Supporting Information for

AI-assisted inverse design of sequence-ordered high intrinsic thermal conductivity polymers

Xiang Huang[a], C. Y. Zhao[a], Hong Wang[b], and Shenghong Ju[a, b, *]
[a] China-UK Low Carbon College, Shanghai Jiao Tong University, Shanghai, China
[b] Materials Genome Initiative Center, School of Material Science and Engineering, Shanghai Jiao Tong University, Shanghai, China

* Corresponding email: shenghong.ju@sjtu.edu.cn

This PDF file includes:

Supporting text
Figures S1 to S14
Tables S1 and S2
SI References

Supporting Information Text

A. Morgan fingerprints with frequency (MFF) for polymer representation

Morgan fingerprints with frequency (MFF) is an expansion of Morgan fingerprints to overcome the high dimensional limitations of vectors [1]. In the Morgan algorithm, different atoms (depending on the type and nearest neighbors under a predefined radius) are given unique hashed identifiers (bit vectors). Thus, each identifier corresponds to a specified substructure. For the benchmark dataset, we can count the number and frequency of different identifiers and set a frequency threshold to determine the composition of the MFF vector. In this work, we counted 6926 chemical substructures with the radius of 3 from the repeating units of 1144 polymers, of which the 194 most popular substructures with a frequency no less than 1100 times were retained as input features. For example, a monomer with PI813 (Figure S1a) has 32 types of substructures, where parts of which are shown in Figure S1b. Moreover, 15 substructures are valid, corresponding to the 194 most frequent segments. In Figure S1c, the positions of the 15 substructures were sequentially assigned frequencies (integers), i.e., the MFF for the PI813.

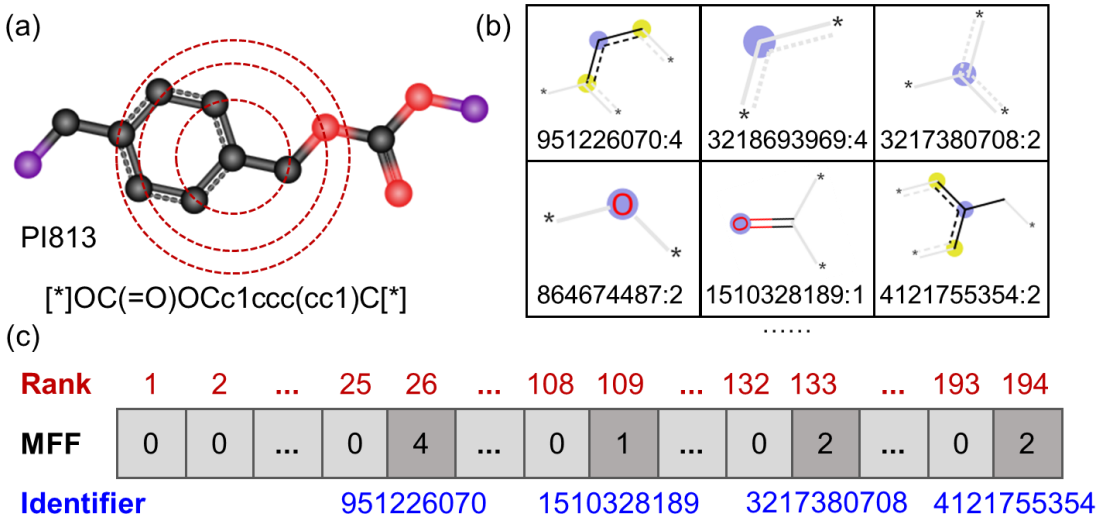

Fig. S1. Example of MFF generation. (a) Repeating unit of polymer with ID of PI813. (b) Library of polymer substructures. (c) MFF vector

B. Machine learning models trained by 1144 polymers with known thermal conductivity

We evaluated the performance of five machine learning (ML) algorithms of deep neural networks (DNN), random forests (RF), eXtreme gradient boosting (XGBoost), multilayer perceptron (MLP), and Gaussian process regression (GPR) separately. The RF, XGBoost and MLP were executed in the Scikit-learn repository with 10-fold cross-validation [2], and the GPR was repeated 10 runs using different training sets in the Gpytorch toolkit [3]. Figure S2a statistics of polymer thermal conductivity (TC) distribution based on Gaussian kernel density estimation. The thermal conductivity of polymers is mainly distributed in the range of 0.1-0.4 W $m^{-1}K^{-1}$, and achieving a high thermal conductivity is quite difficult, with only 4.63%. The root-mean-square error (RMSE) and R-square (R^2) for five ML models are shown in Figure S2b and c. Overall, the performance of the five models is comparable, but the DNN model is more stable in 10 repetitions of the trial.

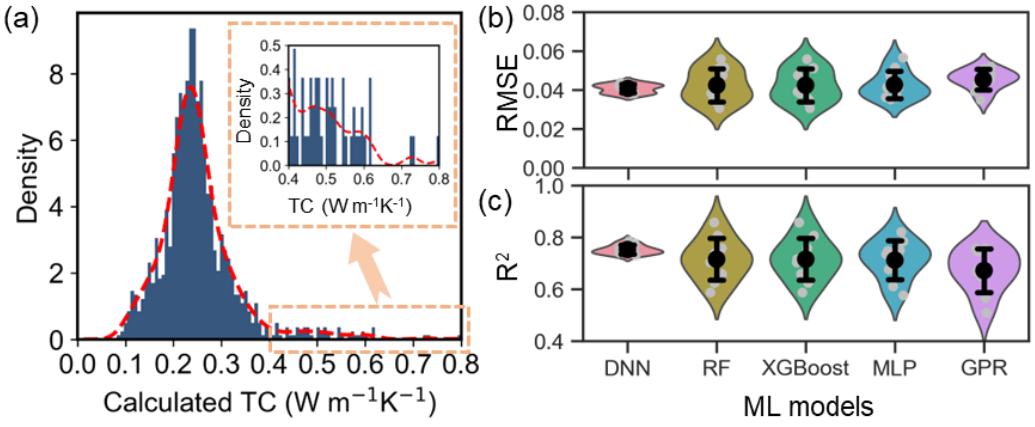

Fig. S2. Evaluation of machine learning models trained by 1144 polymer data. (a) Kernel density estimate plot visualizes the distribution of TC among 1144 polymers. (b) and (c) Root-mean-square error (RMSE) and R-square (R^2) for deep neural networks (DNN), random forests (RF), eXtreme gradient boosting (XGBoost), multilayer perceptron (MLP), as well as Gaussian process regression (GPR) models. To acquire statistical results, each model was repeated 10 runs using different training sets. Violins represent the distributions of the subsampling results, mean and standard deviation of MSE are shown in black, and individual subsample results are in gray.

The training and test R^2 of the surrogate DNN model are 0.95 and 0.79, respectively. We performed an additional five-fold cross-validation (CV) to evaluate the accuracy of the DNN model, as shown in Figure S3. In the five-fold cross-validation, the training R^2 of the DNN models ranged from 0.90 to 0.97, and the test R^2 ranged from 0.67 to 0.76. The accuracy of the surrogate prediction model utilized in this work is basically consistent with the results from five-fold cross-validation, reflecting the fact that the prediction model is reliable.

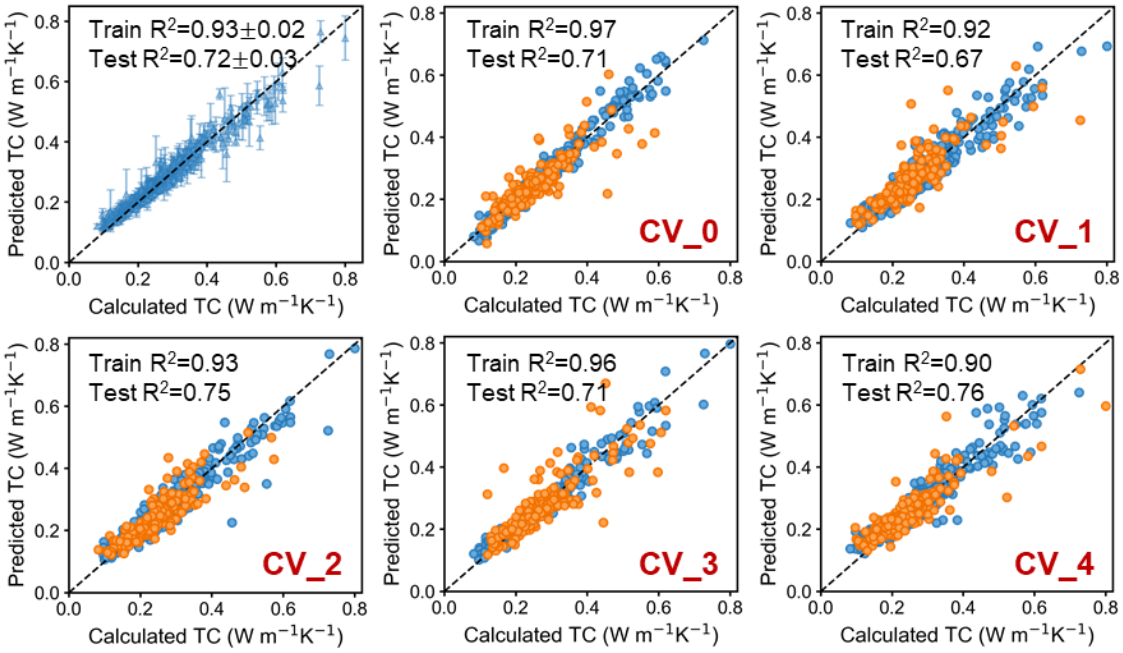

Fig. S3. Performance evaluation of DNN models based on five-fold cross-validation

C. Demonstration of known high thermal conductive polymers and construction of polymer-unit library

Among the 1144 training data, 53 polymers have thermal conductivities exceeding 0.40 W $m^{-1}K^{-1}$, as listed in Table S1 and shown in Figure S4. Combining the structural characteristics of highly thermal conductive polymers and the outcomes from the shapley additive explanations (SHAP) [4], a polymer-unit library was generated as shown in Figure S5.

Table S1. List of amorphous polymers with high intrinsic thermal conductivity (TC > 0.40 W $m^{-1}K^{-1}$), where the corresponding monomer structures are illustrated in Figure S4.

ID	SMILES	TC (W $m^{-1}K^{-1}$)
AP1	[*]c1ccc(-c2nc3cc4nc([*])[nH]c4cc3[nH]2)cc1	0.800
AP2	[*]c1ccc(-c2ccc(-c3nc4cc5nc([*])oc5cc4o3)cc2)cc1	0.729
AP3	[*]c1ccc(-n2c(=O)c3cc4c(=O)n([*])c(=O)c4cc3c2=O)cc1	0.725
AP4	[*]c1ccc(-n2c(=O)c3cc4c(=O)n([*])c(=O)c4cc3c2=O)c(C)c1	0.619
AP5	[*]c1ccc2[nH]c([*])nc2c1	0.619
AP6	[*]c1ccc(-c2nc3cc4nc([*])oc4cc3o2)c(O)c1	0.618
AP7	[*]NC(=O)C=CC(=O)Nc1nc([*])nc(N)n1	0.606
AP8	[*]c1ccc([*])s1	0.597
AP9	[*]C1=CC2=NC([*])=CC2=N1	0.594
AP10	[*]Nc1ccc(C#Cc2ccc(NC(=O)c3ccc(C([*])=O)cc3)cc2)cc1	0.588
AP11	[*]c1ccc(-c2ccc(-n3c(=O)c4cc5c(=O)n([*])c(=O)c5cc4c3=O)c(C)c2)cc1C	0.581
AP12	[*]NC(=O)c1ccc(cc1)C(=O)Nc1ccc(cc1)c1ccc(cc1)[*]	0.576
AP13	[*]/C=C\[*]	0.573
AP14	[*]c1ccc2c(c1)Cc1cc(-n3c(=O)c4cc5c(=O)n([*])c(=O)c5cc4c3=O)ccc1-2	0.566
AP15	[*]c1ccc2cc(-c3nc4ccc(-c5ccc6nc([*])[nH]c6c5)cc4[nH]3)ccc2c1	0.553
AP16	[*]c1nc2cc3nc(-c4ccc([*])o4)[nH]c3cc2[nH]1	0.547
AP17	[*]Nc1ccc(NC(=O)c2ccc(C([*])=O)cc2)cc1	0.545
AP18	[*]c1ccc(-c2ccc(-c3nc4ccc(-c5ccc6nc([*])oc6c5)cc4o3)cc2)cc1	0.542
AP19	[*]Nc1ccc(C([*])=O)cc1	0.527
AP20	[*]C=C([*])F	0.522
AP21	[*]c1ccc([*])[nH]1	0.517
AP22	[*]c1ccc(-c2ccc(-n3c(=O)c4cc5c(=O)n([*])c(=O)c5cc4c3=O)c(OC)c2)cc1OC	0.515
AP23	[*]c1c(C)c(C)c(-n2c(=O)c3cc4c(=O)n([*])c(=O)c4cc3c2=O)c(C)c1C	0.514
AP24	[*]c1ccc(-c2ccc(-c3ccc(N4C(=O)c5ccc(-c6ccc7c(c6)C(=O)N([*])C7=O)cc5C4=O)cc3)cc2)cc1	0.511
AP25	[*]N1C(=O)c2c(C1=O)cc(cc2)c1cc2c(C(=O)N(C2=O)c2ccc(cc2)c2ccc(cc2)[*])cc1	0.509
AP26	[*]NNC(=O)C([*])=O	0.504
AP27	[*]C(O)C([*])O	0.503
AP28	[*]c1ccc(C(=O)Nc2ccc(-n3c(=O)c4cc5c(=O)n([*])c(=O)c5cc4c3=O)cc2)cc1	0.501
AP29	[*]NC(=O)c1ccc(C(=O)Nc2cnc([*])nc2)cc1	0.491
AP30	[*]c1ccc(N2C(=O)c3ccc(-c4ccc5c(c4)C(=O)N([*])C5=O)cc3C2=O)nc1	0.483
AP31	[*]c1ccc(N2C(=O)c3ccc(-c4ccc5c(c4)C(=O)N([*])C5=O)cc3C2=O)cc1	0.482
AP32	[*]Nc1ccc(NC(=O)C=CC([*])=O)cc1	0.479
AP33	[*]NC(=O)c1ccc(cc1)C(=O)Nc1ccc(cc1)[*]	0.473
AP34	[*]c1ccc(-c2nc3cc(-n4c(=O)c5cc6c(=O)n([*])c(=O)c6cc5c4=O)ccc3[nH]2)cc1	0.472

ID	SMILES	TC (W $m^{-1}K^{-1}$)
AP35	[*]CNC(=O)N[*]	0.470
AP36	[*]c1ccc(NC(=O)c2ccc(C(=O)Nc3ccc(-n4c(=O)c5cc6c(=O)n([*])c(=O)c6cc5c4=O)cc3)cc2)cc1	0.468
AP37	[*]c1ccc(-c2ccc(N3C(=O)c4ccc(-c5ccc6c(c5)C(=O)N([*])C6=O)cc4C3=O)c(C)c2)cc1C	0.463
AP38	[*]c1ccc2c(c1)Cc1cc(N3C(=O)c4ccc(-c5ccc6c(c5)C(=O)N([*])C6=O)cc4C3=O)ccc1-2	0.461
AP39	[*]NNC(=O)C=CC(=O)Nc1ccc(NC(=O)C=CC([*])=O)cc1	0.460
AP40	[*]CC[*]	0.456
AP41	[*]Nc1nnc(Nc2n[nH]c([*])n2)[nH]1	0.451
AP42	[*]CC([*])C(N)=O	0.445
AP43	[*]C(=O)Nc1ccc(NC(=O)c2ccc(N3C(=O)c4ccc([*])cc4C3=O)cc2)cc1	0.440
AP44	[*]C(C[*])O	0.439
AP45	[*]c1ccc2oc([*])nc2c1	0.436
AP46	[*]c1ccc(Oc2ccc(N3C(=O)c4cc5C(=O)N([*])C(=O)c5cc4C3=O)cc2)cc1	0.425
AP47	[*]c1ccc(Oc2ccc(-c3ccc(Oc4ccc(-n5c(=O)c6cc7c(=O)n([*])c(=O)c7cc6c5=O)cc4)cc3)cc2)cc1	0.419
AP48	[*]NNC(=O)c1ccc(-c2ccc(C(=O)NNC(=O)c3cccc(C([*])=O)c3)cc2)cc1	0.419
AP49	[*]c1ccc(Nc2ccc(-n3c(=O)c4cc5c(=O)n([*])c(=O)c5cc4c3=O)cc2)cc1	0.419
AP50	[*]C(=O)NNC(=O)c1ccc([*])nc1	0.415
AP51	[*]c1ccc(-c2nc3cc(-c4ccc5oc([*])nc5c4)ccc3o2)cc1	0.410
AP52	[*]c1ccc(C(=O)Oc2ccc(-n3c(=O)c4cc5c(=O)n([*])c(=O)c5cc4c3=O)cc2)cc1	0.403
AP53	[*]c1ccc(-c2ccc(N3C(=O)c4ccc(-c5ccc6c(c5)C(=O)N([*])C6=O)cc4C3=O)c(OC)c2)cc1OC	0.401

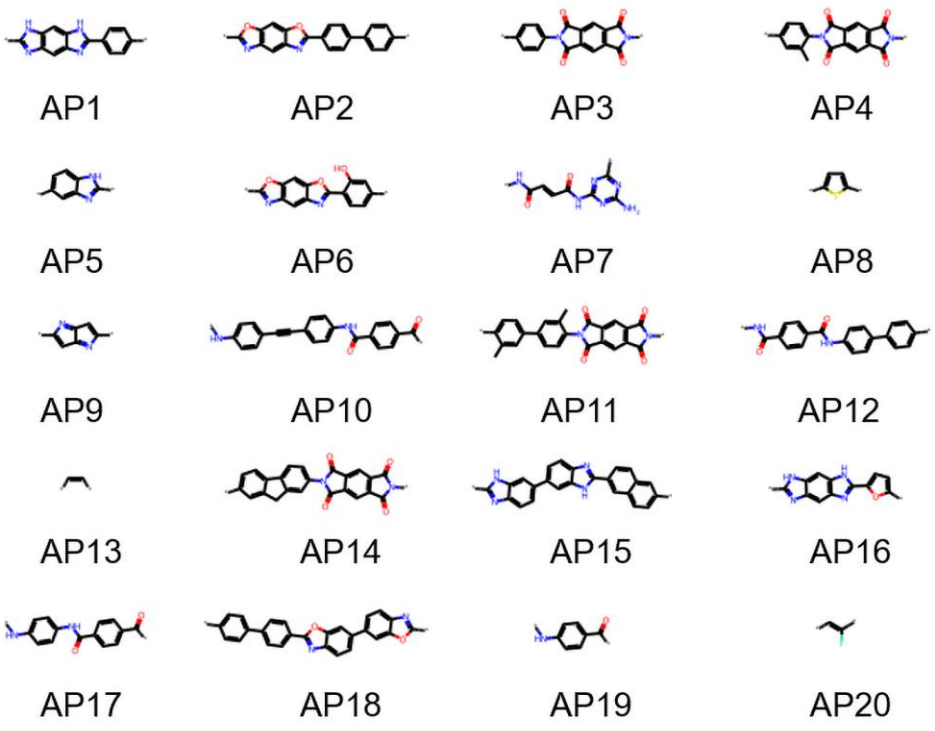

Fig. S4. Continued.

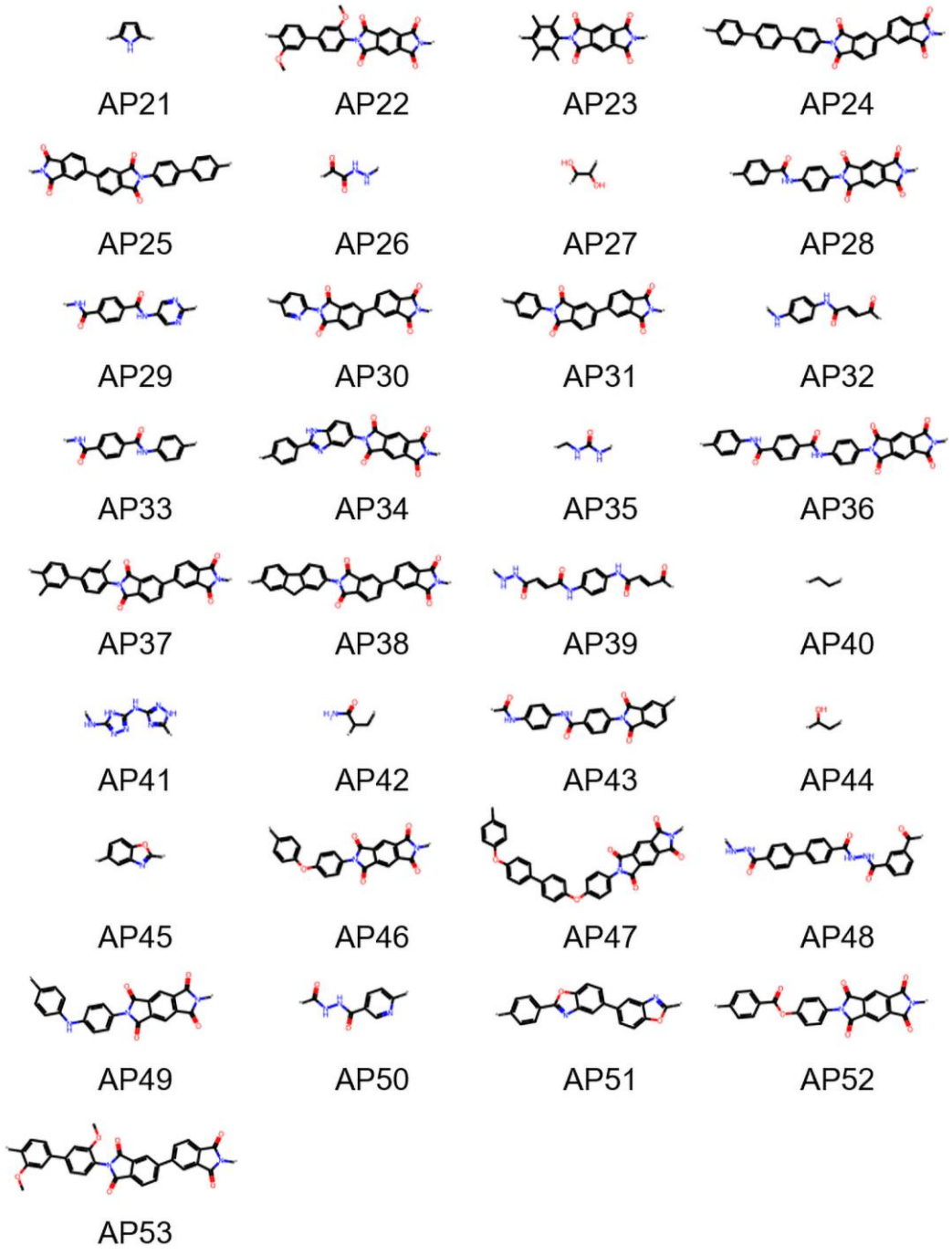

Fig. S4. Structures of polymer repeating units with thermal conductivity greater than 0.40 W/(mK).

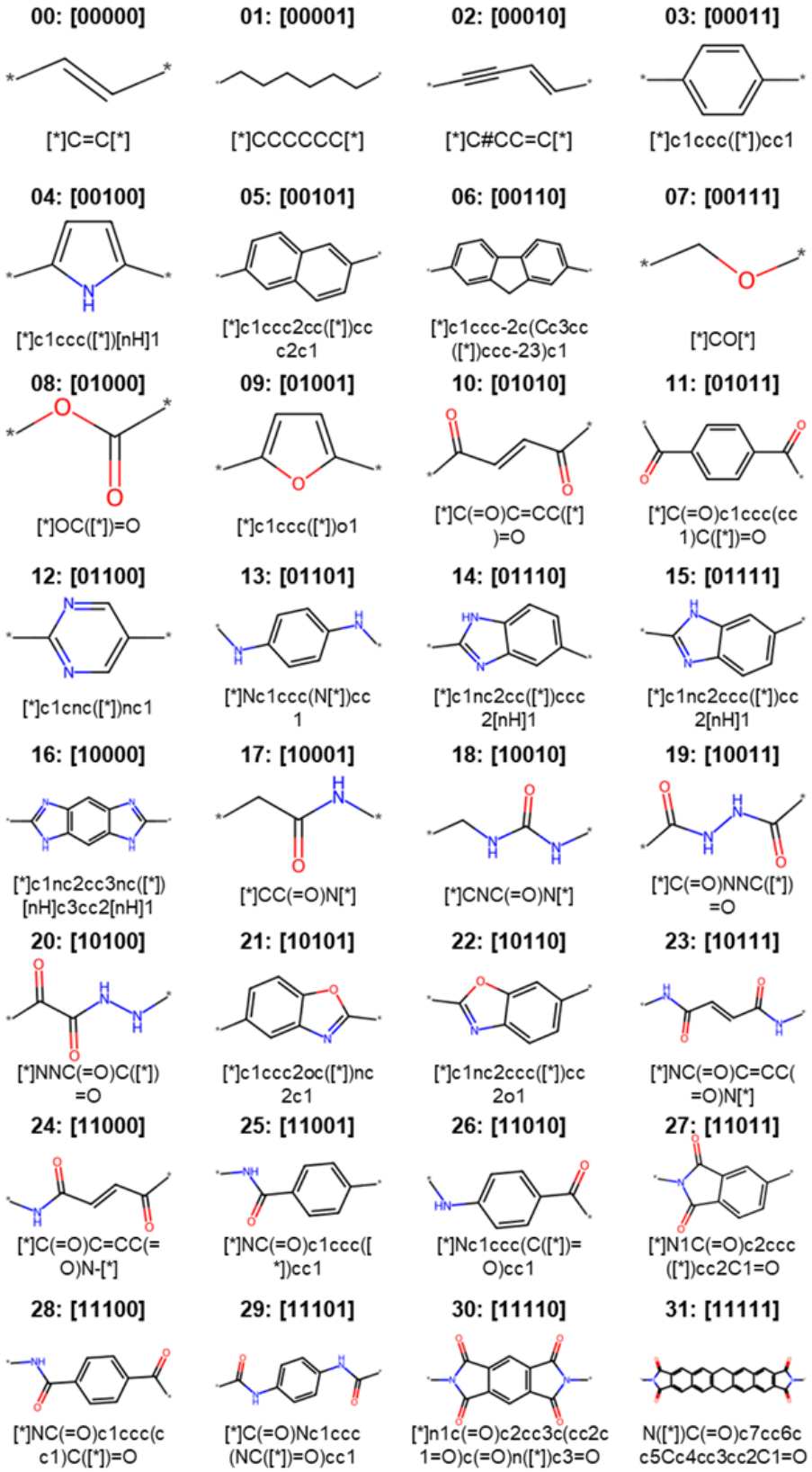

Fig. S5. Structures of 32 polymer fragments as basic units for high thermal conductivity polymer design.

D. Convergence ability assessment of MOEA and MOBO algorithms

For obtaining statistical results we performed 20 runs of the MOEA and MOBO algorithms with different initial candidates, respectively, and the HV convergence curves are displayed in Figure S6a-b. Hypervolumes (HVs) of U-NSGA-III can rapidly rise to a certain level (within 20 generations), but it is difficult to increase again in subsequent. However, there are three qNEHVI runs that identified nine global optimal polymers within 200 generations and almost all of the HVs get a secondary boost after the first time to a certain level. The difference in this enhancement depends on the stochastic nature of QMC sampling. All the HVs of optimization algorithms reach a referred value that is calculated by the five ideal global optimal Pareto polymers and the referred point, although the mean HV of MOBO is greater than that of MOEA (see Figure S6c-d).

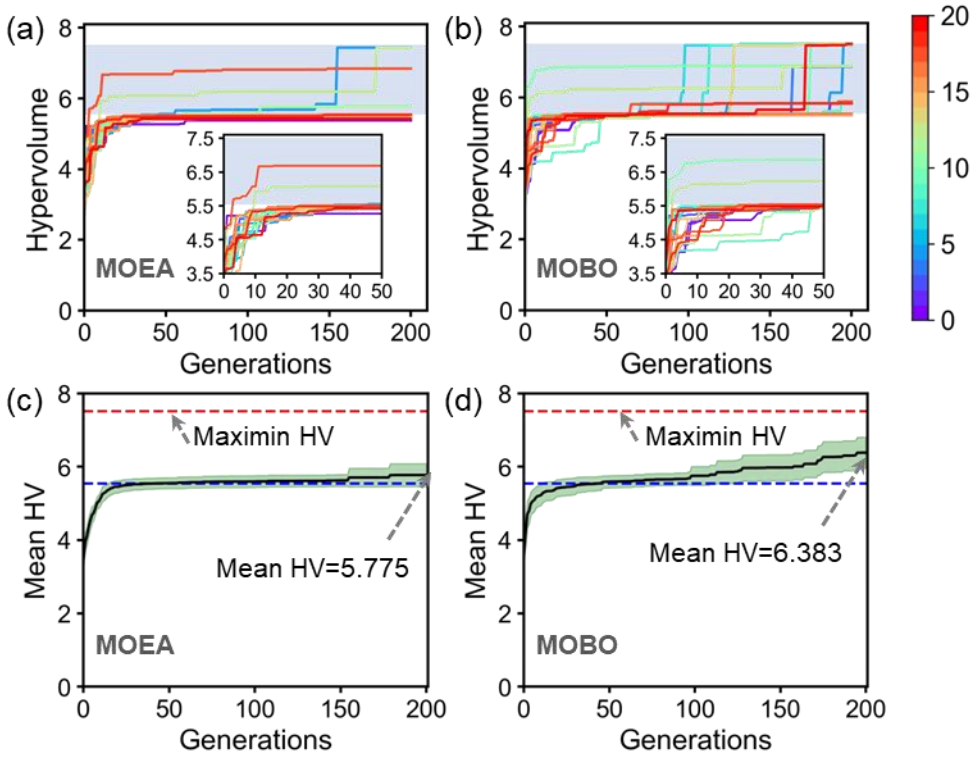

Fig. S6. Comparison of multi-objective evolutionary algorithm (MOEA) and multi-objective Bayesian optimization (MOBO) in triblock high thermal conductivity polymers inverse design. (a) and (b) Convergence curves for 20 runs of MOEA and MOBO. Each optimization run with 10 random initial structures and 200 iterations × 10 candidates per batch. (c) and (d) Mean hypervolume curves for 20 MOEA and MOBO runs. The upper edge of the blue strip or the red dashed line corresponds to the global optimal HV, and the lower edge of the blue strip or the blue dashed line indicates the HV computed from the five ideal global optimal Pareto polymers with the reference points

E. Impact of initial structures on MOEA convergence performance

Starting from different initial structures, the convergence level of the HV curves of MOEAs varies widely, as shown in Fig. S7a. The red, green and blue lines correspond to 3 different sets of initial structures with random seeds of 12, 17 and 19, respectively. Each MOEA run has 10 initial structures and 10 candidates × 200 generations. Their optimization trajectories are displayed in Figure S7b~d. We noticed that various initial structures lead to differences in the optimization direction of the polymers. Among them, the initial structures controlled by seeds 12 and 17 are able to converge towards the entire Pareto front, while MOEA with seed 19 is only close to the four ideal Pareto polymers (TC > 0.40 W $m^{-1}K^{-1}$ and SA < 3.0). Therefore, MOEAs with seeds 12 and 17 have relatively large HVs (7.43 and 6.84) after 200 optimization iterations, while MOEA with seed 19 is only 5.43.

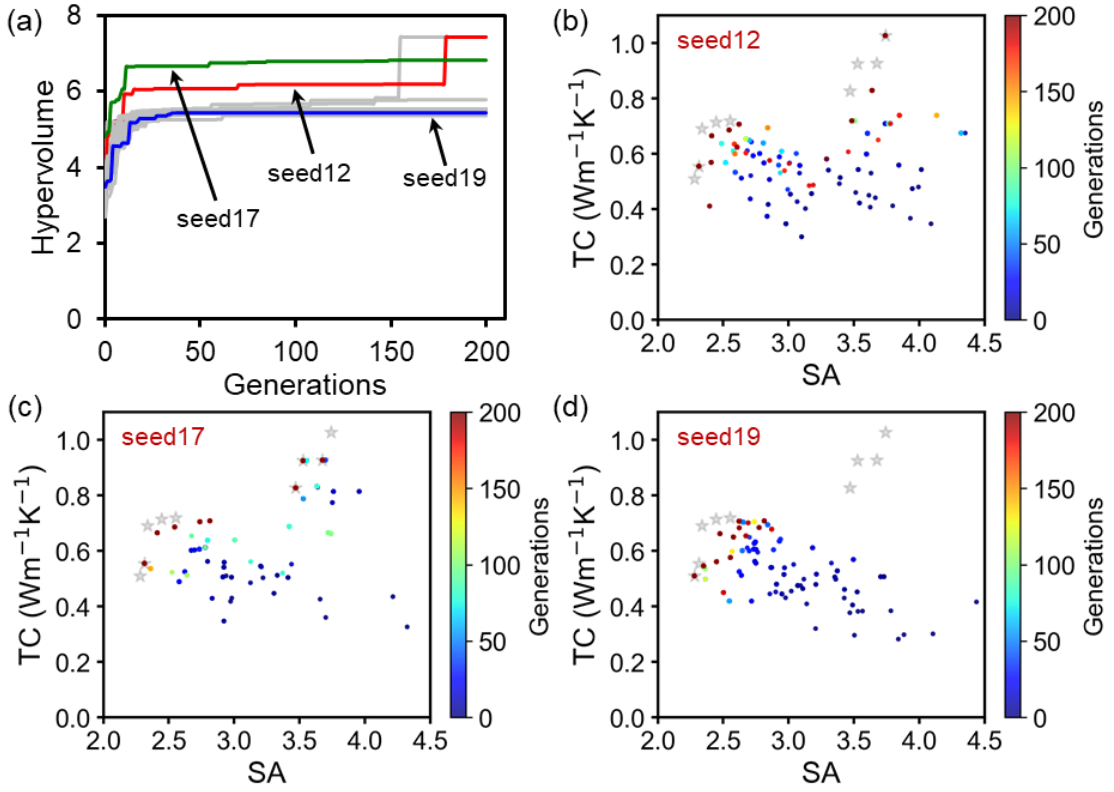

Fig. S7. Effect of initial structures on MOEA behavior. (a) HV convergence curves for MOEAs with different initial structures, where red, green, and blue lines correspond to MOEAs trained on the initial structures generated by the three random seeds of 12, 17 and 19, respectively. Their optimization trajectories are plotted in (b)~(d).

To demonstrate this point more intuitively, we counted the chemical blocks at different MOEA optimization stages, including 10 initial polymers with seed 12/17/19, and local Pareto polymers after 100 iterations and 200 iterations, as shown in Figure S8~S10. Moreover, chemical blocks of global Pareto polymers across the whole triblock polymers are presented for comparison. The 10 initial structures have diverse chemical blocks. Along with the optimization iterations, the types of chemical blocks of local Pareto polymers decrease and aggregate towards some promising blocks. In addition, chemical blocks in initial structures cover more blocks involved in the global Pareto polymers, which contributes to the convergence efficiency and capacity of MOEA.

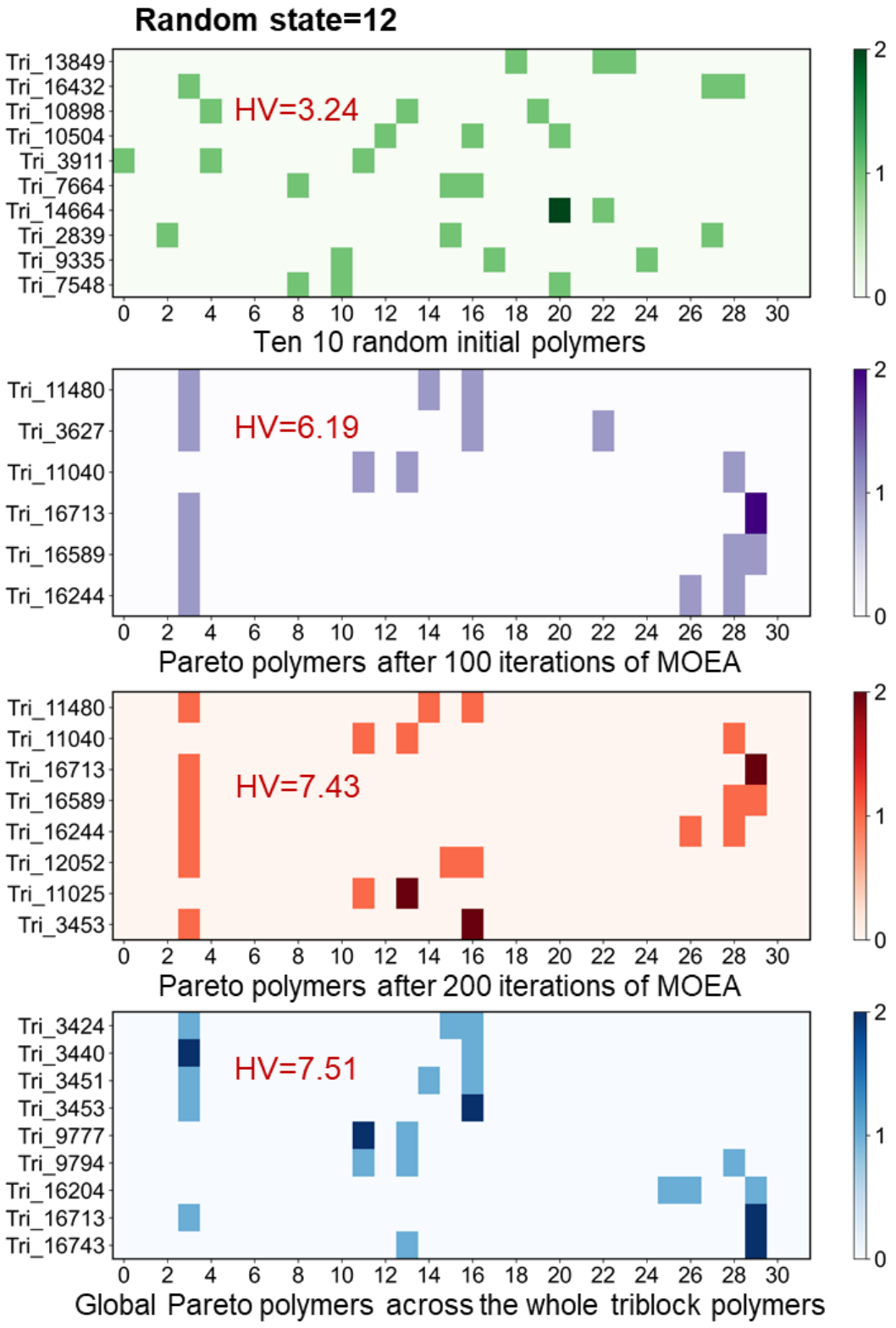

Fig. S8. Chemical blocks statistics for polymers at different stages, where 10 initial polymers were generated at the random seed of 12.

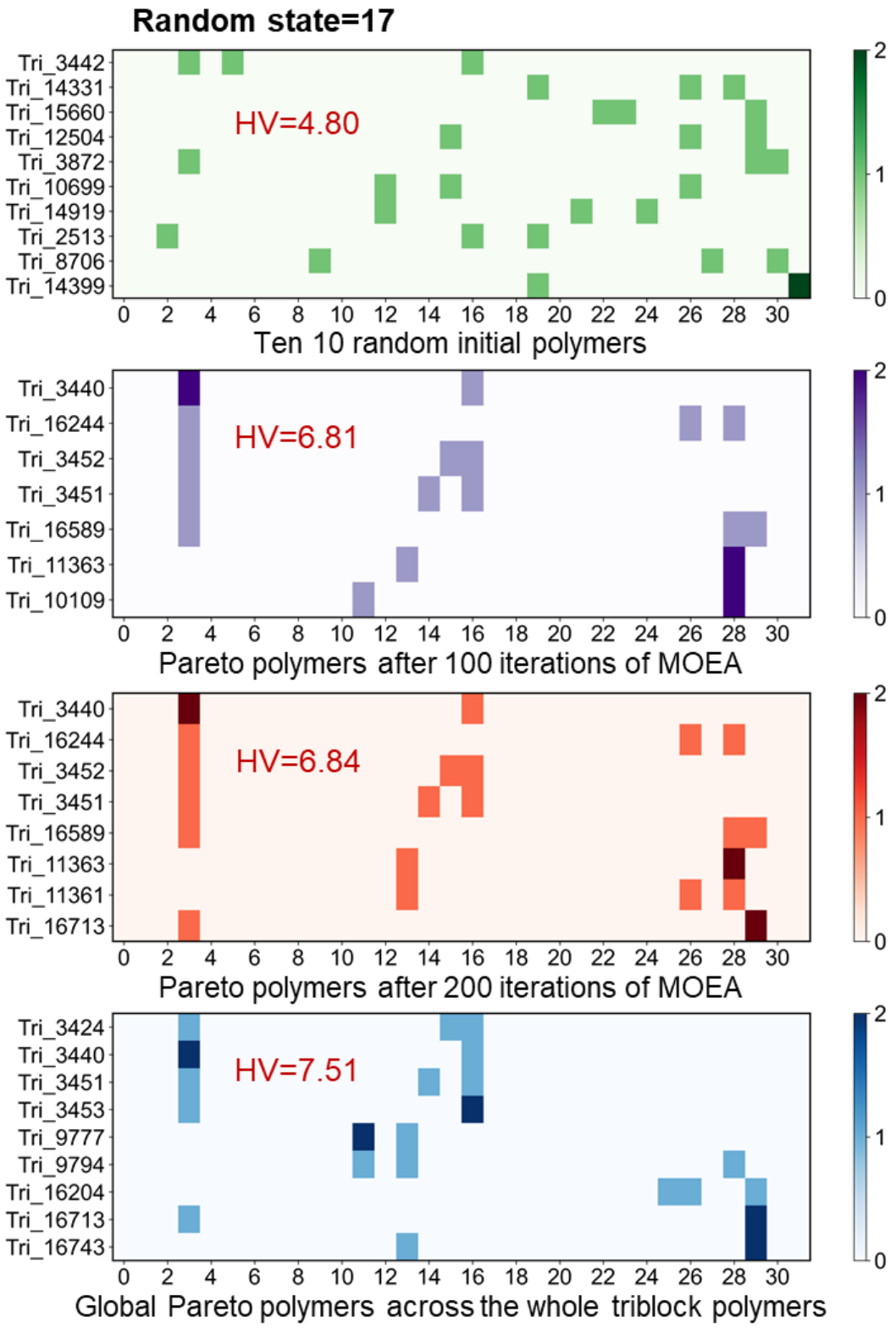

Fig. S9. Chemical blocks statistics for polymers at different stages, where 10 initial polymers were generated at the random seed of 17.

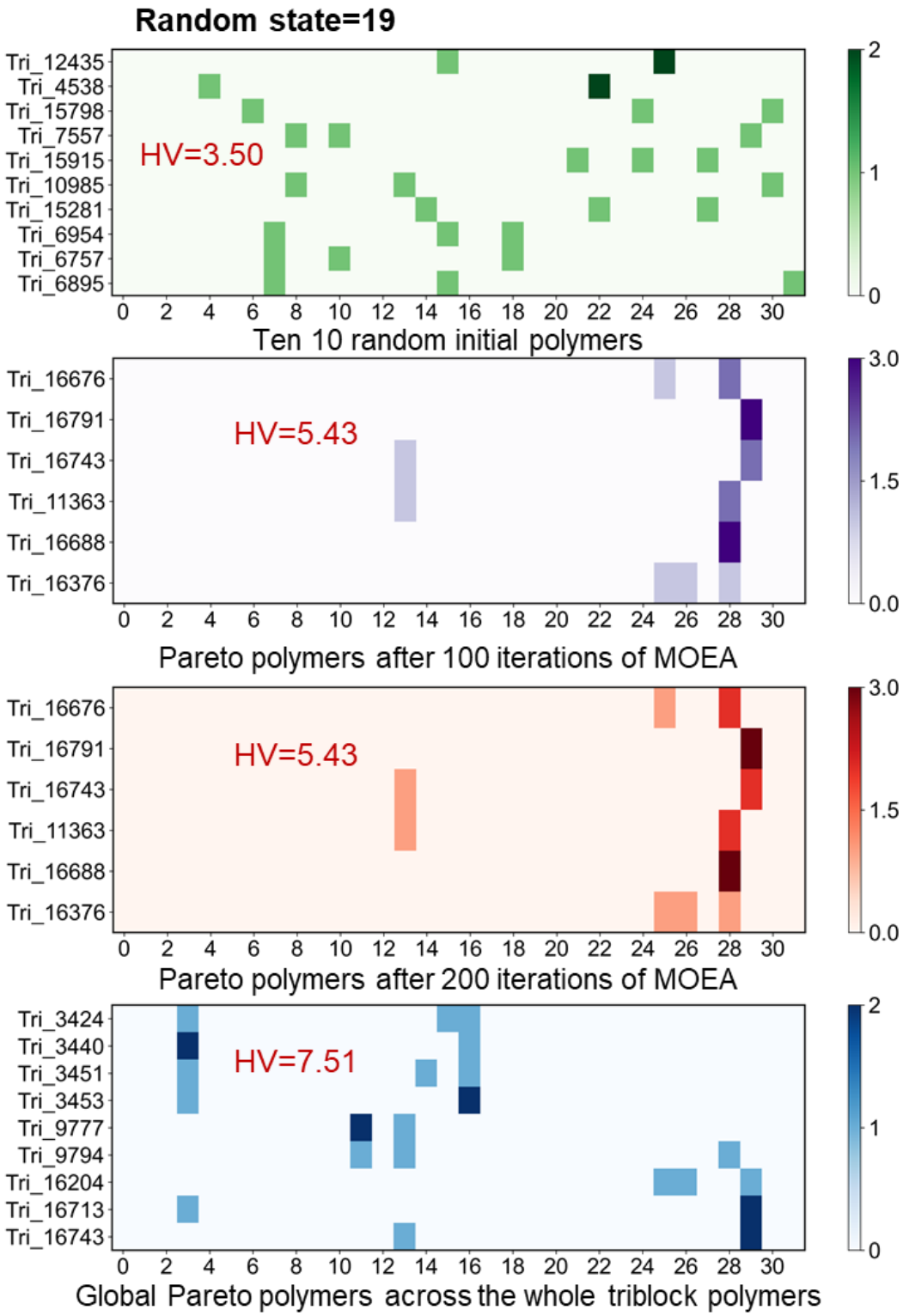

Fig. S10. Chemical blocks statistics for polymers at different stages, where 10 initial polymers were generated at the random seed of 19.

F. Parallel MOEAs versus MOBO for the design of high thermal conductivity pentablock polymers

Figure S11 demonstrates the distributions of generated pentablock polymers from 20 parallel MOEAs (blue dots) and a MOBO run (grey dots). Among 1921 MOEA-derived polymers, half of the candidates satisfy predefined requirements, i.e., SA≤3.0 and TC≥0.40 W $m^{-1}K^{-1}$. However, only 338 of 2005 polymers meet the above conditions in a MOBO run

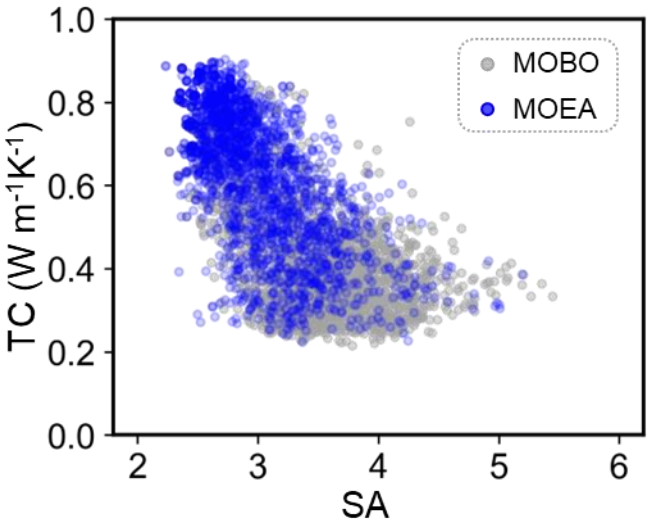

Fig. S11. Distributions of generated pentablock polymers from 20 parallel MOEAs (blue dots) and a MOBO run (grey dots).

G. Demonstration of novel polymers with MD calculated TC in this work

Table S2 lists the 50 promising polymers designed in this work, and their TCs were calculated by molecular dynamics simulation. The structures of repeating units can be viewed in the online tool of Marvin JS (https://marvinjs-demo.chemaxon.com/latest/) using SMILES as input.

Table S2. List of novel amorphous polymers (NAP) with MD calculated thermal conductivity, of which 20 triblock polymers and 30 pentablock polymers.

No.	SMILES	TC ($W\ m^{-1}K^{-1}$)	Rg (Å)	SA score	Source
NAP_1	[*]c1ccc(c2nc3cc4nc(c5cnc([*])nc5)[nH]c4cc3[nH]2)cc1	1.005	72.965	3.634	Triblock polymer
NAP_2	[*]c1ccc(c2nc3cc4nc(c5ccc([*])cc5)[nH]c4cc3[nH]2)cc1	0.777	76.485	3.471	
NAP_3	[*]c1ccc(c2nc3cc4nc(c5nc6cc7nc([*])[nH]c7cc6[nH]5)[nH]c4cc3[nH]2)cc1	0.753	74.992	3.741	
NAP_4	[*]c1ccc(c2nc3cc4nc(c5ccc6nc([*])oc6c5)[nH]c4cc3[nH]2)cc1	0.684	53.128	3.555	
NAP_5	[*]c1ccc(c2nc3cc4nc(c5ccc6nc([*])[nH]c6c5)[nH]c4cc3[nH]2)cc1	0.631	49.282	3.527	
NAP_6	[*]c1ccc(c2nc3cc(c4nc5cc6nc([*])[nH]c6cc5[nH]4)ccc3[nH]2)cc1	0.583	51.589	3.546	
NAP_7	[*]c1ccc(c2nc3cc4nc(c5nc6cc([*])ccc6o5)[nH]c4cc3[nH]2)cc1	0.571	54.191	3.699	
NAP_8	[*]c1ccc(c2nc3ccc(c4nc5cc6nc([*])[nH]c6cc5[nH]4)cc3[nH]2)cc1	0.570	50.309	3.527	
NAP_9	[*]c1ccc(c2ccc3nc(c4nc5cc6nc([*])[nH]c6cc5[nH]4)oc3c2)cc1	0.537	55.492	3.638	
NAP_10	O=C([*])Nc1ccc(NC(=O)Nc2ccc(NC(=O)Nc3ccc(NC(=O)[*])cc3)cc2)cc1	0.532	37.589	2.280	
NAP_11	[*]c1ccc(c2nc3cc4nc(c5nc6cc([*])ccc6[nH]5)[nH]c4cc3[nH]2)cc1	0.523	52.847	3.675	
NAP_12	O=C([*])Nc1ccc(NC(=O)c2ccc(C(=O)Nc3ccc(NC(=O)[*])cc3)cc2)cc1	0.494	50.556	2.314	
NAP_13	[*]c1ccc(c2ccc3oc(c4nc5cc6nc([*])[nH]c6cc5[nH]4)nc3c2)cc1	0.410	46.680	3.622	
NAP_14	O=C([*])c1ccc(C(=O)Nc2ccc(NC(=O)c3ccc(C(=O)N[*])cc3)cc2)cc1	0.408	29.009	2.443	
NAP_15	O=C([*])c1ccc(Nc2ccc(C(=O)Nc3ccc(C(=O)N[*])cc3)cc2)cc1	0.389	27.095	2.631	
NAP_16	O=C(Nc1ccc(NC(=O)Nc2ccc(C(=O)[*])cc2)cc1)c1ccc(C(=O)N[*])cc1	0.374	27.683	2.552	
NAP_17	O=C([*])c1ccc(NC(=O)c2ccc(Nc3ccc(C(=O)N[*])cc3)cc2)cc1	0.353	23.934	2.631	
NAP_18	O=C([*])c1ccc(NC(=O)c2ccc(C(=O)NC(=O)c3ccc(N[*])cc3)cc2)cc1	0.344	29.039	2.622	
NAP_19	O=C([*])c1ccc(C(=O)NC(=O)c2ccc(C(=O)Nc3ccc(N[*])cc3)cc2)cc1	0.321	27.350	2.814	
NAP_20	O=C([*])c1ccc(C(=O)Nc2ccc(NC(=O)c3ccc(C(=O)[*])cc3)cc2)cc1	0.315	26.576	2.336	

No.	SMILES	TC ($W\ m^{-1}K^{-1}$)	Rg (Å)	SA score	Source
NAP_21	O=C(NN[*])C(=O)c1ccc(c2ccc(C(=O)Nc3nc4ccc(c5nc6cc7nc([*])[nH]c7cc6[nH]5)cc4[nH]3)cc2)o1	0.904	32.468	3.647	Pentablock polymer
NAP_22	O=C([*])c1ccc(Nc2ccc(C(=O)NNC(=O)c3ccc(C(=O)OCC(=O)c4ccc(N[*])cc4)cc3)cc2)cc1	0.897	33.379	2.851	
NAP_23	O=C(Nc1ccc(NC(=O)OCC(=O)c2ccc(N[*])cc2)cc1)c1ccc(NC(=O)c2ccc(C(=O)[*])cc2)cc1	0.888	33.456	2.751	
NAP_24	O=C(O[*])Nc1ccc(NC(=O)c2ccc(C(=O)Nc3ccc(NC(=O)c4ccc(C(=O)[*])cc4)cc3)cc2)cc1	0.884	30.615	2.447	
NAP_25	O=C([*])c1ccc(C(=O)C=Cc2ccc(c3ccc4nc(c5ccc([*])cc5)[nH]c4c3)cc2)cc1	0.882	29.734	3.073	
NAP_26	O=C([*])Nc1ccc(NC(=O)C(=O)Nc2ccc(NC(=O)c3ccc(C(=O)NNC(=O)c4ccc(C(=O)Nc5ccc(C(=O)[*])cc5)cc4)cc3)cc2)cc1	0.880	28.939	2.685	
NAP_27	O=C(Nc1ccc(NC(=O)c2ccc(C(=O)N[*])cc2)cc1)c1ccc(C(=O)NNC(=O)c2ccc(C(=O)Nc3ccc(N[*])cc3)cc2)cc1	0.871	33.116	2.512	
NAP_28	O=C(Nc1ccc(NC(=O)c2ccc(C(=O)N[*])cc2)cc1)c1ccc(C(=O)NNC(=O)c2ccc(C(=O)Nc3ccc(C(=O)[*])cc3)cc2)cc1	0.862	40.238	2.538	
NAP_29	O=C(Nc1ccc(NC(=O)C(=O)Nc2ccc(NC(=O)c3ccc(C(=O)N[*])cc3)cc2)cc1)NC(=O)c1ccc(C(=O)Nc2ccc(C(=O)[*])cc2)cc1	0.840	27.541	2.783	
NAP_30	O=C(Nc1ccc(NC(=O)NC(=O)c2ccc(C(=O)Nc3ccc(NC(=O)c4ccc(C(=O)N[*])cc4)cc3)cc2)cc1)NC(=O)c1ccc(C(=O)[*])cc1	0.838	36.181	2.726	
NAP_31	O=C(c1ncc(Nc2ccc(Nc3cnc([*])nc3)cc2)cn1)c1ccc(NC=CC#C[*])cc1	0.801	32.844	3.666	
NAP_32	[*]c1ccc(CCCCCCc2ccc(CCCCCCCCCCCC[*])cc2)cc1	0.756	35.150	2.622	
NAP_33	[*]c1ccc(c2ccc3cc(c4ccc(c5ccc(c6ccc([*])cc6)cc5)cc4)ccc3c2)cc1	0.746	52.658	2.344	
NAP_34	O=C([*])c1ccc(NC(=O)c2ccc(C(=O)NNc3cc(NC(=O)c4ccc(NC(=O)c5ccc(C(=O)N[*])cc5)cc4)cc3)cc2)cc1	0.715	32.926	2.627	
NAP_35	O=C(Nc1ccc(NC(=O)c2ccc(C(=O)Nc3ccc(NC(=O)c4ccc(C(=O)N[*])cc4)cc3)cc2)cc1)NC(=O)c1ccc(C(=O)[*])cc1	0.694	39.057	2.619	
NAP_36	O=C([*])c1ccc(C(=O)Nc2ccc(NC(=O)c3ccc(C(=O)Nc4ccc(NC(=O)c5ccc(C(=O)[*])cc5)cc4)cc3)cc2)cc1	0.680	29.537	2.234	
NAP_37	O=C(NNC(=O)C(=O)OC(=O)c1ccc(N[*])cc1)c1ccc(C(=O)c2ccc(C(=O)[*])cc2)cc1	0.676	28.847	3.099	
NAP_38	O=C([*])Oc1ccc(NC(=O)c2ccc(C(=O)Nc3ccc(NC(=O)c4ccc(C(=O)[*])cc4)cc3)cc2)cc1	0.669	46.822	2.462	

No.	SMILES	TC ($W\ m^{-1}K^{-1}$)	Rg (Å)	SA score	Source
NAP_39	O=C(C=CC(=O)c1ccc(NC(=O)c2ccc(C(=O)c3nc4cc5nc(c6ccc7nc([*])oc7c6)[nH]c5cc4[nH]3)cc2)cc1)N[*]	0.666	32.180	3.581	
NAP_40	O=C(Nc1ccc(NC(=O)OC[*])cc1)c1ccc(NC(=O)c2ccc(NC(=O)c3ccc(C(=O)[*])cc3)cc2)cc1	0.653	30.014	2.542	
NAP_41	O=C(Nc1nc2cc3nc(c4ccc(c5ccc([*])cc5)cc4)[nH]c3cc2[nH]1)c1ccc(C(=O)c2nc3ccc([*])cc3[nH]2)cc1	0.614	48.977	3.534	
NAP_42	O=C(Nc1nc2cc3nc(c4ccc(c5ccc([*])cc5)cc4)[nH]c3cc2[nH]1)c1ccc(C(=O)Nc2ccc(N[*])cc2)cc1	0.597	42.386	3.229	
NAP_43	O=C(Nc1ccc(NC(=O)c2ccc(C(=O)N[*])cc2)cc1)C(=O)Nc1ccc(NC(=O)C(=O)Nc2ccc(NC(=O)c3ccc([*])cc3)cc2)cc1	0.576	48.112	2.685	
NAP_44	O=C(Nc1ccc(NC(=O)C(=O)Nc2ccc(NC(=O)c3ccc(C(=O)N[*])cc3)cc2)cc1)Nc1ccc(NC(=O)c2ccc(C(=O)[*])cc2)cc1	0.537	29.749	2.656	
NAP_45	O=C(Nc1ccc(NC(=O)c2ccc(C(=O)Nc3ccc(C(=O)N[*])cc3)cc2)cc1)NC(=O)c1ccc(Nc2ccc(C(=O)[*])cc2)cc1	0.520	24.420	2.702	
NAP_46	O=C([*])NNC(=O)c1ccc(C(=O)NNC(=O)c2ccc(C(=O)Nc3ccc(NC(=O)c4ccc(C(=O)N[*])cc4)cc3)cc2)cc1	0.465	24.866	2.571	
NAP_47	O=C([*])Nc1ccc(NC(=O)c2ccc(C(=O)Nc3ccc(NC(=O)c4ccc5cc(C(=O)Nc6ccc(NC(=O)[*])cc6)ccc5c4)cc3)cc2)cc1	0.460	24.916	2.543	
NAP_48	O=C(Nc1ccc(NC(=O)c2ccc(C(=O)O[*])cc2)cc1)c1ccc(NC(=O)c2ccc(C(=O)[*])cc2)cc1	0.395	28.573	2.425	
NAP_49	O=C(Nc1ccc(NC(=O)c2ccc(C(=O)Nc3ccc(NC(=O)c4ccc(C(=O)N[*])cc4)cc3)cc2)cc1)c1ccc([*])cc1	0.393	74.187	2.416	
NAP_50	O=C(Nc1ccc(NC(=O)c2ccc(C(=O)Nc3ccc(C(=O)N[*])cc3)cc2)cc1)c1ccc(NC(=O)c2ccc(C(=O)[*])cc2)cc1	0.309	21.521	2.367	

H. Expansion case of parallel MOEAs for designing triblock polymers with TC > 0.40 $Wm^{-1}K^{-1}$ and RI > 1.80

Our proposed scheme is flexible and universal, and can be extended to the design of polymers with other target properties. To demonstrate it, we provide an expansion case of parallel MOEAs for designing triblock polymers with TC > 0.40 $Wm^{-1}K^{-1}$ and RI (refractive index) > 1.80. Polymers with high RI are favorable for flexible displays [5], organic light-emitting diodes [6] and image sensors [7]. Moreover, the computational database records the RI of polymers, which was calculated by Psi4 using Lorentz–Lorenz equation in RadonPy [8,9]. There are 1138 candidates with known RIs out of 1144 polymers in the benchmark dataset. Based on these 1138 polymers, we trained a random forest (RF) model in Scikit-learn using five-fold cross-validation and Bayesian optimization for the determination of hyperparameters [2]. Figure S12 displays the pairs of RI predicted by the RF model versus that calculated by the MD simulation, which suggests that the trained surrogate model has good predictive ability, with R^2 of 0.93±0.02 and RMSE of 0024±0.03.

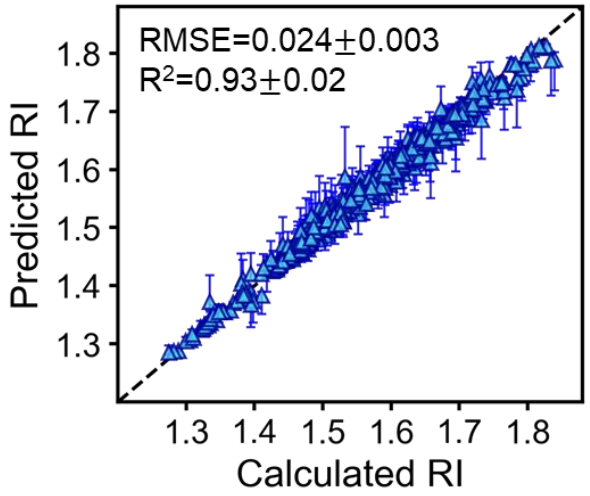

Fig. S12. Random forest model for refractive index prediction

The entire 16896 triblock polymers were configured to the exploration space, and their RIs versus TCs are shown in Figure S13a. There are eight polymers at the global Pareto front, and 7/8 are polyimines, as illustrated in Figure S13d. Our optimization target was set to TC > 0.40 W $m^{-1}K^{-1}$ and RI > 1.80. However, achieving a high refractive index (> 1.80) is not easy, with a percentage of only 0.38% among all triblock polymers (Figure S13b). The Venn diagram of Figure S13c counts the number of target polymers in the whole exploration space as 56, with a ratio of only 0.33%.

We performed 10 MOEAs in parallel for designing triblock polymers with target properties. Each MOEA starts from 10 randomized initial structures, and goes through 10 candidates × 200 generations. Figure S14a exhibits the HV curves of 10 MOEA runs, where the reference point was set to [0.15,1.45] for TC and RI, and the maximum HV value is 3.83 (Calculated from 8 global Pareto polymers and the reference point). The final HVs ranged from 2.80 to 3.50, since different collections of initial candidates. We then employed the Gaussian kernel to estimate the probability density function (PDF) of all searched polymers in 10 MOEAs, as shown in Figure S14b. The high probability region occurs close to the global Pareto front, which reflects the robustness of the parallel MOEA. Figure S14c illustrates 91 unrepeated triblock polymers (blue dots) designed by parallel MOEAs, of which 16 candidates (16.7%, in Figure S14d) are the target polymers. This demonstrates the scalability of our developed inverse design workflow.

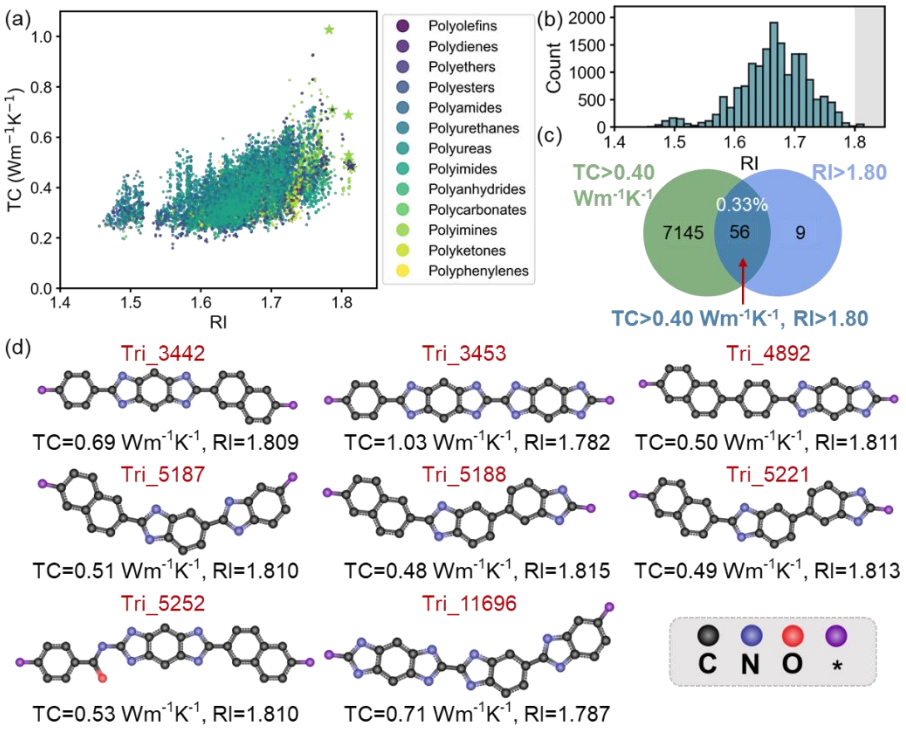

Fig. S13. All triblock polymers with ML predicted TCs and RIs. (a) Relationship between RIs and TCs for 16896 triblock polymers. (b) Distribution of RIs. (c) Venn diagram statistics for polymers with TC>0.40 $Wm^{-1}K^{-1}$ or RI>1.80. (d) Repeating units of eight global Pareto polymers.

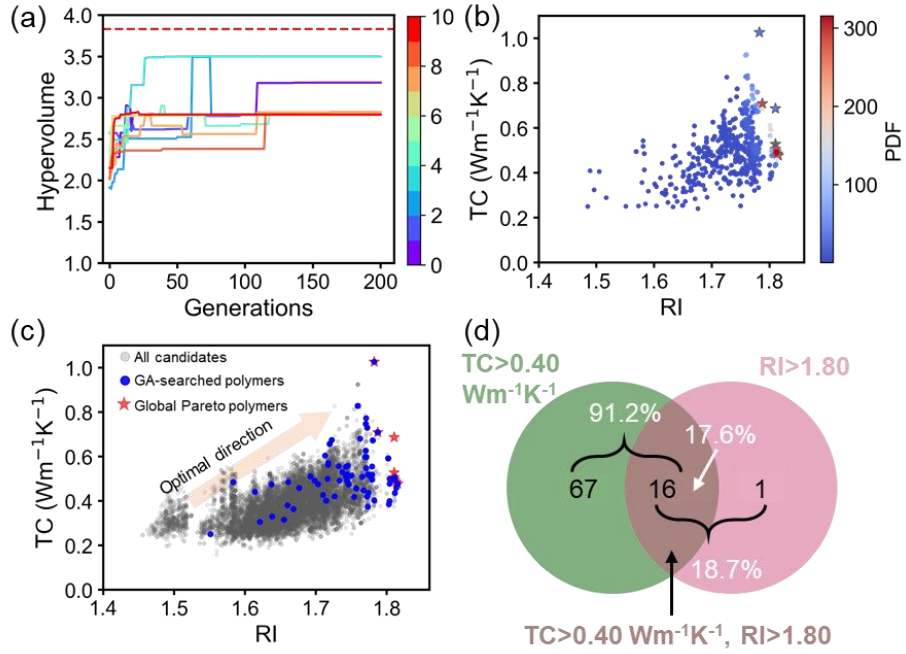

Fig. S14. Parallel MOEAs for the design of innovative triblock polymers. (a) HV curves for 10 MOEAs with different initial structures. (b) Probability density maps in objective space for 10 runs of MOEA. (c) Distributions of 91 non-repeating triblock polymers generated by 10 parallel MOEAs. (d) Venn diagram statistics for MOEA-produced polymers with TC > 0.40 $Wm^{-1}K^{-1}$ or RI > 1.80.

SI References

1. Morgan, H. L., The Generation of a Unique Machine Description for Chemical Structures-A Technique Developed at Chemical Abstracts Service. *J. Chem. Doc.* (1965) 5 (2), 107

2. Pedregosa, F.*, et al.*, Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* (2011) 12, 2825

3. Gardner, J.*, et al.*, Gpytorch: Blackbox matrix-matrix gaussian process inference with gpu acceleration. *Adv. Neural Inf. Process. Syst.* (2018) 31

4. Lundberg, S. M., and Lee, S.-I., A unified approach to interpreting model predictions. *Adv. Neural Inf. Process. Syst.* (2017) 30, 4765−4774

5. Chang, J.-H., Equibiaxially stretchable colorless and transparent polyimides for flexible display substrates. *REVIEWS ON ADVANCED MATERIALS SCIENCE* (2020) 59 (1), 1

6. Kim, J. G.*, et al.*, Modeling of flexible light extraction structure: Improved flexibility and optical efficiency for organic light-emitting diodes. *Organic Electronics* (2020) 85, 105760

7. Fang, L.*, et al.*, Phosphorus- and Sulfur-Containing High-Refractive-Index Polymers with High Tg and Transparency Derived from a Bio-Based Aldehyde. *Macromol.* (2020) 53 (1), 125

8. Hayashi, Y.*, et al.*, RadonPy: automated physical property calculation using all-atom classical molecular dynamics simulations for polymer informatics. *npj Comput. Mater.* (2022) 8 (1), 222

9. Turney, J. M.*, et al.*, Psi4: an open-source ab initio electronic structure program. *WIREs Comput Mol Sci.* (2012) 2 (4), 556