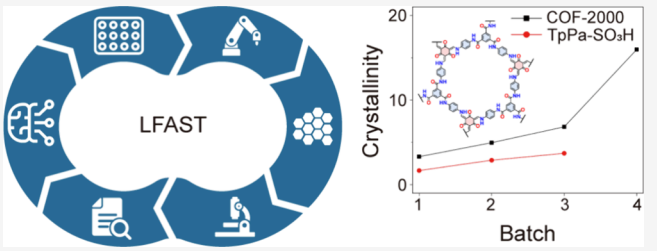
JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY
pubs.acs.org/JACS
Article

Synthesis of Highly Crystalline Covalent Organic Frameworks Using Large Language Models

Kaiyu Wang, Daehyun Daniel Ahn, Nakul Rampal, Jackson Thomassian, Neda S. Sabeva, Om Kannan, Christian Borgs, Jennifer T. Chayes, and Omar M. Yaghi*

Cite This: https://doi.org/10.1021/jacs.5c23233
Read Online

ABSTRACT: Crystallizing covalent organic frameworks (COFs) remain a central challenge in reticular chemistry, as achieving long-range order typically requires extensive trial-and-error optimization over many months or years. Here, we demonstrate that by integrating a deep research agent within ChatGPT, this process can be markedly accelerated, reducing the crystallization timeline to less than one month. Our approach, termed the LLM For Accelerated Synthesis Technique (LFAST), operates through two interlinked cycles. In the first, we formulated a structured, multistep prompt to guide the deep research agent in mining, correlating, and validating synthesis parameters from the relevant chemical literature. This yielded an expanded and refined design space for reaction condition screening. In the second, these conditions were executed by using an automated synthesis platform coupled with high-throughput powder X-ray diffraction analysis. Using a widely reported β-ketoenamine-linked COF, TpPa-SO_3H, as a benchmark, LFAST produced frameworks with diffraction peaks corresponding to a 350% increase in crystallinity index (CI) relative to prior reports. The same protocol enabled the synthesis of an unreported β-ketoenamine-linked COF-2000 with an even higher structural order. To ensure reproducibility and data accessibility, we further introduce a standardized metadata format encompassing synthesis and PXRD data sets. This data-driven methodology transforms the way that COFs are crystallized and significantly accelerates the pace of materials discovery.



INTRODUCTION

Crystallization is a central challenge in chemistry. The attainment of highly crystalline substances suitable for X-ray diffraction requires a delicate balance between thermodynamic and kinetic processes at crystal nucleation and growth stages. These are dependent on many variables, such as concentration, solvent composition, temperature, time, and additives, whose combinations together form a huge experimental space.[1,2] Typically, conditions are explored in a heuristic manner and are rarely representative of the entire space; therefore, such synthesis condition optimization is often a prolonged trial-and-error process.

Recent studies have leveraged the synergistic use of large language models (LLMs) with a high-throughput liquid-handling robotic synthetic platform to accelerate metal–organic framework (MOF) crystallization condition optimization and discovery.[3−8] However, achieving high-throughput characterization and extending such strategies to covalent organic frameworks (COFs) is much more demanding, both chemically and practically, because 1) covalent bond formation affords less reversibility than metal–ligand coordination, and therefore error correction during crystallization is more limited, making COFs generally more challenging to crystallize than MOFs; 2) many COF precursors are poorly soluble, precluding stock solution preparation as in MOF synthesis and necessitating high-throughput solid handling; and 3) crystallinity cannot be triaged by optical microscopy alone, making the use of high-throughput powder X-ray diffraction (PXRD) characterization essential, yet obtaining a pure phase PXRD pattern requires additional washing and drying preparative steps.

Here, we address these challenges by developing a closed-loop high-throughput LLM For Accelerated Synthesis Technique (LFAST) platform that integrates LLMs with robotic synthesis and high-throughput PXRD feedback to accelerate COF synthesis optimization and discovery (Figure 1). Deep research powered by the GPT-4*o* model is chosen, which can learn from and consolidate the literature conditions and insights. This forms an expanded design space that subsumes published recipes and their chemically reasonable extensions, and then it composes screening conditions that efficiently explore this defined space. The conditions are executed on a robotic platform, and then high-throughput

Received: January 4, 2026
Revised: February 8, 2026
Accepted: February 10, 2026
Published: February 23, 2026

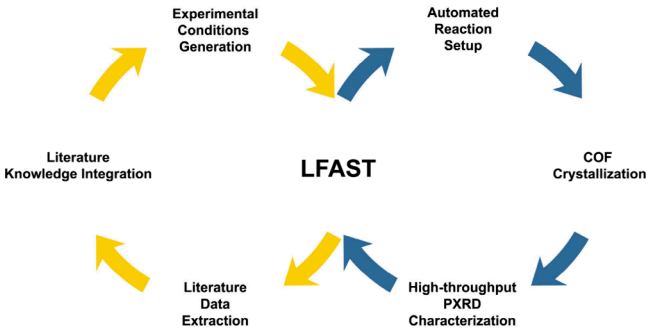


Figure 1. Schematic of the LLM For Accelerated Synthesis Technique (LFAST) platform that couples an *in silico* cycle with high-throughput experimentation. Yellow denotes the *in silico* loop: literature mining, reasoning by an LLM, and condition generation. Blue cycle denotes the experimental loop: high-throughput preparation, incubation, and characterization. Results from the characterization step flow back iteratively to the *in silico* cycle to plan the next screening round.

PXRD is used to provide a quantitative evaluation of the product crystallinity, which informs subsequent optimization rounds of experimental design through deep research. A *β*-ketoenamine-linked COF TpPa-SO_3H[9]—reported in over one hundred studies yet still lacking in high crystallinity—was selected as the benchmark to validate that the closed-loop screening platform can perform crystallinity optimization. On applying the LFAST cycle, we can improve the crystallinity, exceeding the best crystallinity reported to date by 350%. We then apply the same platform to realize a previously unreported *β*-ketoenamine-linked COF, which we name COF-2000, demonstrating that the approach generalizes and extrapolates from optimization of known COFs to *de novo* COF discovery.

RESULTS AND DISCUSSION

LLM Agent for Literature Mining and Screening Reaction Conditions

We established a closed-loop platform that unites literature-informed design, high-throughput synthesis, and PXRD feedback for COF crystallization optimization and discovery. LLMs are generative artificial intelligence (AI) systems pretrained on extensive text corpora and associated metadata, from which it distills and organizes chemical knowledge. This is especially relevant and would be a valuable tool for accelerating chemistry research by consolidating knowledge from the literature, defining an extended relevant synthetic space, and then proposing executable experimental designs under specified laboratory constraints.

In prior studies from our group, GPT-family models have been shown to assist with data mining, materials design, and optimization, yet they rely on a first step of inputting manuscript files from human-selected papers as part of the prompts. Such a manual curation step can introduce selection bias in the number of papers, selection of journals, and types of studies and, in turn, arrives at biased outcomes and can omit important precedents, ultimately leading to potential hallucination.[10] To reduce such bias, we employ GPT-4*o* in an agentic deep research mode. All deep research queries were performed using the GPT-4*o* with deep research option at the time of the study from April 20, 2025, to September 6, 2025. Deep research takes natural language as prompt inputs, retrieves, cross-checks, and synthesizes relevant information from web sources, and then consolidates it into insights. These insights inform the selection of an expanded synthetic space that includes recipes from the literature and any other associated chemically reasonable extensions. This mirrors quite closely how independent researchers would normally embark on a new project, where they often begin by conducting a literature survey to look for domain-specific knowledge.

At the same time, we emphasize that information retrieval using a deep research agent does not eliminate bias. While it can reduce human selection bias by expanding information retrieval beyond a manually curated set of sources, we believe it is still not exhaustive and therefore may not guarantee comprehensive coverage of the relevant literature. In COF synthesis, practical experimental details are often embedded in Supporting Information and other less visible sources and may not be consistently retrievable, since the agent's access depends on what is indexed, accessible at runtime, and machine-readable.

Because deep research is driven by natural language inputs, prompt engineering is pivotal to ensure consistent, high-quality performance that meets scientific research needs. Guidelines for adapting the provided templates to other COF families and laboratory constraints are provided in the Supporting Information Section S3. We set the role at the outset, wherein we first designate the agent as a reticular chemist to put the task into a domain-specific context, followed by stating the scientific objective. We then explicitly impose our high-throughput laboratory configuration so that the experimental design outputs will be commensurate with the 96-well format lab execution.

Having set the roles, we give a broad overview of the scientific task. Two sets of step-by-step prompts were developed for initial screening as well as for consecutive synthesis condition optimization in subsequent batches (Figures S1 and S2). For the generation of synthesis conditions for the first batch, the prompt explicitly requires that an extensive literature investigation be conducted on *β*-ketoenamine-linked COFs and thereafter generation of a table in comma-separated value (*.csv*) format as the first batch screening conditions for high-throughput robotic execution in a 96-well plate format. A step-by-step procedure was then given for the deep research agent to follow to ensure consistency in the chain-of-thought reasoning process, therefore leading to a more stable model performance.

We structure the initial screening prompt as a 10-step workflow (Figure S1). First, we defined a 96-well formatted screening matrix with distinct condition entries. Second, we provided information on the precursors of interest and asked the LLM agent to determine the amount of each precursor to use according to the commonly reported stoichiometric ratios in analogous systems. To ensure the product amount is sufficient for high-throughput PXRD screening, we restricted the theoretical yield to be >5 mg assuming an up to 20% weight loss during the entire sample handling process. In the third step, once the amounts of precursors have been determined, the LLM agent was prompted to determine reaction temperature and time based on the consolidated literature knowledge. Thereafter, reaction solvent selection criteria were defined so that the pure solvents or solvent mixture used for screening would include the literature solvent space as a subset. To expand from the literature-defined solvent space, structurally and physiochemically similar solvents were also explored as potential solvent choices for

https://doi.org/10.1021/jacs.5c23233
J. Am. Chem. Soc. XXXX, XXX, XXX−XXX

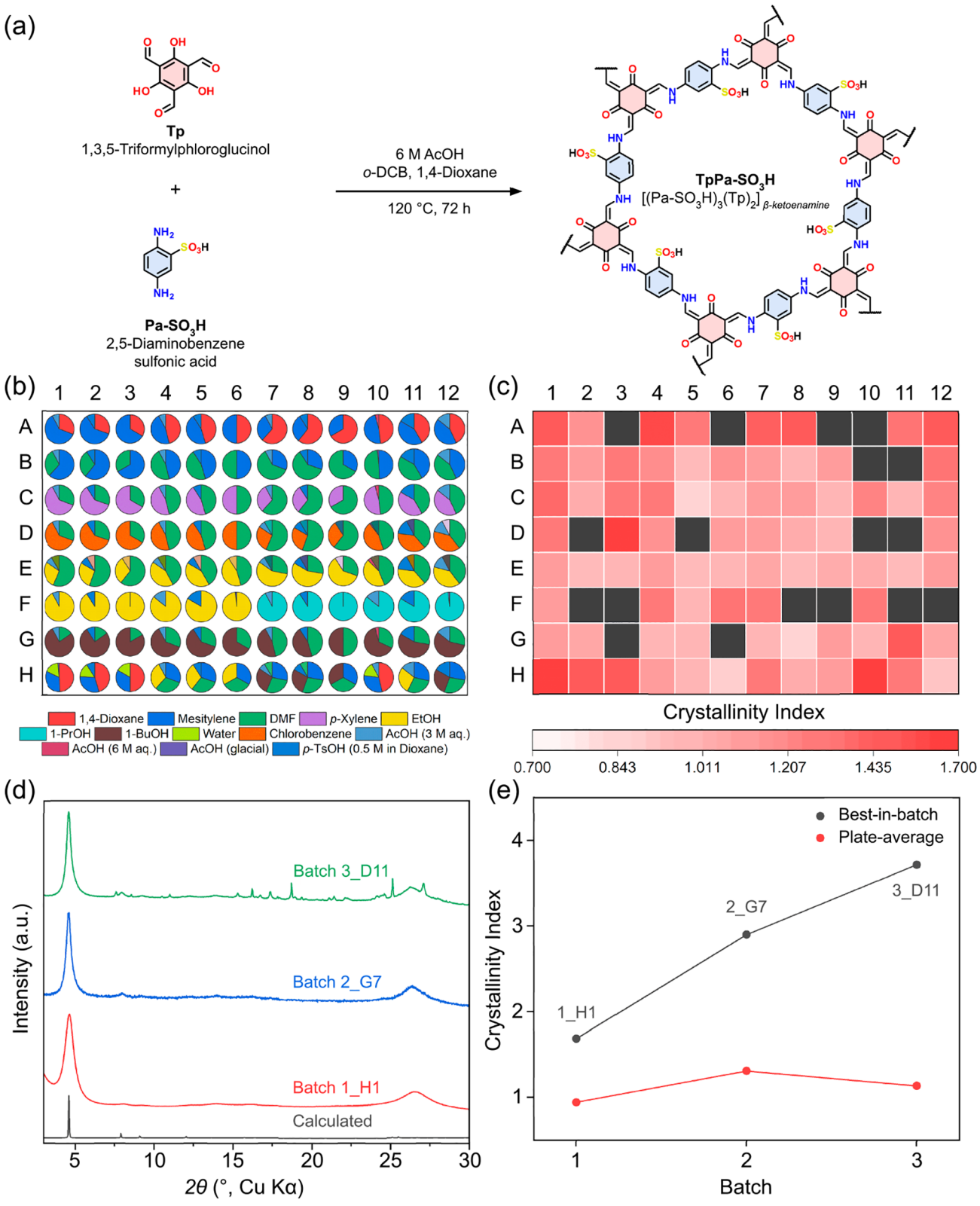


Figure 2. Optimizing TpPa-SO_3H crystallinity using LFAST. (a) The synthesis scheme of TpPa-SO_3H with the optimized crystallization condition. (b) 96-condition screening of solvent and catalyst compositions displayed as pie charts in a well plate layout (A1:H12). The legend below the plate identifies each solvent or catalyst, and slice areas indicate the volumetric fraction. (c) Heat map of CI for the first batch screening; deeper red denotes higher CI values and therefore higher crystallinity, and black squares indicate noncrystalline products. (d) Calculated and representative experimental PXRD patterns from the best condition in each screening batch. (e) The best-in-batch (in black) CI increase steadily from the first to third batch, and the red trace represents plate-average CI.

https://doi.org/10.1021/jacs.5c23233
J. Am. Chem. Soc. XXXX, XXX, XXX−XXX

screening condition generation. As the next step, the catalyst selection criteria were defined in a similar way to those of the solvent selection, except that hygroscopic or overly reactive acids/bases were explicitly prompted to be excluded.

These first five steps define the initial screening space. We then continued to enforce a validation procedure. We first ensured that the total solvent volume adds up to 300 μL and that each liquid aliquot addition is above 10 μL, the lower limit of accurate liquid handling by the robotic platform. We then asked to make sure that all chemicals (powders and solvents) included are safe to handle under the defined reaction conditions. As the eighth step, data documentation was defined in *.csv* format with a specific header that is composed of all reaction variables and an optional note section for providing the rationale of each specific condition. The two final steps include performing a validation of calculation and literature sources and listing all references used in performing the search.

For subsequent batches, the optimization prompt incorporates the prior batch PXRD results as feedback and follows six reasoning steps under the same experimental constraints (Figure S2). To ensure sufficient chemical space is explored in the optimization screening matrix, we direct the LLM agent to allocate 60% of the screening conditions for exploiting local refinement based on the first batch PXRD feedback and 40% for exploration of underrepresented yet plausible chemical space. As we move onto subsequent optimization batches, we progressively tighten the search around the emerging optima by increasing the exploit/explore ratio from 8:2 in batch 3, and to 9:1 in batch 4, providing the right balance necessary for local refinement while retaining a measured degree of exploration.

High-Throughput Synthesis and Characterization Methodology

A high-throughput semiautomated synthesis and characterization platform was established for screening execution. Specifically, 96-well plates with 1 mL vial inserts were used as the reaction vessel for the reagent preparation and synthesis of COFs using solvothermal methods. All sample handling and transfer steps were tracked and standardized across batches for reproducibility.

A typical COF reaction preparation is composed of a first step of solid handling, followed by the addition of solvents and acid/base catalysts. A customized Unchained Labs Junior system was used for dispensing powder and liquid reagents according to the 8 × 12 screening matrices generated by the LLM. Solid reagents were all preground to yield homogeneous particle sizes before loading into 1 dram storage vials with a vibratory dispensing head for small amount gravimetric dispensing. To improve the efficiency of solid dispensing and maximize the potential throughput of the system, solids were dispensed directly into the well plate inside the balance. Tare and measure before and after each dispense allowed for an accurate recording of the actual dispense amounts for every well and thereafter were logged with the corresponding condition identifier (A1:H12). Liquid dispensing was handled by a 1 mL pipet tool with positive displacement tips that are also capable of accurately dispensing difficult liquids with a viscous or volatile nature. A side wall touch off protocol was applied to all aspiration and dispense actions to ensure accurate dispensing, especially with small volume dispenses, as liquid droplets on pipet tips can sometimes be carried over to adjacent vials. The pipet was calibrated for accurate dispensing between 10 and 1000 μL with an absolute deviation of ±2 μL using an eight-point calibration of 10, 25, 75, 100, 250, 500, 750, and 1000 μL.

After the robotic platform has finished the addition of all reagents, the reaction plate was sealed by a layer of perfluoroalkoxy alkane (PFA) film, which has a wide solvent and reagent compatibility, and two layers of silicone rubber mat with nine screws for a tight seal. This ensures that no gas exchange or solvent leakage would happen during the incubation step and that the reaction vessel can withstand the autogenous reaction pressure at elevated temperature. The well plates were then transferred into preheated ovens for incubation.

After reaction, the plates were allowed to cool to room temperature and unsealed. The crude products were then transferred onto polypropylene 96-well filtration plates with pore sizes of 0.45 μm for thorough washing on a vacuum filtration manifold (Figure S14). Each sample was washed by 1.5 mL (300 μL × 5) of *N*,*N*-dimethylformamide (DMF) followed by 1.5 mL (300 μL × 5) of acetone and dried under vacuum for 10 min. The dry powders from the filter plates were then transferred onto a 96-well format customized well plate for high-throughput PXRD measurements. The PXRD plate uses mylar film as substrate, which has a broad feature at 2θ (Cu Kα) = ∼16.2° (Figure S15). The 2θ scan measurements were taken under transmission geometry between 2° and 30° at a scanning speed of 10°/min. Each PXRD pattern was saved as a separate file under the respective condition identifier. For each sample, we quantified crystallinity by using the crystallinity index (CI), defined as the inverse of the fitted full width at half-maximum (FWHM) of the low-angle diffraction peaks. Samples for which no peak was detected or did not pass the sigma cut criterion within the predefined search window were assigned CI = 0 and treated as noncrystalline.

Benchmarking LFAST for Crystallinity Optimization

A challenging β-ketoenamine-linked COF TpPa-SO_3H was chosen to benchmark the performance of the LFAST platform in crystallinity optimization (Figure 2a). TpPa-SO_3H was first reported in 2015 and since then, to the best of our knowledge, has been reported in 118 papers for a wide range of applications, including but not limited to desalination,[11,12] catalysis,[9,13,14] separation,[15−18] sensing,[19] and batteries.[20−23] However, there has been no paper to date that reports this COF with high crystallinity and a sharp PXRD profile.

To achieve highly crystalline TpPa-SO_3H, LFAST was used as aforementioned. Drawing on literature reports,[24−30] the LLM agent arrived at a fixed 2:3 stoichiometric ratio between the two precursors, 1,3,5-triformylphloroglucinol (Tp) and 2,5-diaminobenzenesulfonic acid (Pa-SO_3H), corresponding to a concentration of 3.1 and 4.2 mg, respectively, per 300 μL reaction.

Within the comprehensive deep research report, insights were explicitly provided that span across monomer reactivity, literature protocols, solvent choices, and safety evaluation within the laboratory constraints defined above (Figure S3). First, the deep research agent provided an analysis of the monomer reactivity. It is noteworthy to mention that in addition to the explicit information provided in the references, the LLM also has its own unique chemical insights into the synthesis. It noted that Pa-SO_3H has a sulfonic acid functional

https://doi.org/10.1021/jacs.5c23233
J. Am. Chem. Soc. XXXX, XXX, XXX−XXX

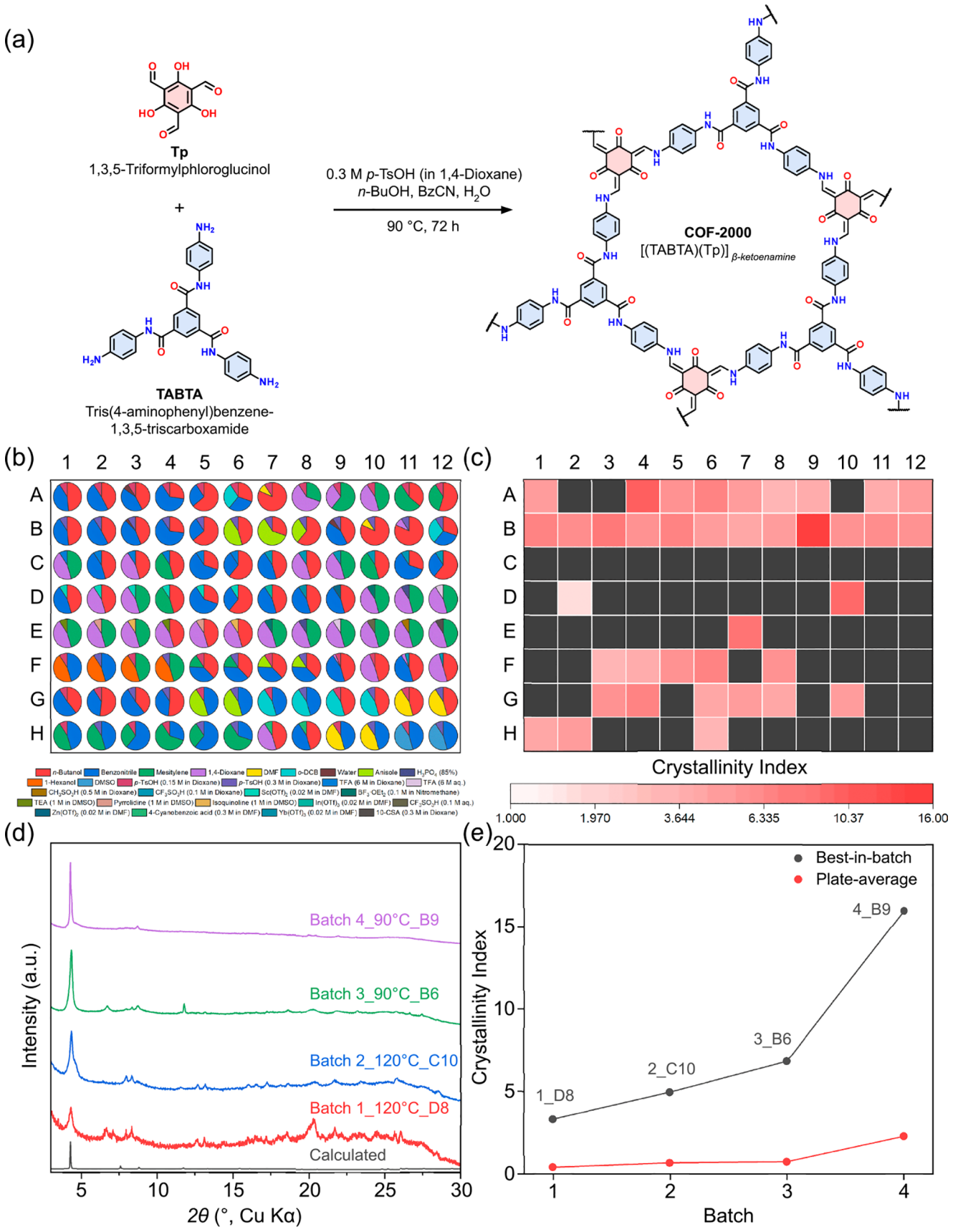


Figure 3. LFAST for the *de novo* discovery of COF-2000. (a) The synthesis scheme of COF-2000 with the optimized crystallization condition. (b) 96-condition screening of solvent and catalyst compositions displayed as pie charts in a well plate layout (A1:H12). The legend below the plate identifies each solvent or catalyst, and slice areas indicate the volumetric fraction. (c) Heat map of CI for the fourth batch screening; deeper red denotes higher CI values and therefore higher crystallinity, and black squares indicate noncrystalline products. (d) Calculated and representative

https://doi.org/10.1021/jacs.5c23233
J. Am. Chem. Soc. XXXX, XXX, XXX−XXX

Figure 3. continued

experimental PXRD patterns from the best condition in each screening batch. (e) Both the best-in-batch (in black) and plate-average (in red) CI increase steadily from the first to fourth batch.

group that could potentially self-catalyze the imine condensation reaction by protonating carbonyl groups on the aldehyde of Tp, but it still acknowledges that most reports on the synthesis of this COF still use Brønsted−Lowry acids as the catalyst. Therefore, the suggested screening matrix systematically explored the catalyst choices. Specifically, no external acid (relying purely on Pa-SO_3H self-catalysis), acetic acid (3 M aq., 6 M aq., glacial), and *p*-toluenesulfonic acid (0.5 M aq.) were suggested in the design. This excluded the use of strong acids that could lead to precipitation of ammonium salt of Pa-SO_3H or bases that could lead to side reactions like deprotonation of phloroglucinol, which is a very reasonable catalyst screening starting point, something even reticular chemists with extensive experience in COF synthesis would consider systematically exploring. As the next step, the LLM agent tried to design the solvent choices so that it balances the kinetics of the imine condensation and crystallization processes. It classified solvents into different categories and adopted various combinations of low-polarity (aromatic organic solvents like mesitylene, xylene, and chlorobenzene) and high-polarity choices (aprotic solvents like 1,4-dioxane, DMF; protic solvents like ethanol, 1-propanol, and 1-butanol). Therefore, the ultimate solvent map would be representative of the entire polarity range of the solvent space (Figure 2b).

The first batch screening was executed and characterized with the high-throughput platform, and the results were compared with the literature PXRD patterns. By using the vision-enabled GPT-4*o* model, the highest CI of the (100) diffraction peak from all reported PXRD patterns is determined to be ∼0.81° (FWHM = 1.23) (Table S3), which is consistent with the pattern obtained with the same PXRD measurement setup from a sample prepared using the literature procedure (Figure S16). From an analysis of the PXRD patterns from the first batch screening, 78 samples among the 96 showed some level of crystallinity ranging from a CI of 0.843 to 1.685, or FWHM values of 1.19° to 0.59° (Figure 2c). To our surprise, reaction well H1 of this first batch of screening already increased the COF crystallinity by 100%, as measured by CI compared with the best reported crystallinity to date.

To assess the reproducibility of this standardized workflow, we set up a full 96-well plate using the best conditions from this batch. The results were very encouraging. The average CI among the 96 wells was 1.54 (FWHM = 0.65°), with a standard deviation of 0.14 (Figure S17 and Table S4). These results not only show reproducibility of the standardized robotic synthetic pipeline but also show that crystallinity is independent of well position on the plate, or in other words, the same reaction run in different wells yields an indistinguishable crystallinity difference, indicating that nonuniform heat transfer across the plate does not introduce positional bias. Specifically, seven diffraction peaks can be clearly identified, which is also indicative of improved crystallinity compared with literature precedents where only three diffraction peaks could be identified (Figure S18). We then organized the CI information from the first batch and used it as a part of the input for the optimization prompt for further narrowing down the optima. A total of 3 batches have been carried out with this iterative process,[24,31−39] and samples with the best crystallinity obtained from each batch have a CI of 1.69, 2.94, and 3.70, showing an additional 119% increase in crystallinity, or FWHM of 0.59°, 0.34°, and 0.27° (Figures 2b−e, Figures S3−7, S19−22). To assess whether the observed crystallinity gains could be achieved by the high-throughput workflow alone, we conducted a baseline study in which we performed synthesis optimization of TpPa-SO_3H using the same robotic platform and PXRD as readouts but with conditions designed through a conventional human literature review without the deep research agent. Across three 96-condition well plates, we obtained a plate-average CI of 1.20 and a best CI of 3.10, whereas LFAST reached a best CI of 3.70 over three batches. Overall, integrating deep research into the high-throughput robotic workflow provides a clear improvement in the efficiency of crystallinity optimization over robotic screening alone.

LFAST for COF Discovery

To test the efficacy of the LFAST platform, we used TpPa-SO_3H as a benchmark material and showed that indeed, using the LFAST cycle, one can reach the highest ever crystallinity reported for this benchmark COF. As a next step, we proposed a new *β*-ketoenamine-linked COF that has not yet been reported in the literature. COF precursors that could adopt many different conformations are typically considered to be more challenging to synthesize. Two linkers, N^1,N^3,N^5-tris(4-aminophenyl)benzene-1,3,5-tricarboxamide (TABTA) and Tp, were used for the synthesis of this novel COF, which we henceforth refer to as COF-2000 (Figure 3a). TABTA contains three amide groups that allow for intrinsic flexibility within its trigonal planar geometry.

Deep research was used for identifying the first batch of conditions in the same way as was applied with TpPa-SO_3H. Importantly, for this optimization, the LLM had no prior literature or precedence to refer to. Therefore, it first started by determining the stoichiometry according to inference from similar literature examples.[24,40−46] It narrows to a 1:1 stoichiometric ratio between Tp and TABTA for the synthesis of COF-2000 by inference from the literature. The reasoning it provides is as follows: "*β*-Ketoenamine COFs form by condensing 1,3,5-triformylphloroglucinol (Tp) with multi-amine linkers, using a ratio that equalizes their functional groups. Tp has three aldehydes per molecule, and TABTA has three amine groups, so an equimolar 1:1 ratio (Tp:TABTA) is typically employed" (Figure S8). Then, it moves on to determining a reasonable precursor concentration to use. For such small-scale high-throughput syntheses, it proposed that 0.01 mmol of each precursor be used for a 300 *μ*L-scale reaction. This corresponds to 2.1 and 4.8 mg of Tp and TABTA, respectively, which meets the constraint of targeting >5 mg when assuming an 80% isolated yield. Immediately after determining the precursor amounts, the LLM agent provided an additional validation focused on yield, acknowledging that achieving high crystallinity can sometimes incur a decrease in yield. However, it correctly pointed out that most literature on *β*-ketoenamine-linked COFs reports high isolated yields in the 70−90% range. These high yields are generally attributed to the low crystallinity reported and irreversible linkage formation

https://doi.org/10.1021/jacs.5c23233
J. Am. Chem. Soc. XXXX, XXX, XXX−XXX

```
data_pxrdif

_exptl_operator 'Yaghi group'                                          experimental
_exptl_pxrd_model 'Rigaku Smartlab'
_exptl_experiment_number 'A7'
_exptl_condition_id 'A7'

_diffrn_radiation_wavelength 1.5406
_diffrn_measurement_scanning_speed 10.0
_diffrn_radiation_power 9.0
_diffrn_measurement_mode 'transmission'

_reaction_time 72.0
_reaction_temperature 100.0

_units_angle 'deg'                                                     units
_units_intensity 'cps'
_units_mass 'mg'
_units_volume 'µL'
_units_concentration 'M'
_units_wavelength 'Å'
_units_scanning_speed 'deg/min'
_units_power 'kW'
_units_time 'hours'
_units_temperature '°C'

loop_                                                                  precursor
_precursor_name
_precursor_cas_number
_precursor_target_mass
_precursor_actual_mass
'1,3,5-Triformylphloroglucinol (Tp)' '34374-88-4' 3.15 3.15
'2,5-Diaminobenzenesulfonic acid (Pa-SO3H)' '88-45-9' 4.23 4.23

loop_                                                                  solvent
_solvent_name
_solvent_volume
'1_4-Dioxane' 100.0
'Mesitylene' 200.0

_catalyst_type 'AcOH'                                                  catalyst
_catalyst_concentration 3.0
_catalyst_volume 22.0

loop_                                                                  PXRD data
_diffrn_reflns_theta
_diffrn_reflns_intensity
2.0 25577.18359375
2.00999999046326 25986.384765625
2.01999998092651 26055.318359375
2.02999997138977 25364.8828125
2.03999996185303 24848.384765625
2.04999995231628 24682.986328125
...
```

Figure 4. Powder X-ray diffraction information file. Example of a structured *.pxrdif* with data section (represented by change in color and label) detailing different metadata (experimental, sample, and unit information) and PXRD data.

following tautomerization. Considering this, the screening uses slightly excess precursor masses to ensure sufficient material for PXRD characterization.

Having determined the precursor amounts, the temperature and time of the reaction were set to 120 °C for 72 h based on commonly reported solvothermal COF synthetic conditions.

https://doi.org/10.1021/jacs.5c23233
J. Am. Chem. Soc. XXXX, XXX, XXX−XXX

The choice of temperature and reaction time is based not only on literature precedent but also on practical considerations. A kinetic study on COFs showed that linkage formation is very fast as compared with crystallization, which can take hours to days.[40] Therefore, extended incubation at elevated temperatures helps enhance long-range order. The 120 °C temperature also increases the saturation vapor pressure of water, which facilitates imine condensation, where water is generated as a byproduct. As the next step, solvents were categorized into classes (nonpolar aprotic, polar aprotic, and polar protic), as in the case of TpPa-SO_3H, to span solvent polarity, hydrogen-bonding capacity, and solubility. Possible Brønsted−Lowry acids and Lewis acids were then proposed as catalyst candidates. This first-batch screening matrix explored a much wider chemical space than in the crystallinity-optimization case: eight different solvents or solvent mixtures were tested along with eight different choices of Brønsted−Lowry and Lewis acid catalysts (Figure S25), which was further expanded to 10 and 18 in the fourth batch (Figure 3b), as compared with the nine solvents and four Brønsted−Lowry acid catalyst choices proposed for the initial batch screening of TpPa-SO_3H.

The first 96-well screening batch yielded 20 crystalline samples out of 96, which was then iteratively improved through LFAST in three subsequent batches. The number of crystalline samples increased from 20 to 27 in the second batch, 33 in the third batch, and to 38 in the fourth batch (Figures 3b and c, Figures S8−13, S25−34). Meanwhile, samples with the highest crystallinity across the four batches improved from CIs of 3.32° to 4.92°, 6.84°, and 15.97, corresponding to FWHMs of 0.30° to 0.20°, 0.15°, and 0.06° (Figures 3d and e). Overall, the crystallinity was improved by 381% across four batches as measured by CI.

The best reaction condition for COF-2000 was found to be a quaternary solvent mixture of water, 1-butanol, and benzonitrile with 0.3 M *p*-toluenesulfonic acid (in 1,4-dioxane) as the catalyst at 90 °C for 72 h. This represents a unique and uncommon combination of solvent and catalyst choices for the synthesis of imine- and *β*-ketoenamine-linked COFs. There is only a handful of literature where ternary solvent mixtures were used for COF synthesis, since the solvent space scales exponentially when moving from binary to ternary mixtures, let alone quaternary mixtures. In the third batch, 80% of the conditions were dedicated to exploitation and the remainder to exploration, and the best condition was found in the exploration trials, where a cosolvent modulation strategy was proposed through literature mining, using nitrile-containing aromatic solvents,[47−49] which enhanced crystallinity and was further optimized in the exploitation conditions in the fourth batch. It is noteworthy to mention that the cited literature provides no mechanistic insight, instead only claiming that the nitrile-based solvents slow the nucleation kinetics. However, the LLM agent adopted this hypothesis because both precursors show increased solubility in benzonitrile, thereby slowing the crystallization kinetics.

To confirm the formation of the *β*-ketoenamine linkage in COF-2000, Fourier-transform infrared (FT-IR) spectroscopy was measured (Figure S35). The FT-IR spectrum of the COF-2000 crystals showed a characteristic C═C stretch at 1634 cm^{-1} and a C−N signal at 1298 cm^{-1}, representative of *β*-ketoenamine linkage formation.[31] Attenuation of the aldehyde stretch at 1634 cm^{-1} from Tp and the amine stretch between 2887 and 3231 cm^{-1} from TABTA was also observed, collectively confirming conversion of the starting materials to the *β*-ketoenamine-linked product. This is further corroborated using ^{13}C solid-state nuclear magnetic resonance (ssNMR) spectroscopy. The obtained ^{13}C cross-polarization magic angle spinning (^{13}C CP-MAS) spectrum showed a prominent peak at 184 ppm that corresponded to the formed carbonyl carbons after tautomerization (Figure S36). Having confirmed the linkage formation, Pawley refinement against the optimized PXRD pattern yielded unit cell parameters of $a = b = 22.031(6)$ Å, $c = 3.510(3)$ Å (profile residual factor R_p = 1.31%, weighted profile residual factor R_{wp} = 2.23%) (Figure S37). We acknowledge that PXRD-based structural analysis of 2D COFs can be nonunique, because different stacking variants and microstructural effects can sometimes yield similar patterns.[50] Therefore, the presented refined structure represents the best-fit model that is consistent with the PXRD pattern and complementary characterization.

Powder X-ray Diffraction Information File

All of the COF data in the literature only include the optimized conditions for crystallization. Failed or poorly crystalline conditions have almost never been reported. In addition, there is also a lack of standardized PXRD raw data, instrument parameters, or complete screening provenance. This in many cases hampers reproducibility and precludes fine-tuning of the LLMs or, in general, AI models.

To address this, we introduced the powder X-ray diffraction information file (*.pxrdif*). This structured file format provides a standardized framework for archiving metadata and experimental data from PXRD studies. The core aim of the format is to ensure that the diffraction data (often meaningless in isolation) are always accompanied by sufficient contextual metadata to allow for reproducibility, interoperability, and data analysis. Without metadata detailing parameters such as the radiation source, scan speed, and sample preparation conditions, even the most precise diffraction patterns lack an interpretive value.

To meet these demands, *.pxrdif* adopts a tag-value architecture inspired by the STAR format.[51] The STAR model supports a flexible yet structured data schema in which each value is paired with a uniquely named data tag identifiable by a preceding underscore character (e.g., _exptl_operator, _reaction_temperature), ensuring that the information remains self-descriptive and machine-readable. This approach accommodates the multidimensional and nested nature of PXRD data, such as multiple precursor inputs, solvent systems, and $\theta/2\theta$ diffraction measurements.

The *.pxrdif* is structured as a sequence of data blocks, each beginning with a data block identifier (e.g., data_pxrdif). Within each block, data items are either defined individually using tag-value pairs or grouped into tabular structures known as loops (Figure 4), which are ideal for storing repetitive or multivariate data, such as theta-intensity pairs from diffraction scans. Each loop begins with a loop_ statement followed by a set of data tags and an ordered list of data values, where each row represents a complete data packet. Regarding the present data, a loop is terminated either by a new data item or by the end of the file, allowing it to accurately represent multidimensional data produced at each point in an experiment.

CONCLUSIONS

Crystallization has long limited the pace of the COF discoveries. Here, we established the LFAST cycle, the LLM For Accelerated Synthesis Technique, that couples literature-

https://doi.org/10.1021/jacs.5c23233
J. Am. Chem. Soc. XXXX, XXX, XXX−XXX

guided design with robotic synthesis and high-throughput PXRD feedback. We demonstrated that the LFAST cycle is not only able to perform crystallinity optimization for TpPa-SO_3H but also can generalize from the literature knowledge to *de novo* discovery. Additionally, we established *.pxrdif*, a machine-readable standard that links full diffraction arrays to synthetic conditions so that all successful and failed results are well-structured and model-ready. These elements collectively accelerate discoveries of new materials from months or years to weeks and provide a reproducible basis for well-documented data-centric reticular chemistry.

■ ASSOCIATED CONTENT

Data Availability Statement

All supplementary files can be accessed in the following GitHub repository: https://github.com/nakulrampal/pxrdif-generator. The complete experimental data sets including all powder X-ray diffraction information files, the reaction condition files for the LFAST experiments, and the human-curated TpPa-SO_3H baseline study are available on Zenodo (DOI: 10.5281/zenodo.17388178).

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/jacs.5c23233.

Detailed experimental procedures, synthesis, and characterization details of the reported compounds, including FT-IR, SEM-EDS, PXRD analysis and refinement, solid state NMR, and TGA (PDF)

Accession Codes

Deposition Number 2527756 contains the supplementary crystallographic data for this paper. These data can be obtained free of charge via the joint Cambridge Crystallographic Data Centre (CCDC) and Fachinformationszentrum Karlsruhe Access Structures service.

■ AUTHOR INFORMATION

Corresponding Author

Omar M. Yaghi − *Department of Chemistry, Kavli Energy NanoScience Institute, Bakar Institute of Digital Materials for the Planet, College of Computing, Data Science, and Society, and Department of Materials Science and Engineering, 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

Kaiyu Wang − *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/0000-0003-2464-2828

Daehyun Daniel Ahn − *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*

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;* orcid.org/0000-0002-6187-5631

Jackson Thomassian − *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*

Neda S. Sabeva − *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/0009-0009-1040-8927

Om Kannan − *Bakar Institute of Digital Materials for the Planet, College of Computing, Data Science, and Society, University of California, Berkeley, California 94720, 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 Mathematics, Department of Statistics, and School of Information, University of California, Berkeley, California 94720, United States*

Complete contact information is available at: https://pubs.acs.org/10.1021/jacs.5c23233

Funding

This research is supported by the Bakar Institute of Digital Materials for the Planet (BIDMaP), Atoco Inc., and the King Abdulaziz City for Science and Technology (KACST).

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We acknowledge Hasan Celik, Raynald Giovine, and UC Berkeley Pines Magnetic Resonance Center's Core NMR Facility for spectroscopic assistance. The solid-state NMR instrument used in material screening is supported by the National Science Foundation under Grant No. 2018784. K.W. would like to thank Zichao Rong and Dr. Ali H. Alawadhi for helpful discussions. This study is also supported by the Kavli Discovery Fellowship (K.W.) and the Bakar Institute of Digital Materials for the Planet (BIDMaP) Emerging Scholars Program (N.R.).

■ REFERENCES

(1) Spingler, B.; Schnidrig, S.; Todorova, T.; Wild, F. Some thoughts about the single crystal growth of small molecules. *CrystEngComm* **2012**, *14* (3), 751−757.

(2) Van Der Sluis, P.; Hezemans, A. M. F.; Kroon, J. Crystallization of low-molecular-weight organic compounds for X-ray crystallography. *J. Appl. Crystallogr.* **1989**, *22* (4), 340−344.

(3) Zheng, Z.; He, Z.; Khattab, O.; Rampal, N.; Zaharia, M. A.; Borgs, C.; Chayes, J. T.; Yaghi, O. M. Image and data mining in reticular chemistry powered by GPT-4V. *Digital Discovery* **2024**, *3* (3), 491−501.

(4) Zheng, Z.; Rampal, N.; Inizan, T. J.; Borgs, C.; Chayes, J. T.; Yaghi, O. M. Large language models for reticular chemistry. *Nature Reviews Materials* **2025**, *10* (5), 369−381.
(5) Zheng, Z.; Alawadhi, A. H.; Chheda, S.; Neumann, S. E.; Rampal, N.; Liu, S.; Nguyen, H. L.; Lin, Y.-h.; Rong, Z.; Siepmann, J. I.; Gagliardi, L.; Anandkumar, A.; Borgs, C.; Chayes, J. T.; Yaghi, O. M. Shaping the Water-Harvesting Behavior of Metal−Organic Frameworks Aided by Fine-Tuned GPT Models. *J. Am. Chem. Soc.* **2023**, *145* (51), 28284−28295.
(6) 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* (46), No. e202311983.
(7) 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 Central Science* **2023**, *9* (11), 2161−2170.
(8) Zheng, Z.; Zhang, O.; Borgs, C.; Chayes, J. T.; Yaghi, O. M. ChatGPT Chemistry Assistant for Text Mining and the Prediction of MOF Synthesis. *J. Am. Chem. Soc.* **2023**, *145* (32), 18048−18062.
(9) Peng, Y.; Hu, Z.; Gao, Y.; Yuan, D.; Kang, Z.; Qian, Y.; Yan, N.; Zhao, D. Synthesis of a Sulfonated Two-Dimensional Covalent Organic Framework as an Efficient Solid Acid Catalyst for Biobased Chemical Conversion. *ChemSusChem* **2015**, *8* (19), 3208−3212.
(10) Sharma, M.; Tong, M.; Korbak, T.; Duvenaud, D.; Askell, A.; Bowman, S. R.; Durmus, E.; Hatfield-Dodds, Z.; Johnston, S. R.; Kravec, S.; Maxwell, T.; McCandlish, S.; Ndousse, K.; Rausch, O.; Schiefer, N.; Yan, D.; Zhang, M.; Perez, E. Towards Understanding Sycophancy in Language Models. *In ICLR 2024* **2024**, DOI: 10.48550/arXiv.2310.13548.
(11) Shen, J.; Yuan, J.; Shi, B.; You, X.; Ding, R.; Zhang, T.; Zhang, Y.; Deng, Y.; Guan, J.; Long, M.; Zheng, Y.; Zhang, R.; Wu, H.; Jiang, Z. Homointerface covalent organic framework membranes for efficient desalination. *Journal of Materials Chemistry A* **2021**, *9* (40), 23178−23187.
(12) Wei, X.; Liu, Y.; Zhao, F.; Wang, T.; Li, Z.; Fan, C.; Yang, Y.; Wang, Y.; Jiang, Z. Covalent organic framework membrane with hourglass-shaped nanochannels for ultrafast desalination. *Nat. Commun.* **2025**, *16* (1), 8125.
(13) Zhu, Z.; Zhu, Y.; Ren, Z.; Liu, D.; Yue, F.; Sheng, D.; Shao, P.; Huang, X.; Feng, X.; Yin, A.-X.; Xie, J.; Wang, B. Covalent Organic Framework Ionomer Steering the CO2 Electroreduction Pathway on Cu at Industrial-Grade Current Density. *J. Am. Chem. Soc.* **2024**, *146* (2), 1572−1579.
(14) Zhang, J.; Zhang, X.; Shen, Y.; Fu, B.; Wu, Y.; Kang, J.; Chen, S.; Wang, G.; Zhang, H.; Yin, H.; Zhao, H. Joule-Heated Interfacial Catalysis for Advanced Electrified Esterification with High Conversion and Energy Efficiency. *Adv. Mater.* **2025**, *37* (4), No. 2413949.
(15) Guo, Z.; Li, W.; Wu, H.; Cao, L.; Song, S.; Ma, X.; Shi, J.; Ren, Y.; Huang, T.; Li, Y.; Jiang, Z. Reverse filling approach to mixed matrix covalent organic framework membranes for gas separation. *Nat. Commun.* **2025**, *16* (1), 3617.
(16) Wang, H.; Zhai, Y.; Li, Y.; Cao, Y.; Shi, B.; Li, R.; Zhu, Z.; Jiang, H.; Guo, Z.; Wang, M.; Chen, L.; Liu, Y.; Zhou, K.-G.; Pan, F.; Jiang, Z. Covalent organic framework membranes for efficient separation of monovalent cations. *Nat. Commun.* **2022**, *13* (1), 7123.
(17) Ying, Y.; Peh, S. B.; Yang, H.; Yang, Z.; Zhao, D. Ultrathin Covalent Organic Framework Membranes via a Multi-Interfacial Engineering Strategy for Gas Separation. *Adv. Mater.* **2022**, *34* (25), No. 2104946.
(18) Shen, J.; Zhang, R.; Su, Y.; Shi, B.; You, X.; Guo, W.; Ma, Y.; Yuan, J.; Wang, F.; Jiang, Z. Polydopamine-modulated covalent organic framework membranes for molecular separation. *Journal of Materials Chemistry A* **2019**, *7* (30), 18063−18071.
(19) Krishnaveni, V.; Dmello, M. E.; Thokala, N.; Bonda, A.; Sriramadasu, V. K.; Akkenapally, S.; Bubanale, S. B.; Basavaiah, K.; Bhattacharyya, S.; Kalidindi, S. B. Catalytic Gas-Driven Exfoliation of Sulfonic Acid-Linked Covalent Organic Framework Nanostructures for Hydrogen Sensing. *ACS Applied Nano Materials* **2024**, *7* (20), 23416−23422.
(20) Xu, J.; An, S.; Song, X.; Cao, Y.; Wang, N.; Qiu, X.; Zhang, Y.; Chen, J.; Duan, X.; Huang, J.; Li, W.; Wang, Y. Towards High Performance Li−S Batteries via Sulfonate-Rich COF-Modified Separator. *Adv. Mater.* **2021**, *33* (49), No. 2105178.
(21) Huang, J.; Cheng, L.; Zhang, Z.; Li, C.; Bang, K.-T.; Liem, A.; Luo, H.; Hu, C.; Lee, Y. M.; Lu, Y.; Wang, Y.; Kim, Y. High-Performance All-Solid-State Lithium Metal Batteries Enabled by Ionic Covalent Organic Framework Composites. *Adv. Energy Mater.* **2024**, *14* (27), No. 2400762.
(22) Yu, G.; Cui, Y.; Lin, S.; Liu, R.; Liu, S.; Zhu, Y.; Wu, D. Ultrathin Composite Separator Based on Lithiated COF Nanosheet for High Stability Lithium Metal Batteries. *Adv. Funct. Mater.* **2024**, *34* (24), No. 2314935.
(23) Liu, J.; Guo, K.; Guo, W.; Chang, J.; Li, Y.; Bao, F. Superconjugated Anthraquinone Carbonyl-Based Covalent Organic Framework as Anode Material for High-Performance Aqueous Ammonium-Ion Batteries. *Angew. Chem., Int. Ed.* **2025**, *64* (14), No. e202424494.
(24) Jaryal, R.; Khullar, S.; Kumar, R. Effect of solvent and acid on the morphology of the β-ketoenamine-linked covalent organic frameworks (COFs). *Materials Today: Proceedings* **2023**, *78*, 885−890.
(25) Das, G.; Balaji Shinde, D.; Kandambeth, S.; Biswal, B. P.; Banerjee, R. Mechanosynthesis of imine, β-ketoenamine, and hydrogen-bonded imine-linked covalent organic frameworks using liquid-assisted grinding. *Chem. Commun.* **2014**, *50* (84), 12615−12618.
(26) Pachfule, P.; Kandmabeth, S.; Mallick, A.; Banerjee, R. Hollow tubular porous covalent organic framework (COF) nanostructures. *Chem. Commun.* **2015**, *51* (58), 11717−11720.
(27) Dang, M.; Deng, Q.-L.; Tian, Y.-Y.; Liu, C.; Shi, H.-P.; Fang, G.-Z.; Wang, S. Synthesis of anionic ionic liquids@TpBd-(SO3)2 for the selective adsorption of cationic dyes with superior capacity. *RSC Adv.* **2020**, *10* (9), 5443−5453.
(28) Jin, Y.-H.; Li, M.-H.; Yang, Y.-W. Covalent Organic Frameworks for Membrane Separation. *Advanced Science* **2025**, *12* (5), No. 2412600.
(29) Li, W.; Wang, Q.; Cui, F.; Jiang, G. Covalent organic framework with sulfonic acid functional groups for visible light-driven CO2 reduction. *RSC Adv.* **2022**, *12* (28), 17984−17989.
(30) Bourda, L.; Krishnaraj, C.; Van Der Voort, P.; Van Hecke, K. Conquering the crystallinity conundrum: efforts to increase quality of covalent organic frameworks. *Materials Advances* **2021**, *2* (9), 2811−2845.
(31) Kandambeth, S.; Mallick, A.; Lukose, B.; Mane, M. V.; Heine, T.; Banerjee, R. Construction of Crystalline 2D Covalent Organic Frameworks with Remarkable Chemical (Acid/Base) Stability via a Combined Reversible and Irreversible Route. *J. Am. Chem. Soc.* **2012**, *134* (48), 19524−19527.
(32) Liu, X.-H.; Guan, C.-Z.; Ding, S.-Y.; Wang, W.; Yan, H.-J.; Wang, D.; Wan, L.-J. On-Surface Synthesis of Single-Layered Two-Dimensional Covalent Organic Frameworks via Solid−Vapor Interface Reactions. *J. Am. Chem. Soc.* **2013**, *135* (28), 10470−10474.
(33) Uribe-Romo, F. J.; Doonan, C. J.; Furukawa, H.; Oisaki, K.; Yaghi, O. M. Crystalline Covalent Organic Frameworks with Hydrazone Linkages. *J. Am. Chem. Soc.* **2011**, *133* (30), 11478−11481.
(34) Matsumoto, M.; Dasari, R. R.; Ji, W.; Feriante, C. H.; Parker, T. C.; Marder, S. R.; Dichtel, W. R. Rapid, Low Temperature Formation of Imine-Linked Covalent Organic Frameworks Catalyzed by Metal Triflates. *J. Am. Chem. Soc.* **2017**, *139* (14), 4999−5002.
(35) van Voorthuizen, T. A.; van der Veen, M. A.; de Smet, L. C. P. M.; Smulders, M. M. J. On the Importance of Balancing the pKa of the Additive in β-Ketoenamine COF Synthesis. *Chemistry − A European Journal* **2025**, *31* (39), No. e202501512.
(36) Li, Y.; Liu, M.; Wu, J.; Li, J.; Yu, X.; Zhang, Q. Highly stable β-ketoenamine-based covalent organic frameworks (COFs): synthesis

and optoelectrical applications. *Frontiers of Optoelectronics* **2022**, *15* (1), 38.
(37) Gruber, C. G.; Frey, L.; Guntermann, R.; Medina, D. D.; Cortés, E. Early stages of covalent organic framework formation imaged in operando. *Nature* **2024**, *630* (8018), 872−877.
(38) Qing, Q.; Luo, J.; Liu, S.; Wang, J.; Wang, Z.; Xiong, X.-G.; Chen, J.; Lu, Y. General synthesis of covalent organic frameworks under ambient condition within minutes via microplasma electrochemistry approach. *Nat. Commun.* **2025**, *16* (1), 2571.
(39) Prieto, T.; Ponte, C.; Guntermann, R.; Medina, D. D.; Salonen, L. M. Synthetic Strategies to Extended Aromatic Covalent Organic Frameworks. *Chemistry − A European Journal* **2024**, *30* (45), No. e202401344.
(40) Laemont, A.; Matthys, G.; Lavendomme, R.; Van Der Voort, P. Mild and Scalable Conditions for the Solvothermal Synthesis of Imine-Linked Covalent Organic Frameworks. *Angew. Chem., Int. Ed. Engl.* **2024**, *63* (51), No. e202412420.
(41) Li, Y.; Chen, W.; Gao, R.; Zhao, Z.; Zhang, T.; Xing, G.; Chen, L. 2D covalent organic frameworks with built-in amide active sites for efficient heterogeneous catalysis. *Chem. Commun.* **2019**, *55* (96), 14538−14541.
(42) Vardhan, H.; Rummer, G.; Deng, A.; Ma, S. Large-Scale Synthesis of Covalent Organic Frameworks: Challenges and Opportunities. *Membranes* **2023**, *13* (8), 696.
(43) Niu, F.; Shao, Z.-W.; Zhu, J.-L.; Tao, L.-M.; Ding, Y. Structural evolution of imine-linked covalent organic frameworks and their NH3 sensing performance. *Journal of Materials Chemistry C* **2021**, *9* (27), 8562−8569.
(44) Ji, W.; Kim, D. M.; Posson, B. M.; Carlson, K. J.; Chew, A. C.; Chew, A. J.; Hossain, M.; Mojica, A. F.; Ottoes, S. M.; Tran, D. V.; Greenberg, M. W.; Hamachi, L. S. COF-300 synthesis and colloidal stabilization with substituted benzoic acids. *RSC Adv.* **2023**, *13* (21), 14484−14493.
(45) Matsumoto, M.; Valentino, L.; Stiehl, G. M.; Balch, H. B.; Corcos, A. R.; Wang, F.; Ralph, D. C.; Mariñas, B. J.; Dichtel, W. R. Lewis-Acid-Catalyzed Interfacial Polymerization of Covalent Organic Framework Films. *Chem.* **2018**, *4* (2), 308−317.
(46) Kang, C.; Yang, K.; Zhang, Z.; Usadi, A. K.; Calabro, D. C.; Baugh, L. S.; Wang, Y.; Jiang, J.; Zou, X.; Huang, Z.; Zhao, D. Growing single crystals of two-dimensional covalent organic frameworks enabled by intermediate tracing study. *Nat. Commun.* **2022**, *13* (1), 1370.
(47) Ji, W.; Hamachi, L. S.; Natraj, A.; Flanders, N. C.; Li, Rebecca L.; Chen, L. X.; Dichtel, W. R. Solvothermal depolymerization and recrystallization of imine-linked two-dimensional covalent organic frameworks. *Chemical Science* **2021**, *12* (48), 16014−16022.
(48) Smith, B. J.; Parent, L. R.; Overholts, A. C.; Beaucage, P. A.; Bisbey, R. P.; Chavez, A. D.; Hwang, N.; Park, C.; Evans, A. M.; Gianneschi, N. C.; Dichtel, W. R. Colloidal Covalent Organic Frameworks. *ACS Central Science* **2017**, *3* (1), 58−65.
(49) Li, R. L.; Flanders, N. C.; Evans, A. M.; Ji, W.; Castano, I.; Chen, L. X.; Gianneschi, N. C.; Dichtel, W. R. Controlled growth of imine-linked two-dimensional covalent organic framework nanoparticles. *Chemical Science* **2019**, *10* (13), 3796−3801.
(50) Van Gele, S.; Bette, S.; Lotsch, B. V. The devil is in the details: pitfalls and ambiguities in the analysis of X-ray powder diffraction data of 2D covalent organic frameworks. *Journal of the American Chemical Society Au* **2025**, *5* (1), 388−398.
(51) Hall, S. R. The STAR file: a new format for electronic data transfer and archiving. *J. Chem. Inf. Comput. Sci.* **1991**, *31* (2), 326−333.

https://doi.org/10.1021/jacs.5c23233
J. Am. Chem. Soc. XXXX, XXX, XXX−XXX