

Automated Assembly Modelling of Metal-Organic Polyhedra

Aleksandar Kondinski^{1,2}, Andrea M. Oyarzún^{1,3}, Simon Rihm¹, Jiaru Bai¹,
Sebastian Mosbach^{1,2}, Jethro Akroyd^{1,2}, Markus Kraft^{1,2,4}

released: November 11, 2024

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

² CARES
Cambridge Centre for Advanced
Research and Education in Singapore
1 Create Way
CREATE Tower, #05-05
Singapore, 138602

³ Departamento de Ingeniería Química
Universidad de Magallanes
Avenida Bulnes 01855
Punta Arenas, 6200000
Chile

⁴ CMCL
No. 9, Journey Campus
Castle Park
Cambridge
CB3 0AX
United Kingdom

Preprint No. 329



Keywords: metal-organic polyhedra, assembly modelling, molecular modelling, iodine capture, pore size, cavity

Edited by

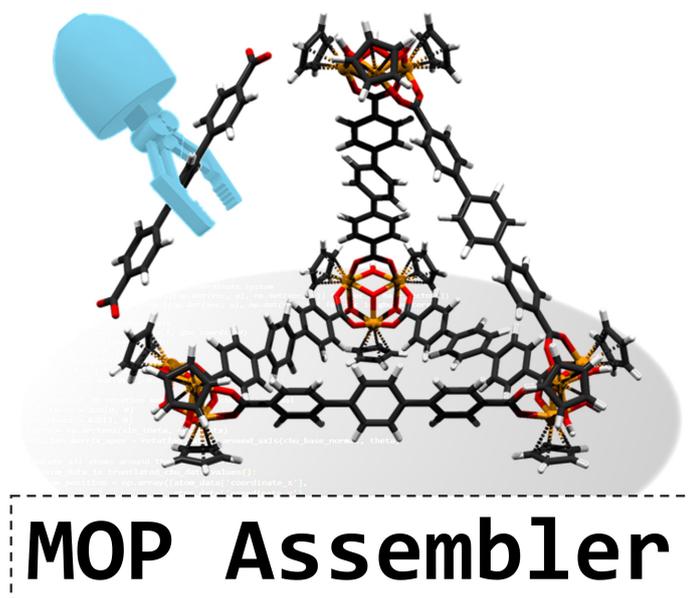
Computational Modelling Group
Department of Chemical Engineering and Biotechnology
University of Cambridge
Philippa Fawcett Drive
Cambridge, CB3 0AS
United Kingdom

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



Abstract

Assembly modelling has been achieved in knowledge AI systems for the automated inference of new and rational metal-organic polyhedra (MOPs) (*J. Am. Chem. Soc.* 2022, 144, 26, 11713–11728). In this work, we implemented an algorithm and data structure that extends the process of assembly modelling to the automated generation of structural information about MOPs, enabling computational approaches to analyse trends in cavity and pore sizing. The structural geometries obtained from this work are semantically integrated as part of *The World Avatar*, a dynamic knowledge ecosystem.



Highlights

- Developed an automated workflow for constructing computation-ready MOPs.
- Extended assembly modeling to include polyoxometalates in MOP design.
- Generated MOP geometries closely matching experimental data.
- Workflow enables identifying MOPs for iodine capture applications.
- Structures are accessible via TWA-Marie agent with interactive 3D visuals.

Contents

| | | |
|----------|--|-----------|
| 1 | Introduction | 3 |
| 2 | MOP Assembler | 4 |
| 3 | Generated Structures and Analysis | 5 |
| 4 | Conclusion and Outlook | 8 |
| | Nomenclature | 9 |
| A | Appendix | 10 |
| A.1 | Examples of Constructed Assemblies | 10 |
| A.2 | Assembly Models | 11 |
| A.3 | CBUs Structure | 12 |
| A.4 | Application of Force Field Methods | 14 |
| A.5 | Selection of Structures for Potential Iodine Capture | 15 |
| | References | 16 |

1 Introduction

Metal-organic polyhedra (MOPs) are discrete, cage-like molecular materials constructed from metal- and organic-based chemical building units (CBUs) [6]. As the name suggests, MOPs resemble the shape of polyhedra; however, their rational construction typically follows structural blueprints referred to as assembly models. Considering that different organic and metal-based CBUs can be involved in the construction of MOPs, there are many different chemical structures that can be rationally derived and post-synthetically modified [3], covering broad stability and property ranges in terms of internal cavity volume and pore diameter but also chemical reactivity deriving from the nature of the involved CBUs. The tunable properties of the metal-organic polyhedra are often harnessed in the development of applications relevant to sustainability technologies such as catalysis, separation technologies (e.g. water purification), and molecular sieving (e.g. carbon, iodine capture) [2, 5, 10].

Molecular engineers rationally and sequentially develop new MOPs by consideration of various factors. A crucial step in this MOP development process involves assessing whether a specific combination of CBUs can construct a MOP according to a predefined assembly model [6]. This assessment reflects on the relationship between physical and abstract entities, similar to the conceptual framework involved in retrosynthetic analysis where synthons, retrons, and transforms are part of the abstract conceptualisation vocabulary [7]. In MOP assembly modelling, the CBUs that form MOP map to generic building units (GBUs) that combine in specific numbers to form an assembly model with a defined symmetry point group [6]. The GBUs are defined in terms of planarity and modularity, which define how GBUs and, more specifically, CBUs connect to one another to form a MOP associated with the assembly model (see Figure 1.a). Forms of virtual assembly modelling are also exhibited by computational chemists as part of their routines in constructing computation-ready structural models. In this line, the assembly modelling appears in cognitive decision-making across different stages of the digitally accelerated MOP development, and thus, its utility extends beyond the rational design process (see Figure 1.b). Knowledge graph technologies, including broad world-centric AI models, enable this form of information cascading *via* inter-entity connectivity and semantic interoperability across chemical domains [9], which in return can facilitate complex decision-making and problem-solving in chemistry.

Over the past decade, various research groups have developed workflows and software packages for constructing cage-like supramolecular entities, with a focus primarily on organic cages. These workflows have typically utilised cheminformatics standards for string processing and universal force fields (UFF) to optimise the cage structures [14, 15, 18, 20]. By contrast, cages with a predominantly inorganic character present challenges for standard cheminformatics tools, often requiring an alternative approach involving the use of direct spatial fragments from CBUs. This method necessitates precise spatial positioning and arrangement of CBUs to maintain their relative sizes. Such direct mapping of CBUs onto framework contours has proven effective in modelling metal-organic frameworks (MOFs) and covalent organic frameworks (COFs) [1]. Nonetheless, this strategy encounters cumulative challenges, including spatial misalignment and the limitations of cost-effective force field optimisation. These issues highlight the need for innovative as-

sembly model approaches that integrate multiple components for complex and automated structural modelling [8].

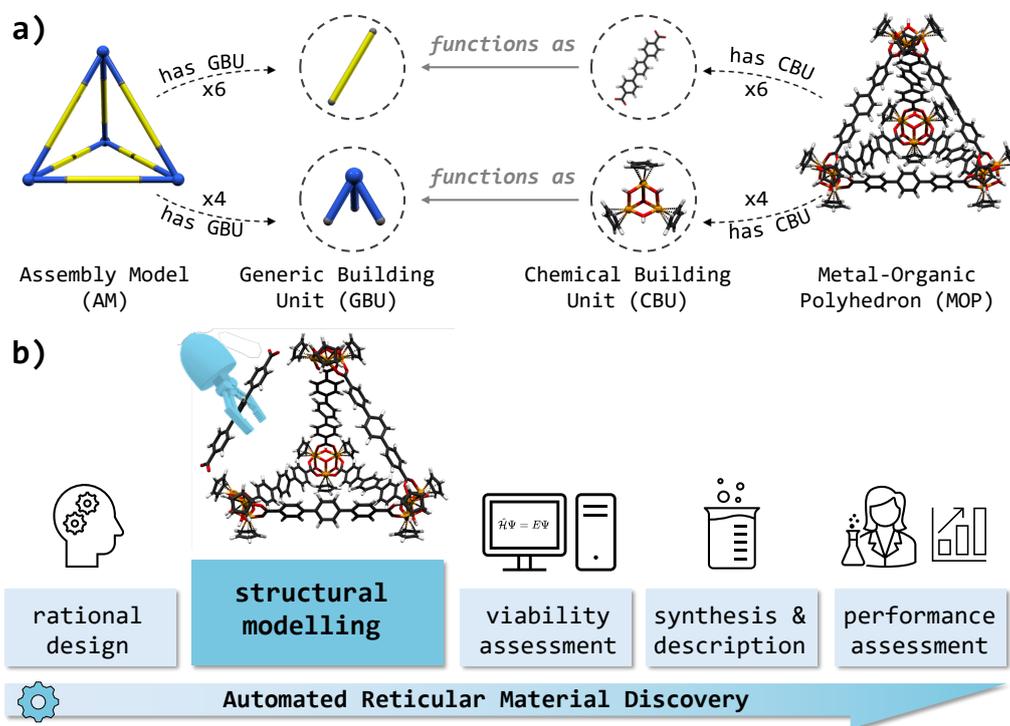


Figure 1: a) Key concepts involved in the cognitive mapping and deduction logic for the derivation of new MOP materials [6]; b) A general workflow for the automated discovery of new MOPs and potentially other reticular materials.

2 MOP Assembler

In our previous work, assembly modelling was effectively utilised in a decision-making process to determine which combination of CBUs, considering assembly constraints, can be rationally used to create MOPs [6]. The output decisions of the previous step are subjected to extended assembly modelling through a multistage workflow that produces computation-ready MOP structures (Figure 2.a). In the current work, critical information about the CBUs and the assembly model is provided through input files specifying atom and GBU connectivities to initiate geometry assembly. Binding sites on CBUs are marked with ‘X’ atoms, with corresponding markers on the GBUs. Next, the assembly model is resized by calculating size parameters based on averaged distances to binding sites for planar/linear CBUs or proportionate scaling from the binding base for bent/pyramidal CBUs. This rescaling aligns GBU centres and contact points to ensure a precise fit in the assembly model, after which each GBU is rotated and translated to align its binding sites with AM contact points, preserving intended geometry and preventing framework distortions. Integrated CBUs are positioned within the assembly using calculated translation and rotation vectors, maintaining centroid alignment and spatial relationships *via*

dummy atoms, resulting in an assembled geometry provided in XYZ format. This flexible workflow supports the modelling of complex MOPs and all-inorganic analogues such as polyoxometalates (POMs) (Figure 2.b), a significant capability given the interest in high-nuclearity POMs for various applications despite challenges in their automated construction and topological exploration [12, 13]. Operating independently of the assembly model design, the workflow accommodates diverse structures, as demonstrated in this study by constructing all eighteen previously reported assembly models (Figure 2.c), while remaining readily adaptable to new assembly models and other structural types [6].

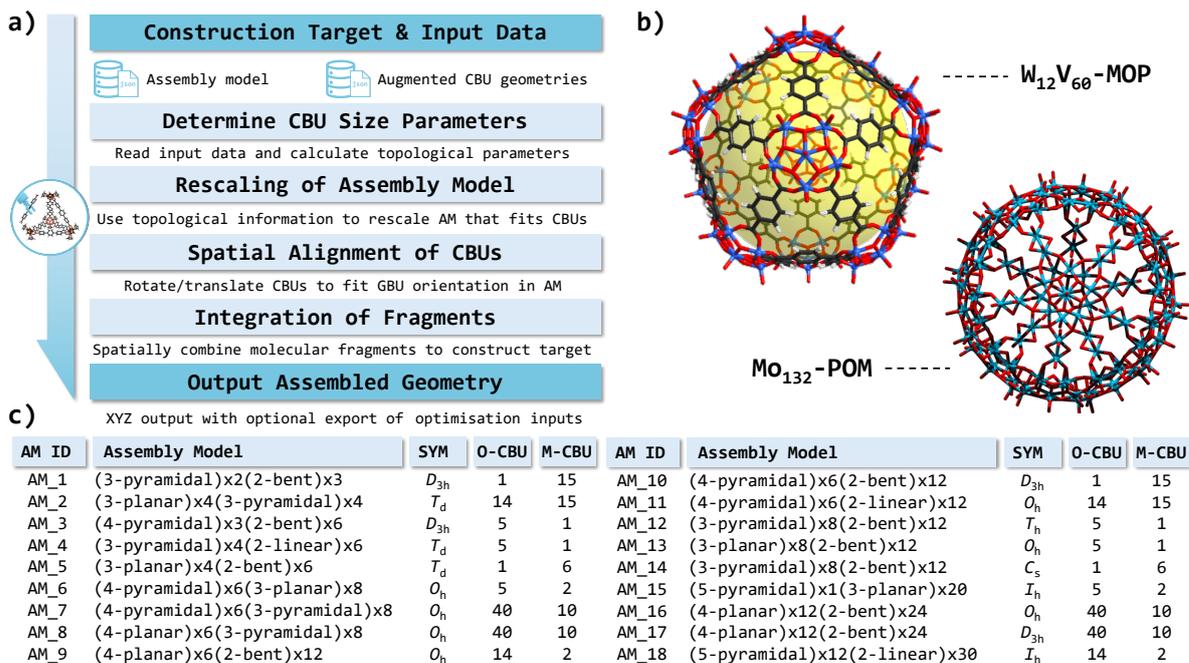


Figure 2: a) Algorithmic workflow for MOP construction, from input processing to geometry output; b) Assembly model examples for MOPs ($W_{12}V_{60}$) and keplerate type (Mo_{132}); c) Overview of assembly models with a particular symmetry (SYM) used in automated rational design of MOPs, [6] contrasted to the number of different organic (O) and metal-based (M) CBUs for structure construction in this work.

3 Generated Structures and Analysis

Post-assembly model validation using experimental data is imperative to ensure the fidelity of computational model predictions [11]. The modelling of metal-oxo clusters within MOPs poses significant challenges due to the complex nature of bonding, charge delocalisation, and structural distortions, which are not easily captured by classical force fields (detailed in the supporting information). Although enhancements to force-field-based geometry optimisations exist, quantum mechanical methods, particularly density functional theory, are indispensable for accurate modelling [11]. Our analysis focuses

on computation-ready geometries that are direct outputs of the assembly process. Given that the coordinates of most of the CBUs are derived from fragments of experimental MOP structures, discrepancies primarily arise at connectivity sites due to the placement of dummy atoms, impacting the precision of internal diameters. An evaluation of 124 modelled geometries against experimental counterparts shows that the largest inner sphere diameters predominantly range between 2 and 15 Å (see Figure 3.a). The estimated inner sphere radii for the assembled structures correlate linearly with those found for the experimental structures, with a relatively low overall deviation (R^2 value of 0.9894). This methodology effectively estimates internal cavity sizes. An analysis of the largest pore diameters (when more than one is present in a model structure) reveals a distribution primarily between 2 and 15 Å, with some exceptional cages exhibiting pores up to 30 Å, as shown in Figure 3.b. Typical pore sizes begin at around 10 Å and can extend to approximately 8 nm in MOPs with elongated side chains, as depicted in Figure 3.c.

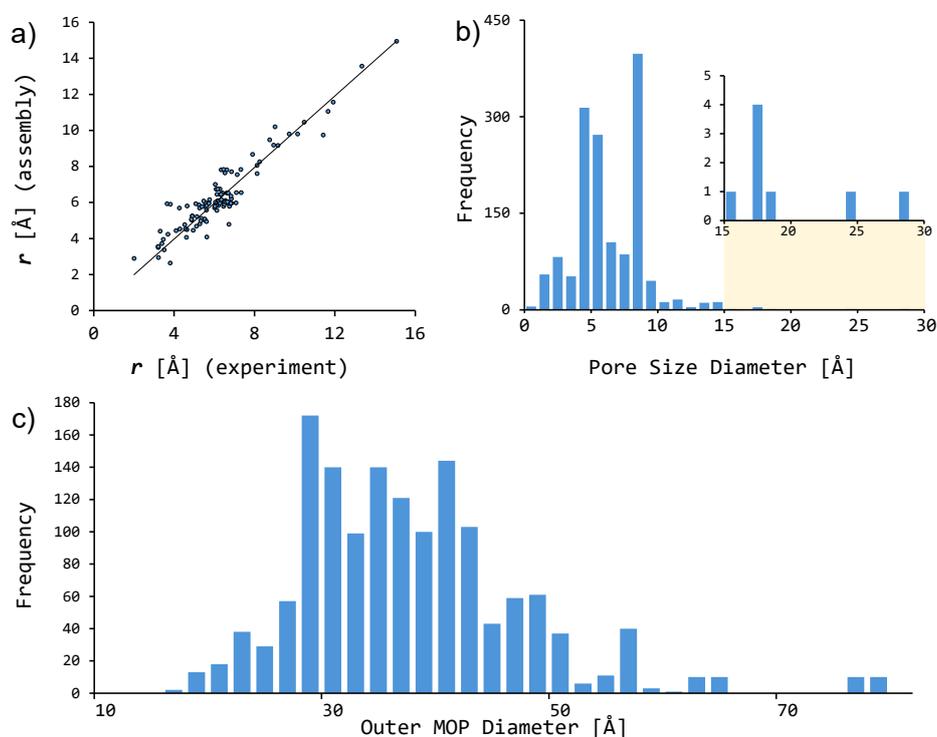


Figure 3: a) Correlation between the largest cavity sphere radii r_{exp} for 124 geometries from crystallographic geometries and their constructed assemblies $r_{assemb.}$, showing a slope of $r_{assemb.} = 0.9923 \cdot r_{exp}$ with $R^2 = 0.9894$. Covalent radii of elements were considered in the calculation of $r_{assemb.}$ and r_{exp} ; Histograms of known and new MOPs showing: b) pore diameters and c) outer MOP diameters.

As polyhedra-derived assembly models increase in complexity—from tetrahedra through octahedra to icosahedra—the ratio of pore diameter to inner sphere diameter decreases, with the steepest decline observed in tetrahedra and the least in icosahedra. By analysing six assembly models using vanadium-based CBUs associated with 3-, 4-, and 5-pyramidal GBUs and organic CBUs expanded incrementally with benzene rings, we observe a con-

sistent trend: as the number of faces converging at each vertex of the MOP increases, the pores shrink relative to the internal sphere (see Figure 4.a). This pattern holds across different assembly models, though pore sizes can be further minimised through the side functionalities (see Figure 4.b). These findings highlight the potential to tailor pore size and internal cavity features by modifying the CBUs, such as by introducing electron-donating groups and redox-active centres [21]. These modifications enhance the selection of MOPs for specific applications, for example, iodine capture, a field gaining significant attention in MOP and supramolecular chemistry [21]. Recent studies show experimental observations consistent with these trends [4, 19], leading to a curated list of MOPs as promising candidates for iodine capture.

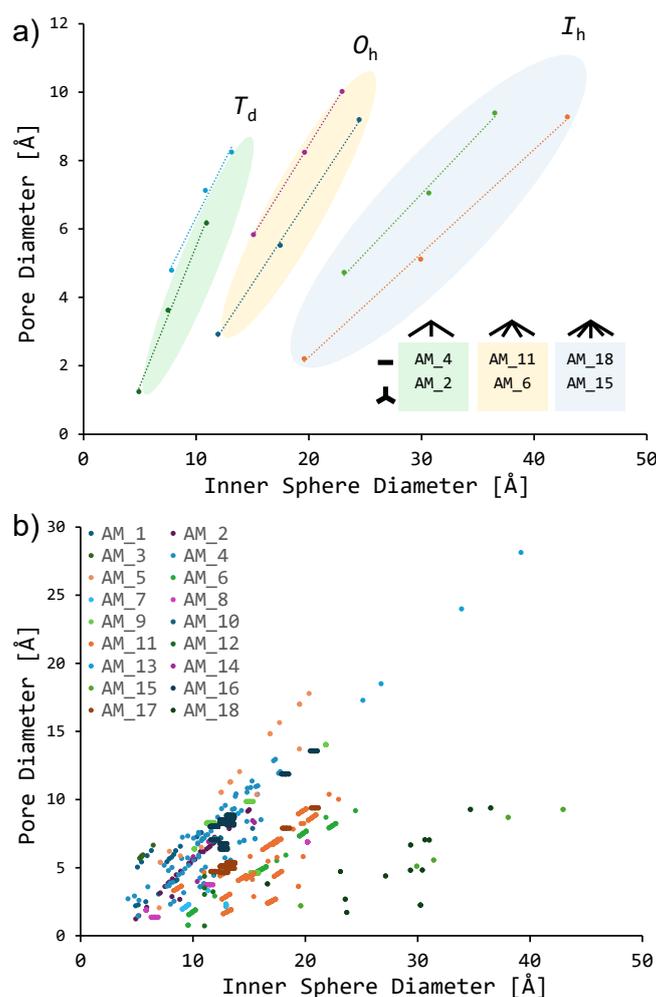


Figure 4: a) Scatter plot of T_d , O_h and I_h symmetric MOPs made of metal-based 3-, 4-, and 5-pyramidal CBUs (V_3 , V_5 , and V_6) and 2-linear/3-planar organic CBUs $[(C_6H_4)_n(CO_2)_2]^{2-}$ and $[(C_6H_3)(C_6H_4)_{3n}(CO_2)_3]^{3-}$, where $n = 1, 2$, and 3; b) Scatter plot of pore vs inner sphere diameter sizes of 1471 automatically constructed MOPs.

All structures constructed with the MOP assembler have been made available via Marie, the language agent of The World Avatar knowledge ecosystem (see Figure 5). This ar-

chitecture employs a fine-tuned Flan-T5 model (an advanced text-to-text transformer) to translate natural language queries into SPARQL (a query language for knowledge graphs), enabling precise, real-time access to structured data. This setup eliminates the need for traditional entity and relation-linking components [16, 17]. Marie provides interactive 3D visualisations and natural language summaries of newly generated MOPs, facilitating an intuitive exploration of their structural properties, symmetry, and assembly models. Such tools streamline the analysis of complex MOP geometries, enabling researchers to conduct rapid comparative assessments and deepen insights into structural nuances critical for advancing materials science applications.

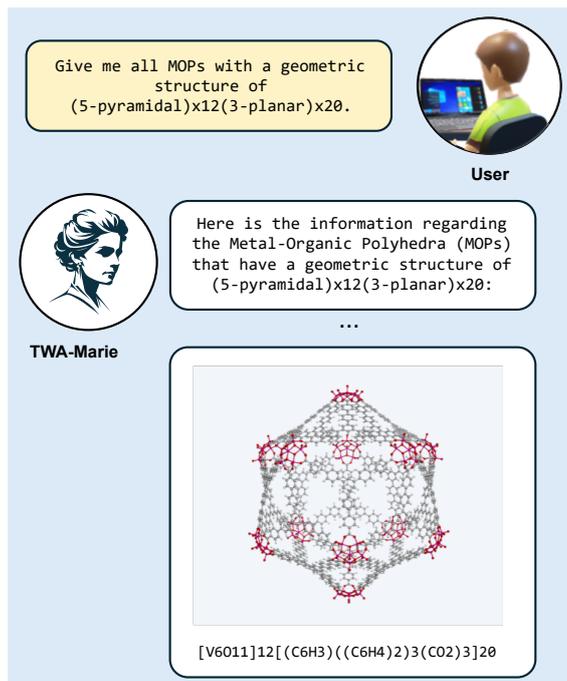


Figure 5: Example of a conversation with Marie obtaining information on new MOPs.

4 Conclusion and Outlook

Our automated assembly modelling approach enables the generation of computationally-ready geometries for MOPs and fully inorganic clusters, such as polyoxometalates, a capability not previously demonstrated by existing methods [1, 18, 20]. The geometries we have generated align closely with experimental data and are primed for further refinement using high-precision density functional theory, offering a significant advantage for optimising large clusters that pose challenges for traditional molecular mechanics. These structures facilitate efficient screening processes to pinpoint candidates with optimal pore-to-cavity ratios for applications in molecular storage. Accessible through the Marie natural language interface, all generated structures support our ongoing efforts to expand assembly modelling into semantic spaces, thus broadening the workflow’s utility across various applications, as outlined in our strategy for automated computational modelling of new materials [8].

Nomenclature

3D Three-Dimensional

CBU Chemical Building Unit

GBU Generic Building Unit

JSON JavaScript Object Notation

MOF Metal-Organic Framework

MOP Metal-Organic Polyhedron

POM Polyoxometalate

SPARQL SPARQL Protocol and RDF Query Language

TWA The World Avatar

UFF Universal Force Fields

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. AO thanks the ANID for Becas Chile Scholarship (74220070). For the purpose of open access, the author has applied a Creative Commons Attribution (CC BY) licence to any arising Author Accepted Manuscript version.

Data and code availability

All data covered in this research can be queried through TWA-Marie interface (<https://theworldavatar.io/demos/marie/>). All codes are accessible via the TWA git repository (https://github.com/cambridge-cares/TheWorldAvatar/tree/main/Agents/MOP_Assembler).

A Appendix

A.1 Examples of Constructed Assemblies

The following figure provides examples of constructed assemblies. Each example represents a unique MOP, with various building units and assembly models. Note that we use 'MOP_ID' codes as provided in the supporting data, where 'AM' followed by the first number refers to the assembly model.

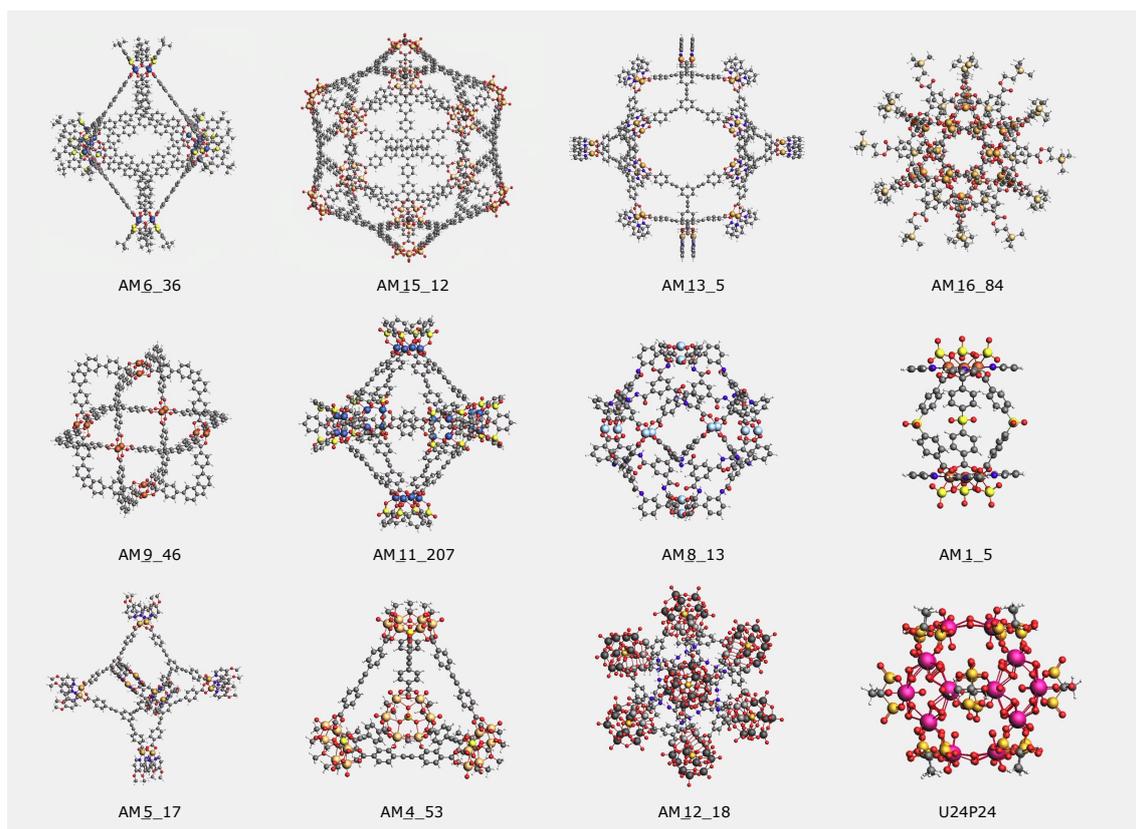


Figure A.1: Example of MOPs involving different types of building units and assembly models. An example of uranium-based polyoxometalate 'U24P24' constructed with the MOP assembler is provided to show the broader utility of the workflow in treating all-inorganic cage-like systems.

A.2 Assembly Models

Each assembly model is represented by a unique label format, such as '(5-pyramidal)x12(2-linear)x30_Ih', indicating the types and counts of units included, along with arrays of positions and dummy atoms. Each 'Position' entry includes a 'Key' identifier, a 'Label' specifying the unit type (e.g., "5-pyramidal"), coordinates ('X', 'Y', 'Z'), a 'Neighbors' list with distance values for adjacent units, and a 'ClosestDummies' array for alignment references. The 'Dummies' entries define auxiliary points, each with a 'Key', coordinates, and associated positions to aid in spatial alignment. An abbreviated JSON representation of an assembly model is provided below to illustrate the data structure and organisation of these components.

```
{
  "(5-pyramidal)x12(2-linear)x30_Ih": [
    {
      "Key": "Position_1",
      "Label": "5-pyramidal",
      "X": 2.6368,
      "Y": 2.7551,
      "Z": 1.2068,
      "Neighbors": [
        {
          "Key": "Position_13",
          "Label": "2-linear",
          "Distance": 2.1029
        },
        {
          "Key": "Position_15",
          "Label": "2-linear",
          "Distance": 2.1029
        },
        ...
      ],
      "ClosestDummies": ["Dummy_1", "Dummy_2", "Dummy_3", ...]
    },
    ...
  ],
  {
    "Key": "Dummy_1",
    "Label": "Dummy",
    "X": 2.8490,
    "Y": 1.7266,
    "Z": 1.2590,
    "Positions": ["Position_13", "Position_1"]
  },
  ...
  {
    "Key": "Center",
```

```

    "Label": "Center",
    "x": 0.0,
    "y": 0.0,
    "z": 0.0
  }
  ...
}

```

A.3 CBU Structure

Each chemical building unit (CBU) is identified by a unique key (UUID) and contains core details such as ‘atom’ type, spatial coordinates (‘coordinate_x’, ‘coordinate_y’, ‘coordinate_z’), and a ‘bond’ array for bonded atoms and bond orders. Additionally, a CBU may include fields that aid in deriving input files for various calculations. For example, the ‘mmttype’ field specifies the molecular mechanics type, using labels such as ‘Cr4+2’ to denote specific atom types with particular oxidation states, coordination environments, or other chemical characteristics. The ‘qmmm’ field indicates whether the entry is intended for quantum or molecular mechanics treatment.

```

{
  "99a00b53-c5a2-4fb0-9d4c-60f9eba4b284": {
    "atom": "Cr",
    "coordinate_x": 0.0,
    "coordinate_y": 0.0,
    "coordinate_z": 0.0,
    "bond": [
      {
        "to_atom": "6572a6ad-021a-4b0c-971e-8f441d890a94",
        "bond_order": 1.0
      },
      {
        "to_atom": "1b7cc2e8-78c4-443b-bb0e-5fca2fff1f55",
        "bond_order": 1.0
      },
      ...
    ],
    "mmttype": "Cr4+2",
    "qmmm": "MM"
  },
  "14db1344-f3a8-4646-8733-915000558618": {
    "atom": "Cr",
    "coordinate_x": 0.0,
    "coordinate_y": 0.0,
    "coordinate_z": -2.141683,
    "bond": [

```

```

    {
      "to_atom": "6572a6ad-021a-4b0c-971e-8f441d890a94",
      "bond_order": 1.0
    },
    {
      "to_atom": "1b7cc2e8-78c4-443b-bb0e-5fca2ffff1f55",
      "bond_order": 1.0
    },
    ...
  ],
  "mmtype": "Cr4+2",
  "qmmm": "MM"
},
...
"c94c5281-faec-43e3-9898-7122555da0d6": {
  "atom": "X",
  "coordinate_x": -0.169096,
  "coordinate_y": 2.13638,
  "coordinate_z": 5.30183,
  "bond": [
    {
      "to_atom": "eabc6e1e-83d0-46e7-b806-cd9f3b37155e",
      "bond_order": 1.0
    },
    {
      "to_atom": "0246dbbc-510f-472f-b4b4-e1443a538ad9",
      "bond_order": 1.0
    }
  ],
  "mmtype": "O_HH",
  "qmmm": "MM"
},
"CENTER": {
  "atom": "CENTER",
  "coordinate_x": -0.16909599999999997,
  "coordinate_y": -1.22602075,
  "coordinate_z": 3.7247382499999997,
  "bond": [],
  "mmtype": "C_R",
  "qmmm": "MM"
}
}

```

A.4 Application of Force Field Methods

Metal oxo clusters are essential building blocks in MOPs, yet their modelling with classical force fields (FFs) encounters significant limitations due to their complex bonding and electronic environments. These clusters, with variable oxidation states, charge delocalisation, and intricate electrostatics, present challenges for FFs that rely on fixed parameters, often leading to inaccuracies in representing the metal-oxo bonding. FFs also struggle with solvent and counter-ion effects, which are crucial in realistic MOP simulations. Despite advancements, like MMFF94x/POMFF-II, that improve interactions with organic components, FFs lack transferability across different metal oxo cluster configurations. In MOPs, the connectivity between metal-based and organic building units (CBUs) needs to be precise; however, structural distortions frequently occur. For instance, in paddlewheel complexes, while Cr and Pd retain their structure, Cu-based centres tend to distort towards tetrahedral coordination, despite copper(II)'s potential stability in a square planar form. Adding terminal hydroxo or aqua ligands can mitigate some distortions, but other issues, like the unexpected rotation of carboxylate groups relative to the benzene ring, further impact structural fidelity. Local symmetry-breaking distortions are also common, especially around oxo groups in polynuclear clusters, where symmetry reductions, such as from C_{4v} to C_{2v} , destabilise the overall structure. MM optimisation generally improves only a limited subset of MOP structures, reinforcing the need for quantum mechanical approaches, such as DFT, for accurate modelling and optimisation.

These computation-ready MOPs geometries, originally derived from crystallographic information, exhibit major FF optimisation discrepancies from their initial structures (CCDC codes): 277343, 974183, 1424875, 1424879, 974181, 1007919, 1424878, 755921, 1465218, 1833526, 224753, 755922, 1478734, 1835131, 288571, 617177, 755928, 1497171, 692012, 766959, 706820, 962336, 1830798, 1830799, 1533032, 706818, 1586600, 1830800, 121469, 964631, 1423404, 1575660, 950331, 998115, 1424706, 1839943, 950332, 1218651, 1469173, 1823162, 950330, 1469175, 1583661, 248306, 950333, 1520038, 1590349, 273616, 965262, 1521975, 1823163, 273621, 965266, 1525508, 1838490, 1528352, 1846389, 759738, 1412183, 1552945, 1849691, 869988, 1425529, 1552946, 1880376, 885932, 1425533, 1552951, 1940761, 885933, 1434669, 1552953, 885934, 1434670, 1576897, 902100, 1469174, and 1576898.

A.5 Selection of Structures for Potential Iodine Capture

To identify suitable candidates for iodine capture, a range of MOPs were shortlisted based on their structural properties. Figure A.2 illustrates examples of the CBUs that form these MOPs, showcasing the diversity of structures considered. Table A.1 provides specific data on selected MOPs, including pore sizes and volumes, highlighting key attributes that make these structures potential candidates for iodine capture applications.

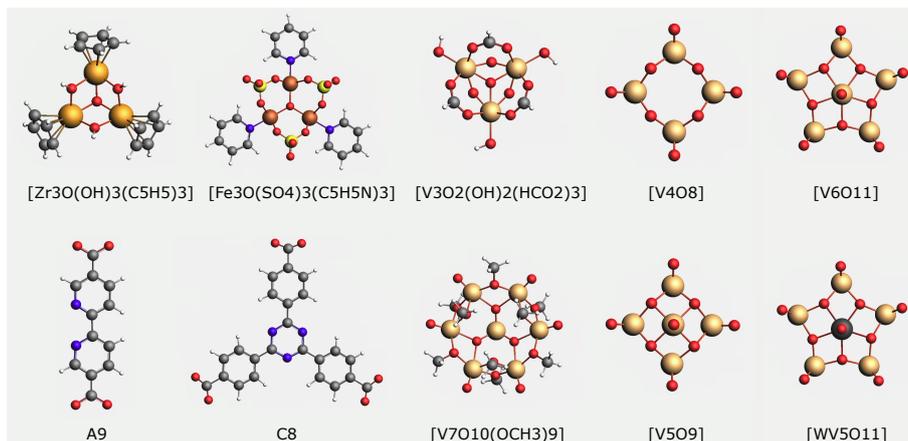


Figure A.2: CBUs considered for potential iodine capture.

| MOP_ID | CBU 1 | CBU 2 | d_{pore} [Å] | d_{sphere} [Å] | V_{sphere} [Å ³] |
|-----------|--|-------|-----------------------|-------------------------|---------------------------------------|
| AM_2_10 | [V ₇ O ₁₀ (OCH ₃) ₉] | C8 | 6.5 | 11.4 | 775.4 |
| AM_4_34 | [V ₃ O ₂ (OH) ₂ (HCO ₂) ₃] | A9 | 7.2 | 11.0 | 694.2 |
| AM_4_20 | [Fe ₃ O(SO ₄) ₃ (C ₅ H ₅ N) ₃] | A9 | 7.4 | 11.3 | 757.7 |
| AM_2_9 | [V ₆ O ₆ (OCH ₃) ₉ (SO ₄)] | C8 | 6.6 | 11.5 | 802.7 |
| AM_4_6 | [Zr ₃ O(OH) ₃ (C ₅ H ₅) ₃] | A9 | 8.1 | 12.2 | 949.7 |
| AM_4_62 | [V ₇ O ₁₀ (OCH ₃) ₉] | A9 | 10.9 | 15.1 | 1816.3 |
| AM_4_48 | [V ₆ O ₆ (OCH ₃) ₉ (SO ₄)] | A9 | 11.0 | 15.8 | 2051.2 |
| AM_6_44 | [V ₅ O ₉] | C8 | 5.5 | 17.5 | 2786.3 |
| AM_11_132 | [V ₄ O ₈] | A9 | 7.0 | 17.6 | 2842.3 |
| AM_11_34 | [V ₅ O ₉] | A9 | 8.3 | 19.9 | 4105.4 |
| AM_15_2 | [V ₆ O ₁₁] | C8 | 5.1 | 29.9 | 14017.9 |
| AM_15_8 | [WV ₅ O ₁₁] | C8 | 5.1 | 29.9 | 14018.3 |
| AM_18_7 | [V ₆ O ₁₁] | A9 | 7.0 | 31.0 | 15642.0 |
| AM_18_21 | [WV ₅ O ₁₁] | A9 | 7.0 | 31.0 | 15642.5 |

Table A.1: Selected properties of MOPs. d_{pore} and d_{sphere} represent the diameter of the pore and the largest inner sphere, respectively, both calculated excluding the covalent radii of the involved elements. V_{sphere} is the calculated volume of the largest encapsulated sphere within the structure.

References

- [1] M. A. Addicoat, D. E. Coupry, and T. Heine. Autografs: Automatic topological generator for framework structures. *J. Phys. Chem. A*, 118(40):9607–9614, 2014. doi:10.1021/jp507643v.
- [2] 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 CO₂ photocatalytic conversions: Current trends and future outlook. *J. CO₂ Util.*, 80:102664, Feb 2024. doi:10.1016/j.jcou.2023.102664.
- [3] 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. *J. Am. Chem. Soc.*, 141(9):4094–4102, 2019. doi:10.1021/jacs.8b13593.
- [4] S. Cheng, W. Chen, L. Zhao, X. Wang, C. Qin, and Z. Su. Synthesis, crystal structure and iodine capture of zr-based metal-organic polyhedron. *Inorg. Chim. Acta*, 516:120174, Feb 2021. doi:10.1016/j.ica.2020.120174.
- [5] A. J. Gosselin, C. A. Rowland, and E. D. Bloch. Permanently microporous metal–organic polyhedra. *Chem. Rev.*, 120(16):8987–9014, 2020. doi:10.1021/acs.chemrev.9b00803.
- [6] A. Kondinski, A. Menon, D. Nurkowski, F. Farazi, S. Mosbach, J. Akroyd, and M. Kraft. Automated rational design of metal-organic polyhedra. *J. Am. Chem. Soc.*, 2022. doi:10.1021/jacs.2c03402.
- [7] A. Kondinski, J. Bai, S. Mosbach, J. Akroyd, and M. Kraft. Knowledge engineering in chemistry: From expert systems to agents of creation. *Acc. Chem. Res.*, 56(2):128–139, 2023. doi:10.1021/acs.accounts.2c00617.
- [8] A. Kondinski, S. Mosbach, J. Akroyd, A. Breeson, Y. R. Tan, S. Rihm, J. Bai, and M. Kraft. Hacking decarbonization with a community-operated creatorspace. *Chem*, 10(4):1071–1083, 2024. doi:10.1016/j.chempr.2023.12.018.
- [9] A. Kondinski, P. Rutkevych, L. Pascazio, D. N. Tran, F. Farazi, S. Ganguly, and M. Kraft. Knowledge graph representation of zeolitic crystalline materials. *Digit. Discov.*, 2024. doi:10.1039/D4DD00166D.
- [10] S. Lee, H. Jeong, D. Nam, M. S. Lah, and W. Choe. The rise of metal–organic polyhedra. *Chem. Soc. Rev.*, 50:528–555, 2021. doi:10.1039/D0CS00443J.
- [11] X. López, J. J. Carbó, C. Bo, and J. M. Poblet. Structure, properties and reactivity of polyoxometalates: A theoretical perspective. *Chem. Soc. Rev.*, 41:7537–7571, 2012. doi:10.1039/C2CS35168D.
- [12] D. Melgar, N. A. G. Bandeira, and C. Bo. Electronic structure studies on the whole keplerate family: Predicting new members. *Chem. Eur. J.*, 23(22):5338–5344, 2017. doi:10.1002/chem.201605981.

- [13] P. Miro and C. Bo. Uranyl-Peroxide Nanocapsules: Electronic Structure and Cation Complexation in $[(\text{UO}_2)_{20}(\mu\text{-O}_2)_{30}]^{20-}$. *Inorg. Chem.*, 51(6):3840–3845, 2012. doi:10.1021/ic300029d.
- [14] T. K. Piskorz, V. Martí-Centelles, T. A. Young, P. J. Lusby, and F. Duarte. Computational modeling of supramolecular metallo-organic cages—challenges and opportunities. *ACS Catal.*, 12(10):5806–5826, 2022. doi:10.1021/acscatal.2c00837.
- [15] A. Tarzia and K. E. Jelfs. Unlocking the computational design of metal–organic cages. *Chem. Commun.*, 58:3717–3730, 2022. doi:10.1039/D2CC00532H.
- [16] D. N. Tran, L. Pascazio, J. Akroyd, S. Mosbach, and M. Kraft. Leveraging text-to-text pretrained language models for question answering in chemistry. *ACS Omega*, 9(12):13883–13896, 2024. doi:10.1021/acsomega.3c08842.
- [17] D. N. Tran, S. D. Rihm, A. Kondinski, L. Pascazio, F. Saluz, S. Mosbach, J. Akroyd, and M. Kraft. Natural language access point to digital metal-organic polyhedra chemistry in the world avatar. Technical Report 327, c4e-Preprint Series, Cambridge, 2024. URL <https://como.ceb.cam.ac.uk/preprints/327/>.
- [18] L. Turcani, A. Tarzia, F. Szczypiński, and K. E. Jelfs. stk: An extendable python framework for automated molecular and supramolecular structure assembly and property prediction. *J. Chem. Phys.*, 154:214102, 2021. doi:10.1063/5.0049708. URL <https://doi.org/10.1063/5.0049708>.
- [19] Y. Yang, Y. Fu, Y. Tian, L. Zhao, C. Qin, X. Wang, and Z. Su. Efficient iodine capture by metal–organic cubes based on hexanuclear vanadium clusters. *Inorg. Chem. Front.*, 10:6221–6228, 2023. doi:10.1039/D3QI01501G.
- [20] T. A. Young, R. Gheorghe, and F. Duarte. cgbind: Rapid construction and binding affinity prediction of host-guest systems. *J. Chem. Inf. Model.*, 60:3546–3557, 2020. doi:10.1021/acs.jcim.0c00519.
- [21] W. Zhou, R. Lavendomme, and D. Zhang. Recent progress in supramolecular iodine capture by macrocycles and cages. *Chem. Commun.*, 60:779–792, 2024. doi:10.1039/D3CC05337G.