University of London
Imperial College of Science, Technology and Medicine
Department of Materials

High Throughput Workflow for the Computational Design of New Thermally Activated Delayed Fluorescence Emitters

Kritam Thapa
Supervisor: Dr. Laura E. Ratcliff

This thesis is submitted for fulfilment of the requirements for the degree of
Doctor of Philosophy
15/03/2023

Statement of Originality

I declare that this thesis is an original report of my research, has been written solely by me under the supervision of Dr. Laura E. Ratcliff, and has not been submitted for any previous degree or qualification. All of the results presented in this thesis are based on my work during the research degree program at Imperial College London, from October 2018 to March 2023.

Copyright Declaration

Abstract

This thesis explores the use of computational methods for discovering new TADF molecules, with a focus on developing a high-throughput virtual screening workflow that reduces costs and time associated with experimental screening. Using methods like STONED and SYBA, diverse molecule libraries were generated and evaluated to identify promising candidates for further investigation. The study also examines the challenges of using computational methods, such as discrepancies and limitations with computationally efficient methods. Modifications were made to parent molecules based on ΔSCF calculations and similarity map analysis. Overall, this study provides valuable insights into the use of computational methods for TADF molecule design and offers guidance for future research aimed at designing new TADF materials.

Acknowledgements

I am deeply grateful to my supervisor, Dr. Laura E. Ratcliff, for giving me the opportunity to delve into the fascinating field of theory and simulation of materials. Throughout my research, she provided a supportive and pleasant work environment, offering constant encouragement, guidance, and feedback. Her patience in mentoring me went above and beyond, making me a better scientist. Her expertise and passion inspire me, and I owe my growth and development as a researcher to her exceptional mentorship. Thank you, Dr. Ratcliff, for this invaluable experience.

I would also like to extend my heartfelt gratitude to Dr. Martina Stella for her valuable support and feedback throughout the years. Furthermore, I would like to thank Prof. Peter Haynes for his insightful feedback on my thesis.

I would also like to extend my gratitude to my Imperial friends for making my time at Imperial College London memorable. Especially, I want to thank Matthias Golomb, Kazuki Morita, and Lucy Whalley for their friendship, camaraderie, and support throughout this journey. Their presence and companionship have not only made the academic experience enjoyable but have also provided me with a sense of belonging within the academic community and outside of it as friends. Their contributions and discussions have been invaluable, and I am fortunate to have had the opportunity to meet with such remarkable colleagues.

I would also like to thank my *Thulo Baba* (Yam Thapa), *Thulo Ama* (Buddhisara Thapa) and *Dudu Bajai* (Tilisara Rana) for their love and support. Special thanks go to my cousins, *Thulo Dai* (Youwan) and *Sano Dai* (Jawan Thapa), their presence during the final stages of my PhD journey was incredibly uplifting, and I am grateful for the camaraderie we share.

Additionally, I want to express my heartfelt gratitude to my *Kancha Baba* (Bikash Thapa) for the moral support provided towards the end of my PhD journey. Your encouragement and belief in my abilities were instrumental in pushing me forward when challenges seemed insurmountable.

Lastly, and most importantly, I want to express my heartfelt appreciation to my family. My *Ama* (Tulasi Thapa), my *Baba* (Krishna Thapa), and my *Bhai* (Pritam Thapa), were there for me through both easy and tough times. Their encouragement and belief in me have been the

driving force behind my achievements, and I owe my success to their constant presence and encouragement. Thank you, *Ama*, *Baba*, and *Bhai*, for being my pillars of strength and for always being there to lift me up when I needed it the most. Your support has meant the world to me, and I could not have reached this milestone without you. I am truly blessed to have such a loving and supportive family.

Last but certainly not least, I acknowledge support from the Engineering and Physical Sciences Research Council (EP/S515085/1). Calculations were performed on the Imperial College High Performance Computing Service and the ARCHER UK National Supercomputing Service.

To all of you, thank you for being an essential part of my life and for your valuable contributions that have brought me to this important milestone. Your support has been invaluable, and I am truly blessed to have such wonderful people in my life.

Contents

Abstract **v**

Acknowledgements **vii**

1 Organic Light-Emitting Diodes **1**

1.1 Introduction 1

1.2 Towards Cheap and Eco-Friendly Display Technology 3

1.3 The Dawn of Organic Electronics 4

1.3.1 Electroluminescence 5

1.3.2 Fluorescence-based OLED 8

1.3.3 Phosphorescence 10

1.4 Thermally Activated Delayed Fluorescence 11

1.4.1 Key TADF Mechanisms 12

1.4.2 The Traditional Design Strategy for TADF molecules 17

1.4.3 Beyond Donor and Acceptor Designs 18

2 Density Functional Theory **21**

2.1 The Schrödinger Equation 23

2.2 The Born-Oppenheimer Approximation 25

2.3 The Hartree Product 28

2.3.1 Limitations of Hartree Product 29

2.4 The Electron Density 30

2.5 Foundations of Density Functional Theory 30

2.5.1 Kohn Sham Formalism of DFT 32

2.5.2 Exchange Correlation Functional 34

2.5.3 Solving Kohn Sham Equation 43

2.6 Density Functional Theory-Based Methods for Simulating Excited States 44

2.6.1 Time Dependent Density Functional Theory 44

2.6.2 ΔSelf-Consistent Field Density Functional Theory 45

2.7 Pseudopotentials 46

2.7.1 Non-linear Core Correction 47

2.8 Basis sets 48

2.9 BigDFT 50

2.9.1 Bromine and Iodine NLCC pseudopotential 51

2.10 Force Field Optimisation 54

2.10.1 The Merck Molecular Force Field 55

3 High Throughput Virtual Screening Workflow 58

3.1 Introduction . . . 58

3.2 Applications . . . 60

3.3 Philosophy of High Throughput Virtual Screening . . . 62

3.4 Our Workflow . . . 63

3.4.1 Library Generation . . . 65

3.4.2 Molecule Representation . . . 66

3.4.3 Superfast Traversal, Optimisation, Novelty, Exploration and Discovery Algorithm . . . 71

3.4.4 Parent Molecules . . . 71

3.4.5 Molecular Diversity . . . 73

3.4.6 Synthetic Accessibility Assessment . . . 76

3.4.7 Force Field Geometry Optimisation . . . 80

3.4.8 BigDFT Parameters . . . 81

3.4.9 DFT on MMFF94 Geometry Optimised Molecules . . . 81

3.4.10 DFT Geometry optimisation . . . 82

3.4.11 Frontier orbitals and Singlet-Triplet Splittings . . . 82

3.5 Summary . . . 82

4 High Through Put Virtual Screening Workflow Results - Part 1 84

4.1 Introduction . . . 84

4.2 Library Generation . . . 85

4.3 Molecule Library Check 88

4.3.1 Diversity and Relevance 88

4.3.2 Stability 93

4.4 Tanimoto Coefficient 95

4.5 Synthesis accessibility 99

4.5.1 SYnthetic Bayesian Accessibility Score 99

4.6 MMFF94 Geometry optimisation 101

4.7 Summary and Conclusion 102

5 High Through Put Virtual Screening Workflow Results - Part 2 105

5.1 Introduction 105

5.2 DFT on Force Field Optimised Geometry 106

5.3 DFT Geometry optimisation 111

5.4 Frontier Orbitals and Singlet-Triplet Splittings 112

5.4.1 Similarity Map 115

5.4.2 Candidates 116

5.5 Summary and Conclusion 129

6 Outlook and Conclusion 131

6.1 Conclusion 131

6.2 Future Developments 132

6.3 Data Availability 134

Bibliography **134**

List of Tables

4.1 SMILES representation of parent molecules used to generate molecule libraries. N_{atoms} represents the number of atoms in a molecule excluding hydrogen atoms. 85

4.2 Total number of SMILES from parent molecules and obtained the corresponding canonical SMILES. The percentage of canonical SMILES obtained is calculated as the ratio of the number of canonical SMILES to the total number of SMILES generated. 87

4.3 The number of molecules with an SYBA score of -20 or greater, along with the number of molecules successfully optimized using MMFF94. 102

4.4 Number of SMILES at each filtering stage in the high-throughput virtual screening workflow discussed in this chapter. 103

5.1 Number of molecules remaining after applying the HOMO-LUMO and density overlap threshold during the high-throughput virtual screening workflow. 111

5.2 Number of molecules remaining after DFT geometry optimisation in the high-throughput virtual screening workflow discussed in this work. 111

5.3 Singlet-triplet energy gap (ΔE_{ST}) and HOMO-LUMO energy gap (eV) and SYnthetic Bayesian Accessibility (SYBA) score of the parent molecules. 116

5.4 Singlet-triplet energy gap (ΔE_{ST}) and HOMO-LUMO energy gap (eV), SYnthetic Bayesian Accessibility (SYBA) score and Tanimoto coefficient (compared to parent molecule) of molecules chosen based on smallest ΔE_{ST}. 117

List of Figures

1.1 Diagram of a simple OLED device structure, featuring multiple layers sandwiched between an anode and cathode electrode, with a substrate layer providing structural support. The hole injection and electron injection layers enable the efficient emission of light when a voltage is applied. 6

1.2 Jablonski diagram of the fluorescence mechanism in OLEDs. ISC, ΔE_{ST}, e-, h+ represent intersystem crossing (ISC), singlet-triplet energy gap ΔE_{ST}, electrons (e-) and holes (h+), respectively. 9

1.3 Jablonski diagram of the phosphorescence mechanism in OLEDs. ISC, ΔE_{ST}, e-, h+ represent intersystem crossing (ISC), singlet-triplet energy gap ΔE_{ST}, electrons (e-) and holes (h+), respectively. 10

1.4 Jablonski diagram of the TADF mechanism in OLEDs. RISC, ISC, ΔE_{ST}, e-, h+ represent reverse intersystem crossing (RISC), intersystem crossing (ISC), singlet-triplet energy gap ΔE_{ST}, electrons (e-) and holes (h+), respectively. . . . 12

2.1 (a) The convergence of LUMO-HOMO absolute relative values with respect to grid spacing h for molecules containing I and Br atoms. The y-axis represents the absolute relative LUMO-HOMO values, which are compared to those obtained at $h = 0.3$ (b) The convergence of the absolute relative Force Norm as a function of h, where y values are relative to the y value obtained at $h = 0.3$. A molecule is represented by a consistent symbol and colour. 53

3.1 Outline of our workflow schematic illustrating the method and design space, with accompanying snapshots of analysis and results from later chapters. 64

3.2 2D SMILES representation of TADF molecules used to generate molecule library. 72

4.1 Relationship between total SMILES generated from parent molecules and canonical SMILES obtained. 86

4.2 Plot between parent molecule size (number of atoms N_{atoms}) and canonical SMILES obtained. 87

4.3 Molecules distribution by the number of atoms (N_{atoms}), showing the frequency of molecules with respect to the number of atoms they contain. The solid red vertical line represents the threshold for excluding molecules below a certain size, while the dotted red line indicates the size of the reference parent molecule. . . . 90

4.4 Molecular size and fingerprint (FP) score displayed in a scatter plot. The solid red line denotes the established threshold for filtering out SMILES, while the dashed red line represents the size of the parent molecules used as a reference point. 92

4.5 Number of canonical molecules that are closed and open shell. 94

4.6 Visualisations of the SMILES distribution, post-filtering of small and open-shell systems, including a histogram, boxplot, and kernel density estimate (KDE) plot (blue line), as discussed in the preceding section of this chapter. 97

4.7 2D similarity maps showing graphical representations of the Tanimoto coefficient between all molecules. A high coefficient indicates that the molecule is more similar to the parent molecule, while a low score indicates greater dissimilarity. . 98

4.8 Molecule distribution by SYnthetic Bayesian Accessibility (SYBA) score, showing the frequency of molecules with respect to their SYBA score. The solid red line indicates the SYBA score of the parent molecule, while the dotted red line represents the threshold of -20 for SYBA score 100

5.1 Boxplots showing the distribution of (a) HOMO-LUMO energy gap and (b) HOMO-LUMO overlap (Λ_T) values across the filtered molecule library. The black diamond represents the value of the corresponding parent molecule. 107

5.2 HOMO-LUMO energy gap (eV) and HOMO-LUMO overlap Λ_T of molecules, with color-coded Tanimoto coefficients. The red vertical dashed line represents the value of the parent molecule's Λ_T and the red horizontal dashed line represents the parents HOMO-LUMO (eV). 110

5.3 Box plots showing the distribution of (a) HOMO-LUMO energy gap and (b) singlet-triplet splitting (ΔE_{ST}) energy for the parent molecules and their derivatives. The black diamond represents the value of the corresponding parent molecule. 112

5.4 The singlet-triplet splitting (ΔE_{ST}) and LUMO-HOMO gap are shown using black markers for the parent molecule and coloured markers for the candidate molecules. 114

5.5 Visualisation of similarity maps generated by comparing the selected TADF molecules with the smallest ΔE_{ST} (eV) values from each D-A TADF family to their parent molecules. The green colour indicates a higher similarity between the parent and child molecules in that particular region, while the red colour indicates a lower similarity. 124

5.6 Visualisation of the similarity maps generated using the chosen TADF molecules with the smallest ΔE_{ST} (eV) values from each type-VII TADF family. The green colour indicates a higher similarity between the parent and child molecules in that particular region, while the red colour indicates a lower similarity. 125

6.1 Difference in HOMO-LUMO energy gap (eV) of MMFF94 and DFT optimised molecules. 133

Chapter 1

Organic Light-Emitting Diodes

1.1 Introduction

Global energy consumption has been rising exponentially since the industrial revolution, and there are no signs of it slowing down [1]. In 2021, zero-carbon energy sources contributed only about 16% of global energy consumption. The impact of greenhouse gases, like CO_2, on the global temperature is well-documented [2, 3]. Simply replacing non-renewable sources with renewable sources, like windmills and solar panels, is only one aspect of addressing the climate and energy crises. Renewables are intermittent and require significant upfront investment and land use [4]. Additionally, meeting global energy demand will require a mix of energy sources, energy-efficient technologies, and policies that promote conservation and innovation. The rising energy demand, fueled by rapid economic growth worldwide, is contributing to the increasing global temperature [5]. The other part of the solution is complex and requires multiple strategies [6]. Nonetheless, this work aims to make everyday technology sustainable and run at peak efficiency.

In 2019, global electric waste (e-waste) generation reached 53.6 million metric tons, with only 17.4% of this waste being collected and recycled [7]. By 2030, this number is projected to grow to 74 million metric tons (Mt). The need for sustainable technology is just as critical as efficiency. The materials used to manufacture a device not only affect its efficiency but

also its sustainability. For example, the size of an electric car's battery (measured in kilowatt-hours, or kWh) determines its maximum energy output, while factors like body shape, tire, motor, and weight can affect its range (efficiency). However, the use of sustainable materials in the manufacturing process can also improve the device's overall environmental impact. By promoting sustainability in technology, we can reduce e-waste and contribute to a cleaner, more efficient future.

Lighting accounts for approximately 20% of the world's electricity consumption, making it crucial to develop sustainable and efficient solid-state lighting (SSL) technology for the future [8]. However, the current generation of SSL devices based on OLED technology relies on heavy metals such as iridium and platinum, which are both harmful to the environment and expensive [8, 9, 10, 11]. Additionally, e-waste generated in 2019 included or was composed of lighting and display technologies, such as lamps, screens, and monitors, contributing to about 23% of the total waste generated that year [7]. Therefore, it's essential to address the environmental impact of SSL devices and prioritise the use of sustainable materials in their manufacture. Just as with electric cars, considering the materials used in developing lighting and display solutions is critical to ensuring their sustainability and efficiency.

Thermally activated delayed fluorescence (TADF) molecules are a highly promising type of organic molecule used in OLEDs. Unlike traditional OLED emitters, which can only harvest singlet excitons, TADF molecules are capable of harvesting both singlet and triplet excitons, leading to highly efficient energy transfer and emission. TADF emitters have attracted significant attention due to their potential to replace heavy metal-based OLED emitters [12]. A recent study [13] has shown that new TADF molecules can be created without the traditional design strategy. However, the large number of possible TADF emitters can make experimental screening time-consuming and expensive, which is one of the major challenges in designing the next-generation OLEDs. Therefore, our work aims to design a blueprint for HTVS that can be used to design new TADF molecules while simultaneously reducing the costs associated with experimental screening. Achieving this goal will help us overcome the existing hurdles.

In this chapter, we will embark on a journey through the history of OLED emitter types, be-

ginning with early fluorescent emitters, advancing to phosphorescent emitters, and ultimately arriving at the latest generation of TADF emitters. Through this journey, we will analyse the key differences between these emitter types. Moving forward, we will discuss traditional design strategies for optimising OLED and TADF emitters, such as molecular structure modification. Moreover, we will go beyond these conventional design approaches and examine advanced strategies that hold the potential to further enhance the performance of these materials.

1.2 Towards Cheap and Eco-Friendly Display Technology

The use of display technology is omnipresent in modern devices, playing a crucial role in essential sectors of daily life, such as work, education, and health. As a primary driver of energy consumption, the demand for more efficient display technology continues to rise. In recent decades, we have transitioned from traditional incandescent bulbs and compact fluorescent lamps (CFLs) to light-emitting diodes (LEDs), which have emerged as a superior lighting technology in terms of lifespan, colour rendering, physical robustness, size, speed, and eco-friendliness [14]. LEDs are classified as solid-state lighting (SSL) technology, utilising semiconductors to convert electricity into light, while traditional bulbs rely on electrical filaments or plasma and gas. Notably, white-light emitting LEDs are currently more than twice as efficient as incandescent bulbs [15].

The first visible LED, emitting only red light, was demonstrated in 1962 [16]. It was not until the late 1960s that red and green LEDs became commercially available, while blue LEDs remained a significant challenge due to the requirement for a wider band gap and a limited number of candidate materials at the time. Three major milestones led to significant improvements in blue LED efficiency in 1993 [17]. First, techniques were developed in the 1980s to produce large, high-quality, gallium nitride (GaN) crystals [18]. Second, techniques were developed to turn GaN into a suitable p-type semiconductor since it is naturally n-type [19, 20]. Finally, in the 1990s, the efficiency of blue LEDs was significantly improved [17]. These breakthroughs were the work of Akasaki, Amano, and Nakamura, and earned them the Nobel Prize in Physics

in 2014 for their contributions to energy-efficient white LED lighting and advances in digital displays.

All in all, LED technology has replaced traditional lighting technologies like CFL and incandescent bulbs across various applications. However, there is still a need to optimise and enhance this solid-state lighting technology in terms of improving its efficiency in converting electricity to light at high operating currents, extending its lifespan, and enhancing its reliability, while also reducing its size and production costs. Therefore, ongoing research and development efforts are focused on improving the performance and cost-effectiveness of LEDs, which will enable further advances in lighting and display technologies.

1.3 The Dawn of Organic Electronics

The invention of LEDs revolutionised the lighting industry and modern society is heavily reliant on this technology. However, LEDs do have some limitations, including their inability to be used as a pixel in high-resolution displays due to their large size, and their narrow band of wavelengths resulting in poor colour production. To address this, a "white light" LED is used as a backlight in displays [21, 22], with different types of phosphors converting monochromatic light to a desired spectrum [23]. This has led to the development of LCD technology, which utilises liquid crystals and colour filters to create pixels. However, this use of additional materials can make LED devices complex, large, and expensive to produce. To overcome these limitations, organic light-emitting diodes (OLEDs) have been developed. OLEDs are self-emissive and do not require additional materials like a backlight or phosphors to produce a broad spectrum of light. This results in simpler and smaller devices, making OLEDs an attractive alternative to LED-based displays.

Nowadays, OLED technology has matured and has been successfully employed in various commercial applications, ranging from small portable systems like smartphones and watches to large televisions and monitors. Compared to their LED counterparts, the latest OLED-based displays are considerably lighter, thinner, and smaller while delivering superior picture quality.

In fact, recent advancements in OLED fabrication techniques [24] have made it possible to create highly efficient displays on ultra-thin fibres, with diameters thinner than that of human hair. This paves the way for wearable displays and other innovative applications in the future.

While OLEDs offer numerous advantages over LEDs, they also have limitations that must be considered. One major drawback is their shorter lifespan compared to LEDs. Due to the organic materials used in their construction, OLEDs degrade faster over time, leading to decreased brightness and colour accuracy that make them unsuitable for long-lasting applications. This is because OLEDs are sensitive to moisture and oxygen, which can damage their performance over time [25]. As a result, the production environment for OLEDs must be highly controlled, which requires complex fabrication methods and increases production costs. Additionally, factors such as the roughness of metal electrodes, poor bonding between organic and inorganic layers, and migration of metal ions from electrodes into organic layers can significantly impact device efficiency and lifespan [26]. LEDs, on the other hand, are more robust and markets claim up to 50,000 hours or more [27], making them a better option for applications that require longevity and durability.

As briefly mentioned in the introduction, TADF is an emerging technology that holds promise in overcoming some of the challenges associated with OLED technology. But before delving into the intricacies of TADF OLED molecules, it is important to understand the origins and key concepts behind previous generations of OLED. Specifically, the first and second generations of OLED, based on fluorescence and phosphorescence, respectively, have paved the way for TADF OLED and will provide useful context for understanding the advancements made by this new technology.

1.3.1 Electroluminescence

The first report of electroluminescence (EL) using organic materials dates back to 1953 when Bernanose *et al.* [28]. A decade later, in 1962, Pope *et al.* demonstrated the first organic EL by observing fluorescence induced by a direct current (DC) to an anthracene ($C_{14}H_{10}$) crystal at high voltage (400 V) [29]. Although this initial research proved the possibility of EL with

organic materials, the low efficiency made it unattractive for practical applications until the groundbreaking work of C. W. Tang and S. A. VanSlyke in 1987 [30]. They achieved green ($\lambda \approx 550$nm) EL at low voltages (≈ 2.5 V) and high luminance (>1000 cd/m^2) at voltages below 10 V, making OLEDs a practical option for display devices. This breakthrough triggered widespread interest from both academia and industry, leading to the commercialisation of OLED displays we see today. Their device structure, consisting of two organic layers sandwiched between an anode and a cathode, was innovative at the time and is now widely used.

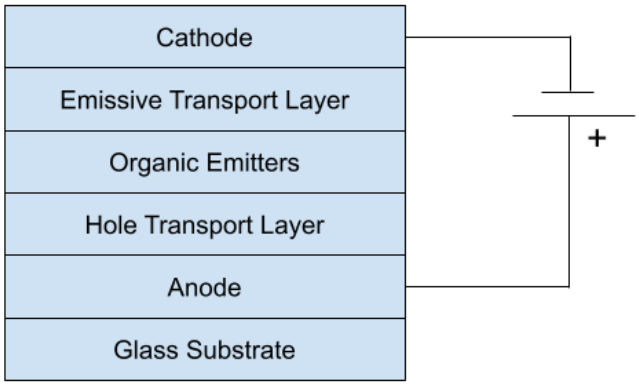


Figure 1.1: Diagram of a simple OLED device structure, featuring multiple layers sandwiched between an anode and cathode electrode, with a substrate layer providing structural support. The hole injection and electron injection layers enable the efficient emission of light when a voltage is applied.

To improve the efficiency of an OLED, it's important to understand the physical processes that occur within it. A basic OLED structure (Figure 1.1) consists of organic and/or inorganic layers sandwiched between the electron transport layer (ETL) and the hole transport layer (HTL) [31]. These layers are sandwiched between a metal cathode layer and a layer of transparent conductive material acting as the anode, such as indium tin oxide (ITO). When a voltage is applied across the sandwiched layers, charge carriers are injected into the structure. There are two types of charge carriers: a hole (positive charge) injected from the anode and an electron (negative charge) injected from the cathode. These charge carriers perturb the covalent (π) bonding structure of the organic molecules in the semiconductor, causing localised distortion [32]. This happens due to the balancing interactions between the charge carriers and the forces that hold the molecules in place.

The distorted position of the molecules creates a local polarisation centred around the charge

carriers, resulting in the formation of a polaron (P). A polaron is a quasiparticle that can be thought of as an electron or hole surrounded by a cloud of phonons. These polarons have a spin $\frac{1}{2}$ and they migrate through the organic layer to meet at the recombination zone [33, 34]. Initially, their spin orientation is random, making them uncorrelated. However, at the recombination zone, they become correlated and form an intermediate state held by their mutual Coulombic interaction [35, 36]. This intermediate state can be described as either a singlet or a triplet charge transfer (CT) state [37].

When polarons combine to form a CT state, they have a probability of either forming a singlet or a triplet state, depending on the specific molecular and electronic configuration. The S_1 state corresponds to the first singlet excited state with zero total electronic spin (S=0). It involves the promotion of an electron from the Highest Occupied Molecular Orbital (HOMO) to the Lowest Unoccupied Molecular Orbital (LUMO) or a higher unoccupied orbital.

On the other hand, the T_1 state represents the first triplet excited state with a total electronic spin equal to one (S=1). It involves the promotion of an electron from the HOMO to a higher unoccupied orbital while changing its spin orientation. Transitions involving triplet states are typically "forbidden" in optical spectroscopy due to spin selection rules.

Both S_1 and T_1 states are examples of excited states that depend on the molecular system's electronic structure. While S_1 states are more frequently observed and relevant for fluorescence and light emission, T_1 states are essential in understanding photochemical reactivity and excited-state dynamics. Molecules can have multiple excited states with different energy levels and spin multiplicities, contributing to their unique photophysical and photochemical properties. The investigation of these excited states is critical in photochemistry, photophysics, and the design of electronic materials.

Based on quantum mechanical spin statistics [38], there are four possible combinations of S_1 or T_1 states:

1. $\uparrow\uparrow T_1$
2. $\downarrow\downarrow T_1$

3. $\frac{1}{\sqrt{2}}[\uparrow\downarrow + \downarrow\uparrow]$ T_1, spin precessing in-phase

4. $\frac{1}{\sqrt{2}}[\uparrow\downarrow - \downarrow\uparrow]$ S_1, spin precessing out-of-phase

The probability of forming a singlet state is 25%, while the probability of forming a triplet state is 75%, resulting in a 1:3 ratio. Understanding these statistics is crucial in addressing the quantum limitations of OLEDs.

Regarding the development of OLED efficiency, it is essential to define two crucial parameters: EQE and IQE. The EQE refers to the ratio of the number of photons emitted by the device to the number of electrons injected into the device. It quantifies the overall efficiency of converting electrical current into emitted light from the OLED, taking into account both the internal and external factors that influence light extraction and outcoupling.

On the other hand, IQE specifically measures the efficiency of the internal processes within the OLED device. It represents the ratio of the number of excitons (electron-hole pairs) formed to the number of electrons injected into the device. IQE focuses solely on the quantum efficiency of exciton formation and recombination within the OLED's active layer, disregarding any losses due to light extraction or other external factors.

It is important to maintain consistency and avoid interchanging EQE and IQE throughout the manuscript to ensure clarity and accuracy in discussing the different aspects of OLED efficiency. By understanding the distinction between EQE and IQE, researchers can comprehensively assess the performance and optimise the efficiency of OLED devices.

1.3.2 Fluorescence-based OLED

Fluorescent OLEDs harvest singlet excitons, which are excited states of molecules with the same spin, resulting in light emission. However, they face limitations in their efficiency due to the under-utilisation of triplet excitons. These excitons are typically classified as spin-forbidden transitions, leading to a lower probability of emitting light compared to singlet excitons. This

reduced probability arises from selection rules that limit the radiative decay of triplet states. Consequently, the emission efficiency of triplet excitons is generally lower in fluorescent OLEDs.

Triplet excitons in fluorescent OLEDs predominantly undergo non-radiative decay, primarily thermal in nature, rather than efficiently emitting light. This non-radiative decay process results in the dissipation of a significant portion of the initially absorbed energy, up to 75%. Such energy losses significantly limit the overall efficiency of OLEDs. Figure 1.2 visually illustrates this non-radiative decay mechanism and its impact. As a result, the maximum IQE of

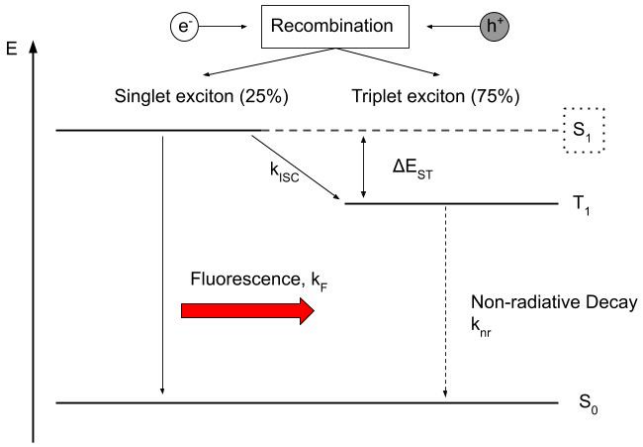


Figure 1.2: Jablonski diagram of the fluorescence mechanism in OLEDs. ISC, ΔE_{ST}, e-, h+ represent intersystem crossing (ISC), singlet-triplet energy gap ΔE_{ST}, electrons (e-) and holes (h+), respectively.

fluorescent-based OLEDs is limited to 25%, as any additional energy that is absorbed beyond this point is lost due to non-radiative decay.

In 1998, Baldo *et al.* [38] reported a breakthrough discovery that revolutionised OLED technology. By harvesting both singlet and triplet excitons using a red-emitting organometallic complex, phosphorescent dye 2,3,7,8,12,13,17,18-octaethyl-21H,23H-porphine platinum(II) (PtOEP), they achieved an impressive IQE of 23%, almost reaching the maximum limit of fluorescence-based devices. This pioneering work opened the door for research into phosphorescence-based OLEDs, ushering in a new era of more efficient and advanced lighting and display technologies.

1.3.3 Phosphorescence

Phosphorescent materials are a type of light-emitting compound that has received significant attention in recent years due to their ability to achieve high efficiency and long operational lifetimes in OLEDs. Unlike fluorescent materials, which can only utilise singlet excitons, phosphorescent materials are able to use both singlet and triplet excitons, which allows them to achieve higher IQEs. This was due to the introduction of heavy metal atoms, such as iridium (Ir) and platinum (Pt). The heavy metal in these organometallic complexes mediates a strong spin-orbit coupling that allows for intersystem crossing (ISC) (Figure 1.3) from S_1 to T_1 [8, 38, 9].

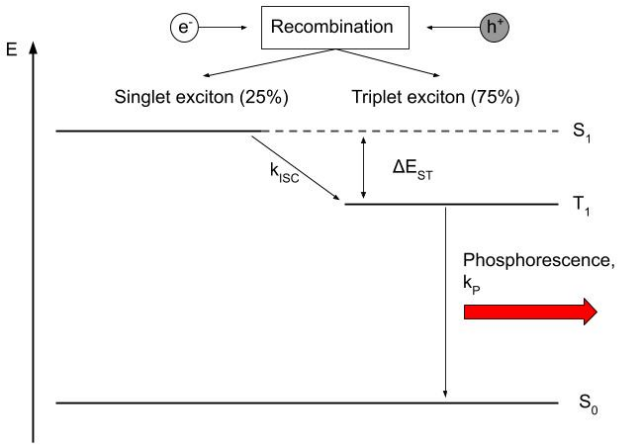


Figure 1.3: Jablonski diagram of the phosphorescence mechanism in OLEDs. ISC, ΔE_{ST}, e-, h+ represent intersystem crossing (ISC), singlet-triplet energy gap ΔE_{ST}, electrons (e-) and holes (h+), respectively.

Since 2001, there have been reports of devices with an IQE of almost 100% [10]. Although green and red-emitting cyclometalated iridium complexes are currently used in commercial OLED devices [39], blue-emitting complexes suffer from poor stability and performance during operation [40, 41]. As a result, phosphorescent-based blue emitters have not yet been commercialised. The poor performance of blue emitters is believed to be due to processes called triplet-triplet annihilation (TTA) and triplet-polaron annihilation (TPA), which reduce efficiency under high current density and degrade the device by creating highly energetic polarons. In addition to this, concerns exist over the use of heavy metal salt reagents, which are low in abundance and increase the overall cost of devices, as well as the potential for environmen-

tal contamination [11]. These factors make current phosphorescent OLEDs unsustainable and polluting.

Today, the challenge is to design OLED emitters that are not only cost-effective but also sustainable and environmentally friendly, without compromising on performance. Several strategies have been proposed to utilise the 75% and avoid the use of expensive metals, such as TTA [42], tuning spin-orbit coupling by side-stepping Kasha's rule [43], hybridized local and charge-transfer (HLCT) [44, 45, 46, 47, 48], and TADF [49, 50, 51]. Recent investigations have shown that TADF has made the most rapid progress among these strategies.

OLED emitters based on TADF properties show great promise in achieving almost 100% IQE, while also addressing the concerns associated with phosphorescent emitters. A TADF-based OLED is considered the third-generation OLED and is expected to be the successor to the current OLED technology.

1.4 Thermally Activated Delayed Fluorescence

A thermally activated delayed fluorescence (TADF) emitter is a type of light-emitting molecule that can use both singlet and triplet excitons, enabling it to achieve a maximum theoretical IQE of 100%, similar to a phosphorescent emitter. Unlike phosphorescent emitters, however, TADF emitters can be purely organic, which makes them more versatile and easier to integrate into various electronic devices.

The concept of TADF emitters has been around since 1961 [52, 53], when an organic chemical compound called 'eosin' was observed to emit delayed fluorescence in ethanol, leading to the term 'E-type' delayed fluorescence. From the 1970s to the 1990s, other organic compounds with this 'E-type' nature were discovered, such as benzophenone [54], 9,10-anthraquinone [55], aromatic thiones [56, 57], and thioketones [58].

The observation of delayed fluorescence in a Cu(I)-complex in 1980 marked the discovery of the first metal-containing TADF material [59]. This discovery expanded the range of materials

capable of exhibiting this phenomenon beyond the original group of materials referred to as 'E-type' delayed fluorescence and led to the adoption of the term TADF to encompass this broader range of materials.

Adachi *et al* conducted a series of studies on the use of TADF molecules in OLED devices in the late 2000s [60] and early 2010s [61][50]. Their research led to the discovery of highly efficient purely organic TADF molecules. In 2012, Adachi *et al* reported a class of TADF molecules based on carbazolyl dicyanobenzene (***CDCB***) with an exceptional external quantum efficiency (EQE) of 19.3% [50]. In contrast, the EQE of a typical inorganic (phosphorescent) OLED device typically ranges from around 20-25% [62]. This groundbreaking discovery sparked an enormous research effort worldwide to improve the efficiency of purely organic TADF emitters [63, 51, 12].

1.4.1 Key TADF Mechanisms

TADF consists of two distinct mechanisms: Prompt fluorescence (PF) and delayed fluorescence (DF), as seen in Figure 1.4).

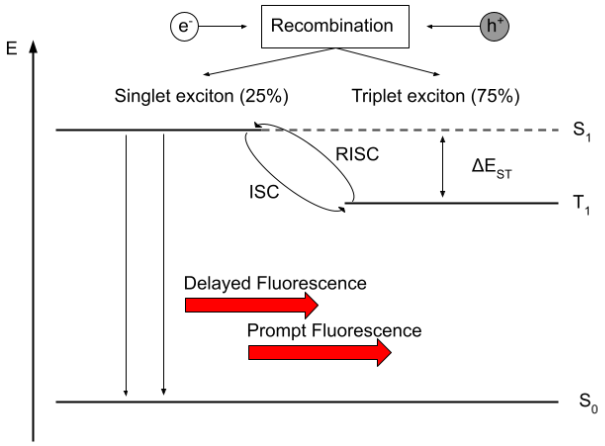


Figure 1.4: Jablonski diagram of the TADF mechanism in OLEDs. RISC, ISC, ΔE_{ST}, e-, h+ represent reverse intersystem crossing (RISC), intersystem crossing (ISC), singlet-triplet energy gap ΔE_{ST}, electrons (e-) and holes (h+), respectively.

The process of PF and DF can be simplified and explained with the following expressions:

$$S_1 \xrightarrow{k_{PF}} S_0 \tag{1.1}$$

$$S_1 \xrightarrow{k_{ISC}} T_1 \xrightarrow{k_{RISC}} S_1 \xrightarrow{k_{DF}} S_0 \tag{1.2}$$

In the expressions 1.1 and 1.2, 'k' represents the rate of the corresponding mechanism indicated in the subscript. The PF process described in expression 1.1 involves the decay of singlets to the ground state (S_0). The DF process represented by expression 1.2 entails initial ISC to T_1, followed by the re-population of S_1 via the reverse intersystem crossing (RISC) mechanism. TADF emitters can efficiently convert between the lowest T_1 and S_1 states due to the small singlet-triplet energy gap (ΔE_{ST}), which is typically < 0.1 eV. The thermal activation mechanism facilitates RISC, hence the name "thermally activated" delayed fluorescence.

According to the Boltzmann distribution relation [50], ΔE_{ST} dictates the rate of RISC (k_{RISC}):

$$k_{RISC} \propto \exp -\frac{\Delta E_{ST}}{k_B T} \tag{1.3}$$

where k_B is the Boltzmann constant and T is the temperature. A small ΔE_{ST} would result in a fast k_{RISC}.

The molecular energy of the lowest singlet (E_{S_1}) and triplet (E_{T_1}) excited states can be determined by the orbital energy (E), electron repulsion energy (K), and exchange energy (J) of the two unpaired electrons in these states, as shown in Equation 1.4 and 1.5 [63].

$$E_{S_1} = E + K + J \tag{1.4}$$

$$E_{T_1} = E + K - J \tag{1.5}$$

Although the singlet and triplet excited states of a molecule have the same electronic configuration, resulting in identical values of the electronic parameters E, K, and J, the energy of the lowest triplet excited state (E_{T_1}) is reduced (as shown in Equation 1.5) due to the same spin states of the unpaired electrons, while the energy of the lowest singlet excited state (E_{S_1}) is increased (as shown in Equation 1.4). Consequently, the energy difference between the two

states (ΔE_{ST}) is twice the exchange energy (J) (as shown in Equation 1.6).

$$\Delta E_{ST} = E_S - E_T = 2J \tag{1.6}$$

At the lowest S_1 excited state and lowest T_1 excited state, the two unpaired electrons are primarily located on the frontier orbitals, HOMO and LUMO. Despite their different spin states, these electrons have the same J value. Consequently, the exchange energy value for these electrons at the HOMO and LUMO can be determined using Equation 1.7 [64]:

$$J = \int\int \phi_H(r_1)\phi_L(r_2)\frac{1}{|r_2 - r_1|}\phi_H(r_2)\phi_L(r_1)dr_1dr_2 \tag{1.7}$$

where ϕ_H and ϕ_L represent wavefunctions of the HOMO and the LUMO, respectively. r_1 and r_2 represent position vectors. Equation 1.7 shows that a smaller (greater) overlap of the HOMO and the LUMO results in a smaller (greater) J.

Equations 1.4 and 1.5 demonstrate that while the exchange energy of electrons destabilises the S_1 state, it stabilises the T_1 state. As a result, in scenarios where the T_1 states are not solely determined by a single HOMO-to-LUMO electronic configuration, research has shown that the triplet state tends to have a more localised excitation character, while the singlet state exhibits more charge-transfer excitation character [65, 66].

Equation 1.7 indicates that ΔE_{ST} can be minimised by maximising the overlap between HOMO and LUMO. This would be true if the S_1 and T_1 states were solely defined by a HOMO-LUMO CT transition. However, in reality, the electronic configurations of T_1 states are often more complex, requiring a more detailed description than the simple HOMO-LUMO picture. In such cases, natural transition orbitals (NTOs) can provide a compact representation of the electronic excitations, particularly when the excited states correspond to a single pair of NTO orbitals for the hole and the electron [67, 63]. By calculating NTOs, it is possible to identify the orbitals that are involved in the electronic transitions that give rise to TADF. This can provide information on the energy level alignment between the donor and acceptor molecules, which is critical for efficient TADF [68].

Understanding the underlying mechanisms of spin-orbit coupling (SOC) and the spin-vibronic mechanism is crucial for optimising TADF materials and device performance.

SOC plays a significant role in TADF due to its influence on ISC and spin dynamics. SOC is the interaction between the electron spin and its orbital motion, resulting in the mixing of singlet and triplet states. In TADF systems, strong SOC facilitates the interconversion between these states, allowing for efficient triplet-to-singlet conversion and subsequent fluorescence emission. Therefore, a high SOC can enhance the TADF efficiency by facilitating the reverse intersystem crossing process. The rate of RICS (k_{RICS}) of a TADF material can be evaluated by using the equation 1.8 [69]:

$$k_{RICS} \propto \left| \frac{\langle S_1 | \hat{H}_{SOC} | T_1 \rangle}{\Delta E_{ST}} \right|^2 \tag{1.8}$$

Where $\hat{H}_{SOC}$ represents SOC operator, which describes the interaction between the electron spin and its orbital motion. $\langle S_1 | \hat{H}_{SOC} | T_1 \rangle$ | denotes the SOC matrix element between S_1 and T_1 states. Similar to equation 1.3, equation 1.8 also shows what reducing ΔE_{ST} can increase the k_{RISC} of TADF materials.

The spin-vibronic mechanism also contributes to the TADF process [70]. It involves the coupling of electronic excitations with vibrational modes within the molecule. Vibrational motions can modify the energy landscape and influence the rates of ISC, internal conversion, and fluorescence processes. In TADF materials, efficient spin-vibronic coupling enables an efficient population of the excited triplet state and subsequent RISC to the singlet state, leading to fluorescence emission. Optimising the spin-vibronic coupling strength and vibrational modes in TADF systems is essential for achieving high TADF efficiency.

The implications of SOC and the spin-vibronic mechanism for TADF extend beyond the understanding of fundamental processes. These factors guide the design and synthesis of TADF materials with tailored molecular structures, energy levels, and vibrational properties. By optimising SOC and spin-vibronic coupling, researchers can enhance the efficiency of TADF materials, reduce non-radiative decay pathways, and minimise efficiency roll-off at high current densities.

The equation equation 1.9 describes the coupling between the singlet and triplet states and involves the SOC and spin-vibronic interactions.

$$k_{RISC} = \frac{2\pi}{\hbar} \left| \frac{\langle {}^1\psi_{CT}| \hat{H}_{SOC} |{}^3\psi_{LE}\rangle \langle {}^3\psi_{LE}| \hat{H}_{vib} |{}^3\psi_{CT}\rangle}{\delta({}^3E_{LE} - {}^3 E_{CT})} \right|^2 \delta({}^3E_{LE} - {}^1 E_{CT}) \tag{1.9}$$

Where $\hbar$ is the reduced Planck's constant, $\langle {}^1\psi_{CT}|$ and $|{}^3\psi_{LE}\rangle$ represent the wavefunctions of the singlet CT state, $\langle {}^3\psi_{LE}|$ and $|{}^3\psi_{CT}\rangle$ represent the wavefunctions of the triplet CT and locally excited (LE) state, $\hat{H}_{vib}$ represents the vibronic coupling operator, which accounts for the coupling between the electronic excitations and vibrational modes within the molecule. $\delta({}^3E_{LE} - {}^3 E_{CT})$ and $\delta({}^3E_{LE} - {}^1 E_{CT})$ are the energy conservation conditions, ensuring that the energy differences between the involved states match. The equation quantifies the RISC rate based on the coupling strengths between the singlet and triplet states mediated by SOC and spin-vibronic interactions. It takes into account the energy conservation conditions for the involved electronic states.

Furthermore, the study of SOC and the spin-vibronic mechanism in TADF provides insights into the fundamental physics of exciton dynamics and electronic transitions in organic materials. This knowledge contributes to the broader field of organic optoelectronics and aids in the development of advanced materials for other applications, such as organic photovoltaics and sensors.

In summary, SOC and the spin-vibronic mechanism play crucial roles in the TADF mechanism. Understanding and optimising these factors are key to enhancing TADF efficiency and improving the performance of TADF-based OLEDs. Further research in this area will not only advance the field of TADF but also contribute to the fundamental understanding of exciton dynamics and pave the way for the development of next-generation organic optoelectronic devices.

Overall, the TADF mechanism provides an alternative route to designing OLEDs with high efficiency and has the potential to rival phosphorescent OLEDs, which are the current commercial standard for efficient OLEDs.

1.4.2 The Traditional Design Strategy for TADF molecules

The traditional design strategy for TADF molecules aims to minimise the key parameter ΔE_{ST} (equation 1.6). This is achieved by reducing the spatial overlap between the HOMO and LUMO frontier orbitals, resulting in the minimisation of exchange energy J. This approach enables the efficient conversion of triplets into singlets, as described by the Boltzmann distribution (equation 1.3).

Donor-acceptor (D-A) compounds are used to reduce spatial overlap between the HOMO and LUMO. In D-A compounds, the HOMO is mainly localized on the donor moiety, while the LUMO is situated on the acceptor moiety. However, a large spatial separation between the HOMO and LUMO frontier orbitals can lead to a strong intramolecular charge-transfer (ICT) character in the lowest singlet and triplet excited states. The strong ICT character of the excited states may not be compatible with this emission requirements [71], leading to lower photoluminescence quantum yields and reduced device performance. Intense emission from the lowest excited state is typically desired for optoelectronic applications. This requires a large transition dipole moment, which describes the strength of the interaction between the excited state and the surrounding environment, as well as a high oscillator strength, which reflects the probability of emission. Therefore, designing TADF molecules with appropriate donor and acceptor groups that balance the ICT character and emission properties is essential. The traditional design strategy extends to intramolecular D-A systems, which involve the use of a large steric hindrance structure or a twisted/spiro/bulky connection between the donor and the acceptor [63].

While there is a wide range of organic donors available for selection [63], N-containing aromatic compounds such as carbazole, diphenyl amine, phenoxazine, and their derivatives are commonly used as they exhibit the strong electron-donating ability and stable, high triplet states. However, to optimise the TADF emission strength, colour, and device performance, a variety of acceptor molecules are utilised in the design of donor-acceptor compounds.

Overall, the traditional D-A design principle offers great flexibility in tuning the structure and

properties of TADF molecules. It has proven to be a very successful strategy as there now exists hundreds of TADF emitters [12]. However, identifying the best TADF molecules can be a time and resource-intensive process that involves both experimental and computational techniques.

1.4.3 Beyond Donor and Acceptor Designs

Numerous studies have demonstrated the effectiveness of the donor-acceptor (DA) design in creating TADF molecules. Nevertheless, researchers are currently exploring alternative design strategies as well [72, 73, 74, 75, 76, 66, 77, 13].

In recent years, multiple-resonance (MR) emitters [72, 73, 74] have emerged as promising candidates for achieving high-efficiency TADF OLEDs. These emitters offer a unique molecular design strategy that goes beyond the conventional electron-donating and electron-accepting groups, enabling efficient interconversion of singlet and triplet excitons through the RISC process.

MR emitters incorporate diverse functional groups within the molecular structure, such as electron-rich nitrogen and electron-deficient boron or carbonylgroups, resulting in multiple resonance structures. These resonances facilitate effective energy transfer and promote efficient RISC between singlet and triplet states. By creating multiple energy minima and maxima along the potential energy surface, these emitters enable enhanced energy transfer pathways and increased harvesting of both singlet and triplet excitons, leading to improved overall TADF efficiency.

The design of MR emitters involves the careful integration of electron-donating and electron-accepting moieties, as well as the precise tuning of energy levels and molecular orbital distributions. This molecular engineering approach enables the optimisation of charge transfer abilities, energy level alignment, and molecular packing. The resulting MR emitters exhibit efficient reverse intersystem crossing, even with relatively small energy gaps between the singlet and triplet excited states.

Beyond donor and acceptor designs, MR emitters provide enhanced flexibility in tailoring the

emission properties and device performance. By exploring various combinations of functional groups and their spatial arrangements, researchers can fine-tune the energy levels and charge transfer characteristics of these emitters. This flexibility allows for the optimisation of TADF efficiency, reduction of efficiency roll-off at high current densities, and improvement of device stability.

The development of MR emitters represents a significant advancement in the field of TADF OLEDs. Their unique molecular design strategies and effective interconversion of singlet and triplet states offer a promising avenue for achieving highly efficient and stable TADF devices. By exploring beyond traditional donor and acceptor concepts, MR emitters open up new opportunities for advancements in organic optoelectronics and pave the way for future breakthroughs in efficient light-emitting devices.

In summary, while MR emitters often utilise a modified D-A structure, the "beyond" aspect refers to their incorporation of multiple resonance structures within the molecule, which allows for enhanced energy transfer and interconversion between singlet and triplet states.

Recently, Zhao and colleagues reported on a group of TADF molecules that depart from the traditional design rules [13], naming them "Type VII." These molecules pose a challenge as their small ΔE_{ST} cannot be attributed to a specific donor or acceptor moiety, suggesting that they were not designed using existing guidelines. In addition, the authors described six other types of TADF molecules, including Type I (known TADF molecules), Type II (molecules with common TADF donors and acceptors), Type III (molecules containing a novel donor or acceptor), Type IV (molecules with non-standard chemical topology), Type V (zwitterionic molecules), and Type VI (molecules consisting of only acceptors or donors and conjugated or non-conjugated linkers to form acceptor or donor molecules). This work by Zhao and co-workers provided inspiration and a foundation for this study.

Type VII molecules represent a novel pathway towards molecular design, as they deviate from traditional TADF design rules. Typically, these molecules have a larger HOMO-LUMO overlap and a slightly larger ΔE_{ST} than D-A molecules. However, Zhao and colleagues were able to enhance the properties of these molecules through their innovative design approach. For

instance, their investigation of ZERJEL02 led them to introduce additional nitrogen atoms into the peripheral aromatic units, which resulted in a remarkable reduction of ΔE_{ST} from 0.35 eV to 0.19 eV.

The discovery of "non-traditional" TADF molecules is intriguing and adds complexity and flexibility to an already rich, yet expensive, design process. To accelerate and improve this process, new strategies are needed. Researchers have turned to high-throughput virtual screening (HTVS) methods to design optoelectronic materials, which have significantly reduced both time and cost. With the existence of non-D-A TADF emitters and HTVS workflows, the prospect of combining these approaches is highly promising. This combination has the potential to identify novel TADF molecules that may have been overlooked using traditional design rules and in a cost and time-effective manner.

In the upcoming chapter 2, we will introduce the theory and methods of density functional theory that were utilised in our workflow.

Chapter 3 will provide an overview and discussion of the techniques employed to create and screen molecule libraries.

Moving on to Chapter 4, we will delve into a comprehensive analysis of the initial filtering process and an examination of the molecule library produced by our workflow.

Chapter 5 will present the outcomes of the DFT calculations performed on the candidate molecules, an extensive analysis of a select few final molecules, and a proposed design strategy for new TADF molecules.

Finally, in Chapter 6, we will conclude our work and highlight potential avenues for future research and development.

Chapter 2

Density Functional Theory

In computational chemistry, the electronic structure of molecules can be studied using both wavefunction-based methods and density functional theory (DFT). While it is true that wavefunction based methods, such as Hartree-Fock (HF) and post-HF methods (e.g., MPn, CCSD) [78, 79], are capable of providing very accurate results, it is important to note that DFT is not inherently less accurate than these methods. The accuracy of DFT depends on the choice of the exchange-correlation functional employed, and there are many functionals available with varying levels of accuracy.

On the other hand, DFT is a widely used method in computational chemistry that models the electronic structure of molecules by approximating the electron density as a function. Compared to wavefunction-based methods, DFT exhibits a favourable scaling with system size and is generally less computationally expensive. This enables the study of larger molecules and systems, as well as the execution of more extensive simulations.

Despite being an approximate method, DFT has been found to be quite accurate in many cases, and it has become a standard tool in computational chemistry. One of the advantages of using DFT is the well-known accuracy-cost trade-off. That is, by choosing an appropriate level of approximation, one can balance the accuracy of the results with the computational cost. This makes DFT a powerful and flexible tool for studying the electronic structure of molecules.

In summary, wavefunction-based methods are highly accurate but computationally expensive and limited to small systems, while DFT is less accurate but more computationally efficient and capable of handling larger systems. The accuracy-cost trade-off makes DFT an attractive choice for many computational chemistry applications.

In recent decades, DFT has emerged as a powerful computational tool in the fields of physics, chemistry, and materials science. Its ability to predict the properties of a system by using the electronic density (rather than the many-body wave function) as the primary variable has revolutionised our understanding of materials and their properties. However, to fully appreciate the value of DFT, it is important to consider its origins and the problem it is trying to solve.

The foundations of DFT can be traced back to the pivotal contributions of Hohenberg and Kohn in the 1960s. They proposed a theoretical framework [80] based on the proof that the electron density contains all the information required to describe the ground-state properties of many-electron systems. The Kohn-Sham approach [81], further developed this theory by introducing a set of non-interacting fictitious particles with an effective potential that reproduces the same electron density as the original interacting system. The Kohn-Sham approach has since become the standard method for performing practical DFT calculations, and its development has played a significant role in the advancement of modern DFT [82, 83, 84].

At its core, DFT aims to solve the problem of accurately predicting the electronic density of a system, which provides a complete description of its ground-state properties. It solves for the electron density rather than the wave function. To comprehend the fundamental principles of DFT, it is necessary to examine the basic formalism originating from the Schrödinger equation.

In this chapter, we will delve into the DFT methods that we utilise in our workflow. Additionally, we will introduce and discuss our force field optimisation approach. To ensure consistency in notation, we denote the total wavefunction as Ψ, which is expressed as the product of the electronic wavefunction Ψ_{elec} and the nuclear wavefunction Ψ_{nuc}.

2.1 The Schrödinger Equation

Quantum mechanics is concerned with determining the solutions to the Schrödinger equation, which describes the behaviour of quantum systems. In most cases, we work with the non-relativistic, time-independent Schrödinger equation (TISE).

$$\hat{H}\Psi = E\Psi \tag{2.1}$$

where $\hat{H}$ is the Hamiltonian, which is a differential operator which represents the total energy of the system, consisting of kinetic ($\hat{T}$) and potential ($\hat{V}$) energies (equation 2.2). Ψ, the total wavefunction is a set of solutions (or eigenstates) of the Hamiltonian and E is the eigenvalue.

$$\hat{H} = \hat{T} + \hat{V} \tag{2.2}$$

Each solution Ψ_n has an associated eigenvalue E_n.

The detailed definition of the Hamiltonian is dependent on the physical system that we aim to describe using the Schrödinger equation. Although there are well-known examples like the particle in a box or the harmonic oscillator, where the Hamiltonian takes on a simple form and can be solved exactly with a shorter equation, the systems we are interested in are typically more complex, such as molecules. In these systems, there are multiple nuclei interacting with numerous electrons, resulting in a more detailed Hamiltonian operator.

Consider a system comprising M nuclei and N electrons in the absence of a strong electric or magnetic field. To simplify matters, the Hamiltonian in equation 2.3 and in all future expressions are in terms of atomic units, where physical quantities are measured as multiples of fundamental constants. In this system, the mass of an electron m_e, the magnitude of its charge e, reduced Planck's constant $\hbar$, and the permittivity of free space $4\pi\epsilon_0$ are all set to unity (set to a value of one). By expressing these constants as multiples of the equation, we can simplify

the equation to the form shown in equation 2.3.

$$\hat{H} = -\frac{1}{2}\sum_{i=1}^{N}\nabla_i^2 - \frac{1}{2}\sum_{A=1}^{M}\frac{1}{m_A}\nabla_A^2 - \sum_{i=1}^{N}\sum_{A=1}^{M}\frac{Z_A}{r_{iA}} + \sum_{i=1}^{N}\sum_{j>1}^{N}\frac{1}{r_{ij}} + \sum_{A=1}^{M}\sum_{B>A}^{M}\frac{Z_A Z_B}{R_{AB}} \tag{2.3}$$

Where the indices A and B run over the M nuclei, while i and j denote the N electrons in the system. The nuclear charge of atom A and atom B is represented by Z_A and Z_B, respectively. This Hamiltonian contains information regarding the kinetic and potential energies of all particles of a system. The first two terms describe the kinetic energy of the electrons and the nuclei, respectively. The mass of nucleus A is represented by m_A and its charge by Z_A. The following three terms represent the potential energy of the system: attractive electrostatic interaction between the nuclei and electrons, repulsive electron-electron and nucleus-nucleus interaction, respectively. r represents the distance between an electron and another electron or the nucleus, while R represents exclusively the distance between the nuclei.

However, applying the Schrödinger equation to real-world problems can be challenging for several reasons. One of the main challenges is the complexity of the systems being modelled. These systems typically involve many interacting particles, which makes it difficult to solve the equation 2.3 analytically or numerically. This gives rise to what is known as a many-body problem, where the interactions between particles have to be considered simultaneously.

In a many-body problem, the behaviour of one particle is affected by the behaviour of all the other particles in the system. This results in complex and sometimes unexpected behaviour that is difficult to predict using the Schrödinger equation alone. As a result, additional methods and approximations are often required to obtain accurate solutions for these types of systems.

One approach to simplify the Schrödinger equation is to use the Born-Oppenheimer approximation, which was introduced in 1927 by Born and Oppenheimer [85]. We will discuss this approximation in the next section.

2.2 The Born-Oppenheimer Approximation

The Born-Oppenheimer approximation [85] simplifies the problem of studying a system by separating the electronic and nuclear motions. This is because the mass difference between electrons and nuclei is at least three orders of magnitude. Therefore, electrons move much faster than nuclei for the same amount of kinetic energy, E_{KE}:

$$velocity = \sqrt{\frac{2E_{KE}}{mass}} \tag{2.4}$$

This approximation assumes that the electrons' motion is significantly faster than that of nuclei, allowing for separate treatment of electronic and nuclear degrees of freedom. The electronic part can be solved first by considering the nuclei as fixed, using the clamped nuclei approximation (2.5).

As the nuclei's positions are changed incrementally, the electronic part is solved repeatedly. This process maps out the electronic energy as a function of the nuclei coordinates, producing a potential energy surface (PES). The nuclear part is then solved separately, using the PES calculated from the electronic part. The separation of this problem into nuclear and electronic parts is known as the Born-Oppenheimer approximation and it simplifies the Hamiltonian equation 2.3 to equation 2.5.

$$\hat{H} = -\frac{1}{2}\sum_{i=1}^{N}\nabla_i^2 - 0 - \sum_{i=1}^{N}\sum_{A=1}^{M}\frac{Z_A}{r_{iA}} + \sum_{i=1}^{N}\sum_{j>1}^{N}\frac{1}{r_{ij}} + Const. \tag{2.5}$$

In the kinetic term for nuclei reduces to 0 at the positions of nuclei are fixed, while positions of election can vary. Additionally, since the nuclei are static, the nuclear-nuclear repulsion does not change, therefore, it stays constant. These two terms can be ignored when only focusing on the electronic part of the problem: The final term *const.* represents any constant term that does not depend on the positions of the electrons and nuclei. It may include contributions from nuclear-nuclear repulsion, nuclear kinetic energy, or other constant terms. When only focusing

on the electronic part of the problem, the Hamiltonian reduces to equation 2.6

$$\hat{H}_{elec} = -\frac{1}{2}\sum_{i=1}^{N}\nabla_i^2 - \sum_{i=1}^{N}\sum_{A=1}^{M}\frac{Z_A}{r_{iA}} + \sum_{i=1}^{N}\sum_{j>1}^{N}\frac{1}{r_{ij}} = \hat{T} + \hat{V}_{Ne} + \hat{V}_{ee} \tag{2.6}$$

Where $\hat{T}$ is the kinetic energy of electrons, $\hat{V}_{Ne}$ is the electrostatic nuclei-electrons attraction and $\hat{V}_{ee}$is electrostatic repulsion between the electrons. the electronic Hamiltonian (equation 2.6) is a powerful tool for describing the electronic structure and dynamics of many-body systems with interacting electrons. This includes a wide range of physical systems such as molecules, solids, surfaces, and nanostructures. By providing a comprehensive description of the electronic behaviour of these systems, the electronic Hamiltonian enables the prediction and understanding of their properties and behaviour, making it a central concept in both theoretical and experimental studies of condensed matter physics and materials science.

The solution to the Schrödinger equation 2.7, with an electronic Hamiltonian $\hat{H}_{elec}$ is the electronic wave function Ψ_{elec}, with the eigenvalues E_{elec}:

$$\hat{H}_{elec}\Psi_{elec}(\mathbf{r}_1, \ldots, \mathbf{r}_N) = E_{elec}\Psi_{elec}(\mathbf{r}_1, \ldots, \mathbf{r}_N) \tag{2.7}$$

The electronic wave function depends only on the electron's spatial coordinates, although a complete description must include the electron spin, up or down.

The wave function solution Ψ_{elec} that corresponds to the lowest energy state E_{elec} is commonly referred to as the ground state. The eigenvalue of this ground state wavefunction is known as the ground state energy. However, accurately determining an exact solution for Ψ_{elec} that describes a system beyond trivial cases is generally impossible. This is due to the complexity of the underlying many-body problem, which scales exponentially with the number of electrons and atomic orbitals in the system.

The term "trivial" refers to simple or easily solvable systems. For example, a system with only two electrons and two atomic orbitals would be considered trivial, as the calculation of the wave function and energy of the ground state would be straightforward. However, for more complex systems, such as those with many interacting particles and orbitals, obtaining an exact solution

becomes much more difficult and may not be possible with current computational resources.

The complexity of the many-body problem in electronic structure calculations scales exponentially with the number of electrons and atomic orbitals in the system. This makes it computationally infeasible to obtain an exact solution for the wave function Ψ_{elec} for even moderately sized systems.

The exponential scaling is primarily due to the $\hat{V}_{ee}$ term in $\hat{H}_{elec}$ (equation 2.6), which represents the electrostatic repulsion between all pairs of electrons in the system. The computational cost of this term scales as $O(N_{elec}^2)$, where N_{elec} is the number of electrons in the system. In addition to the $\hat{V}_{ee}$ term, other factors such as the size of the basis set used to expand the wave function also contribute to the scaling of the electronic structure problem.

As a result of the scaling considerations and the complexity of the many-body problem, various approximations and methods have been developed to obtain accurate solutions for the electronic structure of molecular, solid-state, and nanostructured systems. These methods often involve a combination of analytical approximations, numerical techniques, and computational algorithms that exploit various features of the system, such as symmetries and sparsity of the Hamiltonian matrix, to reduce the computational cost.

The Hartree product approach is a widely used analytical approximation for the many-body wave function. It assumes that the wave function can be expressed as a product of one-electron wave functions, where each electron is treated independently, ignoring electron-electron interactions. Although this approximation reduces the computational cost of the electronic structure problem, it neglects the correlation between electrons and can lead to inaccurate predictions for systems with strong electron-electron interactions. Introducing the Hartree product sets the foundation for understanding how electron correlation is treated in electronic structure calculations and highlights the motivation for developing more advanced methods such as DFT.

2.3 The Hartree Product

In order to overcome the many-body problem, it is necessary to develop an approximation of the true wave function. However, it is important to remember that the wavefunction cannot be directly observed. This is a fundamental concept in quantum mechanics and highlights the need for indirect methods, such as measuring the probabilities of various outcomes, to study quantum systems. Instead, quantum mechanics provides us with a framework to calculate the probability of various outcomes. For example, we can calculate the probability that there are N electrons at a particular set of coordinates $\mathbf{r}_1, \ldots, \mathbf{r}_N$. This probability is given by:

$$|\Psi(\mathbf{r}_1, \ldots, \mathbf{r}_N)|^2 = \Psi^*(\mathbf{r}_1, \ldots, \mathbf{r}_N)\Psi(\mathbf{r}_1, \ldots, \mathbf{r}_N) \tag{2.8}$$

Where Ψ^* denotes the complex conjugate. It is also possible and useful to approximate $\Psi(\mathbf{r}_1, \ldots, \mathbf{r}_N)$ as a product of individual wave functions:

$$\Psi(\mathbf{r}_1, \ldots, \mathbf{r}_N) \cong \psi_1(\mathbf{r})\psi_2(\mathbf{r}) \ldots \psi_N(\mathbf{r}) \tag{2.9}$$

This approximation is known as the Hartree product.

In summary, the Hartree product is a popular method for approximating the many-electron wave function in quantum mechanics. The Hartree product represents the many-electron wave function as a product of one-electron wave functions (orbitals), and it is used to calculate the electron density, which describes the distribution of electrons in space. The electron density (equation 2.10) is related to the Hartree product because it is calculated based on the probability density of finding electrons at specific coordinates in space, which is represented by the one-electron wave functions in the Hartree product. Therefore, the Hartree product provides a framework for calculating the electron density and gaining insights into the electronic structure and properties of molecules, atoms, and solids.

2.3.1 Limitations of Hartree Product

The Hartree product, which represents the many-electron wavefunction as a product of individual one-electron wavefunctions (equation 2.9), has several limitations:

Neglects Electron Correlation: The Hartree product does not account for electron-electron correlation, which is a fundamental aspect of the behaviour of electrons in many-electron systems. Electron correlation refers to the interactions between electrons that cannot be captured by simply considering their independent motion. Neglecting electron correlation can lead to inaccurate descriptions of electronic properties, especially for systems with strong electron-electron interactions.

Violates Pauli Exclusion Principle: The Hartree product does not satisfy the Pauli exclusion principle, which states that two identical fermions (e.g., electrons) cannot occupy the same quantum state simultaneously. As a result, the Hartree product may not correctly describe the antisymmetric nature of the many-electron wavefunction required for fermionic particles.

Not Variationally Stable: The Hartree product is not a variational wavefunction, meaning that it does not necessarily provide an upper bound to the true ground-state energy. Variational stability is a desirable property in wavefunctions used in quantum mechanics, as it ensures that the energy obtained from the wavefunction is always an upper bound to the exact ground-state energy.

Limited Applicability: The Hartree product is typically used as an initial trial wavefunction in variational calculations to obtain better approximations to the true wavefunction. However, as the sole wavefunction for many-electron systems, it has limited applicability for accurately describing complex electronic structures and properties.

Due to these limitations, the Hartree product is not a fully satisfactory description of the many-electron wavefunction, especially for systems with strong electron-electron interactions and significant correlation effects. More sophisticated methods, such as coupled cluster theory [86], density functional theory [80, 81], or post-Hartree-Fock approaches [87], are employed to address these limitations and provide more accurate descriptions of electronic structures and

properties in many-electron systems.

2.4 The Electron Density

In experiments electrons are indistinguishable, thus cannot label them. However, we can measure the probability that any order of a set of N electrons are the coordinates $\mathbf{r}_1$ through $\mathbf{r}_n$. The electron density $\rho(\mathbf{r})$ is closely related as it can be calculated by the following equation:

$$\rho(\mathbf{r}) = 2\sum_i \psi_i^*(\mathbf{r})\psi_i(\mathbf{r}) \tag{2.10}$$

where $\psi_i(\mathbf{r})$ represents individual electron wavefunction located at position $\mathbf{r}$. According to the Pauli exclusion principle, no two electrons in an atom or molecule can have the same set of quantum numbers, which includes their spin states. As a result, each electronic state can be occupied by at most two electrons with opposite spins. The factor of 2 arises from this spin degeneracy. This means that the electron density of a system depends not only on the positions of the electrons but also on their spin states, which can affect the spatial distribution of the electrons.

The electron density, $\rho(\mathbf{r})$, of a system is a function of only three coordinates, but it contains a significant amount of information that is observable from the full wavefunction, which is a function of 3N coordinates, where N is the number of electrons in the system. The electron density provides a way to visualise the distribution of electrons in the system without having to consider the entire wavefunction. In this way, electron density is a powerful tool for understanding the electronic structure and properties of molecules, atoms, and solids.

2.5 Foundations of Density Functional Theory

DFT is built upon the Hohenberg-Kohn theorems [80], which serve as the foundation of the theory. The first theorem states that the ground-state energy E from Schrödinger's equation

is a unique functional of the electron density $\rho(\mathbf{r})$:

$$E[\rho(r)] = E \tag{2.11}$$

In other words, the ground state energy of a system can be expressed as a functional of the electron density. Hence, this theory is known as the density functional theory.

The second theorem states that the electron density that minimises the energy of the overall functional is the true electron density, corresponding to the full solution of the Schrödinger equation.

The Hohenberg-Kohn theorems provide a way to simplify the Schrödinger equation by defining a function of just three variables, the electron density $\rho(\mathbf{r})$, instead of a function of 3N variables (the wavefunction). This reduction greatly simplifies the computational problem, making DFT a powerful tool for predicting the electronic structure and properties of various systems. The true ground state density of the system is the density that minimises the total energy.

It is useful to write the functional described by the Hohenberg-Kohn theorems in terms of single electron wave functions:

$$E[\{\psi_i\}] = E_{known}[\{\psi_i\}] + E_{XC}[\{\psi_i\}] \tag{2.12}$$

It is composed of two major parts: "known" and "eXchange-Correlation" (XC).

$$E_{known}[\{\psi_i\}] = -\sum_i \int \psi_i^* \nabla^2 \psi_i d^3r + \int V(\mathbf{r})\rho(\mathbf{r})d^3r + \frac{1}{2}\iint \frac{\rho(\mathbf{r})\rho(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} d^3r d^3r' + E_{n-n} \tag{2.13}$$

The terms in equation 2.13 are defined as: the non-interacting fictitious particles, elections-nuclei Coulomb interaction, electron-electron pair Coulomb interaction and internuclear interactions, respectively.

While $E_{XC}[\{\psi_i\}]$ includes all quantum mechanical effect not included in "known terms".

2.5.1 Kohn Sham Formalism of DFT

The Kohn-Sham (KS) formalism [81] is a widely used approach in DFT to calculate the electronic structure of molecules, atoms, and solids. It was introduced by Walter Kohn and Lu Sham in 1965 as a way to simplify the problem of finding the ground-state electron density of a system. In this approach, the motion of correlated and interacting electrons under an external potential is modelled as fictitious non-interacting particles moving under an effective potential called the Kohn-Sham potential. The eigensolutions of the Hamiltonian of this modelled system are known as the Kohn-Sham orbitals, which are orthonormal to each other. The orthonormality of the KS orbitals is important because it is a requirement for the eigenstates of a Hermitian operator of the electron density. This property is critical for the accuracy of DFT calculations since the electron density is used to calculate many physical and chemical properties of a system.

The electronic charge density in Kohn-Sham DFT (KS-DFT) is defined as:

$$\rho(r) = \sum_i \phi_i^*(\mathbf{r})\phi_i(\mathbf{r}) \tag{2.14}$$

where $\phi(\mathbf{r})$ represents the KS orbital of the i-th eigenstate and the summation runs over all the occupied states. The electronic charge density is a function of position in space, $\mathbf{r}$, and describes the distribution of electrons in the system.

Kohn and Sham demonstrated that obtaining an accurate electron density involves solving a set of equations, known as the KS equations [81], which are constructed in a way that ensures that the electron density yields the minimum total energy of the electronic system in its ground state:

$$\left[\frac{1}{2}\nabla^2 + V(\mathbf{r}) + V_H(\mathbf{r}) + V_{XC}(\mathbf{r})\right]\psi_i(\mathbf{r}) = \epsilon_i\psi_i(\mathbf{r})) \tag{2.15}$$

$$V_H(\mathbf{r}) = \int \frac{\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d^3r' \tag{2.16}$$

$$V_{XC}(\mathbf{r}) = \frac{\delta E_{XC}(\mathbf{r})}{\delta\rho(\mathbf{r})} \tag{2.17}$$

Where $V(\mathbf{r})$ is the potential due to the interaction between an electron and a collection of atomic nuclei (appeared in full Schrödinger equation 2.7). $V_H(\mathbf{r})$ is the Hartree potential (equation 2.16) which describes the Coulomb repulsion between the electron being considered in one of the KS orbitals and the total electron density defined by all electrons in the problem. As a result, the Hartree potential includes a self-interaction contribution because the electron being described in KS equations is also part of the total electron density. The exchange-correlation $V_{XC}(\mathbf{r})$ (equation 2.17) is the potential that defined the exchange and correlation contributions to the single-electron equations. It is defined as a functional derivative of the exchange-correlation energy with respect to electron density.

The equation 2.15 is superficially similar to $\hat{H}_{elec}$ (equation 2.3). The main difference is that the KS equations are missing the summations inside the full Schrödinger equation. This is because the solutions of KS equations are single electron wave functions that depend only on three spatial variables.

So far, there might be a sense that the KS equations are circular. To recap, solving KS equations requires Hartree potential. Hatree potential requires knowing the electron density. Electron density requires knowing the single electron wave functions and solving this requires solving the KS equations. To break this cycle, the problem is solved iteratively:

1. Define an initial trial electron density $\rho(\mathbf{r})$ which is typically determined using some approximation or heuristic method based on prior knowledge or assumptions about the system being studied.

2. Solve the KS equations defined using the trial electron density to find the single-particle wave functions $\psi_i(\mathbf{r})$

3. Calculate the electron density defined by the Kohn-Sham single-particle wavefunctions from step 2
   $\rho_{KS}(\mathbf{r}) = 2\sum_i \phi_i^*(\mathbf{r})\phi_i(\mathrm{r})$

4. Compare the calculated electron density $\rho_{KS}(\mathbf{r})$ with the trial electron density used to solve the KS equations $\rho(\mathbf{r})$. If these two densities, $\rho_{KS}(\mathbf{r})$ and $\rho(\mathbf{r})$, are equal or nearly

equal, this is the ground state electron density. Otherwise, update the trail electron density and start again at step 2.

2.5.2 Exchange Correlation Functional

The Exchange-Correlation (XC) functional is a critical component of DFT, as it accounts for the effects of electron correlation and exchange that are fundamental to understanding many properties of matter. However, the true form of the XC functional is generally unknown and must be approximated to solve the KS equations. Finding a good approximation for the XC functional is one of the most challenging aspects of DFT. This is because developing accurate approximations involves balancing the need for accuracy with the need for computational efficiency, and there are many different approximations that have been proposed, each with its own strengths and weaknesses. Therefore, finding a good approximation for a particular system often requires significant trial and error, as well as a deep understanding of the underlying physics.

XC functionals are approximations of the true functional and come in various types. These functionals are classified into families based on specific characteristics, creating distinct groups that can be likened to the rungs of "Jacob's ladder". As we climb each rung, we approach closer to achieving chemical accuracy, as Perdew explains [88]. Some examples of XC functionals include the local density approximation (LDA), generalised gradient approximation (GGA) [89], meta-generalised gradient approximation (meta-GGA) [90], hybrid and double hybrid functionals [91]. In most XC functionals, the exchange and correlation parts are formulated separately from each other [92]. This separation makes it easier to design and incorporate different types of effects into the functional, such as the impact of the molecule's environment or the presence of external fields. By formulating the exchange and correlation parts separately, researchers can more easily modify and tune the functional to better represent the electronic properties of different systems.

In addition, there exists Range-separated functionals (RSFs) [93, 92] which are a type of XC functional that divide the calculation of exchange and correlation energies into two regions,

short-range and long-range. These regions are separated by a range-separation parameter that determines the distance at which the exchange and correlation contributions switch from short-range to long-range behaviour. They are designed to address the limitations of traditional XC functionals, which often struggle to accurately capture both short-range and long-range interactions simultaneously. By separating the calculation into two distinct regions, range-separated functionals can more effectively describe both types of interactions and improve the accuracy of electronic structure calculations.

RSFs have been used in the study of TADF materials to accurately calculate the energy levels and excited-state properties of the organic molecules involved [94]. Specifically, these functionals have been shown to improve the accuracy of calculations of excited-state properties, such as singlet and triplet energies, transition dipole moments, and excited-state geometries. However, RSFs need to be optimised for specific systems and applications. The optimal range separation parameter depends on the electronic structure and properties of the system being studied. The range separation parameter determines the distance at which the functional switches from using a semi-local functional to using a non-local functional for the exchange contribution.

There are many different types of XC functionals [93, 92], each with its own strengths and limitations, and choosing the best functional for a specific system can be a challenging task. The choice of functional can significantly impact the accuracy of the calculated electronic properties, such as molecular structure, reactivity, and spectroscopic properties.

While there are some general guidelines and benchmarks [93, 92, 94] for choosing an appropriate functional, the selection process often involves a degree of trial and error, and experience can be a valuable guide in making an informed decision. Researchers may rely on previous studies of similar systems, as well as their own knowledge and intuition, to select a functional that is likely to yield accurate results.

Overall, while the selection of a suitable functional is a complex and challenging task, experience and careful consideration of the system and the available options can aid in the decision-making process.

Local Density Approximation

The LDA is the simplest XC functional (equation 2.18) and assumes that the exchange-correlation energy of an electron is solely determined by the electron density at its position.

$$E_{LDA}(\rho) = \int \epsilon_{xc}\{\rho(\mathbf{r})\}d^3\mathbf{r} \tag{2.18}$$

LDA offers several advantages that make it a popular choice in many applications. Firstly, LDA is known for its simplicity and computational efficiency. It provides a reasonable description of electronic structures while keeping the computational costs relatively low, making it accessible for studying a wide range of systems. Additionally, LDA exhibits a systematic behaviour that is particularly suited for certain types of systems. It can capture trends and qualitative features in electronic properties, making it valuable for investigating bulk materials and solids. Moreover, LDA is known to provide accurate estimates for the binding energies of atoms and molecules. It effectively describes the energetics of chemical reactions and processes involving weakly bound systems.

However, despite its advantages, LDA does have certain limitations [95] that need to be considered. One limitation is the lack of spatial and non-local information. LDA assumes that the exchange-correlation energy depends solely on the local electron density, neglecting the density gradient and other non-local contributions. As a result, it fails to capture spatial variations and non-local effects, limiting its accuracy in systems where such effects are important. Another limitation is the underestimation of dispersion forces, which are critical for accurate predictions of intermolecular interactions. LDA tends to underestimate these forces, leading to incorrect predictions of molecular geometries and properties in systems governed by weak intermolecular forces.

Furthermore, LDA often underestimates band gaps in semiconductors and insulators, which affects its ability to accurately predict electronic excitations and optical properties. The delocalisation and non-local character of excited states are not properly captured by LDA, making it less suitable for studying optical and electronic transitions. Additionally, LDA struggles

to describe strongly correlated systems, such as transition metal compounds or systems with strong electron-electron interactions. It does not account for correlation effects beyond the mean-field level, limiting its accuracy in such cases.

Considering these advantages and limitations is crucial when selecting an appropriate level of theory for a given system or property of interest. While LDA has its merits, it is important to recognise its limitations, especially in systems where non-local effects, dispersion forces, accurate band gaps, or strong correlations play a significant role.

Generalised Gradient Approximation Functional

GGA functionals [89] are a more sophisticated approximation that takes into account the gradient of the electron density (equation 2.19). The inclusion of the gradient allows GGA functionals to capture the variations in electron density, which are crucial in describing bonding, non-local interactions, and other electronic properties. The gradient accounts for how the electron density changes as one moves from one point in space to another. This information is essential for systems with rapidly varying electron densities, such as molecules, surfaces, and interfaces.

By taking into account the gradient of the electron density, GGA functionals can provide more accurate descriptions of molecular properties, such as bond lengths and bond angles. They also offer improved treatment of weak interactions, such as van der Waals forces, which are sensitive to the spatial variations in the electron density.

$$E_{GGA}(\rho) = \int \varepsilon_{xc}\{\rho(\mathbf{r}), |\nabla\rho(\mathbf{r})|\}d^3\mathbf{r} \tag{2.19}$$

GGA functionals exhibit notable benefits in terms of energetic properties and reaction energies. They generally yield improved accuracy in predicting atomisation energies, reaction energies, and reaction barriers. By incorporating the electron density gradient, GGA functionals capture important energetic contributions, leading to more reliable results in a wide range of chemical reactions.

However, it is important to recognise the limitations of GGA functionals [96]. They face challenges in accurately describing systems with strong electron correlation, such as transition metals and strongly correlated materials. These systems often require more sophisticated methods beyond the GGA level, such as hybrid functionals or techniques incorporating explicit correlation effects. Additionally, GGA functionals typically underestimate the strength of weak intermolecular interactions, such as van der Waals forces and hydrogen bonding. Although they partly account for dispersion interactions, their treatment is less accurate compared to dedicated dispersion-corrected functionals or advanced approaches specifically designed for these interactions.

Moreover, the performance of GGA functionals is sensitive to the chosen approximation schemes. Different GGA functionals can yield varying results for the same system, highlighting the sensitivity to the functional used. It is essential to carefully consider the advantages and limitations of GGA functionals when employing them in DFT calculations and to select an appropriate level of theory based on the system of interest and the desired level of accuracy.

In summary, GGA functionals provide valuable improvements over the LDA, particularly in describing molecules, surfaces, and energetic properties. However, they have limitations in capturing strong electron correlation effects and accurately describing weak intermolecular interactions. Awareness of these advantages and limitations is crucial for informed decision-making when employing GGA functionals in DFT studies.

Meta Generalised Gradient Approximation

Meta-GGA functionals [90] are an extension of GGA functionals in DFT, offer several advantages and exhibit certain limitations. In the context of electronic structure calculations, it is essential to understand the strengths and weaknesses of meta-GGA functionals for their successful application in various systems.

One of the primary advantages of meta-GGA functionals is their improved treatment of strong electron correlation effects. By incorporating not only the electron density and its gradient

but also higher density derivatives or the kinetic energy density, meta-GGA functionals go beyond the GGA level of approximation. This additional information enables a more accurate description of electronic systems with complex electronic structures, such as transition metal complexes, open-shell systems, and molecules with significant charge transfer.

Meta-GGA functionals also show enhanced accuracy in capturing weak interactions, such as van der Waals forces and hydrogen bonding. The inclusion of higher-density derivatives in the functional formulation allows for a better description of the spatial variations in electron density, resulting in improved predictions of intermolecular interactions. This is particularly relevant in studying biomolecular systems, molecular aggregates, and supramolecular assemblies.

Despite their advantages, meta-GGA functionals are not without limitations. The increased computational cost associated with meta-GGA calculations compared to standard GGA functionals can be significant. The inclusion of higher-density derivatives requires more complex calculations, which can be a challenge for large systems or high-throughput studies. As a result, careful consideration of the balance between accuracy and computational efficiency is necessary when employing meta-GGA functionals in research.

Another limitation is related to the difficulty in designing universally applicable meta-GGA functionals. Different systems may require different levels of information, and a single meta-GGA functional may not be optimal for all cases. The performance of meta-GGA functionals can vary depending on the nature of the electronic system being studied, necessitating the development of specialised functionals for specific applications.

Overall, meta-GGA functionals offer significant advantages in treating strong electron correlation effects and capturing weak interactions, making them valuable tools for a wide range of electronic structure calculations. However, their increased computational cost and system-dependent performance require careful consideration and further advancements in functional development to fully exploit their potential.

Range-Separated Functional

Range-separated functionals have emerged as a powerful class of density functionals within DFT, offering several advantages in the description of electronic properties. These functionals divide the electron-electron interaction into short-range (SR) and long-range (LR) components, treating them separately. By incorporating two distinct range-dependent exchange-correlation functionals, range-separated functionals provide enhanced accuracy in different electronic regimes. The transition from the SR to the LR behaviour is achieved by "crossfading" using an error function of the interelectronic distance ($\mathbf{r}_{12}$):

$$\frac{1}{\mathbf{r}_{12}} = \frac{1-[\alpha+\beta erf(\omega \mathbf{r}_{12})]}{\mathbf{r}_{12}} \frac{\alpha+\beta erf(\omega \mathbf{r}_{12})}{\mathbf{r}_{12}} \tag{2.20}$$

Where the first and second terms represent SR and LR respectively in equation 2.20. In this equation, ω is the range separation parameter, which can be determined empirically using a training set [97, 98, 99, 100, 101] or by minimising the deviation from the conditions that the exact KS functional must satisfy. The parameters α and β represent the percentages of HF exchange in the short-range and long-range limits, respectively.

The range-separated functional approach offers a balanced treatment of SR and LR electron interactions, enhancing the accuracy of DFT calculations for a wide range of systems. The choice of ω and the values of α and β are critical in achieving reliable and physically meaningful results with range-separated functionals.

One of the key advantages of range-separated functionals is their ability to capture both short-range and long-range electron-electron interactions appropriately. This feature allows for a more accurate description of electronic properties, particularly for systems with significant charge transfer or strongly correlated electrons. By treating SR interactions with a more accurate functional and LR interactions with a different functional, range-separated functionals can better capture the intricate interplay between localised and delocalised electronic behaviour.

Furthermore, range-separated functionals offer improved treatment of charge transfer excita-

tions, such as those occurring in donor-acceptor complexes or charge transfer states in organic semiconductors. The long-range component of these functionals, often modelled based on a Coulombic interaction, effectively accounts for the electron-hole interaction and captures charge transfer phenomena more accurately than traditional functionals. This makes range-separated functionals particularly relevant for the study of charge transport and optoelectronic properties of materials.

However, range-separated functionals also have certain limitations. Choosing appropriate range-separation parameters such as ω can be challenging, as the accuracy of the functional is sensitive to their values. The range-separation parameters must be carefully tuned for different systems to achieve the best compromise between short-range and long-range interactions. Inaccurate parameter selection can lead to erroneous results and compromise the reliability of the calculations.

Additionally, range-separated functionals can be computationally demanding compared to traditional functionals. The separate treatment of short-range and long-range components often involves more complex calculations and additional computational resources. Therefore, careful consideration of the trade-off between computational cost and accuracy is necessary when applying range-separated functionals in large-scale or time-sensitive calculations.

In summary, range-separated functionals provide improved accuracy in capturing both short-range and long-range electronic interactions, making them suitable for systems with charge transfer and strong electron correlation effects. They offer an enhanced description of charge transfer excitations and are particularly relevant for studying optoelectronic properties. However, their accurate application requires careful parameter selection, and their computational cost should be considered in large-scale simulations. Understanding the advantages and limitations of range-separated functionals is crucial for their effective utilisation in DFT studies.

Conclusion

Perdew-Burke-Ernzerhof (PBE) [102] is a widely used GGA functional in DFT. It was developed by Perdew, Burke, and Ernzerhof in 1996 and has become a popular choice for computing the electronic structures of a wide range of materials. The selection of the PBE functional for this work calculations was motivated by the specific objectives of the research, which involved employing a high-throughput virtual screening method for designing OLED molecules. Several factors influenced the choice, including the accuracy of the functional, its treatment of dispersion interactions, and computational efficiency.

The PBE functional has been widely utilised in computational chemistry and demonstrated good performance in describing the electronic structure and energetics of molecular systems. By incorporating both the electron density and its gradient, the PBE functional offers a more accurate representation of molecular properties, such as bond lengths, bond angles, and surface characteristics. Given that the project revolves around the design of OLED molecules, the precise prediction of these molecular properties is crucial for guiding the exploration of the chemical space.

Although the PBE functional does not explicitly include the kinetic energy density associated with dispersion interactions, it still provides an approximate treatment of these weak intermolecular forces. While dedicated dispersion-corrected functionals or range-separated functionals may offer more accurate treatments, their implementation would significantly increase computational costs. In the context of high-throughput virtual screening, computational efficiency is paramount to explore a large chemical space and identify promising candidates. Therefore, the PBE functional strikes a balance between accuracy and computational feasibility.

It is important to acknowledge the limitations of the chosen functional. The PBE functional may not accurately capture strong electron correlation effects or provide precise treatments of dispersion interactions. Thus, caution should be exercised when interpreting the results obtained using the PBE functional. It is advisable to validate the findings using alternative methods or experimental data to confirm the reliability of the outcomes.

In summary, the decision to use the PBE functional for the project calculations was based on its established accuracy for molecular systems, its approximate treatment of dispersion interactions, and its computational efficiency for high-throughput virtual screening. However, the limitations of the PBE functional should be recognised, and additional validation through alternative methods or experimental data is recommended to support and confirm the obtained results.

2.5.3 Solving Kohn Sham Equation

The previous sections introduced and discussed the Kohn-Sham (KS) formalism, which relies on the charge density that is expressed as the sum over squares of KS orbitals. Because the potential depends explicitly on the charge density, a self-consistent solution must be found. Two primary methods are typically used to solve the KS equation: the minimisation scheme and the mixing scheme.

Minimisation Scheme

Solving the Kohn-Sham equation involves minimising the energy in N-dimensional space. To begin, an initial guess for the KS orbitals is generated using a random or predetermined scheme. One possible predetermined scheme could be to use the orbitals from a previous calculation as the initial guess for the current calculation. Alternatively, a predetermined scheme could involve selecting an initial guess based on the symmetry properties of the system being studied. The estimated gradient is then used to guide the search for the optimal direction of minimisation, as illustrated in the equation 2.21 with P as the preconditioner and g_i as the gradient. The iteration continues until the gradient's norm falls below a specified threshold value.

$$\phi_i = \phi_i - Pg_i \tag{2.21}$$

Mixing Scheme

Another way to solve the KS equation is by using a mixing scheme. Initially, an input guess is used to construct the KS potential and the corresponding KS equation, which is an eigenvalue problem. The resulting eigensolutions provide the charge density that replaces a portion of the input guess. This process is repeated until the difference between the input (ρ_{in}) and output (ρ_{out}) charge density is below a certain threshold value. There are various mixing schemes available, including the linear mixing scheme (equation 2.22) which utilises a mixing fraction α. Other commonly used and efficient mixing schemes include Broyden's method [103] and the Pulay mixing scheme [104].

$$\rho_{in} = (1 - \alpha)\rho_{in} + \alpha\rho_{out} \tag{2.22}$$

2.6 Density Functional Theory-Based Methods for Simulating Excited States

DFT's formulation is grounded in the nature of the ground state, which makes it less suitable for describing excited states. To overcome this limitation, a range of DFT-based methodologies has been developed, aimed at improving the treatment of excited electronic states. These methodologies include Δ self-consistent field (ΔSCF) [82] and time-dependent DFT (TDDF) [105, 106] which includes Linear-response (LR-TDDFT) and constrained DFT (CDFT) [107, 108]. In our work, we use ΔSCF to calculate the energies of the lowest S_1 and T_1 excited states.

2.6.1 Time Dependent Density Functional Theory

While the DFT has been highly effective in determining the ground state properties of electrons under time-independent potentials, it has limitations when it comes to solving excitation problems. This is because the density used in DFT is limited to the ground state density. Additionally, when the system involves a time-varying potential, an extension to DFT is needed.

Time-dependent density functional theory (TDDFT) is such an extension that allows for the solution of time-dependent problems in DFT.

Similar to the Hohenberg-Kohn theorem in DFT, the Runge-Gross theorem [105] is a fundamental theory in TDDFT. The theorem proves that for a given initial state, there exists a one-to-one correspondence between the time-dependent density $\rho(\mathbf{r}, t)$ and the time-dependent external potential $V_{ext}(\mathbf{r}, t)$. In other words, if we know the time-dependent external potential at each point in space and time, we can uniquely determine the corresponding time-dependent density at each point in space and time, and vice versa. This result is significant because it allows TDDFT to determine the time-dependent behaviour of quantum systems based on their initial state, which is often more practical than solving the Schrödinger equation directly.

It is worth noting that several commonly used XC functionals have a reputation for performing poorly in predicting the energy of CT-excited states via TDDFT [109]. Although range-separated functionals may alleviate some of these issues, the following DFT methods for excited states are presented as an alternative to avoid this problem.

2.6.2 ΔSelf-Consistent Field Density Functional Theory

The energy splitting of a specific excited state is defined as the difference in energy between that excited state and the ground state of the system. This energy difference is denoted by

$$\Delta E^{n}_{SCF} = E_n – E_0 \tag{2.23}$$

where E_0 is the ground-state energy and E_n is the energy of the excited state, labelled by n. The energy of the excited state is obtained by manually adjusting the occupation of the KS states until the system reaches self-consistency.

The ΔSCF approach has been widely successful due to its simplicity and low computational cost. In particular, it has been effective when the excited state being studied has the same symmetry as the ground state of the system [110]. In such cases, ΔSCF has been a justifiable

and accurate method for determining energy differences. However, the applicability of ΔSCF has also been extended to the general case [111]. Overall, the ΔSCF approach's ease of use, affordability, and ability to determine energy differences accurately has made it a favoured method in computational chemistry.

The ΔSCF method is especially advantageous for high-throughput virtual screening (HTVS) processes, which involve filtering large databases of compounds to identify those with specific desired properties. At the initial stage of molecule filtering, ΔSCF calculations are preferred over TDDFT because they are significantly faster and require fewer computational resources. This speed and efficiency enable the screening of a large number of molecules, which accelerates the discovery of novel compounds with desirable properties.

It should be noted that the ΔSCF method is not valid when multiple electronic configurations significantly contribute to the description of an excited state [112]. In such cases, the electronic structure of the system cannot be adequately captured by a single reference configuration, and the use of a single-reference approach like DeltaSCF may lead to inaccurate results for the excited state properties. Instead, more sophisticated methods, such as multi-configurational approaches [113] like multi-reference perturbation theory (MRPT) or configuration interaction (CI), should be employed to account for the diverse electronic configurations and accurately describe the complex electronic structure of excited states.

2.7 Pseudopotentials

Atomic pseudopotentials are a widely used technique in computational materials science and quantum chemistry. These pseudopotentials are used to approximate the potential energy of an atom's core electrons, which are always at a lower energy level than the valence electrons. By replacing the core electrons with an effective potential, the computational cost of electronic structure calculations can be greatly reduced, allowing for the study of larger and more complex systems.

The use of pseudopotentials in conjunction with a basis set is a common approach to solving

the electronic structure problem. The basis set represents the valence electrons and is used to expand the electronic wavefunctions of the system. The pseudopotential is then used to describe the behaviour of the core electrons, which are typically less important in chemical reactions and other phenomena. By separating the electronic structure calculation into these two components, it becomes possible to use a smaller basis set and achieve a more efficient calculation.

The choice of pseudopotential depends on the system being studied and the level of accuracy required. There are several types of pseudopotentials [114, 115], each with its own strengths and weaknesses. The most commonly used pseudopotentials include norm-conserving pseudopotentials (NCPP) [116] and ultrasoft pseudopotentials(USPP) [117]. NCPPs are designed to preserve the norm of the wave function, and are typically used in calculations where high accuracy is required. USPPs are more flexible and can be used in a wider range of systems, whereas USPPs are more flexible than NCPPs and can accurately describe the core and valence electrons with fewer basis functions. They are designed to have soft cores, which means that the potential is smoothly varying at the core region, allowing for a better description of the core electrons.

2.7.1 Non-linear Core Correction

Nonlocal-core corrected pseudopotentials (NLCCs) [118] are a type of pseudopotential commonly used in electronic structure calculations to address the limitations of traditional pseudopotentials, including norm-conserving pseudopotentials (NCPPs), in the treatment of the core electrons of an atom. NLCCs aim to correct these limitations by including an additional non-linear correction to the effective potential, which takes into account the behaviour of the core electrons more accurately. The correction potential is obtained from a density-functional theory calculation and is added to the norm-conserving potential to produce the final NLCC pseudopotential. This additional correction improves the transferability and accuracy of the pseudopotential by accounting for the non-locality of the core electrons.

2.8 Basis sets

Basis sets are a crucial component in DFT calculations as they are used to represent the wave function of a molecule or material. They are sets of mathematical functions that are chosen to represent the spatial distribution of electrons in a molecule or material. The choice of basis set is a balance between accuracy and computational cost [119, 120]. One might assume that the larger the basis set, the more accurate the approximation of the wavefunction, however, the relationship between basis set size and accuracy is not always straightforward and depends on several factors.

For systematic basis sets such as those based on Gaussian-type functions or plane waves, increasing the number of basis functions usually leads to an improvement in the accuracy of the results. This is because adding more basis functions allows for a more flexible representation of the wavefunction, which can capture more of the complexity of the system being studied.

However, there are cases where adding more basis functions does not necessarily lead to more accurate results. For example, in some cases, adding too many Gaussians can lead to basis set superposition error (BSSE) [121], where the basis functions become overcomplete and interfere with each other, leading to unphysical results.

There are several types of basis sets used in Density Functional Theory (DFT), each with its own advantages and disadvantages. Some of the most common ones are:

1. Gaussian basis sets: These are typically contracted Gaussian functions that are used to describe the electronic wavefunctions of atoms and molecules. They are widely used in quantum chemistry software such as Gaussian [122], NWChem [123], and GAMESS [124].

2. Plane wave basis sets: These are used in periodic DFT calculations and are typically defined on a uniform grid in reciprocal space. They are used in software packages such as VASP [125, 126, 127] and Quantum ESPRESSO [128].

3. Numeric atom-centred orbital (NAO) basis sets: These are based on the radial functions of atomic orbitals and can be used for both molecular and periodic DFT calculations.

They are used in FHI-aims [129]

4. Wavelet basis sets: They are a set of functions that can be used to represent other functions through linear combinations. While they are not typically used as a complete orthonormal basis for a function space, they are still considered basis sets. In addition to wavelet basis sets, real space grid approaches can also be used to represent functions in numerical calculations. They are commonly used in the BigDFT [130].

5. Mixed basis sets: These are combinations of different types of basis sets, such as Gaussian and plane wave basis sets used by CP2K [131], or ONETEP [132] which combines plane waves with sinc functions to accurately represent the electronic wavefunction.

6. Slater-type orbital (STO) basis sets: These are based on the radial functions of hydrogen-like atomic orbitals and can be used to describe the electronic wavefunctions of atoms and molecules. They are commonly used in software packages such as MOLPRO [133] and CFOUR [134].

The choice of basis set is crucial in DFT calculations and depends on the problem being studied and the computational resources available. For molecular systems, Gaussian-type basis sets are popular as they are flexible and accurately describe electron density near nuclei. However, these basis sets are computationally expensive for large systems like solids.

Solid-state systems are better suited for plane-wave basis sets that efficiently describe periodic systems with many atoms. However, these basis sets may be less precise in describing electron density near nuclei.

Wavelet basis sets are less common but offer several advantages, including efficient descriptions of localised electronic states and adaptability to non-uniform grids. They may be useful in modelling complex systems or problems requiring non-uniform grids.

For small-to-medium-sized molecular systems, Gaussian-type basis sets are preferred, while large solid-state systems benefit from plane-wave basis sets. Wavelet basis sets are ideal for complex electronic states or non-uniform grid requirements. By considering the nature of the

problem and the computational resources available, researchers can choose the most suitable basis set to obtain reliable and efficient results in their DFT calculations.

2.9 BigDFT

As briefly introduced in the previous section, BigDFT [135, 130, 136] is a powerful computational software package that uses Density Functional Theory (DFT) as its foundation. It is an open-source software package that is distributed under the terms of the GNU General Public License (GPL), which means it is free to use and can be modified and distributed by anyone as long as the license terms are followed. It utilises a combination of Daubechies wavelets and pseudopotentials, resulting in efficient and precise calculations [137, 138].

The wavelet basis set is constructed by generating a set of orthogonal wavelets, each with a different scale and position, and then combining them to form a basis that can represent the electronic wave function. This approach provides several advantages over traditional DFT methods, including the ability to efficiently and accurately compute the electronic structure of large systems. The Kohn-Sham equations are solved using the resulting wavelet basis set, allowing for the calculation of a wide range of properties for the system under study.

Daubechies wavelets are a type of basis set known for their high level of accuracy due to their orthonormal property. This property allows them to be both orthogonal and normalised. Orthogonality means that the wavelets have zero inner product (i.e., they are perpendicular) unless they are the same function, while normalisation means that the wavelets have a unit norm (i.e., the integral of the square of the function is equal to one). Daubechies wavelets' orthogonality property allows them to be localised in both real space and Fourier space, which is beneficial for accurately representing the electronic wave function. In real space, the wavelets can be focused on specific regions of the system, leading to a more efficient and accurate calculation. In Fourier space, the wavelets can represent high-frequency components of the wave function that are crucial for accurately describing the electronic properties of the system. Daubechies wavelets' adaptivity makes them particularly well-suited for simulating isolated or

homogeneous systems [135, 136].

The wavelet of functions on a uniform grid of points spaced at intervals h in three dimensions (x, y, z). The spacing of grid points can be adjusted to vary the number of basis functions, leading to more accurate calculations. The adaptive grid in BigDFT consists of two levels, a high-resolution region containing chemical bonds and a low-resolution region further away from the atoms where wave functions decay exponentially to zero [135]. The adaptive grid is used to efficiently and accurately represent the electron density in a region of interest while using a coarser representation in regions of lower electronic density. The size and resolution of the adaptive grid are determined by the coarse and fine multipliers, which define the size and resolution of the coarse and fine grids, respectively. The adaptive grid is constructed on top of a uniform grid of points with a fixed spacing, but the resolution of the adaptive grid can be adjusted based on the local electronic density.

2.9.1 Bromine and Iodine NLCC pseudopotential

As previously mentioned in section 2.7.1, the use of non-local core-correction (NLCC) pseudopotentials (PSPs) can improve the accuracy and efficiency of DFT calculations by reducing the number of electrons that need to be explicitly included in the calculation.

In BigDFT, the NLCC method is implemented as an option for the exchange-correlation functional [139]. In other words, NLCC is used as a correction to the standard exchange-correlation functional to account for the non-local correlations in the electron density, which are not captured by the standard DFT exchange-correlation functionals. This can lead to more accurate predictions of the electronic and structural properties of molecules and materials.

NLCC PSPs are typically more computationally expensive than the default HGH-GTH (Norm-conserving) [140, 141] pseudopotentials used in electronic structure calculations based on DFT. However, they are necessary to accurately account for non-local correlations in electron density.

A limiting factor for the use of NLCC pseudopotentials is that they may not be available for all elements, such as bromine (Br) and iodine (I) at the time of a particular study. This can

pose a challenge when studying molecules containing these heavy elements. In such cases, the grid spacing parameter h becomes crucial as it directly affects the computational accuracy and efficiency of the DFT calculations.

In a high-throughput workflow, it may not be feasible to converge the grid spacing on a case-by-case basis for every molecule being studied. Therefore, it is important to find a value for the h that is good enough across molecules. This is especially challenging when dealing with heavy elements, where the accuracy of the DFT calculations can be more sensitive to the choice of the h parameter.

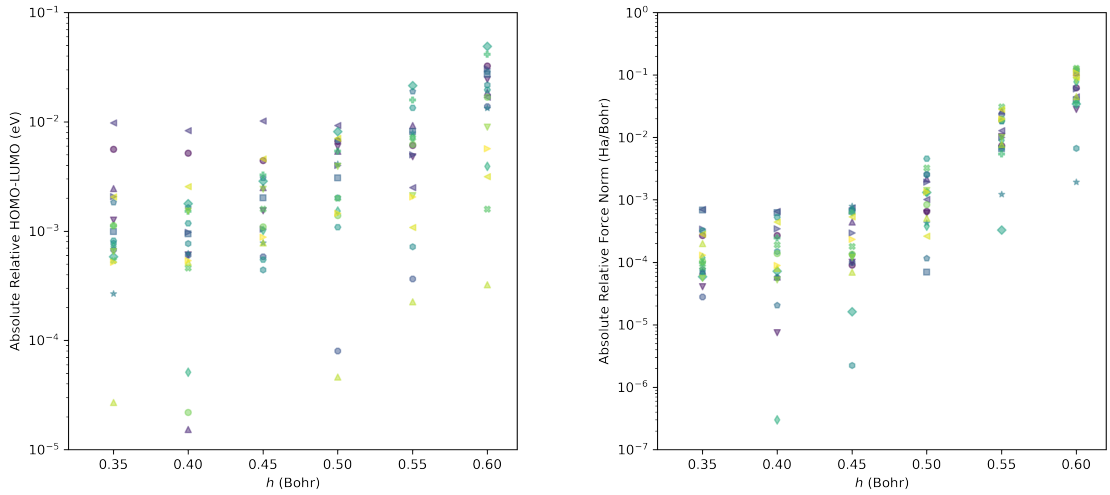
Convergence tests (Figure 2.1) are an essential aspect of electronic structure calculations, as they ensure the accuracy and reliability of the numerical results obtained from a simulation. The accuracy of these calculations is highly dependent on the computational parameters used, such as the grid spacing. Without proper convergence testing, it is difficult to assess the accuracy and reliability of the results obtained.

In this chapter, we randomly selected twenty molecules containing I or Br from a chemical space generated from the molecule TXO-TPA. The details of TXO-TPA and the chemical space will be discussed in the next chapter.

To optimise the gas phase geometry of the ground state, we used the PBE method in BigDFT. We set the maximum force threshold to 0.02 eV/Å and used grid spacings of $h = 0.45$ Bohr (a_0).

To investigate the convergence behaviour of the HOMO-LUMO gap (eV) and normalised force (force Norm) with respect to the h-grid spacing, we performed a convergence test using a range of h values, starting from 0.30 Bohr (a_0) and increasing in increments of 0.05 up to 0.60 Bohr (a_0), as seen in Figure 2.1.

We observed that the absolute difference in the HOMO-LUMO (eV) between two consecutive h values was similar at $\approx 10^{-2}$ eV for $h = 0.5$ and below. For the force, the absolute difference was similar at $\approx 10^{-3}$ Ha/Bohr for $h = 0.45$ and below. However, for the force at $h = 0.5$, we observed an absolute difference of 10^{-2} Ha/Bohr. This suggests that the force has not yet fully



(a) (b)

Figure 2.1: (a) The convergence of LUMO-HOMO absolute relative values with respect to grid spacing h for molecules containing I and Br atoms. The y-axis represents the absolute relative LUMO-HOMO values, which are compared to those obtained at $h = 0.3$ (b) The convergence of the absolute relative Force Norm as a function of h, where y values are relative to the y value obtained at $h = 0.3$. A molecule is represented by a consistent symbol and colour.

converged for this h value, and further calculations with smaller h values may be necessary to achieve convergence.

Overall, our results suggest that the HOMO-LUMO gap has converged for h values of 0.5 and below, while further calculations may be necessary to fully converge the force, particularly for $h = 0.5$.

The choice of computational parameters for virtual screening involves a trade-off between accuracy and efficiency, as well as the need for validation and comparison with previous studies. Based on the results of our convergence testing, we found that $h = 0.5$ may be sufficient for calculating the HOMO-LUMO gap, while $h = 0.45$ may be necessary for geometry optimisation. In the context of high throughput virtual screening, this choice of parameters provides a good balance between accuracy and efficiency.

2.10 Force Field Optimisation

Force field optimisation is a computational method used in molecular dynamics simulations to calculate the interactions between atoms and molecules in a system. A force field is a mathematical model that describes the interactions between atoms and molecules, including electrostatic interactions, van der Waals forces, and bond stretching and bending.

The goal of force field optimisation is to determine the force field parameters that best match experimental data or quantum mechanical calculations so that the simulation accurately predicts the behaviour of the system being studied. This involves adjusting the values of the force field parameters, such as bond lengths, angles, and charges until the simulation results match the experimental data or quantum mechanical calculations.

Force field optimisation can be performed using several techniques [142], such as fitting to experimental data [143], using quantum mechanical calculations to derive parameters [144], or using statistical methods such as Bayesian inference [145]. The optimised force field can then be used to simulate the behaviour of the system under different conditions, such as changes in temperature, pressure, or composition, allowing researchers to investigate the properties and behaviour of molecules and materials in a wide range of scenarios.

However, in the context of HTVS, force fields are often used for the initial screening of molecules [146] due to their computational efficiency and ability to handle larger molecular systems. As force fields are empirical models that are based on simplified descriptions of molecular interactions, it makes them much faster than DFT. By using force fields for initial screening, we can quickly identify a smaller set of promising molecules that can be studied in more detail using more accurate methods like DFT. Additionally, force fields can provide a quick and reliable estimate of some properties, which can be used for initial screening. Overall, force fields are a useful tool for quickly screening a large number of molecules and identifying promising candidates for further study.

2.10.1 The Merck Molecular Force Field

The Merck Molecular Force Field (MMFF) [147, 148] is an empirical force field that is commonly used for studying organic molecules. It was developed by Merck to model the interactions between atoms and molecules, including bond stretching, bending, out-of-plane (OOP) bending and torsion, as well as non-bonded interactions such as Van der Waals (VdW) forces and electrostatic interactions (Q).

The total MMFF energy expression can be written as equation 2.24 [147]. In equation 2.24, $i, j, k, l, \ldots$ represent atoms, E_{ij}^{bond} represents the energy associated with stretching or compressing bonds (equation 2.25), E_{ijk}^{angle} represents angle bending (equation 2.26), E_{ijk}^{bend} represents stretch-bend interactions (equation 2.27), E_{ijkl}^{OOP} represents OOP bending (equation 2.28), $E_{ijkl}^{torison}$ represents torsion interaction (equation 2.29), E_{ij}^{VdW} represents Vander Walls interactions (equation 2.30) and E_{ij}^{Q} represents electrostatic interactions (equation 2.35).

$$E_{MMFF} = \sum E^{bond}ij + \sum E_{ijk}^{angle} + \sum E_{ijk}^{bend} + \sum E_{ojk;l}^{OOP} + \sum E_{ijkl}^{torsion} + \sum E_{ij}^{VdW} + \sum E_{ij}^{Q} \tag{2.24}$$

$$E_{ij}^{bond} = 143.9325 \cdot \frac{k_{IJ}^{bond}}{2} \cdot \Delta r_{ij} \cdot \left(1 + cs \cdot \Delta r_{ij}^2 + \frac{7}{12} cs^2 2 \cdot \Delta r_{ij}^2\right) \tag{2.25}$$

In equation 2.25, $I, J, K, L, \ldots$ denote the corresponding (to $i, j, k, l, \ldots$) numerical MMFF atom types, k_{ij}^{bond} is the force constant, $\Delta r_{ij} = r_{ij} - r_{ij}^{ref}$ is the difference between actual and reference bond lengths and cs is the *cubic stretch* constant.

$$E_{ijk}^{angle} = 0.043844 \cdot \frac{k_{IJL}^{angle}}{2} \cdot \Delta\theta_{ijk}^2 \cdot (1 + cb \cdot \Delta\theta_{ijk}) \tag{2.26}$$

In equation 2.26, k_{IJK}^{angle} is the force constant, $\Delta\theta_{ijk} = \theta_{ijk} - \theta_{ijk}^{ref}$ is the difference between actual and reference bond angles and cb is the *cubic bend* constant.

$$E_{ijk}^{bend} = 2.51210 \cdot (k_{IJK}^{bend} \cdot \Delta r_{ij} + k_{KJI}^{bend} \cdot \Delta r_{kj}) \cdot \Delta\theta_{ijk} \tag{2.27}$$

In equation 2.27, k_{IJK}^{bend} and k_{KJI}^{bend} are force constants which couple the i-j and k-j stretches to

the i-j-k bend. Stretch bend interactions are omitted for linear bond angles.

$$E_{ijk;l}^{OOP} = 0.043844 \cdot \frac{k_{IJK;L}^{OOP}}{2} \cdot X_{ijk;l}^2 \tag{2.28}$$

In equation 2.27, $k_{IJK;L}^{OOP}$ is the force constant, $X_{ijk;l}$ is the Wilson angle [149] between the bond j-l and the plane i-j-k.

$$E_{ijkl}^{torison} = \frac{1}{2} \cdot [V_1(1 + \cos\Phi) + V_2(1 - \cos 2\cos\Phi) + V_3(1 + \cos 3\cos\Phi)] \tag{2.29}$$

In equation 2.29, the constants V_1, V_2 and V_3 depend on the atom types (I, J, J, K, L), where $i-j$, $j-k$ and $k-l$ are bonded pairs. Φ is the $i-j-k-l$ torsion angle.

$$E_{ij}^{VdW} = \varepsilon_{IJ}\left(\frac{1.07R_{IJ}^*}{R_{ij} + 0.07R_{IJ}^*}\right)^7\left(\frac{1.12R_{IJ}^{*7}}{R_{ij}^7 + 0.12R_{IJ}^{*7}} - 2\right) \tag{2.30}$$

MMFF employs *Buffered-14-7* [150] form, in which the potential is described by the equation 2.30. This equation is used together with equation 2.31 which relates the minimum energy separation R_{II}^* to the atomic polarisability α, with combination rules (equation 2.32 and 2.33), and utilising a Slater-Kirkwood expression for the well depth ε_{IJ} (equation 2.34). R_{ij} is the internuclear separation.

$$R_{II}^* = A_I \cdot \alpha_I^{\frac{1}{4}} \tag{2.31}$$

$$R_{IJ}^* = \frac{1}{2} \cdot (R_{II}^* + R_{JJ}^*) \cdot \{1 + 0.2[1 - \exp(-12 \cdot \gamma_{IJ}^2)]\} \tag{2.32}$$

$$\gamma_{IJ} = \frac{(R_{II}^* - R_{JJ}^*)}{(R_{II}^* + R_{JJ}^*)} \tag{2.33}$$

$$\varepsilon_{IJ} = \frac{181.16 G_I G_J \alpha_I \alpha_J}{(\frac{\alpha_I}{N_I})^{\frac{1}{2}} + (\frac{\alpha_J}{N_J})^{\frac{1}{2}}} \frac{1}{R_{IJ}^{*6}} \tag{2.34}$$

$$E_{ij}^Q = \frac{332.0716 q_i q_j}{D(R_{ij} + \delta)^n} \tag{2.35}$$

In equation 2.35, q_i and q_j are partial atomic charges, D is the dielectric constant and δ is electrostatic buffering.

MMFF force field is that it has been widely used and tested for many years [151], so it has

a well-established set of parameters that can provide reasonable results for a wide range of organic molecules, including TADF molecules [152]. The parameters have been calibrated against experimental data, which makes it a reliable tool for predicting the properties of TADF molecules.

MMFF comes in two flavours: MMFF94 and MMFF94S [153]. The difference between them lies in the torsion and out-of-plane bending parameters they use. The 's' in MMFF94S stands for 'static', which means that this set of parameters is better suited for tasks where the output remains static.

A recent benchmark paper [142] compared several empirical force fields for their ability to predict molecular geometries using quantum mechanics (QM) calculations. The study found that MMFF94 and MMFF94S were able to capture QM geometries better than some of the other force fields tested, but still not as well as OPLS3e [154] or OpenFF [155]. While MMFF94 may not be the most accurate force field compared to some of the newer force fields, it is still a well-established and widely used force field that has been shown to provide reasonable predictions for a wide range of organic molecules. Additionally, MMFF94 has been extensively tested and calibrated against experimental data, which makes it a reliable tool for predicting the properties of organic molecules. For these reasons, MMFF94 was chosen for this work.

In summary, we employed the BigDFT code for DFT calculations. Our convergence test revealed that a grid spacing of $h = 0.50a_0$ (0.26Å) was adequate for calculating the HOMO-LUMO gap, whereas $h = 0.45a_0$ (0.24Å) was optimal for geometry optimisation. For the latter, we utilised MMFF94 to initialise molecular geometries, followed by a single-point DFT calculation to determine the HOMO-LUMO energy gap and overlap.

In the upcoming chapter, we will delve into the design and methodology of our HTVS workflow.

Chapter 3

High Throughput Virtual Screening Workflow

3.1 Introduction

The concept of High Throughput Virtual Screening (HTVS) involves using automated algorithms and machine learning techniques to rapidly evaluate large libraries of molecular compounds or materials for specific properties of interest. By harnessing the power of computers, HTVS allows researchers to efficiently identify potential candidates from vast collections of molecules or materials, without the need for time-consuming and resource-intensive physical experimentation.

High-throughput computational approaches have been widely employed in the pharmaceutical industry to identify compounds for experimental study [156]. In recent times, these strategies have also gained popularity for designing molecules and materials intended for optoelectronic applications. Several techniques have been utilised for HTVS [76, 157, 146, 158] designing TADF molecules [159, 160, 13]. However, the techniques involved in HTVS can be summarised in basic steps, which serve as a general framework for guidance and should not be followed strictly, as the process can be highly flexible and adaptable depending on the specific research question and available resources.

Firstly, a database of TADF molecules with known or hypothetical structures is selected or prepared. This database may include various molecular structures and properties, such as HOMO-LUMO energy levels and electron densities. Next, the target for the design of TADF molecules with a small ΔE_{ST} property is identified. The target may be a specific application, such as OLEDs, or a particular material property, such as ΔE_{ST}.

The molecular structures of candidate TADF molecules are then generated using computational methods, such as DFT calculations or molecular dynamics simulations. The candidate molecules are designed to have specific structural features that contribute to small ΔE_{ST} property, such as donor-acceptor or charge-transfer moieties.

The candidate TADF molecules are screened virtually against the target and filtered based on various criteria, such as ΔE_{ST} values and synthetic accessibility. This step is important for identifying the most promising candidate molecules for further experimental validation.

The most promising candidate TADF molecules are synthesised and tested experimentally to evaluate their photophysical properties and performance in the target application, such as OLEDs. The results of these experiments are used to refine the computational models and screening protocols. Based on the results of experimental validation, the computational models and screening protocols are refined, and the process is iterated to improve the efficacy of the candidate molecules. The goal is to identify and optimise TADF molecules with small ΔE_{ST} property and high performance in the target application.

Overall, the HTVS process is a powerful tool for the design of TADF molecules with a small ΔE_{ST} property. The process involves selecting and preparing a database of TADF molecules, identifying a specific target, generating candidate molecules, virtually screening these candidates, filtering them based on various criteria, experimental validation, and iterative refinement. These steps can be adapted to fit specific research objectives and constraints.

3.2 Applications

The application of the HTVS method to the design of optoelectronic materials has yielded numerous successful results in recent times [159, 161, 160, 13, 162, 163].

Shu *et al.* [159] introduced a flexible approach for automated material design that searches a large chemical space to identify potential TADF molecules as new fluorophores for OLEDs. The approach utilises a tree data structure to represent molecules, allowing users to choose fragments and rules to define a specific chemical space. A genetic algorithm (GA) was then employed to identify molecules with desired properties, leading to the discovery of 3792 promising candidate fluorophores from a chemical space comprising 1.26×10^6 molecules. Their GA optimisations required significant computer time, thus, it is required to take advantage of advanced hybrid parallel computer architectures.

The application of genetic algorithms for virtual screening has also been extended to the design of other layers within OLED devices, including Hole Transport, Electron Transport, and Host Layers [161].

Aspuru-Guzik and collaborators made a groundbreaking effort in HTVS of organic TADF materials [160]. Their approach utilised machine learning and TDDFT methods, enabling them to efficiently screen thousands of potential TADF molecules from a pool of 1.6 million candidates. The best molecules identified through this screening process have the potential to be used in the fabrication of OLED devices. Impressively, the resulting external quantum efficiencies of the devices can reach as high as 22%. However, the method described in their work is tailored to generating OLED molecules with TADF character that adhere to a donor-bridge-acceptor structure. Therefore, for TADF molecules lacking distinct donor and acceptor moieties, such as type-VII molecules [13], this method may not be optimal for generating candidate molecules.

Building on the work of Aspuru-Guzik and collaborators, researchers have leveraged Machine Learning-Assisted Virtual Screening techniques [162, 163] to design molecules. They have employed this approach to quickly predict ISC/RISC rates and identify potential TADF candidates

from experimental databases [162]. Additionally, the researchers have introduced the concept of mutation and selection from GAs and combined machine learning algorithms with quantum chemical computations to explore the chemical compound space and obtain organic TADF material candidates [163]. While explicit quantum chemistry calculations can help correct false positive predictions for materials with exceptional properties, these instances are typically rare. However, the use of machine learning methods for virtual screening poses a greater risk of false negatives. This means that exceptional materials may not be identified as such, and could be missed by the screening process.

Zhao *et al.* [13] investigated a diverse database of 40,000 molecules obtained from the Cambridge Structural Database (CSD) which were not intentionally designed for thermally activated delayed fluorescence (TADF) and lacked specific criteria. The team performed ΔSCF excited-state calculations on X-ray geometries and filtered the results based on ΔE_{ST} and oscillator strength of S_1. They then performed ΔSCF calculations on their DFT-optimised geometries. Through a virtual search, the team identified 125 potential TADF candidates without predetermined design rules. The candidates offer new possibilities for TADF material design and exhibit novel donors, acceptors, and molecular structures. Some candidates did not display the typical donor-acceptor (D-A) character, and they were classified into different types based on their novelty compared to existing TADF literature. Overall, the study identifies promising candidates for the development of new TADF materials.

Although Zhao *et al.* [13] achieved impressive results, it is important to note that their methodology can only be applied to databases containing X-ray geometries. However, the team provides illustrative examples at the end of their paper to showcase how their findings can be used to design new molecules. Overall, their work has contributed significantly to the discovery of potential candidates for TADF material design and highlights the importance of exploring diverse databases to identify novel structures.

Overall, HTVS is a powerful tool for designing optoelectronic materials, and researchers have made significant progress in using this approach to identify promising candidates for OLED fabrication. The integration of GAs, machine learning, and TDDFT methods has increased the

efficiency and accuracy of the screening process. However, it's important to note that while these approaches have been successful for traditional TADF molecules, they may not be as effective for type-VII TADF molecules. As mentioned in the introduction of this chapter, the HTVS approach is highly flexible and adaptable depending on the specific research question and available resources.

3.3 Philosophy of High Throughput Virtual Screening

The main philosophy of the HTVS process for designing molecules, including TADF molecules, is to accelerate the discovery of new compounds with desirable properties by screening a large database of potential candidates using computational methods. The HTVS process combines computational chemistry methods, such as DFT calculations, with techniques that are able to rapidly screen a large number of compounds and identify promising candidates with optimised properties.

The goal of HTVS is to reduce the time and cost involved in the traditional trial-and-error approach of designing molecules by using a data-driven approach. This approach can help in identifying potential candidates with desirable properties, such as high TADF efficiency, suitable energy levels for device applications, and good solubility and stability properties.

The HTVS process typically involves the generation of a large database of potential molecules, calculation of molecular descriptors using DFT calculations, training of machine learning models using a subset of the database with known properties, prediction of the properties of the remaining molecules in the database using the trained model, and experimental validation of the predicted properties of the selected candidates. Pyzer-Knapp*et al.* has summarised these ideas into 'four philosophies of High Throughput Virtual screening [146]: Timescale is important, Automated techniques, Data-driven discovery, Computational Funnels.

3.4 Our Workflow

HTVS projects require the selection of candidate molecules to investigate, which is crucial to the success of the project. The successful candidate can come only from the initial library or its successive rounds of growth, putting boundaries on the possible outcomes from the start. Generating novel lead backbones is a harder challenge than pursuing a complete set of substitution patterns through side-chain enumeration. Side-chain modeling [164] is a computational method that involves systematically exploring variations in the substituents attached to a specific site on the core structure of a molecule, while keeping the core structure fixed.

Performing systematic explorations in molecular space is difficult as there are no predefined magnitudes to survey chemical space. Molecular structure obeys a large and complex set of rules that so far defies systematic exploration. Therefore, there is much effort put into methods to avoid explicit enumeration in high-throughput screening, such as optimisation of potentials [165], alchemical transformation [166], generation of electronically equivalent aromatic rings [167], stochastic generation of derivatives [168], recursive substructure searches [169], and morphing of starting molecules of interest [170].

One of the first decisions that need to be made is whether to search for novel properties among pre-existing compounds or completely new chemical motifs. If pre-existing molecules are not to be considered, navigating the vast chemical space of organic molecules becomes the focus. However, this space is too large to sample fully, making the task impossible.

Randomly generating molecules based on organic chemistry rules is not efficient in searching for niche phenomena, as it will produce a large list of uninteresting structures. Instead, synthetic accessibility of hypothetical compounds is an important aspect to consider when designing a de novo library for exploring chemical space. The generation of 3D structural information is often difficult in HTVS, which makes the ease of synthetic accessibility of hypothetical compounds a crucial factor in designing a de novo library for exploring chemical space.

As seen in section 3.2 there are several methods to sample the chemical space, such as Combinatorial modifications involving modifying a starting molecule, structural and chemical databases

screening pre-existing compounds, and generative models using machine learning to generate novel compounds. We also see that these methods are not mutually exclusive and can be used either individually or in combination to effectively explore and sample the vast chemical space in pursuit of compounds possessing desired properties.

The outline of our workflow is summarised in Figure 3.1, highlighting the various steps involved in the method and design space exploration. Each step represents a crucial component of the research process, and the diagram serves as a visual roadmap for the overall investigation.

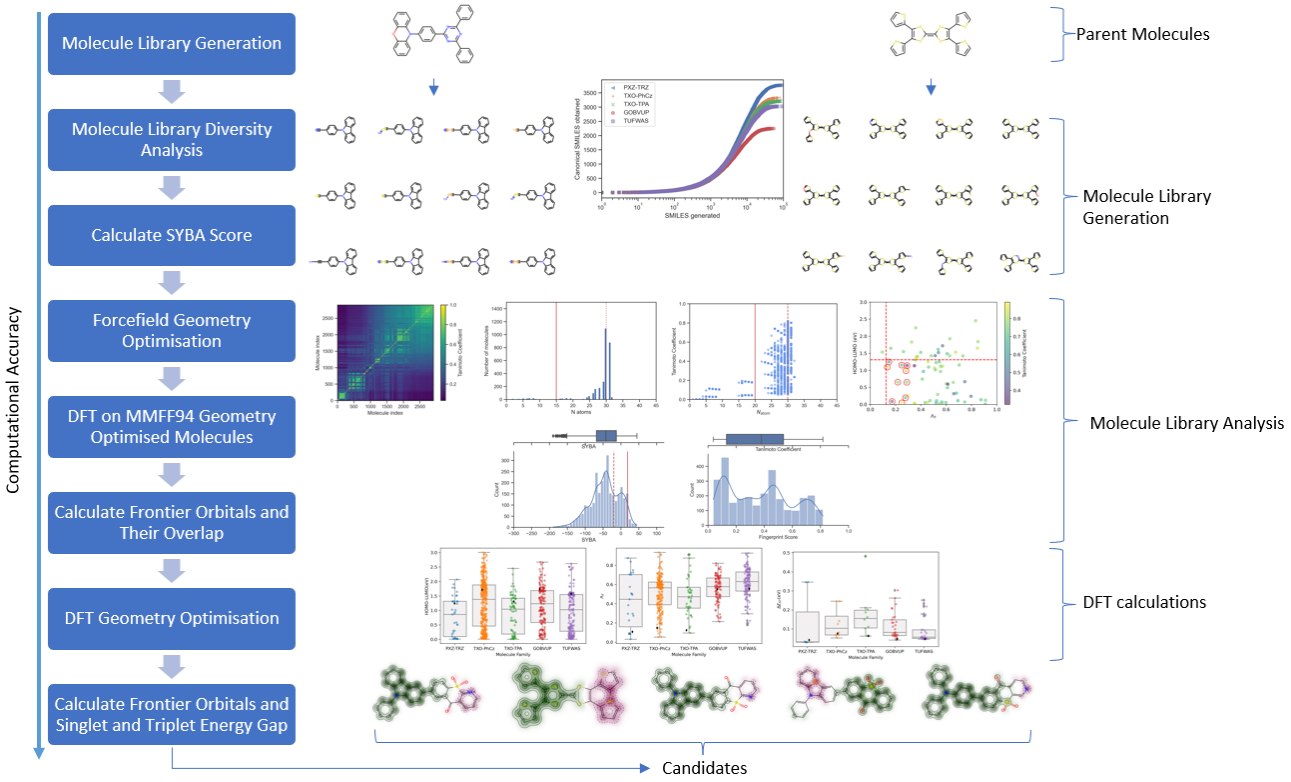


Figure 3.1: Outline of our workflow schematic illustrating the method and design space, with accompanying snapshots of analysis and results from later chapters.

In the following section, we provide a detailed description of the methods employed in our HTVS workflow, along with relevant computational details. This will include library generation, molecular diversity, synthetic accessibility and DFT calculations.

3.4.1 Library Generation

In this work, we aim to design molecules with desirable properties, such as high TADF efficiency, using the HTVS workflow. To achieve this, we used the Self-Referencing Embedded Strings (SELFIES) [171] molecular representation and the Superfast Traversal, Optimisation, Novelty, Exploration and Discovery (STONED) algorithm [172]. Specifically, we used these tools to generate a chemical space from a selected set of TADF molecules. This involves making modifications to the SELFIES representation through point mutations, which correspond to single-character additions, deletions, or replacements. The details regarding the composition and contents of the TADF molecule set are provided in a later section of this chapter, section 3.4.4.

This method allows for the rapid generation of a large number of structurally diverse molecules. This can be particularly useful when searching for novel compounds with specific properties or when exploring a wide range of chemical space. By generating a diverse set of compounds, we can increase the likelihood of finding molecules with desirable properties. The SELFIES molecular representation allows for flexible modification of existing molecules, which can be useful for improving properties or generating new scaffolds. For example, the SELFIES representation can be used to modify a known molecule to improve its potency, selectivity, or other properties. Alternatively, it can also be used to generate new scaffolds by combining different molecular fragments.

Overall, SELFIES representation offers a flexible, robust, and efficient approach to generating and screening a library of molecules.

In this project, we used RDKit, a popular open-source cheminformatics toolkit written in Python, to preprocess a set of TADF molecules and extract their corresponding SELFIES and SMILES (Simplified molecular-input line-entry system) [173] representations. In the next section, we will explain how the molecules are represented with SMILES and SELFIES.

3.4.2 Molecule Representation

Line notations like SMILES [173] and SELFIES [171] are a shorthand way to represent the structure of a molecule in a compact and easily human-readable representation of molecular structures that can be easily processed and analysed by computers. Furthermore, line notations are standardised formats that are widely recognised and supported by a large number of software tools and databases, making it easy to integrate into existing workflows and applications. In contrast, 3D xyz files may have different formats and conventions depending on the software used to generate them, which can make them difficult to work with.

There are other several line notations used to represent molecular structures, such as international Chemical Identifier (InChI) [174], Modular Chemical Descriptor Language (MCDL) [175, 176], SYBYL Line Notation (SLN) [177], Representation of Structure Description Arranged Linearly (ROSDAL) [178] and many more [179, 180, 181, 182, 183, 184, 185, 186]. However, SMILES strings and InChI strings are widely recognised as the most established ones.

The main goal of InChI is to provide a unique and canonical identifier for chemical structures. To achieve this, InChI uses a layered approach where the molecular structure is first standardised into a set of structural features, which are then encoded into a text string. In contrast to SMILES strings, which can have multiple representations for the same molecule, InChI addresses this weakness by providing a single, unambiguous identifier.

On the other hand, SMILES strings are widely utilised for the storage and exchange of chemical structures, but there is currently no standardised approach for generating a canonical SMILES string [187]. This is because the algorithm for canonicalising SMILES strings is proprietary, leading to varying implementations across different companies and research teams [188]. SMILES strings are relatively simple and easy to use, with a well-defined syntax and rules for representing atoms, bonds, and other molecular features. SMILES can represent both small and large molecules and are widely used in cheminformatics and computational chemistry applications.

SELFIES is a novel line notation for representing molecular structures using a predefined set of

symbols and rules, which was introduced in 2020 [171]. Unlike other line notations, SELFIES guarantees a 100% robust molecular string representation, ensuring that any combination of symbols in the SELFIES alphabet corresponds to a chemically valid graph. This makes SELFIES a highly reliable method for representing molecular structures, particularly in applications where accuracy is critical. Moreover, SELFIES allows for the generation of canonical STLFIES strings, which are standardised SELFIES strings for a given molecule that are invariant to different representations of the same molecule.

Overall, the choice between InChI, SMILES, and SELFIES depends on the specific application and requirements. InChI is suitable for global chemical identification and integration with chemical databases and search engines, while SMILES is useful for general molecular structure representation and analysis. SELFIES is a 100% robust notation that is particularly useful for machine learning and other computational applications.

We chose to use SELFIES over InChI for representing molecular structures in our project due to its 100% robustness and other related advantages. In comparison, InChI relies on a fixed set of rules that may not always result in chemically valid structures, leading to errors and inaccuracies. SELFIES, on the other hand, guarantees that every combination of symbols in its alphabet corresponds to a valid molecular graph, making it a highly reliable method for representing molecular structures. Additionally, SELFIES allows for the generation of canonical STLFIES strings that are invariant to different molecular representations. These benefits make SELFIES an ideal choice for our project.

In this work, we utilise two popular line notations, SMILES and SELFIES, in conjunction with RDKit, a powerful cheminformatics toolkit, to represent and manipulate the molecular structures in our computational analyses. In this work, we employ two line notations, SMILES and SELFIES, to represent and manipulate molecular structures in our computational analyses. SMILES was used to for molecular representation and SELFIES for molecular mutation and to generate canonical string representations that can be converted back to SMILES, making them readable and compatible with existing chemical databases. Overall, By utilising the combined power of SMILES and SELFIES in RDKit, we can enhance our ability to represent, modify, and

analyse molecular structures in our computational studies, thus providing a robust platform for conducting effective research.

In the following section, we will give examples of SMILES and SELFIES string representations of molecules.

Simplified Molecular-Input Line-entry System

SMILES is a line notation for representing the structure of a molecule as a text string. It was developed in the 1980s [173] and has become one of the most widely used notations for representing molecular structures in cheminformatics and computational chemistry.

SMILES represent a molecular structure as a text string, where atoms are represented by their elemental symbols and bonds are represented by various symbols and characters. The notation is designed to be easily readable and writable by humans, while also being easily interpretable by computers.

An example of a simple molecule, methane (CH4), is represented in SMILES notation: `C`. In this notation, "`C`" represents the carbon atom and no explicit symbol is used for the hydrogen atoms. The implicit hydrogen atoms are assumed to be attached to the carbon atom. The SMILES string for methane can also be written with explicit hydrogen atoms as follows:

```
C[H][H][H][H]
```

This notation explicitly shows that there are four hydrogen atoms attached to the carbon atom.

Another example of a more complex molecule, aspirin (C9H8O4), is represented in SMILES notation:

```
CC(=O)Oc1ccccc1C(=O)O
```

In this notation, "`C`" represents a carbon atom,"`O`" represents an oxygen atom, "`c`" represents an aromatic carbon atom, and "`()`" represents a functional group. The SMILES string encodes

the molecular structure of aspirin, including the positions of the atoms and the nature of the bonds between them. In general, each unique molecule should have one unique SMILES representation.

However, there are certain cases where a molecule can have multiple SMILES representations. This occurs when there are different valid ways to represent the same molecular graph, due to the various possible arrangements of hydrogen atoms and/or stereochemistry. These alternate SMILES representations can arise from differences in the order of atoms or the assignment of bond types and can be caused by differences in the software used to generate the SMILES strings or the options and settings used.

For example, consider the molecule 1,2-dichloroethene. One valid SMILES string for this molecule is "`ClC=CCl`", which represents the molecule in the transform. However, this same molecule can also be represented in the cis form with the SMILES string "`Cl\C=C/Cl`". In this case, the double bond is represented using the slash notation to indicate the cis configuration of the two chlorine atoms. It is important to note that while these SMILES strings represent the same molecular graph, they do not necessarily represent the same physical molecule, since the cis and trans isomers can have different properties and reactivity. In the context of HTVS and chemical space exploration, the cis and trans isomers may not be as important, as the focus is often on identifying compounds with certain molecular properties or functional groups that could be useful for a particular application. However, it is still important to be aware of the stereochemistry when evaluating the potential properties and reactivity of a molecule.

SELF-referencIng Embedded Strings

SELFIES (SELF-referencIng Embedded Strings) [171] is a recently developed line notation for representing the structure of a molecule as a string of text. It was developed in 2020 as a successor to SMILES, with the goal of addressing some of the limitations of SMILES and other line notations. SELFIES represent a molecular structure using a set of predefined symbols and rules. These symbols and rules are designed to be easily interpretable by computers and can represent a wide range of molecular structures, including those that may be difficult to

represent using other line notations.

An example of a simple molecule, methane (CH4), is represented in SELFIES notation: `[CH4]`. In this sting notation, "`[CH4]`" represents the methane molecule, with the square brackets indicating the start and end of the SELFIES string. The notation is similar to SMILES but uses square brackets to group atoms and functional groups.

Another example of a more complex molecule, aspirin (C9H8O4), is represented in SELFIES notation:

`[O][=C][OH][C][c][c][c][c][c][C][=O][O][C][C][C]`

In this notation,"[O]" represents an oxygen atom, "`[=C]`" represents a double bond, "`[OH]`" represents a hydroxyl group, "`[c]`" represents an aromatic carbon atom, and "`[C]`" represents a non-aromatic carbon atom. The SELFIES string encodes the molecular structure of aspirin, including the positions of the atoms and the nature of the bonds between them.

SELFIES is designed to address some of the limitations of SMILES. One major advantage of SELFIES is that it is a more robust representation of molecular structures. SMILES can produce different representations of the same molecule, which can cause issues with data consistency and reproducibility. SELFIES, on the other hand, is a unique string representation that guarantees a one-to-one correspondence between the SELFIES string and the underlying molecular structure.

Another advantage of SELFIES is that it allows for easier handling of complex molecular structures. SELFIES can represent complex molecular features such as rings, stereochemistry, and charges in a more concise and consistent manner compared to SMILES.

It should be noted that while SELFIES were considered superior to SMILES, the format was still in the process of being standardised. Therefore, to generate a molecule library, we used both SMILES and SELFIES. The SMILES format was converted to SELFIES before point mutations were applied. This approach allowed us to take advantage of the benefits of both formats while mitigating any potential issues with their standardisation.

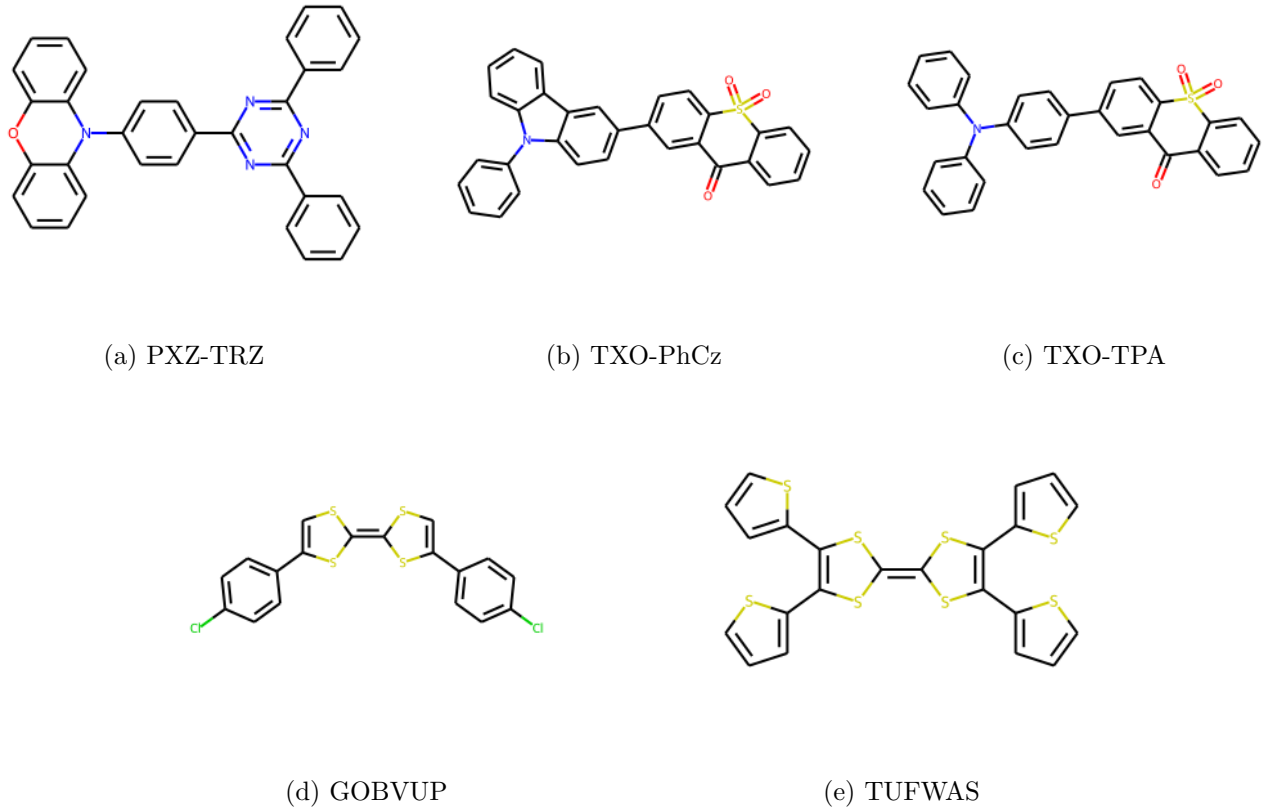
Overall, SELFIES are better than SMILES regarding data consistency, reproducibility, and the representation of complex molecular features. The authors of SELFIES acknowledge that further work is required to standardise the format. They plan to organise a workshop to extend the format to include features present in SMILES, such as stereochemistry, polyvalency, and isotopic substitution, and to define the direct canonicalisation of SELFIES. The authors aim to use Unicode to create readable symbols and to allow for canonical SELFIES strings for unique molecules. Currently, SMILES can be indirectly made canonical by translating them to SELFIES and converting the canonical SMILES back to SELFIES.

3.4.3 Superfast Traversal, Optimisation, Novelty, Exploration and Discovery Algorithm

To generate a diverse library of molecules, we employed part of the STONED algorithm [172] that forms local chemical spaces, which utilises the SMILES string representation of each molecule as input. The algorithm then transformed the SMILES representation into SELFIES, a more recent format that allows for point mutations. In SELFIES, each point mutation corresponds to a single-character addition, deletion, or replacement, with an equal probability (33%) for each type of mutation. The algorithm repeated this process iteratively until it achieved a canonical SELFIES representation of the molecule, ensuring a unique and standardised representation regardless of its initial form. Finally, the canonical SELFIES representation was converted back to the original SMILES format.

3.4.4 Parent Molecules

Our study centres around investigating five TADF molecules (Figure 3.2) to explore their chemical space and provide insights towards the development of better TADF molecule designs. Among these molecules, three are identified as D-A TADF molecules, namely PXZ-TRZ, TXO-PhCZ, and TXO-TPA. The remaining two are classified as type-VII TADF molecules [13], which are unique types of TADF that cannot be designed using traditional D-A design rules.



(a) PXZ-TRZ (b) TXO-PhCz (c) TXO-TPA

(d) GOBVUP (e) TUFWAS

Figure 3.2: 2D SMILES representation of TADF molecules used to generate molecule library.

As discussed in section 1.4.3 (Chapter 1), their small ΔE_{ST} values were found to be dependent on specific cases.

lifetimes, and good stability. For type-VII TADF molecules, we chose them due to their size and small ΔE_{ST} values, which were calculated computationally as they have not been synthesised yet. Additionally, the selection of these molecules was also based on their structure, as some molecules can be challenging or even impossible to represent using SMILES notation.

We selected D-A TADF molecules for their molecule size, high efficiency, long fluorescence lifetimes, and good stability. For type-VII TADF molecules, we chose them due to their size and small ΔE_{ST} values, which were calculated computationally [13]. Additionally, the selection of these molecules was also based on their structure, as some molecules can be challenging or even impossible to represent using SMILES notation. This is because SMILES notation is a linear string of characters that represents the molecular structure based on a set of rules, and not all molecular structures can be described using these rules. For example, molecules with

complex or highly branched structures may be difficult to represent accurately with SMILES notation. In addition, molecules with ambiguous stereochemistry, such as cis/trans isomers, may require additional information beyond what is available in the SMILES string to represent their structure accurately.

It is important to note that the methods used in our HTVS workflow to generate and analyse TADF molecules have not been applied to D-A TADF molecules before. By extending our analysis to include D-A TADF molecules, we can assess the effectiveness of these methods in predicting TADF molecules for both types. This will provide insights into the applicability and robustness of our methods for discovering new TADF molecules.

In the next section, we will discuss how to determine the diversity of the molecules generated.

3.4.5 Molecular Diversity

To ensure the chemical space of interest is well-represented in the library, it is crucial to include a diverse range of compounds. This diversity will allow for the incorporation of compounds with a broad range of properties, including those that are conducive to efficient and effective TADF emission, as well as compounds relevant to this research area.

Structural diversity is a crucial aspect of diversity that measures the differences in chemical structures among molecules in a library. One common approach to assess structural diversity is by computing the molecular fingerprint of each compound and comparing them using similarity metrics, such as Tanimoto. While other similarity metrics have been proposed for specific use cases, the Tanimoto coefficient remains the most extensively used in cheminformatics, as demonstrated by numerous studies [189, 190, 191]. Its popularity is attributed to its simplicity, fast computation, and proven performance across various scenarios. Therefore, we employed the Tanimoto coefficient to analyze our molecule library. However, it is worth noting that some studies suggest using alternative metrics if specific information on molecule sizes is available [192].

Tanimoto Coefficient

The Tanimoto coefficient, defined by equation 3.1, is defined as the ratio of the intersection of the two sets (A, B) over the union of the two sets.

$$T = \frac{N_{AB}}{N_A + N_B - N_{AB}} \tag{3.1}$$

In equation 3.1, N_{AB} represents the number of common elements or features between sets A and B, N_A is the total number of elements or features in set A and N_B is the total number of elements or features in set B.

In the context of chemistry and molecular similarity, sets A and B can represent molecular fingerprints or descriptors, where each element or feature corresponds to a specific molecular property or substructure present in the molecules. These fingerprints or descriptors encode information about the presence or absence of specific molecular features or substructures in each molecule.

To calculate the Tanimoto coefficient for chemical structures, the molecular fingerprints or descriptors of two molecules are first generated. Molecular fingerprints are binary representations of molecular structures, where each bit represents the presence (1) or absence (0) of a specific molecular feature or substructure. Fingerprints can be generated using various algorithms, such as the Extended Connectivity Fingerprints (ECFPs) [193], Morgan fingerprints [194], or MinHash fingerprints (MHFPs) [195]. Each molecule in the dataset is assigned a unique fingerprint, resulting in a set of binary values for each molecule. More detailed discussions on types of fingerprints are present in the next section.

Molecular descriptors are numerical representations of molecular properties, such as molecular weight, topological indices, and physicochemical properties. Each molecule is characterized by a vector of numerical values corresponding to its descriptors.

A Tanimoto coefficient of 1 indicates that the two sets (fingerprints or descriptors) are identical, meaning the molecules have the same molecular features or properties. A coefficient of 0

represents no similarity, indicating that the two sets have no common features or properties.

The Tanimoto coefficient is widely used in cheminformatics and drug discovery to assess the similarity between molecules, screen compound libraries, and perform virtual screening to identify potential candidates with similar properties to the target molecule. It provides a valuable quantitative measure of molecular similarity, aiding in the exploration of chemical space and the identification of promising leads for further experimental investigation.

In the following section, we will discuss the types of fingerprints and the reasons for the fingerprint method we used.

Fingerprint

As briefly mentioned in the previous section, various algorithms can be used to generate fingerprints, such as the Morgan fingerprint [194], ECFPs [193], MHFPs [195], and atom-pair fingerprints [196]. ECFPs are derived from the Morgan fingerprint algorithm and are typically used for small molecules, while atom-pair fingerprints are designed for large molecules. MHFPs are versatile and can be used for both small and large molecules [197].

The terms “small” and “large” are relative and can be used in different contexts with different meanings. In the context of molecular fingerprints, the size of a molecule can refer to the number of atoms it contains. As fingerprint algorithms are typically used in drug design (but are not limited to), TADF molecules can be considered relatively small.

ECFP4 is considered one of the most effective molecular fingerprints for small molecule virtual screening, as well as for target prediction benchmarks, according to several studies [198, 199, 200, 197]. Along with the related MinHashed fingerprint MHFP6, these fingerprints have shown to be highly reliable and efficient in identifying potential drug candidates and predicting their targets.

The numbering of ECFP4 and MHFP6 corresponds to the particular implementation of these algorithms. ECFP4 is a variant of the ECFPs family of fingerprints, where the fingerprints are based on the molecular graph topology up to a certain radius (in the case of ECFP4, up to a

radius of 4 bonds). MHFP6 is a variant of the MHFPs family, where the fingerprints are based on random hashing of molecular substructures with a minimum frequency cutoff (in the case of MHFP6, the minimum frequency cutoff is 6). These numbering schemes help to differentiate between the various versions of these algorithms and provide a reference for researchers using these fingerprints in their studies.

For our work, we chose to use the ECFP4 fingerprint algorithm due to its proven performance and efficiency in small molecule virtual screening and target prediction benchmarks [198, 199, 200]. In the next section, we will discuss the synthetic accessibility assessment method we used for our HTVS workflow.

3.4.6 Synthetic Accessibility Assessment

Synthetic accessibility refers to the ease or difficulty of synthesising a specific chemical compound using currently available chemical reactions and methodologies. Given the vastness of chemical space [159, 161, 160, 13, 162, 163], it is not feasible to synthesise and test every possible compound. Therefore, synthetic accessibility is a measure of the practicality of the synthetic route to obtain a desired molecule.

There are several methods that are used to calculate the synthetic accessibility score of molecules, such as Synthetic Complexity score (SCScore) [201], synthetic accessibility score (SAScore) [202] and SYnthetic Bayesian Accessibility (SYBA) [203]. These methods cannot compare to sophisticated synthetic path-reconstruction methods [204, 74] that can predict and optimise chemical reaction pathways to synthesise a target molecule. However, they are more computationally intensive than fragment-based approaches like SCScore, SAScore and SYBA, thus we will be discussing these.

SCScore is a data-driven metric designed to describe real syntheses and assess the synthetic accessibility of a molecule [201]. It is based on the idea that reaction products are synthetically more complex than reactants. SCScore assigns a synthetic complexity score between 1 and 5 to a molecule using a neural network. The neural network was trained on 22 million reactant-

product pairs from the Reaxys database [205].

SAScore [202] is based on the analysis of ECFP4[193] fragments obtained from one million compounds randomly selected from the PubChem database [206]. Each fragment is assigned a numerical score, where frequent fragments have positive scores and less frequent fragments have negative scores. SAScore also includes a complexity penalty and symmetry bonus, which penalise nonstandard structural motives such as macrocycles, stereo centres, spiro and bridge atoms, but reward the symmetry of a structure. The method produces values between 1 (easy to make) and 10 (very difficult to make), where 6.0 is suggested by the authors as a threshold to distinguish between easy- and hard-to-synthesise compounds.

SYBA [203] is a fragment-based computational method that predicts the feasibility of constructing a molecule from simpler fragments based on their structural and chemical properties. Using a Bayesian statistical framework, prior knowledge and uncertainty are incorporated into the prediction. The model was trained on easy-to-synthesise (ES) molecules from the ZINC15 database [207, 208] and hard-to-synthesise (HS) molecules generated by the Nonpher methodology [170, 209].

The SYBA method has been shown to provide comparable or better performance than other methods like SAScore and SCScore while being less computationally demanding and providing a more straightforward analysis of individual fragment contributions [203]. Specifically, in a study comparing SYBA to SAScore and SCScore, the authors recommended using SYBA over SAScore due to its smaller complexity and more straightforward analysis and noted that SYBA outperformed SAScore when the threshold value was set to 6.0 but had comparable accuracy when the threshold was reduced to ~ 4.5. Additionally, all three methods (SYBA, RF/ECFP4, and SAScore) substantially outperformed SCScore in the same study. Therefore, we have chosen to use SYBA for our HTVS workflow as it has been shown to provide a reliable and efficient classification of organic compounds.

SYnthetic Bayesian Accessibility

The SYBA method provides an accessibility score, which may be positive or negative, depending on the synthetic accessibility of the molecule. A positive score indicates easier synthesis, while a negative score indicates a more challenging synthesis. The magnitude of the score reflects the molecule's ease of synthesis, with larger positive scores indicating easier synthesis and larger negative scores indicating more challenging synthesis.

As previously mentioned, the calculation of SYBA value is based on the Bayes theorem expressed in equation 3.2 [203]. Each compound is described by a binary fingerprint represented as $\mathbf{F} = [f_1, f_2, \cdots, f_M]$, where the length of the fingerprint is denoted by M. The binary values f_i (fragment i) in the fingerprint indicate the presence ($f_i = 1$) or absence ($f_i = 0$) of various molecular building blocks or substructures in the molecule. SYBA uses this fingerprint to assign the molecule to a class $C \in \langle ES, HS \rangle$.

$$p(C|\mathbf{F}) = \frac{p(\mathbf{F}|C)p(C)}{p(\mathbf{F})} \tag{3.2}$$

In equation 3.2, the term $p(C|\mathbf{F})$ represents the posterior probability that a compound, given a specific set of molecular fragments $\mathbf{F}$ belongs to a certain class C. The likelihood denoted as $p(\mathbf{F}|C)$ refers to the conditional probability that a compound from class C contains the set of molecular fragments $\mathbf{F}$. On the other hand, the marginal probabilities $p(\mathbf{F})$ and $p(C)$ express our overall beliefs or probabilities of observing a set of molecular fragments $\mathbf{F}$ and the occurrence of a molecule belonging to class C respectively.

The SYBA score is defined as the logarithm of the ratio of the posterior probabilities that the molecule belongs to the ES and HS classes [203], as expressed in equation 3.3.

$$SYBA(\mathbf{F}) = \ln\left(\frac{p(ES|\mathbf{F})}{p(HS|\mathbf{F})}\right) \tag{3.3}$$

By applying equation 3.2, the SYBA score can be expressed as equation 3.4.

$$SYBA(\mathbf{F}) = \ln\left(\frac{p(ES)}{P(HS)}\right) + \ln\left(\frac{p(\mathbf{F})|ES}{p(\mathbf{F}|HS)}\right) \tag{3.4}$$

The SYBA score (equation 3.4) can be simplified to equation 3.5 [203].

$$SYBA(\mathbf{F}) = \sum_{i=1}^{M} s_i(f_i) \tag{3.5}$$

The score contribution $s_i(f_i)$ from fragment i in equation 3.5 is expressed in equation 3.6.

$$s_i(f_i) = \ln\left(\frac{p(f_i|ES)}{p(f_i|HS)}\right) \tag{3.6}$$

In practical applications, the SYBA score generally falls between -100 and $+100$, according to the authors. However, theoretically, it can assume any value between plus and minus infinity. It is noteworthy that the absolute value of the SYBA score does not measure the degree of synthetic accessibility but rather the level of confidence in the prediction.

Although SYBA has not been specifically tested on TADF molecules, the method's ability to predict synthetic accessibility has broad applications in guiding the design and synthesis of novel compounds. In particular, SYBA scores can be used as a useful tool to filter out molecules that are difficult or expensive to synthesise, potentially saving time and resources in the synthesis and characterisation process.

One advantage of using SYBA scores to screen a molecular library for TADF design is that it can help prioritise compounds that are predicted to be easier to synthesise. This can increase the efficiency of the synthesis and characterisation process and may lead to the discovery of novel TADF molecules that would otherwise have been overlooked.

However, SYBA has certain limitations that merit consideration. First, the accuracy of the SYBA model heavily relies on the quality and diversity of the training data used to estimate the prior probabilities. Additionally, its simplified binary fingerprint representation may not fully capture the complexities of molecular structures, potentially leading to inaccuracies. The

model's static nature might also hinder its adaptability to evolving synthetic methodologies. Moreover, SYBA does not consider specific reaction pathways, overlooking critical synthetic considerations. Rigorous validation of diverse datasets and chemical classes is necessary to assess their specificity and generalisation capabilities. While SYBA provides valuable insights, complementary approaches and considerations, such as quantum mechanical effects, may be warranted to comprehensively assess the synthetic feasibility of novel compounds.

While it is important to acknowledge the limitations of using SYBA scores, this method still offers valuable insight into the synthetic accessibility of compounds, including those with potential TADF properties. Although the actual synthetic accessibility of a compound may differ from what is predicted by SYBA, these scores can help researchers prioritise compounds that are predicted to be easier to synthesise and save valuable time and resources in the process. However, it is important to balance the use of SYBA scores with the consideration of other factors, such as the potential value of compounds that may be more challenging to synthesise, as well as their overall properties and potential.

Overall, while SYBA scores can be a useful tool for screening molecular libraries for TADF design, it is important to use them in conjunction with other considerations and to carefully balance the advantages and limitations of this approach. The details of the chosen threshold value will be explained in chapter 4.

3.4.7 Force Field Geometry Optimisation

In our HTVS workflow, subsequent to filtering the molecules in previous stages, the optimisation step was performed by minimising the energy of each molecule. We use MMFF94 as discussed in section 2.10.1 (Chapter 2) with a maximum of 500 steps used to achieve a suitable conformation for subsequent analyses. The number of optimisation steps varies depending on factors such as the size, complexity, and initial geometry of the molecule. While 500 steps may be sufficient to reach a local minimum energy geometry in some cases, others may require more steps. It is important to strike a balance between accuracy and computational cost when selecting the number of steps since increasing the number of steps can escalate computational

costs and the time required for optimisation. Generally, 500 steps are commonly used and a reasonable choice for geometry optimisation in molecular simulations, which produces an acceptable approximation of the minimum energy geometry.

In the next section, we will present the parameters used for our DFT calculations in BigDFT before moving on to the next step of the workflow.

3.4.8 BigDFT Parameters

We performed BigDFT calculations using HGH-GTH PSPs [140, 141] with NLCC [139]. For the ground state, gas phase geometry optimisation was carried out using PBE in BigDFT, with a maximum force threshold of 0.02 eV/Å. To ensure accurate force calculations, a smaller grid spacing of 0.24Åwas employed for geometry optimisations, while single-point calculations used a wavelet grid spacing of 0.26Å. The gradient convergence threshold was set to 10^{-5}. After the single point calculation, we calculated the HOMO-LUMO spatial overlap Λ_T as a post-processing calculation using the HOMO and LUMO wavefunctions [138] (equation 3.7).

$$\Lambda_T = \int |\psi_{HOMO}(\mathbf{r})||\psi_{LUMO}|d\mathbf{r} \tag{3.7}$$

The parameter Λ_T is calculated as the square root of the product of the overlap of solely the HOMO and LUMO wavefunctions (ψ_{HOMO}, ψ_{LUMO}). $\Lambda_T = 0$ indicates no spatial overlap between the HOMO and LUMO, representing a CT excitation. Conversely, $\Lambda_T = 1$ signifies full spatial overlap, corresponding to a LE state. It is essential to note that Λ_T does not distinguish between singlet or triplet excitations and does not consider additional contributions in the case of mixed excitations.

3.4.9 DFT on MMFF94 Geometry Optimised Molecules

We conducted DFT calculations on both the parent molecules and the generated molecules, which we obtained by optimising their geometry using MMFF94. Specifically, we calculated

the HOMO-LUMO energy gap and HOMO-LUMO density overlap Λ_T, and compared these parameters with those of the parent molecules. In Chapter 4, Section 4.6, we discuss the details of this method and the criteria that we used to filter the molecules.

3.4.10 DFT Geometry optimisation

We optimised the geometry of molecules selected according to the criteria established in the previous step, using DFT.

3.4.11 Frontier orbitals and Singlet-Triplet Splittings

We perform single point and ΔSCF calculations on the DFT geometry optimised molecules to obtain the HOMO-LUMO energy gap and singlet-triplet splitting ΔE_{ST}.

3.5 Summary

In summary, we used the STONED algorithm to generate a library of molecules. This algorithm required a SMILES string representation of the molecule as input, which was converted to SELFIES, a newer format that allowed for point mutations. Each point mutation in SELFIES corresponded to a single character addition, deletion, or replacement with a 33% probability for each mutation. The canonical SELFIES representation was then converted back to SMILES.

The study investigated five TADF molecules to explore their chemical space and provide insights towards the development of better TADF molecule designs. ECFP4 was used to calculate the fingerprint of each molecule, and the diversity of the molecule libraries was analysed using Tanimoto similarity. SYBA score was used to filter molecules based on easy-to-synthesise and hard-to-synthesise.

Finally, single-point DFT calculations were performed on MMFF94-optimised structures, and

single-point DFT calculations and ΔSCF calculations were performed on DFT-optimised structures.

Chapter 4

High Through Put Virtual Screening Workflow Results - Part 1

4.1 Introduction

This chapter presents the results of the molecule generation process and force field optimisation of the generated molecules utilised in our HTVS workflow. The primary objective is to create a comprehensive and diverse molecule library to increase the chances of discovering promising candidates. To achieve this objective, the workflow must ensure that the generated molecules meet stringent quality standards. Designing a chemical library for TADF molecules should look at the following key considerations:

1. Diversity: The library should contain a diverse range of compounds that represent the chemical space of interest. This can help to ensure that the library is representative of the chemical space and that it includes compounds with a range of different properties.

2. Relevance: The library should contain compounds that are relevant to the research area of interest, in this case, TADF molecules, with structural and electronic properties conducive to efficient and effective TADF emission.

3. Stability: The compounds in the library should be stable under a range of conditions and should not be prone to degradation.

4. Synthesis accessibility: The compounds in the library should be theoretically synthesisable, meaning that they can be produced in the laboratory using existing chemical synthesis techniques.

These four key considerations overlap in that they are all important factors to consider when designing a library of TADF molecules for research purposes.

4.2 Library Generation

To generate the initial molecule library, we explored the chemical space of the parent molecules presented in Figure 3.2 which were selected based on specific criteria discussed in Chapter 3, Section 3.4.4. We obtained the SMILES representation of these parent molecules and used them to guide the generation of the initial library. Table 4.1 provides the SMILES representation of the parent molecules used in this process. As discussed in Section 3.4.1, Chapter 3, the

Table 4.1: SMILES representation of parent molecules used to generate molecule libraries. N_{atoms} represents the number of atoms in a molecule excluding hydrogen atoms.

Molecule	N_{atoms}	SMILES
PXZ-TRZ	38	`c7ccc(c6nc(c1ccccc1)nc(c5ccc(n4c2ccccc2oc3ccccc34)cc5)n6)cc7`
TXO-PhCz	36	`O=C6c1ccccc1S(=O)(=O)c7ccc(c2ccc4c(c2)c3ccccc3n4c5ccccc5)cc67`
TXO-TPA	36	`O=C5c1ccccc1S(=O)(=O)c6ccc(c4ccc(N(c2ccccc2)c3ccccc3)cc4)cc56`
GOBVUP	24	`Clc4ccc(c3csc(=c2scc(c1ccc(Cl)cc1)s2)s3)cc4`
TUFWAS	30	`c6csc(c4sc(=c3sc(c1cccs1)c(c2cccs2)s3)sc4c5cccs5)c6`

STONED algorithm is capable of generating thousands of canonical molecules in a matter of minutes [172]. However, there is a limit to the number of canonical molecules that can be generated for a given number of molecules. To determine this limit, it is useful to estimate the minimum number of molecules required to exhaust the chemical space and obtain canonical molecules. This information can be used to optimise the computational resources needed for molecule generation, by considering factors such as the size and number of parent molecules.

Our observations (see Figure 4.1 and Table 4.2) indicate that for parent molecules with sizes ranging from 24 to 38 atoms (excluding hydrogen atoms), generating a chemical space of approximately 10^4 to 10^5 molecules maximises the number of canonical molecules.

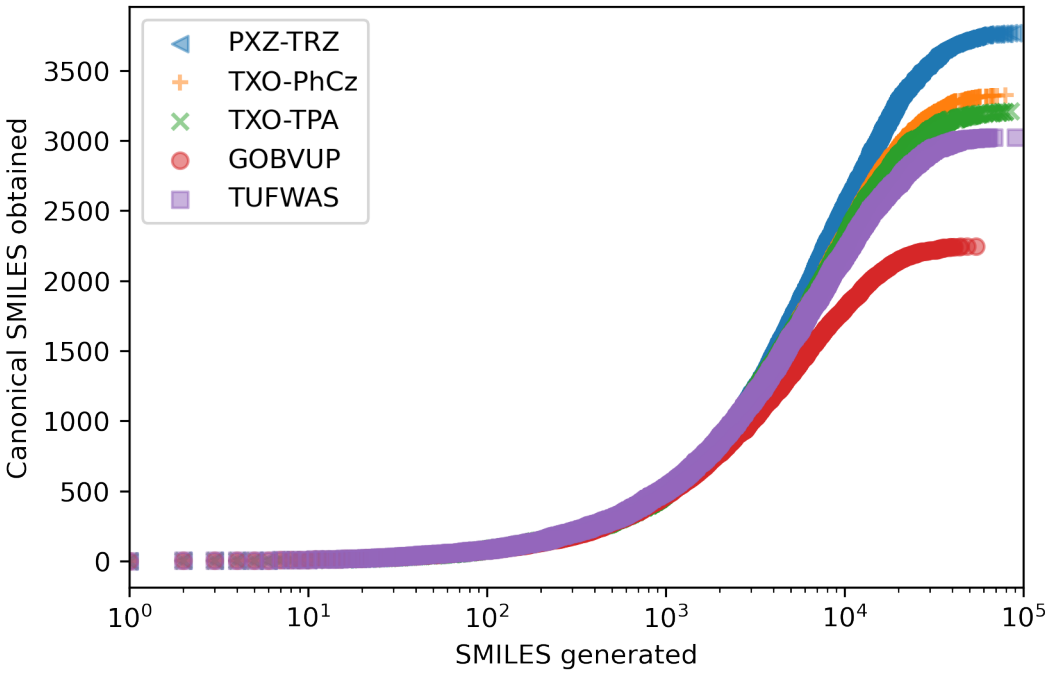


Figure 4.1: Relationship between total SMILES generated from parent molecules and canonical SMILES obtained.

To ensure the reliability of our results, we conducted a verification step where we investigated the relationship between the size of molecules and the number of canonical SMILES generated. As illustrated in Figure 4.2 and summarised in Table 4.2, we observed a positive correlation between the two variables. Specifically, our findings indicate that longer SMILES strings offer more opportunities for point mutations, thereby expanding the chemical subspace and increasing the likelihood of generating canonical molecules. This verification step provides a rigorous validation of our methodology and supports the validity of our conclusions.

After generating the library of canonical SMILES, it is essential to assess its diversity to ensure a thorough exploration of the chemical space. In the following section, we will discuss the method used to verify the library's diversity.

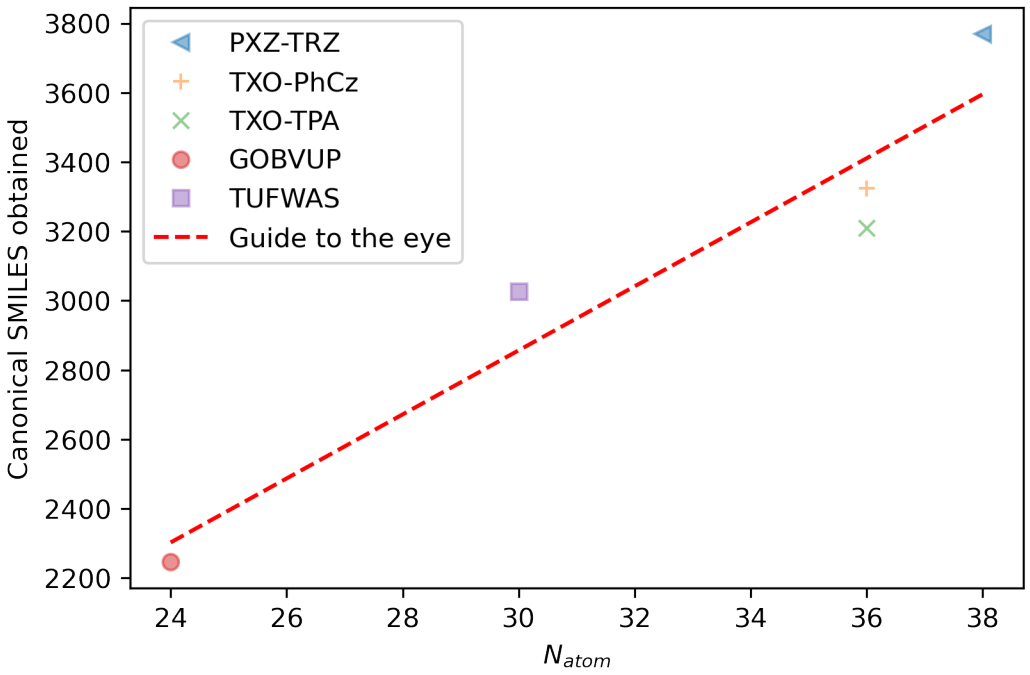


Figure 4.2: Plot between parent molecule size (number of atoms N_{atoms}) and canonical SMILES obtained.

Table 4.2: Total number of SMILES from parent molecules and obtained the corresponding canonical SMILES. The percentage of canonical SMILES obtained is calculated as the ratio of the number of canonical SMILES to the total number of SMILES generated.

Molecule	Canonical SMILES (percentage)	SMILES generated
PXZ-TRZ	3769 (2.79%)	135000
TXO-PhCz	3325(2.89%)	115000
TXO-TPA	3210 (2.38%)	135000
GOBVUP	2247 (2.37%)	95000
TUFWAS	3026 (2.88%)	105000
YAFNOI	1925 (1.83%)	105000

4.3 Molecule Library Check

The canonical molecules generated through the HTVS process should be evaluated for their diversity, relevance, stability, synthesis ability, and predictive power, as discussed in Section 4.1. We believe that the first and foremost requirement is to assess the diversity of the molecules, as this can provide insight into the overall composition of the library. The relevance of the library can then be directly evaluated based on its diversity in terms of TADF molecules.

After assessing the diversity of the generated molecules, the order in which other requirements are checked will depend on the specific goals and applications of the study. For our work, the stability of the molecules was the next important factor to consider. We evaluated each molecule's system type and determined whether it was a closed- or open-shell system. While open-shell TADF molecules offer certain advantages such as a higher spin density resulting in enhanced luminescence and longer excited state lifetimes, leading to potentially higher efficiency and stability of the TADF device, closed-shell TADF molecules have their own benefits. For instance, closed-shell TADF molecules are generally more stable and less reactive than their open-shell counterparts, which can be advantageous for device applications that require long-term stability. Additionally, closed-shell TADF molecules can exhibit excellent performance in certain applications, particularly in cases where their electronic properties match well with the device requirements.

After selecting only the closed-shell systems, the synthesis ability of those molecules was evaluated using SYnthetic Bayesian Accessibility (SYBA) scores (discussed in Section 3.4.6). The chosen threshold value for the SYBA score will be detailed and explained later in this chapter.

4.3.1 Diversity and Relevance

As discussed in Chapter 3, Section 3.4.5, we chose the Tanimoto coefficient to measure and analyse the diversity of our molecule libraries. We did not use other methods, such as molecular descriptors, as it is not easy to identify the donor or the acceptor moiety within type-VII parent molecules, which are highly symmetric. For control, we used the same diversity check for the

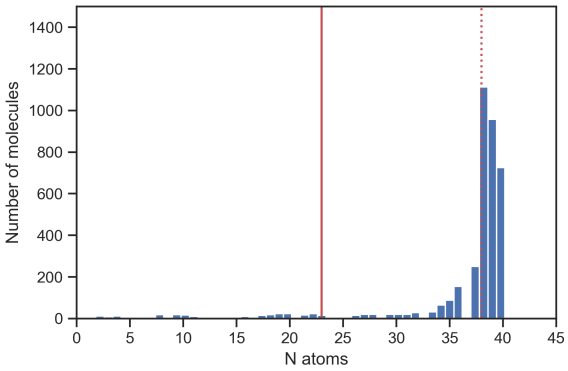
traditional D-A TADF parent molecules, which allowed us to compare the performance of our method with previous studies.

Other methods for diversity analysis often rely on identifying molecular substructures, such as functional groups, that are important for biological activity or chemical properties. These substructures can then be compared and used to assess molecular diversity. However, for highly symmetric molecules like the type-VII parent molecules used in our study, it can be difficult to identify specific substructures that are relevant for diversity analysis. This is because the symmetric nature of these molecules means that they have multiple equivalent substructures that cannot be easily distinguished.

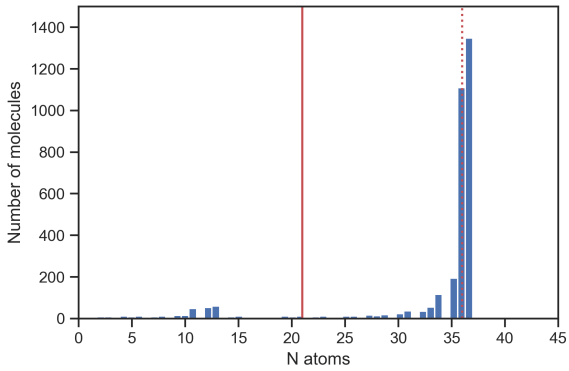
Therefore, for our study, we chose to focus on the Tanimoto coefficient as a measure of diversity, since it is based purely on the molecular structure and does not require the identification of specific substructures. Furthermore, our analysis revealed that the STONED algorithm typically generates a few molecules with significantly fewer atoms than their parent compounds, leading to low fingerprint scores that indicate their insignificance. Nevertheless, we aimed to incorporate molecules with low fingerprints similar in size to the parent molecule. Therefore, we filtered the molecules based on their atom size. Including low-fingerprint molecules that are similar in size to the parent molecule can help provide a more comprehensive representation of the chemical space being analysed, increase sensitivity, and avoid bias. Additionally, it can lead to a more subtle understanding of the results and provide insights into potentially interesting relevant molecules.

Molecule size

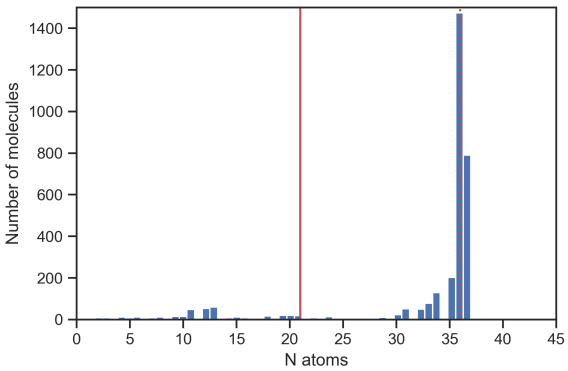
Our analysis (Figure 4.3) showed that most of the canonical molecules generated have sizes similar to those of the parent molecules. However, there are a few exceptions with significantly fewer atoms than the parent molecules. These small molecules are located far from the parent molecules in chemical space and are expected to have limited value for subsequent HTVS analyses. Therefore, we recommend that these small molecules be excluded from the library before the HTVS process, to ensure that the resulting hits are more likely to be structurally



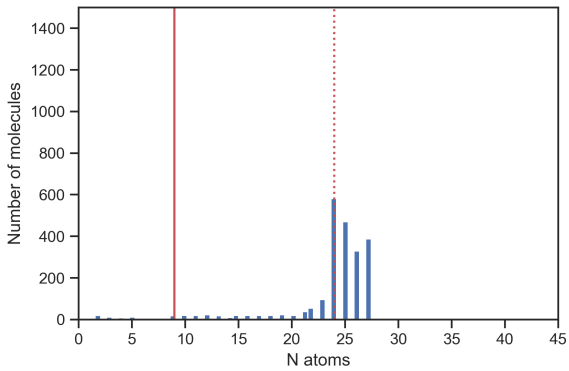
(a) PXZ-TRZ



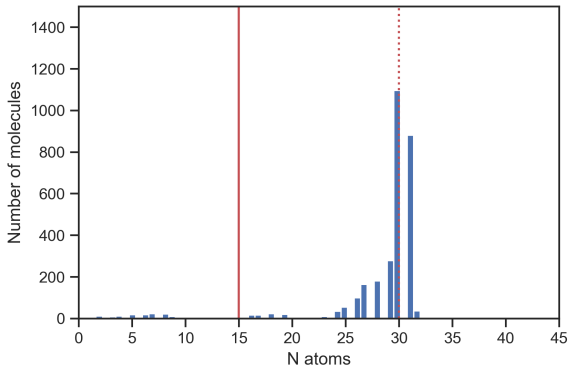
(b) TXO-PhCz



(c) TXO-TPA



(d) GOBVUP



(e) TUFWAS

Figure 4.3: Molecules distribution by the number of atoms (N_{atoms}), showing the frequency of molecules with respect to the number of atoms they contain. The solid red vertical line represents the threshold for excluding molecules below a certain size, while the dotted red line indicates the size of the reference parent molecule.

similar and chemically relevant to the parent molecules.

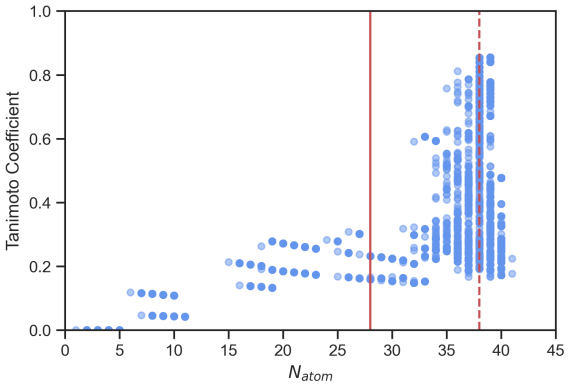
To determine the cut-off based on the number of atoms N_{atom} (molecule size), we compared the fingerprint scores and molecule size (Figure 4.4). While smaller molecules typically have lower fingerprint scores, we also observed some molecules with similar sizes to the parent molecules but with significantly lower fingerprint scores.

However, these molecules are still likely to be structurally similar and chemically relevant to the parent TADF molecules due to their similar size. Therefore, we decided to filter the molecules based solely on their number of atoms (size), using the equation 4.1, to ensure that only molecules with a size similar to that of the parent TADF molecules are retained in the library for subsequent HTVS analyses.

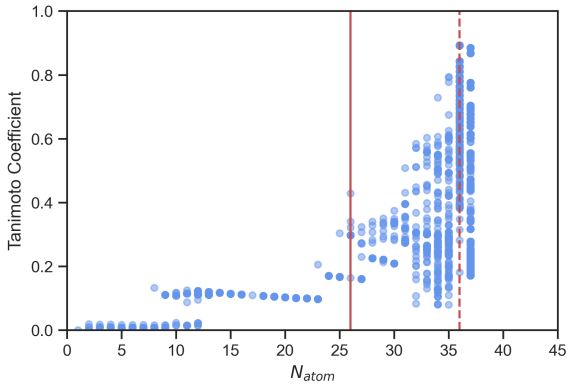
$$N_{atom}^{threshold} = N_{atom}^{parent} - 10 \tag{4.1}$$

To determine the cutoff based on the number of atoms in the molecules, we compared the fingerprint scores and molecule sizes. As expected, smaller molecules generally had lower fingerprint scores. However, we also observed that some molecules with similar or close-to-similar sizes as the parent molecules had substantially lower fingerprint scores. These findings suggest that molecule size alone cannot be relied upon to identify potential compounds for HTVS analyses.

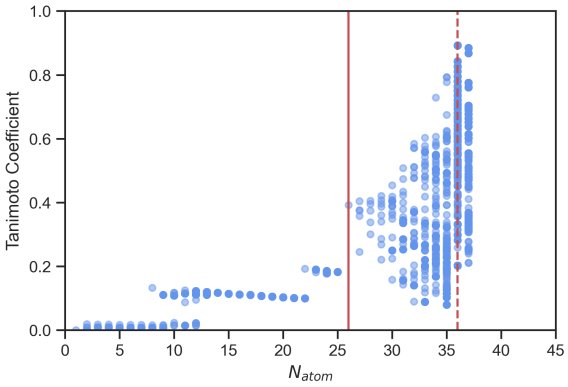
Therefore, we used Equation 4.1 to filter the molecules, and we chose a threshold value of 10 based on our analysis of the Tanimoto coefficient and molecule size N_{atom}. The value of 10 was selected based on the data shown in Figure 4.4. This approach has advantages and disadvantages. One benefit is that it eliminates small, chemically dissimilar molecules that are unlikely to have significant value in subsequent HTVS analyses. Moreover, by requiring a minimum size of $N_{atom}^{threshold}$, it ensures that the library contains a diverse set of molecules with structural similarities to the parent compounds.



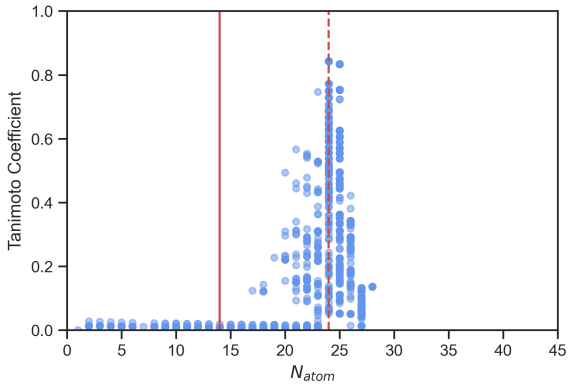
(a) PXZ-TRZ



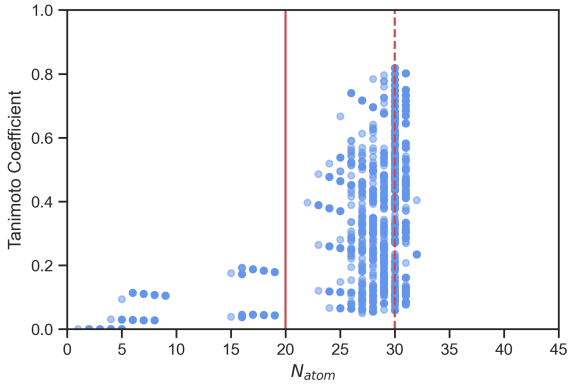
(b) TXO-PhCz



(c) TXO-TPA



(d) GOBVUP



(e) TUFWAS

Figure 4.4: Molecular size and fingerprint (FP) score displayed in a scatter plot. The solid red line denotes the established threshold for filtering out SMILES, while the dashed red line represents the size of the parent molecules used as a reference point.

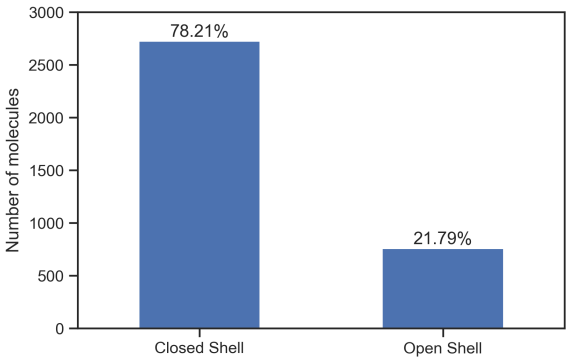
4.3.2 Stability

The stability of molecules is a crucial consideration in the design of chemical libraries for various applications, including TADF molecules. To examine the stability of TADF molecules, we classified them based on their electronic configuration into two groups: open-shell and closed-shell. Organic radicals are highly reactive molecules [210] composed of light elements such as hydrogen (H), carbon (C), nitrogen (N), oxygen (O), and sulfur (S) that have at least one unpaired electron in their outermost shell. This unpaired electron makes them open-shell systems that readily participate in chemical reactions such as hydrogen abstraction, dimerisation, or recombination, leading to a loss of the open-shell character.

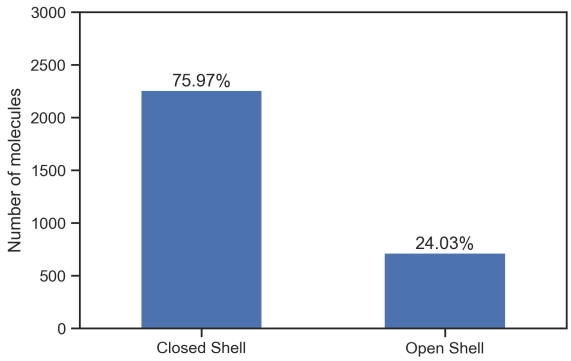
Closed and Open-shell System

This observation is necessary for a molecules library in HTVS because open-shell molecules may have different electronic and optical properties compared to closed-shell molecules, which can affect their performance as TADF materials. Therefore, knowing the percentage or number of open-shell molecules in the library can help in designing and selecting suitable molecules for TADF applications.

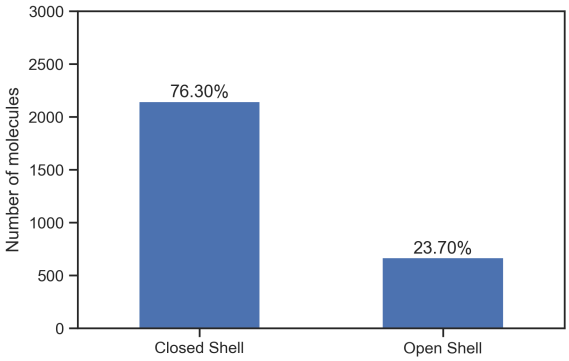
The presence of open-shell systems in a molecule library can significantly affect the outcome of various computational chemistry calculations downstream in an HTVS workflow. For example, open-shell systems often require different computational methods and more sophisticated basis sets to accurately describe their electronic structure [211, 212, 213, 214, 215], which can significantly increase the computational cost and time required for these calculations. In addition, the presence of open-shell systems can also affect the geometry optimisation process [216], as they may exhibit different structural characteristics compared to closed-shell systems, which may impact the final predicted properties of the molecule. Therefore, it is important to take into consideration the electronic configuration of the molecules in a library to ensure that appropriate computational methods and basis sets are used and that accurate and reliable results are obtained.



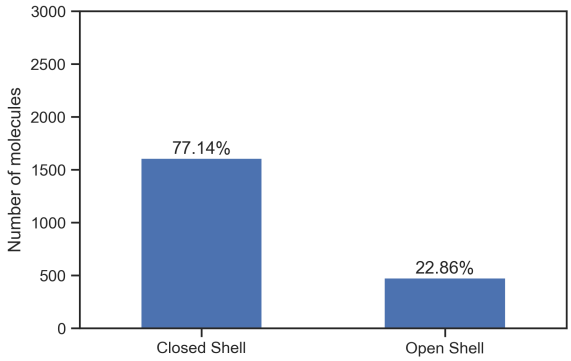
(a) PXZ-TRZ



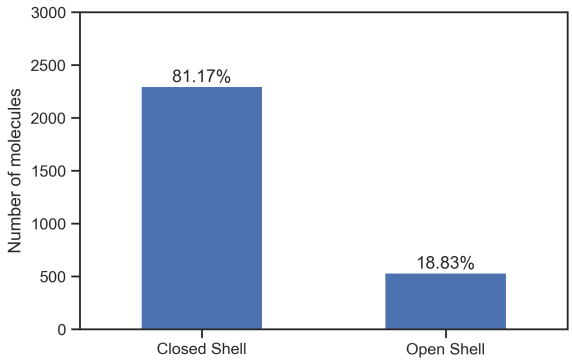
(b) TXO-PhCz



(c) TXO-TPA



(d) GOBVUP



(e) TUFWAS

Figure 4.5: Number of canonical molecules that are closed and open shell.

Figure 4.5 that the majority of the molecules generated through the HTVS workflow resulted in closed-shell species. Specifically, less than $\lesssim 30\%$ of the molecules for each parent molecule were found to be open-shell. This result may have implications for subsequent computational steps in the HTVS workflow, particularly for geometry optimisation and DFT calculations. The presence of unpaired electrons in open-shell systems may lead to difficulties in geometry optimisation and convergence, as well as the additional computational expense associated with DFT calculations. Therefore, this observation should be taken into account when selecting molecules for further study and optimisation.

Studies have revealed the presence of stable neutral radical emitters in OLEDs [210, 217, 218]. To investigate the potential impact of open-shell molecules on the final results for our candidate molecules, we specifically chose to include them in our study, utilising the methods we employed.

4.4 Tanimoto Coefficient

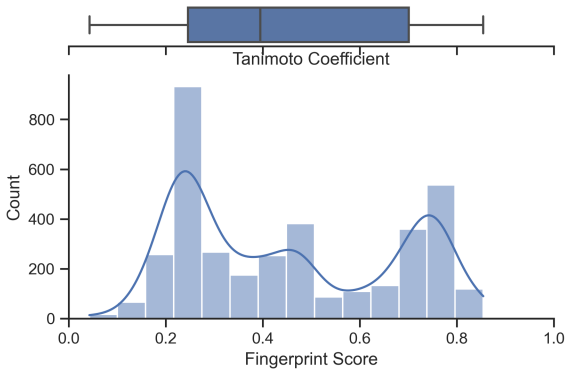
Maintaining diversity in a molecular library is essential to ensure a successful high-throughput virtual screening (HTVS) process. After applying the $N_{atom}^{threshold}$ criteria specified in Section 4.3.1 to filter out certain molecules, we performed an analysis to assess the diversity of the remaining molecules in the library. Specifically, we calculated the Tanimoto coefficient between all pairs of molecules, as well as between each molecule and its parent molecule. The Tanimoto coefficient is a measure of similarity that enabled us to evaluate the differences and similarities between the molecules.

As discussed in chapter 3, a Tanimoto coefficient of 1 indicates an exact match between two molecules, while a score of 0 means there is no similarity between them. A score of 0.5, therefore, represents a moderate level of similarity, indicating that the molecules share some structural features but are not identical. The reason why the range is commonly used as a threshold for identifying structurally similar molecules is that it strikes a balance between including molecules that are sufficiently similar to the parent molecule(s) and excluding those that are too dissimilar.

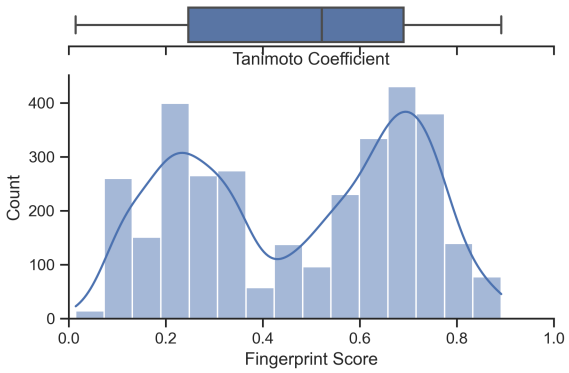
The analysis in Figure 4.6 showed that the remaining molecules in the library were diverse, with Tanimoto coefficient ranging from 0 to 1. This indicated that the library contained a wide range of molecules that were both structurally similar and distinct from the parent molecules. Moreover, this range of coefficients is also compatible with the concept of molecular diversity, where a diverse library should contain molecules that are both similar and dissimilar from the parent molecules. By including molecules with similarity scores between 0 and 1, the library can capture a range of structural variations that may be important for achieving the desired activity or property in the screening process.

Including molecules with a low Tanimoto coefficient (< 0.5) can have several advantages for designing Type-VII TADF molecules. One of these advantages is the increased likelihood of discovering molecules with canonical structural features that may lead to improved TADF properties. Furthermore, it can help to expand the diversity of the library, improving the chances of finding hits during HTVS. However, this approach also has potential drawbacks. For instance, it may increase the computational cost and time required for subsequent HTVS analyses and may increase the likelihood of including ineffective TADF materials. Thus, it is crucial to carefully balance the benefits and drawbacks of including molecules with lower similarity scores to optimise the molecular library for designing Type-VII TADF molecules. As this is the first iteration of using this method, it can be further refined by analysing the final output of the HTVS workflow. Furthermore, Figure 4.7 shows that there was no significant clustering of molecules in the library, indicating that the library was not biased towards any particular structural motif.

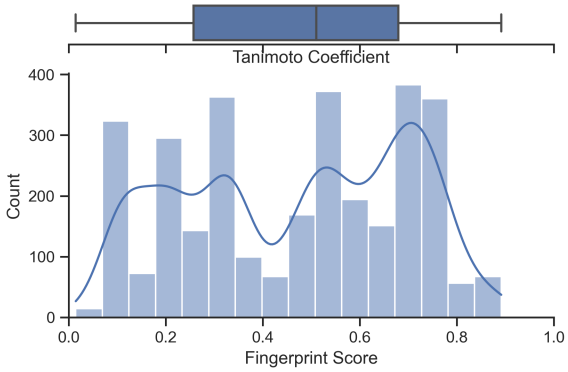
The ability of STONED to explore chemical space was evaluated by analysing the impact of excluding small molecules with $N_{atom}^{threshold}$ below a specified threshold. The analysis of the remaining molecules in the library demonstrated that the exclusion of these molecules did not significantly affect the library's overall diversity, indicating the effectiveness of STONED in navigating chemical space.



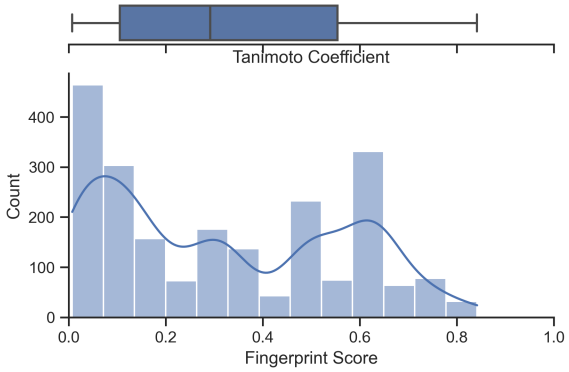
(a) PXZ-TRZ



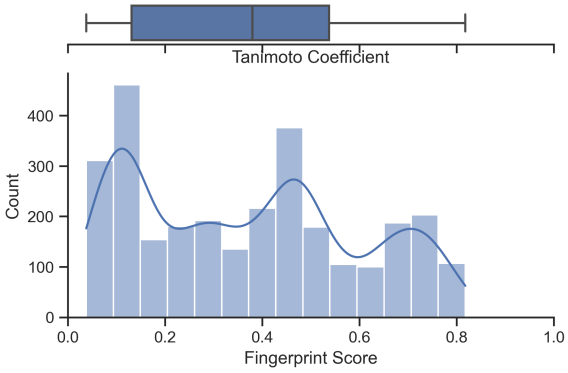
(b) TXO-PhCz



(c) TXO-TPA

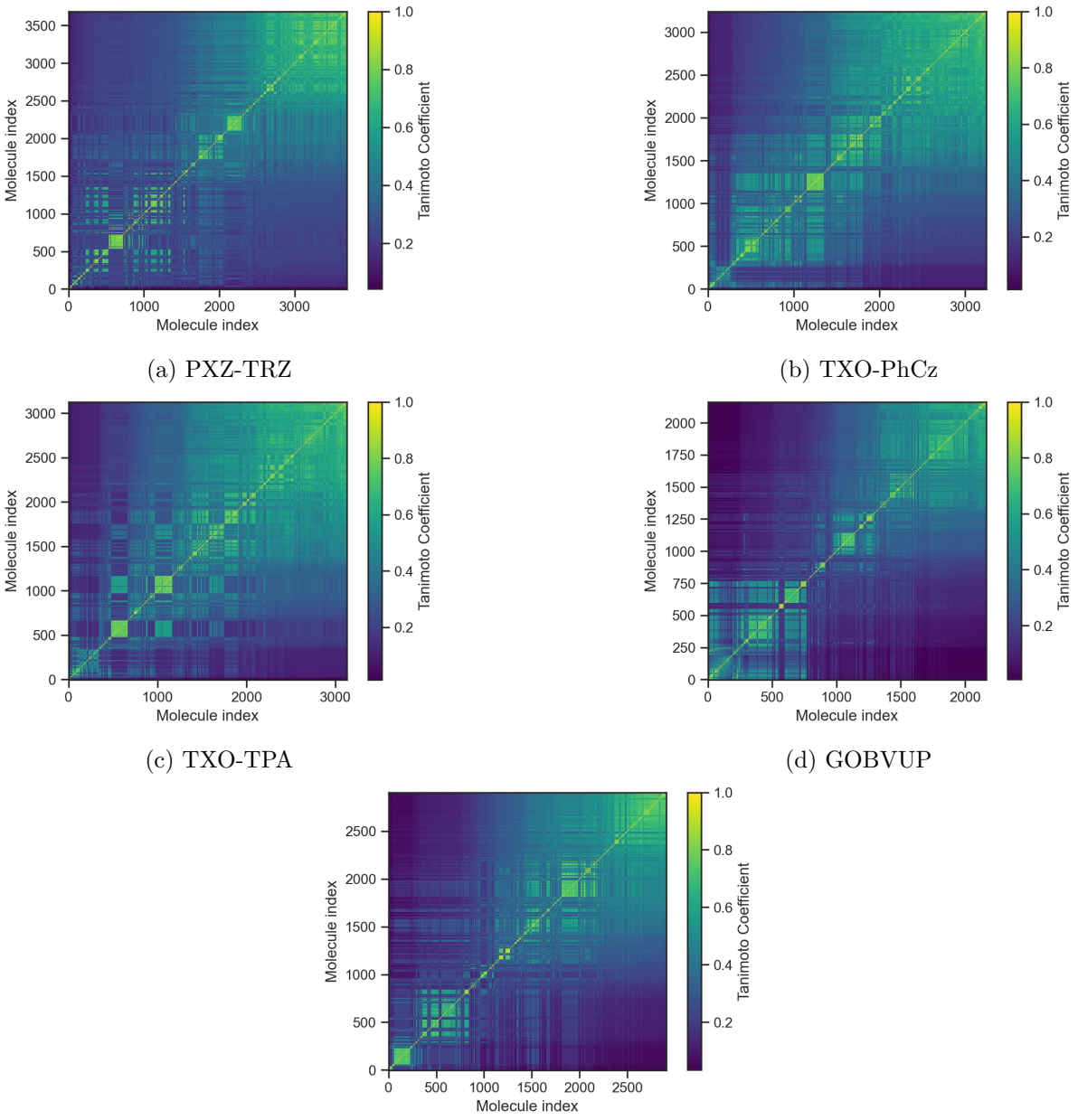


(d) GOBVUP



(e) TUFWAS

Figure 4.6: Visualisations of the SMILES distribution, post-filtering of small and open-shell systems, including a histogram, boxplot, and kernel density estimate (KDE) plot (blue line), as discussed in the preceding section of this chapter.



(a) PXZ-TRZ

(b) TXO-PhCz

(c) TXO-TPA

(d) GOBVUP

(e) TUFWAS

Figure 4.7: 2D similarity maps showing graphical representations of the Tanimoto coefficient between all molecules. A high coefficient indicates that the molecule is more similar to the parent molecule, while a low score indicates greater dissimilarity.

4.5 Synthesis accessibility

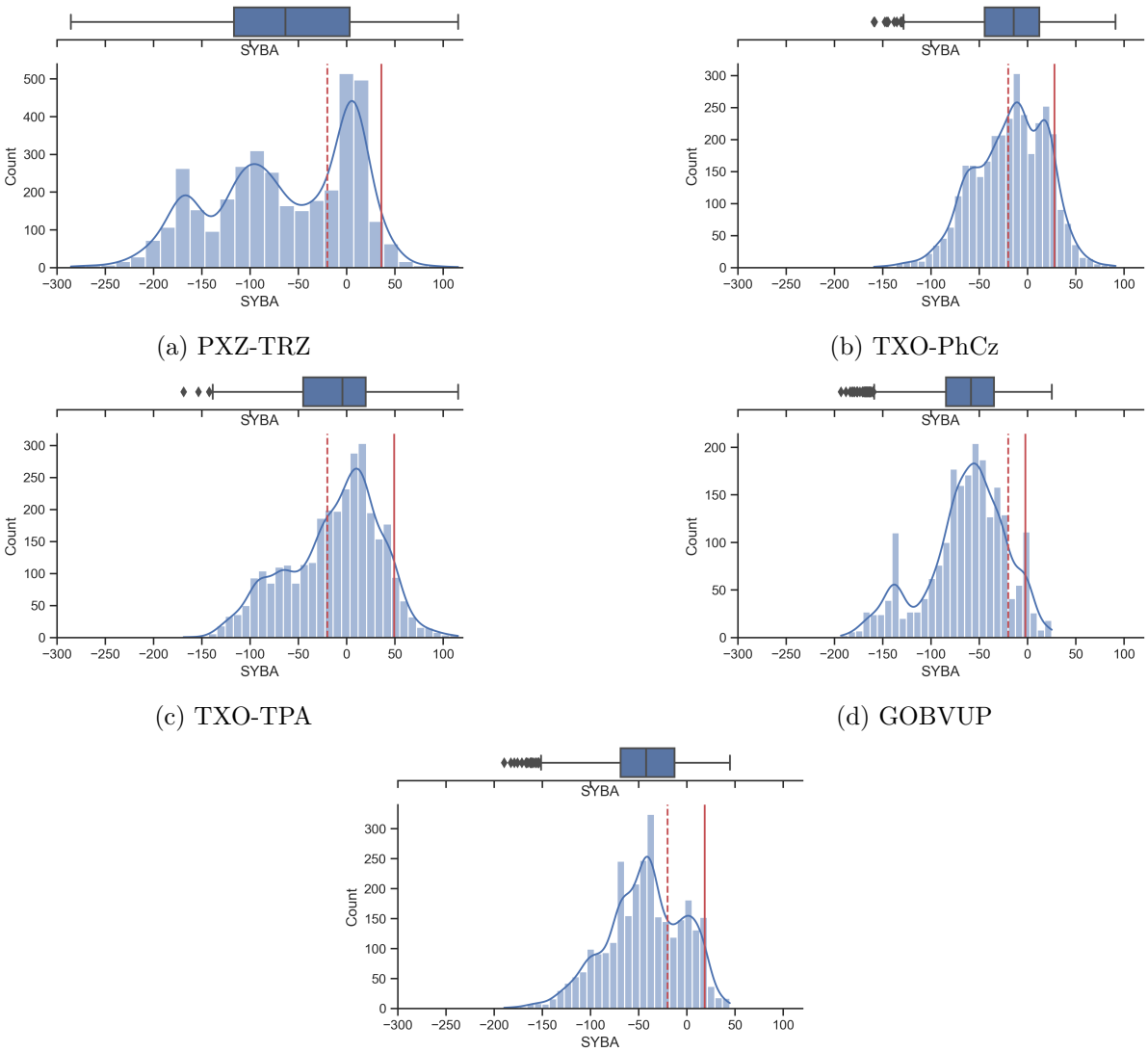
Synthesis accessibility is an important consideration when designing a chemical library for TADF molecules. Compounds that are theoretically synthesisable, using existing chemical synthesis techniques, are preferred as they can be produced in the laboratory with ease. However, the prediction of synthetic accessibility is not always straightforward, especially for large and complex molecules. One method for evaluating the synthetic accessibility of a molecule is by using the SYBA score. This section will explore the use of the SYBA score to assess the synthesis accessibility of the TADF molecules in our library.

4.5.1 SYnthetic Bayesian Accessibility Score

As discussed in Section 3.4.6, the SYBA score [203] can be a useful tool for screening molecular libraries for TADF design. Figure 4.8 shows that the majority of molecules in the library have positive SYBA scores, indicating that they are relatively easy to synthesise. Additionally, most molecules have SYBA scores of > -100 and $< +100$, indicating that they are practical for use in applications, as suggested by the original paper.

In this study, we chose to filter our molecular library based on the SYBA score threshold of ≥ -20. This was done to ensure that the molecules we select for further investigation are likely to be experimentally accessible, as a higher SYBA score implies that a molecule is easier to synthesise. Additionally, the SYBA score is calculated based on a number of factors, such as bond formation and aromaticity, which are important for designing TADF molecules. By selecting molecules with an SYBA score of ≥ -20, we aim to improve the likelihood of identifying promising TADF materials that can be experimentally realised with reasonable effort. The obtained value is in close proximity to the -18.6 threshold value reported in the original paper, which was determined computationally using a test set [203].

It is noted that the parent molecule GOBVUP has an SYBA value in close proximity to zero, albeit negative. This observation does not necessarily imply any concerns regarding the accuracy of SYBA for this molecule. As a predictive measure of synthetic accessibility, a value



(a) PXZ-TRZ

(b) TXO-PhCz

(c) TXO-TPA

(d) GOBVUP

(e) TUFWAS

Figure 4.8: Molecule distribution by SYnthetic Bayesian Accessibility (SYBA) score, showing the frequency of molecules with respect to their SYBA score. The solid red line indicates the SYBA score of the parent molecule, while the dotted red line represents the threshold of -20 for SYBA score

close to zero suggests relative ease of synthesis, while a negative value implies higher synthetic challenges.

The bias towards negative SYBA, as observed in figure 4.8, values observed in the molecules generated by STONED suggest that a significant proportion of the generated molecules may be synthetically challenging. The SYBA score is a measure of synthetic accessibility, with more negative scores indicating lower accessibility. The reduced percentage of molecules remaining after filtering for the type-VII family compared to the D-A family supports this interpretation, as it implies that type-VII molecules may be more challenging to synthesise than D-A molecules.

While the observed bias towards negative SYBA values may be viewed as a limitation of STONED, it is worth noting that the primary objective of the program is to explore chemical space and identify promising compounds for further investigation. STONED does not explicitly incorporate synthetic accessibility during molecule generation, which could explain the observed bias. These findings suggest that STONED generate molecules that are hard-to-synthesise. Therefore, it is necessary to complement the use of STONED with other tools, such as SYBA, that consider synthesise accessibility and other relevant properties to fully evaluate the program's utility in TADF discovery.

4.6 MMFF94 Geometry optimisation

Once an initial library of molecules was generated and filtered based on the previously discussed criteria, their geometries were optimised using the MMFF94 force field. To minimise their energy and prepare them for further analyses (as described below), each molecule was optimised with a maximum of 500 steps.

The number of geometry optimisation steps required for a molecule can vary depending on several factors such as its size, complexity, and initial geometry. In some cases, 500 steps may be sufficient to reach a local minimum energy geometry, while in other cases, more steps may be required. However, using a higher number of steps can also increase the computational cost and time required for the optimisation. Therefore, the choice of the number of steps should

be based on a balance between accuracy and computational cost. In general, 500 steps is a common and reasonable choice for geometry optimisation in molecular simulations and is often sufficient to obtain a good approximation of the minimum energy geometry.

The table presented in 4.3 clearly demonstrates the impact of MMFF94 force field optimization on the number of molecules considered in this study. Our findings show that some molecules, specifically PXZ-TRZ and TXO-PhCz, experienced a significant reduction in number from the initial library due to the inability to set up the MMFF94 force field for these compounds. This difficulty in setting up the force field can be attributed to several factors, including the presence of unusual chemical functionalities, large molecular size, or insufficient structural information. Notably, all molecules without this issue were successfully optimised within 500 steps.

The STONED algorithm is a useful tool for generating initial structures, but it is not a substitute for accurate experimental or theoretical data. It is possible that some of the structures generated by STONED were not chemically reasonable or had unusual conformations that could not be properly optimised using the MMFF94 force field.

Overall, it is likely that both factors may have played a role in the inability to set up the MMFF94 force field for certain molecules, and further investigation or modification of the initial structures and force field parameters may be necessary to improve optimisation results.

Table 4.3: The number of molecules with an SYBA score of -20 or greater, along with the number of molecules successfully optimized using MMFF94.

Parent Molecule	PXZ-TRZ	TXO-PhCz	TXO-TPA	GOBVUP	TUFWAS
Filter Stage	Number of Molecules				
SYBA ≥ -20	1321	1581	1660	257	853
MMFF94	37	397	72	170	204

4.7 Summary and Conclusion

The aim of this part of the workflow was to generate a diverse and comprehensive library of molecules to increase the likelihood of identifying potential TADF candidates. STONED

software was used to generate libraries of molecules based on TADF parent molecules for further investigation. It was observed that by generating around 10^4 to 10^5 molecules for each parent molecule with a size ranging from 24 to 38 atoms (excluding hydrogen atoms), the number of canonical molecules was maximised. Those canonical molecules were then filtered based on a number of atoms using a specific threshold (equation 4.1). After analysing the number of molecules that were closed- and open-shell, we observed that around 30% exhibited the latter property. Based on this finding, we decided to include open-shell molecules in order to assess their potential impact on the final results for our candidate molecule. The molecules were then rated in terms of easy-to-synthesis (ES) and hard-to-synthesis (HS) by using the SYBA method. Molecules with SYBA score ≥ -20 were chosen. Finally, MMFF94 geometry optimisation was performed on the filtered molecules. The number of molecules remaining after each step of the process is summarised in table 4.4. These findings can inform the design of future libraries of molecules for similar areas of research.

Table 4.4: Number of SMILES at each filtering stage in the high-throughput virtual screening workflow discussed in this chapter.

Parent Molecule	PXZ-TRZ	TXO-PhCz	TXO-TPA	GOBVUP	TUFWAS
Filter Stages	Number of Molecules				
Total Generated	135000	115000	135000	95000	105000
Canonical SMILES	3769	3325	3210	2247	3026
$N_{atoms}^{threshold}$	3484	2971	2814	2087	2831
SYBA ≥ -20	1321	1581	1660	257	853
MMFF94	37	397	72	170	204

There are two immediate limitations in this study. First, the MMFF94 geometry optimisation method employed in this research may not accurately represent the conformational energies of all molecules. Furthermore, the number of molecules was significantly reduced after undergoing MMFF94 geometry optimisation, which was likely caused by challenges in configuring the force field. Both STONED molecule generation and MMFF94 optimisation processes may have played a role in the inability to establish the force field for most molecules. Therefore, to improve optimisation results, it is necessary to explore other options, such as conducting further mutations using STONED and using alternative force field methods.

There are several areas of future work that could be explored based on the findings of this study.

One potential avenue for future research is to further refine the filtering criteria used in this study to generate a more targeted and specific library of TADF molecules. This could involve exploring different criteria for filtering molecules based on factors such as electronic properties or structural characteristics. Additionally, future studies could focus on further characterising the properties and potential applications of the TADF molecules identified through this workflow, such as investigating their optical and electronic properties or exploring their potential for use in organic light-emitting diodes (OLEDs). While the library of molecules generated by STONED using a single mutation in the SELFIES representation was successful in identifying potential TADF candidates, in future work, we would like to explore the use of more than one number of mutations in the SELFIES representation to generate a more diverse and comprehensive library of molecules for TADF research.

In conclusion, the findings of this study highlight the importance of a well-designed molecule library for identifying promising TADF candidates. The generation of a diverse and comprehensive library of molecules can increase the likelihood of identifying potential TADF candidates. The study's approach can be applied to other areas of research and can inform the design of future libraries of molecules for similar areas of investigation. However, it is important to note the limitations of the study, including the reliance on computational methods and the lack of experimental validation. Therefore, future work should involve experimental validation of the identified molecules, as well as the exploration of other computational methods and selection criteria to further improve the generation of molecule libraries.

Chapter 5

High Through Put Virtual Screening Workflow Results - Part 2

5.1 Introduction

In the previous chapter, we described and discussed the first part of our HTVS workflow, where we generated a diverse library of molecules using the STONED software and filtered them based on a number of criteria. In this chapter, we focus on the second part of our workflow, which involves the use of first principle calculations, specifically DFT calculations, to study the optimised molecules from the previous chapter at a deeper level. The aim of this chapter is to understand why the molecules that passed the filtering criteria succeeded and to gain insights that can inform the design of novel TADF molecules. The use of DFT calculations allows us to investigate the electronic properties and excited-state energies of the molecules, which are critical factors in TADF. Overall, this chapter is an important step in our efforts to identify promising TADF molecules and contribute to the development of more efficient OLEDs.

5.2 DFT on Force Field Optimised Geometry

In order to gain a deeper understanding of the potential candidate TADF molecules obtained through the MMFF94 geometry optimisation process, further investigation is required. DFT can provide valuable insights into the electronic structure and properties of the molecules. In this section, we will discuss the process of running single-point DFT calculations on the MMFF94 geometry-optimised molecules. This will allow us to obtain electronic energies and properties such as the HOMO and LUMO energy levels, which can aid in filtering further molecules. In addition, we also calculated HOMO-LUMO spatial overlap (Λ_T) [138].

The HOMO-LUMO energy gap is a crucial parameter in determining the electronic properties of molecules. In particular, it is an important metric for identifying candidate molecules with the potential for TADF, a process that can improve the efficiency of OLEDs. As a key determinant of light absorption and emission, the HOMO-LUMO gap influences the wavelength and colour of light emitted by the molecule, making it essential for OLED design and display technologies.

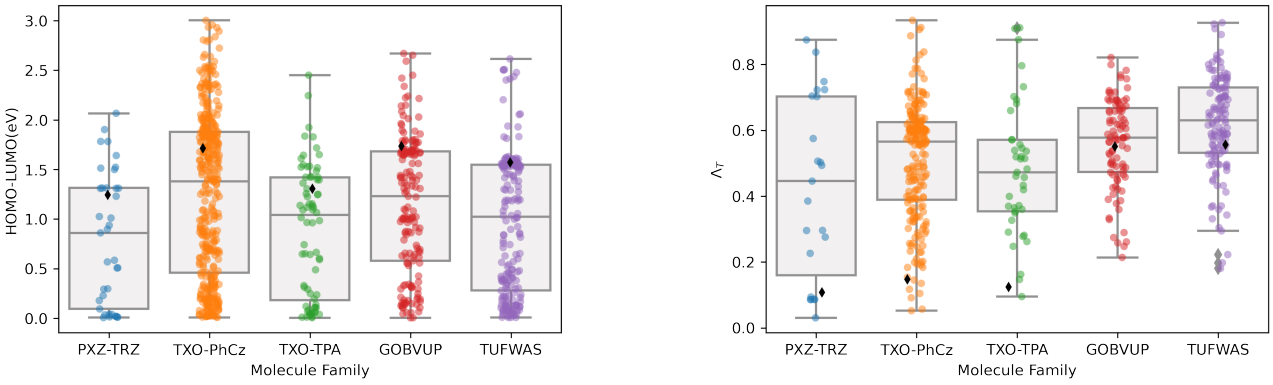
In the context of TADF, the HOMO-LUMO gap plays a pivotal role in determining the ΔE_{ST}. A small energy gap results in a reduced energy difference between the lowest S_1 and the lowest T_1. This characteristic is highly desirable for TADF materials, as it enables the efficient conversion of triplet states to singlet states, leading to delayed fluorescence emission and improved OLED performance. By optimising the HOMO-LUMO energy gap, the radiative recombination rate can be enhanced and the overall emission efficiency of TADF materials, which are critical factors in achieving high-performance OLEDs.

Beyond its impact on light emission and TADF, the HOMO-LUMO gap also influences charge transport properties within the molecule. A narrow energy gap facilitates charge transfer and promotes efficient movement of charge carriers, affecting the electrical conductivity and charge mobility of the material. On the other hand, a large energy gap suggests higher chemical stability, which can be advantageous for material durability and device lifetime.

Considering the relevance of the HOMO-LUMO gap in determining electronic and optical properties, its precise tuning becomes crucial in designing innovative OLED materials. Through

rational molecular design and computational modelling, molecules with an optimal HOMO-LUMO energy gap can be identified, paving the way for the development of highly efficient TADF emitters and the advancement of next-generation OLED technologies. The exploration of novel materials with tailored energy gaps represents a promising avenue for enhancing the performance and versatility of OLED devices for various applications in displays, lighting, and optoelectronics.

One way to predict TADF properties in an HTVS workflow is to calculate the HOMO-LUMO overlap Λ_T, which is related to the probability of exciton transfer between the electron donor and acceptor moieties of a molecule. By using this parameter as a filter in HTVS, it is possible to narrow down the pool of candidate molecules to those with a higher likelihood of exhibiting TADF properties, thereby reducing the number of molecules that need to be experimentally synthesised and tested. This approach can save time and resources in the search for new TADF molecules and can lead to the discovery of more efficient OLEDs. In type-VII TADF systems where it is not easy to distinguish between the donor and acceptor moieties, the HOMO-LUMO overlap integral can still be used as a metric to assess the potential for exciton transfer between the two parts of the molecule [13], which is a key factor in TADF efficiency.



(a) (b)

Figure 5.1: Boxplots showing the distribution of (a) HOMO-LUMO energy gap and (b) HOMO-LUMO overlap (Λ_T) values across the filtered molecule library. The black diamond represents the value of the corresponding parent molecule.

In this section, we compare and visualise the distribution of the HOMO-LUMO energy gap (eV) and the HOMO-LUMO overlap parameter (Λ_T) among the different molecule families in Figures 5.1a and 5.1b, respectively. The HOMO-LUMO energy gap and Λ_T for the parent molecules

were calculated based on their MMFF94-optimised geometry. It is a fair comparison to use these values for both parent and child molecules, as they are optimised using the same method of MMFF94 geometry optimisation. However, using DFT calculations to optimise the geometry of the parent molecules would be unfair, as it significantly increases the computational cost and time required for the screening process. Moreover, using MMFF94-optimised geometry for both parent and child molecules ensures a consistent and unbiased comparison between them. Therefore, the use of MMFF94-optimised geometry for both parent and child molecules enables a more efficient and unbiased screening process in the HTVS approach.

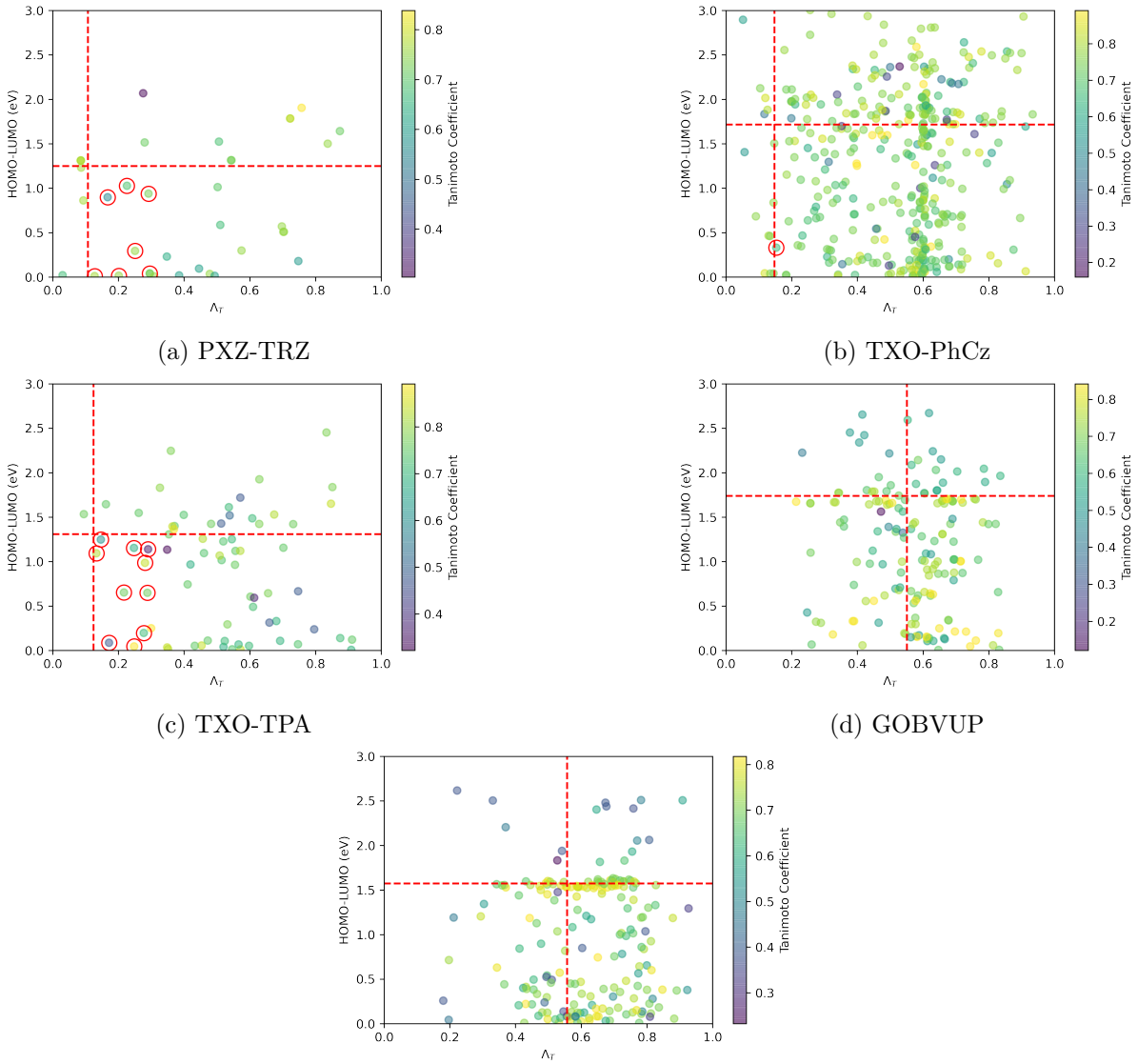
Our analysis revealed an interesting pattern in the HOMO-LUMO values of parent and child molecules (Figure 5.1a). Specifically, the HOMO-LUMO value of D-A TADF parent molecules falls within quartile group 3 (Q3), whereas type-VII parent molecules are positioned just above the edge of Q3, in quartile group 4 (Q4). This suggests that the HOMO-LUMO value of the parent molecules is relatively consistent and falls within a certain range, while the child molecules have a wider range of values. This finding may be useful in the selection of potential TADF molecules, as it suggests that child molecules with HOMO-LUMO values falling within the Q3 range of the parent molecules may have higher chances of exhibiting TADF behaviour. However, other factors such as Λ_T and overall chemical structure should also be taken into consideration during the selection process.

Analysing Λ_T values of the two TADF families, D-A and Type-II, in Figure 5.1b, reveals a consistent pattern in the values of Λ_T. The D-A parent molecules generally have smaller Λ_T values, with all falling within quartile group 1 (Q1) of the distribution, while the Type-VII parent molecules have higher Λ_T values, with most falling within quartile group 2 (Q2). However, the child molecules exhibit a wider range of Λ_T values. These observations can inform the selection of potential TADF molecules, with a focus on child molecules with Λ_T values falling within the Q1 and Q2 range of their respective parent molecules, along with consideration of other factors such as HOMO-LUMO values and overall chemical structure.

As discussed in Section 1.4.2, D-A type molecules are expected to have a small HOMO-LUMO overlap due to their asymmetric electronic structure, which leads to a small singlet-triplet en-

ergy gap and promotes efficient ISC. This is in contrast to Type-VII molecules, which have a symmetric electronic structure and are expected to have a greater HOMO-LUMO overlap. As seen in the literature, the distribution of the HOMO-LUMO overlap parameter (Λ_T) for the D-A parent molecules in our dataset is mostly in Q1, which is consistent with the theoretical prediction. By contrast, Type-VII parent molecules generally exhibit higher values of Λ_T than those of D-A TADF and predominantly occupy Q2, which is consistent with the findings reported in the original paper [13]. The child molecules, on the other hand, exhibit a wider range of values for Λ_T, which suggests that their electronic structures have been significantly modified to achieve TADF properties.

In addition to the box plots, we also plotted the HOMO-LUMO gap as a function of Λ_T (figure 5.2) for each molecule. Using this plot, we performed molecule filtering based on the criteria of having a HOMO-LUMO gap and Λ_T values smaller than their respective parent molecules. This allowed us to focus on molecules with a similar or smaller HOMO-LUMO gap and Λ_T value, which are important parameters for efficient TADF performance. Thus, this plot provided a useful tool for narrowing down the selection of potential TADF candidates for further testing. However, we only had 3 molecules for the PXZ-TRZ family, 9 for the TXO-TPA family, and none for the TXO-TPB family. To ensure a diverse set of compounds for further analysis, we selected a minimum of 10 molecules for these families. The additional molecules were chosen based on their increasing order of Λ_T, while still being smaller than the parent HOMO-LUMO value (Table 5.1).



(a) PXZ-TRZ

(b) TXO-PhCz

(c) TXO-TPA

(d) GOBVUP

(e) TUFWAS

Figure 5.2: HOMO-LUMO energy gap (eV) and HOMO-LUMO overlap Λ_T of molecules, with color-coded Tanimoto coefficients. The red vertical dashed line represents the value of the parent molecule's Λ_T and the red horizontal dashed line represents the parents HOMO-LUMO (eV).

The process of filtering molecules based on the HOMO-LUMO gap and the overlap integral parameter Λ_T resulted in a decreased total number of molecules, as shown in Table 5.1. Furthermore, performing a single-point DFT calculation on the MMFF94 optimised geometry further decreased the number of molecules due to convergence issues that can arise from discrepancies between the MMFF94 force field and DFT methods. These discrepancies can lead to differences in the energy and geometry calculations, potentially causing convergence issues during the DFT calculation.

HTVS typically involves the rapid screening of large chemical libraries to identify potential candidates. Due to the large number of molecules involved, it is often necessary to use computationally efficient methods that can quickly filter out unlikely candidates. However, these methods may also have limitations and discrepancies compared to more accurate but computationally expensive methods like DFT. When using computationally efficient methods like MMFF94-optimised structures for DFT calculations to calculate the HOMO-LUMO gap and overlap integral parameters in HTVS efforts, it is important to acknowledge the potential discrepancies and limitations of these methods. Such discrepancies can lead to differences in the energy and geometry calculations, potentially causing convergence issues during the DFT calculation.

Table 5.1: Number of molecules remaining after applying the HOMO-LUMO and density overlap threshold during the high-throughput virtual screening workflow.

Method	PXZ-TRZ	TXO-PhCz	TXO-TPA	GOBVUP	TUFWAS
MMFF94 Optimisation	37	397	72	170	204
DFT on MMFF94	37	383	72	165	198
HL & LT	10	10	10	30	32

5.3 DFT Geometry optimisation

In this section, we present the data obtained from the geometry optimisation of the selected molecules using the DFT method. As discussed in the previous section, we filtered the molecules based on their HOMO-LUMO gap and Λ_T values obtained from the MMFF94 calculations. All selected molecules were then subjected to DFT geometry optimisation to obtain their optimised geometries. We obtained the optimised geometries for all selected molecules (Table 5.2).

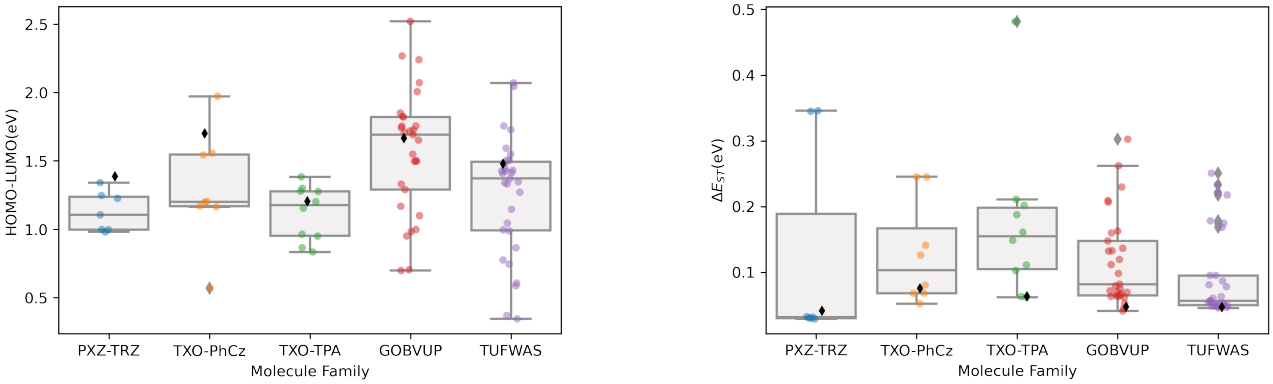
Table 5.2: Number of molecules remaining after DFT geometry optimisation in the high-throughput virtual screening workflow discussed in this work.

Method	PXZ-TRZ	TXO-PhCz	TXO-TPA	GOBVUP	TUFWAS
DFT Geopt.	10	10	10	30	32

5.4 Frontier Orbitals and Singlet-Triplet Splittings

The results section presents the analysis of Frontier Orbitals (HOMO-LUMO) and singlet-triplet splitting (ΔE_{ST}) of the DFT geometry-optimised molecules, obtained from section 5.3. The HOMO-LUMO energy gap is a crucial factor that determines the TADF efficiency of a molecule. A smaller HOMO-LUMO energy gap indicates a higher probability of efficient ISC, which is desirable for TADF applications. Moreover, the singlet-triplet energy splitting (ΔE_{ST}) plays a crucial role in determining the TADF properties of a molecule. A smaller ΔE_{ST} indicates a higher probability of achieving efficient reverse intersystem crossing (RISC) and thus higher TADF efficiency [50]. Therefore, understanding the values of HOMO-LUMO and ΔE_{ST} is essential to identify potential TADF molecules for practical applications. The analysis presented in this section provides valuable insights into the TADF properties of the selected molecules.

DFT geometry optimisation of the parent molecules was an essential step, as it provided a necessary baseline for comparison with the child molecules. The boxplot of HOMO-LUMO



(a) (b)

Figure 5.3: Box plots showing the distribution of (a) HOMO-LUMO energy gap and (b) singlet-triplet splitting (ΔE_{ST}) energy for the parent molecules and their derivatives. The black diamond represents the value of the corresponding parent molecule.

values across each molecule family after DFT geometry optimisation reveals some interesting trends. First, it is important to note that all candidate molecules, both parents and children, were optimised with DFT. This is a positive sign and indicates that the molecules are reasonable candidates for further investigation.

The Figure 5.3a of HOMO-LUMO gaps after DFT geometry optimisation revealed interesting trends among different molecule families. It was observed that the HOMO-LUMO gaps for all the candidate molecules increased overall, indicating improved stability and electronic properties after optimisation. Moreover, it was observed that different molecule families had varying ranges of HOMO-LUMO gaps. The type-VII TADF families, including GOBVUP and TUFWAS, exhibited a larger range of HOMO-LUMO gaps, suggesting that these families may be promising sources for high-performance TADF materials. In contrast, the D-A TADF families, including PXZ-TRZ, TXO-PhCz, and TXO-TPA, had smaller ranges of HOMO-LUMO gaps, indicating that the use of this method to generate molecules may prove less effective for D-A type. However, the type-VII families had more candidate molecules than D-A families, which could have contributed to the result in a wider range of HOMO-LUMO values but also type-VII molecules generally do have a larger HOMO-LUMO gap [13]. Overall, the type-VII TADF families, GOBVUP and TUFWAS, had a wider range of HOMO-LUMO gaps compared to the D-A TADF families, PXZ-TRZ, TXO-PhCz, and TXO-TPA, which had smaller ranges. The type-VII families may be a promising source for high-performance TADF materials.

Figure 5.3b shows the ΔE_{ST} of the child molecules following DFT geometry optimisation. Overall, the majority of child molecules have higher ΔE_{ST} values than their parent molecules. However, some child molecules in the D-A family, namely PXZ-TRZ, TXO-PhCz, and TXO-TPA, exhibit smaller or comparable ΔE_{ST} values, despite being selected based on higher HOMO-LUMO overlap (Λ_T). Λ_T calculated on MMFF94-optimised geometry could still be a useful indicator depending on the research question and context. For instance, when the goal is to rapidly compare different molecules within a specific class of compounds, using the same level of theory for all molecules could provide a consistent basis for comparison. However, as mentioned in the previous section, filtering molecules solely based on the HOMO-LUMO energy gap could be feasible in terms of computational cost, as evidenced by previous studies that have performed DFT geometry optimisation on hundreds of molecules [161, 13, 160]

Overall, the range of ΔE_{ST} values varied among different molecule families. Among the D-A TADF families, TXO-TPA had the smallest range of ΔE_{ST} values, indicating that this method may be less effective in generating D-A type TADF materials. On the other hand, the type-

VII TADF families showed a wider range of ΔE_{ST} values, with some molecules lying beyond the upper whisker of the boxplot. This suggests that these families may contain potential candidates for high-performance TADF materials.

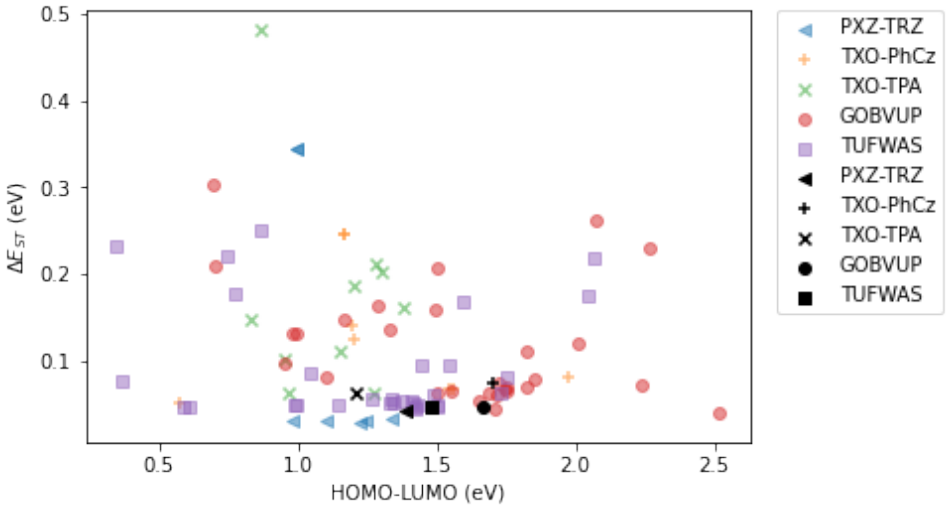


Figure 5.4: The singlet-triplet splitting (ΔE_{ST}) and LUMO-HOMO gap are shown using black markers for the parent molecule and coloured markers for the candidate molecules.

Figure 5.4 does not show any clear trend between the two variables, ΔE_{ST} and HOMO-LUMO. However, there is a cluster of molecules around the ΔE_{ST} values of the parent molecules, indicating that the child molecules maintained some similarity to the parent molecules in terms of their singlet-triplet splitting. This suggests that the design strategy employed to generate the child molecules did not result in a significant change in the singlet-triplet splitting of the molecules.

The study investigated the HOMO-LUMO gap and ΔE_{ST} of candidate molecules for TADF materials. DFT geometry optimisation was performed for both parent and child molecules. The boxplot of HOMO-LUMO gaps after optimisation showed that the type-VII TADF families had a wider range of gaps compared to the D-A TADF families, indicating their potential for high-performance TADF materials. The ΔE_{ST} of the child molecules also showed interesting trends among different molecule families, with some child molecules exhibiting smaller or comparable values despite being selected based on higher HOMO-LUMO overlap. Overall, the study suggests that filtering molecules based on the HOMO-LUMO energy gap could be feasible in terms of computational cost and could be a useful indicator depending on the research

question and context.

In the subsequent stage of our workflow, we focus on exploring candidate molecules that exhibit the smallest $\Delta E_{ST} < 0.1$ eV. This is crucial because a smaller ΔE_{ST} implies an efficient reverse intersystem crossing (RISC), which in turn can facilitate efficient TADF. Hence, molecules with smaller ΔE_{ST} values are more promising in exhibiting TADF behaviour. We use the visualisation of a similarity map to analyse the structure which will be discussed next.

5.4.1 Similarity Map

To visualise the similarities and differences in structure between parent and child molecules, we used the `GetMorganFingerprint` function in RDkit with a radius of 4 and a fingerprint type of 'bv' (bit vector) [219, 194, 220].

The radius value chosen for fingerprint generation can impact the size of the molecular environment considered, but its importance depends on the application and molecule complexity. Larger radii capture more information about the molecule while smaller radii focus on the immediate environment. A suitable radius must capture relevant features and avoid noise.

Two commonly used molecular fingerprint types are count and bit vector. Count generates a vector that counts the occurrence of substructures in the molecule, while bit vector generates a fixed-length vector that represents the presence or absence of each substructure. 'bv' vectors are preferred in machine learning models and similarity searches due to their consistent length, which makes them suitable for fixed-length feature sets.

The fixed-length nature of bv fingerprints is advantageous in reducing noise caused by substructure variations and ensuring the feature sets used for training and prediction are consistent. These fingerprints can be saved in a file format and used as input for various machine learning models and similarity search algorithms. For machine learning, bv fingerprints can be used as input features for predictive models, while in similarity searches, they can be used to search large databases of compounds to identify molecules with similar substructures.

5.4.2 Candidates

In order to understand the structure-activity relationships of TADF molecules, it is important to investigate the similarities and differences between molecules with similar electronic properties. One approach to accomplish this is to generate similarity maps using molecular fingerprints. Here, we focus on three candidate molecules from each TADF family with the smallest ΔE_{ST} and analyse their structures using a similarity map approach, discussed in Section 5.4.1. This will allow us to visually compare the structural similarities and differences between parent and child molecules and potentially identify key substructures that contribute to their TADF properties.

Table 5.3: Singlet-triplet energy gap (ΔE_{ST}) and HOMO-LUMO energy gap (eV) and SYnthetic Bayesian Accessibility (SYBA) score of the parent molecules.

Molecule	HOMO-LUMO (eV)	ΔE_{ST} eV	SYBA
PXZ-TRZ	1.39	0.041	36.2
TXO-PhCz	1.70	0.076	27.8
TXO-TPA	1.21	0.063	49.2
GOBVUP	1.67	0.048	-2.2
TUFWAS	1.48	0.047	18.5

The naming convention of the molecules in Table 5.4 is based on their order of generation. Specifically, the number in the name (e.g. 2086) refers to the index of the molecule in the sequence of generated molecules. For instance, TXO-PhCz 2086 is simply the 2086th molecule generated using this method.

When comparing the HOMO-LUMO energy and ΔE_{ST} values (Table 5.4) of the candidate molecules with parent molecules, among the PXZ-TRZ family, PXZ-TRZ 866, PXZ-TRZ 1461, and PXZ-TRZ 2078 exhibited a decrease in HOMO-LUMO energy compared to their parent molecule by 0.16, 0.14, and 0.40 eV, respectively.

On the other hand, the TXO-PhCz family exhibited an increase in HOMO-LUMO energy compared to their parent molecule, with TXO-PhCz 2424, TXO-PhCz 2086, and TXO-PhCz 1126 showing an increase of 1.13, 0.16 and 0.145 eV, respectively.

GOBVUP 2026 and GOBVUP 1527 molecules showed an increase in HOMO-LUMO energy

by 0.85 and 0.04 eV, respectively. While GOBVUP 2149 showed a decrease in HOMO-LUMO energy by 0.02 eV. The TUFWAS 2720 and TUFWAS 340 molecule showed a decrease in HOMO-LUMO energy 0.06 and 0.89 eV, respectively, while TUFWAS 22 increased by 0.02 eV. Regarding ΔE_{ST} values, all molecules exhibited a decrease in ΔE_{ST} except for TUFWAS

Table 5.4: Singlet-triplet energy gap (ΔE_{ST}) and HOMO-LUMO energy gap (eV), SYnthetic Bayesian Accessibility (SYBA) score and Tanimoto coefficient (compared to parent molecule) of molecules chosen based on smallest ΔE_{ST}.

Molecule	HOMO-LUMO (eV)	ΔE_{ST}eV	SYBA	Tanimoto Coefficient
PXZ-TRZ 866	1.23	0.029	11.5	0.77
PXZ-TRZ 1461	1.25	0.030	-10.7	0.61
PXZ-TRZ 2078	0.98	0.031	8.2	0.77
TXO-PhCz 2424	0.57	0.052	-2.8	0.65
TXO-PhCz 2086	1.54	0.068	-7.5	0.75
TXO-PhCz 1126	1.56	0.068	12.7	0.83
TXO-TPA 1575	1.28	0.062	45.3	0.83
TXO-TPA 380	0.96	0.063	37.6	0.84
GOBVUP 2026	2.52	0.041	-3.2	0.44
GOBVUP 1527	1.71	0.045	8.4	0.64
GOBVUP 2149	1.65	0.053	7.6	0.75
TUFWAS 2720	1.43	0.046	-17.5	0.62
TUFWAS 22	1.50	0.047	9.4	0.80
TUFWAS 340	0.59	0.048	-16.1	0.66

340, GOBVUP 2149 and TXO-TPA 380, which showed a small increase in ΔE_{ST}. The largest decrease in ΔE_{ST} was observed for TXO-PhCz 2424, which showed a decrease of 0.023 eV compared to its parent molecule.

Based on the analysis, it can be concluded that there are notable differences in the electronic properties of the molecules in the D-A type family (PXZ-TRZ, TXO-PhCz, TXO-TPA) and type-VII family (GOBVUP and TUFWAS) compared to their respective parent molecules. The PXZ-TRZ family showed a significant decrease in HOMO-LUMO energy, while the TXO-PhCz family exhibited an increase. The GOBVUP and TUFWAS molecules showed a mix of increase and decrease in HOMO-LUMO energy. Regarding ΔE_{ST} values, all molecules showed a small decrease except for TUFWAS 340, GOBVUP 2149 and TXO-TPA 380, which showed a small increase.

Using similarity maps (Figure 5.5 and 5.6) can be an effective approach for understanding

the observed differences in electronic properties. The similarity maps provide a visual representation of the similarities and differences in electronic distribution between the candidate molecules and their respective parent molecules. By comparing the similarity maps of the candidate molecules with their parent molecules, we can identify the regions where there are differences in electronic distribution that lead to the observed differences in electronic properties such as HOMO-LUMO energy and ΔE_{ST}.

This approach can provide valuable insights into the molecular structure and electronic properties of the candidate molecules, which can guide the design of new molecules with desirable electronic properties. Therefore, the D-A type family (PXZ-TRZ, TXO-PhCz, TXO-TPA) and type-VII family (GOBVUP and TUFWAS) can be fine-tuned by modifying their structures, making them potential candidates for organic optoelectronic devices.

PXZ-TRZ 866

As depicted in Figure 5.5a, the PXZ-TRZ 866 variant exhibits a smaller HOMO-LUMO energy gap of 1.23 eV, a smaller ΔE_{ST} of 0.029, and a smaller SYBA value of 11.5 compared to the parent molecule (as summarised in Table 5.4). The Tanimoto coefficient of 0.77 suggests a higher level of similarity with the parent molecule.

The smaller HOMO-LUMO energy gap in PXZ-TRZ 866 compared to the parent molecule indicates that it may have improved charge transport properties. However, the lower SYBA value suggests that PXZ-TRZ 866 may be less synthetically accessible than the parent molecule.

The Tanimoto coefficient indicates that there are some structural differences between PXZ-TRZ 866 and the parent molecule. In PXZ-TRZ 866, the nitrogen-containing functional group has a positive charge instead of a neutral charge, and the position of the oxygen-containing functional group is different.

In summary, PXZ-TRZ 866 has a different structure compared to the parent molecule PXZ-TRZ, with a positively charged nitrogen-containing functional group and a different position of the oxygen-containing functional group. This results in differences in electronic and structural

properties, including a smaller HOMO-LUMO energy gap and lower SYBA value. This is interesting as prior research has described radical emitters that produce emission originating from a spin doublet, rather than a singlet or triplet exciton, as previously discussed in the literature [210, 217, 218]. Further investigation is necessary to fully understand the implications of these differences on the properties and potential applications of the molecule.

PXZ-TRZ 1461

The candidate molecule PXZ-TRZ 1461 (Figure 5.5b) has a similar structure to the parent molecule but with a bromine atom substituted at TRZ. The HOMO-LUMO value for PXZ-TRZ 1461 is 1.25 eV, which is smaller than the parent molecule (1.39 eV). The ΔE_{ST} for PXZ-TRZ 1461 is 0.030, which is also smaller than the parent molecule. The SYBA value for PXZ-TRZ 1461 is negative (-10.66). The Tanimoto coefficient for PXZ-TRZ 1461 is 0.61 suggests a moderate level of similarity with the parent molecule

Based on these results, it can be concluded that the substitution of the hydrogen atom with a bromine atom at TRZ of the PXZ-TRZ molecule leads to a decrease in the HOMO-LUMO value and the ΔE_{ST}. The negative SYBA value for PXZ-TRZ 1461 suggests it is difficult to synthesise. Therefore, if the goal is to design a molecule with similar properties to PXZ-TRZ, substituting a hydrogen atom with a bromine atom at TRZ may not be an optimal strategy.

PXZ-TRZ 2078

PXZ-TRZ 2078 has a positively charged phosphonium group (P+) attached to one of the aryl groups (Figure 5.5c), while PXZ-TRZ does not have any charged groups. In terms of their electronic properties, PXZ-TRZ 2078 has a smaller HOMO-LUMO energy gap compared to PXZ-TRZ. This indicates that PXZ-TRZ 2078 is more easily ionised compared to PXZ-TRZ. Additionally, PXZ-TRZ 2078 has a smaller ΔE_{ST} than PXZ-TRZ, indicating that it is more easily oxidised. However, despite the difference in their electronic properties, both molecules have positive SYBA values.

In terms of their Tanimoto coefficient, PXZ-TRZ 2078 has a score of 0.77, which is lower than the Tanimoto coefficient of 0.95875264 for PXZ-TRZ 866. This indicates that PXZ-TRZ 2078 is less similar to the parent molecule PXZ-TRZ compared to PXZ-TRZ 866.

Overall, the addition of the positively charged phosphonium group to one of the aryl groups in PXZ-TRZ 2078 introduces a significant structural difference, leading to different electronic properties and lower similarity to the parent molecule PXZ-TRZ.

Design Strategy for Improving PXZ-TRZ

Based on the data provided, there is a design strategy that could potentially improve the TADF properties of PXZ-TRZ:

- Modifying the aryl groups: The aryl groups attached to the pyrazine ring in PXZ-TRZ could be modified to improve its TADF properties. For example, introducing electron-donating or electron-withdrawing groups to the aryl groups could change the energy levels of the molecule and possibly improve its TADF properties.

TXO-PhCz 2424

TXO-PhCz 2424 has a phosphine group (PH) (Figure 5.5d) attached to the carbazole unit, while the parent TXO-PhCz does not have any such group.

In terms of their electronic properties, TXO-PhCz 2424 has a smaller HOMO-LUMO energy gap compared to TXO-PhCz. This indicates that TXO-PhCz 2424 is more easily ionised compared to TXO-PhCz. Additionally, TXO-PhCz 2424 has a smaller ΔE_{ST} than TXO-PhCz. However, despite the difference in their electronic properties, both molecules have positive SYBA values.

Overall, the addition of the PH group to the carbazole unit in TXO-PhCz 2424 introduces a significant structural difference, leading to different electronic properties and lower similarity to the parent molecule TXO-PhCz. The smaller HOMO-LUMO energy gap and smaller ΔE_{ST}

of TXO-PhCz 2424 suggest that this molecule may have better TADF properties compared to TXO-PhCz. However, further experimental studies would be required to confirm this.

TXO-PhCz 2086

In TXO-PhCz 2086 (Figure 5.5e), one of the benzene rings attached to the thiophene ring has been quaternised (i.e. a positively charged nitrogen atom with four organic substituents)), resulting in a positively charged nitrogen atom in the ring. This modification may have an effect on the molecule's reactivity and electronic properties.

The HOMO-LUMO gap of TXO-PhCz 2086 is smaller than that of the parent molecule. The ΔE_{ST} value for TXO-PhCz 2086 is also smaller than that of the parent molecule. The SYBA value for TXO-PhCz 2086 is negative, indicating that it may be more difficult to synthesise than the parent molecule. The Tanimoto similarity value of 0.75 suggests a higher level of similarity with the parent molecule, which could simplify the synthetic process, however, the SYBA value for TXO-PhCz 2086 is negative, indicating that it may be more difficult to synthesise.

We can see that TXO-PhCz 2086 has an additional nitrogen atom that is positively charged. This modification is expected to affect the electronic properties of the molecule, potentially leading to changes in the HOMO-LUMO energy gap and other related properties. The HOMO-LUMO energy gap of TXO-PhCz 2086 is slightly smaller than that of the parent molecule, indicating the lower energy required for electron excitation.

TXO-PhCz 1126

Comparing TXO-PhCz 1126 to the parent molecule TXO-PhCz (Figure 5.5f), we see that both HOMO-LUMO gap, ΔE_{ST} and SYBA of TXO-PhCz 1126 is lower than that of the parent molecule. The Tanimoto value of 0.83 suggests that the two molecules still have a relatively high degree of similarity, indicating that TXO-PhCz 1126 may still have similar properties and potential applications as the parent molecule.

Design Strategy for Improving TXO-PhCz

A design strategy that could potentially improve the TADF properties of TXO-PhCz:

- Modification of the acceptor moiety: In both TXO-PhCz 2086 and TXO-PhCz 1126, the acceptor moiety has been modified by introducing electron-withdrawing groups such as nitrogen, which can improve the TADF properties of the molecule. Therefore, further modification of the acceptor moiety could potentially enhance the TADF properties of TXO-PhCz.

TXO-TPA 1575

The parent molecule TXO-TPA is a donor-acceptor type organic molecule with a thiophene core and two electron-withdrawing groups (anhydride and sulfone).

TXO-TPA 1575 is a derivative of TXO-TPA with a modified acceptor group (Figure 5.5g). It has a similar molecular structure with a HOMO-LUMO energy gap of 1.28 eV, a DEST of 0.062 eV, and an SYBA score of 45.3. Compared to the parent molecule, the modified molecule has a slightly larger HOMO-LUMO energy gap. The ΔE_{ST} of the modified molecule is slightly lower than the parent molecule. The SYBA score (45.3) of the modified molecule is similar to the parent molecule, indicating that it is easy to synthesise.

In terms of similarity, TXO-TPA 1575 has a Tanimoto coefficient of 0.83, indicating that it shares a high degree of structural similarity with the parent molecule.

Overall, the modifications made to the acceptor group in TXO-TPA 1575 seem to have improved the TADF properties of the parent molecule. However, further experimental investigation is necessary to confirm the theoretical predictions.

TXO-TPA 380

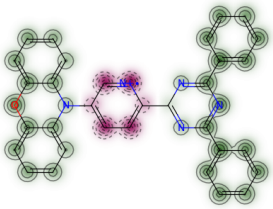
The HOMO-LUMO gap of TXO-TPA 380 (Figure 5.5h) is smaller than that in the parent molecule, indicating that TXO-TPA 380 has a lower excitation energy requirement. The singlet-triplet splitting (ΔE_{ST}) in TXO-TPA 380 is similar to the parent molecule. However, the SYBA of TXO-TPA 380 is lower than that of the parent molecule yet positive value of 37.6 suggests that it may be easy to synthesise.

Despite the differences, TXO-TPA 380 still retains some similarities to the parent molecule, as it has a Tanimoto coefficient of 0.84.

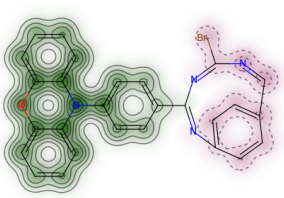
Design Strategy for Improving TXO-TPA

Based on the comparison of TXO-TPA 1575 and TXO-TPA 380, a potential design strategy that could improve TADF properties of TXO-TPA is:

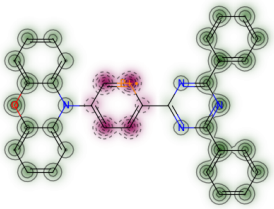
- Introduction of electron-donating groups: The HOMO-LUMO energy gap of TXO-TPA 1575 is higher than that of TXO-TPA 380. The introduction of electron-donating groups such as alkyl or aryl substituents on the aromatic rings may enhance the electron transport properties and reduce the HOMO-LUMO energy gap.



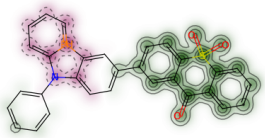
(a) PXZ-TRZ 866
ΔE_{ST} = 0.029 eV
HOMO-LUMO = 1.23 eV
Tanimoto coefficient = 0.77
SYBA = 11.47



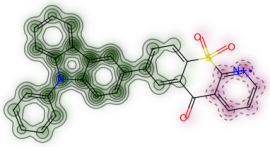
(b) PXZ-TRZ 1461
ΔE_{ST} = 0.030 eV
HOMO-LUMO = 1.25 eV
Tanimoto coefficient = 0.61
SYBA = -10.66



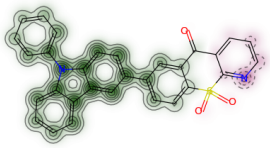
(c) PXZ-TRZ 2078
ΔE_{ST} = 0.031 eV
HOMO-LUMO = 0.98 eV
Tanimoto coefficient = 0.77
SYBA = 8.24



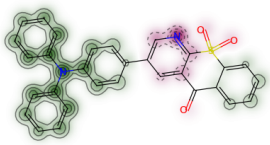
(d) TXO-PhCz 2424
ΔE_{ST} = 0.052 eV
HOMO-LUMO = 0.57 eV
Tanimoto coefficient = 0.65
SYBA = -2.81



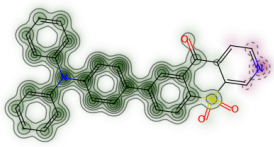
(e) TXO-PhCz 2086
ΔE_{ST} = 0.068 eV
HOMO-LUMO = 1.54 eV
Tanimoto coefficient = 0.75
SYBA = -7.52



(f) TXO-PhCz 1126
ΔE_{ST} = 0.068 eV
HOMO-LUMO = 1.56 eV
Tanimoto coefficient = 0.83
SYBA = 12.69

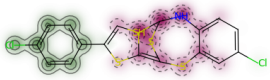


(g) TXO-TPA 1575
ΔE_{ST} = 0.062 eV
HOMO-LUMO = 1.28 eV
Tanimoto coefficient = 0.83
SYBA = 45.33

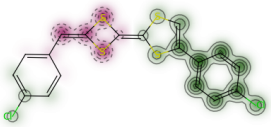


(h) TXO-TPA 380
ΔE_{ST} = 0.063 eV
HOMO-LUMO = 0.96 eV
Tanimoto coefficient = 0.84
SYBA = 37.60

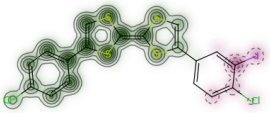
Figure 5.5: Visualisation of similarity maps generated by comparing the selected TADF molecules with the smallest ΔE_{ST} (eV) values from each D-A TADF family to their parent molecules. The green colour indicates a higher similarity between the parent and child molecules in that particular region, while the red colour indicates a lower similarity.



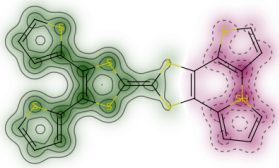
(a) GOBVUP 2026
ΔE_{ST} = 0.041 eV
HOMO-LUMO = 2.52 eV
Tanimoto coefficient = 0.44
SYBA = -3.23



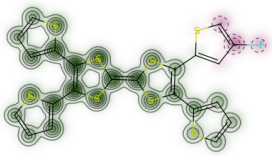
(b) GOBVUP 1527
ΔE_{ST} = 0.045 eV
HOMO-LUMO = 1.71 eV
Tanimoto coefficient = 0.64
SYBA = 8.41



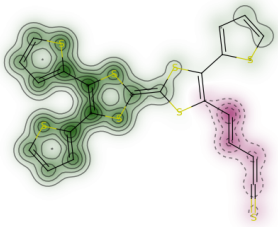
(c) GOBVUP 2149
ΔE_{ST} = 0.053 eV
HOMO-LUMO = 1.65 eV
Tanimoto coefficient = 0.75
SYBA = 7.58



(d) TUFWAS 2720
ΔE_{ST} = 0.046 eV
HOMO-LUMO = 1.43 eV
Tanimoto coefficient = 0.62
SYBA = 37.60



(e) TUFWAS 22
ΔE_{ST} = 0.047 eV
HOMO-LUMO = 1.50 eV
Tanimoto coefficient = 0.80
SYBA = 37.60



(f) TUFWAS 340
ΔE_{ST} = 0.048 eV
HOMO-LUMO = 0.59 eV
Tanimoto coefficient = 0.66
SYBA = 37.60

Figure 5.6: Visualisation of the similarity maps generated using the chosen TADF molecules with the smallest ΔE_{ST} (eV) values from each type-VII TADF family. The green colour indicates a higher similarity between the parent and child molecules in that particular region, while the red colour indicates a lower similarity.

GOBVUP 2026

The ΔE_{ST} of GOBVUP 2026 is 0.041 eV, which is smaller than that of the parent molecule. The SYBA value of GOBVUP 2026 is -3.22, indicating that the molecule did not improve in terms of synthesis difficulty compared to the parent molecule. However, the similarity value of GOBVUP 2026 is 0.44, indicating a smaller degree of similarity to the parent molecule.

Overall, GOBVUP 2026 appears to have reduced ΔE_{ST} of the molecule while significantly reducing structural similarity to the parent molecule. This suggests that introducing similar structural modifications to the parent molecule, such as changing the core structure, maybe a

potential design strategy to improve TADF properties.

GOBVUP 1527

In GOBVUP 1527 (Figure 5.6b), one of the 1,3-Dithiole units has been replaced with a 1,3-Dithietane unit. The HOMO-LUMO energy gap (1.71 eV) is larger compared to the parent molecule. The ΔE_{ST} is also slightly smaller at 0.045 eV. However, the SYBA value is positive at 8.42, indicating that the synthesis ability has been improved over the parent molecule.

GOBVUP 2149

GOBVUP 2149 (Figure 5.6c) is a derivative of the parent molecule GOBVUP with the addition of an iodine atom in one of the phenyl rings. In terms of electronic properties, GOBVUP 2149 has a slightly smaller HOMO-LUMO gap than its parent molecule. However, the ΔE_{ST} is higher than the parent molecule.

In terms of accessibility, GOBVUP 2149 has improved synthesisability over the parent molecule, as it is a positive vale of 7.6. When it comes to molecular structure, GOBVUP 2149 is very similar to its parent molecule, with the only major difference being the presence of an iodine atom in one of the phenyl rings.

Overall, GOBVUP 2149 appears to be a promising derivative of its parent molecule with slightly altered electronic and improved synthesisability.

Design Strategy for Improving GOBVUP

Based on the data of GOBVUP 2026, GOBVUP 1527, and GOBVUP 2149, there is one potential design strategy that could improve the TADF properties of GOBVUP:

- The TADF properties of GOBVUP 2149, which contain an iodine atom, are better than those of GOBVUP 1527, which does not contain iodine. This suggests that incorporating heavy atoms, particularly halogen atoms, into the molecular structure could improve

TADF properties by increasing the spin-orbit coupling and decreasing the energy difference between the singlet and triplet states [221].

TUFWAS 2720

In TUFWAS 2720 (Figure 5.6d), the thiophene rings are substituted with a mixture of sulfur-containing and carbon-based groups.

TUFWAS 2720 has a smaller HOMO-LUMO energy gap and ΔE_{ST} value than that of the parent molecule TUFWAS, indicating that the TADF properties of TUFWAS 2720 may be more favourable than those of TUFWAS. However, TUFWAS 2720 is indicated to be hard-to-synthesis as the SYBA is -17.5. It also has Tanimoto coefficient of 0.62 suggesting there is a moderate degree of similarity with the parent.

TUFWAS 22

TUFWAS 22 (Figure 5.6e) has a fluorine atom substituted on one of the thiophenes, which causes an increase in the HOMO-LUMO energy gap from 1.48 to 1.50 eV (parent structure to TUFWAS 22). This suggests that the fluorine substitution weakly affects the electron-donating ability of the molecule, likely due to the electron-withdrawing nature of the fluorine atom. However, we observed that the ΔE_{ST} of TUFWAS 22 (0.047 eV) is similar to that of the parent molecule TUFWAS (0.047 eV).

The SYBA value of TUFWAS 22 is 9.3 is significantly smaller than that of the parent molecule TUFWAS (18.5), indicating that the fluorine substitution increases the synthesise difficulty of the molecule. The Tanimoto coefficient of TUFWAS 22 is 0.80, suggesting that it has a high degree of similarity to the parent molecule. Overall, the substitution of a fluorine atom in TUFWAS 22 has a minimal impact on the molecule's electronic and optical properties but increases its synthesise difficulty.

TUFWAS 340

TUFWAS 340 (Figure 5.6e) has a linear backbone with a vinyl group, while TUFWAS has a more complex structure with multiple rings and sulfur atoms.

In terms of electronic properties, TUFWAS 340 has a much smaller HOMO-LUMO gap than the parent molecule. However, the ΔE_{ST} of TUFWAS 340 is similar. The SYBA value of TUFWAS 340 is negative which indicates that it may be difficult to synthesise. In terms of similarity, TUFWAS 340 has a moderate similarity to the parent molecule TUFWAS, with a value of 0.66.

Overall, while TUFWAS 340 has a lower HOMO-LUMO gap than the parent molecule TUFWAS but similar ΔE_{ST}, indicates that it may be a good candidate for TADF. However, the negative SYBA value also suggests that it may be difficult to synthesise. On the other hand, the parent molecule TUFWAS has a positive DEST and SYBA value, making it a potentially good candidate for TADF.

Design Strategy for Improving TUFWAS

Based on the data of TUFWAS 2720, TUFWAS 22, and TUFWAS 340, there is one potential design strategy that could improve the TADF properties of TUFWAS:

- The substitution of a fluorine atom in thiophene group in TUFWAS 22 has a minimal impact on the molecule's electronic and has the potential to decrease ΔE_{ST}, however, it can increase synthesis difficulty.

In Zhao *et al.*'s paper [13], it was demonstrated that removing fluorine atoms of the bis(pentafluorophenyl)bo
group in ILUBEY significantly increased ΔE_{ST} from 0.194 eV to 0.587 eV, highlighting the important impact of fluorine atoms on TADF properties.

5.5 Summary and Conclusion

In this chapter, we utilised DFT calculations to assess the HOMO-LUMO energy gap and Λ_T of MMFF94 geometry-optimised molecules. These molecules were generated and screened according to specific criteria outlined in the previous chapter. Subsequently, we narrowed down our selection based on the parent HOMO-LUMO gap and Λ_T values. However, due to limited availability, only three molecules from the PXZ-TRZ family and none from the TXO-TPB family were used, while nine molecules were selected from the TXO-TPA family. To ensure diversity, we included a minimum of 10 molecules from each family and added more molecules based on increasing Λ_T order and parent HOMO-LUMO value. After performing DFT geometry optimisation and ΔSCF calculations, we selected the top three molecules from each family based on the lowest ΔE_{ST} values (< 0.1 eV) for structure analysis using similarity maps. We analysed the structure of each molecule on a case-by-case basis and provided our own design strategies for each.

In our workflow, it is essential to recognise the potential discrepancies and limitations when utilising computationally efficient methods, such as MMFF94-optimised structures, for DFT calculations in HTVS efforts to determine HOMO-LUMO gap and overlap integral parameters. These limitations may cause variations in energy and geometry calculations, which may lead to convergence issues during DFT calculations. Therefore, it is crucial to carefully evaluate the accuracy and reliability of such methods and consider the potential impact on the overall HTVS strategy. However, for a robust HTVS workflow filtering molecules based on the HOMO-LUMO energy gap could be feasible in terms of computational cost and could be a useful indicator depending on the research question and context.

Radical-based molecules have been studied in the literature for their potential application in OLED devices [210, 217, 218], making the radical molecules we produced through our workflow, such as PXZ-TRZ 866, PXZ-TRZ 2078, and TXO-PhCz 2086, interesting.

Our analysis of D-A molecules suggests that modifying either the donor or acceptor by increasing their electron donating or withdrawing capabilities, respectively, can improve their design.

This approach is currently being used to design D-A molecules[222].

To improve the design strategy of type-VII TADF molecules, we suggest introducing electronegative atoms such as fluorine, nitrogen, or iodine to the benzene or thiophene groups in GOBVUP or TUFWAS, respectively. This approach is similar to the tentative design reported in Zhao *et al.*'s paper [13].

Overall, our study demonstrated the utility of using DFT calculations to evaluate the HOMO-LUMO energy gap and overlap of a diverse set of molecules designed for TADF applications. While the limitations of using computationally efficient methods for geometry optimisation should be acknowledged, our findings suggest that filtering molecules based on the HOMO-LUMO energy gap can be a useful strategy in designing TADF materials. We also identified potential modification strategies for donor-acceptor molecules and type-VII TADF molecules, which may guide future efforts in developing high-performance TADF materials.

Chapter 6

Outlook and Conclusion

6.1 Conclusion

The goal of this study was to develop a HTVS workflow that could facilitate the discovery of new TADF molecules, while minimising the time and costs involved in experimental screening. To achieve this, we utilised a range of computational methods, including STONED for generating diverse molecule libraries, SYBA for scoring molecules based on their synthesis difficulty, and DFT calculations for evaluating candidate molecules. By applying these techniques to explore the chemical spaces of five TADF molecules, we were able to identify a set of promising candidates for further investigation and design iteration.

Based on SYBA analysis, STONED generated molecules that were difficult to synthesise. However, SYBA score should be viewed as a confidence indicator rather than a strict guideline. Our study demonstrated the utility and efficiency of using DFT calculations on MMFF94 optimised structures to evaluate parameters such as the HOMO-LUMO energy gap and overlap integral for a diverse set of TADF molecules. These calculations were used to filter molecules for further investigation. However, we also observed that the use of computationally efficient methods such as MMFF94-optimised structures for DFT calculations can lead to discrepancies and limitations that must be acknowledged to prevent convergence issues.

Although SMILES notation is a useful and widely used method for representing the structure of molecules, it should be noted that it has limitations for certain types of molecules. Large or complex molecules with multiple conformations may be challenging to represent with SMILES. Therefore, while SMILES notation was a valuable tool in our work, it may not be suitable for all types of molecules and structures.

Furthermore, we employed ΔSCF calculations on DFT-optimised structures and conducted similarity map analysis to investigate potential modification strategies for parent donor-acceptor molecules and type-VII TADF molecules. Our HTVS workflow generated molecules that led to adjustments in the structure of parent molecules, resembling the designs proposed in previous literature. These findings offer valuable guidance for future research aimed at designing new TADF materials.

In conclusion, our study provides valuable insights into the use of computational methods for TADF molecule design and the challenges associated with using these methods. The findings of this study have important implications for the development of future TADF molecules, as they provide a framework for reducing costs and time associated with experimental screening while also improving the efficiency of virtual screening. Future research should focus on exploring alternative force field methods and incorporating synthetic accessibility considerations into the molecule generation process to overcome the limitations of the current methods.

6.2 Future Developments

The subsequent step in the design process involves exploring the chemical space while preserving a specific substructure of the molecule, similar to the side-chain placement method commonly employed in protein design. For instance, in the case of TUFWAS and GOBVUP, the introduction of a fluorine or iodine atom to one of the thiophene groups resulted in a reduction of ΔE_{ST}. The central tetrathiafulvalene structure of the molecule is to be preserved while mutations are introduced to other portions of the structure. It is recommended to increase the number of mutations per SELFIES to expand the search space and potentially discover more

effective modifications to the molecule. Whereas, in this study, we constrained the number of mutations to one to explore the immediate chemical space from the original molecule and to facilitate further iterations in design.

To further enhance our understanding, we can conduct a thorough comparison between the MMFF94 and DFT optimised structures, specifically in terms of the HOMO-LUMO energy gap and ΔE_{ST}. This information can be used to calculate errors and fine-tune values obtained through MMFF94 geometry optimisation. Figure 6.1 shows a comparison of the HOMO-LUMO energy gap values obtained from DFT and MMFF94 optimised structures in our workflow, but additional data is required to confirm any trends. Future analyses can include a comparison of HOMO-LUMO overlap (Λ_T) and ΔE_{ST}. Moreover, it may be worthwhile to explore the use

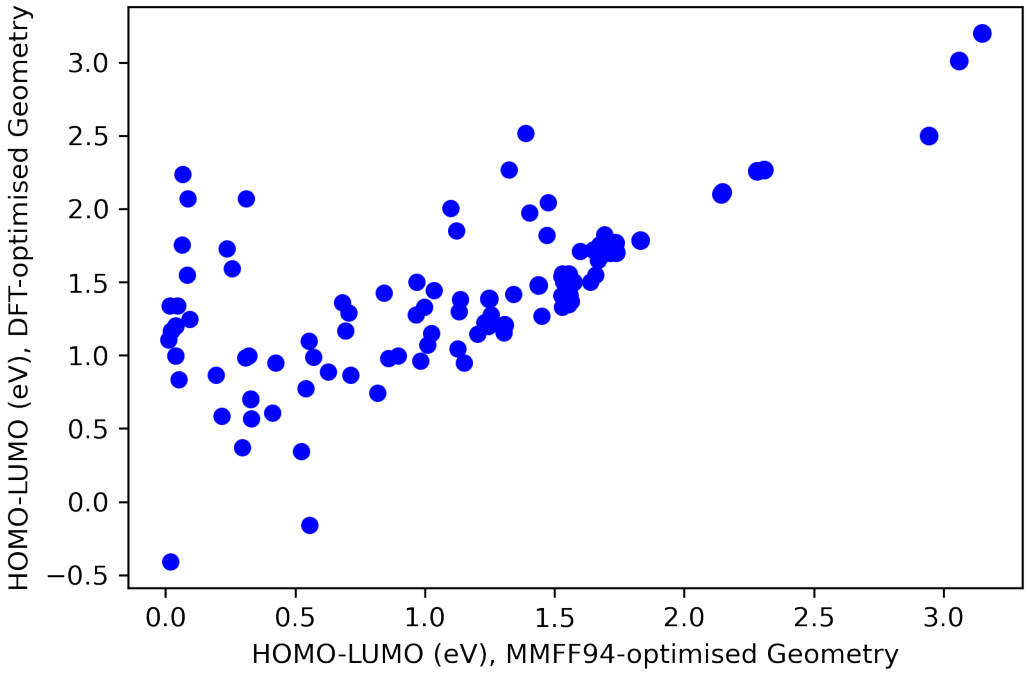


Figure 6.1: Difference in HOMO-LUMO energy gap (eV) of MMFF94 and DFT optimised molecules.

of alternative force field optimisation methods, such as OPLS3e [154] or OpenFF [155], which could potentially outperform MMFF94.

Additional research on different TADF molecules is encouraged, and we will evaluate their feasibility using this workflow.

6.3 Data Availability

All data from this thesis is available at:

`https://gitlab.com/Kritam/HTVS_STONE_SYBA_DFT`

Bibliography

[1] Hannah Ritchie, Max Roser, and Pablo Rosado. Energy. *Our World in Data*, 2022. https://ourworldindata.org/energy.

[2] Robert S. Knox. Physical aspects of the greenhouse effect and global warming. *American Journal of Physics*, 67(12):1227–1238, 1999.

[3] James Hansen, Makiko Sato, Reto Ruedy, Andrew Lacis, and Valdar Oinas. Global warming in the twenty-first century: An alternative scenario. *Proceedings of the National Academy of Sciences*, 97(18):9875–9880, 2000.

[4] IEA. Renewables 2021, Dec 2021.

[5] NOAA National Centers for Environmental information. Climate at a glance: Global time series, Oct 2022.

[6] Trevor M. Letcher. 1 - global warming—a complex situation. In Trevor M. Letcher, editor, *Climate Change (Third Edition)*, pages 3–17. Elsevier, third edition edition, 2021.

[7] Vanessa Forti, Cornelis Baldé, Ruediger Kuehr, and Garam Bel. *The Global E-waste Monitor 2020. Quantities, flows, and the circular economy potential.* 07 2020.

[8] Hisahiro Sasabe and Junji Kido. Recent progress in phosphorescent organic light-emitting devices. *European Journal of Organic Chemistry*, 2013(34):7653–7663, 2013.

[9] Boris Minaev, Gleb Baryshnikov, and Hans Agren. Principles of phosphorescent organic light emitting devices. *Phys. Chem. Chem. Phys.*, 16:1719–1758, 2014.

[10] Chihaya Adachi, Marc A. Baldo, Mark E. Thompson, and Stephen R. Forrest. Nearly 100% internal phosphorescence efficiency in an organic light-emitting device. *Journal of Applied Physics*, 90(10):5048–5051, 2001.

[11] Paul B. Tchounwou, Clement G. Yedjou, Anita K. Patlolla, and Dwayne J. Sutton. *Heavy Metal Toxicity and the Environment*, pages 133–164. Springer Basel, Basel, 2012.

[12] Michael Y. Wong and Eli Zysman-Colman. Purely organic thermally activated delayed fluorescence materials for organic light-emitting diodes. *Advanced Materials*, 29(22):1605444, 2017.

[13] Ke Zhao, Ömer H. Omar, Tahereh Nematiaram, Daniele Padula, and Alessandro Troisi. Novel thermally activated delayed fluorescence materials by high-throughput virtual screening: going beyond donor–acceptor design. *J. Mater. Chem. C*, 9:3324–3333, 2021.

[14] Richard J. Bravo and Nagy Y. Abed. Experimental evaluation of the harmonic behavior of led light bulb. In *2013 IEEE Power & Energy Society General Meeting*, pages 1–4, 2013.

[15] F.M. Steranka, J. Bhat, D. Collins, L. Cook, M.G. Craford, R. Fletcher, N. Gardner, P. Grillot, W. Goetz, M. Keuper, R. Khare, A. Kim, M. Krames, G. Harbers, M. Ludowise, P.S. Martin, M. Misra, G. Mueller, R. Mueller-Mach, S. Rudaz, Y.-C. Shen, D. Steigerwald, S. Stockman, S. Subramanya, T. Trottier, and J.J. Wierer. High power leds – technology status and market applications. *physica status solidi (a)*, 194(2):380–388, 2002.

[16] N. Holonyak, Jr. and S. F. Bevacqua. Coherent (visible) Light Emission from $Ga(As_{1-x}P_x)$ Junctions. *Applied Physics Letters*, 1:82–83, December 1962.

[17] I. Akasaki, H. Amano, S. Sota, H. Sakai, T. Tanaka, and M. Koike. Stimulated emission by current injection from an algan/gan/gainn quantmn well device. *Japanese Journal of Applied Physics*, 34(11):L1517, 1995.

[18] H. Amano, N. Sawaki, I. Akasaki, and Y. Toyoda. Metalorganic vapor phase epitaxial growth of a high quality gan film using an aln buffer layer. *Applied Physics Letters*, 48(5):353–355, 1986.

[19] Hiroshi Amano, Isamu Akasaki, Takahiro Kozawa, Kazumasa Hiramatsu, Nobuhiko Sawaki, Kousuke Ikeda, and Yoshikazu Ishii. Electron beam effects on blue luminescence of zinc-doped gan. *Journal of Luminescence*, 40-41:121 – 122, 1988.

[20] Hiroshi Amano, Masahiro Kito, Kazumasa Hiramatsu, and Isamu Akasaki. P-type conduction in mg-doped GaN treated with low-energy electron beam irradiation (LEEBI). *Japanese Journal of Applied Physics*, 28(Part 2, No. 12):L2112–L2114, dec 1989.

[21] Gerard Harbers and Christoph Hoelen. Lp-2: High performance lcd backlighting using high intensity red, green and blue light emitting diodes. *SID Symposium Digest of Technical Papers*, 32(1):702–705, 2001.

[22] Gerard Harbers, Wim Timmers, and Willem Sillevis-Smitt. Led backlighting for lcd hdtv. *Journal of the Society for Information Display*, 10(4):347–350, 2002.

[23] Yoon Hwa Kim, Noolu S. Viswanath, Sanjith Unithrattil, Ha Jun Kim, and Won Bin Im. Review—phosphor plates for high-power led applications: Challenges and opportunities toward perfect lighting. *ECS Journal of Solid State Science and Technology*, 7(1), 2017.

[24] Seonil Kwon, Hyuncheol Kim, Seungyeop Choi, Eun Gyo Jeong, Dohong Kim, Somin Lee, Ho Seung Lee, Young Cheol Seo, and Kyung Cheol Choi. Weavable and highly efficient organic light-emitting fibers for wearable electronics: A scalable, low-temperature process. *Nano Letters*, 18(1):347–356, 2018. PMID: 29210590.

[25] Sebastian Scholz, Denis Kondakov, Björn Lüssem, and Karl Leo. Degradation mechanisms and reactions in organic light-emitting devices. *Chemical Reviews*, 115(16):8449–8503, 2015. PMID: 26230864.

[26] Sujith Sudheendran Swayamprabha, Deepak Kumar Dubey, Shahnawaz, Rohit Ashok Kumar Yadav, Mangey Ram Nagar, Aayushi Sharma, Fu-Ching Tung, and Jwo-Huei

Jou. Approaches for long lifetime organic light emitting diodes. *Advanced Science*, 8(1):2002254, 2021.

[27] Jessika Luth Richter, Leena Tähkämö, and Carl Dalhammar. Trade-offs with longer lifetimes? the case of led lamps considering product development and energy contexts. *Journal of Cleaner Production*, 226:195–209, 2019.

[28] A. Bernanose, M. Comte, and P. Vouaux. Sur'le mechanisme de i'electroluminescence organique. *J. Chem. Phys.*, 52:396, 1955.

[29] M. Pope, H. P. Kallmann, and P. Magnante. Electroluminescence in organic crystals. *The Journal of Chemical Physics*, 38(8):2042–2043, 1963.

[30] C. W. Tang and S. A. VanSlyke. Organic electroluminescent diodes. *Applied Physics Letters*, 51(12):913–915, 1987.

[31] Sean C. Xia, Raymond C. Kwong, Vadim I. Adamovich, Michael S. Weaver, and Julie J. Brown. Oled device operational lifetime: Insights and challenges. In *2007 IEEE International Reliability Physics Symposium Proceedings. 45th Annual*, pages 253–257, 2007.

[32] Zakya H. Kafafi. *Organic electroluminescence*. CRC Press, 2005.

[33] J. Kalinowski, L. C. Palilis, W. H. Kim, and Z. H. Kafafi. Determination of the width of the carrier recombination zone in organic light-emitting diodes. *Journal of Applied Physics*, 94(12):7764–7767, 2003.

[34] C. Rothe, H. A. Al Attar, and A. P. Monkman. Absolute measurements of the triplet-triplet annihilation rate and the charge-carrier recombination layer thickness in working polymer light-emitting diodes based on polyspirobifluorene. *Phys. Rev. B*, 72:155330, Oct 2005.

[35] Martin Reufer, Manfred J. Walter, Pavlos G. Lagoudakis, Anne Beate Hummel, Johanna S. Kolb, Hartmut G. Roskos, Ullrich Scherf, and John M. Lupton. Spin-conserving carrier recombination in conjugated polymers. *Nature Materials*, 4(4):340–346, 4 2005.

[36] William Barford. Theory of singlet exciton yield in light-emitting polymers. *Phys. Rev. B*, 70:205204, Nov 2004.

[37] Stoyan Karabunarliev and Eric R. Bittner. Spin-dependent electron-hole capture kinetics in luminescent conjugated polymers. *Phys. Rev. Lett.*, 90:057402, Feb 2003.

[38] M. A. Baldo, D. F. Obrien, Y. You, A. Shoustikov, S. Sibley, M. E. Thompson, and S. R. Forrest. Highly efficient phosphorescent emission from organic electroluminescent devices. *Nature*, 395(6698):151–154, 1998.

[39] Hui Xu, Runfeng Chen, Qiang Sun, Wenyong Lai, Qianqian Su, Wei Huang, and Xiaogang Liu. Recent progress in metal–organic complexes for optoelectronic applications. *Chem. Soc. Rev.*, 43:3259–3302, 2014.

[40] Dandan Song, Suling Zhao, Yichun Luo, and Hany Aziz. Causes of efficiency roll-off in phosphorescent organic light emitting devices: Triplet-triplet annihilation versus triplet-polaron quenching. *Applied Physics Letters*, 97(24):243304, 2010.

[41] N. C. Giebink, B. W. D'Andrade, M. S. Weaver, J. J. Brown, and S. R. Forrest. Direct evidence for degradation of polaron excited states in organic light emitting diodes. *Journal of Applied Physics*, 105(12):124514, 2009.

[42] J. Partee, E. L. Frankevich, B. Uhlhorn, J. Shinar, Y. Ding, and T. J. Barton. Delayed fluorescence and triplet-triplet annihilation in π-conjugated polymers. *Phys. Rev. Lett.*, 82:3673–3676, May 1999.

[43] Debangshu Chaudhuri, Eva Sigmund, Annemarie Meyer, Lisa Röck, Philippe Klemm, Sebastian Lautenschlager, Agnes Schmid, Shane R. Yost, Troy VanVoorhis, Sebastian Bange, Sigurd Höger, and John M. Lupton. Metal-free oled triplet emitters by side-stepping kasha's rule. *Angewandte Chemie International Edition*, 52(50):13449–13452, 2013.

[44] Weijun Li, Dandan Liu, Fangzhong Shen, Dongge Ma, Zhiming Wang, Tao Feng, Yuanxiang Xu, Bing Yang, and Yuguang Ma. A twisting donor-acceptor molecule with an

intercrossed excited state for highly efficient, deep-blue electroluminescence. *Advanced Functional Materials*, 22(13):2797–2803, 2012.

[45] Weijun Li, Yuyu Pan, Ran Xiao, Qiming Peng, Shitong Zhang, Dongge Ma, Feng Li, Fangzhong Shen, Yinghui Wang, Bing Yang, and Yuguang Ma. Employing 100% excitons in oleds by utilizing a fluorescent molecule with hybridized local and charge-transfer excited state. *Advanced Functional Materials*, 24(11):1609–1614, 2014.

[46] Yuyu Pan, Weijun Li, Shitong Zhang, Liang Yao, Cheng Gu, Hai Xu, Bing Yang, and Yuguang Ma. High yields of singlet excitons in organic electroluminescence through two paths of cold and hot excitons. *Advanced Optical Materials*, 2(6):510–515, 2014.

[47] Liang Yao, Shitong Zhang, Rong Wang, Weijun Li, Fangzhong Shen, Bing Yang, and Yuguang Ma. Highly efficient near-infrared organic light-emitting diode based on a butterfly-shaped donor–acceptor chromophore with strong solid-state fluorescence and a large proportion of radiative excitons. *Angewandte Chemie International Edition*, 53(8):2119–2123, 2014.

[48] Shitong Zhang, Weijun Li, Liang Yao, Yuyu Pan, Fangzhong Shen, Ran Xiao, Bing Yang, and Yuguang Ma. Enhanced proportion of radiative excitons in non-doped electro-fluorescence generated from an imidazole derivative with an orthogonal donor–acceptor structure. *Chem. Commun.*, 49:11302–11304, 2013.

[49] Q. Zhang, Q. Zhou, Y. Cheng, L. Wang, D. Ma, X. Jing, and F. Wang. Highly efficient green phosphorescent organic light-emitting diodes based on cui complexes. *Advanced Materials*, 16(5):432–436, 2004.

[50] Hiroki Uoyama, Kenichi Goushi, Katsuyuki Shizu, Hiroko Nomura, and Chihaya Adachi. Highly efficient organic light-emitting diodes from delayed fluorescence. *Nature*, 492:234 EP –, Dec 2012.

[51] Chihaya Adachi. Third-generation organic electroluminescence materials. *Japanese Journal of Applied Physics*, 53(6):060101, 2014.

[52] C. A. Parker and C. G. Hatchard. Triplet-singlet emission in fluid solutions. phosphorescence of eosin. *Trans. Faraday Soc.*, 57:1894–1904, 1961.

[53] R.E. Brown, L.A. Singer, and J.H. Parks. Prompt and delayed fluorescence from benzophenone. *Chemical Physics Letters*, 14(2):193 – 195, 1972.

[54] P.F Jones and A.R Calloway. Temperature effects on the intramolecular decay of the lowest triplet state of benzophenone. *Chemical Physics Letters*, 10(4):438 – 443, 1971.

[55] Bill S. Yamanashi and David M. Hercules. Solution luminescence of anthraquinone, naphthoquinone, and some derivatives at room temperature. *Applied Spectroscopy*, 25(4):457–460, 1971.

[56] Andrzej Maciejewski, Marian Szymanski, and Ronald P. Steer. Thermally activated delayed s1 fluorescence of aromatic thiones. *The Journal of Physical Chemistry*, 90(23):6314–6318, 1986.

[57] M. Szymanski, A. Maciejewski, and R. P. Steer. Radiationless decay of aromatic thiones in solution selectively excited to their s3, s2, s1, and t1 states. *The Journal of Physical Chemistry*, 92(9):2485–2489, 1988.

[58] Marek Sikorski, Igor V. Khmelinskii, Wlodzimierz Augustyniak, and Francis Wilkinson. Triplet state decay of some thioketones in solution. *J. Chem. Soc., Faraday Trans.*, 92:3487–3490, 1996.

[59] George Blasse and David R. McMillin. On the luminescence of bis (triphenylphosphine) phenanthroline copper (i). *Chemical Physics Letters*, 70(1):1 – 3, 1980.

[60] Ayataka Endo, Mai Ogasawara, Atsushi Takahashi, Daisuke Yokoyama, Yoshimine Kato, and Chihaya Adachi. Thermally activated delayed fluorescence from sn4+–porphyrin complexes and their application to organic light emitting diodes — a novel mechanism for electroluminescence. *Advanced Materials*, 21(47):4802–4806, 2009.

[61] Ayataka Endo, Keigo Sato, Kazuaki Yoshimura, Takahiro Kai, Atsushi Kawada, Hiroshi Miyazaki, and Chihaya Adachi. Efficient up-conversion of triplet excitons into a sin-

glet state and its application for organic light emitting diodes. *Applied Physics Letters*, 98(8):083302, 2011.

[62] Wolfgang Brütting, Jörg Frischeisen, Tobias D. Schmidt, Bert J. Scholz, and Christian Mayr. Device efficiency of organic light-emitting diodes: Progress by improved light outcoupling. *physica status solidi (a)*, 210(1):44–65, 2013.

[63] Ye Tao, Kai Yuan, Ting Chen, Peng Xu, Huanhuan Li, Runfeng Chen, Chao Zheng, Lei Zhang, and Wei Huang. Thermally activated delayed fluorescence materials towards the breakthrough of organoelectronics. *Advanced Materials*, 26(47):7931–7958, 2014.

[64] Katsuyuki Shizu, Yumi Sakai, Hiroyuki Tanaka, Shuzo Hirata, Chihaya Adachi, and Hironori Kaji. [paper] meta-linking strategy for thermally activated delayed fluorescence emitters with a small singlet-triplet energy gap. *ITE Transactions on Media Technology and Applications*, 3(2):108–113, 2015.

[65] Kyungeon Lee and Dongwook Kim. Local-excitation versus charge-transfer characters in the triplet state: Theoretical insight into the singletâ€“triplet energy differences of carbazolyl-phthalonitrile-based thermally activated delayed fluorescence materials. *The Journal of Physical Chemistry C*, 120(49):28330–28336, 2016.

[66] Pralok K. Samanta, Dongwook Kim, Veaceslav Coropceanu, and Jean-Luc Brédas. Up-conversion intersystem crossing rates in organic emitters for thermally activated delayed fluorescence: Impact of the nature of singlet vs triplet excited states. *Journal of the American Chemical Society*, 139(11):4042–4051, 2017. PMID: 28244314.

[67] Richard L. Martin. Natural transition orbitals. *The Journal of Chemical Physics*, 118(11):4775–4777, 2003.

[68] Jing Zhang, Dongxue Ding, Ying Wei, and Hui Xu. Extremely condensing triplet states of dpepo-type hosts through constitutional isomerization for high-efficiency deep-blue thermally activated delayed fluorescence diodes. *Chem. Sci.*, 7:2870–2882, 2016.

[69] Minghan Cai, Morgan Auffray, Dongdong Zhang, Yuewei Zhang, Ryo Nagata, Zesen Lin, Xun Tang, Chin-Yiu Chan, Yi-Ting Lee, Tianyu Huang, Xiaozeng Song, Youichi

Tsuchiya, Chihaya Adachi, and Lian Duan. Enhancing spin-orbital coupling in deep-blue/blue tadf emitters by minimizing the distance from the heteroatoms in donors to acceptors. *Chemical Engineering Journal*, 420:127591, 2021.

[70] Marc K. Etherington, Jamie Gibson, Heather F. Higginbotham, Thomas J. Penfold, and Andrew P. Monkman. Revealing the spin–vibronic coupling mechanism of thermally activated delayed fluorescence. *Nature Communications*, 7(1), 2016.

[71] Yoann Olivier, Mónica Moral, Luca Muccioli, and Juan-Carlos Sancho-García. Dynamic nature of excited states of donor–acceptor tadf materials for oleds: how theory can reveal structure–property relationships. *J. Mater. Chem. C*, 5:5718–5729, 2017.

[72] Takuji Hatakeyama, Kazushi Shiren, Kiichi Nakajima, Shintaro Nomura, Soichiro Nakatsuka, Keisuke Kinoshita, Jingping Ni, Yohei Ono, and Toshiaki Ikuta. Ultrapure blue thermally activated delayed fluorescence molecules: Efficient homo–lumo separation by the multiple resonance effect. *Advanced Materials*, 28(14):2777–2781, 2016.

[73] Xiao Liang, Zhen-Long Tu, and You-Xuan Zheng. Thermally activated delayed fluorescence materials: Towards realization of high efficiency through strategic small molecular design. *Chemistry – A European Journal*, 25(22):5623–5642, 2019.

[74] Rongchuan Su and Zhenmei Huang. A series of singlet-triplet inverted tadf fluorescent probes with high stability, low molecular weight, and synthesis accessibility. *Advanced Theory and Simulations*, 6(5):2200863, 2023.

[75] Xian-Kai Chen, Youichi Tsuchiya, Yuma Ishikawa, Cheng Zhong, Chihaya Adachi, and Jean-Luc Brédas. A new design strategy for efficient thermally activated delayed fluorescence organic emitters: From twisted to planar structures. *Advanced Materials*, 29(46):1702767, 2017.

[76] Brian K. Shoichet. Virtual screening of chemical libraries. 432(7019):862–865.

[77] J. Gibson and T. J. Penfold. Nonadiabatic coupling reduces the activation energy in thermally activated delayed fluorescence. *Phys. Chem. Chem. Phys.*, 19:8428–8434, 2017.

[78] Chr. Møller and M. S. Plesset. Note on an approximation treatment for many-electron systems. *Phys. Rev.*, 46:618–622, Oct 1934.

[79] Johannes T. Margraf, Ajith Perera, Jesse J. Lutz, and Rodney J. Bartlett. Single-reference coupled cluster theory for multi-reference problems. *The Journal of Chemical Physics*, 147(18):184101, 2017.

[80] P. Hohenberg and W. Kohn. Inhomogeneous electron gas. *Phys. Rev.*, 136:B864–B871, Nov 1964.

[81] W. Kohn and L. J. Sham. Self-consistent equations including exchange and correlation effects. *Phys. Rev.*, 140:A1133–A1138, Nov 1965.

[82] R. O. Jones and O. Gunnarsson. The density functional formalism, its applications and prospects. *Rev. Mod. Phys.*, 61:689–746, Jul 1989.

[83] W. Kohn. Nobel lecture: Electronic structure of matter—wave functions and density functionals. *Rev. Mod. Phys.*, 71:1253–1266, Oct 1999.

[84] R. O. Jones. Density functional theory: Its origins, rise to prominence, and future. *Rev. Mod. Phys.*, 87:897–923, Aug 2015.

[85] M. Born and R. Oppenheimer. Zur quantentheorie der molekeln. *Annalen der Physik*, 389(20):457–484, 1927.

[86] R. F. Bishop. An overview of coupled cluster theory and its applications in physics. *Theoretica Chimica Acta*, 80(2–3):95–148, 1991.

[87] Erin R. Johnson and Axel D. Becke. A post-Hartree–Fock model of intermolecular interactions. *The Journal of Chemical Physics*, 123(2):024101, 07 2005.

[88] John P. Perdew and Karla Schmidt. Jacob's ladder of density functional approximations for the exchange-correlation energy. *AIP Conference Proceedings*, 577(1):1–20, 2001.

[89] Chen Li, Xiao Zheng, Aron J. Cohen, Paula Mori-Sánchez, and Weitao Yang. Local scaling correction for reducing delocalization error in density functional approximations. *Phys. Rev. Lett.*, 114:053001, Feb 2015.

[90] Jianwei Sun, Adrienn Ruzsinszky, and John P. Perdew. Strongly constrained and appropriately normed semilocal density functional. *Phys. Rev. Lett.*, 115:036402, Jul 2015.

[91] Neil Qiang Su and Xin Xu. Toward the construction of parameter-free doubly hybrid density functionals. *International Journal of Quantum Chemistry*, 115(10):589–595, 2015.

[92] Christopher J. Cramer and Donald G. Truhlar. Density functional theory for transition metals and transition metal chemistry. *Phys. Chem. Chem. Phys.*, 11:10757–10816, 2009.

[93] Yan Zhao and Donald G. Truhlar. Density functional theory for reaction energies: Test of meta and hybrid meta functionals, range-separated functionals, and other high-performance functionals. *Journal of Chemical Theory and Computation*, 7(3):669–676, 2011.

[94] Haitao Sun, Cheng Zhong, and Jean-Luc Brédas. Reliable prediction with tuned range-separated functionals of the singlet–triplet gap in organic emitters for thermally activated delayed fluorescence. *Journal of Chemical Theory and Computation*, 11(8):3851–3858, 2015. PMID: 26574466.

[95] Nadia Salami and Aliasghar Shokri. Chapter 5 - electronic structure of solids and molecules. In Mehrorang Ghaedi, editor, *Photocatalysis: Fundamental Processes and Applications*, volume 32 of *Interface Science and Technology*, pages 325–373. Elsevier, 2021.

[96] John P. Perdew. *Can Density Functional Theory Describe Strongly Correlated Electronic Systems?*, pages 237–252. Springer US, Boston, MA, 2003.

[97] Takeshi Yanai, David P Tew, and Nicholas C Handy. A new hybrid exchange–correlation functional using the coulomb-attenuating method (cam-b3lyp). *Chemical Physics Letters*, 393(1):51–57, 2004.

[98] Mary A. Rohrdanz, Katie M. Martins, and John M. Herbert. A long-range-corrected density functional that performs well for both ground-state properties and time-dependent density functional theory excitation energies, including charge-transfer excited states. *The Journal of Chemical Physics*, 130(5):054112, 02 2009.

[99] Hisayoshi Iikura, Takao Tsuneda, Takeshi Yanai, and Kimihiko Hirao. A long-range correction scheme for generalized-gradient-approximation exchange functionals. 115(8):3540–3544. _eprint: https://pubs.aip.org/aip/jcp/article-pdf/115/8/3540/10837377/3540_1_online.pdf.

[100] Roberto Peverati and Donald G. Truhlar. Improving the accuracy of hybrid meta-gga density functionals by range separation. *The Journal of Physical Chemistry Letters*, 2(21):2810–2817, 2011.

[101] Jeng-Da Chai and Martin Head-Gordon. Systematic optimization of long-range corrected hybrid density functionals. *The Journal of Chemical Physics*, 128(8):084106, 02 2008.

[102] John P. Perdew, Kieron Burke, and Matthias Ernzerhof. Generalized gradient approximation made simple. *Phys. Rev. Lett.*, 77:3865–3868, Oct 1996.

[103] C. G. Broyden. A class of methods for solving nonlinear simultaneous equations. *Mathematics of Computation*, 19:577–593, 1965.

[104] Péter Pulay. Convergence acceleration of iterative sequences. the case of scf iteration. *Chemical Physics Letters*, 73(2):393–398, 1980.

[105] Erich Runge and E. K. U. Gross. Density-functional theory for time-dependent systems. *Phys. Rev. Lett.*, 52:997–1000, Mar 1984.

[106] MARK E. CASIDA. *Time-Dependent Density Functional Response Theory for Molecules*, pages 155–192.

[107] P. H. Dederichs, S. Blügel, R. Zeller, and H. Akai. Ground states of constrained systems: Application to cerium impurities. *Phys. Rev. Lett.*, 53:2512–2515, Dec 1984.

[108] Qin Wu and Troy Van Voorhis. Direct optimization method to study constrained systems within density-functional theory. *Phys. Rev. A*, 72:024502, Aug 2005.

[109] Andreas Dreuw, Jennifer L. Weisman, and Martin Head-Gordon. Long-range charge-transfer excited states in time-dependent density functional theory require non-local exchange. *The Journal of Chemical Physics*, 119(6):2943–2946, 2003.

[110] O. Gunnarsson and B. I. Lundqvist. Exchange and correlation in atoms, molecules, and solids by the spin-density-functional formalism. *Phys. Rev. B*, 13:4274–4298, May 1976.

[111] Andreas Görling. Density-functional theory beyond the hohenberg-kohn theorem. *Phys. Rev. A*, 59:3359–3374, May 1999.

[112] Eva Vandaele, Momir Mališ, and Sandra Luber. The SCF method for non-adiabatic dynamics of systems in the liquid phase. *The Journal of Chemical Physics*, 156(13):130901, 04 2022.

[113] Ankit Mahajan, Nick S. Blunt, Iliya Sabzevari, and Sandeep Sharma. Multireference configuration interaction and perturbation theory without reduced density matrices. *The Journal of Chemical Physics*, 151(21), dec 2019.

[114] R. Nazarov, L. Shulenburger, M. Morales, and Randolph Q. Hood. Benchmarking the pseudopotential and fixed-node approximations in diffusion monte carlo calculations of molecules and solids. *Phys. Rev. B*, 93:094111, Mar 2016.

[115] Pedro Borlido, Jan Doumont, Fabien Tran, Miguel A. L. Marques, and Silvana Botti. Validation of pseudopotential calculations for the electronic band gap of solids. *Journal of Chemical Theory and Computation*, 16(6):3620–3627, 2020. PMID: 32407117.

[116] D. R. Hamann, M. Schlüter, and C. Chiang. Norm-conserving pseudopotentials. *Phys. Rev. Lett.*, 43:1494–1497, Nov 1979.

[117] David Vanderbilt. Soft self-consistent pseudopotentials in a generalized eigenvalue formalism. *Phys. Rev. B*, 41:7892–7895, Apr 1990.

[118] Steven G. Louie, Sverre Froyen, and Marvin L. Cohen. Nonlinear ionic pseudopotentials in spin-density-functional calculations. *Phys. Rev. B*, 26:1738–1742, Aug 1982.

[119] Frank Jensen. Atomic orbital basis sets. *WIREs Computational Molecular Science*, 3(3):273–295, 2013.

[120] Stig Rune Jensen, Santanu Saha, José A. Flores-Livas, William Huhn, Volker Blum, Stefan Goedecker, and Luca Frediani. The elephant in the room of density functional

theory calculations. *The Journal of Physical Chemistry Letters*, 8(7):1449–1457, 2017. PMID: 28291362.

[121] Tanja Van Mourik, Angela K. Wilson, Kirk A. Peterson, David E. Woon, and Thom H. Dunning. The effect of basis set superposition error (BSSE) on the convergence of molecular properties calculated with the correlation consistent basis sets. In *Advances in Quantum Chemistry*, volume 31, pages 105–135. Elsevier.

[122] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. V. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. J. Bearpark, J. J. Heyd, E. N. Brothers, K. N. Kudin, V. N. Staroverov, T. A. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. P. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman, and D. J. Fox. Gaussian˜16 Revision C.01, 2016. Gaussian Inc. Wallingford CT.

[123] E. Aprà, E. J. Bylaska, W. A. de Jong, N. Govind, K. Kowalski, T. P. Straatsma, M. Valiev, H. J. J. van Dam, Y. Alexeev, J. Anchell, V. Anisimov, F. W. Aquino, R. Atta-Fynn, J. Autschbach, N. P. Bauman, J. C. Becca, D. E. Bernholdt, K. Bhaskaran-Nair, S. Bogatko, P. Borowski, J. Boschen, J. Brabec, A. Bruner, E. Cauët, Y. Chen, G. N. Chuev, C. J. Cramer, J. Daily, M. J. O. Deegan, T. H. Dunning, M. Dupuis, K. G. Dyall, G. I. Fann, S. A. Fischer, A. Fonari, H. Früchtl, L. Gagliardi, J. Garza, N. Gawande, S. Ghosh, K. Glaesemann, A. W. Götz, J. Hammond, V. Helms, E. D. Hermes, K. Hirao, S. Hirata, M. Jacquelin, L. Jensen, B. G. Johnson, H. Jónsson, R. A. Kendall, M. Klemm, R. Kobayashi, V. Konkov, S. Krishnamoorthy, M. Krishnan, Z. Lin, R. D. Lins, R. J.

Littlefield, A. J. Logsdail, K. Lopata, W. Ma, A. V. Marenich, J. Martin del Campo, D. Mejia-Rodriguez, J. E. Moore, J. M. Mullin, T. Nakajima, D. R. Nascimento, J. A. Nichols, P. J. Nichols, J. Nieplocha, A. Otero-de-la Roza, B. Palmer, A. Panyala, T. Pirojsirikul, B. Peng, R. Peverati, J. Pittner, L. Pollack, R. M. Richard, P. Sadayappan, G. C. Schatz, W. A. Shelton, D. W. Silverstein, D. M. A. Smith, T. A. Soares, D. Song, M. Swart, H. L. Taylor, G. S. Thomas, V. Tipparaju, D. G. Truhlar, K. Tsemekhman, T. Van Voorhis, Á. Vázquez-Mayagoitia, P. Verma, O. Villa, A. Vishnu, K. D. Vogiatzis, D. Wang, J. H. Weare, M. J. Williamson, T. L. Windus, K. Woliński, A. T. Wong, Q. Wu, C. Yang, Q. Yu, M. Zacharias, Z. Zhang, Y. Zhao, and R. J. Harrison. Nwchem: Past, present, and future. *The Journal of Chemical Physics*, 152(18):184102, 2020.

[124] Giuseppe M. J. Barca, Colleen Bertoni, Laura Carrington, Dipayan Datta, Nuwan De Silva, J. Emiliano Deustua, Dmitri G. Fedorov, Jeffrey R. Gour, Anastasia O. Gunina, Emilie Guidez, Taylor Harville, Stephan Irle, Joe Ivanic, Karol Kowalski, Sarom S. Leang, Hui Li, Wei Li, Jesse J. Lutz, Ilias Magoulas, Joani Mato, Vladimir Mironov, Hiroya Nakata, Buu Q. Pham, Piotr Piecuch, David Poole, Spencer R. Pruitt, Alistair P. Rendell, Luke B. Roskop, Klaus Ruedenberg, Tosaporn Sattasathuchana, Michael W. Schmidt, Jun Shen, Lyudmila Slipchenko, Masha Sosonkina, Vaibhav Sundriyal, Ananta Tiwari, Jorge L. Galvez Vallejo, Bryce Westheimer, Marta Wloch, Peng Xu, Federico Zahariev, and Mark S. Gordon. Recent developments in the general atomic and molecular electronic structure system. *The Journal of Chemical Physics*, 152(15):154102, April 2020.

[125] G. Kresse and J. Hafner. Ab initio molecular dynamics for liquid metals. *Phys. Rev. B*, 47:558–561, Jan 1993.

[126] G. Kresse and J. Furthmüller. Efficiency of ab-initio total energy calculations for metals and semiconductors using a plane-wave basis set. *Computational Materials Science*, 6(1):15–50, 1996.

[127] G. Kresse and J. Furthmüller. Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set. *Phys. Rev. B*, 54:11169–11186, Oct 1996.

[128] Paolo Giannozzi, Stefano Baroni, Nicola Bonini, Matteo Calandra, Roberto Car, Carlo Cavazzoni, Davide Ceresoli, Guido L Chiarotti, Matteo Cococcioni, Ismaila Dabo, Andrea Dal Corso, Stefano de Gironcoli, Stefano Fabris, Guido Fratesi, Ralph Gebauer, Uwe Gerstmann, Christos Gougoussis, Anton Kokalj, Michele Lazzeri, Layla Martin-Samos, Nicola Marzari, Francesco Mauri, Riccardo Mazzarello, Stefano Paolini, Alfredo Pasquarello, Lorenzo Paulatto, Carlo Sbraccia, Sandro Scandolo, Gabriele Sclauzero, Ari P Seitsonen, Alexander Smogunov, Paolo Umari, and Renata M Wentzcovitch. QUANTUM ESPRESSO: a modular and open-source software project for quantum simulations of materials. *Journal of Physics: Condensed Matter*, 21(39):395502, sep 2009.

[129] Volker Blum, Mariana Rossi, Sebastian Kokott, and Matthias Scheffler. The fhi-aims code: All-electron, ab initio materials simulations towards the exascale, 2022.

[130] Laura E. Ratcliff, William Dawson, Giuseppe Fisicaro, Damien Caliste, Stephan Mohr, Augustin Degomme, Brice Videau, Viviana Cristiglio, Martina Stella, Marco D'Alessandro, Stefan Goedecker, Takahito Nakajima, Thierry Deutsch, and Luigi Genovese. Flexibilities of wavelets as a computational basis set for large-scale electronic structure calculations. *The Journal of Chemical Physics*, 152(19):194110, 2020.

[131] Thomas D. Kühne, Marcella Iannuzzi, Mauro Del Ben, Vladimir V. Rybkin, Patrick Seewald, Frederick Stein, Teodoro Laino, Rustam Z. Khaliullin, Ole Schütt, Florian Schiffmann, Dorothea Golze, Jan Wilhelm, Sergey Chulkov, Mohammad Hossein Bani-Hashemian, Valéry Weber, Urban Borštnik, Mathieu Taillefumier, Alice Shoshana Jakobovits, Alfio Lazzaro, Hans Pabst, Tiziano Müller, Robert Schade, Manuel Guidon, Samuel Andermatt, Nico Holmberg, Gregory K. Schenter, Anna Hehn, Augustin Bussy, Fabian Belleflamme, Gloria Tabacchi, Andreas Glöß, Michael Lass, Iain Bethune, Christopher J. Mundy, Christian Plessl, Matt Watkins, Joost VandeVondele, Matthias Krack, and Jürg Hutter. Cp2k: An electronic structure and molecular dynamics software package - quickstep: Efficient and accurate electronic structure calculations. *The Journal of Chemical Physics*, 152(19):194103, 2020.

[132] Joseph C. A. Prentice, Jolyon Aarons, James C. Womack, Alice E. A. Allen, Lampros Andrinopoulos, Lucian Anton, Robert A. Bell, Arihant Bhandari, Gabriel A. Bramley, Robert J. Charlton, Rebecca J. Clements, Daniel J. Cole, Gabriel Constantinescu, Fabiano Corsetti, Simon M.-M. Dubois, Kevin K. B. Duff, José María Escartín, Andrea Greco, Quintin Hill, Louis P. Lee, Edward Linscott, David D. O'Regan, Maximillian J. S. Phipps, Laura E. Ratcliff, Álvaro Ruiz Serrano, Edward W. Tait, Gilberto Teobaldi, Valerio Vitale, Nelson Yeung, Tim J. Zuehlsdorff, Jacek Dziedzic, Peter D. Haynes, Nicholas D. M. Hine, Arash A. Mostofi, Mike C. Payne, and Chris-Kriton Skylaris. The onetep linear-scaling density functional theory program. *The Journal of Chemical Physics*, 152(17):174111, 2020.

[133] H.-J. Werner, P. J. Knowles, G. Knizia, F. R. Manby, and M. Schütz. Molpro: a general-purpose quantum chemistry program package. *WIREs Comput Mol Sci*, 2:242–253, 2012.

[134] Devin A. Matthews, Lan Cheng, Michael E. Harding, Filippo Lipparini, Stella Stopkowicz, Thomas-C. Jagau, Péter G. Szalay, Jürgen Gauss, and John F. Stanton. Coupled-cluster techniques for computational chemistry: The cfour program package. *The Journal of Chemical Physics*, 152(21):214108, 2020.

[135] Luigi Genovese, Alexey Neelov, Stefan Goedecker, Thierry Deutsch, Seyed Alireza Ghasemi, Alexander Willand, Damien Caliste, Oded Zilberberg, Mark Rayson, Anders Bergman, and Reinhold Schneider. Daubechies wavelets as a basis set for density functional pseudopotential calculations. *The Journal of Chemical Physics*, 129(1):014109, 2008.

[136] Laura E. Ratcliff, Stephan Mohr, Georg Huhs, Thierry Deutsch, Michel Masella, and Luigi Genovese. Challenges in large scale quantum mechanical calculations. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 7(1):e1290, 2017.

[137] Stephan Mohr, Laura E. Ratcliff, Paul Boulanger, Luigi Genovese, Damien Caliste, Thierry Deutsch, and Stefan Goedecker. Daubechies wavelets for linear scaling density functional theory. *The Journal of Chemical Physics*, 140(20):204110, 2014.

[138] Martina Stella, Kritam Thapa, Luigi Genovese, and Laura E. Ratcliff. Transition-based constrained dft for the robust and reliable treatment of excitations in supramolecular systems. *Journal of Chemical Theory and Computation*, 18(5):3027–3038, 2022. PMID: 35471972.

[139] Alex Willand, Yaroslav O. Kvashnin, Luigi Genovese, Álvaro Vázquez-Mayagoitia, Arpan Krishna Deb, Ali Sadeghi, Thierry Deutsch, and Stefan Goedecker. Norm-conserving pseudopotentials with chemical accuracy compared to all-electron calculations. *The Journal of Chemical Physics*, 138(10):104109, 2013.

[140] S. Goedecker, M. Teter, and J. Hutter. Separable dual-space gaussian pseudopotentials. *Phys. Rev. B*, 54:1703–1710, Jul 1996.

[141] C. Hartwigsen, S. Goedecker, and J. Hutter. Relativistic separable dual-space gaussian pseudopotentials from h to rn. *Phys. Rev. B*, 58:3641–3662, Aug 1998.

[142] Victoria T. Lim, David F. Hahn, Gary Tresadern, Christopher I. Bayly, and David L. Mobley. Benchmark assessment of molecular geometries and energies from small molecule force fields. 9:1390.

[143] Andrea Cesari, Sandro Bottaro, Kresten Lindorff-Larsen, Pavel Banáš, Jiří Šponer, and Giovanni Bussi. Fitting corrections to an rna force field using experimental data. *Journal of Chemical Theory and Computation*, 15(6):3425–3431, 2019. PMID: 31050905.

[144] Alexey Odinokov, Alexander Yakubovich, Won-Joon Son, Yongsik Jung, and Hyeonho Choi. Exploiting the quantum mechanically derived force field for functional materials simulations. 7(1):155.

[145] Jürgen Köfinger and Gerhard Hummer. Empirical optimization of molecular simulation force fields by bayesian inference. 94(12):245.

[146] Edward O. Pyzer-Knapp, Changwon Suh, Rafael Gómez-Bombarelli, Jorge Aguilera-Iparraguirre, and Alán Aspuru-Guzik. What is high-throughput virtual screening? a perspective from organic materials discovery. *Annual Review of Materials Research*, 45(1):195–216, 2015.

[147] Thomas A. Halgren. Merck molecular force field. i. basis, form, scope, parameterization, and performance of mmff94. *Journal of Computational Chemistry*, 17(5-6):490–519, 1996.

[148] Thomas A. Halgren. Merck molecular force field. ii. mmff94 van der waals and electrostatic parameters for intermolecular interactions. *Journal of Computational Chemistry*, 17(5-6):520–552, 1996.

[149] E.B. Wilson, J.C. Decius, and P.C. Cross. *Molecular Vibrations: The Theory of Infrared and Raman Vibrational Spectra*. Dover Books on Chemistry Series. Dover Publications, 1980.

[150] Thorsten Schnabel, Jadran Vrabec, and Hans Hasse. Unlike lennard–jones parameters for vapor–liquid equilibria. *Journal of Molecular Liquids*, 135(1-3):170–178, 2007.

[151] Paolo Tosco, Nikolaus Stiefl, and Gregory Landrum. Bringing the MMFF force field to the RDKit: implementation and validation. 6(1):37.

[152] Zheng Tan, Yan Li, Ziying Zhang, Xin Wu, Thomas Penfold, Weimei Shi, and Shiqing Yang. Efficient adversarial generation of thermally activated delayed fluorescence molecules. *ACS Omega*, 7(21):18179–18188, 2022.

[153] Thomas A. Halgren. Mmff vi. mmff94s option for energy minimization studies. *Journal of Computational Chemistry*, 20(7):720–729, 1999.

[154] Katarina Roos, Chuanjie Wu, Wolfgang Damm, Mark Reboul, James M. Stevenson, Chao Lu, Markus K. Dahlgren, Sayan Mondal, Wei Chen, Lingle Wang, Robert Abel, Richard A. Friesner, and Edward D. Harder. Opls3e: Extending force field coverage for drug-like small molecules. *Journal of Chemical Theory and Computation*, 15(3):1863–1874, 2019. PMID: 30768902.

[155] David L. Mobley, Caitlin C. Bannan, Andrea Rizzi, Christopher I. Bayly, John D. Chodera, Victoria T. Lim, Nathan M. Lim, Kyle A. Beauchamp, David R. Slochower, Michael R. Shirts, Michael K. Gilson, and Peter K. Eastman. Escaping atom types in force fields using direct chemical perception. *Journal of Chemical Theory and Computation*, 14(11):6076–6092, 2018. PMID: 30351006.

[156] Tobias Fink and Jean-Louis Reymond. Virtual exploration of the chemical universe up to 11 atoms of c, n, o, f: assembly of 26.4 million structures (110.9 million stereoisomers) and analysis for new ring systems, stereochemistry, physicochemical properties, compound classes, and drug discovery. *Journal of Chemical Information and Modeling*, 47(2):342–353, 2007. PMID: 17260980.

[157] Stefano Curtarolo, Gus L. W. Hart, Marco Buongiorno Nardelli, Natalio Mingo, Stefano Sanvito, and Ohad Levy. The high-throughput highway to computational materials design. 12(3):191–201.

[158] Ömer H. Omar, Marcos del Cueto, Tahereh Nematiaram, and Alessandro Troisi. High-throughput virtual screening for organic electronics: a comparative study of alternative strategies. *J. Mater. Chem. C*, 9:13557–13583, 2021.

[159] Yinan Shu and Benjamin G. Levine. Simulated evolution of fluorophores for light emitting diodes. *The Journal of Chemical Physics*, 142(10):104104, 2015.

[160] Rafael Gómez-Bombarelli, Jorge Aguilera-Iparraguirre, Timothy D. Hirzel, David Duvenaud, Dougal Maclaurin, Martin A. Blood-Forsythe, Hyun Sik Chae, Markus Einzinger, Dong-Gwang Ha, Tony Wu, Georgios Markopoulos, Soonok Jeon, Hosuk Kang, Hiroshi Miyazaki, Masaki Numata, Sunghan Kim, Wenliang Huang, Seong Ik Hong, Marc Baldo, Ryan P. Adams, and Alán Aspuru-Guzik. Design of efficient molecular organic light-emitting diodes by a high-throughput virtual screening and experimental approach. 15(10):1120–1127.

[161] Shao-Yu Lu, Sukrit Mukhopadhyay, Robert Froese, and Paul M. Zimmerman. Virtual screening of hole transport, electron transport, and host layers for effective OLED design. 58(12):2440–2449. Publisher: American Chemical Society.

[162] Ping Li, Zijie Wang, Wenjing Li, Jie Yuan, and Runfeng Chen. Design of thermally activated delayed fluorescence materials with high intersystem crossing efficiencies by machine learning-assisted virtual screening. *The Journal of Physical Chemistry Letters*, 13(42):9910–9918, 2022. PMID: 36256799.

[163] Chunyun Tu, Weijiang Huang, Sheng Liang, Kui Wang, Qin Tian, and Wei Yan. Combining machine learning and quantum chemical calculations for high-throughput virtual screening of thermally activated delayed fluorescence molecular materials: the impact of selection strategy and structural mutations. *RSC Adv.*, 12:30962–30975, 2022.

[164] Shide Liang and Nick V. Grishin. Side-chain modeling with an optimized scoring function. *Protein Science*, 11(2):322–331, 2002.

[165] Mingliang Wang, Xiangqian Hu, David N. Beratan, and Weitao Yang. Designing molecules by optimizing potentials. *Journal of the American Chemical Society*, 128(10):3228–3232, 2006. PMID: 16522103.

[166] Robert Balawender, Meressa A. Welearegay, Michał Lesiuk, Frank De Proft, and Paul Geerlings. Exploring chemical space with the alchemical derivatives. *Journal of Chemical Theory and Computation*, 9(12):5327–5340, 2013. PMID: 26592270.

[167] Meihua Tu, Brajesh K. Rai, Alan M. Mathiowetz, Mary Didiuk, Jeffrey A. Pfefferkorn, Angel Guzman-Perez, John Benbow, Cristiano R. W. Guimarães, Scot Mente, Matthew M. Hayward, and Spiros Liras. Exploring aromatic chemical space with neat: Novel and electronically equivalent aromatic template. *Journal of Chemical Information and Modeling*, 52(5):1114–1123, 2012. PMID: 22486394.

[168] Aaron M. Virshup, Julia Contreras-García, Peter Wipf, Weitao Yang, and David N. Beratan. Stochastic voyages into uncharted chemical space produce a representative library of all possible drug-like compounds. *Journal of the American Chemical Society*, 135(19):7296–7303, 2013. PMID: 23548177.

[169] Hans-Christian Ehrlich, Angela M. Henzler, and Matthias Rarey. Searching for recursively defined generic chemical patterns in nonenumerated fragment spaces. *Journal of Chemical Information and Modeling*, 53(7):1676–1688, 2013. PMID: 23751070.

[170] David Hoksza, Petr Škoda, Milan Voršilák, and Daniel Svozil. Molpher: a software framework for systematic chemical space exploration. 6(1):7.

[171] Mario Krenn, Florian Häse, AkshatKumar Nigam, Pascal Friederich, and Alan Aspuru-Guzik. Self-referencing embedded strings (selfies): A 100% robust molecular string representation. *Machine Learning: Science and Technology*, 1(4):045024, oct 2020.

[172] AkshatKumar Nigam, Robert Pollice, Mario Krenn, Gabriel dos Passos Gomes, and Alan Aspuru-Guzik. Beyond generative models: Superfast traversal, optimization, novelty, exploration and discovery (stoned) algorithm for molecules using selfies. 2021.

[173] David Weininger. Smiles, a chemical language and information system. 1. introduction to methodology and encoding rules. *Journal of Chemical Information and Computer Sciences*, 28(1):31–36, 1988.

[174] Stephen R. Heller, Alan McNaught, Igor Pletnev, Stephen Stein, and Dmitrii Tchekhovskoi. InChI, the IUPAC international chemical identifier. 7(1):23.

[175] Andrei A. Gakh and Michael N. Burnett. Modular chemical descriptor language (mcdl): composition, connectivity, and supplementary modules. *Journal of Chemical Information and Computer Sciences*, 41(6):1494–1499, 2001. PMID: 11749574.

[176] Andrei A. Gakh, Michael N. Burnett, Sergei V. Trepalin, and Alexander V. Yarkov. Modular chemical descriptor language (MCDL): Stereochemical modules. 3(1):5.

[177] Sheila Ash, Malcolm A. Cline, R. Webster Homer, Tad Hurst, and Gregory B. Smith. Sybyl line notation (sln): a versatile language for chemical structure representation. *Journal of Chemical Information and Computer Sciences*, 37(1):71–79, 1997.

[178] H.-G. Rohbeck. Representation of structure description arranged linearly. In Jürgen Gmehling, editor, *Software Development in Chemistry 5*, pages 49–58, Berlin, Heidelberg, 1991. Springer Berlin Heidelberg.

[179] Wendy A. Warr. Representation of chemical structures. *WIREs Computational Molecular Science*, 1(4):557–579, 2011.

[180] Shungo Koichi, Satoru Iwata, Takeaki Uno, Hiroyuki Koshino, and Hiroko Satoh. Algorithm for advanced canonical coding of planar chemical structures that considers stere-

ochemical and symmetric information. *Journal of Chemical Information and Modeling*, 47(5):1734–1746, 2007. PMID: 17636943.

[181] Stoyan Karabunarliev, Julian Ivanov, and Ovanes Mekenyan. Coding of chemical structures based on a line notation. *Computers Chemistry*, 18(2):189–193, 1994.

[182] Shinsaku Fujita and Nobuya Tanaka. Xym notation for electronic communication of organic chemical structures. *Journal of Chemical Information and Computer Sciences*, 39(6):903–914, 1999.

[183] Cheng Wentang, Zhang Ying, and Yu Feibai. New computer representation for chemical structures: Two-level compact connectivity tables. *Journal of Chemical Information and Computer Sciences*, 33(4):604–608, 1993.

[184] Luciana Quadrelli, Vittorio Bareggi, and Sergio Spiga. A new linear representation of chemical structures. *Journal of Chemical Information and Computer Sciences*, 18(1):37–40, 1978.

[185] Hidetsugu Abe, Yoshihiro Kudo, Tohru Yamasaki, Kazuo Tanaka, Masahiro Sasaki, and Shinichi Sasaki. A convenient notation system for organic structure on the basis of connectivity stack. *Journal of Chemical Information and Computer Sciences*, 24(4):212–216, 1984.

[186] William J. Wiswesser. 107 years of line-formula notations (1861-1968). *Journal of Chemical Documentation*, 8(3):146–150, 1968.

[187] Noel M. O'Boyle. Towards a universal SMILES representation - a standard method to generate canonical SMILES based on the InChI. 4(1):22.

[188] Laurianne David, Amol Thakkar, Rocío Mercado, and Ola Engkvist. Molecular representations in AI-driven drug discovery: a review and practical guide. 12(1):56.

[189] Xin Chen and Charles H. Reynolds. Performance of similarity measures in 2d fragment-based similarity searching: comparison of structural descriptors and similarity coefficients.

Journal of Chemical Information and Computer Sciences, 42(6):1407–1414, 2002. PMID: 12444738.

[190] Naomie Salim, John Holliday, and Peter Willett. Combination of fingerprint-based similarity coefficients using data fusion. *Journal of Chemical Information and Computer Sciences*, 43(2):435–442, 2003. PMID: 12653506.

[191] Peter Willett. Similarity-based virtual screening using 2d fingerprints. *Drug Discovery Today*, 11(23):1046–1053, 2006.

[192] J.D. Holliday, C-Y. Hu, and P. Willett. Grouping of coefficients for the calculation of inter-molecular similarity and dissimilarity using 2d fragment bit-strings. *Combinatorial Chemistry High Throughput Screening*, 5(2):155–166, 2002.

[193] David Rogers and Mathew Hahn. Extended-connectivity fingerprints. *Journal of Chemical Information and Modeling*, 50(5):742–754, 2010. PMID: 20426451.

[194] H. L. Morgan. The generation of a unique machine description for chemical structures-a technique developed at chemical abstracts service. *Journal of Chemical Documentation*, 5(2):107–113, 1965.

[195] Daniel Probst and Jean-Louis Reymond. A probabilistic molecular fingerprint for big data settings. 10(1):66.

[196] Raymond E. Carhart, Dennis H. Smith, and R. Venkataraghavan. Atom pairs as molecular features in structure-activity studies: definition and applications. *Journal of Chemical Information and Computer Sciences*, 25(2):64–73, 1985.

[197] Alice Capecchi, Daniel Probst, and Jean-Louis Reymond. One molecular fingerprint to rule them all: drugs, biomolecules, and the metabolome. 12(1):43.

[198] Sereina Riniker and Gregory A. Landrum. Open-source platform to benchmark fingerprints for ligand-based virtual screening. 5(1):26.

[199] Mahendra Awale and Jean-Louis Reymond. *Web-Based Tools for Polypharmacology Prediction*, pages 255–272. Springer New York, New York, NY, 2019.

[200] Mahendra Awale and Jean-Louis Reymond. Polypharmacology browser ppb2: Target prediction combining nearest neighbors with machine learning. *Journal of Chemical Information and Modeling*, 59(1):10–17, 2019.

[201] Connor W. Coley, Luke Rogers, William H. Green, and Klavs F. Jensen. Scscore: Synthetic complexity learned from a reaction corpus. *Journal of Chemical Information and Modeling*, 58(2):252–261, 2018. PMID: 29309147.

[202] Peter Ertl and Ansgar Schuffenhauer. Estimation of synthetic accessibility score of drug-like molecules based on molecular complexity and fragment contributions. 1(1):8.

[203] Milan Voršilák, Michal Kolář, Ivan Čmelo, and Daniel Svozil. SYBA: Bayesian estimation of synthetic accessibility of organic compounds. 12(1):35.

[204] Elias James Corey. The logic of chemical synthesis: Multistep synthesis of complex carbogenic molecules (nobel lecture). *Angewandte Chemie International Edition in English*, 30(5):455–465, 1991.

[205] Reaxys. Accessed: 2023-01-16.

[206] Sunghwan Kim, Jie Chen, Tiejun Cheng, Asta Gindulyte, Jia He, Siqian He, Qingliang Li, Benjamin A Shoemaker, Paul A Thiessen, Bo Yu, Leonid Zaslavsky, Jian Zhang, and Evan E Bolton. PubChem 2019 update: improved access to chemical data. 47:D1102–D1109. _eprint: https://academic.oup.com/nar/article-pdf/47/D1/D1102/27437306/gky1033.pdf.

[207] John J. Irwin, Teague Sterling, Michael M. Mysinger, Erin S. Bolstad, and Ryan G. Coleman. Zinc: A free tool to discover chemistry for biology. *Journal of Chemical Information and Modeling*, 52(7):1757–1768, 2012. PMID: 22587354.

[208] Teague Sterling and John J. Irwin. Zinc 15 – ligand discovery for everyone. *Journal of Chemical Information and Modeling*, 55(11):2324–2337, 2015. PMID: 26479676.

[209] Milan Voršilák and Daniel Svozil. Nonpher: computational method for design of hard-to-synthesize structures. 9(1):20.

[210] Imma Ratera and Jaume Veciana. Playing with organic radicals as building blocks for functional molecular materials. *Chem. Soc. Rev.*, 41:303–349, 2012.

[211] So Hirata and Martin Head-Gordon. Time-dependent density functional theory for radicals: An improved description of excited states with substantial double excitation character. *Chemical Physics Letters*, 302(5):375–382, 1999.

[212] So Hirata and Martin Head-Gordon. Time-dependent density functional theory within the tamm–dancoff approximation. *Chemical Physics Letters*, 314(3):291–299, 1999.

[213] David C. Graham, Ambili S. Menon, Lars Goerigk, Stefan Grimme, and Leo Radom. Optimization and basis-set dependence of a restricted-open-shell form of b2-plyp double-hybrid density functional theory. *The Journal of Physical Chemistry A*, 113(36):9861–9873, 2009. PMID: 19645437.

[214] Martin Korth and Stefan Grimme. "mindless" dft benchmarking. *Journal of Chemical Theory and Computation*, 5(4):993–1003, 2009. PMID: 26609608.

[215] Leonard R. Maurer, Markus Bursch, Stefan Grimme, and Andreas Hansen. Assessing density functional theory for chemically relevant open-shell transition metal reactions. *Journal of Chemical Theory and Computation*, 17(10):6134–6151, 2021. PMID: 34546754.

[216] C.David Sherrill, Michael S. Lee, and Martin Head-Gordon. On the performance of density functional theory for symmetry-breaking problems. *Chemical Physics Letters*, 302(5):425–430, 1999.

[217] Qiming Peng, Ablikim Obolda, Ming Zhang, and Feng Li. Organic light-emitting diodes using a neutral radical as emitter: The emission from a doublet. *Angewandte Chemie International Edition*, 54(24):7091–7095, 2015.

[218] Xin Ai, Emrys W. Evans, Shengzhi Dong, Alexander J. Gillett, Haoqing Guo, Yingxin Chen, Timothy J. H. Hele, Richard H. Friend, and Feng Li. Efficient radical-based light-emitting diodes with doublet emission. 563(7732):536–540.

[219] Hee-Jun Park, Si Hyun Han, Jun Yeob Lee, Herim Han, and Eung-Gun Kim. Managing orientation of nitrogens in bipyrimidine-based thermally activated delayed fluorescent emitters to suppress nonradiative mechanisms. 30(10):3215–3222.

[220] Noel M. O'Boyle and Roger A. Sayle. Comparing structural fingerprints using a literature-based similarity benchmark. 8(1):36.

[221] Michał Mońka, Illia E. Serdiuk, Karol Kozakiewicz, Estera Hoffman, Jan Szumilas, Aleksander Kubicki, Soo Young Park, and Piotr Bojarski. Understanding the internal heavy-atom effect on thermally activated delayed fluorescence: application of arrhenius and marcus theories for spin–orbit coupling analysis. *J. Mater. Chem. C*, 10:7925–7934, 2022.

[222] Yirang Im, Mounggon Kim, Yong Joo Cho, Jeong-A Seo, Kyoung Soo Yook, and Jun Yeob Lee. Molecular design strategy of organic thermally activated delayed fluorescence emitters. *Chemistry of Materials*, 29(5):1946–1963, 2017.