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

Functional Group Isomerism in Porous Molecular Materials: Describing Everything, Everywhere, All At Once

Predicting something (such as an adsorbate binding energy) for a 0-3D porous material typically involves the selection and modelling of a single isomer. Depending on the size and complexity of the adsorbate, the isomer(s) chosen may not be ideal or representative. Therefore, we would like to model adsorption for all structures in a topology (e.g. all possible isomers of a tetrahedral pore). This is our motivation for this work. If we consider two cases:

Case 1: 3D-periodic MOFs/COFs

From the crystal structure (Refcode: RUBTAK), the chemical formula for UiO-66 is $C_{48}H_{28}O_{32}Zr_6$ and Z=4, so the molecular weight of one unit cell of UiO-66 is 6656.24 g/mol

$\therefore$ 1g of UiO-66 = 1.5023 x 10^{-4} mol

$\therefore$ 9.047 x 10^{19} unit cells / g.

24 linkers per unit cell:

$\therefore$ 2.171 x 10^{21} linkers / g

Each UiO-66 unit cell contains 4 octahedral pores (which can be generated by coordinates (0.5, 0.5, 0.5)) and 8 tetrahedral pores (coordinates (0.25, 0.25, 0.25)).

This means that in 1g of UiO-66, there are:

4 * 9.047 x 10^{19} = 3.169 x 10^{20} octahedral pores

And 8 * 9.047 x 10^{19} = 7.238 x 10^{20} tetrahedral pores

Dividing these (very large!) numbers by the number of isomers of the octahedral (351976) and tetrahedral pores (176) gives the number of times we'd expect to see each pore isomer occurring in 1g of UiO-66:

= 1.028 x 10^{15} times for the octahedral pore

= 4.112 x 10^{18} times for the tetrahedral pore

So in 1g of material, which seems like a fairly modest amount if we are considering an actual application, our guest molecule / adsorbate will see **every** pore isomer **many, many** times and therefore to model adsorption of some molecule in UiO-66 requires somehow accessing the "average" pore, somehow summed over every possible pore isomer – which is what we attempt here.

- Quick order-of-magnitude check / aside / reference: The original Lillerud UiO-66 paper (DOI:10.1021/ja8057953) used 0.227 mmol of both BDC and $ZrCl_4$ in their synthesis vs 0.15 mmol of MOF assumed above.)

Case 2: 0D-periodic cages (MOPs/POCs)

The analysis above applies similarly to 0D cages, with the following caveats:

- For the single and simple type of isomerism considered here (i.e. a linker flip), it is hard to see any entropic reason to end up with only one isomer, especially if the cages were in dilute solution.
- In a crystallisation process there could be a possibility of say, forcing (e.g. the limiting cases of) an "all FG in" or "all FG out" isomer, but this would still leave two choices (out of the original four) for each functional group, and therefore a significant (topology dependent) number of isomers w.r.t the inside of the pore/cage, even where the outside is fully defined.

We intend our descriptor to apply equally whether the pore is discrete (i.e. a MOP/POC) or part of a network (i.e. a MOF/COF), and also regardless of the identity of the connector.

Uniqueness:

The *x* (e.g. for tetrahedral 176) individual histograms are, by definition, unique – if every FG-FG distance was the same in two isomers *n* and *m*, then some symmetry operation(s) would map *n* onto *m*.

For most applications (e.g. adsorption) it is probably the blue / In-In part of our histograms, describing the environment strictly inside the pore, that is most relevant.

We have chosen not to delete the violet (In-Out) and red (Out-Out) parts of our histograms. By so doing, we are in essence leaving the machine learning algorithm to decide what is important. We think this is the safest choice because at this point, we have not defined our chemical use case:

- If the chemical use case involves 0D cages (POCs/MOPs), the In-In histogram should indeed define the pore for an adsorption problem, but if you were considering crystallization or aggregation as well, then the "out" parts may become important
- For the 3D (COF/MOF) case, it is arguably more subtle: We can imagine that, for example, UiO-66 could be represented as the (possibly weighted) sum of tetrahedral (Tri4Di6) and octahedral (Tet6Di12) histograms, but we know that each pore is not independent of its neighbours (i.e. a FG pointing in to one pore, must be out of the neighbouring pore). Given the incredibly large numbers of pores however (i.e. ~ 10^{20} for 1g of material, worked above), we contend that this *local* effect can be ignored.

If we did decide to consider the In-In FG-FG distances only... then it is clear that there are some number of isomers (when considering all FG), that yield the same inner pore structure – which is probably what a guest molecule would consider to be the "same environment".

To illustrate, and to test whether the histogram defines the environment, we've worked this through in full for a D_{4h} square, deliberately ignoring symmetry / point group operations and exhaustively building up each isomer and from there testing uniqueness. The python code (as a jupyter notebook), and the output (data as xlsx, structures as xyz and figures/tables as pdf) are included in this SI.

In brief:
One substitution of each of the four linkers, yields 4^4 = 256 possible isomers, of which 39 are symmetry unique. If one considers only the inside pore environment, then this number reduces to 15.

Number of FG inside pore	Total symmetry-unique isomers	Unique inner-pore isomers
0	4	1
1	8	1
2	15	5
3	8	4
4	4	4

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

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.

Pore topology	Number of Isomers
Tri^2Di^3	10
Tri^4Di^6	176
$Tri^4_2Di^6$	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

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 .

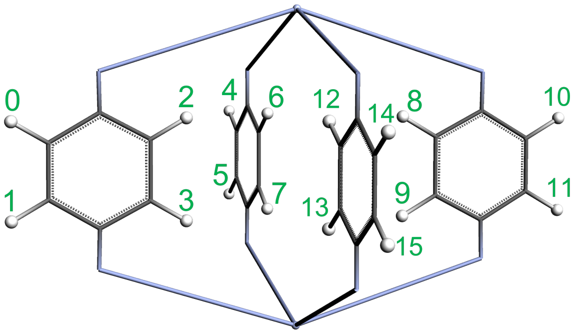
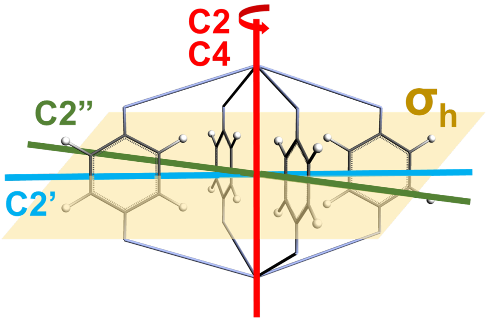
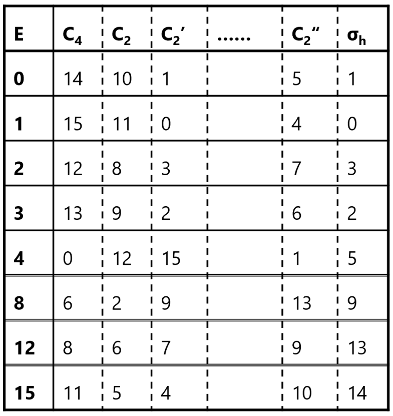


Figure S1: Pore Tet^2Di^4 with assigned index for each functional group, functional group slot is shown in blue circles.

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

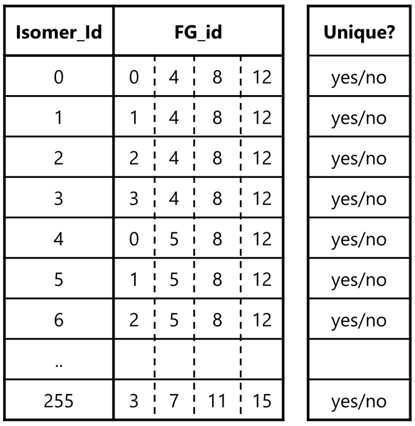
Figure S2: (right) Symmetry elements of pore Tet^2Di^4 (only 5 from the 16 are shown) (left) Transformation library, library of each functional group index and the transformed index, the first column represents the initial position, each subsequent column represent a specified symmetry operation, double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

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

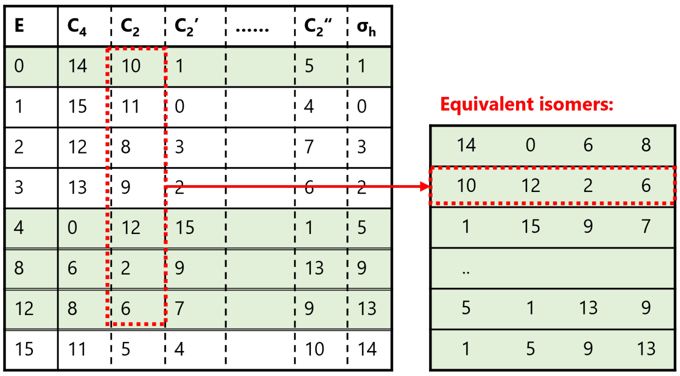


Isomer_Id	FG_id				Unique?
0	0	4	8	12	yes/no
1	1	4	8	12	yes/no
2	2	4	8	12	yes/no
3	3	4	8	12	yes/no
4	0	5	8	12	yes/no
5	1	5	8	12	yes/no
6	2	5	8	12	yes/no
..					
255	3	7	11	15	yes/no

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

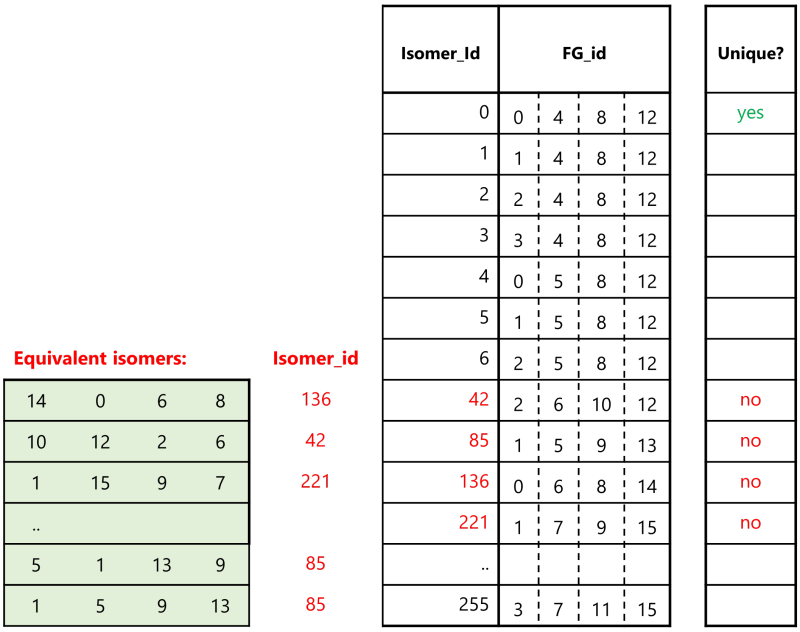
Equivalent isomers:

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.



Equivalent isomers:

				Isomer_id
14	0	6	8	136
10	12	2	6	42
1	15	9	7	221
..				
5	1	13	9	85
1	5	9	13	85

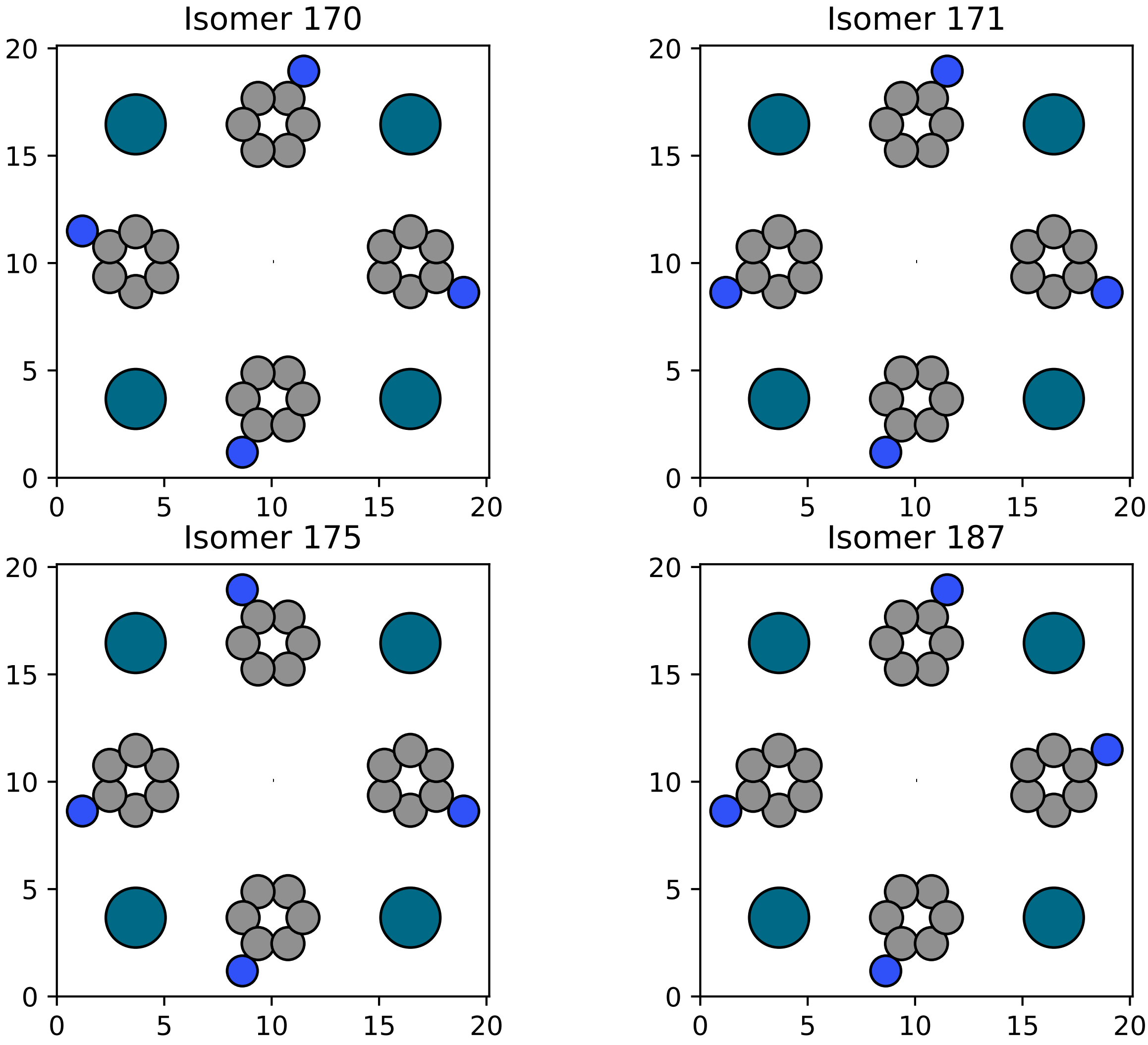
Isomer_Id	FG_id				Unique?
0	0	4	8	12	yes
1	1	4	8	12	
2	2	4	8	12	
3	3	4	8	12	
4	0	5	8	12	
5	1	5	8	12	
6	2	5	8	12	
42	2	6	10	12	no
85	1	5	9	13	no
136	0	6	8	14	no
221	1	7	9	15	no
..					
255	3	7	11	15	

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

Number of FG in pore: 0



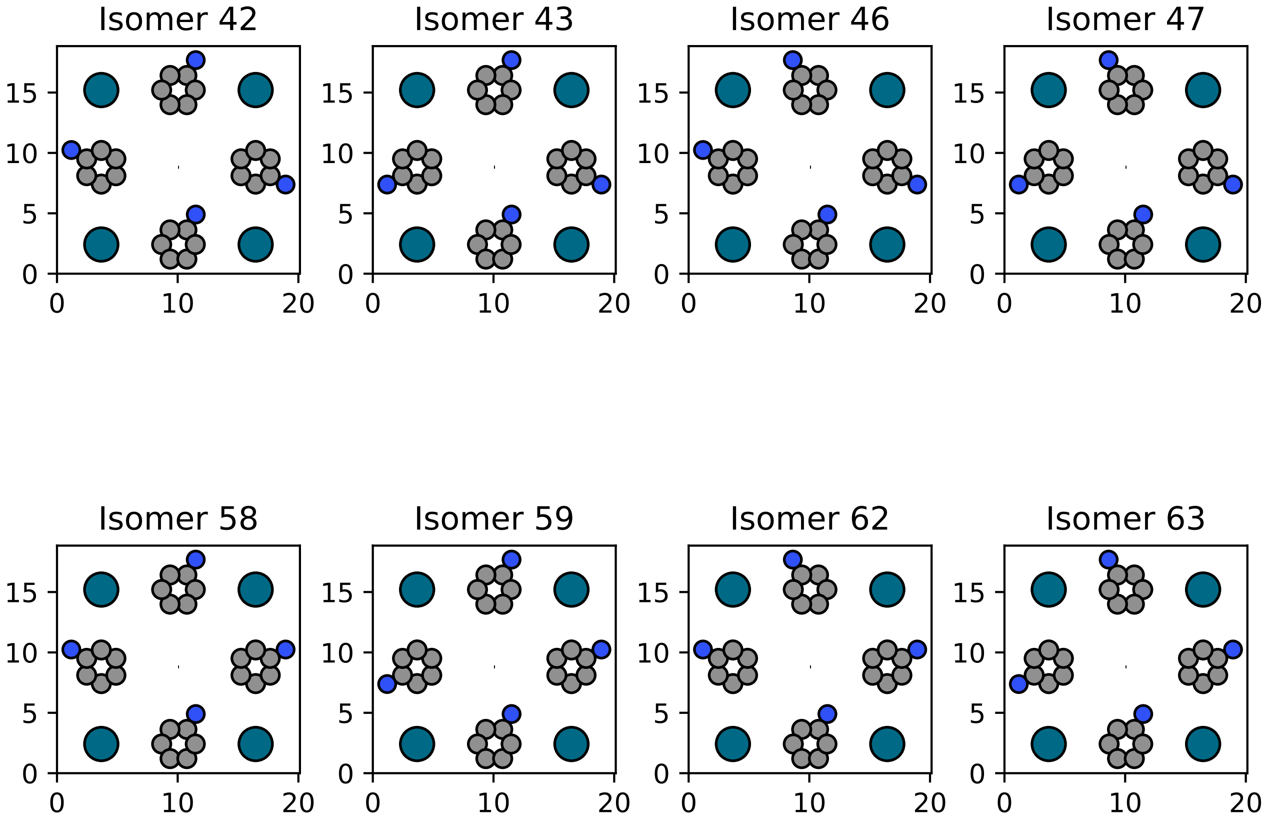
Isomer 170

Isomer 171

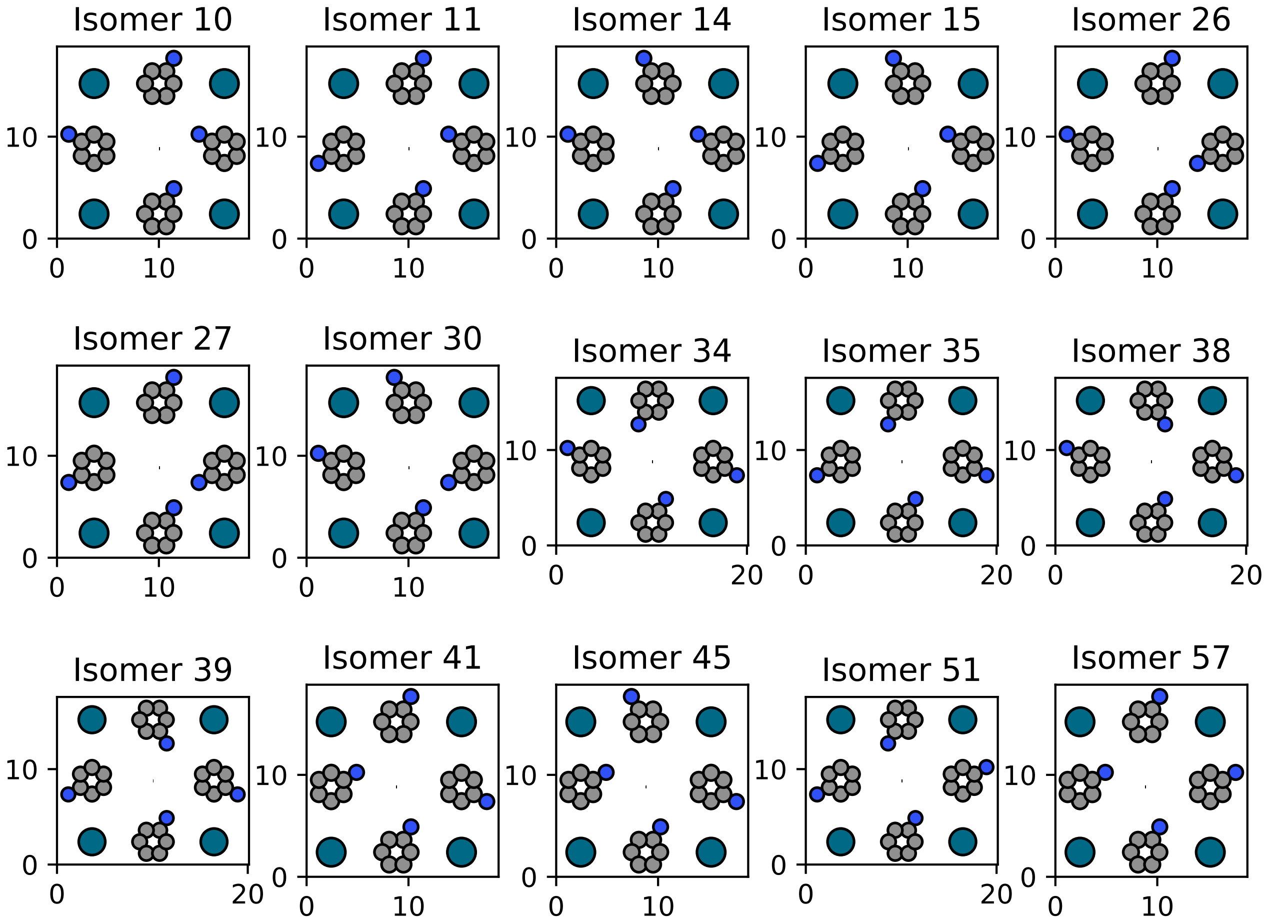
Isomer 175

Isomer 187

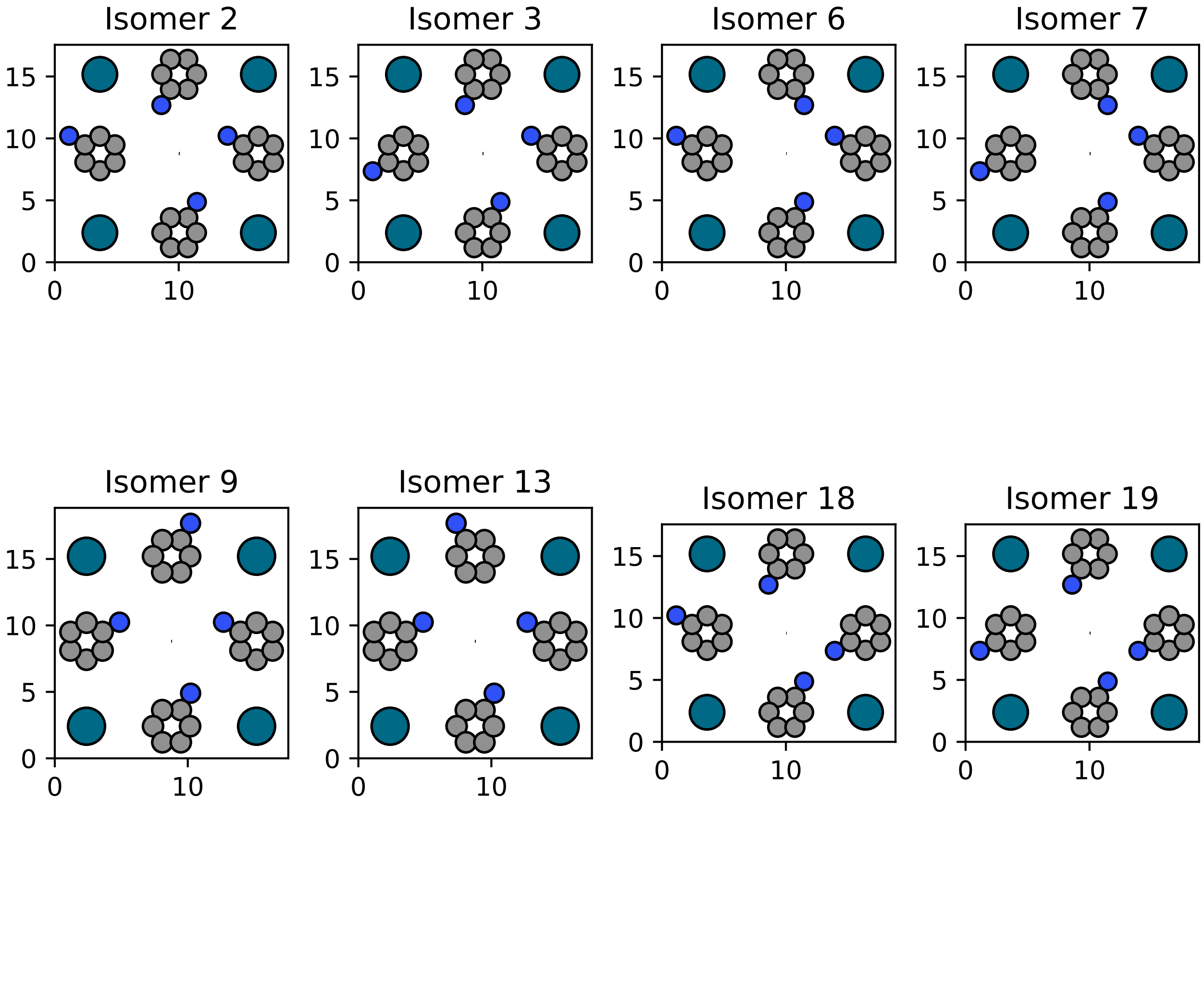
Number of FG in pore: 1



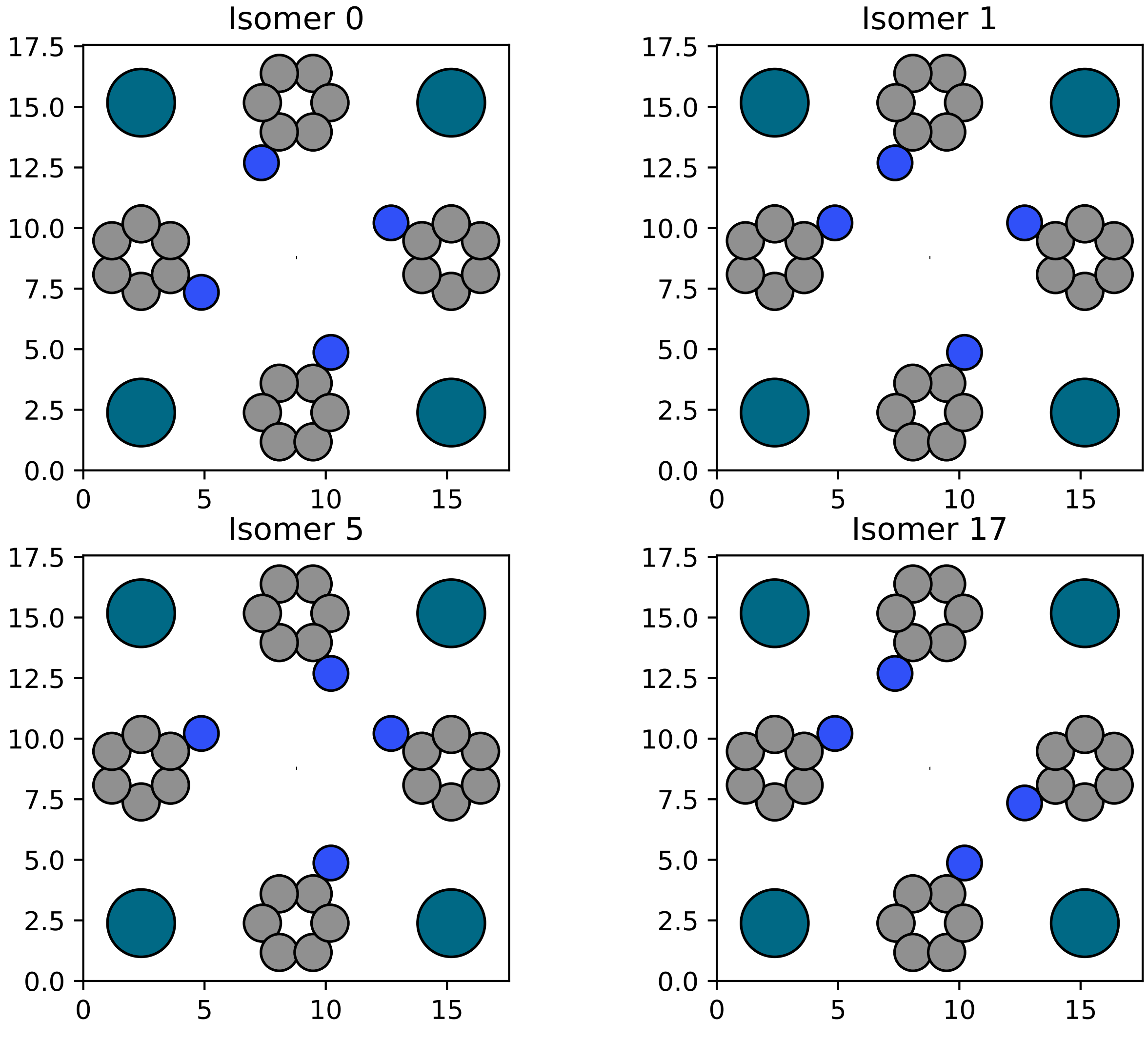
Number of FG in pore: 2



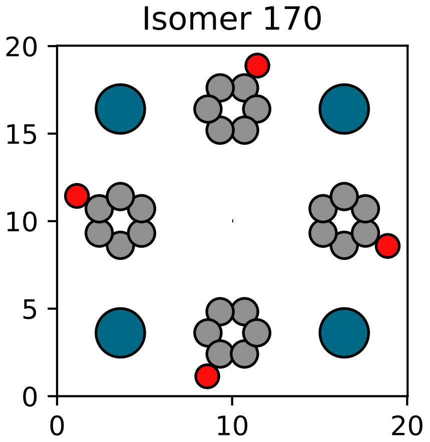
Number of FG in pore: 3



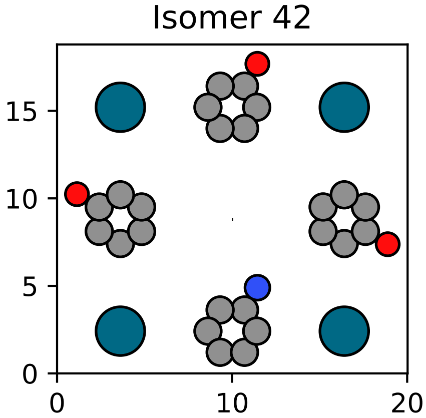
Number of FG in pore: 4



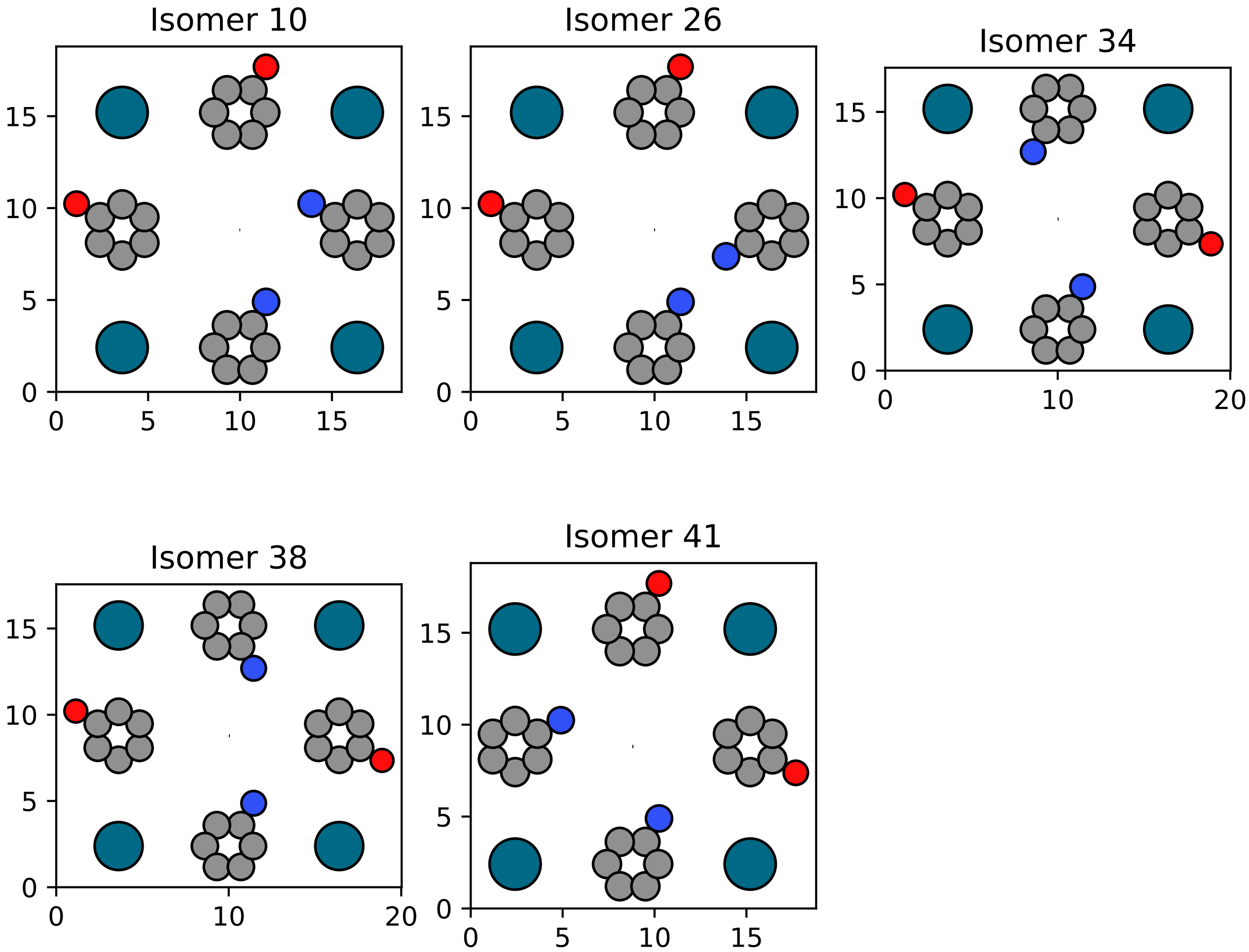
Number of FG in pore: 0



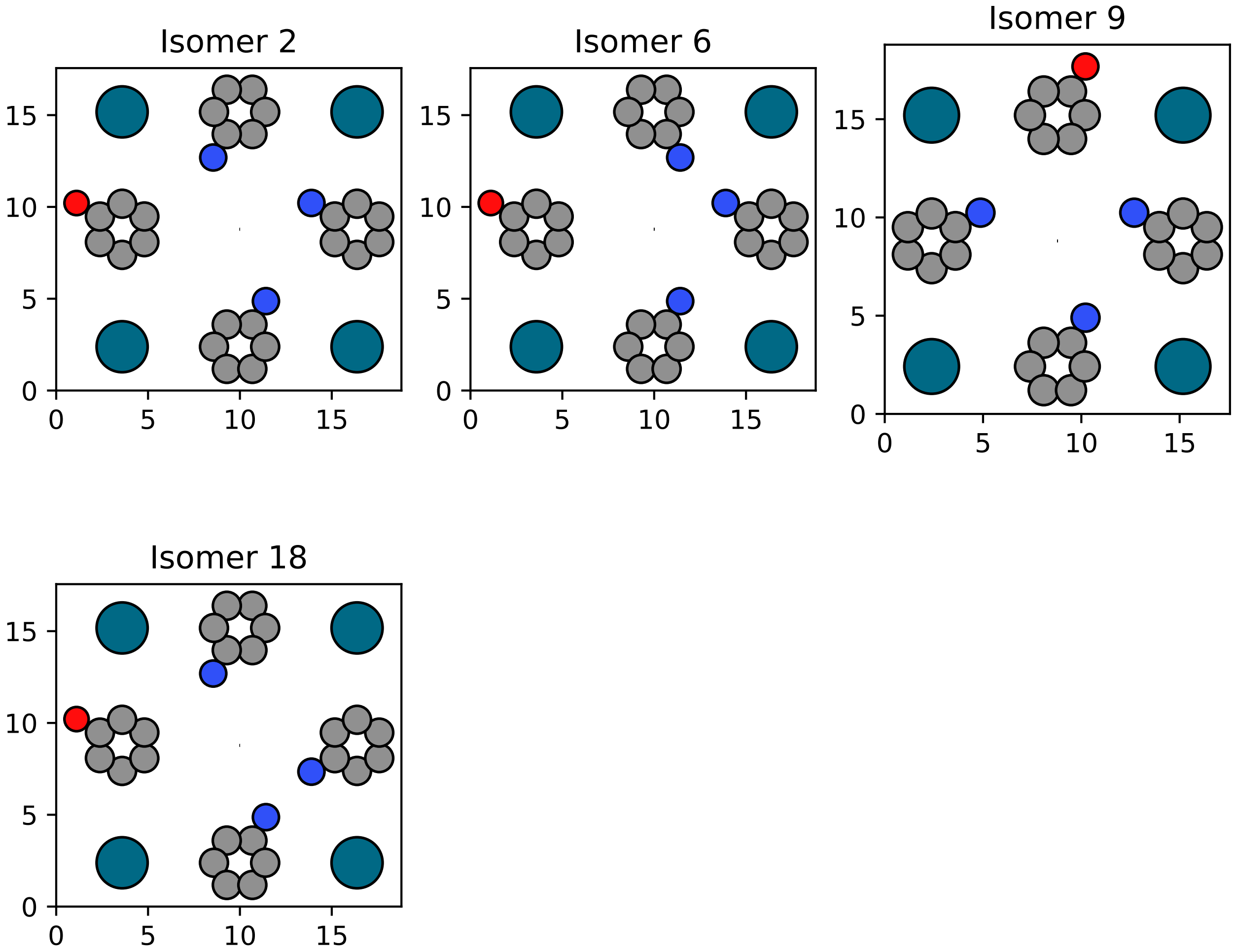
Number of FG in pore: 1



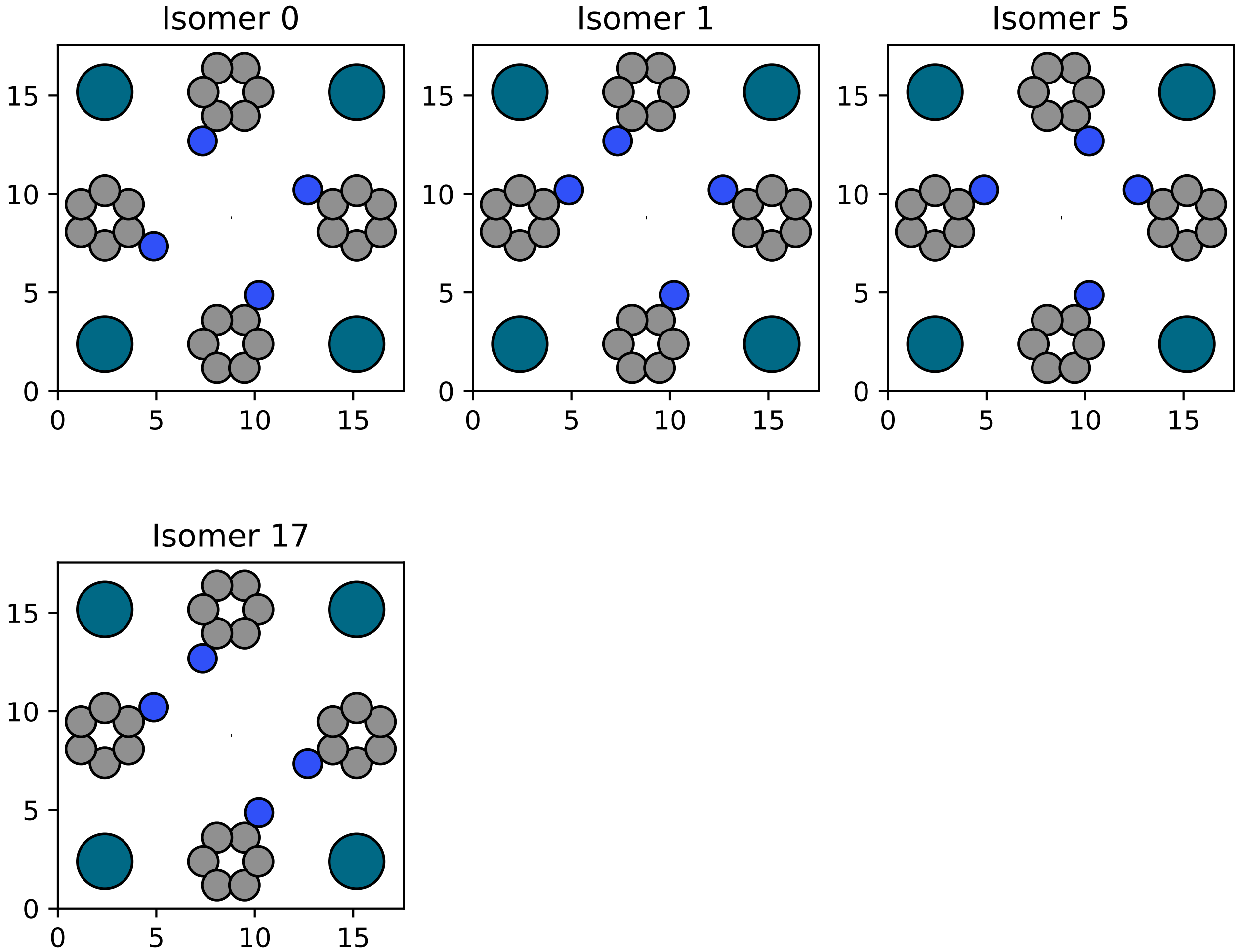
Number of FG in pore: 2



Number of FG in pore: 3



Number of FG in pore: 4


Isomer 0

Isomer 1

Isomer 5

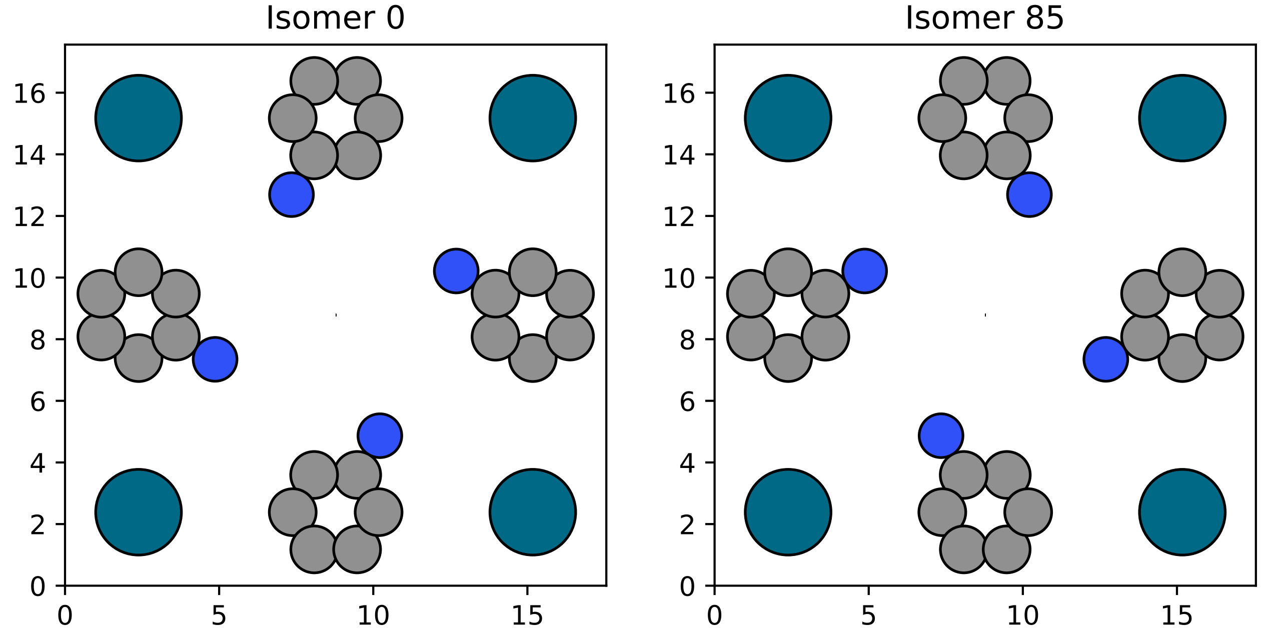
Isomer 17

Isomer number 0

Number of FG in pore: 4

Isomer FG-FG distances (Å):

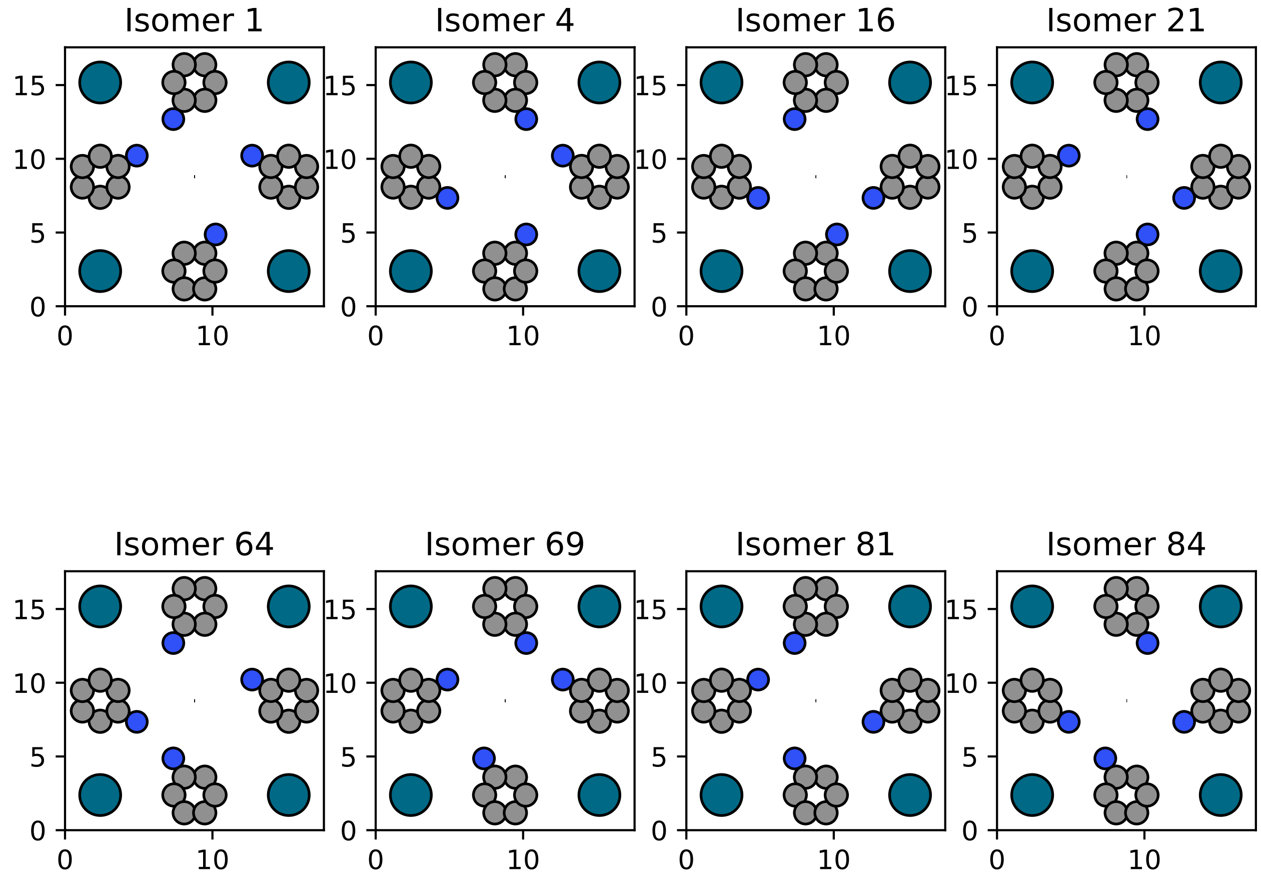
5.8919
8.3324
5.8919
5.8919
8.3324
5.8919



Isomer number 1

Number of FG in pore: 4
Isomer FG-FG distances (Å):

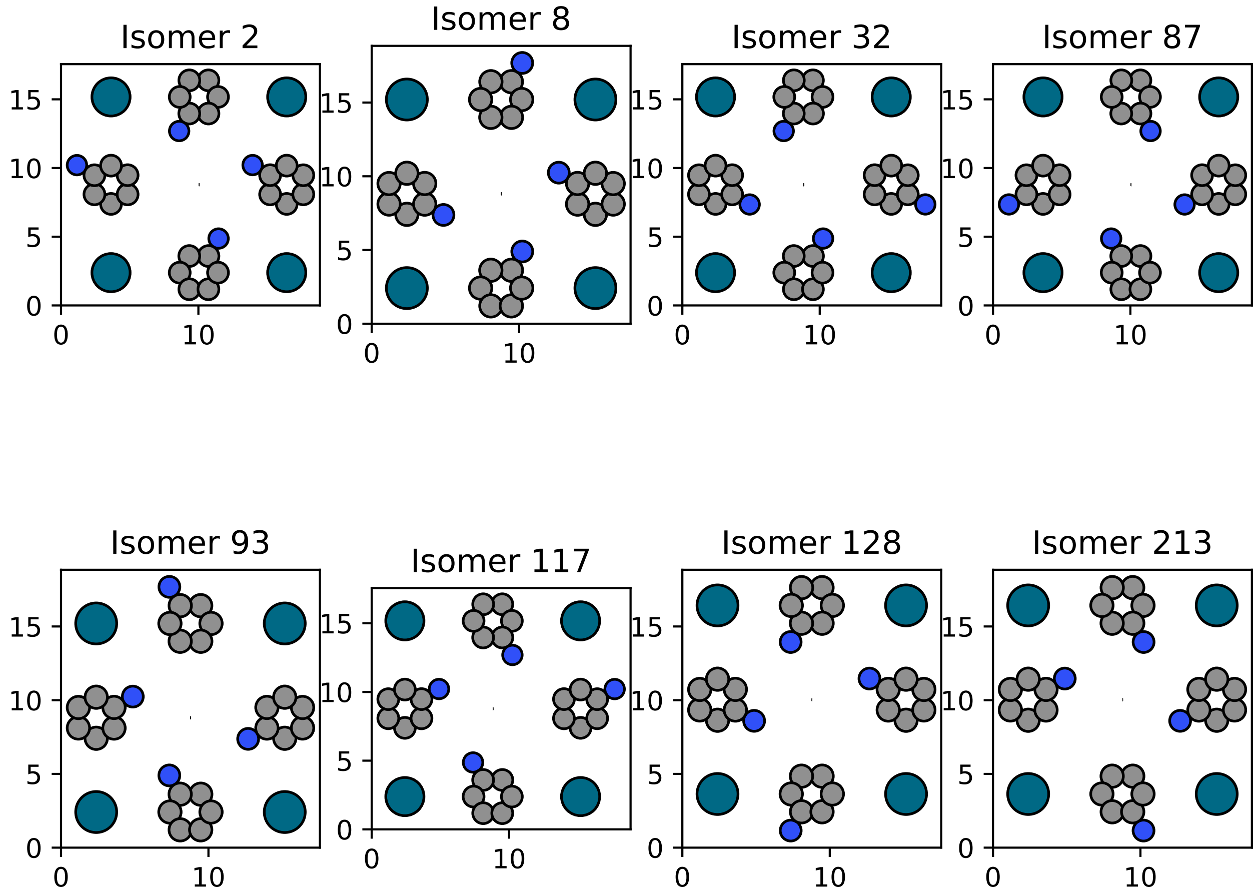
5.8919
8.3324
7.5595
5.8919
7.8236
3.5048



Isomer number 2

Number of FG in pore: 3
Isomer FG-FG distances (Å):

5.8919
8.3324
11.6104
5.8919
12.7870
7.8447

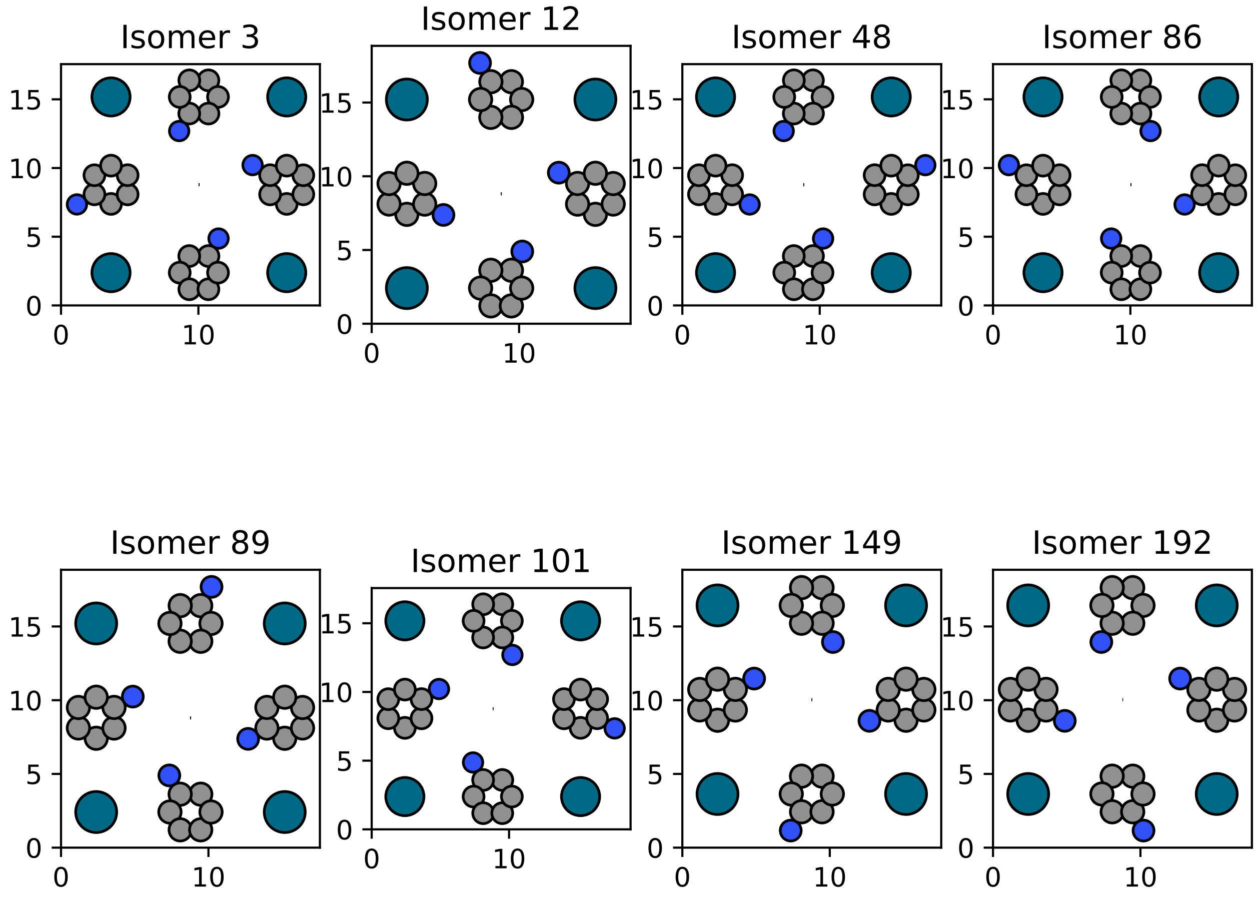


Isomer number 3

Number of FG in pore: 3

Isomer FG-FG distances (Å):

5.8919
8.3324
10.6034
5.8919
13.1037
9.1602

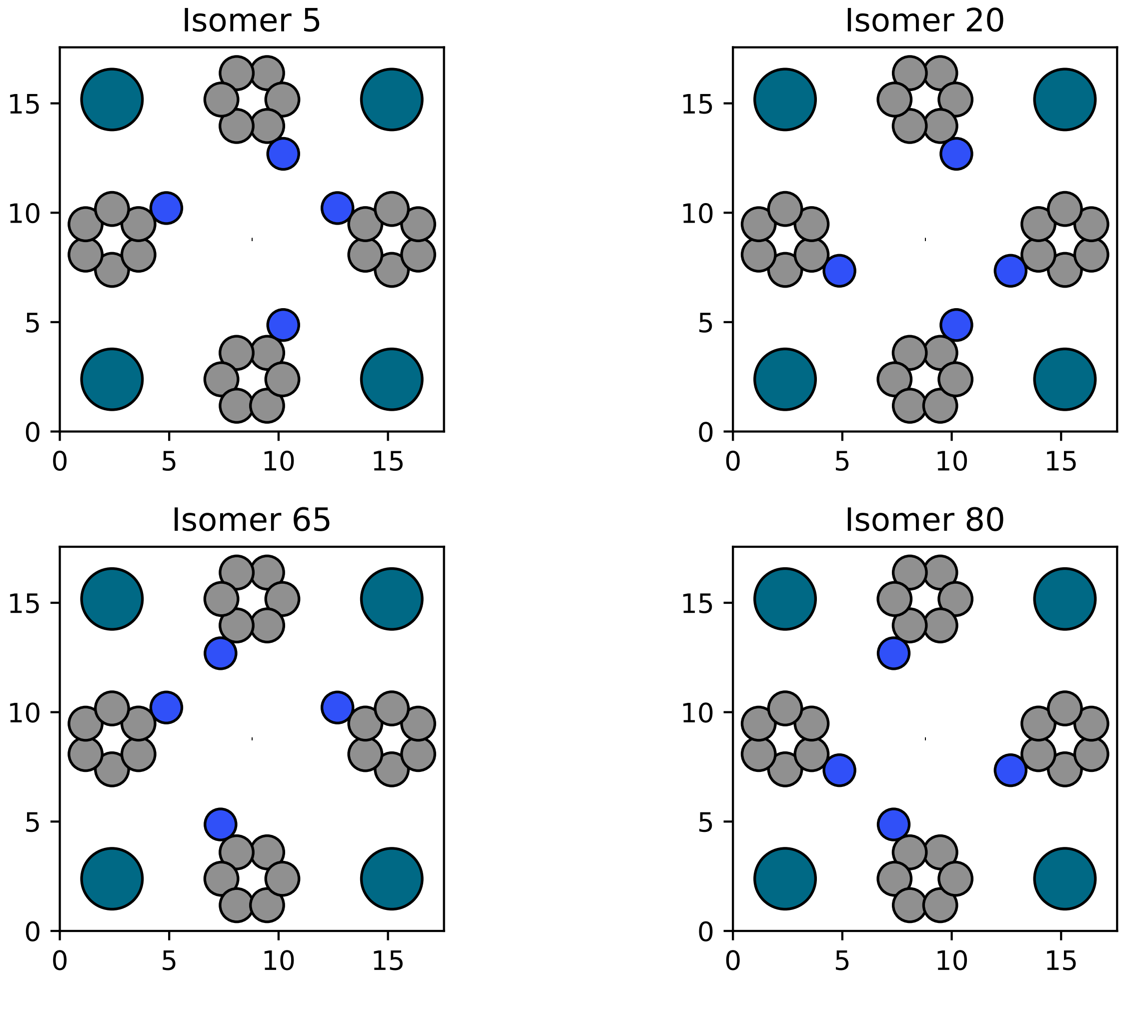


Isomer number 5

Number of FG in pore: 4

Isomer FG-FG distances (Å):

5.8919
7.8236
7.5595
3.5048
7.8236
5.8919

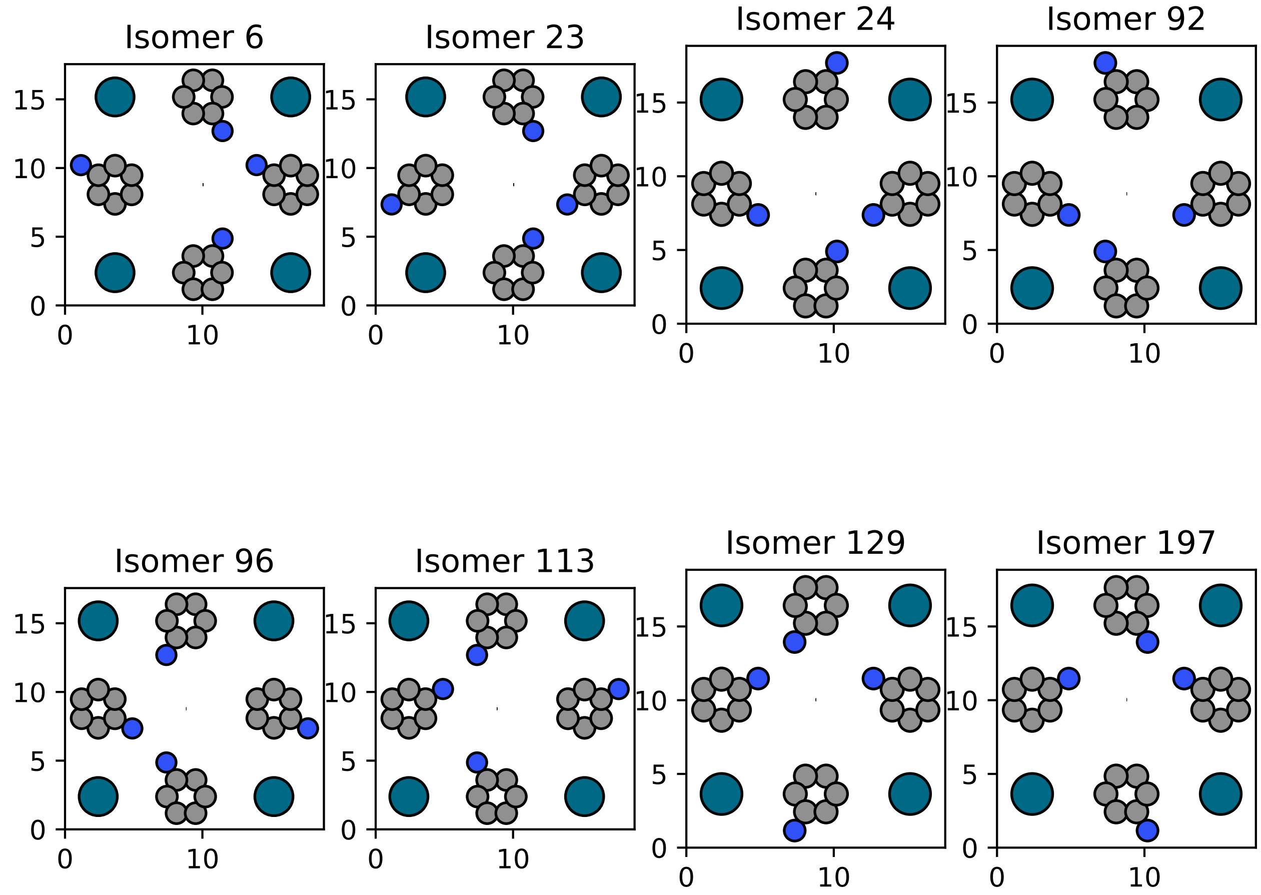


Isomer number 6

Number of FG in pore: 3

Isomer FG-FG distances (Å):

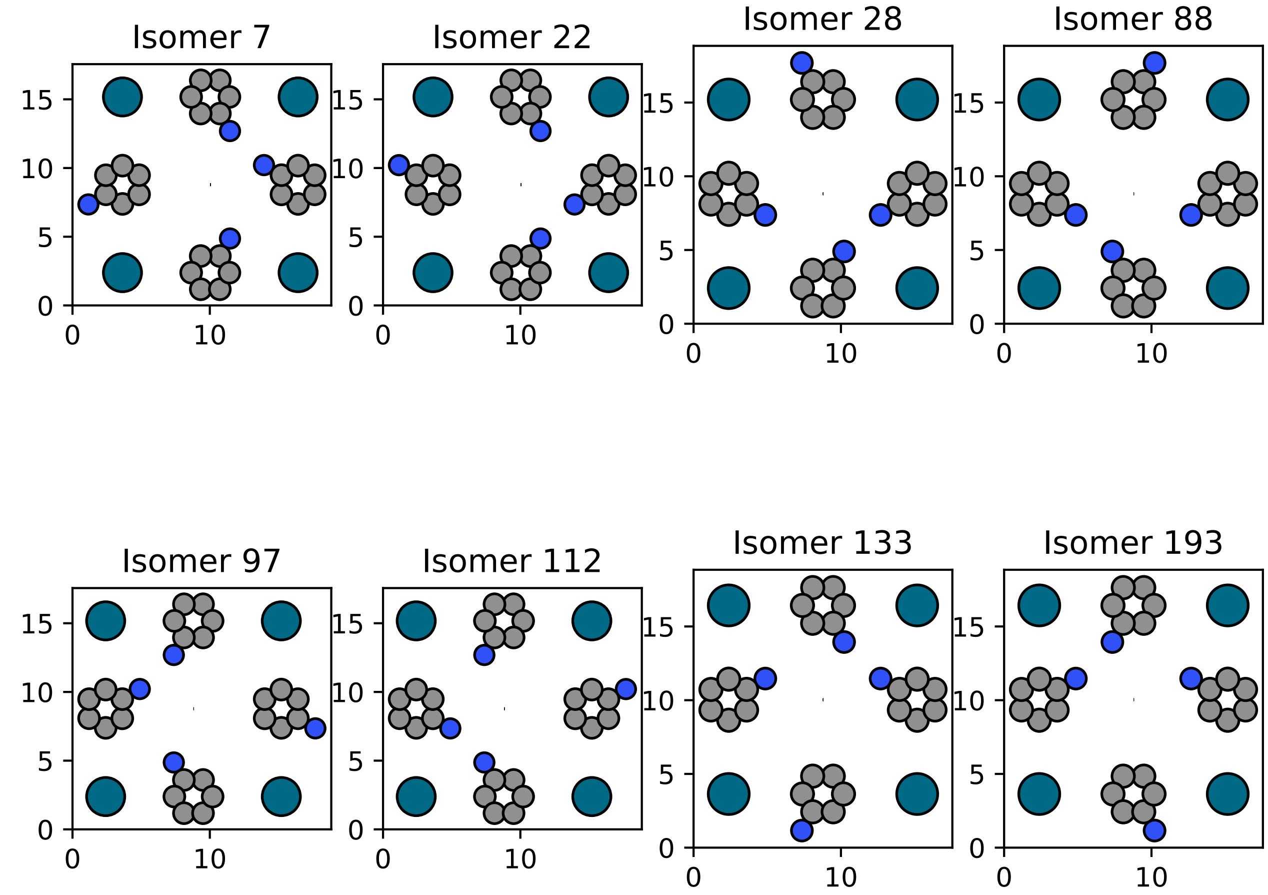
5.8919
7.8236
11.6104
3.5048
12.7870
10.6034



Isomer number 7

Number of FG in pore: 3
Isomer FG-FG distances (Å):

5.8919
7.8236
10.6034
3.5048
13.1037
11.6104



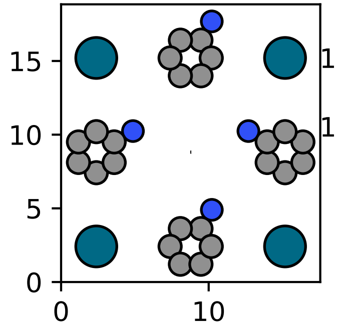
Isomer number 9

Number of FG in pore: 3

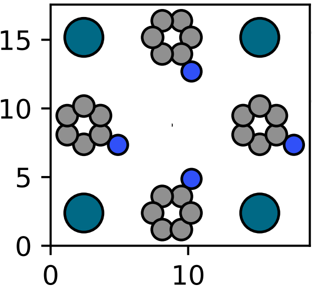
Isomer FG-FG distances (Å):

5.8919
12.7870
7.5595
7.8447
7.8236
9.1602

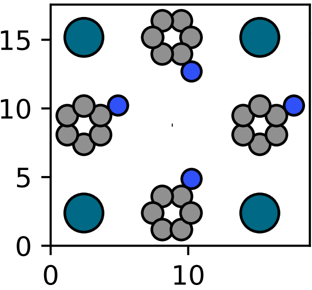
Isomer 9



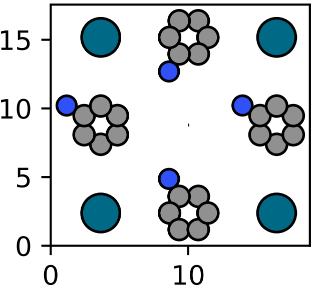
Isomer 36



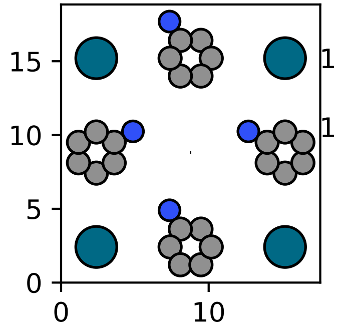
Isomer 53



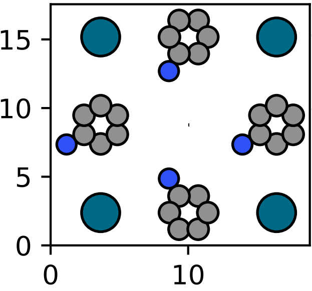
Isomer 66



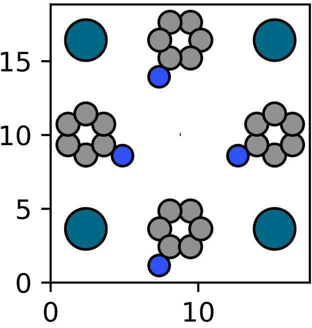
Isomer 77



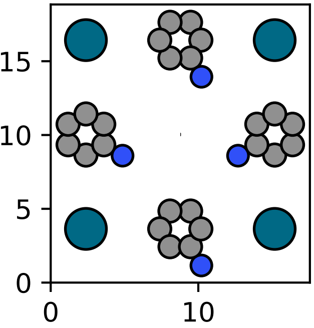
Isomer 83



Isomer 144



Isomer 212

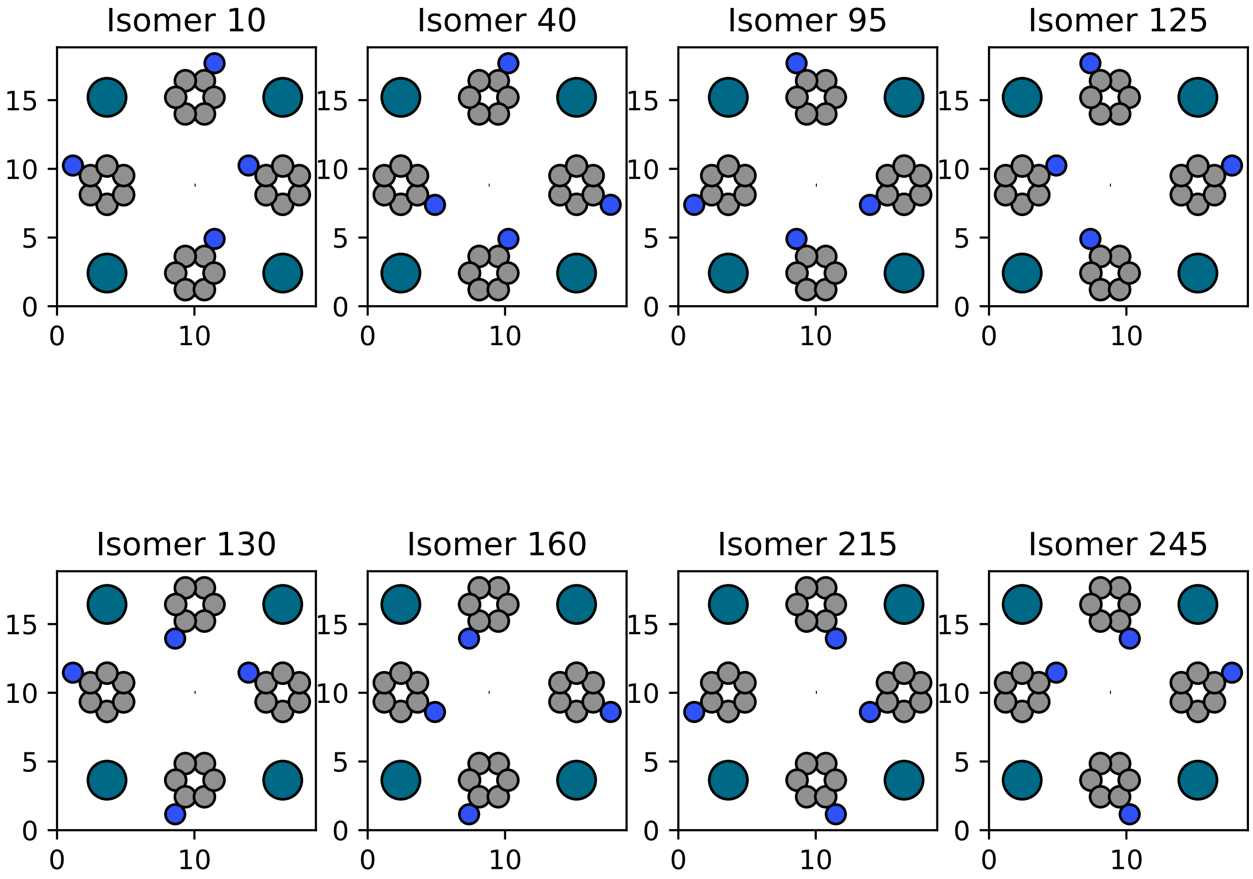


Isomer number 10

Number of FG in pore: 2

Isomer FG-FG distances (Å):

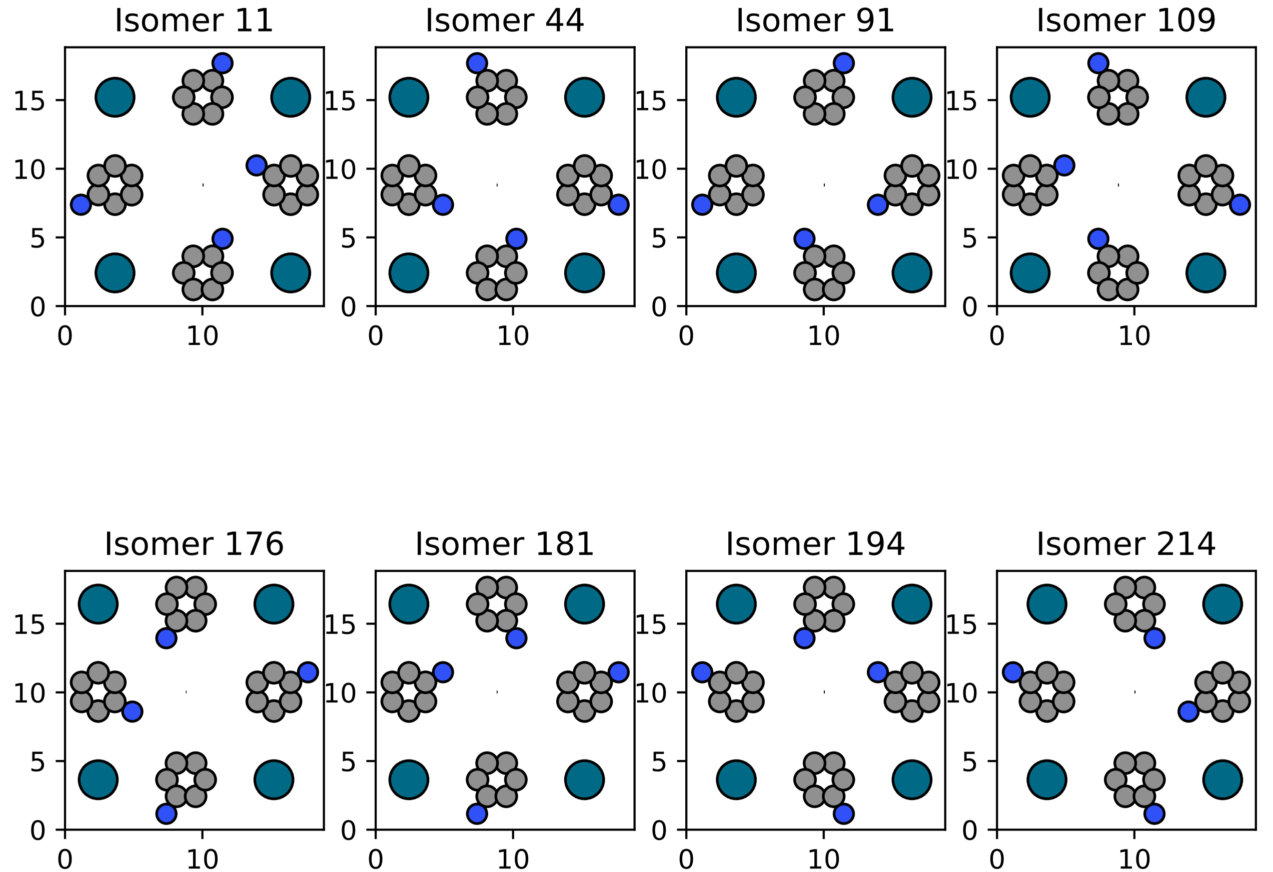
5.8919
12.7870
11.6104
7.8447
12.7870
12.7133



Isomer number 11

Number of FG in pore: 2
Isomer FG-FG distances (Å):

5.8919
12.7870
10.6034
7.8447
13.1037
14.5733



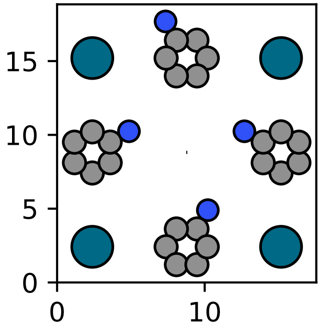
Isomer number 13

Number of FG in pore: 3

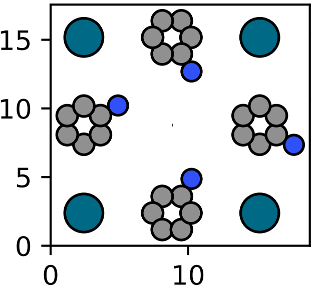
Isomer FG-FG distances (Å):

5.8919
13.1037
7.5595
9.1602
7.8236
7.8447

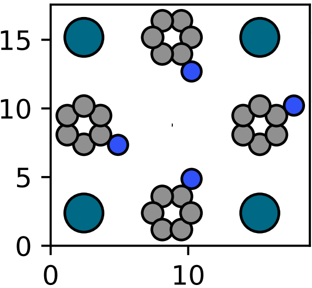
Isomer 13



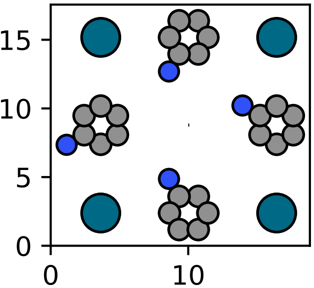
Isomer 37



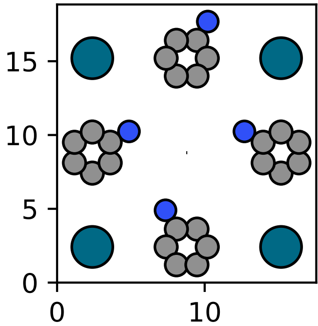
Isomer 52



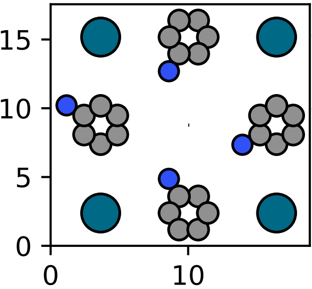
Isomer 67



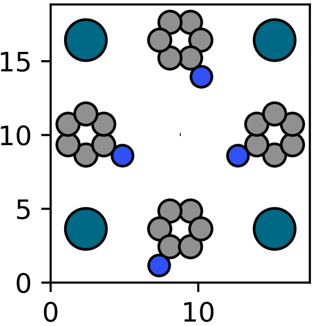
Isomer 73



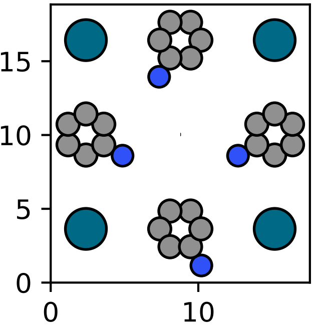
Isomer 82



Isomer 148



Isomer 208

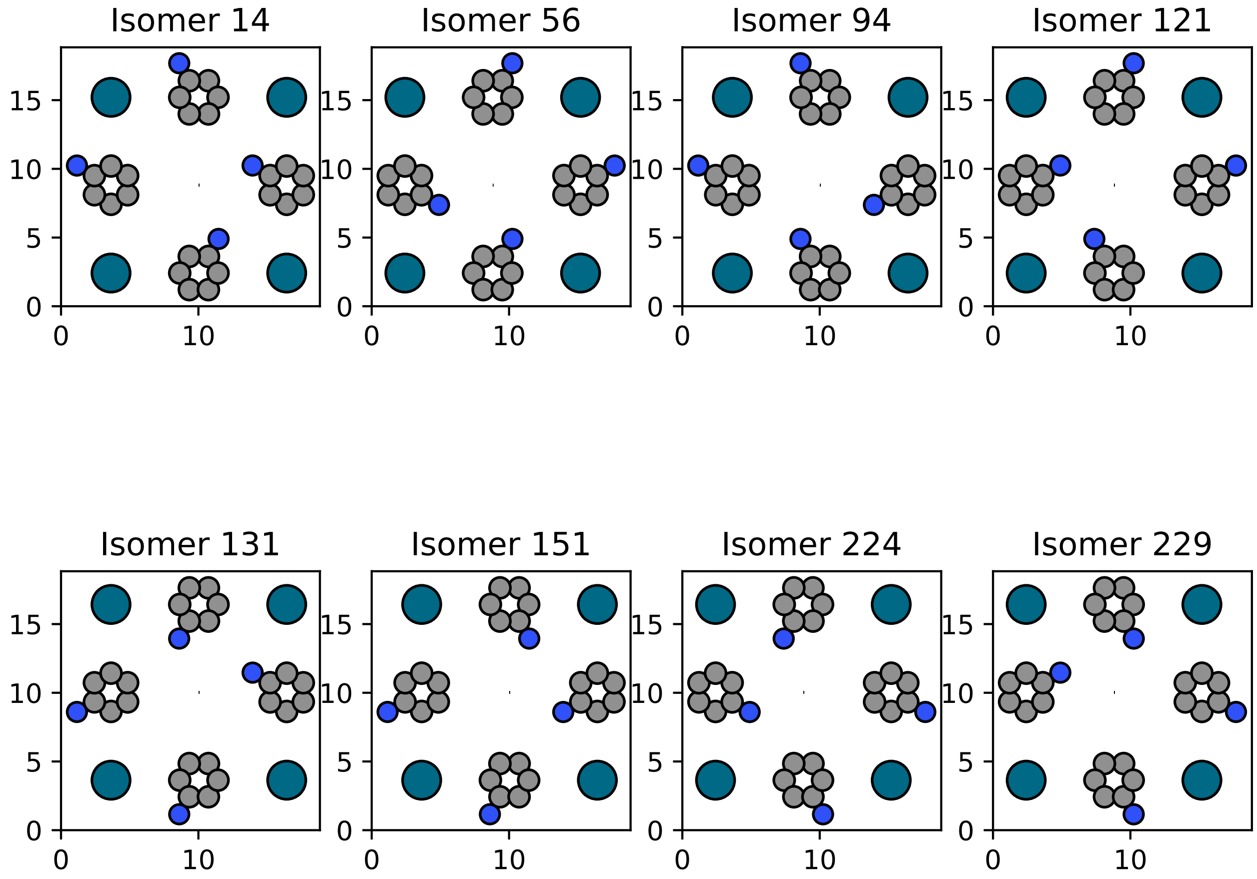


Isomer number 14

Number of FG in pore: 2

Isomer FG-FG distances (Å):

5.8919
13.1037
11.6104
9.1602
12.7870
10.5297

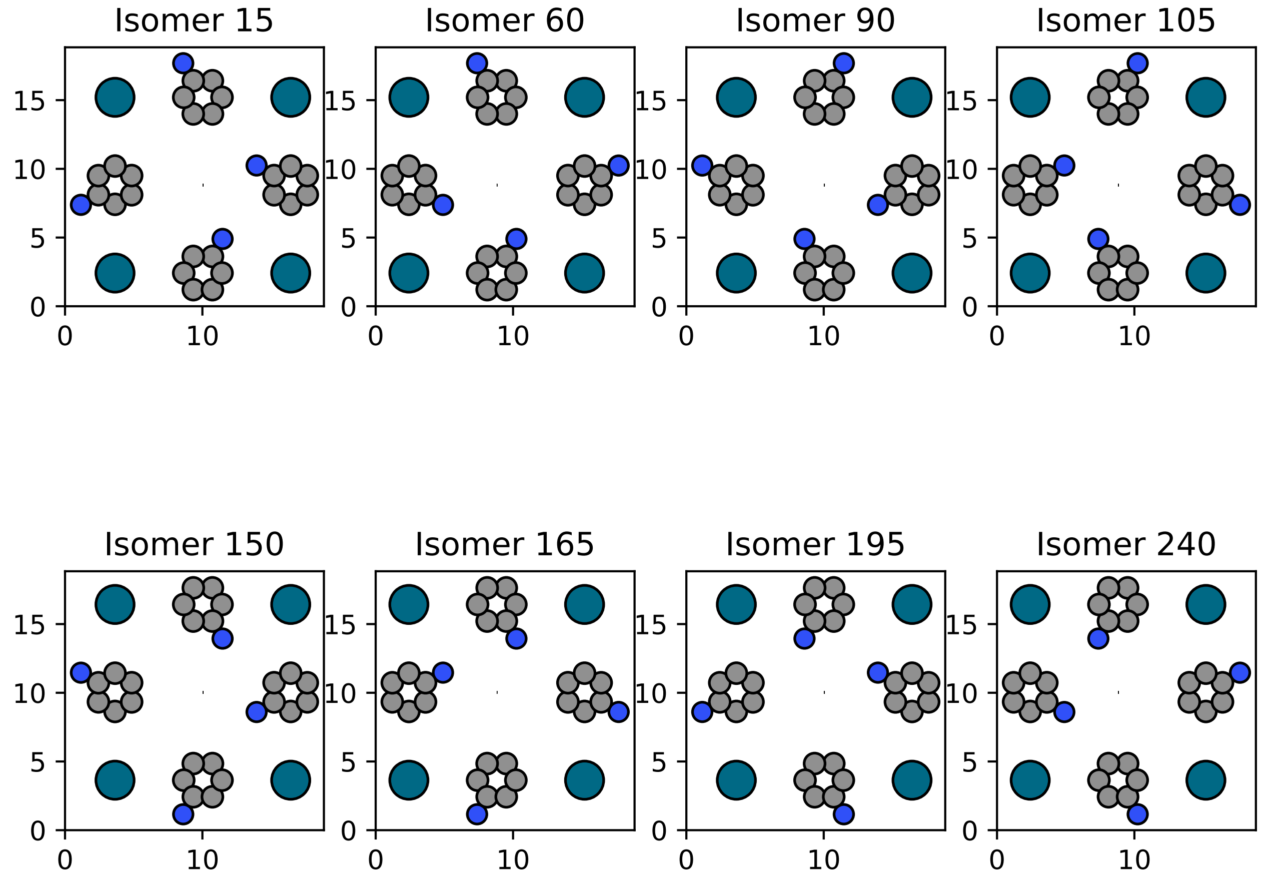


Isomer number 15

Number of FG in pore: 2

Isomer FG-FG distances (Å):

5.8919
13.1037
10.6034
9.1602
13.1037
12.7133

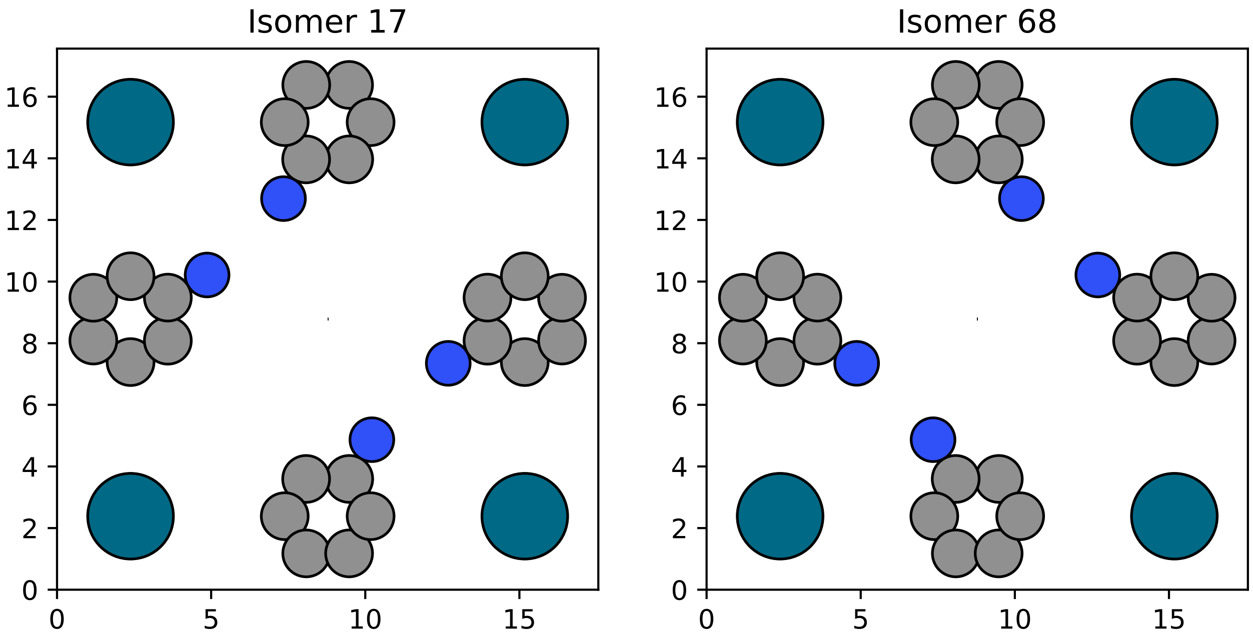


Isomer number 17

Number of FG in pore: 4

Isomer FG-FG distances (Å):

3.5048
8.3324
7.5595
7.5595
8.3324
3.5048

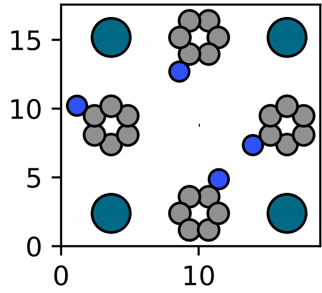


Isomer number 18

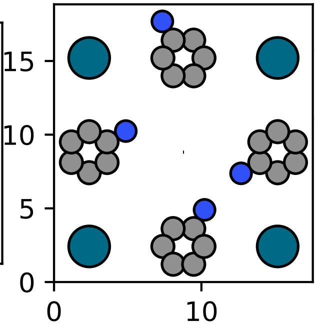
Number of FG in pore: 3
Isomer FG-FG distances (Å):

3.5048
8.3324
11.6104
7.5595
13.1037
7.8447

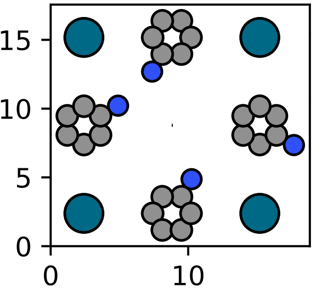
Isomer 18



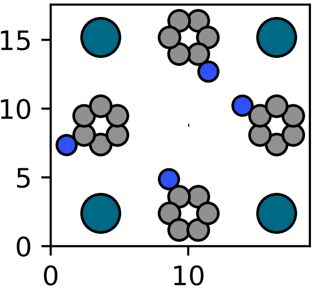
Isomer 29



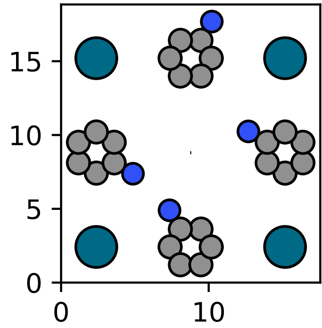
Isomer 33



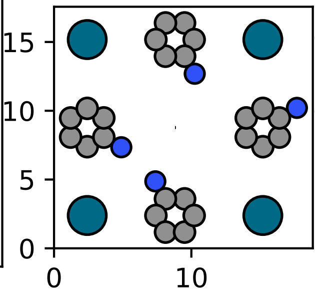
Isomer 71



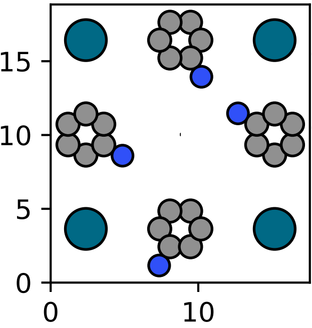
Isomer 72



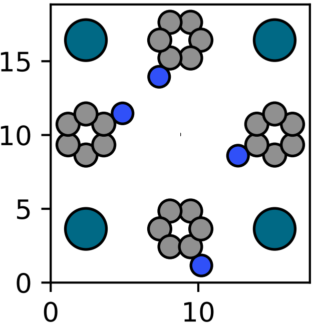
Isomer 116



Isomer 132



Isomer 209

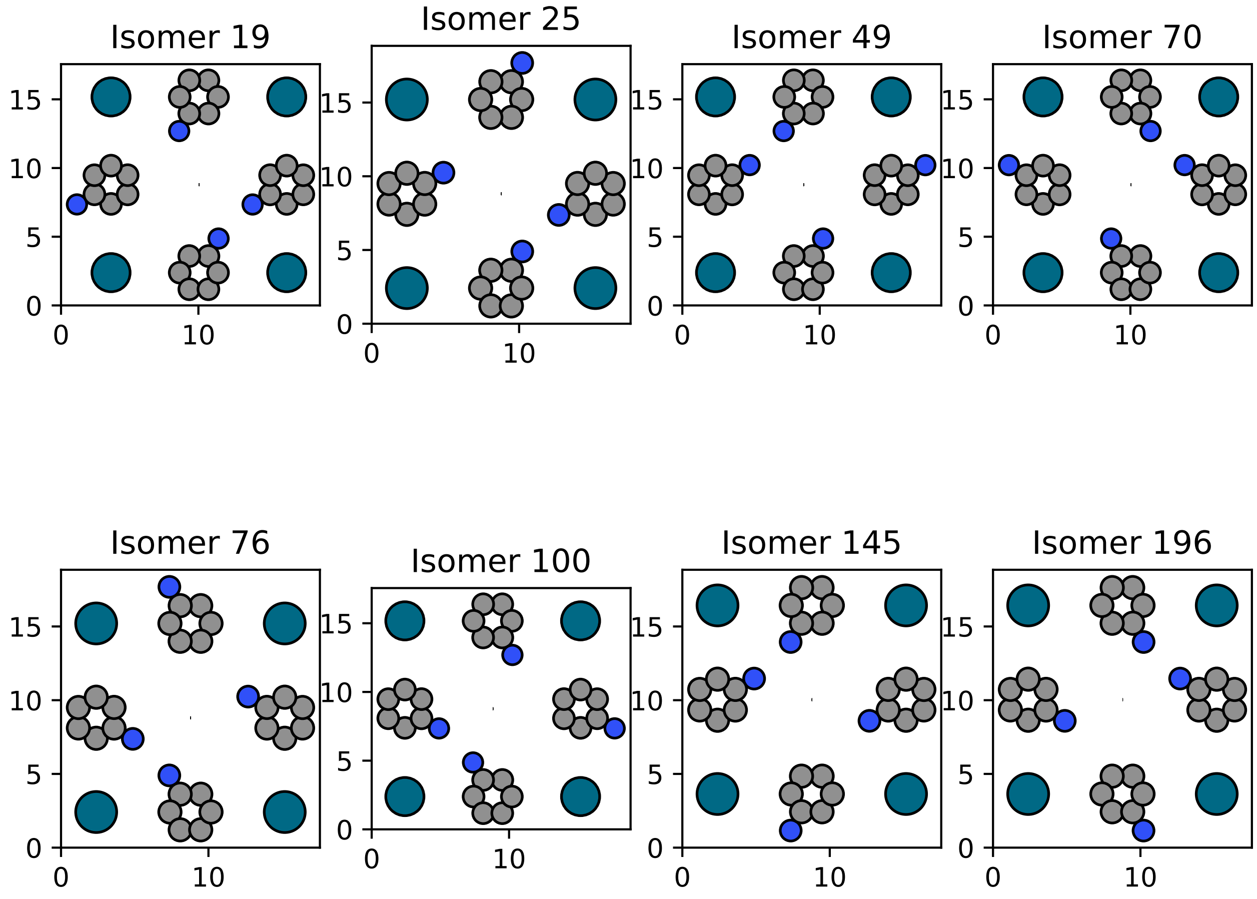


Isomer number 19

Number of FG in pore: 3

Isomer FG-FG distances (Å):

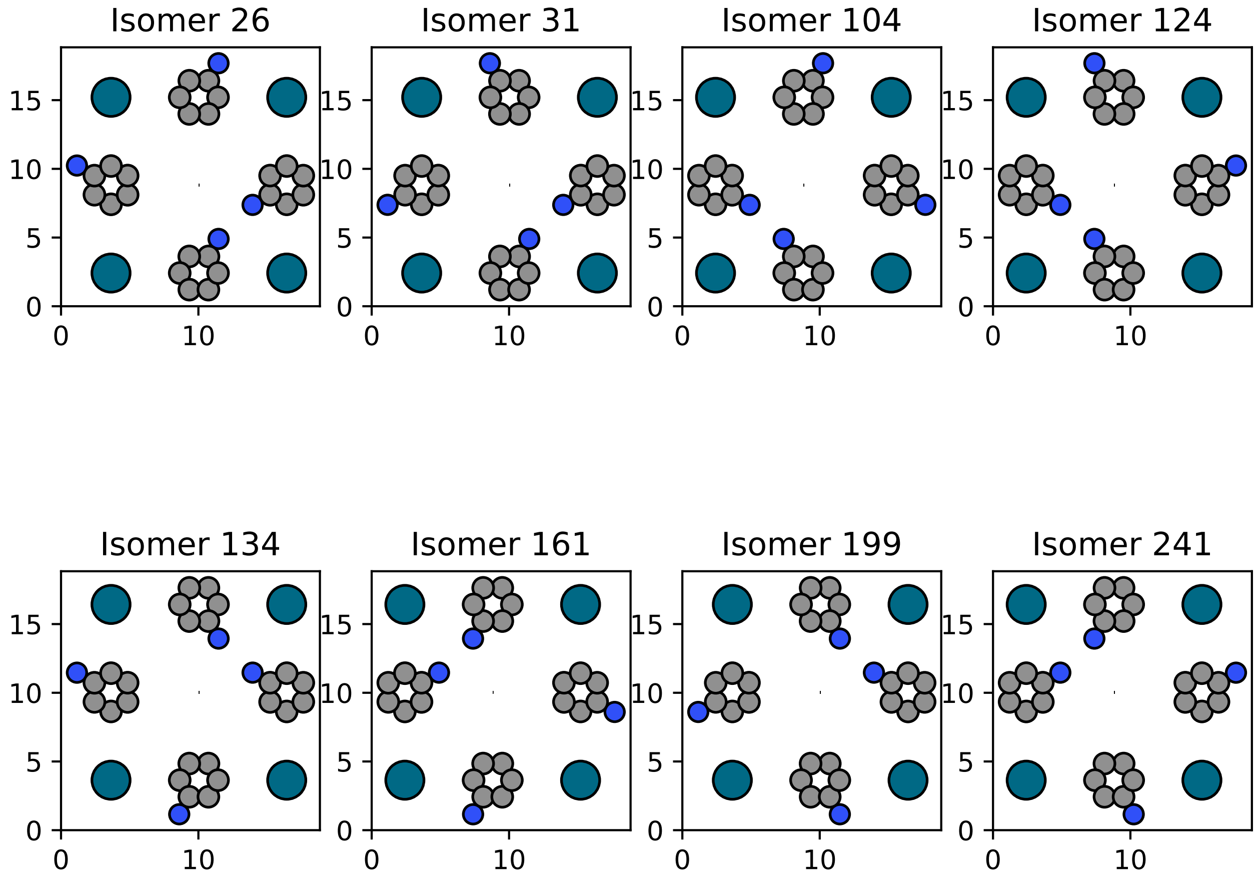
3.5048
8.3324
10.6034
7.5595
12.7870
9.1602



Isomer number 26

Number of FG in pore: 2
Isomer FG-FG distances (Å):

3.5048
12.7870
11.6104
10.6034
13.1037
12.7133

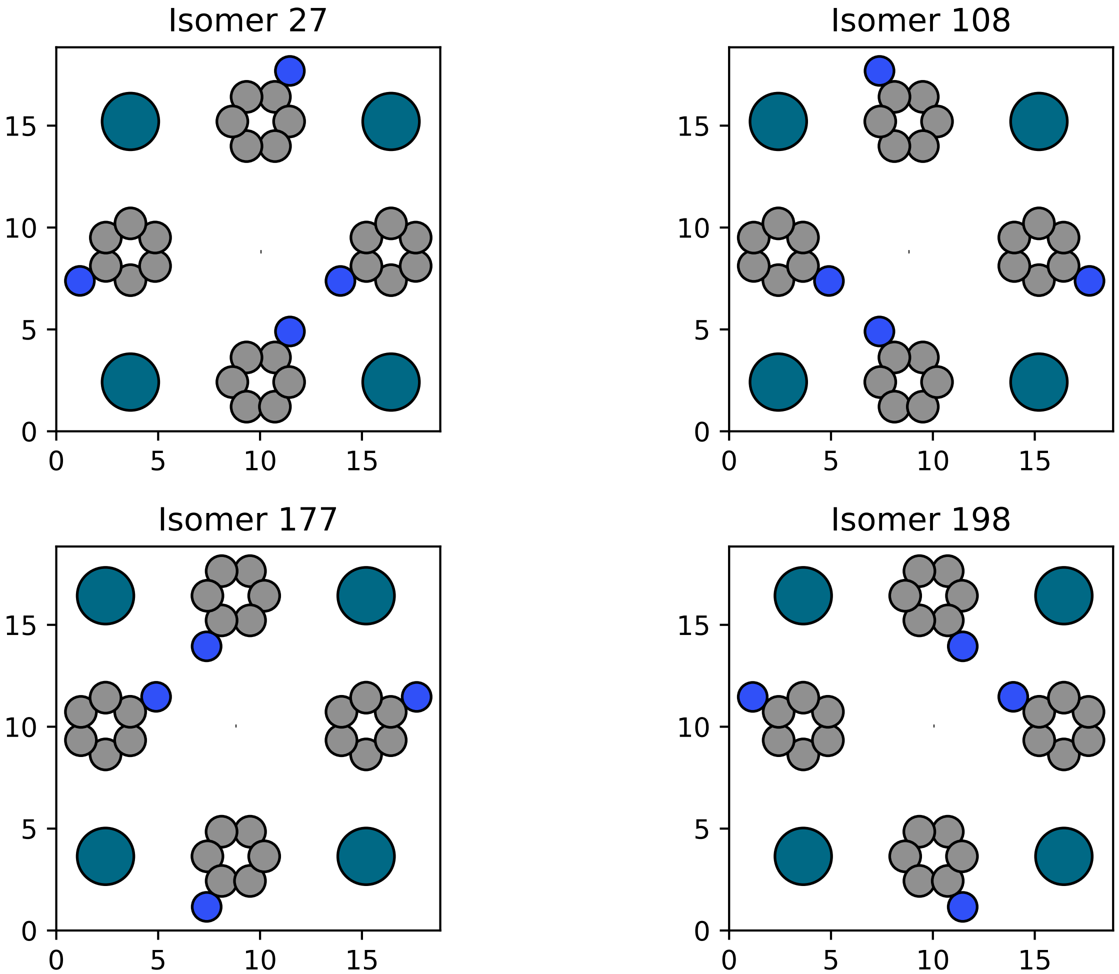


Isomer number 27

Number of FG in pore: 2

Isomer FG-FG distances (Å):

3.5048
12.7870
10.6034
10.6034
12.7870
14.5733


Isomer 27

Isomer 108

Isomer 177

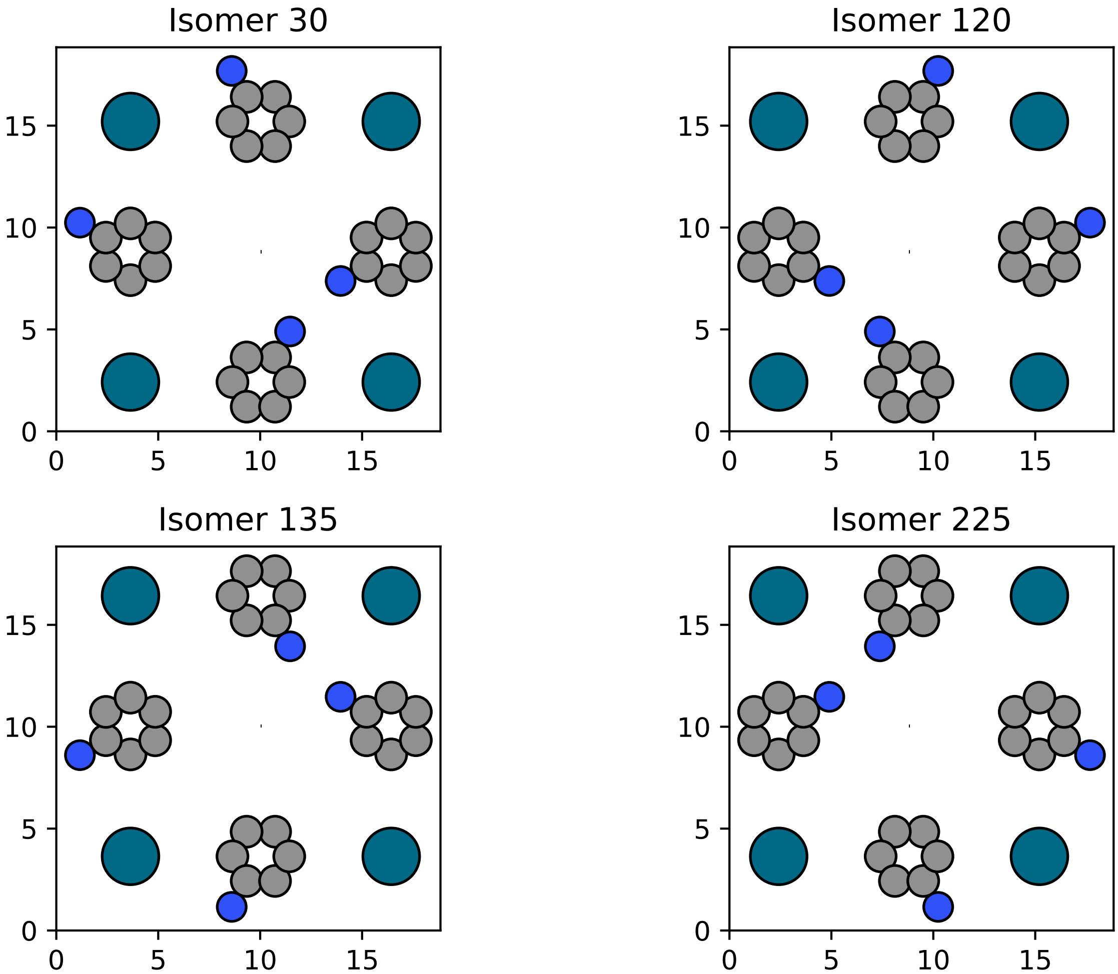
Isomer 198

Isomer number 30

Number of FG in pore: 2

Isomer FG-FG distances (Å):

3.5048
13.1037
11.6104
11.6104
13.1037
10.5297


Isomer 30

Isomer 120

Isomer 135

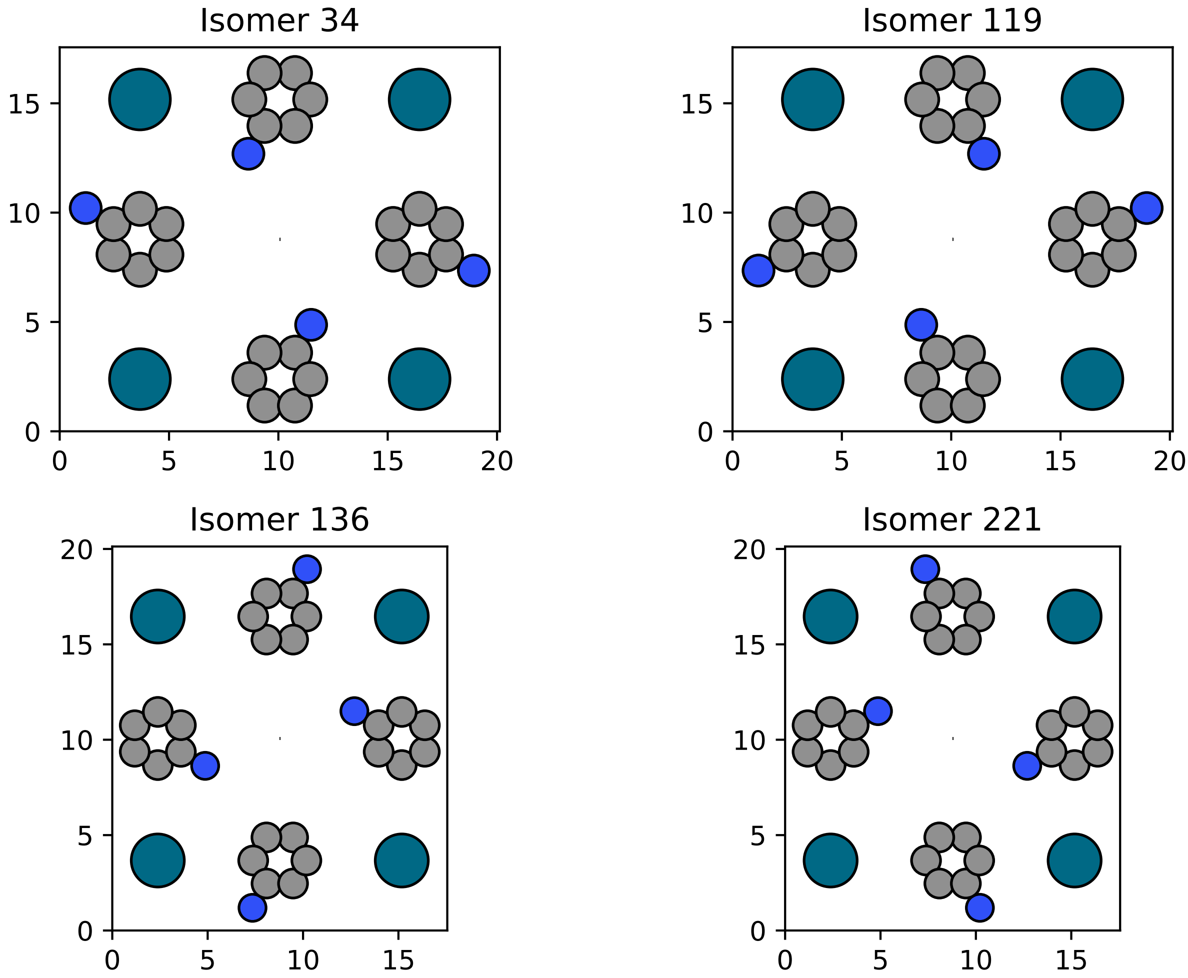
Isomer 225

Isomer number 34

Number of FG in pore: 2

Isomer FG-FG distances (Å):

7.8447
8.3324
11.6104
11.6104
17.9793
7.8447

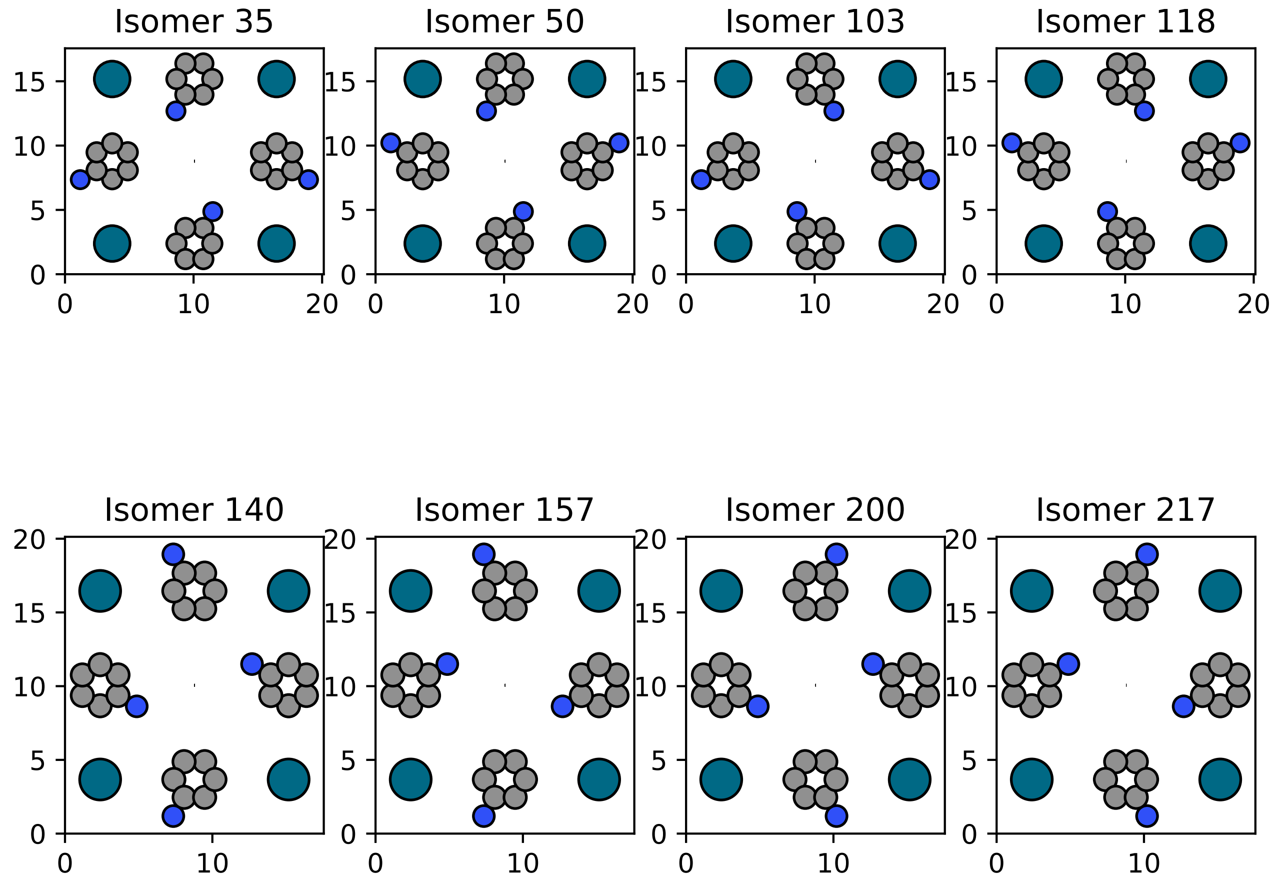


Isomer number 35

Number of FG in pore: 2

Isomer FG-FG distances (Å):

7.8447
8.3324
10.6034
11.6104
17.7505
9.1602

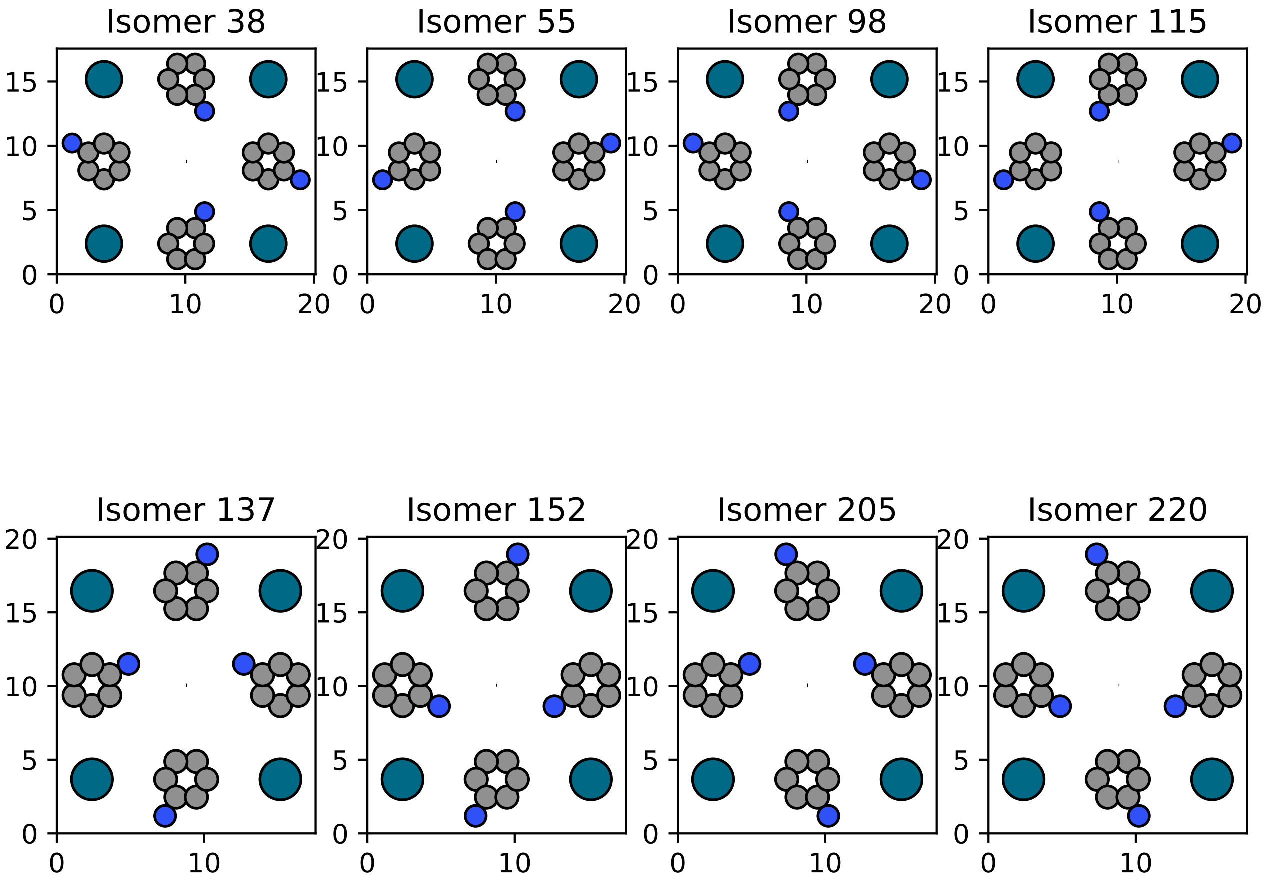


Isomer number 38

Number of FG in pore: 2

Isomer FG-FG distances (Å):

7.8447
7.8236
11.6104
9.1602
17.9793
10.6034

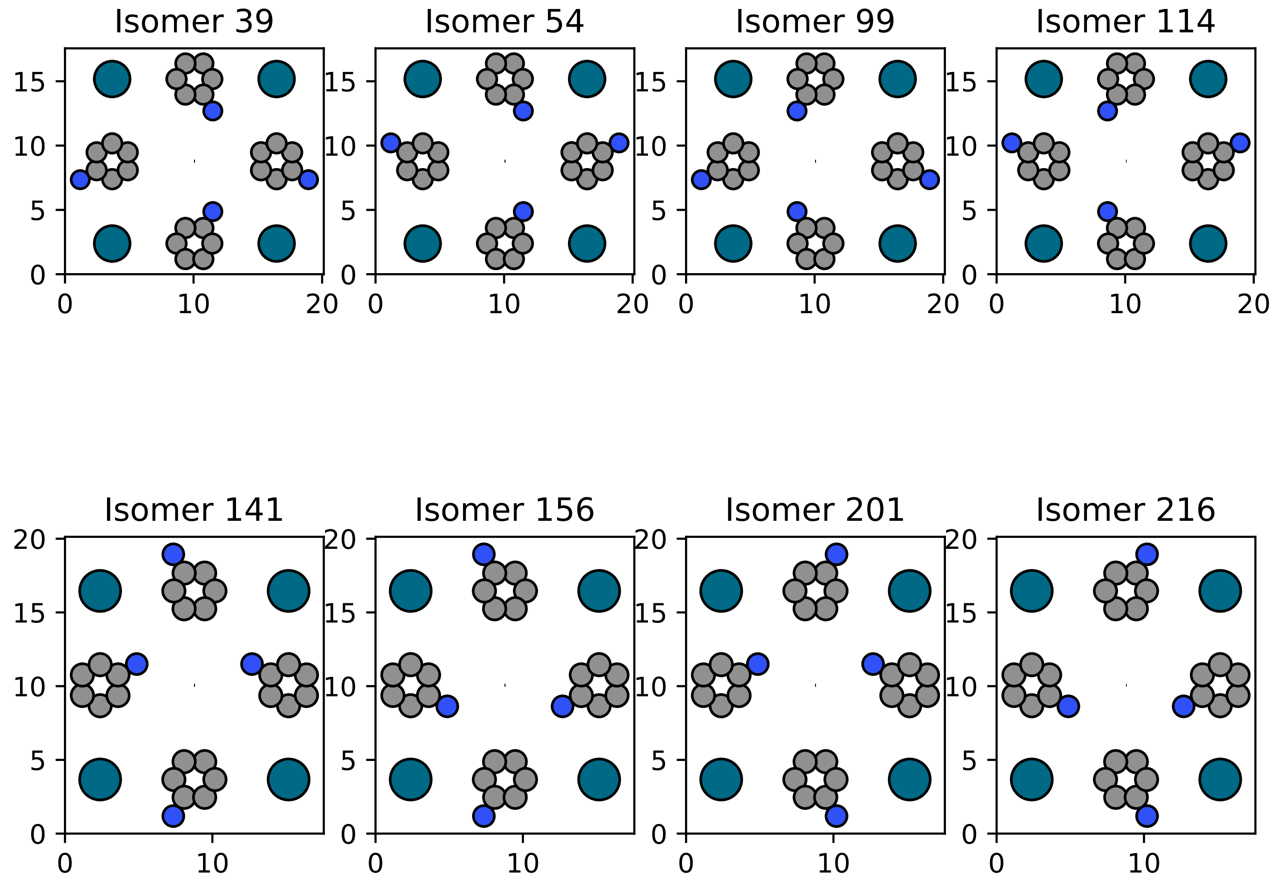


Isomer number 39

Number of FG in pore: 2

Isomer FG-FG distances (Å):

7.8447
7.8236
10.6034
9.1602
17.7505
11.6104



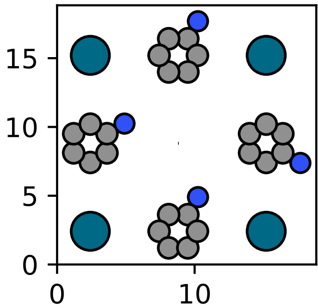
Isomer number 41

Number of FG in pore: 2

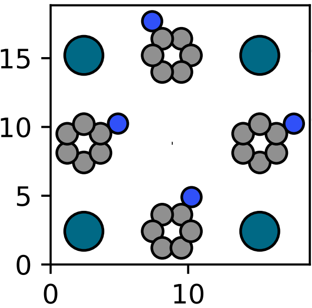
Isomer FG-FG distances (Å):

7.8447
12.7870
7.5595
12.7133
13.1037
9.1602

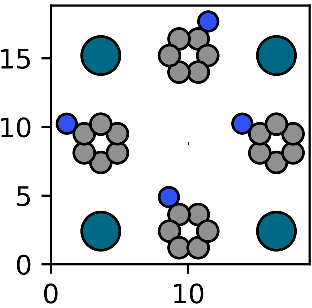
Isomer 41



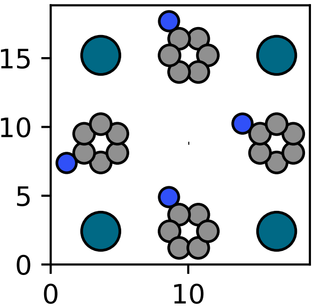
Isomer 61



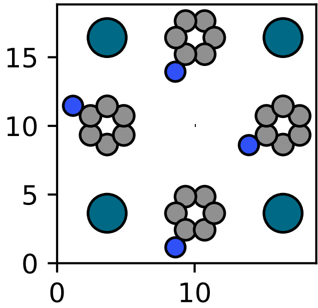
Isomer 74



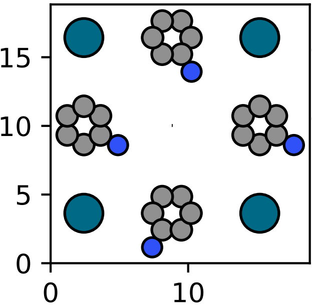
Isomer 79



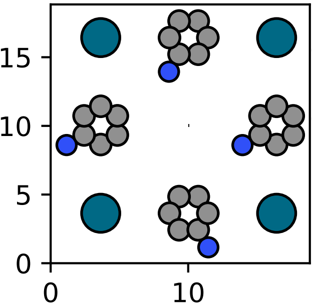
Isomer 146



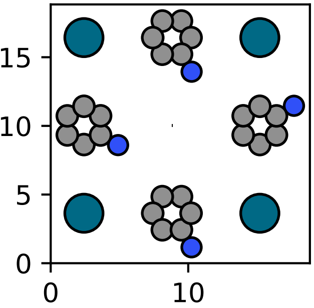
Isomer 164



Isomer 211



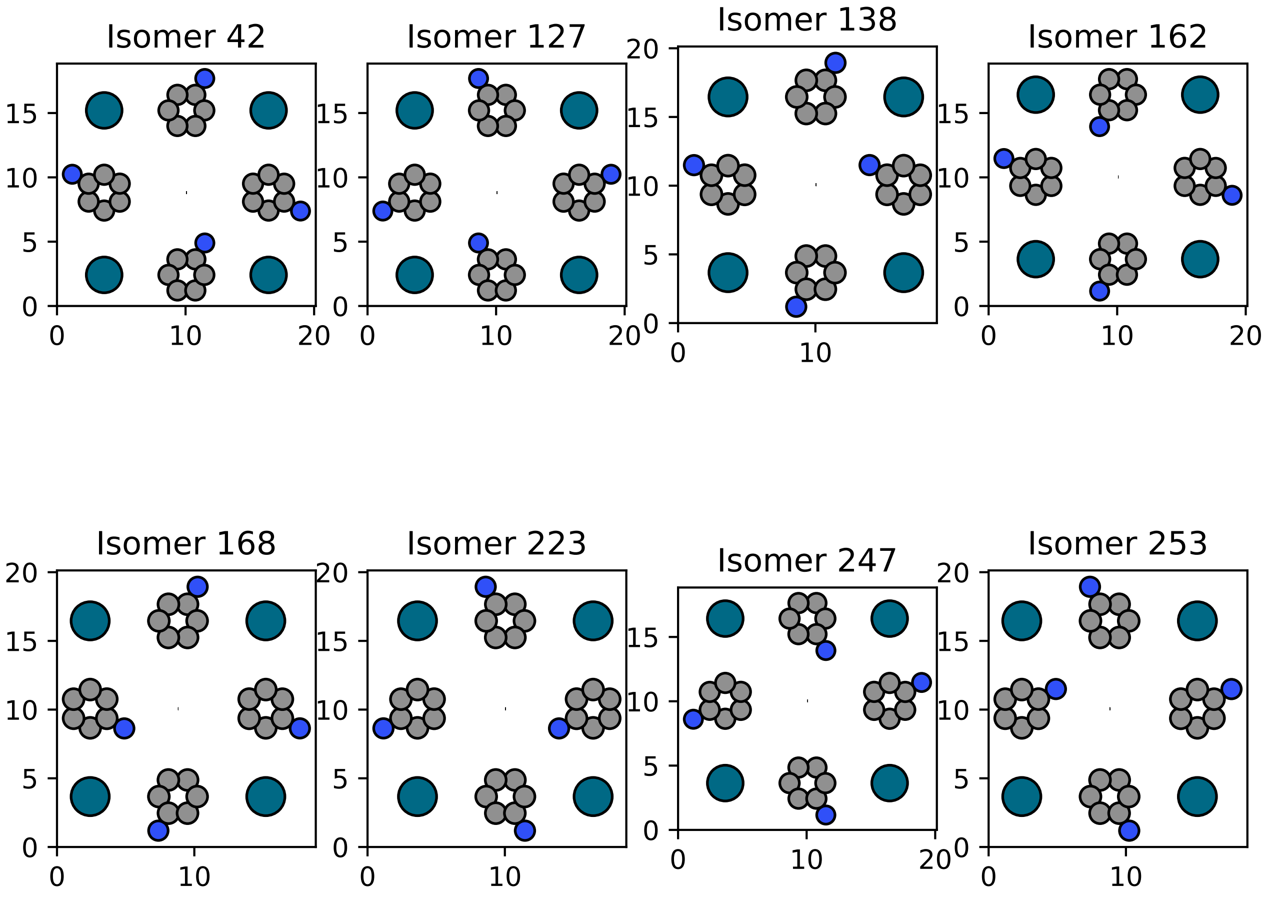
Isomer 244



Isomer number 42

Number of FG in pore: 1
Isomer FG-FG distances (Å):

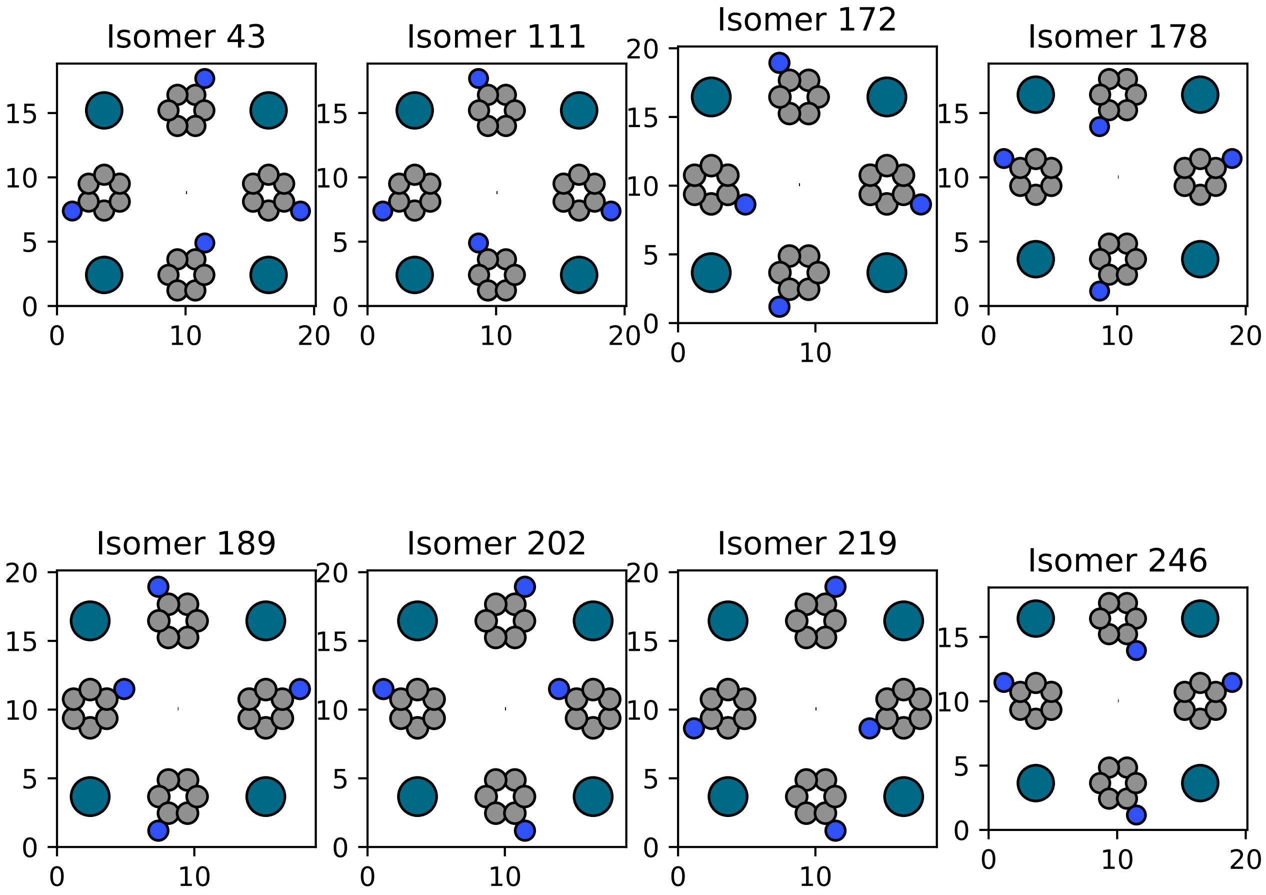
7.8447
12.7870
11.6104
12.7133
17.9793
12.7133



Isomer number 43

Number of FG in pore: 1
Isomer FG-FG distances (Å):

7.8447
12.7870
10.6034
12.7133
17.7505
14.5733

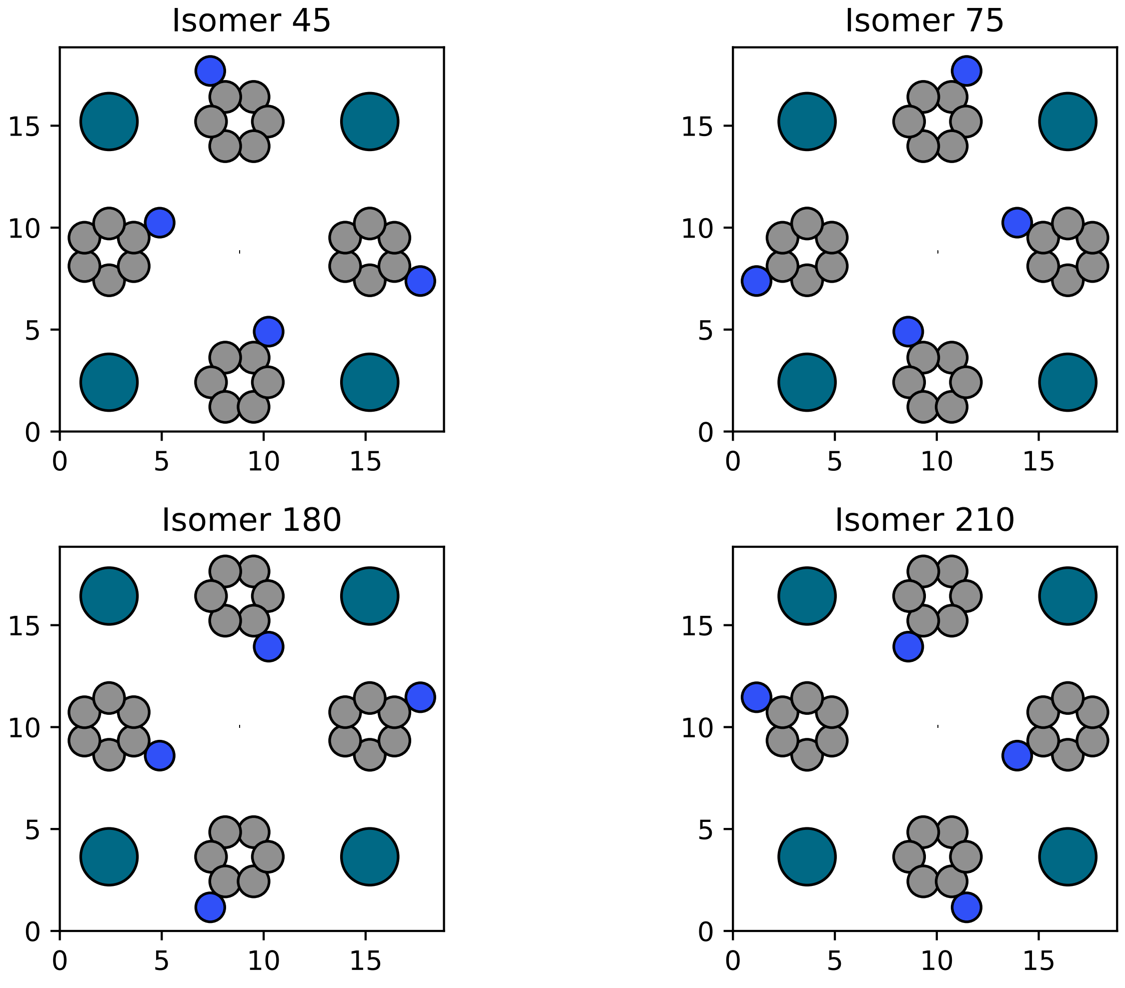


Isomer number 45

Number of FG in pore: 2

Isomer FG-FG distances (Å):

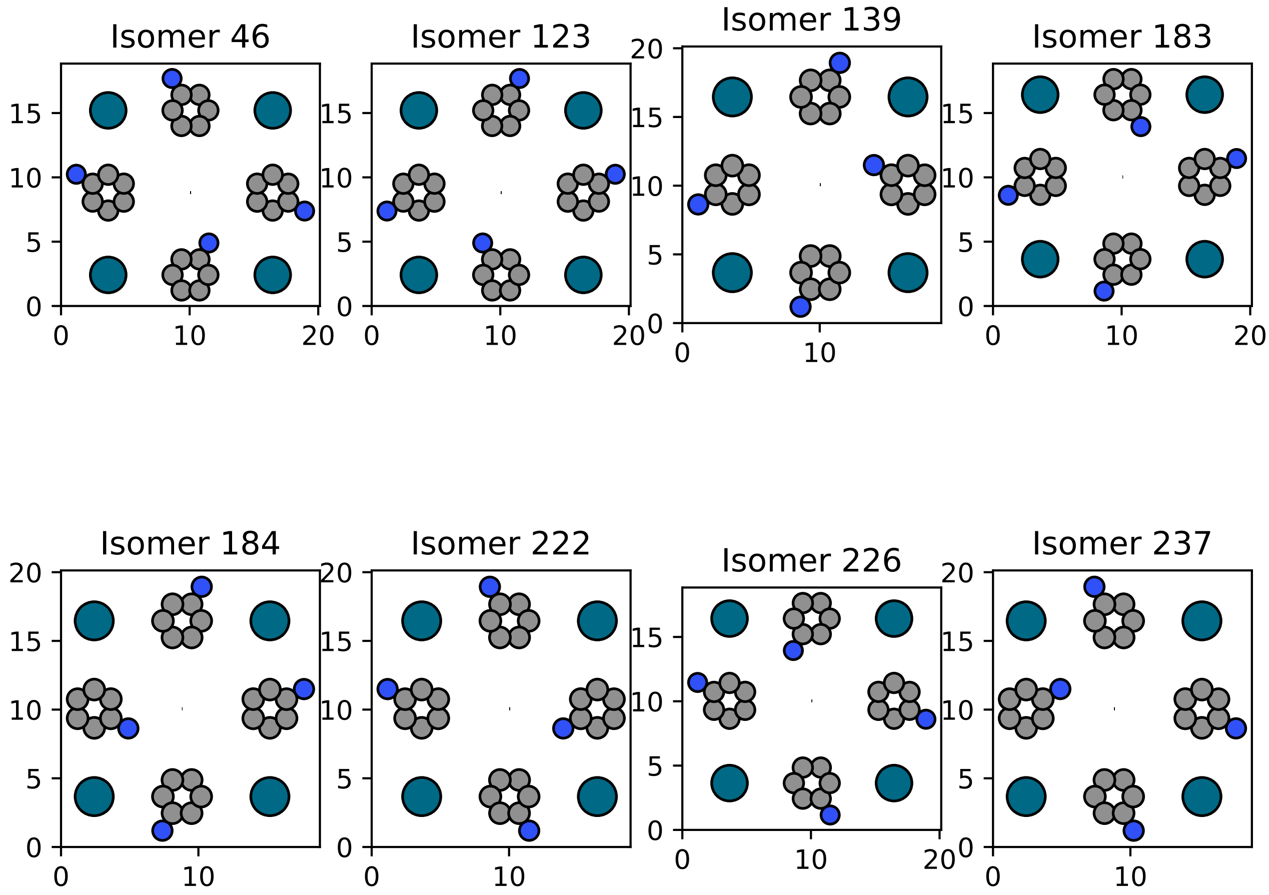
7.8447
13.1037
7.5595
14.5733
13.1037
7.8447



Isomer number 46

Number of FG in pore: 1
Isomer FG-FG distances (Å):

7.8447
13.1037
11.6104
14.5733
17.9793
10.5297

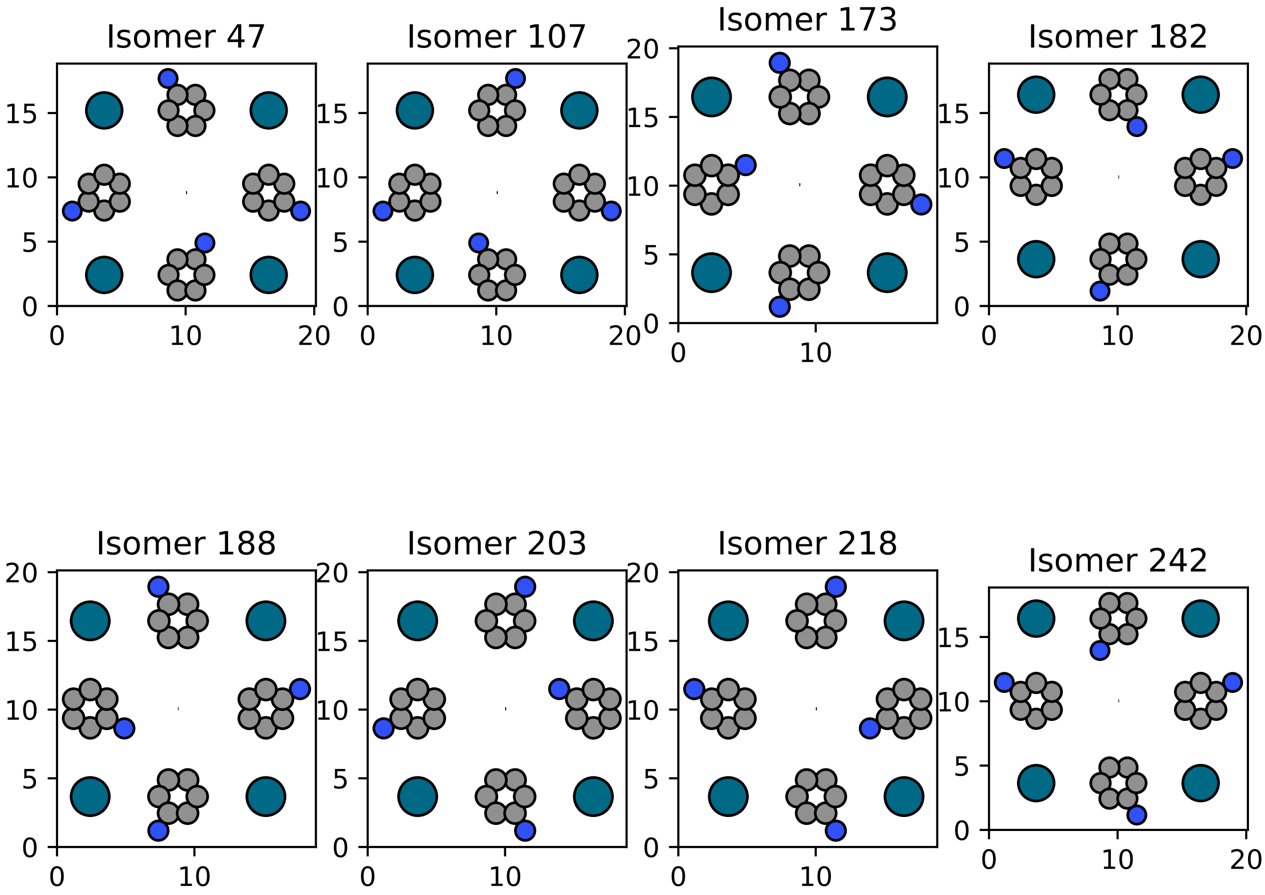


Isomer number 47

Number of FG in pore: 1

Isomer FG-FG distances (Å):

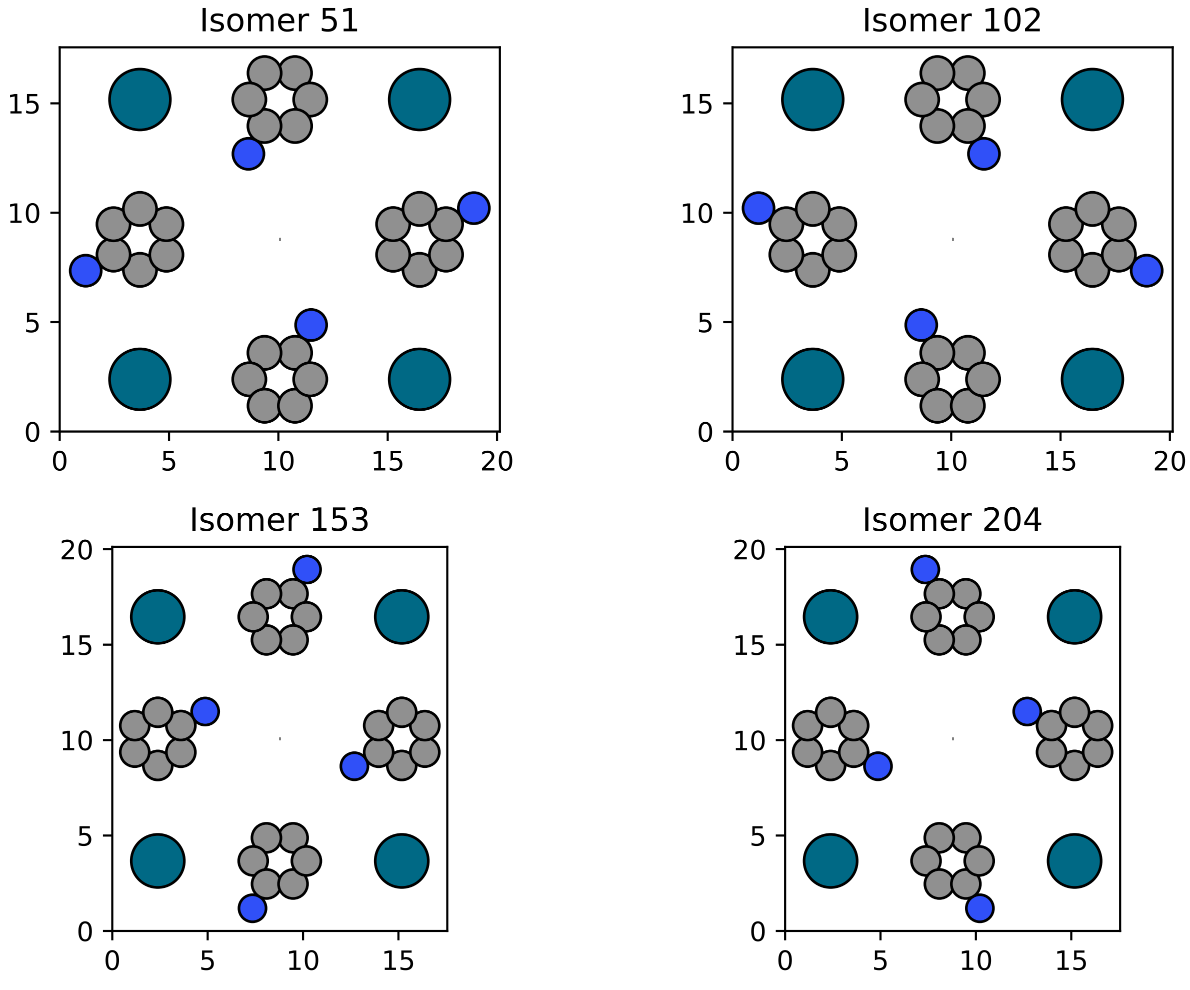
7.8447
13.1037
10.6034
14.5733
17.7505
12.7133



Isomer number 51

Number of FG in pore: 2
Isomer FG-FG distances (Å):

9.1602
8.3324
10.6034
10.6034
17.9793
9.1602

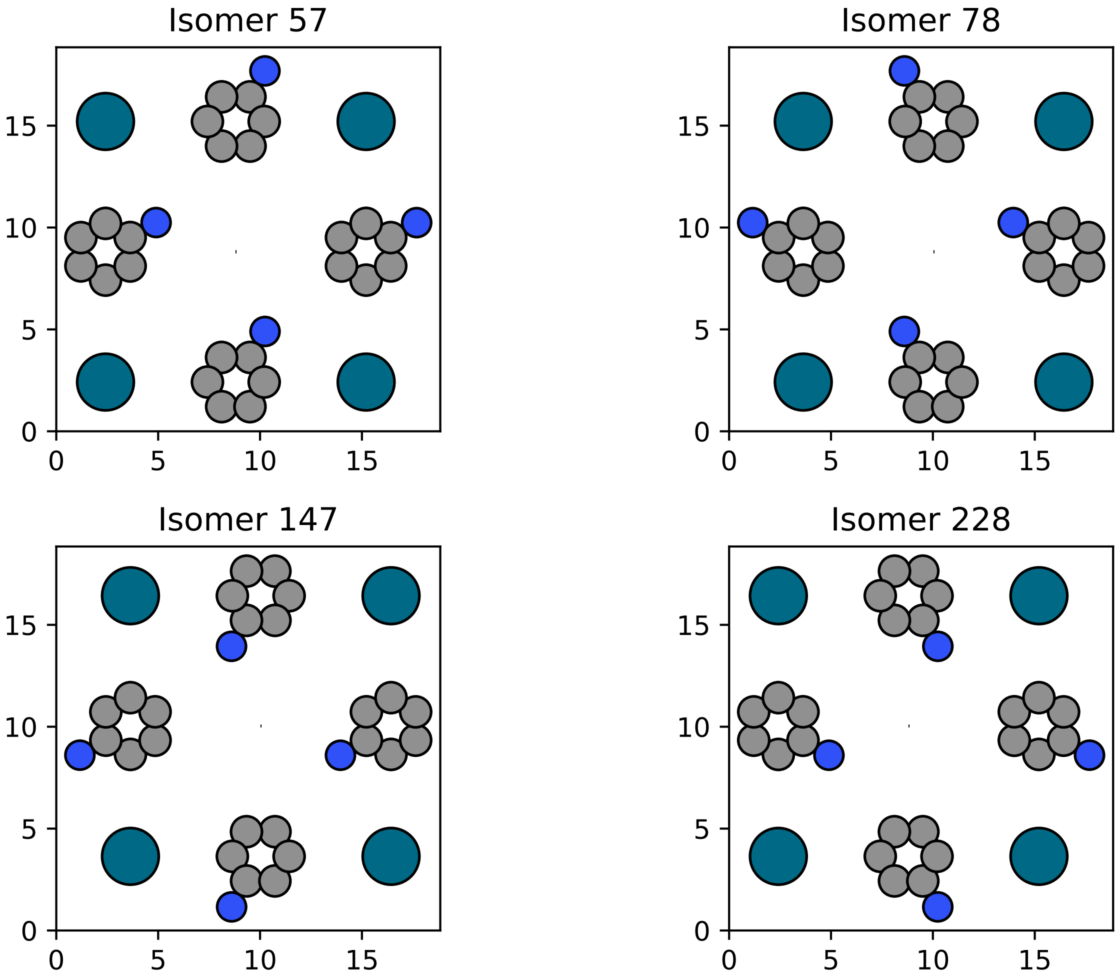


Isomer number 57

Number of FG in pore: 2

Isomer FG-FG distances (Å):

9.1602
12.7870
7.5595
10.5297
12.7870
9.1602


Isomer 57

Isomer 78

Isomer 147

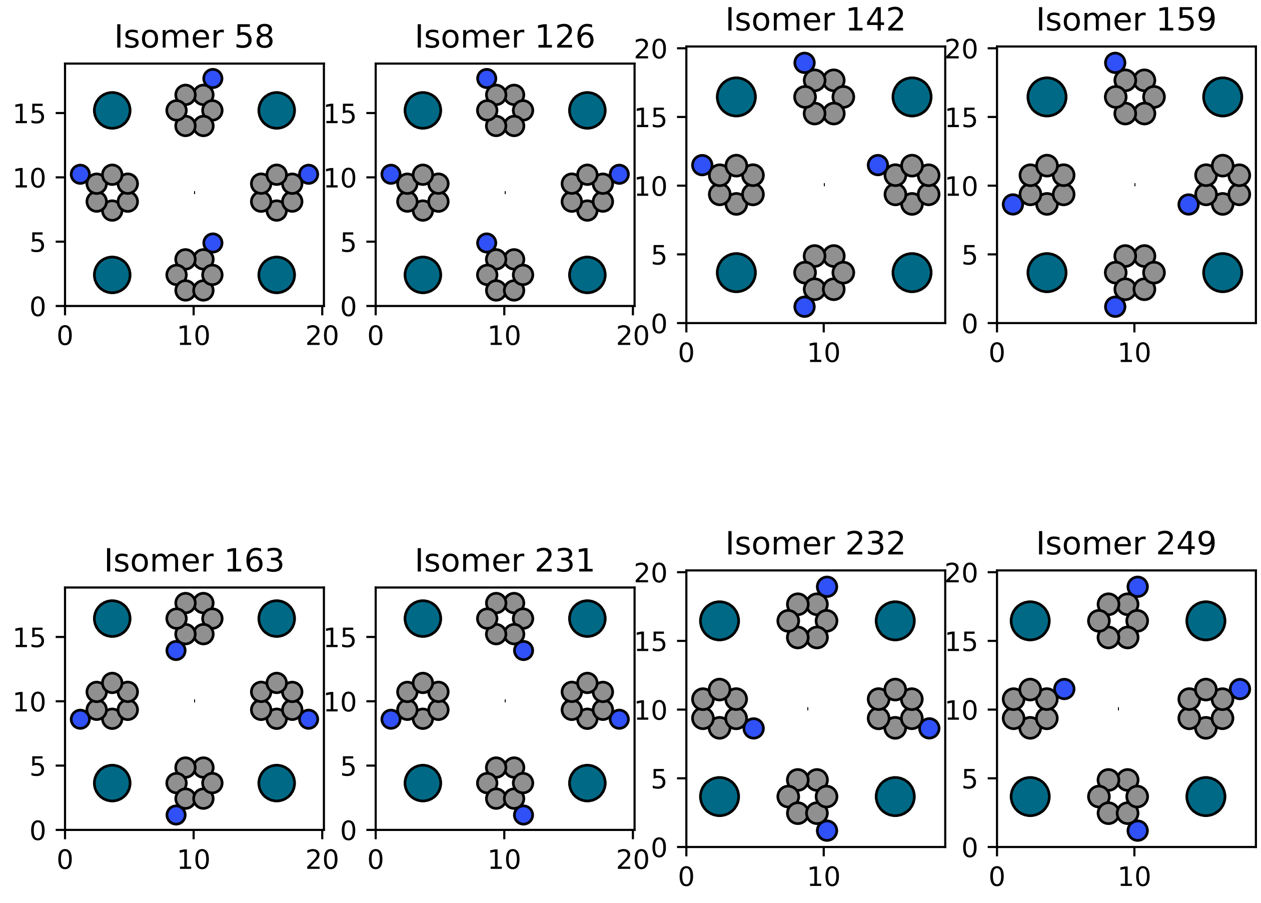
Isomer 228

Isomer number 58

Number of FG in pore: 1

Isomer FG-FG distances (Å):

9.1602
12.7870
11.6104
10.5297
17.7505
12.7133

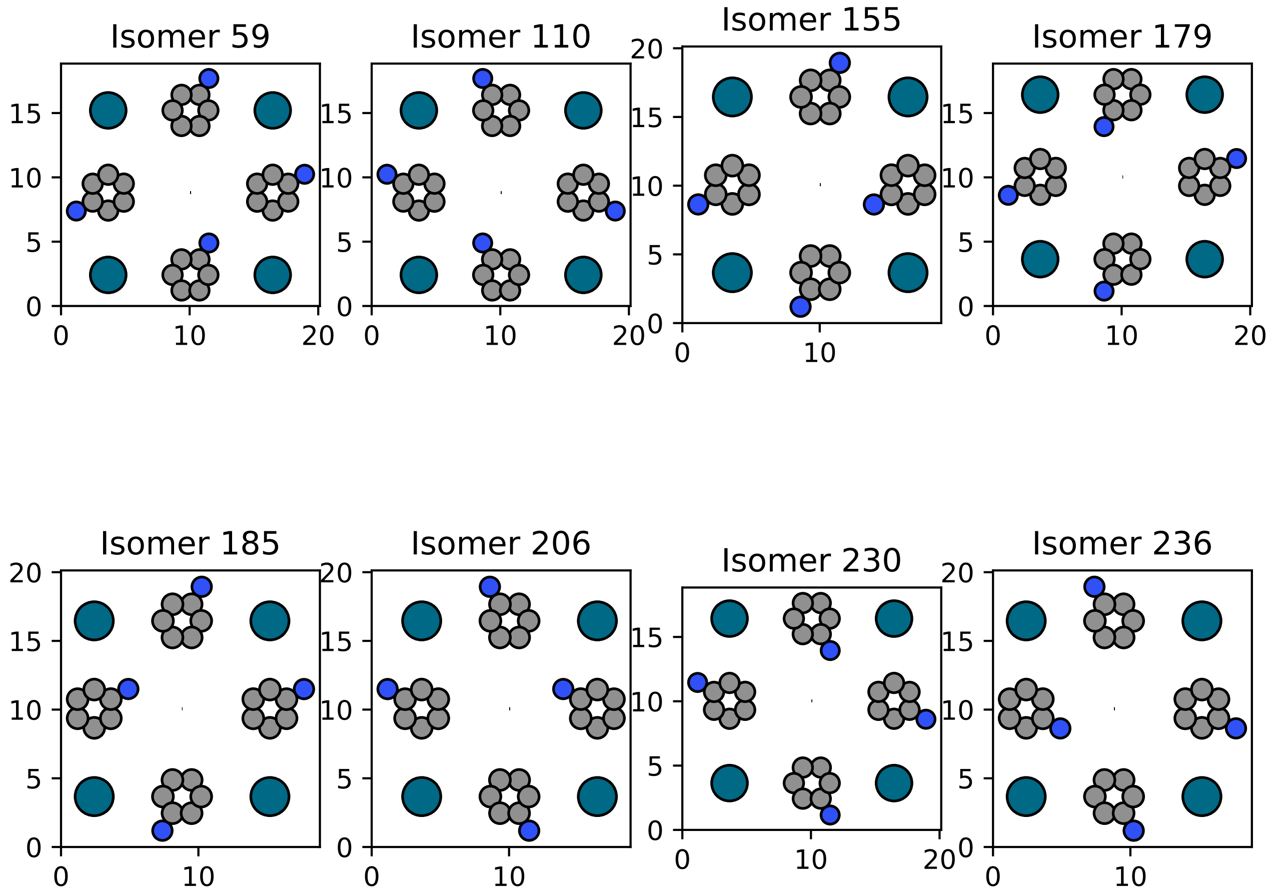


Isomer number 59

Number of FG in pore: 1

Isomer FG-FG distances (Å):

9.1602
12.7870
10.6034
10.5297
17.9793
14.5733



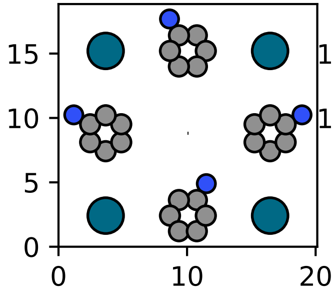
Isomer number 62

Number of FG in pore: 1

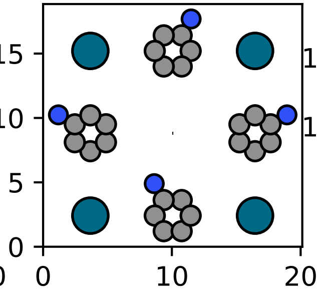
Isomer FG-FG distances (Å):

9.1602
13.1037
11.6104
12.7133
17.7505
10.5297

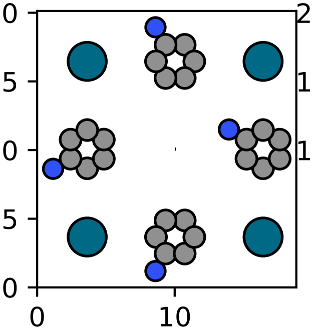
Isomer 62



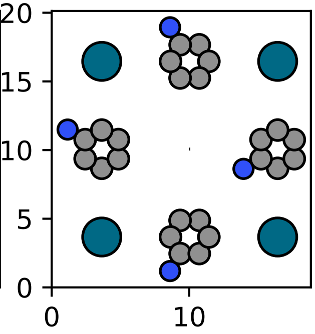
Isomer 122



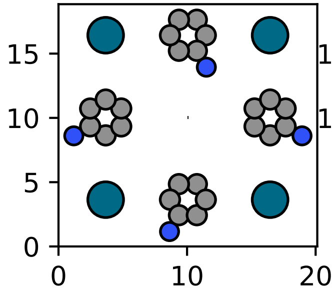
Isomer 143



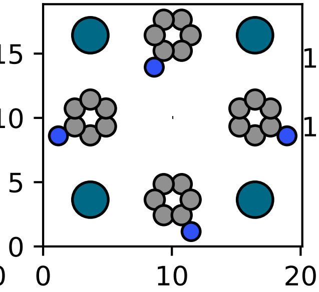
Isomer 158



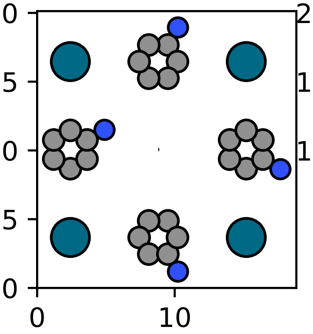
Isomer 167



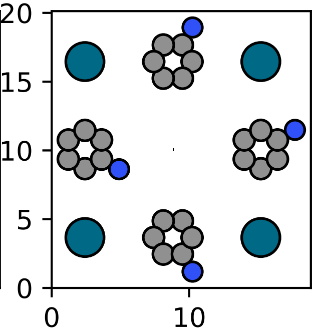
Isomer 227



Isomer 233



Isomer 248

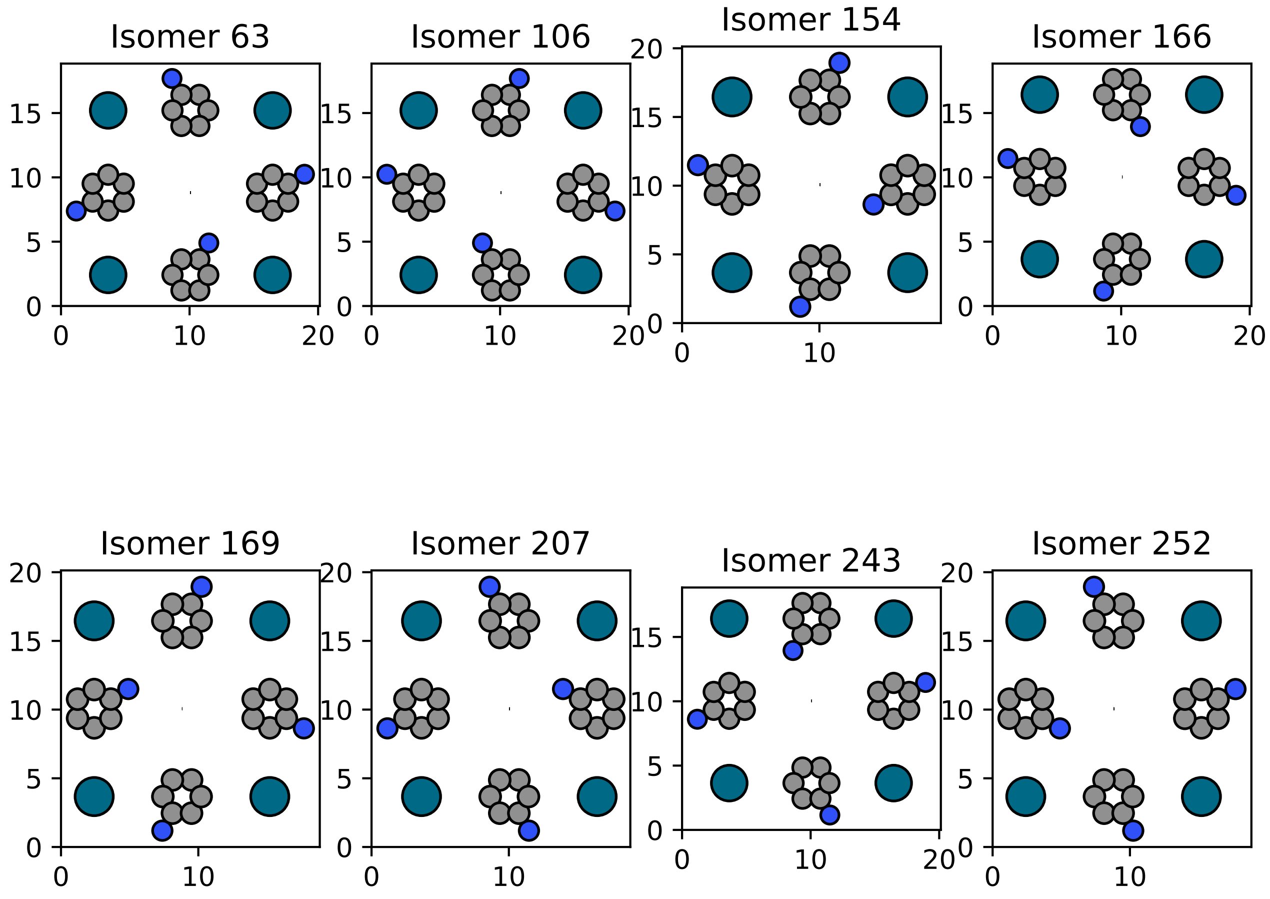


Isomer number 63

Number of FG in pore: 1

Isomer FG-FG distances (Å):

9.1602
13.1037
10.6034
12.7133
17.9793
12.7133

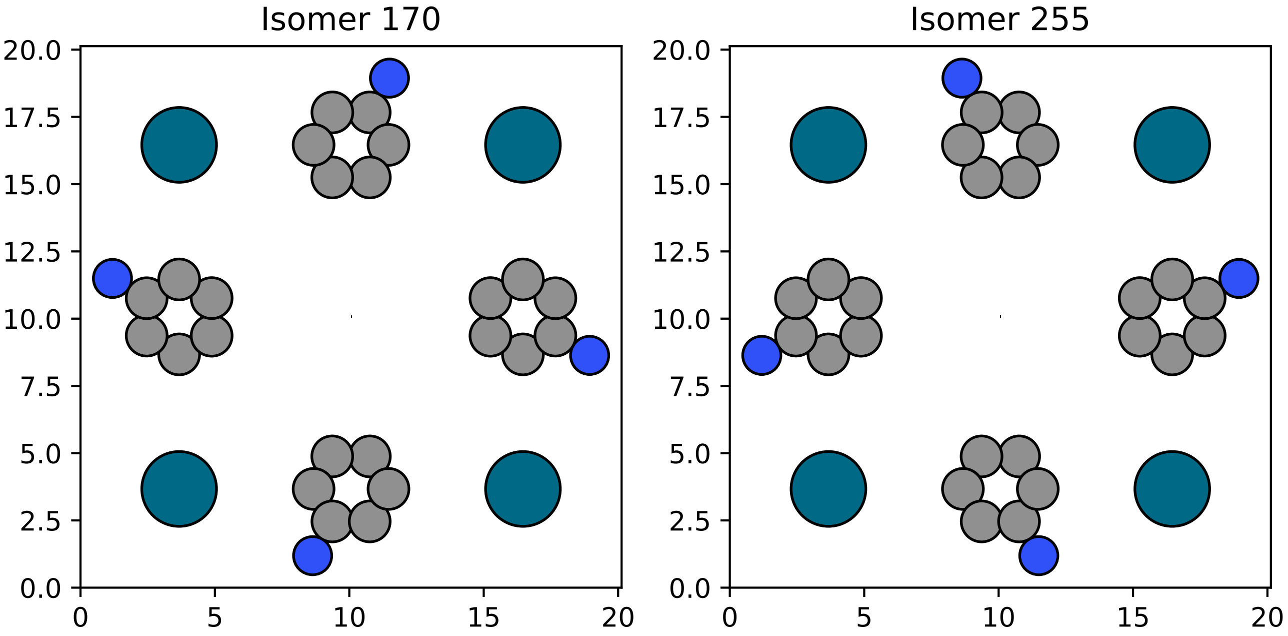


Isomer number 170

Number of FG in pore: 0

Isomer FG-FG distances (Å):

12.7133
17.9793
12.7133
12.7133
17.9793
12.7133

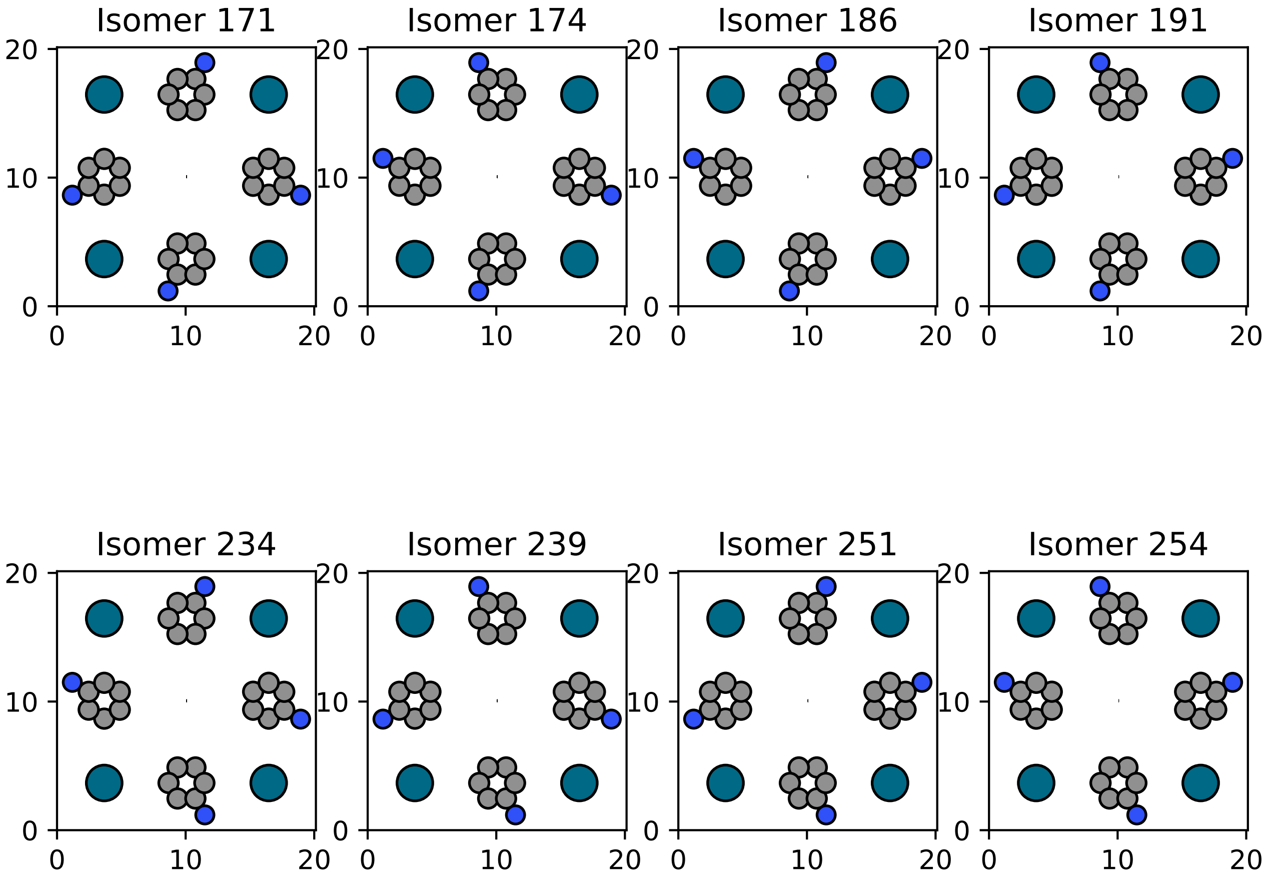


Isomer number 171

Number of FG in pore: 0

Isomer FG-FG distances (Å):

12.7133
17.9793
10.5297
12.7133
17.7505
14.5733

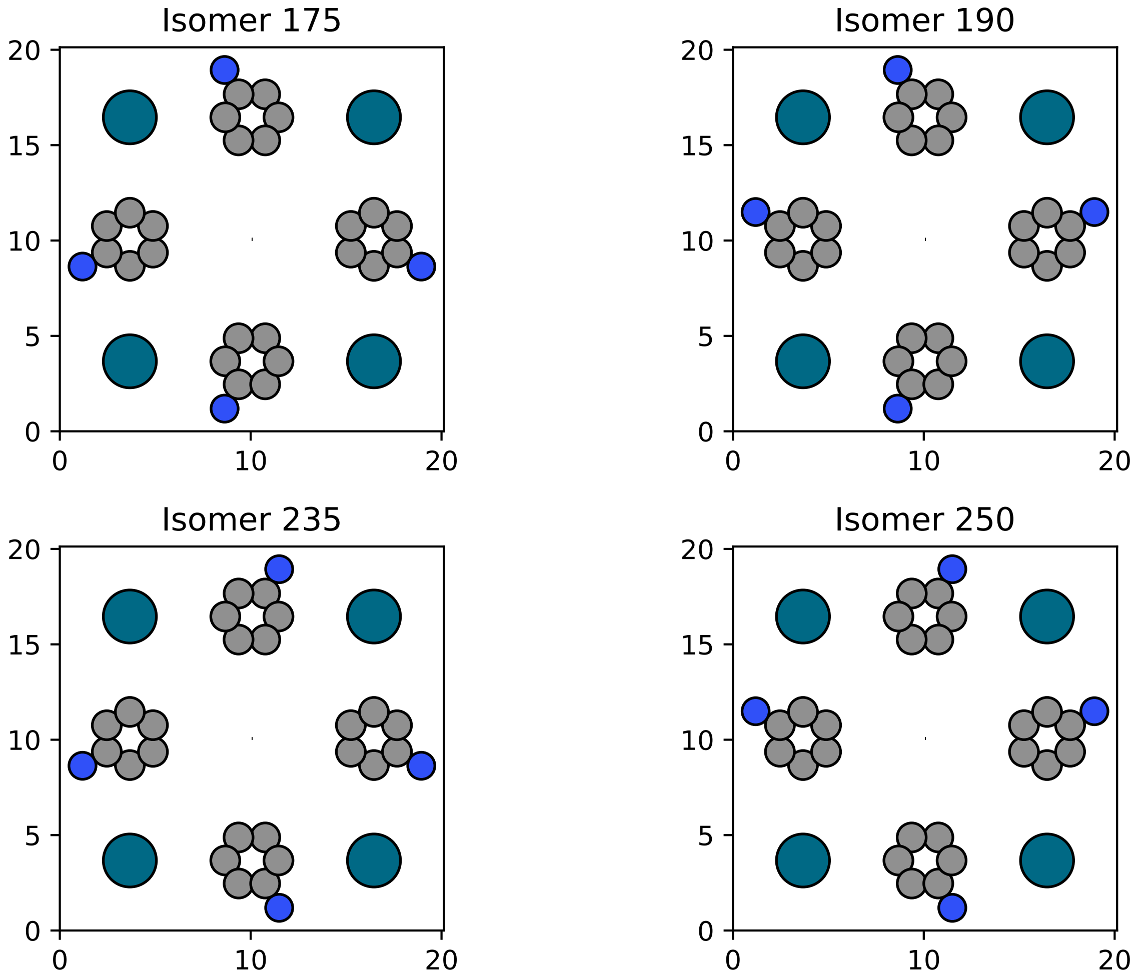


Isomer number 175

Number of FG in pore: 0

Isomer FG-FG distances (Å):

12.7133
17.7505
10.5297
14.5733
17.7505
12.7133

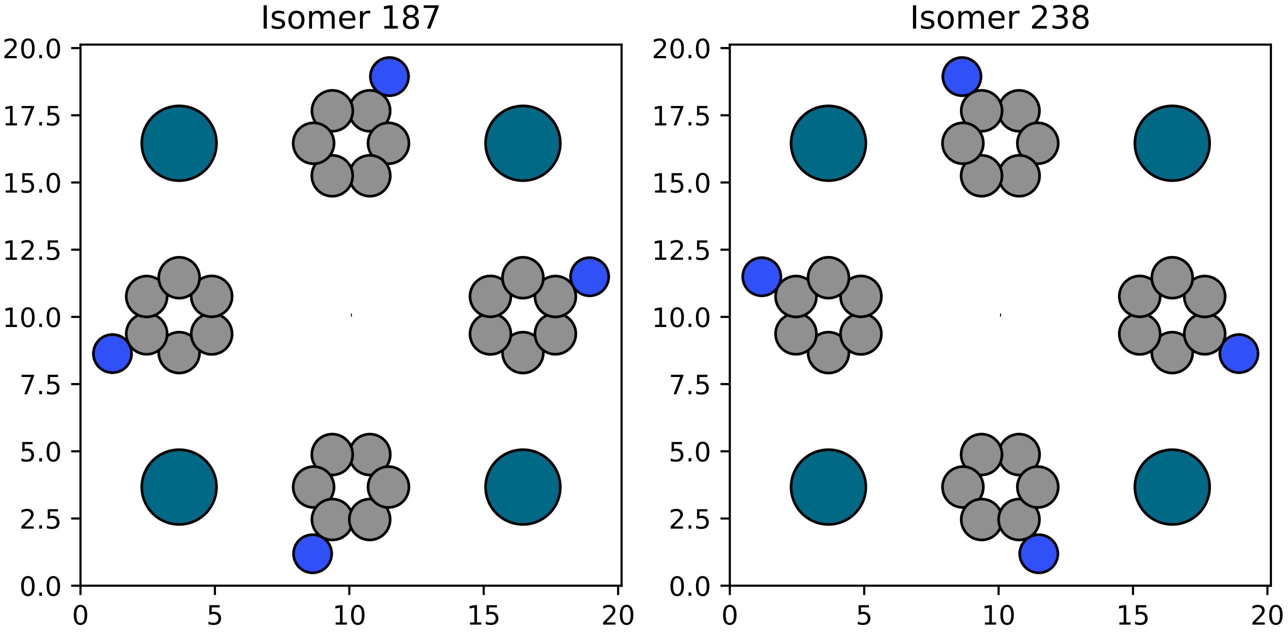


Isomer number 187

Number of FG in pore: 0

Isomer FG-FG distances (Å):

14.5733
17.9793
10.5297
10.5297
17.9793
14.5733

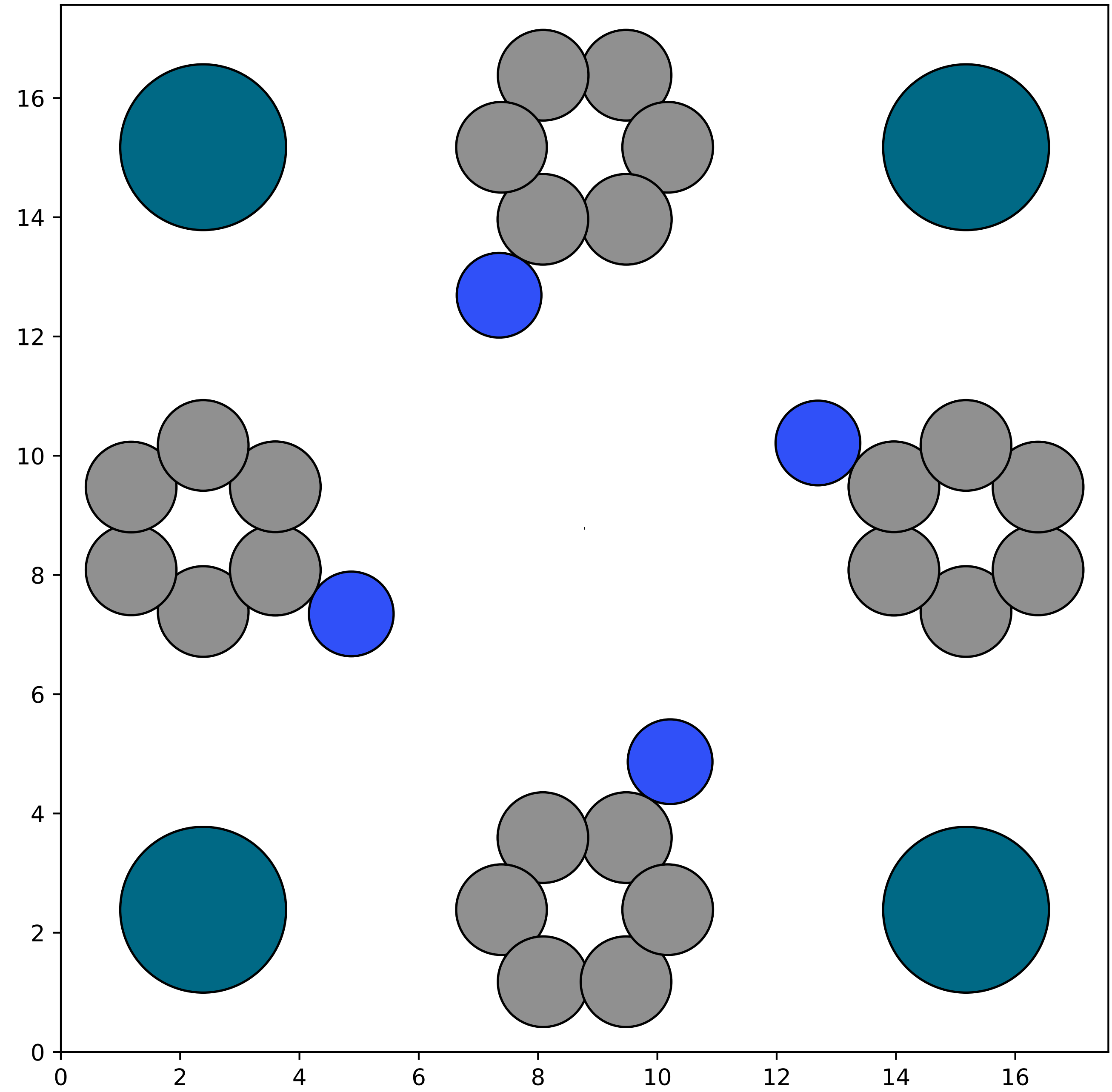


Unique Isomers with 4 FG inside the pore

Isomer number 0

Isomer FG-FG distances (Å):

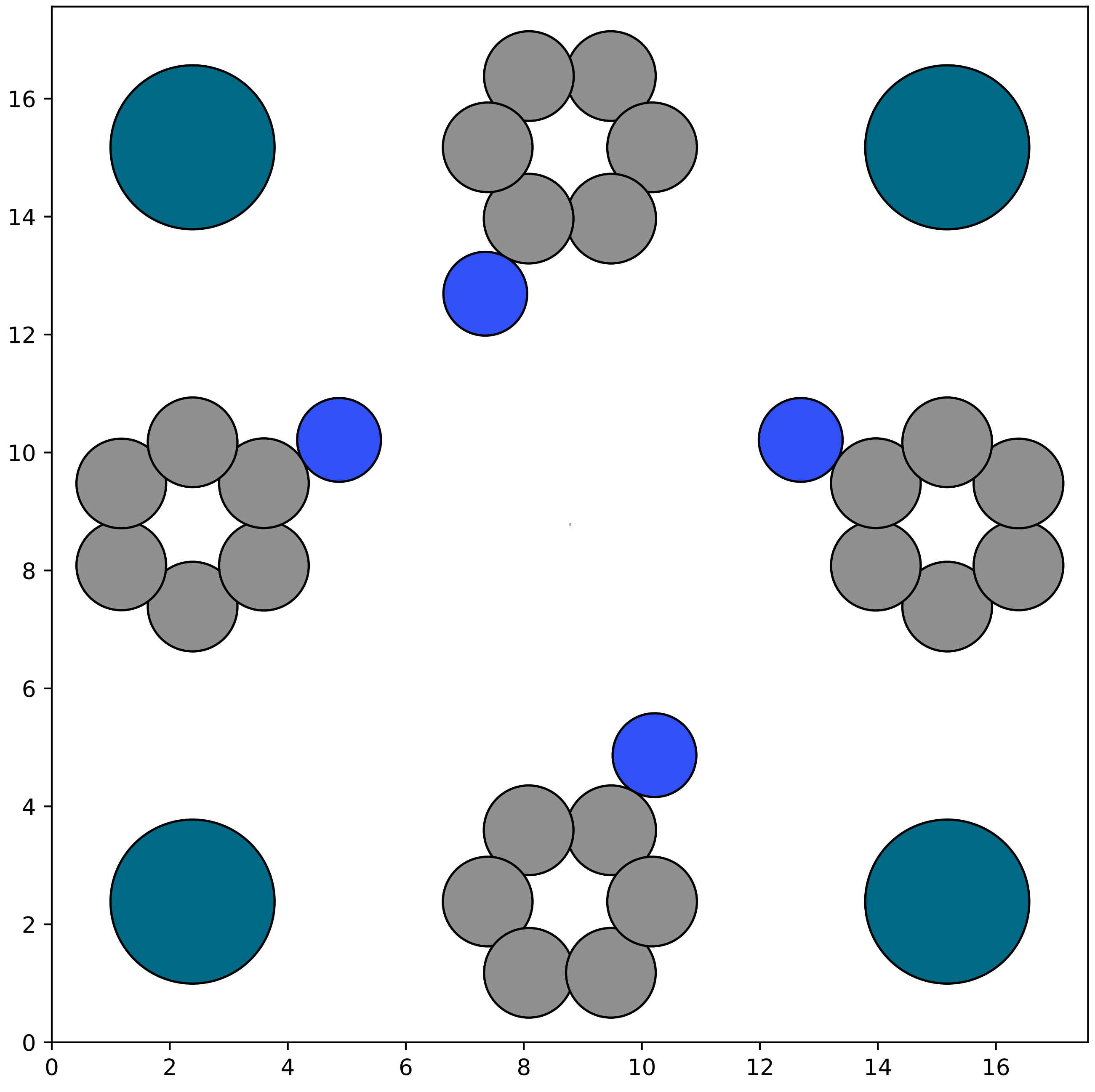
5.8919
8.3324
5.8919
5.8919
8.3324
5.8919



Isomer number 1

Isomer FG-FG distances (Å):

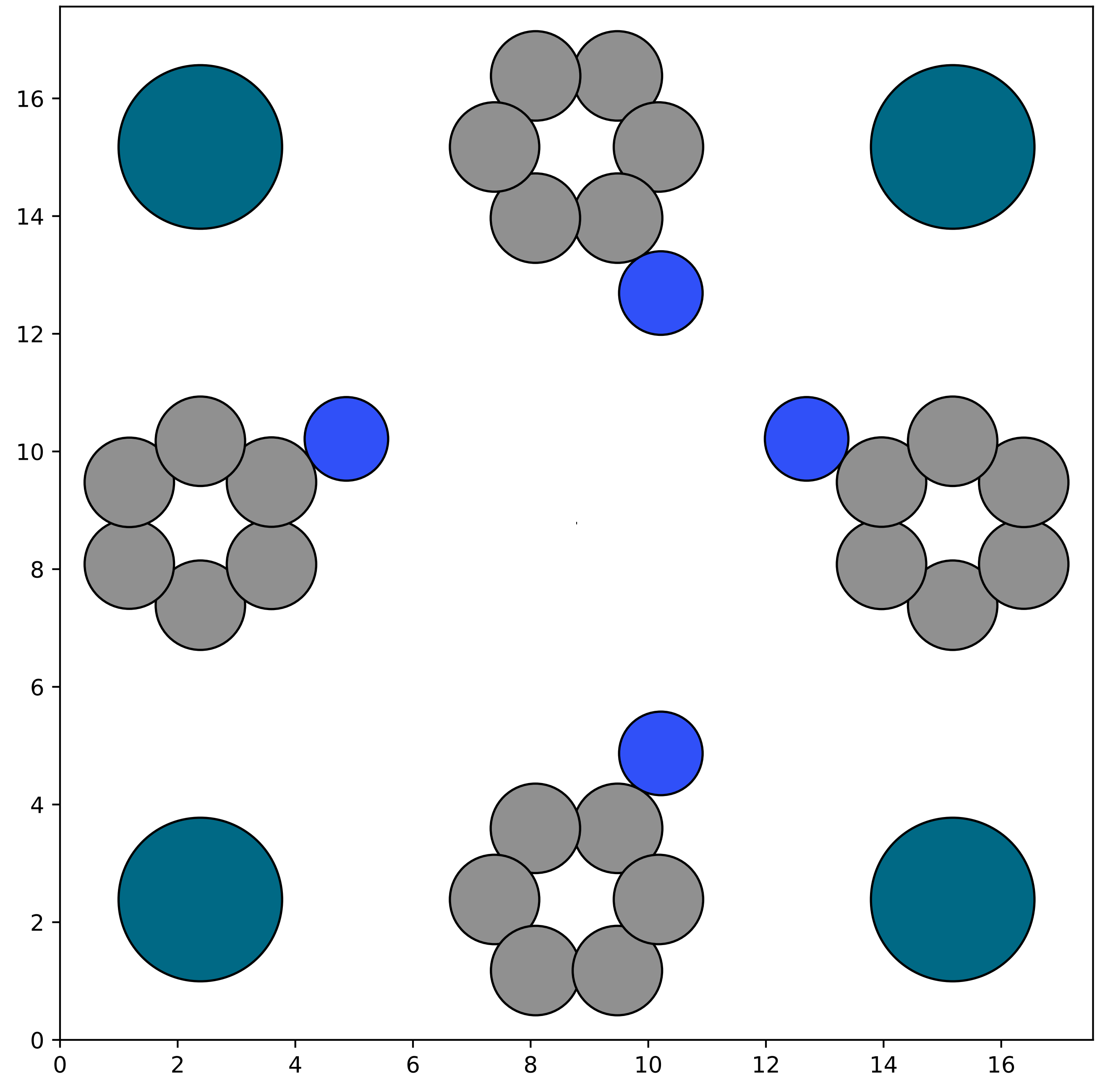
5.8919
8.3324
7.5595
5.8919
7.8236
3.5048



Isomer number 5

Isomer FG-FG distances (Å):

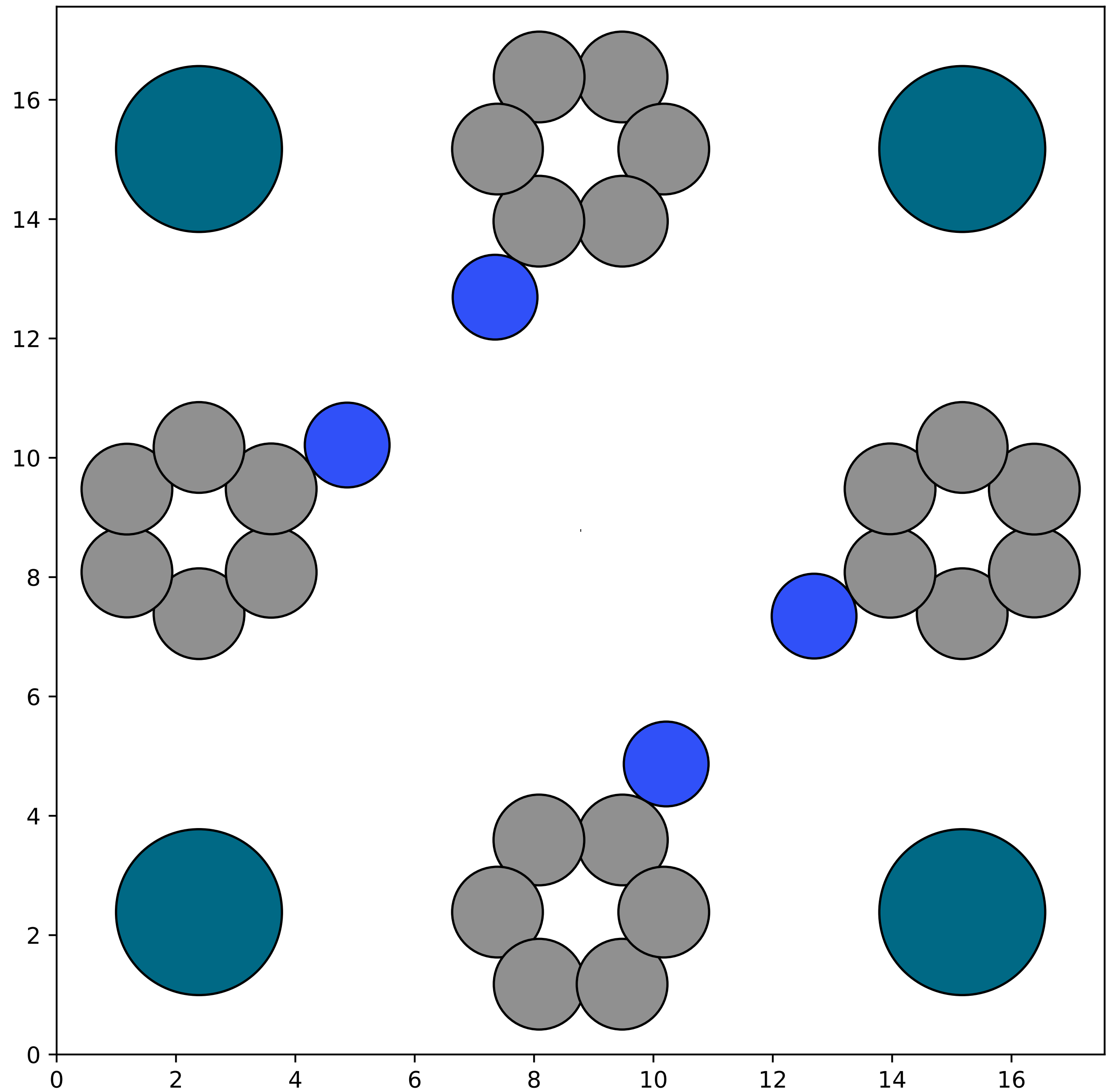
5.8919
7.8236
7.5595
3.5048
7.8236
5.8919



Isomer number 17

Isomer FG-FG distances (Å):

3.5048
8.3324
7.5595
7.5595
8.3324
3.5048

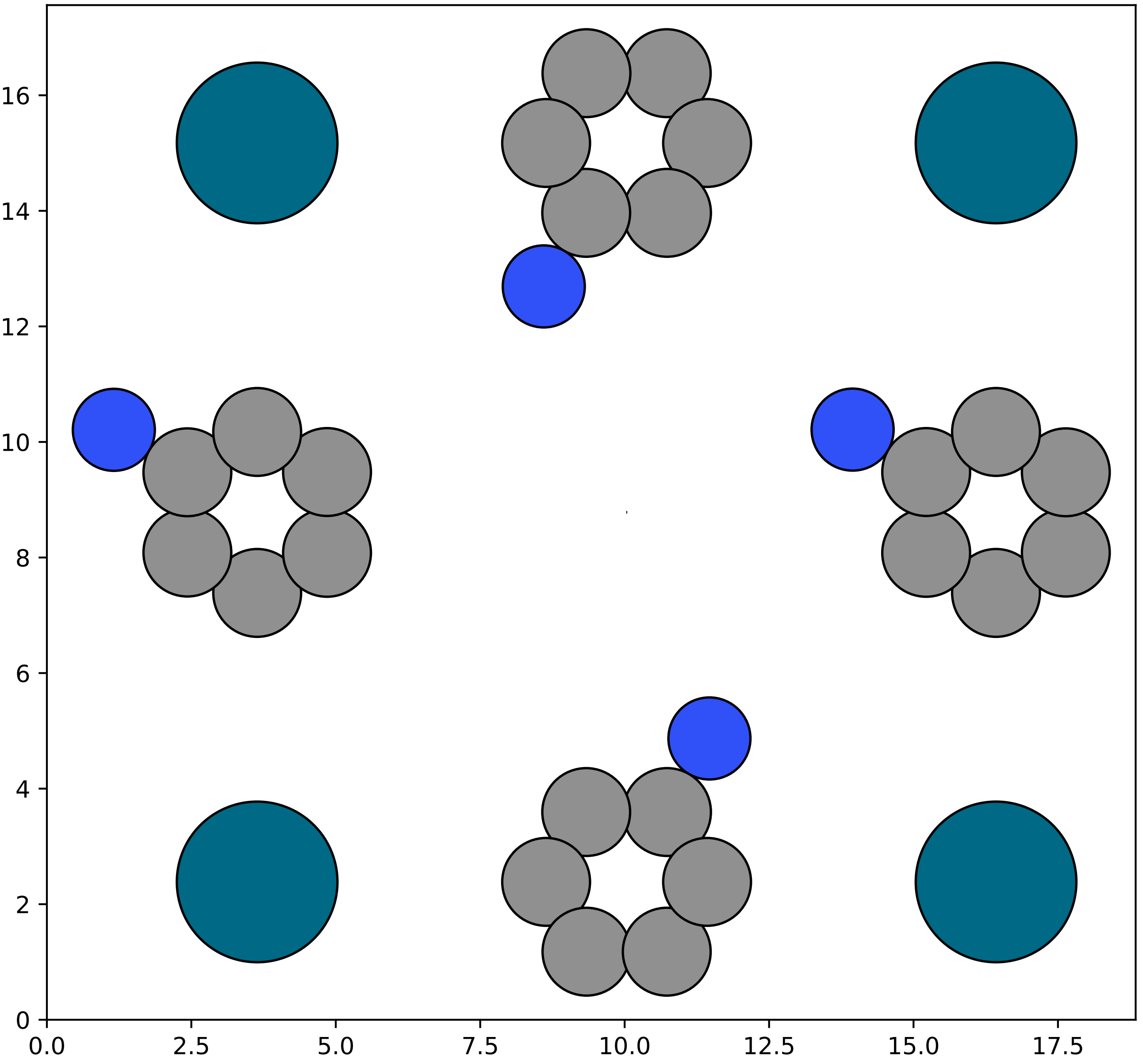


Unique Isomers with 3 FG inside the pore

Isomer number 2

Isomer FG-FG distances (Å):

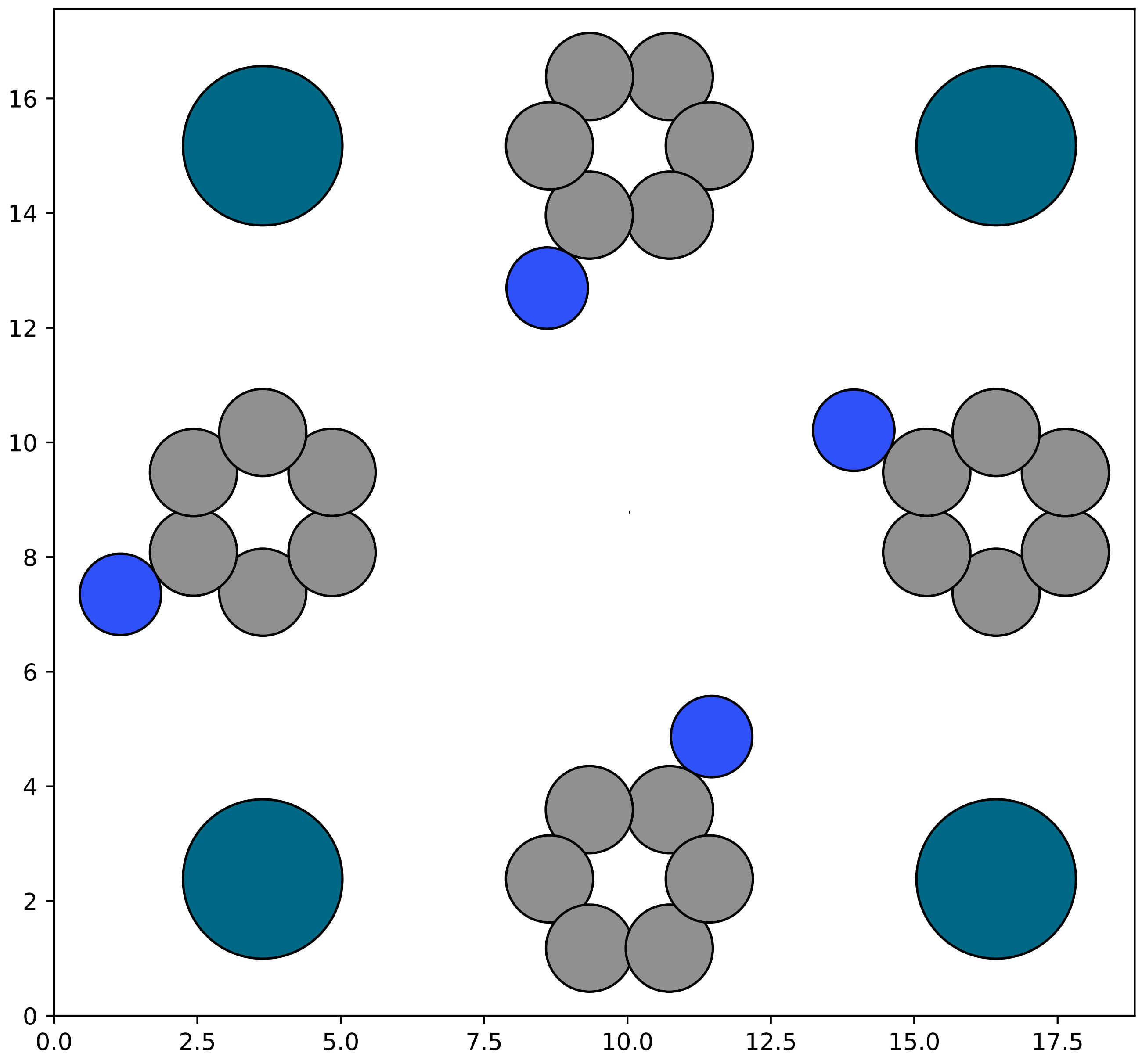
5.8919
8.3324
11.6104
5.8919
12.7870
7.8447



Isomer number 3

Isomer FG-FG distances (Å):

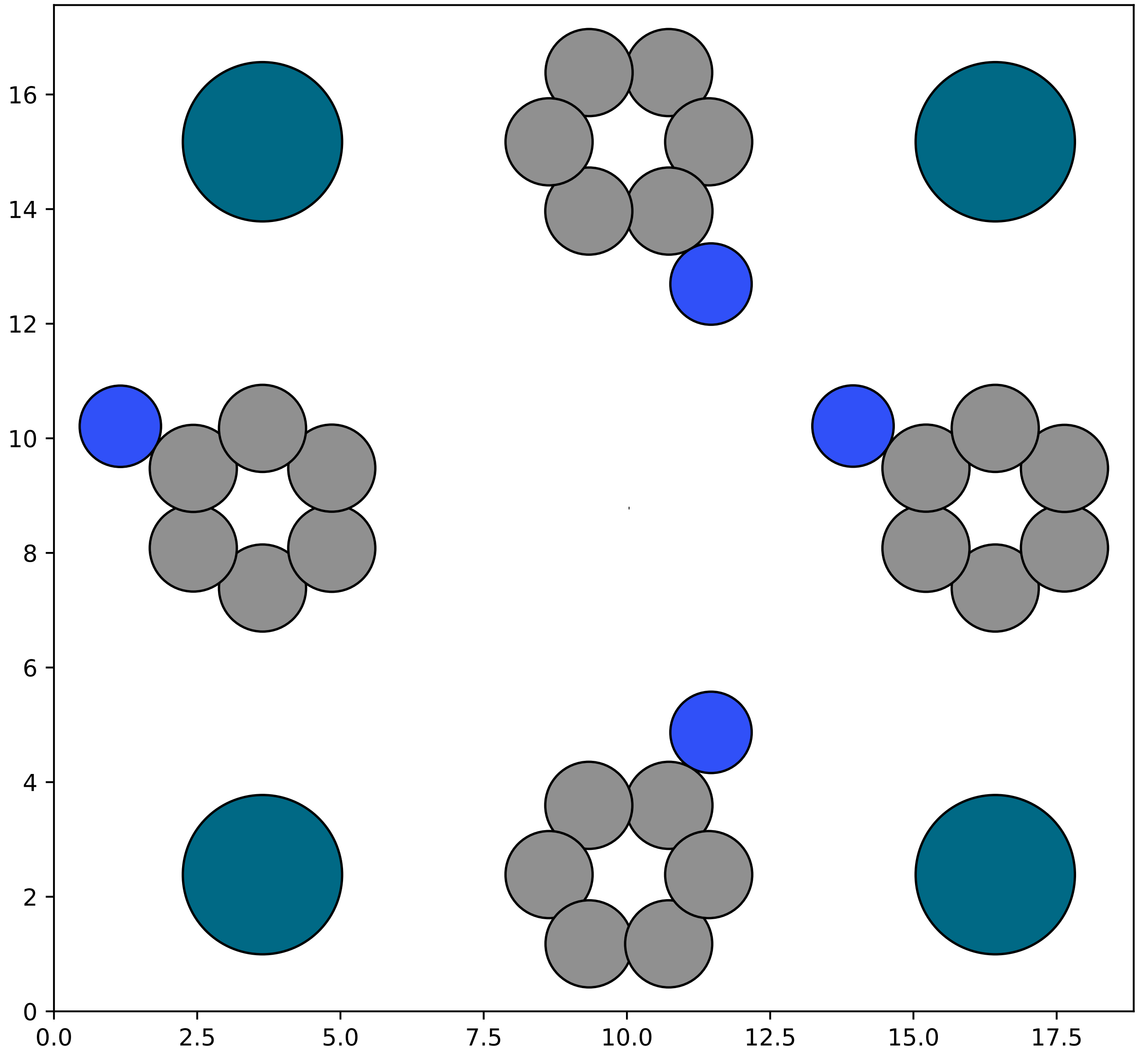
5.8919
8.3324
10.6034
5.8919
13.1037
9.1602



Isomer number 6

Isomer FG-FG distances (Å):

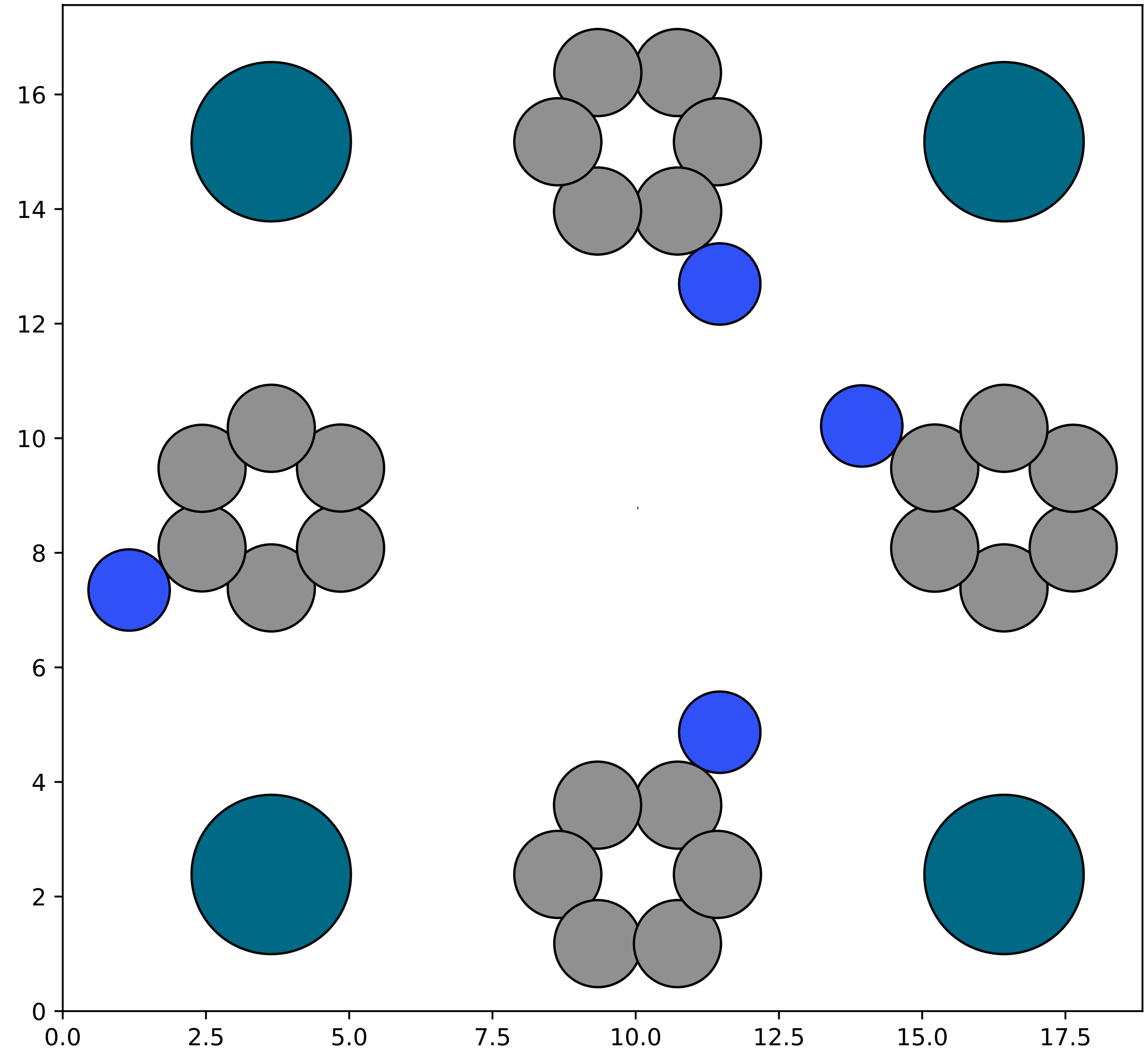
5.8919
7.8236
11.6104
3.5048
12.7870
10.6034



Isomer number 7

Isomer FG-FG distances (Å):

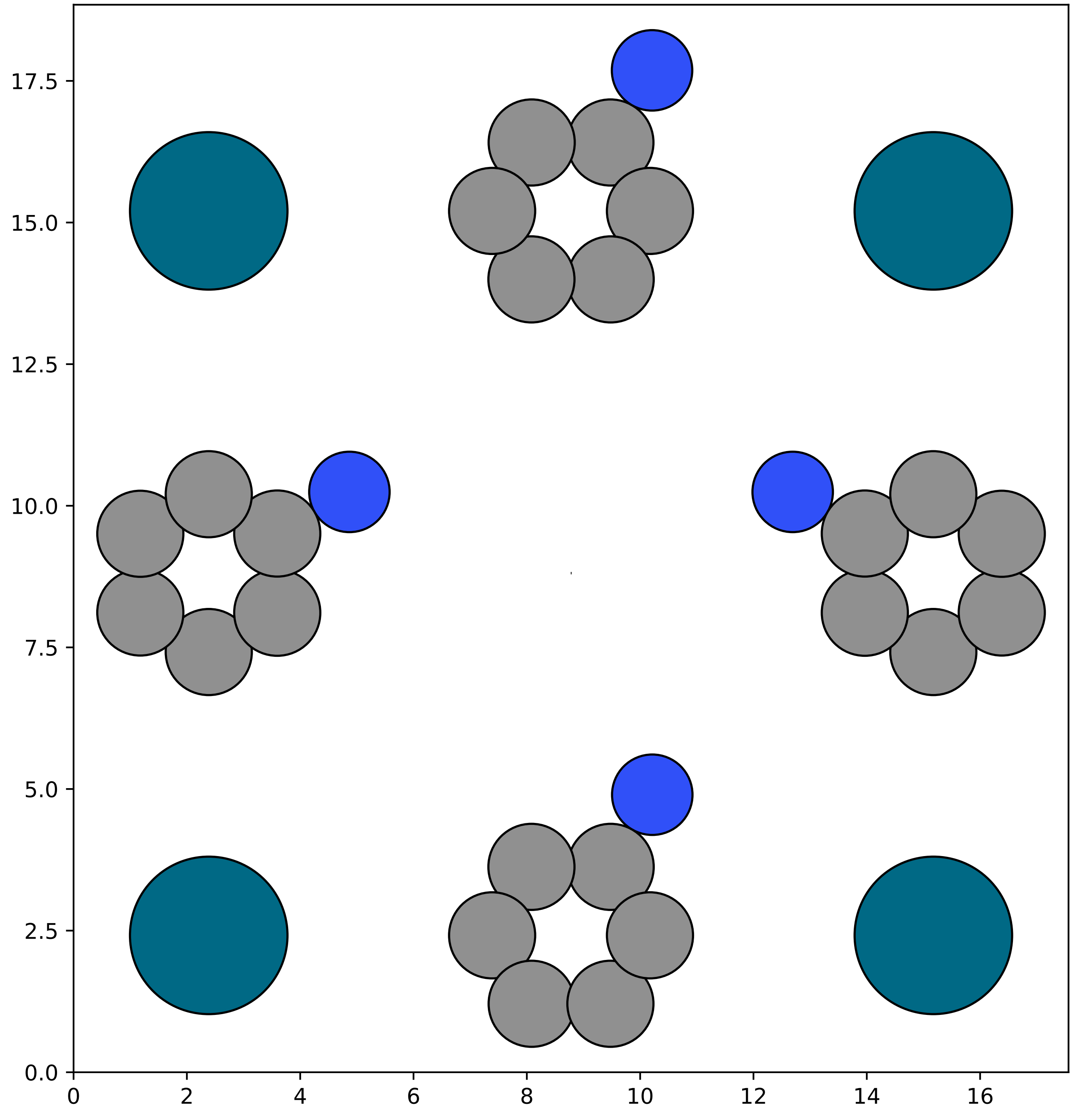
5.8919
7.8236
10.6034
3.5048
13.1037
11.6104



Isomer number 9

Isomer FG-FG distances (Å):

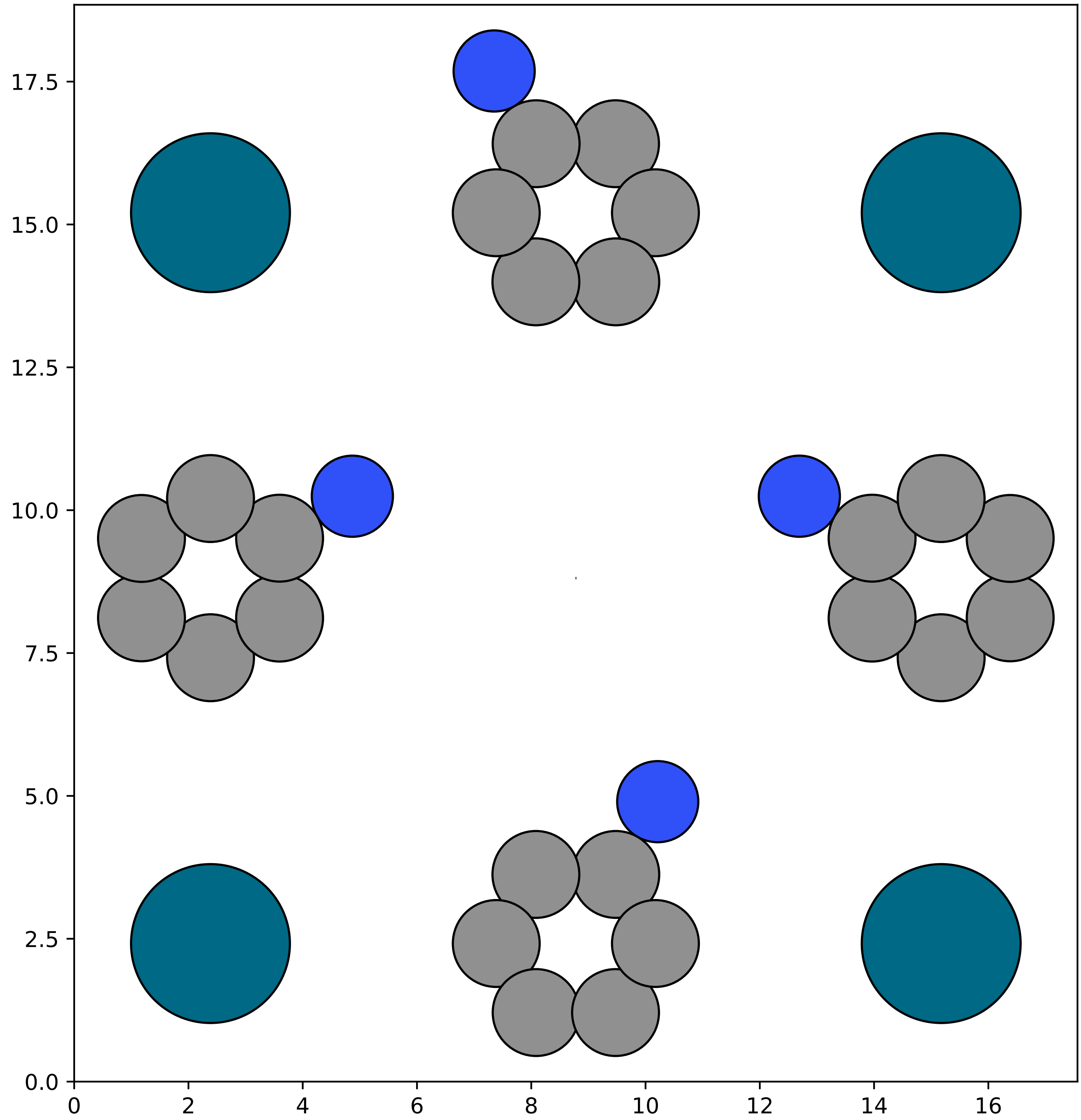
5.8919
12.7870
7.5595
7.8447
7.8236
9.1602



Isomer number 13

Isomer FG-FG distances (Å):

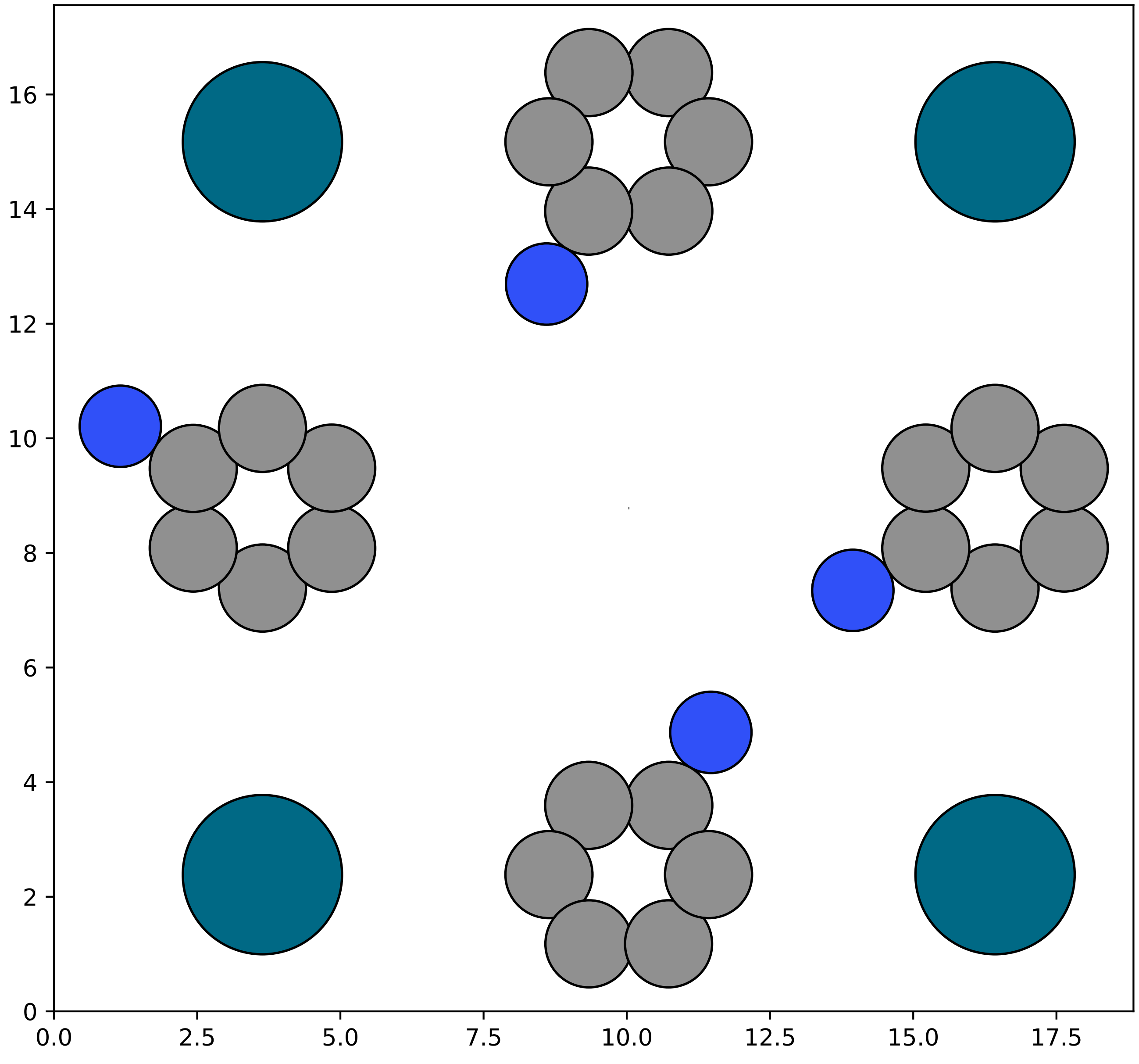
5.8919
13.1037
7.5595
9.1602
7.8236
7.8447



Isomer number 18

Isomer FG-FG distances (Å):

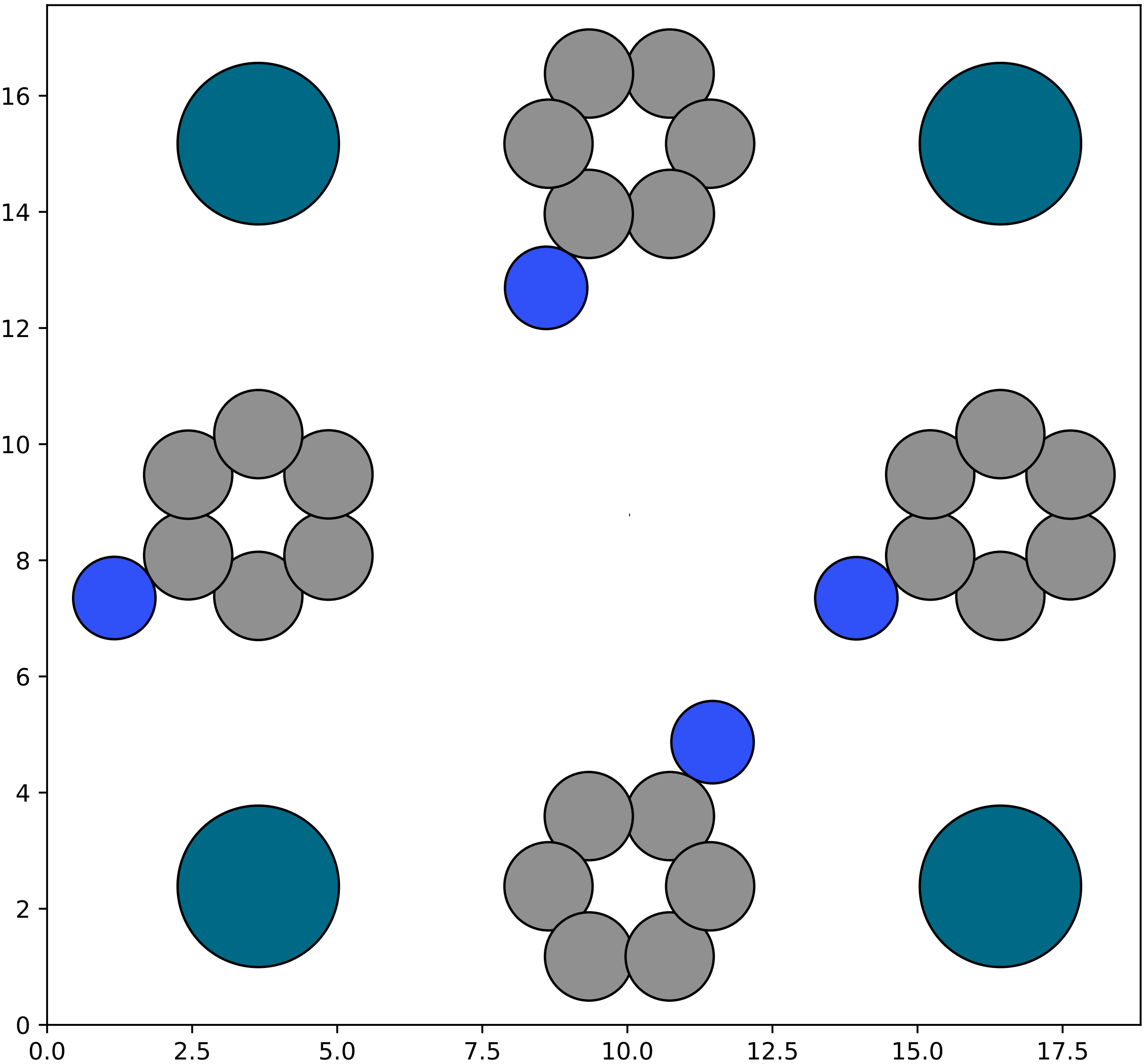
3.5048
8.3324
11.6104
7.5595
13.1037
7.8447



Isomer number 19

Isomer FG-FG distances (Å):

3.5048
8.3324
10.6034
7.5595
12.7870
9.1602

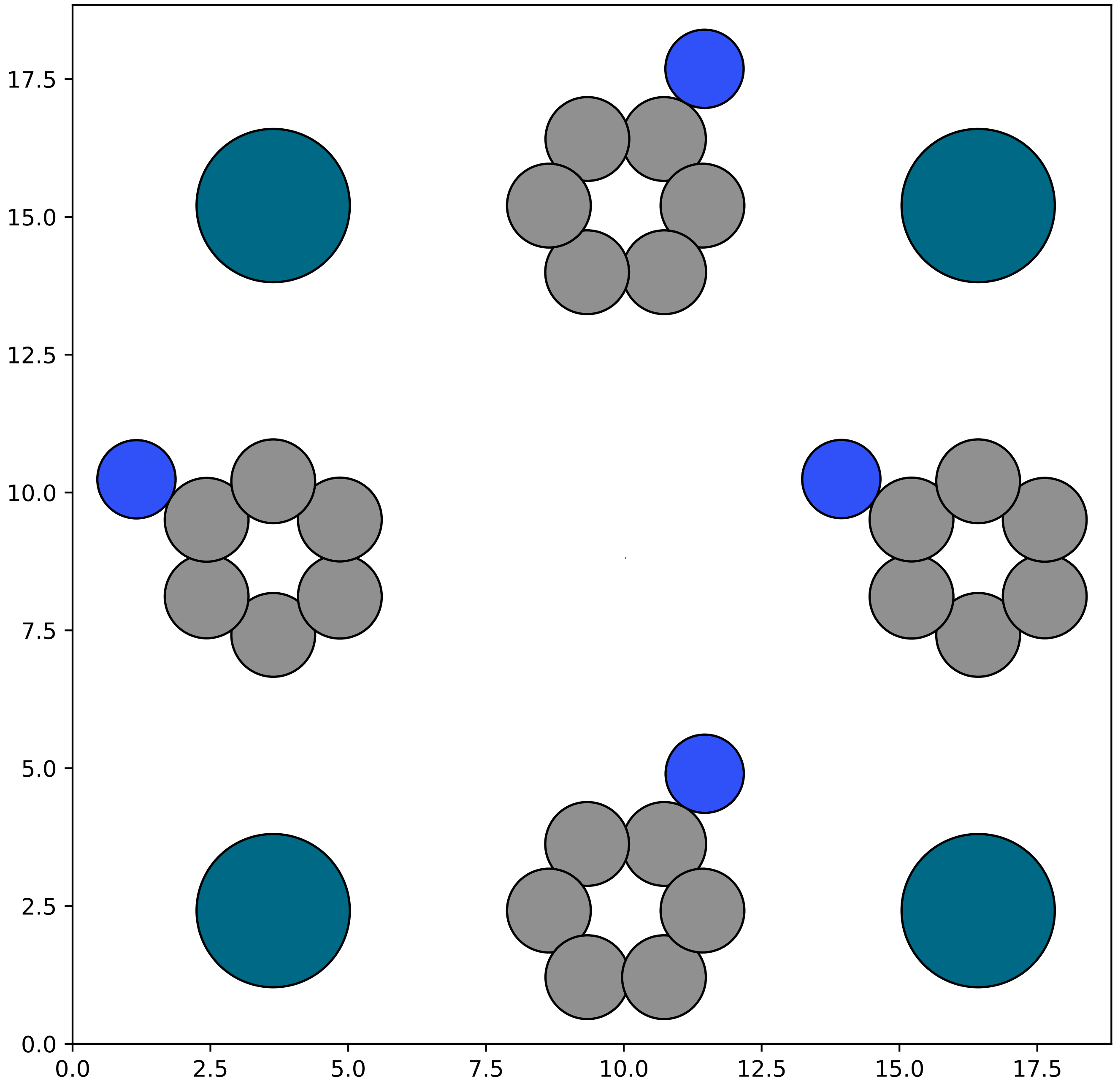


Unique Isomers with 2 FG inside the pore

Isomer number 10

Isomer FG-FG distances (Å):

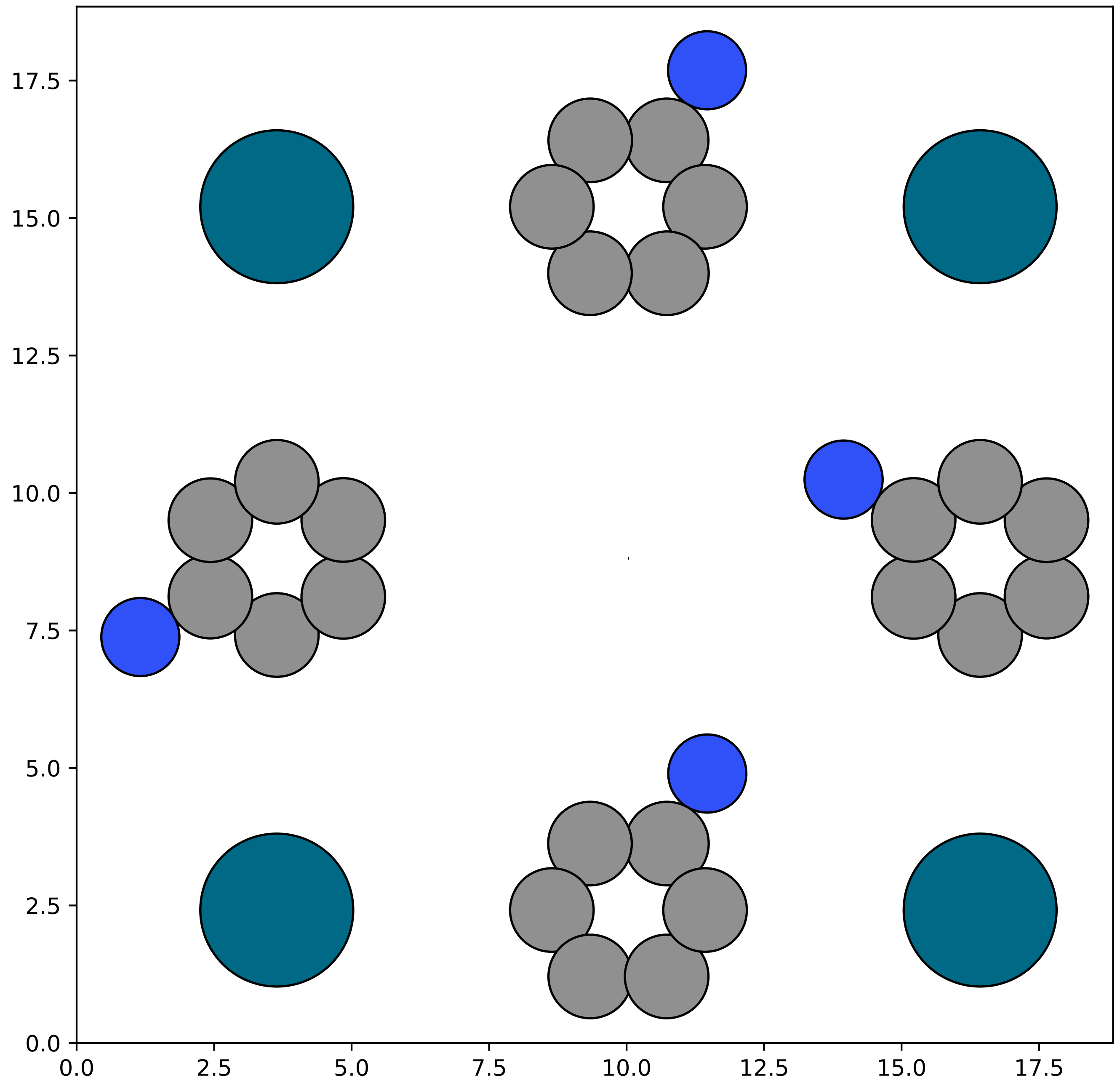
5.8919
12.7870
11.6104
7.8447
12.7870
12.7133



Isomer number 11

Isomer FG-FG distances (Å):

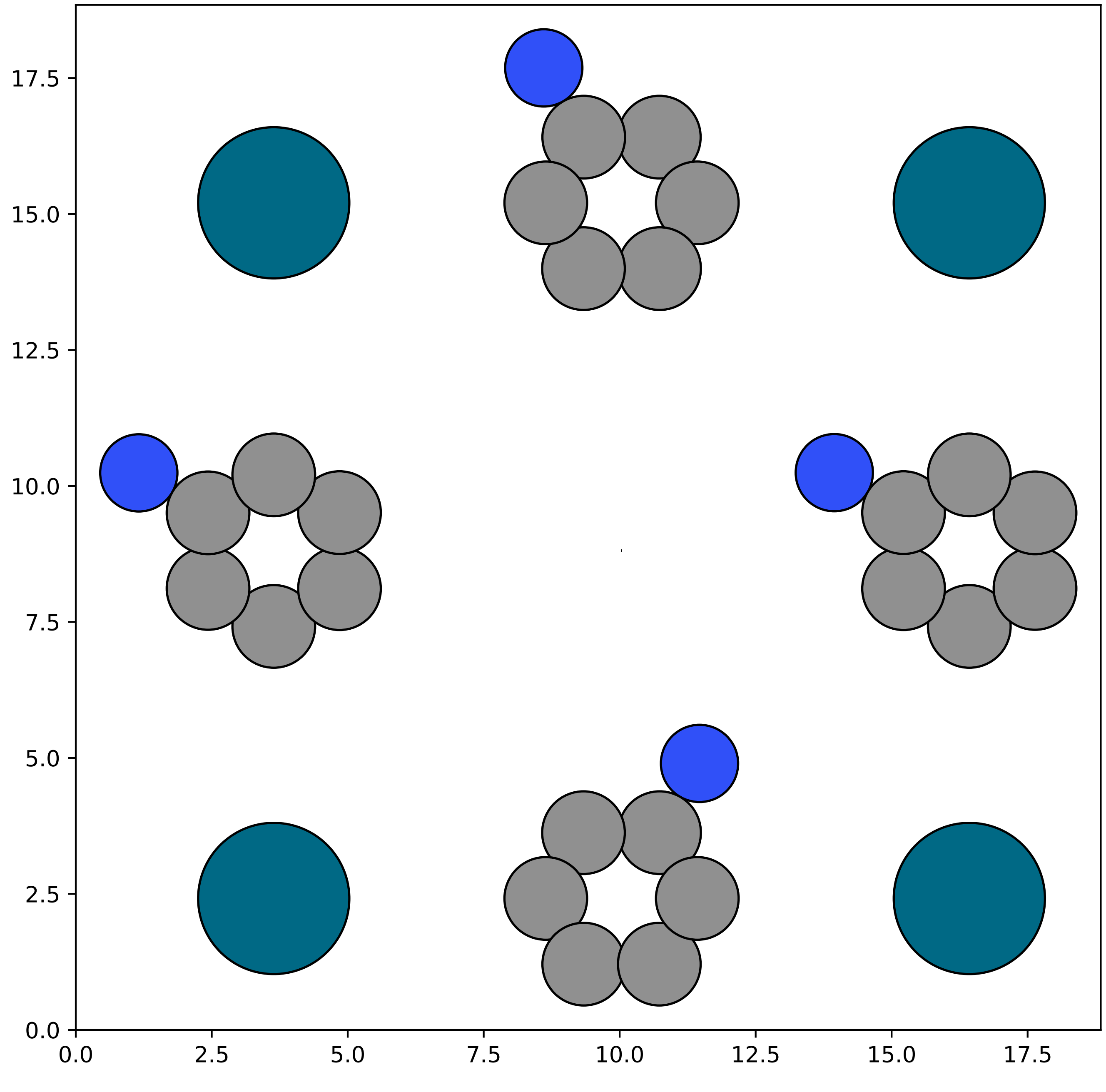
5.8919
12.7870
10.6034
7.8447
13.1037
14.5733



Isomer number 14

Isomer FG-FG distances (Å):

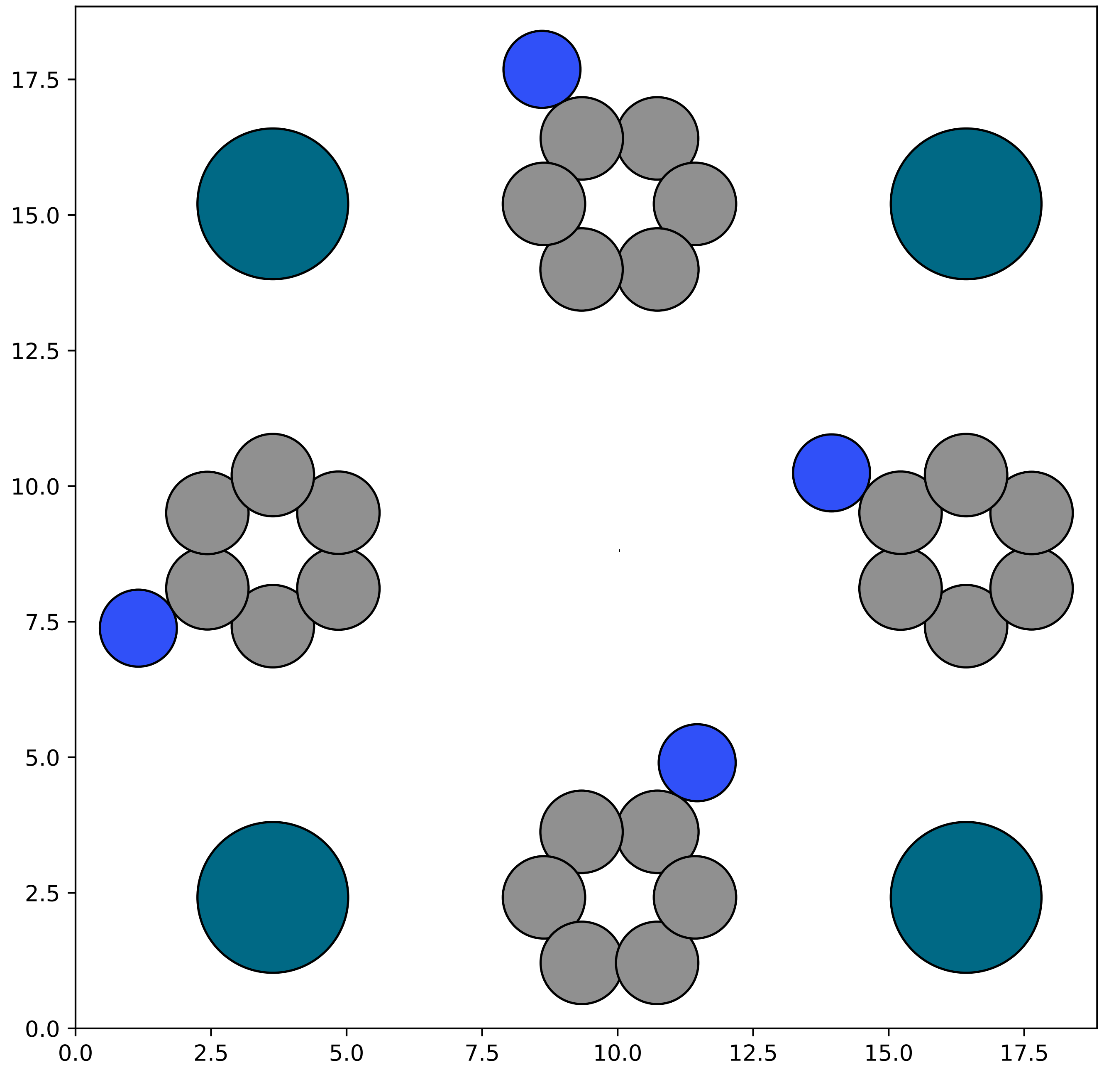
5.8919
13.1037
11.6104
9.1602
12.7870
10.5297



Isomer number 15

Isomer FG-FG distances (Å):

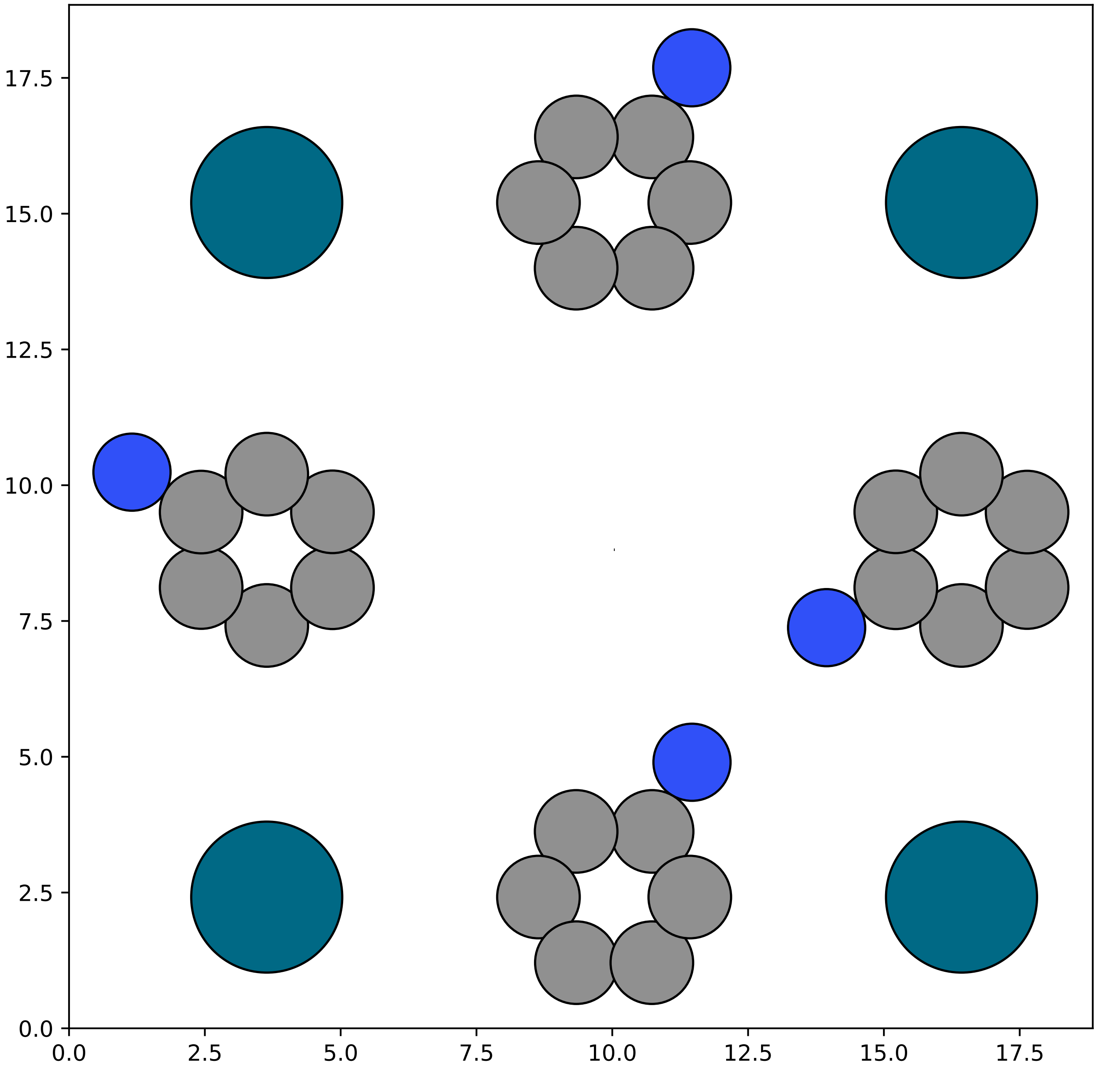
5.8919
13.1037
10.6034
9.1602
13.1037
12.7133



Isomer number 26

Isomer FG-FG distances (Å):

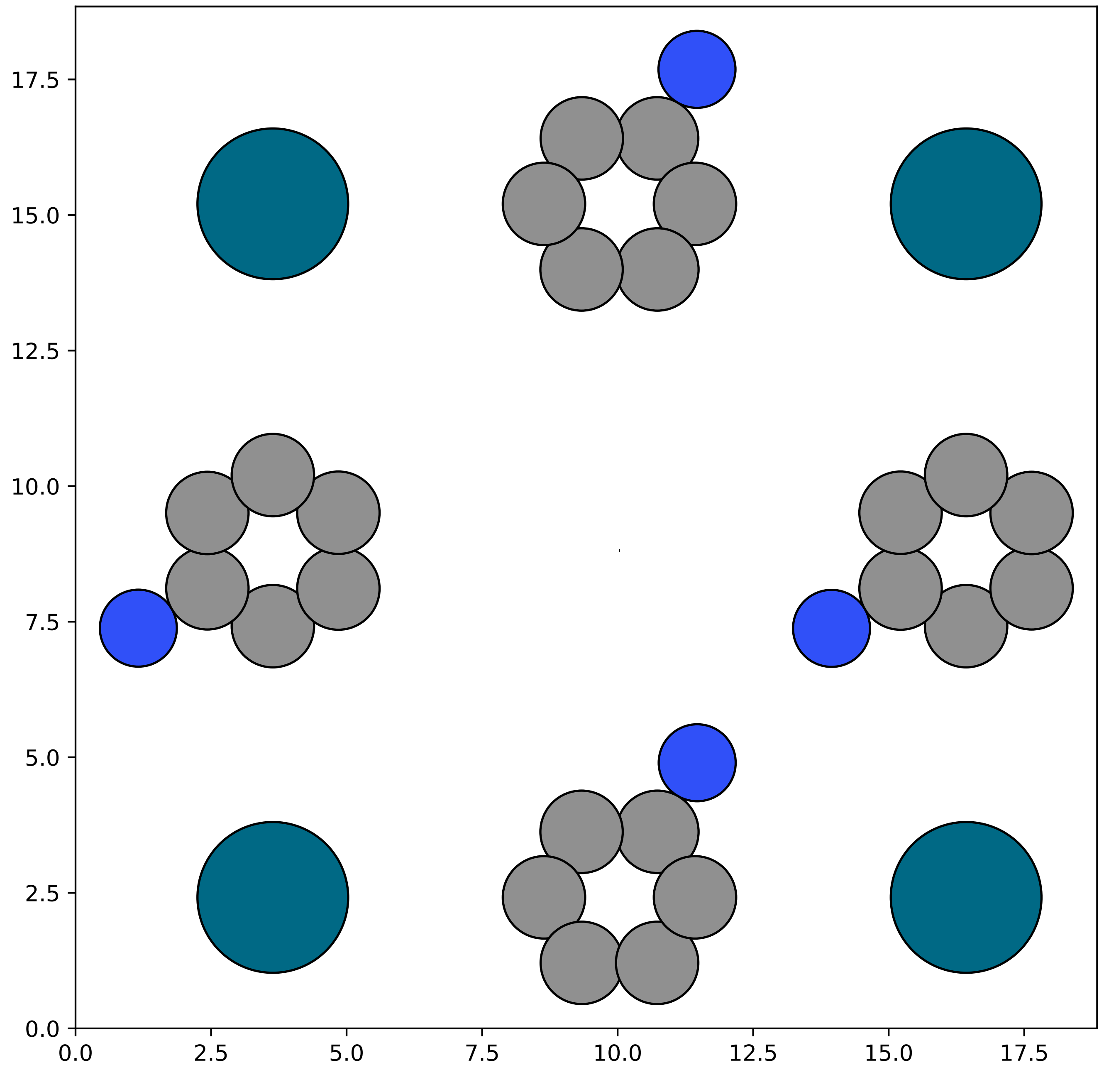
3.5048
12.7870
11.6104
10.6034
13.1037
12.7133



Isomer number 27

Isomer FG-FG distances (Å):

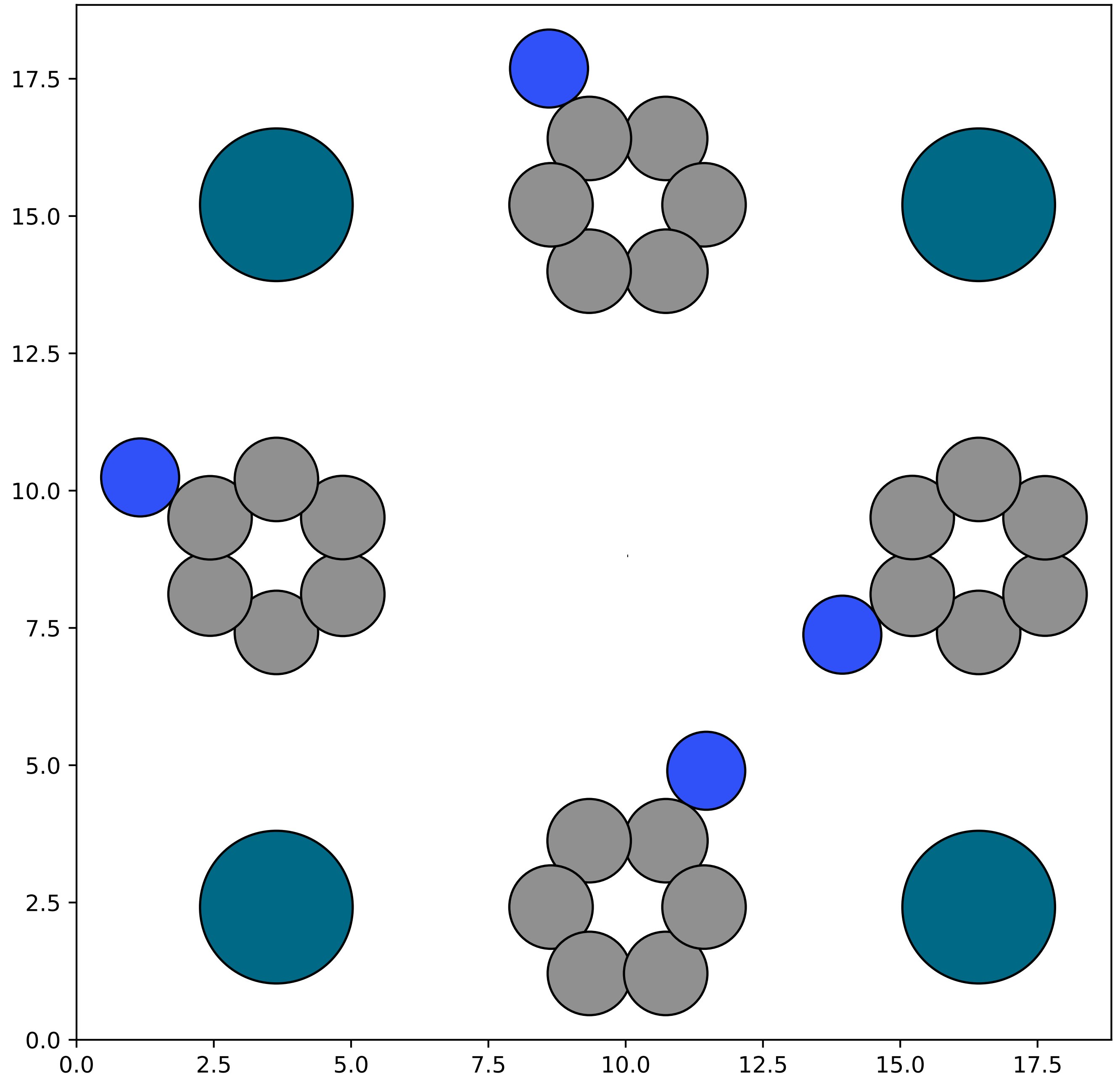
3.5048
12.7870
10.6034
10.6034
12.7870
14.5733



Isomer number 30

Isomer FG-FG distances (Å):

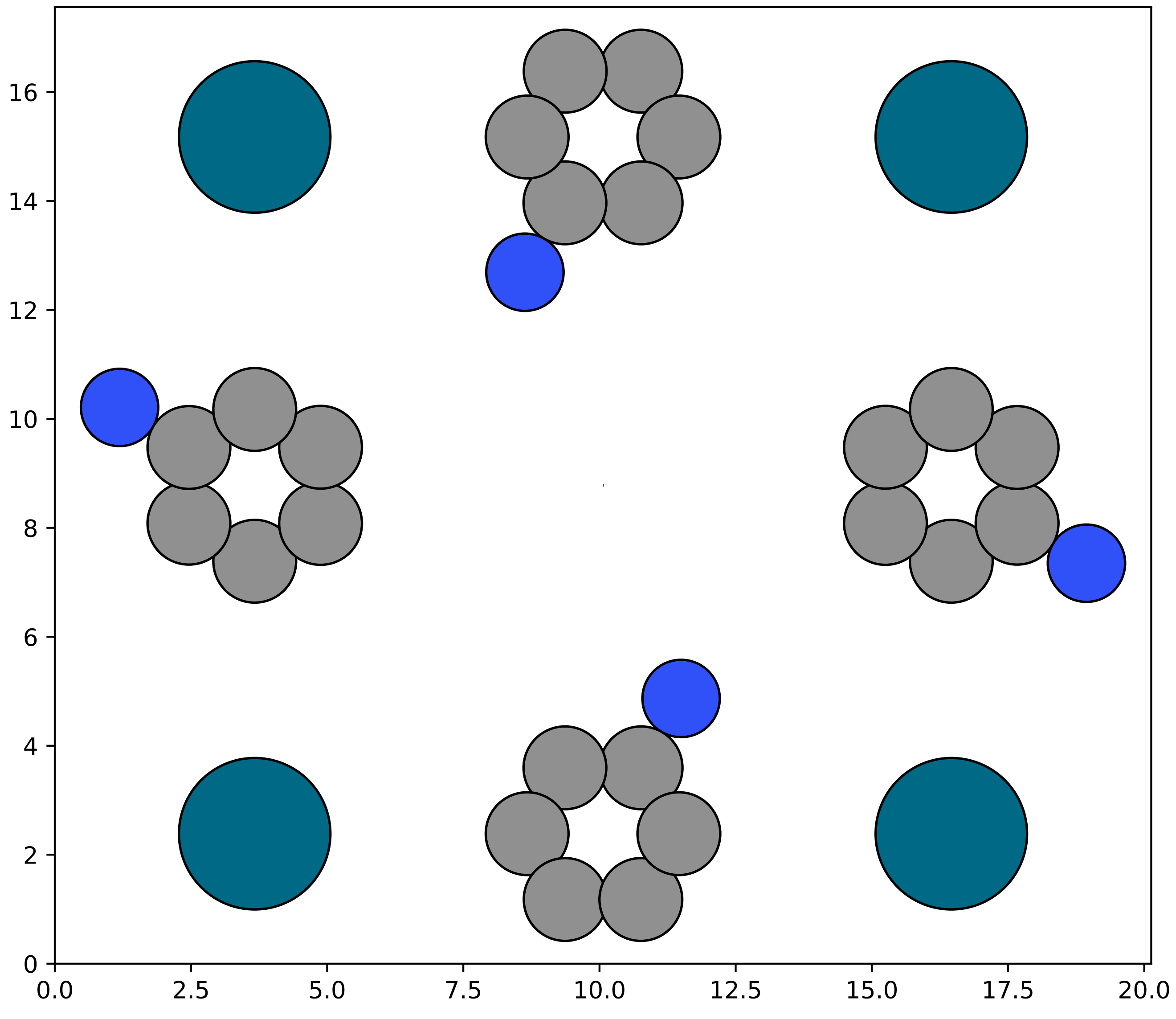
3.5048
13.1037
11.6104
11.6104
13.1037
10.5297



Isomer number 34

Isomer FG-FG distances (Å):

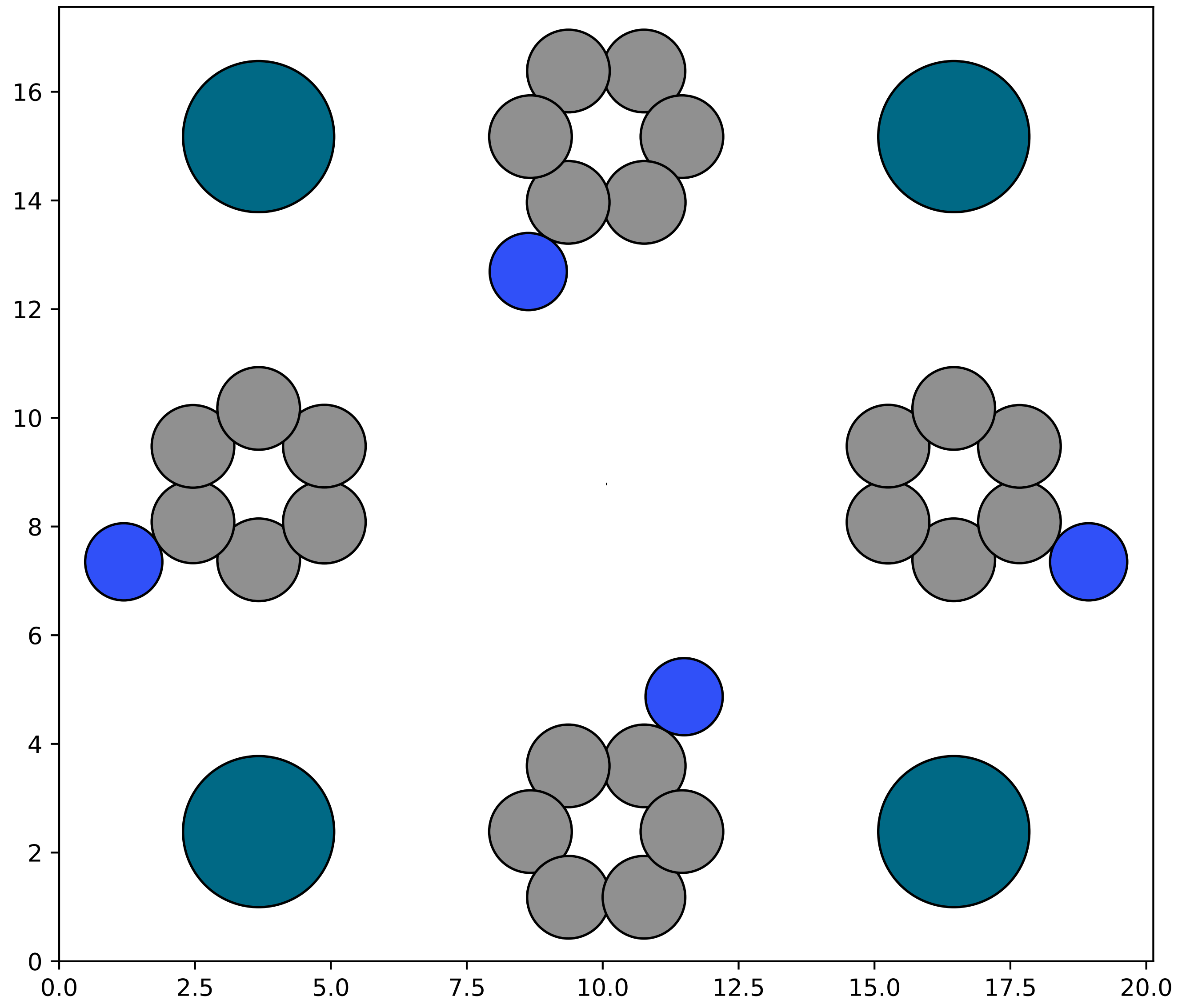
7.8447
8.3324
11.6104
11.6104
17.9793
7.8447



Isomer number 35

Isomer FG-FG distances (Å):

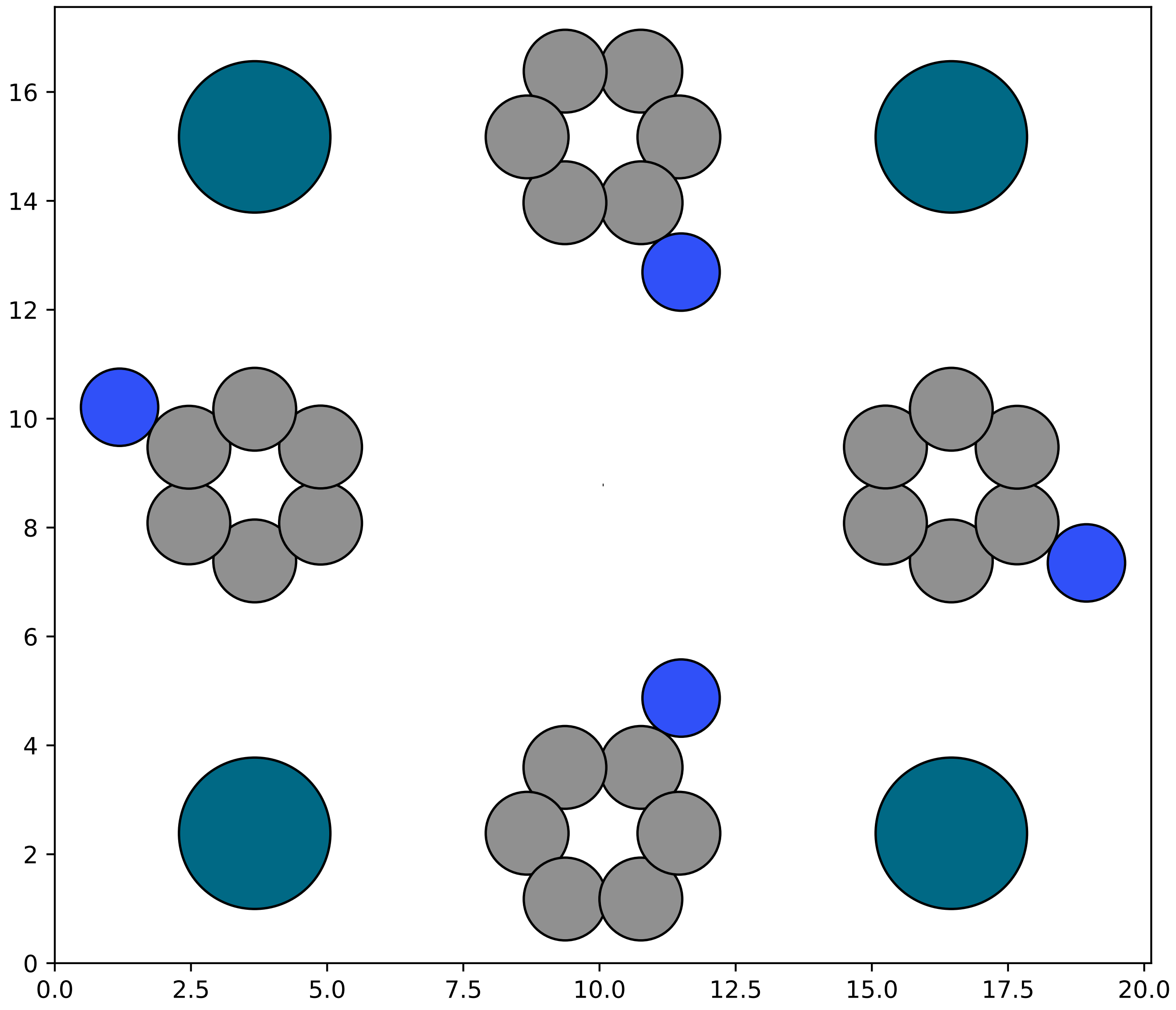
7.8447
8.3324
10.6034
11.6104
17.7505
9.1602



Isomer number 38

Isomer FG-FG distances (Å):

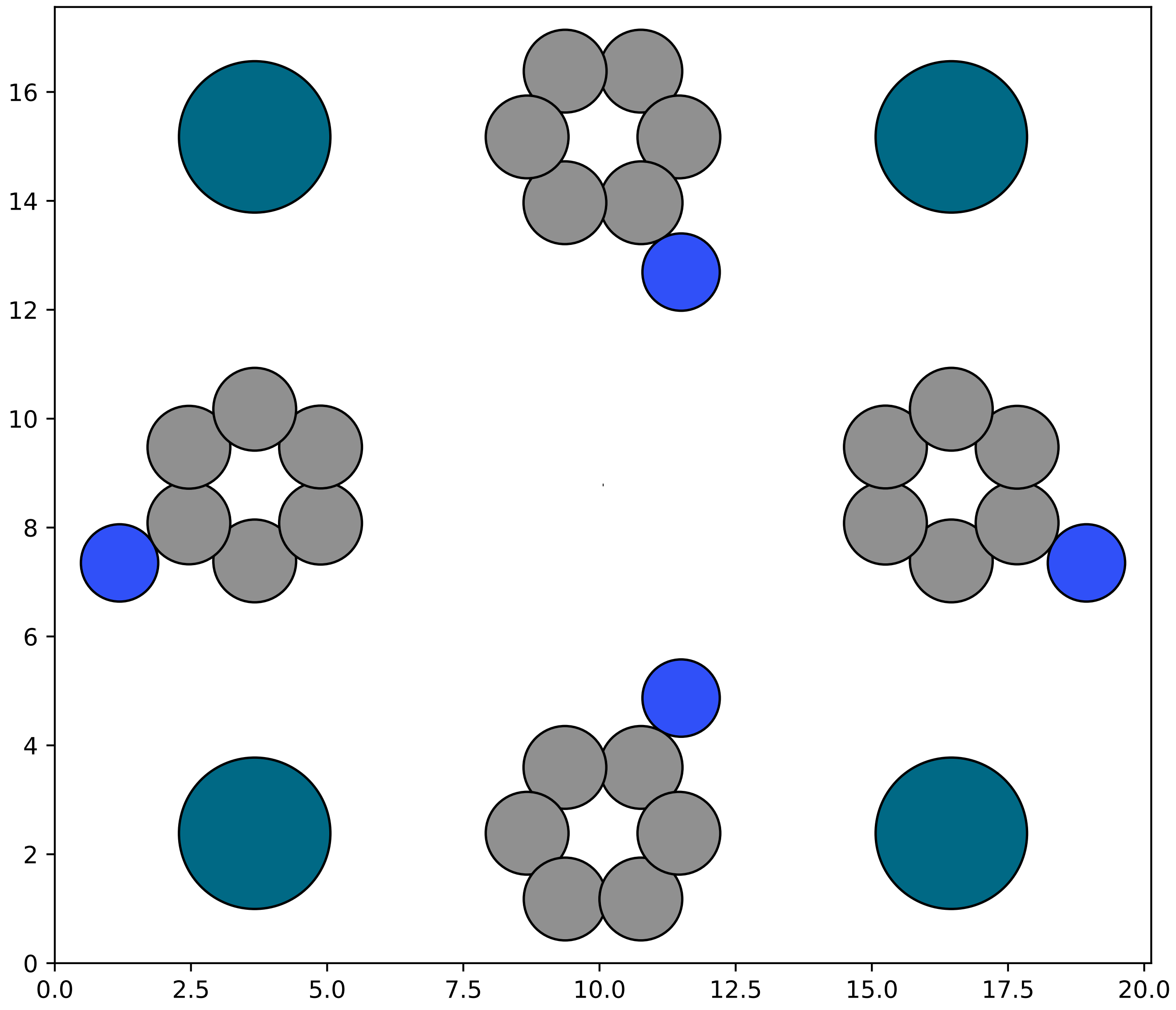
7.8447
7.8236
11.6104
9.1602
17.9793
10.6034



Isomer number 39

Isomer FG-FG distances (Å):

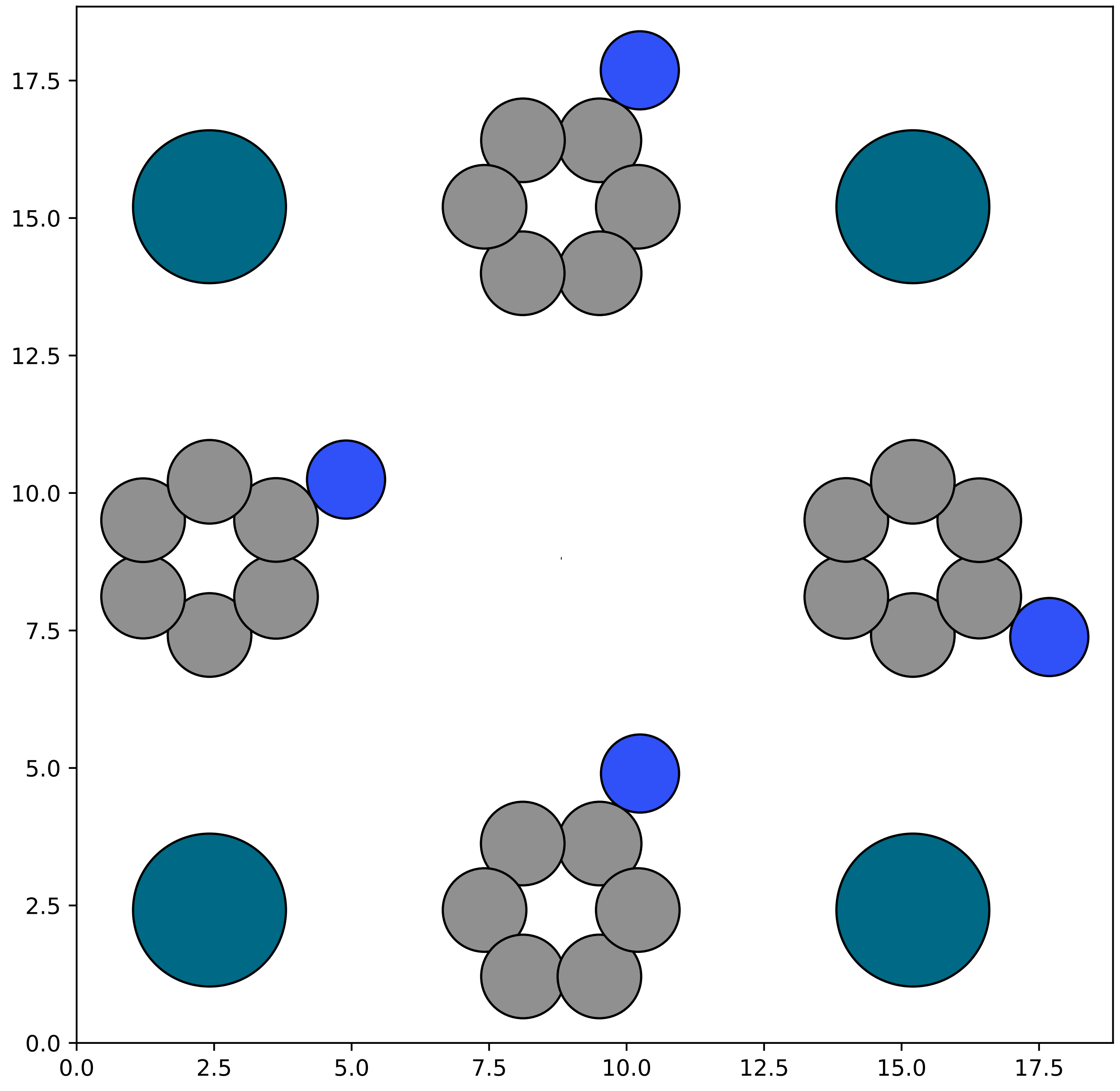
7.8447
7.8236
10.6034
9.1602
17.7505
11.6104



Isomer number 41

Isomer FG-FG distances (Å):

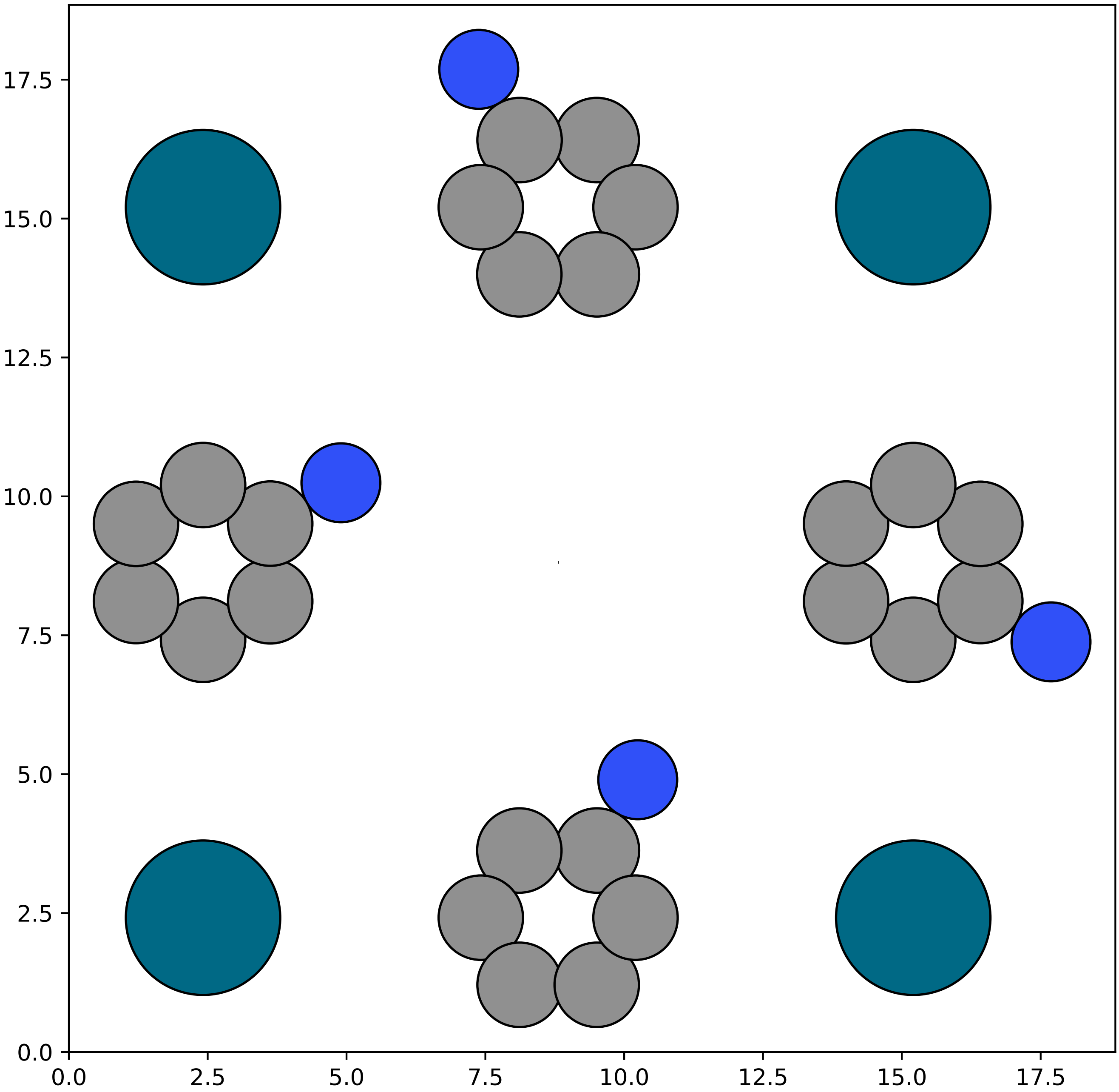
7.8447
12.7870
7.5595
12.7133
13.1037
9.1602



Isomer number 45

Isomer FG-FG distances (Å):

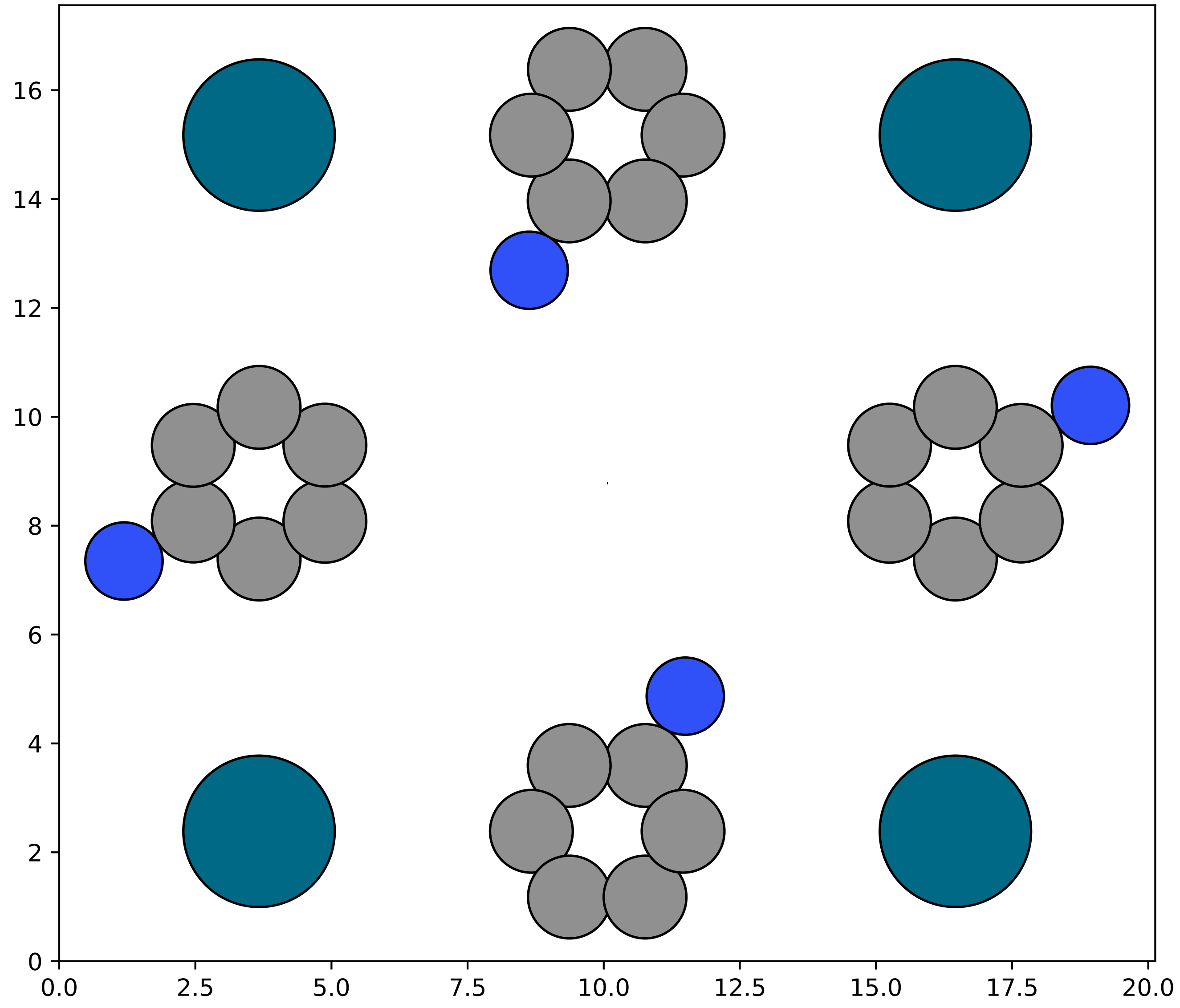
7.8447
13.1037
7.5595
14.5733
13.1037
7.8447



Isomer number 51

Isomer FG-FG distances (Å):

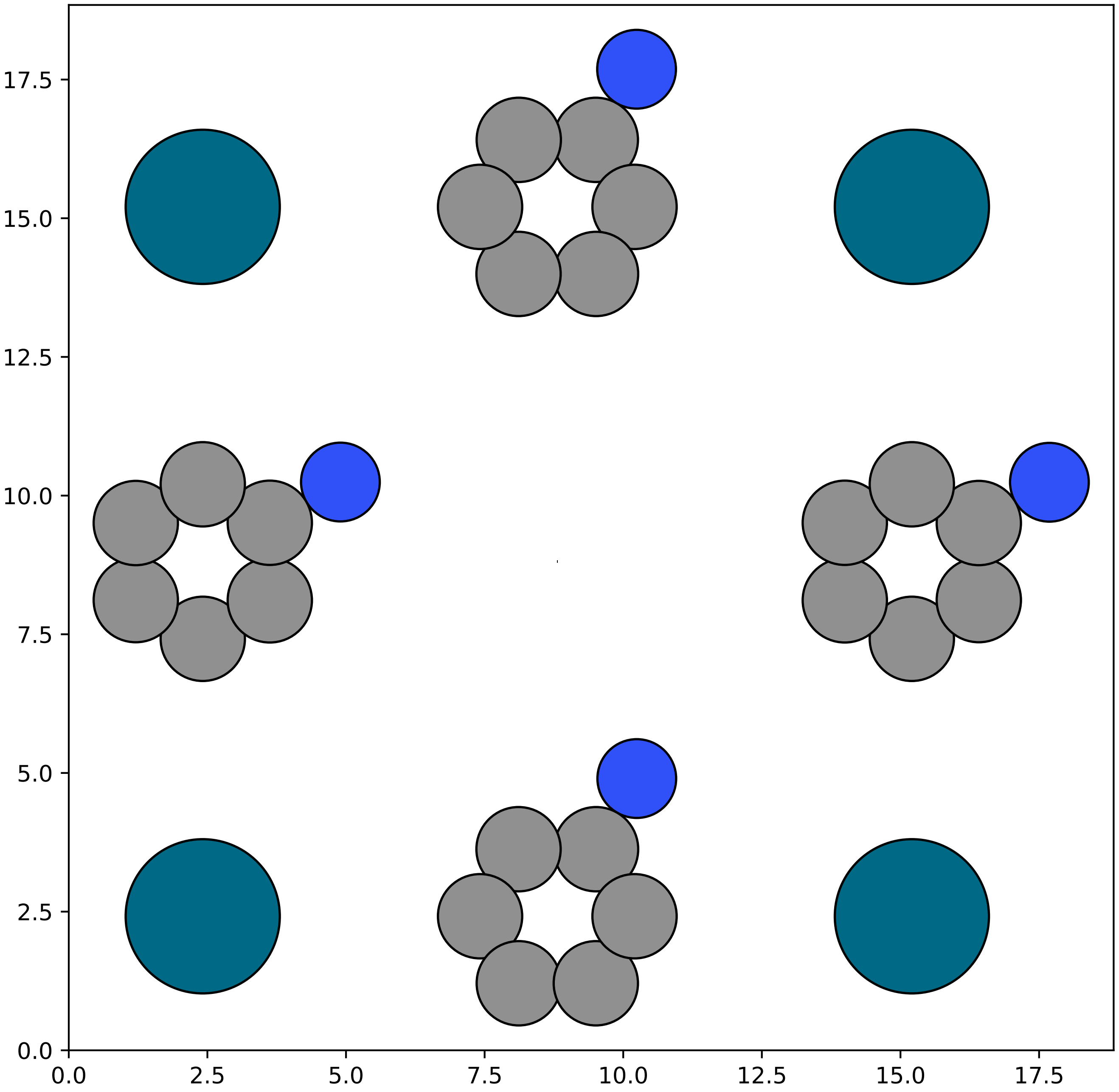
9.1602
8.3324
10.6034
10.6034
17.9793
9.1602



Isomer number 57

Isomer FG-FG distances (Å):

9.1602
12.7870
7.5595
10.5297
12.7870
9.1602

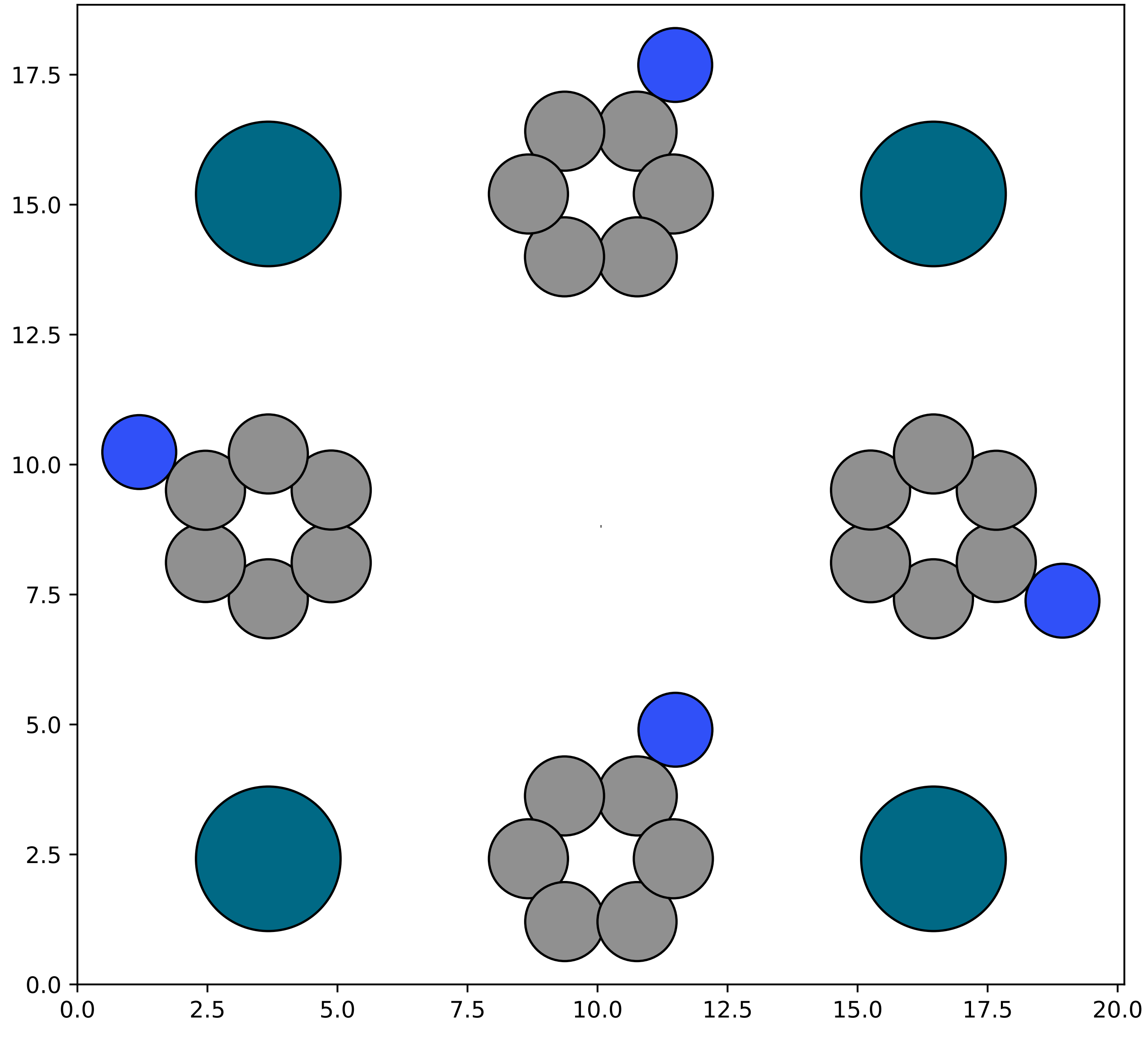


Unique Isomers with 1 FG inside the pore

Isomer number 42

Isomer FG-FG distances (Å):

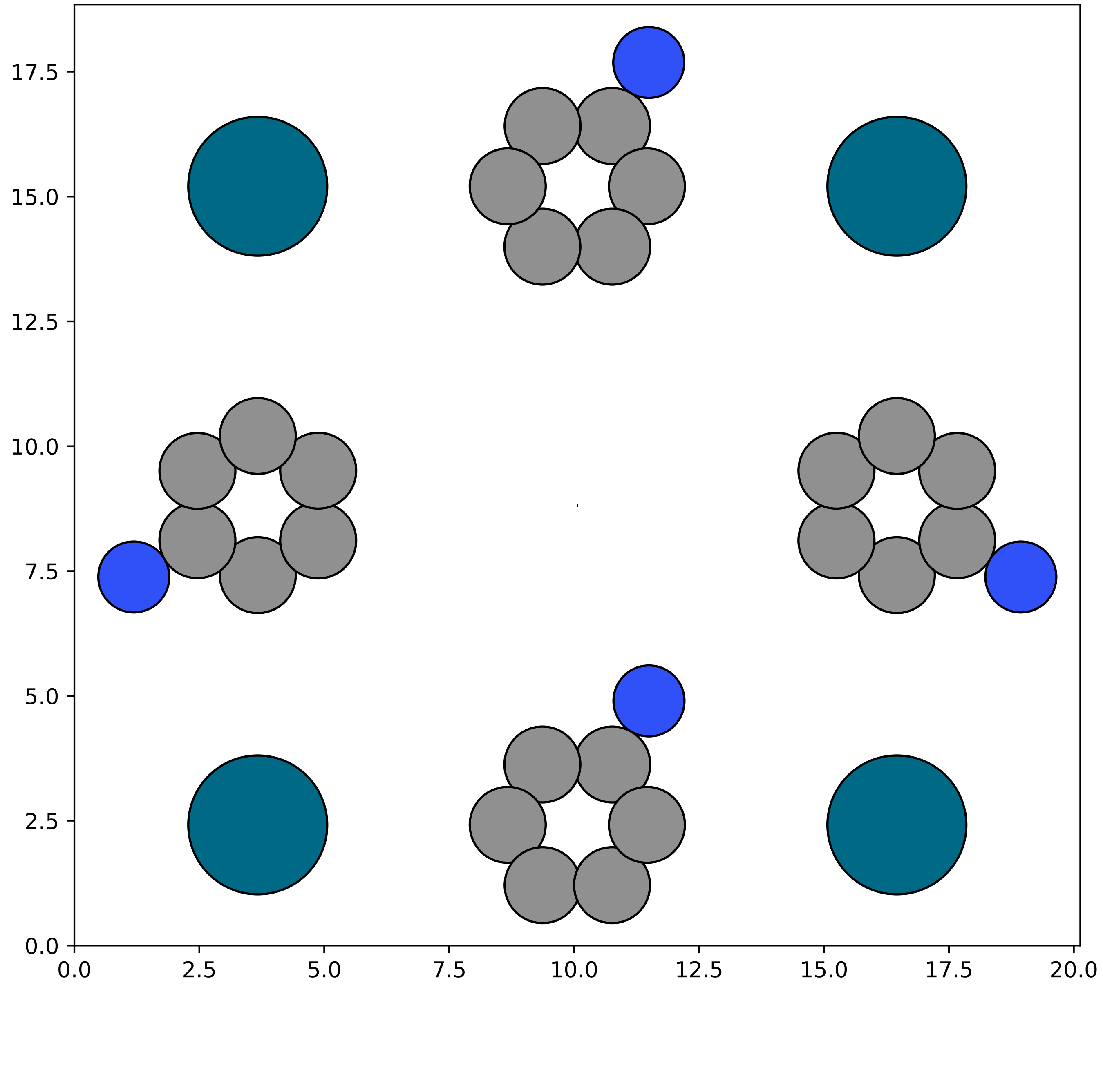
7.8447
12.7870
11.6104
12.7133
17.9793
12.7133



Isomer number 43

Isomer FG-FG distances (Å):

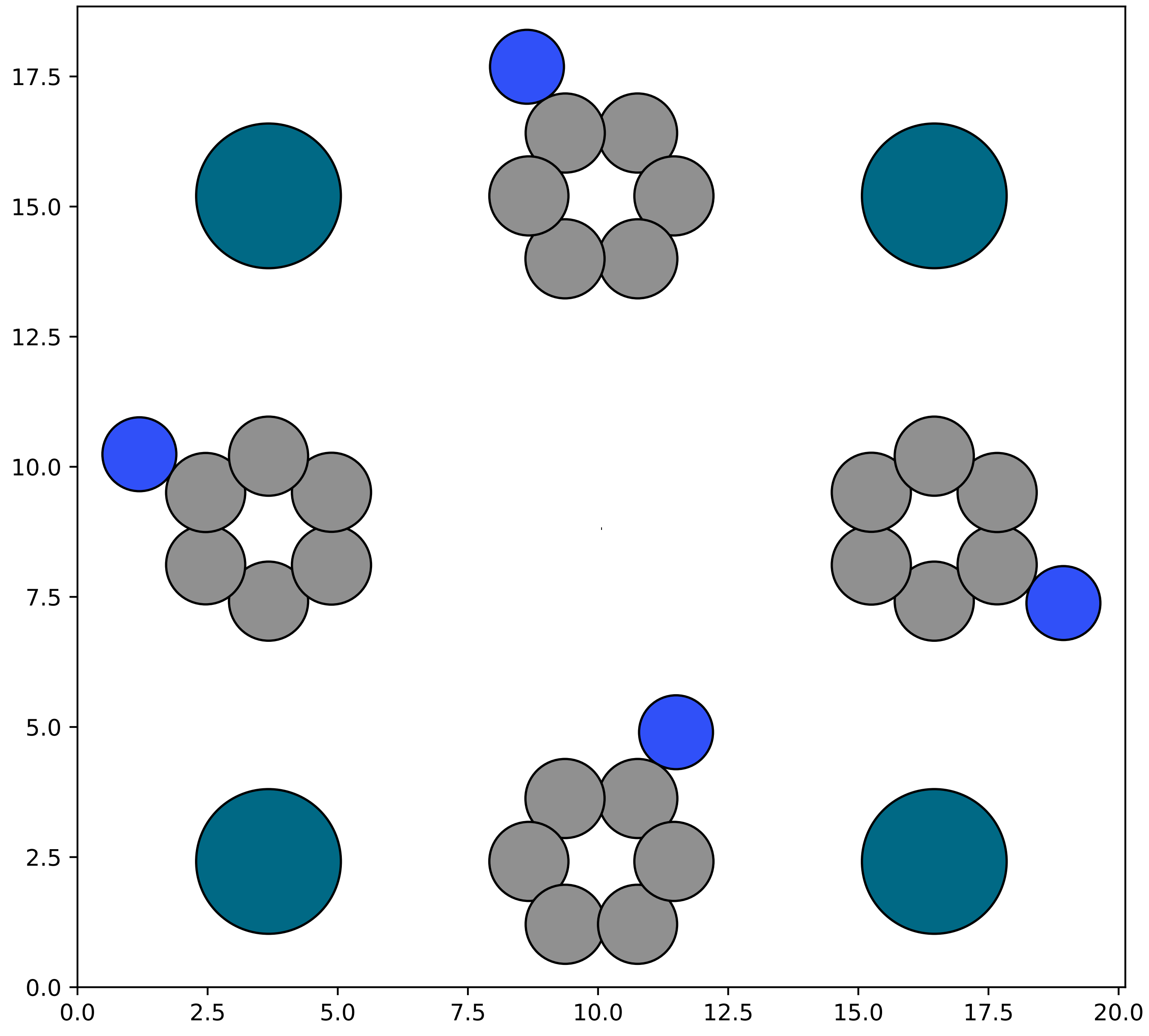
7.8447
12.7870
10.6034
12.7133
17.7505
14.5733



Isomer number 46

Isomer FG-FG distances (Å):

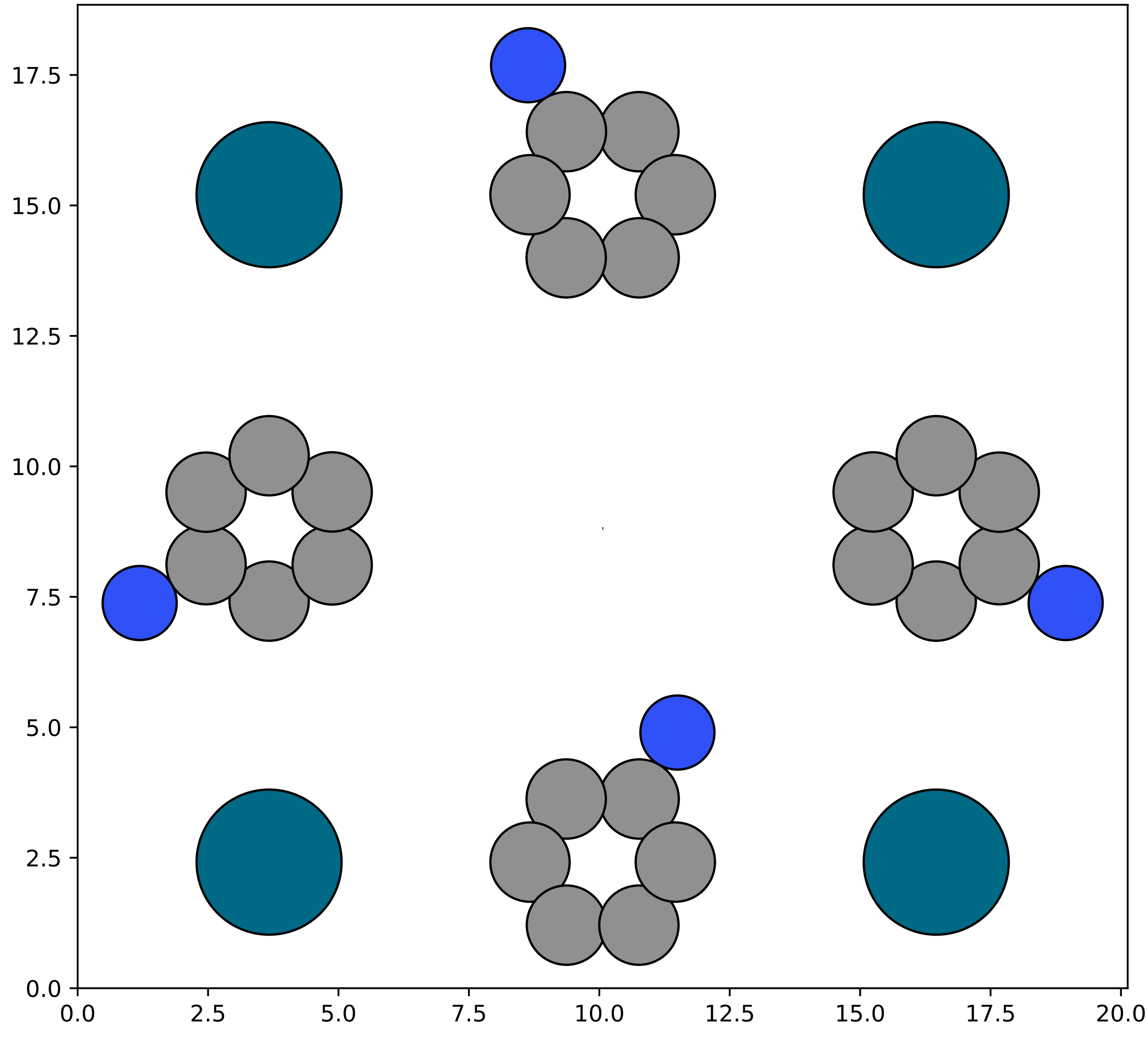
7.8447
13.1037
11.6104
14.5733
17.9793
10.5297



Isomer number 47

Isomer FG-FG distances (Å):

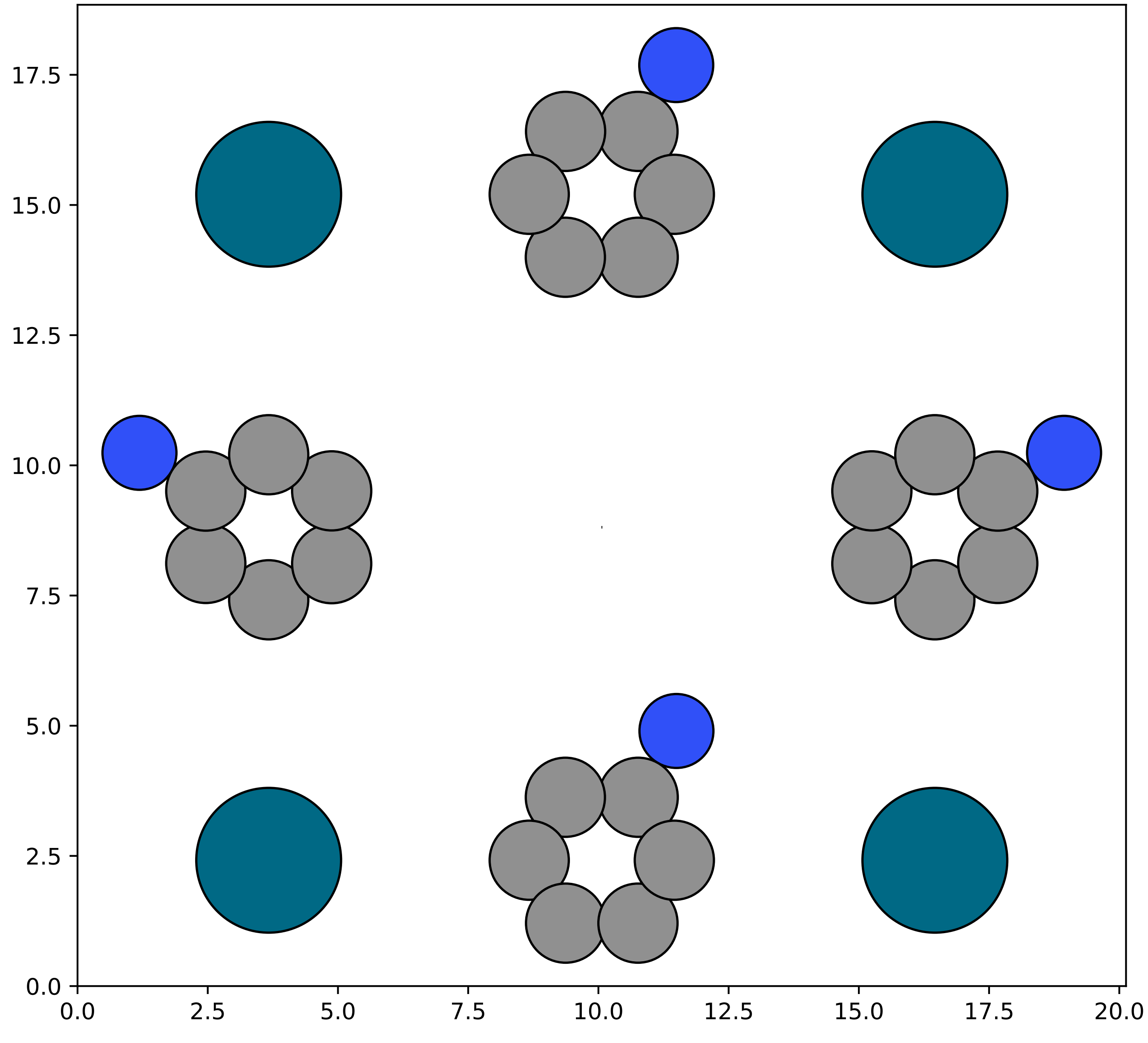
7.8447
13.1037
10.6034
14.5733
17.7505
12.7133



Isomer number 58

Isomer FG-FG distances (Å):

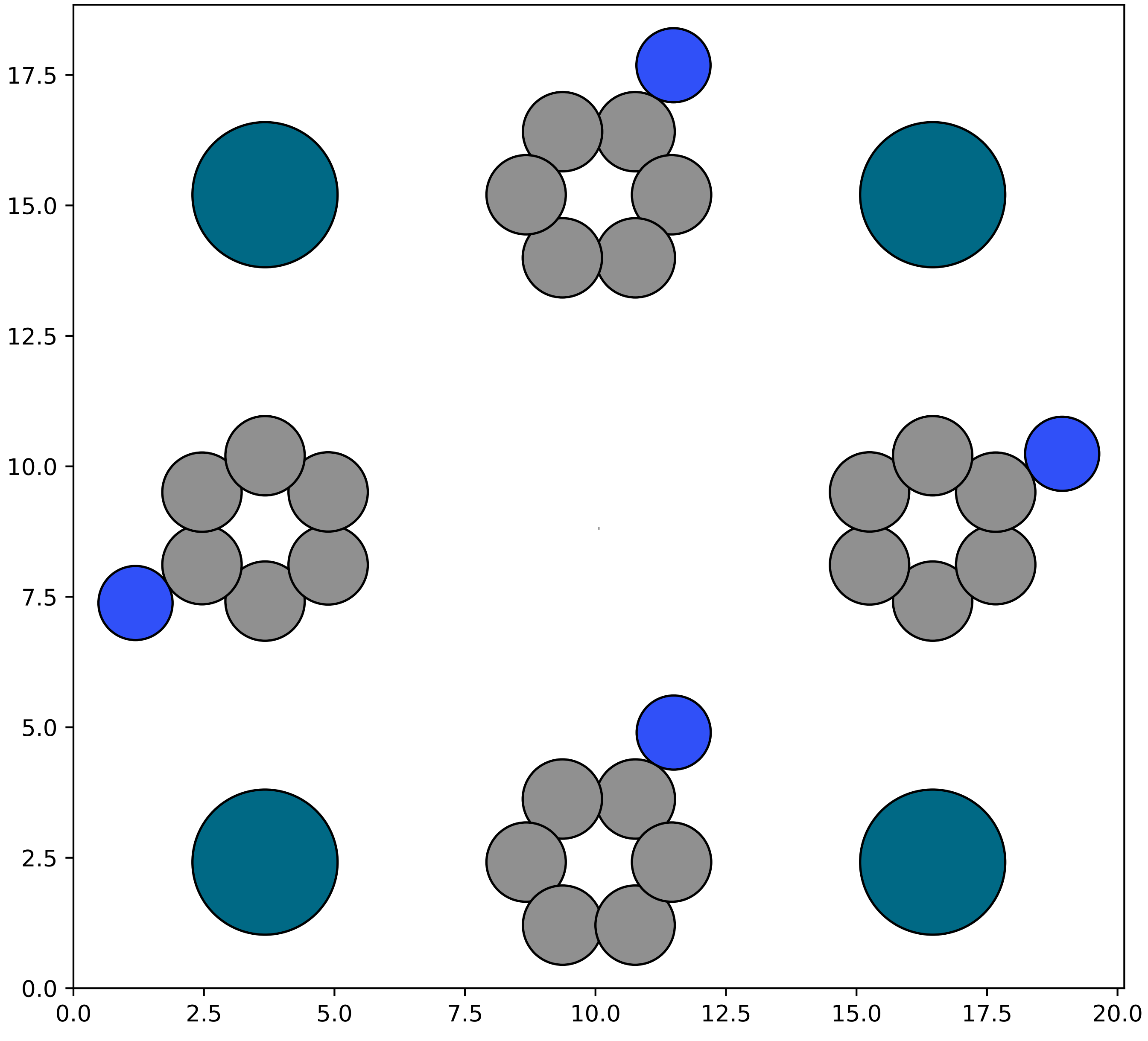
9.1602
12.7870
11.6104
10.5297
17.7505
12.7133



Isomer number 59

Isomer FG-FG distances (Å):

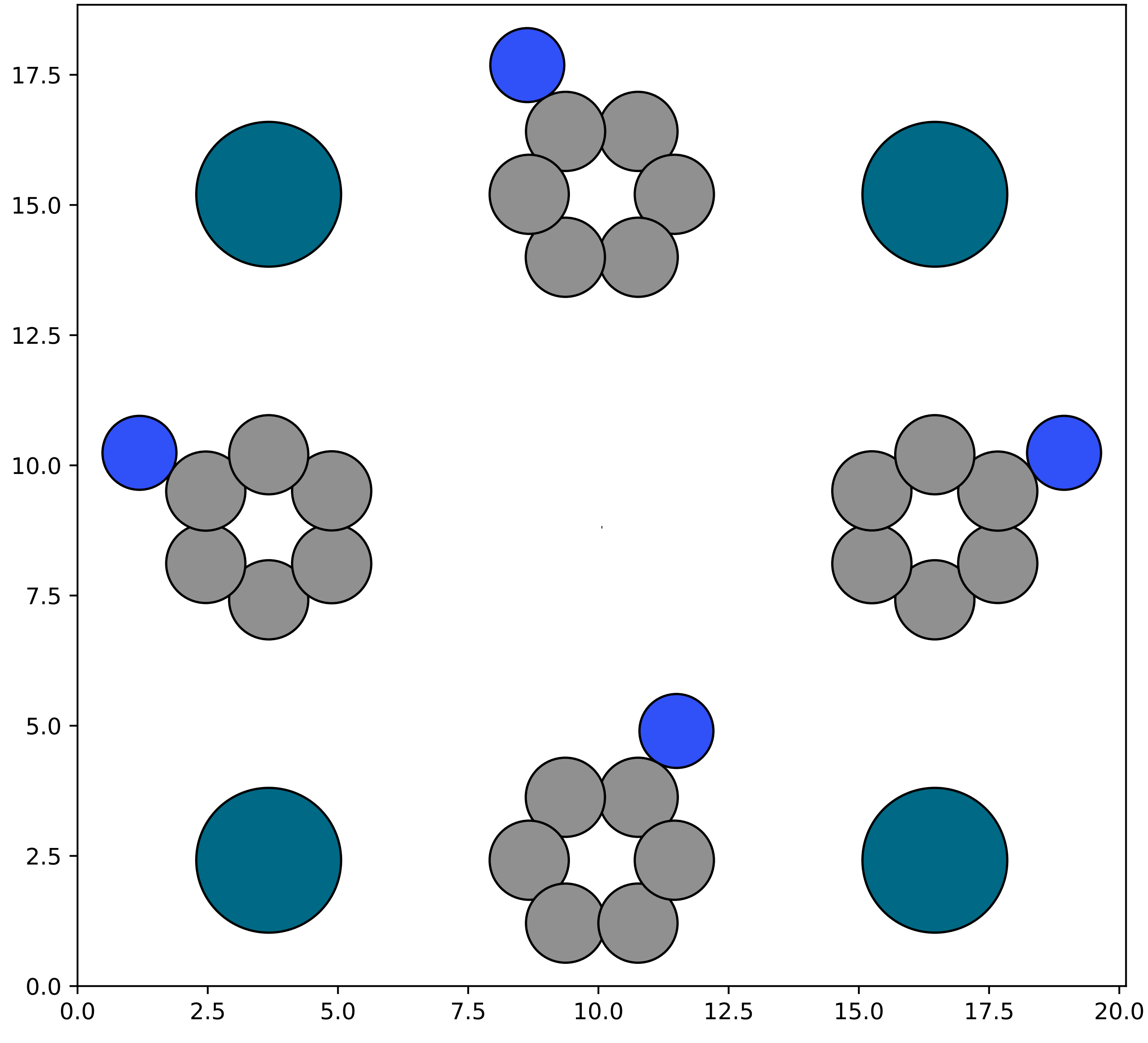
9.1602
12.7870
10.6034
10.5297
17.9793
14.5733



Isomer number 62

Isomer FG-FG distances (Å):

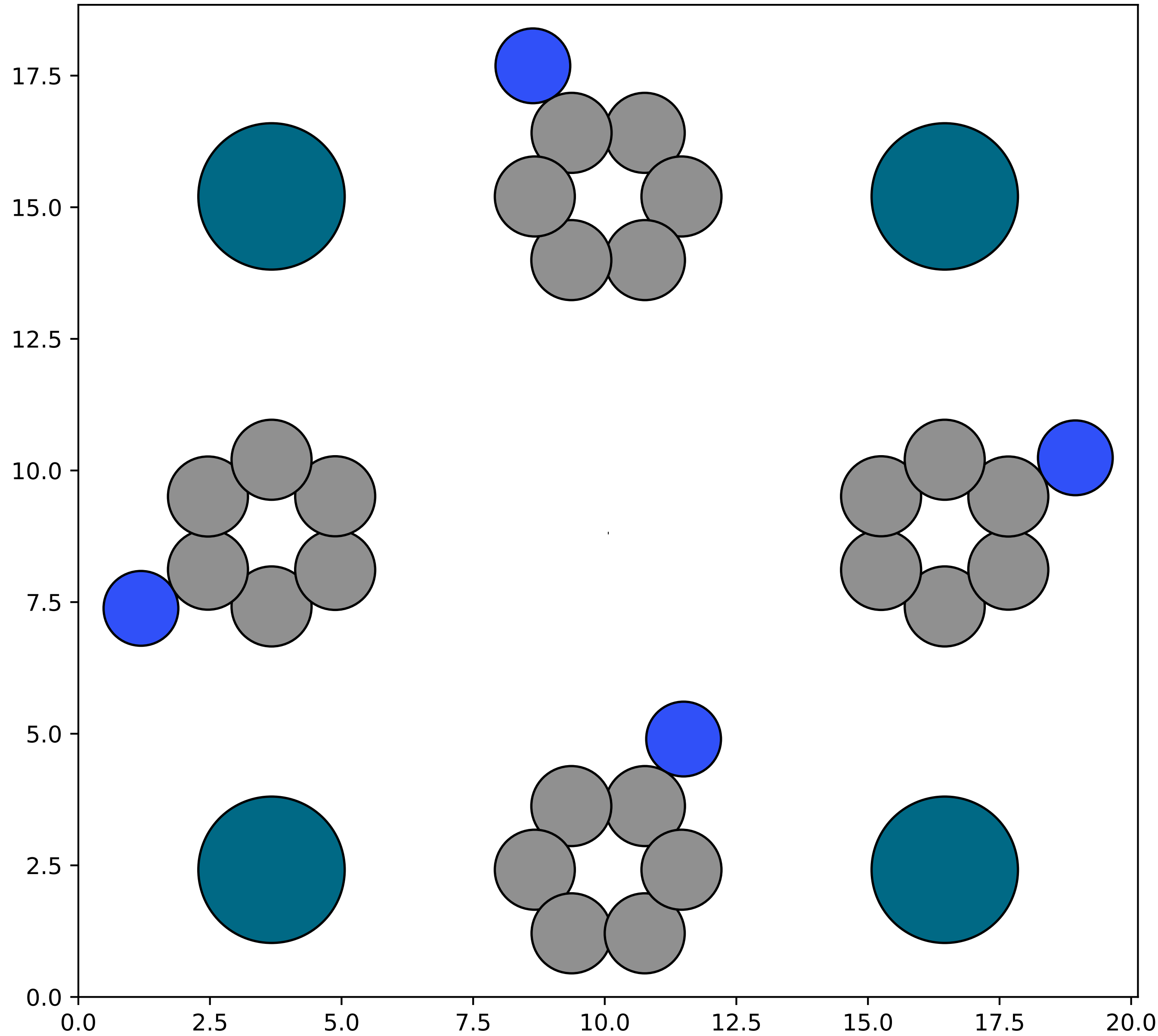
9.1602
13.1037
11.6104
12.7133
17.7505
10.5297



Isomer number 63

Isomer FG-FG distances (Å):

9.1602
13.1037
10.6034
12.7133
17.9793
12.7133

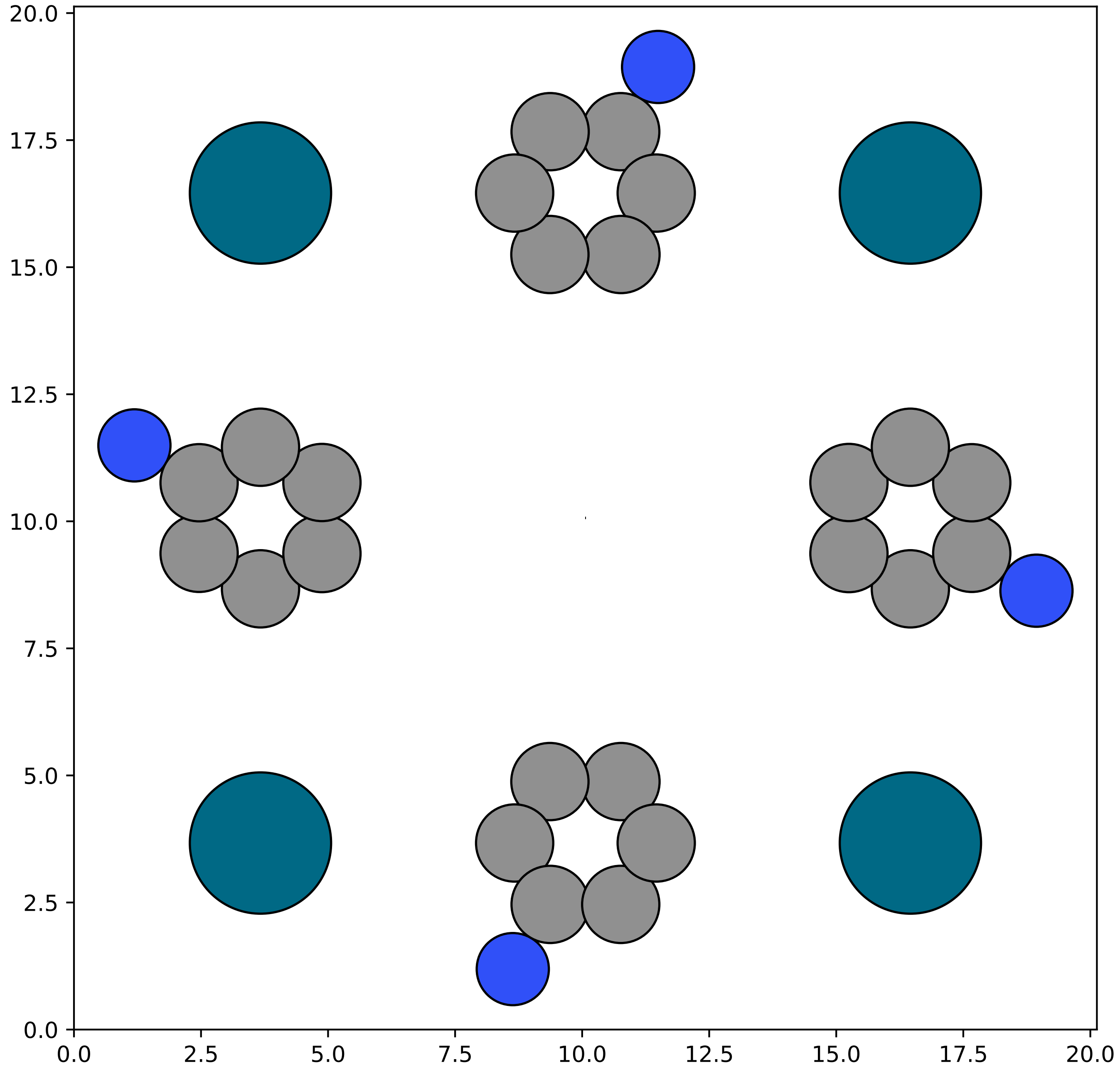


Unique Isomers with 0 FG inside the pore

Isomer number 170

Isomer FG-FG distances (Å):

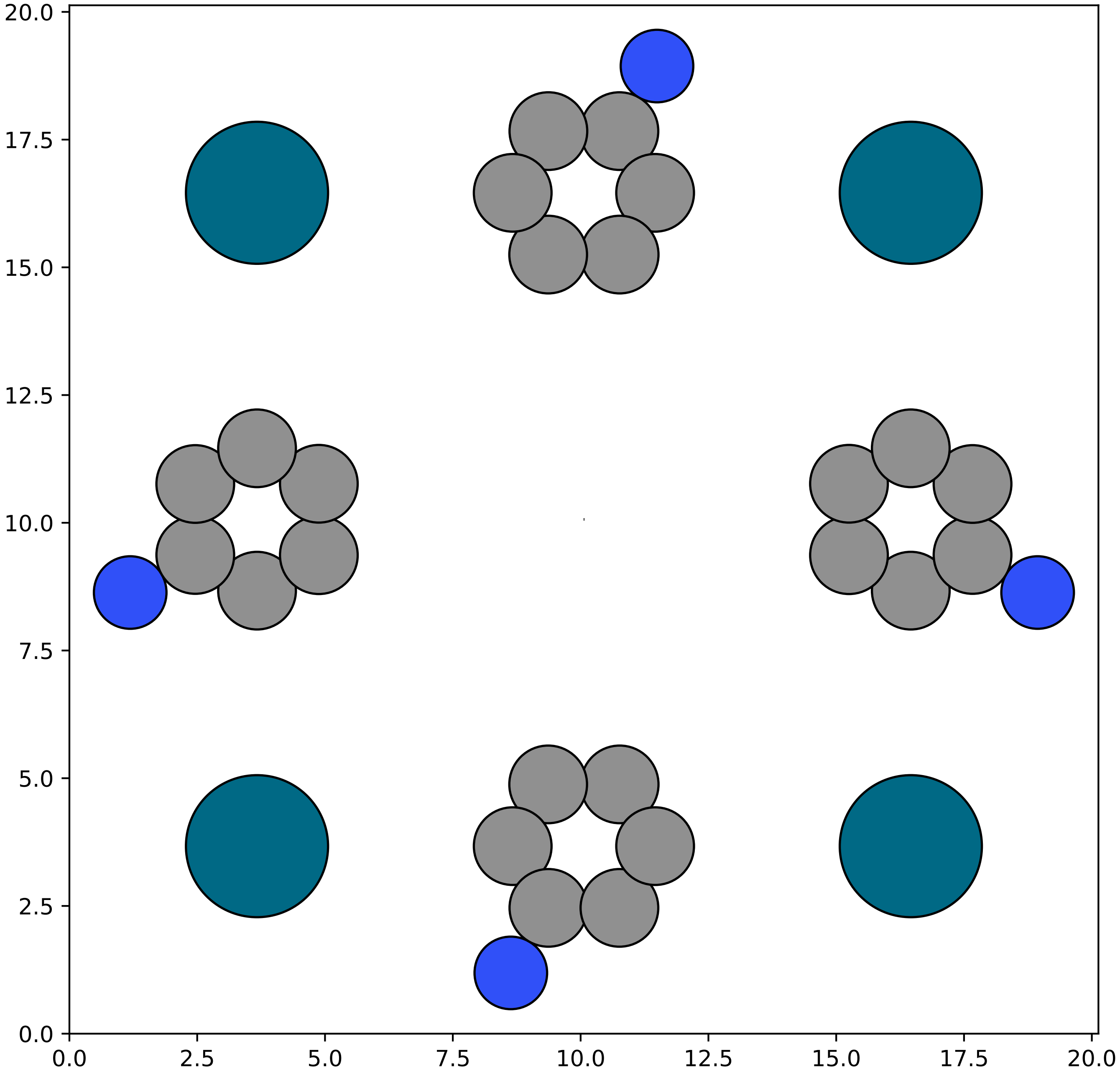
12.7133
17.9793
12.7133
12.7133
17.9793
12.7133



Isomer number 171

Isomer FG-FG distances (Å):

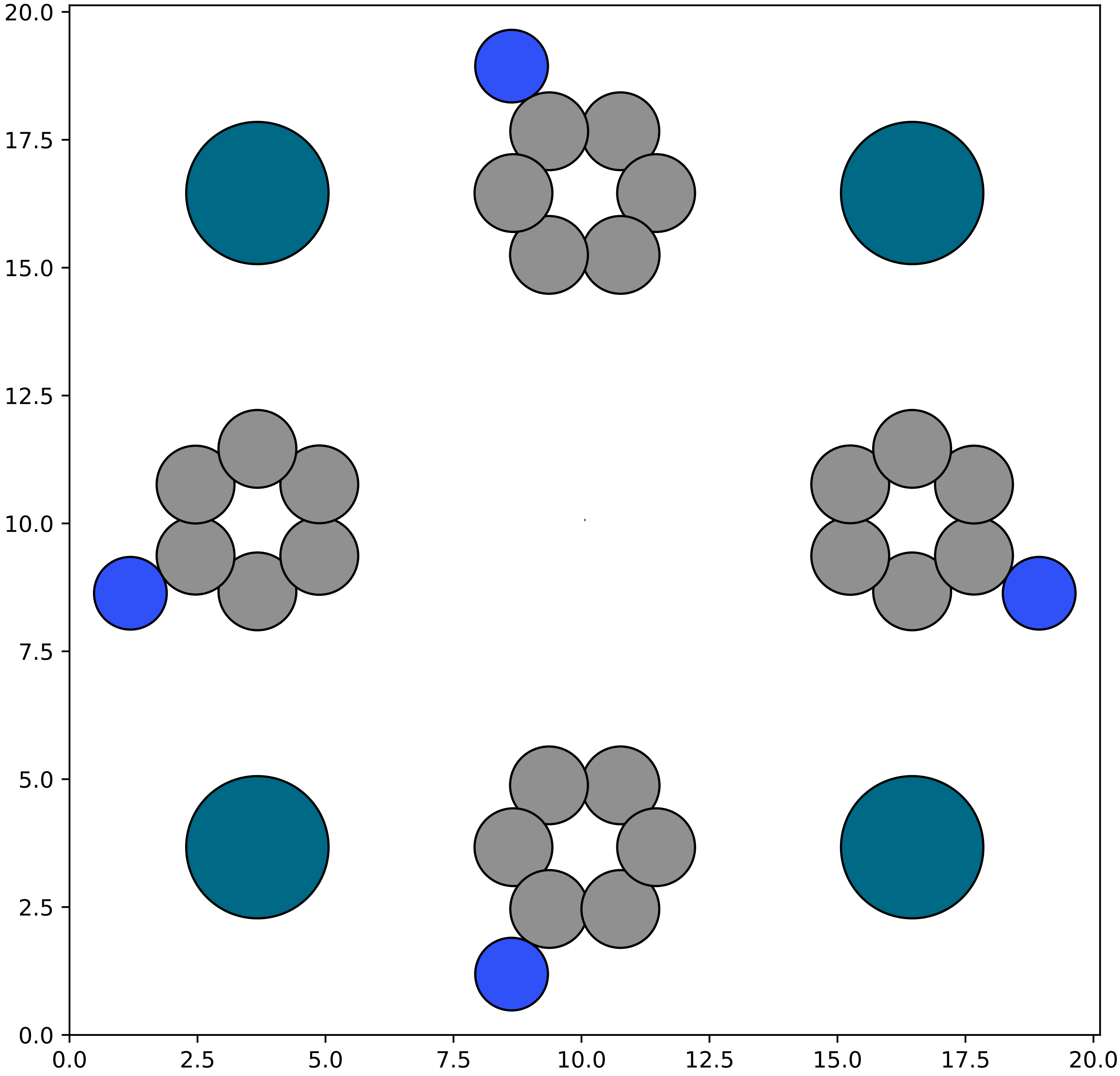
12.7133
17.9793
10.5297
12.7133
17.7505
14.5733



Isomer number 175

Isomer FG-FG distances (Å):

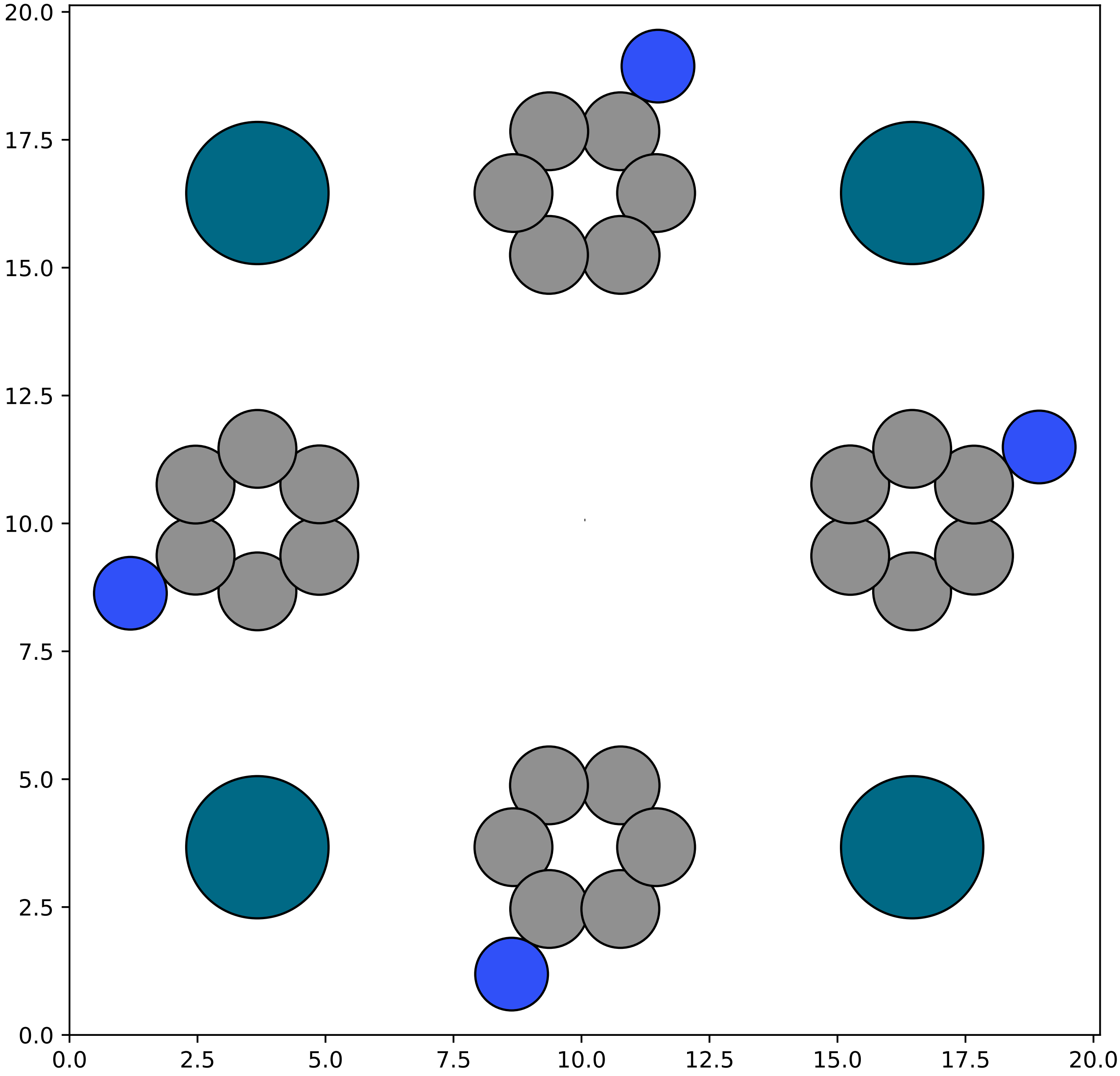
12.7133
17.7505
10.5297
14.5733
17.7505
12.7133



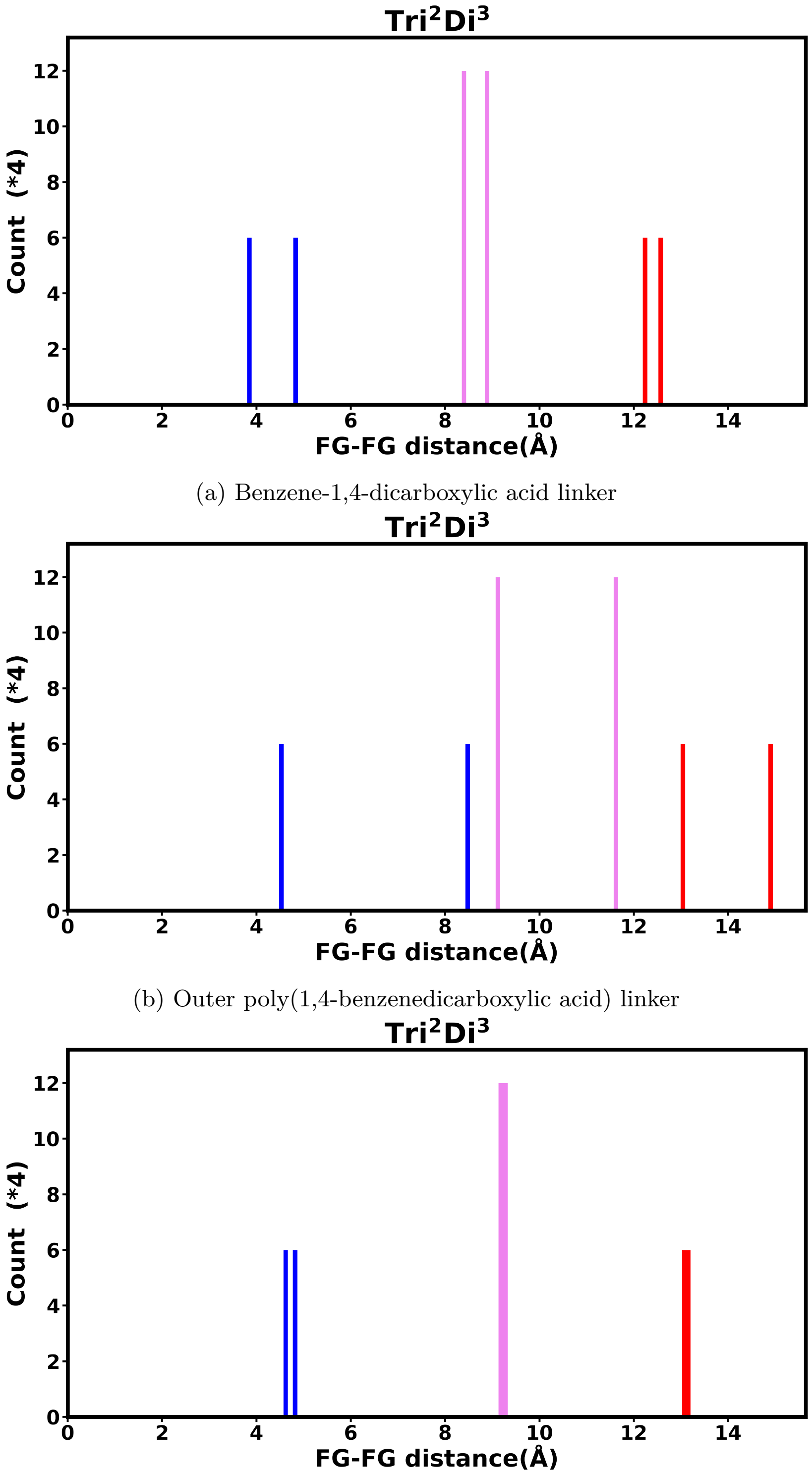
Isomer number 187

Isomer FG-FG distances (Å):

14.5733
17.9793
10.5297
10.5297
17.9793
14.5733



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

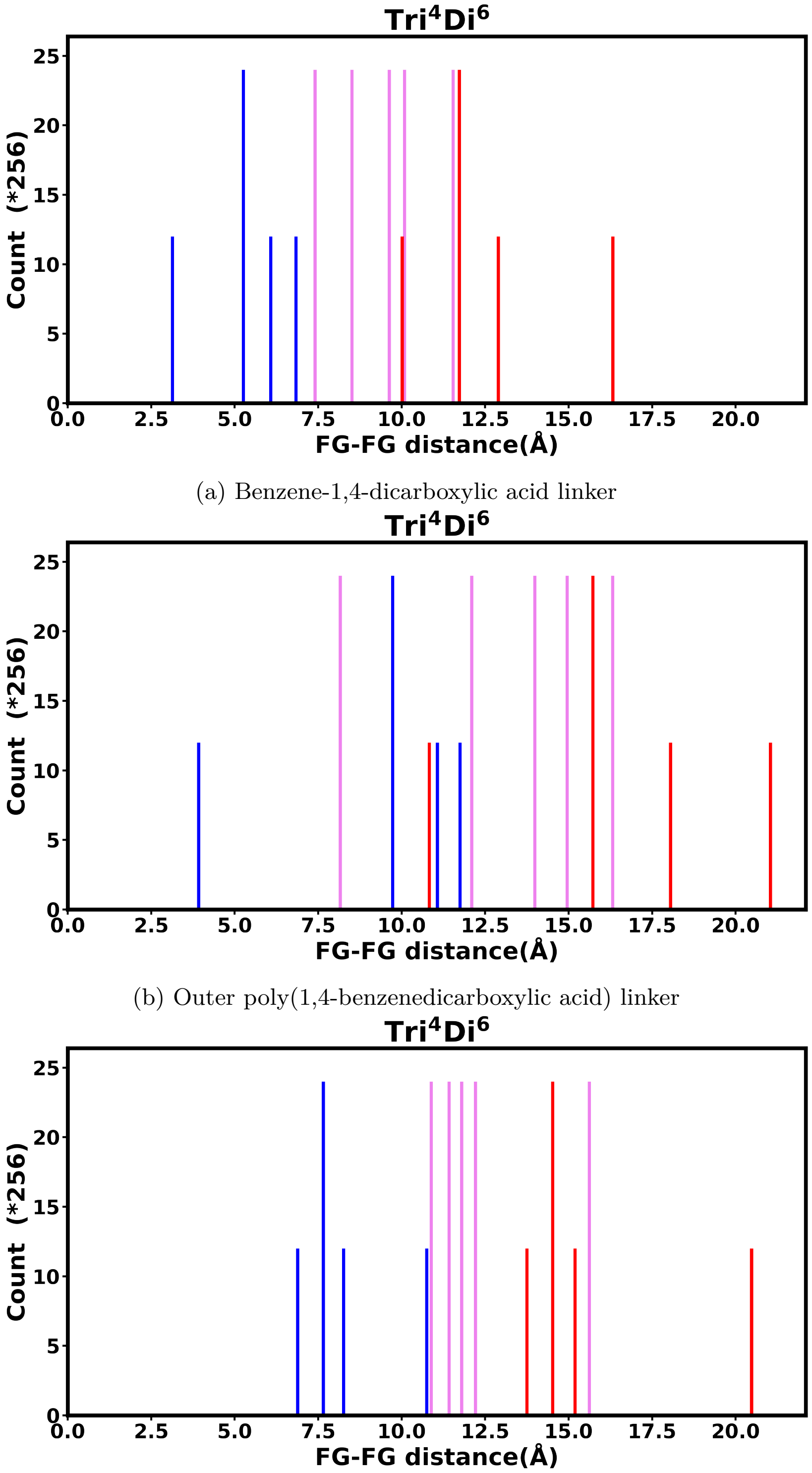


(a) Benzene-1,4-dicarboxylic acid linker

(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.

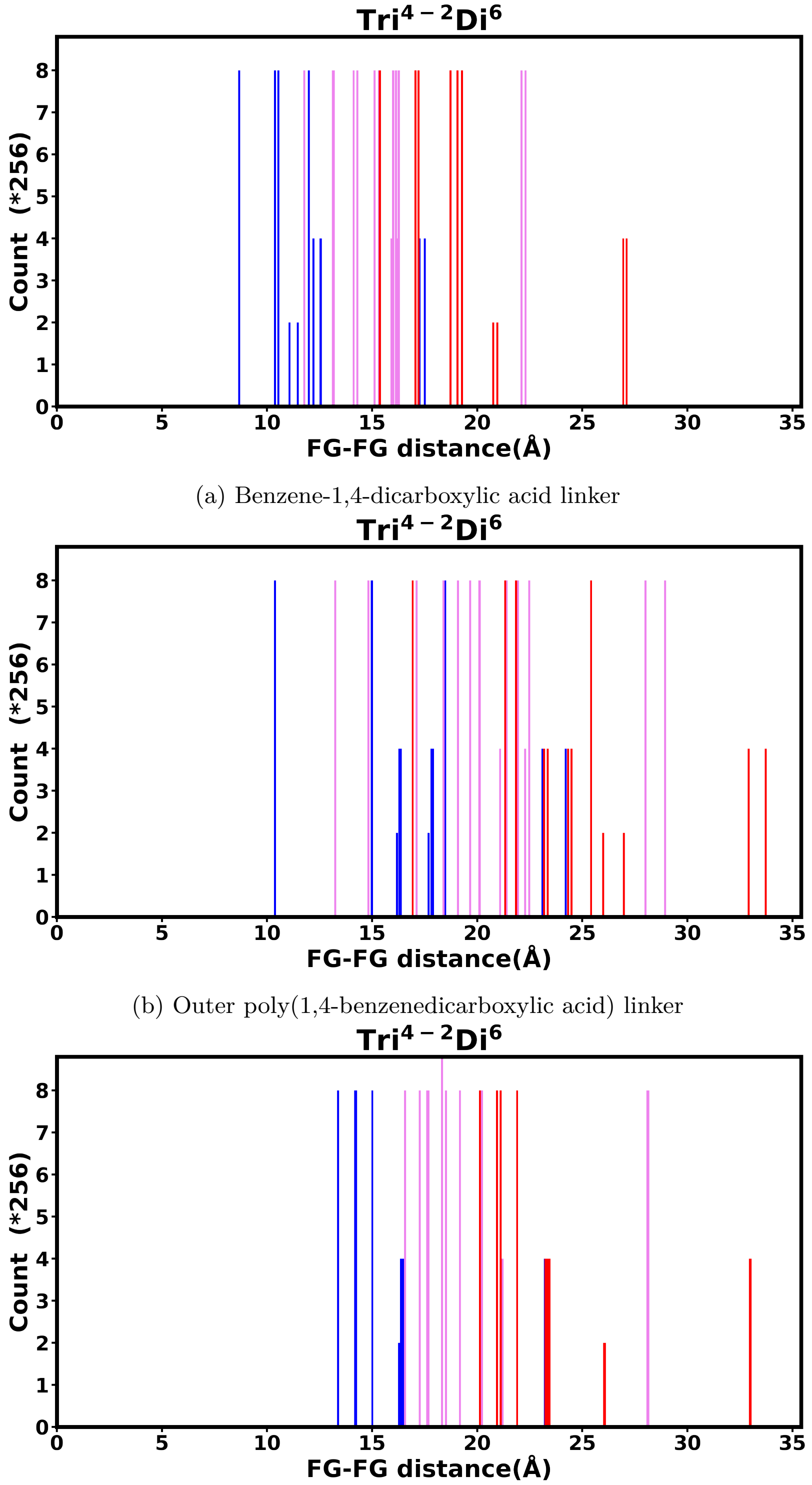


(a) Benzene-1,4-dicarboxylic acid linker

(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.

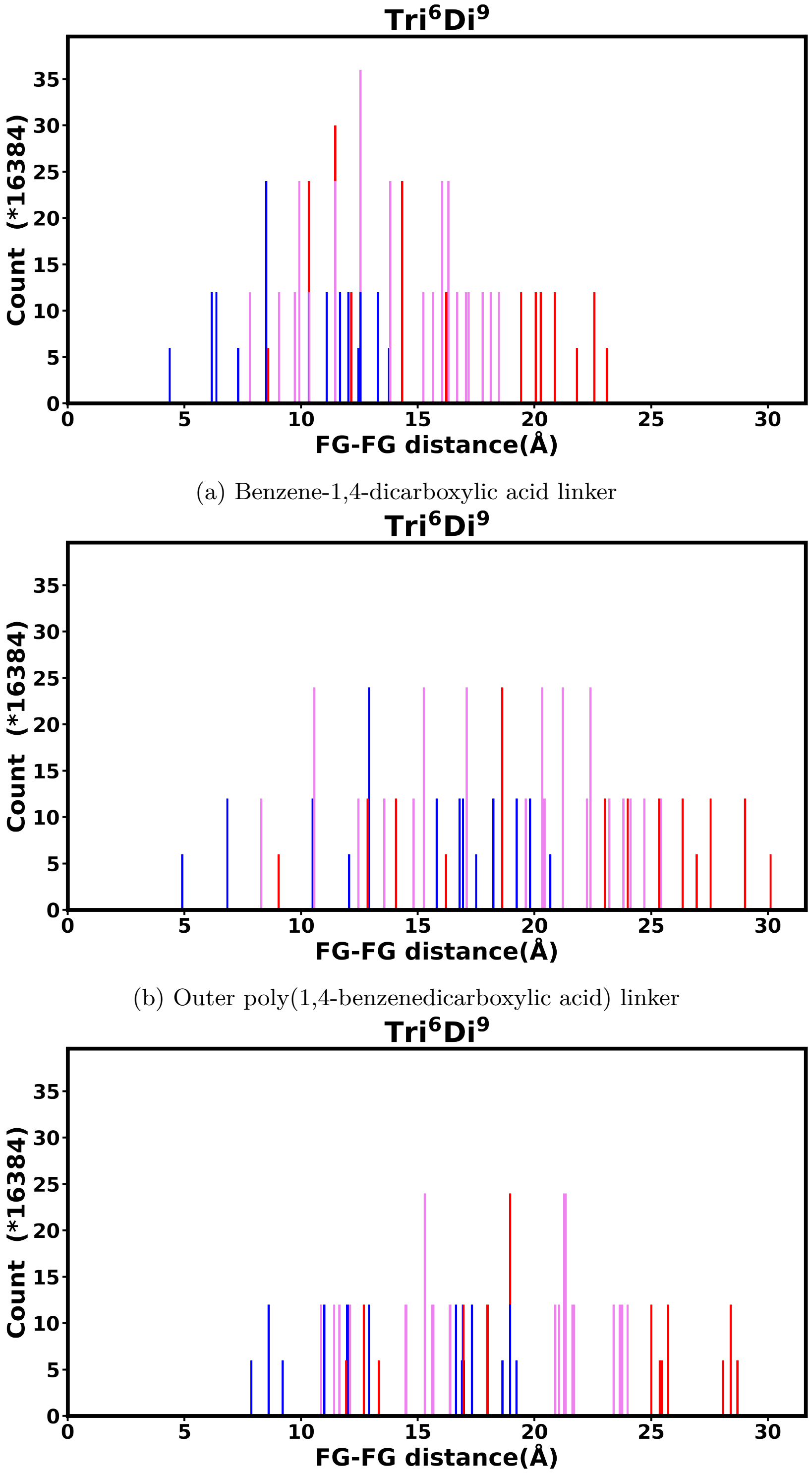


(a) Benzene-1,4-dicarboxylic acid linker

(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 $Tri^4_2Di^6$, 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.

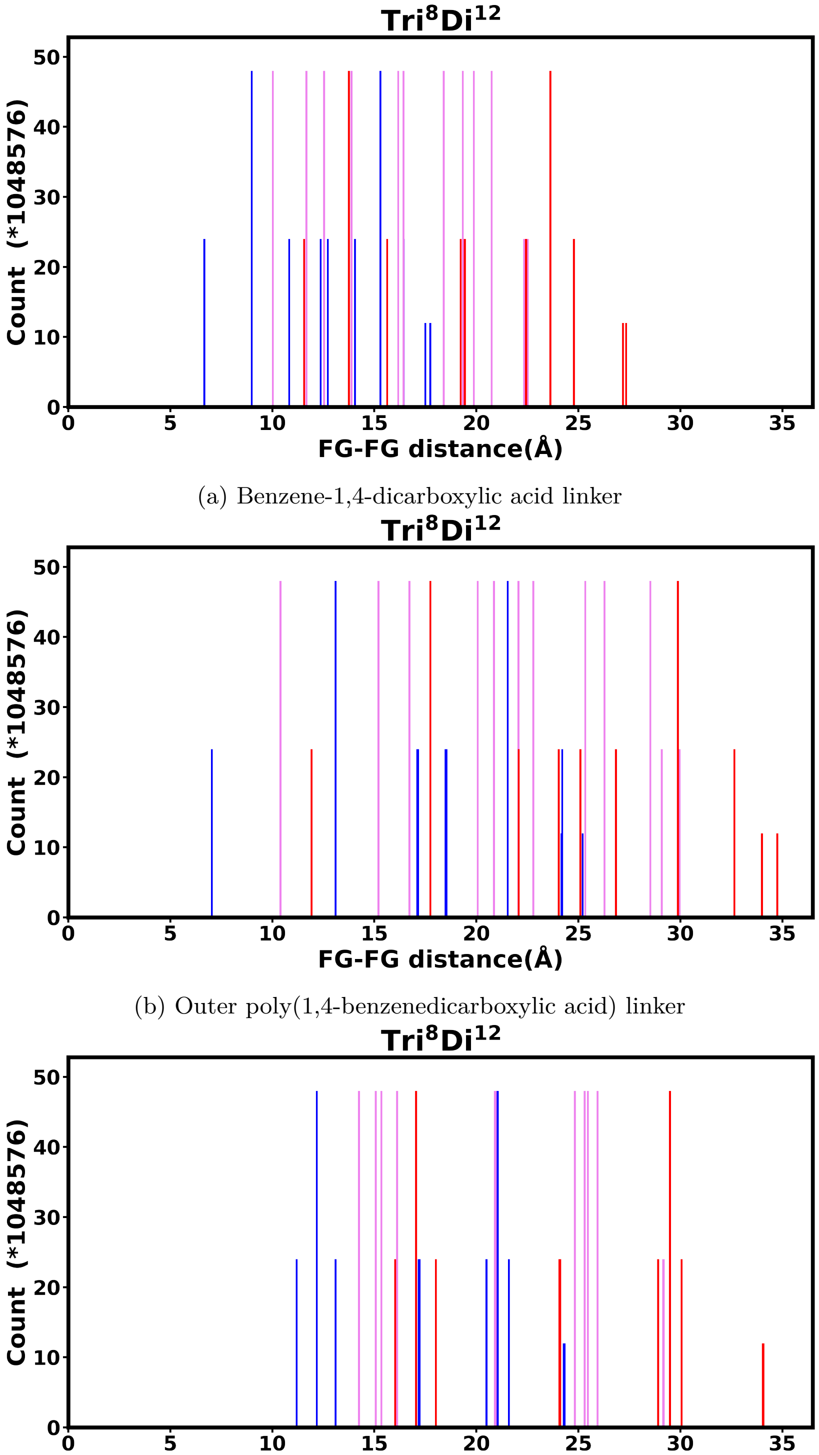


(a) Benzene-1,4-dicarboxylic acid linker

(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.

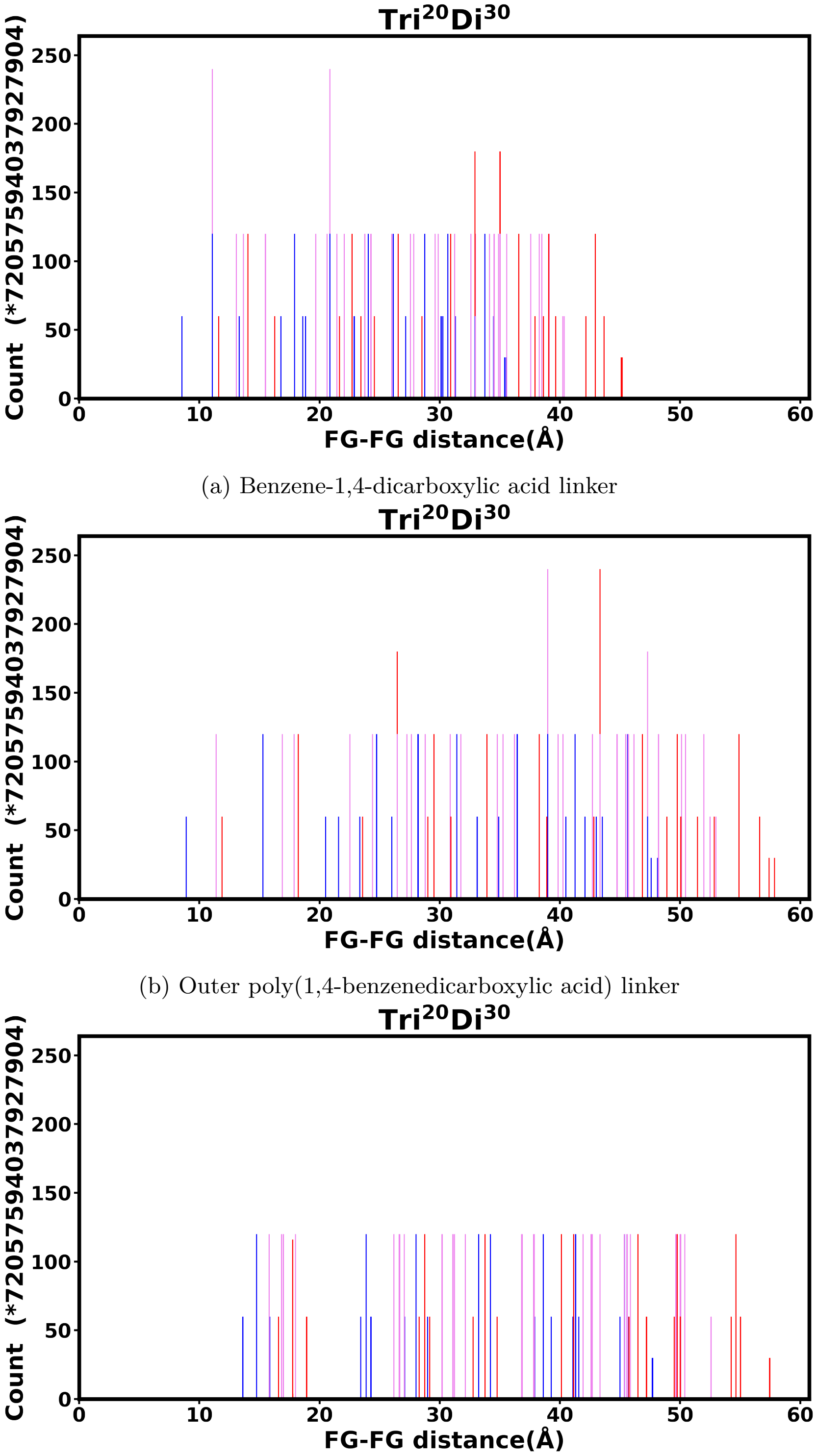


(a) Benzene-1,4-dicarboxylic acid linker

(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.

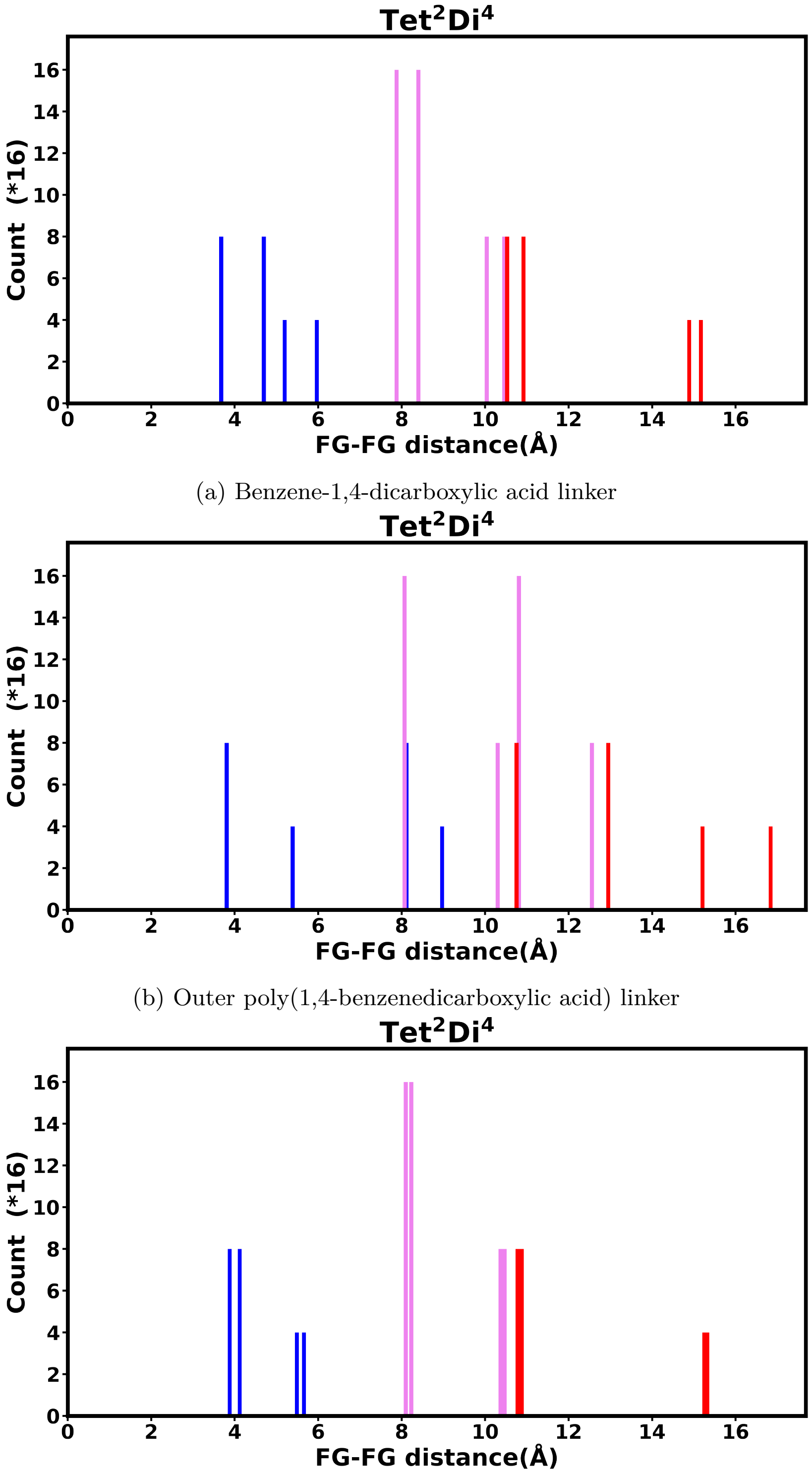


(a) Benzene-1,4-dicarboxylic acid linker

(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.

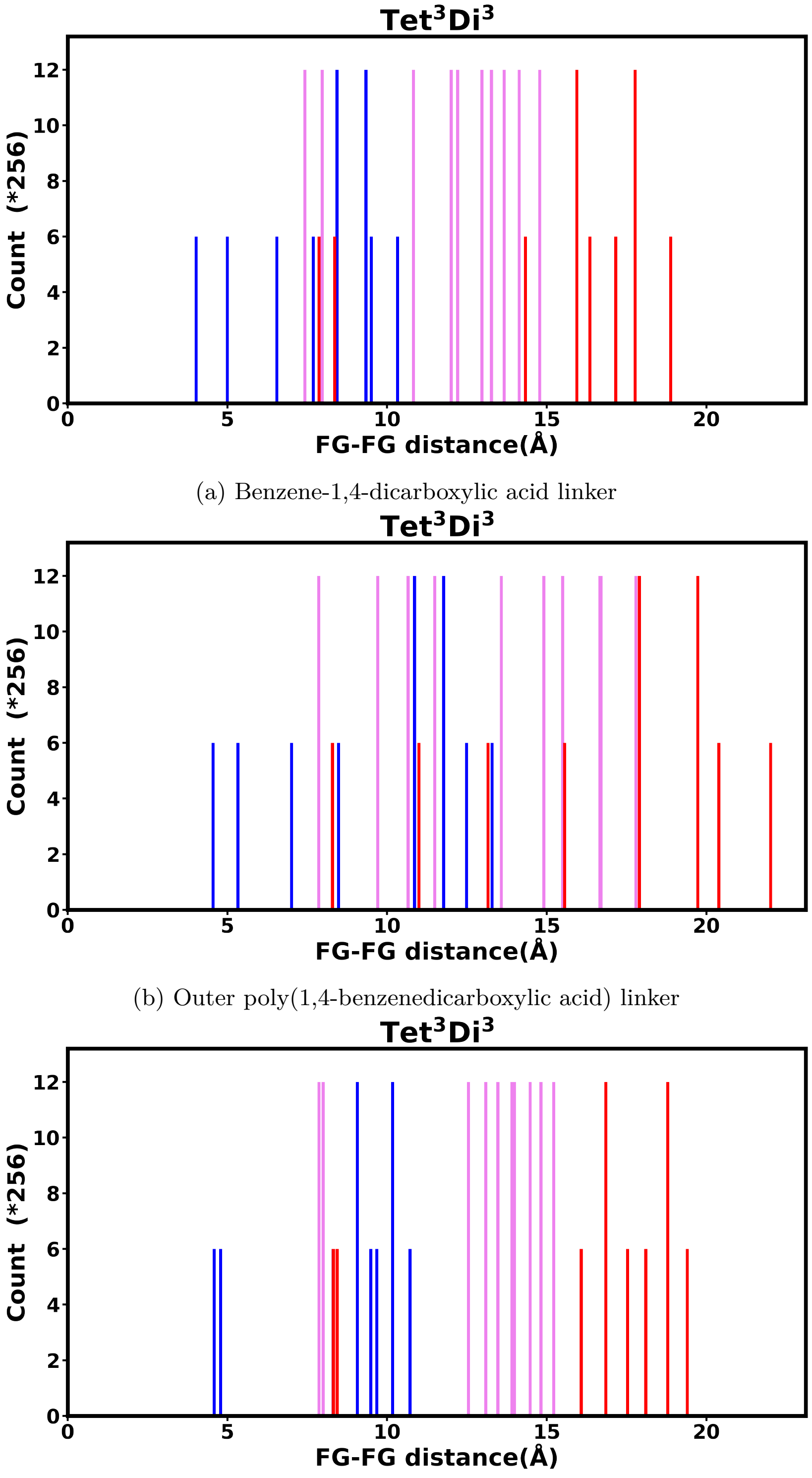


(a) Benzene-1,4-dicarboxylic acid linker

(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.

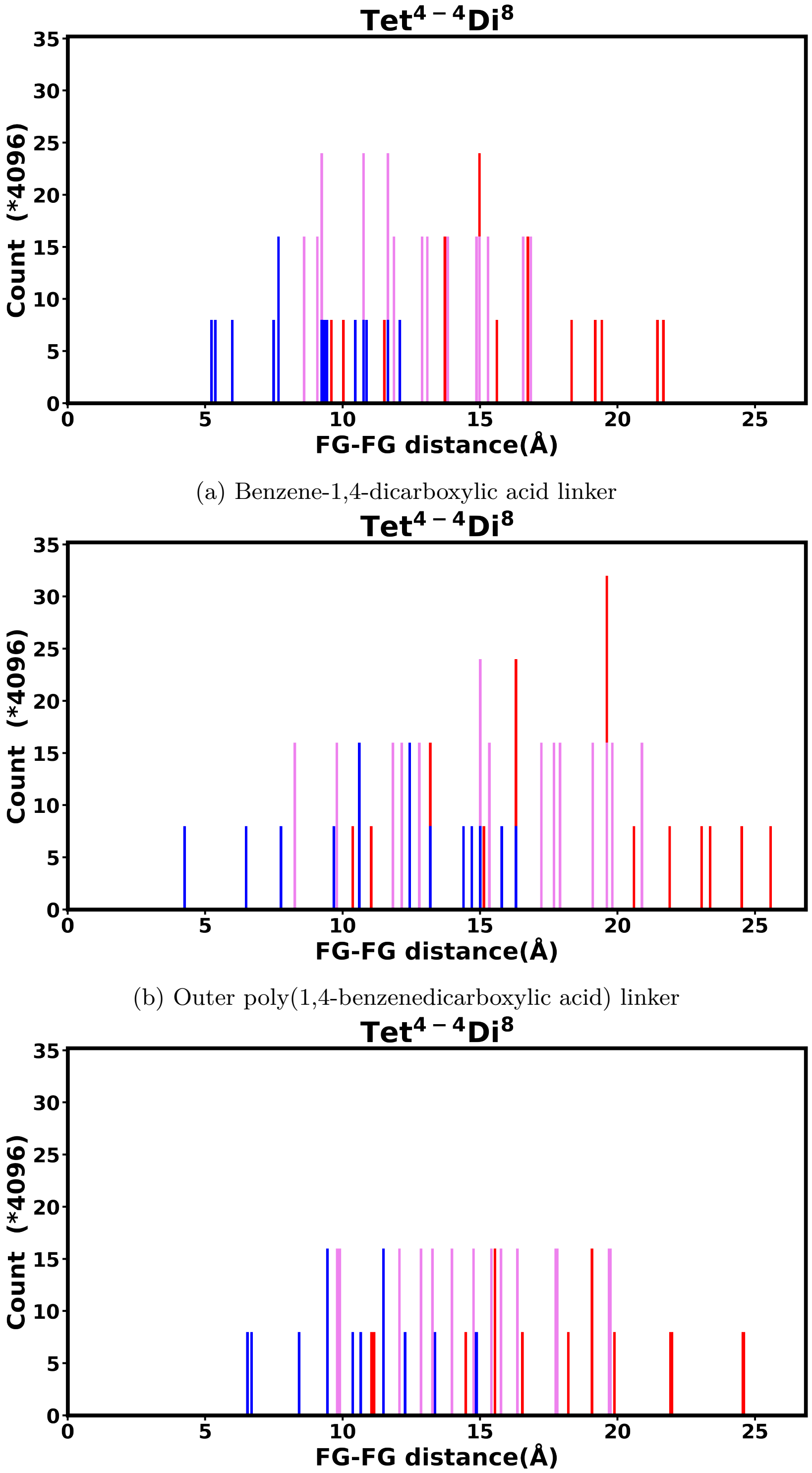


(a) Benzene-1,4-dicarboxylic acid linker

(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.

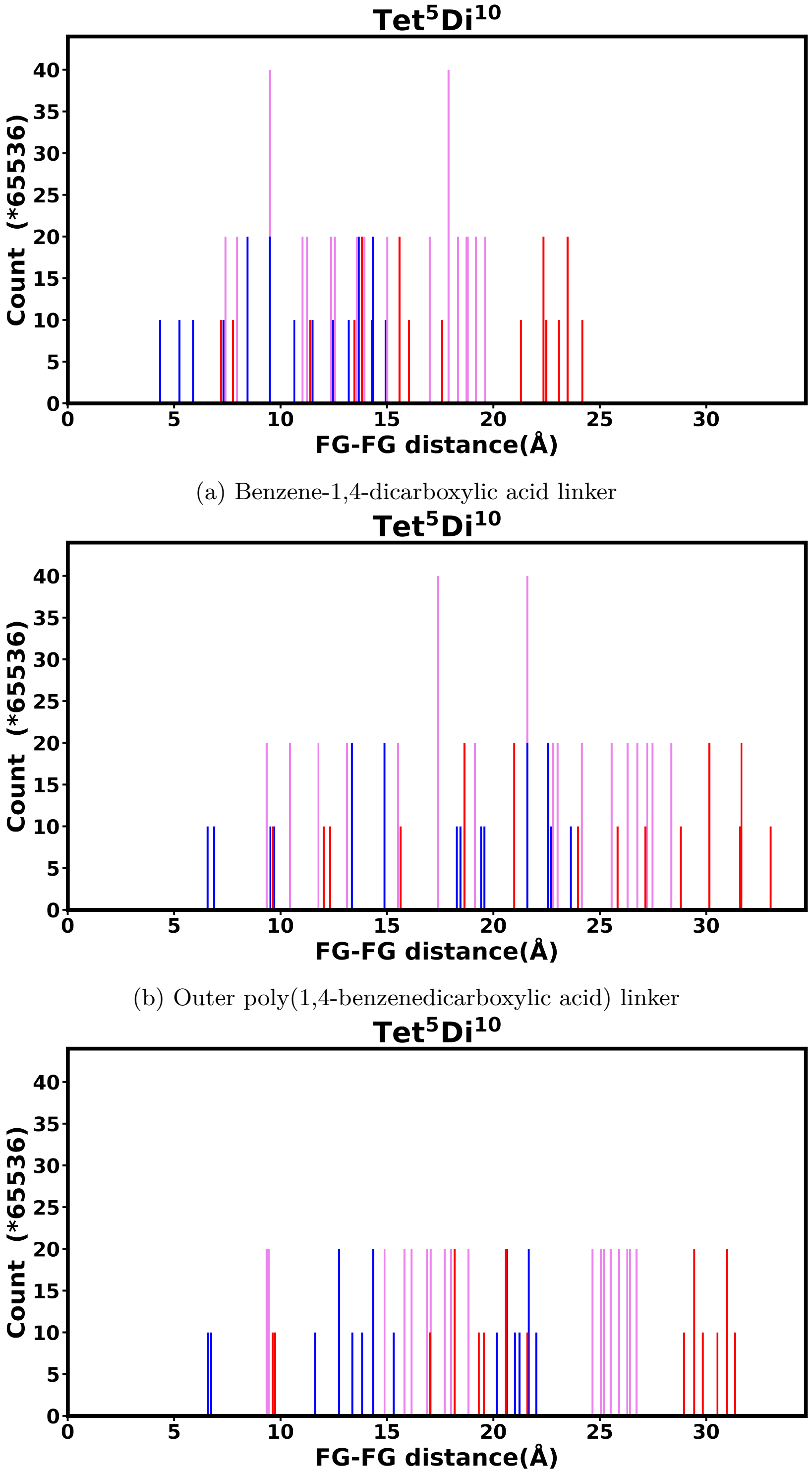


(a) Benzene-1,4-dicarboxylic acid linker

(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.

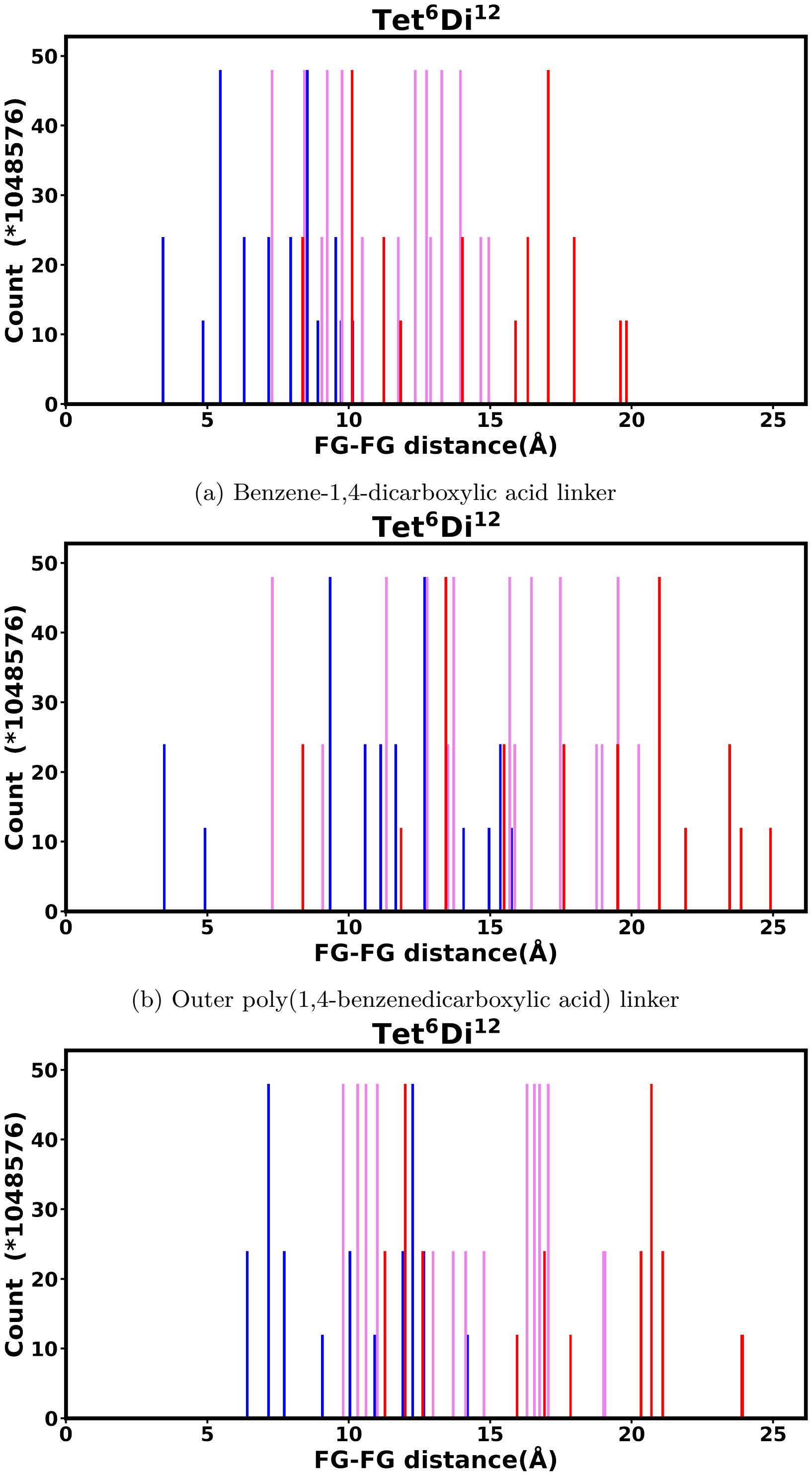


(a) Benzene-1,4-dicarboxylic acid linker

(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.

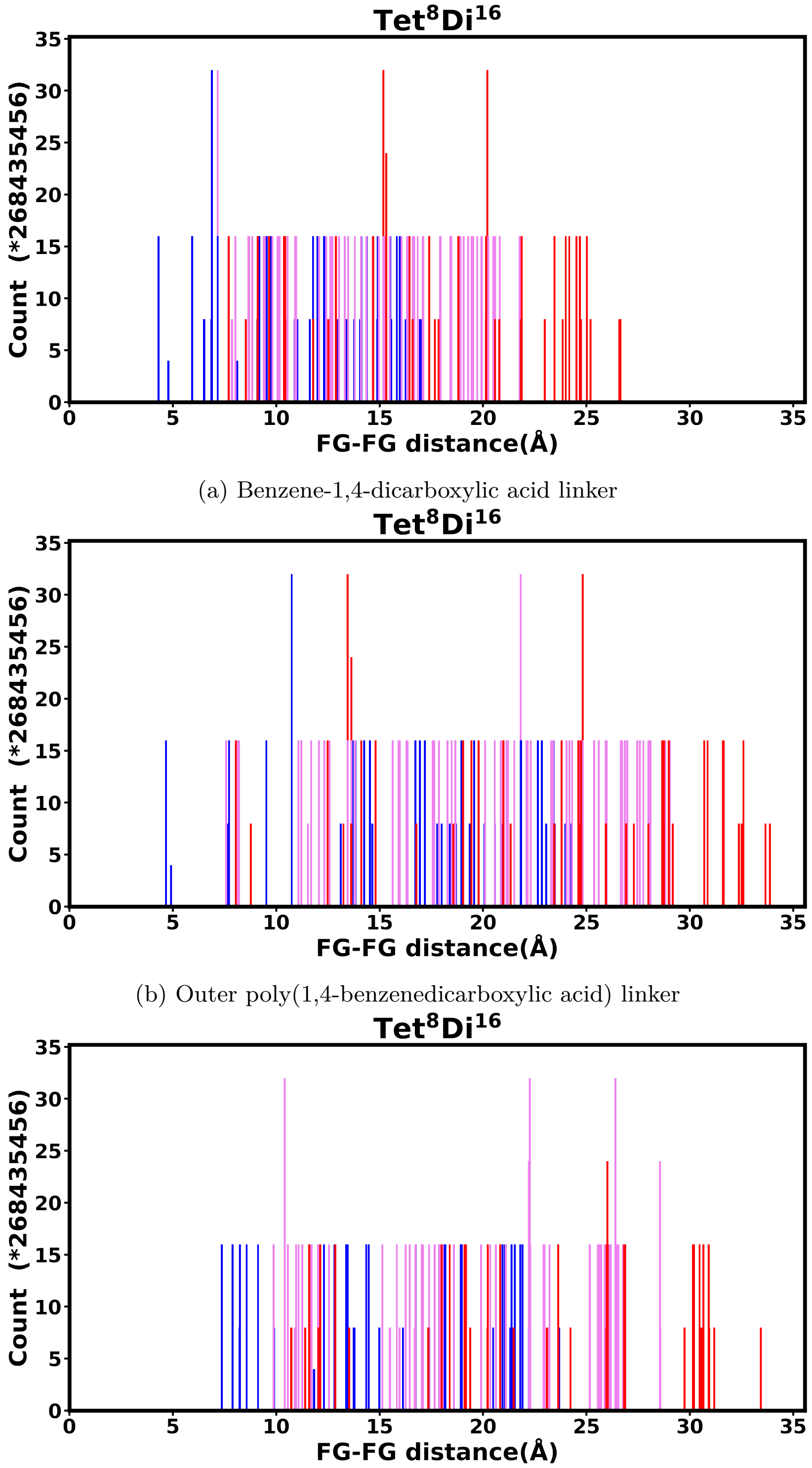


(a) Benzene-1,4-dicarboxylic acid linker

(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.

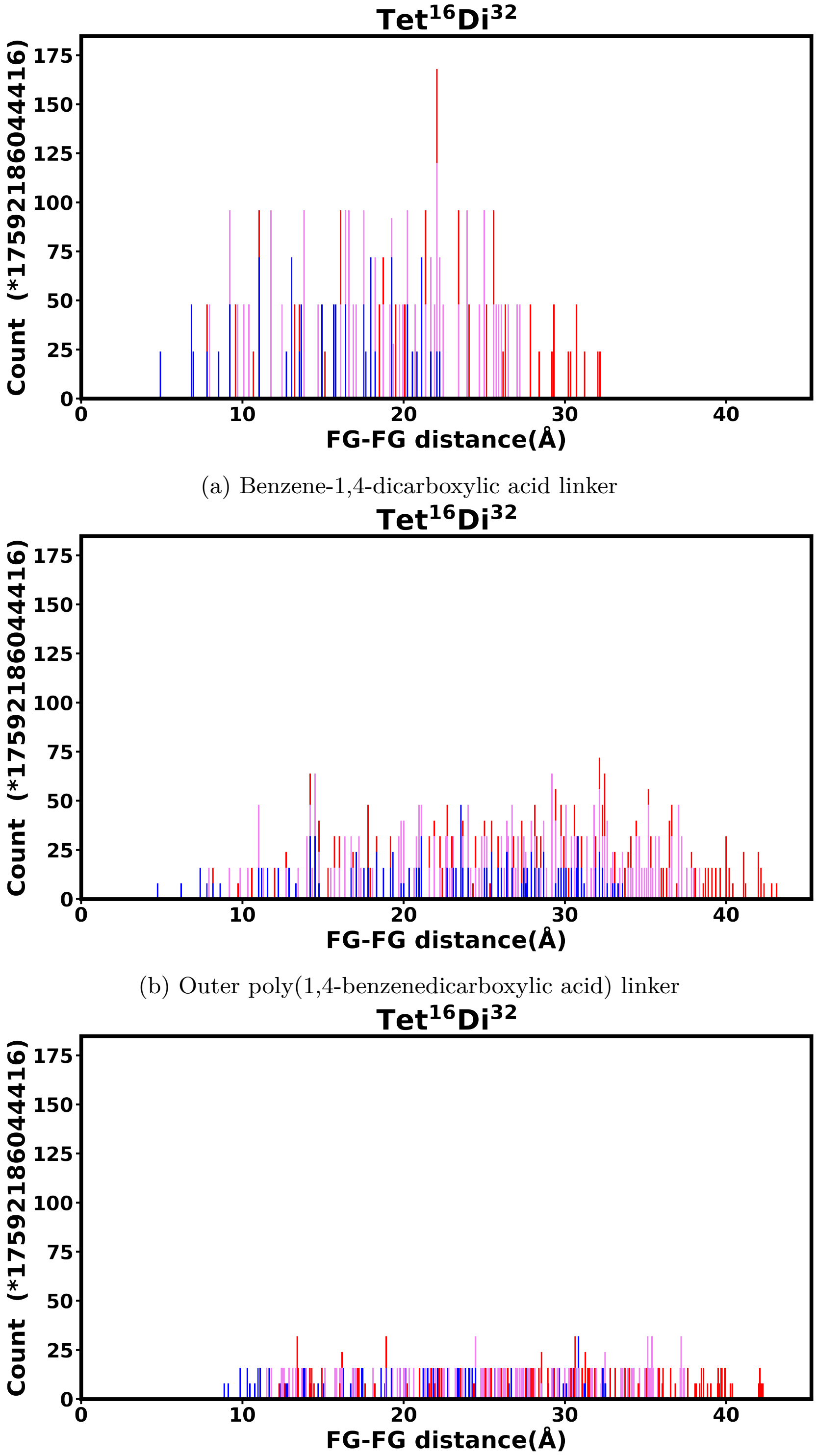


(a) Benzene-1,4-dicarboxylic acid linker

(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.

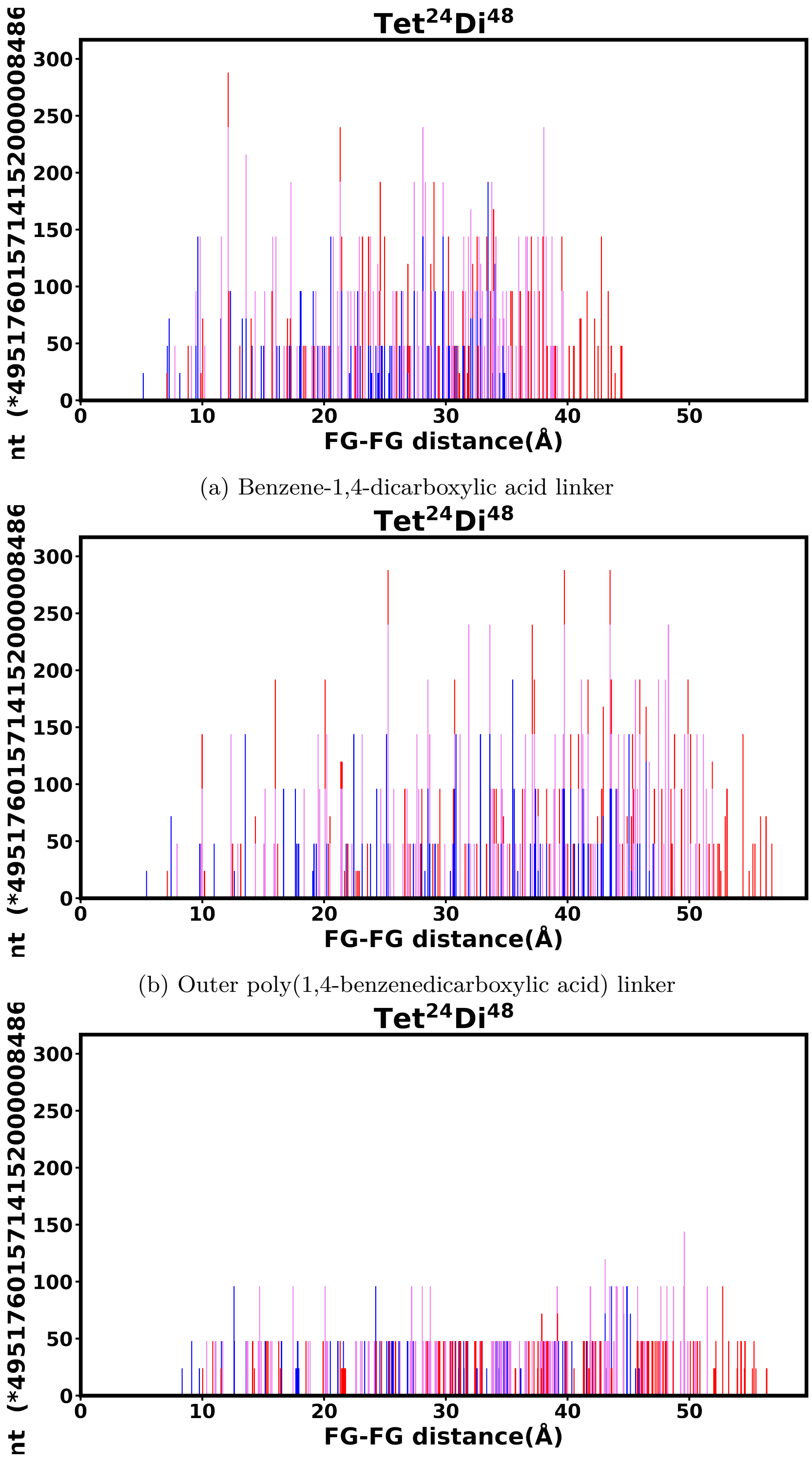


(a) Benzene-1,4-dicarboxylic acid linker

(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

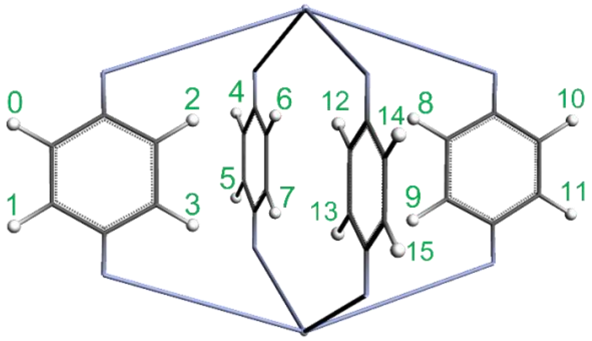
(b) Outer poly(1,4-benzenedicarboxylic acid) linker

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

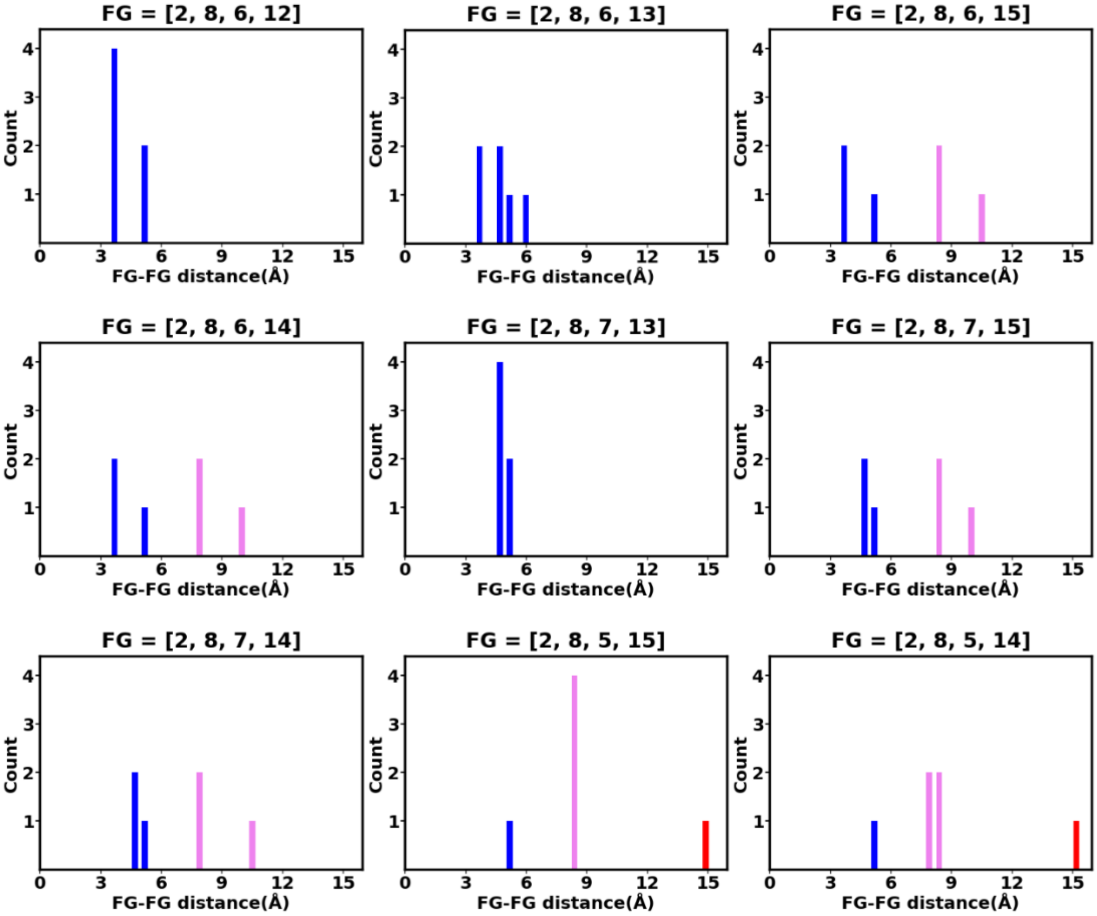
Figure S19: Histogram of FG-FG distances in pore $Tet^{24}Di^{48}$, constructed from a node with radius 5Å and two different linker sizes, (a) benzene-1,4-dicarboxylic acid, (b) poly(1,4-benzenedicarboxylic acid) outer functional group positions (c) poly(1,4-benzenedicarboxylic acid) inner functional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

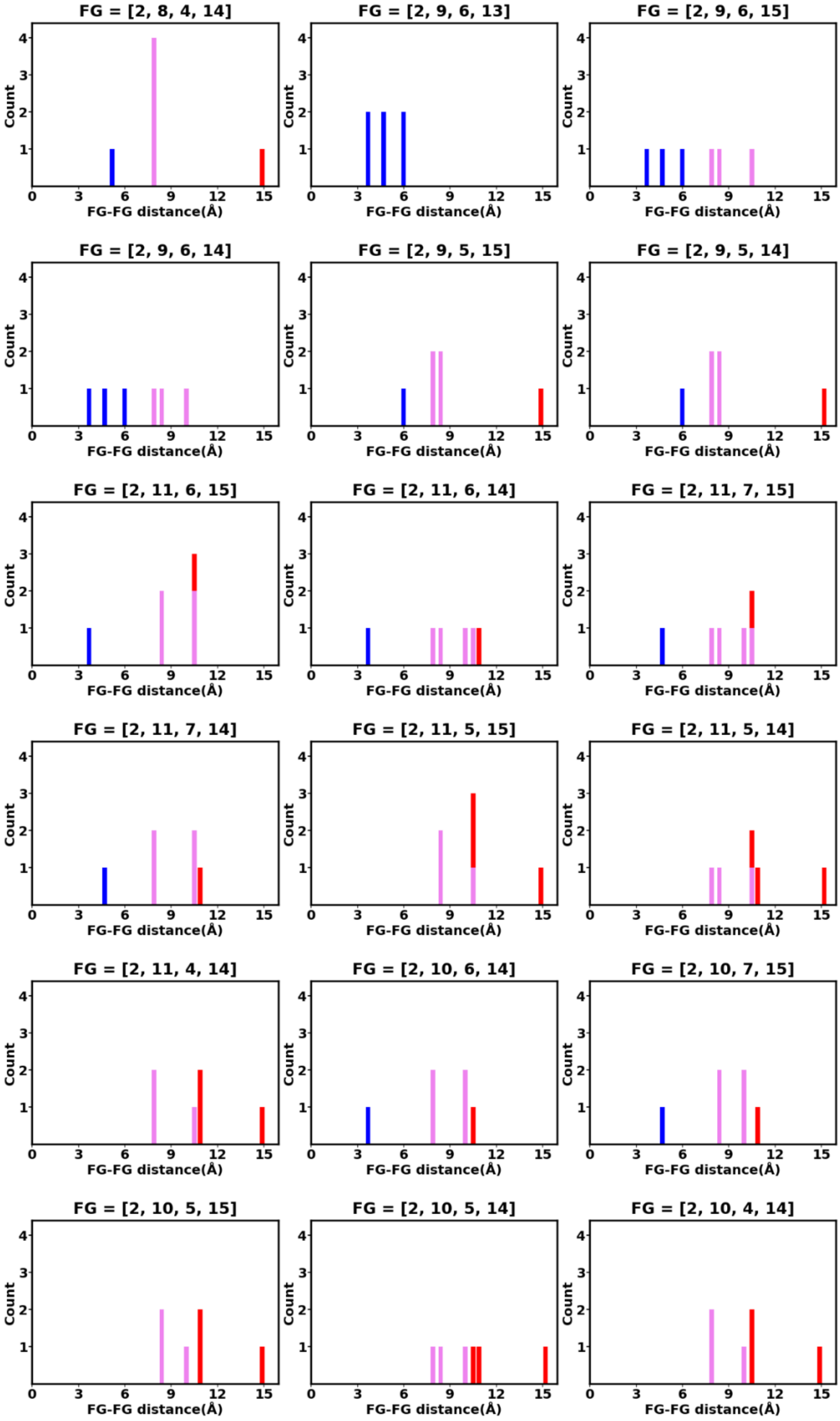
S6 Functional Group - Functional Group distance Histogram of Tet^2Di^4 Isomers.

(a)



(b)




FG = [2, 8, 4, 14]
FG = [2, 9, 6, 13]
FG = [2, 9, 6, 15]
FG = [2, 9, 6, 14]
FG = [2, 9, 5, 15]
FG = [2, 9, 5, 14]
FG = [2, 11, 6, 15]
FG = [2, 11, 6, 14]
FG = [2, 11, 7, 15]
FG = [2, 11, 7, 14]
FG = [2, 11, 5, 15]
FG = [2, 11, 5, 14]
FG = [2, 11, 4, 14]
FG = [2, 10, 6, 14]
FG = [2, 10, 7, 15]
FG = [2, 10, 5, 15]
FG = [2, 10, 5, 14]
FG = [2, 10, 4, 14]
Count
1
2
3
4
FG-FG distance(Å)
0
3
6
9
12
15

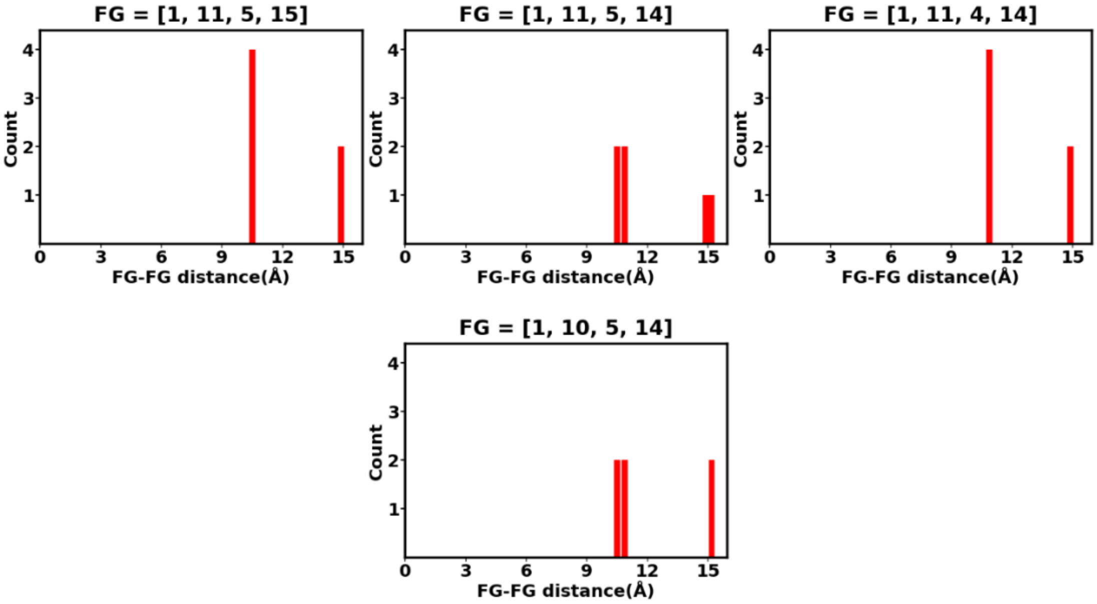


Figure S6.1. (a) A Tet^{2}Di^4 cage is constructed from a node with radius 5Å and linker benzene-1,4-dicarboxylic acid (bdc). The functionalisation based on one per-bdc linker would generate 31 unique isomers. (b) The functional Group - functional group distance histogram of Tet^{2}Di^4 Isomers.

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

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)
3.68	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

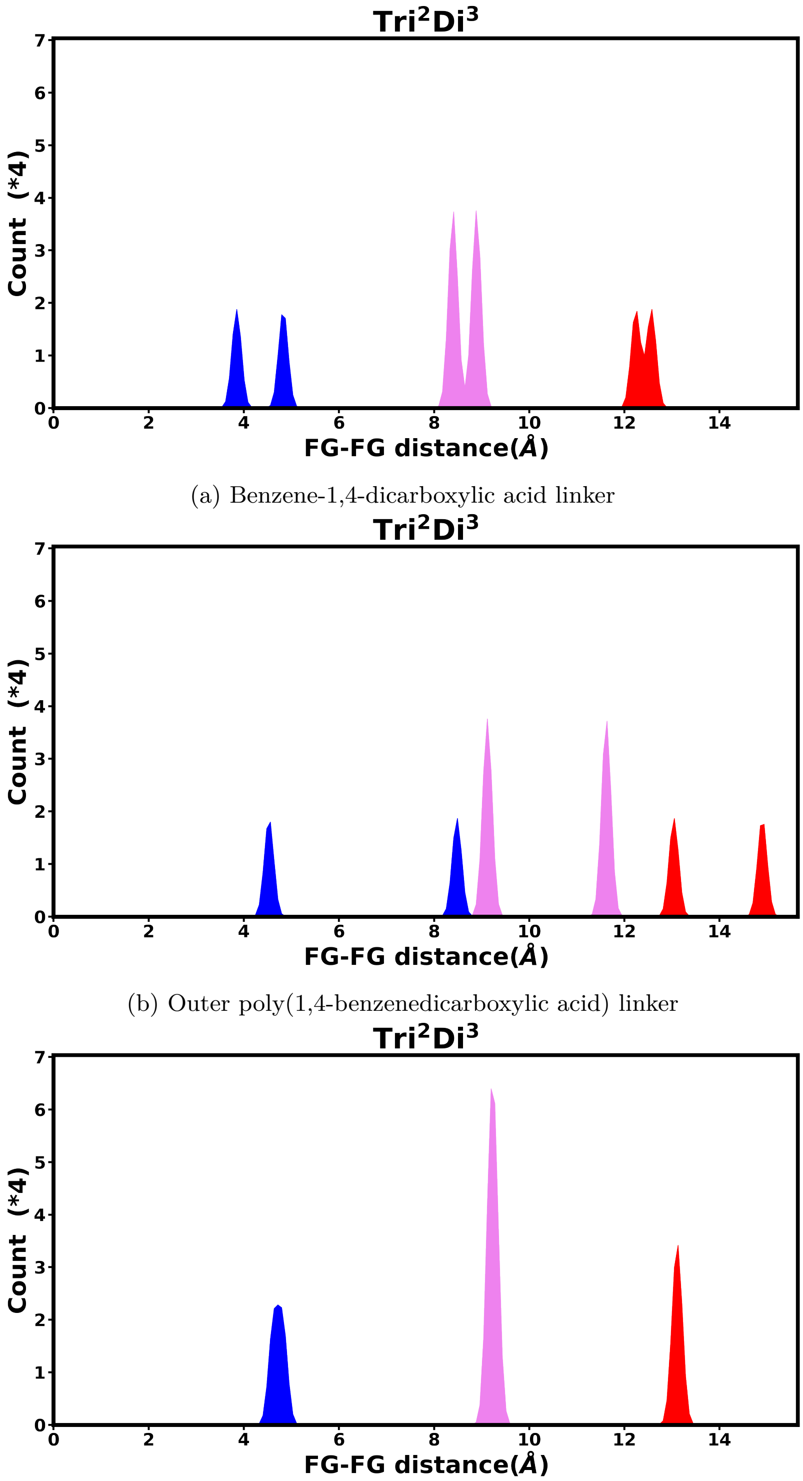
Table S3: Distances found in Tri^4Di^6 pore. The linker rotation deviates the distances from its symmetric position, the maximum and minimum distance is tabulated. Distances are colour-coded: blue represents In-In distances, violet In-out distances, red out-out distances.

FG-FG distance(Å)	Shortest Distance(Å)	Maximum distance(Å)
3.14	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 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.

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

S8 Kernel Density Estimation on the Histograms

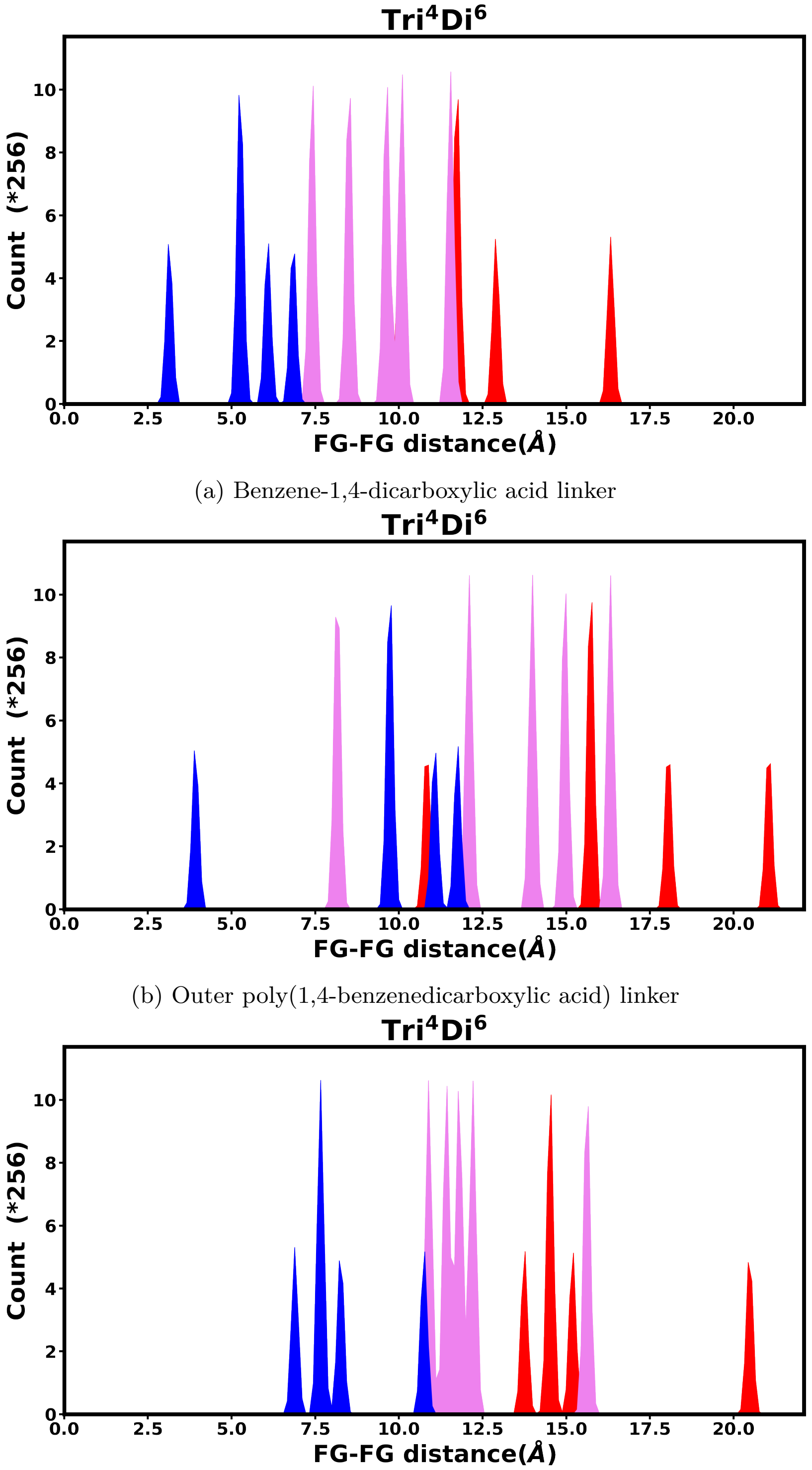


(a) Benzene-1,4-dicarboxylic acid linker

(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.

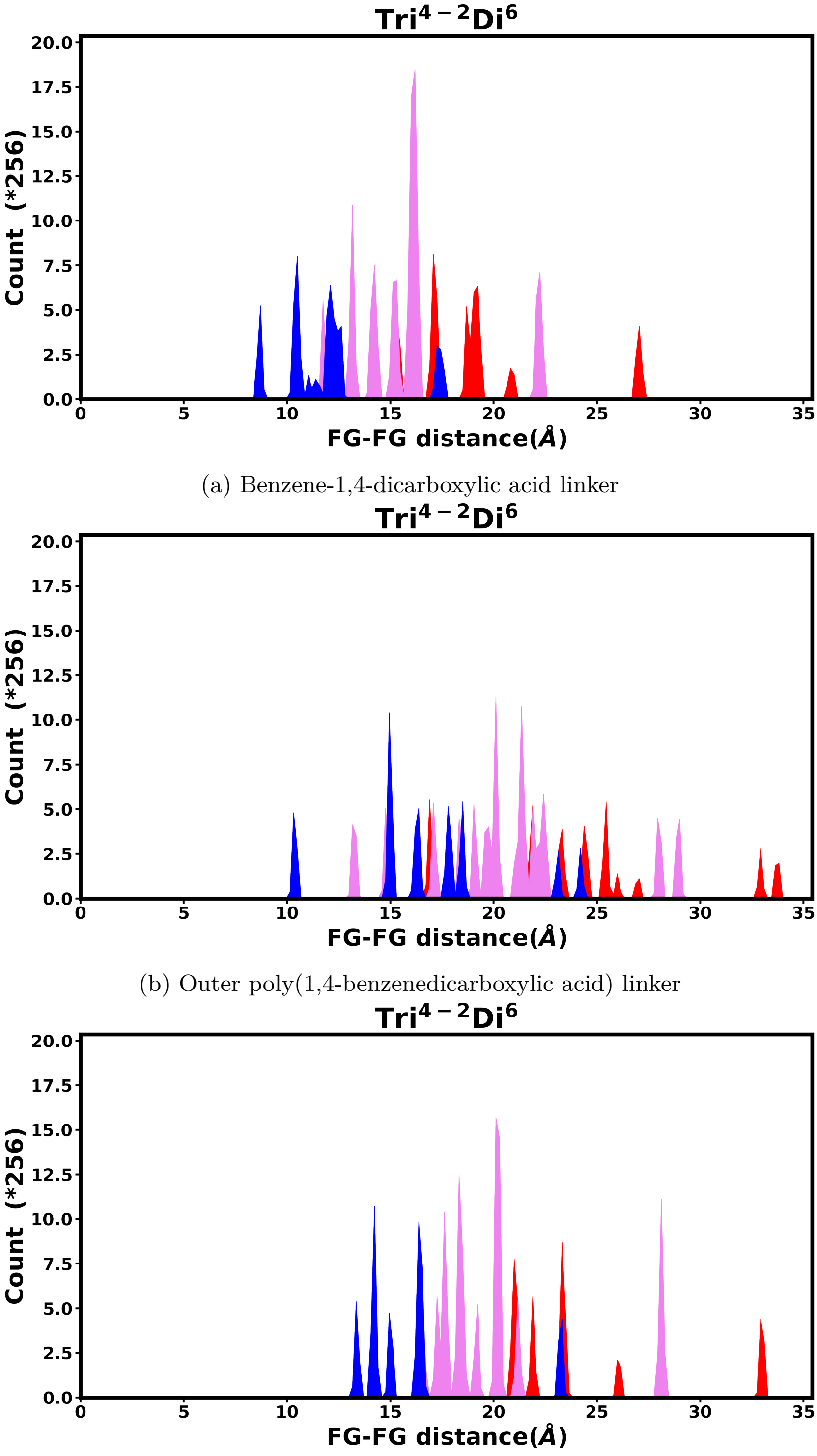


(a) Benzene-1,4-dicarboxylic acid linker

(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.

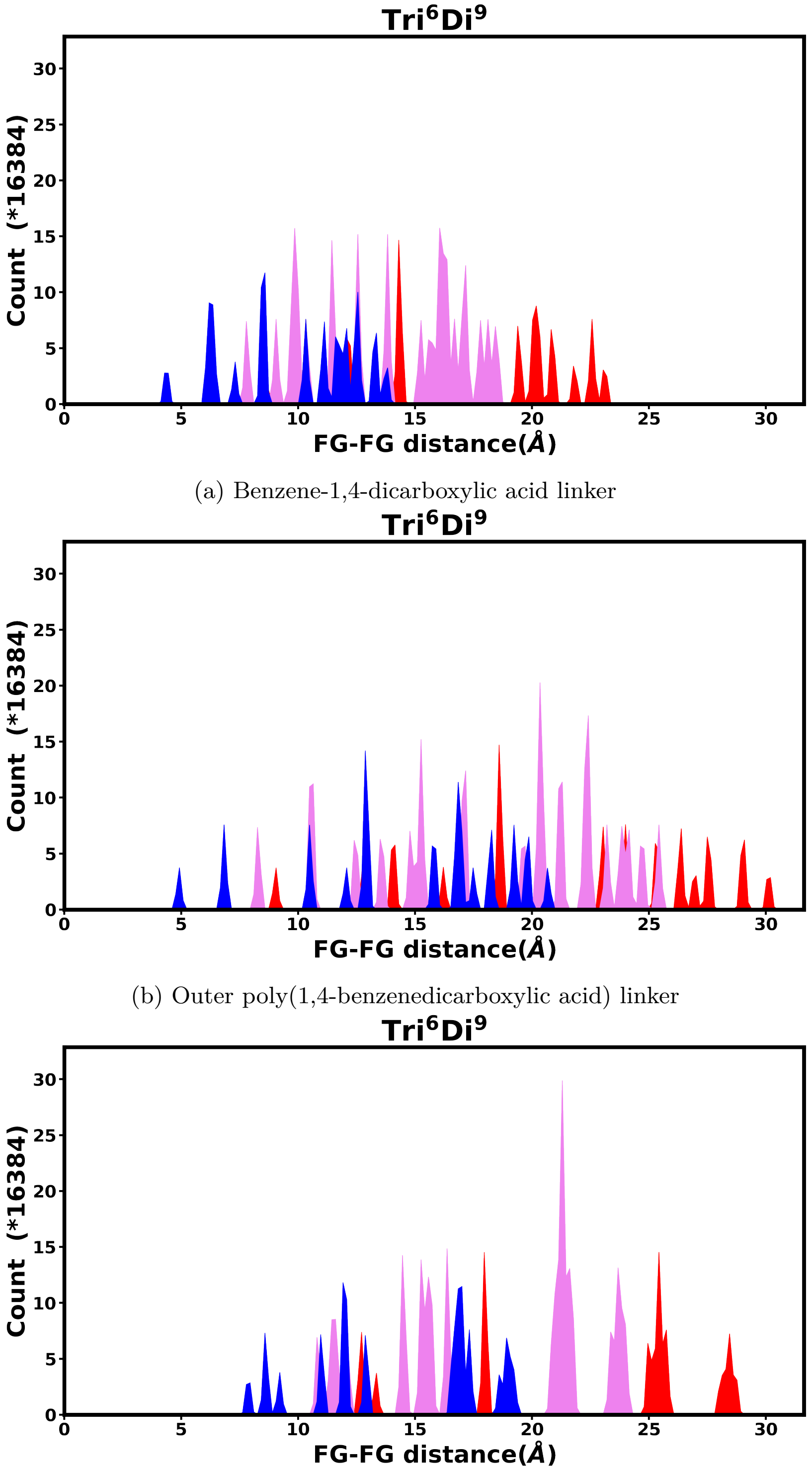


(a) Benzene-1,4-dicarboxylic acid linker

(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.

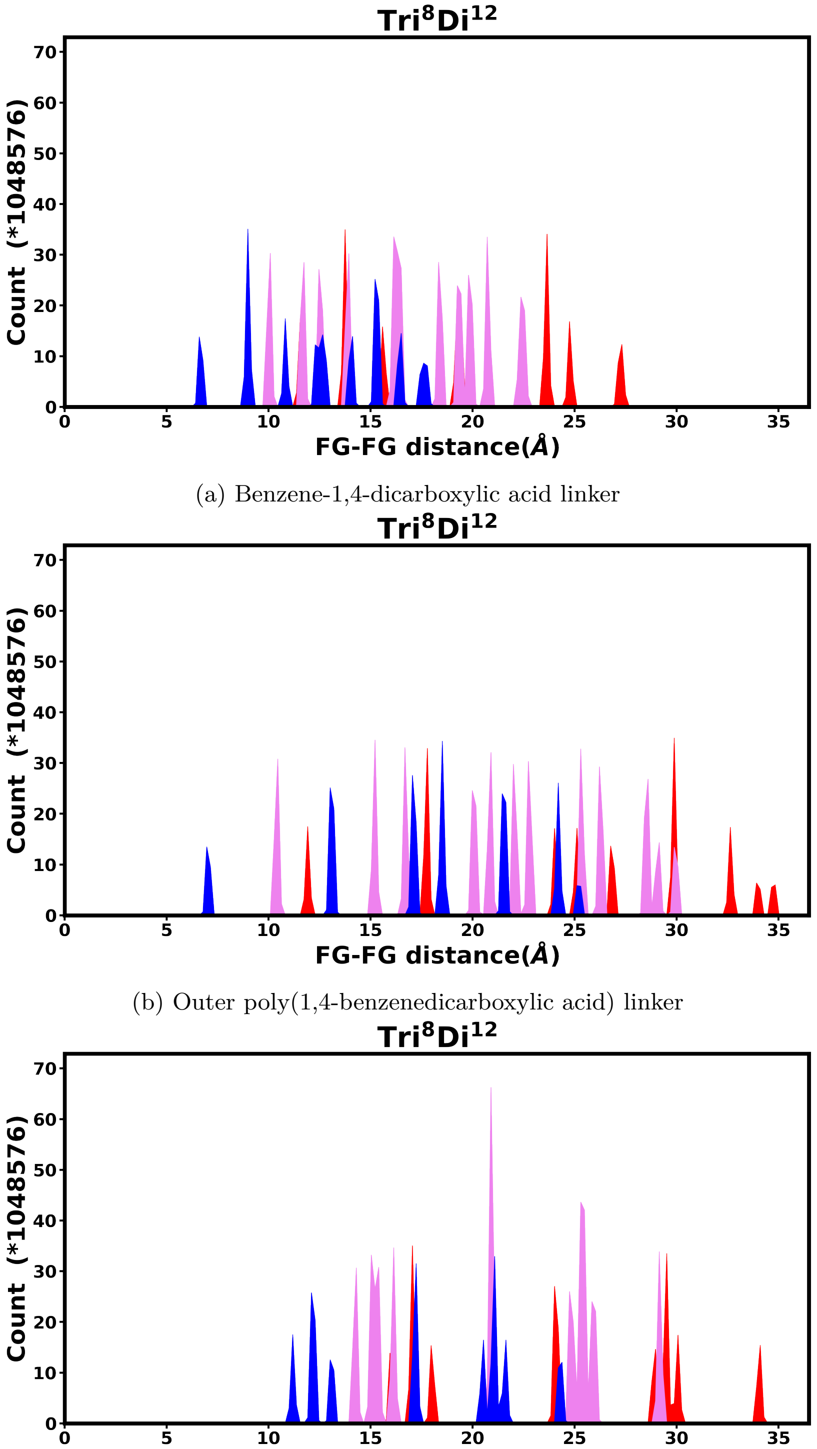


(a) Benzene-1,4-dicarboxylic acid linker

(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.

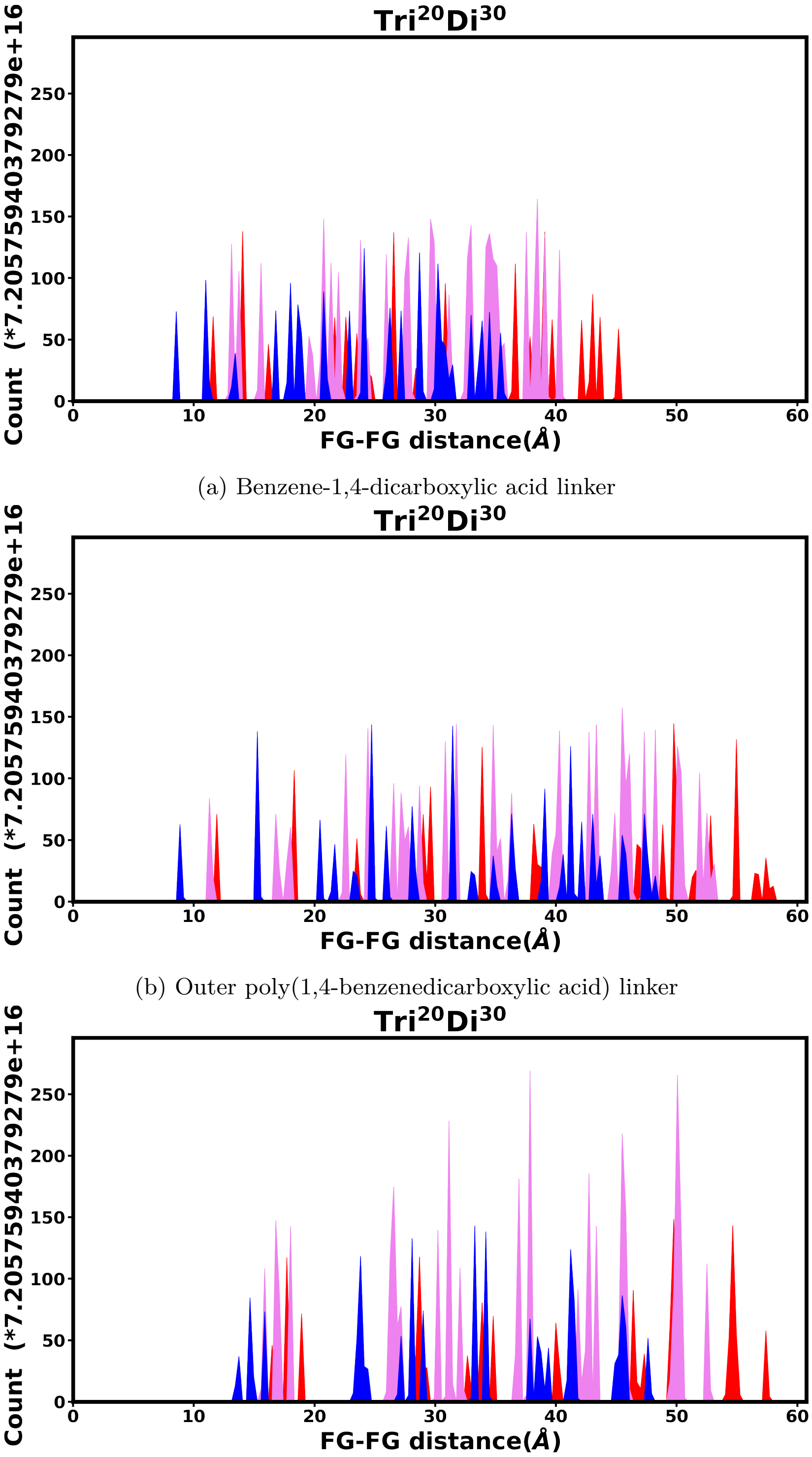


(a) Benzene-1,4-dicarboxylic acid linker

(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.

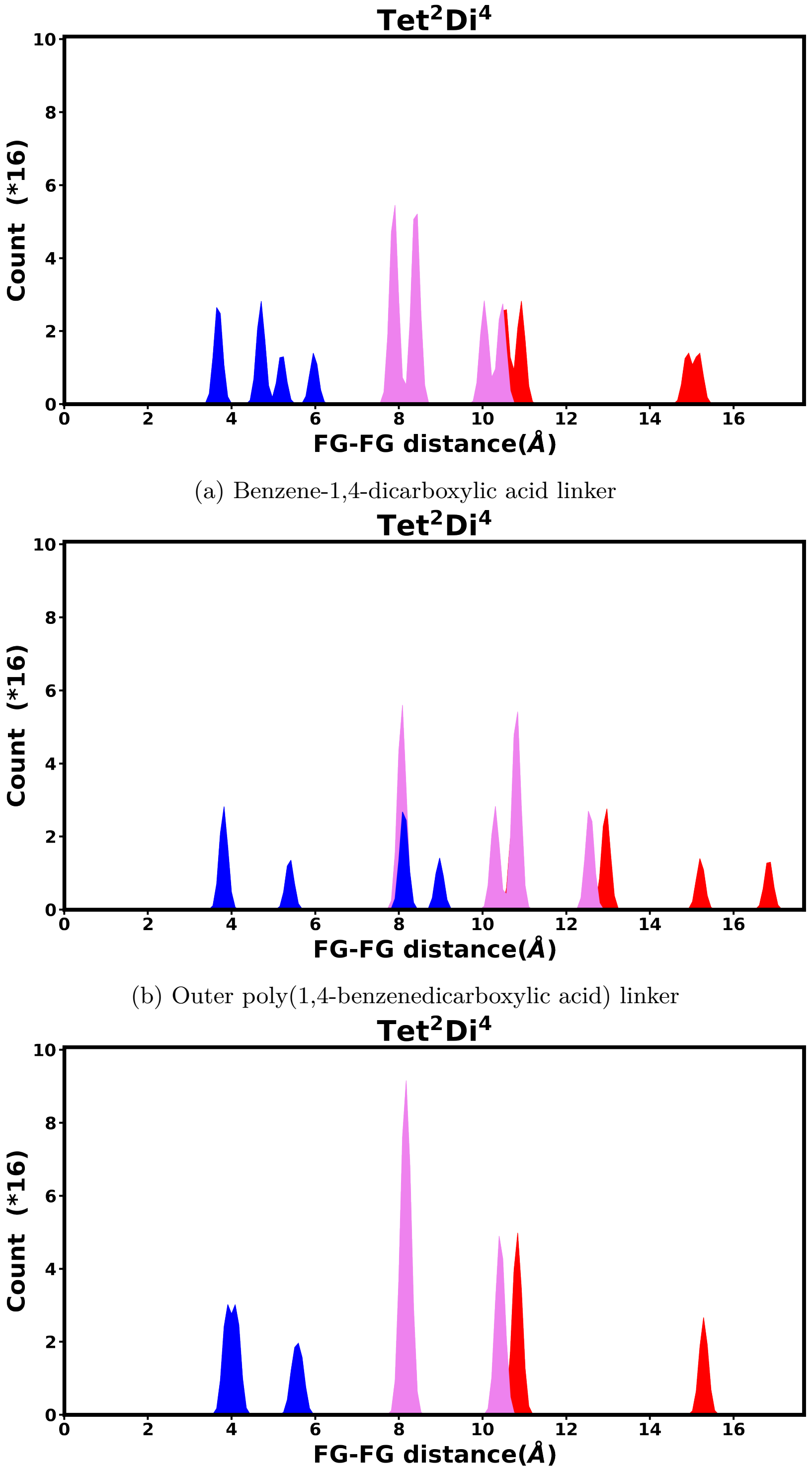


(a) Benzene-1,4-dicarboxylic acid linker

(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.

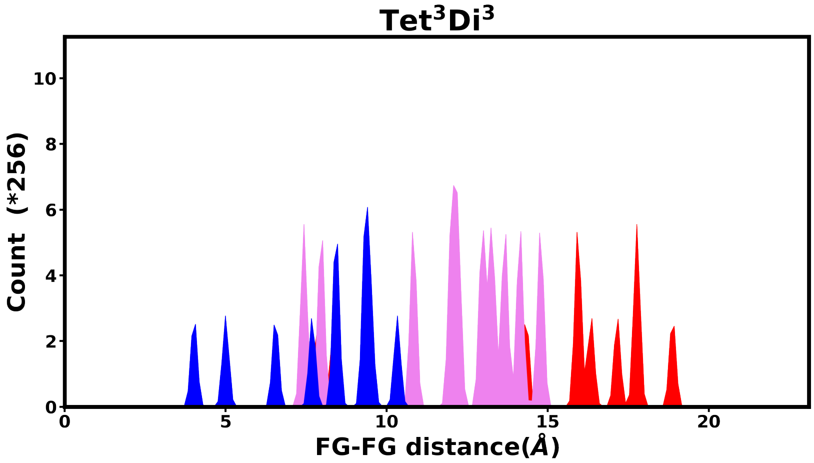


(a) Benzene-1,4-dicarboxylic acid linker

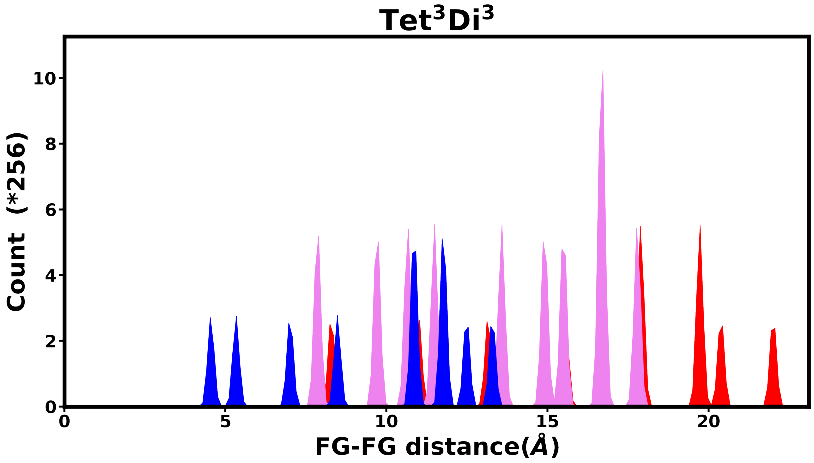
(b) Outer poly(1,4-benzenedicarboxylic acid) linker

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

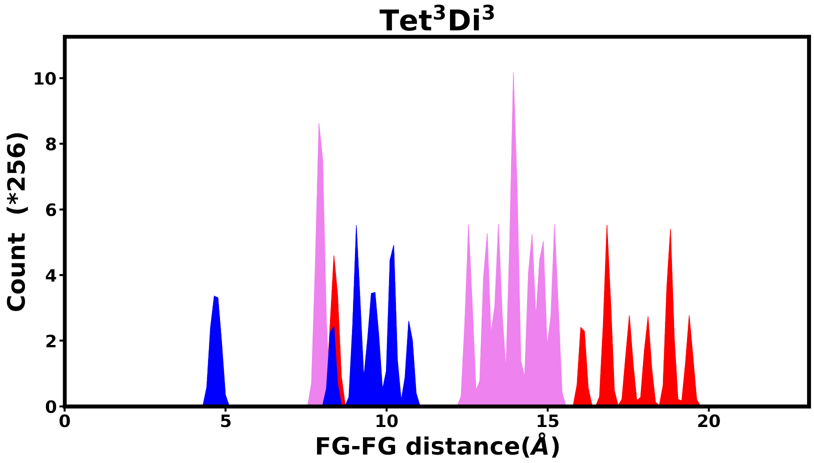
Figure S26: KDE applied on FG-FG distance histograms of pore Tet^2Di^4, constructed from a node with radius 5Å and two different linker sizes, (a) benzene-1,4-dicarboxylic acid, (b) poly(1,4-benzenedicarboxylic acid) outer functional group positions (c) poly(1,4-benzenedicarboxylic acid) inner functional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



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

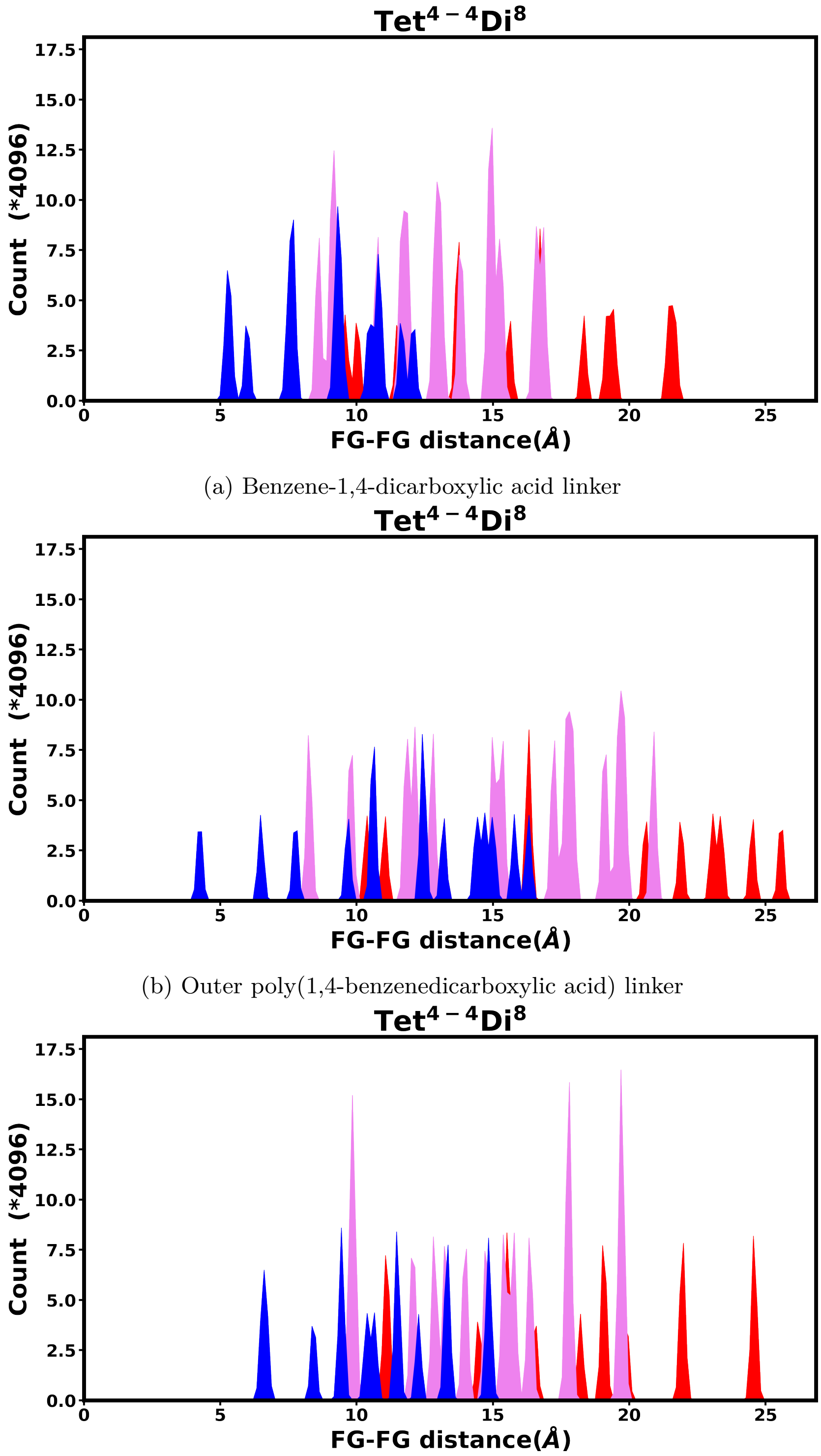


(b) Outer poly(1,4-benzenedicarboxylic acid) linker



(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.

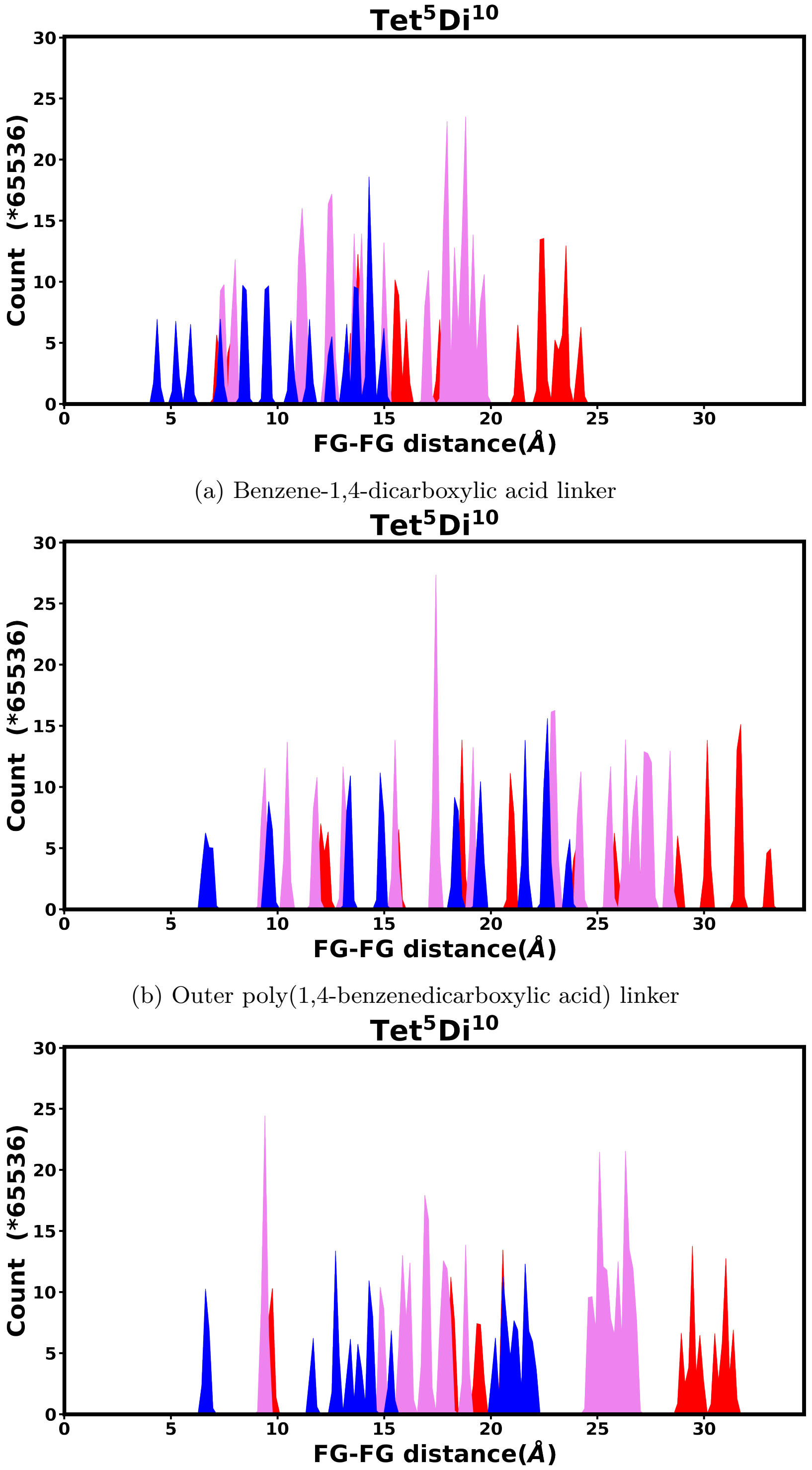


(a) Benzene-1,4-dicarboxylic acid linker

(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.

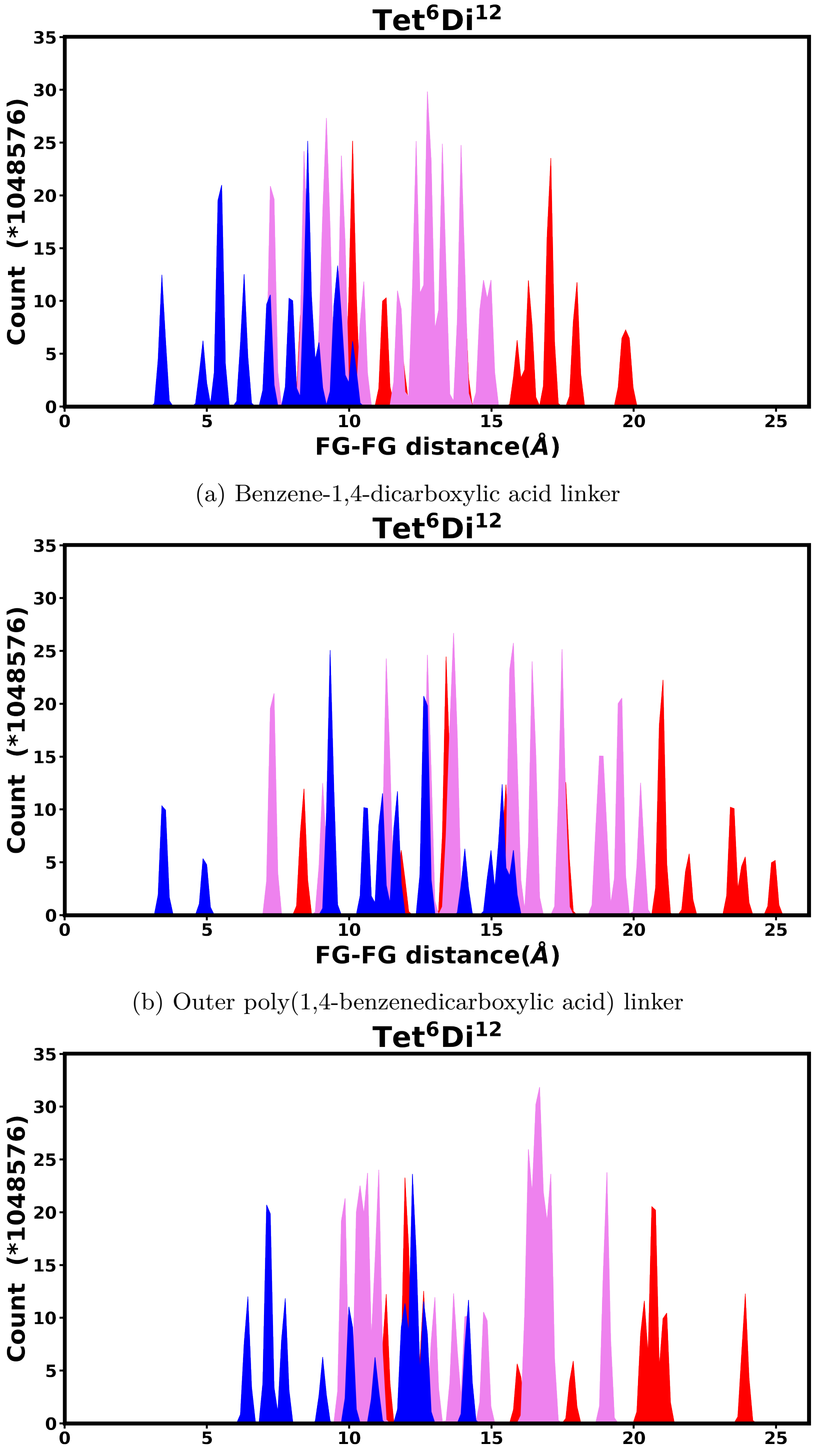


(a) Benzene-1,4-dicarboxylic acid linker

(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.

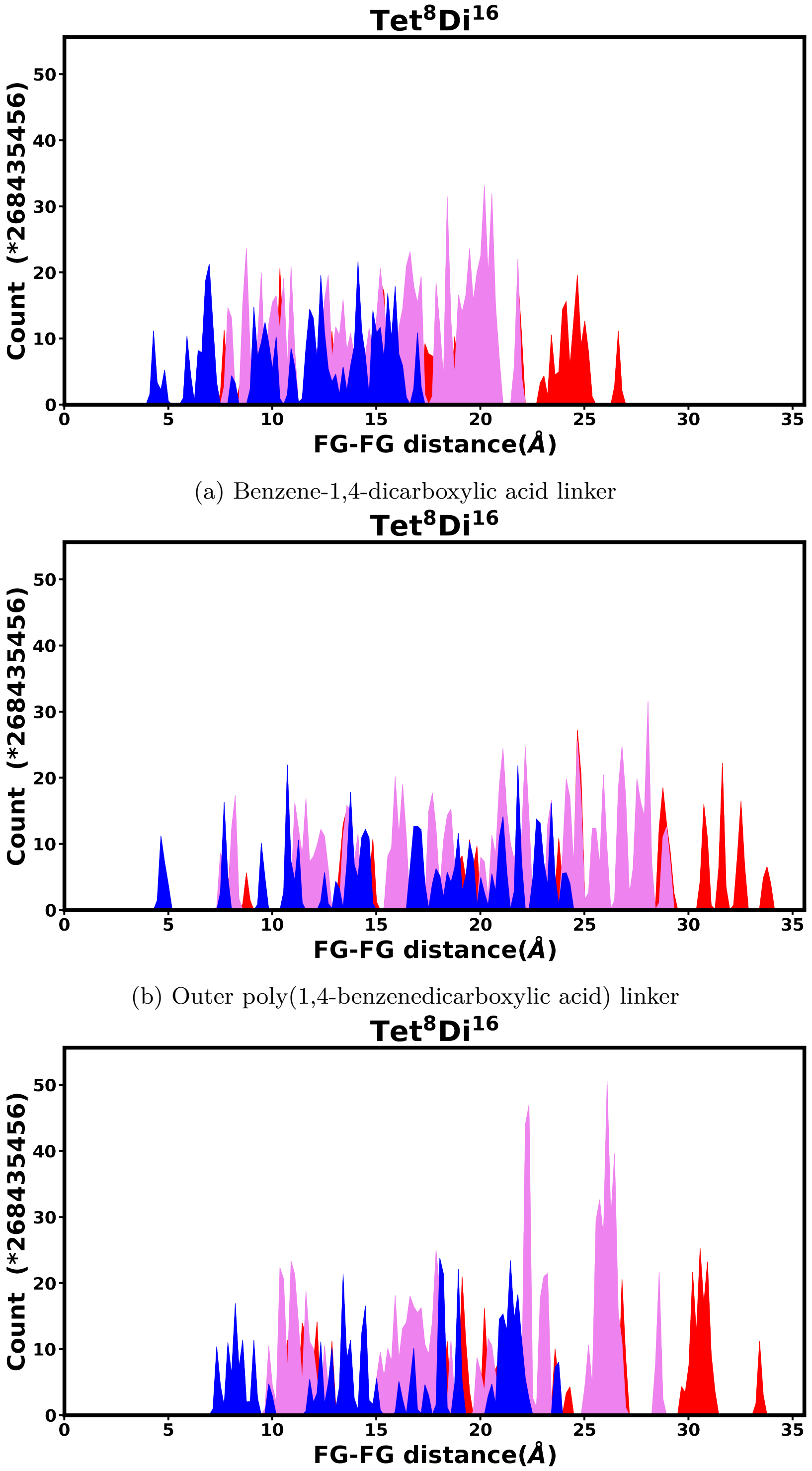


(a) Benzene-1,4-dicarboxylic acid linker

(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.

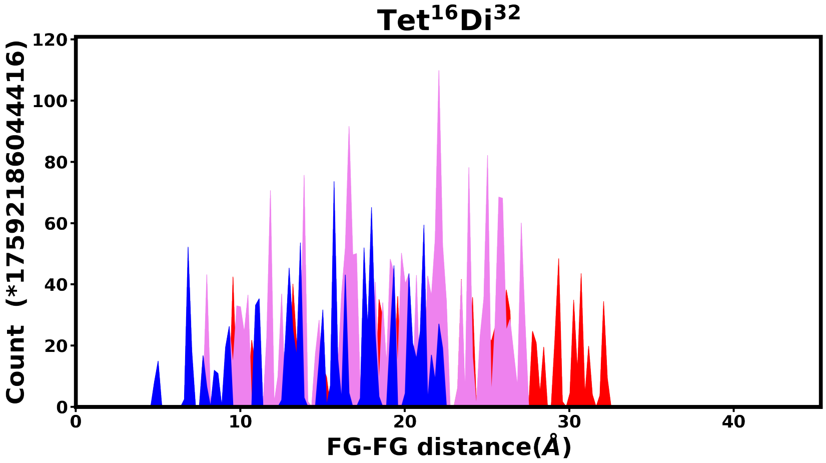


(a) Benzene-1,4-dicarboxylic acid linker

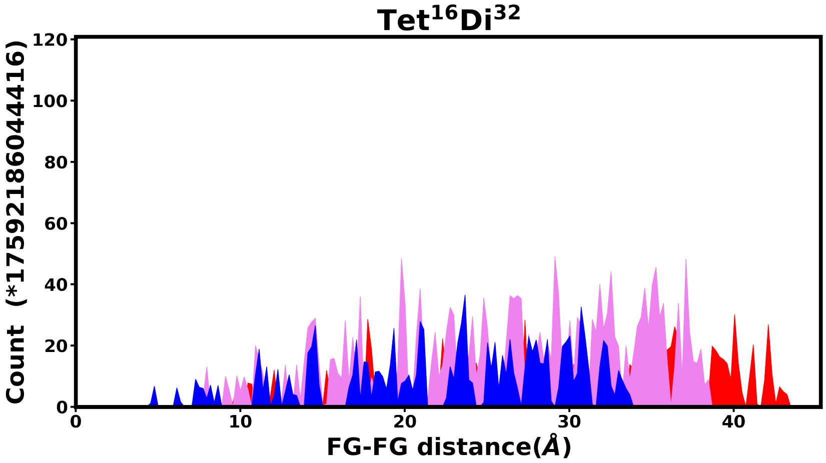
(b) Outer poly(1,4-benzenedicarboxylic acid) linker

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

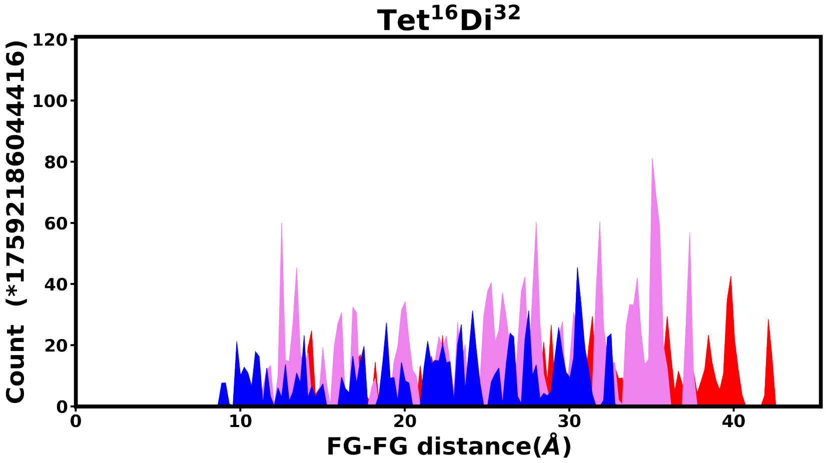
Figure S31: KDE applied on FG-FG distance histograms of pore $Tet^{8}Di^{16}$, constructed from a node with radius 5Å and two different linker sizes, (a) benzene-1,4-dicarboxylic acid, (b) poly(1,4-benzenedicarboxylic acid) outer functional group positions (c) poly(1,4-benzenedicarboxylic acid) inner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



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

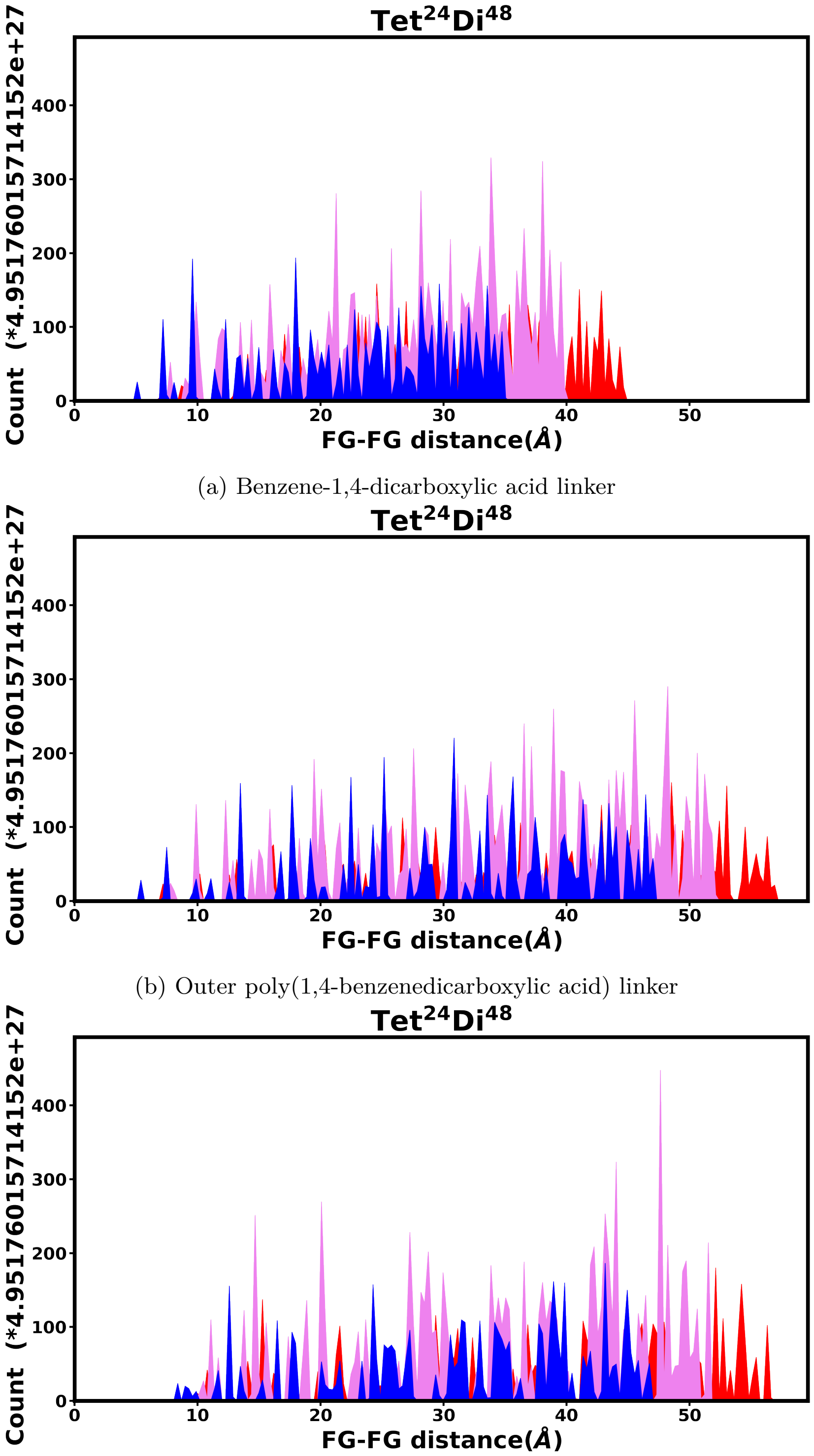


(b) Outer poly(1,4-benzenedicarboxylic acid) linker



(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.

10 isomers of each topology were used for the input data of the Principal Component Analysis (PCA). The histogram of each isomer was generated and kernel density estimation was then applied on the histograms using gaussian kernel and bandwidth 0.1. The histograms were converted to a matrix with upper limit of 60Å and 1Å wide bins.

S10. Representative Isomers of NH2-UiO-66 Oh Cage for Machine Learning Input.

We carefully selected a number of representative structures based on four features of geometrical properties we are particularly interested in:

- The number of functional groups pointing into the centre of the pore (Section S10.1).
- The number of functional groups on the windows of the pore (Section 0).
- The number of functional groups adjacent to one node (Section 0).
- The functional group - functional group distance set (Section 0).

The final choice of representative isomers are computed by characterising each feature by a string of numbers, every unique combination of the four features is one representative isomer.

S10.1. The number of FGs pointing into the centre of the pore.

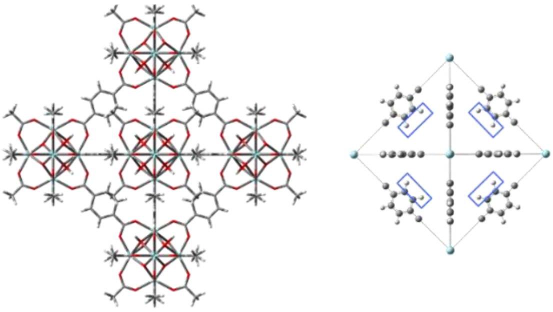


Figure S10.1. (right) Octahedral pore of UiO-66. (left) benzene linkers of octahedral pore, blue rectangle showing functional group slots that is pointing into the centre of the pore.

In the UiO-66 octahedral pore, the bdc planar ring is oriented such that the functional group is either facing into or outwards the pore, as shown in Figure S10.1. The functional group that is facing into the pore is more important to the guest molecule, while the outer FG only gives a small electronic effect (through electron withdrawal / donation with respect to the pristine linker) to the guest molecule inside the pore. Thus we care about how many functional groups are inside the pore. This feature is characterised by a string of one single number, the number of FG pointing into the pore, as shown in Figure S10.2., an example is presented in Figure S10.3.

“ **Number of FG pointing into the pore** ”

Figure S10.2. Descriptor to represent the number of functional groups pointing inside the pore.

Example:

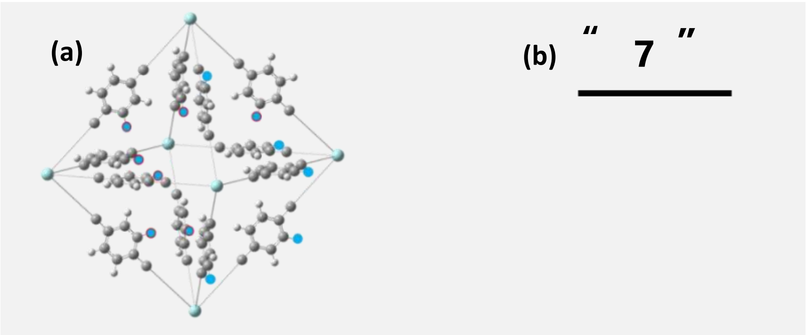


Figure S10.3. (a) An example of functionalised UiO-66 octahedral pore. (b). There are 7 functional groups pointing into the pore, thus the descriptor is ‘’7’’.

S10.2. The number of functional groups on the windows of the pore.

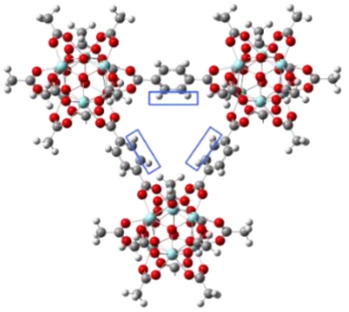


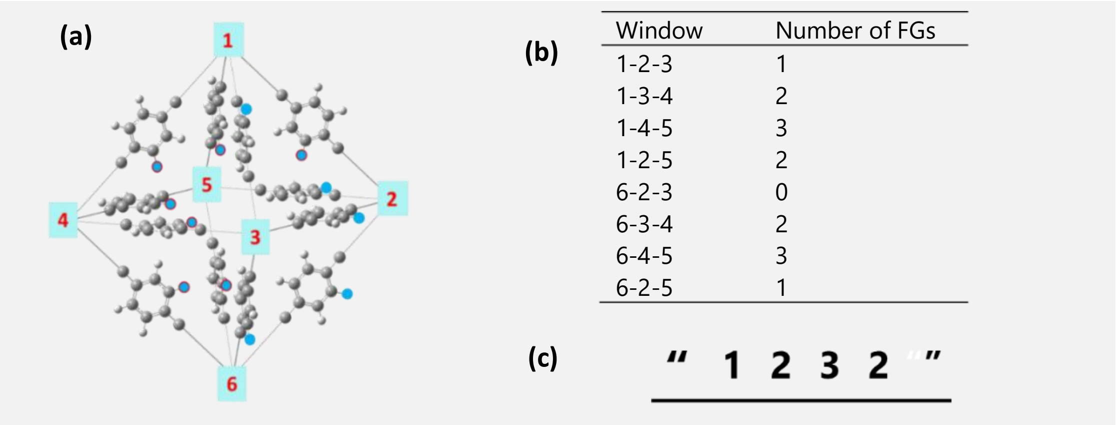
Figure S10.4. A window of UiO-66 octahedral pore. The window is bordered by three linkers, the blue rectangle shows FG slots that is pointing into the pore centre.

The number of FG located on a window is important for the transport property of the guest molecules. The octahedral pore has 8 windows, every window is bordered by 3 linkers (Figure 4.). As one window has 3 linkers thus the number of functional groups on one window is in between 0-3. Figure 5. shows a string of 4 numbers, with each of the numbers representing the total number of FGs. An example of how the descriptor is calculated is in Figure S10.6.

“	Number of windows with 0 FG	Number of windows with 1 FG	Number of windows with 2 FGs	Number of windows with 3 FGs	”

Figure S10.5. Descriptor to represent the arrangement of functional groups on the window.

Example:



Window	Number of FGs
1-2-3	1
1-3-4	2
1-4-5	3
1-2-5	2
6-2-3	0
6-3-4	2
6-4-5	3
6-2-5	1

“ 1 2 3 2 ”

Figure S10.6. (a) An example of functionalised UiO-66 octahedral pore (b) the table of windows and the corresponding number of FG in the window (c) the descriptor used in choosing representative structures.

S10.3. The number of FG adjacent to one node

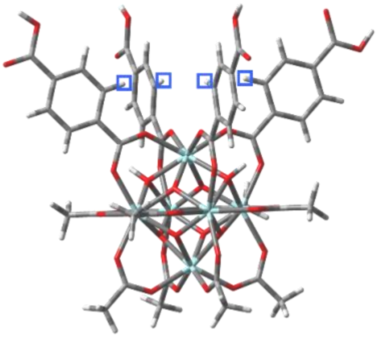


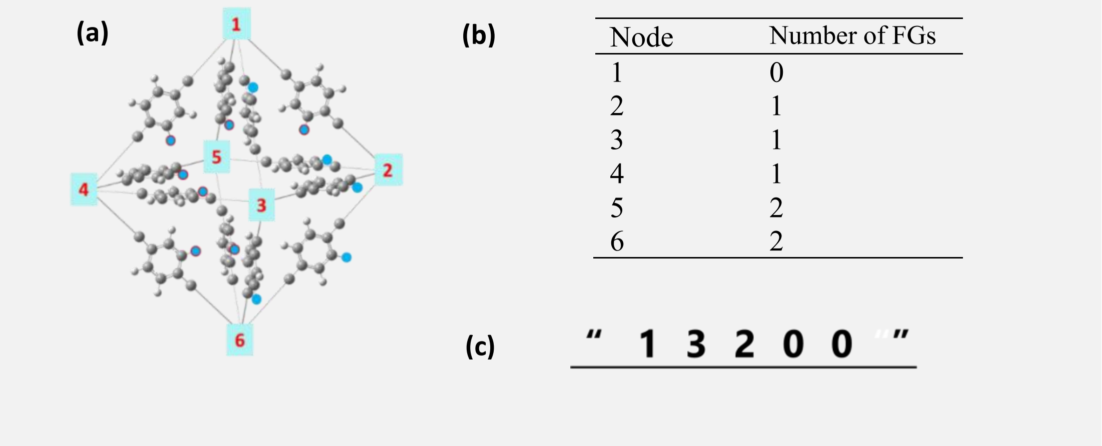
Figure S10.7. One corner of UiO octahedral pore. Blue squares showing the FG slot where functional group is adjacent to the node.

In the UiO-66 octahedral pore, there are 6 Zr_6O nodes, and every node is connected to 4 linkers, as shown in Figure 7. In many absorption situations, binding pockets are preferable for a given guest. We considered the functional groups adjacent to a node to be important because the functional groups in a pore corner are the most likely to be situated closely enough to form a binding pocket. The total number of FG adjacent to one node is in between 0-4, and similarly to the previous implementation, this feature is characterised by a string of 5 numbers, with each of the number representing the number node with the following conditions, as shown in Figure S10.8, an example is presented in Figure S10.9.

“	Number of nodes with 0 adjacent FG	Number of nodes with 1 adjacent FG	Number of nodes with 2 adjacent FGs	Number of nodes with 3 adjacent FGs	Number of nodes with 4 adjacent FGs	”

Figure S10.8. Descriptor to represent the arrangement of functional group adjacent to each node.

Example:



Node	Number of FGs
1	0
2	1
3	1
4	1
5	2
6	2

“ 1 3 2 0 0 ”

Figure S10.9. (a) An example of functionalised UiO-66 octahedral pore (b) the list of nodes and the corresponding number of FG adjacent to the node (c) the descriptor used in choosing representative structures

S10.4. The functional group - functional group distance set

Every structure has its own unique FG-FG distance set. In total, there are three types of FG-FG distances, shown in Figure 11, the first type is the distance which occurs between two FG located inside the pore (in-in), the second type is two FG located outside the pore (out-out), and the third type is where one inside with one outside the pore (in-out).
This feature is characterised by a string of its FG-FG distance set, shown in Figure 10.

“	FG-FG distance 1	FG-FG distance 2	FG-FG distance 3	and so on ..	”

Figure S10.10. Descriptor to represent the arrangement of functional group based on functional group-functional group distances.

So to eliminate similar cage isomers, we rounded the FG-FG distance and only consider the FG-FG distance that occur inside the pore as they are a more prominent influence to the guest molecule. The FG-FG distance we consider is shown by the blue line in the figure below, while the violet and red dotted line are ignored.

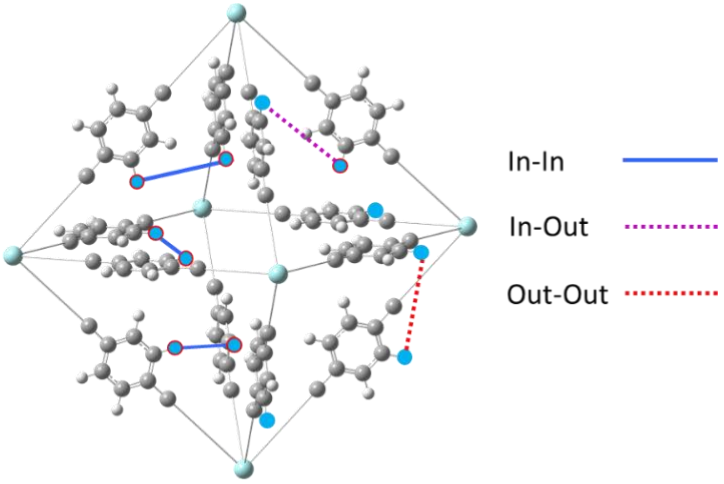


Figure S10.11. An example of functionalised UiO-66 octahedral pore. Blue line shows the distance between two FG that occur inside the pore, violet dotted line shows the distance between two FG that occur across the boundary of the pore, and red dotted line shows the distance that occur outside the pore.

S11 Machine Learning Method

S11.1. Dataset preparation

Before calculating the binding energy, the cages (with propranolol molecule inside) are relaxed to their stable position using GFN-xTB. During the relaxation, the Zr_6O node geometries are constrained, while the carboxylate oxygen atoms are allowed to move but the bond length to their neighbouring Zr atoms are restrained with a harmonic restraint in order to prevent bond breaking, while the rest of the atoms are free to move.

The molecule and cage single point energies are calculated from the relaxed structure. Finally, the binding energies are calculated by subtracting the individual cage and propranolol energies from the total energy (equation S11.1). The geometry optimisation and binding energy calculation are performed using the AMS 2020 software package.

$$E_{BE} = E_{cage\,\&\,propanolol} - \left(E_{cage} + E_{propanolol}\right) \qquad \text{(S11.1)}$$

Each isomer has 10 binding energies corresponding to 10 random propranolol positions, standard error is calculated for each isomer. The isomers with standard error of the binding energies more than 8 kJ/mol are discarded from the machine learning dataset, the total number of isomers become 3127.

S11.2. Machine learning model

The dataset is then trained on a supervised neural network, consisting of one layer with 6 neurons. The machine learning calculation is performed using scikit-learn python package.

The dataset is divided to 80:20 ratio for training data and testing data. The input of the machine learning is the cage isomer functional group pair distance histogram after KDE has been applied (gaussian kernel, bandwidth = 0.1Å, bin size of 0.2Å) and the response value f(x) of each cage isomer is the average binding energies.

After training using 80% of the data, the trained model is tested using the remaining 20% data. Finally, the binding energy of test data is compared, binding energy from machine learning versus binding energy from GFN-xTB calculation.

The dataset and code is available at:
https://github.com/maryamnhd/FG-Pair-Distance-Descriptor
https://github.com/maryamnhd/Cage-Isomers