ACS APPLIED MATERIALS & INTERFACES

www.acsami.org Research Article

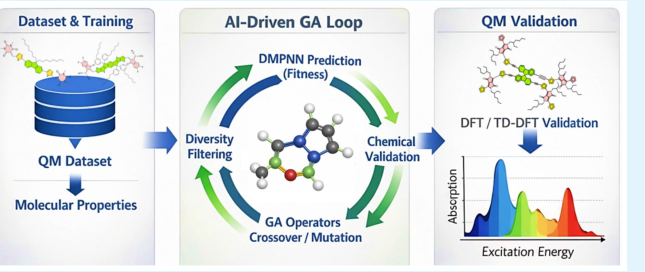
Physics-Guided Inverse Design of Nonfullerene Acceptors via a Deep-Learning-Accelerated Genetic Algorithm

Bibhas Das* and Anirban Mondal*

Cite This: https://doi.org/10.1021/acsami.5c26136 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: The discovery of high-performing nonfullerene acceptors (NFAs) for organic solar cells (OSCs) is challenged by the vastness of chemical space and the need to satisfy multiple tightly coupled electronic criteria. Here, we present a physics-informed generative framework that integrates an evidential message-passing neural network (MPNN) with a constraint-encoded genetic algorithm (GA) to enable inverse molecular design guided directly by quantum-relevant descriptors. Instead of relying on empirical OSC efficiency surrogates, the GA optimizes three key molecular-level properties known to govern charge generation efficiency—oscillator strength (f), exciton binding energy (E_b), and the LUMO−LUMO+1 energy gap (ΔE_{LUMO})—while enforcing structural validity and chemical realism throughout evolution. The combined MPNN−GA workflow efficiently explores a diverse chemical landscape and converges toward synthetically plausible NFAs that satisfy stringent multiobjective constraints. Predicted properties show strong agreement with quantum chemical benchmarks, confirming the reliability of the surrogate model. Pareto analyses further reveal that the generative pipeline captures established quantum-chemical trade-offs and extends the accessible design frontier by identifying candidates that simultaneously exhibit high f, low E_b, and suppressed ΔE_{LUMO}. These results demonstrate a scalable and interpretable approach for physics-driven inverse design of next-generation NFAs, offering a generalizable strategy for molecular discovery in organic electronics.



KEYWORDS: *Nonfullerene Acceptors, Organic Solar Cell, Inverse Molecular Design, Genetic Algorithm, Message-Passing Neural Network, Multi-Objective Optimization, Exciton Binding Energy, Oscillator Strength*

1. INTRODUCTION

Organic solar cells (OSCs) represent a promising renewable energy technology due to their advantages, such as flexibility, low cost, and tunable optical properties. The power conversion efficiency (PCE) of OSCs has significantly improved with the advent of nonfullerene acceptors (NFAs), which offer enhanced light absorption, tunable energy levels, and better stability than traditional fullerene derivatives.[1−10] However, the vast chemical space of potential organic molecules presents a formidable challenge for discovering new, high-performing NFA materials.[11−16]

Traditionally, the discovery of new OSC materials has relied mainly on experimental synthesis or top-down computational approaches. Experimental synthesis is often time-consuming and resource-intensive, making it impractical to explore the immense chemical space of organic molecules.[11−13] Empirical studies and top-down computational methods, such as high-throughput virtual screening (HTVS), aim to predict the properties of a large number of predefined molecules.[16−23] While HTVS can screen millions of candidates, it is inherently limited by the initial library of molecules and often inefficiently expends computational resources on low-performing candidates.[16,22,23] A significant limitation of these data-driven machine learning (ML) approaches is their heavy reliance on comprehensive, high-quality data sets.[16,22,23] Generating such data sets for NFA properties often necessitates expensive and time-consuming quantum mechanical (QM) calculations, including Density Functional Theory (DFT) and time-dependent DFT (TD-DFT), for each molecule.[22,23] Furthermore, these top-down models primarily excel at interpolating within known chemical spaces and struggle to generate truly novel molecular structures, often providing no guarantee of discovering molecules with optimally enhanced properties.[22,23] The 'black box' nature of many machine learning models also hinders the extraction of fundamental design principles, limiting rational molecular design efforts.[11,16,22]

To overcome these limitations, bottom-up generative approaches have emerged as powerful tools for exploring chemical space by generating entirely new molecular structures

Received: December 28, 2025
Revised: February 17, 2026
Accepted: February 23, 2026

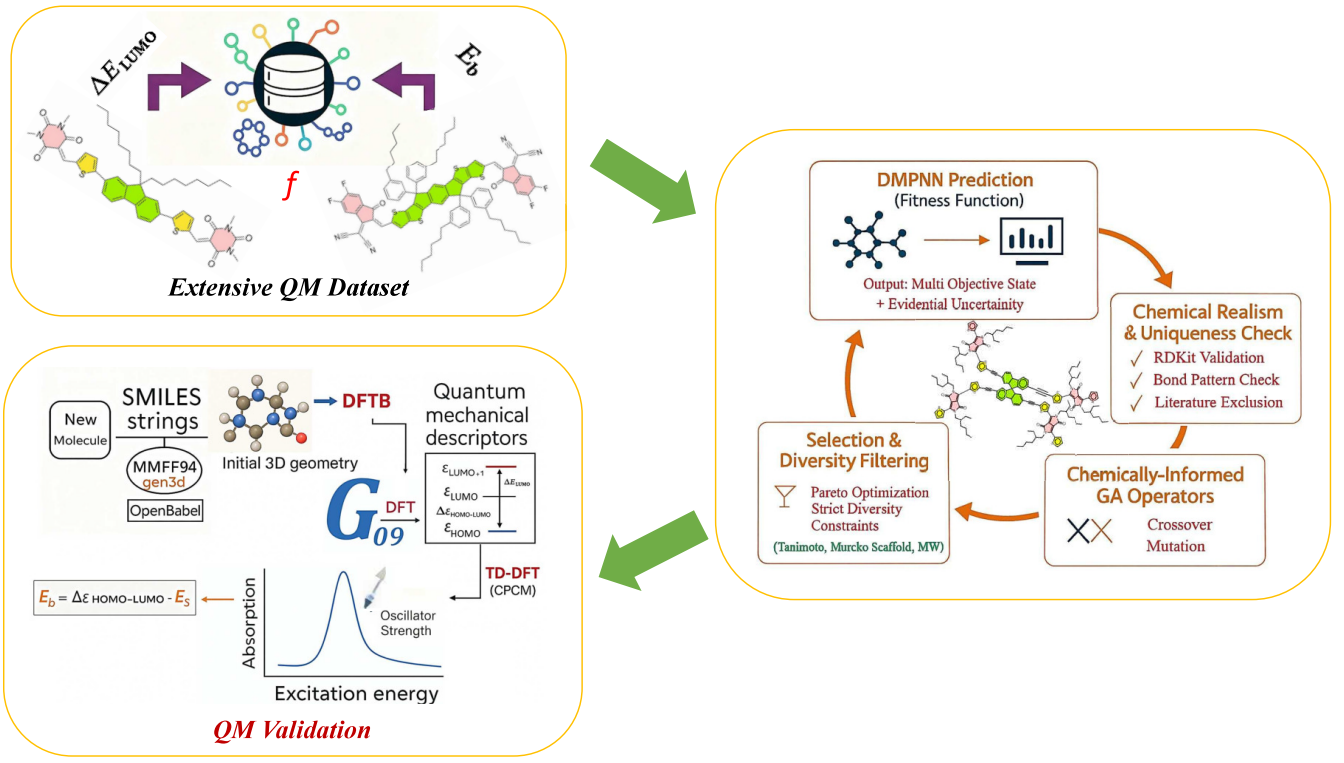


Figure 1. Schematic representation of the genetic algorithm-based molecular discovery framework. An extensive quantum-mechanical (QM) data set provides key optoelectronic descriptors, which serve as inputs for model training and validation. Candidate molecules are encoded into genetic representations and evolved through selection, crossover, and mutation operations to explore chemical space efficiently. The fitness of generated structures is evaluated using a surrogate model trained on QM data, ensuring accurate prediction of properties such as oscillator strength, exciton binding energy, and LUMO-level gaps. Top-performing candidates undergo further QM validation to confirm predicted trends and refine the structure−property landscape for high-performance organic semiconductors.

guided by fitness-driven optimization. Among these, genetic algorithms (GAs), inspired by natural selection, have proven particularly effective for molecular design. Several studies have demonstrated their potential in the context of OSCs. Greenstein et al. employed a GA to identify high-performing unfused NFAs with predicted PCEs exceeding 18%, achieving enormous speedups compared to brute-force searches and revealing structural design rules favoring specific donor cores and acceptor end groups.[24] In subsequent work, they extended this approach for tandem OSCs, incorporating a multistep GA and ensemble ML model (OPEP2) to optimize NFA−donor combinations, achieving predicted PCEs up to 20.8%.[25] Similarly, Cao et al. coupled ML and GA optimization using Random Forest models to design donor−acceptor pairs with predicted PCEs of ∼16.8%, demonstrating the effectiveness of integrated ML−GA frameworks in accelerating molecular discovery.[26] Beyond GAs, neural network-based generative models such as convolutional neural networks (CNNs) and attention-based graph neural networks (GNNs) have also been explored to directly generate novel NFAs with tailored energy levels, significantly broadening chemical design capabilities.[27] Chen et al. further demonstrated the scalability of CNN-based generative models by designing 12,224 novel donor−acceptor pairs predicted to exceed 19% PCE, with a maximum predicted efficiency of 19.20%, underscoring the potential of deep generative networks for high-performance OSC discovery.[28] Parallel efforts have advanced end-to-end pipelines integrating large-scale chemical space enumeration with data-driven screening and physics-informed filtering. Zhang et al. employed fragment-level fingerprints with Random Forest and Extra Trees regression to construct over 24 billion donor−acceptor combinations, identifying candidates with predicted PCEs up to 13.2%.[29] In subsequent work, they curated 547 experimental donor−acceptor pairs and used Morgan/MACCS fingerprints with Random Forest and SHAP analysis to guide recombination, generating ∼3.45 billion pairs, of which >14,000 exceeded 14% PCE and 123 surpassed 15.5% (max ∼ 15.9%).[30] Complementary deep-learning frameworks further reinforced these trends. Cao et al. combined LSTM-based PCE prediction with fragment recombination to generate ∼7,600 candidates above 18% PCE[31] whereas Lv et al. integrated generative LSTM models with symbolic regression, assembling ∼185 billion donor−acceptor pairs and identifying 5,753 candidates above 18.5% PCE while extracting interpretable structure−property relationships beyond purely black-box models.[32] Reinforcement learning approaches, such as Qiu et al.'s transformer-based framework, further expanded the design space, reporting acceptors with predicted PCEs exceeding 21%.[33]

Despite these advances, existing generative approaches face critical challenges. Many workflows rely on empirical PCE surrogates with limited physics-based constraints, which can introduce bias and decouple molecular design from underlying photophysical processes.[24−26] In addition, a lack of integrated chemical validity checks often results in synthetically infeasible or unstable molecular candidates. Furthermore, current frameworks generally optimize a single property and rarely address the simultaneous tuning of multiple, interdependent molecular parameters critical for efficient charge generation and transport.[24−27]

https://doi.org/10.1021/acsami.5c26136
ACS Appl. Mater. Interfaces XXXX, XXX, XXX−XXX

This work presents a constrained bottom-up genetic algorithm framework for the rational design of new nonfullerene acceptors for OSCs. The algorithm directly integrates chemical and physical constraints into evolutionary operations—selection, crossover, and mutation—to ensure the creation of chemically valid and synthetically meaningful molecular structures. Unlike traditional PCE-driven fitness functions, our strategy optimizes molecular-level descriptors directly linked to device performance, including oscillator strength (f), exciton binding energy (E_b), and the LUMO−LUMO+1 energy gap (ΔE_{LUMO}). These three descriptors were selected based on well-established design principles for high-performing NFAs: minimizing E_b reduces the Coulomb barrier for charge separation;[34−36] suppressing ΔE_{LUMO} provides a dense manifold of low-lying charge-transfer states that accelerates exciton dissociation;[37−40] and maximizing f enhances photophysical brightness and mitigates nonradiative voltage losses.[35,41−43] To avoid arbitrary filtering, the screening thresholds were chosen to align with trends consistently reported in the NFA literature, where efficient acceptors typically exhibit $\Delta E_{LUMO} \approx 0.2-0.3$ eV,[37,38,40,44] $E_b \approx 0.25-0.35$ eV,[35,36] and strong transitions with $f \geq 1.5$.[41] Accordingly, we adopt $\Delta E_{LUMO} \leq 0.2$ eV, $E_b \leq 0.28$ eV, and $f \geq 2.0$ as balanced, literature-supported constraints that preserve chemical diversity while capturing optoelectronic performance characteristics essential for high-PCE OSCs. DFT calculations further validate these properties, providing an *ab initio* foundation for the generative workflow. Overall, this integrated, physics-aware GA framework provides a robust and efficient route for discovering chemically sound and functionally optimized NFA molecules, advancing the design of next-generation high-efficiency organic solar cells.

2. METHOD

A compiled data set of 300 nonfullerene acceptors (referred to as Data set-I) forms the basis of this study. This data set, developed in our earlier work,[22] was compiled from a larger collection of NFAs reported in the literature.[16] The selection was designed to capture a broad structural diversity while maintaining a representative distribution of electronic and optical properties relevant to organic photovoltaics. For each molecule, key quantum-mechanical descriptors were extracted, including oscillator strength, exciton binding energy, and the LUMO−LUMO+1 energy difference—parameters directly linked to charge separation and transport efficiency in OSCs. Although the compiled data set comprises 300 nonfullerene acceptors, recent studies have demonstrated that graph neural networks can achieve robust performance in similarly sized chemical and materials data sets when architectural complexity is carefully controlled, and strong regularization is employed.[45−47] Accordingly, we adopted a shallow message-passing neural network (MPNN)-style architecture with constrained hidden dimensions to limit model capacity, combined with dropout, weight decay, and validation-based early stopping to mitigate overfitting. Hyperparameters were systematically optimized using Optuna-based Bayesian search (100 trials per target property), ensuring a balanced trade-off between expressivity and generalization in this low-data regime. While larger data sets would further improve statistical robustness, the present framework is consistent with established best practices for data-efficient graph learning. The overall computational workflow, including data set preparation, model construction, and genetic algorithm optimization, is summarized in Figure 1.

2.1. Model Development-Evidential MPNN Property Predictors

2.1.1. Data Preparation and Graph Construction. Molecules are provided as SMILES strings and are first filtered through a strict validity pipeline that rejects disconnected structures, malformed bracket/parentheses patterns, and chemically invalid graphs (RDKit[48] parsing must succeed), followed by canonicalization to ensure consistent representations. Each accepted SMILES is converted into a molecular graph using the `Deep Graph Library` (DGL) and `dgllife` utilities,[49] where atoms and bonds are featurized using the canonical atom/bond featurizers and mapped to a directed bigraph representation. Target properties are standardized independently using per-property `StandardScaler` objects, enabling stable training across potentially different property scales while retaining invertibility for physical-unit reporting during inference.

We employed a message-passing neural network to predict molecular properties directly from graph representations derived from SMILES strings. In this framework, atoms and bonds are encoded using canonical featurization, and iterative message passing propagates information along chemical bonds to capture both local chemical environments and longer-range conjugation effects relevant to nonfullerene acceptors. A `Set2Set` readout layer generates permutation-invariant graph-level embeddings for variable-sized molecules. This end-to-end graph-based approach avoids handcrafted or quantum-chemical descriptors, enabling computationally efficient and scalable inference suitable for integration with the genetic algorithm optimization loop. Hyperparameters were optimized using Optuna-based[50] Bayesian search (100 trials per property), and the final tuned parameters for each target (f, E_b, and ΔE_{LUMO}) are explicitly reported in Table S1 of Supporting Information. We emphasize, however, that hyperparameters optimized with Optuna are not guaranteed to be globally optimal or stable across independent runs, as its stochastic samplers (e.g., TPE, random sampling, CMA-ES) may explore different regions of the search space. The process can also overfit to a specific validation split, especially with limited data, and performance may vary due to random initialization, data shuffling, and retraining variability. With a finite number of trials, selected configurations may partly reflect statistical noise. For stricter reproducibility, fresh hyperparameter searches or repeated cross-validation (e.g., K-fold CV within Optuna) are recommended. In this work, however, we prioritize dynamic deployment of optimized models within the GA framework for on-the-fly property prediction, rather than exhaustive hyperparameter stabilization.

2.1.2. MPNN Architecture with Evidential Outputs. For each target property, an MPNN is trained using the `dgllife` MPNN backbone (`MPNNGNN`) to compute node-level embeddings, followed by a `Set2Set` readout to obtain graph-level representations. The predictor head maps the readout features to evidential parameters by outputting $4 \times n_{tasks}$ values and splitting them into $(\mu,\lambda,\alpha,\beta)$ per task; positivity and validity constraints are enforced using `Softplus` transforms (with α shifted by +1), yielding a Normal-Inverse-Gamma−style parametrization that can be used to derive both predictions and uncertainty estimates. For optimization and evaluation in this implementation, the mean term μ is used as the primary point prediction, while the additional evidential parameters enable uncertainty quantification when needed for downstream auditing or reliability analysis.

https://doi.org/10.1021/acsami.5c26136
ACS Appl. Mater. Interfaces XXXX, XXX, XXX−XXX

2.1.3. Training Protocol and Bayesian Hyperparameter Optimization. The compiled data set (300 NFAs) was partitioned using an 80:20 training−test split. Owing to the limited data set size, a fixed three-way split would substantially reduce the effective training set and yield a statistically small validation subset. Instead, hyperparameter optimization was performed using Optuna-based Bayesian search (100 trials per property), with each trial using a newly generated 80:20 split and a different random seed. This repeated splitting strategy reduces partition bias and provides more robust model selection than a single fixed validation set. Early stopping with optimized patience was applied during training to further mitigate overfitting.

All models were trained using the Mean Squared Error (MSE) loss function, a standard objective for regression tasks. The loss is defined as

$$L = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2$$

where N is the batch size, y_i represents the reference molecular property, and $\hat{y}_i$ denotes the corresponding prediction from the MPNN. Separate single-task models were trained for each target property (f, E_b, and ΔE_{LUMO}). Molecular graph embeddings generated through message passing and aggregated via the `Set2Set` readout layer were passed to regression heads, and parameters were optimized using the Adam optimizer to minimize MSE between predicted and standardized target values. Early stopping based on held-out performance was applied to mitigate overfitting.

Separate models are trained for each property (e.g., f, E_b, ΔE_{LUMO}) using mini-batched DGL graph loaders and the Adam optimizer with weight decay. The Bayesian search procedure tuned both architectural and training parameters including hidden dimensions, message passing and readout depth, regularization (`dropout`, `weight decay`), learning rate, batch size, random seed, epoch budget, and early stopping patience. The best configuration for each property is saved to a property-specific file, and the resulting trained network is checkpointed along with optimizer state, learning curves, and scalers to ensure full reproducibility and consistent inference in the subsequent evolutionary stage.

Before any new tuning/training is triggered, the workflow checks for the existence and readability of the required checkpoint artifacts (hyperparameter file and model checkpoint) per property. If these files are present and valid, the model is loaded and Optuna is skipped; otherwise, optimization and retraining are automatically executed, guaranteeing that the GA always uses well-defined and traceable predictors.

2.2. Evolutionary Molecular Generation Guided by Property Predictors

2.2.1. Overview and Integration of Predictors into the GA Loop. The generative stage couples the trained property predictors with a genetic algorithm implemented for multiobjective molecular optimization. At each evaluation, candidate molecules are converted to DGL graphs and passed through the property-specific MPNN models; predictions are inverse-transformed back to physical units using the stored scalers. Fitness is computed as a bounded, monotonic multiobjective score combining the three photovoltaic targets and an additional NFA design-quality heuristic, enabling direct ranking of molecules and stable cross-generation comparisons.

Molecular generation was performed using a multiobjective, diversity-aware genetic algorithm. The initial population (200−3000 molecules) was seeded from the curated NFA data set and diversified through fragment recombination and NFA-inspired structural variations to ensure broad chemical coverage. Fitness was evaluated using a bounded, weighted multiobjective scoring function in which predicted f, E_b, and ΔE_{LUMO} were individually mapped through logistic functions (30% weight each) and combined with a domain-informed NFA design heuristic (10%), yielding normalized scores in the [0,1] range. Novelty was enforced using canonical SMILES/InChI representations and Tanimoto similarity thresholds relative to the training set. Parent selection incorporated diversity constraints based on Morgan fingerprint similarity, Murcko scaffold differentiation, and physicochemical variation, alongside elitism to preserve top-performing individuals. Crossover was implemented via chemically meaningful fragment-based recombination using BRICS decomposition, maintaining donor−acceptor motifs and conjugated backbones. Mutation followed a hierarchical strategy, prioritizing NFA-specific structural modifications (e.g., donor/acceptor tuning, side-chain engineering, conjugation adjustment) and general chemical transformations (e.g., substituent replacement and heteroatom substitution), with adaptive mutation rates applied during evolution. Diversity metrics were continuously monitored, and exploratory candidates were injected when necessary to prevent premature convergence. Evolution was conducted for 50 generations with complete logging of population statistics and structural annotations to ensure reproducibility. A detailed section-wise description of each component of the GA workflow is provided below.

2.2.2. Population Initialization and Novelty Bookkeeping. The initial population is seeded from the curated training set (canonical SMILES, duplicates removed) and then aggressively expanded to increase the chemical baseline diversity via training-data-driven variations and NFA-inspired generation strategies. Throughout evolution, the algorithm maintains strict novelty constraints: a global registry of previously seen SMILES, canonical SMILES and InChI-key checks for duplicate prevention, and similarity-based filtering against training structures (Tanimoto thresholds) to prevent rediscovery of training compounds or near-identical variants.

2.2.3. Fitness Function with Physics-Informed Logistic Scaling. The genetic algorithm optimization is driven by a weighted, normalized multiobjective fitness function defined as

$$F = w_1 S_f + w_2 S_{E_b} + w_3 S_{\Delta E_{LUMO}} + w_4 S_{NFA}$$

where $w_1 = 0.30$, $w_2 = 0.30$, $w_3 = 0.30$, and $w_4 = 0.10$. The property-based scores are obtained using logistic transformations to ensure bounded and monotonic scaling:

$$S_f = \sigma[2.0(f - 2.0)]$$

to promote high oscillator strength,

$$S_{E_b} = \sigma[3.0(0.28 - E_b)]$$

to favor low exciton binding energy, and

$$S_{\Delta E_{LUMO}} = \sigma[3.0(0.2 - \Delta E_{LUMO})]$$

where $\sigma(x) = (1 + e^{-x})^{-1}$. These transformations map predicted deviations from desired target values into normalized scores in the range [0,1], enabling balanced multiobjective optimization. The fourth term S_{NFA} represents a normalized structural design score based on chemically informed heuristics inspired by high-

https://doi.org/10.1021/acsami.5c26136
ACS Appl. Mater. Interfaces XXXX, XXX, XXX−XXX

performance nonfullerene acceptors. Property values (f, E_b, and ΔE_{LUMO}) are predicted using independently trained MPNN models, and duplicate structures are removed using canonical SMILES and InChI identifiers. This formulation ensures smooth, generation-independent evolutionary pressure and interpretable trade-offs among competing objectives. We note that the predicted oscillator strength is an intrinsic per-molecule quantity, whereas actual film absorption scales approximately with the molecular number density, i.e., $\alpha_{int} \propto (\rho/MW)\, f$ (or f/V_{mol}).[51−53] In this work, f is incorporated through a saturating scoring function to avoid disproportionate reward of excessively large multichromophore structures. We therefore interpret high f as a necessary intrinsic prerequisite for strong absorption, while recognizing that realistic device performance additionally depends on mass/volume normalization, packing, and morphology effects beyond the scope of the present screening framework.

Fitness is therefore computed from predicted property values using logistic (sigmoid) mappings centered on explicit physical targets ($f > 2.0$, $E_b < 0.28$ eV, $\Delta E_{LUMO} < 0.2$ eV), yielding component scores in [0,1] that saturate smoothly as candidates improve. The final objective aggregates these components with explicit weights (30% each for f, E_b, ΔE_{LUMO}, and 10% for an NFA design-quality score), producing a single bounded fitness used for selection and elitism while still reflecting multiobjective optimization priorities. In addition, the implementation tracks per-generation compliance rates with each target and logs top candidates with their predicted property breakdowns for transparency.

2.2.4. Selection, Diversity Constraints, and Elitism. Parent selection follows a hybrid fitness−diversity strategy that balances the exploitation of high-performing candidates with the preservation of structural diversity. At each generation, all molecules are ranked by their composite fitness score. A parent pool (approximately 50% of the target population size) is constructed by iteratively selecting high-fitness candidates that simultaneously satisfy predefined diversity criteria relative to previously selected parents, including fingerprint-based Tanimoto similarity thresholds (RDKit and Morgan fingerprints), distinct Murcko scaffolds, and sufficient molecular weight variation. If strict diversity constraints limit pool formation, thresholds are adaptively relaxed to ensure adequate parent availability. To promote directional improvement across objectives, property-specialized elitism is incorporated by explicitly retaining top-performing individuals for each target property (f, E_b, and ΔE_{LUMO}) as well as the highest overall fitness candidates (top ∼10%), which are directly propagated to the next generation. During offspring generation, parents are sampled uniformly at random from this curated parent pool, ensuring balanced reproductive opportunity while maintaining the fitness- and diversity-filtered structure of the population. This multilayer selection framework enables broad exploration in early generations and progressively stronger exploitation as the search converges, preventing premature collapse of chemical diversity while sustaining evolutionary pressure toward optimal regions of chemical space.

2.2.5. Crossover via Chemically Meaningful Fragment Recombination. Crossover is performed through fragment-based recombination that decomposes parents into chemically meaningful fragments (BRICS-based operations) and recombines them to preserve key NFA structural motifs such as donor cores, acceptor patterns, and conjugated backbones. The crossover probability is adaptive (approximately 60−80% depending on whether the population already contains successful candidates), balancing exploitation of high-performing motifs with continued exploration.

2.2.6. Mutation Operators with Adaptive Schedules and Entropy Regulation. Mutation is implemented as a hierarchical operator set. A majority of mutation events are NFA-specific, including template- and literature-inspired NFA construction and component-level substitutions targeting donor cores, acceptor motifs, side chains, conjugation patterns, and overall complexity. The remainder uses general medicinal/organic transformations such as substituent edits, functional-group insertions, heteroatom swaps, ring modifications, and coupling-reaction simulations. The mutation rate is annealed using a logistic schedule (high early to promote exploration, decaying toward a nonzero floor for refinement), and multiple mutation applications per offspring are used adaptively (more when no candidates meet constraints; fewer once successful designs emerge).

In this work, entropy is defined as a quantitative measure of population diversity within the genetic algorithm and is conceptually equivalent to population entropy. Specifically, entropy is computed as a normalized composite diversity index bounded in [0, 1], integrating multiple structural and physicochemical diversity descriptors, including RDKit and Morgan fingerprint-based Tanimoto dissimilarity, Murcko scaffold diversity (fraction of unique scaffolds), MACCS key diversity, and the coefficients of variation for molecular weight and LogP. This metric provides an operational measure of structural heterogeneity in chemical space at each generation. Entropy is dynamically regulated throughout evolution to balance exploration and exploitation. Early generations maintain high entropy through elevated mutation rates and diversity-driven selection pressure. In contrast, later generations progressively reduce entropy via mutation-rate annealing, increasing fitness acceptance thresholds, and refinement-focused crossover. Additionally, diversity-triggered replenishment mechanisms inject highly mutated candidates when entropy falls below a predefined target trajectory, preventing premature convergence. Together, this framework establishes controlled entropy management as a quantitatively defined and adaptively enforced diversity regulation strategy rather than a qualitative heuristic.

2.2.7. Quality-Gated Acceptance, Replenishment, and Termination. Candidate offspring are accepted only if they pass SMILES validity checks, chemical sanity checks, and strict uniqueness constraints, and they additionally satisfy a progress-dependent quality gate based on fitness and per-property logistic component thresholds to prevent population drift toward low-quality regions. If the population size drops below a defined fraction of the target, automated replenishment mechanisms are triggered to restore diversity and maintain search momentum. The evolutionary process runs for a fixed number of generations (50 in this implementation), with comprehensive logging of generation-level statistics, diversity metrics, and archiving of successful molecules and design annotations, thereby ensuring reproducibility and interpretability of the GA-driven molecular discovery process. The detailed technical implementation plan is described in Section S1 of the Supporting Information (SI).

2.3. First-Principles Validation of Generated NFAs

An extensive quantum-chemical benchmarking was conducted to validate the reliability of the machine-learning-generated NFA candidates. The goal was to ensure that the predicted

https://doi.org/10.1021/acsami.5c26136
ACS Appl. Mater. Interfaces XXXX, XXX, XXX−XXX

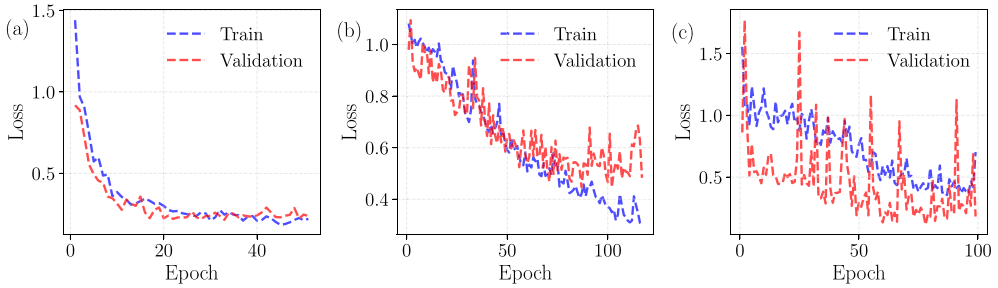


Figure 2. Training and validation learning curves for MPNN models predicting key optoelectronic descriptors: (a) oscillator strength (f), (b) exciton binding energy (E_b), and (c) LUMO−LUMO+1 gap (ΔE_{LUMO}). The rapid and stable convergence for f, gradual yet consistent learning for E_b, and higher but controlled variance for ΔE_{LUMO} highlight descriptor-specific complexity while collectively confirming robust and generalizable model performance for guiding GA-driven molecular design.

properties and design trends from the MPNN−GA framework correspond to physically meaningful, quantum-mechanically consistent molecular behaviors. Accordingly, ground- and excited-state properties were computed using DFT and time-dependent DFT. Each molecule was represented by its SMILES string, from which an initial 3D structure was generated using the MMFF94 force field implemented in the `gen3d` module of OpenBabel.[54] The Merck Molecular Force Field (MMFF) parameters[55−59] provided a physically reasonable starting geometry while maintaining computational efficiency. These preliminary structures were further refined using the self-consistent charge density-functional tight-binding method (DFTB3),[60−62] augmented with many-body dispersion (MBD) corrections[63−66] to accurately capture long-range intermolecular interactions. All DFTB3-MBD calculations were performed with DFTB+,[67] interfaced through the Atomic Simulation Environment (ASE).[68] This two-step geometry preparation ensured that each candidate structure was chemically sound and energetically consistent before higher-level DFT calculations. Final geometry optimizations were carried out using Kohn−Sham DFT within the Gaussian 09 package,[69] employing the B3LYP functional with the 6−31G(d,p) basis set. Harmonic vibrational frequency analysis confirmed that all optimized geometries correspond to true minima on the potential energy surface (no imaginary frequencies).

From the converged structures, ground-state frontier molecular orbital descriptors were extracted. All calculations were carried out in a solvent environment mimicking experimental conditions—chloroform (ε = 4.7113)—modeled with the Conductor-like Polarizable Continuum Model (CPCM). The first singlet excitation energy (E_S) and corresponding oscillator strength were extracted from the TD-DFT calculations using the same functional and basis set. The exciton binding energy (E_b) was subsequently computed to assess the degree of electron−hole separation, a key indicator of charge-transfer efficiency in NFAs. It was estimated from the difference between the HOMO−LUMO gap and the lowest singlet excitation energy using the relation:

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

This systematic multilevel validation—from empirical to tight-binding to hybrid DFT— ensures that the GA-generated molecules are not only algorithmically viable but also physically consistent, providing confidence in their suitability for organic photovoltaic applications.

3. RESULTS AND DISCUSSION

3.1. Training and Validation Dynamics of the MPNN Predictive Models

The training behavior of the directed message passing neural network models provides insight into how well the architecture learns key optoelectronic descriptors that guide the genetic algorithm. Figure 2 summarizes the learning curves for oscillator strength (f), exciton binding energy (E_b), and the LUMO−LUMO+1 energy difference (ΔE_{LUMO}). The training and validation losses for the oscillator strength model decrease rapidly within approximately 15 epochs and converge to values below 0.25 (Figure 2a), indicating stable learning and reliable generalization across chemically diverse scaffolds. The E_b model exhibits a slower but consistent decline in training and validation loss over nearly 120 epochs (Figure 2b), reflecting the more complex and noisy relationship between molecular structure and exciton binding strength. In contrast, the ΔE_{LUMO} model shows occasional fluctuations in validation loss (Figure 2c), suggesting higher sensitivity to subtle geometric or electronic variations; nevertheless, the overall downward trend indicates meaningful signal capture. The spiking behavior in the validation loss for the LUMO−LUMO+1 gap arises from both randomized mini-batch evaluation and the target's intrinsic sensitivity. Because ΔE_{LUMO} is the difference between two closely spaced frontier orbitals, it is susceptible to near-degeneracy and orbital reordering, leading to a heteroscedastic distribution of labels with occasional large-gap outliers. When validation loss is computed over shuffled mini-batches, including such samples can transiently increase the mean-squared error (MSE), leading to visible epoch-to-epoch fluctuations despite overall convergence.

Together, these results demonstrate that the MPNN learns physically interpretable correlations across distinct molecular properties. The model for f effectively identifies optically bright scaffolds, the E_b predictor embeds energetic stability, and the ΔE_{LUMO} model contributes electronic-level discrimination. When integrated into the GA workflow, these complementary predictors bias molecular evolution toward regions of chemical space where multiple photovoltaic criteria converge—thereby enabling rational, physics-aware exploration for high-performing NFAs.

3.2. Predictive Accuracy of the MPNN Models

The predictive performance of the trained MPNNs was evaluated on an independent test set comprising unseen NFA molecules. Figure S1 shows the parity plots comparing predicted and reference (DFT-computed) values for the three target

https://doi.org/10.1021/acsami.5c26136
ACS Appl. Mater. Interfaces XXXX, XXX, XXX−XXX

properties—f, E_b, and ΔE_{LUMO}. The performance metrics (RMSE, MAE, and Pearson correlation coefficient r) for predicting key optoelectronic properties—LUMO offset (ΔE_{LUMO}), oscillator strength (f), and exciton binding energy (E_b)—using the evidential MPNN model on the training and test sets are provided in Table S2. Data points cluster tightly around the diagonal reference line, indicating strong agreement between model predictions and quantum mechanical benchmarks. The oscillator strength model shows the closest alignment (Figure S1c), capturing subtle variations linked to conjugation length and donor−acceptor interactions. Similarly, the ΔE_{LUMO} and E_b predictors maintain high fidelity (Figures S1a,b), confirming that the MPNN generalizes effectively across electronic and excitonic descriptors when tuned per property. The narrow 95% confidence intervals further demonstrate quantitative reliability suitable for property-driven molecular optimization. These results confirm that the MPNN can be an accurate and computationally efficient surrogate for evaluating photovoltaic descriptors during GA-driven molecular generation. This predictive capability bridges the gap between data-driven learning and physics-based validation, enabling rapid yet reliable exploration of the vast chemical landscape of NFAs.

Having established the predictive reliability of the MPNN surrogates, we next integrated them into the physics-informed GA to explore the NFA chemical space. The subsequent subsections discuss the evolution of molecular populations, convergence behavior, and structural trends in the high-fitness candidates.

3.3. Convergence and Stability of GA−MPNN Optimization

The convergence profiles (Figure 3) reveal a smooth and physically consistent optimization process under the GA−MPNN framework. The rolling best fitness increases steadily from about 0.64 to 0.81, while the mean population fitness rises from 0.52 to 0.75 over roughly 50 generations (g), demonstrating effective exploration and stable convergence. The near-sigmoidal evolution indicates a well-balanced multi-objective search, simultaneously maximizing oscillator strength, while minimizing exciton binding energy and the LUMO−LUMO+1 energy gap. This progression arises from three key design features of the genetic search: (i) a balanced scalarization scheme that ensures comparable weighting of optical and electronic targets, (ii) elitism that preserves the most promising candidates across generations, and (iii) a gradually decaying mutation rate that guides the transition from broad exploration to fine-grained optimization. Together, these mechanisms prevent premature convergence and maintain population diversity.

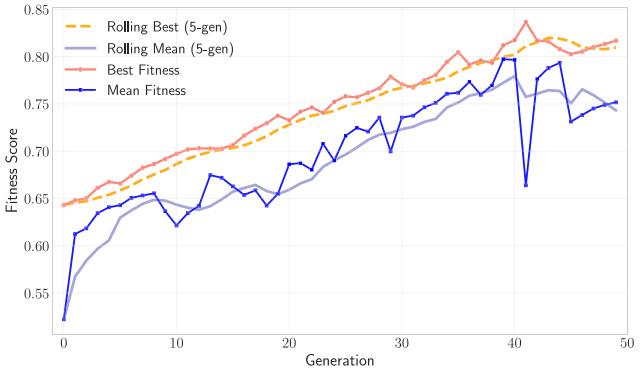


Figure 3. Evolution of fitness over 50 generations during GA−MPNN optimization. The plot shows the instantaneous best (red circles) and mean population fitness (blue squares), along with their five-generation rolling averages (orange dashed and gray solid lines). The steady rise and eventual plateau indicate smooth convergence, maintained population diversity, and balanced trade-offs among oscillator strength, exciton binding energy, and LUMO−LUMO+1 gap objectives.

As the optimization proceeds, the spread between best and mean fitness initially widens, reflecting diversification of elite solutions, and later narrows as the population concentrates around high-performing regions of chemical space. The consistency of the rolling stability metric (see Figure S2 and Section S2) confirms that improvements correspond to genuine property gains rather than stochastic oscillations. Since all three objectives are evaluated using the same MPNN surrogate with uniform molecular featurization, the composite fitness reflects correlated progress in oscillator strength, exciton energetics, and orbital alignment. These results demonstrate that the GA−MPNN framework conducts a physics-aware, sample-efficient, and stable exploration of the nonfullerene acceptor landscape.

3.4. Evolution of Population Diversity and Structural Stability

The population diversity trajectories (Figure 4) provide insight into how the GA−MPNN search balances exploration of chemical space with convergence toward optimal NFA architectures. At initialization, the population exhibits relatively high diversity ($D_g \approx 0.55-0.60$), reflecting wide structural and property heterogeneity. Over successive generations, diversity decreases gradually to around 0.40 by ($g \approx 40$), indicating a controlled transition from exploratory sampling to focused optimization. This decline is consistent with elitism, adaptive novelty reweighting, and an annealed mutation probability [$p_\mu(g)$] that narrows the genotypic search radius while maintaining stochastic variability.

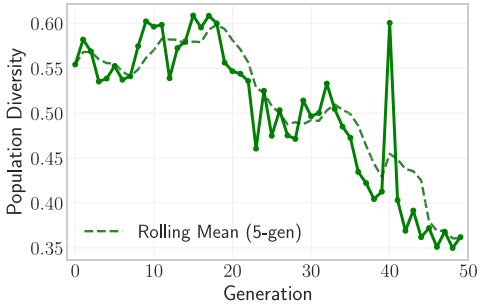


Figure 4. Evolution of population diversity over 50 generations in the GA−MPNN optimization. The solid green line shows instantaneous diversity, and the dashed line represents the 5-generation rolling mean. Diversity decreases steadily as the population converges toward high-fitness NFAs, with a controlled resurgence near generation 40 preventing overconvergence and sustaining structural variety.

A brief, deliberate resurgence in diversity occurs near generation 40, where an adaptive guardrail detects excessive convergence and temporarily reintroduces variation ($\Delta D_g \approx +0.10$). This corrective pulse is followed by stabilization [$D_g \approx 0.35-0.38$], marking entry into a steady-state regime of exploitation. The lack of extended negative runs in the diversity-change rate (ΔD_g; see Figure S3 and Section S3) and the persistence of a diversity floor confirm that the algorithm avoids mode collapse. Parent selection rules—based on cross-metric dissimilarity in fingerprints, scaffolds, molecular weights,

https://doi.org/10.1021/acsami.5c26136
ACS Appl. Mater. Interfaces XXXX, XXX, XXX−XXX

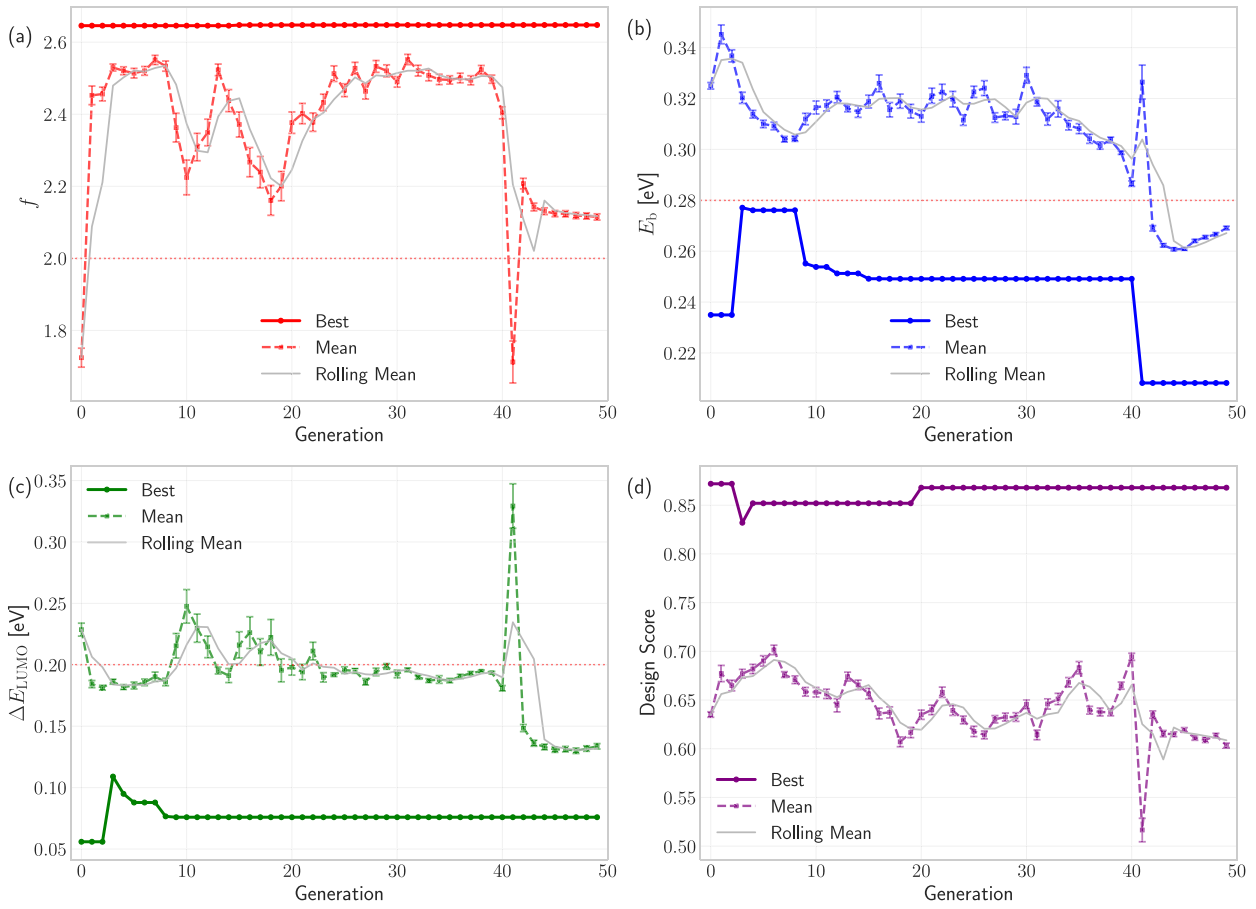


Figure 5. Convergence profiles of key molecular properties during GA−MPNN optimization over 50 generations. (a) f, (b) E_b, (c) ΔE_{LUMO}, and (d) overall design score. Solid lines represent generation-best candidates, dashed lines denote population means, and gray traces show five-generation rolling averages, while dashed horizontal lines indicate the target threshold values used in the fitness function for property selection (minimum f, maximum allowable E_b, and maximum allowable ΔE_{LUMO}). All properties exhibit smooth, physically consistent convergence toward target thresholds, with the composite design score stabilizing around 0.85.

and (log P)—preserve structural heterogeneity while filtering out redundant or chemically implausible solutions.

Overall, the diversity evolution reveals a structured progression through three phases: an early stage of broad exploration sampling varied donor−acceptor topologies, a midstage corrective burst that restores variance, and a late-stage convergence to a chemically meaningful basin. This controlled entropy management ensures that the final population remains compositionally diverse yet fitness-oriented, providing a robust foundation for property-level convergence. As diversity stabilizes, the optimization naturally shifts focus from structural exploration to refining key optoelectronic targets—leading to the emergence of consistently high-performing NFAs with balanced photophysical characteristics, as detailed below.

3.5. Convergence Trajectories of Targeted Molecular Properties

The evolution of property-specific fitness components (Figure 5) illustrates the coordinated, multiobjective convergence achieved through the GA−MPNN optimization. Each panel corresponds to one of the key descriptors—f, E_b, and ΔE_{LUMO}—along with the overall NFA design score that balances them within a physically informed composite objective. For oscillator strength, top-performing candidates exceed the threshold of 2.0 within the initial few generations and stabilize around ∼2.6 (Figure 5a). The population mean remains within the 2.2−2.5 range, signifying strong but controlled selection toward optically bright chromophores. This early saturation reflects the model's capacity to recognize conjugation and donor−acceptor motifs that enhance radiative transitions. E_b shows a more gradual decline—from ∼0.32 eV at initialization to ∼0.28 eV by generation 40—with the mean converging below to the practical threshold of 0.28 eV at later generations (Figure 5b). This progressive reduction indicates that low-E_b architectures emerge naturally under concurrent pressure to maintain strong oscillator strength, highlighting the cooperative rather than conflicting nature of these objectives. For the frontier orbital descriptor (ΔE_{LUMO}), rapid convergence is observed: the best molecules achieve gaps below 0.2 eV within the first ten generations and continue to narrow to <0.1 eV at later stages (Figure 5c). The mean value also approaches this threshold, confirming early discovery and sustained retention of closely spaced LUMO manifolds—an essential feature for charge delocalization in high-performance NFAs. The earlier and stronger convergence of the optical objective f and the frontier-orbital objective ΔE_{LUMO} compared to the exciton binding energy can be rationalized based on their underlying physical nature. Both f and ΔE_{LUMO} are predominantly intramolecular electronic properties that respond directly to common structural motifs in nonfullerene acceptors, such as extended and planar π-conjugation, rigid fused cores, and substantial donor−acceptor

https://doi.org/10.1021/acsami.5c26136
ACS Appl. Mater. Interfaces XXXX, XXX, XXX−XXX

substitution patterns. These features are relatively abundant in the accessible chemical space and tend to monotonically enhance transition dipole moments while compressing low-lying virtual orbitals, making high-f and small-ΔE_{LUMO} solutions comparatively easy to discover and rapidly enriched during early generations. In contrast, E_b is a many-body Coulombic quantity that depends on electron−hole separation and dielectric screening, approximately scaling as $\left(E_b \approx E_g^{QP} - E_{opt} \sim \frac{e^2}{4\pi\varepsilon_0\varepsilon_r r_{eh}}\right)$.[70] In organic semiconductors, weak dielectric screening and localized excitations intrinsically limit how low E_b can be. Achieving very small E_b therefore requires rarer combinations of strong yet balanced intramolecular charge-transfer character, enhanced delocalization, and increased polarizability—conditions that may introduce trade-offs with oscillator strength or orbital alignment. Consequently, low-E_b candidates occupy a narrower region of chemical space, leading to delayed emergence and lower population density relative to the optical and frontier-orbital objectives. Finally, the overall NFA design score, which integrates physical priors and synthetic plausibility, rises steadily and plateaus above 0.85 for the best individuals, with the mean maintained in the 0.60−0.70 range (Figure 5d). This behavior confirms that the prior influences the search trajectory without dominating it, ensuring that high fitness arises from genuine structure−property improvements rather than numerical bias.

Together, these property trajectories demonstrate that the GA−MPNN pipeline achieves self-consistent optimization across multiple, interdependent targets. Oscillator strength acts as the primary driver of optical activity. At the same time, E_b and ΔE_{LUMO} serve as energetic stabilizers, collectively yielding a balanced population of bright, electronically favorable, and physically realistic NFAs ready for quantum chemical validation. This interplay between diversity and fitness—quantitatively examined in Section S4 and Figure S4—underscores that the GA−MPNN framework maintains a controlled balance between exploration and exploitation, enabling convergence without sacrificing structural variety.

3.6. Population Growth and Candidate Enrichment

The evolution of population trajectories (Figure 6) illustrates how the diversity-regulated GA, coupled with the evidential MPNN predictor, progressively channels sampling toward the chemically and electronically optimal design space. The total molecular population (light blue) is tracked alongside the subset of successful molecules—those simultaneously satisfying $f > 2.0$, $E_b < 0.28$ eV, and $\Delta E_{LUMO} < 0.2$ eV. For the first ∼38 generations, the success set remains limited, a consequence of stringent novelty constraints (SMILES/InChI uniqueness, scaffold nonredundancy, and fingerprint diversity thresholds) and logistic fitness shaping that restricts early overexploitation of any single structural motif. Beyond $g \approx 40$, a pronounced inflection emerges: the count of successful candidates rises steeply, exceeding ∼1,300 unique structures by $g = 50$. This surge coincides with the late-stage annealing of mutation probability [μ: 0.6 → 0.15] and the activation of adaptive novelty gates, which together enable recombination of viable scaffolds, controlled heteroatom substitutions, and side-chain diversification while maintaining property coherence. The resulting nonlinear enrichment demonstrates a synergistic transition from exploratory to exploitative search, where population entropy is judiciously converted into chemically valid, property-compliant solutions.

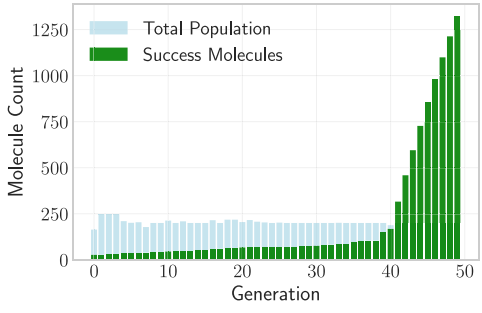


Figure 6. Population dynamics during GA−MPNN optimization showing the evolution of total molecular population (light blue) and cumulative successful molecules (green) over 50 generations. The success population remains sparse through the first 38 generations due to novelty constraints and strict selection thresholds, followed by rapid enrichment beyond $g \approx 40$, reaching over 1,300 high-quality candidates by generation 50. The late-stage surge reflects efficient property-driven filtering and scaffold recombination within a controlled evolutionary regime.

Overall, these dynamics highlight the efficiency of the GA−MPNN coupling in filtering vast combinatorial space into a compact yet diverse pool of optoelectronically favorable NFAs. The monotonic enrichment of successful molecules—without collapse of structural diversity—confirms that the evolutionary process remains both physically guided and statistically balanced, achieving accelerated molecular discovery within a constrained generational budget.

3.7. Property Correlations and Fitness Drivers

Building on the enrichment trends discussed above, we next examine how individual molecular properties interact to shape the overall fitness landscape. The correlation analysis (Figure 7) provides a quantitative view of how f, E_b, and ΔE_{LUMO} collectively influence the MPNN−guided GA optimization.

A strong positive correlation between fitness and oscillator strength ($r = 0.67$) confirms that optical brightness remains the dominant driver of selection, consistent with its leading weight in the scalarized objective (Figure 7a). In contrast, moderate negative correlations with both E_b ($r = -0.35$) and ΔE_{LUMO} ($r = -0.38$) indicate that reduced exciton binding and smaller frontier orbital gaps are systematically favored, playing supporting and stabilizing roles rather than dictating the optimization pathway. The intentionally down-weighted NFA prior ($w = 0.10$) exhibits only a weak relationship with fitness ($r = 0.07$), acting primarily as a regularizer to constrain chemically unrealistic motifs. Low interdescriptor correlations ($|r| < 0.2$) further emphasize the orthogonality of the property set and justify the multiobjective formulation. The evolutionary trajectory of the population reinforces this relationship structure (Figure 7b). As generations progress, high-f candidates increasingly occupy the upper-fitness region, forming a dense band above $F \approx 0.7$ by $g > 35$. Simultaneously, the fraction of molecules satisfying all target thresholds ($f > 2.0$, $E_b < 0.28$ eV, and $\Delta E_{LUMO} < 0.2$ eV) rises steadily, marking the emergence of a consistent, high-quality subpopulation. These coordinated gains illustrate the algorithm's capacity to reinforce axis-specific improvement—oscillator strength elevates the achievable fitness ceiling, while E_b and ΔE_{LUMO} act as energetic filters that ensure optical and electronic coherence. To further elucidate how these objectives shape the evolving molecular population, we analyze the statistical behavior of key properties over successive

https://doi.org/10.1021/acsami.5c26136
ACS Appl. Mater. Interfaces XXXX, XXX, XXX−XXX

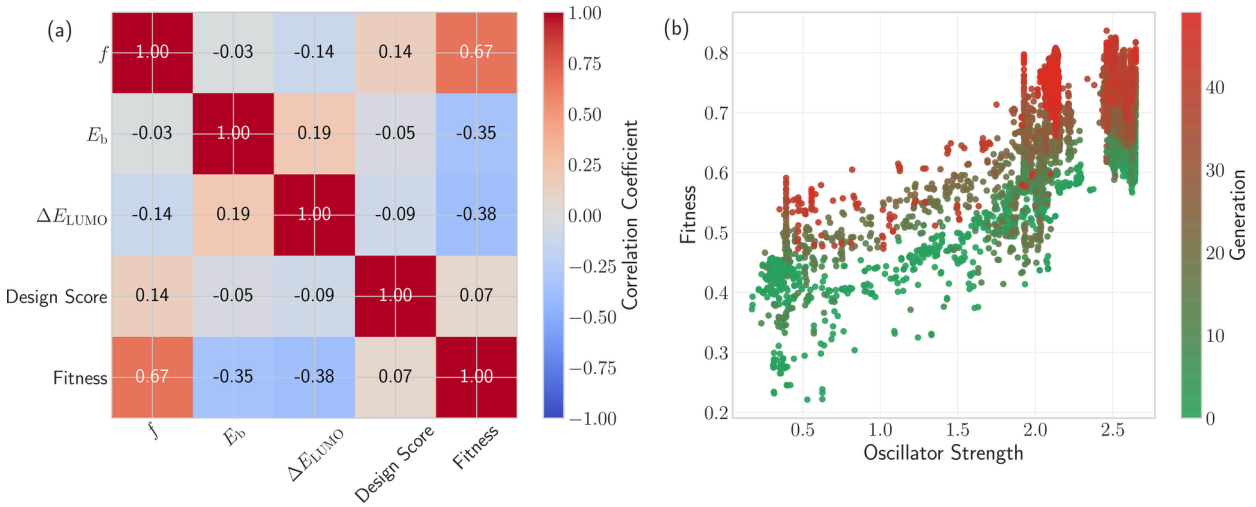


Figure 7. (a) Pairwise correlation matrix showing relationships among oscillator strength (f), exciton binding energy (E_b), LUMO−LUMO+1 gap (ΔE_{LUMO}), NFA prior score, and overall fitness (F). Fitness correlates strongly with f ($r = 0.67$) and moderately anticorrelates with E_b and ΔE_{LUMO}, consistent with the weighted multiobjective design. Low interdescriptor correlations highlight the complementarity of the property set. (b) Fitness versus oscillator strength across generations, with color encoding generation index. Later generations progressively cluster in the joint high-f, high-F region, indicating an evolutionary focus on optically bright and electronically balanced candidates that meet all design thresholds.

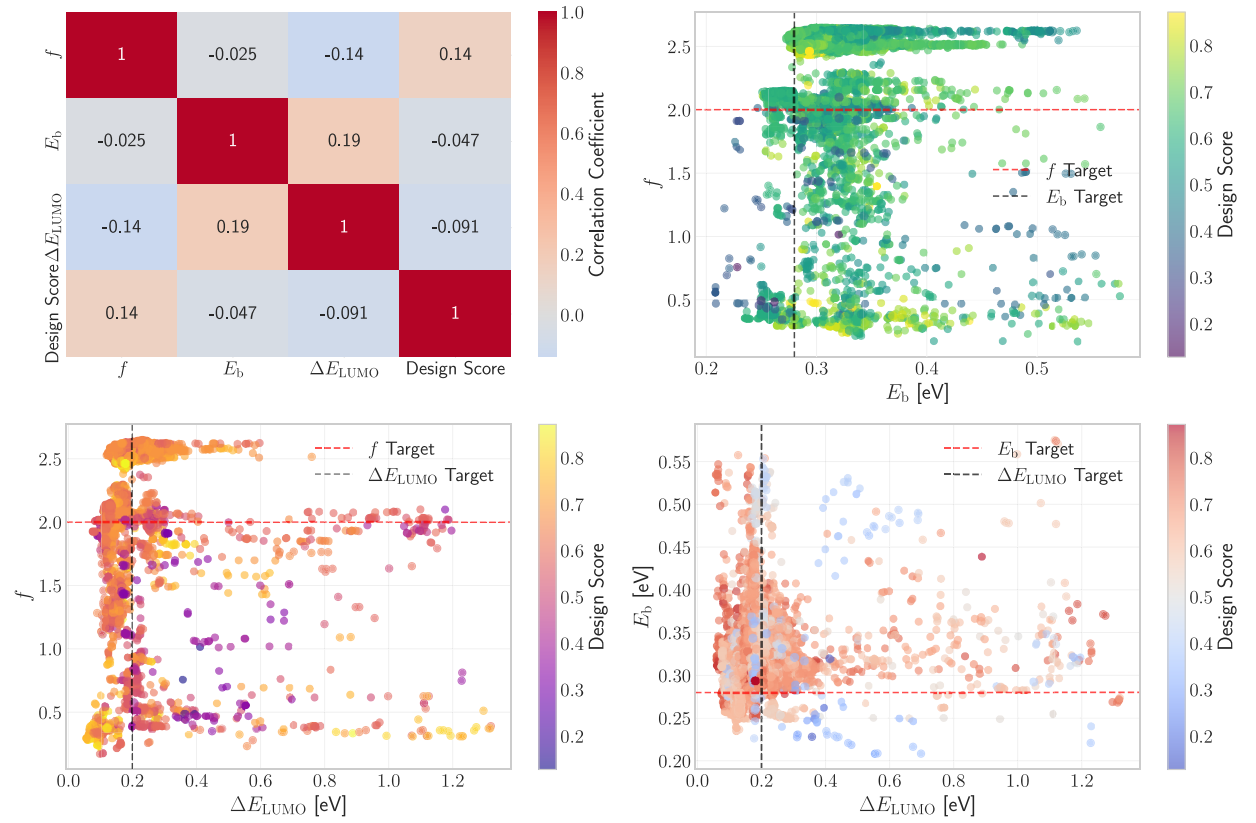


Figure 8. Correlation matrix (top left) and pairwise scatter plots showing relationships among oscillator strength (f), exciton binding energy (E_b), frontier-orbital gap (ΔE_{LUMO}), and NFA design score for the GA-evolved population. Weak interproperty correlations ($|r| < 0.15$) confirm the orthogonality ofdescriptors, supporting multiobjective optimization. Scatter plots highlight constraint thresholds ($f > 2.0$, $E_b < 0.28$ eV, $\Delta E_{LUMO} < 0.2$ eV), with high design scores concentrated along the Pareto-consistent funnel—demonstrating successful evolutionary enrichment of optoelectronically balanced, high-fitness molecular scaffolds.

generations. A comprehensive description of the generational property evolution, including detailed convergence trajectories and statistical analyses, is provided in Section S5 of the Supporting Information (SI).

3.8. Interproperty Correlations and Emergent Pareto Structure

Building on the property-specific success trends, we next examine the global correlation structure among f, E_b, ΔE_{LUMO}, and the composite NFA design score (Figure 8). Across the full

https://doi.org/10.1021/acsami.5c26136
ACS Appl. Mater. Interfaces XXXX, XXX, XXX−XXX

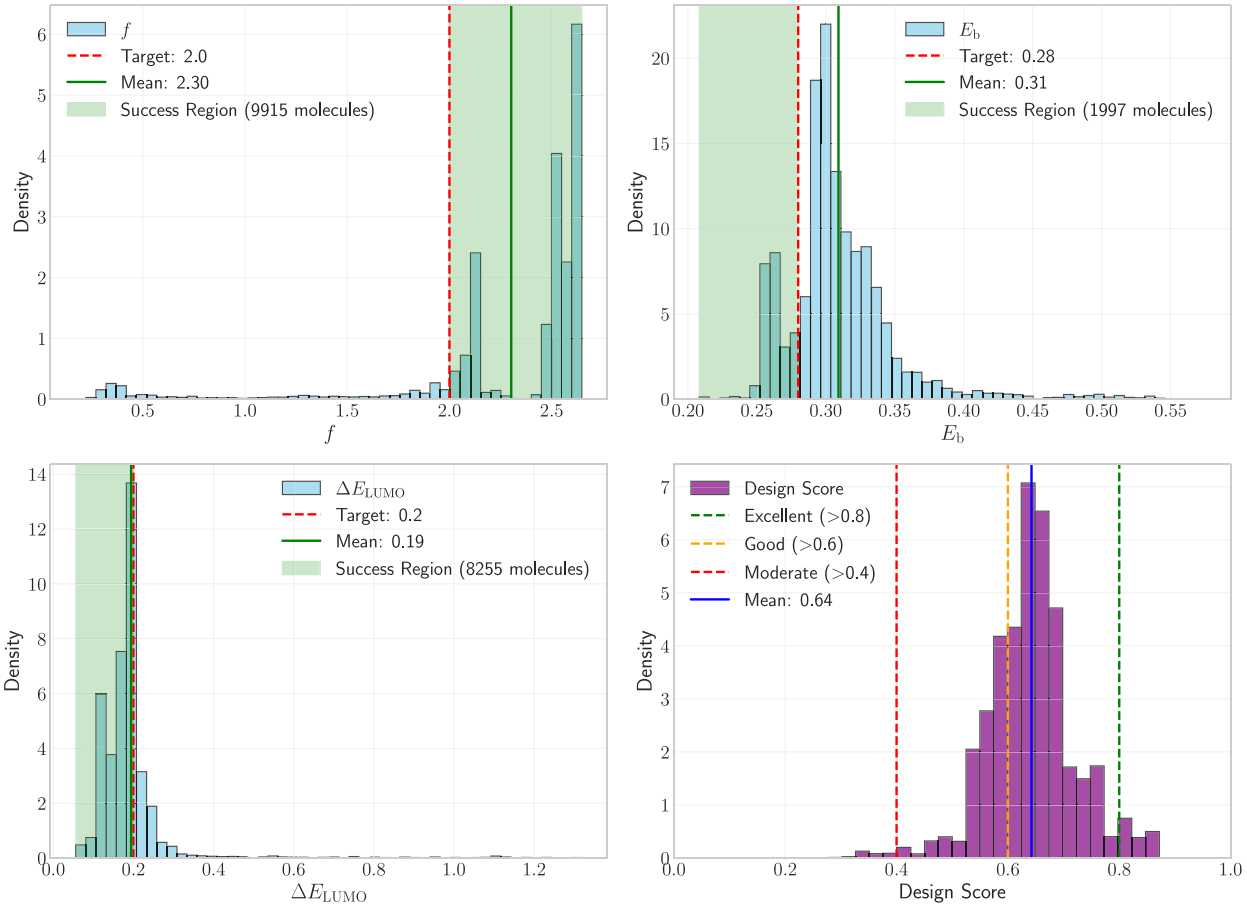


Figure 9. Aggregate distributions of molecular properties across all GA generations for (top left) oscillator strength (f, target >2.0; 9,915 successes), (top right) exciton binding energy (E_b, target <0.28 eV; 1,997 successes), (bottom left) LUMO−LUMO+1 gap (ΔE_{LUMO}, target <0.2 eV; 8,255 successes), and (bottom right) NFA design score. Green shaded regions denote threshold-satisfying populations. The oscillator strength and ΔE_{LUMO} objectives exhibit high success rates (∼99% and ∼83%), whereas the exciton-binding criterion remains the most restrictive (∼20%). The design score distribution peaks near 0.64 with pronounced "good" (>0.6) and "excellent" (>0.8) subsets, confirming that multiobjective optimization preserves chemical realism while advancing optical and energetic performance.

GA-evolved population, pairwise correlations remain uniformly weak ($|r| < 0.15$), indicating that these descriptors encode largely orthogonal information and thereby justify a multiobjective treatment. The weak but systematic correlations observed among f, E_b, and ΔE_{LUMO} can be understood in terms of a common underlying axis of electronic delocalization and intramolecular charge-transfer (ICT) character in conjugated nonfullerene acceptors. Increased conjugation length, strong donor−acceptor interactions, and enhanced electron delocalization promote ICT character in the lowest excited state. Greater delocalization and polarizability reduce the Coulomb attraction between electron and hole, thereby lowering the exciton binding energy E_b.[71,72] Simultaneously, extended conjugation and acceptor-strength effects compress the spacing of low-lying virtual orbitals, leading to a smaller ΔE_{LUMO}. However, while moderate LE−CT hybridization can maintain strong transition dipoles, excessively CT-dominated excitations may reduce oscillator strength f due to weaker optical coupling.[73−75] As a result, improvements in E_b and ΔE_{LUMO} do not translate linearly into higher f, yielding only modest interproperty correlations.[76,77] These trends indicate that the GA navigates a multidimensional design landscape governed by partially coupled, nonredundant electronic structure effects rather than a single dominant structural parameter. The broad distribution of points across all pairwise planes reflects an extensive exploration of the accessible chemical space, with successive generations progressively concentrating within the constraint-satisfying region defined by the target thresholds ($f > 2.0$, $E_b < 0.28$ eV, $\Delta E_{LUMO} < 0.2$ eV). Within the f−E_b landscape, candidates cluster along a plateau near $E_b \sim 0.3$−0.4 eV, while very few candidates exceed the optical activity target ($f > 2.0$) and simultaneously satisfy the $E_b < 0.28$ eV requirement, revealing a selective trade-off between oscillator strength and exciton confinement. In f−ΔE_{LUMO} space, strong oscillators span a wide energetic range, yet the joint attainment of $f > 2.0$ and $\Delta E_{LUMO} < 0.28$ eV remains sparse—highlighting the intrinsic difficulty of co-optimizing optical brightness and charge-transfer character. The E_b−ΔE_{LUMO} projection exhibits a narrow, funnel-like region approaching the Pareto front, where the color-coded NFA design score increases monotonically. This monotonic rise substantiates that the GA−MPNN framework selectively amplifies structurally realistic scaffolds that reconcile optical, electronic, and thermodynamic requirements. Collectively, these observations reveal that while single-property trends appear weak, the underlying landscape is inherently multiobjective and nonlinear. The emergent Pareto structure provides a mechanistic pathway through which the evolutionary search achieves balanced, physically consistent optimization of next-generation NFA candidates.

https://doi.org/10.1021/acsami.5c26136
ACS Appl. Mater. Interfaces XXXX, XXX, XXX−XXX

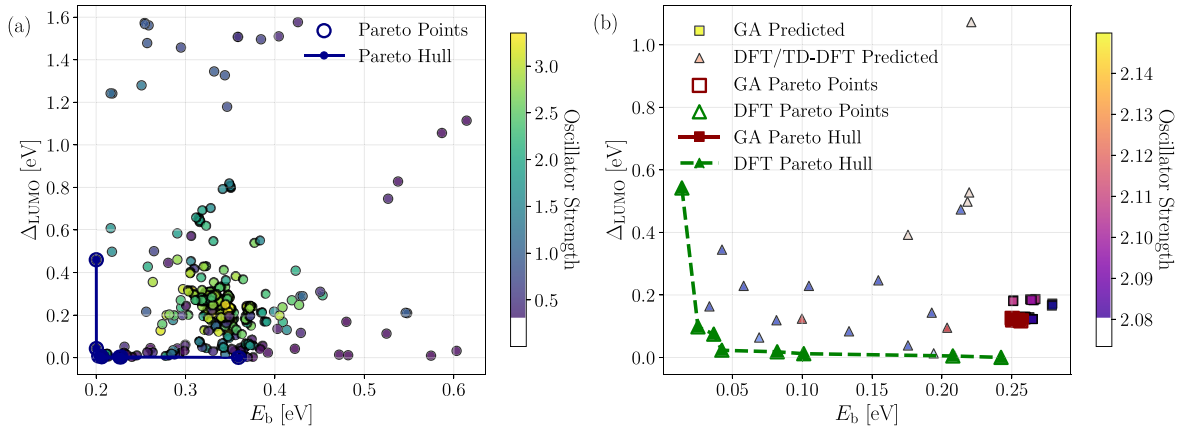


Figure 10. Pareto front comparison between (a) the training data set (left; blue hull) and (b) GA−MPNN-generated candidates validated via DFT/TD-DFT (right). Each point represents a molecule characterized by exciton binding energy (E_b), LUMO−LUMO+1 energy gap (ΔE_{LUMO}), and oscillator strength (f). Green regions mark the DFT-validated Pareto frontier, while red-outlined squares highlight GA Pareto-optimal molecules. The GA population extends the DFT Pareto hull toward lower E_b and ΔE_{LUMO} values while maintaining high f, demonstrating that the evolutionary search reproduces realistic optoelectronic trade-offs and pushes the design boundary into promising, previously unexplored chemical regions.

3.9. Aggregate Property Distributions and Multiobjective Convergence

The Pareto-resolved correlations reveal that physically meaningful convergence in the GA−MPNN search stems from coordinated optimization across orthogonal property axes, rather than dominance by any single descriptor. To consolidate these insights, we next examine the global property distributions aggregated over all generations, providing a population-level validation of how multiobjective constraints collectively shape the emergent chemical space. Figure 9 summarizes these aggregate distributions. The oscillator strength (top left) exhibits a pronounced right skew with a mean of 2.30 and nearly 9900 molecules above the 2.0 threshold, confirming strong evolutionary enrichment toward optically bright scaffolds. The exciton binding energy (top right) centers at 0.31 eV, marginally above the 0.28 eV target, yet includes nearly 2000 successful molecules, reflecting the scarcity of low-E_b chemistries within the sampled design space. The ΔE_{LUMO} distribution (bottom left) displays a narrow peak near 0.19 eV, with 8200+ molecules satisfying the energetic constraint, underscoring intense selection pressure on frontier orbital separation. Meanwhile, the NFA design score (bottom right) shifts steadily toward higher values (mean 0.64), with substantial fractions in the "good" (>0.6) and "excellent" (>0.8) performance tiers—signifying retention of structural plausibility even under aggressive optoelectronic optimization.

Together, these global distributions confirm that the GA−MPNN framework efficiently navigates the multiobjective landscape: while optical and frontier-orbital objectives converge early and strongly, low-exciton-binding solutions emerge later and remain relatively sparse, thereby defining the ultimate constraint in high-performance NFA discovery.

3.10. Pareto Validation of GA-Designed Candidates against DFT Benchmarks

The aggregate property distributions underscore that the GA−MPNN framework systematically concentrates population density into regions of chemical space consistent with high oscillator strength, low exciton binding, and narrow frontier orbital gaps. To confirm that these trends extend beyond surrogate model predictions, we next benchmark representative GA-evolved candidates against DFT and TD-DFT calculations, examining whether their trade-off structures remain physically consistent at the quantum chemical level.

Out of the 56 nonfullerene acceptor molecules generated by the genetic algorithm (Data set - II), we validated the predicted properties of 28 using DFT and TD-DFT calculations. Further details regarding the validation are provided in Section S6. The top-ranked NFAs identified by the GA exhibit substantial structural similarity to previously reported spirobifluorene−diketopyrrolopyrrole (SF−DPP) architectures.[78] While early devices based on such acceptors demonstrated only moderate PCEs (~2−3%), it is important to distinguish between device-level efficiency and intrinsic molecular optoelectronic quality. In this study, "high-performance" refers to intrinsic electronic characteristics targeted by the GA—namely, high oscillator strength, reduced exciton binding energy, favorable frontier-orbital alignment associated with high V_{OC}, and structurally encoded aggregation control—rather than record device efficiencies. Spiro-centered three-dimensional topologies are well-known to provide high-lying LUMO levels and tunable intramolecular charge-transfer absorption while mitigating excessive planar aggregation, thereby offering robust morphology control and voltage retention.[79−81] Device PCE, however, is an emergent property that depends sensitively on donor selection, nanoscale phase separation, crystallinity, charge-transport balance, and processing conditions (e.g., solvent additives and post-treatment).[82−86] Therefore, the rediscovery of SF−DPP-like motifs should be interpreted as identification of an intrinsically favorable electronic design class within the defined search space, rather than as a claim of state-of-the-art device performance. Figure 10 presents a comparative Pareto analysis of exciton binding energy, LUMO−LUMO+1 energy gap, and oscillator strength between the training distribution (left) and the GA−MPNN-generated molecules validated through DFT/TD-DFT (right). In the training set, the Pareto hull (blue) encloses compositions that balance moderate exciton binding (E_b ~ 0.3−0.4 eV), frontier gap suppression (ΔE_{LUMO} < 0.5 eV), and enhanced oscillator strength. High-f cases define the lower boundary of this hull, illustrating the intrinsic trade-off between optical activity and energetic stabilization. In contrast, GA-generated candidates (squares) cluster toward lower E_b (<0.25 eV) and narrower ΔE_{LUMO} (<0.2 eV), extending the DFT Pareto frontier (green) into previously unoccupied regions

https://doi.org/10.1021/acsami.5c26136
ACS Appl. Mater. Interfaces XXXX, XXX, XXX−XXX

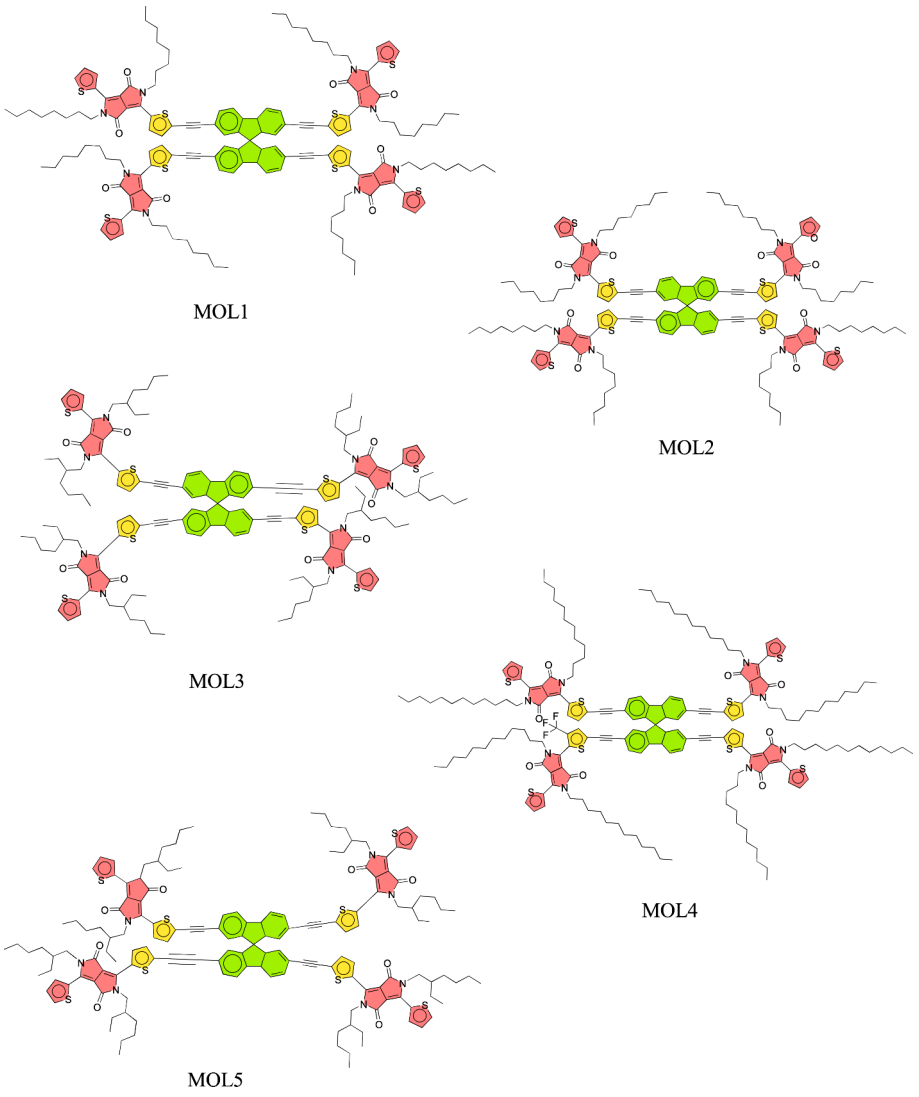


Figure 11. Representative molecular structures of the top-performing nonfullerene acceptors identified by the GA−MPNN optimization framework. These candidates exhibit optimal combinations of high oscillator strength, low exciton binding energy, and favorable frontier orbital alignment, reflecting the framework's ability to discover chemically realistic, high-performance scaffolds.

of design space. Pareto-optimal GA points (red-outlined squares) overlap with the DFT hull—demonstrating predictive fidelity—while also populating new low-E_b, low-ΔE_{LUMO} regions, signifying innovation beyond the training data. The constrained oscillator strength range ($f \approx 2.08-2.14$) confirms that evolutionary pressure preserves photophysical brightness while enabling structural diversification that supports improved charge-transfer balance. Overall, the strong alignment between GA-predicted and DFT-validated property landscapes confirms that the combined MPNN−GA framework captures the essential physics of NFA design, maintaining the fundamental trade-offs among optical excitation, electronic separation, and exciton confinement while efficiently navigating toward optoelectronically optimal scaffolds that define the next-generation Pareto frontier. The molecular structures of the top five best-performing NFAs are shown in Figure 11. To evaluate synthetic plausibility, we computed the Synthetic Accessibility Score (SAscore)[87] and QED[88] for all 28 DFT-verified GA-generated NFAs and benchmarked them against 1,299 experimentally reported NFAs. Although the GA candidates exhibit moderately higher average SAscores—reflecting increased structural complexity associated with extended π-conjugation—their values substantially overlap with the experimental distribution, confirming that the generated molecules remain within a realistic and synthetically precedented chemical space. Additional details regarding the synthetic accessibility analysis are provided in Section S7.

4. CONCLUSION

This work presents a physics-grounded generative framework that couples an evidential message-passing neural network with a constraint-encoded genetic algorithm to enable inverse design of nonfullerene acceptors for organic solar cells. By incorporating key quantum-validated descriptors—oscillator strength, exciton binding energy, and the LUMO−LUMO+1 gap—directly into the evolutionary fitness function, the workflow moves beyond empirical PCE surrogates and directs the search toward molecules that embody established structure−property principles of high-performing NFAs. The integration of scaffold-aware diversity control and chemically informed mutation allows the GA to efficiently explore a large and structurally heterogeneous space while maintaining synthetic plausibility.

https://doi.org/10.1021/acsami.5c26136
ACS Appl. Mater. Interfaces XXXX, XXX, XXX−XXX

Across generations, the algorithm identifies numerous candidates that simultaneously satisfy the targeted multi-objective criteria, and MPNN predictions show strong agreement with DFT/TD-DFT benchmarks, reinforcing the reliability of the surrogate model. Pareto analyses further reveal that the GA not only recovers known quantum-chemical trade-offs but also extends the frontier to previously underexplored regions of chemical space, yielding molecules with concurrently low E_b, suppressed ΔE_{LUMO}, and high f.

Overall, this study demonstrates a robust and interpretable route for the rational design of next-generation NFAs, effectively linking machine-learning-driven exploration with first-principles validation. The framework is general in scope and can be readily adapted to other classes of functional organic materials, providing a versatile foundation for data-guided molecular discovery.

■ ASSOCIATED CONTENT

Data Availability Statement

The data and code to reproduce the results are available in our GitHub repository (https://github.com/Bib569/DMPNN-GA-NFA-Design).

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acsami.5c26136.

Detailed description of the MPNN-GA framework, including convergence and stability metrics, population diversity and fitness, and classification of GA-generated molecules (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

Bibhas Das − *Department of Chemistry, Indian Institute of Technology Gandhinagar, Gandhinagar, Gujarat 382355, India;* orcid.org/0000-0002-9671-5275; Email: dasbibhas@iitgn.ac.in

Anirban Mondal − *Department of Chemistry, Indian Institute of Technology Gandhinagar, Gandhinagar, Gujarat 382355, India;* orcid.org/0000-0003-3029-8840; Email: amondal@iitgn.ac.in

Complete contact information is available at: https://pubs.acs.org/10.1021/acsami.5c26136

Author Contributions

AM and BD conceived the problem. BD conducted all the simulations. BD and AM analyzed the results and prepared the draft.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

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

(2) Lu, L.; Zheng, T.; Wu, Q.; Schneider, A. M.; Zhao, D.; Yu, L. Recent advances in bulk heterojunction polymer solar cells. *Chem. Rev.* **2015**, *115*, 12666−12731.

(3) Abhijith, T.; Suthar, R.; Karak, S. 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.

(4) Keshtov, M. L.; Kuklin, S. A.; Konstantinov, I. O.; Khokhlov, A. R.; Nikolaev, A. Y.; Dou, C.; Zou, Y.; Suhtar, R.; Sharma, G. D. 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. *Solar RRL* **2020**, *4*, 1900475.

(5) Zhao, X.; An, Q.; Zhang, H.; Yang, C.; Mahmood, A.; Jiang, M.; Jee, M. H.; Fu, B.; Tian, S.; Woo, H. Y.; Wang, Y.; Wang, J.-L. 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.

(6) Bai, H.-R.; An, Q.; Zhi, H.-F.; Jiang, M.; Mahmood, A.; Yan, L.; Liu, M.-Q.; Liu, Y.-Q.; Wang, Y.; Wang, J.-L. 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.

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

(8) Bai, H.-R.; An, Q.; Jiang, M.; Ryu, H. S.; Yang, J.; Zhou, X.-J.; Zhi, H.-F.; Yang, C.; Li, X.; Woo, H. Y.; Wang, J.-L. 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.

(9) Suthar, R.; T, A.; Dahiya, H.; Singh, A. K.; Sharma, G. D; Karak, S. 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.

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

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

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

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

(14) Greenstein, B. L.; Hutchison, G. R. Organic photovoltaic efficiency predictor: data-driven models for non-fullerene acceptor organic solar cells. *J. Phys. Chem. Lett.* **2022**, *13*, 4235−4243.

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

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

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

(18) Oh, S.; Kim, Y.; Ahn, T.; Lee, S. K. 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.

(19) Xu, J.; Lin, F.; Zhu, L.; Zhang, M.; Hao, T.; Zhou, G.; Gao, K.; Zou, Y.; Wei, G.; Yi, Y.; Jen, A. K.-Y.; Zhang, Y.; Liu, F. The crystalline

behavior and device function of nonfullerene acceptors in organic solar cells. *Adv. Energy Mater.* **2022**, *12*, 2201338.
(20) Huang, J.; Chen, T.; Mei, L.; Wang, M.; Zhu, Y.; Cui, J.; Ouyang, Y.; Pan, Y.; Bi, Z.; Ma, W.; Ma, Z.; Zhu, H.; Zhang, C.; Chen, X.-K.; Chen, H.; Zuo, L. On the role of asymmetric molecular geometry in high-performance organic solar cells. *Nat. Commun.* **2024**, *15*, 3287.
(21) Goldey, M. B.; Reid, D.; de Pablo, J.; Galli, G. Planarity and multiple components promote organic photovoltaic efficiency by improving electronic transport. *Phys. Chem. Chem. Phys.* **2016**, *18*, 31388−31399.
(22) Das, B.; Mondal, A. Predictive Modeling and Design of Organic Solar Cells: A Data-Driven Approach for Material Innovation. *ACS Appl. Energy Mater.* **2024**, *7*, 9349−9363.
(23) Khatua, R.; Das, B.; Mondal, A. Physics-informed machine learning with data-driven equations for predicting organic solar cell performance. *ACS Appl. Mater. Interfaces* **2024**, *16*, 57467−57480.
(24) Greenstein, B. L.; Hiener, D. C.; Hutchison, G. R. Computational evolution of high-performing unfused non-fullerene acceptors for organic solar cells. *J. Chem. Phys.* **2022**, *156*, 174107.
(25) Greenstein, B. L.; Hutchison, G. R. Screening efficient tandem organic solar cells with machine learning and genetic algorithms. *J. Phys. Chem. C* **2023**, *127*, 6179−6191.
(26) Cao, R.; Zhang, C.-R.; Liu, X.-M.; Gong, J.-J.; Zhang, M.-L.; Liu, Z.-J.; Wu, Y.-Z.; Chen, H.-S. Molecular design of organic photovoltaic donors and non-fullerene acceptors: a combined machine learning and genetic algorithm approach. *J. Mater. Chem. C* **2025**, *13*, 12150−12168.
(27) Peng, S.-P.; Yang, X.-Y.; Zhao, Y. Molecular Conditional Generation and Property Analysis of Non-Fullerene Acceptors with Deep Learning. *Int. J. Mol. Sci.* **2021**, *22*, 9099.
(28) Chen, L.-Q.; Zhang, C.-R.; Sang, C.-C.; Liu, X.-M.; Gong, J.-J.; Zhang, M.-L.; Chen, H.-S. High-Throughput Molecular Design of Donors and Non-Fullerene Acceptors for Organic Solar Cells Based on Convolutional Neural Networks. *J. Chem. Inf. Model.* **2025**, *65*, 10107−10123.
(29) Zhang, C.-R.; Cao, R.; Liu, X.-M.; Zhang, M.-L.; Gong, J.-J.; Liu, Z.-J.; Wu, Y.-Z.; Chen, H.-S. Designing Donors and Nonfullerene Acceptors for Organic Solar Cells Assisted by Machine Learning and Fragment-Based Molecular Fingerprints. *Solar RRL* **2025**, *9*, 2400846.
(30) Zhang, C.-R.; Lv, L.-F.; Li, M.; Liu, X.-M.; Gong, J.-J.; Liu, Z.-J.; Wu, Y.-Z.; Chen, H.-S. 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.
(31) Lv, L.-F.; Zhang, C.-R.; Cao, R.; Liu, X.-M.; Zhang, M.-L.; Gong, J.-J.; Liu, Z.-J.; Wu, Y.-Z.; Chen, H.-S. 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.
(32) Lv, L.-F.; Zhang, C.-R.; Sang, C.-C.; Liu, X.-M.; Zhang, M.-L.; Gong, J.-J.; Chen, Y.-H.; Chen, H.-S. Integrating deep learning and symbolic regression for molecular design and virtual screening of organic solar cells. *npj Computational Materials* **2026**, *12*, 31.
(33) Qiu, J.; Lam, H. H.; Hu, X.; Li, W.; Fu, S.; Zeng, F.; Zhang, H.; Wang, X. Accelerating High-Efficiency Organic Photovoltaic Discovery via Pretrained Graph Neural Networks and Generative Reinforcement Learning. *arXiv preprint arXiv:2503.23766*, 2025.
(34) Han, G.; Yi, Y. Molecular insight into efficient charge generation in low-driving-force nonfullerene organic solar cells. *Acc. Chem. Res.* **2022**, *55*, 869−877.
(35) Khatua, R.; Das, B.; Mondal, A. Rational design of non-fullerene acceptors via side-chain and terminal group engineering: a computational study. *Phys. Chem. Chem. Phys.* **2023**, *25*, 7994−8004.
(36) Zhu, L.; Zhang, J.; Guo, Y.; Yang, C.; Yi, Y.; Wei, Z. Small exciton binding energies enabling direct charge photogeneration towards low-driving-force organic solar cells. *Angew. Chem.* **2021**, *133*, 15476−15481.
(37) Kuzmich, A.; Padula, D.; Ma, H.; Troisi, A. Trends in the electronic and geometric structure of non-fullerene based acceptors for organic solar cells. *Energy Environ. Sci.* **2017**, *10*, 395−401.
(38) Ma, H.; Troisi, A. 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.
(39) Liu, T.; Troisi, A. What makes fullerene acceptors special as electron acceptors in organic solar cells and how to replace them. *Adv. Mater.* **2013**, *25*, 1038−1041.
(40) Liu, T.; Troisi, A. Absolute rate of charge separation and recombination in a molecular model of the P3HT/PCBM interface. *J. Phys. Chem. C* **2011**, *115*, 2406−2415.
(41) Yan, J.; Rodríguez-Martínez, X.; Pearce, D.; Douglas, H.; Bili, D.; Azzouzi, M.; Eisner, F.; Virbule, A.; Rezasoltani, E.; Belova, V.; Dörling, B.; Few, S.; Szumska, A. A.; Hou, X.; Zhang, G.; Yip, H.-L.; Campoy-Quiles, M.; Nelson, J. Identifying structure-absorption relationships and predicting absorption strength of non-fullerene acceptors for organic photovoltaics. *Energy Environ. Sci.* **2022**, *15*, 2958−2973.
(42) Ren, Y.-T.; Zhang, C.-R.; Zhang, M.-L.; Liu, X.-M.; Gong, J.-J.; Chen, Y.-H.; Liu, Z.-J.; Wu, Y.-Z.; Chen, H.-S. 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.
(43) Rehman, F. u.; Waqas, M.; Imran, M.; Ibrahim, M. A.; Iqbal, J.; Khera, R. A.; Hadia, N.; Al-Saeedi, S. I.; Shaban, M. 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.
(44) Khatua, R.; Mondal, A. Design and screening of B-N functionalized non-fullerene acceptors for organic solar cells via multiscale computation. *Mater. Adv.* **2023**, *4*, 4425−4435.
(45) Park, J.; Shim, Y.; Lee, F.; Rammohan, A.; Goyal, S.; Shim, M.; Jeong, C.; Kim, D. S. Prediction and interpretation of polymer properties using the graph convolutional network. *ACS Polymers Au* **2022**, *2*, 213−222.
(46) Zhi, H.-Y.; Zhao, L.; Lee, C.-C.; Chen, C. Y.-C. A novel graph neural network methodology to investigate dihydroorotate dehydrogenase inhibitors in small cell lung cancer. *Biomolecules* **2021**, *11*, 477.
(47) Dai, M.; Demirel, M. F.; Liang, Y.; Hu, J.-M. Graph neural networks for an accurate and interpretable prediction of the properties of polycrystalline materials. *npj Computational Materials* **2021**, *7*, 103.
(48) Landrum, G. R. *A software suite for cheminformatics, computational chemistry, and predictive modeling*, 2013.
(49) Wang, M.; Zheng, D.; Ye, Z.; Gan, Q.; Li, M.; Song, X.; Zhou, J.; Ma, C.; Yu, L.; Gai, Y.; Xiao, T.; He, T.; Karypis, G.; Li, J.; Zhang, Z. Deep graph library: A graph-centric, highly-performant package for graph neural networks. *arXiv preprint arXiv:1909.01315*, 2019.
(50) Akiba, T.; Sano, S.; Yanase, T.; Ohta, T.; Koyama, M. Optuna: A next-generation hyperparameter optimization framework. *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining* **2019**, 2623−2631.
(51) McHale, J. L. *Molecular spectroscopy*, 2nd ed.; CRC Press, 2017.
(52) Zheng, L.; Polizzi, N. F.; Dave, A. R.; Migliore, A.; Beratan, D. N. Where is the electronic oscillator strength? Mapping oscillator strength across molecular absorption spectra. *J. Phys. Chem. A* **2016**, *120*, 1933−1943.
(53) Zhao, C. X.; Xiao, S.; Xu, G. Density of organic thin films in organic photovoltaics. *J. Appl. Phys.* **2015**, *118*, 044510.
(54) O'Boyle, N. M.; Banck, M.; James, C. A.; Morley, C.; Vandermeersch, T.; Hutchison, G. R. Open Babel: An open chemical toolbox. *J. Cheminf.* **2011**, *3*, 33.
(55) Halgren, T. A. Merck molecular force field. I. Basis, form, scope, parameterization, and performance of MMFF94. *J. Comput. Chem.* **1996**, *17*, 490−519.
(56) Halgren, T. A. Merck molecular force field. II. MMFF94 van der Waals and electrostatic parameters for intermolecular interactions. *J. Comput. Chem.* **1996**, *17*, 520−552.
(57) Halgren, T. A. Merck molecular force field. III. Molecular geometries and vibrational frequencies for MMFF94. *J. Comput. Chem.* **1996**, *17*, 553−586.
(58) Halgren, T. A.; Nachbar, R. B. Merck molecular force field. IV. Conformational energies and geometries for MMFF94. *J. Comput. Chem.* **1996**, *17*, 587−615.

https://doi.org/10.1021/acsami.5c26136
ACS Appl. Mater. Interfaces XXXX, XXX, XXX−XXX

(59) Halgren, T. A. Merck molecular force field. V. Extension of MMFF94 using experimental data, additional computational data, and empirical rules. *J. Comput. Chem.* **1996**, *17*, 616−641.
(60) Seifert, G.; Porezag, D.; Frauenheim, T. Calculations of molecules, clusters, and solids with a simplified LCAO-DFT-LDA scheme. *Int. J. Quantum Chem.* **1996**, *58*, 185−192.
(61) Elstner, M.; Porezag, D.; Jungnickel, G.; Elsner, J.; Haugk, M.; Frauenheim, T.; Suhai, S.; Seifert, G. Self-consistent-charge density-functional tight-binding method for simulations of complex materials properties. *Phys. Rev. B* **1998**, *58*, 7260.
(62) Gaus, M.; Cui, Q.; Elstner, M. DFTB3: Extension of the self-consistent-charge density-functional tight-binding method (SCC-DFTB). *J. Chem. Theory Comput.* **2011**, *7*, 931−948.
(63) Tkatchenko, A.; DiStasio, R. A., Jr.; Car, R.; Scheffler, M. Accurate and efficient method for many-body van der Waals interactions. *Phys. Rev. Lett.* **2012**, *108*, 236402.
(64) Ambrosetti, A.; Reilly, A. M.; DiStasio, R. A.; Tkatchenko, A. Long-range correlation energy calculated from coupled atomic response functions. *J. Chem. Phys.* **2014**, *140*, 18A508.
(65) Stöhr, M.; Michelitsch, G. S.; Tully, J. C.; Reuter, K.; Maurer, R. J. Communication: Charge-population based dispersion interactions for molecules and materials. *J. Chem. Phys.* **2016**, *144*, 151101.
(66) Mortazavi, M.; Brandenburg, J. G.; Maurer, R. J.; Tkatchenko, A. Structure and stability of molecular crystals with many-body dispersion-inclusive density functional tight binding. *J. Phys. Chem. Lett.* **2018**, *9*, 399−405.
(67) Aradi, B.; Hourahine, B.; Frauenheim, T. D. a Sparse Matrix-Based Implementation of the DFTB Method. *J. Phys. Chem. A* **2007**, *111*, 5678−5684.
(68) Larsen, A. H.; Mortensen, J. J.; Blomqvist, J.; Castelli, I. E.; Christensen, R.; Dułak, M.; Friis, J.; Groves, M. N.; Hammer, B.; Hargus, C.; Hermes, E. D.; Jennings, P. C.; Jensen, P. B.; Kermode, J.; Kitchin, J. R.; Kolsbjerg, E. L.; Kubal, J.; Kaasbjerg, K.; Lysgaard, S.; Maronsson, J. B.; Maxson, T.; Olsen, T.; Pastewka, L.; Peterson, A.; Rostgaard, C.; Schiøtz, J.; Schütt, O.; Strange, M.; Thygesen, K. S.; Vegge, T.; Vilhelmsen, L.; Walter, M.; Zeng, Z.; Jacobsen, K. W. The atomic simulation environment-a Python library for working with atoms. *J. Phys.: Condens. Matter* **2017**, *29*, 273002.
(69) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, O.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. *Gaussian 09*, revision D.01; Gaussian, Inc.: Wallingford, CT, 2009.
(70) Wang, J.; Xie, Y.; Chen, K.; Wu, H.; Hodgkiss, J. M.; Zhan, X. Physical insights into non-fullerene organic photovoltaics. *Nature Reviews Physics* **2024**, *6*, 365−381.
(71) Knupfer, M. Exciton binding energies in organic semiconductors. *Appl. Phys. A: Mater. Sci. Process.* **2003**, *77*, 623−626.
(72) Sugie, A.; Nakano, K.; Tajima, K.; Osaka, I.; Yoshida, H. Dependence of exciton binding energy on bandgap of organic semiconductors. *J. Phys. Chem. Lett.* **2023**, *14*, 11412−11420.
(73) Coropceanu, V.; Chen, X.-K.; Wang, T.; Zheng, Z.; Brédas, J.-L. Charge-transfer electronic states in organic solar cells. *Nature Reviews Materials* **2019**, *4*, 689−707.
(74) Pelzer, K. M.; Darling, S. B. Charge generation in organic photovoltaics: a review of theory and computation. *Molecular Systems Design & Engineering* **2016**, *1*, 10−24.
(75) Chen, X.-K.; Coropceanu, V.; Brédas, J.-L. Assessing the nature of the charge-transfer electronic states in organic solar cells. *Nat. Commun.* **2018**, *9*, 5295.
(76) Zhang, G.; Chen, X.-K.; Xiao, J.; Chow, P. C.; Ren, M.; Kupgan, G.; Jiao, X.; Chan, C. C.; Du, X.; Xia, R.; et al. Delocalization of exciton and electron wavefunction in non-fullerene acceptor molecules enables efficient organic solar cells. *Nat. Commun.* **2020**, *11*, 3943.
(77) Mahadevan, S.; Liu, T.; Pratik, S. M.; Li, Y.; Ho, H. Y.; Ouyang, S.; Lu, X.; Yip, H.-L.; Chow, P. C.; Brédas, J.-L.; et al. Assessing intra- and inter-molecular charge transfer excitations in non-fullerene acceptors using electroabsorption spectroscopy. *Nat. Commun.* **2024**, *15*, 2393.
(78) Wu, X.-F.; Fu, W.-F.; Xu, Z.; Shi, M.; Liu, F.; Chen, H.-Z.; Wan, J.-H.; Russell, T. P. Spiro linkage as an alternative strategy for promising nonfullerene acceptors in organic solar cells. *Adv. Funct. Mater.* **2015**, *25*, 5954−5966.
(79) Li, S.; Liu, W.; Shi, M.; Mai, J.; Lau, T.-K.; Wan, J.; Lu, X.; Li, C.-Z.; Chen, H. A spirobifluorene and diketopyrrolopyrrole moieties based non-fullerene acceptor for efficient and thermally stable polymer solar cells with high open-circuit voltage. *Energy Environ. Sci.* **2016**, *9*, 604−610.
(80) Qiu, N.; Yang, X.; Zhang, H.; Wan, X.; Li, C.; Liu, F.; Zhang, H.; Russell, T. P.; Chen, Y. Nonfullerene small molecular acceptors with a three-dimensional (3D) structure for organic solar cells. *Chem. Mater.* **2016**, *28*, 6770−6778.
(81) Lee, H.; Oh, S.; Song, C. E.; Lee, H. K.; Lee, S. K.; Shin, W. S.; So, W.-W.; Moon, S.-J.; Lee, J.-C. Stable P3HT: amorphous non-fullerene solar cells with a high open-circuit voltage of 1 V and efficiency of 4%. *RSC Adv.* **2019**, *9*, 20733−20741.
(82) He, D.; Zhao, F.; Wang, C.; Lin, Y. Non-radiative recombination energy losses in non-fullerene organic solar cells. *Adv. Funct. Mater.* **2022**, *32*, 2111855.
(83) Liao, H.-C.; Ho, C.-C.; Chang, C.-Y.; Jao, M.-H.; Darling, S. B.; Su, W.-F. Additives for morphology control in high-efficiency organic solar cells. *Mater. Today* **2013**, *16*, 326−336.
(84) Zhao, F.; Wang, C.; Zhan, X. Morphology control in organic solar cells. *Adv. Energy Mater.* **2018**, *8*, 1703147.
(85) Sandberg, O. J.; Armin, A. Energetics and kinetics requirements for organic solar cells to break the 20% power conversion efficiency barrier. *J. Phys. Chem. C* **2021**, *125*, 15590−15598.
(86) Wang, D. H.; Morin, P.-O.; Lee, C.-L.; Kyaw, A. K. K.; Leclerc, M.; Heeger, A. J. Effect of processing additive on morphology and charge extraction in bulk-heterojunction solar cells. *Journal of Materials Chemistry A* **2014**, *2*, 15052−15057.
(87) Ertl, P.; Schuffenhauer, A. Estimation of synthetic accessibility score of drug-like molecules based on molecular complexity and fragment contributions. *Journal of cheminformatics* **2009**, *1*, 8.
(88) Bickerton, G. R.; Paolini, G. V.; Besnard, J.; Muresan, S.; Hopkins, A. L. Quantifying the chemical beauty of drugs. *Nature Chem.* **2012**, *4*, 90−98.

https://doi.org/10.1021/acsami.5c26136
ACS Appl. Mater. Interfaces XXXX, XXX, XXX−XXX