Chemical Science

ROYAL SOCIETY OF CHEMISTRY

EDGE ARTICLE

Published on 04 September 2026
Licensed under CC-BY 4.0

Check for updates

Cite this: *Chem. Sci.*, 2026, **17**, 20555

All publication charges for this article have been paid for by the Royal Society of Chemistry

Hierarchical crystal structure prediction of zeolitic imidazolate frameworks using DFT and machine-learned interatomic potentials

Yizhi Xu, [ab] Jordan Dorrell, [cd] Katarina Lisac, [b] Ivana Brekalo, [b] James P. Darby, [e] Andrew J. Morris *[c] and Mihails Arhangelskis *[a]

Received 22nd May 2026
Accepted 3rd September 2026

DOI: 10.1039/d6sc04307k

rsc.li/chemical-science

Crystal structure prediction (CSP) is emerging as a powerful method for the computational design of metal–organic frameworks (MOFs). In this article we employ CSP to perform high-throughput exploration of the crystal energy landscape of zinc imidazolate (Zn**Im**$_2$). As the most polymorphic member of the zeolitic imidazolate framework (ZIF) family, Zn**Im**$_2$ has at least 24 reported structural and topological forms, and new polymorphs are still being regularly discovered. With the aid of custom-trained machine-learned interatomic potentials (MLIPs) we have performed a high-throughput sampling of over 3 million randomly generated crystal packing arrangements and identified 9609 energy minima characterized by 1484 network topologies, including 855 topologies that have not been reported before. Of the experimentally reported structures of Zn**Im**$_2$ falling within the search boundaries, all but one were ultimately matched with the predicted structures, demonstrating the power of the CSP method in sampling experimentally relevant ZIF structures. Finally, through a combination of topological analysis and density and porosity considerations, we have identified a set of structures representing promising targets for future experimental screening, as well as demonstrated how structures of mechanochemically synthesized MOFs could be identified *via* matching experimental powder diffraction patterns with simulated patterns from the predicted structures.

Introduction

Metal–organic frameworks (MOFs) are highly versatile materials with applications in gas storage[1] and separation,[2] catalysis,[3,4] water purification,[5] removal of harmful agents from air,[6] energy storage,[7] light harvesting,[8] fuels[9,10] and more.[11,12] Such functional diversity is directly related to the modular nature of MOFs, which are constructed from metal nodes interconnected with organic linker molecules, giving rise to a vast number of node and linker combinations, resulting in materials with diverse short-range interaction geometries, long-range crystal packing and associated functional properties.[13]

The structural and functional variability of MOFs, however, is not limited to node-and-linker variations: a further dimension of structural diversity comes from polymorphism, where the same building blocks give rise to multiple crystallographic arrangements. A prominent class of MOFs renowned for polymorphic and topological diversity are zeolitic imidazolate frameworks (ZIFs),[14] which are geometrically and topologically related to zeolites, thanks to the tetrahedral geometry of the metal nodes and angled coordination geometry of the imidazolate linkers. An archetypal example of ZIF polymorphic diversity is zinc imidazolate (Zn**Im**$_2$) which, to date, has been represented by at least 24 crystallographically distinct forms in 19 topologies, isolated *via* solution crystallization, template-assisted synthesis,[15] solvothermal methods,[16] high-pressure-and-temperature experiments[17] and mechanochemical screening.[18] The regular discovery of new polymorphs of Zn**Im**$_2$ suggests that many more such forms can be discovered in the future. Yet, without knowing the crystal structures of the not-yet-discovered polymorphs of Zn**Im**$_2$ it is difficult to systematically target materials with specific functional characteristics, including surface area and pore volume.

The discovery of new ZIF forms with the desired functional characteristics can be accelerated through the use of crystal structure prediction (CSP), a method which has been widely used for the discovery of new crystal forms of organic molecular materials,[19] including pharmaceutical solids, porous organic materials,[20] inorganic solid electrolytes[21] and high pressure mineral phases.[22] Yet, unlike for purely organic and inorganic materials, where CSP has become an established method for

[a]*Faculty of Chemistry, University of Warsaw, Warsaw 02-093, Poland. E-mail: m.arhangelskis@uw.edu.pl*

[b]*Division of Physical Chemistry, Ruđer Bošković Institute, Zagreb 10000, Croatia*

[c]*School of Metallurgy and Materials, University of Birmingham, Elms Road, Edgbaston, Birmingham B15 2TT, UK. E-mail: a.j.morris.1@bham.ac.uk*

[d]*School of Chemistry and Chemical Engineering, University of Southampton, Southampton, SO17 1BJ, UK*

[e]*Symmetric Group, 71-75 Shelton Street, Covent Garden, London, WC2H 9JQ, UK*

© 2026 The Author(s). Published by the Royal Society of Chemistry

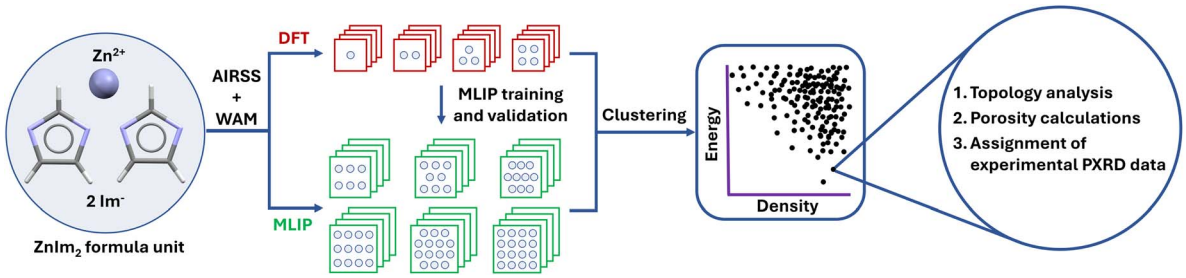


Fig. 1 Diagram depicting the key steps of crystal structure prediction for Zn**Im**$_2$. We begin with the crystal structure generation using the AIRSS + WAM method. The generated structures containing 1–4 formula units of Zn**Im**$_2$ per primitive crystallographic unit cell are then subjected to periodic DFT optimization. The MLIPs are trained against a subset of DFT data and then validated against the remainder of DFT data, not used during training. The resulting MLIPs are then used to optimize and energy-rank all the trial structures, containing up to 16 formula units per cell. Duplicate structures are removed during the clustering step, resulting in a crystal-energy landscape of unique structures in a synthetically relevant energy range of 45 kJ mol^{-1} from the global minimum. The predicted structures are then characterized in terms of network topology and porosity, as well as matched against experimental powder diffraction data.

materials design, the development of CSP methods capable of addressing the hybrid node-and-linker composition of MOFs mainly relied on topology-based[23–27] structure generation, limiting the generated structures only to derivatives of known topologies. To address this issue in 2020 (ref. 28) we developed a new CSP approach for structure generation of MOFs based on the *ab initio* random structure searching (AIRSS) method,[29] supplemented by the Wyckoff alignment of molecules (WAM) procedure,[28] which utilizes the point group symmetry of linkers when generating putative structures. The structures were then optimized using periodic density-functional theory (DFT) calculations, resulting in an energy ranking of the generated structures and thus a prediction of the most thermodynamically stable crystal forms. Emphasizing its utility, this approach soon after allowed for the first CSP-driven discovery of functional hypergolic MOFs.[30]

Our choice in using periodic DFT for the energy ranking was motivated by its excellent accuracy in reproducing experimentally measured MOF polymorph energies,[31–34] yet the high computational cost is a major limitation in terms of the system sizes amenable to CSP. This limitation is particularly relevant in the study of the highly polymorphic Zn**Im**$_2$ materials, where reported system sizes reach up to 40 formula units per primitive crystallographic unit cell.[35] Since our key focus is on the wide adaptation of CSP-based MOF design, as a complementary approach to experimental structure screening, accurate yet computationally more efficient alternatives to DFT-based energy ranking are indispensable. Such an alternative has been presented in the form of machine-learned interatomic potentials (MLIPs), which have recently gained traction in computational materials discovery.[36–43]

Here we present a hierarchical high-throughput CSP study of Zn**Im**$_2$ (Fig. 1), involving training of a custom-made MLIP (Fig. 2). The chemical equivalence of all ZIF structures and low computational cost of MLIP-based geometry optimizations allowed us to target structures with up to 16 formula units of Zn**Im**$_2$ per primitive crystallographic unit cell. This is a significant advancement compared to our previous DFT-based CSP studies that have been limited to 1–4 MOF formula units per cell.[28,30] Inclusion of structures comprising larger unit cells and higher atomic content increased the chances of locating experimentally relevant structures and expanded the topological diversity of the predicted structures.

The validity of our WAM-MLIP approach was tested by comparing the MLIP predictions with the results of select periodic DFT geometry optimizations and by comparing the geometry parameters of the MLIP-generated structures to statistical experimental values.

We verify the robustness of the presented CSP approach by reproducing multiple experimentally observed polymorphs of Zn**Im**$_2$ and use the exploration of the topology and porosity characteristics of the other predicted structures to propose likely targets for future experimental synthesis. Finally, we present the assignment of experimental powder X-ray diffraction (PXRD) data for the mechanochemically synthesized polymorphs of Zn**Im**$_2$ against the predicted structures. The presented protocol, based on the variable cell powder-based similarity index (VC-GPWDF) method,[44] highlights the utility of CSP in analyzing the outcomes of mechanochemical reactions, where the polycrystalline nature of their products makes the experimental structure determination particularly challenging.

Results and discussion

Training and validation of the ML potential

Previously, the AIRSS method has been used to predict structures of a wide variety of materials, including solid electrolytes,[45,46] materials under high pressure,[47–50] extra-terrestrial minerals,[51] hybrid perovskites,[52] organic molecular crystals,[53,54] non-metal organic frameworks based on porous organic salts[20] and MOFs.[28,30,55] This diversity signifies the versatility of AIRSS, which is based on placing structural building blocks at random positions within the trial unit cell with randomly defined unit cell parameters, followed by relaxation of the geometry of such trial structures. The structure generation step is then repeated until the search is converged. The key strength of AIRSS lies in the ability to apply structural constraints suitable for a particular system, *e.g.* defining geometries of the structural building

© 2026 The Author(s). Published by the Royal Society of Chemistry

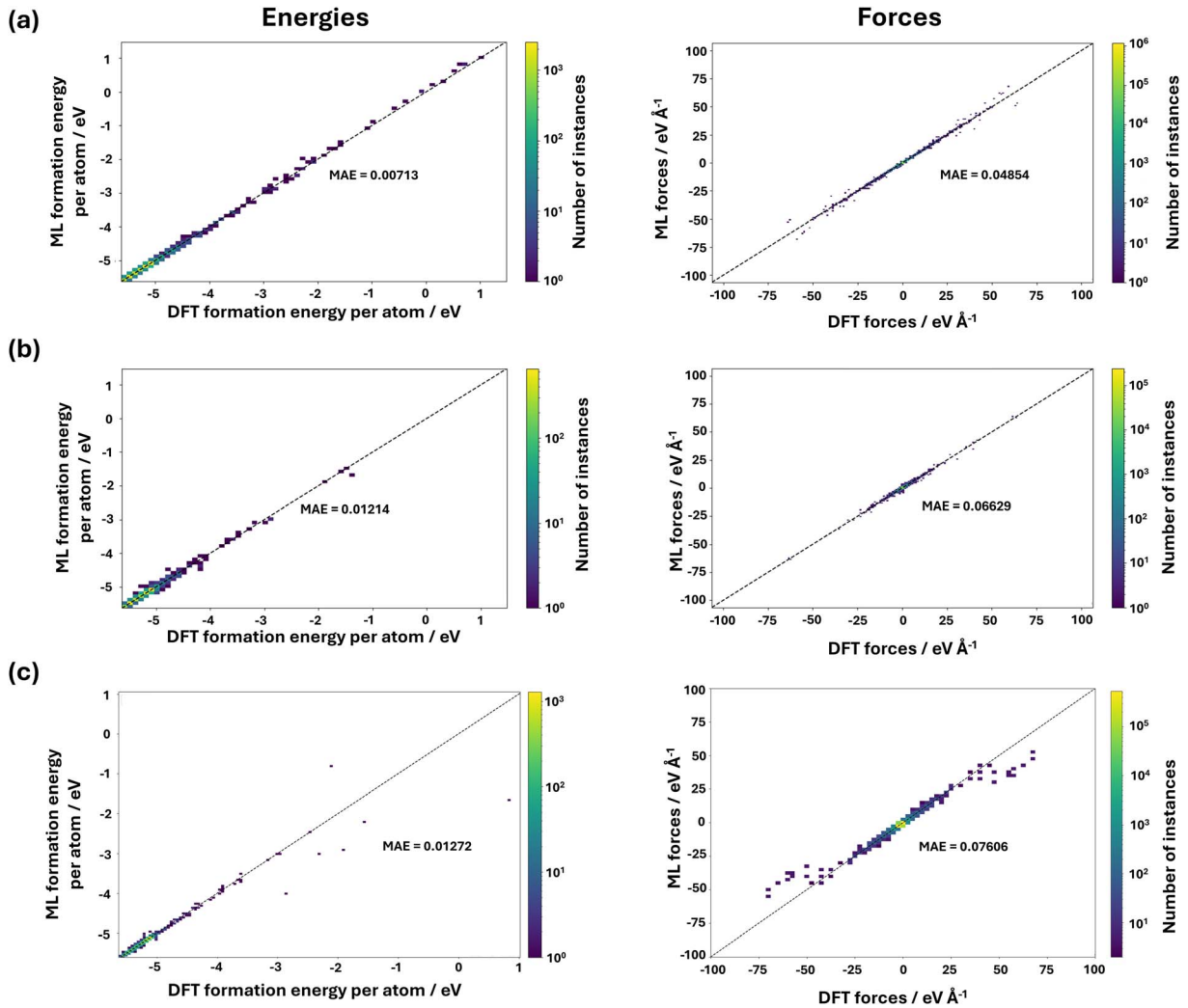


Fig. 2 Correlation between the ML and DFT energies and forces in (a) training set, (b) validation set and (c) test set. The colored vertical bar represents the abundance of energy and force values within the given dataset. The strong correlation between DFT and ML forces is a prerequisite for accurate and reliable structure optimization using the ML potential, while agreement between DFT and ML energies is a prerequisite for accurate energy ranking of the optimized structures. The magnified versions of these plots can be found in SI Fig. S7 and S8. MAE signifies the mean absolute error.

blocks in the form of isolated atoms, atomic clusters or extended molecules.

In the context of MOFs, the natural building blocks for an AIRSS search are metal nodes and organic molecular linkers. In addition, given that MOFs are known for their high crystallographic symmetry, we apply symmetry constraints *via* the WAM[28] method that allow symmetric building blocks to occupy special Wyckoff positions, enabling structures to be realized in unit cells with fewer formula units.

The key feature of our AIRSS + WAM methodology is that we are not making any assumptions about the metal coordination number, coordination geometry or framework topology. The only input information is the atomic composition and geometry of individual nodes and linkers, as well as the overall number of these fragments to be placed in the trial unit cell, subject to minimum separation (MINSEP) constraints. The connectivity between individual building blocks and, ultimately, framework topology, is established during the subsequent structure optimization steps.

This is in stark contrast to a family of structure-building methods,[23,24,26,27] where the nodes and linkers are initially placed at the positions defined by the desired network topology, and this topological connectivity is then preserved during the structure optimization step. Building structures from isolated nodes and linkers allows us to sample structures from a wider range of topologies, as well as discover new topologies, not yet found in databases such as Reticular Chemistry Structure Resource (RCSR)[56] or Topological Types Database (TTD).[57] We also see promise in addressing polymorphism within the same network topology: with the recent discovery of two new forms of zinc imidazolate with **crb** topology,[18] this material now has five crystallographically distinct **crb** polymorphs, emphasizing the importance of considering this type of polymorphism in computational screening of MOF structures.

© 2026 The Author(s). Published by the Royal Society of Chemistry

The major challenge in the development of MOF CSP is directly related to their covalent node-and-linker, hybrid organic–inorganic character. MOFs, being covalent 3D-polymeric structures, cannot be broken into fragments held only by non-covalent interactions. Because of this, MOF CSP cannot utilize force-field potentials in a way they are used for molecular crystals. Our initial strides in CSP for MOFs were, therefore, made using periodic DFT for energy ranking of putative structures, due to its excellent accuracy in reproducing experimentally measured MOF polymorph energies,[31–34] despite the high computational cost.

The CSP calculation process for Zn**Im**$_2$ initially followed a similar approach to our earlier CSP studies for zinc triazolate and tetrazolate,[28] as well as copper(II)-based hypergolic ZIFs.[30] The initial structure search spanned the space of 1–4 formula units per primitive crystallographic cell, with these structures geometry-optimized *via* plane-wave periodic DFT calculations in CASTEP19, using the LDA functional. However, it quickly became apparent that the search limited to 1–4 formula units per cell would not be sufficient to cover the relevant structural space, where many of the previously reported polymorphs of Zn**Im**$_2$ have been found. Indeed the experimentally reported polymorphs span a much larger structural space, including 8 formula units (**gis**, CSD EQOCOC01;[58] **crb**, CSD VEJYEP;[58] **moc**, CSD KUMXEW[59]) and 16 formula units (**coi**, CSD IMIDZB07;[60] **zni**, CSD IMIDZB02;[60] **crb** GITTEJ;[61] **cag**, CSD VEJYUF;[58] **dft**, CSD VEJYOZ;[58] **mer**, CSD VEJZIU;[58] **crb**, CSD VEJYIT[58]) as well as several examples of extra-large structures spanning 20–40 formula units per primitive cell (**nog**, CSD HIFWAV;[62] **zec**, CSD HICGEG;[62] **hlw**, CSD ZAVBUX;[63] **can**, CSD PAJRUQ;[64] **afi**, CSD IMIDZB13;[64] **10mr**, CSD GOQSIQ[35]).

Expanding the standard search method to a higher number of formula units would be met with two major obstacles: first, with the increased number of formula units, the number of atoms in the unit cells and the size of the unit cells themselves get larger, resulting in a higher cost of DFT optimization for each structure; second, a larger number of formula units leads to more structural degrees of freedom, making it necessary to optimize more structures in order to obtain good coverage of the potential energy surface (PES). Overall, the computational cost of exploring the structural landscape of Zn**Im**$_2$ with DFT-based energy ranking would quickly become prohibitively expensive, motivating us to seek a different strategy.

Broadly, there exist two approaches for the use of MLIPs in computational materials design: one can train a custom-made potential, based on DFT data gathered for a particular atomic or molecular composition, using one of the available codes for the construction of interatomic potentials based on artificial neural networks (ANNs), for example ænet[65] or SchNetPack.[66] Alternatively, one can utilize one of the generalized foundation models, such as MACE,[67] CHGnet[68] or NequIP,[69] that are trained on large databases of *ab initio* calculations, such as Materials Project,[70] OMOL[71] and SPICE.[72]

The generalized foundation models, owing to their large training datasets (it is now typical to train models on millions of structures, containing most of the elements of the periodic table), offer better generality over a range of systems they can target, but their accuracy needs to be tested for a particular problem at hand. Additionally, it is often possible to use fine-tuning procedures to improve the performance of these models when describing specific characteristics of a given material class.

On the other hand, custom-made potentials offer a more accurate description for materials closely related to the structures contained in the underlying training set, with no additional fine-tuning needed. As a downside, introduction of different types of chemical bonding can cause difficulties and addition of new elements not present in the original training set is generally impossible.

However, in our particular case, the similarity of all ZIF structures in terms of chemical connectivity (each structure is based on tetrahedral Zn nodes connected by imidazolate linkers *via* Zn–N bonds) and the need for a very accurate model (due to the very small energetic differences between different ZIF polymorphs) encouraged us to use custom-made potentials. This was made easier by the fact that extensive DFT data from the optimizations of 1–4 formula units were available as a basis to train a custom-made MLIP. We hypothesized that a MLIP trained on DFT data sampled from structures containing 1–4 Zn**Im**$_2$ formula units would be capable of predicting structures containing a much larger number of formula units. The rationale for this comes from the fact that all Zn**Im**$_2$ ZIF structures are chemically equivalent, regardless of their size: zinc metal nodes are expected to form four Zn–N coordination bonds, with consistent bonding lengths and angles, with each imidazolate linker utilizing both nitrogen atoms to connect with the metal nodes. Since the connectivity of the Zn**Im**$_2$ building blocks is the same regardless of their total number, a custom-made MLIP should have no scaling problems within this system.

To test this hypothesis, we will check the validity of our MLIP methodology by performing selected periodic DFT geometry optimizations for larger structures containing different numbers of formula units in their crystallographic cells and comparing our MLIP predictions with the results of DFT calculations. We will also check the statistics of the geometric parameters of the MLIP-optimized structures to see if they conform to the ranges expected based on a survey of experimental ZIF structures. To verify that building a custom-made potential instead of using a generalized foundation model was justified, we will evaluate the performance of our MLIP against that of the MACE model (SI Section S2).

We have selected the deep neural network atomistic simulation code SchNetPack[66] with the built-in polarizable interaction neural network (PaiNN) architecture[73] for constructing the MLIPs. The PaiNN neural network allowed us to use both energy and force data from the DFT calculations to train the MLIP models. In the end, we have constructed two separate MLIPs: one potential trained exclusively on DFT forces, which we used to optimize the trial structures, and another one trained on energies, used for energy ranking of the optimized structures (see SI Sections S1 and S2 for details). To validate the accuracies of energy and forces MLIPs, comparisons between ML predicted energies and forces with DFT values were obtained (Fig. 2) from structures in training, validation and test sets. This resulted in

© 2026 The Author(s). Published by the Royal Society of Chemistry

low mean absolute errors (MAEs) of 7.13 meV per atom, 12.14 meV per atom and 12.72 meV per atom from the energy MLIP for the training, validation and test sets, respectively. In the case of force MLIP, the MAEs were 48.54 meV $Å^{-1}$, 66.29 meV $Å^{-1}$ and 76.06 meV $Å^{-1}$ for the training, validation and test sets, respectively. Further statistical analysis of the agreement between DFT and MLIP energies and forces is presented in SI Fig. S1–S8. The MAEs obtained herein are similar to those reported in earlier MLIP-based global structure searches for inorganic materials.[41,74] Closer to the chemical nature of the materials explored in this work, the recent exploration of ZIF phase transformations reported an MLIP model with lower MAEs, namely 1.43 meV per atom for energies and 11.28 meV $Å^{-1}$ for the forces.[75] However, this study focused on the local energy landscape accessible from the ZIF-4 structure, rather than performing a global exploration of the entire ZIF structural space.

Based on the encouraging results from validating the accuracy of our MLIPs, all putative ZIF structures were geometry optimized. The optimized structures were then ranked using the energy MLIP, with the final set of structures ranging up to 45 kJ mol^{-1} above the global energy minimum retained for detailed analysis. Such an energy window was selected based on the prior results of DFT calculations and experimental calorimetric measurements of ZIF polymorph stability.[31–34,76–81]

General trends within the predicted energy landscape

The first, immediately apparent, feature of the energy landscape (Fig. 3) is the trend where lower density structures tend to be higher in energy. This is consistent with our previous calculations on MOFs,[31,32,82] as well as CSP studies of porous molecular crystals.[53,83,84] Based on void fraction analysis from the software PLATON,[85] 8247 structures out of 9609 from the CSP energy landscape are considered porous with non-zero void fractions. However, we must keep in mind that the porous nature of MOF structures poses certain challenges for energy ranking in CSP. While in close-packed materials, predicted structures with higher densities tend to have the lowest energies due to a larger number of short-range interatomic contacts, the situation is more complex with MOFs and other porous materials. The challenge is associated with the possibility of guest inclusion within the voids of the porous structures: while in a conventional CSP calculation, such voids are assumed to be empty, under the conditions of experimental synthesis, the structural voids can be readily occupied by solvent molecules, or other small molecule guests present in the reaction mixture. The inclusion of guests within the voids leads to additional stabilization of the structure *via* host–guest interactions, effectively making porous structures more stable than they appear under the energy calculations based on structures with empty voids. The effect of guest inclusion has been recognized as a challenge in previous CSP studies of porous molecular crystals,[86–89] as well as during the DFT-based energy ranking of MOF polymorphs obtained in the mechanochemical screening *via* liquid-assisted grinding (LAG).[18] Given the known propensity of Zn**Im**$_2$ and other ZIF systems to form porous structures, renowned for their sorption capacity,[90,91] it will be imperative to consider the effect

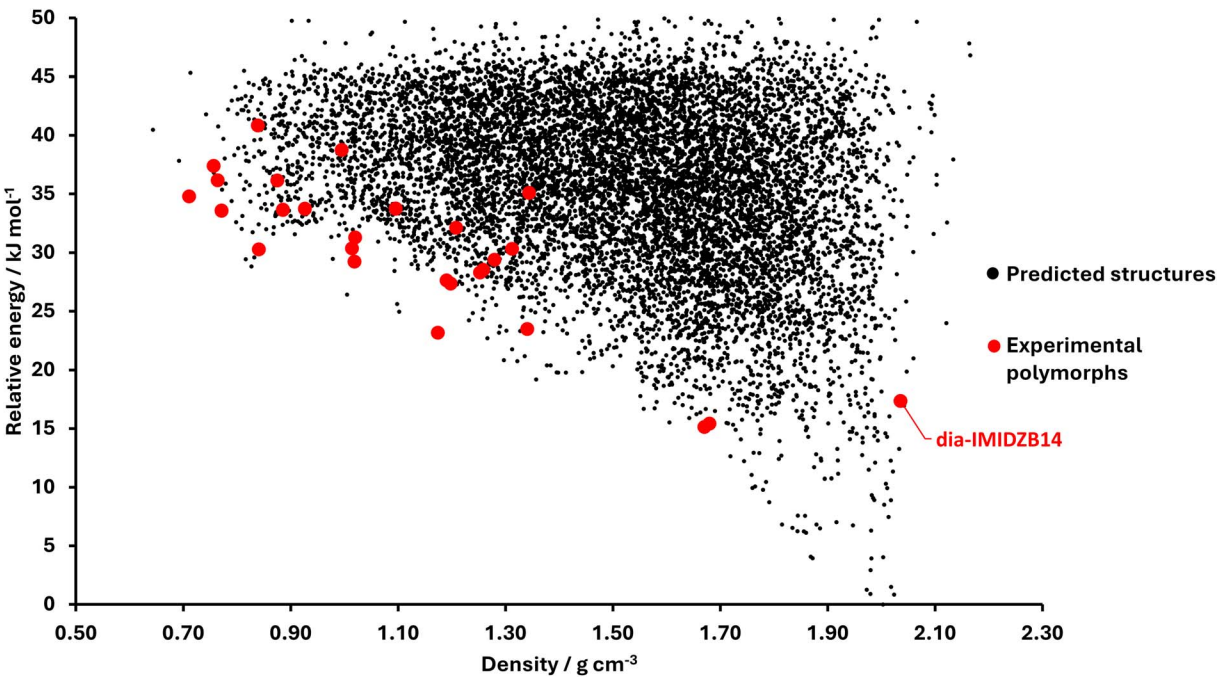


Fig. 3 Crystal energy landscape of Zn**Im**$_2$, where structures were geometry-optimized and energy-ranked with MLIPs, showing the calculated relative energies of predicted crystal structures against their density. The energies and densities of experimental structures of Zn**Im**$_2$ from CSD are shown in red. Experimental structures are clustered along the lower end of the energy-density envelope, with the exception of the highest density structure **dia**-IMIDZB14, which has been only experimentally synthesized under high pressure. This structure is specifically highlighted with a CSD REFCODE.

© 2026 The Author(s). Published by the Royal Society of Chemistry

of host–guest stabilization on the calculated energy landscape of Zn**Im**$_2$ in this study.

The energy-density trend discussed above correlates well with the characteristics of the experimentally reported forms of Zn**Im**$_2$. When placing the MLIP-optimized structures of experimentally reported polymorphs of Zn**Im**$_2$ on the energy landscape of CSP structures (Fig. 3), it is evident that these structures are concentrated at the lower diagonal part of the energy-density plot, signifying that there exists a relationship between relative stability, density and porosity of ZIF structures found experimentally. This suggests that the lower-diagonal region of the energy landscape is where the candidates for the future discovery of new polymorphs of Zn**Im**$_2$ are most likely to be found.

Topology distributions and structures matching experimentally reported forms of ZnIm$_2$

Having investigated the general stability trends throughout the calculated energy landscape, we turned our attention to the topological analysis of the predicted structures and the geometrical matching between the predicted structures and experimentally determined ZIF polymorphs (Fig. 4). The topological analysis revealed a remarkable diversity with 1484 distinct topological nets, of which 855 were found to be new topologies, not contained in the ToposPro[92] TTD database. We performed a statistical analysis of the geometrical parameters of the predicted structures, using the CCDC suite of programs.[93] Our analysis shows that these structures contain node and linker connectivity appropriate for ZIFs, with the Zn–N bond length and N–Zn–N bond angle distributions (SI Fig. S12 and 13) very similar to those of the experimental structures of Zn**Im**$_2$ found in the Cambridge Structural Database (CSD),[94] with the τ geometry index indicative of tetrahedral coordination geometry (SI Fig. S14). Agreement of the geometrical statistics of our predicted structures with experimental data provides confidence in the validity of our MLIP approach.

We next focused on the exploration of individual topologies and their distribution within the overall CSP energy landscape (SI Fig. S17–31), while highlighting the predicted structures matching the experimentally observed polymorphs and suggesting structures that appear as likely candidates for future synthesis of new polymorphs of Zn**Im**$_2$. In order to make such comparisons more robust, the experimental structures from the CSD were optimized with the same MLIP as that used for crystal structure prediction. The optimized structures were then

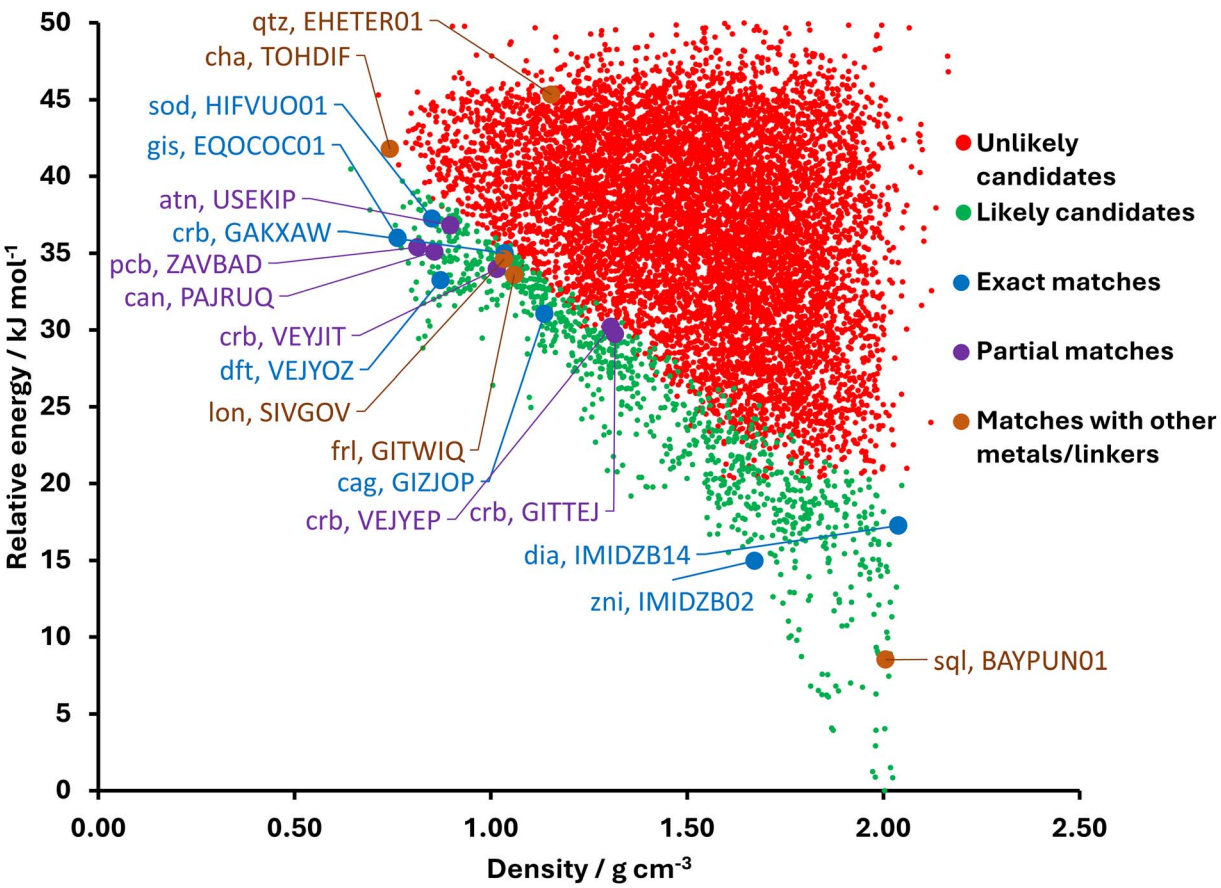


Fig. 4 Crystal energy landscape of Zn**Im**$_2$, where structures were geometry-optimized and energy-ranked with MLIPs. The structures are colored according to their synthetic feasibility, based on the relationship between the energy and void fraction calculated for the experimentally observed polymorphs. The structures colored in green are most likely to be synthesizable, while those colored in red are deemed synthetically less likely, given their higher relative energy and lower calculated porosity. In particular, it is evident that the hypothetical **qtz** structure is unlikely to exist based on the combined energy-porosity criterion for synthetic feasibility.

© 2026 The Author(s). Published by the Royal Society of Chemistry

compared with the structures from the CSP crystal energy landscape (SI Table S2).

The first major observation was the abundance of 2D structures throughout the CSP landscape, with 1294/9626 structures belonging to the **sql** topology, including the global minimum structure and all other structures within 4.1 kJ mol^{-1} above it (SI Fig. S30). To the best of our knowledge, no **sql** polymorph has been isolated for Zn**Im**$_2$ so far, but **sql** structures of unsubstituted imidazolate ZIFs have been reported for other transition metal centers, notably for Ni(**Im**)$_2$ (CSD ALIDUU)[95] and Hg(**Im**)$_2$ (CSD BAYPUN01).[96] In the latter case both experimental simulations and periodic DFT calculations have shown that the **sql**-Hg(**Im**)$_2$ form is more stable than its 3D polymorph with **dia** topology.[96] Interestingly, the Zn**Im**$_2$ structure isomorphous to the reported **sql**-Hg(**Im**)$_2$ form is found in our CSP energy landscape with an energy of 8.50 kJ mol^{-1} above the global minimum and below that of any of the experimentally obtained Zn**Im**$_2$ structures. This implies that it should in principle be possible to prepare a 2D polymorph of Zn**Im**$_2$.

Going up in energy, at 14.98 kJ mol^{-1}, the structure of **zni** topology Zn**Im**$_2$, matching the experimental structure (CSD IMIDZB02), was found. The **zni** form is currently regarded as one of the two densest and most thermodynamically stable reported polymorphs of Zn**Im**$_2$, along with the **coi** form.[76,77,79] The low energy of the **zni** form is evidenced both by experimental dissolution calorimetry measurements[76] and periodic DFT calculations.[18] Notably, the CSP energy landscape contained 15 different crystal structures with the **zni** topology, but it was the lowest energy structure among them that matched the experimentally reported form (SI Fig. S31).

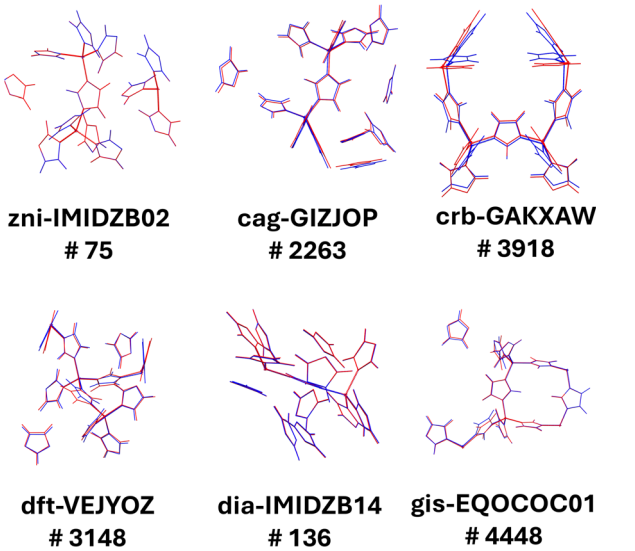


Fig. 5 Overlays of the predicted (red) and experimentally reported (blue) structures from the CSD, where the experimental structures were optimized with the same MLIP as that used for CSP. The topology and CSD REFCODE are written underneath each overlay picture. The CSP-generated structures are shown in blue, and the experimental MLIP-optimized structures are shown in red. The numbers underneath the structures refer to the entries in the supporting Chemiscope file.

Further inspection of the energy landscape revealed several more matches (Fig. 5) to the experimentally observed polymorphs of Zn**Im**$_2$, including the high pressure doubly interpenetrated **dia** polymorph (17.26 kJ mol^{-1} above the global minimum, matching structure CSD IMIDZB14),[17,97] **cag** (31.04 kJ mol^{-1}, matching CSD GIZJOP),[61] **dft** (33.25 kJ mol^{-1}, matching CSD VEJYOZ),[58] **gis** (36.00 kJ mol^{-1}, matching CSD EQOCOC01)[58] and **sod** (37.24 kJ mol^{-1}, matching CSD HIFVUO01).[98] In the latter case, it should be noted that the pure **sod** polymorph of Zn**Im**$_2$ composition has not actually been obtained so far; however, an **sod** material of the composition Zn(**Im**)$_{1.7}$(**mIm**)$_{0.3}$ (where **mIm** = 2-methylimidazolate) has been synthesized through a solvent-assisted linker exchange (SALE) procedure starting from Zn(**mIm**)$_2$ (ZIF-8), achieving 85% replacement of the **mIm**$^-$ linker with **Im**$^-$.[98] In the light of that result, the presence of the **sod** structure in the crystal energy landscape of Zn**Im**$_2$ is fully justified.

Further matches were found among the structures represented by the **crb** topology. This topology is rather unique in the sense that five crystallographically distinct polymorphs of Zn**Im**$_2$ have been reported so far,[18,58,61] making it the highest number of ZIF polymorphs sharing the same topology, to the best of our knowledge. The CSP landscape included a match to one of the forms of **crb**-Zn**Im**$_2$ just recently reported as **crbT** in our earlier publication (CSD GAKXAW).[18] This structure, experimentally obtained by liquid-assisted grinding of zinc oxide with imidazole in the presence of the toluene liquid additive, is found at 35.00 kJ mol^{-1} relative energy. For other existing polymorphs with **crb** topology, exact matches could not be found among the predicted structures; however partial matches were located for three out of four remaining **crb** structures: 29.73 kJ mol^{-1}, matching CSD GITTEJ;[61] 30.20 kJ mol^{-1}, matching CSD VEJYEP[58] and 33.97 kJ mol^{-1}, matching CSD VEJYIT.[58] In addition, a partial match was identified for the **pcb**/**aco** topology (CSD ZAVBAD)[63] at 35.39 kJ mol^{-1} and for the **can** topology (CSD PAJRUQ),[64] with a relative energy of 35.10 kJ mol^{-1}. In all of these cases, some of the imidazolate linkers were oriented differently in the CSP-generated structures, compared to their experimental counterparts, as seen from the overlays in Fig. 6, with the details given in SI Table S3. Imidazolate linker rotation around the Zn–Zn axis can bring the structure to a new energy minimum without breaking the covalent bonds and changing the network topology; therefore such partial matches, whilst less rewarding than complete matches discussed above, are nonetheless instructive from the point of view of structural and topological diversity of the zinc imidazolate crystal energy landscape.

A particularly interesting case is that of the **neb** topology. Experimentally, the neb topology was found in two distinct forms, **neb1** (CSD KUDJOK[99] – with morpholine – and GAKXUQ[18] – with cyclohexane, cHANE) and **neb2** (CSD KEVLEE[79] – with pyridine), primarily dictated by the type of included guest. Our CSP search found three different structures with **neb** topology, one of which, while not an exact match, appears to be structurally related to the **neb1** experimental polymorph. Namely, the structure (Chemiscope file from the SI entry 6574) is in the same *Fdd*2 space group as cHANE@**neb1**-Zn**Im**$_2$

© 2026 The Author(s). Published by the Royal Society of Chemistry

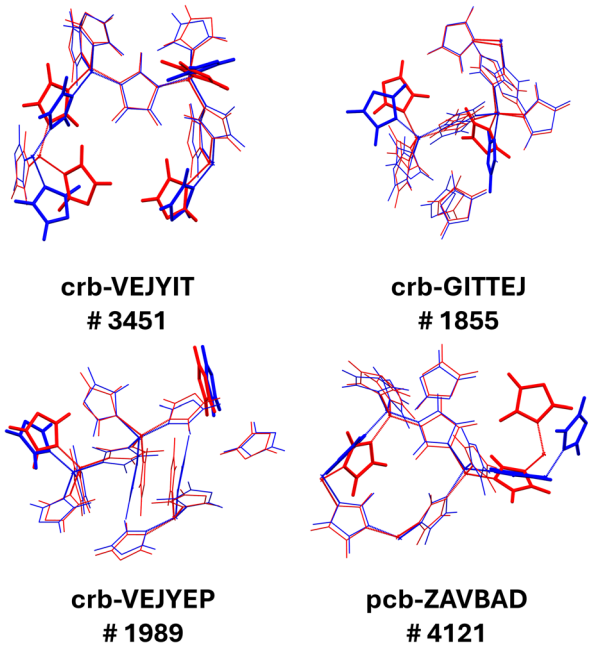


Fig. 6 Partial overlay between the experimental (blue) and predicted (red) structures of **crb** and **pcb** topologies. The imidazolate linkers drawn with thicker bonds are those whose orientations do not agree between the predicted and experimental structures. The numbers underneath the structures refer to the entries in the supporting Chemiscope file.

(GAKXUQ) and has the following unit cell parameters: $a = 16.87$ Å, $b = 26.72$ Å, and $c = 28.48$ Å, while the cHANE@**neb1**-Zn**Im**$_2$ (GAKXUQ) unit cell parameters are $a = 17.74$ Å, $b = 27.46$ Å, and $c = 9.11$ Å. It therefore appears that the predicted structure has similar a and b unit cell axes, but triple the c axis of the cHANE **neb1** polymorph. A graphical inspection of the predicted structure shows that its unit cell can be divided into three roughly repeating layers along the c axis, where each layer is a slightly distorted **neb1** unit cell, with rotations of imidazolate ligands causing differences between the layers. A comparison of the node and linker representations of the predicted **neb** structure with those of the empty and cyclohexane occupied **neb1**-GAKXUQ structure (Fig. 7) shows that the **neb1** cage is preserved in both structures, but is conformationally distorted in the CSP generated structure. We hypothesize that the source of the distortion is the lack of guest modelled in the CSP generated structure. It is very likely that the ordering of **neb** cages in the **neb1**-GAKXUQ structure arises from the incorporated guest. If the guest is not there, like in our CSP calculations, the linkers and nodes have much more freedom to move and distort, resulting in a structural mismatch, despite the fairly accurate crystal structure prediction.

Having identified the matching experimental polymorphs of Zn**Im**$_2$ among the predicted structures, we need to investigate the structures that have not been located and discuss the reasons for their absence within the predicted crystal energy landscape. The main reason for missing some of the existing polymorphs was limiting the search space to 16 formula units per primitive cell, as several experimentally determined

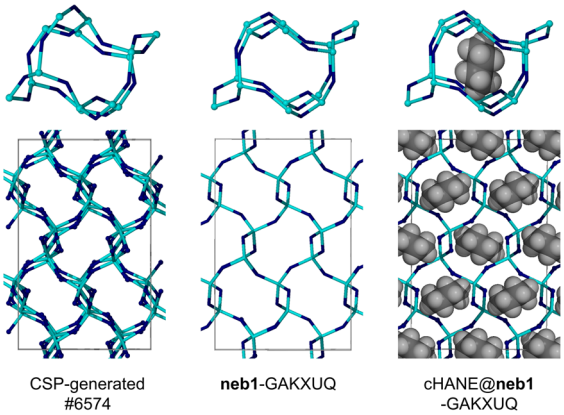


Fig. 7 Comparison of the predicted and experimental **neb1** polymorphs of Zn**Im**$_2$, highlighting the structural similarity when viewing down the c-axis.

structures contain more formula units. Specifically, the polymorphs with **zec** (CSD HICGEG)[62] and **nog** (CSD HIFWAV)[62] topologies contain 20 formula units per primitive cell, while structures with **gme** (CSD DOTCIC),[100] **hlw** (CSD ZAVBUX)[63] and **afi** (CSD IMIDZB13)[64] topologies contain 24 formula units. Finally, the **10mr** framework (CSD GOQSIQ) represents the most complex structure as a 10-nodal net with 40 formula units per primitive cell.

While the limit on the size of the structural search space explains the majority of the missing structures, there were two experimental polymorphs of Zn**Im**$_2$ containing 16 formula units per primitive cell, which have not been located in our CSP search. These were **mer** (CSD DOTBOH)[100] and **coi** (CSD IMIDZB07).[60] The **mer** structure, while representing a uninodal net, with just one Zn atom in the crystallographic asymmetric unit, has all its imidazolate linkers disordered with respect to rotation around the Zn–Zn axis. Since the predicted structures are necessarily ordered, we may suggest the inability to match the disorder of the experimental structure as the reason for our inability to reproduce the structure of the **mer**-Zn**Im**$_2$ framework.

The experimental form with **coi** topology (CSD IMIDZB07), also missing from our CSP landscape, represents a 4-nodal network, meaning that the corresponding crystal structure must contain at least 4 symmetry-independent Zn nodes, resulting in a large asymmetric unit, which is harder to generate during the AIRSS + WAM structure generation. Given the importance of the **coi** form as the lowest energy structure among the experimentally synthesized polymorphs of Zn**Im**$_2$ so far,[79] we decided to perform an additional structural search in order to better understand the challenges associated with the discovery of low symmetry MOF structures by CSP.

The dedicated search for the **coi** polymorph included generation of an additional 100 000 structures containing 16 formula units of Zn**Im**$_2$ in the space group $I4_1$, with the settings consistent with the experimental structure **coi**-IMIDZB07. For comparison, our original CSP search contained 12102 structures in these crystallographic settings; therefore the additional search corresponded to an 8-fold increase in the number of trial

© 2026 The Author(s). Published by the Royal Society of Chemistry

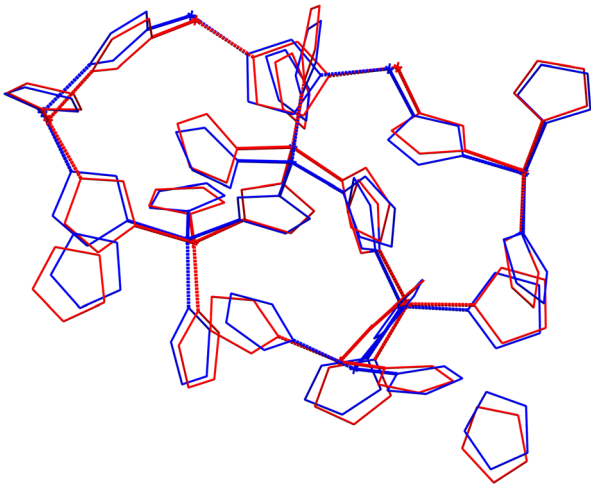


Fig. 8 Overlay of the experimental **coi**-Zn**Im**$_2$ (shown in blue, CSD IMIDZB07) with the structure generated during the additional search in $I4_1$ symmetry (shown in red).

structures. Gratifyingly, this additional search resulted in the location of a **coi** structure as the overall energy minimum among the newly sampled structural space (Fig. 8). This result implies that missing the **coi** structures in the initial CSP search was not caused by the limitations of AIRSS and WAM methods, but rather by the restricted number of generated structures. Increasing the number of trial structures can certainly increase our chances of locating low symmetry structures, yet the benefits of searching more structures have to be balanced with the higher computational cost of the calculation.

To summarize, the presented CSP search located all but two experimental crystal forms of Zn**Im**$_2$ within the imposed 16 formula unit limit, with the **coi** polymorph subsequently recovered in a more targeted search. Given that our previous CSP studies of MOFs, utilizing periodic DFT energy ranking, were limited to 4 formula units per primitive cell, the introduction of MLIP provided for a great expansion of the search space. Indeed, if we had performed this search at the DFT level and kept the search complexity limit to 4 formula units, we would not have found any of the experimental polymorphs of Zn**Im**$_2$, highlighting the importance of MLIPs for geometry optimizations and energy rankings for high-throughput CSP of MOF materials.

Predicted structures that are likely to be found in the future

The primary purpose of performing CSP calculations is to find likely candidates for future synthesis. With over 9000 structures found in the CSP landscape, we need a way to narrow down the search space for future synthetic efforts. The first thing to notice is the presence of multiple **sql** structures near the bottom of the energy landscape, including the global minimum. While no **sql** polymorphs have been isolated for Zn**Im**$_2$ so far, the existence of 2D **sql** structures for Ni**Im**$_2$ and Hg**Im**$_2$ suggests that **sql**-Zn**Im**$_2$ could be isolated, perhaps through seeding experiments.

The next general observation, arising from the location of the experimentally matching structures, is that synthetically viable structures appear at the bottom end of the energy-density envelope, as has been previously observed in CSP energy landscapes of porous molecular materials.[86,101–103] This means that high energy structures can be experimentally feasible, as long as they have low density and, correspondingly, high void volume and surface area, that lead to energy stabilization through host–guest interactions with structural templates or solvent guest molecules. Higher energy non-porous structures, however, cannot benefit from such host–guest stabilization and are therefore less likely to be produced during experimental synthesis.

One promising candidate for future synthesis may be the predicted structure with **lon** topology, with an energy of 34.62 kJ mol^{-1} above the global minimum. This structure is isomorphous to CSD SIVGOV, an experimental framework containing the 2-methyltetrazolate linker. The hypothetical structure of **lon**-Zn**Im**$_2$ has a similar energy to multiple experimentally observed polymorphs (*e.g.* **dft**, **can** and **gis**), a low calculated density of 1.03 g cm^{-3} and a high calculated void fraction of 49%, making it highly accessible to guest inclusion.

Another structure deemed promising based on similar arguments is the one with **cha** topology. With an even lower density of 0.86 g cm^{-3} and 58% calculated void fraction, this structure is isomorphous to CSD TOHDIF, a ZIF based on mixed 2-methylimidazolate and 5-methylimidazolate linkers.

The true value of CSP, however, is not in identifying individual structures isomorphous with CSD entries with different metal nodes or organic linkers, but rather in providing a range of targets that are structurally distinct from anything that has been experimentally obtained before, yet feasible from a synthetic standpoint. In order to narrow down the range of predicted structures and rank the remaining structures in the order of synthetic feasibility, we devised an empirical equation combining the relative lattice energy and calculated void volume:

$$E' = E_{\mathrm{rel}} - k_{\mathrm{V}} \times f_{\mathrm{void}}$$

where E_{rel} is the calculated energy of the structure relative to the global minimum (in kJ mol^{-1}) and f_{void} is the calculated void fraction (range from 0 to 1). The k_{V} parameter was fitted by evaluating the E_{rel} and f_{void} values of the existing experimental polymorphs of Zn**Im**$_2$, resulting in the best fit value of $k_{\mathrm{V}} =$ 34.05 kJ mol^{-1}, with the mean value for $E' = 16.35 \pm 3.66$ kJ mol^{-1} (see SI Section S7 for details). The significance of the descriptor E' is that structures with high lattice energy E_{rel} can be stabilized by solvent inclusion if they contain solvent-accessible void volume; the higher the void fraction, the greater the stabilization offered by guest inclusion. For non-porous structures, E' is equal to E_{rel}, while for more porous structures, the difference between E' and E_{rel} becomes progressively higher. Empirically we considered structures within one standard deviation from the mean E' value for the experimentally observed polymorphs of Zn**Im**$_2$ to be viable candidates for future synthesis. This resulted in 8-fold reduction of the number of structures under consideration from 9626 to 982,

© 2026 The Author(s). Published by the Royal Society of Chemistry

narrowing down the set of structures worthy of consideration for experimental screening.

In the future, such analysis may be expanded by explicitly modelling the host–guest interactions *via* molecular dynamics (MD) and Grand Canonical Monte Carlo (GC-MC) simulations;[104,105] however the force field potentials used in these simulations will first have to be tested for compatibility with the MLIP model used to describe the guest-free MOF structures. Alternatively, a generalized model capable of simultaneously simulating the structures of guest-free and guest-filled MOFs, as well as the bulk guests, needs to be developed. These calculations are beyond the scope of the present study, while the empirical model allows us to predict the guest-induced stabilization of Zn**Im**$_2$ structures and narrow down the range of structures deemed feasible from an experimental synthesis standpoint.

The significance of the synthesizability criterion can best be highlighted by looking at the predicted structure with **qtz** topology that was found to be isomorphous to the experimentally reported **qtz** polymorph of Zn(**EtIm**)$_2$ (CSD EHETER, **EtIm** = 2-ethylimidazolate).[106] While **qtz**-Zn(**EtIm**)$_2$ is known to be a stable dense structure,[31] our predicted **qtz**-Zn**Im**$_2$ analogue is found very deep in the region of non-synthesizable structures on the energy landscape (Fig. 4). It has a high relative energy, yet has only 6% solvent-accessible void volume, suggesting that guest-stabilization is likely to be unsuccessful and that **qtz**-Zn**Im**$_2$ is unlikely to be synthesized in the future.

The likely synthesizable structures were further analyzed for porosity, with 517 structures having non-zero calculated surface area and 291 structures having non-zero network-accessible surface area. Among these, 21 structures exceeded the network accessible surface area of 2000 $m^2\ g^{-1}$, with the maximum surface area found in a predicted structure with **dei** topology, at 2538.62 $m^2\ g^{-1}$. This structure had a three-dimensional pore network with a limiting pore diameter of 7.35 Å and maximum pore diameter of 12.39 Å. Calculated porosity characteristics for all predicted structures can be found in the supporting Chemiscope file.

Finally, state-of-the-art periodic DFT optimizations using the PBE functional[107] with D3 (ref. 108) dispersion correction were performed for the 982 likely synthesizable structures, in order to further verify the accuracy of our MLIPs.

In general, we have observed a strong correlation between relative ML and DFT energies, as shown in SI Fig. S10, where the energy of the experimentally matching **zni** structure (CSD IMIDZB02) was used as a reference. The slope of the linear fit slightly deviated from one, suggesting that our MLIP trained on data obtained by using the LDA functional gives a somewhat "compressed" energy scale, compared to the dispersion-corrected PBE functional. However, the standard error for the MLIP energies compared to the PBE + D3 DFT energies was found to be only 3.09 kJ mol^{-1}, which is quite small, considering that the overall energy landscape spans a window of 45 kJ mol^{-1}.

Furthermore, this new set of DFT-optimized larger structures allowed us to verify that the accuracy of MLIP predictions does not depend on the number of formula units contained in the crystallographic unit cell. Since our potentials were trained exclusively on structures containing 1–4 formula units, we tested whether the accuracy of the predictions remains high for structures with up to 16 formula units. By separately comparing the MLIP and PBE + D3 DFT energies for structures with 1–4, 6, 8, 12 and 16 formula units (there were only three structures with 14 formula units among the likely synthesizable forms of Zn**Im**$_2$, making statistical analysis meaningless for this group), we found that MLIPs perform similarly for all of these groups, both in terms of linear regression parameters and error distributions (SI Fig. S10 and 11). This highly rewarding result validates our training strategy, with the potentials trained on 1–4 formula units performing comparably across the entire range of 1–16 formula units per cell.

Analysis of the new DFT-based energy ranking supports the observation made with MLIP ranking that there are hypothetical structures of Zn**Im**$_2$ more stable than the lowest-energy reported **zni** and **coi** forms. Moreover, the global minimum based on DFT ranking has **sql** topology, further supporting the possibility of discovering **sql**-Zn**Im**$_2$ experimentally.

Assessment of synthetic feasibility remains a major challenge for computational materials design, since it requires analysis of multiple factors, affecting the likelihood of formation of a given structure. The empirical synthesizability criterion presents an initial foray into the much deeper exploration needed for detailed understanding of the synthesizability prediction. The key aspects that need to be incorporated in the future comprehensive analysis of synthetic feasibility for the predicted structures include lattice free energies,[109,110] host–guest energetic stabilization due to solvent inclusion within MOF pores,[111] and assessment of mechanical stability.[112] While individually these effects have been addressed by computational means, a combined simulation of all major factors, controlling the synthetic feasibility of the predicted structures, is beyond the boundaries of present-day simulation-based computational chemistry. Development of comprehensive ML models, trained on a combination of experimental and computational data,[113,114] however, brings exciting future opportunities in addressing this challenge in materials design.

Identification of unknown experimental structures through the assignment of powder diffraction patterns

An important challenge in the synthesis of new materials is structure determination of products of high-throughput syntheses. The synthesis does not always result in diffraction quality single crystals, but instead produces polycrystalline materials. This is often the case during solvothermal syntheses, but especially when using mechanochemistry. In this case, structure solution from powder X-ray diffraction (PXRD) data must be performed, which can be challenging. Recently, advances in electron diffraction methods provided a potential alternative; however this technique is still not widely available, and porous materials are often challenging subjects for ED, as they are extremely susceptible to electron beam damage.[115] Instead, we propose a combination of CSP and a PXRD based structure matching protocol as an alternative to *ab initio* structure solution. We here demonstrate the assignment of experimental

© 2026 The Author(s). Published by the Royal Society of Chemistry

PXRD patterns of mechanochemically synthesized MOFs against the thousands of predicted structures in the CSP landscape, as a way to determine the structures of new materials. The assignment is based on the GPWDF algorithm,[44] implemented in Critic2.[116,117] We first tested the protocol on a selection of experimental PXRD patterns with known crystal structures from our recent publication[18] on the mechanochemical solid form screening of Zn**Im**$_2$, in order to test the sensitivity and precision of this assignment method. Then, we performed an assignment for a pattern with an unknown crystal structure.

In the initial test we included PXRD patterns of four Zn**Im**$_2$ materials that were both identified by CSP and found in our mechanochemical screening, namely **zni** (CSD IMIDZB02), **crbT** (CSD GAKXAW) and two different solvates of the **cag** topology material (CSD VEJYUF01, prepared by milling with dimethylformamide (DMF) and chloroform liquid additives). First, PXRD patterns for all CSP predicted structures were simulated using the CSD Python API.[118] The simulated patterns were then compared to the selected experimental patterns, and the structures were ranked in ascending similarity order based on the variable-cell similarity index (DIFF), from the function of COMPAREVC[44] in Critic2. A variable-cell similarity index of 0 indicates a perfect match between the experimental powder pattern and the predicted structure, while a score of 1 means a full dissimilarity. The detailed results of these PXRD assignments, with the DIFF rankings of the matching predicted structures, are shown in SI Table S5. Among these assignments, the predicted structure with **zni** topology showed the lowest DIFF score of 0.02, being ranked as the 6th best match overall. The CSP structure matching the recently reported **crbT** form synthesized *via* liquid-assisted grinding with toluene was ranked as the 199th best match to the corresponding experimental pattern. Finally, the predicted structure for **cag**-Zn**Im**$_2$ ranked as 473rd and 763rd, respectively, when compared against the patterns for two different preparations for the **cag** form, milled with dimethylformamide and chloroform, respectively.

The results demonstrate considerable variations in the ranking of correct structural matches against the experimental data and give us an indication of what to expect when using this method for PXRD patterns whose true experimental structure is not yet known and needs to be determined. The lowest overall ranking for the matching of the **zni** structure is attributed to the lack of porosity in this structure, resulting in a very good match of the calculated diffraction peak intensities.

Conversely, for the **cag** and **crbT** forms, the experimental patterns were collected on materials with guest molecules occupying the structural voids, while the predicted structures are modelled with empty voids. This leads to a discrepancy between the simulated and experimental diffraction intensities and leads to higher DIFF scores. Given that most synthesized MOFs will have their pores occupied by guest molecules, this will be an important consideration for the future method development for the assignment of experimental PXRD patterns against CSP results for MOFs and porous materials in general.

Additionally, not only do guests inside MOF pores contribute electron density and thus change the intensity profile of the PXRD patterns, but they can also have a direct impact on the MOF itself. This is particularly true in the case of flexible MOFs, such as ZIFs. The two tested **cag** solvates are an excellent example, as we see that the CSP structure matching well with 0.5DMF@**cag**-Zn**Im**$_2$ (CSD: VEJYUF01) has very different DIFF scores when compared to the DMF and $CHCl_3$ solvates of cag-Zn**Im**$_2$. Namely, the DMF solvate provides a much better match.

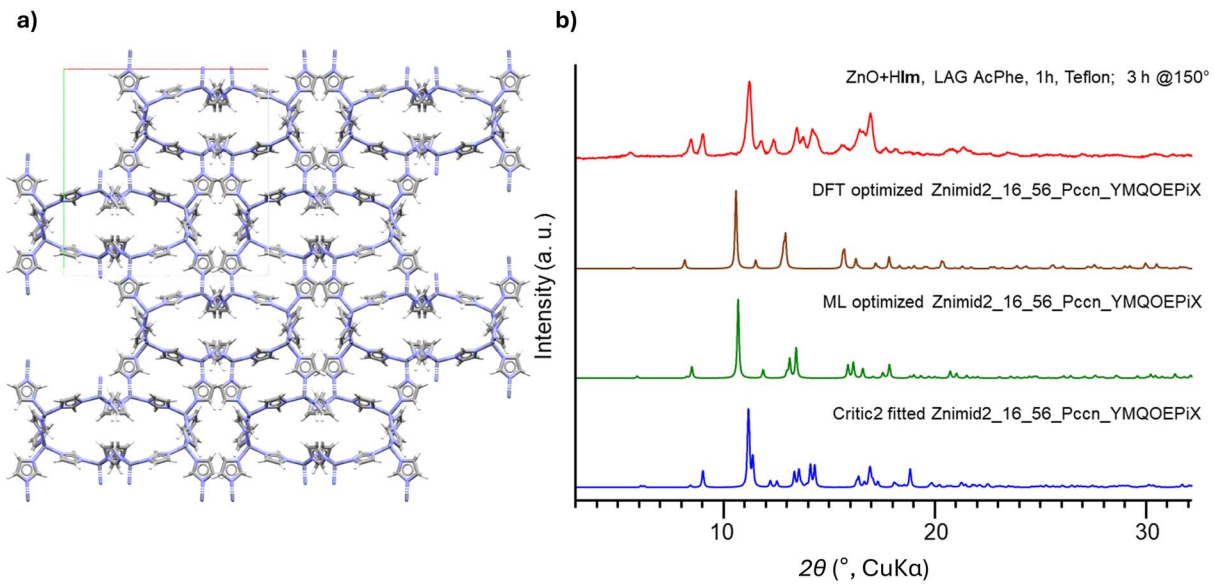


Fig. 9 (a) ML optimized crystal structure with a Chemiscope file data label 2470 (name Znimid2_16_56_Pccn_YMQOEPiX), viewed along the *c* axis; (b) PXRD patterns from top to bottom: (1) (red) experimentally collected unknown phase from our previous work,[18] obtained by heating **crbA**-Zn**Im**$_2$ at 150° for 3 hours; (2) (brown) DFT optimized predicted matching structure; (3) (green) ML optimized matching structure; (4) (blue) predicted matching structure after cell relaxation in Critic2.

© 2026 The Author(s). Published by the Royal Society of Chemistry

This is unsurprising when we take into consideration that the VEJYUF01 structure is exactly a DMF solvate of **cag**-Zn**Im**$_2$. Even without actually modelling the DMF guest in the CSP calculation, the effect of the guest on the conformation of the framework is visible in the quality of the match with the experimental structure.

With these observations in mind, we continued the exploration of our CSP energy landscape, aiming to gain more structural insights for experimentally unknown structures. We selected several PXRD patterns with unknown structures from our internal experimental findings, one of which yielded a CSP match, shown in Fig. 9. The DIFF score of the structure was 0.078 and it was found to be the 33rd lowest ranked structure among the whole list of CSP entries. Experimentally, the unknown structure was obtained by heating the acetophenone solvate of **crbA**-Zn**Im**$_2$ (CSD: GAKXOX) for 3 hours at 150 °C, resulting in a guest-free porous material of unknown structure. Besides the PXRD similarity, several other factors pointed in favor of this assignment: first, the predicted matching structure had the **crb** topology, the same as the parent phase from which the new material originated *via* thermal transformation, and second, the matching CSP structure fell into the set of structures deemed synthesizable based on the energy-porosity criteria described earlier (Fig. 4 and SI Section S7). However, Rietveld refinement[119] against the experimental PXRD data using the CSP matched structure (SI Fig. S36) was not in full agreement with this assignment. Not all the experimental PXRD peaks could be matched against the predicted structure, suggesting that a lower symmetry transformation may be needed to obtain the structural model in full agreement with the experimental data.[120]

The candidate structures described above are certainly not the only possible options for the future synthesis of new polymorphs of Zn**Im**$_2$. The great opportunities presented by CSP are brought by the diversity of structures found in these calculations. However, there also lies a challenge: CSP has been known to produce more structures than could be experimentally isolated,[121] with this being a general phenomenon, applicable to molecular crystals, inorganic materials and, most certainly, MOFs. The reasons for this are both experimental (inability to sample all possible synthetic conditions) and computational: predicted energy minima may be separated by a very high energy barrier, making them kinetically inaccessible, or, alternatively, the barriers may be too low, making some of the predicted polymorphs inherently kinetically unstable at any temperature above 0 K.[122] In line with these thoughts we are releasing the entire CSP dataset, hoping that our results will prove useful to the experimental MOF community, helping in the interpretation of existing results and acting as a guideline for the design of future experiments.

Methods

Crystal structure generation

Structure generation was performed using the *ab initio* random structure search (AIRSS)[29] algorithm, with randomly chosen symmetry constraints. The Wyckoff alignment of molecules (WAM)[28] method was used to exploit the point group symmetry of the individual building blocks and allow them to occupy appropriate special Wyckoff sites.[28]

Structures were generated separately for 1, 2, 3, 4, 6, 8, 10, 12, 14 and 16 formula units per primitive crystallographic unit cell, with each formula unit including one Zn atom and two imidazolate molecular fragments, placed at random positions within the trial unit cell. Further details of the structure generation are given in the SI section S1.1.

Generation of the DFT training dataset

The randomly generated structures containing 1–4 formula units per primitive cell were geometry-optimized using periodic density-functional theory (DFT) calculations within the code CASTEP19.[123] Calculations were performed using the LDA functional, with the plane-wave basis set was truncated at 400 eV cutoff. The ultrasoft pseudopotentials were used from the internal QC5 library of CASTEP. The first electronic Brillouin zone was sampled with a $2\pi \times 0.07$ Å^{-1} Monkhorst–Pack *k*-point grid.[124] Structures were optimized with respect to unit cell parameters and atomic positions, subject to the symmetry constraints imposed by WAM space group assignment. The following convergence criteria were used: a maximum energy change of 2×10^{-5} eV per atom; maximum atomic force of 0.05 eV Å^{-1}; maximum atom displacement of 10^{-3} Å; maximum residual stress of 0.1 GPa.

Training of machine-learned potentials

Periodic DFT optimization of Zn**Im**$_2$ structures, containing 1–4 formula units, provided us with 6000 distinct DFT trajectories, from which the crystal structures, representing individual optimization steps, were extracted. The data were then split into training, validation and test sets, making sure that each of the optimization trajectories was only used in one of those sets. This resulted in the training set, containing 10 972 structures, validation set with 2739 structures and test set with 5600 structures.

The training was performed using SchNetPack[66] v 1.0.0 (GitHub commit 6fe78ef), which allows the use of the polarizable interaction neural network (PaiNN) architecture[73] with a distance cutoff for pairwise interatomic interactions set to 5 Å (other model hyperparameters are described in SI Section S1.3).[125–127] The cutoff radius should not be seen as a local interaction limit: for multi-layer message-passing neural networks, the effective receptive field spans multiple interaction shells.[128,129] In our case, with a three-layer network, the effective receptive field is approximately 15 Å, which is certainly sufficient to distinguish different ZIF polymorphs. Separate MLIPs were constructed for energies and forces, and the reasons for training separate potentials rather than a single one are discussed in SI Section S2, while the statistical analysis of the agreement between DFT and MLIP energies and forces is presented in SI Fig. S1–S8.

© 2026 The Author(s). Published by the Royal Society of Chemistry

Geometry optimization and energy ranking

The MLIPs were used to optimize and rank the energies of structures containing 6–16 formula units, which would be too computationally expensive to study with periodic DFT. The structures generated by AIRSS + WAM were optimized using the MLIP trained on atomic forces, initially with a force tolerance of 0.05 eV $Å^{-1}$. Subsequently, the structures were energy-ranked by performing single point calculations using the energy MLIP. Structures found to be within 45 kJ mol^{-1} from the global energy minimum were then clustered with the aid of the simulated powder diffraction pattern (PXRD) comparison method, implemented in the code Critic2. Structures containing 1 to 4 formula units, which were originally used to train the ML potentials, were subjected to the same procedure, in order to have a complete structural landscape from 1 to 16 formula units per primitive cell.

The final set of structures was re-optimized using the force MLIP with a tighter force tolerance of 0.005 eV $Å^{-1}$, followed by single point calculation using the energy MLIP. At this point the presence of low-density 1D and 2D structures was noticed, with large separations between chains or layers. These structures were subjected to geometry optimization under increased pressure, as explained in SI Section S1.4.

The final combined set of structures was again clustered using the PXRD similarity algorithm, followed by geometry comparison *via* the COMPACK algorithm,[130] accessed through the CCDC Python API.[131] This resulted in 9626 unique structures.

Finally, structures were assessed in terms of chemical connectivity. Structures containing 3-, 5- or 6-coordinate zinc atoms were removed, leaving only frameworks with 4-coordinate nodes. In addition, structures containing Zn–C and N–N bonds were also removed from the final CSP set. The number of such erroneous structures, however, was small, with only 17 structures removed, leaving us with a final set of 9609 structures with proper ZIF connectivity, with 4-coordinate metal centers and doubly coordinated imidazolate linkers.

Finally, the void volumes and packing coefficients for the remaining structures were calculated using PLATON.[85]

Post-processing of the predicted structures

Network topologies were determined for the final set of the predicted Zn**Im**$_2$ structures using ToposPro.[92] Coordination networks were established using the default settings of the AutoCN module. The resulting nets were then simplified and analyzed using the default settings of the ADS module. The topological descriptors were compared against the built-in TTD database. A small subset of 61 structures could not be processed *via* the ToposPro interface; therefore their topologies were determined using the https://topcryst.com/ online server.

The structures were merged into a database using CSD Editor software. This database was then analyzed using ConQuest,[93] in order to extract bond length and bond angle distributions (SI Fig. S12 and 13). The bond length and angle distributions from the predicted structures of Zn**Im**$_2$ were compared with those of the experimental structures found in the Cambridge Structural Database (CSD).[94]

Periodic DFT optimization of the structures deemed likely experimental candidates

The 982 predicted structures deemed to be synthesizable based on the energy-porosity criterion (SI Section S7) were geometry-optimized with periodic DFT. These calculations used the PBE functional,[107] combined with Grimme D3 dispersion correction.[108] The plane-wave basis set truncated at 800 eV cutoff and CASTEP default ultrasoft pseudopotentials were used. The electronic *k*-point grid was sampled with a $2\pi \times 0.06$ $Å^{-1}$ Monkhorst–Pack *k*-point grid.[124] Convergence criteria were set as follows: a maximum energy change of 2×10^{-5} eV per atom; maximum atomic force of 0.05 eV $Å^{-1}$; maximum atom displacement of 10^{-3} Å; maximum residual stress of 0.05 GPa.

Mechanochemical synthesis

All results of mechanochemical syntheses presented herein utilize the data reported in our previous publication.[18] Mechanochemical ball milling reactions were performed by mixing zinc oxide (75.0 mg, 0.92 mmol), imidazole (125.5 mg, 1.84 mmol) and 100 μl of a liquid additive (methanol, toluene, chloroform or dimethylformamide, depending on the experiment) in a milling jar, containing two ball bearings. The samples were milled at 30 Hz for up to 90 min.

Comparison of experimental and simulated PXRD patterns

To start the PXRD comparisons, the experimental PXRD patterns in .raw format were converted into .xy format using the open-source software PowDLL,[132] whereas all predicted structures were supplied in .res format. Comparisons were performed on background-subtracted PXRD patterns, such that each experimental PXRD pattern was compared individually with each predicted structure (9626 in total). Finally, the predicted CSP structures were ranked based on ascending DIFF scores, where lower DIFF structures were accessed further to identify likely experimentally matching structures.

Conclusions

We have presented the crystal structure prediction (CSP) study aimed at uncovering the crystal energy landscape of a highly polymorphic MOF material Zn**Im**$_2$. The major step forward in CSP methodology, presented herein, was the introduction of MLIPs for efficient geometry optimization and energy ranking of the trial structures that allowed us to greatly extend the scope of the structural search, sampling millions of structures of the highest complexity, reaching unit cells containing 16 ZIF Zn**Im**$_2$ formula units, whereas our previous searches, where we utilized periodic DFT calculations for energy ranking, were limited to four formula units.

The large search space enabled by the use of MLIP manifested itself in an unprecedented topological diversity of the predicted structures with 1484 unique topologies. Between full and partial matches, we have located all but one of the

© 2026 The Author(s). Published by the Royal Society of Chemistry

experimentally reported polymorphs of Zn**Im**$_2$ falling within the boundaries of the defined search space. Moreover, the analysis of the energy-density map and exploration of calculated void volume within the predicted structures allowed us to suggest some likely candidates that may lead to future new polymorphs of Zn**Im**$_2$, including a 2D form with **sql** topology.

We have then demonstrated the protocol of using CSP-generated structures for the assignment of mechanochemically synthesized materials, by comparing the experimental and simulated powder diffraction patterns. Given the propensity of mechanochemistry to reveal new MOF solid forms, this assignment approach is particularly important for the interpretation of experimental results.

Finally, releasing the entire CSP dataset will allow the readers to navigate the predicted structures and analyze their structural, topological and porosity characteristics. We hope that this will prove useful in guiding future experimental discovery of new ZIF materials.

To conclude, this work marks a major step in the development of CSP for MOFs, bringing it to the forefront of high-throughput computational discovery of new MOF structures with diverse packing arrangements, topological connectivities and functional properties.

Author contributions

Conceptualization: A. J. M.; M. A.; data curation: Y. X., J. P. D., and M. A.; formal analysis: Y. X., I. B., and M. A.; funding acquisition: I. B., A. J. M., and M. A.; investigation: Y. X. and K. L.; resources: I. B., A. J. M., and M. A.; software: J. D. and J. P. D.; writing – original draft: M. A.; writing – review & editing: all authors; visualization: Y. X., I. B., and M. A.

Conflicts of interest

J. P. D. is a member of Symmetric Group LLP that licenses force fields commercially.

Data availability

Supplementary information (SI): detailed description of computational and experimental procedures; statistical analysis of MLIP energies and forces; CSP plots showing distributions of individual topologies; assignment of experimental PXRD patterns. See DOI: https://doi.org/10.1039/d6sc04307k.

In addition, an accompanying dataset can be accessed *via* https://doi.org/10.5281/zenodo.21513301. This dataset includes the ML potential files; the training, validation and testing datasets; all the predicted crystal structures in RES file format with the associated data (energies, unit cell parameters, calculated porosity characteristics and network topologies) in a separate CSV file. In addition, a Chemiscope[133] file is provided; for convenient data visualization the file can be uploaded to the https://chemiscope.org/ platform.

Finally, output files for the periodic DFT optimizations of structures deemed likely synthesizable are also provided in the Zenodo dataset.

Acknowledgements

We would like to thank Professor Tomislav Friščić for the support and helpful discussions. YX and MA acknowledge the support of the National Science Center (NCN) *via* grant 2018/31/D/ST5/03619. MA further acknowledges the NCN grant 2023/51/B/ST5/01555. This work was supported by the "Developing Research Support" Program of the Croatian Ministry of Science and the Croatian Science Foundation, funded by the European Union from the NextGenerationEU program through grant NPOO.C3.2.R2-I1.06.0049. AJM gratefully acknowledges networking support from CCP-NC (UKRI grant EP/T026642/1), CCP9 (EP/T026375/1), and UKCP (EP/P022561/1). We gratefully acknowledge Poland's high-performance Infrastructure PLGrid ACC Cyfronet AGH for providing computer facilities and support within computational grant PLG/2025/018422. The authors acknowledge computational support from the UK national high performance computing service, ARCHER2, for which access was obtained *via* the UKCP consortium and funded by EPSRC grant ref EP/X035891/1. This work was performed using resources provided by the Cambridge Service for Data Driven Discovery (CSD3) operated by the University of Cambridge Research Computing Service (https://www.csd3.cam.ac.uk/), provided by Dell EMC and Intel using Tier-2 funding from the Engineering and Physical Sciences Research Council (capital grant EP/P020259/1).

Notes and references

1 M. Eddaoudi, H. Li and O. M. Yaghi, *J. Am. Chem. Soc.*, 2000, **122**, 1391–1397.
2 H. Bux, C. Chmelik, R. Krishna and J. Caro, *J. Membr. Sci.*, 2011, **369**, 284–289.
3 T. Wang, L. Gao, J. Hou, S. J. A. Herou, J. T. Griffiths, W. Li, J. Dong, S. Gao, M.-M. Titirici, R. V. Kumar, A. K. Cheetham, X. Bao, Q. Fu and S. K. Smoukov, *Nat. Commun.*, 2019, **10**, 1340.
4 I. Luz, F. X. Llabrés i Xamena and A. Corma, *J. Catal.*, 2010, **276**, 134–140.
5 A. J. Howarth, M. J. Katz, T. C. Wang, A. E. Platero-Prats, K. W. Chapman, J. T. Hupp and O. K. Farha, *J. Am. Chem. Soc.*, 2015, **137**, 7488–7494.
6 Y. Liu, A. J. Howarth, J. T. Hupp and O. K. Farha, *Angew. Chem., Int. Ed.*, 2015, **54**, 9001–9005.
7 P. G. Yot, L. Vanduyfhuys, E. Alvarez, J. Rodriguez, J.-P. Itié, P. Fabry, N. Guillou, T. Devic, I. Beurroies, P. L. Llewellyn, V. Van Speybroeck, C. Serre and G. Maurin, *Chem. Sci.*, 2016, **7**, 446–450.
8 H.-J. Son, S. Jin, S. Patwardhan, S. J. Wezenberg, N. C. Jeong, M. So, C. E. Wilmer, A. A. Sarjeant, G. C. Schatz, R. Q. Snurr, O. K. Farha, G. P. Wiederrecht and J. T. Hupp, *J. Am. Chem. Soc.*, 2013, **135**, 862–869.
9 H. M. Titi, J. M. Marrett, G. Dayaker, M. Arhangelskis, C. Mottillo, A. J. Morris, G. P. Rachiero, T. Friščić and R. D. Rogers, *Sci. Adv.*, 2019, **5**, eaav9044.
10 Y. Xu, Y. Wang, Y. Zhong, G. Lei, Z. Li, J. Zhang and T. Zhang, *Inorg. Chem.*, 2021, **60**, 5100–5106.

© 2026 The Author(s). Published by the Royal Society of Chemistry

11 I. Huskić, N. Novendra, D.-W. Lim, F. Topić, H. M. Titi, I. V. Pekov, S. V. Krivovichev, A. Navrotsky, H. Kitagawa and T. Friščić, *Chem. Sci.*, 2019, **10**, 4923–4929.

12 X. Tian, F. Li, Z. Tang, S. Wang, K. Weng, D. Liu, S. Lu, W. Liu, Z. Fu, W. Li, H. Qiu, M. Tu, H. Zhang and J. Li, *Nat. Commun.*, 2024, **15**, 2920.

13 S. M. Moosavi, A. Nandy, K. M. Jablonka, D. Ongari, J. P. Janet, P. G. Boyd, Y. Lee, B. Smit and H. J. Kulik, *Nat. Commun.*, 2020, **11**, 4068.

14 A. Phan, C. J. Doonan, F. J. Uribe-Romo, C. B. Knobler, M. O'Keeffe and O. M. Yaghi, *Acc. Chem. Res.*, 2010, **43**, 58–67.

15 I. Brekalo, C. M. Kane, A. N. Ley, J. R. Ramirez, T. Friščić and K. T. Holman, *J. Am. Chem. Soc.*, 2018, **140**, 10104–10108.

16 J. Cravillon, C. A. Schröder, H. Bux, A. Rothkirch, J. Caro and M. Wiebcke, *CrystEngComm*, 2012, **14**, 492–498.

17 R. N. Widmer, G. I. Lampronti, S. Chibani, C. W. Wilson, S. Anzellini, S. Farsang, A. K. Kleppe, N. P. M. Casati, S. G. MacLeod, S. A. T. Redfern, F.-X. Coudert and T. D. Bennett, *J. Am. Chem. Soc.*, 2019, **141**, 9330–9337.

18 I. Brekalo, K. Lisac, J. R. Ramirez, P. Pongrac, A. Puškarić, S. Valić, Y. Xu, M. Ferguson, J. M. Marrett, M. Arhangelskis, T. Friščić and K. T. Holman, *J. Am. Chem. Soc.*, 2025, **147**, 27413–27430.

19 P. Cui, D. P. McMahon, P. R. Spackman, B. M. Alston, M. A. Little, G. M. Day and A. I. Cooper, *Chem. Sci.*, 2019, **10**, 9988–9997.

20 M. O'Shaughnessy, J. Glover, R. Hafizi, M. Barhi, R. Clowes, S. Y. Chong, S. P. Argent, G. M. Day and A. I. Cooper, *Nature*, 2024, **630**, 102–108.

21 J. H. Kim, J. S. Kim, Y. H. Kim, B. Jun, Y. J. Jang and S. U. Lee, *J. Am. Chem. Soc.*, 2025, **147**, 47381–47391.

22 Y. Hinuma, T. Hatakeyama, Y. Kumagai, L. A. Burton, H. Sato, Y. Muraba, S. Iimura, H. Hiramatsu, I. Tanaka, H. Hosono and F. Oba, *Nat. Commun.*, 2016, **7**, 11962.

23 L. Marleny Rodriguez-Albelo, A. R. Ruiz-Salvador, A. Sampieri, D. W. Lewis, A. Gómez, B. Nohra, P. Mialane, J. Marrot, F. Sécheresse, C. Mellot-Draznieks, R. Ngo Biboum, B. Keita, L. Nadjo and A. Dolbecq, *J. Am. Chem. Soc.*, 2009, **131**, 16078–16087.

24 D. W. Lewis, A. R. Ruiz-Salvador, A. Gómez, L. M. Rodriguez-Albelo, F.-X. Coudert, B. Slater, A. K. Cheetham and C. Mellot-Draznieks, *CrystEngComm*, 2009, **11**, 2272.

25 M. A. Addicoat, D. E. Coupry and T. Heine, *J. Phys. Chem. A*, 2014, **118**, 9607–9614.

26 S. Lee, D. Nam, D. C. Yang and W. Choe, *Small*, 2023, **19**, 2300036.

27 S. Lee, H. Jeong, S. Jung, Y. Kim, E. Cho, J. Nam, D. ChangMo Yang, D. Y. Shin, J.-H. Lee, H. Oh and W. Choe, *JACS Au*, 2025, **5**, 1460–1470.

28 J. P. Darby, M. Arhangelskis, A. D. Katsenis, J. M. Marrett, T. Friščić and A. J. Morris, *Chem. Mater.*, 2020, **32**, 5835–5844.

29 C. J. Pickard and R. J. Needs, *J. Phys.:Condens. Matter*, 2011, **23**, 053201.

30 Y. Xu, J. M. Marrett, H. M. Titi, J. P. Darby, A. J. Morris, T. Friščić and M. Arhangelskis, *J. Am. Chem. Soc.*, 2023, **145**, 3515–3525.

31 Z. Akimbekov, A. D. Katsenis, G. P. Nagabhushana, G. Ayoub, M. Arhangelskis, A. J. Morris, T. Friščić and A. Navrotsky, *J. Am. Chem. Soc.*, 2017, **139**, 7952–7957.

32 M. Arhangelskis, A. D. Katsenis, N. Novendra, Z. Akimbekov, D. Gandrath, J. M. Marrett, G. Ayoub, A. J. Morris, O. K. Farha, T. Friščić and A. Navrotsky, *Chem. Mater.*, 2019, **31**, 3777–3783.

33 N. Novendra, J. M. Marrett, A. D. Katsenis, H. M. Titi, M. Arhangelskis, T. Friščić and A. Navrotsky, *J. Am. Chem. Soc.*, 2020, **142**, 21720–21729.

34 G. J. Leonel, C. B. Lennox, Y. Xu, M. Arhangelskis, T. Friščić and A. Navrotsky, *J. Phys. Chem. C*, 2023, **127**, 19520–19526.

35 Q. Shi, X. Kang, F.-N. Shi and J. Dong, *Chem. Commun.*, 2015, **51**, 1131–1134.

36 E. V. Podryabinkin, E. V. Tikhonov, A. V. Shapeev and A. R. Oganov, *Phys. Rev. B*, 2019, **99**, 1–7.

37 P. Friederich, F. Häse, J. Proppe and A. Aspuru-Guzik, *Nat. Mater.*, 2021, **20**, 750–761.

38 S. Takamoto, C. Shinagawa, D. Motoki, K. Nakago, W. Li, I. Kurata, T. Watanabe, Y. Yayama, H. Iriguchi, Y. Asano, T. Onodera, T. Ishii, T. Kudo, H. Ono, R. Sawada, R. Ishitani, M. Ong, T. Yamaguchi, T. Kataoka, A. Hayashi, N. Charoenphakdee and T. Ibuka, *Nat. Commun.*, 2022, **13**, 1–6.

39 A. Merchant, S. Batzner, S. S. Schoenholz, M. Aykol, G. Cheon and E. D. Cubuk, *Nature*, 2023, **624**, 80–85.

40 M. Alverson, S. Baird, R. Murdock, H. Enoch Sin-Hang, J. Johnson and T. D. Sparks, *Digital Discovery*, 2024, **69**, 30–33.

41 J. P. Darby, A. F. Harper, J. R. Nelson and A. J. Morris, *Phys. Rev. Mater.*, 2024, **8**, 105002.

42 C. J. Pickard, *Phys. Rev. B*, 2022, **106**, 014102.

43 P. T. Salzbrenner, S. H. Joo, L. J. Conway, P. I. C. Cooke, B. Zhu, M. P. Matraszek, W. C. Witt and C. J. Pickard, *J. Chem. Phys.*, 2023, **159**, 144801.

44 A. Otero-de-la-Roza, *J. Appl. Crystallogr.*, 2024, **57**, 1401–1414.

45 M. Mayo, K. J. Griffith, C. J. Pickard and A. J. Morris, *Chem. Mater.*, 2016, **28**, 2011–2021.

46 B. Zhu and D. O. Scanlon, *ACS Appl. Energy Mater.*, 2022, **5**, 575–584.

47 N. P. Salke, M. M. Davari Esfahani, N. Yedukondalu, Y. Zhang, I. A. Kruglov, J. Zhou, E. Greenberg, V. B. Prakapenka, J. Liu, A. R. Oganov and J.-F. Lin, *Inorg. Chem.*, 2020, **59**, 5303–5312.

48 J. R. Nelson, R. J. Needs and C. J. Pickard, *Phys. Chem. Chem. Phys.*, 2015, **17**, 6889–6895.

49 S. Ninet, F. Datchi, P. Dumas, M. Mezouar, G. Garbarino, A. Mafety, C. J. Pickard, R. J. Needs and A. M. Saitta, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 2014, **89**, 174103.

50 J. Hou, H. Cui, A. R. Oganov, H.-F. Li, X.-J. Weng, X.-F. Zhou, H.-T. Wang and X. Dong, *Phys. Rev. B*, 2024, **109**, 174102.

© 2026 The Author(s). Published by the Royal Society of Chemistry

51 S. Pan, T. Huang, A. Vazan, Z. Liang, C. Liu, J. Wang, C. J. Pickard, H.-T. Wang, D. Xing and J. Sun, *Nat. Commun.*, 2023, **14**, 1165.

52 N. Karimitari, W. J. Baldwin, E. W. Muller, Z. J. L. Bare, W. J. Kennedy, G. Csányi and C. Sutton, *J. Am. Chem. Soc.*, 2024, **146**, 27392–27404.

53 J. T. A. Jones, T. Hasell, X. Wu, J. Bacsa, K. E. Jelfs, M. Schmidtmann, S. Y. Chong, D. J. Adams, A. Trewin, F. Schiffman, F. Cora, B. Slater, A. Steiner, G. M. Day and A. I. Cooper, *Nature*, 2011, **474**, 367–371.

54 A. V. Kazantsev, P. G. Karamertzanis, C. S. Adjiman, C. C. Pantelides, S. L. Price, P. T. A. Galek, G. M. Day and A. J. Cruz-Cabeza, *Int. J. Pharm.*, 2011, **418**, 168–178.

55 N. Kim, S. Kim, M. Kim, J. Park and S. Ahn, in *International Conference on Learning Representations*, ed. Y. Yue, A. Garg, N. Peng, F. Sha and R. Yu, 2025, pp. 98142–98162.

56 M. O'Keeffe, M. A. Peskov, S. J. Ramsden and O. M. Yaghi, *Acc. Chem. Res.*, 2008, **41**, 1782–1789.

57 V. A. Blatov, A. P. Shevchenko and D. M. Proserpio, *Cryst. Growth Des.*, 2014, **14**, 3576–3586.

58 K. S. Park, Z. Ni, A. P. Cote, J. Y. Choi, R. Huang, F. J. Uribe-Romo, H. K. Chae, M. O'Keeffe and O. M. Yaghi, *Proc. Natl. Acad. Sci. U. S. A.*, 2006, **103**, 10186–10191.

59 G. A. V. Martins, P. J. Byrne, P. Allan, S. J. Teat, A. M. Z. Slawin, Y. Li and R. E. Morris, *Dalton Trans.*, 2010, **39**, 1758–1762.

60 E. C. Spencer, R. J. Angel, N. L. Ross, B. E. Hanson and J. A. K. Howard, *J. Am. Chem. Soc.*, 2009, **131**, 4022–4026.

61 R. Banerjee, A. Phan, B. Wang, C. Knobler, H. Furukawa, M. O'Keeffe and O. M. Yaghi, *Science*, 2008, **319**, 939–943.

62 Y.-Q. Tian, Y.-M. Zhao, Z.-X. Chen, G.-N. Zhang, L.-H. Weng and D.-Y. Zhao, *Chem.–Eur. J.*, 2007, **13**, 4146–4154.

63 S. Guo, H.-Z. Li, Z.-W. Wang, Z.-Y. Zhu, S.-H. Zhang, F. Wang and J. Zhang, *Inorg. Chem. Front.*, 2022, **9**, 2011–2015.

64 Q. Shi, W.-J. Xu, R.-K. Huang, W.-X. Zhang, Y. Li, P. Wang, F.-N. Shi, L. Li, J. Li and J. Dong, *J. Am. Chem. Soc.*, 2016, **138**, 16232–16235.

65 N. Artrith and A. Urban, *Comput. Mater. Sci.*, 2016, **114**, 135–150.

66 K. T. Schütt, P. Kessel, M. Gastegger, K. A. Nicoli, A. Tkatchenko and K.-R. Müller, *J. Chem. Theory Comput.*, 2019, **15**, 448–455.

67 I. Batatia, P. Benner, Y. Chiang, A. M. Elena, D. P. Kovács, J. Riebesell, X. R. Advincula, M. Asta, M. Avaylon, W. J. Baldwin, F. Berger, N. Bernstein, A. Bhowmik, F. Bigi, S. M. Blau, V. Cărare, M. Ceriotti, S. Chong, J. P. Darby, S. De, F. Della Pia, V. L. Deringer, R. Elijošius, Z. El-Machachi, E. Fako, F. Falcioni, A. C. Ferrari, J. L. A. Gardner, M. J. Gawkowski, A. Genreith-Schriever, J. George, R. E. A. Goodall, J. Grandel, C. P. Grey, P. Grigorev, S. Han, W. Handley, H. H. Heenen, K. Hermansson, C. H. Ho, S. Hofmann, C. Holm, J. Jaafar, K. S. Jakob, H. Jung, V. Kapil, A. D. Kaplan, N. Karimitari, J. R. Kermode, P. Kourtis, N. Kroupa, J. Kullgren, M. C. Kuner, D. Kuryla, G. Liepuoniute, C. Lin, J. T. Margraf, I.-B. Magdău, A. Michaelides, J. H. Moore, A. A. Naik, S. P. Niblett, S. W. Norwood, N. O'Neill, C. Ortner, K. A. Persson, K. Reuter, A. S. Rosen, L. A. M. Rosset, L. L. Schaaf, C. Schran, B. X. Shi, E. Sivonxay, T. K. Stenczel, C. Sutton, V. Svahn, T. D. Swinburne, J. Tilly, C. van der Oord, S. Vargas, E. Varga-Umbrich, T. Vegge, M. Vondrák, Y. Wang, W. C. Witt, T. Wolf, F. Zills and G. Csányi, *J. Chem. Phys.*, 2025, **163**, 184110.

68 B. Deng, P. Zhong, K. J. Jun, J. Riebesell, K. Han, C. J. Bartel and G. Ceder, *Nat. Mach. Intell.*, 2023, **5**, 1031–1041.

69 S. Batzner, A. Musaelian, L. Sun, M. Geiger, J. P. Mailoa, M. Kornbluth, N. Molinari, T. E. Smidt and B. Kozinsky, *Nat. Commun.*, 2022, **13**, 2453.

70 A. Jain, S. P. Ong, G. Hautier, W. Chen, W. D. Richards, S. Dacek, S. Cholia, D. Gunter, D. Skinner, G. Ceder and K. A. Persson, *APL Mater.*, 2013, **1**, 011002.

71 D. S. Levine, M. Shuaibi, E. W. C. Spotte-Smith, M. G. Taylor, M. R. Hasyim, K. Michel, I. Batatia, G. Csányi, M. Dzamba, P. Eastman, N. C. Frey, X. Fu, V. Gharakhanyan, A. S. Krishnapriyan, J. A. Rackers, S. Raja, A. Rizvi, A. S. Rosen, Z. Ulissi, S. Vargas, C. L. Zitnick, S. M. Blau and B. M. Wood, *arXiv*, 2025, preprint, arXiv:2505.08762, DOI: 10.48550/arXiv.2505.08762.

72 P. Eastman, P. K. Behara, D. L. Dotson, R. Galvelis, J. E. Herr, J. T. Horton, Y. Mao, J. D. Chodera, B. P. Pritchard, Y. Wang, G. De Fabritiis and T. E. Markland, *Sci. Data*, 2023, **10**, 11, DOI: 10.1038/s41597-022-01882-6.

73 K. T. Schütt, O. T. Unke and M. Gastegger, *arXiv*, 2021, preprint, arXiv:2102.03150, DOI: 10.48550/arXiv.2102.03150.

74 K. Ito, T. Yokoi, K. Hyodo and H. Mori, *npj Comput. Mater.*, 2024, **10**, 255.

75 L. Triestram and F.-X. Coudert, *Chem. Sci.*, 2026, **17**, 6734–6745.

76 J. T. Hughes, T. D. Bennett, A. K. Cheetham and A. Navrotsky, *J. Am. Chem. Soc.*, 2013, **135**, 598–601.

77 I. A. Baburin and S. Leoni, *J. Mater. Chem.*, 2012, **22**, 10152–10154.

78 S. Springer, I. A. Baburin, T. Heinemeyer, J. G. Schiffmann, L. van Wüllen, S. Leoni and M. Wiebcke, *CrystEngComm*, 2016, **18**, 2477–2489.

79 C. A. Schröder, I. A. Baburin, L. van Wüllen, M. Wiebcke and S. Leoni, *CrystEngComm*, 2013, **15**, 4036–4040.

80 R. Galvelis, B. Slater, R. Chaudret, B. Creton, C. Nieto-Draghi and C. Mellot-Draznieks, *CrystEngComm*, 2013, **15**, 9603–9612.

81 C. Mellot-Draznieks and B. Kerkeni, *Mol. Simul.*, 2014, **40**, 25–32.

82 M. Arhangelskis, A. D. Katsenis, A. J. Morris and T. Friščić, *Chem. Sci.*, 2018, **9**, 3367–3375.

83 A. G. Slater, P. S. Reiss, A. Pulido, M. A. Little, D. L. Holden, L. Chen, S. Y. Chong, B. M. Alston, R. Clowes, M. Haranczyk, M. E. Briggs, T. Hasell, G. M. Day and A. I. Cooper, *ACS Cent. Sci.*, 2017, **3**, 734–742.

© 2026 The Author(s). Published by the Royal Society of Chemistry

84 C. Zhao, L. Chen, Y. Che, Z. Pang, X. Wu, Y. Lu, H. Liu, G. M. Day and A. I. Cooper, *Nat. Commun.*, 2021, **12**, 817.
85 A. L. Spek, *J. Appl. Crystallogr.*, 2003, **36**, 7–13.
86 A. Pulido, L. Chen, T. Kaczorowski, D. Holden, M. A. Little, S. Y. Chong, B. J. Slater, D. P. McMahon, B. Bonillo, C. J. Stackhouse, A. Stephenson, C. M. Kane, R. Clowes, T. Hasell, A. I. Cooper and G. M. Day, *Nature*, 2017, **543**, 657–664.
87 G. M. Day and A. I. Cooper, *Adv. Mater.*, 2018, **30**, 1704944.
88 E. O. Pyzer-Knapp, H. P. G. Thompson, F. Schiffmann, K. E. Jelfs, S. Y. Chong, M. A. Little, A. I. Cooper and G. M. Day, *Chem. Sci.*, 2014, **5**, 2235–2245.
89 D. P. McMahon, A. Stephenson, S. Y. Chong, M. A. Little, J. T. A. Jones, A. I. Cooper and G. M. Day, *Faraday Discuss.*, 2018, **211**, 383–399.
90 J. Yang, Y.-B. Zhang, Q. Liu, C. A. Trickett, E. Gutiérrez-Puebla, M. Á. Monge, H. Cong, A. Aldossary, H. Deng and O. M. Yaghi, *J. Am. Chem. Soc.*, 2017, **139**, 6448–6455.
91 H. N. Abdelhamid, Z. Huang, A. M. El-Zohry, H. Zheng and X. Zou, *Inorg. Chem.*, 2017, **56**, 9139–9146.
92 V. A. Blatov, A. P. Shevchenko and D. M. Proserpio, *Cryst. Growth Des.*, 2014, **14**, 3576–3586.
93 I. J. Bruno, J. C. Cole, P. R. Edgington, M. Kessler, C. F. Macrae, P. McCabe, J. Pearson and R. Taylor, *Acta Crystallogr.*, 2002, **B58**, 389–397.
94 C. R. Groom, I. J. Bruno, M. P. Lightfoot and S. C. Ward, *Acta Crystallogr.*, 2016, **B72**, 171–179.
95 N. Masciocchi, F. Castelli, P. M. Forster, M. M. Tafoya and A. K. Cheetham, *Inorg. Chem.*, 2003, **42**, 6147–6152.
96 I. R. Speight, I. Huskić, M. Arhangelskis, H. M. Titi, R. S. Stein, T. P. Hanusa and T. Friščić, *Chem.–Eur. J.*, 2020, **26**, 1811–1818.
97 I. A. Baburin and S. Leoni, *CrystEngComm*, 2010, **12**(10), 2809–2816, DOI: 10.1039/b926717d.
98 O. Karagiaridi, M. B. Lalonde, W. Bury, A. A. Sarjeant, O. K. Farha and J. T. Hupp, *J. Am. Chem. Soc.*, 2012, **134**, 18790–18796.
99 C. A. Schröder, S. Saha, K. Huber, S. Leoni and M. Wiebcke, *Z. Kristallogr.–Cryst. Mater.*, 2014, **229**, 807–822.
100 G. F. Turner, I. M. Jones, G. Koutsantonis, C. J. Kepert and S. A. Moggach, *Cryst. Growth Des.*, 2024, **24**, 5179–5192.
101 M. O'Shaughnessy, P. R. Spackman, M. A. Little, L. Catalano, A. James, G. M. Day and A. I. Cooper, *Chem. Commun.*, 2022, **58**, 13254–13257.
102 Q. Zhu, J. Johal, D. E. Widdowson, Z. Pang, B. Li, C. M. Kane, V. Kurlin, G. M. Day, M. A. Little and A. I. Cooper, *J. Am. Chem. Soc.*, 2022, **144**, 9893–9901.
103 B. J. Eckstein, H. R. Martin, M. P. Moghadasnia, A. Halder, P. S. Brodale, P. Le Magueres, P. W. V. Butler, K. A. Forrest, L. Ritter, R. A. Klein, H. J. Moon, C. Li, S. E. Massimi, Y. Cheng, C. H. Hendon, G. M. Day, B. Space, C. M. Brown and C. M. McGuirk, *Nat. Synth.*, 2026, **5**, 1327–1339.
104 F. Fajardo-Rojas, R. Anderson, K. Ardila, A. J. Pak and D. A. Gómez-Gualdrón, *Chem. Mater.*, 2026, **38**, 607–618.
105 A. Niyongabo Rubungo, F. Fajardo-Rojas, D. A. Gómez-Gualdrón and A. B. Dieng, *J. Am. Chem. Soc.*, 2025, **147**, 48035–48045.
106 P. J. Beldon, L. Fábián, R. S. Stein, A. Thirumurugan, A. K. Cheetham and T. Friščić, *Angew. Chem., Int. Ed.*, 2010, **49**, 9640–9643.
107 J. P. Perdew, K. Burke and M. Ernzerhof, *Phys. Rev. Lett.*, 1996, **77**, 3865–3868.
108 S. Grimme, J. Antony, S. Ehrlich and H. Krieg, *J. Chem. Phys.*, 2010, **132**, 154104.
109 A. Niyongabo Rubungo, F. Fajardo-Rojas, D. A. Gómez-Gualdrón and A. B. Dieng, *J. Am. Chem. Soc.*, 2025, **147**, 48035–48045.
110 J. Mao, N. Jiang, A. Darù, A. S. Filatov, J. E. Burch, J. Hofmann, S. M. Vornholt, K. W. Chapman, J. S. Anderson and A. L. Ferguson, *J. Am. Chem. Soc.*, 2025, **147**, 17651–17667.
111 R. Wang, B. C. Bukowski, J. Duan, K. Zhang, R. Q. Snurr and J. T. Hupp, *ACS Appl. Mater. Interfaces*, 2023, **15**, 51854–51862.
112 L. Bouëssel du Bourg, A. U. Ortiz, A. Boutin and F.-X. Coudert, *APL Mater.*, 2014, **2**, 124110.
113 C. G. Livas, P. N. Trikalitis and G. E. Froudakis, *J. Chem. Inf. Model.*, 2024, **64**, 8193–8200.
114 E. R. Antoniuk, G. Cheon, G. Wang, D. Bernstein, W. Cai and E. J. Reed, *npj Comput. Mater.*, 2023, **9**, 155.
115 E. Yörük and P. Naumov, *Small Methods*, 2026, **10**, 1–25.
116 A. Otero-de-la-Roza, M. A. Blanco, A. M. Pendás and V. Luaña, *Comput. Phys. Commun.*, 2009, **180**, 157–166.
117 A. Otero-de-la-Roza, E. R. Johnson and V. Luaña, *Comput. Phys. Commun.*, 2014, **185**, 1007–1018.
118 R. A. Sykes, N. T. Johnson, C. J. Kingsbury, J. Harter, A. G. P. Maloney, I. J. Sugden, S. C. Ward, I. J. Bruno, S. A. Adcock, P. A. Wood, P. McCabe, A. A. Moldovan, F. Atkinson, I. Giangreco and J. C. Cole, *J. Appl. Crystallogr.*, 2024, **57**, 1235–1250.
119 H. M. Rietveld, *J. Appl. Crystallogr.*, 1969, **2**, 65–71.
120 C. B. Lennox, J.-L. Do, J. G. Crew, M. Arhangelskis, H. M. Titi, A. J. Howarth, O. K. Farha and T. Friščić, *Chem. Sci.*, 2021, **12**, 14499–14506.
121 S. L. Price, *Acta Crystallogr.*, 2013, **B69**, 313–328.
122 S. Yang and G. M. Day, *Commun. Chem.*, 2022, **5**, 86.
123 S. J. Clark, M. D. Segall, C. J. Pickard, P. J. Hasnip, M. I. J. Probert, K. Refson and M. C. Payne, *Z. Kristallogr.–Cryst. Mater.*, 2005, **220**, 567–570.
124 H. J. Monkhorst and J. D. Pack, *Phys. Rev. B*, 1976, **13**, 5188–5192.
125 J. Westermayr, M. Gastegger and P. Marquetand, *J. Phys. Chem. Lett.*, 2020, **11**, 3828–3834.
126 W. G. Stark, J. Westermayr, O. A. Douglas-Gallardo, J. Gardner, S. Habershon and R. J. Maurer, *J. Phys. Chem. C*, 2023, **127**, 24168–24182.
127 W. G. Stark, C. van der Oord, I. Batatia, Y. Zhang, B. Jiang, G. Csányi and R. J. Maurer, *Mach. Learn.: Sci. Technol.*, 2024, **5**, 030501.
128 D. Kim and B. Cheng, *J. Chem. Phys.*, 2026, **164**, 060901.

© 2026 The Author(s). Published by the Royal Society of Chemistry

129 E. Annevelink and V. Shankar, *arXiv*, 2026, preprint, arXiv:2602.02228, DOI: 10.48550/arXiv.2602.02228.
130 J. A. Chisholm and W. D. S. Motherwell, *J. Appl. Crystallogr.*, 2005, **38**, 228–231.
131 R. A. Sykes, N. T. Johnson, C. J. Kingsbury, J. Harter, A. G. P. Maloney, I. J. Sugden, S. C. Ward, I. J. Bruno, S. A. Adcock, P. A. Wood, P. McCabe, A. A. Moldovan, F. Atkinson, I. Giangreco and J. C. Cole, *J. Appl. Crystallogr.*, 2024, **57**, 1235–1250.
132 N. Kourkoumelis, *Powder Diffr.*, 2013, **28**, 137–148.
133 G. Fraux, R. Cersonsky and M. Ceriotti, *J. Open Source Software*, 2020, **5**, 2117.

© 2026 The Author(s). Published by the Royal Society of Chemistry