CrystEngComm

Accepted Manuscript

ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/crystengcomm

Isoreticular Isomerism in 4,4-connected Paddle-wheel Metal-Organic Frameworks: Structural Prediction by the Reverse Topological Approach

Sareeya Bureekaew[1,2]*, Vishal Balwani[1], Saeed Amirjalayer[1,3], Rochus Schmid[1]*

[1] Lehrstuhl für Anorganische Chemie 2, Computational Materials Chemistry Group, Ruhr-Universität Bochum, Universitätsstr. 150, D-44780 Bochum, Germany.

[2] Current Address: Division of Chemistry and Biological Chemistry, School of Physical and Mathematical Sciences, Nanyang Technological University, 21 Nanyang Link, 637371, Singapore

[3] Current Address: Physical Institute and Center for Nanotechnology (CeNTech), Westfälische Wilhelms-Universität Münster, 48149 Münster, Germany

* Email: Sbureekaew@ntu.edu.sg, Rochus.Schmid@rub.de

Abstract

The theoretical structure prediction for a series of 4,4-connected copper paddle-wheel metal-organic frameworks has been performed by using the reverse topological approach, starting from the **nbo-b** topology. Since the rectangular shaped tetracarboxylate linkers have a lower symmetry than the square vertices in **nbo-b**, two alternative insertion modes are possible for each linker. This leads in principle to multiple isoreticular isomers, which have been screened by a genetic algorithm global minimum search, using the first principles parametrized force field MOF-FF for structure optimization and ranking. It is found that isoreticular isomerism does in this case not lead to disorder but to a number of well defined, but structurally distinct phases, which all share the same network topology but have substantially different pore shapes and properties. In all cases, the experimentally observed structure is correctly predicted, but in addition a number of other, slightly less stable phases are observed. Only one of these phases has been synthesized, yet. The theoretical analysis of molecular model systems of the pore cages revealed the reasons for the trends in conformational energy. The proof-of-concept study demonstrates that a screening of isoreticular isomerism, using an efficient but accurate force field, allows to predict the atomistic structure of even complex and flexible frameworks.

1 Introduction

Metal-organic frameworks (MOFs) have received much attention in recent years as a new class of crystalline porous material due to their potential in various fields of application, like for example gas storage, -separation and -sensing or catalysis.[1–8] The main advantages of MOFs over other porous materials like zeolites are derived from their properties, e.g., high surface areas, large internal pore volumes (void space), tunable pore size and shape, and structural transformations upon external stimulation. Importantly, these properties depend not only on the chemical composition but also on the geometrical structure of the system. Thus, by tuning the structure, it is in principle possible to generate MOFs with desired properties, like for example appropriate pore shapes and sizes for a desired adsorption selectivity.

With the introduction of new inorganic building blocks, usually referred to as secondary building blocks (SBUs), and polytopic organic linkers the number of accessible network topologies and the overall structural complexity is constantly increasing.[9] In order to analyze, rationalize and compare such crystalline polymers the concept of topological analysis has proven to be very successful.[9–12] The network topology describes the structure of the framework by a graph of nodes or vertices and their connections. Since the topology depends only on the way the nodes are connected, but not on their actual shape or size, it is perfectly suited to reduce the complexity and reveal the underlying structural motifs. The topological approach, which is also referred to as "deconstruction" of a MOF, is schematically given in Fig. 1 from left to right. In the first step, the collection of atoms to be condensed into one vertex needs to be identified. Note that there is no specific rule to define the SBUs for particular MOFs. A usual way is to maximize the symmetry of the nodes.[11] In the second step the topology can then be identified. The computer program TOPOS can be used to facilitate this process.[13] Collections of experimentally observed or theoretically predicted network topologies are available in databases like the Reticular Chemistry Structure Resource (RCSR).[14] It is crucial to realize that this process on purpose reduces the amount of information contained in the description of the system by removing the atomistic resolution, including possible disorder or supramolecular isomerism.

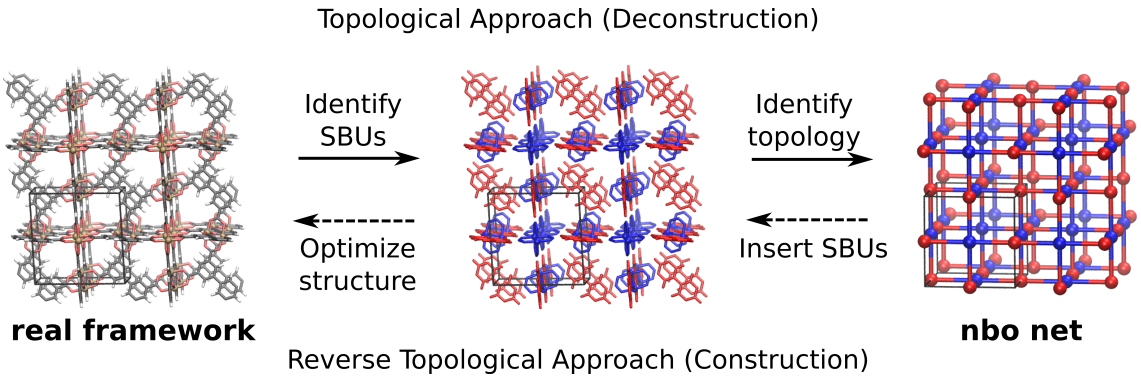


Figure 1: Schematic representation of the reverse topological approach (RTA) as the reversed process of "deconstruction" of a real MOF structure to its underlying topology.

In order to go beyond a simple "trial-and-error" based synthesis of new MOFs, recently a number of theoretical strategies for the generation of hypothetical MOF structures have been proposed, which have nicely been reviewed in ref. 15. Wilmer et al. developed a method to generate MOF structures in a "bottom-up" type of approach by a geometrical assembly algorithm using preformed building blocks. A large database with more than 137,000 hypothetical MOF structures was generated and in turn computationally screened to identify potential candidates for high methane-storage capacity.[16] However, since the frameworks were generated from rigid building blocks and no energy function was used to relax the structures, some chemically not sensible structures have been included and only a rather small number of topologies was formed. Clearly, the target in this study was to achieve a massive computational high-throughput screening.

A viable alternative is to use known topologies as a starting point in a "top-down" fashion for the construction and to enumerate all possible nets for the given connectivity of vertices, starting from the high symmetry cases.[17] As shown in Fig. 1 this means to revert the topological approach (from right to left), constructing the atomistic structure in a two step process. First, the building blocks are inserted at the vertex positions. Then the framework is structurally optimized using an accurate force field, such as the recently developed first principles parameterized MOF-FF.[18] Such force fields, which are fitted to structure and low normal modes of model systems computed on density functional theory (DFT) level, were shown to reproduce structure and deformation of MOFs as well as dynamic properties like negative thermal expansion (NTE).[19,20] This strategy, which we call reverse topological approach (RTA), was previously used to analyze the topological preference of copper paddle-wheel MOFs like HKUST-1 and MOF-14[21] as well as boron based 3D covalent organic frameworks (COFs).[22,23] It was also employed to predict the structures of hypothetical boron based 3D COFs considering both edge transitive and non edge transitive topologies.[17] Here, the relative stabilities of topological isomers predicted by MOF-FF could successfully be validated by a periodic DFT method for the smallest monomer. A somewhat similar approach has been employed by Baburin et al. and later Lewis et al. for the analysis of zeolitic imidazolate frameworks (ZIFs) using known zeolite topologies. In this case a periodic DFT method was used for structure relaxation.[24–26] Very recently, Martin and Haranczyk proposed another "top-down" construction scheme for hypothetical MOFs, which is restricted, however, to a simple collision detection scheme to avoid unphysical structures and is intended for the generation of databases for high-throughput screening.[27]

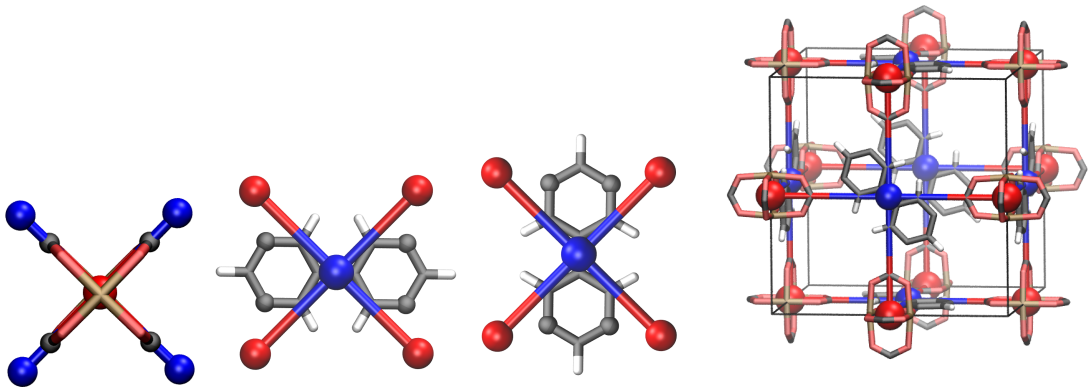
It is crucial to realize that, in parallel to the loss of information upon reducing a real framework to its underlying topology, there is in reverse not always a unique solution for the RTA. This is due to the molecular nature of the building blocks, which can have a lower symmetry than the local symmetry of the vertex. An example is the tetrahedral Zn_4O unit in the IRMOF series,[10,28] which lacks the center of inversion of the octahedral vertex.[29] In this case there is more than one equivalent way to insert the SBU or the linker. As a consequence, besides the topological isomerism, a further form of supramolecular isomerism[30] needs to be considered, which we refer to as *isoreticular isomerism.* Isoreticular isomers share the same topology but are formed during

the self-assembly and cannot be inter-converted without bond breaking. If they are energetically very close this will result in a disorder of the structure. As a consequence, a fundamental part of the RTA is an efficient search strategy for the global energy minimum in the space of possible isoreticular isomers. Note, that this approach makes a numerically efficient but accurate force field method for the structure optimization and ranking inevitable, and none of the other "top-down" strategies described in ref. 15 contain such a conformational screening step. We have developed the computer program *weaver* to automatically detect potential isoreticular isomerism and to solve the global minimum problem via a genetic algorithm (GA),[31] which is briefly introduced in the next section. A detailed description of the underlying algorithms and their implementation in the *weaver* code will be given in a separate publication.[32]

From a wider perspective, the energy based "top-down" approach in the RTA — as compared to the purely geometric "bottom-up" strategy — is more related to the extensive work on the topological screening of zeolite materials,[33–35] which lead to databases of hypothetical zeolites and related materials[36–38] and the famous observation of the density/stability correlation for porous materials. For a in depth discussion on zeolite structure prediction and we refer here to a recent review in ref. 39. Note, however, that in case of zeolite structure prediction, because of the simple tetrahedral atomic building blocks, various ways of a global search for new topologies are performed. In contrast to that, in the RTA for MOFs known (high symmetry) topologies are needed as an input. On the other hand, only because of the more complex molecular building blocks in MOFs, allowing for a lower symmetry then the network vertex, isoreticular isomerism can arise and needs to be screened by a global minimum search.

From an experimental point of view, both forms of supramolecular isomerism have been observed. For example, in case of paddle-wheel based MOFs, Kondo et al. reported the synthesis of topological isomeric forms of $[Zn_2(bdc)_2(dabco)]_n$ and $[Zn_2(bdc)_2(bpy)]_n$ (bdc = 1,4-benzenedicarboxylate, dabco = diazabicyclo[2,2,2]octane, bpy = 4,4-bipyridyl) by tuning the synthesis conditions.[40] The 2D network, pillared by dabco or bipy, can form either a square grid or the Kagomé topology. Another example are the Zn_4O based topological isomers of MOF-177 synthesized by Caskey et al..[41] More importantly, two isomeric MOFs have been synthesized by Sun et al. from Cu(II) and the tetracarboxylate linker 5,5´-(1,2-ethynediyl)bis(1,3-benzenedicarboxylate) (ebdc, L-2 in Fig. 3).[42] The 4-connected paddle-wheel SBUs and the tetratopic linkers form a 4,4-connected network with **nbo-b** topology. Depending on the synthesis conditions, two structurally different phases PCN-16 and PCN-16' with the same space group but different sorption properties were crystallized. This phenomenon was termed "symmetry preserving isomerism". However, because of the rectangular shape of the tetratopic linker, two possible modes of insertion on a square vertex position in the **nbo-b** net are possible, which leads to isoreticular isomerism (see Fig. 2).

It was this finding which motivated us to investigate the isoreticular isomerism for copper paddle-wheel based MOFs in the **nbo-b** topology with three different tetratopic carboxylate linkers of different backbone length (see Fig. 3 (b-d)) by the RTA. Note that the corresponding experimentally known MOF with linker L-1 is either referred to as MOF-505[43] or NOTT-100[44]. The synthesized



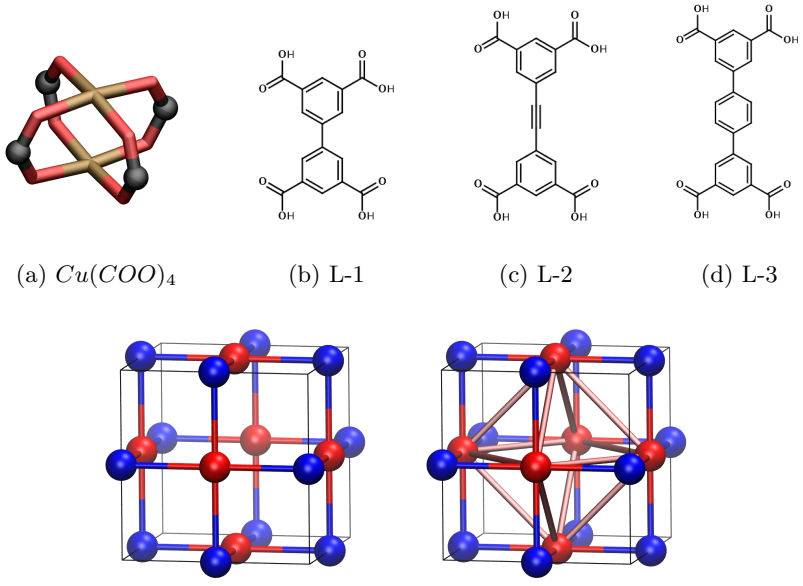
(a) Square SBU (b) Two ways to insert a rectangular linker (c) Building blocks inserted

Figure 2: Insertion of the square SBU (a) and the rectangular linker (b) into the **nbo-b** network (c).

MOF with linker L-3 is called NOTT-101[44]. Because of the open metal sites in these systems, they are of relevance in hydrogen storage applications. To our knowledge, for these systems no mention of different phases or isomeric structures is made. We intended to answer the question, whether further isomeric forms are possible in principle and if our RTA method is able to predict the structures and stabilities correctly. In addition, the theoretical analysis should shed light on the molecular origin for the stability tendencies. At the current stage our method considers only conformational energies of the MOF itself, whereas finite temperature effects (e.g. entropic contributions) and solvation is ignored. It is clear that under solvothermal conditions the isomer with the lower free energy of MOF and solvent will be formed and under proper synthesis conditions a MOF with a somewhat higher strain energy could be formed. We are currently extending our RTA to include also solvation and finite temperature effects in an approximate way, following the ideas recently presented by Gee and Sholl.[45]

2 Computational Methods

All systems have been constructed by our *weaver* code[46] from a $2 \times 2 \times 2$ supercell of the **nbo-b** topology with periodic boundary conditions. Possible insertion modes for both the $Cu_2(O_2C)_4$ SBUs and the tetracarboxylate linkers were detected. For all linkers two modes were detected, whereas for the paddle-wheel only one insertion mode can be realized. The system contains 48 vertices with 24 SBUs and 24 linker units and a total of 24 formula units (f.u.). Thus, in principle 2^{24} (more than 16×10^6) isoreticular isomers are possible, many of them are however related by symmetry. The *weaver* code implements a standard GA (always best individual maintained in the next generation, tournament style selection rules with crossover and constant mutation rate), similar to the one used previously for the conformational isomers in IRMOFs.[29] The different isomers were



(a) $Cu(COO)_4$ (b) L-1 (c) L-2 (d) L-3

(e) **nbo-b** net (f) octahedral cage from 6 red and 12 blue vertices.

Figure 3: Schematic representation of $Cu_2(CO_2)_4$ SBUs, tetracarboxylate linkers, and the **nbo-b** topology considered in this study.

encoded by a binary chromosome of length 24. To evaluate the fitness of an individual (potential energy of the relaxed system), the structure was generated automatically and in turn optimized without any symmetry constraints ($P1$ symmetry, allowing for triclinic systems) with the force field MOF-FF[18] with our *pydlpoly*[18] molecular mechanics code with a threshold of 0.01 kcal/(mol Å) for the root mean force of structure and lattice parameters. All non-bonded interactions were treated by truncated 2-body potentials with a cutoff of 12.0 Å, using the shifted force methods by Fennell and Gezelter[47] for the electrostatic interactions of the Gaussian charge distributions, which are used in MOF-FF. Note that avoiding Ewald summation is beneficial for the large isomer screening needed in this case. It is always possible to re-optimize the global minimum structures by an Ewald summation based method, but we found no changes in the relative ordering of the supramolecular isomers. Note that the $2 \times 2 \times 2$ supercell was chosen to allow the formation of phases with lower symmetry, not possible in a $1 \times 1 \times 1$ setup, and in order to match with the 12.0 Å cutoff in the minimum image convention.

The final structures have been analyzed by the detecting the space group and primitive unit cell using the program library *spglib*.[48] Since the fully atomistic structure (including hydrogen atoms) is used in the symmetry analysis, it is possible that the full space group symmetry is not detected due to disorder, despite a rather loose threshold in the symmetry detection ($symprec = 0.5$). This is especially the case for linker L-3, where the central phenyl unit can easily flip between two positions. Geometric surface areas where computed by sampling the surface, accessible by a spherical probe molecule with the kinetic diameter of nitrogen (3.681 Å).[49] It was shown that

this geometric surface area matches well with the results from a surface area derived from a fit to the nitrogen adsorption isotherm.[50] For the computation we used the code provided by Düren for non-orthorhombic systems.[51]

3 Results and Discussion

3.1 Investigated Systems

In this proof-of concept investigation, employing the RTA for MOFs with isoreticular isomerism, we have intentionally focused on experimentally known systems, and varied the tetracarboxylate linkers systematically from a small to a large aspect ratio. Measured as the ratio of the short to the long carboxylate carbon atom distance, it ranges from 1:1.44 for L-1 over 1:1.97 for L-2 up to 1:2.27 for L-3, L2 and L-3, respectively. The rectangular shape is the reason for the two insertion modes available for each linker on the square vertex position in the **nbo-b** topology, as shown in Fig. 2. We refer to these systems as NBO-1 to NBO-3, labeling the low energy isoreticular isomers or phases by an appended roman number.

Table 1: Relative energies, space group, primitive cell parameters and surface ares of first five stable phases in NBO-1, NBO-2 and NBO-3. Space group and primitive cell parameters are given as detected by *spglib*[48] for the fully optimized systems.

NBO-1	NBO-1-I	NBO-1-II	NBO-1-III	NBO-1-IV	NBO-1-V
$E[kcal\ mol^{-1}]$	0.0	0.4	0.6	2.5	2.5
space group	$R\bar{3}m$	$R\bar{3}m$	$C2/m$	$P\bar{1}$	$P1$
a/b/c [Å]	13.6/13.6/13.6	13.6/13.6/13.6	13.5/13.5/27.1	13.5/27.1/27.1	13.6/27.00/27.0
$\alpha/\beta/\gamma$ [^{o}]	93.1/93.1/86.9	92.7/92.7/92.7	87.5/92.5/ 90.2	93.4/92.1/87.5	93.0/88.2/91.7
S [m^2/g]	1842	1699	1653	1768	1778
NBO-2	NBO-2-I	NBO-2-II	NBO-2-III	NBO-2-IV	NBO-2-V
$E[kcal\ mol^{-1}]$	0.0	1.5	2.4	4.7	5.4
space group	$R\bar{3}m$	$R\bar{3}m$	$P\bar{1}$	$C2/m$	$C2/m$
a/b/c [Å]	15.6/15.6/15.6	15.4/14.5/15.4	15.5/ 15.5/29.2	15.5/30.0/30.0	15.4/30.1/30.1
$\alpha/\beta/\gamma$ [^{o}]	106.4/106.4/73.6	104.9/104.9/104.9	97.1/103.3/74.6	96.1/102.9/77.1	98.7/104.0/104.0
S [m^2/g]	2867	2617	2667	2771	2514
NBO-3	NBO-3-I	NBO-3-II	NBO-3-III	NBO-3-IV	NBO-3-V
$E[kcal\ mol^{-1}]$	0.0	2.9	4.0	6.3	7.3
space group	$R\bar{3}m^a$	$R\bar{3}m$	$P\bar{1}$	$P\bar{1}$	$P\bar{1}$
a/b/c [Å]	16.8/16.8/16.8^a	16.4/16.4/16.4	16.5/16.7/28.4	16.5/30.0/30.0	16.7/30.0/30.0
$\alpha/\beta/\gamma$ [^{o}]	112.1/112.2/67.8^a	108.4/108.4/108.4	99.4/99.2/69.9	108.9/99.0/99.0	93.8/81.7/98.2
S [m^2/g]	2961	2078	2333	2213	2531

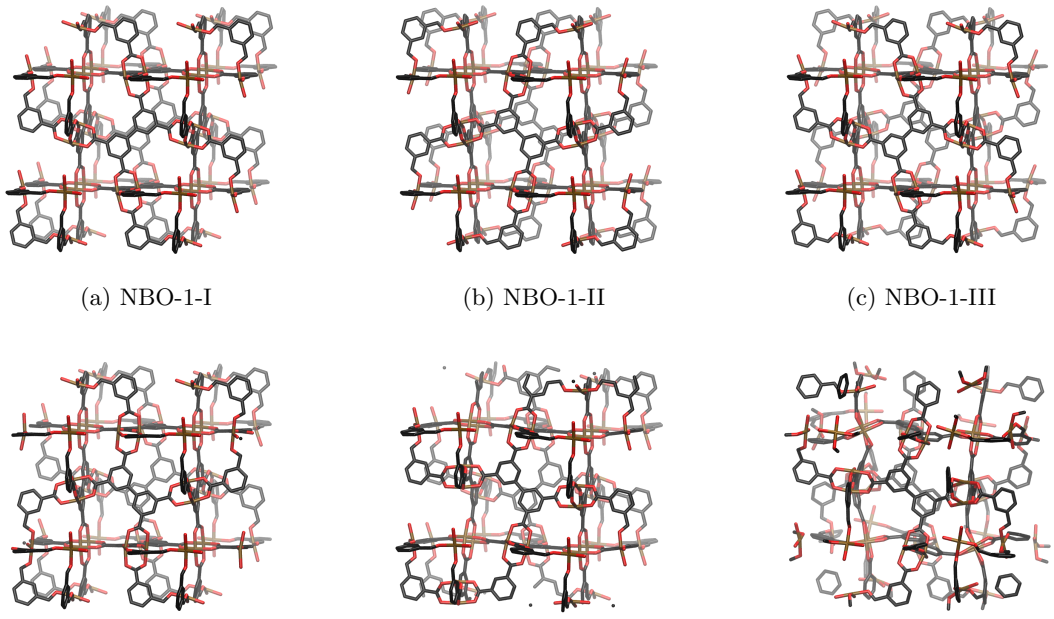
[a] Due to disorder in the linker backbone a space group detection and primitive cell search it *spglib* was only possible for a structure with the linker atoms removed.

3.2 NBO-1

First we report our results for the NBO-1 system, which is constructed from the shortest linker, 3,3',5,5'-biphenyltetracarboxylate (Fig.3(b)). It can be expected that for this short linker with the smallest aspect ratio the effects of different insertion modes is the least dominant. The GA screening was performed starting from a random population of 80 individuals and was run for 100 generations. The global minimum isoreticular isomer NBO-1-I was already found in the 7th generation, whereas the second best structure was bread only in the 97th generation of the GA run. A total of 1863 different chromosomes were screened by full optimization of the corresponding structure, which is substantially less than the theoretical 2^{24} potential structures, indicating the efficiency of the GA global search approach. However, the mere number of completely relaxed structures with around 1000 atoms also demonstrates the numerical effort necessary, if the screening would rely on periodic DFT methods instead of a much more efficient force field.

The five most stable phases or isomers NBO-1-I to NBO-1-V resulting from the GA screening will be discussed now in full detail. The relative energies per f.u. with respect to NBO-1-I, the space group and cell parameters (as detected by *spglib*) and the geometric surface area of all systems are given in Tab. 1. The corresponding structures are shown in Fig. 4. It is evident that both NBO-1-I and NBO-1-II are highly ordered and show both a space group symmetry of $R\bar{3}m$ with a unit cell equivalent to the **nbo-b** the unit cell. The isoreticular isomer NBO-1-II is only 0.4 $kcal\,mol^{-1}$ per f.u. less stable in energy than NBO-1-I. Both structures deviate only slightly from the cubic structure of the parent **nbo-b** topology. In Fig. 4 (a) and (b) one can see that the linkers on one of the faces of the cubic pore, spanned by the paddle-wheel units, are oriented all in the same way. However, in NBO-1-II exactly one of these faces is inserted differently. The difference between NBO-1-I and NBO-1-II will be discussed in more detail in the following. The next stable phase NBO-1-III has already a lower space group symmetry of $C2/m$ and a primitive cell extending over two unit cells of the parent topology. The next phases IV and V, which are energetically almost equal are even lower in symmetry. For comparison, in Fig. 4 (f) the isoreticular isomer with the highest energy found in the screening is shown. This structure, which is part of the initial zeroth generation random species, has a a strain energy of 21.4 $kcal\,mol^{-1}$ per f.u. with respect to NBO-1-I. This is due to the large deformations in the paddle-wheel units for this completely erratic orientation of the linkers. This shows that even for a small aspect ratio of 1:1.44 in L-1 there is a large driving force to insert the linkers properly with respect to each other. On the other hand, if a certain order is maintained, structurally different phases with very similar strain energy can be formed.

At this point it is instructive to compare the computed results with the available experimental data. NBO-1 was first reported as MOF-505 by Chen et al.[43] who presented a single crystal X-ray diffraction study. The structure with space group $R\bar{3}m$ is exactly equivalent to the most stable isoreticular isomer NBO-1-I, demonstrating the predictive capabilities of the RTA. Interestingly, in the later study by the Nottingham group a series of **nbo-b** based paddle-wheel MOFs was synthesized,[44] with NOTT-100 made from the same linker L-1 as MOF-505. However, for NOTT-



(a) NBO-1-I (b) NBO-1-II (c) NBO-1-III

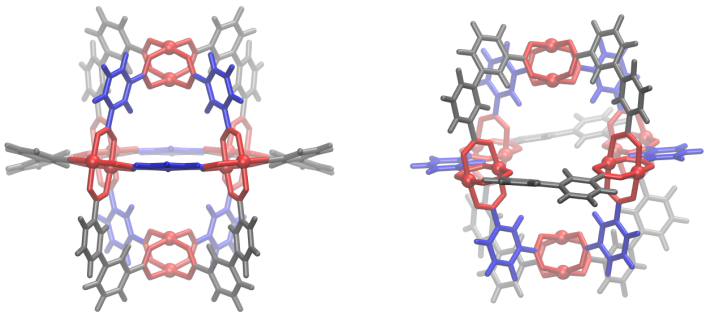
(d) NBO-1-IV (e) NBO-1-V (f) NBO-1 highest energy isomer

Figure 4: Optimized structures of the first five most stable (a-e) and the energetically least stable isomer (f) of NBO-1 (hydrogen atoms omitted for clarity, Cu: brown; C: black; O: red).

100 only powder-XRD (PXRD) data was published, which differs slightly from the PXRD simulated for NBO-1-I or the crystal structure of MOF-505 (see Fig. S1 in the SI). Unfortunately, from a comparison with the simulated PXRD of NBO-1-II to NBO-1-V (Fig. S2) it can not be safely be concluded whether NOT-100 is a pure phase isoreticular isomer of MOF-505. Since the three most stable isomers show a very small energy difference of less than 1.0 $kcal\,mol^{-1}$ per f.u. it should be possible, in principle, to synthesize pure phase crystals of those isomers with optimized synthesis conditions. For the large-scale synthesis of micro-crystalline powders, where kinetic effects on the self-assembly process can become more relevant, it appears likely that mixed phases are formed.

A further interesting result from the screening procedure is the notion that NBO-1-I has with 1842 m^2/g the highest surface area of all isomeric phases. In particular, the slightly less stable NBO-1-II is slightly more dense. This is in contrast to the usual observation that for porous materials stability increases with denser packing. Obviously, the lower surface area in NBO-1-II is accompanied by a less optimal conformation of the building blocks.

In order to understand the fundamental structural differences between NBO-1-I and NBO-1-II one has to focus on the octahedral cage formed from six paddle-wheel units (see Fig. 3 (f)). Because the organic linker is not square planar, a deformed pseudo-octahedral cage is formed in the real network. The cage of six paddle wheels at the tips of the octahedron is held together by twelve bridges forming the edges. These bridges can either be short or long, depending on the orientation of the corresponding linker. For the low energy isoreticular isomers we find that mainly two types of pseudo-octahedral cages, referred to as α-cage and β-cage, are formed, whose pore geometries



(a) α-cage (b) β-cage

Figure 5: The two low energy octahedral PW cages: The α-cage has two small windows opposite to each other, which are constructed from the shorter side of the linkers shown in blue (a). The β-cage possesses two large opposite windows formed by the longer sides of the linkers shown in gray (b).

are completely different. In Fig. 5 these two cages are shown. In both cases six short (blue) and six long (gray) edges are present, and two opposite faces of the pseudo-octahedral cage are constructed from one type of bridge. However, in the α-cage these faces or pore windows are constructed from the short side of the linker, whereas the β cage has two large opposite windows. An analysis of the structures shows that NBO-1-I (as the experimentally observed MOF-505) contains only α-cages, and NBO-1-II is built solely from β-cages. The uniform cage type found in these phases is due to the equal orientation of all linkers for a given axis discussed above. Obviously, the flip of all linkers along one axis from NBO-1-I to NBO-1-II converts all α-cages to β-cages. For NBO-1-III to -V, the orientations of the linkers are less ordered, resulting in various combinations of α- and β-cages. Mixing cage types tends to lower the stability of the frameworks.

3.3 Extended linkers NBO-2 and NBO-3

By the insertion of either $C \equiv C$ or another phenylene unit, the extended linkers L-2 (5,5'-(buta-1,3-diyne-1,4-diyl)-diisophthalate, Fig. 3 (c)) and L-3 (terphenyl-3,3'',5,5''-tetracarboxylate, Fig. 3 (d)) with larger aspect ratios are formed. For the corresponding NBO-2 and NBO-3 systems a similar GA screening of all isoreticular isomers was performed. Again relative strain energies and crystallographic parameters are summarized in Tab. 1 for the five lowest energy isomers. Because of the larger aspect ratio of these linkers we found a larger perturbation due to different insertion modes leading to a quicker convergence of the GA runs. Overall, a similar picture as for NBO-1 is found for both NBO-2 and NBO-3. First of all, two low energy phases with the space group symmetry $R\bar{3}m$ are identified. However, in case of NBO-3 the final structures show a disorder with respect to a flipping of the central phenyl ring of the linker backbone. Therefore it was not possible for the symmetry detection algorithm to find this symmetry for NBO-3-I. By just considering the SBUs atom coordinates the $R\bar{3}m$ space group symmetry was correctly detected. For the higher

energy isomers III to V in both cases the space group symmetry is reduced and the unit cell grows with respect to the unit cell of the underlying **nbo-b** topology.

The most important trend to be observed in the series of NBO-1 to NBO-3 is the fact that the surface area is increasing with increasing linker length and aspect ratio, but also the energy difference between to phase I and II is raising. For NBO-2 with a linker aspect ratio of 1:1.97 it is already 1.5 $kcal\ mol^{-1}$ per f.u. and almost doubles to 2.9 $kcal\ mol^{-1}$ per f.u. for L-3 with an aspect ratio of 1:2.27. In line with that, also the higher energy isomers become less and less stable for the extended linkers. Thus, for NBO-2 and NBO-3 the driving force for a correct linker insertion is increasing and pure phase samples should be achievable. It is interesting to note that in particular for NBO-3 the variations in the geometrical surface area are very dramatic. The more stable NBO-3-I has with 2961 m^2/g nearly a 50% larger surface area as the less stable NBO-3-II.

Experimentally, Sun et. al. were able to synthesize two distinct MOF phases PCN-16 and PCN-16' from L-2 by using different solvent mixtures and a slightly different temperature in the solvothermal synthesis.[42] A comparison of the single crystal XRD results with our computed structures reveals that PCN-16, received at higher temperatures, is equivalent to the more stable NBO-2-I phase and PCN16' is the less stable isoreticular isomer NBO-2-II. The two experimental isomers show different pore shapes and sizes and correspondingly a larger BET surface area of 2270 m^2/g is reported for PCN-16 and a smaller value of 1760 m^2/g for PCN-16'. This is consistent with the tendencies for the geometric surface areas of the computed structures. In case of NBO-3 only one phase has been experimentally synthesized. For L-3 the MOF NOTT-101 with space group $R\bar{3}m$ is again structurally identical to the most stable isoreticular isomer NBO-3-I, computed by the RTA screening procedure. For this system it is questionable, whether the higher energy isomer NBO-3-II is accessible by variations in synthesis conditions, due to the large penalty in strain energy. However, templating effects used in zeolite synthesis might allow the targeted synthesis of such an isoreticular isomer.

3.4 Detailed analysis of the α- and β-cage

In analogy to NBO-1 also for NBO-2 and NBO-3 the energetic preferences can be explained by a detailed investigation of the octahedral cages formed by the paddle-wheel units. Also NBO-2-I and NBO-3-I consist only of α-cages and NBO-2-II and NBO-3-II are formed only from β-cages. Since the energy differences between each isomer in NBO-2 and NBO-3 become larger, it can be concluded that this is due to an increasing deformation of the pseudo octahedral cages. To investigate this trend we performed an investigation of isolated, non-periodic and thus fictitious molecular cages, resembling the α- and β-cage in the periodic frameworks. For this purpose two different bent dicarboxylates were used, which result from a truncation of a C-C bond and replacing it by a C-H bond. The molecular systems were fully optimized by the MOF-FF force field and the resulting structures are shown in Fig. 6. Note that for both cages the same number of six short and six long edges are used and the different systems represent just different conformations of the same molecule.

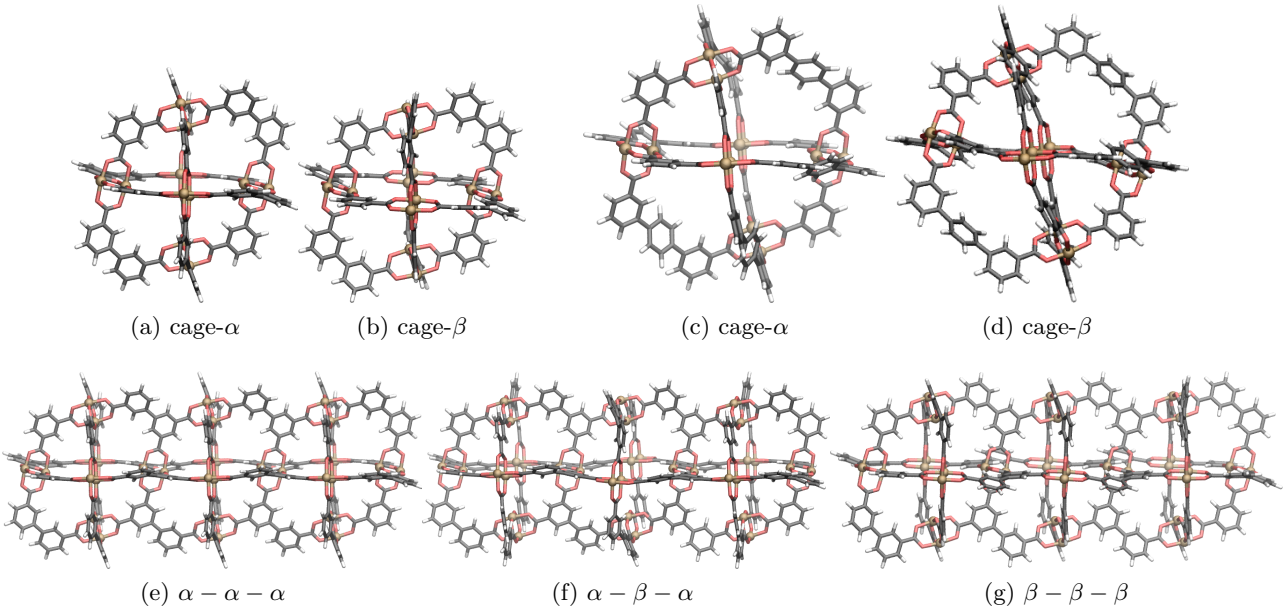
In order to study the stability and the effect of the linker length, we focused first on the α- and β-cages found in NBO-1 and NBO-3. For the fictitious cages of NBO-1, a phenyl unit is used to mimic the short edge linker, while a biphenyl unit represents the long edge linker. The isolated α-cage is by 7.2 $kcal\,mol^{-1}$ more stable than the β-cage. For the extended NBO-3 system, a terphenyl unit is used as the long edge linker. Again the α-cage is more stable than the β-cage with 13.1 $kcal\,mol^{-1}$, which is almost twice the difference for NBO-1. By a visual inspection of the structures shown in Fig. 6 (a) to (d) it is obvious that the paddle wheel in both α-cages are less deformed and the bending for the β-cage of NBO-3 (Fig. 6 (d)) is most pronounced. Thus, it can be concluded that the longer linker introduce more deformations on the paddle-wheel in the β-cage, making these less and less stable.

Moreover, to study the effect of a combination of cages in case of the higher energy isomers, three-cage-clusters where constructed from phenyl and biphenyl units, as they would exist in the NBO-1 phases. As shown in Fig. 6 (e-g) we optimized systems with $\alpha-\alpha-\alpha$, $\beta-\beta-\beta$, and $\alpha-\beta-\alpha$ cages, fused together. As expected, also the $\alpha-\alpha-\alpha$ combination is the most stable cluster, while $\beta-\beta-\beta$, and $\alpha-\beta-\alpha$ are less stable by 17.4 and 20.2 $kcal\,mol^{-1}$, respectively. This trend shows that combining three β-cages is destabilized with respect to $\alpha-\alpha-\alpha$. However, a mixing of cages as $\alpha-\beta-\alpha$ is even less stable. This is also visible in the larger deformations of $\alpha-\beta-\alpha$ shown in Fig. 6 (f). These results explain very well the tendencies in the stabilities we have found for NBO-1, NBO-2 and NBO-3. The most stable isoreticular isomers are always found to consist only of α-cages while the second most stable only consists of β-cages. A mixture of cages leads to higher energy isomeric forms.

The relative stability of these isolated cage structures can not directly be compared to the periodic MOFs, since certain interactions are lost by cutting out the clusters and further long-range interactions are present in the crystalline material. However, the match in the observed tendencies clearly indicate that the local conformation in these cages are the main contribution for the lower conformational energy of the most stable phases.

4 Conclusion

By using the first principles parametrized force field MOF-FF and a systematic isomer screening with a genetic algorithm we succeeded in predicting the atomistic structure of a series of 4,4-connected paddle-wheel MOFs in **nbo-b** topology. Due to the lower symmetry of the rectangular tetracarboxylate linkers with respect to the square vertex in **nbo-b**, different insertion modes are possible. The exact structure is not uniquely defined by the topology alone and a wide variety of isoreticular isomers is possible. Only by an efficient global minimum search a structure prediction is possible in this case. Even for the linker with the smallest aspect ratio the strain energy between the most and least stable isomer differs by more than 20 $kcal\,mol^{-1}$ per formula unit. For all three investigated linkers with increasing aspect ratio always a highly symmetric minimum structure with $R\bar{3}m$ space group symmetry is found, which corresponds to experimentally observed structure.



(a) cage-α (b) cage-β (c) cage-α (d) cage-β

(e) $\alpha-\alpha-\alpha$ (f) $\alpha-\beta-\alpha$ (g) $\beta-\beta-\beta$

Figure 6: Discrete molecular cages of pure and fused α- and β-cages.

Interestingly, a second low energy phase with the same space group symmetry but lower surface area is possible, which has been synthesized only for the medium sized linker L-2.[42] The results demonstrate how the symmetry and size of the organic linkers can affect the stability of isomeric forms. With increasing aspect ratio the second phase becomes more and more unstable. the origin of this behavior can be tracked down to the stability of the pseudo-octahedral cages formed by the paddle-wheel units. The two low energy phases are constructed solely by either an α- or a β-cage, which are formed by a specific arrangement of long or short linker bridges, representing the edges of the octahedron. From calculations of discrete molecular cages it is found that the α-cage is more stable than the β-cage, and for fused cages a mixing of the α and β form is destabilizing.

The reverse topological approach is able to predict the atomistic structures of even complex MOFs on the basis of the geometrical structure of the building blocks and the possible underlying topologies, which are available from databases like RCSR.[14] We are currently extending our investigation of 4,4-connected paddle-wheel MOFs to a wider selection of linkers and other topologies. Note, that the current results indicate that it should be possible to generalize the isomer screening results performed for a specific linker to other, structurally similar linkers, which would allow the incorporation of the RTA into high-throughput screening methods.[15] The approach gives access to structures and other properties like adsorption properties of hypothetical MOFs and can serve as a tool for the future rational design of functional MOFs.

Acknowledgment

This project has financially been supported by the Deutsche Forschungsgemeinschaft (research grants within SFB-558 and Priority Program 1362). VB would like to thank the Research Department IFSC of the Ruhr-University Bochum for financial support.

References

[1] H.-C. Zhou, J. R. Long and O. M. Yaghi, *Chem. Rev.*, 2012, **112**, 673–674.

[2] L. J. Murray, M. Dinca and J. R. Long, *Chem. Soc. Rev.*, 2009, **38**, 1294–1314.

[3] H. Furukawa, K. E. Cordova, M. O'Keeffe and O. M. Yaghi, *Science*, 2013, **341**, 1230444.

[4] M. P. Suh, H. J. Park, T. K. Prasad and D.-W. Lim, *Chem. Rev.*, 2012, **112**, 782–835.

[5] L. E. Kreno, K. Leong, O. K. Farha, M. Allendorf, R. P. Van Duyne and J. T. Hupp, *Chem. Rev.*, 2012, **112**, 1105–1125.

[6] G. Férey, *Chem. Soc. Rev.*, 2008, **37**, 191–214.

[7] S. Horike, S. Shimomura and S. Kitagawa, *Nat Chem*, 2009, **1**, 695–704.

[8] A. Schneemann, V. Bon, I. Schwedler, I. Senkovska, S. Kaskel and R. A. Fischer, *Chem. Soc. Rev.*, 2014, **43**, 6062–6096.

[9] M. Li, D. Li, M. O'Keeffe and O. M. Yaghi, *Chem. Rev.*, 2014, **114**, 1343–1370.

[10] O. M. Yaghi, M. O'Keeffe, N. W. Ockwig, H. K. Chae, M. Eddaoudi and J. Kim, *Nature*, 2003, **423**, 705–714.

[11] D. J. Tranchemontagne, J. L. Mendoza-Cortes, M. O'Keeffe and O. M. Yaghi, *Chem. Soc. Rev.*, 2009, **38**, 1257–1283.

[12] M. O'Keeffe and O. M. Yaghi, *Chem. Rev.*, 2012, **112**, 675–702.

[13] V. A. Blatov, A. P. Shevchenko and D. M. Proserpio, *Cryst. Growth Des.*, 2014, **14**, 3576–3586.

[14] M. O'Keeffe, M. A. Peskov, S. J. Ramsden and O. M. Yaghi, *Acc. Chem. Res.*, 2008, **41**, 1782–1789.

[15] Y. J. Colon and R. Q. Snurr, *Chem. Soc. Rev.*, 2014, **43**, 5735–5749.

[16] 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.

[17] S. Bureekaew and R. Schmid, *CrystEngComm*, 2013, **15**, 1551–1562.

[18] S. Bureekaew, S. Amirjalayer, M. Tafipolsky, C. Spickermann, T. K. Roy and R. Schmid, *Phys. Status Solidi B*, 2013, **250**, 1128–1141.

[19] M. Tafipolsky and R. Schmid, *J. Phys. Chem. B*, 2009, **113**, 1341–1352.

[20] M. Tafipolsky, S. Amirjalayer and R. Schmid, *J. Phys. Chem. C*, 2010, **114**, 14402–14409.

[21] S. Amirjalayer, M. Tafipolsky and R. Schmid, *J. Phys. Chem. C*, 2011, **115**, 15133–15139.

[22] R. Schmid and M. Tafipolsky, *J. Am. Chem. Soc.*, 2008, **130**, 12600–12601.

[23] S. Amirjalayer, R. Q. Snurr and R. Schmid, *J. Phys. Chem. C*, 2012, **116**, 4921–4929.

[24] I. A. Baburin, S. Leoni and G. Seifert, *J. Phys. Chem. B*, 2008, **112**, 9437–9443.

[25] I. A. Baburin and S. Leoni, *CrystEngComm*, 2010, **12**, 2809–2816.

[26] D. W. Lewis, A. R. Ruiz-Salvador, A. Gomez, L. M. Rodriguez-Albelo, F.-X. Coudert, B. Slater, A. K. Cheetham and C. Mellot-Draznieks, *CrystEngComm*, 2009, **11**, 2272–2276.

[27] R. L. Martin and M. Haranczyk, *Cryst. Growth Des.*, 2014, **14**, 2431–2440.

[28] M. Eddaoudi, J. Kim, N. Rosi, D. Vodak, J. Wachter, M. O'Keeffe and O. M. Yaghi, *Science*, 2002, **295**, 469–472.

[29] S. Amirjalayer and R. Schmid, *J. Phys. Chem. C*, 2008, **112**, 14980–14987.

[30] B. Moulton and M. J. Zaworotko, *Chem. Rev.*, 2001, **101**, 1629–1658.

[31] D. E. Goldberg, *Genetic Algorithms in Search, Optimization, and Machine Learning*, Addison-Wesley, New York, NJ, 1989.

[32] R. Schmid, S. Amirjalayer and S. Bureekaew, *Predicting Framework Structures by a Reverse Topological Approach: A Systematic Screening of Topological and Isoreticular Supramolecular Isomers*, manuscript in preparation.

[33] M. W. Deem and J. M. Newsam, *J. Am. Chem. Soc.*, 1992, **114**, 7189–7198.

[34] D. J. Earl and M. W. Deem, *Ind. Eng. Chem. Res.*, 2006, **45**, 5449–5454.

[35] N. J. Henson, A. K. Cheetham and J. D. Gale, *Chem. Mater.*, 1994, **6**, 1647–1650.

[36] M. W. Deem, Deem Database, http://www.hypotheticalzeolites.net/DATABASE/DEEM/index.php.

[37] A. Le Bail, Predicted Crystallography Open Database http://www.crystallography.net/pcod/.

[38] M. M. J. Foster, M. D.; Treacy, Atlas of Prospective Zeolite Structures, http://www.hypotheticalzeolites.net/.

[39] Y. Li and J. Yu, *Chem. Rev.*, 2014, **114**, 7268–7316.

[40] M. Kondo, Y. Takashima, J. Seo, S. Kitagawa and S. Furukawa, *CrystEngComm*, 2010, **12**, 2350–2353.

[41] S. R. Caskey, A. G. Wong-Foy and A. J. Matzger, *Inorg. Chem.*, 2008, **47**, 7751–7756.

[42] D. Sun, S. Ma, J. M. Simmons, J.-R. Li, D. Yuan and H.-C. Zhou, *Chem. Commun.*, 2010, **46**, 1329–1331.

[43] B. Chen, N. W. Ockwig, A. R. Millward, D. S. Contreras and O. M. Yaghi, *Angew. Chem. Int. Ed.*, 2005, **44**, 4745–4749.

[44] X. Lin, I. Telepeni, A. J. Blake, A. Dailly, C. M. Brown, J. M. Simmons, M. Zoppi, G. S. Walker, K. M. Thomas, T. J. Mays, P. Hubberstey, N. R. Champness and M. Schroeder, *J. Am. Chem. Soc.*, 2009, **131**, 2159–2171.

[45] J. A. Gee and D. S. Sholl, *The Journal of Physical Chemistry C*, 2013, **117**, 20636–20642.

[46] R. Schmid, *An object-oriented framework for the automated generation of polymeric 3D network structures*, 2009, Ruhr-Universität Bochum.

[47] C. J. Fennell and J. D. Gezelter, *J. Chem. Phys.*, 2006, **124**, 234104.

[48] A. Togo, *spglib*, 2010, `http://http://spglib.sourceforge.net/`, version 1.6.

[49] H. Frost, T. Düren and R. Q. Snurr, *J. Phys. Chem. B*, 2006, **110**, 9565–9570.

[50] T. Düren, F. Millange, G. Ferey, K. S. Walton and R. Q. Snurr, *J. Phys. Chem. C*, 2007, **111**, 15350–15356.

[51] T. Düren, L. Sarkisov and R. Snurr, Code for computing surface areas (version for nonorthorhombic cells), available from `http://www.see.ed.ac.uk/~tduren/research/surface_area/non_ortho/`.

Table of Content

By a screening of the possible space of isoreticular isomers the structure of copper paddle-wheel based metal-organic frameworks in **nbo-b** topology can be predicted and the formation of different phases can be rationalized.