Supporting Information

Machine Learning Aided Design of Polymer with Targeted Band Gap Based on DFT Computation

Pengcheng Xu[a], Tian Lu [a], Lifei Ju[b], Lumin Tian[b], Minjie Li[b], Wencong Lu[a,b,*]

[a] *Materials Genome Institute, Shanghai University, and Shanghai Materials Genome Institute, Shanghai 200444, China*

[b] *Department of Chemistry, College of Sciences, Shanghai University, Shanghai 200444, China*

AUTHOR INFORMATION

Corresponding Author *Phone: 086-021-66132406. E-mail: wclu@shu.edu.cn.

ORCID Wencong Lu: 0000-0001-5361-6122

Table S1: The repeating units and experimental band gap in paper

NO.	Repeating units	Experimental band gap/eV	DFT calculated band gap/eV
1	NH-CS-NH-C6H4-CH2-C6H4	3.300[1]	3.423
2	NH-CS-NH-C6H4	3.100[2]	3.426
3	NH-CS-NH-C6H4-NH-CS-NH-C6H4	3.070[1]	3.072
4	NH-CO-NH-C6H4	3.900[2]	4.455
5	NH-CS-NH-C6H4-NH-CS-NH-C6H4-CH2-C6H4	3.160[1]	3.052
6	NH-CS-NH-C6H4-NH-CS-NH-C6H4-O-C6H4	3.220[1]	3.099
7	NH-CS-NH-C6H4-NH-CS-NH-CH2-CH2-CH2-CH2	3.5300[1]	3.080
8	CO-NH-CO-C6H4	4.000[2]	3.374
9	CH2-CH2-CH2-CH2	8.800[3]	8.825

Table S2: 284 repeating units and DFT calculated band gap values

NO	Polymer samples	DFT band gap/eV
1	CH2-CO-NH-CS	2.691
2	CH2-CS-C6H4-O	2.041
3	C6H4-C6H4-C4H2S-CS	1.807
4	CO-O-C4H2S-CS	1.918
5	C6H4-CS-C4H2S-O	1.930
6	CO-C6H4-O-CS	2.772
7	C6H4-O-CS-C4H2S	2.537
8	C6H4-C4H2S-C6H4-C4H2S	2.645
9	NH-C4H2S-CO-C4H2S	2.678
10	CH2-NH-C4H2S-CS	2.063
11	NH-C6H4-CO-C6H4	3.042
12	NH-CO-CS-C4H2S	1.701
13	CO-C6H4-O-C4H2S	3.410
14	C6H4-CS-C4H2S-CS	1.332
15	CH2-C4H2S-NH-C4H2S	3.092
16	CH2-C4H2S-C6H4-C4H2S	2.872
17	CH2-NH-C6H4-CO	3.226
18	CO-C6H4-C4H2S-CS	1.699
19	C6H4-C6H4-CS-O	2.623
20	CH2-NH-C4H2S-C6H4	2.951
21	NH-C4H2S-C4H2S-CH2	2.919
22	CH2-C4H2S-C6H4-CS	1.841
23	CH2-NH-CS-C6H4	2.737
24	CH2-C6H4-C4H2S-C6H4	2.960
25	NH-CO-C4H2S-C6H4	3.244
26	NH-CO-CS-CO	1.441
27	CO-C6H4-C6H4-C6H4	2.993
28	CH2-CH2-C4H2S-O	4.186
29	NH-CS-CO-CS	1.512
30	CH2-C6H4-CS-O	2.729
31	CH2-CO-C4H2S-CS	1.647
32	C4H2S-C4H2S-C4H2S-O	2.267

33	CO-C4H2S-C6H4-CS	1.708
34	CH2-C6H4-NH-O	4.141
35	NH-CS-NH-CS	2.491
36	CH2-NH-CS-NH	4.019
37	NH-C4H2S-C4H2S-CO	2.521
38	CH2-C4H2S-CS-O	2.659
39	NH-CO-C4H2S-O	3.567
40	CH2-C4H2S-C6H4-O	3.375
41	C6H4-C6H4-C6H4-C4H2S	3.046
42	NH-C6H4-CO-O	3.667
43	NH-C4H2S-CO-O	3.840
44	NH-CS-NH-C6H4	3.426
45	CO-CS-O-CS	1.532
46	CH2-NH-CH2-CS	2.141
47	CH2-CH2-NH-C6H4	4.005
48	C4H2S-C4H2S-C4H2S-CS	1.686
49	CH2-C6H4-C6H4-O	3.823
50	CH2-CO-C6H4-CS	1.741
51	CO-C6H4-C6H4-CS	1.800
52	CH2-CH2-CH2-CS	2.315
53	CH2-C6H4-CH2-O	4.867
54	CH2-O-CO-O	6.933
55	NH-C6H4-CS-C6H4	1.924
56	CH2-CO-C6H4-O	3.499
57	CH2-NH-C4H2S-NH	3.701
58	CH2-C6H4-CO-O	4.293
59	NH-C4H2S-NH-CS	3.123
60	CO-C6H4-C4H2S-O	2.590
61	NH-C4H2S-O-C4H2S	3.878
62	CH2-NH-CH2-CO	3.842
63	NH-CO-NH-C6H4	4.455
64	CH2-C6H4-NH-CS	2.960
65	CH2-CH2-CH2-O	6.863
66	CH2-CS-CH2-O	2.411

67	NH-C6H4-C6H4-O	3.619
68	CH2-C4H2S-NH-CS	3.025
69	CH2-C6H4-O-C4H2S	4.281
70	CH2-CH2-CO-C6H4	3.407
71	CO-C4H2S-C4H2S-CS	1.599
72	CH2-C6H4-O-C6H4	4.132
73	CO-C4H2S-C6H4-O	2.740
74	CO-C4H2S-CO-O	2.930
75	CH2-CS-C4H2S-O	2.028
76	CH2-O-C4H2S-C6H4	3.240
77	CH2-C4H2S-CO-C4H2S	3.095
78	CO-C6H4-CO-O	2.979
79	CH2-CH2-CO-C4H2S	3.376
80	CH2-CO-CS-C6H4	1.754
81	CH2-CS-CO-O	1.599
82	NH-C4H2S-NH-C4H2S	3.515
83	C6H4-CS-C6H4-CS	1.511
84	CH2-C6H4-CH2-C6H4	4.691
85	CH2-C4H2S-O-CS	3.061
86	CO-C6H4-C6H4-C4H2S	2.705
87	CH2-CH2-NH-C4H2S	3.931
88	CH2-CH2-CO-O	5.794
89	CH2-CO-CS-C4H2S	1.651
90	CH2-C6H4-CH2-CS	2.354
91	CH2-C6H4-NH-C6H4	3.570
92	C6H4-C4H2S-C6H4-O	2.904
93	C6H4-C4H2S-O-CS	2.824
94	CH2-C4H2S-C4H2S-CS	1.762
95	CH2-CO-CH2-O	4.348
96	C6H4-C4H2S-C4H2S-C4H2S	2.321
97	CH2-C4H2S-NH-O	3.964
98	CH2-CO-C6H4-C4H2S	3.001
99	NH-CO-O-C4H2S	4.288
100	CH2-C4H2S-CO-O	3.772

101	CH2-CS-C6H4-NH	2.063
102	CO-O-C6H4-CS	1.932
103	NH-CO-C4H2S-CS	1.773
104	NH-CO-O-CS	2.917
105	CH2-NH-CS-CO	2.093
106	CH2-CO-C4H2S-O	3.320
107	CH2-CH2-C6H4-O	4.351
108	CH2-C6H4-CS-C6H4	1.889
109	CO-C6H4-CO-C6H4	2.826
110	CH2-C4H2S-CH2-O	4.313
111	CH2-NH-C4H2S-CO	2.987
112	CH2-NH-C6H4-NH	3.392
113	NH-CS-O-CS	2.557
114	CH2-C6H4-O-CS	3.054
115	NH-CS-C4H2S-O	2.562
116	CO-C6H4-C6H4-O	2.965
117	CH2-NH-CH2-C4H2S	4.226
118	C4H2S-CS-O-CS	1.912
119	CO-C4H2S-C6H4-C4H2S	2.505
120	CO-C4H2S-C4H2S-C4H2S	2.080
121	CH2-NH-CS-C4H2S	2.571
122	NH-O-CS-O	4.333
123	CH2-NH-CH2-NH	5.942
124	C6H4-CS-C4H2S-C4H2S	1.728
125	NH-CO-C6H4-CS	1.836
126	NH-CO-NH-CO	4.829
127	C4H2S-C4H2S-O-CS	2.588
128	NH-CS-NH-O	3.988
129	CO-O-C4H2S-O	4.016
130	CH2-O-C4H2S-O	4.150
131	NH-O-CO-O	6.731
132	CH2-CO-O-CS	2.705
133	CH2-O-C6H4-O	3.859
134	NH-C6H4-NH-O	3.571

135	CH2-CH2-C6H4-C4H2S	3.526
136	NH-C4H2S-CS-O	2.612
137	C6H4-C6H4-C4H2S-C4H2S	2.615
138	NH-C4H2S-C6H4-C4H2S	2.584
139	CO-C4H2S-C4H2S-O	2.360
140	NH-C4H2S-CO-CS	1.633
141	NH-C4H2S-C4H2S-C4H2S	2.344
142	CH2-CH2-CH2-C6H4	5.108
143	NH-C4H2S-O-CS	2.550
144	CH2-NH-CS-O	4.240
145	CH2-CO-O-C4H2S	4.022
146	CO-C6H4-CS-C6H4	1.753
147	NH-CS-C4H2S-CS	1.614
148	NH-CO-C6H4-C4H2S	3.238
149	NH-C4H2S-CS-C4H2S	1.882
150	C6H4-C4H2S-O-C4H2S	3.310
151	C6H4-C6H4-C6H4-CS	1.879
152	CH2-C4H2S-CH2-C4H2S	4.109
153	CO-CS-CO-O	1.477
154	CO-C6H4-C4H2S-C6H4	2.546
155	CH2-NH-CH2-C6H4	4.318
156	CH2-C6H4-CO-C4H2S	3.144
157	CH2-O-CH2-O	6.436
158	CO-C6H4-C4H2S-C4H2S	2.361
159	CH2-CO-CS-O	2.161
160	NH-CO-CS-O	2.557
161	CH2-CO-CS-CO	1.278
162	CH2-NH-C6H4-C4H2S	3.176
163	NH-CS-C6H4-CS	1.639
164	NH-C6H4-C6H4-CH2	3.569
165	NH-CO-NH-CS	2.941
166	NH-C4H2S-C4H2S-CS	1.764
167	NH-C6H4-NH-C6H4	3.313
168	NH-C6H4-NH-C4H2S	3.149

169	CH2-C6H4-C6H4-C4H2S	3.190
170	CH2-C4H2S-CS-C4H2S	1.802
171	CO-NH-CO-C6H4	3.710
172	CH2-CS-CO-CS	1.492
173	CH2-CH2-NH-CS	3.375
174	CH2-C4H2S-C4H2S-O	2.817
175	CH2-CO-NH-C4H2S	4.130
176	CH2-CH2-CO-CS	1.425
177	NH-C6H4-C6H4-C4H2S	2.930
178	CH2-C6H4-C4H2S-CS	1.834
179	C6H4-O-CS-O	3.903
180	CH2-O-NH-CO	5.357
181	NH-C6H4-C4H2S-C6H4	2.753
182	C6H4-O-C4H2S-O	4.061
183	NH-C6H4-O-C4H2S	3.606
184	CH2-NH-CO-CS	1.827
185	CH2-NH-C6H4-O	3.538
186	C6H4-C6H4-C6H4-C6H4	3.307
187	CH2-CO-NH-C6H4	4.296
188	CH2-CH2-CH2-CH2	8.828
189	NH-CO-C6H4-C6H4	3.656
190	NH-CO-CS-C6H4	1.721
191	CO-C6H4-CS-C4H2S	1.787
192	CH2-CO-CH2-CO	3.761
193	CO-O-C6H4-O	4.193
194	NH-CO-C4H2S-CO	3.029
195	CH2-C6H4-C4H2S-C4H2S	2.737
196	CO-C6H4-CO-C4H2S	2.867
197	CH2-C6H4-CO-CS	1.735
198	CH2-CH2-C6H4-CS	1.965
199	NH-C4H2S-C6H4-CS	1.850
200	CH2-NH-O-C6H4	4.520
201	CH2-C6H4-C6H4-C6H4	3.533
202	CH2-O-NH-O	6.258

203	CH2-CH2-C4H2S-CS	1.939
204	NH-CO-NH-C4H2S	4.069
205	CH2-C4H2S-CH2-CS	2.322
206	NH-CO-NH-O	5.284
207	CO-C6H4-CS-O	2.340
208	CH2-CH2-O-NH	6.746
209	CH2-CO-CH2-C4H2S	3.997
210	C6H4-O-C6H4-O	4.085
211	CH2-O-CS-O	3.958
212	NH-CO-C6H4-O	4.072
213	C4H2S-O-C4H2S-O	4.103
214	CH2-NH-CO-C6H4	3.951
215	CH2-C4H2S-O-C4H2S	3.773
216	CH2-CO-O-CO	4.582
217	CH2-CO-C4H2S-C6H4	3.020
218	CH2-CO-C6H4-CO	2.795
219	NH-C4H2S-C4H2S-O	2.834
220	CH2-CH2-C4H2S-C4H2S	3.124
221	CH2-NH-CH2-O	5.758
222	NH-C6H4-C6H4-C6H4	3.145
223	C6H4-CS-O-CS	1.914
224	CO-C4H2S-CS-O	2.246
225	NH-C6H4-C4H2S-C4H2S	2.587
226	CH2-CH2-CH2-NH	6.207
227	CH2-CO-NH-CO	4.232
228	NH-CS-C6H4-O	2.700
229	CO-C4H2S-CO-C4H2S	2.765
230	NH-C6H4-O-CS	2.945
231	CH2-C4H2S-CO-CS	1.532
232	CH2-CO-O-C6H4	4.756
233	C4H2S-C4H2S-C4H2S-C4H2S	2.019
234	CO-C6H4-O-C6H4	3.441
235	CH2-CH2-CH2-C4H2S	4.506
236	NH-CO-O-C6H4	4.926

237	CO-C4H2S-O-CS	2.823
238	C6H4-C4H2S-C4H2S-O	2.605
239	CH2-CO-C4H2S-C4H2S	2.596
240	CH2-CO-CH2-CS	2.174
241	NH-C6H4-C4H2S-CS	1.856
242	CH2-C4H2S-C4H2S-C4H2S	2.328
243	NH-C6H4-CS-O	2.830
244	NH-C6H4-CO-C4H2S	2.887
245	CO-C4H2S-CO-CS	1.447
246	NH-CS-CO-O	2.267
247	CH2-CS-NH-CS	2.051
248	C6H4-C6H4-C4H2S-O	2.951
249	CH2-CO-C4H2S-CO	2.750
250	CH2-NH-CO-C4H2S	3.946
251	CH2-C6H4-C6H4-CS	1.898
252	CH2-C6H4-NH-C4H2S	3.464
253	C6H4-C4H2S-C6H4-CS	1.806
254	CH2-C6H4-CS-C4H2S	1.845
255	C6H4-C4H2S-CS-C4H2S	1.740
256	C4H2S-O-CS-O	3.820
257	CH2-CS-NH-O	3.256
258	CH2-C6H4-CH2-C4H2S	4.409
259	CO-O-CS-O	3.374
260	NH-C6H4-O-C6H4	3.875
261	CH2-C6H4-CO-C6H4	3.186
262	NH-CO-O-CO	5.163
263	CO-C4H2S-CS-C4H2S	1.633
264	CH2-CS-O-CS	2.214
265	C4H2S-CS-C4H2S-O	1.860
266	CO-C4H2S-O-C4H2S	3.325
267	C6H4-CS-C6H4-O	1.915
268	NH-O-C6H4-O	3.916
269	CH2-CO-CH2-C6H4	4.257
270	CH2-NH-O-CS	3.231

271	CH2-NH-C4H2S-O	3.881
272	CH2-CS-C4H2S-CS	1.319
273	CO-C6H4-CO-CS	1.648
274	CH2-CH2-CH2-CO	4.112
275	CH2-CO-C6H4-C6H4	3.215
276	CH2-CH2-CS-O	3.309
277	NH-C4H2S-CS-C6H4	1.964
278	CH2-NH-O-C4H2S	4.294
279	C6H4-C6H4-C6H4-O	3.396
280	NH-C6H4-CO-CS	1.760
281	CH2-CH2-C6H4-C6H4	4.020
282	CH2-CS-CH2-CS	1.890
283	NH-C6H4-C6H4-CS	1.917
284	CH2-NH-O-NH	5.865

Reference:

1. Ma, R.; Sharma, V.; Baldwin, A. F.; Tefferi, M.; Offenbach, I.; Cakmak, M.; Weiss, R.; Cao, Y.; Ramprasad, R.; Sotzing, G. A., Rational design and synthesis of polythioureas as capacitor dielectrics. *J. Mater. Chem. A* **2015,** *3*, 14845-14852.

2. Sharma, V.; Wang, C.; Lorenzini, R. G.; Ma, R.; Zhu, Q.; Sinkovits, D. W.; Pilania, G.; Oganov, A. R.; Kumar, S.; Sotzing, G. A.; Boggs, S. A.; Ramprasad, R., Rational design of all organic polymer dielectrics. *Nat. Commun.* **2014,** *5*, 4845.

3. Picaud, F.; Smogunov, A.; Corso, A. D.; Tosatti, E., Complex band structures and decay length in polyethylene chains. *J. Phys.: Condens. Matter* **2003,** *15*, 3731-3740.