

## **Data-based *Polymer-Unit Fingerprint (PUFp)*: A Newly Accessible Expression of Polymer Organic Semiconductors for Machine Learning**

**Xinyue Zhang<sup>1</sup>, Genwang Wei<sup>1, 2</sup>, Ye Sheng<sup>3</sup>, Jiong Yang<sup>3, 5</sup>, Caichao Ye<sup>1, 2, 4, \*</sup>, Wenqing Zhang<sup>1, 4, \*</sup>**

<sup>1</sup> *Department of Physics & Guangdong Provincial Key Laboratory of Computational Science and Material Design, Southern University of Science and Technology, Shenzhen, 518055, PR China*

<sup>2</sup> *Academy for Advanced Interdisciplinary Studies, Southern University of Science and Technology, Shenzhen, 518055, PR China*

<sup>3</sup> *Materials Genome Institute, Shanghai University, Shanghai, 200444, PR China*

<sup>4</sup> *Department of Materials Science and Engineering, Southern University of Science and Technology, Shenzhen, 518055, PR China*

<sup>5</sup> *Zhejiang Laboratory, Hangzhou, 311100, PR China*

<sup>\*</sup> Corresponding Authors: C. Ye: [yecc@sustech.edu.cn](mailto:yecc@sustech.edu.cn), W. Zhang: [zhangwq@sustech.edu.cn](mailto:zhangwq@sustech.edu.cn).

## 1. Polymer-unit SMILES

### (1) Branch-chain

11. 'CCCCCCCC(CC)CCCCCCC',  
27. 'CCCCCCCCCCCCCCCCCCCCCCCCCCCC',  
29. 'CCCCCCCCCCCCCCCCCCCC(CCCCCC)CCCCCCCCCCCCCCCC',  
31. 'CCCCCCCCCCCCCCCCCCCC(CCCCCC)CCCCCCCCCCCCCCCC',  
39. 'CCCCCCCCCCCCCCCC(C)CCCCCCCCC',  
42. 'CCCCCCCCCCCCCCCCCCCC(CCC)CCCCCCCCCCCCC',  
46. 'CCCCCCCCCCCCCCCCCCCC(CCCCC)CCCCCCCCCCCCC',  
51. 'CCCCCCCCCCCCCCCC(CC)CCCCCCCCC',  
56. 'CCCCCCCC(CCCCC)CCCC',  
62. 'CCCCC(C)CC',  
65. 'CCCCCCCCCCCCCCCCCCCC(C)CCCCCCCCCCCCC',  
68. 'CCCCCCCCCCCC(C)CCCCCC',  
69. 'CCCCCCCCCCC',  
75. 'CCCCCCCCCCCCCCCC(CCC)CCCCCCC',  
87. 'CCCCCCCCCCCCCCCCCCCCCCCCCCCC',  
109. 'CCCCCCCCCCCCCCCC(CCC)CCCCCCCCCCCCC',  
114. 'CCCCCCCCCCCCCCCCCCCC(C)CCCCCCCCCCCCCCCC',  
144. 'CCC(C)CCCC(C)C',  
147. 'CCCCCCCC(C)CCCCCCC',  
158. 'CCCCCCCCCCCCCCCCCCCC(C)CCCCCCCCCCCCCCCC',  
163. 'CCCCCCCCCCCCCCCCCCCC(CCCCC)CCCCCCCCCCCCCCCCCCCC',  
167. 'CCCCCCCCCCCCCCCC(CCCC)CCCCCCCCCCCCC',  
171. 'CCCCCCCCCCCCCCCC(CCCCC)CCCCCCCCCCC',  
181. 'CCCCCCCC(CCCCC)CCCCCCC',  
191. 'CCCCCCCCCCCCCCCCCCCCCCCC(CCC)CCCCCCCCCCCCCCCCCCCC',  
193. 'CCCCCCCC(C)CC',  
195. 'CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC',  
196. 'CCCCCCCC(CCCCC)CCCCCCC',  
206. 'CCCCCCCC(CCC)CCCC',  
213. 'CCCCCCCCCCCCCCCCCCCC(CC)CCCCCCCCCCCCC',  
214. 'CCCCCCCCCCCCCCCCC',  
216. 'CCCCCCCCCCCCCCCCCCCC(CCCC)CCCCCCCCCCCCC',  
221. 'CCCCCCCCCCCCCCCCCCCCC',  
231. 'CCCCCCCCCCCCCCCCCCCCC',  
238. 'CCCCCCCCCCCCCCCCCCCC(C)CCCCCCCCCCC',  
250. 'CCCCCCCC',  
252. 'CCCCCCCCC',  
257. 'CCCCC',  
258. 'CCCCC',  
265. 'CCCCCCCC=CCCCCCCCC',  
267. 'CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC',  
270. 'CCCCC(CC)CCC',  
276. 'CCCCCCCC(CCC)CCCCCCC',  
283. 'CCCCCCCCCCCCCCCCCCCCCCCCC',  
285. 'CCCCCCCCCCCCCCCCCCCC(C)CCCCCCCCCCCCC',

298.'CCCCCCCCCCCCCCCC',  
 300.'CCCCCCCCCCCC',  
 304.'CCCCCCCCCCCC(C)CCCCCCCCCCCC',  
 310.'CCCCCCCCCCCC(C)CCCCCCCCCCCC',  
 315.'CCCCCCCCCCCCCCCC(C)CCCCCCCCCCCCCCCC',  
 320.'CCCCCCCCCCCCCCCCCCCC(C)CCCCCCCCCCCCCCCC',  
 327.'CCCCCCCCCCCCCCCCCCCCCCCC(CCC)CCCCCCCCCCCCCCCCCCCC',  
 329.'CCCCCCCCCCCCCCCCCCCCCCCCCCCC',  
 333.'CCCC(C)CCCC',  
 334.'CCCCCCCCCCCC(CCCC)CCCCCCCCCCCC',  
 335.'CC(C)CCCC(C)C',  
 337.'CCCCCCCC(C)CCCCCCCC',  
 346.'CC(C)C',  
 352.'CCCCCCCCCCCCCCCCCCCCCCCC(CCCC)CCCCCCCCCCCCCCCCCCCC',  
 355.'CCCCCCC(C)CCCC',  
 357.'CCCCCCCCCCCCCCCCCCCC',  
 366.'CCCCCCCCCCCCCCCCCCCCCCCC(C)CCCCCCCCCCCCCCCCCCCC',  
 368.'CCCCCCCCCCCCCCCCCCCCCCCC(C)CCCCCCCCCCCCCCCCCCCC',  
 373.'CCCCCCCCCCCCCCCC(CCCCCCCC)CCCCCCCCCCCC',  
 375.'CCCCCCCCCCCC(CCC)CCCCCCCC',  
 380.'CCCCCCCCCCCCCCCC(CCCCCC)CCCCCCCCCCCC',  
 382.'CCCCCCCCCCCCCCCC(CCCCCCCC)CCCCCCCCCCCC',  
 390.'CCCCCCCCCCCC',  
 396.'CCCCCCCCCCCCCCCCCCCCCCCC(CCCCCC)CCCCCCCCCCCCCCCCCCCC',  
 397.'CCCCCCCCCCCCCCCCCCCC(CCC)CCCCCCCCCCCC',  
 398.'CCCCCCCCCCCCCCCCCCCC',  
 404.'CCCCCCCCCCCC(CCCCCC)CCCCCCCCCCCC',  
 413.'CCCC(C)CC',  
 428.'CCCCCCCCCCCC(C)CCCCCCCCCCCC',  
 432.'CCCCCCCCCCCC(CCCC)CCCCCCCC',  
 433.'CCCCCCCCCCCCCCCCCCCCCCCCCCCC',  
 439.'CCCCCCCCCCCC(C)CCCCCCCC',  
 444.'CCCCCCCCCCCCCCCCCCCCCCCC(CCCCC)CCCCCCCCCCCCCCCCCCCC'

## (2) Chain-F

94.'CCCCC(F)(F)C(F)(F)C(F)(F)C(F)(F)F',  
 253.'FC(F)F',  
 287.'CCCCCCCCCCCC(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)F',  
 293.'CCCCCCCCCCCC(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)F',  
 318.'CCCCC(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)F',  
 370.'CCCCC(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)F',  
 387.'CCCCC(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)F',  
 435.'CCCCCCCCC(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)F'

## (3) Chain-F-O

197.'CCCCOCCC(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)F',  
 294.'CCCCOCCC(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)F'

## (4) Chain-N

129.'CCC#N',  
146.'CCCCCCCCCCCCCN',  
159.'CCCCCCCCCN',  
248.'C=C(C)C#N',  
305.'CC=CC#N',  
311.'N#CCC#N',  
313.'C=C(C#N)C#N

(5) Chain-N-F

43.'CCCCCCCCCCC(CCCCCCCC)CNB(F)F'  
369.'NB(F)F'

(6) Chain-O

2.'CCCCCCCCCCCCCCCC(CCCCCCCCCCCCCC)OC(=O)CCC',  
6.'CCCCCCC(CCCC)OC=O',  
9.'CCCCCCCCCCO',  
15.'CCCOCCOC',  
21.'COCCOC',  
23.'CCCCCCCCOCCOCCOCCCO',  
25.'CCCCCCCCCCCCCOC(=O)CCC',  
26.'CCCCCCC(CO)CCCC',  
55.'CCCCCCCCOCCOCCOCCOC',  
57.'CCCCCCCCCCCCCO',  
59.'CC=O',  
71.'CCCCCCC(O)CCCC',  
74.'COCCOCCOCCOC',  
77.'CCCCCCCCCO',  
84.'CCOCOCOC',  
96.'CCCCCCCCCCCCCOC(=O)CCCCC',  
97.'CCCOCOCCO',  
105.'CCCCCCCCCCC(CCCO)CCCCCCCC',  
111.'C#CC(C)CO',  
113.'CCCCCCCCCCCCCOC(=O)CCCCC',  
120.'CCCCCCCCOCCOCCOCCOC',  
125.'CCCCCCCCCCCCCCCCCO',  
134.'CCCCCCCCCCCCCOC(=O)CCC',  
139.'CCCCCCCCCCCCC=O',  
157.'CCCCCCCCCCCCCOC=O',  
166.'CCCCCCCCOCCOCCOCCCO',  
174.'CCCCCCCCCCCC(C=O)CCCCCCCCC',  
192.'CCCCCCCCCCCC(CCO)CCCCCCCCC',  
208.'CCCCC(CC)COC=O',  
215.'CCOCCOCCOCCOC',  
217.'CCCCCCCCOC=O',  
225.'CCCCCCCCCCCC(O)CCCCCCCCC',  
227.'CC(C)CCCC(C)COC=O',  
233.'CCCCCCO',  
241.'CCCCCCCCCCCC(CCCO)CCCCCCCCC',  
255.'CCOCCOCCOC',

260.'CCCCCCCCCCCCCCCCO',  
 261.'CCCCCCCCCCCCCCCC(CCCCCCCCCC)COC(=O)CCC',  
 282.'CCCCC(CC)OC=O',  
 292.'CCCCCCCCCCCC(CCCCCCCC)COC(=O)CCC',  
 295.'CCCCCCCCCCCCCCCC(CO)CCCCCCCCCCCC',  
 314.'CCCCCCCCCCCCOC=O',  
 317.'CCCCCCCCCCCC(CO)CCCCCCCC',  
 319.'CCCCCCCCCCCCCCCCCCCCO',  
 321.'CCCCCCCCCCCCCCCCO',  
 323.'CC(=CCCC)CCO',  
 331.'CCCCCCCCCCCCCCCC(CCCCCCCCCC)COC(=O)CCCCC',  
 336.'CCCCC(O)CC',  
 363.'CCCCCCCCCCCCOCCOCCOCCO',  
 364.'CCOCCOC',  
 377.'CCCCCCCCOCCOCCOCCOCC',  
 379.'CCCCC(CC)CO',  
 384.'COCCOCC(C)COCCOC',  
 395.'CCCCCOCOCOCOCOCOC',  
 411.'CCCCOCCOCCOC',  
 425.'CCCCCCCCCCCC(CCCCCCCC)COC(=O)CCCCC',  
 434.'CCOCCOCCOC',  
 437.'CCCCCCCCCCCCCCCC(CCCCCO)CCCCCCCCCCCC',  
 441.'CCCCCCCCOCCCCC'

#### (7) Chain-O-Si

19.C[Si](C)(C)O[Si](C)(C)O[Si](C)(C)C  
 73.CCCCCC[Si](C)(O[Si](C)(C)C)O[Si](C)(C)C  
 93.CCCCCCCCCC[Si](OC)OC  
 118.CCCCC[Si](C)(O[Si](C)(C)C)O[Si](C)(C)C  
 119.CCCCC[Si]O[Si]C  
 154.C[Si](C)(C)O  
 173.CCCCCCCCCC[Si](C)(O[Si](C)(C)C)O[Si](C)(C)C  
 244.CCCCCCCCCC[Si]O[Si]C  
 246.CCCCCCCCCC[Si](C)(O[Si](C)(C)C)O[Si](C)(C)C  
 264.CCCCC[Si]O[Si]C  
 291.CCCCCC[Si]O[Si]C  
 297.CCCCC[Si](C)(O[Si](C)(C)C)O[Si](C)(C)C  
 326.CCCCC[Si](C)(O[SiH](C)C)O[Si](C)(C)C  
 349.CCCC[Si](C)(O[Si](C)(C)C)O[Si](C)(C)C  
 365.CCCC[Si]O[Si]C  
 407.CCCCCCCCC[Si](C)(O[Si](C)(C)C)O[Si](C)(C)C  
 429.CCCCCCCCC[Si](C)(O[Si](C)(C)C)O[Si](C)(C)C

#### (8) Chain-S

130.CCCCCCCCCS  
 143.CCCCC(S)CC  
 420.CCCCCCCCC(S)CCCCC

#### (9) Chain-S-O

180.CC(=O)S

(10) Chain-Si

70.CCCCCC[Si](C)(CCCCC)CCCCC  
148.CCCCCCCC[Si](C)(CCCCCCCC)CCCCCCCC  
150.CCCCCCCC[Si](C)(CCCCC)CCCCC  
168.C#CC#CC#C[Si](C)(CCCCC)CCCCC  
219.CCC[Si](CCC)CCC  
234.CCCCCC[Si]C  
341.CCCCCCCCC[Si](C)(CCCCCCCC)CCCCCCCC

(11) Bicyclic

389.Cc1ccc2ccc(C)cc2c1  
430.Cc1ccc2cc(C)ccc2c1

(12) Bicyclic-N

152.Cc1ccc2cc(C)ncc2c1  
177.Cc1cccc2nc(C)c(C)nc12  
205.CC1=C2CN(C)C(C)=C2CN1  
240.Cc1ccc2ccncc2c1  
249.Cc1cccc2nn(C)nc12

(13) Bicyclic-N-F

40.Cc1c(F)c(F)cc2nn(C)nc12  
340.Cc1c(F)ccc2nc(C)c(C)nc12

(14) Bicyclic-N-O

50.Cc1ccc2c(c1)C(C)C(=O)N2  
101.Cc1cccc2c1C(=O)N(C)C2=O  
201.CC1=C2C(=O)NC(C)=C2C(=O)N1  
268.Cc1cccc2nonc12  
289.CC1=C2NC(=O)C(C)=C2N(C)C1=O  
424.CC1=C2C(=O)N(C)C(C)=C2C(=O)N1  
426.CCCCN1C(=O)C2=C(C)N(CCCC)C(=O)C2=C1C  
431.Cc1ccc2c(c1)N(C)C(=O)C2C

(15)Bicyclic-N-O-Cl

243.CC1C(=O)Nc2ccc(Cl)cc21  
345.Cc1cc2c(cc1Cl)C(C)C(=O)N2C

(16)Bicyclic-O

277.CC1=C2C(=O)C(C)C(C)=C2C(=O)C1

(17) Bicyclic-S

28.Cc1cc2sc(C=C)cc2s1  
66.Cc1cc2sc(C)cc2s1  
102.Cc1cc2ccsc2s1  
108.Cc1sc2ccsc2c1C  
123.Cc1cc2secc2s1

385.Cc1scc2sc(C)cc12

(18)Bicyclic-S-F

199.Cc1scc2c(F)c(C)sc12

(19) Bicyclic-S-N

18.Cc1cncc2nsnc12

24.Cc1ccc(C)c2nsnc12

49.Cc1c(C#N)c(C#N)cc2nsnc12

82.Cc1cc2c(s1)C(C)CN2

95.Cc1nc2esc(C)c2nc1C

116.Cc1scc2c1N=S=N2

128.Cc1nc2snc2s1

165.Cc1cc2c(s1)C(C)CN2C

169.Cc1scc2nccnc12

224.Cc1cccc2nsnc12

230.Cc1scc2nc(C)c(C)nc12

308.Cc1nccc2nsnc12

330.Cc1cc2nsnc2c(C)c1C

(20) Bicyclic-S-N-F

7.C=Cc1c(F)cc(C)c2nsnc12

122.Cc1c(F)ccc2nsnc12

124.Cc1cc(F)cc2nsnc12

278.Cc1c(F)c(F)cc2nsnc12

(21)Bicyclic-S-N-O

4.Cc1cc2c(s1)C(C)C(=O)N2

190.CC1C(=O)Nc2ccsc21

316.Cc1scc2c1C(=O)N(C)C2=O

338.Cc1cc2c(s1)C(C)C(=O)N2C

(22)Bicyclic-S-O

106.CC1=C2SC(=O)C(C)=C2SC1=O

(23)Fused

12.Cc1cc2c(cc1C)C(C)c1cc3c(cc1-2)C(C)c1cc(C(C)(C)C)c(C)cc1-3

20.Cc1ccc2cc3cc(C)ccc3cc2c1

91.Cc1ccc2c(c1)C(CCCC)(CCCC)c1cc(C)ccc1-2

110.Cc1ccc2c(c1)C(C)(C)c1cc(C)ccc1-2

186.Cc1ccc2c(c1)[Ge](CCCC)(CCCC)c1cc(C)ccc1-2

212.Cc1ccc2c(c1)-c1ccc(C)cc1C2

242.Cc1ccc2c(c1)C(C)(C)c1cc3c(cc1-2)C(C)(C)c1cccc1-3

290.CC1=CC2=C(C1)C1=C(C=C(C)C1)C2

348.Cc1ccc2c(c1)C(C)c1cc3c(cc1-2)C(C)c1cccc1-3

401.CC#Cc1c2ccc(C)cc2c(C)c2ccc(C)cc12

436.CC1c2cccc2-c2cccc21

438.Cc1ccc2c(c1)C(C)(C)c1cccc1-2

(24)Fused-N

16.Cc1ccc2c3ccccc3n(C)c2c1  
64.Cc1ccc2c3ccc(C)cc3n(C)c2c1

(25)Fused-N-F-O

3.Cc1ccc2c(c1F)N(C)C(=O)C2=C1C(=O)N(C)c2c1ccc(C)c2F  
13.Cc1c(F)cc2c(c1F)N(C)C(=O)C2=C1C(=O)Oc2cc3c(cc21)OC(=O)C3=C1C(=O)N(C)c2c(F)cc(F)cc21  
117.Cc1ccc2c(c1F)N(C)C(=O)C2=C1C(=O)Oc2cc3c(cc21)OC(=O)C3=C1C(=O)N(C)c2c(F)cccc21  
145.CC=Cc1ccc2c(c1F)N(C)C(=O)C2=C1C(=O)Oc2cc3c(cc21)OC(=O)C3=C1C(=O)N(C)c2c1ccc(C)c2F  
198.Cc1ccc2c(c1F)N(C)C(=O)C2=C1C(=O)N(C)c2c(F)cccc21  
200.Cc1ccc2c(c1F)N(C)C(=O)C2=C1CN(C)c2c1ccc(C)c2F  
309.Cc1ccc2c(c1F)N(C)C(=O)C2=C1C(=O)N(C)c2ncccc21  
322.Cc1ccc2c(c1F)N(C)C(=O)C2=C1C(=O)N(C)c2nc(C)ccc21

(26) Fused-N-O

10.Cc1cc2c(cc1C)nc1c3ccc4c5c(ccc(c(=O)n21)c35)c(=O)n1c2cc(C)c(C)cc2nc41  
14.Cc1c2nc3c(nc2cc2nc4c(nc12)-c1ccc2c5c(ccc-4c15)C(=O)N(C)C2=O)-c1ccc2c4c(ccc-3c14)C(=O)N(C)C2=O  
22.Cc1ccc2c(c1)N(C)C(=O)C2=C1C(=O)N(C)c2cccc21  
34.CC=Cc1ccc2c(n1)N(C)C(=O)C2=C1C(=O)Oc2cc3c(cc21)OC(=O)C3=C1C(=O)N(C)c2nc(C)ccc21  
37.CC=Cc1ccc2c(c1)N(C)C(=O)C2=C1C(=O)Oc2cc3c(cc21)OC(=O)C3=C1C(=O)N(C)c2cc(C)ccc21  
41.Cc1cc2c3c(c(C#C)cc4c3c1C(=O)N(C)C4=O)C(=O)N(C)C2=O  
53.Cc1cc2nc3c4ccc5c6n(c(=O)c7ccc(c(=O)n3c2cc1C)c4c75)C(C)C(C)N=6  
58.Cc1cc2c3c(ccc4c5c(C)cc6c7c(ccc(c1c34)c75)C(=O)N(C)C6=O)C(=O)NC2=O  
76.Cc1ccc2c(n1)N(C)C(=O)C2=C1C(=O)N(C)c2nc(C)ccc21  
78.Cc1c(=O)n2c3ccc(C)cc3c3cc(=O)n4c5ccc(C)cc5c1c4c32  
115.Cc1cc2c3c(ccc4c5ccc6c7c(ccc(c1c34)c75)C(=O)N(C)C6=O)C(=O)N(C)C2=O  
142.Cc1cc2c3c(cc(C=C)c4c3c1C(=O)N(C)C4=O)C(=O)N(C)C2=O  
153.Cc1cc2nc3c4ccc5c6c(ccc(c(=O)n3c2cc1C)c46)c(=O)n1c2cc(C)c(C)cc2nc51  
175.Cc1ccc2c(c1)N(C)C(=O)C2=C1C(=O)N(C)c2cc3c(cc21)N(C)C(=O)C3=C1C(=O)N(C)c2cc(C)ccc21  
185.Cc1cc2c3c(ccc4c3c1C(=O)N(C)C4=O)C(=O)N(C)C2=O  
203.Cc1ccc2c(c1)N(C)C(=O)C2=C1C(=O)Oc2cc3c(cc21)OC(=O)C3=C1C(=O)N(C)c2cc(C)ccc21  
256.CC#Cc1cc2c3c(ccc4c5c(C)cc6c7c(ccc(c1c34)c75)C(=O)N(C)C6=O)C(=O)N(C)C2=O  
279.Cc1ccc2c(n1)N(C)C(=O)C2=C1C(=O)Oc2cc3c(cc21)OC(=O)C3=C1C(=O)N(C)c2nc(C)ccc21  
301.CN=C1c2ccc3c4c2c(ccc4c(=O)n2c4cc(C)c(C)cc4nc32)C(=O)N1C  
306.Cc1cc2c3c(c(C)cc4c3c1C(=O)N(C)C4=O)C(=O)N(C)C2=O  
324.Cc1ccc2c(c1)c1cc(=O)n3c4ccc(C)cc4c4c(C)c(=O)n2c1c43  
328.Cc1cc2c3c(c(C=C)cc4c3c1C(=O)N(C)C4=O)C(=O)N(C)C2=O  
361.Cc1ccc2c(c1)N(C)C(=O)C2=C1C(=O)Oc2c1ccc1c3c(ccc21)C(C)C(=O)O3  
371.Cc1c(=O)n2c3ccccc3c3cc(=O)n4c5ccccc5c1c4c32  
403.Cc1cc2c3c(c(C)cc4c3c1C(=O)NC4=O)C(=O)N(C)C2=O  
427.Cc1ccc2c(c1)N(C)C(=O)C2=C1C(=O)N(C)c2cc(C)ccc21

(27) Fused-N-O-Cl

354.Cc1ccc2c(c1Cl)N(C)C(=O)C2=C1C(=O)N(C)c2c(Cl)cccc21

(28) Fused-O

127.Cc1ccc2c(c1)C(CO)(CO)c1ccc(C)cc1-2  
226.CC1C(=O)Oc2c1ccc1c3c(ccc21)C(C)C(=O)O3



399.Cc1ccc2c(c1)C(=O)c1cc(C)ccc1-2

(29) Fused-S

5.Cc1cc2c(s1)-c1sc(C)cc1C2  
44.Cc1sc2c(sc3c2sc2c4sccc4sc23)c1C  
61.Cc1sc2c(sc3ccsc32)c1C  
98.Cc1cc2c(s1)c1secc1c1c(C)sc(C)c21  
99.Cc1cc2c(s1)-c1secc1[Ge]2  
140.Cc1cc2cc3sccc3cc2s1  
155.Cc1cc2sc3c4sccc4sc3c2s1  
156.Cc1cc2c(s1)-c1sc(C)cc1[Ge]2  
184.Cc1cc2c(s1)-c1secc1C2  
189.Cc1cc2cc3c(ccc4ccsc43)cc2s1  
211.Cc1cc2sc3ccsc3c2s1  
236.CC1CC2=C(S1)c1secc1C2  
239.Cc1cc2c3sc(C)cc3c3ccsc3c2s1  
263.Cc1cc2cc3c(cc2s1)-c1cc2secc2cc1C3  
269.Cc1cc2ccc3c(ccc4ccsc43)c2s1  
272.Cc1sc2c(sc3c4sccc4sc23)c1C  
280.Cc1cc2c(C)c3secc3c(C)c2s1  
325.Cc1cc2c(s1)C(C)c1cccc1-2  
343.Cc1cc2sc3cc(C#C)sc3c2s1  
350.Cc1cc2c3csc3c3ccsc3c2s1  
374.Cc1cc2c(s1)-c1cc3c(cc1C2)-c1secc1C3  
386.Cc1cc2c3ccsc3c3sc(C)cc3c2s1  
388.Cc1cc2c(C)c3sc(C)cc3c(C)c2s1  
393.Cc1cc2sc3c(c2s1)C(C)(C)c1cc2c(cc1-3)C(C)(C)c1c-2sc2cc(C)sc12  
405.Cc1cc2c(s1)c(C)cc1c3ccsc3c(C)cc21  
409.Cc1cc2ccc3ccsc3c2s1  
416.CC1c2cccc2-c2ccsc21

(30) Fused-S-N

48.Cc1c2c(cc3nsnc13)N=S=N2  
90.Cc1cc2c(s1)-c1sc3c4c(sc3c1CN(C)C2)-c1sc2c3c(sc2c1CN(C)C4)-c1sc2c4c(sc2c1CN(C)C3)-  
c1sc2c3c(sc2c1CN(C)C4)-c1sc(C)cc1CN(C)C3  
100.Cc1ccc2c(c1)Sc1cc(C)ccc1N2  
162.CC(C#N)=C1Sc2c(c3c4c(c5c(c6c4c2CN(C)C6)SC(C)S5)CN(C)C3)S1  
183.Cc1sc2nc3c(nc12)-c1cccc2cccc-3c12  
204.Cc1cc2c(ccc3nsnc32)c2nsnc12  
229.CC1=NC2C(C)c3nsnc3C(C)C2N=C1C  
237.Cc1nc2c(C)c3scnc3c(C)c2s1  
266.Cc1nc2cc3scnc3cc2s1  
271.CC1c2nn(C)nc2C(C)C2N=S=NC12  
284.Cc1c2c(cc3nn(C)nc13)N=S=N2  
299.Cc1cc2c3nsnc3c3ccsc3c2s1  
339.Cc1nc2cc3nsnc3c(C)c2nc1C  
406.CC1=C2N=S=NC2=C(C)C2N=S=NC12  
410.Cc1cc2c3nc4cccc4nc3c3ccsc3c2s1  
414.Cc1ccc2c(c1)N(C)c1ccc(C)cc1S2

423.Cc1cc2c(s1)-c1sc3c4c(sc3c1CN(C)C2)-c1sc2c3c(sc2c1CN(C)C4)-c1sc2c4c(sc2c1CN(C)C3)-c1sc(C)cc1CN(C)C4

(31) Fused-S-N-F

136.Cc1cc2c3nc4cc(F)c(F)cc4nc3c3ccsc3c2s1

(32) Fused-S-N-F-O

60.Cc1cc2c(=O)n(C)c(=O)c3c(F)c(C)sc3c2s1

107.Cc1cc2c(=O)n(C)c(=O)c3c(F)csc3c2s1

347.Cc1sc2c(c1F)c(=O)n(C)c(=O)c1c2sc2c1sc1c3scc(F)c3c(=O)n(C)c(=O)c12

(33) Fused-S-N-O

1.Cc1cc2c(=O)n(C)c(=O)c3cc4c(cc3c2s1)c(=O)n(C)c(=O)c1ccsc14

47.Cc1ccc2c(c1)N(C)C(=O)C2=C1C(=O)N(C)c2cc(C)sc21

112.Cc1cc2c(=O)n(C)c(=O)c3c(sc4c3sc3c5sc(C)cc5c(=O)n(C)c(=O)c34)c2s1

121.Cc1ccc2c(c1)N(C)C(=O)C2=C1CCC(C)S1

262.Cc1cc2sc3c(c2s1)N(C)C(=O)C3=C1C(=O)N(C)c2c1sc1cc(C)sc21

274.Cc1cc2c(=O)n(C)c(=O)c3ncsc3c2s1

281.Cc1cc2c3c4c(c5sccc5c5c4c(c2s1)C(=O)N(C)C5=O)C(=O)N(C)C3=O

288.CC1=CC2C(=O)N(C)C(=O)C3N=C(C)SC3C2S1

351.Cc1cc2c(s1)C1=C3C(=O)N(C)C(C)=C3C(=O)N1C=C2

367.CC1=CC2C(=O)N(C)C(=O)C3C=C(C)SC3C2S1

372.CC1=CC2C(=O)N(C)C(=O)C3c4sc5c(c4SC3C2S1)c(=O)n(C)c(=O)c1cc(C)sc15

376.Cc1cc2c(=O)n(C)c(=O)c3nc(C)sc3c2s1

419.Cc1cc2c(=O)n(C)c(=O)c3cc(C)sc3c2s1

422.Cc1nc2c(=O)n(C)c(=O)c3ccsc3c2s1

(34)Fused-S-N-O-Cl

131.CC1C(=O)Nc2c1cc1sccc1c2Cl

222.Cc1cc2c(Cl)c3c(cc2s1)C(C)C(=O)N3C

(35)Fused-S-O

138.Cc1cc2sc(=O)c3ccsc3c2s1

141.Cc1cc2oc(=O)c3ccsc3c2s1

232.Cc1cc2c(s1)-c1sc(C)cc1C(=O)C(C)C2=O

400.Cc1cc2c(s1)-c1cc3c(cc1C2(C)CC(C#C)CO)-c1sc(C)cc1C3

418.Cc1scc2c1C(=O)c1c(C)sc(C)c1C2=O

(36)Fused-S-Si

296.Cc1cc2c(s1)-c1cc3c(cc1[Si]2)-c1sccc1[Si]3

312.Cc1cc2c(s1)-c1sccc1[Si]2

356.Cc1cc2c(s1)-c1sc(C)cc1[Si]2

378.Cc1csc2c1[Si](C)(C)c1sccc1-2

(37)Fused-Si

52.Cc1ccc2c(c1)[Si](C)(C)c1cc(C)ccc1-2

417.CC1=CC2=C(C1)C1=C(C=C(C)C1)[Si]2

(38)Fused-se

182.Cc1cc2c([se]1)-c1[se]c(C)cc1C2

223.Cc1cc2c([se]1)-c1cc3c(cc1C2(C)CCCCCCCCCCCCCCCCC)-c1[se]c(C)cc1C3

(39)Mono

17.C#Cc1ccc(C)cc1

54.Cc1ccc(C)c(C)c1

79.CC=Cc1ccc(C)cc1

92.Cc1ccc(C)cc1

126.C=Cc1ccc(C)cc1

160.C#Cc1ccc(C)c(C)c1

188.Cc1ccccc1C

218.CC1=CC=C(C)C1(C)C

220.c1ccccc1

247.Cc1ccccc1

286.C=CC1=CC=C(C)C1

342.CC=Cc1ccc(C)c(C)c1

440.CC#Cc1ccc(C)c(C)c1

442.CC1=C(C)C(C)=C(C)CC1

(40)Mono-F

67.Cc1ccccc1F

72.Cc1cc(F)cc(F)c1

151.Cc1ccc(F)c1F

179.C=Cc1ccc(C)cc1F

194.Cc1c(F)c(F)cc(F)c1F

251.Cc1ccc(F)c1

273.Cc1ccc(C=C)cc1F

303.Cc1c(F)cccc1F

412.Cc1cc(F)ccc1F

443.Cc1ccc(C)c(F)c1

(41)Mono-N

8.Cc1ccnc1

30.Cc1cc(C#N)ccc1C#N

38.Cc1cnc(C)c(C)c1

80.CNc1ccc(C)cc1

86.Cc1ccc(C)nc1

359.CNc1ccc(C)cc1C

360.Cc1cccn1

383.Cc1ccc(C)cn1

(42)mono-N-O

103.CC1=CC(C)C(=O)N1C

245.CC1=CC(C)C(=O)N1

(43)Mono-O

45.Cc1ccc(C)o1

81.CC1CCC(C)O1

164.Cc1ccc(O)cc1

202.COc1ccc(OC)c(C)c1  
254.Cc1cc(C)c(C)o1  
302.CC=Cc1ccc(C)c(OC)c1  
358.Cc1ccc1

(44)Mono-S

0.CC1CCC(C)S1  
32.Cc1cc(CCCC)c(C)s1  
36.Cc1esc(C)c1  
135.Cc1cc(C)c(C)s1  
137.Cc1sccc1C  
172.Cc1cccs1  
178.Cc1ccsc1C  
210.Cc1ccc(C)s1  
235.Cc1ccc(C=C)s1  
344.C=Cc1ccc(C)s1  
353.C#Cc1ccc(C)s1  
392.Cc1sc(C=C)cc1C  
408.CC=Cc1ccc(C)s1

(45)Mono-S-Cl

259.Cc1cc(Cl)cs1  
381.Cc1cc(Cl)c(C)s1  
402.Cc1sccc1Cl

(46)Mono-S-F

33.C=Cc1sc(C)cc1F  
83.Cc1cc(F)c(C)s1  
89.Cc1sc(C)c(C)c1F  
187.Cc1cc(F)cs1  
228.Cc1scc(F)c1F  
391.Cc1sccc1F  
394.Cc1sc(C=C)c(F)c1F

(47)Mono-S-N

35.Cc1nc(C)c(C)s1  
85.Cc1sccc1C#N  
88.Cc1nccs1  
132.Cc1nncs1  
133.Cc1ncsc1C  
170.Cc1cc(C#N)c(C=C)s1  
209.Cc1cnc(C)s1  
332.CC=NN=Cc1ccc(C)s1  
415.Cc1cnsc1  
421.Cc1sc(C)c(C)c1C#N

(48)Mono-S-N-O

149.C=C1C(=O)N(CC)C(=S)N(CC)C1=O

(49)Mono-S-O

104.CC1CCS(=O)C1

176.CC(=O)Sc1ccc(C)cc1

275.O=S1CCCC1

(50)Mono-se

63.Cc1ccc(C=C)[se]1

161.Cc1ccc(C)[se]1

207.Cc1ccc[se]1

307.C=Cc1ccc(C)[se]1

362.Cc1cc(C)c(C)[se]1

## 2. Polymer-unit types

### 2.1 Bratch chain combination

'+chain+chain-O+chain-O-Si',  
'+chain+chain-N+chain-O+chain-O-Si',  
'+chain-F+chain-O+chain-O-Si',  
'+chain+chain-O-Si',  
'+chain-O-Si',  
'+chain+chain-O',  
'+chain',  
'+chain+chain-N+chain-O',  
'+chain-O',  
'+chain+chain-N-F+chain-O',  
'+chain-O+chain-S-O',  
'+chain-O-Si+chain-Si',  
'+chain+chain-F',  
'+chain-F-O+chain-O',  
'+chain+chain-Si',  
'+chain-F+chain-O-Si',  
'+chain-O+chain-O-Si',  
'+chain+chain-O-Si+chain-Si',  
'+chain+chain-F+chain-O',  
'+chain-F+chain-O',  
'+chain+chain-F-O+chain-O',  
'+chain-O+chain-S',  
'+chain+chain-O+chain-Si',  
'+chain+chain-O+chain-S'

### 2.2 Mono-ring combination

'+mono-se',  
'+mono-S',  
'+mono-N+mono-S',  
'+mono-S+mono-S-F',  
'+mono-S-N',  
'+mono-S+mono-se',  
'no\_mono',  
'+mono-S-F',  
'+mono-F+mono-S',  
'+mono-O',  
'+mono-S+mono-S-N',  
'+mono+mono-se',  
'+mono-N+mono-S-F',  
'+mono-S-Cl',  
'+mono+mono-S',  
'+mono',  
'+mono-O+mono-S',  
'+mono-N-O+mono-S',  
'+mono+mono-S-O',

'+mono+mono-N',  
'+mono-F',  
'+mono-N',  
'+mono+mono-O',  
'+mono-S+mono-S-O',  
'+mono+mono-N+mono-S-N-O',  
'+mono-N-O+mono-S+mono-se',  
'+mono-S+mono-S-Cl',  
'+mono+mono-S-Cl'

## 2.3 Fused/bicyclic ring combination

'+bicyclic-S-N+fused-N-O',  
'+bicyclic-N-O',  
'+fused+fused-N-O',  
'+bicyclic-N-O+fused+fused-S',  
'+bicyclic-N-O+fused',  
'+fused+fused-S-N-O',  
'+fused+fused-N-F-O',  
'+bicyclic-S-N-F+fused+fused-N-O',  
'+fused+fused-S-N-F-O',  
'+bicyclic-N-O+bicyclic-S',  
'+fused-N-O',  
'+bicyclic-S+fused+fused-N-O',  
'+bicyclic-N-O+fused-S-N',  
'+bicyclic-S-N+bicyclic-S-N-O',  
'+bicyclic-N-O+bicyclic-S-N',  
'+bicyclic-N-O-Cl+fused',  
'+bicyclic-S+fused+fused-S-N',  
'+bicyclic-S-N+fused-S-N-O',  
'+bicyclic-N-O+fused+fused-N-O',  
'+fused+fused-N-O+fused-S',  
'+bicyclic-S-N-F+fused',  
'+bicyclic-S-N-F+bicyclic-S-N-O',  
'+bicyclic-S-N+fused',  
'+fused',  
'+bicyclic-S-O+fused',  
'+fused+fused-S-N',  
'+bicyclic-S-N+bicyclic-S-N-O+fused',  
'+bicyclic-N-O+bicyclic-S+fused-N-O',  
'+fused-S-N-O',  
'+fused+fused-N-O+fused-S-N',  
'+fused-S-N-O-Cl',  
'+fused+fused-N-O+fused-O',  
'+fused+fused-S-O',  
'+fused+fused-S-N-O-Cl',  
'+bicyclic-S-N+fused+fused-S',  
'+fused+fused-S',  
'+bicyclic-N-O+bicyclic-S+fused',

'+bicyclic-N-O+bicyclic-S-N+fused',  
'+bicyclic+bicyclic-S-N-O+fused',  
'+bicyclic-S+bicyclic-S-N-O+fused',  
'+bicyclic-S+fused',  
'+fused+fused-N',  
'+bicyclic-S-N+fused+fused-N',  
'+bicyclic-O+bicyclic-S+fused',  
'+bicyclic-N-F+fused',  
'+bicyclic-S-N-F+fused+fused-S',  
'+bicyclic-S-N+fused+fused-S-Si',  
'+bicyclic-S-N-F+fused+fused-S-Si',  
'+bicyclic-S+bicyclic-S-N-F+fused+fused-S',  
'+bicyclic-N+fused',  
'+bicyclic-S+bicyclic-S-N+fused+fused-S',  
'+bicyclic-S-N+fused+fused-N-O',  
'+bicyclic-S-N+fused+fused-se',  
'+bicyclic-S-N-F+fused+fused-S-Si+fused-Si',  
'+bicyclic-N+bicyclic-N-O+fused',  
'+bicyclic-N',  
'+bicyclic-N+bicyclic-N-O+bicyclic-S+fused',  
'+bicyclic-S+bicyclic-S-O',  
'+bicyclic-S-N',  
'+bicyclic-N-O+fused+fused-N',  
'+bicyclic-S+fused+fused-N-F-O',  
'+bicyclic-N-O+bicyclic-S+fused+fused-S-N-O',  
'+bicyclic-N-O+fused+fused-S-N-O',  
'+bicyclic+bicyclic-N-O+fused',  
'+bicyclic-N-O+fused+fused-S-Si',  
'no\_fused',  
'+bicyclic-S-N+fused+fused-O',  
'+bicyclic-S+fused+fused-S',  
'+bicyclic-N-O+fused-O',  
'+bicyclic-N-O+fused-Si',  
'+bicyclic-N-O+bicyclic-S-N-O+fused+fused-S-N-O',  
'+bicyclic-N+bicyclic-N-O',  
'+bicyclic-N-O+bicyclic-S-N-O+fused',  
'+bicyclic-S-N+fused+fused-S-O',  
'+bicyclic-S-F+fused+fused-S',  
'+fused+fused-S+fused-S-O',  
'+bicyclic-N-F+fused+fused-S',  
'+fused+fused-S+fused-S-N',  
'+fused+fused-S+fused-S-N-F',  
'+bicyclic-S-N-O+fused',  
'+fused+fused-S-Si'