MSDE

ROYAL SOCIETY OF CHEMISTRY

PAPER

View Article Online
View Journal

Cite this: DOI: 10.1039/d5me00182j

From fragments to function: data-driven design of high-performance non-fullerene acceptors for organic photovoltaics

Bibhas Das, [a] Kalyani Patrikar,[a] Atharva Sachin Keny[b] and Anirban Mondal *[a]

The rapid advancement of organic photovoltaics (OPVs) depends critically on discovering non-fullerene acceptors (NFAs) with finely balanced optoelectronic properties. Yet, identifying optimal NFAs remains challenging due to the vast chemical space and complex, non-linear structure–property relationships. Here, we present a computational framework integrating physics-guided molecular fragmentation, hierarchical clustering, and combinatorial assembly with uncertainty-aware machine learning to accelerate NFA design. Beginning with 257 experimentally reported NFAs, we assembled a synthetically viable library of 500 000 NFAs with acceptor–donor–acceptor (ADA) structures. An evidential message-passing neural network (MPNN) was trained to predict oscillator strength (f), LUMO offset (ΔE_{LUMO}), absorption maximum (λ_{max}), and exciton binding energy (E_b), achieving high accuracy with built-in uncertainty quantification. Compared to the training distribution, our pipeline produced a deterministic enrichment of candidates with tightly converged, target-optimized values ($f \geq 1.5$, $\Delta E_{LUMO} < 0.25$ eV, $E_b < 0.32$ eV), in line with state-of-the-art OPV performance benchmarks. Quantum chemical validation confirmed prediction fidelity, with all deviations within 22%. This unified and interpretable framework provides a scalable route for rational NFA discovery and establishes a generalizable benchmark for machine learning-guided materials design in organic electronics.

Received 2nd October 2025,
Accepted 8th December 2025

DOI: 10.1039/d5me00182j

rsc.li/molecular-engineering

Design, System, Application

We present a physics-guided, data-driven framework for rationally designing non-fullerene acceptors (NFAs) in organic photovoltaics. Our approach integrates molecular fragmentation, hierarchical clustering, and uncertainty-aware machine learning to construct and screen a synthetically meaningful library of half a million candidate molecules. By embedding chemical intuition at the building-block level, the framework ensures interpretable structure–property relationships while directly targeting critical OPV performance trade-offs such as exciton binding energy, LUMO offset, and oscillator strength. The system we investigate—donor–acceptor–donor type NFAs—represents the current frontier of high-efficiency organic solar cells. Beyond photovoltaics, this methodology provides a generalizable route for accelerating molecular discovery across organic electronics, enabling scalable and rational design of semiconductors for applications ranging from light-emitting diodes to photodetectors and transistors.

1 Introduction

Over the past decade, organic photovoltaics (OPVs) have achieved remarkable gains in power conversion efficiency (PCE), driven by advances in molecular design,[1–3] device engineering,[4–8] and processing technologies.[9–14] A major turning point has been the emergence of non-fullerene acceptors (NFAs),[15–21] which extend the absorption profile of OPVs and enable efficient solar energy capture across a broader spectrum. Unlike their fullerene counterparts, NFAs provide tunable optoelectronic properties, enhanced morphological stability, and superior charge transport, establishing them as the cornerstone of next-generation OPV technologies.[22–31]

Achieving high OPV efficiency requires fine control over a hierarchy of microscopic processes—exciton generation, charge separation, transport, and recombination—each of which depends sensitively on molecular properties such as absorption breadth, exciton binding energy, and charge mobility.[32–39] While empirical studies have uncovered correlations between structure and performance,[40–44] the search for optimal NFAs remains formidable. The underlying challenge stems from the vast chemical space and the complex, non-linear mapping between molecular structure and device-level performance.[45–50]

To navigate this complexity, machine learning (ML) has become a powerful tool for accelerating molecular

[a] Department of Chemistry, Indian Institute of Technology Gandhinagar, Gujarat, 382355, India. E-mail: amondal@iitgn.ac.in
[b] Computer Science and Engineering, Indian Institute of Technology Goa, Goa, 403401, India

discovery.[45,46,48–67] By learning from experimental or computational datasets, ML models can predict key properties of novel candidates and enable efficient pre-screening.[48,58–60,62,65–67] Nevertheless, most existing frameworks suffer from a critical limitation: they rely on random or unconstrained generation of molecular structures, which are then filtered by the algorithm. This approach often overlooks promising candidates and provides limited chemical interpretability. Moreover, relatively little attention has been paid to how the input molecular building blocks themselves can be tuned to guide property optimization.[48,58–60,62,63,65–67]

Here, we introduce a systematic computational framework for the targeted design of NFAs that embeds chemical intuition directly into the generative process. Our approach begins by deconstructing 257 experimentally reported NFAs into modular fragments that preserve donor–acceptor motifs and conjugation pathways. Each fragment is characterized using quantum chemical (QM) calculations to obtain optoelectronic descriptors including LUMO energy (E_{LUMO}), exciton binding energy (E_{b}), electron affinity (EA), HOMO–LUMO gap ($\Delta E_{\mathrm{HOMO-LUMO}}$), oscillator strength ($f$), quadrupole moment ($Q_{20}$), and absorption maximum (λ_{max}). Unsupervised learning methods such as principal component analysis (PCA) and k-means clustering then organize these fragments into chemically meaningful groups, enabling rational design space navigation. By employing a multi-objective optimization strategy, we directly target the central trade-offs in OPV design—minimizing exciton binding energy and LUMO offset (ΔE_{LUMO}) to promote charge separation, while maximizing oscillator strength and optical activity through favorable λ_{max} values.

Numerous machine-learning and deep-learning approaches have been explored for OPV and NFA design, ranging from supervised models trained on molecular descriptors and quantum-mechanical features to high-throughput virtual screening workflows driven by semi-empirical or DFT calculations.[39,50,52,56,59,65–68] Recent studies have introduced NFA-specific representations, including sub-unit descriptors, fragment-based fingerprints, and DL architectures that jointly optimise donor–acceptor pairs.[69–73] While these methods have significantly advanced computational OPV design, they typically operate on whole-molecule fingerprints or latent embeddings, which provide limited interpretability and often lack safeguards against chemically unrealistic structures. In contrast, our framework introduces a chemically guided, electron-withdrawing group (EWG)-aware fragmentation strategy, fragment-level clustering with optoelectronic interpretability, and an uncertainty-aware message-passing neural network (MPNN) for multi-objective optimization. This combination enables synthetically grounded molecular generation and offers a complementary, transparent route to ML-assisted NFA discovery.

From this structured and interpretable fragment space, we generate 500 000 new NFAs through guided recombination across cluster boundaries, ensuring systematic yet chemically relevant exploration. A message-passing neural network trained on the manually compiled and quality-controlled dataset predicts the optoelectronic properties of these candidates, and high-ranking structures are validated with density functional theory (DFT). The excellent agreement between ML predictions and QM benchmarks highlights both accuracy and efficiency. By uniting physics-guided fragmentation, data-driven clustering, and ML-assisted screening, our framework offers a scalable and interpretable route to molecular design. Beyond OPVs, this methodology provides a generalizable blueprint for accelerating the discovery of functional organic semiconductors with tailored electronic and optical properties.

2 Methods

Our study is established on a manually compiled and quality-controlled dataset of 257 non-fullerene acceptor molecules with acceptor–donor–acceptor (A–D–A) type structure, each characterized by quantum mechanical descriptors (dataset-I). This dataset was derived from a broader collection of NFAs reported in the literature[50] and was initially developed as part of a previous study in our group.[66] The selection process was designed to capture the structural diversity of ADA-type NFAs while ensuring a representative distribution of key electronic and optical properties relevant to OPVs.

2.1 Automated fragmentation and structural classification

To better understand how the structure of ADA-type NFAs influences their electronic properties, we developed a cheminformatics-based method for automated molecular fragmentation and classification. This approach, built using the RDKit library,[74] allows us to break down complex molecules into meaningful parts while preserving key electronic features.

Fragmentation procedure and tagging system. We start by converting the molecular structures from their SMILES representations into RDKit objects. The first step in the analysis focuses on locating EWGs, which are known to play an important role in the electronic behavior of NFAs. We use a well-chosen library of SMILES patterns to detect these groups in each molecule. Once identified, the bonds directly connected to the EWGs are marked and excluded from fragmentation to keep the acceptor subgroup intact (see Fig. 1a). Next, we break bonds primarily at junctions between aromatic and non-aromatic regions—usually where the donor core connects with flank groups. We introduce a tagging system to keep track of how the fragments are connected. At each cut point, we label the atoms with special atomic numbers (greater than 49) that act as markers. These tags help us maintain a clear map of how the fragments were initially connected, making it easier to reconstruct or analyze the molecule later.

Identifying the donor core. To find the central donor unit in each molecule, we use two criteria: (i) the number of fused

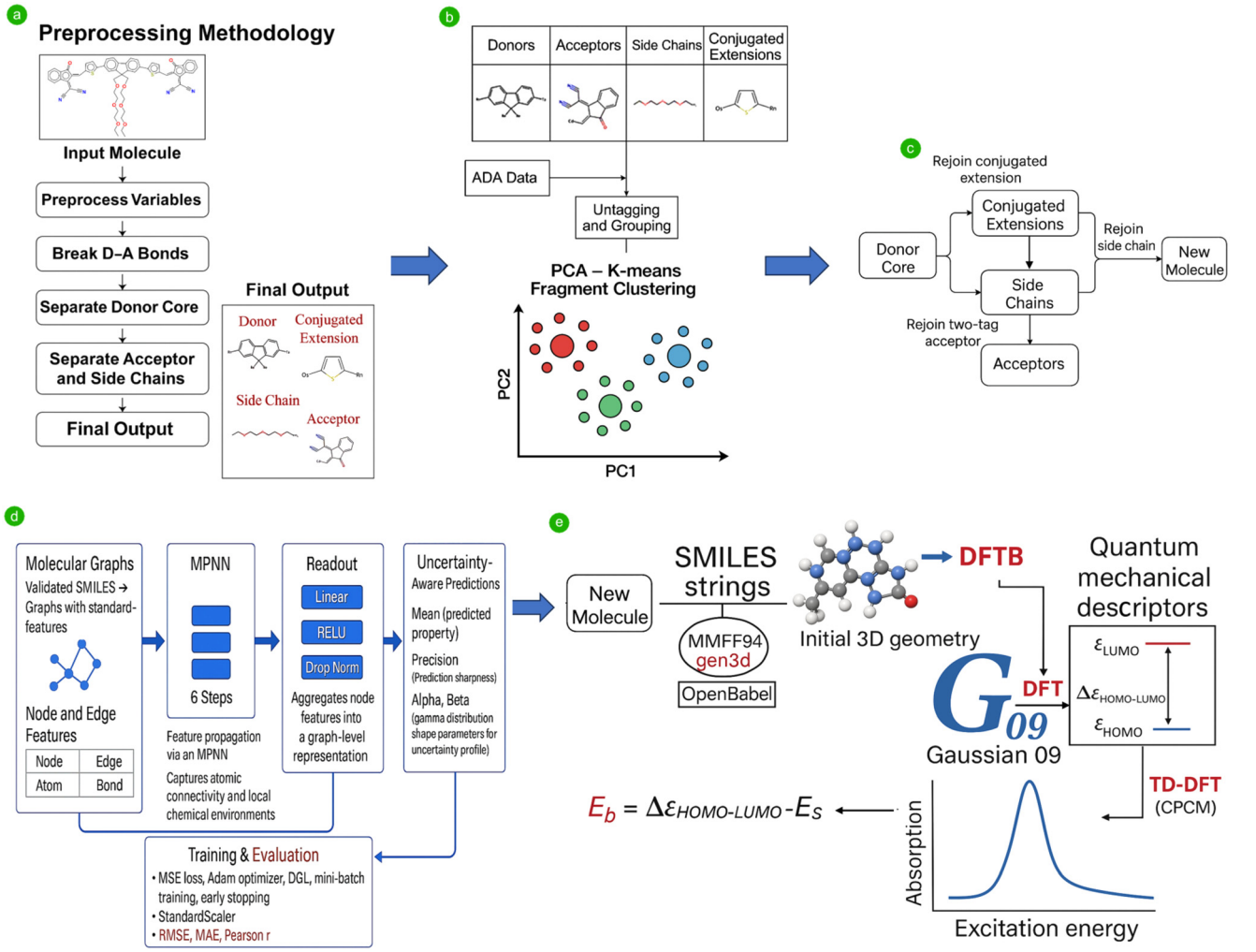

Fig. 1 Schematic overview of the fragment-based molecular design framework. (a) Experimentally validated non-fullerene acceptors are decomposed into modular fragments, which are characterized *via* quantum chemical calculations to derive key optoelectronic descriptors. (b) Dimensionality reduction (PCA) and clustering (*k*-means) identify chemically meaningful fragment families. (c) Strategic recombination of fragments across clusters enables the generation of novel NFAs, (d) which are screened using a message-passing neural network trained on computed properties. (e) Top candidates are further validated *via* DFT calculations to ensure predictive fidelity and optimize structure–property relationships for efficient organic photovoltaics.

aromatic rings in a fragment and (ii) the number of tagged connection points. This helps us reliably identify the central donor structure, especially in complex molecules. Once the donor is isolated, we perform another fragmentation round—this time only at the bonds associated with EWGs. This allows us to cleanly separate the acceptor units from the rest of the molecule while still using the tagging system to keep connectivity information.

Classifying molecular fragments. After fragmentation, the resulting pieces are grouped into four categories:

- Donor cores – aromatic fused ring systems forming the molecule's electronic backbone,
- Acceptor units – fragments that include electron-withdrawing groups,
- Side chains – non-aromatic parts without any EWGs and
- Conjugated extensions – aromatic fragments attached to the donor core but not part of the fused ring system.

To assign these categories, we use a step-by-step classification system that checks ring structures, matches known substructures, and examines aromaticity. Special care was taken to distinguish between true donor units and other aromatic fragments, especially in cases with branching or non-fused rings. The tagging system also tracks how fragments are connected—whether donor-to-acceptor, donor-to-side-chain, or donor-to-conjugated-extension. This ensures that important connectivity information is preserved for later use.

Generating output and ensuring robustness. The final output includes the SMILES strings of all key fragments, along with associated electronic and structural information, including HOMO/LUMO energy levels, oscillator strengths, and properties of the parent molecules. These outputs can be used in downstream analyses like clustering, principal component analysis, or machine learning models. We have

also built in extensive error-checking throughout the pipeline. The system verifies ring detection, validates how fragments are classified, and fixes any issues with tagging. It also handles exceptional cases—like complex fused ring systems or unclear conjugation paths—using fallback rules to ensure consistency and reliability. Overall, this method offers a robust, modular, and scalable approach for breaking down and analyzing ADA-type NFAs.

2.2 Unsupervised analysis of structural and electronic features

To understand how variations in molecular substructure influence electronic properties in ADA-type NFAs, we employed a data analysis pipeline that integrates cheminformatics and machine learning. Molecular SMILES strings were processed using RDKit and fragmented as described earlier. Structural roles—donor, acceptor, side chain, or extension—were assigned to each fragment, and inter-fragment connections and associated physicochemical properties were stored in a structured JSON format for downstream analysis.

We focus on seven key optoelectronic descriptors for each acceptor fragment: the LUMO–LUMO+1 energy difference (ΔE_{LUMO}), exciton binding energy (E_{b}), oscillator strength (f), quadrupole moment (Q_{20}), electron affinity (EA), absorption maximum (λ_{max}), and HOMO–LUMO gap ($\Delta E_{\mathrm{HOMO-LUMO}}$). For each NFA fragment, we evaluated these seven quantum-chemical descriptors because, together, they provide a compact yet comprehensive representation of the electronic, optical, and interfacial characteristics most relevant to NFA performance. To enable a physically guided reduction of this feature space, the descriptors were organized into two conceptually motivated groups: (i) gap- and exciton-related quantities (ΔE_{LUMO}, E_{b}, λ_{max}, and $\Delta E_{\mathrm{HOMO-LUMO}}$) that co-vary with conjugation and define the energetic landscape governing charge generation, and (ii) charge-distribution and optical-response descriptors (EA, f, and Q_{20}) that influence absorption intensity and donor–acceptor electrostatics. Scaling these two groups separately prior to PCA places the dominant design trade-offs along orthogonal axes, making the resulting low-dimensional representation more interpretable and better aligned with the multi-objective property targets used for molecular screening. This strategy ensures that the subsequent clustering analysis reflects meaningful physicochemical distinctions among fragments rather than artifacts of feature scaling. We implemented a two-stage dimensionality-reduction strategy to manage this seven-dimensional space while preserving its most informative variations. First, each descriptor group was independently standardized using the StandardScaler method to ensure uniform scaling. Principal component analysis was then applied separately to the two groups, compressing them into one-dimensional representations that capture their dominant patterns. With the processed features, we performed clustering to identify the underlying structure in the fragment space. Both k-means clustering (with $k = 3$) and affinity propagation were applied to the combined standardized features, and the elbow method, silhouette score, and Calinski–Harabasz index were used to select a robust and well-separated clustering solution that reflects genuine differences in physicochemical behavior.

We generated parallel coordinates plots, PCA scatter plots, and molecular renderings to visualize and interpret the clustering results. For each cluster, we compiled detailed statistical summaries, including mean values and standard deviations for all descriptors. We also examined feature contributions to each PCA axis and assessed how well the clusters are separated. This automated and comprehensive analysis framework allows us to systematically group molecular fragments with similar electronic characteristics.

2.3 Combinatorial assembly of donor–acceptor fragments

We developed a modular framework that assembles molecular fragments—donor cores, acceptors, side chains, and conjugated extensions—through a combinatorial strategy to construct a diverse library of ADA-type NFA molecules. This assembly process was guided by a tagging system that identifies chemically viable connection points and ensured consistent structural integration across fragments. We began by loading picked out sets of SMILES strings representing different fragment types from structured CSV files. Each donor core was analyzed for available connection points using the pre-defined tags described earlier. Connectivity rules were stored in JSON format, enabling flexible and scalable manipulation. Fragment combinations were deliberately chosen from distinct clusters based on their electronic and structural properties to enhance molecular diversity.

The assembly followed a stepwise approach. First, tag-matching algorithms attached zero or more conjugated extensions to specific positions on the donor core. These extensions were bonded to the donor by aligning complementary tags, after which the tags were removed while ensuring molecular integrity. Next, side chains were added to the donor or any attached conjugated extensions, with an internal indexing scheme managing multiple attachment sites during this stage. To avoid generating chemically implausible NFAs, the tagging system was embedded within a multi-stage validation procedure. Fragmentation was performed only at chemically reasonable single-bond junctions (*e.g.*, donor–flank or donor-side-chain linkages), while electron-withdrawing acceptor motifs were protected from cleavage. Tag codes uniquely specify donor–acceptor, donor-side-chain, donor-conjugated-extension, and extension-side-chain junctions, and these tags act as hard constraints during recombination: only matching tag types may be paired, and all resulting bonds correspond to junctions already present in at least one experimental NFA. Structures that violate valence rules, aromaticity, stereochemistry, or charge neutrality—assessed using multiple RDKit sanitization

routines—were removed. This workflow ensures that the generative process is restricted to synthetically realistic connectivity patterns, preventing the formation of chemically unreasonable structures. The final step involved attaching acceptor fragments in one of two ways: single-tag acceptors were joined directly to the donor, while double-tag acceptors were connected across both ends of the donor core, completing the ADA motif. Throughout this process, the framework used counters and dynamic data structures to track tag usage and manage bond formation precisely. Every assembled molecule underwent sanitization and structural checks using RDKit to ensure chemical validity. Successful assemblies were stored in a pandas DataFrame, with each molecule represented by its SMILES string. The final generative library was capped at 500 000 ADA-type NFAs to strike a balance between chemical coverage and computational feasibility. Although the retained fragment clusters could, in principle, produce millions to billions of combinations, exhaustive enumeration is unnecessary and impractical. Instead, a diversity-aware stochastic recombination scheme with a predefined stopping criterion was used to generate a library extensive enough to robustly sample all high-performing fragment clusters and yield smooth multi-objective property distributions, while remaining tractable for complete graph construction and uncertainty-aware MPNN evaluation. This procedure generated a chemically diverse and synthetically viable dataset of 500 000 ADA-type NFA molecules (dataset-II). The framework's modularity allows easy incorporation of new fragments or updated connection rules, making it a flexible tool for high-throughput molecular design in organic electronics.

2.4 Message passing neural network with evidential learning

To predict molecular properties across a chemically diverse library of 500 000 ADA-type non-fullerene acceptors, we employed a deep learning framework based on message passing neural networks enhanced with evidential learning. In this framework, each molecule was encoded as a graph, where atoms served as nodes and bonds as edges, enabling the model to learn from the molecular structure directly.[75] This representation naturally captured atomic connectivity and local chemical environments, providing a rich foundation for structure–property learning. Our custom neural architecture integrates an MPNN for feature propagation with an evidential regression module for uncertainty quantification. Information was exchanged among atomic nodes over six message-passing steps—a value empirically determined to balance prediction accuracy with computational cost. Fewer steps yielded insufficient structural context, while additional steps offered marginal gains at higher computational expense. The evidential learning component outputs four parameters per predicted property: the predicted mean, a precision term dictating prediction sharpness, and two gamma-distribution-derived shape parameters that characterize the uncertainty profile's spread and tail behavior. This allows for confidence-aware predictions, distinguishing between outputs of similar values but differing reliability.

Model training was guided by a mean squared error loss function, using the Adam optimizer[76] with weight decay to ensure robust convergence while controlling overfitting. Mini-batch training was implemented using the Deep Graph Library (DGL),[77] which supports efficient processing of variable-sized molecular graphs. Early stopping was applied based on validation performance to preserve model generalizability. Before training, SMILES-based molecular structures were validated and converted to graph representations with standardized atomic and bond features. Target property values were standardized using scikit-learn's StandardScaler, which uses a *z*-score transformation to stabilize training dynamics. To optimize model performance, we employed Optuna[78] for hyperparameter tuning across both architectural (*e.g.*, number of hidden features, dropout rate) and training parameters (*e.g.*, learning rate, batch size, number of epochs). The model evaluation relied on multiple metrics—including RMSE, MAE, and Pearson correlation coefficient—to assess prediction accuracy and the quality of uncertainty estimates. This graph-based deep learning with uncertainty-aware inference offers a robust, scalable framework for high-throughput screening and rational design of novel NFA candidates in optoelectronics.

2.5 Quantum chemical validation

To validate the MPNN-predicted molecular properties, we selected 10 best-performing representative molecules from the entire 500 000 molecule library and subjected them to quantum chemical calculations. These molecules were defined using their SMILES strings, and their initial 3D geometries were constructed using the MMFF94 force field *via* the gen3d module in OpenBabel.[79] The empirical force field parameters were taken from the Merck Molecular Force Field series.[80–84] These initial geometries were further optimized using the third-order self-consistent charge density-functional tight-binding method (DFTB3),[85–87] combined with many-body dispersion (MBD) corrections.[88–91] Calculations were performed using DFTB+,[92] interfaced *via* the atomic simulation environment (ASE).[93] This DFTB3-MBD refinement provided accurate starting structures for further optimization using Kohn–Sham density functional theory in Gaussian 09.[94] Geometry optimizations were carried out at the B3LYP/6-31G(d,p) level of theory, followed by frequency calculations to confirm the stability of the stationary points. From the optimized geometries, quantum mechanical descriptors such as energy levels (HOMO, LUMO), electron affinity and the HOMO–LUMO gap were extracted. To probe optoelectronic performance, we conducted time-dependent DFT (TD-DFT) calculations using the same functional and basis set. Excitation energies, absorption maxima, and oscillator strengths were evaluated

in a chloroform environment (ε = 4.7113) using the conductor-like polarizable continuum model (CPCM). For all fragments, the oscillator strength and absorption maximum were extracted from excited state 1, corresponding to the lowest singlet excited state in the TD-DFT calculations, ensuring consistency across the dataset. Finally, exciton binding energies were estimated from the HOMO–LUMO gap and the singlet excitation energy (E_S) according to the relation

$$E_b = \Delta E_{HOMO-LUMO} - E_S. \quad (1)$$

3 Results and discussion

3.1 Structural characteristics of NFA fragments

We performed a detailed structural analysis of the NFA fragment library to gain insights into the underlying molecular architecture that governs their electronic behavior. This analysis, summarized in Fig. 2, explores the distribution of key structural features—rings, atoms, bonds, and electron-withdrawing groups—across different fragment types and offers design-relevant trends for optimizing donor and acceptor performance. Fig. 2a presents the distribution of ring counts across the donor fragments. Donors typically contain a median of six rings, with aromatic rings dominating over aliphatic ones. This preference for aromaticity is consistent with established principles of organic electronics, where extended π-conjugation facilitates charge delocalization and enhances mobility. Interestingly, some donor structures exhibit a much higher ring count, reaching up to 24, indicating considerable structural diversity and suggesting that specific design strategies aim to amplify conjugation and fine-tune electronic properties *via* extended fused systems.

Fig. 2b examines atom count distributions across various fragment types. Donor fragments show a relatively narrow range, typically between 20 and 40 atoms, reflecting a standardized approach to constructing electron-rich cores. In contrast, acceptor fragments span a broader range—extending beyond 100 atoms in some cases—highlighting the complexity and diversity of electron-deficient architectures required to modulate electron affinity and optical properties. Side chains, which serve solubilizing and morphological tuning roles, exhibit a mild bimodal distribution centered around 10–15 atoms, likely reflecting the distinction between linear and branched alkyl chains. Notably, extensions show a more pronounced bimodal distribution, suggesting the presence of two distinct classes of structural modifications that are possibly associated with rigid *versus* flexible moieties. Bonding patterns, shown in Fig. 2c, further support the central role of aromatic systems in donor design. Donor

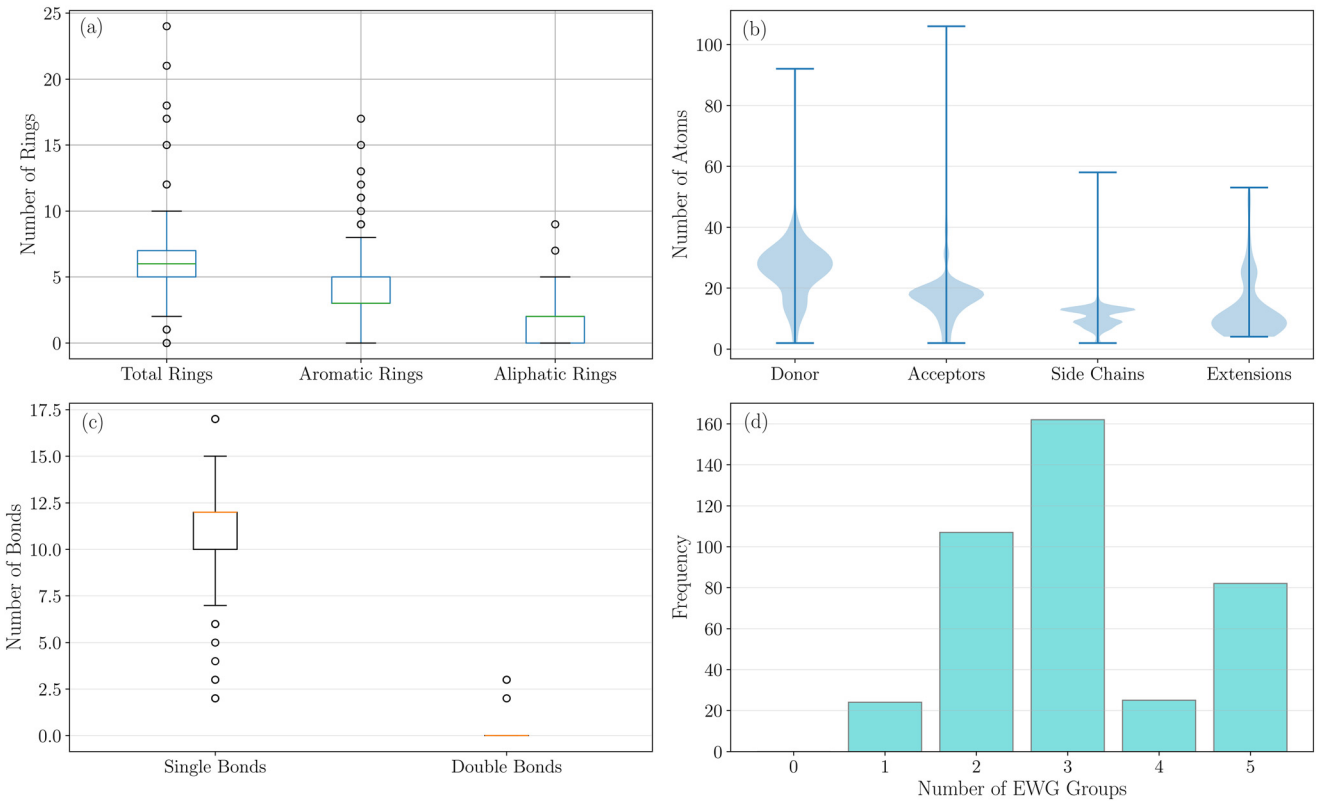

Fig. 2 Structural characteristics of non-fullerene acceptor fragments: (a) box plots showing the distribution of total, aromatic, and aliphatic rings in donor fragments; (b) violin plots of atom counts for donor, acceptor, side chain, and extension fragments; (c) box plots of single and double bond counts in donor fragments; and (d) histogram of electron-withdrawing group (EWG) counts in acceptor fragments.

fragments predominantly feature single bonds (median ~11), with minimal contributions from double bonds outside aromatic systems. This bond distribution implies a design strategy that favors conformational flexibility—imparted by single bonds—alongside electronic continuity provided by fused aromatic cores. The limited presence of isolated double bonds suggests an intentional avoidance of localized electronic features in favor of delocalized conjugation.

Finally, Fig. 2d highlights trends in incorporating electron-withdrawing groups within acceptor fragments. The most frequent configuration includes three EWGs, followed by structures with two or five. This suggests that incorporating three EWGs may represent an optimal balance between increasing electron affinity and maintaining other critical characteristics, such as solubility and morphological compatibility. The reduced occurrence of acceptors with one or four EWGs implies that such configurations may fall outside the desirable range for tuning LUMO levels and achieving efficient exciton dissociation.

Altogether, this fragment-level analysis underscores the structural strategies embedded in NFA design: the dominance of aromaticity in donor fragments to promote π-conjugation,

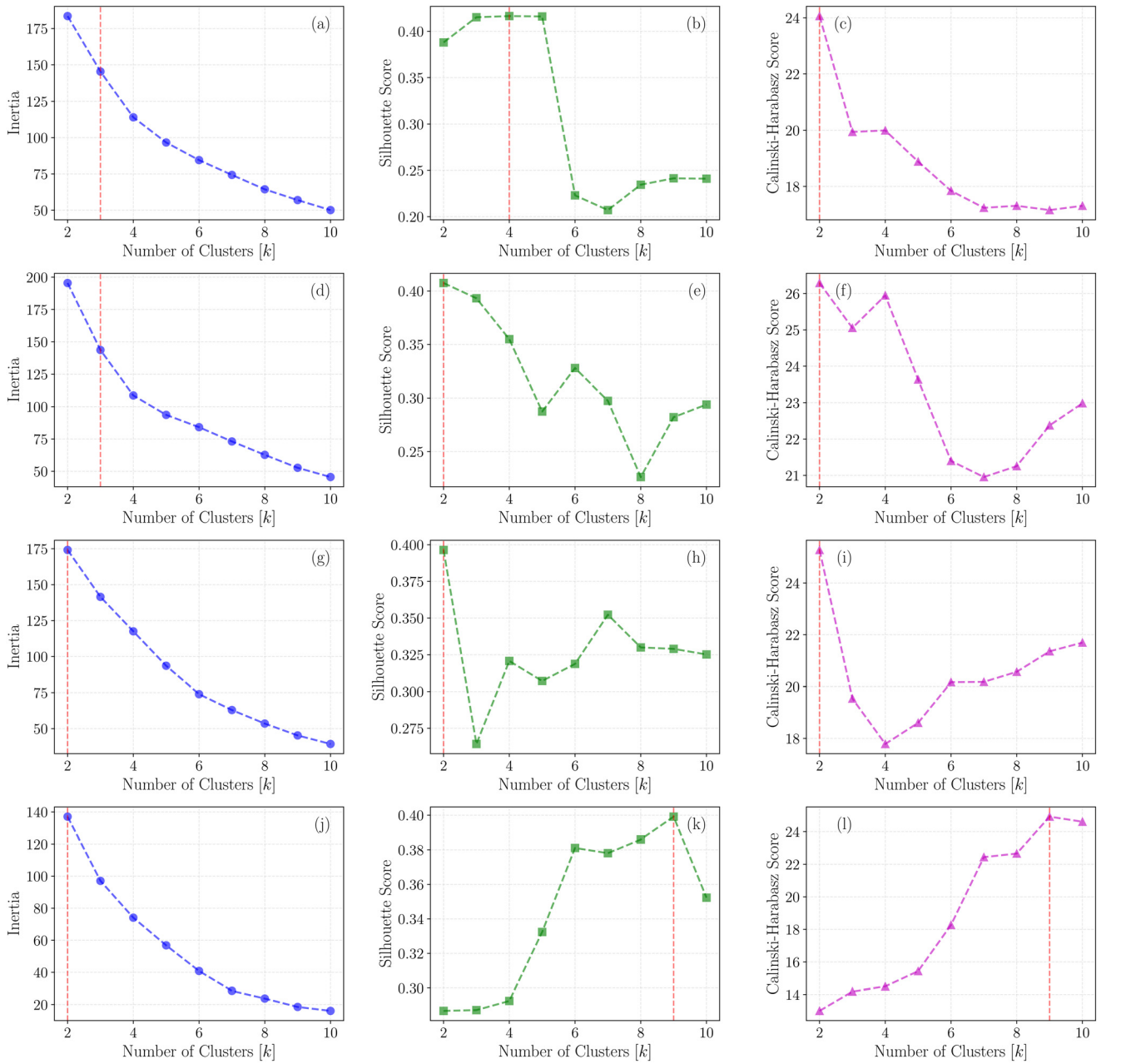

Fig. 3 Unsupervised clustering analysis of molecular fragments using the elbow method, silhouette score, and Calinski–Harabasz (CH) index across four fragment types: donor cores (a–c), acceptor units (d–f), side chains (g–i), and conjugated extensions (j–l). Each row represents a distinct fragment class, with panels depicting inertia (left), silhouette score (middle), and CH index (right) as functions of the number of clusters k. Red vertical lines highlight key inflection points or local maxima.

the broad architectural flexibility of acceptors to accommodate functional requirements, and the careful tuning of EWG content to modulate electronic properties. These insights are foundational design rules for the rational development of high-performance NFAs in organic photovoltaic applications.

3.2 Clustering analysis of fragments

Building upon the fragment-level structural insights, we sought to uncover broader organizational patterns within the molecular space by performing an unsupervised clustering analysis. To systematically classify the diverse set of donor, acceptor, and side chain fragments, we employed three complementary clustering validation techniques: the elbow method,[95] silhouette analysis, and the Calinski–Harabasz (CH) index. Together, these methods offer a statistically grounded approach to identifying the optimal number of clusters, enabling us to capture underlying structural similarities and diversity across the fragment population.

Donor. Fig. 3a–c presents the clustering analysis of donor fragments, offering insight into their underlying structural organization using three different statistical methods. Fig. 3a shows the elbow method, which helps determine the optimal number of clusters by plotting the within-cluster sum of squares (inertia) against different values of k (number of clusters). A noticeable drop in inertia is seen from $k = 2$ to $k = 3$ (from about 180 to 145). This "elbow" at $k = 3$, marked with a red line, suggests that adding more clusters beyond this point does not significantly improve the clustering, as the reduction in variance becomes marginal. This implies that three clusters may be suitable for capturing most of the structure in the donor fragment space. Fig. 3b provides silhouette analysis, which evaluates how well each molecule fits within its assigned cluster. The silhouette score starts at about 0.39 for $k = 2$, reaches a peak around 0.42 for $k = 4$, then drops sharply to about 0.22 at $k = 6$, and remains relatively low beyond that. This pattern suggests that clustering into 3 or 4 groups gives the most apparent separation and internal consistency, while larger clusters may not meaningfully represent the data.

Fig. 3c presents the Calinski–Harabasz index, which compares the spread between clusters to the spread within them. This score is highest at $k = 2$, with a value around 24, and then decreases steadily, leveling off near 17 for $k > 6$. This supports a primary division into two major groups, though the decline after $k = 2$ also suggests some structural subgroups may exist within this broader classification. These results indicate that donor fragments naturally form a few well-defined clusters—most notably two main groups, which can be further refined into three or four subgroups. The agreement across different methods highlights that, despite the structural diversity of donor fragments, their design space can be effectively organized into a small number of meaningful categories.

Acceptor. Fig. 3d–f show the clustering results for acceptor fragments, revealing distinct structural patterns within this molecular class. In Fig. 3d, the curve starts around 195 units at $k = 2$ and shows an apparent elbow at $k = 3$, indicating a sharp drop in the inertia of about 26%. Beyond this point, the curve flattens, suggesting that increasing the number of clusters adds limited value. This elbow suggests that acceptor fragments can be meaningfully grouped into three major types—likely corresponding to rylene diimides, benzothiadiazole, and cyano-based motifs, as commonly reported in the literature.[96–98] The silhouette score (Fig. 3e) peaks at around 0.405 for $k = 2$, then declines steadily, reaching about 0.36 at $k = 4$ and dropping further to 0.29 at $k = 5$. This trend suggests that a two-cluster classification best captures the internal consistency and separation of acceptor fragments, with further divisions providing a less meaningful structure. Fig. 3f displays the CH index, supporting this observation. The CH score reaches its highest value of about 26.5 at $k = 2$, confirming the strong distinction between the two primary groups. While there is a smaller secondary peak near $k = 4$, the decline beyond this point suggests limited benefit from additional clusters. Together, these analyses indicate that acceptor fragments have a clear and hierarchical structure: a strong initial division into two groups, with possible further separation into three or four subtypes.

Side chains. The third row (Fig. 3g–i) explores the clustering of side chain fragments, highlighting how these flexible units organize structurally. In Fig. 3g, the elbow method shows inertia values decreasing sharply from around 175 at $k = 2$ to about 95 at $k = 5$, suggesting the presence of key structural differences among side chains. Beyond this, the curve flattens and gradually drops to around 40 at $k = 10$, indicating limited benefits from further subdivision. Fig. 3h presents the silhouette analysis, which peaks at $k = 2$ with a score of roughly 0.39, pointing to a strong binary grouping. However, it drops sharply at $k = 3$ before rising again, reaching a secondary peak of around 0.35 at $k = 7$. This bimodal trend suggests a two-level structure: an initial broad division followed by finer subcategories, likely reflecting variations in flexibility, polarity, or branching. Fig. 3i shows the CH index, which also peaks at $k = 2$ with a value of 25.5, confirming the primary split. After an initial drop, the CH score increases steadily from $k = 5$ to $k = 10$, reaching about 21.7, suggesting the presence of nested subgroups. These results point to a hierarchical organization within side chain fragments—dominated by a major bifurcation and enriched by additional subtypes. The binary classification may suffice for many design applications, capturing dominant conformational or functional differences. However, the gradual evolution in clustering quality at higher k values supports a more detailed exploration when needed.

Conjugated extension. Fig. 3j–l focus on conjugated extension fragments, revealing a different and more complex organizational behavior. Fig. 3j, showing the elbow method, identifies a sharp drop in inertia from about 138 at $k = 2$ to 95 at $k = 3$, suggesting two dominant structural types. Beyond

this, the curve declines steadily to about 20 at k = 10, indicating a gradual refinement of substructures. Interestingly, Fig. 3k shows that the silhouette score behaves differently than other fragment types. Rather than peaking early, the score steadily increases from k = 2 to a maximum of around 0.40 at k = 9, indicating that finer clustering better captures the subtle structural and electronic features of extension units. Fig. 3l further supports this trend, as the CH index also increases consistently, peaking at about 25 near k = 9. A noticeable jump between k = 6 and k = 7 suggests a shift to more distinct subgroups, possibly due to variations in conjugation length or substitution patterns. While the elbow method hints at a basic two-cluster division, the silhouette, and CH analyses reveal a more subtle picture. The gradual increase in clustering quality points to overlapping structural motifs and a continuum of conjugation effects rather than sharply separated classes. This suggests that extension fragments are best understood through a multi-cluster perspective, where each group reflects distinct conjugation patterns or functional modifications.

3.3 Principal component and explained variance analysis

We employed principal component analysis with k-means clustering to reveal structure–property relationships and uncover distinct molecular architectures within the NFA design space. PCA was first applied for dimensionality reduction, and the cumulative explained variance plot (Fig. 4)

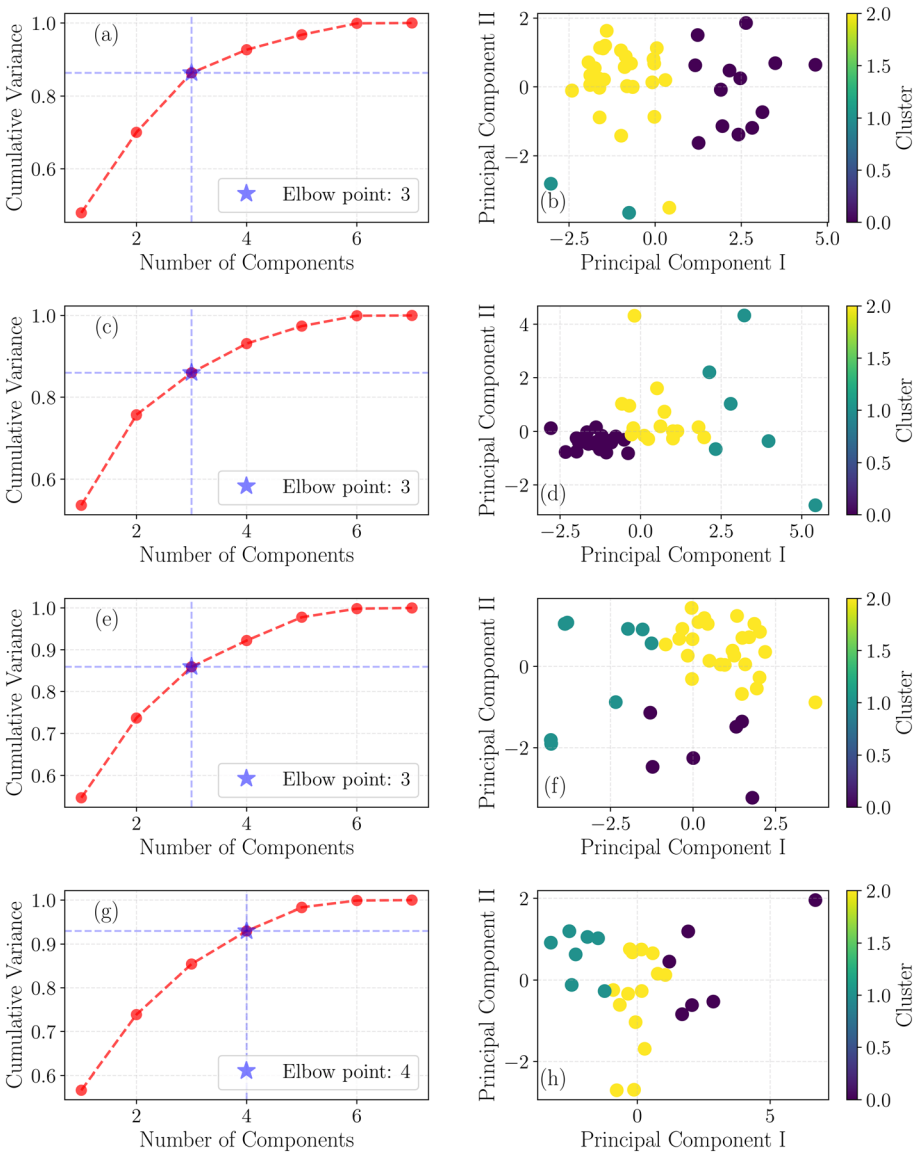

Fig. 4 Principal component analysis (PCA) of molecular fragments used non-fullerene acceptors: (a and b) donor fragments; (c and d) acceptor fragments; (e and f) side-chain fragments; (g and h) conjugated extension fragments. Left panels (a, c, e, g) show the explained variance ratios of the top seven principal components, with elbow points indicating optimal dimensionality thresholds. Right panels (b, d, f, h) display 2D PCA projections along the first two principal components, colored by k-means clustering (k = 3), revealing distinct groupings with interpretable variations in optoelectronic and excitonic properties.

highlights a non-linear accumulation of variance across the first seven principal components, indicating a complex distribution of information across dimensions. To identify the optimal number of components and clusters, we used the KneeLocator algorithm from the kneed package.[99] This method detects the "elbow" point in a curve—where additional components or clusters contribute marginal gains—by analyzing deviations from a baseline connecting the endpoints. The algorithm identifies the point of maximum curvature by normalizing input data and computing point-wise distances from this baseline. Parameters such as sensitivity (S = 1.0), curve shape ("concave"), and direction ("increasing") improve its flexibility across data types. Its automation and robustness make it well-suited for selecting optimal model complexity without manual tuning.

The two-dimensional PCA projection (Fig. 4) condenses the seven-dimensional property space into a more interpretable form, aiding the visualization of fragment distributions and their potential grouping. However, while PCA supports visual insight, k-means clustering was performed in the whole seven-dimensional property space—comprising properties such as $\Delta E_{\mathrm{HOMO-LUMO}}$, ΔE_{LUMO}, E_{b}, $\lambda_{\max}$, Q_{20}, f, and EA—to retain the complete molecular feature set during clustering. This analysis balances interpretability with data fidelity by combining PCA for dimensionality awareness with k-means clustering on the entire property space.

Donor fragments. For donor fragments, the first principal component (PC1) captures 48% of the total variance (Fig. 4a and Table S1), primarily influenced by the HOMO–LUMO gap ($\Delta E_{\mathrm{HOMO-LUMO}}$), exciton binding energy (E_{b}), and LUMO offset (ΔE_{LUMO}). These descriptors are closely interrelated, reflecting the strong coupling between frontier orbital energies and exciton-binding characteristics. The explained variance increases sharply from 48% to 86% across the first three components, after which it levels off—indicating that donor fragments occupy a relatively low-dimensional property space. The elbow point at component three highlights the optimal threshold for dimensionality reduction. The two-dimensional PCA projection (Fig. 4b), overlaid with k-means clustering (k = 3), reveals three well-separated donor fragment clusters. The horizontal axis (PC1) is dominated by the HOMO–LUMO gap ($\Delta E_{\mathrm{HOMO-LUMO}}$, 0.5121) with significant input from exciton binding energy (E_{b}, 0.3532) and a smaller contribution from LUMO offset (ΔE_{LUMO}, 0.0845), spanning roughly −3 to +5 standardized units. The vertical axis (PC2) is shaped by exciton binding energy (E_{b}, 0.5214), quadrupole moment (Q_{20}, 0.4980), and oscillator strength (f, 0.4877), ranging from −3 to +2 units and thus encoding the balance between exciton stability, charge-distribution anisotropy, and light-absorption strength. Cluster 2 (yellow) emerges as particularly promising: it encompasses a broad yet cohesive group of fragments located in the moderate PC1 region, exhibiting intermediate $\Delta E_{\mathrm{HOMO-LUMO}}$ and ΔE_{LUMO}, and a moderate-to-high PC2 region characterized by reduced E_{b} alongside substantial Q_{20} and f, indicating fragments with tunable bandgaps, efficient exciton dissociation, strong transition dipoles, and favorable charge-distribution profiles—all traits that support efficient charge separation and robust light absorption. In contrast, cluster 0 (violet) is more spatially dispersed across both components, suggesting a mix of outliers with extreme $\Delta E_{\mathrm{HOMO-LUMO}}$ (very wide gaps), high E_{b}, or misaligned optical *versus* electronic characteristics. These may include unusual or specialized donors that either push novel design boundaries or fall outside the main performance envelope. Finally, cluster 1 (teal)—comprising only two tightly grouped points—reflects highly uniform fragments with narrow variance in $\Delta E_{\mathrm{HOMO-LUMO}}$, E_{b}, Q_{20}, and f, implying rigid, possibly over-engineered donor motifs with limited structural diversity and generalizability. The high cumulative variance captured within the first four components (over 90%) further validates the robustness of the chosen descriptors and the effectiveness of the dimensionality reduction approach.

Acceptor fragments. Following the donor fragment analysis, we performed PCA on the acceptor fragments to explore their underlying property space (Fig. 4c and Table S2). The first principal component (PC1) accounts for 53.6% of the total variance and is primarily governed by the HOMO–LUMO gap, exciton binding energy, and LUMO offset. These interrelated descriptors highlight the coupling between frontier orbital alignment and exciton stability—key to acceptor functionality. As with donors, the variance rapidly accumulates across the first few components, reaching 86.0% by PC3, indicating that the acceptor space is similarly low-dimensional and defined by a few dominant electronic and optical properties. The two-dimensional PCA projection (Fig. 4d), overlaid with k-means clustering (k = 3), reveals three clear fragment clusters. Cluster 0 (violet), located at low PC1 and near-neutral PC2, includes fragments with narrow $\Delta E_{\mathrm{HOMO-LUMO}}$, low E_{b}, and moderate ΔE_{LUMO}—traits that favor efficient exciton dissociation, strong electron-accepting behavior, and favorable energy level alignment. Cluster 2 (yellow) shows moderate $\Delta E_{\mathrm{HOMO-LUMO}}$ and E_{b}, with a wider PC2 spread due to varied ΔE_{LUMO}, suggesting tunable acceptors that require careful optimization. Cluster 1 (teal) is the most dispersed, with high $\Delta E_{\mathrm{HOMO-LUMO}}$ and E_{b}, and highly variable ΔE_{LUMO}, indicating structural outliers that may be less effective for general exciton dissociation but potentially useful in niche roles.

Side chain fragments. PCA was applied to the side-chain fragment space to identify key variations among descriptors (Fig. 4e and Table S3). The first principal component (PC1) explains 54.7% of the variance and is dominated by electron affinity, $\lambda_{\max}$, and Q_{20}, reflecting the importance of frontier orbital energetics, optical absorption, and charge distribution. The variance captured rises sharply to 86.0% by the third component, beyond which additional components contribute minimally—indicating a low-dimensional descriptor space for side chains. A two-dimensional PCA projection (Fig. 4f), colored by k-means clusters (k = 3), reveals three distinct fragment groups.

Cluster 2 (yellow) occupies the moderate-to-high regions of both PC1 and PC2, comprising fragments with higher EA, red-shifted absorption, moderate Q_{20}, and elevated E_b, f, and $\Delta E_{HOMO-LUMO}$—suggesting strong optical activity and electron affinity, though potentially higher exciton binding. Cluster 0 (violet) shows a widespread along PC1—reflecting variation in EA, λ_{max}, and Q_{20}—but remains relatively tightly grouped at low PC2, marked by low E_b, f, and $\Delta E_{HOMO-LUMO}$. This indicates a heterogeneous set of fragments with favorable exciton dissociation and narrow bandgaps, though lower emission probability. Cluster 1 (teal) is mainly located at low PC1—indicative of lower EA, blue-shifted absorption, and smaller Q_{20}—but spans a broad PC2 range, revealing exciton binding, oscillator strength, and bandgap variability. These fragments are weakly electron-withdrawing but exhibit diverse effects on optoelectronic behavior depending on molecular context.

Conjugated extension fragments. PCA was applied to the conjugated extension fragment space to identify orthogonal directions of variance and capture key structure–property relationships (Fig. 4g, Table S4). The first principal component (PC1) explains 56.6% of the variance, primarily influenced by $\Delta E_{HOMO-LUMO}$, ΔE_{LUMO}, and E_b, underscoring their strong coupling. Including PC2 and PC3 raises the cumulative variance to 73.9% and 85.5%, respectively, and 93.0% with the first four components—beyond which the variance plateaus. This suggests that the conjugated extension space lies in a relatively low-dimensional manifold. The elbow point at PC4 supports this as a threshold for dimensionality reduction. The 2D PCA projection (Fig. 4h), combined with $k = 3$ clustering, reveals three distinct fragment clusters. While the elbow appears at PC4, a 3-cluster solution was adopted for consistency across fragment types and is justified by the 85.5% variance explained by the first three PCs. These components offer chemically interpretable insight into the design space without overcomplicating analysis. PC1 primarily encodes variations in $\Delta E_{HOMO-LUMO}$, ΔE_{LUMO}, and E_b while PC2 is shaped by E_b, f, and Q_{20}. Cluster 0 (violet) spans moderate-to-high PC1 with dispersion along PC2. Fragments with moderate PC1 and lower PC2 (*i.e.*, reduced E_b) are promising, offering narrow bandgaps and favorable exciton dissociation. The widespread suggests structural or electronic diversity, including some potential outliers. Cluster 2 (yellow) occupies low-to-moderate PC1 and is widely distributed along PC2. These fragments exhibit low $\Delta E_{HOMO-LUMO}$ and ΔE_{LUMO} but diverse E_b, f, and Q_{20}—likely stemming from differences in conjugation or substitution—making them well-suited for tuning specific optoelectronic properties. Cluster 1 (teal) forms a compact group at low PC1 and moderate-to-high PC2, comprising low-bandgap fragments with moderate E_b, strong optical activity, and consistent Q_{20}. These structures are promising scaffolds for efficient light harvesting and controlled exciton separation—especially if E_b is further optimized. In sum, the high explained variance (93.0% in four PCs) affirms the validity of PCA for describing the conjugated extension design space, supporting data-driven fragment selection.

3.4 Parallel coordinates analysis of molecular fragments

To complement the PCA-based dimensionality reduction and clustering, we employed parallel coordinates plots (PCPs) to unravel the complex, multidimensional property space associated with the four components of ADA-type NFAs—donors, acceptors, side chains, and conjugated extensions. PCPs are especially well-suited for high-dimensional chemical data, offering a visually intuitive yet rigorous framework for observing correlations, trade-offs, and fragment-level design motifs across multiple properties. In these plots, each vertical axis represents a molecular descriptor (*e.g.*, ΔE_{LUMO}, E_b, f), and each line traces the standardized (z-scored) property profile of an individual fragment. Consistent patterns of line bundling signify clusters with shared property signatures,

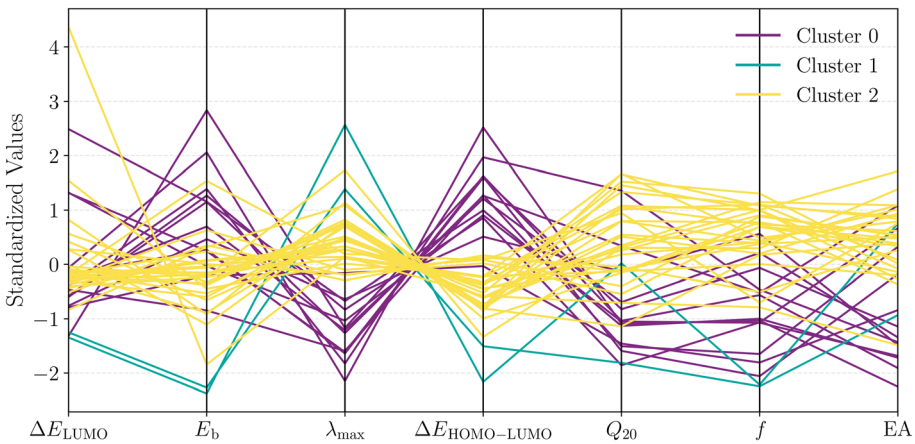

Fig. 5 Parallel coordinates plot representing the distribution of key optoelectronic properties across donor fragments in the ADA-type non-fullerene acceptor dataset. Each line corresponds to an individual donor fragment, and properties are standardized (z-scores) for comparison across different scales. Distinct clusters are highlighted in color, revealing characteristic property patterns and interdependencies. This visualization facilitates the identification of high-performing donor motifs for rational molecular design.

while crossings and divergences denote competing interactions or performance trade-offs.

Across all four fragment classes, PCPs reveal three consistently recurring clusters, each encoding distinct electronic behaviors. For donor fragments, the clusters span a range of energy level alignments and excitonic behaviors (Fig. 5). The yellow cluster exhibits slightly elevated ΔE_{LUMO}, strong oscillator strength, and high EA, with moderately low E_{b} and $\Delta E_{\mathrm{HOMO-LUMO}}$—suggesting good charge-transfer characteristics and strong light absorption. The purple cluster, in contrast, features higher $\Delta E_{\mathrm{HOMO-LUMO}}$ and E_{b} with diminished EA and f, indicating inefficient exciton dissociation and weak absorption. The teal cluster displays low ΔE_{LUMO} and $\Delta E_{\mathrm{HOMO-LUMO}}$ with red-shifted λ_{max} but suffers from low f and EA, pointing to potentially useful bandgap alignment but reduced optical activity. In acceptors, the clusters are even more sharply differentiated, as seen in Fig. S1. The purple cluster stands out with low $\Delta E_{\mathrm{HOMO-LUMO}}$, high f and EA, and red-shifted absorption—an ideal combination for efficient light harvesting and charge separation. The yellow cluster represents an intermediate regime with broader variation in $\Delta E_{\mathrm{HOMO-LUMO}}$ and f, suggesting tunability but with less consistent performance. Meanwhile, the teal cluster shows elevated $\Delta E_{\mathrm{HOMO-LUMO}}$ and E_{b} along with suppressed f, EA, and λ_{max}, making it poorly suited for acceptor applications.

Though often viewed as peripheral, side chains exhibit meaningful property differentiation. As seen in Fig. S2, the yellow cluster shows a somewhat compact distribution with favorable λ_{max}, moderate to low E_{b}, and strong EA, f, and Q_{20}—highlighting their potential for enhancing light absorption and charge distribution. The purple cluster shows moderate-to-low E_{b}, red-shifted λ_{max}, and variable EA, but with lower f and Q_{20}, suggesting weaker emission probability. The teal cluster, by contrast, shows largely unfavorable traits: high $\Delta E_{\mathrm{HOMO-LUMO}}$, elevated E_{b}, and poor absorption metrics. The conjugated extensions profile is displayed on Fig. S3. The teal cluster demonstrates low ΔE_{LUMO} and $\Delta E_{\mathrm{HOMO-LUMO}}$, high EA and f, and red-shifted λ_{max}—traits highly desirable for promoting exciton splitting and photon harvesting. The purple cluster exhibits the opposite—large $\Delta E_{\mathrm{HOMO-LUMO}}$, blue-shifted absorption, and poor f and EA—indicating weak optoelectronic performance. The yellow cluster occupies an intermediate zone with moderate dispersion, suggesting a flexible design space where careful tuning could yield optimized properties.

Collectively, PCPs underscore several key cross-cutting trends: (i) an inverse correlation between $\Delta E_{\mathrm{HOMO-LUMO}}$ and λ_{max} is consistent across fragment types; (ii) electron affinity and oscillator strength show cluster-specific synergies that can be exploited to tune light-harvesting and charge-transfer performance; and (iii) E_{b} and f often operate in tension, particularly in side chains and conjugated extensions, reinforcing the trade-off between strong exciton binding and optical activity. These analyses provide a chemically interpretable view of how fragment-level variations map onto global performance metrics in NFAs. By identifying high-performing fragment clusters—such as the yellow cluster in acceptors and the teal cluster in conjugated extensions—we establish robust design principles for constructing next-generation NFA materials. The most promising fragment clusters identified *via* PCA and PCP were subsequently recombined using the generative protocol outlined in section 2.3. Although our starting set comprises 257 experimentally reported ADA-type NFAs, these molecules were selected to cover all major acceptor families and structural motifs. Automated fragmentation yields several thousand donor, acceptor, extension, and side-chain fragments, each annotated with seven quantum-chemical descriptors. Clustering in this descriptor space reveals that the resulting fragment manifold encompasses the experimentally relevant optoelectronic landscape, allowing for the construction of a 500 000-member library of synthetically plausible NFAs under strict connectivity constraints. This fragment-rich and physically informed representation provides a statistically robust foundation for uncertainty-aware ML screening despite the modest size of the original whole-molecule dataset.

3.5 Evidential MPNN for molecular screening

Although the initial model development began with seven quantum mechanical descriptors, a refined subset of four—oscillator strength (f), maximum absorption wavelength (λ_{max}), LUMO energy-level difference (ΔE_{LUMO}), and exciton binding energy (E_{b})—was ultimately retained based on their relevance and learnability. Among these, f and λ_{max} are traits ideally maximized for enhanced light absorption and efficient charge generation, whereas ΔE_{LUMO} and E_{b} are preferably minimized to facilitate electron transfer and exciton dissociation, respectively. This handpicked selection balances physical insight and model interpretability, enabling a rational screening framework aligned with device performance requirements.

MPNN with evidential learning demonstrates robust predictive performance across these properties, validated using the newly generated diverse library of 500 000 NFAs. An in-depth analysis of property-specific learning dynamics reveals convergence patterns that reflect the complexity of their underlying physical principles. For instance, predictions of ΔE_{LUMO} converge within 110 epochs, with modest oscillations in training loss before settling between 0.2–0.4 (Fig. 6a). E_{b}, being more complex, exhibits early volatility with training losses spiking up to 1.6, gradually stabilizing near 0.4–0.5 after 250 epochs, while test losses remain lower at ~0.2, indicating a challenging but generalizable learning landscape (Fig. 6b). In contrast, f exhibits rapid convergence within 65 epochs, with final training and test losses reaching ~0.13–0.15 and ~0.17–0.20, respectively, highlighting the strong structure–property correlation and high learnability (Fig. 6c). λ_{max} demonstrates intermediate convergence behavior, requiring around 120

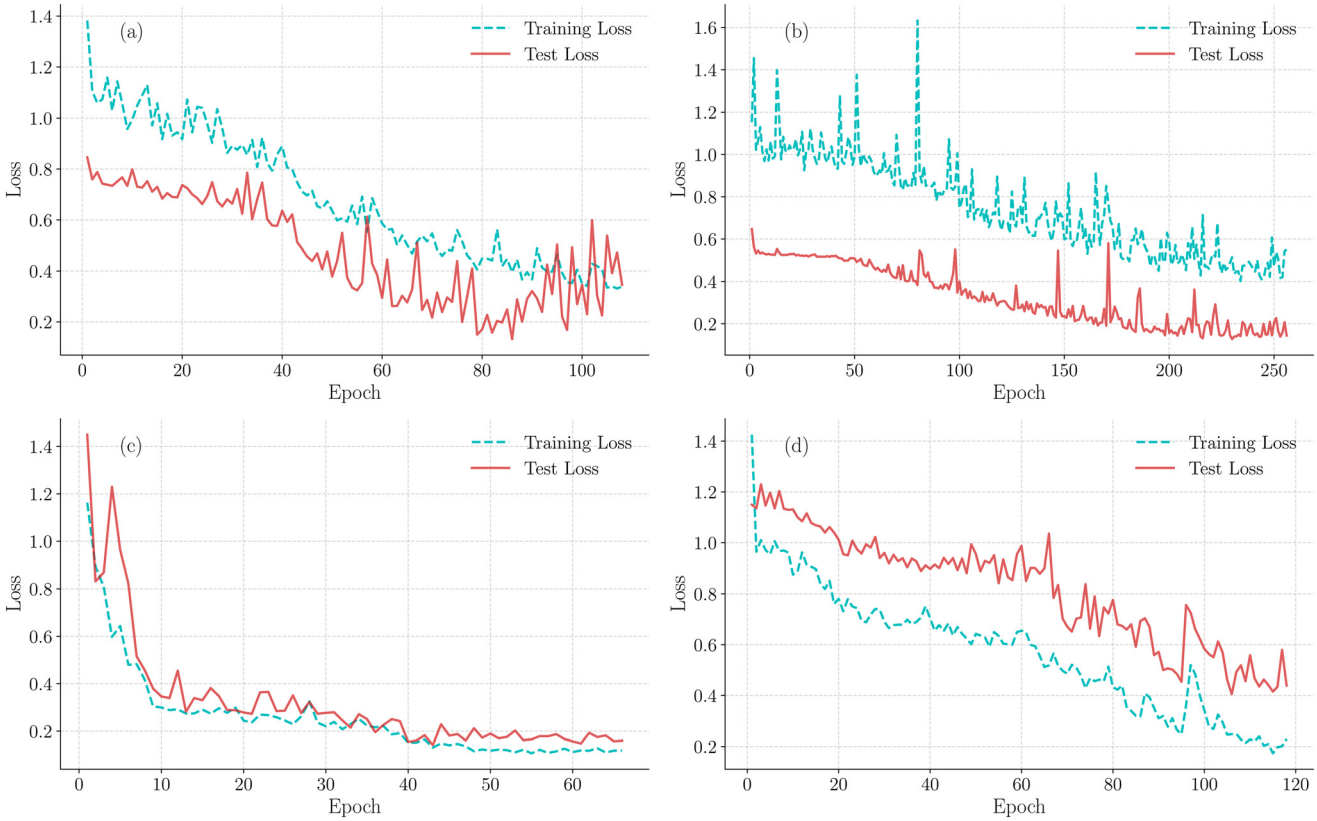

Fig. 6 Training and test loss profiles for the four selected molecular properties: (a) ΔE_{LUMO}, (b) E_b, (c) f, and (d) λ_{max}. The convergence behaviors reflect the relative complexity and learnability of each property, with varying epochs required for stabilization.

epochs to stabilize, with training and test losses of ~0.2 and ~0.4–0.5, respectively (Fig. 6d).

These convergence profiles reflect three common phases across all properties: (1) rapid early learning with sharp loss reduction, (2) property-specific stabilization phases, and (3) final convergence with consistent generalization gaps. The varying training durations (65–250 epochs) and final loss values map coherently onto the complexity hierarchy of the properties, with E_b being most intricate and f the most tractable. The strong generalization performance and absence of overfitting across tasks underscore the effectiveness of our model architecture—featuring six message-passing steps, Set2Set read-out for global feature aggregation, dropout regularization, LayerNorm, and early stopping. The integration of evidential learning further enhances reliability through uncertainty quantification. The model achieves low RMSEs, particularly for E_b (0.03 eV) and ΔE_{LUMO} (0.17 eV), affirming its ability to capture intricate quantum-mechanical phenomena. These results collectively establish the MPNN's robustness and scalability for high-throughput, uncertainty-aware screening in molecular optoelectronics.

The model's performance metrics on the test set, summarized in Table 1, confirm its robust predictive power across all target properties relevant to organic solar cell design. Complementary metrics for the training set are summarized in Table S5. Pearson correlation coefficients range from 0.838 to 0.926, indicating strong agreement between predicted and reference values. Among these, oscillator strength predictions stand out with the highest correlation coefficient of 0.926, along with a low RMSE of 0.438 and MAE of 0.330 (Fig. 7c), highlighting the model's ability to capture the quantum-mechanical characteristics of electronic transitions—critical for identifying efficient light-harvesting materials. The ΔE_{LUMO}, a key metric influencing charge-transfer dynamics, is predicted with a Pearson r of 0.838, RMSE of 0.170 eV, and an impressively low MAE of 0.099 eV (Fig. 7a). Predictions for maximum absorption wavelength also demonstrate strong reliability, with r = 0.841 and RMSE of 48.64 nm (Fig. 7d), despite the inherent complexity of excited-state transitions. Furthermore, E_b is predicted with high precision, yielding r = 0.858, RMSE of

Table 1 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 test set using the evidential MPNN model

Metrics	ΔE_{LUMO}	λ_{max}	f	E_b
RMSE	0.170	48.64	0.438	0.030
MAE	0.099	34.51	0.330	0.019
Pearson r	0.838	0.841	0.926	0.858

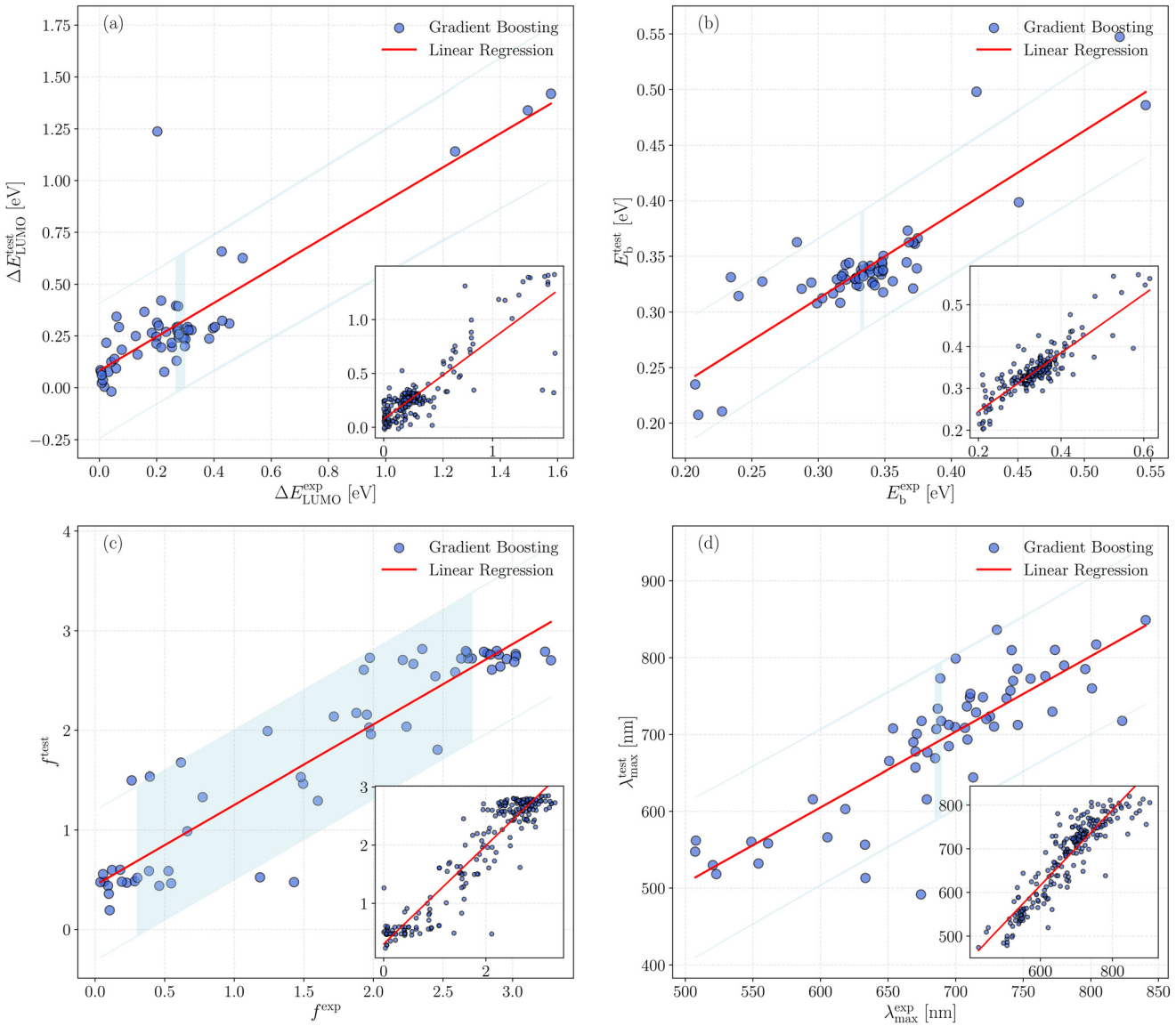

Fig. 7 MPNN models: the plots of predicted properties *versus* their actual values, along with the trained function for both training (inset) and testing steps, illustrate the model's performance across various solar cell properties – (a) ΔE_{LUMO} (b) E_{b}, (c) f, and (d) λ_{max}. Shaded regions around the best-fitted line depict the 95% confidence interval, offering a visual representation of prediction certainty.

0.030 eV, and MAE of 0.019 eV (Fig. 7b), emphasizing the model's sensitivity to fine electronic interactions that govern charge-separation efficiency in OSCs.

Beyond predictive performance, the model significantly enhances high-throughput virtual screening by enabling simultaneous multi-property predictions relevant to molecular design. Its integration of evidential learning not only provides accurate property estimation but also quantifies uncertainty—an essential feature for guiding experimental prioritization. The particularly low RMSEs for E_{b} and ΔE_{LUMO} reflect the model's subtle understanding of quantum-level interactions, while its consistent performance across all properties demonstrates effective extraction of both atomic- and molecular-scale features. Collectively, these strengths position the evidential MPNN as a robust, generalizable, and data-efficient framework for rational discovery of next-generation organic photovoltaic materials.

3.6 Design-space compression and enrichment

Our methodology for designing novel ADA-type non-fullerene acceptors offers a systematic and highly effective route to achieving targeted optoelectronic properties, marking a significant advancement in data-driven materials discovery. By integrating automated molecular fragmentation and structural classification with advanced data analysis—such as clustering and PCA on key descriptors—we identify molecular fragments and connectivity motifs associated with optimal performance, as visualized through 2D PCA and parallel coordinates. A tagging-based combinatorial assembly then

leverages these insights to construct a diverse library of 500 000 new NFA candidates.

The impact of this fragment-based design strategy becomes clear when comparing the property distributions of our computational library with those of experimentally reported molecules used as the training data. A key challenge in high-performance NFA discovery lies in simultaneously optimizing multiple, often competing, electronic properties. Our approach addresses this by coupling intelligent fragment clustering with systematic recombination of high-performing motifs, thereby accessing previously unexplored regions of chemical-property space and significantly broadening the landscape of optimal candidates. Our multi-objective screening strategy is guided by established design principles for high-performing NFAs: (i) minimizing exciton binding energy E_b to reduce the Coulomb barrier for charge separation,[100] (ii) minimizing ΔE_{LUMO} to provide a dense manifold of low-lying CT states that accelerates exciton dissociation,[101,102] and (iii) maximizing oscillator strength f and selecting favorable, red-shifted λ_{max} values to enhance light harvesting and reduce non-radiative voltage losses.[39] These criteria collectively capture the key electronic and optical features known to govern efficient charge generation and high photovoltaic performance in NFA-based OPVs. The contrast in performance metrics exemplifies this: while only three molecules in the training set meet all critical thresholds—$\Delta E_{LUMO} < 0.25$ eV, $E_b < 0.32$ eV, and $f > 1.5$ (Fig. 8a)—our design framework identifies 4812 such candidates, representing an extraordinary 1600-fold enrichment (Fig. 8b). In both panels of Fig. 8, the color bar encodes the oscillator strength (f) associated with each molecule, providing a third dimension to the ΔE_{LUMO}–E_b scatter plots. Warmer colors correspond to larger f values, allowing direct visualization of how the generated NFA library becomes enriched in candidates that simultaneously exhibit low ΔE_{LUMO}, low E_b, and high oscillator strength. This achievement is particularly notable because each parameter governs key aspects of photovoltaic performance: a small LUMO–LUMO+1 gap (ΔE_{LUMO}) Known to facilitate exciton dissociation at the donor–acceptor (D/A) interface, owing to the presence of multiple electron-accepting states in the anionic form;[101,103,104] low exciton binding energy (E_b) enables efficient charge separation;[105,106] and high oscillator strength (f) ensures strong light absorption and radiative efficiency.[106–108] To prevent arbitrary filtering, we anchored the screening thresholds to both the statistics of dataset-I and established design rules in the literature. The highest-performing experimental NFAs cluster around $\Delta E_{LUMO} \approx 0.2$–0.3 eV and $E_b \approx 0.25$–0.35 eV, with strong transitions $f \geq 1.5$. These regions also correspond to the most enriched zones after fragment recombination and MPNN prediction. Prior studies similarly identify $\Delta E_{LUMO} < 0.3$ eV and $E_b \leq 0.3$ eV as key markers of efficient charge separation in high-PCE NFAs,[100–103,109] while strong absorbers typically exhibit $f \geq 1.5$.[39] Accordingly, we selected $\Delta E_{LUMO} \leq 0.25$ eV, $E_b \leq 0.32$ eV, and $f \geq 1.5$ as balanced, literature-consistent thresholds that retain chemically diverse yet optoelectronically competitive candidates.

Our pipeline not only increases the proportion of "triple-hit" candidates ($\Delta E_{LUMO} \leq 0.25$ eV, $E_b \leq 0.32$ eV, $f \geq 1.5$) by more than 1600-fold but also reshapes the entire property landscape to align with the narrow corridor required by state-of-the-art NFAs. Kernel-density overlays in Fig. 9 illustrate this convergence: the experimentally measured ΔE_{LUMO} distribution (0–1.6 eV) collapses to ≈0.30 eV, targeting the <0.4 eV limit that minimizes non-radiative losses; measured E_b values (0.25–0.60 eV) concentrate around ~0.32 eV, close to the ≈0.30 eV threshold for efficient free-charge generation; the absorption maximum (λ_{max}) shifts from 680 nm to a dominant 750–800 nm region, entering the NIR window desirable for tandem and semi-transparent devices; and the low-f tail (<1), which limits photocurrent in the training set, is replaced by a mode at $f \approx 2.2$, consistent with design targets for high J_{SC}.

This sharpening of property distributions arises from the synergy between (i) chemically intuitive fragment recombination—guided by PCA-validated clusters that capture favorable donor–acceptor interactions—and (ii) an

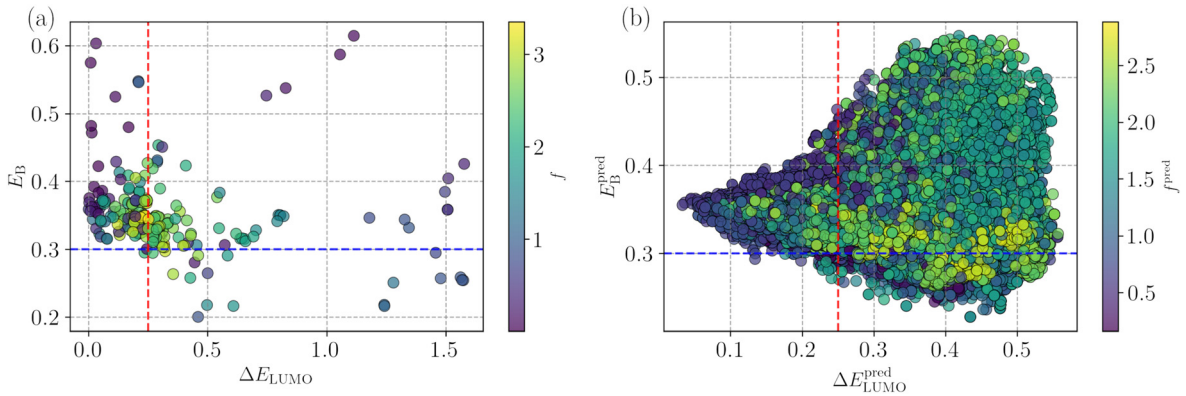

Fig. 8 Scatter plots showing the distribution of three key optoelectronic properties—ΔE_{LUMO}, E_b, and f—in (a) the experimental/training dataset and (b) the generative model's library of 500 000 NFAs. The enriched population in panel (b) highlights the model's ability to discover candidates satisfying all critical thresholds ($\Delta E_{LUMO} < 0.25$ eV, $E_b < 0.32$ eV, and $f > 1.5$), reflecting a dramatic expansion of the optimal property space.

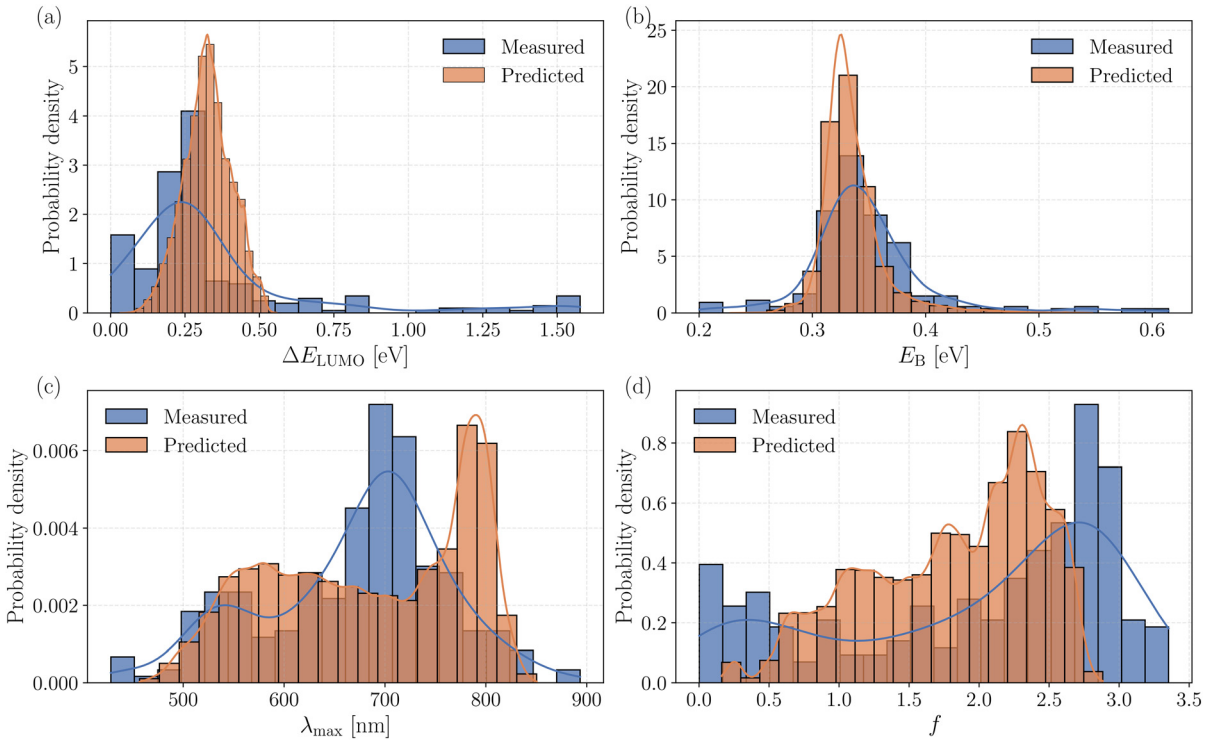

Fig. 9 Kernel-density estimates comparing the distributions of key optoelectronic properties in the experimentally measured dataset (blue) *versus* the predicted molecular library (orange). Panels illustrate the dramatic sharpening and targeted shifts in (a) LUMO–LUMO+1 gap (ΔE_{LUMO}), (b) exciton binding energy (E_b), (c) oscillator strength (f), and (d) absorption maximum (λ_{max}), highlighting convergence toward optimal values for high-performance NFA design.

evidential-MPNN fitness function that incorporates uncertainty-aware prediction while steering optimization toward literature-derived targets for ΔE_{LUMO}, E_b, λ_{max}, and f. By simultaneously compressing the distributions of four orthogonal descriptors, our framework outperforms prior machine-learning-based screens, which typically focus on one or two properties.[69,110] Rather than identifying isolated optima, our method re-engineers the full statistical distribution, transforming a sparse experimental space into a densely populated, application-ready domain and establishing a new benchmark for rational NFA discovery. Further analysis shows that this approach not only saturates the optimal region with unprecedented density but also broadens the accessible wavelength range into the near-infrared (λ_{max} = 750–800 nm), crucial for panchromatic harvesting in next-generation photovoltaics. The concurrent narrowing of the E_b distribution (~0.32 eV) and the abundance of high-f molecules creates a synergistic effect, directly enhancing charge-generation efficiency and reducing recombination losses—key factors in surpassing the 20% power-conversion efficiency milestone in organic solar cells.

What makes this achievement especially significant is that our fragment-based design framework preserves synthetic accessibility while enabling exploration of an expanded optoelectronic property space, thereby bridging the longstanding gap between *in silico* prediction and experimental realization. Recent ML frameworks for OPV design have increasingly adopted fragment- or unit-based representations,[70,71] reflecting a shift toward interpretable and composition-aware molecular design. Building on this direction, our approach introduces automated, chemically meaningful fragmentation of experimentally reported ADA-type NFAs, quantum-chemically informed fragment descriptors, descriptor-space clustering, and strictly constrained recombination governed by tag-encoded connectivity rules. Coupled with an uncertainty-aware message-passing neural network for multi-objective prediction, this framework offers a systematic and synthetically grounded route for navigating NFA chemical space, complementing and extending existing fragment-based ML strategies. The integration of evidential uncertainty quantification ensures confident candidate selection, while the modular assembly enforces chemical realism through experimentally validated connectivity patterns. Together, these elements—synthetic viability, predictive confidence, and interpretable property optimization—establish our methodology as a powerful and generalizable tool for accelerating the discovery of next-generation organic optoelectronic materials.

3.7 Benchmarking against existing fragment-based and ML frameworks

To place our results in context, we compare our methodology with prior fragment-driven and ML-assisted

frameworks for designing NFAs. Earlier studies, such as the fragment-fingerprint and unit-library approaches by Zhang *et al.* and Cao *et al.*, as well as GA/LSTM-based fragment recombination pipelines, have demonstrated the value of decomposing OSC materials into reusable building units.[70–73] However, these methods generally rely on manually picked out donor/acceptor subunits, structural fingerprints, or one-hot fragment encodings, and typically optimize a single scalar metric such as PCE or frontier orbital energies. Moreover, most existing recombination schemes assemble D–π–A or A_1–D–A_2 motifs without imposing chemically rigorous connectivity constraints or integrating quantum-chemical fragment properties, thereby limiting their generalizability across chemical families. In contrast, our workflow begins with fully automated, chemically meaningful fragmentation of 257 experimentally realized ADA-type NFAs, preserving exact connectivity and classifying fragments into donor cores, acceptor units, conjugated extensions, and side chains. Each fragment is characterized by seven quantum-chemical optoelectronic descriptors, enabling a physics-grounded representation rather than a purely symbolic one. Sections 3.2–3.4 show that, unlike earlier methods that cluster fragments using structural heuristics or topological fingerprints, we perform PCA-assisted *k*-means clustering in this quantum-chemical descriptor space to identify fragment families with jointly favorable optoelectronic profiles. Only fragments from these optimal clusters are recombined, and assembly is strictly governed by chemically informed tag-derived rules. SMILES validation, charge/valence checks, and RDKit sanitization ensure structural viability throughout, producing an extensive yet chemically coherent library of 500 000 ADA-type NFAs. Finally, section 3.5, 3.6 and this section compares our evidential message-passing neural network with prior single-target ML models used in fragment-based OSC design. In the screening stage, numerical thresholds were introduced to identify NFAs whose predicted properties fall within a physically and statistically justified optimal regime. Analysis of the 257 experimentally reported NFAs (dataset-I) shows that the highest-performing molecules cluster around $\Delta E_{LUMO} \approx 0.2$–$0.3$ eV, $E_b \approx 0.25$–0.35 eV, and oscillator strengths $f \geq 1.5$. These regions are well supported by OPV design principles that associate small ΔE_{LUMO} and E_b with favorable charge-separation energetics and large f with strong light harvesting.[39,100–103,109] Based on these trends, and allowing for the residual prediction error of the evidential MPNN, we adopt conservative thresholds of $\Delta E_{LUMO} \leq 0.25$ eV, $E_b \leq 0.32$ eV, and $f \geq 1.5$. These criteria are applied only after recombination and prediction, ensuring that the final set of candidates reflects both the statistical structure of dataset-I and established physical considerations relevant to high-efficiency NFAs. Our uncertainty-aware GNN jointly predicts ΔE_{LUMO}, E_b, f, and λ_{max}, providing multi-objective, physics-guided prioritization and achieving an enrichment of high-performing candidates greater than 1600-fold. Together, these elements demonstrate that our integrated framework delivers a substantially more rigorous, physically informed, and

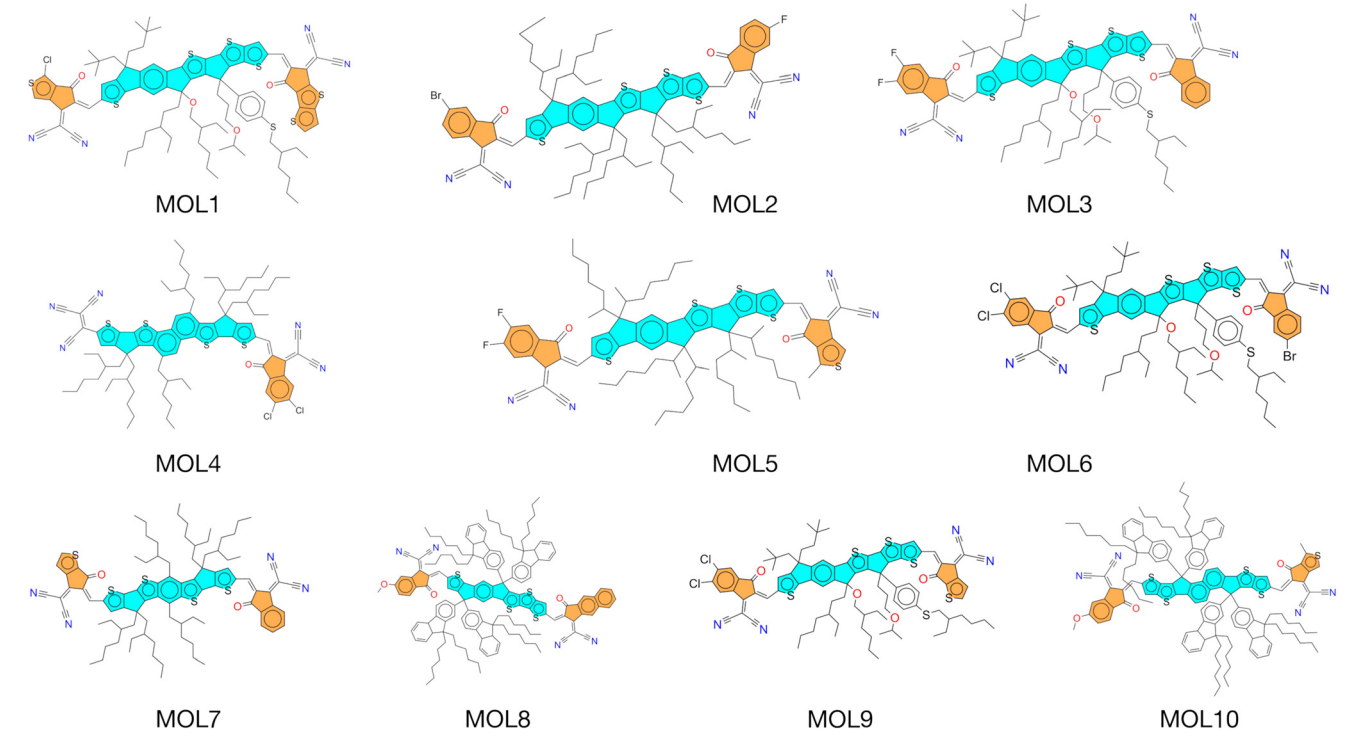

Fig. 10 Molecular structures of the top-performing non-fullerene acceptors identified through our fragmentation-recombination and ML-assisted screening workflow.

Table 2 Comparison of DFT-calculated and MPNN-predicted optoelectronic properties—oscillator strength (f), LUMO energy offset (ΔE_{LUMO}), maximum absorption wavelength (λ_{max}), and exciton binding energy (E_b)—for 10 selected high-performing molecules. The table also summarizes prediction accuracy using RMSE, MAE, and maximum absolute error (MaxAE) metrics

Molecule	f^{DFT}	f^{ML}	ΔE_{LUMO}^{DFT}	ΔE_{LUMO}^{ML}	λ_{max}^{DFT}	λ_{max}^{ML}	E_B^{DFT}	E_B^{ML}
MOL1	2.33	2.21	0.207	0.249	783.9	804.4	0.344	0.317
MOL2	2.84	2.22	0.251	0.232	777.6	796.2	0.337	0.317
MOL3	2.74	2.35	0.264	0.246	767.3	794.4	0.327	0.319
MOL4	2.29	1.92	0.249	0.210	705.7	785.7	0.327	0.319
MOL5	2.91	2.32	0.262	0.247	773.3	810.5	0.335	0.315
MOL6	2.76	2.34	0.259	0.245	781.2	795.5	0.332	0.319
MOL7	2.22	1.96	0.256	0.231	727.8	802.9	0.346	0.314
MOL8	2.70	2.37	0.274	0.246	717.4	666.2	0.327	0.312
MOL9	2.46	2.12	0.282	0.242	771.7	799.7	0.331	0.318
MOL10	2.38	2.36	0.248	0.245	696.5	671.5	0.334	0.316

Performance metrics			
f	RMSE: 0.39	MAE: 0.35	MaxAE: 0.62
ΔE_{LUMO}	RMSE: 0.03	MAE: 0.02	MaxAE: 0.04
λ_{max}	RMSE: 43.77	MAE: 37.68	MaxAE: 80.01
E_b	RMSE: 0.02	MAE: 0.02	MaxAE: 0.03

chemically consistent fragment–property landscape than any previously reported fragment-based ML approach for NFA discovery.

3.8 Quantum chemical validation of predicted properties

To rigorously validate the MPNN model, we performed detailed DFT calculations on a representative subset of 10 top-performing candidate molecules. This selective approach was necessitated by the significant computational cost associated with high-level quantum chemical simulations, which limits their scalability to large datasets. Despite this constraint, the chosen subset captured sufficient structural and electronic diversity, enabling a meaningful assessment of the model's predictive power. The chemical structures of these molecules are displayed in Fig. 10.

The predicted values for key optoelectronic descriptors —ΔE_{LUMO}, λ_{max}, f, and E_b—showed excellent agreement with DFT results, as summarized in Table 2. The RMSE values were notably low: 0.027 for ΔE_{LUMO}, 43.77 nm for λ_{max}, 0.390 for f, and 0.019 for E_b. MAE values further underscored this accuracy: 0.024, 37.68 nm, 0.348, and 0.017, respectively. The model's prediction of ΔE_{LUMO} (0.2071–0.2822 eV) maintained a relative error below 20% across all molecules and demonstrated a strong linear correlation with DFT references. Similarly, predicted λ_{max} values (666.22–810.47 nm) tracked well with DFT trends, with a maximum absolute error of 80.01 nm (≈11.34%). Exciton binding energies were estimated with high fidelity, with relative errors consistently below 10%. The largest deviation—between 0.314 (predicted) and 0.346 (DFT)—amounted to just 9.2%, reflecting the model's ability to capture delicate energetic differences. For oscillator strength, over 80% of the predictions fell within 10% of DFT values, and the largest error remained within ~20%, indicating that most deviations were random rather than systematic.

A comprehensive statistical analysis confirmed that all predicted values deviated by no more than 22% from the corresponding DFT-calculated values. Notably, the model showed particularly high accuracy for molecules with extended π-conjugation, maintaining robust performance across a range of structural complexities. These results collectively demonstrate that the MPNN model can reliably and efficiently approximate expensive quantum mechanical properties, enabling rapid pre-screening of promising molecules for optoelectronic applications. While this study focused on theoretical validation, future work incorporating experimental synthesis and photophysical characterization will be invaluable to fully confirm the model's real-world applicability—although such efforts are beyond the scope of the present investigation.

4 Conclusion

In this work, we introduced a computational framework that integrates cheminformatics, machine learning, and quantum chemical calculations to accelerate the discovery of high-performance non-fullerene acceptors for organic photovoltaics. By combining automated fragmentation, hierarchical clustering, and combinatorial assembly, we generated a synthetically accessible library of 500 000 acceptor–donor–acceptor type NFAs with broad chemical diversity. Principal component analysis and fragment clustering revealed structural motifs that critically influence exciton dissociation, charge transport, and optical absorption.

A message-passing neural network enhanced with evidential learning enabled accurate prediction of key descriptors—oscillator strength (f), LUMO offset (ΔE_{LUMO}), maximum absorption wavelength (λ_{max}), and exciton binding energy (E_b)—while also quantifying predictive uncertainty. Validation against DFT benchmarks confirmed excellent

reliability, with all predictions lying within 22% of reference values. Importantly, the model not only captured property trends but also reshaped the accessible chemical space, compressing broad distributions into narrow regions centered on optimal targets. This deterministic refinement produced a >1600-fold enrichment in "triple-hit" candidates simultaneously satisfying the criteria of $\Delta E_{LUMO} \leq 0.25$ eV, $E_b \leq 0.32$ eV, and $f \geq 1.5$, ultimately yielding 4800 promising NFAs with finely tuned optoelectronic properties.

Unlike prior efforts that optimize individual descriptors in isolation,[69,110] our framework achieves simultaneous multi-objective optimization and leverages fragment-level chemical intuition to ensure interpretability and reproducibility. This integrative strategy establishes a scalable route for rational NFA design, bridging computational predictions with experimental feasibility. Beyond OPVs, the methodology provides a generalizable blueprint for the accelerated discovery of functional organic semiconductors, paving the way for data-driven materials design in optoelectronics and related energy-conversion technologies.

Author contributions

AM conceived the problem. BD led the machine learning model development and carried out the majority of the simulations. KP and ASK contributed to simulations and supported model training and validation. BD and AM analyzed the results and prepared the manuscript draft.

Conflicts of interest

There are no conflicts to declare.

Data availability

The data and code to reproduce the results are available in our GitHub repository (https://github.com/Bib569/MPNN-Frag-ML-Discovery).

Supplementary information (SI): the SI contains parallel coordinate analysis, chemical structures, principal component analysis, and training metrics. See DOI: https://doi.org/10.1039/d5me00182j.

Acknowledgements

The authors gratefully acknowledge the Indian Institute of Technology Gandhinagar, India, for providing research facilities and financial support. We thank PARAM Ananta for computational resources.

References

1 J. Fu, Q. Yang, P. Huang, S. Chung, K. Cho, Z. Kan, H. Liu, X. Lu, Y. Lang, H. Lai, F. He, P. W. K. Fong, S. Lu, Y. Yang, Z. Xiao and G. Li, Rational molecular and device design enables organic solar cells approaching 20% efficiency, *Nat. Commun.*, 2024, **15**, 1830.

2 M. Kim, S. U. Ryu, S. A. Park, Y.-J. Pu and T. Park, Designs and understanding of small molecule-based non-fullerene acceptors for realizing commercially viable organic photovoltaics, *Chem. Sci.*, 2021, **12**, 14004–14023.

3 C. Gu, X. Wang, H. Wang, Y. Tian, J. Ma and R. Yang, Recent advances in small molecular design for high performance non-fullerene organic solar cells, *Mol. Syst. Des. Eng.*, 2022, **7**, 832–855.

4 W. Cao and J. Xue, Recent progress in organic photovoltaics: device architecture and optical design, *Energy Environ. Sci.*, 2014, **7**, 2123–2144.

5 G. Luo, X. Ren, S. Zhang, H. Wu, W. C. Choy, Z. He and Y. Cao, Recent advances in organic photovoltaics: device structure and optical engineering optimization on the nanoscale, *Small*, 2016, **12**, 1547–1571.

6 Y. Cui, H. Yao, L. Hong, T. Zhang, Y. Tang, B. Lin, K. Xian, B. Gao, C. An, P. Bi, W. Ma and J. Hou, Organic photovoltaic cell with 17% efficiency and superior processability, *Natl. Sci. Rev.*, 2020, **7**, 1239–1246.

7 P. Cheng and Y. Yang, Narrowing the band gap: the key to high-performance organic photovoltaics, *Acc. Chem. Res.*, 2020, **53**, 1218–1228.

8 C.-C. Chueh, S.-C. Chien, H.-L. Yip, J. F. Salinas, C.-Z. Li, K.-S. Chen, F.-C. Chen, W.-C. Chen and A. K.-Y. Jen, Toward high-performance semitransparent polymer solar cells: optimization of ultra-thin light absorbing layer and transparent cathode architecture, *Adv. Energy Mater.*, 2013, **3**, 417–423.

9 Y. Sun, L. Meng, X. Wan, Z. Guo, X. Ke, Z. Sun, K. Zhao, H. Zhang, C. Li and Y. Chen, Flexible high-performance and solution-processed organic photovoltaics with robust mechanical stability, *Adv. Funct. Mater.*, 2021, **31**, 2010000.

10 D. Zhang, B. Fan, L. Ying, N. Li, C. J. Brabec, F. Huang and Y. Cao, Recent progress in thick-film organic photovoltaic devices: materials, devices, and processing, *SusMat*, 2021, **1**, 4–23.

11 L. Sun, K. Fukuda and T. Someya, Recent progress in solution-processed flexible organic photovoltaics, *npj Flexible Electron.*, 2022, **6**, 89.

12 X. Xu, K. Fukuda, A. Karki, S. Park, H. Kimura, H. Jinno, N. Watanabe, S. Yamamoto, S. Shimomura, D. Kitazawa, T. Yokota, S. Umezu, T.-Q. Nguyen and T. Someya, Thermally stable, highly efficient, ultraflexible organic photovoltaics, *Proc. Natl. Acad. Sci. U. S. A.*, 2018, **115**, 4589–4594.

13 X. Meng, Z. Xing, X. Hu and Y. Chen, Large-area flexible organic solar cells: printing technologies and modular design, *Chin. J. Polym. Sci.*, 2022, **40**, 1522–1566.

14 K. Liu, T. T. Larsen-Olsen, Y. Lin, M. Beliatis, E. Bundgaard, M. Jørgensen, F. C. Krebs and X. Zhan, Roll-coating fabrication of flexible organic solar cells: comparison of fullerene and fullerene-free systems, *J. Mater. Chem. A*, 2016, **4**, 1044–1051.

15 J. Gao, N. Yu, Z. Chen, Y. Wei, C. Li, T. Liu, X. Gu, J. Zhang, Z. Wei, Z. Tang, X. Hao, F. Zhang, X. Zhang and H. Huang, Over 19.2% efficiency of organic solar cells enabled by precisely tuning the charge transfer state via donor alloy strategy, *Adv. Sci.*, 2022, **9**, 2203606.

16 L. Zhu, M. Zhang, J. Xu, C. Li, J. Yan, G. Zhou, W. Zhong, T. Hao, J. Song, X. Xue, Z. Zhou, R. Zeng, H. Zhu, C.-C. Chen, R. C. I. MacKenzie, Y. Zou, J. Nelson, Y. Zhang, Y. Sun and F. Liu, Single-junction organic solar cells with over 19% efficiency enabled by a refined double-fibril network morphology, *Nat. Mater.*, 2022, **21**, 656–663.

17 D. Li, N. Deng, Y. Fu, C. Guo, B. Zhou, L. Wang, J. Zhou, D. Liu, W. Li, K. Wang, Y. Sun and T. Wang, Fibrillization of Non-Fullerene Acceptors Enables 19% Efficiency Pseudo-Bulk Heterojunction Organic Solar Cells, *Adv. Mater.*, 2023, **35**, 2208211.

18 H. Chen, S. Y. Jeong, J. Tian, Y. Zhang, D. R. Naphade, M. Alsufyani, W. Zhang, S. Griggs, H. Hu, S. Barlow, H. Y. Woo, S. R. Marder, T. D. Anthopoulos, I. McCulloch and Y. Lin, A 19% efficient and stable organic photovoltaic device enabled by a guest nonfullerene acceptor with fibril-like morphology, *Energy Environ. Sci.*, 2023, **16**, 1062–1070.

19 K. Liu, Y. Jiang, F. Liu, G. Ran, F. Huang, W. Wang, W. Zhang, C. Zhang, J. Hou and X. Zhu, Organic Solar Cells with Over 19% Efficiency Enabled by a 2D-Conjugated Non-fullerene Acceptor Featuring Favorable Electronic and Aggregation Structures, *Adv. Mater.*, 2023, **35**, 2300363.

20 L. Zhan, S. Li, Y. Li, R. Sun, J. Min, Y. Chen, J. Fang, C.-Q. Ma, G. Zhou, H. Zhu, L. Zuo, H. Qiu, S. Yin and H. Chen, Manipulating charge transfer and trans-port via intermediary electron acceptor channels enables 19.3% efficiency organic photovoltaics, *Adv. Energy Mater.*, 2022, **12**, 2201076.

21 P. Bi, S. Zhang, J. Ren, Z. Chen, Z. Zheng, Y. Cui, J. Wang, S. Wang, T. Zhang, J. Li, Y. Xu, J. Qin, C. An, W. Ma, X. Hao and J. Hou, A highperformance nonfused wide-bandgap acceptor for versatile photovoltaic applications, *Adv. Mater.*, 2022, **34**, 2108090.

22 A. J. Heeger, 25th anniversary article: bulk heterojunction solar cells: understanding the mechanism of operation, *Adv. Mater.*, 2014, **26**, 10–28.

23 L. Lu, T. Zheng, Q. Wu, A. M. Schneider, D. Zhao and L. Yu, Recent advances in bulk heterojunction polymer solar cells, *Chem. Rev.*, 2015, **115**, 12666–12731.

24 T. Abhijith, R. Suthar and S. Karak, Synergistic Plasmonic Responses of Multi-Shaped Au Nanostructures Hybridized with Few-Layer WS2 Nanosheets for Organic Solar Cells, *ACS Appl. Nano Mater.*, 2023, **6**, 11737–11746.

25 M. L. Keshtov, S. A. Kuklin, I. O. Konstantinov, A. R. Khokhlov, A. Y. Nikolaev, C. Dou, Y. Zou, R. Suhtar and G. D. Sharma, New Conjugated Polymers Based on Dithieno [2, 3-e: 3, 2-g] Isoindole-7, 9 (8H)-Dione Derivatives for Applications in Nonfullerene Polymer Solar Cells, *Sol. RRL*, 2020, **4**, 1900475.

26 X. Zhao, Q. An, H. Zhang, C. Yang, A. Mahmood, M. Jiang, M. H. Jee, B. Fu, S. Tian, H. Y. Woo, Y. Wang and J.-L. Wang, Double Asymmetric Core Optimizes Crystal Packing to Enable Selenophene-based Acceptor with Over 18% Efficiency in Binary Organic Solar Cells, *Angew. Chem., Int. Ed.*, 2023, **62**, e202216340.

27 H.-R. Bai, Q. An, H.-F. Zhi, M. Jiang, A. Mahmood, L. Yan, M.-Q. Liu, Y.-Q. Liu, Y. Wang and J.-L. Wang, A Random Terpolymer Donor with Similar Monomers Enables 18.28% Efficiency Binary Organic Solar Cells with Well Polymer Batch Reproducibility, *ACS Energy Lett.*, 2022, **7**, 3045–3057.

28 A. Armin, W. Li, O. J. Sandberg, Z. Xiao, L. Ding, J. Nelson, D. Neher, K. Vandewal, S. Shoaee, T. Wang, H. Ade, T. Heumüller, C. Brabec and P. Meredith, A history and perspective of non-fullerene electron acceptors for organic solar cells, *Adv. Energy Mater.*, 2021, **11**, 2003570.

29 H.-R. Bai, Q. An, M. Jiang, H. S. Ryu, J. Yang, X.-J. Zhou, H.-F. Zhi, C. Yang, X. Li, H. Y. Woo and J.-L. Wang, Isogenous asymmetric–symmetric acceptors enable efficient ternary organic solar cells with thin and 300 nm thick active layers simultaneously, *Adv. Funct. Mater.*, 2022, **32**, 2200807.

30 R. Suthar, H. Dahiya, A. K. Singh, G. D. Sharma and S. Karak, Role of Exciton Lifetime, Energetic Offsets, and Disorder in Voltage Loss of Bulk Heterojunction Organic Solar Cells, *ACS Appl. Mater. Interfaces*, 2023, **15**, 3214–3223.

31 Z. Zheng, H. Yao, L. Ye, Y. Xu, S. Zhang and J. Hou, PBDB-T and its derivatives: A family of polymer donors enables over 17% efficiency in organic photovoltaics, *Mater. Today*, 2020, **35**, 115–130.

32 G. J. Moore, F. Günther, K. M. Yallum, M. Causa', A. Jungbluth, J. Réhault, M. Riede, F. Ortmann and N. Banerji, Direct visualization of the charge transfer state dynamics in dilute-donor organic photovoltaic blends, *Nat. Commun.*, 2024, **15**, 9851.

33 Y. Dong, R. Zheng, D. Qian, T. H. Lee, H. L. Bristow, P. Shakya Tuladhar, H. Cha and J. R. Durrant, Activationless charge transfer drives photocurrent generation in organic photovoltaic blends independent of energetic offset, *J. Am. Chem. Soc.*, 2024, **146**, 33579–33586.

34 D. B. Riley, O. J. Sandberg, N. Zarrabi, Y. R. Kim, P. Meredith and A. Armin, Efficient Nanoscale Exciton Transport in Non-Fullerene Organic Solar Cells Enables Reduced Bimolecular Recombination of Free Charges, *Adv. Mater.*, 2023, **35**, 2211174.

35 C. C. Chan, C. Ma, X. Zou, Z. Xing, G. Zhang, H.-L. Yip, R. A. Taylor, Y. He, K. S. Wong and P. C. Chow, Quantification of temperature-dependent charge separation and recombination dynamics in non-fullerene organic photovoltaics, *Adv. Funct. Mater.*, 2021, **31**, 2107157.

36 V. Coropceanu, X.-K. Chen, T. Wang, Z. Zheng and J.-L. Brédas, Charge-transfer electronic states in organic solar cells, *Nat. Rev. Mater.*, 2019, **4**, 689–707.

37 J. Wang, Y. Xie, K. Chen, H. Wu, J. M. Hodgkiss and X. Zhan, Physical insights into non-fullerene organic photovoltaics, *Nat. Rev. Phys.*, 2024, **6**, 365–381.

38 J. Wu, H. Cha, T. Du, Y. Dong, W. Xu, C.-T. Lin and J. R. Durrant, A comparison of charge carrier dynamics in organic and perovskite solar cells, *Adv. Mater.*, 2022, **34**, 2101833.

39 J. Yan, X. Rodríguez-Martínez, D. Pearce, H. Douglas, D. Bili, M. Azzouzi, F. Eisner, A. Virbule, E. Rezasoltani, V. Belova, B. Dörling, S. Few, A. A. Szumska, X. Hou, G. Zhang, H.-L. Yip, M. Campoy-Quiles and J. Nelson, Identifying structure–absorption relationships and predicting absorption strength of nonfullerene acceptors for organic photovoltaics, *Energy Environ. Sci.*, 2022, **15**, 2958–2973.

40 S. Li, L. Zhan, N. Yao, X. Xia, Z. Chen, W. Yang, C. He, L. Zuo, M. Shi, H. Zhu, X. Lu, F. Zhang and H. Chen, Unveiling structure-performance relationships from multi-scales in non-fullerene organic photovoltaics, *Nat. Commun.*, 2021, **12**, 4627.

41 S. Oh, Y. Kim, T. Ahn and S. K. Lee, Molecular insights of non-fused nonfullerene acceptor comprising a different central core for high efficiency organic solar cell, *Mol. Cryst. Liq. Cryst.*, 2023, **761**, 68–78.

42 J. Xu, F. Lin, L. Zhu, M. Zhang, T. Hao, G. Zhou, K. Gao, Y. Zou, G. Wei, Y. Yi, A. K.-Y. Jen, Y. Zhang and F. Liu, The crystalline behavior and device function of nonfullerene acceptors in organic solar cells, *Adv. Energy Mater.*, 2022, **12**, 2201338.

43 J. Huang, T. Chen, L. Mei, M. Wang, Y. Zhu, J. Cui, Y. Ouyang, Y. Pan, Z. Bi, W. Ma, Z. Ma, H. Zhu, C. Zhang, X.-K. Chen, H. Chen and L. Zuo, On the role of asymmetric molecular geometry in high-performance organic solar cells, *Nat. Commun.*, 2024, **15**, 3287.

44 M. B. Goldey, D. Reid, J. de Pablo and G. Galli, Planarity and multiple components promote organic photovoltaic efficiency by improving electronic transport, *Phys. Chem. Chem. Phys.*, 2016, **18**, 31388–31399.

45 A. Mahmood and J.-L. Wang, Machine learning for high performance organic solar cells: current scenario and future prospects, *Energy Environ. Sci.*, 2021, **14**, 90–105.

46 X. Rodríguez-Martínez, E. Pascual-San-José and M. Campoy-Quiles, Accelerating organic solar cell material's discovery: high-throughput screening and big data, *Energy Environ. Sci.*, 2021, **14**, 3301–3322.

47 G. Zhang, F. R. Lin, F. Qi, T. Heumüller, A. Distler, H.-J. Egelhaaf, N. Li, P. C. Y. Chow, C. J. Brabec, A. K.-Y. Jen and H.-L. Yip, Renewed prospects for organic photovoltaics, *Chem. Rev.*, 2022, **122**, 14180–14274.

48 B. L. Greenstein and G. R. Hutchison, Organic photovoltaic efficiency predictor: datadriven models for non-fullerene acceptor organic solar cells, *J. Phys. Chem. Lett.*, 2022, **13**, 4235–4243.

49 W. Sun, Y. Zheng, Q. Zhang, K. Yang, H. Chen, Y. Cho, J. Fu, O. Odunmbaku, A. A. Shah, Z. Xiao, S. Lu, S. Chen, M. Li, B. Qin, C. Yang, T. Frauenheim and K. Sun, Artificial Intelligence Designer for Highly-Efficient Organic Photovoltaic Materials, *J. Phys. Chem. Lett.*, 2021, **12**, 8847–8854.

50 Y. Miyake and A. Saeki, Machine learning-assisted development of organic solar cell materials: issues, analyses, and outlooks, *J. Phys. Chem. Lett.*, 2021, **12**, 12391–12401.

51 G.-H. Kim, C. Lee, K. Kim and D.-H. Ko, Novel structural feature-descriptor platform for machine learning to accelerate the development of organic photovoltaics, *Nano Energy*, 2023, **106**, 108108.

52 A. Mahmood, A. Irfan and J.-L. Wang, Machine learning and molecular dynamics simulation-assisted evolutionary design and discovery pipeline to screen efficient small molecule acceptors for PTB7-Th-based organic solar cells with over 15% efficiency, *J. Mater. Chem. A*, 2022, **10**, 4170–4180.

53 K. T. Butler, D. W. Davies, H. Cartwright, O. Isayev and A. Walsh, Machine learning for molecular and materials science, *Nature*, 2018, **559**, 547–555.

54 M. S. Islam, M. T. Islam, S. Sarker, H. Al Jame, S. S. Nishat, M. R. Jani, A. Rauf, S. Ahsan, K. M. Shorowordi, H. Efstathiadis, J. Carbonara and S. Ahmed, Machine learning approach to delineate the impact of material properties on solar cell device Physics, *ACS Omega*, 2022, **7**, 22263–22278.

55 K. Kranthiraja and A. Saeki, Experiment-oriented machine learning of polymer: nonfullerene organic solar cells, *Adv. Funct. Mater.*, 2021, **31**, 2011168.

56 A. Mahmood, A. Irfan and J.-L. Wang, Developing efficient small molecule acceptors with sp2-hybridized nitrogen at different positions by density functional theory calculations, molecular dynamics simulations and machine learning, *Chem. – Eur. J.*, 2022, **28**, e202103712.

57 A. Mahmood, A. Irfan and J.-L. Wang, Machine learning for organic photovoltaic polymers: A minireview, *Chin. J. Polym. Sci.*, 2022, **40**, 870–876.

58 G. Han and Y. Yi, Singlet-triplet energy gap as a critical molecular descriptor for predicting organic photovoltaic efficiency, *Angew. Chem., Int. Ed.*, 2022, **61**, e202213953.

59 Y. Miyake, K. Kranthiraja, F. Ishiwari and A. Saeki, Improved Predictions of Organic Photovoltaic Performance through Machine Learning Models Empowered by Artificially Generated Failure Data, *Chem. Mater.*, 2022, **34**, 6912–6920.

60 Q. Zhao, Y. Shan, C. Xiang, J. Wang, Y. Zou, G. Zhang and W. Liu, Predicting power conversion efficiency of binary organic solar cells based on Y6 acceptor by machine learning, *J. Energy Chem.*, 2023, **82**, 139–147.

61 Z.-W. Zhao, M. del Cueto, Y. Geng and A. Troisi, Effect of increasing the descriptor set on machine learning prediction of small molecule-based organic solar cells, *Chem. Mater.*, 2020, **32**, 7777–7787.

62 R. Suthar, T. Abhijith, P. Sharma and S. Karak, Machine learning framework for the analysis and prediction of energy loss for non-fullerene organic solar cells, *Sol. Energy*, 2023, **250**, 119–127.

63 A. Mahmood, Y. Sandali and J.-L. Wang, Easy and fast prediction of green solvents for small molecule donor-based organic solar cells through machine learning, *Phys. Chem. Chem. Phys.*, 2023, **25**, 10417–10426.

64 C. Liu, L. Lüer, V. M. L. Corre, K. Forberich, P. Weitz, T. Heumüller, X. Du, J. Wortmann, J. Zhang, J. Wagner, L. Ying, J. Hauch, N. Li and C. J. Brabec, Understanding

Causalities in Organic Photovoltaics Device Degradation in a MachineLearning-Driven High-Throughput Platform, *Adv. Mater.*, 2023, 2300259.

65 R. Suthar, T. Abhijith and S. Karak, Machine-learning-guided prediction of photovoltaic performance of non-fullerene organic solar cells using novel molecular and structural descriptors, *J. Mater. Chem. A*, 2023, **11**, 22248–22258.

66 B. Das and A. Mondal, Predictive Modeling and Design of Organic Solar Cells: A Data-Driven Approach for Material Innovation, *ACS Appl. Energy Mater.*, 2024, **7**, 9349–9363.

67 R. Khatua, B. Das and A. Mondal, Physics-informed machine learning with datadriven equations for predicting organic solar cell performance, *ACS Appl. Mater. Interfaces*, 2024, **16**, 57467–57480.

68 Q. Zhang, Y. J. Zheng, W. Sun, Z. Ou, O. Odunmbaku, M. Li, S. Chen, Y. Zhou, J. Li and B. Qin, others High-efficiency non-fullerene acceptors developed by machine learning and quantum chemistry, *Adv. Sci.*, 2022, **9**, 2104742.

69 T. Zhang, J. Yuk Lin Lai, M. Shi, Q. Li, C. Zhang and H. Yan, Data cleansing and sub-unit-based molecular description enable accurate prediction of the energy levels of non-fullerene acceptors used in organic solar cells, *Adv. Sci.*, 2024, **11**, 2308652.

70 C.-R. Zhang, R. Cao, X.-M. Liu, M.-L. Zhang, J.-J. Gong, Z.-J. Liu, Y.-Z. Wu and H.-S. Chen, Designing Donors and Nonfullerene Acceptors for Organic Solar Cells Assisted by Machine Learning and Fragment-Based Molecular Fingerprints, *Sol. RRL*, 2025, **9**, 2400846.

71 R. Cao, C.-R. Zhang, X.-M. Liu, J.-J. Gong, M.-L. Zhang, Z.-J. Liu, Y.-Z. Wu and H.-S. Chen, Molecular design of organic photovoltaic donors and non-fullerene 38 acceptors: a combined machine learning and genetic algorithm approach, *J. Mater. Chem. C*, 2025, **13**, 12150–12168.

72 L.-F. Lv, C.-R. Zhang, R. Cao, X.-M. Liu, M.-L. Zhang, J.-J. Gong, Z.-J. Liu, Y.-Z. Wu and H.-S. Chen, Design and virtual screening of donor and non-fullerene acceptor for organic solar cells using long short-term memory model, *J. Mater. Chem. A*, 2024, **12**, 23859–23871.

73 C.-R. Zhang, L.-F. Lv, M. Li, X.-M. Liu, J.-J. Gong, Z.-J. Liu, Y.-Z. Wu and H.-S. Chen, High throughput molecular design of electron donors and non-fullerene acceptors using machine learning combined with substructure importance, *J. Mater. Chem. C*, 2025, **13**, 14864–14874.

74 G. Landrum, *RDKit: A software suite for cheminformatics, computational chemistry, and predictive modeling*, 2013, vol. 8, p. 5281.

75 D. Trivedi, K. Patrikar and A. Mondal, Graph-Based Networks for Accurate Prediction of Ground and Excited State Molecular Properties from Minimal Features, *Mol. Syst. Des. Eng.*, 2024, **9**, 1275–1284.

76 D. P. Kingma and J. Ba, Adam: A method for stochastic optimization, *arXiv*, 2014, preprint, arXiv:1412.6980, DOI: 10.48550/arXiv.1412.6980.

77 M. Wang, D. Zheng, Z. Ye, Q. Gan, M. Li, X. Song, J. Zhou, C. Ma, L. Yu, Y. Gai, T. Xiao, T. He, G. Karypis, J. Li and Z. Zhang, Deep graph library: A graph-centric, highly-performant package for graph neural networks, *arXiv*, 2019, preprint, arXiv:1909.01315, DOI: 10.48550/arXiv.1909.01315.

78 T. Akiba, S. Sano, T. Yanase, T. Ohta and M. Koyama, Optuna: A next-generation hyperparameter optimization framework, *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*, 2019, preprint, pp. 2623–2631.

79 N. M. O'Boyle, M. Banck, C. A. James, C. Morley, T. Vandermeersch and G. R. Hutchison, Open Babel: An open chemical toolbox, *Aust. J. Chem.*, 2011, **3**, 1–14.

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

81 T. A. Halgren, Merck molecular force field. II., MMFF94 van der Waals and electrostatic parameters for intermolecular interactions, *J. Comput. Chem.*, 1996, **17**, 520–552.

82 T. A. Halgren, Merck molecular force field. III. Molecular geometries and vibrational frequencies for MMFF94, *J. Comput. Chem.*, 1996, **17**, 553–586.

83 T. A. Halgren and R. B. Nachbar, Merck molecular force field. IV. Conformational energies and geometries for MMFF94, *J. Comput. Chem.*, 1996, **17**, 587–615.

84 T. A. Halgren, Merck molecular force field. V., Extension of MMFF94 using experimental data, additional computational data, and empirical rules, *J. Comput. Chem.*, 1996, **17**, 616–641.

85 G. Seifert, D. Porezag and T. Frauenheim, Calculations of molecules, clusters, and solids with a simplified LCAO-DFT-LDA scheme, *Int. J. Quantum Chem.*, 1996, **58**, 185–192.

86 M. Elstner, D. Porezag, G. Jungnickel, J. Elsner, M. Haugk, T. Frauenheim, S. Suhai and G. Seifert, Self-consistent-charge density-functional tight-binding method for simulations of complex materials properties, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 1998, **58**, 7260.

87 M. Gaus, Q. Cui and M. Elstner, DFTB3: Extension of the self-consistent-charge density-functional tight-binding method (SCC-DFTB), *J. Chem. Theory Comput.*, 2011, **7**, 931–948.

88 A. Tkatchenko, R. A. DiStasio Jr., R. Car and M. Scheffler, Accurate and efficient method for many-body van der Waals interactions, *Phys. Rev. Lett.*, 2012, **108**, 236402.

89 A. Ambrosetti, A. M. Reilly, R. A. DiStasio and A. Tkatchenko, Long-range correlation energy calculated from coupled atomic response functions, *J. Chem. Phys.*, 2014, **140**, 18A508.

90 M. Stöhr, G. S. Michelitsch, J. C. Tully, K. Reuter and R. J. Maurer, Communication: Charge-population based dispersion interactions for molecules and materials, *J. Chem. Phys.*, 2016, **144**, 151101.

91 M. Mortazavi, J. G. Brandenburg, R. J. Maurer and A. Tkatchenko, Structure and stability of molecular crystals with many-body dispersion-inclusive density functional tight binding, *J. Phys. Chem. Lett.*, 2018, **9**, 399–405.

92 B. Aradi, B. Hourahine and T. D. Frauenheim, a Sparse Matrix-Based Implementation of the DFTB Method, *J. Phys. Chem. A*, 2007, **111**, 5678–5684.

93 A. H. Larsen, J. J. Mortensen, J. Blomqvist, I. E. Castelli, R. Christensen, M. Du Lak, J. Friis, M. N. Groves, B. Hammer, C. Hargus, E. D. Hermes, P. C. Jennings, P. B. Jensen, J. Kermode, J. R. Kitchin, E. L. Kolsbjerg, J. Kubal, K. Kaasbjerg, S. Lysgaard, J. B. Maronsson, T. Maxson, T. Olsen, L. Pastewka, A. Peterson, C. Rostgaard, J. Schiøtz, O. Schütt, M. Strange, K. S. Thygesen, T. Vegge, L. Vilhelmsen, M. Walter, Z. Zeng and K. W. Jacobsen, The atomic simulation environment—a Python library for working with atoms, *J. Phys.: Condens. Matter*, 2017, **29**, 273002.

94 M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, D. WilliamsYoung, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. A. Montgomery Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, T. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman and D. J. Fox, *09, Revision D. 01*, Gaussian. Inc., Wallingford, CT, 2009.

95 H. Humaira and R. Rasyidah, Determining the appropiate cluster number using elbow method for k-means algorithm, *Proceedings of the 2nd Workshop on Multidisciplinary and Applications (WMA) 2018*, Padang, Indonesia, 2020.

96 A. Casey, Y. Han, Z. Fei, A. J. P. White, T. D. Anthopoulos and M. Heeney, Cyano substituted benzothiadiazole: a novel acceptor inducing n-type behaviour in conjugated polymers, *J. Mater. Chem. C*, 2015, **3**, 265–275.

97 T. C. Parker, D. G. D. Patel, K. Moudgil, S. Barlow, C. Risko, J.-L. Brédas, J. R. Reynolds and S. R. Marder, Heteroannulated acceptors based on benzothiadiazole, *Mater. Horiz.*, 2015, **2**, 22–36.

98 P. Murugan, E. Ravindran, V. Sangeetha, S.-Y. Liu and J. W. Jung, Perylenediimide for organic solar cells: current scenario and prospects in molecular geometric, functionalization, and optoelectronic properties, *J. Mater. Chem. A*, 2023, **11**, 26393–26425.

99 V. Satopaa, J. Albrecht, D. Irwin and B. Raghavan, Finding a "Kneedle" in a Haystack: Detecting Knee Points in System Behavior, *31st International Conference on Distributed Computing Systems Workshops*, 2011, pp. 166–171.

100 L. Zhu, J. Zhang, Y. Guo, C. Yang, Y. Yi and Z. Wei, Small exciton binding energies enabling direct charge photogeneration towards low-driving-force organic solar cells, *Angew. Chem.*, 2021, **133**, 15476–15481.

101 A. Kuzmich, D. Padula, H. Ma and A. Troisi, Trends in the electronic and geometric structure of non-fullerene based acceptors for organic solar cells, *Energy Environ. Sci.*, 2017, **10**, 395–401.

102 T. Liu and A. Troisi, Absolute rate of charge separation and recombination in a molecular model of the P3HT/PCBM interface, *J. Phys. Chem. C*, 2011, **115**, 2406–2415.

103 H. Ma and A. Troisi, Modulating the exciton dissociation rate by up to more than two orders of magnitude by controlling the alignment of LUMO+1 in organic photovoltaics, *J. Phys. Chem. C*, 2014, **118**, 27272–27280.

104 T. Liu and A. Troisi, What makes fullerene acceptors special as electron acceptors in organic solar cells and how to replace them, *Adv. Mater.*, 2013, **25**, 1038–1041.

105 G. Han and Y. Yi, Molecular insight into efficient charge generation in low-drivingforce nonfullerene organic solar cells, *Acc. Chem. Res.*, 2022, **55**, 869–877.

106 R. Khatua, B. Das and A. Mondal, Rational design of non-fullerene acceptors via side-chain and terminal group engineering: a computational study, *Phys. Chem. Chem. Phys.*, 2023, **25**, 7994–8004.

107 Y.-T. Ren, 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, Regulating the Photovoltaic Performance of Organic Solar Cells by Modifying the Y6-Based Non-Fullerene Acceptors: A Quantum Chemistry Study, *Int. J. Quantum Chem.*, 2025, **125**, e70083.

108 F. u. Rehman, M. Waqas, M. Imran, M. A. Ibrahim, J. Iqbal, R. A. Khera, N. Hadia, S. I. Al-Saeedi and M. Shaban, Approach toward low energy loss in symmetrical nonfullerene acceptor molecules inspired by insertion of different π-spacers for developing efficient organic solar cells, *ACS Omega*, 2023, **8**, 43792–43812.

109 R. Khatua and A. Mondal, Design and screening of B–N functionalized non-fullerene acceptors for organic solar cells via multiscale computation, *Mater. Adv.*, 2023, **4**, 4425–4435.

110 S.-P. Peng, X.-Y. Yang and Y. Zhao, Molecular Conditional Generation and Property Analysis of Non-Fullerene Acceptors with Deep Learning, *Int. J. Mol. Sci.*, 2021, **22**, 9099.