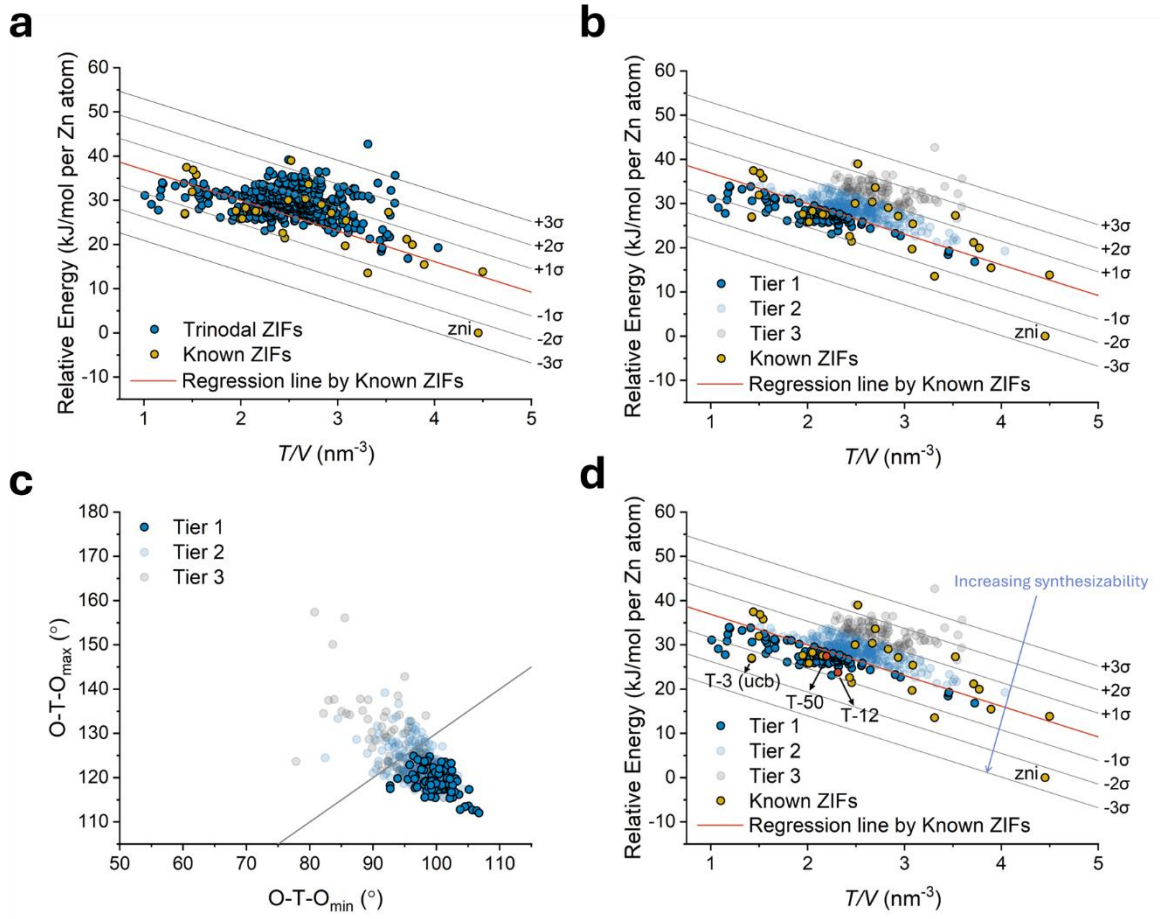


Fig. 2. Synthesizability of f-ZIFs. a, The calculated total energy of trinodal ZIFs with density. The red line represents the regression line set by known ZIF topologies for prediction intervals. **b**, Synthesizability tier of 420 f-ZIFs. Tier 1 (90/420; T-1 to T-90), Tier 2 (233/420; T-91 to T-323), and Tier 3 (97/420; T-324 to T-420) were divided in the section between the prediction intervals. **c**, The O−T−O angle difference (ω) in 420 f-ZIFs with synthesizable Tiers. 90 Tier 1 ZIFs are involved in $\omega < 30°$. **d**, Synthesizability ranking was set by the distance from -3σ line. Realized topologies in this study, T-12, and T-50, are denoted as red dots. T-3 is identical to the realized **ucb** topology.

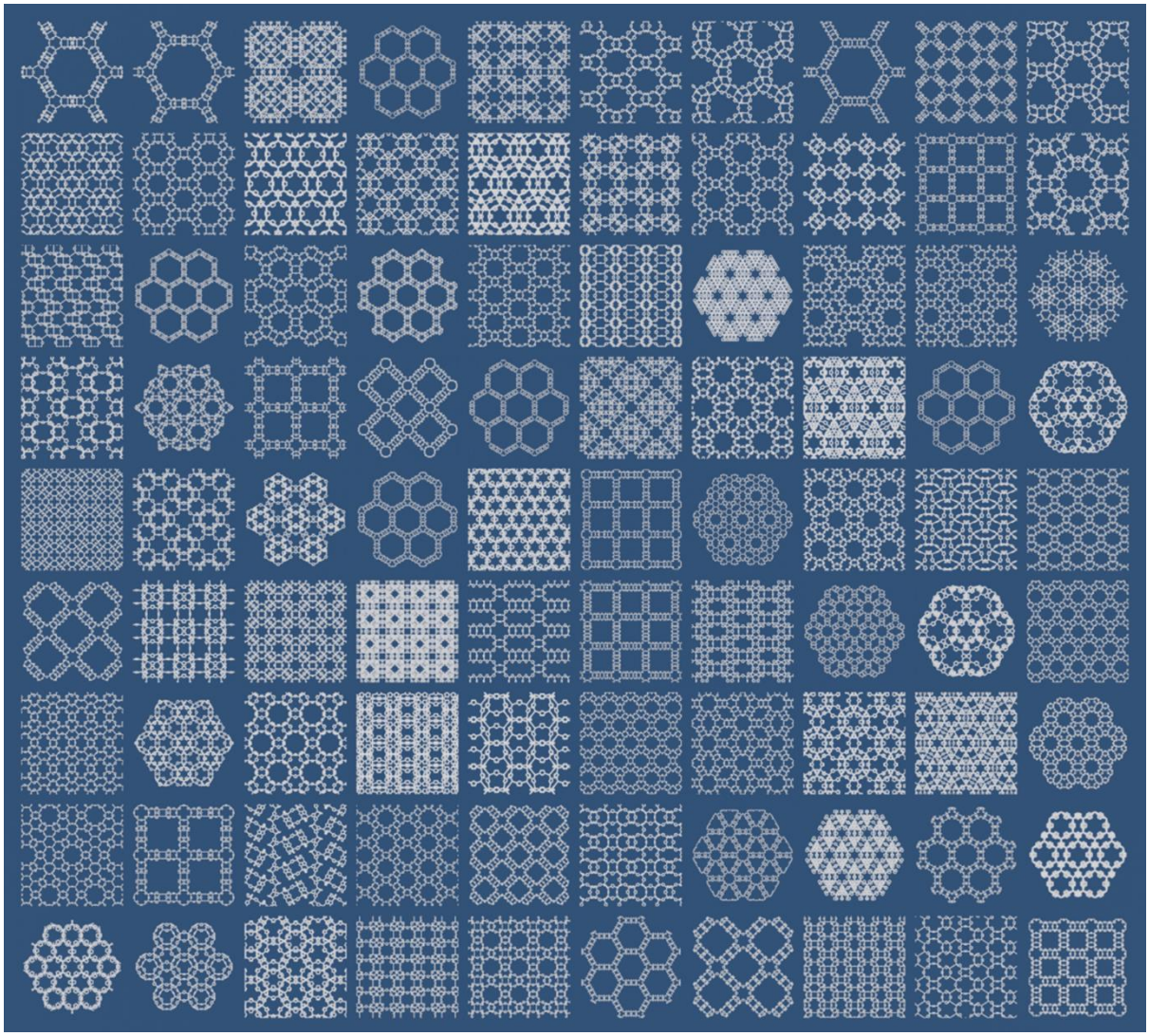


Fig. 3. Tiling projections of 90 f-ZIFs in Tier 1. Schematic illustrations with tiling projections of the generated 90 f-ZIFs (from T-1 to T-90) belong to Tier 1 synthesizability.

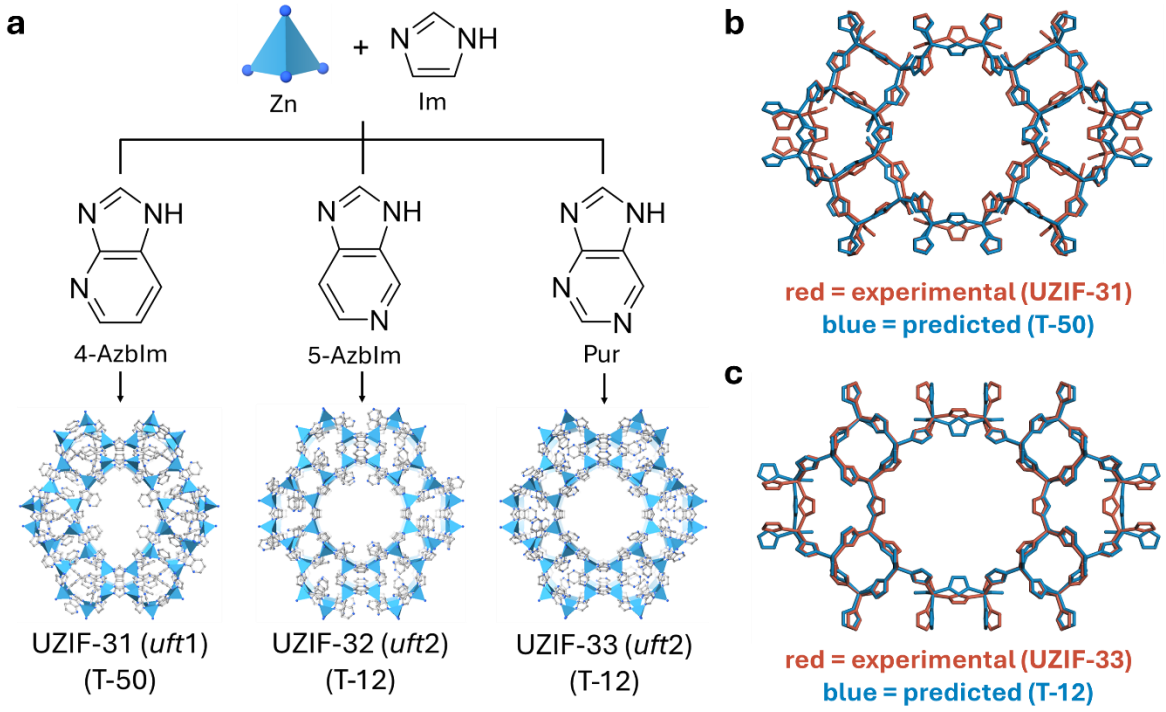
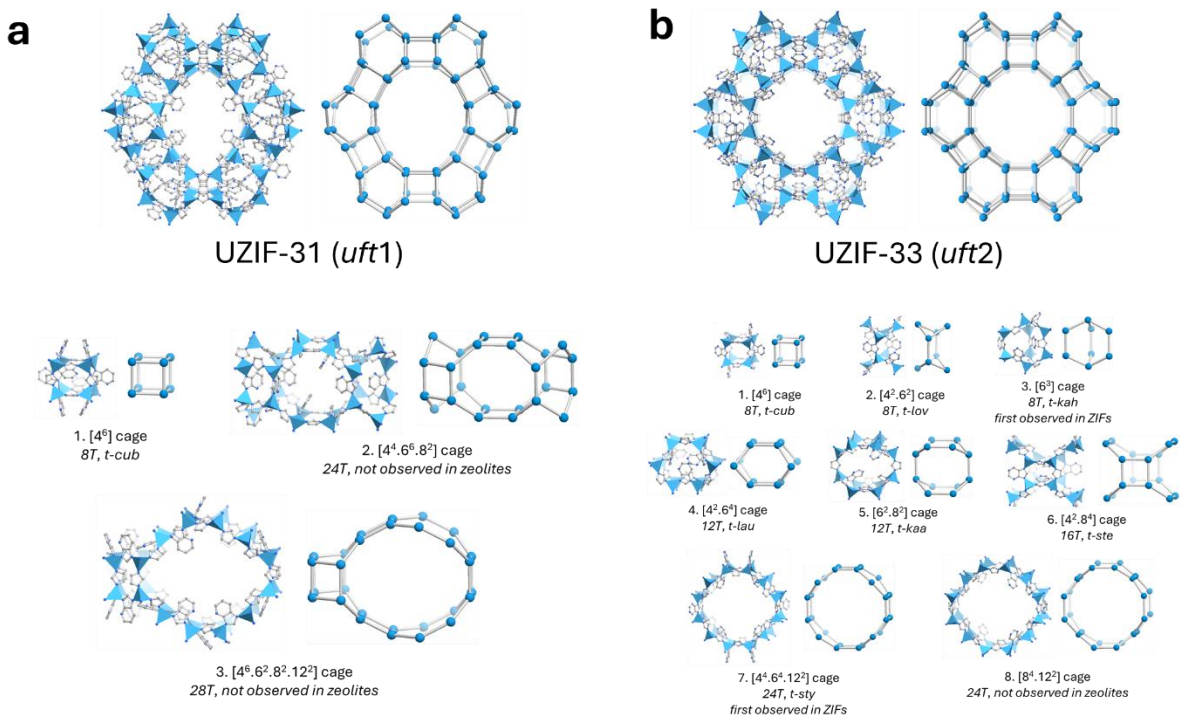
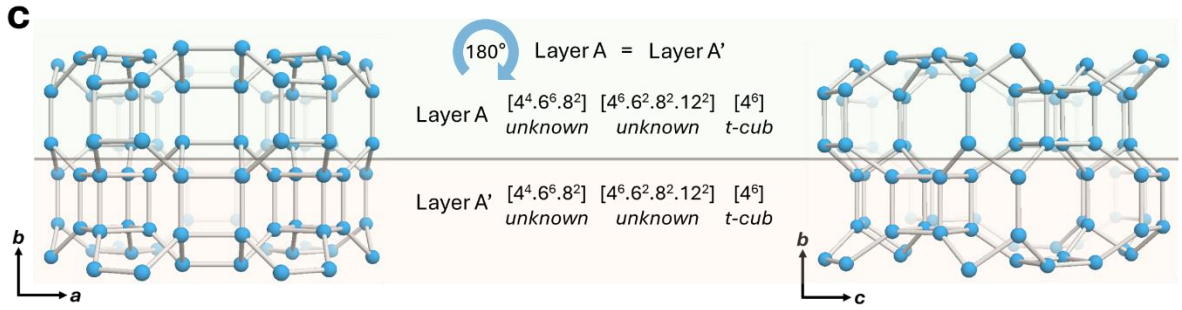


Fig. 4. Synthesis of new ZIFs. a, Synthetic scheme of UZIF-31 (*uft*1), UZIF-32 (*uft*2), and UZIF-33 (*uft*2) involving solvothermal reaction with Zn ion, Im, and AzbIm derivatives. The topologies for *uft*1 and *uft*2 are identical with T-50 and T-12 in Tier 1, respectively. **b**,**c**, Overlay of crystal structures between (**b**) synthesized UZIF-31 (red) versus predicted T-50 (blue) and (**c**) UZIF-33 (red) versus predicted T-12 (blue). Functional groups and hydrogens are omitted for clarity.





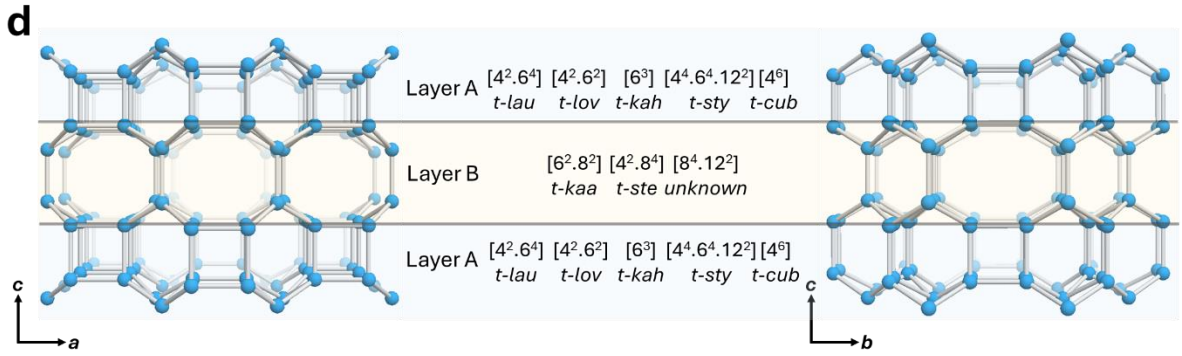


Fig. 5. Structural analysis of new ZIFs. a, Structural illustration of UZIF-31 (*uft*1) along to [010] direction and their constructing three types of cages, [4^6] cage; *t-cub*, [$4^4.6^6.8^2$] cage, and [$4^6.6^2.8^2.12^2$] cage. **b**, Structural illustration of UZIF-33 (*uft*2) along to [001] direction and their constructing eight types of cages, [4^6] cage; *t-cub*, [$4^2.6^2$] cage; *t-lov*, [6^3] cage; *t-kah*, [$4^2.6^4$] cage; *t-lau*, [$6^2.8^2$] cage; *t-kaa*, [$4^2.8^4$] cage; *t-ste*, [$4^4.6^4.12^2$] cage; *t-sty*, and [$8^4.12^2$] cage. Blue polyhedron and ball, Zn; grey, C; blue, N. **c**, Schematic representation of UZIF-31 along to [001] and [100] directions as the assembly to layer stacking pattern, A-A'. **d**, Schematic representation of UZIF-32 and -33 along to [010] and [100] directions as the assembly to layer stacking pattern, A-B-A.

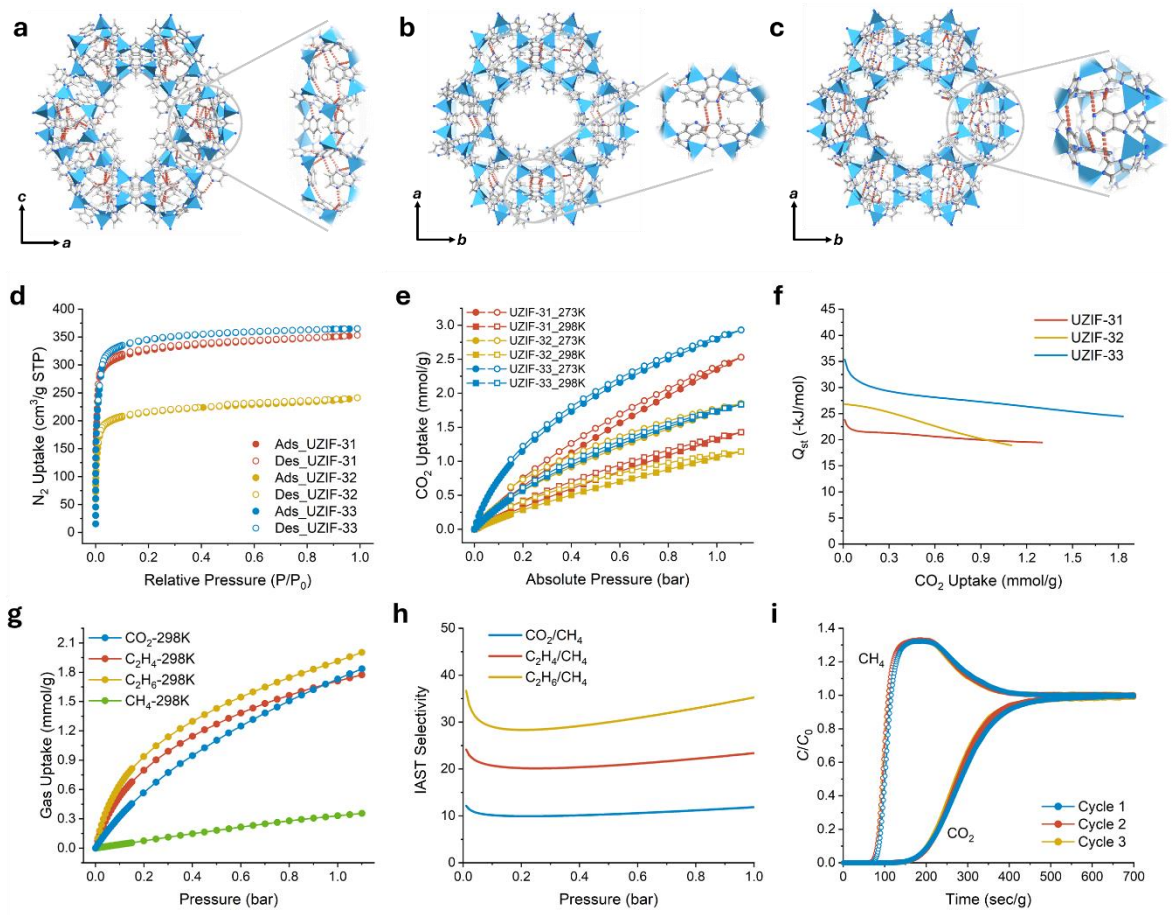


Fig. 6. Hydrogen bonds and gas sorption for ZIFs. a,**b**,**c** Hydrogen bonds (red dotted line; within 3.0 Å) in the crystal structure of (**a**) UZIF-31, (**b**) UZIF-32, and (**c**) UZIF-33. (**a**) The infinite chain of hydrogen bonds along [010] direction is highlighted. (**b**) The hydrogen bond between the adjacent *t-cub* cages is highlighted. (**c**) A cluster of hydrogen bonds in a *t-lau* cage is highlighted. Blue polyhedron, Zn; grey, C; blue, N; white, H. **d**, N_2 sorption isotherms for ZIFs measured at 77 K. **e**, CO_2 sorption isotherms for ZIFs measured at 273 K (circle symbol) and 298 K (square symbol). **f**, The Q_{st} of CO_2 for ZIFs. **g**, Experimental adsorption isotherms of CO_2, C_2H_4, C_2H_6, and CH_4 for UZIF-33 measured at 298 K. **h**, Predicted IAST selectivities of CO_2/CH_4, C_2H_4/CH_4, and C_2H_6/CH_4 for UZIF-33 with equimolar mixtures (50/50) at 298 K. **i**, Dynamic breakthrough curves for 50/50 mixtures of CO_2 and CH_4 at 298 K and 1 bar in a column packed with UZIF-33.

Methods

Computational details

Building ZIF structures. A total of 420 hypothetical ZIF structures were generated using ToBaCCo code[56] based on 420 parent zeolite topologies. The topological information for crystal construction comprises tetrahedral coordination symmetry, number of vertices and edges types, unit cell vectors, and fractional coordinates of vertices and edges. The unit cell vectors are parent zeolite unit cell vectors, and fractional coordinates of the vertices and edges are the coordinates of Si and O, respectively, in the parent zeolite. Zn as the metal node and Im as the organic linker were used as the building blocks.

Energy calculation of ZIFs. As generated, the hypothetical ZIFs are not in their minimum-energy structures (”geometries”). Along with the experimentally known ZIFs, they were subjected to molecular mechanics (MM) optimization as implemented in the LAMMPS package[58] with the force field MOF-FF whose parameters were fit to ZIF[57]. The optimization was done while simultaneously relaxing the unit cell parameters and the atomic coordinates within each unit cell, with the convergence criterion set for the 2-norm (length) of the global (all-atom) force vector being smaller than 10^{-4} kcal mol^{-1} $Å^{-1}$. The data is plotted as the MM energies per Zn atom (relative to that of the MM-optimized ZIF with the **zni** topology) and the number of Zn atoms per volume of the optimized cell[49].

Synthesis of ZIFs.

Synthesis of UZIF-31 (*uft*1). Zinc(II) nitrate hexahydrate (29.7 mg, 0.1 mmol), Im (30.0 mg, 0.44 mmol), and 4-AzbIm (13.2 mg, 0.11 mmol) were dissolved in 2.0 mL of DMF in a 5 mL vial. The vial was sealed and heated at 120 °C for 72 h. Transparent crystals were obtained and washed with DMF and acetone 3 times, respectively.

Synthesis of UZIF-32 (*uft*2). Zinc(II) nitrate hexahydrate (29.7 mg, 0.1 mmol), Im (30.0 mg, 0.44 mmol), and 5-AzbIm (12.0 mg, 0.1 mmol) were dissolved in 2.0 mL of DMF in a 5 mL vial. The vial was sealed and heated at 120 °C for 24 h. Clear solution of the mixture was stored at room temperature for 48 h. Transparent hexagonal prismatic crystals were obtained and washed with DMF and acetone 3 times, respectively.

Synthesis of UZIF-33 (*uft*2). Zinc(II) nitrate hexahydrate (29.7 mg, 0.1 mmol), Im (26.6 mg, 0.39 mmol), and Pur (13.2 mg, 0.11 mmol) were dissolved in 1.5 mL of DMF in a 5 mL vial. The vial was sealed and heated at 120 °C for 96 h. Transparent hexagonal prismatic crystals were obtained and washed with DMF and acetone 3 times, respectively.

Data availability

The data that support the findings of this study are available in the Supplementary Information. Crystallographic information files (T-*n*_zeolitecode.cif) of optimized 420 hypothetical ZIFs are available in the Supplementary Information. Crystallographic data for the structures reported in this article have been deposited at the Cambridge Crystallographic Data Centre under deposition numbers CCDC 2285519 (UZIF-31), 2285520 (UZIF-32), and 2285521 (UZIF-33). Copies of the data can be obtained free of charge via https://www.ccdc.cam.ac.uk/structures/.

Acknowledgments

This research work was financially supported by the National Research Foundation (NRF) of Korea (NRF-2020R1A2C3008226, NRF-2021M3I3A1084909, and RS-2023-00279793) and

Carbon Neutral Institute Research Fund (1.230069.01) of Ulsan National Institute of Science & Technology (UNIST). Computational resources at the Korea Institute of Science and Technology Information (KISTI) were used (KSC-2021-CRE-0386). We would like to acknowledge Prof. Davide M. Proserpio for his assistance in the topological analysis of the newly discovered ZIFs.

Author contributions

W.C. conceptualized and supervised the project. S.L. and W.C. led the project and interpreted the result. S.L. carried out the overall experiments and characterizations. H.J., Y.K., and E.C. helped in the synthesis of crystals. J.N. helped in the analysis of single-crystal diffraction data. D.C.Y. performed the energy calculation. S.J. and H.O. performed the breakthrough experiments. S.L. and W.C. wrote the original draft of the manuscript. All authors reviewed and edited the manuscript.

Competing interests

The authors declare no competing interests.

References

1. Rajan, K. Materials informatics: the materials "gene" and big data. *Annu. Rev. Mater. Res.* **45**, 153−169 (2015).

2. Agrawal, A. & Choudhary, A. Perspective: Materials informatics and big data: Realization of the "fourth paradigm" of science in materials science. *APL Mater.* **4**, 053208 (2016).

3. Butler, K. T., Davies, D. W., Cartwright, H., Isayev, O. & Walsh, A. Machine learning for molecular and materials science. *Nature* **559**, 547−555 (2018).

4. López, C. Artificial intelligence and advanced materials. *Adv. Mater.* **35**, 2208683 (2023).

5. Raccuglia, P. et al. Machine-learning-assisted materials discovery using failed experiments. *Nature* **533**, 73−76 (2016).

6. Bernstein, A. et al. Renewables need a grand-challenge strategy. *Nature* **538**, 30 (2016).

7. Liu, J. et al. Systems integration for global sustainability. *Science* **347**, 1258832 (2015).

8. Tabor, D. P. et al. Accelerating the discovery of materials for clean energy in the era of smart automation. *Nat. Rev. Mater.* **3**, 5−20 (2018).

9. Ongari, D., Talirz, L. & Smit, B. Too many materials and too many applications: an experimental problem waiting for a computational solution. *ACS Cent. Sci.* **6**, 1890−1900 (2020).

10. Mroz, A. M., Posligua, V., Tarzia, A., Wolpert, E. H. & Jelfs, K. E. Into the unknown: how computation can help explore uncharted material space. *J. Am. Chem. Soc.* **144**, 18730−18743 (2022).

11. Szczypiński, F. T., Bennett, S. & Jelfs, K. E. Can we predict materials that can be synthesised? *Chem. Sci.* **12**, 830−840 (2021).

12. Price, S. L. Why don't we find more polymorphs? *Acta Crystallogr. Sect. B* **69**, 313−328 (2013).

13. Oganov, A. R., Pickard, C. J., Zhu, Q. & Needs, R. J. Structure prediction drives materials discovery. *Nat. Rev. Mat.* **4**, 331−348 (2019).

14. Pulido, A. et al. Functional materials discovery using energy–structure–function maps. *Nature* **543**, 657−664 (2017).

15. Davies, I. W. The digitization of organic synthesis. *Nature* **570**, 175−181 (2019).

16. Peng, J. et al. Human- and machine-centred designs of molecules and materials for sustainability and decarbonization. *Nat. Rev. Mat.* **7**. 991−1009 (2022).

17. Moosavi, S. M. et al. Capturing chemical intuition in synthesis of metal-organic frameworks. *Nat. Commun.* **10**, 539 (2019).

18. Bustillo, L., Laino, T. & Rodrigues, T. The rise of automated curiosity-driven discoveries in chemistry. *Chem. Sci.* **14**, 10378−10384 (2023).

19. Bennett, S. et al. Materials precursor score: modeling chemists' intuition for the synthetic accessibility of porous organic cage precursors. *J. Chem. Inf. Model.* **61**, 4342−4356 (2021).

20. Lipinski, C. A., Lombardo, F., Dominy, B. W. & Feeney, P. J. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Delivery Rev.* **46**, 3−26 (2001).

21. Lombardino, J. G. & Lowe III, J. A. The role of the medicinal chemist in drug discovery – then and now. *Nat. Rev. Drug Discov.* **3**, 853−862 (2004).

22. Schneider, P. et al. Rethinking drug design in the artificial intelligence era. *Nat. Rev. Drug Discov.* **19**, 353−364 (2020).

23. Stein, H. S. & Gregoire, J. M. Progress and prospects for accelerating materials science with automated and autonomous workflows. *Chem. Sci.* **10**, 9640−9649 (2019).

24. Jablonka, K. M., Ongari, D., Moosavi, S. M. & Smit, B. Big-data science in porous materials: materials genomics and machine learning. *Chem. Rev.* **120**, 8066−8129 (2020).

25. Tshitoyan, V. et al. Unsupervised word embeddings capture latent knowledge from materials science literature. *Nature* **571**, 95−98 (2019).

26. Sun, W. et al. The thermodynamic scale of inorganic crystalline metastability. *Sci. Adv.* **2**, e1600225 (2016).

27. Ouyang, B. et al. Synthetic accessibility and stability rules of NASICONs. *Nat. Commun.* **12**, 5752 (2021).

28. Park, K. S. et al. Exceptional chemical and thermal stability of zeolitic imidazolate frameworks. *Proc. Natl Acad. Sci. USA* **103**, 10186−10191 (2006).

29. Huang, X.-C., Lin, Y.-Y., Zhang, J.-P. & Chen, X.-M. Ligand-directed strategy for zeolite-type metal–organic frameworks: zinc(II) imidazolates with unusual zeolitic topologies. *Angew. Chem. Int. Ed.* **45**, 1557−1559 (2006).

30. Banerjee, R. et al. High-throughput synthesis of zeolitic imidazolate frameworks and application to CO_2 capture. *Science* **319**, 939−943 (2008).

31. Noh, K., Lee, J. & Kim, J. Compositions and structures of zeolitic imidazolate frameworks. *Isr. J. Chem.* **58**, 1075−1088 (2018).

32. Baerlocher, C. & and McCusker, L. Database of Zeolite Structures. http://www.iza-structure.org/databases/ (accessed: June 2022).

33. Kwon, H. T. & Jeong, H.-K. In situ synthesis of thin zeolitic–imidazolate framework ZIF-8 membranes exhibiting exceptionally high propylene/propane separation. *J. Am. Chem. Soc.* **135**, 10763−10768 (2013).

34. Ma, X. et al. Zeolitic imidazolate framework membranes made by ligand-induced permselectivation. *Science* **361**, 1008−1011 (2018).

35. Knebel, A. et al. Solution processable metal–organic frameworks for mixed matrix membranes using porous liquids. *Nat. Mater.* **19**, 1346−1353 (2020).

36. Wang, S., Yao, W., Lin, J., Ding, Z. & Wang, X. Cobalt imidazolate metal–organic frameworks photosplit CO_2 under mild reaction conditions. *Angew. Chem. Int. Ed.* **53**, 1034−1038 (2014).

37. Tran, U. P., Le, K. K., & Phan, N. T. Expanding applications of metal−organic frameworks: zeolite imidazolate framework ZIF-8 as an efficient heterogeneous catalyst for the Knoevenagel reaction. *ACS Catal.* **1**, 120−127 (2011).

38. Zhou, K. et al. Characterization and properties of Zn/Co zeolitic imidazolate frameworks vs. ZIF-8 and ZIF-67. *J. Mater. Chem. A* **5**, 952−957 (2017).

39. Arafat, Y. et al. Advances in zeolite imidazolate frameworks (ZIFs) derived bifunctional oxygen electrocatalysts and their application in zinc–air batteries. *Adv. Energy Mater.* **11**, 2100514 (2021).

40. Liu, X. et al. Zeolitic imidazolate frameworks as Zn^{2+} modulation layers to enable dendrite-free Zn anodes. *Adv. Sci.* **7**, 2002173 (2020).

41. Xia, B. Y. et al. A metal–organic framework-derived bifunctional oxygen electrocatalyst. *Nat. Energy* **1**, 15006 (2016).

42. Atlas of Prospective Zeolite Structures, http://www.hypotheticalzeolites.net/ (accessed: November 2019).

43. Blatov, V. A., Ilyushin, G. D. & Proserpio, D. M. The zeolite conundrum: why are there so many hypothetical zeolites and so few observed? a possible answer from the zeolite-type frameworks perceived as packings of tiles. *Chem. Mater.* **25**, 412−424 (2013).

44. Moghadam, P. Z. et al. Targeted classification of metal–organic frameworks in the Cambridge structural database (CSD). *Chem. Sci.* **11**, 8373−8387 (2020).

45. CCDC. The Cambridge Structural Database (CSD) MOF collection. https://www.ccdc.cam.ac.uk/Community/csd-community/csd-mof-collection/.

46. Freund, R. et al. 25 Years of reticular chemistry. *Angew. Chem. Int. Ed.* **60**, 23946−23974 (2018).

47. Yang, J. et al. Principles of designing extra-large pore openings and cages in zeolitic imidazolate frameworks. *J. Am. Chem. Soc.* **139**, 6448−6455 (2017).

48. Hayashi, H., Côté, A. P., Furukawa, H., O'Keeffe, M. & Yaghi, O. M. Zeolite A imidazolate frameworks. *Nat. Mater.* **6**, 501−506 (2007).

49. Lee, S., Nam, D., Yang, D. C. & Choe, W. Unveiling hidden zeolitic imidazolate frameworks guided by intuition-based geometrical factors. *Small* **19**, 2300036 (2023).

50. Deem, M. W., Pophale, R., Cheeseman, P. A. & Earl, D. J. Computational discovery of new zeolite-like materials. *J. Phys. Chem. C* **113**, 21353−21360 (2009).

51. Sastre, G. & Corma, A. Predicting structural feasibility of silica and germania zeolites. *J. Phys. Chem. C* **114**, 1667−1673 (2010).

52. Wragg, D. S., Morris, R. E. & Burton, A. W. Pure silica zeolite-type frameworks: a structural analysis. *Chem. Mater.* **20**, 1561−1570 (2008).

53. Belmabkhout, Y. et al. Natural gas upgrading using a fluorinated MOF with tuned H_2S and CO_2 adsorption selectivity. *Nat. Energy* **3**, 1059−1066 (2018).

54. Hepburn, C. et al. The technological and economic prospects for CO_2 utilization and removal. *Nature* **575**, 87−97 (2019).

55. O'Keeffe, M., Peskov, M. A., Ramsden, S. J. & Yaghi, O. M. The reticular chemistry structure resource (RCSR) database of, and symbols for, crystal nets. *Acc. Chem. Res.* **41**, 1782−1789 (2008).

56. Colón, Y.J., Gómez-Gualdrón, D. A. & Snurr, R.Q. Topologically guided, automated construction of metal–organic frameworks and their evaluation for energy-related applications. *Cryst. Growth Des.* **17**, 5801−5810 (2017).

57. Dürholt, J. P., Fraux, G., Coudert, F.-X. & Schmid, R. Ab initio derived force fields for zeolitic imidazolate frameworks: MOF-FF for ZIFs. *J. Chem. Theory Comput.* **15**, 2420−2432 (2019).

58. Thompson, A. P. et al. LAMMPS - a flexible simulation tool for particle-based materials modeling at the atomic, meso, and continuum scales. *Comput. Phys. Commun.* **271**, 108171 (2022).

59. Anderson, R. & Gómez-Gualdrón, D. A. Large-scale free energy calculations on a computational metal−organic frameworks database: toward synthetic likelihood predictions. *Chem. Mater.* **32**, 8106−8119 (2020).

60. Park, J., Lim, Y., Lee, S. & Kim, J. Computational design of metal−organic frameworks with unprecedented high hydrogen working capacity and high synthesizability. *Chem. Mater.* **35**, 9−16 (2023).

61. Lewis, D. W. et al. Zeolitic imidazole frameworks: structural and energetics trends compared with their zeolite analogues. *CrystEngComm* **11**, 2272−2276 (2009).

62. Baburin, I. A., Leoni, S. & Seifert, G. Enumeration of not-yet-synthesized zeolitic zinc imidazolate MOF networks: a topological and DFT approach. *J. Phys. Chem. B* **112**, 9437−9443 (2008).

63. Jones, J. T. A. et al. Modular and predictable assembly of porous organic molecular crystals. *Nature* **474**, 367−371 (2011).

64. Rodriguez-Albelo, L. M. et al. Zeolitic polyoxometalate-based metal-organic frameworks (Z-POMOFs): computational evaluation of hypothetical polymorphs and the successful

targeted synthesis of the redox-active Z-POMOF1. *J. Am. Chem. Soc.* **131**, 16078−16087 (2009).

65. Zhu, Q. et al. Analogy powered by prediction and structural invariants: computationally led discovery of a mesoporous hydrogen-bonded organic cage crystal. *J. Am. Chem. Soc.* **144**, 9893−9901 (2022).

66. Shields, C. E. et al. Experimental confirmation of a predicted porous hydrogen-bonded organic framework. *Angew. Chem. Int. Ed.* **62**, e202303167 (2023).

67. Xu, Y. et al. Experimentally validated ab Initio crystal structure prediction of novel metal−organic framework materials. *J. Am. Chem. Soc.* **145**, 3515−3525 (2023).

68. Steiner, T. The hydrogen bond in the solid state. *Angew. Chem. Int. Ed.* **41**, 48−76 (2002).

69. Blatov, V. A., Shevchenko, A. P. & Proserpio, D. M. Applied topological analysis of crystal structures with the program package ToposPro. *Cryst. Growth Des.* **14**, 3576−3586 (2014).

70. Shevchenko, A. P., Shabalin, A. A., Karpukhin, I., Yu & Blatov, V. A. Topological representations of crystal structures: generation, analysis and implementation in the TopCryst system. *Sci. Technol. Adv. Mater.* **2**, 250−265 (2022). https://topcryst.com/.

71. Pophale, R., Cheeseman, P. A. & Deem, M. W. A database of new zeolite-like materials. *Phys. Chem. Chem. Phys.* **13**, 12407−12412 (2011). https://mwdeem.org/PCOD/; https://zenodo.org/records/8035241.

72. Blatov, V. A., Blatova, O. A., Daeyaert, F. & Deem, M. W. Nanoporous materials with predicted zeolite topologies. *RSC Adv.* **10**, 17760−17767 (2020).

73. Myers, A. L. & Prausnitz, J. M. Thermodynamics of mixed-gas adsorption. *AIChE J.* **11**, 121−127 (1965).

74. Lyu, H., Ji, Z., Wuttke, S. & Yaghi, O. M. Digital reticular chemistry. *Chem* **6**, 2219−2241 (2020).