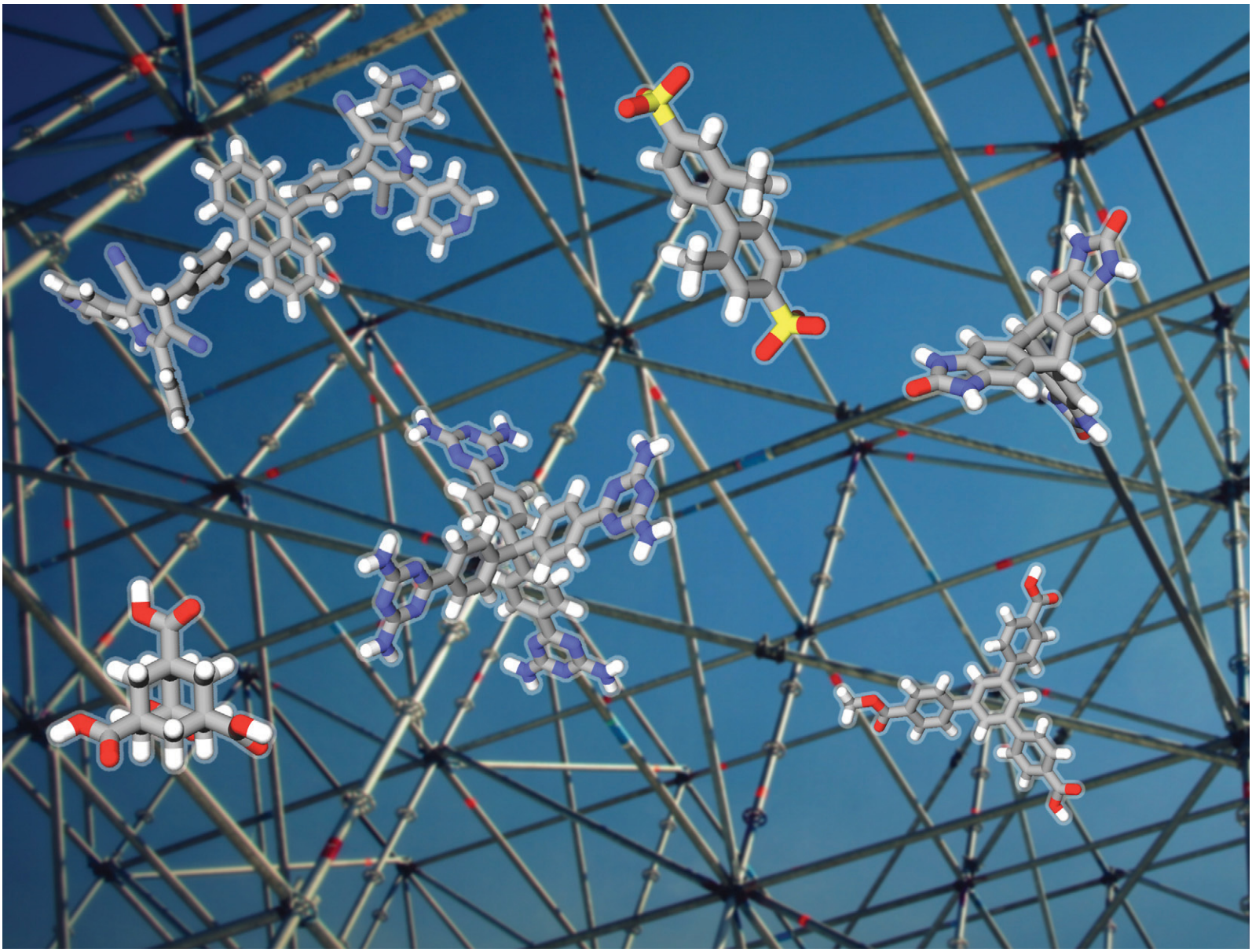
Volume 6
Number 10
October 2021
Pages 751–834

MSDE

Molecular Systems Design & Engineering

rsc.li/molecular-engineering



ISSN 2058-9689

ROYAL SOCIETY OF CHEMISTRY

REVIEW ARTICLE
Michael D. Ward *et al.*
Hydrogen bonded frameworks: smart materials used smartly

IChemE ADVANCING CHEMICAL ENGINEERING WORLDWIDE

EXHIBIT 2009
Novartis v. The Regents of Univ. of MI
IPR2023-01346

MSDE

REVIEW

Check for updates

Cite this: *Mol. Syst. Des. Eng.*, 2021, **6**, 756

Hydrogen bonded frameworks: smart materials used smartly

Anna Yusov, Alexandra M. Dillon and Michael D. Ward *

Received 14th May 2021,
Accepted 12th July 2021

DOI: 10.1039/d1me00055a

rsc.li/molecular-engineering

Hydrogen-bonded frameworks (HBFs) have been studied for decades owing to their fascinating and diverse architectures, always with an eye toward the role of hydrogen bonding in their design as well as their utility in various applications. This review addresses recent advances in HBFs that illustrate their versatility and utility stemming from their unique attributes compared with other classes of molecular frameworks. Guanidinium organosulfonate hydrogen-bonded frameworks, pioneered in our lab and one of the most extensive and versatile collections of HBFs, are used to illustrate molecular design concepts and the principle of architectural isomerism that expands access to a greater structural landscape. Recognizing the growing role of computation in materials design, from *ab initio* methods to machine learning, this review also touches on their emerging use in the design and synthesis of HBFs. The growth of the HBF arsenal promises continuing innovations, with applications ranging from electronic materials and chemical separations to gas adsorption and catalysis.

Design, System, Application

This review centers on hydrogen-bonded frameworks, a growing and versatile class of noncovalent materials created through molecular design and a sophisticated understanding of the structure–property relationships that lead to their use for a variety of applications. Representative applications that illustrate the rational design of hydrogen-bonded frameworks include nonlinear optics, magnetism, molecular structure determination, gas adsorption, chemical separations, and catalysis. Emerging computational and machine learning approaches for accelerating the discovery and design of new hydrogen-bonded frameworks are also on the horizon.

Department of Chemistry and Molecular Design Institute, New York University, 100 Washington Square East, Room 1001, New York, New York 10003, USA.
E-mail: mdw3@nyu.edu

Introduction

The first published discussion of what would later be termed "hydrogen bonds" can be attributed to Moore and Winmill in 1912, who invoked hydrogen bonds to explain the weaker

Anna Yusov

Anna Yusov received her B.A. in Chemistry from the Barnard College at Columbia University, and is currently a PhD candidate supervised by Profs. Michael Ward and Bart Kahr at New York University. Her PhD research is focused on the study of guanidinium organosulfonate hydrogen bonded frameworks, examined by machine learning techniques, as well as their many applications including structure determination and charge transfer complex formation.

Alexandra Dillon

Alexandra Dillon received her B. A. in Chemistry from Skidmore College, and is currently a PhD candidate in the Chemistry Department at New York University, advised by Prof. Michael D. Ward. Her PhD research is centered around a class of hydrogen-bonded guanidinium organosulfonate frameworks, focusing on their use as host materials for conformational analysis of reactive substrates, as well as creating new chiral host frameworks for use in chiral separations.

This journal is © The Royal Society of Chemistry and IChemE 2021

basicity of trimethylammonium hydroxide compared with tetramethylammonium hydroxide.[1] Although hydrogen bonding is often associated with aqueous systems, its role in water was not published until 1920 by Latimer and Rodebush, who attributed apparent water molecule "aggregation" to "the hydrogen nucleus held between 2 octets [constituting] a weak "bond".[2] In 1928, Pauling invoked the concept of "the shared-electron chemical bond", which led to hydrogen bonding becoming an essential descriptor for the structure and properties of water.[3]

In 1930, West proposed that the X-ray analysis of potassium dihydrogen phosphate could be explained by placing hydrogen atoms on a line between two oxygen atoms of adjacent phosphate groups.[4] Three years later, Zachariasen suggested this same idea for the position of hydrogen in atoms in sodium bicarbonate, citing the presence of a hydrogen bond,[5] and in a 1935 publication on liquid methanol stated that "if we wish to characterize the nature of these hydrogen bonds, we should employ the term dipole bonding". This was likely the first suggestion that the hydrogen bond arises from strong dipole interactions.[6]

Examples of hydrogen bonding in the organic solid state gathered throughout the 30s and 40s were well summarized in 1951 by Donohue, who surveyed dozens of structures to gather data regarding N–H···N, N–H···Cl, N–H···O, O–H···N and O–H···O hydrogen bond lengths and angles, clearly demonstrating their versatility.[7] Of course, the role of hydrogen bonding between base pairs was critical to the determination of the crystal structure of DNA in 1953, which can be credited to the contributions from Chargaff,[8,9] Franklin,[10] Wilkins,[11] Watson and Crick.[12]

Following these early reports, the definition of the hydrogen bond expanded. In their 1960 publication, Pimentel and McClellan defined hydrogen bonds more broadly, stating that hydrogen bonds can exist between a "functional group A–H and an atom or a group of atoms B in the same or different molecule when (i) there is evidence of bond formation (association or chelation) and (ii) there is evidence that this new bond linking A–H and B specifically involves the hydrogen atom already bonded to A".[13] In 1963, Sutor reported a survey of crystals with short C–H···O contacts, referring to these as hydrogen bonds.[14] In 1982, Taylor and Kennard provided further evidence for hydrogen bonding in crystals through a systematic analysis of 113 neutron diffraction crystal structures from the Cambridge Structural Database.[15] Their analysis of C–H···X bond distances led to the conclusion that these interactions were attractive rather than repulsive when X = O, N or Cl, consistent with hydrogen bonding. Subsequently, the criteria for hydrogen bonds was reinforced by an analysis of hydrogen bond angles, in addition to distances, defining conical zones within which hydrogen bonds could be assigned in the solid state.[16] Importantly, Taylor and Kennard also noted that "charge assisted" hydrogen bonds were shorter, and therefore likely stronger, than uncharged hydrogen bonds.[16] This is illustrated by the so-called "Speakman salts", compounds which contain hydrogen bonds between a carboxylic acid and a charged carboxylate species.[17] Owing to the strength of charge-assisted hydrogen bonding, using these as structural features of crystalline materials can be a key design consideration.[18]

As early as 1980, Dauber and Hagler acknowledged the importance of hydrogen bonding in influencing packing motifs as well as space group symmetries, noting the implications for the challenging task of crystal structure prediction.[19] Etter later published rules for hydrogen bonding in molecular crystals that provided priority rankings of hydrogen bond pairings in the solid state based on a comprehensive set of observations. These rules significantly advanced the understanding of hydrogen bonding in the organic solid state and created a foundation for predicting molecular aggregation in crystals.[20] These seminal reports eventually led to the use of hydrogen bonds as a design tool for the crystalline solid state, for example, hydrogen-bonded one-dimensional "tapes",[21] and supramolecular "strands"[22] and "ribbons" (Fig. 1).[23] In 1988 Ermer reported the crystal structure of adamantane-1,3,5,7-tetracarboxylic acid, which crystallized into a five-fold interpenetrated diamondoid network, each network stitched together by hydrogen bonds between tetrahedrally disposed carboxylic acid groups.[24] In a subsequent publication, Ermer revealed that functional groups on the adamantane core served as steric blockers that frustrated interpenetration,[25] enabling formation of a low-density two-fold interpenetrated network from 2,6-dimethylideneadamantane-1,3,5,7-tetracarboxylic acid framework accompanied by inclusion of a variety of guest molecules in void spaces.[26] This elegant discovery illustrated that inclusion compounds could be obtained by empirical design of 3D hydrogen-bonded frameworks through reliance on the symmetry and directional hydrogen-bonding of the

Michael D. Ward

Michael D. Ward received his B. S. degree in Chemistry from the William Paterson College of New Jersey and his Ph.D. from Princeton University, after which he was a postdoctoral research fellow at the University of Texas, Austin. He worked as a research scientist at Standard Oil of Ohio and at DuPont Central Research and Development Laboratories, joining the faculty of the Department of Chemical Engineering and Materials Science at the University of Minnesota in 1990. He joined New York University as a Professor of Chemistry in 2006, establishing the Molecular Design Institute. Ward's research group has interests in organic solid-state chemistry, crystal growth, polymorphism, and biominerals in disease.

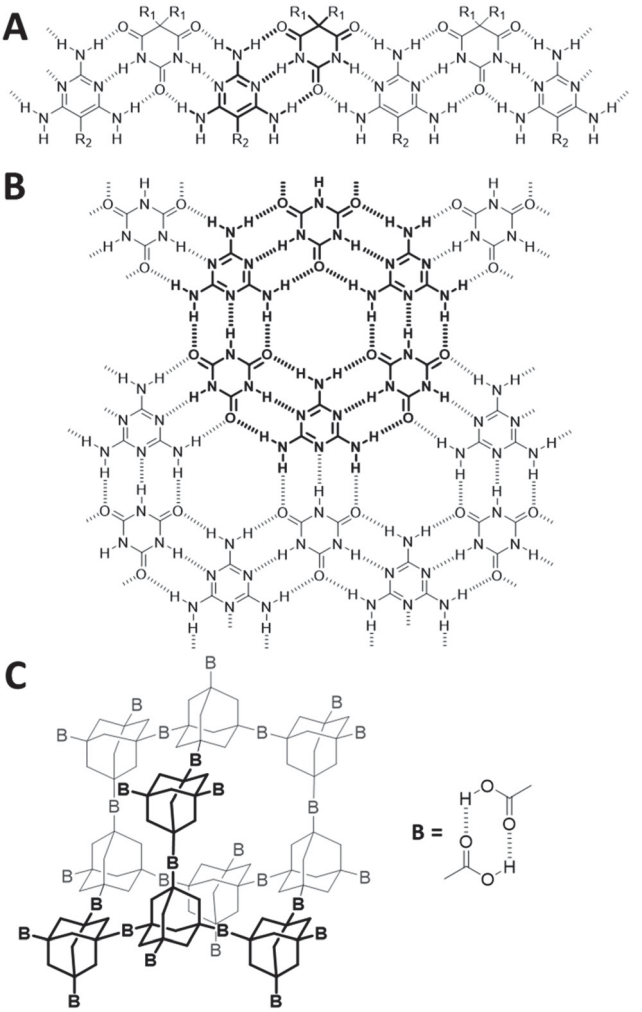


Fig. 1 (A) Schematic representation of a one-dimensional hydrogen bonded tapes. (B) Schematic representation of a two-dimensional hydrogen-bonded sheet.[31] (C) Schematic representation of a three-dimensional network assembled from adamantanetetracarboxylic acid.

self-assembling constituents. Shortly afterward in 1989, "infinite polymeric frameworks" that emulated the adamantane frameworks, but were based on coordination of tetracyanotetraphenylmethane to cuprous ions, were reported.[27] This can be viewed as a progenitor of the now ubiquitous metal–organic frameworks (MOFs), a term first coined by Yaghi in 1995.[28] Other early examples of 3D hydrogen-bonded frameworks include diamondoid networks generated from adamantane-like molecules reported in 1991 by Wuest,[29] who noted their resemblance to inorganic zeolites and the potential for frameworks with large void volumes and adjustable microporosity.[30] These compounds, however, lacked the structural integrity of zeolites as the frameworks were stable only when their voids were occupied by guest molecules. Nonetheless, they illustrated the promise of low-density HBFs, whether for inclusion of guest molecules or as all-organic crystalline materials with persistent porosity.

The International Union of Pure and Applied Chemistry (IUPAC) defines porosity as "a concept related to texture, referring to the pore space in a material",[32] which leaves plenty of room for interpretation as to whether a material is "porous". Porosity can be *intrinsic*, wherein pores are formed from building blocks that already contain voids on their own, or *extrinsic*, wherein pores are a product of the packing of these building blocks.[33] Intrinsic porosity can be exemplified by crystalline materials incorporating calixarenes[34–36] or cyclodextrin cavitands.[37,38] The hydrogen-bonded frameworks described herein are primarily limited to the extrinsic category, some with pores formed with inclusion of guest molecules (a.k.a. inclusion compounds) and others with persistent porosity in the absence of guests. This distinction has been nicely addressed, describing the former as "conventional" porosity and the latter as "virtual" porosity, while providing clear guidance with respect to the proper description of porosity.[39] Porosity often is ill-defined, sometimes conflating persistent porosity with guest exchange, wherein guest molecules are exchanged with guests in the external medium. In most of these cases, exchange actually occurs through depletion of the guest in the inclusion compound, accompanied by collapse of the surrounding framework, then followed by nucleation and growth of the new host–guest composition. Using accepted terminology, hydrogen-bonded frameworks with permanent and persistent porosity can be regarded as "microporous".[40] IUPAC defines micropores as pores with width not exceeding about 2.0 nm.[32]

Although the pursuit of HBFs with true persistent porosity is a worthy goal, this remains rare and technically challenging owing to the tendency of these frameworks to collapse to more dense phases. This is a consequence of non-covalent bonding in these materials, in contrast with zeolites, metalloorganic frameworks, and emerging covalent organic frameworks. The term "HOF", short for hydrogen-bonded organic framework, was recently coined to describe hydrogen-bonded frameworks with persistent porosity (*vide infra*), in the absence of otherwise void-filling guest molecules that typically confer framework stability. Unfortunately, this term does not distinguish between hydrogen-bonded frameworks with persistent pores and those with "virtual pores" that crystallize with interpenetration or with guest incorporation. In fact, the first "HOF" was designated as "HOF-1" in 2011,[41] but this same framework was reported by Wuest in 1997.[42] This nomenclature has the potential to create confusion and mislead newcomers to conclude that hydrogen-bonded frameworks, in general, are a new class of materials. While permanent porosity certainly is essential, and even attractive, for many applications, the overarching challenge of sustaining true porosity in HBFs suggests that attention may be better focused on their unique attributes compared with their metal–organic framework (MOFs) counterparts, for which persistent porosity has been amply demonstrated. These advantages include (i) reversibility of HBF framework assembly, which typically assures a high degree of crystallinity; (ii) a substantial molecular toolkit available through the versatility of organic synthesis; (iii) inherent

 This journal is © The Royal Society of Chemistry and IChemE 2021

flexibility of intermolecular hydrogen bonds that can allow for greater stability and adaptability of framework architecture; (iv) the potential low density of HBF frameworks, which allows for inclusion of a substantial amount of guest molecules, often 50% (or more); (v) potential for systematic manipulation of solid-state properties through the choice of guest molecules. The purpose of this review is to illustrate design pathways for hydrogen-bonded frameworks, particularly those that capture guest molecules during self-assembly rather than relying on adsorption of guests by frameworks with permanent porosity, a property already well demonstrated for MOFs and COFs. Given their distinct characteristics, it is reasonable to pursue the design of inclusion compounds based on HBFs, wherein solid-state architecture is imposed by the framework, and properties and function are tuned systematically through judicious choice of guest molecules – smart materials used smartly.

A case study for molecular design of HBFs

In 1994, our laboratory reported crystalline materials based on a persistent 2D network consisting of guanidinum cations (**G**, $C(NH_2)_3^+$) and a wide range of organomonosulfonate anions (**S**, RSO_3^-),[43] formed by charge-assisted N–H···O hydrogen bonds between 6 protons of **G** and 6 H-bond acceptors of **S** (Fig. 2). The resulting 2D hydrogen bonded sheet network, which can be described generally as "quasi-hexagonal", is a signature of the three-fold symmetry of these ions.[43,44] The "quasi-hexagonal" GS motif is identical to that observed for guanidium nitrate,[45] but the extra valency of the sulfonate ion introduces organic residues on the sulfonate nodes that cement the layers along the third dimension to generate layered architectures. Notably, the GS motif can be dissected into quasi-hexagonal 1D ribbons wherein hydrogen bonds between adjacent ribbons act as a hinge, allowing the sheets to "pucker" at an angle, denoted by θ_{IR}.[46] This puckering is essential to the persistence of the GS sheet for a wide range of sulfonate residues, as it allows the framework to adjust its volume to an optimum packing density. The persistence of this key structural feature has enabled our laboratory, and others, to create a substantial number of GS compounds based on a wide range of framework components and framework architectures, with a remarkable propensity to include guest molecules by design and regulate solid-state function and properties.

Architectural isomerism and guest templating

In 1986, Etter reported "stereoisomeric hydrogen-bonded supermolecules" defined by two different 2D hydrogen bonding motifs of 1,3-cyclohexanedione molecules, appearing in two different crystalline forms, one a guest-free form comprising hydrogen-bonded ribbons and the other an inclusion compound in which a hydrogen-bonded cyclamer motif surrounded a benzene guest. This observation, which suggested the cyclamer was templated by the benzene guest during crystallization, can be viewed as an early account of architectural isomerism in hydrogen-bonded frameworks.[47]

Likewise, guest-free guanidinium organomonosulfonates (GMS) displayed two different architectures, depending on the orientation of the organic residues, which can project from the same side or from opposite sides (Fig. 3). The architecture observed depends on the size and "footprint" of the host. For example (**G**)(2-naphthalenesulfonate) affords a guest-free bilayer architecture in which organic residues are

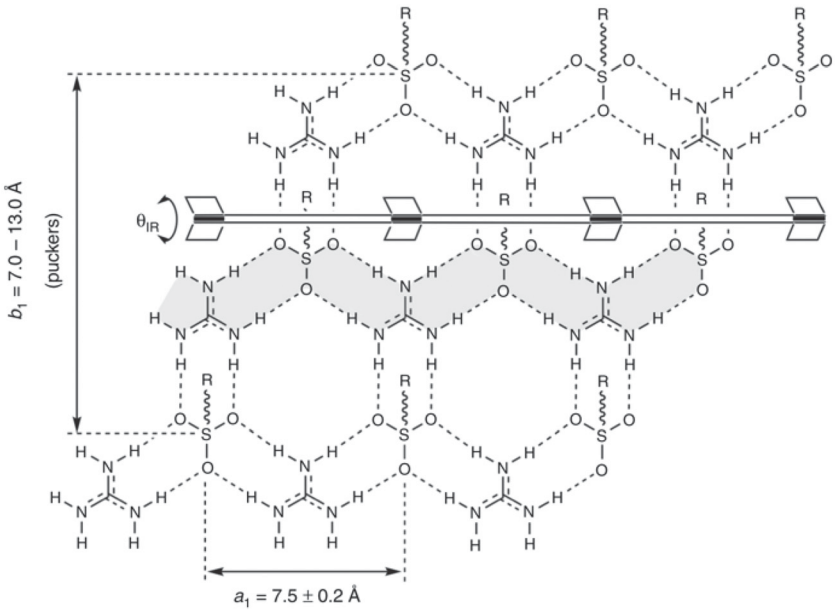


Fig. 2 The quasihexagonal hydrogen bonded GS sheet, which can be viewed as guanidinium-sulfonate ribbons (gray shading) connected by a hydrogen-bonding hinge that permits puckering of the sheet about the hinge over a wide range of angles, denoted as θ_{IR}.

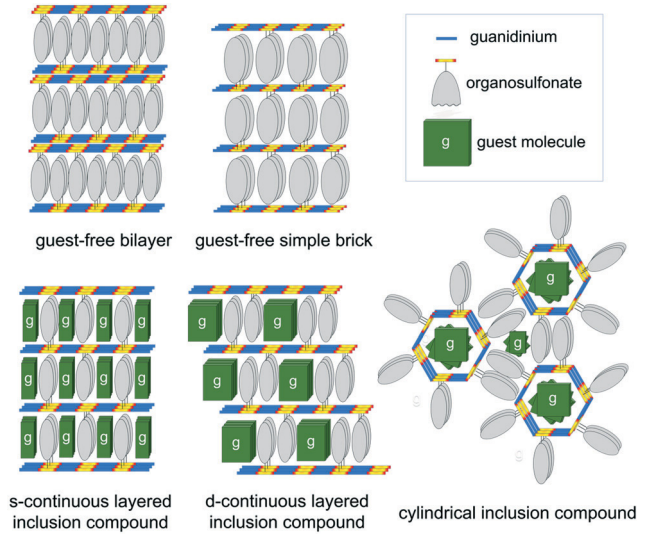


Fig. 3 Architectures for guest-free and guest-containing GS frameworks derived from guanidinium organomonosulfonate hosts.

interdigitated to provide a dispersive glue, but (**G**)(1-naphthalenesulfonate) forms the simple brick architecture as its differently shaped footprint precludes interdigitation as a bilayer. Instead, flipping one half of the organic residues to the opposite side of each sheet creates more volume to allow continuous interdigitation along the stack. Notably, the GS sheet exhibits puckering in the simple brick architecture, which permits the framework to achieve the packing density required for crystallization, but with retention of the quasi-hexagonal hydrogen-bonding connectivity, illustrating that structural robustness can be a consequence of structural compliance rather than rigidity.

In the case of guest-containing GS inclusion compounds, however, the architecture depends on the guest shape and the combined volumes of the sulfonate substituent and guests. The observation of guest inclusion by GMS compounds initially was surprising because inclusion cavities are not preordained. Moreover, guest-free GMS compounds form readily.[48] Nevertheless, our laboratory reported more than 300 inclusion compounds formed from various organomonosulfonates and a moderately sized library of guests. Reminiscent of the aforementioned guest-free continuously layered architecture, two distinct continuously layered inclusion compound architectures were observed that differed with respect to the spacings between the organic "posts" projecting from adjacent hydrogen bonded sheets.[48] The compliance of the GS sheet permitted puckering in these compounds, such that dense packing was achieved through a kind of "shrink wrapping" around guests while retaining the essential features of the GS sheet. Moreover, the selectivity for the two layered architectures was determined by the size and shape of the guest molecules, revealing that the guests acted as templates for the assembly of the GS framework architecture capable of accommodating their inclusion. The compliance of the GS sheet was illustrated further by guanidinium monosulfonates that form cylindrical inclusion compounds in the presence of appropriately sized and shaped guests, wherein the GS sheets wrapped around guest molecules with retention of the quasi-hexagonal motif.[49] The cylinders then assembled further into hexagonal arrays through interdigitation of the organosulfonate substituents, which project form the outer surfaces of the cylinders.

The early work with GMS compounds prompted the design and synthesis of guanidinium disulfonates (GDS), wherein disulfonate pillars span opposing GS sheets anticipated predestined voids between the sheets (Fig. 4). Though these frameworks were expected to have more structural restrictions due to the disulfonate connectivity between opposing sheets, organic substituents can project from either side of the sheet, similar to their monosulfonate analogs, leading to various distinct framework architectures. This resulted in simple brick and bilayer architectures, reminiscent of the guest-free and guest containing guanidinium organomonosulfonates, as well as zig-zag brick, double brick, crisscross bilayer, and chevron brick architectures.[50,51] Like the GMS compounds, the size and shape of the guest molecules dictated the framework architecture, indicating a templating role during self-assembly wherein the arrangement of disulfonate pillars adjusted to accommodate the guest molecules.

The connectivity patterns of the GDS frameworks along the third dimension – as defined by the projections of the organic residues from either side of each sheet – differ among these architectures but the 2D quasihexagonal hydrogen-bonded motif is retained, revealing the structural versatility and robustness of this class of materials. These "architectural isomers", whether for GMS or GDS compounds, can be distinguished by projection topologies of the GS sheet wherein filled and unfilled circles represent sulfonate notes for which the R groups project "up" or "down", respectively (the undecorated nodes of this sheet correspond to guanidinium ions). In principle, an infinite number of projection topologies is possible, although random arrangements are unlikely if long-range crystallographic order is to be preserved.

Architectural isomerism and the role of guest templating can be illustrated by the guanidinium naphthalene disulfonate (G_2NDS) host with two different guest molecules, (3a*R*)-(+)-sclareolide and drospirenone,[52] which template the formation of a simple brick and zig-zag brick architectures, respectively, as a consequence of their different sizes and shapes (Fig. 5 and 6). The framework architecture simply adapts to the more demanding steric requirements of drospirenone by altering the projection topology to create the larger inclusion cavities of the zig-zag brick form.

Architectural isomerism also can be illustrated in the inverse, that is, by altering the size and character of the organodisulfonate pillars for a common guest molecule. The lasing dye coumarin 1 was included in guanidinium 4,4′-azobenzenedisulfonate (G_2ABDS), octanedisulfonate (G_2ODS) and anthracenedisulfonate (G_2ADS) hosts, which form the

 This journal is © The Royal Society of Chemistry and IChemE 2021

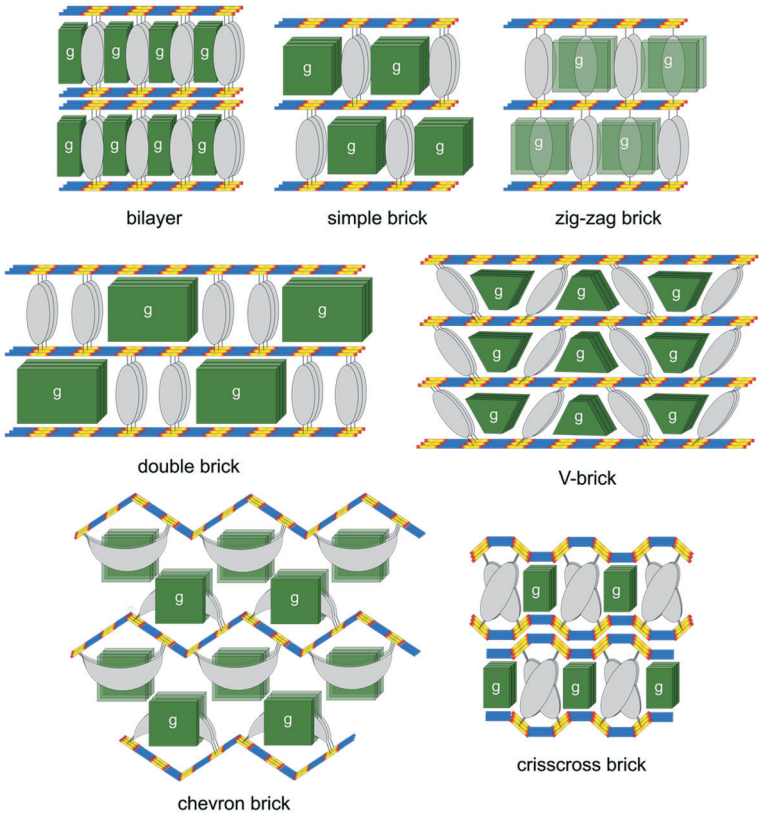


Fig. 4 Some of the GS framework inclusion compound architectures observed for guanidinium organodisulfonate frameworks.

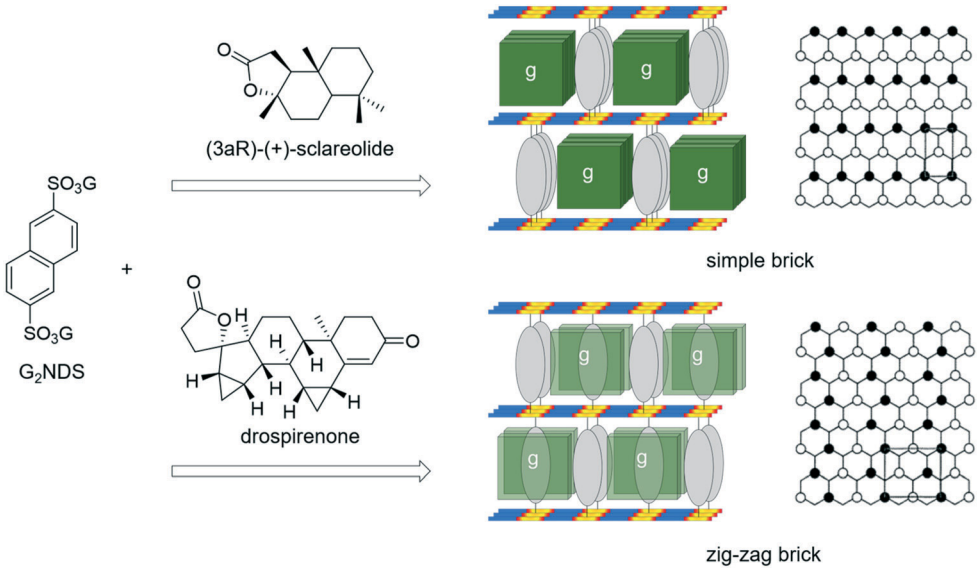


Fig. 5 The G_2NDS host forms two architectural isomers – simple brick and zig-zag brick – when templated by (3a*R*)-(+)-sclareolide and drospirenone during crystallization. The corresponding topological projection maps of the quasihexagonal hydrogen-bonded sheet for each architecture is depicted at the right, with filled and unfilled circles denoting "up" and "down" projection of organic substituents on the sulfonate nodes. Guanidinium ions reside on the undecorated nodes.

bilayer, simple brick and zig-zag brick architectures, respectively (Fig. 7).[51] The selectivity for these frameworks is a consequence of the decreasing lengths of these pillars (G_2ABDS > G_2ODS > G_2ADS), which in turn required architectures with larger cavity volumes to accommodate coumarin 1.[51]

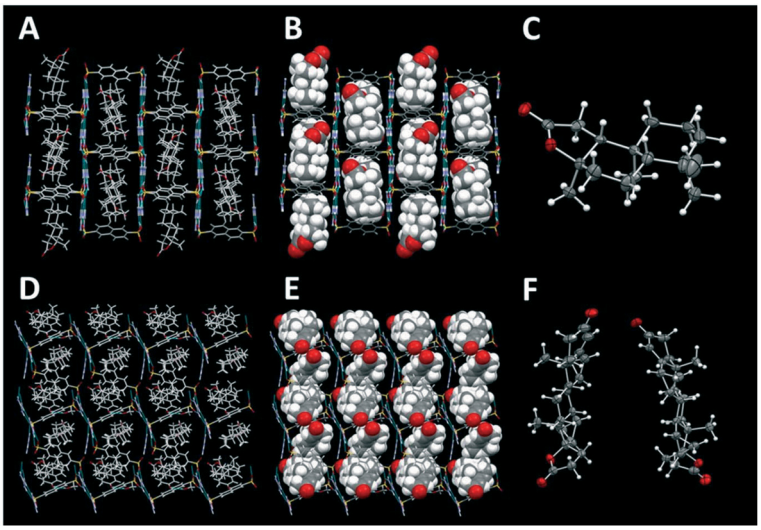


Fig. 6 Illustrative crystal structures of GS inclusion compounds. These are depicted as stick figures that illustrate the 100% occupancy (left panel), the target guest molecules as space filling (center), and ORTEP representations of the guests (right panel). (A–C) The simple brick architecture of $(G_2NDS)\supset(3aR)$-(+)-sclareolide; (D–F) the zig-zag brick architecture of $(G_2NDS)\supset(\text{drospirenone})(\text{methanol})_{0.84}(H_2O)_{0.1}$. The stick renderings in the leftmost panels reveal the 100% occupancy of target guest molecules in the inclusion cavities and the center panel illustrates the substantial contribution of the guests to the total crystal volume. Adapted with permission from Li *et al.*[52] Copyright 2019 Nature Springer.

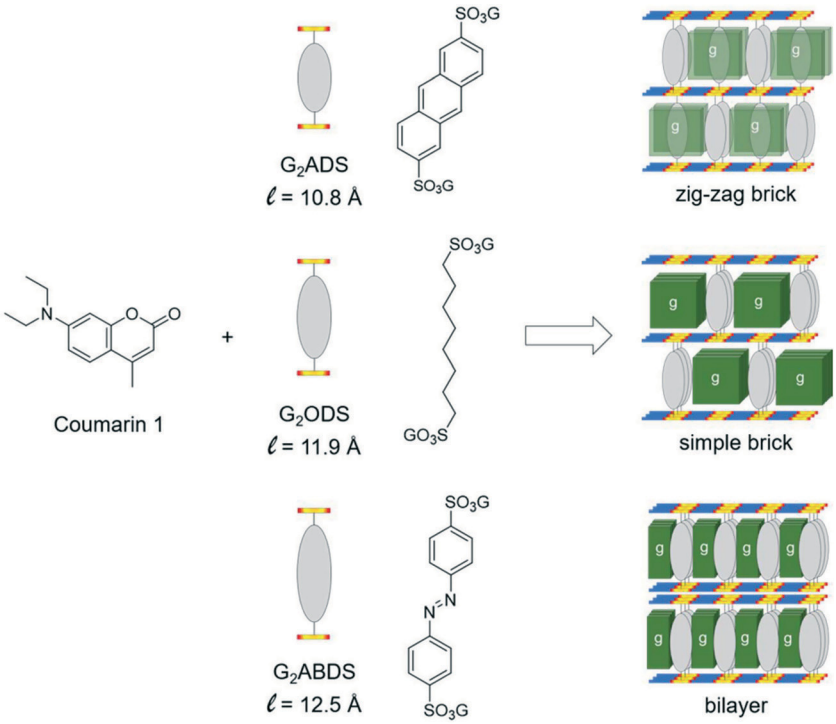


Fig. 7 Coumarin 1 templates the formation of three unique framework architectures for differently sized pillars G_2ADS (top), G_2ODS (middle), and G_2ABDS (bottom), forming inclusion compounds in zig-zag brick, simple brick and bilayer, respectively. The frameworks compensate for the decreasing pillar length by forming architectures with increasing inclusion cavity volumes.

Expanding the valency further, organotri-, tetra- and hexasulfonates afford a variety of architectures, but ones characterized by more rigidity and constraints. Trisulfonates form hexagonal cylindrical inclusion compounds, mimicking the aforementioned GMS cylindrical phases.[54,55] Organotetrasulfonates form architectures that depend on the flexibility of the linkers, with more flexible linkers affording lamellar architectures and more rigid organic linkers forming cylindrical architectures.[55] This trend is exemplified by 1,2,4,5-tetrasulfonatomethylene benzene (TSMB), 1,2,4,5-tetra(4-sulfonatophenyl)benzene (TSPB), and 1,3,6,8-tetrasulfonato-pyrene (TSP) tetrasulfonate host molecules (in

 This journal is © The Royal Society of Chemistry and IChemE 2021

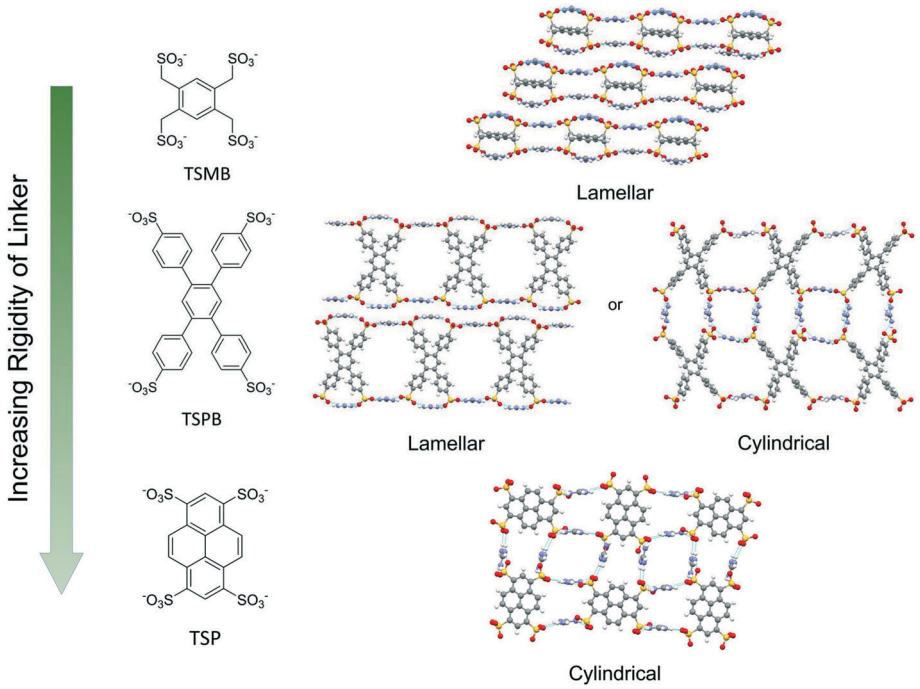


Fig. 8 Cylindrical and lamellar GS framework architectures with organotetrasulfonates depend on the rigidity of the organic linker, illustrated here for G_4TSMB, G_4TSPB and G_4TSP.

order of increasing rigidity) (Fig. 8).[55] The flexible organic arms of G_4TSMB can join opposing sheets to generate a lamellar architecture, whereas the G_4TSP framework is enforced by the rigid arrangement of sulfonate ions, affording a cylindrical architecture. G_4TSPB, which is moderately flexible, adopts both cylindrical and lamellar structures. Similar trends have been observed for organotrisulfonates with flexible arms.[55]

Design to function

Armed with a toolkit of GS hosts, GS frameworks have proven to be extraordinarily versatile with respect to functional materials. For example, GS frameworks constructed with banana-shaped pillars served as scaffolds for inclusion of polar arrays of guest molecules with molecular second-order nonlinear optical hyperpolarizabilities, leading to crystalline materials with tunable second harmonic generation.[44] Tunable optical properties through designed GS hosts and functional guests was further demonstrated by encapsulation of laser dyes with control of aggregation states[51] and sequestration of individual luminophore guests within polyhedral compartments of a crystalline zeolite-like GS framework, demonstrating a unique approach to limiting self-quenching despite the high concentrations of the luminescing guests in the framework.[53,56,57] GS hosts were demonstrated to constrain the geometry, and therefore the optical absorption, of conformationally flexible chromophores as well.[58] In a particularly interesting twist, a GS host containing calixarene "baskets" was used to trap the active form of a fruit fly pheromone from an equilibrating mixture of its monomer and trimer, suggesting a sustainable pathway to pest control for agriculture.[35] GS frameworks also have been deployed for the design of molecular magnets.[59] These examples illustrate that new materials can be realized from HBFs that provide well-controlled and reliable architectures that can guide the arrangement of guests molecules that provide a variety of properties and functions. Conversely, optoelectronic function can be introduced by redox-active donor or acceptor framework components, as suggested by the recent design of a HBF constructed with hydrogen-bonding components containing 1,4,5,8-naphthalenete-tracarboxylicdiimide, a well-known n-type semiconductor,[60] or GS frameworks constructed with azobenzenedisulfonate, a weak electron acceptor that forms charge-transfer complexes with tetrathiafulvalene guests.[61] GS frameworks also have been designed for effective separation of molecular regioisomers – that otherwise are difficult to separate – through an inclusion-crystallization strategy.[62–64]

Further elaboration of GS frameworks and their inclusion compounds beyond mono- and disulfonates to tri-, tetra- and hexasulfonates, all with predictable framework architectures, can mirror soft matter microstructures with lamellar, cylindrical and cubic architectures.[46,48–50,53,55–57] The ability to swap organosulfonate pillars with retention of the GS motif, even in frameworks built from polysulfonates, enables straightforward engineering of the volume, height, shape and chemical environment of inclusion cavities, which is key to materials design.[46] Currently, the GS library, guest-free and guest-included, comprises more than 500 crystalline compounds with various host–guest combinations,

architectures, and properties. Although the GS frameworks and other HBFs have considerable promise, their full potential, yet to be realized, would be aided by universal and reliable design principles, reinforced by emerging computational methods.

Computational design of host frameworks

As computational research has increased during the past several decades, so has its use for the acceleration of materials design and development. During the past five decades organic solid-state chemistry has produced a wide range of molecular crystals, with an equally wide range of solid-state properties, capitalizing on the versatility of organic synthesis. The prediction and control of solid-state structure in these materials, however, often has been thwarted by the delicate, noncovalent forces that govern molecular organization in the solid state, to the extent that even the slightest modification of a constituent can lead to unpredictable changes in crystal architecture. Empirical strategies have been used that rely on structure-directing interactions such as hydrogen bonds (*e.g.* HBFs) and metal coordination (*e.g.* MOFs) to override the cumulative effect of weaker and less directional crystal packing forces. In these cases, crystal structure sometimes can be anticipated from the molecular symmetry of the building blocks. Even under the best conditions, however, trial-and-error combinatorial approaches can be necessary. Computational crystal structure prediction (CSP) methods, while improving significantly, often face limitations with respect to discriminating among polymorphs that are closely spaced with respect to lattice energy as well as computational constraints that leave structures with high Z' values undetected. This can be illustrated by a discovery in our laboratory of new polymorphs of isoniazid,[65] which had previously been described as monomorphic[66] because previous experimental screens failed to find these polymorphs and CSP concluded their energies were either too high or they were not detected because the computations were limited to $Z' \leq 2$ ($Z' = 4$ for one of the polymorphs).[67]

Nonetheless, computational methods continue to improve and new approaches are emerging for structure prediction as well as solid-state properties, including applications to frameworks. For example, a *de novo* synthesis of a MOF (NU-100) was executed *in silico*, followed by its synthesis in the laboratory. The experimentally synthesized material matched the predicted structure and gas adsorption capacity predicted from computation.[68] Dispersion-corrected density functional theory (DFT) was used to characterize guest-framework interactions, gas absorption and absorption selectivity for an existing MOF framework but with a hypothetical analog equipped with aliphatic linkers and reduced pore apertures, qualities that would permit improved H_2 storage potential.[69] DFT has been used to examine the binding enthalpies of small molecules – ranging from H_2 to as large as C_3H_8 – by the isostructural M-MOF-74 series in which the metal was varied but the identity of the organic linker unchanged. This led to the prediction of selective CO_2 absorption by Cu-MOF-74 and selective removal of gas impurities and toxic gases from gas mixtures by Mn-MOF-74, which was corroborated by experimentally available binding enthlapies.[70] DFT analysis of hypothetical MOF-5 analogues with commercially available rigid and flexible organic linkers, which would alleviate the need for the synthesis of the organic ligands, suggested the possibility of "mail-order" MOFs.[71]

As the library of synthesized frameworks increases, so does the need to find a way to process this massive amount of data. One approach relies on computational high throughput screening (HTS) wherein the properties of many MOF candidates are evaluated to identify the best options for a target application, given searchable properties. Tools that have evolved under the impetus of the Materials Genome Initiative or those available in the Cambridge Structural Database have created opportunities for data mining techniques to amass libraries of known structures for HTS. This approach has been used to identify frameworks appropriate for gas separation,[72,73] gas adsorption and storage,[74–76] and synthetic parameters for formation of frameworks.[77]

An alternative to databases created for HTS relies on the generation of framework structures *de novo*. This can be exemplified by the Automated Assembly of Secondary Building Units (AASBU) method, wherein secondary building units (SBUs) are randomly distributed in a unit cell to determine viable MOF structures (Fig. 9).[78] Structures can be generated through top-down or bottom-up approaches, the former involving the mapping of individual SBUs directly onto a topology, and the latter approach connecting SBUs and then mapping them onto a topology.[79]

A new method for dividing a crystalline electron distribution into molecular fragments, based on Hirshfeld's partitioning scheme,[80] was first published by Spackman in 1997.[81] Hirshfeld surfaces were then introduced in 2005, wherein a crystal is divided into regions where the electron distribution of a sum of spherical atoms for the molecule (the promolecule) dominates the sum over the crystal (the procrystal) (Fig. 10).[82] Upon the generation of a Hirshfeld surface, a 2D fingerprint plot can be created that plots the combination of d_i (distance from a point on the surface to the nearest nucleus inside the surface) and d_e (distance from a point on the surface to the nearest nucleus outside the surface). These surfaces are unique for every crystal structure. Hirshfeld surfaces have been used recently for correlating MOF structures and surface areas (which can be important for gas adsorption), wherein the 2D fingerprints were analyzed using non-linear manifold learning methods to find correlations that may otherwise not have been evident.[83]

Although gaining more traction in recent years, principles central to machine learning (ML) were introduced hundreds of years ago. In 1763, Bayes described the probability of an event based on prior knowledge of conditions that might be

 This journal is © The Royal Society of Chemistry and IChemE 2021

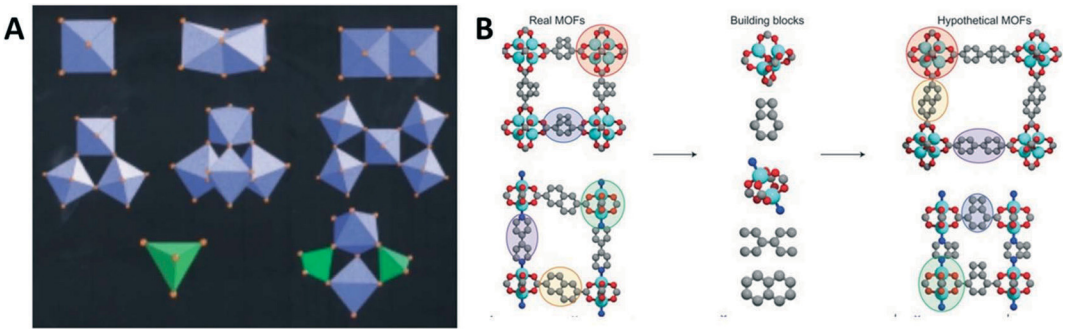


Fig. 9 (A) Examples of building units used in the AASBU method. Reprinted with permission from Mellot Draznieks *et al.*[78] Copyright 2000 John/Wiley & Sons, Inc. (B) crystal structures of existing MOFs obtained from X-ray data, divided into building blocks, and recombined to form new hypothetical MOFs. Reprinted with permission from Wilmer *et al.*[79] Copyright 2012 Springer Nature.

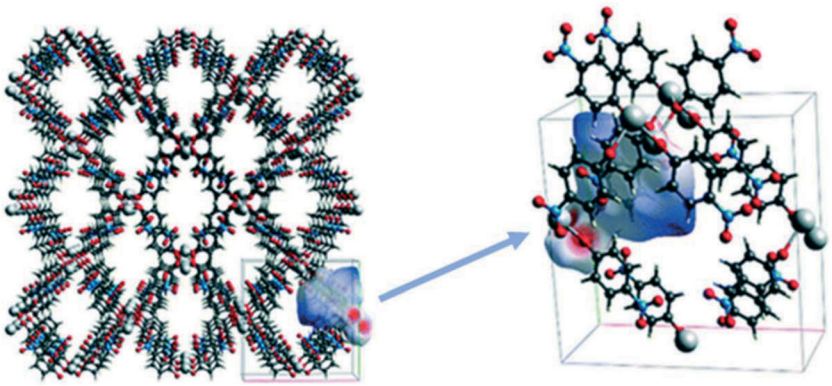


Fig. 10 A MOF structure and its Hirshfeld surface. (CSD REFCODE: FIRPOM) adapted with permission from Shen *et al.*[83] Copyright 2018 Royal Society of Chemistry.

related to the event.[84] Further pursuits of ML, however, did not commence until the mid-1900s, most notably by Markov, who described the techniques by which he analyzed a poem. These were later named Markov Chains, a stochastic model describing a sequence of events in which the probability of the subsequent event depends only on the state attained in the previous event.[85] It was not until the 1950s that the first true examples of ML were demonstrated with Turing's proposal of a 'learning machine' that could learn and become artificially intelligent,[86] the first able-to-learn neural network machine (the SNARC),[87] and a program that could play checkers.[88] In 1957 Rosenblatt invented the perceptron, a machine capable of supervised learning for binary classification.[89] Research slowed in the 1970s and again in the late 1980s due to reduced interest and funding for artificial intelligence research, causing the so-called "AI winters".[90]

As ML has gained popularity in recent years, it has gained traction in the materials world. For example, ML has been used to predict conductive MOF structures *via* transfer learning,[91] framework types of zeolite structures,[92] and MOF structures for methane adsorption.[93] ML has been employed for determining MOF properties for improved methane CO_2 and H_2 storage,[94–97] the design of MOFs with high surfaces areas based on "failed" experiments,[98] and the pursuit of a "universal" machine learning algorithm for materials screening.[99]

Prospects for computational methods and hydrogen-bonded frameworks

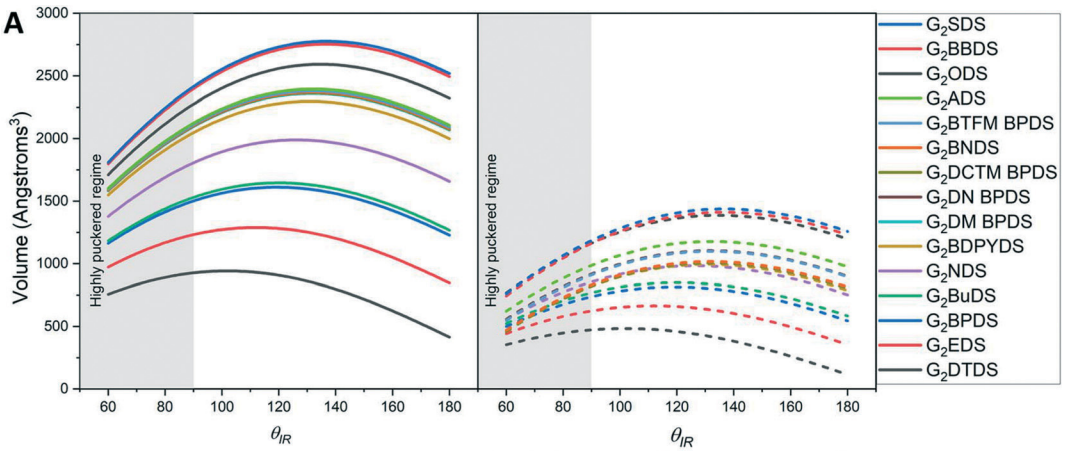
Currently there are very few examples of advanced computational methods, such as DFT or ML, for prediction of the formation or structure of HBFs, whether for inclusion compounds or for structures with persistent porosity. Energy–structure–function (ESF) maps have been constructed for building blocks based on two different organic "cores" and six different hydrogen bonded moieties, demonstrating that the number of hydrogen bonding sites and their positioning influenced the resulting ESF maps.[100] This work focused primarily on studying the hypothetical polymorphism of the frameworks, allowing the discovery of potential framework configurations and considering various pore apertures and accessible surface areas for possible framework polymorphs. DFT and Monte-Carlo methods have been used in HBFs to corroborate experimentally observed properties, such as gas adsorption.[101,102] ML has been employed for the prediction of organic cage structures with various topologies that can be construed as analogs for

cavities in solid-state frameworks, although they lack an extended network.[103]

The limited scope of computational discovery of HBFs is a clearly unmet need, which we are attempting to resolve using various approaches that will replace empirical or rudimentary intuitive ones. One of these approaches relies on a simple arithmetic model that can be employed for certain GS frameworks. This is illustrated by "master curves" for GDS frameworks with the simple brick architecture described above,[50] wherein a relatively simple geometric formula can be used to calculate the unit cell volumes (V_{cell}) for organodisulfonate pillars of different lengths (l) as a function of puckering angle θ_{IR} (Fig. 11). The master curves compare favorably with experimentally determined single crystal structures.[50] V_{guest} can be calculated readily using modeling programs, or a simple easy-to-use additive formula[104] that generates V_{guest} values matching volumes obtained by more computationally intensive modeling techniques. We have found that combining this formula with open-source cheminformatics software (rdkit) can produce accurate guest molecule volumes quickly. The volume available for guest molecules can be obtained by first accounting for a packing fraction of 0.70, which is typical of GS compounds, and then subtracting the host volume (V_{host}) from the total V_{cell}. This permits the selection of guests that can fit in the inclusion cavities of a given host or, conversely, the selection of a host that has sufficient empty volume to accommodate a given host.

Although we have found the master curve approach reliable, it cannot be applied to GS frameworks other than those based on disulfonates with simple brick architectures. This has prompted us to explore a ML approach to examining inclusion behavior. Hirshfeld surfaces, combined with ML, may offer an opportunity for rapid prediction of framework inclusion and inclusion compound structure that goes beyond simple volume-based restraints by identifying specific characteristics that govern successful host–guest inclusion compound formation. We were prompted by the Hirshfeld surfaces and their associated 2D fingerprints reported for MOFs[83] to generate the 2D fingerprints of GS frameworks and employ dimensionality reduction techniques. Notably, Hirshfeld 2D fingerprint plots of d_e *vs.* d_i for GS inclusion compounds display distinct "fingerprint plots" that can be used to code host-guest packing. For example, the Hirshfeld fingerprints in Fig. 12 reveal the differences between the G_2NDS host alone and its G_2NDS⊃(naphthalene) inclusion compound. Moreover, the fingerprint for the organic components only in the inclusion compound reveals a signature differing from that in crystalline naphthalene, even though the packings appear to be identical by visual inspection of the crystal structures.[50]

In principle, the 2D fingerprints of Hirshfeld surfaces of the GS compounds can be used to discern the inter- and intramolecular interactions governing successful inclusion compound formation. The practical use of the numerous Hirshfeld surfaces generated from the extensive library of GS compounds requires dimensionality reduction, which refers to the transformation of data from a high dimensional space into a lower dimension space, so that the lower dimensional space retains the key properties of the original data. Reducing



B

$$V_{cell}(\text{Å}^3) = a_1 b_1 c_1 = (7.5)\left(13.0\, sin\left(\frac{\theta_{IR}}{2}\right)\right)\left(13.0\, cos\left(\frac{\theta_{IR}}{2}\right) + (2\, cos(\phi)\right)$$

$$V_{guest}(\text{Å}^3) = (V_{cell} \times 0.7) - V_{host}$$

Fig. 11 Master curves (A) depicting the relationship between V_{cell} and puckering angle (θ_{IR}), as shown with the solid lines on the left. These curves are described by a simple geometric formula (B), which considers the length of the guanidinium disulfonate host where a_1 is the lattice constant along the GS ribbon direction, b_1 is orthogonal to the ribbon, c_1 is the normal to the GS sheets, ℓ is the pillar length, and φ is the tilt of the pillar (φ = 0 when the pillar is normal to the sheet). Upon multiplication by the packing fraction observed in our GS systems (0.7) and then subtraction of the host volume (where Z = 2 for the orthorhombic setting of the simple brick framework), we obtain the smaller volume available for a guest molecule, as shown by the dashed lines on the right. This remaining volume can be used to select a specific host, by knowing the volume of a target guest molecule, V_{guest}.

 This journal is © The Royal Society of Chemistry and IChemE 2021

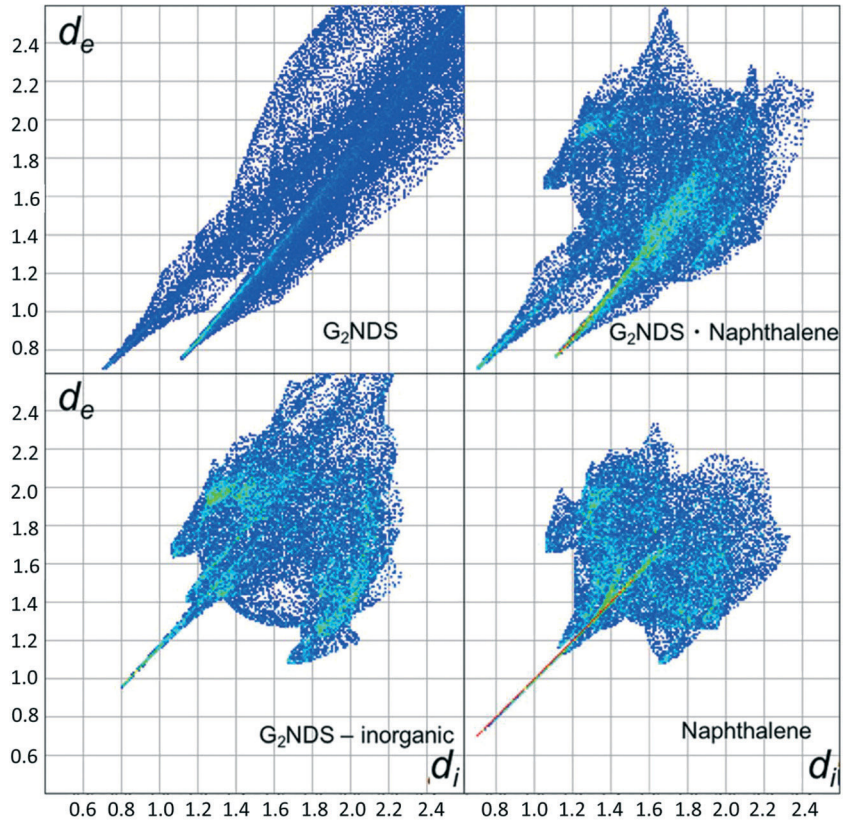


Fig. 12 Hirshfeld fingerprint plots for the naphthalene family. Upon solely visual inspection of the crystal structures, one would expect the bottom two fingerprint plots to be essentially identical, however a clear difference is observed.

data to its key features allows for better results in training ML algorithms, as well as less computational expense. The dimensionality of a dataset can be reduced by using both manifold learning, a dimensionality reduction technique used for datasets with very high dimensionality, and deep learning, a type of machine learning that uses artificial neural networks to learn. We first considered the ISOMAP algorithm, provided by scikit, which is the manifold learning method previously discussed with applications to studying MOFs.[83] The ISOMAP algorithm, first defines neighbors for each data point (corresponding to each fingerprint plot), finding interpoint distances (preserving geodesic distances), represented as matrix M, and then finding eigenvectors of M. ISOMAP was selected for its ability to locate points on a non-linear manifold, allowing for the generation of a low dimensional representation, with distance between points preserved. Autoencoders are artificial neural networks which can learn without supervision. Stacked autoencoders take an input, encode it to a smaller, more efficient internal representation of the data (an encoded representation), and then reconstruct by decoding. The smallest internal representation is information poor. Consequently, the network needs to decide which features are most characteristic, obviating the noise that does not contribute to characteristic differences. The accuracy of the autoencoder is measured by how closely the decoded data (output) matches the original (input).

Using this approach, ML has the potential to allow a user to predict suitable GS hosts for inclusion of a particular target guest molecule. Moreover, ML has considerable potential for the design of hydrogen bonded organic frameworks, in general. Combining experimentally determined qualities with computational tools can enable the rational design of HBFs with key characteristics for specific applications, accelerating discovery and development in the laboratory in a manner that obviates Edisonian approaches.

HBFs for structure determination of stubborn molecules

Molecular structure is central to chemistry, recognized since the pioneering efforts of 19th century scientists such as Jacobus Henricus van't Hoff,[105–107] who formulated the early principles of stereochemistry and the structure of the tetrahedral carbon atom, and Louis Pasteur,[108–110] who unraveled the relationship between molecular chirality, optical isomers and crystal form. It was not until 1951, however, that Johannaes Martin Bijvoet, determined the absolute configuration of a molecule, sodium rubidium tartrate, using single crystal X-ray diffraction (SCXRD).[111] Determination of molecular structure by single crystal X-ray diffraction is regarded as definitive, but it often can be frustrated by (i) the inability to grow sufficiently large single

crystals for conventional X-ray diffraction analysis, (ii) the tendency of some complex molecules to form oils or amorphous phases rather than crystals, (iii) low melting points that preclude solidification at convenient temperatures, and (iv) reactivity or decomposition under ambient conditions. While NMR spectroscopy and, more recently, micro electron diffraction (microED),[112] have been used to circumvent these obstacles posed by "stubborn" molecules, these methods are not always reliable,[113,114] and X-ray structure determination remains most definitive.[115]

One innovative strategy for determining the molecular structure of such "stubborn" target employs metal–organic frameworks (MOFs) that serve as "crystalline sponges" capable of absorbing target molecules from solution,[116,117] relying on the affinity of the pores through specific π–π, CH–π, covalent, and charge-transfer interactions that fix their positions so that their structure can be determined by X-ray diffraction.[118] This has allowed structure determination for very small amounts target compounds (≥80 ng), many with multiple stereogenic centers.[117,119–124] Crystalline sponges also have been used to characterize reaction pathways[125–128] as well as reactive molecules.[129] Alternative networks have been used to encapsulate liquid molecules through dispersive forces and hydrophilicity.[130,131] The crystalline sponge approach has been extended to crystalline sponges comprised of solely organic materials or sugars.[130,131]

While the crystalline sponge approach is undeniably innovative, its use can be limited by the need for specific intermolecular host–guest interactions or covalent fixation to minimize disorder, a requirement for "activation" of the sponge, slow absorption kinetics, an upper size limit on target molecules imposed by the size of the pore apertures, and challenges in structure determination presented by low occupancy and disordered solvent molecules. Recent developments have attempted to address some of these challenges, including optimization of guest-exchange conditions using high-throughput methods, data collection at higher diffraction angles that increase the number of reflections, and refinement protocols to mitigate the contribution from disorder of solvent molecules.[132]

Co-crystallization, which also has been used to determine the molecular structure of steroids,[133] pharmaceuticals,[134] natural products[135] and ladderanes,[136] can be particularly useful for compounds unable to crystalize on their own, but these typically require specific interactions with a co-crystal former, reflecting the absence of a more universal co-crystallizing agent or family of agents.

Our laboratory has amply demonstrated structure determination of guest molecules encapsulated within GS frameworks, but recently we extended this to more complex guests, including those with multiple stereogenic centers (Fig. 13).[52] This approach, which relies on a straightforward

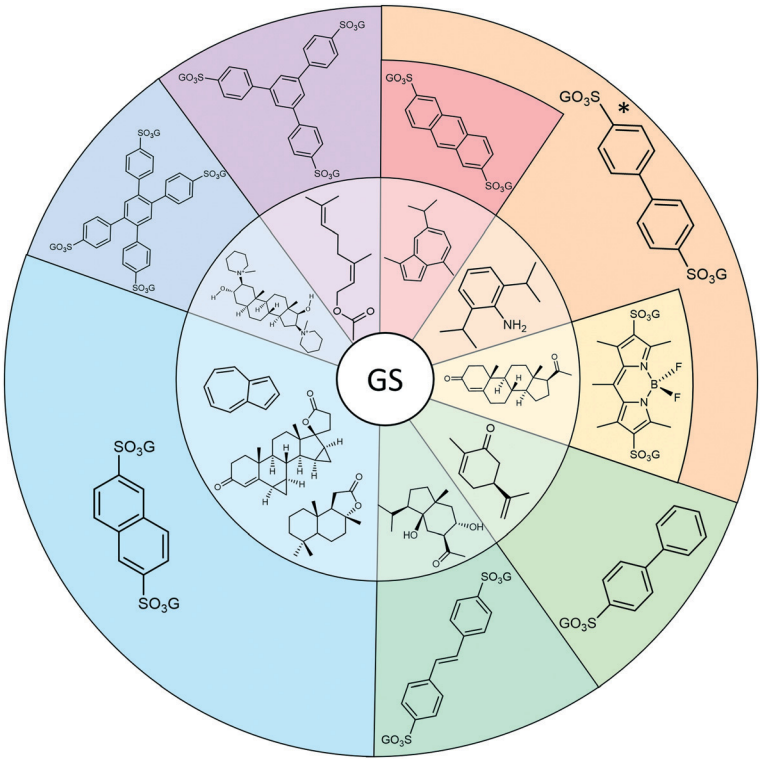


Fig. 13 Visual representation of the inclusion compounds obtained for structure determination ref. 52, guest molecules in the inner circle, with hosts used on the outer circle. All inclusion compounds allowed for definitive structure determination, and assignment of absolute configuration of stereogenic centers. *Please note that one of the hosts, biphenyl disulfonate (BPDS) formed inclusion compounds with several guests that had also formed inclusion compounds with anthracene disulfonate (ADS) and bodipy disulfonate (BDPYDS).

 This journal is © The Royal Society of Chemistry and IChemE 2021

and relatively rapid single-step crystallization of an inclusion compound upon adding a target guest of interest to a solution containing dissolved GS framework components presents many advantages: (i) GS frameworks are inherently versatile with respect to the size, shape and physicochemical character of the inclusion cavities, features that are adjustable through selection of the organosulfonate; (ii) access to an indefinite number of framework isomers that can accommodate a wide range of guests, without the need for specific interactions; (iii) the frameworks can "shrink wrap" around trapped guest molecules through puckering of the GS sheet, enabling close-packing and reduction, if not elimination, of guest disorder and solvent inclusion; (iv) GS inclusion compounds typically are stoichiometric, affording 100% occupancy of the host framework, which is tantamount to approximately 50% of the total volume, both features enabling reliable structure determination; (v) GS frameworks can include a wide range of guests, from non-polar to polar, from aliphatic to aromatic, and they tolerate many guest functional groups; (vi) diffraction data sets can be collected with a conventional diffractometer on crystals as small as 100 μm on a side, equivalent to a volume of 10^{-6} cm^3, which translates to <1 μg of target compound; and (vii) the sulfur atoms of the frameworks provide strong anomalous scattering, mitigating the challenges in determining absolute configuration of chiral centers in molecules consisting of light atoms.[137,138]

Various GS frameworks based on mono-, di-, tri- and tetrasulfonate anions were successful hosts for formation of inclusion compounds suitable for structure determination of a wide range of guest molecules. Structure determination of guaiazulene, chosen for comparison with the crystalline sponge method, was achieved using two GS frameworks, guanidinium biphenyldisulfonate (G_2BPDS) and G_2ADS.[52] The guaiazulene guests in G_2BPDS were substantially disordered, but the rigidity of the ADS pillars in G_2ADS effectively eliminated guest disorder, demonstrating that selection of organosulfonate host can lead to improved structure determination. A particularly noteworthy example of host engineering involves the structure determination of progesterone. Progesterone crystallized with G_2BPDS to produce (G_2BPDS)⊃(progesterone)(ethanol), adopting the simple brick architecture with a highly puckered GS sheet. The structure refinement of the progesterone guest was satisfactory, even though it was accompanied by an ethanol solvent molecule. Unlike solvent that persists in crystalline sponges, however, solvent included in GS compounds is typically stoichiometric and readily refined. Solvent inclusion was eliminated completely, however, by using the BDPYDS pillar, which crystallized as G_2BDPYDS⊃(progesterone) in the same framework architecture, but with improved refinement. This was attributed to the larger volume of BDPYDS (80 $Å^3$ greater than that of BPDS) displacing the volume otherwise occupied by an ethanol molecule (50 $Å^3$). Notably, absolute configuration for guest molecules containing one or more chiral centers was readily determined with high quality Flack parameters, ensuring confidence in stereochemical assingement.[52]

Interestingly, recent reports have demonstrated that the organosulfates, which can be regarded as congeners of organosulfonates, readily form crystalline salts with the guanidinium ion. Although these salts do not adopt the quasihexagonal motif of the GS compounds, they form interesting sheet- and cage-like structures, many with chiral or polar space symmetry. Moreover, the organosulfates could be prepared by conversion of an alcohol substituent of a chiral organic compound, some that were liquids that could frustrate crystallization.[139] Subsequent crystallization of the guanidinium organosulfates permitted reliable assignment of absolute configuration. Although this protocol requires a chemical step prior to crystallization, the facile formation of a salt may prove more convenient and reliable compared with the Mosher method, which involves conversion of a chiral alcohol to diastereomeric esters that permit NMR assignment of absolute configuration.

The definition of "cocrystal" can be rather contentious, and our laboratory has refrained from describing GS inclusion compounds as such. Yet a recent approach to molecular structure determination based on adamantane-based hosts, but absent hydrogen bonding, invoked "co-crystallization" to determine the structures of small molecule guests with molecular weights <222 g mol^{-1}.[140] These compounds were reminiscent of the earlier inclusion compounds reported by Ermer *et al.*[26] based on hydrogen-bonded adamantane host frameworks. Although this new family of adamantane hosts are not assembled through hydrogen bonds, they serve as an interesting contrast to the GS compounds. Many of the crystal structures were disordered, which can be attributed to the absence of strong directional interactions in the adamantane hosts. Moreover, the absence of a heavy atom like sulfur, which provides anomalous dispersion in the GS compounds, resulted in poor Flack parameters, which reduces confidence in the assignment of absolute configuration of stereogenic centers of the guest molecules.[140]

Can persistent porosity exist in HBFs?

The design of molecular-based frameworks often focuses on gas adsorption and separations, as exemplified by thousands of publications based on MOFs. In the case of HBFs, the role of guest molecules is significant, as these frameworks often collapse upon loss of guest molecules to the ambient environment.[141] Guest exchange between parent inclusion compounds and suitably sized guests in surrounding media has been observed, but this behavior usually can be attributed to loss of guest from the parent phase accompanied by collapse of the framework, followed by regeneration of the framework with a new guest through nucleation and growth in the solid state, as exemplified recently for the formation of G_2BDS inclusion compounds with small guest molecules.[142] Guest exchange also can occur

through single crystal-single crystal transformations wherein guest molecules in a parent phase are replaced with others, with retention of the framework throughout the exchange process. For example, guest exchange in guanidinium organosulfonate frameworks can be achieved through single crystal-single crystal transformations even when porosity is not sustainable in the absence of guests, or when void spaces are seemingly inaccessible. In these cases, guest exchange occurs through complex mechanisms that likely involve defects or lattice dynamics that allow for apparent single-file diffusion during exchange,[141] or the formation of lamella that shorten the diffusion pathway for guest migration, followed by a spontaneous annealing of the crystal to the new single crystalline composition.[143] It is likely that many reports of "guest exchange" occur through such mechanisms rather than ones involving a framework with persistent porosity. One motivation for designing porous HBFs is their potential for structural compliance wherein they can "expand" and "shrink" to accommodate gas adsorption, not unlike recent examples of flexible MOFs. If achievable, porous HBFs would have lower mass density than MOFs, an important consideration for gas storage materials.

Recently, a limited number of HBFs, based on aromatic building blocks equipped with hydrogen bonding groups, have been described as porous. These frameworks, have been assigned the moniker "HOF-X" (for hydrogen-bonded organic frameworks), apparently in an attempt to evoke a relationship to MOFs, where X is either an acronym denoted a molecular core or simply a number in a series (the first designation, HOF-1, was reported in 2011 (ref. 41)). As mentioned above, this nomenclature is somewhat misleading as it obscures the hundreds of HBFs known for decades. We note, however, that many of the so-called porous HOFs lack definitive single-crystal X-ray structural characterization of their porosity and the physicochemical environment of their pores, relying instead on inference from a combination of powder X-ray diffraction and gas adsorption isotherms. Moreover, in many cases these materials still contain guest molecules in their voids, usually disordered, and their role in stabilizing the porosity is either ignored or unknown.

Supramolecular organic frameworks, which have been denoted as "SOFs," and their applications for gas adsorption, have been reported. The assembly of the framework denoted SOF-1 was attributed to intermolecular N—H···$N_{pyridine}$ hydrogen bonds and π–π stacking between the arene rings in the core (Fig. 14). The porosity of SOF-1 was persistent upon removal of methanol, although dimethylformamide (DMF) from the crystallization medium was retained in the pores. The π–π interactions appeared to be important, as SOF-2, with a smaller arene in the core, collapsed to a non-porous phase upon removal of solvate molecules. Nonetheless, a selectivity for C_2H_2 over CH_4 was observed in the guest-free SOF-1a, which was attributed to the Lewis basicity of the pyridyl-decorated channels.[144]

A unique route to the design of porous HBFs relied on a search of known structures in the Cambridge Structural Database to identify compounds that form flat ordered sheets assembled by hydrogen-bonding.[145] This led to the determination that almost all 4,5-disubstituted benzimidazolones formed nearly planar ribbon-like structures through hydrogen bonding between the imidazolone units, suggesting their potential in molecular precursors to extrinsically porous compounds. This in turn led to the synthesis of a crystalline HBF, based on a

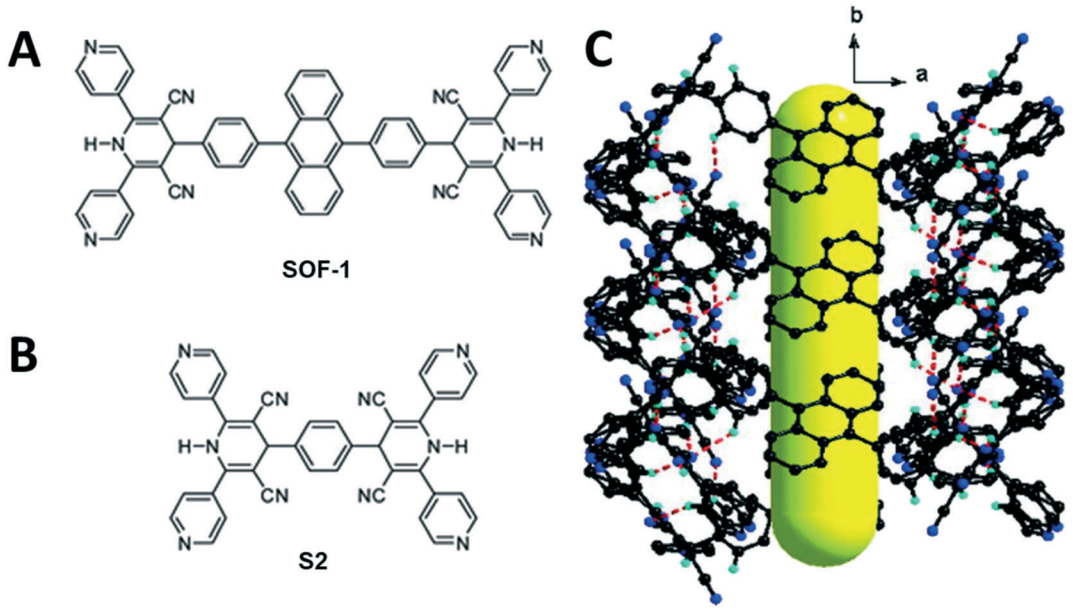


Fig. 14 (A) The building block used to assemble SOF-1. (B) The smaller building block used to assemble SOFs S2A– S2C. (C) A representation of the single crystal structure of SOF-1, with the cylindrical channels highlighted in yellow. (CSD REFCODE: APADIF) reprinted with permission from Yang *et al.*[144] Copyright 2010 American Chemical Society.

 This journal is © The Royal Society of Chemistry and IChemE 2021

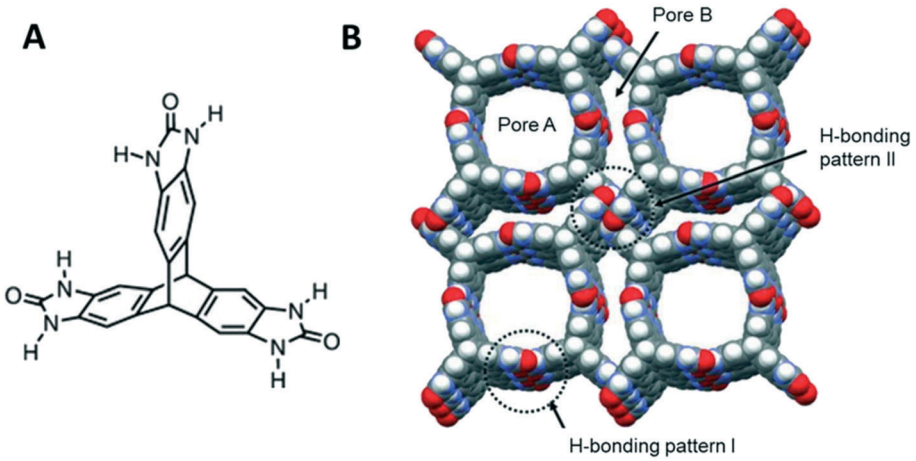


Fig. 15 Trisbenzimidazolone molecules (A) assemble into a predicted HBF architecture, as determined from a single crystal X-ray structure, with one-dimensional cylindrical channels (pore A in (B)) by a ribbon-like self-assembling of the imidazolone subunits by hydrogen bonds (H-bonding pattern I in (B)). Two of three benzimidazolone units are involved in forming the ribbon-like assembly, whereas one subunit forms another one-dimensional chainlike H-bonding structure with dihedral angles of two adjacent molecules of 90° (H-bonding pattern II in (B)). By this pattern, additional slit-like pores (pore B in (B)) are generated in between the cylindrical channels. (CSD REFCODE: DEBXIT) reprinted with permission from Mastalerz *et al.*[145] Copyright 2012 John/Wiley & Sons, Inc.

4,5-disubstituted benzimidazole with a triptycene core, which generated cylindrical pores filled with disordered solvent molecules (Fig. 15), not unlike some of the GS compounds described above. Successive exchange of the DMSO solvent molecules with more volatile acetone and *n*-pentane, followed by degassing to remove *n*-pentane (rather than more aggressive thermal treatment) afforded a compound with a high Brunauer–Emmett–Teller (BET) surface area. Although the structure of the degassed material was not reported, the gas adsorption characteristics suggested retention of porosity. The high surface area was attributed to hydrogen bond "point contacts," which minimized the molecular surface

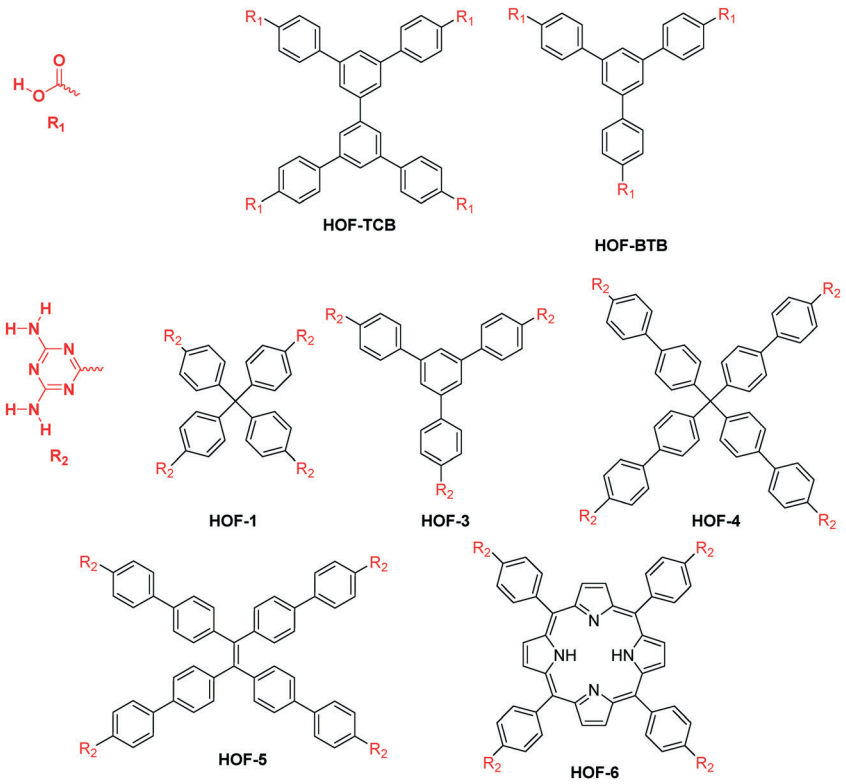


Fig. 16 Examples of molecular building blocks of so-called porous HOFs. The hydrogen-bonding groups responsible for assembling the frameworks are denoted in red.

area "sacrificed" for framework assembly, combined with a negligible loss of accessible surface area by the π–π stacking of the building blocks.[145]

HOF-TCBP was assembled by hydrogen bonding between carboxylic acid functional groups on the tetrahedral 3,3′,5,5′-Tetrakis-(4-carboxyphenyl)-1,1′-biphenyl building block. The single crystal structure of a solvent-filled framework revealed 5-fold interpenetrated structure, and powder patterns of the desolvated form were consistent with retention of the framework.[146] This framework was reported to adsorb nitrogen and separate small hydrocarbons selectively, removing C_3 and C_4 hydrocarbons from CH_4. A related framework, HOF-BTB, was reported to separate small hydrocarbons was created by using carboxylic acid based building blocks, which then stack by π–π stacking (Fig. 16).[147] The single crystal structure of the guest-filled version of HOF-BTB revealed an eightfold interpenetration of its hexagonal sheets matching a previously reported structure,[148] which created a 1D channel that produced highly accessible surface areas for gas adsorption.

Another class of HOFs has been reported recently using a combination of (amide)N–H···N(pyridine) hydrogen bonds and π–π interactions to achieve the goal of persistent porosity. HOF-1, originally synthesized by Wuest in 1997,[42] crystallized with guest molecules, which could be removed by evacuation to form HOF-1a, which adsorbed CO_2 and was selective for C_2H_2 adsorption over C_2H_4. The powder X-ray diffraction pattern of HOF-1a suggested crystallinity was retained, which was attributed to an inherent framework flexibility due to hydrogen bonding combined with stabilizing aromatic π–π interactions, as well as the existence of "molecular gates".[41] HOF-3, a guest-filled framework generated by triangular organic building blocks, was reported to form a porous HOF-3a upon evacuation with retention of crystallinity; this framework was capable of selective absorption of C_2H_2 over CO_2, both binding at gas adsorption sites in a "pocket" between two hydrogen bonding groups.[149] The PXRD data of HOF-3a framework suggested a decrease in the framework volume as compared with the parent HOF-3 with retention of hydrogen-bonding connectivity. HOF-4 was synthesized with the aim of increasing gas adsorption through the use of longer linkers to produce larger voids,[150] with gas adsorption comparable with or even superior to some MOFs and zeolites. The persistent porosity reported for HOF-5, constructed from 2,4-diaminotriazinyl (DAT) with a tetraphenylethylene (TPE) building blocks, was attributed to the participation of all amino groups of the organic components in the hydrogen bonding that supported the framework channels (HOF-5a, CSD REFCODE: BUPHUR).[151] These channels, which appear to be filled with disordered guest molecules, preferentially adsorbed CO_2 over C_2H_2, CH_4 and N_2. Notably, neutron diffraction revealed that 1.44 CO_2 molecules were adsorbed per organic linker, consistent with experimental gas adsorption values of 1.5. Neutron diffraction revealed the formation of "pseudo-one-dimensional arrays" of CO_2 molecules at two binding sites through dispersive interactions, which was corroborated by dispersion-corrected DFT calculations.

HOF-7 was constructed from a zinc porphyrin-based building block, motivated by the use of the porphyrin moiety in materials for catalysis and electronic applications. This framework consisted of two-dimensional layers of these building blocks assembled by intermolecular hydrogen-bonding and π–π interactions. CO_2 was adsorbed selectively from a mixture with N_2, confirming calculations performed by ideal adsorbed solution theory (IAST).[152] HOF-6 exhibited a similar structure motif but with a metal-free porphyrin. IAST calculations suggested selectivity for CO_2 over CH_4 and N_2 as well as C_2H_2 over CH_4.[153] The metal-free porphyrin framework was found to have similar BET surface area to the Zn containing structure, and had similar CO_2 adsorption, as well as IAST modeled CO_2/N_2 separation, suggesting that the framework can be just as effective even upon removal of the metal component.

The HOF series relies on the combination of hydrogen bonding and π–π interactions for the persistence of porosity. HOF-8, built from N_{amide}—H···$N_{pyridine}$ hydrogen bonds, is stable even upon heating to over 350 °C, surpassing that of MOF materials.[154] This exceptional thermal stability was attributed to three pairs of highly symmetric hydrogen bonds. Comparison of thermal gravimetric analysis of HOF-8, with amide hydrogen atoms replaced by deuterium, was consistent with significant contributions of the hydrogen bonds to the stability of the frameworks. Gas adsorption studies revealed that the framework was highly selective for CO_2 over N_2 or O_2, which was attributed to the quadrupole moment of CO_2 and to pyridyl amide nitrogen atoms decorating the interior walls of the framework.[155] HOF-8 also demonstrated selectivity for benzene adsorption over various other hydrocarbons, including *n*-hexane, cyclohexane, toluene, *p*-xylene.[154]

In 2016 guanidinium arenesulfonate frameworks were reported as a "new class of porous crystalline proton-conducting materials", describing "two porous two-dimensional (2D) HOFs based on arene sulfonic acid, that is, 4,4'-biphenyldisulfonic acid and 1,5-Napthalenedisulfonic acid, non-covalently bonded to guanidinium ions to form infinite pillar-brick type arrangement". The authors noted a "structural dynamism" associated with collapse of the (guanidinum)(4,4′-biphenyldisulfonate) framework to a dense structure upon loss of *p*-xylene guests and regeneration to the original form upon immersion in *p*-xylene, a well-known phenomenon for GS inclusion compounds. This most likely occurs through a straightforward nucleation and growth mechanism, like the aforementioned behavior observed for G_2BDS.[142] Although no direct structural evidence for persistent porosity in these frameworks was reported, it was inferred for one of the frameworks based on the observation of CO_2 adsorption. The two compounds above were denoted as "HOF-GS-11" and "HOF-GS-10", respectively, assigning numbers for two frameworks among many in the GS series, both of which were reported by our laboratory roughly twenty

 This journal is © The Royal Society of Chemistry and IChemE 2021

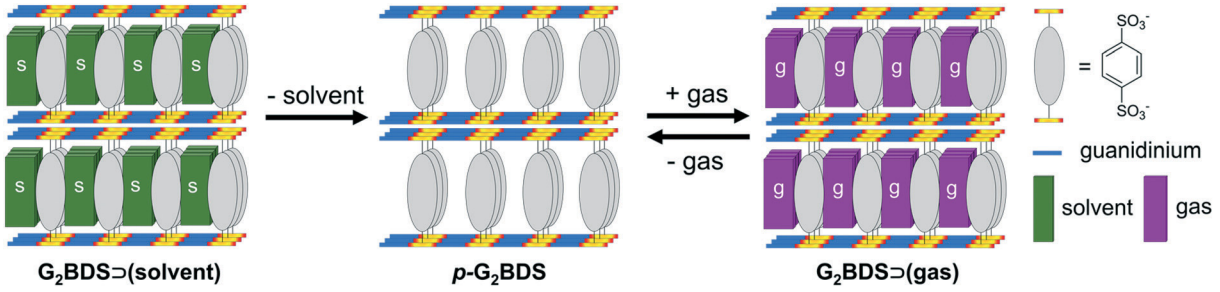


Fig. 17 Transformation of $G_2BDS\supset$(solvent) (left) upon removal of the solvent, which retains its crystallinity and becomes the microporous structure, *p*-G_2BDS (middle), which is capable of reversible adsorption of gases to become $G_2BDS\supset$(gas) (right).

years earlier.[156,157] At least two review articles have perpetuated this misrepresentation, illustrating how "branding" can lead to a propagation of misinformation.[158,159] Ironically, these review articles, which use the recent "HOF-X" nomenclature, acknowledge the existence of HOFs decades ago, one citing the earliest "HOF" dating back to 1969, while erroneously stating that "research progress on HOFs was stagnant during the following decades" despite hundreds of hydrogen-bonded frameworks being reported during that time.

Recently, however, gas adsorption by a truly porous guanidinium 1,4-benzenedisulfonate, G_2BDS, formed by a single crystal-to-single crystal transformation upon removing acetone solvent guests from the framework, was reported (Fig. 17).[40] In an attempt to distinguish the hundreds (if not thousands) of HBFs reported over decades – most without persistent porosity – from frameworks with persistent porosity, a different nomenclature was suggested wherein G_2BDS framework was denoted as a "*p*-HOF" (*p* = porous). The porous phase *p*-G_2BDS was metastable relative to a "collapsed" nonporous polymorph denoted *np*-G_2BDS, but a two-decades-old sample was found intact as the open, porous form, indicative of robust kinetic stability against transformation to the nonporous form. Notably, *p*-G_2BDS was capable of adsorbing N_2, CO_2, Xe and acetone, at 30 bar for CO_2 and 21.5 bar of Xe. This is a rare example of a definitive structural determination of porosity in a hydrogen-bonded framework.

Catalysis

Frameworks, whether MOFs or HOFs, represent a promising alternative to traditional catalysts due to high surface areas, which allow for high catalytic reaction rates, and customizable pores that can allow for shape-selective catalysis, chemical selectivity, rapid transport and confinement effects.[160] The study of hydrogen bonded organic frameworks for catalysis has been limited, largely attributed to the difficulties in establishing persistent porosity.

Catalytic behavior in organic inclusion compounds was observed as early as 1987, when a 1,1,6,6-tetraphenylhexa-2,4-diyne-1,6-diol converted an olefinic chalcone to its dimer, a cyclobutane derivative, when irradiated with white light. The photochemical dimerization of the chalcone in the complex was significantly accelerated compared with the chalcone alone, and the host could be recycled, effectively meeting key criteria for a catalyst.[161] Roughly ten years later a hydrogen bonded network host material was reported to catalyze a Diels–Alder reaction, with stereoselectivity, catalytically for an aldehyde dienophile and stoichiometrically for ester dienophiles.[162] Mechanistic studies demonstrated that the catalytic activity was associated with the confinement of the reaction within the internal cavities of the host.

Inspired by the permanently porous porphyrin framework mentioned above,[152] a framework constructed from a Co(II) 5,10,15,20-tetra(4-(4-acetateethyl)phenoxy)phenylporphyrin (CoTCPp) core equipped with hydrogen-bonding groups was evaluated for catalytic activity in alkylbenzene oxidation (Fig. 18).[163] This building block was selected because of its potential for modifying catalytic activity through introduction of different substituents on the porphyrin ring as well as different metal ions. Catalysis was achieved with only slight

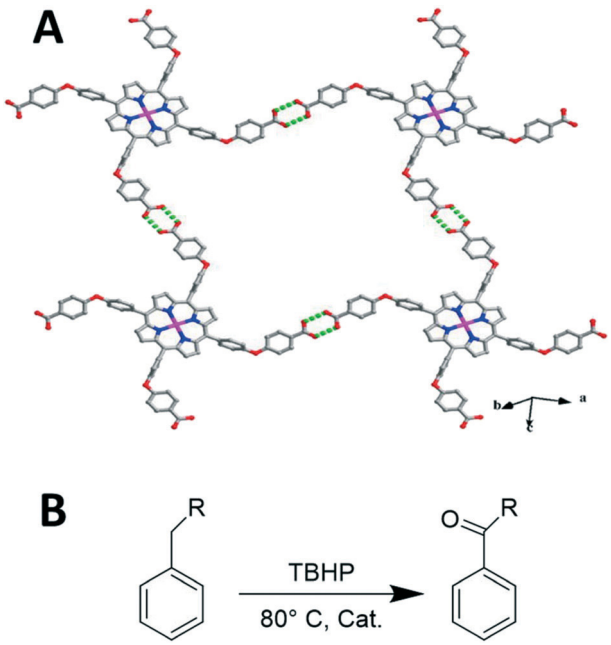


Fig. 18 (A) Crystal structure of the hydrogen bonded framework (hydrogen bonds are depicted in green) studied for the catalysis of an oxidation reaction with *tert*-butylhydroperoxide (TBHP) (B). (CSD REFCODE: NURCEK) reprinted with permission from Zhang *et al.*[163] Copyright 2015 American Chemical Society.

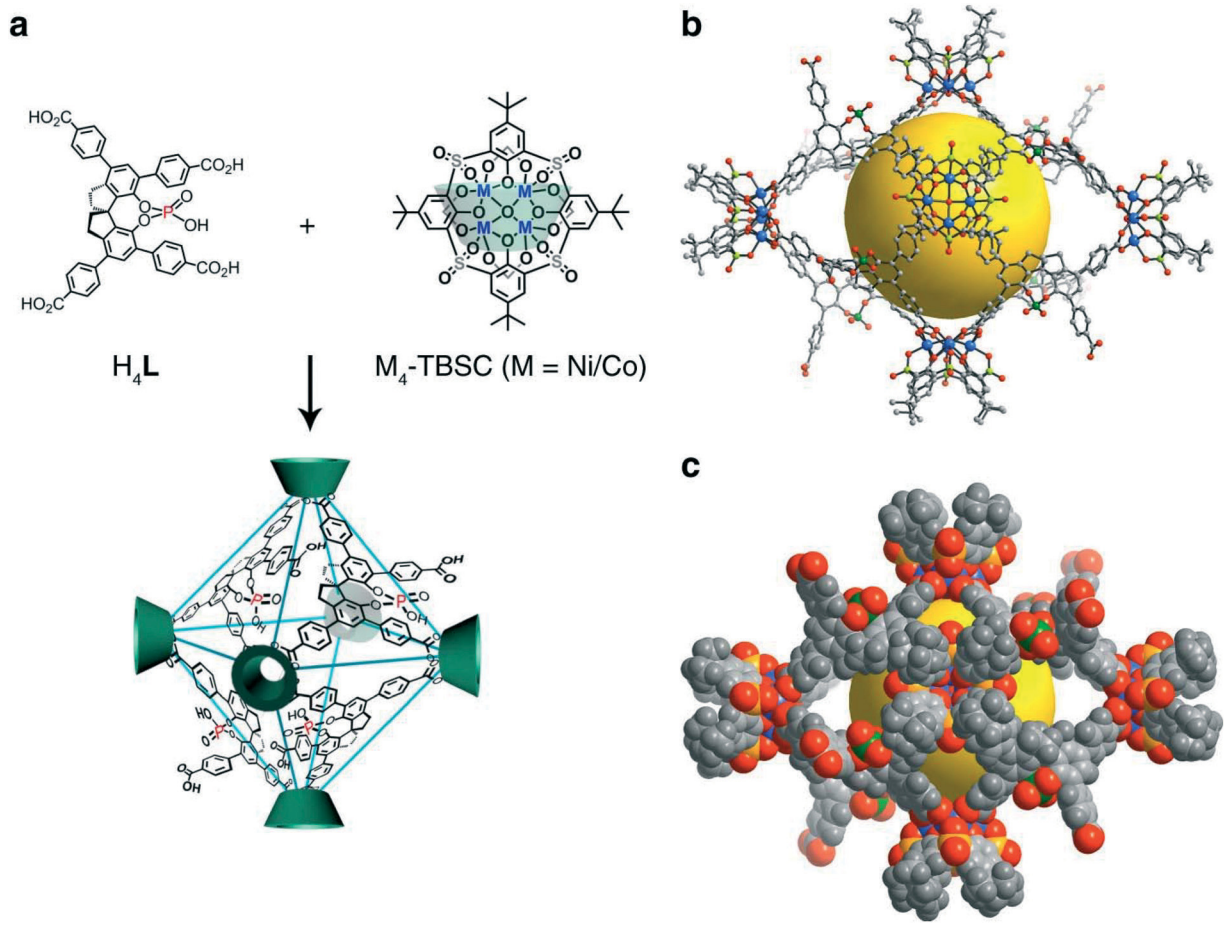


Fig. 19 Assembly procedures. (a) Self-assembly of cages **1**-Ni and **1**-Co (only half of the HL^{3-} ligand are shown in the cage for clarity). (b) The single-crystal structure of the octahedral cage in **1**-Ni and (c) the space-filling model with an elliptical shape viewed along the short axis (sky-blue, Ni; green, P; yellow, S; gray, C; red, O). The cavities are highlighted by colored spheres. (CSD REFCODES: HIZQIT, HIZVOE, HIZVUK) reprinted with permission from Gong *et al.*[165] Copyright 2019 Springer Nature.

decrease in activity after three turnovers, and the framework was easily recoverable by centrifugation and filtration so that it could be recycled.

A cage-based hydrogen bonded framework, HOF-19, was constructed from amino-substituted bis(tetraoxacalix[2]arene[2]triazine) cagelike building blocks that were thought to coerce assembly through π–π interactions and multiple hydrogen bonds between the building blocks.[164] Post-synthetic modification of the framework with palladium(II) generated HOF-19⊃Pd(II) with retention of the framework structure, as revealed by PXRD. The HOF-19⊃Pd(II) catalyst demonstrated superior performance for the catalysis of a Suzuki–Miyaura coupling reaction when compared with palladium acetate, a mixture of HOF-19 and palladium acetate, and a Pd/C catalyst, demonstrating the vital importance of the palladium(II) ion along the framework channels. HOF-19⊃Pd(II) remained crystalline for four turnovers, after which the crystallinity and activity began to decrease. The catalytic activity, however, was recovered after recrystallization of the deactivated catalyst.

Organic cages have been used as building blocks to generate frameworks for catalysis, as demonstrated by octahedral coordination cages constructed from using the enantiopure 4,4′,6,6′-tetra(benzoate) ligand of 1,1′-spirobiindane-7,7′-phosphoric acid and Ni_4/Co_4-*p-tert*-butylsulfonylcalix[4]arene clusters, which single crystal X-ray diffraction revealed was assembled through strong hydrogen bonds (Fig. 19).[165] The host was evaluated as a heterogeneous Brønsted acid catalyst for asymmetric [3 + 2] coupling of indoles with quinone monoamine, as well as Friedel–Crafts alkylation of indole with aryl aldimines. High enantioselectivities, with *ee* values up to 99.9%, combined with high catalytic activity, were attributed to the chiral environment of the framework pores, as validated by lower enantioselectivities observed for analogous homogenous catalysts. The catalyst was active for at least ten turnovers with minimal loss of activity and enantioselectivity, and with retention of porosity, crystallinity, and structure.

Summary and outlook

The role of hydrogen bonds in guiding the organization of molecular constituents of organic crystalline solids, including those with well-defined frameworks, has been recognized for

 This journal is © The Royal Society of Chemistry and IChemE 2021

decades. It may surprise some to learn that hydrogen-bonded frameworks preceded the first reports of metal–organic frameworks, which can be viewed as the predecessors of the more contemporary covalent organic frameworks. These different classes of materials share many common features, including the use of design principles that rely on the versatility of organic synthesis to construct building blocks with molecular symmetries and functional groups that guide the assembly of the frameworks with rather predictable topology and architecture in the solid state. They differ, however, with respect to several key attributes. Unlike MOFs and COFs, materials based on HBFs typically form under near-equilibrium conditions, favoring a higher degree of crystallinity that often enables more reliable determination of solid-state structure and its correspondence with materials properties. As illustrated by the examples above, HBFs differ with respect to their stability as truly porous materials, more often than not relying on the space-filling of guest molecules for framework support, a so-called virtual porosity. Recent reports suggest that persistent porosity in HBFs, and ensuing applications such as gas adsorption, gas separations, and chiral separations, are within reach. A question remains, however, as to whether truly porous HBFs will compete with the more robust MOFs and COFs, or their inorganic zeolite counterparts. Instead, the community may want to consider expanding activity in the design of well-defined inclusion compounds that have utility stemming from the solid-state architecture imposed by a framework, wherein properties and function can be tuned through judicious choice of guest molecules – smart materials used smartly. HBF inclusion compounds are almost always stoichiometric, with well-defined ordered arrangements of guest molecules that are essential to many solid-state properties. This has been amply demonstrated for the guanidinium organosulfonate frameworks, which can accommodate a range of guest molecules that enable tunable molecular magnetism, second harmonic generation and light emission. Guanidinum organosulfonate frameworks also have been used for structure determination of guest molecules, chemical separations, including enantioselective separations through the use of chiral framework components, not unlike that reported for a homochiral HBF based on twisted bis(napthyl) building blocks.[166] This range of properties suggests that HBFs, in general, are sound platforms for the design and synthesis of useful materials with many more frameworks waiting to be discovered. Indeed, guanidinum organosulfonate frameworks have been emulated in reports of analogous pillared networks constructed from organosulfonates and transition metals equipped with amine ligands,[167–170] and hydrogen bonded guanidinium borate networks in which the guanidinium ion acts as a 3-fold connecting node to generate a cubic boracite network.[171] We anticipate other classes of HBFs will illuminate further the underlying principles governing their structure and can deliver new functional materials. The ability to control framework topology and alignment of guest molecules with uniformity and precision suggests opportunities for use in molecular electronics, such as field effect transistors, wherein transport properties can be manipulated by the selection of hosts as well as guests. The use of computations, from density functional theory to machine learning, for the design of functional MOFs has exploded in the past few years. We anticipate that computer-aided design of HBFs promises to accelerate their discovery and development as well.

Appendix. Explanation of acronyms

Abbreviation

G_2SDS	G_2 4,4′-stilbenedisulfonate
G_2BBDS	G_2 4,4′-bibenzyldisulfonate
G_2ODS	G_2 octanedisulfonate
G_2ADS	G_2 anthracenedisulfonate
G_2BTFM BPDS	G_2 2,2′-bis(trifluoromethyl)-4,4′-biphenyldisulfonate
G_2BNDS	G_2 4,4′-binaphthyldisulfonate
G_2DCTM BPDS	G_2 3,3′-dichloro-2,2′,6,6′-tetramethyl-4,4′-biphenyldisulfonate
G_2DN BPDS	G_2 2,2′-dinitro-4,4′-biphenyldisulfonate
G_2DM BPDS	G_2 2,2′-dimethyl-4,4′-biphenyldisulfonate
G_2BDPYDS	G_2 bodipydisulfonate
G_2NDS	G_2 2,6-napthalenedisulfonate
G_2BuDS	G_2 butanedisulfonate
G_2BPDS	G_2 biphenyldisulfonate
G_2EDS	G_2 ethanedisulfonate
G_2DTDS	G_2 dithionatedisulfonate

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

This work was supported by the National Science Foundation through award number DMR-2002964.

References

1 T. S. Moore and T. F. Winmill, *J. Chem. Soc., Trans.*, 1912, **101**, 1635–1676.
2 W. M. Latimer and W. H. Rodebush, *J. Am. Chem. Soc.*, 1920, **42**, 1419–1433.
3 L. Pauling, *Proc. Natl. Acad. Sci. U. S. A.*, 1928, **14**, 359–362.
4 J. West, *Ferroelectrics*, 1930, **71**, 1–9.
5 W. H. Zachariasen, *J. Chem. Phys.*, 1933, **1**, 634–639.
6 W. H. Zachariasen, *J. Chem. Phys.*, 1935, **3**, 158–161.
7 J. Donohue, *J. Phys. Chem.*, 1952, **56**, 502–510.
8 D. Elson and E. Chargaff, *Experientia*, 1952, **8**, 143–145.
9 E. Chargaff, R. Lipshitz and C. Green, *J. Biol. Chem.*, 1952, **195**, 155–160.
10 R. E. Franklin and R. G. Gosling, *Nature*, 1953, **171**, 740–741.
11 M. H. F. Wilkins, W. E. Seeds, A. R. Stokes and H. R. Wilson, *Nature*, 1953, **172**, 759–762.

12 J. D. Watson and F. H. C. Crick, *Nature*, 1953, **171**, 737–738.
13 G. C. Pimentel and A. L. McClellan, *The Hydrogen Bond*, San Francisco, 1960.
14 D. J. Sutor, *J. Chem. Soc.*, 1963, 1105–1110.
15 R. Taylor and O. Kennard, *J. Am. Chem. Soc.*, 1982, **104**, 5063–5070.
16 R. Taylor and O. Kennard, *Acc. Chem. Res.*, 1984, **17**, 320–326.
17 J. Clare Speakman, *Acid Salts of Carboxylic Acids, Crystals with some "Very Short" Hydrogen Bonds*, Springer Berlin Heidelberg, Berlin, Heidelberg, 1972, pp. 141–199.
18 M. D. Ward, *Chem. Commun.*, 2005, 5838–5842.
19 P. Dauber and A. T. Hagler, *Acc. Chem. Res.*, 1980, **13**, 105–112.
20 M. C. Etter, *Acc. Chem. Res.*, 1990, **23**, 120–126.
21 J. A. Zerkowski, C. T. Seto, D. A. Wierda and G. M. Whitesides, *J. Am. Chem. Soc.*, 1990, **112**, 9025–9026.
22 J.-M. Lehn, M. Mascal, A. Decian and J. Fischer, *J. Chem. Soc., Chem. Commun.*, 1990, 479.
23 J.-M. Lehn, M. Mascal, A. DeCian and J. Fischer, *J. Chem. Soc., Perkin Trans. 2*, 1992, 461.
24 O. Ermer, *J. Am. Chem. Soc.*, 1988, **110**, 3747–3754.
25 O. Ermer and A. Eling, *Angew. Chem., Int. Ed. Engl.*, 1988, **27**, 829–833.
26 O. Ermer and L. Lindenberg, *Helv. Chim. Acta*, 1991, **74**, 825–877.
27 B. F. Hoskins and R. Robson, *J. Am. Chem. Soc.*, 1989, **111**, 5962–5964.
28 O. M. Yaghi, G. Li and H. Li, *Nature*, 1995, **378**, 703–706.
29 M. Simard, D. Su and J. D. Wuest, *J. Am. Chem. Soc.*, 1991, **113**, 4696–4698.
30 X. Wang, M. Simard and J. D. Wuest, *J. Am. Chem. Soc.*, 1994, **116**, 12119–12120.
31 C. T. Seto and G. M. Whitesides, *J. Am. Chem. Soc.*, 1990, **112**, 6409–6411.
32 *IUPAC Compendium of Chemical Terminology*, ed. M. Nič, J. Jirát, B. Košata, A. Jenkins and A. McNaught, IUPAC, Research Triagle Park, NC, 2nd edn, 2009.
33 J. D. Evans, C. J. Sumby and C. J. Doonan, *Chem. Lett.*, 2015, **44**, 582–588.
34 R. Gramage-Doria, D. Armspach and D. Matt, *Coord. Chem. Rev.*, 2013, **257**, 776–816.
35 W. Xiao, C. Hu and M. D. Ward, *Cryst. Growth Des.*, 2013, **13**, 3197–3200.
36 Y. Liu and M. D. Ward, *Cryst. Growth Des.*, 2009, **9**, 3859–3861.
37 L. R. MacGillivray and J. L. Atwood, *Nature*, 1997, **389**, 469–472.
38 S. J. Dalgarno, P. K. Thallapally, L. J. Barbour and J. L. Atwood, *Chem. Soc. Rev.*, 2007, **36**, 236–245.
39 L. J. Barbour, *Chem. Commun.*, 2006, 1163.
40 I. Brekalo, D. E. Deliz, L. J. Barbour, M. D. Ward, T. Friščić and K. T. Holman, *Angew. Chem., Int. Ed.*, 2020, **59**, 1997–2002.
41 Y. He, S. Xiang and B. Chen, *J. Am. Chem. Soc.*, 2011, **133**, 14570–14573.
42 P. Brunet, M. Simard and J. D. Wuest, *J. Am. Chem. Soc.*, 1997, **119**, 2737–2738.
43 V. A. Russell, M. C. Etter and M. D. Ward, *J. Am. Chem. Soc.*, 1994, **116**, 1941–1952.
44 V. A. Russell, M. C. Etter and M. D. Ward, *Chem. Mater.*, 1994, **6**, 1206–1217.
45 A. Katrusiak and M. Szafrański, *Acta Crystallogr., Sect. C: Cryst. Struct. Commun.*, 1994, **50**, 1161–1163.
46 K. T. Holman, S. M. Martin, D. P. Parker and M. D. Ward, *J. Am. Chem. Soc.*, 2001, **123**, 4421–4431.
47 M. C. Etter, Z. Urbanczyk-Lipkowska and D. A. Jahn, *J. Am. Chem. Soc.*, 1986, **108**, 5871–5876.
48 M. J. Horner, K. T. Holman and M. D. Ward, *J. Am. Chem. Soc.*, 2007, **129**, 14640–14660.
49 M. J. Horner, K. T. Holman and M. D. Ward, *Angew. Chem., Int. Ed.*, 2001, **40**, 4045.
50 K. T. Holman, A. M. Pivovar, J. A. Swift and M. D. Ward, *Acc. Chem. Res.*, 2001, **34**, 107–118.
51 A. C. Soegiarto and M. D. Ward, *Cryst. Growth Des.*, 2009, **9**, 3803–3815.
52 Y. Li, S. Tang, A. Yusov, J. Rose, A. N. Borrfors, C. T. Hu and M. D. Ward, *Nat. Commun.*, 2019, **10**, 4477.
53 Y. Liu, C. Hu, A. Comotti and M. D. Ward, *Science*, 2011, **333**, 436–440.
54 A. Comotti, S. Bracco, P. Sozzani, S. M. Hawxwell, C. Hu and M. D. Ward, *Cryst. Growth Des.*, 2009, **9**, 2999–3002.
55 Y. Liu, W. Xiao, J. J. Yi, C. Hu, S.-J. Park and M. D. Ward, *J. Am. Chem. Soc.*, 2015, **137**, 3386–3392.
56 M. Handke, Y. Wu, Y. Li, C. T. Hu and M. D. Ward, *CrystEngComm*, 2020, **22**, 3749–3752.
57 M. Handke, T. Adachi, C. Hu and M. D. Ward, *Angew. Chem., Int. Ed.*, 2017, **56**, 14003–14006.
58 S. Tang, A. Yusov, Y. Li, M. Tan, Y. Hao, Z. Li, Y.-S. Chen, C. T. Hu, B. Kahr and M. D. Ward, *Mater. Chem. Front.*, 2020, **4**, 2378–2383.
59 A. C. Soegiarto, W. Yan, A. D. Kent and M. D. Ward, *J. Mater. Chem.*, 2011, **21**, 2204–2219.
60 L. Wang, L. Yang, L. Gong, R. Krishna, Z. Gao, Y. Tao, W. Yin, Z. Xu and F. Luo, *Chem. Eng. J.*, 2020, **383**, 123117.
61 A. C. Soegiarto, A. Comotti and M. D. Ward, *J. Am. Chem. Soc.*, 2010, **132**, 14603–14616.
62 J. Kim, J. Yi, M. D. Ward and W.-S. Kim, *Sep. Purif. Technol.*, 2009, **66**, 57–64.
63 J. Kim, S.-O. Lee, J. Yi, W.-S. Kim and M. D. Ward, *Sep. Purif. Technol.*, 2008, **62**, 517–522.
64 A. M. Pivovar, K. T. Holman and M. D. Ward, *Chem. Mater.*, 2001, **13**, 3018–3031.
65 K. Zhang, N. Fellah, A. G. Shtukenberg, X. Fu, C. Hu and M. D. Ward, *CrystEngComm*, 2020, **22**, 2705–2708.
66 S. R. Kennedy, C. D. Jones, D. S. Yufit, C. E. Nicholson, S. J. Cooper and J. W. Steed, *CrystEngComm*, 2018, **20**, 1390–1398.
67 C. R. Taylor, M. T. Mulvee, D. S. Perenyi, M. R. Probert, G. M. Day and J. W. Steed, *J. Am. Chem. Soc.*, 2020, **142**, 16668–16680.
68 O. K. Farha, A. Özgür Yazaydın, I. Eryazici, C. D. Malliakas, B. G. Hauser, M. G. Kanatzidis, S. T. Nguyen, R. Q. Snurr and J. T. Hupp, *Nat. Chem.*, 2010, **2**, 944–948.

 This journal is © The Royal Society of Chemistry and IChemE 2021

69 M. Witman, S. Ling, A. Gladysiak, K. C. Stylianou, B. Smit, B. Slater and M. Haranczyk, *J. Phys. Chem. C*, 2017, **121**, 1171–1181.
70 K. Lee, J. D. Howe, L.-C. Lin, B. Smit and J. B. Neaton, *Chem. Mater.*, 2015, **27**, 668–678.
71 R. L. Martin, L.-C. Lin, K. Jariwala, B. Smit and M. Haranczyk, *J. Phys. Chem. C*, 2013, **117**, 12159–12167.
72 S. Han, Y. Huang, T. Watanabe, Y. Dai, K. S. Walton, S. Nair, D. S. Sholl and J. C. Meredith, *ACS Comb. Sci.*, 2012, **14**, 263–267.
73 S. Li, Y. G. Chung and R. Q. Snurr, *Langmuir*, 2016, **32**, 10368–10376.
74 Y. G. Chung, J. Camp, M. Haranczyk, B. J. Sikora, W. Bury, V. Krungleviciute, T. Yildirim, O. K. Farha, D. S. Sholl and R. Q. Snurr, *Chem. Mater.*, 2014, **26**, 6185–6192.
75 N. S. Bobbitt, J. Chen and R. Q. Snurr, *J. Phys. Chem. C*, 2016, **120**, 27328–27341.
76 Y. J. Colón, D. Fairen-Jimenez, C. E. Wilmer and R. Q. Snurr, *J. Phys. Chem. C*, 2014, **118**, 5383–5389.
77 E. Biemmi, S. Christian, N. Stock and T. Bein, *Microporous Mesoporous Mater.*, 2009, **117**, 111–117.
78 C. Mellot Draznieks, J. M. Newsam, A. M. Gorman, C. M. Freeman and G. Férey, *Angew. Chem., Int. Ed.*, 2000, **39**, 2270–2275.
79 C. E. Wilmer, M. Leaf, C. Y. Lee, O. K. Farha, B. G. Hauser, J. T. Hupp and R. Q. Snurr, *Nat. Chem.*, 2012, **4**, 83–89.
80 F. L. Hirshfeld, *Theor. Chim. Acta*, 1977, **44**, 129–138.
81 M. A. Spackman and P. G. Byrom, *Chem. Phys. Lett.*, 1997, **267**, 215–220.
82 J. J. McKinnon, M. A. Spackman and A. S. Mitchell, *Acta Crystallogr., Sect. B: Struct. Sci.*, 2004, **60**, 627–668.
83 X. Shen, T. Zhang, S. Broderick and K. Rajan, *Mol. Syst. Des. Eng.*, 2018, **3**, 826–838.
84 T. Bayes, *Philos. Trans. R. Soc. London*, 1763, **53**, 370–418.
85 B. Hayes, *Am. Sci.*, 2013, **101**, 252.
86 A. M. Turing, *Mind*, 1950, **LIX**, 433–460.
87 W. S. Bainbridge, *Science*, 1993, **261**, 1186–1187.
88 G. Moser, *AI Mag.*, 1990, vol. 11, pp. 10–11.
89 F. Rosenblatt, *Psychol. Rev.*, 1958, **65**, 386–408.
90 D. Crevier, *AI: The Tumultuous History of the Search for Artificial Intelligence*, Basic Books, London and New York, 1993.
91 Y. He, E. D. Cubuk, M. D. Allendorf and E. J. Reed, *J. Phys. Chem. Lett.*, 2018, **9**, 4562–4569.
92 S. Yang, M. Lach-hab, I. I. Vaisman and E. Blaisten-Barojas, *J. Phys. Chem. C*, 2009, **113**, 21721–21725.
93 M. Pardakhti, E. Moharreri, D. Wanik, S. L. Suib and R. Srivastava, *ACS Comb. Sci.*, 2017, **19**, 640–645.
94 X. Wu, S. Xiang, J. Su and W. Cai, *J. Phys. Chem. C*, 2019, **123**, 8550–8559.
95 G. Borboudakis, T. Stergiannakos, M. Frysali, E. Klontzas, I. Tsamardinos and G. E. Froudakis, *npj Comput. Mater.*, 2017, **3**, 40.
96 M. Fernandez, P. G. Boyd, T. D. Daff, M. Z. Aghaji and T. K. Woo, *J. Phys. Chem. Lett.*, 2014, **5**, 3056–3060.
97 B. J. Bucior, N. S. Bobbitt, T. Islamoglu, S. Goswami, A. Gopalan, T. Yildirim, O. K. Farha, N. Bagheri and R. Q. Snurr, *Mol. Syst. Des. Eng.*, 2019, **4**, 162–174.
98 S. M. Moosavi, A. Chidambaram, L. Talirz, M. Haranczyk, K. C. Stylianou and B. Smit, *Nat. Commun.*, 2019, **10**, 539.
99 G. S. Fanourgakis, K. Gkagkas, E. Tylianakis and G. E. Froudakis, *J. Am. Chem. Soc.*, 2020, **142**, 3814–3822.
100 C. Zhao, L. Chen, Y. Che, Z. Pang, X. Wu, Y. Lu, H. Liu, G. M. Day and A. I. Cooper, *Nat. Commun.*, 2021, **12**, 817.
101 P. Tholen, C. A. Peeples, R. Schaper, C. Bayraktar, T. S. Erkal, M. M. Ayhan, B. Çoşut, J. Beckmann, A. O. Yazaydin, M. Wark, G. Hanna, Y. Zorlu and G. Yücesan, *Nat. Commun.*, 2020, **11**, 3180.
102 J. Liang, S. Xing, P. Brandt, A. Nuhnen, C. Schlüsener, Y. Sun and C. Janiak, *J. Mater. Chem. A*, 2020, **8**, 19799–19804.
103 L. Turcani, R. L. Greenaway and K. E. Jelfs, *Chem. Mater.*, 2019, **31**, 714–727.
104 Y. H. Zhao, M. H. Abraham and A. M. Zissimos, *J. Org. Chem.*, 2003, **68**, 7368–7373.
105 J. H. van't Hoff, *La Chemie dans L'Espace*, Bazendijk, Rotterdam, 1875.
106 T. M. van der Spek, *Ann. Sci.*, 2006, **63**, 157–177.
107 E. W. Meijer, *Angew. Chem., Int. Ed.*, 2001, **40**, 3783–3789.
108 H. D. Flack, *Acta Crystallogr., Sect. A: Found. Crystallogr.*, 2009, **65**, 371–387.
109 L. Pasteur, *C. R. Hebd. Séances Acad. Sci.*, 1878, **86**, 1037–1043.
110 P.-S. Laplace, *Théorie analytique des probabilitiés*, 1812.
111 J. M. Bijvoet, A. F. Peerdeman and A. J. van Bommel, *Nature*, 1951, **168**, 271–272.
112 C. G. Jones, M. W. Martynowycz, J. Hattne, T. J. Fulton, B. M. Stoltz, J. A. Rodriguez, H. M. Nelson and T. Gonen, *ACS Cent. Sci.*, 2018, **4**, 1587–1592.
113 M. W. Lodewyk, M. R. Siebert and D. J. Tantillo, *Chem. Rev.*, 2012, **112**, 1839–1862.
114 P. H. Willoughby, M. J. Jansma and T. R. Hoye, *Nat. Protoc.*, 2014, **9**, 643–660.
115 K. C. Nicolaou and S. A. Snyder, *Angew. Chem., Int. Ed.*, 2005, **44**, 1012–1044.
116 Y. Inokuma, T. Arai and M. Fujita, *Nat. Chem.*, 2010, **2**, 780–783.
117 Y. Inokuma, S. Yoshioka, J. Ariyoshi, T. Arai, Y. Hitora, K. Takada, S. Matsunaga, K. Rissanen and M. Fujita, *Nature*, 2013, **495**, 461.
118 Y. Inokuma, S. Yoshioka, J. Ariyoshi, T. Arai and M. Fujita, *Nat. Protoc.*, 2014, **9**, 246–252.
119 Y. Inokuma, T. Ukegawa, M. Hoshino and M. Fujita, *Chem. Sci.*, 2016, **7**, 3910–3913.
120 S. Yoshioka, Y. Inokuma, M. Hoshino, T. Sato and M. Fujita, *Chem. Sci.*, 2015, **6**, 3765–3768.
121 S. Urban, R. Brkljača, M. Hoshino, S. Lee and M. Fujita, *Angew. Chem., Int. Ed.*, 2016, **55**, 2678–2682.
122 S. Lee, M. Hoshino, M. Fujita and S. Urban, *Chem. Sci.*, 2017, **8**, 1547–1550.
123 Y. Matsuda, T. Mitsuhashi, S. Lee, M. Hoshino, T. Mori, M. Okada, H. Zhang, F. Hayashi, M. Fujita and I. Abe, *Angew. Chem., Int. Ed.*, 2016, **55**, 5785–5788.

124 S. Lee, E. A. Kapustin and O. M. Yaghi, *Science*, 2016, **353**, 808–811.
125 K. Ohara, M. Kawano, Y. Inokuma and M. Fujita, *J. Am. Chem. Soc.*, 2010, **132**, 30–31.
126 A. Burgun, C. J. Coghlan, D. M. Huang, W. Chen, S. Horike, S. Kitagawa, J. F. Alvino, G. F. Metha, C. J. Sumby and C. J. Doonan, *Angew. Chem., Int. Ed.*, 2017, **56**, 8412–8416.
127 V. Duplan, M. Hoshino, W. Li, T. Honda and M. Fujita, *Angew. Chem., Int. Ed.*, 2016, **55**, 4919–4923.
128 A. B. Cuenca, N. Zigon, V. Duplan, M. Hoshino, M. Fujita and E. Fernández, *Chem. – Eur. J.*, 2016, **22**, 4723–4726.
129 S. Yoshioka, Y. Inokuma, V. Duplan, R. Dubey and M. Fujita, *J. Am. Chem. Soc.*, 2016, **138**, 10140–10142.
130 E. Sanna, E. C. Escudero-Adán, A. Bauzá, P. Ballester, A. Frontera, C. Rotger and A. Costa, *Chem. Sci.*, 2015, **6**, 5466–5472.
131 G.-H. H. Ning, K. Matsumura, Y. Inokuma and M. Fujita, *Chem. Commun.*, 2016, **52**, 7013–7015.
132 M. Hoshino, A. Khutia, H. Xing, Y. Inokuma and M. Fujita, *IUCrJ*, 2016, **3**, 139–151.
133 P. M. Bhatt and G. R. Desiraju, *CrystEngComm*, 2008, **10**, 1747.
134 M. Khan, V. Enkelmann and G. Brunklaus, *J. Am. Chem. Soc.*, 2010, **132**, 5254–5263.
135 A. L. Albright and J. M. White, *Determination of Absolute Configuration Using Single Crystal X-Ray Diffraction*, Humana Press, Totowa, NJ, 2013, pp. 149–162.
136 M. B. J. Atkinson, S. V. S. Mariappan, D.-K. Bucar, J. Baltrusaitis, T. Friscic, N. G. Sinada and L. R. MacGillivray, *Proc. Natl. Acad. Sci. U. S. A.*, 2011, **108**, 10974–10979.
137 S. Parsons, H. D. Flack and T. Wagner, *Acta Crystallogr., Sect. B: Struct. Sci., Cryst. Eng. Mater.*, 2013, **69**, 249–259.
138 A. L. Thompson and D. J. Watkin, *Tetrahedron: Asymmetry*, 2009, **20**, 712–717.
139 B. R. Brummel, K. G. Lee, C. D. McMillen, J. W. Kolis and D. C. Whitehead, *Org. Lett.*, 2019, **21**, 9622–9627.
140 F. Krupp, W. Frey and C. Richert, *Angew. Chem., Int. Ed.*, 2020, **59**, 15875–15879.
141 W. Xiao, C. Hu and M. D. Ward, *J. Am. Chem. Soc.*, 2014, **136**, 14200–14206.
142 H. Abe, T. Kobayashi, N. Hoshino, T. Takeda, Y. Suzuki, J. Kawamata and T. Akutagawa, *CrystEngComm*, 2021, **23**, 1149–1157.
143 Y. Li, M. Handke, Y.-S. Chen, A. G. Shtukenberg, C. T. Hu and M. D. Ward, *J. Am. Chem. Soc.*, 2018, **140**, 12915–12921.
144 W. Yang, A. Greenaway, X. Lin, R. Matsuda, A. J. Blake, C. Wilson, W. Lewis, P. Hubberstey, S. Kitagawa, N. R. Champness and M. Schröder, *J. Am. Chem. Soc.*, 2010, **132**, 14457–14469.
145 M. Mastalerz and I. M. Oppel, *Angew. Chem., Int. Ed.*, 2012, **51**, 5252–5255.
146 F. Hu, C. Liu, M. Wu, J. Pang, F. Jiang, D. Yuan and M. Hong, *Angew. Chem., Int. Ed.*, 2017, **56**, 2101–2104.
147 T.-U. Yoon, S. Bin Baek, D. Kim, E.-J. Kim, W.-G. Lee, B. K. Singh, M. S. Lah, Y.-S. Bae and K. S. Kim, *Chem. Commun.*, 2018, **54**, 9360–9363.
148 C. A. Zentner, H. W. H. Lai, J. T. Greenfield, R. A. Wiscons, M. Zeller, C. F. Campana, O. Talu, S. A. FitzGerald and J. L. C. Rowsell, *Chem. Commun.*, 2015, **51**, 11642–11645.
149 P. Li, Y. He, Y. Zhao, L. Weng, H. Wang, R. Krishna, H. Wu, W. Zhou, M. O'Keeffe, Y. Han and B. Chen, *Angew. Chem., Int. Ed.*, 2015, **54**, 574–577.
150 P. Li, Y. He, H. D. Arman, R. Krishna, H. Wang, L. Weng and B. Chen, *Chem. Commun.*, 2014, **50**, 13081–13084.
151 H. Wang, B. Li, H. Wu, T.-L. Hu, Z. Yao, W. Zhou, S. Xiang and B. Chen, *J. Am. Chem. Soc.*, 2015, **137**, 9963–9970.
152 W. Yang, B. Li, H. Wang, O. Alduhaish, K. Alfooty, M. A. Zayed, P. Li, H. D. Arman and B. Chen, *Cryst. Growth Des.*, 2015, **15**, 2000–2004.
153 W. Yang, F. Yang, T.-L. Hu, S. C. King, H. Wang, H. Wu, W. Zhou, J.-R. Li, H. D. Arman and B. Chen, *Cryst. Growth Des.*, 2016, **16**, 5831–5835.
154 X.-Z. Luo, X.-J. Jia, J.-H. Deng, J.-L. Zhong, H.-J. Liu, K.-J. Wang and D.-C. Zhong, *J. Am. Chem. Soc.*, 2013, **135**, 11684–11687.
155 C. Graham, D. A. Imrie and R. E. Raab, *Mol. Phys.*, 1998, **93**, 49–56.
156 V. A. Russell, C. C. Evans, W. Li and M. D. Ward, *Science*, 1997, **276**, 575–579.
157 K. T. Holman and M. D. Ward, *Angew. Chem., Int. Ed.*, 2000, **39**, 1653–1656.
158 B. Wang, R.-B. Lin, Z. Zhang, S. Xiang and B. Chen, *J. Am. Chem. Soc.*, 2020, **142**, 14399–14416.
159 J. Luo, J.-W. Wang, J.-H. Zhang, S. Lai and D.-C. Zhong, *CrystEngComm*, 2018, **20**, 5884–5898.
160 D. Yang and B. C. Gates, *ACS Catal.*, 2019, **9**, 1779–1798.
161 F. Toda, K. Tanaka and A. Sekikawa, *J. Chem. Soc., Chem. Commun.*, 1987, 279–280.
162 K. Endo, T. Koike, T. Sawaki, O. Hayashida, H. Masuda and Y. Aoyama, *J. Am. Chem. Soc.*, 1997, **119**, 4117–4122.
163 Z. Zhang, J. Li, Y. Yao and S. Sun, *Cryst. Growth Des.*, 2015, **15**, 5028–5033.
164 B. Han, H. Wang, C. Wang, H. Wu, W. Zhou, B. Chen and J. Jiang, *J. Am. Chem. Soc.*, 2019, **141**, 8737–8740.
165 W. Gong, D. Chu, H. Jiang, X. Chen, Y. Cui and Y. Liu, *Nat. Commun.*, 2019, **10**, 600.
166 P. Li, Y. He, J. Guang, L. Weng, J. C.-G. Zhao, S. Xiang and B. Chen, *J. Am. Chem. Soc.*, 2014, **136**, 547–549.
167 D. S. Reddy, S. Duncan and G. K. H. Shimizu, *Angew. Chem., Int. Ed.*, 2003, **42**, 1360–1364.
168 X.-Y. Wang, R. Justice and S. C. Sevov, *Inorg. Chem.*, 2007, **46**, 4626–4631.
169 S. A. Dalrymple and G. K. H. Shimizu, *J. Am. Chem. Soc.*, 2007, **129**, 12114–12116.
170 X.-Y. Wang and S. C. Sevov, *Chem. Mater.*, 2007, **19**, 4906–4912.
171 B. F. Abrahams, M. G. Haywood and R. Robson, *J. Am. Chem. Soc.*, 2005, **127**, 816–817.

 This journal is © The Royal Society of Chemistry and IChemE 2021