Supporting Information for

Data Science-Centric Design, Discovery, and Evaluation of Novel Synthetically Accessible Polyimides with Desired Dielectric Constant

Mengxian Yu[1], Qingzhu Jia[1], Qiang Wang[1], Zheng-Hong Luo[2], Fangyou Yan[1*], Yin-Ning Zhou[2*]

1 School of Chemical Engineering and Material Science, Tianjin University of Science and Technology, Tianjin 300457, P. R. China

2 Department of Chemical Engineering, School of Chemistry and Chemical Engineering, State Key Laboratory of Metal Matrix Composites, Shanghai Jiao Tong University, Shanghai 200240, P. R. China

Corresponding authors

*Correspondence to: Fangyou Yan (yanfangyou@tust.edu.cn), Yin-Ning Zhou (zhouyn@sjtu.edu.cn)

Supplementary Tables

Table S1. The property matrices (*MP*).

MP	**Note**	***MP***	**Note**
$MP_{\text{noe}} = \left[noe_j \right]$	*noe*: number of outermost electrons	$MP_{\text{rad}} = \left[rad_j \right]$	*rad*: atomic radius
$MP_{\text{ion}} = \left[ion_j \right]$	*ion*: ionization energy/10	$MP_{\text{bd}} = \left[bd_j \right]$	*bd*: branched degree
$MP_{\text{wei}} = \left[wei_j \right]$	*wei*: atom weight	$MP_{\text{aro}} = \left[aro_j \right]$	*aro*: aromaticity
$MP_{\text{fwei}} = \left[\frac{(wei_j)}{\sum (wei_j)} \right]$			

Table S2. Statistical parameters.

Parameters	Definition
Squared correlation coefficient (R^2)	$R^2 = \frac{\sum_{i=1}^{n}\left((y_{i,\mathrm{exp}} - \bar{y}_{\mathrm{exp}}) \times (y_{i,\mathrm{cal}} - \bar{y}_{\mathrm{cal}})\right)^2}{\sum_{i=1}^{n}(y_{i,\mathrm{exp}} - \bar{y}_{\mathrm{exp}})^2 \times \sum_{i=1}^{n}(y_{i,\mathrm{cal}} - \bar{y}_{\mathrm{cal}})^2}$
Mean absolute error (AAE)	$\mathrm{AAE} = \frac{\sum_{i=1}^{n}\left\lvert y_{i,\mathrm{exp}} - y_{i,cal}\right\rvert}{n}$
Root mean squared error (RMSE)	$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}(y_{i,\mathrm{exp}} - y_{i,cal})^2}{n}}$
Squared correlation coefficient for leave-one-out cross validation(Q^2_{LOO})	$Q^2_{\mathrm{LOO}} = \frac{\sum_{i=1}^{n_{\mathrm{train}}}\left((y_{i,\mathrm{exp}} - \bar{y}_{\mathrm{exp}}) \times (y_{i,\mathrm{cal}} - \bar{y}_{\mathrm{cal}})\right)^2}{\sum_{i=1}^{n_{\mathrm{train}}}(y_{i,\mathrm{exp}} - \bar{y}_{\mathrm{exp}})^2 \times \sum_{i=1}^{n}(y_{i,\mathrm{cal}} - \bar{y}_{\mathrm{cal}})^2}$
Leverage value (h)	$h = x_i(X^T X)^{-1} x_i^T$
Leverage value threshold (h^*)	$h^* = \frac{3(n_{\mathrm{train}} + 1)}{n_{\mathrm{cal.}}}$

Note: x_i is the descriptor row-vector of the i-th compound, x_i^T is the transpose of x_i, X is the descriptor matrix, and X^T is the transpose of X. n, n_{train}, and $n_{\mathrm{cal.}}$ are the numbers of data points, data points in the training set, and objects used to calculate the model, respectively.

Table S3. Norm descriptors (I) and coefficients (b) for the ε-QSPR model.

i	H-suppressed	*I*	*b*
1	*no*	$\left\|MS_{\mathrm{AB}}\cdot\times MP_{\mathrm{noe}}^{\mathrm{T}}\right\|_3$	0.3936
2	*yes*	$\left\|MS_{\mathrm{AB}}\cdot\times MP_{\mathrm{ion}}^{\mathrm{T}}\right\|_3$	6.2901
3	*yes*	$\left\|MS_{\mathrm{AB}}\cdot\times MP_{\mathrm{noe}}^{\mathrm{T}}\right\|_3$	-0.6117
4	*yes*	$\left\|MS_{\mathrm{ABC}}\cdot\times MP_{\mathrm{fwei}}^{\mathrm{T}}\right\|_5$	1.3301
5	*yes*	$\left\|MS_{\mathrm{bond}}\cdot\times\left\|MP_{\mathrm{fwei}}-MP_{\mathrm{fwei}}^{\mathrm{T}}\right\|\right\|_1$	237.6102
6	*no*	$\left\|MS_{\mathrm{ABCaro}}\times\left\|MP_{\mathrm{fwei}}-MP_{\mathrm{fwei}}^{\mathrm{T}}\right\|\right\|_6$	-4.6332
7	*no*	$\left\|MS_{\mathrm{bondA}}\times\left\|MP_{\mathrm{aro}}-MP_{\mathrm{aro}}^{\mathrm{T}}\right\|\right\|_6$	1.7601
8	*yes*	$\left\|MS_{\mathrm{ABC}}\cdot\times MP_{\mathrm{wei}}^{\mathrm{T}}\right\|_3$	-1.0878
9	*yes*	$\left\|MS_{\mathrm{AB}}\cdot\times MP_{\mathrm{wei}}^{\mathrm{T}}\right\|_3$	1.2742
10	*no*	$\left\|MS_{\mathrm{A}}\times\left\|MP_{\mathrm{aro}}-MP_{\mathrm{aro}}^{\mathrm{T}}\right\|\right\|_6$	-0.3697
11	*yes*	$\left\|MS_{\mathrm{bond}}\cdot\times\left\|MP_{\mathrm{ion}}-MP_{\mathrm{ion}}^{\mathrm{T}}\right\|\right\|_1$	265.5382
12	*no*	$\left\|MS_{\mathrm{AB}}\cdot\times\left\|MP_{\mathrm{ion}}-MP_{\mathrm{ion}}^{\mathrm{T}}\right\|\right\|_2$	-2.2814
13	*yes*	$\left\|MS_{\mathrm{ABC}}\cdot\times\left\|MP_{\mathrm{bd}}-MP_{\mathrm{bd}}^{\mathrm{T}}\right\|\right\|_4$	-5.1375
14	*no*	$\left\|MS_{\mathrm{bondA}}\cdot\times\left\|MP_{\mathrm{bd}}-MP_{\mathrm{bd}}^{\mathrm{T}}\right\|\right\|_5$	-11.3024
15	*no*	$\left\|MS_{\mathrm{bond}}\times\left\|MP_{\mathrm{aro}}-MP_{\mathrm{aro}}^{\mathrm{T}}\right\|\right\|_2$	-0.0248
16	*yes*	$\left\|MS_{\mathrm{F}}\cdot\times MP_{\mathrm{rad}}\right\|_3$	-0.3824
17	*no*	$\left\|MS_{\mathrm{A}}\cdot\times\left\|MP_{\mathrm{aro}}-MP_{\mathrm{aro}}^{\mathrm{T}}\right\|\right\|_1$	-23.8595
18	*yes*	$\left\|MS_{\mathrm{bond}}\cdot\times\left\|MP_{\mathrm{bd}}-MP_{\mathrm{bd}}^{\mathrm{T}}\right\|\right\|_1$	-17.0902
19	*yes*	$\left\|MS_{\mathrm{bond}}\cdot\times\left\|MP_{\mathrm{fwei}}-MP_{\mathrm{fwei}}^{\mathrm{T}}\right\|\right\|_5$	-35.5794

Table S4. Statistical results of the ε-QSPR model.

Method	Variable	Data points	Values
External validation	$R^2_{training}$	1013	0.9754
	$R^2_{testing}$	243	0.9821
	$AAE_{training}$	1013	0.1175
	$AAE_{testing}$	243	0.1138
	$RMSE_{training}$	1013	0.1652
	$RMSE_{testing}$	243	0.1653
Overall data set	R^2	1256	0.9768
	AAE	1256	0.1168
	RMSE	1256	0.1652
LOPO-CV	Q^2_{LOO}	1013	0.9654
	AAE_{LOO}	1013	0.1424
	$RMSE_{LOO}$	1013	0.1960
Y-random validation	$\overline{R_Y^2}$	10000(times)	0.0151
	$\overline{Q_Y^2}$	10000(times)	0.0331

Supplementary Figures

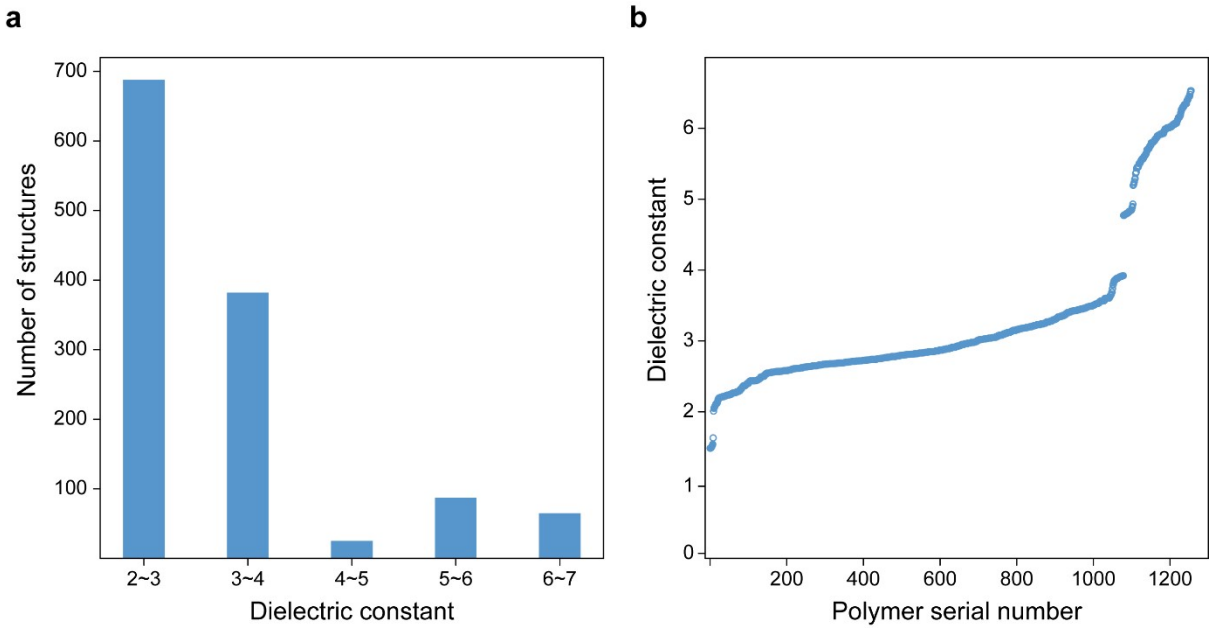

Fig. S1. Summary of the dataset used in this work. a Distribution of ε dataset; **b** The applied ε values are plotted in ascending order.

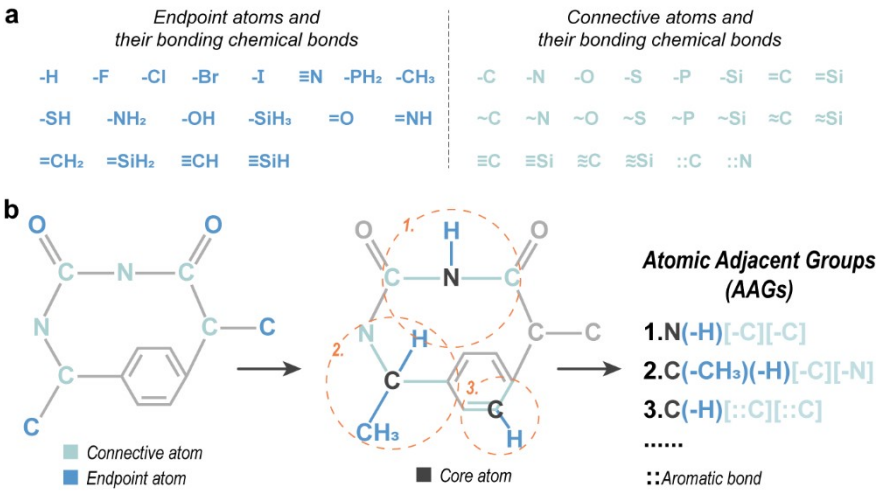

Fig. S2. a Examples of endpoint and connective atoms and their binding chemical bond types in the atomic adjacent group (AAG) representation rule; **b** the core atom, connective atoms and bond type in AAGs and AAG examples.

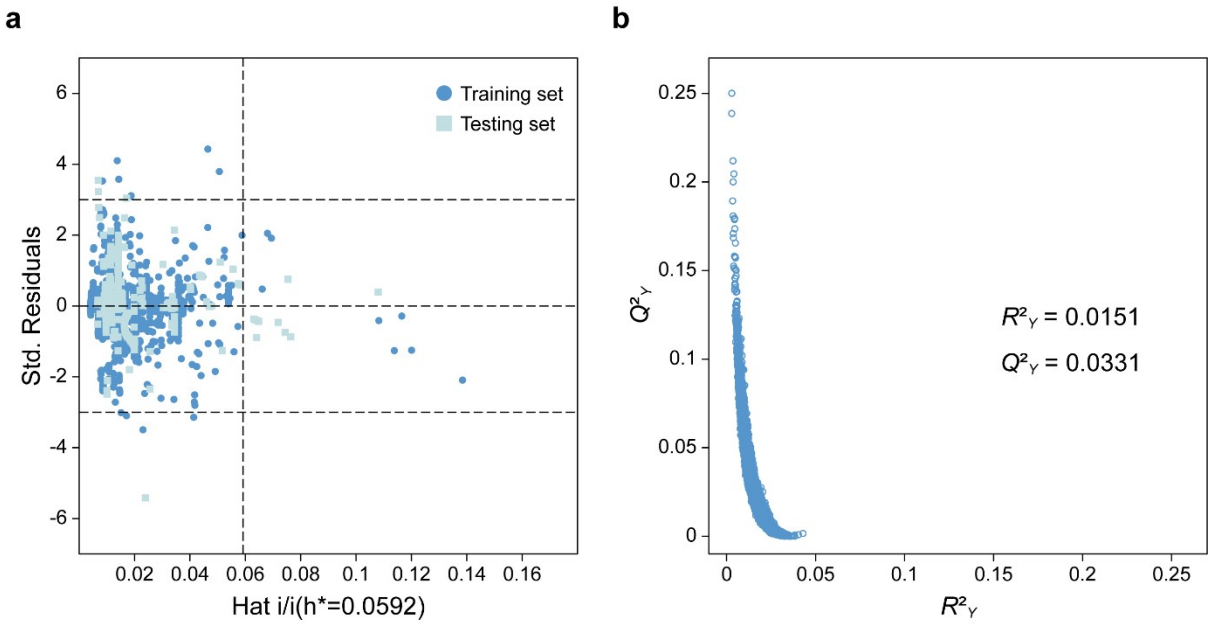

Fig. S3. a William plot and **b** R_Y^2 *vs.* Q_Y^2 of the 10000 *Y*-randomization tests for PI-ε model.

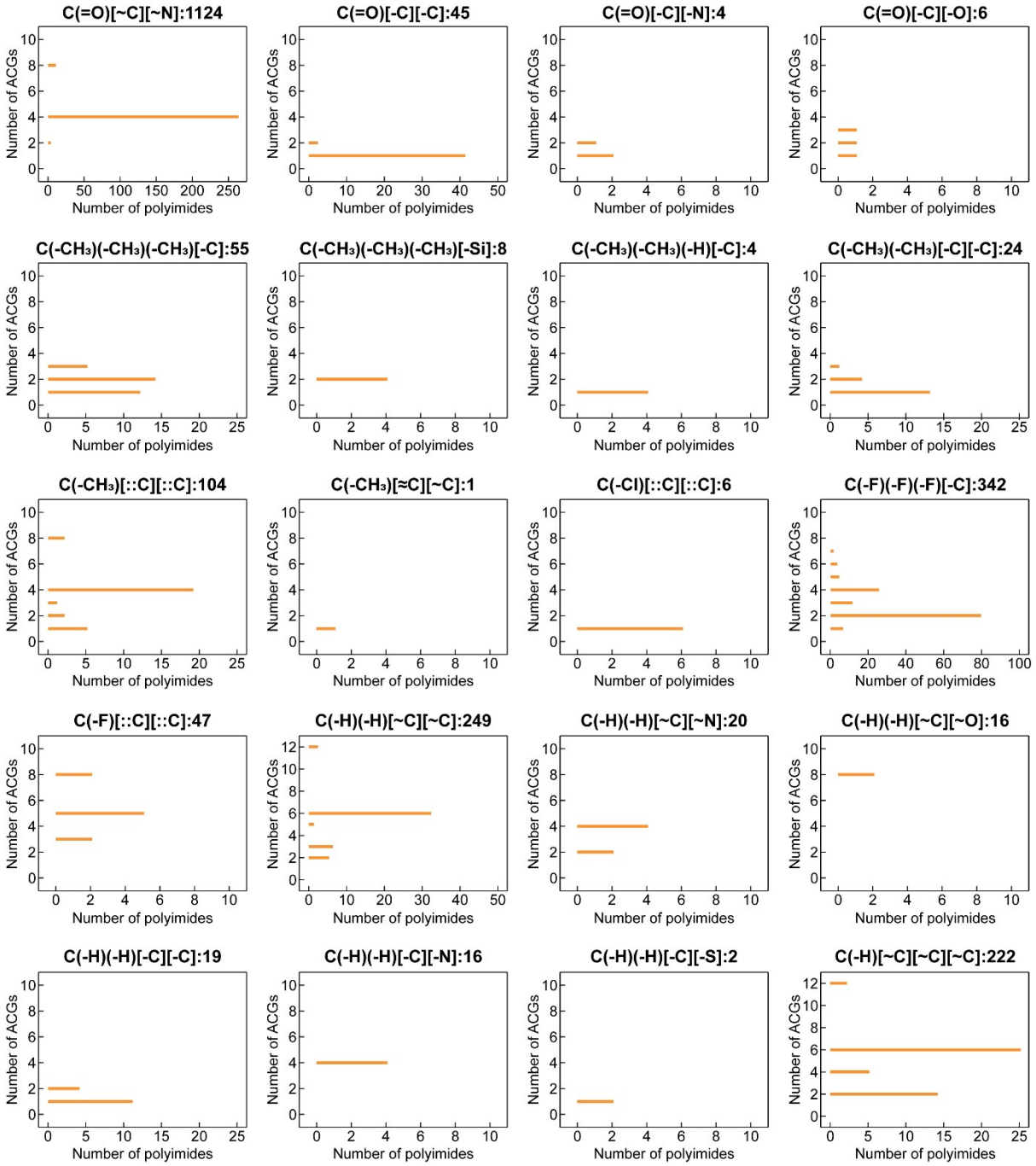

Fig. S4. Analysis of the 20 atomic adjacent group (AAG) of polyimides (PIs) in the modelled dataset with carbon (C) atom as the core atom.

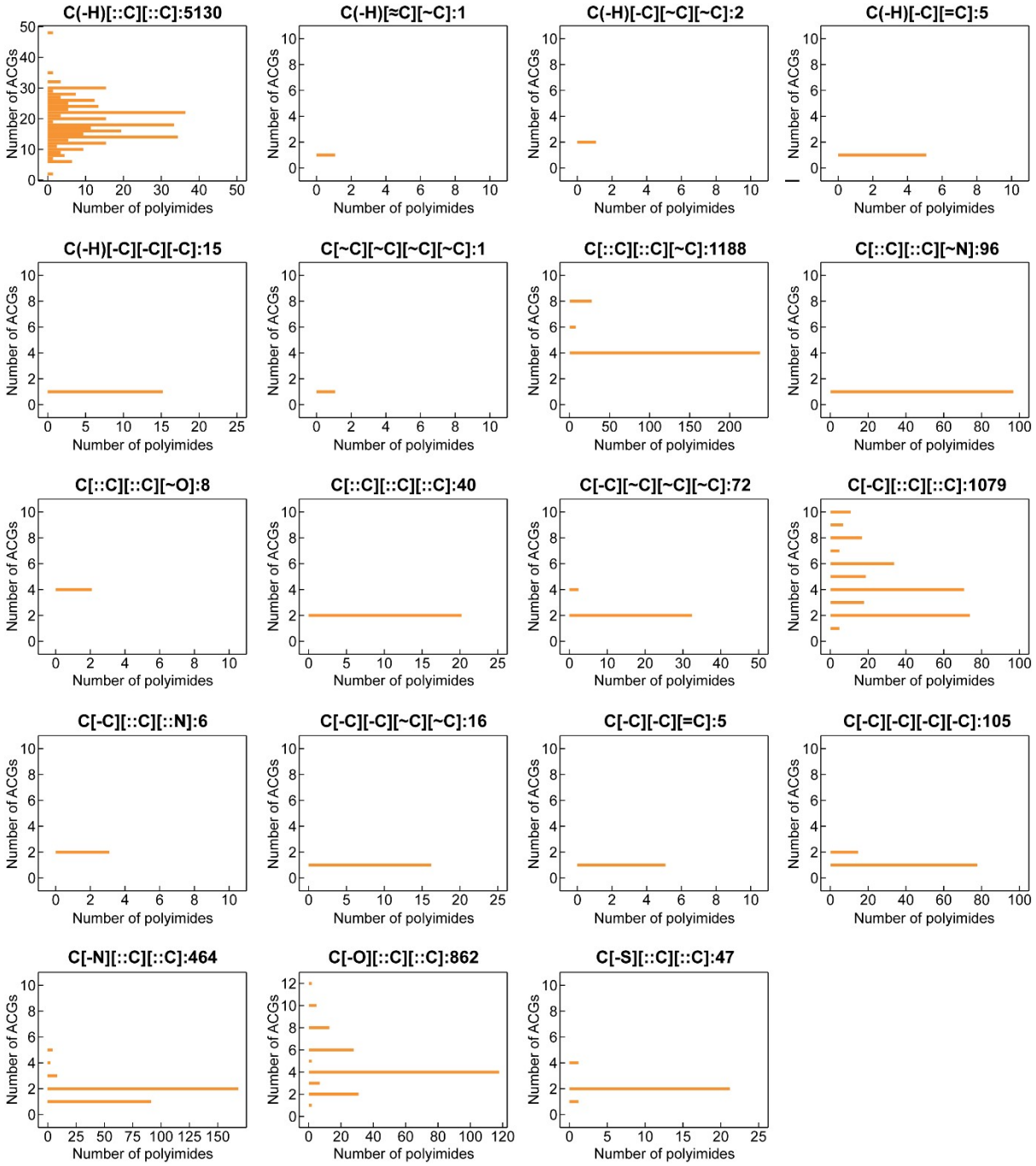

Fig. S5. Analysis of the 19 atomic adjacent group (AAG) of polyimides (PIs) in the modelled dataset with carbon (C) atom as the core atom.

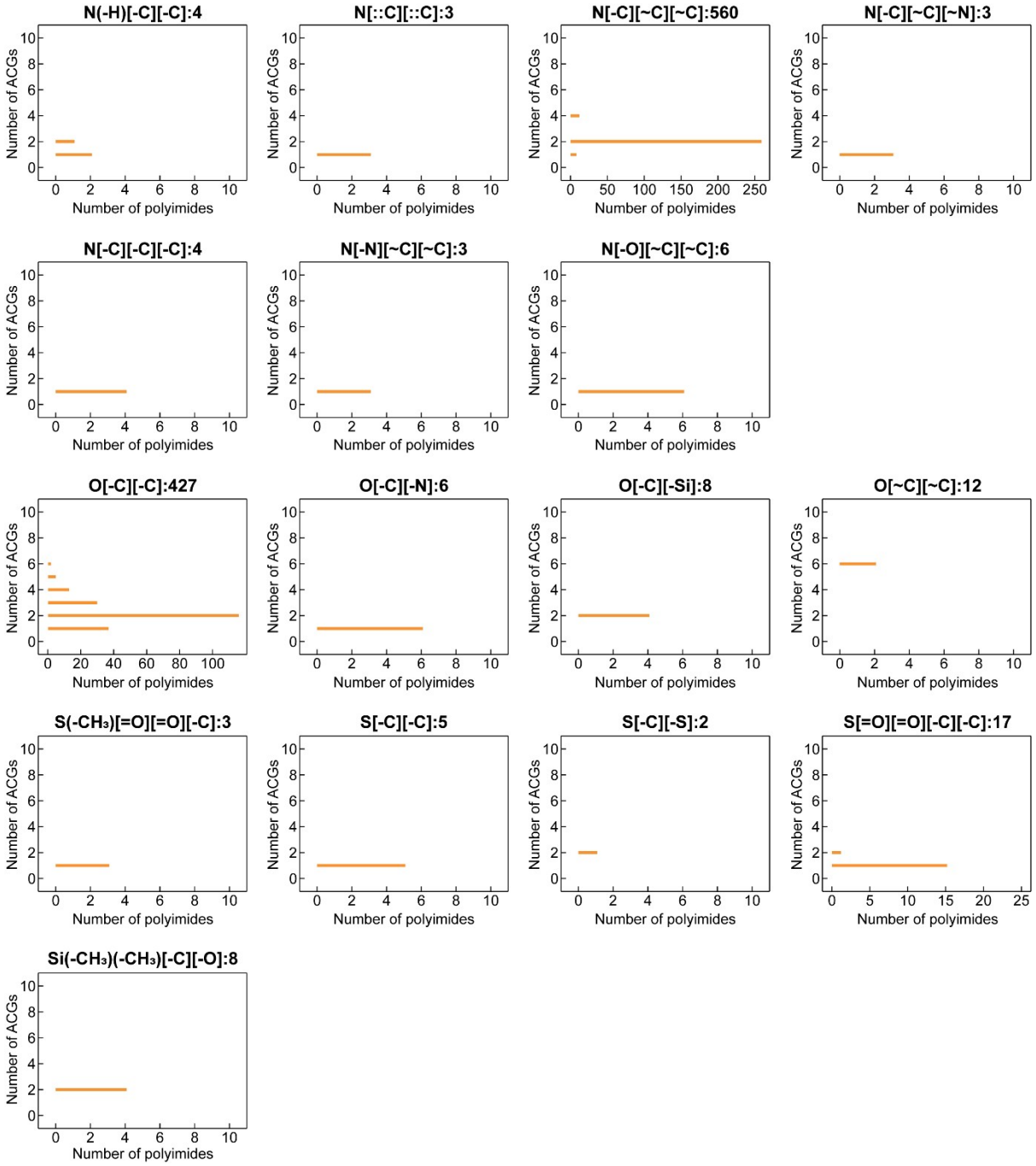

Fig. S6. Analysis of the 16 atomic adjacent group (AAG) of polyimides (PIs) in the modelled dataset with nitrogen (N), oxygen (O), sulfur (S) or silicon (Si) Atoms as the core atom.

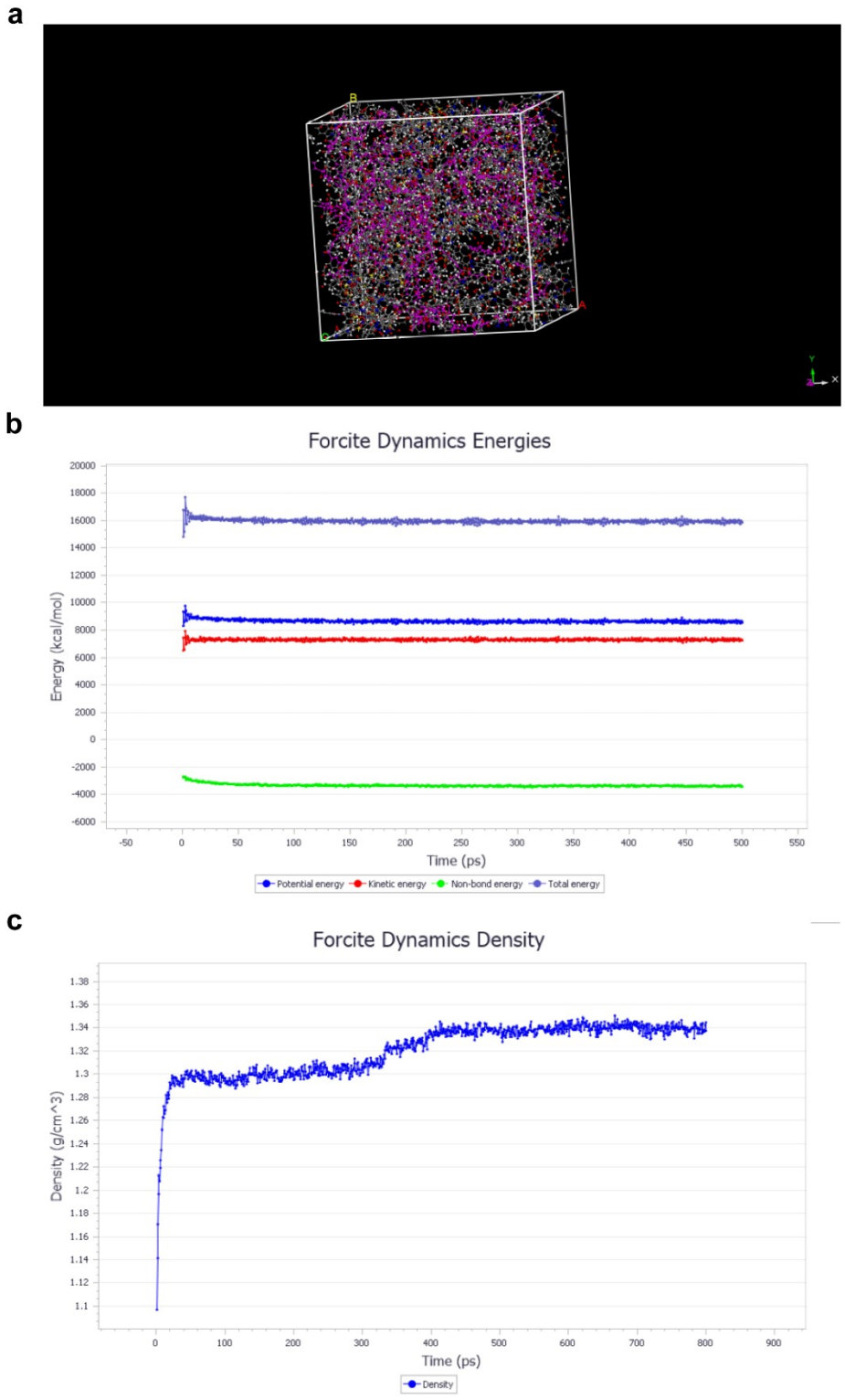

Fig. S7. All-atom molecular dynamics simulation results of PI_CD24351. a MD amorphous cell; **b** energy fluctuation in the 500 ps NVT dynamic equilibrium process; **c** density fluctuation in the 800 ps NPT.

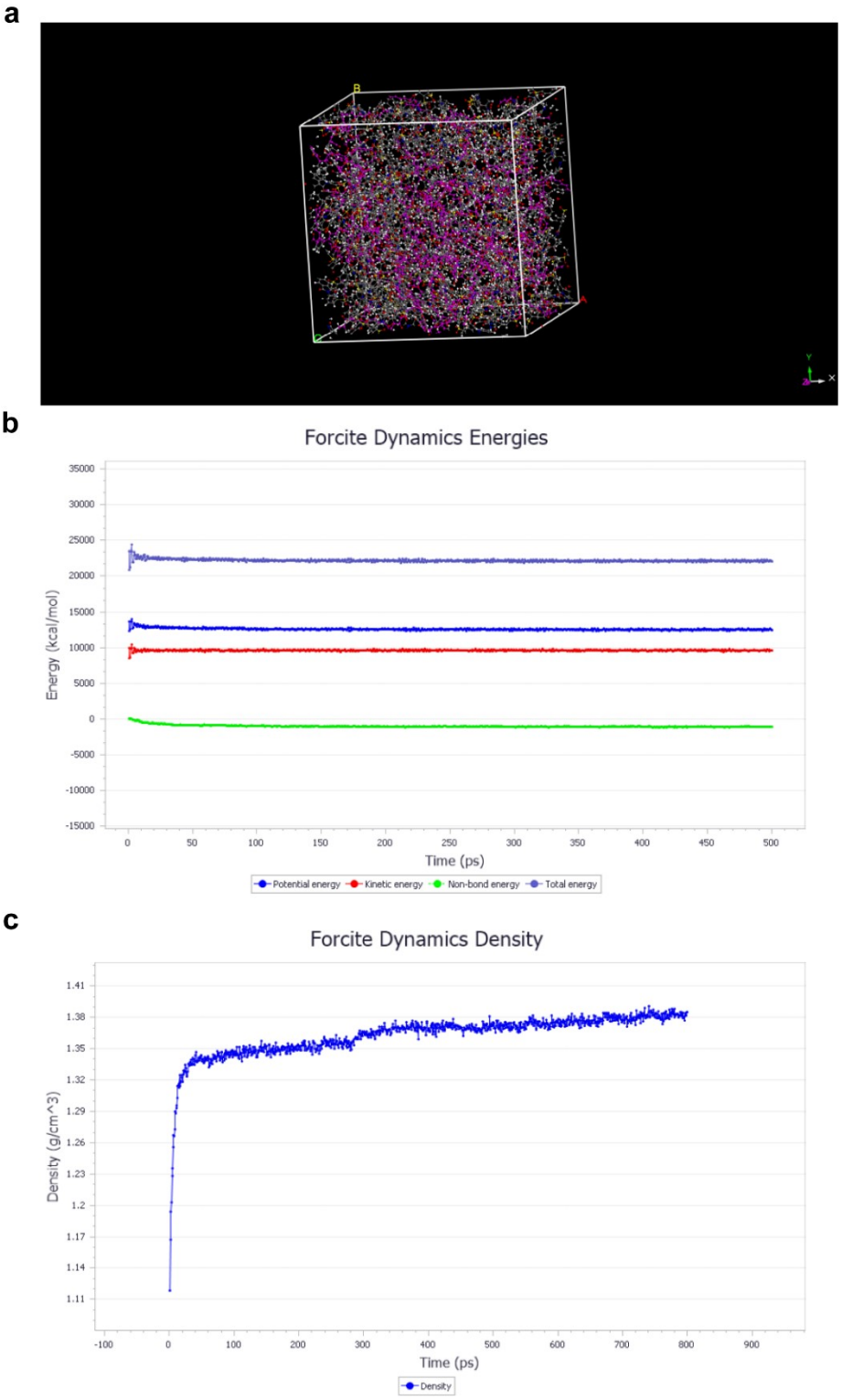

Fig. S8. All-atom molecular dynamics simulation results of PI_CD24416. a MD amorphous cell; **b** energy fluctuation in the 500 ps NVT dynamic equilibrium process; **c** density fluctuation in the 800 ps NPT.

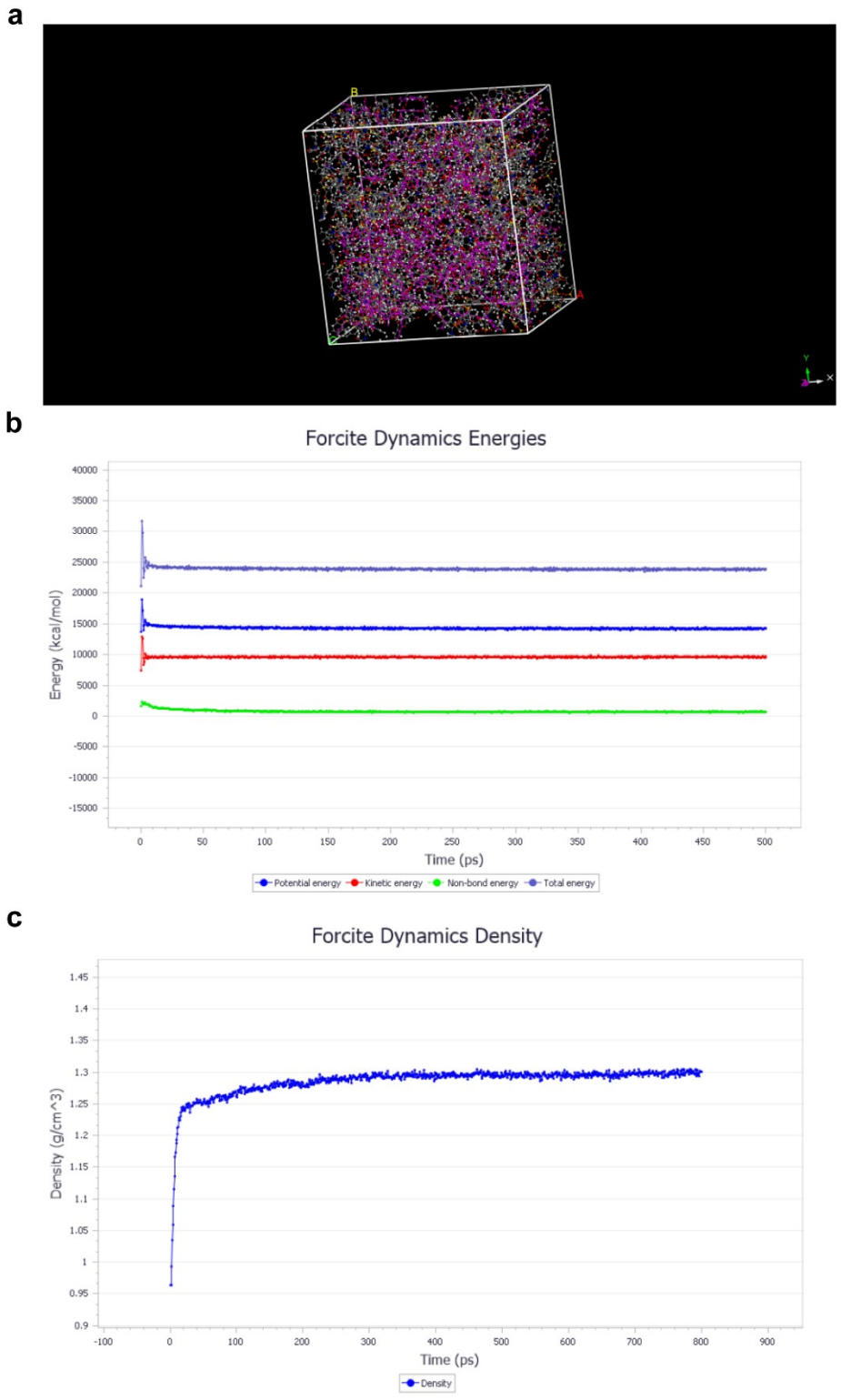

Fig. S9. All-atom molecular dynamics simulation results of PI_CD24366. a MD amorphous cell; **b** energy fluctuation in the 500 ps NVT dynamic equilibrium process; **c** density fluctuation in the 800 ps NPT.

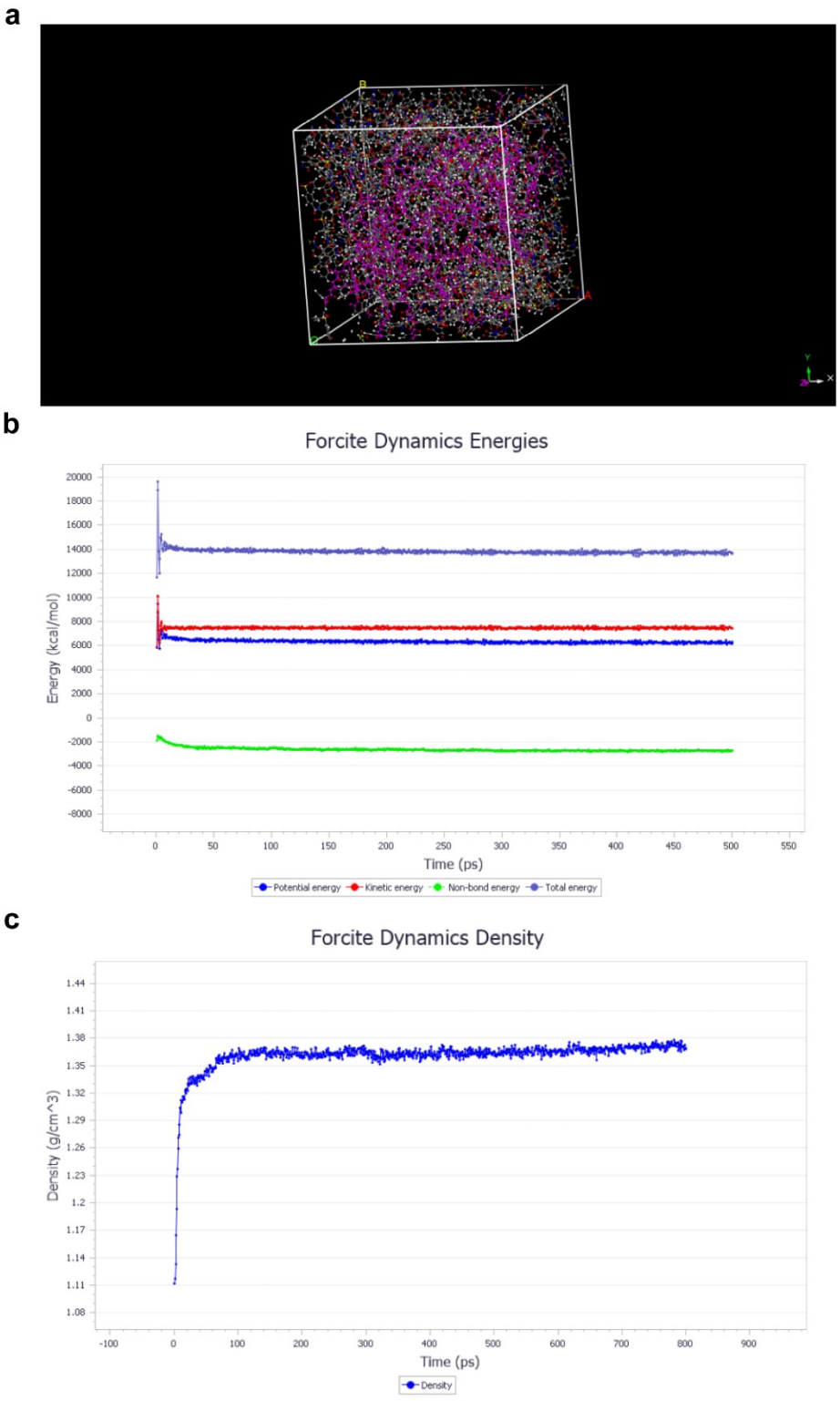

Fig. S10. All-atom molecular dynamics simulation results of PI_CD24719. a MD amorphous cell; **b** energy fluctuation in the 500 ps NVT dynamic equilibrium process; **c** density fluctuation in the 800 ps NPT.

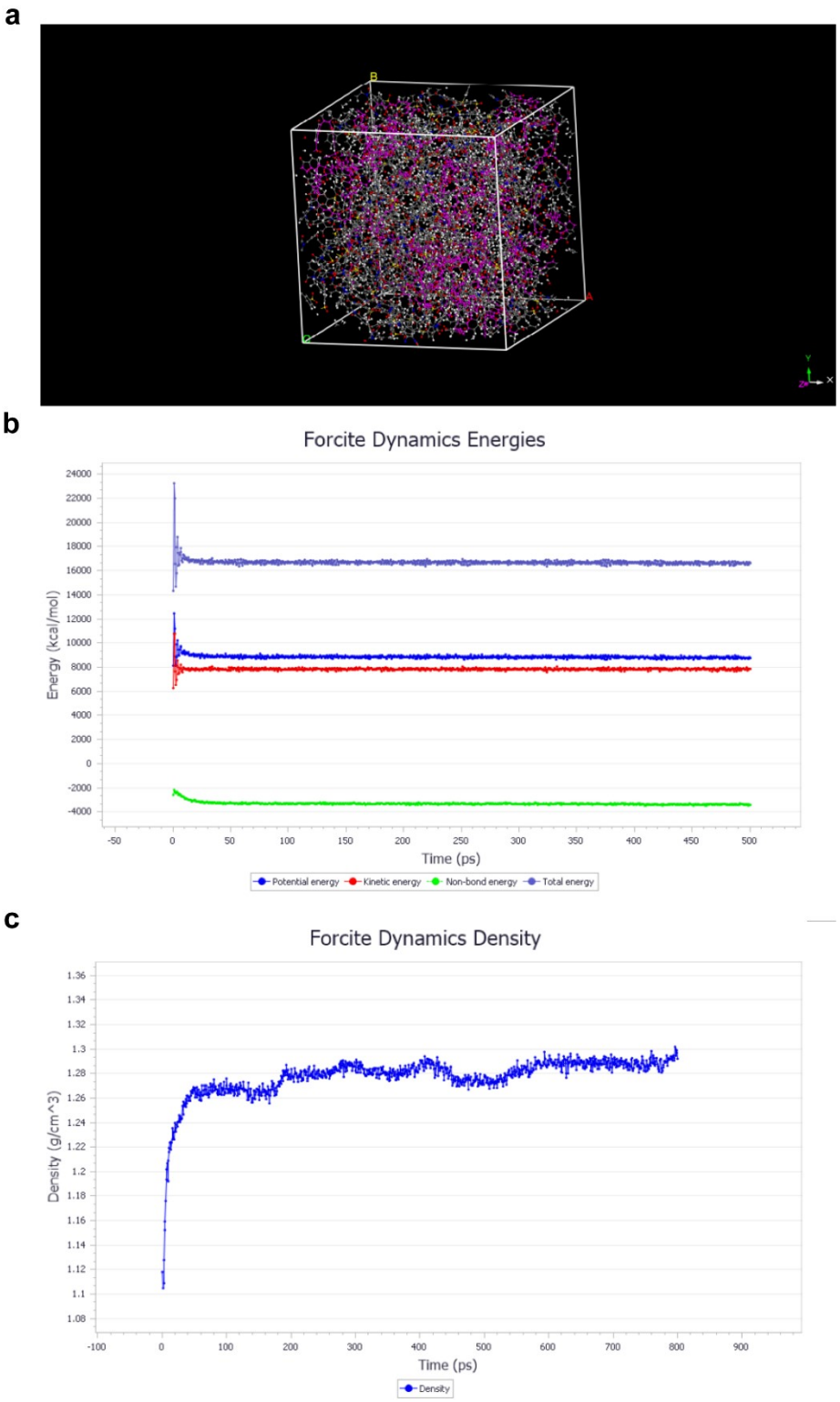

Fig. S11. All-atom molecular dynamics simulation results of PI_CD24529. a MD amorphous cell; **b** energy fluctuation in the 500 ps NVT dynamic equilibrium process; **c** density fluctuation in the 800 ps NPT.

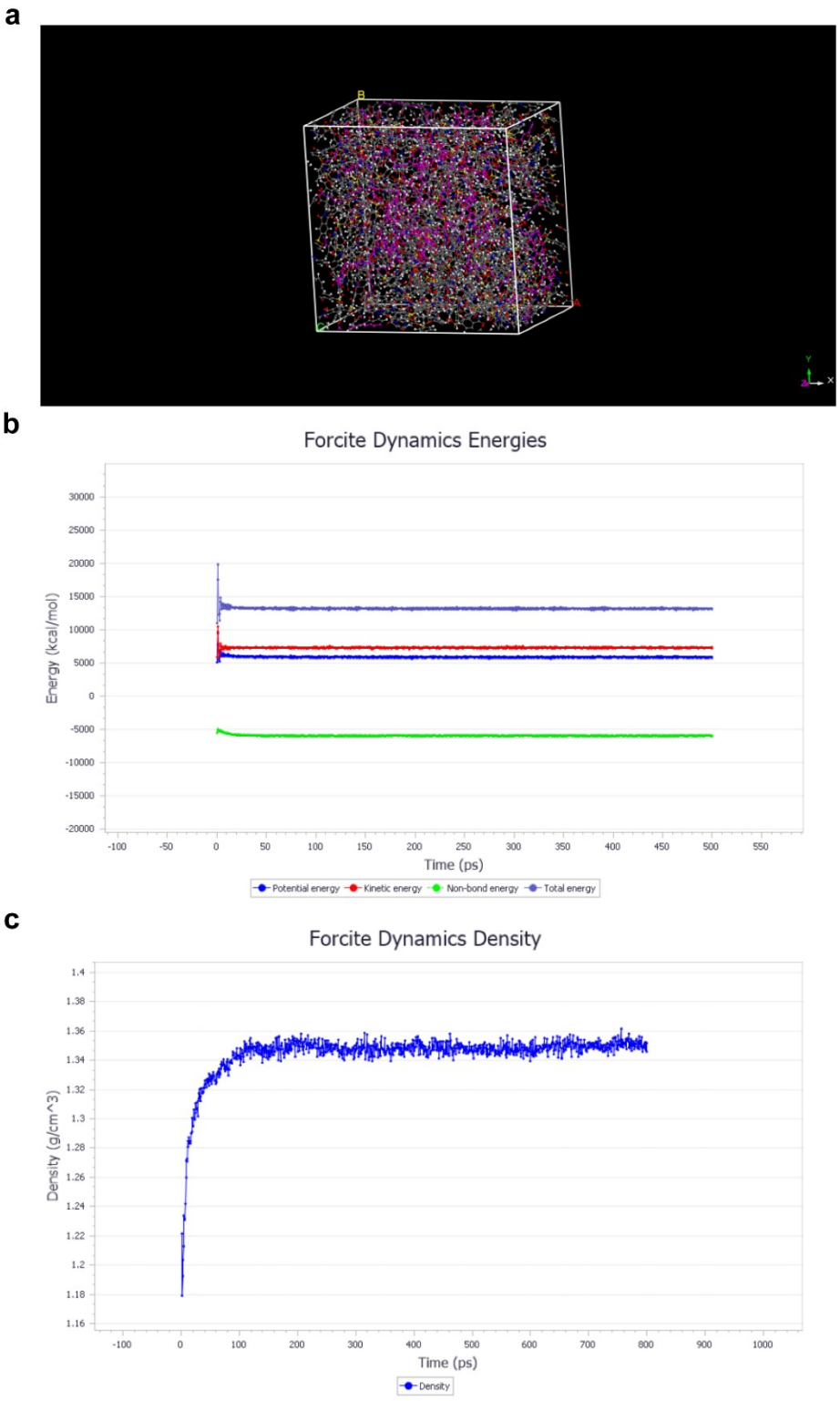

Fig. S12. All-atom molecular dynamics simulation results of PI_CD24352. a MD amorphous cell; **b** energy fluctuation in the 500 ps NVT dynamic equilibrium process; **c** density fluctuation in the 800 ps NPT.

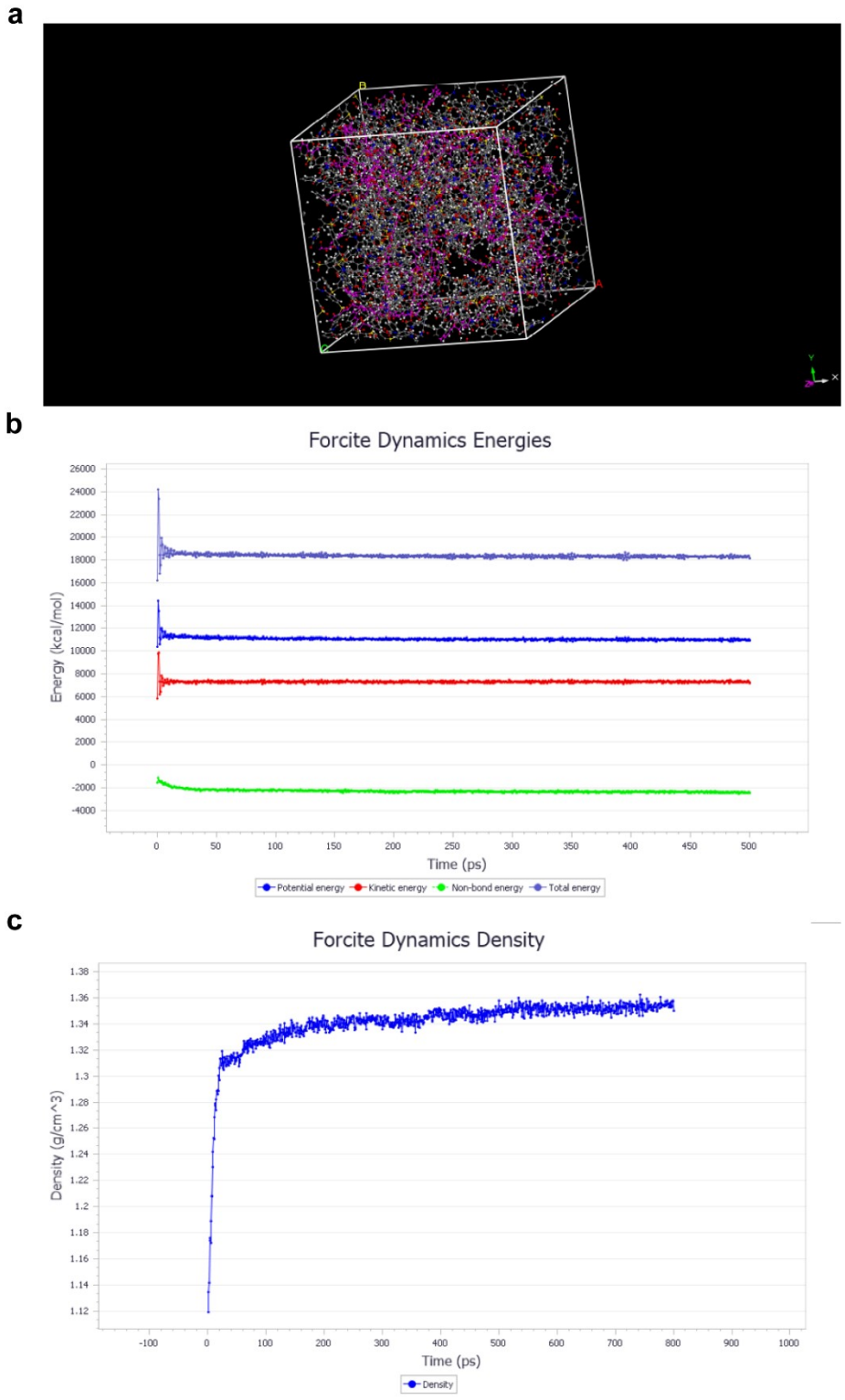

Fig. S13. All-atom molecular dynamics simulation results of PI_CD24603. a MD amorphous cell; **b** energy fluctuation in the 500 ps NVT dynamic equilibrium process; **c** density fluctuation in the 800 ps NPT.

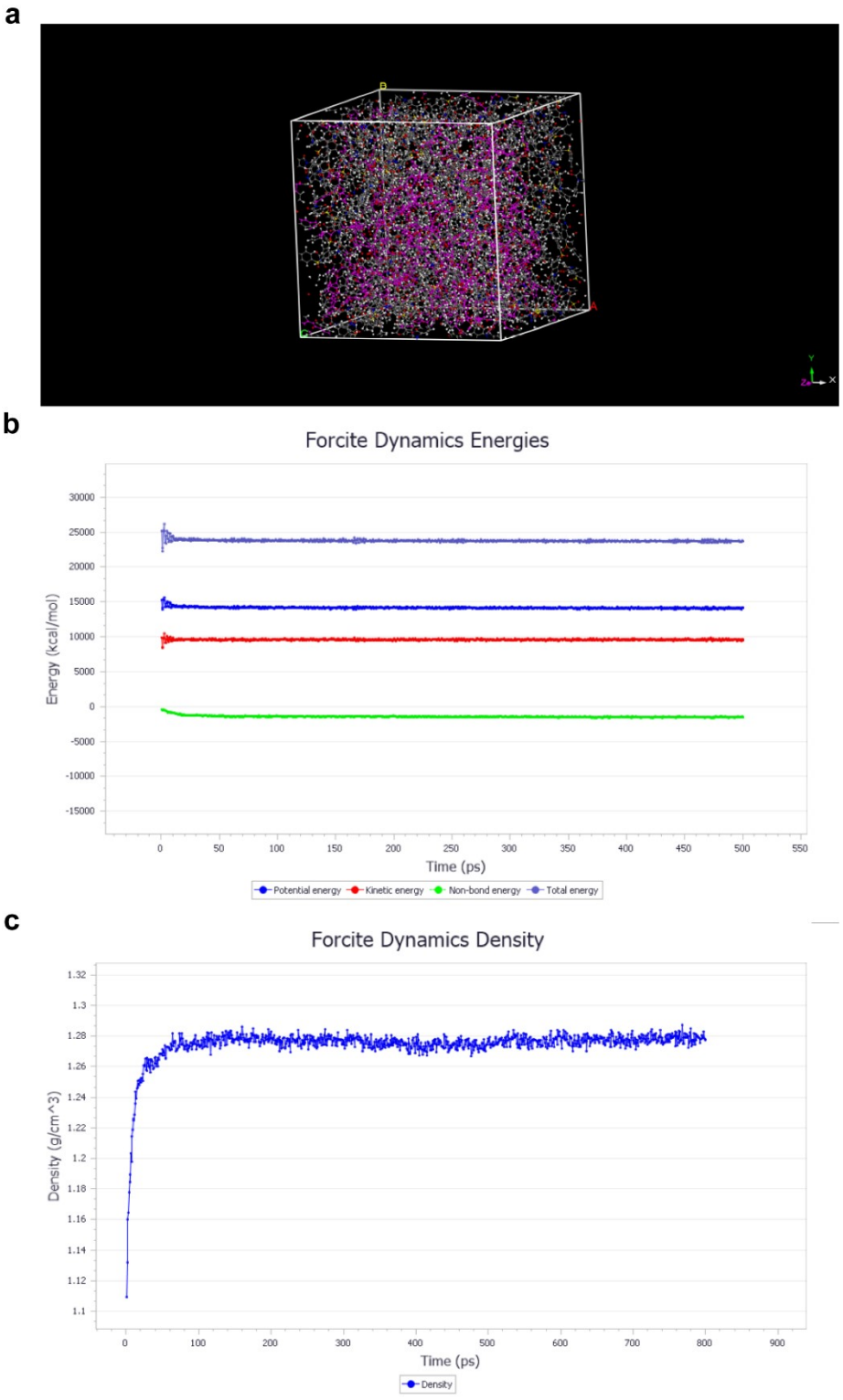

Fig. S14. All-atom molecular dynamics simulation results of PI_CD24495. a MD amorphous cell; **b** energy fluctuation in the 500 ps NVT dynamic equilibrium process; **c** density fluctuation in the 800 ps NPT.

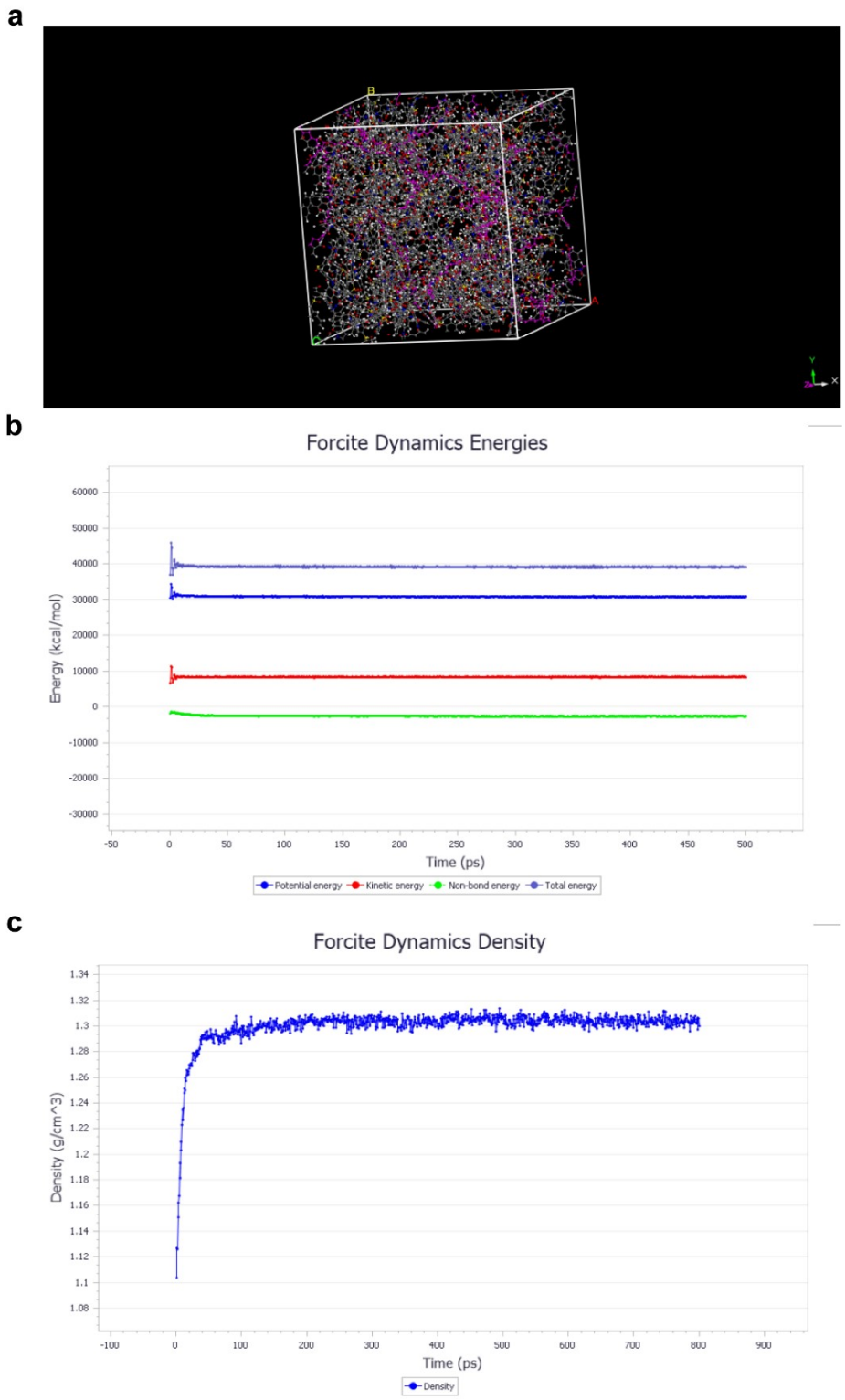

Fig. S15. All-atom molecular dynamics simulation results of PI_CD24378. a MD amorphous cell; **b** energy fluctuation in the 500 ps NVT dynamic equilibrium process; **c** density fluctuation in the 800 ps NPT.

Supplementary Texts

Gaussian input coordinates and connectivity of PI_CD24351:

0 1

C	-3.27580000	-1.01500000	-2.51380000
C	-3.17920000	0.34730000	-2.81810000
C	-1.96620000	0.98300000	-3.02800000
C	-0.84400000	0.14890000	-2.91050000
C	-0.92400000	-1.22700000	-2.60620000
C	-2.17110000	-1.84600000	-2.40070000
C	-4.61120000	-1.21200000	-2.39390000
N	-5.47970000	-0.15180000	-2.57550000
C	-4.47020000	0.75590000	-2.83450000
O	-5.02610000	-2.31920000	-2.13080000
O	-4.71740000	1.91600000	-3.08050000
C	-6.89600000	-0.04670000	-2.52810000
S	0.53190000	-2.15800000	-2.50650000
C	1.45700000	-1.39830000	-1.25430000
C	2.86320000	-1.34490000	-1.35990000
C	3.69110000	-0.77960000	-0.37820000
C	3.01900000	-0.28110000	0.72620000
C	1.62680000	-0.33510000	0.85050000
C	0.80220000	-0.89110000	-0.11680000
C	3.41660000	0.32170000	1.87220000
N	2.46560000	0.70460000	2.81850000
C	1.39410000	0.24160000	2.05490000
O	0.24270000	0.36420000	2.41510000
O	4.60250000	0.52110000	2.03370000
C	2.56490000	1.29910000	4.02220000
O	0.18970000	-3.43520000	-1.95710000
O	1.23100000	-1.89340000	-3.72820000
C	3.48970000	0.79230000	4.94170000
C	3.64500000	1.38800000	6.19460000
C	2.87790000	2.50230000	6.53550000
C	1.94720000	3.00780000	5.62660000
C	1.78310000	2.40870000	4.37410000
N	0.85540000	2.94520000	3.55900000
H	-1.88840000	2.05590000	-3.26990000
H	0.14230000	0.61650000	-3.07080000
H	-2.29200000	-2.91670000	-2.16900000
H	-7.28540000	-0.54430000	-1.61100000
H	-7.21450000	1.02020000	-2.51290000
H	-7.33780000	-0.53830000	-3.42410000

H	3.36620000	-1.77370000	-2.24300000
H	4.78930000	-0.74680000	-0.47330000
H	-0.29070000	-0.93280000	0.02250000
H	4.10910000	-0.08500000	4.68970000
H	4.37650000	0.98210000	6.91380000
H	3.00130000	2.97730000	7.52340000
H	1.33820000	3.88290000	5.91090000
H	0.69730000	2.46360000	2.67770000
H	0.36290000	3.74270000	3.95600000

1 2 1.5 6 1.5 7 1.0
2 3 1.5 9 1.0
3 4 1.5 34 1.0
4 5 1.5 35 1.0
5 6 1.5 13 1.0
6 36 1.0
7 8 1.0 10 2.0
8 9 1.0 12 1.0
9 11 2.0
10
11
12 37 1.0 38 1.0 39 1.0
13 14 1.0 26 2.0 27 2.0
14 15 1.5 19 1.5
15 16 1.5 40 1.0
16 17 1.5 41 1.0
17 18 1.5 20 1.0
18 19 1.5 22 1.0
19 42 1.0
20 21 1.0 24 2.0
21 22 1.0 25 1.0
22 23 2.0
23
24
25 28 1.5 32 1.5
26
27
28 29 1.5 43 1.0
29 30 1.5 44 1.0
30 31 1.5 45 1.0
31 32 1.5 46 1.0
32 33 1.0
33 47 1.0 48 1.0
34

35
36
37
38
39
40
41
42
43
44
45
46
47
48

Gaussian input coordinates and connectivity of PI_CD24416:

0 1

C	-2.32480000	-2.48530000	-4.07830000
C	-2.27850000	-1.25290000	-4.73920000
C	-1.10970000	-0.73160000	-5.26940000
C	0.02490000	-1.53380000	-5.07630000
C	-0.00390000	-2.77980000	-4.41330000
C	-1.20940000	-3.28870000	-3.89450000
C	-3.62750000	-2.59720000	-3.72220000
N	-4.51550000	-1.58760000	-4.04430000
C	-3.55930000	-0.81760000	-4.67870000
O	-3.99310000	-3.58140000	-3.11810000
O	-3.84360000	0.24940000	-5.17660000
C	-5.90410000	-1.40620000	-3.80410000
S	1.48080000	-3.64000000	-4.16340000
C	2.14950000	-2.72700000	-2.84910000
C	3.36460000	-2.02490000	-2.99750000
C	3.92610000	-1.23110000	-1.98490000
C	3.19380000	-1.18590000	-0.80930000
C	1.99230000	-1.88450000	-0.64530000
C	1.43200000	-2.67060000	-1.63980000
C	3.35430000	-0.55220000	0.37760000
N	2.40280000	-0.71570000	1.38200000
C	1.61140000	-1.56740000	0.61540000
O	0.55930000	-2.00230000	1.03440000
O	4.31580000	0.17160000	0.52970000
C	2.17920000	-0.08990000	2.55090000
O	1.14690000	-4.91250000	-3.59910000
O	2.28930000	-3.38080000	-5.31560000

C	2.18350000	-0.81980000	3.73960000
C	1.92410000	-0.16370000	4.94330000
C	1.65040000	1.20570000	4.95190000
C	1.64730000	1.94660000	3.76330000
C	1.91480000	1.28120000	2.56040000
S	1.27230000	3.63590000	3.78900000
O	1.91670000	4.14990000	4.95990000
O	1.42550000	4.11770000	2.44900000
C	-0.43170000	3.68330000	4.07860000
C	-1.26910000	2.73130000	3.48120000
C	-2.65180000	2.76220000	3.67070000
C	-3.22550000	3.76550000	4.45050000
C	-2.40670000	4.72890000	5.03960000
C	-1.02110000	4.68740000	4.85740000
N	-0.30810000	5.65520000	5.46130000
H	-1.06900000	0.24430000	-5.78090000
H	0.98080000	-1.13310000	-5.45400000
H	-1.28650000	-4.24470000	-3.35140000
H	-6.19660000	-1.87920000	-2.83920000
H	-6.15140000	-0.32190000	-3.74740000
H	-6.49120000	-1.86930000	-4.62880000
H	3.91890000	-2.07090000	-3.94930000
H	4.87040000	-0.67790000	-2.11920000
H	0.47620000	-3.19540000	-1.47590000
H	2.37830000	-1.90560000	3.73210000
H	1.91590000	-0.73390000	5.88760000
H	1.42330000	1.68700000	5.91840000
H	1.89960000	1.81870000	1.59760000
H	-0.86000000	1.93710000	2.83490000
H	-3.29320000	2.00230000	3.19270000
H	-4.31880000	3.80120000	4.59310000
H	-2.86730000	5.52540000	5.64900000
H	0.70110000	5.62020000	5.34450000
H	-0.86670000	6.32030000	5.99250000

1 2 1.5 6 1.5 7 1.0
2 3 1.5 9 1.0
3 4 1.5 43 1.0
4 5 1.5 44 1.0
5 6 1.5 13 1.0
6 45 1.0
7 8 1.0 10 2.0
8 9 1.0 12 1.0
9 11 2.0

10
11
12 46 1.0 47 1.0 48 1.0
13 14 1.0 26 2.0 27 2.0
14 15 1.5 19 1.5
15 16 1.5 49 1.0
16 17 1.5 50 1.0
17 18 1.5 20 1.0
18 19 1.5 22 1.0
19 51 1.0
20 21 1.0 24 2.0
21 22 1.0 25 1.0
22 23 2.0
23
24
25 28 1.5 32 1.5
26
27
28 29 1.5 52 1.0
29 30 1.5 53 1.0
30 31 1.5 54 1.0
31 32 1.5 33 1.0
32 55 1.0
33 34 2.0 35 2.0 36 1.0
34
35
36 37 1.5 41 1.5
37 38 1.5 56 1.0
38 39 1.5 57 1.0
39 40 1.5 58 1.0
40 41 1.5 59 1.0
41 42 1.0
42 60 1.0 61 1.0
43
44
45
46
47
48
49
50
51
52
53

54
55
56
57
58
59
60
61

Gaussian input coordinates and connectivity of PI_CD24366:

0 1

C	-3.66840000	-2.39140000	-4.60180000
C	-3.51500000	-1.19500000	-5.31130000
C	-2.27790000	-0.70560000	-5.69700000
C	-1.19190000	-1.50740000	-5.31620000
C	-1.32980000	-2.71780000	-4.60350000
C	-2.59970000	-3.19130000	-4.22520000
C	-5.01020000	-2.48410000	-4.43710000
N	-5.83300000	-1.49400000	-4.94180000
C	-4.78710000	-0.75640000	-5.46300000
O	-5.46950000	-3.43860000	-3.84960000
O	-4.98640000	0.28070000	-6.05610000
C	-7.24230000	-1.31240000	-4.94080000
S	0.09300000	-3.61730000	-4.19610000
C	0.93880000	-2.51870000	-3.15660000
C	2.24920000	-2.09830000	-3.46830000
C	2.99900000	-1.23830000	-2.65130000
C	2.35230000	-0.82660000	-1.49700000
C	1.05080000	-1.22960000	-1.17760000
C	0.30310000	-2.07710000	-1.98120000
C	2.70550000	-0.03880000	-0.45310000
N	1.80560000	0.16950000	0.59060000
C	0.81510000	-0.62210000	0.01050000
O	-0.25010000	-0.81080000	0.55910000
O	3.82360000	0.43110000	-0.42900000
C	1.97720000	0.62870000	1.84330000
O	-0.31130000	-4.69000000	-3.33840000
O	0.84050000	-3.70810000	-5.41360000
C	1.84240000	-0.24700000	2.92180000
C	2.03620000	0.21250000	4.22530000
C	2.37380000	1.54860000	4.47160000
C	2.49800000	2.41910000	3.38090000
C	2.30650000	1.96520000	2.07340000
S	2.63930000	2.12080000	6.08280000

O	2.40860000	1.01780000	6.96660000
O	3.82890000	2.91450000	6.00420000
C	1.32780000	3.20560000	6.38160000
C	0.05360000	2.92270000	5.87240000
C	-1.02600000	3.77320000	6.11680000
C	-0.85020000	4.91920000	6.89180000
C	0.40810000	5.20190000	7.42330000
C	1.48830000	4.35130000	7.17140000
N	2.66610000	4.69420000	7.72360000
H	-2.15420000	0.23720000	-6.25510000
H	-0.18690000	-1.15040000	-5.59940000
H	-2.76260000	-4.12790000	-3.66710000
H	-7.68910000	-1.74670000	-4.01770000
H	-7.49760000	-0.22890000	-4.97620000
H	-7.68380000	-1.81700000	-5.82940000
H	2.74360000	-2.45660000	-4.38640000
H	4.02710000	-0.92760000	-2.90070000
H	-0.71520000	-2.38720000	-1.69240000
H	1.59330000	-1.30840000	2.75210000
H	1.92440000	-0.51220000	5.04870000
H	2.75780000	3.48040000	3.53470000
H	2.42180000	2.66360000	1.22710000
H	-0.13200000	2.01470000	5.27440000
H	-2.02160000	3.53430000	5.70670000
H	-1.70260000	5.59040000	7.09270000
H	0.53970000	6.10300000	8.04610000
H	3.45560000	4.08370000	7.53030000
H	2.63970000	5.54890000	8.27630000

1 2 1.5 6 1.5 7 1.0
2 3 1.5 9 1.0
3 4 1.5 43 1.0
4 5 1.5 44 1.0
5 6 1.5 13 1.0
6 45 1.0
7 8 1.0 10 2.0
8 9 1.0 12 1.0
9 11 2.0
10
11
12 46 1.0 47 1.0 48 1.0
13 14 1.0 26 2.0 27 2.0
14 15 1.5 19 1.5
15 16 1.5 49 1.0

16 17 1.5 50 1.0
17 18 1.5 20 1.0
18 19 1.5 22 1.0
19 51 1.0
20 21 1.0 24 2.0
21 22 1.0 25 1.0
22 23 2.0
23
24
25 28 1.5 32 1.5
26
27
28 29 1.5 52 1.0
29 30 1.5 53 1.0
30 31 1.5 33 1.0
31 32 1.5 54 1.0
32 55 1.0
33 34 2.0 35 2.0 36 1.0
34
35
36 37 1.5 41 1.5
37 38 1.5 56 1.0
38 39 1.5 57 1.0
39 40 1.5 58 1.0
40 41 1.5 59 1.0
41 42 1.0
42 60 1.0 61 1.0
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59

60
61

Gaussian input coordinates and connectivity of PI_CD24719:

0 1

C	-3.53300000	-1.16960000	-2.61800000
C	-3.38460000	0.21740000	-2.72630000
C	-2.14700000	0.83400000	-2.81240000
C	-1.05500000	-0.04600000	-2.77900000
C	-1.18770000	-1.44580000	-2.65860000
C	-2.45810000	-2.04580000	-2.58370000
C	-4.87710000	-1.33540000	-2.56790000
N	-5.70950000	-0.22730000	-2.62540000
C	-4.66010000	0.67190000	-2.72240000
O	-5.32500000	-2.45780000	-2.47200000
O	-4.86970000	1.86280000	-2.80090000
O	-6.98060000	-0.07560000	-2.58030000
S	0.23750000	-2.42930000	-2.61250000
C	1.19820000	-1.67140000	-1.38630000
C	2.56310000	-1.39870000	-1.61610000
C	3.39360000	-0.77140000	-0.67460000
C	2.76940000	-0.44010000	0.51740000
C	1.41950000	-0.71640000	0.76360000
C	0.59350000	-1.33820000	-0.16030000
C	3.18150000	0.16310000	1.65820000
N	2.27960000	0.35640000	2.70190000
C	1.22300000	-0.23860000	2.01600000
O	0.11540000	-0.32610000	2.50270000
O	4.33060000	0.54360000	1.73600000
C	2.35560000	1.01400000	3.87140000
O	-0.13820000	-3.68990000	-2.04720000
O	0.89250000	-2.18570000	-3.86190000
C	3.03770000	0.42480000	4.93730000
C	3.12040000	1.09570000	6.15710000
C	2.51500000	2.34490000	6.29280000
C	1.83760000	2.91860000	5.21480000
C	1.74490000	2.27030000	3.97590000
O	1.09200000	2.79640000	2.90000000
N	0.55190000	3.97290000	2.88090000
H	-2.02900000	1.92640000	-2.90210000
H	-0.05100000	0.40590000	-2.84850000
H	-2.61450000	-3.13370000	-2.50190000
H	-7.35980000	0.78490000	-2.62410000
H	3.02880000	-1.67880000	-2.57560000

H	4.45710000	-0.55720000	-0.87040000
H	-0.46670000	-1.53990000	0.06690000
H	3.50550000	-0.56770000	4.82230000
H	3.65280000	0.63930000	7.00860000
H	2.56790000	2.87780000	7.25730000
H	1.36620000	3.90380000	5.36240000
H	0.26870000	4.41550000	2.00820000
H	0.85790000	4.71870000	3.50300000

1 2 1.5 6 1.5 7 1.0
2 3 1.5 9 1.0
3 4 1.5 35 1.0
4 5 1.5 36 1.0
5 6 1.5 13 1.0
6 37 1.0
7 8 1.0 10 2.0
8 9 1.0 12 1.0
9 11 2.0
10
11
12 38 1.0
13 14 1.0 26 2.0 27 2.0
14 15 1.5 19 1.5
15 16 1.5 39 1.0
16 17 1.5 40 1.0
17 18 1.5 20 1.0
18 19 1.5 22 1.0
19 41 1.0
20 21 1.0 24 2.0
21 22 1.0 25 1.0
22 23 2.0
23
24
25 28 1.5 32 1.5
26
27
28 29 1.5 42 1.0
29 30 1.5 43 1.0
30 31 1.5 44 1.0
31 32 1.5 45 1.0
32 33 1.0
33 34 1.0
34 46 1.0 47 1.0
35

36
37
38
39
40
41
42
43
44
45
46
47

Gaussian input coordinates and connectivity of PI_CD24529:

0 1

C	-3.32980000	-1.18410000	-2.72730000
C	-3.22700000	0.18360000	-3.00540000
C	-2.01130000	0.81890000	-3.20070000
C	-0.89270000	-0.02130000	-3.09570000
C	-0.98060000	-1.40090000	-2.81090000
C	-2.22930000	-2.02130000	-2.62210000
C	-4.66630000	-1.37860000	-2.61720000
N	-5.53030000	-0.31230000	-2.78360000
C	-4.51640000	0.59720000	-3.01850000
C	-6.94680000	-0.20590000	-2.74580000
O	-5.08580000	-2.48910000	-2.37710000
O	-4.75770000	1.76350000	-3.24010000
S	0.47430000	-2.33190000	-2.69680000
C	1.37800000	-1.53570000	-1.45130000
C	2.77690000	-1.40060000	-1.57550000
C	3.58310000	-0.79480000	-0.60000000
C	2.89690000	-0.33810000	0.51380000
C	1.51090000	-0.46690000	0.65330000
C	0.70860000	-1.06870000	-0.30490000
O	0.13030000	-3.59820000	-2.12420000
O	1.17390000	-2.08460000	-3.92160000
C	3.27440000	0.28090000	1.65780000
N	2.31500000	0.61030000	2.61490000
C	1.26130000	0.09500000	1.86120000
O	0.10980000	0.15650000	2.23670000
O	4.45130000	0.53230000	1.81210000
C	2.40210000	1.14890000	3.84530000
C	3.31540000	0.58650000	4.74130000
C	3.45630000	1.12080000	6.02130000

C	2.68590000	2.21870000	6.40400000
C	1.75900000	2.79160000	5.52560000
C	1.61940000	2.24380000	4.24050000
N	0.71410000	2.79830000	3.41130000
C	0.92940000	3.97390000	5.97100000
H	-1.92870000	1.89600000	-3.42160000
H	0.09600000	0.44540000	-3.24480000
H	-2.35280000	-3.09490000	-2.40440000
H	-7.34230000	-0.70090000	-1.82980000
H	-7.26550000	0.86100000	-2.73610000
H	-7.38110000	-0.70150000	-3.64360000
H	3.29170000	-1.79190000	-2.46910000
H	4.67630000	-0.69910000	-0.70780000
H	-0.37860000	-1.17310000	-0.15290000
H	3.93130000	-0.28090000	4.44970000
H	4.17790000	0.67710000	6.72850000
H	2.81400000	2.62900000	7.41920000
H	0.59160000	2.34280000	2.51070000
H	0.20270000	3.58370000	3.80770000
H	-0.15510000	3.72140000	5.93780000
H	1.16300000	4.29030000	7.01290000
H	1.11580000	4.85050000	5.30930000

1 2 1.5 6 1.5 7 1.0
2 3 1.5 9 1.0
3 4 1.5 35 1.0
4 5 1.5 36 1.0
5 6 1.5 13 1.0
6 37 1.0
7 8 1.0 11 2.0
8 9 1.0 10 1.0
9 12 2.0
10 38 1.0 39 1.0 40 1.0
11
12
13 14 1.0 20 2.0 21 2.0
14 15 1.5 19 1.5
15 16 1.5 41 1.0
16 17 1.5 42 1.0
17 18 1.5 22 1.0
18 19 1.5 24 1.0
19 43 1.0
20
21

22 23 1.0 26 2.0
23 24 1.0 27 1.0
24 25 2.0
25
26
27 28 1.5 32 1.5
28 29 1.5 44 1.0
29 30 1.5 45 1.0
30 31 1.5 46 1.0
31 32 1.5 34 1.0
32 33 1.0
33 47 1.0 48 1.0
34 49 1.0 50 1.0 51 1.0
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51

Gaussian input coordinates and connectivity of PI_CD24352:

0 1

C	-3.45670000	-0.98160000	-2.49630000
C	-3.33100000	0.34500000	-2.92370000
C	-2.11190000	0.91250000	-3.25720000
C	-1.01380000	0.04920000	-3.12640000
C	-1.12410000	-1.29150000	-2.69900000
C	-2.37650000	-1.84260000	-2.37080000
C	-4.78820000	-1.11370000	-2.28240000
N	-5.62820000	-0.03800000	-2.50350000
C	-4.60550000	0.80240000	-2.90090000
O	-5.22500000	-2.17800000	-1.90300000
O	-4.82650000	1.94620000	-3.23260000

C	-7.03420000	0.13010000	-2.38310000
S	0.30970000	-2.25470000	-2.57030000
C	1.22740000	-1.43130000	-1.35280000
C	2.61110000	-1.21820000	-1.52730000
C	3.41860000	-0.57410000	-0.57740000
C	2.75160000	-0.16320000	0.56570000
C	1.38120000	-0.37380000	0.75540000
C	0.57740000	-1.01120000	-0.17780000
C	3.13490000	0.46790000	1.70160000
N	2.19410000	0.73680000	2.69380000
C	1.14540000	0.16350000	1.97700000
O	0.01500000	0.12410000	2.41490000
O	4.29890000	0.78710000	1.82250000
C	2.27360000	1.30000000	3.91180000
O	-0.05850000	-3.49540000	-1.95840000
O	1.00470000	-2.06280000	-3.80750000
C	3.14050000	0.76650000	4.86890000
C	3.22110000	1.34760000	6.13620000
C	2.43770000	2.46060000	6.44850000
C	1.57190000	2.98820000	5.48850000
C	1.48480000	2.41180000	4.21940000
N	0.81630000	4.05830000	5.79060000
H	-2.01170000	1.95790000	-3.59360000
H	-0.02210000	0.46610000	-3.37200000
H	-2.51800000	-2.88110000	-2.02990000
H	-7.40150000	-0.36080000	-1.45350000
H	-7.30050000	1.21040000	-2.33690000
H	-7.54100000	-0.32930000	-3.26160000
H	3.11320000	-1.57320000	-2.44290000
H	4.49960000	-0.41720000	-0.72640000
H	-0.49640000	-1.17450000	0.01150000
H	3.76050000	-0.11490000	4.63290000
H	3.90290000	0.92420000	6.89330000
H	2.50470000	2.91450000	7.45170000
H	0.79840000	2.83390000	3.46530000
H	0.21000000	4.38000000	5.03920000
H	0.94050000	4.41610000	6.73530000

1 2 1.5 6 1.5 7 1.0
2 3 1.5 9 1.0
3 4 1.5 34 1.0
4 5 1.5 35 1.0
5 6 1.5 13 1.0
6 36 1.0

7 8 1.0 10 2.0
8 9 1.0 12 1.0
9 11 2.0
10
11
12 37 1.0 38 1.0 39 1.0
13 14 1.0 26 2.0 27 2.0
14 15 1.5 19 1.5
15 16 1.5 40 1.0
16 17 1.5 41 1.0
17 18 1.5 20 1.0
18 19 1.5 22 1.0
19 42 1.0
20 21 1.0 24 2.0
21 22 1.0 25 1.0
22 23 2.0
23
24
25 28 1.5 32 1.5
26
27
28 29 1.5 43 1.0
29 30 1.5 44 1.0
30 31 1.5 45 1.0
31 32 1.5 33 1.0
32 46 1.0
33 47 1.0 48 1.0
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48

Gaussian input coordinates and connectivity of PI_CD24603:

0 1

C	-3.62910000	-0.90330000	-2.52590000
C	-3.49790000	0.43110000	-2.92610000
C	-2.27530000	1.00300000	-3.23840000
C	-1.17960000	0.13500000	-3.11950000
C	-1.29540000	-1.21300000	-2.71750000
C	-2.55080000	-1.76740000	-2.40650000
C	-4.96290000	-1.03830000	-2.32880000
N	-5.79990000	0.04160000	-2.53990000
C	-4.77210000	0.88930000	-2.90770000
C	-7.20820000	0.20510000	-2.44010000
O	-5.40450000	-2.10910000	-1.97420000
O	-4.98850000	2.03960000	-3.21950000
S	0.13460000	-2.18280000	-2.60370000
C	1.06880000	-1.36460000	-1.39550000
C	2.45530000	-1.17700000	-1.57840000
C	3.27870000	-0.54220000	-0.63580000
C	2.62390000	-0.10990000	0.50610000
C	1.24990000	-0.29150000	0.70230000
C	0.43090000	-0.92270000	-0.22180000
O	-0.23310000	-3.42230000	-1.98900000
O	0.81810000	-1.99190000	-3.84730000
C	3.02320000	0.52440000	1.63460000
N	2.09110000	0.82240000	2.62590000
C	1.02920000	0.26310000	1.91880000
O	-0.10090000	0.25920000	2.35970000
O	4.19290000	0.82350000	1.75080000
C	2.18210000	1.43360000	3.81940000
C	2.95030000	0.86020000	4.83550000
C	3.04800000	1.49620000	6.07490000
C	2.37720000	2.70210000	6.28910000
C	1.60630000	3.27430000	5.27530000
C	1.50800000	2.63760000	4.03620000
N	2.47370000	3.31730000	7.48050000
H	-2.17140000	2.05380000	-3.55560000
H	-0.18620000	0.55230000	-3.35710000
H	-2.69580000	-2.81250000	-2.08760000
H	-7.58850000	-0.28940000	-1.51770000
H	-7.48070000	1.28390000	-2.39590000
H	-7.70010000	-0.25400000	-3.32730000
H	2.94680000	-1.54540000	-2.49400000
H	4.36200000	-0.40580000	-0.78970000
H	-0.64520000	-1.06340000	-0.02640000
H	3.47990000	-0.09240000	4.66510000

H	3.65510000	1.04360000	6.87720000
H	1.07670000	4.22690000	5.44590000
H	0.90230000	3.09140000	3.23350000
H	1.95570000	4.19050000	7.55340000
H	3.05540000	2.83720000	8.16390000

1 2 1.5 6 1.5 7 1.0
2 3 1.5 9 1.0
3 4 1.5 34 1.0
4 5 1.5 35 1.0
5 6 1.5 13 1.0
6 36 1.0
7 8 1.0 11 2.0
8 9 1.0 10 1.0
9 12 2.0
10 37 1.0 38 1.0 39 1.0
11
12
13 14 1.0 20 2.0 21 2.0
14 15 1.5 19 1.5
15 16 1.5 40 1.0
16 17 1.5 41 1.0
17 18 1.5 22 1.0
18 19 1.5 24 1.0
19 42 1.0
20
21
22 23 1.0 26 2.0
23 24 1.0 27 1.0
24 25 2.0
25
26
27 28 1.5 32 1.5
28 29 1.5 43 1.0
29 30 1.5 44 1.0
30 31 1.5 33 1.0
31 32 1.5 45 1.0
32 46 1.0
33 47 1.0 48 1.0
34
35
36
37
38

39
40
41
42
43
44
45
46
47
48

Gaussian input coordinates and connectivity of PI_CD24495:

0 1

C	4.60270240	0.57847246	-4.66489704
C	5.17528405	-0.35638123	-5.53355937
C	4.55542046	-1.55136815	-5.86073607
C	3.30768324	-1.74936742	-5.25004511
C	2.71770476	-0.81688443	-4.37043471
C	3.37039207	0.39018242	-4.05797328
C	5.52426852	1.57066548	-4.62518406
N	6.68810725	1.45529298	-5.36262542
C	6.34992486	0.22010867	-5.88234345
C	7.81403350	2.30390322	-5.53869303
O	5.31147289	2.54395689	-3.93610872
O	7.08900832	-0.36232062	-6.64503648
S	1.17930000	-1.14130000	-3.64060000
C	1.60120000	-1.12880000	-1.95980000
C	1.70160000	-2.33600000	-1.23600000
C	1.98290000	-2.39090000	0.13720000
C	2.16170000	-1.15600000	0.73810000
C	2.07580000	0.05010000	0.03320000
C	1.79030000	0.11150000	-1.32200000
O	0.39221006	0.02675914	-3.89583817
O	0.84569985	-2.50055426	-3.94107164
C	2.42140000	-0.76280000	2.00750000
N	2.51300000	0.59510000	2.30480000
C	2.29860000	0.98790000	0.98580000
O	2.28990000	2.15810000	0.66600000
O	2.55300000	-1.61040000	2.86530000
C	2.68000000	1.26230000	3.45940000
C	3.92070000	1.14770000	4.09260000
C	4.15650000	1.80330000	5.30070000
C	3.14080000	2.56790000	5.87340000
C	1.90450000	2.67550000	5.23440000

C	1.64350000	2.02260000	4.02200000
C	0.29040000	2.20380000	3.36250000
C	-0.15810000	3.63900000	3.19200000
C	0.69490000	4.54970000	2.55590000
C	0.32890000	5.88470000	2.37530000
C	-0.91630000	6.33300000	2.81580000
C	-1.78820000	5.43630000	3.43310000
C	-1.41170000	4.10380000	3.61890000
N	-2.29870000	3.29510000	4.22600000
H	5.00683224	-2.28860987	-6.54503133
H	2.78947972	-2.69236371	-5.49028299
H	2.95156164	1.14622425	-3.37308511
H	7.97060587	2.94019121	-4.63841785
H	8.73450371	1.69930939	-5.70366834
H	7.65033910	2.96112936	-6.42253767
H	1.53520000	-3.30260000	-1.73910000
H	2.03920000	-3.33940000	0.69640000
H	1.69770000	1.08060000	-1.84000000
H	4.72290000	0.53790000	3.64290000
H	5.13670000	1.71670000	5.79920000
H	3.31780000	3.09180000	6.82800000
H	1.12390000	3.29770000	5.70280000
H	0.29190000	1.75830000	2.34100000
H	-0.43130000	1.60330000	3.95840000
H	1.68170000	4.21270000	2.19690000
H	1.02480000	6.58580000	1.88500000
H	-1.21120000	7.38690000	2.67690000
H	-2.77350000	5.78960000	3.77920000
H	-1.99690000	2.33600000	4.37550000
H	-3.17920000	3.73680000	4.48300000

1 2 1.5 6 1.5 7 1.0
2 3 1.5 9 1.0
3 4 1.5 41 1.0
4 5 1.5 42 1.0
5 6 1.5 13 1.0
6 43 1.0
7 8 1.0 11 2.0
8 9 1.0 10 1.0
9 12 2.0
10 44 1.0 45 1.0 46 1.0
11
12
13 14 1.0 20 2.0 21 2.0

14 15 1.5 19 1.5
15 16 1.5 47 1.0
16 17 1.5 48 1.0
17 18 1.5 22 1.0
18 19 1.5 24 1.0
19 49 1.0
20
21
22 23 1.0 26 2.0
23 24 1.0 27 1.0
24 25 2.0
25
26
27 28 1.5 32 1.5
28 29 1.5 50 1.0
29 30 1.5 51 1.0
30 31 1.5 52 1.0
31 32 1.5 53 1.0
32 33 1.0
33 34 1.0 54 1.0 55 1.0
34 35 1.5 39 1.5
35 36 1.5 56 1.0
36 37 1.5 57 1.0
37 38 1.5 58 1.0
38 39 1.5 59 1.0
39 40 1.0
40 60 1.0 61 1.0
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57

58
59
60
61

Gaussian input coordinates and connectivity of PI_CD24378:

0 1

C	-3.15870000	-1.58020000	-3.11790000
C	-3.06340000	-0.24400000	-3.52280000
C	-1.85220000	0.37470000	-3.78610000
C	-0.72970000	-0.44800000	-3.61050000
C	-0.80820000	-1.79710000	-3.20350000
C	-2.05330000	-2.40020000	-2.94580000
C	-4.49340000	-1.76630000	-2.97560000
N	-5.36220000	-0.72100000	-3.22820000
C	-4.35410000	0.16360000	-3.56110000
C	-6.77750000	-0.60740000	-3.17320000
O	-4.90770000	-2.85030000	-2.62860000
O	-4.60210000	1.30290000	-3.88970000
S	0.65570000	-2.70600000	-3.02270000
C	1.50380000	-1.78850000	-1.82180000
C	2.75290000	-1.20010000	-2.11310000
C	3.48130000	-0.43750000	-1.18730000
C	2.87980000	-0.30520000	0.05390000
C	1.64410000	-0.88780000	0.35910000
C	0.91730000	-1.64000000	-0.55120000
O	0.31550000	-3.93610000	-2.37430000
O	1.35460000	-2.53510000	-4.26010000
C	3.22460000	0.32490000	1.20220000
N	2.37560000	0.26510000	2.30430000
C	1.43270000	-0.52120000	1.64630000
O	0.41580000	-0.88800000	2.19670000
O	4.27310000	0.93320000	1.24300000
C	2.45270000	0.75340000	3.55400000
C	3.23910000	0.06000000	4.47490000
C	3.34310000	0.53250000	5.78170000
C	2.65820000	1.68860000	6.15780000
C	1.86360000	2.38990000	5.24170000
C	1.76500000	1.91760000	3.92200000
C	1.16380000	3.54800000	5.60810000
C	0.37900000	4.23570000	4.68070000
C	0.28810000	3.76920000	3.37060000
C	0.98000000	2.61720000	2.99690000
N	1.21400000	4.04560000	6.85710000

H	-1.77530000	1.42730000	-4.10530000
H	0.25520000	0.00900000	-3.80650000
H	-2.17090000	-3.44870000	-2.62670000
H	-7.16390000	-1.08820000	-2.24600000
H	-7.08900000	0.46170000	-3.16900000
H	-7.22860000	-1.11000000	-4.05860000
H	3.20750000	-1.32650000	-3.10970000
H	4.45580000	0.01840000	-1.42840000
H	-0.05580000	-2.08220000	-0.27920000
H	3.77590000	-0.85690000	4.17800000
H	3.96340000	-0.00890000	6.51640000
H	2.76610000	2.02600000	7.20120000
H	-0.16970000	5.14700000	4.97390000
H	-0.32890000	4.30980000	2.63280000
H	0.88780000	2.27890000	1.95180000
H	0.66070000	4.88810000	6.99970000
H	1.79130000	3.52800000	7.51460000

1 2 1.5 6 1.5 7 1.0
2 3 1.5 9 1.0
3 4 1.5 38 1.0
4 5 1.5 39 1.0
5 6 1.5 13 1.0
6 40 1.0
7 8 1.0 11 2.0
8 9 1.0 10 1.0
9 12 2.0
10 41 1.0 42 1.0 43 1.0
11
12
13 14 1.0 20 2.0 21 2.0
14 15 1.5 19 1.5
15 16 1.5 44 1.0
16 17 1.5 45 1.0
17 18 1.5 22 1.0
18 19 1.5 24 1.0
19 46 1.0
20
21
22 23 1.0 26 2.0
23 24 1.0 27 1.0
24 25 2.0
25
26

27 28 1.5 32 1.5
28 29 1.5 47 1.0
29 30 1.5 48 1.0
30 31 1.5 49 1.0
31 32 1.5 33 1.5
32 36 1.5
33 34 1.5 37 1.0
34 35 1.5 50 1.0
35 36 1.5 51 1.0
36 52 1.0
37 53 1.0 54 1.0
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54

Supplementary Equations

Norm indexes are defined as follows:

$$\|M\|_1 = \frac{1}{n}\left(\sum_j |m_{ij}|\right) \tag{S1}$$

$$\|M\|_2 = \frac{1}{n}\left(\sum_j \sum_i |m_{ij}|\right) \tag{S2}$$

$$\|M\|_3 = \sqrt{\max\left(\sum_j m_{ij}^{\,2}\right)} \tag{S3}$$

$$\|M\|_4 = \sqrt{\sum_j \sum_i |m_{ij}|^2} \tag{S4}$$

$$\|M\|_5 = \max\left(\sum_j |m_{ij}|\right) \tag{S5}$$

$$\|M\|_6 = \max\left(\sum_i |m_{ij}|\right) \tag{S6}$$

where, λ_i is the eigenvalue of matrix, the M^H is the Hermite matrix.

The seven *MS* (containing MS_F, MS_A, MS_{AB}, MS_{ABC}, MS_{bond}, MS_{bondA}, MS_{ABCaro}) are defined as follows:

$$MS_F = [a_{ij}] \qquad a_{ij} = \begin{cases} s_{ij} & s_{ij} \neq 0 \\ 0 & otherwise \end{cases} \tag{S7}$$

$$MS_A = [a_{ij}] \qquad a_{ij} = \begin{cases} s_{ij} & s_{ij} = 1 \\ 0 & otherwise \end{cases} \tag{S8}$$

$$MS_{AB} = [a_{ij}] \qquad a_{ij} = \begin{cases} s_{ij} & s_{ij} = 1 \; or \; 2 \\ 0 & otherwise \end{cases} \tag{S9}$$

$$MS_{ABC} = [a_{ij}] \qquad a_{ij} = \begin{cases} s_{ij} & s_{ij} = 1 \; or \; 2 \; or \; 3 \\ 0 & otherwise \end{cases} \tag{S10}$$

$$MS_{bond} = [a_{ij}] \qquad a_{ij} = \begin{cases} b_{ij} & s_{ij} = 1, b_{ij} = 1 \; or \; 2 \; or \; 3 \; or \; 1.5 \\ 0 & otherwise \end{cases} \tag{S11}$$

$$MS_{bondA} = [a_{ij}] \qquad a_{ij} = \begin{cases} b_{ij} & s_{ij} = 1, b_{ij} = 1 \\ 0 & otherwise \end{cases} \tag{S12}$$

$$MS_{ABCaro} = [a_{ij}] \qquad a_{ij} = \begin{cases} s_{ij} & s_{ij} = 1 \; or \; 2 \; or \; 3, \text{and atom } j \text{ is on the benzene ring} \\ 0 & otherwise \end{cases} \tag{S13}$$

where s_{ij} is the step between atoms *i* and *j* and b_{ij} is the type of chemical bond between atoms *i* and *j*. Specifically, for b_{ij}, single, double, and triple bonds are denoted as 1.0, 2.0, and 3.0, while bonds between aromatic atoms are denoted as 1.5.