JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY
pubs.acs.org/JACS
Article

Shaping the Water-Harvesting Behavior of Metal–Organic Frameworks Aided by Fine-Tuned GPT Models

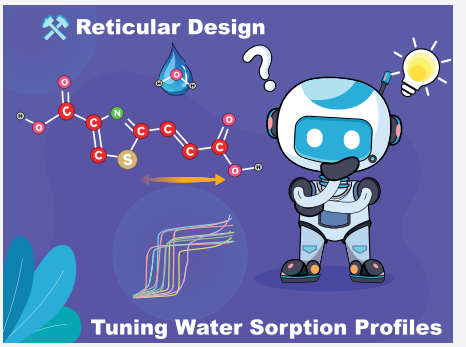
Zhiling Zheng, Ali H. Alawadhi, Saumil Chheda, S. Ephraim Neumann, Nakul Rampal, Shengchao Liu, Ha L. Nguyen, Yen-hsu Lin, Zichao Rong, J. Ilja Siepmann, Laura Gagliardi, Anima Anandkumar, Christian Borgs, Jennifer T. Chayes, and Omar M. Yaghi*

Cite This: https://doi.org/10.1021/jacs.3c12086

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: We construct a data set of metal–organic framework (MOF) linkers and employ a fine-tuned GPT assistant to propose MOF linker designs by mutating and modifying the existing linker structures. This strategy allows the GPT model to learn the intricate language of chemistry in molecular representations, thereby achieving an enhanced accuracy in generating linker structures compared with its base models. Aiming to highlight the significance of linker design strategies in advancing the discovery of water-harvesting MOFs, we conducted a systematic MOF variant expansion upon state-of-the-art MOF-303 utilizing a multidimensional approach that integrates linker extension with multivariate tuning strategies. We synthesized a series of isoreticular aluminum MOFs, termed Long-Arm MOFs (LAMOF-1 to LAMOF-10), featuring linkers that bear various combinations of heteroatoms in their five-membered ring moiety, replacing pyrazole with either thiophene, furan, or thiazole rings or a combination of two. Beyond their consistent and robust architecture, as demonstrated by permanent porosity and thermal stability, the LAMOF series offers a generalizable synthesis strategy. Importantly, these 10 LAMOFs establish new benchmarks for water uptake (up to 0.64 g g^{-1}) and operational humidity ranges (between 13 and 53%), thereby expanding the diversity of water-harvesting MOFs.



Downloaded via Zhiling Zheng on December 13, 2023 at 17:08:07 (UTC).
See https://pubs.acs.org/sharingguidelines for options on how to legitimately share published articles.

INTRODUCTION

As the 21st century unfolds, humanity faces an unprecedented water crisis exacerbated by climate change and population growth.[1−5] Freshwater reserves are dwindling, and existing technologies for water generation are either energy-intensive or geographically limited.[6,7] We believe that water harvesting from air provides access to an inexhaustible source of water.[8−11] However, the challenge lies in developing materials that can efficiently capture this water, particularly in arid regions where the need is most urgent.[12−15]

Metal–organic frameworks (MOFs) have shown great promise as sorbents for atmospheric water harvesting.[16−24] While earlier research has produced MOFs with success in field tests to produce potable water under desert environments,[25−27] the task of designing new and better performing water-harvesting MOFs is complex and labor-intensive.[28−32] In addition to limited guiding principles on searching for candidate linkers,[33−35] the permutations for linker structural variations are virtually infinite, even when starting with a known effective linker.[36,37] This complexity transforms the manual search for new MOF variants as candidate water-harvesting sorbents into a daunting and time-consuming endeavor.

In this report, we present a multipronged approach to revolutionize the reticular design and synthesis of water-harvesting MOFs (Figure 1). First, we report a family of 10 MOFs known as LAMOFs, synthesized through a synergistic integration of linker extension[16,19,38−41] and multivariate (MTV) tuning strategies.[42−46] Second, we discuss how artificial intelligence (AI) can significantly accelerate the pace of research by efficiently proposing potential linker variants with minimum hallucinations, thereby reducing the number of labor-intensive tasks. In this work, we build a MOF linker mutation database, which consists of 3943 molecular editing examples, and use it to train large language models (LLMs) like GPT-3.5 to propose new MOF linker structures with high accuracy through a simple fine-tuning method. Third, we showcase the power of MOF linker editing, which allows us to shape the water-harvesting behaviors of the resulting MOFs. Specifically, through analysis, the obtained 10 different LAMOFs demonstrate distinct gas sorption and water sorption behavior due to just small changes in one or two atoms on the heterocyclic ring. Finally, we offer a combined experiment-computation analysis of their gas and water sorption behavior, showcasing significantly enhanced

Received: October 30, 2023
Revised: November 27, 2023
Accepted: November 28, 2023

© XXXX American Chemical Society

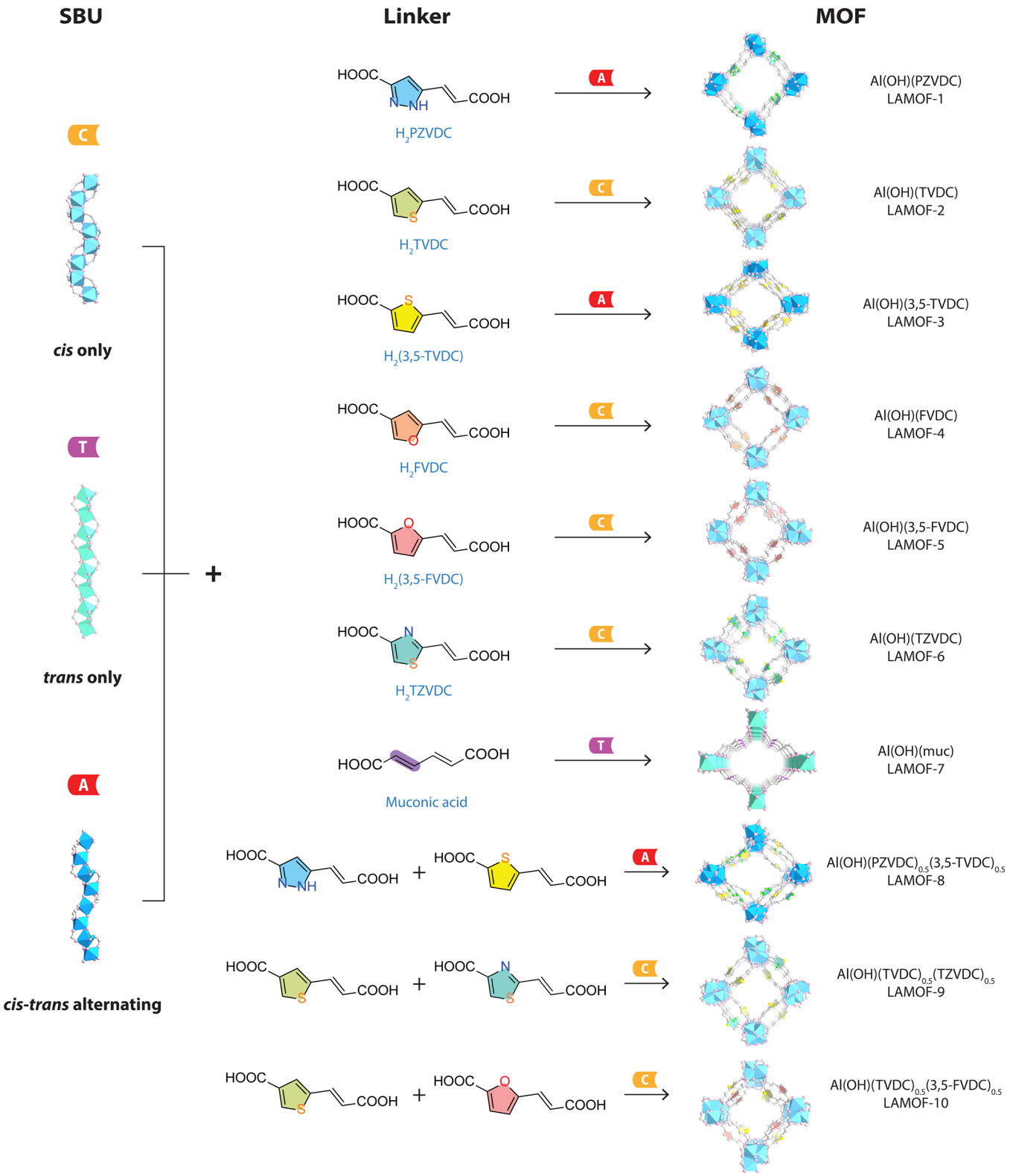


Figure 1. Structural diversity of long-armed metal−organic frameworks (LAMOFs) for atmospheric water harvesting. Displayed are various long-armed organic linkers (middle) and the corresponding LAMOFs (right), derived either from an individual or combination of two linkers. The depicted LAMOFs all exhibit infinite rod secondary building units composed of *cis*-connected, *trans*-connected, or alternating *cis*−*trans* corner-sharing AlO_6 octahedra (left). For illustrative clarity: Al is represented by light blue, cyan, or deep blue octahedra, each color corresponding to a distinct SBU; O, pink; C, light gray; N, green; and S, yellow. H atoms are omitted for clarity.

water uptake and reduced heats of adsorption compared to those of traditional water-harvesting MOFs. These newly developed LAMOFs set new benchmarks for water uptake and operational humidity ranges, thereby broadening the scope of real-world applications. Taken together, these advances aim to significantly

https://doi.org/10.1021/jacs.3c12086
J. Am. Chem. Soc. XXXX, XXX, XXX−XXX

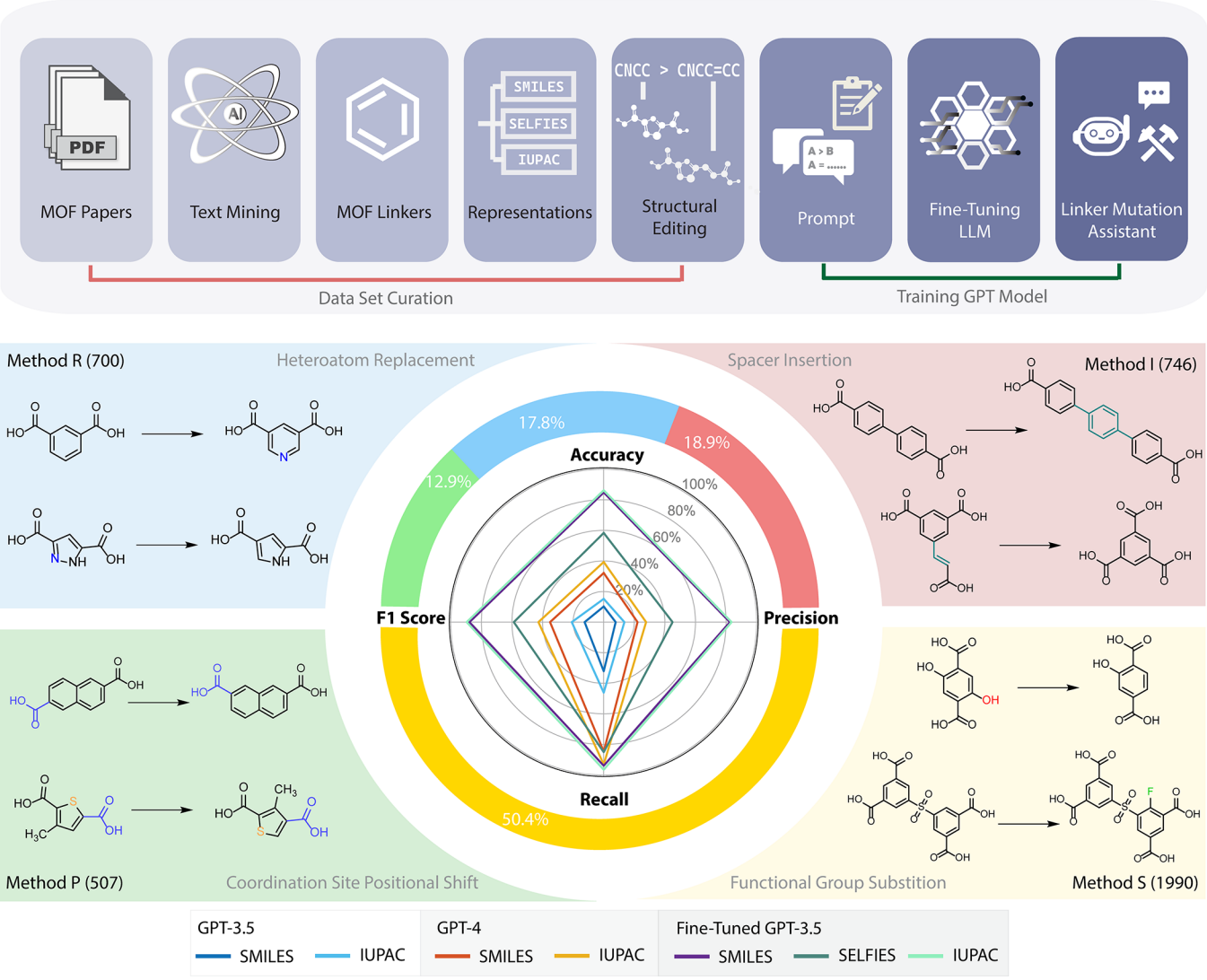


Figure 2. Schematic illustration of data set preparation for fine-tuning GPT-3.5, along with the evaluation metrics for the fine-tuned models. Details regarding each step of data set preparation and the model fine-tuning are given in the Supporting Information, Section S2. The central spider chart compares the performance of the fine-tuned GPT models (purple, deep green, and light green) with the base model *gpt-3.5-turbo* (deep blue and light blue) and the more advanced, yet not fine-tuned, model *gpt-4* (red and yellow). It illustrates improvements in all metrics, including accuracy, precision, recall, and F1 score, highlighting enhancements in editing the SMILES strings and IUPAC names. The surrounding pie chart illustrates the percentage representation of each specific type of mutation method. In total, there are up to 3943 linker-to-linker transformations from the four major categories, with the number of each type given in parentheses. For illustrative clarity, two examples are provided for each type of mutation method. Note that each method can encompass two to five different actions, and the provided examples represent just two of them.

improve the field of water-harvesting MOFs, heralding a new era of rapid and inclusive innovation in sustainable water harvesting.

RESULTS AND DISCUSSION

GPT-Assisted MOF Linker Design. We initiated our study by demonstrating the capability of fine-tuning GPT models for MOF linker editing by using designated mutation methods. The process is presented in a conversation format, where users input MOF linker structures in text-based chemical representations like SMILES codes,[47] SELFIES strings,[48] or IUPAC names.[49] The AI agent then responds with the corresponding mutated structures in the same chemical representation (Supporting Information, Section S2.1). Notably, this iterative procedure can be executed repeatedly, generating an array of edited structures, each branching out based on the original structure provided.

To train the LLM on molecular structure editing principles, we constructed a MOF linker data set encompassing 3943 examples of input linker structures and their respective mutated outputs (Figure 2 and Supporting Information, Section S2.2). We categorized mutations into four methods (Table S1): Substitution (S), where different functional groups are introduced or moved on the ring; Insertion (I), involving the choice of different spacers, such as a phenyl ring, vinyl group, or azo group; Replacement (R), where different heteroatoms like sulfur, nitrogen, and oxygen are substituted on the ring; and Positioning (P), implying different coordination sites such as carboxylates or nitrogen donors to positional shift within the ring. Each category encompasses two to five specific mutation actions. Our journey began with over 200 linkers, extracted from the literature using a ChatGPT-based assistant we previously reported.[50] From this, we concentrated on 142 carboxylate

https://doi.org/10.1021/jacs.3c12086
J. Am. Chem. Soc. XXXX, XXX, XXX−XXX

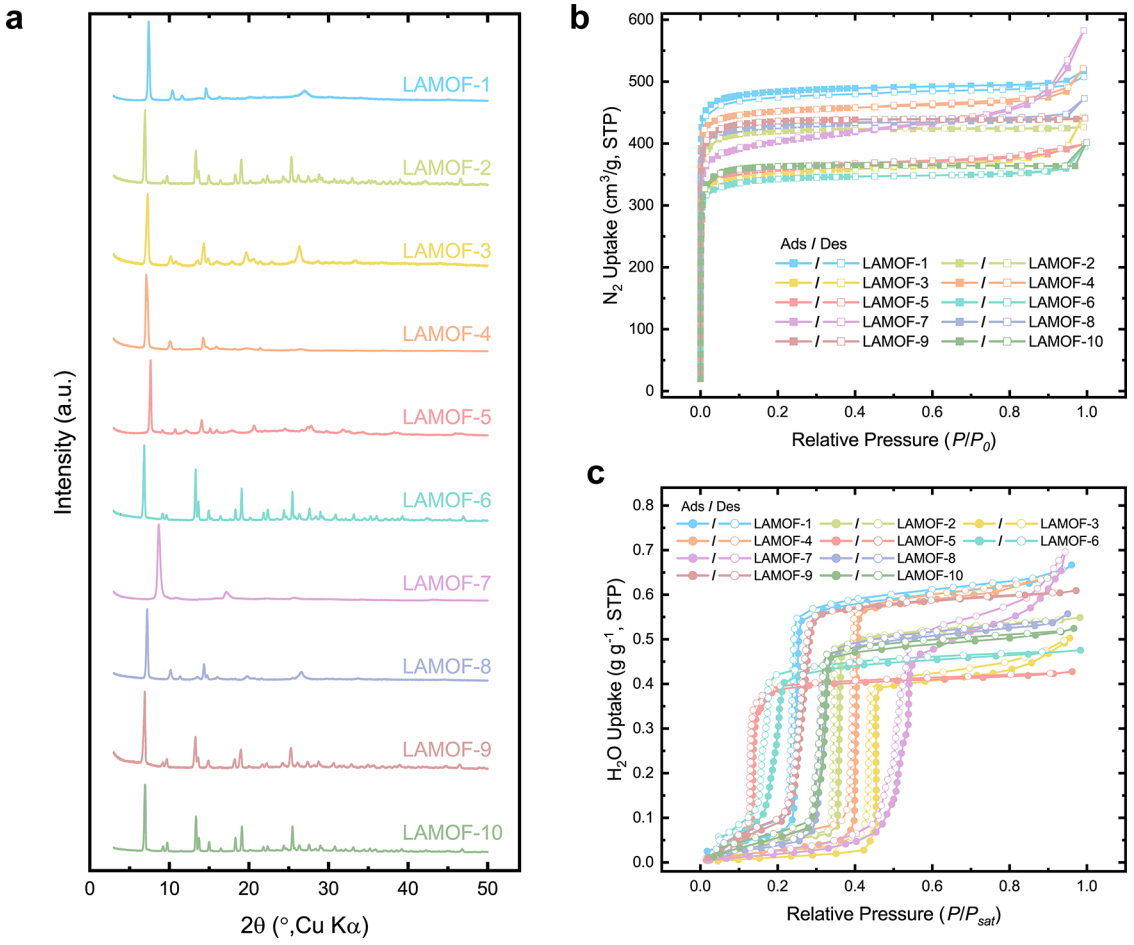


Figure 3. Characterization and sorption measurement of LAMOFs for shaping water-harvesting behavior. (a) PXRD analysis using Cu K_α radiation. (b) Nitrogen sorption isotherms at 77 K. (c) Water sorption isotherms at 25 °C. Filled circles depict adsorption; open circles indicate desorption.

linkers, which are more commonly seen in MOF synthesis. Subsequently, human volunteers manually went over each input linker structure and drew the outcomes by applying all 14 mutation actions across 4 mutation categories, resulting in 3943 outcomes. Note that one input linker can yield zero, one, or multiple outcomes depending on the mutation action applied (Table S1 and Figure S1).

A crucial consideration in AI-assisted generation of chemical molecules is the necessity for the output to be both syntactically and semantically coherent, conforming precisely to the assigned editing instructions (Supporting Information, Section S2). Simply put, output representations should be valid chemical structures, respecting all rules of valency and bonding, and should accurately reflect any assigned mutations or modifications from the original structure. While LLMs like GPT-3.5, GPT-4, and their chatbot adaptations, known as "ChatGPT", offer the advantage of interpreting human instructions in a conversational format, which makes it simpler to convey abstract mutations and modifications, the early performance evaluations of these models have shown their limitations.[51−55] Despite demonstrating certain levels of understanding of the underlying syntax and chemistry, these models sometimes suffer from "hallucinations" in their generated SMILES strings, which appear correct in formatting but are either chemically invalid or slightly misaligned when closely examined.

To address this challenge, we utilized our constructed data set to fine-tune the *gpt-3.5-turbo* model, the base model for ChatGPT, exposing it to thousands of examples of valid molecular edits between original and mutated chemical representations (like SMILES and SELFIES strings and IUPAC names). In this study, this process can be completed through the application programming interface (API) provided by OpenAI (Supporting Information, Section S2.2). The objective is to provide various examples to teach the model the intricacies behind such edits. In this case, this approach is more powerful than the zero-shot or few-shot prompt engineering strategies,[54,56−59] as the MOF linker mutation methods and actions are diverse, and the principles behind molecular editing encompass not only chemical rules like bonding and valency but also the construction of syntactically correct text-based chemical representations (Figures S3−S5). Fine-tuning through API, in comparison to in-context learning in prompt engineering, is not constrained by token limits, allowing the training of models on a substantial volume of examples and case studies (Table S2 and Figure S6). This approach is instrumental in adapting generalized pretraining models to more specific tasks in fields like chemistry.[60−62]

Leveraging a diverse training set and explicit human-guided instructions in the prompt, our fine-tuned model achieved a remarkable 84.8% accuracy and 93.9% recall in generating SMILES codes of mutated structures (Tables S3 and S4). This performance far outstrips the standard GPT-3.5 model (with 10.2% accuracy and 32.1% recall) and the GPT-4 model (with 32.1% accuracy and 84.6% recall). A parallel notable improve-

https://doi.org/10.1021/jacs.3c12086
J. Am. Chem. Soc. XXXX, XXX, XXX−XXX

ment was observed in IUPAC name generation (Figure S7). These enhancements are not confined to a specific kind of molecular editing; fine-tuning has proven to be an effective strategy in educating the LLM to manage all different kinds of linker mutation methods shown in this study and improving all performance metrics, including accuracy, precision, recall, and F1 scores, compared with the base GPT-3.5 and GPT-4 models (Figure S8). We also demonstrate that each of the output SMILES code and IUPAC name from the fine-tuned model can be iteratively fed to and analyzed by another GPT-4-based chemistry assistant, which assists researchers in assessing the commercial availability of the generated compounds by accessing a database (ZINC20,[63] Supporting Information, Section S2.4). It is paramount to acknowledge that while experienced researchers can propose similar structures and verify them by hand, the integration of AI serves to automate and expedite this process with less human bias, conserving valuable time and resources for researchers. In other words, many structures, whether proposed by AI or humans, often require time-intensive cross-referencing regarding the availability of precursors or synthetic feasibility, and therefore, automating the verification of hundreds of proposed structures can be instrumental for both experienced chemists and newcomers to the field. Fine-tuning enhances the capabilities of LLMs, making them more adept at proposing and verifying hypothetical linker structures and becoming efficient chemistry assistants. Researchers can thus explore potential variants of existing MOFs more efficiently, replacing successful linkers with new, feasible ones to modulate properties.

Synthesis and Structural Characterization. Central to our study is the premise that a strategic linker design can markedly influence MOF water-harvesting properties. In previous work, originating from the linker structure of the state-of-the-art water-harvesting adsorbent MOF-303,[27] a linker extension strategy can be employed to yield the new H2PZVDC linker through the introduction of a vinyl spacer to yield LAMOF-1 (Figure 1).[19] Further iterations on this structure, employing heteroatom replacement over several generations, yielded a suite of variants via multivariate tuning strategies.[42−46] This linker "evolution" introduced heterogeneity into the backbone, incorporating moieties such as pyrazole, thiophene, furan, thiazole, or an additional vinyl group. Cumulatively, this design paradigm represents a systematic expansion of linkers previously employed in the synthesis of certain aluminum rod MOFs.[29,30,64] Once a pool of linker candidates was proposed, they were selected based on commercial availability and evaluated by humans to propose routes for synthesis. Remarkably, we found that all 10 LAMOFs can be prepared in a similar manner by dissolving the respective linkers in DMF and mixing with aqueous $AlCl_3$, followed by solvothermal synthesis at 120 °C overnight. Furthermore, by applying the protocol we previously developed for the synthesis of aluminum rod MOFs,[17] we observed that some of the LAMOFs were amenable to green synthesis methods giving higher space-time-yield, or microwave-assisted synthesis to reduce reaction time to 1 h or less (Supporting Information, Section S3.2), both of which increase potential of practical use of LAMOFs in an industrial setting. We also note that the design and synthesis of multivariate LAMOF-7, LAMOF-8, and LAMOF-9, which contain more than one linker, were conducted following the discovery of the previous seven single-linker LAMOFs, using conditions similar to those of their parent single-linker LAMOFs. This strategy for making multivariate MOFs with mixed linkers was further discussed in our previous publications.[29,42,65]

The obtained microcrystalline powder was first characterized by powder X-ray diffraction (PXRD) analysis (Figure 3a). All of the resulting LAMOF series compounds showed high crystallinity. Previous investigations into the linker conformation of LAMOF-1 and insights from periodic density functional theory (DFT) calculations allowed us to implement similar strategies, revealing good agreement between experimental PXRD data and simulated structures (Figures S10−S19). Although prior endeavors yielded LAMOF-1 crystals, the linker configuration remained elusive, with only conformation insights from the aluminum secondary building unit (SBU) discerned. In this work, through extensive efforts, we used different crystallization conditions and obtained LAMOF-2 single crystals (approximately $20 \times 20 \times 10\ \mu m^3$) suitable for single-crystal X-ray diffraction (SCXRD) analysis (Supporting Information, Section S7). The results reveal that LAMOF-2 crystallizes in the $I4_1/a$ (no. 88) space group with unit cell parameters ($a = b = 26.259(3)$ Å, $c = 10.4728(16)$ Å), and more importantly, both the linker configuration and SBU conformation were elucidated from the SCXRD results. This empirical evidence is in agreement with prior computational predictions on SBU formation and linker configurations (Supporting Information, Section S11).[19] This alignment substantiates the validity of our PXRD simulated patterns and bolsters our confidence in constructing analogous models for the entire LAMOF series (Supporting Information, Sections S4 and S11).

Compositional Analysis, Thermal Stability, and Porosity. Among the 10 unique LAMOFs, seven are single-linker MOFs, while the remaining three have two linkers incorporated in the framework. We employed elemental analysis (EA), NMR spectroscopy, and scanning electron microscopy (SEM) integrated with energy-dispersive X-ray spectroscopy (EDS) to investigate the LAMOFs' composition. The resultant chemical formulas derived from the EA aligned well with theoretical calculations, underscoring the precision of the utilized LA linkers and their ratios. Significantly, for three multivariate MOFs (MTV-MOFs) LAMOF-8, LAMOF-9, and LAMOF-10, which contain two linkers—one with nitrogen or sulfur and one without—the nitrogen and sulfur percentage, inferred from an assumed 50:50 input ratio, matched closely with the EA results. This lends support to the hypothesis that each linker in these three MTV-MOFs maintains a balanced presence (Supporting Information, Section S5).

Next, all 10 LAMOFs were subjected to digestion proton NMR by dissolving in 5% NaOD in D_2O solution to determine their linker purity (Figures S20−S29). By comparing integral peak values in the NMR spectra from parent MOFs LAMOF-1 and LAMOF-3 for LAMOF-8, LAMOF-2 and LAMOF-6 for LAMOF-9, and LAMOF-2 and LAMOF-5 for LAMOF-10, we discerned a nearly equal distribution of each compositional linker in the three MTV-MOFs. This homogeneity reaffirms the consistency between input and output linker compositions, a phenomenon we have also noticed in other aluminum MTV-MOFs.[29,42] Further investigation using SEM−EDS measurements on the microcrystalline powder of all members of LAMOFs was conducted (Supporting Information, Section S6). SEM−EDS measurements conducted on the separated microcrystalline powder of LAMOFs revealed uniform signals associated with Al, O, and C in all crystallites, with additional N and S signals appearing in LAMOF-8, LAMOF-9, and LAMOF-10, which provides empirical evidence for the

https://doi.org/10.1021/jacs.3c12086
J. Am. Chem. Soc. XXXX, XXX, XXX−XXX

homogeneous distribution of the two linkers within the three multivariate LAMOFs (Figure S30).

In addition, the thermal stability and permanent porosity of the LAMOF series were investigated by using thermogravimetric analysis (TGA) and nitrogen sorption analysis, respectively. TGA under a nitrogen atmosphere evidenced a commendable thermal resilience, with no substantial weight loss noted below 300 °C (Figures S32−S41). This highlights their suitability for thermal regeneration processes. Subsequent nitrogen sorption analysis at 77 K offered insights into the Brunauer−Emmett−Teller (BET) surface area and pore volume of this class of materials, which were found to range from 1115 to 1895 $m^2 g^{-1}$ and 0.5 to 0.69 $cm^3 g^{-1}$, respectively (Figure 3b and Table 1). It is noteworthy that these values signify a substantial enhancement compared to the parent MOFs made from V-shape or linear dicarboxylate linker without vinyl expansion (Supporting Information, Sections S9.1 and S9.2). Moreover, the analysis of the nitrogen adsorption−desorption isotherms provided the pore size distribution for all 10 LAMOF compounds, which are detailed in the analyses (Figures S52−S61). These distributions align with expectations for frameworks constructed with LA linkers of similar length and without additional ring substitutions that might reduce the pore diameters.

Table 1. Summary of the Sorption, Physical, and Pore Structure Properties of the Water-Harvesting LAMOFs

material	BET surface area ($m^2 g^{-1}$)	pore volume ($cm^3 g^{-1}$)	pore diameter (Å)	isotherm step (% RH)	water uptake ($g g^{-1}$)
LAMOF-1	1895	0.69	10.9	25	0.64
LAMOF-2	1691	0.61	11.5	36	0.54
LAMOF-3	1423	0.50	11.0	45	0.47
LAMOF-4	1847	0.65	11.2	40	0.63
LAMOF-5	1115	0.56	9.4, 15.2	13	0.42
LAMOF-6	1372	0.50	10.4	18	0.47
LAMOF-7	1584	0.55	11.1	53	0.61
LAMOF-8	1738	0.62	11.0	31	0.54
LAMOF-9	1755	0.62	11.3	26	0.60
LAMOF-10	1469	0.54	12.2, 15.4	32	0.51

LAMOFs for Water Harvesting from Air. The water sorption profiles of the LAMOF series were initially examined by conducting water sorption measurements under isothermal conditions at room temperature. While interpreting the water sorption isotherms, in this study, three critical parameters warrant attention:[12,35,46]

1. *Isotherm shape*: A step-shaped water sorption isotherm is advantageous, enabling the release of substantial quantities of water through minor perturbations in temperature or pressure; the sharper, the more optimal.
2. *Humidity cutoff position*: This position on the isotherm indicates the capability of the material to adsorb water at a specific relative humidity (RH). For applications in desert environments, the RH is typically low, and materials that can function effectively at or below 40% RH are considered desirable.
3. *Water uptake*: This parameter represents the volume of water that can be captured and stored in the pore. Higher values are preferable.

The measurements reveal that all LAMOFs exhibit a single-step isotherm with isotherm inflection points at 13, 18, 25, 26, 31, 32, 36, 40, 45, and 53% RH (Figure 3c and Table 1). This range effectively encompasses nearly all desirable operational RH environments for water harvesting from arid air, allowing for the material selection to be more flexible and tailored to specific real-world environments. This is attributed to the reticular design and heterogeneity within the linker backbone; the diversity of linkers leads to varying pore environments with different levels of hydrophilic pocket units. Notably, only a minimal degree of hysteresis was observed for these LAMOFs, an essential characteristic of energy-efficient atmospheric water harvesting.

We observed that the single-step isotherm serves as more robust evidence for confirming the identity of the MTV-MOFs, ensuring that they are not merely physical mixtures of single-linker MOFs. This method appears to be more definitive than both PXRD and nitrogen isotherms, which often gives no direct evidence on whether a given compound qualifies as an MTV-MOF.[29] As demonstrated in this study, when two linkers integrate to form an MTV-MOF, the resultant water isotherm exhibits a single step, with the position of this step falling between the steps of its two parent MOFs. Conversely, if two linkers result in two MOFs that are merely physically combined, the water isotherm manifests two distinct steps and overall is the average of the two (Figures S74 and S75). Additionally, the water isotherm proves invaluable in determining if two MOFs are isoreticular, especially when one MOF has an unidentified SBU conformation and lacks crystal structures for verification (Figure 4). As discussed earlier, we identified LAMOF-1 and LAMOF-2 as having a *cis*−*trans*- or *cis*−*cis*-connected Al-rod SBU, respectively, based on the SCXRD study, while the SBU of the remaining LAMOFs can only be inferred by indirect evidence from PXRD. By combining linkers in a 50:50 ratio with those known to form specific SBUs, and subsequently analyzing the resulting water isotherm for either one or two steps, we were able to deduce the SBUs for the rest of the compounds. For instance, when the LAMOF-4 linker was combined with LAMOF-1, the resulting compound presented a two-step isotherm (Figure S72). This led us to determine that LAMOF-4 did not possess a *cis*−*trans* SBU. On the other hand, a subsequent experiment combining the LAMOF-3 linker with LAMOF-1 yielded a compound with a single-step isotherm, confirming the SBU of LAMOF-4 as the *cis*−*trans* SBU (Figure S73).

Regarding the water uptake of LAMOFs, a discernible trend exists, correlating water uptake with BET surface areas and specific pore volumes—the larger the surface area and pore volumes, the higher the gravimetric water uptake.[32] Notably, the highest water uptake was recorded in LAMOF-1 (0.64 $g g^{-1}$), previously reported, followed by LAMOF-4 (0.63 $g g^{-1}$) and LAMOF-9 (0.60 $g g^{-1}$) from this study (Figure 5a). Interestingly, LAMOF-9, developed with two linkers from LAMOF-3 (H_2TVDC) and LAMOF-6 (H_2TZVDC), emerged as a promising water-harvesting MOF. Its parent MOFs, LAMOF-2 and LAMOF-6, display contrasting properties. In particular, LAMOF-2 has higher water uptake yet less desirable humidity cutoff (>40%), while LAMOF-6 has more hydrophilic pockets but has lower uptake. In a synergistic fashion, LAMOF-9 inherits optimal properties from its parent MOFs, achieving an isotherm step at 26% RH and a water uptake (0.60 $g g^{-1}$) comparable to LAMOF-1. An added benefit is the wider commercial availability of precursors for synthesis of LAMOF-9

https://doi.org/10.1021/jacs.3c12086
J. Am. Chem. Soc. XXXX, XXX, XXX−XXX

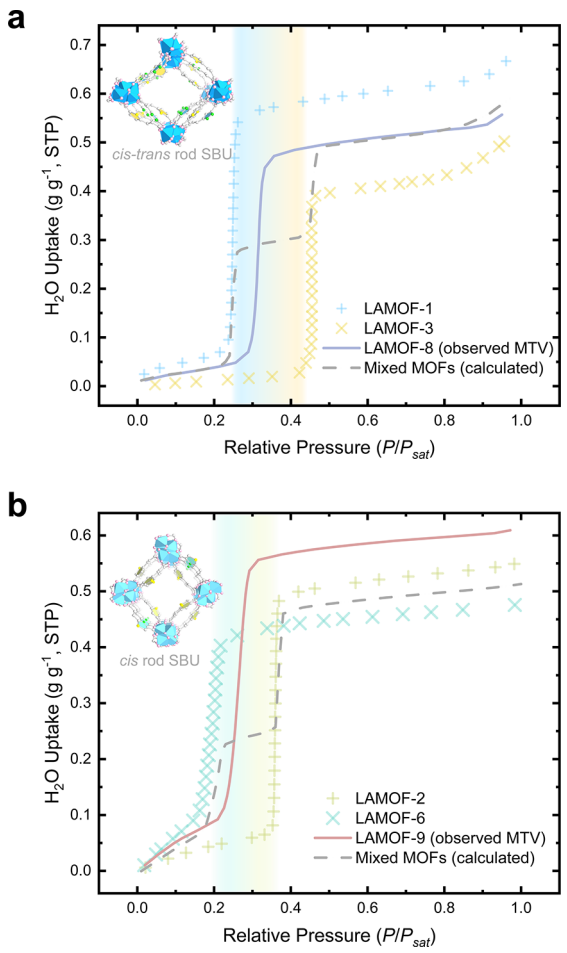


Figure 4. Comparison of the water sorption isotherms for single-linker MOFs, MTV-MOFs consisting of two linkers, and a physical mixture of two MOFs. The single-step water isotherms of (a) LAMOF-8 and (b) LAMOF-9 are shown in a solid line, and each has inflection points that lie between those of the two single-linker MOFs, confirming that the two single-linker MOFs share the same SBU. Conversely, if a physical mixture of two MOFs were formed, it would exhibit a two-step water isotherm shown in a dashed line, where the first step would align with one of the component MOFs and the second step would overlap with the step position from the other MOF.

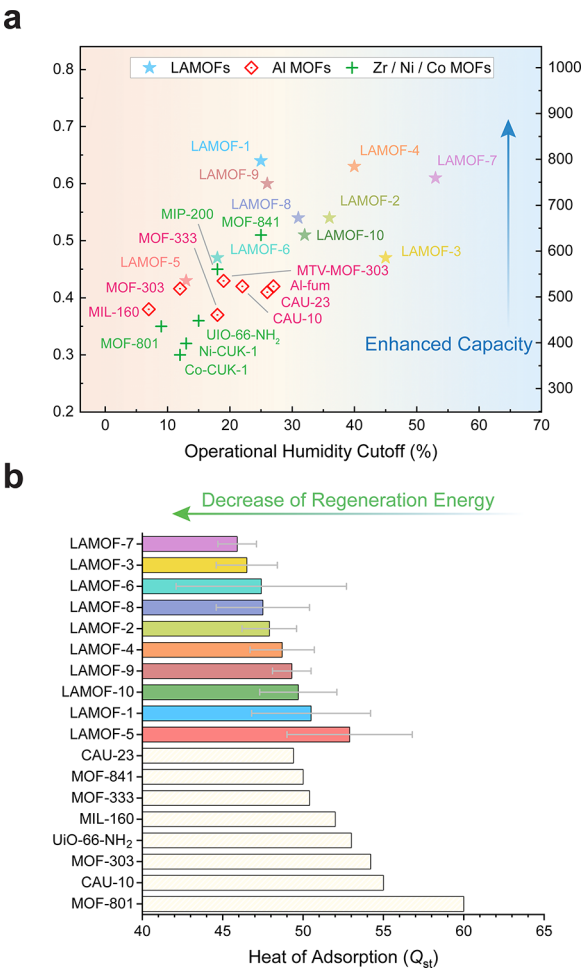


Figure 5. Water sorption properties of the LAMOF series. (a) Illustration of the step position of the water vapor isotherm and the associated water capacity, obtained from the water isotherm recorded at room temperature, along with corresponding data from other water-harvesting MOFs reported in the literature.[29,32,42,66−71] (b) Computed average heat of adsorption values for LAMOFs based on water sorption isotherms between 25 and 45 °C, compared with other water-harvesting MOFs. The loading-dependent trends and the respective standard errors can be extracted from Figures S62−S71.

linkers in contrast to those for LAMOF-1 (Supporting Information, Section S3.1), offering up to a substantial reduction in synthesis costs. This outcome accentuates the merits of reticular and multivariate linker designs in striking a balance between commercial feasibility and water-harvesting performance.

Additional water sorption data were collected on all compounds at 35 and 45 °C, which indicated consistent behavior across different temperatures (Figures S62−S71). These data sets allowed the calculation of the isosteric heat of water adsorption Q_{st} using the Clausius−Clapeyron relation (Figure 5b and Supporting Information, Section S10.2). The hydrophilicity of the MOF pore environment, as indicated by the position of the step of the isotherm, was reflected in the differences in Q_{st} values ranging between 45.9 and 53.6 kJ mol^{-1}. Furthermore, it was observed that, upon linker expansion, LAMOFs demonstrated a decrease in heat of adsorption, reflecting an average reduction of around 3 kJ mol^{-1} when compared to their parent MOFs measured under similar conditions (Table S13).[29,42,66,67] This reduction translates into a potential boost in the atmospheric water-harvesting efficiency. Notably, LAMOF-9 showcased an average Q_{st} of 49.3 kJ mol^{-1}, which translates to 23% less energy penalty than LAMOF-1 (50.5 kJ mol^{-1}) during the regeneration while still maintaining nearly identical water uptake and operational humidity cutoff, reaffirming the beneficial impact of the multivariate strategy on water adsorption enthalpies and the potential for finding promising candidates of water-harvesting sorbents.

To rationalize the observed effects of the MOF reticular design on its water adsorption behavior, periodic DFT calculations (Figure 6) were performed on the stable polymorphs identified for the single-linker LAMOFs synthesized in this work (Supporting Information, Section S11.1). The number and binding strength of different framework adsorption sites capable of seeding water molecules in the MOF pores were investigated, which collectively with the pore size of the MOF govern the overall water adsorption behavior. LAMOF-1 and LAMOF-3 exhibit the alternating *cis*−*trans* corner-sharing Al-rod SBU,[17,19,29,42,72] while LAMOF-2, LAMOF-4, LAMOF-5,

https://doi.org/10.1021/jacs.3c12086
J. Am. Chem. Soc. XXXX, XXX, XXX−XXX

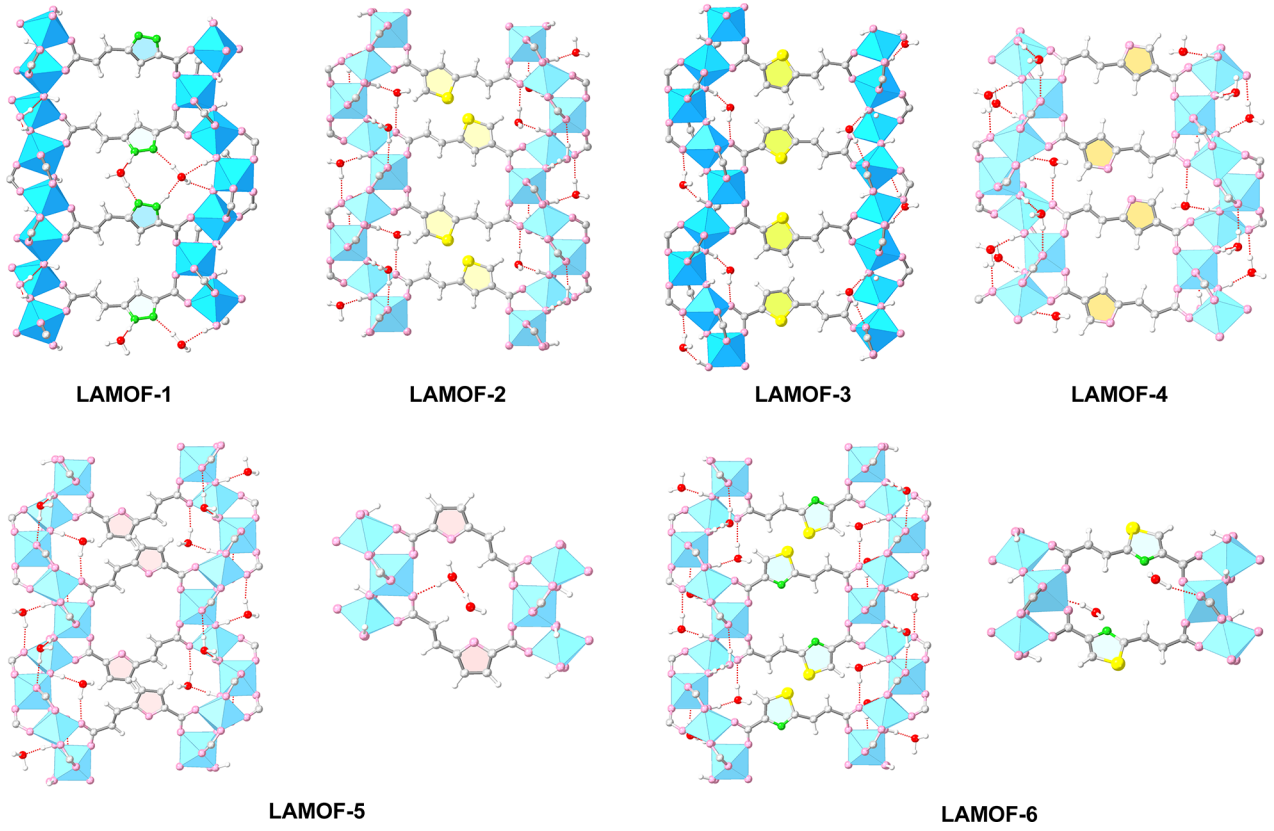


Figure 6. Identification of water adsorption sites in LAMOFs using periodic DFT calculations. LAMOF-1 and LAMOF-3 exhibit the alternating *cis*−*trans* Al-rod SBU, whereas the remaining LAMOFs exhibit the alternating cis-connected rod SBU. The reticular design of the LA linker results in different water sorption behaviors of the corresponding LAMOFs. For illustrative clarity: Al, blue octahedra; O, pink (SBU) or red (water); C, light gray; N, green; S, yellow; H, white.

and LAMOF-6 exhibit the *cis*-only corner-sharing Al-rod SBU.[30,44,64,73,74] First, we compare the water adsorption behavior of LAMOF-5 and LAMOF-6, which exhibit the smallest pore sizes (pore diameter = 7.3 and 8.3 Å, respectively, computed in Zeo++[75] using a nitrogen probe with kinetic diameter = 3.64 Å). Note that LAMOF-6 may coexist in two nearly isoenergetic *cis*-connected rod SBU polymorphs, which differ in the orientation of the vinyl group of the linker ($\Delta E = 0.8$ kJ mol^{-1} per [Al(μ_2−OH)(linker)] unit). The pore size of LAMOF-6 in its other polymorph is 7.3 Å. Our investigations focused on the water adsorption sites in the most stable polymorph, although pockets of the alternate polymorph may also be present. Two different water adsorption sites were identified in both LAMOF-5 and LAMOF-6. The first water molecule adsorbs in similar sites constituted of proximal μ_2-(OH)$_{rod}$ and O$_{carboxylate}$ functionalities in the two MOFs (Figures S90 and S91). The binding energy of a water molecule in this adsorption site is −60 kJ mol^{-1} in LAMOF-5 and −65 kJ mol^{-1} in LAMOF-6. A second water molecule adsorbs in the cavity formed by the opposite-facing linkers in the MOF through hydrogen bond interactions with the aromatic group of the linker (O$_{furan}$ for LAMOF-5 and N$_{linker}$ for LAMOF-6) and O$_{carboxylate}$ of the opposite linker, and intermolecular hydrogen bonds with the adsorbed water molecules (Figures S90 and S91). The binding energy of a water molecule in this second adsorption site is −44 kJ mol^{-1} in LAMOF-5 and −51 kJ mol^{-1} in LAMOF-6. The presence of two water "seeding" sites in the framework along with the smallest pore diameters computed results in isotherm step positions observed at low RH for these MOFs.

Next, we compare the water adsorption behavior of LAMOF-2 and LAMOF-4, which exhibit the same rod SBU topology[45,73,76,77] (*cis*-connected Al-rod SBU) and comparable pore sizes (pore diameter = 8.5 and 8.6 Å, respectively). Similar to LAMOF-5 and LAMOF-6, the first water molecule adsorbs through hydrogen bond interactions with adjacent μ_2-(OH)$_{rod}$ and O$_{carboxylate}$ groups in LAMOF-2 and LAMOF-4 (binding energy = 65 kJ mol^{-1}; Figures S87 and S89). Notably, in contrast to both LAMOF-5 and LAMOF-6, subsequent water molecules were not found to interact strongly with the linker aromatic groups (S$_{linker}$ and O$_{furan}$, binding energy = −27 kJ mol^{-1}) consistent with the previously reported studies.[29,78] Thus, the larger pore size of these MOFs coupled with only one strong adsorption site to seed water molecules results in the isotherm step position at higher RH for these MOFs compared to those of LAMOF-5 and LAMOF-6. The higher RH cutoff for LAMOF-4 compared to LAMOF-2 can be attributed to the smaller "effective" pore size of this MOF resulting from the larger van der Waals radii for S$_{linker}$ compared to O$_{furan}$. Finally, we compare the water adsorption behavior of LAMOF-1 and LAMOF-3, which exhibit an alternating *cis*−*trans* corner-sharing Al-rod SBU and pore sizes of 8.9 and 8.5 Å, respectively. The primary water molecules bind strongly to the pyrazole groups of the linker and μ_2-(OH)$_{rod}$ in LAMOF-1 by forming up to three hydrogen bonds (binding energy = −84 to −57 kJ mol^{-1}).[19] In contrast, water molecules adsorb in LAMOF-3 through hydrogen bond interactions with the μ_2-(OH)$_{rod}$ and O$_{carboxylate}$

https://doi.org/10.1021/jacs.3c12086
J. Am. Chem. Soc. XXXX, XXX, XXX−XXX

groups of the MOF (binding energy = −57 to −52 kJ mol^{-1}) and do not interact strongly with the S_{linker} group of the aromatic linker (Figure S88). Thus, the presence of very strong adsorption sites in LAMOF-1 leads to its isotherm step position at intermediate RH, despite its largest pore size among the LAMOFs investigated here, while LAMOF-3 exhibits the isotherm step at the highest RH (among the single-linker LAMOFs) collectively due to its larger pore size and weaker primary adsorption sites.

CONCLUDING REMARKS

Chemistry, with its intricate yet diverse structures and representations, is inherently a language in itself. LLMs can be perceived as motivated learners, while their foundational versions might grasp only the basics of chemistry. Through meticulous fine-tuning and coherent human instruction, LLMs' proficiency in a range of chemistry-related tasks can be significantly enhanced. This learning curve is attributed to their exposure to extensive examples provided by training data set in this specialized language, allowing them to write chemistry-related text fluently. As a proof of concept, in this study, we establish a data set of MOF linker mutations and demonstrate how a GPT-based chemistry assistant, via fine-tuning, can proficiently propose MOF linker designs by adeptly modifying existing linker structures, outperforming its base models. This methodology enables the GPT model to grasp the complex grammar of the chemical language and molecular representations, attaining good accuracy in generating linker structures. Building on this, there will be an exciting avenue in the integration of chemical structures alongside text in open-source LLMs.[79] Furthermore, the AI's potential is not limited to proposing high-level linker designs; it can also act as a surrogate for speeding up intricate simulations like DFT.[80] We envision that the intersection of chemistry and AI, including LLMs and machine learning methods, holds immense potential.

The 10 LAMOFs presented in this study underscore the pivotal role of linker design strategies in propelling the development of water-harvesting MOFs. A considerable spectrum of tunability is accomplished under operational RH conditions (between 13 and 53% RH), heats of adsorption (between 46 and 53 kJ mol^{-1}), and water uptake (up to 0.64 g g^{-1}) in this new series of aluminum-based MOFs, attributed to the variations in hydrophilicity in pore pockets resulting from the diversity in linker design. This versatility extends to reduced synthesis costs and potentially better commercial availability. Overall, these advancements are crucial for the journey of the development of next-generation water-harvesting MOF sorbents and hold substantial promise for the commercialization of water-harvesting technologies. The incorporation of these materials in large-scale atmospheric moisture extraction units could potentially provide practical water solutions, addressing the global water scarcity challenge universally and efficiently.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/jacs.3c12086.

Additional descriptions of general methods and materials; synthetic procedures of LA linkers and LAMOFs; detailed information in MOF linker mutation data set and additional discussions on fine-tuning GPT-3.5 model and evaluation on fine-tuned models; detailed information on the computational study of LAMOFs; characterization details of the MOF compounds, including elemental analysis, digestion proton NMR spectra, PXRD patterns, SEM images, crystallography data, TGA curves, nitrogen sorption isotherms, as well as water sorption isotherms (PDF)

Experimental nitrogen and water adsorption−desorption isotherms data in the adsorption information file (AIF) format; machine-readable files for the DFT-optimized structures for the initial water adsorption sites in LAMOFs in the crystallographic information file (CIF) format (ZIP)

Accession Codes

CCDC 2302011 (LAMOF-2) contains the supplementary crystallographic data for this paper. These data can be obtained free of charge via www.ccdc.cam.ac.uk/data_request/cif, or by emailing data_request@ccdc.cam.ac.uk, or by contacting The Cambridge Crystallographic Data Centre, 12 Union Road, CambridgeCB2 1EZ, U.K.; fax: +44 1223 336033.

AUTHOR INFORMATION

Corresponding Author

Omar M. Yaghi − *Department of Chemistry, Kavli Energy Nanoscience Institute, and Bakar Institute of Digital Materials for the Planet, College of Computing, Data Science, and Society, University of California, Berkeley, California 94720, United States; KACST−UC Berkeley Center of Excellence for Nanomaterials for Clean Energy Applications, King Abdulaziz City for Science and Technology, Riyadh 11442, Saudi Arabia;* orcid.org/0000-0002-5611-3325; Email: yaghi@berkeley.edu

Authors

Zhiling Zheng − *Department of Chemistry, Kavli Energy Nanoscience Institute, and Bakar Institute of Digital Materials for the Planet, College of Computing, Data Science, and Society, University of California, Berkeley, California 94720, United States;* orcid.org/0000-0001-6090-2258

Ali H. Alawadhi − *Department of Chemistry and Kavli Energy Nanoscience Institute, University of California, Berkeley, California 94720, United States;* orcid.org/0000-0003-2680-5221

Saumil Chheda − *Department of Chemistry, Kavli Energy Nanoscience Institute, and Bakar Institute of Digital Materials for the Planet, College of Computing, Data Science, and Society, University of California, Berkeley, California 94720, United States; Department of Chemical Engineering and Materials Science, Department of Chemistry, and Chemical Theory Center, University of Minnesota—Twin Cities, Minneapolis, Minnesota 55455, United States;* orcid.org/0000-0002-0989-5707

S. Ephraim Neumann − *Department of Chemistry and Kavli Energy Nanoscience Institute, University of California, Berkeley, California 94720, United States;* orcid.org/0000-0002-8515-9621

Nakul Rampal − *Department of Chemistry, Kavli Energy Nanoscience Institute, and Bakar Institute of Digital Materials for the Planet, College of Computing, Data Science, and Society, University of California, Berkeley, California 94720, United States*

Shengchao Liu − *Bakar Institute of Digital Materials for the Planet, College of Computing, Data Science, and Society and Department of Electrical Engineering and Computer Sciences,*

https://doi.org/10.1021/jacs.3c12086
J. Am. Chem. Soc. XXXX, XXX, XXX−XXX

University of California, Berkeley, California 94720, United States

Ha L. Nguyen − *Department of Chemistry and Kavli Energy Nanoscience Institute, University of California, Berkeley, California 94720, United States;* orcid.org/0000-0002-4977-925X

Yen-hsu Lin − *Department of Chemistry and Kavli Energy Nanoscience Institute, University of California, Berkeley, California 94720, United States*

Zichao Rong − *Department of Chemistry, Kavli Energy Nanoscience Institute, and Bakar Institute of Digital Materials for the Planet, College of Computing, Data Science, and Society, University of California, Berkeley, California 94720, United States;* orcid.org/0000-0002-9014-9540

J. Ilja Siepmann − *Department of Chemical Engineering and Materials Science, Department of Chemistry, and Chemical Theory Center, University of Minnesota—Twin Cities, Minneapolis, Minnesota 55455, United States;* orcid.org/0000-0003-2534-4507

Laura Gagliardi − *Department of Chemistry, Pritzker School of Molecular Engineering, Chicago Center for Theoretical Chemistry, University of Chicago, Chicago, Illinois 60637, United States;* orcid.org/0000-0001-5227-1396

Anima Anandkumar − *Computing and Mathematical Sciences, California Institute of Technology, Pasadena, California 91125, United States; NVIDIA Corporation, Santa Clara, California 95051, United States*

Christian Borgs − *Bakar Institute of Digital Materials for the Planet, College of Computing, Data Science, and Society and Department of Electrical Engineering and Computer Sciences, University of California, Berkeley, California 94720, United States*

Jennifer T. Chayes − *Bakar Institute of Digital Materials for the Planet, College of Computing, Data Science, and Society, Department of Electrical Engineering and Computer Sciences, Department of Statistics, and School of Information, University of California, Berkeley, California 94720, United States; Department of Mathematics, University of California, Berkeley, California 94720, United States*

Complete contact information is available at: https://pubs.acs.org/10.1021/jacs.3c12086

Author Contributions

The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

Notes

The authors declare the following competing financial interest(s): Omar M. Yaghi is co-founder of ATOCO Inc., aiming at commercializing related technologies.

■ ACKNOWLEDGMENTS

This material is based upon work supported by the Defense Advanced Research Projects Agency (DARPA) under contract HR0011-21-C-0020. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of DARPA. The computational work is partially supported by the Department of Energy (DOE), Office of Basic Energy Sciences, Division of Chemical Sciences, Geosciences, and Biosciences, under award DE-SC0023454. In addition, the National Science Foundation (NSF), Division of Chemistry, Chemical Structure, Dynamics, and Mechanisms A (CSDM−A), provided support for the computational resources, award number: CHE-2223442. The authors also extend their gratitude to the Research Computing Center at the University of Chicago for providing computational resources. Additionally, this research utilized the facilities of the Advanced Light Source, a DOE Office of Science User Facility, under contract no. DE-AC02-05CH11231. The study made use of instruments located in the College of Chemistry Nuclear Magnetic Resonance (NMR) Facility, partially supported by NIH S10OD024998. The authors are grateful to Dr. Seth Cohen (DARPA) and Dr. David Moore (General Electric) for their helpful comments and suggestions on this work. Moreover, Z.Z. expresses gratitude to Drs. Nikita Hanikel and Daria Kurandina, Ms. Oufan Zhang, and Mr. Boyu Qie for their valuable discussions. Z.Z. also acknowledges financial support from a Kavli ENSI Graduate Student Fellowship.

■ REFERENCES

(1) McDonald, R. I.; Green, P.; Balk, D.; Fekete, B. M.; Revenga, C.; Todd, M.; Montgomery, M. Urban growth, climate change, and freshwater availability. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108* (15), 6312−6317.

(2) Flörke, M.; Schneider, C.; McDonald, R. I. Water competition between cities and agriculture driven by climate change and urban growth. *Nat. Sustain* **2018**, *1* (1), 51−58.

(3) Jaeger, W. K.; Plantinga, A. J.; Chang, H.; Dello, K.; Grant, G.; Hulse, D.; McDonnell, J. J.; Lancaster, S.; Moradkhani, H.; Morzillo, A. T.; Mote, P.; Nolin, A.; Santelmann, M.; Wu, J. Toward a formal definition of water scarcity in natural-human systems. *Water Resour. Res.* **2013**, *49* (7), 4506−4517.

(4) Malmqvist, B.; Rundle, S. Threats to the running water ecosystems of the world. *Environ. Conserv.* **2002**, *29* (2), 134−153.

(5) Water, U. N. *2018 UN World Water Development Report, Nature-based Solutions for Water*; https://www.unwater.org/publications/world-water-development-report-2018/.

(6) Shannon, M. A.; Bohn, P. W.; Elimelech, M.; Georgiadis, J. G.; Marinas, B. J.; Mayes, A. M. Science and technology for water purification in the coming decades. *Nanosci. Technol.* **2009**, 337−346.

(7) Oelkers, E. H.; Hering, J. G.; Zhu, C. Water: is there a global crisis? *Elements* **2011**, *7* (3), 157−162.

(8) Wahlgren, R. V. Atmospheric water vapour processor designs for potable water production: a review. *Water Res.* **2001**, *35* (1), 1−22.

(9) Lord, J.; Thomas, A.; Treat, N.; Forkin, M.; Bain, R.; Dulac, P.; Behroozi, C. H.; Mamutov, T.; Fongheiser, J.; Kobilansky, N.; Washburn, S.; Truesdell, C.; Lee, C.; Schmaelzle, P. H. Global potential for harvesting drinking water from air using solar energy. *Nature* **2021**, *598* (7882), 611−617.

(10) Lu, H.; Shi, W.; Guo, Y.; Guan, W.; Lei, C.; Yu, G. Materials engineering for atmospheric water harvesting: progress and perspectives. *Adv. Mater.* **2022**, *34* (12), 2110079.

(11) Wang, J.; Hua, L.; Li, C.; Wang, R. Atmospheric water harvesting: critical metrics and challenges. *Energy Environ. Sci.* **2022**, *15* (12), 4867−4871.

(12) Lin, H.; Yang, Y.; Hsu, Y.-C.; Zhang, J.; Welton, C.; Afolabi, I.; Loo, M.; Zhou, H.-C. Metal−Organic Frameworks for Water Harvesting and Concurrent Carbon Capture: A Review for Hygroscopic Materials. *Adv. Mater.* **2023**, 2209073.

(13) Entezari, A.; Esan, O. C.; Yan, X.; Wang, R.; An, L. Sorption-based Atmospheric Water Harvesting: Materials, Components, Systems, and Applications. *Adv. Mater.* **2023**, *35*, 2210957.

(14) Xu, W.; Yaghi, O. M. Metal−organic frameworks for water harvesting from air, anywhere, anytime. *ACS Cent. Sci.* **2020**, *6* (8), 1348−1354.

(15) Zhou, X.; Lu, H.; Zhao, F.; Yu, G. Atmospheric water harvesting: a review of material and structural designs. *ACS Mater. Lett.* **2020**, *2* (7), 671−684.

(16) Ntep, T. J. M. M.; Wahiduzzaman, M.; Laurenz, E.; Cornu, I.; Mouchaham, G.; Dovgaliuk, I.; Nandi, S.; Knop, K.; Jansen, C.; Nouar, F.; Florian, P.; Füldner, G.; Maurin, G.; Janiak, C.; Serre, C. When Polymorphism in Metal−Organic Frameworks Enables Water Sorption Profile Tunability for Enhancing Heat Allocation and Water Harvesting Performance. *Adv. Mater.* **2023**, 2211302.
(17) Zheng, Z.; Nguyen, H. L.; Hanikel, N.; Li, K. K.-Y.; Zhou, Z.; Ma, T.; Yaghi, O. M. High-yield, green and scalable methods for producing MOF-303 for water harvesting from desert air. *Nat. Protoc.* **2023**, *18*, 136−156.
(18) Hu, Y.; Wang, Y.; Fang, Z.; Yao, B.; Ye, Z.; Peng, X. Ca-MOF-Derived Porous Sorbents for High-Yield Solar-Driven Atmosphere Water Harvesting. *ACS Appl. Mater. Interfaces* **2023**, *15* (38), 44942−44952.
(19) Hanikel, N.; Kurandina, D.; Chheda, S.; Zheng, Z.; Rong, Z.; Neumann, S. E.; Sauer, J.; Siepmann, J. I.; Gagliardi, L.; Yaghi, O. M. MOF Linker Extension Strategy for Enhanced Atmospheric Water Harvesting. *ACS Cent. Sci.* **2023**, *9* (3), 551−557.
(20) Rieth, A. J.; Yang, S.; Wang, E. N.; Dincă, M. Record atmospheric fresh water capture and heat transfer with a material operating at the water uptake reversibility limit. *ACS Cent. Sci.* **2017**, *3* (6), 668−672.
(21) Zheng, Z.; Alawadhi, A. H.; Yaghi, O. M. Green Synthesis and Scale-Up of MOFs for Water Harvesting from Air. *Mol. Front. J.* **2023**, *7* (1), 1−20.
(22) Logan, M. W.; Langevin, S.; Xia, Z. Reversible atmospheric water harvesting using metal-organic frameworks. *Sci. Rep.* **2020**, *10* (1), 1492.
(23) Neumann, S. E.; Neumann, K.; Zheng, Z.; Hanikel, N.; Tsao, J.; Yaghi, O. M. Harvesting Water in the Classroom. *J. Chem. Educ.* **2023**, *100* (11), 4482−4487.
(24) Abtab, S. M. T.; Alezi, D.; Bhatt, P. M.; Shkurenko, A.; Belmabkhout, Y.; Aggarwal, H.; Weseliński, Ł. J.; Alsadun, N.; Samin, U.; Hedhili, M. N.; Eddaoudi, M. Reticular chemistry in action: A hydrolytically stable MOF capturing twice its weight in adsorbed water. *Chem* **2018**, *4* (1), 94−105.
(25) Song, W.; Zheng, Z.; Alawadhi, A. H.; Yaghi, O. M. MOF water harvester produces water from Death Valley desert air in ambient sunlight. *Nat. Water* **2023**, *1* (7), 626−634.
(26) Hanikel, N.; Prévot, M. S.; Fathieh, F.; Kapustin, E. A.; Lyu, H.; Wang, H.; Diercks, N. J.; Glover, T. G.; Yaghi, O. M. Rapid cycling and exceptional yield in a metal-organic framework water harvester. *ACS Cent. Sci.* **2019**, *5* (10), 1699−1706.
(27) Fathieh, F.; Kalmutzki, M. J.; Kapustin, E. A.; Waller, P. J.; Yang, J.; Yaghi, O. M. Practical water production from desert air. *Sci. Adv.* **2018**, *4* (6), No. eaat3198.
(28) Wang, B.; Zhou, X.; Guo, Z.; Liu, W. Recent advances in atmosphere water harvesting: Design principle, materials, devices, and applications. *Nano Today* **2021**, *40*, No. 101283.
(29) Hanikel, N.; Pei, X.; Chheda, S.; Lyu, H.; Jeong, W.; Sauer, J.; Gagliardi, L.; Yaghi, O. M. Evolution of water structures in metal-organic frameworks for improved atmospheric water harvesting. *Science* **2021**, *374* (6566), 454−459.
(30) Cho, K. H.; Borges, D. D.; Lee, U.-H.; Lee, J. S.; Yoon, J. W.; Cho, S. J.; Park, J.; Lombardo, W.; Moon, D.; Sapienza, A.; Maurin, G.; Chang, J.-S. Rational design of a robust aluminum metal-organic framework for multi-purpose water-sorption-driven heat allocations. *Nat. Commun.* **2020**, *11* (1), 5112.
(31) Cadiau, A.; Lee, J. S.; Borges, D. D.; Fabry, P.; Devic, T.; Wharmby, M. T.; Martineau, C.; Foucher, D.; Taulelle, F.; Jun, C.-H.; Hwang, Y. K.; Stock, N.; De Lange, M. F.; Kapteijn, F.; Gascon, J.; Maurin, G.; Chang, J.-S.; Serre, C. Design of hydrophilic metal organic framework water adsorbents for heat reallocation. *Adv. Mater.* **2015**, *27* (32), 4775−4780.
(32) Furukawa, H.; Gandara, F.; Zhang, Y.-B.; Jiang, J.; Queen, W. L.; Hudson, M. R.; Yaghi, O. M. Water adsorption in porous metal−organic frameworks and related materials. *J. Am. Chem. Soc.* **2014**, *136* (11), 4369−4381.
(33) Burtch, N. C.; Jasuja, H.; Walton, K. S. Water stability and adsorption in metal−organic frameworks. *Chem. Rev.* **2014**, *114* (20), 10575−10612.
(34) Rieth, A. J.; Wright, A. M.; Dincă, M. Kinetic stability of metal−organic frameworks for corrosive and coordinating gas capture. *Nat. Rev. Mater.* **2019**, *4* (11), 708−725.
(35) Hanikel, N.; Prévot, M. S.; Yaghi, O. M. MOF water harvesters. *Nat. Nanotechnol.* **2020**, *15* (5), 348−355.
(36) Ghasempour, H.; Wang, K.-Y.; Powell, J. A.; ZareKarizi, F.; Lv, X.-L.; Morsali, A.; Zhou, H.-C. Metal−organic frameworks based on multicarboxylate linkers. *Coord. Chem. Rev.* **2021**, *426*, No. 213542.
(37) Soni, S.; Bajpai, P. K.; Arora, C. A review on metal-organic framework: Synthesis, properties and ap-plication. *Characterization and Application of Nanomaterials* **2020**, *3* (2), 87−106.
(38) Loiseau, T.; Mellot-Draznieks, C.; Muguerra, H.; Férey, G.; Haouas, M.; Taulelle, F. Hydrothermal synthesis and crystal structure of a new three-dimensional aluminum-organic framework MIL-69 with 2, 6-naphthalenedicarboxylate (ndc), Al (OH)(ndc)· H2O. *C. R. Chim.* **2005**, *8* (3−4), 765−772.
(39) Krüger, M.; Reinsch, H.; Inge, A. K.; Stock, N. Effect of partial linker fluorination and linker extension on structure and properties of the Al-MOF CAU-10. *Microporous Mesoporous Mater.* **2017**, *249*, 128−136.
(40) Zhang, X.; Zhang, X.; Johnson, J. A.; Chen, Y.-S.; Zhang, J. Highly porous zirconium metal−organic frameworks with β-UH3-like topology based on elongated tetrahedral linkers. *J. Am. Chem. Soc.* **2016**, *138* (27), 8380−8383.
(41) Deng, H.; Grunder, S.; Cordova, K. E.; Valente, C.; Furukawa, H.; Hmadeh, M.; Gándara, F.; Whalley, A. C.; Liu, Z.; Asahina, S.; Kazumori, H.; O'Keeffe, M.; Terasaki, O.; Stoddart, J. F.; Yaghi, O. M. Large-pore apertures in a series of metal-organic frameworks. *Science* **2012**, *336* (6084), 1018−1023.
(42) Zheng, Z.; Hanikel, N.; Lyu, H.; Yaghi, O. M. Broadly Tunable Atmospheric Water Harvesting in Multivariate Metal−Organic Frameworks. *J. Am. Chem. Soc.* **2022**, *144* (49), 22669−22675.
(43) Truong, B. N.; Borges, D. D.; Park, J.; Lee, J. S.; Jo, D.; Chang, J.-S.; Cho, S. J.; Maurin, G.; Cho, K. H.; Lee, U.-H. Tuning Hydrophilicity of Aluminum MOFs by a Mixed-Linker Strategy for Enhanced Performance in Water Adsorption-Driven Heat Allocation Application. *Adv. Sci.* **2023**, *10*, No. 2301311.
(44) Schlüsener, C.; Xhinovci, M.; Ernst, S.-J.; Schmitz, A.; Tannert, N.; Janiak, C. Solid-solution mixed-linker synthesis of isoreticular Al-based MOFs for an easy hydrophilicity tuning in water-sorption heat transformations. *Chem. Mater.* **2019**, *31* (11), 4051−4062.
(45) Schlüsener, C.; Jordan, D. N.; Xhinovci, M.; Ntep, T. J. M. M.; Schmitz, A.; Giesen, B.; Janiak, C. Probing the limits of linker substitution in aluminum MOFs through water vapor sorption studies: mixed-MOFs instead of mixed-linker CAU-23 and MIL-160 materials. *Dalton Trans.* **2020**, *49* (22), 7373−7383.
(46) Ma, A.; Cong, H.; Deng, H. Multivariate MOF for optimizing atmospheric water harvesting. *Green Energy & Environment* **2022**, *7* (4), 575−577.
(47) Weininger, D. SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. *J. Chem. Inf. Comput. Sci.* **1988**, *28* (1), 31−36.
(48) Krenn, M.; Häse, F.; Nigam, A.; Friederich, P.; Aspuru-Guzik, A. Self-referencing embedded strings (SELFIES): A 100% robust molecular string representation. *Mach. Learn.: Sci. Technol.* **2020**, *1* (4), No. 045024.
(49) Favre, H. A.; Powell, W. H. *Nomenclature of organic chemistry: IUPAC recommendations and preferred names 2013*; Royal Society of Chemistry, 2013.
(50) Zheng, Z.; Zhang, O.; Borgs, C.; Chayes, J. T.; Yaghi, O. M. ChatGPT Chemistry Assistant for Text Mining and Prediction of MOF Synthesis. *J. Am. Chem. Soc.* **2023**, *145* (32), 18048−18062.
(51) Jablonka, K. M.; Schwaller, P.; Ortega-Guerrero, A.; Smit, B. Is GPT-3 all you need for low-data discovery in chemistry? *ChemRxiv*; Cambridge Open Engage: Cambridge, DOI: 10.26434/chemrxiv-2023-fw8n4 (accessed 2023−05−16), 2023.

https://doi.org/10.1021/jacs.3c12086
J. Am. Chem. Soc. XXXX, XXX, XXX−XXX

(52) Guo, T.; Guo, K.; Liang, Z.; Guo, Z.; Chawla, N. V.; Wiest, O.; Zhang, X. What indeed can GPT models do in chemistry? A comprehensive benchmark on eight tasks. arXiv preprint arXiv:2305.18365; 2023, (accessed 2023−09−10).
(53) White, A. D. The future of chemistry is language. *Nat. Rev. Chem.* **2023**, *7*, 457−458.
(54) Hatakeyama-Sato, K.; Yamane, N.; Igarashi, Y.; Nabae, Y.; Hayakawa, T. Prompt engineering of GPT-4 for chemical research: what can/cannot be done? *ChemRxiv*; Cambridge Open Engage: Cambridge, 2023, DOI: 10.26434/chemrxiv-2023-s1x5p (accessed 2023−06−05).
(55) White, A. D.; Hocky, G. M.; Gandhi, H. A.; Ansari, M.; Cox, S.; Wellawatte, G. P.; Sasmal, S.; Yang, Z.; Liu, K.; Singh, Y.; Ccoa, W. J. P. Assessment of chemistry knowledge in large language models that generate code. *Digital Discovery* **2023**, *2* (2), 368−376.
(56) OpenAI. *GPT-4 technical report*; DOI: 10.48550/arXiv:2303.08774 (accessed 2023−03−27).
(57) Zheng, Z.; Rong, Z.; Rampal, N.; Borgs, C.; Chayes, J. T.; Yaghi, O. M. A GPT-4 Reticular Chemist for Guiding MOF Discovery. *Angew. Chem., Int. Ed.* **2023**, *62*, No. e202311983.
(58) Bubeck, S.; Chandrasekaran, V.; Eldan, R.; Gehrke, J.; Horvitz, E.; Kamar, E.; Lee, P.; Lee, Y. T.; Li, Y.; Lundberg, S.; Nori, H.; Palangi, H.; Ribeiro, M. T.; Zhang, Y.; Sparks of artificial general intelligence: Early experiments with gpt-4. arXiv preprint arXiv:2303.12712; 2023, DOI: 10.48550/arXiv:2303.12712 (accessed 2023−04−13).
(59) Liu, S.; Wang, J.; Yang, Y.; Wang, C.; Liu, L.; Guo, H.; Xiao, C. ChatGPT-powered Conversational Drug Editing Using Retrieval and Domain Feedback. arXiv preprint arXiv:2305.18090; DOI: 10.48550/arXiv.2305.18090 (accessed 2023−05−29).
(60) Radiya-Dixit, E.; Wang, X. How fine can fine-tuning be? learning efficient language models. In *International Conference on Artificial Intelligence and Statistics*; PMLR, 2020; pp 2435−2443.
(61) Thirunavukarasu, A. J.; Ting, D. S. J.; Elangovan, K.; Gutierrez, L.; Tan, T. F.; Ting, D. S. W. Large language models in medicine. *Nat. Med.* **2023**, *29* (8), 1930−1940.
(62) Xie, T.; Wa, Y.; Huang, W.; Zhou, Y.; Liu, Y.; Linghu, Q.; Wang, S.; Kit, C.; Grazian, C.; Hoex, B. Large Language Models as Master Key: Unlocking the Secrets of Materials Science with GPT. arXiv preprint arXiv:2304.02213; DOI: 10.48550/arXiv:2304.02213 (accessed 2023−04−12).
(63) Irwin, J. J.; Tang, K. G.; Young, J.; Dandarchuluun, C.; Wong, B. R.; Khurelbaatar, M.; Moroz, Y. S.; Mayfield, J.; Sayle, R. A. ZINC20—a free ultralarge-scale chemical database for ligand discovery. *J. Chem. Inf. Model.* **2020**, *60* (12), 6065−6073.
(64) Tannert, N.; Jansen, C.; Nießing, S.; Janiak, C. Robust synthesis routes and porosity of the Al-based metal−organic frameworks Al-fumarate, CAU-10-H and MIL-160. *Dalton Trans.* **2019**, *48* (9), 2967−2976.
(65) Yaghi, O. M.; Hanikel, N.; Hao, L. Multivariate and other metal-organic frameworks, and uses thereof. U.S. Patent No. 62/771,537, University of California, 2020.
(66) Teo, H. W. B.; Chakraborty, A.; Kitagawa, Y.; Kayal, S. Experimental study of isotherms and kinetics for adsorption of water on Aluminium Fumarate. *Int. J. Heat Mass Transfer* **2017**, *114*, 621−627.
(67) Silva, M. P.; Ribeiro, A. M.; Silva, C. G.; Nogueira, I. B. R.; Cho, K.-H.; Lee, U.-H.; Faria, J. L.; Loureiro, J. L.; Chang, J.-S.; Rodrigues, A. E.; Ferreira, A. MIL-160(Al) MOF's potential in adsorptive water harvesting. *Adsorption* **2021**, *27* (2), 213−226.
(68) Fröhlich, D.; Pantatosaki, E.; Kolokathis, P. D.; Markey, K.; Reinsch, H.; Baumgartner, M.; van der Veen, M. A.; De Vos, D. E.; Stock, N.; Papadopoulos, G. K.; Henninger, S. K.; Janiak, C. Water adsorption behaviour of CAU-10-H: a thorough investigation of its structure−property relationships. *J. Mater. Chem. A* **2016**, *4* (30), 11859−11869.
(69) Lee, J. S.; Yoon, J. W.; Mileo, P. G. M.; Cho, K. H.; Park, J.; Kim, K.; Kim, H.; de Lange, M. F.; Kapteijn, F.; Maurin, G.; Humphrey, S. M.; Chang, J.-S. Porous metal−organic framework CUK-1 for adsorption heat allocation toward green applications of natural refrigerant water. *ACS Appl. Mater. Interfaces* **2019**, *11* (29), 25778−25789.
(70) Wang, S.; Lee, J. S.; Wahiduzzaman, M.; Park, J.; Muschi, M.; Martineau-Corcos, C.; Tissot, A.; Cho, K. H.; Marrot, J.; Shepard, W.; Maurin, G.; Chang, J.-S.; Serre, C. A robust large-pore zirconium carboxylate metal−organic framework for energy-efficient water-sorption-driven refrigeration. *Nat. Energy* **2018**, *3* (11), 985−993.
(71) Lu, F.-F.; Gu, X.-W.; Wu, E.; Li, B.; Qian, G. Systematic evaluation of water adsorption in isoreticular UiO-type metal−organic frameworks. *J. Mater. Chem. A* **2023**, *11* (3), 1246−1255.
(72) Lenzen, D.; Zhao, J.; Ernst, S.-J.; Wahiduzzaman, M.; Inge, A. K.; Fröhlich, D.; Xu, H.; Bart, H.-J.; Janiak, C.; Henninger, S.; Maurin, G.; Zou, X.; Stock, N. A metal−organic framework for efficient water-based ultra-low-temperature-driven cooling. *Nat. Commun.* **2019**, *10* (1), 3025.
(73) Zheng, Z.; Zhang, O.; Nguyen, H. L.; Rampal, N.; Alawadhi, A. H.; Rong, Z.; Head-Gordon, T.; Borgs, C.; Chayes, J. T.; Yaghi, O. M. ChatGPT Research Group for Optimizing the Crystallinity of MOFs and COFs. *ACS Cent. Sci.* **2023**, *9* (11), 2161−2170.
(74) Alawadhi, A. H.; Chheda, S.; Stroscio, G. D.; Rong, Z.; Kurandina, D.; Nguyen, H. L.; Rampal, N.; Zheng, Z.; Gagliardi, L.; Yaghi, O. Harvesting Water from Air with High-Capacity, Stable Furan-Based Metal−Organic Frameworks. *ChemRxiv*; Cambridge Open Engage: Cambridge, 2023; DOI: 10.26434/chemrxiv-2023-frr55 (accessed 2023−10−31).
(75) Willems, T. F.; Rycroft, C. H.; Kazi, M.; Meza, J. C.; Haranczyk, M. Algorithms and tools for high-throughput geometry-based analysis of crystalline porous materials. *Microporous Mesoporous Mater.* **2012**, *149* (1), 134−141.
(76) Noa, F. M. A.; Abrahamsson, M.; Ahlberg, E.; Cheung, O.; Göb, C. R.; McKenzie, C. J.; Öhrström, L. A unified topology approach to dot-, rod-, and sheet-MOFs. *Chem* **2021**, *7* (9), 2491−2512.
(77) Schoedel, A.; Li, M.; Li, D.; O'Keeffe, M.; Yaghi, O. M. Structures of metal−organic frameworks with rod secondary building units. *Chem. Rev.* **2016**, *116* (19), 12466−12535.
(78) Silva, W. G.; van Wijngaarden, J. Characterization of Large-Amplitude Motions and Hydrogen Bonding Interactions in the Thiophene−Water Complex by Rotational Spectroscopy. *J. Phys. Chem. A* **2021**, *125* (16), 3425−3431.
(79) Liu, S.; Nie, W.; Wang, C.; Lu, J.; Qiao, Z.; Liu, L.; Tang, J.; Xiao, C.; Anandkumar, A. Multi-modal molecule structure-text model for text-based retrieval and editing. arXiv preprint arXiv:2212.10789; 2022; DOI: 10.48550/arXiv.2212.10789 (accessed 2022−12−21).
(80) Qiao, Z.; Welborn, M.; Anandkumar, A.; Manby, F. R.; Miller, T. F., III OrbNet: Deep learning for quantum chemistry using symmetry-adapted atomic-orbital features. *J. Chem. Phys.* **2020**, *153* (12), No. 124111.

https://doi.org/10.1021/jacs.3c12086
J. Am. Chem. Soc. XXXX, XXX, XXX−XXX