



HAL
open science

Multiscale Modeling of Physical Properties of Nanoporous Frameworks: Predicting Mechanical, Thermal, and Adsorption Behavior

Arthur Hardiagon, François-Xavier Coudert

► **To cite this version:**

Arthur Hardiagon, François-Xavier Coudert. Multiscale Modeling of Physical Properties of Nanoporous Frameworks: Predicting Mechanical, Thermal, and Adsorption Behavior. *Accounts of Chemical Research*, 2024, 57 (11), pp.1620-1632. 10.1021/acs.accounts.4c00161 . hal-04578293

HAL Id: hal-04578293

<https://hal.science/hal-04578293v1>

Submitted on 16 May 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

Multi-Scale Modeling of Physical Properties of Nanoporous Frameworks: Predicting Mechanical, Thermal and Adsorption Behavior

Arthur Hardiagon and François-Xavier Coudert*

Chimie ParisTech, PSL University, CNRS, Institut de Recherche de Chimie Paris, 75005 Paris, France

E-mail: fx.coudert@chimieparistech.psl.eu

Conspectus

Nanoporous frameworks are a large and diverse family of supramolecular materials, whose chemical building units (organic, inorganic, or both) are assembled into a three-dimensional architecture with well-defined connectivity and topology, featuring intrinsic porosity. These materials play a key role in various industrial processes and applications such as energy production and conversion, fluid separation, gas storage, water harvesting, and many more. The performance and suitability of nanoporous materials for each specific application are directly related to both its physical and chemical properties, and their determination is crucial for process engineering and optimization of performances. In this Account, we focus on some recent developments in the multi-scale modeling of physical properties of nanoporous frameworks, highlighting

the latest advances in three specific areas: mechanical properties, thermal properties, and adsorption. In the study of the mechanical behavior of nanoporous materials, the last few years have seen a rapid acceleration of research. For example, computational resources have been pooled to create a public large-scale database of elastic constants as part of the Materials Project initiative to accelerate innovation in materials research: those can serve as a basis for data-based discovery of materials with targeted properties, as well as the training of machine learning predictor models. The large-scale prediction of thermal behavior, in comparison, is not yet routinely performed at such large scale. Tentative databases have been assembled at the DFT level on specific families of materials, like zeolites, but prediction at larger scale currently requires the use of transferable, classical force fields, whose accuracy can be limited. Finally, adsorption is naturally one of the most studied physical properties of nanoporous frameworks, as fluid separation or storage is often the primary target for these materials. We highlight the recent achievements and open challenges for adsorption prediction at large scale, focusing in particular on the accuracy of computational models and the reliability of comparisons with experimental data available. We detail some recent methodological improvements in the prediction of adsorption-related properties: in particular, we describe the recent research efforts to go beyond the study of thermodynamic quantities (uptake, adsorption enthalpy, thermodynamic selectivity) and predict transport properties using data-based methods and high-throughput computational schemes. Finally, we stress the importance for data-based methods of addressing all sources of uncertainty. The Account concludes with some perspectives about the latest developments and open questions in data-based approaches, and the integration of computational and experimental data together in the materials discovery loop.

Key References

- Gaillac, R.; Chibani, S.; Coudert, F.-X. Speeding Up Discovery of Auxetic Zeolite Frameworks by Machine Learning. *Chem. Mater.* **2020**, *32*, 2653–2663.¹ *A multi-scale workflow, integrating a classical force field, DFT calculations and machine learning can identify materials with anomalous mechanical properties (negative Poisson’s ratio) among a database of hypothetical zeolites.*
- Ducamp, M.; Coudert, F.-X. Prediction of Thermal Properties of Zeolites through Machine Learning. *J. Phys. Chem. C* **2022**, *126*, 1651–1660.² *A data set of thermal expansion of zeolites was created through DFT calculations in the quasi-harmonic approximation, and used to train a machine learning model.*
- Ren, E.; Coudert, F.-X. Rapid Adsorption Enthalpy Surface Sampling (RAESS) to Characterize Nanoporous Materials. *Chem. Sci.* **2023**, *14*, 1797–1807.³ *A new algorithm that quickly calculates adsorption enthalpies by sampling the surface of the material instead of the whole porous space, speeding up the high-throughput screening of materials databases for adsorption applications.*
- Ren, E.; Coudert, F.-X. Prediction of the Diffusion Coefficient through Machine Learning Based on Transition-State Theory Descriptors. *J. Phys. Chem. C* **2024**, *128*, 6917–6926.⁴ *Going beyond standard methodologies focusing on adsorption thermodynamics, an algorithm to allow high-throughput screening of guest transport by predicting self-diffusion coefficients.*

1 Introduction

Nanoporous materials are a key family of systems in the domain of materials science, with high scientific and technological importance. Nanoporous frameworks are a large and diverse family of supramolecular materials, whose chemical building units (organic, inorganic, or both) are assembled into a three-dimensional architecture with well-defined connectivity and topology, featuring intrinsic porosity. Because of the presence of pores of controllable dimensions in their structure, they present a solid/fluid interface with high surface area and tunable chemistry. Many industrial processes are based on adsorption in these nanoporous frameworks and are particular key to addressing contemporary challenges in the areas such as energy production, conversion and storage, fluid separation, environmental sciences, medicine, etc.⁵⁻⁸

In some cases, the adsorbed molecules form chemical bonds with the host material, and some applications rely on this chemisorption, such as gas sensing, chemical hydrogen storage, and probably most commonly, heterogeneous catalysis. On the other hand, many industrial applications are based on the phenomenon of physisorption, where no chemical bonds are formed and relatively weak interactions are involved (van der Waals attraction, Coulombic interactions, etc). Such is the case of applications like gas storage, gas purification, fluid separation, drug release, etc.^{6,9,10}

Molecular simulation of physisorption is a research field with a long history, and a variety of computational methods have been developed to study both the thermodynamics and dynamics of the confined phase.¹¹ However, because physisorption does not create chemical bonds between the host material and guest molecules, most of these methods rely on a “rigid host” approximation.¹² In this approach, the nanoporous material is considered as rigid and the presence of the adsorbed molecules does not change its structure or its electronic state: the adsorbent acts as an “external confinement field” on the adsorbed phase. However, this approach finds important limitations with the ever-increasing number of nanoporous frameworks exhibiting large-scale flexibility (also called “soft porous crystals”¹³), whether

due to thermal motions, or through strain induced by external stimulation.

Adsorption in these soft porous crystals can lead to increased materials performance and novel applications, including some eye-catching cases for gas storage with heat control¹⁴ and increased gas separation performance.^{15,16} It is a complex phenomenon that couples strong host-guest interactions against the flexibility of the host framework, and the resulting effect depends on a subtle balance between those.¹⁷ Therefore, in order to fully rationalize the effect of adsorption at the molecular scale, it is necessary to better understand the physical properties of the host framework itself. This includes several aspects, like the intra-framework dynamics, but also the response of the structure to thermodynamic parameters such as temperature and pressure.

This realization has led to a shift in focus from the computational chemistry community, with a stronger interest in physical properties such as mechanical and thermal behavior. There has been a growing effort in the past ten years to develop new modelling methodologies, at the microscopic scale, to provide accurate predictive models of the physical properties of nanoporous frameworks based on their structure, and to explore the structure-property relationships at play in this family of materials. This Account highlights some of the efforts in this direction, focusing specifically on three types of physical properties: mechanical behavior, thermal behavior, and adsorption properties.

2 Predicting Mechanical Behavior

2.1 Calculating mechanical properties

Various terms have been used to describe framework materials that can undergo large-scale changes in their structure, such as dynamic or flexible materials, stimuli-responsive frameworks, or soft porous crystals.¹³ The number of such materials has been growing, and include many members from supramolecular families such as the now ubiquitous metal-organic frameworks (MOFs),¹⁸ but also covalent organic frameworks (COFs)¹⁹ and supramolecular

organic frameworks (SOFs).²⁰ While the common occurrence of this flexibility induced by guest adsorption or solvent removal was noted relatively early in the development of MOFs and related materials, it was only later that researchers in the field fully realized that soft porous crystals often also exhibit a structural response to other physical or chemical stimuli.

We refer the reader to Refs. 21 and 22 for comprehensive reviews of the mechanical properties of MOFs and their applications. We note, in particular, many soft porous crystals are also “soft” in the mechanical sense of the word: they often have strong anisotropy in their elastic properties, and exhibit directions of low Young’s and shear modulus. Many framework materials display mechanical behavior not commonly found in dense crystalline materials, among which negative linear compressibility²³ and auxeticity (negative Poisson’s ratio).²⁴ While anisotropic properties can be difficult to access experimentally, their determination has become relatively routine in atomistic calculations, whether at the level of Density Functional Theory (DFT) or through molecular dynamics (MD) based on a classical force field. Both methods allow the determination of the full second-order elastic tensor, and through tensorial analysis the calculation of the various directional moduli and their polycrystalline averages.²⁵

2.2 Towards databases of elastic data

Due to the availability of the calculation of elastic tensors in quantum chemistry packages and the increase in high performance computing (HPC) resources available to researchers, several research groups started to perform systematic calculations on structural databases. Our group published in 2013 the first computational database of mechanical properties of zeolites at the DFT level, which was of modest size with 121 pure silica frameworks. In 2015, de Jong et al. calculated and published elastic information for 1,181 inorganic compounds,²⁶ which they deposited into the Materials Project database.²⁷ Since then, that elastic database in the Materials Project has continued to grow and in April 2024, it contained elastic tensors for 12,128 crystalline structures.

The use envisioned for this newly available information was twofold: the statistical study

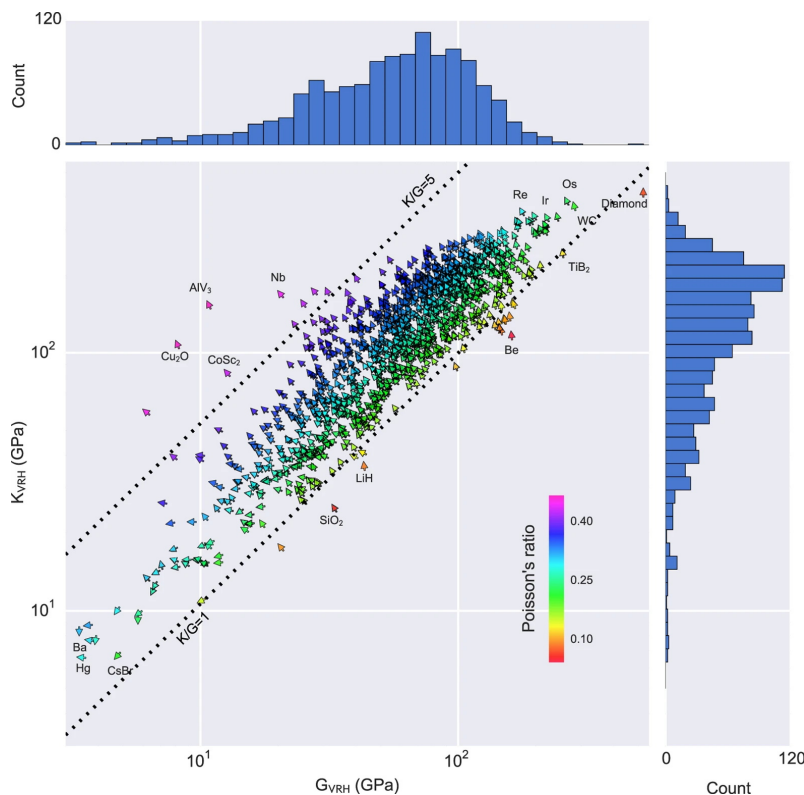


Figure 1: Distribution of bulk and shear modulus, Poisson’s ratio and atomic volume for 1,181 inorganic crystals. Arrows indicate the volume per atom: pointing at 12 o’clock corresponds to the minimum value, moving anti-clockwise in the direction of maximum value, which is located at 6 o’clock. Reproduced from Ref. 26 under the CC BY 4.0 license.

of the mechanical properties of the materials, and their use as a target property for high-throughput screening in the process of materials discovery and design. To illustrate the first point, de Jong et al.²⁶ have used the data obtained to obtain a map of the distribution of bulk moduli, putting boundaries on what values are possible and likely across the wide chemical space of inorganic materials. They have also illustrated the existence of correlations between bulk modulus, shear modulus, Poisson’s ratio and atomic volume of the crystal, as illustrated on Figure 1): such correlations serve to reinforce some classical laws of elasticity in continuous media, as well as to highlight how and how much crystalline materials (anisotropic by nature) can deviate from these ideal laws.

Moreover, the integration of the data into the Materials Project web app (available at <https://www.materialsproject.org>) allows researchers to browse through materials and

performing simple screening tasks based on several criteria, including the elastic properties — without being experts in this type of calculation. A simple interface is available, as highlighted in Figure 2, where you can select possible compositions (elements present or absent) as well as target values for various elastic moduli — among other structural and calculated properties. Once a materials is selected, its elastic properties can be visualized through the ELATE web app developed in our group²⁵ (available at <https://progs.coudert.name>), which is interlinked with the Materials Project portal. Finally, we note that the same elastic information can also be queried programmatically through an API, for integration into high-throughput screening workflows.

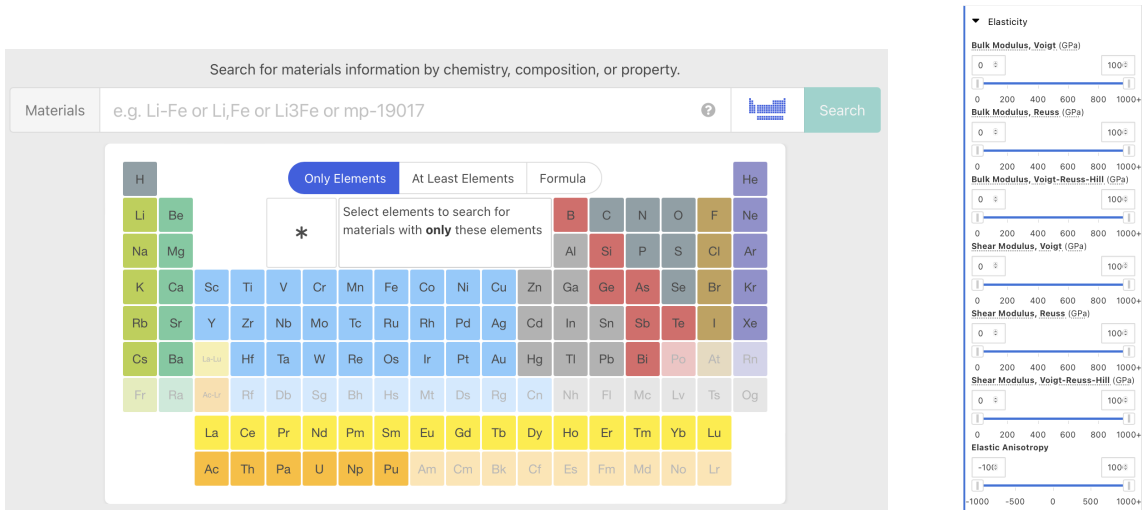


Figure 2: Top panel: Materials Project interface. Bottom panel: ELATE interface.

2.3 Identifying materials with anomalous behavior

With an increasing number of framework materials with predicted elastic constants published in the literature, it became clear that some properties considered as “rare” in crystalline materials actually have a higher prevalence than expected. For example, mechanical metamaterials²⁸ appear to be relatively common among MOFs, because of their dual organic–inorganic nature. In these meta-MOFs, the microscopic origin of the mechanical response is linked, for the most part, to the topology of the material: the framework acts like a macroscopic

truss, where the organic linkers are the rigid struts, and metal centres act as hinges²⁹ — in a very similar fashion to how macroscopic metamaterials are designed, for example by additive manufacturing techniques.³⁰

Based on large-scale databases like the one described above, it becomes possible to quantify how rare (or common) specific mechanical properties are among crystalline materials. In our group, Chibani et al. looked at 13,621 crystals from the Materials Project database and analyzed the anisotropy of their elastic properties: bulk modulus, shear modulus, Young’s modulus, Poisson’s ratio, and linear compressibility.³¹ Among other conclusions drawn from this large amount of data, we could highlight the importance of elastic anisotropy and show that mechanics play a role in the experimental feasibility of inorganic compounds: materials with higher elastic anisotropy are mechanically more fragile. It was therefore possible to quantify, for the first time in a study of that scale, exactly how frequently those properties are encountered: negative linear compressibility is found in 3% of inorganic materials, while partial auxeticity (negative Poisson’s ratio) is found in 30% — and therefore cannot be considered exceptional, as is often stated. Total auxeticity is the rarest of these phenomena, being observed in 0.3% of crystals studied, corresponding to a total of 30 crystals in the entire database.

2.4 Accelerating calculations with machine learning models

The availability of large databases of mechanical properties occurred at the same time as a growth in interest for data-based models and the increased availability of software libraries for statistical learning. In the past five years, this has led to the development and publication of machine learning models trained in this data sets obtained at the microscopic level (typically through DFT calculations). The goal there is to identify structure–property relationships and to predict mechanical behavior based on a material’s structure, sidestepping any actual atomistic modelling — i.e., without relying on physics-informed molecular simulations.

Our group originally proposed this approach on a subset of nanoporous materials with

consistent chemistry, in order to probe directly the effect of geometry and topology on mechanical properties, without influence of the chemical composition of the materials. Pure-silica zeolites offer a suitable playground for this kind of study, with a large number of known and hypothetical SiO_2 polymorphs: 237 experimentally known framework topologies, and more than 500,000 hypothetical structures.³² In a first study, we demonstrated with the work of Evans et al.³³ that it was possible to predict average (polycrystalline) mechanical moduli such as the bulk modulus or shear modulus, through a machine learning approach. We showed that a gradient boosting regressor (GBR) could be coupled to a set of *ad hoc* geometric descriptors, combining local features (Si–O distances, Si–O–Si angles, etc), global features (framework density, atomic volume, ring sizes) and porosity information (surface area, pore volume, pore size); see Figure 3. We found that the accuracy of this predictor, even when trained on a limited amount of DFT data (121 zeolites), was better than that of classical force fields for the prediction of bulk and shear moduli.

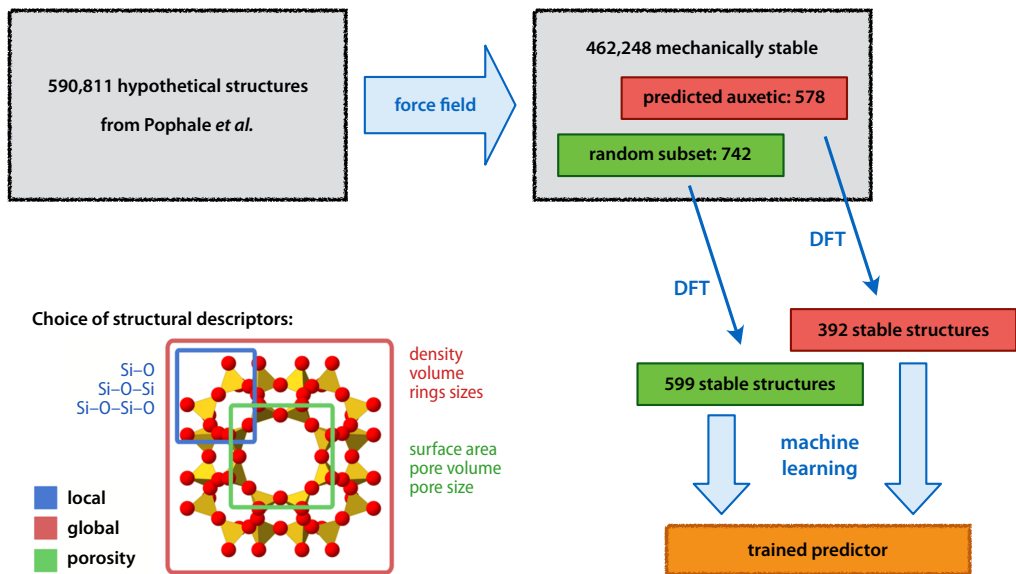


Figure 3: Description of the multi-scale modeling strategy used to discover auxetic zeolite frameworks, combining the use of a force field, DFT calculations, and machine learning. Bottom left: Summary of the structural descriptors used as entries in the machine learning algorithm, classified in local properties, global properties, and porosity-related properties. Adapted with permission from Ref. 1. Copyright 2020 American Chemical Society.

In the later work of Gaillac et al.,¹ we pushed this approach to a new level by combining

modelling at multiple levels of theory with supervised machine learning. The application in that case was the discovery of novel zeolitic structures with auxetic behavior (i.e., the presence of negative Poisson’s ratio). Because that behavior is rare, to find the proverbial needle in the haystack, it was not possible to train a ML predictor on a randomly chosen subset of zeolites, but we needed to incorporate more materials with potential auxeticity. To do so, we screened a database of nearly 600,000 hypothetical zeolitic structures at the classical level, through the use of a force field. Eliminating mechanically unstable structures, we could identify 578 frameworks predicted to be auxetic by the force field: we selected those for further study, and combined these with another 742 chosen at random. On this balanced subset of structures, we performed DFT calculations, obtaining a data set of 991 DFT-stable zeolitic frameworks. This was used as training set to optimize a machine learning predictor (again through the use of a gradient boosting regressor). This entire workflow, illustrated on Figure 3, shows a good example of the combination of molecular simulations at two different levels of theory to generate a representative data set for training a ML model. We expect such multi-scale modelling strategies to play an important role in the materials discovery pipeline, combined with experimental feedback and the use of optimal experimental design, or approaches such as multi-fidelity Bayesian optimization.³⁴

More recently, some authors have proposed to use deep learning methods in order to predict elastic properties across the entire chemical space of crystals, a harder problem because it needs to take into account both geometrical and chemical considerations. Some authors have used this approach to predict scalar quantities, such as Mazhnik et al.³⁵ who targeted hardness and fracture toughness in a study in order to identify new superhard materials. Similarly, Tsuruta et al.³⁶ used a representation of crystal structures as 3-dimensional tetrahedral meshes to predict average (isotropic) bulk and shear moduli. Finally, the last months have seen the application of equivariant graph neural networks, which are independent of the frame of reference and preserve material symmetry and therefore allow for the prediction of a full elastic tensor, going beyond average properties.^{37,38} All these approaches, while relying

on the power of neural networks, differ in their encoding and representation of the materials' structure. Such tensorial approaches are powerful, but require training data sets of very large size (in this case, more than 10,000 elastic tensors): it remains to be seen how accurate they would be for other tensorial materials properties, where such large data sets are not currently available.

3 Predicting Thermal Behavior

Like their mechanical properties, the thermal behavior of materials is a key piece of information for their practical use in industrial applications. Quantities like thermal expansion, heat capacity, thermal conductivity and thermal diffusivity are routinely used in chemical engineering for process design, simulation and optimization. It is therefore important to be able to predict the thermal behavior of framework materials by computational chemistry and to study the structure/property relationships in order to find optimal materials (or design new structures) for applications.

Our group has looked early on at the diversity of thermal expansion among MOFs, and in particular at uncoupling the impacts of reticular topology and framework chemistry. Bouëssel du Bourg et al. used molecular dynamics simulations combined with classical force fields to show that, for a similar chemistry (zinc-based zeolitic imidazolate frameworks, or ZIFs), the thermal expansion could vary both in sign and magnitude with the topology of the materials.³⁹ We could also show that some ZIFs exhibit very large negative thermal expansion (NTE), i.e., they contract upon heating. That phenomenon is counter-intuitive in dense materials, but actually quite common in framework materials such as some oxides,⁴⁰ zeolites, MOFs, etc.⁴¹ The identification of materials with tunable thermal expansion is an important axis of research, in particular in the design of novel nanostructure composites.

It is important to note that, compared to the number of computational studies focusing on mechanical properties, thermal behavior is still a much more open question. The root

cause for this is to be found in the greater computational expense of the methodology. Thermal properties are typically calculated through long molecular dynamics (MD) simulations: those require the use of a classical force field, accurately describing the full flexibility of the framework. Such calculations are therefore difficult to perform at large scale, as “universal” or “transferable” force fields have limited accuracy in their description of intra- and inter-molecular interactions. In another approach, thermal properties such as heat capacity and thermal expansion can be approximated through the determination of the vibrational modes of the framework (or phonons). Such harmonic or quasi-harmonic calculations can be performed at the DFT level, but are very expensive for large frameworks: they computational cost scales with the number of atoms.^{42,43}

In order to push in this direction, our group published in Ducamp et al.⁴⁴ the first systematic study of the thermal properties of 134 all-silica zeolites (SiO_2 polymorphs) based on DFT calculations in the quasi-harmonic approximation — i.e., in the limit of low temperature. We showed that zeolitic frameworks feature a wide range of thermal expansion coefficients, highlighting the great influence of the framework topology over this physical property. This approach also allowed us to compute the bulk modulus for each structure, but also its pressure and temperature dependence, something that was not accessible in previous linear response calculations.⁴⁵ We confirmed that zeolitic frameworks can display pressure-induced softening, something which had been seen in a few cases experimentally, and could quantify this behavior, showing it is in fact the case for most zeolites. Moreover, this statistical exploration of the thermo-mechanical behavior of zeolites gave some hints at the challenging question of experimental feasibility of zeolitic frameworks, identifying that experimentally synthesized structures correspond to a restricted range of the bulk modulus, K_0 , and its first derivative, K'_0 .

In later work, we used this newly created database of thermal expansion coefficients at the DFT level to train machine learning models, in a workflow summarized on Figure 4. Our main goal was to compare the statistical accuracy of trained models, all based on the same

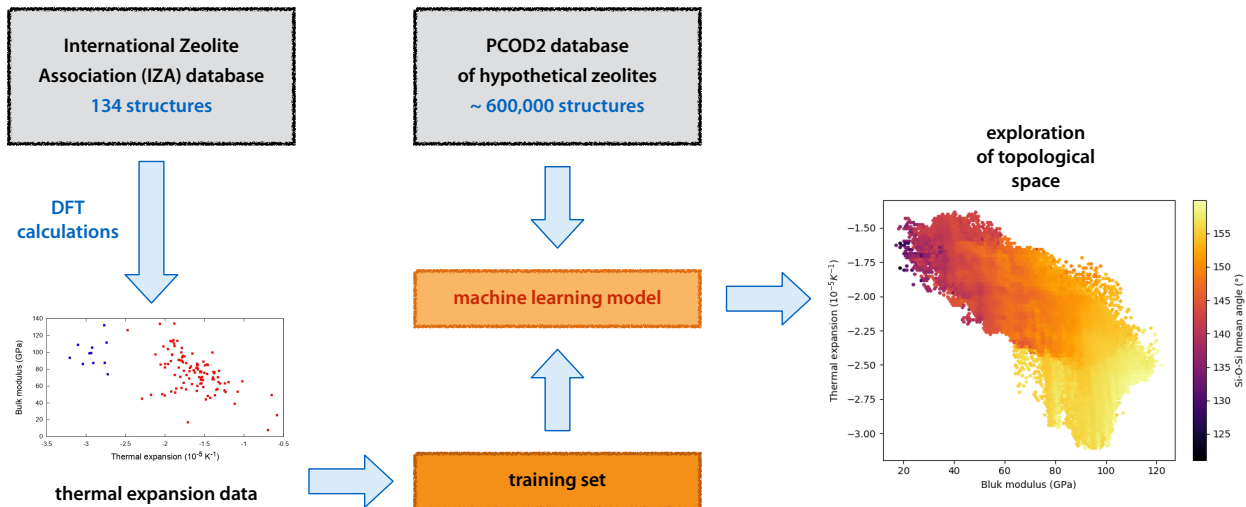


Figure 4: Workflow used to generate thermal expansion data of zeolitic frameworks through DFT calculations, train a machine learning model based on that data, and apply it to a database of hypothetical structures to explore topological space. Adapted with permission from Ref. 2. Copyright 2022 American Chemical Society.

gradient-boosting regressor algorithm, depending on the type of descriptors used to encode the geometrical features of the zeolites.² In particular, we compared *ad hoc* geometrical features, topology, pore space, to the performance of general geometric descriptors based on smooth overlap of atomic position (SOAP), which are not specific to zeolites. We showed that generic geometric descriptors performed best, and demonstrated that the predictions were not very sensitive to small changes in the input structure, i.e., that the outcome of the prediction did not necessarily require the refinement of framework structures at a high level of theory. This finding is very important for the generalizability of the technique, showing that the ML predictor obtained can be used on lower-quality structures. We then proceeded to apply the ML predictor to the PCOD2 database of zeolites containing around 600,000 hypothetical structures. This large-scale screening study confirmed a general statistical correlation between the thermal expansion coefficient and bulk modulus of the materials, with the Si–O–Si angles playing a key role in both.

Other groups have experimented different data-based approaches for the determination of thermal behavior of MOFs. Nandy et al. have used the natural language processing to

mine information about solvent removal and thermal stability of MOFs from the published literature,⁴⁶ coming up with a data set of 3,000 MOFs and associated thermal decomposition temperatures. In 2023, Islamov et al. used a computational high-throughput screening setup to study the thermal conductivity of 10,194 hypothetical MOFs created using the ToBaCCo software.⁴⁷ While this represents a systematic study over a large set of structures, to work at this scale the authors had to rely on a very generic force field (UFF4MOF), which limits the accuracy of their results obtained.

In a different approach, Moosavi et al. devised a computational workflow for the prediction of heat capacities of nanoporous frameworks spanning different families: zeolites, MOFs and COFs. Their approach is based on the calculation of heat capacities through DFT calculations for 230 structures with diverse chemical environments, used for training and validating machine learning models.⁴⁸ Their ML models are based on a sum of atomic contributions, where each atom in the framework is featurized in terms of atomic identity, local geometry and chemistry. Their final model demonstrated a very good accuracy (ca. 3% relative error) against DFT data, and in most cases a good agreement with available experimental data. This approach is very promising for the study of other thermal properties, focusing on a data set that is limited in size, but has good diversity and high accuracy.

4 Predicting Adsorption

Adsorption is another important physical property of nanoporous materials. Based on adsorption data, one can rank materials for a targeted application (such as gas storage, gas separation, catalysis, etc.).⁴⁹ However, while prediction of fluid adsorption has been a mainstream computational technique for decades, adsorption predictions suffer from two main limitations: the accuracy of the models used in computational screening and the reliability of the comparison with experimental data.

4.1 Limitations of experimental data

There is a lot of experimental adsorption data available in the published literature, but it is hard to mine. There have been important efforts to digitize data from experimental adsorption measurements, in particular adsorption isotherms, like in the ISODB/NIST online database.⁵⁰ In that case, all data have been digitized manually over the last decade but the database can be fed by users via data standardization.⁵¹ Isotherm data can be freely and easily visualized on the database website, or through an API which facilitates a more systematic access. However, despite this effort there are still a number of difficulties to overcome, such as uncertainties in the digitization process, lack of systematic metadata, and the identification of materials and samples.⁵²

Previous studies about reproducibility in the field of equilibrium adsorption indicated that single-component adsorption uptakes have been fully validated through consensus only for a few well-studied materials.⁵³ In most cases, different research groups use different experimental protocols, such that synthesized materials published under the same name and/or composition often behave differently in terms of adsorption characterization. Efforts are needed to publish the metadata associated with preparation conditions, handling, and activation — as well as standardization in the way it is published, for example by using the recently proposed AIF format,⁵⁴ a project supported by the International Union of Pure and Applied Chemistry (IUPAC).

4.2 Grand Canonical Monte Carlo (GCMC) simulations

Furthermore, data on adsorption in porous materials is sparse in the literature, as some materials have been specifically studied with a limited the number of gases adsorbed. This is why computational methods have long been used in order to supplement experimental databases — either through systematic studies, or more recently by using a machine learning-based recommendation system.⁵⁵ To this end, Grand Canonical Monte Carlo (GCMC) simulations have been used extensively to predict fluid uptake and gas separation. Within a limited

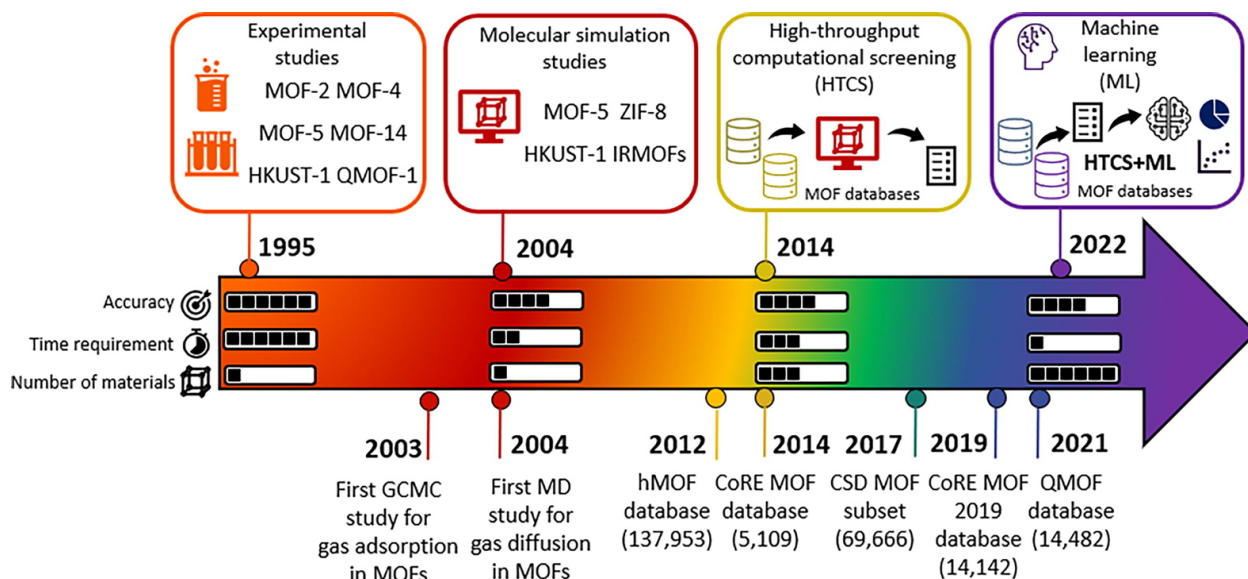


Figure 5: Timeline of major milestones in computational MOF research, highlighting the development of materials databases (and the number of materials in each database), as well as the recent shift to high-throughput screening studies and data-based statistical learning methods for prediction of properties. Reprinted from Ref. 58, Copyright 2023, with permission from Elsevier.

set of assumptions (rigid adsorbent, accurate classical force field, no open metal sites), these predictions are relatively accurate and computationally cheap, providing absolute gas uptake values in the case of single adsorption, or co-adsorption uptakes at a fixed temperature and mixture composition. Therefore GCMC data have been used in many screening studies on MOFs,^{56,57} and are particularly useful for co-adsorption characterization, where experiments are particularly difficult and the dimensionality of the thermodynamic space is higher.

Beyond the identification of top-performing materials, two goals can be achieved with those studies: (i) studying the boundaries of possible performance across the chemical space, and getting insight into the relationships between physical or chemical properties; (ii) understanding structure/property relationships in order to train faster models for prediction. Gas uptakes, thermodynamic properties like Henry coefficients and heat of adsorption have been extensively used to train machine learning model that can infer structure/properties relationships^{58,59} (see Figure 5).

Recently, our group proposed a methodology for the identification of separation selectivity

at different operating conditions (temperature and pressure) on a large set of nanoporous structures. To avoid the necessity for systematic GCMC calculations, at least in a first phase of the screening process, we developed a screening process (illustrated in Figure 6) based a machine learning predictor of selectivity, which is fed by both geometrical descriptors as well as energy-relation properties, obtained by a fast grid calculation of host–guest interaction energies. As a proof of concept, we tested our methodology for the separation of a 20:80 Xe/Kr mixture at 298 K and ambient pressure, screening through the CoRE MOF 2019 database of structures.⁶⁰ The resulting workflow can evaluate the selectivity of a single structure in about one minute, while retaining good accuracy, whereas a GCMC calculations for the same task took on average 40 minutes.

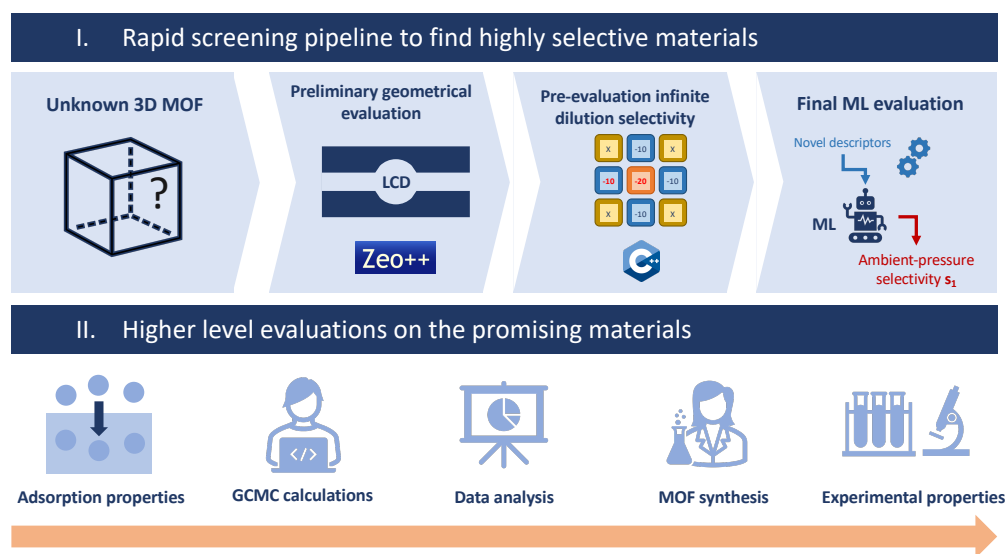


Figure 6: Schematic representation of the screening procedure used in Ren et al.⁵⁹ to identify highly selective materials for adsorption-based separation selectivity at different operating conditions of temperature and pressure, on a large set of nanoporous structures. Reproduced with permission from Ref. 59. Copyright 2023 American Chemical Society.

4.3 Current limitations

However, there are a number of limitations introduced by GCMC simulations that the reader should bear in mind. Ranking materials for a given application in a database based on GCMC

data assumes that host–guest interactions are equivalently described with the same accuracy for all host atoms. However this is rarely the case, in particular for polar gas molecules and/or open metal sites on frameworks. Most studies rely on “general” or “universal” classical forcefield, such as UFF, Dreiding and TraPPE, to model van der Waals interactions between the atoms of the frameworks and the atoms of the gas. Electrostatic interactions are then calculated from partial atomic charges derived using semi-empirical methods, ML methods, or DFT calculations on large set of structures. Ab initio FFs might be necessary to improve the quality of generic FFs:⁶¹ in particular, ab initio FFs have been used to describe accurately the adsorption uptakes of polar molecules in frameworks with open-sites metal.⁶² They are, however, computationally very expensive.

The effect of the host structure’s geometry has been less studied, though it has a significant effect on the thermodynamic properties computed in GCMC simulations.⁶³ To overcome these limitations, one generally proceeds in two steps: first, screening all structures of a database for a targeted properties using generic FFs and clean experimental structures; secondly, refining the ranking after FFs and geometry optimization on the best-performing materials. Another approach is to rank materials in a screening study for a given application and varying the FFs, then assigning uncertainty on the outcomes related to the underlying model.⁶⁴

Since numerous screening studies rely on shared assumptions and database structures, the community would benefit from disseminating reference data derived from simulations.^{65,66} This would improve computational reproducibility and facilitate the evaluation of accuracy for training datasets, particularly in the context of recent machine learning models that leverage GCMC results for fast predictions of MOF properties.

4.4 Beyond thermodynamics

Finally, let us note that most large-scale studies on adsorption have focused on thermodynamic equilibrium and avoided considerations of kinetics, such as transport properties inside

the nanopores. Yet these properties are key parameters in practical applications. Brute-force Molecular Dynamics simulations allow to compute diffusion coefficients, yet it is computationally expensive compared to GCMC simulations, explaining why only a few studies apply this methodology to large databases.⁶⁷ Another approach to compute diffusion coefficients uses Transition State Theory and kinetic Monte-Carlo simulations and allows to decrease considerably the computational cost.⁶⁸ Recently, our group computed diffusion coefficients of xenon in a large subset of the CoRE MOF structural database⁶⁰ and showed that diffusion coefficients can be efficiently predicted using ML models with fast calculations of activation energies and other geometrical descriptors.⁴ These studies pave the way to the extension of porous materials databases with kinetic descriptors, and we think this area of research has not seen yet its full potential, and many methodological developments can be expected in the near future.

5 Perspectives

Over the last 20 years, there has been a notable acceleration in the discovery of porous structures. This growth is illustrated in the timeline of Figure 5. The view of our field, once focused on experimental characterization of a select group of few commonly available materials, has since been strengthened by the use of simulation techniques for property prediction based on atomic structures, high-throughput computational screening studies, and the recent advances in artificial intelligence (AI) for faster materials discovery. Since it is standard (in for most journals even mandated) for new crystalline structures to be published in a database, such as the CSD (Crystallography Structural Database), we now have access a large set of porous materials, including > 70,000 crystalline MOFs.⁶⁹ The large scale of such data sets is a novel dimension in the context of materials discovery: it can help accelerate the identification of existing materials for targeted applications through data-driven approaches, and serve as a basis for the prediction of novel hypothetical structures.

The first requirement to perform computational materials discovery is to have access to open databases with cleaned and computation-ready structures, as well as associated parameters.⁶⁰ Given the size of the databases (typically ten of thousands of structures), screening studies are computationally expensive, so two approaches are favored. In the first approach, efforts are made to reduce the number of structures in the database while increasing the diversity of structures (from a structural perspective, or based on the underlying chemistry or reagents used in synthesis).^{70,71} An alternative method is to perform multi-stage screening, with a first filter with low computational cost to calculate properties that encode the most pertinent information for the application chosen.³ Screening studies often rely on assumptions in material models to accommodate a wide range of diverse structures, which tends to decrease precision in the measured properties. Consequently, research groups advocate for providing datasets with high levels of precision, such as DDEC charge sets for adsorption calculations with GCMC.^{72,73}

Despite their promising opportunity to accelerate the process of discovering new materials based on structural features, the data-driven approaches still suffer from a number of shortcomings, including the availability of structural data, its accuracy and synthesizability of the hypothetical structures. Screening studies often currently neglect the synthesis data and promote materials that might not be relevant for designing real materials. With the advent of text-based information processing methods, it is however becoming possible to automatically review a large corpus of scientific articles to extract synthetic properties from the published literature (see Figure 7).⁷⁴⁻⁷⁶ In the near future, the aim of these directed approaches would be to propose synthetic routes to experimental chemists, given a limited number of constraints such as a set of reagents, the material available and the desired properties.

Currently, the growing amount of structural and simulation data is generating a lot of enthusiasm for creating machine learning models capable of predicting the properties of new materials. Over the last few years, a series of classes of ML models (simple ML models for

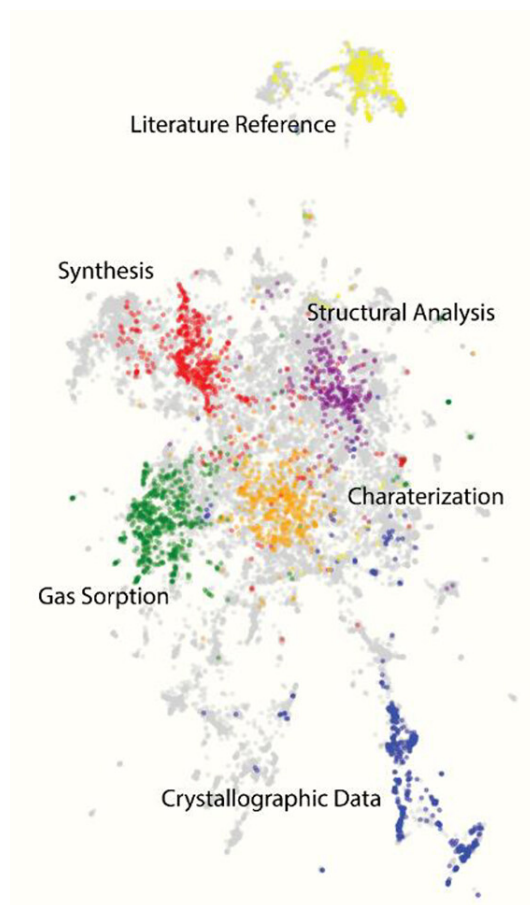


Figure 7: A 2D representation of labeled data extracted from mining literature on MOFs. Large Language Models (LLMs) allow to classify text sequences and extract relevant information about synthesis conditions, characterization results, structural data and sorption data. These data can be used to create databases with heterogeneous data for accurate predictions related to synthesis conditions. Reproduced with permission from Ref. 74. Copyright 2023 American Chemical Society.

regression, convolutional neural networks, graph neural networks and transformers for more complex data) have emerged, demonstrating their growing effectiveness in predicting properties such as adsorption uptakes in MOFs,^{77–83} see Figure 8. Authors typically concentrate on assessing their model’s performance using various metrics on a validation dataset. However, comparing models becomes challenging when they are trained and validated on different datasets. Establishing reference datasets, a common practice in other artificial intelligence fields, will likely be a key challenge in the coming years. To this purpose, researchers that focus on developing new predictive models should use available datasets, namely using cur-

rent online portals that emphasize FAIR practices.^{65,84} Structural datasets evolve over time due to new experimental discoveries, improved cleaning procedures, and the discarding of outliers. Any study aiming to enhance prediction performance in machine learning models should therefore publish their training, test, and validation datasets, along with metadata and lineage to previously published datasets.⁸⁵

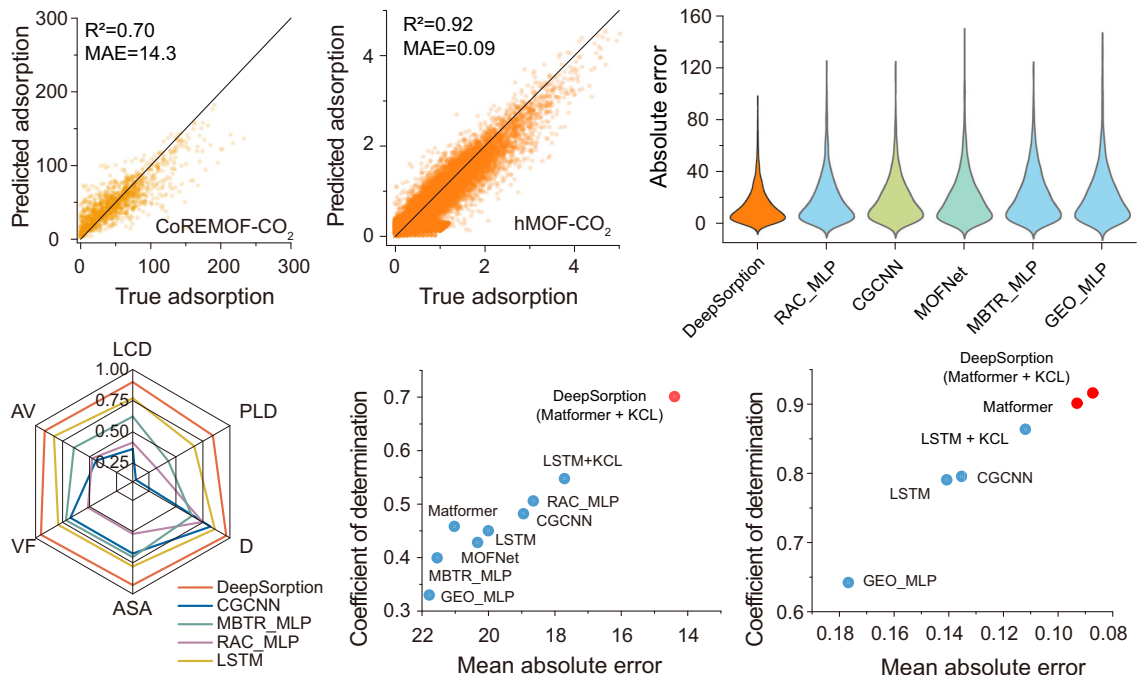


Figure 8: Performance of CO₂ adsorption uptake predictions on CoRE MOF database using various deep learning/ML models. Reproduced from Ref. 83 under the CC BY 4.0 license.

Although the accuracy of machine learning models might keep increasing in the coming years, the quality of training data intrinsically limits the accuracy of models for applications. Indeed, most ML studies assume that materials properties are fully encoded in their crystalline atomic-scale structures. However, porous materials experimentally have physical and chemical properties that depend on other factors, such as synthesis conditions, purification steps and material formulation, textural properties, presence of defects, etc.^{86–89} Model validation steps need to be carried out directly on high-quality experimental data sets, which can be guaranteed by the use of FAIR data processing tools and standardized best practices.⁹⁰ Some studies have already established prediction-guided protocols based on powerful

language models using mined synthesis data.⁹¹ However, all data can not be mined with the same degree of accuracy when metadata are missing or non-standardized measurements are used.^{52,92,93} As an alternative, porous materials discovery projects could combine both experimental and modeling approaches to build up consistent datasets. Automation of the synthesis and characterization steps with active learning feedback and including negative results will open up new perspectives in this direction, however it is still in the very early stages of its development.⁹⁴⁻⁹⁶

Author Information

The authors declare no competing financial interest.

Biography

Arthur Hardiagon is a post-doctoral researcher at the French Centre for Scientific Research (CNRS). He currently works on metal-organic frameworks (MOF) using multiscale simulations and data science approaches. He received his PhD in physical chemistry from the Université Paris Cité (France) in 2022 for his work on the molecular dynamics of artificial water channels in lipid membranes.

François-Xavier Coudert is a senior researcher at the French National Centre for Scientific Research (CNRS) and a professor at PSL University/École normale supérieure. His group applies computational chemistry methods at various scales to investigate the physical and chemical properties of nanoporous materials, their responses to external stimuli, and the behavior of fluids at fluid/solid interfaces. He obtained his PhD from the Université Paris-Sud (France) in 2007 for work on the properties of water and solvated electrons confined in zeolite nanopores, and worked as a postdoctoral researcher at University College London (UK) on the growth of metal-organic frameworks on surfaces.

Acknowledgement

This work was funded under the France 2030 framework by Agence Nationale de la Recherche (project ANR-22-PEXD-0009 DIADEM). Although no original calculations were performed in the writing of this paper, our group’s work on this topic benefits from regular access to high-performance computing platforms provided by GENCI under grant A0150807069.

References

- 1 Gaillac, R.; Chibani, S.; Coudert, F.-X. Speeding Up Discovery of Auxetic Zeolite Frameworks by Machine Learning. *Chem. Mater.* **2020**, *32*, 2653–2663.
- 2 Ducamp, M.; Coudert, F.-X. Prediction of Thermal Properties of Zeolites through Machine Learning. *J. Phys. Chem. C* **2022**, *126*, 1651–1660.
- 3 Ren, E.; Coudert, F.-X. Rapid adsorption enthalpy surface sampling (RAESS) to characterize nanoporous materials. *Chem. Sci.* **2023**, *14*, 1797–1807.
- 4 Ren, E.; Coudert, F.-X. Prediction of the Diffusion Coefficient through Machine Learning Based on Transition-State Theory Descriptors. *J. Phys. Chem. C* **2024**,
- 5 Bell, A. T. The Impact of Nanoscience on Heterogeneous Catalysis. *Science* **2003**, *299*, 1688–1691.
- 6 Sholl, D. S.; Lively, R. P. Seven chemical separations to change the world. *Nature* **2016**, *532*, 435–437.
- 7 Kerry, F. G. *Industrial Gas Handbook: Gas Separation and Purification*; CRC Press, 2007.
- 8 National Academies of Sciences, Engineering, and Medicine, *A Research Agenda for Transforming Separation Science*, National Academies Press, 2019.

- 9 Li, J.-R.; Kuppler, R. J.; Zhou, H.-C. Selective gas adsorption and separation in metal-organic frameworks. *Chem. Soc. Rev.* **2009**, *38*, 1477.
- 10 Ding, M.; Flaig, R. W.; Jiang, H.-L.; Yaghi, O. M. Carbon capture and conversion using metal-organic frameworks and MOF-based materials. *Chem. Soc. Rev.* **2019**, *48*, 2783–2828.
- 11 Jelfs, K. E. *Inorganic Materials Series*; Royal Society of Chemistry, 2021; Chapter Introduction to Computational Modelling of Microporous Materials, pp 1–26.
- 12 Fuchs, A. H.; Cheetham, A. K. Adsorption of Guest Molecules in Zeolitic Materials: Computational Aspects. *J. Phys. Chem. B* **2001**, *105*, 7375–7383.
- 13 Horike, S.; Shimomura, S.; Kitagawa, S. Soft porous crystals. *Nature Chem.* **2009**, *1*, 695–704.
- 14 Mason, J. A.; Oktawiec, J.; Taylor, M. K.; Hudson, M. R.; Rodriguez, J.; Bachman, J. E.; Gonzalez, M. I.; Cervellino, A.; Guagliardi, A.; Brown, C. M.; Llewellyn, P. L.; Masciocchi, N.; Long, J. R. Methane storage in flexible metal–organic frameworks with intrinsic thermal management. *Nature* **2015**, *527*, 357–361.
- 15 Hiraide, S.; Sakanaka, Y.; Kajiro, H.; Kawaguchi, S.; Miyahara, M. T.; Tanaka, H. High-throughput gas separation by flexible metal–organic frameworks with fast gating and thermal management capabilities. *Nature Commun.* **2020**, *11*, 435.
- 16 Bonneau, M.; Lavenn, C.; Zheng, J.-J.; Legrand, A.; Ogawa, T.; Sugimoto, K.; Couderc, F.-X.; Reau, R.; Sakaki, S.; Otake, K.-i.; Kitagawa, S. Tunable acetylene sorption by flexible catenated metal–organic frameworks. *Nature Chem.* **2022**, *14*, 816–822.
- 17 Schneemann, A.; Bon, V.; Schwedler, I.; Senkowska, I.; Kaskel, S.; Fischer, R. A. Flexible metal–organic frameworks. *Chem. Soc. Rev.* **2014**, *43*, 6062–6096.

- 18 Furukawa, H.; Cordova, K. E.; O’Keeffe, M.; Yaghi, O. M. The Chemistry and Applications of Metal-Organic Frameworks. *Science* **2013**, *341*, 673.
- 19 Feng, X.; Ding, X.; Jiang, D. Covalent organic frameworks. *Chem. Soc. Rev.* **2012**, *41*, 6010.
- 20 Pedrini, A.; Marchetti, D.; Pinalli, R.; Massera, C. Stimuli-Responsive, Dynamic Supramolecular Organic Frameworks. *ChemPlusChem* **2023**, *88*.
- 21 Burtch, N. C.; Heinen, J.; Bennett, T. D.; Dubbeldam, D.; Allendorf, M. D. Mechanical Properties in Metal–Organic Frameworks: Emerging Opportunities and Challenges for Device Functionality and Technological Applications. *Adv. Mater.* **2018**, *30*, 737.
- 22 Redfern, L. R.; Farha, O. K. Mechanical properties of metal–organic frameworks. *Chem. Sci.* **2019**, *10*, 10666–10679.
- 23 Cairns, A. B.; Goodwin, A. L. Negative linear compressibility. *Phys. Chem. Chem. Phys.* **2015**, *17*, 20449–20465.
- 24 Huang, C.; Chen, L. Negative Poisson’s Ratio in Modern Functional Materials. *Adv. Mater.* **2016**, *28*, 8079–8096.
- 25 Gaillac, R.; Pullumbi, P.; Coudert, F.-X. ELATE: an open-source online application for analysis and visualization of elastic tensors. *J. Phys.: Condens. Matter* **2016**, *28*, 275201.
- 26 de Jong, M.; Chen, W.; Angsten, T.; Jain, A.; Notestine, R.; Gamst, A.; Sluiter, M.; Krishna Ande, C.; van der Zwaag, S.; Plata, J. J.; Toher, C.; Curtarolo, S.; Ceder, G.; Persson, K. A.; Asta, M. Charting the complete elastic properties of inorganic crystalline compounds. *Sci Data* **2015**, *2*, 345.
- 27 Jain, A.; Ong, S. P.; Hautier, G.; Chen, W.; Richards, W. D.; Dacek, S.; Cholia, S.; Gunter, D.; Skinner, D.; Ceder, G.; Persson, K. A. Commentary: The Materials Project:

- A materials genome approach to accelerating materials innovation. *APL Mater.* **2013**, *1*, 1049.
- 28 Mechanical metamaterials are artificial materials, engineered through the design of their microstructure to display novel mechanical properties not typically found in nature.
- 29 Coudert, F.-X.; Evans, J. D. Nanoscale metamaterials: Meta-MOFs and framework materials with anomalous behavior. *Coord. Chem. Rev.* **2019**, *388*, 48–62.
- 30 Zadpoor, A. A. Mechanical meta-materials. *Mater. Horiz.* **2016**, *3*, 371–381.
- 31 Chibani, S.; Coudert, F.-X. Systematic exploration of the mechanical properties of 13 621 inorganic compounds. *Chem. Sci.* **2019**, *10*, 8589–8599.
- 32 Pophale, R.; Cheeseman, P. A.; Deem, M. W. A database of new zeolite-like materials. *Phys. Chem. Chem. Phys.* **2011**, *13*, 12407.
- 33 Evans, J. D.; Coudert, F.-X. Predicting the Mechanical Properties of Zeolite Frameworks by Machine Learning. *Chem. Mater.* **2017**, *29*, 7833–7839.
- 34 Gantzler, N.; Deshwal, A.; Doppa, J. R.; Simon, C. M. Multi-fidelity Bayesian optimization of covalent organic frameworks for xenon/krypton separations. *Digital Discovery* **2023**, *2*, 1937–1956.
- 35 Mazhnik, E.; Oganov, A. R. Application of machine learning methods for predicting new superhard materials. *J. Appl. Phys.* **2020**, *128*, 040901.
- 36 Tsuruta, H.; Katsura, Y.; Kumagai, M. DeepCrysTet: A Deep Learning Approach Using Tetrahedral Mesh for Predicting Properties of Crystalline Materials. *arXiv* **2023**, 2310.06852.
- 37 Pakornchote, T.; Ektarawong, A.; Chotibut, T. StrainTensorNet: Predicting crystal structure elastic properties using SE(3)-equivariant graph neural networks. *arXiv* **2023**, 2306.12818.

- 38 Wen, M.; Horton, M. K.; Munro, J. M.; Huck, P.; Persson, K. A. An equivariant graph neural network for the elasticity tensors of all seven crystal systems. *arXiv* **2024**, 2307.15242.
- 39 Bouëssel du Bourg, L.; Ortiz, A. U.; Boutin, A.; Coudert, F.-X. Thermal and mechanical stability of zeolitic imidazolate frameworks polymorphs. *APL Mater.* **2014**, *2*, 10186.
- 40 Baise, M.; Maffettone, P. M.; Trouselet, F.; Funnell, N. P.; Coudert, F.-X.; Goodwin, A. L. Negative Hydration Expansion in ZrW_2O_8 : Microscopic Mechanism, Spaghetti Dynamics, and Negative Thermal Expansion. *Phys. Rev. Lett.* **2018**, *120*.
- 41 Shi, N.; Song, Y.; Xing, X.; Chen, J. Negative thermal expansion in framework structure materials. *Coord. Chem. Rev.* **2021**, *449*, 214204.
- 42 Bristow, J. K.; Skelton, J. M.; Svane, K. L.; Walsh, A.; Gale, J. D. A general forcefield for accurate phonon properties of metal-organic frameworks. *Phys. Chem. Chem. Phys.* **2016**, *18*, 29316–29329.
- 43 Ryder, M. R.; Maul, J.; Civalleri, B.; Erba, A. Quasi-Harmonic Lattice Dynamics of a Prototypical Metal-Organic Framework. *Adv. Theory Simul.* **2019**, *2*.
- 44 Ducamp, M.; Coudert, F.-X. Systematic Study of the Thermal Properties of Zeolitic Frameworks. *J. Phys. Chem. C* **2021**, *125*, 15647–15658.
- 45 Coudert, F.-X. Systematic investigation of the mechanical properties of pure silica zeolites: stiffness, anisotropy, and negative linear compressibility. *Phys. Chem. Chem. Phys.* **2013**, *15*, 16012.
- 46 Nandy, A.; Terrones, G.; Arunachalam, N.; Duan, C.; Kastner, D. W.; Kulik, H. J. MOFSimplify, machine learning models with extracted stability data of three thousand metal–organic frameworks. *Sci Data* **2022**, *9*, 1230444.

- 47 Islamov, M.; Babaei, H.; Anderson, R.; Sezginel, K. B.; Long, J. R.; McGaughey, A. J. H.; Gomez-Gualdron, D. A.; Wilmer, C. E. High-throughput screening of hypothetical metal-organic frameworks for thermal conductivity. *npj Comput. Mater.* **2023**, *9*, 2334.
- 48 Moosavi, S. M.; Novotny, B. A.; Ongari, D.; Moubarak, E.; Asgari, M.; Kadioglu, O.; Charalambous, C.; Ortega-Guerrero, A.; Farmahini, A. H.; Sarkisov, L.; Garcia, S.; Noé, F.; Smit, B. A data-science approach to predict the heat capacity of nanoporous materials. *Nature Mater.* **2022**, *21*, 1419–1425.
- 49 Altintas, C.; Avci, G.; Daglar, H.; Nemati Vesali Azar, A.; Velioglu, S.; Erucar, I.; Keskink, S. Database for CO₂ Separation Performances of MOFs Based on Computational Materials Screening. *ACS Appl. Mater. Interfaces* **2018**, *10*, 17257–17268.
- 50 Siderius, D. NIST/ARPA-E Database of Novel and Emerging Adsorbent Materials. 2020; <https://data.nist.gov/od/id/FF429BC1787D8B3EE0431A570681E858236>, last accessed 26 April 2024.
- 51 NIST-ISODB/isodbtools, software repository available online at <https://github.com/NIST-ISODB/isodbtools>, last accessed 26 April 2024.
- 52 Ongari, D.; Talirz, L.; Jablonka, K. M.; Siderius, D. W.; Smit, B. Data-Driven Matching of Experimental Crystal Structures and Gas Adsorption Isotherms of Metal-Organic Frameworks. *J. Chem. Eng. Data* **2022**, *67*, 1743–1756.
- 53 Park, J.; Howe, J. D.; Sholl, D. S. How Reproducible Are Isotherm Measurements in Metal-Organic Frameworks? *Chem. Mater.* **2017**, *29*, 10487–10495.
- 54 Evans, J. D.; Bon, V.; Senkovska, I.; Kaskel, S. A Universal Standard Archive File for Adsorption Data. *Langmuir* **2021**, *37*, 4222–4226.
- 55 Sturluson, A.; Raza, A.; McConachie, G. D.; Siderius, D. W.; Fern, X. Z.; Simon, C. M.

- Recommendation System to Predict Missing Adsorption Properties of Nanoporous Materials. *Chem. Mater.* **2021**, *33*, 7203–7216.
- 56 Erucar, I.; Keskin, S. High-Throughput Molecular Simulations of Metal Organic Frameworks for CO₂ Separation: Opportunities and Challenges. *Front. Mater.* **2018**, *5*.
- 57 Ren, E.; Guilbaud, P.; Coudert, F.-X. High-throughput computational screening of nanoporous materials in targeted applications. *Digital Discovery* **2022**, *1*, 355–374.
- 58 Demir, H.; Daglar, H.; Gulbalkan, H. C.; Aksu, G. O.; Keskin, S. Recent advances in computational modeling of MOFs: From molecular simulations to machine learning. *Coord. Chem. Rev.* **2023**, *484*, 215112.
- 59 Ren, E.; Coudert, F.-X. Enhancing Gas Separation Selectivity Prediction through Geometrical and Chemical Descriptors. *Chem. Mater.* **2023**, *35*, 6771–6781.
- 60 Chung, Y. G.; Haldoupis, E.; Bucior, B. J.; Haranczyk, M.; Lee, S.; Zhang, H.; Vogiatzis, K. D.; Milisavljevic, M.; Ling, S.; Camp, J. S.; Slater, B.; Siepmann, J. I.; Sholl, D. S.; Snurr, R. Q. Advances, Updates, and Analytics for the Computation-Ready, Experimental Metal-Organic Framework Database: CoRE MOF 2019. *J. Chem. Eng. Data* **2019**, *64*, 5985–5998.
- 61 McDaniel, J. G.; Li, S.; Tylianakis, E.; Snurr, R. Q.; Schmidt, J. R. Evaluation of Force Field Performance for High-Throughput Screening of Gas Uptake in Metal-Organic Frameworks. *J. Phys. Chem. C* **2015**, *119*, 3143–3152.
- 62 Lin, L.-C.; Lee, K.; Gagliardi, L.; Neaton, J. B.; Smit, B. Force-Field Development from Electronic Structure Calculations with Periodic Boundary Conditions: Applications to Gaseous Adsorption and Transport in Metal-Organic Frameworks. *J. Chem. Theory Comput.* **2014**, *10*, 1477–1488.

- 63 Nazarian, D.; Camp, J. S.; Chung, Y. G.; Snurr, R. Q.; Sholl, D. S. Large-Scale Refinement of Metal-Organic Framework Structures Using Density Functional Theory. *Chem. Mater.* **2017**, *29*, 2521–2528.
- 64 Cleeton, C.; Oliveira, F. L. d.; F. Neumann, R.; H. Farmahini, A.; Luan, B.; Steiner, M.; Sarkisov, L. A process-level perspective of the impact of molecular force fields on the computational screening of MOFs for carbon capture. *Energy Environ. Sci.* **2023**, *16*, 3899–3918.
- 65 Talirz, L.; Kumbhar, S.; Passaro, E.; Yakutovich, A. V.; Granata, V.; Gargiulo, F.; Borelli, M.; Uhrin, M.; Huber, S. P.; Zoupanos, S.; Adorf, C. S.; Andersen, C. W.; Schütt, O.; Pignedoli, C. A.; Passerone, D.; VandeVondele, J.; Schulthess, T. C.; Smit, B.; Pizzi, G.; Marzari, N. Materials Cloud, a platform for open computational science. *Sci Data* **2020**, *7*, 299.
- 66 NIST Standard Reference Simulation Website, available online at <https://www.nist.gov/programs-projects/nist-standard-reference-simulation-website>, last accessed 26 April 2024.
- 67 Altintas, C.; Avci, G.; Daglar, H.; Gulcay-Ozcan, E.; Erucar, I.; Keskin, S. Computer simulations of 4240 MOF membranes for H₂/CH₄ separations: insights into structure-performance relations. *J. Mater. Chem. A* **2018**, *6*, 5836–5847.
- 68 Mace, A.; Barthel, S.; Smit, B. Automated Multiscale Approach To Predict Self-Diffusion from a Potential Energy Field. *J. Chem. Theory Comput.* **2019**, *15*, 2127–2141.
- 69 Moghadam, P. Z.; Li, A.; Wiggin, S. B.; Tao, A.; Maloney, A. G. P.; Wood, P. A.; Ward, S. C.; Fairen-Jimenez, D. Development of a Cambridge Structural Database Subset: A Collection of Metal-Organic Frameworks for Past, Present, and Future. *Chem. Mater.* **2017**, *29*, 2618–2625.

- 70 Moosavi, S. M.; Nandy, A.; Jablonka, K. M.; Ongari, D.; Janet, J. P.; Boyd, P. G.; Lee, Y.; Smit, B.; Kulik, H. J. Understanding the diversity of the metal-organic framework ecosystem. *Nature Commun.* **2020**, *11*, 4068.
- 71 Majumdar, S.; Moosavi, S. M.; Jablonka, K. M.; Ongari, D.; Smit, B. Diversifying Databases of Metal Organic Frameworks for High-Throughput Computational Screening. *ACS Appl. Mater. Interfaces* **2021**, *13*, 61004–61014.
- 72 Rosen, A. S.; Fung, V.; Huck, P.; O'Donnell, C. T.; Horton, M. K.; Truhlar, D. G.; Persson, K. A.; Notestein, J. M.; Snurr, R. Q. High-throughput predictions of metal-organic framework electronic properties: theoretical challenges, graph neural networks, and data exploration. *npj Comput. Mater.* **2022**, *8*, 1–10.
- 73 Burner, J.; Luo, J.; White, A.; Mirmiran, A.; Kwon, O.; Boyd, P. G.; Maley, S.; Gibaldi, M.; Simrod, S.; Ogden, V.; Woo, T. K. ARC-MOF: A Diverse Database of Metal-Organic Frameworks with DFT-Derived Partial Atomic Charges and Descriptors for Machine Learning. *Chem. Mater.* **2023**, *35*, 900–916.
- 74 Zheng, Z.; Zhang, O.; Borgs, C.; Chayes, J. T.; Yaghi, O. M. ChatGPT Chemistry Assistant for Text Mining and the Prediction of MOF Synthesis. *J. Am. Chem. Soc.* **2023**, *145*, 18048–18062.
- 75 Glasby, L. T.; Gubsch, K.; Bence, R.; Oktavian, R.; Isoko, K.; Moosavi, S. M.; Cordiner, J. L.; Cole, J. C.; Moghadam, P. Z. DigiMOF: A Database of Metal-Organic Framework Synthesis Information Generated via Text Mining. *Chem. Mater.* **2023**, *35*, 4510–4524.
- 76 Park, H.; Kang, Y.; Choe, W.; Kim, J. Mining Insights on Metal-Organic Framework Synthesis from Scientific Literature Texts. *J. Chem. Inf. Model.* **2022**, *62*, 1190–1198.
- 77 Altintas, C.; Altundal, O. F.; Keskin, S.; Yildirim, R. Machine Learning Meets with

- Metal Organic Frameworks for Gas Storage and Separation. *J. Chem. Inf. Model.* **2021**, *61*, 2131–2146.
- 78 Sarikas, A. P.; Gkagkas, K.; Froudakis, G. E. Gas adsorption meets deep learning: voxelizing the potential energy surface of metal-organic frameworks. *Sci Rep* **2024**, *14*, 2242.
- 79 Jalali, M.; Wonanke, A. D. D.; Wöll, C. MOFGalaxyNet: a social network analysis for predicting guest accessibility in metal-organic frameworks utilizing graph convolutional networks. *J. Cheminform.* **2023**, *15*, 94.
- 80 Cao, Z.; Magar, R.; Wang, Y.; Barati Farimani, A. MOFormer: Self-Supervised Transformer Model for Metal-Organic Framework Property Prediction. *J. Am. Chem. Soc.* **2023**, *145*, 2958–2967.
- 81 Kang, Y.; Park, H.; Smit, B.; Kim, J. A multi-modal pre-training transformer for universal transfer learning in metal-organic frameworks. *Nature Mach. Intell.* **2023**, *5*, 309–318.
- 82 Choudhary, K.; Yildirim, T.; Siderius, D. W.; Kusne, A. G.; McDannald, A.; Ortiz-Montalvo, D. L. Graph neural network predictions of metal organic framework CO₂ adsorption properties. *Comput. Mater. Sci.* **2022**, *210*, 111388.
- 83 Cui, J.; Wu, F.; Zhang, W.; Yang, L.; Hu, J.; Fang, Y.; Ye, P.; Zhang, Q.; Suo, X.; Mo, Y.; Cui, X.; Chen, H.; Xing, H. Direct prediction of gas adsorption via spatial atom interaction learning. *Nature Commun.* **2023**, *14*, 7043.
- 84 Scheidgen, M.; Himanen, L.; Ladines, A. N.; Sikter, D.; Nakhaee, M.; Fekete, A.; Chang, T.; Golparvar, A.; Márquez, J. A.; Brockhauser, S.; Brückner, S.; Ghiringhelli, L. M.; Dietrich, F.; Lehmborg, D.; Denell, T.; Albino, A.; Näsström, H.; Shabih, S.; Dobener, F.; Kühbach, M.; Mozumder, R.; Rudzinski, J. F.; Daelman, N.; Pizarro, J. M.; Kuban, M.; Salazar, C.; Ondračka, P.; Bungartz, H.-J.; Draxl, C. NO-MAD: A distributed web-based platform for managing materials science research data. *J. Open Source Softw.* **2023**, *8*, 5388.

- 85 Artrith, N.; Butler, K. T.; Coudert, F.-X.; Han, S.; Isayev, O.; Jain, A.; Walsh, A. Best practices in machine learning for chemistry. *Nature Chem.* **2021**, *13*, 505–508.
- 86 Xiang, W.; Zhang, Y.; Chen, Y.; Liu, C.-j.; Tu, X. Synthesis, characterization and application of defective metal-organic frameworks: current status and perspectives. *J. Mater. Chem. A* **2020**, *8*, 21526–21546.
- 87 Wu, Y.; Duan, H.; Xi, H. Machine Learning-Driven Insights into Defects of Zirconium Metal-Organic Frameworks for Enhanced Ethane-Ethylene Separation. *Chem. Mater.* **2020**, *32*, 2986–2997.
- 88 Yeskendir, B.; Dacquin, J.-P.; Lorgouilloux, Y.; Courtois, C.; Royer, S.; Dhainaut, J. From metal-organic framework powders to shaped solids: recent developments and challenges. *Mater. Adv.* **2021**, *2*, 7139–7186.
- 89 Senkovska, I.; Bon, V.; Abylgazina, L.; Mendt, M.; Berger, J.; Kieslich, G.; Petkov, P.; Luiz Fiorio, J.; Joswig, J.-O.; Heine, T.; Schaper, L.; Bachetzky, C.; Schmid, R.; Fischer, R. A.; Pöppel, A.; Brunner, E.; Kaskel, S. Understanding MOF Flexibility: An Analysis Focused on Pillared Layer MOFs as a Model System. *Angew. Chem. Int. Ed.* **2023**, *62*, e202218076.
- 90 Jablonka, K. M.; Patiny, L.; Smit, B. Making the collective knowledge of chemistry open and machine actionable. *Nature Chem.* **2022**, *14*, 365–376.
- 91 Zheng, Z.; Rong, Z.; Rampal, N.; Borgs, C.; Chayes, J. T.; Yaghi, O. M. A GPT-4 Reticular Chemist for Guiding MOF Discovery. *Angew. Chem. Int. Ed.* **2023**, *62*, e202311983.
- 92 Kojima, S.; Park, J.; Carter, E. A.; Walton, K. S.; Realff, M. J.; Sholl, D. S.; Yajima, T.; Fujiki, J.; Kawajiri, Y. Discrepancy quantification between experimental and simulated data of CO₂ adsorption isotherm using hierarchical Bayesian estimation. *Sep. Purif. Technol.* **2022**, *296*, 121371.

- 93 Osterrieth, J. W. M.; Rampersad, J.; Madden, D.; Rampal, N.; Skoric, L.; Connolly, B.; Allendorf, M. D.; Stavila, V.; Snider, J. L.; Ameloot, R.; Marreiros, J.; Ania, C.; Azevedo, D.; Vilarrasa-Garcia, E.; Santos, B. F.; Bu, X.-H.; Chang, Z.; Bunzen, H.; Champness, N. R.; Griffin, S. L.; Chen, B.; Lin, R.-B.; Coasne, B.; Cohen, S.; Moreton, J. C.; Colón, Y. J.; Chen, L.; Clowes, R.; Coudert, F.-X.; Cui, Y.; Hou, B.; D'Alessandro, D. M.; Doheny, P. W.; Dincă, M.; Sun, C.; Doonan, C.; Huxley, M. T.; Evans, J. D.; Falcaro, P.; Ricco, R.; Farha, O.; Idrees, K. B.; Islamoglu, T.; Feng, P.; Yang, H.; Forgan, R. S.; Bara, D.; Furukawa, S.; Sanchez, E.; Gascon, J.; Telalović, S.; Ghosh, S. K.; Mukherjee, S.; Hill, M. R.; Sadiq, M. M.; Horcajada, P.; Salcedo-Abraira, P.; Kaneko, K.; Kukobat, R.; Kenvin, J.; Keskin, S.; Kitagawa, S.; Otake, K.-i.; Lively, R. P.; DeWitt, S. J. A.; Llewellyn, P.; Lotsch, B. V.; Emmerling, S. T.; Pütz, A. M.; Martí-Gastaldo, C.; Padial, N. M.; García-Martínez, J.; Linares, N.; MasPOCH, D.; Suárez del Pino, J. A.; Moghadam, P.; Oktavian, R.; Morris, R. E.; Wheatley, P. S.; Navarro, J.; Petit, C.; Danaci, D.; Rosseinsky, M. J.; Katsoulidis, A. P.; Schröder, M.; Han, X.; Yang, S.; Serre, C.; Mouchaham, G.; Sholl, D. S.; Thyagarajan, R.; Siderius, D.; Snurr, R. Q.; Goncalves, R. B.; Telfer, S.; Lee, S. J.; Ting, V. P.; Rowlandson, J. L.; Uemura, T.; Iiyuka, T.; van der Veen, M. A.; Rega, D.; Van Speybroeck, V.; Rogge, S. M. J.; Lamaire, A.; Walton, K. S.; Bingel, L. W.; Wuttke, S.; Andreato, J.; Yaghi, O.; Zhang, B.; Yavuz, C. T.; Nguyen, T. S.; Zamora, F.; Montoro, C.; Zhou, H.; Kirchon, A.; Fairen-Jimenez, D. How Reproducible are Surface Areas Calculated from the BET Equation? *Adv. Mater.* **2022**, *34*, 2201502.
- 94 Domingues, N. P.; Moosavi, S. M.; Talirz, L.; Jablonka, K. M.; Ireland, C. P.; Ebrahim, F. M.; Smit, B. Using genetic algorithms to systematically improve the synthesis conditions of Al-PMOF. *Commun. Chem.* **2022**, *5*, 1–8.
- 95 Pilz, L.; Natzeck, C.; Wohlgemuth, J.; Scheuermann, N.; Weidler, P. G.; Wagner, I.; Wöll, C.; Tsotsalas, M. Fully Automated Optimization of Robot-Based MOF Thin Film

Growth via Machine Learning Approaches. *Adv. Mater. Interfaces* **2023**, *10*, 2201771.

96 Xie, Y.; Zhang, C.; Deng, H.; Zheng, B.; Su, J.-W.; Shutt, K.; Lin, J. Accelerate Synthesis of Metal-Organic Frameworks by a Robotic Platform and Bayesian Optimization. *ACS Appl. Mater. Interfaces* **2021**, *13*, 53485–53491.

TOC Graphic

