Supporting Information

Symmetry-Guided Monomer Design Enables the Combinatorial Synthesis and Targeted Screening of Polyesters

Xiaojie Feng[+][a], Xiaoying He[+][b], Jiayi Zhu[b], Li-Hong Lin[c], Qiaoyan Shang[b], Zheng-Hong Luo[c], Yin-Ning Zhou*[c], and Fangyou Yan*[a]

[a] X. Feng[+], F Yan

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

E-mail: yanfangyou@tust.edu.cn

[b] X. He[+], J. Zhu, Q. Shang

School of Marine and Environmental Science, Tianjin University of Science and Technology, Tianjin 300457, P.R. China.

[c] L.-H. Lin, Z.-H. Luo, Y.-N Zhou

State Key Laboratory of Synergistic Chem-Bio Synthesis, School of Chemistry and Chemical Engineering, Shanghai Jiao Tong University, Shanghai 200240, People's Republic of China.

E-mail: zhouyn@sjtu.edu.cn

[+] Both authors contributed equally to this work.

Analysis of polyesters from PoLyInfo

In our previous work, 695 collected polyesters were decomposed into 267 diacids and 358 diols. By removing the carboxyl groups from the diacids and the hydroxyl groups from the diols, and subsequently eliminating duplicates, a total of 460 unique small molecules were obtained. Structural analysis of these small molecules (Figure S1) reveals that the majority contain fewer than 50 heavy atoms, with 455 out of 460 molecules (98.91%) meeting this criterion. The molecular weights of all small molecules are below 1000, with the largest being 988.80. The maximum value of the maximum step (i.e., longest topological distance) is 37. Almost all small molecules contain fewer than 50 carbon (C) atoms, with only one molecule exceeding this number. Additionally, all small molecules contain fewer than 10 oxygen (O) and nitrogen (N) atoms.

Information distribution of the small molecules

Figure S2 shows the information distribution of the small molecules (H-suppressed structure) used for monomer design. Most molecules have moderate heavy atom counts and molecular weights, and contain 2-30 carbon atoms with few heteroatoms, reflecting chemical diversity while maintaining synthetic accessibility. The predominance of carbon-rich scaffolds aligns with typical polyester monomers, ensuring both feasibility for combination and providing structural variability for systematic exploration of the chemical space.

Pairwise atomic symmetry index

Algorithm S1 illustrates the computational procedure for calculating the pairwise atomic symmetry index (PASI). Where, Latom is a list of atomic symbols; Sa, SF, and Sbon represent the adjacency matrix, the topological distance matrix, and bond-order matrix, respectively;

Ladj_H is a list of hydrogen atoms for each atom in a molecule; notH and iH record the index of heavy atoms and hydrogen atoms; Sa_notH, SF_notH, and Sbon_notH stand for the Sa, SF, and Sbon for the hydrogen-suppressed structure, respectively; bra, ∑bds, and ∏bds denote the branched degree, the sum of bond orders, and the product of bond orders; Z records the atomic number of each atom, identifying the chemical element; Mpasi is a matrix of the PASI. Furthermore, the complete source code has been uploaded to the public GitHub repository (https://github.com/FangyouYan/PairwiseAtomicSymmetryIndex) for full transparency and reproducibility.

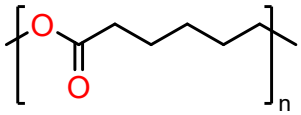
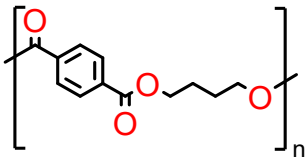
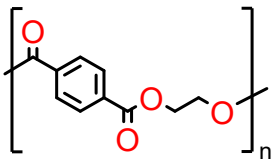
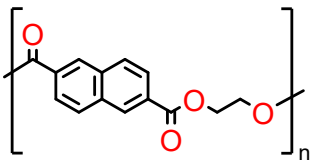
Algorithm S1 Generation of pairwise atomic symmetry index (PASI)

1: **input** molfile
2: (Latom, Sa, SF, Sbon, Ladj_H) ← **extract** molecular information
3: notH, iH ← **identify** heavy atoms and hydrogen atoms
4: Sa_notH, SF_notH, Sbon_notH ← **remove** hydrogen-related rows and columns
5: bra ← **calculate** branched degree
6: ∑bds ← **calculate** the sum of bond orders
7: ∏bds ← **calculate** the product of bond orders
8: Z ← **record** the atomic number
9: #H ← **record** the number of bonded hydrogens
10: initial_graph ← **construct** initial graph according to (SF, Z, bra, ∑bds, ∏bds, #H)
11: ascending_order_graph ← **sort** initial_graph in lexicographic order
12: for each atom pair (i, j), where i ≠ j:
13: $R_{i,j}$ ← **calculate** the identical information between i and j after ascending order
14: Mpasi[i,j] ← **calculate** the pairwise atomic symmetry index
15: return Mpasi

High-throughput screening of polyesters

To provide a clear reference for the T_g range of our designed polyesters, we compiled representative T_g values of commonly used commercial polyesters (Table S1). For example, poly-ε-caprolactone (PCL) exhibits a T_g of approximately -69.0 to -59.4 °C[1-3], that of polybutylene terephthalate (PBT) ranges from 32.0 to 41.0 °C[4-6], that of polyethylene terephthalate (PET) is 58.3-68.0 °C[7-9], and that of polyethylene naphthalate (PEN) is 114-132.6 °C[10-12]. The predicted T_g values reasonably match the experimentally reported ranges, further supporting the reliability of the T_g-QSPR model within conventional polyester chemistry. Moreover, the predicted T_g values of our designed polyesters span this entire range, encompassing both low-T_g flexible polyesters and high-T_g rigid polyesters. These results demonstrate that our design strategy can generate polyesters covering a broad chemical space and performance spectrum, providing potential candidates for diverse applications.

Table S1. T_g of commonly used commercial polyesters.

Name	Structure	Reported T_g (°C)[a]	Predicted T_g (°C)[b]
poly-ε-caprolactone (PCL)		-69.0[1] -62.6[2] -59.4[3]	-69.67
polybutylene terephthalate (PBT)		32.0[4] 39.85[5] 41.0[6]	25.84
polyethylene terephthalate (PET)		58.3-59.4[7] 61.0[8] 68.0[9]	51.84
polyethylene naphthalate (PEN)		114.0[10] 119.1[11] 132.6[12]	101.70

[a] Reported values were measured by differential scanning calorimetry (DSC)

[b] The used QSPR model is featured with R^2 = 0.9060, average absolute error (AAE) = 17.7197 °C, and cross-validation Q^2 = 0.8889

Molecular dynamics simulations

We employed the PCFF force field for molecular dynamics simulations due to its proven reliability in modeling chemically diverse polymers, as demonstrated in previous studies[13-16]. While more sophisticated all-atom or polarizable force fields could in principle capture additional

interactions (e.g., polarization effects), these models are computationally more demanding. In this work, the average absolute relative deviation (AARD, Eq. (S4)) between the MD-predicted and experimental values is 6.45%, and the maximum absolute relative deviation (ARD, Eq. (S3)) is 14.1%. These results indicate that the PCFF provides a reasonable balance between computational efficiency and accuracy, and the expected errors relative to all-atom or polarizable models are within acceptable limits for screening purposes.

Each amorphous cell contained ten polyester chains, with each chain comprising twenty repeating units.[17] The initial molecular configuration was first generated and employed as input for MD simulations in LAMMPS.[15] Before the dynamics simulation, energy minimization was carried out using the conjugate gradient algorithm to remove unfavorable contacts and relax internal stress. Subsequently, the system was equilibrated in the constant temperature and volume (NVT) ensemble at 800 K for 800 ps. To gradually release volumetric stress and approach a realistic density, a series of simulations was conducted in the constant temperature and pressure (NPT) ensemble under descending pressures of 1000 atm, 100 atm, 10 atm, and finally 1 atm, with each stage lasting 1000 ps.

To further optimize chain conformations and packing, the system underwent ten annealing cycles. Each cycle consisted of a 500 ps cooling run from 800 K to 300 K, followed by a 500 ps heating run from 300 K back to 800 K. All of these steps were performed under the NVT ensemble. After annealing, the T_g was investigated through a series of simulations at temperatures ranging from ($T_{g,pred}$ + 150) K to ($T_{g,pred}$ - 150) K in 15 K intervals, where T_g refers to the T_g value predicted by the T_g-QSPR model[18]. At each temperature, the system was equilibrated via an 800 ps NVT run followed by a 1000 ps NPT simulation at 1 atm. The T_g value

was determined by analyzing the temperature-dependent density profile using linear fitting methods.

All simulations were conducted with a timestep of 1.0 fs. Van der Waals and Coulombic interactions were calculated using the lj/class2/coul/long pair style with a cutoff of 9.5 Å. Long-range electrostatic interactions were treated using the particle-particle particle-mesh (PPPM) algorithm with an accuracy of 1.0×10^{-4}. The neighbor list was updated every timestep, and atomic trajectories were saved every 1000 steps for subsequent structural and thermodynamic analysis.

Figure S3 shows the temperature-density relationship obtained from molecular dynamics simulations. For several polyesters, the first one or two data points at the highest temperatures were excluded from the analysis. These initial points correspond to the early equilibration stage, during which the system may not have fully relaxed, resulting in transient density fluctuations. The excluded points are marked with red "×" symbols in the graph.

Experimental synthesis and characterization of polyesters

Polyesters were synthesized by a two-step polycondensation process.[19] First, the titanium butoxide catalyst (0.1 mol%, based on the diacid) was added to a three-neck round-bottom flask (250 mL) together with the requisite molar ratio of diacid and diol. The mixture was then esterified in an oil bath at 190-230 °C for 2-3 hours. The entire reaction process was carried out under nitrogen (N_2) protection at a stirring speed of 100 rpm to prevent oxidation. The esterification stage was considered complete when the amount of water produced reached 90% of the theoretical value. Subsequently, the temperature was raised to 240-260 °C, and the system pressure was gradually reduced to below 100 Pa. Under these conditions,

polycondensation was performed for 3-4 hours. Finally, the pressure was restored to atmospheric pressure by introducing N_2, yielding the desired product. It should be noted that some of the synthesized polyesters were highly viscous and partially adhered to the walls of the reaction vessel, preventing complete recovery and accurate yield determination. According to our observations throughout the experimental process and experiment record, these polyesters were generally obtained in yields of 85-95%.

The thermal transition behavior of the polyester was characterized using differential scanning calorimetry (DSC) to determine its T_g. The DSC was performed at temperatures ranging from -70 to 200 °C. The measurement procedure was as follows: (a) The sample was heated to a specified temperature under an N_2 atmosphere at a rate of 20 °C/min and held for 3 minutes. (b) The sample was then cooled at a rate of 20 °C/min to a second specified temperature and maintained for three minutes. (c) The sample was then reheated to the specified temperature at the same heating rate. The temperature range for each polyester was set in reference to the predicted value.

Equations

Absolute error:

$$\mathrm{AE}_i = \left| X_i^{\mathrm{pred}} - X_i^{\mathrm{exp}} \right| \qquad \text{S(1)}$$

Average absolute error:

$$\mathrm{AAE} = \frac{\sum \left| X_i^{\mathrm{pred}} - X_i^{\mathrm{exp}} \right|}{N} \qquad \text{S(2)}$$

where, X_i^{exp} (°C) and X_i^{pred} (°C) are the experimental and predicted values of the glass transition temperature (T_g) for polyester i, and N is the number of polyesters.

Absolute relative deviation:

$$\mathrm{ARD}_i = \frac{\left|T_i^{\mathrm{MD}} - T_i^{\mathrm{exp}}\right|}{T_i^{\mathrm{exp}}} \times 100\% \qquad \text{S(3)}$$

Average absolute relative deviation:

$$\mathrm{AARD} = \frac{1}{N}\sum_{i=1}^{N} \frac{\left|T_i^{\mathrm{MD}} - T_i^{\mathrm{exp}}\right|}{T_i^{\mathrm{exp}}} \times 100\% \qquad \text{S(4)}$$

where, T_i^{MD} (K) and T_i^{exp} (K) represent the MD-predicted and experimental values for the i-th sample, respectively, and N is the total number of samples.

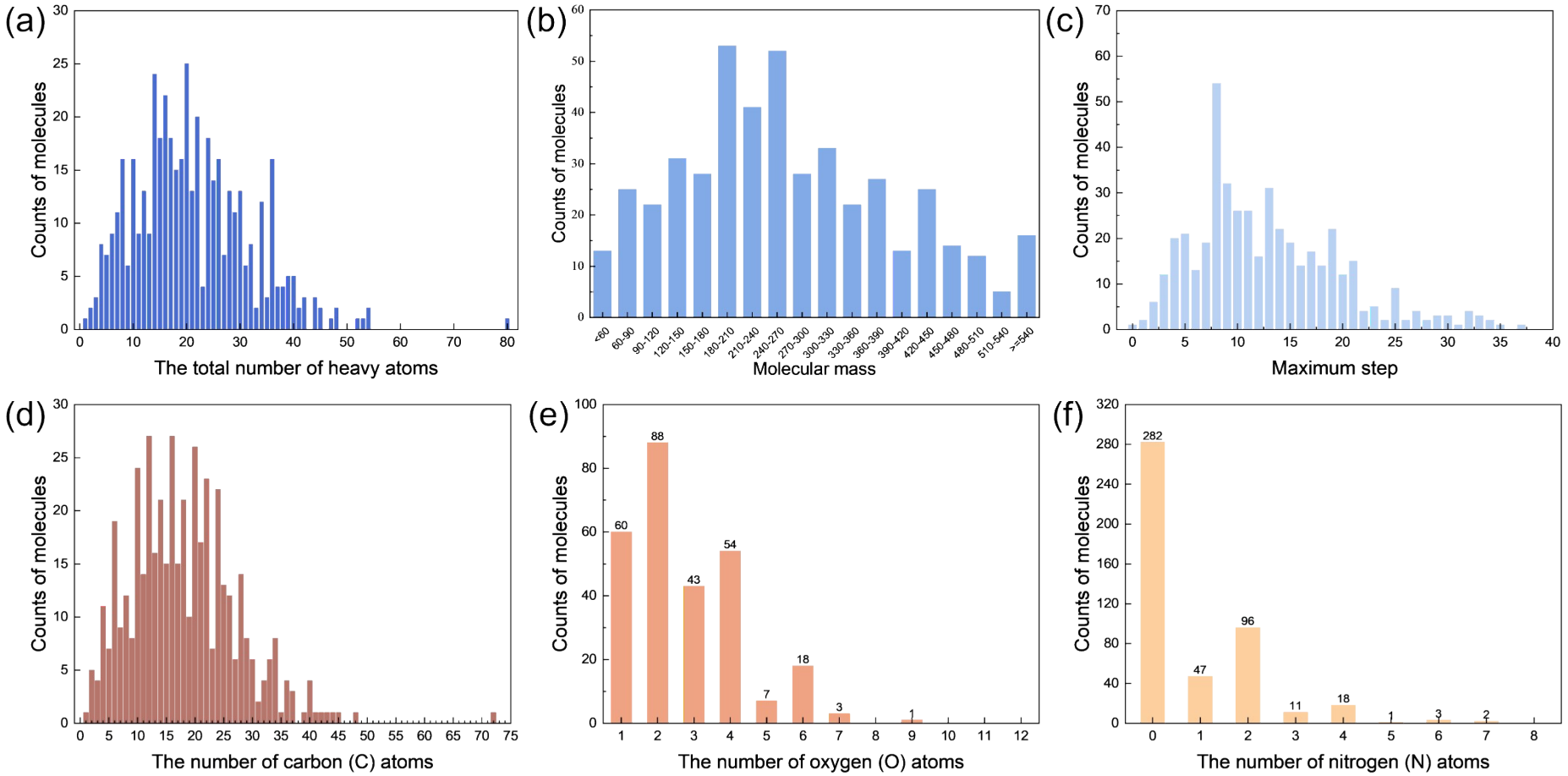


Figure S1. Information distribution of the 460 unique small molecules decomposed from 695 polyesters collected from PoLyInfo (H-suppressed structure). (a) Counts of molecules by heavy atom number, (b) counts of molecules by molecular mass, (c) counts of molecules by maximum step, (d) counts of molecules by the number of carbon (C) atoms, (e) counts of molecules by the number of oxygen (O) atoms, and (f) counts of molecules by the number of nitrogen (N) atoms.

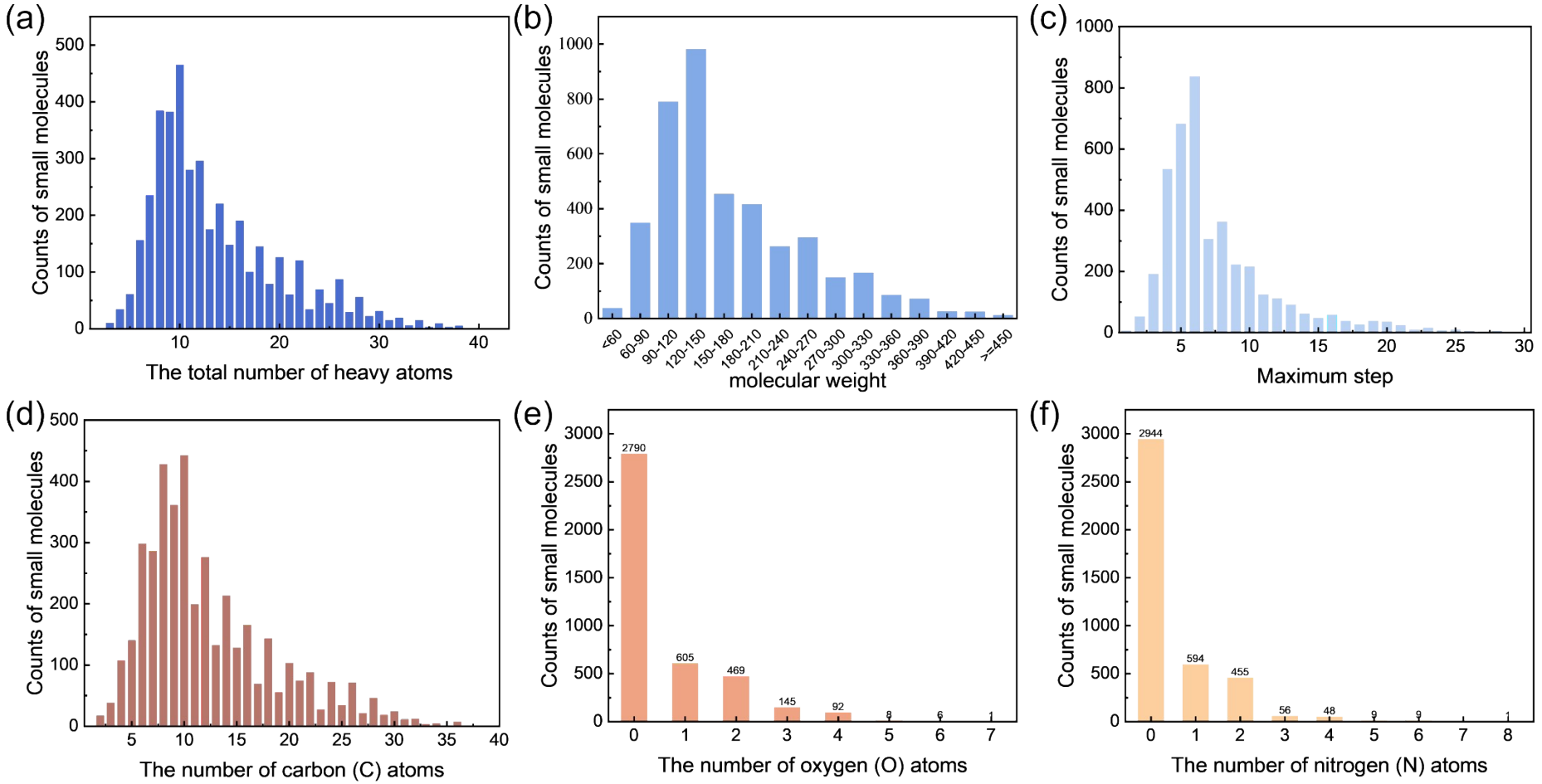


Figure S2. Information distribution of 4,116 small molecules used this work (H-suppressed structure). (a) Counts of molecules by heavy atom number. (b) Counts of molecules by molecular weight, (c) Counts of molecules by maximum step, (d) Counts of molecules by the number of carbon (C) atoms, (e) Counts of molecules by the number of oxygen (O) atoms, and (f) Counts of molecules by the number of nitrogen (N) atoms.

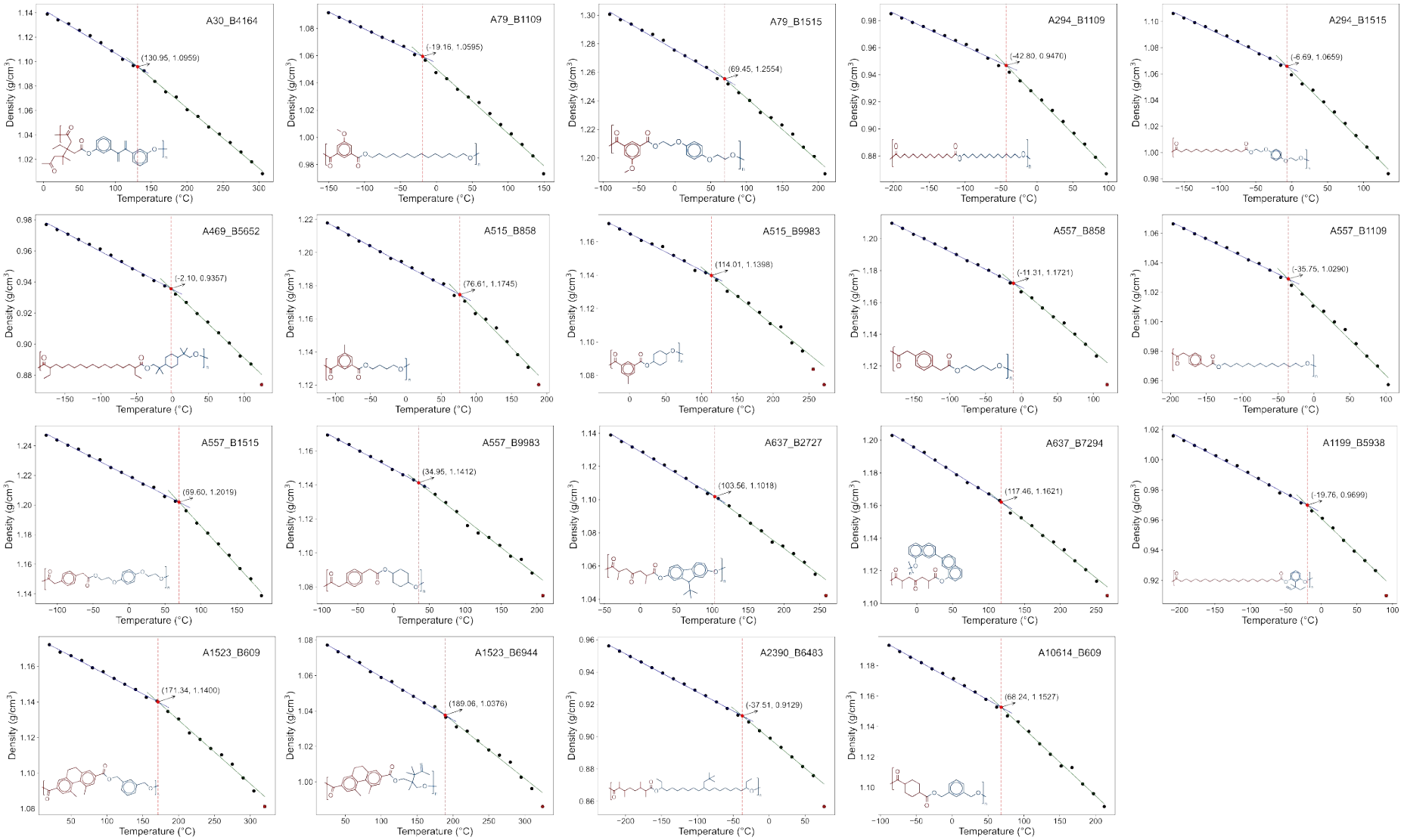


Figure S3. Temperature–density relationships obtained from molecular dynamics simulations for the selected polyesters.

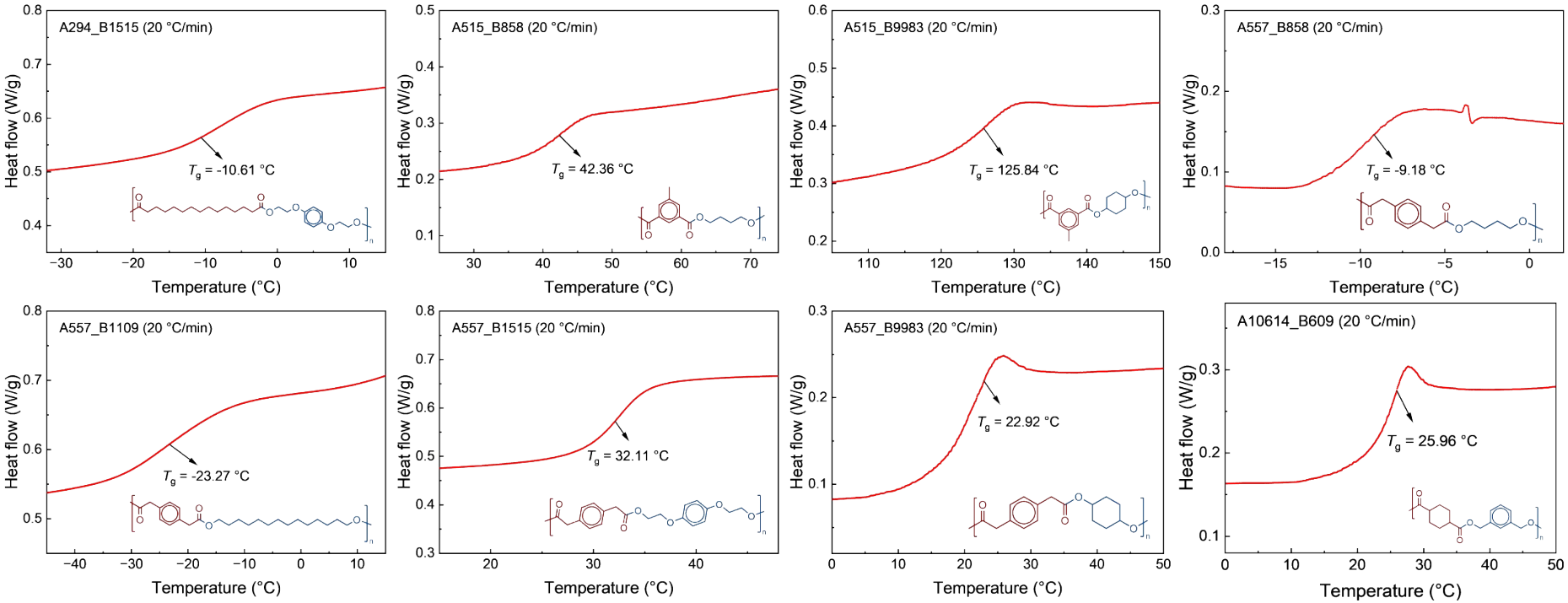


Figure S4. DSC curves of eight polyesters synthesized in this work.

References

[1] S. Li, M. Pignol, F. Gasc, M. Vert, *Macromolecules* **2004**, *37*, 9798-9803.
[2] S. Miura, N. Teramoto, M. Shibata, *J. Polym. Res.* **2016**, *23*, 19.
[3] M. Hong, X. Tang, B. S. Newell, E. Y. X. Chen, *Macromolecules* **2017**, *50*, 8469-8479.
[4] F. Gómez, R. Quintana, A. M. De Ilarduya, E. Rudé, S. Muñoz-Guerra, *J. Polym. Sci., Part A:Polym. Chem.* **2005**, *43*, 92-100.
[5] Y. Higaki, H. Kabayama, D. Tao, A. Takahara, *Macromol. Chem. Phys.* **2015**, *216*, 1103-1108.
[6] C. Lei, D. Chen, Y. Gou, W. Huang, *J. Appl. Polym. Sci.* **2009**, *113*, 87-95.
[7] B. Demirel, *Polym. Test.* **2017**, *60*, 220-228.
[8] J.-P. He, H.-M. Li, X.-Y. Wang, Y. Gao, *Eur. Polym. J.* **2006**, *42*, 1128-1134.
[9] T. Chen, J. Zhang, H. You, *RSC Adv.* **2016**, *6*, 102778-102790.
[10] F. Mebarki, E. David, *Polym. Eng. Sci.* **2018**, *58*, 701-712.
[11] J. Y. Kim, S. I. Han, D. K. Kim, S. H. Kim, *Compos. Part A: Appl. Sci. Manuf.* **2009**, *40*, 45-53.
[12] S. Can, N. G. Karsli, S. Yesil, A. Aytac, *J. Polym. Eng.* **2016**, *36*, 615-624.
[13] K. K. Bejagam, C. N. Iverson, B. L. Marrone, G. Pilania, *Phys. Chem. Chem. Phys.* **2020**, *22*, 17880-17889.
[14] J. L. Suter, W. A. Müller, M. Vassaux, A. Anastasiou, M. Simmons, D. Tilbrook, P. V. Coveney, *J. Chem. Theory Comput.* **2025**, *21*, 1405-1421.
[15] L.-H. Lin, J.-J. Li, Y.-X. Pan, F. Yan, Z.-H. Luo, Y.-N. Zhou, *ACS Appl. Mater. Interfaces* **2025**, *17*, 55347-55359.
[16] J. Zhang, A. Fan, Z. Wang, L. Zheng, Y. Liu, Z. Wang, P. Kang, *J. Phys. Chem. B* **2025**, *129*, 8821-8831.
[17] M. Yu, Q. Jia, Q. Wang, Z.-H. Luo, F. Yan, Y.-N. Zhou, *Chem. Sci.* **2024**, *15*, 18099-18110.
[18] X. He, M. Yu, J.-P. Han, J. Jiang, Q. Jia, Q. Wang, Z.-H. Luo, F. Yan, Y.-N. Zhou, *AIChE J.* **2024**, *70*, e18409.
[19] Y. Zhu, J. Wang, S. Mahmud, X. Zhang, X. Liu, J. Zhu, *Eur. Polym. J.* **2022**, *181*, 111654.