Supplementary Materials

Thermal stability prediction of copolymerized polyimides via an interpretable transfer learning model

Yu Zhang[1], Yating Fang[1], Ling Li[1], Tongle Xu[1], Fang Peng[2], Xiong Li[1], Guangrui Xu[3], Wei Lv[3], Minjie Li[1], Peng Ding[1,*]

[1]Department of Chemistry, College of Sciences, Shanghai University, Shanghai 200444, China.
[2]School of Materials Science and Engineering, Shanghai University, Shanghai 200444, China.
[3]Shanghai Plastics Research Institute Co., Ltd, Shanghai Huayi Holdings Group, Shanghai 200090, China.

[*]**Correspondence to:** Prof. Peng Ding, Department of Chemistry, College of Sciences, Shanghai University, 99 Shangda Road, Baoshan District, Shanghai 200444, China.
E-mail: dingpeng@shu.edu.cn

Supplementary Material

General Information

Article Retrieval

Homo- or copolymerized polyimide films synthesized from dianhydrides and diamines are considered valid candidates for our dataset. We formed a Temperature at 5% weight loss data of 271 data from "Polyimide - Chemistry, Relationship between Structure and Properties and Materials" [1] and 57 papers in different journals. The papers were published between 2014 and 2023 and are available for download in PDF format. [2-44] Supplementary Figures 2 and 3 show that the dataset includes 22 types of dianhydrides and 72 types of diamines in both powder and film forms. Currently, we have considered most types of polyimides. The chemical structural diversity in the dataset helps the model capture a broader range of structure-property relationships, enhancing the model's ability to predict the properties of unknown chemical structures.

Morgan fingerprints with frequency (MFF)[45] for polymer representation

MFF is an extension of Morgan Fingerprints designed to overcome the high dimensionality limitation of vectors 1[46]. In the Morgan algorithm, different atoms (depending on their type and nearest neighbors within a predefined radius) are assigned unique hash identifiers (bit vectors). Thus, each identifier corresponds to a specific substructure. For the benchmark dataset, we can calculate the number and frequency of different identifiers and set a frequency threshold to determine the composition of the MFF vector. In this work, with a radius set to 3, the 22 dianhydrides and 67 diamines molecular structures generated 327 and 897 Morgan Fingerprints with Frequency (MFF), respectively. For example, diamine A70 (Supplementary Figure 2a) has 40 types of substructures, some of which are shown in Supplementary Figure 2b. In Supplementary Figure 2c, the substructures are assigned frequencies (integers) sequentially, which constitutes the MFF of A70.

Experimental section

Raw materials. The s-BPDA (3,3',4,4' -biphenyl tetracarboxylic acid di anhydride), a-BPDA (2,3,4' -biphenyl tetracarboxylic acid di anhydride, analytically pure), PDA, m-PDA, p-phenylenediamine are all purchased from Sigma Aldrich (Shanghai) Trading Co., LTD. N, N-dimethylacetamide (DMAc, chemically pure), xylene is purchased from China National Pharmaceutical Group Co., LTD. The raw materials are not yet purified and need to be dried with molecular sieve before experiment.

Characterizations. The structure and chemical composition of polyimides (PIs) powder was tested by Fourier transform infrared (FTIR, AVATAR370, Nicolet, USA).

The scanning range of the FTIR was 400~4000 cm^{-1}. Thermo-gravimetric analysis (TGA, TA Q500, USA) at a heating rate of 10℃·min−1 in the Nitrogen atmosphere. The temperature range of the TGA was 20~1000 ℃.

Synthesis and structure analysis. The FTIR spectrum of PI-a and PI-b powder was shown in Supplementary Figure 11. The PI powder prepared by terpolymer showed obvious infrared absorption peaks around 1774 cm^{-1}, 1708 cm^{-1}, 1349 cm^{-1} and 734 cm^{-1}, corresponding to the asymmetric C=O stretching vibration peak, symmetric stretching vibration peak, C-N stretching vibration peak and C=O bending vibration peak in the chemical structure of PI respectively. The results are consistent with the previous research, indicating that the imide transformation of PI powder has been basically completed.

Supplementary Tables

Supplementary Table 1. The structural codes of homopolymer and copolymer polyimide films, along with the experimental $T_{5\%}$ (N_2) values and the $T_{5\%}$ values predicted by the CatBoost model (outliers are indicated by "*"), are presented. The corresponding structural codes are depicted in Supplementary Figures 2 and 3. The ratio is expressed in terms of molar ratio.

No.	A1	ratio	A2	ratio	G1	ratio	G2	ratio	$T_{5\%}^{exp}$	$T_{5\%}^{pred}$
1	17	1			12	0.8	6	0.2	556	555
2	17	1			12	0.5	6	0.5	537	541
3	22	1			12	0.8	6	0.2	553	555
4	1	0.5	20	0.5	12	1			553	495
5	1	0.7	20	0.3	12	1			550	495
6	23	0.5	20	0.5	12	1			546	555
7	1	0.5	17	0.5	12	1			551	546
8	22	0.5	17	0.5	12	1			535	495
9			6	1	3	0.2	7	0.8	565	570
10	18	0.01	6	0.99	3	0.2	7	0.8	559	546
11	18	0.025	6	0.975	3	0.2	7	0.8	555	546
12	18	0.05	6	0.95	3	0.2	7	0.8	551	541
13	18	0.075	6	0.925	3	0.2	7	0.8	537	548
14	18	0.1		0.9	3	0.2	7	0.8	535	495
15	24	1			6	1			525	534
16	24	1			2	1			485	508
17	24	1			3	1			526	508
18	24	1			7	1			538	549
19	24	1			5	1			515	508
20	1	1			1	1			559	559
21	1	0.5	14	0.5	1	1			541	552
22	1	0.4	14	0.6	1	1			552	552
23	1	0.3	14	0.7	1	1			556	545
24	1	0.2	14	0.8	1	1			564	560
25			14	1	1	1			558	530
26	25	1			6	1			526.06	517
27	25	1			2	1			522.28	517
28	5	1			3	1			502.45	517
29	1	100			1	1			553	540
30	1	0.05	2	0.95	1	1			565	534
31	1	0.1	2	0.9	1	1			563	557

32	1	0.15	2	0.85	1	1			568	556
33	1	0.2	2	0.8	1	1			569	558
34	3	1			5	1			513	516
35	3	1			5	2			482	533
36	3	1			2	1			516	516
37	3	1			2	2			518	509
38	3	1			3	1			510	505
39	3	1			3	2			511	509
40	5	1			3	1			543	556
41	5	1			6	1			558	556
42	5	1			7	1			566	556
43	5	1			2	1			564	556
44	5	1			1	1			570	556
45	4	1			3	1			584	556
46	4	1			6	1			587	556
47	4	1			7	1			570	556
48	4	1			2	1			568	553
49	4	1			1	1			571	556
50	1	0.8	6	0.2	6	1			564.2	551
51	1	0.5	6	0.5	6	1			566.2	540
52	1	0.2	6	0.8	6	1			576.4	557
53			6	1	6	1			547.5	488
54	7	1			8	1			523	527
55	9	1			1	1			532	545
56	10	1			1	1			523	546
57	26	1			1	1			482	509
58	8	1			4	1			420	408
59	8	1			9	1			352	368
60	8	1			3	1			400	398
61	11	0.2			11	0.08	10	0.92	540	495
62	13	0.2			11	0.08	10	0.92	548	548
63	1	0.2			11	0.08	10	0.92	553	539
64	5	0.2			11	0.08	10	0.92	525	508
65	27	1			1	1			607	585
66	41	1			3	1			501	512
67	41	1			1	1			487	475
68	41	1			2	1			472	477
69	41	1			6	1			491	512

70	41	1			6	2			472	512
71	1	1			7	1			539	495
72	7	1			7	1			559	495
73	7	0.5	1	0.5	7	1			551	559
74	7	1			7	1.3			544	549
75	7	1			7	1.3			578	555
76	7	0.5	1	0.5	7	1.3			569	560
77	1	0.2	7	0.8	12	0.1	7	0.9	590	592
78	1	0.1	7	0.9	12	0.2	7	0.8	602	596
79	1	0.2	7	0.8	12	0.2	7	0.8	582	596
80	15	1			1	1			445	456
81	15	1			7	1			462	456
82	15	1			6	1			461	456
83	15	1			2	1			440	450
84	15	1			3	1			454	456
85	31	1			3	1			453.9	488
86	1	0.5	32	0.5	6	1			438	429
87	21	0.5	32	0.5	6	1			430	429
88	1	0.5	32	0.5	2	1			413	429
89	21	0.5	32	0.5	2	1			421	429
90	1	0.5	32	0.5	7	1			433	444
91	21	0.5	32	0.5	7	1			423	429
92	30	1			2	1			608	533
93	29	1			2	1			500	511
94	28	1			2	1			490	495
95	12	1			6	1			371	347
96	12	1			3	1			346	347
97	12	1			7	1			335	347
98	12	1			2	1			340	347
99			1	1	6	1			478	495
100	12	0.05	1	0.95	6	1			498	495
101	12	0.1	1	0.9	6	1			462	468
102	12	0.2	1	0.8	6	1			438	422
103	12	0.3	1	0.7	6	1			407	399
104	12	0.5	1	0.5	6	1			376	382
105			1	1	3	1			503	495
106	12	0.2	1	0.8	3	1			394	422
107			1	1	2	1			498	503

108	12	0.2	1	0.8	2	1		414	422
109			1	1	2	1		473	495
110	12	0.2	1	0.8	2	1		425	419
111	33	1			15	1		425	422
112	34	1			15	1		424	419
113	1	1			15	1		439	426
114	38	1			15	1		442	455
115	37	1			15	1		445	431
116	36	1			15	1		433	429
117	35	1			15	1		431	447
118	39	1			1	1		483	477
119	39	1			7	1		493	477
120	39	1			6	1		491	477
121	39	1			2	1		477	475
122	39	1			3	1		489	481
123	40	1			5	1		443.8	475
124	40	1			6	1		490.3	475
125	40	1			2	1		479.1	475
126	40	1			5	2		463.1	472
127	40	1			6	2		464.2	472
128	40	1			2	2		489.4	472
129			1	1	1	1		508	499
130	16	0.1	1	0.9	1	1		514	512
131	16	0.3	1	0.7	1	1		525	547
132	16	0.5	1	0.5	1	1		528	533
133	16	0.7	1	0.3	1	1		538	508
134	16	0.9	1	0.1	1	1		544	548
135	41				1	1		530	512
136	41				7	1		533	512
137	41				2	1		514	497
138	41				6	1		509	512
139	41				3	1		523	512
140	42				3	1		507	493
141	43				3	1		471	481
142	45				3	1		484	503
143	46				3	1		489	481
144	1	1			1	1		548	495
145	1	0.75	19	0.25	1	1		535	552

146	1	0.8	19	0.2	1	1		549	549
147	44	1			3	1		467	478
148	47	1			3	1		438	459
149	48	1			2	1		510	511
150	48	1			6	1		544	555
151	48	1			3	1		500	511
152	21	1			2	1		511	495
153	21	1			6	1		513	495
154	21	1			3	1		490	495
155	50	1			14	1		400	387
156	49	1			14	1		400	398
157	50	1			1	1		445	495
158	50	1			2	1		550	495
159	49	1			2	1		535	539
160	50	1			6	1		540	495
161	50	1			7	1		490	495
162	50	1			13	1		445	495
163	49	1			13	1		395	495
164	54	1			10	3		511	515
165	54	1			18	3		480	505
166	54	1			5	3		533	512
167	54	1			19	3		517	505
168	54	1			20	3		510	516
169	51	1			6	1		440	488
170	52	1			16	1		435	434
171	53	1			6	1		406	433
172	51	1			16	1		416	421
173	52	1			6	1		428	495
174	53	1			16	1		424	410
175	33	1			17	1		375	366
176	34	1			17	1		374	366
177	35	1			17	1		390	395
178	36	1			17	1		389	392
179	55	1			17	1		384	392
180	1	1			17	1		379	388
181	11	1			17	1		387	388
182	56	1			1	1		475	456
183	56	1			7	1		457	456

184	56	1	2	1	451	456
185	56	1	6	1	445	495
186	56	1	3	1	455	456
187	37	1	3	1	522	500
188	63	1	22	1	323	352
189*	65	1	22	1	286	-
190	64	1	22	1	382	352
191	66	1	22	1	376	352
192	21	1	5	2	440	456
193	21	1	5	1.33	442	441
194	21	1	5	1.14	450	459
195	21	1	5	1	452	495
196	57	1	1	1	465	461
197	57	1	7	1	479	461
198	57	1	6	1	449	461
199	57	1	3	1	486	461
200	57	1	2	1	477	461
201	57	1	6	1	473	461
202	57	1	5	1	464	461
203	42	1	1	1	530	520
204	42	1	7	1	533	505
205	42	1	2	1	514	535
206	42	1	6	1	509	512
207	42	1	3	1	523	512
208	43	1	3	1	507	493
209	45	1	3	1	471	481
210	46	1	3	1	484	503
211	44	1	3	1	489	477
212	26	1	1	1	541	533
213	26	1	6	1	524	520
214	26	1	2	1	542	533
215	26	1	3	1	534	533
216	26	1	7	1	546	533
217	58	1	1	1	514	520
218	58	1	6	1	505	520
219	58	1	2	1	519	520
220	58	1	3	1	524	533
221	58	1	7	1	530	512

222	57	1			1	1	445	456
223	57	1			7	1	454	461
224	57	1			3	1	463	461
225	57	1			2	1	439	461
226	57	1			6	1	450	461
227	57	1			5	1	447	461
228	16	1			1	1	544	511
229	16	1			7	1	609	548
230	16	1			6	1	549	548
231	16	1			1	1	515	548
232	68	1			3	1	497	493
233	68	0.7	69	0.3	3	1	514	512
234	68	0.5	69	0.5	3	1	512	515
235	68	0.3	69	0.7	3	1	511	505
236			69	1	3	1	514	525
237			1	1	3	1	559	495
238	59	0.1	1	0.9	3	1.06	557	536
239	59	0.21	1	0.79	3	1.13	508	495
240	59	0.34	1	0.66	3	1.2	488	492
241	59	0.48	1	0.52	3	1.29	455	475
242	59	0.65	1	0.35	3	1.39	447	457
243	59	0.85	1	0.15	3	1.5	434	439
244	59	1			3	1.6	432	424
245	62	1			6	1	477	477
246	62	1			21	1	465	475
247	60	1			6	1	471	475
248	60	1			21	1	453	475
249	61	1			6	1	487	512
250	61	1			21	1	465	475
251	48	1			2	1	510	511
252	48	1			6	1	544	511
253	48	1			3	1	500	516
254	21	1			2	1	511	495
255	21	1			6	1	513	495
256	21	1			3	1	490	495
257	50	1			14	1	400	398
258	49	1			14	1	400	398
259	50	1			1	1	445	495

260	50	1	7	1	550	543
261	49	1	7	1	535	539
262	50	1	2	1	540	544
263	50	1	6	1	490	495
264	50	1	13	1	445	461
265	49	1	13	1	395	495
266	67	1	6	1	442	459
267	67	1	1	1	456	459
268	67	1	7	1	444	459
269	67	1	2	1	438	488
270	67	1	3	1	450	492

Supplementary Table 2. The structural codes of homopolymer and copolymer polyimide powders, along with the experimental $T_{5\%}$ (N_2) values and the $T_{5\%}$ values predicted by the CatBoost model, are presented. The corresponding structural codes are depicted in Supplementary Figures 2 and 3. The ratio is expressed in terms of molar ratio.

No.	A1	ratio	A2	ratio	G1	ratio	G2	ratio	$T_{5\%}^{exp}$	$T_{5\%}^{pred}$
1	1	1			1	1			589	590
3	1	0.5	70	0.5	1	1			595	595
4	1	0.3	70	0.7	1	1			604	604
5	1	0.7	70	0.3	1	1			589	590
6	71	1			1	1			595	592
7	1	0.5	71	0.5	1	1			590	590
8	1	0.3	71	0.7	1	1			592	592
9	1	0.7	71	0.3	1	1			590	590
10	7	0.3	71	0.7	1	1			603	604
11	7	0.5	71	0.5	1	1			605	605
12	7	0.7	71	0.3	1	1			608	608
13	7	0.5	70	0.5	1	1			607	607
14	7	0.7	70	0.3	1	1			608	607
15	7	0.3	70	0.7	1	1			609	607
16	7	0.3	1	0.7	1	1			608	607
17	7	0.7	1	0.3	1	1			607	609
18	7	0.5	1	0.5	1	1			608	608
19	7	1			7	1			610	611
32	7	0.97	72	0.03	7	1			616	615
33	7	0.95	72	0.05	7	1			616	615
34	7	0.99	72	0.01	7	1			616	615
20	7	1			7	0.93	12	0.07	610	610
21	7	1			7	0.9	12	0.1	610	610
23	7	1			7	0.85	12	0.15	612	611
24	7	1			7	0.7	12	0.3	610	612
25	7	1			7	0.6	12	0.4	610	611
26	7	1			7	0.5	12	0.5	610	611
27	7	1			7	0.4	12	0.6	608	607

28	7	1			7	0.3	12	0.7	608	609
29	7	0.5	23	0.5	7	1			608	608
30	7	0.3	23	0.7	7	1			608	607
35	70	1			1	0.7	12	0.3	598	600
36	70	1			1	0.3	12	0.7	591	592
37	70	1			7	1			604	604
38	70	0.5	7	0.5	7	1			602	603
39	70	0.3	7	0.7	7	1			603	603
40	70	0.7	7	0.3	7	1			602	602
41	70	0.3	23	0.7	7	1			589	590
42	70	0.7	23	0.3	7	1			597	598
43	70	0.5	23	0.5	7	1			592	593
44	71				7	1			585	586
45	70				7	1			604	604

Supplementary Table 3. Four descriptor subsets generated by Ridge, RF, XGBoost and CatBoost algorithms: Ridge (18 descriptors), RF (12 descriptors), XGBoost (18 descriptors) and CatBoost (9 descriptors). Descriptors filtered by any three algorithms are highlighted in bold.

algorithm	number	Descriptors
XGBoost	18	copolymer, **A3431531268**, **A864662311**, G2223687999, **G2976033787,** A2245273601, A1063794729, A1446633534, **A3975295864**, A4256002933, A847957139, **A951226070**, A3372254436, A269959362, A1538830519, A2599598242, A997097697, **A4216335232**
CatBoost	9	Copolymer, **A864662311**, A3217380708, **A951226070, G2976033787, A3975295864, A3431531268**, G994485099, **A4216335232**
RF	12	**A864662311**, **A3975295864**, A3217380708, **A951226070, G2976033787**, A3372254436, G4290259127, copolymer, **A3431531268, A4216335232**, A1083852209, G3559852389
Ridge	18	A2439737417, A4256002933, G2960894223, A644322065, G3559852389, A1040949262, G2117068077, A1101907775, A3957411950, A2245384272, A4121755354, A174348455, G275607895, A1063794729, A1992087727, A2245273601, A10565946, G229241619

Supplementary Table 4. Ranking from 10 to 3 descriptors obtained by RFE-CatBoost method. 6 descriptors are the optimal subset.

No.	**Descriptors**
10	copolymer, A2720313463, A1637023899, A668520641, A2041434490, A1295649264, A2084364935, A951226070, G3692055567, G3276511768
9	copolymer, A2720313463, A1637023899, A668520641, A2041434490, A2084364935, A951226070, G3692055567, G3276511768
8	copolymer, A2720313463, A1637023899, A2041434490, A2084364935, A951226070, G3692055567, G3276511768
7	A2720313463, A1637023899, A2041434490, A2084364935, A951226070, G3692055567, G3276511768
6	**A2720313463, A2041434490, A2084364935, A951226070, G3692055567, G3276511768**
5	A2720313463, A2041434490, A2084364935, A951226070, G3276511768
4	A2041434490, A2084364935, A951226070, G3276511768
3	A2041434490, A2084364935, A951226070

Supplementary Table 5. Table of optimal hyperparameters for model-A, model-B, model-C and mode-D.

hyper-parameter	Model-A	Model-B	Model-C	Model-D
Learning rate	0.3	0.3	0.1	0.3
subsample	0.7	0.7	0.7	0.9
depth	3	3	3	7
iterations	100	100	100	100

Supplementary Table 6. Table of coefficient of determination (R^2) and Root Mean Square Error (RMSE) results across 20 random splits into training and test sets: assessing the stability and generalization ability of model-A.

	train		test	
Datasets1	RMSE	R^2	RMSE	R^2
1	21.91	0.85	26.54	0.81
2	20.90	0.87	30.35	0.62
3	22.63	0.85	21.75	0.82
4	22.42	0.85	26.80	0.78
5	21.20	0.86	29.69	0.76
6	20.00	0.88	35.16	0.68
7	22.39	0.86	29.60	0.59
8	22.37	0.85	23.96	0.84
9	22.87	0.84	19.51	0.88
10	21.14	0.86	31.85	0.71
11	21.35	0.86	31.55	0.68
12	21.78	0.86	23.50	0.79
13	22.40	0.86	26.93	0.63
14	20.37	0.87	33.29	0.71
15	21.88	0.86	28.29	0.72
16	21.78	0.86	24.26	0.81
17	21.25	0.86	29.46	0.72
18	19.71	0.87	36.20	0.74
19	22.56	0.84	24.27	0.86
20	21.00	0.86	35.03	0.67
μ	21.60	0.86	28.40	0.74

Supplementary Table 7. External validation set (not in the numerical set of Supplementary Table 1). The structural codes of copolymer polyimide films, along with the experimental $T_{5\%}$ (N_2)[47] values and the $T_{5\%}$ values predicted by the CatBoost model. The corresponding structural codes are depicted in Supplementary Figures 2 and 3. The ratio is expressed in terms of molar ratio.

No.	A1	ratio	A2	ratio	G1	ratio	G2	ratio	$T_{5\%}^{exp}$	$T_{5\%}^{pred}$	*RMSE*
1	73	0.25	4	0.75	1	1			500	509	19.19
3	73	0.17	4	0.83	1	1			522	523	
4	73	0.125	4	0.875	1	1			532	517	
5	73	0.1	4	0.9	1	1			551	531	
6	73	0.08	4	0.92	1	1			543	536	
7	73	0.07	4	0.93	1	1			562	536	

Supplementary Table 8. The 9 descriptors from model-A were completely transferred to model-B, with the descriptors ranked from 9 to 1.

No.	Descriptors
9	copolymer, A951226070, A3431531268, A864662311, A3975295864, A4216335232, A3217380708, G994485099, G2976033787
8	copolymer, A951226070, A864662311, A3975295864, A4216335232, A3217380708, G994485099, G2976033787
7	copolymer, A951226070, A3975295864, A4216335232, A3217380708, G994485099, G2976033787
6	copolymer, A951226070, A4216335232, A3217380708, G994485099, G2976033787
5	copolymer, A951226070, A4216335232, A3217380708, G994485099
4	A951226070, A4216335232, A3217380708, G994485099
3	A951226070, A4216335232, A3217380708
2	**A951226070, A4216335232**
1	A4216335232

In model-C, A951226070 is selected, and thus A951226070 and A4216335232 are migrated at the same time during transfer learning.

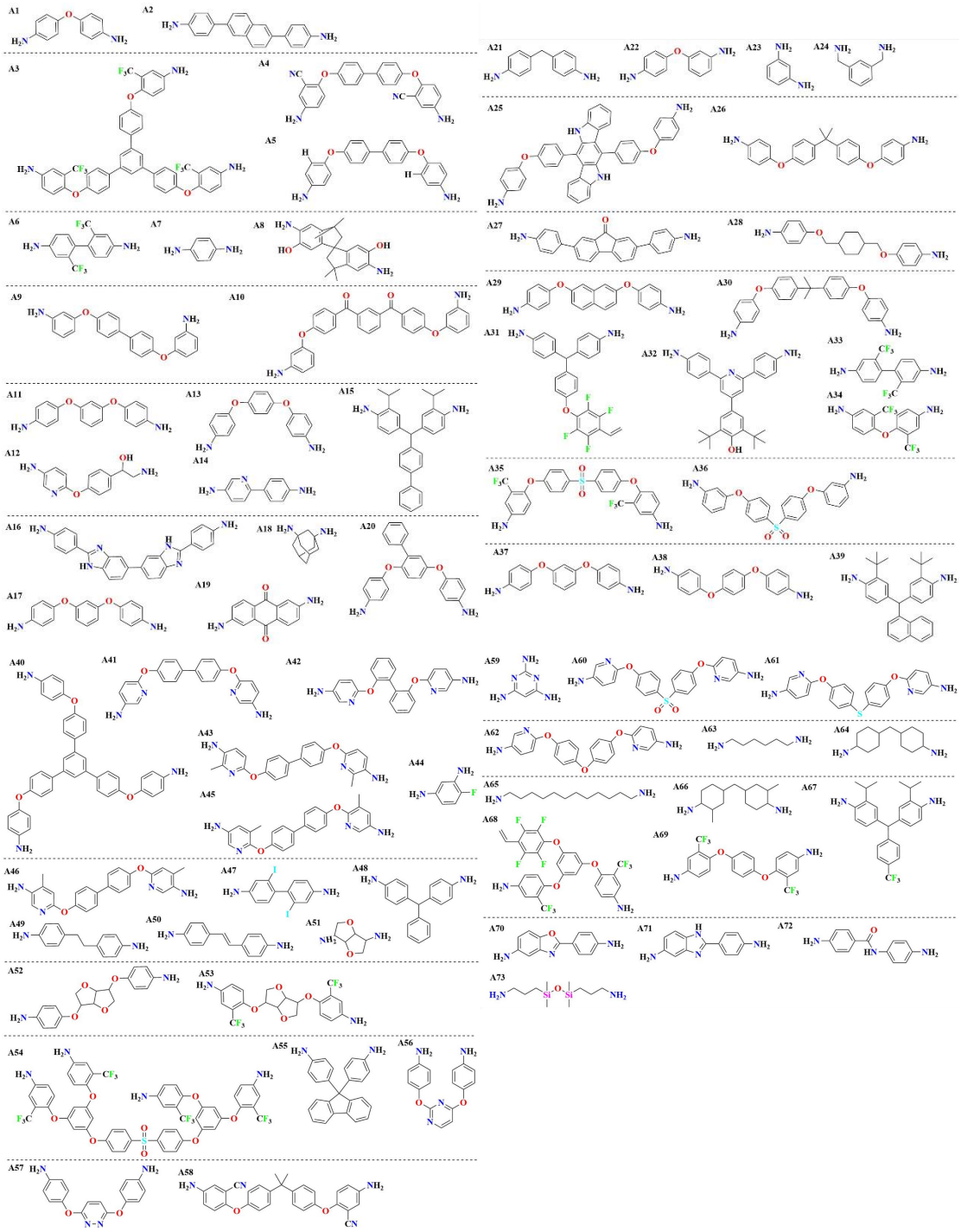
Supplementary Table 9. Table of R^2 and RMSE results across 6 random splits into training and test sets: assessing the stability and generalization ability of model-D.

	train		test	
Datasets2	RMSE	R^2	RMSE	R^2
1	0.06	0.99	2.31	0.91
2	0.06	0.99	1.91	0.91
3	0.05	0.99	1.89	0.93
4	0.07	0.99	3.74	0.80
5	0.07	0.99	1.84	0.89
6	0.08	0.99	2.96	0.86
μ	0.07	0.99	2.44	0.88

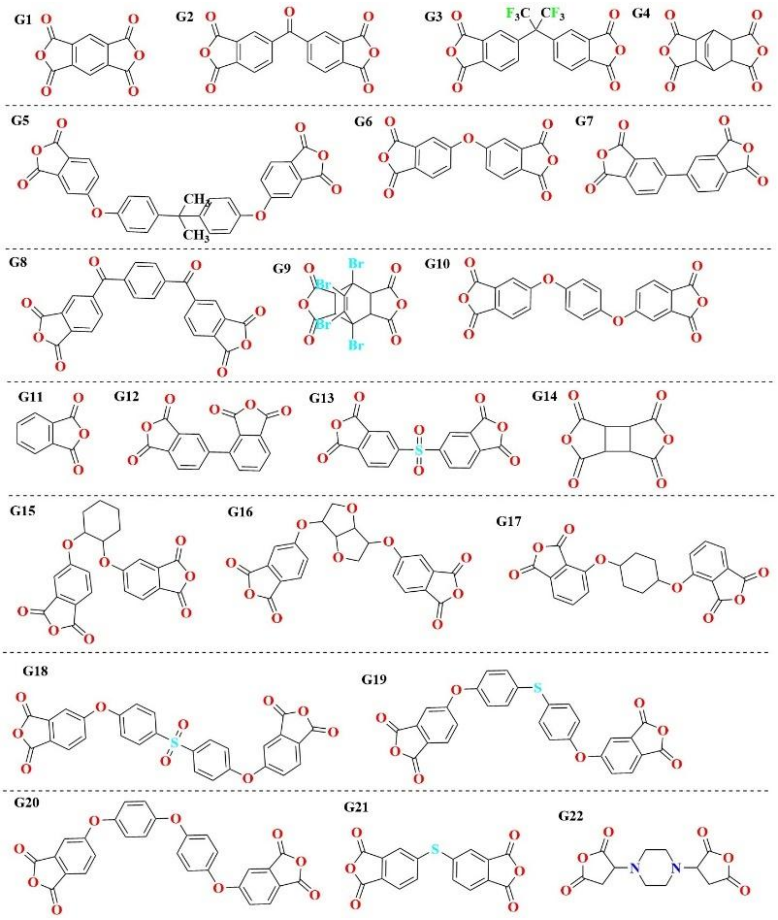
Supplementary Table 10. Statistical results of PI powder thermogravimetric analysis.

Sample name	Nitrogen atmosphere		
	$T_{5\%}$/°C	$T_{10\%}$/°C	1000°C residual /%
PI-1	600	620	60.4
PI-2	606	623	57.1

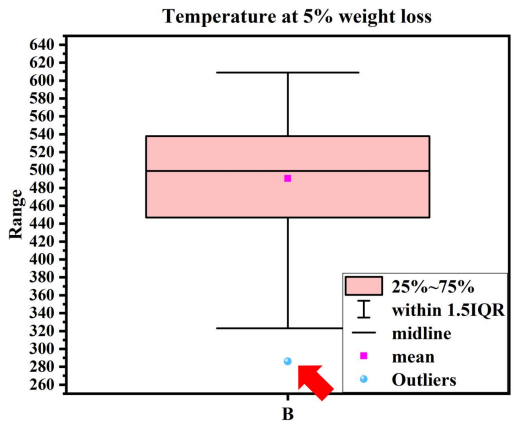
Supplementary Figures



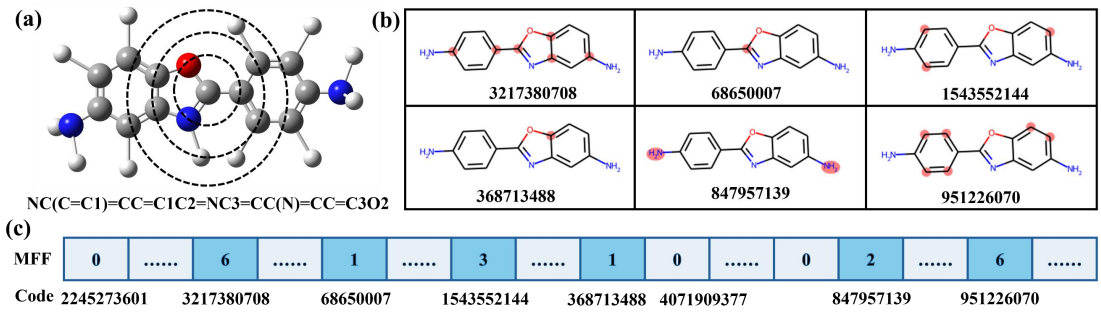
Supplementary Figure 1. Molecular structure diagrams of diamines.



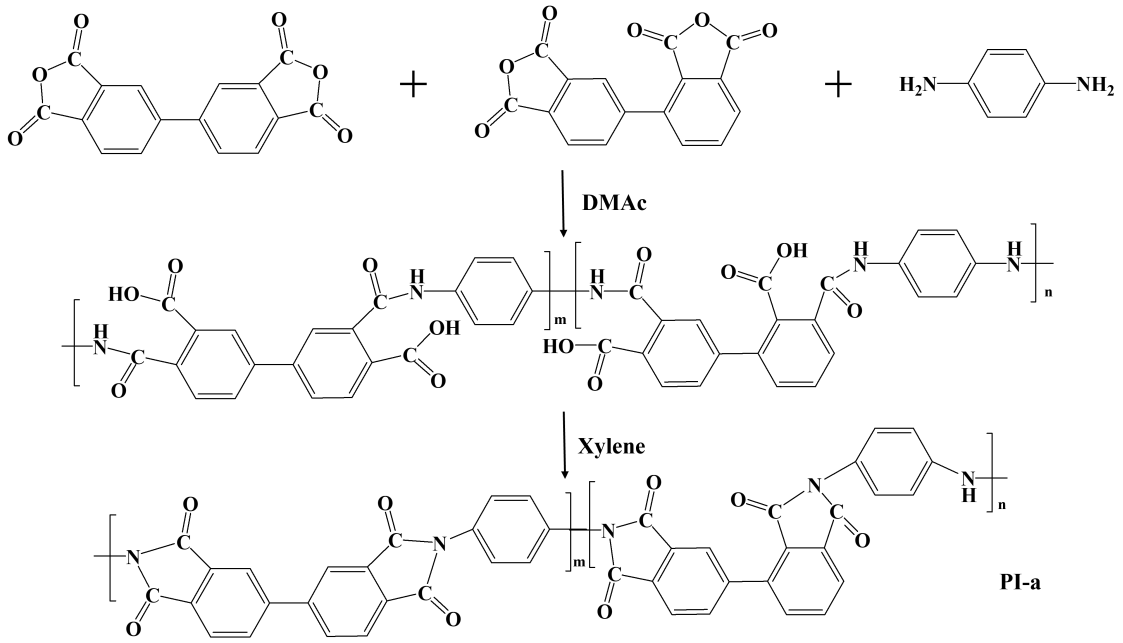
Supplementary Figure 2. Molecular structure diagrams of dianhydrides.



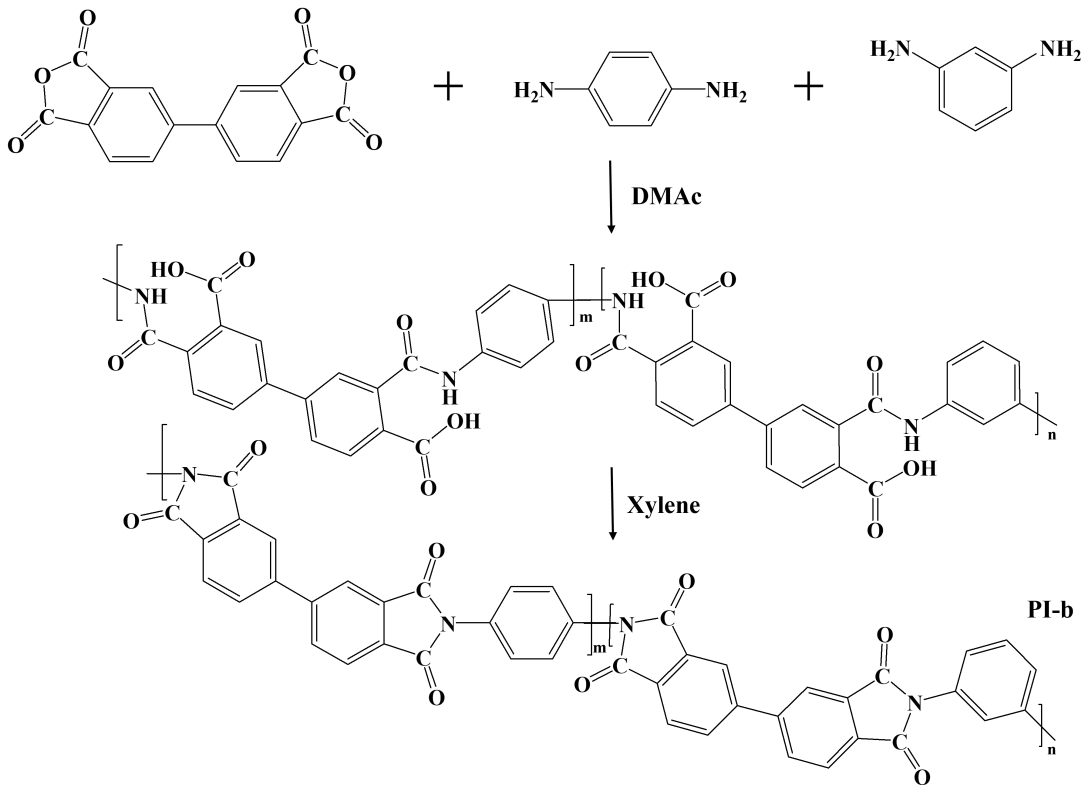
Supplementary Figure 3. Box plots of polyimides Temperature at 5% weight loss. The bins usually represent the first quartile (Q1) and the third quartile (Q3) of the data, which are the bottom and top of the bin. The line in the middle of the box represents the median (Q2), which is the center point of the data. The 1.5 IQR rule means whiskers extend to the maximum and minimum values, but not beyond Q1 - 1.5IQR and Q3 + 1.5IQR (IQR is the interquartile range, i.e. Q3-Q1). Data points outside this range are considered outliers. The calculation of whiskers can also be based on the standard deviation of the data, such that the maximum and minimum values do not exceed the mean plus or minus 2 or 3 standard deviations. Points outside the above range are usually considered outliers and marked as points in the boxplot.



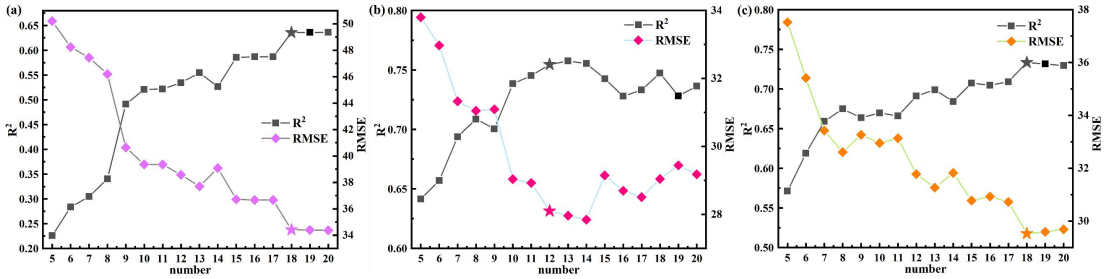
Supplementary Figure 4. Example of MFF generation. (a) Diamine with ID of A70. (b) Part of the substructure of diamine A70. (c) MFF vector.



Supplementary Figure 5. Synthesis route of the PI-a.



Supplementary Figure 6. Synthesis route of the PI-b.

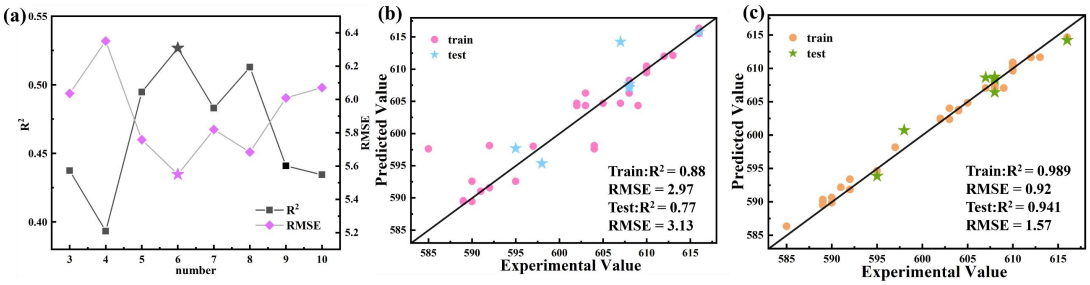


Supplementary Figure 7. Descriptor sniping. The R^2 and RMSE changes of models (5-fold-CV) built with RFE nested (a) Ridge, (b)RF, (c) XGBoost hitting different numbers of descriptors were accurately described.

A higher R^2 and a lower RMSE indicate better model performance, making it the most suitable choice for the optimal subset.

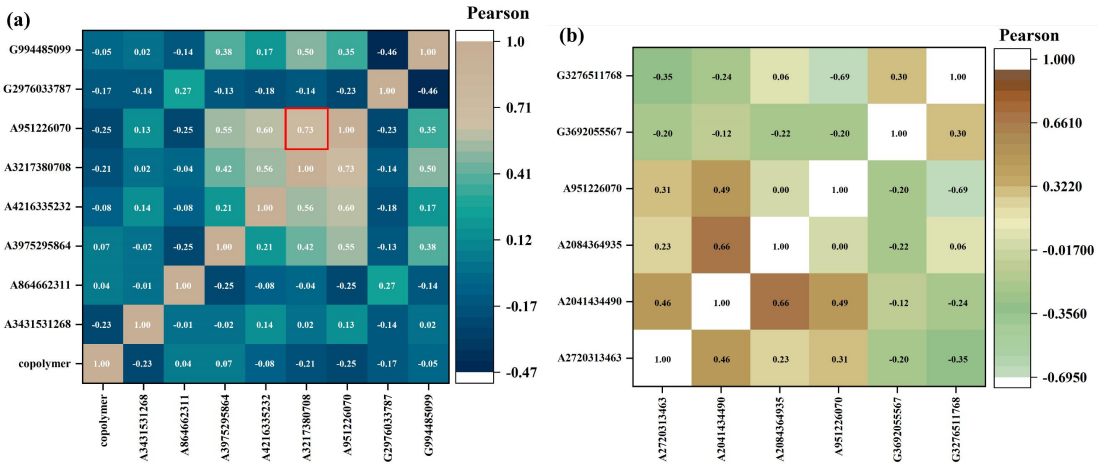
According to Supplementary Figure 7a, the model performance consistently improves with an increasing number of descriptors, showing no inflection point (R^2 increases, RMSE decreases). At 20 descriptors, the R^2 is 0.64, and the RMSE is 34.36. However, in feature selection process, four models were compared. The other three models demonstrating superior performance. The RFE-CatBoost model (Figure 3a) identified the optimal subset with 9 descriptors, achieving an R^2 of 0.75 and RMSE of 28.84. Similarly, the RFE-RF model (Supplementary Figure 7b) found the optimal subset at 12 descriptors, with an R^2 of 0.75 and RMSE of 28.10. The RFE-XGBoost model selected 18 descriptors, yielding an R^2 of 0.73 and RMSE of 29.69. A good descriptor must be simpler to obtain than the target property and of as low dimensionality as possible[48]. Therefore, even though Supplementary Figure 7a shows no inflection point, we decided not to further increase the number of descriptors to achieve better model results.

In addition, based on Supplementary Figure7b, the model using the 13-feature subset and 12-feature subset achieved R^2 values of 0.76 and 0.75, and RMSE values of 27.96 and 28.10, respectively. Although the subset of 13 descriptors showed the highest R^2 value and the lowest RMSE value in terms of predictive performance, slightly better than the subset of 12 descriptors, we still opted a smaller number of features to build the model. Because we follow the "Occam's razor"[49] when choosing the optimal feature set, that is, favoring a simpler and more interpretable model while achieving sufficient predictive accuracy.



Supplementary Figure 8. Descriptor sniping and best models. (a) The R^2 and RMSE changes of models (5-fold-CV) built with RFE nested CatBoost hitting different numbers of descriptors were accurately described for model-C. Scatter plots of the best model to predict the $T_{5\%}$ values and their experimental values for the train and test sets for (b) model-B and (c) model-C.

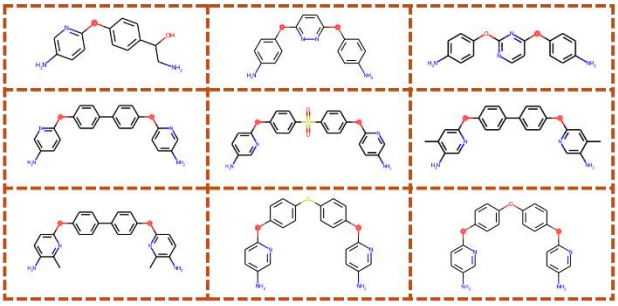
For the model-C (Fig. S8 a and b), the R^2 is 0.989 and the RMSE is 0.92. In the independent test set, R^2 of 0.941 and an RMSE of 1.57 are recorded.



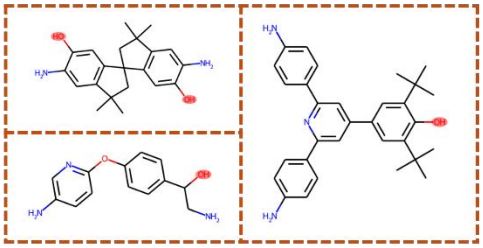
Supplementary Figure 9. The correlation coefficient heatmap for the descriptors selected by the RFE-CatBoost model, comprising(a) 9 descriptors for model-A and (b) 6 descriptors for model-C.

The successful removal of redundant and unrelated descriptors can be observed from the correlation heatmap of descriptors filtered by RFE-CatBoost. Notably, A951226070 demonstrates a relatively strong positive correlation with A3217380708. However, the correlations between these pairs are relatively weak, indicating their "global" representativeness.

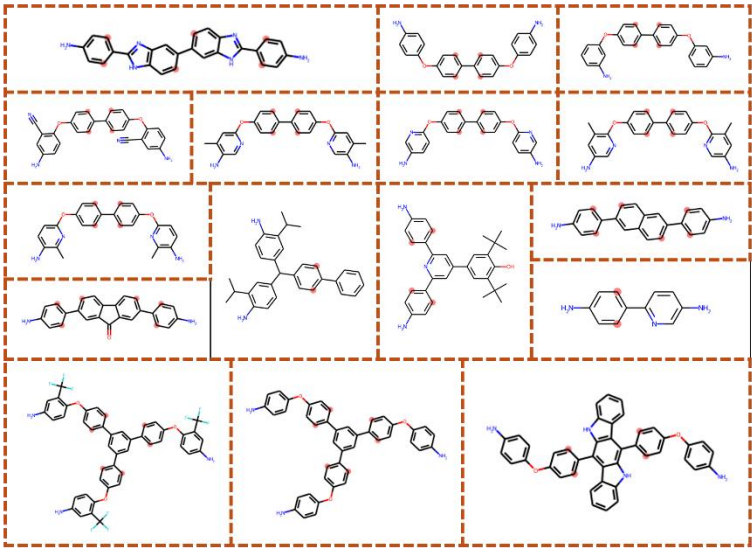
A3431531268



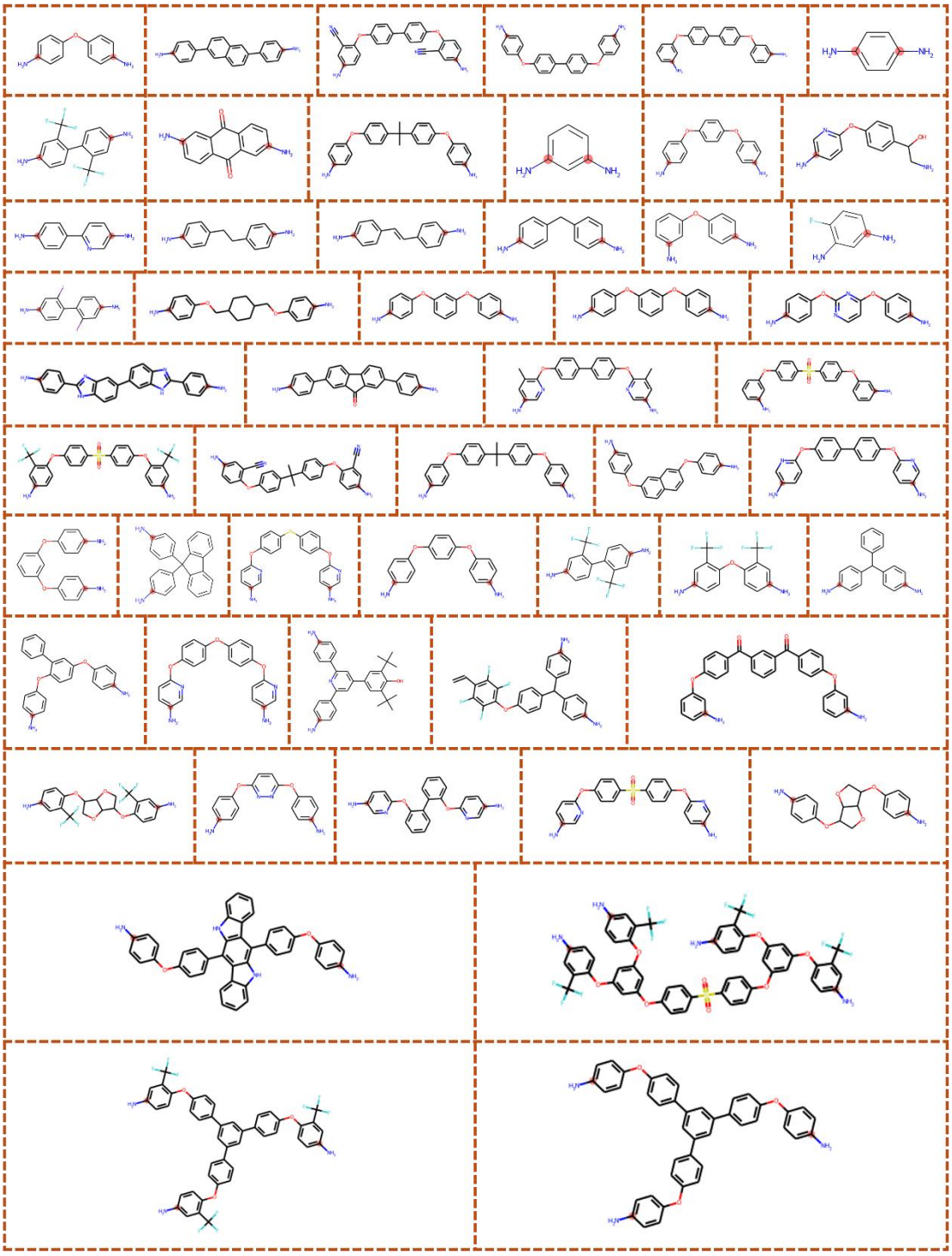
A864662311

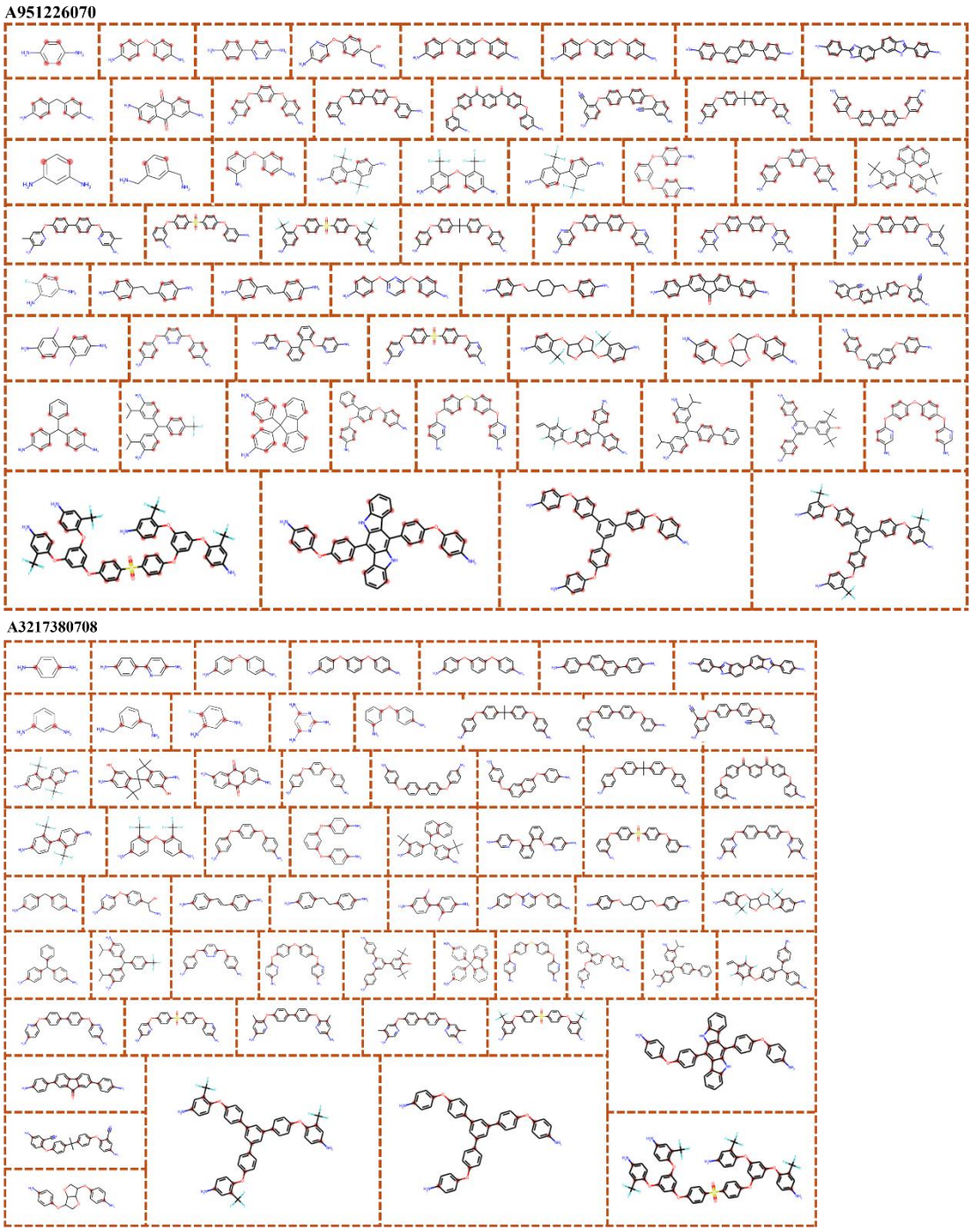


A4216335232

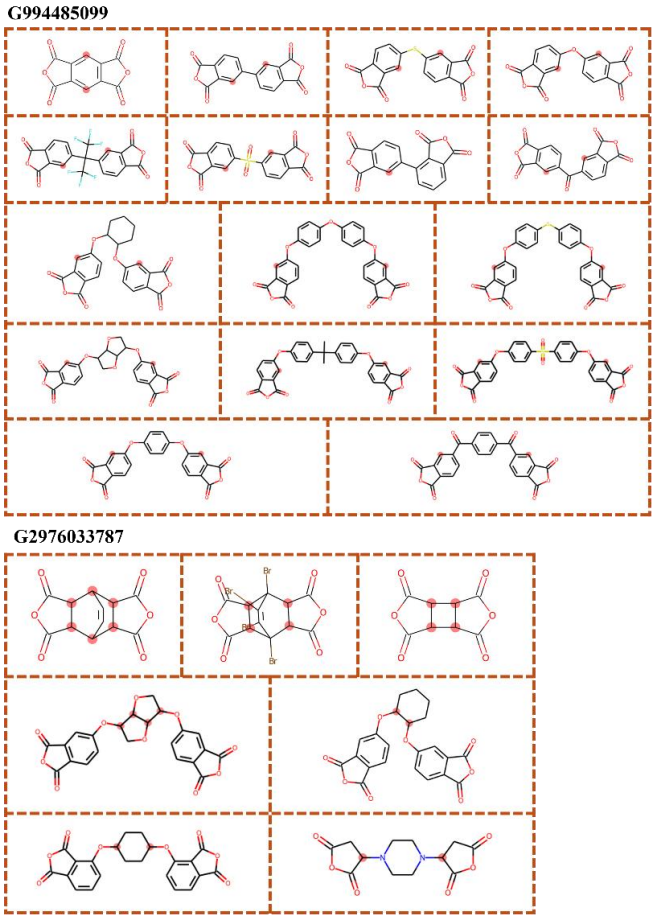


A3975295864

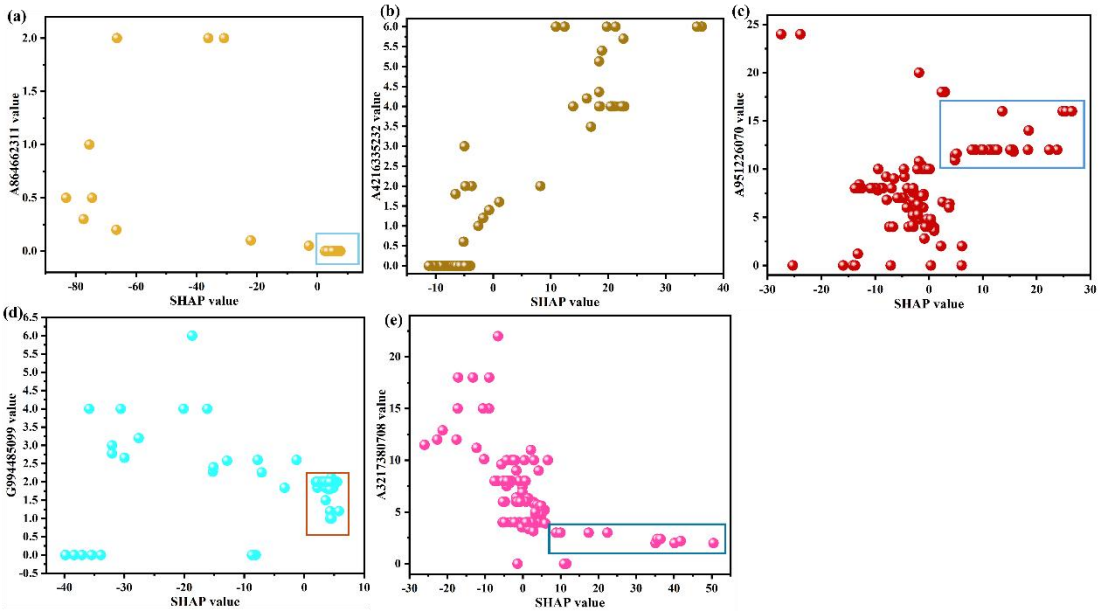




Supplementary Figure 10. Highlighted image of 6 descriptors of diamine molecular structure.



Supplementary Figure 11. Highlighted image of 2 descriptors of the dianhydride molecular structure.



Supplementary Figure 12. For the remaining five descriptors of model-A (a) A864662311 (b) A4216335232 (c) A951226070 (d) G994485099 (e) A3217380708 the relationship between the descriptor value and the SHAP value output by the SHAP nested CatBoost algorithm.

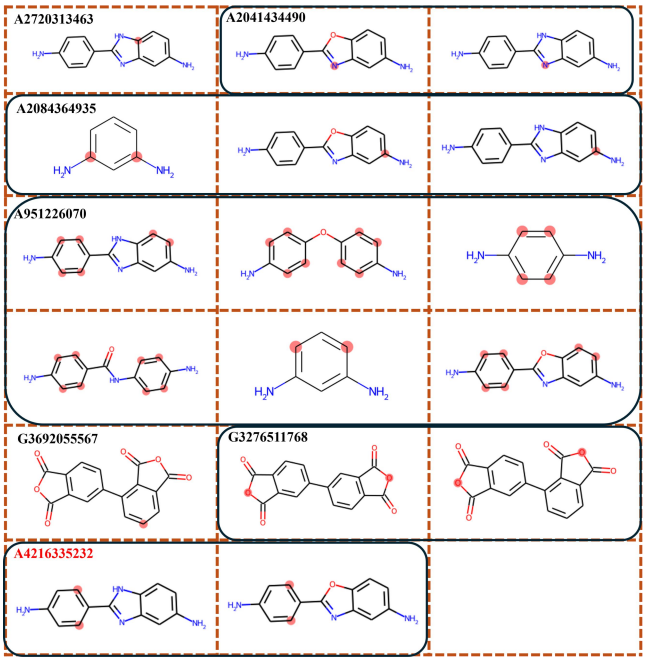
A864662311 refers to hydroxyl groups. Supplementary Figure 10 highlights diamine molecules containing the A864662311 structure. In Supplementary Figure 9a, the selected region within the blue box demonstrates that as A864662311 approaches zero, the higher SHAP values increasingly favor $T_{5\%}$. Conversely, PI-60 and PI-96, which are synthesized from the same anhydride G3 and amines A8 and A12, exhibit low $T_{5\%}$. Additionally, as depicted in Supplementary Table 1, PI-106 shows an enhanced $T_{5\%}$ when blended with a high proportion of A1 (A864662311 = 0). Hydroxyl groups in side-chain aliphatic hydrocarbons facilitate intermolecular cross-linking through condensation and secondary weak cross-linking via hydrogen bonds, which subsequently increase solubility[22] and reduce thermal stability[27]. Consequently, hydroxyl groups are deemed unsuitable for incorporation into diamine structures for improved $T_{5\%}$.

A4216335232 represents strong π-π interactions involving linked benzene, naphthalene, and pyridine rings, as illustrated in Supplementary Figure 10. Supplementary Figure 12b presents a noticeable positive correlation for A3975295864, where an increase in its value corresponds to enhanced SHAP values, mirroring the behavior of A3975295864. Stronger intermolecular π-π interactions[6] promote denser molecular packing, resulting in ordered arrangements that restrict internal rotation and enhance structural rigidity, thereby boosting thermal stability. In Supplementary Table 1, PI-41, produced using the same anhydride G6 and amine A5, shows a 7% higher $T_{5\%}$ compared to PI-213, which is synthesized using A26. Additionally, PI-40, PI-215,

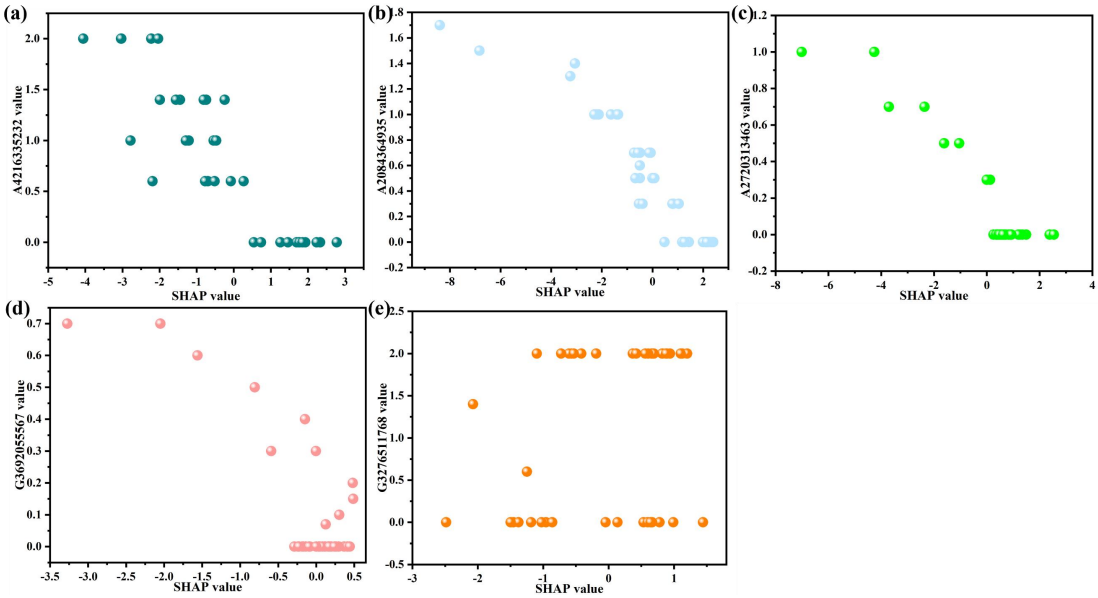
PI-42, and PI-216 demonstrate similar trends. Therefore, conjugated functional groups are advantageous for inclusion in diamine structures.

As depicted in Supplementary Figure 10, A951226070 denotes the number of unsubstituted hydrogens on the benzene ring (vacant sites), while A3217380708 represents the number of substituted hydrogens on the benzene ring. The heatmap in Fig. S4 reveals a positive Pearson correlation coefficient of 0.73 between them. The selected regions within the blue boxes in Supplementary Figure 12c and e indicate that A951226070 values between 10-15 and A3217380708 values between 2-3 yield positive SHAP values, favoring $T_{5\%}$. Excessive substitution of hydrogen atoms on the benzene ring leads to increased side chains, irregular PI molecular chains, spatial hindrance, low molecular packing density, and decreased thermal stability.[6] In Supplementary Table 1, for the same anhydride G1 and amine A1 (A951226070 = 8, A3217380708 = 4), the inclusion of A16 (A951226070 = 12, A3217380708 = 12) results in PI-129 to PI-134, featuring benzene rings, thereby increasing $T_{5\%}$. Additionally, excessively high values of A951226070 and A3217380708 lead to low molecular packing density, loose molecules, and reduced $T_{5\%}$, as observed in PI-126 to PI-128 and PI-42 to PI-216, none of which exceed 490°C. Thus, controlling the frequency of side-chain appearances on the benzene ring in diamine structures is crucial.

G994485099 represents anhydrides that need to be connected to the benzene ring, as depicted in Supplementary Figure 10, which highlights the adjacent positions of the anhydride on the benzene ring. The selected region within the red box in Supplementary Figure 12d shows that G994485099 values between 1-2 yield positive SHAP values. Both excessive and insufficient G994485099 adversely affect $T_{5\%}$, indicating that the connection of anhydrides to the benzene ring without adjacent side chains is critical. In Supplementary Table 1, for the same diamine A1 but different anhydrides (G2, G5, G6), at a molar ratio of 1: 2 (G994485099 = 4), PI-126 to PI-128 do not exceed 490 °C. Additionally, a low G994485099 (A24) also leads to low $T_{5\%}$, with PI-180 not exceeding 400 °C, similar to G2976033787, where the addition of aromatic compounds enhances the thermal stability of polyimides.[31]



Supplementary Figure 13. Highlighted images of diamine and dianhydride molecular structures for the 6 descriptors selected by Model-B.



Supplementary Figure 14. For the remaining four descriptors of model-B (a) A4216335232 (b) A2084364935 (c) A2720313463 (d) G3692055567 (d) G3276511768 the relationship between the descriptor value and the SHAP value output by the SHAP nested CatBoost algorithm.

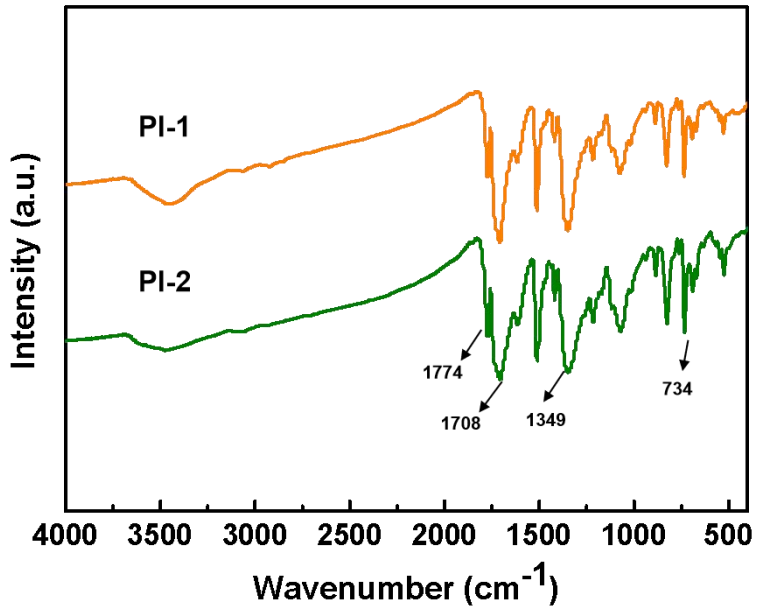
A4216335232 is a descriptor migrated from model-A. A4216335232 (Fig. S13) represents the strong π-π interactions connecting the benzene rings, etc., but contrary to its performance in films, this feature does not seem to enhance the powder model's performance (Fig. S14a). In data 10, 11 and 12 of Supplementary Table 2, increasing the amount of A70 resulted in a decrease in $T_{5\%}$. A different pattern emerges here and an effect due to too small a volume of data cannot be ruled out. More data will be provided to continue to investigate this point.

Supplementary Figure 13 highlights amine linkage points where the meta positions on the benzene rings are occupied. Supplementary Figure 14b demonstrates that A2084364935 exhibits a clear negative correlation, with SHAP values increasing as it decreases. The increased steric hindrance at the meta positions leads to lower packing density of the molecular chains, thereby reducing the thermal stability of the PI.[27] This is consistent with experimental results. In Supplementary Table 2, with the same dianhydride G7, the difference between the diamine components G23 (A2084364935 = 2) and G7 (A2084364935 = 0) results in PI-36 having a higher $T_{5\%}$ than PI-38. For the thermal stability of polyimide powders, it is important to avoid occupancy at the meta positions of the benzene rings in the diamine structures to prevent excessive steric hindrance.

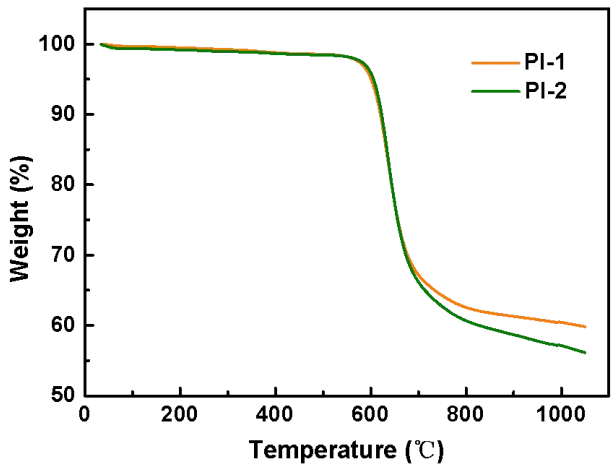
Supplementary Figure 13 highlights the structure of A2720313463, but with only one diamine molecule, it is difficult to directly illustrate structural characteristics. However, according to Supplementary Figure 14c, A2720313463 shows a clear negative correlation. In Supplementary Table 2, polyimides PI-5 and PI-41, which contain the diamine A71, both exhibit $T_{5\%}$ values that do not exceed 600 °C.

G3692055567 and G994485099 both indicate that anhydrides need to be connected to the benzene ring, but Supplementary Figure 13 highlights the meta positions of anhydrides on the benzene rings. In Supplementary Figure 14d, G3692055567 is shown near 0.1, where the SHAP values are more likely to be positive. In Supplementary Table 2, with the same diamine A70, an increased proportion of dianhydride G12 (G3692055567 = 1) results in a lower $T_{5\%}$ for PI-33 compared to PI-32. The introduction of aromatic compounds can enhance the thermal stability of polyimides, but it is advisable to avoid occupancy at the meta positions of the benzene rings in dianhydrides to prevent increased steric hindrance.

G3276511768 represents dianhydrides in Supplementary Figure 13. Figure 5b shows that G3276511768 has the least impact on $T_{5\%}$, and consequently, there is no trend for the effect of G3276511768 on $T_{5\%}$ in Supplementary Figure 14e. Due to limited data, a detailed discussion of G3276511768 is deferred at this time.



Supplementary Figure 15. FTIR spectrum of PI-a and PI-b powders.



Supplementary Figure 16. the thermo-gravimetric diagram of PI powders.

REFERENCES

1. Ding M. Polyimides: chemistry, structure-property relationships and materials. 2nd ed.;2012.
2. Lian R, Lei X, Xiao Y, Xue S, Xiong G, et al. Synthesis and properties of colorless copolyimides derived from 4,4′-diaminodiphenyl ether-based diamines with different substituents. *Polymer Chemistry* 2021;12:4803-11. doi.org/10.1039/d1py00633a
3. Qin A, Wang X, Qiao W, Wang Z. Synthesis and properties of polyimides derived from m-xylylenediamine monomer. *Fine Chemicals* 2022;39:1141-7. <Go to ISI>://CSCD:7268579
4. Yang Z-h, Kang C-q, Guo H-q, Gao L-x. Synthesis and Properties of Polyimide Films for Flexible OLED Displays. *Acta Polymerica Sinica* 2021;52:1308-15. doi.org/10.11777/j.issn1000-3304.2021.21077
5. Ting W, Yu-Miao S, Feng C, Wen-Mu L. Synthesis and Characterization of Polyimides Based on Twisted Non-coplanar Backbone Containing Indolocarbazole. *Chinese Journal of Structural Chemistry* 2021;40:1611-20. doi.org/10.14102/j.cnki.0254-5861.2011-3212
6. Luo J-R, Liu Y-D, Liu H, Chen W-P, Cui T-T, et al. Synthesis and Characterization of Polyimides with Naphthalene Ring Structure Introduced in the Main Chain. *Materials* 2022;15. doi.org/10.3390/ma15228014
7. Li Q, Zhang S, Liao G, Yi C, Xu Z. Novel fluorinated hyperbranched polyimides with excellent thermal stability, UV-shielding property, organosolubility, and low dielectric constants. *High Performance Polymers* 2018;30:872-86. doi.org/10.1177/0954008317734034
8. Liu Z, Shen Y, Li X, Shi Q, Liu B, Matsumoto T. Synthesis and properties of the novel polyimides containing cyano and biphenyl moieties. *High Performance Polymers* 2018;30:1183-92. doi.org/10.1177/0954008317744033
9. Yu B, Jiang C, Wang C, Zhao X, Li J, Ren Q. Synthesis and Characterization of Highly Transparent Fluorine-containing Copolymer Polyimide. *Fine Chemicals* 2019;36:2406-10,30. <Go to ISI>://CSCD:6697871
10. Liu Y, Zhou L, Sheng S, Song C, Hou H, Song C. Synthesis and Properties of Super-high Temperature Diketone Anhydride Polyimides. *Chinese Journal of Applied Chemistry* 2019;36:658-63. <Go to ISI>://CSCD:6508283
11. Cui X, Yang J. Synthesis and Characterization of High Performance Thermoplastic Polyimide Used the Anti-riot Bomb. In. International Conference on Electronical, Mechanical and Materials Engineering (ICE2ME). Wuhan, PEOPLES R CHINA; 2019, p. 207-10. (ISBN No. 978-94-6252-685-3)
12. Abdulhamid MA, Ma X, Ghanem BS, Pinnau I. Synthesis and Characterization of Organo-Soluble Polyimides Derived from Alicyclic Dianhydrides and a Dihydroxyl-

Functionalized Spirobisindane Diamine. *Acs Applied Polymer Materials* 2019;1:63-9. doi.org/10.1021/acsapm.8b00036
13. Zhang H, Wang W, Chen G, Zhang A, Fang X. Melt-Processable Semicrystalline Polyimides Based on 1,4-Bis(3,4-dicarboxyphenoxy) benzene Dianhydride (HQDPA): Synthesis, Crystallization, and Melting Behavior. *Polymers* 2017;9. doi.org/10.3390/polym9090420
14. Liu Y-W, Huang J, Tan J-H, Zeng Y, Ding Q, et al. Synthesis and Characterization of Intrinsic High-Barrier Polyimide Derived from a Novel Diamine Monomer Containing Rigid Planar Moiety. *Journal of Polymer Science Part a-Polymer Chemistry* 2017;55:2373-82. doi.org/10.1002/pola.28626
15. Wang C, Zhao X, Tian D, Wang D, Chen C, Zhou H. Synthesis and characterization of novel polyimides derived from 4,4′-bis(5-amino-2-pyridinoxy)benzophenone: effect of pyridine and ketone units in the main. *Designed Monomers and Polymers* 2017;20:97-105. doi.org/10.1080/15685551.2016.1231036
16. Wu F, Zhou X, Yu X. Synthesis and characterization of novel star-branched polyimides derived from 2,2-bis 4-(2,4-diaminophenoxy) phenyl hexafluoropropane. *Rsc Advances* 2017;7:35786-94. doi.org/10.1039/c7ra05643e
17. Lei Y, Shu Y, Peng J, Tang Y, Huo J. Synthesis and properties of low coefficient of thermal expansion copolyimides derived from biphenyltetracarboxylic dianhydride with p-phenylenediamine and 4,4′-oxydialinine. *E-Polymers* 2016;16:295-302. doi.org/10.1515/epoly-2015-0267
18. Huang X, Pei X, Wang L, Mei M, Liu C, Wei C. Design and synthesis of organosoluble and transparent polyimides containing bulky substituents and noncoplanar structures. *Journal of Applied Polymer Science* 2016;133. doi.org/10.1002/app.43266
19. Li B, Yan Z, Zhang T, Jiang S, Wang K, et al. Synthesis and properties of novel colorless and thermostable polyimides containing cross-linkable bulky tetrafluorostyrol pendant group and organosoluble triphenylmethane backbone structure. *Journal of Polymer Science* 2020;58:2355-65. doi.org/10.1002/pol.20200388
20. Yuan C, Sun Z, Wang Y. Synthesis and characterization of a novel organo-soluble polyimide containing hydroxyl and bis-tert-butyl substituted triphenylpyridine units. *Journal of Polymer Research* 2020;27. doi.org/10.1007/s10965-020-02208-z
21. Mirsamiei A. Synthesis and properties of polyimides derived from bis-(Aminophenoxy) containing naphthalene, Phenyl propane and Methyl cyclohexane segment and 4, 4-carbonyldiphthalic anhydride. *Journal of Macromolecular Science Part a-Pure and Applied Chemistry* 2018;55:519-25. doi.org/10.1080/10601325.2018.1476823
22. Wang Z, Guo L, Han S, Qi H, Cheng Y, Liu F. Polyimides from an asymmetric

hydroxyl-containing aliphatic-aromatic diamine synthesized via henry reaction. *Journal of Polymer Science Part a-Polymer Chemistry* 2017;55:3413-+. doi.org/10.1002/pola.28720
23. Zhao H, Chen G, Zhou Y, Li X, Fang X. Synthesis and characterization of organosoluble and transparent polyimides derived from trans-1,2-bis(3,4-dicarboxyphenoxy)cyclohexane dianhydride. *Journal of Applied Polymer Science* 2015;132. doi.org/10.1002/app.42317
24. Liu C-j, Mei M, Pei X-l, Huang X-h, Wei C. Aromatic polyimides with tertbutyl-substituted and pendent naphthalene units: synthesis and soluble, transparent properties. *Chinese Journal of Polymer Science* 2015;33:1074-85. doi.org/10.1007/s10118-015-1658-3
25. Li Q, Xiong H, Pang L, Li Q, Zhang Y, et al. Synthesis and characterization of thermally stable, hydrophobic hyperbranched polyimides derived from a novel triamine. *High Performance Polymers* 2015;27:426-38. doi.org/10.1177/0954008314555243
26. Chen Y, Zhang Q. Synthesis, characterization and properties of aromatic copolyimides containing Bi-benzimidazole moiety. *Journal of Polymer Research* 2015;22. doi.org/10.1007/s10965-015-0726-2
27. Guan Y, Wang C, Wang D, Dang G, Chen C, et al. High transparent polyimides containing pyridine and biphenyl units: Synthesis, thermal, mechanical, crystal and optical properties. *Polymer* 2015;62:1-10. doi.org/10.1016/j.polymer.2015.02.009
28. Wang Y, Wang N, Yu Z, Li G, Zhang X. Novel dye-containing copolyimides: synthesis, characterization and effect of chain entanglements on developed electrospun nanofiber morphologies. *Journal of Polymer Research* 2015;22. doi.org/10.1007/s10965-015-0713-7
29. Tapaswi PK, Choi M-C, Nagappan S, Ha C-S. Synthesis and Characterization of Highly Transparent and Hydrophobic Fluorinated Polyimides Derived from Perfluorodecylthio Substituted Diamine Monomers. *Journal of Polymer Science Part a-Polymer Chemistry* 2015;53:479-88. doi.org/10.1002/pola.27461
30. Aguilar-Lugo C, Santiago-Garcia JL, Loria-Bastarrachea MI, Guzman-Lucero D, Alexandrova L, Aguilar-Vega M. Synthesis, characterization, and structure-property relationships of aromatic polyimides containing 4,4′-diaminotriphenylmethane. *Journal of Polymer Research* 2016;23. doi.org/10.1007/s10965-016-0939-z
31. Kumar A, Tateyama S, Yasaki K, Ali MA, Takaya N, et al. Ultrahigh performance bio-based polyimides from 4,4′-diaminostilbene. *Polymer* 2016;83:182-9. doi.org/10.1016/j.polymer.2015.12.008
32. Liu S, Zhang Y, Wang X, Tan H, Song N, Guan S. Synthesis and properties of hyperbranched polyimides derived from tetra-amine and long-chain aromatic dianhydrides. *Rsc Advances* 2015;5:107793-803. doi.org/10.1039/c5ra23227a

33. Yang G, Zhang R, Huang H, Liu L, Wang L, Chen Y. Synthesis of novel biobased polyimides derived from isomannide with good optical transparency, solubility and thermal stability. *Rsc Advances* 2015;5:67574-82. doi.org/10.1039/c5ra14526k
34. Zhou Y, Chen G, Zhao H, Song L, Fang X. Synthesis and properties of transparent polyimides derived from trans-1,4-bis(2,3-dicarboxyphenoxy)cyclohexane dianhydride. *Rsc Advances* 2015;5:53926-34. doi.org/10.1039/c5ra07676e
35. Wang Y. Synthesis and characterization of novel polyimides derived from 2,4-bis(4-aminophenoxy)pyrimidine. *High Performance Polymers* 2014;26:978-85. doi.org/10.1177/0954008314536216
36. Tapaswi PK, Choi M-C, Jung YS, Cho HJ, Seo DJ, Ha C-S. Synthesis and Characterization of Fully Aliphatic Polyimides from an Aliphatic Dianhydride with Piperazine Spacer for Enhanced Solubility, Transparency, and Low Dielectric Constant. *Journal of Polymer Science Part a-Polymer Chemistry* 2014;52:2316-28. doi.org/10.1002/pola.27242
37. Li Y, Liu C, Jiao L, Song G, Zhao X, Dang G. Synthesis of new autophotosensitive semiaromatic hyperbranched polyimides with excellent mechanical properties and low birefringences. *High Performance Polymers* 2014;26:569-77. doi.org/10.1177/0954008314523055
38. Guan Y, Wang D, Song G, Zhang X, Chen C, et al. Synthesis and characterization of novel polyimides derived from 3,6-bis(4-aminophenoxy)pyridazine. *High Performance Polymers* 2014;26:455-62. doi.org/10.1177/0954008313518566
39. Cao X, Song Y, Matsumoto T, Liu B. Synthesis and properties of cyano group-containing polyimides with high peel strength. *High Performance Polymers* 2016;28:953-61. doi.org/10.1177/0954008315611469
40. Chen Y, Zhang Q. Synthesis and properties of polyimides derived from diamine monomer containing bi-benzimidazole unit. *Journal of Polymer Research* 2014;21. doi.org/10.1007/s10965-014-0424-5
41. Yao H, Zhang Y, Liu Y, You K, Liu S, et al. Synthesis and Properties of Cross-Linkable High Molecular Weight Fluorinated Copolyimides. *Journal of Polymer Science Part a-Polymer Chemistry* 2014;52:349-59. doi.org/10.1002/pola.27007
42. Chen Y, Zhang Q, Sun W, Lei X, Yao P. Synthesis and gas permeation properties of hyperbranched polyimides membranes from a novel (A₂ + B₂B′ + B₂)-type method. *Journal of Membrane Science* 2014;450:138-46. doi.org/10.1016/j.memsci.2013.09.003
43. Guan Y, Wang D, Wang Z, Dang G, Chen C, et al. Synthesis and characterization of novel polyimides from 4,4′-bis(5-amino-2-pyridinoxy)diphenyl ether, 4,4′-bis(5-amino-2-pyridinoxy)diphenyl thioether and 4,4′-bis(5-amino-2-pyridinoxy)diphenyl sulfone. *Rsc Advances* 2014;4:50163-70. doi.org/10.1039/c4ra07855a

44. Liu C, Pei X, Mei M, Chou G, Huang X, Wei C. Synthesis and characterization of organosoluble, transparent, and hydrophobic fluorinated polyimides derived from 3,3-diisopropyl-4,4-diaminodiphenyl-4-trifluoromethyltoluene. *High Performance Polymers* 2016;28:1114-23. doi.org/10.1177/0954008315617230
45. Huang X, Zhao CY, Wang H, Ju S. AI-assisted inverse design of sequence-ordered high intrinsic thermal conductivity polymers. *Materials Today Physics* 2024;44:101438. doi.org/https://doi.org/10.1016/j.mtphys.2024.101438
46. Morgan HL. The Generation of a Unique Machine Description for Chemical Structures-A Technique Developed at Chemical Abstracts Service. *Journal of Chemical Documentation* 1965;5:107-13. doi.org/10.1021/c160017a018
47. Chen Y, Liu Y, Min Y. Synthesis and Properties Comparison of Low Dielectric Silicon Containing Polyimides. *Materials* 2022;15:2755. https://www.mdpi.com/1996-1944/15/8/2755
48. Butler KT, Davies DW, Cartwright H, Isayev O, Walsh A. Machine learning for molecular and materials science. *Nature* 2018;559:547-55. doi.org/10.1038/s41586-018-0337-2
49. Wen C, Wang C, Zhang Y, Antonov S, Xue D, et al. Modeling solid solution strengthening in high entropy alloys using machine learning. *Acta Materialia* 2021;212. doi.org/10.1016/j.actamat.2021.116917