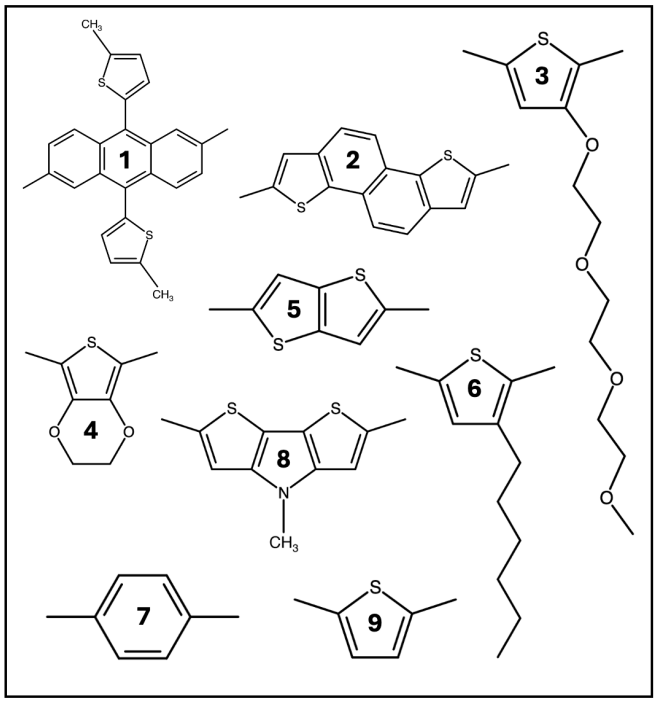
Supplemental Information: A Bayesian approach providing design choices and chemical insight for solution-processed thermoelectric polymers

Connor Ganley[1], Yuqi Feng[1], Himadri Shekhar Karmakar[2], Howard E. Katz[2], and Paulette Clancy[1,*]
[1]*Department of Chemical and Biomolecular Engineering*
Johns Hopkins University
Baltimore, Maryland, USA and
[2]*Department of Materials Science and Engineering*
Johns Hopkins University
Baltimore, Maryland, USA

[1] T. Mukhopadhyaya, J. Wagner, T. D. Lee, C. Ganley, S. Tanwar, P. Raj, L. Li, Y. Song, S. J. Melvin, Y. Ji, P. Clancy, I. Barman, S. Thon, R. S. Klausen, and H. E. Katz, enSolution-Doped Donor–Acceptor Copolymers Based on Diketopyrrolopyrrole and 3, 3-Bis (2-(2-(2-Methoxyethoxy) Ethoxy) ethoxy)-2, 2-Bithiophene Exhibiting Outstanding Thermoelectric Power Factors with p-Dopants, Advanced Functional Materials **34**, 2309646 (2024).



#	IUPAC Name	Label
1	2-methyl-5-[10-(5-methylthiophen-2-yl)anthracen-9-yl]thiophene	B3TA
2	[1]benzothiolo[7,6-g][1]benzothiole	NDT
3	3-[2-[2-(2-methoxyethoxy)ethoxy]ethoxy]thiophene	MEET
4	2,3-dihydrothieno[3,4-b][1,4]dioxine	EDOT
5	thienothiophene	TT
6	3-hexylthiophene	T6C
7	benzene	BZ
8	7-methyl-3,11-dithia-7-azatricyclo[6.3.0.02,6]undeca-1(8),2(6),4,9-tetraene	SNS
9	thiophene	T

Fig. S1. Functional group structures, labels, and IUPAC names of the conjugated species studied in this work.

* Corresponding author email: pclancy3@jh.edu

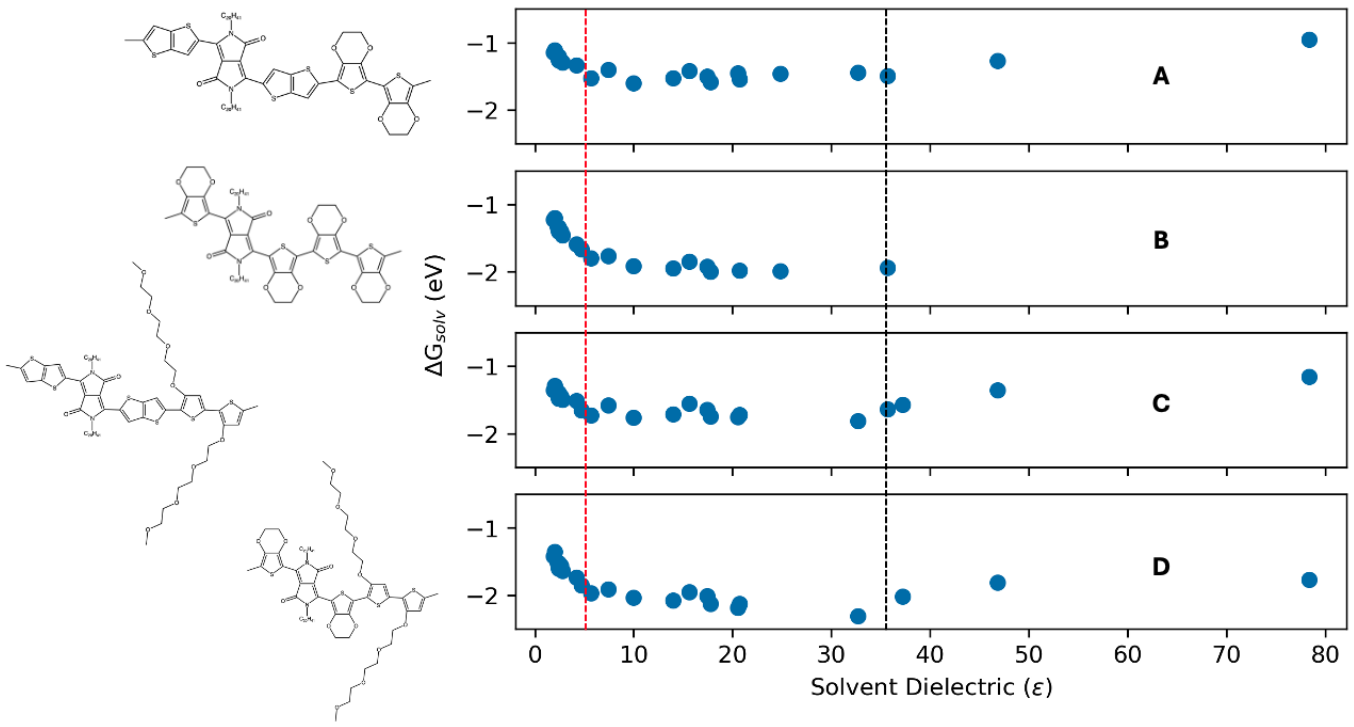


Fig. S2. Free energy of solvation ΔG_{solv} versus solvent dielectric for the four doped DPP-based polymers synthesized in work by Mukhopadhyaya *et al.* [1] against which our PAL-predictions of polymer design were validated. Graphs A,B,C,D refer to the polymer repeat units shown to the left of the graphs. The dotted red line at $\epsilon = 5.7$ denotes that of chlorobenzene, the solvent used for the polymer. The dotted black line at $\epsilon = 35.7$ denotes that of acetonitrile, the solvent used for the dopant. The solution dielectric is a linear combination of the component solvents and varies with dopant concentration.

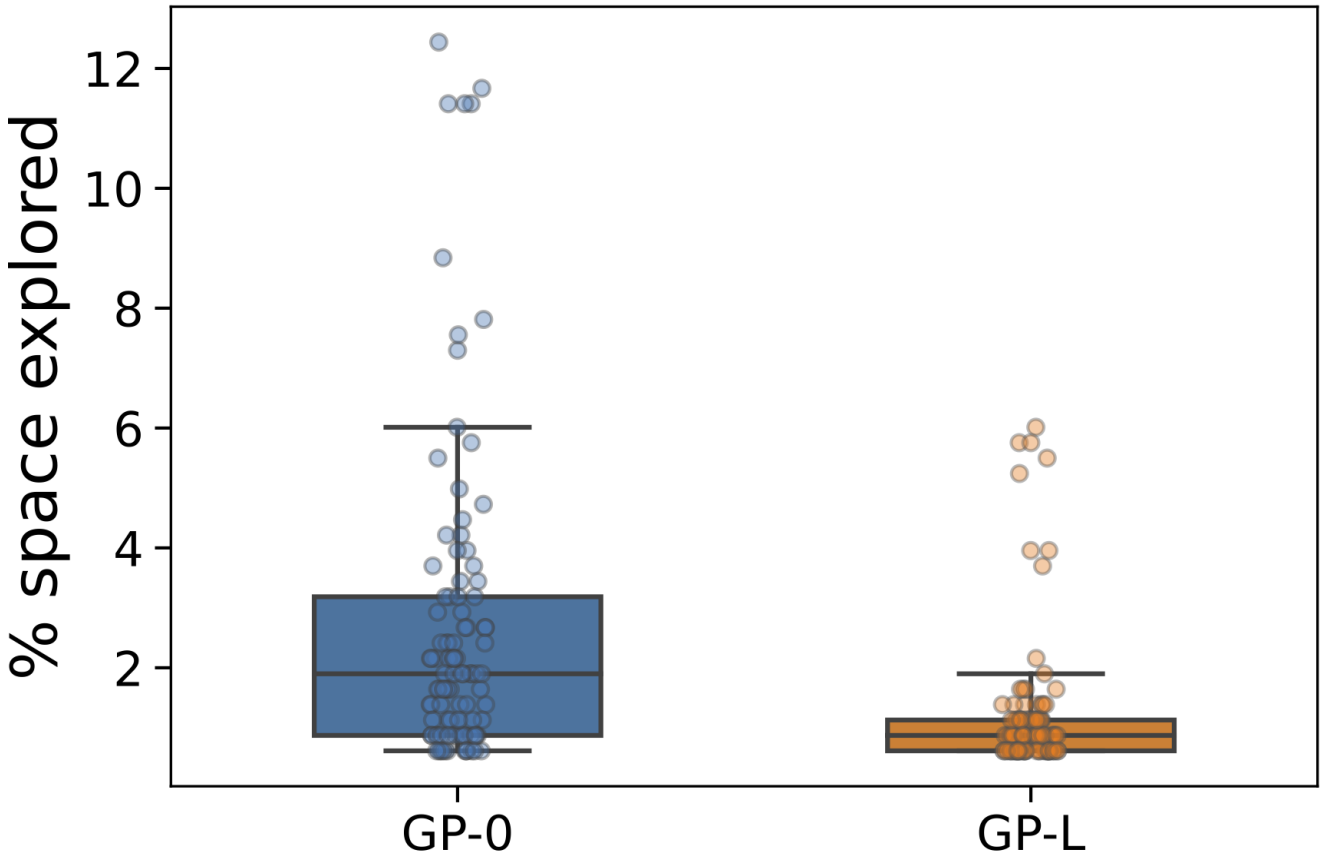


Fig. S3. Comparison between the performance of two surrogate models used in our work. One (shown in blue) represents the use of a zero-mean prior GP (GP-0). The other (shown in orange) employs a linear-mean prior GP (GP-L). The results show the number of objective function evaluations required before arriving at an optimum design. The GP-L model (orange) is able to find the optimal objective function more quickly (in fewer evaluations) than the GP-0 model (blue) with smaller uncertainty. Each circle represents the number of GP objective function evaluations that the PAL 2.0 Bayesian Optimization (BO) algorithm required to find the optimum ΔG_{solv}, and the quartiles of all BO runs are shown by the boxes and whiskers.

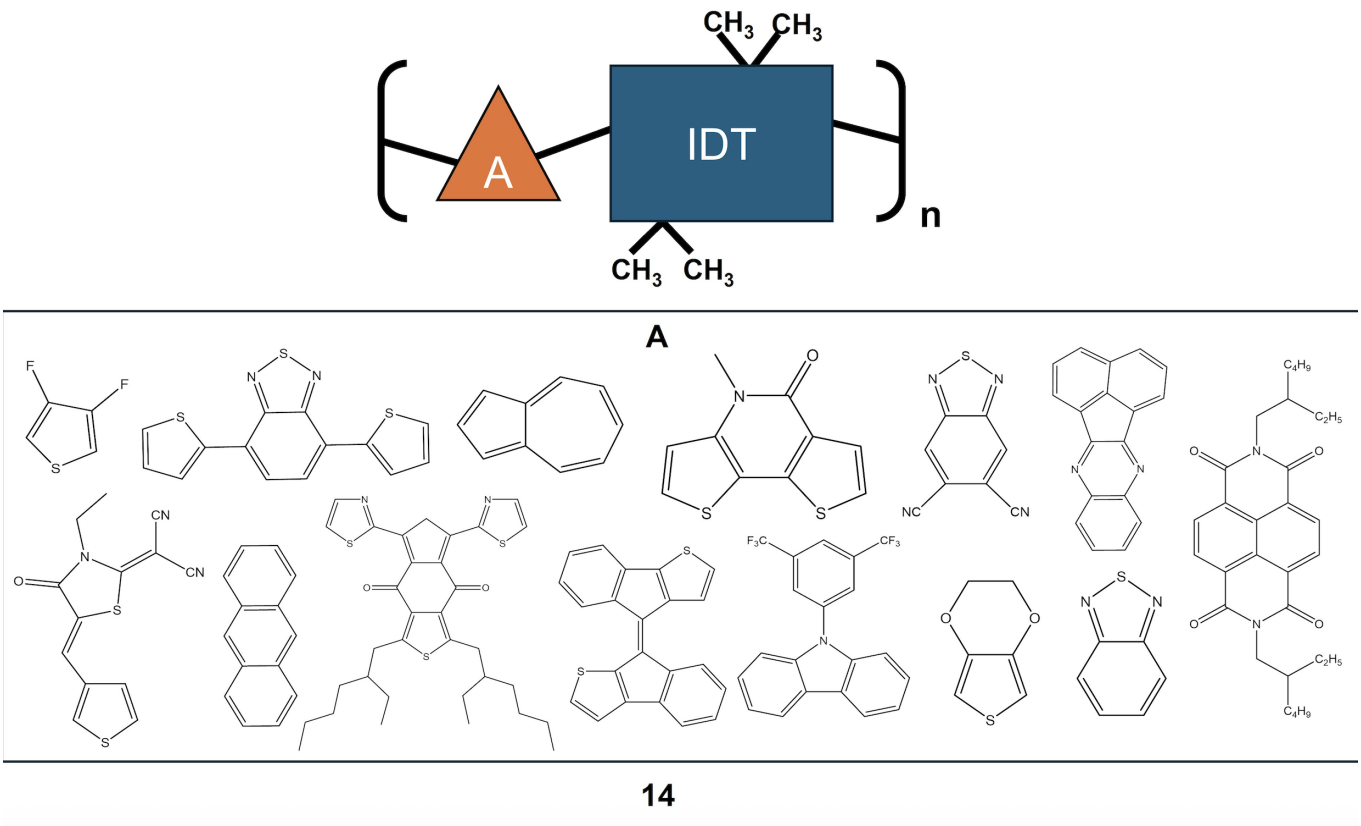


Fig. S4. Curated design space of the hypothetical IDT-based polymers studied in this work, with 14 possible choices for A.

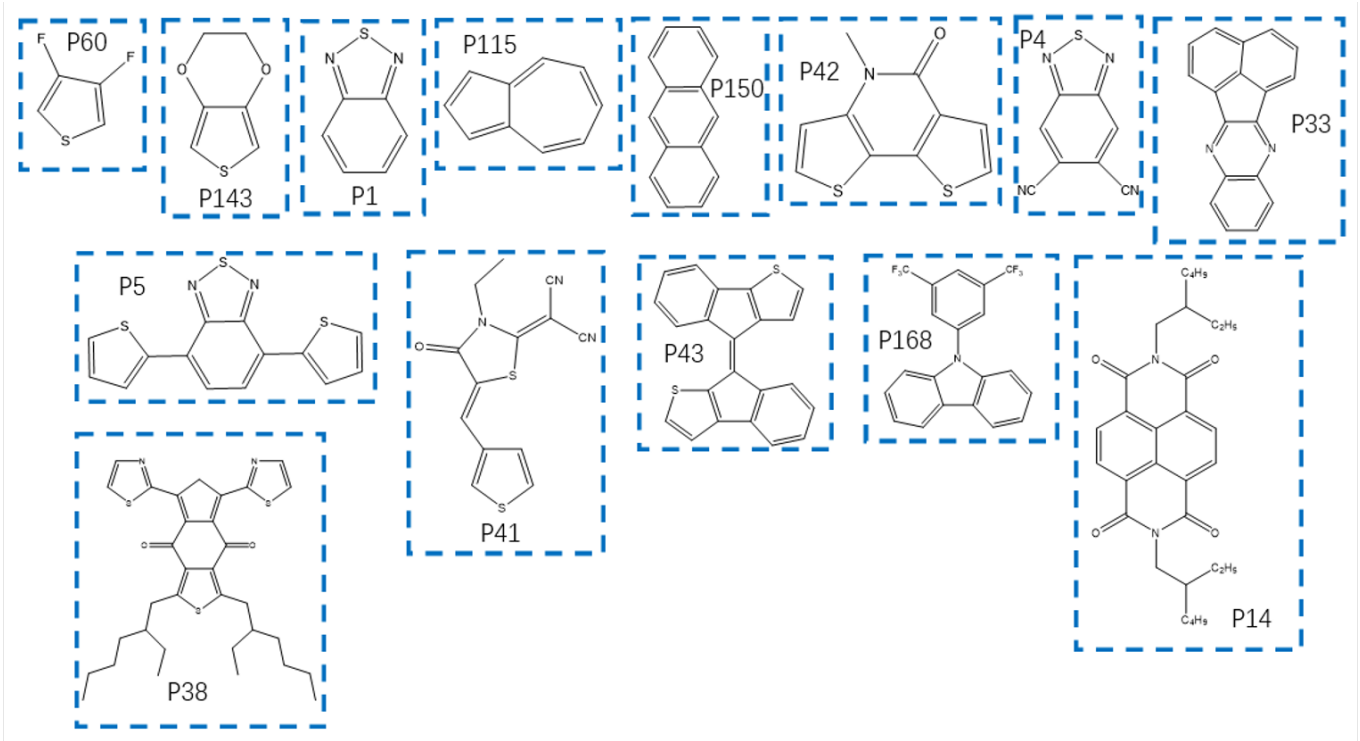


Fig. S5. Labeled referential nomenclature of structures polymerized with IDT in an A-IDT motif.

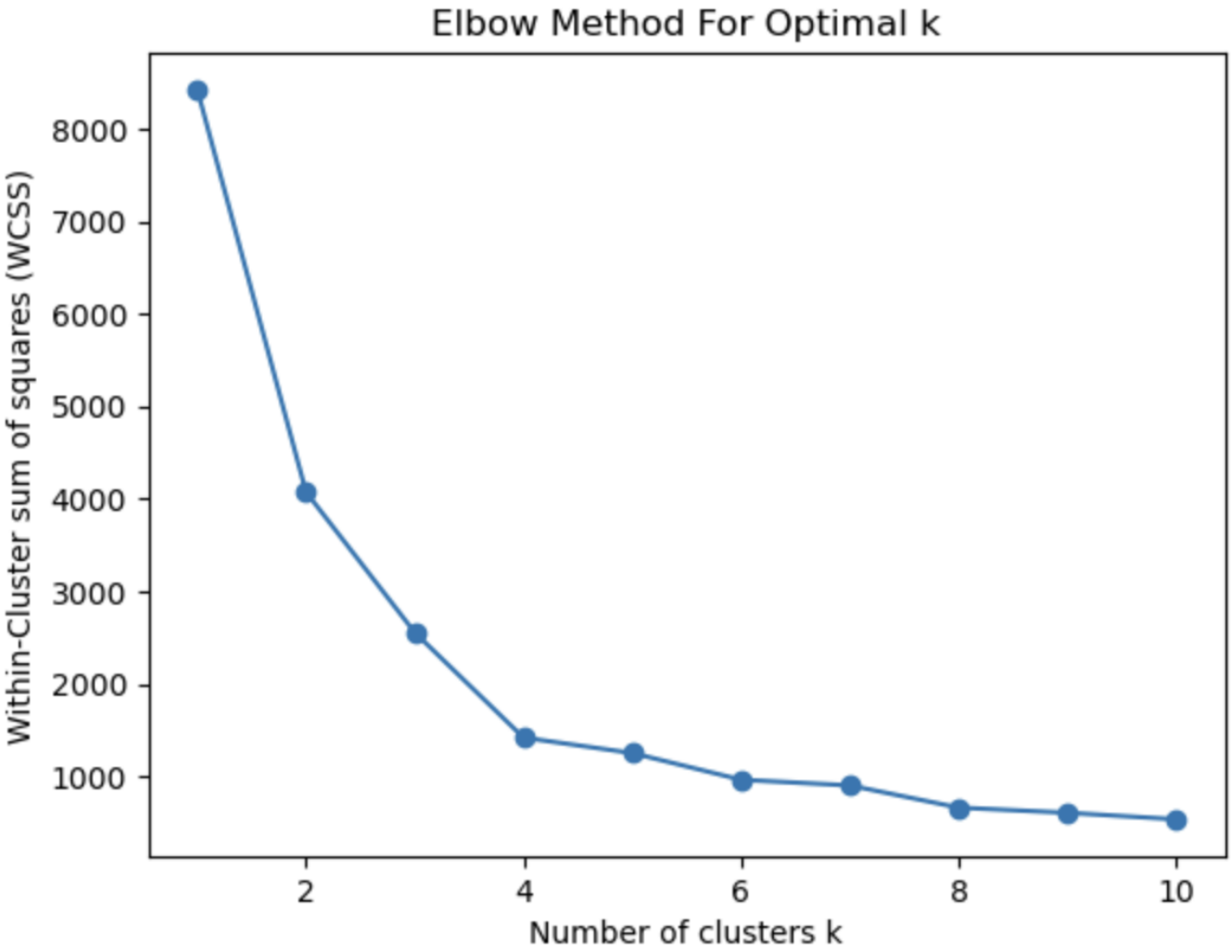


Fig. S6. Elbow plot shows the within-cluster sum of squares (WCSS) decays asymptotically and that 4 clusters approaches this asymptote. The DFT solvation data of DPP-based polymers, grouped by solvent, was used for this analysis.

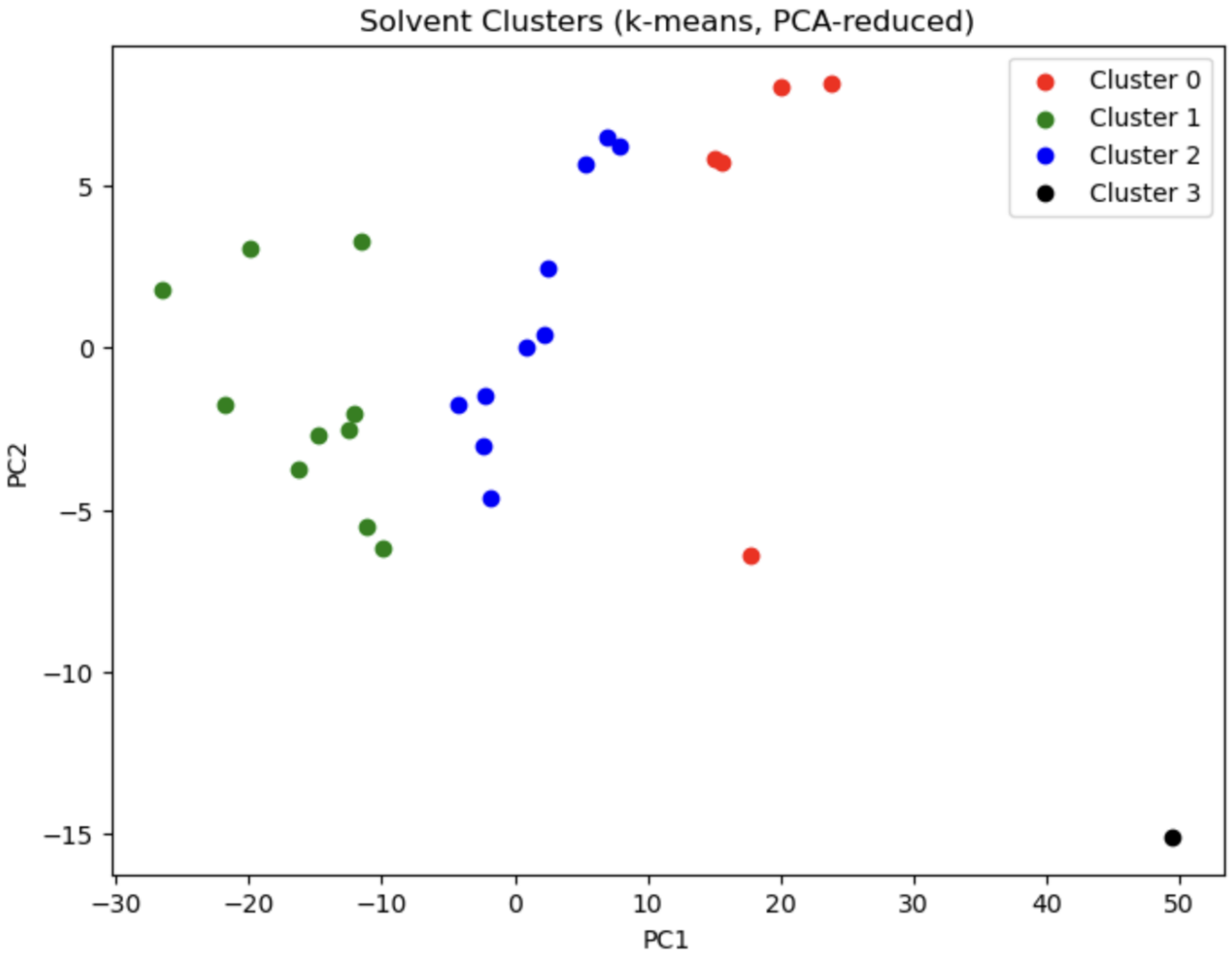


Fig. S7. When grouped by solvent, the ΔG_{solv} data is represented in 324-dimensional space, with one dimension per DPP-based polymer. Principal Component Analysis (PCA) decomposes these 324-dimensional vectors into 2-dimensional vectors characterized by maximal variation. Each data point represents a solvent in this 2-dimensional principal component space. The colors of the dots are determined by the k-means clustering algorithm. The clusters appear to occupy discrete space along PC1.

Functional Group Subunit	Free Volume (Å^3)
T	31.1
BZ	35.5
TT	63.3
NDT	80.8
EDOT	91.6
SNS	121.2
DPP	130.5
T6C	144.3
MEET	488.9
B3TA	1739.9

Fig. S8. Calculated free volume quantities of the functional group subunits studied in this work. Nomenclature is the same as in Figure S1. The free volume was calculated by subtracting the isosurface volume from the DFT calculations described in the Methods section from the volume of the smallest rectangular prism that enclosed the functional group subunit.

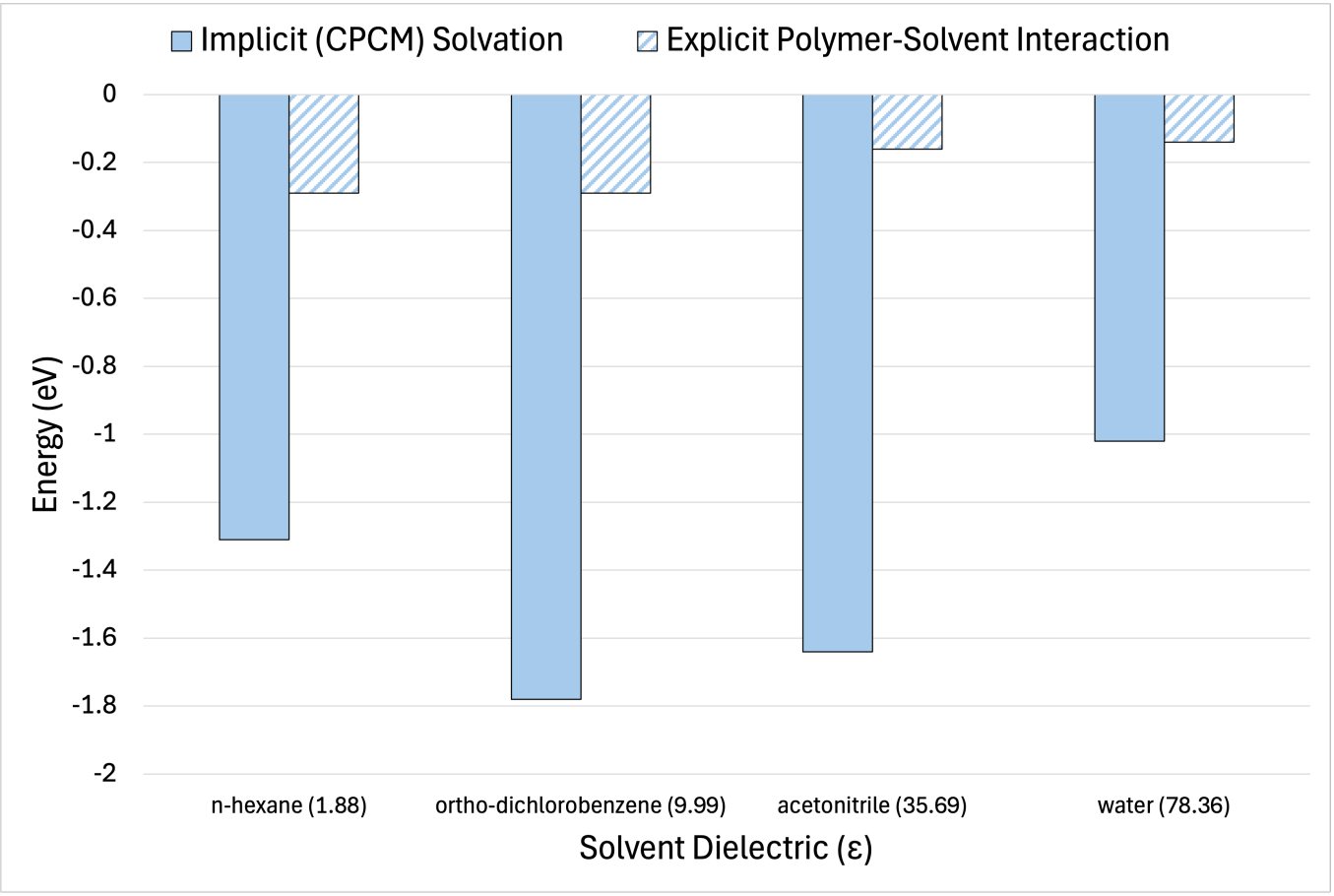


Fig. S9. Combination plot of the mean energies for implicitly modeled solvents, in that case the *free energy of solvation* (solid bars), to the mean energies of the explicit polymer-solvent interaction, in this case shown as *enthalpies* (striped). Results are shown for the 324 polymer candidates studied in this work that were immersed in four choice of solvents: *n*-hexane, *ortho*-dichlorobenzene, acetonitrile, and water. To reiterate the point, it is important to note that the energies for implicit and explicit relate to different properties: free energy versus enthalpy. Hence a direct comparison is not intended here. This image uses the same data as in Figures 2 and 5. The implicit (solid) data suggests greater variability in the interaction enthalpy, likely due to the dielectric mean field approach that underlies implicit solvation calculations. It is possible that explicit solvation calculations with more than one solvent molecule would exhibit closer variability. Nevertheless, the important point is that both data sets indicate a regime of greater solvation at lower solvent dielectric values, *n*-hexane and *ortho*-dichlorobenzene), and a regime of lesser solvation (at higher solvent dielectric values: acetonitrile and water).