Supporting Information: From Fragments to Function: Data-Driven Design of High-Performance Non-Fullerene Acceptors for Organic Photovoltaics

Bibhas Das[1], Kalyani Patrikar[1], Atharva Sachin Keny[2], and Anirban Mondal[1,*]

[1]Department of Chemistry, Indian Institute of Technology Gandhinagar, Gujarat, 382355, India

[2]Computer Science and Engineering, Indian Institute of Technology Goa, Goa, 403401, India

December 1, 2025

E-mail: amondal@iitgn.ac.in

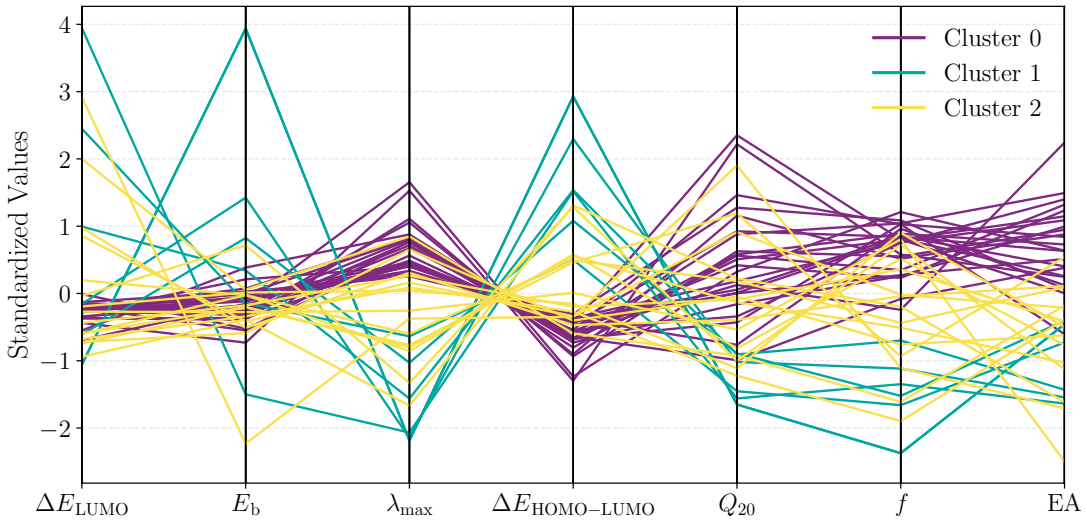

Figure S1: Parallel coordinates plot representing the distribution of key optoelectronic properties across acceptor fragments in the ADA-type non-fullerene acceptor dataset. Each line corresponds to an individual acceptor fragment, and properties are standardized (z-scores) for comparison across different scales. Distinct clusters are highlighted in color, revealing characteristic property patterns and interdependencies. This visualization facilitates the identification of high-performing acceptor motifs for rational molecular design.

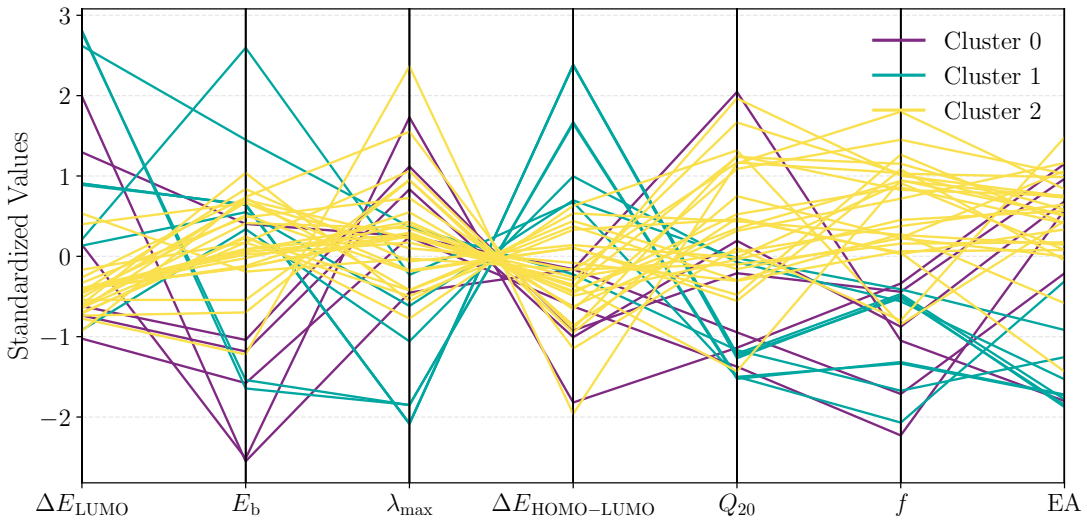

Figure S2: Parallel coordinates plot representing the distribution of key optoelectronic properties across side chain fragments in the ADA-type non-fullerene acceptor dataset. Each line corresponds to an individual side chain fragment, and properties are standardized (z-scores) for comparison across different scales. Distinct clusters are highlighted in color, revealing characteristic property patterns and interdependencies. This visualization facilitates the identification of high-performing acceptor motifs for rational molecular design.

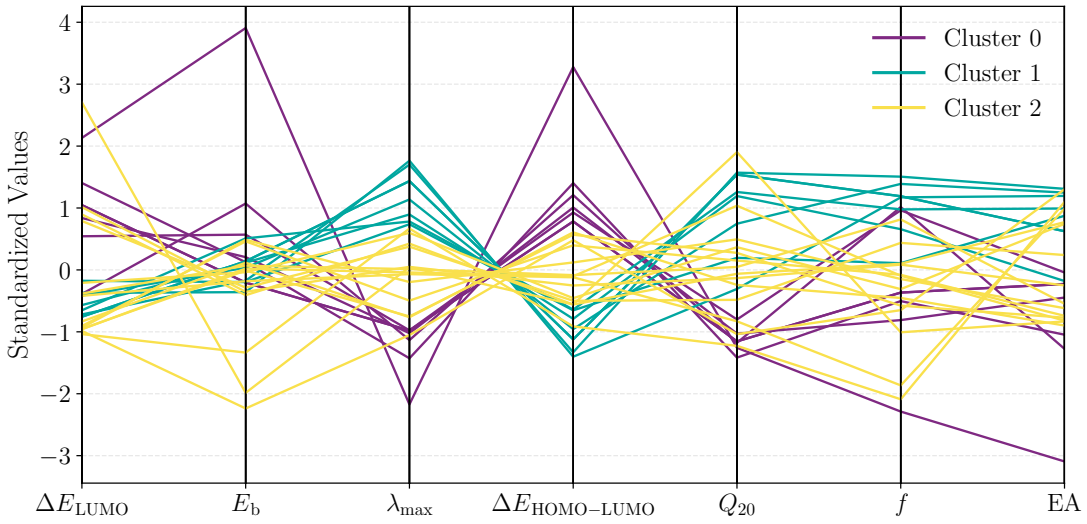

Figure S3: Parallel coordinates plot representing the distribution of key optoelectronic properties across conjugated extension fragments in the ADA-type non-fullerene acceptor dataset. Each line corresponds to an individual conjugated extension fragment, and properties are standardized (z-scores) for comparison across different scales. Distinct clusters are highlighted in color, revealing characteristic property patterns and interdependencies. This visualization facilitates the identification of high-performing acceptor motifs for rational molecular design.

Table S1: Principal component analysis (PCA) of donor fragment descriptors. The table lists the top seven principal components, their explained variance ratios, and the three most contributing features to each component.

Component	Explained Variance Ratio	Top Features (Contribution)
Component 1	0.479	[$\Delta E_{\mathrm{HOMO-LUMO}}$ (0.51), E_{b} (0.35), ΔE_{LUMO} (0.08)]
Component 2	0.221	[E_{b} (0.52), Q_{20} (0.49), f (0.49)]
Component 3	0.163	[ΔE_{LUMO} (0.80), f (0.45), Q_{20} (0.23)]
Component 4	0.064	[EA (0.79), ΔE_{LUMO} (0.38), E_{b} (0.19)]
Component 5	0.041	[f (0.63), $\Delta E_{\mathrm{HOMO-LUMO}}$ (0.08), λ_{max} (-0.26)]
Component 6	0.031	[E_{b} (0.59), λ_{max} (0.46), f (0.29)]
Component 7	0.001	[$\Delta E_{\mathrm{HOMO-LUMO}}$ (0.76), λ_{max} (0.63), f (0.06)]

Table S2: Principal component analysis (PCA) of acceptor fragment descriptors. The table lists the top seven principal components, their explained variance ratios, and the three most contributing features to each component.

Component	Explained Variance Ratio	Top Features (Contribution)
Component 1	0.536	[$\Delta E_{\mathrm{HOMO-LUMO}}$ (0.49), E_{b} (0.32), ΔE_{LUMO} (0.07)]
Component 2	0.221	[ΔE_{LUMO} (0.72), Q_{20} (0.08), λ_{max} (-0.07)]
Component 3	0.103	[EA (0.59), ΔE_{LUMO} (0.36), λ_{max} (0.26)]
Component 4	0.071	[Q_{20} (0.75), EA (0.29), $\Delta E_{\mathrm{HOMO-LUMO}}$ (0.13)]
Component 5	0.042	[f (0.59), EA (0.42), ΔE_{LUMO} (0.46)]
Component 6	0.025	[E_{b} (0.68), λ_{max} (0.44), ΔE_{LUMO} (0.36)]
Component 7	0.001	[$\Delta E_{\mathrm{HOMO-LUMO}}$ (0.76), λ_{max} (0.61), f (0.02)]

Table S3: Principal component analysis (PCA) of side-chain fragment descriptors. The table lists the top seven principal components, their explained variance ratios, and the three most contributing features to each component.

Component	Explained Variance Ratio	Top Features (Contribution)
Component 1	0.547	[EA (0.47), λ_{max} (0.46), Q_{20} (0.33)]
Component 2	0.189	[E_{b} (0.72), f (0.43), $\Delta E_{\mathrm{HOMO-LUMO}}$ (0.32)]
Component 3	0.123	[Q_{20} (0.64), f (0.41), $\Delta E_{\mathrm{HOMO-LUMO}}$ (0.14)]
Component 4	0.063	[Q_{20} (0.62), E_{b} (0.43), λ_{max} (0.18)]
Component 5	0.056	[ΔE_{LUMO} (0.72), f (0.52), E_{b} (0.25)]
Component 6	0.021	[EA (0.83), ΔE_{LUMO} (0.31), $\Delta E_{\mathrm{HOMO-LUMO}}$ (0.26]
Component 7	0.002	[$\Delta E_{\mathrm{HOMO-LUMO}}$ (0.73), λ_{max} (0.67), EA (0.03)]

Table S4: Principal component analysis (PCA) of conjugated-extension fragment descriptors. The table lists the top seven principal components, their explained variance ratios, and the three most contributing features to each component.

Component	Explained Variance Ratio	Top Features (Contribution)
Component 1	0.566	[$\Delta E_{\mathrm{HOMO-LUMO}}$ (0.49), ΔE_{LUMO} (0.28), E_{b} (0.27)]
Component 2	0.172	[E_{b} (0.67), f (0.48), Q_{20} (0.41)]
Component 3	0.116	[ΔE_{LUMO} (0.86), Q_{20} (0.40), λ_{max} (0.12)]
Component 4	0.076	[f (0.69), ΔE_{LUMO} (0.36), EA (0.31)]
Component 5	0.054	[E_{b} (0.60), λ_{max} (0.46), EA (0.34)]
Component 6	0.016	[EA (0.68), Q_{20} (0.46), $\Delta E_{\mathrm{HOMO-LUMO}}$ (0.41)]
Component 7	0.001	[$\Delta E_{\mathrm{HOMO-LUMO}}$ (0.73), λ_{max} (0.64), f (0.01)]

Table S5: Performance metrics (RMSE, MAE, and Pearson correlation coefficient r) for predicting key optoelectronic properties—LUMO offset (ΔE_{LUMO}), maximum absorption wavelength (λ_{max}), oscillator strength (f), and exciton binding energy (E_{b})—on the training set using the evidential MPNN model.

Metrics	ΔE_{LUMO}	λ_{max}	f	E_{b}
RMSE	0.169	39.33	0.339	0.028
MAE	0.095	29.25	0.263	0.019
Pearson r	0.862	0.907	0.945	0.909