

## Review



**Cite this article:** Fraux G, Chibani S, Coudert F-X. 2019 Modelling of framework materials at multiple scales: current practices and open questions. *Phil. Trans. R. Soc. A* **377**: 20180220. <http://dx.doi.org/10.1098/rsta.2018.0220>

Accepted: 22 February 2019

One contribution of 10 to a theme issue 'Mineralomimesis: natural and synthetic frameworks in science and technology'.

**Subject Areas:**

materials science, computational chemistry

**Keywords:**

molecular simulation, modelling, framework materials, machine learning, databases

**Author for correspondence:**

François-Xavier Coudert

e-mail: [fx.coudert@chimieparistech.psl.eu](mailto:fx.coudert@chimieparistech.psl.eu)

# Modelling of framework materials at multiple scales: current practices and open questions

Guillaume Fraux, Siwar Chibani and

François-Xavier Coudert

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

GF, 0000-0003-4824-6512; F-XC, 0000-0001-5318-3910

The last decade has seen an explosion of the family of framework materials and their study, from both the experimental and computational points of view. We propose here a short highlight of the current state of methodologies for modelling framework materials at multiple scales, putting together a brief review of new methods and recent endeavours in this area, as well as outlining some of the open challenges in this field. We will detail advances in atomistic simulation methods, the development of material databases and the growing use of machine learning for the prediction of properties.

This article is part of the theme issue 'Mineralomimesis: natural and synthetic frameworks in science and technology'.

## 1. Introduction

Nanoporous materials with high specific surface area are extensively used in a wide range of applications, including catalysis, ion exchange, gas storage, gas or liquid separations, sensing and detection, electronics and drug delivery. The last 15 years have seen the emergence of entire new classes of crystalline nanoporous materials, based on weaker bonds (coordination bonds,  $\pi$ - $\pi$  stacking, hydrogen bonds, ...). The most studied of these new materials are the metal-organic frameworks (MOFs): these nanoporous hybrid organic-inorganic materials, built from metal centres interconnected by organic linkers, have been the subject of an intensive research effort since the pioneering work done by R. Robson in the

1990s, with thousands of structures synthesized. Other classes of crystalline nanoporous materials that have emerged in the past decade include covalent organic frameworks, porous molecular organic solids and other porous molecular framework materials.

Among these nanoporous materials, an interesting family of materials has recently started to emerge, named ‘stimuli-responsive materials’ or ‘soft porous crystals’ [1], which exhibit large or anomalous responses to external physical or chemical stimulation [2]. These modifications of framework structure and pore dimensions also involve, in turn, a modification of other physical and chemical properties, making such materials multifunctional (or ‘smart materials’). Stimuli-reactive crystals include a wide diversity of eye-catching phenomena such as negative adsorption [3], negative linear compressibility or negative area compressibility [4], pressure-induced bond rearrangement and framework topology changes [5], photoresponsive frameworks [6] and intrusion-induced polymorphism [7], to name a few. Each of these properties can be leveraged for applications in several fields, for example to make sensors and actuators, to store mechanical energy, to engineer composite materials with targeted mechanical and thermal properties, etc.

Soft framework materials, because they are built from weaker interactions, have large-scale complex supramolecular architectures, and can exhibit many dynamic phenomena such as those just described, are a particular challenge in terms of computational modelling. Compared to ‘traditional’ dense materials, such as oxides, they can require additional computational power (due to the increased time and length scales involved), or even novel simulation methodologies. In this paper, we propose a brief review of new methods and recent endeavours in this area, of the perspectives opened, as well as outline some of the open challenges in this field. We will first detail recent advances in atomistic simulation methods for framework materials, going beyond structural properties of perfect crystals to address their behaviour under stimulation and in a large range of working conditions, as well as the emergence of defects and disordered phases. We will then highlight the recent development of material databases, and within this the specific place of framework materials. Finally, our last section will focus on the growing use of machine learning techniques for the prediction of complex material structures and properties.

## 2. Computational methods for framework materials

### (a) Classical and *ab initio* simulations

If one wants to understand the properties and behaviour of a crystalline material using computational methods, the usual starting point is to compute ‘static’ properties of the perfect infinite crystal, using quantum chemistry methods, such as Kohn–Sham density functional theory. Starting from an energy-minimized (relaxed) structure, researchers can then compute zero Kelvin properties, at or around that energy minimum: structural and electronic properties, such as the band gap and the band structure; vibrations of the atoms around their equilibrium position, computed as phonons; and infinitesimal deformations of the system can yield elastic properties. For ‘traditional’ materials, such as oxides, metallic alloys or other dense inorganic materials, most of the behaviour and properties of a system can be computed using such methodology. In stark contrast, for complex framework materials with highly dynamic behaviour, this might not be enough and one has to resort to more complex and more demanding simulation methods. Specifically, for soft porous materials, their dynamic properties and response to various external stimuli play a crucial part in their properties and possible applications. In this case, exploring the behaviour of the system in the vicinity of its energy-minimal structure is not sufficient, and molecular dynamics (MD) simulations can be necessary to adequately describe the behaviour of the material—as well as providing important insights into the atomistic processes governing the macroscopic behaviour.

The so-called classical MD simulations, relying on parameterized force-fields to represent intra- and intermolecular interactions, have the advantage of being usable for big simulations,

either in the duration of the simulated events or the size of the system. This means that we can study rare events such as crystal nucleation or reactions—as well as systems where a large simulation box is needed, for example the effects of disorder and defects (a topic which we will discuss more below). The issue here is that there are very few reliable, well-tested and transferable force-fields for use with framework materials. One has to choose between: (i) force fields derived for a single material, which describe the potential energy surface of the system with high accuracy, but are not transferable to other materials; (ii) generic force-fields, whose analytical expressions and parameters are transferable among a large class of material, but that poorly reproduce physical properties. The second approach has been widely used, by relying on generic force-fields such as AMBER [8] or UFF [9]—possibly with adjustments or extensions—to get a consistent treatment of all frameworks, and therefore to compare different materials when searching out the best candidate for a given application in high-throughput studies [10–12]. One problem arising from this approach is that these force-fields might not contain adequate terms to describe the delicate balance of intra- and intermolecular interactions in framework materials. In particular, one can think of the metal coordination bonds,  $\pi$ - $\pi$  stacking and other soft intermolecular interactions. On the other hand, deriving new force-fields for a specific systems, while useful to investigate the behaviour of a given material thanks to higher accuracy of the potential energy surface, fails to allow for comparisons with other systems and is not suitable for large-scale screening.

Another choice of methodology is to use an *ab initio* description of the interactions in the system, where a quantum chemistry method is used at every time step of the MD simulation—this approach is also called first-principles molecular dynamics (FPMD). This has a much higher computational cost, and thus limits the length and time scales that can be reached, but does not make any assumption on the nature of the interactions. This was used by Chaplais *et al.* [13] to describe how the adsorbed phase arranges inside a fully flexible ZIF-8, without needing to create a classical force-field that would be able to reproduce the full flexibility of ZIF-8. Furthermore, FPMD allows the description of bond breaking and formation, which can be crucial in some dynamic phenomena: as an example, Howe *et al.* [14] used it to analyse the stability of MOFs in the presence of water.

We note that the question of the ‘level of description’ applied to the systems (quantum chemistry versus empirical potentials) is relevant not only for MD but also for Monte Carlo simulations, which stochastically generate representative configurations of the system in a given thermodynamic ensemble, by the application of random moves weighted by the appropriate Boltzmann probabilities. However, while *ab initio* Monte Carlo simulations are possible, the large number of energy evaluations necessary make them relatively rare in the literature [15,16]. In the context of framework materials, Monte Carlo simulations are used at various scales. First, simulated annealing and biased Monte Carlo simulations are extensively used in the areas of structure solution and for localization of extra-framework ions and adsorbed species [17,18]. Secondly, Grand Canonical Monte Carlo is very often used to describe the thermodynamics of adsorption of fluids and fluid mixtures in nanoporous frameworks [19]. Finally, mesoscale Monte Carlo modelling methods can be used to assess the large-scale ordering (or disorder) in supramolecular frameworks, based on carefully constructed Hamiltonians that describe the local interactions [20–22].

## (b) Make force-fields great again

Despite the rather strong limitations of force-fields described above for their application to framework materials, there have been several recent developments in that area, which we want to highlight here. Deriving a new force-field for a material is a hard and long task, where one needs not only to gather or generate reference data, but also to adapt parameters and check every time that the physical properties predicted by the force-field are right. In the past few years, novel methodologies for force-field fitting have been proposed, relying on machine learning algorithms. They aim to make the process more automatic, more reproducible, and also reduce its reliance on

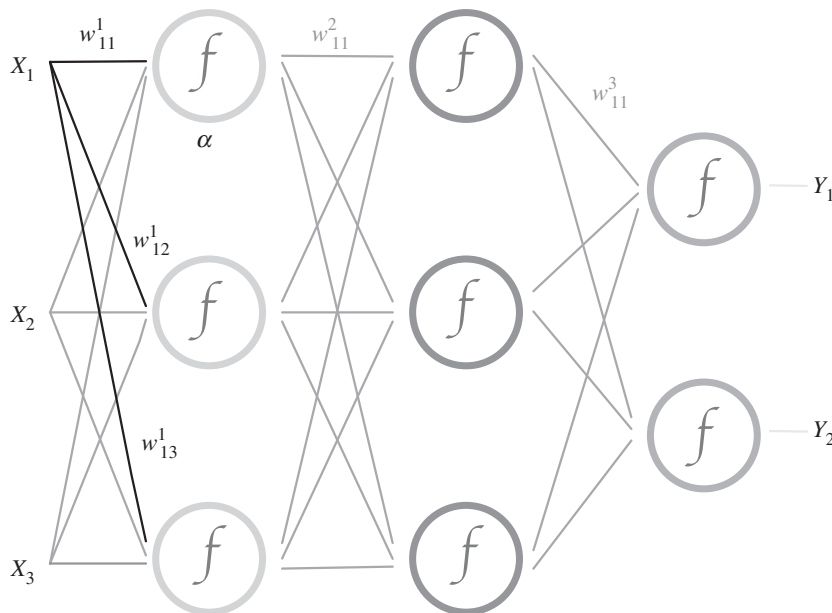
human input. Starting from the structure optimized with *ab initio* calculations and the Hessian at this energy minimum, a machine-learning procedure (for example, a genetic algorithm coupled with a least-squares minimization) finds the optimal set of parameters matching the structure and the Hessian. Some implementations of this idea are the MOF-FF [23] and QuickFF [24] force-fields—or maybe more accurately, force-field optimization methodologies. While they use slightly different input data and fitting procedures, they share the common goal of parameterizing force-fields in a systematic and consistent fashion, from first-principles reference data.

To give an example of the use of these new force-field methodologies, MOF-FF was recently used to predict the most stable structure and topology for copper paddle-wheel MOFs depending on the linker [25]. The authors generated all the structure by combining simple building blocks (linkers and copper paddle-wheel) with different topologies, and were then able to use the same force-field to optimize and study them all. Finally, we note that these methods were originally developed by relying on reference data gathered on (finite) clusters representative of the MOF structures, and were later extended to periodic input data. The use of periodic structures as a reference was shown to be essential for a correct description of structural, vibrational and thermodynamic properties of soft framework materials like MIL53(Al) by QuickFF [26].

Despite this progress, classical force-fields remain fundamentally limited by the analytical form they choose to represent interactions, even when parameterized in an optimal fashion. For example, a force-field using a Lennard–Jones dispersion potential will be unable to reproduce any long-range interaction that does not follow this functional form. A promising alternative, in order to be able to reproduce any possible interaction profile coming from the reference data (i.e. quantum chemistry calculations), is the use of neural-networks force-field. Neural networks are algorithms that map a set of input values to a set of output values by associating adjustable weights with each value, and then using a nonlinear function (called the activation function) to map the weighted inputs to the outputs. If the outputs are then fed to another neural network, the resulting network is said to have multiple layers—see figure 1 for a graphical representation with three layers. One property of neural networks is their ability to reproduce arbitrary multidimensional  $\mathbb{R}^n \rightarrow \mathbb{R}^m$  functions with arbitrary accuracy [27]. This makes them very appealing to reproduce energy or forces from *ab initio* calculations, using only the atomic position as input—effectively functioning like a force-field, without any assumptions on the nature of the interactions. Before being usable, the network must be ‘trained’ with data representative of the system of interest. During this training, the weights are adapted to ensure a correct mapping from the input (the atomic positions) to the output (forces and energy). Using atomic positions in Cartesian coordinates as the input is not optimal, as the generated network will only be usable with the exact same system used for training. An alternative is to rely solely on the local environment of an atom up to a cutoff distance, represented in a translation and rotation-independent manner [28]. Neural network force-fields are a very promising approach to cheap simulation with high accuracy, and they are already used for small organic molecules [29], water [30,31]; as well as classical dense crystalline materials [32], and amorphous inorganic materials [33,34]. They are especially helpful with amorphous materials such as silicon or glasses, where the usual classical potential are complex multi-body potentials. This approach still remains to be extended to porous frameworks materials.

### (c) Simulating complex systems

One of the biggest current challenges in the simulation of framework materials lies in the complexity of these systems. The computational cost of our tools imposes limits on the systems we can model, in terms of length scale (and thus number of atoms) and time scales. For crystalline phases, the use of periodic boundary conditions, where the simulated system is repeated in all spatial dimensions, is a very effective way to describe infinite systems within computers with limited memory and CPUs. However, this approach falls short when we want to study phenomena involving large correlation lengths, such as dynamic properties of soft materials. Another difficult area is the computational modelling of disordered phases, where



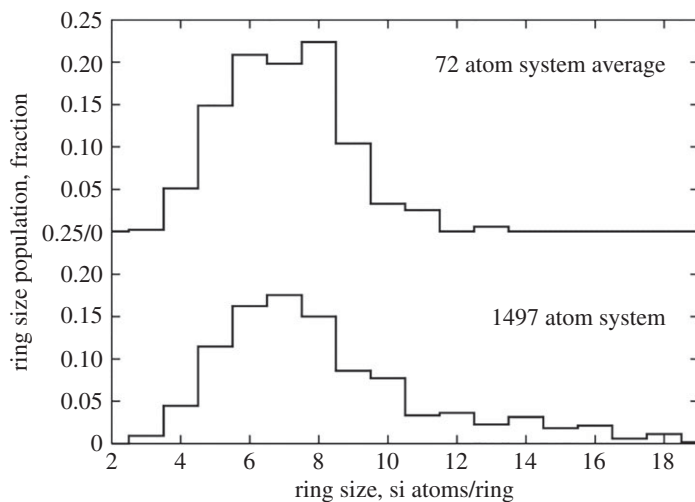
**Figure 1.** A representation of a neural network mapping from  $X_i$  to  $Y_j$  using three layers. Circles represent neurons using  $f$  as the activation function, and lines are adjustable weights  $w_{ij}^\alpha$  for the input values. In this case, the output of the neuron  $\alpha$  would be  $f(\sum_i w_{i1}^1 X_i)$ . (Online version in colour.)

a very large simulation box would be necessary to correctly describe the system. Yet, within the field of framework materials, such disordered systems are attracting a lot of interest due to their properties that differ from their crystalline counterparts. We can here cite as examples systems such as MOF glasses [35–37] and liquids [38], or framework materials with defects and correlated disorder [39]. There is thus an important drive to model these materials, because of their properties (e.g. amorphous phases can have more appealing mechanical and optical properties than crystals) or because catalysis, nucleation or adsorption can occur preferentially around defects.

A strategy that can be used in this case—if the brute force approach of using a very large simulation box is not feasible—is to use multiple realizations (or ‘copies’) of the system of interest, and average the measured properties over those replicas. This approach has been extensively used in the past for the study of amorphous systems such as silica glasses or disordered carbons. For example, Van Ginhoven *et al.* [40] used DFT calculations on 18 different configurations of silica glass created using a classical force-field, and were thus able to obtain a good statistical representation of static and dynamic properties with comparable or better accuracy than a longer, bigger simulation (figure 2).

Another strategy to study large-scale systems is to change the level of description, moving closer to mesoscopic methods and using coarse-grained force-fields instead of atomistic ones. Dürholt *et al.* [41] have generated such a coarse-grained force-field for the HKUST-1 MOF, based on copper paddle-wheels. These authors showed that even a very coarse model is able to reproduce the low-energy deformations of the system, with only one coarse-grained bead for 30 atoms. Another mesoscopic approach, in the field of adsorption, is the use of Lattice Boltzmann methods to describe the coupling between fluid flow and adsorption in porous media with complex geometries [42].

Beyond this scale, it can also be useful to turn to macroscopic modelling methods to simulate even larger systems. Indeed, many potential applications of framework materials are based on their use not as single crystals, but are expected to construct nanostructured or composite systems: common examples include monoliths, supported thin films and mixed-matrix membranes. In



**Figure 2.** Ring size distribution in silica glass, as computed from 18 small (72 atoms) replicas, or a single big (1497 atoms) simulation. Reprinted with permission from [40]. Copyright © 2005 by the American Physical Society.

order to describe these composite systems, one has to turn to conventional microscopic modelling methods: finite elements for solid mechanics, computational fluid dynamics to describe transport, etc. In this vein, Evans *et al.* [43] used a macroscopic description and finite-element methods to compute deformation properties of mixed-matrix membrane and other composite of framework materials and polymers. The use of finite-element methods allowed them to study sizes up to 400  $\mu\text{m}$ , which is five orders of magnitude bigger than typical atomistic simulations.

Finally, we note that while we are starting to see new techniques and methods that go from one level of description to the next (quantum to classical, micro to meso, meso to macro), the bridging of those various scales of simulations into a coherent multi-scale simulation methodology is still a widely open research question. How can one use data from *ab initio* simulation to fit atomistic classical force-field? [23,44] Or leverage force-field-based data to create a coarse-grained model? [41] Or transfer microscopic properties into input for a finite-element method? [43,45] Every time we go up a level of description, we are able to work with bigger systems at longer timescales, at the cost of some accuracy and precision, but we still lack a systematic way to create and validate these novel models for performance and accuracy.

#### (d) Describing excited states

We note here that a particularly challenging area of the modelling of framework materials is that of the description of their excited states, in order to better understand, e.g. their optical properties and photocatalytic activity. Such phenomena involve transitions between the system's ground state and another state of higher energy (the excited state) upon photon absorption or emission. The energy difference involved in the electronic transitions is directly related to the position of absorption and emission bands. Theoretical models can give insight into the properties of electronically excited states, and are therefore a useful complement to experimental measurements. In that framework, density functional theory (DFT), and more precisely its time-dependent form (TD-DFT) [46,47], is the *ab initio* method of choice for most of the cases [48,49], as it may treat structures containing up to *ca* 300 atoms. To study framework materials, Wilbraham *et al.* have developed a computational protocol in order to simulate the optical signatures of two MOF structures based on the 4,4'-bis((3,5-dimethyl-1H-pyrazol-4-yl)methyl)-biphenyl ( $\text{H}_2\text{DMPMB}$ ) linker. The developed protocol was successfully applied to characterize and to rationalize the adsorption and the emission behaviour on the interchange of zinc and cadmium

as metal cation [48]. Another important optical property in hybrid materials is the nature of the electronic excitations that could present ligand-to-metal charge transfer (LMCT) characteristics. Very recently, Wu *et al.* showed that from different cations, electronic excitations occur in the linker of the UiO-66(Ce) MOF upon light absorption. These authors showed that incorporation of the cerium cation presents an effective way not only to stabilize the LMCT, but also to increase the photocatalytic activity of UiO-66 MOF [49]. For applications in photocatalysis, the magnitude of the band gap and the absolute positions of the band edges are of high importance [50,51]. As an example, based on the mixing of organic linkers, Ricardo *et al.* have designed new ZIF materials with a narrower band gap in order to allow the absorption of the visible range solar spectrum. They showed that by introducing a transition metal (copper) in the tetrahedral position of the mixed-linker ZIFs, it is possible to increase photo-adsorption [51].

### 3. Material databases

As stated in the introduction, the last decade has seen an important increase in the number of studies on various families of framework materials, with the goal of discovering or designing novel materials with targeted properties. Given the large number of materials synthesized, characterized and reported, three important series of questions arise:

- (i) Where and how is the information on these materials stored? What are the available **data**?
- (ii) Under what form is it stored, how can it be queried, retrieved and interpreted? That is, issues of **Application Programming Interface (API), format and interoperability**.
- (iii) What is the extent of information and properties provided for each structure? How were they determined? Those are questions about the **metadata** associated with each structure.

In this section, we will briefly review the current state of the art and describe some of the existing material databases for framework materials, contrasting the situation with that of inorganic materials.

#### (a) Zeolites

Let us start with the grandparent of this family of databases, namely the database of zeolite structures from the International Zeolite Association (IZA), which is freely available on the Internet at <http://www.iza-structure.org/databases/>. Most of the information is also available in the printed form, as the *Atlas of zeolite framework types* book [52]. Zeolites belong to the class of nanoporous materials and are composed of oxygen, silicon and aluminium. They have widespread applications at the industrial level in the fields of catalysis, adsorption and separation [53–56]. At the current date, the corresponding database provides structural information for 230 zeolite framework types reported experimentally, among which 67 are natural zeolites. Ten years ago, only 176 zeolite frameworks were known, showing that even among ‘conventional’ porous materials, progress is steady and the synthesis of new zeolites remains a considerable challenge. The IZA database is heavily curated, as all the zeolitic structures it includes have been approved by the Structure Commission of the IZA, to verify that it is unique and that the structure has been satisfactorily proved.

The nomenclature for these materials is recognized by IUPAC (the International Union of Pure and Applied Chemistry) and is assigned by a three letter code—such as FAU, for the faujasite framework, or MOR for the mordenite framework. Data associated with each framework type code include crystallographic data: space group, cell parameters, positions of vertices in the idealized framework, but also topological density, ring size, channel dimensions, maximum diameter of an included sphere, accessible volume and composite-building units. Moreover, going beyond idealized framework structure and topological properties, the database features detailed information for building models, and simulated powder diffraction patterns for representative materials, as well as all corresponding literature references.

At this stage, the reader unfamiliar with zeolites may be surprised that only 230 zeolitic structures have been identified experimentally. Indeed, at the molecular scale, zeolites are constituted of  $\text{TO}_4$  tetrahedra (where typically  $\text{T} = \text{Al}$  or  $\text{Si}$ ), connected by their corners. It is mathematically possible to create an infinite number of such four-connected nets that have three-dimensional periodicity. The question of why only a few structures are experimentally realized, known as ‘zeolite feasibility’, is still wide open [57,58]. Nevertheless, researchers have used theoretical and computational tools to develop databases of hypothetical zeolitic structures—based on four-connected nets, but usually with added constraints such as an upper bound on the lattice energy or topology. Compared to the experimental zeolites, the number of hypothetical zeolitic structures is much larger and rapidly growing. In the first such database published, by Li *et al.* [59–61] and available at <http://mezeopor.jlu.edu.cn/hypo/>, two sets of hypothetical zeolite structures are provided [61,62]. The first set is generated by the FraGen algorithm, which is based on Monte Carlo direct space structure modelling [63]. The second set is composed of so-called ABC-6 structures, which are enumerated through a material genome approach [64]. The number of all the ABC-6 structures is 84 292. Besides their structures in CIF format, all hypothetical structures are assembled in an Excel spreadsheet listing their properties, such as stacking layers, stacking sequences, space groups, cell dimensions, channel openings, framework energies, framework densities, stacking compactness and the constituent cages [61].

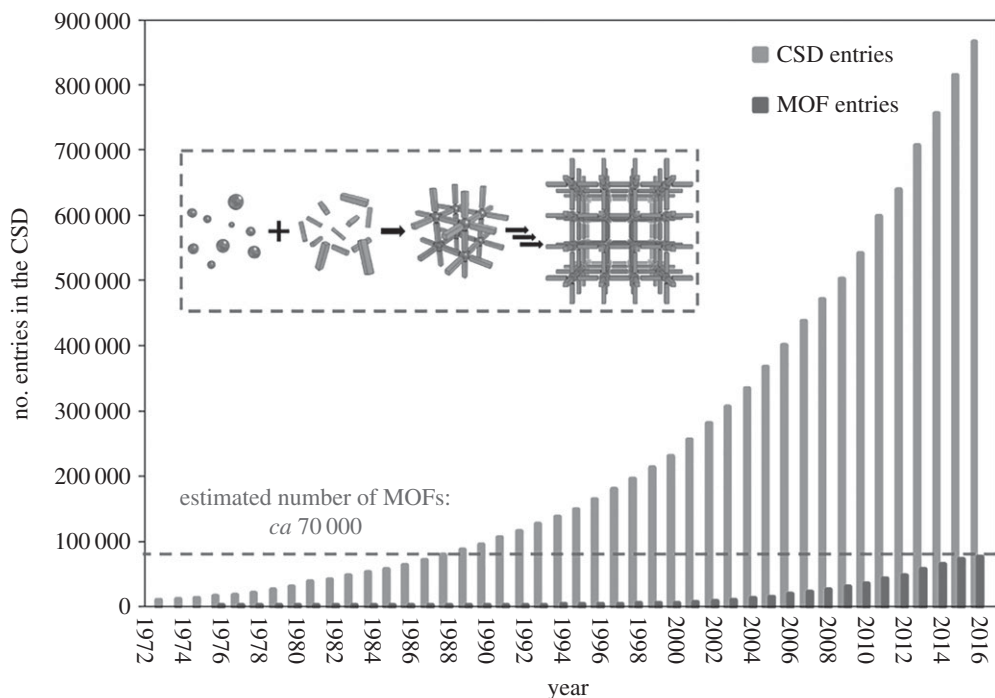
A second hypothetical zeolite structure database, available at <http://www.hypotheticalzeolites.net>, was generated by Treacy and Foster [65,66]. It contains 5 million different frameworks, triaged into ‘bronze’ and ‘silver’ sets, depending on their feasibility based respectively on a specifically designed cost function and force-field energy minimization. The two sets contain 5 389 408 bronze and 1 270 921 silver structures, and have been used as starting points for a series of theoretical surveys of zeolitic frameworks [67] and related four-connected frameworks [68,69]. Using the Monte Carlo approach, Earl *et al.* have developed a systematic computational procedure to search through unit cells with different space group symmetries [70], called the symmetry-constrained intersite bonding search (SCIBS) approach. They have used it to generate a third database of 2.6 million zeolite-like materials that have topological, geometrical and diffraction characteristics that are similar to those of known zeolites [71]. All three hypothetical zeolite databases are maintained by individual research groups, and are not open to external submissions of new structures.

Besides the aluminosilicate zeolites, open-framework aluminophosphates, or AIPOs, constitute an important class of microporous inorganic materials with a variety of structures ranging from neutral zeolites to anionic frameworks. The AIPO framework is not only limited to Al and Si as tetrahedral atoms: the upper limit of pore size can go beyond 12-membered rings, and the primary building units are not restricted to tetrahedra. This gives the AIPO family a rich variety of structural architectures and physico-chemical properties. There is an AIPO database, available online at <http://mezeopor.jlu.edu.cn/alpo/>, developed by Y. Li, J. Yu and R. Xu. It contains over 200 experimental AIPO structures reported in the literature [72]. In addition to general information, such as formula, space group, cell parameters and atomic coordinates, this database also includes more detailed structural information, such as coordination environment, Al/P ratio, stacking sequences for two-dimensional structures and coordination sequences. Simulated XRD reflections and references are also included to aid the identification of samples of users.

## (b) Metal-organic frameworks

MOFs appeared almost 30 years ago, and designate a class of materials composed of inorganic nodes linked by organic ligands. These are a novel generation of materials, with promising applications to follow zeolites in catalysis and adsorption-related applications. Since their discovery, the growth in the number of MOF structures reported in the Cambridge Structural Database (CSD) has been staggering, as shown in figure 3. The latter contains more than 900 000 structures of small molecule crystal structures and materials, among which 70 000 MOF materials



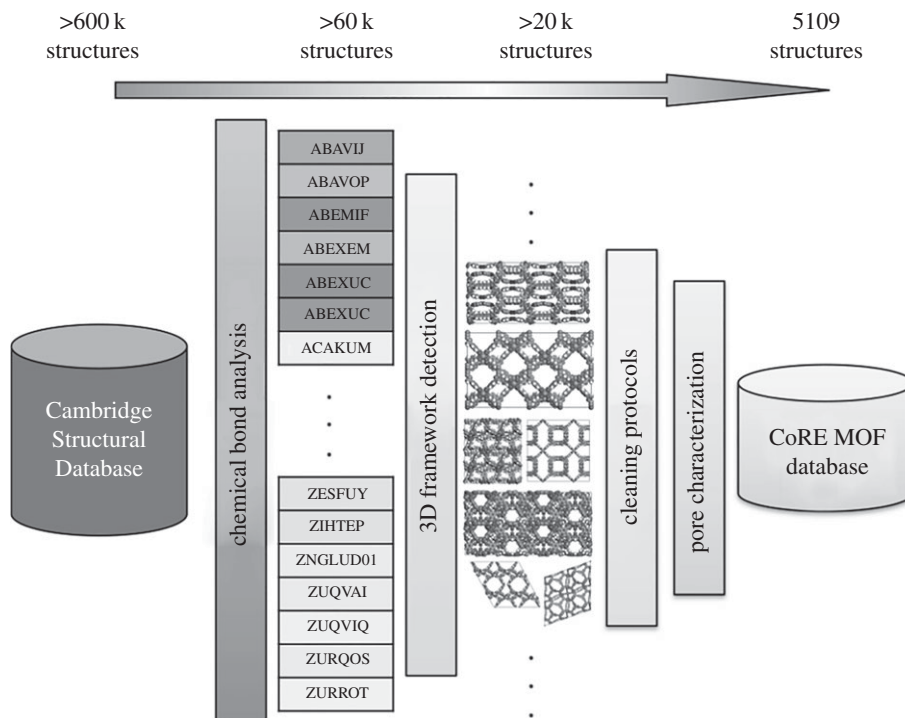


**Figure 3.** Schematic of the CSD and MOF entries since 1972. The inset shows the MOF self-assembly process from building blocks: metal nodes (red spheres) and organic ligands (blue struts). Reprinted with permission from [73]. Copyright © 2017 American Chemical Society. (Online version in colour.)

can be found. Each crystal structure undergoes extensive validation and cross-checking by expert chemists and crystallographers to ensure that the database is maintained to the highest possible standards. Apart from X-ray, neutron diffraction analyses and three-dimensional structure, every entry is enriched with bibliographic, chemical and physical information. Even though all published MOF structures are collected in the CSD, it is not easy to distinguish them from the rest of the structures in the CSD. In this vein, Watanabe *et al.* have extracted 30 000 extended MOF compounds from the CSD, among which 1163 MOF materials were applied for CO<sub>2</sub>/N<sub>2</sub> separation [74]. In 2013, Goldsmith *et al.* published an automated approach for screening 20 000 porous structures in the CSD useful for hydrogen storage [75]. This requires the use of algorithms for virtual solvent removal, and relies on an established empirical correlation between excess hydrogen uptake and surface area.

In 2014, Chung *et al.* developed a curated database of MOF structures, named the ‘Computation-Ready Experimental MOFs’ (CoRE MOF) database; it is available at <https://gregchung.github.io/CoRE-MOFs/>. It contains over 6000 three-dimensional MOFs, with solvents and templating agents cleaned, and with a pore limiting diameter (PLD) larger than 2.4 Å [76]. The protocol used to generate the database, represented in figure 4, is the following: (i) identify and extract MOF structures from CSD, based on atomic types and bonds present; (ii) remove solvent molecules and included templates; (iii) in some cases, remove disorder. Several recent studies have used this database as a starting point [77,78]. Additional computational data can also be added to the database, as did the Sholl group by computing and publishing point charges derived from periodic DFT calculations for more than 2000 structures in the CoRE MOF database [79]. This allows for easier reuse by other research groups, as a starting point for adsorption calculations of polar molecules, for example.

Despite the importance of these CSD-derived databases, they are not integrated within the CSD, and thus require manual updates over time, as new entries are added to the CSD. To



**Figure 4.** Schematic illustration of the CoRE MOF database construction. Reprinted with permission from [76]. Copyright © 2014 American Chemical Society. (Online version in colour.)

address this deficiency, Moghadam *et al.* have recently implemented seven criteria for MOFs embedded within a custom CSD Python Application Programming Interface (API) workflow [73]. The constructed CSD MOF is currently integrated into the CCDC's (Cambridge Crystallographic Data Centre) structure search program ConQuest, which allows for tailored structural queries and visualization. CSD MOF thus presents the most complete collection of MOFs, and will stay synchronized with the CSD as time goes by. The authors have also developed an array of computational algorithms in order to remove the solvent molecules from the CSD MOF subset, and then to calculate the geometric and physical properties for all the structures in the database.

Finally, we should note here that some effort has also been devoted to designing hypothetical MOFs structures. In this quest, Wilmer *et al.* have generated a database of 137 953 hypothetical MOF structures from 102 different building blocks, containing secondary building units (SBU) and organic linkers [80]. The authors then used this database as a starting point for computational screening, with the goal of identifying the best candidates for specific applications. This was applied, for example, to the cases of hydrogen storage, methane storage and adsorption/stability of water [81–83]. However, once a computational screening approach has identified possible targets, the design of synthesis protocol for these hypothetical materials, as well as their feasibility, is still often a complex issue.

### (c) The Materials Project

For other crystalline compounds, and for inorganic solids in particular, there have been a large number of databases, often with a specific focus on a particular class of materials. Most—but not all—are dedicated to experimental structures and properties. They are briefly reviewed in [84], for the interested reader. We want to focus here on a recent development, the development of the Materials Project, which provides a material database as well as an open API (and web portal) to

computed information on known and predicted materials. As we are writing this, it includes information about 86 371 inorganic compounds, and it is regularly updated with additional entries. It also aggregates nanoporous structures from several databases, including CoRE MOFs, hypothetical MOFs and zeolites described in the previous sections, as well as computational predicted porous polymeric networks (PPNs). The main goal of this database is to accelerate advanced material discovery and deployment [85]. Classes of materials that feature a specific focus include battery materials, intercalation electrodes and conversion electrodes.

The database is open—after registration—and accessible through its own open-source API. A high-quality reference implementation of this API is provided as part of the open-source Python Materials Genomics (pymatgen) material analysis package, available at <http://pymatgen.org/>. In addition to the Materials Project API, pymatgen is a generic material-oriented Python library, with classes for the representation of elements, sites, molecules and periodic structures, input/output support for several common file formats, analysis tools for electronic structure and physical properties, etc. For non-programmers, the Materials Project also includes a web front-end at <https://materialsproject.org/>, through which one can access the large dataset. Properties, such as space group, X-ray diffraction, band structures and elastic properties, can be browsed or searched. This architecture is extensible; for example, our group has recently provided an integration of the online ELATE application for the analysis and visualization of elastic tensors [86]. This ELATE analysis and visualization is linked from every Materials Project entry that contains elastic data, i.e. every crystalline solid for which the elastic stiffness tensor has been computed by DFT calculations. In 2015, Jong *et al.* reported elastic properties for 1181 inorganic compounds [87]. This number has since grown, and the database currently contains elastic information for 13 934 inorganic compounds—and this number is still growing.

## 4. Machine learning for property prediction

While the databases of structures, both experimentally determined and hypothetical, grow at a fast pace, the efforts to add physical and chemical properties of these materials in databases are happening on a longer timescale. The current theoretical chemistry methods, using microscopic (quantum chemistry and classical molecular modelling) and mesoscopic scales, make it possible to predict and understand the physical and chemical behaviour of given materials that already exist. However, these methods are computationally intensive, and their use on a very large scale is somewhat limited. Computational screening studies based on existing databases, as we have described below, are often limited to very simple descriptors of a material's performance for a given application. They are often used in a multi-stage strategy, where filters of increasing complexity and computational cost are applied successfully. For example, in the case of adsorption, studies will focus first on pore space and accessible area (geometric descriptors), then identify among those best-performing candidates the ones suitable for adsorption based on Grand Canonical Monte Carlo simulations. A similar strategy was applied by Davies *et al.* for the screening of stoichiometric inorganic materials for water splitting, where low-computational cost filters based on electronegativity, electronic chemical potential and atomic solid-state energy [88].

To go beyond these methods and identify novel materials for targeted applications, there is thus a need to develop active methods for property prediction based on structure and chemical composition, bypassing quantum calculations and classical molecular simulations—at least during an initial high-throughput screening step. In order to develop such methods, databases are useful in two different ways: first, databases of physical and chemical properties are necessary in order to train, benchmark and validate the new prediction methods. Second, larger databases of hypothetical structures are needed as a basis for large-scale screening, once the property prediction methods are adequate.

With this goal in mind, machine learning appears as a powerful tool for predicting chemical and physical properties for large number of materials, i.e. at low computational cost. Neural networks—already presented in §2—are a class of machine learning algorithms, but many others

exist. Machine learning is the generic term used for algorithms that generate another algorithm, in order to progressively improve their performance for a task they have not been explicitly programmed to perform. In the most commonly used family of machine learning methods, called *supervised learning*, the algorithm generated is called the predictor. It takes a set of input descriptors, and maps them to the required output. This output is usually the numeric value of a physical property in our case, but it can also be the classification of the input in a given class. When using machine learning on chemical systems, the descriptors can take multiple forms: local descriptors such as atomic positions, bond length, angle or dihedral angles; global descriptors like mass density, largest included sphere in a porous framework or elastic properties; and topological descriptors such as ring size distribution. As we said, machine learning algorithms generate predictor algorithms from a set of reference input and output data. The idea is to train the machine learning algorithm on a subset of the data, and then test the generated predictor on the remaining part of the reference data. This allows us to evaluate the accuracy of the predictor. For more information on machine learning and its usage in molecular and materials science, we refer the interested reader to the very pedagogical review by Butler *et al.* [89].

Within the fields of physics and chemistry, machine learning has been applied to a large diversity of applications. On the computational side, research is ongoing on the use of machine learning to improve electronic structure calculations by bypassing the Kohn–Sham equations [90], developing machine-learned functionals [91] and creating adaptive basis sets [92]. Other applications in chemistry include the extraction of chemical data (structures, reactions, etc.) from published work [93], the prediction of novel synthetic pathways [94], the design of catalysts [95], etc. In 2016, Jong *et al.* used machine learning techniques to predict elastic properties (bulk and shear modulus) for inorganic compounds in order to accelerate material discovery and design [96]. However, few studies have focused so far on framework materials and their physical properties. Recently, Evans *et al.* [97] used a machine learning algorithm to predict elastic properties (such as the bulk modulus and shear modulus) of 590 448 hypothetical pure-silica zeolites, using an accurate training set of elastic properties determined with DFT calculations [98]. Evans combined the GBR (gradient boosting regressor) approach using regression trees and a set of local, structural and porosity-related descriptors, and their results highlighted several important correlations and trends in terms of stability for zeolitic structures. Romain Gaillac extended this to predict the auxeticity and the Poisson's ratio of more than 1000 zeolites [99]. These recent advances, combined with the availability of DFT-computed elastic tensors for a large number of inorganic materials within the Materials Project, create new opportunities for computationally assisted material discovery and design. We should also note here, for the sake of completeness, that *unsupervised* machine learning has also been applied to chemical questions: such techniques take a dataset as input and identify hidden structures in the data—e.g. clustering of data points or structures by similarity [100,101].

## 5. Conclusion

We have given here a short overview of the current state of methodologies for modelling framework materials at multiple scales and tried to highlight some of the common themes as well as differences between this rapidly expanding class of materials and other inorganic solids. It is clear from the examples listed that the diversity of modelling methods is also growing to match the rapid pace of experimental developments, and the increasing complexity of the systems and phenomena studied. However, while modelling strategies develop at all length and time scales, from the microscopic to the macroscopic, the links between these simulation scales are still rather *ad hoc*, and comprehensive, coherent multi-scale simulation strategies are still the exception, rather than the norm. Just as experimental and computational tools are complementary in providing a large variety of viewpoints on a given material, studies containing multiple simulations strategies at different scales are appearing, which provide a very deep understanding of the macroscopic properties of a material and its microscopic origins.

**Data accessibility.** Supporting data are available online in our data repository at <https://github.com/fxcoudert/citable-data>.

**Competing interests.** We declare we have no competing interests.

**Funding.** No funding has been received for this article.

**Acknowledgements.** A large part of the work reviewed here requires access of scientists in the field to large supercomputer centres. Although no original calculations were performed in the writing of this review, we acknowledge GENCI for high-performance computing CPU time allocations (grant no. A0050807069). We sincerely thank colleagues from Université de France (UNIV France) for their support. We thank Cory Simon and Martijn Zwijnenburg for their insightful feedback on a first version of this manuscript which appeared on the chemRxiv preprint server.

## References

1. Horike S, Shimomura S, Kitagawa S. 2009 Soft porous crystals. *Nat. Chem.* **1**, 695–704. (doi:10.1038/nchem.444)
2. Coudert FX. 2015 Responsive metal–organic frameworks and framework materials: under pressure, taking the heat, in the spotlight, with friends. *Chem. Mater.* **27**, 1905–1916. (doi:10.1021/acs.chemmater.5b00046)
3. Krause S *et al.* 2016 A pressure-amplifying framework material with negative gas adsorption transitions. *Nature* **532**, 348–352. (doi:10.1038/nature17430)
4. Cairns AB, Goodwin AL. 2015 Negative linear compressibility. *Phys. Chem. Chem. Phys.* **17**, 20 449–20 465. (doi:10.1039/C5CP00442J)
5. Spencer EC, Kiran MSRN, Li W, Ramamurty U, Ross NL, Cheetham AK. 2014 Pressure-induced bond rearrangement and reversible phase transformation in a metal-organic framework. *Angew. Chem. Int. Ed.* **53**, 5583–5586. (doi:10.1002/anie.v53.22)
6. Lyndon R, Konstas K, Ladewig BP, Southon PD, Kepert PCJ, Hill MR. 2013 Dynamic photo-switching in metal-organic frameworks as a route to low-energy carbon dioxide capture and release. *Angew. Chem. Int. Ed.* **52**, 3695–3698. (doi:10.1002/anie.201206359)
7. Lapidus SH, Halder GJ, Chupas PJ, Chapman KW. 2013 Exploiting high pressures to generate porosity, polymorphism, and lattice expansion in the nonporous molecular framework zn(cn). *J. Am. Chem. Soc.* **135**, 7621–7628. (doi:10.1021/ja4012707)
8. Cornell WD *et al.* 1995 A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.* **117**, 5179–5197. (doi:10.1021/ja00124a002)
9. Rappe AK, Casewit CJ, Colwell KS, Goddard WA, Skiff WM. 1992 Uff, a full periodic table force field for molecular mechanics and molecular dynamics simulations. *J. Am. Chem. Soc.* **114**, 10 024–10 035. (doi:10.1021/ja00051a040)
10. Greathouse JA, Ockwig NW, Criscenti LJ, Guilinger TR, Pohl P, Allendorf MD. 2010 Computational screening of metal–organic frameworks for large-molecule chemical sensing. *Phys. Chem. Chem. Phys.* **12**, 12621. (doi:10.1039/c0cp00092b)
11. Ryan P, Farha OK, Broadbelt LJ, Snurr RQ. 2010 Computational screening of metal-organic frameworks for xenon/krypton separation. *AIChE J.* **57**, 1759–1766. (doi:10.1002/aic.v57.7)
12. Colón YJ, Snurr RQ. 2014 High-throughput computational screening of metal–organic frameworks. *Chem. Soc. Rev.* **43**, 5735–5749. (doi:10.1039/C4CS00070F)
13. Chaplais G, Fraux G, Paillaud JL, Marichal C, Nouali H, Fuchs AH, Coudert FX, Patarin J. 2018 Impacts of the imidazolate linker substitution (CH<sub>3</sub>, cl or br) on the structural and adsorptive properties of ZIF-8. *J. Phys. Chem. C* **122**, 26 945–26 955. (doi:10.1021/acs.jpcc.8b08706)
14. Howe JD, Morelock CR, Jiao Y, Chapman KW, Walton KS, Sholl DS. 2016 Understanding structure, metal distribution, and water adsorption in mixed-metal MOF-74. *J. Phys. Chem. C* **121**, 627–635. (doi:10.1021/acs.jpcc.6b11719)
15. McGrath MJ, Siepmann JI, Kuo IFW, Mundy CJ, VandeVondele J, Hutter J, Mohamed F, Krack M. 2005 Isobaric-isothermal monte carlo simulations from first principles: application to liquid water at ambient conditions. *ChemPhysChem* **6**, 1894–1901. (doi:10.1002/(ISSN)1439-7641)
16. Leiding J, Coe JD. 2014 An efficient approach to *ab initio* monte carlo simulation. *J. Chem. Phys.* **140**, 034106. (doi:10.1063/1.4855755)

17. Falcioni M, Deem MW. 1999 A biased monte carlo scheme for zeolite structure solution. *J. Chem. Phys.* **110**, 1754–1766. (doi:10.1063/1.477812)
18. Maurin G, Senet P, Devautour S, Gaveau P, Henn F, Van Doren VE, Giuntini JC. 2001 Combining the Monte Carlo technique with  $^{29}\text{Si}$  nmr spectroscopy: simulations of cation locations in zeolites with various si/al ratios. *J. Phys. Chem. B* **105**, 9157–9161. (doi:10.1021/jp011789i)
19. Coudert FX, Fuchs AH. 2016 Computational characterization and prediction of metal-organic framework properties. *Coord. Chem. Rev.* **307**, 211–236. (doi:10.1016/j.ccr.2015.08.001)
20. Paddison JAM, Agrestini S, Lees MR, Fleck CL, Deen PP, Goodwin AL, Stewart JR, Petrenko OA. 2014 Spin correlations in  $\text{Ca}_3\text{Co}_2\text{O}_6$ : polarized-neutron diffraction and Monte Carlo study. *Phys. Rev. B* **90**, 31. (doi:10.1103/PhysRevB.90.014411)
21. Sapnik AF, Geddes HS, Reynolds EM, Yeung HHM, Goodwin AL. 2018 Compositional inhomogeneity and tuneable thermal expansion in mixed-metal zif-8 analogues. *Chem. Commun.* **54**, 9651–9654. (doi:10.1039/C8CC04172E)
22. Cairns AB, Cliffe MJ, Paddison JAM, Daisenberger D, Tucker MG, Coudert FX, Goodwin AL. 2016 Encoding complexity within supramolecular analogues of frustrated magnets. *Nat. Chem.* **8**, 442–447. (doi:10.1038/nchem.2462)
23. Bureekaew S, Amirjalayer S, Tafipolsky M, Spickermann C, Roy TK, Schmid R. 2013 MOF-FF - a flexible first-principles derived force field for metal-organic frameworks. *Phys. Status Solidi B* **250**, 1128–1141. (doi:10.1002/pssb.v250.6)
24. Vanduyfhuys L, Vandenbrande S, Verstraelen T, Schmid R, Waroquier M, Speybroeck VV. 2015 QuickFF: a program for a quick and easy derivation of force fields for metal-organic frameworks fromab initioinput. *J. Comput. Chem.* **36**, 1015–1027. (doi:10.1002/jcc.v36.13)
25. Impeng S, Cedeno R, Dürholt JP, Schmid R, Bureekaew S. 2018 Computational structure prediction of (4, 4)-connected copper paddle-wheel-based MOFs: influence of ligand functionalization on the topological preference. *Cryst. Growth Des.* **18**, 2699–2706. (doi:10.1021/acs.cgd.8b00238)
26. Vanduyfhuys L, Vandenbrande S, Wieme J, Waroquier M, Verstraelen T, Van Speybroeck V. 2018 Extension of the quickff force field protocol for an improved accuracy of structural, vibrational, mechanical and thermal properties of metal-organic frameworks. *J. Comput. Chem.* **39**, 999–1011. (doi:10.1002/jcc.v39.16)
27. Hornik K, Stinchcombe M, White H. 1989 Multilayer feedforward networks are universal approximators. *Neural Netw.* **2**, 359–366. (doi:10.1016/0893-6080(89)90020-8)
28. Behler J, Parrinello M. 2007 Generalized neural-network representation of high-dimensional potential-energy surfaces. *Phys. Rev. Lett.* **98**, 146401. (doi:10.1103/PhysRevLett.98.146401)
29. Smith JS, Isayev O, Roitberg AE. 2017 ANI-1: an extensible neural network potential with DFT accuracy at force field computational cost. *Chem. Sci.* **8**, 3192–3203. (doi:10.1039/C6SC05720A)
30. Cho KH, No KT, Scheraga HA. 2002 A polarizable force field for water using an artificial neural network. *J. Mol. Struct.* **641**, 77–91. (doi:10.1016/S0022-2860(02)00299-5)
31. Morawietz T, Behler J. 2013 A density-functional theory-based neural network potential for water clusters including van der waals corrections. *J. Phys. Chem. A* **117**, 7356–7366. (doi:10.1021/jp401225b)
32. Hellström M, Behler J. 2018 Neural network potentials in materials modeling. In *Handbook of materials modeling*, pp. 1–20. Berlin: Springer International Publishing.
33. Deringer VL, Csányi G. 2017 Machine learning based interatomic potential for amorphous carbon. *Phys. Rev. B* **95**, 094203. (doi:10.1103/PhysRevB.95.094203)
34. Deringer VL, Bernstein N, Bartók AP, Cliffe MJ, Kerber RN, Marbella LE, Grey CP, Elliott SR, Csányi G. 2018 Realistic atomistic structure of amorphous silicon from machine-learning-driven molecular dynamics. *J. Phys. Chem. Lett.* **9**, 2879–2885. (doi:10.1021/acs.jpcllett.8b00902)
35. Bennett TD *et al.* 2016 Melt-quenched glasses of metal-organic frameworks. *J. Am. Chem. Soc.* **138**, 3484–3492. (doi:10.1021/jacs.5b13220)
36. Zhao Y, Lee SY, Becknell N, Yaghi OM, Angell CA. 2016 Nanoporous transparent mof glasses with accessible internal surface. *J. Am. Chem. Soc.* **138**, 10818–10821. (doi:10.1021/jacs.6b07078)

37. Zhou C *et al.* 2018 Metal-organic framework glasses with permanent accessible porosity. *Nat. Commun.* **9**, 19. (doi:10.1038/s41586-018-0513-4)
38. Gaillac R, Pullumbi P, Beyer KA, Chapman KW, Keen DA, Bennett TD, Coudert FX. 2017 Liquid metal-organic frameworks. *Nat. Mater.* **16**, 1149–1154. (doi:10.1038/nmat4998)
39. Cheetham AK, Bennett TD, Coudert FX, Goodwin AL. 2016 Defects and disorder in metal organic frameworks. *Dalton Trans.* **45**, 4113–4126. (doi:10.1039/C5DT04392A)
40. Ginhoven RMV, Jónsson H, Corrales LR. 2005 Silica glass structure generation for ab initio calculations using small samples of amorphous silica. *Phys. Rev. B* **71**, 024208. (doi:10.1103/PhysRevB.71.024208)
41. Dürholt JP, Galvelis R, Schmid R. 2016 Coarse graining of force fields for metal-organic frameworks. *Dalton Trans.* **45**, 4370–4379. (doi:10.1039/C5DT03865K)
42. Vanson JM, Coudert FX, Rotenberg B, Levesque M, Tardivat C, Klotz M, Boutin A. 2015 Unexpected coupling between flow and adsorption in porous media. *Soft Matter* **11**, 6125–6133. (doi:10.1039/C5SM01348H)
43. Evans JD, Coudert FX. 2017 Macroscopic simulation of deformation in soft microporous composites. *J. Phys. Chem. Lett.* **8**, 1578–1584. (doi:10.1021/acs.jpcllett.7b00397)
44. Vanduyfhuys L, Verstraelen T, Vandichel M, Waroquier M, Van Speybroeck V. 2012 *Ab initio* parametrized force field for the flexible metal-organic framework mil-53(al). *J. Chem. Theory Comput.* **8**, 3217–3231. (doi:10.1021/ct300172m)
45. Coudert FX, Fuchs AH, Neimark AV. 2016 Adsorption deformation of microporous composites. *Dalton Trans.* **45**, 4136–4140. (doi:10.1039/C5DT03978A)
46. Runge E, Gross EKV. 1984 Density-functional theory for time-dependent systems. *Phys. Rev. Lett.* **52**, 997–1000. (doi:10.1103/PhysRevLett.52.997)
47. Casida ME. 1995 Time-dependent density functional response theory for molecules. In *Recent advances in density functional methods*, pp. 155–192. World Scientific.
48. Wilbraham L, Coudert FX, Ciofini I. 2016 Modelling photophysical properties of metal-organic frameworks: a density functional theory based approach. *Phys. Chem. Chem. Phys.* **18**, 25 176–25 182. (doi:10.1039/C6CP04056J)
49. Wu XP, Gagliardi L, Truhlar DG. 2018 Cerium metal-organic framework for photocatalysis. *J. Am. Chem. Soc.* **140**, 7904–7912. (doi:10.1021/jacs.8b03613)
50. Butler KT, Hendon CH, Walsh A. 2014 Electronic chemical potentials of porous metal-organic frameworks. *J. Am. Chem. Soc.* **136**, 2703–2706. (doi:10.1021/ja4110073)
51. Grau-Crespo R, Aziz A, Collins AW, Crespo-Otero R, Hernández NC, Rodríguez-Albelo LM, Ruiz-Salvador AR, Calero S, Hamad S. 2016 Modelling a linker mix-and-match approach for controlling the optical excitation gaps and band alignment of zeolitic imidazolate frameworks. *Angew. Chem.* **128**, 16 246–16 250. (doi:10.1002/ange.201609439)
52. Baerlocher C, McCusker LB, Olson D. 2007 *Atlas of zeolite framework types*, 6th edn. Amsterdam, The Netherlands: Elsevier.
53. Speybroeck VV, Hemelsoet K, Joos L, Waroquier M, Bell RG, Catlow CRA. 2015 Advances in theory and their application within the field of zeolite chemistry. *Chem. Soc. Rev.* **44**, 7044–7111. (doi:10.1039/C5CS00029G)
54. Chapman KW, Chupas PJ, Nenoff TM. 2010 Radioactive iodine capture in silver-containing mordenites through nanoscale silver iodide formation. *J. Am. Chem. Soc.* **132**, 8897–8899. (doi:10.1021/ja103110y)
55. Bučko T, Benco L, Hafner J, Ángyá JG. 2011 Monomolecular cracking of propane over acidic chabazite: an ab initio molecular dynamics and transition path sampling study. *J. Catal.* **279**, 220–228. (doi:10.1016/j.jcat.2011.01.022)
56. Nenoff TM, Rodríguez MA, Soelberg NR, Chapman KW. 2014 Silver-mordenite for radiologic gas capture from complex streams: dual catalytic CH<sub>3</sub>i decomposition and i confinement. *Micropor. Mesopor. Mat.* **200**, 297–303. (doi:10.1016/j.micromeso.2014.04.041)
57. Bushuev YG, Sastre G. 2010 Feasibility of pure silica zeolites. *J. Phys. Chem. C* **114**, 19 157–19 168. (doi:10.1021/jp107296e)
58. Blatov VA, Ilyushin GD, Proserpio DM. 2013 The zeolite conundrum: why are there so many hypothetical zeolites and so few observed? A possible answer from the zeolite-type frameworks perceived as packings of tiles. *Chem. Mater.* **25**, 412–424. (doi:10.1021/cm303528u)

59. Li Y, Yu J, Liu D, Yan W, Xu R, Xu Y. 2003 Design of zeolite frameworks with defined pore geometry through constrained assembly of atoms. *Chem. Mater.* **15**, 2780–2785. (doi:10.1021/cm0213826)
60. Yu J, Xu R. 2010 Rational approaches toward the design and synthesis of zeolitic inorganic open-framework materials. *Acc. Chem. Res.* **43**, 1195–1204. (doi:10.1021/ar900293m)
61. Li Y, Yu J. 2014 New stories of zeolite structures: their descriptions, determinations, predictions, and evaluations. *Chem. Rev.* **114**, 7268–7316. (doi:10.1021/cr500010r)
62. Li J, Corma A, Yu J. 2015 Synthesis of new zeolite structures. *Chem. Soc. Rev.* **44**, 7112–7127. (doi:10.1039/C5CS00023H)
63. Li Y, Yu J, Xu R. 2012 FraGen: a computer program for real-space structure solution of extended inorganic frameworks. *J. Appl. Cryst.* **45**, 855–861. (doi:10.1107/S002188981201878X)
64. Li Y, Li X, Liu J, Duan F, Yu J. 2015 In silico prediction and screening of modular crystal structures via a high-throughput genomic approach. *Nature Commun.* **6**, 011002. (doi:10.1038/ncomms9328)
65. Treacy MMJ, Randall KH, Rao S, Perry JA, Chadi DJ. 1997 Enumeration of periodic tetrahedral frameworks. *Z. Kristallogr. Cryst. Mater.* **212**, 50. (doi:10.1524/zkri.1997.212.11.768)
66. Treacy M, Rivin I, Balkovsky E, Randall K, Foster M. 2004 Enumeration of periodic tetrahedral frameworks. ii. polynodal graphs. *Micropor. Mesopor. Mater.* **74**, 121–132. (doi:10.1016/j.micromeso.2004.06.013)
67. Zwijnenburg MA, Bell RG. 2008 Absence of limitations on the framework density and pore size of high-silica zeolites. *Chem. Mater.* **20**, 3008–3014. (doi:10.1021/cm702175q)
68. Zwijnenburg MA, Jelfs KE, Bromley ST. 2010 An extensive theoretical survey of low-density allotropy in silicon. *Phys. Chem. Chem. Phys.* **12**, 8505. (doi:10.1039/c004375c)
69. Zwijnenburg MA, Illas F, Bromley ST. 2010 Apparent scarcity of low-density polymorphs of inorganic solids. *Phys. Rev. Lett.* **104**, 768. (doi:10.1103/PhysRevLett.104.175503)
70. Earl DJ, Deem MW. 2006 Toward a database of hypothetical zeolite structures. *Ind. Eng. Chem. Res.* **45**, 5449–5454. (doi:10.1021/ie0510728)
71. Pophale R, Cheeseman PA, Deem MW. 2011 A database of new zeolite-like materials. *Phys. Chem. Chem. Phys.* **13**, 12407. (doi:10.1039/c0cp02255a)
72. Yu J, Xu R. 2006 Insight into the construction of open-framework aluminophosphates. *Chem. Soc. Rev.* **35**, 593. (doi:10.1039/b505856m)
73. Moghadam PZ, Li A, Wiggin SB, Tao A, Maloney AGP, Wood PA, Ward SC, Fairen-Jimenez D. 2017 Development of a Cambridge Structural Database subset: a collection of metal–organic frameworks for past, present, and future. *Chem. Mater.* **29**, 2618–2625. (doi:10.1021/acs.chemmater.7b00441)
74. Watanabe T, Sholl DS. 2012 Accelerating applications of metal–organic frameworks for gas adsorption and separation by computational screening of materials. *Langmuir* **28**, 14 114–14 128. (doi:10.1021/la301915s)
75. Goldsmith J, Wong-Foy AG, Cafarella MJ, Siegel DJ. 2013 Theoretical limits of hydrogen storage in metal–organic frameworks: opportunities and trade-offs. *Chem. Mater.* **25**, 3373–3382. (doi:10.1021/cm401978e)
76. Chung YG *et al.* 2014 Computation-ready, experimental metal–organic frameworks: a tool to enable high-throughput screening of nanoporous crystals. *Chem. Mater.* **26**, 6185–6192. (doi:10.1021/cm502594j)
77. Barthel S, Alexandrov EV, Proserpio DM, Smit B. 2018 Distinguishing metal–organic frameworks. *Cryst. Growth Des.* **18**, 1738–1747. (doi:10.1021/acs.cgd.7b01663)
78. Park S, Kim B, Choi S, Boyd PG, Smit B, Kim J. 2018 Text mining metal–organic framework papers. *J. Chem. Inf. Model.* **58**, 244–251. (doi:10.1021/acs.jcim.7b00608)
79. Nazarian D, Camp JS, Sholl DS. 2016 A comprehensive set of high-quality point charges for simulations of metal–organic frameworks. *Chem. Mater.* **28**, 785–793. (doi:10.1021/acs.chemmater.5b03836)
80. Wilmer CE, Leaf M, Lee CY, Farha OK, Hauser BG, Hupp JT, Snurr RQ. 2011 Large-scale screening of hypothetical metal–organic frameworks. *Nat. Chem.* **4**, 83–89. (doi:10.1038/nchem.1192)



81. Gomez DA, Toda J, Sastre G. 2014 Screening of hypothetical metal–organic frameworks for h2 storage. *Phys. Chem. Chem. Phys.* **16**, 19 001–19 010. (doi:10.1039/C4CP01848F)
82. He Y, Zhou W, Qian G, Chen B. 2014 Methane storage in metal–organic frameworks. *Chem. Soc. Rev.* **43**, 5657–5678. (doi:10.1039/C4CS00032C)
83. Burtch NC, Jasuja H, Walton KS. 2014 Water stability and adsorption in metal–organic frameworks. *Chem. Rev.* **114**, 10 575–10 612. (doi:10.1021/cr5002589)
84. Zakutayev A, Wunder N, Schwarting M, Perkins JD, White R, Munch K, Tumas W, Phillips C. 2018 An open experimental database for exploring inorganic materials. *Sci. Data* **5**, 180053. (doi:10.1038/sdata.2018.53)
85. Jain A *et al.* 2013 Commentary: the materials project: a materials genome approach to accelerating materials innovation. *APL Mater.* **1**, 011002. (doi:10.1063/1.4812323)
86. Gaillac R, Pullumbi P, Coudert FX. 2016 Elate: an open-source online application for analysis and visualization of elastic tensors. *J. Phys. Condens. Matter* **28**, 275201. (doi:10.1088/0953-8984/28/27/275201)
87. de Jong M *et al.* 2015 Charting the complete elastic properties of inorganic crystalline compounds. *Sci. Data* **2**, 150009. (doi:10.1038/sdata.2015.9)
88. Davies DW, Butler KT, Jackson AJ, Morris A, Frost JM, Skelton JM, Walsh A. 2016 Computational screening of all stoichiometric inorganic materials. *Chemistry* **1**, 617–627. (doi:10.1016/j.chempr.2016.09.010)
89. Butler KT, Davies DW, Cartwright H, Isayev O, Walsh A. 2018 Machine learning for molecular and materials science. *Nature* **559**, 547–555. (doi:10.1038/s41586-018-0337-2)
90. Brockherde F, Vogt L, Li L, Tuckerman ME, Burke K, Müller KR. 2017 Bypassing the kohn-sham equations with machine learning. *Nat. Commun.* **8**, A1133. (doi:10.1038/s41467-017-00839-3)
91. Hollingsworth J, Baker TE, Burke K. 2018 Can exact conditions improve machine-learned density functionals? *J. Chem. Phys.* **148**, 241743. (doi:10.1063/1.5025668)
92. Schütt O, VandeVondele J. 2018 Machine learning adaptive basis sets for efficient large scale density functional theory simulation. *J. Chem. Theory Comput.* **14**, 4168–4175. (doi:10.1021/acs.jctc.8b00378)
93. Kim E, Huang K, Saunders A, McCallum A, Ceder G, Olivetti E. 2017 Materials synthesis insights from scientific literature via text extraction and machine learning. *Chem. Mater.* **29**, 9436–9444. (doi:10.1021/acs.chemmater.7b03500)
94. Gómez-Bombarelli R *et al.* 2018 Automatic chemical design using a data-driven continuous representation of molecules. *ACS Cent. Sci.* **4**, 268–276. (doi:10.1021/acscentsci.7b00572)
95. Goldsmith BR, Esterhuizen J, Liu JX, Bartel CJ, Sutton C. 2018 Machine learning for heterogeneous catalyst design and discovery. *AIChE J.* **64**, 2311–2323. (doi:10.1002/aic.v64.7)
96. de Jong M, Chen W, Notestine R, Persson K, Ceder G, Jain A, Asta M, Gamst A. 2016 A statistical learning framework for materials science: application to elastic moduli of k-nary inorganic polycrystalline compounds. *Sci. Rep.* **6**, 34256. (doi:10.1038/srep34256)
97. Evans JD, Coudert FX. 2017 Predicting the mechanical properties of zeolite frameworks by machine learning. *Chem. Mater.* **29**, 7833–7839. (doi:10.1021/acs.chemmater.7b02532)
98. Coudert FX. 2013 Systematic investigation of the mechanical properties of pure silica zeolites: stiffness, anisotropy, and negative linear compressibility. *Phys. Chem. Chem. Phys.* **15**, 16012. (doi:10.1039/c3cp51817e)
99. Gaillac R. 2018 Molecular modeling of physico-chemical properties in microporous solids. PhD thesis, Chimie ParisTech, PSL University. See <https://tel.archives-ouvertes.fr/tel-01820463>.
100. Gómez-Bombarelli R *et al.* 2018 Automatic chemical design using a data-driven continuous representation of molecules. *ACS Cent. Sci.* **4**, 268–276. (doi:10.1021/acscentsci.7b00572)
101. Sturluson A, Huynh MT, York AHP, Simon CM. 2018 Eigencages: learning a latent space of porous cage molecules. *ACS Cent. Sci.* **4**, 1663–1676. (doi:10.1021/acscentsci.8b00638)