Digital Discovery

ROYAL SOCIETY OF CHEMISTRY

PAPER

View Article Online
View Journal | View Issue

Cite this: *Digital Discovery*, 2025, **4**, 3774

Design of simple-structured conjugated polymers for organic solar cells by machine learning-assisted structural modification and experimental validation

Shogo Tadokoro,[a] Ryosuke Kamimura,[a] Fumitaka Ishiwari[abc] and Akinori Saeki *[ab]

Improving the performance of organic photovoltaics (OPVs) depends on the development of new p-type polymers and n-type non-fullerene acceptor (NFA) molecules. However, conventional experimental and theoretical methods are inefficient for exploring the vast chemical space. In this report, we use machine learning (ML) to explore simple-structured p-type polymers. The structural simplicity is associated with a small synthesis step relevant for low-cost, large-scale production. By considering the structural simplicity (primitively based on the molecular weight of its repeating unit) of the 200 thousand virtually generated polymers, together with synthetic accessibility, we focus on copolymers composed of benzoxadiazole as an acceptor and thiophene (or phenylene) as a donor. Although the structures of these copolymers resemble a high-performance simple-structured PTQ10, their structural symmetries (regioregularity) are modified for synthetic reasons. Through the characterization of the synthesized polymers, their OPV devices blended with Y6 NFA, and resultant synthetic complexity scores, we show that our polymer with a minor manual modification of the donor and alkyl chain exhibits a power conversion efficiency of 5.56%, which closely aligns with that predicted by ML and provides a basis for the further development of novel polymers with low synthesis and search costs.

Received 18th September 2025
Accepted 25th October 2025

DOI: 10.1039/d5dd00418g

rsc.li/digitaldiscovery

Introduction

Organic photovoltaics (OPVs) offer the advantages of light weight, flexibility, and low toxicity, along with their solution processing, at potentially low cost. The mixture of p-type and n-type organic semiconductors forms a bulk heterojunction (BHJ) and is used as an active layer, which is responsible for light absorption, charge separation, and charge transport.[1,2] The p-type materials are usually conjugated polymers, while the n-type materials are non-fullerene acceptor (NFA) molecules.[3–5] With the advent of NFAs, the power conversion efficiency (PCE) of OPV has dramatically increased.[6–8] Similar to NFAs, p-type polymers are designed by combining electron-donating (D) and electron-accepting (A) units mostly in an alternating fashion.[9–11] Representative polymers, such as PM6,[12] and D18,[13] have achieved high PCE with Y6-type NFAs. However, these p-type polymers have complex chemical structures and require an increased number of synthesis steps, which is irrelevant for a low-cost mass production. In contrast, PTQ10,[14] which has a simple chemical structure, can be synthesized in only two steps, thus enabling reduced synthesis costs. Along with its high PCE value, PTQ10 blended with the Y6 analogue is one of the plausible candidates for large-scale production.[15–18] Accordingly, the development of conjugated polymers that simultaneously satisfy simplicity (low-cost synthesis) and high PCE is desired, while a conventional approach is difficult to cover the vast chemical space of polymers due to their complexity and weak structure–property relationship.

As such, machine learning (ML) is increasingly expected to be a viable approach for the development of high-throughput materials.[19–24] In particular, energy-related materials such as solar cells are common targets as they involve a complex interplay of the material properties and device performance.[25–28] We previously reported an ML model for predicting the PCE of fullerene[29] and NFA[30–32] OPVs based on the experimental data taken from the literature. Furthermore, by incorporating failure data,[33] we demonstrated ML-driven polymer design and experimental validation of the virtually generated 200 932 polymers in combination with ITIC (or IT4F)[31–33] and Y6.[34] Many of these polymers contain halogenated benzodithiophene (BDT) flanked with alkylthiophenes as the D unit,[35,36] which is a promising building block but detrimental to a cost-effective synthesis due to its relatively more synthetic steps (distannyl BDT requires 2–5 steps to be synthesized from benzo[1,2-*b*:4,5-*b*′]dithiophene-4,8-dione).[37,38]

[a]*Department of Applied Chemistry, Graduate School of Engineering, Osaka University, 2-1 Yamadaoka, Suita, Osaka 565-0871, Japan. E-mail: saeki@chem.eng.osaka-u.ac.jp*
[b]*Innovative Catalysis Science Division, Institute for Open and Transdisciplinary Research Initiatives (ICS-OTRI), Osaka University, 1-1 Yamadaoka, Suita, Osaka 565-0871, Japan*
[c]*PRESTO, Japan Science and Technology Agency (JST), Kawaguchi, Saitama 332-0012, Japan*

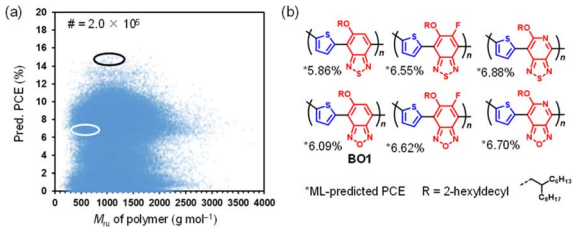
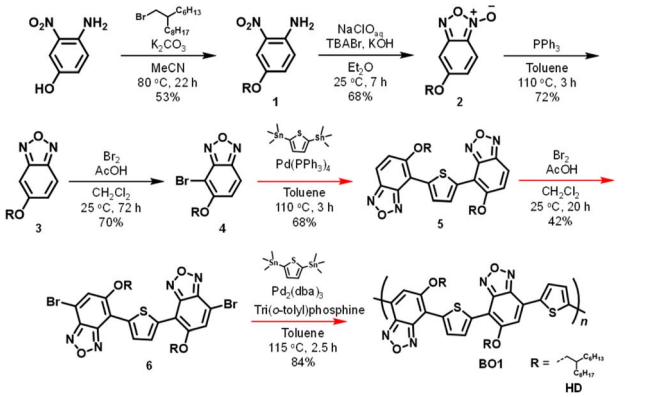


Fig. 1 (a) Plot of ML-predicted PCE and molecular weight of the repeating unit (M_{ru}) of the polymer. The number of these virtually generated polymers is ~200 thousand. The area of the black ellipse is the high PCE region and that of the white ellipse is the target in this study. (b) Selected candidates of simple-structured polymers. The appended values are ML-predicted PCE. BO1 (the bottom left) was chosen as the target in this study.



Scheme 1 Synthetic route of BO1.

In this study, we attempted to develop p-type polymers with a simple structure from our dataset of ~200 thousand virtual polymers, along with their ML-predicted PCE (Fig. 1a). In brief, this ML model uses a random forest algorithm and a dataset of 2169, of which 1295 are experimental data taken from the literature (provided as a text file[30]) and 874 are virtually generated failure data to improve the ML model.[33] Five-fold cross-validation of the model yielded a correlation coefficient of 0.84. Approximately 200 thousand virtual polymers were generated by combining the D (#=382) and A (#=566) units fragmented from reported polymer structures.[32] Here, we primitively defined a simple chemical structure as "a small molecular weight of the repeating unit (M_{ru})". As shown in Fig. 1b, we focused on the M_{ru} range that is about half (the white circle) of the typical high-performing polymers (the black circle) and selected six candidates by stepwise manual screening Fig. 1b. In all compounds, the D unit is thiophene, and the A unit is benzoxadiazole (BO) or benzothiadiazole (BT) (the synthesis details: Fig. S1–S47(SI)). Considering the ease of synthesis, BO1 was selected as the first target for experimental investigation. BO was chosen over BT because it has been reported less frequently than BT. The alkyl side chain of BO1 is 2-hexyldecyl, which is the same as PTQ10 and expected to provide good solubility. In addition to the simple chemical structure, synthetic accessibility (SA)[39] is one of the synthetic cost indicators. The SA assesses the ease of synthesis only from the chemical structure and gives a value from 1 (simple) to 10 (complex). We calculated the SA score for the six candidates and found that BO1 (SA score = 3.28) was the simplest in synthesis among them (Fig. S48(SI)). The SA scores among the 120 thousand polymers range from ~2 to ~9 (Fig. S49(SI)) and are mostly correlated with M_{ru}, and thus, BO1 can be considered as a simple structure with a low SA. The predicted PCE of BO1 : Y6 calculated using our ML model[33] is 6.09%, which is not high, but it is worth investigating to verify our approach.

Results and discussion

Synthesis and OPV characterization of BO1

The synthesis of BO1 from 4-amino-3-nitrophenol was planned in five steps (Scheme S1(SI)). First, a 2-hexyldecyloxy side chain was introduced into 4-amino-3-nitrophenol using Williamson ether synthesis to yield **1**. Next, **3** was synthesized *via* **2** based on the previously reported synthesis method of benzoxadiazole;[40] however, under several conditions for the bromination of **3** to **4** (Fig. S50(SI)),[41–44] **4** was not obtained. We modified the synthesis route (Scheme 1), where two equivalents of **4** and 2,5-bis(trimethylstannyl)thiophene were cross-coupled *via* the Stille reaction to yield **5**. After the dibromination reaction to obtain **6**, the Stille coupling copolymerization of **6** and 2,5-bis(trimethylstannyl)thiophene was applied to yield BO1. By changing the chemical structure of the repeating unit (Fig. 2a), the SA value of BO1 was increased to 4.53. In addition, the molecular weight of BO1 could not be accurately measured by size exclusion chromatography (SEC, $CHCl_3$ at 40 °C) because BO1 showed unexpected strong aggregation in $CHCl_3$. Nevertheless, its highest occupied molecular orbital (HOMO) and the

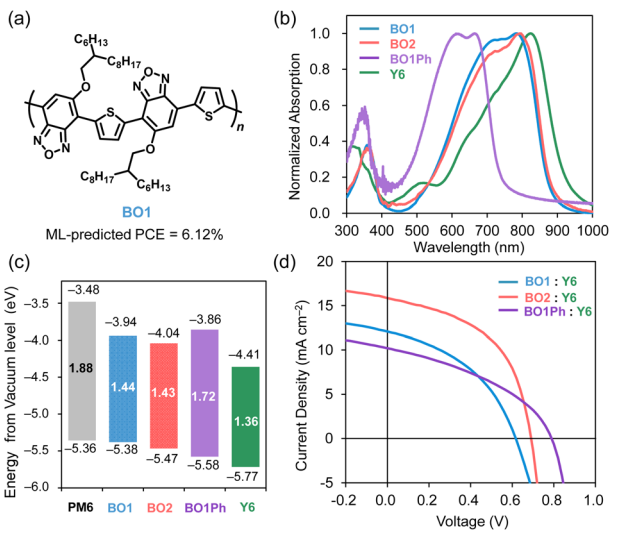


Fig. 2 (a) Structure of the BO1 polymer that was modified from the original structure Fig. 1b because of the synthetic reason (see the main text). (b) UV-vis absorption spectra of BO1, BO2, BO1Ph (*vide infra*) and Y6 films. (c) Energy diagram of PM6, BO1, BO2, BO1Ph, and Y6; data for PM6 and Y6 were taken from the literature.[45] (d) *JV* curves of BO1 : Y6, BO2 : Y6, and BO1Ph : Y6 OPVs.

Table 1 Summary of the OPV device performance[a]

Polymer : NFA (p : n ratio)	Structure	PCE [%]	*Jsc* [mA cm^{-2}]	*Jsc*^{EQE}[b] [mA cm^{-2}]	*Voc* (V)	FF (%)	*L*[c] (nm)
BO1 : Y6 (1 : 0.5)	Inverted[d]	3.04 ± 0.12 (3.12)	11.83 ± 0.15 (12.07)	9.18	0.610 ± 0.013 (0.617)	42.1 ± 0.5 (41.9)	124 ± 12
BO2 : Y6 (1 : 1.5)	Inverted[d]	5.34 ± 0.22 (5.56)	15.68 ± 0.38 (15.85)	14.49	0.687 ± 0.002 (0.690)	49.5 ± 1.0 (50.8)	85 ± 5
BO1Ph : Y6 (1 : 1.5)	Inverted[d]	3.05 ± 0.18 (3.18)	9.96 ± 0.40 (10.21)	9.17	0.784 ± 0.004 (0.789)	39.0 ± 0.7 (39.4)	96 ± 7
BO2 : Y6 (1 : 1.5)	Normal[d]	4.28 ± 0.18 (4.53)	15.12 ± 0.43 (15.65)	15.11	0.682 ± 0.002 (0.685)	41.5 ± 0.5 (42.3)	90 ± 10

[a] The values in parentheses are the average and standard deviation for at least three devices. [b] J_{SC} value calculated by integrating the EQE spectrum. [c] Thickness of the BHJ layer. [d] Inverted: glass/ITO/ZnO/BHJ/MoO3/Ag, normal: glass/ITO/PEDOT:PSS/BHJ/PDINO/Ag.

lowest unoccupied molecular orbital (LUMO) energy levels and the bandgap energy (E_g) were measured using ultraviolet-visible (UV-vis) absorption spectroscopy (Fig. 2c and S51) and photoelectron yield spectroscopy (PYS, Fig. S52(SI)). BO1 and Y6 showed relatively close optical E_g values of 1.44 and 1.36 eV, respectively. Thus, BO1 could not achieve complementary absorption such as in the PM6 and Y6 systems,[45] while the HOMO–HOMO and LUMO–LUMO offsets of BO1 and Y6 were sufficiently large, more than 0.3 eV, required for efficient charge separation.[46,47] The current density (*J*) voltage (*V*) curves of the optimal BO1 : Y6 devices are shown in Fig. 2d (the device parameters are listed in Table 1). The maximum PCE values of BO1 : Y6 were 3.12%, approximately half of the predicted value (6.09%).

Alkyl and donor modification of BO1

Manual modification of the ML-suggested polymer structure based on experimental feedback is a reasonable approach for further progress.[32] We considered two possible reasons for the much lower PCE value of BO1 : Y6 OPV than predicted. The first is the low solubility of BO1, probably caused by the inadvertent addition of regioregularity not present in the original structure. The regioregular (RR) chemical structure generally increases the crystallinity and improves PCE, while the solubility of the polymer may be decreased. Conversely, some reports suggest improved PCE with regiorandom (RRa) chemical structures.[48,49] Although the improvement in crystallinity would improve the charge transport properties of the polymer, it does not necessarily lead to an improved PCE of BHJ OPVs. We therefore designed BO2 with extended side alkyl chains of 2-octyldodecyl to improve the polymer solubility, which showed a slightly decreased predicted PCE of 6.02% (Fig. 3). Despite the mostly unchanged prediction, an empirical parameter[30] (the ratio of the number of aromatic rings to the number of alkyl carbons) of BO2 (6/40 = 0.15) becomes closer to the optimal value (~0.09) than that of BO1 (6/36 = 0.19), suggesting the improved balance between solubility and crystallinity for BO2. The second reason is the light absorption region. As shown in Fig. 2b, BO1 and Y6 have similar light absorption (the former is slightly blue-shifted) and are unable to achieve complementary absorption with Y6, as is the case with efficient p-type polymers such as PM6. We therefore attempted to adjust the wavelength of light absorption by replacing one of the thiophene D units with phenylene in the repeating unit, as previously reported.[50] The resultant polymer : BO1Ph, which has 2-hexyldecyl chains similar to those in BO1, shows an improved predicted PCE of 6.98% Fig. 3. The SA values of BO2 and BO1Ph were 4.66 and 4.44, respectively.

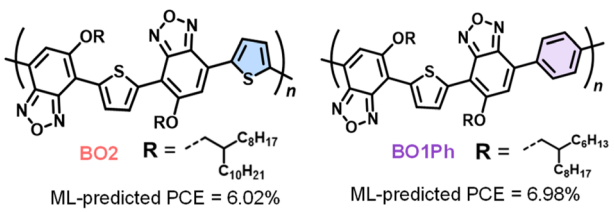


Fig. 3 Manually revised polymer structure of BO2 and BO1Ph. The updated ML-predicted PCE are appended.

BO2 and BO1Ph were successfully synthesized without major changes in the synthetic route of BO1 (Scheme S2(SI)). BO1Ph was obtained by the Suzuki–Miyaura coupling polymerization of **6** and 1,4-benzenediboronic acid bis(pinacol) ester. BO2 was obtained using the same procedure as BO1. As expected, they are well soluble in the eluent of SEC, giving the number- and weight-averaged molecular weights (M_n and M_w, respectively) as 32.7 kDa and 62.1 kDa (M_w/M_n = 1.9) for BO2 and 31.5 kDa and 81.9 kDa (M_w/M_n = 2.6) for BO1Ph, respectively (Table S1(SI)).

OPV characterization of BO2 and BO1Ph

The UV-vis spectra of the three polymers in the film state are provided in Fig. 2b. While the absorption spectra of BO1 and BO2 are very similar, the latter shows enhanced intensity in the longer wavelength peak, indicating a more planar conformation. The BO1Ph spectrum shows blue-shifted peaks and complementary absorption with Y6. The UV-vis spectra of each polymer in chloroform solution, along with those of the films, are shown in Fig. S51(SI). Comparing the absorption spectra between the solution and film states, BO2 and BO1Ph exhibit slightly red-shifted and enhanced 0–0 peaks in their film states, indicating a transition from random to aligned conformations from the solution to the film. However, the difference is very small for BO1, suggesting that BO1 is partially aggregated in solution. As shown in Fig. 2c, we constructed the energy diagram of BO2 and BO1Ph from the E_g values calculated from the UV-vis absorption edges and the HOMO energy levels measured by PYS measurements (Fig. S52(SI)). The E_gs of BO1 and BO2 are nearly the same (1.44 and 1.43 eV, respectively), while BO1Ph has a wider E_g of 1.72 eV. The deeper HOMO level of BO1Ph (−5.58 eV) than BO1 (−5.38 eV) and BO2 (−5.47 eV) is readily explained by the low-lying HOMO of less electron-

donating nature of the benzene ring. Interestingly, the HOMO level of BO2 is 0.09 eV deeper than that of BO1, suggesting that the elongation of the alkyl side chain reduces the crystallinity and deepens the HOMO level, analogous to the previous report on conjugated molecules.[51]

DFT calculations were performed on three repeating units of BO1(or BO2) and BO1Ph (Fig. S53(SI)). As evident from their side views and HOMO/LUMO lobes, BO1(2) has a planar conformation and delocalized HOMO/LUMO, whereas BO1Ph has a curved shape at the phenylene bridge (~30°) and localized HOMO/LUMO. The obtained HOMO energy level of BO1Ph was 0.16 eV deeper than that of BO1(2), and the HOMO–LUMO gap of the former is 0.39 eV narrower than that of BO1(2). The trend in the calculations is in agreement with the experimental results.

The *JV* curves of the best performing OPV devices are shown in Fig. 2d (the parameters are listed in Table 1), where BO2 : Y6 recorded an improved PCE of 5.56% (inverted structure), which is close to the predicted PCE of 6.02% within the mean absolute error (1.59%) and root mean square error (2.12%) of our ML model.[30] In contrast, BO1Ph : Y6 shows a low PCE of 3.18% (ML-predicted PCE = 6.98%), although its photoabsorption spectrum matches that of Y6, and its open-circuit voltage (V_{OC}) is increased to 0.789 V compared to BO2 : Y6 (0.690 V) due to the deeper HOMO level of BO1Ph. As shown in the short-circuit current density (J_{SC}) and external quantum efficiency (EQE) spectra in Fig. 4, the low J_{SC} (10.21 mA cm^{-2}), together with the low fill factor (FF, 39.4%), is responsible for the low PCE of BO1Ph : Y6. We examined a normal structure OPV for BO2 : Y6, but its PCE of 4.53% was lower than that of the inverted structure (Fig. S54(SI)).

Photoluminescence (PL) measurements were performed for both polymer and BHJ (Fig. S55(SI)). Under the predominant excitation of polymers (600 nm for BO1 and BO2 and 530 nm for BO1Ph), BO1 : Y6 showed the lowest quench of 85%, followed by 90% for BO2 : Y6 and 100% for BO1Ph : Y6. In contrast, PL quenching of Y6 using 700 nm excitation showed the highest quenching for BO2 : Y6 (Fig. S56(SI)), which may be linked to its prominent π–π stacking of Y6 observed in 2D-GIXRD (*vide infra*). Flash-photolysis time-resolved microwave conductivity

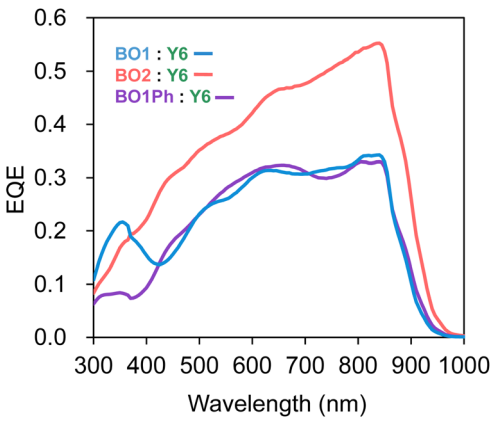


Fig. 4 EQE spectra of BO1 : Y6, BO2 : Y6, and BO1Ph : Y6 OPVs.

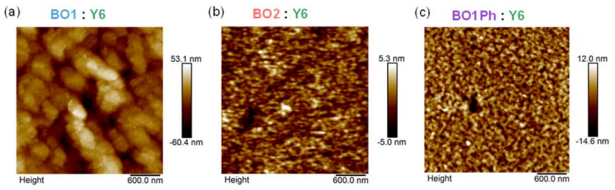


Fig. 5 AFM height images of (a) BO1 : Y6, (b) BO2 : Y6, and (c) BO1Ph : Y6 films. The scale bar is 600 nm (the size is 2 × 2 μm^2). RMS values are 11.9, 1.2, and 1.5 nm for (a)–(c), respectively.

(TRMC)[52,53] measurements revealed that the BHJ of BO1 : Y6 showed the largest signal, while the charge carrier lifetime (τ) was the shortest at 0.79 μs (Table S2 and Fig. S57(SI)). The high TRMC signal despite the low PCE of BO1 : Y6 is explained by its highly crystalline nature, despite the unpreferred orientation (*vide infra*) of BO1 associated with its solubility problem and inefficient PL quenching. The shortest τ probably leads to low charge transport efficiency and low J_{SC} and FF values. The resultant low miscibility between BO1 and Y6 would also cause an inefficient energy cascade at the polymer/NFA interface.[54,55] On the other hand, the BO2 : Y6 and BO1Ph : Y6 showed much smaller TRMC signals (about 1/4 of BO1 : Y6), but much longer τ_S (3.5 and 4.6 μs) than that of BO1 : Y6, which contributes to the improvement of J_{SC} and/or FF.

Fig. 5 shows the height images obtained by atomic force microscopy (AFM) of BHJ surfaces. Aggregated objects were observed in BO1 : Y6, which yielded the largest root mean square roughness (RMS) of 11.9 nm. Such large aggregations are detrimental to efficient exciton diffusion to the p/n interface in BHJ and to charge separation/transport,[56–58] consistent with its low PL quenching and fast decay of the TRMC transient. Domain purity in the p and n phases also affects the fate of geminate recombination.[59] Contrary to BO1 : Y6, BO2 : Y6 and BO1Ph : Y6 show a good mixing and small RMS values of 1.2 and 1.5 nm, respectively. The good miscibility of BO2 (and BO1Ph) with Y6 is partly related to the contact angle (CA) of water and

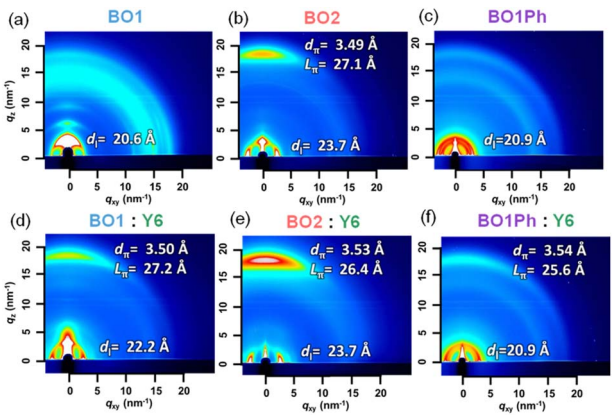


Fig. 6 2D-GIXRD images of (a) BO1, (b) BO2, (c) BO1Ph, (d) BO1 : Y6, (e) BO2 : Y6, and (f) BO1Ph : Y6 films. The calculated π–π distance ($d\pi$), its crystallite coherence length ($L\pi$), and interlamellar distance (d_l) are appended.

Table 2 Characterization of BHJ films and calculated synthesis cost

	Y6-blended BHJ film					Polymer[a]			
Polymer	SCLC μ_h [$cm^2\ V^{-1}\ s^{-1}$]	SCLC μ_e [$cm^2\ V^{-1}\ s^{-1}$]	μ_e/μ_h	PL quench [%]	RMS [nm]	SA	SC	PCE [%]	FOM[b]
BO1	9.1×10^{-5}	8.6×10^{-4}	9.5	~85	11.9	4.53	48.7	3.12	15.6
BO2	2.9×10^{-5}	1.6×10^{-4}	5.5	~90	1.2	4.66	49.5	5.56	8.9
BO1Ph	2.2×10^{-4}	1.4×10^{-5}	0.061	~100	1.5	4.44	52.7	3.18	16.6
PM6	—	—	—	—	—	5.98	73.1[c]	15.7[d]	4.7
PTQ10	—	—	—	—	—	3.44	27.7[c]	16.2[e]	1.7

[a] Calculated synthetic cost of polymer. SA: synthetic accessibility, SC: synthetic complexity. See the main text for their details. [b] FOM = SC/PCE. [c] Taken from ref. 64. [d] Taken from ref. 5. [e] Taken from ref. 65.

glycerol (Fig. S58(SI)) for the pristine polymers, where the CAs of BO2 and BO1 are close to that of Y6,[34] although the difference in CAs is marginal.

The 2-dimensional (2D) grazing-incidence X-ray diffraction (2D-GIXRD) measurements were performed on the pristine polymer film (Fig. 6a–c) and Y6-blended BHJ films (Fig. 6d–f). Their in-plane (IP) and out-of-plane (OOP) profiles are provided in Fig. S59(SI). The pristine BO1 film showed the diffraction of (100), (200), and (300) in the OOP direction (the scattering vector q: 3.05, 6.02, and 8.90 nm^{-1}, respectively), indicating that BO1 is highly crystalline and edge-on oriented (Fig. 6a). The pristine BO2 film showed a strong (010) diffraction in the OOP direction (q = 17.98 nm^{-1}), concomitant with the (100) diffraction in the IP direction (q = 2.65 nm^{-1}), demonstrating its face-on orientation (Fig. 6b). The calculated π–π distance (d_π) and crystallite coherence length (L_π) are 3.49 and 27.1 Å, respectively. These crystalline properties of BO2 are advantageous for vertical charge transport in OPVs.[60–62] In sharp contrast, the pristine BO1Ph film was isotropic and with low crystallinity (Fig. 6c). As shown in Fig. 6d–f, the polymer–Y6 blend films exhibited the superposition of the polymer and Y6 diffraction patterns (the 2D-GIXRD data of the pristine Y6 film is provided in Fig. S60(SI)). Notably, BO2 : Y6 showed a clear face-on orientation, whereas BO1 : Y6 showed a reduced edge-on-originated diffraction intensity of BO1 (disappearance of higher-order diffraction). The BO1Ph : Y6 film exhibited low intensity diffraction attributed to Y6, which could decrease electron mobility in this blend.

The hole mobility (μ_h) and electron mobility (μ_e) in the BHJ films were measured using the space-charge-limited current (SCLC) method. The results are summarized in Tables 2 and S4(SI), and the *JV* curves are shown in Fig. S61(SI). The μ_h and μ_e values in these blend films were *ca.* 0.29–2.2 × 10^{-4} $cm^2\ V^{-1}\ s^{-1}$ and 0.14–8.6 × 10^{-4} $cm^2\ V^{-1}\ s^{-1}$, respectively. Interestingly, the largest μ_h was obtained for BO1Ph : Y6, whereas the smallest μ_h was obtained for BO2 : Y6. The isotropic π–π stacking observed in the BO1Ph pristine film may merit the efficient 3D hole transport even in its Y6 blend, while the face-on orientation observed in the BO2 pristine film may be weakened in its Y6 blend. Although the μ_h and μ_e values are unlikely to directly affect the PCE of OPVs, we found that their ratio (μ_e/μ_h) well explained the device performance. The μ_e/μ_h of BO2 : Y6 was the most balanced value of 5.5, followed by 9.5 of BO1 : Y6 and 0.061 of BO1Ph : Y6. In particular, the lowest μ_e in BO1Ph : Y6 is closely associated with the degraded crystalline feature and orientation of Y6 in the 2D-GIXRD image of the blend. We therefore conclude that the highest PCE of BO2 : Y6 is due to its balanced charge-carrier mobilities and the well-mixed BHJ. Furthermore, the low PCE of BO1Ph : Y6 is due to the unbalanced charge-carrier mobility associated with its low orientation and crystallinity, although it showed well-suited photoabsorption, energy matching, and mixing (high PL quenching and good AFM morphology).

Evaluation of synthetic complexity

Finally, we evaluated the experimental synthetic cost of p-type polymers by calculating their synthetic complexity (SC).[63] The SC value is based on five parameters, namely, the number of reaction steps required to synthesize a compound, the reciprocal overall yield, the number of unit operations for isolation/purification, the number of column chromatographic purification steps, and the use of hazardous chemicals. A low SC means a low synthesis cost. The SC values of our three polymers were calculated based on our synthesis (Fig. S62–S64(SI)), and those for PM6 and PTQ10 were taken from the literature.[64,65] Due to the small synthesis step, PTQ10 has the lowest SC value of 27.7, while the SC values of our polymers were 48.7 for BO1, 49.7 for BO2, and 52.7 for BO1Ph (Table 2). Note that these values are lower than those of PM6 (73.1). However, the figure of merit (FOM), which is defined as the SC value divided by PCE, was still higher for our polymers (8.9–16.6) than those for PM6 (4.7) and PTQ10 (1.7), due to the low PCE values of the former (Table 2). Nevertheless, we demonstrated an effective approach for low-cost, simple-structured polymers based on ML prediction and manual search and modification. Given the good agreement between the experimental and predicted PCEs, this approach is vital for further development of new p-type polymers.

Conclusions

We developed a new p-type polymer with a simple chemical structure, designated as BO1, through a meticulous ML search. The PCE of BO1 : Y6 was 3.12%, which is approximately half of the predicted value. We systematically modified the chemical structure of BO1, resulting in BO2 (extended length of alkyl

chains) and BO1Ph (replacement of thiophene by phenylene). The OPV of BO2 : Y6 improved the PCE to 5.56% in line with the ML-predicted value of 6.02%. The most significant effects are improved solubility and miscibility with Y6, as well as a balanced charge carrier mobility. They are related to the predominant face-on orientation of the polymer film. However, the PCE of BO1Ph : Y6 remained low at 3.18% (prediction = 6.98%), despite a good photoabsorption spectrum complementary to that of Y6 and an increase in V_{OC} due to the deeper HOMO. Mobility balance and orientation of polymer/Y6 could not be well controlled in BO1Ph : Y6, resulting in a low FF and J_{sc}. We demonstrated that the SA and SC values associated with the synthetic cost, as well as ML exploration, are useful for further exploring simple-structured, efficient polymers for large-scale, low-cost OPVs.

Experimental

Machine learning

The PCE prediction of the ~200 000 virtually generated polymers was performed using the reported random forest ML model.[30,32] The training dataset is composed of the reported experimental data of polymer : NFA (1295) and virtually generated failure data (874).[33] Five-fold cross-validation (CV) of this model yielded a correlation coefficient of 0.84, excluding the failure data for CV, the mean absolute error of 1.59% (PCE), and root mean square error of 2.12% (PCE). The synthetic accessibility (SA) score of a polymer repeating unit was calculated from its SMILES using the "calculate_sascore" function in the RDKit library in the Python environment.

OPV fabrication and measurement

A ZnO layer was deposited on a cleaned and UV ozone-treated ITO layer by spin-coating with a ZnO precursor solution (0.1 g per mL zinc acetate dihydrate and 0.028 g mL^{-1} ethanolamine in 2-methoxyethanol) at 4000 rpm for 15 s. The precursor solution was heated at 60 °C for 1 h and at room temperature for 2 h with stirring at ~300 rpm. Before use, the precursor solution was filtered through a 0.22 μm filter. The substrate was annealed on a hot plate at 200 °C for 30 min. An active layer (p + n = 16 mg mL^{-1} of $CHCl_3$ with/without 0.5 vol% CN) was dynamically spin-coated onto the ZnO layer in a nitrogen glove box (O_2 < 0.1 ppm; H_2O < 0.1 ppm). The optimized spin rate for each polymer was 2000–3000 rpm for 30 s. The active layers were annealed on a hot plate at 110 °C for 10 min. An anode consisting of 10 nm MoO_3 and 100 nm Ag layers was sequentially deposited on the active layers, through a shadow mask, by thermal evaporation in a vacuum chamber. The resulting device configuration was ITO (120–160 nm)/ZnO (30 nm)/active layer (around 100 nm)/MoOx (10 nm)/Ag (100 nm) with an active area of 7.14 mm^2. The normal structure device was fabricated according to the same procedure, except for the bottom hole transport layer (Clevious P VP AI 4083 PEDOT:PSS, 5000 rpm, 60 s, anneal at 150 °C for 20 min) and top electron transport layer (1.5 mg mL^{-1} ethanol solution of PDINO, 3000 rpm, 30 s), yielding ITO/PEDOT:PSS/BHJ/PDINO/Ag. Current density–voltage curves were measured using a source-measure unit (ADCMT Corp., 6241A) under AM 1.5G solar illumination at 100 mW cm^{-2} (1 sun, monitored by a calibrated standard cell, Bunko Keiki BS-520BK) from a 300 W solar simulator (SAN-EI Corp., XES-301S). The EQE spectra were measured using a Bunko Keiki model SM-250KD equipped with a Keithley model 2401 source meter. The monochromated light power was calibrated using a silicon photovoltaic cell (Bunko Keiki model S1337-1010BQ).

SCLC hole-only device

A PEDOT:PSS layer was deposited on a cleaned and UV ozone-treated ITO layer by spin-coating with a PEDOT:PSS dispersion solution (Clevios P VP AI 4083) at 5000 rpm for 60 s. The precursor solution was filtered through a 0.45 μm filter before use. The substrate was annealed on a hot plate at 120 °C for 20 min. The procedure for the active layer was the same as for OPV devices. The resulting device configuration was ITO/PEDOT:PSS/active layer/MoOx/Ag, with an active area of 7.1 mm^2.

SCLC electron-only device

A ZnO layer was deposited on a cleaned and UV ozone-treated ITO layer by spin-coating with a ZnO precursor solution (0.1 g per mL zinc acetate dihydrate and 0.028 g per mL ethanolamine in 2-methoxyethanol) at 4000 rpm for 15 s. The precursor solution was heated at 60 °C for 1 h and at room temperature for 2 h with stirring at 280 rpm. The substrate was annealed on a hot plate at 200 °C for 30 min. The procedure for the active layer was the same as for OPV devices. The 20 nm C_{60} and 100 nm Ag layers were sequentially deposited on top of the active layers, through a shadow mask, by thermal evaporation in a vacuum chamber. The resulting device configuration was ITO/ZnO/active layer/C_{60}/Ag, with an active area of 7.1 mm^2.

General measurements

NMR (400 or 600 MHz for ^1H 151 MHz for ^{13}C) spectra were measured on a JEOL JNM-ECS400 (400 MHz) and a Bruker AVANCE III (600 MHz) spectrometer at 25 °C. Chemical shifts (δ) were relative to the resonances of tetramethylsilane (TMS). Absolute values of the coupling constants are given in Hertz (Hz), regardless of their sign, and multiplicities are abbreviated as singlet (s), doublet (d), multiplet (m), and broad (br). High-resolution matrix-assisted laser desorption direct analysis (DART) was performed on a JEOL JMS-T100LP. UV-vis absorption spectra, PL spectra, and FT-IR spectra were recorded on a Jasco V-730 UV-vis spectrophotometer, a Jasco FP-8300 spectrometer, and a Jasco FT/IR-4700AC spectrometer, respectively. PYS measurements were performed using a Bunko Keiki BIP-KV202GD in vacuum. TRMC measurements were performed using 550 nm light (intensity = 7.1×10^{15} photons per cm^2 per pulse) from an optical parametric oscillator (Continuum Inc., Panther) seeded by the third harmonic generation of a Nd:YAG laser (Continuum Inc., Surelite II, 5–8 ns pulse duration, 10 Hz) as the excitation and continuous microwave (~9.1 GHz, ~3 mW) as the probe. Thickness measurements were performed

with a Bruker Dektak XT. AFM was performed using a Bruker Innova. Contact angle was measured with an Asumi Giken CAME1. 2D-GIXRD was performed on the BL13XU beamline at SPring-8 (Japan Synchrotron Radiation Research Institute, JASRI), using 12.39 keV ($\lambda = 1$ Å) X-rays at a grazing incidence angle of 0.12°. The 2D diffraction patterns were recorded with a 2D detector (Dectris PILATUS 300 K).

Conflicts of interest

The authors declare no competing financial interests.

Data availability

The dataset of polymer:NFA OPV is available on the following sites: Zenodo **https://doi.org/10.5281/zenodo.17430843**.[30] For machine learning, R studio (ver. 2025.05.1+513) or Jupyter Notebook (ver. 7.3.2) with Python (ver. 3.10.18) were used.

Supplementary information: synthesis of monomers and polymers (Fig. S1–S47); SI Tables S1–S4; SI Schemes S1 and S2; SI Fig. S48–S64; SI references S1 (PDF). See DOI: **https://doi.org/10.1039/d5dd00418g**.

Acknowledgements

We acknowledge the financial support from KAKENHI from the Japan Society for the Promotion of Science (JSPS) (JP20H05836 and JP24H00484 for A. S. and JP22H04541 and JP21H00400 for F. I.), along with CREST (JPMJCR23O2 for A. S.) and PRESTO (JPMJPR21A2 for F. I.) from the Japan Science and Technology Agency (JST). We thank Dr Ryosuke Nishikubo at Osaka University for the fruitful discussions. We are grateful to Dr Tomoyuki Koganezawa at JASRI for his support with the 2D-GIXRD experiments, which were conducted at SPring-8, JASRI, with proposal numbers of 2023A1824, 2023B1758, 2023B1933, 2024A1666, 2024B1037, and 2024B1970.

Notes and references

1 J. J. M. Halls, C. A. Walch, N. C. Greenham, E. A. Marseglia, R. H. Friend, S. C. Moratti and A. B. Holmes, *Nature*, 1995, **376**, 498–500.
2 G. Yu and A. J. Heeger, *J. Appl. Phys.*, 1995, **78**, 4510–4515.
3 A. Armin, W. Li, J.-O. Sandberg, Z. Xiao, L. Ding, J. Nelson, D. Neher, K. Vandewal, S. Shoaee, T. Wang, H. Ade, T. Heumuller, C. Brabec and P. Meredith, *Adv. Energy Mater.*, 2021, **11**, 20003570.
4 W. Zhao, S. Li, H. Yao, S. Zhang, Y. Zhang, B. Yang and J. Hou, *J. Am. Chem. Soc.*, 2017, **139**, 7148–7151.
5 J. Yuan, Y. Zhang, L. Zhou, G. Zhang, H. L. Yip, T. K. Lau, X. Lu, C. Zhu, H. Peng, P. A. Johnson, M. Leclerc, Y. Cao, J. Ulanski, Y. Li and Y. Zou, *Joule*, 2019, **3**, 1140–1151.
6 H. Fu, Z. Wang and Y. Sun, *Angew. Chem., Int. Ed.*, 2019, **58**, 4442–4453.
7 S. Guan, Y. Li, C. Xu, N. Yin, C. Xu, C. Wang, M. Wang, Y. Xu, Q. Chen, D. Wang, L. Zuo and H. Chen, *Adv. Mater.*, 2024, **36**, 2400342.
8 P. Bi, J. Wang, Y. Cui, J. Zhang, T. Zhang, Z. Chen, J. Qiao, J. Dai, S. Zhang, X. Hao, Z. Wei and J. Hou, *Adv. Mater.*, 2023, **35**, 2210865.
9 M. Wei and D. F. Perepichka, *J. Mater. Chem. A*, 2025, **13**, 12785–12807.
10 Y.-W. Su, C.-E. Tsai, T.-C. Liao and K.-H. Wei, *Sol. RRL*, 2024, **8**, 2300927.
11 X. Kong, T. He, H. Qiu, L. Zhan and S. Yin, *Chem. Commun.*, 2023, **59**, 12051–12064.
12 M. Zhang, X. Guo, W. Ma, H. Ade and J. Hou, *Adv. Mater.*, 2015, **27**, 4655–4660.
13 Q. Liu, Y. Jiang, K. Jin, J. Qin, J. Xu, W. Li, J. Xiong, J. Liu, Z. Xiao, K. Sun, S. Yang, X. Zhang and L. Ding, *Sci. Bull.*, 2020, **65**, 272–275.
14 C. Sun, F. Pan, H. Bin, J. Zhang, L. Xue, B. Qiu, Z. Wei, Z.-G. Zhang and Y. Li, *Nat. Commun.*, 2018, **9**, 743–753.
15 X. Kong, J. Zhang, L. Meng, C. Sun, S. Qin, C. Zhu, J. Zhang, J. Li, Z. Wei and Y. Li, *CCS Chem.*, 2022, **5**, 841–850.
16 X. Zhang, X. Wu, J. Zhang, X. Kong, J. Li, A. Li, Z. Li, X. Li, M. Zhang, G. Yang, Y. Li and C. Sun, *ACS Nano*, 2024, **19**, 900–910.
17 X. Wu, X. Zhang, J. Zhang, Y. Wu, J. Li, X. Kong, Z. Li, X. Zhang, X. Li, A. Li, G. Yang and C. Sun, *Adv. Funct. Mater.*, 2024, **34**, 2405168.
18 Y.-M. Chang, Y.-T. Hsiao and K.-W. Tsai, *Adv. Energy Mater.*, 2024, **14**, 2400064.
19 K. Sodeyama, Y. Igarashi, T. Nakayama, Y. Tateyama and M. Okada, *Phys. Chem. Chem. Phys.*, 2018, **20**, 22585–22591.
20 Y. Kobayashi, Y. Miyake, F. Ishiwari, S. Ishiwata and A. Saeki, *RSC Adv.*, 2023, **13**, 15107–15113.
21 R.-P. Joshi, J. Eickholt, L. Li, M. Fornari, V. Barone and J.-E. Peralta, *ACS Appl. Mater. Interfaces*, 2019, **11**, 18494–18503.
22 J.-H. Li, C.-R. Zhang, M.-L. Zhang, X.-M. Liu, J.-J. Gong, Y.-H. Chen, Z.-J. Liu, Y.-Z. Wu and H.-S. Chen, *Org. Electron.*, 2024, **125**, 106988.
23 R. Cao, C.-R. Zhang, M. Li, X.-M. Liu, M.-L. Zhang, J.-J. Gong, Y.-H. Chen, Z.-J. Liu, Y.-Z. Wu and H.-S. Chen, *Sol. RRL*, 2024, **8**, 2400370.
24 S. G. Jung, G. Jung and J. M. Cole, *J. Chem. Phys.*, 2023, **159**, 194106.
25 F. Häse, L. M. Roch, P. Friederich and A. Aspuru-Guzik, *Nat. Commun.*, 2020, **11**, 4587.
26 A. Mahmood and J.-L. Wang, *Energy Environ. Sci.*, 2021, **14**, 90–105.
27 T. Lu, M. Li, W. Lu and T.-Y. Zhang, *J. Mater. Inform.*, 2022, **2**, 7.
28 S. Shafian, F. N. Mohd Salehin, S. Lee, A. Ismail, S. Mohamed Shuhidan, L. Xie and K. Kim, *ACS Appl. Energy Mater.*, 2025, **8**, 699–722.
29 S. Nagasawa, E. Al-Naamani and A. Saeki, *J. Phys. Chem. Lett.*, 2018, **9**, 2639–2646.
30 Y. Miyake and A. Saeki, *J. Phys. Chem. Lett.* 2021, 12, 12391–12401–, DOI: **10.1021/acs.jpclett.1c03526**.
31 Y. –C. Lin, Y. –J. Lu, C. –S. Tsao, A. Saeki, J. –X. Li, C. –H. Chen, H. –C. Wang, H. –C. Chen, D. Meng, K. –H. Wu,

Y. Yang and K. –H. Wei, *J. Mater. Chem. A*, 2019, **7**, 3072–3082.
32 K. Kranthiraja and A. Saeki, *Adv. Funct. Mater.*, 2021, **31**, 2011168.
33 Y. Miyake, K. Kranthiraja, F. Ishiwari and A. Saeki, *Chem. Mater.*, 2022, **34**, 6912–6920.
34 K. Kranthiraja and A. Saeki, *ACS Appl. Mater. Interfaces*, 2022, **14**, 28936–28944.
35 B. Zheng, L. Huo and Y. Li, *NPG Asia Mater.*, 2020, **12**, 3.
36 C. An and J. Hou, *Acc. Mater. Res.*, 2022, **3**, 540–551.
37 S. Zhang, Y. Qin, J. Zhu and J. Hou, *Adv. Mater.*, 2018, **30**, 1800868.
38 M. Zhang, X. Guo, W. Ma, H. Ade and J. Hou, *Adv. Mater.*, 2015, **27**, 4655–4660.
39 P. Ertl and A. Schuffenhauer, *J. Cheminform.*, 2009, **1**, 8.
40 T. E. A. Frizon, A. A. Vieira, F. N. D. Silva, S. Saba, G. Farias, B. D. Souza, E. Zapp, M. N. Lôpo, H. D. C. Braga, F. Grillo, S. F. Curcio, T. Cazati and J. Rafique, *Front. Chem.*, 2020, **8**, 360.
41 J.-M. Jiang, P.-A. Yang, T.-H. Hsieh and K.-H. Wei, *Macromolecules*, 2011, **44**, 9155–9163.
42 T. E. A. Frizon, A. A. Vieira, F. N. D. Silva, S. Saba, G. Farias, B. D. Souza, E. Zapp, M. N. Lôpo, H. D. C. Braga, F. Grillo, S. F. Curcio, T. Cazati and J. Rafique, *Front. Chem.*, 2020, **8**, 360–366.
43 Q. Peng, S.-L. Lim, I. H.-K. Wong, J. Xu and Z.-K. Chen, *Chem. Eur. J.*, 2012, **18**, 12140–12151.
44 Y. Nishii, M. Ikeda, Y. Hayashi, S. Kawauchi and M. Miura, *J. Am. Chem. Soc.*, 2020, **142**, 1621–1629.
45 P. Pananusorn, H. Sotome, H. Uratani, F. Ishiwari, K. Phomphrai and A. Saeki, *J. Chem. Phys.*, 2024, **161**, 184710.
46 J. Bertrandie, J. Han, C. S. P. D. Castro, E. Yengel, J. Gorenflot, T. Anthopoulos, F. Laquai, A. Sharma and D. Baran, *Adv. Mater.*, 2022, **34**, 2202575.
47 Y. Liu, Y. Wu, Y. Geng, E. Zhou and Y. Zhong, *Adv. Funct. Mater.*, 2022, **32**, 2206707.
48 P. J. Leenaers, H. V. Eersel, J. Li, M. M. Wienk and R. A. J. Janssen, *Macromolecules*, 2020, **53**, 7749–7758.
49 Z. Hu, Q. Huang, C. Liu, A. Song, L. Shao, Y. Bai, Z. Hu, K. Zhang, F. Huang and Y. Cao, *Org. Mater.*, 2022, **4**, 18–27.
50 J. C. Bijleveld, B. P. Karstern, S. G. J. Mathijssen, M. M. Wienk, D. M. Leeuw and R. A. J. Janssen, *J. Mater. Chem.*, 2011, **21**, 1600–1606.
51 M. Suzuki, K. Terai, C. Quinton, H. Hayashi, N. Aratani and H. Yamada, *Chem. Sci.*, 2020, **11**, 1825–1831.
52 A. Saeki, *Polym. J.*, 2020, **52**, 1307–1321.
53 C. Nishikawa, R. Nishikubo, F. Ishiwari and A. Saeki, *JACS Au*, 2023, **3**, 3194–3203.
54 A. Natsuda, T. Saito, R. Shirouchi, Y. Sakamoto, T. Takeyama, Y. Tamai and H. Ohkita, *Energy Environ. Sci.*, 2022, **15**, 1545–1555.
55 H. Tamura, *J. Chem. Phys.*, 2025, **162**, 054712.
56 M. Gao, W. Wang, J. Hou and L. Ye, *Aggregate*, 2021, **2**, e46.
57 T. Zhang, C. An, Q. Lv, J. Qin, Y. Cui, Z. Zheng, B. Xu, S. Zhang, J. Zhang, C. He and J. Hou, *J. Energy Chem.*, 2021, **59**, 30–37.
58 J. Zhou, L. Wang, C. Liu, C. Guo, C. Chen, Y. Sun, Y. Yang, J. Cheng, Z. Gan, Z. Chen, W. Sun, J. Zhou, W. Xia, D. Liu, W. Li and T. Wang, *J. Am. Chem. Soc.*, 2024, **146**, 34998–35006.
59 M. E. Ziffer, S. B. Jo, H. Zhong, L. Ye, H. Liu, F. Lin, J. Zhang, X. Li, H. W. Ade, A. K.-Y. Jen and D. S. Ginger, *J. Am. Chem. Soc.*, 2018, **140**, 9996–10008.
60 I. Osaka and K. Takimiya, *Polymer*, 2015, **59**, A1–A15.
61 Y. Fu, T. H. Lee, Y.-C. Chin, R. A. Pacalaj, C. Labanti, S. Y. Park, Y. Dong, H. W. Cho, J. Y. Kim, D. Minami, J. R. Durrant and J.-S. Kim, *Nat. Commun.*, 2023, **14**, 1870.
62 S. Doi, T. Mikie, K. Yamanaka, Y. Sato, H. Ohkita, M. Saito and I. Osaka, *Polym. J.*, 2024, **56**, 1051–1059.
63 R. Po, G. Bianchi, C. Carbonera and A. Pellegrino, *Macromolecules*, 2015, **48**, 453–461.
64 J. Xiao, X. Jia, C. Duan, F. Huang, H.-L. Yip and Y. Cao, *Adv. Mater.*, 2021, **33**, 2008158.
65 C. Sun, F. Pan, S. Chen, R. Wang, R. Sun, Z. Shang, B. Qiu, J. Min, M. Lv, L. Meng, C. Zhang, M. Xiao, C. Yang and Y. Li, *Adv. Mater.*, 2019, **31**, 1905480.