Theoretical Screening of Organic Conjugated Materials

University College London

Isabelle Heath-Apostolopoulos

July 23, 2021
Abstract

As the urgency to address the effects of climate change increases, so does the need to discover new materials for the generation of renewable energy, to replace the environmentally damaging combustion of fossil fuels. Two main areas of research are focused on the design and discovery of photoactive materials for photovoltaics and for photocatalysts for water splitting. Although there are currently high-efficiency photovoltaics commercially available, there is motivation to replace them with materials with non-toxic and earth-abundant compositions. As they generally meet these criteria, organic conjugated materials are very desirable candidates for these applications. Computational chemistry methods can accelerate materials discovery, eliminating the need to synthesize large libraries of molecules in the preliminary screening stages. Both high-accuracy, expensive methods and fast, cheap, lower-accuracy methods have their merits and in conjunction with one another can provide a detailed and informative description of chemical systems. The high-throughput virtual screening methodology used throughout this thesis provides the opportunity to efficiently explore property space and highlight potential candidates for given applications, such as polymeric photocatalysts, organic photovoltaics and dye sensitizers in solar cells. In this thesis this methodology is explored for a small aromatic molecules, diketopyrrolopyrrole-based dyes and both ordered and disordered polymers. Through the high-throughput virtual screening, large datasets of chemical compounds were investigated and analysed, highlighting the patterns in the optical and electronic properties influenced by building block sequence, conformerism and composition. The use of high-accuracy, expensive methods is also explored in this thesis, demonstrating the difficulties in pushing such methods to larger chemical structures.
Impact Statement

A lot of research in the scientific community is currently focused on the discovery of novel materials for renewable energy applications, with organic conjugated materials being a large area of interest. Computational chemistry methods can provide many advantages over experiment in the preliminary stages of materials design. The high-throughput virtual screening methodology used in most of the work in this thesis, established by our group, has been validated for a variety of organic conjugated systems and as such, has been shown to be a useful method for the screening stage of materials discovery. This could eliminate time spent on synthesizing materials that turn out to be unsuitable for a particular application, and also accelerate the process of arriving to useful candidate materials.

Methodology aside, the work in this thesis has revealed a number of useful general design principles that can be taken into consideration in the future, either in further computational investigations or in experimental work. The effect of building block sequence on the optical properties has been explored extensively, and a number of routes to fine-tuning the optical gap have been demonstrated in many of the chapters. For example, it has been shown that by modifying the building block sequence in polymers, or using asymmetric rather than symmetric structures in dye molecules, the optical gap can be tuned. Additionally, patterns related to the compositions of the target materials were uncovered, for instance, how the relative properties of the constituent homopolymers of a binary co-polymer lead to the properties in the resulting co-polymer. In the same vein, different types of monomers, varying in their electron-richness/poorness, have been shown to combine in different types of materials to obtain different electronic potentials. The vast analysis of property space provides a general guideline that could aid researchers in subsequent investigations of organic conjugated materials. Most of the research in this thesis has been published in peer-reviewed journals.
Contents

Abstract iii
Impact Statement v
List of Figures xviii
List of Tables xx
List of Abbreviations xxi
List of Publications xxv

1 Introduction 1
  1.1 Optical and Electronic Properties of Materials . . . . . . . . . . . . . . . . . . . 3
  1.2 Conjugated Molecules and Polymers as Photoactive Materials . . . . . . . . . . 5
    1.2.1 Polymer Structure and Properties . . . . . . . . . . . . . . . . . . . . . 5
    1.2.2 Applications of Conjugated Organic Materials and Polymers as Pho-
        toactive Materials . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 8
  1.3 Multi-Level Computational Analysis of Materials . . . . . . . . . . . . . . . . . 13
    1.3.1 High-Throughput Virtual Screening . . . . . . . . . . . . . . . . . . . . . 14

2 Background and Theory 17
  2.1 Quantum Chemistry methods . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 17
    2.1.1 The Schrödinger Equation . . . . . . . . . . . . . . . . . . . . . . . . . . . 17
    2.1.2 The Born-Oppenheimer Approximation . . . . . . . . . . . . . . . . . . . 17
    2.1.3 Potential Energy Surfaces . . . . . . . . . . . . . . . . . . . . . . . . . . . 18
    2.1.4 Hartree-Fock Methods . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 20
  2.2 Density Functional Theory . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 22
    2.2.1 Early Density Functional Theory . . . . . . . . . . . . . . . . . . . . . . . 23
    2.2.2 Hohenberg and Kohn Theory . . . . . . . . . . . . . . . . . . . . . . . . . 24
    2.2.3 Kohn-Sham Theory . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 25
    2.2.4 The Exchange-Correlation Functional . . . . . . . . . . . . . . . . . . . . 26
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.2.5</td>
<td>Time-Dependent Density Functional Theory</td>
<td>29</td>
</tr>
<tr>
<td>2.3</td>
<td>Basis Sets</td>
<td>33</td>
</tr>
<tr>
<td>2.4</td>
<td>$\Delta$SCF</td>
<td>36</td>
</tr>
<tr>
<td>2.5</td>
<td>The GW method</td>
<td>36</td>
</tr>
<tr>
<td>2.6</td>
<td>Semi-empirical methods</td>
<td>37</td>
</tr>
<tr>
<td>2.7</td>
<td>Solvation Models</td>
<td>39</td>
</tr>
<tr>
<td>2.7.1</td>
<td>Explicit Solvation</td>
<td>39</td>
</tr>
<tr>
<td>2.7.2</td>
<td>Implicit Solvent Models</td>
<td>40</td>
</tr>
<tr>
<td>2.8</td>
<td>Simplified Molecular-Input Line Entry System (SMILES)</td>
<td>41</td>
</tr>
<tr>
<td>3</td>
<td>Accessing Near-DFT Results by Calibration of Properties from Semi-</td>
<td>43</td>
</tr>
<tr>
<td></td>
<td>Empirical Calculations Against DFT</td>
<td></td>
</tr>
<tr>
<td>3.1</td>
<td>Introduction</td>
<td>43</td>
</tr>
<tr>
<td>3.2</td>
<td>Methodology and Computational Details</td>
<td>45</td>
</tr>
<tr>
<td>3.2.1</td>
<td>Small Molecules Workflow</td>
<td>45</td>
</tr>
<tr>
<td>3.2.2</td>
<td>Dyes Workflow</td>
<td>46</td>
</tr>
<tr>
<td>3.3</td>
<td>Results and Discussion</td>
<td>46</td>
</tr>
<tr>
<td>3.3.1</td>
<td>Small Molecules</td>
<td>46</td>
</tr>
<tr>
<td>3.3.2</td>
<td>Dyes</td>
<td>51</td>
</tr>
<tr>
<td>3.4</td>
<td>Conclusions</td>
<td>55</td>
</tr>
<tr>
<td>4</td>
<td>Finding Diketopyrrolopyrrole Dyes for Dye-Sensitized Water Splitting</td>
<td>57</td>
</tr>
<tr>
<td></td>
<td>and Solar Cells</td>
<td></td>
</tr>
<tr>
<td>4.1</td>
<td>Introduction</td>
<td>57</td>
</tr>
<tr>
<td>4.2</td>
<td>Methodology and Computational Details</td>
<td>59</td>
</tr>
<tr>
<td>4.3</td>
<td>Results and Discussion</td>
<td>60</td>
</tr>
<tr>
<td>4.3.1</td>
<td>Property-Property Relationships</td>
<td>60</td>
</tr>
<tr>
<td>4.3.2</td>
<td>Property-Property Relationships</td>
<td>61</td>
</tr>
<tr>
<td>4.3.3</td>
<td>Effect of Asymmetry in ACA' Dyes</td>
<td>64</td>
</tr>
<tr>
<td>4.3.4</td>
<td>Dye Candidates for Dye-Sensitized Water Proton Reduction and Dye-</td>
<td>66</td>
</tr>
<tr>
<td></td>
<td>Sensitized Solar Cells</td>
<td></td>
</tr>
<tr>
<td>4.3.5</td>
<td>Structure-Property Relationships</td>
<td>68</td>
</tr>
<tr>
<td>4.4</td>
<td>Conclusions</td>
<td>76</td>
</tr>
<tr>
<td>5</td>
<td>Exploring the Effect of Composition, Sequence Isomerism and Conforma-</td>
<td>79</td>
</tr>
<tr>
<td></td>
<td>tional Degrees of Freedom with High-Throughput Virtual Screening</td>
<td></td>
</tr>
<tr>
<td>5.1</td>
<td>Introduction</td>
<td>79</td>
</tr>
<tr>
<td>5.2</td>
<td>Methodology and Computational Details</td>
<td>80</td>
</tr>
<tr>
<td>5.3</td>
<td>Results and Discussion</td>
<td>81</td>
</tr>
<tr>
<td>5.3.1</td>
<td>Conformational Degrees of Freedom</td>
<td>81</td>
</tr>
</tbody>
</table>
CONTENTS

5.3.2 Effect of Co-polymer Sequence ........................................ 83
5.4 Conclusions ............................................................. 85

6 High-Throughput Virtual Screening of Block Co-polymers .............. 89
6.1 Introduction ............................................................ 89
6.2 Methodology and Computational Details .................................. 90
6.3 Results and Discussion ................................................... 91
6.3.1 Calibration ........................................................... 91
6.3.2 Conformer Sampling .................................................... 92
6.3.3 Property-Property Relationships ....................................... 99
6.3.4 Intramolecular Charge Separation in Block Co-polymers ......... 100
6.3.5 Structure-Property Relationships ..................................... 101
6.4 Conclusions ............................................................. 110

7 Benchmarking DFT against GW .............................................. 113
7.1 Introduction ............................................................ 113
7.2 Methodology and Computational Details .................................. 114
7.3 Results and Discussion ................................................... 115
7.3.1 Trends in $-\text{IP}$ and $-\text{EA}$ .......................................... 116
7.3.2 Trends in Fundamental Gap and Optical Gap ....................... 117
7.3.3 Comparison of Exciton Binding Energies ............................ 119
7.3.4 Comparison of Materials ............................................. 120
7.4 Conclusions ............................................................. 122

8 Conclusions and Perspective .................................................. 125

A Appendix ..............................................................................
A.1 Calibration ............................................................ 129
A.2 Dyes ................................................................. 132
A.3 High-Throughput Virtual Screening of Block Co-polymers ............ 136
A.4 Benchmarking GW against DFT ......................................... 139

Bibliography .......................................................................... 149

Acknowledgements .................................................................. 151
List of Figures

1.1 Gaps relevant to the study of photoactive materials: −IP (HOMO), −EA (LUMO), fundamental gap (Δ_f), exciton binding energy (EBE) and optical gap (Δ_o). .............................................................. 4

1.2 The p-orbitals on the carbon atoms not involved in σ-bonding in 1,3,5-hexatriene interact to create a series of π-orbitals, which form a conjugated π-system. .............................................................. 6

1.3 Examples of different types of co-polymers (from top to bottom): fully alternating, alternating, block, periodic and random. .............................................................. 7

1.4 Process taking place during photocatalytic water splitting. Light hits the photoactive material, creating an exciton which separates into a hole and electron. The electron travels to the LUMO of the material eventually reducing H^+ to H_2 and the holes oxidise water to O_2. ................................. 9

2.1 1-dimensional energy profile for a ground state (S_0) and excited state (S_1). The arrows represent a vertical excitation (green), an adiabatic excitation (blue) and fluorescence (orange); A = local minimum, B = excited state local minimum, C = saddle point, D = local maximum, E = local minimum, F = conical intersection. 19

2.2 Radial distribution of hydrogen 1s basis functions in atom-centred coordinates; the dashed lines are the tight (blue), medium (orange) and loose (green) Gaussian functions, which are linearly combined to form the STO-3G (solid red) orbital. These are compared to the 1s-STO function (solid purple). 99 .............................. 34

2.3 Hexa-2,4-diene 2-D representation and SMILES string ........................................ 42

2.4 2-D representation and SMILES strings of bicyclo[4.3.0]nonane, cyclohexane, benzene and anthracene. .............................................................. 42

3.1 Twenty of the small molecules used in the calibration of the unsubstituted cores shown in the white box and the 12 substituents used to modulate these cores in the blue box. .............................................................. 44
3.2 The three dye sequences used in the following calibrations and the structures of the diketopyrrolopyrrole (DPP) core and a selection of the 179 building blocks used for the A, B and A’ positions. .................................................. 44
3.3 DFT-calculated (B3LYP/aug-cc-pVTZ) v. (sTDA-)xTB-calculated $-\text{IP}$, $-\text{EA}$ and $\Delta_o$ for unsubstituted cores. The data prior to calibration is shown in blue and post-calibration in green, the red dashed line is the $x=y$ line and the line of best fit used to calibrate the data is black. .................................................. 48
3.4 (TD-)DFT-calculated (B3LYP/aug-cc-pVTZ/$\omega$B97X) v. (GFN/IPEA/sTDA-)xTB-calculated $-\text{IP}$, $-\text{EA}$ and $\Delta_o$ for substituted cores calibrated using the fitting data from the unsubstituted cores in fig. 3.3. The data prior to calibration is shown in blue and post-calibration in green, the red dashed line is the $x=y$ line. .................................................. 49
3.5 The three outlying quinones with electron-donating substituents in the $\omega$B97X/aug-cc-pVTZ datasets. .................................................. 50
3.6 sTDA-xTB v. TD-DFT calculated optical gaps, using the B3LYP (left) and $\omega$B97X (right) functionals and aug-cc-pVTZ basis set. The pink data points are the quinones with electron-withdrawing substituents, the orange data points are the quinones with electron-donating substituents and the blue data points are the rest of the dataset of substituted cores. The black line is the line of best fit for the entire dataset of substituted cores. .................................................. 50
3.7 DFT-calculated v. xTB-calculated $-\text{IP}$ and $-\text{EA}$ for the ABCBA, ACA’ and ACA dyes, calibrated using the fitting data obtained from the dyes alone solvated in water. The data prior to calibration is in orange and post-calibration is in blue, the red dashed line is the $x=y$ line and the line of best fit used to calibrate the data in black. .................................................. 52
3.8 DFT-calculated v. sTDA-calculated optical gaps for the ABCBA, ACA’ and ACA dyes with sulfur (left) and without sulfur (right), calibrated using the fitting data obtained from the dyes alone. The colour scheme is the same as in fig. 3.7 .................................................. 53
3.9 DFT-calculated v. sTDA-calculated and DFTbaby-calculated optical gaps for the ABCBA, ACA’ and ACA dyes with calibrated using the fitting data obtained from the dyes alone. The colour scheme is the same as in fig. 3.7 .......................... 53
3.10 DFT-calculated v. sTDA-calculated and DFTbaby-calculated optical gaps for the ABCBA, ACA’ and ACA dyes with calibrated using the fitting data obtained from the dyes alone. .................................................. 55
LIST OF FIGURES

4.1 (a) Definition of $-\text{IP}$, $-\text{EA}$, fundamental gap ($\Delta_f$), optical gap ($\Delta_o$) and exciton binding energy (EBE), (b) composition of the different dyes studied in terms of their building block sequence, (c) an example of an ABCBA dye and (d) a selection of the dye building blocks, including the diketopyrrolopyrrole (DPP) core. .................................................. 58

4.2 Two-dimensional histograms of the property spaces spanned by $-\text{IP}$ v. $-\text{EA}$ (top row), $-\text{IP}$ v. $\Delta_o$ (centre row), $-\text{EA}$ v. $\Delta_o$ (bottom row) for the ABCBA, ACA and ACA' dyes (left to right). .................................................. 61

4.3 Convex hulls enclosing the property space of $-\text{IP}$ v. $-\text{EA}$ (left), $-\text{IP}$ v. $\Delta_o$ (centre), $-\text{EA}$ v. $\Delta_o$ (right) for the ABCBA, ACA and ACA' dyes. .................................................. 62

4.4 Property-property relationships between the ABCBA dyes and their constituent building blocks as described by the averaging model (top row) and the max/min model (bottom row) using the building block properties (left four panels) and AACAA/BBCBB dyes (right four panels) as input. Points for the ACA dyes are shown in black. .................................................. 62

4.5 Property-property relationships between the ACA' dyes and their constituent building blocks as described by the averaging model (top row) and the max/min model (bottom row) using the building block properties (left four panels) and ACA dyes (right four panels) as input. Points for the ACA dyes are shown in black. .................................................. 63

4.6 Kernel density plots and histograms for the DFT-calculated properties of the building blocks with the value for diketopyrrolopyrrole (DPP) for each property indicated by a black dashed line. .................................................. 63

4.7 ACA' structures with the smallest fundamental (left) and optical (right) gaps relative to their corresponding ACA and A'CA' dyes. .................................................. 65

4.8 (a) Cumulative histogram of ACA' dyes for which the optical gap (blue) or fundamental gap (orange) is less than that of both related ACA and A'CA' dyes, negative values, or larger than at least one of the related dyes, positive values), (b) zoomed in version of the cumulative histogram, (c) ACA dyes with an fundamental gap lower (yellow) or larger (blue) than their ACA and A'CA' counterparts (yellow) as a function of the offset between the $-\text{IPs}$ and $-\text{EAs}$ of the ACA and A'CA' dyes, (d) Plot of ACA dyes with an optical gap lower (yellow) or larger (blue) than their ACA and A'CA' counterparts as a function of the offset between the $-\text{IPs}$ and $-\text{EAs}$ of the ACA and A'CA' dyes, (e) example of ACA' dyes where the ACA/A'CA' $-\text{IP}$ and $-\text{EA}$ values are staggered, and (f) example of ACA' dyes where the ACA/A'CA' $-\text{IP}$ and $-\text{EA}$ values are straddled. .................................................. 66

4.9 Possible alignments of dye potentials: staggered (left) and straddled (right) . . 67
4.10 Scatter plot showing the $-\text{IP}/-\text{EA}$ slice of property space for the most common A and B units of the ABCBA dyes, when residing on the B-site. The building blocks shown are the most common A and B building blocks in (a) the top corner ($\text{IP} \geq -4.8, \text{EA} \geq -3.3$) (b)-(c) the bottom corner ($\text{IP} \leq -6, \text{EA} \leq -4.3$) and (d)-(f) the middle section of the grid ($\text{IP} = -6, -4.8; \text{EA} = -4.3, -3.3$). The colours of the building blocks correspond to the positions in the scatter plot of the dyes they reside in.

4.11 Most common building blocks on the A and B site within the regions of the $-\text{IP}/-\text{EA}$ slice of property space of the ABCBA dyes discussed in the text.

4.12 Convex hulls for the building block in section (d) in fig. 4.11 (top left), building block i in fig. 4.13(a) and (b) (top right) (blue = occurrence on B site, yellow = occurrence on A site, red = entire ABCBA set) and the two most common building blocks in the proton reduction region for the ACA$^\prime$ dyes (blue = benzene, yellow = 1,6-naphthyridine, green = entire ACA$^\prime$ set). The unshaded regions on the top right and bottom depict the $-\text{IP}$ and $-\text{EA}$ limits where proton reduction is thermodynamically feasible.

4.13 Most common building blocks on (a) the A and (b) B sites for the proton reduction region.

4.14 Most common building blocks on B sites for the ABCBA dyes in the dye-sensitized solar cell region.

4.15 Convex hulls for selected building blocks in the dye-sensitized solar cell region. Building block (i) in fig. 4.14 occurring on the A (yellow) and B (blue) sites of ABCBA dyes in this region (left); most common building blocks in the ACA$^\prime$ non-sulfur containing dyes in the DSSC region (yellow = 9H-carbozole, blue = difuro[3,2-b:2',3'-d]furan) (right). The unshaded regions depict the region of the $-\text{IP}$ v. $-\text{EA}$ property space where charge injection from the dye into the TiO$_2$ and oxidation of I$^- \rightarrow$ I$_3^-$ is possible.

5.1 Workflow followed for high-throughput screening method: (a) Construction of polymer with RDKit and stk, (b) Conformer search with RDKit, (c) Property calculation with (GFN/IPEA/sTDA)-xTB and (d) property calibration against DFT data.

5.2 Calibrated IP, EA and optical gap v. Boltzmann factor for the 500 conformers of the polymers of the repeat units shown on the right. Coloured chemical structures indicate data shown in the same colour. The lowest energy conformers are those with Boltzmann factors close to 1.
5.3 Distributions of properties (IP, EA and optical gap) of disordered (a) phenylene–thiophene, (b) phenylene–pyridine and (c) pyridine–thiophene co-oligomers varying with differing co-monomer ratios. For each ratio, the properties of all possible monomer sequences were calculated, using a fixed oligomer length of 12 monomer units in total. .................................................. 83

5.4 (a) Formula for calculation of degree of segregation and depiction of its variables, 0 being fully segregated and 1 being fully alternating; (b) Variation of properties (IP, EA and optical gap) as a function of D for all possible sequences for 1:1 dodecamer. ........................................... 86

6.1 Calibration curves for $-\text{IP}$ (left), $-\text{EA}$ (centre) and optical gap (left) for a subset of the oligomers, as calculated by (IPEA/sTDA-)xTB and B3LYP/DZP in water. Here the uncalibrated data points (blue), the calibrated points (orange), the line of best fit (black) and the $x=y$ line (red) are shown. ............... 91

6.2 The library of 61 monomers used to construct the oligomers. The dashed lines indicate the site of connection in the (co-)oligomers. .................................................. 93

6.3 Calibrated properties in water v. Boltzmann factor for the 500 conformers of each of the homo-oligomers (dodecamers) composed of monomers 35, 26, 15 and 46. The lowest energy conformers are those with Boltzmann factors close to 1. .................................................. 94

6.4 Calibrated properties in water v. Boltzmann factor for the 500 conformers of each of the co-oligomers (dodecamers) composed of monomers 35, 26, 15 and 46. The colours of the points correspond to the co-oligomers of the monomer pairs of the same colour. The lowest energy conformers are those with Boltzmann factors close to 1. .................................................. 95

6.5 Violin plots describing the effect of both sequence and conformerism on the $-\text{IP}$ and $-\text{EA}$ of both the octamers and the dodecamers of 35, 26, 15 and 46 in water. There are 3 and 4 unique sequences for the octamers and dodecamers, respectively. The pink/blue horizontal lines correspond to the $-\text{IP}$/$-\text{EA}$ of the homo-oligomer of the first monomer in each pair, and the red/green dashed line to the potentials of the homo-oligomer of the second monomer in each pair. 97

6.6 Violin plots describing the effect of both sequence and conformerism on the optical gap for both the octamers and the dodecamers of 35, 26, 15 and 46 in water. There are 3 and 4 unique sequences for the octamers and dodecamers, respectively. .................................................. 98

6.7 2D histograms of the slices of property space spanned by $-\text{IP}$ v. $-\text{EA}$, $-\text{IP}$ v. $\Delta_o$ and $-\text{EA}$ v. $\Delta_o$ (left to right). .................................................. 100
6.8 Convex hulls enclosing the property space of $-\text{IP}$ v. $-\text{EA}$, $-\text{IP}$ v. $\Delta_o$ and $-\text{EA}$ v. $\Delta_o$ (left to right), for the different co-oligomer sequences and homo-oligomers (octamers) in water; pink = homo-oligomers, red = AB, blue = AABB and green = AAAABBBB. 100

6.9 Example of an AABB octamer (top) and an AAAABBBB octamer (bottom) with optical gaps below 1.7 eV. 100

6.10 (a) Scatter plot of the difference in IP between the two constituent homo-oligomers of each block dodecamer v. the difference in EA of the two constituent homo-oligomers in water. The quadrants correspond to block co-oligomers where the potentials of their corresponding homo-oligomers are either straddled or staggered. The yellow points represent the block co-oligomers exhibiting the discussed charge-separated effect and the points in blue represent those that do not; (b) examples of two dodecamers that exhibit intramolecular charge separation; (c) example of staggered and straddled alignments of potentials. 102

6.11 Scatter plot of the $-\text{IP}/-\text{EA}$ property slice showing (a) all octamers, (c) all $(AB)_4$ octamers, (d) all $(AABB)_2$ octamers and (e) all AAAABBBB octamers containing the most common monomers in the different regions described in table 6.2. Each colour represents the oligomers containing the relevant monomer. A selection of the monomers shown on these plots are shown in (b), coloured to match the data points describing them on the scatter plots. 103

6.12 Scatter plot of the $-\text{IP}/\Delta_o$ property slice showing (a) all octamers, (c) all $(AB)_4$ octamers, (d) all $(AABB)_2$ octamers and (e) all AAAABBBB octamers containing the most common monomers in the different regions described in table 6.3. Each colour represents the oligomers containing the relevant monomer. A selection of the monomers shown on these plots are shown in (b), coloured to match the data points describing them on the scatter plots. 104

6.13 Scatter plot of the $-\text{EA}/\Delta_o$ property slice showing (a) all octamers, (c) all $(AB)_4$ octamers, (d) all $(AABB)_2$ octamers and (e) all AAAABBBB octamers containing the most common monomers in the different regions described in table 6.4. Each colour represents the oligomers containing the relevant monomer. A selection of the monomers shown on these plots are shown in (b), coloured to match the data points describing them on the scatter plots. 105

6.14 $-\text{IP}/-\text{EA}$ of homo-oligomer A v. $-\text{IP}/-\text{EA}$ of homo-oligomer B corresponding to each co-oligomer within the electronic constraints and driving force (top row) and within the optical and electronic constraints including a driving force (bottom row), with respect to the acceptor PCBM. 109

6.15 Optical gap v. open circuit voltage for oligomers in combination with PCBM, PC$_{70}$BM and PN1 (left to right) within the electronic (blue), electronic + driving force (orange) and optical, electronic + driving force (green) constraints. 109
LIST OF FIGURES

7.1 −IP/−EA v. one over oligomer length; blue = B3LYP DFT, red = $G_0W_0$, green = B3LYP KS. ........................................... 117
7.2 Fundamental gap v. oligomer length; blue = DFT, red = $G_0W_0$, green = KS. .................................................. 118
7.3 Optical gap v. oligomer length; blue = (TD-)DFT, red = $G_0W_0$/BSE, green = KS. ........................................... 119
7.4 Correlation between the exciton binding energy values calculated with $\Delta$B3LYP/TD-B3LYP and $G_0W_0$/BSE; yellow = pyrazine, blue = pyridine, red = pyrrole, green = phenyl, black = thiophene. .............. 119
7.5 Comparison of the trends in −IP and −EA values of the different materials as a function of one over the oligomer length, as predicted by the different methods; yellow = pyrazine, blue = pyridine, red = pyrrole, green = phenyl, black = thiophene. .................................................. 120
7.6 Fundamental gaps calculated by DFT (top left) and $G_0W_0$ (top right), and optical gaps calculated by TD-DFT (bottom left) and $G_0W_0$/BSE (bottom right); yellow = pyrazine, blue = pyridine, red = pyrrole, green = phenyl, black = thiophene. .................................................. 121
A.1 HOMO (left) and LUMO (right) of 2-dimethylamino-3-amino-1,4-benzoquinone, one of the outlying quinones in fig. 3.6. ................. 129
A.2 HOMO (left) and LUMO (right) of 2,5-bis(dimethylamino)-1,4-benzoquinone, one of the outlying quinones in fig. 3.6. ..................... 130
A.3 HOMO (left) and LUMO (right) of 2-amino-3-hydroxy-1,4-benzoquinone, one of the outlying quinones in fig. 3.6. ........................... 130
A.4 DFT-calculated v. xTB-calculated −IP and −EA for the ABCBA, ACA′ and ACA dyes, calibrated using the fitting data obtained from the dyes alone solvated in benzene. The data prior to calibration is in orange and post-calibration is in blue, the red dashed line is the x=y line and the line of best fit used to calibrate the data in black. ........................................... 130
A.5 DFT-calculated v. sTDA-calculated optical gaps for the ABCBA, ACA′ and ACA dyes with sulfur (left) and without sulfur (right), calibrated using the fitting data obtained from the dyes alone and solvated in benzene. The colour scheme is the same as in fig. A.4 ........................................... 131
A.6 Two-dimensional histograms of the property spaces spanned by −IP v. −EA (top row), −IP v. $\Delta_o$ (centre row), −EA v. $\Delta_o$ (bottom row) for the ABCBA, ACA and ACA′ dyes (left to right). ........................................... 131
A.7 Convex hulls enclosing the property space of −IP v. −EA (left), −IP v. $\Delta_o$ (centre), −EA v. $\Delta_o$ (right) for the ABCBA, ACA and ACA′ dyes. ......................... 133
A.8 Convex hulls of the −IP and −EA of the ABCBA dyes containing the building block in fig. 4.11(a) ........................................... 133
A.9 Convex hulls of the $-\text{IP}$ and $-\text{EA}$ of the ABCBA dyes containing the building block in fig. 4.13(a)(iii) .................................................. 133
A.10 Scatter plot showing the $-\text{IP}$ v. $-\text{EA}$ slice of property space for the most common A and B units of the ABCBA dyes, when residing on the A-site. ... 134
A.11 Calibration curves for $-\text{IP}$ (left), $-\text{EA}$ (centre) and optical gap (left) for a subset of the oligomers, as calculated by (IPEA/sTDA-$x$)TB and B3LYP/DZP in benzene. Here the uncalibrated data points (blue), the calibrated points (yellow), the line of best fit (black) and the $x=y$ line (red) are shown. ........... 136
A.12 Property-property relationships for calibrated $x$TB-calculated properties in benzene; $-\text{IP}$ v. $-\text{EA}$, $-\text{IP}$ v. $\Delta_o$ and $-\text{EA}$ v. $\Delta_o$ (left to right). ............. 136
A.13 Convex hulls for the calibrated properties of the different co-polymer sequences and homopolymers (dodecamers) in water; pink = homopolymers, red = AB, blue = AABB and green = AAAABBBB. ................................. 136
A.14 Convex hulls for the calibrated properties of the different co-polymer sequences and homopolymers (dodecamers) in benzene; pink = homopolymers, red = AB, blue = AABB, yellow = AAAABB and green = AAAAAAAABB. ......... 137
A.15 Convex hulls for the calibrated properties of the different co-polymer sequences and homopolymers (octamers) in benzene; pink = homopolymers, red = AB, blue = AABB, yellow = AAAABB and green = AAAAAAAABBB. ............ 137
A.16 Scatter plots of the difference in IP between the two constituent homopolymers of each block co-polymer v. the difference in EA of the two constituent homopolymers for the octamers in water, the octamers in benzene and the dodecamers in benzene (left to right). The quadrants correspond to block co-polymers where the potentials of their corresponding homopolymers are either straddled or staggered. The yellow points represent the block co-polymers exhibiting the discussed charge-separated effect and the points in blue represent those that do not. ................................................. 137
A.17 $-\text{IP}$/$-\text{EA}$ of homopolymer A v. $-\text{IP}$/$-\text{EA}$ of homopolymer B corresponding to each co-polymer within the electronic constraints and driving force (top row) and within the optical and electronic constraints including a driving force (bottom row), with respect to the acceptor PC$_{70}$BM. ......................... 138
A.18 $-\text{IP}$/$-\text{EA}$ of homopolymer A v. $-\text{IP}$/$-\text{EA}$ of homopolymer B corresponding to each co-polymer within the electronic constraints and driving force (top row) and within the optical and electronic constraints including a driving force (bottom row), with respect to the acceptor PN1. ......................... 138
A.19 $-\text{IP}$/$-\text{EA}$ v. one over oligomer length; blue = BHLYP DFT, red = $G_0W_0$, green = B3LYP KS. ............................... 139
A.20 $-\text{IP}$/$-\text{EA}$ v. one over oligomer length; blue = BLYP DFT, red = $G_0W_0$, green = B3LYP KS. ............................... 140
List of Tables

2.1 Orbital functions included for different basis sets for a carbon atom with electronic configuration $1s^22s^22p^2$ .......................... 35

3.1 The $R^2$ values from the calibration curves of the unsubstituted cores and MAE between the DFT and calibrated (GFN/IPEA/sTDA)-xTB properties of these structures. ................................................................. 47

3.2 The MAEs between the (GFN/IPEA/sTDA)-xTB properties of the substituted cores, calibrated using the fitting parameters from the unsubstituted cores calibration curve, and the DFT calculated properties. ................................. 48

3.3 Comparison of the $R^2$ and MAE values from the substituted cores dataset, fitted using the parameters from the substituted cores, including and excluding the problematic quinone structures. ................................................................. 50

3.4 MAEs between DFT-calculated properties and (GFN/IPEA/sTDA)-xTB properties of the substituted cores excluding the problematic quinone structures, calibrated using the fitting parameters of the unsubstituted cores. .......... 51

3.5 MAEs for the properties following the different calibrations performed, MAEs of the properties using the validation set and the fitting parameters from the different calibrations, and the $R^2$ values of each fit. ................................. 54

4.1 Number of OPV polymer candidates for each co-polymer type relative to the experimentally determined potentials of the three acceptors PCBM, PC$_{70}$BM and PN1. ................................................................. 75

6.1 Fitting parameters and $R^2$ obtained from the linear fit between the (sTDA-)xTB and DFT results in water, and the MAE between the DFT-calculated properties and the calibrated xTB-calculated properties. ................................. 92

6.2 Most commonly found monomers in each of the −IP v. −EA grid sections. 103

6.3 Most commonly found monomers in each of the −IP v. $\Delta_o$ grid sections. 106

6.4 Most commonly found monomers in each of the −EA v. $\Delta_o$ grid sections. 106
6.5 Number of OPV polymer candidates for each co-oligomer sequence relative to the experimentally determined potentials of the three acceptors PCBM, PC$_{70}$BM and PN1. 110

7.1 $\Delta$IP/$\Delta$EA and fundamental gap values for the benzene, pyridine, pyrrole, thiophene and pyrazine monomers with the def2-SVPD basis set and at the CBS limit, as calculated with $G_0W_0$. 115

7.2 Optical gap values for the benzene, pyridine, pyrrole, thiophene and pyrazine monomers with the def2-SVPD basis set and at the (pseudo-)CBS limit. 116

A.1 Effect of phosphonic acid, carboxylic acid and cyanocarboxylic acid anchoring groups on the properties of dyes. SMILES for each dye ID can be found in the supplementary information of the published work. 135
List of Abbreviations

ASCII American Standard Code for Information Interchange.

BODIPY 4,4-difluoro-4-bora-3a,4a-diaza-3-indacene.

BSE Bethe-Salpeter equations.

CB conduction band.

CBS complete basis set.

CI configuration interaction.

CI conical intersection.

COF covalent organic framework.

COSMO COnductor-like Screening MOdel.

CT charge-transfer.

CTF covalent triazine framework.

DFT density functional theory.

DNA deoxyribonucleic acid.

DPP diketopyrrolopyrrole.

DS-PC dye-sensitized photocatalysis.

DSSC dye-sensitized solar cell.

DTS(PTTh2)2 4,4’-|4,4-Bis(2-ethylhexyl)-4H-silolo[3,2-b:4,5-b’]dithiophene-2,6-diyl|bis[7-(5’-hexyl-2,2’-bithiophen-5-yl)-1,2,5-thiadiazolo[3,4-c]pyridine]-5,5’- Bis[4-(7-hexylthiophen-2-yl)thiophen-2-yl]-1,2,5-thiadiazolo[3,4-c]pyridine-3,3’-di-2-ethylhexylsilylene-2,2’-bithiophene.
EA  electron affinity.
EBE exciton binding energy.
EQE external quantum efficiency.
ETKDG Experimental-Torsion Distance Geometry.
FIrpic Bis[2-(4,6-difluorophenyl)pyridinato-C₂,N](picolinato)iridium.
GGA generalized gradient approximation.
GTO Gaussian type orbital.
HK Hohenberg-Kohn.
HOMO highest occupied molecular orbital.
HTS high-throughput screening.
HTVS high-throughput virtual screening.
IP ionization potential.
ISC intersystem crossing.
KS Kohn-Sham.
LDA local density approximation.
LR linear-response.
LR-TD-DFT linear-response time-dependent density functional theory.
LSDA local spin density approximation.
LUMO lowest unoccupied molecular orbital.
MAE mean absolute error.
MBGF many-body Green’s function.
MBPT many-body perturbation theory.
MMFF Merck Molecular Force Field.
NIR near infrared.
**LIST OF ABBREVIATIONS**

**OLED** organic light emitting diode.

**OPV** organic photovoltaic.

**P3HT** poly(3-hexylthiophene-2,5-diyl).

**PBDB-TF** Poly[4,8-bis[5-(2-ethylhexyl)-4-fluoro-2-thienyl]benzo[1,2-b:4,5-b’]dithiophene-2,6-diyl]-2,5-thiophenediyl[5,7-bis(2-ethylhexyl)-4,8-dioxo-4H,8H-benzo[1,2-c:4,5-c’]dithiophene-1,3-diyl]-2,5-thiophenediyl] (same as PM6).

**PCBM** [6,6]-phenyl-C_{61}-butyric.

**PCDTBT** Poly[N-9’-heptadecanyl-2,7-carbazole-alt-5,5-(4’,7’-di-2-thienyl-2’,1’,3’-benzothiadiazole)].

**PCE** power conversion efficiency.

**PCM** polarisable continuum solvation model.

**PES** potential energy surface.

**PIC-TRZ** 2-biphenyl-4,6-bis(12-phenylindolo[2,3-a]carbazol-11-yl)-1,3,5-triazine.

**PM6** Poly[4,8-bis[5-(2-ethylhexyl)-4-fluoro-2-thienyl]benzo[1,2-b:4,5-b’]dithiophene-2,6-diyl]-2,5-thiophenediyl[5,7-bis(2-ethylhexyl)-4,8-dioxo-4H,8H-benzo[1,2-c:4,5-c’]dithiophene-1,3-diyl]-2,5-thiophenediyl] (same as PBDB-TF).

**PN1** Poly(3,9-bis(2-methylene-(3-(1,1-dicyanomethylene)-indenone))-5,5,11,11-tetakis[3-(2-hexyldecyl)oxy]benzene]-dithieno[2,3-d:20,30-d0]-s-indaceno[1,2-b:5,6-b0]dithiophene).

**PPP** poly(p-phenylene).

**PPV** poly-(para-phenylene vinylene).

**PTB7** Poly[4,8-bis[[2-ethylhexyl]oxy]benzo[1,2-b:4,5-b’]dithiophene-2,6-diyl]-3-fluoro-2-[(2-ethylhexyl)carbonyl]thieno[3,4-b]thiophenediyl].

**SCF** self-consistent field.

**SEA** sacrificial electron acceptor.

**SED** sacrificial electron donor.

**SHE** standard hydrogen electrode.

**SIE** self-interaction error.
SMILES simplified molecular-input line entry system.

STO Slater type orbital.

TADF thermally activated delayed fluorescence.

TD-DFT time-dependent density functional theory.

TDSE time-dependent Schrödinger equation.

TEA triethylamine.

TEOA triethanolamine.

TISE time-independent Schrödinger equation.

TPA tris(2-pyridylmethyl)amine.

TTA triplet-triplet annihilation.

UEG uniform electron gas.
List of Publications


Chapter 1

Introduction

Climate change is currently considered by many to be one of the most pertinent crises faced globally. Over the last few years, increased media coverage has brought into the spotlight the damage that climate change has caused, the risks that will ensue in the future and efforts to reverse or limit said damage. As such, this has brought about a greater concern in environmental issues, global heating in particular, which is vastly being attributed to man-made carbon dioxide emissions. The urgent need for the global transition to clean, sustainable energy usage is more relevant than ever. In recent years, enormous efforts and progress have been made in the scientific community towards the discovery and design of materials for clean energy conversion, however, there is still a long way to go.

Between 1950 and 2017, global fossil fuel consumption increased by over 560%, reaching 133,853.83 TWhs in 2017. That year, 93.2% of CO$_2$ emissions were caused by fossil fuel combustion.$^1$ The burning of fossil fuels is not only detrimental to our planet and its inhabitants but is also a finite source of energy. Given the lack of sustainability in our ever-increasing energy demands, it is of paramount importance that fossil fuels are replaced by alternative energy sources.

Solar power is one of the most promising forms of renewable energy due to the abundance of sunlight available. Even after taking into consideration the sunlight lost by reflection, about $1 \times 10^9$ TWhs of energy reach the Earth’s surface and are absorbed by the atmosphere, land, oceans and clouds. This is almost 8000 times the global fossil fuel consumption in 2017. Unfortunately, despite the Sun being an enormous resource, solar power has yet to be universally adopted. Materials for conversion of light energy into chemical energy have attracted a lot of research attention with the aim of making solar power a ubiquitous source of energy.

Photovoltaic power production, the direct conversion of sunlight into electricity through solar cells, has proven to be very successful. Although the highest performers do very well in terms of power conversion efficiency (PCE), they tend to be expensive (containing crystalline silicon or rare metals such as tellurium, indium or gallium)$^{2,3}$ or contain toxic elements such as cadmium.$^3$ The discovery of equally efficient organic polymer photovoltaics is an attractive
prospect as these materials provide the benefits of having easily tuned properties, earth-abundant and non-toxic compositions, mechanical flexibility and solution processability.

Hydrogen as a fuel is also a very attractive prospect for reduction of CO\textsubscript{2} emissions that is seldom used – it is the most abundant element on the planet and simply oxidises to water, emitting no greenhouse gases. While attention on hydrogen fuelled vehicles is growing, such vehicles are not commonplace at this point in time. This is due to the burning of petrol simply being cheaper than any of the processes currently available for hydrogen fuel production.

Hydrogen fuel for cars would provide an emission-free alternative to petrol and diesel, however, the process of obtaining hydrogen fuel is highly energy intensive. Most hydrogen fuel is obtained \textit{via} steam reforming of methane or natural gas, requiring temperatures of around $900^\circ\text{C}$.\textsuperscript{4} It is not only that attaining the high temperatures is expensive; the process of methane reforming produces CO\textsubscript{2} as a by-product, which is counterproductive from an environmental perspective. Other routes to hydrogen fuel production are coal and biomass gasification, and electrolysis, however, methane reforming is the most common.

Water splitting is a clean route to hydrogen production, producing simply hydrogen and oxygen from water molecules. This currently can be done \textit{via} electrolysis. However, despite being the state of the art fossil fuel-free method for hydrogen production, only 4\% of hydrogen produced globally is obtained in this way due to the significantly lower cost of steam reforming.\textsuperscript{4,5} Electrolysis provides the advantage of being entirely free of fossil fuels if powered by electricity from solar or wind energy. An alternative to electrolysis is photocatalytic water splitting (PC-WS); using solar energy to drive the water oxidation process and split water molecules into hydrogen and oxygen at ambient temperatures. This would bring the use of hydrogen fuelled cars much closer to achieving carbon neutrality as the entire process, from extracting the fuel to burning it, would be emission-free. Organic photoactive polymers such as the ones discussed above have also been in focus as photocatalysts in the PC-WS process, providing the same benefits mentioned above.

In this thesis I have computationally investigated different types of organic photoactive materials, mostly for (but not limited to) the application of renewable energy sources. Computational chemistry provides many advantages over experimental chemistry in the early screening stages of materials design. One of these is that computational methods for the calculation of properties provide the benefits of not having to synthesize the materials before analysing them. This is advantageous for a number of reasons; firstly, the lack of complete control over the synthesis of large molecules such as polymers limits the types of isomers that can be studied. It is trivial to build a large variety of well-defined polymers differing only in their monomer sequence on a molecular visualisation program, whereas experimentally, it could prove rather challenging if not impossible. Secondly, conformer isomerism can also be explored to a greater extent computationally as usually, only the lowest energy conformer and a small amount of others can be accessed experimentally. Computations also allow for a large library of materials to be studied at a fraction of the time and effort. Another advantage is that when searching
for new materials for a given application it is easy to go down familiar design routes that are known to lead to the appropriate materials; however, by searching property space with computational chemistry methods, materials that had not been previously considered or discovered for the specific application may be revealed as useful. Finally, there are certain properties that are simply easier to calculate than to measure experimentally. One such example is the electron affinity, which experimentally, is measured by the particularly complex method of inverse photoemission spectroscopy.

1.1 Optical and Electronic Properties of Materials

The determination of a material’s suitability for an application such as photovoltaics or PC-WS requires knowledge on the material’s optical and electronic properties. Optical properties refer to the properties related to the absorption and emission of light, which involve the creation and annihilation of excitons - electrostatically bound electron-hole pairs. Electronic properties involve the physical ability to donate and accept electrons, the separation of excitons into free charge carriers and the properties involved in transporting charge carriers. Often the combination of optical and electronic properties is referred to as the optoelectronic properties, a term used throughout this thesis.

Certain key quantities such as the ones mentioned above need to be known for the analysis of photoactive materials – materials that have the ability to interact with light and through this interaction exhibit interesting physical properties. To determine the electronic properties of a material, knowledge of the ionization potential (IP) and electron affinity (EA) is essential. The IP is defined as the energy required to remove an electron from the highest occupied molecular orbital (HOMO) and is often represented as the negative of the energy of the HOMO. The EA is the energy required to add an electron and is often represented as the negative of the energy of the lowest unoccupied molecular orbital (LUMO). The difference between the IP and the EA is the fundamental gap and is often approximated as the difference between the HOMO and LUMO, in line with Koopman’s theorem (further discussed in section 2.4). Often throughout this thesis the results are analysed in terms of the $-\text{IP}$ and $-\text{EA}$.

To form an understanding of the material’s optical properties the optical gap must be known. This is the first electronic transition that occurs following the absorption of a photon. For an optical transition to occur, the light absorbed must exceed the energy of the optical gap therefore, the optical gap provides information on what wavelengths of light a photoactive material can absorb. The exciton binding energy (EBE) may also be of interest, representing the ease with which an exciton can separate into single charge carriers; this is the energetic difference between the fundamental gap and the optical gap. This property is particularly relevant for organic conjugated polymers. In a silicon solar cell for instance, the very low EBE (around 15 meV) means that excitons need not be considered in the analysis of the electron and hole transport mechanism. In polymers however, the EBE is significantly higher...
(at around 0.5 – 1 eV), necessitating consideration of exciton behaviour. Knowledge of the EBE can reveal whether a material will have suitable properties for the application, such as the ability to separate into free charge carriers in a photovoltaic cell. All of these quantities are depicted in fig. 1.1.

An example in which these properties allow for the characterization of a photoactive material is in the design of an ideal photovoltaic absorber. In photovoltaics, the delicate balance between a good open-circuit voltage and short-circuit current is struck by having a fundamental gap in the appropriate range. Photocatalysts for water splitting require appropriate optical properties for optimal light absorption in addition to an electronic structure in which the IP and EA of the photocatalyst straddle the potentials of the proton reduction and water oxidation reactions in order to reduce protons and oxidise water, producing hydrogen and oxygen gas.

![Energy Vacuum](image)

Figure 1.1: Gaps relevant to the study of photoactive materials: −IP (HOMO), −EA (LUMO), fundamental gap (Δf), exciton binding energy (EBE) and optical gap (Δo).

Although it is possible to determine these properties experimentally, there are significant advantages to calculating them using computational methods. The IP and EA are measured in the laboratory using cyclic voltammetry and (inverse) photoemission spectroscopy. The first disadvantage of measuring these properties experimentally rather than through computation is that carrying out controlled synthesis of the samples is more challenging than computationally modelling them; secondly, the execution of these experiments is rather complex – for inverse photoemission spectroscopy in particular. The optical gap is much easier to measure using UV-Vis spectroscopy however, computational methods still provide the same benefit as above of not requiring prior synthesis of the materials.

There is a huge selection of computational methods to choose from for the calculation of the aforementioned properties depending on computational resources and desired level of accuracy. The IP, EA and fundamental gap can be easily calculated by standard density functional theory (DFT) calculations, as they have been throughout this thesis. All that is
required are the optimized ground state, anionic and cationic energies and the use of the ΔSCF method described in section 2.4. There are also many semi-empirical calculations in this thesis where the IP and EA are calculated using IP/EA-GFN-xTB. An example of a higher accuracy calculation that can be used is the GW method.

For the calculation of optical excitations, ground state DFT is not sufficient and time-dependent DFT (TD-DFT) is used in its place. Higher-accuracy methods such as the Bethe-Salpeter equations (BSE, in conjunction with GW) can also be used to calculate these properties and/or corroborate the TD-DFT results. Semi-empirical methods that are used for these calculations in this thesis are DFTbaby and sTDA-xTB. These are just a few of the many existing computational methods that are able to calculate IP, EA and optical gap.

1.2 Conjugated Molecules and Polymers as Photoactive Materials

1.2.1 Polymer Structure and Properties

In this thesis the materials of interest are mainly large, conjugated molecules and polymers. The focus on conjugated molecules and polymers is due to their ability to absorb light i.e. they are photoactive – making them useful for applications such as photocatalysis and photovoltaics. Conjugation in an organic molecule arises by the interaction of neighbouring π-orbitals in alternating single and double bonds, which form due to the interaction of p-orbitals on adjacent carbon atoms, as shown in fig. 1.2. The interaction of these orbitals causes electrons to move more freely throughout the conjugated π-system resulting in interesting optoelectronic properties. As the number of π-orbitals increases, the spacing between these energy levels decreases and it becomes easier to excite electrons to higher energetic states, which is why conjugated molecules are photoactive. As the conjugated system extends over a greater number of atoms, the optical gap reduces and longer wavelengths of light can be absorbed, leading to an excitation. The materials that have been investigated in this thesis all possess π-conjugation, whether it is in the small aromatic molecules in chapter 3, the dyes in chapter 4 or the conjugated organic polymers in chapter 6, and as a result exhibit interesting photoactivity.

Polymers are materials composed of monomers covalently bonded to form long-chained molecules and can be categorized according to the number of repeating monomers and their sequential ordering. A homopolymer for example, is a polymer composed of a repeating unit based on a single monomer, whereas the term co-polymer refers to a polymer made up of multiple monomers.

Co-polymers can be sub-categorized into different types, for example, statistical co-polymers, alternating co-polymers, periodic co-polymers and block co-polymers, to name a few. A statistical co-polymer is one where the distribution of repeat units is dependent on statistical laws – a specific type of which is a random co-polymer where the repeat units are
CHAPTER 1. INTRODUCTION

1,3,5-hexatriene

Figure 1.2: The p-orbitals on the carbon atoms not involved in σ-bonding in 1,3,5-hexatriene interact to create a series of π-orbitals, which form a conjugated π-system.

distributed in a truly random fashion (fig. 1.3 (bottom)). Alternating co-polymers, as the name suggests, are composed of alternating monomer units e.g. \((AB)_n\), or \((AABB)_n\) if the repeat unit is taken to be AABB (fig. 1.3 (top two)). Periodic co-polymers form repeating sequences of monomers e.g. \((ABAABB)_n\) in a polymer containing monomers A and B (fig. 1.3 (fourth from the top)). Block co-polymers consist of alternating sections made up of non-alternating monomer units (fig. 1.3 (middle)) e.g. \((AAAAABBBBB)_n\). This type of co-polymer is investigated in chapter 6.

Co-polymerisation allows for the modulation of polymer properties, such as the optical and electronic properties that are of interest for photocatalysis and photovoltaics, as the overall properties will be a non-trivial combination of the properties of the starting homopolymers. The overall polymer properties in a statistical, random or alternating co-polymer are usually in between the properties of the constituent homopolymers; block co-polymers on the other hand, tend to exhibit properties more similar to the corresponding homopolymers. In this thesis I focus on the tunability of polymers by probing into the effect of varying compositions, composition ratios, sequence isomerism and conformer isomerism on the overall material properties.

As exemplified by some of the cases discussed below in section 1.2.2, co-polymerisation makes it possible to easily fine-tune the properties of polymers. Changes in the overall polymer properties can be made by varying the composition of the polymer via co-polymerisation, for example, by changing the type of monomer unit or linker between the monomer units in the polymer. Further tuning of the properties can also be achieved through modification of the site of substitution or the number of substituents per monomer unit, leading to changes in planarity or porosity, which ultimately modify the optoelectronic properties. Structural
modifications can also be made in order to modulate the properties of a polymer such as nanostructuring,\textsuperscript{14,15} doping\textsuperscript{16} and by controlling the crystallinity of the material.

Clearly, the tunability of conjugated organic polymers is a very attractive feature, however, aside from tunability, there are many other advantages that organic polymer photocatalysts and photovoltaics provide over their inorganic counterparts. Firstly, is the fact that they are composed of earth-abundant and often, non-toxic elements. The inorganic semiconductors used in photochemical water splitting and solar absorbers, often contain expensive and toxic metals such as gallium, indium and tantalum,\textsuperscript{17} and may often require precious metal co-catalysts such as RuO\textsubscript{2}, IrO\textsubscript{2} or Pt.\textsuperscript{18} Organic photoactive materials are therefore a much more sustainable alternative. Secondly, they require a much more facile manufacturing process, free of extreme temperatures and pressures which drive up costs, due to their solution-processability.\textsuperscript{19,20} Finally, polymers are in general, mechanically flexible materials which makes them easier to incorporate into device architecture.\textsuperscript{13,20}
1.2.2 Applications of Conjugated Organic Materials and Polymers as Photoactive Materials

Conjugated polymers in photocatalytic water splitting

Photocatalytic water splitting is the process of using solar energy to drive the dissociation of water molecules into hydrogen and oxygen for the extraction of hydrogen fuel. This process is inspired by photosynthesis and is therefore, often referred to as artificial photosynthesis. Equation (1.1) is the redox equation resulting from the two half reactions in eq. (1.2) and eq. (1.3).

\[
H_2O \leftrightarrow H_2 + \frac{1}{2}O_2 \quad \text{(1.1)}
\]

\[
2H^+ + 2e^- \rightarrow H_2 \quad \text{(1.2)}
\]

\[
2H_2O \rightarrow O_2 + 4H^+ + 4e^- \quad \text{(1.3)}
\]

The hydrogen reduction half reaction involves two electrons and the water oxidation reaction occurs via a 4 electron process. The electronic requirements for these redox reactions to take place at pH 0 are that the photocatalyst possesses an EA more negative than 0 V and an IP more positive than 1.23 V, with respect to a standard hydrogen electrode (SHE) i.e. the IP and EA of the photocatalyst must straddle the water oxidation and hydrogen reduction potentials. The basic operation mechanism in colloidal photocatalytic water splitting starts with the absorption of incident light by a photocatalyst resulting in the formation of an exciton (a bound electron and hole pair). The exciton then separates into a free electron, which is excited to the conduction band, and a free hole in the valence band where they respectively reduce protons and oxidise water. A schematic of this process is shown in fig. 1.4.

In 1972, Fujishima and Honda made the breakthrough discovery of TiO\(_2\) as a photocatalyst for water splitting in a cell simply consisting of a TiO\(_2\) photoanode and a platinum cathode. The problem with TiO\(_2\) is that due to its band-gap of \(\sim \) 3.2 eV,\(^{21}\) absorption is limited to the narrow UV region and as most of the solar energy available to us is in the form of visible light, it is not a very efficient photocatalyst for this application. Smaller band-gap materials are required in order to achieve visible light absorption and while there are many examples of suitable inorganic semiconductors, it is much more favourable to use organic materials to eliminate the need for expensive, rare and toxic metals. The implementation of polymers for this purpose is particularly desirable as they possess a lot of advantageous features such as solution processability, mechanical flexibility and high tunability.

The first reported photoactivity in conjugated organic polymers was for poly(\(p\)-phenylene) (PPP) in the presence of a sacrificial electron donor (SED).\(^{22,23}\) Much of the research carried out in the field of photocatalytic water splitting uses an SED, as the photocatalysts are often...
only able to catalyse the proton reduction reaction and not the water oxidation reaction. An example of a commonly used SED is triethyamine (TEA). Sacrificial electron acceptors (SEAs) are also used for the study of water oxidation catalysis, an example of which are Ce⁴⁺ salts. Despite the low activity of PPP, limited to the UV light region, it provided a starting point for research into conjugated polymers for photocatalytic applications. In polymeric photocatalytic water splitting, unlike in Fujishima and Honda’s breakthrough paper, the water splitting does not take place in a photoelectrochemical cell, but on the surface of photocatalyst particles instead. This is known as colloidal photocatalysis and the discussion hereafter is focused on this rather than photoelectrochemical cells, unless otherwise stated.

Multiple cases of conjugated polymers exhibiting water splitting were reported since the discovery of PPP’s activity, however, progress plateaued until the discovery of the photocatalytic activity of carbon-nitride in 2009. Carbon-nitride, a polymeric organic material, was found to catalyse hydrogen production from water. Thus far, polymeric photocatalysts required a metal co-catalyst in order to facilitate the transport of charge carriers from the bulk to the surface of the water. More ground-breaking however, was the fact that carbon-nitride was able to catalyse both half-reactions in the water splitting process. These exceptional features (despite the modest levels of activity) made carbon-nitride an attractive material for further investigation.

Plenty of research thereafter focused on increasing the rate of hydrogen evolution and fine-tuning the properties of carbon nitride. One of the main strategies used amongst others, included co-polymerisation to explore the effect of the polymeric structure, which indeed led to increased rate of H₂ evolution. For example, the tunability of covalent triazine frameworks (CTFs), which have a structure based on carbon nitride, was explored modifying the linker between phenyl rings. Polymers similar to CTFs were also developed, where the
triazine rings were connected via disulphide bridges. As alluded to above, numerous modifications can be made to the structure of polymers by co-polymerisation in order to achieve a continuous range of properties.

Covalent organic frameworks (COFs) have also achieved good levels of activity owing to their high crystallinity, porosity and surface area – properties that reduce charge carrier recombination and provide more surface for catalysis to take place on. Their adjustable morphology presents the advantage of highly tunable optoelectronic properties. Photoactive COFs reportedly capable of water splitting include ones based on hydrazones, azines and more recently, sulfones.

Another example of linear or conjugated microporous polymers (CMPs) that evolve hydrogen gas upon absorption of light are fluorene polymers, as reported by Cooper et al. By co-polymerisation of different monomer units and experimenting with different linkers they synthesized both linear polymers and 3-dimensional microporous networks, depending on the positions and number of substitutions. They found that by switching from fluorene to dibenzob,dthiophene sulfone units, the polymers tended towards a more positive IP. Other polymers based on PPP, poly(pyridine) and pyrenes were found to exhibit hydrogen evolution activity. The photocatalysts discussed can be improved by increasing the conjugation within the polymer and therefore boosting the charge carrier separation, which is essential to a good photocatalyst. It was found that enforcing planarity and limiting the torsional angle between repeat units in polyphenylene-based structures to form non-porous fluorene structures led to a higher rate of exciton dissociation and thus formation of free charge carriers, ultimately leading to a higher hydrogen evolution rate (HER).

Dye-sensitized solar cells

Solar cells come in many forms, the two most common being silicon cells and solid-state inorganic cells. Another main family of solar cells in research is dye-sensitized solar cells (DSSCs), differing from the other two in their electron transport mechanism. In DSSCs, the solar absorber is separate from the charge carrier allowing the dye adsorbed onto the TiO$_2$ surface to absorb light of the correct wavelength (that TiO$_2$ would be unable to absorb due to its wide band-gap) and the TiO$_2$ layer to be used to transport electrons. Providing that the HOMO and LUMO of the dye align with the conduction band edge of TiO$_2$ as required, its optical gap can be fine-tuned to absorb a range of wavelengths of light in the visible light and preferably even near-infrared (NIR) regions. Therefore, the dye allows for the absorption of light that would not be possible by TiO$_2$ itself due to its band-gap being too wide (≈ 3.2 eV) – this is the advantage of DSSCs.

The working principle involves the initial absorption of light by a dye with a band gap small enough to absorb visible light. The exciton generated separates and electrons move to the LUMO of the dye. Subsequently, the electrons are injected into the TiO$_2$ layer, which transports the electrons around a circuit, propagating an electrical current. A redox reaction in
the electrolyte, typically I$^-$/I$_3^-$, returns the electrons to the dye ground state. Electronically, it is required that the $-\text{EA}$ of the dye is more positive than the band-edge of TiO$_2$ with respect to vacuum and that the $-\text{IP}$ of the dye is more negative than the redox potential of the electrolyte.

In 1991, Grätzel and O'Regan made the first breakthrough in DSSC technology with a cell composed of a TiO$_2$ nanocrystalline anode and a ruthenium-based dye - the well-known Grätzel cell.$^{37}$ While the PCE of 7.9% was far off the theoretical solar cell efficiency limit of 33.7%,$^9$ this research paved the way for many of the discoveries that followed. Within a few years, PCEs of up to 14% were being attained, albeit with low thermal stability,$^{38}$ however, the developments were promising. Much of the research into increasing the DSSC efficiency is focused on the optimisation of the dye sensitizer, which plays the role of absorbing light and injecting electrons into the photoactive TiO$_2$ layer. It follows that fine-tuning the composition of these dyes to achieve the appropriate photophysical properties is an area of focus for researchers.

Most of the high PCE-yielding dyes contain expensive ruthenium. It is therefore preferable to move towards metal-free photoactive materials such as the conjugated organic molecules mentioned above. Progress has been made in the development of purely organic dyes possessing the key qualities of cheap and non-toxic elemental composition, tunability, solution processability and a high extinction coefficient.$^{38}$ The organic dyes currently being researched are typically made up of a combination of electron-withdrawing and electron-donating groups – the properties of which can be modified by the addition or removal of aromatic rings and other functional groups.

Dyes based on porphyrins,$^{39}$ perylene,$^{40}$ indoline,$^{41,42}$ BODIPY,$^{43}$ oligothiophenes,$^{44}$ coumarin$^{45,46}$ and triphenylamine (TPA)$^{47,48}$ are just a few among the organic dyes being researched. Interestingly, the highest achieved PCE of 14% was with a metal-free dye, a TPA and silyl anchor dye.$^{49}$ Another example of a high efficiency dye is the C219 dye (composed of a TPA donor and hydrophilic cyanoacrylic acid electron-acceptor), achieving 10.1% PCE.$^{50}$

The same principle as in DSSCs can be applied to photocatalytic water splitting too. A dye sensitizer can be used to absorb light of a longer wavelength to the electron-transporting medium where the same process described in fig. 1.4 is followed to dissociate water molecules. The benefit of this, of course, is the fact that organic molecules can be used to absorb the optimum wavelength of energy without the need for inorganic semiconductors that possess the undesirable features discussed earlier. Through the use of a dye sensitizer, TiO$_2$ can be used as the electron transporting material.

**Organic photovoltaics**

Another huge field of research related to such materials is that of organic photovoltaics (OPV). The research of OPV is a rapidly emerging field aiming to find a competitive alternative to current commercial solar cells. Consisting of thin layers of flexible and inexpensive materials,
they possess ideal structural properties for incorporation into device architecture, however, they are yet to attain the high PCEs of silicon and inorganic thin film photovoltaics. In the simplest OPV cell, the architecture consists of a single layer of a small molecule or polymer that absorbs light to form excitons. These excitons migrate to their respective electrodes, generating electricity. The design of new OPV materials is often tailored to the bilayer (donor-acceptor) architecture in which the excitons form at the interface between a donor material and an acceptor material, greatly improving the charge separation in comparison to the single layer cell. As such, research is aimed at the discovery of efficient donor and acceptor materials. The beginning of OPV research was largely based on materials used in OLEDs, namely the donor material poly-(p-phenylene vinylene) (PPV).

The efficiency of OPV materials has been optimised in a number of ways; some examples include increasing the regioselectivity leading to improved $\pi-\pi$-stacking, minimising the band-gap by alternating donor and acceptor moieties within the polymers or molecules and increasing the molecule’s planarity in order to improve the charge carrier mobility. One of the most studied donor materials is poly(3-hexylthiophene-2,5-diyl) (P3HT), which in conjunction with the fullerene-derived acceptor molecule PCBM achieved PCEs nearing 5%. Similar donor materials have been developed with even better efficiencies, for example, PCDTBT with a PCE of 7.5%, PTB7, which was optimised to achieve a PCE of over 9%, and DTS(PTTh2)2 which has a similar structure to the molecules investigated in chapter 4. Fullerenes are often the go-to structure for the acceptor layer. They usually have a solubilising moiety attached to them to improve on their poor solubility in organic solution. Much of the OPV research focuses on alternative acceptor molecules, for instance, a high PCE of over 16% was achieved with PBDB-TF (also referred to as PM6 below) with a non-fullerene, chlorine containing acceptor in 2016. There is a greater interest however, in replacing the small molecule donor layer with a polymeric acceptor, constituting an all-polymer OPV. The motivation for this is the greater absorption of polymers with respect to fullerenes, the tunability of polymers (as discussed previously) and the facilitated solution processability due to greater control over the solution viscosity. The attempts at all-polymer OPVs have proven to be successful with PCEs of up to 10.5% being achieved for the combination of the two thiophene-based polymers PM6 and PN1. Other successes in all-polymer solar cells have been made in recent years with other polymers aiming to overcome the 10% PCE limit. While issues with stability and relatively low efficiency compared to the main competitors remain, continual progress is being made.

Organic light-emitting diodes

One final example of the practical application of conjugated organic materials is organic light-emitting diodes (OLEDs). OLEDs operate under the opposite mechanism to OPVs – a voltage is applied and electrons and holes combine to form an exciton, which then de-excites, emitting light. They are used in full colour display panels in devices such as televisions, mobile phones
and digital cameras, and in energy-saving indoor lighting. The main attractive quality of OLEDs is the potential for flexible displays. OLEDs can be made of small molecules or polymers, with the most basic OLED being a single PPV layer. While OLEDs are in use commercially there is still room for improvement and there is a wide field of research aimed at this. One of the main issues is their limited efficiency. In accordance with spin statistics, the ratio of singlets to triplets formed is 1:3 and since triplet decay is non-radiative (i.e. will not emit light) the OLED efficiency is limited to 25%. A lot of research, therefore, is centred around triplet up-conversion, the conversion of a triplet exciton to a singlet exciton.

There are a few common routes to achieving this. Triplet-triplet annihilation (TTA) involves the combination of two triplet states in order to form a singlet, which can then decay and emit light. Another common solution to the inherent efficiency problem in OLEDs is to introduce a heavy metal element into the OLED material in order to utilise all generated excitons.\(^{61}\) The spin-orbit coupling caused by the heavy metal element allows both the transition from \(T_1\) to \(S_0\) and the intersystem crossing (ISC) from the \(S_1\) to the \(T_1\) state. The metals used in this way are usually \(\text{Au(III)}, \text{Os(II)}, \text{Pt(III)}, \text{Ir(III)}, \text{Re(I)}, \text{Ru(III)}\) and \(\text{Cu(I)}\), the most efficient of these being iridium, which, for instance, in a green-blue OLED device using the iridium complex \(\text{FIrpic}\) can give an external quantum efficiency (EQE) of up to 30.3\%.\(^{62,63}\)

Clearly, these are complexes containing mostly rare and expensive metals and it would be desirable to achieve high efficiencies without them. Copper is indeed cheaper, however, it displays poorer efficiency and stability within OLED devices. Another promising route is thermally activated delayed fluorescence (TADF), which yields a theoretical internal quantum efficiency of 100%. In TADF emitters, the \(T_1\) state can convert to the \(S_1\) state via ISC and then decay to the ground state, \(S_0\). The added benefit of TADF emitters is that no metal is required and all-organic molecules can be used.\(^{62}\) The first example of this was \(\text{PIC-TRZ}\),\(^{64}\) where an EQE of 5.3% was achieved, however, there are now examples with EQE up to 19.3%. Although OLEDs show a lot of promise so far there are many obstacles that are yet to be overcome, one of the most pressing being the low stability of blue OLEDs relative to red and green ones due to high energy excitons degrading the devices. It is evident how metal-free, conjugated organic systems would be beneficial in all of these fields and therefore, organic photoactive materials are a huge area of research when it comes to cheaper and renewable sources of energy.

1.3 Multi-Level Computational Analysis of Materials

While ultimately, materials must be synthesized and characterized experimentally, computational methods for materials discovery can provide a multitude of benefits over practical experimental chemistry at the preliminary screening stage. The number of methods available for the characterization of the properties of chemical systems is huge and the choice of method is always dependent on the balance between computational resources and the desired level of
At one end of the spectrum are the ‘high-level’ methods – methods with a high level of theory and higher accuracy results that are more computationally expensive. Such methods include DFT, wavefunction-based methods such as coupled-cluster theory and many-body perturbation theory-based methods such as the GW method and Bethe-Salpeter. These methods solve for the electronic structures of the systems of interest, describing in great detail the underlying quantum chemistry and physics, and requiring plentiful computational resources. Aside from the computational cost, such calculations require a larger timescale for completion and are limiting in terms of the system size and the number of molecules that can be analysed.

On the other hand, there are the ‘low-level’ methods. These are low-cost methods that produce results with lower accuracy by using a greater number of approximations in the theory. Such methods include basic force-fields, which rely on classical mechanics, and semi-empirical methods, which use greatly simplified approaches in the description of electronic structures. The advantage of opting for a method as such is the increased speed in results output and the sheer volume of results that can be obtained at a fraction of the time that would be taken by the higher-level methods.

Both approaches clearly have their benefits and can complement each other to form detailed analyses in research. In this thesis, I have performed a multi-level analysis on organic materials, firstly, to benchmark the methods that are so often used when researching the chemical systems that I have focused on and secondly, to give a well-rounded idea of properties of the materials of interest.

1.3.1 High-Throughput Virtual Screening

High-throughput screening (HTS) is an experimental method for the screening of large libraries of chemical compounds, used predominantly in drug discovery. Using automation and robotics, large volumes of chemical assays can be carried out in order to identify compounds suitable for given applications. HTS has also been used for the study of photoactive polymers, however, the process of producing the data obtained by HTS can be greatly simplified by the use of high-throughput virtual screening (HTVS) as it excludes the need for laborious and expensive materials synthesis. While highly accurate synthesis can be achieved by automation, the issue of the lack of absolute control over conformer, sequence and geometric isomerism can persist. As such, using HTVS provides a route to the study of a much larger dataset including all the different types of isomers mentioned. Once the search has been narrowed down by HTVS, select materials can be synthesized for experimental analysis.

The HTVS method used extensively throughout this thesis involves the creation of a molecule library and the use of semi-empirical methods to calculate their properties. These properties are then calibrated to data obtained by high-level methods, namely (TD-)DFT. Often, the libraries are simply made up of monomers that are combined in different permutations to yield complex sequences and compositions. In some cases different conformers are
screened for – a task that experimentally is not possible due to the inability to force a molecule into the desired conformation. Usually, a mixture of conformers are formed in experiment, a large proportion of which will likely be the lowest energy conformer and it is, therefore, not possible to observe as many conformational changes as it is via computational methods.

Overall, libraries of molecules on the order of $10^4$ can be calculated using the HTVS methods. As mentioned above, the high-level and low-level methods can complement one another to obtain a detailed analysis of a family of materials. For instance, one can use HTVS to identify interesting candidates that can then be further investigated to a high degree of accuracy, avoiding situations in which long and expensive calculations provide results with no relevance to the target application.
Chapter 2

Background and Theory

2.1 Quantum Chemistry methods

2.1.1 The Schrödinger Equation

In order to computationally model a chemical system, including all of its electronic properties, a quantum chemical description rather than a classical description must be used. All quantum chemistry methods are attempts at solving the time-independent Schrödinger equation (TISE):

\[ H\psi = E\psi \]  

(2.1)

where \( H \) is the Hamiltonian operator:

\[ \hat{H} = \hat{T}_N + \hat{T}_e + \hat{V}_{NN} + \hat{V}_{eN} + \hat{V}_{ee} \]  

(2.2)

in which \( \hat{T}_N \) is the nuclear kinetic energy operator, \( \hat{T}_e \) is the electronic kinetic energy operator, \( \hat{V}_{NN} \) is the internuclear potential energy operator, \( \hat{V}_{eN} \) is electron-nucleus interaction operator and \( \hat{V}_{ee} \) is the electronic repulsion operator. Considering all of these terms when attempting to solve the Schrödinger equation becomes increasingly difficult as the system size increases, and as such, approximations must be made.

2.1.2 The Born-Oppenheimer Approximation

At the core of computational chemistry is the Born-Oppenheimer approximation. Within this approximation the nuclear and electronic parts of the Hamiltonian are separated, making it easier to mathematically manage the many-body problem. The underlying principle is that since the nuclei have a much lower velocity than the electrons due to their much greater mass, they can be assumed to maintain fixed positions with respect to the electrons moving in the field of their potential. The consequence of this, therefore, is that in the Hamiltonian of the
system the nuclear kinetic energy can be ignored and the internuclear interaction treated as an additive constant, a non-varying external potential.

The resulting electronic Hamiltonian is as follows:

\[ \hat{H}_e = \hat{T}_e + \hat{V}_{eN} + \hat{V}_{ee} \]  

(2.3)

\[ \hat{H}_e = \sum_{i=1}^{N} \frac{1}{2} \nabla_i^2 - \sum_{A=1}^{M} \sum_{i=1}^{N} \frac{Z_A}{r_{iA}} + \sum_{i>j}^{N} \frac{1}{r_{ij}} \]  

(2.4)

\[ \nabla^2 = \left\{ \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \right\} \]  

(2.5)

The wavefunction resulting from the electronic Hamiltonian depends explicitly on the electronic coordinates, but parametrically on the nuclear coordinates. This means that although the nuclear coordinates do not appear explicitly in the expression for the electronic wavefunction, the wavefunction does depend on the positions of the nuclei and will differ as the nuclear positions vary.\(^6\)

As stated above, the total energy of a system with fixed nuclei can therefore be calculated as the electronic energy plus the constant nuclear repulsion term. By performing this calculation for the electronic energy at different nuclear coordinates we can obtain an averaged potential for the electronic states on which the nuclei can move. This is known as a potential energy surface.

2.1.3 Potential Energy Surfaces

A chemical system, such as a molecule or a pair of molecules, can exist in many different geometries simply by differing in its bond lengths, bond angles and positions of molecules with respect to others. Different combinations of these parameters lead to energy changes and in most cases, chemical systems will exist in their most stable form i.e. their lowest energy geometry. Therefore, the experimentally studied systems will usually be in their most stable conformation. Although the experimental conditions can be chosen in order to obtain alternative geometric configurations, it is not possible to observe every single variation and the corresponding energies, obscuring a lot of information about the chemical system.

The Born-Oppenheimer approximation makes it possible to explore the so-called energy landscape by calculating the mean-field electronic energy over a number of fixed nuclear arrangements. The energy landscape or potential energy surface (PES) is a 3-dimensional contour of the potential energy of a chemical system as a function of atomic parameters such as bond length, bond angle or a combination of these. Obtaining such a landscape can reveal several points of interest.

One of these is the local minima in the PES. These are points where the gradient is equal to zero and the second derivative equal to +1. Local maxima are also stationary points (\( \nabla = 0 \))
however, in this case the second derivative is equal to $-1$. As a molecule changes its conformation it will end up in local maxima and minima depending on how energetically favourable the particular conformer is. Prior to any DFT calculations on polymers in this thesis, a conformer search using a semi-empirical method was performed. Whilst conformer isomerism may not matter so much for more rigid molecules, in large, flexible organic molecules, such as the organic polymers focused on in this thesis, it is instructive to perform a conformer search. Usually, one is interested only in the lowest energy conformer however, in some cases the investigation of higher energy conformers may be informative.

Two other notable regions in the PES are saddle points and conical intersections. Saddle points are relevant as they usually correspond to a transition state along a reaction pathway. They are stationary points that are defined by being a maximum in one direction and a minimum in the other. The diagram in fig. 2.1 is of a 1D energy profile and although the point $C$ is labelled as a saddle point, if this were a real 1D energy profile one would not be able to come to this conclusion without further information about the gradient of the surface in the other directions. Conical intersections (CIs) lie between multiple surfaces. These are points of degeneracy i.e. the same energy level in two different electronic states.$^{67}$

In other words, CIs are the line of the intersection of two potential energy surfaces, which looks like a double cone from a 1-dimensional representation exemplified in fig. 2.1 as $F$. At a CI the Born-Oppenheimer approximation (see section 2.1.2) breaks down and the nuclear-electronic potential energy becomes relevant. The surprising stability of DNA base pairs to UV exposure can be explained by a CI in the excited states and ground state of the base pairs. The existence of a CI between the excited state accessed upon UV absorption and a
dark charge-transfer state means that the photoexcited system remains in the excited state for an extremely short period of time before reverting back to the ground state via a CI between the charge-transfer state and the ground state.\textsuperscript{68}

Finally, there is the global minimum. It has the same mathematical properties as the local minima but is the lowest energy point in the whole PES of the chemical system. As fig. 2.1 shows the 1-dimensional PES, the maxima and minima cannot be labelled as ‘global’ since we do not know the gradient of the peaks or troughs in other directions. When a global optimisation is performed, the aim is to relax the system until the global minimum is reached. Examples of such calculations are most of the $\Delta$SCF calculations in the following chapters where the IPs and EAs are calculated by optimising the neutral, anionic and cationic species of the molecules. If the energy of a molecule at a fixed given geometry is needed, a single point calculation on that geometry can be performed instead.

In the study of excited states, knowledge of the excited state PES is useful. When a vertical excitation occurs from the minimum of the ground state PES to a higher energy state, the geometry may not (and probably will not) necessarily be the minimum on the excited state surface. This is depicted in fig. 2.1. The vertical excitation energy can be calculated by TD-DFT single point calculations. Often it is desirable to know what happens when an excitation occurs and the excited state is allowed to relax i.e. when an adiabatic excitation occurs. This is what is being calculated when an excited state optimisation is performed. Optimisations are much more computationally demanding than single point calculations in both the ground and excited state.

2.1.4 Hartree-Fock Methods

The Independent-Particle Model

The Hartree model was the original attempt at solving the Schrödinger equation for a polyatomic system simplified through the use of a one electron Hamiltonian to describe single electron orbitals called spin orbitals.

\begin{equation}
    h(i) = -\frac{1}{2} \nabla_i^2 - \sum_A Z_A \frac{1}{r_{iA}} \tag{2.6}
\end{equation}

The solution to the independent-particle model is a series of spin orbitals that satisfy the following eigenequation:

\begin{equation}
    h(i)\chi_j(x_i) = \epsilon_j \chi_j(x_i) \tag{2.7}
\end{equation}

This model, however, is physically unrealistic as the use of singly occupied spin orbitals ignores electron repulsion, failing, therefore, to produce accurate descriptions of complex systems. An improvement on this was the Hartree self-consistent field (SCF) method where the electrons are taken to be moving in the electrostatic field of the nucleus and neighbouring
2.1. QUANTUM CHEMISTRY METHODS

Electrons. An averaged electron-electron interaction is therefore used in the corresponding Hamiltonian. The Hartree self-consistent field Schrödinger equation for two or more electron systems is:

\[
\left( -\frac{1}{2}\nabla_i^2 - \sum_A^M Z_A \frac{1}{r_{iA}} + \int \left[ \frac{\chi_j^*(x_j)}{r_{ij}} \right] \right) \chi_i(x_i) = \epsilon_i \chi_i(x_i) \tag{2.8}
\]

An initial guess at the orbitals is made and the electron-electron interaction calculated from this. This produces a new set of orbitals, which in turn are used to recalculate the electron-electron interaction. By iterating until the electron-electron interactions and orbitals converge using the variational method under the constraints of orthonormality, the lowest energy solution to the Schrödinger equation is obtained within the approximation.\(^{69}\)

Even though the electron-electron interaction has now been accounted for in the Hartree SCF method, the form of the interaction term is inherently flawed as it fails to incorporate the Pauli exclusion principle. In the Pauli exclusion principle, interchanging the orbitals in which electron resides should reverse the sign of the wavefunction, however, this is not reproduced in the Hartree wavefunction.\(^{69}\)

**Hartree-Fock method**

As stated above, the true wavefunction of a chemical system must be antisymmetric with respect to the interchange of the spin and position of electrons in their orbitals and the absence of this property in the Hartree SCF method had to be rectified. By replacing the wavefunction in its product form by a single Slater determinant and a normalisation constant, Fock incorporated the Pauli exclusion principle and the exchange energy into the expression for the total energy.\(^{70}\)

The overall wavefunction is thus written as:

\[
\Psi(x; R) = \frac{1}{\sqrt{N_e!}} |\chi_1(1)\chi_2(2)\ldots\chi_N(N_e)| \tag{2.9}
\]

Such that:

\[
\Psi(x_1, x_2) = -\Psi(x_2, x_1) \tag{2.10}
\]

Expansion of the Slater determinants returns two important terms, the exchange integral \(K\), which describes the repulsion between electrons of like spin and the Coulomb integral \(J\), which describes the interaction between each electron and the cloud of neighbouring electrons.

\[
\hat{J}_{ij} = \int \int [\chi_i^*(x_1)\chi_j^*(x_2)] \frac{1}{r_{ij}} [\chi_i(x_1)\chi_j(x_2)] \partial r_1 \partial r_2 \tag{2.11}
\]

\[
\hat{K}_{ij} = \int \int [\chi_i^*(x_1)\chi_j^*(x_2)] \frac{1}{r_{ij}} [\chi_i(x_2)\chi_j(x_1)] \partial r_1 \partial r_2 \tag{2.12}
\]
These two terms \( J \) and \( K \) are included in the so-called Fock operator, which operates on the closed-shell orbital wavefunctions and returns the atomic orbital energies. The Fock operator has the form:

\[
f_1 = h_1 + \sum_{ij} \{2 \hat{J}_{ij}(1) - \hat{K}_{ij}(1)\}
\]

Due to the complexity of the equations, it is difficult to solve them analytically and so this is done by the linear variational method where each atomic orbital is formed by a linear combination of basis functions. These basis functions may be Slater functions (functional forms with the correct exponential decay of the electron density and a cusp at the nucleus) or other functional forms such as Gaussians – the use of which is further discussed in section 2.3. Addition of multiple Slater functions describing the excited states of the system and virtual orbitals with a complete basis set would eventually converge to the exact solution, however, this would be computationally intractable in practice for large systems. Correlated wavefunction methods such as configuration interaction (CI) and coupled-cluster theory use this approach of adding more determinants and do produce results with higher accuracy, however, they come at a much higher computational cost.

Within the HF formalism the expression for the exchange energy is formally exact, however, an important component of the electron-electron interaction – the electron correlation energy – is neglected. Electron correlation energy is the effect of electrons being repelled under the influence of other electrons; it acts to decrease the Coulombic interaction between electrons and thus lowers the total energy. Consequently, energies predicted by HF are too high. While the absence of correlation energy in the HF formalism is not so problematic for calculating conformational changes, it proves to be detrimental when calculating reactions involving the making and breaking of bonds. Therefore, higher-level theory methods are required for such calculations.

### 2.2 Density Functional Theory

Aside from the wavefunction-based methods already discussed, one of the most popular methods used in quantum chemistry is density functional theory (DFT). DFT tackles the many-body problem via an alternative approach in which the energy of a system is described in terms of the electron density, the inner product of the wavefunction with itself, rather than the wavefunction. The electron density describes the probability amplitude of finding an electron at a given point in space and contains all the relevant information in order to solve the Schrödinger equation. For an N-particle system it takes the form:

\[
\rho(r) = N \int \Psi^*(r, r_2, ..., r_N)\Psi(r, r_2, ..., r_N)dr_2...dr_N
\]

Integration over the density yields the number of electrons and at the cusps at the maxima of the density are the nuclear positions. Additionally, the slope of the decay of the density...
2.2. DENSITY FUNCTIONAL THEORY

away from the nuclei reveals information on the atomic number, and therefore, the type of atom. Knowing the positions and types of nuclei and number of electrons provides all the information needed to form the Hamiltonian of the system and subsequently derive the overall energy. The aim, however, is to step away from wavefunction theory and to be able to use the density alone to describe the electronic system; this is what DFT does. Manipulation of the electron density is a simpler task, as it is a 3-dimensional object, as opposed to the 3N-dimensional wavefunction. This should in theory make DFT a less computationally expensive option in comparison to wavefunction-based methods such as Hartree-Fock, however, in the practical application of DFT orbitals are re-introduced, and therefore, the number of variables increases to 3N again. This is discussed in further detail in 2.2.3.

2.2.1 Early Density Functional Theory

Thomas-Fermi DFT

The earliest practical attempts at DFT were by Thomas and Fermi in 1927. In the early approximations the nuclear-electron attraction and electron-electron repulsion were both treated classically. The kinetic energy term was based on a uniform electron gas (UEG) – a constant non-zero density, unlike the true electron density of a many-body system. Thomas and Fermi attempted to minimise the energy with respect to the density by assuming that a variational principle applies.

The energy of the Thomas-Fermi system is given by:

\[ E = C_{\text{kin}} \int [\rho(r)]^{5/3} dr + \int \rho(r)V_{\text{ext}}(r)dr + \frac{1}{2}e^2 \int \frac{\rho(r)\rho(r')}{|r-r'|} drdr' \]  

(2.15)

These equations, however, are of no practical use as they predict the total dissociation of the system into atoms.

Thomas-Fermi-Dirac DFT

It turns out that the treatment of the interelectronic repulsion as a classical repulsion is intrinsically flawed, resulting in large errors in the energetics. Firstly, a self-repulsion error occurs for a one-electron system and secondly, the Pauli exclusion principle is not accounted for as repulsion of an electron occurs regardless of spin. As discussed in section 2.1.4, this is known as the exchange energy. Dirac introduced an exchange term in 1930:

\[ E_x[\rho(r)] = -\frac{9\alpha}{8} \left( \frac{3}{\pi} \right)^{1/3} \int \rho^{4/3}(r) dr \]  

(2.16)

Despite this modification, Thomas-Fermi-Dirac DFT still did not perform well. The insufficiently accurate kinetic energy term played a large role in the poor energetics and lack of bonding.
2.2.2 Hohenberg and Kohn Theory

In 1964, Hohenberg and Kohn proved the existence of an external potential that uniquely describes the ground state density of a many-electron system. This proof is central to modern DFT as it states that the electron density of the system alone is required in order to derive its Hamiltonian and therefore solve the corresponding Schrödinger equation. The total Hohenberg-Kohn (HK) energy is given by:

$$E[\rho(r)] = F[\rho(r)] + \int \rho(r)v_{ext}(r)dr$$

(2.17)

where $F[\rho(r)]$ is the HK functional and $v_{ext}$ the external potential.

Its Hamiltonian can be written as:

$$\hat{H} = \hat{F} + \hat{V}_{ext}$$

(2.18)

The proof states that two different external potentials $v_a$ and $v_b$ cannot have the same ground state wavefunction and therefore density. Calculation of the expectation value of $\hat{H}_a$ over $\Psi_b$ has to result in an energy greater than the ground state, according to the variational principle:

$$E_{0,a} < \langle \Psi_b | \hat{H}_a - \hat{H}_b + \hat{H}_b | \Psi_b \rangle$$

(2.19)

$\hat{H}_b$ can be added and subtracted to give eq. (2.20), which can then be rearranged to eq. (2.21) and then simplified to eq. (2.22) as follows:

$$E_{0,a} < \langle \Psi_b | \hat{H}_a - \hat{H}_b + \hat{H}_b | \Psi_b \rangle$$

(2.20)

$$E_{0,a} < \langle \Psi_b | \hat{H}_a - \hat{H}_b | \Psi_b \rangle + \langle \Psi_b | \hat{H}_b | \Psi_b \rangle$$

(2.21)

$$E_{0,a} < \langle \Psi_b | \hat{H}_a - \hat{H}_b | \Psi_b \rangle + E_{0,b}$$

(2.22)

This final equation may be written as:

$$E_{0,a} < \int [v_a(r) - v_b(r)]\rho_0(r)dr + E_{0,b}$$

(2.23)

This also follows for $v_b$ therefore the following is also true:

$$E_{0,b} < \int [v_b(r) - v_a(r)]\rho_0(r)dr + E_{0,a}$$

(2.24)

Adding eq. (2.23) and eq. (2.24) together yields the impossible result:
2.2. DENSITY FUNCTIONAL THEORY

\[ E_{0,a} + E_{0,b} < \int [v_a(r) - v_b(r)]\rho_0(r)dr + E_{0,b} + \int [v_b(r) - v_a(r)]\rho_0(r)dr + E_{0,a} \quad (2.25) \]

\[ E_{0,a} + E_{0,b} < E_{0,b} + E_{0,a} \quad (2.26) \]

This proves that two different external potentials must have different wavefunctions and therefore electron densities associated with them i.e. there is one-to-one mapping between the external potential and the electron density. The second part of the HK theorem is that a variational principle applies, meaning that the functional will return the lowest possible energy if and only if the true ground state density has been found.

While the HK functional is formally exact, the exact form of the universal functional is unknown – if it were, the Schrödinger equation could be solved exactly. While this proof was ground-breaking in the field of DFT, no method for mathematically approximating the universal functional was proposed until Kohn and Sham a year later.\(^73\)

2.2.3 Kohn-Sham Theory

The advent of Kohn-Sham theory was a major breakthrough in the practical implementation of DFT. In 1965, Kohn and Sham proved that a non-interacting system of electrons with the same ground state density as the real, fully-interacting system can be used to solve for the energy of the system.

Kohn-Sham (KS) theory makes use of one-electron molecular orbitals (spin orbitals) to model a fictitious, non-interacting electronic system. Whilst using an orbital-free model, as in Thomas-Fermi/Thomas-Fermi-Dirac DFT, provides the benefit of only using 3 variables, in practice KS-DFT is used, which reintroduces 3N variables instead. Although this increases the complexity of the problem, KS theory greatly improves the description of the kinetic energy and provides a route to the numerical solution of the HK energy.\(^74\)

The basic expression for the KS equations is:

\[ E_{DFT}[\rho(r)] = T_s[\rho(r)] + V_{Ne}[\rho(r)] + J[\rho(r)] + E_{xc}[\rho] \quad (2.27) \]

The kinetic energy of a non-interacting system of electrons can easily be calculated as the product of the orbital densities, using Slater determinants (hence the subscript \(s\)). The nuclear-electronic interaction is treated as a classical Coulomb interaction between the nuclei and electrons. Part of the electron-electron interaction \((J[\rho(r)])\) can be simply expressed as a Coulomb interaction of the electron density with itself. The final term is the exchange-correlation potential \((V_{xc}[\rho])\), which is a functional of the exchange-correlation energy. This term rectifies the errors associated with using a non-interacting density in lieu of the real system and can be expressed as follows:
\[ E_{xc} = (T[\rho] - T_s[\rho]) + (V_{\text{ee}}[\rho(r)] - J[\rho(r)]) \] (2.28)

Firstly, a correction to the kinetic energy is required, since the repulsion in a fully interacting system would affect the movement of the electrons. Secondly, the \( V_{\text{ee}}[\rho] \) contains the correction to the electron-electron interaction. The classical Coulomb term means that every electron is interacting with itself, as well as the other electrons. The second part of the \( V_{\text{xc}}[\rho] \) therefore, includes effects of exchange and correlation in order to eliminate this error.

The one electron KS equations in greater detail are shown below:

\[
E[\rho(r)] = \sum_i^N \left( \langle \chi_i | - \frac{1}{2} \nabla_i^2 | \chi_i \rangle - \langle \chi_i | \sum_k^{\text{nuclei}} \frac{Z_k}{|r_i - r_k|} | \chi_i \rangle \right) + \sum_i^N \langle \chi_i | - \frac{1}{2} \int \frac{\rho(r')}{|r_i - r'|} dr' \langle \chi_i | \right) + E_{xc}[\rho(r)] \] (2.29)

Where \( N \) is the number of electrons in the system and the density is taken to be the product of orbital densities:

\[
\rho = \sum_{i=1}^N \langle \chi_i | \chi_i \rangle \] (2.30)

In practice, therefore, these equations are solved self-consistently, computing the orbitals and density until the energy of the system is minimised. The difficulty in solving these equations is in the description of the exchange-correlation functional. In HF theory, the exchange energy is calculated exactly but the correlation energy is absent. The KS energy, on the other hand, is formally exact as it contains all the essential information for the solution of the many-body problem, however, the exchange-correlation energy must now be approximated. By knowing the exact form of the exchange-correlation functional it is theoretically possible to find the exact solution for the energy of the system. Unfortunately, the exact form of this functional is likely highly complicated and unknown, therefore, approximations to it must be made in order to make use of the KS equations.

### 2.2.4 The Exchange-Correlation Functional

Although the use of DFT has revolutionised the field of computational chemistry, it does experience some severe pitfalls as well. These pitfalls can be traced back to the approximation of the exchange-correlation functional and cause a range of issues such as underestimation of the band-gap and over-delocalisation of electronic systems, to name a few. Modelling the complex exchange-correlation functional is arguably the greatest challenge faced when using DFT and many attempts have been made to tackle it. There is a large variety of functionals,
2.2. DENSITY FUNCTIONAL THEORY

differing in how and whether they are parameterised, their description of the exchange energy and other features, making different functionals appropriate for different chemical systems.

The local density approximation (LDA)\textsuperscript{76} was the first and simplest attempt at modelling the exchange-correlation functional. LDA and its spin-polarised analogue, local spin density approximation (LSDA)\textsuperscript{77} for open shell systems, work under the assumption that the system is homogeneous, meaning that the electron density varies uniformly away from the nucleus. This is a physically unrealistic property for most systems of interest, resulting in the failure to accurately describe systems with abrupt changes in their electron density. In L(S)DA the exchange energy is given by the Dirac formula for a UEG, and the correlation energy is fitted to accurate quantum Monte Carlo data for a UEG.\textsuperscript{76} While the use of LSDA is now almost always favoured over LDA, they are essentially the same thing for closed-shell systems. The equation for L(S)DA is shown below:

\[ E_{xc}^{\text{L(S)DA}}[\rho] = \int \rho(r) \epsilon_{xc}(\rho(r)) dr \tag{2.31} \]

where \( \rho \) is the electron density and \( \epsilon_{xc} \) is the exchange-correlation energy per particle of a homogeneous electron gas.

LSDA exchange energy is given by:

\[ E_x^{\text{LSDA}}[\rho] = -\frac{2}{3} C_x \int (\rho_a^{4/3} + \rho_b^{4/3}) dr \tag{2.32} \]

Indeed, LSDA is an exact method when dealing with a UEG, however, in the case of molecular systems, large underestimations of the exchange energy and overestimations of the correlation energy are encountered. While LSDA appears to perform rather well overall, it is actually this cancelling out of errors that leads to apparently accurate results.

Having evolved from the LSDA method is the improved generalised gradient approximation (GGA),\textsuperscript{78} which includes some terms from the Taylor expansion of the density in attempt to describe the inhomogeneity of the system. Here not only the density, but also the gradient of the density is taken into consideration (which, as mentioned in section 2.2, provides information on the type of atom), often in the form of a corrective term added to an LDA functional.\textsuperscript{78} This correction does in fact yield better looking results to describe atoms, molecules and materials. An obvious flaw in both the LSDA and GGA methods, however, is that they are based on localised models of the exchange-correlation holes while the true density is non-local. The exchange-correlation interaction ends up decaying exponentially rather than exhibiting the correct asymptotic behaviour. The poor description of the functional means that these methods suffer from severe self-interaction error, one of the biggest problems faced in DFT.

Self-interaction error (SIE) is the spurious interaction of a single electron with itself. The absence of exact exchange within the exchange-correlation functionals such as L(S)DA and GGA is the cause of this. In HF theory the exact expression for the exchange energy exactly cancels out the term for the Coulomb interaction. If the exact exchange-correlation functional
was used the SIE would disappear entirely. There are a few negative repercussions of this error, the most relevant to the work in this thesis being the so-called delocalisation error.\textsuperscript{79}

Delocalisation error occurs as a result of many-electron self-repulsion. Due to self-repulsion, DFT has a tendency to erroneously predict lower energies for fractional charges and higher energies at integer numbers of electrons, leading to more delocalised electron densities.\textsuperscript{75,80,81} As such, certain systems with localised electron densities fail to be properly described; these include Rydberg states, anionic species and in the case of time-dependent DFT in particular, charge-transfer (CT) excitations. Conversely to DFT, HF often experiences the opposite problem of over-localisation due to the lack of correlation energy in the formalism, energetically favouring integer numbers of electrons. Often these issues in DFT can be worked around by the use of hybrid functionals – functionals containing varying proportions of HF exchange. A commonly used hybrid functional is B3LYP,\textsuperscript{77,82–84} which is based on functionals by Becke, Lee, Yang and has been used extensively throughout the work in this thesis. It takes the form:

\[
E_{xc}^{B3LYP} = E_x^{LDA} + \alpha_0 (E_x^{HF} - E_x^{LDA}) + \alpha_x (E_x^{GGA} - E_x^{LDA}) + E_x^{LDA} + \alpha_c (E_x^{GGA} - E_x^{LDA}) \tag{2.33}
\]

Hybrid functionals contain mixed but fixed proportions of HF exact exchange and DFT exchange (e.g. some amount of LDA or GGA exchange). B3LYP contains LDA exchange in addition to a small fraction of exact HF exchange with an LDA correction, GGA exchange with an LDA correction, LDA correlation energy, and a fraction of GGA correlation energy with an LDA correction. As discussed, the exchange-correlation potential should exhibit the correct $1/r$ dependence. Global hybrids like B3LYP and PBE will have an $-\alpha/r$ decay, where $\alpha$ is the fraction of exact exchange (e.g. $x$ in B3LYP). This means that at large internuclear separation, the electron-electron interaction is poorly described. Although B3LYP is widely used and performs well for many systems, it does still suffer from over-delocalisation issues at times. In the case of molecules where charge-transfer excitations may occur it is often necessary to opt for a range-separated hybrid functional.

Range-separated hybrid functionals can be used to circumvent this issue. In these functionals, the electron-electron interaction is partitioned by some smooth separation function, such as an exponential function, in the form of a complementary error function.\textsuperscript{85}

\[
\frac{1}{r} = \frac{\xi \omega(r)}{r} - \frac{1 - \xi \omega(r)}{r} \tag{2.34}
\]

On the left hand side of eq. (2.34) a DFT expression is used as it allows for good exchange-correlation error cancellation at short internuclear distances. At long distances, the right hand side of the equation, which uses exact HF exchange, dominates. Charge-transfer states are inherently long-range, and as such, range-separated hybrid functionals are useful in the description of systems where charge-transfer states are prevalent. In chapters 3 and 4, the $\omega$B97X functional was employed, which contains 16% of exact short-range exchange.\textsuperscript{86,87}
2.2. DENSITY FUNCTIONAL THEORY

Another noteworthy downfall of DFT is the so-called band-gap problem. The difference between the HOMO and LUMO defines the fundamental gap (or band-gap when referring to solids). In DFT, however, the difference in the KS orbital energies is much smaller than the expected fundamental gap. This is due to the discontinuity of the derivative of the energy with respect to the number of electrons. Each time an electron is added to the KS system there is a constant upshift of the eigenvalues. This means that a single particle excitation raises the HOMO by a constant, resulting in the KS gap being smaller than the fundamental gap. This is often referred to as a problem as it fails to physically reproduce the fact that the HOMO-LUMO gap is the fundamental gap, however, there is an argument for the KS gap actually being a good model for the lowest optical excitation of a system.

In KS theory, all orbitals experience the same field of \(N - 1\) electrons i.e. a virtual orbital interacts with the same potential as the occupied orbital. This is in contrast to HF theory where the unoccupied/virtual orbitals and occupied orbitals experience the potential of \(N\) and \(N - 1\) electrons, respectively. When a single particle excitation occurs, the orbital from which the electron was removed will in fact shift to a higher energy. As such, the KS gap emulates the optical gap of a material. This approximation and the approximation of the KS gap as the fundamental gap are compared and discussed further in chapter 7.

2.2.5 Time-Dependent Density Functional Theory

For the calculation of the excited states and optical properties of materials, ground state DFT is not suitable and time-dependent DFT (TD-DFT) must be used instead. TD-DFT attempts to solve the time-dependent Schrödinger equation (TDSE), in which the external potential is time-dependent. Specifically, the time-dependent Kohn-Sham equations are solved in the TD-DFT calculations throughout this thesis within the linear-response formalism. This is done by finding the density as it varies with an external scalar field. The time-dependent Kohn-Sham equation is shown below:

\[
\frac{i\hbar}{\partial t} \Psi(r,t) = \left[ -\frac{\hbar^2}{2m} \nabla^2 + v_{\text{KS}}(r,t) \right] \Psi(r,t)
\] (2.35)

Runge-Gross theorem

At the core of TD-DFT is the Runge-Grosse theorem, which is analogous to HK theory in ground state DFT. In 1984, Runge and Grosse proved the existence of one-to-one mapping between the time-dependent external potential and the resultant time-dependent electron density, evolving from a fixed initial state \(\Psi_0\). It shows that the difference between two differing external potentials evolving in time from the same initial state must be more than an additive time-dependent constant. With this condition it follows that the two systems of uniquely different potentials must give rise to uniquely different densities. As in HK theory, the implication of this relationship is that with the electron density alone the system’s external
CHAPTER 2. BACKGROUND AND THEORY

potential and, therefore, the corresponding Hamiltonian can be derived and thus the TDSE may be solved.

Van-Leeuwen theorem

As discussed in section 2.2.3, the practical implementation of DFT relies on the KS formalism in which the fully interacting density is modelled as a non-interacting system of electrons. The same approach is taken in TD-DFT and the validity of using such an approximation was proven by van Leeuwen in 1999. In van Leeuwen’s paper, the non-interacting \( v \)-representability of many-particle systems was demonstrated. It stated that for a given density \( \rho(r,t) \) obtained from a many-body system with \( \hat{H} \) and initial state \( |\Psi_0\rangle \), a different initial state may be chosen with an interaction potential \( w' \) that will reproduce the density \( \rho(r,t) \) and for it a unique external potential will exist that differs by an additive time-dependent constant. Therefore, there is a valid and unique external potential that exists for \( w' = 0 \) that will yield the correct, specified density at all times. Given that the initial density and initial time derivative of the density are correct, the issue of non-interacting \( v \)-representability is solved. A notable case of the van Leeuwen theory is for two systems with identical interaction potentials \( (w = w') \) and initial states \( |\Psi_0\rangle = |\Phi_0\rangle \), where a unique external potential will produce the correct density of the interacting system, neatly encapsulating the Runge-Gross theorem.

Adiabatic approximation

Modelling of the exchange-correlation functional is complicated enough in the ground state without having to account for the time dependence of the density. Under the adiabatic approximation used in TD-DFT the potential is taken to only depend on the density at each instant in contrast to the true situation where the potential is dependent on the entire history of the density. Under this assumption the use of ground state functionals is justified, although range-separated hybrid functionals may be required to produce accurate results (discussed below).

Linear response TD-DFT

The implementation of TD-DFT in quantum chemistry codes is usually based on linear-response (LR) theory. All of the TD-DFT calculations in this thesis used this method. Within the LR-TD-DFT method, the time-dependent density’s response to a weak perturbation is derived in order to calculate the frequencies of the present electronic transitions. This weak perturbation is analogous to a weak electric field in spectroscopic experiments and justifies the truncation of the equations at the first order (linear) perturbation term – hence the name linear response. In LR-TD-DFT the time-dependent system is weakly perturbed and the time evolved potentials are evaluated. Usually only the densities close to the ground state are of interest and as such, the first order perturbation is usually sufficient for calculation of the
optical absorption spectrum.\(^{92}\) The time evolved density may be written as:

\[
\rho(r, t) = \rho_0(r, t) + \delta \rho(r, t)
\]  

(2.36)

And the time-dependent exchange correlation functional:

\[
v_{xc}[\rho_0 + \delta \rho](r, t) = v_{xc}[\rho_0] + \int dt' \int dr' f_{xc}[\rho_0](r, r', t - t') \delta \rho(r, t)
\]  

(2.37)

Where \(f_{xc}\) is the exchange-correlation kernel:

\[
f_{xc}[\rho_0](r, r', t - t') = \frac{\delta v_{xc}(r, t)}{\delta \rho(r', t')} \bigg|_{\rho=\rho_0}
\]  

(2.38)

As alluded to above, truncation of the terms at the first order simplifies the problem at hand and the exchange correlation kernel or potential can be written as a functional derivative of the ground state density. The point-wise susceptibility \(\chi[\rho_0](r, r', t - t')\), an important quantity in LR-TD-DFT, describes the change in density at position \(r\) and at time \(t\), when the external potential experiences a small change at a point \(r'\) and time \(t'\). In other words, it quantifies the response of the density of the system to the perturbation.\(^{92}\) It can be written as:

\[
\delta \rho(r, t) = \int dt' \int dr' \chi[\rho_0](r, r', t - t') \delta v_{ext}(r, t)
\]  

(2.39)

The point-wise susceptibilities of the non-interacting KS system and the fully-interacting system must equal each other because the densities must respond to the perturbation in the same way. So, in order to obtain the excitation energies of the system, the density change of the interacting system is equated to that of the KS system. Following a Fourier transform to convert the equation into the frequency domain and substituting in the exchange-correlation kernel term, this yields the central equation of LR-TD-DFT:

\[
\chi(r, r', \omega) = \chi_{KS}(r, r', \omega) + \int dr_1 \int dr_2 \chi_{KS}(r, r_1, \omega) \left\{ \frac{1}{|r_1 - r_2|} + f_{xc}(r_1, r_2, \omega) \right\} \chi(r_2, r', \omega)
\]  

(2.40)

where every object is a functional of the ground state density.

When the frequency \(\omega\) corresponds to the frequency of a true transition, there is a spike in the response function – the \(\chi_{KS}\) has a pole as a function of \(\omega\), corresponding to the single-particle excitations of the KS system.

For the practical implementation of these equations a few approximations worth mentioning are made. The simplest method for solving the equations in LR-TD-DFT is to reduce them down to an eigenvalue problem, which most quantum chemistry codes solve \textit{via} the so-called Casida equations.\(^{93}\)
\[
\begin{pmatrix}
A & B \\
B^* & A^*
\end{pmatrix}
\begin{pmatrix}
X \\
Y
\end{pmatrix} = \omega
\begin{pmatrix}
-1 & 0 \\
0 & 1
\end{pmatrix}
\begin{pmatrix}
X \\
Y
\end{pmatrix}
\]

where

\[
A_{ia,jb} = \delta_{ij}\delta_{ab}(\epsilon_a - \epsilon_i) + 2 \hat{d}_3\hat{d}_3 r \Phi^*_q(r) f_{Hxc}(r, r') \Phi_q(r')
\]

\[
B_{ia,jb} = 2 \hat{d}_3\hat{d}_3 r \Phi^*_q(r) f_{Hxc}(r, r') \phi_{-q}(r')
\]

for occupied orbitals \(i, j\) and unoccupied orbitals \(a, b\), where \(f_{Hxc}\) is the Hartree exchange-correlation kernel:

\[
f_{Hxc}(r, r', \omega) = \frac{1}{|r - r'|} + f_{xc}(r, r', \omega)
\]

Often, the Tamm-Dancoff approximation is employed, under which backwards transitions are ignored, setting the B-matrix elements to zero. Where excited state optimisations (adiabatic excitations) have been calculated in this thesis, the Tamm-Dancoff approximation is applied due to its lesser propensity for deviating from the ground state geometry.

**Performance of TD-DFT**

Despite the success of TD-DFT in describing excited states of many types of systems, the method experiences some major pitfalls for certain systems. A well-known failure of TD-DFT is in the description of Rydberg states – states in which an excitation to a high principle quantum number has occurred i.e. the electron resides far away from the nucleus. The cause of this is the poor description of the virtual orbitals in ground state density functionals. As touched upon in section 2.2.4, the failure of LDA and GGA functionals to exhibit the correct asymptotic decay of the density leads to a lack of diffusivity in the virtual orbitals, resulting in an underestimation of these excitations energies.\(^{93}\)

Another issue in TD-DFT is the description of charge-transfer (CT) excitations, which occur when there is a large separation between the electron and hole. The energy of the lowest charge-transfer excitation in the large separation limit tends to:

\[
\omega_{CT}^{exact} \rightarrow IP^D - EA^A - \frac{1}{r}
\]

where \(IP^D\) is the ionization energy of the donor moiety, \(EA^A\) is the electron affinity of the acceptor moiety and \(-1/r\) (where \(r\) is the separation between the acceptor and donor orbitals) describes the electrostatic attraction between the two. Again, the \(1/r\) dependence of the density is poorly described and the electrostatic attraction between the two charges is absent. By referring to eq. (2.44) and considering that the spatial overlap between the donor and acceptor moieties tends to zero, the CT excitation energy simply becomes the difference
between the two orbitals. For the reasons discussed in section 2.2.4, functionals without exact exchange lead to underestimation of the KS gap (if taken to be equivalent to the fundamental gap) and therefore, the CT excitations end up being greatly underestimated.\textsuperscript{93,94}

The issues highlighted can be overcome by the use of range-separated hybrid functionals, which contain varying amounts of exact HF exchange for short and long-range densities, such as those discussed in section 2.2.4.

### 2.3 Basis Sets

In Hartree-Fock and DFT, the choice of functions used to expand the wavefunction and represent the complex differential equations as simple algebraic equations is referred to as the basis set. There are various types of basis sets suited to different applications such as atom-centred ones for molecular systems and plane-wave basis sets for periodic, crystalline systems.\textsuperscript{95,96} In the calculations in this thesis atomic orbital basis sets were employed.

A Slater-type function is the ideal functional representation of an atomic orbital as it possesses the cusp at $r = 0$ and the correct exponential decay ($e^{-ax}$).\textsuperscript{69} However, it is these characteristics that make this function hard to integrate, and therefore, an impractical choice for describing the atomic orbitals. More often, Gaussian type orbitals (GTOs) are employed instead. While Gaussian functions fail to mimic the cusp at the nucleus and proper exponential decay (using $e^{-ax^2}$ instead of $e^{-ax}$), they are much easier to manipulate mathematically.\textsuperscript{97} The reason for this is the Gaussian product rule, which states that the product of two Gaussian functions equates to another Gaussian. Therefore, the product of two Gaussians A and B would be evaluated as a single Gaussian (a linear combination of A and B) centred on the mid-point of A and B, greatly simplifying the matter at hand.\textsuperscript{98}

A GTO takes the functional form:

$$\phi_{l_x, l_y, l_z}^{GTO}(x, y, z) = N x^{l_x} y^{l_y} z^{l_z} e^{-\zeta r^2}$$ (2.45)

Where $x$, $y$ and $z$ are the atomic Cartesian coordinates, $l_x$, $l_y$, and $l_z$ describe the angular quantum number i.e. the type of orbital being modelled, $\zeta$ is the orbital coefficient and $r$ is the distance from the nucleus. One Gaussian function, however, is not enough to describe the atomic orbitals to the desired level of accuracy. A linear combination of multiple functions are therefore usually used to represent the atomic orbitals, coming closer to the shape of a Slater type orbital by combining the more STO-like decay of the loose Gaussians and the higher peak of the tight ones, as shown in fig. 2.2. At larger $r$ the STO-3G function closely follows that of the 1s-STO function but deviates from it closer to the nucleus.

Typically, contracted Gaussian basis sets are constructed, where multiple primitive functions are used to describe each of the atomic orbital. In a so-called double-zeta basis set there are two primitive Gaussians describing each atomic orbital, each with a different value for the zeta exponent and a different coefficient according to the contribution of each function.\textsuperscript{69,98}
CHAPTER 2. BACKGROUND AND THEORY

Figure 2.2: Radial distribution of hydrogen 1s basis functions in atom-centred coordinates; the dashed lines are the tight (blue), medium (orange) and loose (green) Gaussian functions, which are linearly combined to form the STO-3G (solid red) orbital. These are compared to the 1s-STO function (solid purple).

Accuracy is increased by systematically increasing the number of primitive Gaussians to give basis sets with triple and quadruple-zeta. However, as always, this comes at an increased computational cost. Split-valence basis sets are commonly used, where a larger number of functions are assigned to the valence electrons than the core electrons. A well-known split-valence basis set is 6-311G. In this basis set the core orbitals are described by a function formed of 6 linearly combined Gaussian functions; each successive valence orbital is split into three functions formed of three, one and one Gaussian functions. For example, the 6-311G description of a carbon atom with electron configuration $1s^22s^22p^2$ would be: one function consisting of 6 Gaussian functions for the 1s orbital, three functions consisting of 3, 1 and 1 Gaussian functions for the 2s orbital and three more functions for the 2p orbital made up of 3, 1 and 1 Gaussian functions.

Polarisation and diffuse functions are also essential for accurately modelling large conjugated molecules such as those investigated in this thesis. Polarisation functions add character of orbitals of higher angular momentum to the atomic orbital to model the polarisation of the atoms electron density. So, a p-orbital would have d-orbital angular momentum added to it if a polarisation function was used. In Pople basis sets the inclusion of a polarisation function is denoted by a * or by explicitly stating the angular momentum added e.g. 6-311G** or 6-311G(d,p), whereas in Ahlrich’s basis sets the addition of ‘P’ indicates polarisation functions. Diffuse functions involve the use of different zeta exponents in the Gaussian functions, which
alter the diffusivity of the orbital. Diffuse functions allow us to improve upon the long-range electron density and to also include some vital properties of an atom such as polarisability. Diffuse functions are especially relevant for description of anionic species, where the additional electron resides further from the nucleus.\textsuperscript{101} Examples of such functions are the Ahlrich’s basis set def2-SVPD, where the presence of diffuse functions is signified by the ‘D’, and the Dunning basis set aug-cc-pVDZ. In Pople basis sets a + sign is used to signify diffuse functions instead (++ where there are even more diffuse functions).

In the higher-accuracy calculations in this thesis (chapter 7), the basis sets mainly employed were in the Ahlrich’s def2 family of basis sets ranging from the smallest def2-SVP (split-valence polarised) to def2-QZVP (quadruple-zeta polarised),\textsuperscript{102} apart from a few calculations that were ran with the DZP (double-zeta polarised) and the aug-cc-pVTZ basis set. For the GW calculations, larger basis sets with diffuse functions were used, namely def2-SVPD, def2-TZVP and def2-QZVP. The extra polarisation functions make a negligible difference to the energies calculated with DFT, but are required when running MBPT calculations.\textsuperscript{103} In the (TD-)DFT calculations run with Gaussian 16 in chapter 4, the 6-311G** and 6-311+G** basis sets were employed.

In table 2.1, some well-known basis sets and the ones used in this thesis for a carbon atom are tabulated to highlight the differences in orbital character included.\textsuperscript{99}

<table>
<thead>
<tr>
<th>Basis Set</th>
<th>Number of s-functions</th>
<th>Number of p-functions</th>
<th>Number of d-functions</th>
<th>Number of f-functions</th>
<th>Number of g-functions</th>
</tr>
</thead>
<tbody>
<tr>
<td>STO-3G</td>
<td>6</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6-311G</td>
<td>4</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DZP</td>
<td>4</td>
<td>2</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6-311+G**</td>
<td>5</td>
<td>4</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>aug-cc-pVDZ</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>def2-SVPD</td>
<td>4</td>
<td>2</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>def2-TZVP</td>
<td>5</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>def2-TZVPP</td>
<td>5</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>def2-QZVPP</td>
<td>7</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
</tbody>
</table>
2.4 $\Delta$SCF

A very simple approach for the calculation of the ionization potential (IP) and electron affinity (EA) of a material is the $\Delta$SCF approach. In this method, the IP and EA are determined by calculating the energetic difference between the $N$ electron system and the $N+1$ or $N-1$ electron systems, respectively, as calculated by DFT.

\[
\text{IP} = E(N-1) - E(N) \quad (2.46)
\]

\[
\text{EA} = E(N) - E(N+1) \quad (2.47)
\]

Although this method is simple, it provides an accurate description of the IP and EA as the $N+1$ or $N-1$ systems are allowed to relax – a property that is ignored in Koopmans’s theorem. Koopman’s theorem states that the negative of the HOMO in closed shell HF theory is analogous to the IP, under the assumption that the orbitals of a neutral molecule are identical to those of the ionised systems.\textsuperscript{104,105} In HF, the virtual orbitals feel the field of $N-1$ electrons, whereas the occupied orbitals feel the field of $N$ electrons. Since the IP is the energy required to remove an electron from an occupied orbital, the negative of this can be approximated as the HOMO. Conversely, since the EA is the addition of electron into an unoccupied (virtual) orbital from vacuum, this can be approximated as the negative of the LUMO – although the approximation for EA is commonly made, there is no reference to this in the original paper by Koopman. Despite the $\Delta$SCF method being formally correct, the accuracy of the results is dependent on the basis set and, as always in DFT, on the quality of the exchange-correlation functional.

2.5 The GW method

Although DFT is universally used for the calculation of properties of materials, there are a number of downfalls (discussed in section 2.2) that are usually a result of the approximate exchange-correlation functional. As such, even though DFT has been shown to produce excellent results for many systems, it is unknown how good it actually is; it is therefore instructive for it to be benchmarked against methods that are inherently more accurate.\textsuperscript{106} One such method is the GW method, which is based around the many-body Green’s function.

The exact many-body Green’s function (MBGF) has its poles at exactly the HOMO and LUMO of the electronic system and as such, the IP and EA can be deduced from it. In the GW method, a Taylor expansion of the self-energy $\Sigma$ in terms of $G$ (the Green’s function) and $W$ (the screened Coulomb potential) is carried out. The Green’s function $G$, is a matrix, the elements of which are obtained from the Hedin equations shown below. The coordinates $(r_i t_i)$ are replaced with the index $i$ to simplify the notation, so $G(r_1 t_1, r_2 t_2)$ becomes $G(1, 2)$.\textsuperscript{107,108}
\[ G(1, 2) = G(1, 2) \boxplus \int d(34) G(1, 3) \Sigma(3, 4) G(4, 2) \] (2.48)

\[ \Sigma(1, 2) = i \int d(34) G(1, 3^+) W(1, 4) \Gamma(3, 2, 4) \] (2.49)

\[ W(1, 2) = v(1, 2) + \int d(34) v(1, 3) P(3, 4) W(4, 2) \] (2.50)

\[ P(1, 2) = -i \int d(34) G(1, 3) \Gamma(3, 2, 4) G(4, 1^+) \] (2.51)

\[ \Gamma(1, 2, 3) = \delta(1 - 2) \delta(2 - 3) + \int d(4567) \frac{\delta \Sigma(1, 2)}{\delta G(4, 5)} \times G(4, 6) G(7, 5) \Gamma(6, 7, 3) \] (2.52)

The Hedin equations up until eq. (2.51) must be solved self-consistently in order to obtain the self-energy. In practice, a DFT starting point is used; the KS quasiparticle spectrum is used as the starting point to solve to the independent particle Green’s function, \( G_0 \).

The solution of the above equations to self-consistency is of course, very computationally expensive and usually further approximations are made. One such approximation is the \( G_0 W_0 \) method in which the Hedin equations are solved for once and the calculation is not self-consistent. This method is used throughout chapter 7 and it is the simplest flavour of the GW approximation.

### The Bethe-Salpeter equations

The last of the Hedin equations, which is not involved in the self-consistent solution of the first four equations, is the Bethe-Salpeter equation (BSE). While solving the first four equations yields the IP and EA of the system of interest, solving the BSE allows for the calculation of optical excitation energies.\(^{110}\) This method was used in chapter 7 for comparison against TD-DFT calculated optical gaps and the KS gaps as a proxy for optical gap.

### 2.6 Semi-empirical methods

Although it can be useful to perform high-level calculations using methods such as GW when we want a highly accurate description of a chemical system, it can become challenging or even impossible once the system size is increased. For calculations on larger chemical systems semi-empirical methods can be employed instead.

Semi-empirical quantum chemical methods are low-level methods that are computationally inexpensive and are parameterised using data from experiment or ab initio data. They can
be roughly three orders of magnitude faster than DFT or HF calculations and are useful in the calculation of properties of large datasets or large molecules (such as proteins).\textsuperscript{111,112}

Semi-empirical methods are typically methods built upon HF theory. In semi-empirical quantum chemical methods, the wavefunction of a molecule is taken to be a linear combination of atomic orbitals (LCAO) with the electrons bound to the atom centre. Many of the relevant integrals are either described by empirically derived parameters or neglected to varying degrees depending on the method used. An example of a semi-empirical method is Hückel theory, in which an LCAO is performed in order to describe only the $\pi$ orbitals and the overlap integral completely neglected. In this way the resulting secular equations can easily be solved self-consistently.\textsuperscript{113}

Another early semi-empirical method, more sophisticated than Hückel theory as it includes in its description of the wavefunction the sigma orbitals as well as the $\pi$ orbitals, is CNDO (complete neglect of differential overlap) in which differential overlap between different orbitals is entirely excluded.

Tight-binding DFT (DFTB) can also be considered to be a semi-empirical method and uses the same LCAO principle as a starting point.\textsuperscript{111} DFTB is based around a Taylor expansion of the KS representation of the energy in terms of the electron density of the system.\textsuperscript{111} A reference density is chosen so as to minimise the energy of the KS orbitals, as per the KS equations, however, this is not self-consistent as the density is not recomputed. This reference density is usually the atomic orbital density (as calculated by DFT) and therefore, the energy calculated is not dependent on the chemical environment and is the reason DFTB is transferable to different molecules. The energy in terms of the reference is DFTB1, however, inclusion of higher terms from the Taylor expansion lead to DFTB2 and DFTB3.

Two variants of tight-binding DFT were used in the work in this thesis: xTB (extended tight-binding) and DFTbaby, which both build upon DFTB3.\textsuperscript{114} Two versions of the first of these methods is used extensively in this thesis, namely, IPEA-xTB (which is built upon GFN-xTB (Geometry, Frequency, Non-covalent – xTB) and sTDA-xTB (Simplified Tamm-Dancoff Approximation – xTB). These methods use an augmented atomic orbital basis set which includes diffuse and polarization functions. The GFN-xTB variant is known to produce good geometries, employing a minimal basis set of atom centred Slater functions with s-polarization on hydrogen atoms for an improved description of hydrogen-bonding and d-polarization for heavy elements. The IPs and EAs were calculated by the IPEA-xTB version and the excitation energies by sTDA-xTB.

The sTDA-xTB method is an extension of sTDA – a method developed as an improvement to the Tamm-Dancoff approximation (TDA) that is employed in TD-DFT calculations.\textsuperscript{115,116} It involves a simple monopole type approximation for some of the two-electron integrals and truncation of single excitation space. This procedure was shown to work well, with the slow step of the calculation being the self-consistent solutions for the creation of the molecular orbitals. Incorporation of xTB method into this method led to sTDA-xTB, which sped up the
2.7 Solvation Models

As discussed in the preceding sections, there are many parameters to consider when attempting to model materials in a realistic way. While the level of accuracy chosen for the calculation of material properties can be improved by choosing quantum chemical methods over classical ones and judicious choice of parameters such as basis set or functionals (in DFT), the situation modelled is often unrealistic if the effect of solvent is ignored. The PES for a polymer in the gas-phase, for instance, may differ either qualitatively or quantitatively to its solvated counterpart. Reactions that may not be thermodynamically favourable in the gas-phase may be possible in solvent due to the stabilisation of the products. In the context of photoactive materials, the optical gap is often affected by the solvent, giving rise to different results and observations. The inclusion of solvent effects in calculations, particularly in the case of photocatalytic water splitting and dye-sensitised solar cells, where the polymers have to operate under solvated conditions, is therefore essential. There are two main ways of incorporating the effects of solvation in a chemical simulation, either by explicitly including solvent molecules in the simulation or by the use of implicit solvation models, both of which are discussed in detail below.

2.7.1 Explicit Solvation

Explicit solvation is simply the explicit inclusion of solvent molecules to the simulation. This is the most accurate method as it allows for the observation of how the orientation and structure of the solvent molecules are actually affected by being in the presence of the material being modelled. It also means that it is possible to analyse the density of the solvent and whether the solvent density is greater in particular regions of the material.

However, for the solvent effect to be modelled accurately, the amount of solvent molecules required is enormous. It is not sufficiently accurate to include only the first solvation sphere,
i.e. the first layer of solvent molecules surrounding the materials; the effect of the outer solvation spheres is also of high importance as it affects how the inner solvent molecules interact amongst themselves and with the material. As such, in order to get the level of accuracy to a satisfactory level with an explicit solvent model, the computational cost of simulating a large system becomes prohibitively high.

2.7.2 Implicit Solvent Models

Implicit solvation models are a more computationally efficient way to account for solvent effects. Although solvent structure effects cannot be investigated with implicit solvation models, the effect of the solvent on the properties and structure of the material being simulated can be modelled to a satisfactory degree at a much lower cost. Implicit or continuum solvent models involve the insertion of the solute into a cavity in a dielectric continuum that possesses the same dielectric constant as the chosen solvent. The terms that make up the free energy of solvation can be separated as:

\[
\Delta G_{\text{solvation}} = \Delta G_{\text{cavity}} + \Delta G_{\text{dispersion}} + \Delta G_{\text{polarisation}}
\]  

A variety of continuum solvation models exist, differing in ways such as how the shape and size of the cavity are defined and how the dielectric constant is described.

**Born solvation model**

Although more advanced solvation models are used throughout this thesis, the following section first discusses the Born model, the most basic implicit solvation model, in order to introduce the basic concepts. In the Born model the solute is enclosed in a spherical cavity; it is, therefore, only a good model for a single atom. The Born free energy of solvation is given by:

\[
\Delta G_{\text{elec}}(q) = - \left( 1 - \frac{1}{\epsilon} \right) \frac{q^2}{2a}
\]  

where \( q \) is a net charge in a cavity of radius \( a \) and \( \epsilon \) is the dielectric constant of the medium. Not only is this not a satisfactory model of the solute in solution due to the spherical cavity, but there is a lesser discrepancy between different high dielectric constants than there is for lower dielectric constants i.e. two high \( \epsilon \) solvents give rise to similar free energies of solvation. The model also excludes the free energy of cavity formation and van der Waal’s effects.\(^{98}\)

In the xTB calculations carried out in chapters 3 to 6, the generalised Born solvation model (GBSA) - a variation of the above model - is used. The GBSA model expands on the Born model by including the Coulomb interaction between partial charges and can be extended to
polyatomic molecules. It is given by:

$$G_{\text{elec}} = -\frac{1}{8\pi\epsilon_0} \left( 1 - \frac{1}{\epsilon} \right) \sum_{i,j} \frac{q_i q_j}{f_{GB}}$$

(2.55)

where

$$f_{GB} = \sqrt{r_{ij}^2 + a_{ij}^2 e^{-D}}$$

(2.56)

and

$$D = \left( \frac{r_{ij}}{2a_{ij}} \right)^2 a_{ij} = \sqrt{a_i a_j}$$

(2.57)

This model more accurately describes high $\epsilon$ solvents. The function $f_{ij}$ is dependent on the Born radius and internuclear distance between two atoms $a_i$ and $a_j$. The solvent-solute boundary is usually a scaled van der Waals surface. These methods can also be extended to ellipsoid cavities with the Kirkwood-Onsager approach, however, although the use of such cavities allows for the analytical solution of the equations, they are not the best descriptor for a solute in solution.

### Polarisable continuum solvation models

Polarisable continuum solvation models (PCM) are improved solvation models with molecular shaped cavities that solve the equation for free energy of solution self-consistently rather than analytically. The charge of the solvent on the solute’s wavefunction is calculated and in turn the solute’s response on the solvent is re-calculated until reaching a self-consistent solution. In a PCM the equations for a solute embedded in a dielectric medium are developed and solved self-consistently.\(^\text{121}\) Throughout the work in this thesis a COnductor-like Screening MOdel (COSMO) is used instead.\(^\text{122}\) For COSMO the equations are developed for a solute embedded in a metal (i.e. a medium with $\epsilon$ of infinity) but re-scaled for a dielectric constant that represents the chosen solvent.\(^\text{69}\) As in PCM, the resulting wavefunction of the solute in response to this medium is calculated. Essentially, the equations for the electrostatic energy are easier to solve for a metal, which is why many quantum chemical programs make use of this method.

### 2.8 Simplified Molecular-Input Line Entry System (SMILES)

Throughout this thesis the use of SMILES strings is often mentioned. SMILES is a system for encoding a chemical structure as an ASCII string, to be read by molecular editors and converted into two or three-dimensional chemical representations.\(^\text{123}\) SMILES strings can be used to describe any chemical structure by following a set of rules. Atoms belonging to the
organic subset of atoms (C, N, O, B, P, S, F, Cl, Br) do not require brackets. If, however, atoms have a formal charge, are not in the above subset of atoms, have an abnormal number of hydrogen atoms attached to them, have a chiral centre or are not in their normal isotopic form, they are enclosed in square brackets. Single bonds are typically not included explicitly but double, triple and quadruple bonds are denoted by =, # and $, respectively. Below is an example of a SMILES string representation of hexa-2,4-diene.

\[
\begin{align*}
\text{CC} &= \text{CC} = \text{CC} \\
\end{align*}
\]

Figure 2.3: Hexa-2,4-diene 2-D representation and SMILES string

Rings are written by breaking the ring up and labelling the break point with a number. There is no single correct format for describing rings and in general any point in the ring can be selected, however, some representations do make more intuitive sense than others. Aromatic rings are written in the same way as alphatic rings but with lowercase letters instead, although a Kekule style structure can be achieved by explicitly including alternating double bonds. For example:

\[
\begin{align*}
\text{C1CCCCC1} &\quad \text{C1CCCCC1} \\
\text{c1ccccc1} &\quad \text{c3ccc2cc1cccccc1cc2c3} \\
\end{align*}
\]

Figure 2.4: 2-D representation and SMILES strings of bicyclo[4.3.0]nonane, cyclohexane, benzene and anthracene.

SMILES strings can also include information about branches on the main skeleton and stereochemistry. Overall, they are a very versatile and useful format for describing chemical structure. The only caveat is that a structure can have many corresponding SMILES strings, which may lead to complications in data analysis. When parsing through data where the SMILES strings are an important feature, inconsistency in the SMILES strings can be problematic. This can be overcome by canonicalisation of the SMILES strings by an algorithm that returns unique SMILES strings for each unique structure.

In all of the high-throughput screening work in this thesis, the original input has been SMILES strings. This allowed for a much faster input of the structure as they can be fed in as a manageable Python list of SMILES strings. Furthermore, analysis of the results was greatly simplified by identifying structures by the SMILES strings, which was included in the output files.
Chapter 3

Accessing Near-DFT Results by Calibration of Properties from Semi-Empirical Calculations Against DFT

3.1 Introduction

In this thesis, a variety of different methods, from high-level to low-level, were employed in order to delve into the optoelectronic properties of organic materials. Methods such as time-dependent density functional theory ((TD-)DFT) and higher accuracy GW/BSE are certainly useful for detailed studies of a small number of structures, however, when the focus shifts from numbers of structures on the order of $10^2$ to $10^4$ it is no longer computationally tractable to use these methods to obtain results within a reasonable time frame. In high-throughput virtual screening (HTVS), the properties of tens of thousands of materials can be calculated in a fraction of the time that it would take to complete higher accuracy calculations allowing potentially interesting materials or molecules to be identified preliminarily. Subsequently, high-level calculations on selected structures of interest may be carried out to obtain higher accuracy results or to confirm their adherence to the property requirements for a given application one might have in mind, e.g. to confirm whether a material has the right properties to be a suitable dye in dye-sensitized photocatalytic water splitting.

The HTVS method used throughout this thesis involves the calculation of properties by a low-cost, semi-empirical method, performing high-level calculations ((TD-)DFT) on a subset of these to yield their ionization potential (IP), electron affinity (EA) and optical gap ($\Delta_o$), and then performing a linear fit between the semi-empirical and DFT results. The fitting parameters obtained are then used to transform the semi-empirical results allowing one to access near-DFT accuracy for a significantly larger number of molecules. This linear calibration is the subject of this chapter. The HTVS method that was developed by our group is demonstrated through the use of two different semi-empirical methods ((GFN/IPEA/sTDA)-xtB and DFTbaby) and two different chemical datasets. These two datasets are the small
CHAPTER 3. CALIBRATION

conjugated molecules and the larger organic dyes. The small molecules library consisted of 143 unique cyclic, conjugated molecules; these are the unsubstituted cores. Of these, 63 were modified by incorporation of the substituents shown in fig. 3.1, incorporating up to 2 substituents per core.

Three different dye structures were used in this work, ABCBA, ACA′ and ACA, where the A and B building blocks came from the same library of 179 molecules and the C building block was a diketopyrrolopyrrole (DPP) core for all dyes. Examples of some of the building blocks used, the structure of DPP and the sequences used are depicted in fig. 3.2.

Figure 3.1: Twenty of the small molecules used in the calibration of the unsubstituted cores shown in the white box and the 12 substituents used to modulate these cores in the blue box.

Figure 3.2: The three dye sequences used in the following calibrations and the structures of the diketopyrrolopyrrole (DPP) core and a selection of the 179 building blocks used for the A, B and A′ positions.

In this chapter I discuss the validity of this method and whether the calibration parameters fitted to one type of chemical system can ubiquitously be used to improve semi-empirically
calculated properties for different chemical systems. While the two datasets largely differ in terms of molecule size, they are both within the family of organic, conjugated molecules.

Some of the content of this chapter has been published in:


### 3.2 Methodology and Computational Details

The IPs, EAs and optical gaps for the organic dyes and small molecules were calculated by the standard procedure used in all of the HTVS investigations carried out in this thesis, which is described in detail below.

#### 3.2.1 Small Molecules Workflow

For the small molecules investigation, a library of 143 small, cyclic, conjugated molecules was created. A subset of these skeletons were functionalised with up to 2 out of 12 electronically different substituents, which are shown in fig. 3.1. An exhaustive search to calculate every permutation of the positions of substituents was performed, which is why it was not possible to extend the study to a greater number of substituents within our computational resources.

A conformer search was performed on a selection of the substituted cores, where for each structure 30 conformers were randomly generated using the Experimental-Torsion Distance Geometry with additional basic knowledge (ETKDG). The lowest energy conformer for each structure was then determined using the Merck Molecular Force Field (MMFF), as implemented in RDKit. As the unsubstituted skeletons are relatively rigid, a conformer search was not performed on them. The IP, EA and optical gap of each structure in the unsubstituted cores and the lowest energy conformer in the substituted cores datasets were calculated using the semi-empirical tight-binding method (GFN/IPEA/sTDA)-xTB using the code xtb 5.6 and stda. A generalised Born solvation model is available within GFN-xTB and was therefore used for the optimisation and IP/EA calculations, with benzene as the chosen solvent. For these molecules the (TD-)DFT calculations were run using Turbomole 7.01 and a few basis set/functional combinations were attempted. In the first instance the standard B3LYP/DZP option was chosen for both DFT and TD-DFT calculations. Consequent (TD-)DFT calculations used B3LYP with the aug-cc-pVTZ basis set in both the gas-phase and in benzene, and also the range-separated hybrid functional ωB97X for the TD-DFT calculations. Finally, a linear fit was applied in order to access the new near-DFT values.
CHAPTER 3. CALIBRATION

3.2.2 Dyes Workflow

Starting from a library of 179 monomers, 3 dye classes (ABCBA, ACA and ACA$'$) were constructed, all possessing a common core of diketopyrrolopyrrole (DPP), denoted by the C position in the sequences ABCBA, ACA and ACA$'$. The A and B (A$'$, in the ACA$'$ case) building blocks were taken from the same library of 179 building blocks. Using a Python pipeline that I wrote, all possible combinations within these sequences were constructed using stk,\textsuperscript{128} a Python library that takes functionality from RDKit.\textsuperscript{128} For each structure 30 conformers were randomly generated using the Experimental-Torsion Distance Geometry with additional basic knowledge (ETKDG).\textsuperscript{126} The lowest energy conformer for each structure was then determined using the Merck Molecular Force Field (MMFF)\textsuperscript{127} as implemented in RDKit.\textsuperscript{128}

Subsequently, the properties of each lowest energy conformer were calculated with the semi-empirical tight-binding method (GFN/IPEA/sTDA)-xTB\textsuperscript{129–131} using the code xtb 5.6.\textsuperscript{132} The same generalised Born solvation model available within GFN-xTB was used for the optimisation and IP/EA calculations, with water and benzene as the chosen solvents. In the following discussion the water results are the primary focus. As there is no solvation model implemented for sTDA-xTB, the optical gap calculations were performed without one. For dyes not containing sulfur, long-range corrected tight-binding DFT (LR-TDDFTB) calculations were run using DFTbaby 0.1.0, employing the default element-specific parameters to calculate the lowest six excited states.

For the calibration of the xTB-calculated properties of the dyes, the IP, EA and optical gap were calculated using (TD-)DFT. The ground state, anionic and cationic structures were optimised using B3LYP/6-311G** and the IP and EA were calculated using the $\Delta$SCF method discussed in section 2.4 and chapter 7. For the TD-DFT calculations the $\omega$B97X/6-311+G** functional/basis set setup was used. All DFT calculations were run using Gaussian 16. A linear fit was finally applied, as for the small molecules.

3.3 Results and Discussion

3.3.1 Small Molecules

Calibration

For the small molecules, the dataset used consisted of 143 unsubstituted cores and 63 substituted cores. At the first stage of the calibration a variety of combinations of DFT functional and basis set were used in order to determine the calibration with the highest accuracy at the lowest computational cost. The mean absolute error (MAE) between the calibrated properties and the DFT-calculated properties was used to quantify the accuracy of the calibrated properties relative to the DFT calculations, for each set of fitting parameters obtained from
the different linear fits. Table 3.1 shows the MAE between the calibrated properties and the
DFT-calculated properties of the unsubstituted cores for each computational setup and the
R² of the corresponding calibration curves. The fitting parameters obtained from the linear
regression models are also tabulated.

Table 3.1: The R² values from the calibration curves of the unsubstituted cores and MAE between
the DFT and calibrated (GFN/IPEA/sTDA-xTB) properties of these structures.

<table>
<thead>
<tr>
<th>Computational setup for DFT</th>
<th>IP</th>
<th>EA</th>
<th>Δo</th>
</tr>
</thead>
<tbody>
<tr>
<td>B3LYP/DZP benzene</td>
<td>R²</td>
<td>0.88</td>
<td>0.91</td>
</tr>
<tr>
<td></td>
<td>MAE / eV</td>
<td>0.20</td>
<td>0.20</td>
</tr>
<tr>
<td></td>
<td>Slope</td>
<td>1.05</td>
<td>1.04</td>
</tr>
<tr>
<td></td>
<td>Intercept</td>
<td>0.02</td>
<td>1.47</td>
</tr>
<tr>
<td>B3LYP/aug-cc-pVTZ benzene</td>
<td>R²</td>
<td>0.89</td>
<td>0.96</td>
</tr>
<tr>
<td></td>
<td>MAE / eV</td>
<td>0.20</td>
<td>0.12</td>
</tr>
<tr>
<td></td>
<td>Slope</td>
<td>1.08</td>
<td>0.82</td>
</tr>
<tr>
<td></td>
<td>Intercept</td>
<td>0.15</td>
<td>0.62</td>
</tr>
<tr>
<td>B3LYP/aug-cc-pVTZ no solvent</td>
<td>R²</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>MAE / eV</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Slope</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Intercept</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>ωB97X/aug-cc-pVTZ benzene</td>
<td>R²</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>MAE / eV</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Slope</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Intercept</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

In all cases, a strong correlation (R² > 0.82) was observed between the (GFN/IPEA/sTDA-
)xTB- and (TD-)DFT-calculated properties, which improved upon increasing the basis set, as
indicated by the increased R² values down table 3.1. As a solvation model is not implemented
in sTDA-xTB, the TD-DFT calculations were also run in the gas-phase in order to make a
closer comparison to the semi-empirical calculations. The effect on the MAE, however, was
minimal, as seen in table 3.1. In GFN-xTB, any charge-transfer excitations are meant to be
well-described, unlike in standard (TD-)DFT functionals such as B3LYP. As such, the use
of a range-separated functional to better describe the long-range behaviour in a material would
be expected to produce more accurate results and therefore, a lower MAE. This improvement
in the DFT description of the molecules is reflected in the decrease in MAE upon introduction
of the ωB97X functional and the improved basis set. As the discrepancies between MAE for
the B3LYP and ωB97X calculated properties were modest, the second computational setup
was selected for the calibration of other small molecules.
Below, in fig. 3.3, the calibration curves for the B3LYP/aug-cc-pVTZ calculations are shown for the DFT and xTB calculated properties of the unsubstituted skeletons, with \( R^2 \) values of 0.89, 0.96 and 0.86 for the \(-\text{IP}, -\text{EA}\) and optical gap, respectively.

![Calibration curves](image)

Validation

The aim of the HTVS method that I use is to ultimately be able to more accurately predict properties of other molecules using a semi-empirical method and the linear fit obtained from the more expensive (TD-)DFT calculations. In order to confirm this, the (GFN/IPEA/sTDA-)xTB data for the substituted cores was calibrated using the fitting parameters from the linear fit of the unsubstituted cores discussed above. The MAEs following this calibration are shown below in table 3.2.

Table 3.2: The MAEs between the (GFN/IPEA/sTDA-)xTB properties of the substituted cores, calibrated using the fitting parameters from the unsubstituted cores calibration curve, and the DFT calculated properties.

<table>
<thead>
<tr>
<th>Computational setup</th>
<th>IP MAE / eV</th>
<th>EA MAE / eV</th>
<th>( \Delta_o ) MAE / eV</th>
</tr>
</thead>
<tbody>
<tr>
<td>B3LYP/DZP benzene</td>
<td>0.27</td>
<td>0.40</td>
<td>0.43</td>
</tr>
<tr>
<td>B3LYP/aug-cc-pVTZ benzene</td>
<td>0.28</td>
<td>0.24</td>
<td>0.30</td>
</tr>
<tr>
<td>B3LYP/aug-cc-pVTZ no solvent</td>
<td>-</td>
<td>-</td>
<td>0.29</td>
</tr>
<tr>
<td>( \omega )B97X/aug-cc-pVTZ benzene</td>
<td>-</td>
<td>-</td>
<td>0.34</td>
</tr>
</tbody>
</table>

These MAEs confirm that the fitting parameters from the second calibration are most suitable for the prediction of the properties of the substituted cores. The plots for the B3LYP/aug-cc-pVTZ are shown in fig. 3.4.
This model performs reasonably well in the prediction of the substituted skeletons with the mean absolute errors (MAEs) being 0.28, 0.24 and 0.30, for IP, EA and optical gap, respectively. While these values are not as small as in the calibration step, this can be expected as the transferability of the fitting parameters from one type of molecule to another is not guaranteed. The more surprising result was for the calculations using a range-separated functional. Upon changing the B3LYP functional to the range-separated ωB97X functional, the MAE would be expected to decrease, however, this is not what is observed. The MAE between the optical gap data of the unsubstituted cores calibrated using the unsubstituted cores-only fitting parameters and the TD-DFT-calculated optical gaps did in fact decrease as expected (see table 3.1). This was not reproduced when the fitting parameters for the cores ωB97X calibration curve were used to calibrate the sTDA-xTB calculated optical gaps for the substituted cores. Upon inspection of the sTDA-xTB v. TD-DFT curves for the B3LYP/aug-cc-pVTZ and ωB97X/aug-cc-pVTZ datasets of the substituted cores, it seems that the increase in this MAE may be attributed to three outliers that become more apparent upon improving the DFT functional.

These three outliers are all of the quinones with electron-donating substituents in the dataset (shown below in fig. 3.5). Three other quinones with electron-withdrawing groups are present, however, these are not outliers in the calibration curves. The lowest excited state in all cases corresponded to a HOMO→LUMO excitation and visualisation of these orbitals showed no evidence of charge-transfer character that could be causing this behaviour (see figs. A.1 to A.3). The calculations carried out could be repeated for other quinones with electron-donating substituents in order to investigate this further. The two calibration curves are shown below.
CHAPTER 3. CALIBRATION

Figure 3.5: The three outlying quinones with electron-donating substituents in the \( \omega \text{B97X}/\text{aug-cc-pVTZ} \) datasets.

Figure 3.6: sTDA-xTB v. TD-DFT calculated optical gaps, using the B3LYP (left) and \( \omega \text{B97X} \) (right) functionals and \text{aug-cc-pVTZ} basis set. The pink data points are the quinones with electron-withdrawing substituents, the orange data points are the quinones with electron-donating substituents and the blue data points are the rest of the dataset of substituted cores. The black line is the line of best fit for the entire dataset of substituted cores.

Table 3.3: Comparison of the \( R^2 \) and MAE values from the substituted cores dataset, fitted using the parameters from the substituted cores, including and excluding the problematic quinone structures.

<table>
<thead>
<tr>
<th>Computational setup</th>
<th>Property</th>
<th>( R^2 )</th>
<th>MAE / eV</th>
<th>( R^2 )</th>
<th>MAE / eV</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>all</td>
<td>(all)</td>
<td>(no quinones)</td>
<td>(no quinones)</td>
</tr>
<tr>
<td>B3LYP/DZP benzene</td>
<td>IP 0.93</td>
<td>0.26</td>
<td>0.93</td>
<td>0.26</td>
<td></td>
</tr>
<tr>
<td></td>
<td>EA 0.85</td>
<td>0.44</td>
<td>0.84</td>
<td>0.45</td>
<td></td>
</tr>
<tr>
<td></td>
<td>( \Delta_o ) 0.81</td>
<td>0.41</td>
<td>0.80</td>
<td>0.40</td>
<td></td>
</tr>
<tr>
<td>B3LYP/aug-cc-pVTZ benzene</td>
<td>IP 0.93</td>
<td>0.27</td>
<td>0.93</td>
<td>0.26</td>
<td></td>
</tr>
<tr>
<td></td>
<td>EA 0.91</td>
<td>0.24</td>
<td>0.91</td>
<td>0.24</td>
<td></td>
</tr>
<tr>
<td></td>
<td>( \Delta_o ) 0.88</td>
<td>0.31</td>
<td>0.89</td>
<td>0.27</td>
<td></td>
</tr>
<tr>
<td>B3LYP/aug-cc-pVTZ no solvent</td>
<td>IP -</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td></td>
<td>EA -</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td></td>
<td>( \Delta_o ) 0.87</td>
<td>0.30</td>
<td>0.88</td>
<td>0.26</td>
<td></td>
</tr>
<tr>
<td>( \omega \text{B97X}/\text{aug-cc-pVTZ} ) benzene</td>
<td>IP -</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td></td>
<td>EA -</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td></td>
<td>( \Delta_o ) 0.86</td>
<td>0.32</td>
<td>0.89</td>
<td>0.27</td>
<td></td>
</tr>
</tbody>
</table>
3.3. RESULTS AND DISCUSSION

The MAEs of the properties for the dataset not including the problematic quinones are shown in Table 3.3. The removal of the outliers had little effect on the $R^2$ and MAEs of the B3LYP/DZP calibration curves. The B3LYP/aug-cc-pVTZ IP and EA were similarly, minimally affected however, the optical gap MAE did decrease by 0.04 eV. Upon introduction of the range-separated functional $\omega$B97X, this change seen by the removal of the outliers became slightly more pronounced, the $R^2$ increasing to 0.89 and the MAE lowering by 0.05 eV. Interestingly, the MAE for the optical gap DFT calculation in the gas-phase is also 0.02 eV lower than the MAE of the analogous $\omega$B97X calculation when the outliers are present and 0.01 eV lower when they are not. The lack of a large discrepancy between the $\omega$B97X and B3LYP calculated properties upon removal of outliers indicates that in this case, increasing the computational cost by using a more complex functional does not provide an added benefit in terms of accuracy. Therefore, the aug-cc-pVTZ/B3LYP computational setup is the one I would use in any further investigations on these and similar systems.

Finally, Table 3.4 shows the MAE between the DFT-calculated and calibrated properties of the substituted cores dataset, excluding the three outlying data points, using the fitting parameters from the unsubstituted cores (Table 3.1). Indeed, removal of these points lowered the MAEs significantly, with the B3LYP/aug-cc-pVTZ setup still yielding the results closest to the DFT values.

Table 3.4: MAEs between DFT-calculated properties and (GFN/IPEA/sTDA-)xTB properties of the substituted cores excluding the problematic quinone structures, calibrated using the fitting parameters of the unsubstituted cores.

<table>
<thead>
<tr>
<th>Computational setup</th>
<th>IP MAE / eV</th>
<th>EA MAE / eV</th>
<th>$\Delta_{o}$ MAE / eV</th>
</tr>
</thead>
<tbody>
<tr>
<td>B3LYP/DZP benzene</td>
<td>0.27</td>
<td>0.39</td>
<td>0.38</td>
</tr>
<tr>
<td>B3LYP/aug-cc-pVTZ benzene</td>
<td>0.28</td>
<td>0.25</td>
<td>0.26</td>
</tr>
<tr>
<td>B3LYP/aug-cc-pVTZ no solvent</td>
<td>-</td>
<td>-</td>
<td>0.26</td>
</tr>
<tr>
<td>$\omega$B97X/aug-cc-pVTZ benzene</td>
<td>-</td>
<td>-</td>
<td>0.29</td>
</tr>
</tbody>
</table>

3.3.2 Dyes

The diketopyrrolopyrrole (DPP)-based dyes investigated are a vastly different class of molecules to the small aromatic molecules discussed above. It does not make sense to use the fitting parameters from the small molecules data to calibrate semi-empirically obtained property data for the dyes and expect accurate and realistic results. As before, the ultimate aim is to be able to use the inexpensive and fast HTVS workflow to predict near DFT values.
for these dyes specifically and eventually use these results to investigate their applicability to dye-sensitized photocatalytic water splitting (DS-WS), dye-sensitized solar cells (DSSCs) and organic photovoltaics (OPV). Firstly, however, the appropriate calibration curve must be identified and its accuracy and suitability to this dataset assessed. The dye dataset consisted of 48151 dyes, 32041 of which were of the ABCBA class, 15931 of the ACA’ and 179 of the ACA.

The IP and EA of the dyes was calculated using (GFN/IPEA)-xTB and 105 DFT calculations were run on a subset of these. Figure 3.7 shows the correlation between the xTB and DFT-calculated properties, with the post-calibration values shown in yellow. The red dashed line shows the x=y line to highlight the strong linear correlation between the two methods. The $R^2$ values for $-\text{IP}$ and $-\text{EA}$ were 0.83 and 0.86, respectively. Upon inclusion of the monomer dataset in the calibration set, these $R^2$ values improved to 0.90 and 0.96. Although this appears to be an improvement, the MAE between the DFT values and the calibrated xTB values for the dyes-only subset of structures worsened when using these fitting parameters. This raises the question of whether it is indeed correct to include these in the calibration dataset when we are aiming to improve the prediction for a very specific class of molecules.

Although for $-\text{IP}$ and $-\text{EA}$ the difference in correlation was minimal when switching from a dataset including solely the dyes to a dataset including the monomers as well, this was not the case for the discussion of the optical gap. The optical gap was firstly calculated with sTDA-xTB revealing a poor correlation between these values and their DFT counterparts, with an $R^2$ of 0.47. The absence of a solvation model in sTDA-xTB was eliminated as a possible cause of this, as optimising the structures with a solvation model and then running the TD-DFT calculations in the gas-phase did not rectify this problem, exhibiting equally poor correlation. In contrast to the $-\text{IP}$ and $-\text{EA}$ results, the correlation strengthened significantly once the monomers were included in the optical gap calibration curve. The $R^2$ value for this curve was 0.91, a great improvement, however, as before, it does not necessarily make this a more
accurate option as the monomers differ a lot from the dyes both structurally and in terms of properties. As such, an alternative semi-empirical method was chosen – DFTbaby. The optical gaps in the stda calculations were taken to be the first excitation with an oscillator strength above 0.1, this cutoff was therefore used in the DFTbaby calculations.

There was an apparent improvement in correlation in the DFTbaby v. TD-DFT calibration curves ($R^2 = 0.69$) for the dye optical gaps. However, as sulfur atoms are not implemented in the code, sulfur-containing structures had to be removed from the dataset. When these dyes were also removed from the sTDA-xTB v. TD-DFT calibration curve, a similar improvement ($R^2 = 0.59$) was observed, suggesting that the sulfur-containing dyes may be problematic for sTDA-xTB. Figure 3.8 compares the calibration with all dyes and with all non-sulfur containing dyes and the plots in fig. 3.9 show the sTDA-xTB and DFTbaby results before and after calibration for the dyes not containing sulfur.

![Figure 3.8](image1.png)

Figure 3.8: DFT-calculated v. sTDA-calculated optical gaps for the ABCBA, ACA’ and ACA dyes with sulfur (left) and without sulfur (right), calibrated using the fitting data obtained from the dyes alone. The colour scheme is the same as in fig. 3.7

![Figure 3.9](image2.png)

Figure 3.9: DFT-calculated v. sTDA-calculated and DFTbaby-calculated optical gaps for the ABCBA, ACA’ and ACA dyes with calibrated using the fitting data obtained from the dyes alone. The colour scheme is the same as in fig. 3.7

As in section 3.3.1, in order to assess how suitable the chosen fitting parameters are for calibrating a particular dataset, the MAE between the calibrated data and the (TD-)DFT-
calculated data was calculated for each method investigated. Table 3.5 shows the MAEs when the fitting parameters were obtained from the calibration curve of the dyes and monomers combined, and the dyes alone.

Additionally, a final set of DFT calculations were run in order to validate the calibration, analogous to validating the linear fit obtained by the small molecule cores by evaluating the accuracy of the prediction for the substituted cores in section 3.3.1. The results of this final validation are also shown below in table 3.5.

Table 3.5: MAEs for the properties following the different calibrations performed, MAEs of the properties using the validation set and the fitting parameters from the different calibrations, and the $R^2$ values of each fit.

<table>
<thead>
<tr>
<th>Fitting Data</th>
<th>Property</th>
<th>$R^2$</th>
<th>MAE / eV (calibration set)</th>
<th>MAE / eV (validation set)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dyes &amp; Monomers</td>
<td>IP</td>
<td>0.90</td>
<td>0.16</td>
<td>0.22</td>
</tr>
<tr>
<td></td>
<td>EA</td>
<td>0.96</td>
<td>0.13</td>
<td>0.17</td>
</tr>
<tr>
<td></td>
<td>$\Delta_\sigma$ (sTDA)</td>
<td>0.91</td>
<td>0.20</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>$\Delta_\sigma$ (DFTbaby)</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Dyes</td>
<td>IP</td>
<td>0.83</td>
<td>0.13</td>
<td>0.12</td>
</tr>
<tr>
<td></td>
<td>EA</td>
<td>0.86</td>
<td>0.12</td>
<td>0.16</td>
</tr>
<tr>
<td></td>
<td>$\Delta_\sigma$ (sTDA)</td>
<td>0.47</td>
<td>0.12</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>$\Delta_\sigma$ (DFTbaby)</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Dyes &amp; Monomers</td>
<td>IP</td>
<td>0.76</td>
<td>0.19</td>
<td>-</td>
</tr>
<tr>
<td>(no sulfur)</td>
<td>EA</td>
<td>0.93</td>
<td>0.15</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>$\Delta_\sigma$ (sTDA)</td>
<td>0.92</td>
<td>0.20</td>
<td>0.18</td>
</tr>
<tr>
<td></td>
<td>$\Delta_\sigma$ (DFTbaby)</td>
<td>0.89</td>
<td>0.27</td>
<td>0.21</td>
</tr>
<tr>
<td>Dyes</td>
<td>IP</td>
<td>0.86</td>
<td>0.13</td>
<td>-</td>
</tr>
<tr>
<td>(no sulfur)</td>
<td>EA</td>
<td>0.92</td>
<td>0.12</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>$\Delta_\sigma$ (sTDA)</td>
<td>0.59</td>
<td>0.12</td>
<td>0.13</td>
</tr>
<tr>
<td></td>
<td>$\Delta_\sigma$ (DFTbaby)</td>
<td>0.69</td>
<td>0.12</td>
<td>0.14</td>
</tr>
</tbody>
</table>

The validation set consisted of 17 dyes (9 ABCBA, 4 ACA and 4 ACA'), none of them containing sulfur to allow the same 17 dyes to be used for the validation of the IP, EA and optical gaps. The first column in table 3.5 indicates from which dataset the fitting parameters were obtained. At the validation stage, the semi-empirical optical gap results were calibrated using the fitting parameters from the sulfur-free datasets alone. For all dye properties, the MAE decreased when the fitting parameters were obtained from the datasets solely containing the dyes and not the monomers as well. Although the correlations in these calibration curves were weaker than for the datasets with both monomers and dyes, overall, performing the fit on only the dyes produces better results for the prediction of dye properties, highlighting the
importance of transferability in the fitting model.

For both datasets (dyes & monomers, and dyes only), DFTbaby performed worse than sTDA-xTB, although only by 0.03 eV in the case of the dataset containing dyes and monomers, and 0.01 eV in the dataset containing dyes alone. The calibration producing the best results was therefore the one performed on the dyes alone using (GFN/IPEA/sTDA-)xTB, excluding the sulfur containing dyes for the optical gap calculations. These fitting parameters were selected for the analysis that follows in chapter 4. The plots for the DFT v. (sTDA-)xTB-calculated properties for the validation set, before and after the calibrations, are shown in fig. 3.10.

![Figure 3.10: DFT-calculated v. sTDA-calculated and DFTbaby-calculated optical gaps for the ABCBA, ACA′ and ACA dyes with calibrated using the fitting data obtained from the dyes alone.](image)

### 3.4 Conclusions

The results of the studies in this chapter demonstrate the success of the method of calibrating tight-binding DFT property data to DFT, to achieve near-DFT quality properties at a fraction of the time and computational cost. The small molecule calibration curves exhibited strong, positive correlations and therefore, reasonable MAEs when the data on which the fit was performed was calibrated to produce new values. The MAE between (sTDA-)xTB and (TD-)DFT properties was not as good when the same fitting data was applied to a new dataset (the substituted cores). This, of course, was expected and was improved by the removal of outlying data points. The basis set and functional that achieved the best balance between MAE and computational cost was found to be aug-cc-pVTZ and B3LYP.

The calibration for the dyes was less straightforward. While the IP and EA produced calibration curves with strong correlations and low MAEs, the optical gap calculations were much worse and necessitated further inspection. It was found that the sulfur-containing dyes may be the cause of the poor correlation between sTDA-xTB and TD-DFT after using DFTbaby in attempts to improve this correlation. Indeed, removing these structures led to an
improvement in $R^2$ and MAE for the sTDA-xtB properties, however, it is unsure why these structures are problematic.

Overall, the calibration procedure described provides a route to near-DFT quality properties to a satisfactory degree for the systems investigated in this chapter. This method has been used in most of the following chapters. The results and discussion of the study on the dyes discussed in this chapter are continued in the following chapter (chapter 4).
Chapter 4

Finding Diketopyrrolopyrrole Dyes for Dye-Sensitized Water Splitting and Solar Cells

4.1 Introduction

Dyes and pigments have a multitude of applications owing to their interesting and useful chemical properties. These uses are often related to adding colour to other materials in the form of inks, paints or pigments, or harnessing the photoactivity of the dyes to drive photoelectrochemical processes. Organic dyes in dye-sensitized solar cells (DSSCs) are one such use, where the ability of the dyes to absorb light of a particular range of wavelengths ultimately leads to the propagation of an electrical current. Similarly, the dye sensitizer in dye-sensitized photocatalytic water splitting (DS-WS) absorbs light and starts the electron transfer process resulting in the dissociation of water molecules into hydrogen and oxygen gas.

In order to characterize photoactive materials as appropriate for particular applications such as DS-WS, DSSCs or organic photovoltaics (OPVs), the ionization potential (IP), electron affinity (EA) and optical gap ($\Delta_o$) must be known. Comparison of these gaps are visualised in fig. 4.1. There are many computational methods available to obtain these properties, ranging from high-accuracy, expensive methods to cheaper, faster and lower-accuracy methods.

In previous works by our group, a high-throughput virtual screening (HTVS) method was successfully used to produce near-DFT quality results for the properties of large datasets of polymers and small conjugated molecules in a short time frame. The method involves calculating the IP, EA and optical gap of a large library of materials (on the order of $10^4$) using a semi-empirical method (GFN/IPEA/sTDA-)xTB, developed by Grimme et al. and calibrating these results to the DFT-calculated properties of a subset of these molecules using a linear regression model. In the previous chapter, the transferability of this method to a set of conjugated organic dyes was investigated and a range of calibration procedures were discussed. This chapter continues with the study of the dyes using the appropriate calibration method (as determined in chapter 3), probing into the properties of these dyes and considering...
CHAPTER 4. DYES

Figure 4.1: (a) Definition of $-\text{IP}$, $-\text{EA}$, fundamental gap ($\Delta_f$), optical gap ($\Delta_o$) and exciton binding energy (EBE), (b) composition of the different dyes studied in terms of their building block sequence, (c) an example of an ABCBA dye and (d) a selection of the dye building blocks, including the diketopyrrolopyrrole (DPP) core.

Dyes with their suitability for applications such as dye-sensitized photocatalytic water splitting (DS-WS), dye-sensitized solar cells (DSSCs) and organic photovoltaics (OPV).

I focus on three classes of dyes with the sequences ABCBA, ACA and ACA', all possessing a diketopyrrolopyrrole (DPP) core. DPP$^{153,154}$ and its derivatives are well known semi-conducting molecules with a wide range of applications such as in inks, paints, pigments (e.g. car coating Ferrari Red), fluorescent probes and as building blocks in polymers used in field effect transistors and photovoltaics. The attractive features of DPP include good thermal stability and photostability, strong acceptor properties and that it can be easily functionalised to improve solubility (e.g. by introduction of long alkyl chains for better solubility in organic solvents), to provide an anchoring group (e.g. carboxylic acid or phosphonic acid groups) to allow for adsorption of the dye on a substrate, such as a TiO$_2$ photoanode in a DSSC, and to tailor its optoelectronic properties.

Typically, the dyes used in applications such as DSSCs or DS-WS are composed of a conjugated chromophore, donor and acceptor moieties and often, conjugated $\pi$ linkers. A variety of designs are commonly adopted including D-$\pi$-A-$\pi$-D, D-$\pi$-A and D-$\pi$-A-D (where D = donor, A = acceptor and $\pi$ = $\pi$ linker)$^{38,155}$. These designs of dye exhibit an electronic push-pull character, allowing for greater charge separation within the molecule – a property that is essential in the aforementioned applications in order to separate excitons into holes and electrons.
4.2 METHODOLOGY AND COMPUTATIONAL DETAILS

Functionalisation of these structures for the fine-tuning of their properties or the introduction of an anchoring group, for example, in addition to modifying the building block sequence, can rapidly increase the number of structures in a dataset. Using the HTVS approach it becomes possible to investigate the multitude of ways in which substituents and building block sequences can be combined within these dye classes. In the ABCBA class for instance, 179 unique building blocks are used in the A and B positions, in conjunction with the DPP core, to form 32041 unique dyes (15931 for ACA’ and 179 for ACA). The classification and investigation of 10,000+ different structures would be practically impossible experimentally and very computationally expensive using (TD-)DFT. Using the HTVS method described above, the task at hand becomes trivial.

At this stage, I limited the dyes to these three classes, despite asymmetrical dyes proving to be more suitable in many cases, so as to keep the dataset from becoming prohibitively large. For instance, a dye with the sequence ABCDE using the same library of 179 building blocks for the A, B, D and E positions and DPP as the core would increase the dataset to 41,356,876 dyes. Obtaining improved properties for a dataset that large would likely require machine learning and is beyond the scope of this study.

The work in this chapter has been published in:

4.2 Methodology and Computational Details

Three different classes of dyes were investigated, all possessing a diketopyrrolopyrrole (DPP) core. The three different dye sequences (shown in fig. 4.1) were ABCBA, ACA’ and ACA, in which the A and B building blocks were taken from the same library of 179 building blocks and the C building block was DPP. Using a Python pipeline that I wrote, all possible combinations within these sequences were constructed using stk, a Python library that takes functionality from RDKit.

For each structure 30 conformers were randomly generated using the Experimental-Torsion Distance Geometry with additional basic knowledge (ETKDG).126 The lowest energy conformer for each structure was then determined using the Merck Molecular Force Field (MMFF),127 as implemented in RDKit.128 The lowest energy conformers from the MMFF calculations were re-optimised using the tight-binding DFT method GFN-xTB.129 A variant of this method (IPEA-xTB)130 was used to calculate the IP and EA of the dyes. The optical gaps were calculated using the sTDA-xTB method.131 The GFN-xTB and sTDA-xTB calculations were run using the codes xtb132 and stda, respectively. A generalised Born solvation model is available within GFN-xTB and was therefore used for the optimisation and IP/EA calculations, with water and benzene as the chosen solvents. As there is no solvation model implemented for stda, the optical gap calculations were performed without one.
CHAPTER 4. DYES

For the calibration of the xTB-calculated properties of the dyes, the IP, EA and optical gap were calculated using (TD-)DFT. The ground state, anionic and cationic structures were optimised using B3LYP/6-311G**77,82–84 in order to obtain the IP and EA through the \( \Delta \text{SCF} \) method discussed in section 2.4 and chapter 7. The functional/basis set combination \( \omega \text{B97X}/6-311+G^{**}86,87 \) was chosen for the TD-DFT calculations. All DFT calculations were run using Gaussian 16.156

4.3 Results and Discussion

4.3.1 Property-Property Relationships

In fig. 4.2, the 2D histograms depict the predicted property space spanned by the different dye classes dissolved in water for \(-\text{IP} \) and \(-\text{EA} \) (top row), \(-\text{IP} \) and \( \Delta_o \) (centre row) and \(-\text{EA} \) and \( \Delta_o \) (bottom row). The convex hulls enclosing all data points for each of the dye families and property combinations are also shown in fig. 4.3. Where optical gap results are discussed, the dataset is limited to dyes containing only carbon, hydrogen, oxygen and nitrogen atoms (14641 ABCBA, 7260 ACA’ and 121 ACA dyes), in line with the findings of the calibration in chapter 3 (section 3.3.2).

Although quantitative differences exist between the optoelectronic data of these dyes dissolved in water and benzene, qualitatively they look similar (see fig. A.6). Upon observation of figs. 4.2 and 4.3, it is clear that all of the dyes basically have the same energy ranges for \(-\text{IP} \), \(-\text{EA} \) and optical gap. The ABCBA dyes allowed for deeper \(-\text{EA} \) values, shallower \(-\text{IP} \) values and lower optical gap values than the ACA and ACA’ dyes. The convex hulls of the ACA’ and ACA dyes occupied a similar energy range for all of the property combinations, however, the ACA’ dyes’ minimum optical gap reached 0.5 eV lower than ACA. Interestingly, the position of DPP in each of these sections of property space was not within the same regions as any of the dyes. While the \(-\text{IP} \) and \(-\text{EA} \) for the DPP building block lay separately within the \(-\text{IP} \) and \(-\text{EA} \) ranges of the dyes, the combination of these two properties did not. In terms of optical gap, the DPP core lay outside of the optical gap energy space spanned by the dyes. This is unsurprising as the extension of the conjugated \( \pi \)-system should always lead to a lower optical gap.

In general, the properties in figs. 4.2 and 4.3 appeared to be weakly correlation to one another. While the \(-\text{IP} \) v. \(-\text{EA} \) plots appeared to be somewhat positively correlated, where the deeper \(-\text{IP} \) values were usually combined with deeper \(-\text{EA} \) values and \textit{vice versa}, there was no apparent trend in the plots for \(-\text{IP} \) v. \( \Delta_o \). However, the \(-\text{EA} \) v. \( \Delta_o \) histograms and convex hulls indicate that dyes with shallower \(-\text{EA} \) values are more likely to have larger optical gaps than dyes with deep \(-\text{EA} \) values.
Figure 4.2: Two-dimensional histograms of the property spaces spanned by $-\text{IP}$ v. $-\text{EA}$ (top row), $-\text{IP}$ v. $\Delta_o$ (centre row), $-\text{EA}$ v. $\Delta_o$ (bottom row) for the ABCBA, ACA and ACA$'$ dyes (left to right).

### 4.3.2 Property-Property Relationships

Relationships between the properties of the dyes and those of their building blocks

Intuitively, one would assume that the properties of a dye would be correlated with those of its constituent building blocks. Two empirical models were used to explore this idea, based on earlier work in my group,\textsuperscript{157} referred to as the averaging model and the max/min model. In the averaging model, the average $-\text{IP}/-\text{EA}$ between the A and B building block in each dye (e.g. $-\text{IP}_{\text{average}} = (-\text{IP}_A + -\text{IP}_B)/2$) is plotted against the $-\text{IP}/-\text{EA}$ of the dye as determined by the earlier calibration. In the max/min model, the maximum $-\text{IP}/-\text{EA}$ out of the A and B building block for a particular dye is plotted against the dye $-\text{IP}/-\text{EA}$. For both models the calibrated $-\text{IP}$ and $-\text{EA}$ of the dyes are used as the overall dye properties and the DFT-calculated values for the A and B units are used for the building block properties. The
strength of the correlation was assessed by performing a simple linear fit between the results of the models and the dye properties from the calibration and evaluating the $R^2$ values.

In the case of the averaging model, the $-\text{IP}_{\text{average}}$ and $-\text{EA}_{\text{average}}$ values appear reasonably well correlated with the $-\text{IP}$ and $-\text{EA}$ of the dye, albeit with significant discrepancies in absolute values. In the case of the ABCBA dyes, the $R^2$ values of the line of best fit for the $-\text{IP}$ and $-\text{EA}$ plots were 0.65 and 0.46, respectively. For the ACA' dyes, the averaging model also appears to be a reasonable predictor of the dye $-\text{IP}$ and $-\text{EA}$ values, yielding $R^2$ values of 0.70 and 0.63, respectively. In the case of the max/min model, for both the ABCBA and ACA' dyes the model performs similarly to the averaging model with $R^2$ values after a linear fit of 0.70 ($-\text{IP}$) and 0.50 ($-\text{EA}$) for the ABCBA dyes and 0.77 ($-\text{IP}$) and 0.69 ($-\text{EA}$) for the ACA' dyes.

The ACA dyes do not require either of these models as they only have one unique building block.
4.3. RESULTS AND DISCUSSION

Figure 4.5: Property-property relationships between the ACA' dyes and their constituent building blocks as described by the averaging model (top row) and the max/min model (bottom row) using the building block properties (left four panels) and ACA dyes (right four panels) as input. Points for the ACA dyes are shown in black.

block, however, the ACA dye properties v. the properties of the A building block are overlaid on the left panel plots in figs. 4.4 and 4.5. It can be seen that again, the $-\text{IP}$ and $-\text{EA}$ values are correlated reasonably well with those of the A building block with $R^2$ values of 0.72 for $-\text{IP}$ and 0.67 for $-\text{EA}$.

Overall, it is clear that the dyes inherit their electronic character from their constituent A and B building blocks, dominating the effect of the electronic properties of the DPP core. This dominance likely arises from the fact that the $-\text{IP}$ and $-\text{EA}$ values of the DPP core lie near the average of the distributions of the $-\text{IP}$ and $-\text{EA}$ values of the building block dataset from which A and B were taken (fig. 4.6).

Figure 4.6: Kernel density plots and histograms for the DFT-calculated properties of the building blocks with the value for diketopyrrolopyrrole (DPP) for each property indicated by a black dashed line.
Properties of the ABCBA and ACA′ dyes v. the corresponding AACAA and ACA dyes

The models described above were used in a similar fashion to describe the relationship between the dyes in the ABCBA and ACA′ classes and their corresponding dyes with the same building block in every position other than the DPP core. The results of these models are shown in fig. 4.4 and fig. 4.5 in the panels to the right. The right four panels in fig. 4.4 show the relationship between each ABCBA dye’s \(-\text{IP}/-\text{EA}\) and the corresponding AACAA and BBCBB IP/EA of the dye. The averaging model, where the average \(-\text{IP}/-\text{EA}\) of the equivalent AACAA and BBCBB dyes is plotted against that of the ABCBA dye, performs very well with \(R^2\) values of 0.90 and 0.79 for \(-\text{IP}\) and \(-\text{EA}\), respectively - a strong improvement in comparison to the building block-based model. The averaging model of the ACA′ v. ACA/A′CA′ dyes comparison in fig. 4.5 also yielded improved \(R^2\) values of 0.93 and 0.92 for \(-\text{IP}\) and \(-\text{EA}\), respectively. For the max/min model the maximum \(-\text{IP}/\text{minimum }-\text{EA}\) between either the two relevant AACAA and BBCBB or ACA and A′CA′ dyes was taken for the ABCBA and ACA′ dye comparisons, respectively. The model performed slightly worse than the averaging model, yielding \(R^2\) values of 0.86 and 0.68 for the ABCBA dyes’ \(-\text{IP}\) and \(-\text{EA}\), respectively and 0.90 and 0.86 for the ACA′ dyes’ \(-\text{IP}\) and \(-\text{EA}\), respectively.

Evidently, the use of the AACAA and ACA dye properties as input, instead of the building block properties, provides a more robust model for the overall properties of the ABCBA and ACA′ dyes. This was, of course, expected due to the dyes having a greater similarity in electronic properties to the other dyes, rather than their constituent building blocks alone.

4.3.3 Effect of Asymmetry in ACA′ Dyes

I also explored how different building block combinations can lead to either a reduction or increase in the optical gap of the ACA′ dyes with respect to their ACA and A′CA′ dye equivalents. The differences in the optical and fundamental gaps between the ACA′ and ACA/A′CA′ dyes are shown on the left of fig. 4.8. A small but significant proportion of the ACA′ dyes had optical and fundamental gaps lower than their corresponding ACA and A′CA′ dyes. In terms of the optical gap, the largest negative shifts tended to occur for dyes that combine electron-poor and electron-rich A and A′ building blocks.

The two ACA′ dyes with the largest negative shifts in fundamental gap and optical gap are shown in fig. 4.7. While the structure corresponding to the largest negative shift in fundamental gap has fairly similar A and A′ building blocks, the one corresponding to the largest negative shift in optical gap has two very different building blocks on either side of the core, with the tetrazine building block being very electron-poor and the thiophene containing structure being very electron-rich.

As discussed in the introduction to this chapter, many of the dyes for DSSCs and DS-WS in the literature have an asymmetric structure in the form of D-\(\pi\)-A dyes, in order to create
Figure 4.7: ACA’ structures with the smallest fundamental (left) and optical (right) gaps relative to their corresponding ACA and A’CA’ dyes.

a push-pull electronic effect, which facilitates charge separation. The majority of the dyes studied here, however, were in fact symmetrical (i.e. the ABCBA and ACA dyes). This is because the number of possible structures would become prohibitively large within this computational method if the dyes were asymmetrical with the same number of modifiable sites as the ABCBA dyes (e.g. ABCDE dyes). It was computationally tractable however, to investigate the effect of asymmetry on dye properties and the likelihood of them possessing a push-pull nature by studying the ACA’ dyes.

By considering the relative positions of the $-\text{IP}$ and $-\text{EA}$ of the equivalent ACA and A’CA’ dyes it is possible to form an idea of whether the corresponding ACA’ dye is likely to have D-$\pi$-A character. Two alignments of the $-\text{IP}$ and $-\text{EA}$ are possible: the staggered arrangement (fig. 4.9 (left)), in which the $-\text{IP}$ and $-\text{EA}$ of one dye are both lower (or higher) than the $-\text{IP}$ and $-\text{EA}$ of the other, and the straddled arrangement (fig. 4.9 (right)), where the $-\text{IP}$ and $-\text{EA}$ of one dye straddle the $-\text{IP}$ and $-\text{EA}$ of the other.

When, for example, the $-\text{IP}$ and $-\text{EA}$ of the ACA dye straddle those of the A’CA’ dye, according to the max/min model the $-\text{IP}$ and $-\text{EA}$ of the resulting ACA’ dye would be closest to the $-\text{IP}$ and $-\text{EA}$ of the A’CA’ dye. As such, the ACA’ dye inherits character from only the A’CA’ dyes, making it likely to be a D-$\pi$-D or A-$\pi$-A dye. Conversely, if the ACA and A’CA’ $-\text{IP}$ and $-\text{EA}$ are staggered, the highest $-\text{IP}$ and lowest $-\text{EA}$ come from different dyes, likely leading to D-$\pi$-A character.

In fig. 4.8, the ACA’ dyes are plotted according to their differences in the $-\text{IP}$ and $-\text{EA}$ of the corresponding ACA and A’CA’ dyes (denoted $\Delta IP_{AA'}$ and $\Delta EA_{AA'}$, for brevity). The ACA’ dyes with fundamental or optical gaps lower than their analogous ACA and A’CA’ dyes are shown in yellow. The cases in which $\Delta IP_{AA'}$ and $\Delta EA_{AA'}$ are either both positive or both negative correspond to those where the ACA and A’CA’ $-\text{IP}$s and $-\text{EAs}$ are staggered. In the plot showing the fundamental gaps, the yellow points are found only in the quadrants occupied by dyes with staggered $-\text{IP}$s and $-\text{EAs}$ relative to each other, following the expected trend. This trend is predicted slightly worse in the case of the optical gap (fig. 4.8 (d)). This is likely due to the influence of other relevant properties such as the exciton binding energy.
Figure 4.8: (a) Cumulative histogram of ACA′ dyes for which the optical gap (blue) or fundamental gap (orange) is less than that of both related ACA and A′CA′ dyes, negative values, or larger than at least one of the related dyes, positive values), (b) zoomed in version of the cumulative histogram, (c) ACA dyes with an fundamental gap lower (yellow) or larger (blue) than their ACA and A′CA′ counterparts (yellow) as a function of the offset between the −IPs and −EAs of the ACA and A′CA′ dyes, (d) Plot of ACA dyes with an optical gap lower (yellow) or larger (blue) than their ACA and A′CA′ counterparts as a function of the offset between the −IPs and −EAs of the ACA and A′CA′ dyes, (e) example of ACA′ dyes where the ACA/A′CA′ −IP and −EA values are staggered, and (f) example of ACA′ dyes where the ACA/A′CA′ −IP and −EA values are straddled.

4.3.4 Dye Candidates for Dye-Sensitized Water Proton Reduction and Dye-Sensitized Solar Cells

As mentioned in the introduction to this chapter, organic dyes such as these DPP-based ones often find use as dye sensitizers in solar cells or in colloidal photocatalysts. As such, the suitability of these dyes as sensitizers in DS-WS and DSSCs was investigated. In a colloidal photocatalyst combining DPP dyes with TiO$_2$ nanoparticles to drive DS-WS, certain optical and electronic criteria must be fulfilled, in addition to the obvious requirements of stability to water under illumination. Two redox processes must take place for full water dissociation to occur: water oxidation and proton reduction. The following analysis focuses on the electronic ability of the dyes to reduce protons.

The first requirement is that the IP$^*$ of the dye, the potential associated with the ability of the excited dye to donate an electron, should be more negative than the conduction band (CB) of TiO$_2$ to allow for the injection of electrons into the nanoparticles. Secondly, the IP of the dye should be more positive than that of the sacrificial electron donor (SED) ((e.g. ascorbic acid or triethanolamine (TEOA)) or the water oxidation potential. Finally, the optical gap of the dye should be small enough to absorb as much visible light as possible and preferably absorb in the near-infrared region (NIR).
4.3. RESULTS AND DISCUSSION

The description centred around the IP$^*$ of the dye works under the assumption that the injection of electrons into TiO$_2$ by the exciton is faster than the extraction of holes by an SED or water from the dye. If this assumption is incorrect, the discussion would be based on the EA instead of IP$^*$ and EA$^*$ instead of IP. The difference between IP$^*$ and EA and EA$^*$ and IP is the exciton binding energy (EBE), which is likely to be small (0.1 – 0.3 (e)V) in the presence of a high dielectric permittivity solvent like water, based on previous work by our research group$^{24,158,159}$. Therefore, in the following analysis, the EA is used as a proxy for IP$^*$ and the IP as a proxy for EA$^*$.

The conduction band of TiO$_2$ lies at $\sim −0.3$ V$^{160}$ with respect to an SHE ($\sim −4.1$ V with respect to vacuum) and the reduction potential for the one-hole oxidation of ascorbic acid, a commonly used SED, at $\sim 0.7$ V ($\sim −5.1$ V with respect to vacuum) at pH 2.5, the likely pH of an ascorbic acid solution$^{32,161}$. The dye must also be able to absorb visible light and therefore, the optical gap must be small enough to absorb light with energy below 2.95 eV. As the optical gap is relevant to this discussion, only the non-sulfur containing dyes are considered throughout this section. Within the region defined by the above optical and electronic constraints, there were 10941 ABCBA dyes. Of the 121 unique building blocks in the library, 118 were present on the A site and 113 on the B site of the ABCBA dyes present in this region. All of the building blocks on the B sites were present on the A sites. In the case of the ACA$'$ dyes the proton reduction region contained 6527 dyes. Of the 121 building blocks considered, only 4 were not found on the ACA$'$ dyes in this region. The building blocks in this region are explored in greater detail in section 4.3.5.

Supposing an additional offset of 0.4 V was considered, to take into account the kinetically required driving force and/or to correct for the approximation of IP$^*$/EA$^*$ as EA/IP, making the IP and EA thresholds 1.09 V and −0.70 V, respectively, the dye class is narrowed down to 849 non-sulfur containing ABCBA dyes in total. Of these, the maximum optical gap is 2.77 eV and the minimum 2.51 eV – higher than what would be preferable for adequate visible light absorption. Within this set of dyes there were 82 A building blocks and 44 B building blocks. For the ACA$'$ dye class, the number of dyes under these constraints was 1518, with the maximum and minimum optical gaps being 2.81 eV and 2.44 eV, respectively. In terms of thermodynamics, it is much more difficult to oxidise water rather than an SED.
such as ascorbic acid or TEOA, requiring an IP at pH 7 larger than 1.12 V, when including the additional driving force (0.82 V without), meaning that even fewer dyes can drive this half-reaction.

The suitability of the ABCBA and ACA’ dyes in TiO$_2$-based DSSCs can similarly be considered. Solvents used in DSSCs, such as acetonitrile, are not that dissimilar to water, given the inverse dependence of the electronic properties on the dielectric permittivity. It is therefore justifiable to use the screening results in water to investigate the suitability of these dyes for use in DSSCs. In acetonitrile solutions containing Li$^+$ cations, Redmond and Fitzmaurice$^{162}$ report that the conduction band edge of TiO$_2$ lies at $\sim -0.7$ V vs. SHE ($-3.2$ V with respect to vacuum). Boschloo and Hagfeldt$^{163}$ report that the standard reduction potential of the I$_3^-$/I$^-\text{redox}$ couple lies at $\sim +0.3$ V vs. SHE ($-4.7$ V with respect to vacuum).

These values were used to define the boundaries of the electronic requirements for dyes in TiO$_2$-based DSSCs. Within these boundaries lay 8952 ABCBA dyes, with 119 unique building blocks out of the library of 121 on the A site and 107 on the B site. Of these, the maximum and minimum optical gaps were 2.84 eV and 2.2 eV – higher than the ideal range for adequate absorption of visible light. For the ACA’ dye class, 519 dyes lay in this region, containing 119 unique building blocks, with maximum and minimum optical gap values of 2.98 and 2.00 eV, respectively.

By consideration again of an offset of 0.4 V, the number of dyes in the TiO$_2$-based DSSC region reduced to 2303 (105 unique building blocks on the A site and 50 on the B site) and 803 (97 unique building blocks) dyes in the ABCBA and ACA’ dye classes, respectively.

As discussed in section 4.1, in addition to the optical, electronic and kinetic properties accounted for above, it is necessary for a dye to have an anchoring group through which to adsorb the dye onto the semiconductor surface. While this effect has not been included in the dye screening, a small selection of each dye class (3 ABCBA, 2 ACA and 2 ACA’) was used to demonstrate the electronically benign nature of typical anchoring groups. The results of the IP, EA and optical gap calculations with anchoring groups in the para position are shown in table A.1 and compared to the analogous dyes without any substituents present. As the ACA’ dyes are asymmetrical, the effect of the anchoring group when attached to either side was calculated for this dye class, resulting in 4 different values.

4.3.5 Structure-Property Relationships

The identification of the dominant building blocks in particular regions of chemical space can give an idea of what combinations of chemical moieties lead to certain dye properties. Having sectioned the plot of the $-\text{IP}/-\text{EA}$ slice of property space of the ABCBA dyes (see fig. 4.10) and determined which building blocks occurred most often on the A and B sites, I focus below on a few areas of interest: the extremities of the plot – the top right and bottom left corners, the middle sections and also the region that is relevant to the application of dyes in DS-WS.
Figure 4.10: Scatter plot showing the $-\text{IP}/-\text{EA}$ slice of property space for the most common A and B units of the ABCBA dyes, when residing on the B-site. The building blocks shown are the most common A and B building blocks in (a) the top corner ($\text{IP} \geq -4.8$, $\text{EA} \geq -3.3$) (b)-(c) the bottom corner ($\text{IP} \leq -6$, $\text{EA} \leq -4.3$) and (d)-(f) the middle section of the grid ($\text{IP} = -6$, $\text{EA} = -4.3$, $-3.3$). The colours of the building blocks correspond to the positions in the scatter plot of the dyes they reside in.

Top right corner

The top right section of the grid contained dyes with $-\text{IP}$ more positive or equal to $-4.8$ eV and $-\text{EA}$ more positive than or equal to $-3.3$ eV. The fused-pyrrole building block shown in fig. 4.10 (a) (see also Table 3) was the most common building block on both the A and B site in this region containing 1529 dyes, occurring 73 times on the A site and 162 times on the B site, meaning that 41% and 91% of the dyes with this building block on the A or B site, respectively, lay in this region. The implication of this is that the building block on the B site has greater control of the dye’s overall properties. This observation is even more apparent upon inspection of the convex hulls of this building block on the A site (yellow) and on the B site (blue), shown in fig. A.8, in which the points in blue are more localised in the top right corner than the yellow points. This feature is even more pronounced in other regions of the $-\text{IP}/-\text{EA}$ slice of property space, discussed further below. The very electron-rich character of this building block is in line with the shallow $-\text{IP}$ and $-\text{EA}$ of the resulting dyes.
CHAPTER 4. DYES

Bottom left corner

The limits for this region were $-\text{IP}$ more negative or equal to $-6$ eV and $-\text{EA}$ more negative or equal to $-4.3$ eV. In this region the most dominant building block on the A and B sites differed, as shown in fig. 4.11. The most common A building block is shown in fig. 4.10(b) and occurs 10 times on the A site and 33 times on the B site. The building block shown in fig. 4.10(c) occurs most often on the B site in this region. Again the latter occurred more often on the B site (34, 19%) than on the A site (8, 0.4%). The lower percentages of dyes with these building blocks occurring in this region relative to the case for the top right corner is in part the result of the fact that the bottom left corner region contained only 129 dyes in total. The very electron-poor nature of these building blocks is in line with the deep $-\text{IP}$ and $-\text{EA}$ of the resulting dyes.

Middle region

For the middle section, two slices of the $-\text{IP}$ range were taken, $-5.3$ to $-5.8$ eV and $-6.0$ to $-5.3$ eV (see Table 3). The $-\text{EA}$ limits for this section were $-3.3$ and $-4.3$ eV. The most commonly occurring building blocks are shown above in fig. 4.10(d)-(f). For the first slice, containing 9926 dyes, the most common A and B, fig. 4.10(d) occurred 140 times (84%) on the A site and 169 times (93%) on the B site, and 147 times (82%) on the A site and 170 (95%) on the B site, respectively. The convex hull enclosing the data point of the dyes containing this building block on the A site (yellow) and the B site (blue) is shown in the top left panel of fig. 4.12. In this plot, the blue convex hull clearly spans a smaller region of the $-\text{IP}/-\text{EA}$ property space, again supporting the notion that the B site building block is more influential over the overall ABCBA dye properties.

6485 dyes resided within the second $-\text{IP}$ slice, of which the most common building block on the A site is shown in fig. 4.10(e), which occurred 89 times (50%) on the A site and 124 times (69%) on the B site. The most common B building block shown in fig. 4.10(f) occurred 138 times (77%) on the B site and 87 times (49%) on the A site. As can be seen from the percentages most of the dyes containing these building blocks were found within this region. However, this is to be expected since such a large proportion of all the ABCBA dyes were contained in this region. The nature of the building blocks themselves, less electron-rich than the most common building blocks for the top right corner and less electron-poor than the most common building blocks in the bottom left corner, is in line with the properties of the resulting dyes.

Dye-sensitized proton-reduction region

I also focused in on the $-\text{IP}$ and $-\text{EA}$ region relevant to dye-sensitized proton reduction using the experimental TiO$_2$ conduction band edge and solution potential values discussed above. These values translate into electronic constraints of $-\text{IP}$ more negative than or equal to $-5.13$
4.3. RESULTS AND DISCUSSION

<table>
<thead>
<tr>
<th>−IP/−EA limits (eV)</th>
<th>Dominant A building block</th>
<th>Dominant B building block</th>
</tr>
</thead>
<tbody>
<tr>
<td>Top corner</td>
<td><img src="imageA" alt="A building block" /></td>
<td><img src="imageB" alt="B building block" /></td>
</tr>
<tr>
<td>−IP &gt;= −4.8;</td>
<td></td>
<td></td>
</tr>
<tr>
<td>−EA &gt;= −3.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(a)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| Bottom corner       | ![A building block](imageC) | ![B building block](imageD) |
| −IP < −6.0;         |                           |                           |
| −EA < −4.3          |                           |                           |
| (b)                 |                           |                           |

| Middle region       | ![A building block](imageE) | ![B building block](imageF) |
| −IP: −5.3, −4.8;     |                           |                           |
| −EA: −4.3, −3.3      |                           |                           |
| (d)                 |                           |                           |

| Middle region       |                           |                           |
| −IP: −6, −5.3;       |                           |                           |
| −EA: −4.3, −3.3      |                           |                           |
| (e)                 |                           |                           |

Figure 4.11: Most common building blocks on the A and B site within the regions of the −IP/−EA slice of property space of the ABCBA dyes discussed in the text.

eV and −EA more positive than −4.14 eV for dye-sensitized proton reduction.

I have, as before, limited the discussion to dyes containing only carbon, hydrogen, oxygen and nitrogen atoms, as the optical gap is relevant to these applications. Within this dye-sensitized proton reduction region lay 10941 non-sulfur containing dyes (74% of the total number of non-sulfur containing ABCBA dyes), with 118 unique building blocks in the A position and 113 in the B position. There were 6 dominant A building blocks and 5 dominant B building blocks (see fig. 4.13) each occurring 105 and 116 times, respectively, out of the 121 unique non-sulfur containing building blocks in this truncated dataset. Two of these building blocks were both dominant on the A site and on the B site. In line with previous observations, the most common building blocks for either the A or the B site occurred a larger number of times on the B site — this was observed for every building block in fig. 4.13. For example, building block (vi) in fig. 4.13(a) occurred 105 times on the A site and 111 times on the B site. Building block (iii) in fig. 4.13(b) occurred 103 times on the A site and 116 times on the B site. Even where the frequency of occurrence on the A site and B site were closest, for building block (iii) in fig. 4.13(a) (105 times on the A site and 109 times on the B site),
Figure 4.12: Convex hulls for the building block in section (d) in fig. 4.11 (top left), building block i in fig. 4.13(a) and (b) (top right) (blue = occurrence on B site, yellow = occurrence on A site, red = entire ABCBA set) and the two most common building blocks in the proton reduction region for the ACA$'$ dyes (blue = benzene, yellow = 1,6-naphthyridine, green = entire ACA$'$ set). The unshaded regions on the top right and bottom depict the $-IP$ and $-EA$ limits where proton reduction is thermodynamically feasible.

Qualitatively it can be seen that the convex hulls for the B site span a more compact region of the plot (see fig. A.9). Structurally the most common building blocks in this region overlap with the most common building blocks found for the middle region (see fig. 4.13) discussed above – an unsurprising observation, as the regions themselves overlap. In line with what was observed for the middle region, at least one pyridinic nitrogen was present in these building blocks in almost all cases, making these structures all relatively electron-poor. In the case of ACA$'$ dyes, the water splitting region contained 6526 dyes (90% of the total number of ACA$'$ non-sulfur containing dyes). Of the 121 building blocks considered, only 3 of these were not found in the ACA$'$ dyes in this region. The two most dominant building blocks for ACA$'$ dyes in the water splitting region were benzene and 1,6-naphthyridine. The convex hulls for ACA$'$ dyes based on these two building blocks are also shown in fig. 4.12 (labelled as A for benzene and A$'$ for 1,6-naphthyridine) and have a similar spread of data points.

In general, there appears to be a trend in which the B site on the ABCBA dyes has a greater influence on the overall properties of the dye. This can be inferred from firstly, the convex hull plots, in which most of the time the A (yellow) convex hull spans a larger area than the B (blue) convex hull, secondly, from the frequency of occurrence of a particular building block within the highlighted regions and finally, from the comparison of the scatter plots of
4.3. RESULTS AND DISCUSSION

(a) Most common A building blocks  (b) Most common B building blocks

Figure 4.13: Most common building blocks on (a) the A and (b) B sites for the proton reduction region.

all highest occurring building blocks per grid section on the A site and the B site. In the scatter plot showing the most common building blocks per grid section on the B site (shown in fig. 4.10), the points belonging to dyes with a particular building block on the B site were more localised, resulting in well-defined segments for each building block. Conversely, the same plot considering dyes with those same building blocks on the A site had more widely spread data points, creating a less segmented effect (see fig. A.10).

Dye-sensitized solar cell region

For the dyes to be suitable for TiO$_2$-based dye-sensitized solar cells (DSSCs) they must possess $-\text{IP}$ more negative than or equal to $-4.74$ eV and $-\text{EA}$ more positive than $-3.74$ eV. Within the DSSC region lay 8952 ABCBA non-sulfur containing dyes, with 119 and 107 unique building blocks on the A site and B site, respectively. While there was one dominant building block on the A site, pyrrole, there were four dominant building blocks on the B site (see fig. 4.15). Pyrrole occurred 95 times on the A site and 99 times on the B site. Each of the dominant building blocks on the B sites occurred 116 times on the B sites and, in line with the observations discussed above, in all cases fewer times on the A site. In terms of ACA’ dyes, there were 5812 non-sulfur containing dyes in the DSSC region, containing a total of 119 unique building blocks out of the 121 possible non-sulfur containing building blocks. The two most common building blocks were 9H-carbazole and difuro[3,2-b:2’,3’-d]furan, the convex hulls of which are shown in fig. 4.15 in yellow and blue, respectively, occupying very similar regions of chemical space.

Dye candidates for organic photovoltaics

So far, only the results of the screening in water have been discussed. Aside from their potential suitability as dye sensitizers in DS-WS and DSSCs, these dyes are similar in structure to some of the organic molecules used as donor molecules in organic photovoltaics (OPVs). For the
CHAPTER 4. DYES

Figure 4.14: Most common building blocks on B sites for the ABCBA dyes in the dye-sensitized solar cell region

Figure 4.15: Convex hulls for selected building blocks in the dye-sensitized solar cell region. Building block (i) in fig. 4.14 occurring on the A (yellow) and B (blue) sites of ABCBA dyes in this region (left); most common building blocks in the ACA’ non-sulfur containing dyes in the DSSC region (yellow = 9H-carbozole, blue = difuro[3,2-b:2’,3’-d]furan) (right). The unshaded regions depict the region of the \( -\text{IP} \) vs. \( -\text{EA} \) property space where charge injection from the dye into the \( \text{TiO}_2 \) and oxidation of \( \text{I}^- \) to \( \text{I}^{3^-} \) is possible.

analysis of the suitability of the dyes as OPV donor materials, an organic dielectric medium is more relevant, so for this brief investigation I focus on the results in benzene.

For a donor material to be suitable for OPV, its \( -\text{IP} \) and \( -\text{EA} \) must be staggered with respect to the acceptor molecule. Therefore, I have chosen three known OPV acceptor molecules for reference: PCBM, PC\textsubscript{70}BM and PN1. The first two are solubilised fullerene molecules, which as mentioned in section 1.2.2 are commonplace as OPV acceptors.\textsuperscript{56} PN1 was also mentioned in the introduction and is a polymer that has recently been used alongside other organic polymers to yield high PCEs.\textsuperscript{60} The reference \( -\text{IP} \) and \( -\text{EA} \) values were \(-5.85\) eV and \(-3.89\) eV for PCBM,\textsuperscript{164} \(-5.96\) eV and \(-3.9\) eV for PC\textsubscript{70}BM\textsuperscript{164} and \(-5.71\) eV and \(-3.85\) eV for PN1.\textsuperscript{56} In general, a donor material with a LUMO at least 0.3 eV higher than that of the acceptor is chosen so as to provide a driving force for the separation of an exciton. This additional 0.3 eV has also been considered in the following analyses. Table 4.1 shows the number of dyes belonging to each dye class that are good candidates as OPV donors against the three chosen acceptors under the electronic conditions, the electronic conditions and a minimum driving force, and the optoelectronic conditions including a minimum driving force.

As in DSSCs and DS-WS, the range of energies that can be absorbed should ideally be as
large as possible, utilising as much visible light and preferable even near-IR (NIR) radiation as possible. As such, it is desirable to fine-tune the properties of a material so as to minimise the optical gap. A commonly used OPV donor material is P3HT, which is based on polythiophene and has an optical gap of $\sim$1.9 eV.\textsuperscript{165} For these reasons, only dyes with optical gaps between 1.1 and 2.2 eV are considered to be appropriate donor materials for this application. This means that once again this discussion has to be limited to simply non-sulfur containing structures due to the issues encountered with either the code \texttt{stda} or the \texttt{sTDA} method.

Table 4.1: Number of OPV polymer candidates for each co-polymer type relative to the experimentally determined potentials of the three acceptors PCBM, PC\textsubscript{70}BM and PN1.

<table>
<thead>
<tr>
<th>Acceptor</th>
<th>PCBM</th>
<th>PC\textsubscript{70}BM</th>
<th>PN1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>10857 (49.3)</td>
<td>14508 (65.9)</td>
<td>6664 (30.3)</td>
</tr>
<tr>
<td>No. dyes under IP/EA constraints (%) of total dyes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ABCBA</td>
<td>8356 (57.1)</td>
<td>10590 (72.3)</td>
<td>5514 (37.7)</td>
</tr>
<tr>
<td>ACA'</td>
<td>2501 (34.4)</td>
<td>3918 (54.0)</td>
<td>1150 (15.8)</td>
</tr>
<tr>
<td>ACA</td>
<td>38 (31.4)</td>
<td>56 (46.3)</td>
<td>23 (19.0)</td>
</tr>
<tr>
<td>Total</td>
<td>10713 (48.6)</td>
<td>14231 (64.6)</td>
<td>6617 (30.0)</td>
</tr>
<tr>
<td>No. dyes under IP/EA constraints + minimum driving force (%) of total dyes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ABCBA</td>
<td>8212 (56.1)</td>
<td>10313 (70.4)</td>
<td>5467 (37.3)</td>
</tr>
<tr>
<td>ACA'</td>
<td>2501 (34.4)</td>
<td>2918 (40.2)</td>
<td>1150 (15.8)</td>
</tr>
<tr>
<td>ACA</td>
<td>28 (23.1)</td>
<td>56 (46.3)</td>
<td>23 (19.0)</td>
</tr>
<tr>
<td>Total</td>
<td>741 (3.4)</td>
<td>928 (4.2)</td>
<td>501 (2.3)</td>
</tr>
<tr>
<td>No. dyes under electronic and optical gap constraints + minimum driving force (%) of total dyes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ABCBA</td>
<td>656 (4.5)</td>
<td>748 (5.1)</td>
<td>467 (3.2)</td>
</tr>
<tr>
<td>ACA'</td>
<td>85 (1.2)</td>
<td>180 (2.5)</td>
<td>34 (0.5)</td>
</tr>
<tr>
<td>ACA</td>
<td>1 (0.8)</td>
<td>2 (1.7)</td>
<td>0 (0.0)</td>
</tr>
</tbody>
</table>

Overall, a greater proportion of the ABCBA dyes were within the appropriate region for acting as an OPV donor material in comparison to the other two dye classes. However, this was still a very low proportion of the overall dyes in the class once the optical, electronic and driving force constraints were applied, with only 3.4%, 4.2% and 2.3% of the dyes being suitable as donors with respect to PCBM, PC\textsubscript{70}BM and PN1, respectively. For the ABCBA dye class, the same trend as before of higher occurrences of the B building block than the A building block per region was observed; e.g. the highest occurring building block within the
PCBM boundaries was naphtho[2,3-c][1,2,5]oxadiazole, occurring 89 times on the B site and 19 on the A site.

The full list of dyes appropriate for OPV with respect to each set of constraints is available online at:
https://github.com/isabelleapo/dyes_OPV

4.4 Conclusions

In this work, I have demonstrated the transferability of our previously established (sTDA/IPEA-)xTB-based high-throughput virtual screening approach to these dye structures. Although the calibration performed well in general, the problems encountered with the sulfur-containing dyes unfortunately limited the analysis. The property results post-calibration showed that the ABCBA, ACA and ACA’ families of dyes occupied the same region of property space, although the ABCBA dyes occupied a larger part of it. The properties of the dyes were found to be correlated with the properties of their constituent building blocks, using the max/min and averaging models. In the case of the ABCBA and ACA’ dyes, their properties were similarly correlated to the equivalent AACAA/BBCBBA and ACA/A’CA’ dyes. Similarly, I also demonstrated the successful prediction of which ACA’ dyes are likely to have smaller gaps than their ACA/ACA’ counterparts using the method originally developed for co-polymers, which takes the $-\text{IP}$ and $-\text{EA}$ of the ACA/A’CA’. It was found that a fraction of ACA’ dyes have an optical and/or fundamental gap that is smaller than their ACA/A’CA’ counterparts and that those dyes tend to have D-$\pi$-A character.

The analysis of the building blocks that most commonly occurred in dyes in particular areas of property space revealed that the deepest $-\text{IP}$s and $-\text{EAs}$ can be accessed by including electron-poor building blocks in dyes, with the opposite being true for shallow $-\text{IP}$s and $-\text{EAs}$, as expected. For most of the building blocks, the region of property space spanned by ABCBA dyes when a particular building block was on the A site was consistently larger, with the points more dispersed in property space. Conversely, the region of property space spanned by these dyes with a particular building block on the B site was consistently more compact, with the data points being more localised to a specific area in the $-\text{IP}/-\text{EA}$ property slice. Additionally, there were more unique A monomers than B monomers per region of property space and therefore, greater variety in the A position.

This indicates that the building block in the B position is more influential in terms of the overall dye properties and where they reside in property space than the building block on the A site. I speculate that this can be attributed to the closer proximity of building block B to the core. These observations could be particularly useful in the design of dyes for a particular purpose, as knowledge of how the overall properties are influenced by a particular building block’s position relative to the core could allow for judicious fine-tuning of the properties as desired.
4.4. CONCLUSIONS

In the region of property space where proton reduction in the presence of ascorbic acid as a sacrificial electron donor is optically and thermodynamically possible, more than half of all ABCBA and ACA' dyes were found (not including the sulfur-containing structures). The most common building blocks on the B site for ABCBA dyes in this region all possessed at least one pyridinic nitrogen, as did most of the most common building blocks on the A site. The suitability of the dyes as sensitizers in dye-sensitised solar cells (DSSCs) was also explored. Again, more than half of the non-sulfur containing dyes were found suitable to be suitable for both the ABCBA and ACA' dye classes. Finally, the ABCBA dyes were screened for their suitability as donors in organic photovoltaics with reference to three different acceptor materials: PCBM, PC$_{70}$BM and PN1. Under the electronic, optical and kinetic constraints, only a small percentage of the dyes in any of the classes were found to be suitable as OPV donors.

The main issue with the analyses carried out in terms of specific applications was the inability to use the sulfur-containing structures. The removal of these dyes likely led to a lot of appropriate candidates for each of the applications being ignored. The thiophene units in a lot of the building blocks in the original library would likely lead to shallower $-$IPs, which could lead to a reduction in fundamental and optical gaps. Materials based on polythiophene are often used in such applications and therefore, this investigation is weakened by the exclusion of such building blocks. The limitations in computational resources also meant that only one of the orientations in the asymmetrical building blocks was considered and this could have been worth exploring.

The work in this chapter does, however, highlight some interesting points that could be considered in future work related to similar dyes. Firstly, the demonstrated correlation between the heterodye properties and their corresponding homodyes opens up the possibility for a pre-screening phase when repeating this method. For example, the calculation of the properties of 179 homodyes is a trivial task, easily tractable with DFT and could provide insight into the resulting heterodyes without having to run an entire high-throughput screening. Following this, a full screening could be performed, guided by the results of this pre-screening. Secondly, the apparent stronger influence of the building block in the B position may be something to consider in future attempts to fine-tune the properties of ABCBA dyes.

Additionally, in future studies, the building block library could be modified to include variants of the best building blocks in each section, depending on which applications are being researched. The analysis of the ACA' dye properties revealed that asymmetry in the building block sequence can give rise to a number of interesting properties, which is certainly worth further exploration. It would be ideal to extend this study to a larger library of asymmetrical dyes, beyond the relatively simple ACA' dye class, attempting, for example, a building block sequence such as ABCA, if possible with the computational resources available. Overall, despite some of the difficulties encountered throughout this work, some useful insights can be drawn that could improve any future related research by both experimental and computational
organic chemists.
Chapter 5

Exploring the Effect of Composition, Sequence Isomerism and Conformational Degrees of Freedom with High-Throughput Virtual Screening

5.1 Introduction

As discussed in the introduction to this thesis, computationally driven materials discovery has a multitude of advantages over experimental design. Some of these include greater efficiency in compound screening as the need to synthesize the materials is eliminated, the ability to study chemical structures that are difficult to synthesize and the potential to discover new materials for particular applications by expanding the search beyond familiar design routes. High-throughput virtual screening (HTVS) has been used in numerous areas of materials discovery, including in previous work by my group on organic polymers\textsuperscript{152} and diketopyrrolopyrrole-based dyes, allowing for the efficient exploration of chemical space. In this chapter, the previous work on polymers is extended with a focus on conformer and sequence isomerism.

One of the many attractive features of polymers is the tunability of the properties via co-polymerisation however, the number of polymers available is greatly increased when considering all of the possible monomer combinations that have been or have yet to be found. Aside from all of the structures that arise from the vast array of possible monomer combinations, the library of possible polymers is further expanded when sequence isomerism and conformerism are taken into account. For instance, for just two unique monomers at a 1:1 ratio in a dodecamer, the number of distinct sequences is 472. Both conformerism and sequence isomerism are features that are interesting yet much more difficult to explore with experimental HTS than with HTVS. In the case of sequence isomerism, depending on the synthetic route taken, an ordered or disordered structure may be synthesized in experiment and thorough analysis is required to ensure that the synthesis of the desired monomer sequence has actually been achieved.

Conformerism is also difficult to probe as often, the studies carried out experimentally will
be limited to the most stable conformers. With the HTVS method used here, calculations on many different polymer conformations are possible and can give us insight into how different conformers affect overall polymer properties.

In this chapter, the HTVS method that has been previously used by our group\textsuperscript{152} was implemented for the exploration of the effect of sequence and conformerism on the overall properties of conjugated co-polymers composed of a small set of monomers. The work in this chapter aimed to firstly, investigate the chemical space occupied by a large amount of conformers of a polymer and secondly, to investigate the properties of all possible sequences – ordered and disordered – of a set of conjugated organic oligomers. Three different co-polymer compositions, based on benzene, pyridine and thiophene monomers, were explored at different monomer ratios for all possible sequences.

The work in this chapter has been published in:


5.2 Methodology and Computational Details

The workflow I followed involved a series of steps and made use of the semi-empirical method described in Chapter 2, GFN-xTB by Grimme et al.,\textsuperscript{129} for the calculation of the properties calibrated against DFT data. The polymeric system was modeled as a single-polymer strand in a water solvent; a model which was similarly employed in previous work by our group with successful results.\textsuperscript{152}

The workflow, as shown in fig. 5.1, started with a simplified molecular-input line-entry system (SMILES)\textsuperscript{123} representation of the chosen monomer units and then the polymers built using \texttt{stk},\textsuperscript{133,134} a Python library that takes functionality from RDKit.\textsuperscript{128} The polymer chain lengths were restricted to oligomers composed of 12 aromatic rings, as it has been shown in previous work by our group that it is at this oligomer length that the properties converge. As such, this oligomer length provides a reasonable model for a full-length polymer.\textsuperscript{24}

The conformer search in each case was performed by firstly randomly generating 500 polymers using the Experimental-Torsion Distance Geometry with additional basic knowledge (ETKDG) method.\textsuperscript{126} Out of these the lowest energy structure was found via an optimisation and energy-ranking procedure using the Merck Molecular Force Field (MMFF),\textsuperscript{127} as implemented in RDKit. The properties calculated were IP, EA and optical gap. The former two were calculated using an extension of GFN-xTB parametrised for the calculation of IP and EA – namely IPEA-xTB\textsuperscript{130} – using the code \texttt{xtb}.\textsuperscript{132} The optical gaps were calculated using the simplified Tamm-Dancoff approach (sTDA) applied to the wavefunctions obtained through \texttt{xtb} (sTDA-\texttt{xtb})\textsuperscript{116} and this was done using the \texttt{stda} code. The generalised Born solvation model was used for the optimisation and IP/EA calculations, with water as the
5.3. RESULTS AND DISCUSSION

chosen solvent. As there is no solvation model implemented for stda, the optical gap calculations were performed without one. Based on evidence from previous work by our group of the strong linear correlation between (TD-)DFT and (GFN/sTDA)-xTB derived IP, EA and optical gap value, a simple linear model was fitted to translate the xTB results such that they correspond as closely as possible to the previous (TD-)DFT based approach. As such, the lack of solvent model in stda was corrected for within the methodology.

As the aim of this study was to determine the effect of sequence, all possible permutations of each set of monomers within 5 co-monomer ratios were generated using a simple Python script that I wrote that made use of combinatorics. For each of the three monomer compositions explored (phenylene-thiophene, phenylene-pyridine and pyridine-thiophene), all possible co-polymer sequences were generated for the ratios 1:3, 1:2, 1:1, 2:1 and 3:1. An additional study on the variation of properties for the 500 randomly generated conformers was carried out by my colleague, the results of which will also be discussed below as they complement the work I did.

5.3 Results and Discussion

5.3.1 Conformational Degrees of Freedom

Prior to running the xTB calculations on the chosen oligomers, the effect of conformerism on the optoelectronic properties was investigated by Liam Wilbraham. In this investigation the
properties were calculated for 500 randomly generated conformers of oligomers of length 12 based on four homopolymers and three co-polymers shown in fig. 5.2. After the optimisation of each conformer using GFN-xTB, the IP, EA and optical gap were calculated using IP/EA-xTB and sTDA-xTB, respectively.

Figure 5.2 shows the variation in properties with Boltzmann factor of each conformer, where the lowest energy conformers have Boltzmann factors close to one. This plot shows that the maximal variation in properties is on the order of 0.1 (e)V and even lower for the lowest energy conformers. Thus, I opted for a minimal conformer search in the following calculations due to the minimal effect of conformerism on the properties.

Figure 5.2: Calibrated IP, EA and optical gap v. Boltzmann factor for the 500 conformers of the polymers of the repeat units shown on the right. Coloured chemical structures indicate data shown in the same colour. The lowest energy conformers are those with Boltzmann factors close to 1.
5.3. RESULTS AND DISCUSSION

5.3.2 Effect of Co-polymer Sequence

This study was primarily aimed at assessing the effect of sequence isomerism in co-polymers. In general, full control over polymer sequence during synthesis is often not possible and typically the structure is either fully disordered or well-defined. As such, it is desirable to elucidate whether the details of the sequence are in fact vital to the overall polymer properties. There are many possible unique orderings of monomers in a co-polymer of fixed composition e.g. the alternating (AB)_n and block A_nB_n sequence isomers. For instance, for a phenylene-thiophene co-polymer at a 1:1 ratio there exist 472 unique sequences. Although the prediction of the properties of different sequence isomers would offer useful insights and is therefore an attractive prospect, it is a computationally demanding task. It is, therefore, not possible to sample 3 co-polymers at 5 different composition ratios for all possible unique sequence without an efficient HTVS approach like the one used here.

![Graphs showing distributions of properties (IP, EA and optical gap) of disordered co-oligomers with varying co-monomer ratios.](image)

Figure 5.3: Distributions of properties (IP, EA and optical gap) of disordered (a) phenylene–thiophene, (b) phenylene–pyridine and (c) pyridine–thiophene co-oligomers varying with differing co-monomer ratios. For each ratio, the properties of all possible monomer sequences were calculated, using a fixed oligomer length of 12 monomer units in total.

In fig. 5.3, the variation of properties over different monomer concentration ratios is depicted, as well as the spread in properties at each concentration for all possible sequences. Focusing on the median values of each of the properties, it can be seen that for the phenylene-thiophene co-oligomers, increased concentration of thiophene led to a shallower IP, a deeper
EA and a decrease in the optical gap.

For the phenylene-pyridine co-oligomers, increased pyridine content was predicted to result in a deeper median IP and EA, and a decrease in the median optical gap. Finally, for the thiophene-pyridine co-oligomer, increasing the pyridine content again led to a deeper median IP and EA, however, the change was much more dramatic in the case of the IP. In these co-oligomers, an increase in the median optical gap was observed upon increasing the pyridine concentration.

The change in predicted median properties was not constant between compositions. For example, the difference between the maximum and minimum median IP for the phenylene-thiophene co-oligomers was $\sim 0.5$ V, whereas for the phenylene-pyridine co-oligomers it was less than $\sim 0.25$ V. The range in optical gap for the former composition was $\sim 1$ eV and less than 0.5 eV for the latter. It is apparent that the strength of these changes is related to the difference between the properties of the corresponding homopolymers. In fig. 5.3(a), the greater difference in homopolymer potentials and optical gaps compared to those in (b) can be seen. The same trend is observed for the pyridine-thiophene co-oligomers, which exhibit a much stronger variation in properties than the phenylene-pyridine co-oligomers. This result seems fairly intuitive as one would expect the properties of a co-oligomer to be an intermediate between its composite homopolymers.

In the violin plots the length of each violin is indicative of the value range for that composition ratio and the width of each violin is indicative of the frequency of a particular value i.e. the highest frequency of values is at the widest point of the violin. These attributes of the violins can allow one to evaluate the sequence dependence of each property.

For all three compositions, it appeared that the EA was the property least affected by changes in sequence. The maximum range in values was less than 0.5 V for the phenylene-pyridine co-oligomers and the minimum range was for the pyridine-thiophene co-oligomers. This suggests a low dependence of the EA on the ordering of the monomers. Larger ranges of values were calculated for the IPs than for the EAs, but still not exceeding 0.5 V in the case of the phenylene-thiophene co-oligomers where the violins were longest. It seems as though for the IP, the presence of thiophene leads to a greater sequence dependence.

In terms of the optical gaps, the thiophene-containing co-oligomers exhibited the greatest variation in values with sequence, as seen by the longer violins in fig. 5.3. This means that the range in the optical gap as a result of a change in the distribution of the monomers in the sequence was larger for the thiophene-containing co-oligomers. Additionally, the range of optical gap values reduced as the thiophene content was increased.

Interestingly, this implies that again, the greater dissimilarity in the optical properties of the co-oligomers’ corresponding homo-oligomers led to the optical gaps being affected to a greater extent when the ordering of the monomers in the sequence was varied. The phenylene-pyridine co-oligomer optical gaps varied by less than 0.5 eV at most with sequence changes.

Looking at the optical gap plots in fig. 5.3(a) and (c), it is also clear that at lower thio-
phenyl or pyridine (and potentially other equally or more dissimilar monomers) is more disruptive to the optical properties than a few phenylene or pyridine monomers would be in a polymer containing mostly thiophene monomers. Looking at the 25% thiophene data for the phenylene-thiophene co-oligomer in particular (where the co-oligomer contained 3 thiophene monomers and 9 benzene monomers), it is evident that varying the positions of the three thiophenes in the dodecamer did in fact make a significant difference to the calculated properties.

Within this study the question of how the degree of segregation between identical and non-identical monomers affects the overall polymer properties was also explored. Figure 5.4 shows how the properties of a subset of the polymers (specifically, the 1:1 polymers) vary with this parameter. The degree of segregation \(D\) was described using the formula shown in fig. 5.4, where \(n\) is the number of non-identical neighbours and \(N\) is the length of the polymer. The formula yields a value between 0 and 1 corresponding respectively to a fully alternating polymer with no identical neighbours and a polymer with only one non-identical neighbour i.e. fully segregated. Nearing full segregation \((D \rightarrow 0)\) the properties of the co-oligomers lay closest to those of the homopolymers with the deepest IP, shallowest EA and lowest optical gap.

**Overall composition**

The three monomers explored in this chapter are in the set of monomers used in chapter 7. In line with the results in chapter 7, it as found that oligomers of electron-rich monomers, such as thiophene or pyrrole, had shallower IPs and deeper EAs, resulting in lower fundamental and optical gaps. In relation to photocatalytic water splitting, while a low optical gap is preferable for maximal photon absorption, having a shallower IP inhibits the oxidation of water or a sacrificial electron donor (SED) (such as TEA). On other hand, electron-poor monomers such as pyrazine and pyridine make up polymers with deeper IPs and shallower EAs. Here, in the context of water splitting, the reverse problem occurs – the deeper IPs are more suitable for accepting electrons however, this comes at the cost of a wider optical gap, limiting photon absorbance. A situation as such truly exemplifies difficulties encountered in property optimisation.

**5.4 Conclusions**

Overall, this study allowed us to probe the effect of monomer concentration and sequence isomerism on the optical and electronic properties of co-polymers, in addition to demonstrating the transferability of our previously established HTVS method for this purpose. The overall
Figure 5.4: (a) Formula for calculation of degree of segregation and depiction of its variables, 0 being fully segregated and 1 being fully alternating; (b) Variation of properties (IP, EA and optical gap) as a function of D for all possible sequences for 1:1 dodecamer.

The properties of the co-oligomers were intermediates of the properties of the composite homo-oligomers and leaned towards the properties of the monomer that was at a higher concentration in the composition.

For example, a higher concentration of phenylene monomers led to more phenylene-like properties overall. As expected, the greatest change in properties from one concentration to
another was proportional to how different the properties of the corresponding homo-oligomers were. It was also shown that the co-oligomers did exhibit sequence dependence in terms of IP, EA and optical gap. Once again, the effect of sequence isomerism was more dramatic for the co-oligomers composed of the most dissimilar monomers (i.e. ones containing thiophene). This sequence dependence of the overall co-oligomer properties was more pronounced at lower concentrations of thiophene. Clearly, thiophene’s electronics were disruptive to homo-oligomers of pyridine and phenylene more so than phenyl or pyridine monomers were to a thiophene homopolymer.

The outcome of this case study showed that it is indeed useful to conduct a detailed study on the monomer sequencing in polymers, particularly for applications to do with photovoltaics and photocatalysis as the optical properties were the most affected. This led to the similar work in chapter 6, in which I investigated the properties of ordered co-polymers and the effect of block co-polymer sequences, using a larger monomer library.
Chapter 6

High-Throughput Virtual Screening of Block Copolymers

6.1 Introduction

Organic conjugated polymers are often used in applications such as organic photovoltaics (OPV), organic light emitting diodes (OLEDs) and organic hydrogen evolution photocatalysts, owing to their inherent visible-light photoactivity. Aside from their visible-light photoactivity, which results from their conjugated $\pi$-system, conjugated polymers possess a number of other attractive properties. Typical inorganic photovoltaics, photocatalysts and LEDs are often based on materials containing expensive or toxic elements. For example the semiconductor CdTe, which is commonly used in thin-film solar cells, contains toxic cadmium and expensive tellurium. Organic polymers have the advantage of being composed of earth-abundant elements, in addition to their mechanical flexibility (facilitating their incorporation into devices), solution processability and tunability.

The tunability of their optical and electronic properties is one of the biggest advantages of using organic polymers in the applications mentioned above. Co-polymerisation is one of the main ways that polymer properties can be fine-tuned, allowing monomers to be combined in a multitude of ways in order to achieve the desired properties. Through judicious choice of composition and sequence, it is possible to combine the attractive properties that different homopolymers possess in the form of a co-polymer. As discussed in further detail in chapter 1, co-polymers can be categorised by their sequences. Different types of co-polymers include random co-polymers, alternating co-polymers and block co-polymers, with each sequence giving rise to different properties. Previous work by our group explores the effect of co-polymerisation on the hydrogen evolution activity for different families of monomers and polymers, probing the tunability of these materials. For instance, in the collaboration with Sprick et al. it was demonstrated that polymer composition is not the only significant variable affecting the overall polymer properties, highlighting that the segregation of differing
monomers within a polymer may lead to increased hydrogen evolution. Similarly, in another of our previous works\textsuperscript{166} where the effect of monomer sequence on the overall properties of oligomers was explored for a small number of monomers (discussed in chapter 5), we observed that the degree of segregation did indeed affect where the optoelectronic properties lay relative to their corresponding homo-oligomers.

In this chapter, this work is extended to form an understanding of the effect of the variation in ordered sequences, from a fully alternating co-oligomer sequence to a block co-oligomer sequence, composed of a library of 61 organic monomers. The results of this study are then used to analyse the applicability of these materials as donor materials for OPV and colloidal polymeric photocatalysts for water splitting. I investigate how the effect of conformerism and the overall optoelectronic properties are influenced by changes in the monomer sequences.

The datasets for this work can be found at:
https://github.com/isabelleapo/block_copolymers

6.2 Methodology and Computational Details

To carry out the polymer screening I wrote a Python pipeline that builds co-oligomers using the Python libraries RDKit\textsuperscript{128} and stk\textsuperscript{133,134} in different ordered sequences for different oligomer lengths. For example, if an octamer is created from two monomers A and B, co-oligomers in three unique sequences would be created by my code: AAAABBBB, AABBAABB and ABABABAB. For a dodecamer, four possible sequences exist: AAAAAABBBBB, AAABB-BAAAABB, AABBAABAABBB, ABABABABABAB.

Starting from a library of 61 monomers, a dataset of 5551 octamers (5490 co-oligomers - 1830 of each sequence, and 61 homo-oligomers) was created. A truncated library of 53 monomers was used to form a dataset of 5565 dodecamers (5512 co-oligomers - 1378 of each sequence, and 53 homo-oligomers), as some monomers resulted in oligomers too large to build using RDKit. For all oligomers a conformer search was carried out in which 500 conformers were randomly generated using the Experimental-Torsion Distance Geometry with additional basic knowledge (ETKDG).\textsuperscript{126} The lowest energy conformer for each structure was then determined using the Merck Molecular Force Field (MMFF),\textsuperscript{127} as implemented in RDKit.\textsuperscript{128}

The ionization potential (IP) and electron affinity (EA) of all of the lowest energy conformers of all of the oligomers were calculated using (IPEA-xTB).\textsuperscript{130} The optical gaps were calculated using the sTDA-xTB method.\textsuperscript{131} The GFN-xTB and sTDA-xTB calculations were run using the xtb\textsuperscript{132} and stda codes,\textsuperscript{167} respectively. The IP and EA calculations were performed using the generalised Born solvation model available within GFN-xTB, in water and benzene solvents, and the optical gap calculations were performed in the gas-phase. For a subset of octamers and dodecamers, the properties of all 500 conformers were calculated in order to probe the effect of conformerism on the oligomer properties.

For the calibration of the xTB-calculated properties of the oligomers, DFT calculations
were run on 60 of them. The oligomers used to obtain the fitting parameters for the calibration procedure were chosen so as to span the whole range for each of the properties. The ground state, anionic and cationic structures were optimised using B3LYP/DZP and the COSMO solvation model with dielectric constants of 80.1 and 2, to model water and benzene solvation, respectively. The IP and EA were obtained as the difference between the energy of the neutral and cationic, and anionic and neutral structures of each oligomers, respectively, i.e. using the ∆SCF method described in previous chapters (see eqs. (2.46) and (2.47)). All of the DFT calculations were run using Turbomole 7.01. Finally, a linear fit was performed and the fitting parameters used to obtain new calibrated values from the semi-empirically derived properties.

6.3 Results and Discussion

6.3.1 Calibration

The properties of all of the oligomers were calculated using (IPEA/sTDA-)xTB. The properties of 60 of these oligomers were also calculated using (TD-)DFT and used to fit linear models using Python. The structures selected for the fitting step were selected so as to span the entire range of values for each of the properties; the results of the calibration are shown in fig. 6.1. Table 6.1 contains the fitting parameters used in the calibration, as well as the $R^2$ value between the uncalibrated xTB properties and DFT-calculated properties, and the mean absolute error (MAE) between the calibrated and DFT-calculated properties.

Figure 6.1: Calibration curves for $-\text{IP}$ (left), $-\text{EA}$ (centre) and optical gap (left) for a subset of the oligomers, as calculated by (IPEA/sTDA-)xTB and B3LYP/DZP in water. Here the uncalibrated data points (blue), the calibrated points (orange), the line of best fit (black) and the $x=y$ line (red) are shown.
Table 6.1: Fitting parameters and $R^2$ obtained from the linear fit between the (sTDA-)xTB and DFT results in water, and the MAE between the DFT-calculated properties and the calibrated xTB-calculated properties.

<table>
<thead>
<tr>
<th>Property</th>
<th>Slope</th>
<th>Intercept</th>
<th>$R^2$</th>
<th>MAE / eV</th>
</tr>
</thead>
<tbody>
<tr>
<td>IP</td>
<td>0.9386</td>
<td>0.0722</td>
<td>0.91</td>
<td>0.13</td>
</tr>
<tr>
<td>EA</td>
<td>0.9219</td>
<td>0.5522</td>
<td>0.87</td>
<td>0.17</td>
</tr>
<tr>
<td>$\Delta_o$</td>
<td>0.5965</td>
<td>0.7915</td>
<td>0.82</td>
<td>0.27</td>
</tr>
</tbody>
</table>

6.3.2 Conformer Sampling

The effect of conformation was explored for the co-oligomers of only four unique monomers from the monomer library. Octamers and dodecamers were built from monomers 35, 26, 15, and 46 (see fig. 6.2), resulting in 6 unique overall compositions, with 3 and 4 unique monomer sequences for the octamers and dodecamers, respectively (i.e. AAAABBBB, (AABB)$_2$ and (AB)$_4$ for the octamers and AAAAABBBBBB, (AAABBB)$_2$, (AABB)$_3$ and (AB)$_6$). For each of these oligomers (18 octamers and 24 dodecamers), the properties of all of the 500 embedded conformers were calculated. By calculating the Boltzmann factor for each conformer relative to the lowest energy conformer and plotting them against their calibrated properties it was confirmed that for both the octamers and dodecamers the lowest energy conformers - those with Boltzmann factors close to 1 - showed little variation in the overall properties. This justifies the use of the lowest energy conformers in subsequent analyses. The scatter plots showing this relationship for the dodecamers are included in fig. 6.3 and fig. 6.4.
Figure 6.2: The library of 61 monomers used to construct the oligomers. The dashed lines indicate the site of connection in the (co-)oligomers.
Figure 6.3: Calibrated properties in water v. Boltzmann factor for the 500 conformers of each of the homo-oligomers (dodecamers) composed of monomers 35, 26, 15 and 46. The lowest energy conformers are those with Boltzmann factors close to 1.
Figure 6.4: Calibrated properties in water v. Boltzmann factor for the 500 conformers of each of the co-oligomers (dodecamers) composed of monomers 35, 26, 15 and 46. The colours of the points correspond to the co-oligomers of the monomer pairs of the same colour. The lowest energy conformers are those with Boltzmann factors close to 1.
The combined effect of sequence and conformerism on the $-IP$ and $-EA$ of the same 6 compositions can be seen in the violin plots in fig. 6.5. In each plot, each violin represents the type of sequence (ranging from fully alternating to block, with the repeat unit labelled on the x-axis), the width of the violins signifies the number of conformers at that energy level and the y-axis is the energy in eV. The violins at the top of the plots represent the $-EA$ and the bottom ones the $-IP$. The results for both the octamers and dodecamers are shown in fig. 6.5, where in each plot the horizontal lines represent the $-IP$ and $-EA$ of the homo-oligomers that would be formed from the two monomers that make up the co-oligomers in question. In each plot, the pink/blue lines correspond to the $-IP/-EA$ of the homo-oligomer composed of the first monomer in the pair, and the red/green lines correspond to the $-IP/-EA$ of the homo-oligomer composed of the second monomer in the pair. So, for example, in fig. 6.5(c), the pink and blue lines correspond to the $-IP$ and $-EA$ of the homo-oligomer of 35, the red and green lines correspond to the IP and EA of the homo-oligomer of 46, and the violins represent the octamer co-oligomers of 35 and 46 for each sequence.

In all cases, when varying the oligomer sequence from fully alternating to block, the violins tended towards the shallowest $-IP$ and deepest $-EA$ of the two corresponding homo-oligomers, with the effect being more pronounced where there was a larger difference between the potentials of the homo-oligomers of the relevant monomers. For example, although this trend is followed for both fig. 6.5(a) and (g) and fig. 6.5(c) and (i), it is much more apparent for the latter pair, where the constituent monomers are much more structurally (and therefore electronically) distinct. Additionally, although the $-IP$s and $-EAs$ of the alternating oligomers were more of an intermediate between the homo-oligomers of the constituent monomers, they were still closer to the shallowest $-IP$ and deepest $-EA$ in all cases.

There is also a discernible trend in the effect of conformerism. The $-IP$ violins all became narrower upon changing the sequence from alternating to block, indicating an increase in the influence of conformerism on the overall $-IP$ of the oligomer. This trend is again, more pronounced for oligomers composed of more electronically distinct monomers – see fig. 6.5(g) in comparison to fig. 6.5(l).

Trends related to the alignment of the homo-oligomer potentials – i.e. whether they are straddled or staggered – can also been seen in these plots. As mentioned in chapter 4, when the IP and EA of one material are either both higher or both lower than the IP and EA of the other, we say they are staggered. Conversely, if the potentials of one material both lie in between the potentials of the other, we say that one material straddles the other. This is described in fig. 6.10(c).

In the cases where the potentials of the homo-oligomers of the monomers in question were staggered, the shallowest $-IP$ and deepest $-EA$ belonged to different homo-oligomers. An example of this case can be seen in fig. 6.5(c) and (i), where both the $-IP$ and $-EA$ of 35 are less negative than the $-IP$ and $-EA$ of 46. Interestingly, this likely means that block co-polymers with staggered homopolymer potentials exhibit intramolecular charge separation,
6.3. RESULTS AND DISCUSSION

Figure 6.5: Violin plots describing the effect of both sequence and conformerism on the $-\text{IP}$ and $-\text{EA}$ of both the octamers and the dodecamers of 35, 26, 15 and 46 in water. There are 3 and 4 unique sequences for the octamers and dodecamers, respectively. The pink/blue horizontal lines correspond to the $-\text{IP}/-\text{EA}$ of the homo-oligomer of the first monomer in each pair, and the red/green dashed line to the potentials of the homo-oligomer of the second monomer in each pair.
a useful property in applications such as OPV and photocatalysis worth further exploration. Intramolecular charge separation is discussed in greater detail in section 6.3.4.

**Figure 6.6:** Violin plots describing the effect of both sequence and conformerism on the optical gap for both the octamers and the dodecamers of 35, 26, 15 and 46 in water. There are 3 and 4 unique sequences for the octamers and dodecamers, respectively.

Figure 6.6 shows the violin plots for optical gaps of the same co-oligomers, with the pink horizontal line corresponding to the first homo-oligomer composed of the first monomer in each pair and the blue line corresponding to the second one. Three main trends can be seen in these plots. Firstly, the optical gap of a co-oligomer was, for most compositions and sequences, closest to the lowest optical gap of the corresponding homo-oligomers. For instance, all of the violins in fig. 6.6(b) lay closer to the blue dashed line. While there was variation with sequence, overall, the optical gaps of the co-oligomers (i.e. the violins in these plots) resided either in between the two homo-oligomer optical gaps or nearer to the lower optical gap homo-oligomer line.

The second trend expressed in these violin plots is that block co-oligomers tended towards the lower optical gap homo-oligomer, relative to their fully alternating counterparts. In al-
most all cases, the median optical gap value decreased upon increasing the block character of the sequence. Although still present, the trend was less pronounced for structurally similar monomers, as in the case of the monomers 35 and 26 in fig. 6.6(a) and (g). Although there were some deviations from this trend (e.g. fig. 6.6(f) and (l)), the difference in the highest and lowest median optical gaps was less than 1 eV, indicating that this deviation was insignificant.

Finally, there appears to be a weak correlation between the effect of conformer isomerism on the optical gap and the sequence of the co-oligomers. In most cases, there was a gradual narrowing of the violins from the alternating sequence to the block sequences and a greater spread of values with increasing block character. This signifies a greater influence on the optical gap by a change in conformation for the block co-oligomers, which does make intuitive sense as the distinct blocks in a sequence can lead to vastly different conformations, whereas for the fully alternating sequences the conformers do not differ as dramatically in terms of structure.

6.3.3 Property-Property Relationships

Figures 6.7 and 6.8 show the 2D histograms and convex hulls of the calibrated water octamers, respectively. Although the $-\text{IP}$ and $-\text{EA}$ appeared to have a weak, positive correlation, the properties generally did not have a discernible correlation. The 2D histograms in fig. 6.7 show that for the optical gap property slices, the oligomers were in the same general region of property space, save for a few outliers with lower optical gaps. Upon inspection of the convex hulls in fig. 6.8, it is apparent that it was the block and (AABB)$_2$ octamers that led to these lower optical gap values. Below 1.7 eV, there were 14 octamers - 10 with the AAAABBBB sequence and 4 with the (AABB)$_2$ sequence. Interestingly, all of the octamers combined an electron-rich monomer with an electron-poor monomer. All but one of these octamers contained monomer 8 in combination with electron-rich monomers containing either thiophene units or pyrrolic nitrogens. The remaining octamer was composed of the monomers 48 and 16, an electron-rich fused pyrrole monomer and an electron-poor benzoxadiazole monomer, in the AAAABBBB sequence.

Supported by TD-DFT-calculated optical gap data, it appears that there is a relationship between polymers possessing electronically distinct blocks and low optical gaps. This is an interesting result as a materials design strategy that allows for the fine-tuning of the optical gap is particularly desirable for the applications discussed in this work. Two of the structures found in the low optical gap region are shown in fig. 6.9.
Figure 6.7: 2D histograms of the slices of property space spanned by $-\text{IP}$ v. $-\text{EA}$, $-\text{IP}$ v. $\Delta_o$ and $-\text{EA}$ v. $\Delta_o$ (left to right).

Figure 6.8: Convex hulls enclosing the property space of $-\text{IP}$ v. $-\text{EA}$, $-\text{IP}$ v. $\Delta_o$ and $-\text{EA}$ v. $\Delta_o$ (left to right), for the different co-oligomer sequences and homo-oligomers (octamers) in water; pink = homo-oligomers, red = AB, blue = AABB and green = AAAABBBB.

Figure 6.9: Example of an AABB octamer (top) and an AAAABBBB octamer (bottom) with optical gaps below 1.7 eV.

### 6.3.4 Intramolecular Charge Separation in Block Co-polymers

Following the observation of the trends in fig. 6.6, where the staggered homo-oligomers led to block co-oligomers with $-\text{IP}$s closer to that of one of the two corresponding homo-oligomers
and $-E_A$s closer to that of the other, I searched for the occurrence of this feature in all of the octamer and dodecamer block co-oligomers.

The implication of this observation was that these materials would likely exhibit intramolecular charge separation. The term ‘charge separation’ refers to the photoinduced excitation and separation of an exciton into an electron, which transfers to an acceptor, and a hole, which transfers to a donor. An example of this would be in a p-n junction in a photovoltaic device, the interface between a donor and acceptor material at which opposite charges separate. In the context of intramolecular charge separation in block co-polymers, this would refer to the electron and hole of an exciton transferring to different blocks in the co-polymer. If, for example, a polymer displayed this electronic behaviour, it could play both the role of the donor and the acceptor in a p-n junction in an OPV device, which is normally made up of two separate materials.

In fig. 6.10, the differences between homo-oligomer $-IP$s are plotted against the differences in homo-oligomer $-E_A$s, with the points coloured according to whether they exhibit the right electronic properties that could lead to intramolecular charge separation. The axes are labelled $\Delta IP_{AB}$ and $\Delta E_A_{AB}$ to denote the differences between homo-oligomers A and B. It follows that in the quadrants where the differences in $\Delta IP_{AB}$ and $\Delta E_A_{AB}$ are both positive or both negative, the potentials are staggered; the opposite quadrants contain the co-oligomers with straddled homo-oligomer potentials.

Indeed, almost all of the block co-oligomers in the dodecamers dataset with staggered homo-oligomer potentials exhibited properties likely leading to intramolecular charge separation, which is a promising result. As mentioned in the previous example, in a typical polymer OPV both a donor and acceptor molecule are required to form a p-n junction and drive the separation of free charge carriers. One advantage of using a polymer that could act as a single molecule p-n junction is that having fewer components means not having to ensure that both of the materials are compatible and have the appropriate alignment, if for example, a new material was to be incorporated. Having one polymer as a p-n junction would, therefore, allow for greater simplicity in building device architectures. Intramolecular charge separation is also relevant to photocatalysts for water splitting, as it would facilitate the separation of excitons into free charge-carriers. A disadvantage of using these materials, however, would be the difficulty in achieving the controlled synthesis of block co-polymers.

### 6.3.5 Structure-Property Relationships

The structure-property relationships for the different property combinations were also explored, firstly, by splitting the property plots into sections and searching for the most common monomers in each section. The results are displayed in the following scatter plots (figs. 6.11 to 6.13), as well as a selection of the dominant monomers in particular areas. The colours of the points in the sections of property space in which the selected monomers reside match the colours of the monomers in the figure.
Figure 6.10: (a) Scatter plot of the difference in IP between the two constituent homo-oligomers of each block dodecamer v. the difference in EA of the two constituent homo-oligomers in water. The quadrants correspond to block co-oligomers where the potentials of their corresponding homo-oligomers are either straddled or staggered. The yellow points represent the block co-oligomers exhibiting the discussed charge-separated effect and the points in blue represent those that do not; (b) examples of two dodecamers that exhibit intramolecular charge separation; (c) example of staggered and straddled alignments of potentials.

The limits in eV for each property that made up the grids on each plot were:

\[ -\text{IP}: < -5.5; -5.5, -5; -5, -4.5; -4.5, -4; -4 < \]
\[ -\text{EA}: \leq -3; -3, -2.5; -2.5, -2; -2 \leq \]
\[ \Delta_o: \leq -1.5; -1.5, -2.5; -2.5, -2.8; -2.8, -3; -3, -3.2; -3.2 \leq \]
Figure 6.11: Scatter plot of the $-\text{IP}/-\text{EA}$ property slice showing (a) all octamers, (c) all (AB)$_4$ octamers, (d) all (AABB)$_2$ octamers and (e) all AAAABBBB octamers containing the most common monomers in the different regions described in table 6.2. Each colour represents the oligomers containing the relevant monomer. A selection of the monomers shown on these plots are shown in (b), coloured to match the data points describing them on the scatter plots.

Table 6.2: Most commonly found monomers in each of the $-\text{IP} - \text{EA}$ grid sections.

<table>
<thead>
<tr>
<th>$-\text{IP}$</th>
<th>$-\text{EA}$</th>
<th>Monomer</th>
</tr>
</thead>
<tbody>
<tr>
<td>$&lt;-5.5$</td>
<td>$&lt;-3$</td>
<td>8</td>
</tr>
<tr>
<td>$-5.5 \leq$</td>
<td>$&lt;-3$</td>
<td>17</td>
</tr>
<tr>
<td>$&lt;-5.5$</td>
<td>$-3, -2.5$</td>
<td>33</td>
</tr>
<tr>
<td>$-5.5 \leq$</td>
<td>$-3, -2.5$</td>
<td>2, 52, 53</td>
</tr>
<tr>
<td>$&lt;-5.5$</td>
<td>$-2.5, -2$</td>
<td>27</td>
</tr>
<tr>
<td>$-5.5, -4.5$</td>
<td>$-2.5, -2$</td>
<td>59</td>
</tr>
<tr>
<td>$-4.5 &lt;$</td>
<td>$-2.5, -2$</td>
<td>48</td>
</tr>
<tr>
<td>$-5.5, -4.5$</td>
<td>$-2 &lt;$</td>
<td>14</td>
</tr>
<tr>
<td>$-4.5 &lt;$</td>
<td>$-2 &lt;$</td>
<td>48</td>
</tr>
</tbody>
</table>

In the plots for $-\text{IP} - \text{EA}$, the most common monomer in the top right corner was
monomer 48, an electron-rich fused pyrrole molecule. The monomers comprising the oligomers occupying the regions of property space with deeper $-\text{IP}$ and $-\text{EA}$ mostly possessed pyridinic nitrogens, an electron-poor group, for example monomers 8, 17, 29, 27 and 31. These trends are in line with what has previously been observed in my group and in other chapters of this thesis.

In fig. 6.12 the same trend is apparent, as the oligomers with deeper $-\text{IPs}$ generally consisted of electron-poor monomers such as 47, 22, 33, 61 and 8. The most dominant monomers within the optical gap boundaries of 1.5 and 2.5 eV when searching the $-\text{IP}$ and $\Delta_o$ property space, were 8, 11, 16, 48 and 49 (not pictured in fig. 6.12), for the whole range of $-\text{IP}$ values.

For the property space mapped out by $-\text{EA}$ and $\Delta_o$ (pictured in fig. 6.13), the dominant monomers were the almost the same: 8, 11, 48, 49 and 50 (not pictured in fig. 6.13). The common feature in these monomers is the electron-rich heteroatoms on fused five membered rings, features that one would intuitively expect to result in a reduced optical gap.

In general, as seen in the convex hulls in fig. 6.8, block co-oligomers occupied a larger range of the $-\text{IP}$ v. $\Delta_o$ and $-\text{EA}$ v. $\Delta_o$ property slices than their fully alternating counterparts. As mentioned in section 6.3.3, is useful to be able to access lower optical gaps by co-polymerisation and this data suggests that combining the correct monomers in a block sequence may be a route to achieving this.

Figure 6.12: Scatter plot of the $-\text{IP}/\Delta_o$ property slice showing (a) all octamers, (c) all (AB)$_4$ octamers, (d) all (AABB)$_2$ octamers and (e) all AAAABBBB octamers containing the most common monomers in the different regions described in table 6.3. Each colour represents the oligomers containing the relevant monomer. A selection of the monomers shown on these plots are shown in (b), coloured to match the data points describing them on the scatter plots.
Figure 6.13: Scatter plot of the $-\text{EA}/\Delta\sigma$ property slice showing (a) all octamers, (c) all (AB)$_4$ octamers, (d) all (AABB)$_2$ octamers and (e) all AAAABBBB octamers containing the most common monomers in the different regions described in table 6.4. Each colour represents the oligomers containing the relevant monomer. A selection of the monomers shown on these plots are shown in (b), coloured to match the data points describing them on the scatter plots.
Table 6.3: Most commonly found monomers in each of the $-\text{IP} \times \Delta_o$ grid sections.

<table>
<thead>
<tr>
<th>$-\text{IP}$</th>
<th>$\Delta_o$</th>
<th>Monomer</th>
</tr>
</thead>
<tbody>
<tr>
<td>$&lt;-5.5$</td>
<td>1.5, 2.5</td>
<td>8</td>
</tr>
<tr>
<td>$&lt;-5.5$</td>
<td>2.5, 2.8</td>
<td>58</td>
</tr>
<tr>
<td>$&lt;-5.5$</td>
<td>2.8, 3</td>
<td>47</td>
</tr>
<tr>
<td>$&lt;-5.5$</td>
<td>3, 3.2</td>
<td>22, 33</td>
</tr>
<tr>
<td>$&lt;-5.5$</td>
<td>3.2 &lt;</td>
<td>20, 56, 61</td>
</tr>
<tr>
<td>$-5.5$, $-5$</td>
<td>&lt; 1.5</td>
<td>8, 49</td>
</tr>
<tr>
<td>$-5.5$, $-5$</td>
<td>1.5, 2.5</td>
<td>8, 16</td>
</tr>
<tr>
<td>$-5.5$, $-5$</td>
<td>2.5, 2.8</td>
<td>60</td>
</tr>
<tr>
<td>$-5.5$, $-5$</td>
<td>2.8, 3</td>
<td>59</td>
</tr>
<tr>
<td>$-5.5$, $-5$</td>
<td>3, 3.2</td>
<td>13, 54, 55</td>
</tr>
<tr>
<td>$-5.5$, $-5$</td>
<td>3.2 &lt;</td>
<td>18, 57</td>
</tr>
<tr>
<td>$-5$, $-4.5$</td>
<td>&lt; 1.5</td>
<td>8, 50</td>
</tr>
<tr>
<td>$-5$, $-4.5$</td>
<td>1.5, 2.5</td>
<td>11, 49</td>
</tr>
<tr>
<td>$-5$, $-4.5$</td>
<td>2.5, 2.8</td>
<td>15</td>
</tr>
<tr>
<td>$-5$, $-4.5$</td>
<td>2.8, 3</td>
<td>12, 53</td>
</tr>
<tr>
<td>$-5$, $-4.5$</td>
<td>3, 3.2</td>
<td>12</td>
</tr>
<tr>
<td>$-4.5$, $-4$</td>
<td>1.5, 2.5</td>
<td>11</td>
</tr>
<tr>
<td>$-4.5$, $-4$</td>
<td>2.5, 3</td>
<td>51</td>
</tr>
<tr>
<td>$-4$, $&lt; 1.5$</td>
<td>1.5, 2.5</td>
<td>48</td>
</tr>
<tr>
<td>$-4$, $&lt; 2.5$</td>
<td>2.5, 2.8</td>
<td>48</td>
</tr>
</tbody>
</table>

Table 6.4: Most commonly found monomers in each of the $-\text{EA} \times \Delta_o$ grid sections.

<table>
<thead>
<tr>
<th>$-\text{EA}$</th>
<th>$\Delta_o$</th>
<th>Monomer</th>
</tr>
</thead>
<tbody>
<tr>
<td>$&lt;-3$</td>
<td>$\leq 2.5$</td>
<td>8</td>
</tr>
<tr>
<td>$&lt;-3$</td>
<td>2.5, 2.8</td>
<td>17</td>
</tr>
<tr>
<td>$&lt;-3$</td>
<td>2.8, 3</td>
<td>43</td>
</tr>
<tr>
<td>$&lt;-3$, $-2.5$</td>
<td>1.5, 2.5</td>
<td>11</td>
</tr>
<tr>
<td>$&lt;-3$, $-2.5$</td>
<td>2.5, 2.8</td>
<td>58</td>
</tr>
<tr>
<td>$&lt;-3$, $-2.5$</td>
<td>2.8, 3</td>
<td>30, 40</td>
</tr>
<tr>
<td>$&lt;-3$, $-2.5$</td>
<td>3, 3.2</td>
<td>22, 46</td>
</tr>
<tr>
<td>$&lt;-3$, $-2.5$</td>
<td>3.2 &lt;</td>
<td>29</td>
</tr>
<tr>
<td>$-2.5$, $-2$</td>
<td>1.5, 2.5</td>
<td>49, 50</td>
</tr>
<tr>
<td>$-2.5$, $-2$</td>
<td>2.5, 2.8</td>
<td>15, 60</td>
</tr>
<tr>
<td>$-2.5$, $-2$</td>
<td>2.8, 3</td>
<td>59</td>
</tr>
<tr>
<td>$-2.5$, $-2$</td>
<td>3, 3.2</td>
<td>13, 37</td>
</tr>
<tr>
<td>$-2.5$, $-2$</td>
<td>3.2 &lt;</td>
<td>56, 57, 61</td>
</tr>
<tr>
<td>$-2$, $&lt; 1.5$</td>
<td>1.5, 2.8</td>
<td>48</td>
</tr>
<tr>
<td>$-2$, $&lt; 2.8$</td>
<td>2.8, 3</td>
<td>51</td>
</tr>
<tr>
<td>$-2$, $&lt; 3.2$</td>
<td>12, 13</td>
<td>14, 54</td>
</tr>
</tbody>
</table>
Co-polymer candidates for photocatalytic water splitting

There is currently a rising interest in using organic polymers such as the ones investigated in this chapter in photocatalysts for water splitting as they possess desirable characteristics such as tunability and earth-abundance, alongside their inherent photoactivity. This motivated the search for ideal candidates for this application within the dataset solvated in water. For a polymer to work successfully as a colloidal photocatalyst, it must possess a $-\text{IP}$ deeper than the oxidation potential of water and a $-\text{EA}$ shallower than reduction potential of hydrogen. In order to do this, I selected the data for the suitable materials by searching for the co-oligomers with $-\text{EA}$ shallower than the reduction potential of hydrogen ($-4.025$ eV at pH 7), $-\text{IP}$ deeper than the oxidation potential of water ($-5.26$ eV at pH 7) and optical gaps below 2.2 eV, in order to absorb light from as much of the spectrum as possible. An additional 0.3 eV was included in the boundaries in order to account for a potential kinetic driving force required for the reaction.

Within these boundaries lay 297 octamers, 5.41% of the total number of octamers in the dataset, with all 61 monomers occurring at least once. Of these octamers, 28.28% were fully alternating, 30.3% had the AABB repeat unit and 43.4% were block oligomers. The most common monomer in this region was monomer 8, as 100% of the oligomers containing this monomer in the entire dataset were found within this region.

Co-polymer candidates for donor materials in organic photovoltaics

The investigated co-oligomers are also of interest as potential donor or acceptor molecules in OPV devices. Their suitability as donors was explored relative to three known acceptor molecules commonly used in OPV devices; PCBM, PC$_{70}$BM and PN1, which have been discussed in section 1.2.2. The reference $-\text{IP}$ and $-\text{EA}$ values were $-5.85$ eV and $-3.89$ eV for PCBM,$^{164}$ $-5.96$ eV and $-3.9$ eV for PC$_{70}$BM$^{164}$ and $-5.71$ eV and $-3.85$ eV for PN1.$^{56}$

In an OPV device, the donor material should have both a higher $-\text{IP}$ and $-\text{EA}$ than the acceptor material, where the $-\text{EA}$ of the donor is at least 0.3 eV higher than that of the acceptor, so as to provide a driving force for the separation of an exciton. These constraints were initially applied, narrowing the oligomer candidates down to 2763 (50.3% of the total oligomers), 3118 (56.8%) and 2316 (42.2%) relative to PCBM, PC$_{70}$BM and PN1, respectively. However, the range of wavelengths absorbed by the material must of course be considered as well. As such, an additional constraint of optical gaps below 2.2 eV was imposed on the candidate oligomers, yielding a total of 156 (2.8% of the total oligomers), 167 (3.0%) and 148 (2.7%) oligomers relative to PCBM, PC$_{70}$BM and PN1, respectively. This optical gap boundary was chosen in consideration of the commonly used donor material P3HT, which possesses an optical gap of $\sim$1.9 eV.$^{165}$

It appears that more of the co-oligomers with a sequence of AAAABBBBB, i.e. those with more block character, tend to be more appropriate for this application, as a greater proportion
of these oligomers lay within these given boundaries - these numbers are shown in table 6.5. It was seen in fig. 6.8 that materials with the (AABB)$_2$ and AAAABBBBB sequences gave rise to lower optical gaps and therefore, it makes sense that co-oligomers based on these sequences would be more appropriate in OPV donor materials, as they would allow for greater light absorption. The block co-oligomers were also the ones that displayed intramolecular charge separation in section 6.3.4, which, as discussed previously, would be a very attractive feature in an OPV device as it could potentially allow for a single molecule p-n junction.

Figure 6.14 shows the relationship between the $-\text{IP}$s and $-\text{EAs}$ of the two homo-oligomers corresponding to each co-oligomer and which ones were suitable as donors in OPV relative to PCBM. The points in yellow represent those within the given OPV boundaries and those in blue are the co-oligomers outside of the boundaries; with the top two panels corresponding to only the electronic requirements and the driving force, and the bottom two panels corresponding to the dataset under optical gap constraints as well. The equivalent plots for PC$_{70}$BM and PN1 are shown in figs. A.17 and A.18, respectively.

In the top left panel of fig. 6.14, the dominating effect on the co-oligomer properties of the homo-oligomer with the shallower $-\text{IP}$ can be seen. For example, a data point corresponding to a co-oligomer where the $-\text{IP}$ of homo-oligomer A is deep but the $-\text{IP}$ of homo-oligomer B is shallow, still resides within the required electronic boundaries for an OPV donor. This is in line with previous observations in work by my group, and with the discussion of the max/min model in chapter 4.

In OPV devices, and all photovoltaics in general, it is desirable to maximise the open circuit voltage ($V_{oc}$), whilst still minimising the optical gap. A simple model for the calculation of the $V_{oc}$ was proposed by Scharber et al. and has been used here to investigate which oligomers possess high $V_{ocs}$ but low optical gaps. These results are shown in fig. 6.15, with the different colours corresponding to the aforementioned electronic and optical constraints. The ideal donor material would possess the properties discussed above, in addition to a high $V_{oc}$. A lower threshold of 0.7 V was used to narrow down the set of potential donors with respect to the three aforementioned acceptor materials. This threshold was chosen since the $V_{oc}$ of the widely used silicon solar cells is $\sim 0.6/0.7$ V.

Within these constraints lay 99 octamers (made up of 44 unique monomers), 107 octamers (43 unique monomers) and 105 octamers (50 unique monomers), with respect to PCBM, PC$_{70}$BM and PN1, respectively. The most common monomer was 11, a monomer which previously was shown to give rise to low optical gaps. There was a higher occurrence of block sequence octamers, again a result of the optical gap constraints discussed earlier.
Figure 6.14: $-\text{IP}/-\text{EA}$ of homo-oligomer A v. $-\text{IP}/-\text{EA}$ of homo-oligomer B corresponding to each co-oligomer within the electronic constraints and driving force (top row) and within the optical and electronic constraints including a driving force (bottom row), with respect to the acceptor PCBM.

Figure 6.15: Optical gap v. open circuit voltage for oligomers in combination with PCBM, PC$_{70}$BM and PN1 (left to right) within the electronic (blue), electronic + driving force (orange) and optical, electronic + driving force (green) constraints.
Table 6.5: Number of OPV polymer candidates for each co-oligomer sequence relative to the experimentally determined potentials of the three acceptors PCBM, PC$_{70}$BM and PN1.

<table>
<thead>
<tr>
<th>Acceptor</th>
<th>Total</th>
<th>PCBM (50.3)</th>
<th>PC$_{70}$BM (56.8)</th>
<th>PN1 (42.2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. polymers under $-\text{IP}/-\text{EA}$ constraints</td>
<td>(AB)$_4$</td>
<td>867 (47.4)</td>
<td>991 (54.2)</td>
<td>699 (38.2)</td>
</tr>
<tr>
<td>(% of total polymers)</td>
<td>(AABB)$_2$</td>
<td>925 (50.5)</td>
<td>1036 (56.6)</td>
<td>769 (42.0)</td>
</tr>
<tr>
<td>AAAABBBB</td>
<td>971 (53.1)</td>
<td>1091 (59.6)</td>
<td>848 (46.3)</td>
<td></td>
</tr>
<tr>
<td>No. polymers under $-\text{IP}/-\text{EA}$ and optical gap constraints</td>
<td>Total</td>
<td>156 (2.8)</td>
<td>167 (3.0)</td>
<td>148 (2.7)</td>
</tr>
<tr>
<td>(% of total polymers)</td>
<td>(AB)$_4$</td>
<td>29 (1.6)</td>
<td>33 (1.8)</td>
<td>27 (1.5)</td>
</tr>
<tr>
<td>(AABB)$_2$</td>
<td>42 (2.3)</td>
<td>47 (2.6)</td>
<td>39 (2.1)</td>
<td></td>
</tr>
<tr>
<td>AAAABBBB</td>
<td>85 (4.6)</td>
<td>87 (4.8)</td>
<td>82 (4.5)</td>
<td></td>
</tr>
<tr>
<td>No. polymers under $-\text{IP}/-\text{EA}$, optical gap and $V_{\text{oc}}$ constraints</td>
<td>Total</td>
<td>99 (1.8)</td>
<td>107 (1.9)</td>
<td>105 (1.9)</td>
</tr>
<tr>
<td>(% of total polymers)</td>
<td>(AB)$_4$</td>
<td>20 (1.1)</td>
<td>24 (1.3)</td>
<td>18 (1.0)</td>
</tr>
<tr>
<td>(AABB)$_2$</td>
<td>30 (1.6)</td>
<td>35 (1.9)</td>
<td>29 (1.6)</td>
<td></td>
</tr>
<tr>
<td>AAAABBBB</td>
<td>49 (2.7)</td>
<td>48 (2.6)</td>
<td>58 (3.2)</td>
<td></td>
</tr>
</tbody>
</table>

6.4 Conclusions

The co-oligomers of 61 monomers were investigated in this chapter, focusing on the octamers in all cases and the dodecamers where possible. This high-throughput virtual screening (HTVS) allowed for the observation of trends relating to monomer sequence, through the investigation of 3 different sequences for the octamers and 4 for the dodecamers, with the aim of probing the effect of sequence on co-polymer properties.

A number of interesting insights were drawn from this study. The conformer sampling performed on a subset of the compositions revealed some interesting patterns. The aim of this part of the chapter was to validate the use of the lowest energy conformer to draw conclusions on the likely properties of the real-life polymer. It was found that the lowest energy conformers exhibited a very small range of properties, justifying the focus on the lowest energy conformer alone, in most of the chapter. This data also revealed the effect each unique sequence had on the overall properties, as analysed in terms of the violin plots. It was shown that the $-\text{IP}$
6.4. CONCLUSIONS

and −EA of the block co-oligomers tended towards the shallowest −IP and deepest −EA of their corresponding homo-oligomers, whereas the alternating co-oligomers lay in between the homo-oligomer potentials. Where the homo-oligomer potentials were staggered, the trend in which the corresponding block co-oligomer tended to the shallowest homo-oligomer −IP and deepest homo-oligomer −EA meant that the −IP and −EA that the block co-oligomer was closest to belonged to different homo-oligomers, indicating the possibility of intramolecular charge separation.

The equivalent analysis for the optical gaps highlighted the dominance of the monomer of the lowest optical gap homo-oligomer on the overall properties of the co-oligomers, for all of the sequences. However, the block co-oligomers tended to have the lowest optical gaps. This was an interesting result as reduction of the optical gap is a common aim in materials design for photoactive applications.

The property-property relationships revealed a weak correlation between the −IP and −EA, but no distinct correlation for the other two property slices. However, the convex hulls and histograms demonstrated that lower optical gaps can be accessed by increasing the block character of the sequence, in line with what was observed in the conformer sampling.

Following the observations from the conformer sampling, the possibility for intramolecular charge separation in the block co-oligomers for which the potentials of the corresponding homo-oligomers were staggered was investigated. Interestingly, the co-oligomer data showed that the block co-oligomers with staggered homo-oligomer potentials were those that displayed a potential for intramolecular charge separation. This was seen for both water and benzene, with the effect being more pronounced for the dodecamers than the octamers, indicating that this property would be present in a true polymer.

I also looked at the relationship between the structures of the oligomers and their properties by focusing on different sections in each property slice and finding the dominant monomers in each region. As expected, this search demonstrated that electron-poor monomers gave rise to co-oligomers with deeper −IP and −EA, and that conversely, electron-rich monomers gave rise to shallower −IP and −EA. Monomer 8 consistently appeared on co-oligomers in the low optical gap regions, suggesting that this monomer (and similar ones) may be a useful building block for low optical gap polymers.

I also focused on the areas where photocatalytic water splitting would be possible with respect to optical and electronic property constraints. A low percentage (2.84%) of the total oligomers was found to be suitable, with 8 again being the most common monomer as it led to low optical gap oligomers, which also have the correct potentials to drive water splitting.

A final application for which the structure-property relationships were studied were as donors in organic photovoltaics (OPV) devices. The suitability of the oligomers as donors was analysed relative to three common acceptor molecules: PCBM, PC_{70}BM and PN1. Here I focused on oligomers within three levels of constraints: i) electronic requirements plus an additional kinetic driving force, ii) optical and electronic requirements plus an additional
kinetic driving force and iii) electronic, optical and kinetic requirements and an open circuit voltage greater than 0.7 eV. Although around half of the oligomers were suitable donors with regards to the electronic and kinetic constraints, this number decreased dramatically to below 3% once the optical gap was also considered and below 2% once the minimum open circuit voltage requirement was taken into account. In all cases, there was a higher occurrence of block co-oligomers within the suitable region for an OPV donor, which makes intuitive sense as these generally had lower optical gaps.

A number of insights can be drawn from the screening in this chapter. Firstly, I have demonstrated the strength of this HTVS method in predicting the effect of varying the sequence in ordered co-polymers. This method could be extended in future work to a larger library of monomers, including more monomers closely related to those that led to desired properties in the oligomers in this work. Secondly, the potential for block co-polymers to be used in applications where intramolecular charge separation is desirable was revealed - for example, in a single-molecule p-n junction in an OPV device. Finally, I was able to deduce how different polymer compositions and sequences give rise to different properties and how these relate to common photoactive applications such as OPV and photocatalytic water splitting.
Chapter 7

Benchmarking DFT against GW

7.1 Introduction

As discussed in the introduction to this thesis, there are many reasons why it is desirable to calculate the IP, EA and optical gap of a material. Knowledge of these properties is essential for the informative classification and investigation of photoactive materials and experimental data is often either unavailable due to compounds not having been synthesized or difficult to obtain.

The go-to method for the calculation of the optical and electronic properties of molecules and materials is density functional theory (DFT) in combination with time-dependent density functional theory (TD-DFT) for excited state properties. While these methods are well-established and formally exact, the approximation of the exchange-correlation functional means that the calculated energies are not exact. Additionally, when approximated as the $-\text{HOMO}$ and $-\text{LUMO}$, the expected accuracy of the obtained IP and EA values is even lower.\cite{90} It is therefore instructive to benchmark (TD-)DFT against inherently more accurate methods based on correlated wavefunction or many-body perturbation theory. Such methods, for instance coupled-cluster theory and Green’s function based approaches, yield more accurate results but are typically too computationally expensive for routine calculations, especially in materials screening applications. They can, however, prove useful for the validation of the (TD-)DFT method.

In this chapter, the single-particle and optical excitations of oligomers of $p$-phenylene, 2,5-pyridine, 2,5-pyrrole, 2,5-thiophene and pyrazine are investigated using (TD-)DFT and the Green’s function based approach ($G_0W_0$/BSE). These conjugated materials are photoactive and find use in applications such as light emitting layers in organic light emitting diodes (OLEDs), solar absorbers in organic solar cells and organic photocatalysts for water splitting and organic synthesis reactions. The aim of this chapter was to benchmark the relatively inexpensive (TD-)DFT calculations that are used throughout this report against methods of a higher level of theory.
7.2 Methodology and Computational Details

The vertical IP, EA and optical gap ($\Delta_o$) values of a series of oligomers of $p$-phenylene, 2,5-pyridine, 2,5-pyrrole, 2,5-thiophene and pyrazine were calculated using a range of methods including DFT/TD-DFT (using the $\Delta$SCF approach, which will be referred to as $\Delta$DFT or $\Delta$B3LYP where appropriate throughout the rest of this chapter) and $G_0W_0$/BSE. The limitations on the computational resources in the $G_0W_0$ calculations meant that the maximum oligomer length focused on was four monomer units.

The computational workflow consisted of three steps. Firstly, the structures of all of the oligomers were optimised using DFT with the B3LYP functional and the def2-TZVP basis set. These DFT-optimised structures were then used as input to calculate the IP and EA values of the oligomers using a range of approaches including $G_0W_0$. Finally, the vertical optical gaps of the oligomers were investigated using TD-DFT (with B3LYP usually, as in the ground state calculations) and BSE. The BSE calculations used the quasiparticle energies obtained from the $G_0W_0$ calculations. For the excited state calculations the first 4 roots were solved for, the lowest of which was taken to be the vertical optical gap. All (TD-)DFT calculations were performed using Turbomole 7.01 and all $G_0W_0$ and BSE calculations using Turbomole 7.3. In most calculations Ahlrich’s basis sets were used. For the oligomer calculations the def2-SVPD basis set was used and for select systems, where computationally tractable, larger basis sets such as def2-TZVPP and def2-QZVPP were used as well. For the monomers I also used Dunning basis sets (aug-cc-pVDZ, aug-cc-pVTZ, aug-cc-pVQZ and aug-cc-pV5Z) in order to extrapolate the IP and EA to the complete basis set (CBS) limit.
7.3 Results and Discussion

Due to the computational costs of running Green’s function-based calculations they are usually limited to small basis sets. In the oligomer $G_0W_0$ calculations I used the def2-SVPD basis set instead of the recommended def2-TZVPP or def2-QZVPP basis sets, as calculation using the latter were too computationally expensive for most of the systems I looked at. With this basis set it was possible to calculate the IP and EA of oligomers up to the tetramers, however, in order to confirm that this is a satisfactory level of theory for such calculations, I ran a series of calculations on the monomers with larger Dunning basis sets in order to obtain the CBS limit. The CBS limit was obtained by extrapolation from the linear regression between one over the cubed cardinality number (where cardinality number is the number of basis functions in that basis set) and the relevant property.

The $-\text{IP}$ and $-\text{EA}$ values predicted for the monomers using def2-SVPD and those from the linear extrapolation are tabulated below in table 7.1.

<table>
<thead>
<tr>
<th>Molecule</th>
<th>$-\text{IP} / \text{eV}$ (CBS limit)</th>
<th>$-\text{IP} / \text{eV}$ (def2-SVPD)</th>
<th>$-\text{EA} / \text{eV}$ (CBS limit)</th>
<th>$-\text{EA} / \text{eV}$ (def2-SVPD)</th>
<th>$\Delta\gamma / \text{eV}$ (CBS limit)</th>
<th>$\Delta\gamma / \text{eV}$ (def2-SVPD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Benzene</td>
<td>-9.41</td>
<td>-8.82</td>
<td>1.02</td>
<td>1.82</td>
<td>10.4</td>
<td>10.6</td>
</tr>
<tr>
<td>Pyridine</td>
<td>-9.70</td>
<td>-8.96</td>
<td>0.44</td>
<td>1.26</td>
<td>10.1</td>
<td>10.2</td>
</tr>
<tr>
<td>Pyrrole</td>
<td>-8.44</td>
<td>-7.83</td>
<td>0.50</td>
<td>2.29</td>
<td>8.9</td>
<td>10.1</td>
</tr>
<tr>
<td>Thiophene</td>
<td>-9.11</td>
<td>-8.50</td>
<td>0.89</td>
<td>1.74</td>
<td>10.0</td>
<td>10.2</td>
</tr>
<tr>
<td>Pyrazine</td>
<td>-9.72</td>
<td>-8.99</td>
<td>-0.23</td>
<td>0.62</td>
<td>9.5</td>
<td>9.6</td>
</tr>
</tbody>
</table>

The $-\text{IP}$ predicted by def2-SVPD was consistently less negative than that at the CBS limit, where the difference ranged from 0.6 – 0.75 eV. The $-\text{EA}$ values predicted by def2-SVPD were around 0.8 – 0.85 eV more positive than the CBS limit results, except in the case of pyrrole where the difference was much larger. The larger discrepancy in values for pyrrole was likely due to an issue with the DFT starting point and not the $G_0W_0$ calculations as the LUMO values from the B3LYP/def2-SVPD calculation lay in the positive spectrum. This only occurred for the pyrrole monomer and the issue was not encountered by the longer oligomers of pyrrole, the LUMOs of which were negative. In table 7.1, the fundamental gap values calculated by def2-SVPD and extrapolated to the CBS limit are also shown. The fundamental gap appears to converge much more rapidly with basis set size and the basis set size increase led to an overall shift of the absolute $-\text{IP}$ and $-\text{EA}$ values. Apart from in the case of the pyrrole monomer, the difference between the def2-SVPD fundamental gap values and those at the CBS limit was less than 0.25 eV.

A similar analysis was performed for the optical gap values calculated with $G_0W_0$/BSE, shown in table 7.2 below. Here instead of extrapolating to the CBS, the aug-cc-pV5Z values were used as proxy for the CBS limit, as the difference between those values and the aug-
Table 7.2: Optical gap values for the benzene, pyridine, pyrrole, thiophene and pyrazine monomers with the def2-SVPD basis set and at the (pseudo-)CBS limit.

<table>
<thead>
<tr>
<th>Molecule</th>
<th>$\Delta_o$ / eV (def2-SVPD)</th>
<th>$\Delta_o$ / eV (aug-cc-pVDZ)</th>
<th>$\Delta_o$ / eV (aug-cc-pVTZ)</th>
<th>$\Delta_o$ / eV (aug-cc-pVQZ)</th>
<th>$\Delta_o$ / eV (aug-cc-pV5Z)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Benzene</td>
<td>4.77</td>
<td>4.69</td>
<td>4.66</td>
<td>4.66</td>
<td>4.66</td>
</tr>
<tr>
<td>Pyridine</td>
<td>4.12</td>
<td>4.08</td>
<td>4.06</td>
<td>4.07</td>
<td>4.08</td>
</tr>
<tr>
<td>Pyrrole</td>
<td>5.32</td>
<td>4.30</td>
<td>4.53</td>
<td>4.62</td>
<td>4.64</td>
</tr>
<tr>
<td>Thiophene</td>
<td>5.23</td>
<td>4.71</td>
<td>4.71</td>
<td>4.71</td>
<td>4.68</td>
</tr>
<tr>
<td>Pyrazine</td>
<td>3.32</td>
<td>3.27</td>
<td>3.24</td>
<td>3.25</td>
<td>3.25</td>
</tr>
</tbody>
</table>

cc-pVDZ was always less than 0.05 eV, making it unnecessary to perform an extrapolation; however, this was not true for pyrrole. The overall difference in these values and the def2-SVPD was below 0.15 eV for benzene, pyridine and pyrazine. In the case of pyrrole and thiophene, the difference was considerably larger. While in the case of pyrrole, this can be attributed to the positive LUMO in the DFT calculation, this does not occur for thiophene and thus the problem must stem from the BSE calculation and not the underlying $G_0W_0$ quasiparticle spectrum.

7.3.1 Trends in $-\text{IP}$ and $-\text{EA}$

The trends in predicted $-\text{IP}$ and $-\text{EA}$ values were analysed by plotting them as a function of one over the oligomer length. These trends are shown below in fig. 7.1. The expected trend in which the $-\text{IP}$ became more positive and $-\text{EA}$ more negative as the oligomer length increased was followed by all of the materials. The $G_0W_0$ and $\Delta B3LYP$ methods predicted $-\text{IP}$ and $-\text{EA}$ values with relatively close agreement. The KS-B3LYP HOMO and LUMO energies however, though displaying the same trend as $G_0W_0$ and $\Delta B3LYP$, lay much further away in energy. This sheds light on the validity of the commonly made approximation of the HOMO and LUMO as $-\text{IP}$ and $-\text{EA}$ as in this case, it does not appear to be particularly accurate. There was a small but discernible difference between the slopes in the $G_0W_0$ and $\Delta B3LYP$ trend lines; $\Delta B3LYP$ consistently predicted a larger change in $-\text{IP}$ and $-\text{EA}$ with oligomer length than $G_0W_0$ did. This may be a consequence of the $\Delta B3LYP$ exchange-correlation functional decaying faster than the $1/r$ dependence of the exact functional and thus, the resulting over-delocalisation error. To test this hypothesis the same calculations were run using two functionals analogous to $\Delta B3LYP$ but with different amounts of Hartree-Fock exchange, BLYP (0%) and BHLYP (50%). As expected, BLYP suffered more dramatically from the over-delocalisation error as seen by a steepening of the slope in the $-\text{IP}/-\text{EA}$ v. one over oligomer length plot lines in fig. A.20. $\Delta BHLYP$ improved upon $\Delta B3LYP$ trendlines, decreasing the slope and bringing the trendlines closer to those calculated by $G_0W_0$, shown in fig. A.19.
7.3. RESULTS AND DISCUSSION

7.3.2 Trends in Fundamental Gap and Optical Gap

In fig. 7.2 the fundamental gap values predicted by KS-B3LYP, ΔB3LYP and $G_0W_0$ as a function of the oligomer length are shown. Below it are the optical gaps predicted by TD-DFT and $G_0W_0$/BSE. Again, the expected trend of fundamental and optical gap decrease with increasing oligomer length was followed by all of the materials with all of the methods. The fundamental gap predicted by ΔB3LYP and $G_0W_0$ were very similar. However, in line with the previous observation for −IP and −EA, the KS gap appears to be too small to be comparable to the fundamental gap. This is a common problem in Kohn-Sham DFT, as discussed in section 2.2. The fact that the virtual orbitals feel the same KS potential, i.e. the field of $N-1$ electrons, as the occupied orbitals means that the KS HOMO-LUMO gap more closely resembles the energy required to create an interacting exciton than a non-interacting electron-hole pair. It follows that the KS gap is a better model for the lowest excitation energy that the fundamental gap. In the optical gap calculations, the $G_0W_0$/BSE-predicted optical gaps were consistently lower than those calculated with TD-DFT for all materials except phenylene. Despite this, the predicted optical gaps were relatively close and overall, there was a reasonably strong linear correlation between the optical gaps predicted by the two methods as shown in fig. 7.4. In line with the discussion above, the use of the KS-B3LYP
HOMO-LUMO gap as an approximation for the optical gap performed well and although the optical gaps were higher than those calculated with TD-DFT and $G_0W_0$, they were in much closer agreement than the fundamental gap predictions.

Figure 7.2: Fundamental gap v. oligomer length; blue = DFT, red = $G_0W_0$, green = KS.
7.3. RESULTS AND DISCUSSION

Figure 7.3: Optical gap v. oligomer length; blue = (TD-)DFT, red = $G_0W_0$/BSE, green = KS.

Figure 7.4: Correlation between the exciton binding energy values calculated with $\Delta$B3LYP/TD-B3LYP and $G_0W_0$/BSE; yellow = pyrazine, blue = pyridine, red = pyrrole, green = phenyl, black = thiophene.

7.3.3 Comparison of Exciton Binding Energies

The exciton binding energy (EBE), as discussed in chapter 1, is the energy required to separate an interacting electron-hole pair and the difference between the fundamental gap and optical gap. The relevance of this quantity to photovoltaics and photocatalysis is that it controls the
likelihood of an exciton separating into the free charge-carriers following an optical excitation. As the $G_0W_0$ fundamental gap is the starting point for the $G_0W_0$/BSE calculations, the predicted fundamental gaps and optical gaps are internally consistent. This is not the case for the (TD-)DFT calculations as they use the DFT orbital energies as input and not the actual fundamental gap as input. Nevertheless, the two methods predicted very similar exciton binding energies and followed the same overall trends (fig. 7.4).

7.3.4 Comparison of Materials

While the purpose of this chapter is to validate a method that has been used throughout the work in this thesis, the calculations can also reveal properties associated with the investigated materials. Figure 7.5 compares the evolution of $-\text{IP}$ and $-\text{EA}$ values of the different materials with increasing oligomer length. $G_0W_0$ and $\Delta$B3LYP predict the same ordering of materials. Intuitively one would expect the most electron-poor monomers to have the deepest (most negative) $-\text{IP}$s and indeed, this is what was seen for pyrazine and pyridine. Conversely, the electron-rich pyrrole and thiophene oligomers led to the shallowest, least negative $-\text{IP}$ values. The deepest $-\text{EA}$ values were for electron-poor pyrazine and the shallowest were for electron-rich pyrrole, however, not all of the materials $-\text{EAs}$ were in the exact order one would expect based on how electron-rich or electron-poor the materials are. Overall, however, the order of IP and EA for the phenylene, thiophene and pyridine oligomers followed the same trends seen previously in experiment.\textsuperscript{169} However, these experimental values should be tentatively compared to these results due to them being calculated by two different methods, ultraviolet PES (UV-PES) and CV.

![Graph showing comparison of IP and EA values](image1)

Figure 7.5: Comparison of the trends in $-\text{IP}$ and $-\text{EA}$ values of the different materials as a function of one over the oligomer length, as predicted by the different methods; yellow = pyrazine, blue = pyridine, red = pyrrole, green = phenyl, black = thiophene.

The comparison of the fundamental gaps of the different materials with oligomer length
7.3. RESULTS AND DISCUSSION

is shown in fig. 7.6. Both $G_0W_0$ and $\Delta$B3LYP predicted that thiophene had the lowest fundamental gap and the same general ordering was followed by the two methods although all of the fundamental gaps were very similar in energy. In fig. 7.6 the optical gaps are also shown. While there was good agreement in most of the trendlines between methods, phenylene exhibited peculiar behaviour. The optical gap trendline followed by phenylene when calculated by TD-DFT plateaued similarly to that of pyridine. However, upon switching to $G_0W_0$/BSE, the optical gap of the phenylene monomer increased and the general trendline was much straighter.

![Graphs showing fundamental and optical gaps calculated by different methods](image)

Figure 7.6: Fundamental gaps calculated by DFT (top left) and $G_0W_0$ (top right), and optical gaps calculated by TD-DFT (bottom left) and $G_0W_0$/BSE (bottom right); yellow = pyrazine, blue = pyridine, red = pyrrole, green = phenyl, black = thiophene.

From such plots one can usually deduce the so-called conjugation length – the length at which the calculated properties of a polymer converge with respect to chain length. The properties at the polymer limit can then be extrapolated. Whether a conjugated polymer has a long or short conjugation length has a lot to do with the coupling between monomers. For instance, a polymer such as poly-$p$-phenylene experiences strong repulsion and high torsional
angle between monomers. Therefore, the inclusion of an additional monomer only makes a
difference to the overall properties at small chain-lengths and the properties converge quite
quickly. In such a case, one would say that the polymer has a short conjugation length.
On the other hand, an oligomer such as poly-thiophene exhibits strong coupling between
monomer units and as such, increasing the chain length modifies the properties; therefore, it
has a long conjugation length. In the TD-DFT plots in fig. 7.6 this is what is observed, with
the black thiophene optical gap line extending without plateauing and conversely, the green
phenylene line beginning to flatten at $1/n = 0.5$. For the BSE-calculated results, however,
these trends are not the same. Phenylene appears to continue in a straight line like thiophene,
indicating a longer conjugation length. Pyrrole (red) would similarly be expected to have a
longer conjugation length with respect to the optical gap, however, begins to plateau faster
than phenylene.

These observations really highlight the obstacles I encountered whilst carrying out this
work. When studying polymers, especially with high-level methods such as GW and BSE,
often a smaller chain (oligomer) polymer is taken instead of using periodic boundary conditions
to reduce the computational costs. GW was particularly limiting so only up to the tetramer
for all of the materials was studied. In addition to the fact that the analysis is based on only
four data points, the less than satisfactory basis set that was used – def2-SVPD instead of the
recommended def2-TZVPP – meant that any fitting performed would be lacking in reliability.
Overall, the comparison of the optical gaps of materials is not that informative when the high-
level calculations cannot be carried out to the required standard. Usually GW/BSE would be
considered the gold-standard method for optoelectronic calculations but unfortunately, this
was not the case under the restrictions enforced by the available resources and the limitations
of the code used.

7.4 Conclusions

For this study, I set out to validate the (TD-)DFT method for calculating IP, EA and optical
gaps using the $G_0W_0$ method in Turbomole. Firstly, the effect of the choice of basis set
was tested, comparing the small basis set (def2-SVPD) results to those at the CBS limit for
the monomers of the investigated materials. Although the recommended basis set for these
calculations are def2-TZVPP or def2-QZVPP, in order to perform them on larger molecules
I had to opt for the smaller def2-SVPD basis set. While there was a shift in absolute values
when using the def2-SVPD basis set rather than the CBS limit, the overall trends in IP, EA
and optical gap were the same. The pyrrole monomer appeared to be problematic for these
calculations, which was possibly related to the negative LUMO in its quasiparticle spectrum.

For the oligomers, the IP and EA calculated by $\Delta_{DFT}$ and $G_0W_0$ were in close agreement,
becoming respectively shallower and deeper with increasing oligomer length. The steeper slope
of the DFT results highlighted the over-delocalisation error encountered by B3LYP, which was
7.4. CONCLUSIONS

confirmed through the use of the BHLYP and BLYP functionals. The fundamental and optical gaps were shown to decrease with oligomer length when predicted by $\Delta$DFT, $G_0W_0$ and from the KS orbital energies. These results clearly showed that the KS orbital energies are in fact a poor approximation for the fundamental gap. However, as suggested by Baerends, the KS gap can be a good approximation for the optical gap and indeed this was reflected in these results.

Finally, the five different materials were compared, displaying the expected trend of oligomers containing more electron-rich monomers leading to shallower $-\text{IPs}$ and those containing electron-poor monomers leading to deeper $-\text{EAs}$. However, overall, the expected order in the $-\text{IPs}$ and $-\text{EAs}$ relative to how electron-rich/electron-poor the material was not adhered to. It is difficult however, to form insightful conclusions about these materials as polymers when limited to such short oligomer lengths. As the maximum oligomer size was a tetramer, any conclusions about the conjugation length of the properties will be inherently flawed. Despite this, the results did show the expected trends when changing the oligomer length.

Carrying out the work in this chapter was far from simple, as a number of issues were encountered along the way. The aim was to validate the DFT method, however, the limiting computational cost of the $G_0W_0$ calculations and the inconsistencies encountered using the algorithm implemented in Turbomole 7.3 were large obstacles in achieving this goal. One of the main issues I encountered was that the recommended minimum basis sets (def2-TZVPP and def2-QZVPP) could not be used to calculate the properties of the long oligomers. I wanted to ultimately validate the $\Delta$DFT method for extended conjugated systems however, the most expensive calculations I could perform using $G_0W_0$ were on tetramers with the def2-SVPD basis-set. Additionally, when performing the extrapolation to the CBS limit for the different properties, it was only possible to use the monomers of the oligomers. Unfortunately, attempting to describe the electronic structure of small molecules can often be more problematic than for larger systems. Therefore, the trade-offs that had to be made in order to actually run the $G_0W_0$ calculations would have been detrimental to the results. As a result of these obstacles, I was limited to short oligomers, which were not really representative of the materials I was aiming to investigate, and to a lower level of accuracy than desired in the calculations.

In addition to the challenges I faced in performing the calculations, there was no solvation model implemented in the program. I carried out the work in the previous chapters of this thesis with real-life applications in mind, where modelling the materials in the gas-phase would not be sufficiently realistic. The aim was to use an alternative program in which a solvation model for $G_0W_0$ was implemented once the gas-phase calculations had been completed to a satisfactory level, however, the project never reached a stage that I deemed successful.

In addition to the limitations related to computational resources, I encountered a lot of issues with apparent bugs in the $G_0W_0$ implementations in different Turbomole versions. There were many inconsistencies in the results when I originally used Turbomole 7.1, such
as a lack of reproducibility in the results when applying symmetry to the calculations. This necessitated a lot of going back and forth between the developers of the code before finally redoing the calculations using the latest version of Turbomole at the time (Turbomole 7.3).

Overall, even though performing these calculations was not straightforward and the desired level of accuracy was not attainable with the computational resources available to me, both the $G_0W_0$/BSE and (TD-)DFT calculations produced mostly consistent results (with some exceptions) and followed the expected trends for different materials and oligomer lengths. Since the time this work was carried out, a newer version of the program has been released with an improved algorithm for GW calculations, therefore, future investigations using this version may prove less problematic and allow for an even more accurate and informative validation of the (TD-)DFT methods.
Chapter 8

Conclusions and Perspective

In this thesis, a multi-level, theoretical analysis of organic conjugated materials has been conducted, bringing to light the advantages and disadvantages of both highly accurate yet computationally expensive methods and inexpensive and fast yet potentially lower accuracy methods.

In chapters 3 to 6, the high-throughput virtual screening (HTVS) method previously established in our group\textsuperscript{152} was employed to take a zoomed out and more general view of some organic conjugated materials that may be used in photoactive applications. While this method worked very well in some cases, for example, in the prediction of properties of oligomers (chapters 5 and 6) and small molecules (chapter 3), in others, such as the dyes (chapters 3 and 4), it was less successful and encountered a range of problems.

Chapter 3 discussed the consequences of using different fitting parameters to perform the calibrations and the importance of ensuring the transferability of the chosen calibration parameters to the structures of interest. For example, if both dyes and monomers are included in the linear regression, the accuracy of using these fitting parameters to calibrate the properties of dyes must be examined in order to ensure the output of accurate results. In general, the IP and EA calibrations worked well for the small molecules, dyes and polymers, however, the optical gap calculations proved to be less straightforward.

When performing the calibration on the optical gaps of the diketopyrrolopyrrole (DPP) dyes, the strong linear correlation that was seen for the IP and EA of the dyes and previously in the calibration of small molecules and polymers was not present. Modifications in the calibration procedure (i.e. changes in basis set/functional in the DFT calculations, inclusion of the monomers, using a different semi-empirical method) led to results that indicated that dyes containing sulfur were causing the deviation from the expected linear relationship.

A solution to this problem would likely require a model that takes into account chemical structure, for instance, a model that uses each material’s Morgan fingerprint and xTB-calculated properties as variables.

The removal of the sulfur-containing dyes meant their exclusion from the subsequent analyses in terms of photoactive applications, limiting the extent of the study. Despite the obstacles
met, some interesting conclusions about the building block sequence were elucidated from the results. A good proportion of the dyes were found to be good potential candidates for dye-sensitized proton reduction and dye-sensitized solar cells (DSSCs). The ABCBA dye class was particularly interesting, with the results highlighting the dominating influence of the building block in the B position on the overall dye properties. The effect of asymmetry also revealed some interesting trends, which in combination with the former finding could be useful in future materials discovery for the fine-tuning of properties.

The HTVS method worked very well for the small aromatic molecules and polymers. Here, the benefits of using an HTVS method are truly exemplified. Focusing on the polymers, it was possible to explore a wide range of chemical properties efficiently and to a reasonably good degree of accuracy. I therefore was able to explore different polymer compositions by variation of the types of monomers and their concentrations in the oligomers, and also the effects of conformerism and sequence isomerism. This wider, more general view allows one to zoom out and observe what features in composition give rise to particular properties, and then zoom in on areas of interest depending on what the desired properties are.

In chapter 6, the effect of sequence in ordered co-oligomers was successfully investigated and some interesting results obtained. The block co-oligomer sequence was found to give rise to lower optical gaps and the potential for intramolecular charge-separation, dependant on the alignment of the corresponding homo-oligomer potentials, indicating that it is worth investigating block co-polymers further for photoactive applications. In future works the library could be expanded, focusing on monomers based on the ones that gave rise to desired properties in this work. If the aim was to screen much larger numbers of materials than in the HTVS chapter in this thesis, it would likely require the use of machine learning, such as in previous work by our group. This would help overcome issues in computational resources that, in some cases, necessitated focusing on the octamers rather than the dodecamers.

Chapter 7 used the high-level many-body perturbation theory method $G_0 W_0$ in an attempt to benchmark DFT-calculated properties and validate the $\Delta \text{DFT}$ method that is used extensively in our group and in this thesis. Unfortunately, this chapter really underlined the pitfalls of using high-level methods beyond DFT. The problems experienced arose for a number of reasons, mainly due to the inability to perform such computationally demanding calculations within the computational resources available. The $G_0 W_0$ calculations required a lot of memory, meaning that firstly, the output of results was extremely slow in comparison to the other methods. In many cases, these calculations simply crashed, limiting the study that was supposed to be on polymers to oligomers of only 4 monomer units. The results obtained could not, therefore, be used to derive any useful conclusions on the properties of polymers. It was also only possible to carry out calculations on the tetramers with a smaller basis set than recommended in the literature, which again, affected the expected accuracy of the calculation. Furthermore, the validation of this small basis set, by comparison of the results to those extrapolated to the CBS limit, was not very informative, as it could only be
carried out using monomers. Computational resources aside, potential bugs in the algorithm in different versions of the software led to inconsistencies in results. Since completing the work in chapter 7, a newer version of Turbomole with a new algorithm for the GW methods has been released, possibly providing a solution to at least some of the issues I experienced. It is also worth noting that the lack of a solvation model in the Turbomole implementation of $G_0W_0$ and most other easily accessible codes meant that any results obtained would not be a satisfactory description of polymer properties near the polymer-water interface, which is relevant to a lot of the applications I was interested in.

Additionally, it would be ideal to compare the values for the optoelectronic properties that have been calculated throughout this thesis to equivalent experimental measurements, however, it is not that simple when looking at such large molecule libraries. In chapter 6, finding the experimental HOMO and LUMO values for PCBM, PC$_{70}$BM and PN1 was not trivial, as firstly, it was necessary to find samples that were prepared under the same conditions and were in the same form as each other, e.g. they all had to be thin-film samples, and secondly, they needed to be measured in a similar solvent as the one the calculations were performed in. In chapter 7, the optical properties of some of the oligomers were reported, however, the references measured the properties in different ways and therefore, were not as comparable as one would ideally like them to be. In chapter 3, chapter 4 and chapter 6 I wanted to make comparisons with experiment, however, the vastness of the libraries meant that it was challenging to find comparable references. This would really have strengthened the case for the reliability of the methods discussed, however, the minimal comparisons that have been made and the fact that the high-throughput screening method has been benchmarked previously are still sufficient for these investigations.

Some key takeaways from the results of this thesis could be applied to future research by both experimental and computational chemists. From a materials design perspective, this work has highlighted many useful building blocks/monomers for organic materials for tuning for particular properties; for example, the tetrazine monomer leading to low optical gaps in chapter 6 could be useful to experimentalists aiming to fine tune the IP whilst maintaining a low optical gap. The revelation of the effects of different sequences on the overall properties of dyes and polymers could also be very useful for particular applications. The contrast in the properties of the symmetric and asymmetric dyes in chapter 4, and the possibility for intramolecular charge separation in block co-polymers in chapter 6 are two such examples, as these findings could inform chemists aiming to discover materials for particular photoactive applications. From a solely computational standpoint, the methods and results in this thesis are also rather valuable. In chapter 4, the linear relationship between the properties of the ABCBA/ACA$'$ dyes and their corresponding homodyes opens up the possibility for performing fewer calculations in order to construct an ideal monomer library for an HTVS study, as the number of dyes is greatly reduced when only the homodyes are considered. Calculating the properties of the homodyes alone could provide insight into the heterodye properties and
allow researchers to make better informed decisions about future screenings using the HTVS method, which would help prevent the possibility of performing thousands of calculations without producing fruitful results. The HTVS method established by our group has also been tested on a number of different organic systems, and this has in turn highlighted where the method works very well and where it experiences some downfalls. This knowledge would certainly be useful to researchers interested in implementing this method. Even in chapter 7, although the calculations were indeed challenging and computationally expensive, the correct trends were demonstrated and the caveats in using such high-level theory methods were highlighted. Ideally, high-level and low-level methods could be used in tandem, for example, using the efficient and large output of HTVS to narrow down a list of potential candidates for a desired application and then corroborating these results and obtaining a detailed picture of the candidate properties using a GW method. In this way one could carry out a well-rounded and detailed analysis of a particular class of materials, providing an excellent starting point for subsequent experimental investigations.

Overall, this work has provided a multitude of insights with a range of applications in both computational and experimental chemistry, truly highlighting the wide range of strengths in computational chemistry methods.
Appendix A

A.1 Accessing Near-DFT Results by Calibration of Properties from Semi-Empirical Calculations Against DFT

Figure A.1: HOMO (left) and LUMO (right) of 2-dimethylamino-3-amino-1,4-benzoquinone, one of the outlying quinones in fig. 3.6.
Figure A.2: HOMO (left) and LUMO (right) of 2,5-bis(dimethylamino)-1,4-benzoquinone, one of the outlying quinones in fig. 3.6.

Figure A.3: HOMO (left) and LUMO (right) of 2-amino-3-hydroxy-1,4-benzoquinone, one of the outlying quinones in fig. 3.6.

Figure A.4: DFT-calculated v. xTB-calculated $-\text{IP}$ and $-\text{EA}$ for the ABCBA, ACA’ and ACA dyes, calibrated using the fitting data obtained from the dyes alone solvated in benzene. The data prior to calibration is in orange and post-calibration is in blue, the red dashed line is the $x=y$ line and the line of best fit used to calibrate the data in black.
Figure A.5: DFT-calculated vs. sTDA-calculated optical gaps for the ABCBA, ACA’ and ACA dyes with sulfur (left) and without sulfur (right), calibrated using the fitting data obtained from the dyes alone and solvated in benzene. The colour scheme is the same as in fig. A.4
A.2 Finding Diketopyrrolopyrrole Dyes for Dye-Sensitized Water Splitting and Solar Cells

Figure A.6: Two-dimensional histograms of the property spaces spanned by $-\text{IP}$ v. $-\text{EA}$ (top row), $-\text{IP}$ v. $\Delta_o$ (centre row), $-\text{EA}$ v. $\Delta_o$ (bottom row) for the ABCBA, ACA and ACA’ dyes (left to right).
Figure A.7: Convex hulls enclosing the property space of $-\text{IP}$ v. $-\text{EA}$ (left), $-\text{IP}$ v. $\Delta_o$ (centre), $-\text{EA}$ v. $\Delta_o$ (right) for the ABCBA, ACA and ACA$'$ dyes.

Figure A.8: Convex hulls of the $-\text{IP}$ and $-\text{EA}$ of the ABCBA dyes containing the building block in fig. 4.11(a)

Figure A.9: Convex hulls of the $-\text{IP}$ and $-\text{EA}$ of the ABCBA dyes containing the building block in fig. 4.13(a)(iii)
Figure A.10: Scatter plot showing the $-\text{IP}$ v. $-\text{EA}$ slice of property space for the most common A and B units of the ABCBA dyes, when residing on the A-site.
Table A.1: Effect of phosphonic acid, carboxylic acid and cyanocarboxylic acid anchoring groups on the properties of dyes. SMILES for each dye ID can be found in the supplementary information of the published work.\textsuperscript{225}

<table>
<thead>
<tr>
<th>Property</th>
<th>Dye ID</th>
<th>Anchoring Group</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>No anchor</td>
</tr>
<tr>
<td>IP</td>
<td>37382</td>
<td>3.48</td>
</tr>
<tr>
<td></td>
<td>4386</td>
<td>3.46</td>
</tr>
<tr>
<td></td>
<td>7121</td>
<td>3.21</td>
</tr>
<tr>
<td></td>
<td>ACA 174</td>
<td>3.48</td>
</tr>
<tr>
<td></td>
<td>ACA 174</td>
<td>3.48</td>
</tr>
<tr>
<td></td>
<td>ACA 190</td>
<td>3.48</td>
</tr>
<tr>
<td></td>
<td>ACA' 2361</td>
<td>3.48</td>
</tr>
<tr>
<td></td>
<td>ACA' 296</td>
<td>3.48</td>
</tr>
<tr>
<td></td>
<td>ACA' 2361</td>
<td>3.48</td>
</tr>
<tr>
<td></td>
<td>ACA' 296</td>
<td>3.48</td>
</tr>
<tr>
<td>EA</td>
<td>37382</td>
<td>3.48</td>
</tr>
<tr>
<td></td>
<td>4386</td>
<td>3.48</td>
</tr>
<tr>
<td></td>
<td>7121</td>
<td>3.48</td>
</tr>
<tr>
<td></td>
<td>ACA 174</td>
<td>3.48</td>
</tr>
<tr>
<td></td>
<td>ACA 190</td>
<td>3.48</td>
</tr>
<tr>
<td></td>
<td>ACA' 2361</td>
<td>3.48</td>
</tr>
<tr>
<td></td>
<td>ACA' 296</td>
<td>3.48</td>
</tr>
<tr>
<td></td>
<td>ACA' 2361</td>
<td>3.48</td>
</tr>
<tr>
<td>Optical gap</td>
<td>37382</td>
<td>3.48</td>
</tr>
<tr>
<td></td>
<td>4386</td>
<td>3.48</td>
</tr>
<tr>
<td></td>
<td>7121</td>
<td>3.48</td>
</tr>
<tr>
<td></td>
<td>ACA 174</td>
<td>3.48</td>
</tr>
<tr>
<td></td>
<td>ACA 190</td>
<td>3.48</td>
</tr>
<tr>
<td></td>
<td>ACA' 2361</td>
<td>3.48</td>
</tr>
<tr>
<td></td>
<td>ACA' 296</td>
<td>3.48</td>
</tr>
<tr>
<td></td>
<td>ACA' 2361</td>
<td>3.48</td>
</tr>
<tr>
<td></td>
<td>ACA' 296</td>
<td>3.48</td>
</tr>
</tbody>
</table>
A.3 High-Throughput Virtual Screening of Block Copolymers

Figure A.11: Calibration curves for $-IP$ (left), $-EA$ (centre) and optical gap (left) for a subset of the oligomers, as calculated by (IPEA/sTDA-)xTB and B3LYP/DZP in benzene. Here the uncalibrated data points (blue), the calibrated points (yellow), the line of best fit (black) and the $x=y$ line (red) are shown.

Figure A.12: Property-property relationships for calibrated xTB-calculated properties in benzene; $-IP$ v. $-EA$, $-IP$ v. $\Delta_o$ and $-EA$ v. $\Delta_o$ (left to right).

Figure A.13: Convex hulls for the calibrated properties of the different co-polymer sequences and homopolymers (dodecamer) in water; pink = homopolymers, red = AB, blue = AABB and green = AAAABBBB.
A.3. BLOCK CO-POLYMERS

Figure A.14: Convex hulls for the calibrated properties of the different co-polymer sequences and homopolymers (dodecamers) in benzene; pink = homopolymers, red = AB, blue = AABB, yellow = AAABBB and green = AAAAAABBBBBB.

Figure A.15: Convex hulls for the calibrated properties of the different co-polymer sequences and homopolymers (octamers) in benzene; pink = homopolymers, red = AB, blue = AABB, yellow = AAABBB and green = AAAAAABBBBBB.

Figure A.16: Scatter plots of the difference in IP between the two constituent homopolymers of each block co-polymer v. the difference in EA of the two constituent homopolymers for the octamers in water, the octamers in benzene and the dodecamers in benzene (left to right). The quadrants correspond to block co-polymers where the potentials of their corresponding homopolymers are either straddled or staggered. The yellow points represent the block co-polymers exhibiting the discussed charge-separated effect and the points in blue represent those that do not.
Figure A.17: $-\text{IP}/-\text{EA}$ of homopolymer A v. $-\text{IP}/-\text{EA}$ of homopolymer B corresponding to each co-polymer within the electronic constraints and driving force (top row) and within the optical and electronic constraints including a driving force (bottom row), with respect to the acceptor PC$_{70}$BM.

Figure A.18: $-\text{IP}/-\text{EA}$ of homopolymer A v. $-\text{IP}/-\text{EA}$ of homopolymer B corresponding to each co-polymer within the electronic constraints and driving force (top row) and within the optical and electronic constraints including a driving force (bottom row), with respect to the acceptor PN1.
A.4 Benchmarking GW against DFT

Figure A.19: $-\text{IP} / \text{EA} v. \text{one over oligomer length}$; blue = BHLYP DFT, red = $G_0W_0$, green = B3LYP KS.
Figure A.20: $-\text{IP}/-\text{EA}$ v. one over oligomer length; blue = BLYP DFT, red = $G_0W_0$, green = B3LYP KS.
Bibliography


(7) Green, M. A. AIP Advances 2013, 3, 112104.


(64) Wong, M. Y.; Zysman-Colman, E. Advanced Materials 2017, 29, 1605444.


BIBLIOGRAPHY


(99) Basis Set Exchange (BSE).


BIBLIOGRAPHY


(130) Ásgeirsson, V.; Bauer, C. A.; Grimme, S. Chemical Science 2017, 8, 4879–4895.


(136) Li, S.-L.; Jiang, K.-J.; Shao, K.-F.; Yang, L.-M. *Chemical Communications* **2006**, *2792–2794.


Acknowledgements

I would above all like to thank my supervisor Professor Martijn Zwijnenburg for his supervision, scientific guidance and pastoral support over the past three years. For always making time to provide me with explanations, answering every single one of my many questions, making time to read all of my writing, motivating me and helping me progress as an individual. I am extremely grateful for all of the knowledge, skills and opportunities I gained under the supervision of Martijn and it has been a pleasure working with him.

I would like to thank Dr. Liam Wilbraham for the constant help he offered me during his time in our group, introducing me to new concepts and ideas, taking me onto projects that allowed me to develop a huge range of skills, without which I would not have the same opportunities I do now, and for being an overall great desk-side companion. His contribution to the group’s research really helped shape my PhD into work that I am truly proud of – and it was great to have someone to chat to from day one.

To Dr. Louise Price, for all of the great conversations we had and all of the emotional support she provided throughout the entirety of my PhD, particularly the first half. At times where I was really down and thought I could not continue Louise was there to offer advice and motivation. I will miss our chats completely unrelated to anything scientific and distracting the entire office.

I want to thank the rest of the past and current members of Zwijnenburg Group and Sally Price Research Group. This doctorate was funded by the UCL Department of Chemistry.

To my undergraduate lecturers at Queen Mary that inspired me and made me so passionate about my subject, Dr. Lebrasseur, Dr. Arseniyadis and Dr. Abrahams.

To my friends, especially Isaac, Andreas, Geena, Hannah, Waleed, Harman and Mahima, for their continual support. To Alberto, for the meticulous grammar and punctuation checks over the years. To my grandparents, who always take an interest in everything I am involved in – including computational chemistry! To Liam, for always being there for me. Lastly, to my mum for everything she has ever done for me.