Supplementary Information

Polymer genome approach for rational design of poly(aryl ether)s with high glass transition temperature

Ce Song[a,b], Hongjian Gu[b], Linyan Zhu[b], Wanyuan Jiang[b], Zhihuan Weng[b], Lishuai Zong[b], Cheng Liu[b], Fangyuan Hu[a,b], Yuxi Pan[b,*], and Xigao Jian[a,b,*]

[a] *School of Materials Science and Engineering, Key Laboratory of Energy Materials and Devices (Liaoning Province), Dalian University of Technology, Dalian 116024, China*
[b] *State Key Laboratory of Fine Chemicals, Department of Polymer Science & Engineering, Frontiers Science Center for Smart Materials Oriented Chemical Engineering, Technology Innovation Center of High Performance Resin Materials (Liaoning Province), Dalian University of Technology, Dalian 116024, China*

*Corresponding author.
E-mail address:* panyx@dlut.edu.cn (Yuxi Pan); jian4616@dlut.edu.cn (Xigao Jian)

Table of content

1. Additional information for the database.
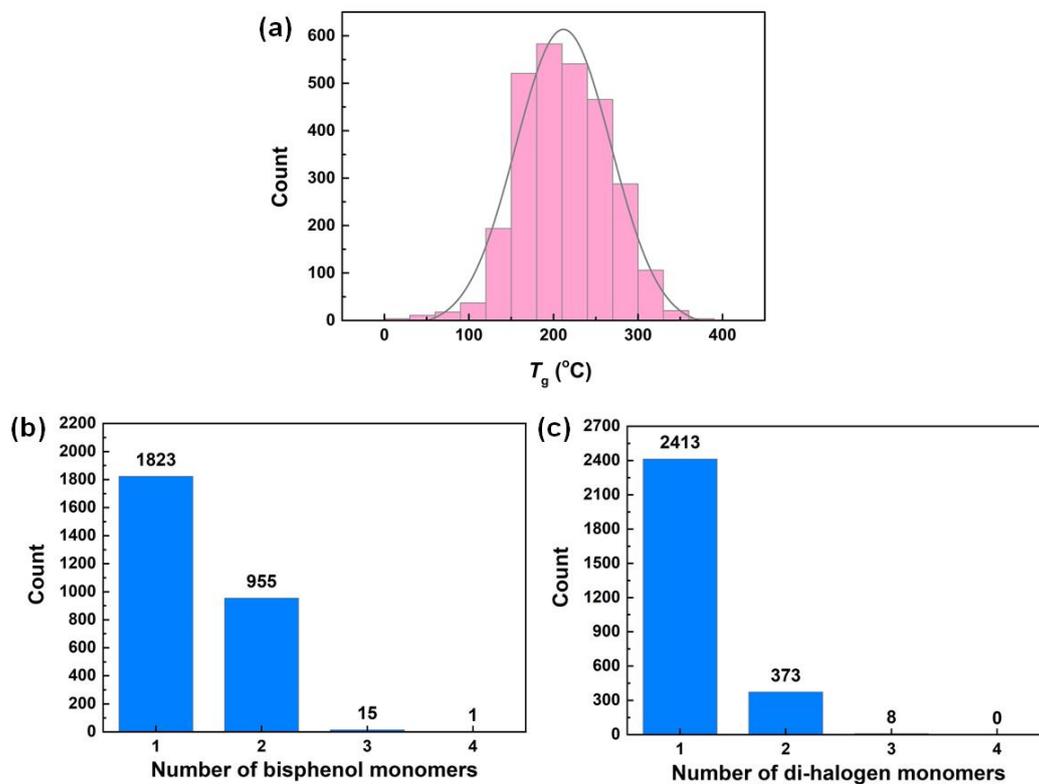
Figure S1. (a) Distribution of experimental T_g values for PAEs in the database established in this work. Distribution of the number of (b) bisphenol monomers and (c) di-halogen monomers contained in a single PAE in the database.
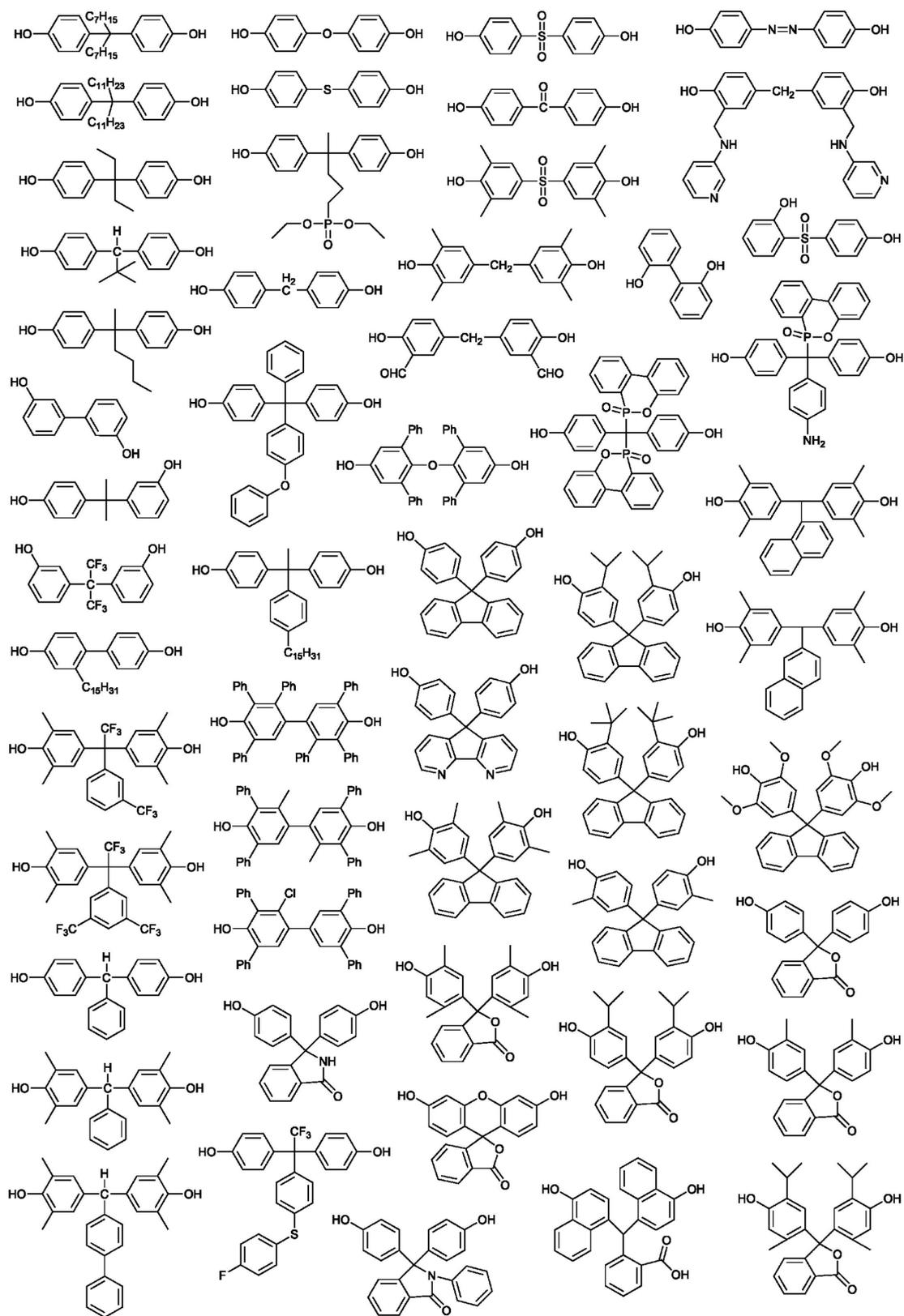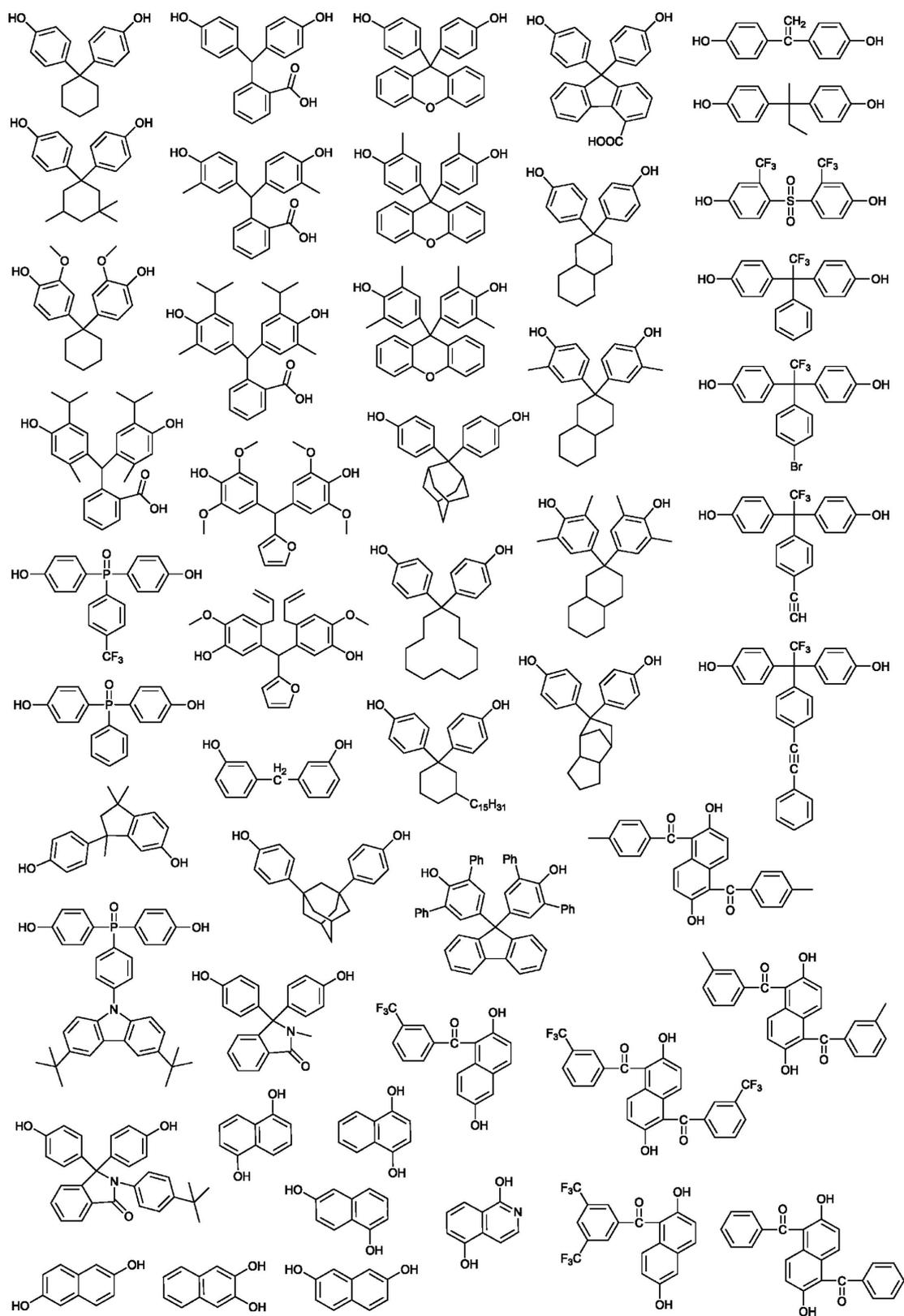
Bisphenol and bisphenol-like monomers

Di-halogen monomers

16

Figure S2. Chemical structures of bisphenol/bisphenol-like and di-halogen monomers involved in the database. "X" in the di-halogen monomer indicates F/Cl/Br element.
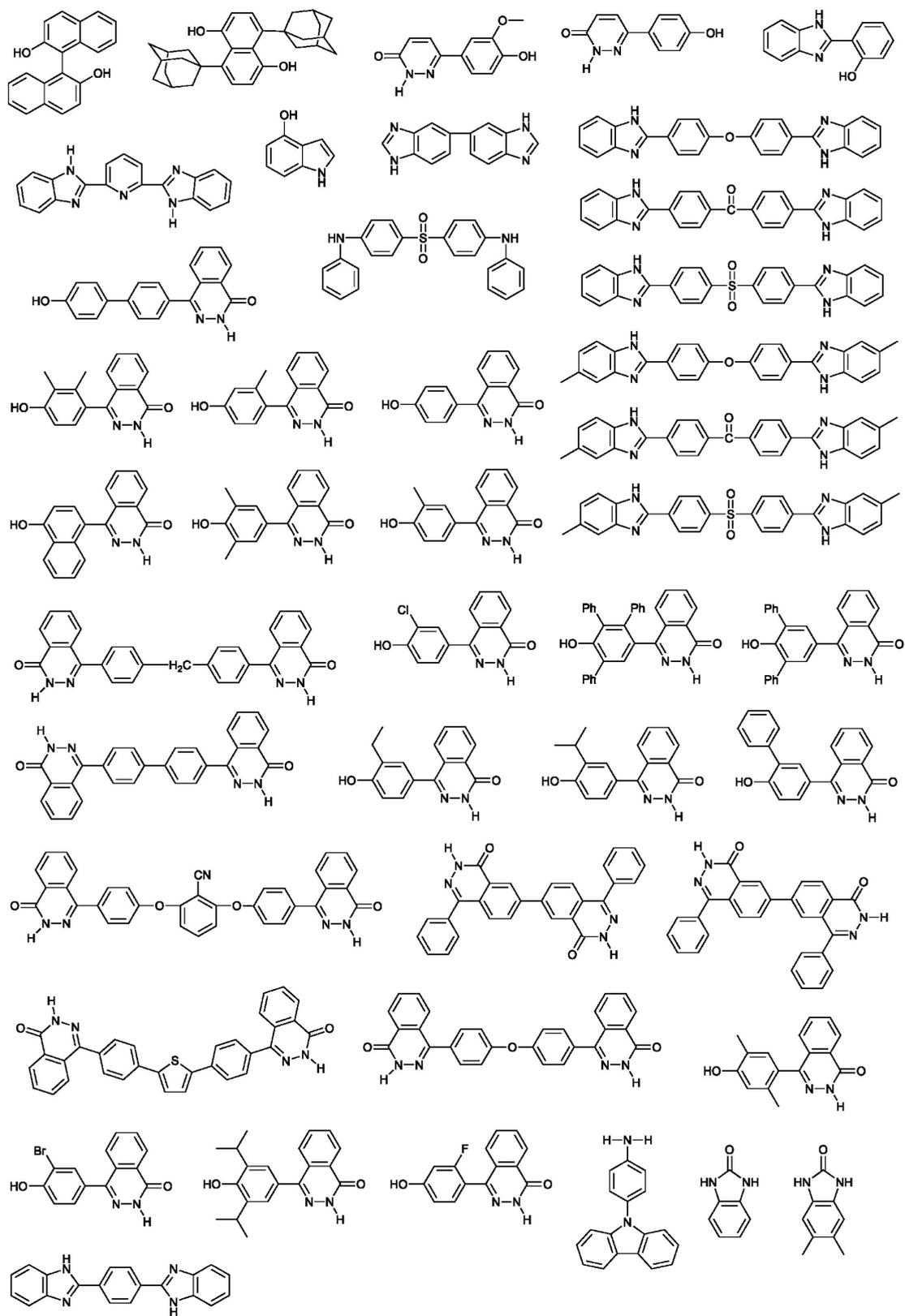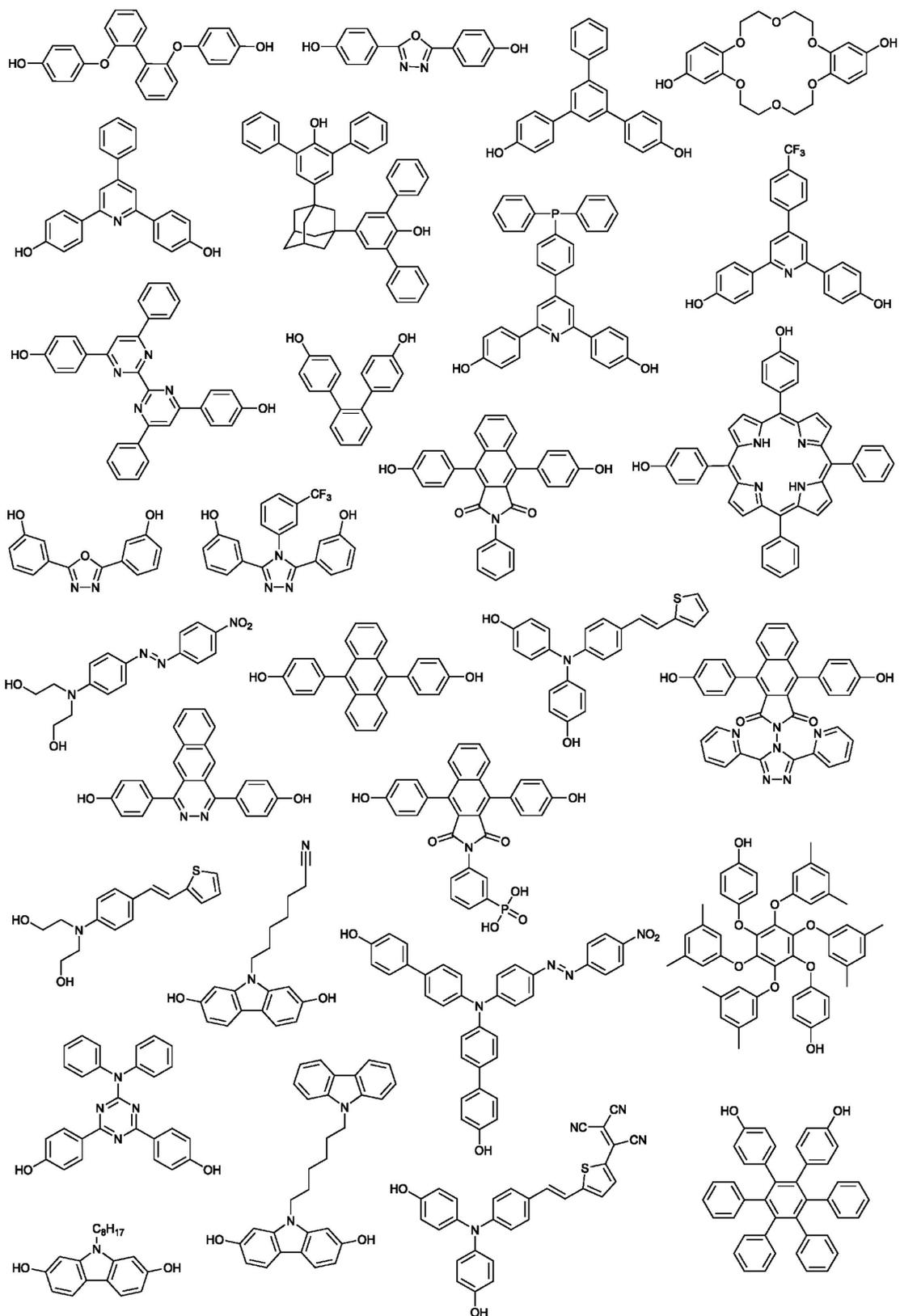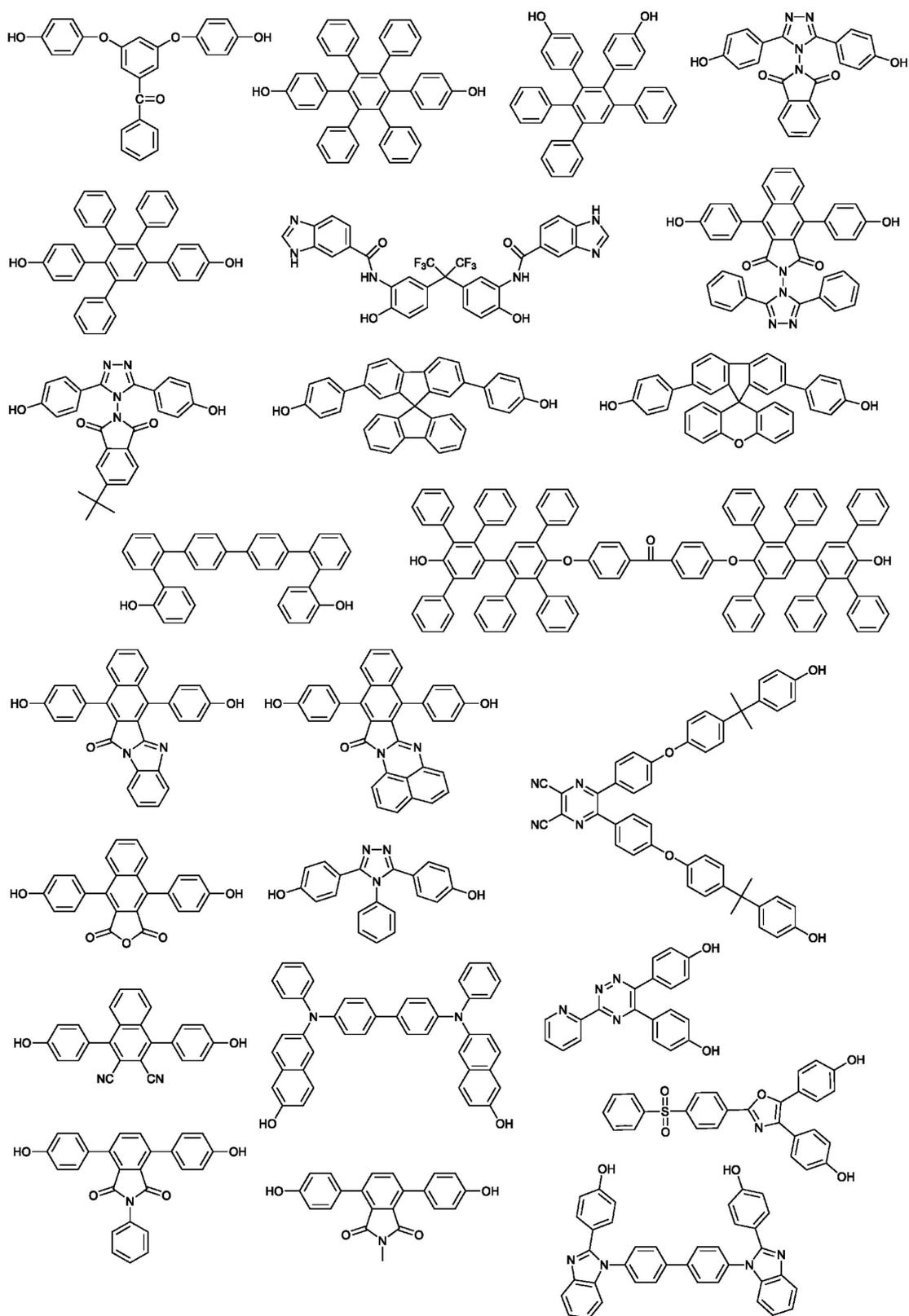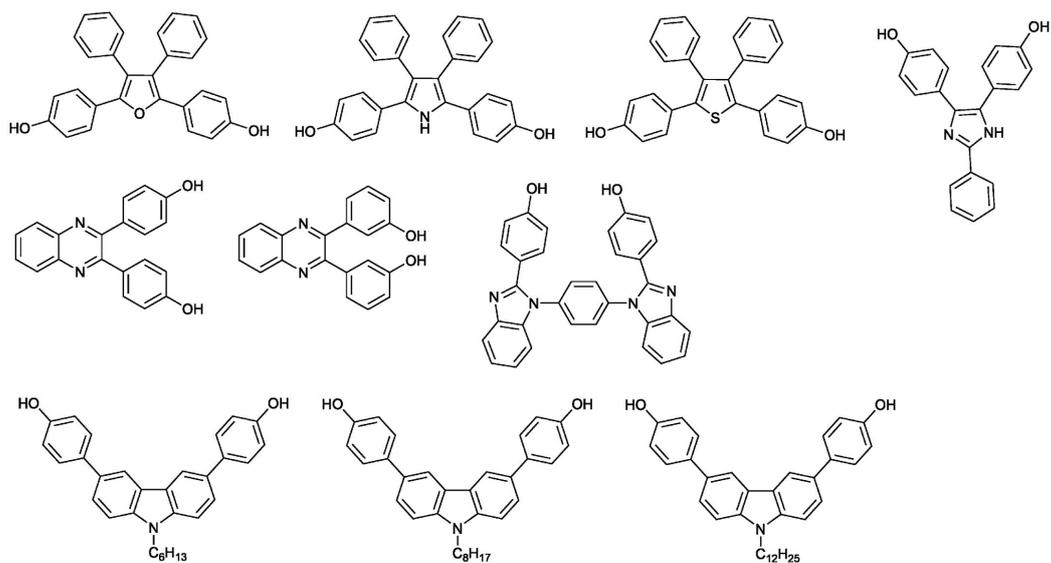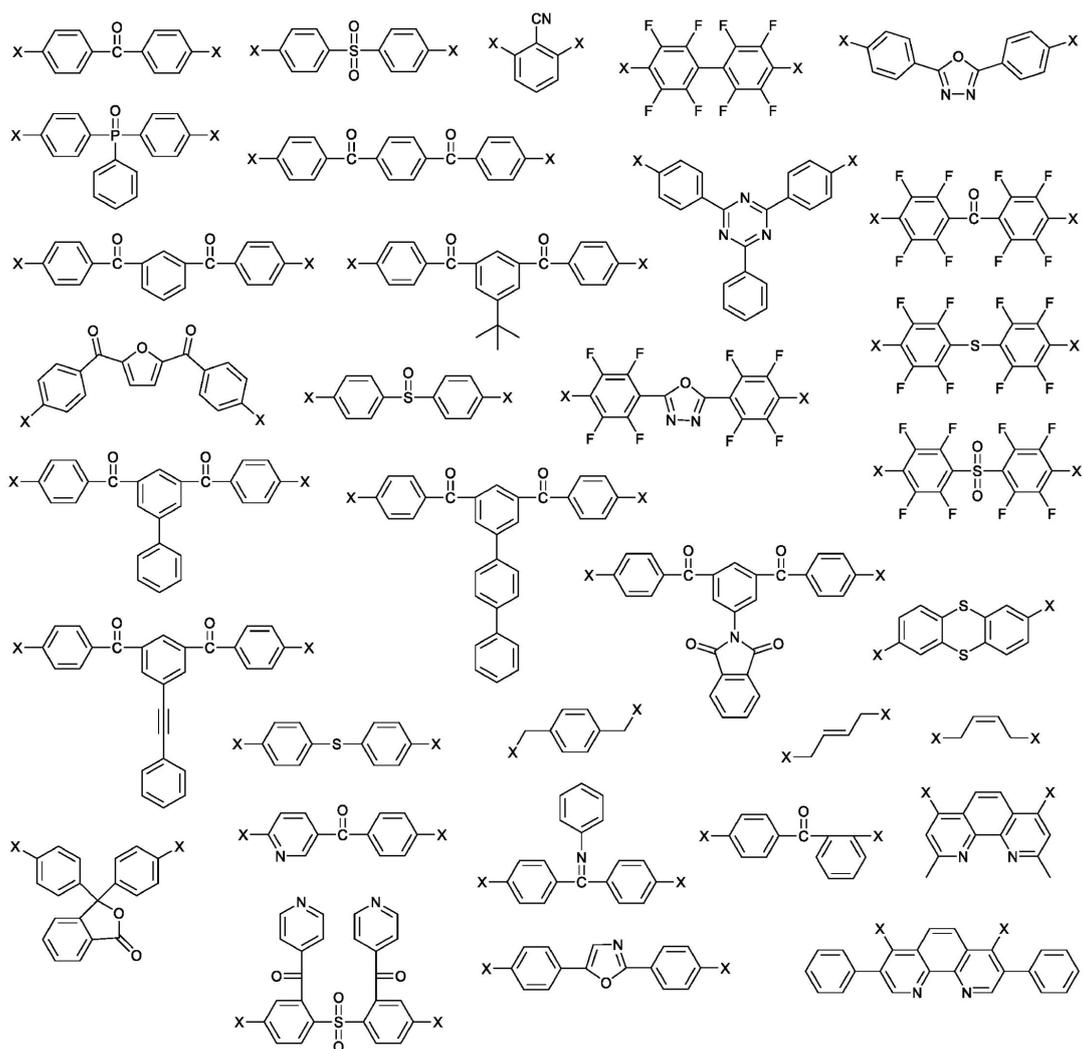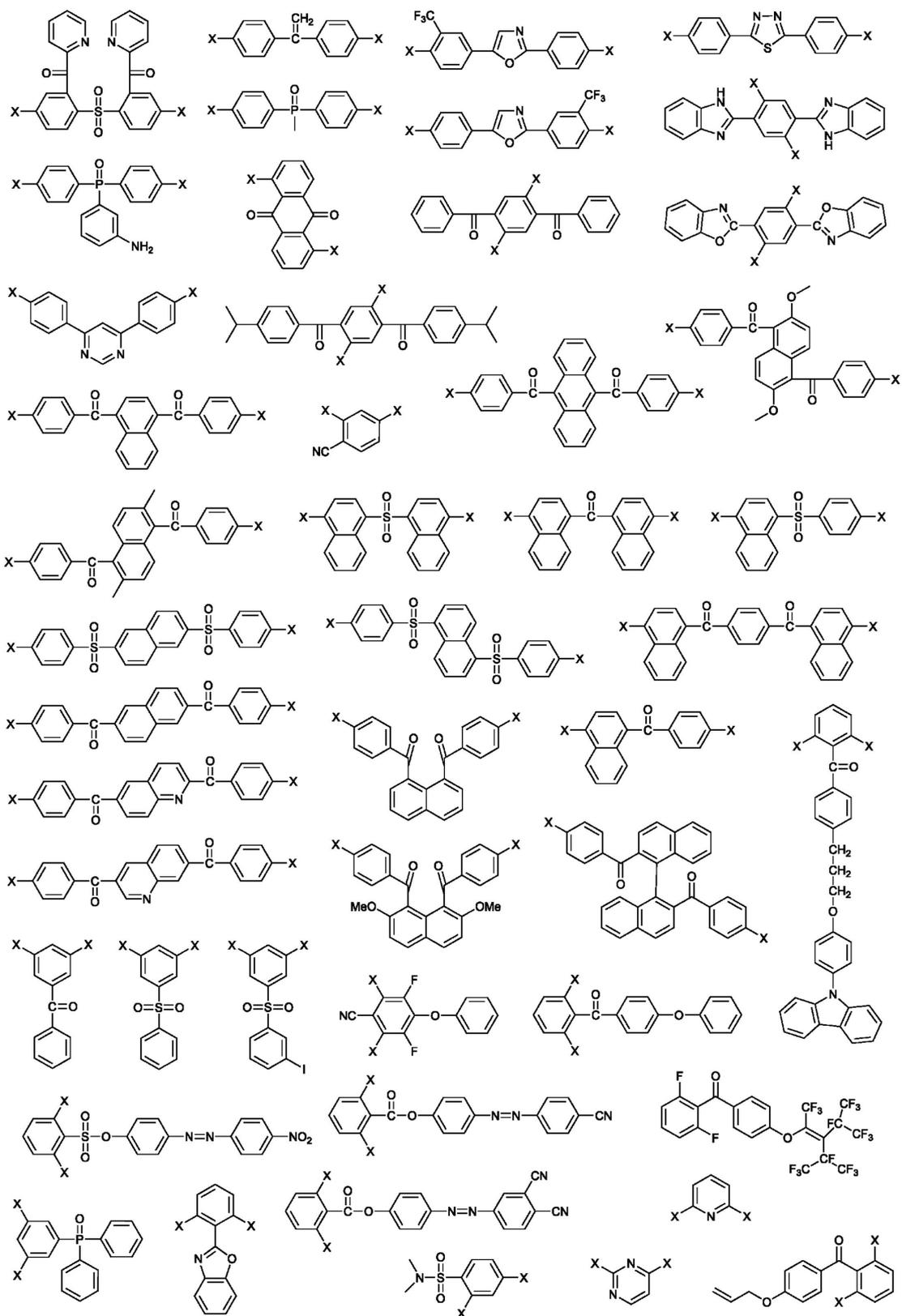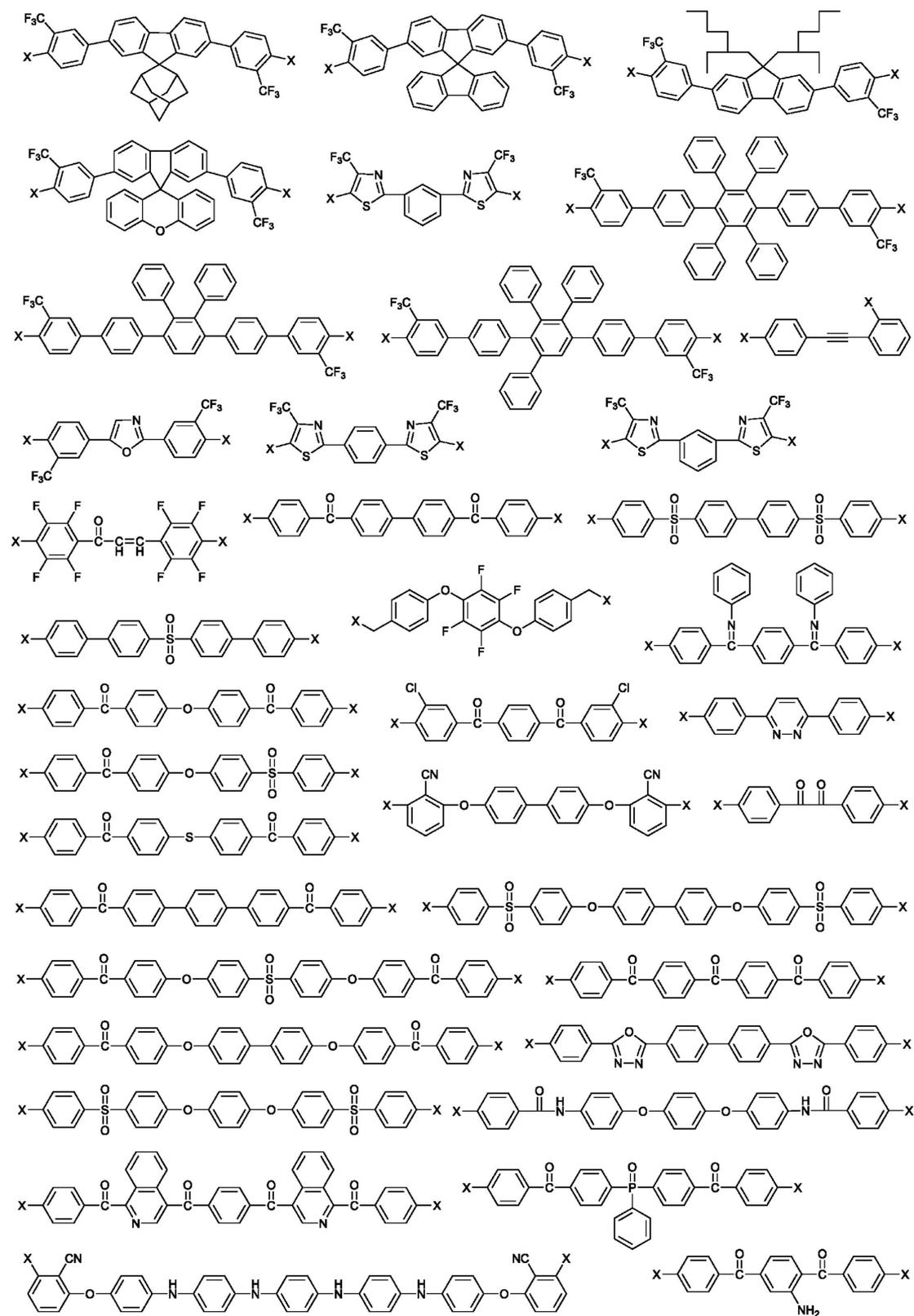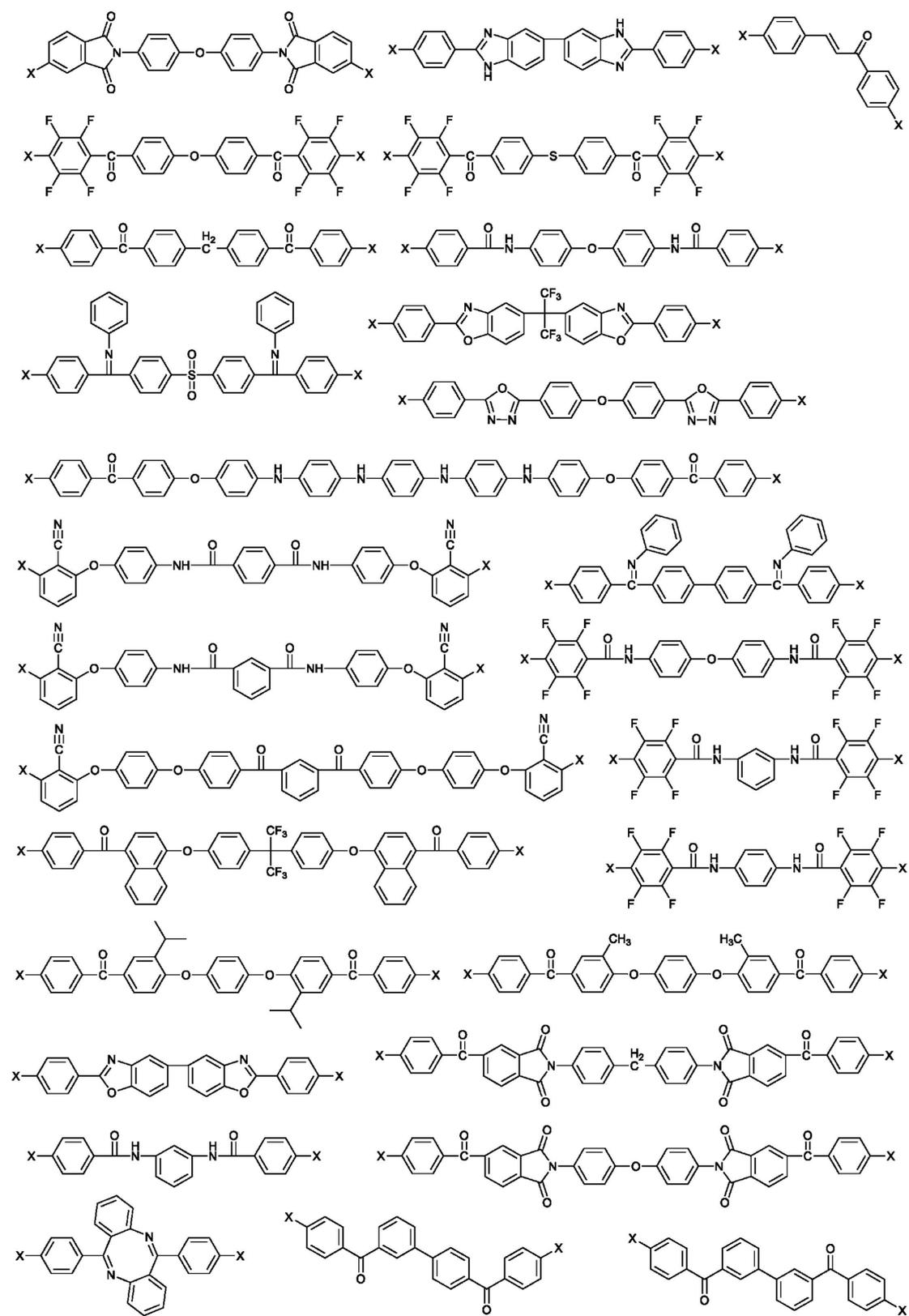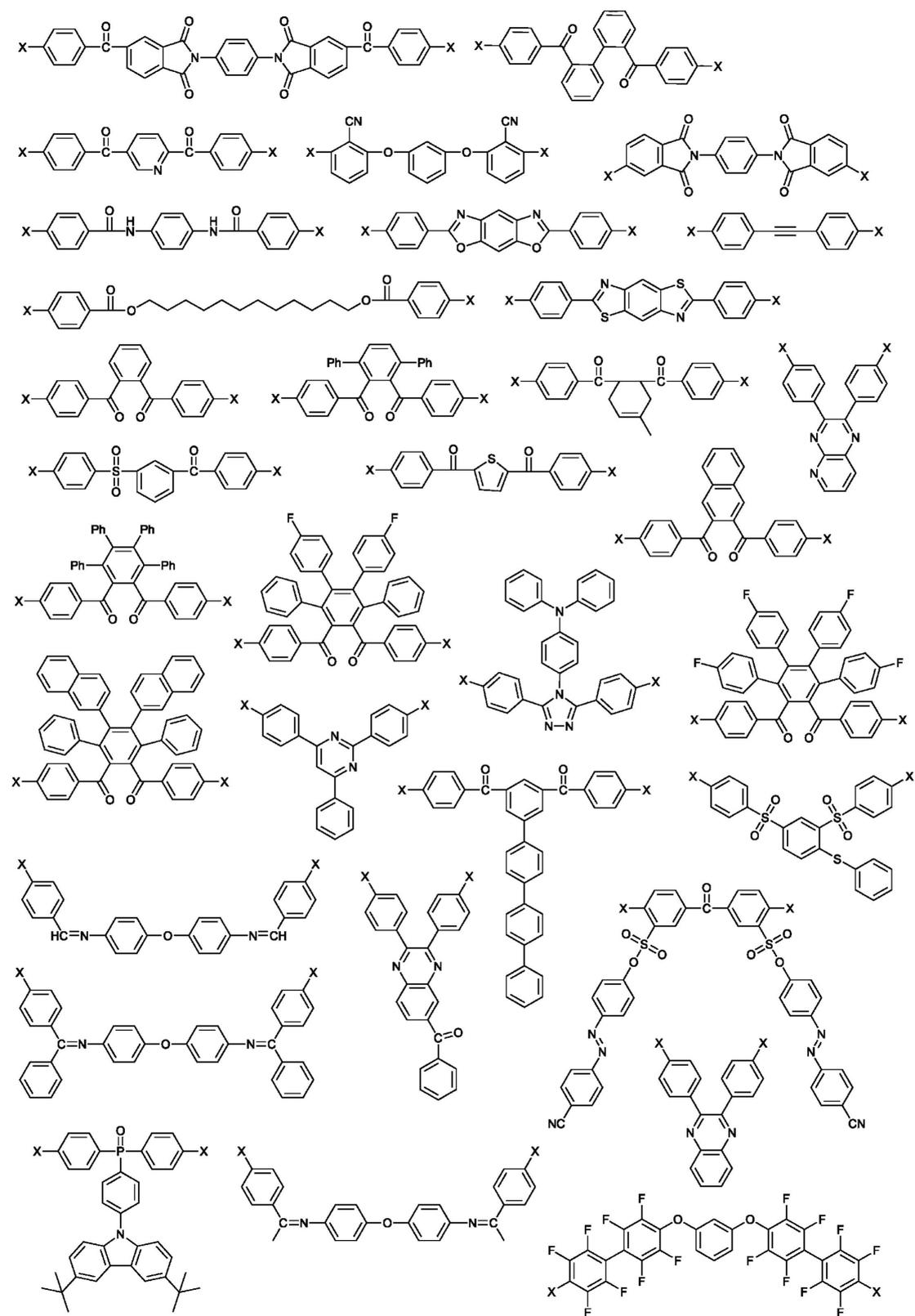
2. Details for the copolymer representation and featurization framework.
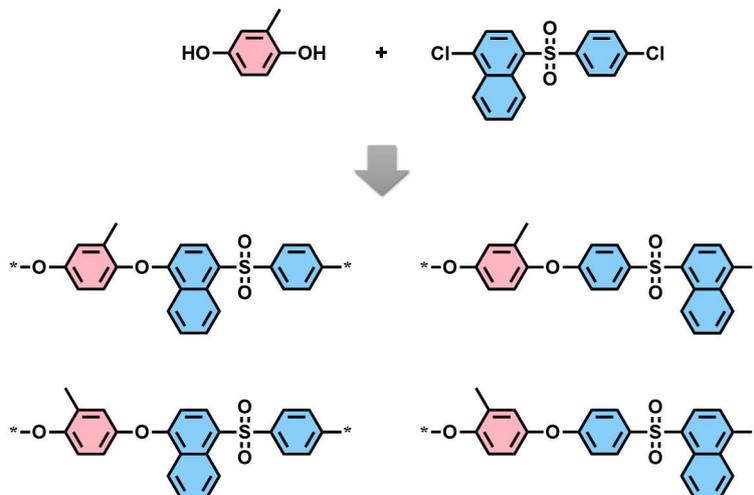
Figure S3. Graphical illustration of the possible structures of repeat units derived from asymmetric bisphenol and di-halogen monomers.

Table S1. List of function names for descriptors calculated from RDKit. Names of 3D descriptors are in bold font.

MaxEStateIndex	MinEStateIndex	MaxAbsEStateIndex	MinAbsEStateIndex
qed	MolWt	HeavyAtomMolWt	FpDensityMorgan1
BalabanJ	ExactMolWt	MaxPartialCharge	FpDensityMorgan2
BertzCT	Chi1n	MinPartialCharge	FpDensityMorgan3
Chi0	Chi3n	MaxAbsPartialCharge	NumRadicalElectrons
Chi0n	Chi0v	Chi1	MinAbsPartialCharge
Chi1v	Chi2n	Chi2v	NumValenceElectrons
Chi3v	Chi4n	Chi4v	HallKierAlpha
Ipc	Kappa1	Kappa2	Kappa3
LabuteASA	PEOE_VSA1	PEOE_VSA10	PEOE_VSA11
PEOE_VSA12	PEOE_VSA13	PEOE_VSA14	PEOE_VSA2
PEOE_VSA3	PEOE_VSA4	PEOE_VSA5	PEOE_VSA6
PEOE_VSA7	PEOE_VSA8	PEOE_VSA9	SMR_VSA1
SMR_VSA10	SMR_VSA2	SMR_VSA3	SMR_VSA4
SMR_VSA5	SMR_VSA6	SMR_VSA7	SMR_VSA8
SMR_VSA9	SlogP_VSA1	SlogP_VSA10	SlogP_VSA11
SlogP_VSA12	SlogP_VSA2	SlogP_VSA3	SlogP_VSA4
SlogP_VSA5	SlogP_VSA6	SlogP_VSA7	SlogP_VSA8
SlogP_VSA9	TPSA	EState_VSA1	EState_VSA10
EState_VSA11	EState_VSA2	EState_VSA3	EState_VSA4
EState_VSA5	EState_VSA6	EState_VSA7	EState_VSA8
EState_VSA9	VSA_EState1	VSA_EState10	VSA_EState2
VSA_EState3	VSA_EState4	VSA_EState5	VSA_EState6
VSA_EState7	VSA_EState8	VSA_EState9	NumAliphaticRings
FractionCSP3	NOCount	NumAliphaticHeterocycles	NumAliphaticCarbocycles
NHOHCount	HeavyAtomCount	NumAromaticCarbocycles	NumAromaticHeterocycles
MolLogP	NumHAcceptors	NumSaturatedCarbocycles	NumSaturatedHeterocycles
MolMR	NumHDonors	NumHeteroatoms	NumSaturatedRings
RingCount	fr_Al_COO	NumRotatableBonds	NumAromaticRings
fr_Al_OH	fr_ArN	fr_Al_OH_noTert	fr_Ar_COO
fr_Ar_N	fr_Ar_NH	fr_Ar_OH	fr_COO
fr_COO2	fr_C_O	fr_C_O_noCOO	fr_C_S
fr_HOCCN	fr_Imine	fr_NH0	fr_NH1
fr_NH2	fr_N_O	fr_Ndealkylation1	fr_Ndealkylation2
fr_Nhpyrrole	fr_SH	fr_aldehyde	fr_alkyl_carbamate
fr_alkyl_halide	fr_allylic_oxid	fr_amide	fr_amidine
fr_aniline	fr_aryl_methyl	fr_azide	fr_azo
fr_barbitur	fr_benzene	fr_benzodiazepine	fr_bicyclic
fr_diazo	fr_epoxide	fr_dihydropyridine	fr_ester
fr_ether	fr_furan	fr_guanido	fr_halogen
fr_hdrzine	fr_hdrzone	fr_imidazole	fr_imide
fr_isocyan	fr_isothiocyan	fr_ketone	fr_ketone_Topliss
fr_lactam	fr_lactone	fr_methoxy	fr_morpholine
fr_nitrile	fr_nitro	fr_nitro_arom	fr_nitro_arom_nonortho
fr_nitroso	fr_oxazole	fr_oxime	fr_para_hydroxylation
fr_phenol	fr_phos_acid	fr_phenol_noOrthoHbond	fr_phos_ester
fr_piperdine	fr_piperzine	fr_priamide	fr_prisulfonamd
fr_pyridine	fr_quatN	fr_term_acetylene	fr_sulfonamd
fr_sulfone	fr_sulfide	fr_tetrazole	fr_thiazole
fr_thiocyan	fr_thiophene	fr_unbrch_alkane	fr_urea
PBF	**PMI1**	**PMI2**	**PMI3**
NPR1	**NPR2**	**RadiusOfGyration**	**InertialShapeFactor**
Eccentricity	**Asphericity**	**SpherocityIndex**	

Table S2. Parameter settings for the RF model during the RFECV process.

Parameter	Setting
n_estimators	90
criterion	'squared_error'
max_depth	10
min_samples_split	2
min_samples_leaf	1

3. Supplementary information for fragment analysis.

The Z-score (Z_i) of each fragment i was calculated by

$$Z_i = \frac{k_i - m\frac{K_i}{M}}{\sigma_i}$$

with

$$\sigma_i = \sqrt{\frac{mK_i(M - K_i)(M - m)}{M^2(M - 1)}}$$

where M is the total number of PAE copolymers in the database, m represents the number of polymers whose T_g are higher than 250 °C in the database, K_i is the number of occurrences of fragment i in M polymers, and k_i is the occurrence of fragment i in m polymers.
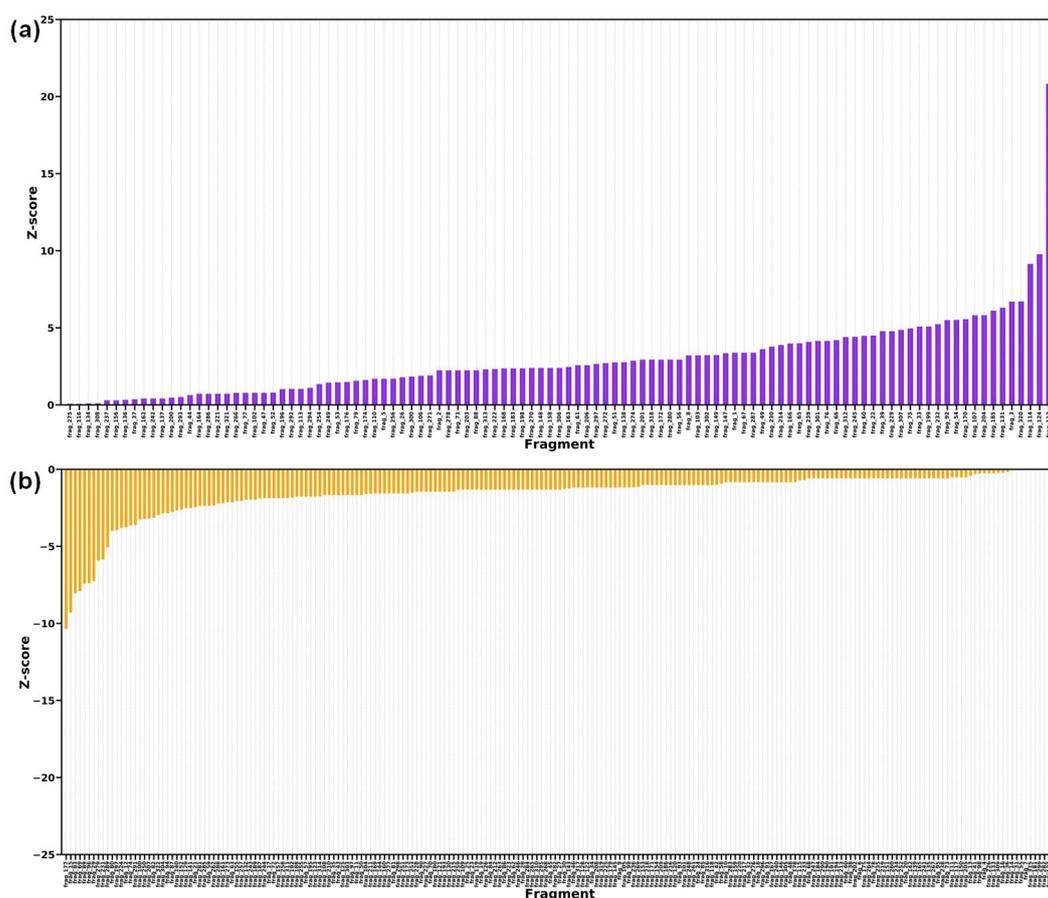
Figure S4. (a) Positive and (b) negative Z-scores of the 322 constituting fragments in the subset of PAEs with T_g higher than 250 °C.

21

4. Additional details and analysis for machine learning models.

Table S3. List of the selected descriptors used as inputs for Model_1.

Descriptor name	Description
MaxEStateIndex	Maximum EState index from J. Chem. Inf. Comput. Sci., vol 31, 76-82 (1991).
MinEStateIndex	Minimum EState index from J. Chem. Inf. Comput. Sci., vol 31, 76-82 (1991).
MinAbsEStateIndex	Minimum absolute EState index from J. Chem. Inf. Comput. Sci., vol 31, 76-82 (1991).
qed	Quantitative estimate of drug-likeness from Nat. Chem., vol 4, 90-98 (2012).
MaxAbsPartialCharge	Maximum absolute partial Gasteiger atomic charge of the molecule.
FpDensityMorgan1	Morgan fingerprint, radius 1
BalabanJ	A chemical distance-based topological index from Chem. Phys. Lett. vol 89, 399-404 (1982).
Chi4v	A valence molecular connectivity index calculated from equations (5), (15) and (16) of Rev. Comp. Chem., vol 2, 367-422 (1991).
Chi4n	Similar to Hall Kier Chi4v, but uses nVal instead of valence.
Kappa3	Third-order shape/connectivity index from Rev. Comp. Chem. vol 2, 367-422 (1991).
PEOE_VSAN, N = 1, 3, 6, 7, 8, 9, 13	Molecular operating environment (MOE)-type descriptors using partial charges and surface area contributions from J. Mol. Graph. Model., Vol 18, 464-477 (2000).
SMR_VSAN, N = 1, 6, 9, 10	MOE-type descriptors using molar refractivity contributions and surface area contributions from J. Mol. Graph. Model., Vol 18, 464-477 (2000).
SlogP_VSAN, N = 4, 5, 8, 11	MOE-type descriptors using LogP contributions and surface area contributions from J. Mol. Graph. Model., Vol 18, 464-477 (2000).
TPSA	Sum of topological surface area of polar atoms in the molecule from J. Med. Chem., vol 43, 3714-7 (2000).
EState_VSAN, N = 9, 10 VSA_EStateN, N = 2, 3, 4, 5, 7, 8, 9	MOE-type descriptors using electrotopological state indices and surface area contributions from J. Chem. Inform. Comput. Sci., vol 31, 76-81 (1991).
FractionCSP3	Fraction of sp^3 hybridized carbon atoms.
NumAliphaticRings	Number of aliphatic (containing at least one non-aromatic bond) rings.
NumRotatableBonds	Number of rotatable bonds.
fr_bicyclic	Number of bicyclic rings.
fr_ketone	Number of ketones.
fr_para_hydroxylation	Number of para-hydroxylation sites.
fr_sulfone	Number of sulfone groups.
NPR2	Normalized principal moments ratio 2 ($=I_2/I_3$) from J. Chem. Inf. Comput. Sci., Vol. 43, 987-1003 (2003).
RadiusOfGyration	Radius of gyration from handbook of Chemoinformatics (https://doi.org/10.1002/9783527618279.ch37)
Eccentricity	Molecular eccentricity from handbook of chemoinformatics (https://doi.org/10.1002/9783527618279.ch37)
SpherocityIndex	Spherocity index from handbook of chemoinformatics (https://doi.org/10.1002/9783527618279.ch37)

Table S4. Parameter settings for the SVR predictive model.

Parameter	Setting
kernel	'rbf'
C	200
epsilon	0.1
gamma	'scale'
tol	0.001
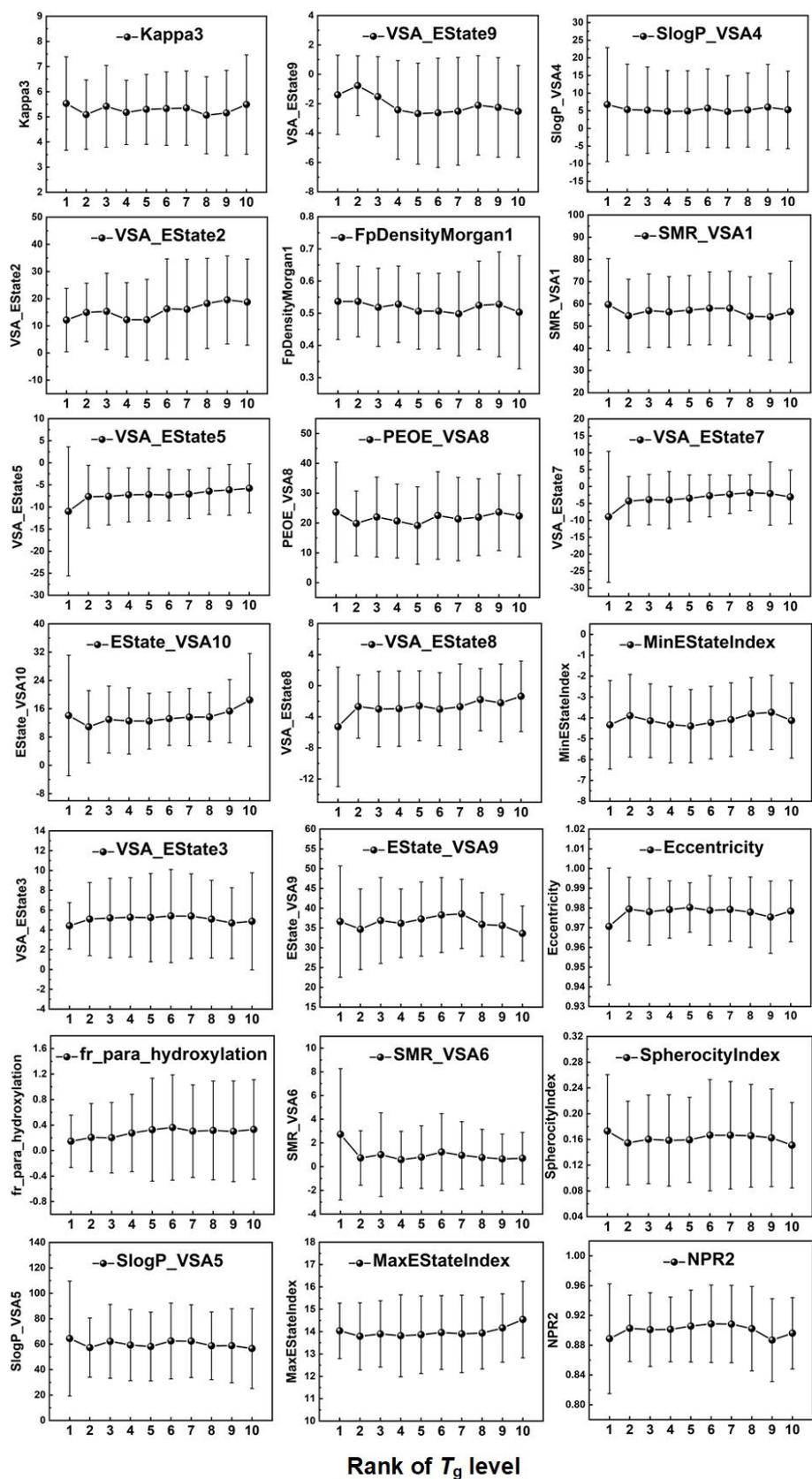
Figure S5. Average values with standard deviations of the selected descriptors in 10 subgroups of increasing T_g levels. The smaller number of the x-axis indicates the lower T_g level.

5. Experimental synthesis and validation of new PAEs.

Materials. 4-(4-hydroxyphenyl)phthalazin-1(2*H*)-one (DHPZ) was supplied by Dalian Polymer New Materials Co., Ltd. (Dalian, China). 4,4'-(9*H*-fluorene-9,9-diyl)diphenol (BHPF) was purchased from Wuhan Jiakailong Co. Ltd., 2,6-difluorobenzonitrile (DFBN) and 4,4'-sulfonylbis(chlorobenzene) (DCS) were obtained from Macklin Co. Ltd., and DHPZ, BHPF, DFBN, and DCS were used without further purification.

Monomer Synthesis. Monomer TBPZ was synthesized according to the synthetic route illustrated in Figure S6. Firstly, 0.33 mol of phthalic anhydride, 0.825 mol of anhydrous aluminum chloride, and 250 mL of 1,2-dichloroethane were added in a 500 mL three-neck flask. Subsequently, 0.15 mol of diphenyl sulfide was added slowly at 0 °C. The reaction was heated to 50 °C and held at this temperature for 6 h. The mixture was poured onto 500 g of crushed ice containing 15 mL of HCl. The 1,2-dichloroethane was evaporated and the 4,4′-bis(2-carboxybenzoyl)diphenyl sulfide was filtered and dried. Secondly, 0.2 mol of hydrazine monohydrate was added into 0.1 mol of 4,4′-bis(2-carboxybenzoyl)diphenyl sulfide and 500 mL of ethanol. The reaction mixture was heated to reflux and stirred for 6 h. A white powder precipitated, which was washed with ethanol and dried at 80 °C. A total of 44.1 g of TBPZ was obtained as a white powder.
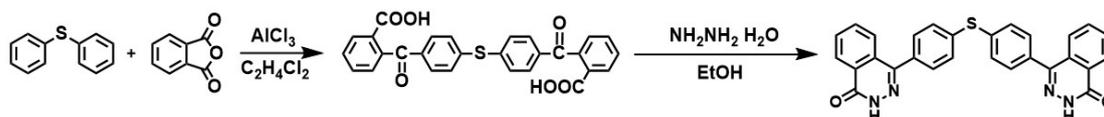
Figure S6. Synthetic route of monomer TBPZ.

Polymer Synthesis. PPFES, PPFEN, and PPFENS were synthesized by high temperature solution nucleophilic substitution of DHPZ and BHPF with DCS and DFBN. The detailed feed ratios of these monomers are presented in the following Table S5. PBPS was synthesized by high temperature solution nucleophilic substitution of TBPZ with DCS. Detailed synthesis process can refer to the literatures [1]. No unexpected or unusually high safety hazards were encountered.

Table S5. Feed ratios of monomers for preparations of PPFES, PPFEN, and PPFENS.

PAEs	DHPZ [a] (mol%)	BHPF [b] (mol%)	DCS [c] (mol%)	DFBN [d] (mol%)
PPFES	50	50	100	0
PPFEN	50	50	0	100
PPFENS	50	50	50	50

[a] Feed molar ratio of DHPZ relative to the total bisphenols.

[b] Feed molar ratio of BHPF relative to the total bisphenols.

[c] Feed molar ratio of DCS relative to the total di-halogens.

[d] Feed molar ratio of DFBN relative to the total di-halogens.

Structural Characterization. The ^{1}H-NMR and Fourier transform infrared (FT-IR) techniques were conducted to characterize chemical structures of PPFES, PPFEN, PPFENS, and PBPS. The ^{1}H-NMR spectra of these PAEs were recorded on a Bruker Vaian DLG400 spectrometer. Samples were composed of a solution of 5–10 mg of each compound in 0.5 ml of deuterated chloroform using tetramethyl silane (TMS) as the internal standard at room temperature. The ^{1}H-NMR spectra of synthesized samples are shown in the following Figure S7. FT-IR measurements were conducted with a Thermo Nicolet IS50 FT-IR spectrometer. Figure S8 plots the FT-IR spectra of PPFES, PPFEN, PPFENS, and PBPS.

Thermal Property Characterization. Glass transition temperatures (T_g) were performed with a METTLER DSC instrument under 50 ml min^{-1} nitrogen flows at 10 °C min^{-1} ramping rate. After the first heating and quenching, 30 °C-400 °C-30 °C-400 °C, the T_g data were selected from the second scan.

Table S6. Experimentally measured T_g obtained via DSC tests compared with the predicted results from the SVR model.

PAEs	Predicted T_g (°C) [a]	Experimental T_g (°C)
PPFES	282	288
PPFEN	283	289
PPFENS	290	292
PBPS	323	328

[a] Predictions on new PAEs were performed by the Model_1 with the highest predictive performance (i.e., the lowest RMSE and the highest R^2, as shown in Fig. 4 in the manuscript) among the ten SVR models.
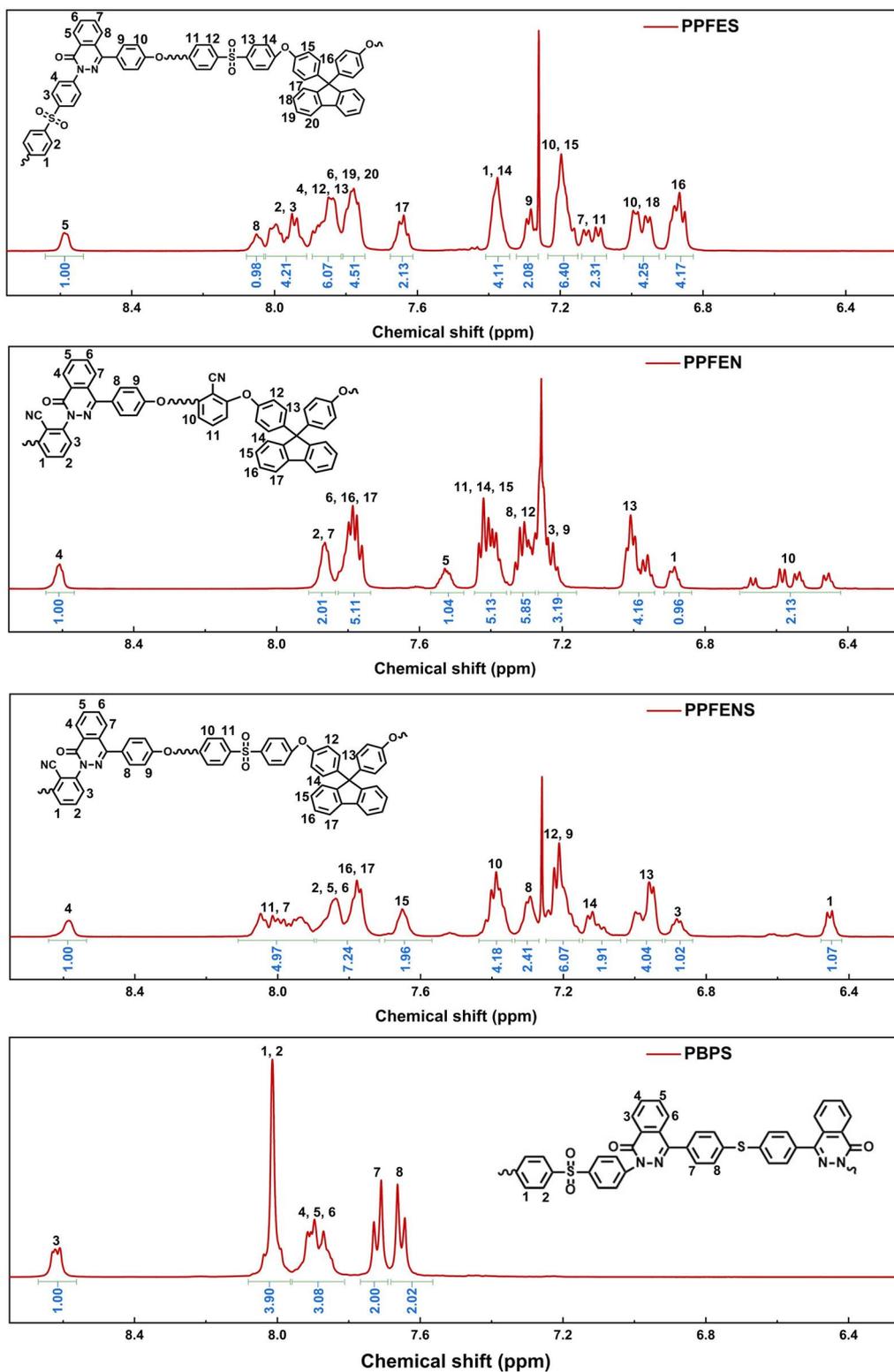
Figure S7. ^{1}H NMR spectrums of PPFES, PPFEN, PPFENS, and PBPS.
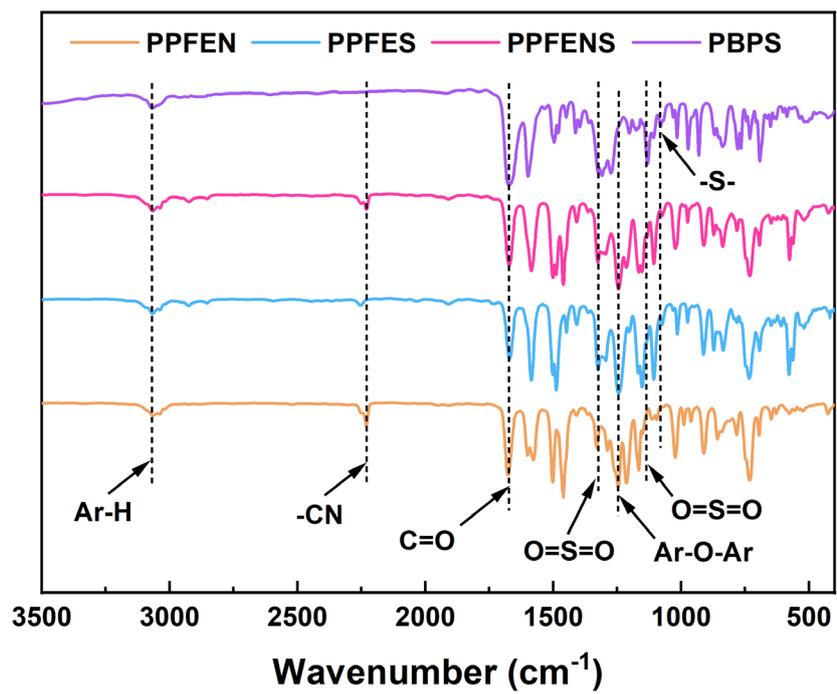
Figure S8. Fourier-transform infrared (FT-IR) spectra of PPFES, PPFEN, PPFENS, and PBPS.

References

[1] Zhu L, Zong L, Wang C, et al. Effect of di-halogen monomers embraced in main chain of low-dielectric colorless fluorene-based poly (aryl ether) s on their performance. Polymers for Advanced Technologies, 2022, 33(6): 1846-1854.