Please cite this article in press as: Lyu et al., Digital Reticular Chemistry, Chem (2020), https://doi.org/10.1016/j.chempr.2020.08.008

Chem

CellPress

Review

Digital Reticular Chemistry

Hao Lyu,[1,2,3,*] Zhe Ji,[1,2,3] Stefan Wuttke,[4,5] and Omar M. Yaghi[1,2,3,*]

SUMMARY

Reticular chemistry operates in an infinite space of compositions, structures, properties, and applications. Although great progress has been made in exploring this space through the development of metal-organic frameworks and covalent organic frameworks, there remains a gap between what we foresee as being possible and what can actually be accomplished with the current tools and methods. The establishment of digital reticular chemistry, where digital tools are deployed, in particular laboratory robotics and artificial intelligence, will fundamentally change the current workflow to enable discovery of this untapped chemical space and to go beyond the limits of human capacity. In so doing, long-standing challenges in reticular chemistry can finally be addressed faster and better, and more significantly, new questions, unimagined before digitization, can be articulated. The interface between human and "machine" is an integral part of this endeavor and one whose quality is critical to uncovering science transcending intellectual and physical borders.

INTRODUCTION

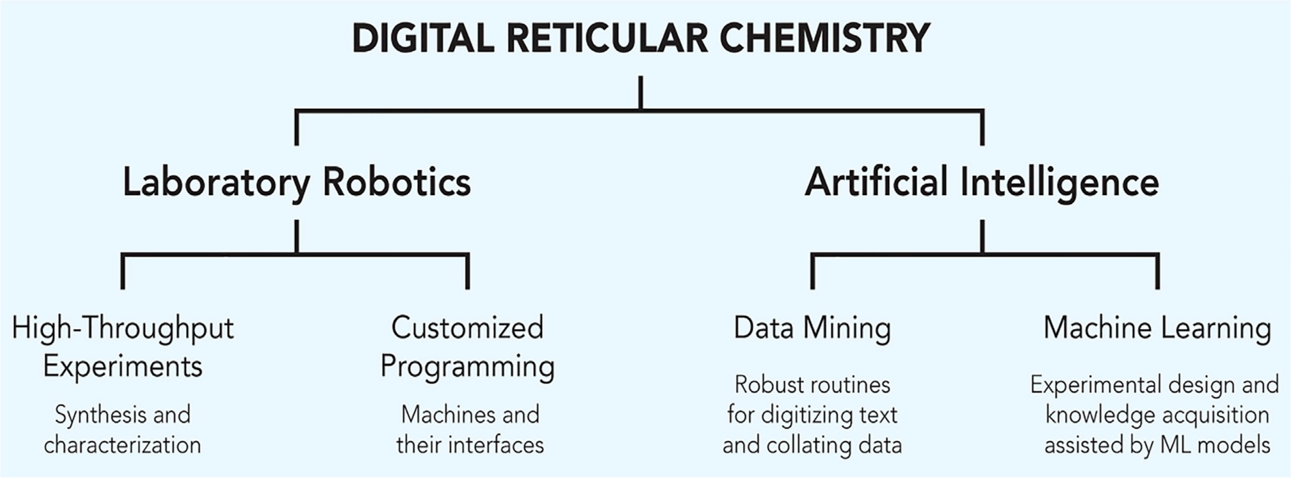
Reticular chemistry is defined as linking molecular building units by strong bonds into crystalline extended structures.[1] This definition encompasses three features contributing to the current expansion of the field. First, the building units are well-defined organic molecules and metal-containing entities whose geometry and rigidity can be used to synthesize extended structures by design. The size of the building units ensures that the resulting constructions are open frameworks—an approach that has been used to reticulate a large number of inorganic and organic building units into extensive classes of porous metal-organic frameworks (MOFs) and covalent organic frameworks (COFs).[2–20] Second, linking these building units by "strong" covalent or metal-organic bonds (specifically, charge-assisted metal-to-ligand bonds, such as metal-carboxylate) has resulted in architecturally and chemically robust structures amenable to chemical functionalization, without compromising their overall integrity and utility.[21–28] Third, crystallinity is essential to achieving atomic-level characterization and atomic precision design of these structures.[29–32] These three defining characteristics of reticular chemistry have extended metal-complex and organic chemistries into infinite 2D and 3D porous frameworks within which the pore environment can be further manipulated and controlled to realize numerous applications.[33–48]

Given the vast number of building units and their corresponding frameworks, and the chemical modifications that can be made to tailor the pores, it is worthwhile to recognize that reticular chemistry operates in an infinite space of composition, structure, property, and application. To date, we have only exploited a tiny fraction of this infinite space and, although we have accessed remarkable properties for MOFs and

The Bigger Picture

The full potential of linking molecular building units with strong bonds to make extended chemical structures (reticular chemistry) can be unlocked by the development of digital tools of laboratory robotics and artificial intelligence. This digitization leads to faster discoveries that have an impact on society and the articulation of previously unimagined questions. Turning what is typically an empirical practice into a data-driven one promises to transform the current state of affairs in reticular chemistry: researchers making chemical structures and investigating their possible properties rather than targeting a structure for a specific property. The four pillars of digital reticular chemistry—a comprehensive database, computational and experimental discovery cycles, and the human-digital interface—are ideally suited to efficiently generate, test, and implement research ideas into impactful discoveries that improve society at large.

Please cite this article in press as: Lyu et al., Digital Reticular Chemistry, Chem (2020), https://doi.org/10.1016/j.chempr.2020.08.008



Scheme 1. Tools for Developing Digital Reticular Chemistry

COFs, the vast potential of reticular chemistry remains untapped. How do we realize this potential and in doing so how will reticular chemistry be transformed? The current methods of reticular chemistry, and for that matter chemistry in general, are inadequate to the task of fully exploiting the vast reticular space.

There is a disparity between what we think is possible in reticular chemistry and what we are able to achieve using the current methods and tools. This is likely to continue and grow, hampering the intellectual and practical advancement of not only our field but also chemistry as a whole. Our conjecture is that the disparity can be addressed by adopting and developing digital tools. As shown in Scheme 1, these are laboratory robotics and artificial intelligence (AI). These tools will enable high-throughput experiments involving synthesis and characterization of compounds and programming the robot machinery and interfaces to automate the process. In addition, the digital tools also include data-mining efforts where robust routines for digitizing text and collating data will be established, as well as using machine learning to help in making data-driven decisions (Scheme 1). Such tools will fundamentally change how we carry out research and how researchers formulate questions. Although, these digital tools are being rapidly developed by specialists in various fields,[49–53] very little is permeating as routine tools for chemistry researchers to be able to transform their experimental inquiries. This "digital gap" can be addressed by using reticular chemistry as an excellent starting point for digitization because of its vast chemical space that is ripe for exploration. We wish to remark that society has already created digital tools to routinely and easily conduct affairs in the business sector, such as commerce, banking, and trade, thereby greatly expanding the creation of new opportunities and their reach worldwide. We see no insurmountable problems in deploying digital tools in reticular chemistry and thereby finding new vistas for exploration.

We propose the establishment of "digital reticular chemistry," which we define as being concerned with the use of laboratory robotics and AI (Scheme 1) to develop intellectual and practical tools for gaining speed and sophistication in the chemistry and applications of framework structures. As experienced in other fields where digitization is more prevalent,[54–56] inevitably there will be two components to digital reticular chemistry. The first is a human component, where our hope is to encourage the practice of reticular chemistry beyond the limitation of human capacity and, in so

[1]Department of Chemistry, University of California, Berkeley, Berkeley, CA 94720, USA

[2]Kavli Energy Nanoscience Institute, University of California, Berkeley, Berkeley, CA 94720, USA

[3]Berkeley Global Science Institute, Berkeley, CA 94720, USA

[4]BCMaterials, Basque Center for Materials, UPV/EHU Science Park, Leioa 48940, Spain

[5]Ikerbasque, Basque Foundation for Science, Bilbao 48013, Spain

*Correspondence: hao.lyu@berkeley.edu (H.L.), yaghi@berkeley.edu (O.M.Y.)

https://doi.org/10.1016/j.chempr.2020.08.008

Please cite this article in press as: Lyu et al., Digital Reticular Chemistry, Chem (2020), https://doi.org/10.1016/j.chempr.2020.08.008

doing, provide new opportunities to map this untapped chemical space and make impactful discoveries. The other is the digital component, where robust, reliable tools must be developed on the basis of objective data. Most of the work will be at the interface joining these two components. The exchange at the interface between humans and computers in producing meaningful discoveries and analysis will be paramount to the success of digitization of reticular chemistry.

In this contribution, strategies for how to close the digital gap and further develop reticular chemistry are described. We examine the current operating model of reticular chemistry research and contrast it with what we propose for its digitization. A discussion is given for how some outstanding fundamental questions in the field can be answered by this effort, and new questions might be articulated and addressed. Understanding the interface between researcher and "machine" becomes critical to the success of this endeavor, and as such the human element of digitization is also explored.

A VAST RETICULAR CHEMICAL SPACE

Reticular chemistry and the framework structures have flourished over the last two decades. This resulted in the creation of more than 80,000 MOFs and 500 COFs,[1,57–60] whose chemical properties have been studied and explored. Already with this number of new structures, reticular chemistry spans a large chemical space where important questions of societal impact are being addressed. By developing digital tools, we believe that this field will continue to make significant intellectual and practical contributions. However, we also believe that these digital tools could be applied to exploring the potential for asking questions yet unimagined. In this section, we discuss the creation of a more extensive chemical space than we have today and present a view of the components of reticular structures, which will figure prominently in studying this space through digitization.

Anatomy and Variations of Frameworks

The "anatomy" of a porous reticular framework is based on the intermingling of matter (linked atoms) and pore space (voids), each of which is vital to their chemistry. Matter in reticular frameworks can be dissected into (1) a framework backbone, where the building units are joined through a strong linkage to make a continuous network of bonded molecules; (2) functionalities, which include organic units, metal complexes, or polymers covalently bonded to the backbone; and (3) guests, a range of molecules, clusters, or even nanoparticles occupying specific or disordered positions in the pore space and interacting with the framework through intermolecular forces. Extensive efforts have been dedicated to making variations in the kind of building units used to construct the framework's backbone and the functionalities employed to design the pore environment into which many different kinds of guests can be introduced. A large majority of variations being made in reticular chemistry today lead to "simple" structures, where a framework is constructed with fewer variety than potentially possible for each one of the three major components.[61–64] Even within this simple structure space, diversity is created by just varying the parameters pertaining to composition, size, linkage, and connectivity of the building units to make frameworks based on different structure types. Furthermore, as we discuss below, a new level of diversity is achieved by applying variations to one or more of these components within a single framework (Figure 1).

Here, we present some of the variations that can be achieved in the building units. The composition can be varied with the majority of metals in the periodic table.

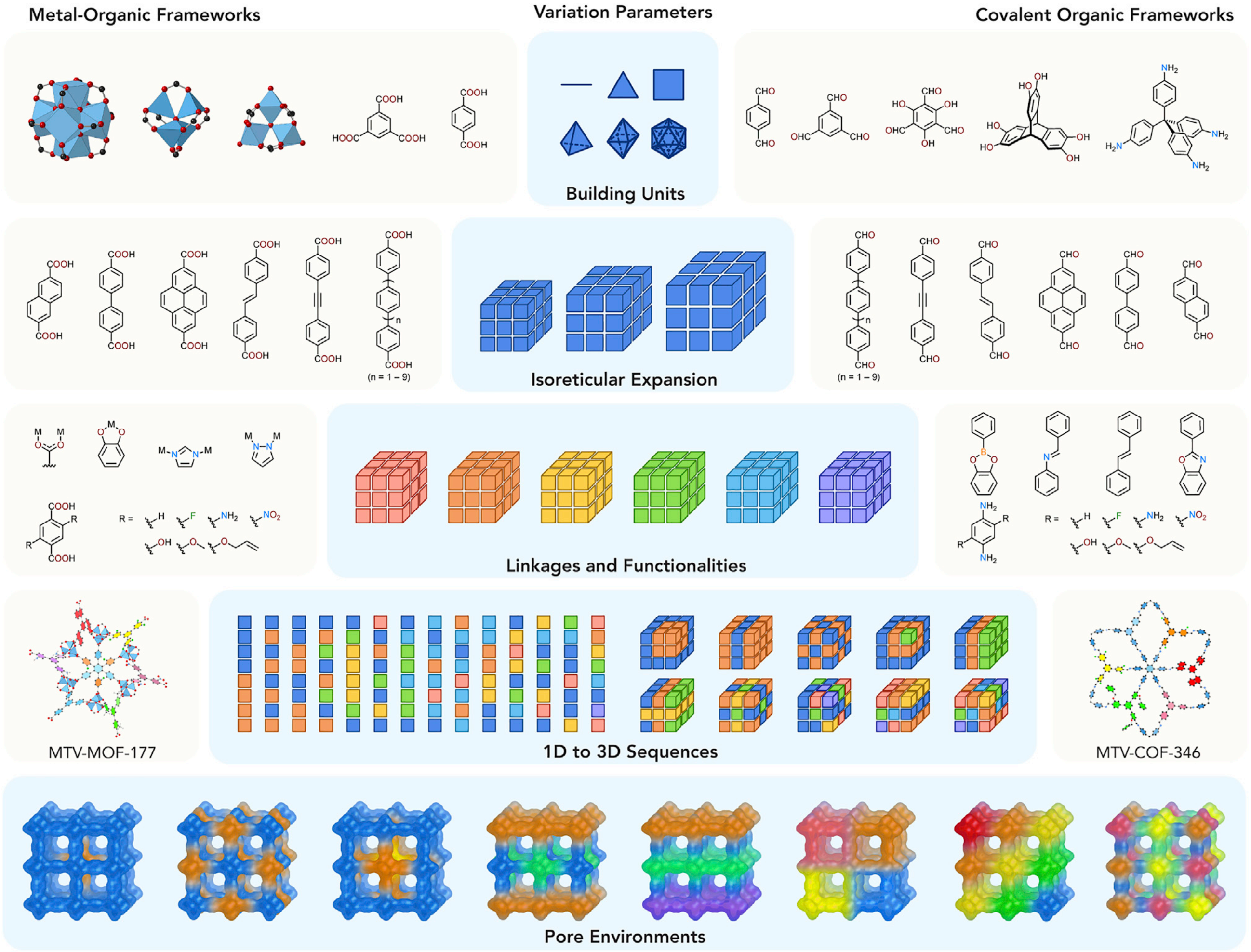


Figure 1. Vast Reticular Structural Space

The diversity of reticular structures is manifested by the multiplicity at each design level and its combinatorial progression across all levels. Rigid molecules with specific geometries are first identified as the building units for making frameworks of targeted topologies. For a chosen geometry (rod) of the molecular building unit, its size can be tuned by varying the length of the organic struct without altering its connectivity and shape. The linkage groups between the connected building units are then specified for completing the construction of the framework backbone, onto which functionalities of desired chemical identities can be appended. When multiple functionalities of different kind are incorporated into a single framework, their spatial arrangement constitutes rich chemical sequences extending in 1D to 3D space. Given a functionality sequence, complex and diverse pore environments that interact differently with guest molecules can be created.

The sizes of building units have been shown to range from 1 to 16 metal centers for discrete units[15,30–32,65,66] and infinity for rods[67,68] and sheets[69] in MOFs and from 1 to 11 phenylene units in MOFs and COFs.[70] In addition, non-phenyl[71] as well as chiral[47,72] and naturally available[4,18,73] organic linkers have been used. In MOFs, the linkages have metals bound to carboxylates,[30] sulfonates,[74,75] phosphonates,[75,76] catecholates,[77] diamines,[38] dithiolates,[78] imidazolates,[79] pyrazolates,[17,80] triazolates,[81,82] and tetrazolates,[83] whereas in COFs, linkages are boroxines,[5] boronate esters,[5] borosilicates,[84] aminals,[85] imines,[86] azines,[87] olefins,[88–90] hydrazones,[91] ureas,[92] amides,[23] imides,[10] phenazines,[93] dioxins,[94] oxazoles,[95] thiazoles,[26] imidazoles,[96] triazines,[97] and borazines[98] (Figure 1).

The connectivity of building units gives different framework types (topologies) whose number can be quite large. Just for linking triangles, there are almost a million possible topologies that could be reticulated, as well as many more for other shapes and combinations thereof.[62–64] At this point, it is useful to remark that even with the simple reticular frameworks known to date, they do occupy a front and center place in solving some of the most vexing energy and water problems facing society.[33–48]

The extent to which the vast reticular space can have multiplicity and diversity may be appreciated by contrasting how these attributes are imparted in inorganic solids and biological macromolecules. A common strategy for making variations on inorganic solid-state compounds is to substitute constituent elements with those belonging to the same group in the periodic table. One example is substituting sulfur with selenium in the hope that the obtained derivatives have the same structure type and thus exhibit properties relevant to making comparisons. In general, the chemical space of inorganic solids is largely defined by the combination of possible constituent elements within significant chemical constraints (e.g., electronegativity[99] and size matching[100,101]) and limited choices of elements obeying those constraints.

The richness of reticular structures is based on the fact that the metal ion is not the point of extension because it is embedded in the building unit contributing to its stability but not its connectivity as such. For example, in MOF-5, far more variations can be implemented at the carbon atoms making up the corners of the $Zn_4O(O_2C–)_6$ octahedron than at zinc atoms in the Zn_4O core. Thus, although some diversity can be created in MOFs by varying the metal ions,[102] far more diversity can be achieved by the organic linkers.[43] Clearly, when these two types of variations are working together, another level of diversity is realized (Figure 1). Conversely, in inorganic solids, the site of variation is typically the metal, which is based on relatively fewer options. In addition, inorganic solids often have challenges in adapting large varieties of structure types as a result of the expensive energy penalty associated with the loss of packing efficiency.[103]

Multivariate Character and Biological Structures

It is natural to think about what will come beyond the present state of affairs. The reticular chemical space will expand explosively once we start increasing the number of different kind of building units and functionalities incorporated into what will become more appropriately termed "multivariate" structures.[104,105] Multivariate is used when (1) multiple linkers that are topologically equivalent are identical in their binding groups and metrics but different in their chemical composition,[106–108] (2) multiple topologically equivalent metal ions form the same multi-metallic building unit,[105,109,110] and (3) aperiodic vacancies are present on or in the backbone of a single reticular structure.[111,112] How to precisely describe the outcome of this increase in multiplicity, diversity, and complexity and how to study it meaningfully become central questions with overwhelming possibilities and challenges, ideally suited for digitization.

Biological molecules such as proteins and DNA are based on very few different kinds of amino acids and nucleotides, but these are linked into specific spatial arrangements (sequences). Their amino acids or nucleotides are covalently attached to a regularly repeating amide (proteins) and a polyphosphate-sugar (DNA) backbone. The principle of sequences that are central to biology can be transferred into reticular structures. For example, this can be done by the covalent attachment of almost an infinite number of multiple different functionalities onto the backbone of the framework. Here, the variance exists not only in the spatial arrangement of the functionalities but also in the backbone itself when the number of its constituent building units is increased. This scenario of increased variation in functionality and backbone constituents coupled with their spatial arrangement brings unparalleled diversity to the reticular chemical space. The variance in such reticular structures is clearly described by the sequences of constituents, whether they be building units in the backbone or functionalities bound to that backbone (Figure 1).[113] However, instead of adopting a discrete linear form as in biological molecules, the sequences in reticular structures are infinite 1D, 2D, and 3D according to the dimensionality of the void space and the framework.[114] A recent report involving multivariate MOFs with mixed-metal rods showed that the metal arrangement along the rods follows unique sequences.[115]

A Continuum of Matter and Pore Space

Now, we turn to the pore space that is being created and architecturally maintained by the backbone and where its environment is diversified by the functionalities. The shape of pores and what an incoming guest "experiences" electronically and sterically can vary as much as the composition of the whole does. Thus, the same infiniteness described for matter applies to the pore space it encompasses. In other words, an infinite number of possible backbone building units, functionalities, and also space configurations can be described in terms of sequences (Figure 1).

The vast reticular chemical space can be appreciated by illustrating the infiniteness of variations achievable in only one framework. The power of our variation capability is manifested by combinatorial mathematics and the magnitude of the exponential growth that results from a cubic structure. How many unique chemical environments may result from a cube where each of its 12 edges has a functionality chosen from among eight different types? There exist 2,863,870,080[116,117] unique pore environments that can be "sampled" by, for example, a chiral guest molecule, far exceeding the total number of compounds that have ever been made by humankind. When these cube units are further joined into an extended cubic framework (Figure 1), the connection of the pores of varying chemical environments into channels creates an even more massive environment space belonging to only this one structure type! Intermolecular forces acting within these chemical environments translate into equally variable and extensive complexity.

On a fundamental level, this immense composition, structure, and chemical environment space coupled with our ability to make infinite number of variations on each of the components leads us to propose that reticular frameworks will no longer occupy discrete states of matter and pore space. Instead, the vast number of states that can be created in even one structure means that the characteristics of these structures will lie on a continuum of states.[107] Such states are intrinsic to reticular structures because the multivariate character can be created on a framework without altering its integrity. This feature is harder to achieve in inorganic solids, proteins, and polymers because the outcome of their multivariate modifications is much less predictable than reticular structures. This continuum of compositional, physical, and chemical properties defines our vast reticular chemical space and motivates the need for digitization to identify, understand, and control the possible topologies, sequences, and pore environments promised by multivariate frameworks.

On the basis of the foregoing discussion, we believe that two parameters amenable to digitization will be important for digital exploration: topology is a global descriptor serving as a blueprint for how the building units are connected to make frameworks,[62–64] and sequences are chemical information superimposed onto the framework and used in the description of the environment within the pores.[114]

Please cite this article in press as: Lyu et al., Digital Reticular Chemistry, Chem (2020), https://doi.org/10.1016/j.chempr.2020.08.008

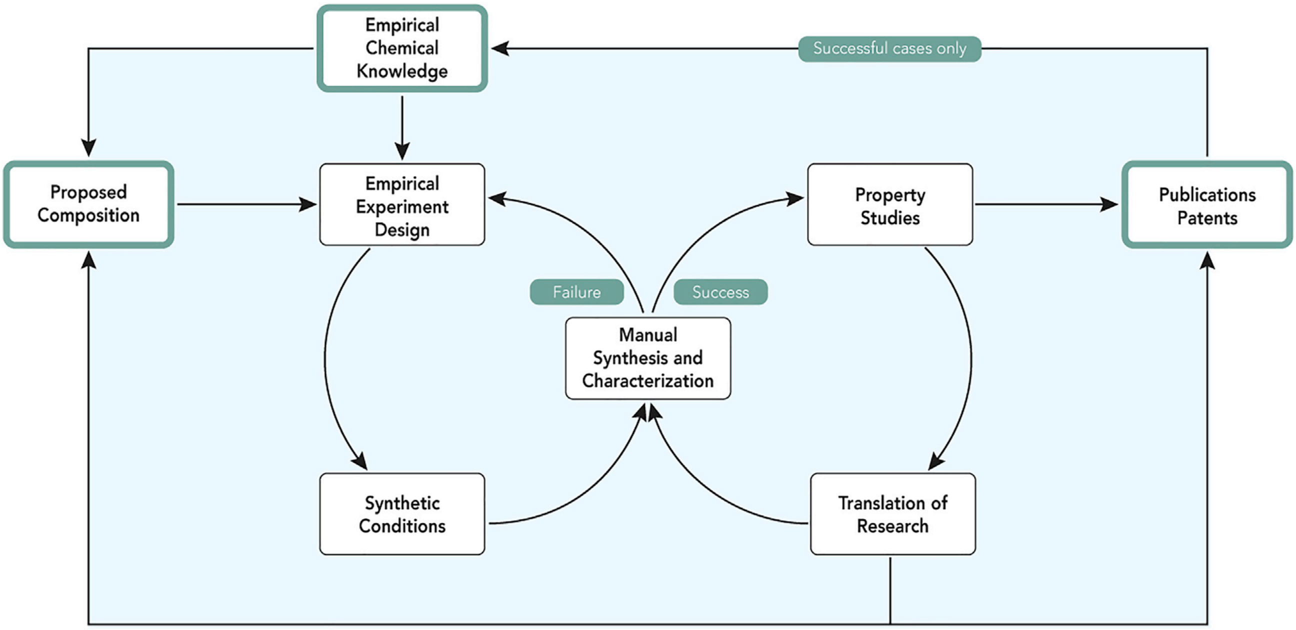


Figure 2. The Current Workflow of Reticular Research

To date, reticular synthesis has largely relied on empirical knowledge in the design of experiments. Synthetic discovery involves trials of manual synthesis and characterization in which those that "fail" to achieve crystallinity are unpublished and fed back for the redesign of synthetic conditions. Successful trails enter the optimization and proper studies cycle and are eventually translated into publications and new knowledge.

CURRENT WORKFLOW OF RETICULAR CHEMISTRY

Despite the exciting developments in reticular chemistry over the last two decades, the digital gap persists mainly because of the high dependence of the research workflow on manual experimentation. Here, we choose the synthesis of new reticular structures as an illustrative aspect of how the present workflow is generally implemented (Figure 2). The process begins with an idea of linking a set of building units of specific proposed composition and geometry with only empirical knowledge of what structure(s) might be produced. Thus, in the experimental design, there is a heavy reliance on experience and development of empirical approaches for the prediction of the outcome of reticular synthesis. After choosing the building units, synthetic conditions are considered, and the synthesis and subsequent characterizations are executed manually,[118] again on the basis of the researcher's empirical knowledge and experience. "Success" is realized by the attainment of a crystalline material. If the reaction is not successful, an iterative process is performed whereby the researcher redesigns the experiments and varies the synthetic conditions again, continually testing the outcome of each reaction by X-ray diffraction to assess the crystallinity of the material. The most critical reaction parameters leading to crystallinity are narrowed down to a few, which are then optimized by feeding the findings back into the cycle until ultimately the desired crystalline product is obtained. Throughout this process, the researcher is screening conditions for crystallization and making decisions on the validity of the approach employed and also on the course of action to be taken for achieving a successful result. These decisions are made on the spot and are guided by observations and, in some cases, the aid of routine definitive measurements, but for the most part they rely on experience. The quality of the decisions a researcher makes is directly related to the precision of their observations. When success is achieved, the properties of the crystalline solids are studied, and if they are found to be unique, the research is translated into application. The process and its outcome are documented and published, but the less successful or ''failed'' attempts are not given the same intense attention; therefore, these are not well documented and often not published.

We wish to remark that for digitization to be successful as a scientific endeavor, data obtained from ''failed'' experiments are equally important as those obtained from ''successful'' experiments. Accordingly, reporting both is an integral part of the experiment and a necessary condition for understanding the chemical space that digital reticular chemistry intends to explore.

Although this workflow has led to one of the most productive fields of science, namely reticular chemistry, its digitization is urgently needed to address some issues with its current practice. First, the heavy reliance on manual processes limits the speed with which information can be uncovered about the system, and this hampers efforts of gaining the maximum knowledge to guide informed decisions. This in turn detracts from idea generation and creativity. Second, poor documentation of those failed attempts to achieve the desired results is detrimental to forming a complete picture of the innerworkings of the system and deprives others of the benefit of having such information.[119] Third, the lack or primitiveness of generic, systematic, and shared databases for efficient knowledge sharing works against the need for systematic passing down of information and ultimately the diversification of approaches for breaking new grounds in the field. Fourth, discovery by empirical guidance and without the full benefit of having systematic knowledge is disadvantageous in tackling complex problems and makes it difficult for others outside reticular chemistry to engage in the process.

OUTSTANDING QUESTIONS AND EXPLORATION WITH DIGITAL TOOLS

Here, we discuss some unanswered elementary questions that we expect digital tools to be valuable in providing opportunities to address. Among the almost infinite number of possible structures that could form from a combination of building unit geometries, which of these will form in the synthesis? In this regard, our thesis has been that the most symmetric structures are the most likely to form, and this has served as a rough guide in helping to identify target structures.[120] However, this thesis does not always work, especially when the building units deviate from high symmetry and when one uses an increasing number of different kinds of units of different geometries to make multicomponent frameworks.[121] To date, only an empirical approach has been attempted to answer this question; no systematic, comprehensive efforts have yet been reported in the field of reticular chemistry. This raises the question, how can one design a structure if its ''blueprint'' is unknown?

The tendency at present is to make a framework structure and then ask what property that structure might have. It would be more productive and perhaps more meaningful to ask, what framework design could provide a specific property? Thus, going from structure to property is the present practice, but rarely the reverse is true as it has not been a straightforward exercise.

Even when a framework is targeted, and its appropriate building units are identified and used to make it, extensive efforts by the researcher are applied to finding the answer to the question, under what conditions can the targeted framework be obtained in crystalline form? As explained in the previous section, this is an arduous task requiring not only rigorous and extensive variation of synthetic conditions but also the need for accurate observations so that correct decisions are made concerning the course of action to be taken for ultimately obtaining crystals.

Please cite this article in press as: Lyu et al., Digital Reticular Chemistry, Chem (2020), https://doi.org/10.1016/j.chempr.2020.08.008

Digital reticular chemistry should be able to address these questions by using robotics for high-throughput synthesis, characterization, and screening of the properties of frameworks, and this, as we discuss below, should be closely coupled to data mining and machine learning (Scheme 1). Specifically, if we succeed in bringing these tools to routine utility, reticular chemistry will benefit in a multitude of ways. It will be possible to make discoveries unrealized with the current empirical approach because of the speed with which the vast reticular space can be explored, mapped, and analyzed. For example, finding a targeted framework in such a space, which is normally like finding "a needle in a haystack," will no longer be done by an empirical approach but rather by a methodical and well-thought-out process provided by digitization. As researchers become adept at this method of inquiry, they will gain new intellectual skills where they begin to see new possibilities for discovery and perhaps view their results in ways otherwise not possible. Indeed, as observations of chemical reactions and their results (e.g., color and crystallinity) become progressively more accurate with digitization, so will the ideas that researchers formulate and the path they choose for their realization.

Fundamentally, as in the development and use of any new tools, their strengths and weaknesses must be evaluated objectively, understood, and appropriately managed. A reliable digital model needs to be developed for reticular chemistry so that we do not risk falling in some of the less attractive aspects of digitization. We can minimize some of these by ensuring that we are critical in excluding possible data poisoning and biasing, overfitting, or falling into false minima.[122] It will be important to keep in check any tendency to over-trust the results coming from digital tools. We wish to emphasize that although digital reticular chemistry will expand the scope and diversity of frameworks and their applications, including perhaps transforming how we do research, still questions dealing with the possible blunting of serendipity and dampening the excitement of discovery as a result of digitization may emerge as less desirable outcomes. However, in the last section of this contribution, we outline how digitization, when employed properly, will in fact add to the intellectual vitality and rigor of those practicing it. Nevertheless, in the fullness of time, we expect that these new tools will advance the fundamental knowledge of reticular chemistry and lead to faster and better solutions to societal problems.

PROPOSED WORKFLOW FOR DIGITAL RETICULAR CHEMISTRY

The workflow of digital reticular chemistry is shown in Figure 3, and it is based on four pillars: (1) a reticular chemistry database that serves as the medium for all the storage and conveyance of information, (2) a computational discovery cycle that serves as a predictive workforce by analyzing information created by humans and/or robotics, (3) an experimental discovery cycle that generates real experimental data on the basis of AI suggestions by using high-throughput methods, and (4) a human-digital interface for queries and providing solutions to intellectual and practical problems in reticular chemistry. Each of these pillars is discussed below in the context of digitizing reticular chemistry.

Reticular Chemistry Database

Although there has been productive establishment of databases for MOF structures (much less has been done for COFs),[58,60,123–125] we propose the creation of a general and all-encompassing database platform. This would include structures as well as experimental and theoretical findings concerning synthetic methods, properties, and applications. Our vision is to create an active platform allowing researchers to deposit data and interrogate the body of knowledge in the field. This would be

Please cite this article in press as: Lyu et al., Digital Reticular Chemistry, Chem (2020), https://doi.org/10.1016/j.chempr.2020.08.008

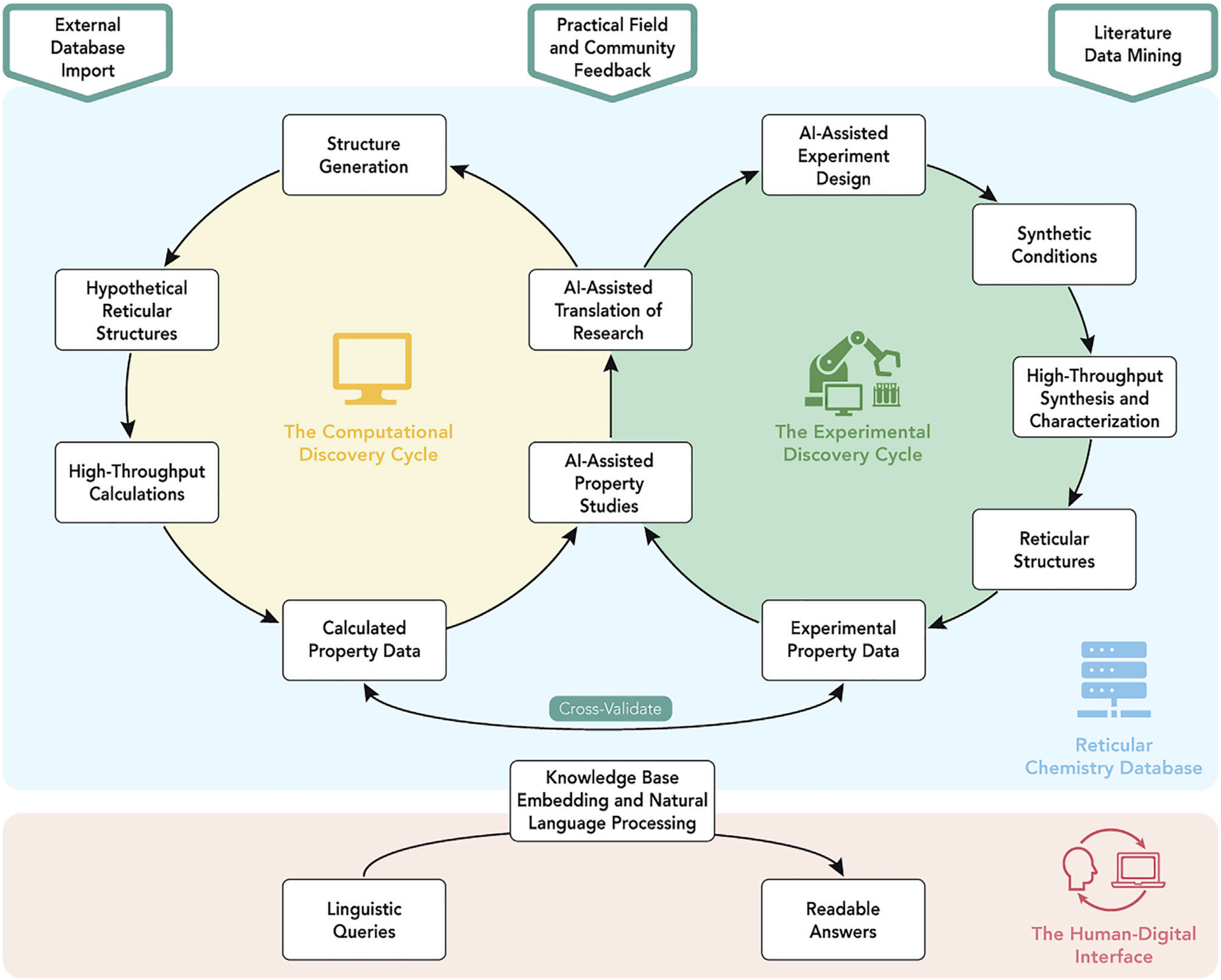


Figure 3. The Proposed Workflow of Digital Reticular Chemistry

A reticular chemistry database houses all information extracted from external databases, community feedback, and literature data mining. Such chemical information is fed to and exchanged between the computational discovery cycle and the experimental discovery cycle, the former of which proposes structures and calculates properties and the latter of which inquires AI for suggestions and generates high-throughput experimental data. Finally, a human-digital interface is designed to interact with users by digesting linguistic queries and providing readable answers.

designed as a widely accessible digital reticular chemistry tool organized in a manner amenable to routine use by researchers worldwide. The hope is that it would constitute a reliable destination for researchers' inquiries with complete and up-to-date information. Specifically, the platform would store and integrate chemical information obtained from both computation and experiments. These would be calculated data from the computational discovery cycle, imported data from outside databases (e.g., the Materials Project126 and Materials Cloud[126]), extracted data from existing experimental literature reports, data and observation from experimental discovery cycle, and feedback and deposit of data from the community.

Progress is already being made in documenting experimentally discovered MOFs and COFs with unique identifiers (MOFid and MOFkeys)[127] and single-crystal structures (Cambridge Structural Database MOF subset,[58,128] U-M MOF database,[129] computation-ready, experimental [CoRE] MOF database,[123,130] and CoRE COF database[125]). These databases are already being used for simulation of the reticular frameworks' physical parameters (e.g., surface area and pore volume)[131,132] and gas sorption properties (e.g., methane uptake and selectivity)[133] and the subsequent establishment of simulated property databases (MOFDB[134] and CURATED COFs[60]). These efforts have been significantly accelerated by computational screening of MOFs and COFs; however, our proposed reticular chemistry database will be required for proceeding well beyond these efforts.

The comprehensive summary of chemical information from literature reports, such as publications and patents, and their subsequent storage in an organized database format is the next logical and necessary step toward a reticular chemistry database. This is complicated by the issue of text-based media that are discrete and not readily accessible in digital form and thus cannot be directly imported to a database. Text mining and natural language processing (NLP) techniques should be implemented in the extraction of chemical information from such sources. Available tools include rule-based text mining[135–137] and machine-learning-based NLP (e.g., recently developed mat2vec for inorganic materials).[138] To date, the collection of literature reports of reticular frameworks has reached an enormous size, and their processing now and into the future requires large-scale deployment of text mining and NLP protocols, which is achievable only through the coordinated efforts of the global science community. Another important aspect will be the documentation of all conducted experimental data, especially "failed" experiments—a practice that is not currently done but we view to be essential as an input of the two AI-assisted discovery cycles discussed in the next two sections.

Our proposed efforts in the documentation of data from all sources mentioned above paves the way for further mining and storage of chemical information through techniques of knowledge base embedding (KBE).[139,140] We anticipate that the database will be represented in the hybrid form of factual entities (e.g., identifier of compounds, structures, and properties) and quantitative parameters (e.g., density, specific surface area, and gas uptake). KBE involves the mapping of these entities in a Euclidian space through vectorizing their coordinates to preserve their structure. In this way, chemical knowledge can be properly represented as vectors (relationships) between entities, and undiscovered chemical knowledge can be proposed through missing vector prediction by KBE models. We anticipate that the establishment of such a vectorized database to be a useful form, representing the reticular chemical space and enabling machine-accelerated discovery.

In addition, building a comprehensive database would finally allow researchers to compare different data sets and, importantly, put their own results in the context of the entire field. In doing so, it would address one major challenge that is caused by the current disparity of data. With the establishment of KBE and machine-learning models in the workflow, data cross-validation and correlation-based discernment will be enabled to gradually enhance the robustness of the database (Figure 3). In the future, the database's enormous body and decentralized nature will require reticular scientists to work as an integral part of the global science community,[141] contributing data for populating the chemical space under uniform guidelines as in the FAIR (findable, accessible, interoperable, and reusable) Data Principles.[142] Only in this way can a standardized, high-quality, rich, and healthy reticular chemistry database be used for AI-assisted discovery.

The Computational Discovery Cycle

This involves the entirely computation-based chemical discovery processes in digital reticular chemistry. The cycle is characterized by a high level of reproducibility and scalability endowed by computation technologies.[122] It provides extremely fast discovery cycle, especially in comparison with laboratory synthetic screening and characterization, thereby reducing the time and cost, which otherwise would be unavoidable.

Computer-based structure generation is the first step in this cycle. Currently, several geometry-based tools and algorithms (e.g., RCSR,[62] Systre,[143] ToposPro,[64] AuToGraFs,[144] ToBaCCo,[145] and TopoFF[146]) have been developed and used to generate hypothetical structural databases (hMOFs[124] and hypothetical COFs[147,148]). Future research should include machine-learning-based algorithms to incorporate chemical information into such processes. In this way, the resulting hypothetical reticular structures will be combined with experimental structures (e.g., CoRE MOFs) and subjected to high-throughput calculations for property data prediction on the level of molecular dynamics (e.g., Monte Carlo simulations for gas sorption),[147,149,150] density functional theory (e.g., point charges and band structures),[148,151,152] *ab initio* calculations,[153–155] and machine-learned models.[154,155] These calculated property data can be further used for statistical analysis (e.g., linear or non-linear regression and principal-component analysis)[149,156] or machine-learning methods (e.g., support vector machine [SVM],[157,158] random forest [RF],[155,159,160] genetic algorithm [GA],[161,162] *k*-nearest neighbors [*k*-NNs],[163,164] artificial neural networks [ANNs],[154,165,166] and other higher-level models) to decipher the properties of materials (AI-assisted property studies in Figure 3).

The generated results serve as a basis for the experimental discovery cycle and represent a fundamental new discovery step toward optimizing material performance and translating research. At this point, the knowledge gained is fed back into the structure generation step of the cycle for further reiteration and optimization. Such new candidates continue to be subjected to calculations for the computational screening until they are fit to enter the experimental discovery cycle. Inevitably, the computational discovery cycle is contemporarily the fastest way to expand the landscape of the sparsely populated reticular chemical space and has made the most significant progress with the cooperative efforts of researchers.[122,124,167] However, these algorithmic models are meanwhile not exempted from the ‘‘no free lunch’’ theorem[168] in that it is impossible to achieve cost efficiency and calculation accuracy simultaneously. A current example is that most of the existing reports of grand canonical Monte Carlo (GCMC)-based screening of reticular structures for CO_2 capture[149,169,170] are particularly fast but with insufficient representation of chemisorption,[171] such as for alkylamine-functionalized materials, and may thus lead to incorrect decisions on sorbent candidates. Hence, the mutual connection between the computational and experimental discovery cycles is essential to making robust progress in reticular discoveries: the theoretical calculation results are validated with laboratory experiments[172] to calibrate and further evolve the calculation and models. In the meantime, the experiments are validated with theoretical predictions to check for experimental errors undetected otherwise. On the whole, the computational discovery cycle should be the first tool available for researchers in exploring and predicting the unidentified ‘‘ocean’’ of reticular chemical space and distinguishing it from the ‘‘islands’’ of known chemical information for the purpose of optimization of materials properties.

The Experimental Discovery Cycle

This represents the discovery processes based on synthesis, characterization, and performance measurements done experimentally. The cycle is similar to the current reticular chemistry workflow but includes the extensive use of high-throughput procedures under AI-assisted guidance. This involves implementation of automated

Please cite this article in press as: Lyu et al., Digital Reticular Chemistry, Chem (2020), https://doi.org/10.1016/j.chempr.2020.08.008

high-throughput synthetic screening and automated structural and property characterization in order to dramatically accelerate the exploration of the large reticular chemical space. To date, these techniques are widely implemented in organic synthesis,[173–175] biology,[176,177] and materials science[119,178–181] but are only just beginning to be used in reticular chemistry.[182,183]

In a typical experimental discovery cycle, the reticular framework candidates from the computational discovery cycle will be subjected to synthesis, whose conditions are proposed by the AI-assisted experimental design (Figure 3). The proposed synthetic conditions are systematically screened by high-throughput experimental tools, for example, automatic liquid handlers and solid dispensers for the dosing of chemicals in each synthetic screening trials or robotic arms for sample handling, which are all programmed through interfacing the machines to the output (recipes) of AI-assisted experimental design. This is followed by high-throughput characterization,[184] in most cases autosampler-equipped instruments, such as powder X-ray diffraction and infrared spectroscopy, for unattended and unbiased data acquisition. During these processes, ‘‘failed’’ experiments can be set at the same importance level as successful samples and logged as a part of the reticular chemistry database without biases.[119] The experimental property data are obtained through assessing the physicochemical behavior of the candidate structures, and this is fed into the AI-assisted property studies and later into the AI-assisted translation of research. Ideally, the result of the synergy between the two cycles is the most desirable reticular structure candidates for a specific application. Importantly, the experimental discovery cycle is the only component of digital reticular chemistry that generates real experimental data and observations, and it is thus the only way to fill the reticular chemical space with experimental rather than just hypothetical data.

The Human-Digital Interface

It will be useful to deploy an interactive portal between the reticular chemistry database and the users (academic, private and public industry, and government), which converts the structured, vectorized information of computers into the semantic, linguistic information of humans. This requires the achievement of an easy-to-access linguistic interface (APIs[185] and interactive websites[186,126]) that wraps around the reticular database and allows for natural language queries without prerequisite expertise in reticular chemistry. It also requires the general embedding of the database for reticular chemical space and the knowledge that are capable of reasoning and provide accurate output. It is necessary to deploy NLP techniques to fill the gap between frontier discoveries of reticular chemistry and societal needs.

The four pillars of the digital reticular chemistry workflow are the basis for populating the reticular chemical space with continuous information, but more importantly, they will efficiently find frameworks with globally optimized performance for targeted applications. Although we have emphasized the sequence of going from basic science to applications, various components of the two cycles would work very well for carrying out basic research into new synthetic and characterization techniques as well as discovery of unusual chemical architectures. We believe the establishment of this workflow requires the joint work of scientists in chemistry, physics, mathematics, computer science, and engineering.[187]

THE PRACTICE OF DIGITAL RETICULAR CHEMISTRY

We anticipate that the new level of integrating digital tools into the workflow of reticular chemistry will stimulate discovery and promote the availability of significantly richer data. These tools will further evolve into a means of asking and answering questions otherwise not possible. In this section, we outline as examples three directions where digital reticular chemistry will make significant contributions.

Discovery of the "Infinite" Reticular Chemical Space

The prediction and design of structures with known composition can be generally viewed as an extended crystal structure prediction question.[188] Conventional crystal structure predictions involve the generation of trial structures (arrangements of isolated molecules, ions, or atoms) and calculation of the whole system's energy to give the minimum configuration.[189,190] Reticular chemistry is even more demanding because of the complexity and variety of possible interactions holding the structure together, such as covalent bonds, coordination bonds, and intermolecular forces. Further to be considered are host-guest interactions,[191] molecular weaving,[192] intercatenation,[193] and framework dynamics.[194] Since reticular structures are extended, their topological tessellation constraints add to the complexity of achieving strong bonding of building units. Although computational tools for generating simple structures are available, only the linker geometry and topological constraints are accounted for without the benefit of having any chemical information. Using the reticular chemistry database, we can now consider important factors such as reactivity, preferred conformation, secondary interaction between building units, and interactions with guests. We note that this information could be contributed through experimental and calculated data.

Furthermore, various downstream factors in the workflow can become design criteria through established quantitative structure-property relationships (QSPR) models,[133] such as predicted application performances and mechanical processability. This can be integrated with the development of next-generation reticular specialized machine-learning models to enhance the models in utilizing the chemical knowledge for tackling optimization problems for practical applications. For example, based on the existing machine-learning model, chemical reaction networks,[195] the reactants, products, and reactions in the original model can be specifically adapted into the reticular chemistry scenario as building units, crystals, and linkage formation reactions. The algorithm will then be redesigned to mathematically represent the crystallization process, where building units are treated as networks. The evaluation function can be deployed considering the global metrics of entropy, system energy, crystallinity, porosity, and QSPR functions generated from the computational discovery cycle.

Structure to Property and Property to Structure

The quantitative study of the relationship between structures and their properties is one of the key aspects for understanding the underlying chemical interactions in molecular and materials chemistry.[133] The prediction of optimum structures for a given property is the reciprocal of such processes, and is essential for the field to significantly speed-up the discovery process. One practical method is the high-throughput screening experiments of known reticular structures with the criteria of computational or experimental characterization methods, using ranking, statistical, or machine-learning methods to reach the optimum. However, the majority of current models are black-box methods, yielding little to no information for understanding the underlying relationships. We expect that the next stage of QSPR in digital reticular chemistry to use highly and fully interpretable AI models[196,197] for understanding chemical information. From these, the information of the model's decision flow can be used for deciphering the underlying physical and chemical interactions and cross-validating with observations and learning results from other processes. Accordingly, correlations that are too subtle or too complicated for human perceptibility can be discovered.

Ultimately, digital reticular chemistry aims to revolutionize chemistry discovery by achieving inverse design, which means going from desired properties to the discovery of frameworks with properties far beyond what can be found in nature. Concurrently, this process is hindered by the low level of exploration of the reticular chemical space. The situation will fundamentally change once a continuum of this space is more thoroughly explored and chemical relationships understood. An example is the new scheme of searching for compounds capable of CO_2 capture. Reticular frameworks are promising candidates and have been shown to have record uptake for CO_2 capture and storage.[198–200] However, achieving a practical material requires heat capacity, thermal conductivity, mechanical processability, H_2O and O_2 stability, and selectivity, yet they are mostly undocumented. Further complicating this process is often the need to balance parameters whose behavior is unknown under the industrially relevant conditions, such as the physical form of the material and heat dissipation. In the current workflow, finding such optimal material is highly impractical.

This problem potentially can be handled well by the digital workflow. First, the embedded reticular chemistry database will be capable of vectorizing the requirements into a group, for instance, (1) high dynamic capture capacity, (2) high selectivity of CO_2 against gases such as N_2, O_2, SO_x, and NO_x, (3) appropriate stability under the operating conditions, and (4) abundant thermal conductivity for fast heat dissipation. Fuzzy search of the reticular chemical space and relationship prediction are conducted accordingly to yield the first generation of candidates. AI-assisted synthetic discoveries are then carried out through the experimental cycle, and data generated are used for training the QSPR and structural generation model on the fly. Next, the computational discovery cycle focuses on the virtual screening of reticular frameworks by using data-improved simulation algorithms and sampling the rest of reticular chemical space to make sure of the globality. The workflow runs until a satisfactory candidate is consolidated.

Synthesis Conditions for a Target Framework

Discovering the optimal synthetic conditions for a target reticular framework is one of the most essential quests for reticular chemists. The synthesis of reticular solids involves the achievement of linkage bond formation, crystallization, and pore formation at the same time. This is accomplished by fine-tuning the starting compound(s), solvent composition, temperature, and even feeding procedures, modulator usage, heating schedule, and hydrolysis kinetics. The delicate balance required to optimize these interconnected parameters involves a synthetic space of extraordinarily high dimensionality. Needless to say, the majority of this parameter space has not been discovered. As such, this inverse synthetic design (going from structure to synthetic conditions) is difficult to implement at the current stage of research and thus calls for the deployment of digital tools. Preliminary results along these lines have already been reported for the synthesis of HKUST-1.[183]

An illustrative example of this scenario is the synthesis of MOFs based on titanium-oxo clusters. The synthetic chemistry of this class of structures features harsh conditions, sensitivity to moisture, and a large variety of Ti-oxo clusters (with different geometry and coordination numbers) possible to form in similar conditions.[14,201,202] Such complexity has hindered the field in the development of Ti-oxo MOFs, and to date, the variety of reported structures remains limited. However, because of their unique band structure and photocatalytic activity, Ti-oxo MOFs are still highly desired in the field. The high-throughput synthesis technology has preliminarily shown its power in tackling this grand challenge. Successful utilization of human-guided high-throughput synthesis was achieved in 2019,[203] where new photoactive MOFs based on $Ti_3(\mu_3\text{-}O)$ clusters were achieved. However, the experimental design in this process was reported to be majorly guided by the forward design of synthesis, and a systematic inverse reasoning was not present until a successful product was obtained. We emphasize that in the next level of digital reticular chemistry, the powerful tool of high-throughput synthesis and characterization will be additionally guided by AI models. By employing them, we will finally be capable of systematically establishing the underlying chemical relationships, in this case, the correlation between the hydrolysis of Ti species and the final structure or the initial concentration and nucleation kinetics.

Digitization of Multivariate Arrangements and Their "Evolution" in the Pores

Multivariate MOFs provide a platform where digitalization addresses a high level of structural complexity and functionality. Here, we borrow the concept from the evolution of enzymes to describe and conduct the optimization of multivariate reticular systems for catalysis. Multivariate MOFs, where chosen combinations of functionalities are precisely positioned inside framework pores, harbor complex and functional chemical environments that resemble an enzyme pocket.[204] While the structure and function of enzymes are dictated by the sequence of amino acids, the pore chemistry of multivariate MOFs is encoded by the spatial arrangement of the installed functionalities. The high definitiveness of the distance, orientation, and chemical identity of functionalities provides quantitative metrics that can be used to digitize the chemical environment within the multivariate pores. Such metrics become what we refer to as the "genome" of multivariate MOFs, which features vast diversity and tunability as alluded above and, more importantly, can be expressed into a specific function.

A fundamental question in this field is what arrangement of functionalities encodes for the best catalytic performance. Such correlations have yet to be established, but they become immediately addressable given the advancement of digital tools. In this context, each set of the geometrical parameters of functionality corresponds to an experimentally determined catalytic performance, including yield, selectivity, efficiency, and cyclability. While the evolution of an enzyme pocket occurs when the fittest survives, the "evolution" of a multivariate pore is conducted by human selection, where its genome is screened within the frame of high-throughput experimental discovery cycle for optimal catalytic performances (Figure 4). This process is expected to be largely accelerated by the deployment of deep reinforcement learning methods,[205] where deep neural networks[165] are used for decision making, guiding the navigation of the evolutionary trajectories of MOFs. It is thus envisioned that the robotics and machine-learning algorithms developed for digital reticular chemistry will provide substantial reduction of the natural evolutionary time scale into a laboratory one, accelerating the discovery of efficient catalysts.

The gradual accumulation of data in this process in turn enables the exploration of multivariate MOFs to be performed on the computational level. Machine-learning methods such as self-organizing maps (SOMs),[208] *k*-NNs, and convolutional neural networks (CNNs)[49] can be used for the extraction of significant structural features from the experimental dataset, which provides the basis for structure generation and development of new simulation methods in the computational discovery cycle. In addition, interfacing the reticular chemistry database with enzyme databases[209,210] will allow KBE to utilize "experiences" learned from the enzyme field by importing the biological structure-activity relationships into the reticular chemical space. The integration of multi-disciplinary knowledge will extend our ability to understand catalysis and offers a unifying framework to assess the general properties of catalysts.

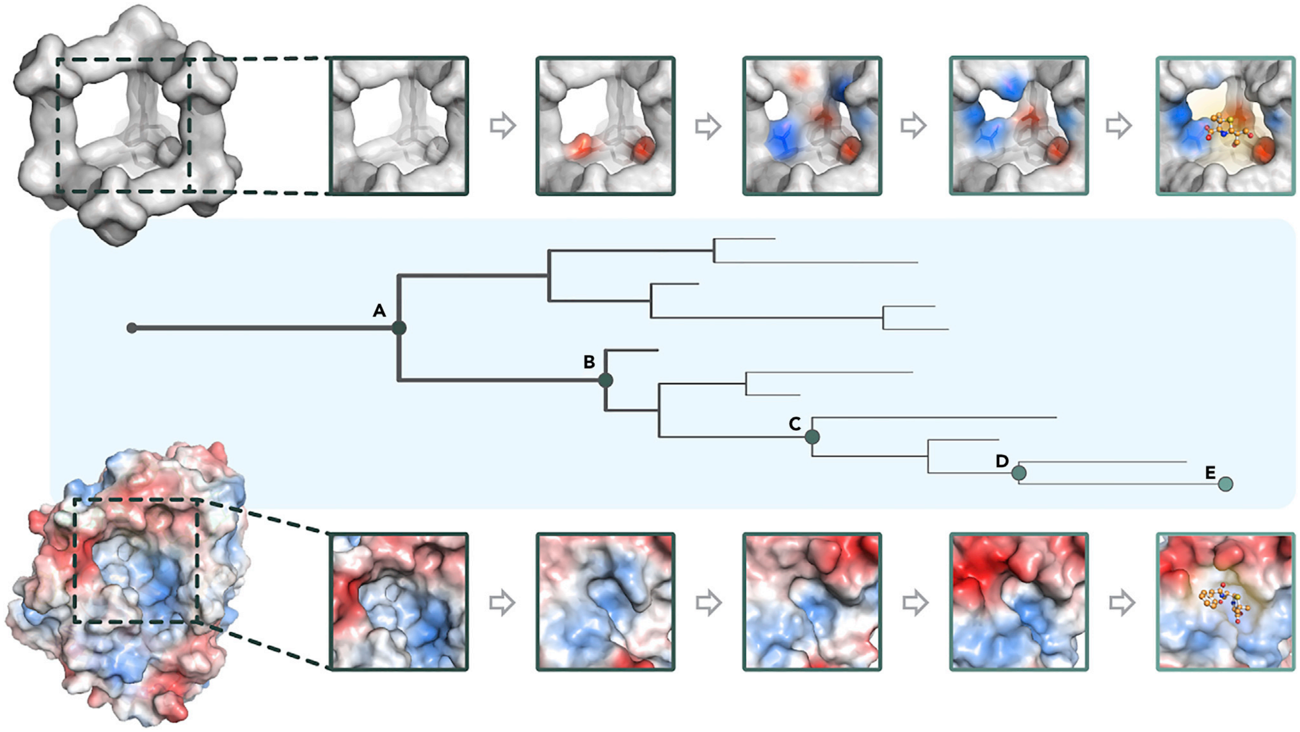


Figure 4. A Proposed Strategy for the Optimization of Multivariate MOF Catalysts and Its Analogy to Enzyme Evolution
To find an efficient MOF catalyst, functionalities are installed on the MOF backbone and screened for generating a variety of pore environments with diverse charge and steric configurations. Promising candidates for catalysis are selected for further functionality variation over multiple iteration cycles. For illustrative purposes, five arbitrary nodes (A–E) on the optimization trajectory are presented in surface filling of their pore environments, where functionalities are colored qualitatively by electrostatic potential (negative, red; positive, blue). The final MOF catalyst is therefore optimized to best stabilize the transition state of the substrate (orange, arbitrary structure in ball-and-stick model). In comparison, the natural evolutionary history of β-lactamases is demonstrated. Under the survival threat of antibiotics, a peptidase was repurposed as β-lactamases[206] and selected for increased catalytic function by evolution on a planetary timescale.[207] These ancestral enzymes, as well as the extant one, corresponding to the five nodes along the evolutionary pathways, are shown in surface filling and colored by electrostatic potential to highlight the transformation of their functional pockets.

THE HUMAN ELEMENT IN DIGITAL RETICULAR CHEMISTRY

Our review expressed in this contribution comes from a real need for a transformative change in the way we conduct chemical research exemplified by that being done in reticular chemistry. Our proposal for the digitization of our chemistry is strengthened by the large chemical space already accessed by current research, as well as the potential that its findings indicate for a much more extensive and diversified space of composition, structure, and properties. The digital tools we wish to deploy in the exploration of this space for the objective of asking better questions and solving more complex problems will inevitably affect not just the research as outlined above but also the researcher.

Arguably, we as scientists do research for the thrill of discovery and to be recognized for those discoveries. The more closely we adhere to the scientific values of honesty, objectivity, curiosity, critical thinking, creativity, and unbiased view of data, the better scientists we will become. We believe digitization of reticular chemistry will contribute to the enhancement of all these values. It will increase drastically transparency due to sharing of data. The reliance on data-driven observations rather than just empirical and experiential notions will increase the precision in making observations and therefore arrive at sound and scientifically valuable conclusions about the system under study. The speed with which problems can be addressed will translate into better and faster decisions in determining the options for experiments to be undertaken and ultimately the course of research. The impact on emerging scholars in our laboratories could be immense: the availability of comprehensive and organized data will empower new researchers to generate ideas of their own and lessen their reliance on the more senior researchers. This will boost their confidence and, in a way, "democratize" the process of idea generation, thereby changing it from the current state of being largely a top-down to a bottom-up process. This will be a boon for bringing further innovation to fundamental and applied science of reticular chemistry. When we have a full view of the vast reticular space, not only will we be able to address convergent problems (specific optimization challenges) better and faster, but we will also be able to pursue divergent problems (exploratory and curiosity-driven inquiries) with the confidence of having the full benefit of the organized available data. The latter may very well lead researchers to ask questions not yet imaginable because of the limited view of the currently available data.

Beyond the researcher, the culture of science is bound to be transformed. For example, in having all experimental results and theoretical determination in the field within every scientist's reach, it will be much easier for researchers to consult with other experts worldwide on their findings, thereby creating a rapid feedback loop, an aspect not yet practiced efficiently under the current workflow. The impact of creating such an instant feedback loop will help in building a global science culture by meaningfully connecting scientists across borders.

We close by remarking that, since digitization potentially makes data available to everyone, any temptation for less scholarly activities, such as unfair practices, hype, misinformation, poor citations, and even "political" networking will be more easily challenged and discouraged by the community at large. The net effect will be a more transparent and professional scientific enterprise, leading to wider acceptance and support by our society. Fundamentally, digital reticular chemistry will transform a largely empirical science to a data-driven science, and in doing so, we will achieve a new level of sophistication in addressing larger and more complex problems and hopefully make more significant discoveries.

ACKNOWLEDGMENTS

We thank Mr. Ralph Freund for his help in graphic illustrations. We acknowledge KACST (Riyadh, Saudi Arabia) as part of a joint KACST-UC Berkeley Center of Excellence for Nanomaterials for Clean Energy Applications. S.W. acknowledges funding from the Basque Government Industry Department under the ELKARTEK and HAZITEK programs.

AUTHOR CONTRIBUTIONS

Conceptualization, O.M.Y.; Writing – Original Draft, H.L., Z.J., and O.M.Y.; Writing – Review & Editing, H.L., Z.J., S.W., and O.M.Y.; Visualization, H.L., Z.J., and S.W.

REFERENCES

1. Yaghi, O.M., Kalmutzki, M.J., and Diercks, C.S. (2019). Introduction to Reticular Chemistry (Wiley-VCH).
2. Yaghi, O.M., Li, H., Davis, C., Richardson, D., and Groy, T.L. (1998). Synthetic strategies, structure patterns, and emerging properties in the chemistry of modular porous solids. Acc. Chem. Res. *31*, 474–484.
3. Kalmutzki, M.J., Hanikel, N., and Yaghi, O.M. (2018). Secondary building units as the turning point in the development of the reticular chemistry of MOFs. Sci. Adv. *4*, eaat9180.
4. Imaz, I., Rubio-Martínez, M., An, J., Solé-Font, I., Rosi, N.L., and Maspoch, D. (2011). Metal–biomolecule frameworks (MBioFs). Chem. Commun. *47*, 7287–7302.
5. Côté, A.P., Benin, A.I., Ockwig, N.W., O'Keeffe, M., Matzger, A.J., and Yaghi, O.M. (2005). Porous, crystalline, covalent organic frameworks. Science *310*, 1166–1170.
6. El-Kaderi, H.M., Hunt, J.R., Mendoza-Cortés, J.L., Côté, A.P., Taylor, R.E., O'Keeffe, M., and Yaghi, O.M. (2007). Designed synthesis of 3D covalent organic frameworks. Science *316*, 268–272.

Please cite this article in press as: Lyu et al., Digital Reticular Chemistry, Chem (2020), https://doi.org/10.1016/j.chempr.2020.08.008

7. Colson, J.W., Woll, A.R., Mukherjee, A., Levendorf, M.P., Spitler, E.L., Shields, V.B., Spencer, M.G., Park, J., and Dichtel, W.R. (2011). Oriented 2D covalent organic framework thin films on single-layer graphene. Science *332*, 228–231.

8. Wan, S., Guo, J., Kim, J., Ihee, H., and Jiang, D. (2008). A belt-shaped, blue luminescent, and semiconducting covalent organic framework. Angew. Chem. Int. Ed. *47*, 8826–8830.

9. Ding, S.Y., Gao, J., Wang, Q., Zhang, Y., Song, W.G., Su, C.Y., and Wang, W. (2011). Construction of covalent organic framework for catalysis: Pd/COF-LZU1 in Suzuki-Miyaura coupling reaction. J. Am. Chem. Soc. *133*, 19816–19822.

10. Fang, Q., Zhuang, Z., Gu, S., Kaspar, R.B., Zheng, J., Wang, J., Qiu, S., and Yan, Y. (2014). Designed synthesis of large-pore crystalline polyimide covalent organic frameworks. Nat. Commun. *5*, 4503.

11. Vyas, V.S., Haase, F., Stegbauer, L., Savasci, G., Podjaski, F., Ochsenfeld, C., and Lotsch, B.V. (2015). A tunable azine covalent organic framework platform for visible light-induced hydrogen generation. Nat. Commun. *6*, 8508.

12. Dogru, M., Handloser, M., Auras, F., Kunz, T., Medina, D., Hartschuh, A., Knochel, P., and Bein, T. (2013). A photoconductive thienothiophene-based covalent organic framework showing charge transfer towards included fullerene. Angew. Chem. Int. Ed. *52*, 2920–2924.

13. Chui, S.S.Y., Lo, S.M.F., Charmant, J.P.H., Orpen, A.G., and Williams, I.D. (1999). A chemically functionalizable nanoporous material $[Cu_3(TMA)_2(H2O)_3]$. Science *283*, 1148–1150.

14. Dan-Hardi, M., Serre, C., Frot, T., Rozes, L., Maurin, G., Sanchez, C., and Férey, G. (2009). A new photoactive crystalline highly porous titanium(IV) dicarboxylate. J. Am. Chem. Soc. *131*, 10857–10859.

15. Cavka, J.H., Jakobsen, S., Olsbye, U., Guillou, N., Lamberti, C., Bordiga, S., and Lillerud, K.P. (2008). A new zirconium inorganic building brick forming metal organic frameworks with exceptional stability. J. Am. Chem. Soc. *130*, 13850–13851.

16. Zacher, D., Shekhah, O., Wöll, C., and Fischer, R.A. (2009). Thin films of metal–organic frameworks. Chem. Soc. Rev. *38*, 1418–1429.

17. Colombo, V., Galli, S., Choi, H.J., Han, G.D., Maspero, A., Palmisano, G., Masciocchi, N., and Long, J.R. (2011). High thermal and chemical stability in pyrazolate-bridged metal–organic frameworks with exposed metal sites. Chem. Sci. *2*, 1311.

18. An, J., Geib, S.J., and Rosi, N.L. (2009). Cation-triggered drug release from a porous zinc−adeninate metal−organic framework. J. Am. Chem. Soc. *131*, 8376–8377.

19. Feng, D., Gu, Z.Y., Li, J.R., Jiang, H.L., Wei, Z., and Zhou, H.C. (2012). Zirconium-metalloporphyrin PCN-222: mesoporous metal-organic frameworks with ultrahigh stability as biomimetic catalysts. Angew. Chem. Int. Ed. *51*, 10307–10310.

20. Carter, J.H., Han, X., Moreau, F.Y., da Silva, I., Nevin, A., Godfrey, H.G.W., Tang, C.C., Yang, S., and Schröder, M. (2018). Exceptional adsorption and binding of sulfur dioxide in a robust zirconium-based metal–organic framework. J. Am. Chem. Soc. *140*, 15564–15567.

21. Morris, W., Doonan, C.J., Furukawa, H., Banerjee, R., and Yaghi, O.M. (2008). Crystals as molecules: postsynthesis covalent functionalization of zeolitic imidazolate frameworks. J. Am. Chem. Soc. *130*, 12626–12627.

22. Wang, Z., and Cohen, S.M. (2009). Postsynthetic modification of metal–organic frameworks. Chem. Soc. Rev. *38*, 1315–1329.

23. Waller, P.J., Lyle, S.J., Osborn Popp, T.M., Diercks, C.S., Reimer, J.A., and Yaghi, O.M. (2016). Chemical conversion of linkages in covalent organic frameworks. J. Am. Chem. Soc. *138*, 15519–15522.

24. Waller, P.J., Alfaraj, Y.S., Diercks, C.S., Jarenwattananon, N.N., and Yaghi, O.M. (2018). Conversion of imine to oxazole and thiazole linkages in covalent organic frameworks. J. Am. Chem. Soc. *140*, 9099–9103.

25. Lyle, S.J., Osborn Popp, T.M., Waller, P.J., Pei, X., Reimer, J.A., and Yaghi, O.M. (2019). Multistep solid-state organic synthesis of carbamate-linked covalent organic frameworks. J. Am. Chem. Soc. *141*, 11253–11258.

26. Haase, F., Troschke, E., Savasci, G., Banerjee, T., Duppel, V., Dörfler, S., Grundei, M.M.J., Burow, A.M., Ochsenfeld, C., Kaskel, S., and Lotsch, B.V. (2018). Topochemical conversion of an imine- into a thiazole-linked covalent organic framework enabling real structure analysis. Nat. Commun. *9*, 2600.

27. Liu, H., Chu, J., Yin, Z., Cai, X., Zhuang, L., and Deng, H. (2018). Covalent organic frameworks linked by amine bonding for concerted electrochemical reduction of CO_2. Chem *4*, 1696–1709.

28. Li, X., Zhang, C., Cai, S., Lei, X., Altoe, V., Hong, F., Urban, J.J., Ciston, J., Chan, E.M., and Liu, Y. (2018). Facile transformation of imine covalent organic frameworks into ultrastable crystalline porous aromatic frameworks. Nat. Commun. *9*, 2998.

29. Jiang, J., Zhao, Y., and Yaghi, O.M. (2016). Covalent chemistry beyond molecules. J. Am. Chem. Soc. *138*, 3255–3265.

30. Yaghi, O.M., Li, G., and Li, H. (1995). Selective binding and removal of guests in a microporous metal–organic framework. Nature *378*, 703–706.

31. Li, H., Eddaoudi, M., Groy, T.L., and Yaghi, O.M. (1998). Establishing microporosity in open metal−organic frameworks: gas sorption isotherms for Zn(BDC) (BDC = 1,4-benzenedicarboxylate). J. Am. Chem. Soc. *120*, 8571–8572.

32. Li, H., Eddaoudi, M., O'Keeffe, M., and Yaghi, O.M. (1999). Design and synthesis of an exceptionally stable and highly porous metal-organic framework. Nature *402*, 276–279.

33. Rosi, N.L., Eckert, J., Eddaoudi, M., Vodak, D.T., Kim, J., O'Keeffe, M., and Yaghi, O.M. (2003). Hydrogen storage in microporous metal-organic frameworks. Science *300*, 1127–1129.

34. Dinča, M., Dailly, A., Liu, Y., Brown, C.M., Neumann, D.A., and Long, J.R. (2006). Hydrogen storage in a microporous metal-organic framework with exposed Mn^{2+} coordination sites. J. Am. Chem. Soc. *128*, 16876–16883.

35. Hanikel, N., Prévot, M.S., and Yaghi, O.M. (2020). MOF water harvesters. Nat. Nanotechnol. *15*, 348–355.

36. Lee, S., Kapustin, E.A., and Yaghi, O.M. (2016). Coordinative alignment of molecules in chiral metal-organic frameworks. Science *353*, 808–811.

37. Pei, X., Bürgi, H.B., Kapustin, E.A., Liu, Y., and Yaghi, O.M. (2019). Coordinative alignment in the pores of MOFs for the structural determination of N-, S-, and P-containing organic compounds including complex chiral molecules. J. Am. Chem. Soc. *141*, 18862–18869.

38. Sheberla, D., Sun, L., Blood-Forsythe, M.A., Er, S., Wade, C.R., Brozek, C.K., Aspuru-Guzik, A., and Dincă, M. (2014). High electrical conductivity in in Ni_3(2,3,6,7,10,11-hexaiminotriphenylene)$_2$, a semiconducting metal-organic graphene analogue. J. Am. Chem. Soc. *136*, 8859–8862.

39. Ploetz, E., Engelke, H., Lächelt, U., and Wuttke, S. (2020). The chemistry of reticular framework nanoparticles: MOF, ZIF, and COF materials. Adv. Funct. Mater. https://doi.org/10.1002/adfm.201909062.

40. Xu, W., Pei, X., Diercks, C.S., Lyu, H., Ji, Z., and Yaghi, O.M. (2019). A metal–organic framework of organic vertices and polyoxometalate linkers as a solid-state electrolyte. J. Am. Chem. Soc. *141*, 17522–17526.

41. Flaig, R.W., Osborn Popp, T.M., Fracaroli, A.M., Kapustin, E.A., Kalmutzki, M.J., Altamimi, R.M., Fathieh, F., Reimer, J.A., and Yaghi, O.M. (2017). The chemistry of CO_2 capture in an amine-functionalized metal-organic framework under dry and humid conditions. J. Am. Chem. Soc. *139*, 12125–12128.

42. McDonald, T.M., Mason, J.A., Kong, X., Bloch, E.D., Gygi, D., Dani, A., Crocellà, V., Giordanino, F., Odoh, S.O., Drisdell, W.S., et al. (2015). Cooperative insertion of CO_2 in diamine-appended metal-organic frameworks. Nature *519*, 303–308.

43. Eddaoudi, M., Kim, J., Rosi, N., Vodak, D., Wachter, J., O'Keeffe, M., and Yaghi, O.M. (2002). Systematic design of pore size and functionality in isoreticular MOFs and their application in methane storage. Science *295*, 469–472.

44. Peng, Y., Krungleviciute, V., Eryazici, I., Hupp, J.T., Farha, O.K., and Yildirim, T. (2013). Methane storage in metal-organic frameworks: current records, surprise findings, and challenges. J. Am. Chem. Soc. *135*, 11887–11894.

Please cite this article in press as: Lyu et al., Digital Reticular Chemistry, Chem (2020), https://doi.org/10.1016/j.chempr.2020.08.008

45. Chen, Z., Li, P., Anderson, R., Wang, X., Zhang, X., Robison, L., Redfern, L.R., Moribe, S., Islamoglu, T., Gómez-Gualdrón, D.A., et al. (2020). Balancing volumetric and gravimetric uptake in highly porous materials for clean energy. Science *368*, 297–303.

46. Lin, S., Diercks, C.S., Zhang, Y.B., Kornienko, N., Nichols, E.M., Zhao, Y., Paris, A.R., Kim, D., Yang, P., Yaghi, O.M., and Chang, C.J. (2015). Covalent organic frameworks comprising cobalt porphyrins for catalytic CO_2 reduction in water. Science *349*, 1208–1213.

47. Seo, J.S., Whang, D., Lee, H., Jun, S.I., Oh, J., Jeon, Y.J., and Kim, K. (2000). A homochiral metal-organic porous material for enantioselective separation and catalysis. Nature *404*, 982–986.

48. Kim, H., Yang, S., Rao, S.R., Narayanan, S., Kapustin, E.A., Furukawa, H., Umans, A.S., Yaghi, O.M., and Wang, E.N. (2017). Water harvesting from air with metal-organic frameworks powered by natural sunlight. Science *356*, 430–434.

49. Lecun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. Nature *521*, 436–444.

50. Silver, D., Hubert, T., Schrittwieser, J., Antonoglou, I., Lai, M., Guez, A., Lanctot, M., Sifre, L., Kumaran, D., Graepel, T., et al. (2018). A general reinforcement learning algorithm that masters chess, shogi, and Go through self-play. Science *362*, 1140–1144.

51. Alpaydin, E. (2020). Introduction to Machine Learning, Fourth Edition (MIT Press).

52. Duvenaud, D., Maclaurin, D., Aguilera-Iparraguirre, J., Gómez-Bombarelli, R., Hirzel, T., Aspuru-Guzik, A., and Adams, R.P. (2015). Convolutional networks on graphs for learning molecular fingerprints. Proceedings of Advances in Neural Information Processing Systems, *28* (NIPS 2015), pp. 2224–2232.

53. King, R.D., Rowland, J., Oliver, S.G., Young, M., Aubrey, W., Byrne, E., Liakata, M., Markham, M., Pir, P., Soldatova, L.N., et al. (2009). The automation of science. Science *324*, 85–89.

54. Janet, J.P., and Kulik, H.J. (2020). Machine Learning in Chemistry (American Chemical Society).

55. Zdeborová, L. (2017). New tool in the box. Nat. Phys. *13*, 420–421.

56. Gromski, P.S., Henson, A.B., Granda, J.M., and Cronin, L. (2019). How to explore chemical space using algorithms and automation. Nat. Rev. Chem. *3*, 119–128.

57. Furukawa, H., Cordova, K.E., O'Keeffe, M., and Yaghi, O.M. (2013). The chemistry and applications of metal-organic frameworks. Science *341*, 1230444.

58. Moghadam, P.Z., Li, A., Wiggin, S.B., Tao, A., Maloney, A.G.P., Wood, P.A., Ward, S.C., and 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.

59. Diercks, C.S., and Yaghi, O.M. (2017). The atom, the molecule, and the covalent organic framework. Science *355*, eaal1585.

60. Ongari, D., Yakutovich, A.V., Talirz, L., and Smit, B. (2019). Building a consistent and reproducible database for adsorption evaluation in covalent-organic frameworks. ACS Cent. Sci. *5*, 1663–1675.

61. Ockwig, N.W., Delgado-Friedrichs, O., O'Keeffe, M., and Yaghi, O.M. (2005). Reticular chemistry: occurrence and taxonomy of nets and grammar for the design of frameworks. Acc. Chem. Res. *38*, 176–182.

62. O'Keeffe, M., Peskov, M.A., Ramsden, S.J., and Yaghi, O.M. (2008). The Reticular Chemistry Structure Resource (RCSR) database of, and symbols for, crystal nets. Acc. Chem. Res. *41*, 1782–1789.

63. Ramsden, S.J., Robins, V., and Hyde, S.T. (2009). Three-dimensional Euclidean nets from two-dimensional hyperbolic tilings: kaleidoscopic examples. Acta Crystallogr. A Found. Crystallogr. *65*, 81–108.

64. Blatov, V.A., Shevchenko, A.P., and Proserpio, D.M. (2014). Applied topological analysis of crystal structures with the program package topospro. Cryst. Growth Des. *14*, 3576–3586.

65. Alhilaly, M.J., Huang, R.W., Naphade, R., Alamer, B., Hedhili, M.N., Emwas, A.H., Maity, P., Yin, J., Shkurenko, A., Mohammed, O.F., et al. (2019). Assembly of atomically precise silver nanoclusters into nanocluster-based frameworks. J. Am. Chem. Soc. *141*, 9585–9592.

66. Nohra, B., El Moll, H., Rodriguez Albelo, L.M., Mialane, P., Marrot, J., Mellot-Draznieks, C., O'Keeffe, M., Ngo Biboum, R., Lemaire, J., Keita, B., et al. (2011). Polyoxometalate-based metal organic frameworks (POMOFs): structural trends, energetics, and high electrocatalytic efficiency for hydrogen evolution reaction. J. Am. Chem. Soc. *133*, 13363–13374.

67. Rosi, N.L., Kim, J., Eddaoudi, M., Chen, B., O'Keeffe, M., and Yaghi, O.M. (2005). Rod packings and metal-organic frameworks constructed from rod-shaped secondary building units. J. Am. Chem. Soc. *127*, 1504–1518.

68. Schoedel, A., Li, M., Li, D., O'Keeffe, M., and Yaghi, O.M. (2016). Structures of metal-organic frameworks with rod secondary building units. Chem. Rev. *116*, 12466–12535.

69. Zhao, S., Wang, Y., Dong, J., He, C.T., Yin, H., An, P., Zhao, K., Zhang, X., Gao, C., Zhang, L., et al. (2016). Ultrathin metal-organic framework nanosheets for electrocatalytic oxygen evolution. Nat. Energy *1*, 1–10.

70. Deng, H., Grunder, S., Cordova, K.E., Valente, C., Furukawa, H., Hmadeh, M., Gándara, F., Whalley, A.C., Liu, Z., Asahina, S., et al. (2012). Large-pore apertures in a series of metal-organic frameworks. Science *336*, 1018–1023.

71. Wißmann, G., Schaate, A., Lilienthal, S., Bremer, I., Schneider, A.M., and Behrens, P. (2012). Modulated synthesis of Zr-fumarate MOF. Micropor. Mesopor. Mater. *152*, 64–70.

72. Wang, X., Han, X., Zhang, J., Wu, X., Liu, Y., and Cui, Y. (2016). Homochiral 2D porous covalent organic frameworks for heterogeneous asymmetric catalysis. J. Am. Chem. Soc. *138*, 12332–12335.

73. Yang, J., Trickett, C.A., Alahmadi, S.B., Alshammari, A.S., and Yaghi, O.M. (2017). Calcium L-lactate frameworks as naturally degradable carriers for pesticides. J. Am. Chem. Soc. *139*, 8118–8121.

74. Côté, A.P., and Shimizu, G.K.H. (2001). The first example of a functional pillared metal sulfonate network. Chem. Commun. *2001*, 251–252.

75. Shimizu, G.K., Vaidhyanathan, R., and Taylor, J.M. (2009). Phosphonate and sulfonate metal organic frameworks. Chem. Soc. Rev. *38*, 1430–1449.

76. Taylor, J.M., Mahmoudkhani, A.H., and Shimizu, G.K.H. (2007). A tetrahedral organophosphonate as a linker for a microporous copper framework. Angew. Chem. Int. Ed. *46*, 795–798.

77. Hmadeh, M., Lu, Z., Liu, Z., Gándara, F., Furukawa, H., Wan, S., Augustyn, V., Chang, R., Liao, L., Zhou, F., et al. (2012). New porous crystals of extended metal-catecholates. Chem. Mater. *24*, 3511–3513.

78. Huang, X., Sheng, P., Tu, Z., Zhang, F., Wang, J., Geng, H., Zou, Y., Di, C.A., Yi, Y., Sun, Y., et al. (2015). A two-dimensional π–d conjugated coordination polymer with extremely high electrical conductivity and ambipolar transport behaviour. Nat. Commun. *6*, 7408.

79. Park, K.S., Ni, Z., Côté, A.P., Choi, J.Y., Huang, R., Uribe-Romo, F.J., Chae, H.K., O'Keeffe, M., and Yaghi, O.M. (2006). Exceptional chemical and thermal stability of zeolitic imidazolate frameworks. Proc. Natl. Acad. Sci. USA *103*, 10186–10191.

80. He, J., Yin, Y.-G., Wu, T., Li, D., and Huang, X.-C. (2006). Design and solvothermal synthesis of luminescent copper(I)-pyrazolate coordination oligomer and polymer frameworks. Chem. Commun. *2006*, 2845–2847.

81. Zhang, J.-P., Zheng, S.-L., Huang, X.-C., and Chen, X.-M. (2004). Two unprecedented 3-connected three-dimensional networks of copper(I) triazolates: in situ formation of ligands by cycloaddition of nitriles and ammonia. Angew. Chem. Int. Ed. *43*, 206–209.

82. Gándara, F., Uribe-Romo, F.J., Britt, D.K., Furukawa, H., Lei, L., Cheng, R., Duan, X., O'Keeffe, M., and Yaghi, O.M. (2012). Porous, conductive metal-triazolates and their structural elucidation by the charge-flipping method. Chemistry *18*, 10595–10601.

83. Dinca, M., Yu, A.F., and Long, J.R. (2006). Microporous metal−organic frameworks incorporating 1,4-benzeneditetrazolate: syntheses, structures, and hydrogen storage properties. J. Am. Chem. Soc. *128*, 8904–8913.

84. Hunt, J.R., Doonan, C.J., LeVangie, J.D., Côté, A.P., and Yaghi, O.M. (2008). Reticular synthesis of covalent organic borosilicate frameworks. J. Am. Chem. Soc. *130*, 11872–11873.

85. Jiang, S.Y., Gan, S.X., Zhang, X., Li, H., Qi, Q.Y., Cui, F.Z., Lu, J., and Zhao, X. (2019). Aminal-linked covalent organic frameworks through condensation of secondary amine

Please cite this article in press as: Lyu et al., Digital Reticular Chemistry, Chem (2020), https://doi.org/10.1016/j.chempr.2020.08.008

with aldehyde. J. Am. Chem. Soc. *141*, 14981–14986.

86. Uribe-Romo, F.J., Hunt, J.R., Furukawa, H., Klöck, C., O'Keeffe, M., and Yaghi, O.M. (2009). A crystalline imine-linked 3-D porous covalent organic framework. J. Am. Chem. Soc. *131*, 4570–4571.

87. Dalapati, S., Jin, S., Gao, J., Xu, Y., Nagai, A., and Jiang, D. (2013). An azine-linked covalent organic framework. J. Am. Chem. Soc. *135*, 17310–17313.

88. Zhuang, X., Zhao, W., Zhang, F., Cao, Y., Liu, F., Bi, S., and Feng, X. (2016). A two-dimensional conjugated polymer framework with fully sp^2-bonded carbon skeleton. Polym. Chem. *7*, 4176–4181.

89. Jin, E., Asada, M., Xu, Q., Dalapati, S., Addicoat, M.A., Brady, M.A., Xu, H., Nakamura, T., Heine, T., Chen, Q., and Jiang, D. (2017). Two-dimensional sp^2 carbon–conjugated covalent organic frameworks. Science *357*, 673–676.

90. Lyu, H., Diercks, C.S., Zhu, C., and Yaghi, O.M. (2019). Porous crystalline olefin-linked covalent organic frameworks. J. Am. Chem. Soc. *141*, 6848–6852.

91. Uribe-Romo, F.J., Doonan, C.J., Furukawa, H., Oisaki, K., and Yaghi, O.M. (2011). Crystalline covalent organic frameworks with hydrazone linkages. J. Am. Chem. Soc. *133*, 11478–11481.

92. Zhao, C., Diercks, C.S., Zhu, C., Hanikel, N., Pei, X., and Yaghi, O.M. (2018). Urea-linked covalent organic frameworks. J. Am. Chem. Soc. *140*, 16438–16441.

93. Guo, J., Xu, Y., Jin, S., Chen, L., Kaji, T., Honsho, Y., Addicoat, M.A., Kim, J., Saeki, A., Ihee, H., et al. (2013). Conjugated organic framework with three-dimensionally ordered stable structure and delocalized π clouds. Nat. Commun. *4*, 2736.

94. Zhang, B., Wei, M., Mao, H., Pei, X., Alshmimri, S.A., Reimer, J.A., and Yaghi, O.M. (2018). Crystalline dioxin-linked covalent organic frameworks from irreversible reactions. J. Am. Chem. Soc. *140*, 12715–12719.

95. Pyles, D.A., Crowe, J.W., Baldwin, L.A., and McGrier, P.L. (2016). Synthesis of benzobisoxazole-linked two-dimensional covalent organic frameworks and their carbon dioxide capture properties. ACS Macro Lett *5*, 1055–1058.

96. Wang, P.L., Ding, S.Y., Zhang, Z.C., Wang, Z.P., and Wang, W. (2019). Constructing robust covalent organic frameworks via multicomponent reactions. J. Am. Chem. Soc. *141*, 18004–18008.

97. Kuhn, P., Antonietti, M., and Thomas, A. (2008). Porous, covalent triazine-based frameworks prepared by ionothermal synthesis. Angew. Chem. Int. Ed. *47*, 3450–3453.

98. Jackson, K.T., Reich, T.E., and El-Kaderi, H.M. (2012). Targeted synthesis of a porous borazine-linked covalent organic framework. Chem. Commun. *48*, 8823–8825.

99. Allen, L.C. (1989). Electronegativity is the average one-electron energy of the valence-shell electrons in ground-state free atoms. J. Am. Chem. Soc. *111*, 9003–9014.

100. Pearson, W.B. (1958). A Handbook of Lattice Spacings and Structures of Metals and Alloys (Elsevier).

101. Urban, A., Abdellahi, A., Dacek, S., Artrith, N., and Ceder, G. (2017). Electronic-structure origin of cation disorder in transition-metal oxides. Phys. Rev. Lett. *119*, 176402.

102. Brozek, C.K., and Dincă, M. (2012). Lattice-imposed geometry in metal-organic frameworks: lacunary Zn_4O clusters in MOF-5 serve as tripodal chelating ligands for Ni^{2+}. Chem. Sci. *3*, 2110–2113.

103. Kitaĭgorodskiĭ, A.I. (1973). Molecular Crystals and Molecules, First Edition (Academic Press).

104. Deng, H., Doonan, C.J., Furukawa, H., Ferreira, R.B., Towne, J., Knobler, C.B., Wang, B., and Yaghi, O.M. (2010). Multiple functional groups of varying ratios in metal-organic frameworks. Science *327*, 846–850.

105. Wang, L.J., Deng, H., Furukawa, H., Gándara, F., Cordova, K.E., Peri, D., and Yaghi, O.M. (2014). Synthesis and characterization of metal–organic framework-74 containing 2, 4, 6, 8, and 10 different metals. Inorg. Chem. *53*, 5881–5883.

106. Zhang, Y.B., Furukawa, H., Ko, N., Nie, W., Park, H.J., Okajima, S., Cordova, K.E., Deng, H., Kim, J., and Yaghi, O.M. (2015). Introduction of functionality, selection of topology, and enhancement of gas adsorption in multivariate metal-organic framework-177. J. Am. Chem. Soc. *137*, 2641–2650.

107. Dong, Z., Sun, Y., Chu, J., Zhang, X., and Deng, H. (2017). Multivariate metal-organic frameworks for dialing-in the binding and programming the release of drug molecules. J. Am. Chem. Soc. *139*, 14209–14216.

108. Xia, Q., Li, Z., Tan, C., Liu, Y., Gong, W., and Cui, Y. (2017). Multivariate metal-organic frameworks as multifunctional heterogeneous asymmetric catalysts for sequential reactions. J. Am. Chem. Soc. *139*, 8259–8266.

109. Liu, Q., Cong, H., and Deng, H. (2016). Deciphering the spatial arrangement of metals and correlation to reactivity in multivariate metal-organic frameworks. J. Am. Chem. Soc. *138*, 13822–13825.

110. Aguirre-Díaz, L.M., Gándara, F., Iglesias, M., Snejko, N., Gutiérrez-Puebla, E., and Monge, M.Á. (2015). Tunable catalytic activity of solid solution metal-organic frameworks in one-pot multicomponent reactions. J. Am. Chem. Soc. *137*, 6132–6135.

111. Fang, Z., Bueken, B., De Vos, D.E., and Fischer, R.A. (2015). Defect-engineered metal-organic frameworks. Angew. Chem. Int. Ed. *54*, 7234–7254.

112. Canossa, S., Ji, Z., and Wuttke, S. (2020). Circumventing wear and tear of adaptive porous materials. Adv. Funct. Mater. https://doi.org/10.1002/adfm.201908547.

113. Kong, X., Deng, H., Yan, F., Kim, J., Swisher, J.A., Smit, B., Yaghi, O.M., and Reimer, J.A. (2013). Mapping of functional groups in metal-organic frameworks. Science *341*, 882–885.

114. Osborn Popp, T.M., and Yaghi, O.M. (2017). Sequence-dependent materials. Acc. Chem. Res. *50*, 532–534.

115. Ji, Z., Li, T., and Yaghi, O.M. (2020). Sequencing of metals in multivariate metal-organic frameworks. Science *369*, 674–680.

116. Pólya, G. (1937). Kombinatorische anzahlbestimmungen für gruppen, graphen und chemische verbindungen. Acta Math *68*, 145–254.

117. Redfield, J.H. (1927). The theory of group-reduced distributions. Am. J. Math. *49*, 433.

118. Gropp, C., Canossa, S., Wuttke, S., Gándara, F., Li, Q., Gagliardi, L., and Yaghi, O.M. (2020). Standard practices of reticular chemistry. ACS Cent. Sci. https://doi.org/10.1021/acscentsci.0c00592.

119. Raccuglia, P., Elbert, K.C., Adler, P.D.F.F., Falk, C., Wenny, M.B., Mollo, A., Zeller, M., Friedler, S.A., Schrier, J., and Norquist, A.J. (2016). Machine-learning-assisted materials discovery using failed experiments. Nature *533*, 73–76.

120. O'Keeffe, M., Eddaoudi, M., Li, H., Reineke, T., and Yaghi, O.M. (2000). Frameworks for extended solids: geometrical design principles. J. Solid State Chem. *152*, 3–20.

121. Furukawa, H., Müller, U., and Yaghi, O.M. (2015). "Heterogeneity within order" in metal-organic frameworks. Angew. Chem. Int. Ed. *54*, 3417–3430.

122. Jablonka, K.M., Ongari, D., Moosavi, S.M., and Smit, B. (2020). Big-data science in porous materials: materials genomics and machine learning. Chem. Rev. https://doi.org/10.1021/acs.chemrev.0c00004.

123. Chung, Y.G., Camp, J., Haranczyk, M., Sikora, B.J., Bury, W., Krungleviciute, V., Yildirim, T., Farha, O.K., Sholl, D.S., and Snurr, R.Q. (2014). Computation-ready, experimental metal–organic frameworks: a tool to enable high-throughput screening of nanoporous crystals. Chem. Mater. *26*, 6185–6192.

124. Wilmer, C.E., Leaf, M., Lee, C.Y., Farha, O.K., Hauser, B.G., Hupp, J.T., and Snurr, R.Q. (2011). Large-scale screening of hypothetical metal–organic frameworks. Nat. Chem. *4*, 83–89.

125. Tong, M., Lan, Y., Yang, Q., and Zhong, C. (2017). Exploring the structure-property relationships of covalent organic frameworks for noble gas separations. Chem. Eng. Sci. *168*, 456–464.

126. Talirz, L., Kumbhar, S., Passaro, E., Yakutovich, A.V., Granata, V., Gargiulo, F., Borelli, M., Uhrin, M., Huber, S.P., Zoupanos, S., et al. (2020). Materials Cloud, a platform for open computational science. arXiv, arXiv:2003.12510.

127. Bucior, B.J., Rosen, A.S., Haranczyk, M., Yao, Z., Ziebel, M.E., Farha, O.K., Hupp, J.T., Siepmann, J.I., Aspuru-Guzik, A., and Snurr, R.Q. (2019). Identification schemes for metal-organic frameworks to enable rapid search and cheminformatics analysis. Cryst. Growth Des. *19*, 6682–6697.

128. Li, A., Bueno-Perez, R., Wiggin, S., and Fairen-Jimenez, D. (2020). Enabling efficient exploration of metal–organic frameworks in the Cambridge Structural Database. CrystEngComm. https://doi.org/10.1039/D0CE00299B.

129. Goldsmith, J., Wong-Foy, A.G., Cafarella, M.J., and Siegel, D.J. (2013). Theoretical limits of hydrogen storage in metal–organic frameworks: opportunities and trade-offs. Chem. Mater. *25*, 3373–3382.

130. Chung, Y.G., Haldoupis, E., Bucior, B.J., Haranczyk, M., Lee, S., Zhang, H., Vogiatzis, K.D., Milisavljevic, M., Ling, S., Camp, J.S., et al. (2019). Advances, updates, and analytics for the computation-ready, experimental metal-organic framework database: CoRE MOF 2019. J. Chem. Eng. Data *64*, 5985–5998.

131. Ongari, D., Boyd, P.G., Barthel, S., Witman, M., Haranczyk, M., and Smit, B. (2017). Accurate characterization of the pore volume in microporous crystalline materials. Langmuir *33*, 14529–14538.

132. Willems, T.F., Rycroft, C.H., Kazi, M., Meza, J.C., and Haranczyk, M. (2012). Algorithms and tools for high-throughput geometry-based analysis of crystalline porous materials. Micropor. Mesopor. Mater. *149*, 134–141.

133. Fernandez, M., Woo, T.K., Wilmer, C.E., and Snurr, R.Q. (2013). Large-scale quantitative structure-property relationship (QSPR) analysis of methane storage in metal-organic frameworks. J. Phys. Chem. C *117*, 7681–7689.

134. Metal Organic Framework Database | 163037 Mofs. https://mof.tech.northwestern.edu/.

135. Park, S., Kim, B., Choi, S., Boyd, P.G., Smit, B., and Kim, J. (2018). Text mining metal–organic framework papers. J. Chem. Inf. Model. *58*, 244–251.

136. Hawizy, L., Jessop, D.M., Adams, N., and Murray-Rust, P. (2011). ChemicalTagger: a tool for semantic text-mining in chemistry. J. Cheminform. *3*, 17.

137. Townsend, J., Copestake, A., Murray-Rust, P., Teufel, S., and Waudby, C. (2005). Language technology for processing chemistry publications. Proceedings of the Fourth UK e-Science All Hands Meet *17*, 79–80.

138. Tshitoyan, V., Dagdelen, J., Weston, L., Dunn, A., Rong, Z., Kononova, O., Persson, K.A., Ceder, G., and Jain, A. (2019). Unsupervised word embeddings capture latent knowledge from materials science literature. Nature *571*, 95–98.

139. Wang, Q., Mao, Z., Wang, B., and Guo, L. (2017). Knowledge graph embedding: a survey of approaches and applications. IEEE Trans. Knowl. Data Eng. *29*, 2724–2743.

140. Xu, P., and Barbosa, D. (2018). Investigations on knowledge base embedding for relation prediction and extraction. arXiv, arXiv:1802.02114.

141. Öhrström, L., Weiderud, P., Youssef, M.A., and Yaghi, O.M. (2018). Global engagement in science: the university's fourth mission? Science & Diplomacy 7. https://www.sciencediplomacy.org/perspective/2018/global-engagement-in-science-universitys-fourth-mission.

142. Wilkinson, M.D., Dumontier, M., Aalbersberg, I.J.J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L.B., Bourne, P.E., et al. (2016). The FAIR Guiding Principles for scientific data management and stewardship. Sci. Data *3*, 160018.

143. Delgado-Friedrichs, O., and O'Keeffe, M. (2003). Identification of and symmetry computation for crystal nets. Acta Crystallogr. A Found. Crystallogr. *59*, 351–360.

144. Addicoat, M.A., Coupry, D.E., and Heine, T. (2014). AuToGraFS: automatic topological generator for framework structures. J. Phys. Chem. A *118*, 9607–9614.

145. Boyd, P.G., and Woo, T.K. (2016). A generalized method for constructing hypothetical nanoporous materials of any net topology from graph theory. CrystEngComm *18*, 3777–3792.

146. Keupp, J., and Schmid, R. (2018). TopoFF: MOF structure prediction using specifically optimized blueprints. Faraday Discuss. *211*, 79–101.

147. Mercado, R., Fu, R.S., Yakutovich, A.V., Talirz, L., Haranczyk, M., and Smit, B. (2018). In silico design of 2D and 3D covalent organic frameworks for methane storage applications. Chem. Mater. *30*, 5069–5086.

148. Lukose, B., Kuc, A., Frenzel, J., and Heine, T. (2010). On the reticular construction concept of covalent organic frameworks. Beilstein J. Nanotechnol. *1*, 60–70.

149. Lin, L.C., Berger, A.H., Martin, R.L., Kim, J., Swisher, J.A., Jariwala, K., Rycroft, C.H., Bhown, A.S., Deem, M.W., Haranczyk, M., and Smit, B. (2012). In silico screening of carbon-capture materials. Nat. Mater. *11*, 633–641.

150. Simon, C.M., Kim, J., Gomez-Gualdron, D.A., Camp, J.S., Chung, Y.G., Martin, R.L., Mercado, R., Deem, M.W., Gunter, D., Haranczyk, M., et al. (2015). The materials genome in action: identifying the performance limits for methane storage. Energy Environ. Sci. *8*, 1190–1199.

151. Fu, J., Liu, Y., Tian, Y., and Wu, J. (2015). Density functional methods for fast screening of metal-organic frameworks for hydrogen storage. J. Phys. Chem. C *119*, 5374–5385.

152. Nazarian, D., Camp, J.S., and Sholl, D.S. (2016). A comprehensive set of high-quality point charges for simulations of metal–organic frameworks. Chem. Mater. *28*, 785–793.

153. Hübner, O., Glöss, A., Fichtner, M., and Klopper, W. (2004). On the interaction of dihydrogen with aromatic systems. J. Phys. Chem. A *108*, 3019–3023.

154. Moghadam, P.Z., Rogge, S.M.J., Li, A., Chow, C.M., Wieme, J., Moharrami, N., Aragones-Anglada, M., Conduit, G., Gomez-Gualdron, D.A., Van Speybroeck, V., and Fairen-Jimenez, D. (2019). Structure-mechanical stability relations of metal-organic frameworks via machine learning. Matter *1*, 219–234.

155. He, Y., Cubuk, E.D., Allendorf, M.D., and Reed, E.J. (2018). Metallic metal-organic frameworks predicted by the combination of machine learning methods and *ab initio* calculations. J. Phys. Chem. Lett. *9*, 4562–4569.

156. Moghadam, P.Z., Islamoglu, T., Goswami, S., Exley, J., Fantham, M., Kaminski, C.F., Snurr, R.Q., Farha, O.K., and Fairen-Jimenez, D. (2018). Computer-aided discovery of a metal–organic framework with superior oxygen uptake. Nat. Commun. *9*, 1378.

157. Cortes, C., and Vapnik, V. (1995). Support-vector networks. Mach. Learn. *20*, 273–297.

158. Fernandez, M., Boyd, P.G., Daff, T.D., Aghaji, M.Z., and Woo, T.K. (2014). Rapid and accurate machine learning recognition of high performing metal organic frameworks for CO_2 capture. J. Phys. Chem. Lett. *5*, 3056–3060.

159. Ho, T.K. (1995). Random decision forests. In Proceedings of 3rd International Conference on Document Analysis and Recognition (IEEE Computer Society)), pp. 278–282.

160. Pardakhti, M., Moharreri, E., Wanik, D., Suib, S.L., and Srivastava, R. (2017). Machine learning using combined structural and chemical descriptors for prediction of methane adsorption performance of metal organic frameworks (MOFs). ACS Comb. Sci. *19*, 640–645.

161. Sadeghi, J., Sadeghi, S., and Niaki, S.T.A. (2014). Optimizing a hybrid vendor-managed inventory and transportation problem with fuzzy demand: an improved particle swarm optimization algorithm. Inf. Sci. *272*, 126–144.

162. Gustafson, J.A., and Wilmer, C.E. (2019). Intelligent selection of metal-organic framework arrays for methane sensing via genetic algorithms. ACS Sens *4*, 1586–1593.

163. Altman, N.S. (1992). An introduction to Kernel and nearest-neighbor nonparametric regression. The American Statistician *46*, 175–185.

164. Shen, X., Zhang, T., Broderick, S., and Rajan, K. (2018). Correlative analysis of metal organic framework structures through manifold learning of Hirshfeld surfaces. Mol. Syst. Des. Eng. *3*, 826–838.

165. Schmidhuber, J. (2015). Deep learning in neural networks: an overview. Neural Netw *61*, 85–117.

166. Yıldız, Z., and Uzun, H. (2015). Prediction of gas storage capacities in metal organic frameworks using artificial neural network. Micropor. Mesopor. Mater. *208*, 50–54.

167. Aspuru-Guzik, A., Baik, M.H., Balasubramanian, S., Banerjee, R., Bart, S., Borduas-Dedekind, N., Chang, S., Chen, P., Corminboeuf, C., Coudert, F.X., et al. (2019). Charting a course for chemistry. Nat. Chem. *11*, 286–294.

168. Rustad, J.R. (2010). Theoretical and Computational Inorganic Chemistry (Academic Press).

169. Ramsahye, N.A., Maurin, G., Bourrelly, S., Llewellyn, P.L., Devic, T., Serre, C., Loiseau, T., and Ferey, G. (2007). Adsorption of CO_2 in metal organic frameworks of different metal centres: grand canonical Monte Carlo simulations compared to experiments. Adsorption *13*, 461–467.

170. Boyd, P.G., Chidambaram, A., García-Díez, E., Ireland, C.P., Daff, T.D., Bounds, R., Gładysiak, A., Schouwink, P., Moosavi, S.M., Maroto-Valer, M.M., et al. (2019). Data-driven design of metal–organic frameworks for wet flue gas CO_2 capture. Nature *576*, 253–256.

171. Pantatosaki, E., Psomadopoulos, D., Steriotis, T., Stubos, A.K., Papaioannou, A., and Papadopoulos, G.K. (2004). Micropore size distributions from CO_2 using grand canonical Monte Carlo at ambient temperatures: cylindrical versus slit pore geometries. Colloids Surf. A Physicochem. Eng. Asp. *241*, 127–135.

172. Oberkampf, W.L., and Barone, M.F. (2006). Measures of agreement between computation and experiment: validation metrics. J. Comput. Phys. *217*, 5–36.

173. Granda, J.M., Donina, L., Dragone, V., Long, D.L., and Cronin, L. (2018). Controlling an organic synthesis robot with machine learning to search for new reactivity. Nature *559*, 377–381.

174. Coley, C.W., Green, W.H., and Jensen, K.F. (2018). Machine learning in computer-aided synthesis planning. Acc. Chem. Res. *51*, 1281–1289.

175. Ahneman, D.T., Estrada, J.G., Lin, S., Dreher, S.D., and Doyle, A.G. (2018). Predicting reaction performance in C–N cross-coupling using machine learning. Science *360*, 186–190.

176. Tarca, A.L., Carey, V.J., Chen, X.W., Romero, R., and Drăghici, S. (2007). Machine learning and its applications to biology. PLoS Comput. Biol. *3*, e116.

177. Stevens, R.C. (2000). High-throughput protein crystallization. Curr. Opin. Struct. Biol. *10*, 558–563.

178. Forster, P.M., Stock, N., and Cheetham, A.K. (2005). A high-throughput investigation of the role of pH, temperature, concentration, and time on the synthesis of hybrid inorganic-organic materials. Angew. Chem. Int. Ed. *44*, 7608–7611.

179. Butler, K.T., Davies, D.W., Cartwright, H., Isayev, O., and Walsh, A. (2018). Machine learning for molecular and materials science. Nature *559*, 547–555.

180. Kusne, A.G., Gao, T., Mehta, A., Ke, L., Nguyen, M.C., Ho, K.M., Antropov, V., Wang, C.Z., Kramer, M.J., Long, C., and Takeuchi, I. (2014). On-the-fly machine-learning for high-throughput experiments: search for rare-earth-free permanent magnets. Sci. Rep. *4*, 6367.

181. Burger, B., Maffettone, P.M., Gusev, V.V., Aitchison, C.M., Bai, Y., Wang, X., Li, X., Alston, B.M., Li, B., Clowes, R., et al. (2020). A mobile robotic chemist. Nature *583*, 237–241.

182. Banerjee, R., Phan, A., Wang, B., Knobler, C., Furukawa, H., O'Keeffe, M., and Yaghi, O.M. (2008). High-throughput synthesis of zeolitic imidazolate frameworks and application to CO_2 capture. Science *319*, 939–943.

183. Moosavi, S.M., Chidambaram, A., Talirz, L., Haranczyk, M., Stylianou, K.C., and Smit, B. (2019). Capturing chemical intuition in synthesis of metal-organic frameworks. Nat. Commun. *10*, 539.

184. Potyrailo, R.A., and Takeuchi, I. (2005). Role of high-throughput characterization tools in combinatorial materials science. Meas. Sci. Technol. *16*, 1–4.

185. Ong, S.P., Richards, W.D., Jain, A., Hautier, G., Kocher, M., Cholia, S., Gunter, D., Chevrier, V.L., Persson, K.A., and Ceder, G. (2013). Python Materials Genomics (pymatgen): a robust, open-source python library for materials analysis. Comput. Mater. Sci. *68*, 314–319.

186. Jain, A., Ong, S.P., Hautier, G., Chen, W., Richards, W.D., Dacek, S., Cholia, S., Gunter, D., Skinner, D., Ceder, G., and Persson, K.A. (2013). Commentary: the materials project: a materials genome approach to accelerating materials innovation. APL Mater *1*, 011002.

187. Littmann, M., Selig, K., Cohen-Lavi, L., Frank, Y., Hönigschmid, P., Kataka, E., Mösch, A., Qian, K., Ron, A., Schmid, S., et al. (2020). Validity of machine learning in biology and medicine increased through collaborations across fields of expertise. Nat. Mach. Intell. *2*, 18–24.

188. Desiraju, G.R. (2002). Cryptic crystallography. Nat. Mater. *1*, 77–79.

189. Woodley, S.M., and Catlow, R. (2008). Crystal structure prediction from first principles. Nat. Mater. *7*, 937–946.

190. Meredig, B., Agrawal, A., Kirklin, S., Saal, J.E., Doak, J.W., Thompson, A., Zhang, K., Choudhary, A., and Wolverton, C. (2014). Combinatorial screening for new materials in unconstrained composition space with machine learning. Phys. Rev. B Condens. Matter. *89*, 094104.

191. Dybtsev, D.N., Chun, H., and Kim, K. (2004). Rigid and flexible: a highly porous metal–organic framework with unusual guest-dependent dynamic behavior. Angew. Chem. Int. Ed. *43*, 5033–5036.

192. Liu, Y., Ma, Y., Zhao, Y., Sun, X., Gándara, F., Furukawa, H., Liu, Z., Zhu, H., Zhu, C., Suenaga, K., et al. (2016). Weaving of organic threads into a crystalline covalent organic framework. Science *351*, 365–369.

193. Reineke, T.M., Eddaoudi, M., Moler, D., O'Keeffe, M., and Yaghi, O.M. (2000). Large free volume in maximally interpenetrating networks: the role of secondary building units exemplified by $Tb_2(ADB)_3[(CH_3)_2SO]_4 \cdot 16[(CH_3)_2SO]$. J. Am. Chem. Soc. *122*, 4843–4844.

194. Fairen-Jimenez, D., Moggach, S.A., Wharmby, M.T., Wright, P.A., Parsons, S., and Düren, T. (2011). Opening the gate: framework flexibility in ZIF-8 explored by experiments and simulations. J. Am. Chem. Soc. *133*, 8900–8902.

195. Aris, R. (1965). Prolegomena to the rational analysis of systems of chemical reactions. Arch. Rational Mech. Anal. *19*, 81–99.

196. Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nat. Mach. Intell. *1*, 206–215.

197. Doshi-Velez, F., and Kim, B. (2017). Towards a rigorous science of interpretable machine learning. arXiv, arxiv.org/abs/1702.08608.

198. Trickett, C.A., Helal, A., Al-Maythalony, B.A., Yamani, Z.H., Cordova, K.E., and Yaghi, O.M. (2017). The chemistry of metal-organic frameworks for CO_2 capture, regeneration and conversion. Nat. Rev. Mater. *2*, 17045.

199. Ding, M., Flaig, R.W., Jiang, H.L., and Yaghi, O.M. (2019). Carbon capture and conversion using metal-organic frameworks and MOF-based materials. Chem. Soc. Rev. *48*, 2783–2828.

200. Oschatz, M., and Antonietti, M. (2018). A search for selectivity to enable CO_2 capture with porous adsorbents. Energy Environ. Sci. *11*, 57–70.

201. Assi, H., Mouchaham, G., Steunou, N., Devic, T., and Serre, C. (2017). Titanium coordination compounds: from discrete metal complexes to metal–organic frameworks. Chem. Soc. Rev. *46*, 3431–3452.

202. Nguyen, H.L. (2017). The chemistry of titanium-based metal–organic frameworks. New J. Chem. *41*, 14030–14043.

203. Castells-Gil, J., Padial, M.N., Almora-Barrios, N., Da Silva, I., Mateo, D., Albero, J., García, H., and Martí-Gastaldo, C. (2019). *De novo* synthesis of mesoporous photoactive titanium(IV)-organic frameworks with MIL-100 topology. Chem. Sci. *10*, 4313–4321.

204. Ji, Z., Wang, H., Canossa, S., Wuttke, S., and Yaghi, O.M. (2020). Pore chemistry of metal–organic frameworks. Adv. Funct. Mater. https://doi.org/10.1002/adfm.202000238.

205. François-Lavet, V., Henderson, P., Islam, R., Bellemare, M.G., and Pineau, J. (2018). An introduction to deep reinforcement learning. Found. Trends Mach. Learn. *11*, 219–354.

206. Massova, I., and Mobashery, S. (1998). Kinship and diversification of bacterial penicillin-binding proteins and β-lactamases. Antimicrob. Agents Chemother. *42*, 1–17.

207. Risso, V.A., Gavira, J.A., Mejia-Carmona, D.F., Gaucher, E.A., and Sanchez-Ruiz, J.M. (2013). Hyperstability and substrate promiscuity in laboratory resurrections of precambrian β-lactamases. J. Am. Chem. Soc. *135*, 2899–2902.

208. Kohonen, T. (1982). Self-organized formation of topologically correct feature maps. Biol. Cybern. *43*, 59–69.

209. Jeske, L., Placzek, S., Schomburg, I., Chang, A., and Schomburg, D. (2019). BRENDA in 2019: a European ELIXIR core data resource. Nucleic Acids Res. *47*, D542–D549.

210. Brenda: The Comprehensive Enzyme Information System. Release 2020.2 (July 2020). https://www.brenda-enzymes.org/.