Digital Exploration of Metal-Organic Polyhedra in The World Avatar

Aleksandar Kondinski[a], **Simon D. Rihm**[b], **Pei Chong**[c], **Yong Ren Tan**[d], **Srishti Ganguly**[d], **Sebastian Mosbach**[b], **Jethro Akroyd**[b], **Ari Fischer**[d], **Xu Rong**[c], **Tej Choksi**[c], **Markus Kraft**[d]

[a] *Graz University of Technology, Graz, Austria*

[b] *Department of Chemical Engineering and Biotechnology (CEB), University of Cambridge, Cambridge, United Kingdom.*

[c] *Nanyang Technological University (NTU), Singapore*

[d] *Cambridge Centre for Advanced Research and Education in Singapore (CARES), Singapore* *Corresponding author: mk306@cam.ac.uk

1. Introduction

Metal-organic polyhedra (MOPs) are discrete, cage-like assemblies comprised of metal- and organic-based chemical building units (CBUs) that resemble the shape of finite, well-defined polyhedral architectures. [1, 2] Because each metal- and organic-based CBU can be selected from a broad library of building units, and they can be developed *de novo*, MOPs exhibit tunable pore sizes, internal cavities, and chemical environments. [3] As a result, they have garnered interest for applications in water purification, gas separation, and catalysis, owing to their structural versatility and ability to sequester target molecules. [4, 5] These sustainability-driven uses have spurred extensive research on how to systematically optimise MOP design features, ranging from coordination geometry to linker functionality.

2. Substantial section

MOPs can be viewed as a subclass of analogues of classical coordination cages, where discrete topologies are realised by the formation of coordinate bonds between the metal- and organic-based CBUs. [4] The concept of "assembly models" (AMs) captures essential topological constraints by mapping each CBU to a "generic building unit" (GBU) (Figure 1a). [1] Represented in the OntoMOPs ontology, the involvement of AMs enables the development of inductive reasoning algorithms based on set operations that can systematically explore the immediate chemical space of rationally designed and constructible combinations of MOP assemblies (i.e. the primary step in an automated workflow as shown in Figure 1b). Using the OntoMOPs ontology and data describing 151 experimentally known MOPs, 18 AMs and 7 GBU types, our algorithm has rationally designed 1418 new MOP structures. [1] The overall approach narrows a vast combinatorial landscape to high-value candidates that can be targeted in the laboratory with minimal trial and error.

Following the rational design step, we developed an automated assembly-modeling strategy that uses geometric information for metal- and organic-based CBUs, along with predefined assembly motifs (AMs), to construct structural models of both existing and newly conceived MOP assemblies suitable for quantum chemical calculations. In contrast to purely rational design (which provides conceptual guidelines for MOP formation), this automated method systematically generates atomic models by specifying connectivity and binding sites, then rescaling, translating, and rotating each fragment to form the intended polyhedral geometry. This structural modelling step (Figure 1b) avoids many pitfalls of standard cheminformatic methods, whose scope is often limited for metal-rich systems, by preserving experimental local geometric parameters around the metal clusters and thereby mitigating geometry distortions. We validate the resulting MOP structures against existing data and then disseminate them via The World Avatar, [6] enabling further analyses (e.g., pore dimensions, internal volumes). Because this hybrid approach does not rely on high-level electronic-structure computations, it also permits high-throughput screening of prospective MOP candidates.

These pre-screened candidates can now be examined more deeply with complex computational methods. Such *viability assessment* can include high-throughput DFT calculations to screen structural stability and predicted properties prior to experimental trials (an ongoing work). Beyond geometric properties such as pore sizes, important applications such as CO_2 reduction via electro- or photocatalysis are based on the complex electronic structure of MOPs need to be explored. This way, promising synthesis candidates can be found in a fast and highly economical fashion.

An additional challenge in the digital exploration of MOPs lies in capturing and organising the synthesis protocols of existing MOPs, which are essential for drawing analogies and interpreting them to establish synthesis protocols for newly proposed MOP designs. We address this by applying large language models (LLMs) to the unstructured experimental sections of scientific publications, extracting essential details (reagents, vessels, yields) and representing them using the "OntoSyn" ontology. [7] Advanced prompt-engineering techniques, including chain-of-thought, in-context learning, and

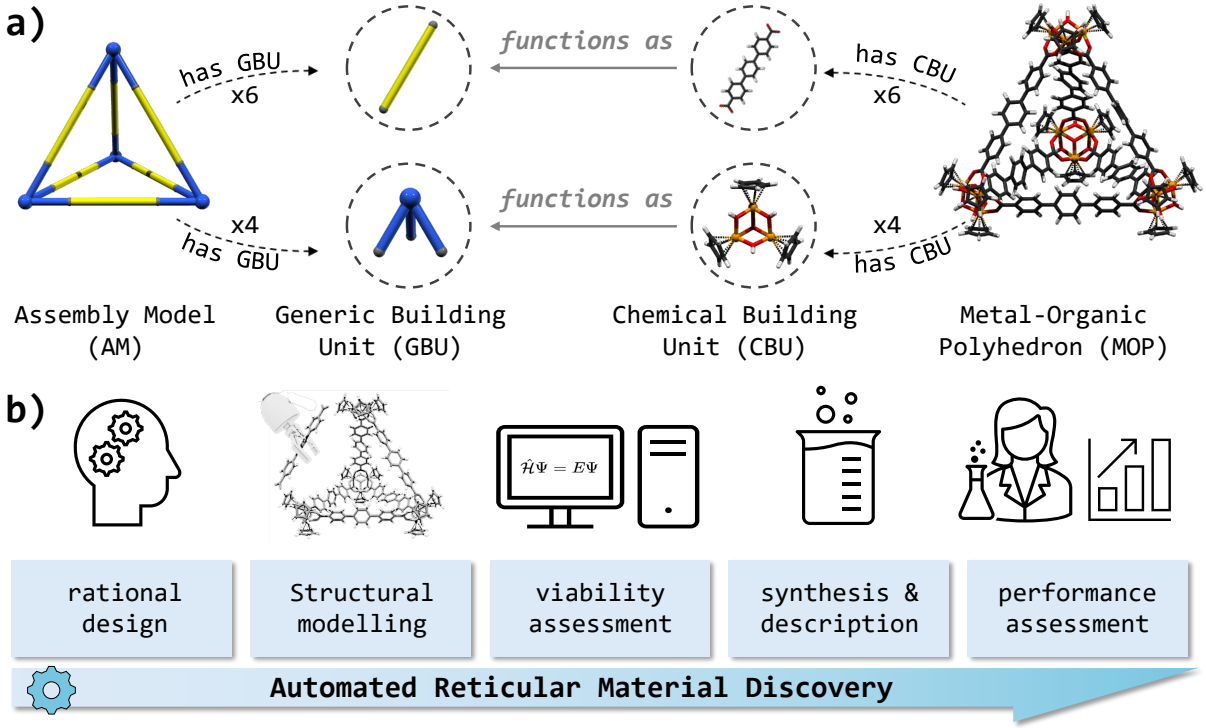

Fig. 1: (a) Concepts involved in the cognitive mapping and rational design of new MOP materials [1]; (b) A general workflow for the automated discovery of new MOPs and other reticular materials.

retrieval-augmented generation, resolve ambiguities in chemical nomenclature and workflow steps. Once serialised into structured dictionary-like format files, these data seamlessly can be instantiated in The World Avatar.

User access to the MOP chemistry is enabled through the specialised "Marie" semantic AI-agent as part of The World Avatar (`https://theworldavatar.io/demos/marie`). This agent, utilising GPT-4o-mini via in-context few-shot learning for semantic parsing and entity recognition, interprets natural language queries and translates them into SPARQL, thereby simplifying data retrieval for end users. [8] Researchers can compare pore metrics, investigate synthetic feasibility, or retrieve condensed procedures for a chosen MOP. The *viability assessment* stage also benefits from these interfaces, as users can promptly gauge which MOP might be most suitable for a targeted application and even export reaction instructions in machine-actionable scripts (e.g., XDL). By bridging theoretical design and practical inquiry in real-time, Marie significantly streamlines iterative decision-making in MOP research.

Overall, these synergistic efforts in assembly modelling, semantic data integration, and automated protocol extraction demonstrate the power of combining computational design with knowledge-graph-assisted experimentation. As illustrated in Figure 1b, our workflow culminates in a forthcoming *performance assessment* stage, where newly discovered MOPs can be tested for application-specific metrics such as adsorption selectivity or catalytic efficiency. By uniting *(i)* rational design using AMs, *(ii)* semantic linkage via OntoMOPs and OntoSyn, and *(iii)* robust LLM-driven extraction, we enable autonomous discovery pipelines that propose novel MOPs, evaluate their synthetic viability, and offer the derivation of new promising synthetic protocols enhancing new material discovery efficiency. Through a human-in-the-loop approach, these experimental protocols can be iteratively executed, and performance measurements inform the next MOP proposed. [9] These advances accelerate material realisation for environmental or energy-related applications, paving the way for a highly optimised and automated MOP research landscape.

Acknowledgments

This research was supported by the National Research Foundation, Prime Minister's Office, Singapore, under its Campus for Research Excellence and Technological Enterprise (CREATE) programme. During the preparation of this work, the author(s) used ChatGPT in order to improve readability and language. After using this tool/service, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the content of the publication.

References

[1] Aleksandar Kondinski, Angiras Menon, Daniel Nurkowski, Feroz Farazi, Sebastian Mosbach,

Jethro Akroyd, and Markus Kraft. Automated rational design of metal-organic polyhedra. *J. Am. Chem. Soc.*, 2022.

[2] Arnau Carné-Sánchez, Jorge Albalad, Thais Grancha, Inhar Imaz, Judith Juanhuix, Patrick Larpent, Shuhei Furukawa, and Daniel Maspoch. Postsynthetic covalent and coordination functionalization of rhodium(ii)-based metal–organic polyhedra. *J. Am. Chem. Soc.*, 141(9):4094–4102, 2019.

[3] Soochan Lee, Hyein Jeong, Dongsik Nam, Myoung Soo Lah, and Wonyoung Choe. The rise of metal–organic polyhedra. *Chem. Soc. Rev.*, 50:528–555, 2021.

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

[5] 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 co2 photocatalytic conversions: Current trends and future outlook. *J. CO2 Util.*, 80:102664, Feb 2024.

[6] Aleksandar Kondinski, Jiaru Bai, Sebastian Mosbach, Jethro Akroyd, and Markus Kraft. Knowledge engineering in chemistry: From expert systems to agents of creation. *Acc. Chem. Res.*, 56(2):128–139, 2023.

[7] Fabio Saluz. Leveraging ai and dynamic knowledge graphs to advance the synthesis of metal-organic polyhedra. Master's thesis, Department of Chemical Engineering and Biotechnology, University of Cambridge, November 2024.

[8] Dan N. Tran, Simon D. Rihm, Aleksandar Kondinski, Laura Pascazio, Fabio Saluz, Sebastian Mosbach, Jethro Akroyd, and Markus Kraft. Natural language access point to digital metal-organic polyhedra chemistry in the world avatar. Technical Report 327, c4e-Preprint Series, Cambridge, 2024. Technical Report 327, c4e-Preprint Series, Cambridge.

[9] Simon D. Rihm, Jiaru Bai, Aleksandar Kondinski, Sebastian Mosbach, Jethro Akroyd, and Markus Kraft. Transforming Research Laboratories with Connected Digital Twins. *Nexus*, 1(1):100004, 2024.