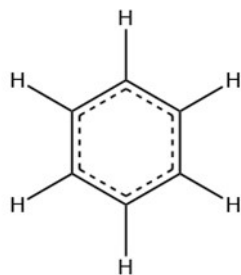


Supplementary Information for

Connectivity Stepwise Derivation (CSD) method: A Generic Chemical Structure Information Extraction Method for the Full Step Matrix

Jialiang Xiong¹, Xiaojie Feng¹, Jingxuan Xue², Yueji Wang², Haoren Niu¹, Yu Gu², Qingzhu

Jia², Qiang Wang¹, Fangyou Yan^{1*}



Benzene



L_c : [[2, 6, 7], [1, 3, 8], [2, 4, 9], [3, 5, 10], [4, 6, 11],
[1, 5, 12], [1], [2], [3], [4], [5], [6]]



```
0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0
```

All-zero matrix



```
0 1 0 0 0 1 1 0 0 0 0 0
1 0 1 0 0 0 0 1 0 0 0 0
0 1 0 1 0 0 0 0 1 0 0 0
0 0 1 0 1 0 0 0 0 1 0 0
0 0 0 1 0 1 0 0 0 0 1 0
1 0 0 0 1 0 0 0 0 0 0 1
1 0 0 0 0 0 0 0 0 0 0 0
0 1 0 0 0 0 0 0 0 0 0 0
0 0 1 0 0 0 0 0 0 0 0 0
0 0 0 1 0 0 0 0 0 0 0 0
0 0 0 0 1 0 0 0 0 0 0 0
0 0 0 0 0 1 0 0 0 0 0 0
```

Adjacency matrix

Figure S1. Complete process of extracting atomic connectivity relationships from benzene molecules containing hydrogen (H) atoms and storing it in list form (i.e., L_c) which in turn generates the adjacency matrix.

0	1	0	0	0	1	1	0	0	0	0	0
1	0	1	0	0	0	0	1	0	0	0	0
0	1	0	1	0	0	0	0	1	0	0	0
0	0	1	0	1	0	0	0	0	1	0	0
0	0	0	1	0	1	0	0	0	0	1	0
1	0	0	0	1	0	0	0	0	0	0	1
1	0	0	0	0	0	0	0	0	0	0	0
0	1	0	0	0	0	0	0	0	0	0	0
0	0	1	0	0	0	0	0	0	0	0	0
0	0	0	1	0	0	0	0	0	0	0	0
0	0	0	0	1	0	0	0	0	0	0	0
0	0	0	0	0	1	0	0	0	0	0	0

Adjacency matrix



0	1	2	3	2	1	1	2	3	4	3	2
1	0	1	2	3	2	2	1	2	3	4	3
2	1	0	1	2	3	3	2	1	2	3	4
3	2	1	0	1	2	4	3	2	1	2	3
2	3	2	1	0	1	3	4	3	2	1	2
1	2	3	2	1	0	2	3	4	3	2	1
1	2	3	4	3	2	0	3	4	5	4	3
2	1	2	3	4	3	3	0	3	4	5	4
3	2	1	2	3	4	4	3	0	3	4	5
4	3	2	1	2	3	5	4	3	0	3	4
3	4	3	2	1	2	4	5	4	3	0	3
2	3	4	3	2	1	3	4	5	4	3	0

Full step matrix(MS_F)

Figure S2. The process of generating the full connectivity matrix (MS_F) from the adjacency matrix is exemplified by benzene molecules containing hydrogen (H) atoms.

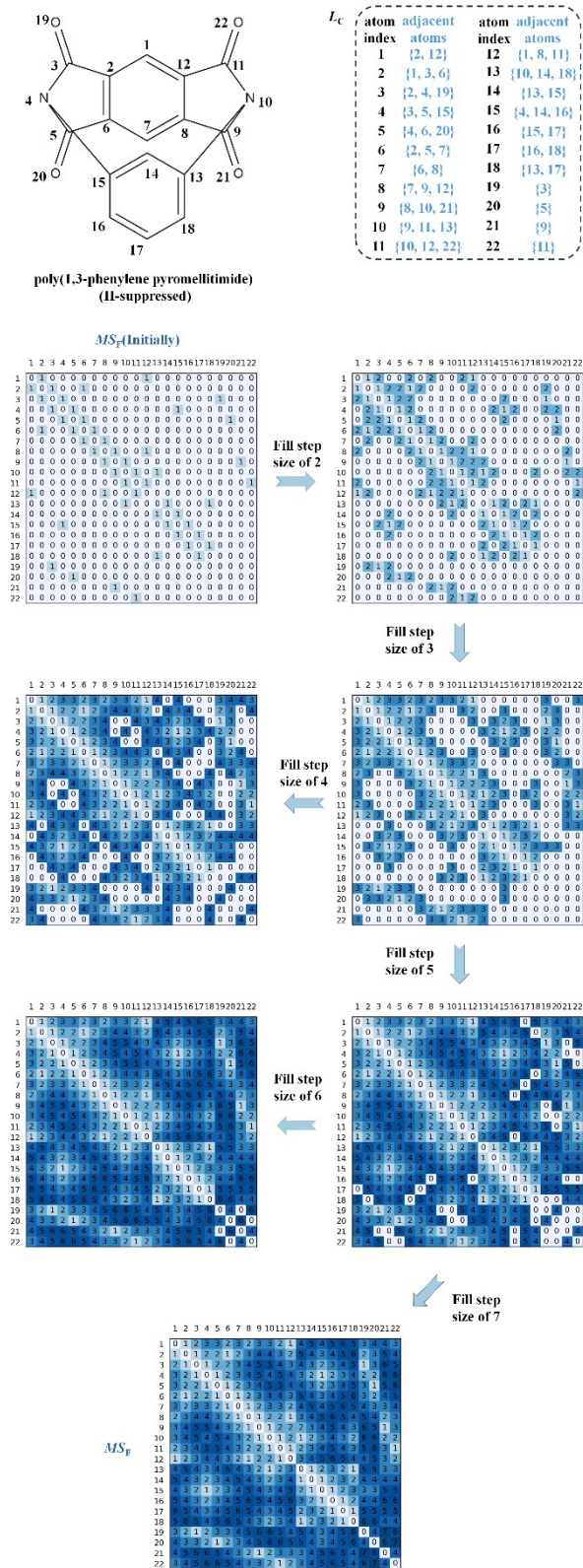


Figure S3. Illustration of the full step matrix (MS_F) generation from the H-suppressed poly (1,3-phenylene pyromellitimide) molecule.

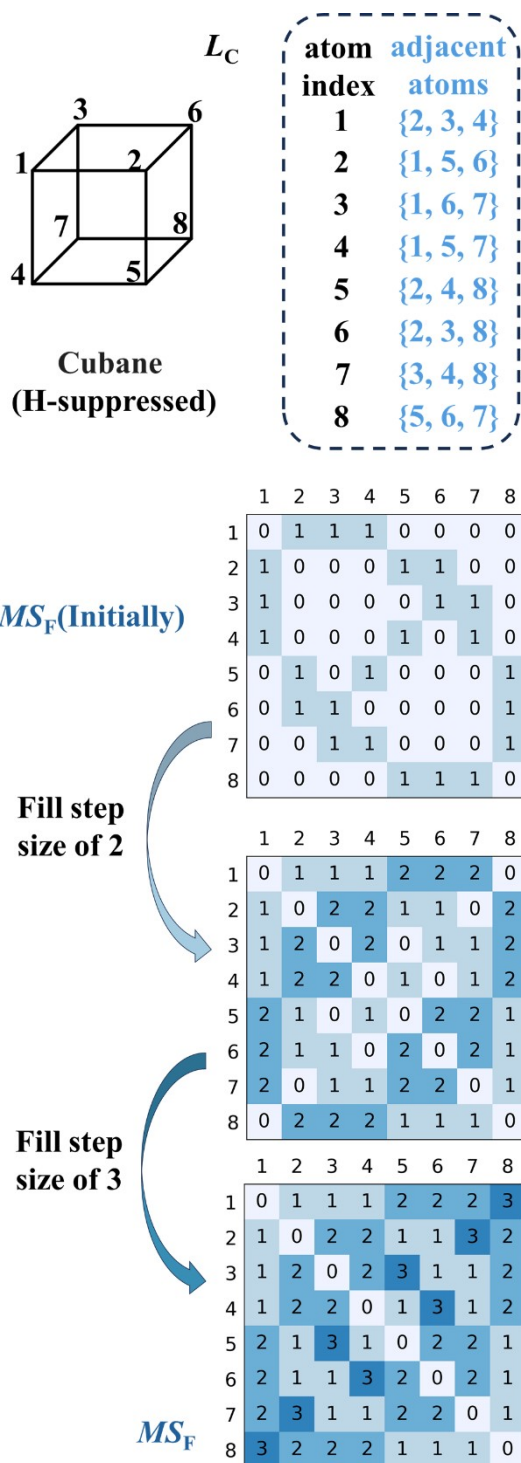


Figure S4. Illustration of the full step matrix (MS_F) generation from the H-suppressed cubane molecule. Note: The process of generating MS_F of the three-dimensional structural is demonstrated using cubane as an example.