JCTC Journal of Chemical Theory and Computation

pubs.acs.org/JCTC Article

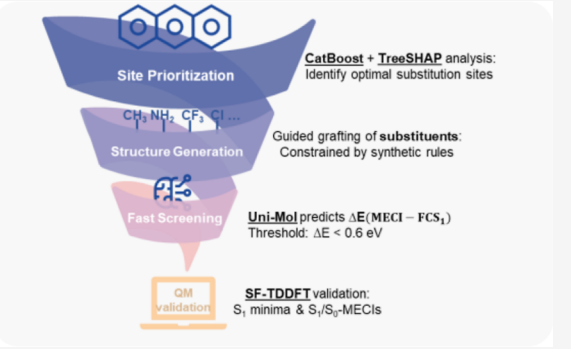
Machine-Learning-Accelerated Rational Design of Aggregation-Induced Emission Luminogens

Wenjie Zhang,[‖] Ping-An Yin,[‖] Qi Ou,* and Zhigang Shuai*

Cite This: *J. Chem. Theory Comput.* 2026, 22, 8825−8835 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: The rational design of aggregation-induced emission (AIE) luminogens presents a significant challenge in molecular photophysics, requiring approaches that connect mechanistic understanding with practical molecular screening. This work uses the electronic-energy difference between the S_1/S_0 minimum-energy conical intersection (MECI) and the vertically accessed Franck−Condon S_1 state, ΔE(MECI − FCS_1), as an efficient descriptor of conical-intersection accessibility. We compile quantum-chemical labels for 228 structure-matched polycyclic aromatic molecules and develop a dual-model strategy that combines an interpretable fingerprint-based model with a Uni-Mol model for rapid property prediction. In an external panel of literature luminogens, the predicted values are significantly lower for AIE molecules than those for aggregation-caused-quenching (ACQ) molecules, supporting an empirical operating threshold near 0.6 eV. This threshold provides an efficient first-pass guide for prioritizing candidates with accessible CI channels, substantially narrowing the pool for subsequent validation. The resulting workflow connects mechanistic insight, interpretable design rules, and high-throughput screening, providing an efficient platform for accelerating the discovery of novel AIEgens.



INTRODUCTION

The discovery of aggregation-induced emission (AIE) in 2001 by Tang's group represents a landmark achievement in photophysics and materials science.[1] In stark contrast to the conventional aggregation-caused quenching (ACQ) effect, AIE luminogens (AIEgens) exhibit weak or no emission in the dissolved state but demonstrate significantly enhanced luminescence upon aggregation or in the solid state.[2,3] This unique phenomenon has unlocked immense application potential across diverse fields,[4−8] including solid-state organic light-emitting devices,[9−13] bioimaging,[14−17] chemical sensing,[18−20] and environmental monitoring.[21−23] The performance of AIEgens, however, is critically governed by their molecular structure and the precise packing modes in the aggregated state.[24−29] Consequently, transcending traditional trial-and-error approaches to enable the rational and efficient design of AIE molecules remains a central challenge for advancing the field.[30−32]

The mechanistic understanding of AIE is predominantly attributed to the restriction of intramolecular motion (RIM) in the aggregate state, which suppresses nonradiative decay pathways and promotes radiative transitions.[2,33] It is noteworthy, however, that other photophysical processes, such as the formation of J-aggregates with specific stacking modes, can also lead to enhanced emission in the solid state, broadening the conceptual framework of AIE.[25,34] Quantum chemical calculations, particularly time-dependent density functional theory (TDDFT) and its advanced variants such as spin-flip TDDFT (SF-TDDFT), coupled with theories such as the thermal vibration correlation function (TVCF), have proven indispensable for elucidating these complex photophysical processes.[35−37] They also permit direct characterization of the relative electronic energy of an S_1/S_0 conical-intersection region with respect to the Franck−Condon excited state, providing a computable measure of the energetic accessibility of this nonradiative channel.[38−40] Despite their precision, these first-principles methods are computationally prohibitive and require significant expertise, rendering them impractical for large-scale chemical space exploration and constituting a major bottleneck for the high-throughput screening and rational design of AIEgens.

Machine learning (ML) has emerged as a transformative paradigm for accelerating the discovery of functional molecules by learning intricate structure−property relationships from existing chemical data.[41,42] The utility of different ML algorithms in molecular design, however, varies profoundly,

Received: June 8, 2026
Revised: August 12, 2026
Accepted: August 13, 2026
Published: August 21, 2026

https://doi.org/10.1021/acs.jctc.6c01107
J. Chem. Theory Comput. 2026, 22, 8825−8835

rooted in their fundamental approaches to molecular representation. On one hand, tree-based descriptor models such as XGBoost and CatBoost, which rely on human-engineered molecular descriptors (e.g., Morgan fingerprints), can be interpreted using feature-attribution methods.[43−46] They facilitate the tracing of atomic or functional group contributions, providing chemists with intuitive guidance for structural modification. Conversely, three-dimensional deep learning models, such as Uni-Mol, learn feature representations directly from atomic coordinates.[47] This approach inherently captures critical three-dimensional structural and conformational information, typically leading to superior predictive accuracy,[48] albeit at the cost of interpretability, often resulting in "black-box" models.[49] Thus, for the complex task of AIE molecular design, interpretable models excel at guiding structural modifications, while high-accuracy deep learning models are superior at providing precise, quantitative predictions of key photophysical properties.

Informed by this complementarity, we present an integrated computational workflow for the rational design of AIEgens, underpinned by 228 structure-matched values of the electronic-energy difference between the S_1/S_0 minimum-energy conical intersection (MECI) and the vertically accessed Franck−Condon S_1 state, i.e., $\Delta E(\mathrm{MECI} - \mathrm{FCS}_1)$. This descriptor places the MECI electronic energy relative to the vertical S_1 state at the optimized S_0 geometry. While our previous studies used the Gibbs free-energy barrier as a descriptor,[39,40] the present work uses this purely electronic property since it avoids computationally expensive frequency analysis and, in fact, it is strongly correlated to $\Delta G(\mathrm{MECI} - \mathrm{FCS}_1)$ as detailed in our calculation. Our workflow employs two complementary models: a fingerprint-based model for interpretable structure−property analysis and a Uni-Mol model for rapid prediction from three-dimensional ground-state geometries. The Uni-Mol predictions further reproduce a lower-ΔE trend for an external panel of literature AIE molecules relative to ACQ molecules, which validates its generalizability to various organic fluorophores. Together, these components provide a practical platform for molecular prioritization that is designed to feed into subsequent quantum-chemical and experimental validation.

METHOD

Physical Basis of AIE and the $\Delta E(\mathrm{MECI} - \mathrm{FCS}_1)$ Descriptor

The AIE phenomenon arises when radiative decay competes more favorably against nonradiative channels upon aggregation. In solution, low-frequency out-of-plane and torsional motions open efficient internal-conversion pathways, quenching fluorescence; in the aggregate/solid, these motions are sterically restricted, nonradiative channels are throttled, and the photoluminescence quantum yield (PLQY) increases.[2,4,33] Among nonradiative pathways, the S_1/S_0 minimum-energy conical intersection (MECI) provides a particularly efficient funnel for radiationless decay, typically accessed by a large out-of-plane "puckering/twist" along a specific geometric coordinate (e.g., central-ring pyramidalization).[50] The basic kinetic picture is

$$k_{\mathrm{tot}} = k_{\mathrm{r}} + k_{\mathrm{nr}}^{\mathrm{TVCF}} + k_{\mathrm{nr}}^{\mathrm{MECI}}, \qquad \Phi_{\mathrm{fl}} = \frac{k_{\mathrm{r}}}{k_{\mathrm{tot}}} \tag{1}$$

where $k_{\mathrm{nr}}^{\mathrm{TVCF}}$ captures harmonic-region internal conversion (and, when relevant, ISC), and $k_{\mathrm{nr}}^{\mathrm{MECI}}$ denotes relaxation through the conical-intersection region.

We quantify the relative electronic energy of this crossing region by

$$\Delta E(\mathrm{MECI} - \mathrm{FCS}_1) = E_{\mathrm{S}_0}(\mathbf{R}_{\mathrm{MECI}}) - E_{\mathrm{S}_1}(\mathbf{R}_{\mathrm{S}_0\text{-min}}) \tag{2}$$

where FCS_1 is the vertically excited S_1 state evaluated at the optimized S_0 geometry. The S_0 root is used for the MECI energy; the S_0 and S_1 roots are degenerate at an ideally converged intersection. A smaller $\Delta E(\mathrm{MECI} - \mathrm{FCS}_1)$ places the crossing region lower relative to the initially accessed excited-state configuration and therefore indicates greater electronic accessibility of a CI-mediated relaxation channel.

Note that in the gas-phase isolated-molecule model used here, a smaller ΔE is consistent with a potentially efficient nonradiative pathway in the unconstrained state. It does not, by itself, determine the molecular emission efficiency or aggregate-state photoluminescence quantum yield. AIE becomes plausible only when access to this channel requires a large-amplitude deformation that is selectively restricted upon aggregation. Rigorous validation requires the computation of the same quantity in the aggregated or crystallized state, which is unfortunately hard to be obtained both theoretically and experimentally for novel molecules. The gas-phase energy difference we choose in this work can therefore serve as a quick first-pass screening protocol to improve the efficiency of novel AIE molecule design.

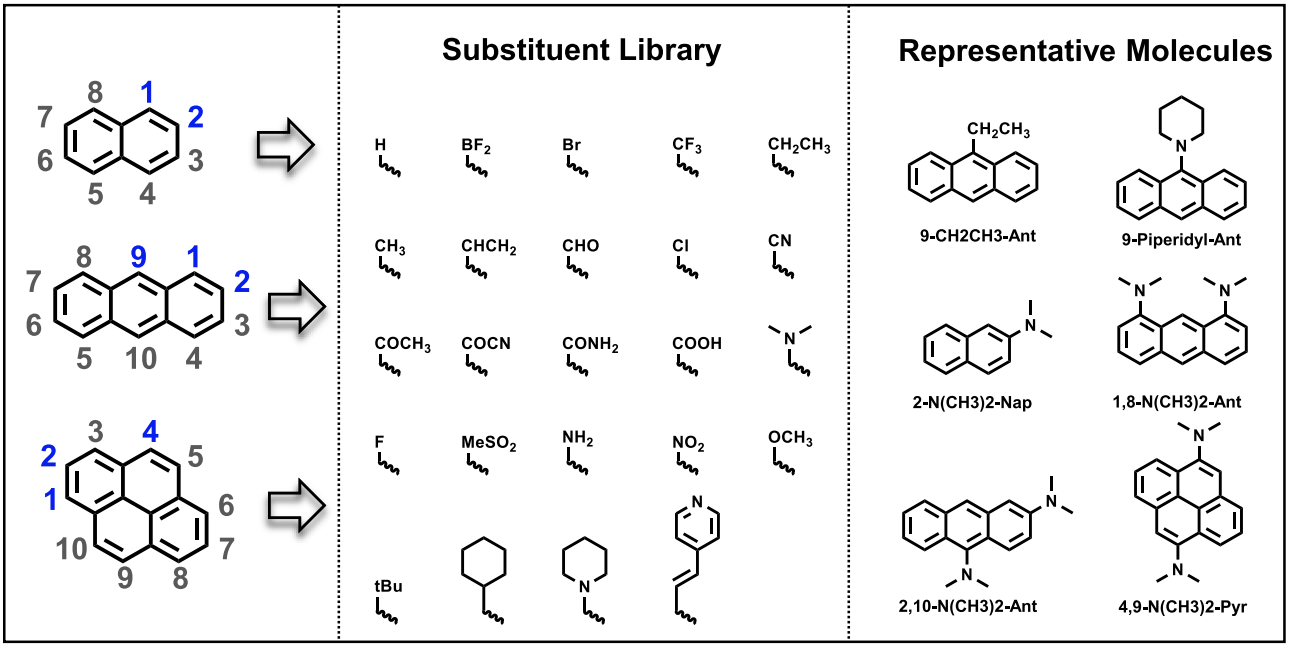
All electronic energies required for the labels were computed with spin-flip time-dependent density functional theory (SF-TDDFT)[35] using the def2-SVP basis and the long-range-corrected ωB97X-D functional, with the range-separation parameter ω tuned per molecular system so that the Koopmans' theorem is guaranteed.[51,52] Calculations were performed in Q-CHEM 5.3.[53,54] The optimized S_0 geometry supplies both the molecular coordinates used by the machine-learning model and the geometry for the vertical S_1 energy. The S_1/S_0 MECI was located with the penalty-function approach using α = 0.02 a.u. and σ = 3.5. The MECI search was initiated from the Franck−Condon point, i.e., the vertical S_1 geometry at the optimized S_0 structure.[55] Further computational details are provided in the Supporting Information.

Data Set Construction

Following earlier findings,[39,40] we focus on the S_1/S_0 MECI of fused aromatic rings. Detailed illustrations and a quantitative conformational analysis of anthracene-derived MECI structures were provided in our previous study,[40] in which the noncoplanarity of the anthracene framework was characterized using three interplanar dihedral angles. That analysis showed that the dominant MECI deformation of anthracene and most monosubstituted derivatives is localized around the central ring, particularly along the C9−X bond, where X denotes hydrogen or a substituent, producing pronounced out-of-plane puckering. It is this large-amplitude central-ring deformation that provides the mechanistic basis considered in the present work. The present target, $\Delta E(\mathrm{MECI} - \mathrm{FCS}_1)$, quantifies the electronic accessibility of this crossing region from the initially excited configuration. Aggregation can suppress the associated nonradiative channel by restricting the required out-of-plane deformation. Note that this purely electronic property ΔE is computationally more efficient and is strongly correlated with

https://doi.org/10.1021/acs.jctc.6c01107
J. Chem. Theory Comput. 2026, 22, 8825−8835

Scheme 1. Molecular Library Used in This Work[a]



[a]The left panel shows the naphthalene, anthracene, and pyrene parent scaffolds and their atom-numbering conventions; blue numbers denote the selected substitution sites. The center panel shows the substituent library. The right panel shows representative substituted molecules constructed by combining the parent scaffolds and substituents.

the Gibbs free-energy barrier ΔG used in our previous studies for the anthracene derivatives (Figure S1). The vibrational and thermal corrections do not significantly alter the relative ordering of MECI accessibility for the polycyclic systems examined here. This correlation justifies the use of ΔE as a computationally efficient proxy for the more expensive ΔG in the present screening workflow.

For clarity, the molecular library is summarized in Scheme 1, which presents the three parent polycyclic aromatic hydrocarbon scaffolds and their atom-numbering conventions, the substituent library, and representative substituted molecules constructed from these components.

Guided by these principles, we first design **17** disubstituted anthracene derivatives. We then extend this molecular framework to a broader data set of polycyclic aromatics, including naphthalene, anthracene, and pyrene, chosen for their versatile substitution sites and inherently high thermal stability. To systematically explore isomerism, we include various $N(CH_3)_2$ disubstitution patterns on naphthalene and pyrene, covering ortho-, meta-, and para-relationships. This enumeration produced an initial pool of 244 designed structures; 228 structures with matched ΔE labels and ground-state geometries were retained for the present model.

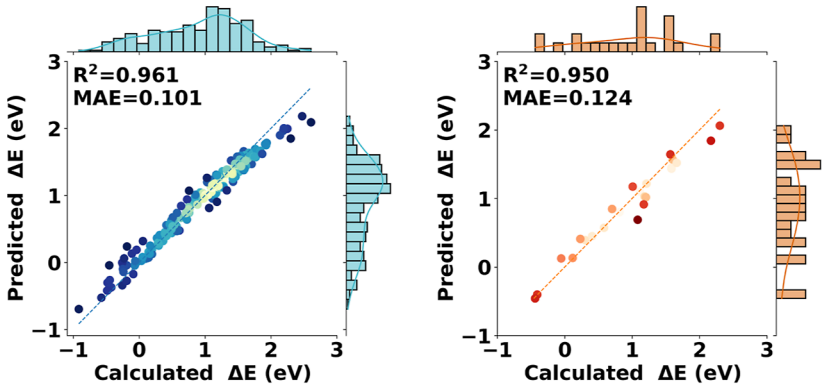
For each structure-matched molecule, the supervised-learning label is $\Delta E(\mathrm{MECI} - \mathrm{FCS}_1)$ from eq 2. After matching the electronic-energy records to optimized S_0 structures and applying the model-input checks, 228 molecules were retained for the present Uni-Mol study. Their labels span −0.914 to 2.609 eV. The data set was randomly divided at an approximately 9:1 ratio using stratification across the ΔE distribution, yielding a training set of 204 molecules and an independent held-out test set of 24 molecules.

Pretrained SE(3) 3D Encoder for End-to-End ΔE Prediction

Direct evaluation of $\Delta E(\mathrm{MECI} - \mathrm{FCS}_1)$ requires a vertical excited-state calculation and an S_1/S_0 MECI optimization for every molecule, which is prohibitive for high-throughput screening. We therefore adopt an end-to-end regressor that maps a relaxed ground-state geometry ($\mathbf{Z}$, $\mathbf{R}$) (atomic types and Cartesian coordinates) to ΔE. Because the labeled set contains 228 molecules, we leverage a large pretrained 3D encoder to improve data efficiency.[47]

We use the Uni-Mol model, a 3D molecular representation learner whose backbone is a transformer with pair representations and an SE(3)-equivariant coordinate head. Uni-Mol is pretrained by self-supervision on **209M** molecular conformations (and, separately, on 3M protein pockets), then fine-tuned on downstream tasks.[47] The molecular pretraining corpus was constructed from approximately 19 million normalized and deduplicated organic molecules collected from large molecular databases, including ZINC and ChEMBL. For each molecule, ten three-dimensional conformations were generated using RDKit and supplemented by one two-dimensional conformation derived from the molecular graph, yielding approximately 209 million conformations in total. The published description of the corpus does not provide a molecule-by-molecule inventory; therefore, it is not possible to determine whether anthracene and pyrene were explicitly included. Importantly, even if these parent scaffolds were present in the unlabeled pretraining corpus, no $\Delta E(\mathrm{MECI} - \mathrm{FCS}_1)$, MECI, excited-state, or AIE/ACQ labels were used during pretraining. Pretraining combines three losses on *corrupted* conformations: masked-atom prediction, pair-distance regression, and SE(3)-equivariant coordinate denoising; the total loss is the (rescaled) sum of these components. For scalar property prediction, a learnable `[CLS]` token is prepended to the atom sequence. After processing by the transformer encoder, its final hidden state aggregates the

https://doi.org/10.1021/acs.jctc.6c01107
J. Chem. Theory Comput. 2026, 22, 8825−8835



(a) Training set. (b) Test set.

Figure 1. Predicted versus calculated ΔE(MECI − FCS_1) values for the training and test sets. Panel (a) displays 204 training molecules. Panel (b) contains the 24 test molecules. The MAE values shown in the figures are in the unit of eV.

atomic and three-dimensional pair-representation information on the entire molecule and serves as the molecule-level input to the regression head. In this regression task, `[CLS]` functions as a global summary token rather than a classification label.

Formally, let E_ϕ denote the pretrained encoder producing a molecular embedding $\mathbf{z} = E_\phi(\mathbf{Z}, \mathbf{R})$. For rigid motions $(\mathbf{Q}, \mathbf{t}) \in \mathrm{SO}(3) \times \mathbb{R}^3$, Uni-Mol's representation used for scalar prediction is *invariant*

$$E_\phi(\mathbf{Z}, \mathbf{R}\mathbf{Q}^{\mathrm{T}} + \mathbf{1}\mathbf{t}^{\mathrm{T}}) = E_\phi(\mathbf{Z}, \mathbf{R}) \tag{3}$$

whereas its coordinate head (used during pretraining) is *equivariant*

$$C_\phi(\mathbf{R}\mathbf{Q}^{\mathrm{T}} + \mathbf{1}\mathbf{t}^{\mathrm{T}}) = C_\phi(\mathbf{R})\mathbf{Q}^{\mathrm{T}} + \mathbf{1}\mathbf{t}^{\mathrm{T}} \tag{4}$$

consistent with the update rule of the SE(3)-equivariant head.[56,57] We then learn a regressor g_ψ such that

$$\widehat{\Delta E} = g_\psi(E_\phi(\mathbf{Z}, \mathbf{R})) \tag{5}$$

optimized by mean-squared error

$$\mathcal{L}_{\mathrm{MSE}} = \frac{1}{|\mathcal{D}|} \sum_{i \in \mathcal{D}} (\Delta E_i - g_\psi(E_\phi(\mathbf{Z}_i, \mathbf{R}_i)))^2 \tag{6}$$

where $\boldsymbol{\theta} = (\boldsymbol{\phi}, \boldsymbol{\psi})$ collects trainable parameters. In practice, we initialize from the Uni-Mol v1 molecular checkpoint, feed relaxed S_0 coordinates, pool via `[CLS]`, and fine-tune a default head (`dropout-linear-tanh-dropout-linear`) using `unimol_tools`. This transfer setting imports general geometry priors learned from a large unlabeled 3D corpus, improving data efficiency and regularization on the present $\{(\mathbf{Z}_i, \mathbf{R}_i, \Delta E_i)\}_{i=1}^{228}$ set while preserving the correct SE(3) symmetries for a scalar target.

To enable library-scale screening of conical-intersection accessibility, we fine-tuned Uni-Mol to predict ΔE(MECI − FCS_1) directly from relaxed ground-state S_0 geometries using a mean-squared-error loss. The 3D transformer encoder generates a whole-molecule embedding that is passed to a lightweight regression head. The complete training protocol and hyperparameters are provided in the Supporting Information, Section "Uni-Mol model training protocol" and Table S1. To complement the 3D model with chemical interpretation, we also trained descriptor-based regressors on Morgan fingerprints. SHAP feature attributions can be mapped through fingerprint provenance to local atomic environments; negative contributions lower the predicted ΔE, whereas positive contributions raise it.

RESULTS AND DISCUSSION

ML Model Performance and External Validation

Using the same approximately 9:1 stratified random split, we trained the Uni-Mol model and the descriptor-based regression models on 204 molecules and evaluated them on an independent held-out test set of 24 molecules, thereby ensuring a consistent comparison across all models. The model architecture of Uni-Mol takes the relaxed S_0 geometry, represented by atomic symbols and their 3D coordinates, as input.

The learning rate and batch size were selected through a compact 4 × 3 hyperparameter search on the training data, which evaluated learning rates of 1×10^{-4}, 3×10^{-5}, 1×10^{-5}, and 3×10^{-6} together with batch sizes of 8, 16, and 32. Each search run was limited to 60 epochs with an early stopping patience of 15 epochs. Based on the comparative model performance, a learning rate of 3×10^{-5} and a batch size of 16 were selected for the final training. The final Uni-Mol model was trained using `unimol-tools` for a maximum of 100 epochs with an early stopping patience of 25 epochs. A 10-fold predefined group split was used within the training set, and validation R^2 was used for checkpoint selection within each fold. The complete settings are reported in Table S1.

On the full 204-member training set, the fitted Uni-Mol predictions give R^2 = 0.961 and MAE = 0.101 eV. The independent 24-member held-out test set gives R^2 = 0.950 and MAE = 0.124 eV (Figure 1). The comparable training and test errors indicate that the geometry-aware pretrained model maintains its predictive accuracy on the held-out molecules. To assess its relative performance, we compared Uni-Mol against a representative set of established descriptor-based regression methods suitable for relatively small molecular data sets: RandomForest (bagging-based tree ensemble), XGBoost, LightGBM, and CatBoost (gradient-boosted decision trees with different optimization strategies), as well as SVR (kernel-based nonlinear regression). All five baseline models used the same binary radius-3 ECFP representation and the same data partition, enabling a consistent comparison among conventional two-dimensional fingerprint-based approaches.[44,45,58−61] As shown in Table 1, Uni-Mol substantially outperforms these

https://doi.org/10.1021/acs.jctc.6c01107
J. Chem. Theory Comput. 2026, 22, 8825−8835

Table 1. Model Comparison for ΔE(MECI − FCS_1) Prediction Using the Approximately 9:1 Stratified Random Training/Test Split (204/24 Molecules)[a]

model	$R^2_{train}(\uparrow)$	$R^2_{test}(\uparrow)$	$MAE_{train}(\downarrow)$	$MAE_{test}(\downarrow)$
RandomForest	0.669	0.673	0.325	0.337
LightGBM	0.461	0.590	0.388	0.332
XGBoost	0.609	0.631	0.360	0.376
CatBoost	0.810	0.704	0.256	0.331
SVR	0.953	0.623	0.032	0.389
Uni-Mol (ours)	**0.961**	**0.950**	**0.101**	**0.124**

[a]Descriptor baselines use the selected binary ECFP representation with radius 3. MAEs are in eV.

baselines, with the performance gap reflecting the combined benefits of three-dimensional molecular representation, geometry-aware pretraining, and the transformer-based regression architecture. This reliable gas-phase ΔE prediction enables efficient first-pass prioritization of candidates with accessible CI channels, substantially narrowing the candidate pool and making subsequent quantum-chemical verification and experimental validation significantly more manageable. Nevertheless, explicit quantum-chemical confirmation remains essential for individual screening hits.

To examine whether ΔE(MECI − FCS_1) remains informative beyond the labeled training scaffolds, we applied the Uni-Mol predictor to a literature panel containing 15 AIE and 15 ACQ luminogens. The AIE group has a mean predicted ΔE of 0.393 ± 0.134 eV, whereas the ACQ group has a mean of 0.812 ± 0.137 eV (Figure 2). The observed separation motivates an empirical first-pass operating threshold near 0.6 eV: lower values indicate a relatively accessible crossing region and are enriched in the AIE panel, whereas higher values are enriched in the ACQ panel. This external validation confirms that ΔE is an effective molecular descriptor for prioritizing AIE candidates, as it cleanly identifies molecules with accessible CI-mediated relaxation channels. That accessibility is the first mechanistic condition for AIE; the second, whether aggregation restricts the deformation needed to reach the CI, requires separate validation. Table S2 further compares the DFT-calculated ΔE values against the Uni-Mol predictions for the screened subset, demonstrating good agreement and supporting the reliability of the model for first-pass screening.

Note that our labeled training set is focused on naphthalene, anthracene, pyrene, and their substituted derivatives. Despite this focused chemical space, the Uni-Mol model successfully transfers three-dimensional chemical priors to the present ΔE prediction task, as demonstrated by its ability to discriminate AIE from ACQ molecules in an external literature panel. This indicates that our workflow provides a practical and transferable descriptor for prioritizing AIE candidates within the widely explored polycyclic aromatic chemical space. While the present model is expected to be most reliable for organic luminogens whose nonradiative decay is governed by similar large-amplitude out-of-plane distortions and S_1/S_0 MECI accessibility, additional quantum-chemical labels and further validation would be required for substantially different classes, such as strongly charge-transfer systems, metal complexes, or packing-dominated emitters.

Branching-Plane Analysis of 9-COCN-Ant

To connect the scalar ΔE descriptor with the local topology of the crossing region, we analyzed the S_1/S_0 MECI of the representative anthracene derivative 9-COCN-Ant. The S_1/S_0 MECI of 9-COCN-Ant was optimized in the gas phase at the SF-TDDFT/ωB97X-D/def2-SVP level using Q-CHEM 5.3. The range-separation parameter was tuned specifically for this molecule following the same system-specific procedure used for the remaining data set. A high-spin triplet reference was employed, and the MECI was located using the penalty-function method with α = 0.02 a.u. and σ = 3.5. The gradient-difference vector **g** and derivative-coupling vector **h** were evaluated at the optimized MECI using the same electronic-structure level. The general computational protocol is described in the Supporting Information. For a two-state crossing, the degeneracy is lifted along two nuclear directions: the gradient-difference vector **g** and the derivative-coupling vector **h**. Together, these two vectors span the branching plane, within which the adiabatic energy gap opens linearly near the crossing point.[38,62] For a small displacement δ**R** from the MECI, the two-state energy splitting can be approximated as

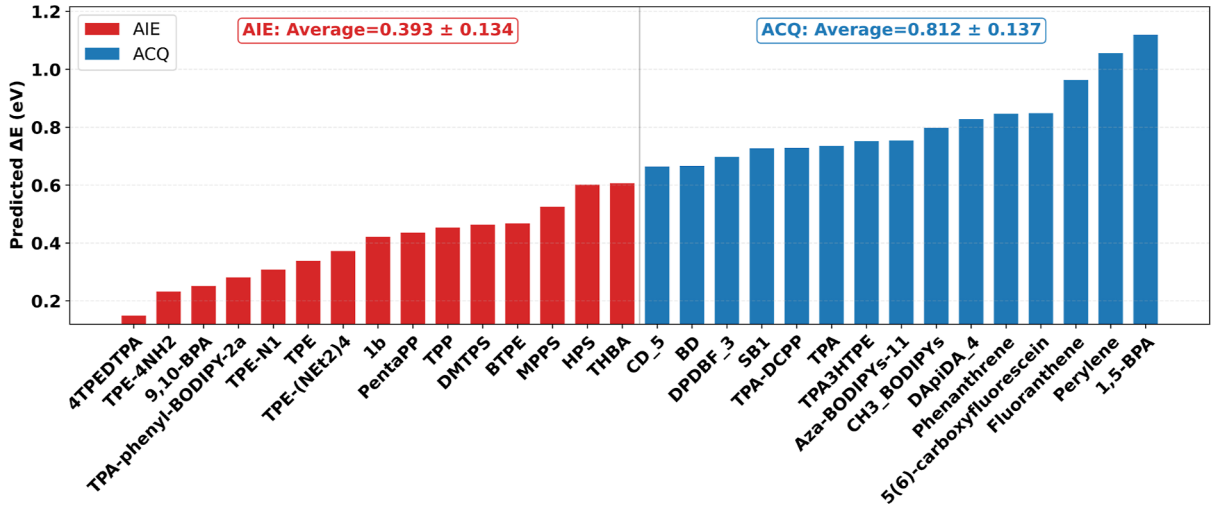
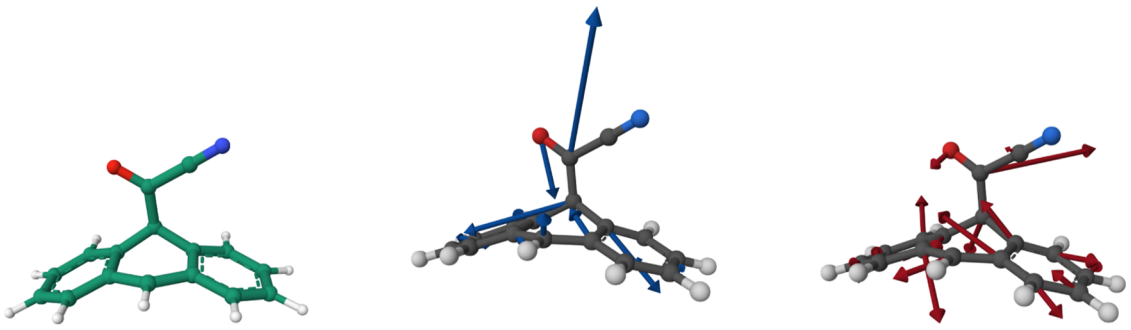


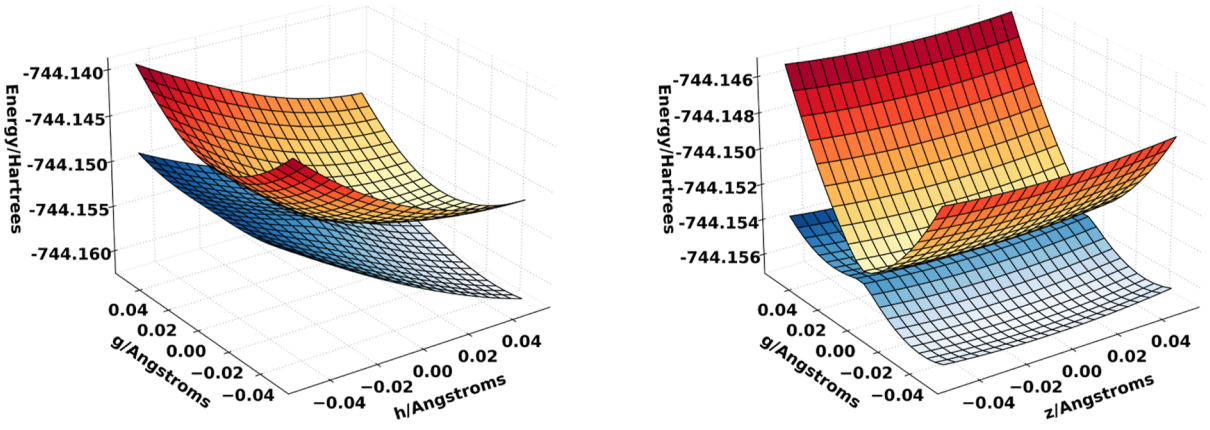
Figure 2. Predicted ΔE(MECI − FCS_1) values for the external literature panel of AIE (red, n = 15) and ACQ (blue, n = 15) luminogens. The group means and sample standard deviations are 0.393 ± 0.134 and 0.812 ± 0.137 eV, respectively. Values near 0.6 eV are treated as an empirical borderline region rather than a universal classification boundary.

https://doi.org/10.1021/acs.jctc.6c01107
J. Chem. Theory Comput. 2026, 22, 8825−8835



(a) Structure. (b) **g** vector. (c) **h** vector.

Figure 3. Structure (a) and branching-plane vectors of 9-COCN-Ant near the S_1/S_0 MECI. The (b) **g** vector is mainly associated with in-plane bond-length alternation and conjugated-framework reorganization, whereas the (c) **h** vector contains stronger out-of-plane components involving the carbonyl-containing substituent and local puckering motions.



(a) g-h branching plane. (b) g-z cross section.

Figure 4. Local potential-energy topology of 9-COCN-Ant around the S_1/S_0 MECI. The (a) g–h plane shows the characteristic double-cone splitting around the crossing, whereas the (b) g–z cross section is flatter along the seam-related direction.

$$E_{\pm}(\delta\mathbf{R}) \approx E_{\mathrm{MECI}} + \mathbf{s}\cdot\delta\mathbf{R} \pm \sqrt{(\mathbf{g}\cdot\delta\mathbf{R})^2 + (\mathbf{h}\cdot\delta\mathbf{R})^2} \quad (7)$$

where **s** is the average-gradient direction of the two states. Thus, the accessibility of the MECI channel depends not only on the relative crossing energy ΔE but also on whether the molecule can deform efficiently along the branching-plane modes that open the gap.

Figure 3 shows that, for 9-COCN-Ant, the **g** vector mainly reflects in-plane bond reorganization of the anthracene framework, while the **h** vector involves more pronounced out-of-plane deformation. As shown in Figure 4a, the potential energy surfaces of the S_0 and S_1 states form a double-cone shape in the branching plane, whereas the degeneracy is retained along the seam-related direction in Figure 4b. The topology in the vicinity of the S_1/S_0 MECI is correctly recovered with the SF-TDDFT method. The g–h branching plane identifies the few nuclear motions that most directly control the opening of the S_1/S_0 gap and the associated nonradiative channel, while $\Delta E(\mathrm{MECI} - \mathrm{FCS}_1)$ quantifies the energetic cost of reaching that crossing from the FC-relaxed state. Training the model on this well-defined molecular property—rather than on a phenomenological AIE/ACQ label—ensures that the learned relationship is grounded in the underlying photophysics, making it transferable across chemically diverse systems and suitable for first-pass screening.

Interpretable Atom-Level Attributions via CatBoost–SHAP

Among the descriptor baselines, CatBoost provides the best balanced train/test performance on the present split (R^2_{train} = 0.810 and R^2_{test} = 0.704; Table 1). Its combination with RDKit Morgan fingerprints and native TreeSHAP therefore provides the chemically interpretable companion to Uni-Mol. We use CatBoost–SHAP to examine atom- and group-resolved attributions to the predicted ΔE, while reserving Uni-Mol for higher-accuracy screening.

Specifically, each molecule is represented by the same 4096-dimensional binary radius-3 ECFP used in the descriptor benchmark. The fingerprints are generated with RDKit and include chirality information. The associated `bitInfo` map records every occurrence of an activated Morgan bit as a center-atom and radius pair, allowing feature-level SHAP attributions to be projected back to the corresponding center atoms and visualized as influential local "hotspot" regions.

To interpret the trained CatBoost ensemble, we use CatBoost's native TreeSHAP implementation to generate additive feature-level attributions. For each activated fingerprint feature, its TreeSHAP attribution is divided equally among the occurrences recorded in the RDKit `bitInfo`

https://doi.org/10.1021/acs.jctc.6c01107
J. Chem. Theory Comput. 2026, 22, 8825–8835

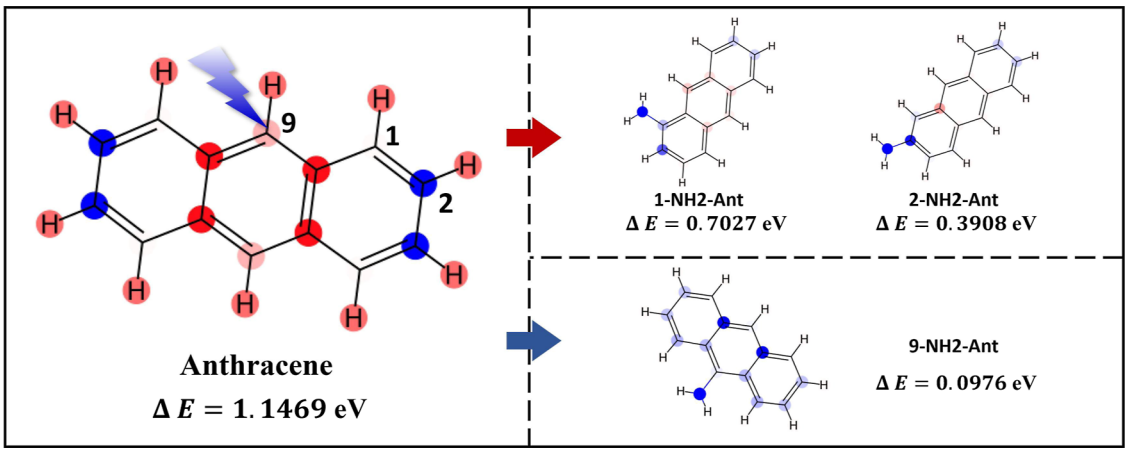


Figure 5. Atom-level CatBoost–SHAP attribution map for anthracene and representative amino-substituted derivatives. Red and blue overlays on the parent scaffold denote positive and negative local attributions to the predicted ΔE, respectively; the displayed derivative values are explicitly calculated $\Delta E(\mathrm{MECI} - \mathrm{FCS}_1)$ values. The examples illustrate substantial site-dependent tuning while emphasizing that a parent-scaffold attribution is not itself a finite-difference substitution effect.

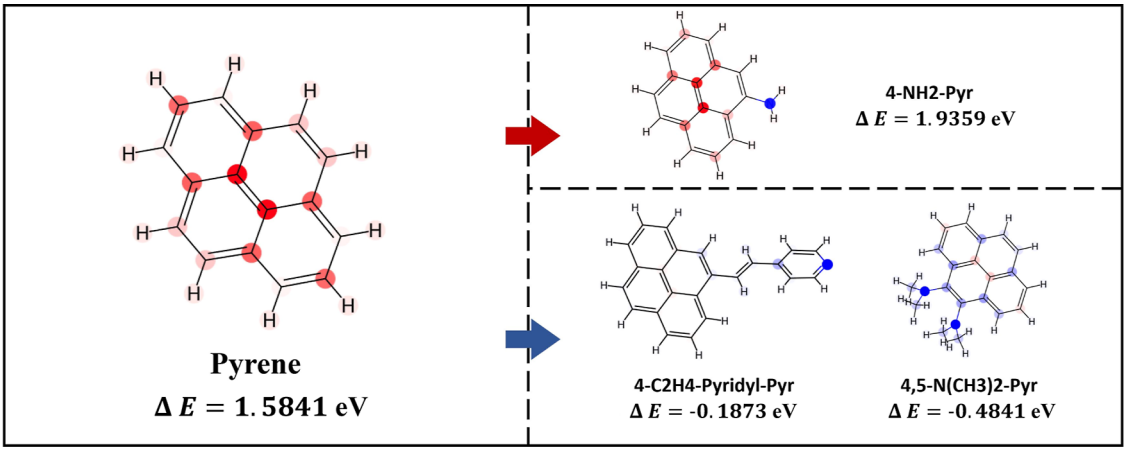


Figure 6. Atom-level CatBoost–SHAP attribution map for pyrene and representative substituted derivatives. Red and blue overlays on the parent scaffold denote positive and negative local attributions to the predicted ΔE, respectively; the displayed derivative values are explicitly calculated $\Delta E(\mathrm{MECI} - \mathrm{FCS}_1)$ values. The examples show that modifications of responsive regions can either raise or lower ΔE, depending on the combined site and substituent.

map, and each share is assigned to the center atom of the corresponding occurrence. Contributions from all occurrences centered on the same atom are then summed to obtain its atom-level attribution. This projection preserves the summed attribution of the mapped active features; the model baseline and contributions from absent or unmapped features are treated as global residual terms and are not assigned arbitrarily to atoms. The complete definition is provided in the Supporting Information. For the present target, a negative atom-level attribution indicates that the associated local fingerprint environments lower the predicted ΔE and thus increase the modeled electronic accessibility of the MECI, whereas a positive attribution raises the prediction. These maps convert high-dimensional fingerprint contributions into chemically recognizable local regions and support the prioritization of modification hypotheses. Because hashed Morgan environments can overlap, the resulting values are model attributions rather than unique intrinsic properties, quantum-mechanical atomic energies, or causal predictions of substitution effects. Proposed sites and substituents must therefore be verified by explicit ΔE calculations and by assessing whether aggregation can restrict the required MECI deformation.

For anthracene-derived scaffolds, the qualitative site-level pattern assigns a pronounced positive attribution to the C9 local environment and the symmetry-related C10 region, a negative attribution to C2, and a near-neutral attribution to C1. Under the present target definition, these signs indicate that the corresponding local environments tend to raise, lower, or minimally affect the predicted ΔE, respectively. The prominent attribution around C9/C10 highlights this central-ring region as a chemically relevant modification region for tuning MECI accessibility. Because C9/C10 lies within the dominant out-of-plane MECI deformation region, substitution at these positions can affect both the electronic accessibility of the crossing region and the structural deformation required to reach it. The experimental relevance of this site selection is supported by reports of AIE behavior in 9,10-dithienylanthracene derivatives and of pronounced substitution-dependent crystal packing and solid-state photophysical properties in 9/10-functionalized anthracene carboxylic acid derivatives.[63,64] These experimental precedents support C9/C10 as a photophysically responsive modification region, but they do not

https://doi.org/10.1021/acs.jctc.6c01107
J. Chem. Theory Comput. 2026, 22, 8825–8835

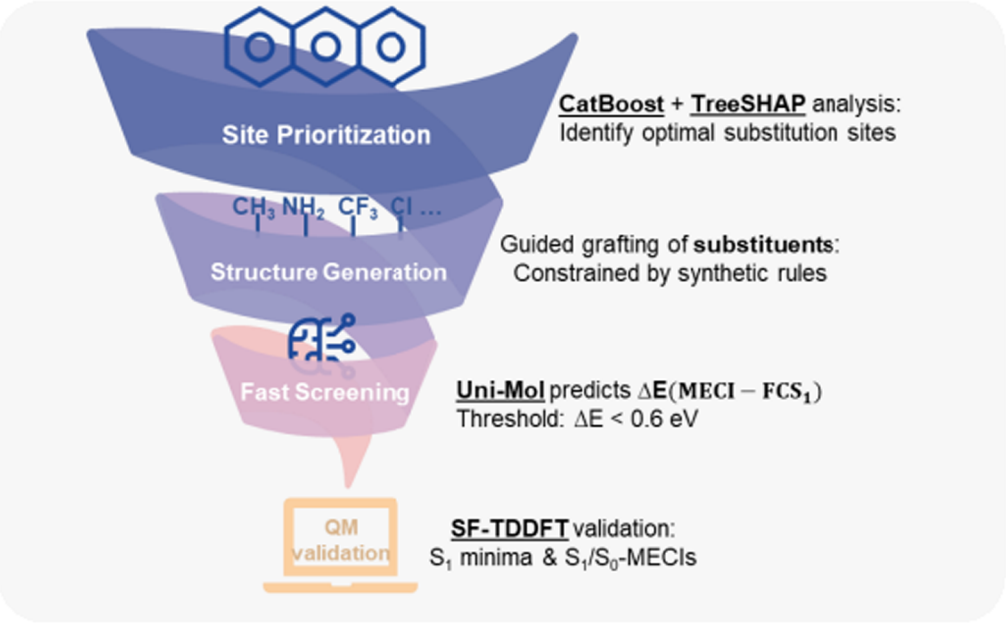


Figure 7. Mechanism-guided, AI-accelerated workflow for AIE design: interpretable site prioritization (CatBoost + SHAP) → guided library assembly → fast ΔE(MECI − FCS_1) screening (Uni-Mol) → CI-level quantum-chemical vetting → multicriterion lead selection.

imply that every substituent changes ΔE in the same direction or that the molecular attribution alone establishes AIE.

Figures 5 and 6 connect the signed parent-scaffold attribution maps with representative explicitly calculated substitution outcomes. For anthracene, amino substitution at C1, C2, or C9 lowers ΔE relative to the parent, with the largest decrease occurring for 9-NH_2-Ant. For pyrene, by contrast, the representative substitutions span both directions: 4-NH_2-Pyr raises ΔE, whereas the pyridyl- and dimethylamino-containing derivatives lower it. These examples demonstrate that the attribution maps identify chemically responsive local regions, while the direction and magnitude of a finite substitution effect depend jointly on the site and substituent and must be confirmed by explicit calculation.

Mechanism-Guided, AI-Accelerated Workflow for AIE Discovery

We combine the above components into an end-to-end, conical-intersection (CI)-centric workflow that turns mechanistic priors into testable design hypotheses and prioritized candidates (summarized in Figure 7). Given a set of core scaffolds and substituent lists, the pipeline proceeds as follows.

Step 1 **Site prioritization via interpretable baseline.** Use the CatBoost companion model on binary radius-3 ECFP features and perform native TreeSHAP attribution to obtain atom- and group-level attributions to the predicted ΔE of the target core. Rank substitution *sites* by signed SHAP aggregates, prioritizing local environments that consistently lower the prediction across related molecules.

Step 2 **Library generation by guided grafting.** Enumerate a focused combinatorial set by grafting selected substituents onto the prioritized sites (single substitutions, disubstitutions, and mixed pairs), using basic valence and connectivity checks to remove chemically invalid structures. Record provenance (`core`, `site`, `substituent`, `pattern`) for downstream attribution.

Step 3 **Fast screening of CI accessibility.** Predict ΔE(MECI − FCS_1) for the enumerated library using the fine-tuned 3D pretrained encoder (Uni-Mol, `unimol:v1`). Prioritize molecules with ΔE_{pred} < 0.6 eV for subsequent calculations. Because 0.6 eV is an empirical operating value inferred from the present external panel, molecules close to this value are treated as borderline rather than assigned by a hard universal rule.

Step 4 **Mechanistic vetting at the quantum level.** For the low-ΔE_{pred} subset, recompute the optimized S_0 structure, the vertical FCS_1 energy, and the S_1/S_0 MECI at the SF-TDDFT/ωB97X-D/def2-SVP level. Confirm the electronic-energy difference from eq 2 and quantify the geometry changes required to reach the crossing region, e.g.

$$\Delta\phi = \phi_{\mathrm{MECI}} - \phi_{S_0\text{-min}}, \qquad \Delta\theta = \theta_{\mathrm{MECI}} - \theta_{S_0\text{-min}}$$

and retain candidates that combine a low crossing energy with a substantial, aggregation-restrictable deformation. These candidates are then advanced to aggregate-state and experimental validation. In this workflow, the CatBoost−SHAP model proposes interpretable modification hypotheses and Uni-Mol accelerates ΔE screening; neither model replaces the quantum-chemical or experimental assessment of AIE.

CONCLUSION

In summary, this work uses ΔE(MECI − FCS_1), the electronic energy of the S_1/S_0 MECI relative to the vertically accessed FCS_1 state, as an efficient descriptor of conical-intersection accessibility. A Uni-Mol regressor predicts this quantity from relaxed ground-state geometries, while a CatBoost−SHAP companion model provides signed, local structure-prediction attributions.

For the present external panel, AIE luminogens have lower predicted values than ACQ luminogens (0.393 ± 0.134 versus 0.812 ± 0.137 eV), supporting an empirical first-pass threshold

https://doi.org/10.1021/acs.jctc.6c01107
J. Chem. Theory Comput. 2026, 22, 8825−8835

near 0.6 eV. This separation does not make ΔE a universal AIE classifier: low ΔE indicates an electronically accessible CI channel, whereas AIE additionally requires aggregation to restrict the deformation that reaches that channel.

The resulting four-step workflow combines interpretable site prioritization, guided library generation, rapid ΔE screening, and explicit quantum-chemical and experimental validation. It provides a scalable route for prioritizing AIE candidates, offering a practical route to accelerated discovery of AIE-active materials.

■ ASSOCIATED CONTENT

Data Availability Statement

The data set generated and the machine learning models developed during the current study are permanently archived in the Zenodo repository. They are publicly available at https://zenodo.org/records/17356940.

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jctc.6c01107.

> The Supporting Information provides detailed methodological descriptions, including: (1) the quantum-chemical protocol and definition of ΔE(MECI − FCS_1); (2) the generation and configuration of extended-connectivity fingerprints used in the descriptor models; (3) the complete Uni-Mol training protocol and hyperparameters; and (4) calculated and predicted ΔE values and optimized ground-state coordinates for the screened molecular subset (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

Qi Ou − *Sinopec Research Institute of Petroleum Processing Co., Ltd., Beijing 100083, P. R. China;* orcid.org/0000-0002-6400-7522; Email: ouqi.ripp@sinopec.com

Zhigang Shuai − *Department of Chemistry, MOE Key Laboratory of Organic OptoElectronics and Molecular Engineering, Tsinghua University, Beijing 100084, P. R. China; Guangdong Basic Research Center of Excellence for Aggregate Science, School of Science and Engineering, The Chinese University of Hong Kong, Shenzhen, Shenzhen, Guangdong 518172, P. R. China;* orcid.org/0000-0003-3867-2331; Email: shuaizhigang@cuhk.edu.cn

Authors

Wenjie Zhang − *Department of Chemistry, MOE Key Laboratory of Organic OptoElectronics and Molecular Engineering, Tsinghua University, Beijing 100084, P. R. China*

Ping-An Yin − *Department of Chemistry, MOE Key Laboratory of Organic OptoElectronics and Molecular Engineering, Tsinghua University, Beijing 100084, P. R. China*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jctc.6c01107

Author Contributions

‖W.Z. and P.-A.Y. contributed equally to this work. Z.S. and Q.O. conceived the study. W.Z. and P.-A.Y. performed the quantum-chemical calculations and model training. All authors contributed to writing the manuscript and approved its final version.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work is supported by the National Natural Science Foundation of China (grant Nos. T2350009, 22433007, and 22573124) and the Guangdong Provincial Natural Science Foundation grant No. 2024A1515011185, as well as Shenzhen city Peacock Team Project KQTD2024072910202801. Q.O. acknowledges financial supports from China Petroleum & Chemical Corp (funding numbers 124014 and E25014) and Beijing Municipal Science & Technology Commission (grant number Z251100007525011). Z.S. is supported by the Guangdong Basic Research Center of Excellence for Aggregate Science.

■ REFERENCES

(1) Luo, J.; Xie, Z.; Lam, J. W.; Cheng, L.; Tang, B. Z.; Chen, H.; Qiu, C.; Kwok, H. S.; Zhan, X.; Liu, Y.; et al. Aggregation-induced emission of 1-methyl-1, 2, 3, 4, 5-pentaphenylsilole. *Chem. Commun.* **2001**, *37*, 1740−1741.

(2) Mei, J.; Leung, N. L.; Kwok, R. T.; Lam, J. W.; Tang, B. Z. Aggregation-induced emission: together we shine, united we soar. *Chem. Rev.* **2015**, *115*, 11718−11940.

(3) Liu, Y.; Chen, X.; Liu, X.; Guan, W.; Lu, C. Aggregation-induced emission-active micelles: synthesis, characterization, and applications. *Chem. Soc. Rev.* **2023**, *52*, 1456−1490.

(4) Wang, H.; Li, Q.; Alam, P.; Bai, H.; Bhalla, V.; Bryce, M. R.; Cao, M.; Chen, C.; Chen, S.; Chen, X.; et al. Aggregation-induced emission (AIE), life and health. *ACS Nano* **2023**, *17*, 14347−14405.

(5) Wu, H.; Fang, Y.; Tian, L.; Liu, X.; Zhou, X.; Chen, X.; Gao, H.; Qin, H.; Liu, Y. AIE nanozyme-based long persistent chemiluminescence and fluorescence for POCT of pathogenic bacteria. *ACS Sens.* **2023**, *8*, 3205−3214.

(6) Duo, Y.; Luo, G.; Zhang, W.; Wang, R.; Xiao, G. G.; Li, Z.; Li, X.; Chen, M.; Yoon, J.; Tang, B. Z. Noncancerous disease-targeting AIEgens. *Chem. Soc. Rev.* **2023**, *52*, 1024−1067.

(7) Cai, Q.; Musiol, R.; Shubhra, Q. T. Advancing fluorescence imaging with dual-mode AIE nanoparticles. *Chem* **2024**, *10*, 429−432.

(8) Barman, D.; Barman, D.; Bhattacharyya, K.; Iyer, P. K. Engineering Single Component Luminogens to Multicomponent Charge-transfer Co-crystal Substrate as New Frontiers for Sensitive SERS Detection. *Adv. Opt. Mater.* **2024**, *12*, 2401352.

(9) Zeng, J.; Guo, J.; Liu, H.; Lam, J. W.; Zhao, Z.; Chen, S.; Tang, B. Z. Aggregation-Induced Delayed Fluorescence Luminogens for Efficient Organic Light-Emitting Diodes. *Chem.—Asian J.* **2019**, *14*, 828−835.

(10) Du, S.; Luo, M.; Li, D.; Lyu, L.; Li, W.; Zhao, M.; Wang, Z.; Zhang, J.; Liu, D.; Li, Y.; et al. Hot-Exciton Mechanism and AIE Effect Boost the Performance of Deep-Red Emitters in Non-Doped OLEDs. *Adv. Mater.* **2023**, *35*, 2303304.

(11) Han, P.; Lin, C.; Xia, E.; Cheng, J.; Xia, Q.; Yang, D.; Qin, A.; Ma, D.; Tang, B. Z. Non-Doped Blue AIEgen-Based OLED with EQE Approaching 10.3%. *Angew. Chem.* **2023**, *135*, No. e202310388.

(12) Xia, Y.-D.; Cheng, Y.; Wu, Y.-J.; Xia, Y.; Yin, X.-B. Driving fluorescence by forming AIEgens in a hollow azine-linked covalent organic framework. *Chem. Mater.* **2023**, *35*, 2579−2587.

(13) Petdee, S.; Chantanop, N.; Arunlimsawat, S.; Saenubol, A.; Nalaoh, P.; Sudyoadsuk, T.; Promarak, V. Excited-state intramolecular proton-transfer solid-state fluorophores with aggregation-induced emission as efficient emitters for electroluminescent devices. *CrystEngComm* **2024**, *26*, 4698−4707.

(14) Feng, Z.; Li, Y.; Chen, S.; Li, J.; Wu, T.; Ying, Y.; Zheng, J.; Zhang, Y.; Zhang, J.; Fan, X.; et al. Engineered NIR-II fluorophores

with ultralong-distance molecular packing for high-contrast deep lesion identification. *Nat. Commun.* **2023**, *14*, 5017.
(15) Mohammed Hashim, K. K.; Manoj, E. Aminoguanidine-based bioactive proligand as AIEE probe for anticancer and anticovid studies. *RSC Adv.* **2024**, *14*, 13654−13668.
(16) Li, C.; Du, J.; Jiang, G.; Gong, J.; Zhang, Y.; Yao, M.; Wang, J.; Wu, L.; Tang, B. Z. White-light activatable organic NIR-II luminescence nanomaterials for imaging-guided surgery. *Nat. Commun.* **2024**, *15*, 5832.
(17) Song, S.; Zhao, Y.; Kang, M.; Zhang, F.; Wu, Q.; Niu, N.; Yang, H.; Wen, H.; Fu, S.; Li, X.; et al. An NIR-II excitable AIE small molecule with multimodal phototheranostic features for orthotopic breast cancer treatment. *Adv. Mater.* **2024**, *36*, 2309748.
(18) Asthana, S.; Mouli, M. V.; Tamrakar, A.; Wani, M. A.; Mishra, A. K.; Pandey, R.; Pandey, M. D. Recent advances in AIEgen-based chemosensors for small molecule detection, with a focus on ion sensing. *Anal. Methods* **2024**, *16*, 4431−4484.
(19) Huang, Y.; Li, Y.; Li, Y.; Zhong, K.; Tang, L. An "AIE+ ESIPT" mechanism-based benzothiazole-derived fluorescent probe for the detection of Hg 2+ and its applications. *New J. Chem.* **2023**, *47*, 6916−6923.
(20) Pei, S.; Li, H.; Chen, L.; Nie, G.; Wang, H.; Liu, C.; Zhang, C. Dual-functional AIE fluorescent probe for visualization of lipid droplets and photodynamic therapy of cancer. *Anal. Chem.* **2024**, *96*, 5615−5624.
(21) Karn-Orachai, K.; Wattanasin, P.; Ngamaroonchote, A. Colorimetric sensor for cr (VI) Ion Detection in tap Water using a combination of AuNPs and AgNPs. *ACS Omega* **2024**, *9*, 26472−26483.
(22) Yi, L.; Zhang, Y.; Xie, Y.; Zhang, X.; Jiao, Z.; Jiang, G.; Chen, H.; Zhang, P. Portable AIE hydrogel sensor for rapid and visual field detection of heavy metal residue in food. *Dyes Pigm.* **2023**, *212*, 111117.
(23) Cao, H.; Zhang, X.; Tang, M.; Dong, W.; Shi, W. Amplified fluorescence sensing of Cr (VI) enabled by AIE-active copper nanoclusters functionalized hydrogels to afford a smartphone-enabled colorimetric platform. *Sens. Actuators, B* **2023**, *392*, 134066.
(24) Bailey, A. D.; Deshmukh, A. P.; Bradbury, N. C.; Pengshung, M.; Atallah, T. L.; Williams, J. A.; Barotov, U.; Neuhauser, D.; Sletten, E. M.; Caram, J. R. Exploring the design of superradiant J-aggregates from amphiphilic monomer units. *Nanoscale* **2023**, *15*, 3841−3849.
(25) Yang, L.-L.; Wang, H.; Zhang, J.; Wu, B.; Li, Q.; Chen, J.-Y.; Tang, A.-L.; Lam, J. W.; Zhao, Z.; Yang, S.; et al. Understanding the AIE phenomenon of nonconjugated rhodamine derivatives via aggregation-induced molecular conformation change. *Nat. Commun.* **2024**, *15*, 999.
(26) Yang, S.-Y.; Chen, Y.; Kwok, R. T.; Lam, J. W.; Tang, B. Z. Platinum complexes with aggregation-induced emission. *Chem. Soc. Rev.* **2024**, *53*, 5366−5393.
(27) Wang, B.; Liu, Y.; Chen, X.; Liu, X.-T.; Liu, Z.; Lu, C. Aggregation-induced emission-active supramolecular polymers: from controlled preparation to applications. *Chem. Soc. Rev.* **2024**, *53*, 10189−10215.
(28) Jindal, S.; Wang, J.-X.; Wang, Y.; Thomas, S.; Mallick, A.; Bonneau, M.; Bhatt, P. M.; Alkhazragi, O.; Nadinov, I.; Ng, T. K.; et al. Aggregation induced emission-based covalent organic frameworks for high-performance optical wireless communication. *J. Am. Chem. Soc.* **2024**, *146*, 25536−25543.
(29) Hecht, M.; Wurthner, F. Supramolecularly engineered J-aggregates based on perylene bisimide dyes. *Acc. Chem. Res.* **2021**, *54*, 642−653.
(30) Wang, L.; Salguero, C.; Lopez, S. A.; Li, J. Machine learning photodynamics uncover blocked non-radiative mechanisms in aggregation-induced emission. *Chem* **2024**, *10*, 2295−2310.
(31) Xie, H.; Wang, J.; Lou, Z.; Hu, L.; Segawa, S.; Kang, X.; Wu, W.; Luo, Z.; Kwok, R. T.; Lam, J. W.; et al. Mechanochemical fabrication of full-color luminescent materials from aggregation-induced emission prefluorophores for information storage and encryption. *J. Am. Chem. Soc.* **2024**, *146*, 18350−18359.
(32) Zhu, L.; Song, G.; Zhang, W.; Wu, Y.; Chen, Y.; Song, J.; Wang, D.; Li, G.; Tang, B. Z.; Li, Y. Aggregation induced emission luminogen bacteria hybrid bionic robot for multimodal phototheranostics and immunotherapy. *Nat. Commun.* **2025**, *16*, 2578.
(33) Tu, Y.; Zhao, Z.; Lam, J. W.; Tang, B. Z. Mechanistic connotations of restriction of intramolecular motions (RIM). *Natl. Sci. Rev.* **2021**, *8*, nwaa260.
(34) Hestand, N. J.; Spano, F. C. Expanded theory of H-and J-molecular aggregates: the effects of vibronic coupling and intermolecular charge transfer. *Chem. Rev.* **2018**, *118*, 7069−7163.
(35) Shao, Y.; Head-Gordon, M.; Krylov, A. I. The spin−flip approach within time-dependent density functional theory: Theory and applications to diradicals. *J. Chem. Phys.* **2003**, *118*, 4807−4818.
(36) Bernard, Y. A.; Shao, Y.; Krylov, A. I. General formulation of spin-flip time-dependent density functional theory using non-collinear kernels: Theory, implementation, and benchmarks. *J. Chem. Phys.* **2012**, *136*, 204103.
(37) Shuai, Z.; Peng, Q. Organic light-emitting diodes: theoretical understanding of highly efficient materials and development of computational methodology. *Natl. Sci. Rev.* **2017**, *4*, 224−239.
(38) Domcke, W.; Yarkony, D. R. Role of conical intersections in molecular spectroscopy and photoinduced chemical dynamics. *Annu. Rev. Phys. Chem.* **2012**, *63*, 325−352.
(39) Ou, Q.; Peng, Q.; Shuai, Z. Toward quantitative prediction of fluorescence quantum efficiency by combining direct vibrational conversion and surface crossing: BODIPYs as an example. *J. Phys. Chem. Lett.* **2020**, *11*, 7790−7797.
(40) Yin, P.-A.; Ou, Q.; Shuai, Z. Computational Design Strategy for Aggregation-Induced Emission Luminogens: Modulating the S1/S0 minimum Energy Conical Intersection of Anthracene Derivatives through Substituent Effects. *J. Chem. Theory Comput.* **2025**, *21*, 4992−5002.
(41) Wu, Z.; Ramsundar, B.; Feinberg, E. N.; Gomes, J.; Geniesse, C.; Pappu, A. S.; Leswing, K.; Pande, V. MoleculeNet: a benchmark for molecular machine learning. *Chem. Sci.* **2018**, *9*, 513−530.
(42) Merchant, A.; Batzner, S.; Schoenholz, S. S.; Aykol, M.; Cheon, G.; Cubuk, E. D. Scaling deep learning for materials discovery. *Nature* **2023**, *624*, 80−85.
(43) Rogers, D.; Hahn, M. Extended-connectivity fingerprints. *J. Chem. Inf. Model.* **2010**, *50*, 742−754.
(44) Chen, T.; Guestrin, C. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd Acm Sigkdd International Conference on Knowledge Discovery and Data Mining*, 2016; pp 785−794.
(45) Prokhorenkova, L.; Gusev, G.; Vorobev, A.; Dorogush, A. V.; Gulin, A. CatBoost: unbiased boosting with categorical features. *Adv. Neural Inf. Process. Syst.* **2018**, *31*, 6639−6649.
(46) Lundberg, S. M.; Lee, S.-I. A unified approach to interpreting model predictions. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 4768−4777.
(47) Zhou, G.; Gao, Z.; Ding, Q.; Zheng, H.; Xu, H.; Wei, Z.; Zhang, L.; Ke, G. Uni-Mol: A Universal 3D Molecular Representation Learning Framework. In *International Conference on Learning Representations (ICLR)*, 2023.
(48) Morehead, A.; Cheng, J. Geometry-complete diffusion for 3D molecule generation and optimization. *Commun. Chem.* **2024**, *7*, 150.
(49) Gallegos, M.; Vassilev-Galindo, V.; Poltavsky, I.; Martín Pendás, Á.; Tkatchenko, A. Explainable chemical artificial intelligence from accurate machine learning of real-space chemical descriptors. *Nat. Commun.* **2024**, *15*, 4345.
(50) Peng, J.; He, X.; Li, Y.; Guan, J.; Wu, B.; Li, X.; Yu, Z.; Liu, J.; Zheng, J. Restriction of crossing conical intersections: the intrinsic mechanism of aggregation-induced emission. *Phys. Chem. Chem. Phys.* **2023**, *25*, 12342−12351.
(51) Refaely-Abramson, S.; Baer, R.; Kronik, L. Fundamental and excitation gaps in molecules of relevance for organic photovoltaics from an optimally tuned range-separated hybrid functional. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2011**, *84*, 075144.

(52) Karolewski, A.; Kronik, L.; Kümmel, S. Using optimally tuned range separated hybrid functionals in ground-state calculations: Consequences and caveats. *J. Chem. Phys.* **2013**, *138*, 204115.
(53) Chai, J.-D.; Head-Gordon, M. Long-range corrected hybrid density functionals with damped atom−atom dispersion corrections. *Phys. Chem. Chem. Phys.* **2008**, *10*, 6615−6620.
(54) Shao, Y.; Gan, Z.; Epifanovsky, E.; Gilbert, A. T.; Wormit, M.; Kussmann, J.; Lange, A. W.; Behn, A.; Deng, J.; Feng, X.; et al. Advances in molecular quantum chemistry contained in the Q-Chem 4 program package. *Mol. Phys.* **2015**, *113*, 184−215.
(55) Levine, B. G.; Coe, J. D.; Martínez, T. J. Optimizing conical intersections without derivative coupling vectors: Application to multistate multireference second-order perturbation theory (MS-CASPT2). *J. Phys. Chem. B* **2008**, *112*, 405−413.
(56) Fuchs, F. B.; Worrall, D.; Fischer, V.; Welling, M. Se (3)-transformers: 3d roto-translation equivariant attention networks. *Adv. Neural Inf. Process. Syst.* **2020**, *33*, 1970−1981.
(57) Thomas, N.; Smidt, T.; Kearnes, S.; Yang, L.; Li, L.; Kohlhoff, K.; Riley, P. Tensor Field Networks: Rotation- and Translation-Equivariant Neural Networks for 3D Point Clouds. *arXiv* **2018**, arXiv:1802.08219 preprint. . Early E(3)-equivariant neural networks for 3D data
(58) Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5−32.
(59) Ke, G.; Meng, Q.; Finley, T.; Wang, T.; Chen, W.; Ma, W.; Ye, Q.; Liu, T.-Y. Lightgbm: A highly efficient gradient boosting decision tree. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 3149−3157.
(60) Drucker, H.; Burges, C. J.; Kaufman, L.; Smola, A.; Vapnik, V. Support vector regression machines. *Adv. Neural Inf. Process. Syst.* **1996**, *9*, 155−161.
(61) Smola, A. J.; Schölkopf, B. A tutorial on support vector regression. *Stat. Comput.* **2004**, *14*, 199−222.
(62) Ou, Q.; Subotnik, J. E. Electronic relaxation in benzaldehyde evaluated via TD-DFT and localized diabatization: Intersystem crossings, conical intersections, and phosphorescence. *J. Phys. Chem. C* **2013**, *117*, 19839−19849.
(63) Wang, R.; Diao, L.; Zhang, J.; Chen, Z.; Pu, S. Aggregation-induced emission compounds based on 9, 10-dithienylanthracene and their applications in cell imaging. *Dyes Pigm.* **2020**, *175*, 108112.
(64) Zhu, L.; Al-Kaysi, R. O.; Dillon, R. J.; Tham, F. S.; Bardeen, C. J. Crystal structures and photophysical properties of 9-anthracene carboxylic acid derivatives for photomechanical applications. *Cryst. Growth Des.* **2011**, *11*, 4975−4983.

https://doi.org/10.1021/acs.jctc.6c01107
J. Chem. Theory Comput. 2026, 22, 8825−8835