AI-guided experimental design of zirconium MOPs with The World Avatar for sustainable photocatalysis

Yong Ren Tan [*1] **Kumaran Elumalai** [*1 2] **Simon D. Rihm** [3] **Patrick Butler** [4] **Jethro Akroyd** [1 4]
Tej S. Choksi [1 2] **Rong Xu** [1 2] **Markus Kraft** [1 3 5 4]

[*]Equal contribution [1]*CARES, Cambridge Centre for Advanced Research and Education in Singapore, 138602 Singapore* [2]*School of Chemistry, Chemical Engineering and Biotechnology, Nanyang Technological University, 62 Nanyang Drive, Singapore 637459, Singapore* [3]*CMPG, GRIPS – Gründerinnenzentrum Pirmasens, Delaware Avenue 1–3, 66953, Pirmasens, Germany* [4]*Department of Chemical Engineering and Biotechnology, University of Cambridge, Cambridge CB3 0AS, United Kingdom* [5]*MIT, Chemical Engineering, 77 Massachusetts Avenue, Cambridge, MA 02139 USA*. Correspondence to: M. Kraft mk306@cam.ac.uk.

1. Introduction

Metal-organic polyhedra (MOPs) are a class of materials characterised by self-assembled, cage-like discrete architectures constructed from metal centres and organic-based chemical building units (CBUs) [1, 2]. Due to their unique structure and chemical properties, they can enable wide-ranging applications in water purification, gas separation, drug delivery, and catalysis [1, 2]. For a sustainable future, they have been increasingly explored in decarbonisation processes to reduce atmospheric CO_2 through capture, separation, and conversion into commercially valuable chemicals, *e.g.* , methane, methanol, formic acid, and ethanol [3, 4, 5]. Notably, the recent awarding of the 2025 Nobel Prize in Chemistry highlights the transformative impact of reticular chemistry, recognising breakthroughs in understanding and controlling the self-assembly principles that govern MOP formation and function [6].

Following the first synthesis of zirconium-based metal–organic frameworks (MOFs) in 2008, a diverse array of Zr-based MOFs have been synthesised and investigated for photocatalytic CO_2 conversion [7]. However, the synthesis of MOPs (and MOFs) remains profoundly influenced by subtle synthetic conditions and the intricate coordination preferences of metal ions, rendering the field heavily reliant on empirical, trial-and-error methodologies [8, 1, 2]. Despite MOPs (and MOFs) structures being theoretically tunable, the experimental procedures for MOPs and MOFs synthesis remain a challenge for codification. This is because experimental techniques is often embedded within the experience of the chemist, encompassing manual processes, conditional adjustments, and contextual decision-making. These are seldom formally reported in the literature. Critical parameters such as solvent effects, protocols, and *in-situ* modifications constitute a hidden knowledge layer that typical computational representation is lacking. This creates a bottleneck for rational design of MOPs (and MOFs) and automated scaling up of MOPs (and MOFs) synthesis.

The **purpose of this work** is to illustrate a *knowledge-first* AI framework that generates detailed and functional synthetic procedure of functional MOPs synthesis for new MOPs structures. Specifically, the knowledge graph (KG)-based architecture powered by The World Avatar (TWA) enables autonomous discovery of Zr-MOPs by systematically representing the entire synthesis pipeline, *i.e.* from CBU selection, assembly model prediction to procedural optimisation. By encoding expert knowledge, experimental protocols, and chemical reasoning into a machine-actionable semantic layer, we generated and experimentally validated synthesis procedures for new MOP structures, accelerating the development of next-generation reticular materials.

2. Background

2.1 The World Avatar

The vision of The World Avatar (TWA) is to investigate how formalised knowledge models can create a comprehensive universal knowledge representation spanning all possible concepts, powered by a self-contained system of autonomous semantic computational agents that continuously maintain and expand the knowledge representation. In this context, 'world' refers not to planet Earth, but to the entire spectrum of conceivable ideas [9].

In reticular chemistry, TWA addresses the challenge of MOP discovery *via* the OntoMOPs ontology, which formally captures relationships between CBUs, generic building units (GBUs), and assembly models. Starting from 151 experimentally verified MOPs, TWA's reasoning algorithms predicted 1,418 rationally designed structures, which were later expanded by 799 new MOPs through combinatorial assembly models, creating a comprehensive repository of over 2,200 candidates [10]. These structures are semantically enriched with geometry data and linked to provenance information, enabling automated screening and synthesis planning. A pipeline to compute MOPs structures at an increasing level of theory has also been developed with the potential to predict MOPs crystal structures based on their geometries [11]. Notably, large language models (LLMs) are integrated within this pipeline to accelerate the extraction of synthesis protocols from literature, bridging unstructured data with structured KGs [11].

2.2 AI MOPs design

Molecular engineering, *i.e.* , the rational design of molecules for programmed assembly into functional

materials, has demonstrated strong multi-domain innovation potential. For instance, high-throughput screening of MOFs for efficient separation of CO_2 from CO_2-rich natural gas has been achieved through integration of machine learning with Grand Canonical Monte Carlo simulations, which can enable rapid identification of high-performance materials [12, 13]. This has demonstrated the potential of AI in novel MOPs (and MOFs) discovery.

The key advancement of this work lies in the complete ontologisation of the discovery pipeline through TWA: from chemical building units and assembly models to complex synthesis protocols. The semantic framework of TWA captures expert knowledge rarely codified in literature. For example, solvent effects, ageing conditions, and *in-situ* modifications. This enable AI agents to reason across both chemical and procedural domains, thereby bridging the gap between computational design and experimental execution.

3. Methodology

Automated generation of MOP-1 synthesis procedure. A novel MOP synthesis procedure (MOP-1) was dynamically generated within TWA *via* an analogy-based algorithm leveraging literature-extracted protocols instantiated with LLMs (see Figure 1 for the illustration of prediction framework) [11]. Formalised TWA representations enabled mapping of ≈800,000 MOP structures and genetic algorithm optimisation of CO_2 interaction energies (estimated *via* ML simulations), yielding structures that outperform benchmarks within few generations [14]. TWA automatically generate end-to-end workflow for new MOP discovery, from quaternion-based alignment maps CBU binding sites to GBU positions, producing validated 3D geometries.

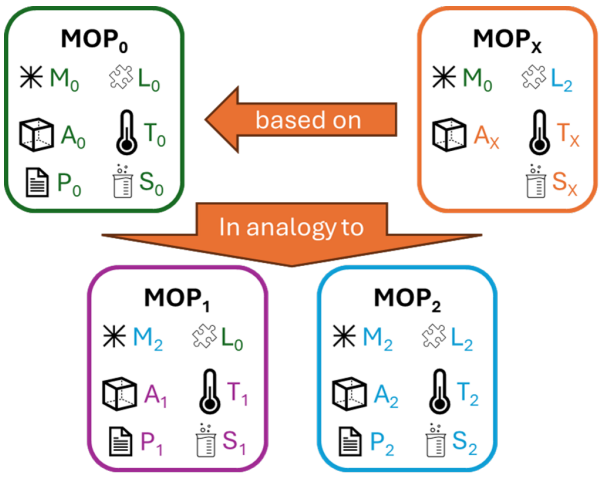

Fig. 1: Prediction framework of synthesis procedures based on analogy to known materials.

Synthesis of MOP-1. Zirconium dichloride ($ZrCl_2$, 100 mg, 0.34 mmol) and 4,4'-(ethyne-1,2-diyl)dibenzoic acid (H_2EDBA, 25 mg, 0.094 mmol) were suspended in N,N-Dimethylformamide (DMF, 10 mL) and distilled water (H_2O, 0.5 mL). The reaction mixture was heated at 60°C for 5 min until it became homogeneous, and then maintained at 40°C for 20 h. The resulting precipitate was collected by centrifugation (5,000 rpm, 6 min), and the supernatant was decanted. The solid was washed by resuspension in DMF (10 mL), followed by centrifugation and decantation. This washing procedure was repeated three times with dichloromethane (CH_2Cl_2, 10 mL), and the off-white solid (yield 45 mg, 38%) obtained was dried under vacuum overnight at room temperature.

Characterisation of MOP-1. Phase purity and crystallinity were confirmed by powder X-ray diffraction (pXRD) using a Bruker D2 Phaser diffractometer equipped with a LYNXEYE linear position-sensitive detector and Ni-filtered Cu Kα radiation (λ = 1.5406 Å). Diffraction patterns were collected over a 2θ range of 3—40° with a step size of 0.02° and a counting time of 1 s per step. Samples were prepared as flat plate specimens on zero-background silicon holders.

4. Results and discussion

The TWA-generated synthesis protocol for MOP-1, was successfully validated through experimental realisation of the target material, demonstrating TWA's predictive capability. Figure 2 shows a comparative analysis of the theoretically predicted XRD pattern and experimentally obtained pXRD pattern. Both patterns exhibit sharp, well-defined Bragg diffraction maxima, indicating decent crystallinity and long-range structural order in the synthesised product. The experimental pXRD pattern shows pXRD peaks at *c.a.* , 9°, 11° and 16° (2θ), which exhibit correlation with the corresponding predicted XRD. This confirms the successful synthesis of the desired MOP topology. There is slight variations in relative peak intensities are attributed to preferred orientation effects common in MOP powders.

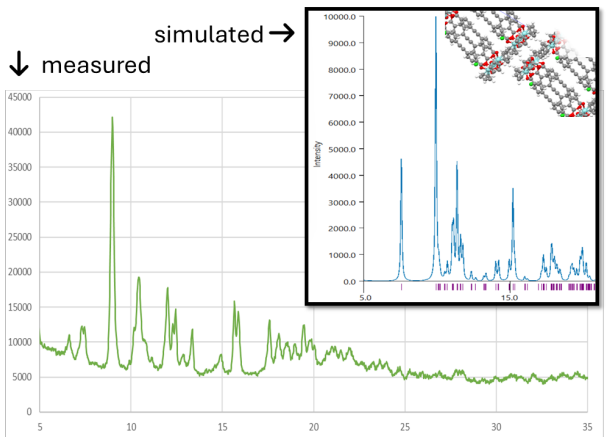

Fig. 2: Comparison of (top right) simulated XRD pattern for the TWA-predicted crystal structure (*i.e.* , MOP-1) and (bottom left) measured pXRD pattern. Difference score = 0.09.

Acknowledgments

This research was part of the Sustainable Manufacture of Molecules and Materials in Singapore (SM_3) programme, supported by the National Research Foundation, Prime Minister's Office, Singapore under its Campus for Research Excellence and Technological Enterprise (CREATE) programme.

References

[1] Adedapo O. Adeola, Joshua O. Ighalo, Pannan I. Kyesmen, and Philiswa N. Nomngongo. Metal-organic polyhedra (MOPs) as emerging class of metal-organic frameworks for CO_2 photocatalytic conversions: Current trends and future outlook. *Journal of CO2 Utilization*, 80:102664, 2024.

[2] Aeri J. Gosselin, Casey A. Rowland, and Eric D. Bloch. Permanently microporous metal–organic polyhedra. *Chemical Reviews*, 120(16):8987–9014, 2020.

[3] Andrea Álvarez, Atul Bansode, Atsushi Urakawa, Anastasiya V. Bavykina, Tim A. Wezendonk, Michiel Makkee, Jorge Gascon, and Freek Kapteijn. Challenges in the greener production of formates/formic acid, methanol, and DME by heterogeneously catalyzed CO_2 hydrogenation processes. *Chemical Reviews*, 117(14):9804–9838, 2017.

[4] Kim Larmier, Wei-Chih Liao, Shohei Tada, Erwin Lam, René Verel, Atul Bansode, Atsushi Urakawa, Aleix Comas-Vives, and Christophe Copéret. CO_2-to-methanol hydrogenation on zirconia-supported copper nanoparticles: Reaction intermediates and the role of the metal–support interface. *Angewandte Chemie International Edition*, 56(9):2318–2323, 2017.

[5] Shuxing Bai, Qi Shao, Pengtang Wang, Qiguang Dai, Xingyi Wang, and Xiaoqing Huang. Highly active and selective hydrogenation of CO_2 to ethanol by ordered Pd–Cu nanoparticles. *Journal of the American Chemical Society*, 139(20):6827–6830, 2017.

[6] Joan Serrano-Plana. A framework for chemists. *Nature Chemistry*, 17(12):1802, 2025.

[7] Jasmina Hafizovic Cavka, Søren Jakobsen, Unni Olsbye, Nathalie Guillou, Carlo Lamberti, Silvia Bordiga, and Karl Petter Lillerud. A new zirconium inorganic building brick forming metal organic frameworks with exceptional stability. *Journal of the American Chemical Society*, 130(42):13850–13851, 2008.

[8] Nadia Gholampour, Yingrui Zhao, François Devred, Capucine Sassoye, Sandra Casale, and Damien P. Debecker. CO_2 methanation over cobalt nanoparticles embedded in ZIF-L–derived porous carbon. *ChemCatChem*, 15(5):e202201338, 2023.

[9] Yong Ren Tan, Markus Hofmeister, Shin Zert Phua, George Brownbridge, Kushagar Rustagi, Jethro Akroyd, Sebastian Mosbach, Amit Bhave, and Markus Kraft. Beyond connected digital twins – can digital twins really deliver sustainable cities? *Sustainable Cities and Society*, 131:106596, 2025.

[10] Aleksandar Kondinski, Angiras Menon, Daniel Nurkowski, Faris Farazi, Sebastian Mosbach, Jethro Akroyd, and Markus Kraft. Automated rational design of metal–organic polyhedra. *Journal of the American Chemical Society*, 144(26):11713–11728, 2022.

[11] Simon D. Rihm, Fabio Saluz, Aleksandar Kondinski, Jiaru Bai, Patrick W. V. Butler, Sebastian Mosbach, Jethro Akroyd, and Markus Kraft. Extraction of chemical synthesis information using the World Avatar. *Digital Discovery*, 4:2893–2909, 2025.

[12] Michael Baldea, Linda J. Broadbelt, Marianthi G. Ierapetritou, Thomas A. Kwan, Chunzhong Li, Zheng-Hong Luo, Xinbin Ma, Massimo Morbidelli, Kirti C. Sahu, Aaron M. Scurto, Zachary P. Smith, Huanting Wang, Feng-Shou Xiao, Victor M. Zavala, and Dan Zhao. 2024 in Retrospective: Trends in chemical engineering. *Industrial & Engineering Chemistry Research*, 64(24):11615–11623, 2025.

[13] Yinjie Zhou, Sibei Ji, Songyang He, Wei Fan, Liang Zan, Li Zhou, Xu Ji, and Ge He. Machine learning-assisted high-throughput screening of metal–organic frameworks for CO_2 separation from CO_2-rich natural gas. *Industrial & Engineering Chemistry Research*, 63(38):16497–16508, 2024.

[14] Patrick Butler, Simon D. Rihm, Sebastian Mosbach, Jethro Akroyd, and Markus Kraft. Inverse design of metal-organic polyhedra through molecular fragmentation and evolutionary optimisation. *c4e-Preprint Series*, Technical Report 340, 2025.