A Hammett's analysis of the Substituent Effect in Functionalized Diketopyrrolopyrrole (DPP) Systems: optoelectronic properties and intramolecular charge transfer effects

Gabriel Monteiro-de-Castro;[a] Itamar Borges Jr.[a,b,*]

Instituto Militar de Engenharia (IME), Praça Gen. Tibúrcio 80, Rio de Janeiro, RJ, 22290-270, Brazil.
[a] Departamento de Química, IME
[b] Programa de Pós-Graduação em Engenharia de Defesa, IME
* E-mail: itamar@ime.eb.br

Abstract

Diketopyrrolopyrrole (DPP) systems have promising applications in different organic electronic devices. In this work, we investigated the effect of 20 distinct substituent groups on the optoelectronic properties of DPP-based derivatives as the donor (D)-material in an organic photovoltaics (OPV) device. For this purpose, we employed Hammett's theory, which quantifies the electron-donating or -withdrawing properties of a given substituent group. Machine-Learning (ML)-based σ_m, σ_p, σ_m^0, σ_p^0, σ_p^+, σ_p^-, σ_I, and σ_R Hammett's constants previously determined were used. Mono- (DPP-X_1) and di-functionalized (DPP-X_2) DPPs, where X is a substituent group, were investigated using density functional theory (DFT), time-dependent DFT (TDDFT), and *ab initio* methods. Several properties were computed using CAM-B3LYP and the second-order algebraic diagrammatic construction, ADC(2), ab initio wave function method, including the adiabatic ionization potential (IP_A), the electro affinity (EA_A), the HOMO-LUMO gaps (E_g), the maximum absorption wavelengths (λ_{max}), the first excited state transition $^1S_0 \rightarrow {}^1S_1$ energies (ΔE) (the optical gap), and exciton binding energies. From the optoelectronic properties and employing typical acceptor systems, the power conversion efficiency (PCE), open-circuit voltage (V_{OC}), and fill factor (FF) were predicted for a DPP-based OPV device. These photovoltaic properties were also correlated with the ML-based Hammett's constants. Overall, good correlations between all properties and the different types of σ constants were obtained, except for the σ_I constants, which are related to inductive effects. This scenario suggests that resonance is the main factor controlling electron donation and withdrawal effects. We found that substituent groups with large σ values can produce higher photovoltaic efficiencies. It was also found that electron-withdrawing groups reduced E_g and ΔE considerably compared to the unsubstituted DPP-H. Moreover, for every decrease (increase) in the values of a given optoelectronic property of DPP-X_1 systems, a more significant decrease (increase) in the same values was observed for the DPP-X_2, thus showing that the addition of second substituent results in a more extensive influence on all electronic properties. For the exciton binding energies, an unsupervised machine learning algorithm identified groups of substituents characterized by average values (centroids) of Hammett's constants that can drive the search for new DDP-derived materials. Our work presents a promising approach by applying Hammett's theory on molecular engineering DPP-based molecules and other conjugated molecules for applications on organic optoelectronic devices.

Keywords: Diketopyrrolopyrrole; Hammett theory; Substituent effects; Optoelectronic properties; Organic electronics; Donor properties

1. Introduction

The accurate theoretical investigation of conjugated organic molecules for organic electronic devices is essential for basic research and for developing practical applications.[1–3] Diketopyrrolopyrrole (DPP) has drawn much attention since it was discovered in 1974[4] for use as dyes and plastics.[5] Later, several theoretical and experimental studies on DPP appeared due to its potential in optoelectronic applications, especially in organic light-emitting diodes (OLED), organic solar cells (OSC), and field-effect transistors (FET).[6,7] DPP-based materials possess strong absorption and emission ability, excellent thermo- and photostability, good charge mobility, and a significant Stokes shift.[8–11]

Many theoretical works have been carried out on DPP systems. Density Functional Theory (DFT)[12] and Time-Dependent DFT (TDDFT)[13] were used to predict optoelectronic and intra-charge transfer (ICT) properties of a series of DPP derivatives functionalized with aromatic substituent groups.[6] It was found that these aromatic sidechains attached to a DPP core can broaden the absorption spectra and increase ICT, thus, making these DPP systems a promising set of compounds for use in OLEDs and OSCs. A 2016 TDDFT work fine-tuned DPP and its derivatives by lactam-lactim and alkoxy-thioalkoxy exchange to improve the efficiency of bulk heterojunction (BHJ) solar cells. The highest occupied molecular orbital (HOMO) energies and the gaps between the HOMO and the lowest unoccupied molecular orbital (LUMO) were reduced, the absorption spectra were redshifted, and hole transport improved.[14] A recent paper[15] studied the photophysical properties of small DPP-based molecules in a D-π-A-π-D-type structure, where donors (D) and acceptor (A) are chemically bonded by π-bridges. DFT and TDDFT calculations suggested that different aromatic side groups attached to DPP can influence molecular photophysical properties. Moreover, ICT effects were predicted, indicating a potential for OLEDs and OSCs.

Homopolymers and copolymers based on DPP were synthesized and characterized theoretically.[16] TDDFT calculations on the large structures indicated an extensive absorption range from 300 nm (4.1 eV) to 1100 nm (1.1 eV), confirmed by UV-vis spectroscopy. However, the extended crystallinity, observed by x-ray diffraction, caused phase separation, thus, low photovoltaic performance of these DPP-based polymers.

Introducing a given substituent into a DPP core can be easily achieved[10] and can change the optical and charge transport properties of the entire system.[17] In a donor-acceptor (D-A) structure, properties such as energy gaps and ICT can be tuned by including different electron donating (EDG) and/or withdrawing groups (EWG). The effects of substituent groups can be studied employing Hammett's theory (HT), where the σ constants quantify the ability of a given substituent to donate or withdraw electrons.[18,19] HT provides a basis for quantitatively studying structure-reactivity in organic chemistry.[20–22] Initially developed for describing the ionization of benzoic acid derivatives with different substitutions at *para*, *meta*, and *ortho* positions, Hammett's constants measure how the charge (electron) distribution of a benzene-type core is modified when different substituents are attached to it, have a much more general character and different applications.[23,24]

To design and predict optoelectronic properties of functionalized polypyrrole (PPy) derivatives, Batagin-Neto and coworkers[25] used the resonance and inductive Hammett's constants σ_R and σ_I, respectively, of a set of ten common substituent groups (–H, –CCH, –Ph, –Me, –CN, –F, –NO2, –NH2, –OH, and –OMe) to analyze their effects on the optical properties of PPy. Correlation between the σ constants and optoelectronic parameters (e.g., HOMO and LUMO energies and the maximum absorption wavelength λ_{max}), was found, and a set of new PPy derivatives with potential application in optoelectronic devices was proposed. Our approach here is similar and was extended for a much larger and new set of Hammett's constants focusing on DPP-based systems.

In this work, we investigated theoretically the correlation between Hammett's substituent constants and different opto- and electrochemical properties of DPP-based systems. The purpose was to understand and systematically investigate the chemical behavior of these systems for engineering DPP molecules and other conjugated molecules for applications on organic electronic devices.

2. Computational methods

2.1. Hammett's constants and Machine Learning approach

Hammett's original constants consisted of *meta* (σ_m) and *para* (σ_p) parameters obtained from ionizing functionalized benzoic acids in water at 25ºC. However, due to the strong influence of electronic phenomena, such as the inductive (I) and resonance (R) effects, new constants describing the delocalization of positive and negative charges have been proposed. The σ_R and σ_I constants represent the I and R components, respectively, of Hammett's original σ_m and σ_p set.[26,27] Considering the series of 4-substituted bicyclo[2,2,2]octane synthesized in 1953,[28] substituents with a negative R-effect – meaning in this case that the electron delocalization flows from the molecule to the substituent group, have approximately the same values for σ_m and σ_I, which indicates that σ_m has a significant inductive (I) effect rather than a resonance (R) one. Therefore, it was suggested the following decomposition of σ_p, [22]

$$\sigma_p = \sigma_I + \sigma_R \tag{3}$$

Obtaining σ_R values becomes then a matter of accurately determining σ_p and σ_I.[27]

In 1957, a set of parameters named σ^+ and σ^- was developed based on the solvolysis of substituted phenyldimethylcarbinyl chlorides.[29] These parameters describe the R-effect when the substituent is in direct conjugation with the molecule's reaction center – e.g., the carboxyl group in the benzoic acid.[20,30] It was soon realized, however, that this new set would apply only to *para*-substituted systems due to the more prominent R-effect at this position.[29,31–33] For this reason, the parameters σ_p^+ and σ_p^- were created, with the first being associated with groups that can stabilize positive charges through resonance. The second describes those substituents that stabilize negative charges through the same effect.[21,22,34]

Another set of Hammett's constants was created in 1959, the σ^0 set. It was based on the idea that one should not consider the R-effect between the substituent and the reaction center.[21,35] The σ_m^0 values have the attractive property of being essentially equal

to the original σ_m values since direct substituent-reaction center resonance is not very significant.[21,27] Moreover, for EDG, which would not participate in strong resonance at the *para*-position, equality was usually observed.[21]

Despite the success of Hammett's constants for rationalizing different phenomena, not every constant is determined experimentally, and consistent values of every type of σ are unavailable. Thus some efforts have been made to find ways to calculate them. In this work, we used our Machine Learning (ML)-based method described in Ref. 36 to predict Hammett's constants for substituents with unknown σ and recomputed the other values to produce more consistent sets. Our ML procedure employed the quantum chemical Hirshfeld atomic charges[37] of different carbon atoms in 180 *meta*- and *para*-substituted benzoic acid derivatives.

Table 1 below lists the values of Hammett's constants for the 20 substituent groups plus the hydrogen used in this work.[36] The available values of σ in the literature[18–20,22,23,25,30,38,39] are collected in Table S1 of the Supporting Information (SI). Also, in the SI, one can find the structures of the investigated substituent groups in Figure S1. The substituent groups were bonded to a DPP core, thus creating the DPP-X_1 or DPP-X_2 systems where X represents the substituent group - see Figure 1.

Table 1. ML-based Hammett's constants for a set of different substituents.[36]

X	σ_m	σ_p	σ_R	σ_I	σ_p^+	σ_p^-	σ_m^0	σ_p^0
–H	0.03	0.06	-0.02	0.07	-0.09	0.25	0.05	0.06
–Br	0.38	0.28	-0.19	0.47	0.01	0.42	0.41	0.37
–Ph	0.07	0.01	-0.10	0.11	-0.19	0.14	0.07	0.00
–CCH	0.33	0.29	-0.01	0.31	0.20	0.47	0.34	0.26
$–CF_3$	0.42	0.53	0.10	0.42	0.52	0.79	0.45	0.54
–Me	-0.02	-0.09	-0.15	0.06	-0.35	-0.01	-0.02	-0.09
$–CHF_2$	0.32	0.33	0.02	0.31	0.26	0.54	0.34	0.34
–CHO	0.39	0.60	0.31	0.28	0.73	0.91	0.39	0.55
–Cl	0.38	0.25	-0.22	0.48	-0.04	0.36	0.40	0.33
–CN	0.67	0.73	0.13	0.60	0.79	0.98	0.69	0.68
–COMe	0.32	0.44	0.20	0.24	0.52	0.73	0.31	0.41
–COOMe	0.32	0.45	0.21	0.23	0.54	0.71	0.31	0.37
–COOH	0.38	0.54	0.25	0.28	0.67	0.81	0.38	0.45
–F	0.40	0.17	-0.40	0.56	-0.24	0.14	0.41	0.26
$–NMe_2$	-0.26	-0.75	-0.79	0.05	-1.58	-0.86	-0.25	-0.52
$–NH_2$	-0.11	-0.58	-0.76	0.18	-1.39	-0.66	-0.09	-0.34
–NHMe	-0.19	-0.69	-0.80	0.11	-1.55	-0.80	-0.18	-0.48
$–NO_2$	0.72	0.87	0.23	0.64	1.05	1.09	0.73	0.78
–OMe	0.06	-0.29	-0.56	0.27	-0.91	-0.36	0.05	-0.15
–OH	0.14	-0.23	-0.60	0.37	-0.90	-0.27	0.14	-0.02
$–SO_2Me$	0.52	0.64	0.16	0.48	0.69	0.94	0.55	0.65

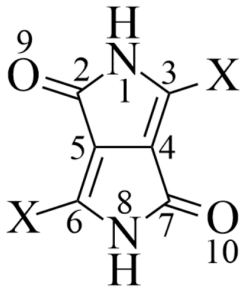

Figure 1. DPP core with the position of X substituent group (Table 1).

2.2. Quantum Chemical approach

Unless otherwise stated, all the calculations were performed using the ORCA package version 5. 0. 3.[40] The DPP-X systems were divided into two sets: one of mono- (DPP-X_1) and another of di-substituted (DPP-X_2) structures. Each set then comprises 20 molecules. DPP-H is the unsubstituted DPP system where X = hydrogen.

The first step in the calculations was pre-optimizing all 41 different molecular geometries using the Avogadro[41] MMFF94s force field[42] because it efficiently models trigonal nitrogen centers given the altered out-of-plane bending parameters compared to its predecessor MMFF94.[43] Singlet ground state ($^1S_0$) geometry optimizations in the gas phase of all DPP-X systems were carried out with the B3LYP-D3/Def2-TZVP method. Afterward, geometry optimizations of the same structures in both cationic ($[DPP\text{-}X]^+$) and anionic ($[DPP\text{-}X]^-$) states were carried out at the same level of theory. The B3LYP exchange-correlation functional[44] is widely used in geometry optimizations of organic molecular systems due to its low computational cost and accurate converged structures. Grimme's D3 semiempirical dispersion corrections were also employed.[45]

To check that the converged structures are minima on the potential energy surface, vibrational frequency calculations were done at the same B3LYP-D3/Def2-TZVP level. When necessary, the *TightOPT* keyword was used for eliminating, in an additional optimization process, any eventual imaginary frequency.

Finally, TDDFT calculations were performed using the CAM-B3LYP functional [46] and the Def2-TZVP basis set[47] on the optimized neutral $^1S_0$ structures, namely, the CAM-B3LYP/Def2-TZVP//B3LYP-D3/Def2-TZVP method. The same optimized $^1S_0$ structures were also used for the second-order algebraic diagrammatic construction, ADC(2)[48,49] calculations of optical oscillator strengths and electronic transition energies using the Turbomole 6.6 package.[50]

2.3. Optoelectronic and photovoltaic properties

We computed several optoelectronic properties of the DPP-X systems as the donor (D) material in a theoretical OPV device. These properties characterize the performance of OPV devices and are described below.[51]

From the TDDFT calculations, we collected the energies of the HOMO (E_{HOMO}), LUMO (E_{LUMO}) orbitals, and obtained the HOMO-LUMO gaps E_g ($= E_{LUMO} - E_{HOMO}$).

We computed the maximum absorption wavelengths (λ_{max}), the transition energies ΔE (the optical gap), and the optical oscillator strengths f of the first singlet-singlet electronic excitation $^1S_0 \rightarrow {}^1S_1$.

Additionally, we calculated the adiabatic ionization potential (IP_A) and the electro affinity (EA_A) according to the following equations [52]

$$IP_A = E^+_{DPPX} - E_{DPPX} \quad (4a)$$
$$EA_A = E_{DPPX} - E^-_{DPPX} \quad (4b)$$

where E^+_{DPPX}, E^-_{DPPX}, and E_{DPPX} are the total energy of the DFT-optimized charged [DPP-X]$^+$, [DPP-X]$^-$, and neutral DPP-X geometries, respectively.

Since DPP-based derivatives have promising applications as a donor (D) material in OSCs, the following parameters were calculated for a theoretical OPV device comprised of our DPP-X systems as the D material. The open-circuit voltage (V_{OC}), according to Scharber's equation, is given in volts by[14,53,54]

$$V_{OC} = \frac{1}{q}(|E^D_{HOMO}| - |E^A_{LUMO}|) - 0.3 \quad (8)$$

where q is the elementary charge, 0.3 V is an empirical factor related to the efficiency of charge separation between donor and acceptor materials,[54] E^D_{HOMO} is the energy of the HOMO of the D-material (DPP-X) and E^A_{LUMO} is the LUMO energy of the acceptor (A) material. We considered here the standard [6,6]-phenyl-C_{60}-butyric acid methyl ester ($PC_{61}BM$) as the acceptor (A) in a typical bulk heterojunction, so E^A_{LUMO} is a constant throughout this work with a value equal to -3.66 eV.[55] It is worth mentioning that newer non-fullerene-type materials, such as the 2,2'-((2Z,2'Z)-((4,4,9,9-tetrahexyl-4,9-dihydro-s-indaceno[1,2-b:5,6-b']dithiophene-2,7-diyl)bis(methanylylidene))bis(3-oxo-2,3-dihydro-1H-indene-2,1-diylidene))dimalononitrile (IDIC) and the 3,9-bis(2-methylene-(3-(1,1-dicyanomethylene)-indanone))-5,5,11,11-tetrakis(4-hexylphenyl)-dithieno[2,3-d:2',3'-d']-s-indaceno[1,2-b:5,6-b']dithiophene (ITIC), could also be used for this calculation, but the only difference in the analysis would be a linear variation of the absolute values of V_{OC} and other related properties.

The fill factor (FF) is related to the efficiency of the free carrier removal from a photovoltaic device. It can be obtained from the expression [14,53]

$$FF = \frac{\frac{qV_{OC}}{k_BT} - ln\left(\frac{qV_{OC}}{k_BT} + 0.72\right)}{\frac{qV_{OC}}{k_BT} + 1} \quad (9)$$

where k_B is the Boltzmann constant, T = 298.15 K, and V_{OC} is calculated by Eq. (8). Finally, the power conversion efficiency (PCE) given by[53]

$$PCE = \frac{V_{OC} \cdot FF \cdot J_{SC}}{P_{inc}} \quad (10)$$

where P_{inc} is the incident photon-to-current efficiency (considered 136.7 mW/cm^2)[56] and J_{SC} is the short-circuit current. Since J_{SC} indirectly depends on parameters such as the light absorption efficiency, exciton diffusion efficiency, charge separation efficiency, and charge collection efficiency,[53] it depends on the solar cells' physical characteristics (such as the properties of the active layer).[57] Therefore, we base our analysis on a few theoretical and experimental works on aromatic DPP-based OPVs. These works report values of J_{SC} for aromatic DPP-based materials ranging from ~1 to ~16 mA/cm^2.[5,58,59] We considered an average value of J_{SC} = 10 mA/cm^2 for obtaining reasonable values of PCE compared to those reported in the literature.

2.3. Excitonic and intramolecular charge transfer analysis between molecular fragments

Following the approach developed in Ref. 60, we analyzed the excitonic and intramolecular charge transfer (ICT) properties of all DPP-X systems calculated with the TDDFT and ADC(2) methods. The analysis includes the calculation of the charge transfer number (Ω) given by

$$\Omega = \frac{1}{2} \sum_{\substack{a \in F_A \\ b \in F_B}} \left(D_{ab}^{0n,[LO]} \right)^2 \tag{11}$$

where the a and b are orbitals situated in the molecular fragments F_A and F_B, respectively, n is a given excited state, and $D_{ab}^{0n,[LO]} = \langle 0 | E_{ab} | n \rangle$ is the one-electron transition density matrix of the ground state in a basis of localized orbitals (LO). E_{ab} represents the excitation energy from orbital a to b.[61] The charge transfer character is given by

$$CT = \frac{1}{\Omega} \sum_{\substack{F_A \\ F_B \neq F_A}} \Omega_{F_A F_B} \tag{12}$$

which includes the weight of configuration for the case where orbitals a and b are on different fragments (say, F_A and F_B, respectively). A CT = 1 value represents a complete charge-separated state while CT = 0 indicates a Frenkel excitonic state, i.e., a locally excited transition.[60]

The participation ratio (PR) quantifies the fragments involved during the excitation and is defined as

$$PR = \frac{\left(PR_i + PR_f \right)}{2} \tag{13}$$

where PR_i and PR_f represent the initial and final numbers, respectively, of fragments where the orbitals (or holes) are localized. PR = 1 represents localized excitations.[60]

Finally, the coherence length (COH) corresponds to the average electron-hole separation and is given by

$$COH = \frac{\Omega^2}{PR \sum_A \sum_B \Omega_{AB}^2} \quad (14)$$

This value is a Frenkel exciton while $COH = PR$ refers to loosely bound excitons.[60]

Figure **2** below depicts the division of each fragment used in the excitonic and intramolecular charge transfer analysis.

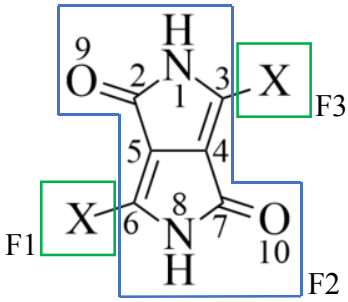

Figure 2. The fragments F1, F2, and F3 of DPP used for the excitonic and intramolecular charge transfer analysis.

2.4. Correlation between Hammett's constants and optoelectronic properties

After computing the optoelectronic properties, we plotted several curves for the different types of Hammett's σ values (Table 1) to find correlations. The Python NumPy library [62] was used for this purpose. The *polyfit* and *poly1d* algorithms were used to fit linear equations of the type $P(\sigma) = m\sigma + n$ where $P(\sigma)$ is an optoelectronic or photovoltaic property. Whenever a linear fit was not found, we used the OriginPro16 software to fit nonlinear functions to the data points.

3. Results and discussion

3.1. Geometries

The optimized geometries of all DPP-X systems are collected in Table S2 of the Supporting Information (SI). Considering θ to be the dihedral angle between the DPP core and the substituent plane, most of the optimized geometries are planar, except for the following molecules: DPP-Ph (θ = 2.1°), DPP-Ph_2 (θ = 2.2°), DPP-SO_2Me (θ = 41.4°), DPP-$(SO_2Me)_2$ (θ = 42.5°), DPP-$(NHMe)_2$ (θ = 12.2°), DPP-NMe_2 (θ = 10.1°), and DPP-$(NMe_2)_2$ (θ = 16.3°), which show primarily minor deviations from a complete planar structure.

The DPP-X systems having the –SO_2Me, –NMe_2, and –NHMe substituent groups showed deviations from planarity probably due to steric effects between the methyl groups and the hydrogen and/or oxygen atoms in the DPP core (see Figure 3a and c). For both mono- and di-substituted DPP-phenyl systems, the minor out-of-plane distortion could also result from the steric effect between the *ortho*-hydrogen atoms of the phenyl group and the nearby hydrogen bonded to the nitrogen atoms of the DPP core (Figure 3b). Experimental data shows that, without an alkyl chain bonded to the nitrogen atom in the DPP core, the phenyl group would be almost in the same plane of the DPP center,[17] thereby confirming our results.

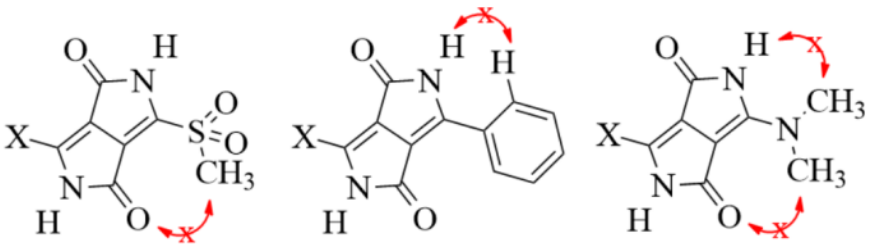

Figure 3. Examples of steric effects in DPP-based systems.

Planarity is essential when designing DPP-based or other conjugated materials for optoelectronic devices because it facilitates charge transport. The presence of substituents like the phenyl group in the same plane of the DPP core improves the electronic coupling in the system, thus favoring a resonance effect.[17] A stronger π-π-stacking between DPP units in an actual OPV arrangement enhances the charge carrier mobility, leading to greater efficiency of OLEDs, OSCs, and FETs.[7,63]

3.2. Optoelectronic properties and Hammett's constants

We found interesting results concerning the influence of substituent groups on the properties of organic-based optoelectronic devices. The PCE, V_{OC}, IP_A, EA_A, E_{HOMO}, E_{LUMO}, and ΔE values show for the DPP-X_1 and DPP-X_2 systems a good correlation with the values of Hammett's σ_m, σ_p, σ_m^0, σ_p^0, σ_p^+, σ_p^-, and σ_R constants, except for σ_I (*vide infra*). We discuss these correlations between the ML-based Hammett's constants and optoelectronic properties; the remaining plots are in the Supporting Information (SI) in Figures S2-S3, S6-S9, and S11-S12.

Frontier molecular orbital (FMO) energies. Figure 4 and Figure 5 show the correlation between the ML-based Hammett's constants and FMO energies calculated with the CAM-B3LYP functional. Tables S3-S4 in the SI collect the values of the FMO energies and HOMO-LUMO gaps (E_g).

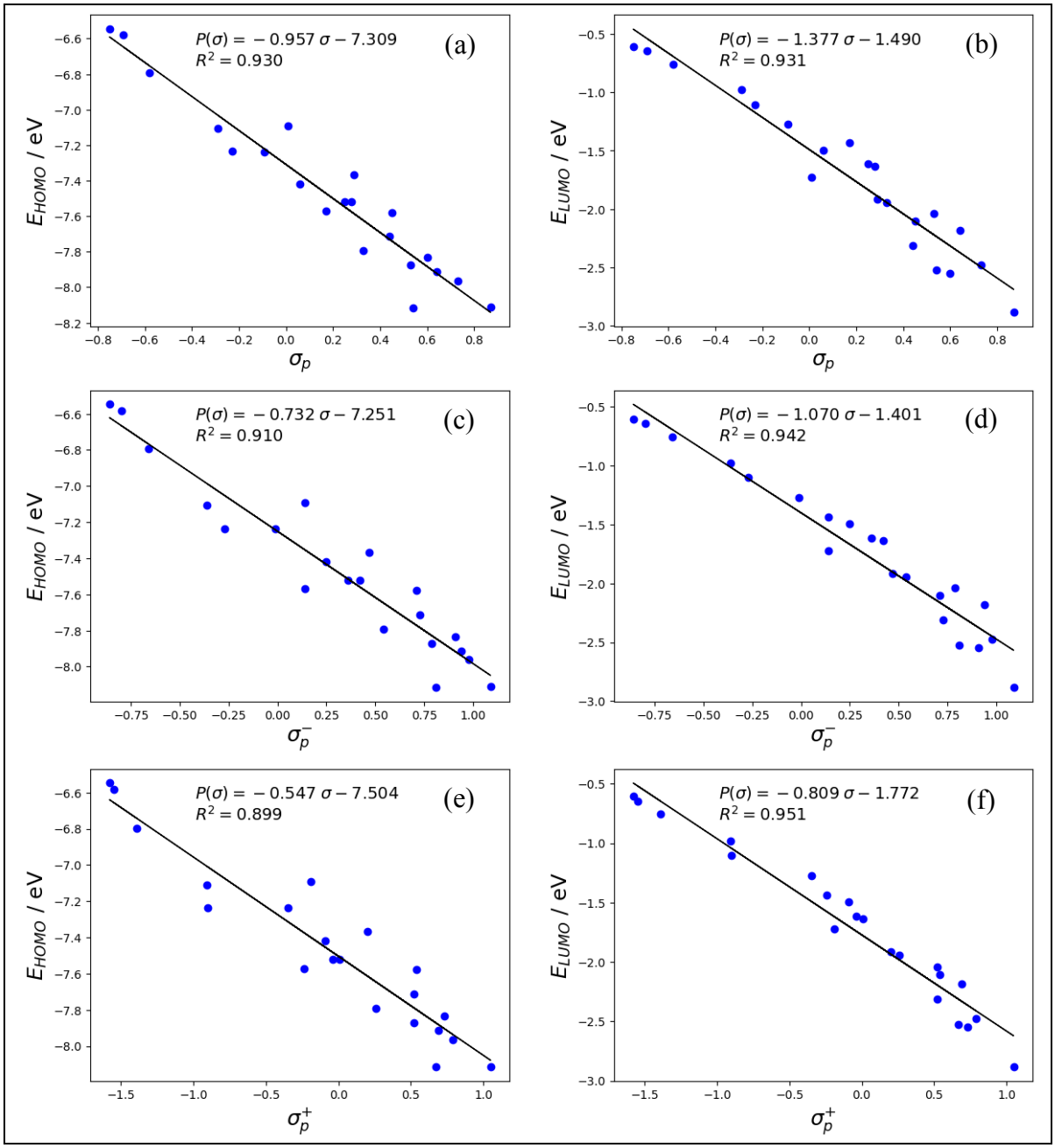

Figure 4. Best correlations between ML-based Hammett's constants and DFT FMO energies (E_{HOMO} – right panels and E_{LUMO} – left panels) of the mono-substitued DPP-X_1 derivatives.

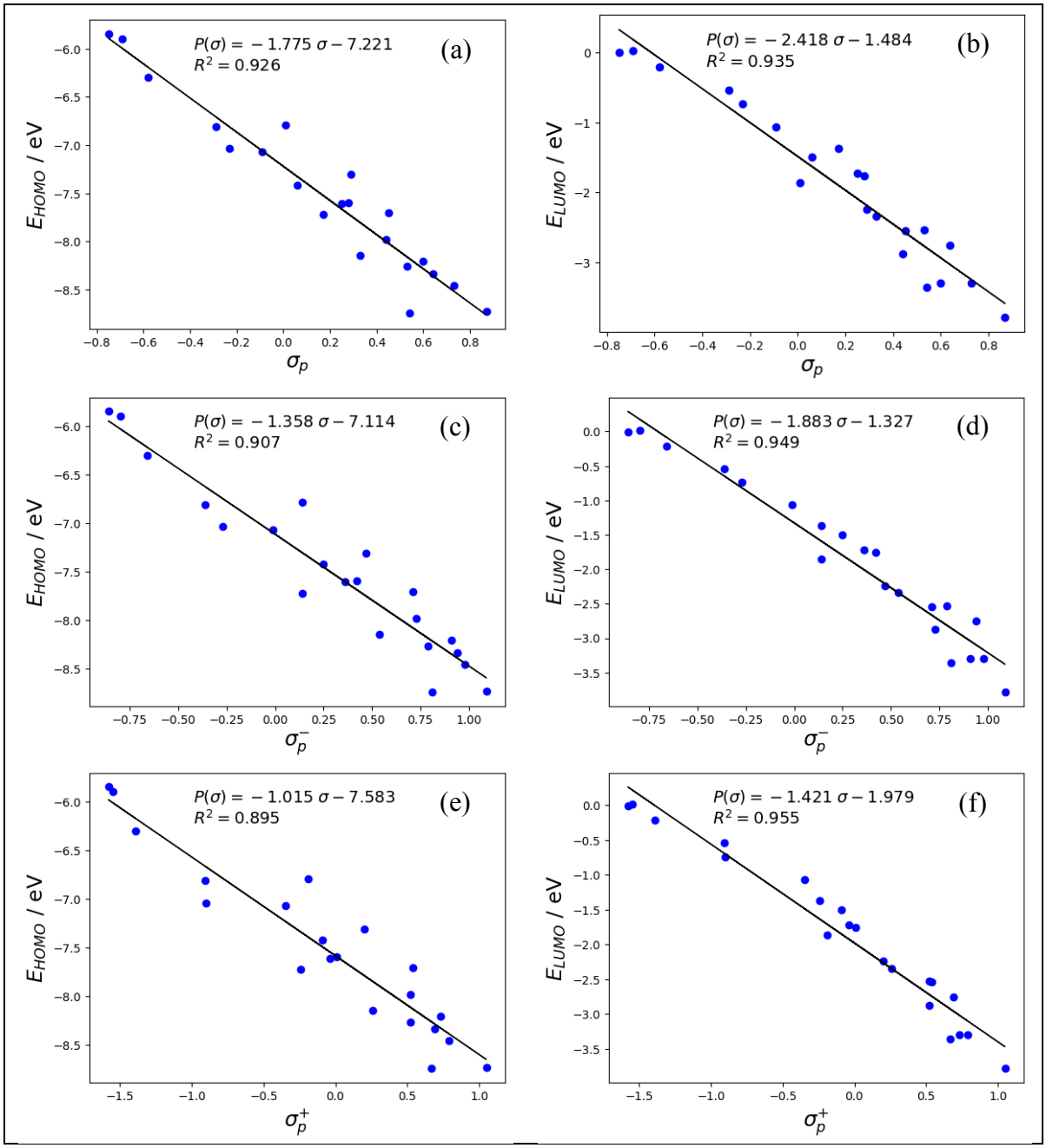

Figure 5. Best correlations between ML-based Hammett's constants and DFT FMO energies (E_{HOMO} – right panels and E_{LUMO} – left panels) of the di-substitued DPP-X_2 derivatives.

Overall, both DFT and Hartree-Fock (HF) orbitals showed that EDGs increase, in different degrees, the energies of both HOMO and LUMO while EWGs decrease, also in different degrees, the energies of the FMOs – see Figure 6 (DFT) and, Figure S2 for the HF orbitals. This effect is due to the increase or decrease of electron density in the DPP core that depends on the nature of substituent.[64,65]

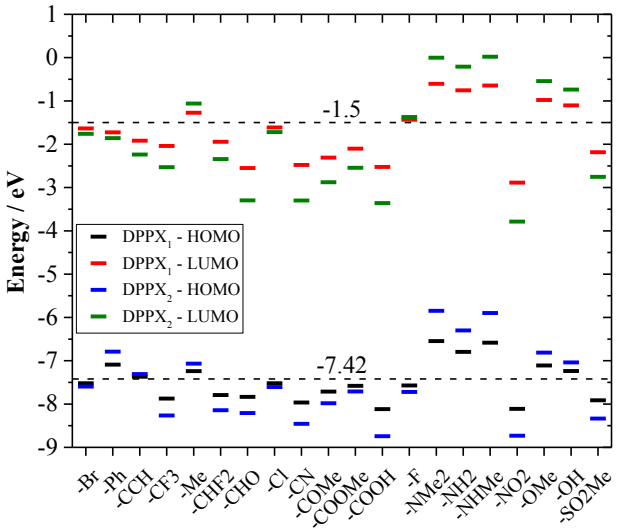

Figure 6. Molecular orbital energies of all DPP-X systems. Calculations at the CAM-B3LYP/Def2-TZVP//B3LYP-D3/Def2-TZVP level. The dashed lines show the HOMO and LUMO energies of the unsubstituted DPP-H system.

The correlation between the ML-based Hammett's constants and the HOMO-LUMO gaps (E_g) was similar for the optical gap (ΔE) – i.e., to the $^1S_0 \rightarrow {}^1S_1$ transition energy, which is expected, given that the former is an approximation for the latter. We will then focus on the values of E_g only to examine these correlations. When E_g values are compared with the value of the unsubstituted DPP-H, decreases occur to -1.58 eV for most EWGs and increases up to 0.72 eV for the EDGs (Table S5).

Furthermore, both at the DFT and HF levels (ADC(2) for the E_g values), for every decrease (or increase) of the E_g value of the mono-substituted DPP-X_1 systems, a more significant decrease (or increase) in the value of the gap is observed for the DPP-X_2 systems (see Figure 7 below), which shows the intensification of the electronic effects when more than one substituent group is present. Note that for both methods, the variations of E_g follow the same trend.

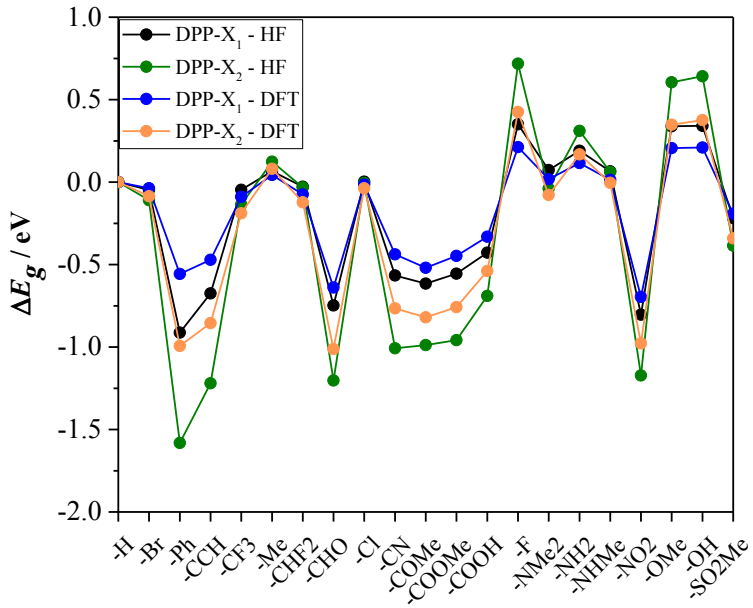

Figure 7. HOMO-LUMO gap (E_g), in eV, of all mono-substituted DPP-X systems. Calculations at the CAM-B3LYP/Def2-TZVP and ADC(2)/Def2-TZVP//B3LYP-D3/Def2-TZVP levels.

Adiabatic ionization potential (IP_A) and electro affinity (EA_A). The results regarding IP_A and EA_A show a direct correlation with different σ's. Figure 8 illustrates the best correlations for DPP-X_1 and DPP-X_2, while Figures S5-S6 display the remaining correlations. All values of IP_A and EA_A are collected in Table S6 of the SI.

The ionization potential and electron affinity are related to hole and electron transport, respectively, and thus are essential for the design and operation of photovoltaic devices.[66] It is observed that almost all the σ values (except σ_I) are directly proportional to both IP_A and EA_A. As the EDGs decrease their sigma values, both IP_A and EA_A, and the EWGs sigmas, increase along with both IP_A and EA_A, with the nitro group (the electron withdrawal group with the largest Hammett constant) having the largest of IP_A and EA_A values. This is expected given that the HOMO of the EDGs interacts with the HOMO of the DPP-X, destabilizing the whole system. At the same time, the interaction between the LUMO of the EWGs and the HOMO of the DPP-X stabilizes it.[67] Therefore, as it will be discussed in the next subsection, for the PCE, V_{OC}, and FF properties, the higher the σ values (thus, the higher IP_A and EA_A), the better the performance of DPP-based OPVs since these groups show the lowest values of E_{HOMO}.

A favorable driving force in a BHJ appears when a D-material (DPP-X) donates an electron to the A-material (e.g., $PC_{61}BM$), avoiding electron-hole recombination.[68] This can be achieved with a lower IP_A coupled to a HOMO of the D-material higher in energy than in the A-material, which facilitates hole transport. With a higher EA_A, the LUMO of the D-material is higher than that of the A-material, which avoids electron-hole recombination.[52] These interesting results suggest that Hammett's theory can guide the appropriate choice of substituent group (one that could balance a low IP_A and a high EA_A) in a DPP-based system to favor both hole and electron transport.

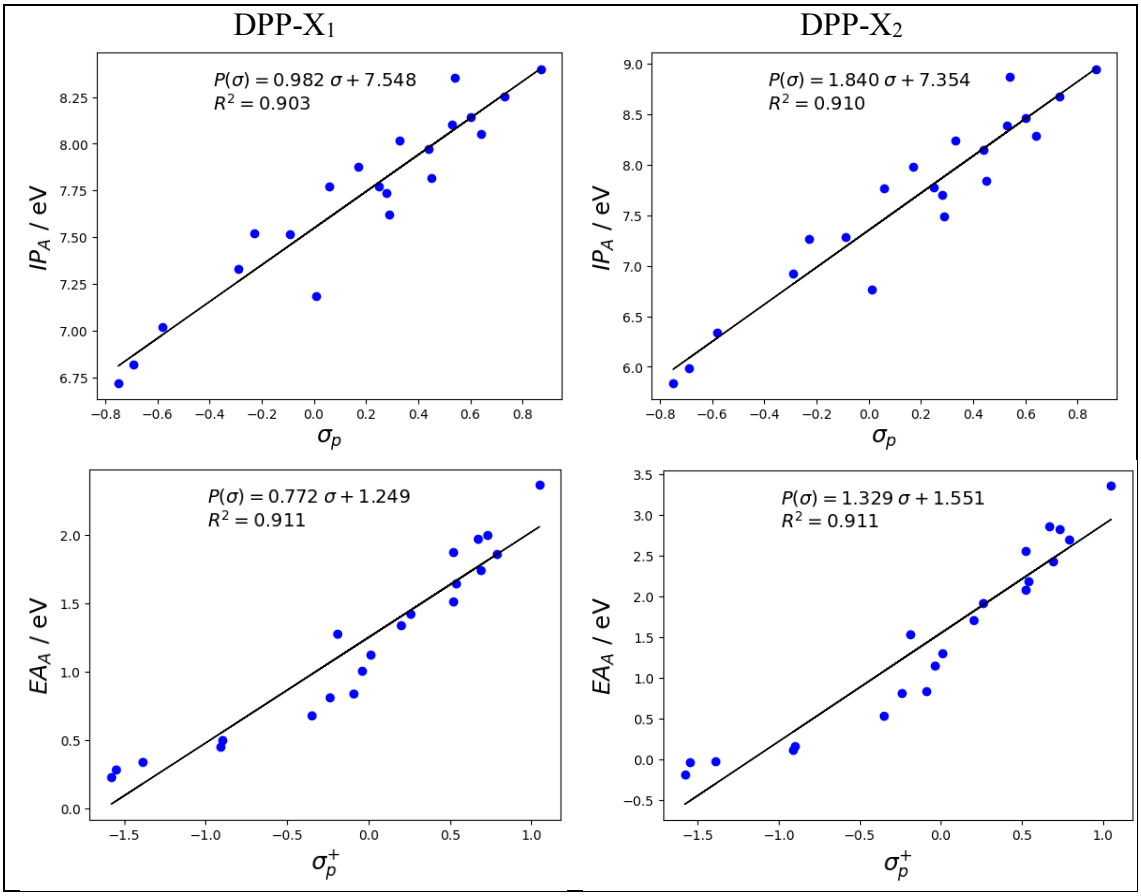

Figure 8. Best correlations between the adiabatic ionization potentials (IP_A) and electron affinities (EA_A), in eV, and the ML-based Hammett's constants. Calculations at the CAM-B3LYP/Def2-TZVP//B3LYP-D3/Def2-TZVP level.

Optical gaps (ΔE) and maximum absorption wavelength (λ_{max}). We computed singlet-singlet transition energies between the ground and the first excited states ($^1S_0 \rightarrow {}^1S_1$), and the ΔE values are the optical gaps. The $S_0 \rightarrow {}^1S_1$ is a HOMO-LUMO transition. Tables S7-S8 in the SI show the maximum absorption wavelength (λ_{max}), and the oscillator strengths of the optical gap (f), computed by the DFT and HF methods.

Concerning the correlations between the optical gap, ΔE (or, equivalently, λ_{max}), and the ML-based Hammett's parameters, most of the plots indicate that some σ constants (σ_p^-, σ_p^+, and σ_R) exponentially correlate with the values of the transitions energies, as seen before for PPy-based systems.[25] Therefore, we conclude that resonance effects play a significant role in the determination of these properties of DPP-X systems. Figure 9 depicts the best correlations between ΔE and the ML-based σ_R constants. The higher the σ values, the lower the transition energies, which leads to redshifts in the absorption spectra discussed below. This behavior indicates that improved visible absorption is achieved by employing EWGs with higher σ values and/or substituents that resonate with the DPP core, such as the phenyl group. The correlation between the HOMO-LUMO gap (E_g) and the σ constants show the same trends seen in Figure 9; however, the only difference is the absolute value of the energy gap.

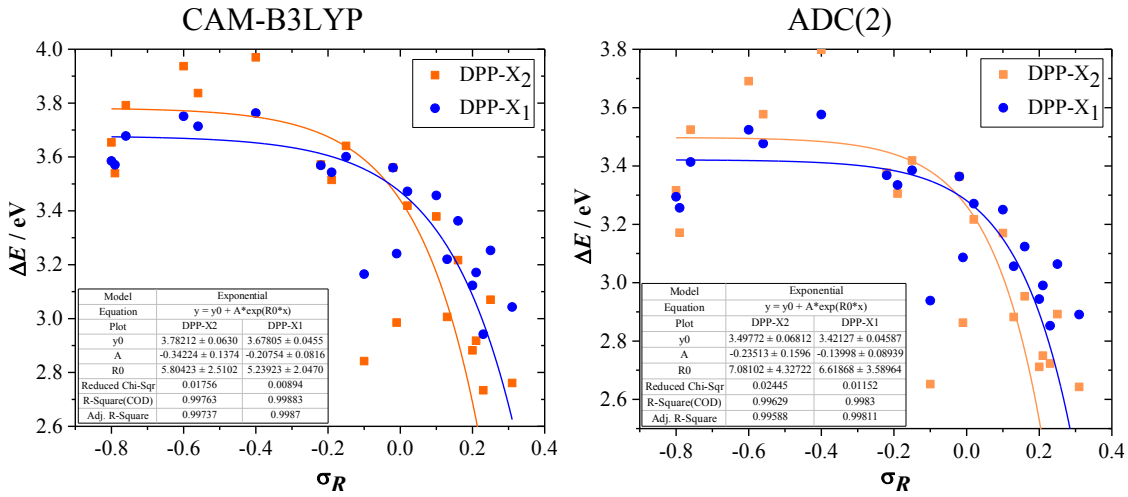

Figure 9. Correlations between ML-based Hammett's parameters and the first transition energy (ΔE), the optical gap.

Figure 10 plots ΔE and f of the first electronic transition of the absorption spectra, calculated at both the TDDFT and ADC(2) methods. The plots show a noticeable increase in the values of f for the di-substituted DPP-X_2 systems compared to the mono-substituted DPP-X_1 due to a synergetic influence of the substituents onto the whole system when more than one group is present. Moreover, one can also see that the presence of two X groups results in larger red and blue shifts compared with the unsubstituted DPP-H system, which appears quite constant between 3.5 and 3.6 eV (CAM-B3LYP) and 3.3. and 3.4 eV (ADC(2)) – see Figure 10.

Although the ADC(2) method overall shows lower values of both f and ΔE compared with CAM-B3LYP, both methods provide the same general trends regarding blue and red shifts, oscillator strength intensities, and the ordering of the X groups.

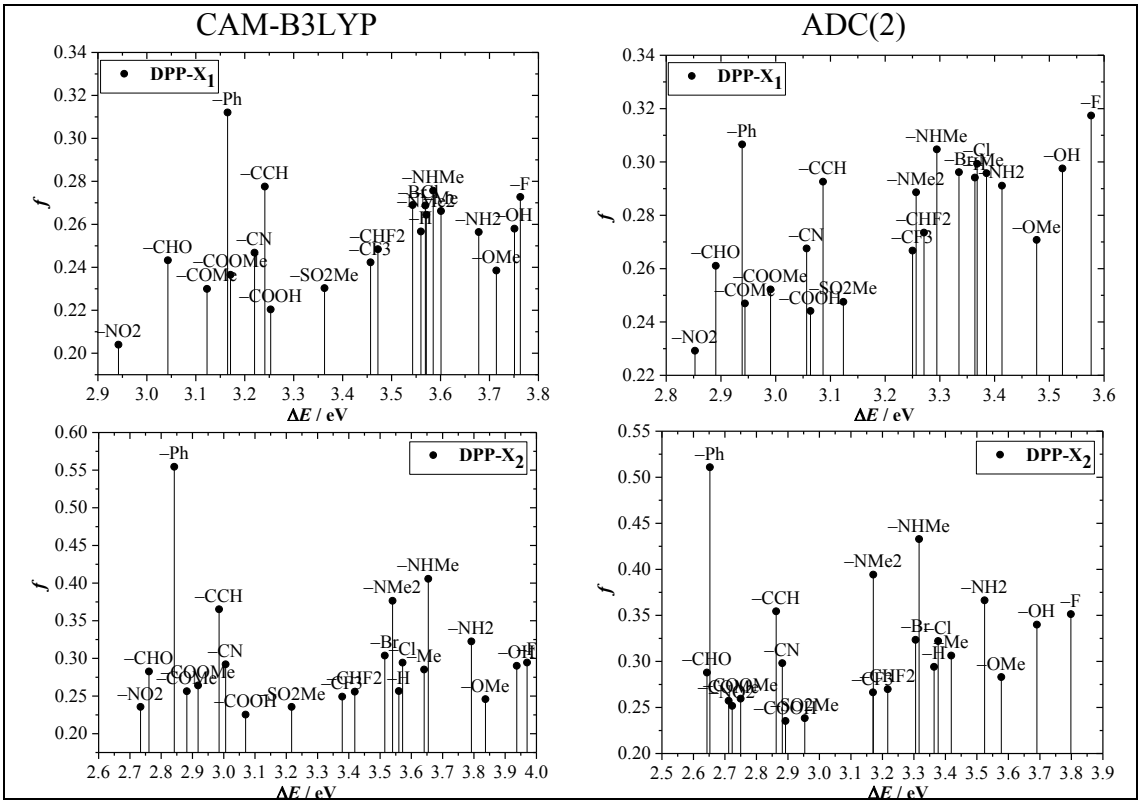

Figure 10. Optical gap (ΔE) and oscillator strength (f) of the first electronic transition energy for both types of DPP-X systems. Calculations at the CAM-B3LYP/Def2-TZVP// and ADC(2)/Def2-TZVP//B3LYP-D3/Def2-TZVP levels.

We found that most DPP-X systems have the maximum absorption in the UV region of the spectrum (3–12 eV). Compared with unsubstituted DPP-H, the systems with redshifts in the visible range of the spectrum (< 3 eV) are DPP-NO_2, DPP-$(CHO)_2$, DPP-Ph, DPP-Ph_2, DPP-$(COMe)_2$, DPP-$(COOMe)_2$, DPP-$(CCH)_2$, DPP-COOH, and DPP-$(COOH)_2$, most of which are typical EWGs except for the phenyl group. The largest red shifts were -0.62 eV (DPP-NO_2, TDDFT), -0.83 eV (DPP-$(NO_2)_2$, TDDFT), -0.65 eV (DPP-COMe, ADC(2)), and -0.72 eV (DPP-$(CHO)_2$, ADC(2)). For the TDDFT and ADC(2) methods, appreciable f values between 0.20 and 0.55 (see Tables S7-S8) were found for all mono- and di-substituted DPP-X.

These results indicate that the substituents capable of withdrawing electrons (for instance, through resonance) from a DPP core can redshift the absorption spectra to the visible range of the spectrum. A similar result was obtained for di-functionalized DPP systems containing different aromatic groups. DPP thus shows potential for application in OPVs due to shifts in the spectrum towards longer wavelengths.[6]

Power conversion efficiency (PCE), open-circuit voltage (V_{OC}), and fill factor (FF). Three of the essential parameters for designing photovoltaic systems are the PCE, FF, and V_{OC}, which allows the evaluation of the potential application in an OPV device. By analyzing the relationship between PCE, V_{OC}, , and other factors, such as material properties, including the ones via Hammett's constants, device architecture, and processing conditions, researchers can optimize the design and fabrication of photovoltaic devices to achieve higher efficiency. Additionally, by using theoretical models to

examine experimental data and predict the behavior of new materials before the synthesis, time and resources can be saved by identifying the most promising candidates for further study.69

According to our calculations, the values of V_{OC} can be directly related to the different σ constants; the best results are in Figure 11. Similar to the results involving IP_A and EA_A, larger values of a given σ lead to larger values of V_{OC}. This is due to the direct proportionality between V_{OC} and the absolute value of the HOMO energies of D-materials (E_{HOMO}) – see Eq. (8).

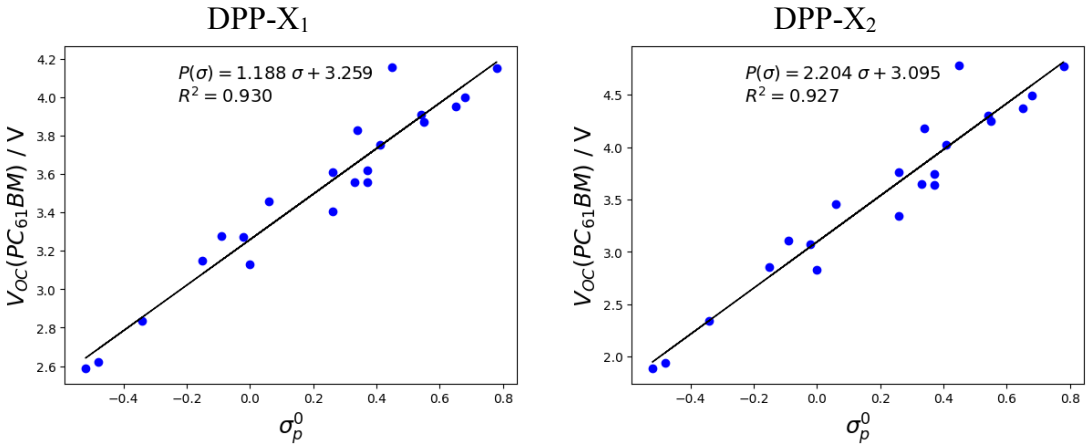

Figure 11. Best correlations between the open-circuit voltage (V_{OC}) and ML-based Hammett's constants. Calculations at the CAM-B3LYP/Def2-TZVP//B3LYP-D3/Def2-TZVP level and considering $PC_{61}BM$ as the acceptor material.

Related to the open-circuit voltage through Eq. (9), the fill factor (FF) measures how efficiently the free carriers can be removed from the photovoltaic device into the external circuit. Additionally, from a typical current-voltage (J-V) curve, one can also calculate FF by dividing the product of the maximum current (J_m) and voltage (V_m) – values for which the device's maximum power can be obtained, by the product of J_{SC} and V_{OC}.[70] Recently, FF values up to ~83% have been achieved in perovskite solar cells.[71–73] This is larger than previously reported for organic solar cells (OSCs), which have FF usually ranging between 50 and 70%.[70] Here, the calculated FF for the DPP-X systems by Eq. (9) showed rather high values of FF up to 98% (see Table S9 in the SI).

The PCE values are functions of V_{OC} (and consequently, E^D_{HOMO} of the DPP-X systems) according to Eq. (8), of FF calculated by Eq. (9), and the short-circuit current (J_{SC}) constant. The correlations between the ML-based Hammett's constants and the PCE values for a theoretical DPP-based OPV device illustrate the power of our approach based on Hammett's theory for designing D-materials in organic electronics. Once again, the σ_p^0 showed the best results, which are illustrated in Figure 12. The correlations with other σ are depicted in Figure S9 in the SI. Similarly to the previous analysis of V_{OC}'s, higher values of σ lead to higher values PCE (see Figure 4, Figure 5, and Figure 11).

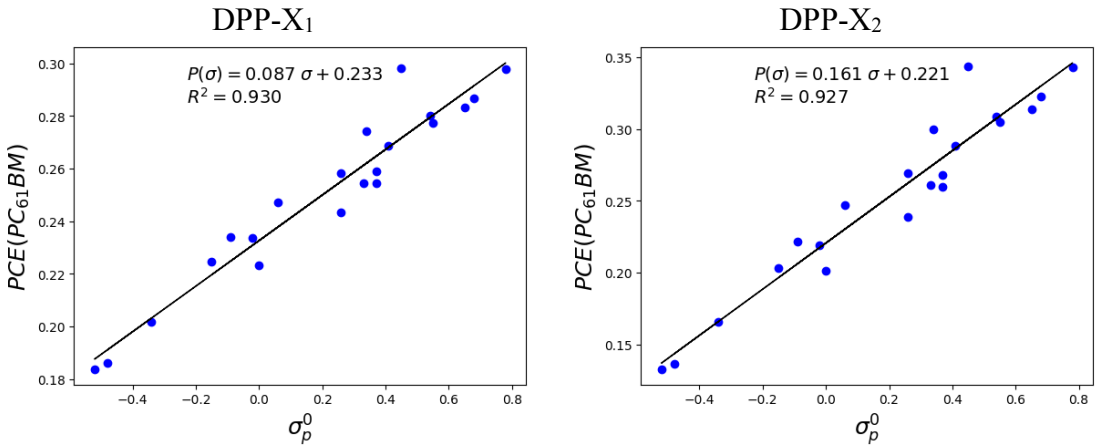

Figure 12. Best correlations between the power conversion efficiency (PCE) and the ML-based σ_p^0 Hammett's constants. Calculations at the CAM-B3LYP/Def2-TZVP//B3LYP-D3/Def2-TZVP level and considering $PC_{61}BM$ as the acceptor (A)-material.

Exciton binding energy (E_{bind}) analysis. The binding energy E_{bind} of an exciton is an essential parameter for the operation of organic photovoltaic devices[74] since it determines how efficiently a bonded hole and an electron dissociate. Therefore, in a donor molecule for an OPV we are looking for the smallest value of E_{bind}. From the plots of Figure 13, we see three amine groups –NH_2, –NHMe, and –NMe_2, which bear the most significant negative Hammett's constants and are EDGs, have the smallest exciton binding energies. Hence, these EDG substituents are the ones most favoring a small E_{bind}. This is valid for the mono- and di-substituted DPPs.

From the inspection of the plots, correlations between the ML-based Hammett's constants and the computed exciton binding energies (E_{bind}) of the DPP-X systems are not apparent. Despite that, interesting insights emerge from clustering analysis.

We used a machine learning technique, the *MeanShift* algorithm,[75] to investigate the eventual clustering of the E_{bind} values and Hammett's constants. The algorithm identifies the number of clusters in a particular data set. Figure 13 shows examples of these clustering for the ML-based σ_R, σ_p^0, and σ_p^- Hammett's constants. The centroids (the geometric center of the region) of each distribution are also indicated. The other plots are depicted in Figure S10 of the SI.

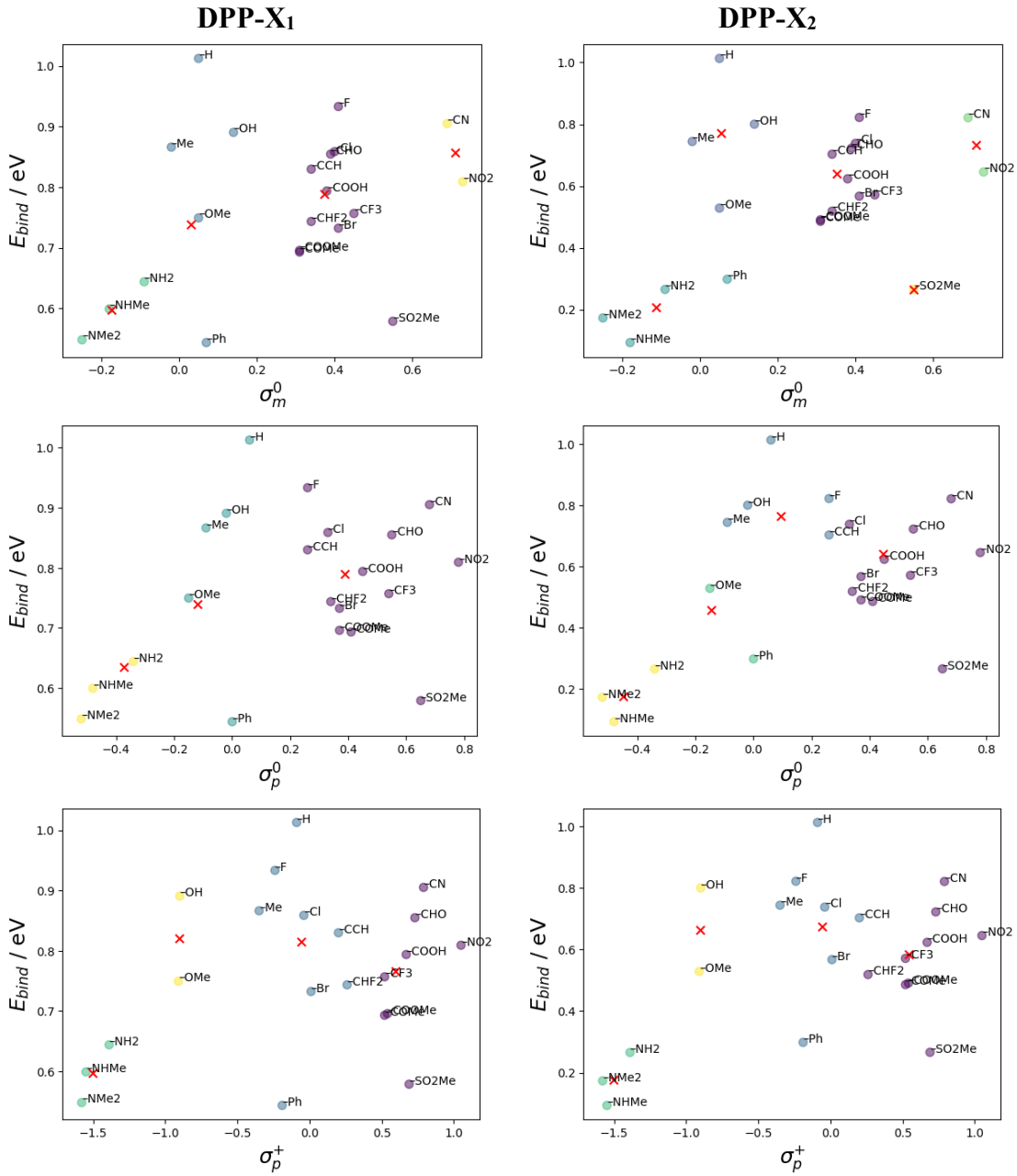

Figure 13. Machine learning analysis employing the *MeanShift* algorithm of exciton binding energies (E_{bind}) correlated with the ML-based Hammett's constants for all DPP-X systems. The centroids of each cluster are indicated by a red cross.

The *MeanShift* clustering algorithm is an unsupervised machine learning algorithm that locates arbitrarily shaped regions in the samples of a data set with uniform density. The number of clusters is not pre-set as in the more popular ML clustering algorithms (e.g., K-means).[76] Candidates for the centroids of each region are constantly updated to be the averages of the points in the given regions until convergence is reached. A final set of centroids is formed and filtered in a post-processing step to eliminate near duplicates.[75]

For all types of Hammett's constants, three to five clusters of substituents were found (Figure 13 and Figure S10). Table 2 e Table 3 list the centroid coordinates of all

Hammett's constants clusters depicted in both figures. A centroid characterizes the center of a region, and indicate an average electron-donating or -withdrawing behavior of a group of substituents selected by the algorithm as having similar properties. The centroids show then an average Hammett constant associated to an average binding energy of the exciton. Therefore, these pairs of values could be used for searching or choosing substituents for optimizing (i.e., reducing) exciton binding energies of DPP-based materials. This kind of analysis would undoubtedly be even more helpful for much larger data sets.

Table 2. Coordinates of all centroid calculated be the *MeanShift* algorithm for the DPP-X_1 systems.

σ_m	σ_p	σ_R	σ_I	σ_p^+	σ_p^-	σ_m^0	σ_p^0
0.36, 0.79	0.46, 0.77	0.16, 0.77	0.33, 0.78	0.60, 0.77	0.73, 0.77	0.37, 0.79	0.39, 0.79
0.03, 0.74	-0.08, 0.89	-0.45, 0.86	0.51, 0.84	-0.06, 0.82	0.29, 0.82	0.03, 0.74	-0.12, 0.74
-0.19, 0.60	-0.58, 0.64	-0.73, 0.64	0.15, 0.62	-1.51, 0.60	-0.13, 0.86	-0.17, 0.60	-0.37, 0.64
0.70, 0.86	0.01, 0.54		0.07, 0.94	-0.91, 0.82	-0.67, 0.64	0.71, 0.86	

Table 3. Coordinates of all centroid calculated be the *MeanShift* algorithm for the DPP-X_2 systems.

σ_m	σ_p	σ_R	σ_I	σ_p^+	σ_p^-	σ_m^0	σ_p^0
0.32, 0.63	0.50, 0.60	0.07, 0.59	0.42, 0.67	0.54, 0.58	0.73, 0.60	0.35, 0.64	0.45, 0.64
0.05, 0.77	0.13, 0.74	-0.35, 0.70	0.15, 0.30	-0.06, 0.68	-0.13, 0.72	0.06, 0.77	0.09, 0.76
-0.12, 0.21	-0.67, 0.18	-0.73, 0.27	0.48, 0.27	-1.51, 0.18	-0.67, 0.27	-0.11, 0.21	-0.15, 0.46
0.70, 0.73			0.07, 1.01	-0.91, 0.66		0.71, 0.73	-0.45, 0.18
0.52, 0.27						0.55, 0.27	

3.3. Intramolecular charge transfer (ICT) analysis

The electron density flows were quantified by the CT, PR, and COH descriptors and exciton analysis. For this analysis, the DPP-X systems are divided into three fragments: two consisting of the X substituent groups (for the DPP-X_1 systems, one of these fragments is a hydrogen atom) and one central fragment consisting of the DPP core – see Figure 2. Tables S10-S14 list all the values obtained from the CAM-B3LYP and ADC(2) methods.

According to the results, the CT numbers increase to $0.39e$ (DPP-CHO, TDDFT), $0.46e$ (DPP-$(NO_2)_2$, TDDFT), $0.35\ e$ (DPP-Ph, ADC(2)), and $0.51e$ (DPP-Ph_2, ADC(2)) in comparison with the DPP-H system, which has $CT = 0.02e$ at both TDDFT and ADC(2)

levels. This suggests that for every DPP-X system, the electron excitations would increase (even if small) ICT processes, with electron transfer occurring from the DPP core to the X groups. Similarly to the analysis of IP_A and EA_A, this could be due to EDGs destabilizing the HOMO of the DPP core and, thus, facilitating the ICT process toward the substituent group. At the same time, EWGs stabilize the HOMO of the DPP core through an interaction with the LUMO of the substituent, as in previous situations discussed above.

The PR and COH descriptors vary according with the CT number. PR numbers show the maximum values of 1.87 (DPP-CHO, TDDFT), 2.67 (DPP-$(CHO)_2$, TDDFT), 1.47 (DPP-Ph, ADC(2)), and 1.96 (DPP-Ph_2, ADC(2)). As for the COH numbers, all substituents have $COH \approx PR$ (see Tables S10-S14), which suggests that the hole-electron separation is rather large; in other words, the excitons are weakly bonded. These results confirm the conclusions from the CT number analysis, namely that the initial and final orbital of a one-electron excitation are generally located in different molecule fragments. Different substituted naphthalimides and naphthalic anhydrides [77] and phenyl-substituted *p*-dimethylaminobenzoates [78] also have properties related to ICT transitions that are significantly affected by the electron donating/withdrawing strengths of the substituents. In photoactive molecules, ICT transitions are essential for applications of such molecules in optoelectronic devices.[79,80] Here, we found that, by adding especially EWGs or substituent groups able to resonate with the DPP core, \facilitates ICT. Moreover, transition density matrix (TDM) analysis with the Multiwfn software [81] was carried out on all DPP-X systems to corroborate the previous ICT analysis. The TDM technique allow the evaluation of solar cells' efficiencies by calculating charge carriers (i.e., holes and electrons), the interaction between moieties with different chemical natures (EDG or EWG), and electronic excitation.[82,83]

The TDM plots of all our DPP-X systems can be seen in Figure S13 of the SI. An example of the results obtained from this analysis is in Figure 14. In these figures, the *y*-axis and *x*-axis represent the electron and the hole positions, respectively. The color shift from blue to red characterizes the intensity (minimum to maximum) of the electron transfer between different regions in the molecule. The DPP core is represented by the numbers 1-10 in both axes.

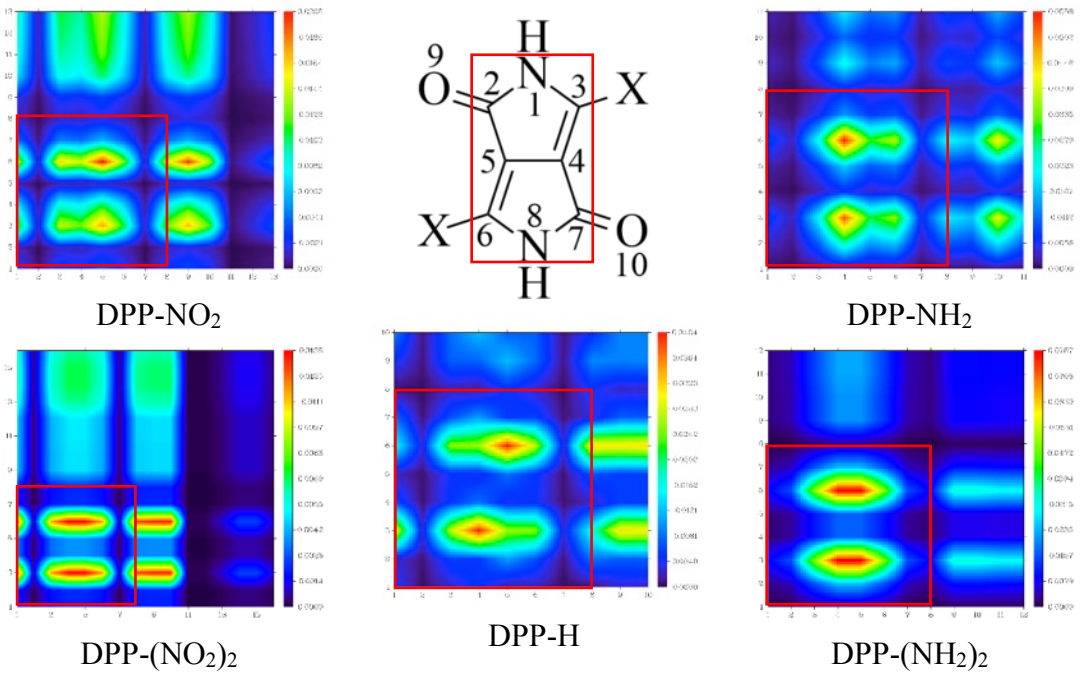

Figure 14. Transition density matrix (TDM) plots for DPP-H, DPP-$(NO_2)_{1,2}$, and DPP-$(NH_2)_{1,2}$. Calculations at the CAM-B3LYP/Def2-TZVP//B3LYP-D3/Def2-TZVP level.

The TDM results show trends similar to the ones discussed above, i.e., the charge density flows from the center of the DPP core, more precisely from carbon atoms 4 and 5 that are common to both rings (Figure 1) to the carbon atoms bonded to the substituent groups (atoms number 3 and 6, Figure 1). Moreover, the charge distribution is more symmetrical for the di-substituted rather than the mono-substituted ones. This is expected considering the higher symmetry of DPP-X_2 systems.

4. Conclusions

In this paper, we investigated correlations between Machine Learning (ML)-based Hammett's constants (σ)[36] and different opto- and electrochemical properties of mono- and di-substituted diketopyrrolopyrrole (DPP) derivatives. Density Functional Theory (DFT), Time-Dependent DFT (TDDFT), and the ab initio second-order algebraic diagrammatic construction (ADC(2)) method were employed. In addition to a fundamental understanding of these systems, we aimed to develop theoretical approaches to develop molecular-engineered DPP-based systems as donor (D)-materials in organic photovoltaics (OPV). Our approach is based on Hammett's theory, which quantifies the donor or withdrawal power of substituents via a numerical constant.

A set of 20 common substituent groups (X) and two groups of functionalized DPP-X systems were modeled. These included mono- (DPP-X_1) and di-substituted (DPP-X_2) systems. Several OPV propertieswere investigated, including the power conversion efficiency (PCE), open-circuit voltage (V_{OC}), HOMO and LUMO energies (E_{HOMO} and E_{LUMO}), HOMO-LUMO gaps (E_g), energies of the first singlet excited states (ΔE) – optical gap , adiabatic ionization potentials (IP_A), and electron affinities (EA_A). We found relevant correlations with the σ_m, σ_p, σ_m^0, σ_p^0, σ_p^+, σ_p^-, and σ_R constants, in particular, that

the higher the values of Hammett's constants, the higher the values of PCE and V_{OC}. The induction constant σ_I did not have a reasonable correlation with the investigated properties, which indicates that resonance is the main factor driving electron donation or withdrawal in DPP derivatives.

The presence of more than one substituent group modified all optoelectronic properties, indicating a synergistic effect when there are two substituents. For every decrease (increase) of the values of a given optoelectronic property in DPP-X_1, a greater decrease (increase) in the same values was observed in DPP-X_2. This larger influence of additional substituent groups bonded to the DPP core also displayed relevant correlations the Hammett's constants.

For the exciton binding energies for which correlations were not apparent correlations, the unsupervised machine learning algorithm *MeanShift* identified three to five clusters of Hammett's constants. Each region is characterized by a centroid, an average pair of values (binding energy, Hammett constant), which can guide the search for substituents that can optimize the exciton binding energies of a donor molecule in an OPV. This approach should be even more helpful for large data sets and similar situations of an apparent lack of correlations.

In this work, we presented a general systematic theoretical approach employing Machine Learning (ML)-based Hammett's constants (σ) for evaluating a large set of electronic properties relevant for assessing donor molecules in organic electronic devices including OPVs. This approach that extended previous works, was applied here to mono and -di-substituted DPPs shows potential for assessing and developing new materials for organic electronics.

Data availability statement

Data are available upon friendly request.

Supporting Information availability statement

The Supporting Information (SI) contains: Hammett's constants from the literature; cartesian coordinates of all DPP-X systems; substituent structures; MO energies; excitation energies, oscillator strengths, and orbital assignments; calculated photovoltaic properties of DPP-X systems; calculated intramolecular charge transfer properties; and transition density matrices of the DPP-X systems.

Authors contributions

Gabriel Monteiro-de-Castro – Data curation; Formal Analysis; Investigation; Methodology; Software; Validation; Visualization; Writing - original draft; Writing & editing.

Itamar Borges Jr. – Conceptualization; Data curation; Formal Analysis; Funding Acquisition; Methodology; Project administration; Resources; Supervision; Validation; Visualization; Writing, review & editing.

Conflicts of Interest

There are no conflicts of interest to declare.

Acknowledgments

I. B. thanks the Brazilian agencies CNPq (Grant numbers 304148/2018-0 and 409447/2018-8) and FAPERJ (Grant number E26/201. 197/2021) for support. G. M. C thanks CAPES for a Ph. D. fellowship.

References

[1] I. Borges, A. J. A. Aquino, A. Köhn, R. Nieman, W. L. Hase, L. X. Chen and H. Lischka, Ab Initio Modeling of Excitonic and Charge-Transfer States in Organic Semiconductors: The PTB1/PCBM Low Band Gap System, *J Am Chem Soc*, 2013, **135**, 18252–18255.

[2] L. Modesto-Costa, I. Borges, A. J. A. Aquino and H. Lischka, Electronic structure theory gives insights into the higher efficiency of the PTB electron-donor polymers for organic photovoltaics in comparison with prototypical P3HT, *J Chem Phys*, 2018, **149**, 184905.

[3] I. Borges, E. Uhl, L. Modesto-Costa, A. J. A. Aquino and H. Lischka, Insight into the Excited State Electronic and Structural Properties of the Organic Photovoltaic Donor Polymer Poly(thieno[3,4- *b*]thiophene benzodithiophene) by Means of *ab Initio* and Density Functional Theory, *The Journal of Physical Chemistry C*, 2016, **120**, 21818–21826.

[4] D. G. Farnum, G. Mehta, G. G. I. Moore and F. P. Siegal, Attempted reformatskii reaction of benzonitrile, 1,4-diketo-3,6-diphenylpyrrolo[3,4-C]pyrrole. A lactam analogue of pentalene., *Tetrahedron Lett*, 1974, **15**, 2549–2552.

[5] S. Qu and H. Tian, Diketopyrrolopyrrole (DPP)-based materials for organic photovoltaics, *Chemical Communications*, 2012, **48**, 3039.

[6] R. Jin, X. Zhang, W. Xiao and A. Irfan, Rational design of diketopyrrolopyrrole-based multifunctional materials for organic light-emitting diodes and organic solar cells, *Theor Chem Acc*, 2018, **137**, 145.

[7] J. C. Bijleveld, A. P. Zoombelt, S. G. J. Mathijssen, M. M. Wienk, M. Turbiez, D. M. de Leeuw and R. A. J. Janssen, Poly(diketopyrrolopyrrole−terthiophene) for Ambipolar Logic and Photovoltaics, *J Am Chem Soc*, 2009, **131**, 16616–16617.

[8] D. C. Young, M. Tasior, A. D. Laurent, Ł. Dobrzycki, M. K. Cyrański, N. Tkachenko, D. Jacquemin and D. T. Gryko, Photostable orange-red fluorescent unsymmetrical diketopyrrolopyrrole-BF2hybrids, *J Mater Chem C Mater*, 2020, **8**, 7708–7717.

[9] L. Wang, L. Zhu, L. Li and D. Cao, Tetraphenylethene-functionalized diketopyrrolopyrrole solid state emissive molecules: Enhanced emission in the solid state and as a fluorescent probe for cyanide detection, *RSC Adv*, 2016, **6**, 55182–55193.

[10] W. Li, L. Wang, H. Tang and D. Cao, Diketopyrrolopyrrole-based fluorescent probes for detection and bioimaging: Current progresses and perspectives, *Dyes and Pigments*, 2019, **162**, 934–950.

[11] Y. Gao, G. Feng, T. Jiang, C. Goh, L. Ng, B. Liu, B. Li, L. Yang, J. Hua and H. Tian, Biocompatible Nanoparticles Based on Diketo-Pyrrolo-Pyrrole (DPP) with

Aggregation-Induced Red/NIR Emission for In Vivo Two-Photon Fluorescence Imaging, *Adv Funct Mater*, 2015, **25**, 2857–2866.

[12] W. Kohn and L. J. Sham, Self-consistent equations including exchange and correlation effects, *Physical Review*, 1965, **140**, A1133–A1138.

[13] E. Runge and E. K. U. Gross, Density-Functional Theory for Time-Dependent Systems, *Phys Rev Lett*, 1984, **52**, 997–1000.

[14] L. Zhang, W. Shen, R. He, X. Liu, X. Tang, Y. Yang and M. Li, Fine structural tuning of diketopyrrolopyrrole-cored donor materials for small molecule-fullerene organic solar cells: A theoretical study, *Org Electron*, 2016, **32**, 134–144.

[15] R. Jin, X. Zhang and W. Xiao, Theoretical Studies of Photophysical Properties of D−π−A−π−D-Type Diketopyrrolopyrrole-Based Molecules for Organic Light-Emitting Diodes and Organic Solar Cells, *Molecules*, 2020, **25**, 667.

[16] P. Murugan, V. Raghavendra, S. Chithiravel, K. Krishnamoorthy, A. B. Mandal, V. Subramanian and D. Samanta, Experimental and Theoretical Investigations of Different Diketopyrrolopyrrole-Based Polymers, *ACS Omega*, 2018, **3**, 11710–11717.

[17] J. Dhar, D. P. Karothu and S. Patil, Herringbone to cofacial solid state packing via H-bonding in diketopyrrolopyrrole (DPP) based molecular crystals: influence on charge transport, *Chemical Communications*, 2015, **51**, 97–100.

[18] L. P. Hammett, Some Relations between Reaction Rates and Equilibrium Constants, *Chem Rev*, 1935, **17**, 125–136.

[19] L. P. Hammett, The effect of structure upon the reactions of organic compounds. temperature and solvent influences, *J Chem Phys*, 1937, **4**, 613–617.

[20] F. A. Carey and R. J. Sundberg, *Advanced Organic Chemistry*, Springer US, Boston, MA, 5th edn., 1990.

[21] C. D. Johnson, *The Hammett Equation*, Cambridge University Press, Cambridge, 1973.

[22] C. Hansch, A. Leo and R. W. Taft, A survey of Hammett substituent constants and resonance and field parameters, *Chem Rev*, 1991, **91**, 165–195.

[23] L. Echegoyen, Modern Physical Organic Chemistry, *J Phys Org Chem*, 2011, **24**, 743–743.

[24] A. Jezuita, K. Ejsmont and H. Szatylowicz, Substituent effects of nitro group in cyclic compounds, *Struct Chem*, 2021, **32**, 179–203.

[25] A. P. Coleone, L. G. Lascane and A. Batagin-Neto, Polypyrrole derivatives for optoelectronic applications: A DFT study on the influence of side groups, *Physical Chemistry Chemical Physics*, 2019, **21**, 17729–17739.

[26] N. B. Chapman and J. Shorter, Eds., *Advances in Linear Free Energy Relationships*, Springer US, Boston, MA, 1st edn., 1972.

[27] C. Hansch and A. Leo, *Exploring QSAR.: Fundamentals and applications in chemistry and biology*, American Chemical Society, Washington, 1995, vol. 1.

[28] J. D. Roberts and W. T. Moreland, Electrical Effects of Substituent Groups in Saturated Systems. Reactivities of 4-Substituted Bicyclo [2.2.2]octane-1-carboxylic Acids 1, *J Am Chem Soc*, 1953, **75**, 2167–2173.

[29] H. C. Brown and Y. Okamoto, Substituent Constants for Aromatic Substitution 1-3, *J Am Chem Soc*, 1957, **79**, 1913–1917.

[30] N. B. Chapman and J. Shorter, *Correlation Analysis in Chemistry*, Springer US, Boston, MA, 1978.

[31] H. C. Brown and Y. Okamoto, Electrophilic Substituent Constants, *J Am Chem Soc*, 1958, **80**, 4979–4987.

[32] Y. Yukawa and Y. Tsuno, Resonance Effect in Hammett Relationship. II. Sigma Constants in Electrophilic Reactions and their Intercorrelation, *Bull Chem Soc Jpn*, 1959, **32**, 965–971.

[33] Y. Yukawa and Y. Tsuno, Resonance Effect in Hammett Relationship. III. The Modified Hammett Relationship for Electrophilic Reactions, *Bull Chem Soc Jpn*, 1959, **32**, 971–981.

[34] E. v. Anslyn and D. A. Dougherty, *Modern Physical Organic Chemistry*, University Science Books, Sausalito, CA, 2005.

[35] H. van Bekkum, P. E. Verkade and B. M. Wepster, A simple re-evaluation of the hammett ϱσ relation, *Recueil des Travaux Chimiques des Pays-Bas*, 1959, **78**, 815–850.

[36] G. Monteiro-de-Castro, J. C. Duarte and I. Borges, Machine learning determination of new Hammett's constants for meta-and para-substituted benzoic acid derivatives employing quantum chemical atomic charge methods, , DOI:10.26434/chemrxiv-2023-5g6gp.

[37] F. L. Hirshfeld, Bonded-atom fragments for describing molecular charge densities, *Theor Chim Acta*, 1977, **44**, 129–138.

[38] V. Nikolova, Di. Cheshmedzhieva, S. Ilieva and B. Galabov, Atomic Charges in Describing Properties of Aromatic Molecules, *J Org Chem*, 2019, **84**, 1908–1915.

[39] A. Yett and P. R. Rablen, A G4 approach to computing the Hammett substituent constants σ_p, σ_m, σ^-, σ^+, and σ^+_m, *J Phys Org Chem*, , DOI:10.1002/poc.4436.

[40] F. Neese, Software update: The ORCA program system—Version 5.0, *WIREs Computational Molecular Science*, 2022, **12**, 1–15.

[41] M. D. Hanwell, D. E. Curtis, D. C. Lonie, T. Vandermeersch, E. Zurek and G. R. Hutchison, Avogadro: an advanced semantic chemical editor, visualization, and analysis platform, *J Cheminform*, 2012, **4**, 17.

[42] T. A. Halgren, MMFF VI. MMFF94s option for energy minimization studies, *J Comput Chem*, 1999, **20**, 720–729.

[43] T. A. Halgren, Merck molecular force field. I. Basis, form, scope, parameterization, and performance of MMFF94, *J Comput Chem*, 1996, **17**, 490–519.

[44] P. J. Stephens, F. J. Devlin, C. F. Chabalowski and M. J. Frisch, Ab Initio Calculation of Vibrational Absorption and Circular Dichroism Spectra Using Density Functional Force Fields, *J Phys Chem*, 1994, **98**, 11623–11627.

[45] S. Grimme, J. Antony, S. Ehrlich and H. Krieg, A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu, *Journal of Chemical Physics*, , DOI:10.1063/1.3382344.

[46] T. Yanai, D. P. Tew and N. C. Handy, A new hybrid exchange–correlation functional using the Coulomb-attenuating method (CAM-B3LYP), *Chem Phys Lett*, 2004, **393**, 51–57.

[47] F. Weigend and R. Ahlrichs, Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: Design and assessment of accuracy, *Physical Chemistry Chemical Physics*, 2005, **7**, 3297.

[48] J. Schirmer, Beyond the random-phase approximation: A new approximation scheme for the polarization propagator, *Phys Rev A (Coll Park)*, 1982, **26**, 2395–2416.

[49] A. B. Trofimov and J. Schirmer, An efficient polarization propagator approach to valence electron excitation spectra, *Journal of Physics B: Atomic, Molecular and Optical Physics*, 1995, **28**, 2299–2324.

[50] S. G. Balasubramani, G. P. Chen, S. Coriani, M. Diedenhofen, M. S. Frank, Y. J. Franzke, F. Furche, R. Grotjahn, M. E. Harding, C. Hättig, A. Hellweg, B. Helmich-Paris, C. Holzer, U. Huniar, M. Kaupp, A. Marefat Khah, S. Karbalaei Khani, T. Müller, F. Mack, B. D. Nguyen, S. M. Parker, E. Perlt, D. Rappoport, K. Reiter, S. Roy, M. Rückert, G. Schmitz, M. Sierka, E. Tapavicza, D. P. Tew, C. van Wüllen, V. K. Voora, F. Weigend, A. Wodyński and J. M. Yu, TURBOMOLE: Modular program suite for ab initio quantum-chemical and condensed-matter simulations, *J Chem Phys*, 2020, **152**, 184107.

[51] H. Sahu and H. Ma, Unraveling Correlations between Molecular Properties and Device Parameters of Organic Solar Cells Using Machine Learning, *Journal of Physical Chemistry Letters*, 2019, **10**, 7277–7284.

[52] H. Roohi and N. Mohtamadifar, The role of the donor group and electron-accepting substitutions inserted in π-linkers in tuning the optoelectronic properties of D-π-A dye-sensitized solar cells: a DFT/TDDFT study, *RSC Adv*, 2022, **12**, 11557–11573.

[53] R. Zaier and S. Ayachi, DFT molecular modeling studies of D-π-A-π-D type cyclopentadithiophene-diketopyrrolopyrrole based small molecules donor

materials for organic photovoltaic cells, *Optik (Stuttg)*, , DOI:10.1016/j.ijleo.2021.166787.

[54] M. C. Scharber, D. Mühlbacher, M. Koppe, P. Denk, C. Waldauf, A. J. Heeger and C. J. Brabec, Design Rules for Donors in Bulk-Heterojunction Solar Cells—Towards 10 % Energy-Conversion Efficiency, *Advanced Materials*, 2006, **18**, 789–794.

[55] M. A. Pan, T. K. Lau, Y. Tang, Y. C. Wu, T. Liu, K. Li, M. C. Chen, X. Lu, W. Ma and C. Zhan, 16.7%-efficiency ternary blended organic photovoltaic cells with PCBM as the acceptor additive to increase the open-circuit voltage and phase purity, *J Mater Chem A Mater*, 2019, **7**, 20713–20722.

[56] ASTM G173-03, Standard Tables for Reference Solar Spectral Irradiances: Direct Normal and Hemispherical on 37° Tilted Surface, *ASTM International*, 2020.

[57] L. Bertoluzzi and S. Ma, On the methods of calculation of the charge collection efficiency of dye sensitized solar cells, *Physical Chemistry Chemical Physics*, 2013, **15**, 4283–4285.

[58] W. Li, K. H. Hendriks, A. Furlan, W. S. C. Roelofs, M. M. Wienk and R. A. J. Janssen, Universal correlation between fibril width and quantum efficiency in diketopyrrolopyrrole-based polymer solar cells, *J Am Chem Soc*, 2013, **135**, 18942–18948.

[59] W. Li, K. H. Hendriks, A. Furlan, W. S. C. Roelofs, S. C. J. Meskers, M. M. Wienk and R. A. J. Janssen, Effect of the fibrillar microstructure on the efficiency of high molecular weight diketopyrrolopyrrole-based polymer solar cells, *Advanced Materials*, 2014, **26**, 1565–1570.

[60] F. Plasser and H. Lischka, Analysis of excitonic and charge transfer interactions from quantum chemical calculations, *J Chem Theory Comput*, 2012, **8**, 2777–2789.

[61] A. v. Luzanov and O. v. Prezhdo, Irreducible charge density matrices for analysis of many-electron wave functions, *Int J Quantum Chem*, 2005, **102**, 582–601.

[62] C. R. Harris, K. J. Millman, S. J. van der Walt, R. Gommers, P. Virtanen, D. Cournapeau, E. Wieser, J. Taylor, S. Berg, N. J. Smith, R. Kern, M. Picus, S. Hoyer, M. H. van Kerkwijk, M. Brett, A. Haldane, J. F. del Río, M. Wiebe, P. Peterson, P. Gérard-Marchant, K. Sheppard, T. Reddy, W. Weckesser, H. Abbasi, C. Gohlke and T. E. Oliphant, *Nature*, 2020, 585, 357–362.

[63] C. Kanimozhi, N. Yaacobi-Gross, K. W. Chou, A. Amassian, T. D. Anthopoulos and S. Patil, Diketopyrrolopyrrole–Diketopyrrolopyrrole-Based Conjugated Copolymer for High-Mobility Organic Field-Effect Transistors, *J Am Chem Soc*, 2012, **134**, 16532–16535.

[64] P. Löwdin, Approximate Formulas for Many-Center Integrals in the Theory of Molecules and Crystals, *J Chem Phys*, 1953, **21**, 374–375.

[65] P. Löwdin, On the Non-Orthogonality Problem Connected with the Use of Atomic Wave Functions in the Theory of Molecules and Crystals, *J Chem Phys*, 1950, **18**, 365–375.

[66] Z. Liang, L. Yan, J. Si, P. Gong, X. Li, D. Liu, J. Li and X. Hou, Rational Design and Characterization of Symmetry-Breaking Organic Semiconductors in Polymer Solar Cells: A Theory Insight of the Asymmetric Advantage, *Materials*, 2021, **14**, 6723.

[67] G. A. DiLabio, D. A. Pratt and J. S. Wright, Theoretical Calculation of Ionization Potentials for Disubstituted Benzenes: Additivity vs Non-Additivity of Substituent Effects, *J Org Chem*, 2000, **65**, 2195–2203.

[68 J. C. Bernède, Organic photovoltaic cells: History, principle and techniques, *Journal of the Chilean Chemical Society*, , DOI:10.4067/S0717-97072008000300001.

[69] B. Qi and J. Wang, Open-circuit voltage in organic solar cells, *J Mater Chem*, 2012, **22**, 24315–24325.

[70] B. Qi and J. Wang, Fill factor in organic solar cells, *Physical Chemistry Chemical Physics*, 2013, **15**, 8972.

[71] D. Grabowski, Z. Liu, G. Schöpe, U. Rau and T. Kirchartz, Fill Factor Losses and Deviations from the Superposition Principle in Lead Halide Perovskite Solar Cells, *Solar RRL*, , DOI:10.1002/solr.202200507.

[72] M. Wang, H. Wang, W. Li, X. Hu, K. Sun and Z. Zang, Defect passivation using ultrathin PTAA layers for efficient and stable perovskite solar cells with a high fill factor and eliminated hysteresis, *J Mater Chem A Mater*, 2019, **7**, 26421–26428.

[73] J. Zhang, J. Qin, M. Wang, Y. Bai, H. Zou, J. K. Keum, R. Tao, H. Xu, H. Yu, S. Haacke and B. Hu, Uniform Permutation of Quasi-2D Perovskites by Vacuum Poling for Efficient, High-Fill-Factor Solar Cells, *Joule*, 2019, **3**, 3061–3071.

[74] M. Knupfer, Exciton binding energies in organic semiconductors, *Applied Physics A*, 2003, **77**, 623–626.

[75] D. Comaniciu and P. Meer, Mean shift: a robust approach toward feature space analysis, *IEEE Trans Pattern Anal Mach Intell*, 2002, **24**, 603–619.

[76] J. Grus, *Data science from scratch: first principles with Python*, O'Reilly Media, Inc., 1st edn., 2015.

[77] M. S. Alexiou, V. Tychopoulos, S. Ghorbanian, J. H. P. Tyman, R. G. Brown and P. I. Brittain, The UV–visible absorption and fluorescence of some substituted 1,8-naphthalimides and naphthalic anhydrides, *J. Chem. Soc., Perkin Trans. 2*, 1990, **Issue, 2**, 837–842.

[78] W. Huang, X. Zhang, L.-H. Ma, C.-J. Wang and Y.-B. Jiang, Intramolecular charge transfer dual fluorescence of substituted-phenyl p-

dimethylaminobenzoates with comparable electron acceptors, *Chem Phys Lett*, 2002, **352**, 401–407.

[79] P.-F. Loos, M. Comin, X. Blase and D. Jacquemin, Reference Energies for Intramolecular Charge-Transfer Excitations, *J Chem Theory Comput*, 2021, **17**, 3666–3686.

[80] L. Modesto-Costa and I. Borges, Discrete and continuum modeling of solvent effects in a twisted intramolecular charge transfer system: The 4-N,N-dimethylaminobenzonitrile (DMABN) molecule, *Spectrochim Acta A Mol Biomol Spectrosc*, 2018, **201**, 73–81.

[81] T. Lu and F. Chen, Multiwfn: A multifunctional wavefunction analyzer, *J Comput Chem*, 2012, **33**, 580–592.

[82] M. Khalid, M. U. Khan, S. Ahmed, Z. Shafiq, M. M. Alam, M. Imran, A. A. C. Braga and M. S. Akram, Exploration of promising optical and electronic properties of (non-polymer) small donor molecules for organic solar cells, *Sci Rep*, 2021, **11**, 21540.

[83] M. Waqas, N. M. A. Hadia, A. M. Shawky, R. F. Mahmood, M. Essid, Z. Aloui, N. S. Alatawi, J. Iqbal and R. A. Khera, Theoretical framework for achieving high V_{oc} in non-fused non-fullerene terthiophene-based end-capped modified derivatives for potential applications in organic photovoltaics, *RSC Adv*, 2023, **13**, 7535–7553.