System of sequences in multivariate reticular structures

Stefano Canossa, Zhe Ji, Cornelius Gropp, Zichao Rong, Evelyn Ploetz, Stefan Wuttke and Omar M. Yaghi

Abstract | When variance is encountered in reticular structures such as metal–organic frameworks, it occurs as multiple kinds of functionalities bound and distributed in aperiodic lattices. In these multivariate systems, unknown spatial arrangements of functionalities create properties that go beyond those of their simple sum. It is therefore essential that we learn how to recognize, study and use these arrangements, and for this purpose we propose using the concept of sequences. Accordingly, we propose a classification system, outline a method for characterizing sequences and describe the application to self-propelled reticular machines. On a fundamental level, this contribution transforms the chemist's thinking from the usual 'make, characterize, use' protocol of doing chemistry to a discovery routine based on finding correlations between input synthesis parameters and output performance. This approach provides an alternative and widely accessible pathway for determining the properties that the system of sequences confers on multivariate reticular structures.

Chemists consider two molecules to be the same when they meet three basic criteria: being described by the same chemical formula; having the same structure (pattern of connectivity and geometry); and sharing the same spatial atomic arrangement (FIG. 1a). These three criteria also apply to crystalline, extended structures such as compositionally periodic metal–organic frameworks (MOFs) or covalent organic frameworks (COFs)[1,2], which can be reasonably considered different forms if they have the same formula and structure but different spatial atomic arrangements. In introducing the topic of this contribution, we present a fundamental question: are there chemical constructs for which the three 'sameness' criteria are insufficient for their distinction?

Considering the gamut of substances in existence, we believe that DNA is a prominent example of such a construct (FIG. 1b). Indeed, two DNA strands having the same chemical formula, sugar–phosphate backbone structure and handedness, but different sequences of nucleotides, would not be considered the same compound. The concept of sequences in DNA can be likened to that of spatial atomic arrangement, but where the role of atoms is taken over by molecules. However, when describing macromolecules, the term 'sequences' is preferred because of its practicality and clarity, leaving 'spatial atomic arrangement' to the structural description of molecular fragments. In fact, a sequence of well-defined molecules can be also described in terms of sequence of their atoms in space, but this is much less convenient, as the individual sequences of atoms within each given molecule of the same type do not change. Therefore, once the molecules involved are known, describing their supramolecular sequences in terms of each one of their atoms does not add important structural information but makes the description considerably more complex.

Other examples of constructs that can be described by sequences are multivariate forms of reticular structures such as MOFs and COFs, wherein various chemical species occupy equivalent positions across different parts of the framework without compromising its long-range architecture[3] (FIG. 1c,d). These species are spatially arranged in a multivariate manner — that is, creating a series of different combinations throughout the lattice, thus not forming a compositionally periodic crystal in which the chemical composition of all unit cells is identical. Because these species are strongly linked to the periodic backbone of the framework, as nucleotides are to the polyphosphate backbone of DNA, it is instructive to think of their non-periodic arrangements as sequences.

It is the presence of a periodic backbone — the underlying scheme of repetition — that fundamentally separates the concept of 'sequence' from the more general one of 'arrangement'. Sequences in MOFs can adopt random arrangements[4,5]; however, conditions can be created for achieving non-random sequences containing short and long repeats[5–9]. We have in the past referred to this complexity as 'heterogeneity within order' to emphasize the multivariate nature of sequences, the known distances between their components and the underlying order of the backbone[10]. A misconception arising from this mixture of structural aspects is to generally equate multivariate frameworks to solid solutions. They both consist of crystalline materials having compositional variation across their structure. However, multivariate frameworks differ in two important ways. First, the spacing of functionalities is fixed along the backbone and negligibly affected by the multivariation; second, the strong bonds in reticular structures keep the various functionalities in place, so that these cannot be easily swapped to make a different multivariate compound. On the other hand, like solid solutions, some of these multivariate frameworks — being endowed with non-random sequences — give rise to properties where the whole performs better than the sum of the parts[11,12]. The emergence of properties with framework multivariation arises from the relationships that sequences may have with each other and with the periodic architecture of the framework. At the fundamental core of these relationships are chemical and physical interactions, functional synergies, and their relative positions, which altogether constitute the complex 'organism' that is the system of sequences. Based on these

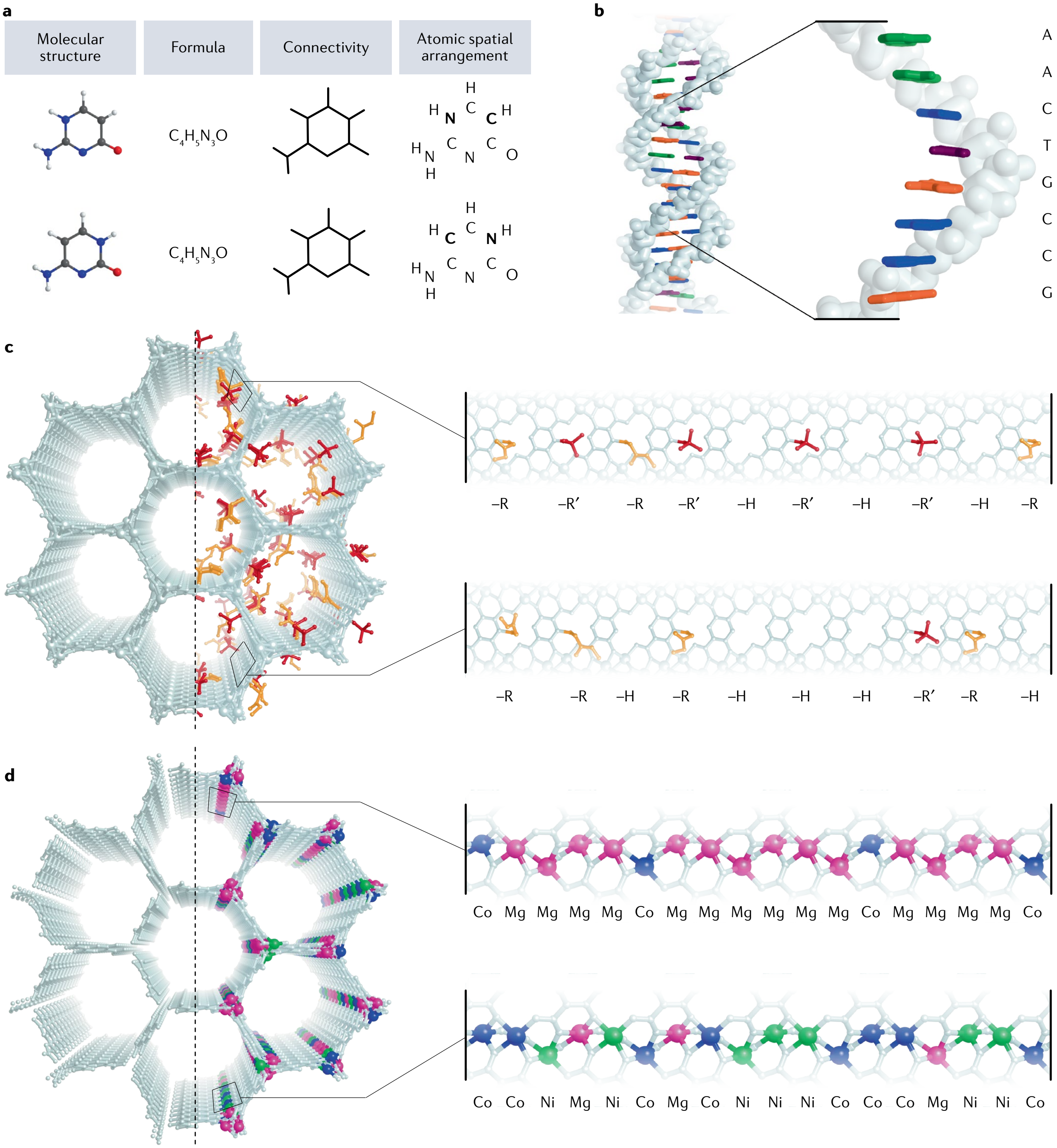


Fig. 1 | **Atomic spatial arrangement and sequences in classifying molecules and extended molecular networks. a** | Cytosine (top) and isocytosine (bottom), which have the same chemical formula and connectivity, can be distinguished based on their different alternation of atoms in the structure. **b–d** | Similarly, it is useful to distinguish periodic backbone (light grey) and sequences of functionalities (coloured) in aperiodic compounds such as DNA (**b**) and multivariate metal–organic frameworks (MOFs, such as the MOF-74 shown here) with mixed functional groups (**c**) or metal species (**d**). The MOFs are shown highlighting the only periodic fraction (left side of the dashed line), the complete framework, and two of the various sequences that populate the lattice. The letters C, G, T and A represent, respectively, the nucleotides cytosine, guanine, thymine and adenine; R and R′ represent different terminal groups.

considerations, we propose to introduce the concept of sequences as a fourth sameness criterion in chemistry for the importance that this aspect has in defining the structural identity and distinctive properties of multivariate crystals.

In this contribution, we provide a unifying groundwork for understanding the system of sequences and thereby the unique nature of multivariate frameworks, and for adopting a new sequences-oriented mindset to their analysis and design.

We start by proposing a set of definitions for the essential aspects of multivariation, and we address crucial terminological challenges posed by its presence in the context of solid-state and materials chemistry. We make the point that, to advance the synthesis of these materials, we must be able to characterize them and, in the absence of direct analysis of the multivariation, its properties can be used as a litmus test to indirectly differentiate between frameworks containing detectably different sequences. Accordingly, we introduce 'molecular sorting' as an approach to assess multivariation by exploiting how sequences affect the host–guest chemistry across different crystal directions. Finally, we advance the concept of using direction-specific sequences by presenting a potential application involving the coupling of molecular machinery and sequences within multivariate frameworks to give 'drivable' porous crystals.

Fundamentals of multivariation

Reticular structures such as MOFs and COFs are 'multivariate' when there is a compositional variability across unit cells, making the average unit cell not representative of each cell in the crystal. In particular, the prefix 'multi-' in multivariate refers to the wide variety of chemical or physical landscapes that even a single type of variation can create. It is important to point out that the variability of a multivariate framework cannot result in an entirely periodic structure, which would instead be categorized correctly as a conventional MOF or COF.

A compositionally periodic structure can also be defined as a multivariate framework if non-chemical but physical variations are considered, such as lattice distortions or dynamics. Nonetheless, in most cases, the compositional variation of the framework, which includes not only different chemical species but also their absence (vacancy defects), is the main origin of all the other variations that are found in multivariate crystals (upper part of the figure in BOX 1). These variations include differences in the internal surface chemistry[13,14], as well as diverse pore size and shape[15–17], flexibility[14,18] and thermal conductivity[19], to mention a few. In the following analysis, we outline the general concepts for describing and classifying multivariate frameworks consistently. Although we deliberately focus on MOFs because of their stronger presence in current literature, the introduced concepts are equally valid for multivariate COFs or other framework materials.

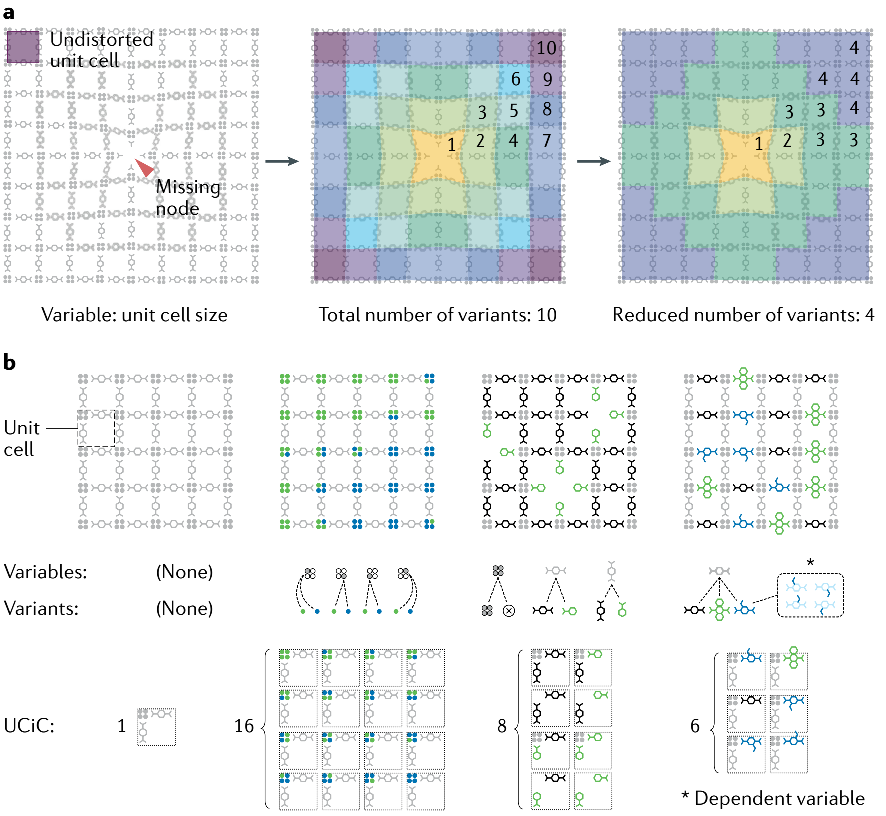
Box 1 | Circumventing pitfalls in the definition of framework variables

Reduction of framework variants

Although the arbitrary definition of framework variables is often straightforward, there might be cases where specific choices are more advisable than others. For instance, the multiplicity of a variable can be difficult to assess when a defect-induced lattice deformation creates a large number of differently distorted unit cells and the variable of choice is 'unit cell size'. In such cases, it is convenient to change the definition of the variable to 'unit cell size interval', thus obtaining fewer variants and simplifying the description. The number of framework variants is conveniently reduced by clustering unit cell sizes into size intervals, thereby simplifying the description of multivariation (see the figure, part **a**).

Assessing framework variables, variants and UCiC

Another preferable choice is to avoid defining variables that are not independent of one another. For instance, if a metal–organic framework has three alternative types of linkers and only one of these can be found in four different orientations, the variable 'linker orientation' exists only for one specific state (variant) of the variable 'linker type'. When such variable–variable dependency occurs, the formula provided in the text for the unit cell information capacity (UCiC) parameter is not valid. Therefore, it is advantageous to adopt a definition that combines both variables into a single one, for example where the linker variants are distinguished based on both their type and orientation. Some examples of multivariation parameter recognition in three multivariate forms of the same framework structure are shown (see the figure, part **b**). From left to right, reference periodic framework and three multivariate forms: with mixed metals, mixed linkers/ligands and defects, and mixed linkers. The mixed linker case shows the presence of linker orientation as an additional variable, which only occurs when a specific linker is present.



Framework variables and variants. A fundamental concept for understanding multivariate frameworks is the 'framework variable': an unambiguously identified characteristic, which can be present in different states, called variants, whose number defines the multiplicity of the variable. In the example of a Fe/Al mixed-metal MOF, Fe and Al are the possible variants of the binary variable 'type of metal'. The achievable combinations of variants break the monotonicity of what in periodic frameworks is the simplest, smallest repeating unit of the crystal, the primitive unit cell, splitting its unique state into multiple alternatives distributed across the lattice. Depending on the research purpose, a certain number of relevant variables are identified. These are not necessarily all the possible variables present in the material, but only those that will provide a clear, purposely limited view of the material under study. The open-minded researcher remains conscious of the other possible variables.

Once the variables that can be manipulated synthetically and their multiplicities (number of possible variants for the variable) are identified, it is possible to derive how many different versions of a single primitive unit cell can be obtained. For this purpose, we define this parameter 'unit cell information capacity' (UCiC), which provides a measure of the multivariate MOF's capability for storing tunable chemical or physical information. For a material having a number n of tunable variables, UCiC can be calculated by using combinatorics, as:

$$\mathrm{UCiC} = m_1 \times m_2 \times \ldots \times m_n = \prod_{i=1}^{n} m_i,$$

where m_i is the multiplicity of the ith variable.

The arbitrary definition of framework variables and variants uniquely defines the value of UCiC. Therefore, the identification of these aspects should allow a description of the multivariation that is as simple as possible and devoid of ambiguities (BOX 1).

As we explore the forms of chemical information that can be encoded into multivariate frameworks, it is important to clarify what 'information' means in this context. At the most basic level, it refers to structural information, meaning the arrangement of chemically distinct species in space. However, structures are important because of their associated behaviour, since no pair of different structures can be found that behave identically in absolute terms. We conclude that the structure constitutes the atomic code whereby properties are written, similar to a series of symbols, numbers and letters encoding a function in a computer program. In this way, the kind of information that is ultimately encoded in multivariate frameworks by tailoring their complex structure is their functional behaviour. This general definition of information differs from its rigorous mathematical meaning in Shannon information theory and its recent elegant application to crystal structure complexity[20,21]. Most importantly, the latter identifies the information content of average unit cells as a function of their atomic structure and symmetry, whereas the definition of UCiC considers discrete (not averaged) unit cells of a crystal as distinct entities, of the same or different type in their succession across the lattice. In this view, each unit cell occupying a specific position in the crystal represents a specific bit of information within sequences formed by a series of unit cells.

Box 2 | Non-ideal sequences

The classification we provided should be considered mostly ideal, as obtaining a multivariate framework that contains precise repeating sequences with no outliers at all is unrealistic in most cases. For this reason, it is important to introduce the concept of 'error' in a sequence, so that deviations that are deemed negligible do not impede the description of the material's multivariation. Indeed, the presence of sparse violations in sequences can have negligible effects on the properties of interest that are conferred on the material, akin to the presence of minor mutations in DNA. Therefore, the expected properties computed by theoretical models considering the ideal structure may approach, to a large extent, those measured on the real material. Similarly, the concept of 'approximate sequence' is useful to describe the averaged situation of a polycrystalline powder that might feature nearly — yet not — identical sequences both within and among different crystals.

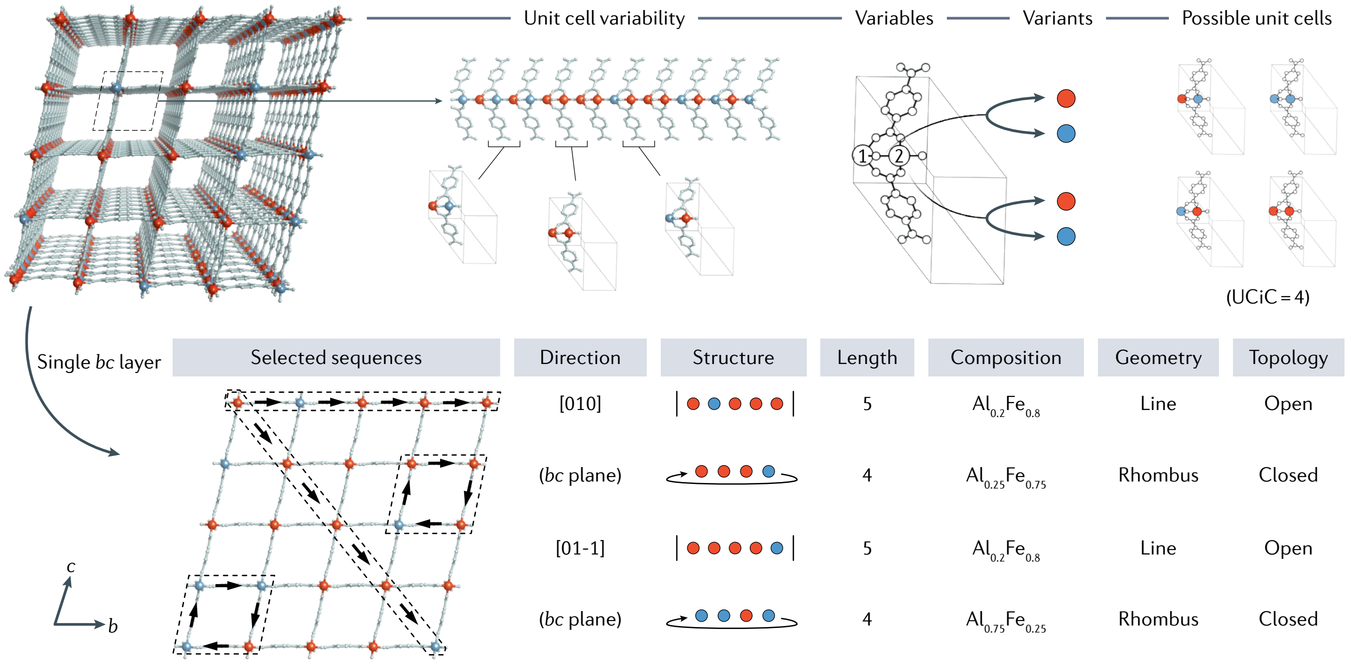
From variants to sequences. The uniqueness of the structure–property relationship would not be fully exploited if the variants of a unit cell were used to build frameworks one at a time, so that different frameworks each had only one of these unit cell types. Multivariate MOFs or COFs instead combine several variants, and the resulting properties are therefore defined not only by the number of variants but also by their spatial distribution in precise sequences. For this reason, the concept of sequence is at the heart of multivariate reticular chemistry and underlies its scope, methods and vast potential.

A sequence is a succession of variants — either identical or different — across the framework. Sequences can be characterized based on six main descriptors: length (that is, number of its constituent variants), structure, chemical composition, crystallographic direction, geometrical shape, and topology (most importantly, closed or open, depending on whether in a sequence the last variant precedes the first one). Because, as discussed, the definition of variants and variables is arbitrary, the identification of their sequences in a given multivariate framework is consequently affected, and in some cases the discrimination of sequences based on their distinguishing features requires some degree of flexibility (BOX 2). An overview of the analysis of a multivariate framework, including the identification and classification of some of its sequences, is provided in FIG. 2.

In the present state of reticular chemistry, the classification and understanding of multivariate frameworks is substantially hindered by the challenge of acquiring accurate structural information on framework multivariation and classifying materials based on their distinctive sequences (BOX 3). Given this difficulty, one might wonder whether a humbler trial-and-error strategy might be a more viable choice for their synthesis than a rational approach, until better analytical means are at hand. Although trial-and-error is arguably valid, a more precisely oriented method to create and control multivariation in MOFs and COFs can be conceived from focusing on their properties.

Function-based reticular design

The 'characterization problem'. Multivariate frameworks vastly expand the scope of chemistry by introducing heterogeneity and aperiodicity (short-range order of sequences) into otherwise homogeneous and periodic structures. Here, the UCiC gives a measure of the extent to which information, in the form of chemical and physical variability, can be stored in a single structure[4,22] (BOX 4). However, the mere possibility of storing vast amounts of information in one structure leads to the question: why introduce inhomogeneity into a periodic structure at the cost of losing the predictability and characterizability of the product? The reason is that this compositional variability provides an additional design handle to create materials composed of virtually countless combinations of different components that could work together cooperatively, outperforming — as an ensemble — their homogeneous and periodic counterparts[23]. Introducing such compositional variability into crystalline compounds, however, is accompanied by new synthetic and analytical challenges to the field. The latter represents what we recognize as the 'characterization problem': how can we distinguish non-random sequences from statistical disorder and different sequences from one another? This is an experimentally non-trivial task, largely because common crystallographic techniques for structural analysis are optimized for providing only average structural information, which is not sensitive to the local sequence of components in a framework[24].



Direction	Structure	Length	Composition	Geometry	Topology
[010]	| 🔴 🔵 🔴 🔴 🔴 |	5	$Al_{0.2}Fe_{0.8}$	Line	Open
(*bc* plane)	↻ 🔴 🔴 🔴 🔵	4	$Al_{0.25}Fe_{0.75}$	Rhombus	Closed
[01-1]	| 🔴 🔴 🔴 🔴 🔵 |	5	$Al_{0.2}Fe_{0.8}$	Line	Open
(*bc* plane)	↻ 🔵 🔵 🔴 🔵	4	$Al_{0.75}Fe_{0.25}$	Rhombus	Closed

Fig. 2 | **Proposed classification of framework multivariation applied to a mixed-metal nanocrystal with a known structure.** In the top row, the elements of metal multivariation are identified. In the bottom row, a selected plane of the crystal (a mixed-metal nanocrystal, MIL-53($Fe_{0.5}Al_{0.5}$)) is used to highlight arbitrary sequences, which are characterized based on the six main descriptors. Although in most real cases such complete structural knowledge is unavailable, crystallographic and chemical analyses can be used to assess the presence of unit cell variability and identify its variables, variants and possible configurations. Similarly, local structure analysis can help to elucidate the most frequent sequences and their features within a certain level of accuracy. UCiC, unit cell information capacity.

Optimizing synthesis via performance. The key to circumventing the characterization problem of multivariance in frameworks is through their properties. Towards this goal, chemists will have to break with the well-established structure-based strategy of making programmable crystalline structures and optimizing them for a specific function through structural design[2,25,26], and instead give way to a function-based strategy. Importantly, the latter does not require full characterization of the precise sequences in a structure at first but instead builds on a feedback loop between synthetic input and output properties, which are directly explored by testing the framework performance instead of being deduced from structural information. Once this iterative optimization process yields a satisfactory performance, the system of sequences can be gleaned by reproducing in silico the effects of various distributions of functionalities and identifying those performing according to the experiments[27–29]. Consequently, as chemists, we are faced with the unusual task of introducing variants synthetically or post-synthetically to target particular functions. In other words, the relationship between synthetic input and performance output — for instance, the ability to separate different species from one another — reveals distinctive traits of the sequences.

MOFs and COFs provide the ideal ground for realizing this performance-based approach. On the one hand, the underlying organic or inorganic framework guarantees a rigid and periodic backbone, while on the other hand, the compositional, configurational and conformational degrees of freedom of the components attached to the backbone provide the entropic driving force to obtain disordered states. In theory, at infinite temperature, these components are completely disordered. Under synthetic conditions, however, kinetic and thermodynamic parameters, such as attractive or repulsive interactions between components, influence their degrees of freedom and can thus lead to local ordering and sequences. One can even go so far as to say that sequences naturally occur in MOFs and COFs, but have been long overlooked because of our focus on high-symmetry, homogeneous structures.

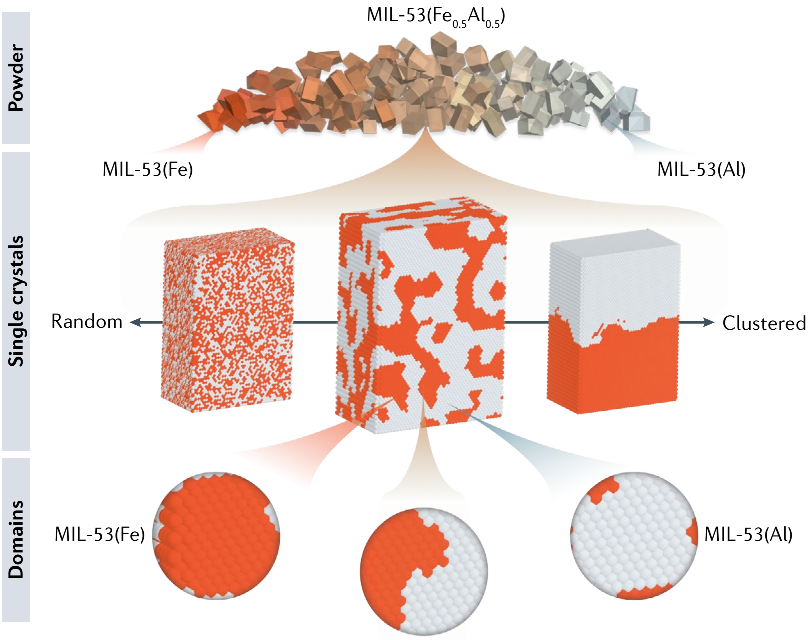
Synthesizing multivariation. In practice, introducing variants in frameworks first requires the identification of possible variables within a pristine, homogeneous structure, such as varying linker and/or metal components. Importantly, only components are considered that preserve the structural integrity of the framework by ensuring its long-range architecture. For instance, the cubic structure of MOF-5 (REF.[30]) allows for extensive variations of the ditopic carboxylic acid linker without compromising its structural integrity, permitting a variety of functionalities ($-NH_2$, –Br, $-(Cl)_2$, $-NO_2$, $-(CH_3)_2$, $-C_4H_4$, $-(OC_3H_5)_2$ and $-(OC_7H_7)_2$) to be introduced in the *ortho*-position to the carboxylic acid units of the benzene linker[4]. In another example, MOF-74 can accommodate divalent metal centres as framework variables, allowing the introduction of several variants such as Mg^{2+}, Ca^{2+}, Sr^{2+}, Ba^{2+}, Mn^{2+}, Fe^{2+}, Co^{2+}, Ni^{2+}, Zn^{2+} or Cd^{2+}, thereby forming sequences within its metal-oxide rods[31]. More generally, once a set of variables compatible with a specific structure is identified, such as a set of linkers in MOF-5 or metals in MOF-74, a library of multivariate frameworks can be made by varying at least one of the many synthetic parameters such as stoichiometry or reaction conditions[9,15,32]. Alternatively, multivariation can be introduced after synthesis by altering the chemical composition locally through selective chemical transformations by external stimuli, such as light, temperature or ion beams, or by post-synthetic approaches such as linker exchange or functional group installation[33–37].

Box 3 | A terminology glitch

The definition of the term ‘crystal phase’ is rather elusive in the terminology guidelines of the international unions of crystallography, chemistry and physics. A general classification of the word ‘phase’ by the International Union of Pure and Applied Chemistry reads “an entity of a material system, which is uniform in chemical composition and physical state”[69]. Once we have acknowledged that the physical state includes most importantly the arrangement of chemical species, we conclude that a polycrystalline powder is made of a single phase when the chemical composition and structural features of a given crystal match those of all the others. Naturally, this requires some flexibility as all crystals are likely not to be identical owing to growth defects, lattice impurities and many other differences. For example, colourless corundum, ruby and sapphire are isostructural Al_2O_3 solids in which metal impurities affect appearance and common name, but not their crystal phase identity as α-alumina.

This need for flexibility is a relatively negligible flaw, but when applied to multivariate frameworks it leads to a terminology glitch. For instance, let us consider a polycrystalline sample of an Fe/Al mixed-metal MIL-53 containing the same stoichiometric amount of both metals. As shown in the figure, although the overall metal composition is $Fe_{0.5}Al_{0.5}$, the concentration of the two metals can vary from 0% to 100% in any single crystal, but also within every single crystal and even in their inner domain structures. In such cases, it is reasonable to wonder whether these single crystals should be considered as being composed of more than one phase.

Similar scenarios are well known to the scientific community working on multicomponent metal alloys[70]. Here, phases are generally distinguished based on the Bragg features of their diffraction patterns, and an ensemble of reflections agreeing with a given unit cell metric is assigned to a single phase based on its average atomic arrangement rather than the composition or the aperiodic sequences formed by its components.



Experimentally, both thermodynamic and kinetic parameters are equally important to modulate — and eventually gain control over — sequences in a framework. Examples of thermodynamic contributors to sequences are molecular recognition events, such as interactions between neighbouring linkers and metal clusters or templating effects with solvents, but also the overall physical properties of the framework, such as charge distributions[38–40]. Kinetic contributors to sequences are different nucleation rates of metal clusters and the stability or solubility of metal precursors and organic linkers[8,41–44]. Although both kinetic and thermodynamic parameters are essential, their control is a lasting challenge. Consequently, recursive experimental screening and simultaneous performance testing are key to obtaining a precise framework multivariation. Ideally, this process is theoretically aided by computational techniques and experimentally accelerated by high-throughput synthesis[15,45].

Sequencing by molecular sorting

The next step is to develop a characterization method that can distinguish sequences from one another by directly measuring how the material’s performance is affected by their presence. Conventional methods such as screening of catalytic conversion rates can only probe local structures, which limits tracing of sequences to a very short range[11,46]. We propose a molecular sorting approach that probes sequences by using travelling molecules to report the variation of chemical environments along their trajectories.

Developing a sorting experiment. Here, we describe how molecular sorting could work in an illustrative case of a cubic multivariate MOF crystal. The interconnected porous space of the framework forms a circuitry of diffusion pathways, allowing guest molecules to travel from one region to another. Because we have multivariation, each diffusion pathway may consist of a unique sequence of chemical environments, making the travelling of a molecule along a given trajectory an integral of various adsorption–desorption events. The varying affinity experienced by the molecule across the framework constitutes a three-dimensional energy surface, where molecular diffusion is not random but a journey through lower-energy paths defined by the structure and distribution of specific sequences. Therefore, the facet of the cube that is most energetically favourable to reach has the highest likelihood of being the destination of molecular diffusion. Once molecules reach a crystal facet, their presence and identity will be measured by mass spectrometry or spectroscopy methods such as surface-enhanced Raman and fluorescence spectroscopies[47,48]. By recording the destination of many diffusion processes, we obtain the distribution of diffusion destinations for a specific molecule, which is a measured property of the multivariate porous crystal.

To bring this experiment into reality, one facet of the cube must function as an inlet for injecting the molecular guests. Therefore, the measured distribution of destinations must be weighted to take into account that the four facets neighbouring the inlet facet have more chances to collect diffusing molecules compared with the facet on the opposite side. After one measurement, the six facets rotate between their roles of inlet and outlet until diffusion processes starting from all six facets have been examined. This precludes artificial choice of inlet and biased sampling of diffusion destinations.

Variants for guest recognition. Let us consider the sorting of CO_2 and H_2O from a gaseous mixture. When both species travel through the multivariate pores, they may experience strongly different interaction energy surfaces because a pore environment highly affinitive to CO_2 may be the opposite for H_2O. By synthesizing a MOF wherein sequences of variants are optimized for promoting the transport of CO_2 in a specific direction and H_2O in another, it becomes possible to collect the two species from distinct facets of the cube (FIG. 3).

The starting step in synthesizing this kind of framework is recognizing which framework variables should be used to create the desired pore environment. Intuitively, a MOF could be designed such that the functional groups occupying equivalent framework sites are variable with two variants, each one having a distinct interaction with CO_2 and H_2O: binding and repellent, or repellent and binding, respectively. When these two variants are distributed appropriately, the diffusion of CO_2 and H_2O are guided towards opposite directions, serving as the site of divergence on the diffusion pathway. The variant that binds CO_2 more than H_2O — which could be polarizable yet hydrophobic functionalities, such as halides and aromatics — attracts the former while diverting the latter. The variant that binds H_2O more than CO_2 functions in the opposite way and could be hydrophilic groups, such as hydroxyls and carboxylates. It may also be useful to have a third variant that is non-bonding to both CO_2 and H_2O, which can quickly transport either CO_2 or H_2O with less resistance after the point of divergence. Lastly, a variant that is repellent to both CO_2 and H_2O should fill the rest of the crystal so that CO_2 and H_2O do not deviate from the pathways constituted by the other three variants. Many other variant possibilities can be conceived, leading to crystals of sophisticated sorting performance (for example, sorting three or more molecules).

When using molecular sorting for analysing multivariation, variants and sequences become analytes, and guest molecules must be judiciously selected to achieve the best specificity of the sorting outcome. Ideally, this practice needs to be adapted to the chemistry of the functional groups of interest. Apart from targeting the sorting of CO_2 and H_2O, mixtures of other guest molecules can be used to improve the precision of probing sequences by amplifying the sorting effect. Molecules with strongly different polarities, such as hexane and dimethylsulfoxide (DMSO), can probe sequences of variants with contrasting polarities, whereas molecules of close polarities, such as DMSO and *N*,*N*′-dimethylformamide (DMF), might distinguish between more subtle polarity differences. Another possibility could be using molecules of similar polarity yet different sizes, such as DMF and *N*,*N*′-diethylformamide, to probe steric effects that can manifest in different sequences of bulky variants. Additionally, sorting of D_2O and H_2O provides a unique opportunity to interrogate which sequences can recognize only differences in hydrogen bonding. Another parameter in a sorting experiment is the compositional ratio of the guest mixtures, which can always be varied to investigate the corresponding change in sorting performance.

Screening of sorting performance. Molecular sorting provides a strategy for circumventing the characterization problem, since the sorting performance can correlate with a given synthesis procedure, suggesting that specific types of sequences are formed, before unravelling their exact structure. This strategy can be implemented by combining high-throughput framework synthesis[49] and automated sorting measurements. The latter is best achieved by integrating robotic crystal harvesting and manipulation[50], microfabrication and gas chromatography–mass spectrometry. Once these techniques are established, a large dataset of synthesis input and sorting output can be obtained with high efficiency and low time costs. To further accelerate the screening of multivariate crystals, machine learning algorithms can be incorporated to analyse real-time data and guide the synthesis optimization[51–53].

By performing cycles of screening and identifying the synthetic conditions that lead to the best-sorting crystals, we aim to identify which sequences are the most relevant. The first clue comes from the composition difference between the crystals, which might correlate with outstanding sorting capabilities. When the composition is identified, the next step is acquiring a structural understanding of sequences on the molecular level. Characterization methods include solid-state NMR spectroscopy[7], fluorescence lifetime imaging[36] and atom probe tomography[9]. Alternatively, computation methods may provide insights into how a molecule interacts with its surroundings. By performing molecular dynamics studies, a time-evolving picture of the adsorption and desorption of diffusing molecules can be obtained[54–56]. The interactions between a guest molecule and its multivariate framework environment can be systematically examined and ranked by the calculated energy, from which the actual sequence can be inferred. A particularly interesting possibility could be to explore the synergistic effect by which the alignment of different adsorption sites into a sequence creates a nonlinear behaviour of a diffusing molecule.

Multivariate reticular machines

Other than passive diffusion, active transport or shuttling of molecules becomes possible when molecular machines are incorporated into the crystal[57]. Active transport in porous structures comes at an entropic cost as it turns chaotic motion of molecules into an organized one. This chaotic motion originates from the thermal energy of the molecules but can be counteracted by a directional motion from the host framework. The rotational behaviour of linkers is a well-known example[58], as the rotating parts can push guests that come into their proximity, thereby generating a molecular transport with a specific directionality[59]. These rotating parts, usually called rotors, perform unidirectional motion but only at an extremely short time interval, whereas over longer time intervals the motion averages to a random rotation[60,61]. To fully utilize such

Box 4 | **Practical relevance of unit cell information capacity**

On the theoretical level, the unit cell information capacity (UCiC) parameter represents the number of all possible forms in which a unit cell can be found, according to the presence of a given set of framework variables and their variants. However, reticular chemists might wonder how UCiC might influence their synthetic practice. When it comes to the products obtained in a laboratory, this parameter does not necessarily correspond to the actual number of unit cell types present in a framework. Indeed, UCiC is an expression of capacity, which is a theoretical limit and not a real state of a material. Of course, this capacity can be fully exploited by aiming to synthesize crystals with as many types of unit cells as allowed by the UCiC. However, reticular chemists should not be necessarily concerned with increasing the UCiC or using the full reticular diversity it allows, but rather with gaining control over the spatial arrangement of variants and the resulting properties of the multivariate metal–organic or covalent organic framework. These two purposes conflict, because the simultaneous presence of all possible combinations allowed by the UCiC leads to a more entropically favourable chaotic distribution of variants, thus making their control much less manageable and precise. It is also true that a more diverse multivariation achievable by pushing the types of unit cells to the UCiC limit can afford more sophisticated functions, control of which is indeed challenging but can be seen as the final goal of the most advanced reticular chemistry.

rotors, their rotation should be permanently constrained to be unidirectional, and the effects of forces acting in other directions should be minor[62]. Towards this goal, we introduce rotary motors into the framework as an ensemble that can perform cooperative unidirectional rotation triggered by external stimuli, such as light irradiation[63–65]. By endowing a framework with such molecular machines, the framework becomes a molecular machinery itself[66].

Unique directionality in reticular structures can be achieved when crystallographic directions are unique — a prerequisite that is true only for a limited group of MOFs or COFs. This is a limitation as it strongly reduces the diversity of materials that can be transformed into reticular machinery, whereas ideally it should be possible to apply such transformation to any MOF or COF having the optimal physical and chemical characteristics for an application of interest. To extend this possibility to any given framework, we use sequences to break the symmetry of framework structures and properties. In a cubic framework, for example, the three main crystal directions are symmetry-equivalent. If we introduce unidirectional rotary motors randomly, the resulting molecular transports will average out to a chaotic motion. However, when rotary motors are aligned in non-random sequences with defined directionality, direction-specific cooperative action of these molecular machines can be achieved and therefore lead to a unidirectional flow[67]. In molecular sorting, this provides an additional kinetic handle, as sorting of molecules by chemical interactions is accelerated by anisotropic transport.

This application of rotary motors in molecular sorting takes advantage of the motion of parts of the framework (the rotary motors) to influence the motion of guests within. At the same time, the latter can also lead to the motion of the entire framework crystal in the surrounding medium. This phenomenon occurs because of the conservation of momentum when the resulting force reaches a sufficient magnitude. When rotary motors are activated by light, for instance, we can initiate the guest mass flow and therefore the movement of the framework in a controlled manner (FIG. 4).

Depending on the number of rotary motors, the mass of the guests, their travelling speed and the overall density of the framework, these forces can reach magnitudes high enough to allow the crystal to travel in space, especially in a low-viscosity medium[68]. Here, sequences formed by two or more variants are key because randomly arranged rotary motors eventually result in an overall net-zero mass flow. In addition to achieving a unidirectional mass flow, sequences allow us to selectively control the directions of the framework's movements. For instance, if we have two rotary motor variants activated by two distinct wavelengths and each type forms a homogeneous sequence of itself across specific crystallographic directions, it becomes possible to control the direction of the guest mass flow depending on the wavelength that is used. By expanding this concept, the construction of porous multivariate crystals driven by multiwavelength light pulses or other stimuli empowers an ensemble of molecular machines to perform cooperatively as a single reticular machine.

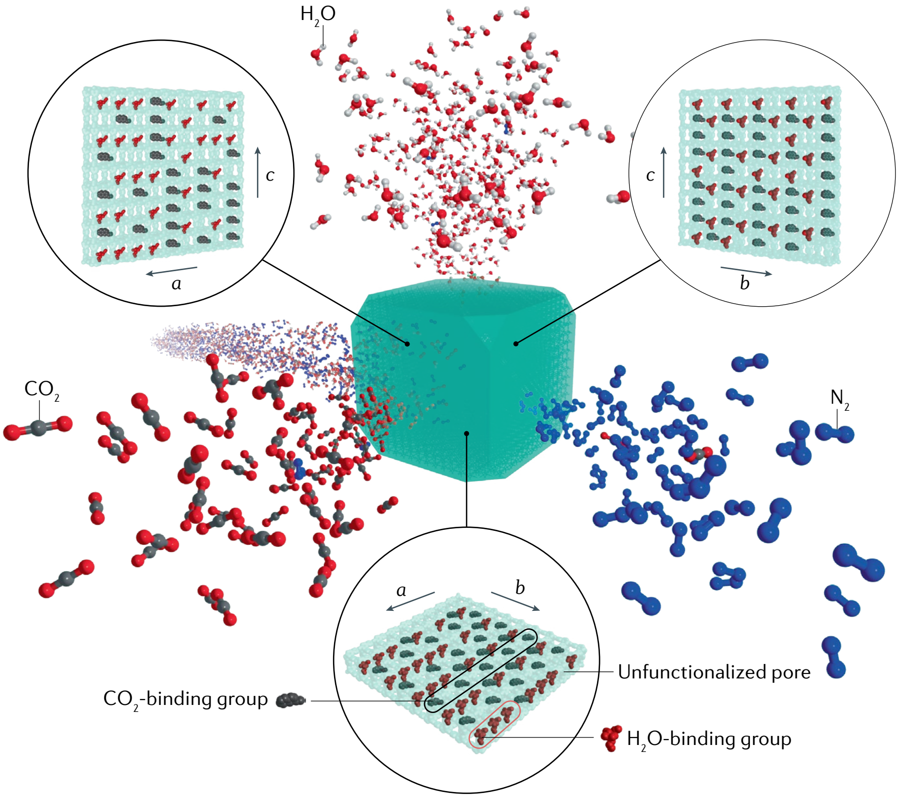


Fig. 3 | **Molecular sorting performed on an $N_2/CO_2/H_2O$ gas mixture by a multivariate MOF-5 framework.** Each pore is present as one of three variants: unfunctionalized, or functionalized with CO_2- or H_2O-binding groups (highlighted in grey or red, respectively). Along the *a* direction, homogeneous repeats of H_2O-binding groups (red oval) prevent water molecules from reaching the surface, while sequences of alternating CO_2-binding groups and voids (black oval) attract CO_2 without hindering its transport and lead to a CO_2-enriched stream. Similarly, an H_2O-enriched stream is achieved along the *c* direction, leaving an H_2O- and CO_2-depleted gas flowing along *b*. Because N_2 molecules interact poorly with both groups, they pass through the framework in every direction and will be dominant along the *b* direction.

Concluding remarks

In this Perspective, we outline how the system of sequences can be a pivotal concept for overcoming the most prominent challenges in the understanding and design of multivariate reticular structures. These multivariate structures break the dogma of traditional reticular chemistry that links framework identity to chemical composition and average structure. Indeed, by considering only these aspects, different multivariate frameworks matching in these two criteria are considered identical, even though they may have radically different properties.

The dogma remains valid for conventional MOFs and COFs, within the approximation that their structures are periodic. However, every reticular structure normally contains unknown structural heterogeneities with certain local arrangements in the lattice. This suggests that a sequence-oriented view is the most general approach to overall reticular chemistry, especially considering that frameworks with undetected defects may have properties that are considered

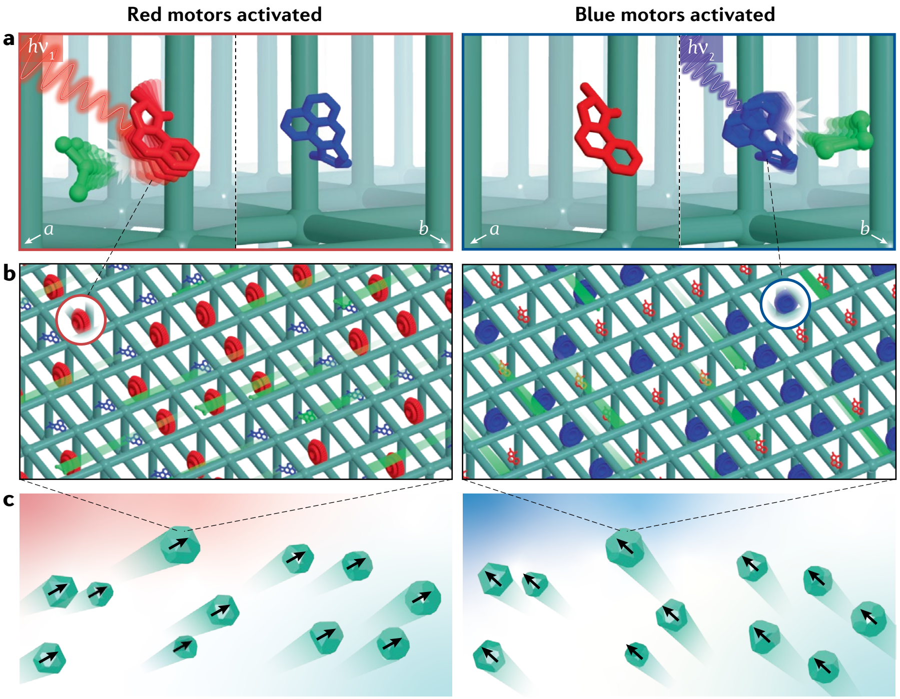


Fig. 4 | **Multivariate reticular machines at work in suitable conditions.** Two types of helical rotary motors (red or blue) activated by distinct irradiation wavelengths are grafted onto a cubic framework backbone. Each type of motor is aligned to form homogeneous sequences along a unique direction: red motors along *a*, blue motors along *b*. **a** | Depending on the irradiation wavelength, only the red (left column) or blue (right column) motor is activated. **b**,**c** | As solvent molecules (green) travel across the framework cavities, the rotational motion of the activated rotary motor turns the molecules' transport from random to unidirectional, resulting in (part **b**) collective unidirectional flow, and ultimately in (part **c**) unidirectional motion of the framework crystals in the surrounding medium. *h*, Planck's constant; *ν*, frequency of irradiation.

typical of their 'periodic' phase, while being influenced by aperiodicity in unknown ways. Moreover, these lattices may be periodic in their composition but still feature multivariation of practical importance. For instance, functional groups can be present on every linker, but disordered in different positions, thereby creating various local arrangements interacting differently with guest species. From this new viewpoint, framework aperiodicity comes into the spotlight as an opportunity, with sequences as its most prominent characteristic. Sequences thus lie at the heart of the approach we present for the classification, synthesis, characterization and application of multivariate frameworks, which we anticipate to be only a starting point for the bright future of this field.

We have aimed to outline the conceptual and methodological blueprint for developments in multivariate reticular chemistry, which will be key to answering its most crucial open questions: how and whether sequences interact and serve to encode specific properties, how and to what precision we can achieve synthetic reproducibility of specific sequences, and what methods we can develop that allow us to reliably distinguish and characterize sequences.

These challenges are complicated by what we described as the characterization problem, which can be circumvented by combining the structure-based approach to framework design with characterization methods such as molecular sorting. In future, when the characterization problem can be solved by more advanced analytical techniques, molecular sorting will be a crucial complementary analysis to access sequence-dependent properties without structural information. Similarly, the establishment of function-based reticular design will allow it to merge with traditional structure-based design in a more advanced hybrid approach to reticular synthesis.

In essence, the concept of sequences changes reticular chemistry at a fundamental level by making framework properties no longer discrete and periodic, but a spectrum of spatially varying states encoding chemical and physical functions. Similarly, periodic MOFs and COFs can now be seen as ground-state forms that can be elevated to bands of countless structural states, giving new meaning to the word 'multifunctional' in reticular chemistry.

Stefano Canossa[1,9]✉, *Zhe Ji*[2], *Cornelius Gropp*[3], *Zichao Rong*[3,4], *Evelyn Ploetz*[5], *Stefan Wuttke*[6,7] *and Omar M. Yaghi*[3,4,8]✉

[1] *EMAT, Department of Physics, University of Antwerp, Antwerp, Belgium.*

[2] *Department of Chemistry, Stanford University, Stanford, CA, USA.*

[3] *Department of Chemistry, University of California–Berkeley, Berkeley, CA, USA.*

[4] *Bakar Institute of Digital Materials for the Planet (BIDMaP), Division of Computing, Data Science, and Society, University of California–Berkeley, Berkeley, CA, USA.*

[5] *Department of Chemistry and Center for NanoScience, Ludwig-Maximilians-Universität München, Munich, Germany.*

[6] *Ikerbasque, Basque Foundation for Science, Bilbao, Spain.*

[7] *Basque Center for Materials, University of the Basque Country, UPV/EHU Science Park, Leioa, Spain.*

[8] *Joint UAEU–UC Berkeley Laboratories for Materials Innovations, UAE University, Al Ain, United Arab Emirates.*

[9] *Present address: Max Planck Institute for Solid State Research, Stuttgart, Germany.*

✉*e-mail: s.canossa@fkf.mpg.de; yaghi@berkeley.edu*

https://doi.org/10.1038/s41578-022-00482-5

1. Freund, R. et al. 25 years of reticular chemistry. *Angew. Chem. Int. Ed.* **60**, 23946–23974 (2021).
2. Jiang, H., Alezi, D. & Eddaoudi, M. A reticular chemistry guide for the design of periodic solids. *Nat. Rev. Mater.* **6**, 466–487 (2021).
3. Xu, W. et al. Anisotropic reticular chemistry. *Nat. Rev. Mater.* **5**, 764–779 (2020).
4. Deng, H. et al. Multiple functional groups of varying ratios in metal–organic frameworks. *Science* **327**, 846–850 (2010).
5. Liu, Q., Cong, H. & Deng, H. Deciphering the spatial arrangement of metals and correlation to reactivity in multivariate metal–organic frameworks. *J. Am. Chem. Soc.* **138**, 13822–13825 (2016).
6. Meekel, E. G. & Goodwin, A. L. Correlated disorder in metal–organic frameworks. *CrystEngComm* **23**, 2915–2922 (2021).
7. Kong, X. et al. Mapping of functional groups in metal–organic frameworks. *Science* **341**, 882–885 (2013).
8. Sue, A. C.-H. et al. Heterogeneity of functional groups in a metal–organic framework displays magic number ratios. *Proc. Natl Acad. Sci. USA* **112**, 5591–5596 (2015).
9. Ji, Z., Li, T. & Yaghi, O. M. Sequencing of metals in multivariate metal–organic frameworks. *Science* **369**, 674–780 (2020).
10. Furukawa, H., Müller, U. & Yaghi, O. M. 'Heterogeneity within order' in metal–organic frameworks. *Angew. Chem. Int. Ed.* **54**, 3417–3430 (2015).
11. Xia, Q. et al. Multivariate metal–organic frameworks as multifunctional heterogeneous asymmetric catalysts for sequential reactions. *J. Am. Chem. Soc.* **139**, 8259–8266 (2017).
12. Zhai, Q.-G., Bu, X., Mao, C., Zhao, X. & Feng, P. Systematic and dramatic tuning on gas sorption performance in heterometallic metal–organic frameworks. *J. Am. Chem. Soc.* **138**, 2524–2527 (2016).
13. Feng, Y., Chen, Q., Jiang, M. & Yao, J. Tailoring the properties of UiO-66 through defect engineering: a review. *Ind. Eng. Chem. Res.* **58**, 17646–17659 (2019).
14. Taddei, M. When defects turn into virtues: the curious case of zirconium-based metal–organic frameworks. *Coord. Chem. Rev.* **343**, 1–24 (2017).

15. Cadman, L. K. et al. Compositional control of pore geometry in multivariate metal–organic frameworks: an experimental and computational study. *Dalton Trans.* **45**, 4316–4326 (2016).
16. Yuan, S. et al. Continuous variation of lattice dimensions and pore sizes in metal–organic frameworks. *J. Am. Chem. Soc.* **142**, 4732–4738 (2020).
17. Ehrling, S. et al. Adaptive response of a metal–organic framework through reversible disorder–disorder transitions. *Nat. Chem.* **13**, 568–574 (2021).
18. Bennett, T. D., Cheetham, A. K., Fuchs, A. H. & Coudert, F.-X. Interplay between defects, disorder and flexibility in metal–organic frameworks. *Nat. Chem.* **9**, 11–16 (2016).
19. Islamov, M., Babaei, H. & Wilmer, C. E. Influence of missing linker defects on the thermal conductivity of metal–organic framework HKUST-1. *ACS Appl. Mater. Interfaces* **12**, 56172–56177 (2020).
20. Krivovichev, S. V. Which inorganic structures are the most complex? *Angew. Chem. Int. Ed.* **53**, 654–661 (2014).
21. Krivovichev, S. V. Structural and topological complexity of zeolites: an information-theoretic analysis. *Microporous Mesoporous Mater.* **171**, 223–229 (2013).
22. Simonov, A. & Goodwin, A. L. Designing disorder into crystalline materials. *Nat. Rev. Chem.* **4**, 657–673 (2020).
23. Ejsmont, A. et al. Applications of reticular diversity in metal–organic frameworks: an ever-evolving state of the art. *Coord. Chem. Rev.* **430**, 213655 (2021).
24. Gándara, F. & Bennett, T. D. Crystallography of metal-organic frameworks. *IUCrJ* **1**, 563–570 (2014).
25. Ockwig, N. W., Delgado-Friedrichs, O., O'Keeffe, M. & Yaghi, O. M. Reticular chemistry: occurrence and taxonomy of nets and grammar for the design of frameworks. *Acc. Chem. Res.* **38**, 176–182 (2005).
26. Zhang, Y.-B. et al. Introduction of functionality, selection of topology, and enhancement of gas adsorption in multivariate metal-organic framework-177. *J. Am. Chem. Soc.* **137**, 2641–2650 (2015).
27. Li, S., Chung, Y. G., Simon, C. M. & Snurr, R. Q. High-throughput computational screening of multivariate metal-organic frameworks (MTV-MOFs) for CO$_2$ capture. *J. Phys. Chem. Lett.* **8**, 6135–6141 (2017).
28. Drummond, M. L., Cundari, T. R. & Wilson, A. K. Cooperative carbon capture capabilities in multivariate MOFs decorated with amino acid side chains: a computational study. *J. Phys. Chem. C* **117**, 14717–14722 (2013).
29. Wu, Y., Duan, H. & Xi, H. Machine learning-driven insights into defects of zirconium metal–organic frameworks for enhanced ethane–ethylene separation. *Chem. Mater.* **32**, 2986–2997 (2020).
30. Li, H., Eddaoudi, M., O'Keeffe, M. & Yaghi, O. M. Design and synthesis of an exceptionally stable and highly porous metal–organic framework. *Nature* **402**, 276–279 (1999).
31. Wang, L. J. et al. Synthesis and characterization of metal–organic framework-74 containing 2, 4, 6, 8, and 10 different metals. *Inorg. Chem.* **53**, 5881–5883 (2014).
32. Liu, Q., Cong, H. & Deng, H. Deciphering the spatial arrangement of metals and correlation to reactivity in multivariate metal–organic frameworks. *J. Am. Chem. Soc.* **138**, 13822–13825 (2016).
33. Feng, L. et al. Imprinted apportionment of functional groups in multivariate metal–organic frameworks. *J. Am. Chem. Soc.* **141**, 14524–14529 (2019).
34. Feng, L. et al. Creating hierarchical pores by controlled LinkerThermolysis in multivariate metal–organic frameworks. *J. Am. Chem. Soc.* **140**, 2363–2372 (2018).
35. Dodson, R. A., Kalenak, A. P. & Matzger, A. J. Solvent choice in metal–organic framework linker exchange permits microstructural control. *J. Am. Chem. Soc.* **142**, 20806–20813 (2020).
36. Schrimpf, W. et al. Chemical diversity in a metal–organic framework revealed by fluorescence lifetime imaging. *Nat. Commun.* **9:1647**, 1–10 (2018).
37. Fracaroli, A. M. et al. Seven post-synthetic covalent reactions in tandem leading to enzyme-like complexity within metal–organic framework crystals. *J. Am. Chem. Soc.* **138**, 8352–8355 (2016).
38. Svane, K. L., Bristow, J. K., Gale, J. D. & Walsh, A. Vacancy defect configurations in the metal-organic framework UiO-66: energetics and electronic structure. *J. Mater. Chem. A* **6**, 8507–8513 (2018).
39. Trousselet, F., Archereau, A., Boutin, A. & Coudert, F. X. Heterometallic metal–organic frameworks of MOF-5 and UiO-66 families: insight from computational chemistry. *J. Phys. Chem. C* **120**, 24885–24894 (2016).
40. Taddei, M. et al. Mixed-linker UiO-66: structure–property relationships revealed by a combination of high-resolution powder X-ray diffraction and density functional theory calculations. *Phys. Chem. Chem. Phys.* **19**, 1551–1559 (2017).
41. Guo, W. et al. Kinetic-controlled formation of bimetallic metal–organic framework hybrid structures. *Small* **13**, 1–8 (2017).
42. Fukushima, T. et al. Modular design of domain assembly in porous coordination polymer crystals via reactivity-directed crystallization process. *J. Am. Chem. Soc.* **134**, 13341–13347 (2012).
43. Shearer, G. C. et al. Tuned to perfection: Ironing out the defects in metal–organic framework UiO-66. *Chem. Mater.* **26**, 4068–4071 (2014).
44. Ye, G. et al. Boosting catalytic performance of metal–organic framework by increasing the defects via a facile and green approach. *ACS Appl. Mater. Interfaces* **9**, 34937–34943 (2017).
45. Lyu, H., Ji, Z., Wuttke, S. & Yaghi, O. M. Digital reticular chemistry. *Chem* **6**, 2219–2241 (2020).
46. Choi, K. M., Na, K., Somorjai, G. A. & Yaghi, O. M. Chemical environment control and enhanced catalytic performance of platinum nanoparticles embedded in nanocrystalline metal–organic frameworks. *J. Am. Chem. Soc.* **137**, 7810–7816 (2015).
47. Kotnala, A., Ding, H. & Zheng, Y. Enhancing single-molecule fluorescence spectroscopy with simple and robust hybrid nanoapertures. *ACS Photonics* **8**, 1673–1682 (2021).
48. Choi, H.-K. et al. Single-molecule surface-enhanced raman scattering as a probe of single-molecule surface reactions: promises and current challenges. *Acc. Chem. Res.* **52**, 3008–3017 (2019).
49. Banerjee, R. et al. High-throughput synthesis of zeolitic imidazolate frameworks and application to CO$_2$ capture. *Science* **319**, 939–943 (2008).
50. Viola, R. et al. Operator-assisted harvesting of protein crystals using a universal micromanipulation robot. *J. Appl. Crystallogr.* **40**, 539–545 (2007).
51. Moosavi, S. M. et al. Capturing chemical intuition in synthesis of metal–organic frameworks. *Nat. Commun.* **10**, 1–7 (2019).
52. Raccuglia, P. et al. Machine-learning-assisted materials discovery using failed experiments. *Nature* **533**, 73–76 (2016).
53. Butler, K. T., Davies, D. W., Cartwright, H., Isayev, O. & Walsh, A. Machine learning for molecular and materials science. *Nature* **559**, 547–555 (2018).
54. Skoulidas, A. I. Molecular dynamics simulations of gas diffusion in metal–organic frameworks: argon in CuBTC. *J. Am. Chem. Soc.* **126**, 1356–1357 (2004).
55. Witherspoon, V. J. et al. Combined nuclear magnetic resonance and molecular dynamics study of methane adsorption in M$_2$(dobdc) metal–organic frameworks. *J. Phys. Chem. C* **123**, 12286–12295 (2019).
56. Düren, T., Bae, Y. S. & Snurr, R. Q. Using molecular simulation to characterise metal–organic frameworks for adsorption applications. *Chem. Soc. Rev.* **38**, 1237–1247 (2009).
57. Karlen, S. D. & Garcia-Garibay, M. A. in *Molecular Machines* (ed. Kelly, T. R.) 179–227 (Springer, 2005).
58. Gonzalez-Nelson, A., Coudert, F.-X. & van der Veen, M. A. Rotational dynamics of linkers in metal–organic frameworks. *Nanomaterials* **9**, 330 (2019).
59. Evans, J. D., Krause, S. & Feringa, B. L. Cooperative and synchronized rotation in motorized porous frameworks: impact on local and global transport properties of confined fluid. *Faraday Discuss.* **225**, 286–300 (2021).
60. Vogelsberg, C. S. et al. Ultrafast rotation in an amphidynamic crystalline metal organic framework. *Proc. Natl Acad. Sci. USA* **114**, 13613–13618 (2017).
61. Perego, J. et al. Fast motion of molecular rotors in metal–organic framework struts at very low temperatures. *Nat. Chem.* **12**, 845–851 (2020).
62. Danowski, W. et al. Unidirectional rotary motion in a metal–organic framework. *Nat. Nanotechnol.* **14**, 488–494 (2019).
63. Kistemaker, J. C. M., Štacko, P., Visser, J. & Feringa, B. L. Unidirectional rotary motion in achiral molecular motors. *Nat. Chem.* **7**, 890–896 (2015).
64. Koumura, N., Geertsema, E. M., Meetsma, A. & Feringa, B. L. Light-driven molecular rotor: unidirectional rotation controlled by a single stereogenic center. *J. Am. Chem. Soc.* **122**, 12005–12006 (2000).
65. Koumura, N., Zijlstra, R. W. J., van Delden, R. A., Harada, N. & Feringa, B. L. Light-driven monodirectional molecular rotor. *Nature* **401**, 152–155 (1999).
66. Krause, S. & Feringa, B. L. Towards artificial molecular factories from framework-embedded molecular machines. *Nat. Rev. Chem.* **4**, 550–562 (2020).
67. Kolodzeiski, E. & Amirjalayer, S. Collective structural properties of embedded molecular motors in functionalized metal–organic frameworks. *Phys. Chem. Chem. Phys.* **23**, 4728–4735 (2021).
68. Terzopoulou, A. et al. Metal–organic frameworks in motion. *Chem. Rev.* **120**, 11175–11193 (2020).
69. IUPAC. *Compendium of Chemical Terminology* 2nd edn (Blackwell, 1997).
70. Kozak, R., Sologubenko, A. & Steurer, W. Single-phase high-entropy alloys — an overview. *Z. Kristallogr. Cryst. Mater.* **230**, 55–68 (2015).

Acknowledgements

S.C. acknowledges the Research Foundation Flanders (FWO) for financial support (grant ID: 12ZV120N), and thanks H.-B. Bürgi and R. Frison for introducing him to the crystallography of correlated disorder. O.M.Y. thanks current and former students whose experiments have helped to shape the ideas presented. C.G., Leopoldina postdoctoral fellow of the German National Academy of Science (LPDS 2019-02), acknowledges the receipt of a fellowship of the Swiss National Science Foundation (P2EZP2-184380). E.P. thanks the Center for NanoScience Munich (CeNS) and the Deutsche Forschungsgemeinschaft (PL 696/4-1) for financial support.

Author contributions

S.C. contributed to writing, supervision, review and editing, and graphics. Z.J., C.G. and Z.R. contributed to writing, review and editing. E.P. and S.W. contributed to review and editing. O.M.Y. contributed to writing, review and editing, and supervision.

Competing interests

The authors declare no competing interests.

Peer review information

Nature Reviews Materials thanks Vladislav Blatov, Dariusz Matoga and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Publisher's note