PCCP

Accepted Manuscript

This is an *Accepted Manuscript*, which has been through the Royal Society of Chemistry peer review process and has been accepted for publication.

Accepted Manuscripts are published online shortly after acceptance, before technical editing, formatting and proof reading. Using this free service, authors can make their results available to the community, in citable form, before we publish the edited article. We will replace this *Accepted Manuscript* with the edited and formatted *Advance Article* as soon as it is available.

You can find more information about *Accepted Manuscripts* in the **Information for Authors**.

Please note that technical editing may introduce minor changes to the text and/or graphics, which may alter content. The journal's standard **Terms & Conditions** and the **Ethical guidelines** still apply. In no event shall the Royal Society of Chemistry be held responsible for any errors or omissions in this *Accepted Manuscript* or any consequences arising from the use of any information it contains.

www.rsc.org/pccp

Page 1 of 14

Physical Chemistry Chemical Physics
CREATED USING THE RSC LaTeX PCCP ARTICLE TEMPLATE – SEE www.rsc.org/electronicfiles FOR DETAILS

ARTICLE TYPE

www.rsc.org/xxxxxx | XXXXXXXX

In Silico Prediction of MOFs with High Deliverable Capacity or Internal Surface Area[†]

Yi Bao,[a] **Richard L. Martin,**[b,e] **Maciej Haranczyk,**[b] **and Michael W. Deem,**[*acd]

Metal-organic frameworks (MOFs) offer unprecedented atom-scale design and structural tunability, largely due to the vast number of possible organic linkers which can be utilized in their assembly. Exploration of this space of linkers allows identification of ranges of achievable material properties as well as discovery of optimal materials for a given application. Experimental exploration of the linker space has to date been quite limited due to the cost and complexity of synthesis, while high-throughput computational studies have mainly explored MOF materials based on known or readily available linkers. Here an evolutionary algorithm for *de novo* design of organic linkers for metal-organic frameworks is used to predict MOFs with either high methane deliverable capacity or methane accessible surface area. Known chemical reactions are applied *in silico* to a population of linkers to discover these MOFs. Through this design strategy, MOF candidates are found in the ten symmetric networks **acs**, **cds**, **dia**, **hxg**, **lvt**, **nbo**, **pcu**, **rhr**, **sod**, and **tbo**. The correlation between deliverable capacities and surface area is network dependent.

1 Introduction

Metal-organic frameworks (MOFs) are coordination networks comprised of secondary building units (SBUs) of metal/metal-organic ions or clusters and organic linkers which connect SBUs. The periodicity of a MOF in three dimensions arises from assembly of SBUs with organic linkers to form a crystalline topology, termed network or net in the field of reticular chemistry[1]. MOFs as a distinct class of porous frameworks have existed for less than 20 years[2,3]. Several thousand MOFs have been synthesized up to date[4], while a computational enumeration method has predicted over 100,000 hypothetical MOFs[5]. The possibility of tailoring building blocks and high porosities allow MOFs to be a vital group of tunable materials[6]. MOFs are considered to be highly promising materials for a wide variety of industrial applications, including gas storage and separation[7–10], gas purification[11–13], catalysis[14–16], CO_2 capture[17], sensing[18,19], and drug delivery[20].

The properties of a MOF depend intimately on the linkers from which the structure is created[4]. Thus, design of MOF materials for any application necessarily includes design of the linkers. New designs would allow discovery of MOFs with potentially higher performance. Experimental studies tend to be limited in the number of linkers explored, due to resource limitations, while high-throughput computational studies have typically restricted their search to either known MOF building blocks[5] or commercially available molecules exhibiting linker-like properties[21]. Future experimental advances in the MOF field may involve construction of complex chemical microenvironments in the MOF cages, resulting from the use of many different linker functionalities[4].

In this contribution we use an *in silico* evolutionary approach to design MOFs with tailored methane accessible surface areas from both predicted organic ligands and known metal-containing SBUs. The method applies known chemical transformations to a precursor library of commercially available organic synthetic precursors and produces transformed molecules. Thus, this procedure generates a population of predicted organic linkers by *in silico* evolution. We consider two different precursor libraries containing $\sim 50,000$ compounds each and compare their performance. We apply 84 chemical reactions that have been successfully applied in drug design[22]. Each of the predicted linkers is produced by these chemical reactions applied to the precursors, and we typically limit the number of steps in the synthetic routes to five or fewer. With such a massive number of precursors and reactions, the resulting chemical space of organic linkers, if all combinations of which are exhausted, is enormous and cannot be simply enumerated. Instead, we use a biologically-inspired algorithm[23] to automate the identification of linker designs which lead to MOFs with optimized certain properties, i.e. methane accessi-

[a] *Physics & Astronomy Department, Rice University, 6100 Main Street, MS-61, Houston, TX, 77005, USA.*
[b] *Computational Research Division, Lawrence Berkeley National Laboratory, One Cyclotron Road, MS 50F-1650, Berkeley, CA, 94720, USA.*
[c] *Department of Bioengineering, Rice University, 6100 Main Street, MS-142, Houston, TX, 77005, USA. Fax: (713) 348-5811; Tel: (713) 348-5852; E-mail: mwdeem@rice.edu*
[d] *Graduate Program in Systems, Synthetic, and Physical Biology, Rice University, 6500 Main Street, MS-180, Houston, TX, 77005, USA.*
[e] *Present address: Watson Group, IBM Almaden Research Center, San Jose, California 95120, USA.*

ble surface area or deliverable capacity in this work. This approach has been used to predict chemically synthesizable organic structure directing agents for zeolites[24], and the method has successfully predicted an organic structure directing agent for synthesis of the zeolite STW, a very challenging synthesis problem[25].

Self-assembly of the MOFs imposes constraints on the linkers in terms of flexibility and geometry. Linkers are typically fairly rigid molecules such that the built MOFs maintain lasting structures and stable property. Also the linkers must contain a proper number of binding sites to be consistent with SBUs under given geometry of the MOF. Finally, the linker itself should be either commercially available or synthesizable from commercially available precursors. Our approach ensures that the predicted linkers satisfy the topological properties required by the MOF network and are synthesizable in a limited number of synthetic steps from commercially available precursors.

In this work we consider the nine well-known MOF nets **acs**, **cds**, **dia**, **hxg**, **lvt**, **nbo**, **pcu**, **rhr**, and **sod**, which constitute some of the possible regular, semiregular, and minimal nets[26] based on combining selected SBUs with linear, two-connected linkers. We also consider the **tbo** net constructed from three-connected linkers. Three sets of results are obtained using the evolutionary algorithm: 1) we evolve a population of linkers to optimize a measure of methane accessible surface area for each net; 2) we evolve a population of linkers to optimize deliverable capacity at 65 bar loading pressure, 5.8 bar delivery pressure, and 298 K for each net; and 3) we evolve a population of linkers to optimize deliverable capacity at 35 bar loading pressure, 5.8 bar delivery pressure, and 298 K for each net. The **tbo** net is considered only in set 3. For set 2, results for the nine symmetric networks at 65–5.8 bar have been previously presented[23]. The results presented here for set 2 are novel and complement to those previously presented. We point out that in comparison to the previous work[23], the three sets of results presented here are generated using the same algorithm but in a larger scale, i.e. by optimizing deliverable capacities at either of the two pressure conditions or by optimizing surface area. Also, one additional precursor library (library A) is applied to be compared with library B probing the differences of the performance. Additionally, one more network, **tbo**, is considered in the optimization of deliverable capacity. For each set of results, we investigate the relationship between the measure of surface area and deliverable capacity. For example, even though set 2 is optimized for deliverable capacity at 65–5.8 bar, we also computed the deliverable capacity at 35–5.8 bar and the accessible surface area for each of the compounds identified. We give an overview of the methods in the Methods section. We present details of the designed MOFs in the Results section. We discuss features of the predicted molecules and their suitability as linkers in the

Discussion section. We summarize the results in the Conclusion section.

2 Methods

In these nets, each linker has two connecting sites to SBUs. The only exception is **tbo** which requires three connecting sites. We consider each MOF network with a single type of node and explore the space of possible linkers. Although it is possible for a MOF to contain more than one type of linker molecule, we here assume that all linker molecules in one MOF are identical.

We use a computational *de novo* algorithm to generate synthesizable linker molecules for MOFs[23]. In brief, the linkers are optimized for their methane accessible surface area, so that linkers providing high surface area in the MOF network are evolved over time in a population of 100 MOFs. Surface area is measured by the product of $GSA \cdot VSA$, where GSA (surface area per unit mass, m²/g) denotes gravimetric surface area, and VSA (surface area per unit volume, m²/cm³) denotes volumetric surface area. The method applies a set of 84 known chemical reactions[24] to an evolving population of predicted linkers. In each generation, one linker is randomly picked from the population and first minimized in terms of the local configurational energy. A subsequent exploration then finds an approximate global energy minimum with the ant algorithm[27]. After these first level geometry optimizations, a linker undergoes one of seven evolution operations. Each move has equal probability, and each operation performs a specific type of *in silico* chemical transformation on the linker, i.e. a specific modification of the synthetic route. The transformed linker is then evaluated by four filters that gauge the geometry and rigidity. For all networks except **tbo**, the linker molecule is first required to have exactly two carboxylic sites to connect with SBUs. The linker is required to have eight or fewer torsions. Then, a molecular dynamics simulation of 30 ps is carried out at 298 K on the linker to produce a set of 300 conformations. For all networks except **tbo**, a filter is applied to these conformations to require that the two carboxylic groups point in nearly opposite directions: for surface area optimization we require angles $\geq 155°$ and for deliverable capacity optimization we require angles $\geq 160°$. We finally require the pairwise distance between the carbons in the two opposing carboxylic groups to have a standard deviation of less than 0.5 Å in the conformations.

Since **tbo** net requires each linker to have three connecting sites to the SBU, the two filters that gauge the linker geometry are slightly different. The linker molecule is first required to have exactly three carboxylic sites to connect with SBUs. Then the distances d_i are calculated, where $i = 1, 2, 3$, between each pairwise carboxylic groups over the set of 300 linker conformations. We require $\sigma^m / \overline{d_i^m} < 2\%$ for each conformation

m, where σ^m is the standard deviation of the three distances of the linker for conformation m and $\overline{d_i^m}$ is the distance averaged over three pairwise carboxylic sites for conformation m. This filter enforces the symmetry assumption of the **tbo** network.

If the child linker successfully passes all filters, the algorithm builds the framework with it under the chosen net, and calculates the surface area by averaging a few MOF conformations. The child linker is inserted in rank order in the population, and the worst one in the population is discarded. The population is initialized by *add* or *multiple add* operations applied to compounds drawn from the precursor library. By construction, each linker in the population is fully described by a chain of synthetic reactions and corresponding reactants, and it is thus chemically synthesizable. Two precursor libraries are compared for surface area results, both containing next day available compounds from Sigma-Aldrich. Precursor library A, which contains additional compounds with exactly one carboxylic group and one benzyl group, has 42,284 molecules in total. These compounds can react to form dicarboxylic molecules. Conversely, precursor library B, which contains additional compounds from Sigma-Aldrich that are used in medicinal chemistry or as organic building blocks, has 57,815 molecules in total. A run with a given precursor library on a specific net performs 40,000 generations of evolutionary steps on proposed linkers and produces a final population of 100 evolved linkers.

To calculate surface area, a MOF is constructed from the proposed linker and the metal SBU in the chosen network with the software suite Zeo++[28]. MOFs are crystalline materials, and thus we construct the unit cell of the MOFs from the metal SBU and organic linker as input based on the net information[29]. The metal SBUs are selected according to simplicity and charge balancing[30]. The SBUs chosen for each of the nine networks are shown in Table 1 and Fig. 1. The MOF is constructed by positioning the SBUs to distinct vertices of the net. After aligning the set of SBUs, the linkers are added to the framework, and the unit cell dimensions are scaled to accommodate the linkers.

Net	SBU	Description
acs	Pd_3	trigonal prism
cds, lvt, nbo, rhr, tbo	Cu_2	square "paddlewheel"
dia, sod	$V_4(OH)_4$	tetrahedral
hxg	Ti_6O_6	hexagonal
pcu	Zn_4O	octahedral

Table 1 SBUs for the nine well-known MOF networks. One unique metal SBU is chosen for each network. SBUs are selected to be charge neutral.

When optimizing deliverable capacity, rather than surface area, we only use precursor library B. The loading pressure, the
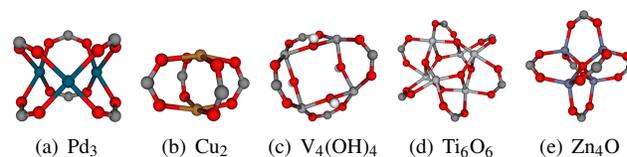

(a) Pd_3 (b) Cu_2 (c) $V_4(OH)_4$ (d) Ti_6O_6 (e) Zn_4O

Fig. 1 Spatial configurations of secondary building units (SBUs) with O-C-O connecting sites. Red: oxygen atoms; dark grey: carbon atoms; white: hydrogen atoms; other colors: respective metal atoms.

delivery pressure, and operating temperature are fixed. The minimum supercell to accommodate an interaction radius cutoff 12.5 Å is built from the SBU, linker, and network using the software suite Zeo++[28]. Then we simulate methane gas adsorption with grand canonical Monte Carlo for 300 steps. We calculate deliverable capacities every 50 MOF conformations over a set of 300 conformations at 298 K and two pressures, either 65 bar and 5.8 bar, or 35 bar and 5.8 bar. The universal force field with a tail correction is used to calculate interactions. Deliverable capacity, a measure of net expendable volume of gas per volume of adsorbent, is finally given in the unit of v(STP)/v.

3 Results

3.1 Methane Accessible Surface Area Optimized

MOFs were evolved to optimize the methane accessible surface area. This set of results contains 1793 MOFs. In this set of results, linkers for each of the ten MOF nets were designed using two different precursor libraries. The best 100 MOFs are retained for each net each precursor library. We compare the computed surface areas with a set of previous results which applies gradient-based optimization in an abstract shape-space of MOF linkers[31,32]. Results are shown in Table 2. In most cases, the present evolution approach identifies linkers which produce greater $GSA \cdot VSA$ compared with values from previously identified optimal linker shapes; furthermore, the linkers in this work are chemically synthesizable following the suggested routes. Particularly favorable results are found for the nets of **acs**, **dia**, **nbo**, and **lvt**. The superior results achieved in this work indicate exploration of linker shapes that fall outside the bounds in the previous optimization work[31]. Previous work imposed symmetry constraints on the linkers, while in this work the best results are typically achieved with asymmetrical molecules. The **sod** net with precursor library B gives the best instance of this, e.g. as shown in Fig. 6(c) and Fig. 7(i), where the very large $GSA \cdot VSA$ is achieved using a short linker with a long side group.

In Fig. 2(a), the GSA and VSA values for the final population of linkers in both precursor libraries and for each network are shown. The Pareto optimal front for each network is shown
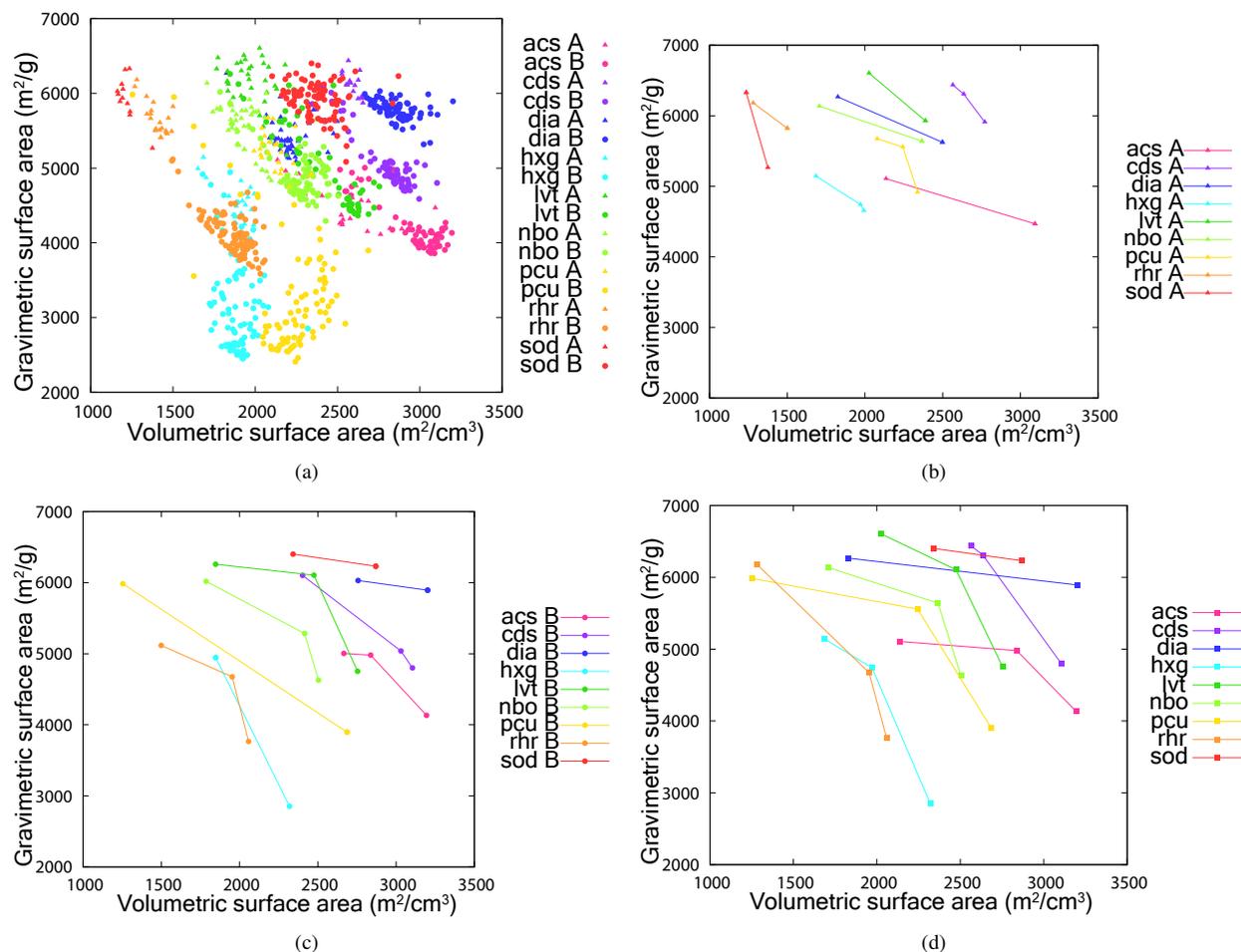
Fig. 2 Distributions of linker molecules in *GSA–VSA* space. (a) *GSA* and *VSA* of evolved linkers for final populations in each of the nine MOF nets and two precursor libraries. (b) The Pareto optimal fronts for each MOF net of precursor library A. (c) The Pareto optimal fronts for each MOF net of precursor library B. (d) The Pareto optimal fronts for each MOF net of both libraries. In all cases, only molecules with pairwise angle $\geq 155°$ are shown.

Net	Precursor Library A			Precursor Library B			$GSA \cdot VSA$ from previous optimization[31]
	GSA	VSA	$GSA \cdot VSA$	GSA	VSA	$GSA \cdot VSA$	
acs	4,468	3,094	13,826,100	4,980	2,838	14,132,400	10,440,000
cds	6,307	2,636	16,624,500	5,039	3,031	15,274,800	17,630,000
dia	5,625	2,498	14,053,700	5,894	3,202	18,870,200	10,065,000
hxg	4,740	1,971	9,342,410	4,945	1,848	9,137,620	10,560,000
lvt	5,929	2,389	14,165,000	6,104	2,475	15,107,800	5,795,000
nbo	5,641	2,366	13,347,100	5,285	2,416	12,769,400	10,875,000
pcu	5,558	2,244	12,471,900	3,897	2,687	10,469,300	11,500,000
rhr	5,823	1,501	8,737,070	4,675	1,953	9,131,470	8,030,000
sod	6,329	1,236	7,820,700	6,231	2,870	17,885,400	7,930,000
tbo				4,099	2,381	8,794,350	

Table 2 Evolved best values of GSA, VSA, and $GSA \cdot VSA$ from two precursor libraries of commercially available compounds. Linkers were selected to optimize $GSA \cdot VSA$. Only results from linkers with pairwise angle $\geq 155°$ are shown. Previous results were calculated via gradient-based optimization in an abstract chemical shape space[31]. Precursor libraries A and B provide consistent results for **acs**, **cds**, **hxg**, **lvt**, **nbo**, and **rhr**. Also shown is the **tbo** net results from deliverable capacity at the 35–5.8 bar condition set using library B. GSA: m²/g; VSA: m²/(cm³); $GSA \cdot VSA$: m⁴/(g cm³).

in Figs. 2(b)–2(d).

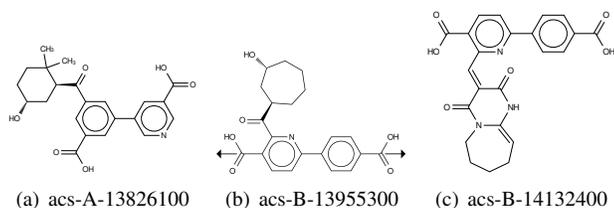Some chemically promising linkers identified from both libraries are shown in Figs. 3–6.

(a) acs-A-13826100   (b) acs-B-13955300   (c) acs-B-14132400

Fig. 3 Representative molecules of **acs** net. Each molecule is named as net-library-$GSA \cdot VSA$. Each black arrow in (b) is a vector orienting from the carbon atom and passing through the midpoint of two oxygen atoms for the carboxylic group metal attachment site. The pairwise angle of the linker molecule is defined as that between these two vectors. We require the pairwise angle $\geq 155°$ or $\geq 160°$.

(a) dia-A-14053700   (b) dia-B-18266400   (c) dia-B-18870200
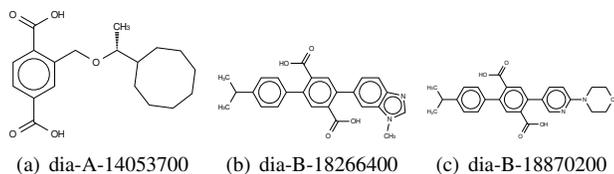
Fig. 4 Representative molecules of **dia** net.

Fig. 7 shows ten sample MOFs with predicted linkers for each net. Fig. 8 shows an example of synthetic route of molecule nbo-A-13347100.

The deliverable capacities at the 65–5.8 bar and 35–5.8 bar conditions for the results using precursor library B are shown in Fig. 9(a)–(d).
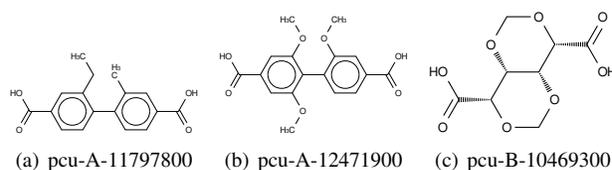

(a) pcu-A-11797800   (b) pcu-A-12471900   (c) pcu-B-10469300

Fig. 5 Representative molecules of **pcu** net.

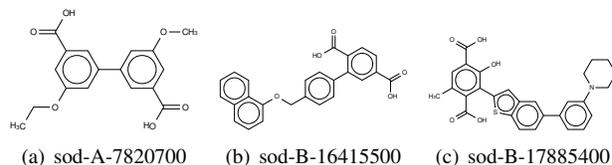(a) sod-A-7820700   (b) sod-B-16415500   (c) sod-B-17885400

Fig. 6 Representative molecules of **sod** net.

3.2 Deliverable Capacity at 65–5.8 bar Condition Optimized

MOFs were evolved to optimize the methane deliverable capacity at 65–5.8 bar and 298 K. This set of results contains 3035 MOFs, and only precursor library B is used. The deliverable capacity at the 35–5.8 bar condition and $GSA \cdot VSA$ are also calculated for each MOF, as summarized in Table. 3 and plotted in Fig. 10.

3.3 Deliverable Capacity at 35–5.8 bar Condition Optimized

MOFs were evolved to optimize the methane deliverable capacity at 35–5.8 bar and 298 K. This set of results contains 4141 MOFs (3820 MOFs for the nine nets and 321 MOFs for **tbo** net), and only precursor library B is used. The deliver-
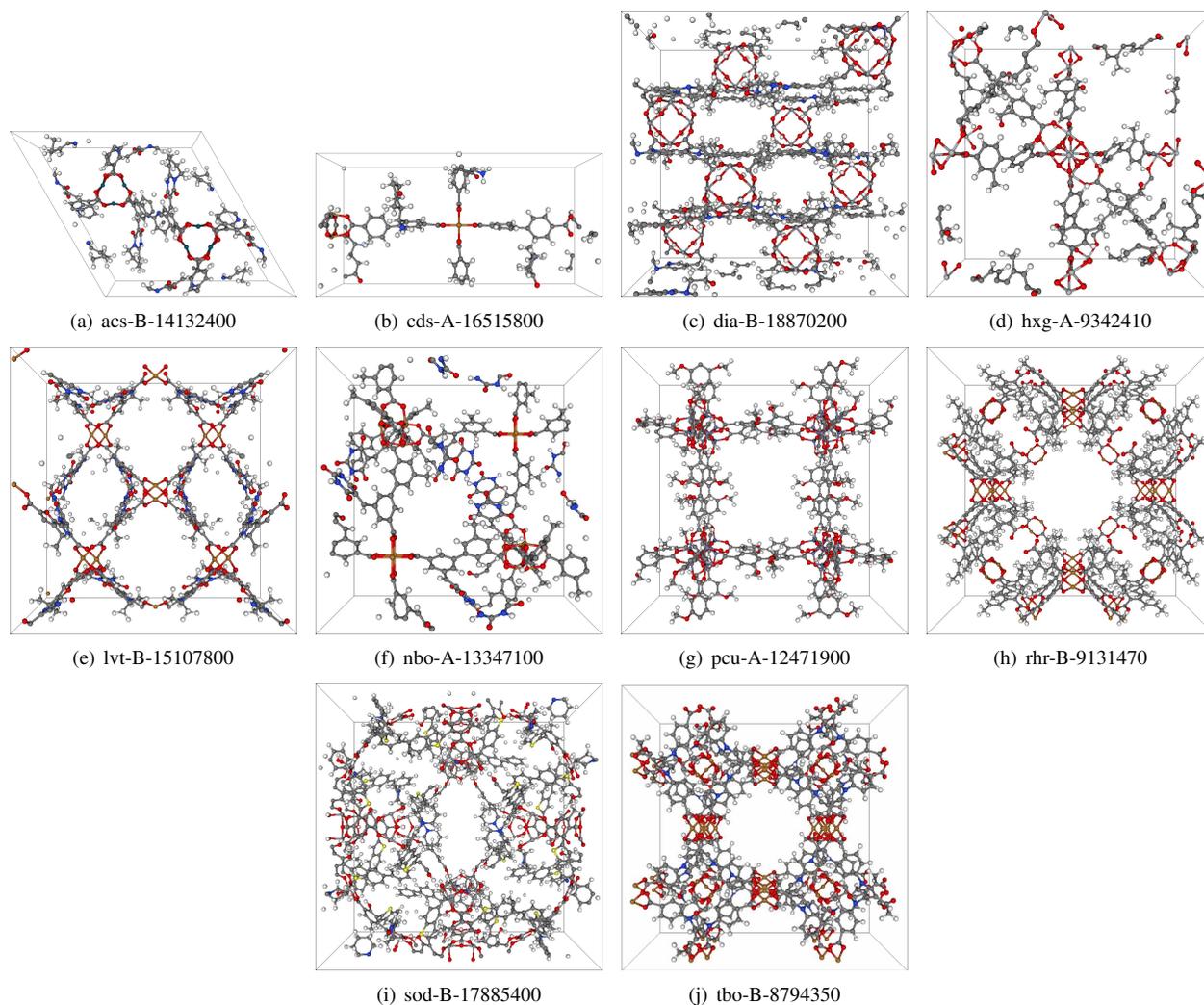
(a) acs-B-14132400   (b) cds-A-16515800   (c) dia-B-18870200   (d) hxg-A-9342410

(e) lvt-B-15107800   (f) nbo-A-13347100   (g) pcu-A-12471900   (h) rhr-B-9131470

(i) sod-B-17885400   (j) tbo-B-8794350

Fig. 7 MOF conformations in each of ten nets. Each conformation shows the geometry of the MOF for a given linker and SBUs, as noted in Table 1 and Fig. 1. The tbo-B-8794350 structure is from optimization of deliverable capacity at the 35–5.8 bar condition. The other nine conformations are from optimization of $GSA \cdot VSA$.

able capacity at the 65–5.8 bar condition and $GSA \cdot VSA$ are also calculated for each MOF, as summarized in Table. 4 and plotted in Fig. 11.

At the 35–5.8 bar condition, HKUST-1, a **tbo** net, has been a well-tested MOF with a deliverable capacity of 141 v(STP)/v experimentally[33]. To compare our results with HKUST-1, we perform the algorithm on the **tbo** net in addition to the nine nets at the 35–5.8 bar condition. Taking the requirement of three-connecting linkers into account, we customize the filters for **tbo** net as discussed in Methods section. The distribution of deliverable capacities of MOFs in ten nets is shown in Fig. 12. We find 319 MOFs in the nets of **cds**, **lvt**, and **pcu**, and 2 MOFs in the net of **tbo** with a higher deliverable capac-

ity than HKUST-1.

4 Discussion

Achievable surface areas for nine of these MOF networks have previously been reported[31]. In that study, shapes of potential MOF linkers in an abstract chemical space were optimized for surface area. The present results are exciting because they validate the high surface areas that were previously predicted using shape only, and are now exemplified with molecular structure. Interestingly, the evolution procedure developed here to identify chemically-synthesizable linkers leads to surface area values, $GSA \cdot VSA$, superior than those previously reported for
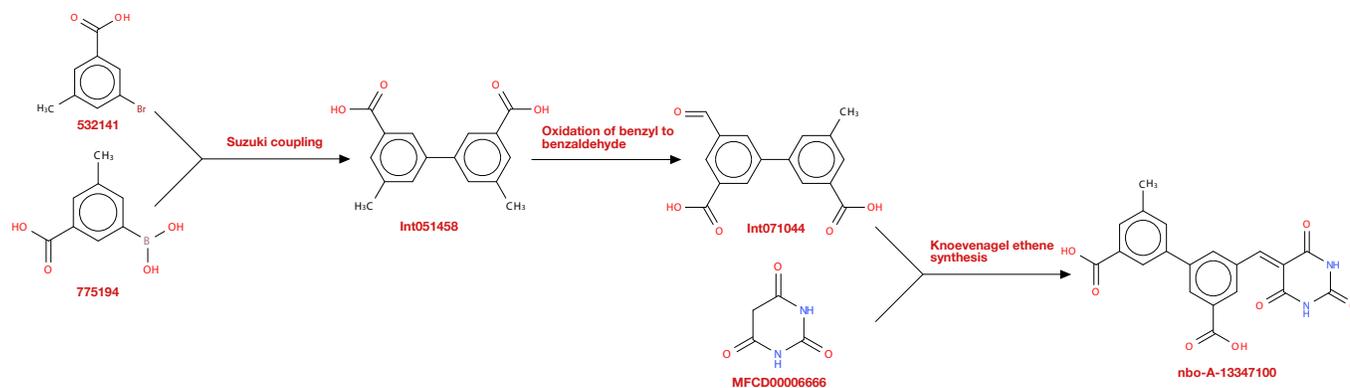
Fig. 8 Synthetic route of molecule nbo-A-13347100 in **nbo** network produced from precursor library A. Compounds 532141 and 775194 from the precursor library are Suzuki coupled to produce the intermediate product Int051458. The benzyl of Int051458 is oxidized into benzaldehyde, and Int071044 is formed. Int071044 and compound MFCD00006666 from the precursor library undergo Knoevenagel ethene synthesis and produce the final product nbo-A-13347100.

Net	Number of Candidates	Highest D.C. at the 65 bar condition	D.C. at the 35 bar condition	$GSA \cdot VSA$	Pearson Correlation
acs	315	183.62	128.45	8,601,250	0.826
cds	409	197.87	143.62	12,090,800	0.925
dia	150	177.32	144.77	13,471,500	0.209
hxg	516	171.99	115.67	2,840,570	0.457
lvt	935	189.53	127.95	296,427	0.318
nbo	374	182.08	121.28	6,728,520	0.580
pcu	201	186.22	132.66	1,964,840	0.748
rhr	42	158.75	99.60	3,739,120	-0.486
sod	93	153.93	97.51	3,230,480	0.301

Table 3 Summary of MOFs with top deliverable capacity at the 65 bar condition for each net. Only results from linkers with pairwise angle $\geq 160°$ are shown. The best predicted MOF at the 65–5.8 bar condition is a **cds** net with a deliverable capacity of 197.87 v(STP)/v. Deliverable capacity at the 35–5.8 bar condition and $GSA \cdot VSA$ are shown for the best optimized MOF in each net. The Pearson correlation coefficient is calculated between highest deliverable capacity at the 65 bar condition and $GSA \cdot VSA$ (column 3 and column 5) for each net, and heavily depends on the net. The average Pearson correlation coefficient for all nine nets is 0.481. High deliverable capacity at the 65 bar condition does not necessarily indicate high $GSA \cdot VSA$. Deliverable capacity: v(STP)/v; $GSA \cdot VSA$: m^4/(g cm^3)

Net	Number of Candidates	Highest D.C. at the 35 bar condition	D.C. the at 65 bar condition	$GSA \cdot VSA$	Pearson Correlation
acs	259	123.72	157.57	11,574,200	0.872
cds	926	148.09	180.12	2,844,200	0.778
dia	45	118.34	172.82	8,271,250	0.536
hxg	84	121.03	160.46	3,228,680	0.678
lvt	848	134.11	177.76	396,580	-0.220
nbo	1,223	124.89	167.40	3,351,830	0.228
pcu	122	134.37	168.52	417,578	0.406
rhr	280	103.62	149.76	2,247,810	-0.357
sod	33	90.51	145.08	9,031,330	-0.049
tbo	321	127.62	170.02	5,288,910	0.860

Table 4 Summary of MOFs with top deliverable capacity at the 35 bar condition for each net including **tbo** net. Only results from linkers with pairwise angle $\geq 160°$ are shown. The best predicted MOF at the 35–5.8 bar condition is a **cds** net with a deliverable capacity of 148.09 v(STP)/v. Deliverable capacity at the 65–5.8 bar condition and $GSA \cdot VSA$ are shown for the best optimized MOF in each net. The Pearson correlation coefficient is calculated between highest deliverable capacity at the 35 bar condition and $GSA \cdot VSA$ (column 3 and column 5) for each net, and heavily depends on the net. The average Pearson correlation coefficient for all nine nets is 0.532. High deliverable capacity at the 35 bar condition does not necessarily indicate high $GSA \cdot VSA$. Deliverable capacity: v(STP)/v; $GSA \cdot VSA$: m^4/(g cm^3)
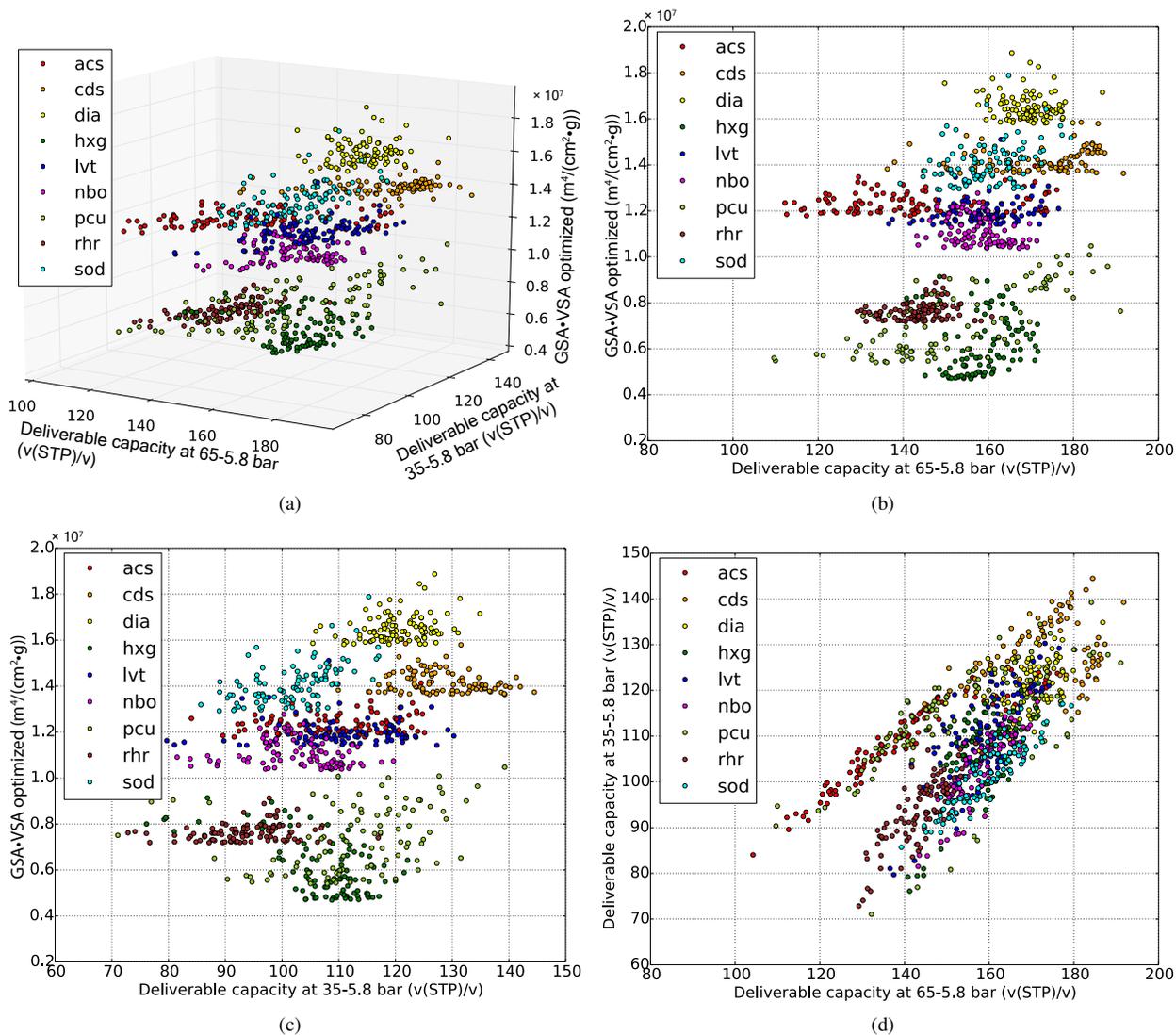
Fig. 9 Relationship between optimized surface area and deliverable capacities using precursor library B. (a) The points represent the surface area optimized results in the space of $GSA \cdot VSA$ and deliverable capacities. (b,c) MOFs with a certain $GSA \cdot VSA$ can have a range of deliverable capacities, although high $GSA \cdot VSA$ tend to indicate high deliverable capacities on average. (d) Deliverable capacities at two pressure conditions have a positive correlation.
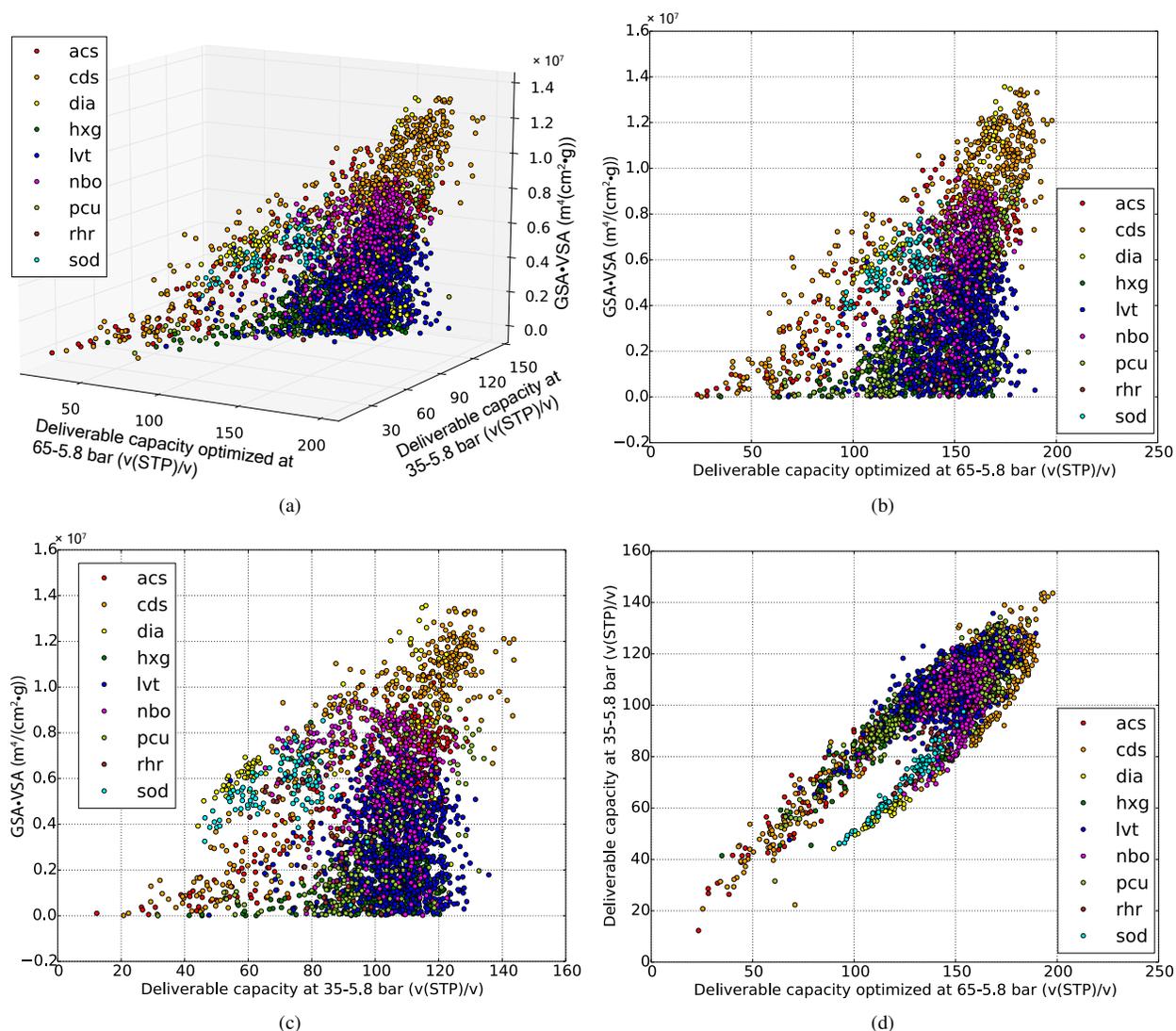
Fig. 10 Relationship between optimized deliverable capacity at the 65 bar condition, deliverable capacity at the 35 bar condition and surface area. (a) **cds** tends to occupy the highest end in the 3-dimensional phase space, but has a broad distribution. A large number of MOFs in **nbo** and **lvt** are located in a similar concentrated region, while other nets, for example **acs** and **sod**, exhibit a long-band pattern. (b, c) For a certain deliverable capacity, the distribution of $GSA \cdot VSA$ strongly depends on network. For instance, high deliverable capacities tend to indicate high $GSA \cdot VSA$ in **cds**; while $GSA \cdot VSA$ is almost independent from deliverable capacities in **nbo**. (d) Deliverable capacities at the two pressure conditions have a correlation.
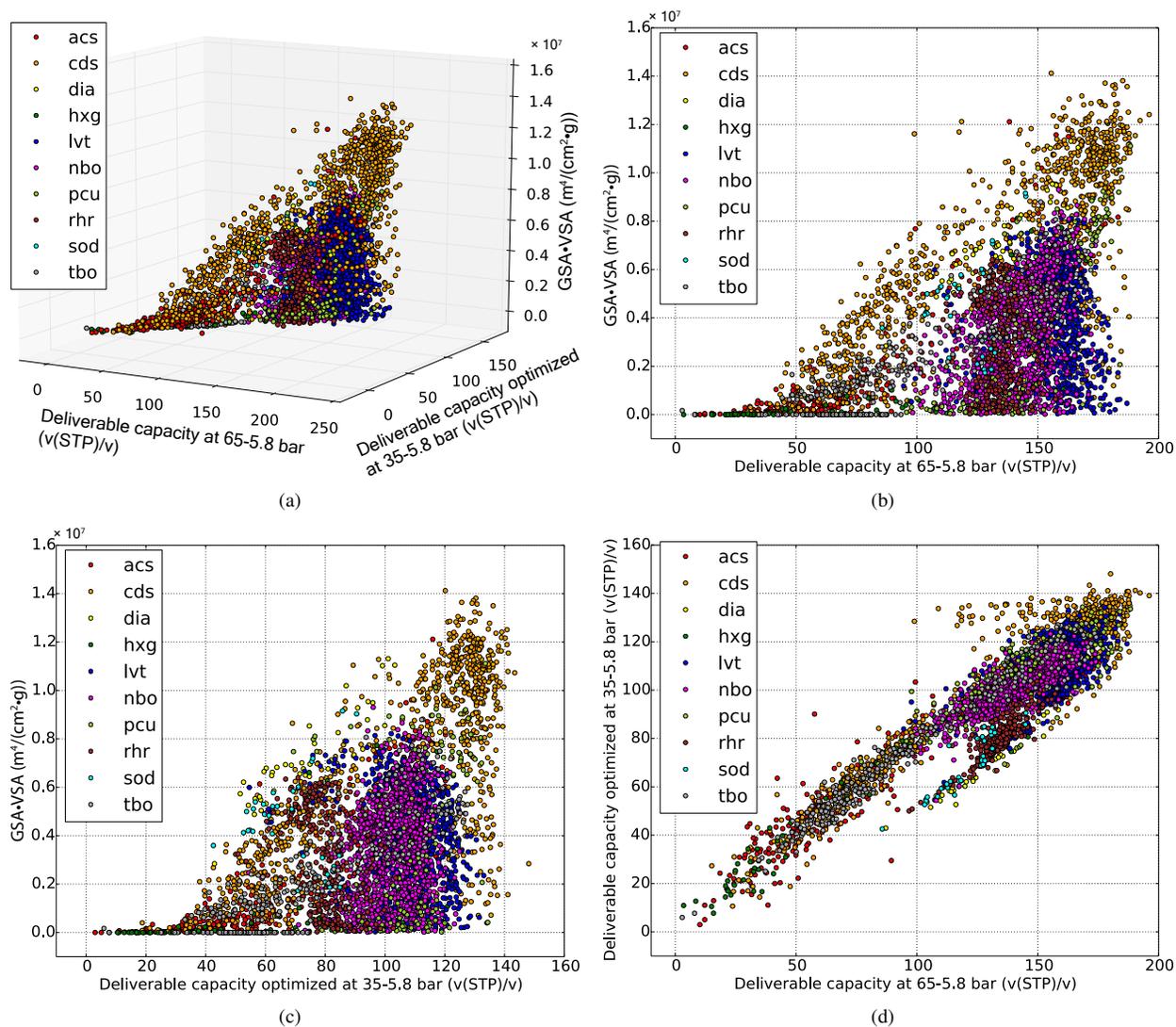
Fig. 11 Relationship between optimized deliverable capacity at the 35 bar condition, deliverable capacity at the 65 bar condition, and surface area. (a, b, c, d) The distributions show similar patterns to the results of optimized deliverable capacity at the 65–5.8 bar condition.
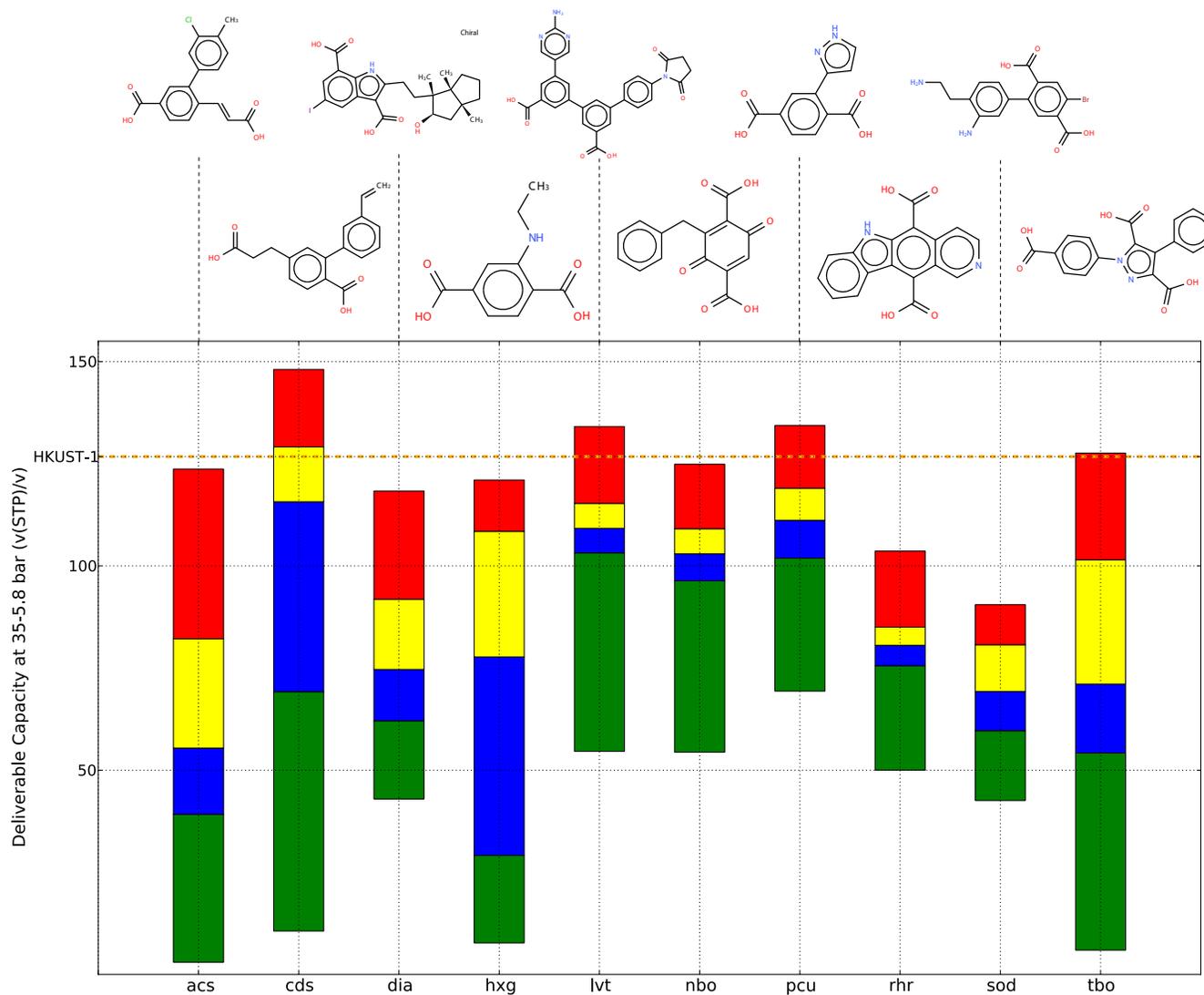
Fig. 12 Distribution of deliverable capacity at the 35–5.8 bar condition of all MOFs identified in the ten nets. Only results from linkers with pairwise angle $\geq 160°$ are shown. HKUST-1 is found in the results of **tbo** net with a computed deliverable capacity of 126.77 v(STP)/v. The best organic linker of each net is shown above each bar. Each color of the bars represents one quartile of the results.

optimized linker shapes for the networks **acs**, **dia**, **lvt**, **nbo**, **pcu**, **rhr** and **sod**, as shown in Table 2.

The Pareto fronts of different nets in Fig. 2(d) shows that **lvt** net dominates **nbo** net, i.e. the maximal GSA, VSA, and $GSA \cdot VSA$ of **lvt** are bigger than those of **nbo**, while **pcu** net dominates **hxg** net. Furthermore, **dia** dominates many nets, as does **sod**. Fig. 2(a) provides even greater insight, particularly into the comparative behaviors of the precursor libraries. Clearly, the **sod** libraries have radically different performance. A less drastic differentiation may also be drawn for the other nets. For instance, the MOFs produced from the different libraries occupy different regions of the property space. MOFs in the **pcu** net produced from precursor library A are consistently in one central region, while those produced from precursor library B are just outside this region and mainly extend in the decreasing GSA direction as well as explore a few outlying positions.

Since the procedure is stochastic, different runs on the same net and with the same precursor library can lead to different final populations of linkers and different optimized values. Nevertheless, the tendency to improve optimizing values as the generation increases appears general[23]. A strict set of selection criteria tends to lead to small populations of elite compounds, while a less strict set of selection criteria leads to a bigger and more diverse population of compounds. For this reason, in the set of surface area optimized results, some of the runs use a looser angle filter to facilitate exploration of the linker chemical space. For example, an angle filter of $\geq 125°$ is used for the networks **dia**, **sod**, and **lvt** with precursor library A. This filter may not be necessary if a strict filter could lead to many good candidates. A tighter criterion should be applied if too many mediocre candidates are generated. This is the case with all the networks when using precursor library B, for which we use an angle filter of $\geq 155°$. It is worth noting that mean pairwise angle of linkers cannot be strictly $180°$ because molecules' conformations fluctuate at finite temperature. For instance, the organic linker BPDC in MOF-106, a known "linear" linker, is observed in a bent conformation, which leads to corrugated sheet structures[34].

Several of the identified linkers have long side groups that could be important to increase the surface area. Generally, high GSA linkers tend to be long or branched, so as to maximize surface area with as much light-weight organic material as possible. Conversely, high VSA is usually the result of utilizing short linkers, so as to minimize the volume of material and increase the surface area by growing linker width. The product of GSA and VSA represents a compromise between these two competing effects[31] and has been used in study of hydrogen storage potential[35]. Intuitively, we can anticipate that the highest $GSA \cdot VSA$ will be the result of short organic linkers, to minimize material volume, with significant branching, to maximize the contribution of the organic to the surface area.

We constrain the total number of reaction steps for any of our predicted linkers to ten. Nevertheless, it is rare to identify predicted linkers that are synthesized with five or more than five steps in our results. Fig. 13 shows the statistics of different numbers of reaction steps in the final population in the **dia** and **pcu** nets with precursor library A for the set of surface area optimized results. The **dia** net significantly favors more reaction steps than does the **pcu** net. We ascribe this distinction to two reasons: 1) the nature of a network determines how difficult it is to grow a complicated molecule within the network; and 2) values of $GSA \cdot VSA$ may set an upper limit in which molecules with only a few reaction steps may already be sufficiently outstanding, as is probably the case for the **pcu** net. Predicted molecules with too many synthesis steps will not be attractive as practical candidates because a long synthetic route leads to low yield and high cost of synthesis[22]. We note that mean $GSA \cdot VSA$ for the linkers with different number of reaction steps are approximately equal, as shown in Fig. 13. This result suggests that high $GSA \cdot VSA$ can be achieved with relatively few reaction steps.
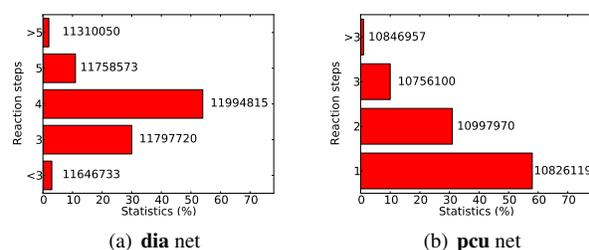

Fig. 13 Statistics of observing linkers with different numbers of reaction steps in the final populations of MOFs optimized for surface area, for two nets with precursor library A. The numbers to the right side of the bars are mean $GSA \cdot VSA$ for each number of reaction steps.

We used two precursor libraries of commercially available compounds in this study. This was done not only to provide statistics but also to judge which precursor library might lead to better performance. Overall, both precursor libraries A and B lead to similar results, except for the **dia** and **sod** network, in which precursor library B achieves higher values, as shown in Table 2. Precursor library A contains specific molecules with one carboxylic group and one benzyl group, while precursor library B is a balanced library, containing medicinal chemistry compounds and organic building blocks from Sigma-Aldrich. Library B appears to be a fairly universal one for use under general conditions. Therefore, we used precursor library B for the sets of deliverable capacity optimized results. It is noteworthy that besides library difference, **dia** and **sod** used an angle filter of $\geq 125°$ in precursor library A and $\geq 155°$ in

precursor library B. Using an angle filter of 155° in precursor library A leads to $GSA \cdot VSA$ values of 13,270,500 m⁴/(g cm³) for **dia** and 13,650,400 m⁴/(g cm³) for **sod**, which suggests that the effect of the angle filter on the evolution of the population of linkers is network dependent.

Measures of surface area, such as BET surface area, are broadly discussed in the literature, as it is generally believed that high surface area should be positively correlated with good performance of materials in gas-related applications[36–38]. Nevertheless, using $GSA \cdot VSA$ and deliverable capacities as metrics for surface area and gas adsorption performance respectively, we find that the correlation between them in the optimized materials presented here strongly depends on the network. For example, **rhr** in Table 3, and **lvt**, **rhr** and **sod** in Table 4 show a negative correlation of surface area and deliverable capacity. Thus, can must be exercised when surface area of MOFs is used to forecast potential adsorption capacity. The deliverable capacities at the 65–5.8 bar and 35–5.8 bar conditions are strongly positively correlated as shown in Fig. 9(d), Fig. 10(d) and Fig. 11(d), because these two quantities are derived from the same isothermal curve at 298 K. A similar relation between these two deliverable capacities was recently observed for covalent organic frameworks (COFs)[39], and porous polymer networks (PPNs)[40].

One intriguing aspect is to see whether experimentally synthesized MOFs, which have outstanding deliverable capacity, have high surface area. MOF-5, a **pcu** net, has been one benchmark MOF with a reported high methane deliverable capacity[33,41,42]. It is found in our results with a $GSA \cdot VSA$ of 10,007,100 m⁴/(g cm³), which is not among the highest. For instance, within the set of $GSA \cdot VSA$ optimized results, 103 MOFs of the **pcu** net have superior $GSA \cdot VSA$ values, and three are shown in Fig. 5(a)–(c) with the highest $GSA \cdot VSA$ of 12,471,900 m⁴/(g cm³). In total, 1,206 results have superior $GSA \cdot VSA$ values than MOF-5 in all nine nets. HKUST-1 is a **tbo** net with a high deliverable capacity, and it is found in the set of predicted results when deliverable capacity is optimized at the 35–5.8 bar condition. This MOF has a $GSA \cdot VSA$ of 4,914,780 m⁴/(g cm³), compared to which 56 MOFs of the **tbo** net have higher values.

We notice that the highest deliverable capacity at 65–5.8 bar condition obtained in this work is 197.87 v(STP)/v, which agrees with a recent proposed upper limit of approximately 200 v(STP)/v[43]. Some structures from the combinatorially created database approach this limit[5], and structures produced here approach that limit as well. The distribution of deliverable capacities of the structures produced with the present method is centered closer to this limit than are those of the combinatorial database[43]. In other words, optimization of deliverable capacity by the present method produces a larger set of MOFs with high deliverable capacity than does screening of a combinatorially created database.

Synthesizing these predicted MOFs requires consideration of several additional practical matters. For example the solubility of a proposed linker in a specific solvent and at a synthesis temperature is an important variable. Solubility is easily estimated, and it may be included in the scoring function of the evolutionary algorithm presented here. Alternatively, since multiple solvents and co-solvents are available for MOF synthesis, solvent may be chosen to best accommodate a predicted linker and SBU self-assembly.

5 Conclusion

Many chemically synthesizable linkers, which have considerable rigidity, suitable metal node binding sites, and promising $GSA \cdot VSA$, were identified for nine MOF networks. The **dia** network appears to be an excellent choice for compromising $GSA \cdot VSA$ surface area, with **cds** and **sod** also being good choices. The networks **lvt**, **cds**, and **sod** stand out for GSA alone, while **dia**, **cds**, and **acs** stand out for VSA alone.

Two sets of MOFs with optimized deliverable capacities at the 65–5.8 bar condition and the 35–5.8 bar condition were also identified. MOFs with high deliverable capacities are not necessary to have high $GSA \cdot VSA$. Conversely, MOFs with high $GSA \cdot VSA$ tend to have high deliverable capacities on average, but exceptions occur for specific networks, and negative correlations between $GSA \cdot VSA$ and deliverable capacities are observed.

In summary, we have used an evolutionary algorithm to efficiently explore the space of organic linkers to design MOFs that have optimized methane deliverable capacity at 65–5.8 or 35–5.8 bar or optimized methane accessible surface area. This algorithm is likely to aid experimental MOF design efforts. The procedure searches the space of chemically-synthesizable MOF linkers by the application of known chemical transformations to a population of evolving molecules. Additionally, it can be extended to consider MOFs with more than one type of SBU or more than one type of linker per structure.

Acknowledgments

Supplemental Information

The sets of predicted MOFs with high predicted surface area, deliverable capacity at 65–5.8 bar, and deliverable capacity at 35–5.8 bar are available as supplemental materials.

References

1 O. M. Yaghi, M. OKeeffe, N. W. Ockwig, H. K. Chae, M. Eddaoudi and J. Kim, *Nature*, 2003, **423**, 705–714.

2 O. M. Yaghi, H. Li, C. Davis, D. Richardson and T. L. Groy, *Acc. Chem. Res.*, 1998, **31**, 474–484.

3 M. Eddaoudi, D. B. Moler, H. Li, B. Chen, T. M. Reineke, M. O'Keeffe and O. M. Yaghi, *Accounts of Chemical Research*, 2001, **34(4)**, 319–330.

4 H. Furukawa, K. E. Cordova, M. O'Keeffe and O. M. Yaghi, *Science*, 2013, **341**, 974–986.

5 C. E. Wilmer, M. Leaf, C. Y. Lee, O. K. Farha, B. G. Hauser, J. T. Hupp and R. Q. Snurr, *Nature Chemistry*, 2012, **4**, 83–89.

6 J.-R. Li, R. J. Kuppler and H.-C. Zhou, *Chemical Society Reviews*, 2009, **38**, 1477–1504.

7 U. Mueller, M. Schubert, F. Teich, H. Puetter, K. Schierle-Arndta and J. Pastréa, *Journal of Materials Chemistry*, 2006, **16**, 626–636.

8 A. U. Czaja, N. Trukhan and U. M´uller, *Chemical Society Reviews*, 2009, **38**, 1284–1293.

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

10 R. B. Getman, Y.-S. Bae, C. E. Wilmer and R. Q. Snur, *Chem. Rev.*, 2012, **112**, 703–723.

11 J.-R. Li, J. Sculley and H.-C. Zhou, *Chem. Rev.*, 2012, **112**, 869–932.

12 Z. R. Herm, B. M. Wiers, J. A. Mason, J. M. van Baten, M. R. Hudson, P. Zajdel, C. M. Brown, N. Masciocchi, R. Krishna and J. R. Long, *Science*, 2013, **340**, 960–964.

13 K. A. Cychosz, R. Ahmadab and A. J. Matzger, *Chem. Sci.*, 2010, **1**, 293–302.

14 M. H. Alkordi, Y. Liu, R. W. Larsen, J. F. Eubank and M. Eddaoudi, *Journal of the American Society*, 2008, **130(38)**, 12639–12641.

15 J.-Y. Lee, O. K. Farha, J. Roberts, K. A. Scheidt, S.-B. T. Nguyen and J. T. Hupp, *Chem. Soc. Rev.*, 2009, **38**, 1450–1459.

16 H.-L. Jiang and Q. Xu, *Chemical Communications*, 2011, **47**, 3351–3370.

17 K. Sumida, D. L. Rogow, J. A. Mason, T. M. McDonald, E. D. Bloch, Z. R. Herm, T.-H. Bae and J. R. Long, *Chem. Rev.*, 2012, **112**, 724–781.

18 L. Sarkisov, *J. Chem. Phys. C*, 2012, **116**, 3025–3033.

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

20 P. Horcajada, R. Gref, T. Baati, P. K. Allan, G. Maurin, P. Couvreur, G. Férey, R. E. Morris and C. Serre, *Chem. Rev.*, 2012, **112**, 1232–1268.

21 R. L. Martin, L.-C. Lin, K. Jariwala, B. Smit and M. Haranczyk, *J. Phys. Chem. C.*, 2013, **117**, 12159–12167.

22 H. M. Vinkers, M. R. de Jonge, F. F. D. Daeyaert, J. Heeres, L. M. H. Koymans, J. H. van Lenthe, P. J. Lewi, H. Timmerman, K. V. Aken and P. A. J. Janssen, *J. Med. Chem.*, 2003, **46(13)**, 2765–2733.

23 Y. Bao, R. L. Martin, C. M. Simon, M. Haranczyk, B. Smit and M. W. Deem, *Journal of Physical Chemistry C*, 2015, **119(1)**, 186–195.

24 R. Pophale, F. Daeyaert and M. W. Deem, *J. Mater. Chem. A*, 2013, **1**, 6750–6760.

25 J. E. Schmidt, M. W. Deem and M. E. Davis, *Angew. Chem. Int. Ed.*, 2014, **53**, 8372–8374.

26 O. Delgado-Friedrichs, M. O'Keeffe and O. M. Yaghi, *Phys. Chem. Chem. Phys.*, 2007, **9**, 1035–1043.

27 F. Daeyaert, M. De Jonge, L. Koymans and M. Vinkers, *Journal of Computational Chemistry*, 2007, **28**, 890–898.

28 T. F. Willems, C. H. Rycroft, M. Kazi, J. C. Meza and M. Haranczyk, *Micropor. Mesopor. Mater.*, 2012, **149**, 134–141.

29 R. L. Martin and M. Haranczyk, *Cryst. Growth Des*, 2014, **14(5)**, 2431–2440.

30 D. J. Tranchemontagne, J. L. Mendoza-Cortés, M. O'Keeffe and O. M. Yaghi, *Chemical Society Reviews*, 2009, **38**, 1257.

31 R. L. Martin and M. Haranczyk, *Crystal Growth & Design*, 2013, **13**, 4208–4212.

32 R. L. Martin and M. Haranczyk, *Chem. Sci.*, 2013, **4**, 1781–1785.

33 J. A. Mason, M. Veenstra and J. R. Long, *Chemical Science*, 2014, **5**, 32–51.

34 M. Eddaoudi, J. Kim, D. Vodak, A. Sudik, J. Wachter, M. O'Keeffe and O. M. Yaghi, *PNAS*, 2002, **99**, 4900–4904.

35 J. Goldsmith, A. G. Wong-Foy, M. J. Cafarella and D. J. Siegel, *Chem. Mater.*, 2013, **25**, 3373–3382.

36 B. Kesanli, Y. Cui, M. R. Smith, E. W. Bittner, B. C. Bockrath, and W. Lin, *Angewandte Chemie International Edition*, 2005, **44**, 72–75.

37 O. K. Farha, A. Ö. Yazaydin, I. Eryazici, C. D. Malliakas, B. G. Hauser, M. G. Kanatzidis, S. T. Nguyen and R. Q. S. J. T. Hupp, *Nature Chemistry*, 2010, **2**, 944–948.

38 H.-L. Jiang, B. Liu, Y.-Q. Lan, K. Kuratani, T. Akita, H. Shioyama, F. Zong and Q. Xu, *Journal of the American Chemistry Society*, 2011, **133(31)**, 11854–11857.

39 R. L. Martin, C. M. Simon, B. Medasani, D. K. Britt, B. Smit and M. Haranczyk, *Journal of Physical Chemistry C*, 2014, **118(41)**, 23790–23802.

40 R. L. Martin, C. M. Simon, B. Smit and M. Haranczyk, *Journal of Physical Chemistry C*, 2014, **136(13)**, 5006–5022.

41 H. Li, M. Eddaoudi, M. O'Keeffe and O. M. Yaghi, *Nature*, 1999, **402**, 276–279.

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

43 C. M. Simon, J. Kim, D. A. Gomez-Gualdron, J. S. Camp, Y. G. Chung, R. L. Martin, R. Mercado, M. W. Deem, D. Gunter, M. Haranczyk, D. S. Sholl, R. Q. Snurr and B. Smit, *Energy & Envi. Sci.*, 2015.