Electronic Supplementary Material (ESI) for Inorganic Chemistry Frontiers.

Supporting Information

Functional Group Pair Distance based Descriptor for Isomerisation in Porous Molecular Framework Materials

Maryam Nurhuda[a], Yusuf Hafidh[b], Cansu Dogan[a], Daniel Packwood[c], Carole C. Perry[a], Matthew A. Addicoat[a]*

[a] *School of Science and Technology, Nottingham Trent University, Clifton Lane, Nottingham NG11 8NS*

[b] *Department of Mathematics, Bandung Institute of Technology, Jl. Ganesa 10 Bandung - Jawa Barat, Indonesia*

[c] *Kyoto Univ, Inst Integrated Cell Mat Sci, Inst Adv Study, Kyoto 6068501, Japan*

Supporting Information:

S0 Motivation

S1 Number of isomers of each pore topology arising from different functional group arrangements.

S2 Number of Tet^2Di^4 Unique Isomer Calculation Using Group Theory

S3 Algorithm for Isomer enumeration.

S4 Enumeration and analysis of 2D D_{4h} cage

S5 Histograms of the functional group - functional group distances of every pore topology.

S6 Functional Group - Functional Group distance Histogram of Tet^2Di^4 Isomers.

S7 The effect of rotation on functional group distances.

S8 Kernel Density Estimation on Histograms

S9 Principal Component Analysis

S10 Representative isomers of NH_2-UiO-66 Oh Cage for the Machine Learning Input.

S11 Machine Learning method

S1 Number of isomers of each pore topology arising from different functional group arrangements

Pore topology	Number of Isomers
Tri^2Di^3	10
Tri^4Di^6	176
$Tri_2^4Di^6$	568
Tri^6Di^9	21856
Tri^8Di^{12}	351976
$Tri^{20}Di^{30}$	9607679214180672
Tet^2Di^4	31
$Tet_3^3Di^3$	366
$Tet_4^4Di^8$	4244
Tet^5Di^{10}	52740
Tet^6Di^{12}	354024
Tet^8Di^{16}	268439588
$Tet^{16}Di^{32}$	17592195482624
$Tet^{24}Di^{48}$	1.650586719047232e27

Table S1: Number of isomers arising from different arrangement of functional group position of pore cage structure, functionalisation is based on one functional group per linker.

S2 Example of Unique Isomer Calculation Using Group Theory

we consider Tet^2Di^4, the pore has point group D_{4h} with 16 symmetry elements:

$$\begin{aligned} G =& E, 2 \cdot C_4, C_2, 2 \cdot C_2', 2 \cdot C_2'', \\ & i, 2 \cdot S_4, \sigma_h, 2 \cdot \sigma_v, 2 \cdot \sigma_d \end{aligned}$$

and the number of possible structures is:

$$H = 4^{n_{linkers}} = 4^4 = 256$$

Then the total number, n of unique structures is calculated as:

$$\begin{aligned} n =& \frac{1}{16}[\chi(E) + 2 \bullet \chi(C_4) + \chi(C_2) + 2 \bullet \chi(C_2') \\ & + 2 \bullet \chi(C_2'') + \chi(i) + 2 \bullet \chi(S_4) + \chi(\sigma_h) \\ & + 2 \bullet \chi(\sigma_v) + 2 \bullet \chi(\sigma_d)] = 31 \end{aligned} \tag{S1}$$

with each term of $\chi(x)$ of the symmetry elements equal to:

$\chi(E) = 256$	$\chi(C_4) = 4$
$\chi(C_2) = 16$	$\chi(C_2') = 16$
$\chi(C_2'') = 4$	$\chi(i) = 16$
$\chi(S_4) = 16$	$\chi(\sigma_h) = 0$
$\chi(\sigma_v) = 0$	$\chi(\sigma_d) = 64$

S3 Algorithm of Isomer Enumeration

Step 1: Preparation

Step 1a: Labelling each functional group slot with an index.

In this step, the pore structures are installed with full functionalisation, which means all hydrogen atoms in the benzene linker are replaced by a functional group (FG). Then each functional group slot is labelled by an index named as FG_id. FG_id in the same linker need to be in a sequence of 4 consecutive numbers, as shown in Figure S1 .

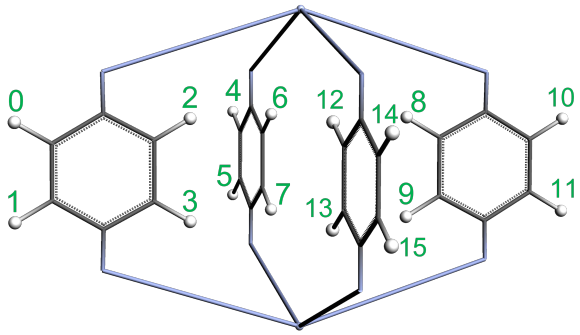
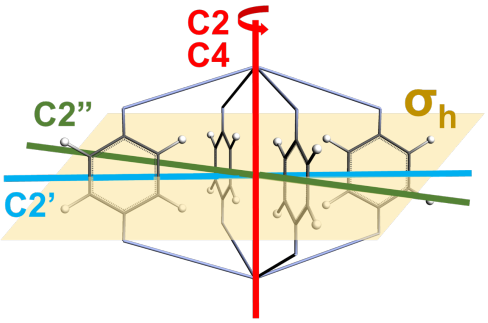


Figure S1: Pore Tet^2Di^4 with assigned index for each functional group, functional group slot is shown in blue circles.

Step 1b: Generating a transformation library.

After indexing each FG slot, symmetry operations are applied to the pore. Again, Tet^2Di^4 is used as an example, Tet^2Di^4 has point group of D_{4h} with 16 symmetry elements. Therefore, the pore is transformed 16 times. After every transformation, the new structure is compared to the original pore structure to list the FG_id position of the new structure, this information is stored in a table which we refer to as the transformation library (will be used in the next step). In the transformation library, the first column represents the initial position of FG. The columns after are the transformed (and projected) FG_id for each symmetry operation. For the Tet2Di4 pore, the transformation library will be a table of dimension 16x16 (Figure S2). The 16 rows correspond to the 16 possible FG position, and the 16 columns correspond to the 16 symmetry elements.



E	C_4	C_2	C_2'	.....	C_2''	σ_h
0	14	10	1		5	1
1	15	11	0		4	0
2	12	8	3		7	3
3	13	9	2		6	2
4	0	12	15		1	5
8	6	2	9		13	9
12	8	6	7		9	13
15	11	5	4		10	14

Figure S2: (right) Symmetry elements of pore Tet^2Di^4 (only 5 from the 16 are shown) (left) Transformation library, library of each functional group index and the transformed index, the first column represents the initial position, each subsequent column represent a specified symmetry operation, double lines represent hidden rows.

Generating a transformation library beforehand is useful for tracking the possible symmetry transformations of the pore, specifically in cutting down repetitive transformations (of the pore) each of which involve multiplication of the xyz coordinate matrices and transformation matrices which can became highly computationally costly.

Step 2: Enumeration of all possible functional group arrangement.

Using the example of Tet^2Di^4 pore, as each linker is installed with one functional group, the pore isomers will have 4 functional groups. The isomers are encoded to a string of 4 FG_id and also encoded to an equivalent single integer isomer_id, defined by Equation S2. All possible combinations of FG_id are listed (an example is shown in S3). Along with that, a unique flag list is prepared to mark the unique isomers Figure 5.10.

$$\text{isomer\_id} = \sum_{\text{FG\_id}} (\text{FG\_id}\%4) \bullet 4^{int(\frac{\text{FG\_id}}{4})} \tag{S2}$$

Isomer_Id	FG_id				Unique?
0	0	4	8	12	yes/no
1	1	4	8	12	yes/no
2	2	4	8	12	yes/no
3	3	4	8	12	yes/no
4	0	5	8	12	yes/no
5	1	5	8	12	yes/no
6	2	5	8	12	yes/no
..					
255	3	7	11	15	yes/no

Figure S3: Example of functional group enumeration unique list

Step 2a: Finding equivalent symmetry related isomers.

For each isomer, equivalent symmetry related isomers are determined by looking at the transformation library generated in step 2. For example, as illustrated in Figure S4, in the first isomer (isomer_id=0), functionalisation is located at FG_id 0-4-8-12, so the FG_id rows are highlighted in green. Then, the highlighted green rows are copied into a new matrix and transposed. Each of the rows from the transpose matrix are the equivalent isomers.

E	C_4	C_2	C_2'	.....	C_2''	σ_h
0	14	10	1		5	1
1	15	11	0		4	0
2	12	8	3		7	3
3	13	9	2		6	2
4	0	12	15		1	5
8	6	2	9		13	9
12	8	6	7		9	13
15	11	5	4		10	14

Equivalent isomers:

14	0	6	8
10	12	2	6
1	15	9	7
..			
5	1	13	9
1	5	9	13

Figure S4: Illustration of equivalent isomers determined from transformation list

Step 2b: Marking the unique flag list.

As illustrated in Figure S5, Isomer_id 0 will be marked as unique in the unique flag list, while the equivalent isomers are marked as a duplicate. To search for the Ids of the equivalent isomers, the string of 4 FG_ids is converted back to its isomer_Id by Equation S2. Steps 2a and 2b are repeated for the next unsigned unique flag list isomer until all isomer_id in the unique flag list are marked.

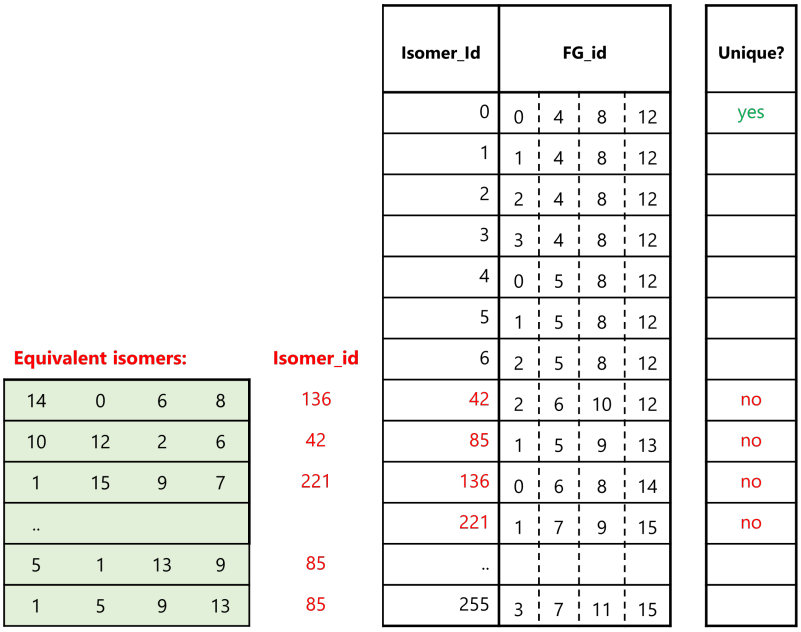
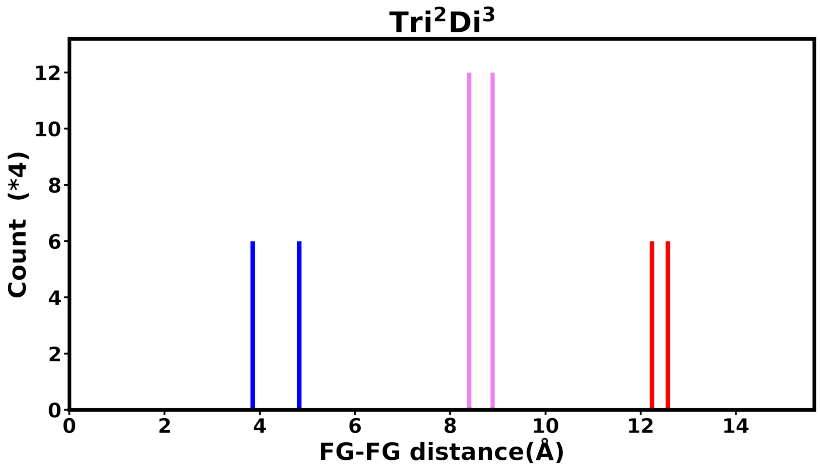


Figure S5: Example of functional group enumeration unique list

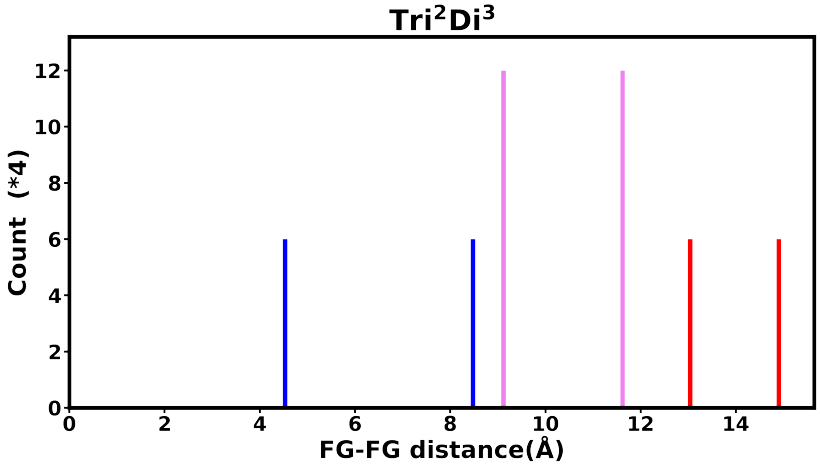
In huge pores with more than 12 ditopic linkers, the number of isomers exceeds the maximum size of a python list. Encoding and decoding the isomers has significantly reduced the computational load. The encoding also cuts the computational time that would be required to search for string containing n_{FG} numbers in a list of $4^{nlinker}$ isomers.

S4 Enumeration and analysis of 2D D_{4h} cage

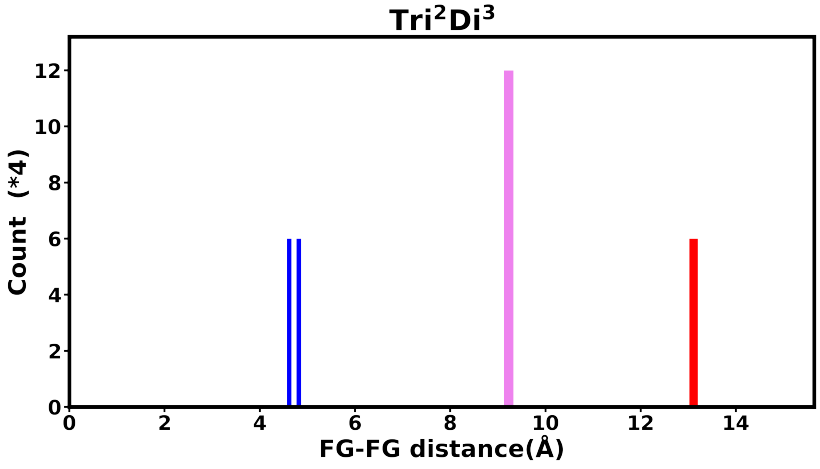
S5 Histograms of the functional group - functional group distances of every pore topology.



(a) Benzene-1,4-dicarboxylic acid linker

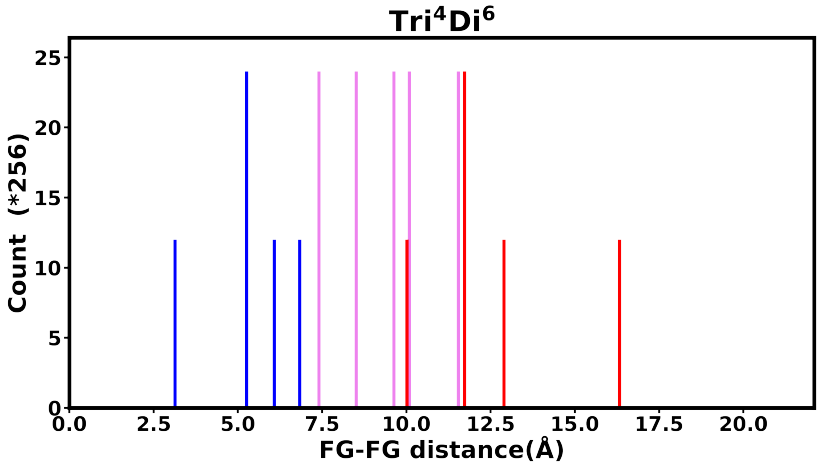


(b) Outer poly(1,4-benzenedicarboxylic acid) linker

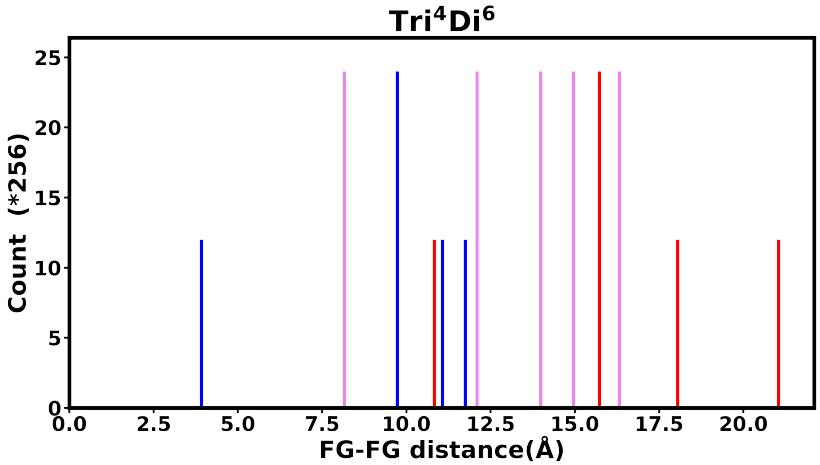


(c) Inner poly(1,4-benzenedicarboxylic acid) linker

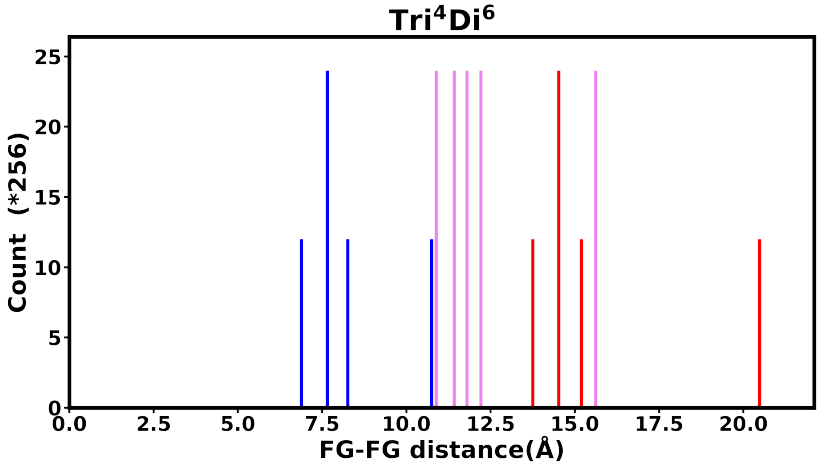
Figure S6: Histogram of FG-FG distances in pore Tri^2Di^3, constructed from a node with radius 5Å and two different linker sizes, (a) benzene-1,4-dicarboxylic acid, (b) poly(1,4-benzenedicarboxylic acid) outer functional group positions (c) poly(1,4-benzenedicarboxylic acid) inner functional group positions.



(a) Benzene-1,4-dicarboxylic acid linker

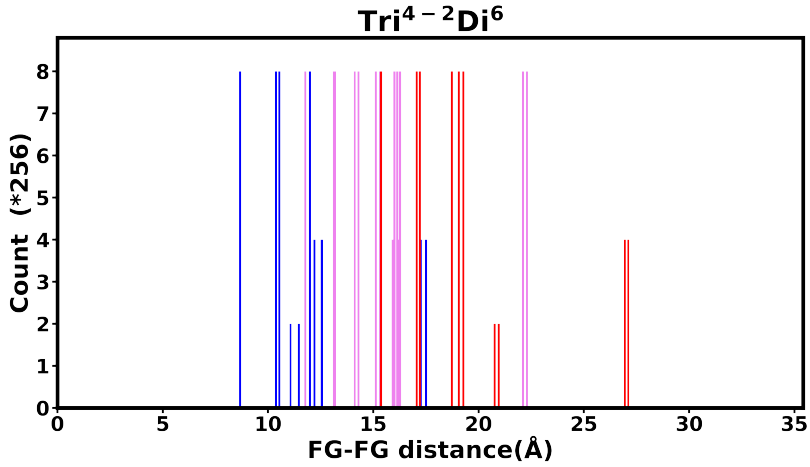


(b) Outer poly(1,4-benzenedicarboxylic acid) linker

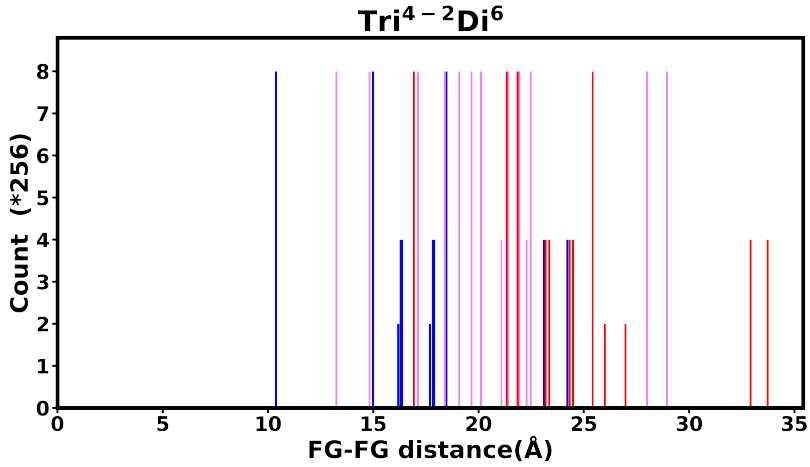


(c) Inner poly(1,4-benzenedicarboxylic acid) linker

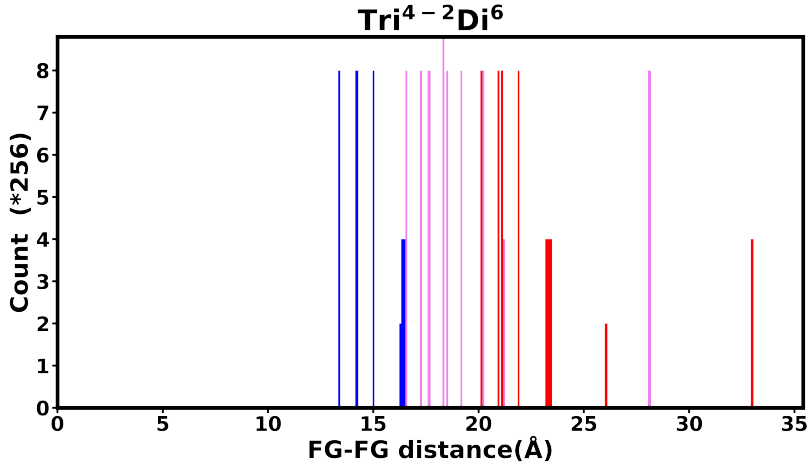
Figure S7: Histogram of FG-FG distances in pore Tri^4Di^6, constructed from a node with radius 5Å and two different linker size, (a) benzene-1,4-dicarboxylic acid, (b) poly(1,4-benzenedicarboxylic acid) outer functional group positions (c) poly(1,4-benzenedicarboxylic acid) inner functional group positions.



(a) Benzene-1,4-dicarboxylic acid linker

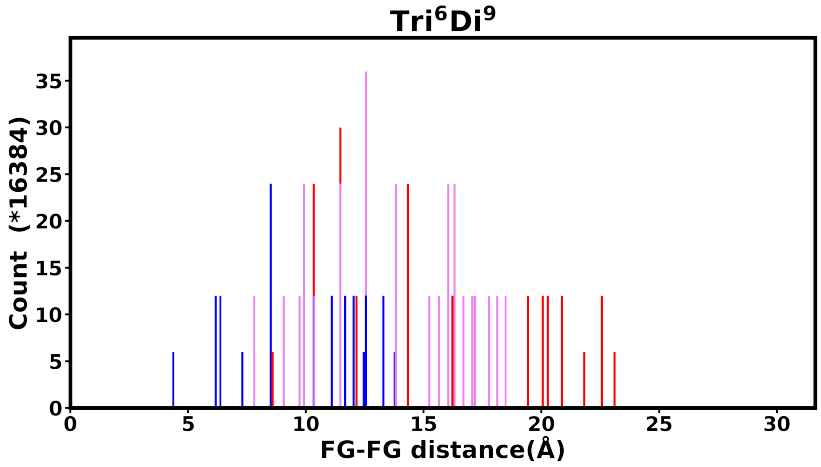


(b) Outer poly(1,4-benzenedicarboxylic acid) linker

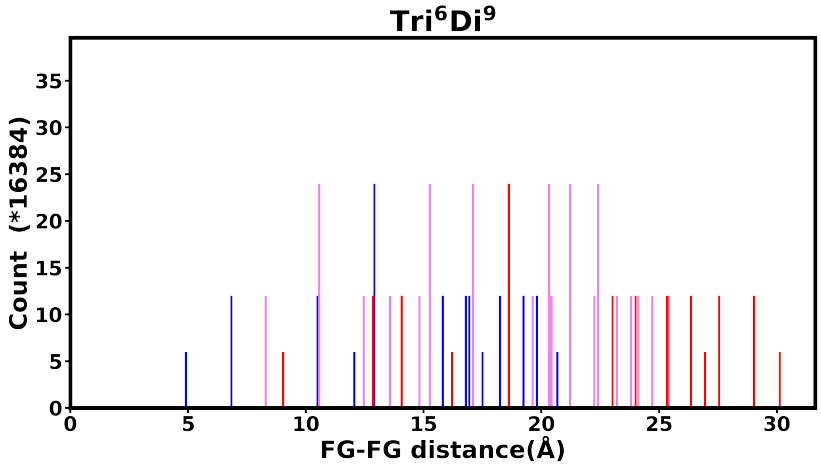


(c) Inner poly(1,4-benzenedicarboxylic acid) linker

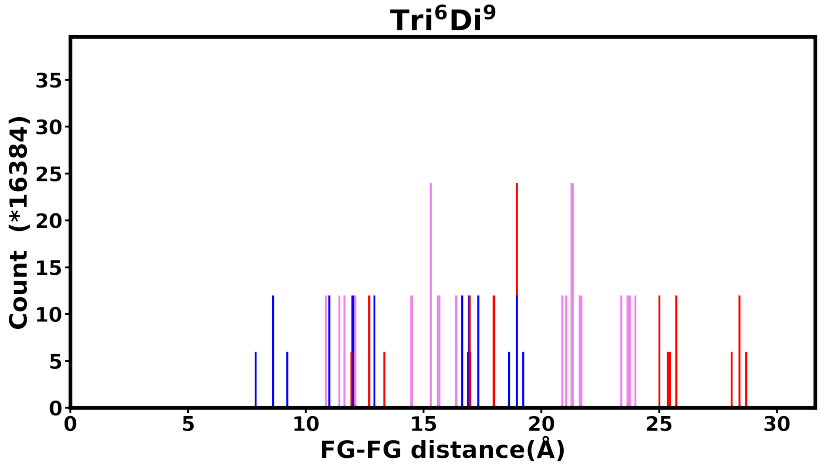
Figure S8: Histogram of FG-FG distances in pore $Tri_2^4Di^6$, constructed from a node with radius 5Å and two different linker size, (a) benzene-1,4-dicarboxylic acid, (b) poly(1,4-benzenedicarboxylic acid) outer functional group positions (c) poly(1,4-benzenedicarboxylic acid) inner functional group positions.



(a) Benzene-1,4-dicarboxylic acid linker

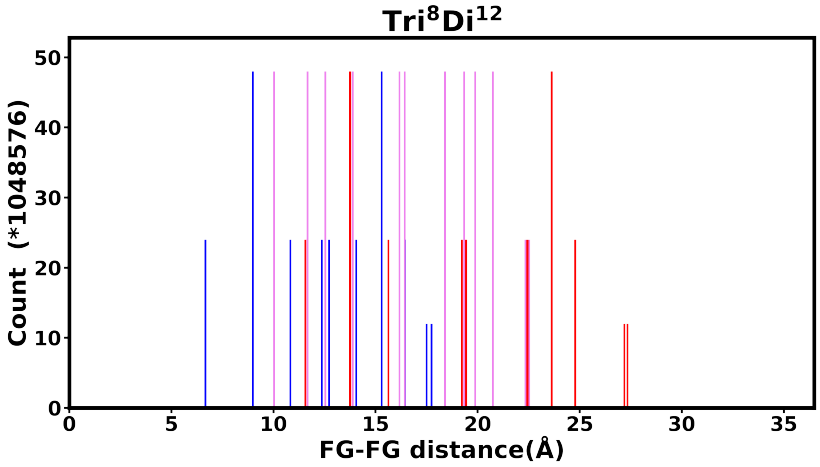


(b) Outer poly(1,4-benzenedicarboxylic acid) linker

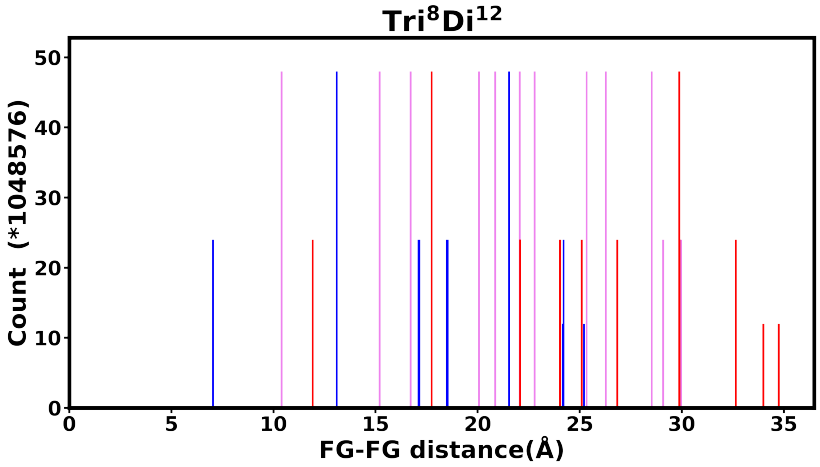


(c) Inner poly(1,4-benzenedicarboxylic acid) linker

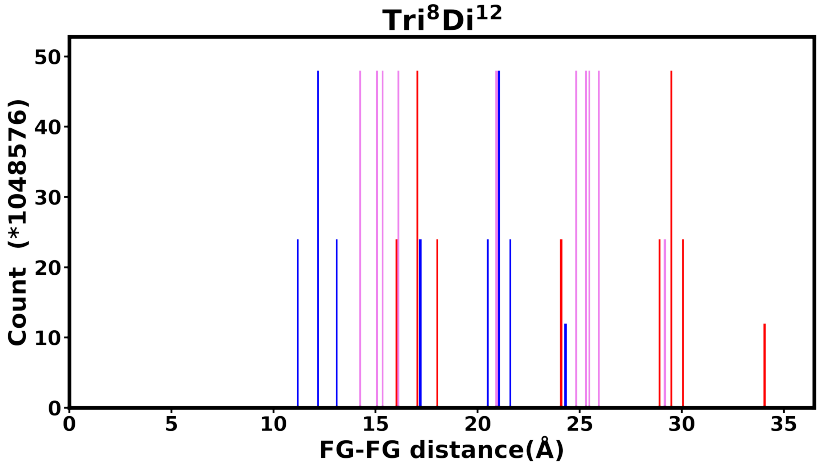
Figure S9: Histogram of FG-FG distances in pore Tri^6Di^9, constructed from a node with radius 5Å and two different linker sizes, (a) benzene-1,4-dicarboxylic acid, (b) poly(1,4-benzenedicarboxylic acid) outer functional group positions (c) poly(1,4-benzenedicarboxylic acid) inner functional group positions.



(a) Benzene-1,4-dicarboxylic acid linker

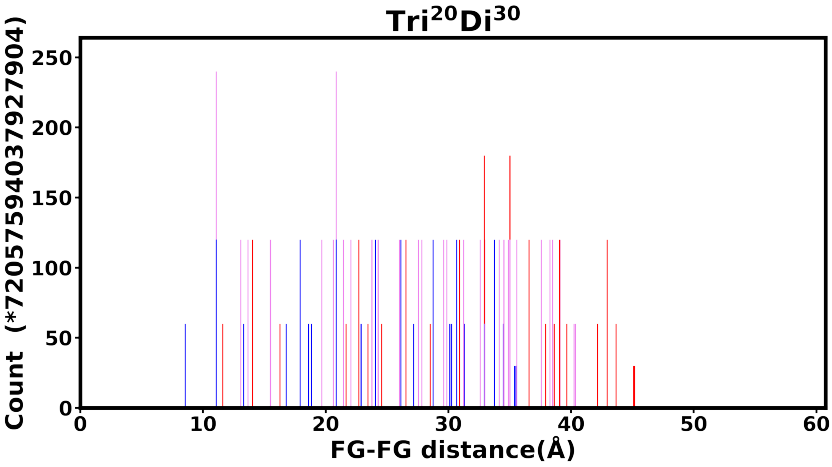


(b) Outer poly(1,4-benzenedicarboxylic acid) linker

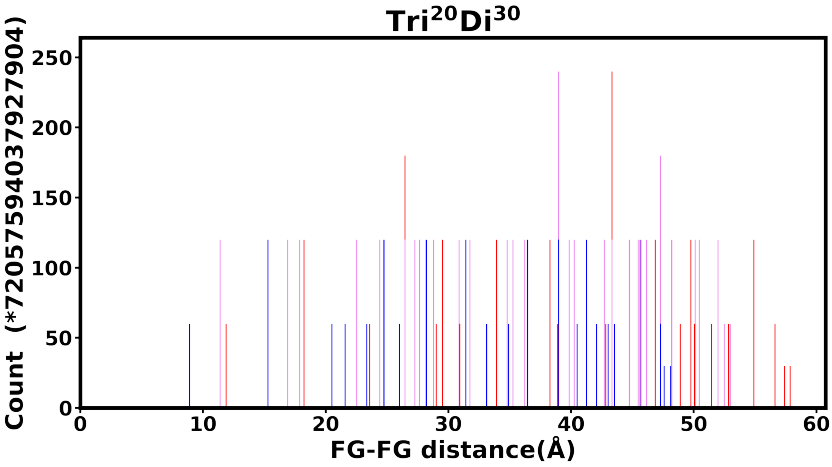


(c) Inner poly(1,4-benzenedicarboxylic acid) linker

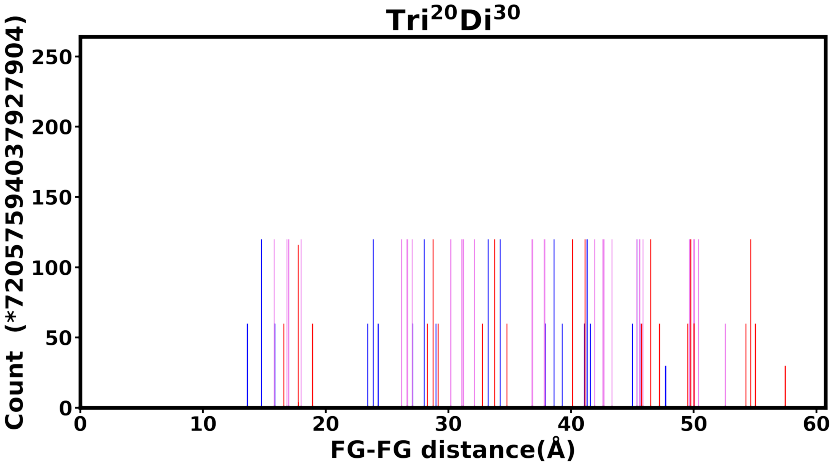
Figure S10: Histogram of FG-FG distances in pore Tri^8Di^{12}, constructed from a node with radius 5Å and two different linker sizes, (a) benzene-1,4-dicarboxylic acid, (b) poly(1,4-benzenedicarboxylic acid) outer functional group positions (c) poly(1,4-benzenedicarboxylic acid) inner functional group positions.



(a) Benzene-1,4-dicarboxylic acid linker

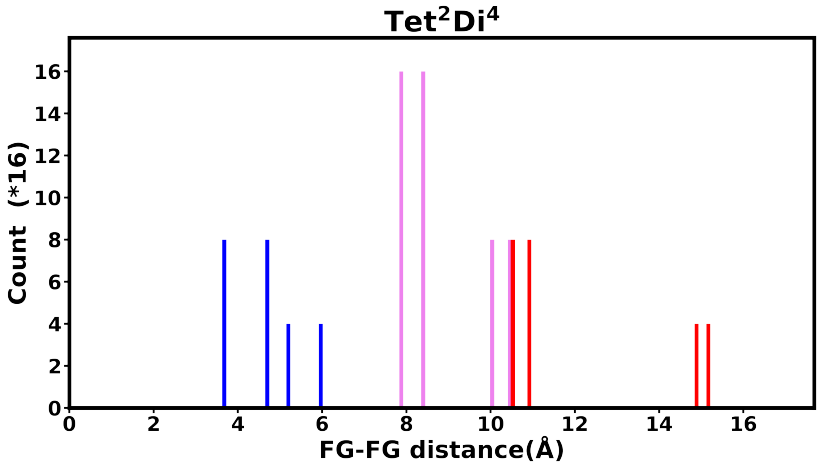


(b) Outer poly(1,4-benzenedicarboxylic acid) linker

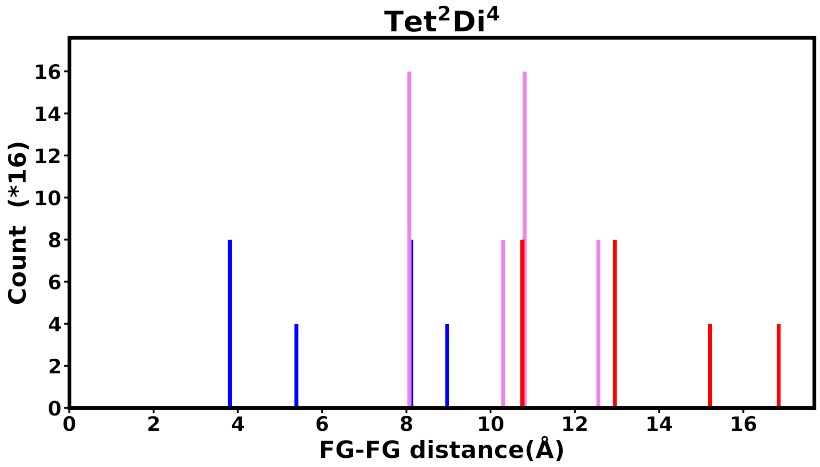


(c) Inner poly(1,4-benzenedicarboxylic acid) linker

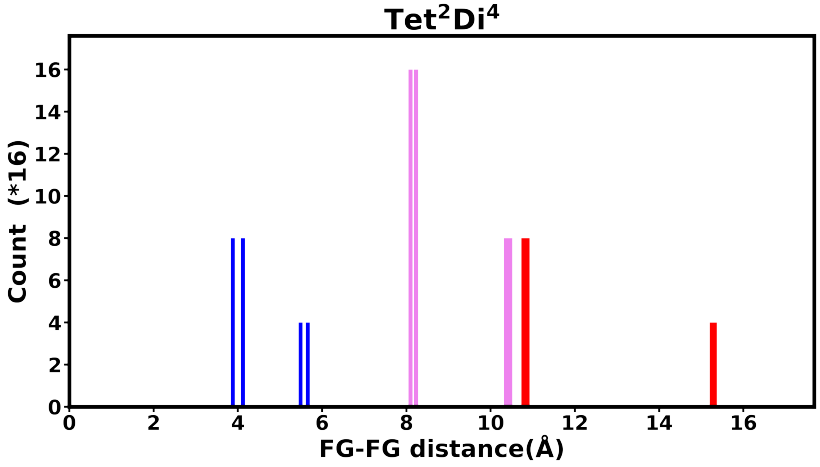
Figure S11: Histogram of FG-FG distances in pore $Tri^{20}Di^{30}$, constructed from a node with radius 5Å and two different linker sizes, (a) benzene-1,4-dicarboxylic acid, (b) poly(1,4-benzenedicarboxylic acid) outer functional group positions (c) poly(1,4-benzenedicarboxylic acid) inner functional group positions.



(a) Benzene-1,4-dicarboxylic acid linker

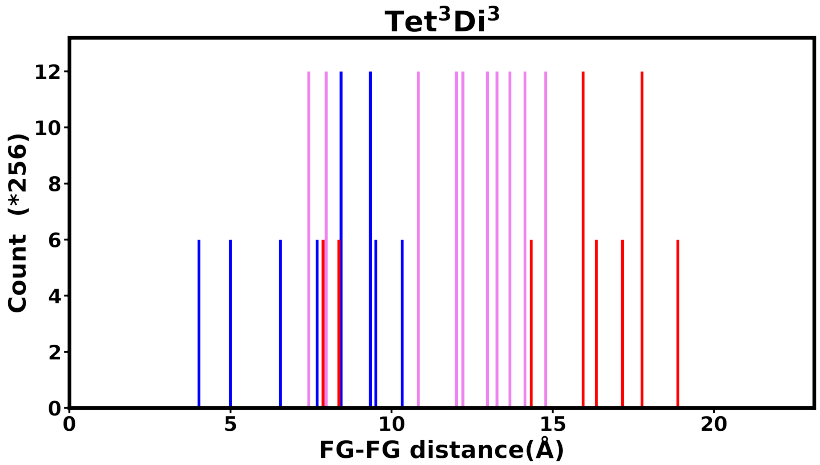


(b) Outer poly(1,4-benzenedicarboxylic acid) linker

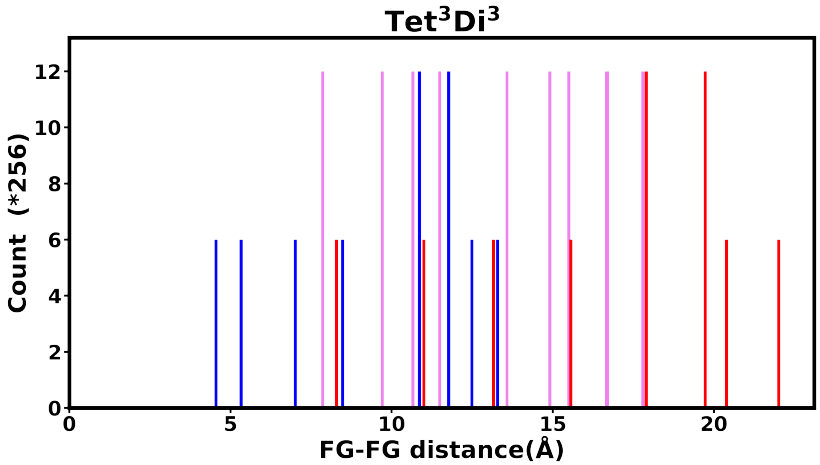


(c) Inner poly(1,4-benzenedicarboxylic acid) linker

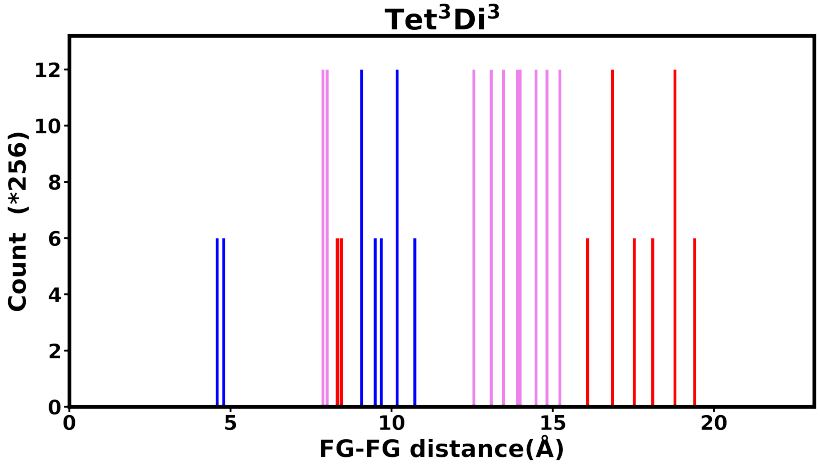
Figure S12: Histogram of FG-FG distances in pore Tet^2Di^4, constructed from a node with radius 5Å and two different linker sizes, (a) benzene-1,4-dicarboxylic acid, (b) poly(1,4-benzenedicarboxylic acid) outer functional group positions (c) poly(1,4-benzenedicarboxylic acid) inner functional group positions.



(a) Benzene-1,4-dicarboxylic acid linker

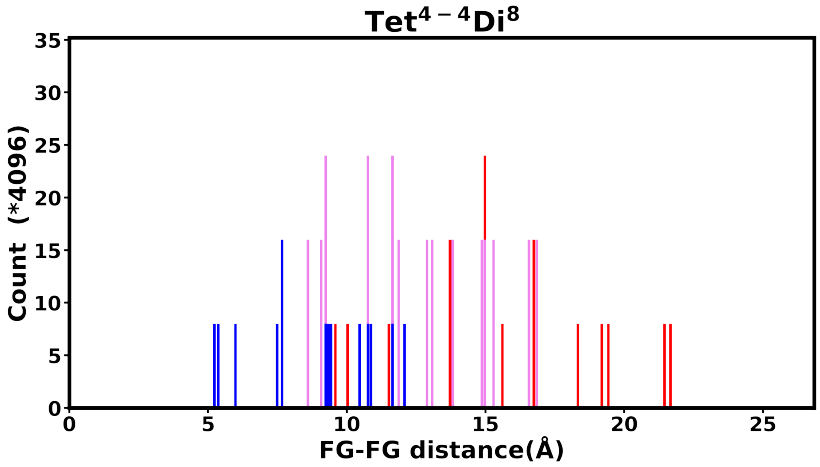


(b) Outer poly(1,4-benzenedicarboxylic acid) linker

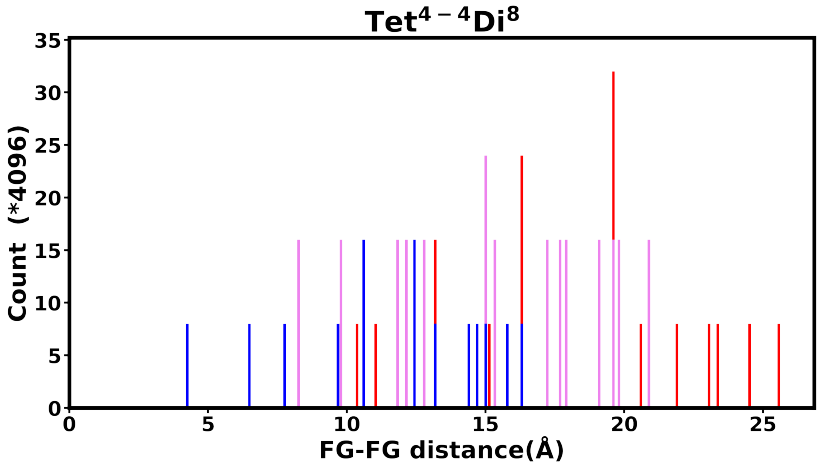


(c) Inner poly(1,4-benzenedicarboxylic acid) linker

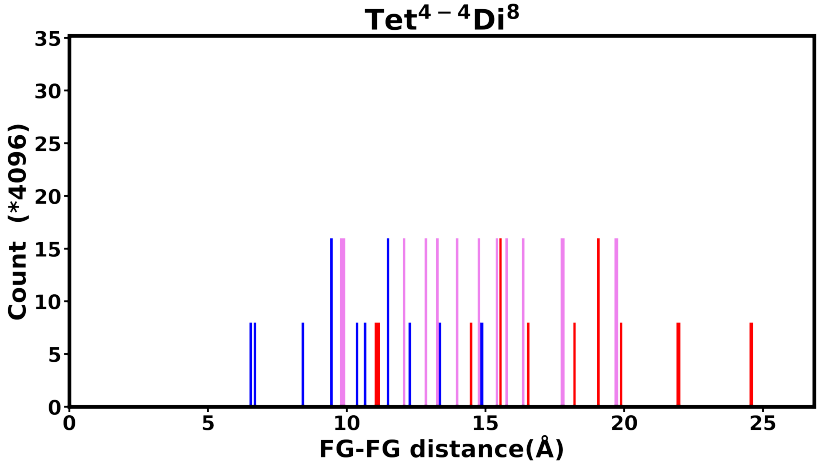
Figure S13: Histogram of FG-FG distances in pore $Tet_3^3Di^3$, constructed from a node with radius 5Å and two different linker sizes, (a) benzene-1,4-dicarboxylic acid, (b) poly(1,4-benzenedicarboxylic acid) outer functional group positions (c) poly(1,4-benzenedicarboxylic acid) inner functional group positions.



(a) Benzene-1,4-dicarboxylic acid linker

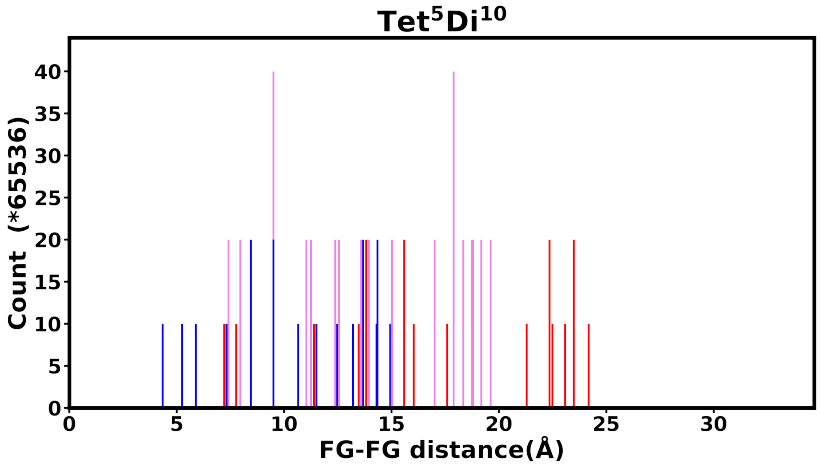


(b) Outer poly(1,4-benzenedicarboxylic acid) linker

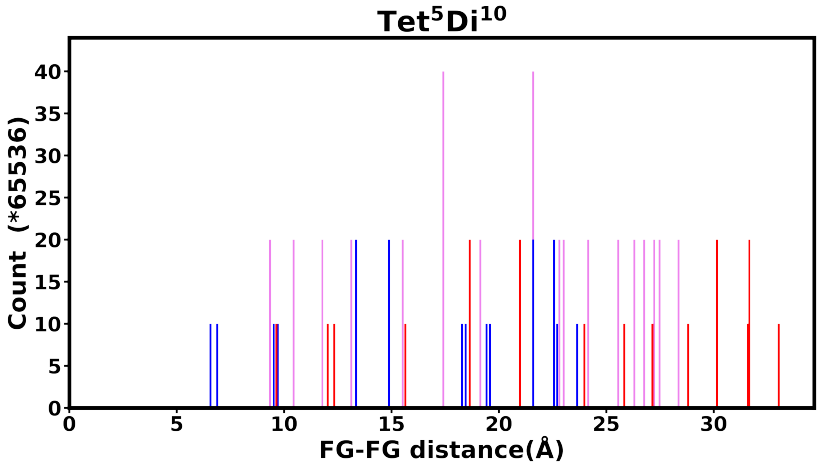


(c) Inner poly(1,4-benzenedicarboxylic acid) linker

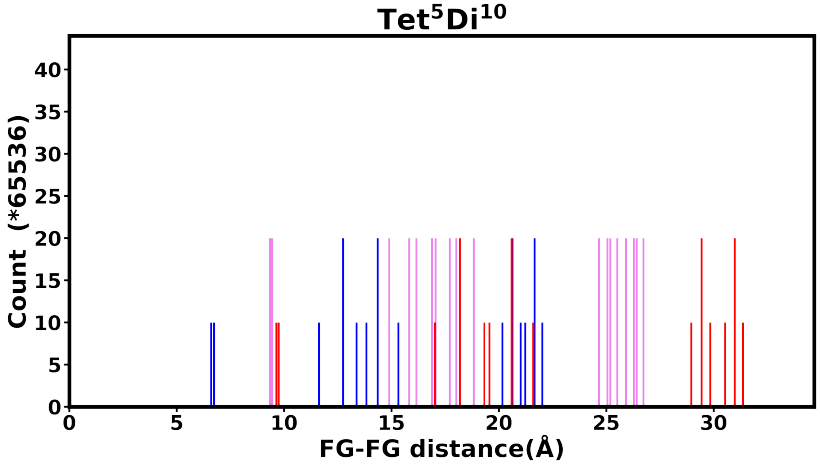
Figure S14: Histogram of FG-FG distances in pore $Tet_4^4Di^8$, constructed from a node with radius 5Å and two different linker sizes, (a) benzene-1,4-dicarboxylic acid, (b) poly(1,4-benzenedicarboxylic acid) outer functional group positions (c) poly(1,4-benzenedicarboxylic acid) inner functional group positions.



(a) Benzene-1,4-dicarboxylic acid linker

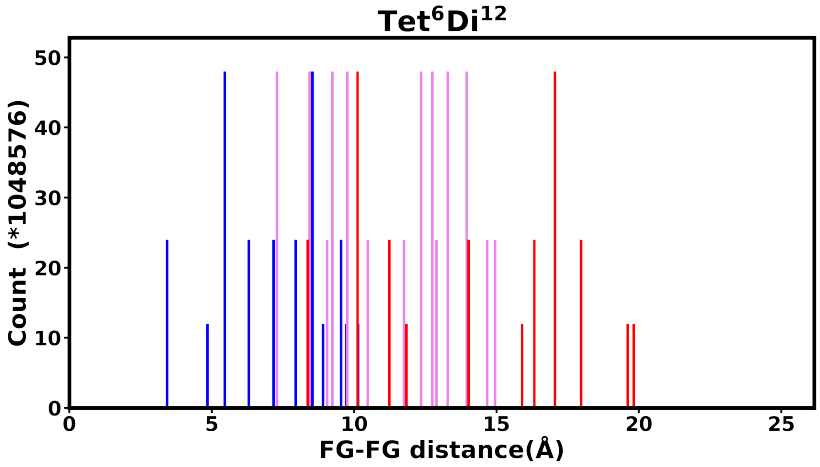


(b) Outer poly(1,4-benzenedicarboxylic acid) linker

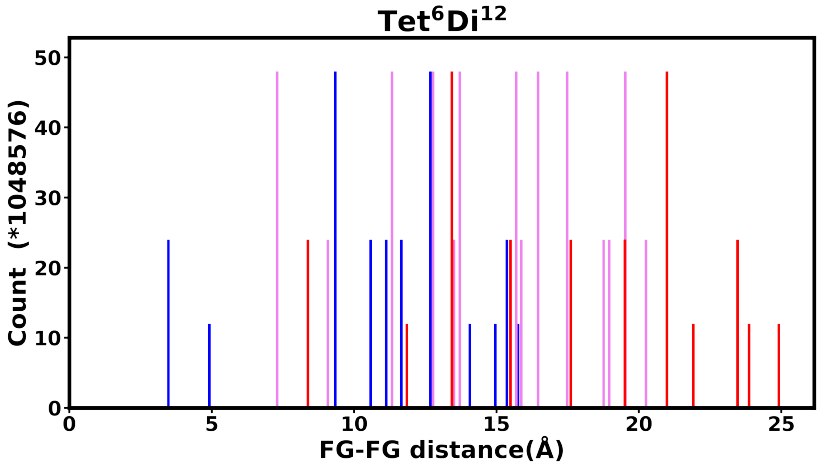


(c) Inner poly(1,4-benzenedicarboxylic acid) linker

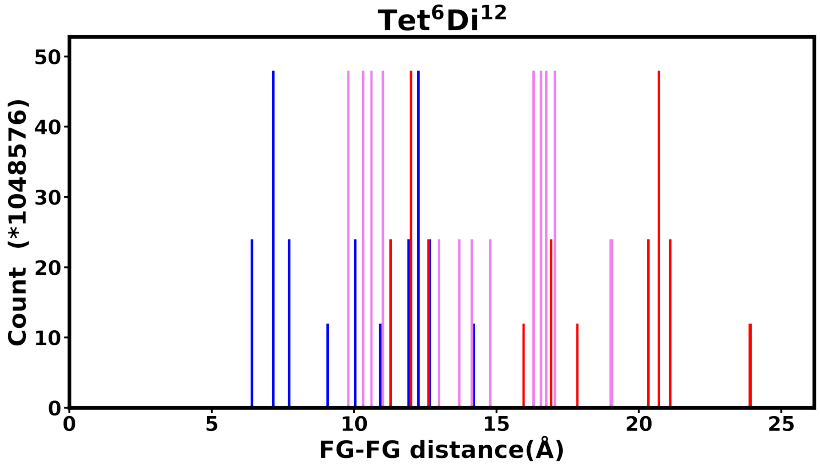
Figure S15: Histogram of FG-FG distances in pore Tet^5Di^10, constructed from a node with radius 5Å and two different linker sizes, (a) benzene-1,4-dicarboxylic acid, (b) poly(1,4-benzenedicarboxylic acid) outer functional group positions (c) poly(1,4-benzenedicarboxylic acid) inner functional group positions.



(a) Benzene-1,4-dicarboxylic acid linker

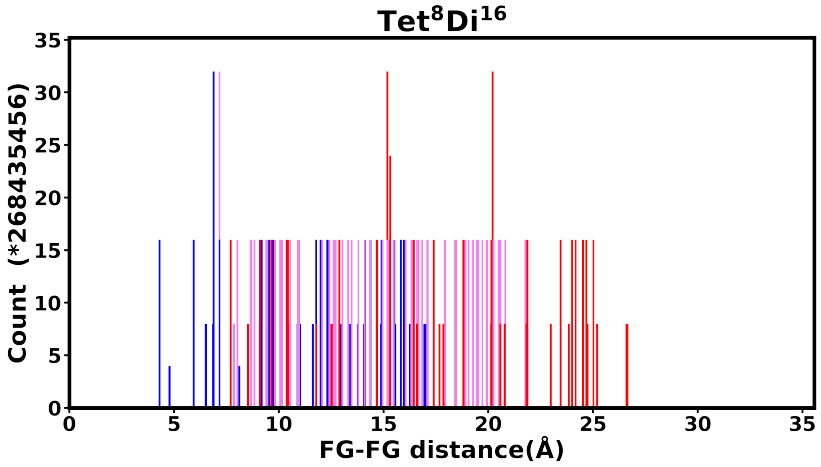


(b) Outer poly(1,4-benzenedicarboxylic acid) linker

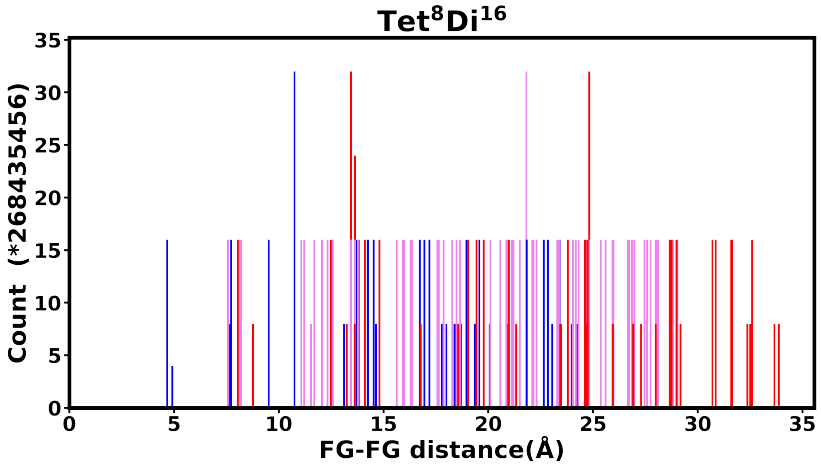


(c) Inner poly(1,4-benzenedicarboxylic acid) linker

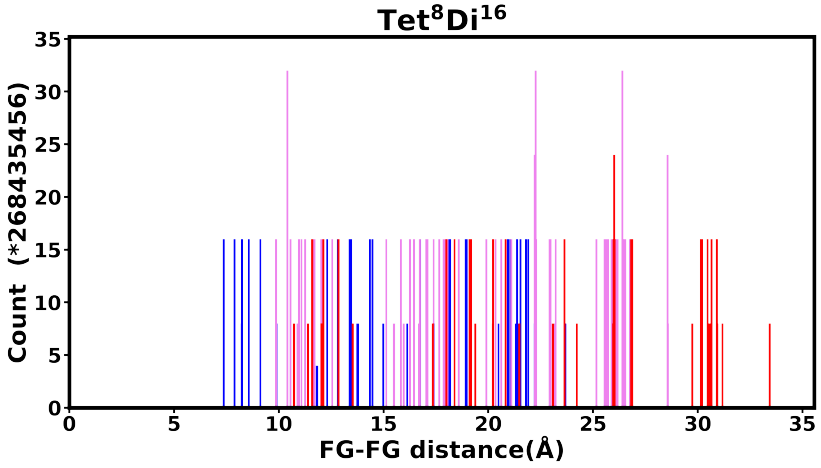
Figure S16: Histogram of FG-FG distances in pore Tet^6Di^{12}, constructed from a node with radius 5Å and two different linker sizes, (a) benzene-1,4-dicarboxylic acid, (b) poly(1,4-benzenedicarboxylic acid) outer functional group positions (c) poly(1,4-benzenedicarboxylic acid) inner functional group positions.



(a) Benzene-1,4-dicarboxylic acid linker

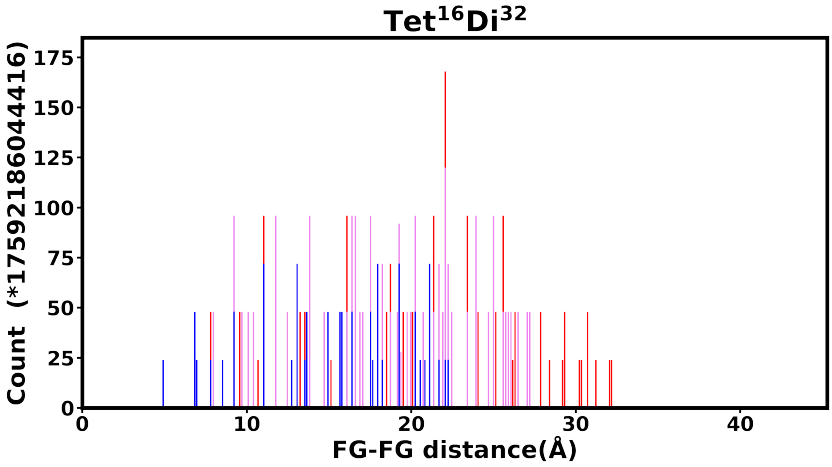


(b) Outer poly(1,4-benzenedicarboxylic acid) linker

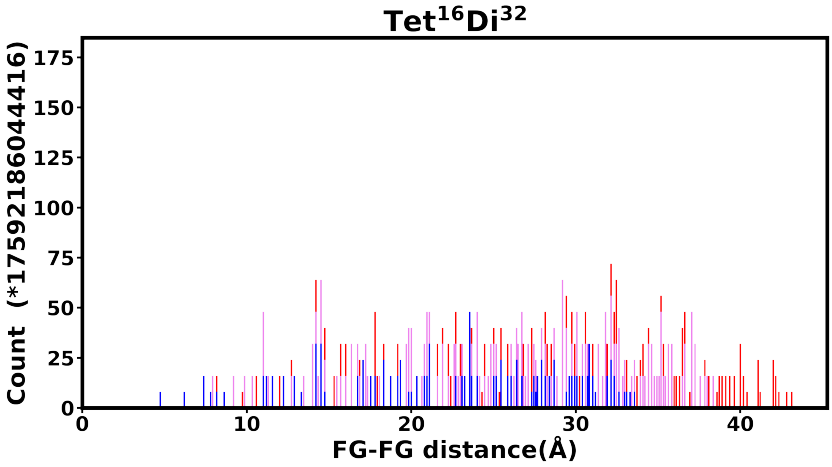


(c) Inner poly(1,4-benzenedicarboxylic acid) linker

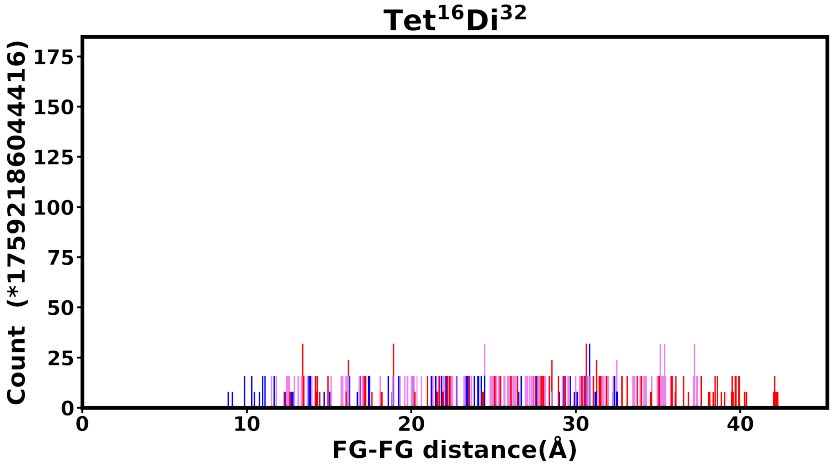
Figure S17: Histogram of FG-FG distances in pore Tet^8Di^{16}, constructed from a node with radius 5Å and two different linker sizes, (a) benzene-1,4-dicarboxylic acid, (b) poly(1,4-benzenedicarboxylic acid) outer functional group positions (c) poly(1,4-benzenedicarboxylic acid) inner functional group positions.



(a) Benzene-1,4-dicarboxylic acid linker

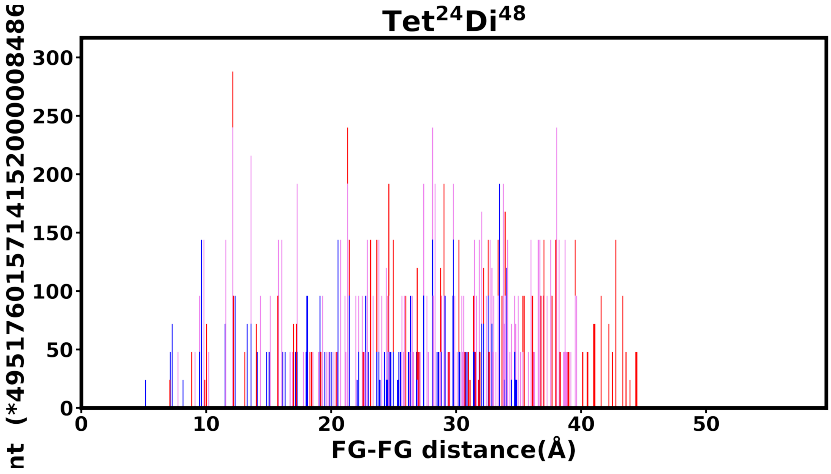


(b) Outer poly(1,4-benzenedicarboxylic acid) linker

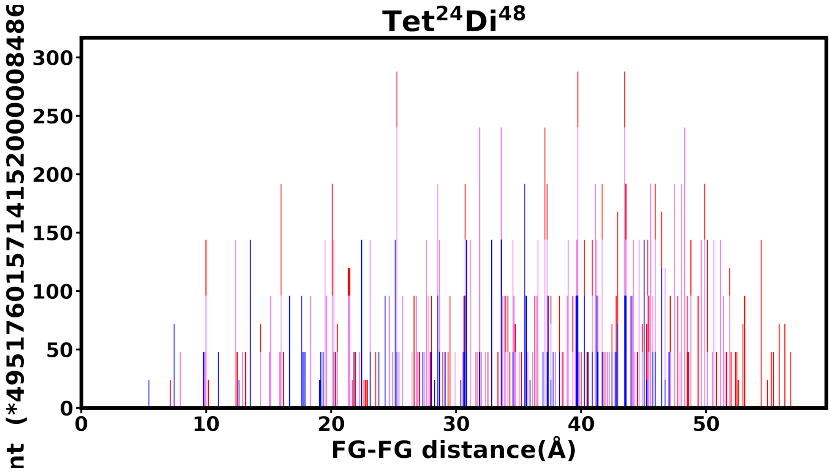


(c) Inner poly(1,4-benzenedicarboxylic acid) linker

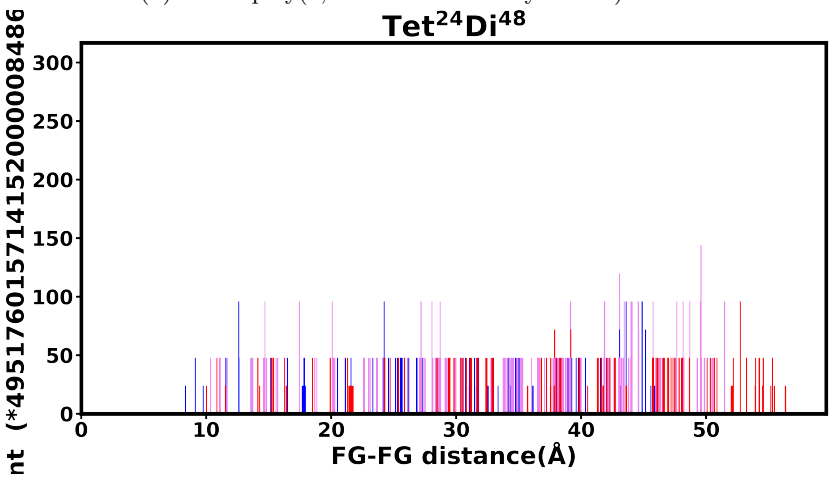
Figure S18: Histogram of FG-FG distances in pore $Tet^{16}Di^{32}$, constructed from a node with radius 5Å and two different linker sizes, (a) benzene-1,4-dicarboxylic acid, (b) poly(1,4-benzenedicarboxylic acid) outer functional group positions (c) poly(1,4-benzenedicarboxylic acid) inner functional group positions.



(a) Benzene-1,4-dicarboxylic acid linker



(b) Outer poly(1,4-benzenedicarboxylic acid) linker



(c) Inner poly(1,4-benzenedicarboxylic acid) linker

Figure S19: Histogram of FG-FG distances in pore $Tet^{24}Di^{48}$, constructed from a node with radius 5Å and two different linker sizes, (a) benzene-1,4-dicarboxylic acid, (b) poly(1,4-benzenedicarboxylic acid) outer functional group positions (c) poly(1,4-benzenedicarboxylic acid) inner functional group positions.

S6 Functional Group - Functional Group distance Histogram of Tet^2Di^4 Isomers.

S7 The effect of linker rotation on the functional group - functional group distances

FG-FG distance(Å)	Shortest Distance(Å)	Maximum distance(Å)
3.68	2.44	5.82
4.70	3.81	6.52
5.20	5.20	6.31
5.96	5.96	6.96
7.91	6.18	9.32
8.43	6.82	9.76
10.07	9.80	10.48
10.48	10.22	10.88
10.57	8.34	11.85
10.95	8.83	12.19
14.94	14.28	14.94
15.22	14.57	15.22

Table S2: Distances found in Tet^2Di^4 pore. The linker rotation deviates the distances from its symmetric position, the maximum and minimum distance is tabulated. Colour code blue represents In-In distances, violet In-out distances, red out-out distances.

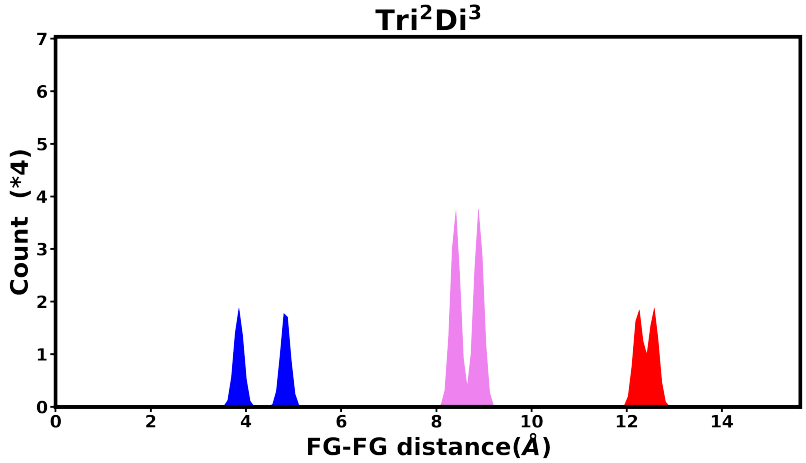
FG-FG distance(Å)	Shortest Distance(Å)	Maximum distance(Å)
3.14	2.39	4.79
5.26	4.63	6.75
6.08	5.33	7.73
6.84	6.75	8.09
7.41	6.13	8.39
8.51	7.27	9.63
9.63	8.40	10.49
10.02	8.33	10.78
10.09	8.80	11.13
11.54	11.11	12.09
11.73	10.08	12.47
12.90	11.21	13.66
16.33	15.55	16.35

Table S3: Distances found in Tri^4Di^6 pore. The linker rotation deviates the distances from its symmetric position, the maximum and minimum distance is tabulated. Distances are colour-coded: blue represents In-In distances, violet In-out distances, red out-out distances.

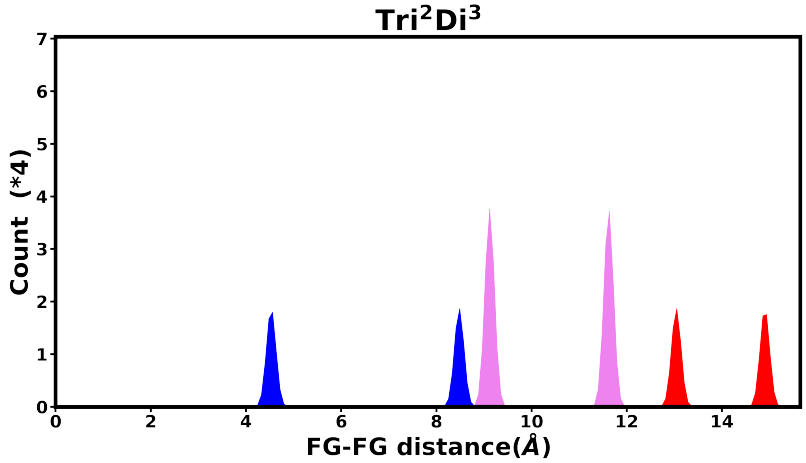
FG-FG distance(Å)	Shortest Distance(Å)	Maximum distance(Å)
3.43	2.01	4.15
4.85	4.85	5.86
5.46	4.24	5.59
6.3	4.88	8.38
7.17	7.17	8.01
7.29	5.57	8.5
7.95	7.22	9.92
8.37	6.29	9.79
8.44	6.71	9.81
8.54	8.24	9.003
8.91	8.91	9.7
9.05	8.82	9.26
9.24	7.57	10.6
9.54	9.54	10.87
9.73	9.73	10.68
9.76	7.94	11.15
10.12	8.12	11.5
10.14	10.14	11.06
10.48	10.28	10.71
11.24	9.16	12.66
11.75	11.75	13.11
11.84	11.37	11.84
12.35	11.11	13.6
12.75	11.8	13.81
12.89	12.66	13.11
13.29	12.31	14.28
13.95	13.25	14.72
14.02	13.55	14.02
14.67	14.39	14.87
14.95	14.68	15.15
15.9	15.43	15.9
16.33	14.6	16.69
17.05	15.5	17.4
17.97	18.27	16.66
19.61	18.95	19.61
19.82	19.33	19.82

Table S4: Distances found in Tet^6Di^{12} pore. The linker rotation deviates the distances from its symmetric position, the maximum and minimum distance is tabulated. Colour code blue represents In-In distances, violet In-out distances, red out-out distances.

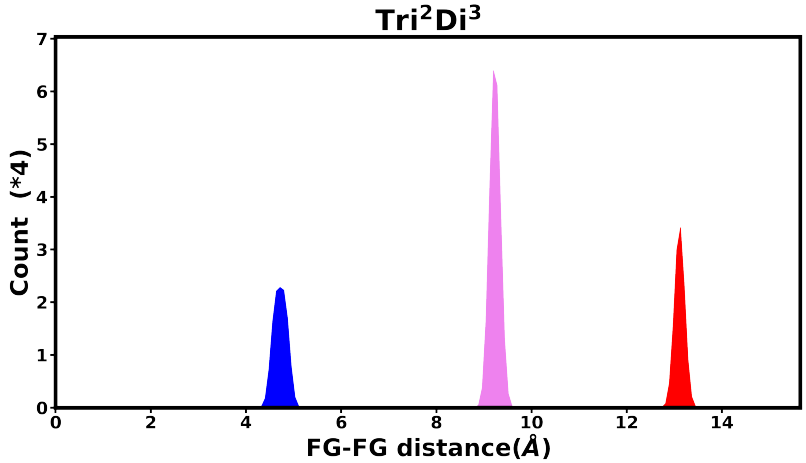
S8 Kernel Density Estimation on the Histograms



(a) Benzene-1,4-dicarboxylic acid linker

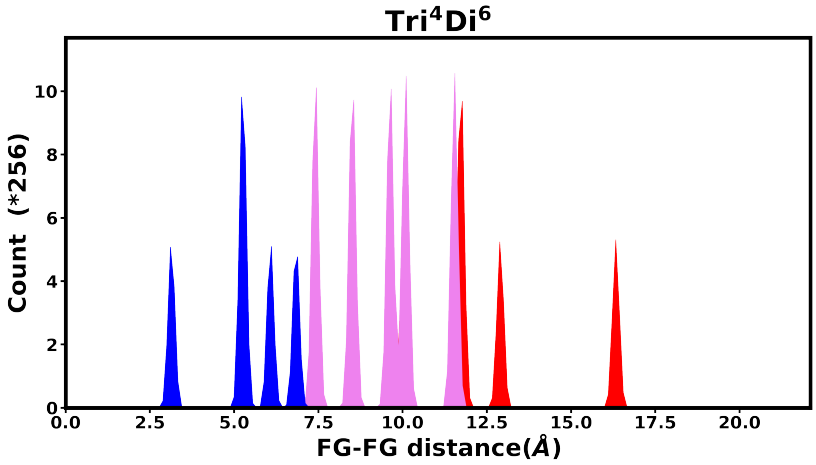


(b) Outer poly(1,4-benzenedicarboxylic acid) linker

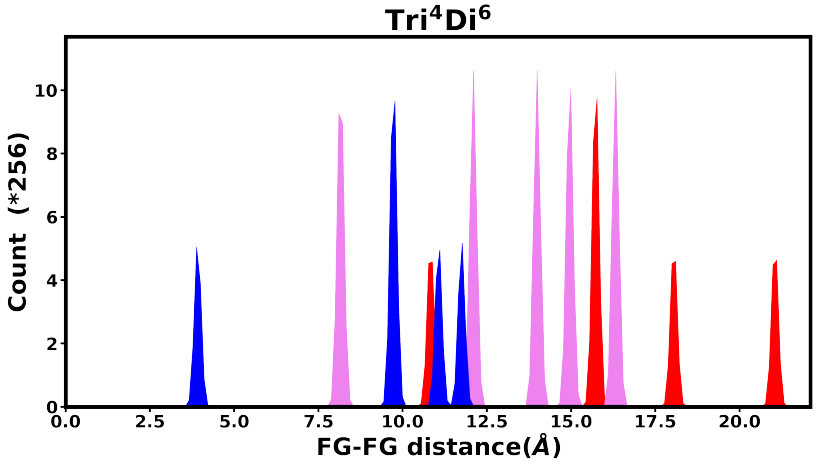


(c) Inner poly(1,4-benzenedicarboxylic acid) linker

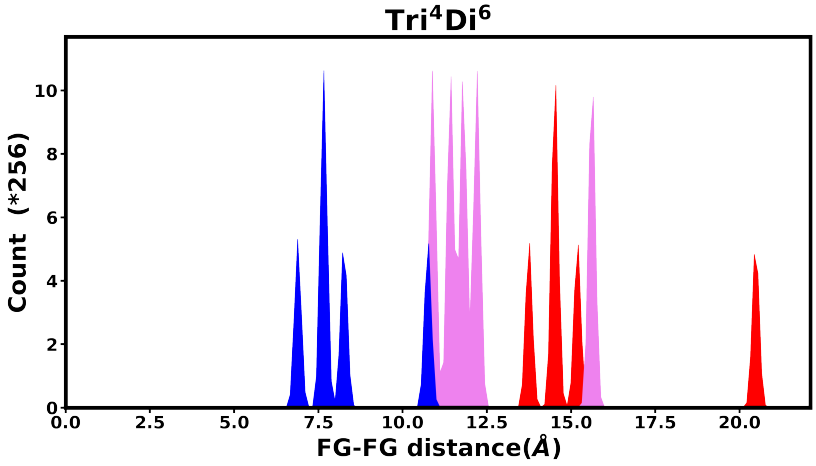
Figure S20: KDE applied on FG-FG distance histograms of pore Tri^2Di^3, constructed from a node with radius 5Å and two different linker sizes, (a) benzene-1,4-dicarboxylic acid, (b) poly(1,4-benzenedicarboxylic acid) outer functional group positions (c) poly(1,4-benzenedicarboxylic acid) inner functional group positions.



(a) Benzene-1,4-dicarboxylic acid linker

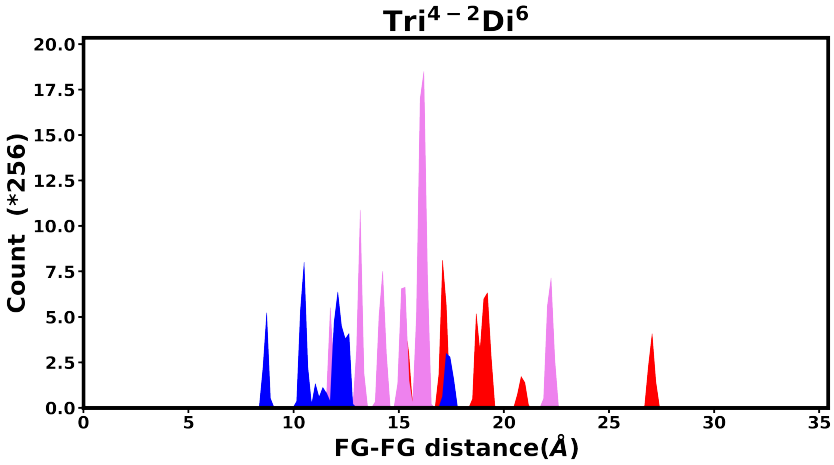


(b) Outer poly(1,4-benzenedicarboxylic acid) linker

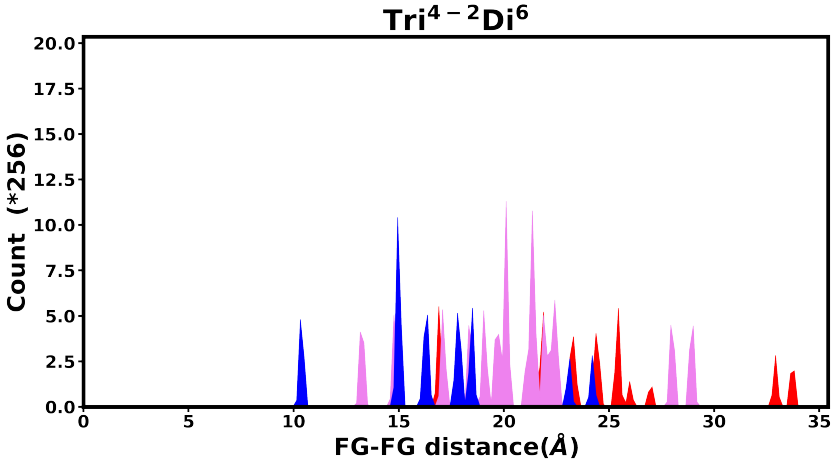


(c) Inner poly(1,4-benzenedicarboxylic acid) linker

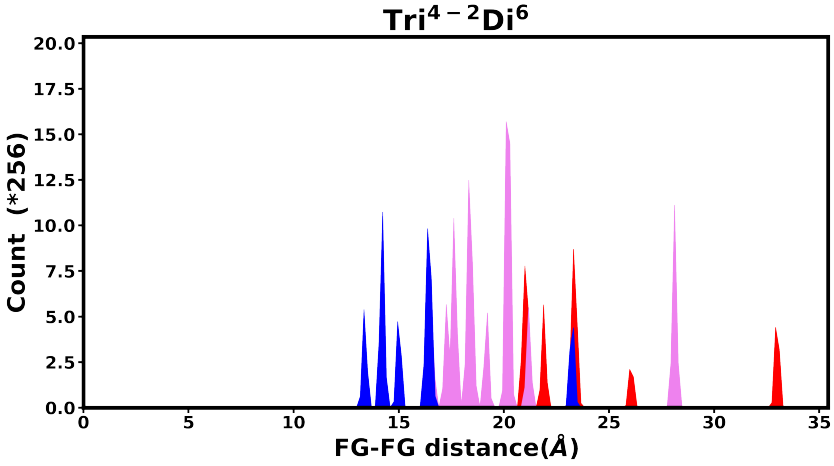
Figure S21: KDE applied on FG-FG distance histograms of pore Tri^4Di^6, constructed from a node with radius 5Å and two different linker size, (a) benzene-1,4-dicarboxylic acid, (b) poly(1,4-benzenedicarboxylic acid) outer functional group positions (c) poly(1,4-benzenedicarboxylic acid) inner functional group positions.



(a) Benzene-1,4-dicarboxylic acid linker

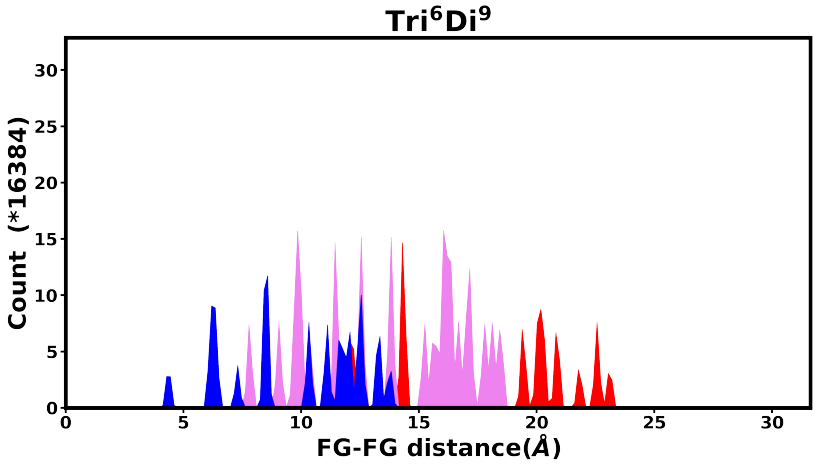


(b) Outer poly(1,4-benzenedicarboxylic acid) linker

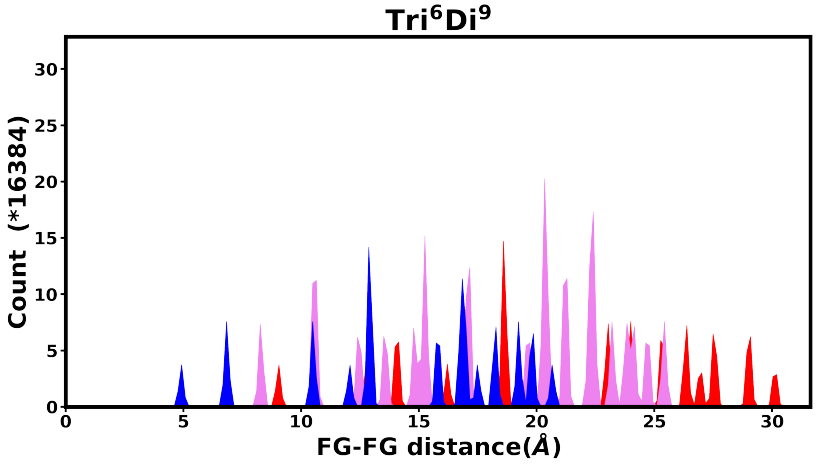


(c) Inner poly(1,4-benzenedicarboxylic acid) linker

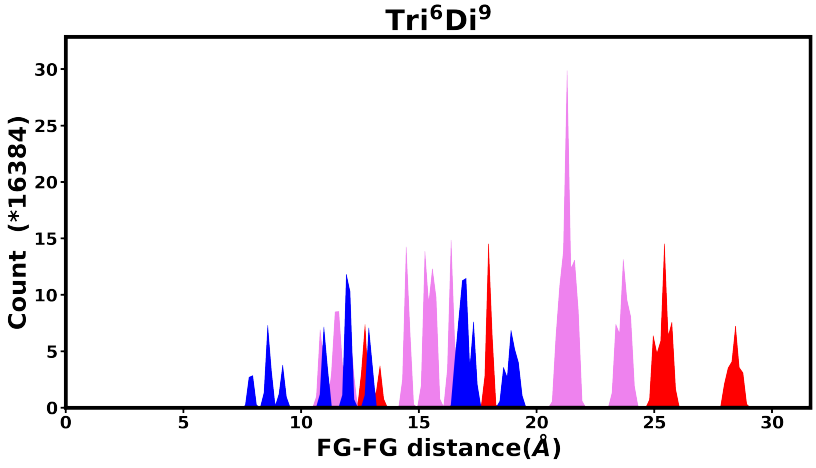
Figure S22: KDE applied on FG-FG distance histograms of pore $Tri_2^4Di^6$, constructed from a node with radius 5Å and two different linker size, (a) benzene-1,4-dicarboxylic acid, (b) poly(1,4-benzenedicarboxylic acid) outer functional group positions (c) poly(1,4-benzenedicarboxylic acid) inner functional group positions.



(a) Benzene-1,4-dicarboxylic acid linker

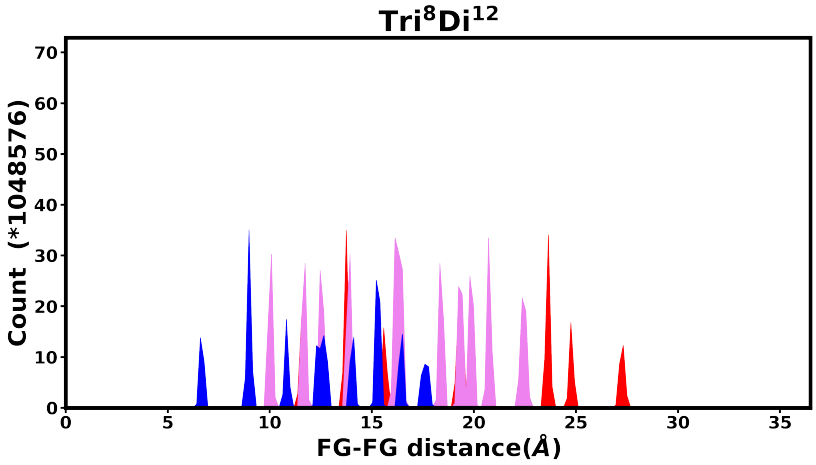


(b) Outer poly(1,4-benzenedicarboxylic acid) linker

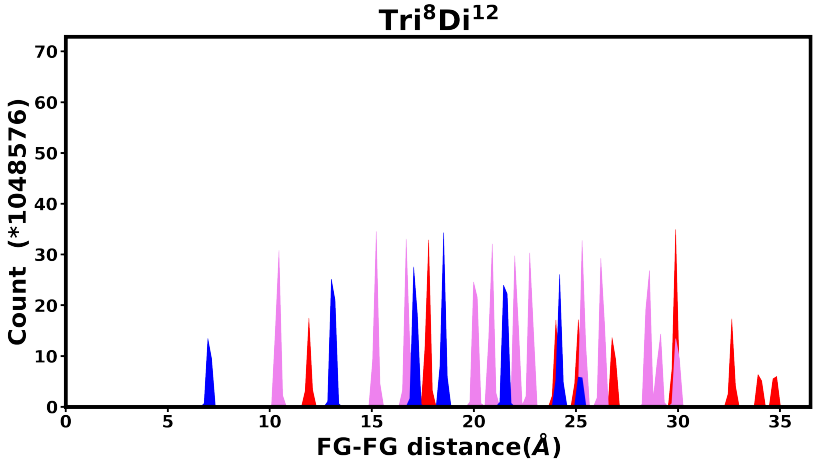


(c) Inner poly(1,4-benzenedicarboxylic acid) linker

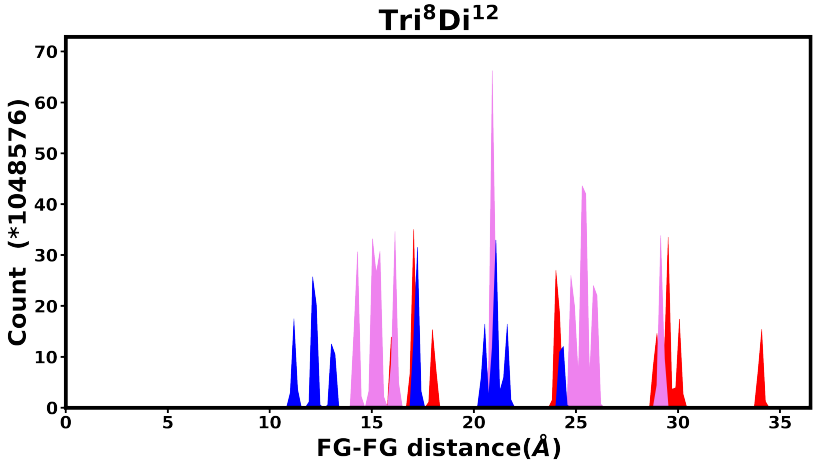
Figure S23: KDE applied on FG-FG distance histograms of pore Tri^6Di^9, constructed from a node with radius 5Å and two different linker sizes, (a) benzene-1,4-dicarboxylic acid, (b) poly(1,4-benzenedicarboxylic acid) outer functional group positions (c) poly(1,4-benzenedicarboxylic acid) inner functional group positions.



(a) Benzene-1,4-dicarboxylic acid linker

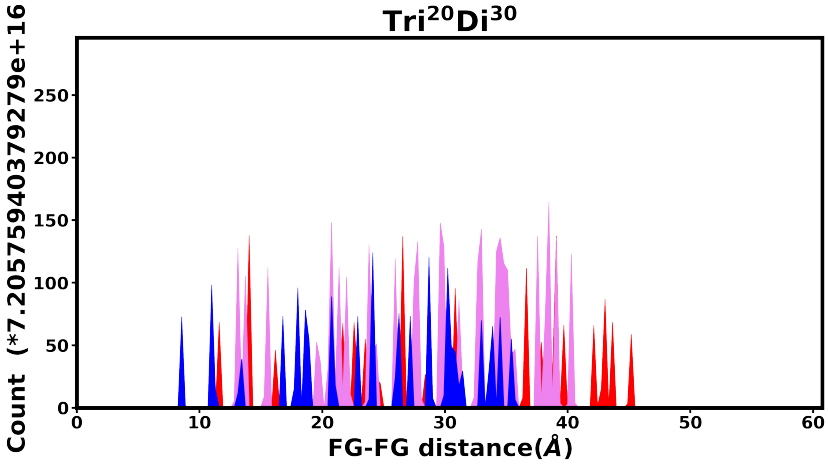


(b) Outer poly(1,4-benzenedicarboxylic acid) linker

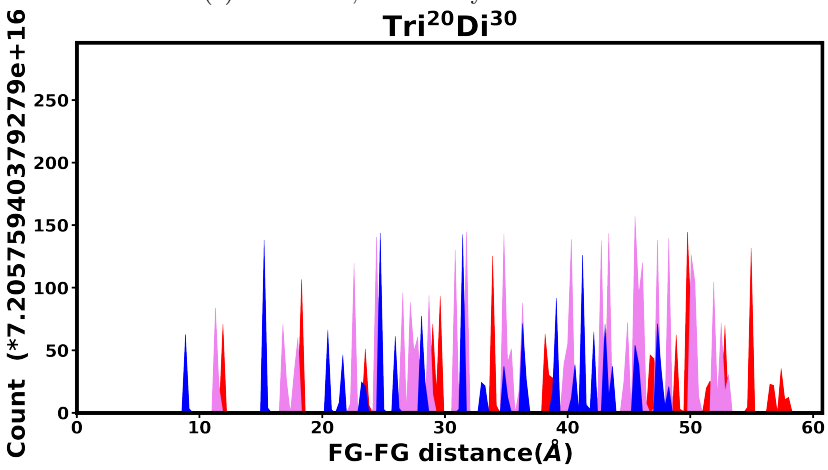


(c) Inner poly(1,4-benzenedicarboxylic acid) linker

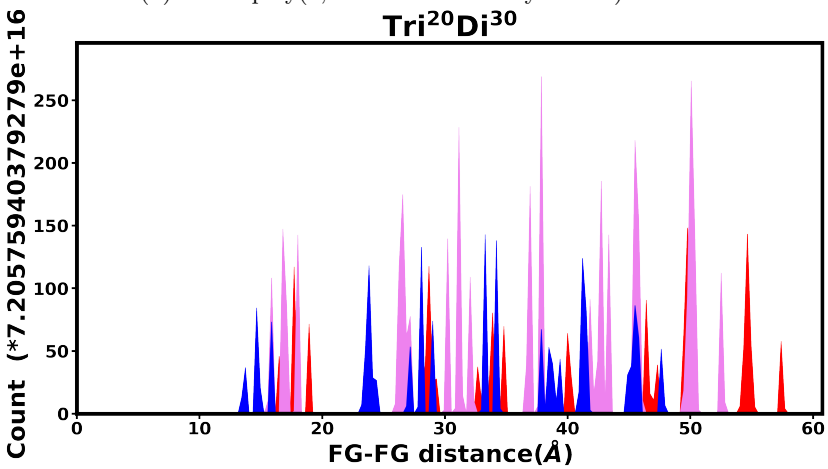
Figure S24: KDE applied on FG-FG distance histograms of pore Tri^8Di^{12}, constructed from a node with radius 5Å and two different linker sizes, (a) benzene-1,4-dicarboxylic acid, (b) poly(1,4-benzenedicarboxylic acid) outer functional group positions (c) poly(1,4-benzenedicarboxylic acid) inner functional group positions.



(a) Benzene-1,4-dicarboxylic acid linker

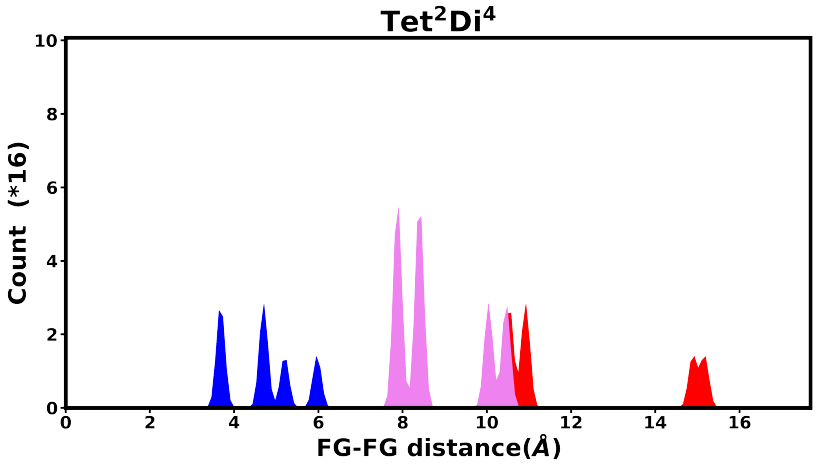


(b) Outer poly(1,4-benzenedicarboxylic acid) linker

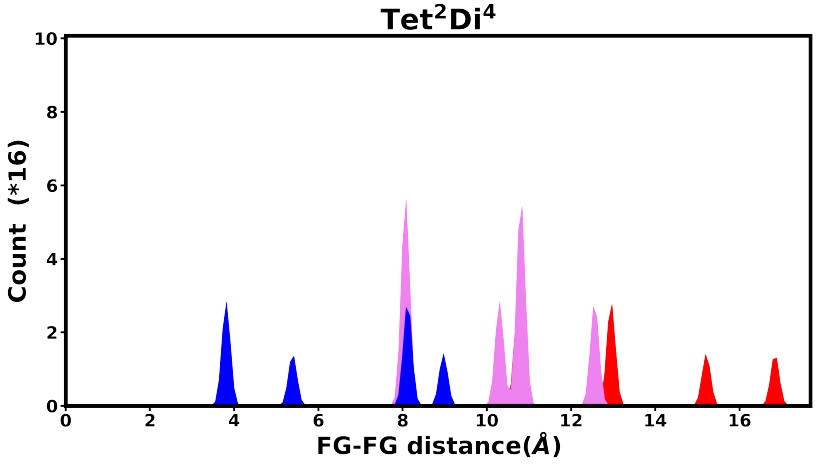


(c) Inner poly(1,4-benzenedicarboxylic acid) linker

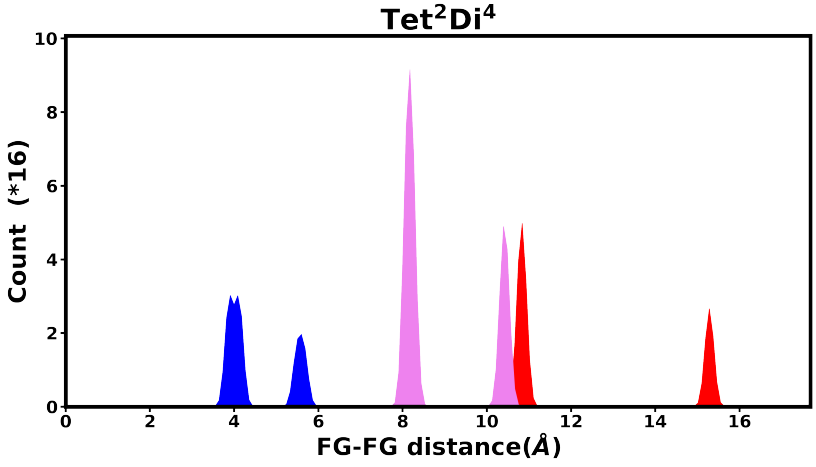
Figure S25: KDE applied on FG-FG distance histograms of pore $Tri^{20}Di^{30}$, constructed from a node with radius 5Å and two different linker sizes, (a) benzene-1,4-dicarboxylic acid, (b) poly(1,4-benzenedicarboxylic acid) outer functional group positions (c) poly(1,4-benzenedicarboxylic acid) inner functional group positions.



(a) Benzene-1,4-dicarboxylic acid linker

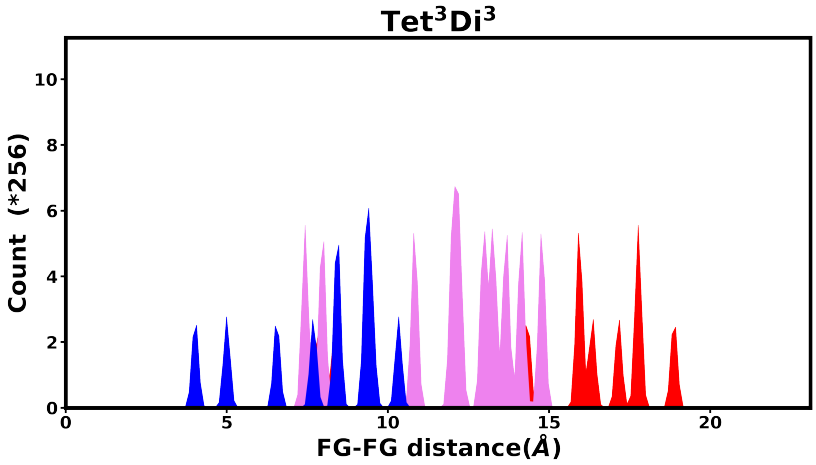


(b) Outer poly(1,4-benzenedicarboxylic acid) linker

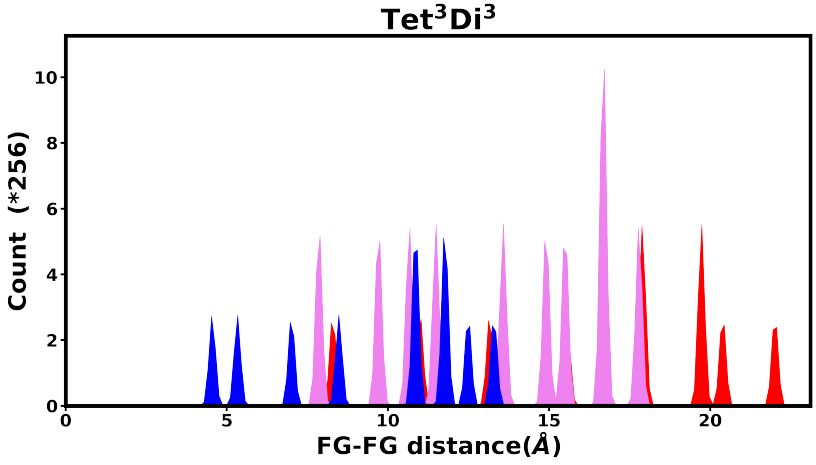


(c) Inner poly(1,4-benzenedicarboxylic acid) linker

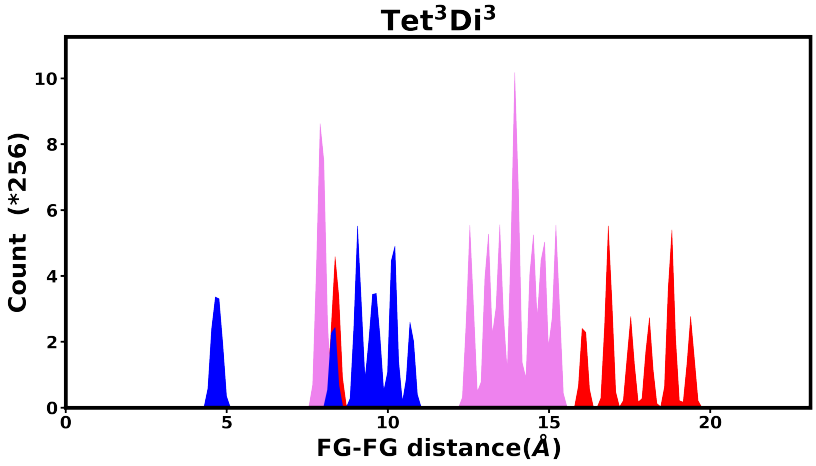
Figure S26: KDE applied on FG-FG distance histograms of pore Tet^2Di^4, constructed from a node with radius 5Å and two different linker sizes, (a) benzene-1,4-dicarboxylic acid, (b) poly(1,4-benzenedicarboxylic acid) outer functional group positions (c) poly(1,4-benzenedicarboxylic acid) inner functional group positions.



(a) Benzene-1,4-dicarboxylic acid linker

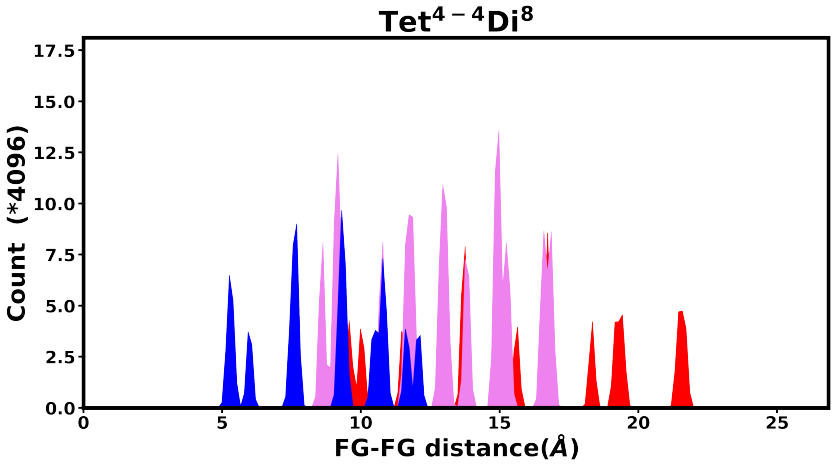


(b) Outer poly(1,4-benzenedicarboxylic acid) linker

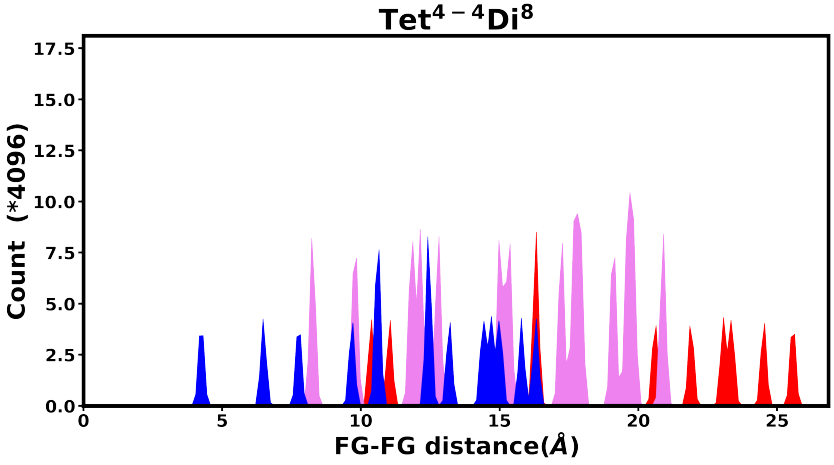


(c) Inner poly(1,4-benzenedicarboxylic acid) linker

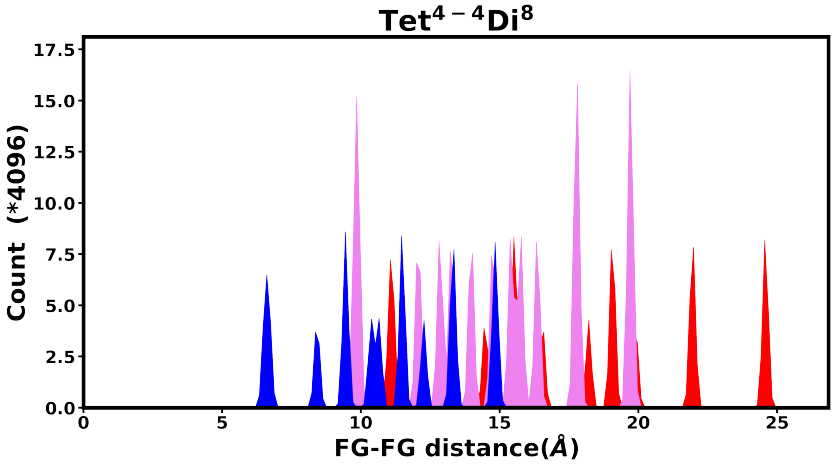
Figure S27: KDE applied on FG-FG distance histograms of pore $Tet_2^3Di^3$, constructed from a node with radius 5Å and two different linker sizes, (a) benzene-1,4-dicarboxylic acid, (b) poly(1,4-benzenedicarboxylic acid) outer functional group positions (c) poly(1,4-benzenedicarboxylic acid) inner functional group positions.



(a) Benzene-1,4-dicarboxylic acid linker

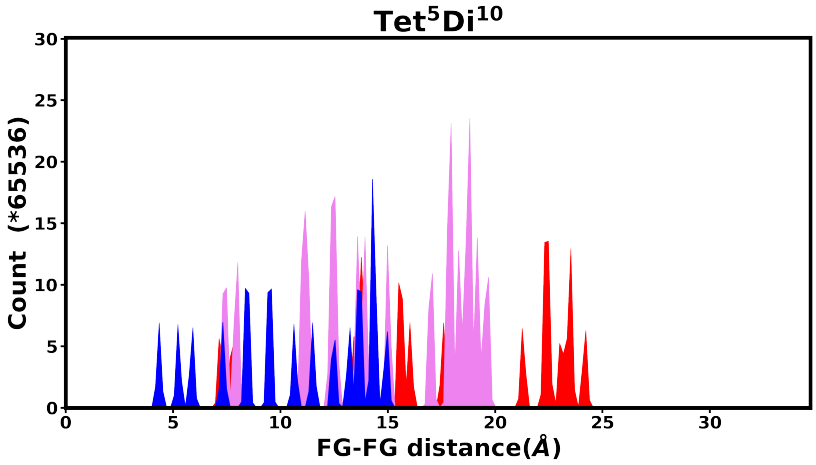


(b) Outer poly(1,4-benzenedicarboxylic acid) linker

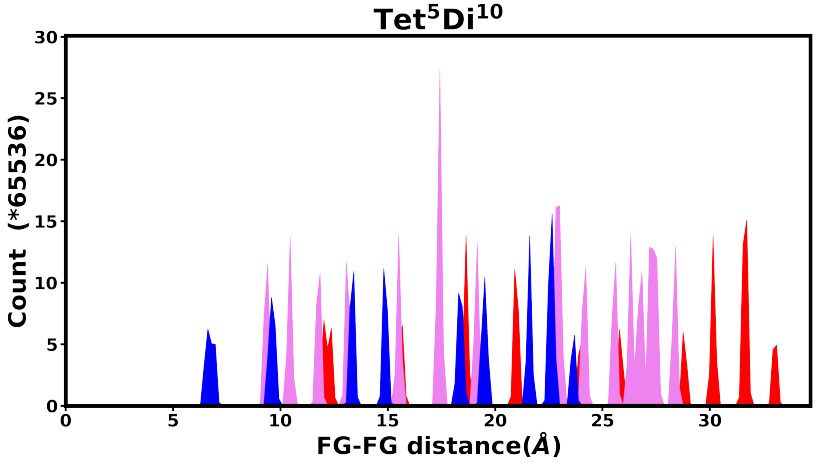


(c) Inner poly(1,4-benzenedicarboxylic acid) linker

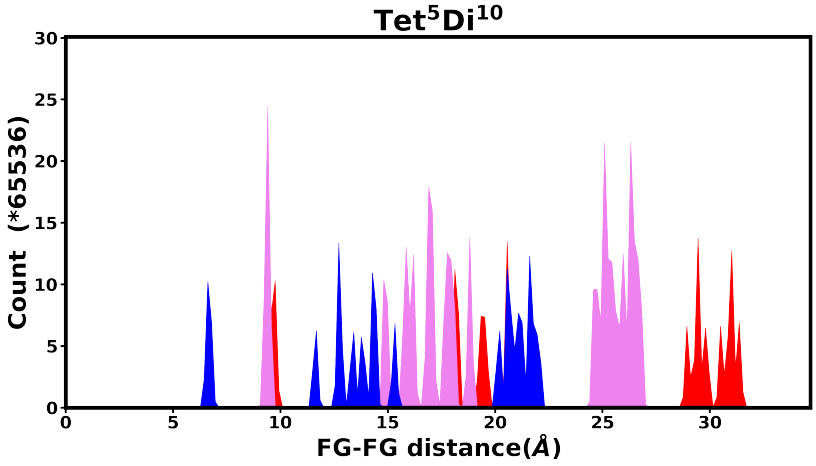
Figure S28: KDE applied on FG-FG distance histograms of pore $Tet_4^4Di^8$, constructed from a node with radius 5Å and two different linker sizes, (a) benzene-1,4-dicarboxylic acid, (b) poly(1,4-benzenedicarboxylic acid) outer functional group positions (c) poly(1,4-benzenedicarboxylic acid) inner functional group positions.



(a) Benzene-1,4-dicarboxylic acid linker

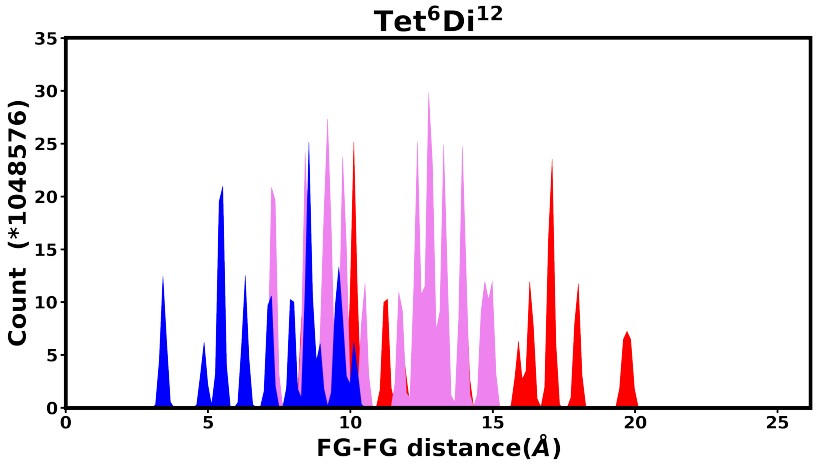


(b) Outer poly(1,4-benzenedicarboxylic acid) linker

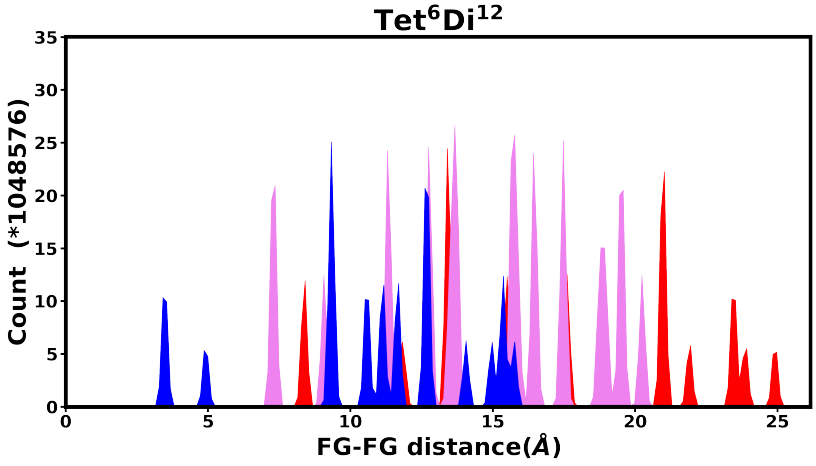


(c) Inner poly(1,4-benzenedicarboxylic acid) linker

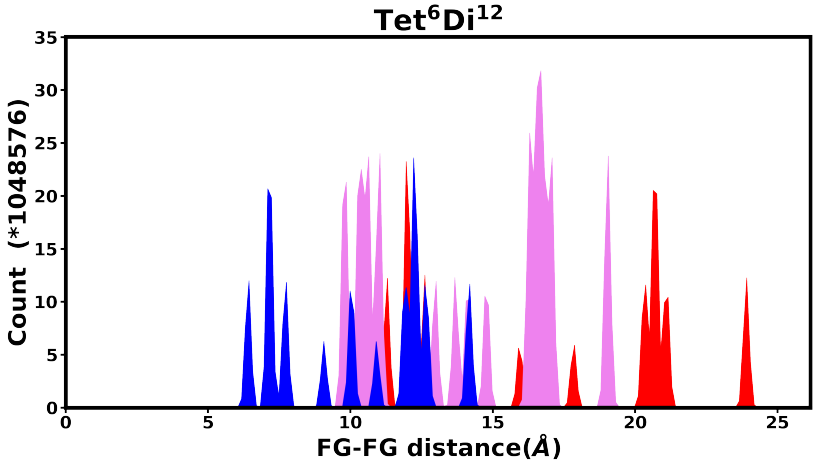
Figure S29: KDE applied on FG-FG distance histograms of pore Tet^5Di^10, constructed from a node with radius 5Å and two different linker sizes, (a) benzene-1,4-dicarboxylic acid, (b) poly(1,4-benzenedicarboxylic acid) outer functional group positions (c) poly(1,4-benzenedicarboxylic acid) inner functional group positions.



(a) Benzene-1,4-dicarboxylic acid linker

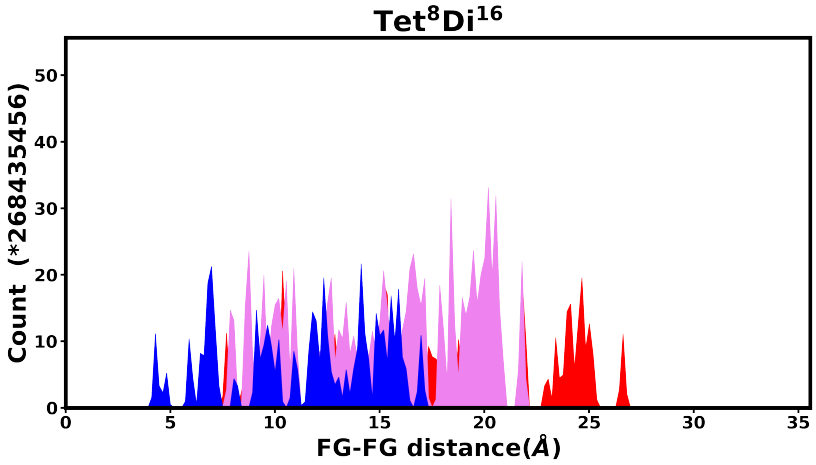


(b) Outer poly(1,4-benzenedicarboxylic acid) linker

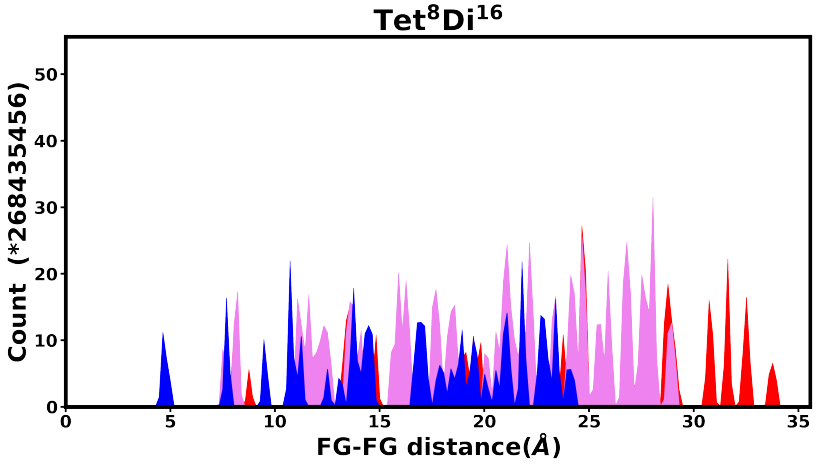


(c) Inner poly(1,4-benzenedicarboxylic acid) linker

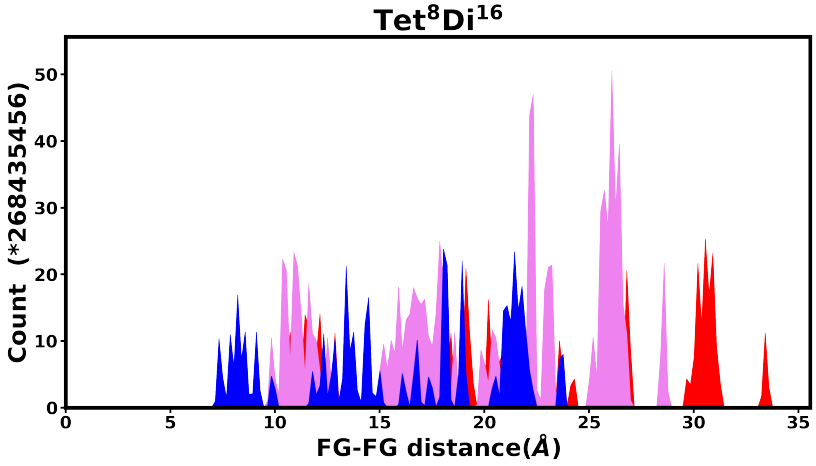
Figure S30: KDE applied on FG-FG distance histograms of pore Tet^6Di^{12}, constructed from a node with radius 5Å and two different linker sizes, (a) benzene-1,4-dicarboxylic acid, (b) poly(1,4-benzenedicarboxylic acid) outer functional group positions (c) poly(1,4-benzenedicarboxylic acid) inner functional group positions.



(a) Benzene-1,4-dicarboxylic acid linker

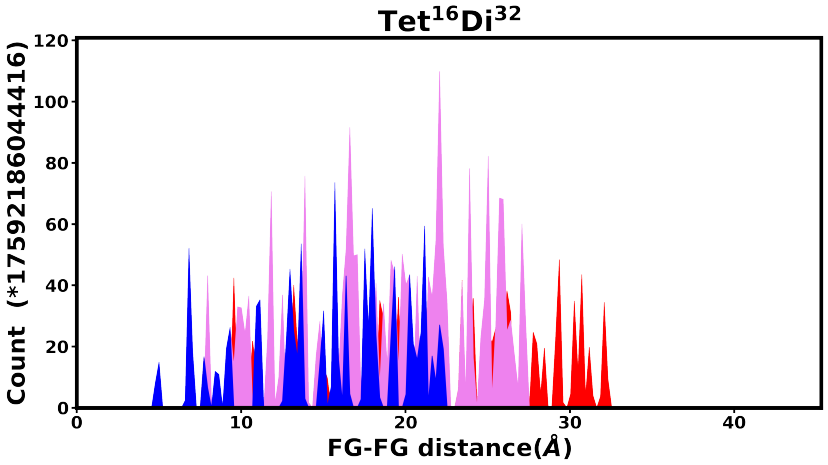


(b) Outer poly(1,4-benzenedicarboxylic acid) linker

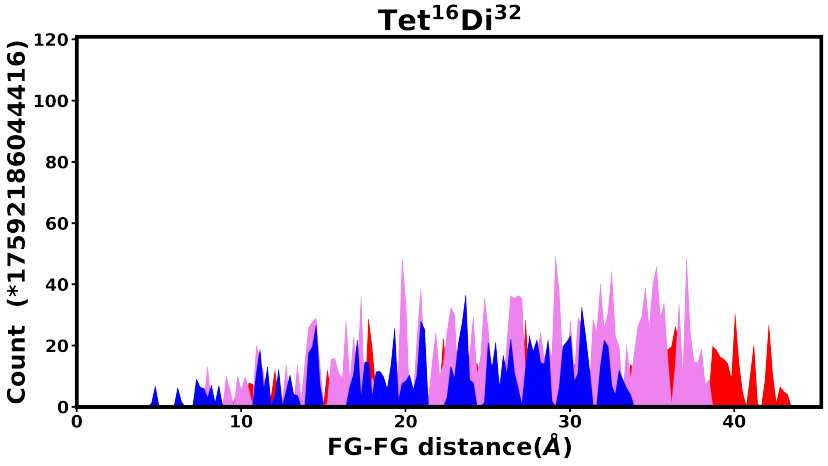


(c) Inner poly(1,4-benzenedicarboxylic acid) linker

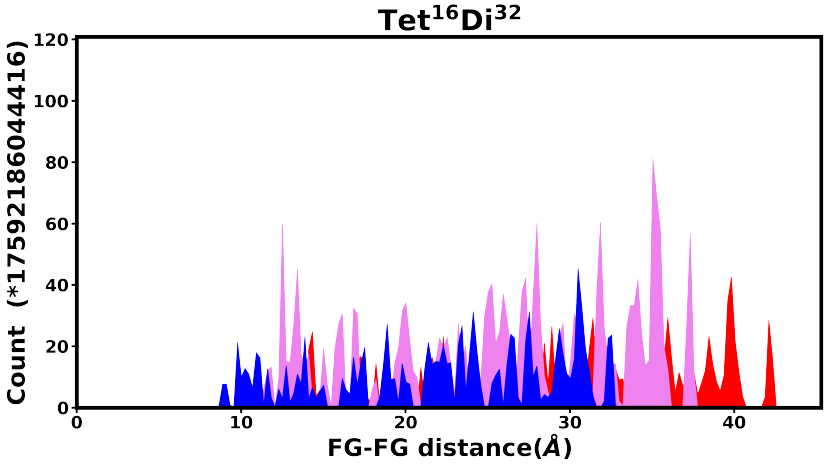
Figure S31: KDE applied on FG-FG distance histograms of pore Tet^8Di^{16}, constructed from a node with radius 5Å and two different linker sizes, (a) benzene-1,4-dicarboxylic acid, (b) poly(1,4-benzenedicarboxylic acid) outer functional group positions (c) poly(1,4-benzenedicarboxylic acid) inner functional group positions.



(a) Benzene-1,4-dicarboxylic acid linker

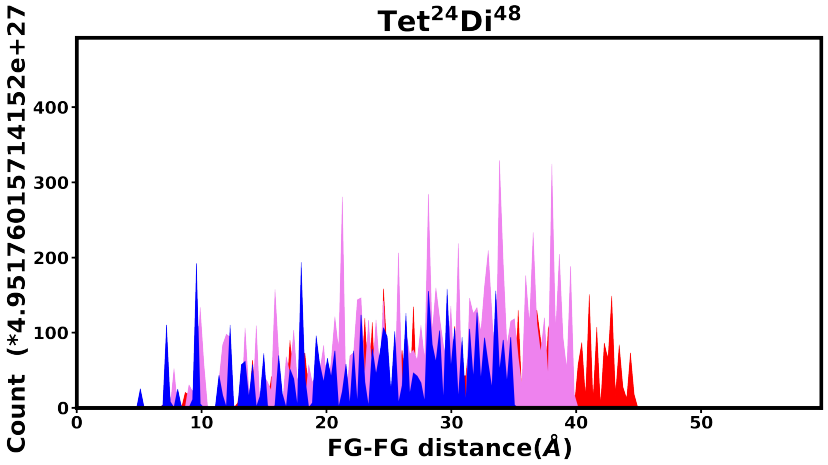


(b) Outer poly(1,4-benzenedicarboxylic acid) linker

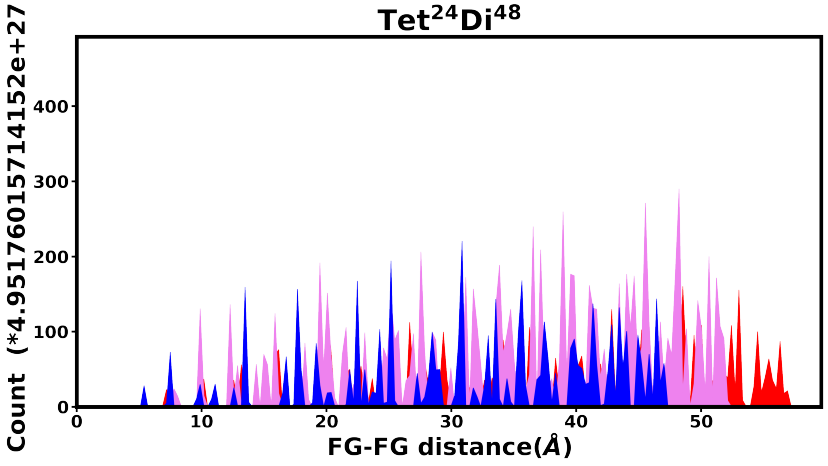


(c) Inner poly(1,4-benzenedicarboxylic acid) linker

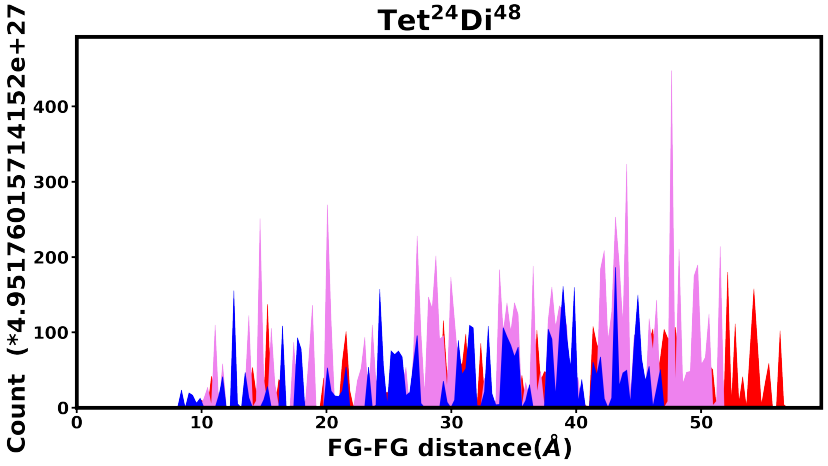
Figure S32: KDE applied on FG-FG distance histograms of pore $Tet^{16}Di^{32}$, constructed from a node with radius 5Å and two different linker sizes, (a) benzene-1,4-dicarboxylic acid, (b) poly(1,4-benzenedicarboxylic acid) outer functional group positions (c) poly(1,4-benzenedicarboxylic acid) inner functional group positions.



(a) Benzene-1,4-dicarboxylic acid linker



(b) Outer poly(1,4-benzenedicarboxylic acid) linker



(c) Inner poly(1,4-benzenedicarboxylic acid) linker

Figure S33: KDE applied on FG-FG distance histograms of pore $Tet^{24}Di^{48}$, constructed from a node with radius 5Å and two different linker sizes, (a) benzene-1,4-dicarboxylic acid, (b) poly(1,4-benzenedicarboxylic acid) outer functional group positions (c) poly(1,4-benzenedicarboxylic acid) inner functional group positions.

S9 Principal Component Analysis

S10 Machine Learning

S11 Machine Learning