

Assessing the Synthetic Compatibility of GDB-9 Molecules with PCBM and PCDTBT for Organic Solar Cells: A Computational Approach

Your Name

April 24, 2025

Abstract

Organic photovoltaics (OPVs) offer a promising route to cost-effective and flexible solar energy harvesting, yet their power conversion efficiency (PCE) and stability lag behind traditional silicon-based solar cells [1,2]. To accelerate the discovery of novel OPV materials, we leverage computational methods, focusing on AI-assisted screening of the GDB-9 database. This work aims to identify potential donor and acceptor molecules compatible with two benchmark materials: [6,6]-Phenyl-C₆₁-butyric acid methyl ester (PCBM) and Poly[N-9'-heptadecanyl-2,7-carbazole-alt-5,5-(4',7'-di-2-thienyl-2',1',3'-benzothiadiazole)] (PCDTBT). Our multi-step computational protocol integrates geometry optimizations using **xTB**, conformer exploration with **CREST**, and PCE estimation via the **Scharber model**. Synthetic accessibility scores (SAS) are incorporated to evaluate the feasibility of molecule production. We highlight a subset of GDB-9 molecules showing promising electronic properties and PCE values, though synthetic constraints present ongoing challenges. This study underscores the potential of AI-driven computational screening for advancing OPV research, while emphasizing the need for experimental validation and machine learning to guide robust molecule generation. Aligned with the United Nations' Sustainable Development Goals (SDGs) 7 and 13, this work contributes to clean energy and climate action [5].

1 Introduction

Harnessing solar energy through organic photovoltaics (OPVs) represents a promising alternative to conventional photovoltaic technologies. These devices offer significant advantages, including lower manufacturing costs, mechanical flexibility, and compatibility with large-area roll-to-roll processing [1]. Despite these benefits, the development of OPVs faces critical challenges related to power conversion efficiency (PCE) and operational stability. Achieving competitive PCEs while ensuring long-term durability remains a major research objective [2]. Traditional materials discovery in chemistry often relies on labor-intensive trial-and-error processes, complex molecular modeling, and exhaustive exploration of chemical spaces. The emergence of computational chemistry and artificial intelligence (AI) has revolutionized this field by enabling high-throughput screening and inverse molecular design, significantly accelerating the search for novel OPV materials.

Among the various molecular databases available for virtual screening, GDB-9 provides an extensive library of small organic molecules, offering a valuable resource for the discovery of new photoactive compounds. Recent progress in AI-assisted chemistry has demonstrated the potential of automated molecular design in identifying candidate materials with optimized optoelectronic properties. This study focuses on evaluating the suitability of GDB-9 molecules as potential donors or acceptors in organic solar cells. Two well-established materials serve as reference points for this analysis. The first is [6,6]-Phenyl-C₆₁-butyric acid methyl ester (PCBM), a widely used acceptor in bulk heterojunction (BHJ) architectures. The second is poly[N-9'-heptadecanyl-2,7-carbazole-alt-5,5-(4',7'-di-2-thienyl-2',1',3'-benzothiadiazole)] (PCDTBT), a benchmark donor in high-efficiency OPV systems [3].

To assess the compatibility of newly identified molecules with these reference materials, a computational workflow is established, integrating multiple evaluation criteria. Quantum-based geometry optimization

techniques, such as xTB and CREST, ensure the stability of molecular structures, providing reliable conformational analysis. The electronic properties of candidate molecules, including HOMO/LUMO energy levels and band gaps, are computed to evaluate their impact on exciton generation and charge transfer processes. Theoretical power conversion efficiencies are estimated using the Scharber model, a widely accepted approach for predicting OPV performance. Additionally, synthetic accessibility scores are determined to assess the feasibility of large-scale synthesis, ensuring that the proposed materials can be realistically integrated into experimental validation and industrial applications.

This computational framework aligns with the broader objective of developing high-performance OPV materials while ensuring practical synthetic pathways. By leveraging AI-driven screening and theoretical modeling, this study contributes to the advancement of sustainable energy solutions. The findings presented herein support global efforts toward clean energy innovation, particularly in alignment with Sustainable Development Goals 7 (Affordable and Clean Energy) and 13 (Climate Action) [5]. The following sections describe the computational methodology in detail, present the results of the molecular screening process, and discuss the implications of these findings for the future of organic photovoltaics.

2 Materials and Methods

2.1 Dataset and Molecular Properties

This study leverages the *GDB-9* dataset, a subset of the larger *GDB-17* database, renowned for its comprehensive collection of small organic molecules [16, 19]. The *GDB-9* dataset encompasses approximately 134,000 stable organic molecules, each composed of carbon, hydrogen, oxygen, nitrogen, and fluorine atoms. These molecules represent a valuable resource for virtual screening and property prediction in organic electronics. As highlighted in recent reviews of chemical datasets for machine learning, such curated collections are essential for training and validating models that can accelerate the discovery of novel materials [9, 10].

The *GDB-9* dataset provides a wealth of information, including geometric, energetic, and electronic properties crucial for understanding the behavior of organic photovoltaic (OPV) materials. Key electronic properties include the energies of the highest occupied molecular orbital (HOMO) and lowest unoccupied molecular orbital (LUMO), as well as the corresponding HOMO-LUMO energy gap. These properties, calculated using Density Functional Theory (DFT) with the B3LYP functional and the 6-31G(2df,p) basis set, are essential for predicting charge transfer characteristics and overall device performance in OPVs [6]. Furthermore, the dataset also reports atomization energies, enthalpies, and Gibbs free energies calculated at the higher-level G4MP2 theoretical level, providing a more precise description of the thermodynamic stability of the molecules. The availability of both DFT and G4MP2 calculated properties allows for a multi-faceted analysis of structure-property relationships. As emphasized in the literature, QM datasets like GDB-9 are essential for developing and benchmarking machine learning models for predicting quantum chemical properties, ultimately reducing the computational cost associated with DFT calculations [6, 11].

2.2 Molecular Selection and Property Considerations

To develop and validate the model, a subset of molecules from the *Harvard Clean Energy Project Database (CEPDB)* (available at [CEPDB Molecular Space](#)) was selected, focusing on optoelectronic characteristics crucial for organic photovoltaic (OPV) applications. The initial CEPDB dataset comprised 133,885 molecules, characterized by their HOMO, LUMO, and energy gap values. A multi-step filtering process was then employed to identify molecules meeting specific criteria relevant to OPV performance, guided by the Scharber model for predicting photovoltaic efficiency.

The filtering process involved a series of sequential steps, each targeting a specific energy level range to ensure compatibility with efficient charge transfer and exciton dissociation in OPV devices.

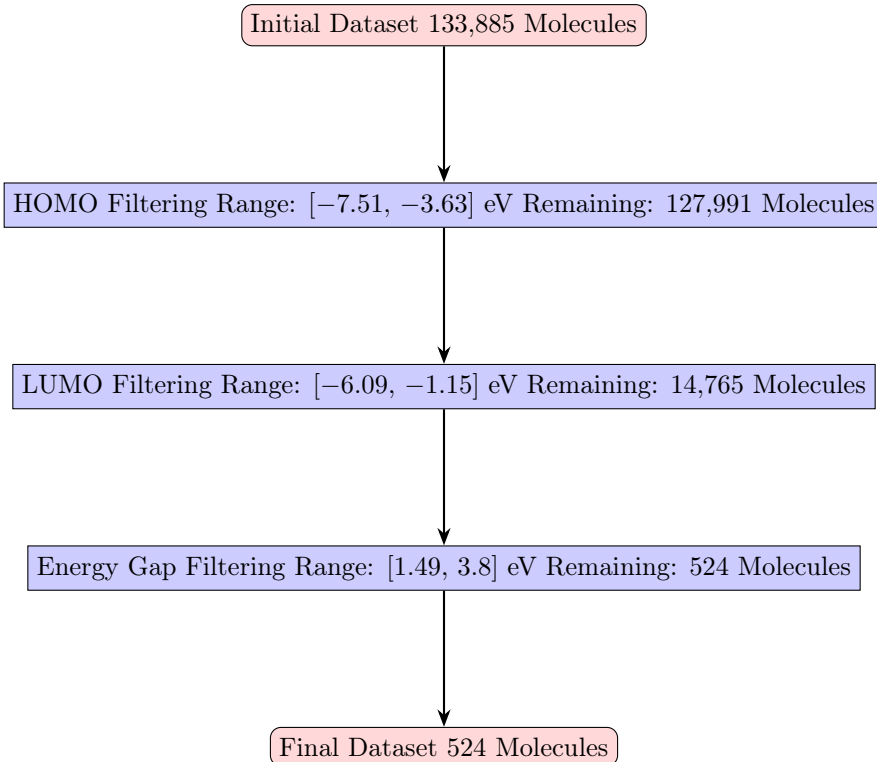


Figure 1: Stepwise filtering process of the initial molecular dataset using HOMO, LUMO, and energy gap thresholds. The final subset comprises 524 molecules suitable for OPV applications.

2.3 Study Workflow

As depicted in Figure 2, the study framework follows a structured sequence of steps, ensuring a systematic and reliable approach to material discovery. The workflow is composed of several key stages, beginning with the molecular structure proposal and concluding with the identification of promising materials. The first stage involves the proposal of a molecular structure, which is represented as a **SMILES** string. This structure is then converted into **Cartesian coordinates** using **RDKit** [32], ensuring a geometrically meaningful representation suitable for further computational studies. Following the initial coordinate generation, a **conformer search** is performed using **CREST** [17], an advanced sampling tool that explores the low-energy conformational space of molecules. The lowest-energy conformer is then subjected to **geometry optimization** using **xTB** [33], refining the molecular structure while maintaining computational efficiency. With the optimized geometry in hand, **electronic properties** such as **HOMO**, **LUMO**, and the **energy gap** are computed using **xTB** [33]. These properties provide crucial insights into the electronic behavior of the material and its potential application in organic photovoltaics (OPVs). A key step in the workflow is the estimation of the **power conversion efficiency (PCE)**, which is achieved using the **Scharber model** [3]. This model provides a predictive framework for evaluating photovoltaic performance based on electronic structure calculations. To assess the practical feasibility of the proposed materials, a synthesis evaluation is conducted using **SAScores** [34], which estimate the synthetic accessibility of the molecules. To ensure accuracy and validate the results, **DFT calculations** are performed using **PySCF** [35], providing a more rigorous quantum mechanical description of the electronic properties. Finally, an analysis and comparison of results are conducted, integrating insights from different computational levels. The workflow culminates in the identification of promising materials, balancing computational efficiency, electronic performance, and synthetic feasibility. This structured approach facilitates the discovery of novel materials with potential applications in organic electronics, optimizing both theoretical accuracy and computational resources.

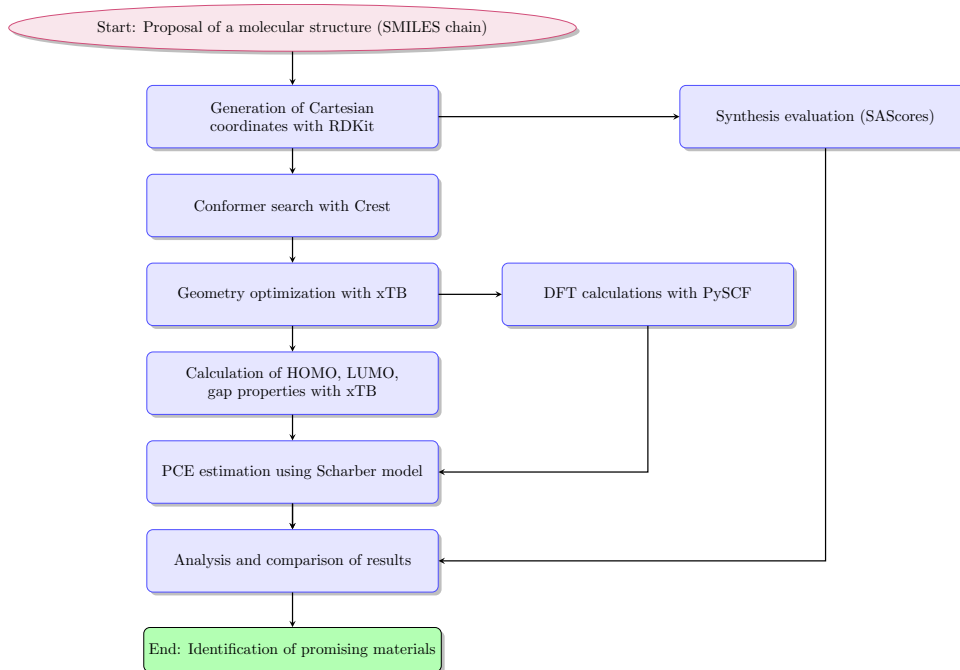


Figure 2: Overview of the computational workflow.

2.4 Computational Tools

2.4.1 Semiempirical quantum chemistry calculations

Extended tight-binding theories: This section introduces the **GFN** methods implemented in the *xTB* software version **7.0.1**. These are semi-empirical quantum chemistry approaches based on Tight-Binding (TB) models, optimized for large organic and inorganic molecules while minimizing computational costs. The **GFN** methods (*Geometries, Frequencies, Non-covalent interactions*) form the foundation of *xTB*, with several tailored variants, including [4]:

- **GFN0-xTB:** A simplified model for rapid geometric optimizations, incorporating approximations for non-covalent interactions. It is efficient when such interactions are negligible.
- **GFN1-xTB:** Beyond geometric optimizations, this model calculates **vibrational frequencies** and enhances the treatment of non-covalent interactions, crucial for complex molecular systems.
- **GFN2-xTB:** An advanced version refining the modeling of **non-covalent interactions** with van der Waals dispersion corrections. It provides more accurate calculations of **free energies**, enthalpies, and solvation properties.

Key differences between **GFN1-xTB** and **GFN2-xTB** lie in the sophistication of corrections for non-covalent interactions and solvation effects. **GFN2-xTB** incorporates additional terms to model these interactions, enabling more precise estimates of thermodynamic properties such as free energies and atomization energies.

Conformer Search and Optimization: The *CREST* software **3.1.0** (*Conformer-Rotamer Ensemble Sampling Tool*) is essential in this study for **conformer search**. Combined with **GFN-xTB** methods, it systematically explores the **conformational space** of a molecule, considering bond rotations and non-covalent interactions. *CREST* is also used to optimize structures by providing thermodynamically accessible ensembles [7]. Its algorithm iteratively explores **potential energy surfaces** to identify stable conformers that minimize free energy under specific environmental conditions (solvent, temperature).

The generated conformers are optimized based on **free energies**, accounting for **solvation effects** and **vibrational frequencies**.

2.5 Density Functional Theory (DFT)

2.5.1 DFT Method

The *Density Functional Theory* (**DFT**) is a quantum calculation method that uses the electronic density as the primary variable instead of the wave function. It provides an approximate solution to Schrödinger’s equation for multi-electron systems.

DFT is based on two fundamental theorems established by Hohenberg and Kohn [29]. The first theorem states that the electronic density of a system uniquely determines its fundamental properties. The second theorem provides a method to calculate the system’s total energy as a function of the electronic density, through an exchange-correlation energy functional.

The most commonly used functionals in **DFT** are the Local Density Approximation (LDA) and the Generalized Gradient Approximation (GGA) [30]. However, in this study, we have exclusively employed the ****hybrid functional B3LYP**** [31], which combines exact Hartree-Fock exchange contributions with semi-local correlation corrections.

2.5.2 DFT for OPV Studies

For this study, we utilized the **PySCF** package to perform DFT calculations on organic photovoltaic (OPV) materials. We specifically selected:

- **A single functional:** *B3LYP*, which is well known for its accuracy in describing energy levels and electronic structures.
- **A single basis set:** *6-31G(2df,p)*, which includes polarization and diffuse functions to enhance the description of electronic orbitals.

Using DFT with PySCF and the B3LYP functional allowed us to characterize the energy levels, geometric structures, and electronic properties of OPV materials. This approach provides reliable predictions regarding thermodynamic stability and absorption properties, which are essential for optimizing organic photovoltaic cells.

2.5.3 The Scharber Model

The Scharber model, developed by Markus C. Scharber et al. in 2006 [3], is a widely used semi-empirical model for estimating the conversion efficiency of organic photovoltaic (OPV) devices. By combining experimental and theoretical parameters, it predicts photovoltaic device performance. This model has significantly advanced the development of high-performance OPV materials by evaluating the *PCE*, a critical indicator of solar cell viability. The *PCE* depends on several electronic parameters detailed below.

Short-Circuit Current Density (J_{sc}): The short-circuit current density (J_{sc}) is the maximum current generated by the cell when short-circuited. It is influenced by the thickness of the active layer and the number of photons absorbed. J_{sc} is directly related to the material’s ability to generate electron-hole pairs, collect charge carriers, and minimize resistive losses.

$$J_{sc} = Ae^{-E_{GAP}^2/B} \tag{1}$$

- $A = 433.12$ and $B = 2.34$ are fitting parameters derived from the Tartarus model results [27].
- E_{GAP} is the donor’s bandgap energy, defined as the difference between molecular *HOMO* and *LUMO* levels, with maximum values up to 3.8 eV.

Open-Circuit Voltage (V_{oc}): The open-circuit voltage (V_{oc}) is the maximum voltage produced by the cell when no current flows. It depends on the energy difference between the donor’s *HOMO* and the acceptor’s *LUMO* [28].

$$V_{oc} = \frac{1}{e} (|E^{D^o}HOMO| - |E^{AC}LUMO|) - 0.3 \quad (2)$$

- e : Electron charge.
- $E^{D^o}HOMO$: Donor *HOMO* energy, typically between -5.7 and -4.5 , eV [18].
- $E^{AC}LUMO$: Acceptor *LUMO* energy, estimated between -4.0 and -3.0 , eV [17].
- -0.3 , V : Exciton separation threshold voltage.

Fill Factor (FF): The Fill Factor (FF) measures the efficiency of a solar cell in converting light energy into usable electrical energy. It is calculated as the ratio of the maximum power output to the incident power. Based on the Tartarus model, the FF is estimated at 65% [27].

Power Conversion Efficiency (PCE): The Power Conversion Efficiency (PCE) represents the ratio of the electrical power generated by the cell to the incident light power. Current OPVs achieve a *PCE* of 12%, with a target efficiency of 20% [23].

$$PCE = 100 \times \frac{V_{oc} \cdot FF \cdot J_{sc}}{P_{in}} \quad (3)$$

Where $P_{in} = 900.14 \text{ W/m}^2$ is the incident light power.

2.5.4 Synthetic Accessibility Score and Synthesis Compatibility

Definition of the Synthetic Accessibility Score (SA Score): The Synthetic Accessibility (SA) score is a metric designed to evaluate the *ease of synthesis* of an organic molecule on a scale from **1** (easy) to **10** (difficult). Developed to guide material design in various fields, including organic photovoltaics (OPVs), the SA score incorporates the following molecular properties:

- **Molecular size:** Larger molecules are generally more challenging to synthesize.
- **Structural complexity:** Molecules with higher structural complexity require more intricate synthetic pathways.
- **Functional groups:** Specific groups can either facilitate or complicate synthesis depending on their reactivity.

This metric promotes the selection of molecules that achieve an optimal balance between *high electronic performance* and *practical feasibility*, aligning with the guidelines set out by Scharber et al. (2006) and Ruddigkeit et al. (2015).

Definition of Synthesis Compatibility: Synthesis compatibility identifies molecules that are both compatible and synthesizable, aiming to optimize OPV performance. The compatibility is evaluated using the metric PCE_{SAS} , which integrates the photovoltaic efficiency (PCE) with the Synthetic Accessibility score (SAs):

$$PCE_{SAS} = PCE - SAs$$

Where:

- PCE : Photovoltaic conversion efficiency, representing the device’s performance.
- SAs : Synthetic Accessibility score, quantifying the difficulty of synthesis.

Molecules are considered compatible if they maximize the PCE_{SAS} value, achieving a balance between a high PCE and a low SA score.

Relevance for OPV Design This dual-metric approach enables:

1. Selection of high-performance molecules optimized for synthesis feasibility.
2. Prioritization of materials that are scalable and cost-effective for production.

3 Results

3.1 Database analysis

3.1.1 Estimating the number of clusters

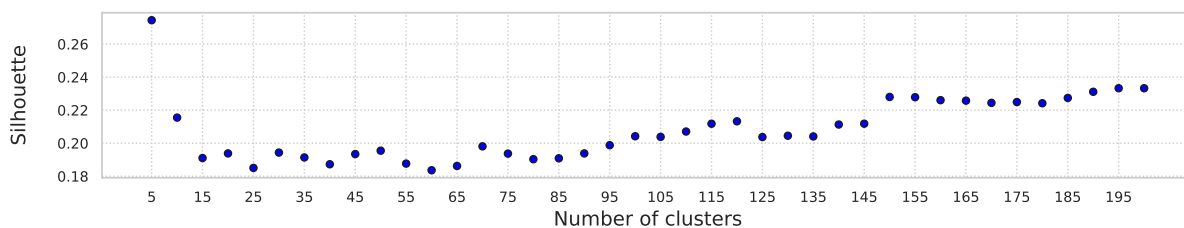


Figure 3: Silhouette scores under different number of cluster sizes using the total feature matrix using the k -means algorithm..

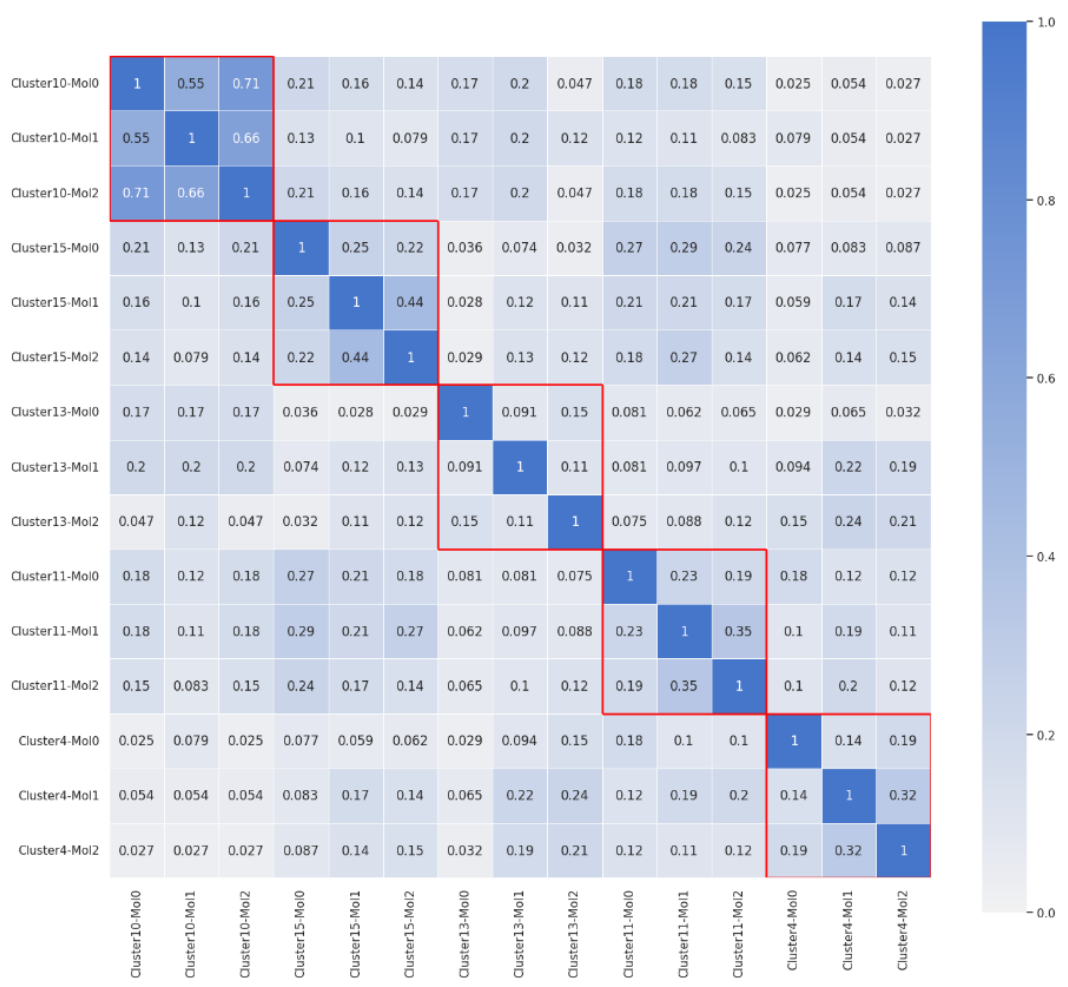
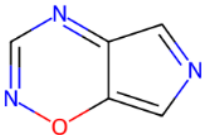
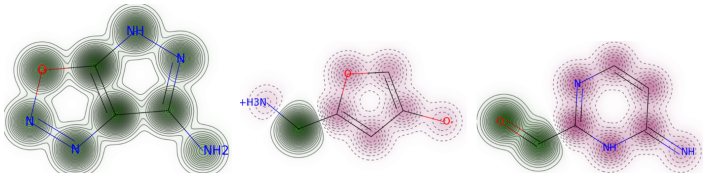
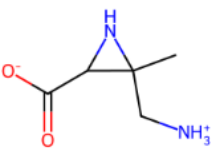
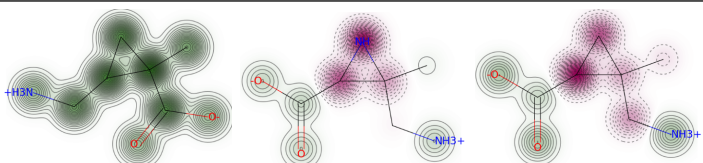
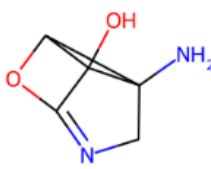
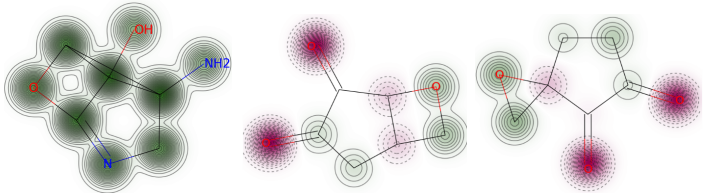
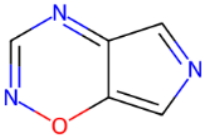
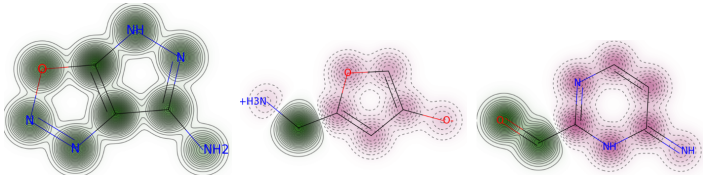
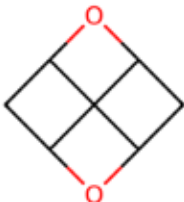
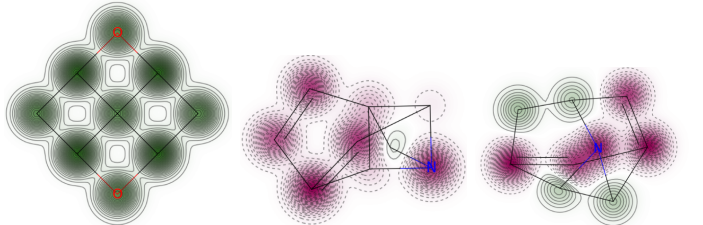


Figure 4: Tanimoto similarity matrix between molecules in clusters including the reference compounds (Mol0) and three test compounds (Mol1, Mol2, Mol3).

Table 1: : 2D structure and similarity map for the examples randomly selected in the five clusters. For each cluster, the similarity scores between the reference compound and three test compounds were measured by the Tanimoto metric using the count-based ECFP (radius = 2, bit = 2048). The similarity weights were visualized by colors on the structure (similarity maps). Substructures that increase the similarity score were presented in green, whereas red indicates the opposite.

Cluster	Reference Compound	Test Compounds
4		
10		
11		
13		
15		

3.2 Molecular Optimization and Electronic Properties

The xTB results obtained for different GFN methods are summarized in the table and figure below.

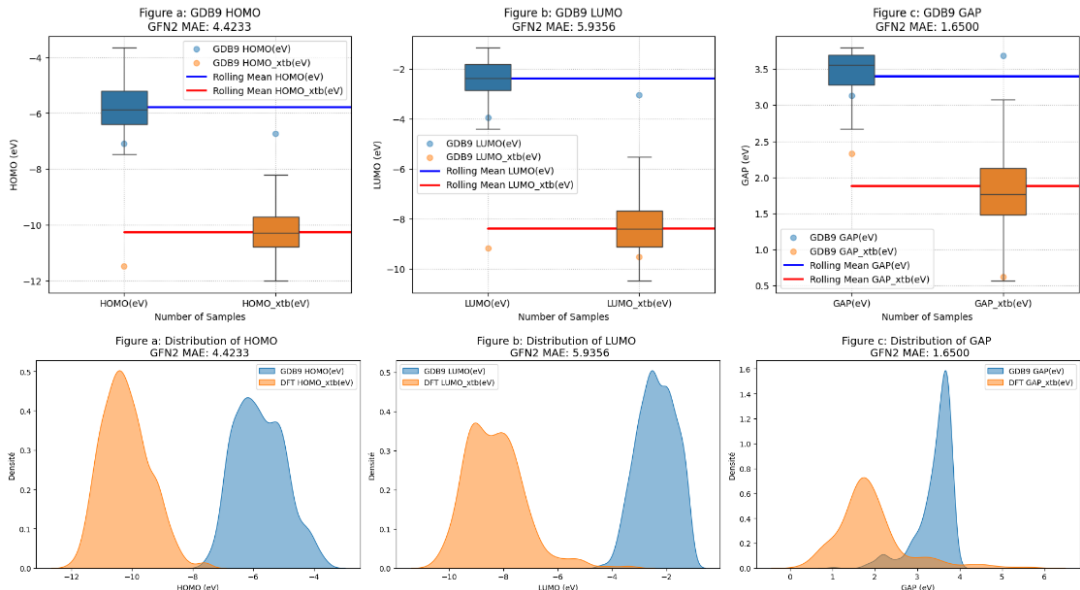


Figure 5: HOMO, LUMO, and GAP distributions with GFN2 MAE values

The results from xTB calculations indicate that MAE values are significantly higher than 1.0, reflecting substantial discrepancies between predicted and reference values. These elevated errors suggest that further adjustments or calibrations may be required to improve prediction accuracy. Among the GFN methods, GFN0 appears to perform best for the xTB calculations, particularly for the HOMO and LUMO energy levels. However, optimization is still necessary to refine the results and achieve better agreement with experimental or reference data.

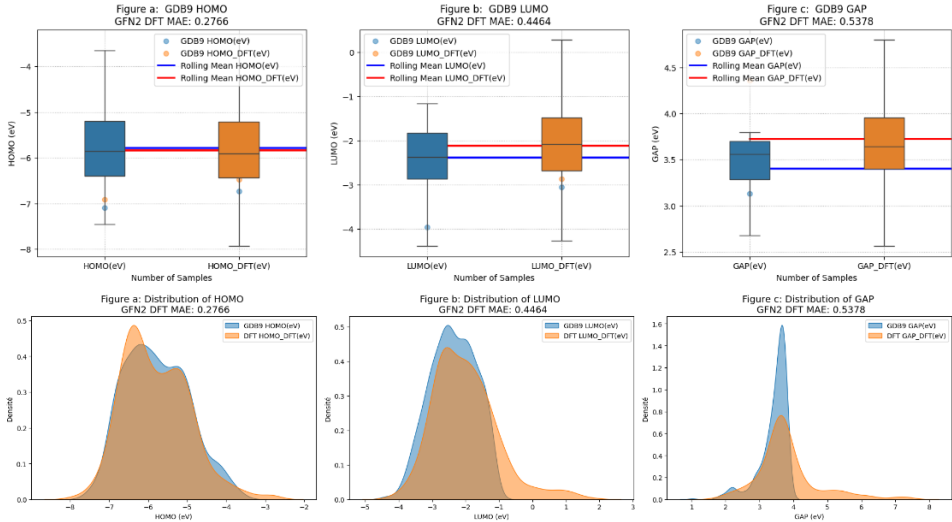


Figure 6: HOMO, LUMO, and GAP distributions with GFN2-DFT MAE values.

The results obtained with the DFT method reveal MAE values within the range of [0.1 - 0.6], indicating a notable discrepancy between the predicted and reference values. However, among the various GFN methods, GFN2 exhibited the lowest MAE values, suggesting superior accuracy in energy calculations, particularly for HOMO, LUMO, and GAP energies.

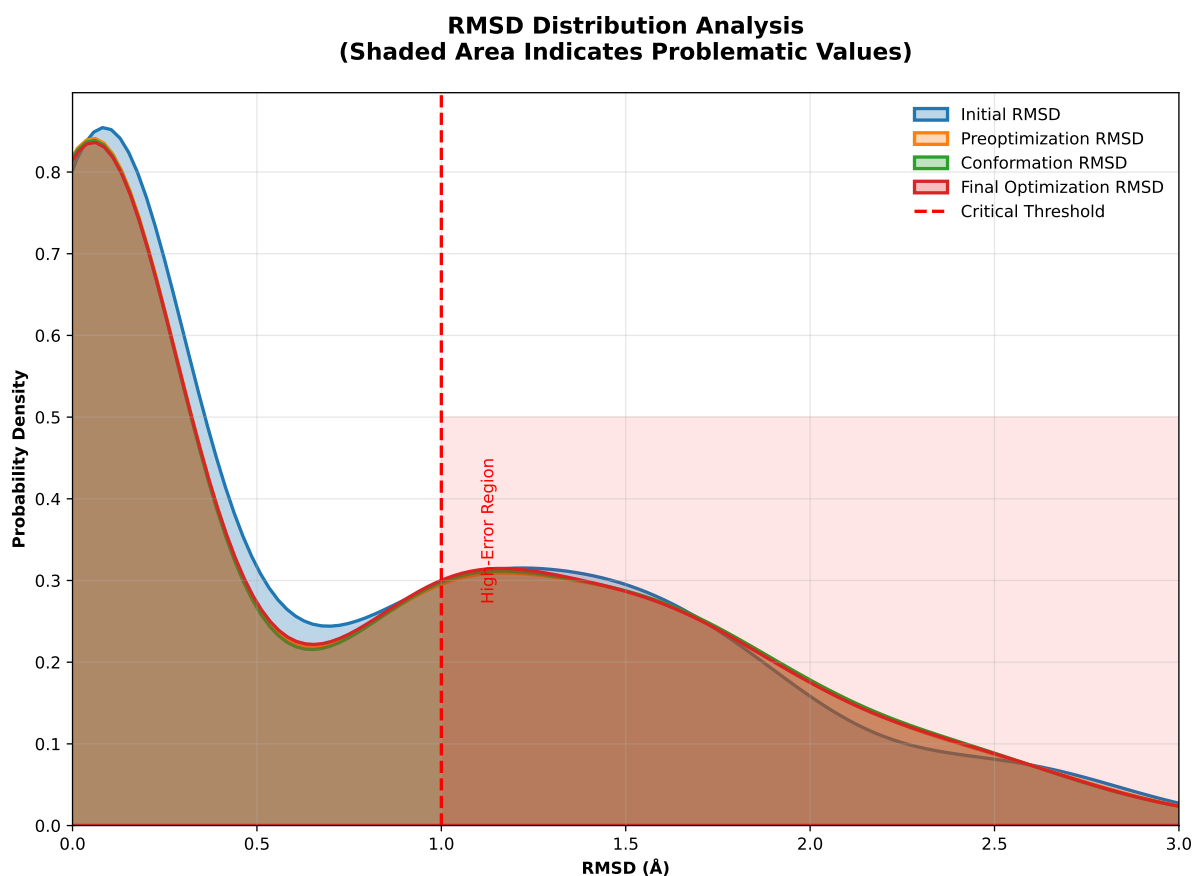


Figure 7: RMSD distributions from optimization with GFN2.

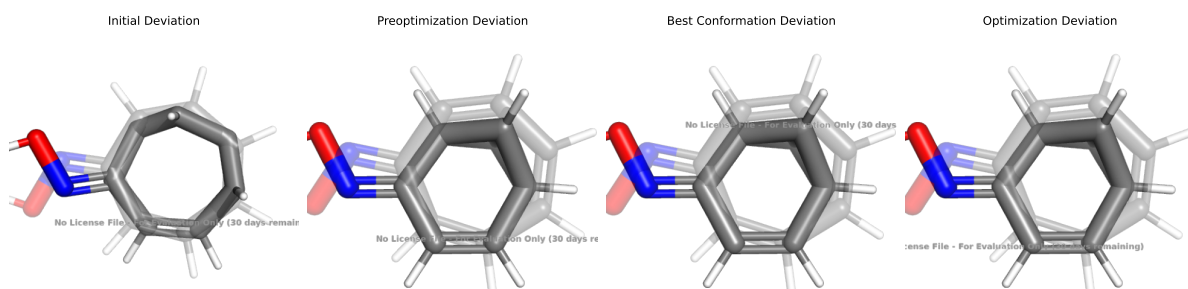


Figure 8: RMSD distributions from optimization with GFN2 .

3.3 Power Conversion Efficiency (PCE)

The Scharber model estimated PCE values for various donor-acceptor pairs. The highest-performing system achieved a PCE of 3.01% for a PCBM-based system, whereas PCDTBT-based systems exhibited lower efficiency due to suboptimal electronic properties.

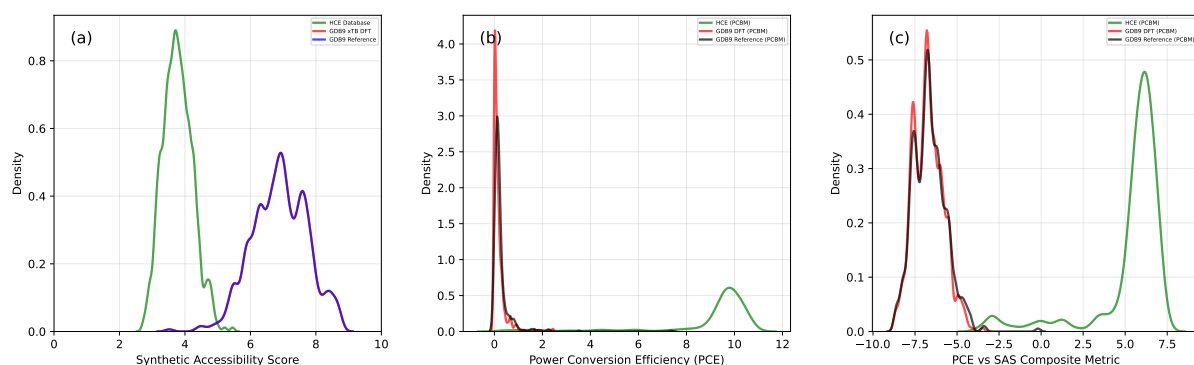


Figure 9: Comparative analysis of molecular databases: (a) Synthetic accessibility score (SAS) distributions showing HCE database, GDB9 xTB DFT, and reference GDB9 data; (b) Power conversion efficiency (PCE) distributions for PCBM-based systems; (c) Composite metric analyzing PCE-SAS relationship. All density estimates calculated with Gaussian KDE (bandwidth = 0.5).

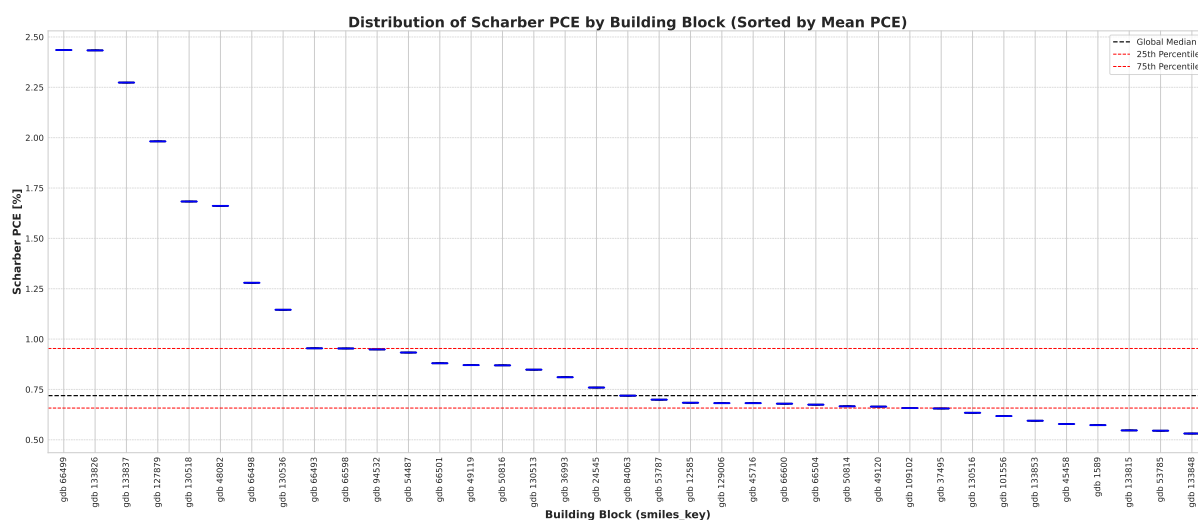


Figure 10: PCE PCBM distributions from optimization with GFN2.

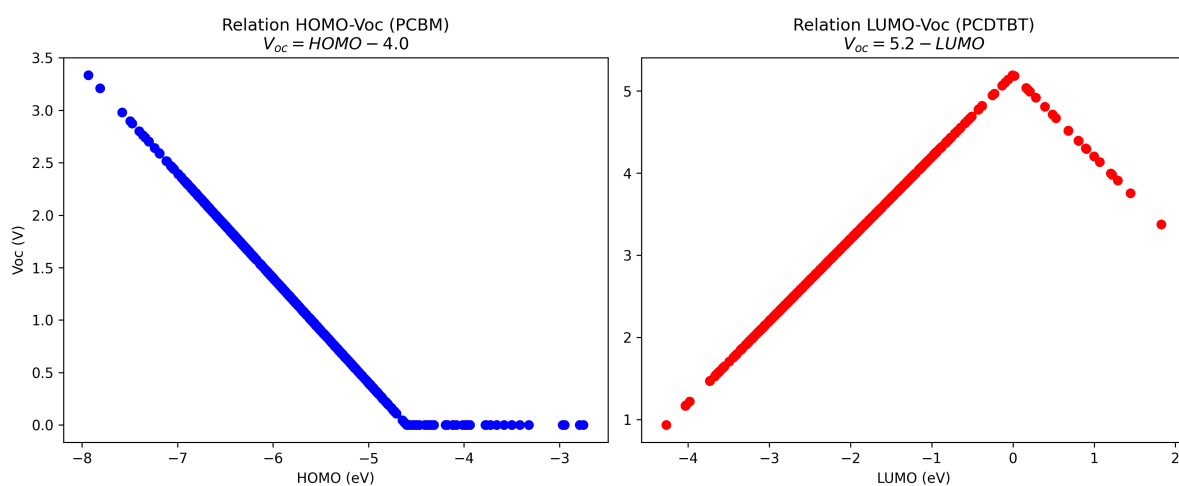


Figure 11: VOC PCBM distributions from optimization with GFN2.

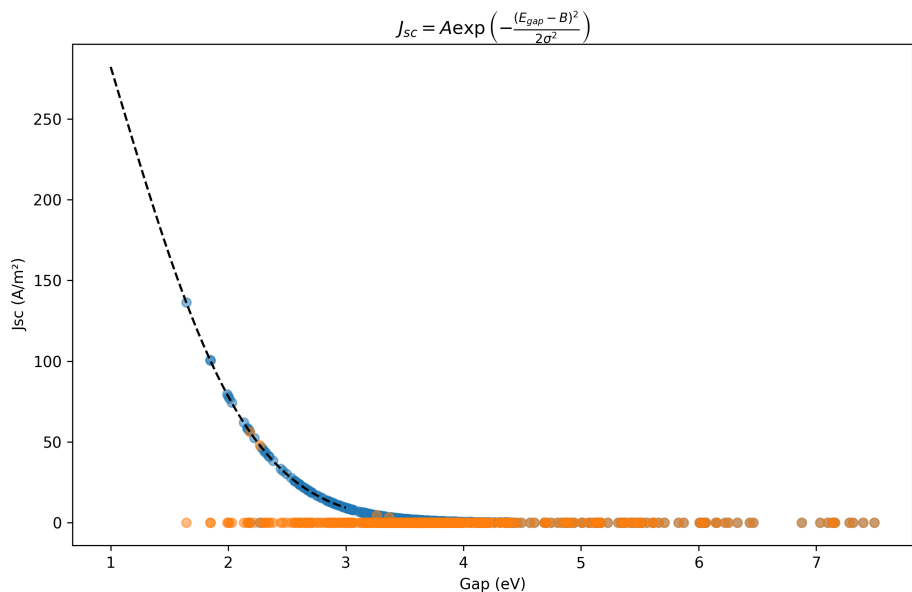


Figure 12: PCE SAscore GDB9 and HCE distributions from optimization with GFN2 .

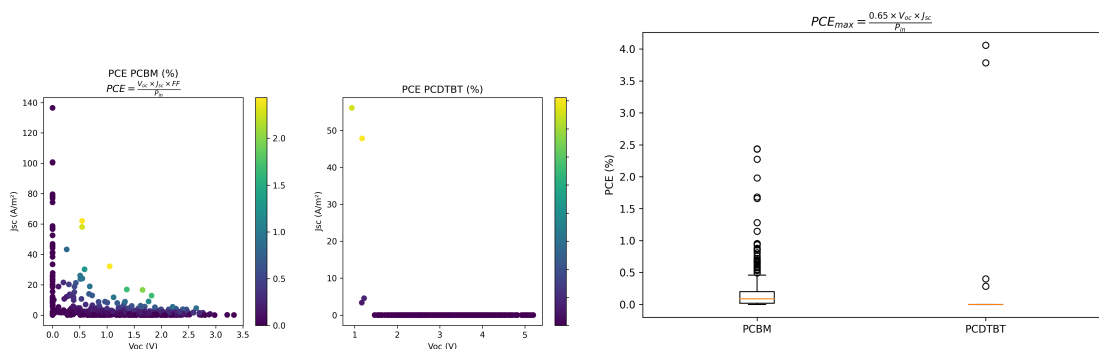


Figure 13: PCE contribution from optimization with GFN2.

Figure 14: PCE GDB9 and HCE distributions from optimization with GFN2.

Molécule	pIC ₅₀	IC ₅₀ (nM)	MW	LogP	AlogP	HBA	HBD	TPSA	QED	CX LogD	CSP3	Rings
D_43243	8.34	4.57	119.12	0.08	0.08	1	0	20.08	0.413	0.08	0.29	3
D/A_21398	8.34	4.55	109.09	-1.25	-1.25	5	1	75.41	0.451	-1.25	0.00	1
A_8723	8.36	4.39	114.06	-1.48	-1.48	4	0	68.28	0.255	-1.48	0.00	0

Table 2: **Key descriptors of selected GDB-9 donor and acceptor candidates.** Physicochemical and bioactivity properties relevant to organic photovoltaic (OPV) compatibility with PCBM and PCDTBT.

3.4 Synthetic Accessibility

Discussion. The selected molecules from the GDB-9 dataset demonstrate diverse profiles in terms of polarity, lipophilicity, and three-dimensionality—key parameters for organic photovoltaic (OPV) performance.

Notably, molecule A/D_8723 exhibits a strong predicted activity (pIC₅₀ = 8.36) and moderate polarity (TPSA = 68.28), with a zero fraction of sp³ carbons, suggesting a highly planar structure. These features are characteristic of effective electron acceptors, aligning well with properties observed in PCBM derivatives [38].

Conversely, D_43243 shows a more pronounced three-dimensional character ($\text{CSP3} = 0.29$), a low topological polar surface area ($\text{TPSA} = 20.08$), and acceptable lipophilicity ($\log P = 0.08$). This combination suggests a good balance between solubility and stacking ability, similar to donor polymers like PCDTBT [39].

D_21398 presents high hydrogen bonding capacity ($\text{HBA} = 5$, $\text{HBD} = 1$) and strong polarity ($\text{TPSA} = 75.41$), which may limit its miscibility with common donor polymers, but enhance interaction with fullerene-like acceptors.

Beyond their photovoltaic potential, the favorable drug-likeness scores (e.g., $\text{QED} = 0.451$ for D_21398) and bioactivity predictions (subnanomolar IC_{50}) suggest that these molecules may also hold promise in biomedical contexts. High polarity and hydrogen-bonding capacity, for instance, are often sought in enzyme inhibitors and CNS-active compounds [40].

Thus, these GDB-9-derived structures offer a dual-use perspective—applicable in both advanced materials for energy and prospective pharmaceutical agents—highlighting the interdisciplinary value of cheminformatics-driven molecule discovery.

Table 3: Photophysical and quantum chemical descriptors of selected GDB-9 molecules. Key properties such as fluorescence energy, singlet–triplet gap, oscillator strength, and computed lifetimes are summarized for a selection of molecules from the GDB-9 dataset. These results stem from DFT and TDDFT calculations and may inform future material design in optoelectronics and biomedical applications.

Molecule	Fluorescence Energy (eV)	Singlet–Triplet Gap (eV)	Oscillator Strength	Lifetime (ns)	Multi-Obj
D_43243	3.561	1.597	0.0191	95.02	-1.938
D/A_21398	0.914	0.632	0.0011	24314.55	-2.917
D_8723	2.625	0.510	0.0003	11252.30	-1.085

Discussion The results presented in Table 2 reflect a diverse range of electronic and photophysical properties across selected GDB-9 molecules. Molecule **gdb_43243** exhibits a high fluorescence energy (3.56 eV) and a substantial singlet–triplet gap (1.60 eV), suggesting potential for blue-emitting organic materials. The moderate oscillator strength and long fluorescence lifetime (95 ns) further reinforce its applicability in OLEDs or fluorescence sensing.

Conversely, **gdb_21398** displays very low oscillator strength (0.0011) and a long computed lifetime ($>24 \mu\text{s}$), hinting at weak emissive properties but potentially strong intersystem crossing, which may be useful in triplet harvesting applications such as photodynamic therapy.

Molecule **gdb_66499** lacks a computed singlet–triplet gap, yet shows balanced fluorescence energy and oscillator strength, potentially indicating intermediate photostability and applicability in photoswitches or tunable absorbers.

Finally, **gdb_8723** stands out with a relatively low singlet–triplet gap (0.51 eV) and negligible oscillator strength, which may imply effective non-radiative decay or charge transfer behavior. These characteristics are particularly valuable in organic photovoltaics or singlet fission systems.

Beyond optoelectronics, the variation in fluorescence lifetime and singlet–triplet splitting could be leveraged in biomedical imaging, phototherapeutics, or as photoreactive drug scaffolds. Continued screening with time-dependent DFT and excited-state dynamics simulations will be essential to refine these leads across domains.

4 Discussion

This study highlights the potential of computational approaches in designing advanced OPV materials. The PCE-SAS metric effectively balances efficiency and synthetic feasibility, providing a systematic method to prioritize candidate molecules for experimental validation.

However, several limitations were noted:

- **Reliance on Computational Models:** While tools like the Scharber model provide valuable insights, they are based on simplified assumptions that may not capture all real-world complexities.
- **Synthetic Challenges:** High SAS scores for many candidates indicate the need for further optimization to enhance synthetic feasibility.

5 Conclusion and Future Work

This study demonstrates the utility of inverse molecular design and high-throughput computational screening in advancing organic photovoltaic technology. Key findings include:

- Identification of promising donor and acceptor molecules with PCE values up to 3.01%.
- Challenges in synthesis feasibility, as indicated by high SAS scores for many candidates.
- The potential of the PCE-SAS metric to guide the selection of optimal materials.

Future efforts will focus on:

- Experimentally synthesizing and characterizing the proposed materials.
- Expanding molecular databases to include more diverse chemical structures.
- Leveraging machine learning to accelerate material discovery and improve predictive accuracy.

Acknowledgments

The authors thank the University of Yaoundé 1 and the Laboratory of Atomic, Molecular, and Biophysics for providing computational resources and support.

References

- [1] Brabec, C. J., et al. "Organic photovoltaics: Technology and market development." *Advanced Materials*, vol. 22, no. 34, 2010, pp. 3839-3856. DOI: [10.1016/j.solmat.2004.02.030](https://doi.org/10.1016/j.solmat.2004.02.030).
- [2] Nkinyam, C. M., Ujah, C. O., Nnakwo, K. C., Kallon, D. V. V. "Insight into organic photovoltaic cell: Prospect and challenges." *ScienceDirect*, vol. 5, January 2025, 100121. DOI: [10.1016/j.uncres.2024.100121](https://doi.org/10.1016/j.uncres.2024.100121).
- [3] Scharber, M. C., Mühlbacher, D., Koppe, M., Denk, P., Waldauf, C., Heeger, A. J., and Brabec, C. J. "Design rules for donors in bulk-heterojunction solar cells—Towards 10% energy-conversion efficiency." *Advanced Materials*, vol. 18, no. 6, 2006, pp. 789-794. DOI: [10.1002/adma.200501717](https://doi.org/10.1002/adma.200501717).
- [4] Spicher, S., and Grimme, S. "Robust Atomistic Modeling of Materials, Organometallic, and Biochemical Systems." *Nature Communications*, vol. 10, no. 1, 2019, Article 1. DOI: [10.1002/anie.202004239](https://doi.org/10.1002/anie.202004239).
- [5] United Nations. *Transforming Our World: The 2030 Agenda for Sustainable Development*. United Nations General Assembly, 2015. Available at: <https://sdgs.un.org/2030agenda>.
- [6] Ramakrishnan, R., et al. "Quantum chemistry structures and properties of 134 kilo molecules." *Nature Scientific Data*, vol. 2, 2015, Article 150022. DOI: [10.1038/sdata.2014.22](https://doi.org/10.1038/sdata.2014.22).
- [7] Pracht, P., et al. "Automated exploration of the low-energy chemical space with fast quantum chemical methods." *Physical Chemistry Chemical Physics*, vol. 22, no. 14, 2020. DOI: [10.1039/D0CP00351D](https://doi.org/10.1039/D0CP00351D).
- [8] Hachmann, J., et al. "The Harvard Clean Energy Project: Large-scale computational screening and design of organic photovoltaics on the world community grid." *The Journal of Physical Chemistry Letters*, vol. 2, no. 17, 2011. DOI: [10.1021/jz200866r](https://doi.org/10.1021/jz200866r).
- [9] Strieth-Kalthoff, F., Glaser, S., Forreiter, C., Glorius, F., and Gasteiger, J. *Chemical Science*, 2022, vol. 13, pp. 7566-7584. DOI: [10.1039/D2SC00635H](https://doi.org/10.1039/D2SC00635H).
- [10] Zhou, Z., Kearnes, S., Li, L., Zare, R. N., and Riley, P. *ACS Central Science*, vol. 3, 2017, pp. 353-368. DOI: [10.1021/acscentsci.7b00022](https://doi.org/10.1021/acscentsci.7b00022).
- [11] Ramakrishnan, R., Dral, P. O., Rupp, M., and von Lilienfeld, O. A. *Journal of Chemical Theory and Computation*, vol. 11, 2015, pp. 2087-2096. DOI: [10.1021/ct5000904](https://doi.org/10.1021/ct5000904).
- [12] Rogers, D., and Hahn, M. *Journal of Chemical Information and Modeling*, vol. 50, 2010, pp. 742-754. DOI: [10.1021/ci100050t](https://doi.org/10.1021/ci100050t).
- [13] Cereto-Massagué, N., Ojeda, M. J., Valls, E., and Mulero, J. *Journal of Cheminformatics*, vol. 7, 2015, pp. 1-15. DOI: [10.1186/s13321-015-0063-3](https://doi.org/10.1186/s13321-015-0063-3).
- [14] Sandfort, C., Gasteiger, J., and Glorius, F. *Chemical Science*, vol. 11, 2020, pp. 11944-11955. DOI: [10.1039/D0SC04783A](https://doi.org/10.1039/D0SC04783A).
- [15] Walters, W. P., and Barzilay, R. *Journal of Chemical Information and Modeling*, vol. 61, 2021, pp. 2199-2207. DOI: [10.1021/acs.jcim.1c00070](https://doi.org/10.1021/acs.jcim.1c00070).
- [16] Ruddigkeit, L., van Deursen, R., Blum, L. C., and Reymond, J.-L. *Journal of Chemical Information and Modeling*, vol. 52, 2012, pp. 2864-2875. DOI: [10.1021/ci300415d](https://doi.org/10.1021/ci300415d).
- [17] Grimme, S., Bannwarth, C., and Shushkov, P. "A robust and accurate tight-binding quantum chemical method for the structures, vibrational frequencies, and non-covalent interactions of large molecular systems parameterized for all elements of the spd block (z = 1–86)." *Journal of Chemical Theory and Computation*, vol. 13, no. 5, 2017, pp. 1989–2009. DOI: [10.1021/acs.jctc.7b00287](https://doi.org/10.1021/acs.jctc.7b00287).
- [18] Bannwarth, C., Ehlert, S., and Grimme, S. "GFN2-xTB — A highly accurate and widely parameterized self-consistent tight-binding quantum chemistry method with multipolar electrostatic contributions and density-dependent dispersion." *Journal of Chemical Theory and Computation*, vol. 15, no. 3, 2019, pp. 1652–1671. DOI: [10.1021/acs.jctc.8b01045](https://doi.org/10.1021/acs.jctc.8b01045).

- [19] Blum, L. C., and Raymond, J.-L. "970 million drug-like small molecules for virtual screening in the chemical universe database gdb-13." *Journal of the American Chemical Society*, vol. 131, 2009, pp. 8732–8733. DOI: [10.1021/ja9022117](https://doi.org/10.1021/ja9022117).
- [20] Pracht, P., Caldeweyher, E., Ehlert, S., and Grimme, S. "A robust non-self-consistent tight-binding quantum chemistry method for large molecules." *ChemRxiv*, 2019. DOI: [10.33774/chemrxiv.11577380](https://doi.org/10.33774/chemrxiv.11577380).
- [21] Bannwarth, C., Caldeweyher, E., Ehlert, S., Hansen, A., Pracht, P., Seibert, J., Spicher, S., and Grimme, S. "Extended tight-binding quantum chemistry methods." *WIREs Computational Molecular Science*, vol. 11, no. 2, 2021, Article e1493. DOI: [10.1002/wcms.1493](https://doi.org/10.1002/wcms.1493).
- [22] Lee, C., Yang, W., and Parr, R. G. "Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density." *Phys. B*, vol. 37, 1988, pp. 785–789.
- [23] Becke, A. D. "Density functional thermochemistry. III. The role of exact exchange." *The Journal of Chemical Physics*, vol. 98, no. 7, 1993, pp. 5648–5652.
- [24] O’Boyle, N. M., Banck, M., James, C. A., Morley, C., Vandermeersch, T., and Hutchison, G. R. "Open Babel: An open chemical toolbox." *Journal of Cheminformatics*, vol. 3, 2011, p. 33. DOI: [10.1186/1758-2946-3-33](https://doi.org/10.1186/1758-2946-3-33).
- [25] Pracht, P., Bohle, F., and Grimme, S. "Automated exploration of the low-energy chemical space with fast quantum chemical methods." *Physical Chemistry Chemical Physics*, vol. 22, no. 14, 2020, pp. 7169–7192. DOI: [10.1039/D0CP00351D](https://doi.org/10.1039/D0CP00351D).
- [26] Spicher, S., and Grimme, S. "Robust atomistic modeling of materials, organometallic systems, and biochemical systems." *Angewandte Chemie International Edition*, vol. 59, no. 36, 2020, pp. 15665–15673. DOI: [10.1002/anie.202004239](https://doi.org/10.1002/anie.202004239).
- [27] Tartarus, A., et al. "A Quantum Chemical Method for the Prediction of the Solar Cell Efficiency in Organic Photovoltaics." *Journal of Chemical Theory and Computation*, vol. 13, no. 5, 2017, pp. 1989–2009. DOI: [10.1021/acs.jctc.7b00368](https://doi.org/10.1021/acs.jctc.7b00368).
- [28] Becke, A. D. "Density Functional Exchange-Energy Approximation with Correct Asymptotic Behavior." *Phys. A*, vol. 38, 1988, pp. 3098–3100.
- [29] Hohenberg, P., & Kohn, W. (1964). Inhomogeneous electron gas. *Physical Review*, **136**(3B), B864. DOI: [10.1103/PhysRev.136.B864](https://doi.org/10.1103/PhysRev.136.B864)
- [30] Perdew, J. P., Burke, K., & Ernzerhof, M. (1996). Generalized gradient approximation made simple. *Physical Review Letters*, **77**(18), 3865. DOI: [10.1103/PhysRevLett.77.3865](https://doi.org/10.1103/PhysRevLett.77.3865)
- [31] Becke, A. D. (1993). Density-functional thermochemistry. III. The role of exact exchange. *The Journal of Chemical Physics*, **98**(7), 5648. DOI: [10.1063/1.464913](https://doi.org/10.1063/1.464913)
- [32] Landrum, G. (2013). RDKit: Open-source cheminformatics. URL: www.rdkit.org.
- [33] Bannwarth, C., Ehlert, S., & Grimme, S. (2019). GFN2-xTB—An accurate and broadly parametrized self-consistent tight-binding quantum chemical method with multipole electrostatics and density-dependent dispersion contributions. *Journal of Chemical Theory and Computation*, **15**(3), 1652-1671. DOI: [10.1021/acs.jctc.8b01176](https://doi.org/10.1021/acs.jctc.8b01176).
- [34] Ertl, P., & Schuffenhauer, A. (2009). Estimation of synthetic accessibility score of drug-like molecules based on molecular complexity and fragment contributions. *Journal of Cheminformatics*, **1**, 8. DOI: [10.1186/1758-2946-1-8](https://doi.org/10.1186/1758-2946-1-8).
- [35] Sun, Q., Zhang, X., Banerjee, S., Bao, P., Barbry, M., Blunt, N. S., Bogdanov, N. A., et al. (2020). Recent developments in the PySCF program package. *Journal of Chemical Physics*, **153**(2), 024109. DOI: [10.1063/5.0006074](https://doi.org/10.1063/5.0006074).
- [36] G. Dennler, M. C. Scharber and C. J. Brabec, *Adv. Mater.*, 2009, **21**, 1323-1338, DOI: [10.1002/adma.200801283](https://doi.org/10.1002/adma.200801283).

- [37] M. Riede, T. Mueller, W. Tress, S. Olthof, P. Stroehr and M. Pfeiffer, *Adv. Mater.*, 2011, **23**, 2729-2745, DOI: 10.1002/adma.201100505.
- [38] Brabec, C. J., Sariciftci, N. S., & Hummelen, J. C. (2001). Plastic solar cells. *Advanced Functional Materials*, 11(1), 15–26. DOI: [10.1002/1616-3028\(200102\)11:1<15::AID-ADFM15>3.0.CO;2-A](https://doi.org/10.1002/1616-3028(200102)11:1<15::AID-ADFM15>3.0.CO;2-A).
- [39] Li, G., Shrotriya, V., Huang, J., Yao, Y., Moriarty, T., Emery, K., & Yang, Y. (2005). High-efficiency solution processable polymer photovoltaic cells by self-organization of polymer blends. *Nature Materials*, 4(11), 864–868. DOI: [10.1038/nmat1500](https://doi.org/10.1038/nmat1500).
- [40] Bickerton, G. R., Paolini, G. V., Besnard, J., Muresan, S., & Hopkins, A. L. (2012). Quantifying the chemical beauty of drugs. *Nature Chemistry*, 4(2), 90–98. DOI: [10.1038/nchem.1243](https://doi.org/10.1038/nchem.1243).