Preprint Cambridge Centre for Computational Chemical Engineering ISSN 1473 – 4273

Prediction and synthesis of novel metal-organic polyhedra in The World Avatar

Simon D. Rihm[1,2], Patrick Butler[1], Kumaran Elumalai[3,4], Ari Fischer[3], Pei Chong Lim[3], Xu Rong[3,4], Tej S. Choksi[3,4], Aleksandar Kondinski[5], Markus Kraft[1,2,4,6]

released: August 6, 2026

[1] Department of Chemical Engineering and Biotechnology
University of Cambridge
Philippa Fawcett Drive
Cambridge, CB3 0AS
United Kingdom

[2] CMPG
GRIPS – Gründerinnenzentrum Pirmasens
Delaware Avenue 1–3
66953 Pirmasens
Germany

[3] School of Chemistry
Chemical Engineering and Biotechnology
Nanyang Technological University
62 Nanyang Drive
Singapore, 637459

[4] CARES
Cambridge Centre for Advanced Research and Education in Singapore
1 Create Way
CREATE Tower, #05-05
Singapore, 138602

[5] Institute of Physical and Theoretical Chemistry
Graz University of Technology
Stremayrgasse 9
Graz, 8010
Austria

[6] MIT, Chemical Engineering
77 Massachusetts Avenue, Room E17-504
Cambridge, MA 02139 USA

Preprint No. 347

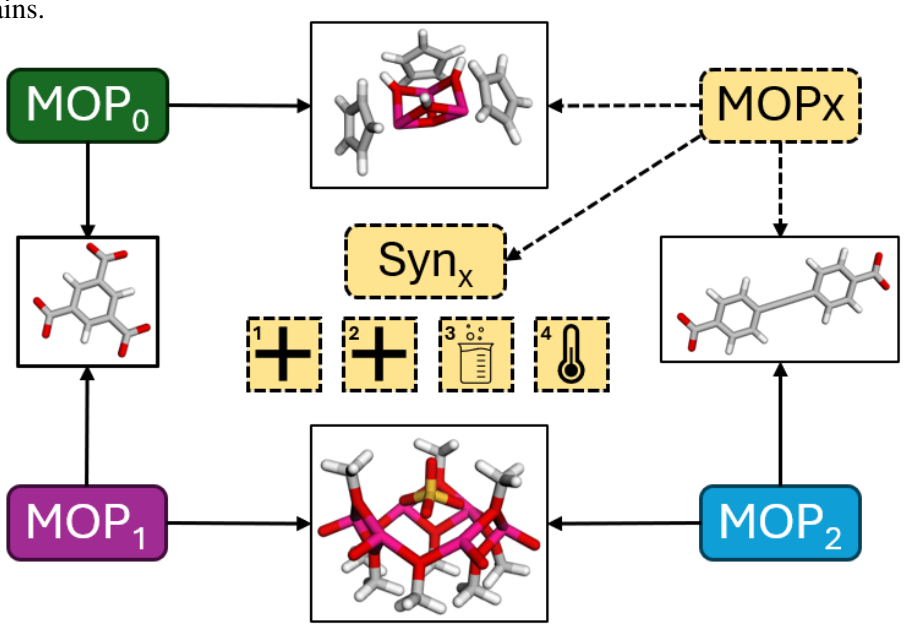
Keywords: automated material discovery, metal-organic polyhedra, digital reticular chemistry

Edited by
Computational Modelling Group
Department of Chemical Engineering and Biotechnology
University of Cambridge
Philippa Fawcett Drive
Cambridge, CB3 0AS
United Kingdom

E-Mail: mk306@cam.ac.uk
World Wide Web: https://como.ceb.cam.ac.uk/

Abstract

Metal-organic polyhedra (MOPs) offer a modular route to porous molecular materials, but their discovery is limited by the difficulty of translating digital designs into experimentally accessible compounds. Here, we demonstrate an integrated workflow for predicting and synthesising novel MOPs in The World Avatar. To our knowledge, this is the first reticular material discovery workflow executed and persisted entirely in a shared machine-readable knowledge representation, without human-specified initial reaction conditions. This domain's small literature corpus hinders data-driven synthesis prediction, yet its high symmetry provides exploitable structural analogies. We therefore encode auditable chemical reasoning as explicit rules rather than learning it statistically. MOP candidates are linked to synthesis knowledge, so procedures for new targets are inferred by analogy to known materials and rendered as human-readable instructions and machine-executable protocols. Candidates are assessed through hierarchical geometry optimisation, electronic-structure evaluation, crystal structure prediction, and simulated powder X-ray diffraction. We evaluate the workflow retrospectively against reported zirconium MOP syntheses and prospectively by synthesising a previously unreported Zr-EDB-MOP. Infrared spectroscopy, high-resolution mass spectrometry, and powder X-ray diffraction support formation of the targeted cage in agreement with the predicted model. This establishes a proof of concept for knowledge-integrated discovery in sparse, highly structured materials domains.



Highlights

- A semantic workflow within The World Avatar predicts synthesis of novel MOPs.
- First complete MOP discovery cycle based in a single persistent knowledge base.
- Synthesis conditions are inferred by analogy to known MOP procedures.
- Computational screening links molecular design to crystal verification.
- A novel Zr-EDB-MOP experimentally validates the end-to-end workflow.
- IR, HRMS, and PXRD jointly verify the predicted MOP structure.

Contents

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. [11, 27] 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. [31] 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. [1, 20] These sustainability-driven uses have spurred extensive research on how to systematically optimise MOP design features, ranging from coordination geometry to linker functionality.

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. [20] To support the systematic representation and exploration of this chemistry, MOP-related knowledge has been integrated into The World Avatar, a semantic framework for the world at large that connects domain ontologies with computational agents. Within this framework, the concept of "assembly models" (AMs) captures essential topological constraints by mapping each CBU to a "generic building unit" (GBU) as shown in Fig. 1a). [27] These concepts are represented in the `OntoMOPs` ontology, enabling inductive reasoning algorithms based on set operations to systematically explore the immediate chemical space of rationally designed and constructible MOP assemblies. This rational-design step forms the starting point of the semantic agentic workflow implemented within TWA, as illustrated in Fig. 1b). Using the `OntoMOPs` ontology and data describing experimentally known MOPs, AMs and GBU types from less than 100 publications, our algorithm has rationally designed thousands of new MOP structures. [5, 9, 27] The overall approach narrows a vast combinatorial landscape to high-value candidates that can be targeted in the laboratory with minimal trial and error.

To complement rational design, an automated assembly-modeling strategy was developed to generate initial geometries of both known and newly proposed MOPs based on their AM and the geometric properties of their CBUs. [28] This method assembles atomic structures by specifying connectivity and binding sites, and then applies rescaling, translation, and rotation operations to construct the intended polyhedral architecture. Unlike generic cheminformatics approaches, this method preserves experimental local geometry – particularly around metal clusters – resulting in high-quality structures suitable for subsequent quantum mechanical calculations. These structural models are key enablers for the next step in the pipeline shown in Fig. 1b, namely computational viability assessment.These structural models provide the computational basis for assessing whether proposed assemblies are geometrically and energetically plausible before experimental investigation.

Building on this foundation, the present study focuses on connecting digitally designed MOPs with experimentally actionable synthesis procedures. We therefore develop a knowledge-based synthesis prediction strategy that leverages analogies between the structural and topological properties of novel MOPs and those of previously synthesised compounds whose experimental procedures have been semantically extracted and encoded using the `OntoSyn` ontology [49]. Drawing on this structured synthesis knowledge base, the pre-

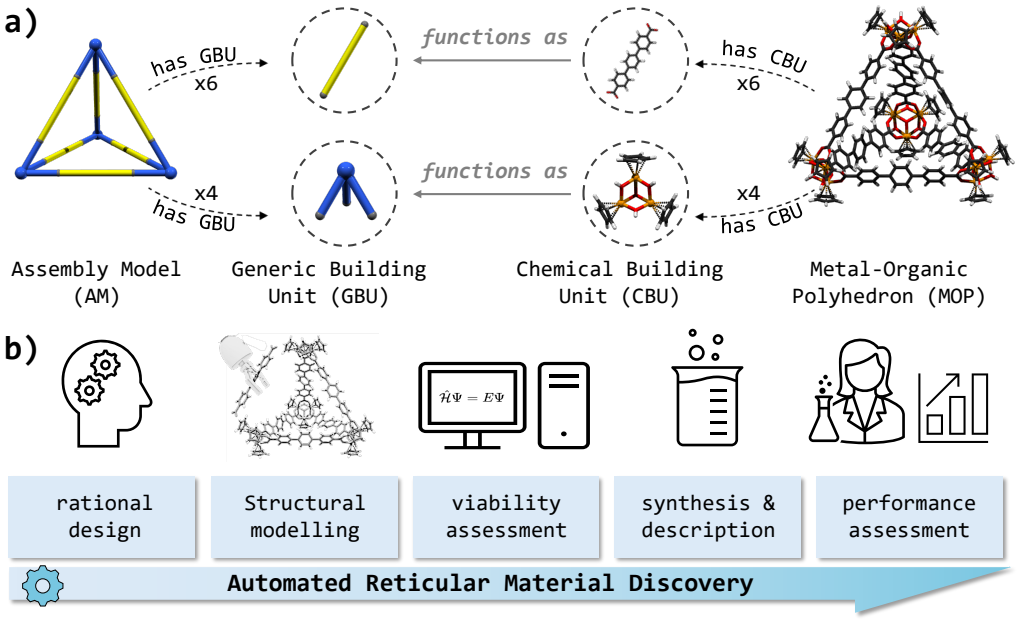


Figure 1: *(a) Concepts involved in the cognitive mapping and rational design of new MOP materials [27]; (b) A general workflow for the automated discovery of new MOPs and other reticular materials.*

diction algorithm proposes plausible synthetic protocols, including choices of solvents, reagents, and reaction conditions. These suggestions are informed by contextual similarity, such as shared AM and GBU compositions, along with heuristic rules derived from known synthesis pathways.

To complete the pipeline, we synthesise a novel MOP using a predicted procedure and validate the resulting material against its computationally screened counterpart. This closes the loop between digital design and experimental realisation within a single machine-readable knowledge representation. To our knowledge, this is the first reticular material discovery workflow in which every stage is executed through, and persisted in, one shared semantic framework. The paper makes three key contributions:

1. **An end-to-end discovery cycle, executed rather than proposed.** Candidate generation, structural modelling, computational viability assessment, synthesis condition prediction, and experimental verification are carried out as a single connected sequence, with inputs and outputs of every step persisted in a shared knowledge representation. Prior work in reticular materials has closed subsets of this loop but the experimental conditions were often supplied by a human expert or searched within human-specified bounds (see section 2.4). Here the conditions themselves are derived from structured prior knowledge without requiring an expert to propose the initial reaction conditions

2. **A knowledge-based method targeted at a regime where machine learning does not apply.** Data-driven synthesis prediction rests on corpora of 10^5 to 10^7 procedures (see section 2.4). In contrast, the corresponding corpus for MOPs is of

order 10^2. We therefore encode chemical reasoning explicitly, as auditable rules over structural analogy, rather than learning it. This is particularly suited to highly symmetric reticular assemblies, where a small set of building units and assembly models generates many structurally analogous precedents despite sparse synthesis data.

3. **Protocols that are simultaneously human- and machine-executable.** A predicted procedure is stored as an `OntoSyn` instance, from which a human-readable instruction sheet and machine-executable laboratory code are alternative renderings of the same object. [47, 49] The syntheses reported here were performed manually, but this is an implementation choice rather than a property of the workflow.

2 Reticular Chemistry in The World Avatar

2.1 The World Avatar

The World Avatar (TWA) is a dynamic digital twin framework designed to integrate data, knowledge, and computational agents into a virtual single living knowledge graph (KG) that mirrors and interacts with the real world.[2, 32] Unlike traditional Semantic Web approaches, TWA embeds real-time data flows, inference logic, and computational processes directly within a distributed architecture powered by autonomous agents. This allows not only for continuous synchronisation with the physical world but also for transparent propagation of new knowledge, facilitated through metadata tracking and structured derivations.[3]

In the chemistry domain, TWA supports a modular suite of ontologies that represent specific types of chemical information in a machine-interpretable format. For reticular materials such as MOPs, the `OntoMOPs` ontology models topological features like assembly models and their constituent building blocks, enabling rational design and structural analysis. The `OntoSyn` ontology complements this by capturing synthesis procedures in a structured form, facilitating reasoning and retrieval. Together with a network of agents that perform tasks such as structural generation, property calculation, and protocol extraction, this semantic infrastructure enables an end-to-end workflow for MOP discovery, from digital conception to experimental realisation.

2.2 Digital Reticular Chemistry

The term "Digital Reticular Chemistry" was introduced by Lyu et al. [37] to describe a vision for transforming the largely empirical field of reticular material discovery into a data-driven discipline powered by automation, computation, and integrated data ecosystems. Their proposed framework outlines a digitally orchestrated discovery process focused primarily on two- and three-dimensional frameworks such as metal-organic frameworks (MOFs) and covalent organic frameworks (COFs). In this work, we extend that vision by including zero-dimensional analogues – namely, metal-organic polyhedra – and

by integrating a broader suite of digital technologies such as KGs, quantum chemical calculations, and semantic reasoning within a unified infrastructure.

We argue that The World Avatar provides a comprehensive digital environment capable of realising this expanded vision. It addresses the four core pillars outlined by Lyu et al. [37] as essential for digital reticular chemistry. First, TWA supports structured knowledge representation through dedicated ontologies such as `OntoMOPs` [5, 28] for structural models and `OntoSyn` [49] for synthesis protocols, effectively forming a semantic database for storing and communicating domain-specific knowledge. Second, TWA enables a computational discovery cycle by linking these ontologies with simulation results, including automated potential energy surface scans [39] and quantum chemistry outputs encoded in `OntoCompChem`[30]. Third, the platform supports an experimental discovery cycle *via* integration with high-throughput experimentation and autonomous laboratories, as demonstrated in the Digital Lab Framework and self-driving lab initiatives [4, 47]. Fourth, TWA facilitates rich human–machine interaction through SPARQL query interfaces [43], intuitive user interfaces, and AI assistants like Marie [50], enabling users to interrogate and leverage the graph-based chemical knowledge.

In addition to covering the four pillars, TWA also unifies the five categories of data identified as necessary for automated reticular material discovery. These include calculated data from the computational cycle (*via* `OntoCompChem`), imported molecular data from public databases (*e.g.*, PubChem and ChEBI through `OntoSpecies` [43]), synthesis data extracted from literature (`OntoSyn` [49]), experimental observations from high-throughput labs (`OntoLab` and `OntoDOE` [4]), and contributions from the broader scientific community. TWA's distributed architecture inherently supports this open, extensible framework for community-driven enrichment.

Moreover, TWA employs knowledge graphs that blend factual entities (*via* object properties) and quantitative parameters (*via* data properties), making them amenable to advanced techniques such as KG embeddings. This aligns closely with the graph-theoretic and machine learning (ML) driven approaches advocated by Lyu et al. [37]. TWA has already demonstrated progress in the three strategic directions they identify for digital reticular chemistry. The exploration of "infinite" reticular chemical space has been advanced through the rational design of over 2,000 new MOP structures using automated topology-based algorithms [5, 27]. The challenge of connecting structure to property and vice versa is addressed through high-throughput experimentation in automated labs. While concerns have been raised about such systems being "black boxes," TWA's Digital Lab Framework explicitly encodes the surrounding experimental context and procedural logic to ensure interpretability and reproducibility [47]. Finally, the prediction of synthesis conditions for novel assemblies – a central goal of digital reticular chemistry – is tackled directly in this work through a novel knowledge-based synthesis prediction algorithm, discussed in detail in Section 3.

Ultimately, TWA's ability to integrate multidisciplinary knowledge *via* a system of distributed, interoperable ontologies and semantic agents makes it a robust platform for realising the vision of Digital Reticular Chemistry -— not only for frameworks like MOFs and COFs but also for discrete molecular assemblies such as MOPs.

2.3 Automated discovery of MOPs

As outlined above, The World Avatar offers a powerful infrastructure for implementing the principles of digital reticular chemistry. Metal-organic polyhedra serve as an ideal testbed for such a framework because their discrete, zero-dimensional architectures make their topological description and automated assembly more tractable than those of extended frameworks, while still exhibiting rich functional behaviour and potential applications. Our goal is to implement a full digital pipeline for MOP discovery and synthesis, from data extraction and knowledge representation to structure prediction, viability assessment, and experimental validation. The components of this workflow are summarised in Figure 2.

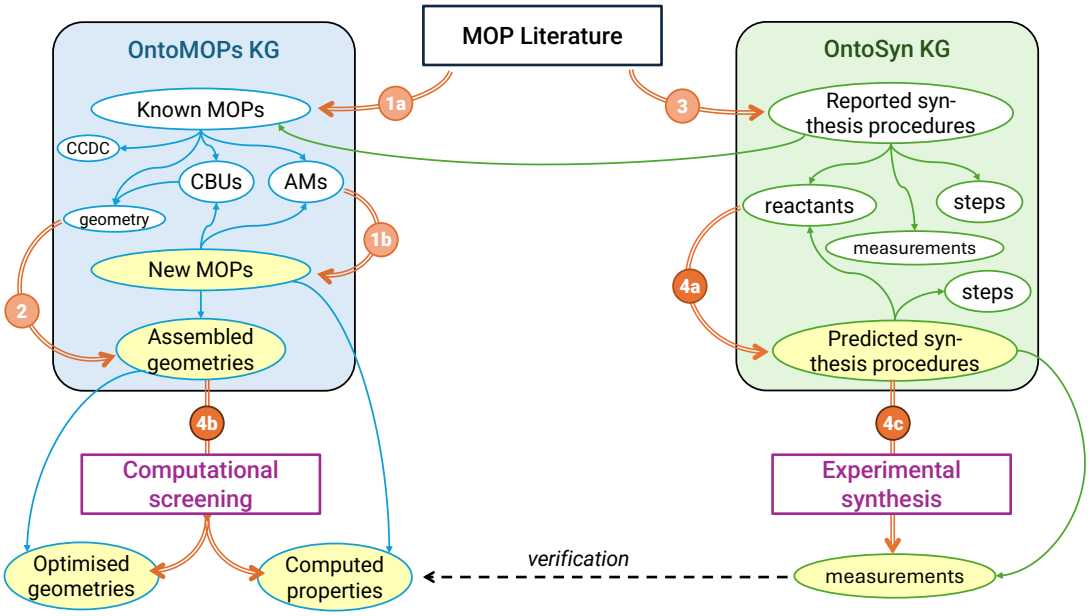


Figure 2: *Workflow of automated MOPs material discovery within TWA. Large orange arrows indicate important steps, (1) - (3) have previously been published. Elliptic bubbles indicate relevant concepts in the* `OntoMOPs` *and* `OntoSyn` *knowledge graphs, of which the ones highlighted in yellow are added in this work.*

The first part of the workflow, encompassing steps (1) to (3), has been developed in earlier work. In step (1), known MOPs are extracted from the literature and encoded into the `OntoMOPs` KG as structured representations involving assembly models, chemical building units, and associated geometric information. These structured inputs then enable the rational design of new MOP candidates *via* automated set-theoretic reasoning over AM-GBU combinations (steps 1a and 1b; see Section 2.3.1). In step (2), initial geometries of these new MOPs are assembled from their building blocks using a rule-based approach that preserves the geometric integrity of experimentally derived metal-organic fragments (see Section 2.3.2). Step (3) involves the extraction and structuring of synthesis procedures from published MOP literature, which are instantiated in the `OntoSyn` KG using large language models and prompt-engineering strategies (see Section 2.3.3).

This paper focuses on the second half of the pipeline, depicted as step (4) in Figure 2.

Here, we introduce three new components. Step (4a) involves a knowledge-based algorithm that predicts synthesis procedures for newly designed MOPs by drawing analogies to previously reported protocols encoded in `OntoSyn` (see Section 3). Step (4b) performs computational screening of the assembled geometries to evaluate their structural and electronic viability, providing optimised structures and predicted properties that feed back into both design and synthesis decisions (see Section 4). Finally, step (4c) demonstrates the experimental synthesis of a selected MOP following a predicted procedure, with characterisation results compared against computed properties to verify the accuracy of the pipeline (see Section 5).

2.3.1 Rational design

The rational design of reticular materials benefits greatly from digital workflows that integrate chemical structures, geometric principles, and curated material data. The `OntoMOPs` ontology provides a semantic framework that links MOPs to their constituent building blocks and design logic. It encodes relationships between MOPs, their CBUs, and abstract geometric analogues known as GBUs, enabling informed material design *via* structured queries and computational exploration. Central to this framework are assembly models: geometric blueprints composed of GBUs that define the overall shape and connectivity of the target MOP. CBUs, representing real chemical species such as ligands or metal clusters, are assigned to GBUs based on geometric and binding characteristics. This layered representation supports both topological reasoning and chemically meaningful design validation.

Using this framework, new MOPs were generated by combining chemically complementary and geometrically compatible building units. [27] Stable assemblies were constructed by matching CBUs with compatible binding sites and mapping them onto GBUs with corresponding spatial roles (*e.g.*, linear, trigonal, pyramidal). Assembly models such as tetrahedra or octahedra were systematically populated with appropriate CBUs, enabling automated construction of valid MOP candidates. This combinatorial reasoning strategy led to the design of over 2,000 previously unreported structures, demonstrating how ontological modelling can drive material discovery through structured, scalable exploration of chemical space. More recently, this design space was expanded by fragmenting reported organic CBUs and recombining the resulting molecular fragments using template-based assembly [9], widening the `OntoMOPs` chemical design space to nearly 800,000 MOP configurations.

2.3.2 Geometry assembly

To enable downstream computational screening of candidate MOPs, we previously developed a geometry assembly algorithm that builds three-dimensional structures directly from assembly models and their constituent building units. [28] Unlike string-based tools typically used for purely organic cages, this method handles inorganic, organic, and hybrid CBUs by positioning them in 3D space according to predefined connectivity rules and local geometries. The output structures are chemically reasonable and immediately suitable for higher-level quantum chemical calculations.

The assembled geometries align closely with experimentally known MOPs and serve as reliable initial guesses for density functional theory (DFT) refinement. This approach is particularly valuable for metal-rich systems where traditional molecular mechanics often fails to preserve correct coordination environments. All generated structures are semantically linked within The World Avatar and accessible through natural language interfaces such as Marie, supporting further integration with automated reasoning and high-throughput screening workflows.

2.3.3 Synthesis extraction

To address the lack of structured synthesis data for MOPs, we previously developed an automated pipeline for extracting and representing synthesis procedures from unstructured literature sources. [49] This approach applies large language models combined with advanced prompt engineering techniques to identify and interpret relevant information – such as reactants, solvents, vessels, and workflow steps – embedded in the experimental sections of scientific publications.

The extracted data are standardised using the `OntoSyn` ontology, which formalises synthesis procedures in a machine-readable format and links them to related entities in The World Avatar. Nearly 300 synthesis protocols were automatically processed and instantiated within the KG, enabling advanced querying, data-driven retrosynthetic reasoning, and integration with rational design tools. This capability provides the foundation for inferring synthesis routes for newly proposed MOPs, bridging the gap between digital design and experimental realisation.

2.4 Synthesis planning: from reaction templates to data-driven prediction

The appropriate synthesis-planning strategy depends on two axes: the number of available procedures, which determines how far statistical inference can be trusted, and the structural regularity of the target class, which determines how far explicit analogy can reach. The following survey positions existing approaches along these axes and explains why MOPs fall outside the effective range of the data-driven methods that dominate the field.

Organic synthesis planning represents the mature, data-rich end of the spectrum. Expert knowledge has been encoded in rule-based systems such as Chematica and validated through laboratory execution, [26] while molecular-similarity methods retrieve applicable precedents directly from known reactions. [14] Neural policies combined with tree search subsequently learned to prioritise retrosynthetic transformations from large reaction collections, [52] alongside related advances in condition recommendation, robotic execution, template-free reaction prediction, and reaction-data standardisation. [15, 23, 51] These systems learn from 10^6–10^7 reaction records and explicit bond transformations between discrete molecules; however, solvothermal self-assembly has no natural bond-disconnection graph, and its decisive variables (solvent, modulator, temperature, and time) are treated as contextual information rather than the primary prediction object.

Materials synthesis research therefore shifts the prediction target from reaction routes to experimental conditions. Text-mining and natural-language-processing workflows have extracted precursors, operations, temperatures, and other synthesis parameters from the inorganic materials literature, particularly for oxide and solid-state synthesis. [25, 29] These developments established literature extraction as an important foundation for data-driven materials synthesis and are situated within the broader progress of materials-focused information extraction reviewed by Olivetti et al. [42]. The resulting corpora contain approximately 10^4–10^5 recipes and can support statistical analysis of synthesis conditions; however, they are dominated by solid-state inorganic materials and are generally represented as flat, composition-centred records in which product structure and topology are not linked richly enough for structural analogy to be expressed.

Reticular materials provide the closest precedent to the present work. Moosavi et al. [40] showed that successful, partially successful, and failed experiments collectively encode transferable chemical intuition for MOF crystallisation, motivating the inclusion of unsuccessful experiments discussed further in Section 5. Luo et al. [36] subsequently trained structure-conditioned synthesis models using the 983-entry SynMOF database, while Zheng et al. [61] extracted 26,257 synthesis parameters for approximately 800 MOFs and used them to predict crystallisation outcomes. DigiMOF expanded the available corpus to 15,501 MOFs, [18] and more recent work has moved towards multimodal synthesis–structure–property resources and automated evaluation of reticular synthesis datasets. [46, 60] Even the largest datasets used in this area therefore remain within the 10^3–10^4 range and are already considered marginal for robust statistical prediction; `OntoSyn`, by comparison, contains approximately 300 MOP synthesis procedures. [49] MOP synthesis prediction must consequently generalise from roughly one to two orders of magnitude fewer examples than these already data-limited MOF approaches.

Closed-loop and autonomous laboratories address the complementary problem of searching an experimental condition space efficiently. Computation-guided robotic screening has accelerated the discovery of organic cages, [21] mobile robotic platforms have autonomously optimised functional materials, [8] and language-model agents now connect information retrieval, synthesis planning, experimental control, and reaction optimisation. [7] Autonomous materials laboratories similarly combine literature-derived recipes with active learning, [55] while agentic MOF-discovery systems integrate generative design, synthetic-feasibility assessment, and experimental validation. [22] Complementing these operational demonstrations, the Digital Lab Framework introduced in Section 2.2 provides a systems-level semantic architecture for linking experimental data, laboratory processes, digital twins, and knowledge-based decision-making within TWA. [47] These platforms can efficiently refine conditions once a target, precursor set, reaction family, or bounded search space has been supplied; however, generating the initial condition set and its chemically defensible bounds from structured precedent remains largely manual.

MOPs occupy the low-data, high-regularity corner of these two axes. Their limited synthesis corpus constrains high-capacity statistical models, but their recurring building units, coordination geometries, topologies, and ligand-substitution relationships provide strong foundations for explicit knowledge-based transfer. The same balance arises in coordination cages, polyoxometalate clusters, and substantial parts of framework chemistry. The approach developed here therefore uses linked structural and procedural knowledge to

identify closely matched precedents and transfer synthesis conditions between them, allowing structural regularity to compensate for data scarcity.

3 Predicting synthesis procedures by analogy

The insights gained from the newly extracted data set of detailed synthesis procedures allows us to hypothesise a few general rules that enable the prediction of previously unreported syntheses. To demonstrate the viability of automated MOP synthesis planning, we focus on a proof-of-concept study involving zirconium-based metal-organic polyhedra (Zr-MOPs). Zirconium nodes offer a well-understood coordination environment, high chemical stability, and broad synthetic accessibility, making them ideal candidates for method development and validation. This study remains within the bounds of what has been termed the "simple structure space" [37], where a limited set of node geometries and reaction types still permits a wide range of structural outcomes. Notably, prior work has shown that greater diversity in MOP architectures can be achieved by varying organic linkers rather than metal centres [37], reinforcing the design flexibility available even within this constrained space. The following sections detail how novel Zr-MOP candidates were selected and how synthetic routes were identified and validated through integration of computational and semantic tools.

3.1 General Idea

The key idea behind synthesis prediction is analogy, building on the rational design approach introduced by Kondinski et al. [27] and illustrated in Fig. 3. Similarity-based retrosynthesis methods have previously demonstrated that structural resemblance between known and target compounds can be used to guide the prediction of viable reaction pathways without relying on exhaustive rule sets or template libraries [14]. Adapting this principle to MOPs, the analogy is constructed from two pairs of MOPs with different metal CBUs. The target pair consists of the reproduced reference MOP_0 and the novel MOP_x, which both contain the same metal CBU M_0 but differ in their organic ligand, L_0 versus L_2. The reference pair consists of two known MOPs, MOP_1 and MOP_2, which instead share a different metal CBU $M_1 = M_2$, with $M_1 \neq M_0$, but exhibit the same ligand substitution from L_0 to L_2. Based on this ligand switch observed in the reference metal family, we infer a synthesis route for MOP_x by transferring the procedural changes from MOP_1 to MOP_2 onto the verified synthesis of MOP_0.

To reduce variability introduced by lab-to-lab differences, MOP_0 should be a structure that has been successfully reproduced at the present site. Based on MOP_0, we identify sets of MOPs satisfying the following boundary conditions implicit in Fig. 3:

- $M_x = M_0$: only the ligand is switched between MOP_0 and MOP_1.
- $L_x = L_2$ and $L_1 = L_0$: same ligand switch occurs from MOP_1 to MOP_2.
- $M_1 = M_2$ and $M_1 \neq M_0$: MOP_1 and MOP_2 are based on a different metal CBU.

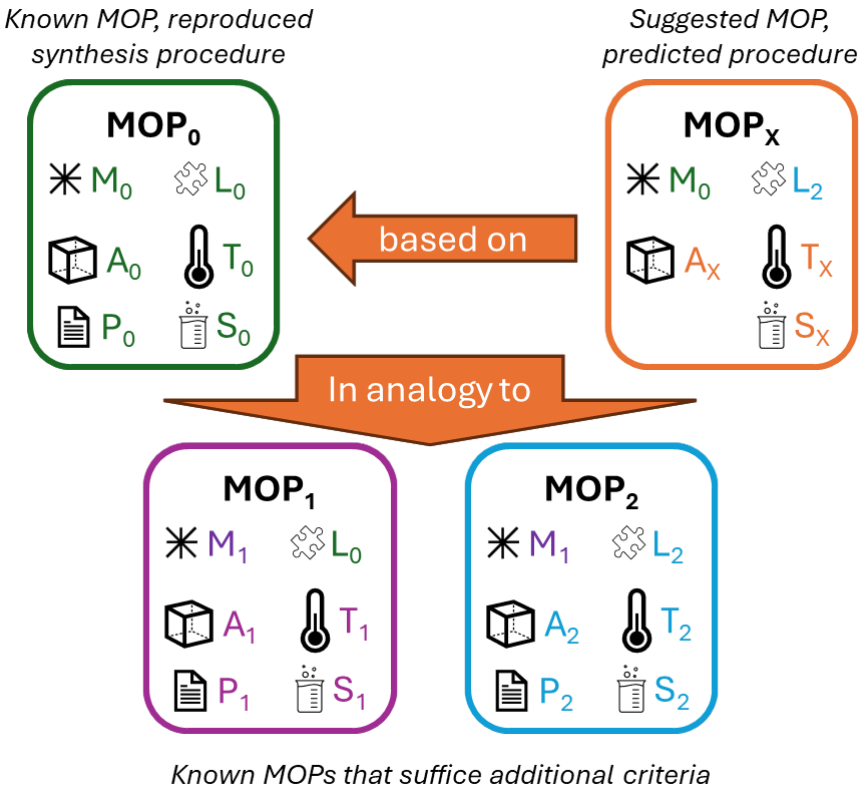


Figure 3: *Illustration of the general idea of predicting MOPs synthesis procedures by analogy. M: metallic CBU, L: organic Ligand, A: assembly model, T: heating regime, P: provenance, S: solvent.*

Each of the resulting tuples $\{MOP_0, MOP_1, MOP_2\}$ represents a potential route for predicting the synthesis procedure of MOP_x. Multiple such routes can often be derived even from a single MOP_0. We then score these routes using additional criteria that either further strengthen the analogy of the ligand switch or simplify synthesis prediction in a different way.

Strong criteria are enforced through the prediction agent:

1. $A_1 = A_0$ and $A_2 = A_x$ to ensure the ligand switch from L_0 to L_2 is actually possible (*i.e.*, has been observed before) within the respective assembly models.

2. $T_2 \approx T_1$ to confirm our assumption that the heating regime is dictated by the metal CBU, which informs our prediction.

3. $R(M_2) = R(M_1)$ to confirm, that the same reactants were used for the metal CBU, which lets us use a simple rule of three for predicting amounts.

Weak criteria are desirable but not required:

4. $P_1 = P_2$ to ensure comparability of synthesis procedures for MOP_2 and MOP_1, as procedural changes for switching ligands from M_0 can be more reliably predicted

if analogous changes from M_2 happened within the same lab respectively. This is specifically important for predicting amounts.

5. $A_1 = A_2$ and $A_0 = A_x$ further simplifies analogies: If criterion 1 applies as well, all MOPs involved have the same assembly model. If not, a (slightly weaker) analogy can be made now as the ligand switch occurs within a single assembly model respectively.

6. $B(L_2) = B(L_0)$ further narrows down the list of candidates by demanding ligands with the same binding sites (*e.g.*, carboxylate-based). This is less relevant if criterion 4 applies.

7. $S_2 \approx S_0$ which drastically simplifies the prediction of S_x.

The assembled and rated routes now offer multiple combinations for synthesis prediction by analogy. Ideally, a new MOP is available for synthesis prediction *via* different routes of which many suffice most criteria defined above. These different routes towards MOP_xcan now be used for verification as well as screening purposes: first, we check if the predicted synthesis procedures are similar between the different routes. If they are, the small differences between predicted conditions are suitable as screening ranges for experimentation. If they are not, screening should be initially done separately with small variances around the different sets of conditions – this should give enough data to focus on one of the predicted condition sets for further experimentation.

Lastly, if few or only a single route is available to predict conditions to synthesise MOP_1, the conditions for synthesis of MOP_0 can offer different permutations: if MOP_0 was successfully synthesised under different sets of conditions or at least one that differs from the literature-reported ones, these can be used for predicting different sets of conditions for MOP_1. Ideally, we want to stick to the procedure used for reproduction of MOP_0 in the same lab that presented the largest yield.

3.2 Prediction algorithm details

This section outlines the knowledge-based algorithm used to predict synthesis procedures for novel MOP candidates, based on the analogy framework introduced above. The algorithm decomposes the synthesis procedure into four distinct components – reactants, solvents, heating conditions, and additional steps – each of which is inferred by comparing the reference MOPs and applying a defined set of heuristics. These rules are designed to capture common chemical practices and observable trends in published synthesis data, while also incorporating contextual adjustments based on local experimental knowledge. For each component, we describe the logic used for prediction, any fallback strategies, and how the confidence of the prediction is assessed.

3.2.1 Reactants

Predicting reactants for MOPs is typically straightforward, as the mapping between CBUs and their corresponding reactants is often unique – particularly for organic ligands. In this

study, the dataset primarily includes ditopic, tritopic, and tetratopic carboxylate-based ligands (sufficing criterion 6), for which the corresponding reactants are the respective di-, tri-, or tetracarboxylic acids.

The situation can be more complex for metal-based CBUs. For example, as illustrated in Fig. 4, Polyoxovanadate (POV) based MOPs can be synthesised from a range of Vanadium-containing reactants and their mixtures. This added variability is largely avoided here by focusing exclusively on Zr-MOPs, which are consistently synthesised using zirconocene dichloride. Furthermore, the application of criterion 3 to MOP_2 and MOP_1 ensures consistent metal precursor usage across reference MOPs. Accordingly, the prediction yields: $\mathrm{R}(\mathrm{M}_x) = \mathrm{R}(\mathrm{M}_0)$ and $\mathrm{R}(\mathrm{L}_x) = \mathrm{R}(\mathrm{L}_2)$

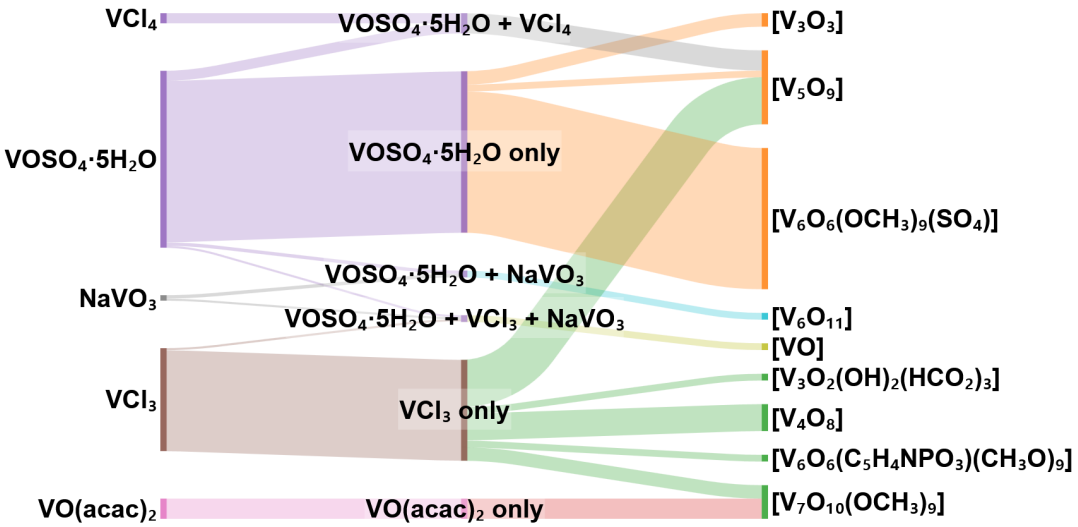


Figure 4: *Sankey diagram of metal CBUs and their corresponding reactants used in synthesis procedures for Vanadium-based MOPs as reported by Rihm et al. [49].*

Predicting **reactant amounts** is a crucial part of the automated synthesis procedure, as manual estimation and trial-and-error can be time-consuming. The approach used here first defines an anchor value and then calculates the remaining absolute amounts based on relative ratios. An anchor value is chosen based on either the desired product mass *via* expected yield or constraints on reactant availability. In this study, we fix the anchor value as $m_{\mathrm{M}_1} = m_{\mathrm{Cp_2ZrCl_2}} = 50\,\mathrm{mg}$.

Relative amounts can be predicted by stoichiometry, but literature shows that often amounts not close to stoichiometric ratios are used. Especially small molecules that serve as metal complex activators are often used in excess, for example water in Zr-MOPs ($3\,\mathrm{ZrCl_2(C_5H_5)_2} + 4\,\mathrm{H_2O} \longrightarrow \mathrm{ZrO(OH)_3(C_5H_5)_3}^{4+} + 3\,\mathrm{C_5H_6} + 2\,\mathrm{HCl} + 4\,\mathrm{Cl}^-$).

Even the key metallic or organic reactants are often present in non-stoichiometric amounts:

$$\frac{m_{M,i}}{m_{L,i}} \neq \frac{n_{M,i}}{n_{L,i}} \cdot \frac{\tilde{M}_{M,i}}{\tilde{M}_{L,i}}. \tag{1}$$

To account for this, we introduce excess factors f_i defined as

$$\frac{m_{M,i}}{m_{L,i}} = f_i \cdot \frac{n_{M,i}}{n_{L,i}} \cdot \frac{\tilde{M}_{M,i}}{\tilde{M}_{L,i}}. \tag{2}$$

For MOP_x, the predicted excess factor is given by

$$f_x = \frac{f_2}{f_1} \cdot f_0. \tag{3}$$

The rationale is to assume a similar excess factor as for the reference MOP with the same metal core, adjusted proportionally based on the effect of the ligand change as observed in the MOP_1 to MOP_2 transformation.

An important advantage of this approach is that it avoids the need to explicitly define balanced reactions, stoichiometric coefficients, or even most molar masses. The resulting expression is:

$$\frac{m_{M,x}}{m_{L,x}} = \frac{f_2}{f_1} \cdot f_0 \cdot \frac{\left(n\tilde{M}\right)_{M,0}}{\left(n\tilde{M}\right)_{L,x}} = \frac{\frac{m_{M,2}}{m_{L,2}} \cdot \frac{\left(n\tilde{M}\right)_{L,x}}{\left(n\tilde{M}\right)_{M,2}}}{\frac{m_{M,1}}{m_{L,1}} \cdot \frac{\left(n\tilde{M}\right)_{L,0}}{\left(n\tilde{M}\right)_{M,2}}} \cdot \frac{m_{M,0}}{m_{L,0}} \cdot \frac{\left(n\tilde{M}\right)_{L,0}}{\left(n\tilde{M}\right)_{M,0}} \cdot \frac{\left(n\tilde{M}\right)_{M,0}}{\left(n\tilde{M}\right)_{L,1}} = \frac{m_{M,2} \cdot m_{L,1} \cdot m_{M,0}}{m_{M,1} \cdot m_{L,2} \cdot m_{L,0}} \tag{4}$$

Alternatively – or when criterion 3 is not satisfied – stoichiometric amounts can be calculated using $f_x = 1$, providing a baseline prediction. When using excess factors, we allow a screening range of $\pm 20\%$.

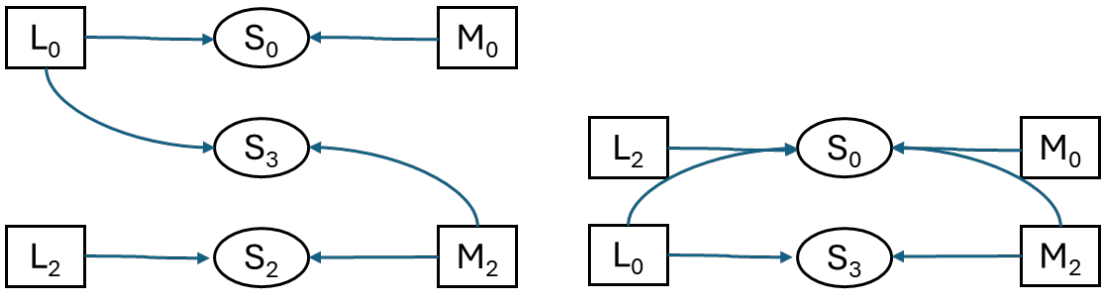
3.2.2 Solvents

Solvent selection remains one of the most critical challenges in chemical synthesis and materials discovery – a challenge that TWA can help address, as demonstrated in appendix A. Within TWA, solvents can be identified either through explicitly defined `Dissolve` steps or as chemicals classified as solvents in the `OntoSpecies` knowledge graph.

As illustrated in Fig. 5, two cases can be distinguished based on the similarity of solvents used in the reference synthesis procedures:

- Fig. 5a shows the general case where the synthesis procedures for MOP_0, MOP_1, and MOP_2 might use different solvents. In this case, we assign a mixture of solvents S_2 and S_0 as we know that S_0 can dissolve $M_0 = M_1$ and S_2 can dissolve $L_2 = L_1$.
- A simpler case arises when synthesis procedures for MOP_0 and MOP_2 use the same solvent as shown in Fig. 5b. In this case, we use solvent S_0 as it apparently dissolves $M_0 = M_1$ as well as $L_2 = L_1$.

In practice, solvents S_0, S_1, and S_2 are often mixtures. Furthermore, nominally different solvents, such as DMF and DMA, may function interchangeably.



(a) *General case.* **(b)** *Special case* $S_0 = S_2$.

Figure 5: *Different cases of solvent similarity between* MOP_0, MOP_2, *and* MOP_1. *Arrows indicate solubility of reactants* $\mathrm{R(L)}_i$ *and* $\mathrm{R(M)}_i$ *in solvent* S_i.

Solvent volume is predicted based on its ratio to the total reactant mass:

$$Y_i = \frac{V_{S,i}}{m_{M,i} + m_{L,i}} \tag{5}$$

The predicted ratio Y_x is derived analogous to reactant excess factors:

$$Y_x = \frac{Y_2}{Y_1} \cdot Y_0 \tag{6}$$

This approach assumes that criterion 3 is satisfied. If not, Y_i is no longer transferable and predictions should be treated with caution. As an additional check, individual solvent-to-reactant ratios $\frac{V_{S,i}}{m_{M,i}}$ and $\frac{V_{S,i}}{m_{L,i}}$ are also evaluated for each synthesis procedure.

3.2.3 Heating

The heating conditions for $\mathrm{MOP_x}$are predicted directly from those of $\mathrm{MOP_0}$, *i.e.*, $T_x = T_0$. This decision is grounded in the observation that the heating regime is primarily dictated by the metal-based CBU, rather than the organic ligand. As shown by the analysis of available synthesis procedures in Fig. 6, temperatures cluster consistently for MOPs sharing the same metal CBU, regardless of ligand identity. This supports the use of $\mathrm{MOP_0}$ as a reliable reference for thermal conditions in the prediction of $\mathrm{MOP_x}$.

From a chemist's perspective, this is a reasonable assumption. The role of heating in MOP formation is largely to enable the formation of the metal complex, which serves as a stable coordination centre. Once the metal-ligand framework is established under the appropriate thermal conditions, the organic ligands typically undergo self-assembly to the thermodynamically favoured cage structure. Therefore, if the metal CBU remains unchanged between $\mathrm{MOP_0}$ and $\mathrm{MOP_1}$, the required heating regime can be expected to remain the same as well.

3.2.4 Additional steps

To ensure completeness and experimental practicality, generic instructions are appended at the end of each predicted synthesis procedure. These steps are commonly applied and

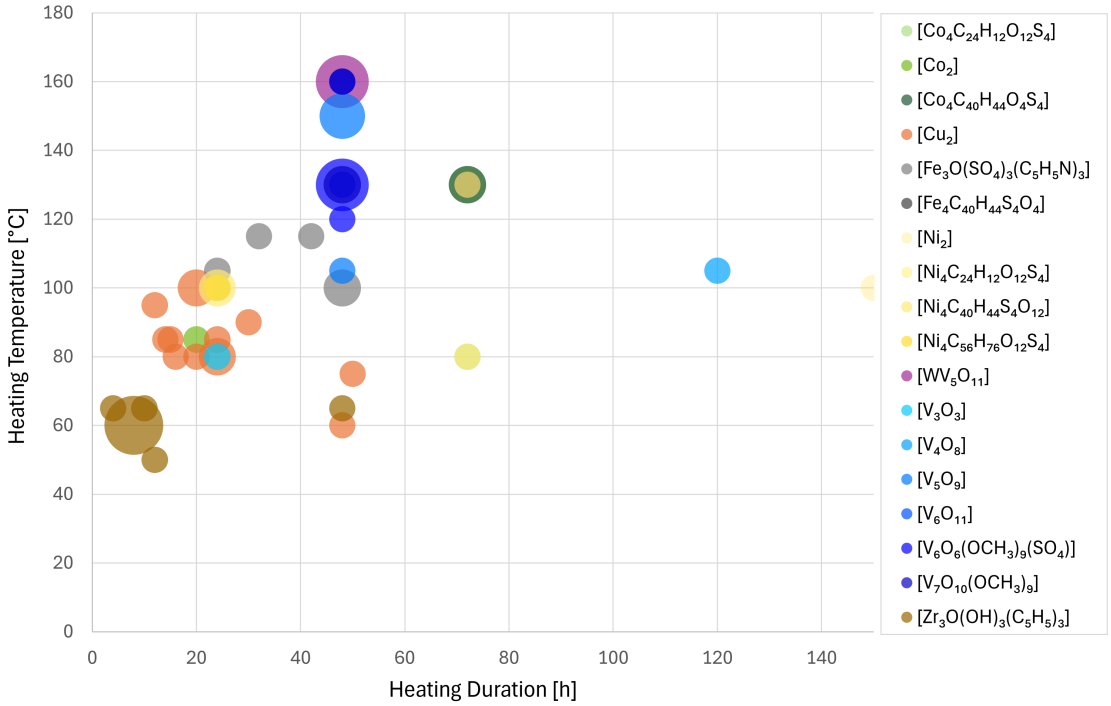


Figure 6: *Heating temperature and duration used for crystallisation of selected MOPs as reported by Rihm et al. [49].*

serve to increase the likelihood of successful synthesis:

- If the ligand does not fully dissolve during preparation, sonication should be applied to aid dissolution.
- After the reaction, wash the resulting solid three times using the same solvent employed in the synthesis procedure.

These additions standardise post-processing and pre-reaction preparation, helping to address common experimental issues while maintaining consistency across predicted protocols.

3.3 Verification

For testing this algorithm, we employed a tetrahedral Zr-MOP using the ditopic linker terephthalic acid (H_2BDC) as MOP_0 as this structure is one of the simplest and well-known Zr-based MOPs which we were able to reproduce as discussed in section 5.1. To assess the reliability of the synthesis prediction framework, we applied it to MOPs with known synthesis procedures. Tab. 1 lists these candidates, including details on ligands and prediction route. For each candidate, the algorithm may identify several prediction routes, each corresponding to a different MOP_1/MOP_2 analogy pair; the criteria fulfilment ratio in brackets refers to the best-ranked route.

Table 1: *Already known or reported MOPs which were identified as candidates for synthesis prediction. Ligand nomenclature is adapted from Gosselin et al. [20].*

Predicted MOP	Ligand	Routes	Status
$[Zr_3O(OH)_3(C_5H_5)_3]_4[(C_6H_3)(C_6H_4)_3(CO_2)_3]_4$	H_3BTB	2 (80%)	Exists in `OntoMOPs` [33]
$[Zr_3O(OH)_3(C_5H_5)_3]_4[(C_6H_3NH_2)(CO_2)_2]_6$	H2BDC-NH_2	3 (90%)	Exists in `OntoMOPs` [41]
$[Zr_3O(OH)_3(C_5H_5)_3]_4[(C_6H_4)_2(CO_2)_2]_6$	H_2BPDC	1 (100%)	Exists in `OntoMOPs` [33]
$[Zr_3O(OH)_3(C_5H_5)_3]_4[(C_{10}H_6)(CO_2)_2]_6$	H_2NDC	1 (70%)	Reported elsewhere [13]
$[Zr_3O(OH)_3(C_5H_5)_3]_4[(C_6H_3Br)(CO_2)_2]_6$	H_2BDC-Br	2 (90%)	Reported elsewhere [56]
$[Zr_3O(OH)_3(C_5H_5)_3]_4[(C_6H_4)_3(CO_2)_2]_6$	H_2TPDC	1 (100%)	Reported elsewhere [19]

We selected two of the MOPs shown in Tab. 1 to compare predicted against reported procedures and provide a first benchmark of the algorithm. We selected Zr-BDC-NH_2-MOP based on H_2BDC-NH_2, whose reported synthesis was part of the original data set analysed in previous work [49], and Zr-TPDC-MOP based on H_2TPDC, which was not represented in TWA at the time of prediction and thus constitutes an out-of-sample test. To simplify comparison, we used the amount of zirconocene dichloride in the reported procedure as the anchor value for our prediction. Suggested screening ranges for reactant amounts are shown behind the predicted values.

3.3.1 Synthesis of Zr-BDC-NH_2-MOP

The synthesis procedure was reported by Nam et al. [41]. The prediction was based on the extracted synthesis procedures of V_7O_{10}-based MOP_2 [59] and MOP_1 [59].

Reported: To a 5 mL vial, Cp_2ZrCl_2 (17.5 mg), 2-aminoterephthalic acid (5.4 mg), DEF (1 mL) and H_2O (150 µL) were added. The mixture became clear yellow solution after sonication. The solution was heated in preheated oven at 60 °C for 8h followed by 4h cooling. Yellow cubic crystals were synthesised and washed with fresh DEF.

Predicted: Dissolve 17.5 mg of zirconocene dichloride and 6.3 mg (5.0 mg to 7.6 mg) of 2-amino-1,4-benzenedicarboxylate in 0.5 mL (0.4 mL to 0.6 mL) of DEF along with a few drops of distilled water. Heat the mixture at 60 °C for 8 hours. Crystallise under static conditions, then allow to cool to room temperature in air. Filter the resulting crystals and wash repeatedly with DEF.

The two alternative routes identified for this target yielded similar predictions at equal or lower criteria fulfilment, with the lowest-ranked route deviating most from the reported procedure

3.3.2 Synthesis of Zr-TPDC-Br-MOP

The synthesis procedure was reported by Gosselin et al. [19]. The prediction was based on the extracted synthesis procedures of Fe_3O-based MOP_2 and MOP_1. [54]

Reported: In a 20 mL vial zirconocene dichloride (350 mg) and p-terphenyl-4,4"dicarboxylic acid (H2TPDC) (186.4 mg) were dissolved in dry DMF (20 mL). The stock solution

(3 mL) and DI water (450 µL) were transferred to a 7 mL vial and heated at 65 °C for 8 hours. White crystalline material was collected [...] [and washed] with fresh DMF three times.

Predicted: Dissolve 350 mg of zirconocene dichloride and 192.7 mg (154 mg to 231 mg) of p-terphenyl-4,4"dicarboxylic acid in 6.1 mL (4.5 mL to 7.8 mL) of DMF along with a few drops of distilled water. Heat the mixture at 60 °C for 8 hours. Crystallise under static conditions, then allow to cool to room temperature in air. Filter the resulting crystals and wash repeatedly with DMF.

3.3.3 Discussion

The predicted and reported procedures are highly similar in terms of stoichiometry, solvent system, heating regime, and work-up steps. This comparison therefore provides a retrospective benchmark of the procedure-prediction logic against known or subsequently reported syntheses, rather than a prospective experimental validation. The latter is provided in Section 6, where the algorithmically suggested conditions are used to guide the synthesis of a previously unreported MOP.

The degree of criteria fulfilment correlates with the closeness of prediction and reported procedure. For Zr-TPDC-MOP, where all criteria were met, the predicted ligand amount deviates by less than 4% from the reported one, with matching solvent and reaction duration; for Zr-BDC-NH_2-MOP at 90% fulfilment, solvent and heating regime are recovered exactly while the ligand amount deviates moderately. This supports using criteria fulfilment to rank candidate routes when several are found. At the same time, predictions that deviate from a reported procedure are not necessarily invalid, as a given MOP can typically be synthesised via multiple distinct procedures. Zr-BPDC-MOP provides direct evidence: its predicted procedure differs from the reported one in solvent (DMF vs. DMA), temperature profile, and ligand amount, yet experimentally yielded the target MOP, albeit with lower product quality than the reported procedure.

The largest systematic discrepancy concerns solvent volumes: for Zr-TPDC-MOP, 6.1 mL were predicted against 20 mL reported, even though all other parameters were closely matched. While substitution among DMF, DEF, and DMA is unlikely to fundamentally alter the outcome for these Zr-based MOP syntheses [24, 33, 56], the comparatively low predicted solvent volumes indicate that this part of the algorithm requires further refinement. Nonetheless, the close agreement in the main synthetic parameters supports the use of simple analogy-based prediction for initial screening. This is consistent with prior work showing that even minimal structural similarity can support effective synthesis prediction [14], and is particularly promising for niche compound classes such as MOPs, where the available data are too sparse for robust machine-learning models but structurally analogous examples are abundant.

4 Computational screening

A high-throughput computational modelling workflow for metal–organic polyhedra was established in previous work [10]. This workflow extends the automated assembly of computation-ready MOP geometries [5, 28] by applying post-assembly optimisation and property calculations. The assembled structures provide chemically meaningful starting geometries because the local coordination environments of the chemical building units are preserved during construction, while the subsequent computational steps refine these geometries and assess their suitability for downstream screening.

In the present work, this strategy is used as a computational feasibility filter before experimental synthesis. The initially assembled geometries are first checked for structural consistency, including missing atoms, atom overlaps, and chemically unreasonable arrangements. Structures that pass this initial validation are then optimised using the GFN2-xTB method implemented in the xTB package, with an implicit solvent continuum parameterised for DMF where appropriate. These semi-empirical optimisations provide a computationally efficient assessment of whether the assembled MOP geometry relaxes to a stable local minimum without major distortion of the intended assembly model.

For selected candidates, the xTB-optimised geometries can be further refined or analysed at a higher level of theory using DFT. In particular, single-point DFT calculations on the optimised structures provide access to electronic properties and relative energies that are not available from the purely geometric assembly workflow. These calculations can be used to evaluate ligand-exchange reactions and related stability metrics, thereby providing an additional estimate of synthetic viability. Overall, the computational screening therefore proceeds through a hierarchy of increasing accuracy and cost: assembly-model validation, xTB-based geometry optimisation, and DFT-based electronic-structure evaluation.

4.1 Crystal structure prediction

The crystal structure prediction (CSP) of the Zr-MOPs was performed using quasi-random sampling of specified space groups as implemented in mol-CSPy. [12] This involved generating trial structures by sampling the appropriate asymmetric unit with a given MOP geometry and corresponding number of chloride counter ions. In the case of Zr-BDC and Zr-BPDC, the MOP geometry was extracted from the experimentally reported structures. In the case of Zr-EDB, the MOP geometry was assembled from the chemical building units and either the M_2L_3 or M_4L_6 assembly model using our MOP Assembler [5, 28] and then energy minimised by the GFN2-xTB method combined with the polarisable continuum model using the DMF parameters. [6] The space groups sampled were based on the experimentally observed structures for the Zr-BDC and Zr-BPDC CSPs (in the case that the experimentally observed structures was Z' < 1, subgroups were identified that yielded Z' = 1). For Zr-EDB, the space groups sampled were $P1$, $P\bar{1}$, $P2_1$, $P2_1/n$, and $Pmmm$. The trial structures were subsequently energy minimised using a custom interface to GULP and the UFF4MOF method. [16, 17] The optimisation was performed with rigid molecules and an applied pressure of 0.1 GPa. Partial charges were determined by

fitting to distributed multipoles, which were calculated by a distributed multipole analysis (DMA) of the B3LYP/6-31G(d,p)/LANL2DZ electron density. [53] The final UFF4MOF energies were often unphysical (likely due to the constraints and approximations) and hence final energy-density landscapes were determined based on single point calculations of the optimised structures using the UMA-s1p1 machine learning interatomic potential with the OMAT task setting. [57]

Matching the predicted structures against those experimentally observed was performed using the VC-PWDF method. [38] A successful match was determined by a score less than 0.035. Structures were further examined visually to confirm the crystal packing.

4.2 Alternative cage topologies of Zirconium-based MOPs

Many of the predicted procedures concern Zr-MOPs with a metallic-to-organic CBU ratio of 2:3 which can produce either a tetrahedral M_4L_6 configuration or a capsule-shaped M_2L_3 configuration. As reviewed in detail by Gosselin et al. [19], the resulting topology appears to be governed largely by the width-to-length ratio of the ligand. Fig. 7 illustrates this empirical relationship and provides an estimate of whether a novel Zr-MOP is more likely to form a tetrahedron or a capsule (resp. "cigar"). For the purpose of synthesis prediction, the two configurations can therefore be treated as variants of the same assembly model as they involve the same coordination chemistry. This interpretation is supported by Zr-BDC-MOP, which lies in the overlapping region of Fig. 7: the reported procedures for its tetrahedral [33] and capsule-shaped forms [19] are essentially identical.

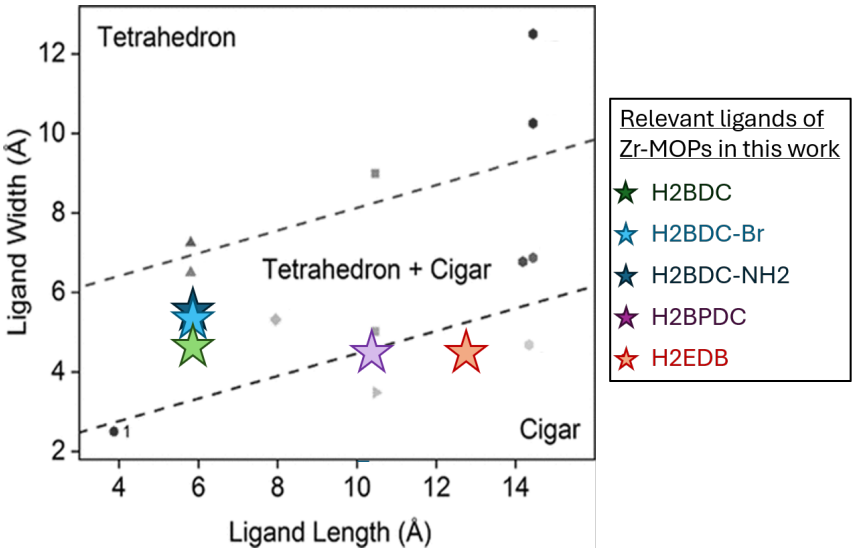
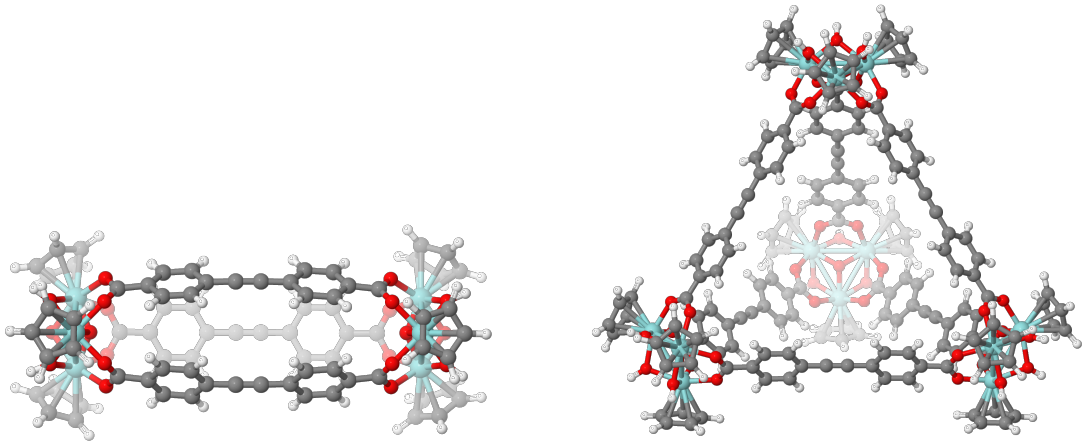


Figure 7: *Empirical relationship between ligand width and length and the observed cage topology of Zr-MOPs, adapted from Gosselin et al. [19]. Relevant MOPs in this work are highlighted.*

For the novel Zr-EDB-MOP synthesised in section 6, the ligand dimensions shown in Fig. 7 suggest that the M_2L_3 capsule topology is the more likely assembly outcome. However, the empirical ligand-geometry relationship alone does not exclude formation of the

M_4L_6 tetrahedral topology. Calculations on the isolated cages likewise do not provide a clear energetic distinction between the two configurations, as both are constructed from the same building units and coordination motifs. We therefore consider both the M_2L_3 capsule shown in Fig. 8a and the M_4L_6 tetrahedron shown in Fig. 8b in the subsequent structural verification.



(a) *Capsule M_2L_3.* **(b)** *Tetrahedron M_4L_6.*

Figure 8: *Possible cage topologies of Zr-EDB-MOP.*

4.3 Verification of predicted structures

Before applying CSP and experimental validation, it is important to assess whether the molecular geometries generated by the assembly workflow constitute realistic representations of the target MOPs. Previous work demonstrated that the assembly algorithm preserves the local coordination environments of the chemical building units while constructing the overall polyhedral architecture from the corresponding assembly model [10]. The resulting structures therefore serve as physically meaningful starting points for subsequent quantum-chemical optimisation.

To evaluate the quality of these assembled geometries, we compared structures obtained at different levels of theory: the initial assembly-model geometry, geometries optimised using GFN2-xTB, and geometries further refined by DFT calculations. For known zirconium-based MOPs with experimentally reported crystal structures, characteristic geometric descriptors such as metal–metal distances, linker orientations, and overall cage dimensions were analysed. As shown in Fig. 9, the optimised structures remain close to the original assembly-model geometries, indicating that the assembly procedure already produces geometrically realistic cages. At the same time, the progressive reduction in deviation from experimental structures when moving from assembly-model geometries to xTB and finally DFT optimisation demonstrates systematic improvement with increasing computational accuracy. The smallest deviations are obtained for the solvent-corrected DFT geometries, providing confidence that the predicted molecular structures represent physically realistic local minima and are suitable starting points for CSP and subsequent experimental comparison.

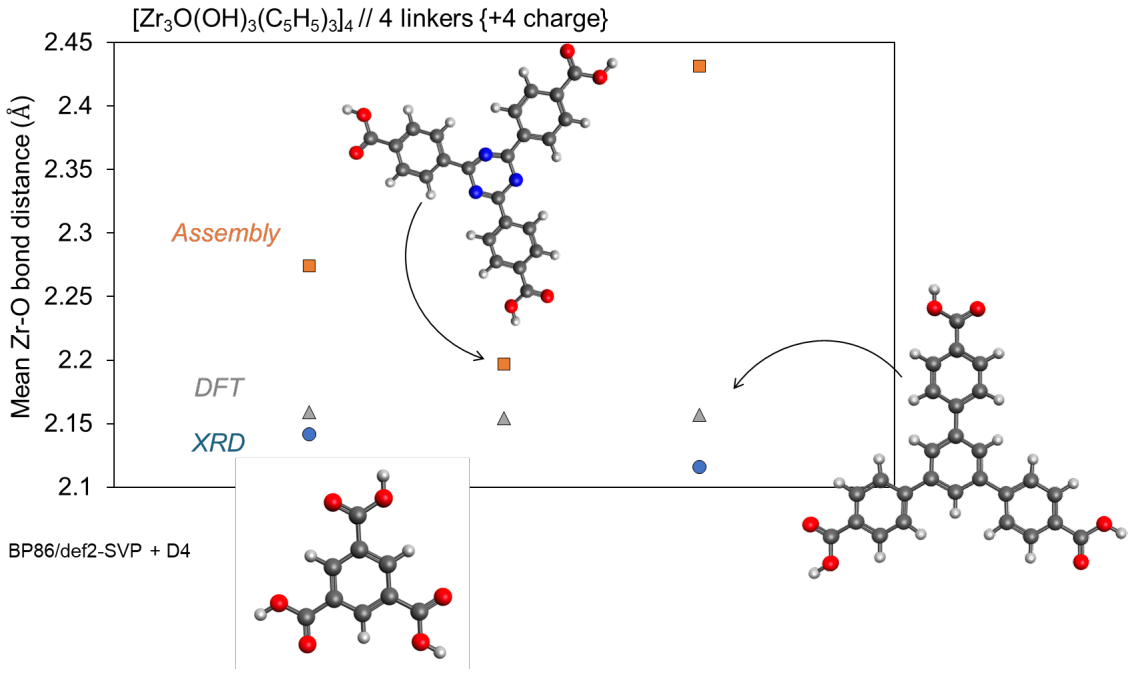


Figure 9: *Geometry metrics for Zr-based MOPs (tetrahedral with 4-linkers)*

The CSP workflow was verified by applying it to zirconium-based MOPs for which experimental crystal structures have previously been reported and deposited as CIF files in the Cambridge Structural Database. For these known systems, predicted crystal structures were compared directly with the reported structures by visual inspection of the crystal packing and quantitatively through simulated powder XRD (PXRD) patterns. This provides a stringent validation of the workflow, as successful prediction requires not only a realistic molecular geometry but also the correct packing arrangement and unit-cell-level periodic structure.

A representative example is provided by the tetrahedral Zr-BDC-MOP in Fig. 10. Visual comparison of the predicted and reported crystal structures showed close agreement in the arrangement of the MOP cages. This qualitative match was supported by quantitative PXRD comparison, where the simulated pattern of the predicted crystal structure showed excellent agreement with the pattern generated from the reported CIF structure, yielding a difference score of 0.0004. The same comparison also includes the measured PXRD spectrum of the Zr-BDC-MOP synthesised in this work as MOP0, demonstrating that the reproduced material is consistent with the expected crystalline phase and can therefore serve as a reliable reference structure for subsequent synthesis prediction.

Additional validation was performed for the capsule-shaped Zr-BDC-MOP and the tetrahedral Zr-BPDC-MOP seen in Fig. 11. In both cases, the main diffraction features of the predicted structures reproduce those of the reported structures, although the agreement is less exact than for the highly symmetric tetrahedral Zr-BDC case. This is expected because the lower symmetry and more complex packing of these structures make both crystal structure prediction and PXRD-based matching more challenging. Nevertheless, the predicted patterns capture the principal peak positions and overall diffraction signa-

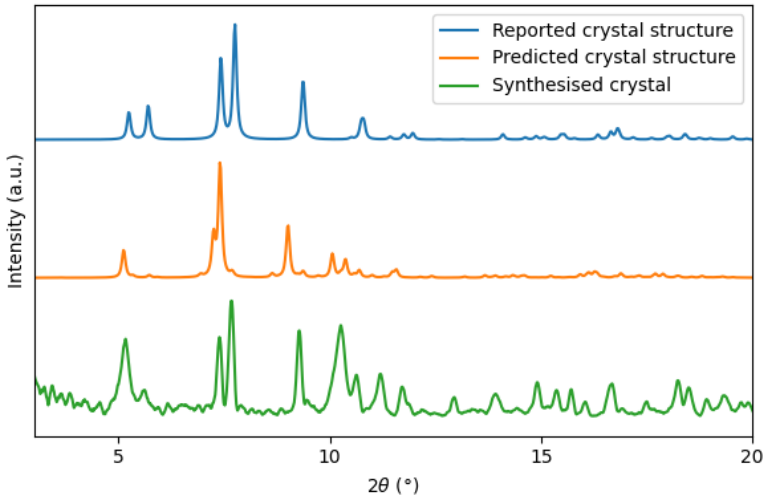
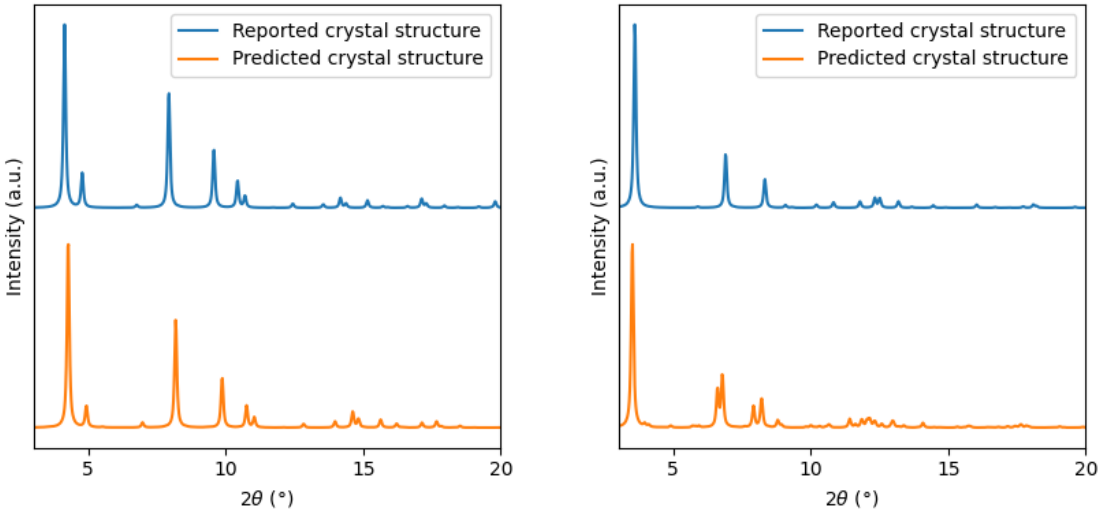


Figure 10: *PXRD patterns of tetrahedral Zr-BDC-MOP generated from the reported and predicted crystal structures, together with the measured PXRD pattern of the material synthesised in this work.*

tures of the reported phases, supporting the use of the CSP workflow for distinguishing plausible crystal structures of novel MOP candidates.



(a) *Capsule structure.* **(b)** *Tetrahedral structure.*

Figure 11: *Comparison of simulated PXRD patterns for Zr-BDC-MOPs based on the predicted and reported crystal structures.*

5 Experimental synthesis

Experimental work was conducted within the Digital Lab Framework of TWA [47], which connects synthesis procedures, laboratory infrastructure, and experimental results through interoperable KGs. A synthesis procedure is represented once as an `OntoSyn` instance and rendered on demand either as a formatted instruction sheet for a human experimentalist or as executable code for laboratory hardware. [49] The same representation can reference the experimental setup through `OntoDevice` [48] and link outcomes to structural and spectroscopic data from sources such as PubChem [43]. Unsuccessful experiments are documented alongside successful ones, in line with best practices that emphasise the value of negative results[37, 40].

The procedures reported here were executed manually. This limits the experimental throughput of the present study, but not the interface between prediction and execution: the predicted `OntoSyn` instances used to generate the human-readable instructions are directly addressable by automated laboratory equipment. Manual and automated execution are therefore alternative renderings of the same machine-readable procedure rather than separate workflow implementations.

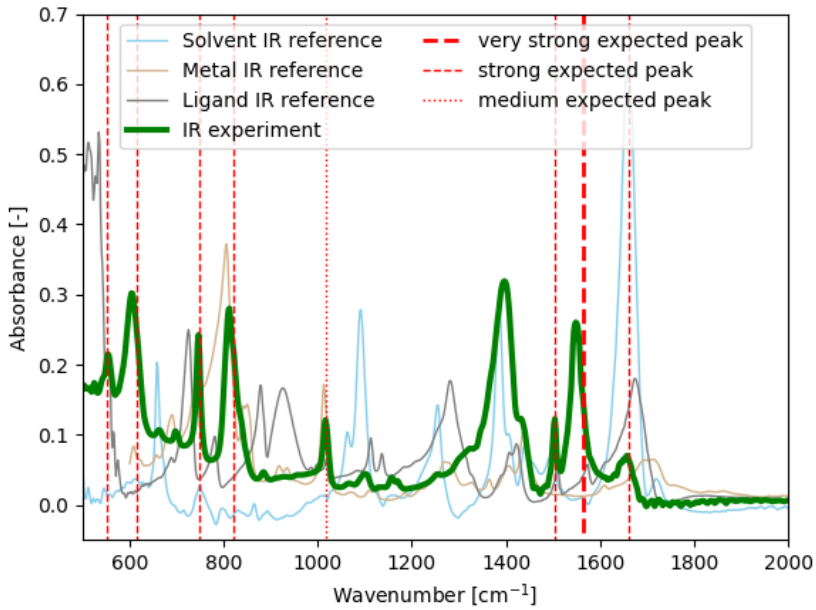
In this chapter, we first describe the reproduction of known MOPs – both to gain familiarity with experimental conditions and to satisfy the prerequisites for the synthesis prediction algorithm presented in section 3. We then detail the screening and synthesis of novel MOP candidates.

5.1 Reproduction of known MOPs

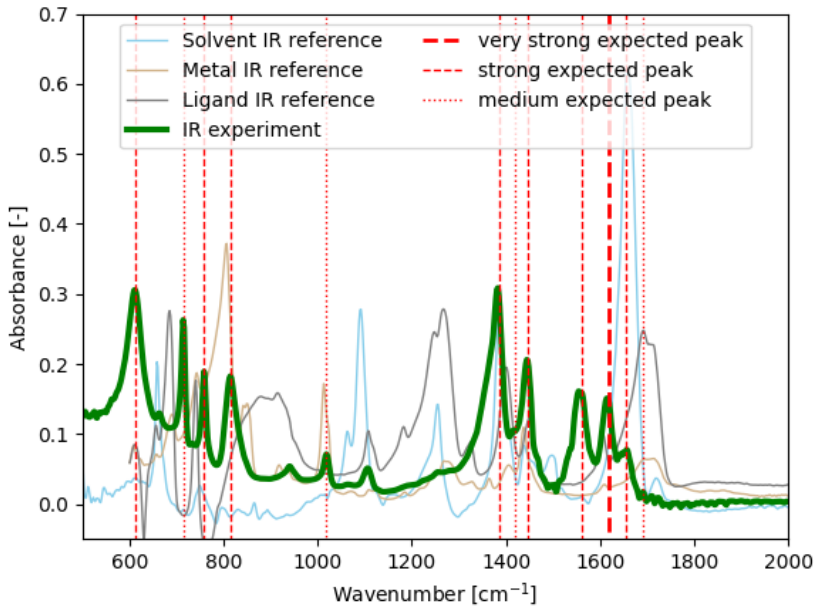
Verification of reproduced MOPs in this work primarily relies on IR spectroscopy, with experimental spectra compared against expected peak positions queried directly from The World Avatar (TWA). These reference spectra were extracted from the literature and instantiated in knowledge graphs during earlier work [49]. PXRD measurements are additionally employed to provide further structural confirmation (see Fig. 10).

Zr-BDC-MOP and Zr-BTC-MOP were selected because they are well-known structures that are often-reproduced throughout the literature [24, 33, 34, 56, 62]. If used as MOP_0 according to the analogy-based algorithm introduced in section 3, they open up a large range of novel MOP structures as their corresponding ligands H_2BDC and H_3BTC are widely used in reticular chemistry and MOP synthesis in particular.

The structures were synthesised according to the procedures reported by Liu et al. [33]. In a previous work, these synthesis procedures were automatically instantiated in TWA [49], and were now automatically queried and presented to the experimentalist in a formatted way. To provide sufficient material for subsequent characterisation and screening, both syntheses were scaled up by a factor of five relative to the reported procedures, increasing reactant and solvent amounts proportionally. The resulting crystalline powders were analysed using Fourier-transform IR (FTIR) for confirmation. Fig. 12 shows the resulting absorbance spectra, which were also automatically compared to existing IR data on reactants and products present in TWA.



(a) *Zr-BDC-MOP.*



(b) *Zr-BTC-MOP.*

Figure 12: *IR spectra of reproduced MOPs and relevant references.*

For both materials, all of the major peaks expected were present, indicating successful reproduction. For the Zr-BDC-MOP in Fig. 12a there are some minor unexpected peaks which can be attributed to residual ligand and solvent still present in the dried product.

5.2 Screening of novel MOPs

The screening of novel MOP candidates builds on the synthesis procedures predicted by the algorithm presented in section 3.2, with reaction parameters varied within defined ranges. These experiments are carried out in the laboratory following a human-in-the-loop approach, enabling experimental feedback to guide subsequent iterations. Structural verification begins with powder XRD measurements, which are compared to simulated diffraction patterns derived from predicted crystal structures based on our computational pipeline presented in section 4.1. This comparison helps identify promising candidates whose experimental patterns match theoretical predictions. For such candidates, additional analyses are performed: IR spectroscopy is used to assess ligand incorporation and detect the presence of multiple phases, while mass spectrometry (MS) is employed to confirm the self-assembly of metal–organic products.

As Zr-BDC-MOP and Zr-BTC-MOP were reproduced in section 5.1, these were used as MOP_0, resulting in a number of novel MOP candidates (MOP_x) listed in Tab. 2 together with the number of prediction routes, the criteria fulfilment of the best-ranked route, and the experimental outcome.

Table 2: *Novel MOPs which were identified as candidates for synthesis prediction. Routes are given as the number of prediction routes via Zr-BDC-MOP + via Zr-BTC-MOP; score in brackets refers to criteria fulfilment of the best-ranked route.*

Predicted MOP	Ligand	Routes	Outcome
$[(C_3N_3)(C_6H_4)_3(CO_2)_3]_4[Zr_3O(OH)_3(C_5H_5)_3]_4$	H_3TATB	3+2 (70%)	ligand unavailable
$[(C_5NH_3)_2(CO_2)_2]_4[Zr_3O(OH)_3(C_5H_5)_3]_4$	H_2BPyDC (b)	1+1 (40%)	rejected by screening
$[Zr_3O(OH)_3(C_5H_5)_3]_4[(C_{16}H_{12})(CO_2)_2]_6$	H_2HPDC	1+0 (70%)	ligand unavailable
$[Zr_3O(OH)_3(C_5H_5)_3]_4[(C_6H_4C)_2(CO_2)_2]_6$	H_2EDB	1+1 (90%)	successful (section 6)

Which candidates could be pursued experimentally was primarily dictated by ligand availability: 4,4',4"-s-Triazine-2,4,6-triyl-tribenzoic acid (H3TATB), despite yielding the largest number of prediction routes, could not be sourced and was not attempted. The ligand 2,2'-bipyridine-4,4'-dicarboxylic acid (H_2BPyDC-bent) was available, but the assembled structure was deemed unrealistic by the computational screening presented in section 4 due to excessive strain; screening experiments were nevertheless carried out and were, as expected, unsuccessful. This negative result is documented in TWA alongside the successful ones. H2HPDC was also not available. Finally, for 4,4'-(ethyne-1,2-diyl)dibenzoic acid (H_2EDB) both predicted routes were tested experimentally. Both gave promising results, with the route derived from Zr-BTC-MOP producing the clearer PXRD pattern; this route led to the successful synthesis of a novel MOP, as detailed in sections 5.3 and 6.

5.3 Verification of novel MOP synthesis

The verification of novel MOP syntheses represents the final step of the digital discovery workflow shown in Fig. 2. While the experimental measurements in this study were

performed manually, their interpretation can be partially automated through integration with TWA. Experimental data are analysed in a hierarchical manner, progressing from confirmation of reaction success to verification of the target composition and ultimately the predicted crystal structure. This workflow combines literature-derived reference data, computational predictions, and experimental measurements to provide increasing confidence in the successful synthesis of a novel MOP.

The verification workflow employed in this study is illustrated in Fig. 13. Following synthesis using a predicted procedure, experimental products are analysed through a sequence of increasingly specific characterisation methods. Each step addresses a distinct question and provides a decision point before proceeding to the next level of verification. This hierarchical approach enables efficient screening of experimental outcomes while maintaining a clear connection between computational predictions and laboratory observations.

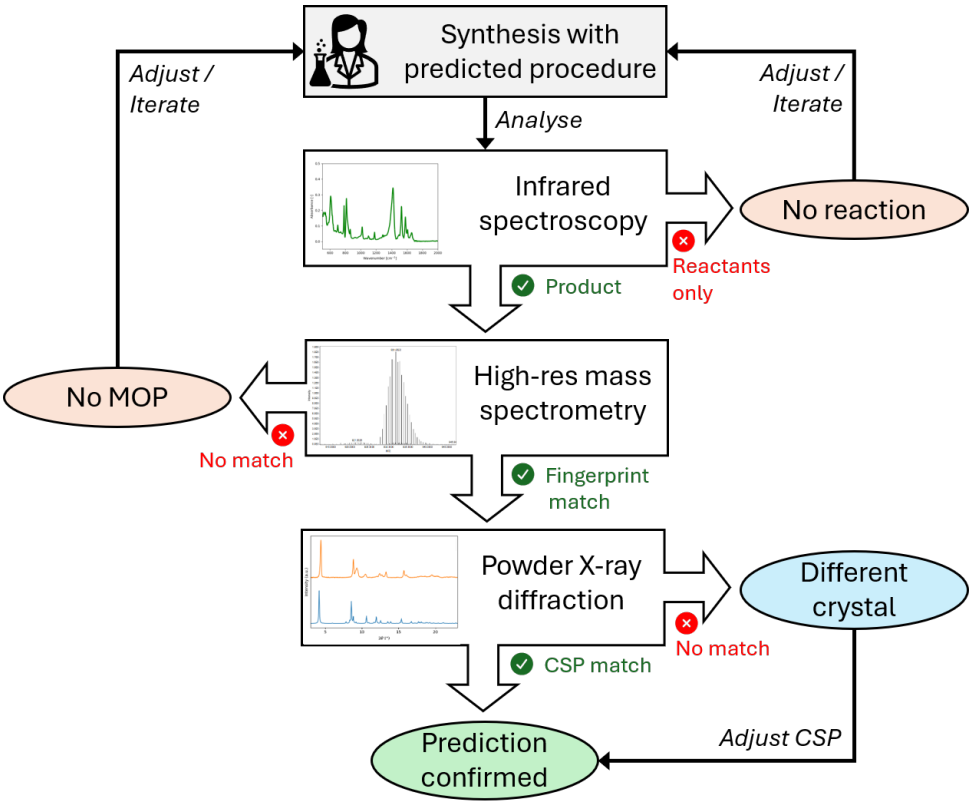


Figure 13: *Hierarchical verification workflow for predicted MOP syntheses. IR spectroscopy verifies reaction success, HRMS verifies formation of the target MOP, and PXRD verifies agreement with the predicted crystal structure obtained from crystal structure prediction.*

The first stage consists of IR spectroscopy. The measured spectrum is automatically compared against reference spectra of reactants, solvents, and previously reported MOPs stored within TWA. The primary objective of this step is to determine whether a chemical reaction has occurred and whether the product differs substantially from the starting materials. If the spectrum is dominated by reactant signatures and no characteristic product peaks are observed, the experiment is classified as unsuccessful and excluded from further

analysis. Conversely, the appearance of characteristic metal–ligand coordination features provides evidence that a metal–organic product has formed and warrants additional investigation.

Products that pass the IR screening are subsequently analysed by high-resolution MS (HRMS). At this stage, the expected molecular composition of the target MOP is already known. Theoretical isotope distributions are therefore generated directly from the predicted molecular formula and compared with the experimentally observed isotope pattern. Agreement between predicted and measured isotope envelopes provides a highly specific fingerprint for the target assembly. A lack of correspondence indicates that although a metal–organic product may have formed, it is unlikely to be the intended MOP. Only products exhibiting a satisfactory fingerprint match are advanced to structural verification. Furthermore, HRMS allows us to distinguish between alternative MOP configurations. For example, the M_2L_3 and M_4L_6 architectures discussed in Section 4.2 exhibit different molecular masses and isotope distributions, allowing the experimentally observed spectrum to identify cage topology prior to crystal-structure verification by PXRD.

The final verification stage employs powder X-ray diffraction (PXRD). Simulated diffraction patterns generated from the CSP workflow described in Section 4.1 are compared with experimentally measured diffraction data. This comparison allows direct assessment of whether the experimentally obtained crystalline phase corresponds to the predicted structure. A poor match suggests either the formation of a different crystal packing arrangement, an alternative MOP topology, or the presence of multiple phases. In contrast, strong agreement between simulated and experimental diffraction patterns provides evidence that both the molecular composition and the crystal structure correspond to the computational prediction.

Together, these three analytical levels provide progressively stronger validation of the synthesis outcome. IR spectroscopy verifies that a reaction has taken place, HRMS verifies the formation of the predicted molecular assembly, and PXRD verifies the predicted crystal structure. Ideally, this assignment would be complemented by single-crystal X-ray diffraction, which provides the most direct structural determination; however, obtaining diffraction-quality single crystals is often highly challenging for MOPs and should therefore not be treated as a prerequisite for every successful synthesis. Successful completion of the three-stage workflow nevertheless provides a robust proof-of-concept validation of the complete prediction pipeline, linking computational design, synthesis prediction, and experimental realisation within a unified digital workflow.

6 Results and discussion

This chapter presents the outcomes of our integrated prediction–synthesis workflow for MOPs within The World Avatar. We begin by detailing the synthesis and characterisation of a novel MOP obtained through the proposed pipeline, thereby validating its predictive capability. We then discuss the broader significance of these findings in the context of related efforts in digital reticular chemistry, followed by an exploration of potential applications, with a focus on targeted CO_2 photocatalysis . Finally, we outline future directions to further enhance the scope, automation, and predictive accuracy of this approach.

6.1 Synthesis of novel MOP

The central result of this work is the experimental realisation of a previously unreported zirconium-based MOP, denoted here as Zr-EDB-MOP, through the integrated prediction–screening–verification workflow developed in The World Avatar. As illustrated in Fig. 14, the target structure was obtained by applying the analogy-based synthesis prediction algorithm to reproduced reference MOPs. In particular, Zr-BDC-MOP and Zr-BTC-MOP were used as experimentally validated starting points, while known ligand-switching relationships in related MOP families provided the basis for inferring the new synthesis procedure. Both predicted routes were pursued experimentally; the route derived from Zr-BTC-MOP, which provides a direct analogy between the known linker H3BTC and the target linker H2EDB, yielded the clearer PXRD pattern and led to the successful synthesis.

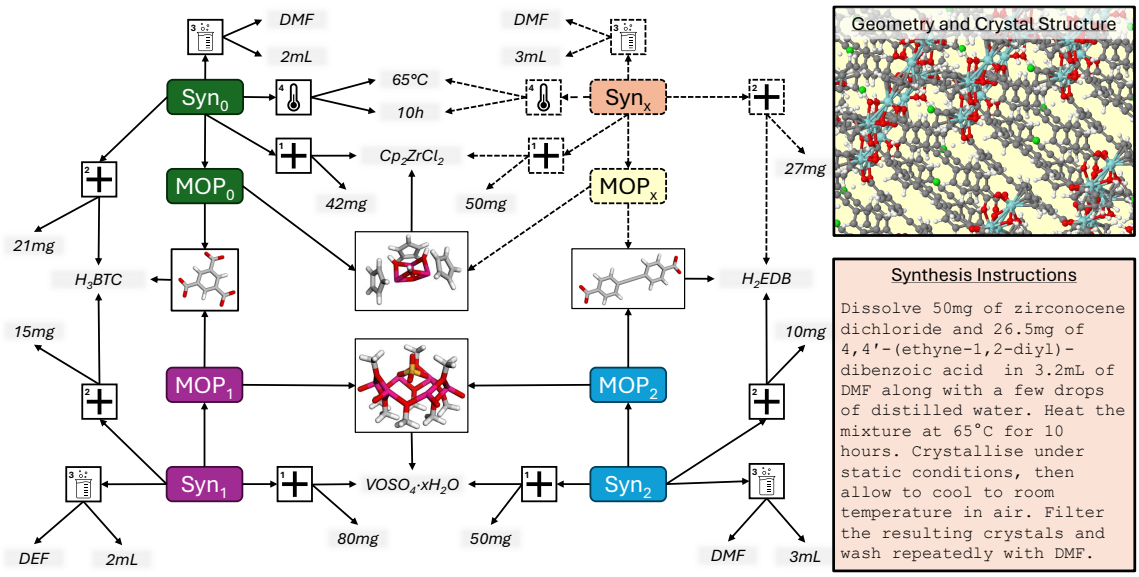


Figure 14: *Overview of the prediction and synthesis workflow leading to the novel Zr-EDB-MOP within TWA. Instantiated reference MOPs and their synthesis procedures are used to infer the target structure and synthesis steps by analogy.*

This prediction did not merely provide a single suggested recipe, but defined a chemically meaningful experimental window around reactant ratios, solvent volume, water content, and heating conditions. Within this window, crystalline material could be obtained and subsequently subjected to the verification workflow described in Section 5.3. The successful experiment used zirconocene dichloride and H_2EDB in DMF/water under solvothermal conditions, yielding crystalline product after filtration and washing. The formation of the target MOP was supported by complementary characterisation: IR spectroscopy indicated conversion of the linker environment, HRMS provided molecular-level evidence consistent with the expected Zr–EDB assembly, and PXRD showed agreement between the measured diffraction pattern and the simulated pattern derived from the predicted crystal structure.

Together, these results demonstrate the full pipeline proposed in this paper: a novel

MOP was first generated by rational design, its molecular and crystal structures were assessed computationally, a synthesis procedure was inferred automatically by analogy to known and reproduced MOPs, and the material was experimentally synthesised and verified. The resulting product was characterised by IR, HRMS, and PXRD, providing a proof-of-concept validation of the integrated prediction-synthesis workflow. Rather than establishing general predictive accuracy across the full MOP design space, this experiment demonstrates that a synthesis procedure generated from the knowledge-graph-based analogy algorithm can lead to a novel Zr-based MOP whose molecular composition and crystalline phase are consistent with the computationally predicted target.

6.2 Analysis of novel MOP

The product obtained from the synthesis conditions identified in Section 6.1 was analysed using the verification workflow introduced in Section 5.3. The aim of this analysis was to determine whether the isolated material was consistent with the predicted Zr-EDB-MOP, first at the level of reaction outcome, then molecular composition, and finally crystalline phase.

IR spectroscopy provided the first indication that a new metal-organic product had formed. As shown in Fig. 15, the spectrum of the isolated solid differs from that of the free ligand, metal precursor, as well as solvent and exhibits characteristic changes expected upon coordination to the zirconium cluster. This supports successful reaction of the linker under the predicted screening conditions, although IR spectroscopy alone cannot distinguish unambiguously between the possible capsule and tetrahedral cage configurations.

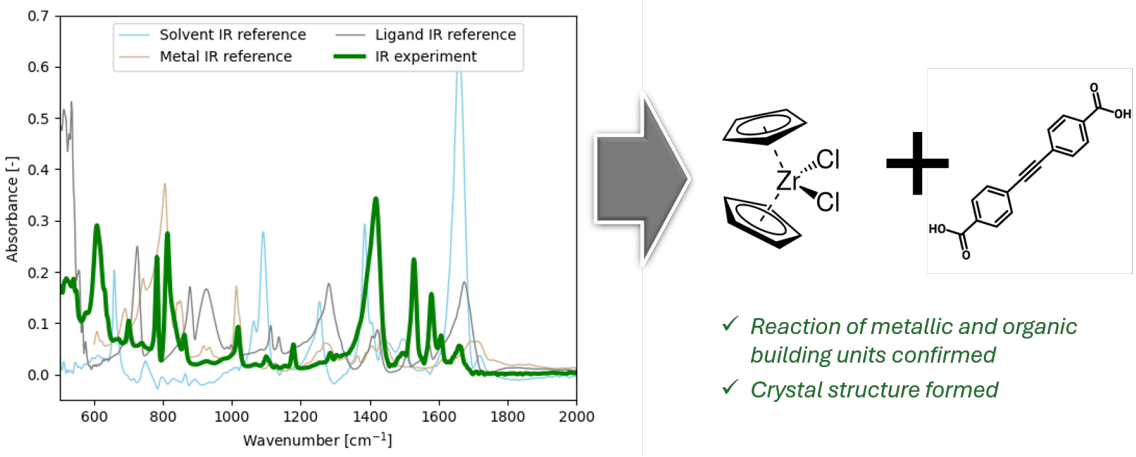


Figure 15: *Comparison of the measured IR spectrum of the synthesised Zr-EDB-MOP product with reference spectra of the ligand, metal precursor, and solvent.*

HRMS was subsequently used to verify the molecular composition of the synthesised product and distinguish between the alternative cage configurations discussed in Section 4.2. Fig. 16 compares the experimentally observed isotope patterns with the theoretical distributions calculated for the capsule-shaped M_2L_3 assembly and the tetrahedral M_4L_6 assembly. The measured spectra exhibit a significantly closer correspondence to the

predicted isotope envelopes of the M_2L_3 species, while no convincing signals attributable to the larger M_4L_6 architecture were observed. These results indicate that the reaction preferentially yielded the capsule-shaped Zr-EDB-MOP and demonstrate the ability of HRMS to discriminate between competing MOP topologies.

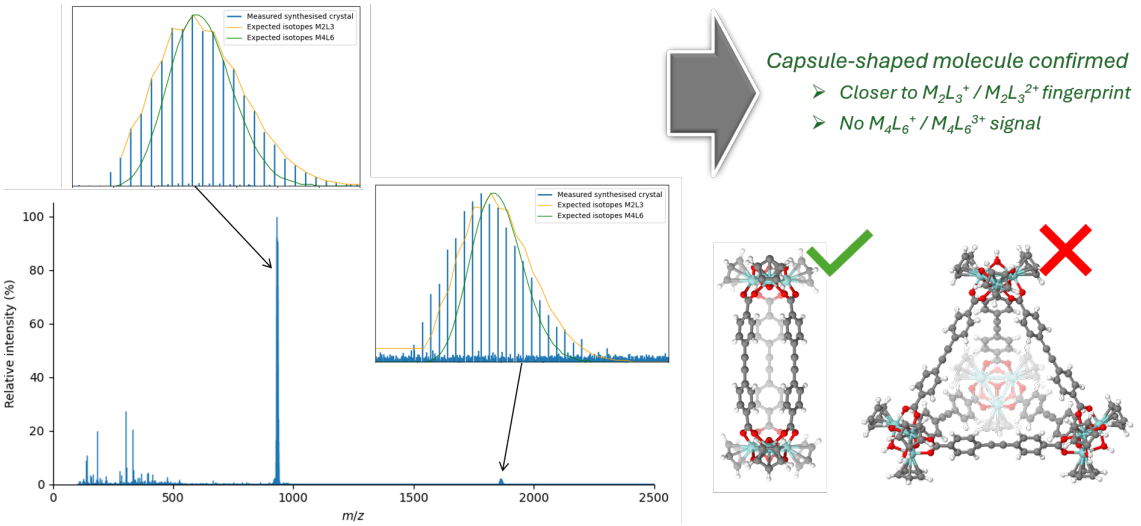


Figure 16: *HRMS analysis of the synthesised Zr-EDB-MOP. Experimental isotope patterns are compared with theoretical distributions calculated for the capsule-shaped M_2L_3 and tetrahedral M_4L_6 cage configurations.*

Finally, PXRD was used to compare the crystalline product with the structures predicted by the crystal structure prediction workflow described in Section 4.1. As shown in Fig. 17, the measured diffraction pattern agrees with the simulated pattern of the predicted crystal structure, particularly with respect to the principal peak positions. This agreement provides evidence that the isolated product adopts the predicted crystalline phase. Taken together, the IR, HRMS, and PXRD results provide a first experimental proof-of-concept validation of the integrated prediction-synthesis workflow for Zr-EDB-MOP.

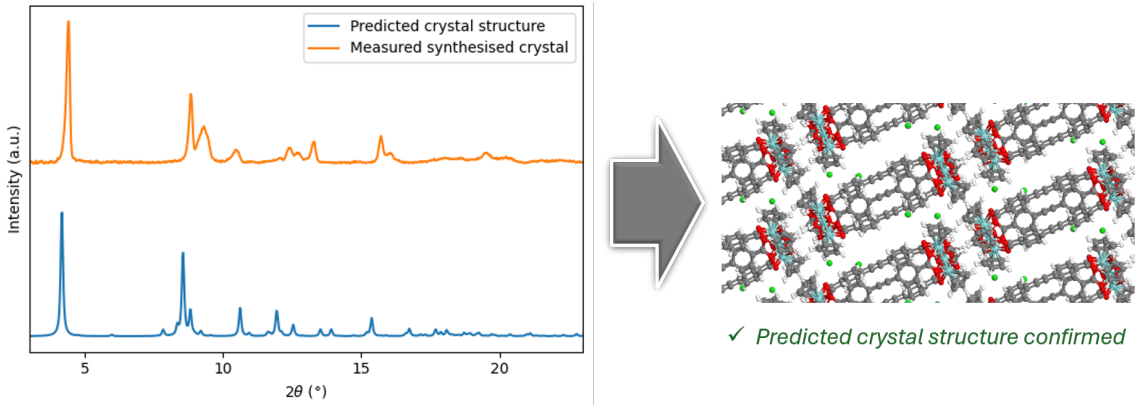


Figure 17: *Comparison of the experimental PXRD pattern of the synthesised Zr-EDB-MOP with the diffraction pattern simulated from the predicted crystal structure.*

6.3 Discussion

The central outcome of this study is not simply a new zirconium-based MOP, but the manner in which it was obtained: the conditions used to synthesise Zr-EDB-MOP were generated from structured prior knowledge rather than supplied by an expert or searched within expert-defined bounds. Candidate generation, structural modelling, computational viability assessment, synthesis-condition prediction, and experimental verification were carried out as a single connected sequence, with each intermediate persisted in a shared knowledge representation. To our knowledge, this is the first demonstration for a reticular material in which every stage of the discovery cycle, up to and including the initial synthesis conditions, is derived from a common machine-readable source rather than initiated by a human-chosen starting point.

This distinction is what separates the present workflow from the closed-loop and autonomous efforts surveyed in Section 2.4. Those platforms explore an experimental space efficiently, but they begin from a supplied target, reaction family, precursor set, or admissible range of conditions; the A-Lab, which comes closest to autonomous initiation, still seeds its recipes from literature similarity before active-learning optimisation. [55] Agentic frameworks for MOFs such as MOFGen likewise combine generative design, computational screening, and experimental agents, and their ligand-switching design strategy parallels our own rational-design approach, [22, 27] yet the experimental conditions that drive synthesis remain externally specified. Here, by contrast, the analogy algorithm of Section 3 produces a small, chemically defensible set of conditions directly from linked structural and procedural precedent, addressing precisely the step that these workflows leave to human judgement.

The result also provides a first concrete instance in support of the argument set out in Section 2.4: that in the low-data, high-regularity corner of synthesis space, explicit structural analogy can substitute for statistical inference. A purely data-driven MOP model would have to generalise from roughly one to two orders of magnitude fewer procedures than even the data-limited MOF studies discussed there, as the approximately 300 procedures in OntoSyn [49] are far too few to train a high-capacity model. What makes the knowledge-based route viable is not the absence of alternatives but the structure of the problem itself: recurring building units, assembly models, and ligand-substitution relationships mean that structurally analogous precedents remain abundant even where procedures are scarce. The same reasoning extends to coordination cages, polyoxometalate clusters, and substantial parts of framework chemistry, and the interpretability of the resulting rules is a practical advantage in its own right, since each predicted condition can be traced back to the precedent and the criterion that produced it.

Implementing this reasoning within The World Avatar, rather than as a standalone script, is what makes the analogy expressible in the first place. A static reaction table records conditions; the dynamic knowledge graph instead links each procedure to the structure, building units, computed geometry, and reference spectra of its product, so that the similarity relationships the algorithm depends on can be queried directly rather than reconstructed by hand. The same representation underlies the practicality of the workflow at the bench: a predicted procedure is stored once as an `OntoSyn` instance and rendered on demand either as a human-readable instruction sheet or as machine-executable laboratory

code, two renderings we have previously shown to be interchangeable. [49] The syntheses reported here were performed manually, which fixes the throughput of this study but not the interface between prediction and execution.

These results should nonetheless be read as a proof of concept rather than a general validation of the prediction model. The prospective demonstration rests on a single zirconium-based system, and the strongest chemical assumptions in the algorithm were developed for a comparatively narrow subset of carboxylate-based MOP chemistry. Establishing how far the approach generalises will require repeated prospective application across a more diverse set of targets, spanning different linkers, assembly models, and metal-based building units, together with quantitative rather than qualitative measures of agreement. Such tests will reveal which of the present rules transfer unchanged, which require refinement, and which will ultimately be better served by data-driven components as the underlying synthesis corpus grows.

6.4 Applications

This pipeline can now be applied to design, compute, verify, and synthesise MOPs for specific applications by targeting and optimising relevant properties. Such optimisation can be achieved either by sampling from the space of already predicted MOPs [27] or by deploying evolutionary algorithms to adapt existing MOPs or their constituent CBUs.

One compelling application is electrochemical CO_2 reduction, where MOFs have already demonstrated strong performance [45], and MOPs have recently been explored for photocatalytic CO_2 reduction [1, 35]. MOPs can be utilised in this context in several ways: as direct catalysts *via* their catalytic metal cores, through functionalised organic ligands that promote catalysis, or as nanoscale containers encapsulating other catalytic materials.

A key challenge for such electronically driven applications is the cost and reliability of computational screening, since relevant descriptors generally require at least semi-empiricial-level electronic-structure calculations. Surrogate models trained on strategically selected calculations may therefore be needed to explore large candidate sets efficiently [58]. By contrast, applications such as gas capture and separation may be more immediately accessible, as many relevant descriptors are textural properties, including pore and cavity dimensions [58]. Preliminary studies in this direction already indicate the potential for rapid screening of MOP candidates [28].

6.5 Future work

Future work should address the main limitations identified in this proof-of-concept study and convert the current semi-manual workflow into a more autonomous discovery cycle within The World Avatar. A first priority is the refinement of synthesis-condition prediction, particularly for solvents and solvent amounts. While the present analogy-based rules produced experimentally useful screening conditions, the retrospective comparisons in Section 3.3 showed that solvent transferability remains one of the least robust parts of the algorithm. Future versions should therefore incorporate additional solvent descriptors,

such as polarity, hydrogen-bond donor and acceptor counts, boiling point, and water–octanol partition coefficient, following related approaches for synthesis prediction [36]. Similar extensions could be made for modulators, feeding procedures, hydrolysis kinetics, and additive acidity, as recommended for digital reticular chemistry by Lyu et al. [37].

A second priority is to automate the discovery and refinement of the prediction rules themselves. In the present work, the rules were derived manually from analysis of the structured synthesis data and chemical reasoning. Within TWA, this process could be supported by agents that query the KG, identify statistically recurring relationships between MOP structures and synthesis procedures, propose candidate rules, and test them against known or newly generated experimental results. Such agent-based rule discovery would preserve the interpretability of the current approach while reducing reliance on manual inspection and enabling the rule base to evolve as additional synthesis data become available.

The experimental side of the workflow should also be made more autonomous. In this study, the predicted protocols were executed manually, with characterisation data interpreted through a hierarchical verification workflow. Future integration with automated laboratory infrastructure would allow predicted conditions to be translated directly into executable experiments, while both successful and unsuccessful outcomes are instantiated back into the KG. This would enable active learning over failed experiments, allowing the system to refine synthesis windows, avoid unproductive regions of parameter space, and prioritise follow-up experiments based on accumulated evidence rather than only successful examples.

Broader validation will require a larger benchmark set covering different linkers, assembly models, metal-based CBUs, and less closely analogous reference procedures. This should include not only additional prospective syntheses but also more quantitative scoring of experimental agreement. For example, PXRD comparisons could be standardised using automated peak matching or difference-score metrics, while HRMS verification could be strengthened by quantitative isotope-pattern matching and confidence scoring. These metrics would allow the workflow to move beyond qualitative pass/fail assessments and provide a more systematic basis for comparing candidate structures, predicted crystal phases, and experimental products.

In the longer term, these developments could support inverse MOP design, analogous to inverse MOF design strategies based on generative and evolutionary methods [44, 58]. However, such approaches will only be useful if combined with a reliable synthetic feasibility metric. The knowledge-graph-driven workflow developed here provides a starting point for such a metric, because it links structural design, synthesis precedent, computational viability, and experimental verification within a single semantic framework. Once validated across a broader chemical space, the same principles could also be extended to other reticular materials, including MOFs and COFs, where synthesis planning would benefit from similarly rich and connected knowledge representations.

7 Conclusions

In this work, we developed and demonstrated an integrated workflow for the prediction and synthesis of novel metal–organic polyhedra within The World Avatar. Building on previously developed knowledge-graph representations of MOP structures and synthesis procedures, we introduced a knowledge-based algorithm that predicts synthesis conditions by analogy to known procedures. The approach uses topological, chemical, and procedural similarity criteria to infer experimentally actionable protocols for new MOP candidates. Retrospective comparison against known or subsequently reported Zr-based MOP syntheses showed promising agreement between predicted and reported conditions, particularly in terms of reactant stoichiometry, heating regime, and general work-up procedure.

The central result of this study is the prospective synthesis and characterisation of a novel Zr-EDB-MOP using conditions derived from this prediction framework. Computational screening and CSP were used to assess plausible molecular and crystalline structures, while IR spectroscopy, HRMS, and PXRD provided complementary experimental evidence for product formation, target composition, and agreement with the predicted crystalline phase. This provides a first experimental proof-of-concept validation of the integrated workflow, demonstrating that knowledge-graph-based synthesis prediction, computational screening, and experimental verification can be combined into a single discovery cycle.

More broadly, the results show that for highly regular reticular materials, chemical reasoning based on structural analogy can be encoded and automated to a significant degree. Such rule-centric approaches are especially valuable for compound classes such as MOPs, where data remain too sparse for robust black-box machine-learning models, but where recurring building units, assembly models, and synthesis motifs provide a strong basis for interpretable prediction. In future work, these explicit rules can be refined, extended, and combined with stochastic or generative models, while remaining grounded in semantic knowledge representation.

Overall, this study closes the loop from knowledge-graph-based MOP design and synthesis prediction to computational evaluation, experimental synthesis, and structural verification of a novel MOP. It therefore represents a step towards inverse design in digital reticular chemistry, where desired structures or properties can be linked not only to hypothetical candidates, but also to plausible and experimentally testable synthesis routes.

Acknowledgements

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. For the purpose of open access, the author has applied a Creative Commons Attribution (CC BY) licence to any Author Accepted Manuscript version arising.

Declaration of Generative AI and AI-assisted technologies in the writing process

During the preparation of this work the authors used ChatGPT in order to enhance the readability and language of the manuscript. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

Data and code availability

The synthesis prediction software code developed within this work is publicly available on Github *via* https://github.com/TheWorldAvatar/MOPTools.

The modified mol-CSPy with the GULP interface is available at https://gitlab.com/pwvbutler/mol-cspy. The input files for generating the trial structures and performing the optimisations are provided in the supplementary data along with the final predicted crystal structures.

Materials and methods

All chemicals and solvents were obtained from commercial suppliers, MOPs were synthesised and isolated under the conditions specified in Appendix B. FTIR spectra were recorded using a Bruker VERTEX 80 spectrometer equipped with the solid-state sampling technique. High-resolution mass spectra were acquired in positive electrospray ionisation mode on a Waters UPLC-Q-TOF time-of-flight mass spectrometer. PXRD patterns were recorded at room temperature using a Rigaku MiniFlex 600C X-ray diffractometer equipped with a D/teX100 detector using Cu $K\alpha$ radiation (40 kV, 30 mA) over a 2θ range of 3° to 60° with a scan rate of $10^\circ\,\mathrm{min}^{-1}$ and a step size of 0.01°.

Nomenclature

Upper-case Roman

A	assembly model
L	organic ligand
M	metallic CBU
$\tilde{M}$	molar mass
P	provenance
R	reactant
S	solvent
T	heating regime
Y	specific reactant volume

Lower-case Roman

f	excess factor
m	reactant mass
n	amount of substance

Subscripts

x	Pertaining to the novel MOP
0	Pertaining to known MOP with same metal as MOP_x
1	Pertaining to known MOP with same ligand as MOP_x
2	Pertaining to known MOP with different metal and ligand than MOP_x

Abbreviations

AM	Assembly model
BDC	Benzene-dicarboxylic
BPDC	Biphenyl-dicarboxylic
BPTC	Biphenyl-tetracarboxylic
BPyDC	Bipyridine-dicarboxylic
BTB	Benzene-triyl-tribenzoic
CBU	Chemical building unit
COF	Covalent organic framework
CSP	Crystal structure prediction
DEF	Diethylformamide
DFT	Density-functional theory
DMA	Dimethylacetamide
DMF	Dimethylformamide
EDB	Ethyne-diyl-dibenzoic
[FT]IR	[Fourier-transformed] Infrared (spectroscopy)
GBU	Generic building unit
[HR]MS	High-resolution mass spectrometry
KG	Knowledge graph
ML	Machine Learning
MOF	Metal-organic framework
MOP	Metal-organic polyhedron
POV	Polyoxovanadate
[P]XRD	[Powder] X-ray diffraction (spectroscopy)
TATB	Triazine-trizl-tribenyoic
TWA	The World Avatar
xTB	Extended tight-binding (DFT)

A Solvent prediction

Predicting suitable solvents for synthesis is one of the most challenging tasks of the presented workflow. Thorough analysis of the known synthesis procedures regarding solvent use requires double-federated queries across 3 KGs as shown in listing 1.

1. `OntoSyn`: Knowledge about reactants used for synthesis of certain MOPs
2. `OntoMOPs`: Knowledge about related CBUs, GBUs and assembly models
3. `OntoSpecies`: Knowledge about reactants

Listing 1: *SPARQL query for Fig. A.1.*

```
SELECT DISTINCT ?mop ?cbu ?gbu ?species
WHERE {

  SERVICE <http://.../OntoMOPs/sparql> {
    ?mop a mops:MetalOrganicPolyhedron ;
           ...
           mops:hasChemicalBuildingUnit ?cbu .
    ?cbu mops:isFunctioningAs ?gbu .
    ...
    }

  ?chemOut syn:isRepresentedBy ?mop .
  ?chemTrans syn:hasChemicalOutput ?chemOut ;
             syn:isDescribedBy ?chemSyn .
  ?chemSyn a syn:ChemicalSynthesis ;
             syn:hasChemicalInput/.../.../.../...
              ?species .
  ...

  SERVICE <http://.../ontospecies/sparql> {
    ?species os:hasUse ?use          .
    ?use rdfs:label ?useLabel .
    FILTER regex(lcase(str(?useLabel)), "solvent") .
    }

}
```

The resulting overview of solvents used for known MOP synthesis procedures is given in Fig. A.1.

CBU	occurences	acetone	Acetonitrile	AcOEt	CH3OH	Ethanol	chloroform	cyclohexane	DCM	DEF	diethyl ether	DMA	DMF	Triethylamine	water
[Co4C40H44O12S4]	3	0%	0%	0%	67%	0%	33%	0%	0%	0%	0%	33%	67%	33%	0%
[Co4C40H44O4S4]	3	0%	0%	33%	33%	0%	0%	0%	0%	0%	33%	33%	67%	67%	0%
[Cr2]	3	0%	0%	0%	33%	0%	0%	0%	0%	67%	0%	33%	33%	0%	0%
[Cu2]	24	8%	4%	0%	58%	50%	0%	25%	4%	21%	8%	25%	67%	4%	13%
[Fe3O(SO4)3(C5H5N)3]	5	0%	0%	0%	0%	0%	0%	80%	20%	0%	0%	0%	100%	100%	0%
[Ni4C40H44S4O12]	4	25%	0%	0%	50%	25%	50%	0%	25%	0%	0%	0%	75%	0%	50%
[Rh2]	3	100%	0%	0%	100%	67%	0%	0%	33%	0%	0%	100%	33%	0%	33%
[Ru2]	2	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	100%	0%	0%
[V5O9]	4	0%	0%	0%	25%	75%	0%	0%	0%	0%	0%	25%	75%	0%	25%
[V6O6(OCH3)9(SO4)]	6	0%	0%	0%	100%	17%	0%	0%	0%	67%	0%	0%	33%	0%	0%
[V7O10(OCH3)9]	3	0%	0%	0%	100%	33%	0%	0%	0%	67%	0%	0%	33%	0%	67%
[WV5O11]	4	0%	75%	0%	25%	50%	0%	0%	0%	0%	0%	0%	100%	0%	25%
[Zr3O(OH)3(C5H5)3]	10	0%	0%	0%	20%	0%	0%	0%	30%	10%	0%	30%	70%	0%	100%

(a) *Metal CBUs.*

CBU	occurences	acetone	Acetonitrile	benzene	CH3OH	Ethanol	chloroform	cyclohexane	DCM	DEF	DMA	DMF	Triethylamine	water
[((C6H3)C(CH3)3)(CO2)2]	4	0%	0%	0%	25%	0%	0%	0%	0%	0%	100%	0%	0%	0%
[(C10H6)(CO2)2]	2	0%	50%	0%	0%	100%	0%	50%	0%	50%	0%	100%	0%	0%
[(C3N3)(C6H4)3(CO2)3]	3	0%	33%	0%	100%	67%	0%	0%	0%	0%	0%	100%	0%	0%
[(C4H2S)(CO2)2]	2	0%	0%	0%	100%	50%	100%	0%	50%	0%	0%	0%	0%	50%
[(C5NH3)2(CO2)2]	2	0%	0%	0%	100%	0%	0%	0%	0%	0%	0%	100%	100%	0%
[(C6H3)(C6H4)3(CO2)3]	3	0%	0%	0%	33%	0%	0%	33%	0%	33%	33%	33%	33%	33%
[(C6H3)(CO2)3]	12	0%	42%	0%	25%	0%	25%	0%	33%	8%	17%	50%	8%	33%
[(C6H3)2NH(CO2)2]	4	0%	0%	0%	75%	25%	0%	0%	0%	25%	0%	75%	0%	0%
[(C6H3Br)(CO2)2]	3	0%	0%	0%	33%	33%	0%	33%	0%	33%	0%	67%	33%	0%
[(C6H3NH2)(CO2)2]	5	0%	0%	0%	40%	40%	0%	20%	0%	60%	0%	40%	0%	40%
[(C6H3OH)(CO2)2]	3	67%	0%	33%	100%	0%	0%	0%	0%	0%	100%	0%	0%	33%
[(C6H4)(CO2)2]	10	10%	10%	0%	40%	40%	0%	10%	0%	20%	20%	100%	10%	20%
[(C6H4)2(CO2)2]	2	0%	0%	0%	0%	0%	0%	50%	0%	0%	50%	50%	50%	50%
[C6H3C2Si(C3H7)3(CO2)2]	2	0%	0%	50%	50%	0%	0%	0%	0%	50%	0%	0%	0%	0%

(b) *Organic CBUs.*

Figure A.1: *Correlation of solvents and respective CBUs.*

A slight correlation is apparent but further analysis needed. For example, material amounts and concentrations need to be taken into account. Furthermore, species properties like dielectric constants should be compared.

B Experimental conditions

The chemicals and solvents were purchased from commercial sources and used without further purification.

All syntheses followed the same general procedure: zirconocene dichloride and the respective ligand were dissolved in the solvent along with a few drops of distilled water. The mixture was stirred, heated in an oil bath at 60 °C to 65 °C for 8 h to 12 h , and then cooled slowly. The resulting precipitate was collected by centrifugation (5000 rpm, 6 min), and the supernatant was decanted. The solid was washed by resuspension in the synthesis solvent (10 mL), followed by centrifugation and decantation. This washing procedure was repeated three times with dichloromethane (10 mL), and the solid obtained was dried under vacuum.

The individual reaction conditions are summarised in Tab. B.1. Procedures labelled as *reported* were queried from literature-derived synthesis procedures instantiated in TWA, while procedures labelled as *predicted* were generated by the analogy-based algorithm presented in section 3.

Table B.1: *Experimental synthesis conditions for all MOPs prepared in this work.* Reported *procedures originate from the literature (see subsections below for sources and applied scale-up),* predicted *procedures were generated by the synthesis prediction algorithm presented in section 3.*

Name	Origin	Cp_2ZrCl_2	Ligand	Solvent	Heating	Yield
Zr-BDC	Reported	87.5 mg	25 mg H_2BDC	2.5 mL DMF	60 °C for 8 h	23 mg
Zr-BTC	Reported	120 mg	100 mg H_3BTC	10 mL DMF	65 °C for 10 h	95 mg
Zr-BPDC-1	Reported	120 mg	60 mg H_2PDC	4.5 mL DMA	65 °C for 12 h	< 5 mg
Zr-BPDC-2	Predicted	50 mg	21 mg H_2BPDC	1.2 mL DMF	60 °C for 8 h	< 5 mg
Zr-BPyDC-1	Predicted	50 mg	14.7 mg H_2BPyDC	1.5 mL DMF	60 °C for 8 h	< 5 mg
Zr-BPyDC-2	Predicted	50 mg	61.7 mg H_2BPyDC	3.6 mL DMF	65 °C for 10 h	< 5 mg
Zr-EDB-1	Predicted	50 mg	11.5 mg H_2EDB	1.9 mL DMF	60 °C for 8 h	39 mg
Zr-EDB-2	Predicted	50 mg	26.5 mg H_2EDB	3.2 mL DMF	65 °C for 10 h	50 mg

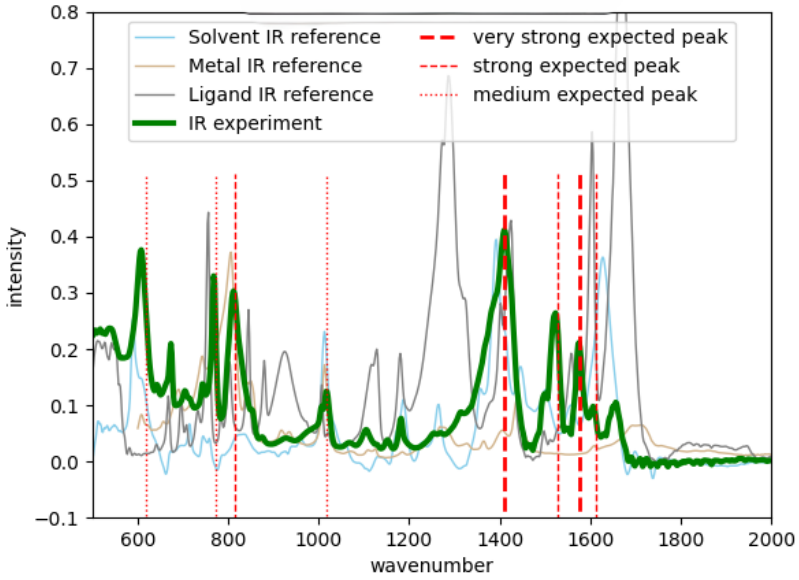
B.1 Zr-BDC-MOP and Zr-BTC-MOP

The procedures were taken from Liu et al. [33] and scaled up by a factor of five (ligands and solvents) to provide sufficient material for characterisation and screening. These two structures serve as the experimentally validated references (MOP_0) for the synthesis prediction algorithm, as described in sections 3 and 5.1.

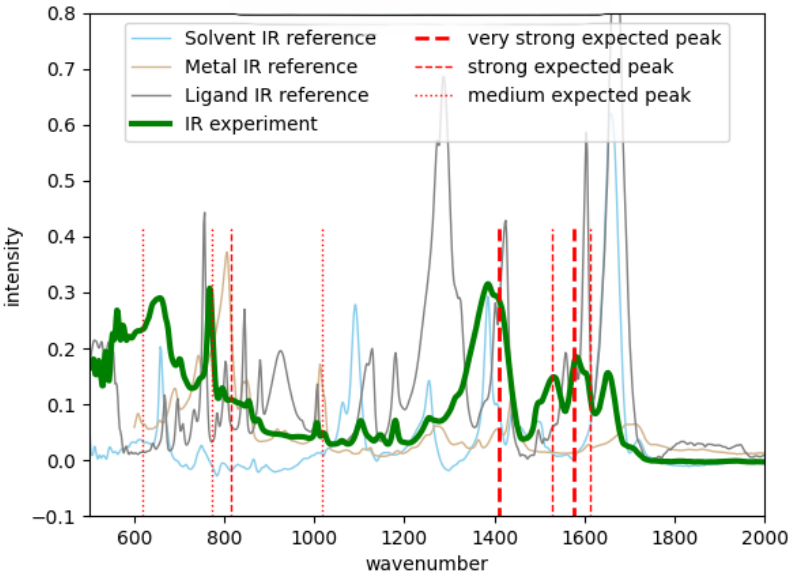
B.2 Zr-BPDC-MOP

The Zr-BPDC-1 procedure was taken from Liu et al. [33] and scaled up by a factor of three (ligands and solvents). The reported procedure specifies CCl_4, which was found

ineffective in our experiments and therefore omitted. Zr-BPDC-2 follows the procedure predicted by our algorithm. Fig. B.1 shows the FTIR spectra of both products. While the reported procedure yields the better result, the predicted procedure also appears to produce the desired material in some capacity, despite differing in solvent and heating regime. This supports the argument made in section 3.3 that predicted procedures deviating from a reported one may nevertheless constitute valid syntheses.



(a) *Zr-BPDC-1*

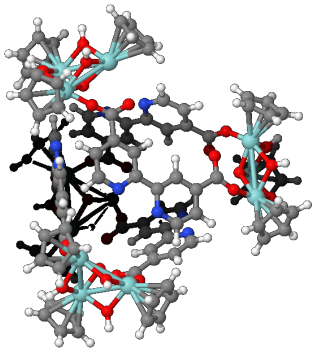


(b) *Zr-BPDC-2*

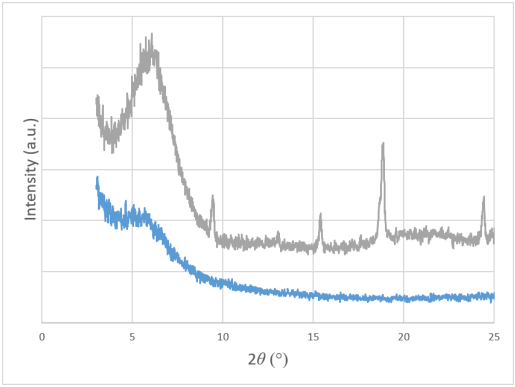
Figure B.1: *FTIR spectra of the products obtained from the reported procedure Zr-BPDC-1 and the predicted procedure Zr-BPDC-2.*

B.3 Zr-BPyDC-MOP

Computational screening flagged the assembled geometry as problematic (Fig. B.2a; cf. section 4). Consistent with this assessment, experimental screening across the suggested ranges of both predicted synthesis routes did not result in crystal formation: the PXRD patterns in Fig. B.2b show no peaks at low angles. The patterns shown are representative; a number of further experiments with varying reactant amounts, solvents, and additives were carried out with the same outcome.



(a) *Assembled geometry.*



(b) *Zr-BPyDC-1 (blue) and Zr-BPyDC-2 (gray).*

Figure B.2: *Attempted Zr-BPDyDC-MOP synthesis with assembled geometry flagged as problematic by the computational screening and representative PXRD patterns of the products obtained from the two predicted synthesis routes showing no reflections at low angles.*

B.4 Zr-EDB-MOP

Both predicted synthesis routes yielded crystalline product. Fig. B.3 compares the PXRD patterns of the two procedures: Zr-EDB-2 exhibits clearer peaks than Zr-EDB-1 and was therefore selected for the detailed verification presented in section 6.

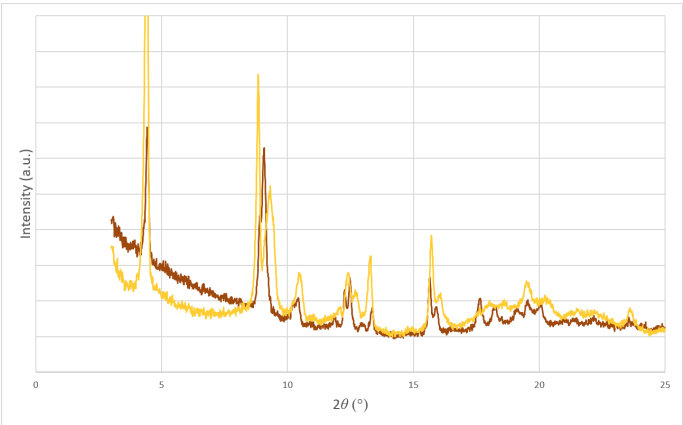


Figure B.3: *PXRD patterns of the products obtained from the two predicted procedures Zr-EDB-1 (brown) and Zr-EDB-2 (yellow).*

C Predicted crystal structures

Below are the energy density landscapes of the predicted structures. The match to the experimental structure is indicated by a red square.

C.1 Zr-BDC-MOP

Energy density of capsule structure is shown in Fig. C.1

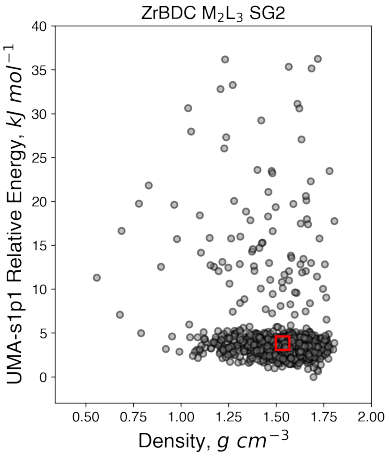
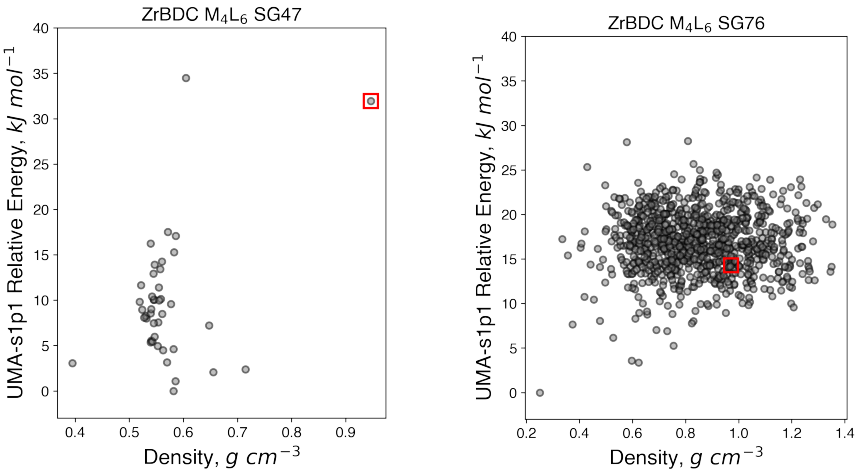


Figure C.1: *Energy landscape of potential crystal structure for capsule Zr-BDC-MOP with space group 2.*

Energy density of tetrahedral structure is shown in Fig. C.2.



(a) *Space group 47.* **(b)** *Space group 76.*

Figure C.2: *Energy landscapes of potential crystal structure for tetrahedral Zr-BDC-MOP with different space groups.*

C.2 Zr-BPDC-MOP

Energy density of tetrahedral structure is shown in Fig. C.3

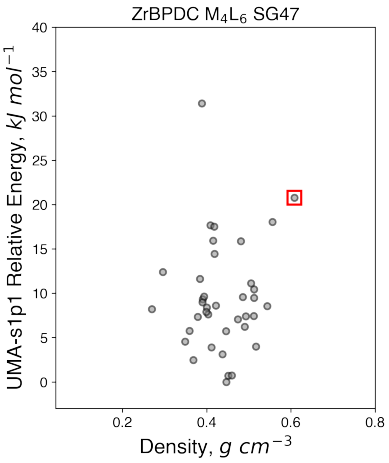
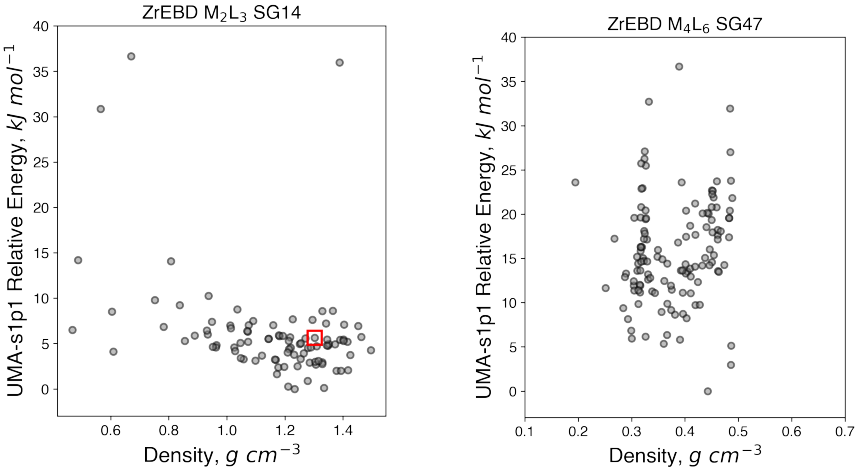


Figure C.3: *Energy landscape of potential crystal structure for tetrahedral Zr-BPDC-MOP with space group 47.*

C.3 Zr-EDB-MOP

Energy density of capsule structure is shown in Fig. C.4a. Energy density of a theoretical tetrahedral structure is shown in Fig. C.4b (not experimentally observed).



(a) *Capsule (space group 14).* **(b)** *Tetrahedral (space group 47).*

Figure C.4: *Energy landscapes of potential crystal structures for different Zr-EDB-MOP morphologies and respective space groups.*

References

[1] A. O. Adeola, J. O. Ighalo, P. I. Kyesmen, and P. 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, Feb 2024. doi:10.1016/j.jcou.2023.102664.

[2] J. Akroyd, S. Mosbach, A. Bhave, and M. Kraft. Universal Digital Twin - A Dynamic Knowledge Graph. *Data-Centric Engineering*, 2(4):1–25, 2021. doi:10.1017/dce.2021.10.

[3] J. Bai, K. F. Lee, M. Hofmeister, S. Mosbach, J. Akroyd, and M. Kraft. A derived information framework for a dynamic knowledge graph and its application to smart cities. *Future Generation Computer Systems*, 152(August 2023):112–126, 2024. doi:10.1016/j.future.2023.10.008.

[4] J. Bai, S. Mosbach, C. J. Taylor, D. Karan, K. F. Lee, S. D. Rihm, J. Akroyd, A. A. Lapkin, and M. Kraft. A dynamic knowledge graph approach to distributed self-driving laboratories. *Nature Communications 2024*, 15(1):1–14, 2024. doi:10.1038/s41467-023-44599-9.

[5] J. Bai, S. D. Rihm, A. Kondinski, F. Saluz, X. Deng, G. Brownbridge, S. Mosbach, J. Akroyd, and M. Kraft. twa: The World Avatar Python package for dynamic knowledge graphs and its application in reticular chemistry. *Digital Discovery*, 4(8): 2123–2135, 2025. doi:10.1039/D5DD00069F.

[6] C. Bannwarth, S. Ehlert, and S. Grimme. GFN2-xTB - An Accurate and Broadly Parametrized Self-Consistent Tight-Binding Quantum Chemical Method with Multipole Electrostatics and Density-Dependent Dispersion Contributions. *Journal of Chemical Theory and Computation*, 15:1652–1671, 3 2019. doi:10.1021/acs.jctc.8b01176.

[7] D. A. Boiko, R. MacKnight, B. Kline, and G. Gomes. Autonomous chemical research with large language models. *Nature*, 624(7992):570–578, Dec. 2023. doi:10.1038/s41586-023-06792-0.

[8] B. Burger, P. M. Maffettone, V. V. Gusev, C. M. Aitchison, Y. Bai, X. Wang, X. Li, B. M. Alston, B. Li, R. Clowes, N. Rankin, B. Harris, R. S. Sprick, and A. I. Cooper. A mobile robotic chemist. *Nature*, 583(7815):237–241, July 2020. doi:10.1038/s41586-020-2442-2.

[9] P. W. Butler, S. D. Rihm, S. Mosbach, J. Akroyd, and M. Kraft. Inverse Design of Metal-Organic Polyhedra through Molecular Fragmentation and Evolutionary Optimisation. *Journal of Chemical Information and Modeling*, 66(7):3933–3943, Apr 2026. doi:10.1021/acs.jcim.5c02956.

[10] P. W. Butler, A. Subramanian, S. D. Rihm, A. F. Fischer, T. S. Choksi, A. Kondinski, S. Mosbach, J. Akroyd, and M. Kraft. Computational Modelling Workflows for Metal-Organic Polyhedra in The World Avatar. *Physical Chemistry Chemical Physics*, 28(24):14828–14838, 06 2026. doi:10.1039/d5cp04871k.

[11] A. Carné-Sánchez, J. Albalad, T. Grancha, I. Imaz, J. Juanhuix, P. Larpent, S. Furukawa, and D. Maspoch. Postsynthetic covalent and coordination functionalization of rhodium(ii)-based metal–organic polyhedra. *Journal of the American Chemical Society*, 141(9):4094–4102, 2019. doi:10.1021/jacs.8b13593.

[12] D. H. Case, J. E. Campbell, P. J. Bygrave, and G. M. Day. Convergence properties of crystal structure prediction by quasi-random sampling. *Journal of Chemical Theory and Computation*, 12:910–924, 2 2016. doi:10.1021/acs.jctc.5b01112.

[13] X. Chen, S.-B. Li, Z.-Y. Liu, and Y.-T. Zhang. Solvent-directed assembly of Zr-based metal-organic cages for dye adsorption from aqueous solution. *Journal of Solid State Chemistry*, 296:121998, 2021. doi:10.1016/j.jssc.2021.121998.

[14] C. W. Coley, L. Rogers, W. H. Green, and K. F. Jensen. Computer-Assisted Retrosynthesis Based on Molecular Similarity. *ACS Central Science*, 3(12):1237–1245, 2017. doi:10.1021/acscentsci.7b00355.

[15] C. W. Coley, D. A. Thomas, J. A. M. Lummiss, J. N. Jaworski, C. P. Breen, V. Schultz, T. Hart, J. S. Fishman, L. Rogers, H. Gao, R. W. Hicklin, P. P. Plehiers, J. Byington, J. S. Piotti, W. H. Green, A. J. Hart, T. F. Jamison, and K. F. Jensen. A robotic platform for flow synthesis of organic compounds informed by AI planning. *Science*, 365(6453):eaax1566, Aug. 2019. doi:10.1126/science.aax1566.

[16] D. E. Coupry, M. A. Addicoat, and T. Heine. Extension of the Universal Force Field for Metal-Organic Frameworks. *Journal of Chemical Theory and Computation*, 12: 5215–5225, 10 2016. doi:10.1021/acs.jctc.6b00664.

[17] J. D. Gale and A. L. Rohl. The General Utility Lattice Program (GULP). *Molecular Simulation*, 29:291–341, 1 2003. doi:10.1080/0892702031000104887.

[18] L. T. Glasby, K. Gubsch, R. Bence, R. Oktavian, K. Isoko, S. M. Moosavi, J. L. Cordiner, J. C. Cole, and P. Z. Moghadam. DigiMOF: A Database of Metal–Organic Framework Synthesis Information Generated via Text Mining. *Chemistry of Materials*, 35(11):4510–4524, June 2023. doi:10.1021/acs.chemmater.3c00788.

[19] A. J. Gosselin, G. E. Decker, B. W. McNichols, J. E. Baumann, G. P. A. Yap, A. Sellinger, and E. D. Bloch. Ligand-based phase control in porous zirconium coordination cages. *Chemistry of Materials*, 32(13):5872–5878, 2020. doi:10.1021/acs.chemmater.0c01965.

[20] A. J. Gosselin, C. A. Rowland, and E. D. Bloch. Permanently Microporous Metal–Organic Polyhedra. *Chemical Reviews*, 120(16):8987–9014, Aug 2020. doi:10.1021/acs.chemrev.9b00803.

[21] R. L. Greenaway, V. Santolini, M. J. Bennison, B. M. Alston, C. J. Pugh, M. A. Little, M. Miklitz, E. G. B. Eden-Rump, R. Clowes, A. Shakil, H. J. Cuthbertson, H. Armstrong, M. E. Briggs, K. E. Jelfs, and A. I. Cooper. High-throughput discovery of organic cages and catenanes using computational screening fused with robotic synthesis. *Nature Communications*, 9(1):2849, July 2018. doi:10.1038/s41467-018-05271-9.

[22] T. J. Inizan, S. Yang, A. Kaplan, Y. hsu Lin, J. Yin, S. Mirzaei, M. Abdelgaid, A. H. Alawadhi, K. Cho, Z. Zheng, E. D. Cubuk, C. Borgs, J. T. Chayes, K. A. Persson, and O. M. Yaghi. System of Agentic AI for the Discovery of Metal-Organic Frameworks, 2025. URL https://arxiv.org/abs/2504.14110.

[23] S. M. Kearnes, M. R. Maser, M. Wleklinski, A. Kast, A. G. Doyle, S. D. Dreher, J. M. Hawkins, K. F. Jensen, and C. W. Coley. The Open Reaction Database. *Journal of the American Chemical Society*, 143(45):18820–18826, Nov. 2021. doi:10.1021/jacs.1c09820.

[24] K. Kiaei, K. Brunson, A. Gładysiak, K. Smith, K. Hunter, A. Thomas, D. Radke, T. Zuehlsdorff, and K. C. Stylianou. Instantaneous degradation of nerve agent simulants using zirconium-based metal–organic polyhedra. *Journal of Materials Chemistry A*, 11:14265–14271, 2023. doi:10.1039/D3TA02014B.

[25] E. Kim, K. Huang, A. Tomala, S. Matthews, E. Strubell, A. Saunders, A. McCallum, and E. Olivetti. Machine-learned and codified synthesis parameters of oxide materials. *Scientific Data*, 4(1):170127, Sept. 2017. doi:10.1038/sdata.2017.127.

[26] T. Klucznik, B. Mikulak-Klucznik, M. P. McCormack, H. Lima, S. Szymkuć, M. Bhowmick, K. Molga, Y. Zhou, L. Rickershauser, E. P. Gajewska, A. Toutchkine, P. Dittwald, M. P. Startek, G. J. Kirkovits, R. Roszak, A. Adamski, B. Sieredzińska, M. Mrksich, S. L. Trice, and B. A. Grzybowski. Efficient Syntheses of Diverse, Medicinally Relevant Targets Planned by Computer and Executed in the Laboratory. *Chem*, 4(3):522–532, Mar. 2018. doi:10.1016/j.chempr.2018.02.002.

[27] A. Kondinski, A. Menon, D. Nurkowski, F. Farazi, S. Mosbach, J. Akroyd, and M. Kraft. Automated Rational Design of Metal-Organic Polyhedra. *Journal of the American Chemical Society*, 144(26):11713–11728, Jul 2022. doi:10.1021/jacs.2c03402.

[28] A. Kondinski, A. M. Oyarzún-Aravena, S. D. Rihm, J. Bai, S. Mosbach, J. Akroyd, and M. Kraft. Automated Assembly Modelling of Metal-Organic Polyhedra. *European Journal of Inorganic Chemistry*, 28(26):e202500115, 2025. doi:10.1002/ejic.202500115.

[29] O. Kononova, H. Huo, T. He, Z. Rong, T. Botari, W. Sun, V. Tshitoyan, and G. Ceder. Text-mined dataset of inorganic materials synthesis recipes. *Scientific Data*, 6(1): 203, Oct. 2019. doi:10.1038/s41597-019-0224-1.

[30] N. Krdzavac, S. Mosbach, D. Nurkowski, P. Buerger, J. Akroyd, J. Martin, A. Menon, and M. Kraft. An ontology and semantic web service for quantum chemistry calculations. *Journal of Chemical Information and Modeling*, 59(7):3154–3165, 2019. doi:10.1021/acs.jcim.9b00227.

[31] S. Lee, H. Jeong, D. Nam, M. S. Lah, and W. Choe. The rise of metal–organic polyhedra. *Chemical Society Reviews*, 50:528–555, 2021. doi:10.1039/D0CS00443J. URL https://doi.org/10.1039/D0CS00443J.

[32] M. Q. Lim, X. Wang, O. Inderwildi, and M. Kraft. The World Avatar – A World Model for Facilitating Interoperability. In O. Inderwildi and M. Kraft, editors, *Intelligent Decarbonisation: Can Artificial Intelligence and Cyber-Physical Systems Help Achieve Climate Mitigation Targets?*, pages 39–53. Springer International Publishing, 2022. doi:10.1007/978-3-030-86215-2_4.

[33] G. Liu, Z. Ju, D. Yuan, and M. Hong. In situ construction of a coordination zirconocene tetrahedron. *Inorganic Chemistry*, 52(24):13815–13817, 2013. doi:10.1021/ic402428m.

[34] G. Liu, Y. Di Yuan, J. Wang, Y. Cheng, S. B. Peh, Y. Wang, Y. Qian, J. Dong, D. Yuan, and D. Zhao. Process-tracing study on the postassembly modification of highly stable zirconium metal–organic cages. *Journal of the American Chemical Society*, 140(20):6231–6234, 2018. doi:10.1021/jacs.8b03517.

[35] Y. Liu, L. Huang, L. Qin, and T. Zhou. Metal-organic cages: Synthetic strategies and photocatalytic application. *ChemCatChem*, 17(1):e202401487, 2025. doi:10.1002/cctc.202401487.

[36] Y. Luo, S. Bag, O. Zaremba, A. Cierpka, J. Andreo, S. Wuttke, P. Friederich, and M. Tsotsalas. MOF Synthesis Prediction Enabled by Automatic Data Mining and Machine Learning. *Angewandte Chemie - International Edition*, 61(19): e202200242, 2022. doi:10.1002/anie.202200242.

[37] H. Lyu, Z. Ji, S. Wuttke, and O. M. Yaghi. Digital Reticular Chemistry. *Chem*, 6(9): 2219–2241, 2020. doi:10.1016/j.chempr.2020.08.008.

[38] R. A. Mayo, A. O.-D. la Roza, and E. R. Johnson. Development and assessment of an improved powder-diffraction-based method for molecular crystal structure similarity. *CrystEngComm*, 24:8326–8338, 11 2022. doi:10.1039/d2ce01080a.

[39] A. Menon, L. Pascazio, D. Nurkowski, F. Farazi, S. Mosbach, J. Akroyd, and M. Kraft. Ontopesscan: An ontology for potential energy surface scans. *ACS Omega*, 8(2):2462–2475, 2023. doi:10.1021/acsomega.2c06948.

[40] S. M. Moosavi, A. Chidambaram, L. Talirz, M. Haranczyk, K. C. Stylianou, and B. Smit. Capturing chemical intuition in synthesis of metal-organic frameworks. *Nature Communications*, 10(1):539, Feb. 2019. doi:10.1038/s41467-019-08483-9.

[41] D. Nam, J. Huh, J. Lee, J. H. Kwak, H. Y. Jeong, K. Choi, and W. Choe. Cross-linking Zr-based metal–organic polyhedra via postsynthetic polymerization. *Chemical Science*, 8(11):7765–7771, 10 2017. doi:10.1039/C7SC03847J.

[42] E. A. Olivetti, J. M. Cole, E. Kim, O. Kononova, G. Ceder, T. Y.-J. Han, and A. M. Hiszpanski. Data-driven materials research enabled by natural language processing and information extraction. *Applied Physics Reviews*, 7(4):041317, Dec. 2020. doi:10.1063/5.0021106.

[43] L. Pascazio, S. Rihm, A. Naseri, S. Mosbach, J. Akroyd, and M. Kraft. Chemical Species Ontology for Data Integration and Knowledge Discovery. *Journal of Chemical Information and Modeling*, 63(21):6569–6586, 2023. doi:10.1021/ACS.JCIM.3C00820.

[44] R. Pollice, G. dos Passos Gomes, M. Aldeghi, R. J. Hickman, M. Krenn, C. Lavigne, M. Lindner-D'Addario, A. Nigam, C. T. Ser, Z. Yao, and A. Aspuru-Guzik. Data-driven strategies for accelerated materials design. *Accounts of Chemical Research*, 54(4):849–860, 2021. doi:10.1021/acs.accounts.0c00785.

[45] C. Qin, X. Li, T. Wang, Z. Xu, K.-J. Chen, and F. Pan. Metal-Organic Frameworks-Based Copper Catalysts for CO_2 Electroreduction Toward Multicarbon Products. *Exploration*, 5(3):270011, 2025. doi:10.1002/EXP.70011.

[46] N. Rampal, D. J. Fu, C. Zhao, H. S. Murayshid, A. A. Abaalkhail, N. E. Alhazmi, M. O. Alawad, C. Borgs, J. T. Chayes, and O. M. Yaghi. An automated evaluation agent for Q&A pairs and reticular synthesis conditions. *Digital Discovery*, 5(1): 231–240, 01 2026. ISSN 2635-098X. doi:10.1039/d5dd00413f.

[47] S. D. Rihm, J. Bai, A. Kondinski, S. Mosbach, J. Akroyd, and M. Kraft. Transforming Research Laboratories with Connected Digital Twins. *Nexus*, 1(1):100004, 2024. doi:10.1016/j.ynexs.2024.100004.

[48] S. D. Rihm, Y. R. Tan, W. Ang, H. Y. Quek, X. Deng, M. T. Laksana, J. Bai, S. Mosbach, J. Akroyd, and M. Kraft. The Digital Lab Facility Manager: Automating operations of research laboratories through "The World Avatar". *Nexus*, 1(3):100031, 2024. doi:10.1016/j.ynexs.2024.100031.

[49] S. D. Rihm, F. Saluz, A. Kondinski, J. Bai, P. W. V. Butler, S. Mosbach, J. Akroyd, and M. Kraft. Extraction of chemical synthesis information using the World Avatar. *Digital Discovery*, 4(10):2893–2909, 2025. doi:10.1039/D5DD00183H.

[50] S. D. Rihm, D. N. Tran, A. Kondinski, L. Pascazio, F. Saluz, X. Deng, S. Mosbach, J. Akroyd, and M. Kraft. Natural Language Access Point to Digital Metal-Organic Polyhedra Chemistry in The World Avatar. *Data-Centric Engineering*, 6:e22, 2025. doi:10.1017/dce.2025.12.

[51] P. Schwaller, T. Laino, T. Gaudin, P. Bolgar, C. A. Hunter, C. Bekas, and A. A. Lee. Molecular Transformer: A Model for Uncertainty-Calibrated Chemical Reaction Prediction. *ACS Central Science*, 5(9):1572–1583, Sept. 2019. doi:10.1021/acscentsci.9b00576.

[52] M. H. S. Segler, M. Preuss, and M. P. Waller. Planning chemical syntheses with deep neural networks and symbolic AI. *Nature*, 555(7698):604–610, Mar. 2018. doi:10.1038/nature25978.

[53] A. J. Stone. Distributed multipole analysis: Stability for large basis sets. *Journal of Chemical Theory and Computation*, 1(6):1128–1132, 2005. doi:10.1021/ct050190+.

[54] A. C. Sudik, A. R. Millward, N. W. Ockwig, A. P. Côté, J. Kim, and O. M. Yaghi. Design, Synthesis, Structure, and Gas (N2, Ar, CO2, CH4, and H2) Sorption Properties of Porous Metal-Organic Tetrahedral and Heterocuboidal Polyhedra. *Journal of the American Chemical Society*, 127(19):7110–7118, 04 2005. doi:10.1021/ja042802q.

[55] N. J. Szymanski, B. Rendy, Y. Fei, R. E. Kumar, T. He, D. Milsted, M. J. McDermott, M. Gallant, E. D. Cubuk, A. Merchant, H. Kim, A. Jain, C. J. Bartel, K. Persson, Y. Zeng, and G. Ceder. An autonomous laboratory for the accelerated synthesis of inorganic materials. *Nature*, 624(7990):86–91, Dec. 2023. doi:10.1038/s41586-023-06734-w.

[56] X.-Q. Wang, M. Zhang, J. Yang, J. Yang, and H. Duan. Efficient removal of dyes from water by zirconium-based metal-organic cages with varying functional groups. *Journal of Molecular Structure*, 1291:136018, 2023. doi:10.1016/j.molstruc.2023.136018.

[57] B. M. Wood, M. Dzamba, X. Fu, M. Gao, M. Shuaibi, L. Barroso-Luque, K. Abdelmaqsoud, V. Gharakhanyan, J. R. Kitchin, D. S. Levine, K. Michel, A. Sriram, T. Cohen, A. Das, A. Rizvi, S. J. Sahoo, Z. W. Ulissi, and C. L. Zitnick. Uma: A family of universal models for atoms, 2025. URL https://arxiv.org/abs/2506.23971.

[58] Z. Yao, B. Sánchez-Lengeling, N. S. Bobbitt, B. J. Bucior, S. G. H. Kumar, S. P. Collins, T. Burns, T. K. Woo, O. K. Farha, R. Q. Snurr, and A. Aspuru-Guzik. Inverse design of nanoporous crystalline reticular materials with deep generative models. *Nature Machine Intelligence*, 3(1):76–86, Jan 2021. doi:10.1038/s42256-020-00271-1.

[59] Y. T. Zhang, S.-B. Li, X.-L. Wang, Y.-R. Gong, K.-Z. Shao, and Z.-M. Su. Synthesis, structures, and magnetic properties of metal-organic polyhedra based on unprecedented $\{V_7\}$ isopolyoxometalate clusters. *Dalton Transactions*, 45(38):14898–14901, 2016. doi:10.1039/C6DT02764D.

[60] Z. Zheng. Reprogramming synthesis. *Science*, 393(6808):253–253, July 2026. doi:10.1126/science.aeh4807.

[61] Z. Zheng, O. Zhang, C. Borgs, J. T. Chayes, and O. M. Yaghi. ChatGPT Chemistry Assistant for Text Mining and the Prediction of MOF Synthesis. *Journal of the American Chemical Society*, 145(32):18048–18062, Aug. 2023. doi:10.1021/jacs.3c05819.

[62] Y.-H. Zou, Q.-J. Wu, Q. Yin, Y.-B. Huang, and R. Cao. Self-Assembly of Imidazolium-Functionalized Zr-Based Metal–Organic Polyhedra for Catalytic Conversion of CO_2 into Cyclic Carbonates. *Inorganic Chemistry*, 60(4):2112–2116, 2021. doi:10.1021/acs.inorgchem.0c01199.