

Supplementary Information

Autonomous Synthesis and Inverse Design of Electrochromic Polymers with High Efficiency and Accuracy

Authors: Yukun Wu^{1,2†}, Aikaterini Vriza^{1†}, Doga Ozgulbas³, Rafael Vescovi³, Jianing Zhou^{1,2}, Zhiyang Wang², Shiyu Hu⁴, Yuepeng Zhang⁴, Qiaomu Yang¹, Anna M. Österholm⁵, John R. Reynolds⁵, Subramanian K.R.S. Sankaranarayanan¹, Maria K.Y. Chan¹, Ian T. Foster³, Jianguo Mei^{2*}, Henry Chan^{1*}, Jie Xu^{1,6*}

Affiliations:

¹Nanoscience and Technology Division, Argonne National Laboratory, Lemont, IL 60439, USA.

²Department of Chemistry, Purdue University, West Lafayette, IN 47907, USA.

³Data Science and Learning Division, Argonne National Laboratory, Lemont, IL 60439, USA

⁴Applied Materials Division, Argonne National Laboratory, Lemont, IL 60439, USA.

⁵School of Chemistry & Biochemistry, School of Materials Science & Engineering, Georgia Institute of Technology, Atlanta, GA 30332, USA

⁶Pritzker School of Molecular Engineering, The University of Chicago, Chicago, Illinois 60637, USA

[†]These authors contributed equally to this work.

*Corresponding author. Email: xuj@anl.gov; jgmei@purdue.edu; hchan@anl.gov

Table of content

1.	Materials and instruments	7
2.	Electrochromic polymers (ECPs) in literature	8
2.1	ECP database construction	8
2.2	Demonstration of large language model (LLM)-assisted data mining workflow	16
2.3	Calculating the color values of polymers in ECP database	18
2.4	ECP database summary	19
2.5	Extracting synthetic conditions from literature	20
3.	Reaction condition optimization	22
3.1	Rational in refining the design space of synthesis condition optimization	22
3.2	Stock solution preparation	22
3.3	Automated Reaction workflow	22
4.	Copolymer Set-transformer model design	27
4.1	System overview	27
4.2	Architecture	28
	Encoder.	28
	Decoders.	28
4.3	ECPs representations	29
5.	High-throughput synthesis of ECPs for model refinement	31
5.1	Stock solution preparation	31
5.2	Reaction workflow	31
6.	Electrochromic characterization	37
6.1	CV and spectro-electrochemical measurements	37
7.	Model Tuning and Optimization	39
8.	Autonomous production of the ECPs with targeted colors	40
8.1	Graphical user interface for the initial selections	40
8.2	Lab network and WEI2 settings	41
8.3	Step-by-step explanation of the workflow	42
	Create the initial screening dataset with all the available experiments.	42
	Stock solution preparation.	42
	Robotic workflow.	43
	Measure the L*a*b* values similarity.	43
	Selecting the next experiments to run.	43
	Selecting the next experiments to run.	44
	Results validation and resubmit experiments.	44
	Applying the stopping criteria.	44
8.4	Results of the autonomous discovery	45
	Orange color results.	45

	Green color results.....	46
9.	References	48

List of figures and tables

Table S1. Full ECP database from literature with monomer constructions and $L^*a^*b^*$ values of each polymer. ECP database was constructed based on the selected ECP literature. (Samples with * do not have absorption spectra, reported $L^*a^*b^*$ values were used. Abbreviations used in the table: “EDOT-BTD-EDOT”: 4,7-bis(2,3-dihydrothieno[3,4-b][1,4]dioxin-5-yl)benzo[c][1,2,5]thiadiazole; “DMP”: 3,3-dimethyl-3,4-dihydro-2H-thieno[3,4-b][1,4]dioxepine).	8
Figure S1. Two additional examples of data extracted using the LLM-assisted data mining framework. (a) OMIECs database compiled from key data sources in relevant literature. The extracted information is summarized in text, tables, and images, with a data table presenting the curated results after screening 52 papers related to OMIECs. b) Example of microscopy images extracted and tagged using the data mining pipeline. An example on how to scrape information from PDF or HTML files can be also found in the following link: https://katerinavr.github.io/exsclaim2.0/PDFScraper.html	16
Figure S2. Correlation between the calculated and literature reported $L^*a^*b^*$ values of some representative samples. Samples were chosen from papers providing both absorption spectra and tabulated $L^*a^*b^*$ values (Polymer index 54-58, 83-87 from ECP database were chosen as the samples here).	18
Figure S3. Scatterplot with the a^* , and b^* values of all the datapoints collected from literature (Table S1). The colors of datapoints are represented by the color of the dots.	19
Figure S4. Distribution of the ECP database generated from literature: (a) the distribution in monomer ratios; (b) the distribution of the number of different monomers in ECP.	20
Scheme S1. Model reaction for synthesis condition optimization.	22
Table S2. Reaction optimization parameters and yield.	24
Figure S5. NMR spectra of the products from reaction optimization experiments, with reaction No. noted on the right side of the spectrum. The yields of the reactions (indicated in Table S2) were calculated based on the integration of the peak from unreacted thiophene aromatic proton at $\delta = 6.2$ ppm (as x) and the integration of the peak from all methoxy proton at $\delta = 3.8-4.0$ ppm (as y).	26
Figure S6. Schematic of the model architecture consisting of a sequential encoder (pink part) and a dual decoder (blue part). After decoder two predictions were performed, one is directly to $L^*a^*b^*$ values; the other is predicted to the absorption spectra first, then converted to $L^*a^*b^*$ values.	27
Figure S7. Comparison of prediction accuracy using different monomer representation techniques. The descriptor made by “Fingerprint with DFT Descriptor” gave the best overall performance and was chosen as final descriptor.	29
Table S3. Detailed list of the DFT descriptors which were calculated using the Auto-QChem package ²⁷	29
Table S4. Batch 1 reaction parameters and the resulting $L^*a^*b^*$ values for high-throughput synthesis of ECP for training dataset.	32
Table S5. Batch 2 reaction parameters and the resulting $L^*a^*b^*$ values for high-throughput synthesis of ECP for training dataset.	33
Table S6. Batch 3 reaction parameters and the resulting $L^*a^*b^*$ values for high-throughput synthesis of ECP for training dataset.	34

Figure S8. Normalized UV-vis Spectra of the high-throughput synthesized ECPs for model refinement: (a)-(c) represent results in batch 1-3 respectively; all spectra were normalized based on absorption intensity from 0 to 1, with the polymer colors represented by the color of the spectral lines.	35
Figure S9. Performance comparison of correlation coefficient (bottom) after updating the training set with the three experimental batches (top). We used correlation coefficients to assess the degree of agreement between predicted and experimental values in terms of relative trends. While absolute deviations are already captured by the MSE-based loss function, correlation provides insight into whether the model is able to reproduce the correct ordering and monotonic relationships across the dataset. This complementary metric therefore emphasizes trend fidelity rather than absolute magnitude.....	35
Table S7. Randomly selected ECPs in different color themes. The following polymers were randomly selected based on the representative colors and synthesized according to the methods described in Section 5. After precipitation, the polymer products were purified by additional Soxhlet extraction of acetone and hexane and collected in chloroform.	37
Figure S10. CV Spectra of the selected of ECPs solutions (Table S7): (a)-(f) represent results of polymer 1-6 respectively.	38
Figure S11. Spectro-electrochemical results of the selected of ECPs solutions (Table S7): (a)-(f) represent results of polymer 1-6 respectively.	38
Table S8. Summary of the types and values range of hyperparameters screened with the Set-Transformer-DFT-ECFP architecture. Optimal hyperparameter configurations for models developed using k-fold cross validation are shown on the last column. The hyperparameter optimization was performed using Optuna ²⁸	39
Table S9. Selected hyperparameters for the ML models used for comparison. Each monomer was featured using concatenation of the molecular fingerprint and the DFT descriptors. The ANN architecture involves the first layer with 15 neurons with ReLU activation and an output layer with three neurons with Softmax activation.	39
Figure S12. A live monitoring platform using streamlit library was developed to initialize the experiment and monitor it in real time. The user was selecting the monomers that were available in Chemspeed, the desired ratio step, the number of monomers to be considered and the desired color (star symbol). This generated an initial json (“init.json”) file with the information and kept updating it once the experimental measurements were done. Snapshot of the Streamlit app before the start of closed-loop experiments. The displayed data points represent the initial dataset used to initiate the experiment, prior to incorporating results from the green and orange optimization campaigns.	40
Figure S13. A .yaml file describing the sequence of actions taken to complete a closed-loop experiment. We first set the modules that were used in the workflow, i.e., Chemspeed, Tecan and UR5e, and then we defined the sequential flow with the actions of each equipment.	41
Figure S14. The outline of the initialization json file to set-up the experiment.	42
Table S10. Reaction parameters for the autonomous exploration of Orange ECPs.	45
Figure S15. Normalized UV-Vis spectra of the orange autonomous exploration samples: (a)-(c) represent the results in loop-1 to loop-3 respectively, the polymer colors are represented by the color of the spectra lines.	45
Table S11. Reaction parameters for the autonomous exploration of Green ECPs.	46

Figure S16. Normalized UV-Vis spectra of the green autonomous exploration samples: (a)-(c) represent the results in loop-1 to loop-3 respectively, the polymer colors are represented by the color of the spectra lines.46

Figure S17. Pictures of final close-loop ECP samples with the color calculated from solution state absorption spectra, real solution and thin film pictures. **a.** results of orange target ($L^* = 70$, $a^* = 35$, $b^* = 50$); **b.** results of green target ($L^* = 65$, $a^* = -10$, $b^* = 30$). All samples were spin-coated at 1000 rpm from 10 mg/mL solutions in chloroform on glass substrate. All the polymers were directly used after precipitation.47

Figure S18. Structural characterization of the final target samples in close-loop optimization. (a) NMR characterization result of orange target: $^1\text{H-NMR}$ (CDCl_3 , 500M, ppm): $\delta = 7.84\text{-}7.67$ (3.2H, m), 4.28-4.07 (4.4H, m), 4.06-3.86 (0.6H, m), 3.67-3.44 (4.4H, m), 3.38-3.25 (4.4H, m), 1.64-1.20 (19.8H, m), 0.97-0.8 (13.2H, m); (b) NMR characterization result of green target: $^1\text{H-NMR}$ (CDCl_3 , 500M, ppm): $\delta = 8.49\text{-}7.59$ (5.4H, m), 4.41-4.13 (4H, m), 3.74-3.54 (4H, m), 3.44-3.24 (4H, m), 2.17-1.96 (2.8H, m), 1.61-0.68 (45.4H, m)47

1. Materials and instruments

All chemicals and materials were purchased from common vendors (e.g., Sigma Aldrich, TCI chemical, Ambeed, etc.) and were used without further purification. Automated synthesis was performed on a Chemspeed FLEX ISYNTH synthesis platform equipped with solid dispensing, liquid handling, capping, and filtering systems. An UR5e mobile robot equipped with a two-finger gripper was used for the transfer of 96-wellplates. Nuclear magnetic resonance (NMR) spectroscopy was performed on the Bruker Ultra Shield 500 MHz NMR Spectrometers. Number-average molar mass (M_n), mass-average molar mass (M_w), and dispersity ($PD = M_w/M_n$) were measured on an Agilent HT-GPC system with the Agilent PLgel 5 μ m Mix-D separation column, calibrated using polystyrene standards with a reflective index detector, and THF as the mobile phase at a temperature of 40 °C. Absorption spectra for the automated synthesized samples were measured in chloroform solutions in a quartz 96-wellplate using a Tecan spectrophotometer. Cyclic voltammetry (CV) was performed on a Biologic SP150 potentiostat. Spectroelectrochemical data were obtained by an Agilent Cary 5000 UV/Vis-NIR spectrophotometer. Spin-coatings were performed on a Laurell spin coater. Device logo patterns were printed on PET-ITO substrates with an Optomec Aerosol Jet 5X system.

2. Electrochromic polymers (ECPs) in literature

2.1 ECP database construction

Table S1. Full ECP database from literature with monomer constructions and L*a*b* values of each polymer. ECP database was constructed based on the selected ECP literature. (Samples with * do not have absorption spectra, reported L*a*b* values were used. Abbreviations used in the table: “EDOT-BTD-EDOT”: 4,7-bis(2,3-dihydrothieno[3,4-b][1,4]dioxin-5-yl)benzo[c][1,2,5]thiadiazole; “DMP”: 3,3-dimethyl-3,4-dihydro-2H-thieno[3,4-b][1,4]dioxepine).

	Repeating Unit 1	Ratio 1	Repeating Unit 2	Ratio 2	Repeating Unit 3	Ratio 3	L*	a*	b*
1 ¹	3,3-bis(octyloxymethyl)-3,4-dihydro-2Hthieno[3,4-b][1,4]dioxepine-6-yl)	67	4,7-di (thiophen-2-yl)benzo[c][1,2,5]thiadiazole	33			59	-29	-7
2 ¹	3,3-bis(octyloxymethyl)-3,4-dihydro-2Hthieno[3,4-b][1,4]dioxepine-6-yl)	67	4,7-bis(2,3-dihydrothieno[3,4-b][1,4]dioxin-5-yl)benzo[c][1,2,5]-thiadiazole	33			66	-30	-4
3 ²	3,4-bis(2-ethylhexyloxy)thiophene	67	2,1,3-benzothiadiazole	33			64	-37	-7
4 ²	3,4-bis(2-ethylhexyloxy)thiophene	80	2,1,3-benzothiadiazole	20			57	7	-7
5 ²	3,4-bis(2-ethylhexyloxy)thiophene	86	2,1,3-benzothiadiazole	14			52	34	7
6 ²	3,4-bis(2-ethylhexyloxy)thiophene	100					81	46	15
7 ²	3,3-bis-(2-ethyl-hexyloxymethyl)-3,4-dihydro-2H-thieno[3,4-b][1,4]dioxepine	67	2,1,3-benzothiadiazole	33			64	-21	-33
8 ²	3,3-bis-(2-ethyl-hexyloxymethyl)-3,4-dihydro-2H-thieno[3,4-b][1,4]dioxepine	80	2,1,3-benzothiadiazole	20			55	-5	-17
9 ²	3,3-bis-(2-ethyl-hexyloxymethyl)-3,4-dihydro-2H-thieno[3,4-b][1,4]dioxepine	86	2,1,3-benzothiadiazole	14			51	17	-27
10 ²	3,3-bis-(2-ethyl-hexyloxymethyl)-3,4-dihydro-2H-thieno[3,4-b][1,4]dioxepine	100					68	60	-35
11 ²	3,3-bis-(2-ethyl-hexyloxymethyl)-3,4-	13	2,1,3-benzothiadiazole	7	ethylhexyloxy -	80	37	1	-20

	dihydro-2H-thieno[3,4-b][1,4]dioxepine				substituted 3,4-propylenedioxythiophene				
12 ³	3,4-propylenedioxythiophene	86	2,1,3-benzothiadiazole	14			48	-15	-5
13 ³	3,4-propylenedioxythiophene	80	2,1,3-benzothiadiazole	20			55	-32	-4
14 ³	3,4-propylenedioxythiophene	66	2,1,3-benzothiadiazole	33			68	-27	-20
15 ³	3,4-propylenedioxythiophene	50	2,1,3-benzothiadiazole	50			65	-23	-6
16 ³	3,4-propylenedioxythiophene	50	2,1,3-benzothiadiazole	50			65	-23	-17
17 ³	3,4-propylenedioxythiophene	50	2,1,3-benzothiadiazole	50			64	-36	-6
18* ⁴	3,4-di(2-ethylhexyloxy)thiophene	100					72	42	53
19* ⁴	3,4-di(2-ethylhexyloxy)thiophene	50	3,4-di(methoxy)thiophene	50			60	49	9
20 ⁵	3,3-Bis-(2-ethyl-hexyloxymethyl)-3,4-dihydro-2H-thieno[3,4-b][1,4]dioxepine	60	2,1,3-benzothiadiazole	40			45	-7	-11
21 ⁵	3,3-Bis-(2-ethyl-hexyloxymethyl)-3,4-dihydro-2H-thieno[3,4-b][1,4]dioxepine	70	2,1,3-benzothiadiazole	30			42	0.7	-12
22 ⁵	3,3-Bis-(2-ethyl-hexyloxymethyl)-3,4-dihydro-2H-thieno[3,4-b][1,4]dioxepine	75	2,1,3-benzothiadiazole	25			39	10	-12
23 ⁵	3,3-Bis-(2-ethyl-hexyloxymethyl)-3,4-dihydro-2H-thieno[3,4-b][1,4]dioxepine	76	2,1,3-benzothiadiazole	24			48	50	-30
24* ⁵	3,3-bis-(2-ethyl-hexyloxymethyl)-3,4-dihydro-2H-thieno[3,4-b][1,4]dioxepine	80	2,1,3-benzothiadiazole	20			40	24	-13
25* ⁶	phenylene	50	3,3-bis-(2-ethyl-hexyloxymethyl)-3,4-dihydro-2H-thieno[3,4-b][1,4]dioxepine	50			60	0	85
26 ⁷	3,4-di(2-ethylhexyloxy)thiophene	80	2,1,3-benzothiadiazole	20			90	-20	-45

27 ⁷	3,4-di(methoxy)thiophene	40	3,4-di(2-ethylhexyloxy)thiophene	40	2,1,3-benzothiadiazole	20	53	-26	-2
28 ⁷	3,4-di(2-ethylhexyloxy)thiophene	40	3,4-di(methoxy)thiophene	40	2,1,3-benzothiadiazole	20	54	-16	-1
29* ⁷	3,3-bis-(2-ethyl-hexyloxymethyl)-3,4-dihydro-2H-thieno[3,4-b][1,4]dioxepine	40	3,4-di(methoxy)thiophene	40	2,1,3-benzothiadiazole	20	53	-22	4
30* ⁷	3,3-bis-(2-ethyl-hexyloxymethyl)-3,4-dihydro-2H-thieno[3,4-b][1,4]dioxepine	80	2,1,3-benzothiadiazole	20			59	-23	1
31* ⁷	thiophene	40	3,3-bis(octyloxymethyl)-3,4-dihydro-2H-thieno[3,4-b][1,4]dioxepine-6-yl)	40	2,1,3-benzothiadiazole	20	54	-23	3
32* ⁷	3,3-bis-(2-ethyl-hexyloxymethyl)-3,4-dihydro-2H-thieno[3,4-b][1,4]dioxepine	40	thiophene	40	2,1,3-benzothiadiazole	20	58	-18	-6
33 ⁸	3,3-bis-(2-ethyl-hexyloxymethyl)-3,4-dihydro-2H-thieno[3,4-b][1,4]dioxepine	50	phenylene	50			97	-9	73
34 ⁸	3,3-bis-(2-ethyl-hexyloxymethyl)-3,4-dihydro-2H-thieno[3,4-b][1,4]dioxepine	50	9,9-Dioctyl-2,7-fluorene	50			98	-11	58
35 ⁸	3,3-bis-(2-ethyl-hexyloxymethyl)-3,4-dihydro-2H-thieno[3,4-b][1,4]dioxepine	50	9-octyl-9H-2,7-carbazole	50			97	-10	76
36 ⁸	3,3-bis-(2-ethyl-hexyloxymethyl)-3,4-dihydro-2H-thieno[3,4-b][1,4]dioxepine	50	dimethoxyphenylene	50			86	25	60
37 ⁸	3,3-bis-(2-ethyl-hexyloxymethyl)-3,4-dihydro-2H-thieno[3,4-b][1,4]dioxepine	50	2,7-pyrene	50			92	-10	48
38 ⁸	3,3-bis-(2-ethyl-hexyloxymethyl)-3,4-dihydro-2H-thieno[3,4-b][1,4]dioxepine	50	dimethoxyphenylene	25	phenylene	25	84	15	60

39 ⁸	3,3-bis-(2-ethyl-hexyloxy)methyl)-3,4-dihydro-2H-thieno[3,4-b][1,4]dioxepine	67	2,7-pyrene	33			91	-3	71
40 ⁸	3,3-bis-(2-ethyl-hexyloxy)methyl)-3,4-dihydro-2H-thieno[3,4-b][1,4]dioxepine	67	phenylene	33			86	31	43
41 ⁹	3,3-bis-(2-ethyl-hexyloxy)methyl)-3,4-dihydro-2H-thieno[3,4-b][1,4]dioxepine	67	3,4-ethylenedioxythiophene	33			53	19	-48
42 ⁹	3,3-bis-(2-ethyl-hexyloxy)methyl)-3,4-dihydro-2H-thieno[3,4-b][1,4]dioxepine	50	3,4-ethylenedioxythiophene	50			46	20	-47
43 ⁹	3,3-bis-(2-ethyl-hexyloxy)methyl)-3,4-dihydro-2H-thieno[3,4-b][1,4]dioxepine	50	3,4-ethylenedioxythiophene	50			55	4	-48
44 ⁹	3,4-bis(2-ethylhexyloxy)thiophene	50	3,3-bis-(2-ethyl-hexyloxy)methyl)-3,4-dihydro-2H-thieno[3,4-b][1,4]dioxepine	50			56	60	-24
45 ⁹	3,4-bis(2-ethylhexyloxy)thiophene	50	3,4-propylenedioxythiophene	50			55	57	-34
46 ⁹	3,4-bis(2-ethylhexyloxy)thiophene	50	3,4-ethylenedioxythiophene	50			49	42	-38
47 ⁹	3,3-bis-(2-ethyl-hexyloxy)methyl)-3,4-dihydro-2H-thieno[3,4-b][1,4]dioxepine	100					57	34	-41
48 ⁹	3,4-bis(2-ethylhexyloxy)thiophene	100					65	45	38
49 ⁹	3,4-bis(2-ethylhexyloxy)thiophene	50	3,4-ethylenedioxythiophene	50			45	32	-41
50 ¹⁰	2-hexyldecyloxy (3,4-propylenedioxythiophene)	50	3,4-ethylenedioxyselenophene	50			53	-4	-47
51 ¹⁰	2-hexyldecyloxy (3,4-propylenedioxythiophene)	50	3,4-ethylenedioxythiophene	50			52	12	-48
52 ¹⁰	2-hexyldecyloxy (3,4-propylenedioxythiophene)	100					42	58	-38
53* ¹⁰	3,4-ethylenedioxythiophene	100					35	10	-54
54 ¹¹	3,4-bis(2-ethylhexyloxy)thiophene	75	EDOT-BTD-EDOT	25			26	33	-32

55 ¹¹	3,3-Bis-(2-ethylhexyloxymethyl)-3,4-dihydro-2H-thieno[3,4-b][1,4]dioxepine	75	EDOT-BTD-EDOT	25			52	-14	44
56 ¹¹	3,4-bis(2-ethylhexyloxy)thiophene	37.5	3,3-bis-(2-ethylhexyloxymethyl)-3,4-dihydro-2H-thieno[3,4-b][1,4]dioxepine	37.5	EDOT-BTD-EDOT	25	32	24	-8
57 ¹¹	3,4-bis(2-ethylhexyloxy)thiophene	35	3,3-bis-(2-ethylhexyloxymethyl)-3,4-dihydro-2H-thieno[3,4-b][1,4]dioxepine	35	EDOT-BTD-EDOT	30	39	14	-4
58 ¹¹	3,4-bis(2-ethylhexyloxy)thiophene	32.5	3,3-bis-(2-ethylhexyloxymethyl)-3,4-dihydro-2H-thieno[3,4-b][1,4]dioxepine	32.5	EDOT-BTD-EDOT	35	44	7	0
59 ¹	3,3-bis-(2-ethylhexyloxymethyl)-3,4-dihydro-2H-thieno[3,4-b][1,4]dioxepine	67	dimethoxyphenylene	33			89	-4	85
60 ¹	3,4-bis(2-ethylhexyloxy)thiophene	67	dimethoxyphenylene	33			86	25	74
61 ¹	3,3-bis-(2-ethylhexyloxymethyl)-3,4-dihydro-2H-thieno[3,4-b][1,4]dioxepine	34	dimethoxyphenylene	33	3,4-propylenedioxythiophene	33	87	21	73
62 ¹²	2-hexyldecyloxy(3,4-propylenedioxythiophene)	67	3,4-ethylenedioxythiophene	33			33	32	-63
63 ¹²	2-hexyldecyloxy(3,4-propylenedioxythiophene)	50	3,4-ethylenedioxythiophene	50			30	27	-58
64 ¹²	2-hexyldecyloxy(3,4-propylenedioxythiophene)	50	3,4-ethylenedioxythiophene	50			37	12	-63
65 ¹²	2-hexyldecyloxy(3,4-propylenedioxythiophene)	33	3,4-ethylenedioxythiophene	67			34	10	-56
66 ^{*4}	3,4-dioctylthiophene	50	3,4-ethylenedioxythiophene	50			65	24	36
67 ^{*4}	3,4-dioctylthiophene	50	3,4-propylenedioxythiophene	50			60	68	34
68 ^{*4}	3,4-dioctylthiophene	50	3,4-(1,3-dimethylpropylene)dioxythiophene	50			87	23	59
69 ^{*4}	3,4-dioctylthiophene	50	3,4-dimethoxythiophene	50			82	25	62
70 ¹³	3,4-propylenedioxythiophene	67	phenylene	33			85	9	67

71 ¹³	3,4-dioctylthiophene	50	3,4-(1,3-dimethylpropylene)dioxythiophene	50			82	29	64
72 ¹³	3,4-propylenedioxythiophene	67	3,3-bis-(2-ethylhexyloxymethyl)-3,4-dihydro-2H-thieno[3,4-b][1,4]dioxepine	33			59	53	-23
73 ¹³	3,4-bis(2-ethylhexyloxy)thiophene	67	2,1,3-benzothiadiazole	33			69	-30	-10
74 ¹³	3,4-bis(2-ethylhexyloxy)thiophene	80	2,1,3-benzothiadiazole	20			60	-32	4
75 ¹⁴	3,4-dioctylthiophene	50	3,3-bis-(2-ethylhexyloxymethyl)-3,4-dihydro-2H-thieno[3,4-b][1,4]dioxepine	25	2,2'-bis(3,4-ethylenedioxythiophene)	25	66	20	43
76 ¹⁴	3,4-dioctylthiophene	40	3,3-bis-(2-ethylhexyloxymethyl)-3,4-dihydro-2H-thieno[3,4-b][1,4]dioxepine	40	2,2'-bis(3,4-ethylenedioxythiophene)	20	51	36	6
77 ¹⁴	3,4-dioctylthiophene	40	3,3-bis-(2-ethylhexyloxymethyl)-3,4-dihydro-2H-thieno[3,4-b][1,4]dioxepine	20	2,2'-bis(3,4-ethylenedioxythiophene)	40	48	32	-2
78 ¹⁴	3,4-dioctylthiophene	34	3,3-bis-(2-ethylhexyloxymethyl)-3,4-dihydro-2H-thieno[3,4-b][1,4]dioxepine	33	2,2'-bis(3,4-ethylenedioxythiophene)	33	44	30	-17
79 ¹⁴	3,4-dioctylthiophene	25	3,3-bis-(2-ethylhexyloxymethyl)-3,4-dihydro-2H-thieno[3,4-b][1,4]dioxepine	50	2,2'-bis(3,4-ethylenedioxythiophene)	25	47	30	-28
80 ¹⁴	3,4-dioctylthiophene	20	3,3-bis-(2-ethylhexyloxymethyl)-3,4-dihydro-2H-thieno[3,4-b][1,4]dioxepine	60	2,2'-bis(3,4-ethylenedioxythiophene)	20	46	26	-29
81 ¹⁴	3,4-dioctylthiophene	20	3,3-bis-(2-ethylhexyloxymethyl)-3,4-dihydro-2H-thieno[3,4-b][1,4]dioxepine	40	2,2'-bis(3,4-ethylenedioxythiophene)	40	45	17	-38
82 ¹⁴	3,4-dioctylthiophene	25	3,3-bis-(2-ethylhexyloxymethyl)-3,4-dihydro-2H-thieno[3,4-b][1,4]dioxepine	25	2,2'-bis(3,4-ethylenedioxythiophene)	50	44	18	-40
83 ¹⁵	acyclic dioxythiophene	100					56	45	-37

84 ¹⁵	acyclic dioxothiophene	50	3,4-bis(2-ethylhexyloxy)-thiophene	50			53	55	-18
85 ¹⁵	3,4-bis(2-ethylhexyloxy)-thiophene	100					81	28	68
86 ¹⁵	3,4-(1,3-dimethylpropylene)dioxythiophene	50	3,4-bis(2-ethylhexyloxy)-thiophene	50			51	50	-33
87 ¹⁵	3,4-(1,3-dimethylpropylene)dioxythiophene	50	acyclic dioxothiophene	50			48	35	-41
88 ¹⁶	3,4-dioctylthiophene	50	ran-1,3-DMP	50			91	-7	78
89 ¹⁶	3,4-dioctylthiophene	50	reg-1,3-DMP	50			91	-7	78
90* ¹⁶	3,4-dioctylthiophene	50	DMP	50			95	-14	68
91 ¹⁶	3,4-dioctylthiophene	50	4,7-di(thiophen-2-yl)benzo[c][1,2,5]thiadiazole	50			95	-13	69
92 ¹⁶	3,4-dioctylthiophene	50	4,7-bis(2,3-dihydrothieno[3,4-b][1,4]dioxin-5-yl)benzo[c][1,2,5]-thiadiazole	50			95	-8	77
93 ¹⁷	3,3-bis-(2-ethylhexyloxymethyl)-3,4-dihydro-2H-thieno[3,4-b][1,4]dioxepine (ProDOT)	70	2,1,3-benzothiadiazole	12.5	2,7-fluorene	17.5	84	-14	-7
94 ¹⁷	3,3-bis-(2-ethylhexyloxymethyl)-3,4-dihydro-2H-thieno[3,4-b][1,4]dioxepine	70	2,1,3-benzothiadiazole	15	2,7-fluorene	15	94	-14	-36
95 ¹⁷	3,3-bis-(2-ethylhexyloxymethyl)-3,4-dihydro-2H-thieno[3,4-b][1,4]dioxepine	67.5	2,1,3-benzothiadiazole	15	2,7-fluorene	17.5	56	-12	12
96 ¹⁷	3,3-bis-(2-ethylhexyloxymethyl)-3,4-dihydro-2H-thieno[3,4-b][1,4]dioxepine	62.5	2,1,3-benzothiadiazole	17.5	2,7-fluorene	20	62	-21	21
97 ¹⁷	3,3-bis-(2-ethylhexyloxymethyl)-3,4-dihydro-2H-thieno[3,4-b][1,4]dioxepine	57.5	2,1,3-benzothiadiazole	20	2,7-fluorene	22.5	59	-26	20
98 ¹⁷	3,3-bis-(2-ethylhexyloxymethyl)-3,4-dihydro-2H-thieno[3,4-b][1,4]dioxepine	50	2,1,3-benzothiadiazole	25	2,7-fluorene	25	61	-23	11

2.2 Demonstration of large language model (LLM)-assisted data mining workflow

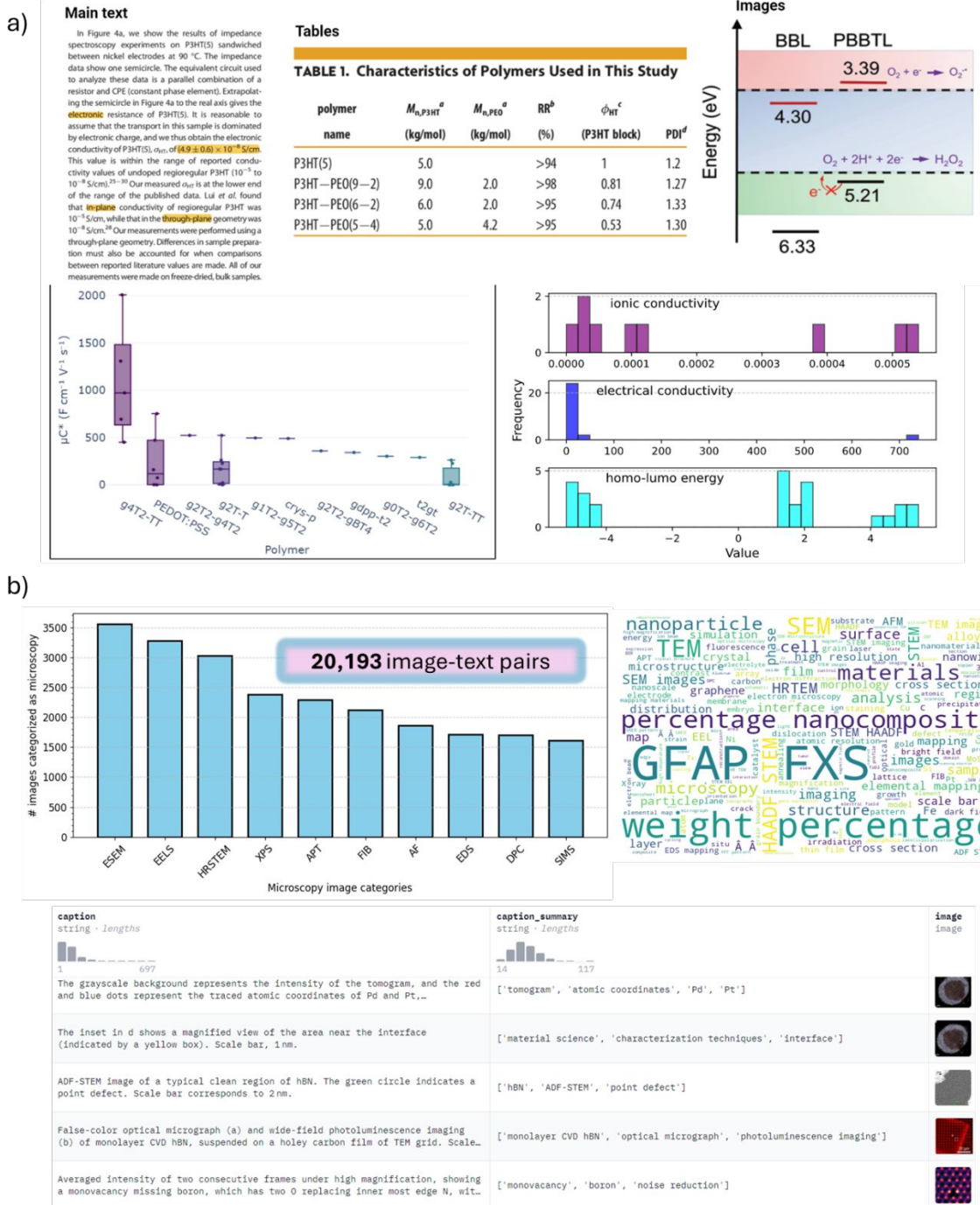


Figure S1. Two additional examples of data extracted using the LLM-assisted data mining framework. (a) OMIECs database compiled from key data sources in relevant literature. The extracted information is summarized in text, tables, and images, with a data table presenting the curated results after screening 52 papers related to OMIECs. b) Example of microscopy images extracted and tagged using the data mining pipeline. An example on how to scrape information from PDF or HTML files can be also found in the following link:

<https://katerinavr.github.io/exsclaim2.0/PDFScraper.html>

To ensure data reliability and improve automated extraction performance, we implemented a quality control and tool-enhancement workflow that extends several existing frameworks. Specifically, EXSCLAIM2.0 was developed to go beyond its original role of microscopy image annotation by incorporating a fine-tuned vision model and large language models (LLMs) for improved image and caption understanding. The Plot2Spectra module was modified to enable color-based separation of multi-line spectra, which are common in the ECP literature. We also acknowledged known limitations of DECIMER.ai in handling molecular structures with R-groups or fragmented components through manual verification and correction of all extracted SMILES strings. After automated extraction, each UV–Vis spectrum was cross-checked against its source figure to confirm the fidelity of peak positions and spectral shapes. Similarly, molecular structures were verified against published depictions, and manual corrections were applied when image quality or formatting inconsistencies caused extraction errors.

2.3 Calculating the color values of polymers in ECP database

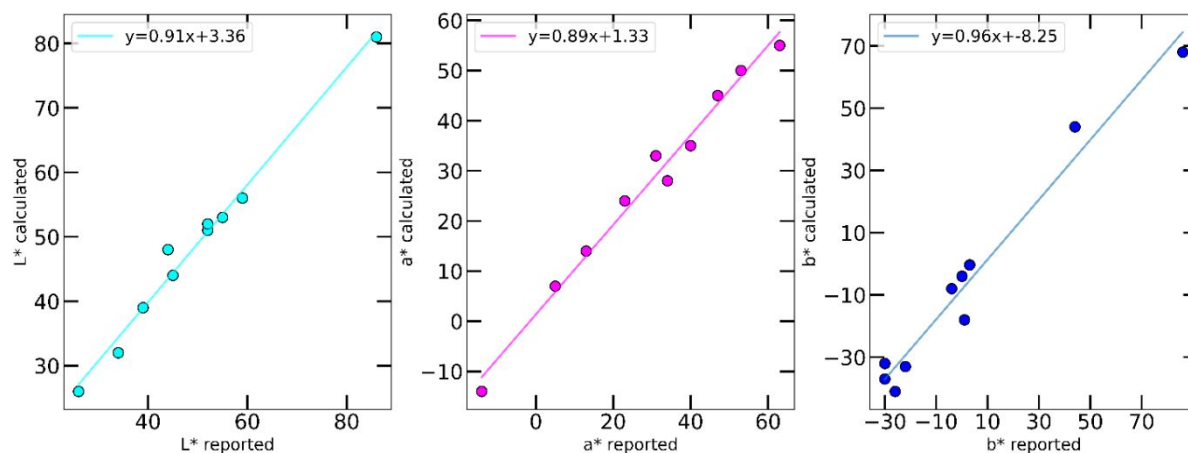


Figure S2. Correlation between the calculated and literature reported L*a*b* values of some representative samples. Samples were chosen from papers providing both absorption spectra and tabulated L*a*b* values (Polymer index 54-58, 83-87 from ECP database were chosen as the samples here).

For the transformation of the absorption spectra to the CIELAB color space, we employed the Python ‘color-science’ library using the following methodology²⁰. Initially the absorption spectra were converted to the transmittance spectra using the equation $\text{Trans} = 10^{(2 - \text{Absorbance})}$. The transmittance data were then normalized and paired with corresponding wavelengths to create a spectral distribution. Subsequently, this spectral distribution was integrated with the CIE 1931 2° Standard Observer color matching functions and the D65 standard illuminant to calculate the CIE XYZ tristimulus values¹⁸, which represent the perceptually uniform color spaces. The XYZ values were normalized by a factor of 100 to align with standard input requirements and were further converted to the CIELAB color space. This methodology was benchmarked using some representation samples from literature where both the absorption spectra and the CIELAB values were reported as shown in Fig. S2.

To ensure consistency in color measurements, all ingested spectra were normalized prior to conversion into L*, a*, and b* values. This normalization step was crucial for mitigating variations in solution concentration and sample thickness, which are inherently difficult to standardize across experimental samples. In this study, L* primarily influences color brightness, while hue and chromaticity—the key parameters of interest—are predominantly defined by a* and b* values.

Normalization constrained L^* values within a narrow range, enabling a more reliable comparison of a^* and b^* values. Without this step, variations in concentration and sample thickness would lead to inconsistencies across L^* , a^* , and b^* values, complicating the accurate analysis of hue-related trends.

2.4 ECP database summary

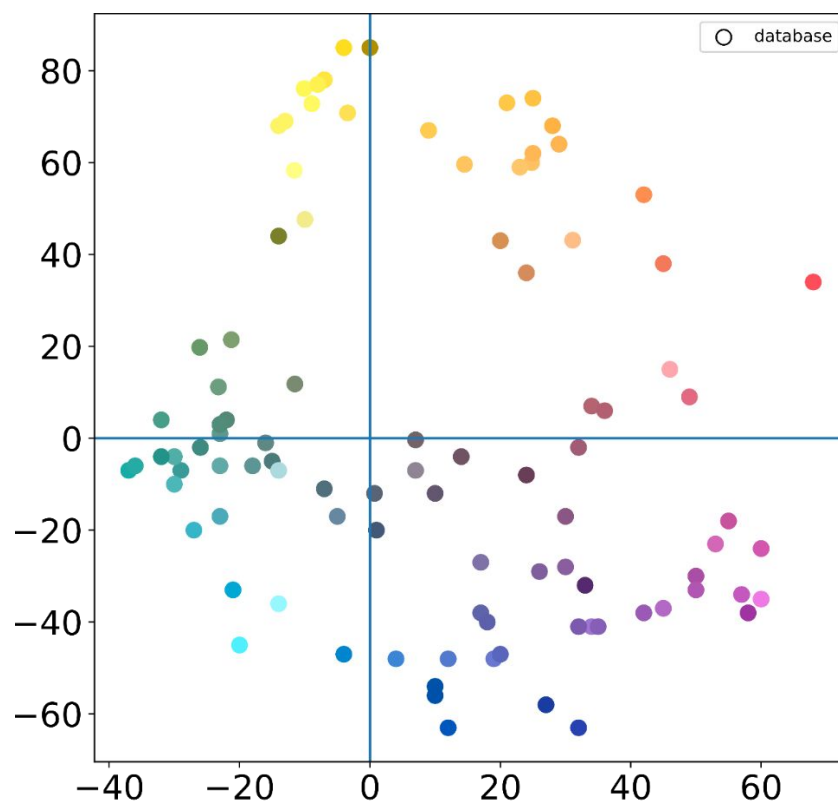


Figure S3. Scatterplot with the a^* , and b^* values of all the datapoints collected from literature (Table S1). The colors of datapoints are represented by the color of the dots.

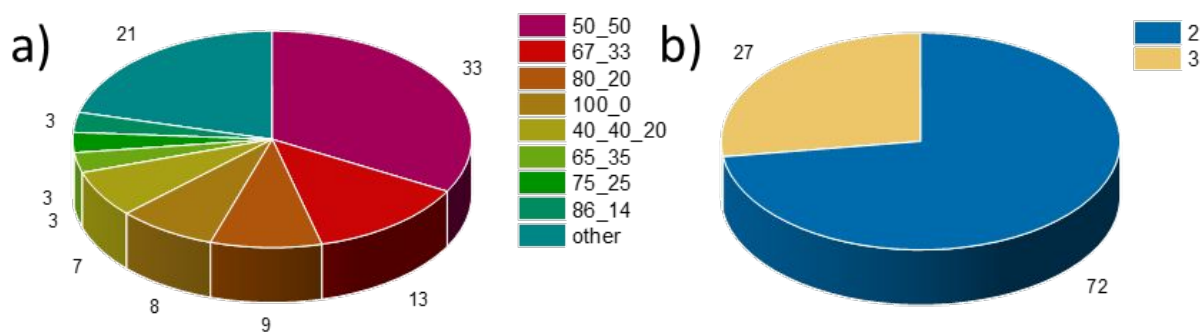


Figure S4. Distribution of the ECP database generated from literature: (a) the distribution in monomer ratios; (b) the distribution of the number of different monomers in ECP.

2.5 Extracting synthetic conditions from literature

The dataset for reaction conditions extraction includes all the papers to construct the ECP database and some additional representative Pd catalysed cross-coupling paper^{1-17, 19-25}.

The process for the named entities extraction was using three sequential Large Language Models (LLM) assistants ('gpt-4-turbo-preview'), each assigned a different task related to data mining. The first LLM assistant (LLM1) was tasked to read the text of each paper and identify only the paragraph that refers to the polymerization reaction. The second LLM assistant (LLM2) was tasked to read the synthesis paragraphs extracted from LLM1 and identify the entities: “base”, “solvents”, “ligands”, and “catalysts”. Each entity is marked using the HTML tags to highlight them. The third LLM assistant was instructed to count all the instances and return a dictionary with the main general categories (base, solvents, ligands, catalysts) as keys, elements that belong to each category as subkeys, and the number of instances as values on the subkeys. The detailed prompting of the three LLM assistants is provided below:

LLM1_prompt:

"This is a research paper related to polymers synthesis: {file_contents}\n\nYour task is to read the whole text and identify all the paragraphs that describe synthesis methods. Your output should be strictly these paragraphs without modifying their context and without adding any other wording."

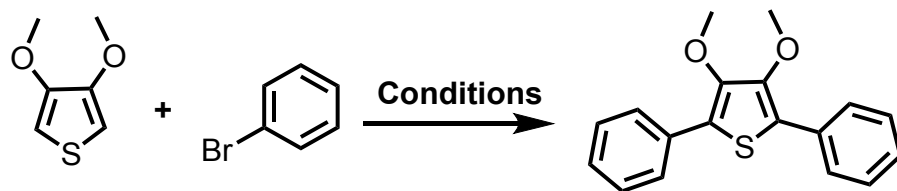
LLM2_prompt:

"This is a paragraph related to polymers synthesis.\n\nContext:\n{paragraph}\n\nYour task is to identify all the chemical elements used in the polymerization reaction only. Then generate an HTML version of the input text, marking up specific entities related to chemical elements. The specific elements that need to be identified are the following: base, solvents, ligands, and catalysts. Use HTML tags to highlight these entities. Each should have a class attribute indicating the type of the entity. Do not count any solvents used for post-reaction processing and for precipitation."

LLM3_prompt:

"This is a list with all the chemical elements used in a polymerization reaction.\n\nContext:\n{list_chemicals}\n\nYour task is to count all the instances and return a dictionary with the main general categories (base, solvents, ligands, catalysts) as keys, elements that belong to each category as subkeys and the number of instances as values on the subkeys. Some elements that are similar should be considered as the same element, e.g., THF and dry THF are the same element and should belong to the same category. Another example, FeCl₃ and iron (III) chloride is the same element and should belong to the same category. Also Pd(OAc)₂ and palladium (II) acetate."

3. Reaction condition optimization



Scheme S1. Model reaction for synthesis condition optimization.

3.1 Rational in refining the design space of synthesis condition optimization

Human input was incorporated to refine the design space, ensuring experimental feasibility while preventing unnecessary expansion. For solvent selection, frequently extracted solvents like ethyl acetate and THF were deemed undesirable due to their high volatility. Since C-H arylation reactions occur near or above their boiling points, their evaporation in the synthesis robot reactors posed a challenge. Therefore, THF, toluene, and N,N-dimethylacetamide (DMAc) were chosen as the solvents in the optimization design space. For base selection, K_3PO_4 , K_2CO_3 , and sodium tert-butoxide (t-BuONa) were included in the optimization design space. KF (which is only suitable for Suzuki reaction and may release hydrofluoric acid) and NaH (highly reactive) were excluded due to safety concerns. For ligands, we chose the ones that are more stable in ambient environment with lower prices. And for catalysts, we selected the most frequently used ones to balance the design space while maintaining experimental feasibility.

3.2 Stock solution preparation

The following reactants were dissolved in chloroform solution at 0.05mmol/mL for 20-30mL: 3,4-dimethoxythiophene, bromobenzene, 1,1'-bis(diphenylphosphino)ferrocene (dppf), tri(o-tolyl)phosphine ($P(o\text{-tol})_3$), and pivalic acid. Palladium acetate ($Pd(OAc)_2$) and Tris(dibenzylideneacetone)dipalladium ($Pd_2(dba)_3$) were dissolved in N,N-Dimethylacetamide (DMAc) and toluene at 0.005mmol/mL (10mL each).

3.3 Automated Reaction workflow

After the preparation of stock solutions, the solution vials were transferred to predetermined zones within the robotic reaction hood and purged with nitrogen gas. Once the workflow was initiated,

and the reaction parameters were imported, stock solutions containing monomer reactants (1 mL each), phosphine ligands (200 μ L), and pivalic acid (200 μ L) were dispensed into the designated reaction vials within the reaction chamber using a liquid handler (Table S2). Upon the successful transfer of stock solutions, the reaction chamber was transitioned to the "close under vacuum" position, facilitating the removal of chloroform under vacuum conditions. Subsequently, either 0.2 mmol of K_2CO_3 or t-BuONa was introduced to the assigned vials through solid gravity dispensing unit. Additionally, 1 mL catalyst/solvent solution was added to the assigned reaction vials before the reactions. Following these additions, the vials underwent nitrogen purging and were subsequently heated to 110 $^{\circ}C$ for 15 hours with refluxing to promote the reactions. After the reactions finished, 10 μ L of the reaction mixture was added to 500 μ L deuterated chloroform ($CDCl_3$). After measurements, the NMR spectra were analyzed manually to calculate the yield based on the 3,4-dimethoxythiophene conversion.

Table S2. Reaction optimization parameters and yield.

<i>Reaction No.</i>	Catalyst	Ligand	Base	Solvent	Yield (%)
1	Pd(OAc) ₂	No Ligand	K ₂ CO ₃	DMAc	94
2	Pd(OAc) ₂	No Ligand	K ₂ CO ₃	Toluene	9
3	Pd ₂ (dba) ₃	No Ligand	K ₂ CO ₃	DMAc	8
4	Pd ₂ (dba) ₃	No Ligand	K ₂ CO ₃	Toluene	6
5	Pd(OAc) ₂	No Ligand	t-BuONa	DMAc	21
6	Pd(OAc) ₂	No Ligand	t-BuONa	Toluene	16
7	Pd ₂ (dba) ₃	No Ligand	t-BuONa	DMAc	2
8	Pd ₂ (dba) ₃	No Ligand	t-BuONa	Toluene	3
9	Pd(OAc) ₂	dppf	K ₂ CO ₃	DMAc	58
10	Pd(OAc) ₂	dppf	K ₂ CO ₃	Toluene	15
11	Pd ₂ (dba) ₃	dppf	K ₂ CO ₃	DMAc	21
12	Pd ₂ (dba) ₃	dppf	K ₂ CO ₃	Toluene	4
13	Pd(OAc) ₂	dppf	t-BuONa	DMAc	30
14	Pd(OAc) ₂	dppf	t-BuONa	Toluene	29
15	Pd ₂ (dba) ₃	dppf	t-BuONa	DMAc	2
16	Pd ₂ (dba) ₃	dppf	t-BuONa	Toluene	0
17	Pd(OAc) ₂	P(o-tol) ₃	K ₂ CO ₃	DMAc	81
18	Pd(OAc) ₂	P(o-tol) ₃	K ₂ CO ₃	Toluene	39
19	Pd ₂ (dba) ₃	P(o-tol) ₃	K ₂ CO ₃	DMAc	0
20	Pd ₂ (dba) ₃	P(o-tol) ₃	K ₂ CO ₃	Toluene	4
21	Pd(OAc) ₂	P(o-tol) ₃	t-BuONa	DMAc	25
22	Pd(OAc) ₂	P(o-tol) ₃	t-BuONa	Toluene	5
23	Pd ₂ (dba) ₃	P(o-tol) ₃	t-BuONa	DMAc	25
24	Pd ₂ (dba) ₃	P(o-tol) ₃	t-BuONa	Toluene	1
25	Pd(OAc) ₂	No Ligand	K ₃ PO ₄	Toluene	59
26	Pd(OAc) ₂	P(o-tol) ₃	K ₃ PO ₄	Toluene	50
27	Pd(OAc) ₂	dppf	K ₃ PO ₄	Toluene	19
28	Pd(OAc) ₂	No Ligand	K ₃ PO ₄	DMAc	98
29	Pd(OAc) ₂	P(o-tol) ₃	K ₃ PO ₄	DMAc	42
30	Pd(OAc) ₂	dppf	K ₃ PO ₄	DMAc	97

31	Pd ₂ (dba) ₃	No Ligand	K ₃ PO ₄	Toluene	17
32	Pd ₂ (dba) ₃	P(o-tol) ₃	K ₃ PO ₄	Toluene	16
33	Pd ₂ (dba) ₃	dppf	K ₃ PO ₄	Toluene	11
34	Pd ₂ (dba) ₃	No Ligand	K ₃ PO ₄	DMAc	1
35	Pd ₂ (dba) ₃	P(o-tol) ₃	K ₃ PO ₄	DMAc	11
36	Pd ₂ (dba) ₃	dppf	K ₃ PO ₄	DMAc	2
37	Pd(OAc) ₂	No Ligand	K ₃ PO ₄	THF	58
38	Pd(OAc) ₂	P(o-tol) ₃	K ₃ PO ₄	THF	42
39	Pd(OAc) ₂	dppf	K ₃ PO ₄	THF	22
40	Pd(OAc) ₂	No Ligand	K ₂ CO ₃	THF	66
41	Pd(OAc) ₂	P(o-tol) ₃	K ₂ CO ₃	THF	66
42	Pd(OAc) ₂	dppf	K ₂ CO ₃	THF	14
43	Pd(OAc) ₂	No Ligand	t-BuONa	THF	68
44	Pd(OAc) ₂	P(o-tol) ₃	t-BuONa	THF	56
45	Pd(OAc) ₂	dppf	t-BuONa	THF	29
46	Pd ₂ (dba) ₃	No Ligand	K ₃ PO ₄	THF	21
47	Pd ₂ (dba) ₃	P(o-tol) ₃	K ₃ PO ₄	THF	6
48	Pd ₂ (dba) ₃	dppf	K ₃ PO ₄	THF	23
49	Pd ₂ (dba) ₃	No Ligand	K ₂ CO ₃	THF	6
50	Pd ₂ (dba) ₃	P(o-tol) ₃	K ₂ CO ₃	THF	24
51	Pd ₂ (dba) ₃	dppf	K ₂ CO ₃	THF	11
52	Pd ₂ (dba) ₃	No Ligand	t-BuONa	THF	41
53	Pd ₂ (dba) ₃	P(o-tol) ₃	t-BuONa	THF	22
54	Pd ₂ (dba) ₃	dppf	t-BuONa	THF	5

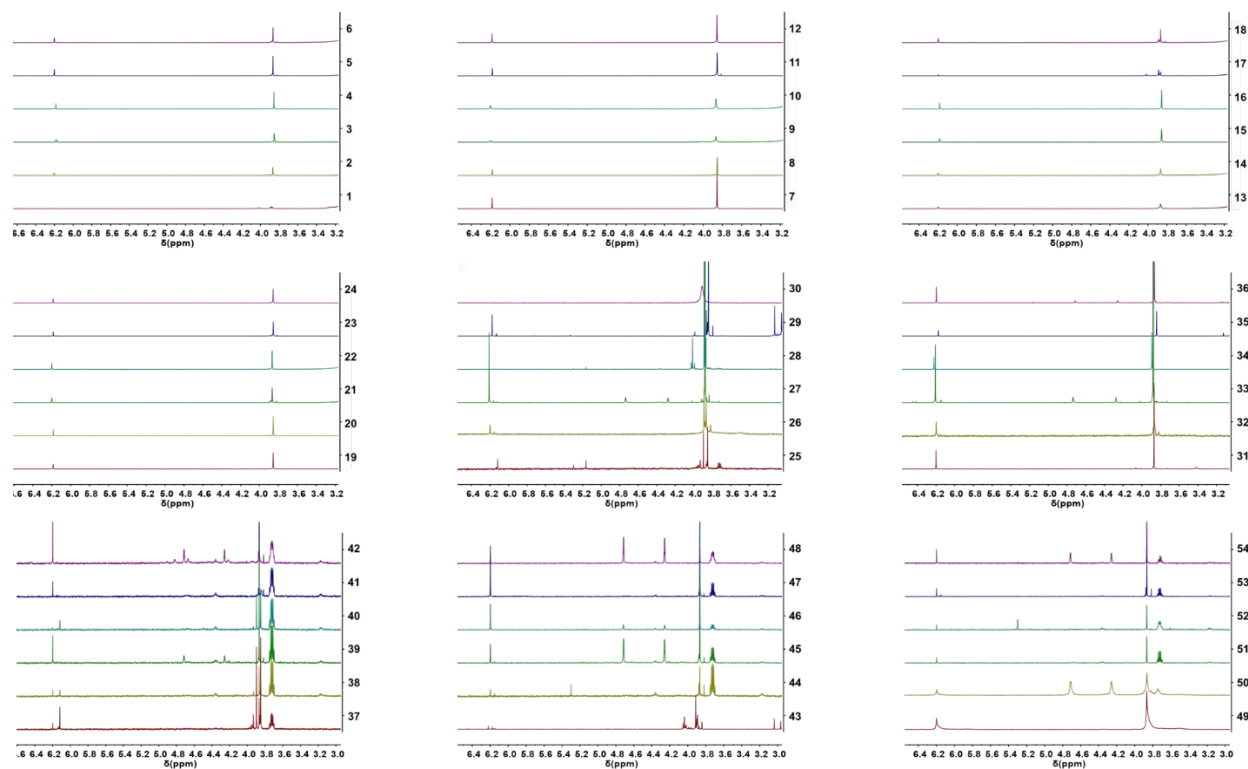


Figure S5. NMR spectra of the products from reaction optimization experiments, with reaction No. noted on the right side of the spectrum. The yields of the reactions (indicated in Table S2) were calculated based on the integration of the peak from unreacted thiophene aromatic proton at $\delta = 6.2$ ppm (as x) and the integration of the peak from all methoxy proton at $\delta = 3.8$ -4.0 ppm (as y).

The formula used to calculate the yield is:

$$\text{Yield (\%)} = \frac{(y-3x)}{y} * 100\%$$

4. Copolymer Set-transformer model design

4.1 System overview

The Copolymer-Set-Transformer is a neural network inspired from the Set Transformer model architecture to be appropriate for processing sets of data points with are order invariant²⁶. The model was designed for a dual-purpose task of predicting physical properties (CIELab values) and optionally absorption spectra of copolymers. As shown in Figure S6, the model involves a sequential encoder and two parallel decoders. The use of the absorption spectra decoder is optional and configurable and is activated setting flag: “*use_abs_decoder*” = True. The model is also providing uncertainty estimation for each prediction. If only the color prediction decoder is used, then the uncertainty is calculated from the average of the outputs using the dropout technique. If both decoders are employed, the uncertainty is calculated by the difference in the Lab values derived directly from the color prediction decoder and the $L^*a^*b^*$ calculated from the absorption spectra which is the output of absorption spectra decoder.

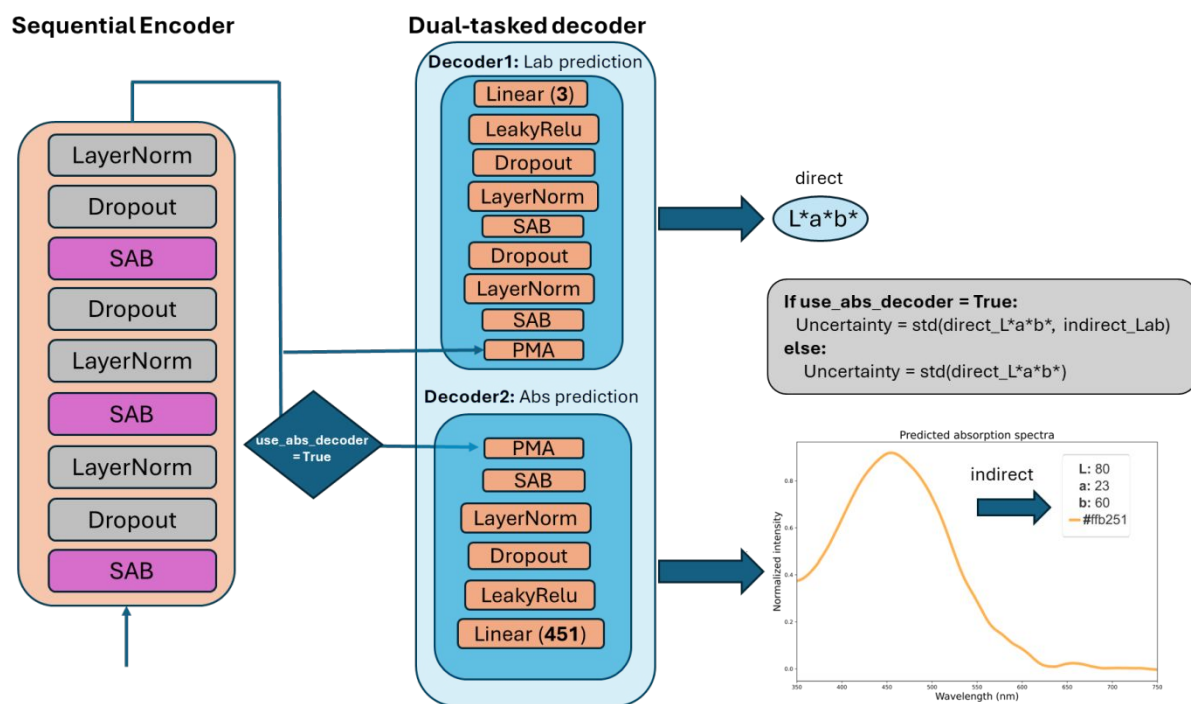


Figure S6. Schematic of the model architecture consisting of a sequential encoder (pink part) and a dual decoder (blue part). After decoder two predictions were performed, one is directly to $L^*a^*b^*$ values; the other is predicted to the absorption spectra first, then converted to $L^*a^*b^*$ values.

4.2 Architecture

Encoder. The encoder consists of a series of Self-Attention Blocks (SAB), normalization layers and dropout layers after the SAB blocks to prevent overfitting and stabilize the training process.

Decoders. The decoders consist of Pooling Multihead Attention (PMA) blocks and SAB blocks. Two separate decoding tasks, i.e., CIELab values and absorption spectrum, are performed by a primary and secondary decoders, respectively. The primary decoder employs linear transformation along with other output layers to output three numbers (L^* , a^* and b^*). The loss of the primary encoder was computed using a weighted MSE to emphasize that the importance of the a^* , b^* components is higher than the L^* . Optionally, the secondary decoder predicts the absorption spectra, but the output was tailored to a 451- dimensional absorption spectrum. The optional secondary decoder was used a standard MSE loss.

The accuracies of a^* and b^* values are prioritized using a weighted mean squared error (MSE) loss function:

$$MSE = \frac{1}{3} [0.5 \cdot (\hat{Y}_L - Y_L)^2 + (\hat{Y}_a - Y_a)^2 + (\hat{Y}_b - Y_b)^2]$$

Where $\hat{Y}$ is the predicted value, Y is the target value, the weight of L^* is 0.5, and the weights of a^* and b^* are 1.

For model evaluation, we employed both the coefficient of determination (R^2) and the Spearman rank correlation coefficient. R^2 was computed for each lab individually and then averaged across all labs to provide the “Overall” R^2 values reported in Fig. S7. While the training loss function is a modified mean squared error (MSE) suited for non-linear models, R^2 remains a useful measure of explained variance and facilitates direct comparison with other regression-based approaches. In addition, Spearman correlation was included as it captures monotonic trends between predicted and experimental values, offering a complementary perspective on model performance.

4.3 ECPs representations

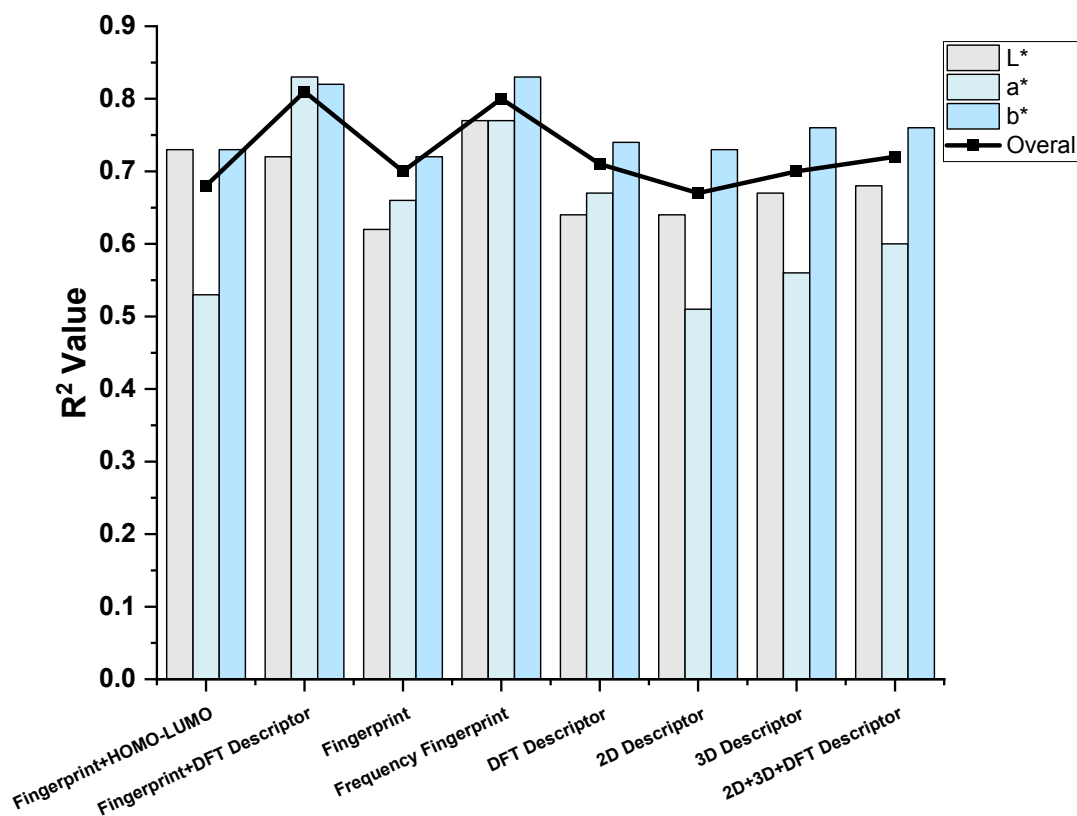


Figure S7. Comparison of prediction accuracy using different monomer representation techniques. The descriptor made by “Fingerprint with DFT Descriptor” gave the best overall performance and was chosen as final descriptor.

Table S3. Detailed list of the DFT descriptors which were calculated using the Auto-QChem package²⁷.

Descriptor Name	Descriptor Explanation
dipole	Dipole moment
molar_mass	Molar mass
molar_volume	Molar volumes
electronic_spatial_extent	Molecular Features
homo_energy	HOMO energy
lumo_energy	LUMO energy
electronegativity	Electronegativity
hardness	Hardness
ES_root_dipole	Root mean square of dipole moment considering the conformers
ES_root_molar_volume	Root mean square of molar volume considering the conformers

ES_root_electronic_spatial_extent	Electronic spatial extent
ES_transition_1	Excitation energy of 1 st state
ES_transition_2	Excitation energy of 2 nd state
ES_transition_3	Excitation energy of 3 rd state
ES_transition_4	Excitation energy of 4 th state
ES_transition_5	Excitation energy of 5 th state
ES_transition_6	Excitation energy of 6 th state
ES_transition_7	Excitation energy of 7 th state
ES_transition_8	Excitation energy of 8 th state
ES_transition_9	Excitation energy of 9 th state
ES_transition_10	Excitation energy of 10 th state
ES_osc_strength_1	Oscillator strength of 1 st state
ES_osc_strength_2	Oscillator strength of 2 nd state
ES_osc_strength_3	Oscillator strength of 3 rd state
ES_osc_strength_4	Oscillator strength of 4 th state
ES_osc_strength_5	Oscillator strength of 5 th state
ES_osc_strength_6	Oscillator strength of 6 th state
ES_osc_strength_7	Oscillator strength of 7 th state
ES_osc_strength_8	Oscillator strength of 8 th state
ES_osc_strength_9	Oscillator strength of 9 th state
ES_osc_strength_10	Oscillator strength of 10 th state

5. High-throughput synthesis of ECPs for model refinement

5.1 Stock solution preparation

The following reactants were dissolved in chloroform at 0.05mmol/mL for 20-30 mL: 3,4-propylenedioxythiophene (ProDOT; Due to reaction constraints, C-H end-capped monomers must be present in a 1:1 ratio with C-Br end-capped monomers to ensure efficient polymerization. Since ProDOT is the only available C-H end-capped monomer in our system, its proportion is necessarily fixed at 50% of the total monomer content, serving as the C-H functional end-group), dibromo-ProDOT, 1,4-dibromobenzene (Ben-2Br), dibromo-benzothiadiazole (BTD-2Br), 2,5-dibromo-3,4-dimethoxythiophene (DMoT-2Br), 2,5-dibromo-3,4-ethylenedioxythiophene (EDOT-2Br), pivalic acid. Pd(OAc)₂ was dissolved in DMAc at 0.005 mmol/mL.

5.2 Reaction workflow

After the preparation of stock solutions, the solution vials were transferred to the predetermined zones within the robotic reaction hood and purged with nitrogen gas. Once the workflow was initiated, and the reaction parameters were imported, stock solutions containing monomer reactants were dispensed into the designated reaction vials within the reaction chamber using a liquid handler (based on Table S3-S5, the ratio of dibromo-ProDOT was calculated by $\text{dibromo-ProDOT\%} = \text{ProDOT\%} - 50\%$, total 2 mL stock solution of monomers (1mL of ProDOT, 1mL of all used dibromo- monomers), and 200 μL of pivalic acid). Upon the successful transfer of stock solutions, the reaction chamber was transitioned to the "close under vacuum" position, facilitating the removal of chloroform under vacuum conditions. Subsequently, 0.2 mmol of K₂CO₃ solid was added dispensing unit to the reaction vials. After solid transfer, 1 mL of Pd(OAc)₂/DMAc solution were added to the reaction vials by syringe pump. Following these additions, the vials underwent nitrogen purging and were then heated to 110 °C for 15 hours with refluxing to promote the reaction. Upon completion of the reactions, 2 mL of chloroform and 15 mL of methanol were successively transferred to each reaction vial. Following the precipitation of the polymer, the reaction mixtures underwent filtration using a vacuum needle and filter. After liquid removal, the resulting solid polymer products were redissolved in 5 mL of chloroform. Subsequently, 50 μL of the polymer solution for each sample was carefully transferred to pre-assigned positions in a 96-well plate, following by adding 350 μL of chloroform to the same position. Once all samples from the entire batch were transferred to the 96-well plate, the plate was autonomously conveyed by the

robotic arm to the UV-vis spectrometer for absorption measurements. Failed reactions (e.g., those yielding minimal or no polymer after precipitation and filtration) are detected by low absorption intensity in the visible range, which automatically initiates the error-handling protocol but cannot be resolved within it. Similarly, when a mechanical error occurs, the system pauses automatically. In both cases, the system alerts the researcher and pauses for human guidance outside the automated workflow. After each batch, GPC characterizations were performed by taking 1mg/mL polymer samples solution in THF after filtration to confirm the Mn and Mw of polymer samples were more than 10k.

Table S4. Batch 1 reaction parameters and the resulting L*a*b* values for high-throughput synthesis of ECP for training dataset.

<i>Batch 1 Reaction</i>	ProDOT (%)	Ben- 2Br	BTD- 2Br	DMoT- 2Br (%)	EDOT- 2Br (%)	L*	a*	b*	Mn (kDa)
1	50	50				93.2	-7.9	81.3	14.9
2	60	40				83.3	21.0	70.3	17.7
3	70	30				70.2	46.8	45.5	10.2
4	80	20				64.6	59.2	17.3	10.6
5	90	10				59.4	66.2	-9.7	10.3
6	100	0				58.0	66.5	-15.9	10.3
7	50		50			52.1	3.4	-42.5	9.7
8	60		40			53.4	-3.6	-37.7	11.8
9	70		30			53.8	-13.0	-26.0	13.3
10	80		20			47.3	-7.6	-14.9	9.8
11	90		10			43.7	27.1	-13.6	10.3
12	100		0			58.0	66.5	-15.9	10.7
13	50			50		57.6	52.0	11.9	9.6
14	60			40		60.1	52.8	14.7	12.5
15	70			30		61.1	53.1	18.0	14.8
16	80			20		59.5	57.9	6.1	13.2
17	90			10		63.0	53.0	22.0	23.1
18	100			0		58.0	66.5	-15.9	10.8
19	50				50	52.5	62.3	-39.9	7.2
20	60				40	52.8	60.0	-44.0	12.4
21	70				30	53.0	59.9	-39.0	13.1
22	80				20	54.0	62.6	-32.8	16.0
23	90				10	56.5	65.5	-19.2	11.6
24	100				0	58.0	66.5	-15.9	10.7

Table S5. Batch 2 reaction parameters and the resulting L*a*b* values for high-throughput synthesis of ECP for training dataset.

Batch 2 Reaction	ProDOT (%)	Ben- 2Br	BTB- 2Br (%)	DMoT- 2Br (%)	EDOT- 2Br (%)	L*	a*	b*	Mn (kDa)
1	87.5		12.5			40.6	18.0	-19.0	16.7
2	85		15			40.6	15.1	-22.2	14.4
3	82.5		17.5			42.3	5.0	-19.4	10.8
4	80		20			47.3	-7.6	-14.9	10.3
5	77.5		22.5			47.5	-6.7	-16.9	12.7
6	75		25			49.5	-9.0	-19.2	16.1
7	72.5		27.5			51.1	-10.7	-21.5	17.6
8	90	5	5			42.1	21.4	-10.8	10.3
9	80	10	10			46.8	-7.5	-13.6	11.0
10	70	15	15			54.6	-14.7	-26.1	9.9
11	60	20	20			57.3	-6.7	-42.9	10.1
12	50	25	25			56.8	-0.6	-48.5	11.1
13	50	40			10	81.8	22.8	67.3	12.2
14	70	15			15	71.9	46.6	41.8	11.8
15	90	5			5	64.6	58.1	16.0	11.3
16	60			20	20	62.7	54.9	20.0	11.3
17	80			10	10	55.6	64.0	-20.8	10.3
18	70		10	20		42.4	25.9	-20.8	9.9
19	50		20	30		43.0	7.9	-21.7	8.9
20	50		30	20		50.2	2.5	-37.2	13.1
21	50		40	10		49.4	11.9	-44.7	10.1
22	50		20		30	58.0	-2.7	-48.5	9.7
23	50		40		10	57.1	-11.4	-35.7	10.5
24	70		10		20	50.3	-12.8	-16.1	9.2

Table S6. Batch 3 reaction parameters and the resulting L*a*b* values for high-throughput synthesis of ECP for training dataset.

Batch 3 Reaction	ProDOT (%)	Ben- 2Br	BTD- 2Br	DMoT- 2Br (%)	EDOT- 2Br (%)	L*	a*	b*	Mn (kDa)
1	80		10		10	51.1	31.9	12.5	17.0
2	90		5		5	55.9	40.2	13.9	14.8
3	50	25		25		61.9	54.6	30.8	17.4
4	50	30		20		57.5	64.0	-10.7	18.7
5	50	20		30		82.4	16.2	68.9	13.5
6	50	10		40		87.3	9.1	76.6	13.0
7	50	40		10		55.9	65.3	-24.1	11.5
8	60	20		20		58.7	60.1	11.1	12.0
9	60	10		30		56.2	65.3	-22.2	19.7
10	60		20	20		42.2	13.6	-16.0	16.3
11	60		30	10		54.5	-9.7	-25.5	14.9
12	60		10	30		49.3	31.5	11.9	9.8
13	70		15	15		41.3	20.7	-20.6	12.3
14	70		20	10		41.6	11.6	-12.2	22.2
15	50	30	20			45.0	1.9	-6.2	15.9
16	50	20	30			48.4	0.1	-28.2	12.9
17	50	10	40			51.9	5.2	-44.3	10.4
18	50	40	10			69.6	5.9	43.2	12.9
19	50	10			40	51.6	56.3	-42.8	11.2
20	60	20			20	59.6	60.5	4.4	17.8
21	60			30	10	53.6	59.4	-34.6	13.3
22	50			20	30	51.7	53.3	-46.6	9.3
23	50			30	20	52.6	56.7	-37.5	15.5
24	50			40	10	53.6	59.2	-30.4	14.6

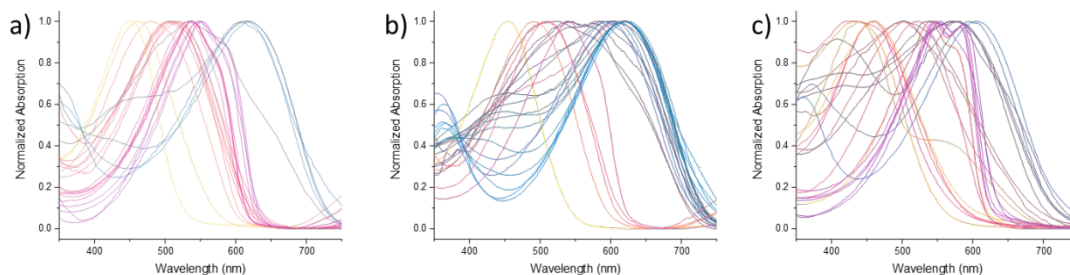


Figure S8. Normalized UV-vis Spectra of the high-throughput synthesized ECPs for model refinement: (a)-(c) represent results in batch 1-3 respectively; all spectra were normalized based on absorption intensity from 0 to 1, with the polymer colors represented by the color of the spectral lines.

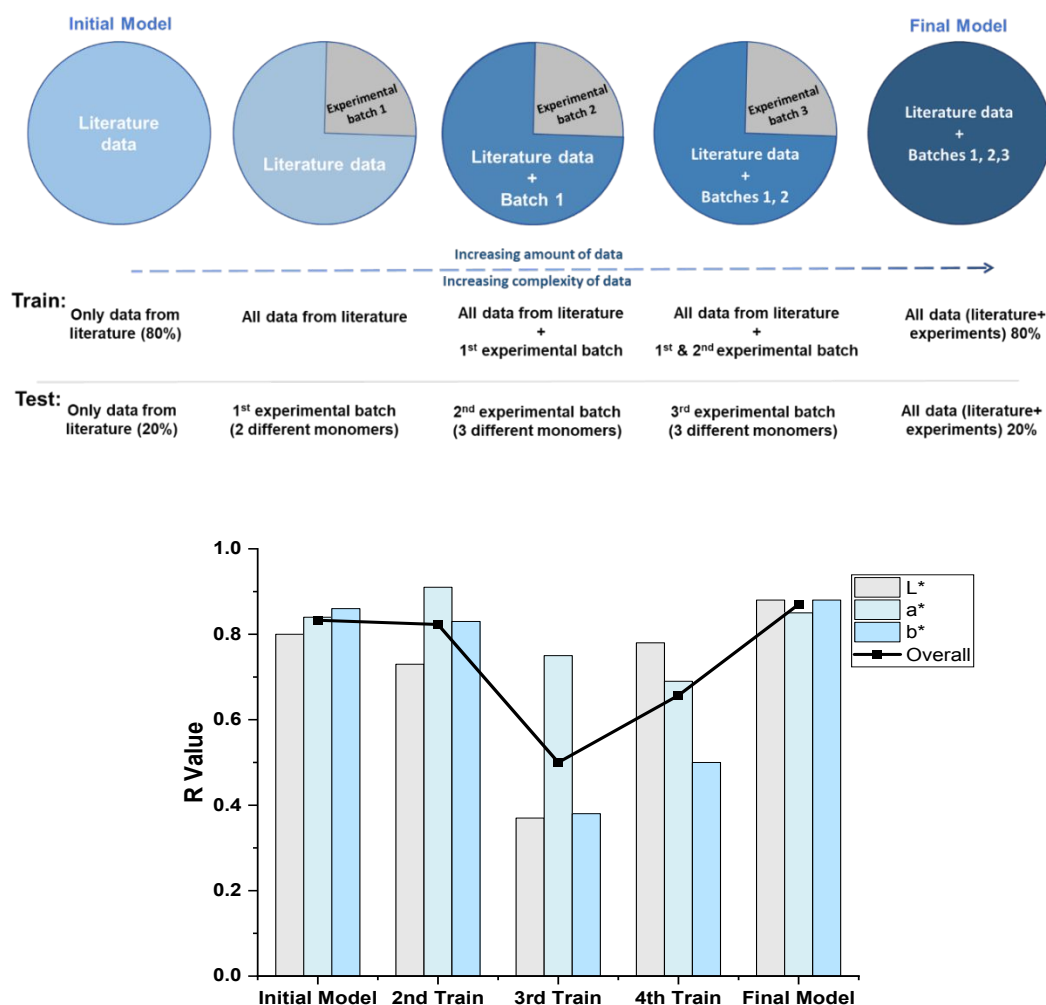


Figure S9. Performance comparison of correlation coefficient (bottom) after updating the training set with the three experimental batches (top). We used correlation coefficients to assess the degree

of agreement between predicted and experimental values in terms of relative trends. While absolute deviations are already captured by the MSE-based loss function, correlation provides insight into whether the model is able to reproduce the correct ordering and monotonic relationships across the dataset. This complementary metric therefore emphasizes trend fidelity rather than absolute magnitude.

6. Electrochromic characterization

Table S7. Randomly selected ECPs in different color themes. The following polymers were randomly selected based on the representative colors and synthesized according to the methods described in Section 5. After precipitation, the polymer products were purified by additional Soxhlet extraction of acetone and hexane and collected in chloroform.

<i>Polymer #</i>	<i>(L*, a*, b*)</i>	ProDOT	Ben-2Br	BTD-2Br	DMoT-2Br
1	(65, 59, 17)	80	20		
2	(93, -8, 81)	50	50		
3	(83, 21, 70)	60	40		
4	(50, -9, -19)	75		25	
5	(58, 67, -16)	100			
6	(59, 60, 11)	60	20		20

6.1 CV and spectro-electrochemical measurements

Electrochemical experiments in solution phase were conducted using a three-electrode system with tetrabutylammonium hexafluorophosphate (TBAPF₆, 0.2 M) in DCM as the supporting electrolyte. The tested polymers were dissolved in DCM at 5 mM. The electrochemical cell comprised a counter electrode of platinum wire, a working electrode of platinum button, and a leakless Ag/AgCl reference electrode. CV was performed at 40 mV/s using Biologic SP150 Potentiostat. Spectro-electrochemical characterizations were performed in-situ in a cuvette with a platinum honeycomb electrode by Pine Research, which incorporated a built-in counter electrode. An identical reference electrode was utilized. Absorption spectra were measured with 0.05 V increment of applied potential.

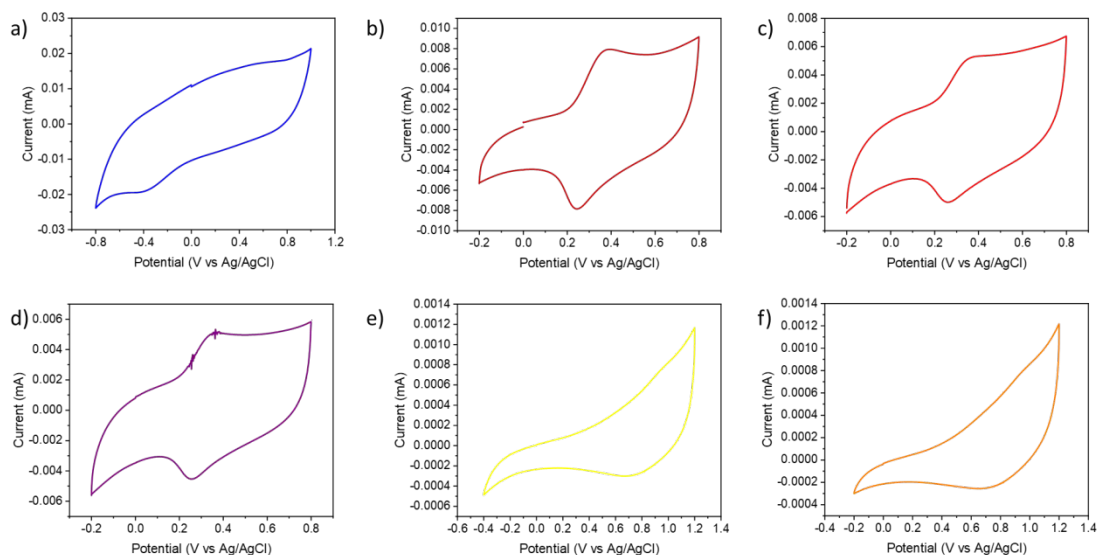


Figure S10. CV Spectra of the selected of ECPs solutions (Table S7): (a)-(f) represent results of polymer 1-6 respectively.

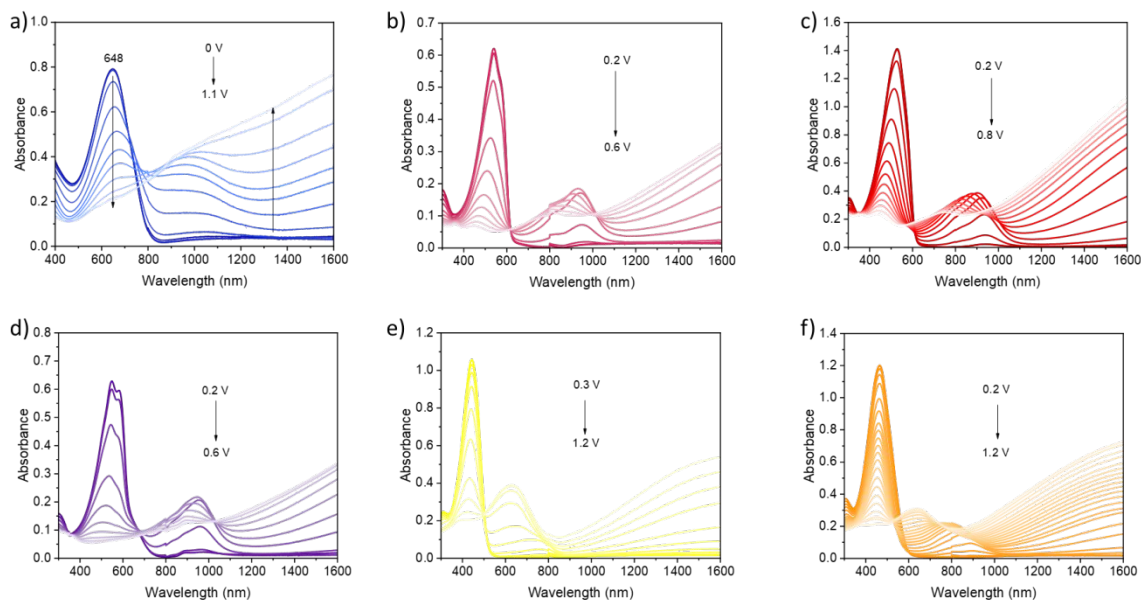


Figure S11. Spectro-electrochemical results of the selected of ECPs solutions (Table S7): (a)-(f) represent results of polymer 1-6 respectively.

7. Model Tuning and Optimization

Table S8. Summary of the types and values range of hyperparameters screened with the Set-Transformer-DFT-ECFP architecture. Optimal hyperparameter configurations for models developed using k-fold cross validation are shown on the last column. The hyperparameter optimization was performed using Optuna²⁸.

Hyperparameter type	Hyperparameter Values	Optimized Hyperparameters
Learning rate	1e-3, 1e-2	1e-3
Number of epochs	100, 800, step = 50	650
Batch size	12, 32, 64, 128, 256	12
Dropout ratio	0.1, 0.3, step = 0.05	0.20

Table S9. Selected hyperparameters for the ML models used for comparison. Each monomer was featured using concatenation of the molecular fingerprint and the DFT descriptors. The ANN architecture involves the first layer with 15 neurons with ReLU activation and an output layer with three neurons with Softmax activation.

Model	Optimized Hyperparameters
ANN	learning_rate = 0.001, epochs = 10, batch_size = 32
Random Forest	n_estimators = 500, max_depth = 10, random_state = 0
XGBoost	n_estimators = 800, max_depth = 10, eta = 0.1, subsample = 0.7, colsample_bytree = 0.8, random_state = 0
Lasso	alpha = 0.1
KNN	n_neighbors = 20
GPR	kernel = C(1.0, (1e-3, 1e3)) * RBF(1.0, (1e-2, 1e2))

8. Autonomous production of the ECPs with targeted colors

8.1 Graphical user interface for the initial selections

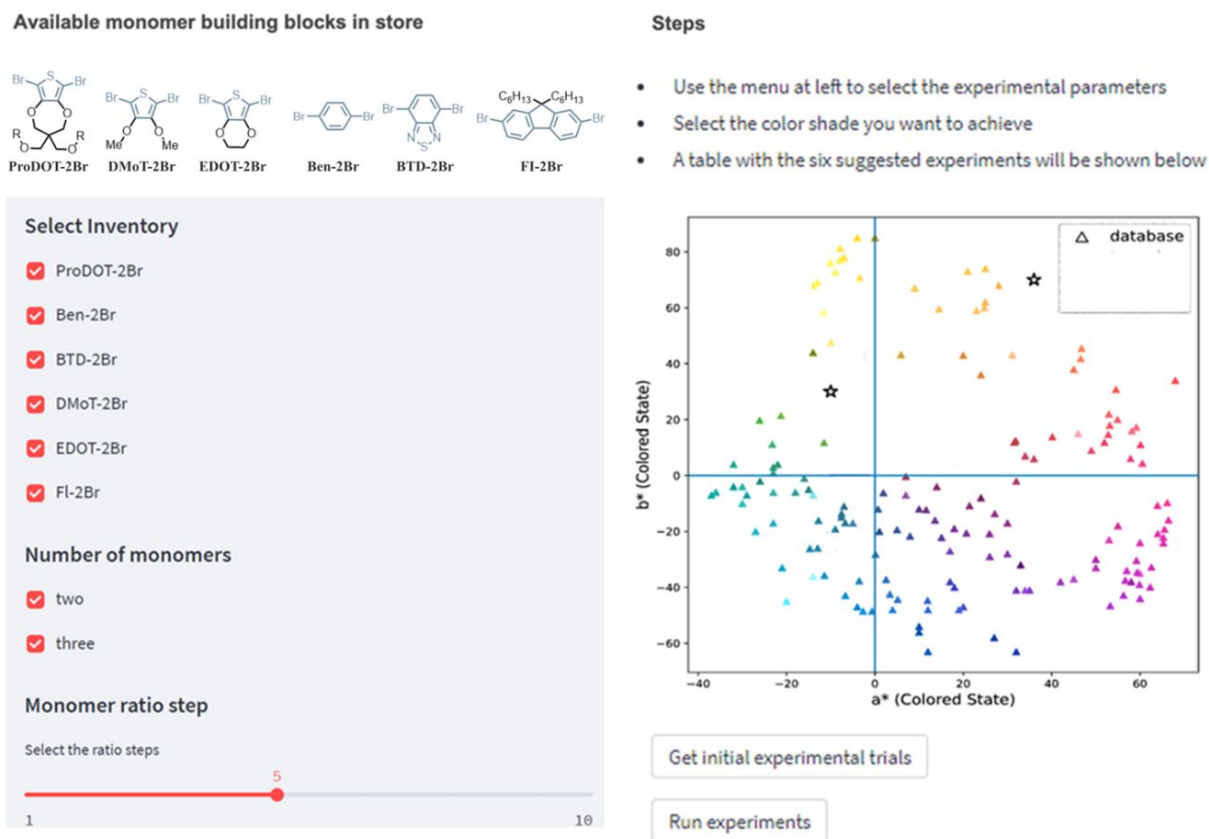


Figure S12. A live monitoring platform using streamlit library was developed to initialize the experiment and monitor it in real time. The user was selecting the monomers that were available in Chemspeed, the desired ratio step, the number of monomers to be considered and the desired color (star symbol). This generated an initial json (“init.json”) file with the information and kept updating it once the experimental measurements were done. Snapshot of the Streamlit app before the start of closed-loop experiments. The displayed data points represent the initial dataset used to initiate the experiment, prior to incorporating results from the green and orange optimization campaigns.

8.2 Lab network and WE12 settings

```
polybot_workcell > workflows > ! demo copy.yaml
1  metadata:
2    name: PCR - Workflow
3    author: Doga Ozgulbas
4    info: Initial PCR workflow for RPL workcell
5    version: 0.1
6
7  workcell: /home/rpl/workspace/polybot_workcell/polybot_workcell/polybot_workcell.yaml
8
9  modules:
10   - name: chemspeed
11   - name: tecan
12   - name: ur5
13
14  flowdef:
15   - name: Mix Reactions
16     module: chemspeed
17     command: open_chemspeed
18     args:
19       foo: "bar"
20     checks: RESOURCE_CHECK
21     comment: resource check not implemented, but will be
22
23   - name: Mix Reactions
24     module: chemspeed
25     command: run_chemspeed
26     args:
27       name_protocol: "closed_loop1"
28     checks: RESOURCE_CHECK
29     comment: resource check not implemented, but will be
30
31   - name: Mix Reactions
32     module: chemspeed
33     command: close_chemspeed
34     args:
35       foo: "bar"
36     checks: RESOURCE_CHECK
37     comment: resource check not implemented, but will be
38
39   - name: Pick up plate from Chemspeed
40     module: ur5
41     command: transfer
42     args:
43       source: platestack
44       target: chemspeed.deck1
45       source_plate_rotation: narrow
46       target_plate_rotation: wide
47     checks: null
48     comment: Pick plate from Chemspeed deck1
```

1. Open the Chemspeed protocol

2. Run the experiment and generate the output files

3. Close the Chemspeed GUI

4. Send a command to UR5e to pick-up the plate for further analysis

Figure S13. A .yaml file describing the sequence of actions taken to complete a closed-loop experiment. We first set the modules that were used in the workflow, i.e., Chemspeed, Tecan and UR5e, and then we defined the sequential flow with the actions of each equipment.

8.3 Step-by-step explanation of the workflow

Create the initial screening dataset with all the available experiments. Before we started the closed loop experimentation, we constructed the full dataset all the possible experiments we can run. The user provided input contains the monomers that were available in the inventory, the number of monomer units, i.e., two or three different components and the ratio step, i.e., the step we wanted to consider the ratio changes of the monomers. Given this information, we created all the two and three monomer combinations. Out of the possible combinations we removed those that already exist in our training dataset to avoid repeating the same experiments. This gave a dataset with 700 possible datapoints, considering the green autonomous discovery as well.

The settings for our closed loop experiment are the following:

```
{
  "monomers": [
    "Cc1sc(C)c2OCC(COCC(CC)CCCC)(COCC(CC)CCCC)COc12",
    "C1=CC(=CC=C1C)C",
    "Cc1ccc(C)c2nsnc12",
    "Cc1sc(C)c(OC)c1OC",
    "Cc1sc(C)c2OCCOc12",
    "Cc1ccc3c(c1)Cc2cc(C)ccc23"
  ],
  "target_Lab": [
    70,
    36.28,
    70.46
  ],
  "number_of_monomers": [
    2,
    3
  ],
  "ratio_step": 5
}
```

Figure S14. The outline of the initialization json file to set-up the experiment.

Stock solution preparation. The stock solutions were prepared with the needs of the monomer selections of the autonomous exploration. Following reactants were dissolved in the chloroform solution at 0.05mmol/mL for 20-30mL: ProDOT, dibromo-ProDOT, Ben-2Br, BTD-2Br, DMoT-2Br, EDOT-2Br, 2,7-dibromo-9,9-dihexyl-9H-fluorene (Fl-2Br, used only for green autonomous discovery), pivalic acid. Pd(OAc)₂ was dissolved in DMAc at 0.005 mmol/mL.

Robotic workflow. After the preparation of stock solutions, the solution vials were transferred to the predetermined zones within the robotic reaction hood and purged with nitrogen gas. Once the workflow was initiated, and the reaction parameters were imported, stock solutions containing monomer reactants were dispensed into the designated reaction vials within the reaction chamber using a liquid handler (Table S9-S10, the ratio of dibromo-ProDOT was calculated by $\text{dibromo-ProDOT\%} = \text{ProDOT\%} - 50\%$, total 2 mL stock solution of monomers (1mL of ProDOT, and 1mL of all used dibromo- monomers) and 200 μL of pivalic acid). Upon the successful transfer of stock solutions, the reaction chamber was transitioned to the "close under vacuum" position, facilitating the removal of chloroform under vacuum conditions. Subsequently, 0.2 mmol of K_2CO_3 solid was added by gravity dispensing unit to the reaction vials. After solid transfer, 1 mL of $\text{Pd}(\text{OAc})_2/\text{DMAc}$ solution were added to the reaction vials by syringe pump. Following these additions, the vials underwent nitrogen purging and were then heated to 110 $^\circ\text{C}$ for 15 hours with refluxing to promote the reaction. Upon completion of the reactions, 2 mL of chloroform and 15 mL of methanol were successively transferred to each reaction vial. Following the precipitation of the polymer, the reaction mixtures underwent filtration using a vacuum needle and filter. After liquid removal, the resulting solid polymer products were redissolved in 5 mL of chloroform. Subsequently, 50 μL of the polymer solution for each sample was carefully transferred to pre-assigned positions in a 96-well plate following by adding 350 μL of chloroform to the same position. Once all samples from the entire batch were transferred to the 96-well plate, the plate was autonomously conveyed by the UR5e robotic arm to the UV-vis spectrometer for absorption measurements. After the measurement, the UR5e arm will transfer the 96-well plate back the original position.

Measure the $L^*a^*b^*$ values similarity. To compare the similarity of the user provided $L^*a^*b^*$ value and the predicted/experimental $L^*a^*b^*$ values, we used ΔE_{Lab} values to calculate which of the predicted values are closer to this desired value. We used the CIEDE2000 Colour difference formula to calculate the ΔE_{Lab} values²⁹.

Selecting the next experiments to run. The selection of the next six experiments was taking into consideration the ΔE_{Lab} value of the predicted $L^*a^*b^*$ to the desired $L^*a^*b^*$ and the uncertainty of the model for that prediction.

We employed expected improvement as the acquisition function for the selection of the next six experiments to run.

Expected improvement with the focus on achieving our goal:

$$EI(x) = (\mu(x) - y_{best} - \xi) \Phi(Z) + \sigma(x) \varphi(Z)$$

where

$\Phi(Z)$: the cumulative distribution function (CDF) of the standard normal distribution

$\varphi(Z)$: is the probability density function (PDF) of the standard normal distribution

$\mu(x)$: predicted mean (mean predicted value of the property)

$\sigma(x)$: standard deviation (uncertainty)

y_{best} : the expected best value

ξ : the parameter that controls exploration vs exploitation. A larger ξ encourages exploration.

During the initial iterations we were focusing more on exploration, as such the ξ value was set to 10. During the last iterations as our goal was exploitation, the ξ value was set to 1.

Selecting the next experiments to run. After measuring the absorption spectra of the six samples, we first evaluated the quality of the samples based on the measured maximum absorption peak. We applied the following criteria to decide whether the sample needed to be diluted, became more concentrated or was fine as it was.

Results validation and resubmit experiments. After the UV-Vis measurements, the absorption spectra would go through an effectiveness validation. If the spectra met the quality requirement, the robots would wait the central computer for the command of the next batch. If the concentration of the sample was too low (<0.2) or too high (>2.5), the synthesis robot would concentrate or dilute the original solution 5 times and rerun the characterization function to measure the corrected values.

Applying the stopping criteria. After the absorption measurements were collected and analysed at the end of each iteration, we check if the stopping criteria are met. We used the ΔE_{Lab} comparison for that measurement and check if the calculated ΔE_{Lab} value shows no further improvement in comparison to the previous batches. If this criterion was met, we exited the loop and stopped the experiments as target was achieved.

8.4 Results of the autonomous discovery

Three loops of autonomous explorations were carried out for both orange and green colors. For the orange color, all formulars in the 3rd loop have been explored in the 1st and 2nd loops.

Orange color results.

Table S10. Reaction parameters for the autonomous exploration of Orange ECPs.

Reaction Index	ProDOT (%)	Ben-2Br (%)	BTB-2Br (%)	DMoT-2Br (%)	EDOT-2Br (%)	L*	a*	b*	Mn (kDa)
Orange 1-1	65	35				81.2	25.2	67.5	9.6
Orange 1-2	55	35			10	73.8	41.1	52.5	9.1
Orange 1-3	60	35			5	74.4	39.8	54.0	8.4
Orange 1-4	50	35			15	71.9	43.4	48.8	9.8
Orange 1-5	55	40			5	81.2	25.4	68.2	10.0
Orange 1-6	55	45				88.9	8.0	78.1	7.7
Orange 2-1	55	30			15	77.4	36.5	60.5	9.7
Orange 2-2	55	40		5		75.9	34.4	59.5	14.6
Orange 2-3	55	30		15		78.6	30.2	62.7	10.9
Orange 2-4	50	40		10		81.8	22.8	67.3	12.9
Orange 2-5	55	40	5			77.3	8.1	58.5	11.1
Orange 2-6	50	45			5	86.8	11.9	75.2	11.9
Orange 3-1	55	30			15	77.4	36.5	60.5	9.7
Orange 3-2	55	40		5		75.9	34.4	59.5	14.6
Orange 3-3	55	30		15		78.6	30.2	62.7	10.9
Orange 3-4	50	40		10		81.8	22.8	67.3	12.9
Orange 3-5	65	35				81.2	25.2	67.5	9.6
Orange 3-6	55	40			5	81.2	25.4	68.2	10.0

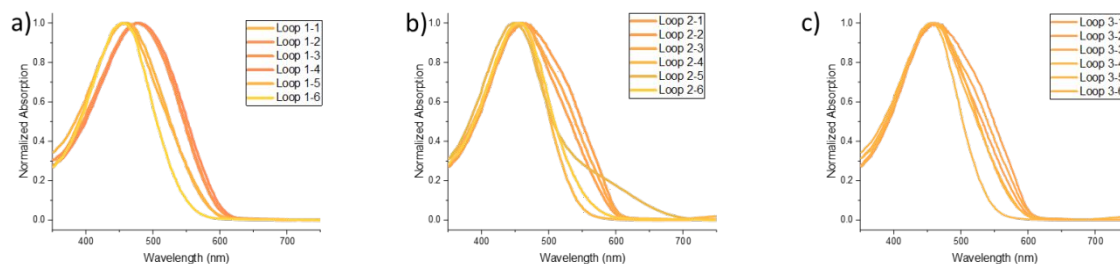


Figure S15. Normalized UV-Vis spectra of the orange autonomous exploration samples: (a)-(c) represent the results in loop-1 to loop-3 respectively, the polymer colors are represented by the color of the spectra lines.

Green color results.

Table S11. Reaction parameters for the autonomous exploration of Green ECPs.

Reaction Index	ProDOT (%)	Ben-2Br	BTD-2Br (%)	DMoT-2Br (%)	EDOT-2Br (%)	Fl-2Br	L*	a*	b*	Mn (kDa)
Green 1-1	70		5			25	46.9	15.4	13.9	11.6
Green 1-2	55		5			40	42.5	16.4	3.3	13.6
Green 1-3	65		5			30	60.7	-4.2	7.4	11.4
Green 1-4	60		5			35	42.4	-6.4	1.1	11.5
Green 1-5	65		10			25	52.8	-3.6	13.6	9.2
Green 1-6	50		25			25	62.9	-4.5	21.6	11.9
Green 2-1	55		10			35	86.4	-4.0	42.7	11.6
Green 2-2	50		15			35	86.5	-7.4	45.1	14.7
Green 2-3	50		10			40	95.9	-15.9	66.5	12.1
Green 2-4	60		10			30	67.6	4.9	36.9	14.0
Green 2-5	60		15			25	58.1	-17.4	14.5	11.0
Green 2-6	55		15			30	66	-8.4	23.0	18.8
Green 3-1	55		10			35	86.4	-4.0	42.7	11.6
Green 3-2	55		25			20	50.9	-10.7	-24.0	9.0
Green 3-3	55		20			25	51.6	-7.5	3.1	12.9
Green 3-4	60		15			25	58.1	-17.4	14.5	11.0
Green 3-5	60		20			20	53.2	-7.5	10.9	10.9
Green 3-6	55		15			30	66.0	-8.4	23.0	18.8

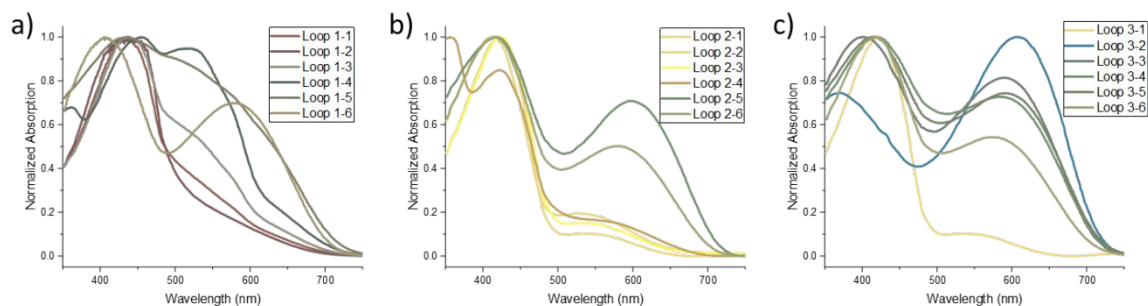


Figure S16. Normalized UV-Vis spectra of the green autonomous exploration samples: (a)-(c) represent the results in loop-1 to loop-3 respectively, the polymer colors are represented by the color of the spectra lines.

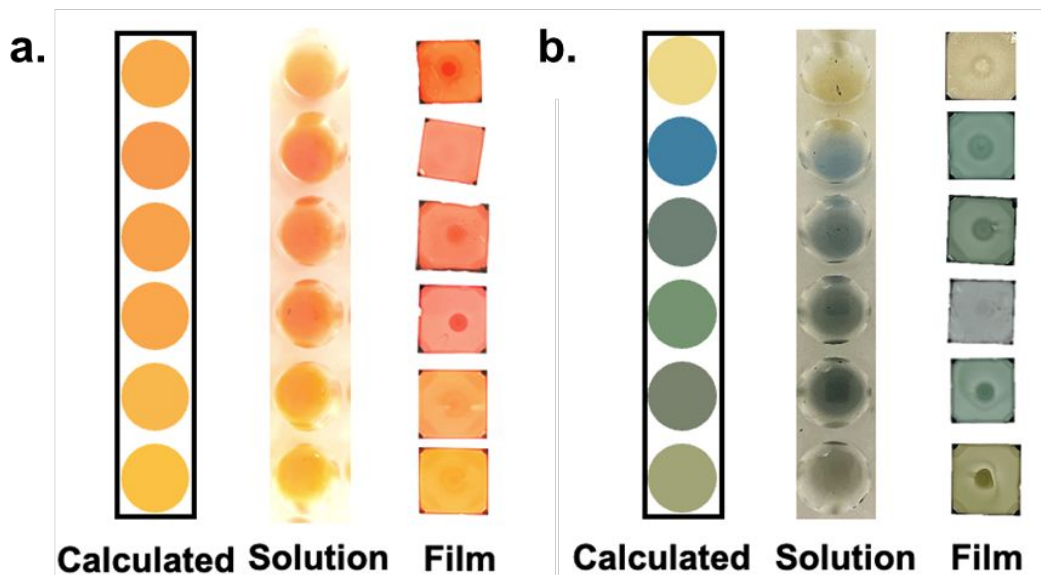


Figure S17. Pictures of final close-loop ECP samples with the color calculated from solution state absorption spectra, real solution and thin film pictures. **a.** results of orange target ($L^* = 70$, $a^* = 35$, $b^* = 50$); **b.** results of green target ($L^* = 65$, $a^* = -10$, $b^* = 30$). All samples were spin-coated at 1000 rpm from 10 mg/mL solutions in chloroform on glass substrate. All the polymers were directly used after precipitation.

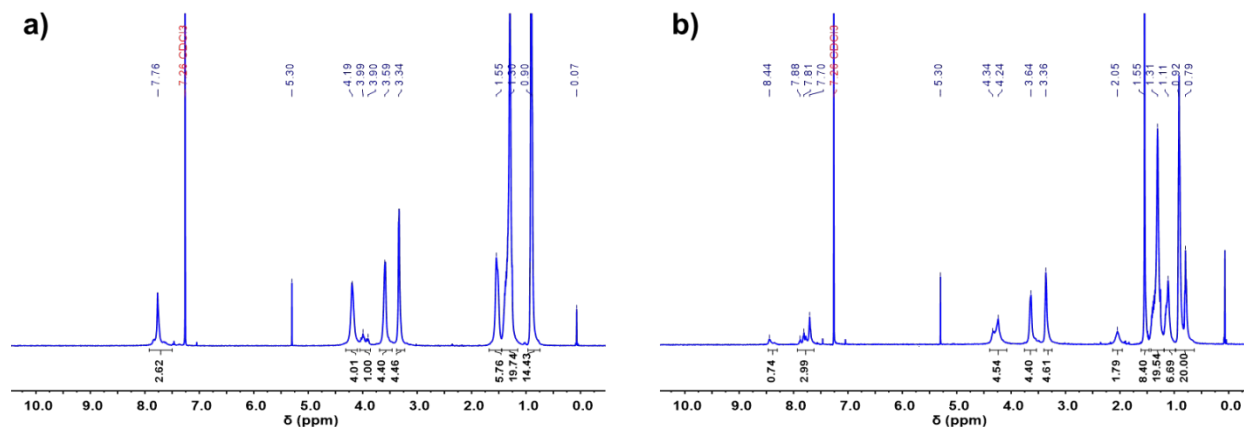


Figure S18. Structural characterization of the final target samples in close-loop optimization. (a) NMR characterization result of orange target: ^1H -NMR (CDCl_3 , 500M, ppm): $\delta = 7.84\text{--}7.67$ (3.2H, m), $4.28\text{--}4.07$ (4.4H, m), $4.06\text{--}3.86$ (0.6H, m), $3.67\text{--}3.44$ (4.4H, m), $3.38\text{--}3.25$ (4.4H, m), $1.64\text{--}1.20$ (19.8H, m), $0.97\text{--}0.8$ (13.2H, m); (b) NMR characterization result of green target: ^1H -NMR (CDCl_3 , 500M, ppm): $\delta = 8.49\text{--}7.59$ (5.4H, m), $4.41\text{--}4.13$ (4H, m), $3.74\text{--}3.54$ (4H, m), $3.44\text{--}3.24$ (4H, m), $2.17\text{--}1.96$ (2.8H, m), $1.61\text{--}0.68$ (45.4H, m)

9. References

- (1) Beaujuge, P. M.; Ellinger, S.; Reynolds, J. R. Spray Processable Green to Highly Transmissive Electrochromics via Chemically Polymerizable Donor-Acceptor Heterocyclic Pentamers. *Advanced Materials* **2008**, *20* (14), 2772-2776, PubMed.
- (2) Beaujuge, P. M.; Ellinger, S.; Reynolds, J. R. The Donor-Acceptor Approach Allows a Black-to-Transmissive Switching Polymeric Electrochrome. *Nat. Mater.* **2008**, *7* (10), 795-799.
- (3) Amb, C. M.; Beaujuge, P. M.; Reynolds, J. R. Spray-Processable Blue-to-Highly Transmissive Switching Polymer Electrochromes via the Donor-Acceptor Approach. *Adv. Mater.* **2010**, *22* (6), 724-728.
- (4) Christiansen, D. T.; Reynolds, J. R. A Fruitful Usage of a Dialkylthiophene Comonomer for Redox Stable Wide-Gap Cathodically Coloring Electrochromic Polymers. *Macromolecules* **2018**, *51* (22), 9250-9258.
- (5) Shi, P.; Amb, C. M.; Knott, E. P.; Thompson, E. J.; Liu, D. Y.; Mei, J.; Dyer, A. L.; Reynolds, J. R. Broadly Absorbing Black to Transmissive Switching Electrochromic Polymers. *Adv. Mater.* **2010**, *22* (44), 4949-4953.
- (6) Amb, C. M.; Kerszulis, J. A.; Thompson, E. J.; Dyer, A. L.; Reynolds, J. R. Propylenedioxythiophene (ProDOT)-Phenylene Copolymers Allow a Yellow-to-Transmissive Electrochrome. *Polym. Chem.* **2011**, *2* (4), 812-814, 10.1039/C0PY00405G.
- (7) Beaujuge, P. M.; Vasilyeva, S. V.; Liu, D. Y.; Ellinger, S.; McCarley, T. D.; Reynolds, J. R. Structure-Performance Correlations in Spray-Processable Green Dioxothiophene-Benzothiadiazole Donor-Acceptor Polymer Electrochromes. *Chem. Mater.* **2012**, *24* (2), 255-268.
- (8) Kerszulis, J. A.; Amb, C. M.; Dyer, A. L.; Reynolds, J. R. Follow the Yellow Brick Road: Structural Optimization of Vibrant Yellow-to-Transmissive Electrochromic Conjugated Polymers. *Macromolecules* **2014**, *47* (16), 5462-5469.
- (9) Kerszulis, J. A.; Johnson, K. E.; Kuepfert, M.; Khoshabo, D.; Dyer, A. L.; Reynolds, J. R. Tuning the Painter's Palette: Subtle Steric Effects on Spectra and Colour in Conjugated Electrochromic Polymers. *J. Mater. Chem. C* **2015**, *3* (13), 3211-3218, 10.1039/C4TC02685C.
- (10) Ponder, J. F., Jr.; Pittelli, S. L.; Reynolds, J. R. Heteroatom Role in Polymeric Dioxyselenophene/Dioxythiophene Systems for Color and Redox Control. *ACS Macro Lett.* **2016**, *5* (6), 714-717.
- (11) Kerszulis, J. A.; Bulloch, R. H.; Teran, N. B.; Wolfe, R. M. W.; Reynolds, J. R. Relax: A Sterically Relaxed Donor-Acceptor Approach for Color Tuning in Broadly Absorbing, High Contrast Electrochromic Polymers. *Macromolecules* **2016**, *49* (17), 6350-6359.
- (12) Ponder, J. F., Jr.; Österholm, A. M.; Reynolds, J. R. Designing a Soluble PEDOT Analogue without Surfactants or Dispersants. *Macromolecules* **2016**, *49* (6), 2106-2111.
- (13) Savagian, L. R.; Österholm, A. M.; Shen, D. E.; Christiansen, D. T.; Kuepfert, M.; Reynolds, J. R. Conjugated Polymer Blends for High Contrast Black-to-Transmissive Electrochromism. *Adv. Opt. Mater.* **2018**, *6* (19), 1800594.
- (14) Christiansen, D. T.; Ohtani, S.; Chujo, Y.; Tomlinson, A. L.; Reynolds, J. R. All Donor Electrochromic Polymers Tunable Across the Visible Spectrum via Random Copolymerization. *Chem. Mater.* **2019**, *31* (17), 6841-6849.
- (15) Ochieng, M. A.; Ponder, J. F.; Reynolds, J. R. Effects of Linear and Branched Side Chains on the Redox and Optoelectronic Properties of 3,4-dialkoxythiophene Polymers. *Polym. Chem.* **2020**, *11* (12), 2173-2181, 10.1039/C9PY01720H.
- (16) Collier, G. S.; Wilkins, R.; Tomlinson, A. L.; Reynolds, J. R. Exploring Isomeric Effects on Optical and Electrochemical Properties of Red/Orange Electrochromic Polymers. *Macromolecules* **2021**, *54* (4), 1677-1692.
- (17) You, L.; He, J.; Mei, J. Tunable Green Electrochromic Polymers via Direct Arylation Polymerization. *Polym. Chem.* **2018**, *9* (43), 5262-5267, 10.1039/C8PY01105B.
- (18) Mansencal, T.; Mauderer, M.; Parsons, M.; Shaw, N.; Wheatley, K.; Cooper, S.; Vandenberg, J. D.; Canavan, L.; Crowson, K.; Lev, O.; et al. Colour 0.4.4. 0.4.4 ed.; Zendo: 2023.
- (19) Cao, K.; Shen, D. E.; Österholm, A. M.; Kerszulis, J. A.; Reynolds, J. R. Tuning Color, Contrast, and Redox Stability in High Gap Cathodically Coloring Electrochromic Polymers. *Macromolecules* **2016**, *49* (22), 8498-8507.
- (20) Dyer, A. L.; Craig, M. R.; Babiarz, J. E.; Kiyak, K.; Reynolds, J. R. Orange and Red to Transmissive Electrochromic Polymers Based on Electron-Rich Dioxythiophenes. *Macromolecules* **2010**, *43* (10), 4460-4467.
- (21) Fujinami, Y.; Kuwabara, J.; Lu, W.; Hayashi, H.; Kanbara, T. Synthesis of Thiophene- and Bithiophene-Based Alternating Copolymers via Pd-Catalyzed Direct C-H Arylation. *ACS Macro Lett.* **2012**, *1* (1), 67-70.
- (22) Luo, X.; Tran, D. T.; Kadlubowski, N. M.; Ho, C. H. Y.; Riley, P.; So, F.; Mei, J. Side-Chain Sequence Enabled Regioisomeric Acceptors for Conjugated Polymers. *Macromolecules* **2018**, *51* (21), 8486-8492.
- (23) Manabe, K.; Ohba, M.; Matsushima, Y. A Repetitive One-Step Method for Oligoarene Synthesis Using Catalyst-Controlled Chemoselective Cross-Coupling. *Org. Lett.* **2011**, *13* (9), 2436-2439.

- (24) Vogt, A.; Stümpges, F.; Bajrami, J.; Baumgarten, D.; Millan, J.; Mena-Osteritz, E.; Bäuerle, P. Tunable Regioselectivity in C–H-Activated Direct Arylation Reactions of Dithieno[3,2-b:2',3'-d]pyrroles. *Chem. – Eur. J.* **2023**, *29* (60), e202301867.
- (25) Zhao, H.; Liu, C.-Y.; Luo, S.-C.; Zhu, B.; Wang, T.-H.; Hsu, H.-F.; Yu, H.-h. Facile Syntheses of Dioxythiophene-Based Conjugated Polymers by Direct C–H Arylation. *Macromolecules* **2012**, *45* (19), 7783-7790.
- (26) Lee, J.; Lee, Y.; Kim, J.; Kosiorek, A.; Choi, S.; Teh, Y. W. Set Transformer: A Framework for Attention-Based Permutation-Invariant Neural Networks. In International conference on machine learning, 2019.
- (27) Żurański, A. M.; Wang, J. Y.; Shields, B. J.; Doyle, A. G. Auto-QChem: An Automated Workflow for the Generation and Storage of DFT Calculations for Organic Molecules. *React. Chem. Eng.* **2022**, *7* (6), 1276-1284, 10.1039/D2RE00030J.
- (28) Akiba, T.; Sano, S.; Yanase, T.; Ohta, T.; Koyama, M. Optuna: A Next-Generation Hyperparameter Optimization Framework. In Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, Anchorage, AK, USA; 2019.
- (29) Sharma, G.; Wu, W.; Dalal, E. N. The CIEDE2000 Color-Difference Formula: Implementation Notes, Supplementary Test Data, and Mathematical Observations. *Color Res. Appl.* **2005**, *30* (1), 21-30.