

Use of Molecular Modelling for the Understanding of Self-Assembly Processes in Supramolecular Polymers

Antiopi Tatiani Politi

November 2020

A thesis submitted in partial fulfilment of the requirements of Liverpool John Moores
University for the degree of Doctor of Philosophy

Date of Submission	02-Nov-2020
--------------------	-------------

Publishing Year	2021
-----------------	------

Prepared with the permission of the School of Pharmacy and Biomolecular Sciences at Liverpool John Moores University.

*“I myself know nothing except just a little, enough to extract an argument from another man
who is wise and to receive it fairly.” -Theatetus, 161b*

Preface

This thesis is a collection of selected research performed by the author in the School of Pharmacy and Biomolecular Sciences at Liverpool John Moores University. The work is original, and no part of it has been submitted for any other degree.

Parts of the work in this thesis has been published in the following journal papers:

1. K. K. Kartha, N. K. Allampally, **A. T. Politi**, D. D. Prabhu, H. Ouchi, R. Q. Albuquerque, S. Yagai and G. Fernandez, *Chemical Science*, 2019, 10, 752-760.
2. D. S. Philips, K. K. Kartha, **A. T. Politi**, T. Kruger, R. Q. Albuquerque and G. Fernandez, *Angewandte Chemie-International Edition*, 2019, 58, 4732-4736.

Conference presentations:

1. Symposium in Computational Chemistry - (Edinburgh, UK), June 2019 (**Poster**)
2. IUPAC conference - (Paris, France), July 2019 (**Poster**)
3. Young Modellers Forum – (London, UK) November 2019 (**Talk**)

Abstract

There is an increasing need to further improve and obtain new materials for biomedical and technological applications. The tools and intricacies of material discovery and design have been exponentially multiplying, especially in recent years, opening the way for new discoveries for next generation materials.

This thesis investigates *in silico* conjugated organic supramolecular assemblies with the use of selected tools at the molecular level. The computational examination entails the geometrical and architectural design of the assemblies, and the study of their energetic details. A directed focus on OPE oligo(p-phenyleneethynylene)s, a photoisomerizing wire aggregate, and a series of porphyrinic assemblies is made in the evolution of this examination. These are versatile structures for supramolecular assembly, able to bind via a combination of non-covalent interactions, with interesting applications. The OPEs form 1D (one dimensional) wires with notable electronic properties. Furthermore, photoisomerizing aggregates have sparked abundant interest due to the control of assembly via light induction. Lastly, porphyrins are highly conjugated and readily functionalised molecules with the ability to form supramolecular assemblies from one to three dimensions. Porphyrin aggregates are also offered in crucial applications such as in photodynamic therapy, as well as an immensely wide and versatile spectrum of applications in further scientific sectors.

The first part of this thesis assesses the recent relevant computational methods through a benchmark study for their accuracy and computational cost for their predictivity of conjugated organic self-assemblies. Subsequently, the chosen assessed method is further evaluated by comparison with a set of published experimental data of OPE assemblies. The newly published method, GFN2-xTB, is then evaluated as the most efficient for this type of entity. The aggregation motifs and polymer properties two of two OPEs and a photoisomerizing 1D wire assembly are initially studied. Subsequently, a set of porphyrinic supramolecular aggregations first in the 1D, and then in 3D are studied by utilising reference experimental data for further validation of the predictive capacity of GFN2-xTB.

Lastly, for the first-time, chemical tuning recommendations are presented for promoting specific aggregations motifs. The combined results produce structure-property trends that purvey model assembly suggestions for supramolecular architectural synthesis and design.

Acknowledgments

Firstly, I would like to thank the School of Pharmacy and Biomolecular Sciences for a full scholarship, the ready resources in computer material, and training that was provided, as well as the support that I received as a student. I also thank my supervisor Dr Linda Seton for attending to my project when my supervisors moved away and advising me on every aspect of research along the way. I especially thank Dr Steve Enoch for firmly deciding on the contents of this thesis and guiding the results, and for choosing the structure for the MD analysis.

I additionally, thank the collaborating Dr Divya Philips and Kalathil Kartha experimentalists from the Fernandez group (Dr Gustavo Fernandez) in Münster University, for the synthesis and analytical experiments carried out for the content of Chapter 5 in this work. Their time, data provided, and expertise have been immensely valued.

Furthermore, a special acknowledgement is dedicated the astrophysics research institute of LJMU, and particularly Dr Daniel Harman for providing additional computational power when I needed it to progress my work. They have very generously given prompt technical support and ample processor power for me to conclude the results of this thesis in half the time that I would otherwise.

Finally, I thank my family and friends in Greece, Scotland and Liverpool for giving me all the material and emotional support I ever needed and more, to help me remain calm and motivated, especially through the turmoil of a pandemic crisis.

I also thank anyone who reads this thesis, I hope you enjoy it.

Nomenclature

Summary of some of the most commonly used terms within this thesis.

Term	Description
Å	Ångstrom (equal to 10^{-10} meters)
ρ	A cooperativity factor (no units)
BE	Binding Energy
B3LYP	Becke, Three-Parameter, Lee-Yang-Parr: DFT Functional
B3LYP-D3	Becke, Three-Parameter, Lee-Yang-Parr: DFT Functional with the D3 dispersion correction scheme
BM-BE	Benchmark Binding Energy
Calc-BE	Calculated Binding Energy
CCSD(T)	Coupled Cluster Double (Triple) level of theory: <i>Ab Initio</i> method
CREST	Conformer Rotamer Ensemble Sampling Tool
DFT	Density Functional Theory
DFTB	Density Functional Theory, Tight Binding
DFT-D3	Density Functional Theory, with the D3 dispersion correction scheme
G	Gibbs free energy
GFN-xTB	Geometry Frequency Non-covalent eXtented Tight Binding
GFN2-xTB	Geometry Frequency Non-covalent version 2 eXtented Tight Binding
H	Enthalpy
HF	Hartree Folk: <i>Ab Initio</i> method
kcal	Kilo calories
LC	Ligand Centered
MD	Molecular Dynamics
mer	Monomer: molecular unit block of a supramolecular polymer
MTD	Meta-Dynamics
MLCT	Metal to Ligand Transfer
MC	Metal Centred
NCI	Non-covalent Interactions
OPE	Oligo(p-phenyleneethynylene)
NMR	Nuclear Magnetic Resonance
PM6	Parametric Method 6
PM6-D3	Parametric Method 6 with the D3 dispersion correction scheme
PM7	Parametric Method 7
PW6B95	Perdew-Wang-91 exchange and Becke-95 correlation: Hybrid meta exchange correlation functional
Pred-BE	Predicted Binding Energy
RMSE	Route Mean Square Error
S	Entropy
TD-DFT	Time Dependent Density Functional Theory
ZINDO	Intermediate Neglect of Differential Orbital by Ridley and Zerner: A semi-empirical method

Table of Contents

Preface	i
Abstract	ii
Acknowledgements	iii
Nomenclature	iv
Table of Contents	v
1. Introduction	1
1.1 Introduction.....	1
1.2 Motivation.....	1
1.2.1 Supramolecular Porphyrin Aggregations.....	1
1.2.2 Computational Predictivity for Supramolecular Aggregates and Synthetic Recommendations	4
1.3 Structure of the Thesis.....	5
2. Background Literature	7
2.1 Introduction.....	7
2.2 Supramolecular Architectural Design.....	8
2.2.1 Molecular Design for Dimensionality Control.....	8
2.3 Cooperativity Studies of Supramolecular Assemblies.....	12
2.4 Computational Approaches and Obstacles for Supramolecular Architectural Design .	13
2.5 Chapter Summary.....	18
3. Theoretical Background	19

3.1 Introduction	19
3.2 Fundamental Concepts of Quantum Chemistry	19
3.3 Hartree-Fock Theory	22
3.4 Density Functional Theory	24
3.4.1 Density Functional Approximations	26
3.4.1.1 <i>Local Density Approximation</i>	27
3.4.1.2 <i>General Gradient Approximation</i>	27
3.4.1.3 <i>Meta-GGA Functionals</i>	28
3.4.1.4 <i>Hybrid Functionals</i>	28
3.4.1.5 <i>Virtual-Orbital Dependent Functionals</i>	29
3.5 Semi-empirical Methods	29
3.5.1 MNDO-type Methods.....	30
3.5.2 Density Functional Tight Binding Methods.....	32
3.6 Dispersion Correction.....	35
3.7 Conformer-Rotamer Ensemble Sampling Tool.....	39
3.8 Chapter Summary.....	40
Benchmarking of Semi-Empirical Methods for the Modelling of Supramolecular olymers	41
4.1 Introduction	41
4.1.1 Semi-Empirical Methods for Predicting Molecular Interactions	41
4.1.2 The Relevant Benchmark Data Sets for NCIs – S66 and S30L	42
4.1.3 Notable Benchmark Literature for Semi-Empirical Methods on NCIs.....	43
4.2 Methodology	45

4.2.1	Computational Methods Investigated	45
4.2.2	Calculation of Binding Energies	45
4.2.3	Statistics	46
4.3	Results and Discussion	46
4.3.1	The S66 Data Set	46
4.3.1.1	<i>Analysis of Benchmark Calculations for the S66 Dataset</i>	47
4.3.2	The S30L Data Set Benchmark	53
4.3.2.1	<i>Analysis of Benchmark Calculations for the S66 Dataset</i>	53
4.4	Chapter Summary	60
5.	Computational Investigations into the Aggregation of One Dimensional Small Organic Conjugated Systems	61
5.1	Introduction	61
5.2	Methodology	64
5.2.1	Computational Methodology	64
5.2.1.1	<i>OPE Optimisations</i>	64
5.2.1.2	<i>C₁ and L₁ Optimisations</i>	64
5.2.1.3	<i>GFN2-xTB Optimisations</i>	65
5.2.1.4	<i>Computed Spectroscopy and Energy Level Diagrams</i>	65
5.2.2	Calculation of BE	66
5.3	Results and Discussion	66
5.3.1	The Role of H-bonding and Pre-Organisation in the Self-Assembly of Aggregates	67
5.3.1.1	<i>The Frontier Molecular Orbital Analysis of OPE</i>	67
5.3.1.2	<i>Geometrical Analysis</i>	69

5.3.1.3 Hydrogen Bonding Analysis of the Assembly.....	70
5.3.1.4 The Mechanism of Aggregation	71
5.3.1.5 The Performance of GFN2-xTB on Evaluating Aggregates Driven by NCIs	72
5.3.2 Influence of Metal Coordination and Light Irradiation on Hierarchical Self- Assembly Processes	73
5.3.2.1 The Effect of a Metal Ion upon Self-Assembly	74
5.3.2.2 The Photoresponsive Aggregation of C ₁	76
5.4 Chapter Summary.....	80
6. Examinations of the Aggregation Motif of 1D Porphyrinic Wires.....	82
6.1 Introduction	82
6.1.2 The Importance of J and H Aggregation	83
6.1.3 Controlling J versus H.....	83
6.2 Methodology	87
6.2.1 Computational Methods	87
6.2.2 Measurement of the Angle θ	87
6.3 Results and Discussion.....	88
6.3.1 A study of Fragmenting the <i>Mer</i> Unit for Computational Efficiency	89
6.3.2 The effect of Steric Hindrance on the θ Angle.....	91
6.3.3 Predicting the Effect of Peripheral Substituents on Symmetrically Substituted Porphyrin Aggregates.....	93
6.4 Reviewing the Data with Conformational Analysis.....	96
6.5 Chapter Summary.....	99
7. An Examination of Substituent and Metal Effects on the Dimensionality of Aggregation	101

7.2 Methodology and Computational Methods	105
7.3 Results and Discussion	105
7.3.1 Variation of the Metal and Peripheral Group Electronic Effects.....	106
7.3.1.1 <i>Evaluation of the Predictivity for Replicating the Experimental Trends</i>	108
7.3.1.2 <i>The Dimensionality Motifs when Altering the Central Metal</i>	109
7.3.2 Variation of the Steric Effects of the Peripheral Groups by Elongation.....	111
7.4 Geometrical Observations.....	113
7.5 Chapter Summary	114
8. Concluding Remarks	116
8.1 Introduction.....	116
8.2 A New Semi-Empirical Method for Improved Supramolecular Predictions.....	117
8.3 The Contributing Factors Affecting Wire Formation	118
8.4 Molecular Design of Wire Porphyrin J and H Stacks.....	119
8.5 Investigations to the Third Dimension.....	120
8.6 Future Work.....	121
8.6.1 Development of Computational Methodology	122
8.6.2 Structure-Property Relation in Wire Assemblies.....	123
8.6.3 The Dimensionality Control in Supramolecular Assemblies.....	124
References.....	126

Chapter 1

Introduction

1.1 Introduction

The motivation and structure of the thesis are presented herein. The collected literature composing Chapter 2 is outlined. Subsequently, the background of the computational methodologies detailed in Chapter 3 is explained. The results are presented in four chapters, structured by computational methodology, combined computational experimental study of 1D (one dimensional) wire aggregates, and a dimensionality motif study of porphyrinic aggregates. A reflection on the summed findings of the results is presented at the end of the thesis, along with a list of relevant future work suggestions.

1.2 Motivation

1.2.1 Supramolecular Aggregates

Supramolecular aggregates are commonly composed of repeating molecular building blocks, called '*mer*' units. Of particular interest in material and engineering sciences are the organic and metalorganic conjugated aggregates, such as supramolecular porphyrins. These conjugated *mer* units (Figure 1.1) list numerous structural and chemical components offered for formation of NCIs (non-covalent interactions), predominantly π - π interactions, van der Waals, and hydrogen bonding. These are weak attractive forces, with the ability to bind reversibly, leading to interesting assembly motifs, and bulk physical and chemical properties. This rich diversity of combinations in chemistry and properties makes them incredibly versatile. Interest in this

field is extensive, as such supramolecular systems have a wide range of applications in a large spectrum of scientific sectors.¹⁻⁷ The most important areas include biomedical gels (photodynamic therapy),^{1,4,8-11} electronic engineering materials (photo-responsive electronics)^{4,12-14} and environmentally applied mediums (metal-organic framework, or MOF, type porous materials).^{7, 15, 16}

The role of NCIs in chemistry has been first examined in 1873 by Johannes Diderik van der Waals.¹⁷ This commenced the investigation of host-guest chemistry and the ‘weak’ interactions in mainly biological entities (e.g. DNA and proteins).¹⁸⁻²¹ During the past three decades the field of supramolecular chemistry has flourished, when their prospect for chemical manipulation and application was recognised.²² With the initial work of Stoddart *et al.* in self-assembly and molecular machinery,²³ the interest and scientific explorations were seeded. Currently, there is a selected range of *mer* units of importance in supramolecular assembly applications. The most versatile are organic based units, commonly include OPE oligo(p-phenyleneethynylene)),²⁴⁻²⁸ phthalocyanines^{10, 29, 30} and most notably porphyrins.^{16, 31, 32}

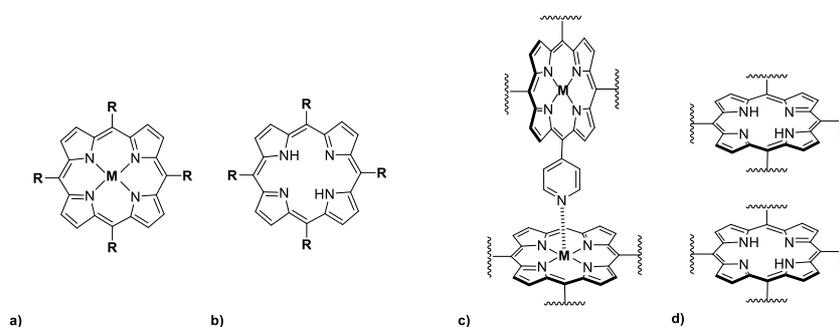


Figure 1.1 The Lewis structure of a generic porphyrin metalized in a), non-metalized in b) a t-shaped dimer in c) and a face-to-face dimer in d).

Supramolecular chemistry has seen an exponential growth in the number of scientific publications reported, and the resulting materials discovered in the recent years, show a rapidly expanding field (Figure 1.2). This research has been very impactful in important fields of science such as cancer research, with porphyrinic assemblies holding a leading role. Notably,

specific aggregation motifs of supramolecular porphyrins is a crucial aspect for optimal singlet oxygen generation for the process of oxygen sensing in cancer research.^{33, 34} Porphyrins are able to hold substituents possessing, almost the complete range of relevant NCIs. Complementary to their ready substitution, they are hollow molecules, also known as ‘*lacuna*’, able to hold a central metal ion, thus featuring even further interactions and properties compared to non-metallated aggregates which do not feature the coordination interaction (Figure 1.1).^{12, 35, 36}

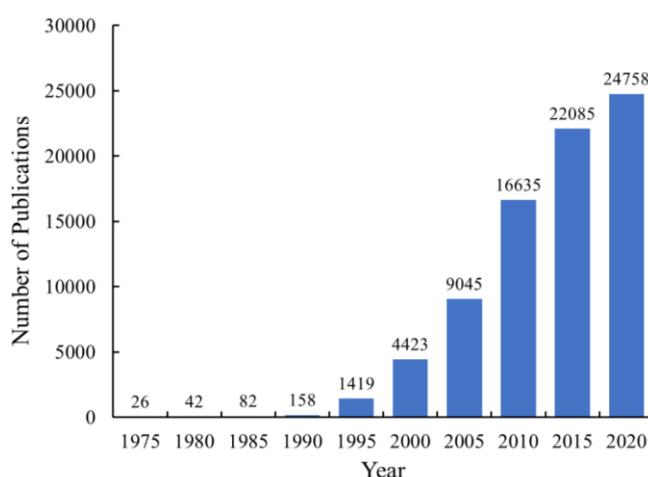


Figure 1.2 Progress in the number of publications produced in the field of supramolecular chemistry from 1970 to 2020.

The chemical and structural versatility of porphyrin molecules offer fertile ground for material and biomedical innovation in the range of zero to three dimensional (3D) assemblies (Figure 1.3).³ Subject to the substituents, or other variations in the structure of the *mer* unit the design of architecture of the supramolecular assembly becomes versatile. The 1D assemblies can form face to face (H), head to tail (J) type stacks,³⁷ or helical stacks. 2D assemblies form arrays in various geometrical shapes, and with 3D assemblies the complexity is even more increased with crystalline materials, dendrimers, porous structures and many more.³⁸⁻⁴⁰ These produce myriads of materials with ranging applications. A list of publications and reviews recommending structural design, such as interplay in hydrogen bonding and experimental

conditions such as solvation and temperature, that control these architectures offer a starting guide for future molecular design.³² Unquestionably, the remarkable versatility and range of applications has caught the attention and imagination of many scientists. The capacity to precisely control and functionalise the chemistry of those entities has allowed for unique multifunctional properties, such as functional porous materials,^{14, 41, 42} optically responsive,⁴³ or with adjusted solubility amongst other properties.^{1, 44}

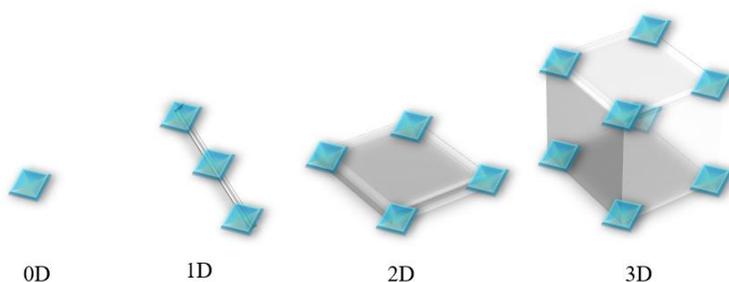


Figure 1.3 The dimensional aggregation motifs of a prototypical *mer* unit represented as a blue chip.

Considering the overview of the impact and interest in supramolecular materials, the need for exploring their design and synthesis is apparent. Particularly, porphyrin materials show promise within some very impactful fields, most importantly biomedical research^{1, 4, 45, 46} and environmental materials.^{7, 15, 47} The demand to predict the mode of aggregation prior to synthesis is high.

1.2.2 Computational Predictivity for Supramolecular Aggregates and Synthetic Recommendations

A systematic approach for the molecular design of supramolecular self-assemblies, requires computational predictivity for cost-effective and efficient synthetic explorations. There are very few studies highlighting computational predictions. The collection of which entails either older unsubstantiated methodologies, with inconsistent predictivity of properties, or potentially intractable computations offered only for small fragment dimers, and selected supramolecular

polymer information. Various *ab initio* and semi-empirical methods have been applied offering different information, without firm findings of a consistently reliable method for the prediction of the motif and properties of supramolecular architectures.⁴⁸ Recent attempts for improving the error in the computations of non-covalently bound species have been especially encouraging. In the last few years new semi-empirical Hamiltonians and theoretical dispersion corrections have been created improving the accuracy and efficiency of predictions for large supramolecular species. However, these have not been applied extensively on conjugated organic supramolecular assemblies. This has created the need for a systematic approach for tackling the computational predictivity of these large species, especially for the purposes of molecular design and mechanism of aggregation of porphyrinic assemblies, that can aid reducing experimental work in some impactful fields of research.

1.3 Structure of the Thesis

This thesis is composed of eight chapters and examines the use of computational methods for the molecular design of supramolecular conjugated organic aggregates with a special focus on porphyrins aggregates in one and three dimensions. Following Chapter 1, a literature background is detailed in Chapter 2. This surveys the relevant studies on supramolecular assemblies of conjugated organic molecules with respect to the structural factors and energetic components for material design. This survey highlights the gap in literature that will be addressed in this work. The theoretical background is outlined in Chapter 3 with an explanation of the physical chemistry relationships that make the chosen methods optimal for this work.

The remaining four chapters discuss the results of this computational analysis. These are categorised by initially validating the available computational methods within the frame of published benchmark and experimental data, evaluating the 1D potential of aggregates by comparison with published findings, and subsequently exploring factors that affect the dimensionality in porphyrin aggregations.

In Chapter 4 a benchmark study for method evaluation is carried out. The appropriate semi-empirical methods capable of computing NCIs were compared to reference benchmark results for a comprehensive data set of non-covalently bound species, relevant to this work. The methods were compared in terms of the binding energies associated with dimer formation.

In Chapter 5 a computational analysis for the cooperativity, energetic and geometrical details, of a series of 1D organic conjugated molecular stacks with interesting optoelectronic applications are undertaken in a combined experimental study. The predictivity of a set of DFT and semi-empirical methods is assessed through this work for a set of properties of the polymers.

Chapter 6 employs the use of published experimental data for 1D-wire motifs (named J and H) of porphyrin in order to investigate the role of peripheral substitution in these aggregations. The importance of aryl substituents when making computational predictions on those systems is highlighted. The use of conformational analysis is shown to provide further clarity concerning these predictions within this Chapter.

Chapter 7 investigates the driving forces behind porphyrin aggregations towards either a 1D-wire or a 3D-square. This is done initially in terms of the electronic nature of a series of σ donor ligands coupled with selected group 10, 11 and 12 metal ions. Furthermore chemical tuning recommendations in terms of the peripheral substitutions for the motif control are made.

Finally, Chapter 8 outlines the key conclusions from the collection of results of the thesis and recommends some future work that would help continue this research given more time and resources were provided. The Appendix electronic folder contains all supplementary data, including the coordinates of the structures computed, figures of the complementary experimental analytical spectra data and relevant validating tables (in MS Excel) of the computations supporting the reported results.

Chapter 2

Background Literature

2.1 Introduction

The purpose of this chapter is to survey the key relevant background literature concerning the molecular design of supramolecular architectures, their binding energetics, and applications, as well as the relevant computational tools for the optimal predictivity of these species. In the past decade it has been evidenced that supramolecular materials are rapidly emerging as promising candidates for a wide range of applications. The impact of these next generation materials can be seen in many crucial applications, such as photodynamic therapy,¹ photovoltaic cells for sustainable engineering,⁴⁹ and the design of interesting novel materials, such as shape memory materials.⁵⁰ The engineering of organic multifunctional materials, with exploitable properties in a wide range of scientific sectors is currently one of the most vital areas of research. Notably porphyrins perform functions in nature, i.e. electron transfer, oxygen transfer, and light harvesting, which are desirable for replication in engineering of novel functional molecular devices or materials via supramolecular assembly.

One of the primary bottlenecks hindering the synthetic progress and production of new supramolecular materials is the current lack of understanding of the specific fundamental molecular tuning resulting in desired supramolecular architectures and material properties. The current need to establish intrinsic and extrinsic mechanistic processes of assembly and the

appropriate chemical tuning that promotes specific material properties has been the motivation of this thesis.

2.2 Supramolecular Architectural Design

2.2.1 Molecular Design for Dimensionality Control

Supramolecular aggregates are entities composed of smaller distinct building blocks, usually molecules or ions, that are able to bind reversibly via NCIs (non-covalent interactions), the *mer* units. These NCIs drive the directionality of the aggregate in space, and their weak nature gives them unique properties, such as reversible spatial organisation, or activation of property on command (e.g., via thermal or photoinduced activation). For the discovery of new materials and control of properties through synthetic design, it is important to be able to control the aggregation architecture and dimensionality motif. In supramolecular assemblies, then, one of the first key challenges in material design is the supramolecular architecture. These architectures range from zero to three dimensions in many geometrical designs, from simple square arrays, to crystalline cubic assemblies, or complex dendrimers. Considering this diversity, the first point to address is to be able to control and define the degree of dimensionality of those aggregations. A particularly researched category of supramolecular assemblies refers to aggregations of conjugated organic molecules, due to their versatile electronic properties and biomedical and material applications.^{8, 43, 51, 52} Many publications and research work have addressed synthetic approaches from zero to three dimensional motifs of such supramolecular assemblies.^{40, 53, 54} With the zeroth dimension referring to the spatially unorganised molecular species, the 1D (one-dimensional) to intermolecularly bound on a 'wire'-like stack,⁵⁵ 2D (two-dimensional) are arrays of those and the more complex 3D (three-dimensional) aggregations entailing almost any given 3D geometry being assembled by *mer* unit building blocks. Each dimensional motif entails its own specific synthetic demands and properties. Though myriads of synthetic investigations have taken place for the creation each

of these supramolecular motifs, little is known about the intrinsic chemical tuning promoting each motif.

One of the key challenges to be addressed for increasing the efficiency of the process of synthetic design of new materials is the understanding of the driving chemical design for the architectural control of supramolecular aggregations.⁵⁶ While the structure and material properties of distinct supramolecular aggregations have been systematically studied and extensively characterised,^{7, 32, 42, 56-58} considerably less attention has been focused on the systematic categorisation of chemical tuning which is so critical for synthetic recommendations when attempting synthesis of new materials. In principle when commencing the synthesis of a new supramolecular material, to be able to predict its long-range structure and properties, some recommendations from literature are followed concerning the structural design of the *mer* unit. However, detailed data, such as a database library, that could provide chemical tuning recommendations for a comprehensive synthetic route remain scarce.^{56, 59}

Particularly, the supramolecular motif design and their electronic properties in conjunction with long-range structural and chemical stability of the assemblies are central for production and design of new materials and are currently limiting the successful deployment of applications in many diverse scientific sectors. The emerging field of supramolecular polymer design, particularly of self-assemblies of conjugated organic molecules and their optoelectronic response applications is an area of intense study. Given the versatile chemical framework of small organic conjugated species, myriads of supramolecular geometrical combinations are formed based on the main NCIs, as well as the compliant metal coordination environments.^{14, 41, 42} This opens exciting new opportunities for yielding novel properties, which can be rare in conventional materials.³⁴ The current supramolecular materials have been developed in various architectures via many synthetic strategies. Many reviews have surveyed notable synthesis of

the 3D motifs of such assemblies.^{3, 31, 40} However, it is scarcely discussed or explained what are the factors promoting each motif for future synthetic consideration.

A survey of the 1D ‘wire’ type assemblies offers the collection of some remarkably interesting applications with particular electronic properties. The preparation of 1D arrays has been approached in various ways, using both covalent and non-covalent approaches. There are various ways to assemble small π -conjugated molecular systems. These modes include fused, conjugated, planar, orthogonal, or helical connections, with respective optoelectronic, or material properties. The ‘wire’ assembly has been categorised in three main stacking modes with distinct spectroscopic and electronic properties. These are a face-to-face mode named the H, a head-to-tail mode named J and a crossed mode, and with rotated *mer* units on the wire axis named X (Figure 2.1).⁵⁵ The distinction between these three is clear through their absorption spectra, with a deeper bathochromic shift (towards longer wavelengths) for the J stack and a hypsochromic (towards shorter wavelengths) for the H, while a crossed stacking mode yields a bathochromic and hypsochromic shift simultaneously with the small rotational angle allowing both splitting transitions. The mode of stacking is definable within the structure by the angle of the intermolecular slippage θ . According to Kasha’s exciton coupling theory, a larger slippage for the J stacks is typically assigned for $\theta < 55.5^\circ$, with similarly assigned for the X that additionally possess intermolecular rotation. Additionally, H is with smaller internal rotation and $\theta > 55.5^\circ$.³⁷ These optically and electronically responsive materials have been the focus of many synthetic approaches, from thermal and solvent aggregation control,⁶⁰ to surface or surface-liquid interface experiments,³² and fine chemical tuning.⁴⁴ However, a systematic categorisation of these factors for a comprehensive set of conjugated organic assemblies has not been proposed for promoting these assembly modes in 1D aggregations, due to the time-consuming complex experimental intricacies needed their study.⁴³

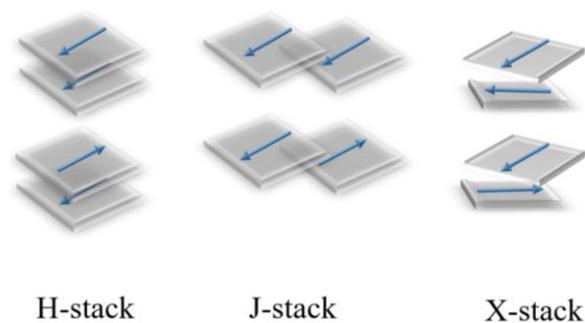


Figure 2.1 The configurations of the H, J and X-type stacking modes from left to right respectively, where the grey tiles represent generic *mer* units, and the blue arrow the direction of the dipole moment in a suggested assembly.

Some literature recommendations can be made for promoting each motif from assessing review surveys. For example, it has been shown that bis(imidazolyl) Mg porphyrins produce only one-dimensional oligomers.⁶¹ Furthermore, the presence of a large bismuth ion in a porphyrin typically leads to a J aggregation.³⁴ More thorough examinations for the molecular design of aggregates promoting a J or H aggregation have been carried out for porphyrin and metalloporphyrin assemblies by various groups.^{2, 37, 43, 60, 62-65} These experimental data additionally include the solvent and thermochemistry factors that essentially promote each motif. It is though crucial as the first synthetic step to establish the correct molecular structure, before attempting synthesis, something that has not been solely reviewed for molecular design.

The self-assembly of conjugated organic molecules has been explored in 2D and 3D materials for different applications. Various methodologies for creating 2D arrays and 3D networks have been developed, for example with surface chemistry and utilisation of metal organic interactions on metal surfaces,³² or in appropriate solvation.³¹ The utilisation of hydrogen bonding, π - π interactions, dipole moment and metal-organic coordinated networks is specifically tuned for each design. Some initial work for tuning the structure in attempt of controlling the dimensionality of the assembly has been carried out by Tanaka *et al.* with some interesting recommendations for the functionalisation of porphyrins and the importance of the

dipole moment of the *mer* unit for promoting a specific dimensionality motif.⁴⁴ Otherwise the extension of organic supramolecular aggregates to the third dimension has been achieved notably by creating vertical columnar stacks perpendicular to a surface substrate,⁶⁶ or more commonly by insertion of a metal ion within the *mer* design when allowed, for creating 3D metal-organic coordinated networks.³⁸ A more global survey of these assemblies of conjugated organic molecules indicates a multifactorial perspective for the control of the aggregation motif. The functional substituents of a conjugated *mer* unit further affect the degree of solubility as well as refine thermodynamic parameters that play an important role in the final assembly mode.^{2, 60} It is however an important first aspect of the molecular design to be able to establish the thermodynamic control before accounting for these additional complexities, this is something that can be carried out efficiently with computational predictions, which compute the binding energies for the evaluation of the thermodynamic control.

2.3 Cooperativity Studies of Supramolecular Assemblies

An important parameter for designing and creating supramolecular materials is the characterisation of the binding mechanism of the self-assembly. This refers to the *cooperativity* and is defined by the binding events or interactions during the self-assembly of the *mer* units. The free energy change ($\Delta\Delta G$) from the first step, the creation of the seed dimer, to the next addition of a *mer* unit is either negative, in which case termed *positively cooperative*, or a positive change, termed *negatively cooperative*. When the change is zero the model is termed *isodesmic*. These categorisations reflect the ease of a supramolecular aggregate to spontaneously assemble or not. Indicative cooperativity factor values have been proposed to quantify the cooperativity. Specifically for 1D homogeneous aggregates Martin *et al.* introduced the ρ factor extracted from the free energy difference of the first and second aggregation step $\Delta\Delta G_{cop} = RT\ln(1/\rho)$, with R the global gas constant and T the temperature.⁶⁷

⁶⁸ These factors extend their complexity to account for heterogeneous structures, valency and

statistical parameters in aggregation,⁶⁸ however this first categorisation describes the key properties for molecular design. Measuring the free energy, that subsequently provides the cooperativity value, is normally carried out experimentally by supramolecular titration methods, NMR (nuclear magnetic resonance), UV-Vis (ultraviolet-visible), CD (circular dichroism) spectroscopy and ITC (isothermal titration calorimetry) methods. Cooperativity can also be measured with computational predictions by evaluating the binding affinity of dimers or larger aggregates, or when it is tractable, and reliable in predictivity, the free energies in thermochemistry computations, normally DFT or *ab initio* methods.⁶⁹ The rate of change of ΔE_{avg} (where, $\Delta E_{avg} = (E_n - nE_1)/(n-1)$, n is the number of *mer* units, E_1 the binding to the monomer, E_n the binding of the oligomer of n *mer* unit size) can indicate a cooperative or other model, when increasing or decreasing respectively with the increase of n . There have been detailed and systematic approaches for controlling and programming the cooperativity in self-assemblies, by creating mathematical models or libraries of molecular and thermodynamic data.⁷⁰⁻⁷² The cooperativity factor of self-assembly defines the conditions in which the material should be synthesized and function. Its definition at a pre-synthetic stage is an important step for cost-efficiency which has not been extensively explored with computational predictions in a molecular design approach.^{12, 26, 73}

2.4 Computational Approaches and Obstacles for Supramolecular Architectural Design

There are many research interests involving supramolecular material design, ranging from bulk material properties to the design at a molecular *mer* unit level. According to each respective focus various computational predictive methodologies have been applied with ranging success. In this thesis the aim targets the molecular design of the *mer* unit for driving the architecture of the long-range assembly within the three dimensions. Previous computational work has focused on establishing the intrinsic interactions that promote the self-assembly via DFT, such as quantifying the strength of hydrogen bonding,⁷⁴ establishing the geometries that drive the

long-range directionality of the assembly,⁷⁵ or reviewing the electronic and photochemistry properties of stacks by excited state computations.^{76, 77} However, for the purpose of deducing a supramolecular architecture motif as function of chemical tuning, *ab initio* and DFT computations are not commonly tractable with these large non-covalently bound species, and a need to compute a larger aggregate segment is present. For this purpose, semi-empirical computations have been developed in combined experimental-theoretical work. The semi-empirical computations are much faster and can be tractable for larger aggregates. Semi-empirical computations have provided interesting information concerning the aggregation mechanism of wire assemblies, specifically the fragmentation pathways preferred during assembly.²⁶ Other semi-empirical results have validated key geometrical information, and cooperative interactions complementing and validating experimental data.^{12, 78, 79} These theoretical predictions vary in precision. The composite nature of supramolecular aggregates, which combines multiple NCIs with large complex entities, has made them a challenging computational problem with a currently non-straightforward solution. Theoretical predictive methodologies for species with NCIs have been suggested with new, still developing corrections, most commonly Grimme's D3 dispersion correction.⁸⁰ This has only been recently applied within semi-empirical methods and has not yet been widely tested on supramolecular aggregates. There have been many attempts to correlate the validity of theoretical predictions with experiment, most of which confirm that DFT computations are only tractable for study of the monomer or at best dimer, and that a dispersion correction is important for improving the accuracy of these predictions.^{73, 81-85} Recent reviews propose post-DFT methods for computational design of new materials, such as diffusion Monte Carlo,⁸⁶ domain-based local pair natural orbital coupled cluster,^{87, 88} and time-dependent DFT⁸⁹ showing improved predictivity.⁹⁰⁻⁹² It is shown that traditional DFT methods have limited accuracy especially for larger systems, or complex supramolecular entities. For example, a critical survey of

experimental and computational correlation of the aromatic stacking interactions using traditional DFT functionals (B3LYP-D3) reviewed the electrostatic, dispersive, and solvent component in these complexes, as a factor of contributing peripheral substitution. These gave some initial indications of the insufficiencies of these methods.⁹³ The trends found in this work were used in the development of the dispersion correction in computational models, as well as making suggestions to the currently obscure computational solvent models.

Further investigations for the computational predictivity of supramolecular systems have outlined a series of notable obstacles currently found in thus far developed methods. A recent review on this subject by Schneider *et al.* lists initially the importance of the free energy and enthalpic values, that are computed often unpredictably by most computational methods.⁹⁴ The evaluation of the binding mechanism, with the dispersive and hydrophobic interactions was reviewed, as well as the factors of medium effects and changes in conformations. It was concluded in this survey that the currently used DFT-D3 and MD methodologies with explicit solvent parameters yield relatively good trends for thermochemistry values, with 1.0 gradient of correlation with experiments and RMSE 4-9 kJ/mol in sets of small organic complexes.⁹⁵ The complexities of experimental conditions have further been reviewed for non-covalently bound systems of biological interest, where the variations in assembly in the gas and solvent phase are detailed as a function of the various intermolecular interactions found in those species (halogen, chalcogen, pnicogen). To address these complexities in computational modelling, systematic correlation with experimental data has been developed in proposed mathematical models.⁹⁶ It is though shown in the majority of these reviews that current thermochemistry data, especially in combination with the available solvent models, present significant errors and inconsistencies and are not always reliable. In order to avoid these discrepancies which are only extrapolated in larger systems, a hypothesis that it can be made that a review of the geometries and binding affinities of supramolecular systems in the gas phase offers sufficient

accuracy. The majority of the benchmark reviews for NCIs has shown encouraging error analysis for systems reviewed in the gas phase, when compared to experimental data. This is a promising indication that ‘expensive’ thermochemistry computations and currently inconsistent solvation models can be omitted all together for reliable computational predictivity. Further in the results Chapter 4 of this thesis, a benchmark study is undertaken to validate this approach with recent literature suggested computational methodologies. While further comparison with experimental data in this work’s results further assesses this hypothesis.

Currently, the ‘golden standard’ for accuracy of non-covalently bound species has been reviewed as the coupled cluster methods CCSD(T)/CBS (single double triple excitations, complete basis set).⁹⁷ This is a computationally demanding approximation that is not tractable for anything larger than a small molecular complex of up to normally 30 atoms. It is commonly used in benchmark studies for high accuracy reference values.⁹⁸

An established computational methodology for evaluating the architecture and cooperativity of large supramolecular aggregates has not yet been confirmed, as the research concerning this remains limited.^{99, 100} One of the first steps in computational predictions for tackling non-covalently bound species has been the development of a dispersion correction.¹⁰¹ In computational studies and benchmarks it has been originally seen that larger errors were produced when NCIs were present.^{80, 102, 103} The most recent version of the dispersion correction D4 gave significant improvements in benchmark studies for non-covalently bound complexes.¹⁰⁴⁻¹⁰⁶ A recent encouraging benchmark, that addressed the validation of the best computational methodology for larger supramolecular species, has been carried out by Sure *et al.* in a set of publications.^{99, 100} The precision and computational cost of a series of DFT and semi-empirical methods were compared for relevant benchmark datasets of small and larger organic complexes. This work highlighted the benefits and shortfalls of each method for gas

phase and solvent phase computations, and notably made use of experimental free energy values as referenced for the S30L benchmark set,¹⁰⁷ which is a data set of larger non-covalently bound complexes. In these results it is concluded that DFT-D3 methods continue to provide a consistent relative low error. It is seen complementary that the less computationally costly semi-empirical methods, such as the PM6-D3¹⁰³ and DFTB,¹⁰⁸ HF-3c,¹⁰⁰ GFN-xTB¹⁰⁹ and GFN2-xTB¹⁰⁵ also provide a comparably low error prediction, occasionally improved to DFT (Figure 2.2). This work continued with developing a new high efficiency tight-binding DFT Hamiltonian, specifically designed for supramolecular species of mainly biologically species, the GFN-xTB and their latest version GFN2-xTB. A complementary solvation model and a conformational analysis package accompanies this model which aids a more thorough analysis of the entailing supramolecular complexities.

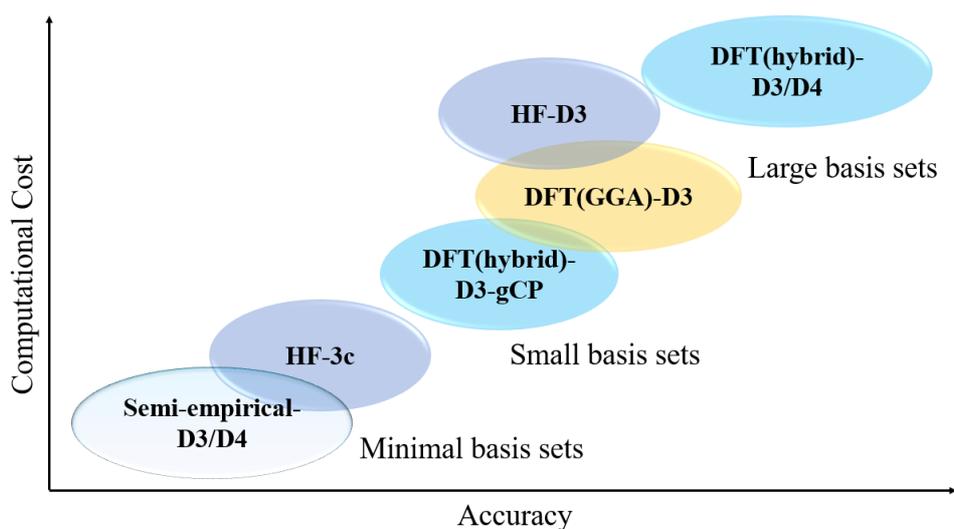


Figure 2.2 Sketch of the cost-accuracy ratio of selected relevant quantum chemical methods where dispersion correction is assumed in all cases, as suggested by the Sure-Grimme benchmark findings.¹⁰⁷

Based on the presented obstacles and advancements concerning the modelling of supramolecular self-assemblies, this thesis has reviewed and collected methodologies which can be tractable for larger aggregations. The precision, relevant advantages and disadvantages

of the selected methods are reviewed in a benchmark in Chapter 4, where a choice of method is made for the application on the remaining results and predictions of this work. It is unquestionable that further advancements and developments are essential for the computational study of supramolecular aggregations. There are multiple factors affecting and controlling the assembly and as of yet little has been assessed and correlated with computational methodology. However, many systematic benchmark reviews have provided firm guiding data suggesting that gas phase computations provide sufficient precision for the molecular design of conjugated organic aggregates.

2.5 Chapter Summary

A representative collection of the most relevant background literature is provided in this chapter. This refers to the recent advancements in architectural design of supramolecular assemblies and the computational methodologies developed for their predictivity. It highlights the relevance of these recent publications with the work performed in the later chapters and the resulting motivation that inspired this project. Through this literature survey it is concluded that there is a need for a better understanding of the forces that drive the preference of motif in aggregation, and it is further shown that the correct choice, as well as further development of an optimal computational methodology is essential.

Chapter 3

Theoretical Background

3.1 Introduction

This chapter provides an overview of the theoretical background of the methods utilized throughout this thesis. The core of these methods is based on the basic concepts of quantum chemistry, here focused on the electronic ground state theory. These methods commenced historically with the HF (Hartree-Fock) approximation which is initially introduced herein. DFT (density functional theory) and their sub-classes then are detailed following these. A spectrum of simplifications and approximations created a range of semi-empirical quantum chemical methods which are then discussed. These are subdivided to MNDO (modified neglect of diatomic orbital) and DFTB (density functional theory tight binding) type. Subsequently, the key predictive tool, for supramolecular entities, dispersion correction is detailed, which is a specific factor applied to systems containing NCIs (non-covalent interactions). Lastly, a CREST (conformer-rotamer ensemble sampling tool) is detailed briefly in relation to its utility for this work.

3.2 Fundamental Concepts of Quantum Chemistry

The fundamental concept of quantum chemistry is the examination of the electronic structure of atoms and molecules and hence their physical and chemical properties through solving the nonrelativistic time-independent Schrödinger equation.¹¹⁰ This is shown compactly in the following relationship:

$$\hat{H}\psi = E\psi \quad \text{Eq. 3.1}$$

where E is the total energy of the entity, and the eigenvalue of the time-independent Hamiltonian \hat{H} operator, while ψ is the wavefunction. This is an eigenvalue equation with a solution being a complete set of orthogonal eigenfunctions ψ and their E eigenvalues. The wavefunction ψ describes a physical concept as the square modulus $|\psi|^2$ giving the probability of finding a particle in a space volume and so the probability density. All the essential property information about a system in the ground state (the lowest energy state) are given by the lowest energy wavefunction. For a system of N -electrons the Hamiltonian operator takes the following form:

$$\hat{H} = -\frac{\hbar^2}{2} \sum_{\alpha} \frac{1}{m_{\alpha}} \nabla_{\alpha}^2 - \frac{\hbar^2}{2m_e} \sum_i \nabla_i^2 + \sum_{\alpha} \sum_{\beta > \alpha} \frac{Z_{\alpha} Z_{\beta} e^2}{4\pi\epsilon_0 r_{\alpha\beta}} - \sum_{\alpha} \sum_i \frac{Z_{\alpha} e^2}{4\pi\epsilon_0 r_{i\alpha}} + \sum_j \sum_{i > j} \frac{e^2}{4\pi\epsilon_0 r_{ij}}$$

$$\text{Eq. 3.2}$$

Therein, \hbar is the reduced Plank's constant, α and β represent two nuclei and i and j two electrons. The right side of the equation is the addition of five summations; The first two reflect the kinetic energy of the nuclei firstly and then of the electrons. The remaining three parts concern the potential energy, starting with the repulsion between the nuclei (with Z_{α} and Z_{β} atomic numbers) at $r_{\alpha\beta}$ distance, following the potential energy of attraction of the electron and the nuclei at $r_{i\alpha}$, and the last the repulsions between the electrons at r_{ij} . Each summation can be replaced by its equivalent symbol giving:

$$\hat{H} = \hat{T}_n + \hat{T}_e + \hat{V}_{nm} + \hat{V}_{ne} + \hat{V}_{ee} \quad \text{Eq. 3.3}$$

The solution of those determines the electronic structure of any chemical entity. For the case, however, of many-particle systems a solution of the exact wavefunction remains inapproachable, due to the correlated motions of particles and the repulsion/attraction terms in \hat{H} suggesting the co-dependent motion of particles. This raised the development of

approximations to tackle this complexity within a tractable accuracy and computer processor cost.

An important initial approximation suggests that the nucleus is heavy enough to be assumed stationary as compared to the moving electrons. This is the Born-Oppenheimer approximation.¹¹¹ This results in a separation of the nucleic and electronic segments of the equation, hence having \widehat{V}_{nm} constant for the repulsion of the nuclei and \widehat{T}_n zero for their kinetic energy. It equates to the electronic Hamiltonian operator \widehat{H}_e and the Schrödinger equation for electronic motion:

$$\widehat{H}_e \psi_e(r; R) = E_e \psi_e(r; R) \quad \text{Eq. 3.4}$$

$$\widehat{H}_e = \widehat{T}_e + \widehat{V}_{ne} + \widehat{V}_{ee} \quad \text{Eq. 3.5}$$

With respectively, ψ_e the electronic wavefunction, r the electronic coordinates parametrically dependent on the nuclear R . The calculation of E_e is applied by the variational principle by utilising the following relationship:

$$E' = \frac{\langle \psi' | \widehat{H} | \psi' \rangle}{\langle \psi' | \psi' \rangle} \geq E_0 = \langle \widehat{H} \rangle = \frac{\langle \psi_0 | \widehat{H} | \psi_0 \rangle}{\langle \psi_0 | \psi_0 \rangle} \quad \text{Eq. 3.6}$$

The variational principle states that an expectation energy value of a trial wavefunction ψ' cannot be lower than the energy of the true wavefunction ψ_0 . Unless explicitly stated, the electronic energies within the remaining of the chapter will only be considered. Therefore, the subscript “e” will be dropped from hereon.

For the ground state of the electronic wavefunction of a system of N electrons the Slater determinant Φ_0 provides an approximation of sufficient accuracy, within the context of the systems studied in this thesis.¹¹²

$$\Phi_0 = \frac{1}{\sqrt{N!}} \sum_n^{N!} (-1)^{p_n} \widehat{P}_n \prod_i^n \phi_i \quad \text{Eq. 3.7}$$

The Slater determinant is then the product of the normalization factor $\frac{1}{\sqrt{N!}}$, the specific summation $\sum_n^N (-1)^{p_n}$, where p_n is the necessary transpositions for the n^{th} permutation, the permutation operator \hat{P}_n to reflect on the exchange of electrons between orbitals, and the product of the independent one electron wave functions ϕ_i . This setup keeps with the Pauli exclusion principle for an anti-symmetric wave function with interchange of two electrons.

3.3 Hartree-Fock Theory

One of the initial approximation procedures for minimizing the Slater determinant was the Hartree-Fock or HF theory.¹¹³ This laid the foundation for developing more sophisticated and advanced approximations for the evaluation of the electronic energy. Consequently, as described in the latter section the energy is minimized for the respective orbitals ϕ_i . The HF eigenvalue is then set with the Lagrange multiplier, constrained for orthonormal orbitals.

$$\hat{f}_i \phi_i = \varepsilon_i \phi_i \quad \text{Eq. 3.8}$$

$$\hat{f}_i = \hat{h}_i + \hat{v}_i^{HF}, \quad \hat{v}_i^{HF} = \sum_j^N \hat{J}_{ij} - \hat{K}_{ij} \quad \text{Eq. 3.9}$$

In HF a mean field theory is in place. The electron-electron Coulomb potential is approximated by setting an average electron field of the combined electrons and giving the interaction of it with a single electron i . The Fock operator here \hat{f}_i is the one-electron operator for which the MOs ϕ_i are the eigenfunctions. \hat{v}_i^{HF} is the sum of the Coulomb operator \hat{J}_{ij} and the exchange operator \hat{K}_{ij} . Where the Coulomb operator includes the electron-electron repulsion and the exchange operator the anti-symmetric wavefunction factor, thus adding some effective electron interaction parameters, for an electron with the same spin.

$$\hat{J}_{ij} |\phi_i\rangle = \left\langle \phi_j \left| \frac{1}{r_{ij}} \right| \phi_j \right\rangle |\phi_i\rangle \quad \text{Eq. 3.10}$$

$$\hat{K}_{ij} |\phi_i\rangle = \left\langle \phi_j \left| \frac{1}{r_{ij}} \right| \phi_i \right\rangle |\phi_j\rangle \quad \text{Eq. 3.11}$$

The above relationships imply that there is no interaction of the electron with itself because $\langle \phi_i | \hat{f}_{ii} | \phi_i \rangle$ is equal to $\langle \phi_i | \hat{K}_{ii} | \phi_i \rangle$. This then addresses the self-interaction error in the original quantum chemical relationships.

Commonly atomic orbitals ψ , expressed as basis functions, compose the linear combination expansion of the MOs (molecular orbitals).

$$\phi_i = \sum_{\mu} C_{\mu i} \psi_{\mu} \quad \text{Eq. 3.12}$$

With $C_{\mu i}$ the LCAO-MO coefficients. With ϕ_i here giving the exact solution for a complete basis. In the case of a finite basis the BSIE (basis set incompleteness error) is produced. The Roothan-Hall relationship gives Eq 3.8 via an expression of a linear matrix.

$$\mathbf{FC} = \mathbf{SC}\varepsilon \quad \text{Eq. 3.13}$$

With \mathbf{F} as the Fock matrix of elements of $\langle \phi_i | \hat{f}_i | \phi_i \rangle$, \mathbf{C} the LCAO-MO coefficients, \mathbf{S} the overlap matrix of elements of $S_{ij} = \langle \phi_i | \phi_j \rangle$, accounting for AO not always being orthogonal. Subsequently, the solution of 3.13 gives the eigenvalues of the orbitals ε . As \mathbf{F} is parametrized with the MO coefficients, the SCF (self-consistent field) method can yield the orbitals iteratively. These can then be used for the HF energy.

$$E^{HF} = \sum_i^N \langle \phi_i | \hat{h}_i | \phi_i \rangle + \sum_i^{N-1} \sum_{j>i}^N (\langle \phi_i | \hat{f}_{ij} | \phi_i \rangle - \langle \phi_i | \hat{K}_{ij} | \phi_i \rangle) \quad \text{Eq. 3.14}$$

The HF approximation is based on the approach of each electron undergoing the average potential of the other electron. This omits the Coulomb repulsion between individual electrons. This then means that the exact solution for the complete basis set cannot be reached. The exact solution E can be defined with an addition of the correlation energy or rather the correlation energy E_{corr} is:

$$E_{corr} = E - E^{HF} \quad Eq. 3.15$$

There are developed approximations for calculating E_{corr} . Notable methods are the MP (Moller-Plesset) theory, CI (configuration interaction) and the CC (couple cluster). The computation of E_{corr} is essential for describing many chemical processes. Nevertheless, these are computationally demanding processes, which are not tractable for systems larger than a hundred atoms. With this account these will not be further described in this chapter, as they cannot be applied for large supramolecular entities.

3.4 Density Functional Theory

Density Functional Theory, DFT, is the most common electronic structure method applied in modern computational chemistry, mostly for organic systems. Originally, Hohenberg and Kohn described a unique functional ρ for the electron density of a system in its ground state.¹¹⁴ The parameter of ρ fundamentally offers the simplicity of being dependent on three spatial coordinates for any number of electrons, as opposed to the $3N$ variable for a N -electron system in HF. This parameter is then used within the Hamiltonian operator for the estimation of the ground state electronic energy. The analogue to the variation principle (in wavefunction theory) is the second Hohenberg-Kohn theorem, giving $E[\tilde{\rho}]$ as the upper bound to the exact ground state energy for a trial $\tilde{\rho}$. $E[\tilde{\rho}]$ can then be defined as a composition of three defined functionals of ρ .

$$E[\rho] = T_e[\rho] + V_{em}[\rho] + V_{ee}[\rho] \quad Eq. 3.16$$

$$V_{ee}[\rho] = J[\rho] + K[\rho] \quad Eq. 3.17$$

$T_e[\rho]$ is the kinetic energy, $V_{em}[\rho]$ is the Coulomb electron-nuclei attractive interaction, $V_{ee}[\rho]$ is the electron-electron interaction, composed of the Coulomb part $J[\rho]$ and the exchange part $K[\rho]$. The simplistic spatial nature of the density means the $V_{em}[\rho]$ and $J[\rho]$ can be expressed with classical relationships.

$$V_{em}[\rho] = - \sum_A^M \int \frac{Z_A \rho(r)}{|R_A - r|} dr \quad Eq. 3.18$$

$$J[\rho] = \frac{1}{2} \int \frac{\rho(r)\rho(r')}{|r - r'|} dr dr' \quad Eq. 3.19$$

The development of the kinetic energy and the exchange part initiated with the Thomas-Fermi model, based on the uniform electron gas, (UEG).

$$T_e^{UEG}[\rho] = \frac{3}{10} (3\pi^2)^{\frac{2}{3}} \int \rho(r)^{\frac{5}{3}} dr \quad Eq. 3.20$$

$$K^D[\rho] = -\frac{3}{4} \left(\frac{3}{\pi}\right)^{\frac{1}{3}} \int \rho(r)^{\frac{4}{3}} dr \quad Eq. 3.21$$

This first approach did not yield an accuracy for a chemical utility, based on the relatively inaccurate kinetic energy of the Thomas-Fermi model. The modern applied approach by Kohn-Sham, KS, gives 98-99% of the true kinetic energy in DFT with the use of a fictitious system, where a set of quasi-particles that are non-interacting have the same density as the true system.¹¹⁵ This approach demands the use of orbitals for calculating the kinetic energy T_{KS} , thus increasing, again, the variables by 3N. The approximation to the complete DFT energy E^{DFT} requires the correlation energy of the independent particles E_C and the exchange effects E_X , summed all together in the exchange-correlation functional E_{XC} .

$$E^{DFT} = T_{KS}[\phi] + V_{en}[\rho] + J[\rho] + E_{XC}[\rho] \quad Eq. 3.22$$

$$E_{XC}[\rho] = E_X[\rho] + E_C[\rho] \quad Eq. 3.23$$

$$T_{KS}[\phi] = -\frac{1}{2} \sum_i^N \langle \phi_i | \hat{V}_i^2 | \phi_i \rangle \quad Eq. 3.24$$

The Kohn-Sham operator \hat{f}_i^{KS} , similarly with the Fock operator, is utilised to produce the electronic energy and the KS-orbitals iteratively.

$$\hat{f}_i^{KS}[\rho]\phi_i = \left[\hat{h}_i[\rho] + \sum_j (\hat{J}_{ij}[\rho] + v_{xc}[\rho]) \right] \phi_i = \varepsilon_i \phi_i \quad Eq. 3.25$$

This is, again, effectively a one-electron operator. The exchange correlation energy $E_{xc}[\rho]$ has the advantage of describing correlation effects. This yields the correlation potential $v_{xc}[\rho]$, in a similar manner as the exchange operator in HF. On the other side, while formally DFT gives an exact solution, the DFA (density functional approximations) are inexact. The improvement of those is less systematic than the WFT (wave function theory) approximations, and it is done by trial and error, rather than a full CI.

3.4.1 Density Functional Approximations

The density functional approximations, DFA, as previously described are developed *via* a trial and error approach. Therefore, their categorisation and classification are carried out in terms of their resulting accuracy and computational cost. ‘Jacob’s ladder’ was formed by Perdew, to reflect the increasing ascension of accuracy. This refers to the increasing separation among rungs when ascending the ladder from the relevant biblical story (a ladder of rungs with progressively increasing separation when ascending). Respectively this indicates the DFA reflected by the rungs in the ‘ladder’, the higher the position of the rung, the more increased the computational cost, and accuracy of yielded energy. This is only a representative, roughly general description of the ranking, and it does not reflect the increase in accuracy faithfully.

3.4.1.1 Local Density Approximation

Jacob’s ladder lowest rung is set with the LDA (local density approximation). This accounts only for the local electron density, while assuming a uniform electron gas behaviour, or alternatively slow electron density variation. In finite systems the exchange-correlation potential $v_{xc}^{LDA}(r)$ decays exponentially. This does not reflect a real system where the decay is

apparently slower as per Coulombic relationships.¹¹⁶ This implies that HOMO energies are overestimated, consequently the ionisation potential of Koopmans' theorem is mis-predicted, as well as most electron rich species (e.g. anions). The LDA exchange functional is a form of the Dirac functional (see Eq. 3.21) with an additional set pre-factor.¹¹⁶ A representative LDA correlation energy is based on interpolating analytically Monte-Carlo calculations, described by Vosko, Wilk and Nusair (VWN).¹¹⁷ LDA shows overbinding predictions for most entities. However, good LDA predictions are made for metal solids, which have a like-UEG electronic structure.

3.4.1.2 General Gradient Approximation

The initial density functionals are parametrized only on the density. This was then improved by the introductions of the gradient density $\nabla\rho$ in the parametrization. This resulted to the GGA (generalised gradient approximation). One of the most popular forms of which is the BLYP functional. This entails gradient corrections within the exchange functions, as developed by Becke in 1988,¹¹⁸ and the correlation part developed by Lee, Yang and Parr (LYP).¹¹⁹ This is a robust functional for a wide range of chemical and physical property prediction, yet it fails in many isomerisation energy and barrier height calculations. Other notable functional include the PBE with the exchange and correlation functional by Perdew, Burke and Ernzerhof,¹²⁰ and Becke's B88 exchange functional.¹²¹

3.4.1.3 Meta-GGA Functionals

With the ascending of the accuracy ladder with functionals, higher order derivatives of the electron density are added. These did not offer consistent numerical output, consequently factors related to the kinetic energy density of the orbitals were included. This resulted in functionals of relatively higher accuracy than the previous GGA on many aspects, yet not as a consistent rule for all systems. A popular meta-GGA functional for NCI, is the TPSS, by Tao,

Perdew, Staroverov and Scuseria,¹²² These GGA and meta-GGA functionals parametrize the proximity and local density of the electron, and are called semi-local functionals.

3.4.1.4 Hybrid Functionals

A further attempt in optimising the DFT functionals was made by specific substitution of the exchange with a non-local exchange based on KS-orbitals, the so-called ‘exact’ exchange, E_X^{HF} , based on the adiabatic connection.¹²³

$$E_{XC}^{hybrid} = E_C^{(meta-)GGA} + (1 - \alpha_X)E_X^{(meta-)GGA} + \alpha_X E_X^{HF} \quad Eq. 3.26$$

These are the hybrid functionals, a widely applied category that utilises the same Fock-exchange part in the space density are the global hybrids, with the most popular B3LYP. This is factored ($\alpha_X = 0.2$), a 20% Fock-exchange, Slater’s LDA at 0.08%, B88 exchange at 0.72%, VWN-LDA 0.19% and LYP correlations at 0.81%.^{124, 125} The Minnesota functionals (M05¹²⁶ and M06¹²⁷) are the prominent hybrid meta-GGA functionals, with an extensive parametrization up to 40 parts. Notably, PW6B95,¹²⁸ is one of the hybrid functionals factored ($\alpha_X = 0.28$) especially successful for species containing NCIs.

This group of global hybrid functionals presented the issue of exponential decay in most meta-GGA based hybrids ($\frac{\alpha_X}{r}$ instead of the accurate $\frac{1}{r}$ decay). The approach to amend this error was set with a range-separation method and the range-separation functionals, that aim to yield a 100% of the Fock exchange in the asymptotic limit. It adds an error function, *erf*, and partitions the two-electron operator $\frac{1}{r_{12}}$ to long and short range.^{129, 130}

$$\frac{1}{r_{12}} = \frac{1 - \text{erf}(\mu r_{12})}{r_{12}} + \frac{\text{erf}(\mu r_{12})}{r_{12}} \quad Eq. 3.27$$

The HF exchange is represented by the long-range component and the exchange by the short-range. This approximation produced functionals with optimised predictions on many charged and non-covalently bound species, such as the ω B97, ω B97X, and their dispersion corrected

ω B97X-D¹³¹ and ω B97X-D3.¹³² Specifically for excited state computations, these range-separated functionals, are able to apply TD-DFT (time-dependent density functional theory), for more accurate orbital energy predictions.

3.4.1.5 Virtual-Orbital Dependent Functionals

Normally assigned on the higher rung of Jacobs ladder are the virtual-orbital dependent functionals. These parametrize the KS-orbitals through the correlation energy. These approximations have been made by several different ways, such as, notably, a perturbation approach by Görling and Levy,¹³³ the RPS (random phase approximation),¹³⁴ and the most popular DHDF (double-hybrid density functional) by Grimme.¹³⁵ These functionals utilise the preceding SCF energy KS orbitals, applying second order Møller-Plesset perturbation theory (MP2) to compute the exchange-correlation.¹³⁶

$$E_{XC}^{DHDF} = (1 - \alpha_X)E_X^{(meta-)GGA} + \alpha_X E_X^{HF} + (1 - \alpha_C)E_C^{(meta-)GGA} + \alpha_C E_C^{MP2} \quad Eq. 3.28$$

Although these functionals are not used in this work, it is noted that their accuracy and computational cost demands a higher percentage of Fock exchange.

3.5 Semi-empirical Methods

As an initial attempt to further reduce computational cost in early eras, where it was scarce for systems of any size, approximations that drastically simplify HF or DFT, were made with semi-empirical methods. These are especially sought after nowadays with a new scientific demand for computational predictions on large supramolecular and biomolecular entities. Pre-set parameters, that are specifically aiming to replicate empirical (traditionally experimental) or high-accuracy *ab initio* data, are inserted to lift the computational cost. This works has made use of the PMx (PM6, PM6-D3H4X, and PM7) Parametric Methods,^{103, 137, 138} which are founded on MNDO type methods. Subsequently, evaluating and applying a new DFTB (tight-binding DFT) approximation the GFN2-xTB.¹⁰⁵ DFTB approximations belong to a different

conceptual category than MNDO type semi-empirical Hamiltonians, and are described in later sections of this Chapter (3.5.2).

3.5.1 MNDO-type Methods

In MNDO methods the standard approximations reduce computations to valence electrons only, and a smaller set of integral by use of minimal basis set STO (Slater type orbitals). One of the founding approximations in this category of semi-empirical Hamiltonians is the NDDO (neglect of diatomic differential overlap). NDDO approximation sets to zero the products of basis functions of the same electron coordinates but on different atoms (A and B).¹³⁸

$$S_{\mu\nu} = \langle \psi_\nu | \psi_\mu \rangle \equiv \langle \mu | \nu \rangle = \delta_{\mu\nu} \delta_{AB} \quad \text{Eq. 3.29}$$

The overlap matrix \mathbf{S} is then simplified to a unity matrix, and the eigenvalue solved is different to that from HF, normally with a large error in the exchange part. A further simplification is the three-centred (from the operator and the two basis functions) electron integrals are set to zero leaving this form:

$$\langle \mu_A | \hat{h} | \nu_A \rangle = \left\langle \mu_A \left| -\frac{1}{2} \nabla^2 - \frac{Z'_A}{|R_A - r|} \right| \nu_A \right\rangle - \sum_{a \neq A}^{nuclei} \left\langle \mu_A \left| \frac{Z'_a}{|R_a - r|} \right| \nu_A \right\rangle \quad \text{Eq. 3.30}$$

$$\langle \mu_A | \hat{h} | \nu_B \rangle = \left\langle \mu_A \left| -\frac{1}{2} \nabla^2 - \frac{Z'_A}{|R_A - r|} - \frac{Z'_B}{|R_B - r|} \right| \nu_B \right\rangle \quad \text{Eq. 3.31}$$

With \hat{h} the one-electron operator, Z'_A the nuclear charge of the completely shielded system without the core electrons. This is then followed by omitting the three/four-electron integrals.

$$\langle \mu_A \nu_B | \lambda_C \sigma_D \rangle = \delta_{AC} \delta_{BD} \langle \mu_A \nu_B | \lambda_A \sigma_B \rangle \quad \text{Eq. 3.32}$$

This lowers the number to 27 one/two-centred two electron integrals when considering s and p functions only, as opposed to 500 when d functions are included. Consequently, leaving five one-centred two-electron integrals for the remaining sp basis.

$$\begin{aligned} \langle ss|ss\rangle &= G_{ss}, \quad \langle sp|sp\rangle = G_{sp}, \quad \langle ss|pp\rangle = H_{pp}, \quad \langle pp|pp\rangle = G_{pp}, \\ \langle pp'|pp'\rangle &= G_{p^2} \end{aligned} \tag{Eq. 3.33}$$

The G parts represent the Coulomb parameters, and H the exchange integral. Specifically, G_{p^2} refers to two different types of p .

The succeeding INDO (intermediate neglect of differential overlap) follows by neglecting all non-Coulombic two-centred two electron integrals, and the CNDO (complete neglect of differential overlap) neglects the Coulombic one and two-centred two electron integrals.

The approach for all the later developed semi-empirical Hamiltonians based on NDDO is to vary the remaining integrals included, by parametrizing from experimental data or fitting the parameters to given experimental data. Historically, the first approximations with MNDO,¹³⁹ AM1 (Austin model 1),¹⁴⁰ and PM3 (parametric method 3)¹⁴¹ applied s and p functions only, with exclusive calculation of the overlap $S_{\mu\nu}$. They all have different parametrization data, with PM3 specifically including parameters from an extended list of experimental results. It is noted that these first Hamiltonians were only parametrized for a handful of basic elements.

The MOPAC2016 sourced PM6, PM6-D3H4X and PM7 are used in Chapters 4 and 5 of this work. These are successors to the original PM3. They mainly apply improvement on core-core interactions, an spd basis set and an extended element list parametrization. PM6 originally tackles the main faults presented in PM3 and improves errors for a wide set of systems. In PM6-D3H4X, the dispersion correction was added as well as improved parametrization for halogenic entities, and long-range intermolecular interactions. The most recent PM7 attempted to produce a robust Hamiltonian which could predict a larger set of entities, while containing a dispersion correction. It is reported to underperform PM6-D3H4X for non-covalently bound species. Other methods under this MDDO category have been developed such as the OMx (OM2, OM3) yet they are not relevant within this work.¹⁴²

3.5.2 Density Functional Tight Binding Methods

Analogous approximations to the MNDO on HF have been made on DFT, giving the DFTB semi-empirical methods. These produce the required electron density by adding a reference density of a system ρ_0 to the deviation from the ground state ρ , $\Delta\rho$. A Taylor series is applied to expand the E_{XC} for up to second order $\Delta\rho$.

$$E_{XC}[\rho(r) + \delta\rho(r)] = E_{XC}[\rho_0(r)] + \int \left[\frac{\delta E_{XC}[\rho(r)]}{\delta[\rho(r)]} \right]_{\rho_0} \Delta\rho(r) dr + \frac{1}{2} \int \left[\frac{\delta^2 E_{XC}[\rho(r)]}{\delta^2[\rho(r')]} \right]_{\rho_0, \rho'_0} \Delta\rho(r) \Delta\rho(r') dr dr' \quad Eq. 3.34$$

The total energy is then given by:

$$E[\rho(r) + \delta\rho(r)] = \underbrace{\sum_i \left\langle \phi_i \left| -\frac{1}{2} \nabla^2 + V_{en} + \int \frac{\rho_0(r')}{|r-r'|} dr + v_{XC} \right| \phi_i \right\rangle}_{E^{H0}} - \underbrace{\frac{1}{2} \int \frac{\rho(r)\rho(r')}{|r-r'|} dr dr' - \int v_{XC}[\rho_0(r)]\rho(r) dr + E_{XC}[\rho_0(r)] + E_{nm}}_{E^{rep}} + \underbrace{\frac{1}{2} \int \left(\frac{1}{|r-r'|} + \frac{\delta^2 E_{XC}}{\delta\rho(r)\delta\rho(r')} \right)_{\rho_0, \rho'_0} \Delta\rho(r) \Delta\rho(r') dr dr'}_{E^V} \quad Eq. 3.35$$

Here, the E^{H0} part is the energy of the reference density, ρ_0 is composed of superimposed atomic densities, thus the $H_{\mu\nu}^0$ is calculated from the atomic orbital basis.

$$E^{H0} = \sum_i \sum_{\mu \in a} \sum_{\nu \in b} n_i c_{\mu i} c_{\nu i} H_{\mu\nu}^0 \quad Eq. 3.36$$

The second term E^{rep} is the approximation of the DFT Coulomb interaction, core repulsion and exchange correlation. The approximation approach is produced by a summation of one-centred, and two-centred (short range) potentials V_{ab}^{rep} . DFTB omits the second order term,

therefore SCC (self-consistent field) is not used to calculate this. The SCC charge is approximated with E^{γ} which is calculated iteratively.

$$E^{\gamma} = \frac{1}{2} \sum_{ab} \Delta q_a \Delta q_b \gamma_{ab} \quad Eq. 3.37$$

An atom α is assigned the net charge $\Delta q_{\alpha} = q_{\alpha} - q_{0,\alpha}$, and the function γ accounts for the electron-electron interaction, via an integral of the normalised product of spherical charge densities. In the large limit of r , γ becomes $1/r$. The self-repulsion is introduced with a Hubbard term (U_{α}) for $\alpha=b$. The sum of E^{H0} , E^{rep} , and E^{γ} as described above give the total energy of the SCC-DFTB, this is the only form of DFTB applied in this work.

3.5.2.1 GFN2-xTB

Recent progress on the DFTB methods produced a less empirical DFTB Hamiltonian, Grimme's GFN-xTB¹⁰⁹ (geometry, frequency, non-covalent, eXtended-TB) and the updated GFN2-xTB. This tight-binding Hamiltonian aimed to tackle the previously mentioned disadvantages, concerning the monopole approximation, by including electrostatic and exchange correlation Hamiltonian terms, and simultaneously keeping the computational cost low. It additionally removed specific halogen and hydrogen bonding corrections. GFN2-xTB is set to be a 'globally' parametrized Hamiltonian, containing an updated dispersion correction D4¹⁰⁴ (applied self-consistently, via second order density fluctuations), parameters specific for 86 elements (up to radon) and further omits element pair-specific parameters. It offers an improvement to its predecessor GFN-xTB with an improved prediction of electrostatic effects. Furthermore, it was evaluated in a benchmark as an improvement from all previous DFTB approximations with superior description of electronic density and therefore organometallic and biomolecular species.

The GFN2-xTB Hamiltonian follows a similar approach to DFTB including up to third order terms with successive alterations/additions to each for enhancing a focused aspect. The zeroth

order terms include a NL (non-local) correlation functional term for the pairwise London dispersion contribution.

$$E^{(0)}[\rho_0] \approx \underbrace{E_{rep}^{(0)} + E_{disp}^{(0)}}_{\text{Buckingham,Lennard-Jones type term}} + \sum_A \left(E_{A,core}^{(0)} + E_{A,valence}^{(0)} \right) \quad Eq. 3.38$$

The atomic core and valence energies summation is added as a constant and is set to zero. $E_{rep}^{(0)}$ and $E_{disp}^{(0)}$ are set as their classical expressions in this method, in the form of the Buckingham/Lennard-Jones potential. This approach treats NCIs of noble gases more accurately. Subsequently, the first order terms of DFTB are augmented with a first order dispersion contribution in GFN2-xTB.

$$E^{(1)}[\rho_0, \delta\rho] \approx E_{disp}^{(1)} + \sum_A E_{A,valence}^{(1)} \quad Eq. 3.39$$

Where $\sum_A E_{A,valence}^{(1)}$, is a description of the electronic valence state.

The second order terms are treated in accordance to DFTB with the addition of anisotropic effects up to second order in the multipole expansion, and the self-consistent treatment of isotropic $\delta\rho$ effects within the D4 correction.

Lastly, the third order terms differ from previous DFTB methods with the Mulliken approximation. This treats the third order terms as partial shell charges. All other third order terms are neglected except an isotropic on-site term to account for short-range Coulomb and exchange-correlation interactions.

The total GFN2-xTB energy is composed as:

$$E_{GFN2-xTB} = E_{rep} + E_{disp} + E_{EHT} + E_{IES+IXC} + E_{AES} + E_{AXC} + F_{Fermi} \quad Eq. 3.40$$

Where the subscripts ‘IES’ indicate the isotropic electrostatic energy, ‘XC and ‘IXC’ the isotropic, ‘AES’ the anisotropic electrostatic, and ‘AXC’ anisotropic term energies. The notable improvement herein is the anisotropic contribution by higher order multipole terms.

This means that in addition to the broad parametrization added for ES and XC terms, GFN2-xTB extends beyond the monopole with up to second order multipole expansion.

The damping function is also adjusted for the anisotropic term, this damps the related terms for short R_{AB} . The exchange-correlation energy is equivalently composed by an isotropic and anisotropic term.

Finally, the D4 correction is added self-consistently with a tight-binding Fock matrix.

$$F_{\kappa\lambda}^{D_4} = -\frac{1}{2}S_{\kappa\lambda}(d_A + d_B), \quad \forall \kappa \in A, \lambda \in B \quad \text{Eq. 3.41}$$

Where κ and λ reflect the AOs and d_A, d_B are the dispersion coefficients.

Although this method is very recent, published results¹⁰⁵ showed improved predictions for non-covalently bound entities. This method is later benchmarked in this thesis (Chapter 4) for validity and subsequently applied throughout the thesis results.

3.6 Dispersion Correction

It was observed for these methods (mean-field) that consistently larger errors were made in predictions of non-covalently bound species. This is because they are unable to yield the dependence $-C_6/R^6$ (R being the inter-atomic distance) for the long-range correlation for the asymptotic limit, which is the London dispersion energy. Grimme's dispersion correction DFT-D3 is the one applied in the methods (DFT and semi-empirical) used in this thesis, and it is simply added within each method to correct the long-range London dispersion energy errors.

$$E^{DFT-D} = E^{DFT} + E_{disp}^D \quad \text{Eq. 3.42}$$

With this correction two intermolecularly bound systems A and B have a dispersion energy potential $V_{disp}(R_{AB})$. This treats the large R distances for this potential with use of perturbation theory, and the multipole expansion of $V(R)$.¹⁴³

$$V_{disp}(R_{AB}) = - \sum_{n \in \mathbb{Z}_+^* - \{1,2,3,4,5\}} \frac{C_n^{AB}}{R_{AB}^n} \quad Eq. 3.43$$

Where n indicates the order of the dispersion coefficient, C_n^{AB} . This is commonly applied as positive; hence a negative sign is applied to this summation to account for the negative dispersion energy. The systems are then considered all as spherical, thus only even n are summed, simplifying the calculation to:

$$E_{disp} = -\frac{1}{2} \sum_{A \neq B} \sum_{n=2\kappa, n \in \mathbb{Z}_+^* - \{1,2\}} \frac{C_n^{AB}}{R_{AB}^n} \quad Eq. 3.44$$

This gives diverting results for any short R_{AB} , consequently a damping function, is added to remove the short-range effect. In the first correction published DFT-D1¹⁴⁴, the dipole-dipole term C_6^{AB} is only accounted for from the multipole, and C_6 dispersion co-efficient values are based on experimental dipole oscillator strength distributions. These are made on averages of individual elements and hybridisation states. The update DFT-D2 incorporated the atomic static polarizabilities and ionization potential (up to element Xe in the periodic table of elements) calculated via PBE0/QZVP for producing C_6 values.

Currently, the most widely applied version for dispersion correction in DFT and semi-empirical methods is the DFT-D3 version.^{145, 146} This was specifically refined to a dipole-quadrupole order coefficient C_8^{AB} to include medium range dispersion. In addition, more empirical parameters were substituted with computed therein.

$$E_{disp}^{D3} = -\frac{1}{2} \sum_{A \neq B} \sum_{n=6,8} s_n \frac{C_n^{AB}}{R_{AB}^n + f_{damp}^{BJ} (R_0^{AB})^n} \quad Eq. 3.45$$

s_n is the global scaling factor, which changes for each functional to reflect asymptotic trends as needed. C_n^{AB} depends of the coordination number of each atom pair AB , and reflects the average isotropic dispersion coefficient. All coefficients with a higher order than $n=6$ are

calculated (through a recursion relationship) based on the Casimir-Polder relationship for C_6^{AB} .¹⁴⁷

$$C_6^{AB} = \frac{3}{4} \int_0^{\infty} \alpha^A(i\omega) \alpha^B(i\omega) d\omega \quad Eq. 3.46$$

Where $\alpha(i\omega)$ is the average dipole polarizability at corresponding reference system, with frequency $i\omega$ (i expressed imaginary) taken from TD-DFT computations for those systems. The original damping function, so-called ‘zero-damping’ function, within the D3(0) scheme has a zero limit in the short-range R_{AB} . This became erroneous for the repulsive forces. A correction was introduced with the Becke-Johnson (BJ) damping function f_{damp}^{BJ} .

$$f_{damp}^{BJ}(R_0^{AB}) = a_1 R_0^{AB} + a_2 \quad Eq. 3.47$$

Where $R_0^{AB} = \sqrt{C_8^{AB} + C_6^{AB}}$, sets the cut-off, and a_1, a_2 are the fitting parameters. The BJ damping function is applied constantly in the dispersion energy for calculating the correlation term.¹⁴⁸

An additional parameter considered for the dispersion correction is the many-body dispersion interactions, which is important for large systems. This opened the path to new improvements published. Notably the Axilrod-Teller-Muto term tackles the dipole-dipole-dipole interaction.^{149, 150} For a (3) three-body system A, B, C , with the averaged R distances between them this is:

$$E_{disp}^{(3)} = -\frac{1}{6} \sum_{A \neq B \neq C} \frac{C_9^{ABC} (3 \cos \theta_a \cos \theta_b \cos \theta_c + 1)}{(R_{AB} R_{BC} R_{AC})^3} f_{damp}^0(R_0^{AB}) \quad Eq. 3.48$$

Where $\theta_{a,b,c}$ are the respective angles formed by the three bodies, and C_9^{ABC} is approached by C_6 as previously described via the following relationship.

$$C_9^{ABC} \approx -\sqrt{C_6^{AB} C_6^{BC} C_6^{AC}} \quad Eq. 3.49$$

The D3 dispersion correction has the advantage of being applicable to most known elements, and incorporated in most widely used HF, DFT and semi-empirical methods, without additional implementation demands. It provides a relatively low computational cost to those methods, and yields consistent output as functioned by an increasing atom number in each system. On the other end, reported disadvantages showed issues in charged and metallized complexes. Further, there is a requirement for detailed parametrization per function in semi-empirical approaches. There is also an apparent absence of the density parameter, which leaves the electronic structure unaffected. This can be re-attributed indirectly through the geometry changes and the coordination number. Finally, C_6 which is the means of estimating the remaining order co-factors can be considered unfounded without a given reference system.

The latest update for a London dispersion scheme is the D4 scheme. This was concisely described as a more widely applied atomic-charge dependent scheme. It is reported to yield a drop in error as compared to D3 (by 0.9% mean relative deviation). This scheme includes wider elemental parametrization, three-body effects via an ATM term, and has included the specific BJ-type damping parameters so far for over 60 functionals. The previously disadvantageous description of metallized entities was improved by specific charge dependence terms. The D4 scheme was included in the mainly applied GFN2-xTB within this thesis.

3.7 Conformer-Rotamer Ensemble Sampling Tool

A key utility tool CREST, for the GFN-xTB and GFN2-xTB Hamiltonians has been created to scan available conformers of mainly large non-covalently bound species, via a tractable, for these entities, procedure.¹⁵¹ The low energy space (up to 6 kcal for conformers, and 30 kcal for tautomers/protomers) is applied through a quantum chemical screening. This produces isomers/conformers via a direct GFN2n-xTB calculation sampling. Additional energy windows are also incorporated in the program to account for rotations or tautomerization. This can be important for chemical processes, such as non-covalent bonding, or reactivity. Whereas

this is more computationally demanding than a chemoinformatics based library scan, it offers the open option of scanning the majority of energetically accessible structures, which are important for mechanistic insights. The CREST algorithm is applicable for most levels of theory, yet it has found its prime utility when applied to non-covalently bound complexes using GFN2n-xTB (and more generally, semi-empirical methods). The algorithm produces a range of conformers that exist on the PES (potential energy surface), still possessing the same covalent topology, while rotations around bonds and types of inversions produce rotamers. These are the thermally accessible minimum-energy structure set, called the CRE (conformer rotamer ensemble). This ensemble usually possesses a useful set of physical property information. There have been two sorting algorithms for the conformational search, the older MF-MD-GC (mode following, molecular dynamics sampling, genetic z-matrix crossing) method, and the newer iMTD-GC (an extensive meta dynamics, genetic z-matrix crossing) approach.¹⁵² The latter has shown superior performance in terms of speed and accuracy, noting minor larger computational demand for larger entities.

The process of the CREST workflow is initiated with an input structure, and a starting hessian calculation takes place on this structure. A MTD (meta dynamics) simulation with an incorporated screening process is then applied, with added biasing contribution factors. A one-dimensional PES is created to sample an ensemble of minima, which is reflective of the conformer set. The physical geometrical shifts between respective conformers correspond to connection in the PES which make this approach valid. Subsequently, an MD (molecular dynamics) type computation is carried out, and selected structures are subjected to a geometry optimisation. An energy threshold is then used to sort the produced isomers, and a subsequent RMSD (route mean square deviation) and rotational constant comparison to define conformers and rotamers, that are distinguished by the relative energy values within the conformer set. At

this point the latest sorting algorithm iMTD-GC works iteratively restating the cycle when a new lower conformer structure is found. This leads to the production of the final CRE.

Overall, the CREST application offers a tractable approach for monitoring conformational changes in a wide range of structures, with a key publication outlining a comparison with experimental data of small organic molecular sets and some selected metalized coordination with organic structures.¹⁵¹ These can only be reflective of the large category of molecules.

3.8 Chapter Summary

There is active developmental progress in the sector of computational predictivity for supramolecular entities. The accuracy and speed of those methods are continuously improved. The dispersion correction implemented within the majority of computational quantum chemistry methods provides improvements in the predictivity of non-covalently bound species. Additionally, reviewed parametrization of semi-empirical methods, and the new GFN2-xTB Hamiltonian offer new tractable and reliable techniques for examining large segments of real supramolecular polymers, which is important for studying their mechanisms and properties. Lastly, CREST provides a new tool for testing the intrinsic interactions and rotations of complexes to ascertain a more reliable prediction. With the focus of this thesis work, being the computational examination of large supramolecular polymers, the tools and methods discussed in this chapter are the most recent and relevant to the aims of this thesis.

Chapter 4

Benchmarking of Semi-Empirical Methods for the Modelling of Supramolecular Polymers

4.1 Introduction

Supramolecular aggregates are large species that assemble in a highly directional manner. Even though extensive experimental work has been carried out to determine the structural and chemical factors driving the assembly architecture and cooperativity, the computational prediction of these elements remains largely unexplored. Supramolecular assemblies stand in the middle ground in terms of size and chemical complexity, between small molecules and large protein or DNA fragments. The monomer units of these aggregates are commonly large, conjugated molecules, able to hold a variety of substituents. Therefore, they are held together mainly by NCIs (non-covalent interactions) these being: hydrogen bonds, π - π interactions and (occasionally) metallophilic interactions. The combination of their size and the “weak” interactions present poses a particular challenge for computational examination.¹⁸

4.1.1 Semi-Empirical Methods for Predicting Molecular Interactions

The appropriate computational tools for predicting supramolecular aggregates are selected by their capacity to make tractable computations in larger species with accuracy, these are, most favourably, semi-empirical quantum chemical methods. Semi-empirical computations apply computational cost-effective approximations, based on high level computations or experimental values with the capacity of effectively modelling systems up to a few thousand

atoms. This makes them ideal for the size of the species in question. In addition to that, an effective dispersion correction has been incorporated in most of the semi-empirical methods, making them suitable for modelling supramolecular aggregations.^{102, 106, 135, 144} The relevant, most recently developed, semi-empirical methods are: PM6¹⁰³, PM6-D3H4X and PM7¹³⁸ and the recently released GFN2-xTB.¹⁰⁵ (see Chapter 3 for a detailed overview of the theory behind these methods and how the dispersion correction is applied).

4.1.2 The Relevant Benchmark Data Sets for NCIs – S66 and S30L

There are numerous benchmark data sets for computational reference values. Each targets a different aspect, such as bond lengths of organic molecules, halogen or metal containing species. For addressing the computational problem of this work, the appropriate benchmark data sets must contain reference values concerning the different NCIs within biologically interesting species, as well as the quantitative description of NCIs in larger complexes. These are respectively addressed by the S66 and S30L data sets.

The S66 benchmark data set consists of 66 non-covalently bound pairs, of biological interest¹⁵³. This set gives a balanced outlook in terms of the examination of the different ‘weak’ interactions directing biological species, representing the dispersion and electrostatic contribution in a comprehensive set. The set was benchmarked for the CCSD(T)/CBS.

The S30L benchmark set consists of 30 supramolecular host-guest complexes.¹⁰⁷ These include non-polar dispersion, π - π interactions, halogen, CH- π , hydrogen bonded and charged species, accompanied by their counter-ions. The set features species of a larger scale than the previous benchmark (up to 200 atoms). The reference benchmark BEs were collected at the PW6B95-D3/def2-QZVP level.

4.1.3 Notable Benchmark Literature for Semi-Empirical Methods on NCIs

The work outlined in this section focuses on the literature in which the ability of semi-empirical Hamiltonians to predict binding energies associated with NCIs was investigated. A number of datasets with binding energies calculated using high-level computational methods have been utilised in these studies, the key ones being as defined in Table 4.1. The work by Christensen *et al.* is considered the most comprehensive for this aim.¹⁵⁴ Within this study, PM6-D3H4X is reported to outperform PM7 and PM6 for the data sets reviewed. Similarly, Korth *et al.* compared a wide variety of methods, including AM1 and the OMx (semi-empirical orthogonalization models) methods on the benchmark data set GMTKN24 (a general database for main group thermochemistry, kinetics and non-covalent interactions).¹⁵⁵ However, as with the work of Christensen *et al.* this work also lacks the newest DFTB methods.¹⁵⁶ Again, in the context of NCIs, PM6-D3 and PM6-D3H4 were found to yield the smallest errors (in terms of RMSE) with respect to the benchmark dataset.

The work of Hostas *et al.* compared the various PMx (PM6, PM6-D3, PM7) methods with each other, again showing the PM6-D3 methods to outperform PM7.¹⁵⁷ Finally, Bannwarth *et al.* carried out a comprehensive benchmark analysis during the development of the GFN2-xTB method.¹⁰⁵ This is the only published study that has compared this DFTB method with the PMx methods, finding GFN2-xTB to be superior, giving the smallest RMSE and best correlation to reference values in terms of binding energies. The key finding in all of the published literature is the importance of the dispersion correction within semi-empirical methods for dealing with the types of interactions found in supramolecular systems.

Aim

This Chapter aims to identify the most predictive semi-empirical method for the calculation of binding energies in non-covalently bound species present in the S66 and S30L sets. Predictivity of each semi-empirical method was assessed in terms of RMSE (route mean square error) following the development of quantitative structure-activity models.

Table 4.1 The relevant benchmark publications listed with the semi-empirical methods they review, and the benchmark data sets utilised for their benchmark.

Publication	Semi-empirical methods compared	Benchmark reference set
Christensen <i>et al.</i> (2016) ¹⁵⁴	DFTB, PM6, PM6-D3H4, OM2, OM2-D3, OM3	S66
Hostas <i>et al.</i> (2013) ¹⁵⁷	PM6, PM6-DH2(X), PM6-DH+, PM6-D3H4(X), PM7	S66
Korth <i>et al.</i> (2011) ¹⁵⁶	OM1, OM2, OM3, PM6, AM1, SCC-DFTB (<i>all with and without D3 correction</i>)	GMTKN24-hnco (<i>includes S66</i>)
Bannwarth <i>et al.</i> (2019) ¹⁰⁵	GFN2-xTB, GFN-xTB, PM6-D3H4X, DFTB3-D3	GMTKN55 (<i>includes S30L, S66</i>)

4.2 Methodology

4.2.1 Computational Methods Investigated

For the geometry optimisations of the S30L data set the cartesian coordinates were collected as provided in the references. The molecular coordinates from the S66 data set were inserted with chemically sensible coordinates as advised by the publication supporting information visualisations.

Semi-empirical computations were carried out with the MOPAC16 software; the Hamiltonians PM6, PM6-D3H4X and PM7 were applied (see Table 4.2). Additionally, the GFN2-xTB semi-empirical method was also utilised. Geometrical optimisations and frequency calculations were computed for all structures in the gas phase. A number of complexes showed small negative eigenvalues that were related to simple geometrical distortions. These were considered minor and thus were ignored. The energy convergence criteria were left as software default for all structures. These criteria are for the energy convergence (Econv), the allowed change in the total energy at convergence 5×10^{-6} Eh, while the gradient convergence (Gconv) the allowed change in the gradient norm at convergence 1×10^{-3} Eh $\times \alpha^{-1}$. All optimised geometries, their coordinates and graphical files are found in A5_G5.

Table 4.2 Semi-empirical methods, and associated dispersion corrections, utilised in this Chapter.

	PM6	PM6-D3H4X	PM7	GFN2-xTB
Type of dispersion correction	None	D3	D3	D4

4.2.2 Calculation of Binding Energies

The BEs of all species were calculated as shown in equation 4.1.

$$\text{BE} = E(\text{complex}) - E(\text{reactant 1}) - E(\text{reactant 2}). \quad \text{Eq. 4.1}$$

Where E is the total electronic energy of the respective optimised structures.

4.2.3 Statistics

The RMSE was calculated to quantify the error of each semi-empirical method. RMSE values were calculated between BM-BE (benchmark binding energies) and Calc-BE (calculated binding energies) for each method in Tables 4.3 and 4.6. The following equation was used to calculate RMSE values:

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (C_i - B_i)^2}{n}} \quad Eq. 4.2$$

Where, n is the number of values, C_i is the computed value (Calc-BE in Tables 4.3 and 4.6) and B_i is the benchmark value (BM-BE in Tables 4.3 and 4.6).

Linear regression analysis was utilised to correlate the calculated binding energies for each semi-empirical method with those from the benchmark datasets. This enabled predicted benchmark binding energies for each semi-empirical method to be generated.

4.3 Results and Discussion

The aim of this Chapter was to undertake a benchmarking exercise to identify the most reliably predictive semi-empirical method for dealing with the types of NCIs present in supramolecular systems. The methods investigated were: GFN2-xTB, PM7, PM6-D3H4X, and the non D3 corrected PM6. The latter was included as a “control” to enable the importance of including a dispersion correction to be assessed. The S66 and S30L datasets were identified for the analysis as these contain a range of chemical species covering NCIs.

4.3.1 The S66 Data Set

This subsection presents the benchmark of the S66 data set. This set features small organic and biologically important molecules, with high-accuracy (CCSD(T)) benchmark reference BEs. CCSD(T) numerical analysis has been established in literature as the “gold standard” for

dealing with NCIs with the highest accuracy within the currently available theoretical chemistry modelling methods.¹⁵⁸ The computed and predicted (with linear regression) BEs of each method are discussed. The data set includes 22 non-covalently bound pairs containing hydrogen bonding interactions, 22 with π - π interactions 22 complexes with the remaining mainly Van der Waals type within 66 small non-covalently bound pairs.

4.3.1.1 Analysis of Benchmark Calculations for the S66 Dataset

Table 4.3 gives the summary of the calculated BE values (BM-BE, Calc-BE and Pred-BE) computed for each of the semi-empirical methods investigated. In addition, Table 4.4 shows the linear regression parameters, and associated statistics, used to correlate the BM-BE values with the Calc-BE values, enabling the generation of the Pred-BE (predicted binding energies) column for each method. Inspection of the gradient and intercept values in Table 4.4 highlights the importance of this step, in terms of ensuring that it is the relative trend in the values that is being compared between each method. Simply stated, the analysis in this Chapter was not concerned with the size of the Calc-BE values, but rather the ability of each semi-empirical method to quantitatively rank the chemicals in the same order as calculated in the benchmark set. The exception to this being the importance of the Calc-BE value being negative – meaning the system is not predicted to bind. The results showed that GFN2-xTB significantly outperformed the other semi-empirical methods in terms of its correlation with the benchmark values and the calculated RMSE values (R^2 and RMSE values respectively in Table 4.4). In addition, GFN2-xTB also calculated fewer positive binding energies values (Calc-BE values taken from Table 4.3 and summarised in Table 4.4).

Table 4.3 Summary of binding energies values for the S66 dataset. The reference benchmark values were calculated at the CCSD(T)/CBS scheme.¹⁵³ The H-bonded pairs are highlighted in light blue, the dispersion-bound pairs are highlighted in yellow, and the remaining pairs bound by any other NCIs are highlighted in light green. BM-BE are benchmark binding energy, Calc-BE the calculated, Pred-BE are the predicted via linear regression parameters shown in Table 4.4. Positive BE values are highlighted in italics and a box. All values are in kcal/mol.

Interaction	Structure	CCSD(T)	PM6		PM6-D3H4X		PM7		GFN2-xTB	
		BM-BE	Calc-BE	Pred-BE	Calc-BE	Pred-BE	Calc-BE	Pred-BE	Calc-BE	Pred-BE
Hydrogen bonds	S66-1	-4.92	-16.32	-7.30	-17.29	-8.36	-7.27	-5.23	-0.18	-5.96
	S66-2	-5.59	-18.29	-7.96	-17.20	-8.33	-7.87	-5.48	-0.18	-5.89
	S66-3	-6.91	-18.29	-7.96	-11.17	-6.48	-17.18	-9.43	-0.20	-6.44
	S66-4	-8.10	-26.65	-10.76	-20.00	-9.19	-15.59	-8.75	-0.27	-8.41
	S66-5	-5.76	-16.20	-7.26	-13.95	-7.33	-4.87	-4.21	-0.18	-5.85
	S66-6	-7.55	-15.33	-6.96	-10.04	-6.13	-13.92	-8.05	-0.21	-6.78
	S66-7	-8.23	-24.37	-9.99	-19.37	-9.00	-11.72	-7.12	-0.27	-8.41
	S66-8	-5.01	-14.52	-6.70	-13.09	-7.07	-4.16	-3.91	-0.18	-5.76
	S66-9	-3.06	-11.60	-5.71	-8.54	-5.67	-8.22	-5.63	-0.10	-3.55
	S66-10	-4.16	-10.00	-5.18	-5.84	-4.84	-10.48	-6.59	-0.12	-4.02
	S66-11	-5.42	-16.26	-7.28	-12.91	-7.01	-13.55	-7.89	-0.15	-5.05
	S66-12	-7.27	-19.45	-8.35	-12.45	-6.87	-18.39	-9.94	-0.20	-6.34
	S66-13	-6.19	-14.50	-6.69	-13.13	-7.08	-11.25	-6.91	-0.17	-5.72

Interaction	Structure	CCSD(T)	PM6		PM6-D3H4X		PM7		GFN2-xTB	
		BM-BE	Calc-BE	Pred-BE	Calc-BE	Pred-BE	Calc-BE	Pred-BE	Calc-BE	Pred-BE
Hydrogen bonds	S66-14	-7.45	-16.03	-7.20	-7.94	-5.49	-18.05	-9.79	-0.23	-7.38
	S66-15	-8.63	-19.86	-8.48	-19.82	-9.14	-15.70	-8.80	-0.25	-7.81
	S66-16	-5.12	-16.38	-7.32	20.36	3.21	-12.39	-7.40	-0.16	-5.44
	S66-17	-17.18	-43.44	-16.38	-43.21	-16.32	-31.12	-15.33	-0.58	-17.57
	S66-18	-6.86	-15.89	-7.15	-8.71	-5.72	-14.69	-8.37	-0.19	-6.05
	S66-19	-7.41	-13.12	-6.22	-0.06	-3.13	-11.37	-6.96	-0.20	-6.50
	S66-20	-19.09	-42.14	-15.95	-40.30	-15.43	-26.25	-13.27	-0.62	-18.78
	S66-21	-16.26	-44.39	-16.70	-40.76	-15.57	-33.57	-16.37	-0.57	-17.39
	S66-22	-19.49	-37.70	-14.46	-32.83	-13.14	-21.67	-11.33	-0.54	-16.55
	S66-23	-19.19	-39.16	-14.95	-32.66	-13.08	-26.44	-13.35	-0.51	-15.72
Dispersion Bound	S66-24	-2.82	0.56	-1.64	1.08	-2.72	-0.89	-2.53	-0.15	-5.01
	S66-25	-3.90	-3.99	-3.17	-3.64	-4.17	-5.49	-4.47	-0.19	-6.06
	S66-26	-9.83	-18.19	-7.92	-14.96	-7.65	-11.29	-6.93	-0.34	-10.62
	S66-27	-3.44	-1.43	-2.31	-0.52	-3.21	-3.38	-3.58	-0.17	-5.46
	S66-28	-5.71	-6.63	-4.05	-6.14	-4.94	-4.46	-4.04	-0.22	-7.21
	S66-29	-6.82	-12.98	-6.18	-12.72	-6.96	-10.84	-6.74	-0.28	-8.76
	S66-30	-1.43	0.45	-1.68	1.15	-2.70	-0.26	-2.26	-0.07	-2.79
	S66-31	-3.38	-3.27	-2.92	-1.79	-3.60	-1.97	-2.98	-0.12	-4.02

Interaction	Structure	CCSD(T)	PM6		PM6-D3H4X		PM7		GFN2-xTB	
		BM-BE	Calc-BE	Pred-BE	Calc-BE	Pred-BE	Calc-BE	Pred-BE	Calc-BE	Pred-BE
Dispersion bound	S66-32	-3.74	-3.69	-3.07	-3.64	-4.17	-2.40	-3.16	-0.12	-4.22
	S66-33	-1.87	-0.81	-2.10	0.67	-2.84	-1.30	-2.70	-0.09	-3.12
	S66-34	-3.78	-2.54	-2.68	-3.38	-4.09	-1.71	-2.87	-0.07	-2.75
	S66-35	-2.61	-1.55	-2.35	-2.69	-3.88	-0.84	-2.50	-0.06	-2.26
	S66-36	-1.78	-2.22	-2.57	-4.43	-4.41	-1.78	-2.90	-0.06	-2.33
	S66-37	-2.40	-2.53	-2.68	-3.40	-4.09	-2.77	-3.32	-0.07	-2.63
	S66-38	-3.00	-1.40	-2.30	-3.14	-4.01	-2.77	-3.32	-0.08	-2.90
	S66-39	-3.58	-3.20	-2.90	-3.20	-4.03	-1.58	-2.82	-0.01	-0.60
	S66-40	-2.90	-3.27	-2.92	-3.29	-4.06	-4.04	-3.86	-0.12	-3.99
	S66-41	-4.85	-2.94	-2.81	7.04	-0.89	-0.92	-2.54	-0.12	-4.21
	S66-42	-4.14	-5.34	-3.62	-1.98	-3.66	-3.38	-3.58	-0.15	-4.97
	S66-43	-3.71	-2.61	-2.70	3.27	-2.05	-1.04	-2.59	-0.11	-3.72
	S66-44	-2.01	-1.88	-2.46	-1.56	-3.53	-1.86	-2.94	-0.05	-2.17
	S66-45	-1.75	-0.75	-2.08	-1.28	-3.44	-0.95	-2.55	-0.05	-2.02
S66-46	-4.26	-4.99	-3.50	0.59	-2.87	-4.01	-3.85	-0.10	-3.41	
Other	S66-47	-2.88	-3.31	-2.94	-3.31	-4.07	-2.71	-3.30	-0.09	-3.11
	S66-48	-3.54	-2.25	-2.58	-2.23	-3.73	-1.30	-2.70	-0.07	-2.74
	S66-49	-3.33	-4.69	-3.40	-4.28	-4.37	-4.05	-3.87	-0.09	-3.26

Interaction	Structure	CCSD(T)	PM6		PM6-D3H4X		PM7		GFN2-xTB		
		BM-BE	Calc-BE	Pred-BE	Calc-BE	Pred-BE	Calc-BE	Pred-BE	Calc-BE	Pred-BE	
Other	S66-50	-2.87	-3.95	-3.15	-3.95	-4.26	-3.63	-3.69	-0.07	-2.79	
	S66-51	-1.52	-1.67	-2.39	-1.67	-3.56	-1.59	-2.82	-0.05	-2.20	
	S66-52	-4.71	-11.17	-5.57	-5.46	-4.72	-9.48	-6.16	-0.15	-4.90	
	S66-53	-4.36	-10.78	-5.44	-9.37	-5.93	-9.21	-6.05	-0.13	-4.53	
	S66-54	-3.28	-10.98	-5.51	-10.12	-6.16	-9.13	-6.02	-0.08	-3.07	
	S66-55	-4.19	-7.66	-4.40	-7.44	-5.33	-7.06	-5.14	-0.12	-4.04	
	S66-56	-3.23	-5.76	-3.76	-6.12	-4.93	-6.76	-5.01	-0.10	-3.57	
	S66-57	-5.28	-8.98	-4.84	-6.95	-5.18	-9.23	-6.06	-0.15	-4.92	
	S66-58	-4.15	-11.45	-5.67	-11.22	-6.50	-9.58	-6.21	-0.11	-3.89	
	S66-59	-2.85	-7.08	-4.20	-6.65	-5.09	-4.93	-4.24	-0.07	-2.66	
	S66-60	-4.87	-8.72	-4.75	-4.71	-4.50	-7.58	-5.36	-0.15	-4.95	
	S66-61	-2.91	-12.19	-5.91	-1.13	-3.40	-3.33	-3.56	-0.12	-4.00	
	S66-62	-3.53	-7.49	-4.34	-5.17	-4.64	-3.71	-3.72	-0.13	-4.29	
	S66-63	-3.80		<i>14.01</i>	<i>2.86</i>	<i>15.59</i>	<i>1.74</i>	<i>18.37</i>	<i>5.63</i>	<i>0.08</i>	<i>1.70</i>
	S66-64	-3.00		-4.70	-3.40	-3.21	-4.04	-3.27	-3.54	-0.08	-2.91
	S66-65	-3.99		-5.88	-3.80	-5.85	-4.84	-4.99	-4.26	-0.12	-4.01
	S66-66	-3.97		-6.86	-4.13	-2.13	-3.70	-8.09	-5.58	-0.12	-4.12

Table 4.4 Linear regression correlation parameters used to relate BM-BE (Y axis) values to Calc-BE (X axis) values for each semi-empirical method. The gradient (M) and intercept (C) values were used to generate the Pred-BE values are shown. RMSE calculated between BM-BE and Pred-BE for each method.

Semi-empirical Method	M	C	R ²	RMSE	No. positive BE
PM7	1.77	-2.15	0.70	2.25	1
PM6-D3H4X	1.29	-3.05	0.69	2.28	8
PM6	1.40	-1.83	0.83	1.69	3
GFN2-xTB	1.08	-0.59	0.91	1.23	1

An overview of the calculated RMSE values is shown in Table 4.5 (colour coding as proposed by Sure *et al.*).¹⁵⁹ GFN2-xTB is shown to outperform all other semi-empirical methods.

Table 4.5 RMSE in kcal/mol, where N shows the number of complexes in each class of NCI. Colour-coded categorization indicates favourable RMSE in green, acceptable in yellow, and high in red as shown in the reference range on the left.

	PM7	PM6-D3H4X	PM6	GFN2-xTB
All complexes N = 66	2.25	2.28	1.69	1.23
H-bonded N = 23	2.87	3.14	2.10	1.12
Dispersion N = 24	1.08	1.45	0.93	1.28
Other N = 19	2.44	1.85	1.84	1.30

RMSE<1.5
1.5<RMSE<2.0
2.0<RMSE

In addition, GFN2-xTB was also calculated to be the best performing within three of the four categories of weak interaction present in the dataset. The exception to this being for complexes within the dispersion class, for which the best performing method was calculated to

be PM6. This is an interesting result, given that the PM6 method is the only method lacking the specific dispersion correction.

4.3.2 The S30L Data Set Benchmark

For the purpose of evaluating the range of error amongst methods and throughout interactions in larger species, the S30L set features an almost complete collection of NCIs in large, mostly organic, host-guest systems, both charged and neutral. The electronic binding energies, extrapolated from experimental free energies, at a PW6B95-D3/def2-QZVP level, are utilised as reference for the statistical error analysis.¹⁰⁷ PW6B95 has shown superior capacity for supramolecular species and the majority of NCIs contained by them and is thus set as the benchmark reference with this set.^{107, 160} The dataset of 30 complexes contained through categorisation indicating their respective NCIs, two non-polar dispersion, ten π - π stacking, two CH- π interaction, two halogen bonded, seven hydrogen bonded and seven charged systems.

4.3.2.1 Analysis of Benchmark Calculations for the S30L Dataset

Table 6 shows a summary of the calculated BE values (BM-BE, Calc-BE and Pred-BE) computed for each of the semi-empirical methods investigated for the S30L dataset. In addition, Table 4.7 shows the linear regression parameters, and associated statistics, used to correlate the BM-BE values with the Calc-BE, enabling the generation of the Pred-BE column for each method. Inspection of the gradient and intercept values in Table 4.7 further highlights the importance of this step as these values are significantly more varied than the equivalent values calculated for the S66 dataset (compare Tables 4.7 and 4.4). The results for GFN2-xTB with the S66 dataset, the results showed to significantly outperform the other semi-empirical methods in terms of its correlation with the benchmark values and the calculated RMSE values (R^2 and RMSE values respectively in Table 4.7). In addition, GFN2-xTB also calculated fewer positive BEs (Calc-BE values highlighted in italics in Table 4.6 and summarised in Table 4.7).

Table 4.6 Summary of the binding energy values for the S30L dataset. The reference benchmark values, were calculated at the PW6B95-D3/def2-QZVP level of theory.¹⁰⁷ The systems are sub-categorized within the second column non-polar dispersion, π - π stacking, CH- π interaction, halogen bonded, hydrogen bonded and charged systems. BM-BE refers to the calculated benchmark binding energy, Calc-BE refers to the calculated binding energies, Pred-BE refers to predicted binding energies using the linear regression parameters shown in Table 4.7. The positive BEs are highlighted in italics. All values are in kcal/mol.

Interaction	Structure number	PW6B95	PM6		PM6-D3H4X		PM7		GFN2-xTB	
		BM-BE	Calc-BE	Pred-BE	Calc-BE	Pred-BE	Calc-BE	Pred-BE	Calc-BE	Pred-BE
non-polar dispersion	S30L-1	-32.08	-10.98	-32.32	-2.97	-35.02	-8.98	-33.23	-26.72	-28.96
	S30L-2	-21.72	-5.76	-28.74	-1.02	-34.20	-3.98	-29.45	-18.85	-21.57
π-π stacking	S30L-3	-26.17	-9.74	-31.47	-3.50	-35.24	-6.81	-31.59	-28.09	-30.24
	S30L-4	-21.07	-9.53	-31.33	-1.88	-34.56	-12.67	-36.01	-23.78	-26.19
	S30L-5	-35.19	-13.77	-34.23	-13.62	-39.50	-10.76	-34.57	-35.85	-37.52
	S30L-6	-31.38	-12.39	-33.29	-6.26	-36.40	-12.34	-35.77	-32.26	-34.15
	S30L-7	-34.93	<i>1.76</i>	-23.58	<i>2.07</i>	-32.90	<i>1.46</i>	-25.35	-46.63	-47.63
	S30L-8	-39.73	<i>2.22</i>	-23.27	-1.47	-34.39	<i>1.37</i>	-25.42	-54.47	-54.99
	S30L-9	-34.98	-3.90	-27.47	-2.51	-34.82	-0.73	-27.01	-39.19	-40.65

Interaction	Structure number	PW6B95	PM6		PM6-D3H4X		PM7		GFN2-xTB	
		BM-BE	Calc-BE	Pred-BE	Calc-BE	Pred-BE	Calc-BE	Pred-BE	Calc-BE	Pred-BE
π-π stacking	S30L-10	-36.05	-4.03	-27.56	-2.81	-34.95	-0.17	-26.58	-40.79	-42.16
	S30L-11	-43.42	-1.46	-25.79	1.83	-33.00	2.74	-24.39	-48.62	-49.50
	S30L-12	-42.88	2.00	-23.42	5.54	-31.44	2.77	-24.36	-48.94	-49.80
CH-π interaction	S30L-13	-28.82	-10.74	-32.16	-39.36	-50.34	-5.60	-30.68	-22.78	-25.26
	S30L-14	-31.32	-10.09	-31.71	-52.66	-55.94	-6.06	-31.03	-26.62	-28.86
Halogen bonds	S30L-15	-18.17	-60.95	-66.58	-24.71	-44.17	-4.78	-30.06	-25.20	-27.53
	S30L-16	-24.51	-42.75	-54.11	-29.50	-46.19	-8.38	-32.78	-26.98	-29.19
Hydrogen bonds	S30L-17	-32.92	-21.01	-39.20	-13.15	-39.31	-15.23	-37.95	-28.81	-30.92
	S30L-18	-21.41	-13.87	-34.31	-13.84	-39.60	-7.97	-32.47	-22.76	-25.24
	S30L-19	-16.64	-9.43	-31.26	-4.98	-35.86	0.45	-26.12	-13.68	-16.71
	S30L-20	-20.18	-14.53	-34.75	-7.33	-36.85	-1.35	-27.47	-19.63	-22.30
	S30L-21	-28.98	-7.16	-29.7	103.83	9.96	5.47	-22.33	-22.48	-24.98
	S30L-22	-33.89	-25.32	-42.15	-22.63	-43.30	-19.93	-41.5	-37.84	-39.39
	S30L-23	-58.67	-46.93	-56.97	-48.82	-54.32	-48.42	-63.00	-61.59	-61.67

Interaction	Structure number	PW6B95	PM6		PM6-D3H4X		PM7		GFN2-xTB	
		BM-BE	Calc-BE	Pred-BE	Calc-BE	Pred-BE	Calc-BE	Pred-BE	Calc-BE	Pred-BE
Charged	S30L-24	-139.23	-128.97	-113.23	-77.27	-66.3	-124.18	-120.17	-140.74	-135.94
	S30L-25	-33.17	-6.84	-29.49	11.04	-29.12	-5.62	-30.70	-29.17	-31.26
	S30L-26	-33.14	-5.64	-28.66	-0.39	-33.93	-5.04	-30.26	-29.32	-31.39
	S30L-27	-84.41	-75.17	-76.34	-75.05	-65.37	-85.5	-90.98	-83.82	-82.53
	S30L-28	-80.68	-73.46	-75.17	-61.24	-59.55	-85.51	-90.99	-80.80	-79.70
	S30L-29	-54.79	-39.99	-52.21	-46.32	-53.27	-33.69	-51.88	-51.73	-52.42
	S30L-30	-50.59	-35.03	-48.81	5.97	-31.25	-27.28	-47.04	-51.29	-52.01
Counter Ions	S30L-23 CI	-63.11	-31.84	-35.78	16.52	-36.27	-41.32	-39.65	-56.54	-54.82
	S30L-24 CI	-74.16	-48.66	-41.8	-8.74	-38.44	-46.31	-39.65	-58.88	-57.04
	S30L-25 CI	-35.82	-2.89	-25.43	-18.64	-39.29	-3.98	-39.63	-31.93	-31.41
	S30L-26 CI	-35.62	-5.62	-26.41	0.32	-37.66	-4.01	-39.63	-31.79	-31.28
	S30L-27 CI	-34.63	-16.58	-30.32	0.50	-37.65	-14.54	-39.64	-23.9	-23.77

Interaction	Structure number	PW6B95	PM6		PM6-D3H4X		PM7		GFN2-xTB	
		BM-BE	Calc-BE	Pred-BE	Calc-BE	Pred-BE	Calc-BE	Pred-BE	Calc-BE	Pred-BE
Counter Ions	S30L-28 CI	-29.41	-14.16	-29.46	-17.9	-39.23	-14.82	-39.64	-21.17	-21.18
	S30L-29 CI	-47.18	-41.72	-39.31	-33.25	-40.55	-17.65	-39.64	-27.54	-27.23
	S30L-30 CI	-47.7	-44.40	-40.27	-23.56	-39.72	-16.97	-39.64	-32.88	-32.32

Table 4.7 Linear regression correlation parameters used to relate BM-BE (Y axis) values to Calc-BE (X axis) values for each semi-empirical method. The gradient (M) and intercept (C) values were used to generate the Pred-BE values shown in Table 4.3. RMSE calculated between BM-BE and Pred-BE for each method.

Y	M	C	R²	RMSE	MD	No. positive BE
Predicted-PM7	0.75	-26.45	0.85	9.08	-0.93	6
Predicted-PM6-D3H4X	0.42	-33.77	0.32	19.86	-1.63	3
Predicted-PM6	0.69	-24.79	0.65	13.51	0.00	8
Predicted-GFN2-xTB	0.94	-3.88	0.91	6.68	0.00	0

The RMSE values in Table 4.8 show GFN2-xTB to have the lowest error when considering all of the complexes in the dataset. In addition, this method also outperforms all the other semi-empirical methods when inspecting the different classes of NCIs present in the dataset. This includes outperforming PM6 for the prediction of the complexes within the dispersion class, in contrast to what was calculated with the S66 dataset. This highlights the need to investigate more than a single dataset when undertaking benchmarking calculations. The exceptions being for CH- π interactions, π - π stacking and CI (counter ion) interactions where PM7 and PM6-D3H4X show the lowest RMSE respectively. It is important to note that both PM7 and PM6 perform poorly when dealing with π - π stacking interactions -a type of interaction that is of particular importance in the aggregation of porphyrin systems, which are the focus of this work.

Interestingly, all the methods perform poorly when dealing with CI interactions. Excluding this type of interaction from the dataset shows a lowering of the RMSE values for all methods, with this improvement being most noticeable for GFN2-xTB (compare the first two rows in Table 4.8). This suggests that none of the semi-empirical methods are capable of dealing with this type of interaction and such complexes are to be examined with caution in this work.

Table 4.8 RMSE in kcal/mol, where N shows the number of complexes in each class of NCI. Colour-coded categorization indicates favourable RMSE in green, acceptable in yellow, and high in red as shown in a reference range on the left.

	PM7	PM6-D3H4X	PM6	GFN2-xTB
All complexes N = 37	9.08	19.86	13.51	6.68
All with no CIs N = 30	9.42	20.1	14.29	4.78
Non-polar dispersion N = 2	5.53	9.06	4.97	2.21
π-π stacking N = 10	11.95	7.63	11.64	7.77
CH-π interaction N = 2	1.33	23.12	2.38	3.06
Halogen bonds N = 2	10.24	23.94	40.12	7.40
Hydrogen bonds N = 7	7.67	19.43	10.02	3.35
Charged Counter Ions N = 7	8.85	30.52	10.79	2.06
Counter Ions N = 7	7.68	18.92	9.89	10.28

RMSE < 6.00
6.00 < RMSE < 12.00
12.00 < RMSE

4.4 Chapter Summary

This chapter aimed to establish the most suitable semi-empirical method for dealing with the NCIs present in conjugated organic supramolecular aggregates. These being: hydrogen bonding; π - π interactions and Van der Waal forces. The S66 and S30L datasets were used for this analysis as they provide BE reference values calculated using high-level computational chemistry methods (CCSD(T)/CBS and PW6B95-D3/def2-QZVP respectively).

Inspection of the RMSE results of the predicted BEs showed the following trends in the data:

- GFN2-xTB showed the best performance overall in terms of the lowest RMSE when taking both data sets into account.
- GFN2-xTB showed the lowest RMSE values for the majority of the different interaction classes present within both datasets. The exceptions being as below.
- PM6 outperformed all the other methods including GFN2-xTB, when dealing with the 24 dispersion dominated complexes within the S66 dataset.
- PM7 outperformed all the other methods, including GFN2-xTB, when dealing with the two complexes containing CH- π interactions. In addition, PM6 also outperformed GFN2-xTB for this type of interaction.
- PM6-D3H4X outperformed all the other methods, including GFN2-xTB, when dealing with the ten complexes containing π - π interactions.
- Finally, none of the semi-empirical methods utilised were capable of accurately dealing with the complexes containing CIs.

In conclusion, the results of this Chapter clearly demonstrate that, in terms of quantitatively predicting the trends on BEs taken from high-level calculations, GFN2-xTB outperforms all other semi-empirical methods. These results are further supported by the analysis in Chapter 5 where the performance of GFN2-xTB is validated further by comparing with the experimental data of a set of small organic conjugated aggregates.

Chapter 5

Computational Investigations into the Aggregation of One Dimensional Small Organic Conjugated Systems

5.1 Introduction

A number of small organic systems have been shown in literature to aggregate extensively due to their ability in forming π - π interactions through their extended aromatic cores, as well as the other NCIs when functionalised appropriately.^{7, 25, 161-163} Given this, it is possible to design their molecular structure with rigidity of the core and nature of the substituents, such as their ability for NCIs, and hence control a supramolecular spatial directionality when they self-assemble (Figure 5.1).

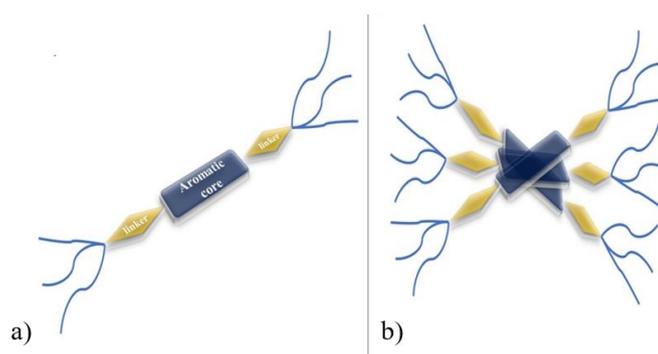


Figure 5.1 Schematic representation of a) a generic monomer building block and b) a spatially directed, helical in this example, trimer assembly.

Recent studies on the intrinsic interactions of short OPE (oligo(p-phenyleneethynylene)) aggregates have provided geometrical and energetic data suitable for computational examination.²⁴⁻²⁸ These systems can act as archetypal units for enabling an initial fundamental understanding of the interplay between the different types of NCIs and the resulting hierarchical self-assembly. This knowledge is likely to promote the rational design of a wider spectrum of supramolecular structures and functional materials. The literature background for OPE aggregation show a variety of experimental to computational analysis.^{24, 26, 28} This was carried out utilising their key structural components offered for self-assembly, which are the aromatic core and the versatile peripheral substitution. By exploration of selected oligomeric assemblies, key publications have utilised OPEs as useful linkages between experiment and theoretical prediction.^{24, 68, 164, 165} In a recent related study, the intrinsic aggregation mechanisms of an OPE were examined for the first time, by showing the preferable energetic pathways of the assembly of the supramolecular fragments.²⁶ The computed BEs (binding energies) of different sized fragments of aggregates were compared and consequently yielded a predicted mechanism of their assembly. Their findings were supported by a complementary experimental analysis to establish the mode of aggregation. This is some representative work for the examination of intrinsic mechanistic details of supramolecular aggregation. Other studies have collected the cooperative aggregation characteristics by evaluating cooperativity factor using combined computational and experimental methods.^{12, 165} These studies utilize small organic entities that are able to form strong intermolecular interactions. A wide collection of NCIs is achieved by OPEs because of their monomeric chemical structures.

Additional structures have been synthesised examining some other factors, such as the effect of metal,¹⁶⁶ electronic charge transfer,¹² and photoisomerization.¹⁶⁷ Molecules containing azo (-N=N-) groups, transition metals, or extended aromaticity pose excellent candidates for studying these effects on the supramolecular assembly mode for either supramolecular

architecture or the cooperativity.^{168, 169} Multiple combinations of structural combinations have been synthesised in an attempt to study these effects. The effect of metal on the aggregation is a recent topical subject of interest. It has been found to generally increase the factor of cooperativity and direct the spatial architecture with, specific to metal, interactions being formed. These findings have been recorded in recent publications showing the effect of metal on cooperativity and geometry.^{26, 71} Furthermore, charge transfer, being strongly associated with the presence of a metal in the assembly, is an important property in large conjugated supramolecular entities, yielding key applications in electronic devices and solar cell assemblies.¹⁷⁰ It has been additionally noted that a scarcely investigated process in supramolecular assemblies is the photoisomerization of supramolecular monomeric units. A photoisomerization can produce important changes in properties and aggregation mode and gives property control with photoinduction.

The assembly modes of supramolecular aggregates have been categorised as isodesmic, anti-cooperative and cooperative (Chapter 2.3). It is clear in literature that there are different approaches for quantifying the cooperativity factor from the experimental, as well as the computational perspective.^{24, 165} Within the experimental frame, polymers are measured with ITC experiments and their k (the reaction constant) values or $\Delta\Delta G$ (the difference in free energy of association) are applied for producing a cooperativity factor.^{68, 71} These values are extracted from experimental conditions with varying concentrations, solvation, and temperatures. On the other hand, computational data are most commonly tractable and consistently accurate in the gas phase, but rarely provide reliable thermochemistry data. The computed electronic energies are directly associated with the BEs with mathematical relationships as shown in the methods section (5.2.2). Importantly, concerning this Chapter's results, a range of studies have shown that computational methods can reproduce trends comparable to experimental trends, including structural details, insights into mechanisms of aggregation, and cooperativity.^{24, 165}

Aim

The aim of this Chapter is to apply computational predictions for a set of chosen 1D aggregates, which offer the structural and chemical parameters essential for molecular design investigation. These are carried out with a set of OPEs where the flip of the amide direction of topology affects the pre-organisation of the aggregation, and a diphenyl-azide based photoisomerizing entity (**L**₁, **C**₁) illustrating the photoisomerization, and the transition metal effect on aggregation. Additionally, it was further aimed when concluding the findings to further evaluate GFN2-xTB by comparing with the experimental data and the applied herein PMx.

5.2 Methodology

5.2.1 Computational Methodology

5.2.1.1 OPE Optimisations

The PM7 semi-empirical method was applied for geometry optimisation and frequency calculation of **OPE1** and **OPE2** aggregates (Figure 5.2), using the MOPAC2016 program. All calculations were applied in the gas phase with no restraints. Monomers to hexamers were optimised with this method. Frequency calculations yielded no imaginary eigenvalues, ensuring a minimum in the PES (potential energy surface). The input coordinates were composed using the Avogadro 1.02 visualisation software and set by estimated average literature values for all bond lengths and non-covalent distances. All data were processed, and plots produced with the use of MS Excel.

5.2.1.2 C₁ and L₁ Optimisations

For the diphenyl-azide photoisomerizing aggregates **C**₁ PM6-D3H4X was preferred, yielding better correlation in geometry and binding affinity as compared to experiment, as shown by preliminary test computations. **L**₁ dimers were computed at a B3LYP/6-31G(d) level of theory, with Gaussian09. Small imaginary frequencies were found in the frequency computations of

C₁ structures attributed to rotations of the hydrocarbon chains, and back corrected to obtain the optimisation energy. As with the OPE structures the input coordinates were composed using the Avogadro 1.02 visualisation software and set by estimated average literature values for all bond lengths and non-covalent distances.

5.2.1.3 GFN2-xTB Optimisations

It is noted that this work was carried out a year before the development of GFN2-xTB, hence, the then readily available, PM7 and PM6-D3H4X methods were mainly applied for this study, despite the conclusions of the preceding benchmark study in Chapter 4. The optimisations of aggregates **OPE1-2** were repeated with the GFN2-xTB Hamiltonian for a further comparison. Geometrical optimisations were computed for all structures in the gas phase, with no restraints. No frequency computations were carried out with GFN2-xTB for this part, as they were not tractable, or time efficient for the size of these structures. However, the energy convergence criteria were left as software default and were achieved by all optimisations. These criteria are for the energy convergence (Econv), the allowed change in the total energy at convergence 5×10^{-6} Eh, while the gradient convergence (Gconv) the allowed change in the gradient norm at convergence 1×10^{-3} Eh $\times \alpha^{-1}$.

5.2.1.4 Computed Spectroscopy and Energy Level Diagrams

The calculated absorption spectrum for monomers, dimers and trimers **OPE1** and **OPE2** were predicted via the ZINDO¹⁷¹ method by means of the ORCA software (version 2.9).^{171, 172} The ZINDO/S semi-empirical method was used to understand the electronic transitions for the complexes and the results provided visibility of the orbitals responsible for these transitions. This method is parameterized to generate the UV–Vis electronic transitions and cannot be used for geometry optimization values. The visualization of spectral results was performed with Gabedit¹⁷³ (spectra plots in Figure A5.2). Molecular orbitals were visualized using Jmol.¹⁷⁴ All

coordinates and graphical data of the optimised entities are contained in the A5_G5 folder of the Appendix.

The energy level diagrams were drawn with the combined use of MS PowerPoint, Ms Excel and GaussView for the plotting of FMOs. The excited states of the Pd^{II}-based isomers, both *cis* and *trans* were calculated via TD-DFT using the respective optimised ground state structures. The TD-DFT calculations were done with the Gaussian09 program¹⁷⁵ using the PBE0 functional¹⁷⁶ and 6-31G(d) basis set¹⁷⁷ for the light atoms, together with the LANL2DZ¹⁷⁸⁻¹⁸⁰ pseudopotential applied on Pd^{II}.

5.2.2 Calculation of the BE

The BE of all species was calculated with the following relationship.

$$BE = E(\text{complex}) - nE(\text{monomer}) \quad \text{Eq. 5.1}$$

Where E is the total electronic energy of the respective optimised structures, and n is the number of monomer units contained in the aggregate.

5.3 Results and Discussion

This chapter outlines the results of a computational study concerning the mechanisms and properties of two conjugates organic systems, using mainly the PM7 and PM6-D3H4X semi-empirical methods, and a complementary TD-DFT analysis. The first section focused on an organic OPE assemblies and the variation of NCIs strength and alignment as a function of the pre-organisation of the amide moiety (i.e. the flipping of its topological direction in structure). The second half of this Chapter's results studied a photo-isomerizing entity, with and without the presence of a transition metal, thus enabling the effect of a transition metal in the assembly to be assessed. These computational results are a part of two combined experimental-computational publications.¹² A short follow-up analysis investigated the OPE1 and OPE2 systems using the newer GFN2-xTB semi-empirical method for further method evaluation.

5.3.1 The Role of H-bonding and Pre-Organisation in the Self-Assembly of Aggregates

The dynamic assembly and directionality of OPE supramolecular structures is driven by the presence of hydrogen bonding and π - π interactions within these assemblies. Therefore, these systems provide key utility tool for understanding the ability of computational methods to aid in supramolecular synthesis and property prediction. With this consideration, it is interesting to investigate computationally the structural and energetic variations as a function of topological pre-organisation in the *mer* unit. This was achieved herein by pre-organising the directionality of the amide linker in the *mer* units of the OPEs as shown in Figure 5.2. This resulted in 1D-wire aggregates with differing geometries and properties. Herein, these species are discussed through a detailed computational analysis, linked to experimental findings (all discussed systems were synthesised and characterised using UV-Vis, NMR and IR spectroscopy by collaborating experimentalists Philips *et al.*⁷⁸ see A5-S5).

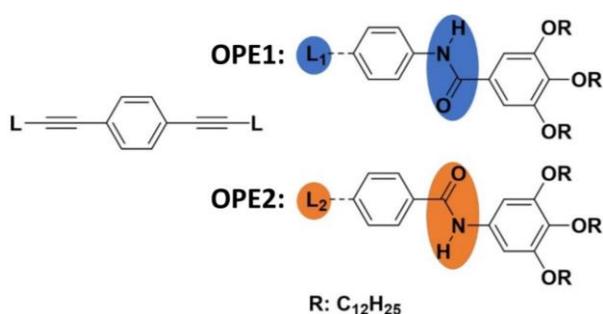


Figure 5.2 Chemical structures of OPE1 and OPE2, where the L₁ linker is refers to OPE1 and L₂ to OPE2, highlighted in blue and red respectively.

5.3.1.1 The Frontier Molecular Orbital Analysis of OPE

The inversion of the amide topology yielded significant differences in the electron density distribution as predicted within the ZINDO scheme (see section 5.2.1.4). The topology of the FMO (frontier molecular orbital) on the molecular structure shows that the LUMO (lowest unoccupied molecular orbital) is mainly located on the aromatic core in both **OPE1** and **OPE2** (Figure 5.3). In contrast, the HOMO (highest occupied molecular orbital) is shifted from the

central aromatic core on **OPE1** to the lateral phenyl rings in **OPE2**. In addition, the HOMO and LUMO orbitals are located largely on the same atoms in **OPE1** compared to **OPE2**. This HOMO-LUMO ‘overlap’ has been suggested to be important for electronic transition and accompanying properties, such as in spectroscopy and photoisomerization.^{163,181} The relevance of the calculated orbital overlap was confirmed by absorption and emission studies in chloroform, where the maxima of the UV-Vis spectrum for **OPE1** is 35% higher than **OPE2** (Figure A5.4a) and b)). The calculated results showed the HOMO (the charge donor orbital) and LUMO (the acceptor orbital) to be located on different atoms (Figure 5.3) implying a charge transfer character.¹⁸²

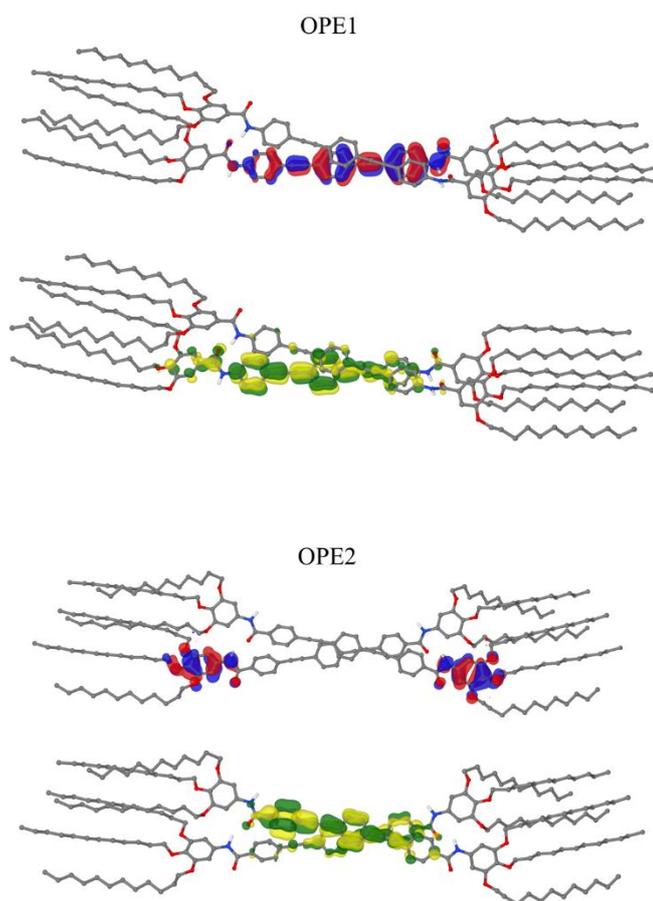


Figure 5.3 OPEs 1-2 FMO topology is shown, where the HOMO and LUMO orbitals are shown in red/blue and green/yellow respectively in each dimer, as computed with PM7.

5.3.1.2 Geometrical Analysis

The PM7 optimised geometries of both aggregates possess a helical alignment, with the amide moiety driving the direction of helicity. In addition, the main aromatic cores stack on top of each other, held together by π - π interactions. Finally, the optimised geometries also indicate that the long hydrocarbon side chains assemble via van der Waals forces (shown structures in Figure 5.4). The comparison of the the optimised geometries of **OPE1** and **OPE2** illustrated the effect of the direction of the amide in the linker. It was seen that **OPE2** had a large distortion in the planarity of the core (Figure 5.4). The optimised geometries also highlighted the core distortion results in a mismatch of the H-bonding alignment. These findings are supported by the results of variable temperature UV-Vis studies carried out on **OPE1** and **OPE2**. These studies showed a shift in the λ_{max} value for **OPE1**, but not **OPE2**, upon aggregation indicating an increasing π - π interaction. In addition, a bathochromic absorption peak and hypsochromic shoulder present in the spectra for **OPE1** also support the computationally predicted helical aggregate geometry driven by face-to-face and hydrogen bonding interactions (Figure A5.3).

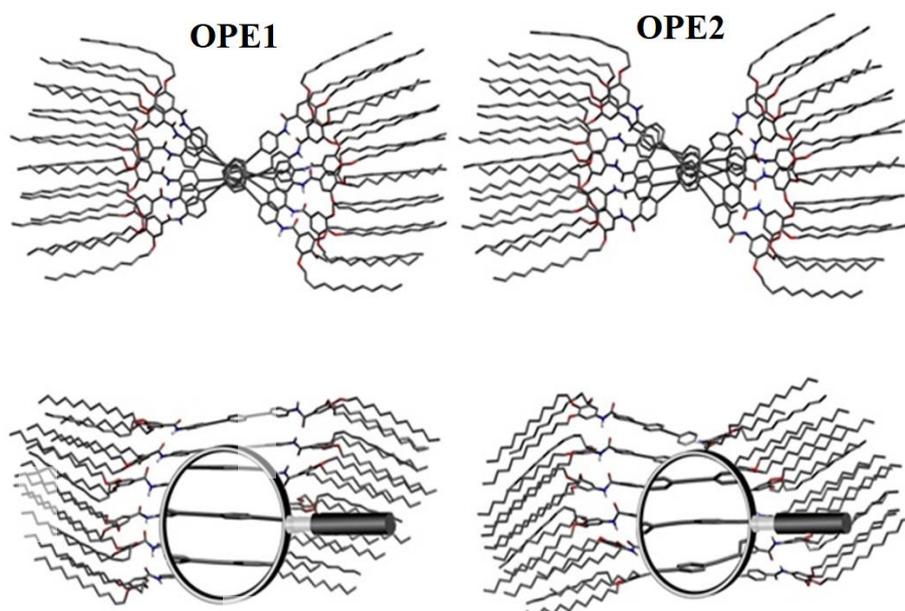


Figure 5.4 Face and side views of the OPE1 and OPE2 PM7 optimised hexamers, where the discrepancies in planarity are magnified.

5.3.1.3 Hydrogen Bonding Analysis of the Assembly

A key consideration when studying these self-assemblies is the strength of the H-bonding network within these systems. This is attributed to the synergistic contributions of the NCIs present (π - π interactions, vdW and H-bonding) that promote a cooperative aggregation mechanism. This is important, as OPE systems without H-bonding linkers do not cooperatively aggregate (e.g. the **OPEM** in Figure 5.5). The strength of H-bonding networks in **OPE1** and **OPE2** were investigated computationally by averaging the consecutive H-bond lengths and the associated BEs upon progressing from a dimer to a hexamer. In addition, the BEs associated with **OPEM** were also calculated for a comparison. The results of this analysis are shown in Table 5.1. These showed that **OPEs 1** and **2** aggregate with exothermic BEs suggesting a cooperative mechanism. As the number of *mer* units in the aggregate grows, it is shown that **OPE2** is becoming more exothermic than **OPE1**. This is supported by the experimentally obtained ΔH_f (enthalpy of formation) found to be -79.3 (OPE1) and -88.6 (OPE2) kcal/mol (Table A5.1). As a complementary quantification of the aggregation force, the length of hydrogen bonds is shown in Table 5.1 from the optimised geometries. This illustrates a progressive decrease in H-bond length as the aggregate grows, reflecting the synergistic effect of the combined intermolecular interactions in the assembly.

For contrast the BE values of **OPEM**, that does not possess H-bonding groups, are also shown in Table 5.1. The non-hydrogen bonded entity **OPEM** possesses a close similar core structure and side chains but without the specific H-bond interaction dominant in the **OPE1-2**. It has been found experimentally that this does not form aggregates.¹⁸³ These findings are confirmed by the PM7 computations, showing an increasingly exothermic trend for **OPEM**. The computed BEs are calculated to be significantly less exothermic than the two OPE systems, thus keeping in with the experimental results. This reinforces the importance of exploring a

hydrogen bonding network in supramolecular assemblies in addition to its pre-organisation in the molecular structure.

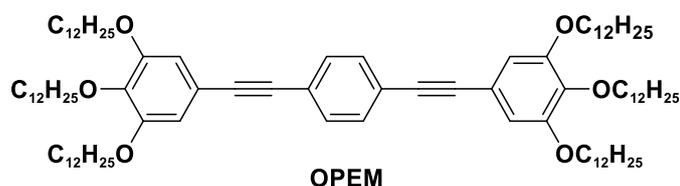


Figure 5.5 The chemical structure of OPEM.

Table 5.1 The average H-bond lengths in Å where n is the number of *mer* units, and the BE are in kcal/mol, as computed with PM7.

	OPE1		OPE2		OPEM
n	H-bond length	BE	H-bond length	BE	BE
2	1.94	-8.16	2.20	-4.83	-1.54
3	1.96	-16.51	2.12	-12.95	-2.46
4	1.93	-20.61	2.06	-19.87	-2.77
5	1.92	-26.29	2.02	-29.61	1.68
6	1.92	-27.53	1.98	-34.89	5.42

5.3.1.4 The Mechanism of Aggregation

Semi-empirical calculations enabled further insight into the mechanism of assembly for **OPE1** and **OPE2**. These calculations involved a series of PM7 single point computations starting from a “V” shaped dimer that has been hypothesised as the aggregation seed structure (Figure 5.6).⁷⁸ Geometry optimisation of the seed structure resulted in the dimer being brought together driven by hydrogen bonding and π - π interactions. An animation of this process illustrates the drive of aggregation, in which the linkers are brought together like snipping scissors (animations A5_G5a and A5_G5b for OPE1 and OPE2 respectively). These computational

findings are supported by the results of VT-NMR experiments showing that, upon cooling, the dimer rearranged into a more compact parallel arrangement driven by the H-bonding and π - π interactions.⁷⁸

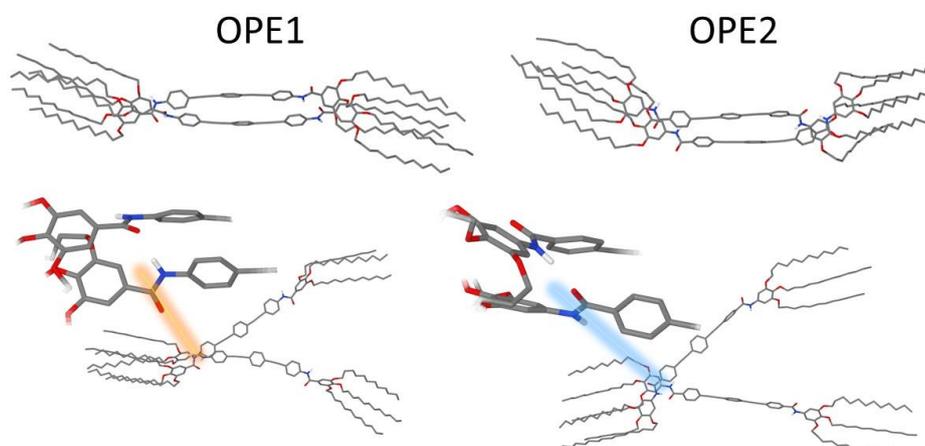


Figure 5.6 The geometries of the proposed pre-nuclei (below) and their dimers (above).

5.3.1.5 The Performance of GFN2-xTB on Evaluating Aggregates Driven by NCIs

An additional analysis was performed upon **OPE1** and **OPE2** to investigate the predictivity of the recently developed GFN2-xTB semi-empirical Hamiltonian, which was not available when this OPE study was published. This was undertaken by comparing aggregate geometries calculated using GFN2-xTB with those calculated using PM7, PM6-D3H4X and, as a control, the non-dispersion corrected PM6 method. The geometries produced by the four methods showed PM6-D3H4X yielded T-shaped π - π alignment (Figure 5.7). In contrast, GFN2-xTB reproduced the previous geometries calculated using the PM7 semi-empirical method with a parallel core alignment. While PM6 produced scattered geometries up to trimer indicating further the importance of the dispersion correction. This is an important finding, that further validates the predictivity of GFN2-xTB. As an additional consideration, the GFN2-xTB computations were carried out approximately 10-fold faster than the PMx methods for these structures.

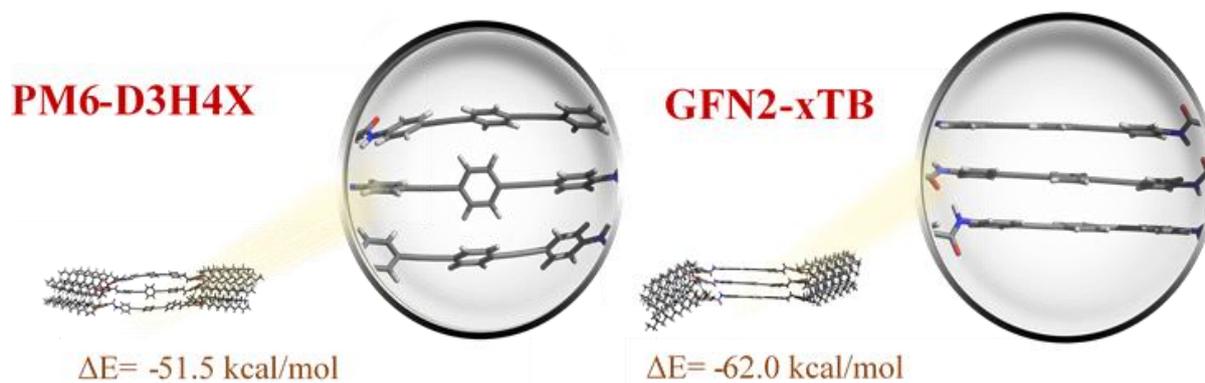


Figure 5.7 The geometry of OPE1 trimer optimised with PM6-D3H4X on the left and GFN2-xTB on the right.

5.3.2 Influence of Metal Coordination and Light Irradiation on Hierarchical Self-Assembly Processes

Herein, an azobenzene based photoisomerizing entity was investigated for the effect of metal complexation on hierarchical aggregation. The featured structure entails the azobenzene unit, an amide group, which guides the H-bond network, and a metal ion responsive pyridyl ligand and dodecyloxy side chains (**C₁**, Figure 5.8). The synthesis and analytical experiments were carried out by Kartha *et al.* A structured computational study defining the geometrical factors and energetic components driving the assembly mechanisms is outlined.

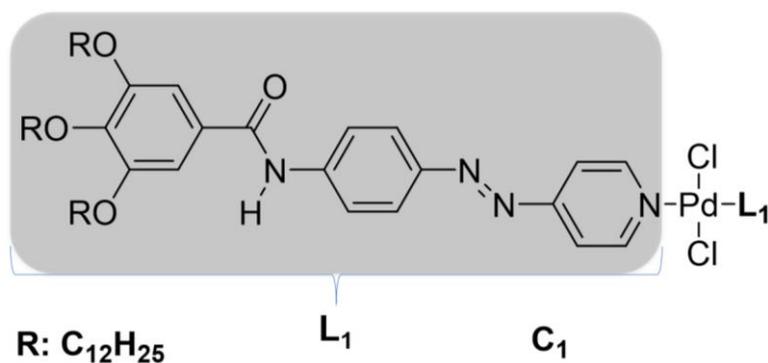
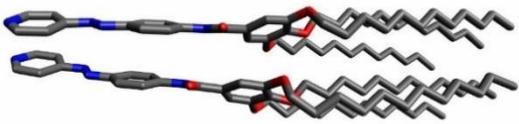


Figure 5.8 The chemical structure of monomer **C₁** and highlighted in grey the ligand **L₁** reviewed as an independent molecule in sections of this work.

5.3.2.1 The Effect of a Metal Ion upon Self-Assembly

The target ligand **L**₁, which compiles the organic part of the monomer (the azobenzene unit, the amide group and the dodecyloxy chains) is assessed as an independent molecular entity computationally for its mode of supramolecular assembly. The monomer and dimers of **L**₁ were computed at a B3LYP/6-31G(d) level of theory. The potential stacking parallel and anti-parallel are shown in Table 5.2. The BEs give a relatively more stable assembly for the anti-parallel stack showing a stronger exothermic BE. In both conformations the amide parts stack antiparallel to each other to create hydrogen bonding interactions. While in the parallel stacked dimer the side chains align to form vdW interactions.

Table 5.2 The optimised **L**₁ dimers at a B3LYP/6-31G(d) level of theory at a parallel and anti-parallel conformation and their BEs in kcal/mol.

L ₁ dimer	BE
 Parallel conformation	-65.7
 Anti-parallel conformation	-66.4

The incorporation of a metal, Pd^{II}, in this aggregation, with the photoresponsive aggregate **C**₁ (Figure 5.8) produces a pronounced difference in the aggregation mechanism. Semi-empirical optimisations at a PM6-D3H4X level show a slipped stacking (Figure 5.9), where the coordinated chlorine closes to interact with the adjacent polar amide hydrogen. These chlorines are also interacting with the neighbouring aromatic protons and the polarized methyl groups of the alkyl chains. The alkyl chains interact with strong vdW interactions, and the core aromatic

rings remain on the close proximity held with π - π interactions. These cooperative forces direct the long-range assembly of the polymer. The cooperativity was modelled with the PM6-D3H4X producing the BEs for up to a tetramer, showing a strong exothermic effect as the aggregate grows (see Table 5.3).

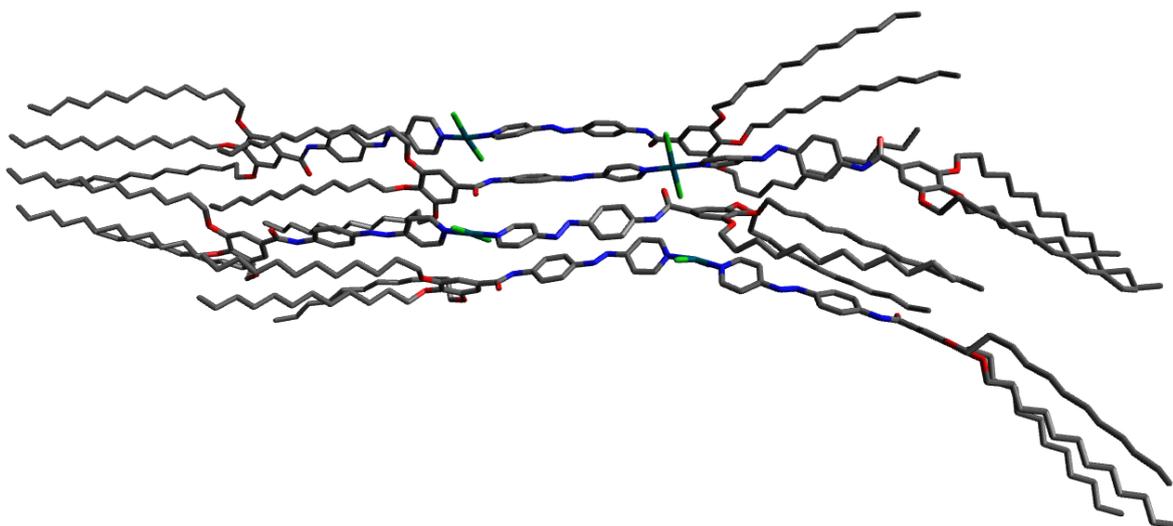


Figure 5.9 The optimised with PM6-D3H4X tetramer of C_1 showing the slipped structure of the monomers which accommodate the interaction between the polar hydrogens and the Pd (in turquoise) and chlorines (in green), here hydrogens are omitted for clarity.

Table 5.3 The BE in kcal/mol of aggregate C_1 up to a tetramer, where n is the number of *mer* units.

n	BE
2	-26.5
3	-70.9
4	-92.0

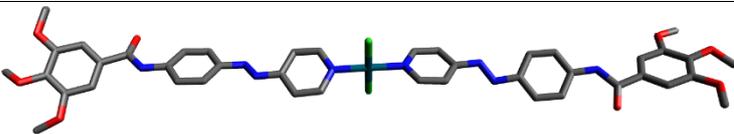
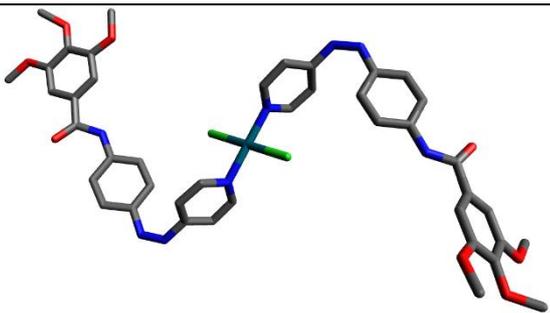
Unlike the L_1 aggregation, experimental analysis when cooling the species shows spontaneous aggregation of C_1 , indicating a stronger cooperative mechanism. The combination

of NMR, FTIR and X-ray analysis confirms the crystallographic conformation produced by the PM6-D3H4X optimisation.¹²

5.3.2.2 The Photoresponsive Aggregation of C₁

For the computational investigation of the photo-isomerization the C₁ was simplified by removing the side alkyl chains. This reduced the computational cost, as they are not practically tractable in TD-DFT, or accurate due to side chain rotations. The *cis* isomers of the two azo-groups on the structure showed the lateral edges of its structure shifted in opposite directions, thus stabilizing the conformation in an overall ‘sigma’ shape (Table 5.4).

Table 5.4 The photo-isomers of C₁ with removed alkyl chains, *trans* and *cis* and their dimerization BE in kcal/mol at a PBE1PBE/6-31G(d) level, hydrogens were omitted for clarity in the structure visualisations.

C ₁ photo-isomers	BE
	-66.9
<i>Trans</i>	
	-31.7
<i>Cis</i>	

Due to the size and complexity of the entity, a transition state optimisation could not be undertaken in this work, yet the essential information of the relative stability of the isomers was determined. This shows the most stable geometry as the *trans* isomer (Table 5.4).

Further TD-DFT optimisations of the *trans* and *cis* monomers produced the important MOs and excitation energies, which were plotted in energy level diagrams for an overview of the important transitions driving the photo-isomerization, (Figures 5.10 and 5.11). In the *trans* isomer the HOMO-LUMO is a MLCT (metal to ligand charge transfer) process, where the HOMO is almost entirely localised on the metal's d orbital, and the LUMO is spread on the lateral ligands. Other pronounced transitions take place as a MC (metal centred) and LC (ligand centred). This π molecular orbital is anti-bonding with respect to the N=N double bond. Light irradiation on the population of this orbital breaks that double bond, allowing the free rotation around an N-N single bond. In comparison, in the *cis*, HOMO is delocalised on the entirety of the molecule, and only on the ligands in the LUMO, thus producing an MLCT and LC effect, in the HOMO-LUMO transition.

The complementary experimental analysis from NMR and UV-Vis showed the kinetically dominant species to be the *trans* isomer, with the *cis* rapidly converting to *trans* after photo-irradiation. The energy level diagram for the *cis* isomer shows the HOMO-LUMO overlap to be more extended than in the *trans* isomer (Figure 5.10), rationalising the experimental observations.

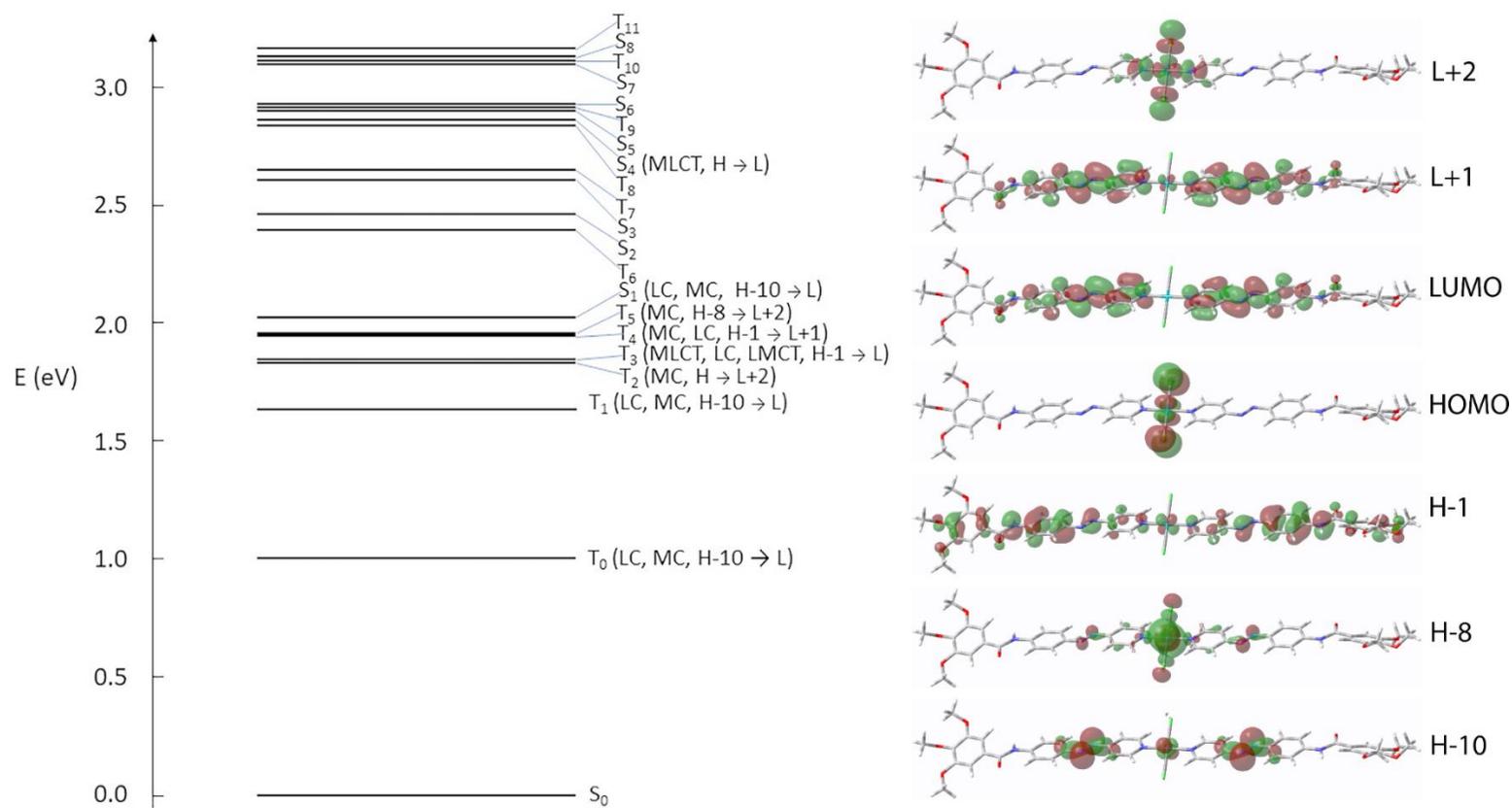


Figure 5.10 Triplet (T) and Singlet (S) excited states and corresponding molecular orbitals predicted for the *trans* isomer from TD-DFT calculations. The numbers beside the levels represent the molecular orbitals with the largest contribution to the state.

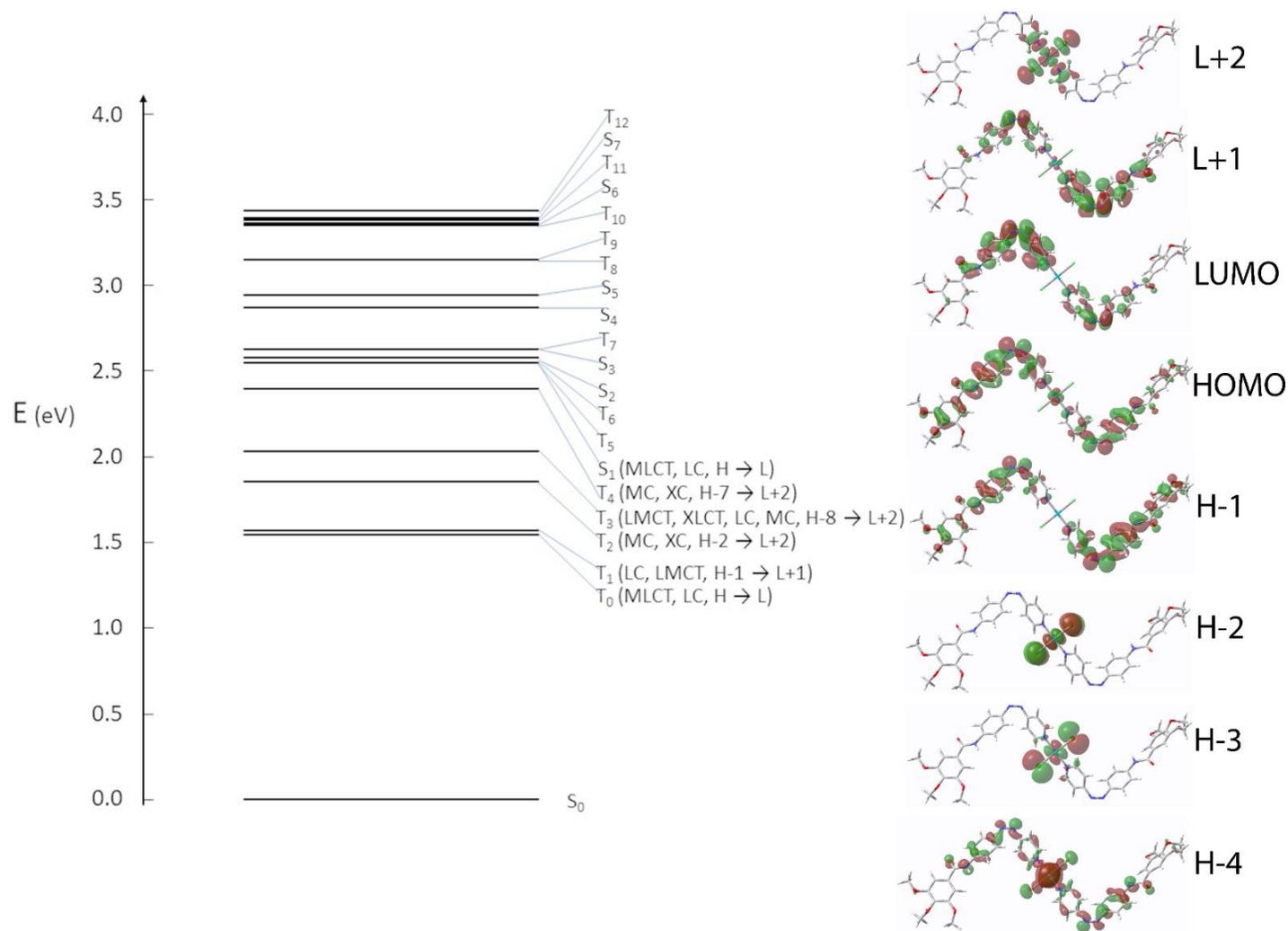


Figure 5.11 Triplet (T) and Singlet (S) excited states and corresponding molecular orbitals predicted for the *cis* isomer, from TD-DFT calculations.

The numbers beside the levels represent the molecular orbitals with the largest contribution to the state.

The collection of this analysis offers insights into the fine mechanistic aspects of the aggregation of this light responsive entity. The thermodynamic and kinetic stability of each isomer combined with the characteristic geometrical aspects led to assembly pathways that were favourable for an organised polymerisation of the *trans* isomer, or unfavourable, as it was for the *cis* isomer. The incorporation of a transition metal as Pd^{II} further quantified the effect on the cooperativity of the aggregation force and provided new electronic effects for the charge transfer processes during the photoisomerization.

5.4 Chapter Summary

This chapter was dedicated to combine a wide collection of the factors affecting aggregation mechanisms within 1D supramolecular assemblies. The following factors were visited through the sections:

- The structural interplay (amide flip) in aggregation driven by H-bonding and π - π interactions.
- The photoisomerization in aggregation driven by H-bonding and π - π interactions.
- The effect of transition metal in the monomeric structure for driving aggregation and electronic processes in supramolecular assemblies.

In the first section it was shown through a set of OPE aggregates that fine structural alteration in the molecule, such as in this case flipping the connectivity of the amide driving the H-bonding network, yields large changes in intrinsic and extrinsic properties. This interplay gave distorted aromatic interactions in the central core, and a crucial H-bond mismatch. These effects were quantified as significant changes in aggregation force and cooperativity, as well as electronic properties, showing important differences in the FMOs. A complementary GFN2-xTB comparison with the older methods showed encouraging validation of the previous benchmark study.

In the second section 5.3.2.2 the *cis-trans* photoisomerization process suggested a complex correlation between kinetic stability of the monomer building block and sterically unfavourable aggregation pathways of the resulting geometry. The experimental data validated that the combination of the latter two effects made the *cis* isomer unfavourable for organised aggregation in **L**₁, and kinetically and sterically unstable in **C**₁.

Through this examination it became apparent that a deep consideration must be made concerning the design of the *mer* unit when developing supramolecular functional entities. Fine changes in structural components and electronic aspects yield profound variations in bulk properties and the potential success of the aggregation. It is additionally noted that computational predictivity with semi-empirical methods (PMx and GFN2-xTB), and DFT for electronic effects analysis, is reliable and recommended for cost effective predictions.

Chapter 6

Examinations of the Aggregation Motif of 1D Porphyrinic Wires.

6.1 Introduction

Porphyrin aggregations in 1D (one dimension) have a directionality driven by their intermolecular interactions.¹⁸⁴⁻¹⁸⁶ The set up and combinations of these interactions can lead to various macromolecular shapes, such as a helix,¹⁸⁷ or a range of angles on the wire-core axis. Specifically, this latter arrangement of the angle of the wire axis produces the most notably J and H aggregates.¹⁸⁸ Where J aggregates refer to the monomer units arranging in a ‘slipped’ fashion, thus an angle axis, and H where the monomers stack face-to-face (Figure 6.1).

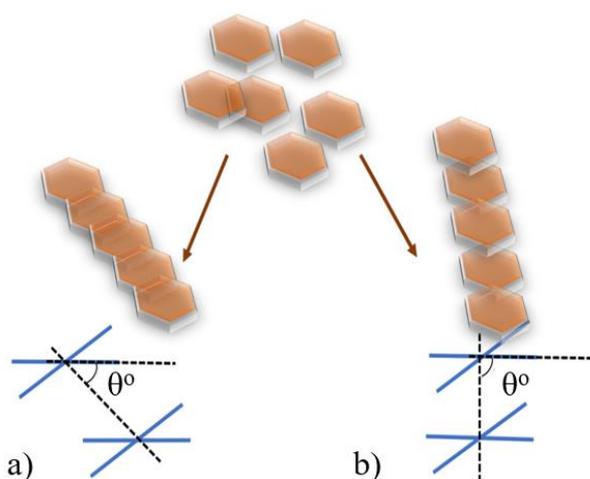


Figure 6.1 Sketch of a) a slipped angle J and b) a face-to-face arrangement, H, where the red tiles represent *mer* units. The porphyrin plane in a dimer is illustrated by the two blue lines in the lower illustration, with the axis angle (θ°) as shown.

6.1.2 The Importance of J and H Aggregation

These J and H modes of aggregation produce polymers with distinctly different spectroscopic or bulk material properties.^{44, 189-192} An important example is the natural antenna systems produced by J aggregations, this is effective because of a strong coupling of the transition dipole moments, and a consequent exciton delocalization phenomenon.¹⁹³⁻¹⁹⁶ A study found that J aggregates of porphyrins promote a significantly increased electro-optic response of excitons in antennas.¹⁹⁷ While another by Parkash *et al.* showed the enhanced resonance of light scattering in the J aggregates of electronically interacting chromophores, such as the porphyrins.¹⁹⁶ This effect was similarly published specifically for J-type aggregations of H₄-TPPS²⁻ (tetrakis(4-sulfonatophenyl)porphyrin diacid), showing the importance of this aggregation motif for these processes.¹⁹⁵ Furthermore, in oxygen sensing, J aggregates are also more effective in turning a singlet to triplet oxygen, as their ‘slipped’ arrangement allows better interaction with the circulating oxygen. This set up was investigated recently with an exotic bismuth porphyrin aggregate, in a ladder (or pseudo-J) arrangement, which was proven to allow the production of singlet oxygen in a much higher quantum yield.³⁴ On this subject, a more topical study specifically illustrates the enhanced capacity of J-type aggregations of porphyrins for a high singlet oxygen yield. Their MOF (metal organic framework) type assemblies of zirconium porphyrins gave evidence of the structure to property relationship from J aggregation to singlet oxygen quantum yield.¹⁹⁸

6.1.3 Controlling J versus H

Porphyrins as supramolecular entities have become the focal point of myriads of studies, because they are able to form multiple NCIs with each other, and additionally because of their versatility in being readily functionalised for controlling those interactions. This control is achieved through introducing peripheral groups,^{5, 6, 52, 199-202} complexation,^{6, 62, 203} or the choice of solvent or other experimental conditions.^{63, 204} A representative publication collecting the

effects of structure to J vs H aggregation by Villari *et al.* highlights the pronounced effect of small structural changes at a molecular level leading to changes in hydrogen bond motifs and by extension the aggregation motif.¹⁸⁸ An additional review by Che *et al.* outlines a survey of the structural alteration concerning the functionalisation of metalloporphyrins via saturated C-H bonds. The review of various mechanistic studies and previous work on functionalisation of metalloporphyrins offered critical conclusions on the importance of structure-property relationship for controlling the architecture, and the J/H type of assembly.²⁰⁵

The seeding work for exploring the intrinsic interactions defining the J and H assembly was made by Kasha's exciton coupling theory showing J aggregates as a red shifted extinction band and H a blue shifted extinction band.^{37, 65} According to this exciton coupling theory a threshold of distinction between a J and H aggregation is suggested as a $\theta^{\circ} < 55.5$ for a J and $\theta^{\circ} > 55.5$ for an H,¹⁸⁸ where θ is the angle drawn between the central points of the *mer* units and the plane of one of the parallel stacked units (Figure 6.1).

Furthermore, Ogi *et al.* compiled three zinc-porphyrin homologue *mer* units, which possess two groups *trans* to each other with long hydrocarbon chains, hydrogen bonded amide groups and a phenoxy group binding them to the main porphyrin (Figure 6.3). The other two *trans* groups then range in size, from hydrogen substituted, to two larger substituents, (**1S**, **1M** and **1L**, in Figure 6.3).⁶⁰ The study revealed the thermodynamic and kinetic control leading to either J or H aggregation. The thermodynamic versus kinetic control was monitored in their study with varied temperature and concentration experiments. The key role of the steric hindrance of the *trans* substituents was recorded. Therefore, importantly for the work within this Chapter, this study outlined that the thermodynamically favoured aggregate for **1S** and **1M** was H-motif, whilst for the most sterically hindered substituent, **1L**, it was a J-motif.

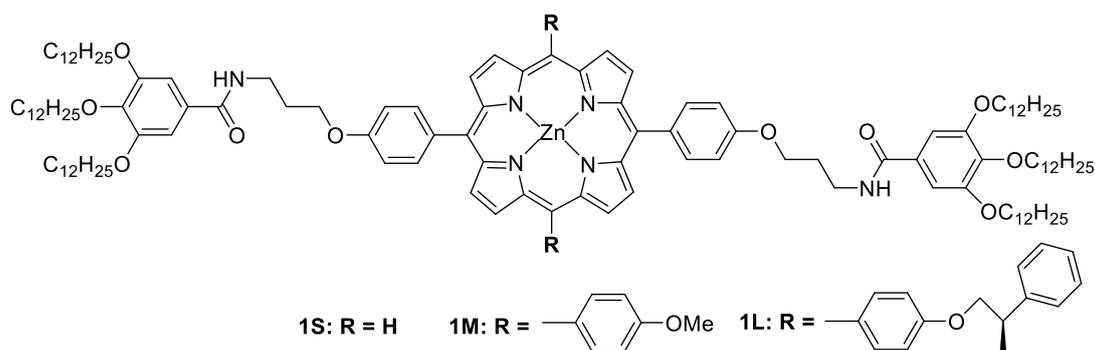


Figure 6.2 The chemical structures of the zinc porphyrin *mer* unit, whole as from the reference study above from ref 60.

A second relevant study reflects the structural interplay of the solubilizing side chains of a tetra-aryl-amidated zinc porphyrin for thermodynamic and kinetic control of J and H aggregation.² The thermodynamic products of the five homologues (Figure 6.3) are stated in their findings. These are offered as model structures for computational investigation on quantifying this structural control.

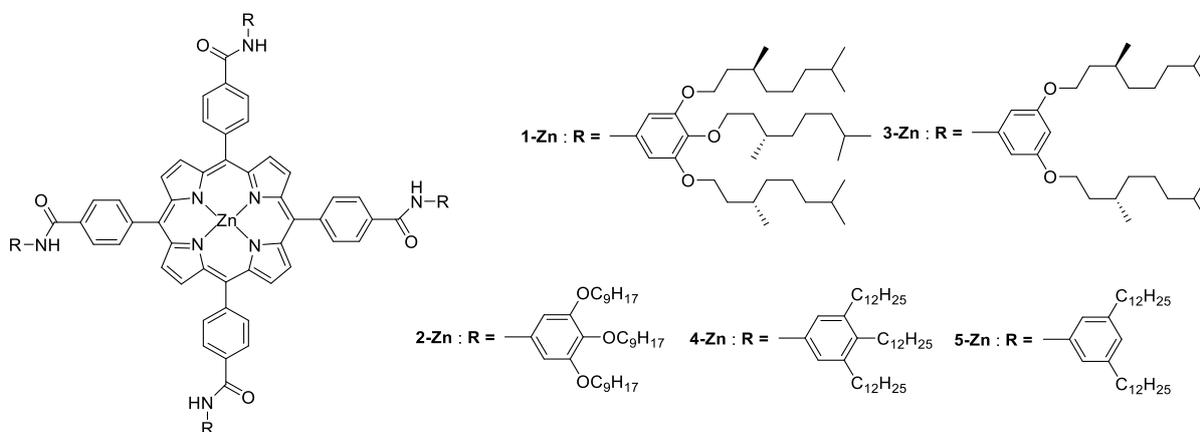


Figure 6.3 The chemical structures of the tetra-amide zinc porphyrin, with the five experimentally studied structures shown from ref 2.

Lastly, a non-metallated porphyrin entity again tetra-aryl-amidated substituted symmetrically, illustrates the effect of the structural design of the side chains by varying their position and length (Figure 6.5).⁶⁴ The conclusions showed the importance of the hydrogen

bonding in the structures to target the H and J aggregation design. Although, a broader discussion is made concerning the solvent effects and some kinetic control, the thermodynamic products are presented, for consideration of further computational evaluation.

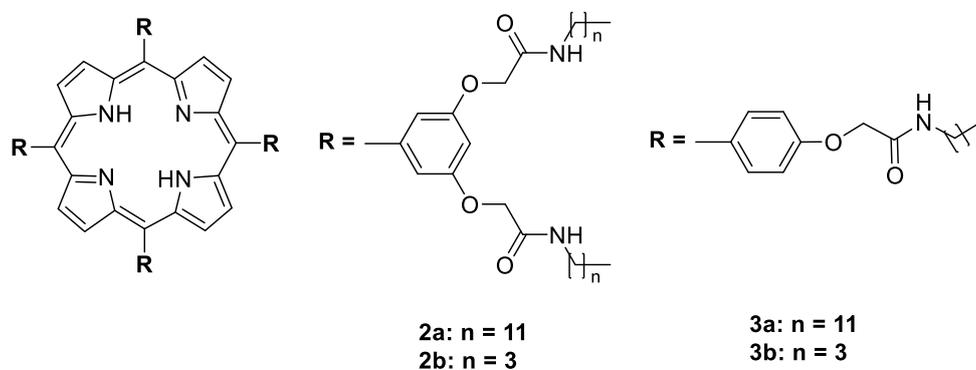


Figure 6.4 The chemical structures of the tetra-substituted porphyrin, with the four experimentally studied structures, from ref 64.

The sum of these sets of porphyrins entails ranging the number of peripheral chains, the existence of oxygen or not in the substituent (ref 64 set), the position of the peripheral group and inclusion of an amide group for hydrogen bonding. The study of this combinations of structural effects can reveal an abundance of information for chemical design of supramolecular aggregates.

Aim

The aim of this chapter is to investigate the ability of GFN2-xTB to correctly predict J vs H type of aggregation, in which J or H formation has been shown to be controlled by side chain effects such as steric hindrance and substitution patterns. This was achieved by investigating the ability of GFN2-xTB to correctly predict the structures of aggregates identified within the three key papers outlined above. In addition, the importance of being able to fragment these larger systems into key sub-sections was also investigated in terms of being able to correctly predict J vs H.

6.2 Methodology

6.2.1 Computational Methods

All coordinates of the structures were formed with the use of Avogadro 1.02 visualisation software.²⁰⁶ The coordinates of the aggregates were set to reflect the reference chemical structures, or fragmented as appropriate where described, up to a tetrameric assembly, and were set as an initial ‘guess’ H aggregation angle in all cases. All optimisations were carried out with GFN2-xTB¹⁰⁵ with no constraints in the gas phase. Though a frequency calculation was not tractable to establish a minimum in the PES (potential energy surface) due to the large size of the entities; all energy criteria were met in the optimisation calculation. For GFN2-xTB these criteria are for the energy convergence (Econv), the allowed change in the total energy at convergence 5×10^{-6} Eh, while the gradient convergence (Gconv) the allowed change in the gradient norm at convergence 1×10^{-3} Eh $\times \text{\AA}^{-1}$.

A conformation analysis search was carried out with CREST¹⁵¹ in section 6.4, where the lowest energy conformer and notable conformers in the search were collected. All data were processed with MS Excel and their detailed analysis is in the A6 file. All coordinates and graphical data of optimised aggregates can be found in A6_G6 in the Appendix.

6.2.2 Measurement of the Angle θ

The threshold between a J and an H aggregation is described in the exciton coupling theory of Kasha.³⁷ The strength of head to tail (in J) interactions produce a varied spectroscopic effect as compared to a face-to-face (in H). The threshold angle that is formed by those planes is suggested at a value of 55.5° . To deduce that from the optimised computed structures, the monomer units were set as planes (each plane considered the planar porphyrin core). The axis of the stack is defined by two Zn-Zn atoms. The angle of the projection of the axis on the below porphyrin plane is the θ angle defining a J or H mode (Figure 6.5). It is shown that θ defines the degree of slippage between adjacent *mer* units in a supramolecular polymer stack.

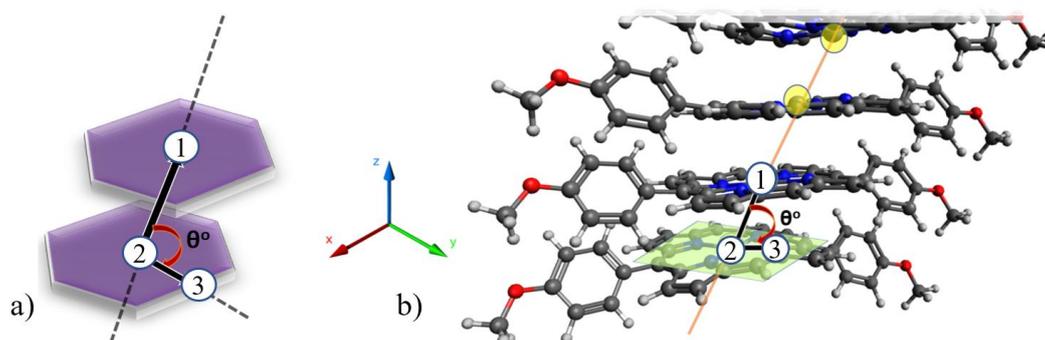


Figure 6.5 The J/H angle θ is shown where in a) the *mer* units are illustrated as purple tiles, the central atoms/points highlighted in white circles, numbered 1 and 2 and the reference third point-nitrogen atom, 3, that forms the angle θ . b) illustrates that in an optimised porphyrin stack.

The angle θ according to Figure 6.5 is then defined in the optimised coordinates as the angle formed between two successive central metal/dummy atoms and the succeeding connected nitrogen forming the narrowest angle gives as numbered $1 - \widehat{2} - 3 = \theta^\circ$.

6.3 Results and Discussion

Herein, the optimised aggregates from three selected publications (by Ogi *et al.*,⁶⁰ van Weegan *et al.*² and Shirakawa *et al.*⁶⁴), illustrating the variation of structural elements as a function of J vs H mode are presented and discussed. The predictive calculations focussed on investigating the ability of GFN2-xTB to correctly predict the experimentally observed J or H structure for the systems in the publications. The analysis also investigated the importance of including the long alkyl side chains, present in these systems, in the calculations by investigating a series of related fragmented systems in which these long substituents were removed. The aim here being to establish the most computationally efficient system size for dealing with these large systems. All calculations used the θ reference threshold angle of 55.5° suggested by the Kasha theory to define an aggregate as a J or H.³⁷

6.3.1 A Study of Fragmenting the *Mer* Unit for Computational Efficiency

The initial set of calculations focused on establishing the importance of including the large alkyl chains within these types of systems. Therefore, a fragment system was developed in which the long alkyl chains were replaced by methyl groups resulting in the model systems shown in Figure 6.6. Comparing the results between the full and fragment system showed the mean θ values to be almost identical (Table 6.1). These results showed that, in terms of predicting such aggregates for a J or H motif, there is no need to include these long alkyl chains in calculations. This is an important finding, as it enables larger aggregates to be investigated, a factor that was found to affect whether the system was predicted to be J or H. This effect was seen upon examining the range of θ values from dimer aggregates of the same set of fragments (1S, 1M, 1L) (Table 6.2). These calculations showed the dimer structures to be significantly distorted compared to the larger tetramer system (Figure 6.7 and A6).

The data in Table 6.1 show the predicted and experimental aggregate geometries (in terms of J versus H character). In both cases the data shown refer to the thermodynamically favoured isomer.

Table 6.1 H and J categorisation of the tetrameric (n=4) aggregates according to the computational θ estimation and the experimental reference value for comparison in the far-right column.

Name	Full system		Fragment system		Full system
	Mean θ°	Motif	Mean θ°	Motif	Experimental
1S	56	H	57	H	H
1M	58	H	57	H	H
1L	58	H	58	H	J

Table 6.2 Comparison of the averaged θ angle (in $^\circ$), from an GFN2-xTB optimised tetramer and the measured (as in the methods section) from the dimer.

Name	Full system		Fragment system	
	Tetramer mean θ	Dimer θ	Tetramer mean θ	Dimer θ
1S	56	53	57	53
1M	58	51	57	54
1L	58	54	58	51

The results showed GFN2-xTB to predict all three systems to favour the H aggregate. This agrees with the experimental findings for the two smaller systems (**1S** and **1M**); however, for the larger **1L** system the results contrast with the experimental findings. Inspection of the structure for system **1L** suggests that the static nature of an energy minimisation may be the cause of this mis-prediction (Figure 6.8). These motifs are assigned clearly by a small margin to the literature threshold $\theta=55^\circ$ and the computational error in geometry is to be additionally considered for these conclusions. However, the small preference shown is thoroughly examined herein in order to reveal further details for consideration.

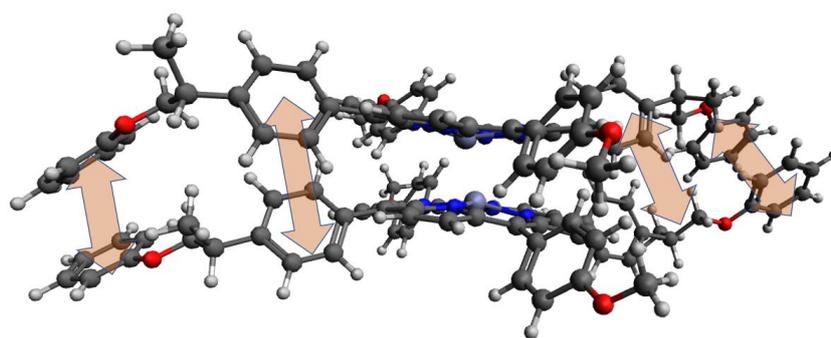


Figure 6.6 Optimised geometry of the dimer for system 1L showing the π - π interactions between the peripheral substituents.

The optimised geometry of the system shows the two aromatic rings of the **1L** substituent to be aligned to maximise the π - π interaction, resulting in the system being constrained to the H isomer. It is possible that within a MD (molecular dynamics) simulation (or CREST conformational analysis) this interaction would be ‘broken’ due to the rotation around the bond linking the phenyl rings of the **1L** peripheral substituents to the porphyrin. This could potentially lead to low energy conformers with J configurations due to the increased steric effect of the rotating phenyl rings.

6.3.2 The Effect of Steric Hindrance on the θ Angle

Given the mis-prediction of the **1L** system outlined above (in which R = aryl, Figure 6.6), further calculations were undertaken to investigate the ability of GFN2-xTB to predict the effect of steric hindrance alone on the type of aggregate formed. The hypothesis was being to decouple the steric effect from the π - π interaction present when R = aryl-type. The results of this analysis showed a reduction in the θ angle from 57° to 43° upon increased steric hindrance for non-arene substituted porphyrins, with **1A** R: a hydrogen, **1B**, R: Me, **1G**, R: isopropyl and **1D**, R: *t-butyl*. This predicted the H to J barrier being crossed when R = Me or larger (Table 6.3). These results indicate that, when dealing with steric effects in isolation, GFN2-xTB can correctly predict whether an aggregate will favour an H or J type formation. In addition, these results also support the previously detailed hypothesis that a simple energy minimisation may not be sufficient to correctly predict systems in which steric hindrance cannot be decoupled from other interactions.

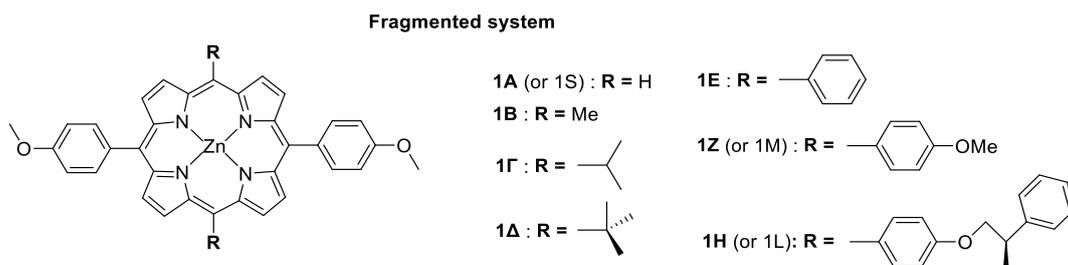


Figure 6.7 The chemical structures of the zinc porphyrin *mer* unit fragmented with a range of additional sterically ranged substituents 1A to 1H.

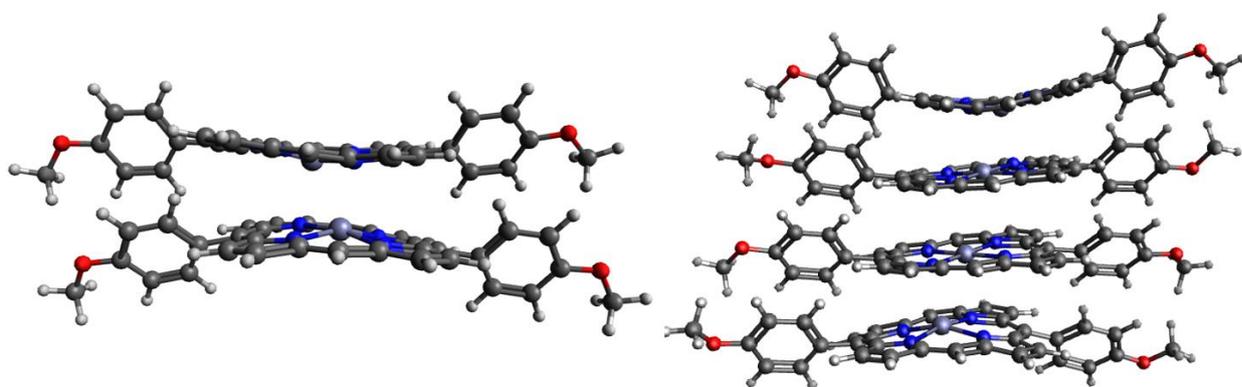


Figure 6.8 Optimised geometries of system 1S-fragment as a dimer (left) and a tetramer (right) showing the distortion of the porphyrin system in the dimer. The plane of the porphyrin is more buckled in small dimers and flattens with a longer aggregate.

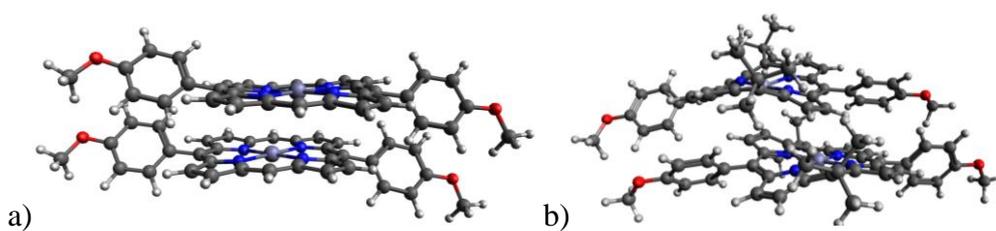


Figure 6.9 The optimised dimers of a) 1A a clear face face to face H motif and b) 1Γ a pronounced shift in θ angle as compared to 1A.

For these systems, it is likely that conformational analysis may be important. Example structures are as shown in Figure 6.9a) (R = H; H type aggregate) and Figure 6.9b) (R = *t*Bu; J type aggregate) with b) showing a pronounced intermolecular slippage as compare to a).

Table 6.3 The estimated mean θ angles (in $^{\circ}$) of the optimised fragmented aggregates and their assigned label as J or H on the right column.

Name	Mean θ	Assigned Motif
1A	57	H
1B	55	H
1 Γ	43	J
1 Δ	43	J

6.3.3 Predicting the Effect of Peripheral Substituents on Symmetrically Substituted Porphyrin Aggregates

Following the above findings, with consideration in developing an optimal computational methodology for studying those systems, the investigation was continued with two more published studies for J versus H aggregation. The symmetrically substituted tetra-amidated zinc porphyrin aggregations of van Weegen *et al.*² and the non-metalized amidated porphyrins by Shirakawa *et al.*⁶⁴ These studies focused on the effect of the side chains for controlling the J vs H mode in a thermodynamic manner. The original structures are listed in Figures 6.3 and 6.4. The fragment systems used in the computational investigations in this study are shown in Figures 6.10 and 6.11. Both of these systems are very similar in that they allow the effect of altering the peripheral chains upon the θ angle to be investigated. Importantly, both systems are symmetrically tetra-substituted with aryl rings and hydrogen bonding amide groups.

The key difference between the two systems being that one contains a coordinated zinc in the centre of the porphyrin with a single aryl ring as a peripheral substituent (Figure 6.10); while the second has no coordinated metal and two aryl rings in each peripheral group (Figure 6.11). As shown in Chapter 5, this interplay of substituent position can play a key role in the geometries and subsequent properties of these types of systems.

The structures were fragmented following the methodology outlined in section 6.3.1, in short, replacing long alkyl chains by methyl groups. For the zinc containing porphyrin system this produced monomeric structures **1A**, **1B**, **1Γ**, and **1Δ** as in Figure 6.6. The peripheral substitution for porphyrin **B-Zn** resembles that of porphyrin **A-Zn** without the para-substituent on the phenyl ring. Porphyrins **Γ-Zn** and **Δ-Zn** bear alkyl chains directly attached to the phenyl ring instead of an alkoxy linkage as in porphyrins **A-Zn** and **B-Zn**. Similarly, as in **A-Zn** and **B-Zn**, **Δ-Zn** lacks the para-methyl substitution on the methyl ring. For the non-metal porphyrin system, the selected fragmented systems resulted in two stacks for computational investigation named **2-T** (as the merged fragmented **2a** and **2b**) and **3-T** (as the merged fragmented **3a** and **3b**) as shown in Figure 6.11. Where **2-T** is doubly meta substituted with the methoxy, amide groups, and **3-T** only in the para position of the lateral aryl ring.

As listed in Table 6.4, the thermodynamic product of these systems have been experimentally determined to be H-type for aggregates **A-Zn** (**1-Zn**, **2-Zn** the equivalent full systems), **B-Zn** (**3-Zn**), and J-type for **Γ-Zn** (**4-Zn**) and **Δ-Zn** (**5-Zn**) for the first set of Zn-porphyrins. And H-type for **2-T** (**2a**, **2b**) and J for **3-T** (**3a**, **3b**). These findings reflect the changes in molecular design with the changes in substituent patterns on the aryl rings driving the preference for H or J-type aggregation.

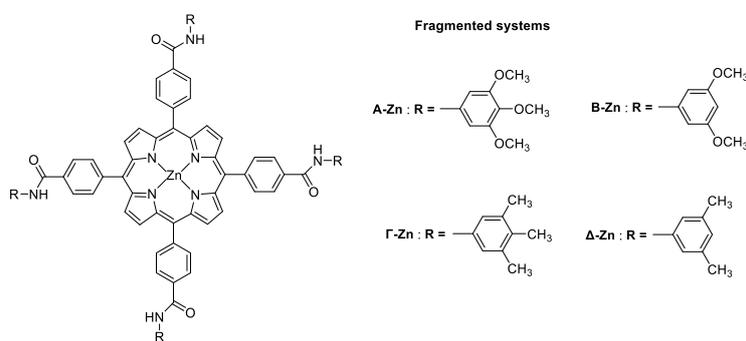


Figure 6.10 The chemical structures of the fragmented Zinc containing porphyrin system.²

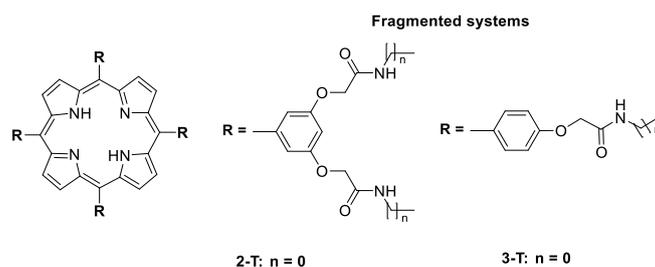


Figure 6.11 The chemical structures of the fragmented non-metal porphyrin system.⁶⁴

In contrast to these experimentally observed differences, the optimised geometries were all assigned as being of the H-type based on the θ angle being consistently greater than 55° (Table 6.4). Interestingly, the calculated results showed no clear trends in terms of J-type character, with experimentally determined J-type aggregates having the highest and lowest calculated θ values (systems **2-T** and **$\Gamma\text{-Zn}$** , **$\Delta\text{-Zn}$** respectively). Closer inspection of the trends illustrated the strength of the H-bonding interactions to affect the θ values. For example, **2-T** was calculated to have a more pronounced H angle at 70° than **3-T** being lower at 62° . System **2-T** possesses twice as many hydrogen bonding connections as **3-T**, with eight amide groups per *mer* unit versus four in **3-T** (Figure 6.12).

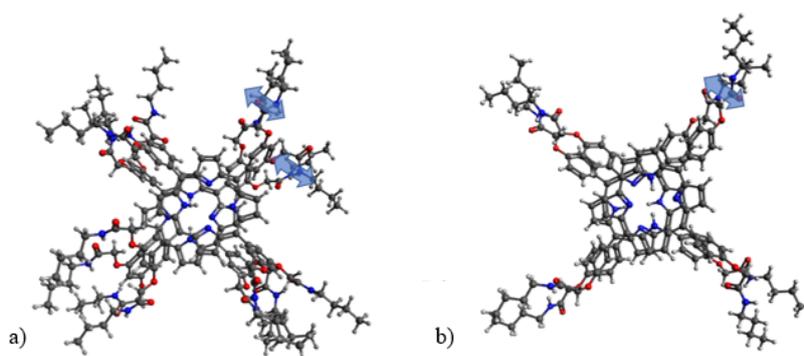


Figure 6.12 The optimised dimers of a) **2-T** and b) **3-T** showing with blue arrows the hydrogen bonding forces, in one of the four substituents in each as representative.

This consistency in the preference for the H-type motif produced in the optimised aggregates is reflective of the findings of the previous section concerning the presence of peripheral aryl

rings (section 6.3.1.2). The calculations upon these systems indicate that this effect is enhanced when the aryl rings are substituted with hydrogen bonding moieties. These findings further reinforce the hypothesis that in order to accurately predict H or J-type motif such systems containing aryl ring systems are likely to need conformation analysis or MD calculations in order to identify a range of low energy conformers.

Table 6.4 Calculated and experimental motif (J versus H) of all entities listed in Figures 6.10 and 6.1. Where θ (in $^{\circ}$) is the estimated from the GFN2-xTB optimised geometries.

Structure (experimental)	Structure (fragmented)	Mean θ	Calculated Motif	Experimental Motif
1-Zn	A-Zn	61	H	H
2-Zn	A-Zn	61	H	H
3-Zn	B-Zn	60	H	H
4-Zn	Γ -Zn	58	H	J
5-Zn	Δ -Zn	58	H	J
2a	2-T	70	H	J
2b	2-T	70	H	J
3a	3-T	62	H	H
3b	3-T	62	H	H

6.4 Reviewing the Data with Conformational Analysis

The results in this chapter have shown consistent trends in data in which the presence of π - π interactions in peripheral aryl rings yield a persistent preference for H-type aggregation (a larger θ angle). It is important to remember that computational optimisations are designed to locate the lowest-energy conformer on the potential energy surface. This static situation is not representative of a real system, which will be dynamic due to bond rotation (and vibration). In order to examine whether such movement would affect the preference these types of systems

display for J or H-type aggregation, a conformational analysis of the dimer of **1M** (Figure 6.13) was undertaken using the CREST software. The aim being to examine the number of H-type and J-type conformers present in the ensemble of structures produced.

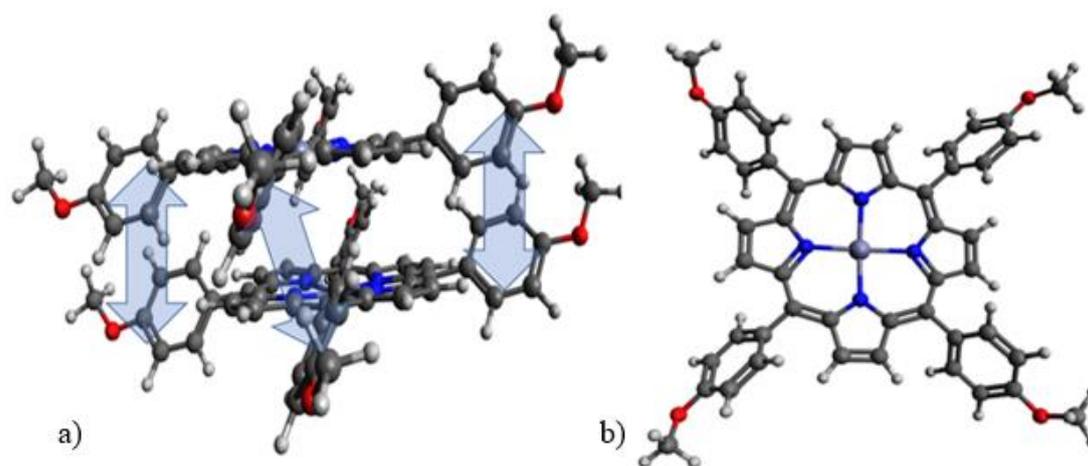


Figure 6.13 The optimised a) aryl substituted dimer, 1M, with the blue arrows showing the π - π forces within the aggregate and b) the monomer.

The range of θ angles for the 102 conformers from the CREST conformational analysis are shown in Figure 6.14. This analysis shows 72.6% of the conformers within the ensemble being a J-type aggregation, which is in contrast to the H-type motif predicted based on the single lowest energy conformer (section 6.3.1). The difference in θ is shown between the GFN2-xTB minimum and a “far-J” conformer in Figure 6.15. This provides the ensemble of conformers for the default energy window of 6 kcal/mol. This may reflect too large a window for the conformers corresponding to the experimental conditions. This conclusion can be reached by noting a majority of J conformers within this energy window, while an H conformer is experimentally recorded. Given this result, an analysis of the ensemble set of the CREST conformers for a 3 kcal/mol energy window was performed. This revealed a contrasting higher H-conformer population, in agreement with the experimental findings (66.7%) (Figure 6.14).⁶⁰ This adjustment of the energy window shows the importance of the choice of the ensemble of structures to correctly reflect the experimental data for porphyrin systems. This can be

attributed to the systems used to develop the CREST algorithm which are mainly small organic molecules, aromatic fragments and some metallated coordinated species.¹⁵¹ These do not reflect the aryl substituted porphyrins that are the focus of the work within this Chapter. However, the results of this analysis indicate that CREST calculations have the potential to correctly predict the J versus H aggregation preference for aryl substituted porphyrin systems. This is an important finding given GFN2-xTB failed to correctly predict the J versus H preference in these systems. Finally, it is worth noting that this work only studied a single system, thus additional systems need to be investigated to identify the optimal energy window that may be required for the correct prediction of J and H structures in porphyrin systems.

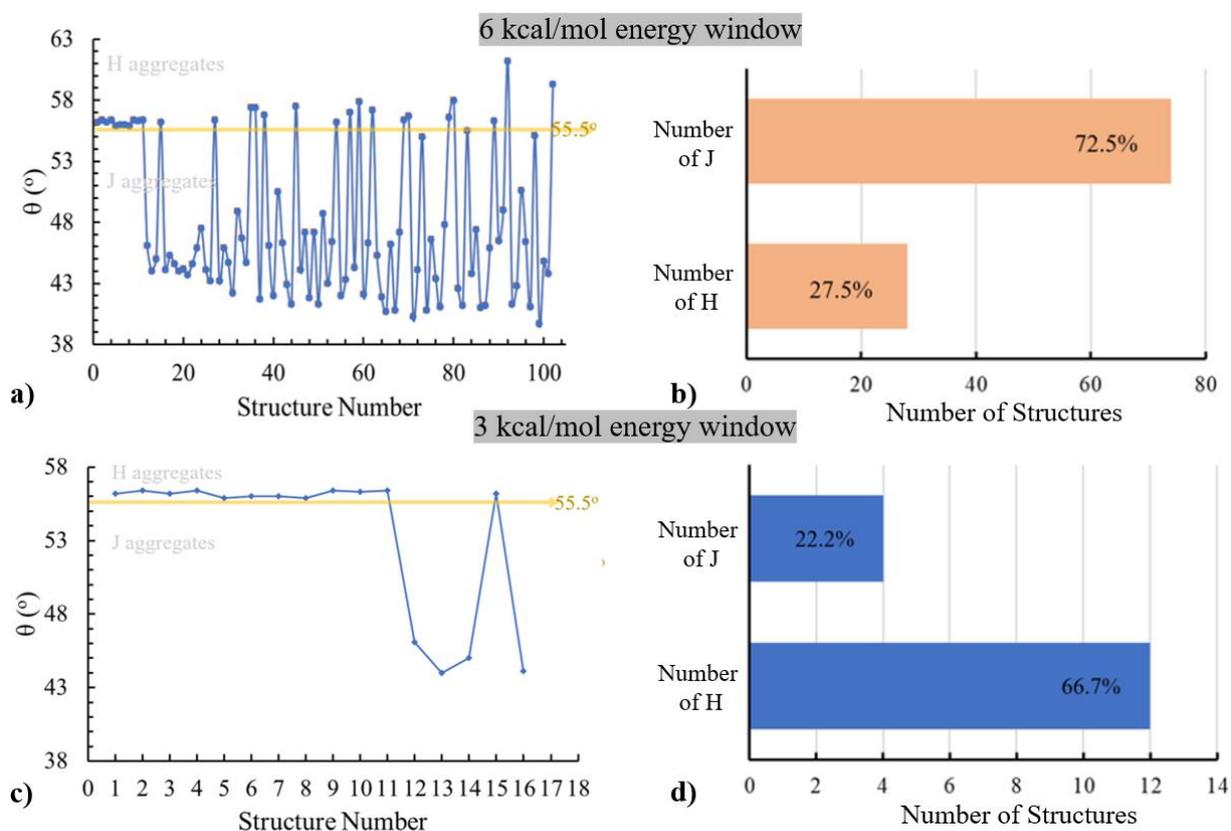


Figure 6.14 The θ angle for the 102 conformers of the CREST evaluation, where the yellow line indicated the threshold between a J (below the line) and H in a) and b), within the default 6 kcal/mol energy window of the CREST analysis, and the equivalent c) and d) for the adjusted 3 kcal/mol energy window for 16 structures.

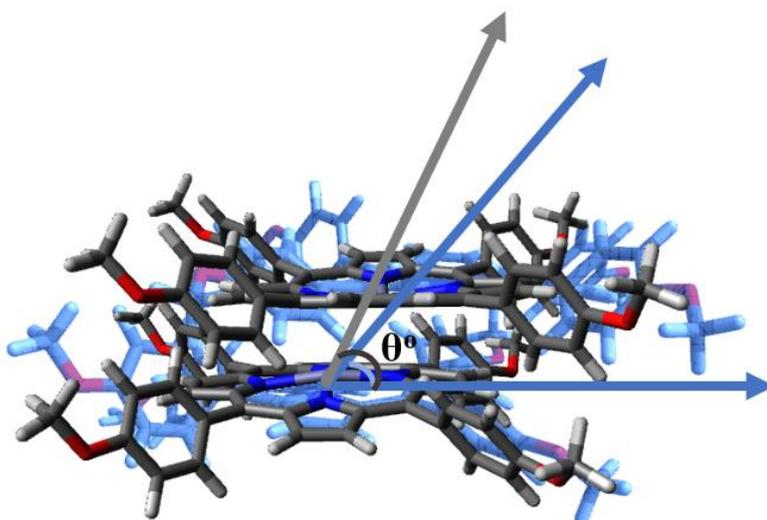


Figure 6.15 The blue highlighted structure is the one with the lowest θ (more J), and the foreground structure is the lowest energy optimised dimer, showing a more pronounced H (higher θ). The structures collected from the CREST analysis.

6.5 Chapter Summary

A J or H type of aggregation in 1D assemblies has distinct spectroscopic and physical properties and plays an important role in impactful applications in many fields. Specifically, in porphyrins those applications have been even more crucial (in the medical sector and oxygen sensing) and diverse (in wider electronic applications) making their determination and study attractive. In this chapter the capacity of the semi-empirical GFN2-xTB predictions was assessed for the J or H angle of axis θ° of those aggregations. The method to fragment the large originally studied systems proved successful as the θ angle was unaffected when comparing the full and fragment systems. The minimum essential size of aggregation was then established as the tetramer. This being since the dimers produced buckled pair geometries, with inconsistent θ values. Given these findings, fragment systems were used to investigate the ability of GFN2-xTB to predict the J or H preference for several porphyrin systems. The results showed this method to be predictive for systems except for those porphyrins substituted with arene functional groups. As reviewed by Schneider *et al.*⁹⁵ arene substituted porphyrins have shown a stronger binding and in addition the dispersion effects are highly dependent on solvation and thermochemistry

conditions which may be critical for analysing those systems. Solvent and thermochemistry calculations are not currently reasonably tractable for this size of aggregates, yet their approach can be considered in different computational strategies in the future. Further calculations using CREST, enabling conformer space to be sampled, confirmed this hypothesis. However, the results of showed the default energy window of 6 kcal/mol is too large to enable a reliable reproduction of experimental findings. Finally, the data presented in this Chapter suggests the size of the energy window may be key for producing a more accurate prediction of the J versus H preference, as exemplified by halving the energy window. It is likely that future work is needed on alternative porphyrinic aggregates which focus on this issue, to determine the correct energy window of CREST and to further reflect on experimental findings.

Chapter 7

An Examination of Substituent and Metal Effects on the Dimensionality of Aggregation

7.1 Introduction

The self-assembly of porphyrins in supramolecular systems is driven by a combination of intermolecular forces. Specifically, in porphyrinic entities the extended π - π network, commonly hydrogen bonding and metal ion coordination promotes supramolecular directionality. The various combinations of points of connections produce 1D, 2D or 3D networks of porphyrins (see Figure 7.1). As the dimensionality factor increases, the number and complexity of architectures becomes more abundant. As shown in Chapter 4 fine interplay in molecular design produces large changes in the properties of the supramolecular architecture and bulk material properties. It is important to review which forces drive the changes in dimensionality of the aggregation. A systematic computational assessment of appropriate divalent metals which coordinate in porphyrins and the electronic and steric effects of the side substituents herein opens the dialogue for a categorisation of those effects and the seeds of a future computational database for synthetic purposes. A carefully selected simple three-dimensional square aggregation from literature was chosen as a model for the evaluation of the metal and substituent effect on the dimensionality aspect of the assembly.

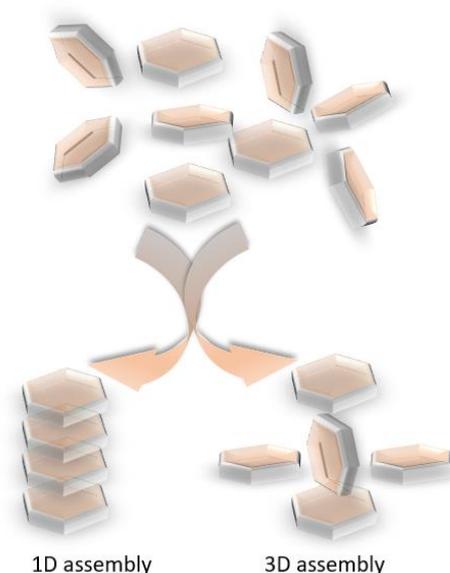


Figure 7.1 Sketch of a porphyrinic supramolecular assembly, where the red tiles represent a generic porphyrin ‘mer’ unit block are shown before organisation on the top and assembled to 1D (bottom left) and 3D-square (bottom right) as shown.

7.1.1 Dimensionality Control of Supramolecular Assemblies

The multiple architectures possible in supramolecular assemblies provide a very wide spectrum of properties from bulk material to electronic. Their synthesis aiming for optimisation of such properties has been the main focus of supramolecular research. Details of the individual interactions and the resulting modes of assembly have been presented by selected reviews, each framing a different aspect concerning the dimensionality.

A chapter by Krische and Lehn²⁰⁷ lists one to 3D motifs of hydrogen bonding networks in supramolecular assemblies. This gives some key starting reference points concerning which groups promote which dimensional motif, whether 1D-wire or 3D-square. The types of materials produced by the respective assemblies are presented in this work laying the base for molecular design in synthesis of supramolecular materials.

A recent review related to the aims of this work, in terms of creating a comprehensive database for a useful synthetic reference by Baburin *et al.*⁵⁶ presents the hydrogen bonding

impact on 3D assemblies of organic species. This work provides a very wide list of 3D architectures extracted from the Cambridge structural database, from single *mer* units to clusters of them. The parameters which particularly affect hydrogen bonding interactions, such as structural interplay geometry, as well as their topological effect on architecture and electronic and crystallographic properties are thoroughly outlined. This publication offers an insight into the effect different synthons may have in the assembly via hydrogen bonding, however not shown for porphyrinic systems. The association energies are also reviewed across these parameters. Though this collection provides a good reference from existing assemblies, it does not provide access information for new unreviewed synthons and combinations of interactions. This is something that can be tractably assembled via a targeted computationally produced database.

Addressing the effect of a metal upon assembly motifs, Saalfrank and Brent produced a starting study for biological supramolecular systems with applications in catalysis, optical, magnetic and molecular recognition.²⁰⁸ The dimensionality in assemblies as driven by the metal coordination was shown in a set of biologically related structures. It is underlined therein, that there is outstanding diversity in the architectural design, and properties, through metal interactions which provide synthetic routes for 3D assemblies.

Concerning exclusively porphyrin aggregations, the survey of one to three-dimensional assemblies of porphyrins on surfaces is made by Mohnani *et al.*³² There is an in depth description of the assembly mechanisms of porphyrinic type blocks in three dimensions. These make a representative collection of the NCIs holding the blocks together in different architectures. It was shown that these are mainly hydrogen bonding and π - π interactions, with it being specifically stated that for achieving the third dimension in the assembly the introduction of a central metal is frequently required.^{38, 209}

Finally, a representative experimental-computational study of a simple square-like 3D aggregation with chirality aspects was carried out by Oliveras *et al.*²¹⁰ This structure utilises the central metal for coordination and extension of the network in the 3D space. The peripheral substituents make an excellent example for highlighting the electronic and steric effects on the assembly motif. The fine substitution choice gives some optical chirality properties which give further insight to more complex architectures. These were shown in the referenced work to exhibit two orders of magnitude span in optical activity when intrinsic structural changes were made (such as the topology of hydrogen bonding groups and steric interactions yielding supramolecular chirality), therefore opening the way to new materials with interesting chromophoric applications.

Little is known concerning the choice of metal or substituent which leads to a specific type of aggregation of architecture. Therefore, it seems a logical first step to commence this study with the 1D-wire vs 3D-square control, featured herein as a wire and a square respectively, as a function of metal and side group, which is the aim of this chapter. The symmetrical square-like (Figure 7.2) assembly of the Oliveras group study can provide clear predictions through computations for the dimensionality aspect, without additional complications. The main queries of a synthetic nature for these assemblies are the electronic and steric properties of the peripheral groups, and the type of metal used. These are assessed in this work in a comprehensive, systematic study.

Aim

Herein, the aim is focused on establishing the preference toward a square or wire aggregation motif of supramolecular porphyrins through the variation of electronic and steric effects of the peripheral groups, and the central metal in a systematic way using the semi-empirical method, GFN2-xTB.

7.2 Methodology and Computational Methods

All coordinates of the structures were formed with the use of Avogadro 1.02 visualisation software.²⁰⁶ The coordinates of the aggregates were collected from the supporting information of the Oliveras publication.²¹⁰ The 1D-wire stack was designed in Avogadro 1.02 by estimation of appropriate intermolecular distances, which printed the cartesian coordinates. The pre-set input geometries of wire stack and square are optimised for binding energy comparison only. All optimisations were carried out with GFN2-xTB¹⁰⁵ with no constrains in the gas phase. All energy criteria were met in the optimisation calculation. The binding energy data were processed and plotted with MS Excel (A5). All coordinates of optimised aggregates can be found in Appendix A7_G7.

7.3 Results and Discussion

The aim of this Chapter was to assess the dimensionality preference, which is wire versus square of a chosen porphyrin aggregation by interchanging the metal and peripheral groups. The following study was based on a model porphyrinic assembly published by Oliveras *et al.*²¹⁰ This structure (Figure 7.2) has been experimentally shown to aggregate in a 3D-square fashion, with each *mer* unit as the side of a square (Figure 7.5 left). Importantly for the aims of this Chapter, each *mer* unit possesses a range of substituents which offer ground for examination in terms of their effect on the aggregation mode. The components of this model *mer* unit entail a central metal, here Zn (highlighted in yellow in Figure 7.2), an electron-withdrawing side group able to form an interaction with the adjacent metal, pyridine (highlighted in blue), and a sterically controlled side group (highlighted in green).

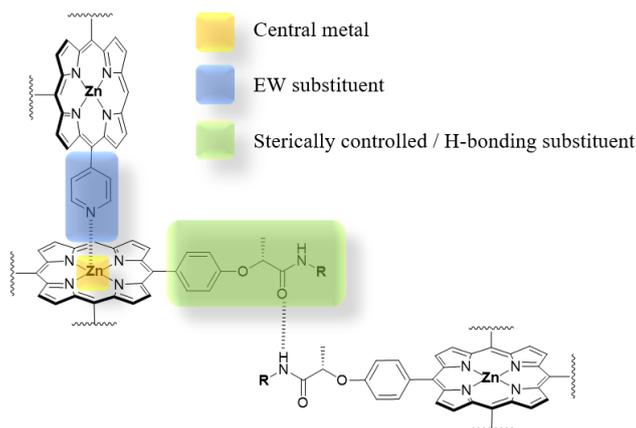


Figure 7.2 Schematic representation of the Oliveras *et al.* aggregation scheme, where **R**: C₁₈-H₃₇. The important interactions in this square aggregation are highlighted and assigned in colour.

In the original work the peripheral sterically controlled groups were utilised not only for steric control in a square assembly, but in alternative motifs explored there for other bonding such as hydrogen bonding with the adjacent unit, this is not explored here,

7.3.1 Variation of the Metal and Peripheral Group Electronic Effects

Initial calculations focused on systematically varying the key interactions of chemical components driving the square aggregation outlined above (Figure 7.2). Divalent metals of groups 10, 11 and 12 of the periodic table were set (Figure 7.3), in keeping with a neutrally charged unit, on a generic porphyrin.

3	4	5	6	7	8	9	10	11	12
21 Sc	22 Ti	23 V	24 Cr	25 Mn	26 Fe	27 Co	28 Ni	29 Cu	30 Zn
39 Y	40 Zr	41 Nb	42 Mo	43 Tc	44 Ru	45 Rh	46 Pd	47 Ag	48 Cd
	72 Hf	73 Ta	74 W	75 Re	76 Os	77 Ir	78 Pt	79 Au	80 Hg
	104 Rf	105 Db	106 Sg	107 Bh	108 Hs	109 Mt	110 Ds	111 Rg	112 Cn

Figure 7.3 Scheme of the systematic choice of metals across and down the periodic table of elements.

Additionally, the peripheral groups were interchanged symmetrically in four positions of the porphyrin unit, to vary the electronic properties, which might affect the strength of binding to the neighbouring porphyrin metal (Figure 7.4).

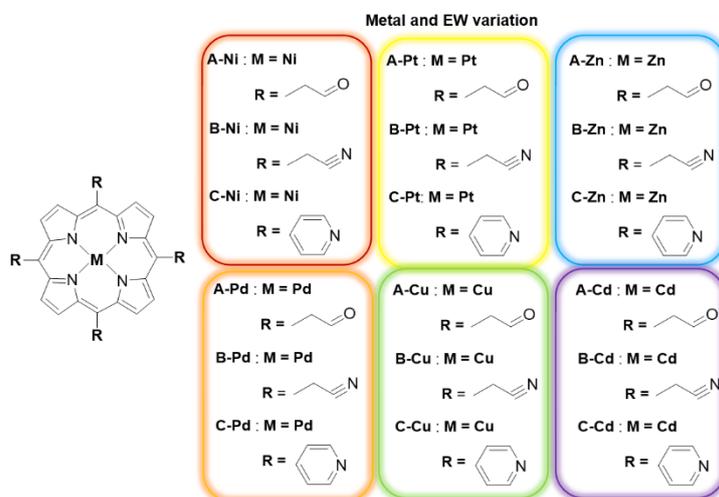


Figure 7.4 Nomenclature and Lewis diagrams of the *mer* unit blocks examined in this work, for 1D-wire (or H) vs 3D-square assembly.

The NCIs between *mer* units in a 3D-square and a 1D-wire aggregation are different (Figure 7.5). In the suggested 3D-square the main drive is formed between a polar nitrogen or oxygen and the central metal of the adjacent unit, which usually sits perpendicularly to the first. In contrast, the wire aggregate is held by metal-metal interactions, π - π interactions (or hydrogen bonding when present) of the peripheral groups as well as the main porphyrinic extended π system.

The smallest unit cell for assessment of the binding energies is a tetramer for square assembly. With this consideration, dimers to tetramers for both the wire and square aggregates were computed at a GFN2-xTB semi-empirical level and the strength of their binding energies was compared to evaluate their relative preferences. Table 7.1 lists all the binding energies for all tetrameric assemblies. The lowest BE indicates the most stable supramolecular conformer. The four composing *mer* units form either a wire stack or a square (as in Figure 7.5).

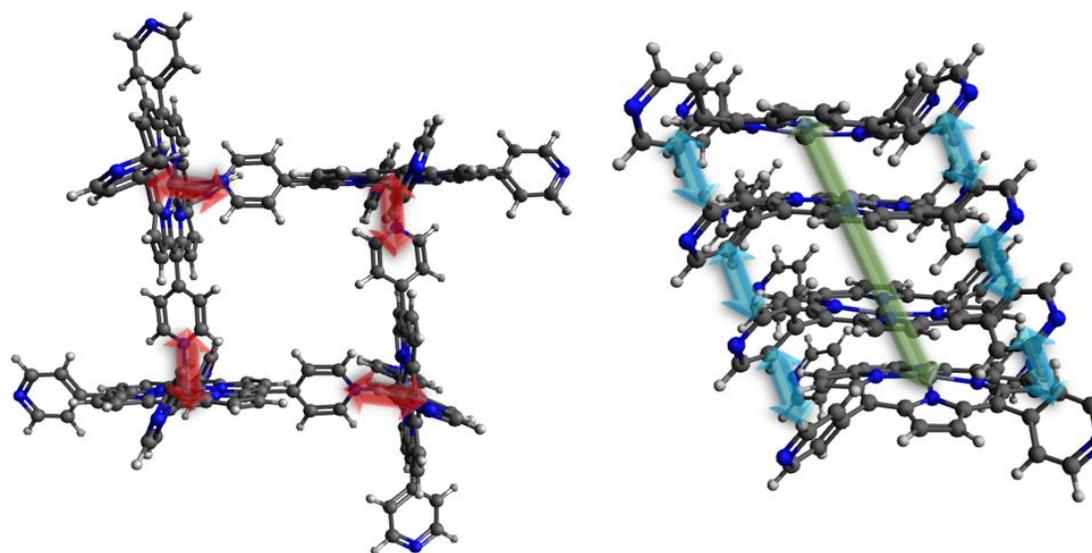


Figure 7.5 Molecular illustrations of the two set modes of aggregation 3D-square on the left and 1D-wire on the right, with representative C-Zn tetramers. The NCIs are shown, where the red is the polar nitrogen with Zn interaction, blue are the π - π interactions of the peripheral pyridine rings and the long green the metal-metal interactions.

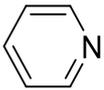
7.3.1.1 Evaluation of the Predictivity for Replicating the Experimental Trends

The Oliveras system (Figure 7.2) was truncated as in Figure 7.5 (structures **A-Zn**, **B-Zn** and specifically **C-Zn**) to reflect the same interactions seen in the original work. This being a pyridine to Zn interaction as the main driving force behind the formation of the square aggregation. The experimental work carried out in this publication shows exclusively square-like assemblies by the porphyrinic entities containing those interactions. Following this template of molecular design aggregation **C-Zn** for 1D-wire and 3D-square motif was assessed computationally for replicating experimental data. In Table 7.1, the BEs and preference of the **C-Zn** are outlined.

Interestingly, the computational data for **C-Zn** predict a wire aggregation motif in contrast to experimental findings. A reflection on these results leads back to the Chapter 6 conclusions, where all ring substituted porphyrins promote an H aggregation with GFN2-xTB computations.

As this prediction is repeated here for a pyridine substituted porphyrin this highlights the difficulty encountered in computations of these systems where π - π interactions are found.

Table 7.1 The binding energies (BE) of tetrameric (n=4) assemblies. All values are in kcal/mol.

Metal	Peripheral group	BE of 1D-wire	BE of 3D-square	Preference
Zn		-105.65	-92.62	1D-wire

7.3.1.2 The Dimensionality Motifs when Altering the Central Metal

With consideration to the above findings where the pyridine remains an obstacle for a reliable predictivity on these systems, alternative substituents were chosen which do not form the problematic π - π interactions. This was the cyano type (Figure 7.4) which provides a similar motif, and as an alternative, a carbonyl possessing substituent with an alternative binding mode for comparison of trends. These are essentially a substituent with an sp hybridized lone pair (the cyano) and one with an sp² hybridized lone pair (the carbonyl). The first forming a synergistic interaction entailing electron donation from ligand to an empty orbital of the metal and electron donation from the filled d orbitals of the metal to the π^* antibonding orbital of the ligand (Figure 7.6).

While the sp² carbonyl ligand is only able to bond to the metal via σ -donation, making the bond weaker. This establishes the hypothesis that the stronger bonding of ligands capable of synergistic bonding to the metal (such as cyano) should promote the formation of the 3D-square assembly.

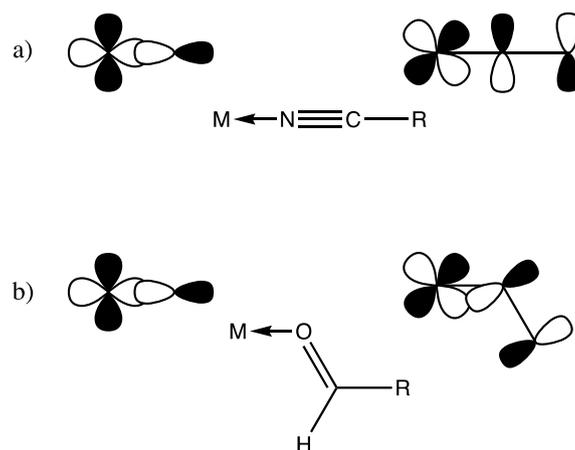
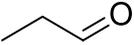
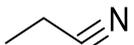
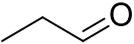
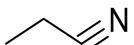
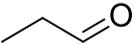
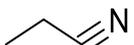
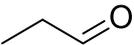
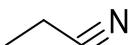
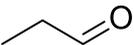
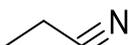
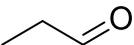
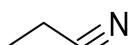


Figure 7.6 Molecular orbital illustrations for a) the synergistic σ -donor and π -back bonding donation between a cyano group and metal (only one set of π -back donation is shown a second orthogonal interaction also exists) and b) σ -donor interaction between a carbonyl and metal (R = porphyrin in both (a) and (b)).

In Table 7.2 the BEs for the 1D-wire and 3D-square assemblies of the tetramers are collected and their preference for a dimensionality motif is assigned. The cyano substituted aggregates show an even distribution between 1D-wire and 3D-square assembly with the cyano containing aggregates. However, there appears to be a preference for first row metals being predicted as 3D-square when substituted with cyano. Upon examination of the trend, when descending the group 10 metals there is no clear pattern, with Ni preferring consistently the square, Pd the wire and Pt a wire with the carbonyl group and a square with the cyano. In this case the strong metal-metal interactions (Figure 7.5) which promote the 1D-wire stack formation, might be at play.

Further, reviewing the carbonyl substituted porphyrins a clear trend is seen with all but the Ni-porphyrin aggregate predicted as a 1D-wire. The collection of these findings can confirm a preference for sp^2 hybridised ligands for 1D-wire formation. Alternatively, sp ligands have more extensive preference for 3D-square assembly, particularly for the first-row metals. The sp hybridised ligand predictions possess, though, some complication as the metal-metal bonding is at play giving unclear results for these trends.

Table 7.2 The BE (binding energies) of all tetrameric (n=4) assemblies. All values are in kcal/mol.

Metal	Peripheral group	BE of 1D-wire	BE of 3D-square	Preference
Ni		-91.79	-100.36	3D-square
		-87.56	-98.39	3D-square
Pd		-136.33	-132.25	1D-wire
		-144.36	-113.01	1D-wire
Pt		-105.10	-103.21	1D-wire
		-98.30	-107.99	3D-square
Cu		-75.11	-61.86	1D-wire
		-79.14	-76.43	1D-wire
Zn		-84.02	-66.76	1D-wire
		-73.53	-100.54	3D-square
Cd		-122.03	-129.16	3D-square
		-113.37	-88.57	1D-wire

7.3.2 Variation of the Steric Effects of the Peripheral Groups by Elongation

A second set of calculations focussed on the steric effect of the peripheral substituents. This was accomplished by gradually elongating the substituent with the addition of a single methylene (CH₂) group at a time, while keeping the central metal parameter constant as the

original (from publication) zinc. The choice of these peripheral groups aimed at the understanding if lengthening these groups can overcome the π - π issue seen at aryl type groups in porphyrins, where an H motif is persistently preferred. This created five different aggregates with cyanide (**1B**, **2B** and **3B**) and pyridine (**1C** and **2C**) containing side groups as detailed in Figure 7.7.

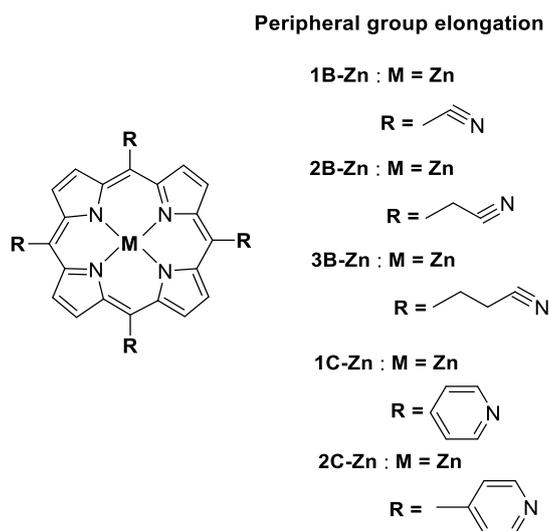
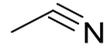
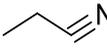
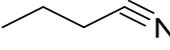
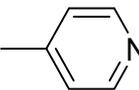


Figure 7.7 Nomenclature and Lewis diagrams of the *mer* unit blocks examined for 1D-wire vs 3D-square assembly.

Table 7.3 shows the calculated 1D-wire versus 3D-square preference for each of the systems investigated. The calculations showed that all pyridine containing aggregates, regardless of elongation parameter, give a 1D-wire preference of aggregation through these semi-empirical computations. While, in the same manner, all cyanide containing aggregates, again regardless of elongation, prefer a 3D-square mode. This additionally underlines the issue of aryl containing peripheral groups previously discussed, which promote a 1D-wire assembly in semi-empirical computations, as shown in Chapter 6.

Table 7.3 The binding energies (BE) of all tetrameric assemblies of zn-porphyrins. All values recorded in kcal/mol.

Peripheral group	BE wire	BE square	Preference
	-108.35	-121.30	3D-square
	-73.53	-100.54	3D-square
	-76.99	-90.68	3D-square
	-105.65	-92.62	1D-wire
	-129.26	-111.14	1D-wire

7.4 Geometrical Observations

Though the focus of this chapter is to solely record 1D-wire vs 3D-square preference of assembly, many other architectures may be formed or preferred with this applied chemical tuning. All optimised geometries (see A7_G7) were examined, in order to monitor any deviations of conformation, that may possibly affect the binding energy. In Figure 7.8 some representative examples are shown.

The Oliveras square assembly is clearly preferred with the specific monomer they present. This is a very symmetrical square-like motif, with each *mer* unit almost completely perpendicular to the adjacent. When making any structural or chemical alterations minor to major deviations are shown. More specifically, as recorded in the optimised tetramers collected, the carbonyl containing **A-Zn** yields the symmetrical square-like 3D-square assembly as in the original structure. The elongated **2C-Zn** promotes a J ladder-like 1D-wire aggregation, while the 3D-square assembly is shown as a distorted asymmetrical geometry.

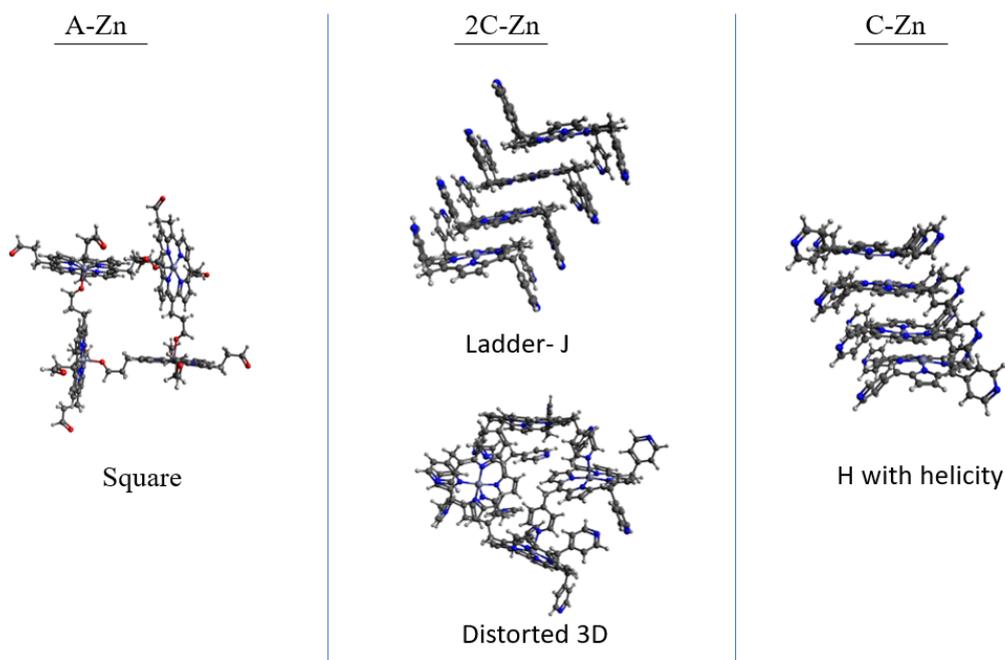


Figure 7.8 Molecular illustrations of A-Zn, on the left, 2C-Zn in the middle and C-Zn on the right tetramers. Indicating the mode of architecture for either 3D-square or 1D-wire for each.

Lastly, as additionally discussed by the Oliveras group, helicity is an important property in aggregates for control and identification purposes; this has been produced with **C-Zn**. The majority of the aggregates has held the pre-set 1D-H and 3D-square (symmetrical) architectures. Yet it is noted that firstly, elongated side substituents promote distorted squares and J-type 1D-wire stacks. Subsequently, pyridine (aromatic ring-type) peripheral groups promote a helicity when added. There are no major differences between metals, with the exception of 1D-wire Zn and Cd stacks showing the central metals in pair formations. These group-12 metals have the ability to form metal-metal dimers possibly affecting the long-range geometry.

7.5 Chapter Summary

The main focus of this chapter was comparing the preference for 1D-wire vs 3D-square aggregation through a systematic increment of metal and peripheral group variation. This was

carried out in two sections where in the first the electronic effect of the peripheral group and the metal were simultaneously varied. In this part it is first revealed Ni gives a persistent square motif through all variations. Additionally, cyano substituents show a preference for square assemblies, especially for first row transition metals. In contrast, carbonyl substituents show a clear preference for a wire. In addition to these substituents, the preference for pyridine peripheral groups was also examined. The calculations showed that these always prefer a wire assembly, this is in contrast to the experimental evidence.

In the second section additional calculations demonstrated that the latter effect is present even upon lengthening the number of linker atoms between the porphyrin and aryl ring system. This further supports the findings from Chapter 6 where it was suggested that GFN2-xTB calculations overestimate the strength of the π - π interaction between the peripheral aryl rings on adjacent *mer* units in these systems. Finally, the effect of the metal upon square or wire formation was investigated, with the calculations suggesting that as the central metal ion becomes larger there is an increased preference for wire formation. This effect being linked to the increasing strength of the metal-metal bond in these systems. Overall, the calculations have shown that the 3D-square versus 1D-wire formation in these types of systems can be predicted based on the nature of the σ -donor ligand, its ability to π -back bond, and complement the size of the central metal ion. However, future work is required to fully explore and categorise the effects of peripheral aryl substituents due to them being mis-predicted in energy minimisation calculations. This work requires the exploration of conformer space using the CREST programme.

Chapter 8

Concluding Remarks

8.1 Introduction

In this chapter the combined results are summarised, and the collected conclusions of the entire thesis are detailed. The findings and novelty of the research is reviewed for making recommendations of how the work could progress in the future, given sufficient time and resources were provided. A new computational method for supramolecular predictions is revealed in Chapter 4, including advantages and disadvantages of previous, traditionally used methods, for supramolecular design. This presents, the recently developed DFTB Hamiltonian GFN2-xTB, as compared with traditionally applied semi-empirical Hamiltonians PMx variations (PM6, PM6-D3H4X and PM7). Improvements in predictions of geometry and error analysis in binding energies are reported through a benchmark chapter. One of the most significant findings presented in this thesis is that the preferred, in this work, methodology has offered superior accuracy and faster computations for geometrical predictions and BEs (binding energies) in supramolecular assemblies of conjugated organic molecules previously reviewed by the PMx Hamiltonians. This has encouraged further computational work to be continued with the focal point of this thesis which is porphyrinic assemblies. This uncovered the intrinsic interactions affecting the fine architectural conformation of supramolecular aggregations, initially for 1D-wire stacks and subsequently in the 3D design. Some of the most relevant chemical tuning exposed so far included the steric effect in promoting J versus H wire stacks of porphyrins, the electronic and steric effect of peripheral groups for a wire versus

square motif of aggregation, and the role of a central metal ion for the preference of motif. The use of a combination of the GFN2-xTB and CREST analysis, for the first time, on supramolecular assemblies has opened the dialogue for further investigations using this combination methodology, for providing mechanistic details previously unexamined.

8.2 A New Semi-Empirical Method for Improved Supramolecular Predictions

A benchmark study was carried out in Chapter 4 of the thesis to compare the capacity of relevant computational methods for predicting properties of supramolecular aggregates. The investigated methods were PM6, PM6-D3H4X, PM7 and GFN2-xTB. These were benchmarked against datasets from the literature (S66 and S30L) with high accuracy reference binding energy values from CCSD(T)/CBS and PW6B95-D3 respectively to carry out an error analysis. These target specifically the NCIs of biologically interesting complexes and larger organic complexes, respectively.

The NCIs (non-covalent interactions) dominant in the conjugated organic assemblies reviewed in this work are hydrogen bonding, π - π interactions and Van de Waals interactions. These were specifically included within both data sets used. An RMSE analysis of the BEs revealed the overall superior performance of GFN2-xTB when accounting for both data sets.

Some notable “strengths” of the remaining methods include the non-dispersion corrected PM6 giving a lower RMSE with the dispersion dominated complexes of S66. PM6-D3H4X yielded the smallest RMSE for the π - π interacting species of S30L. Additionally, PM7 and PM6 outperformed GFN2-xTB with the CH- π interacting species of the latter data set. It is lastly noted that species with counter ions of the S30L set were not predicted accurately by these semi-empirical methods. These findings promote the idea that different methods are best applied in different types of species or applications. Yet it is globally concluded that a robust methodology reliable to an acceptable degree in all NCIs is the GFN2-xTB method.

8.3 The Contributing Factors Affecting Wire Formation

The results reported in Chapter 5, showed that fine structural interplay (a small rearrangement of the topology of the groups or atoms in the molecular structure) within a series of organic conjugated wire assemblies results in distinct changes in geometry and properties of those supramolecular structures. These were reviewed with a combination of experimental spectroscopy methods (carried out by a collaborating experimentalist group) in conjunction with *ab initio* DFT (density functional theory) and semi-empirical calculations. The results showed, that collective NCIs and chemical topology modes can offer new insights into many possible competing pathways governing photochemical properties in these aggregations. The key factors being as follows.

The section relating to the structural interplay in H-bonding and π - π interactions in driving aggregation demonstrated the importance of the former in OPE assemblies. The absence of H-bonding lead to significantly reduced cooperativity. The fine alteration of the direction of the H-bonds with an amide flip showed distortions in the geometry of conjugated core and variations in the H-bonding geometry of interaction. The cooperativity was, additionally, shown altered by this fine interplay. While the computed frontier molecular orbitals in conjunction with the experimental absorption spectra offered new insights about the extrapolated effect that the H-bonding structural topology yields in the photochemical properties of the aggregate. Importantly for the work in this thesis, this study enabled a further analysis of the predictivity of GFN2-xTB, with it illustrating the capacity of this method for reduced computational cost and superior accuracy in geometrical predictions.

In the second section of this chapter the photoisomerization molecular processes were investigated using a combined theoretical and experimental approach. The results provided encouraging insights into the aggregation mechanism pathways of these polymers. It was shown that a sterically bulked photo-isomer yields unfavourable aggregation dynamics as

opposed to a less-sterically hindered linear isomer. These results offer the opportunity for computational design of these systems, in which photo-isomerisation can be predicted.

Finally, the last section on a conjugated organic aggregate, highlights the effect of metal in the structure of the *mer* unit. This shows an increase in cooperativity, different photoisomerization properties than a non-metalized analogue, and valid computed with TD-DFT electronic transitions as compared with experimental spectroscopy. The theoretical approaches applied in this chapter provide an insight into the molecular mechanisms surrounding the assembly modes and dynamics of conjugated organic molecules. Thereby, providing a better understanding of the diverse chemical tuning and structural topological impact highlighted above. Notably, the source of assembly mechanism modes and photochemical phenomena has been explained at the molecular level. Most importantly these predictions at DFT and semi-empirical level in the gas phase have shown good correlation with the experimental findings. In total, the findings reiterate the many undiscovered possibilities offered in the field of supramolecular material design and the exciting opportunities that might be possible in material science and pharmaceutical applications.

8.4 Molecular Design of Wire Porphyrin J and H Stacks

It was reported in Chapter 6 that the ability of GFN2-xTB to predict J or H modes of aggregation in reviewed porphyrinic stacks is strong. The work confirmed the aggregation mode of geometry and connected the thermodynamical preference of aggregation of the porphyrin moieties to the nature of peripheral substitution, as compared to published experimental results. The two main 1D assembly motifs for supramolecular aggregates are J and H, therefore these results offer valuable recommendations for a future synthetic approach of conjugated organic molecular assemblies.

The results showed that GFN2-xTB is capable of correctly predicting J versus H preference for non-aromatic peripheral substituents. Importantly, correctly identifying that the increase in

steric bulk of the peripheral substituents pushes the slippage of *mer* units towards a J aggregation. An interesting discrepancy was met with GFN2-xTB where, when π - π interactions were created amongst the peripheral groups, usually aryl type, the computations consistently converged to the H conformer. It is likely that this effect arises due to the static nature of an energy minimisation calculation, which finds only a single low energy conformer. There are pending questions and obstacles requiring further refinement and development of these newly created Hamiltonians. A key part of those was found by briefly reviewing the CREST analysis on one of those porphyrins, hinting further similar analysis might reveal more information on the subject. This opens the dialogue for further investigations on computations of this type of aggregates.

8.5 Investigations to the Three-Dimensional Aggregate Structures

The final discovery made in this thesis reports structure-property trends linking the chemical tuning of porphyrinic assemblies with their preference for a 1D-wire or 3D-square assembly motif. For the first time, a systematic assessment of the electronic steric effects of the peripheral groups was made for a motif preference. Simultaneously, the central metal ion within a porphyrin in terms of this preference was investigated. This gave valuable information about the effect of chemical tuning in the *mer* unit for controlling dimensionality.

In terms of the electronic effects of the substituents the intuitive expectation that the sp hybridised cyano groups will bind better to the metals than the sp² carboxy was confirmed within the trend of the computations. This enhanced binding leading to a preference for 3D-square formation. Interestingly, this is in keeping with the results from Chapter 6 calculations into the effect of peripheral substitution, which included pyridines hence strong π - π interactions showed a strong preference for a 1D-wire. This however only firmly confirmed the nature of predictions of GFN2-xTB for these species, therefore, future work in this area to test its wider applicability is strongly encouraged. Overall, the trends followed experimental findings

reported for analogues of these assemblies and provide the first seeding work for providing the chemical tuning recommendations for control of dimensionality in supramolecular assemblies.

8.6 Future Work

The results reported in this thesis open the dialogue for many areas of research. The majority of the suggested future work is related to developing computational methodology for supramolecular aggregations and the discovery of trends in the design at a molecular level, as well as identifying supramolecular polymers as significant promising next generation materials. The next step is to focus on the following areas of research.

8.6.1 Development of Computational Methodology

The examples given below will build on the computational predictivity work of the thesis and will be highly advantageous to the field.

- Additional computational prediction work of GFN2-xTB on published experimental data. Further extensive systematic computations for more types of substituents on conjugated organic molecules as well as metal ions when matched with experimental trends will reveal the strengths and weaknesses of the method for these species.
- Addressing the conformational analysis CREST computations for these large aggregates. As supramolecular aggregations, even in the dimer state, are not tractable for any traditional MD (molecular dynamics) computations, the CREST approach must be investigated. The large species contain many internal rotations and conformations which essentially affect the assembly pathways. Notably in this work it was found that the size of the energy window within which CREST computes must be reviewed for better correlation with experimental findings, as

well as that CREST is essential for stacks with strong π - π interactions. It is crucial that these issues are reviewed and reported for improved predictivity.

- Advancing the work on the computational methodology targeted for supramolecular polymers. For the design of supramolecular polymers at the *mer* unit, and the important mechanisms of assembly, a methodology which is tractable for more than 100 atoms and corrected for NCIs is required. From this work it was benchmarked and tested via experimental comparison that GFN2-xTB, with the CREST combination when needed, is the current optimal approach. However, scarce research work or publications have tackled a theoretical approach on the subject to date.^{26, 78, 167} There are still many pending obstacles, one clearly seen in this work is addressing the π - π forces of peripheral groups in porphyrins. These need to be addressed and possibly incorporated in already existing Hamiltonians, or a proposal of a combination of composite methodologies which can treat these issues better.
- The ‘fair’ computational treatment of supramolecular polymers requires an analogue environment to experiment, meaning solvation models and thermochemistry parameters (free energy, enthalpy and entropy). Unfortunately, these are hardly predicted consistently by any theoretical approach developed yet. Further improvement of the solvation models is an approachable project, with clear foundations in the mathematical concepts of it. This can stimulate the imagination and creativity of many researchers in the mathematical and chemistry sector, yielding valuable new models for solvation. An initial step towards this cause would be the review and testing of the optimal combination of the currently available solvation and thermochemistry approximations which best agree with experimental findings.

8.6.2 Structure-Property Relation in Wire Assemblies

The examples given below will build on the computational predictivity work for 1D-wires reported in this thesis.

- The 1D supramolecular stacks discussed throughout the thesis possess distinct electronic and photochemical properties, because of their electron transfer abilities. The FMO (frontier molecular orbital) analysis with TD-DFT showed interesting variations with analogues when structural topological interplay was applied. The photochemical and electronic properties of these materials is key for impactful applications from important environmental engineering²¹¹ to cancer research.¹ Though time consuming synthetic explorations have been made for addressing and controlling these properties, it is important to have at hand recommendations of faster sources for a faster and more accurate synthesis of materials. It would be, then, interesting to advance the TD-DFT results reported herein to include contributions of a wider range of molecular species and compare subtle differences and the presence of any interesting properties that may be discovered.
- The J or H assembly mode plays a crucial role in material applications not only for their differences in spectroscopy but also because their intrinsic intermolecular slippage allows structural accessibility for chemical processes. An interesting example is the interaction with oxygen, being at a higher quantum yield in J conformations.^{34, 198} A synthetic design aiming to specifically obtain one of these motifs is important for a wide spectrum of sectors of research. Computational design at a molecular level will aid for synthetic recommendations promoting a more accurate approach for experimentalists. This entails

examination of a wider range of peripheral substitution and comparison with experimental findings where possible.

8.6.3 The Dimensionality Control in Supramolecular Assemblies

The dimensionality preference of porphyrin aggregations reported in Chapter 7 provided the first example of computational investigations for dimensionality control of organic conjugated species. Therefore, the technique will be used to investigate a variety of additional aggregations and types of materials. The examples below would be interesting materials to study.

- Addressing the effect of metal in supramolecular assemblies has had a key role for synthetic consideration. It has been reported that to reach the third dimension in assemblies, a metal structural parameter is a standard strategy.^{38, 40} The complexity of the chemical nature of the structural components and their topology makes for a cumbersome task concerning the assessment of the metal choice in the *mer* unit. In this work a first systematic approach through the relevant d block metals was made. A more targeted focus on the role of the metal, by removing additional distracting interactions could reveal the nature of metal in dimensionality preference. Continuing this computational work for additional metalized porphyrins in a systematic manner will provide further clarity for a synthetic recommendation.
- Investigating and comparing the electronic structure and properties in and amongst the interacting *mer* units will provide further insights into the nature of factors promoting an assembly mode. As seen in Chapter 5 this information can be provided computationally via TD-DFT computations in the dimers at a first instance. While complementary experimental spectroscopy will be valuable for confirming further these predictions.

- The formulation of a systematic library for synthetic recommendations can be an important tool for cutting down costs and time in experimental work on these materials. The current GFN2-xTB is a semi-empirical Hamiltonian able to provide predictions for large non-covalently bound materials in tractable computations as opposed to traditional *ab initio* and DFT methods that are not tractable for larger entities. By systematically ranging the nature of *mer* unit, the nature of peripheral substituents (steric or electronic), and the nature of aggregation motif input, a library can be built for providing recommendations for future experimental synthesis.

References

1. M. A. Rajora, J. W. H. Lou and G. Zheng, *Chem. Soc. Rev.*, 2017, **46**, 6433-6469.
2. R. van der Weegen, A. J. P. Teunissen and E. W. Meijer, *Chem. -Eur. J.*, 2017, **23**, 3773-3783.
3. E. Iengo, E. Zangrando and E. Alessio, *Eur. J. Inorg. Chem.*, 2003, **2003**, 2371-2384.
4. X. N. Feng, C. X. Liu, X. Q. Wang, Y. Y. Jiang, G. X. Yang, R. Wang, K. S. Zheng, W. X. Zhang, T. Y. Wang and J. Z. Jiang, *Frontiers in Chemistry*, 2019, **7**.
5. C. J. Medforth, Z. Wang, K. E. Martin, Y. Song, J. L. Jacobsen and J. A. Shelnutt, *Chem. Commun.*, 2009, 7261-7277.
6. K. Hosomizu, M. Oodoi, T. Umeyama, Y. Matano, K. Yoshida, S. Isoda, M. Isosomppi, N. V. Tkachenko, H. Lemmetyinen and H. Imahori, *J. Phys. Chem. B*, 2008, **112**, 16517-16524.
7. H. J. R. Schneider, *Applications of Supramolecular Chemistry*, CRC Press, 2012.
8. Y. A. Shieh, S. J. Yang, M. F. Wei and M. J. Shieh, *ACS Nano*, 2010, **4**, 1433-1442.
9. M. Kepczynski, M. Dzieciuch and M. Nowakowska, *Current Pharmaceutical Design*, 2012, **18**, 2607-2621.
10. R. Bonnett, *Chem. Soc. Rev.*, 1995, **24**, 19-33.
11. E. D. Sternberg, D. Dolphin and C. Brückner, *Tetrahedron*, 1998, **54**, 4151-4202.
12. K. K. Kartha, N. K. Allampally, A. T. Politi, D. D. Prabhu, H. Ouchi, R. Q. Albuquerque, S. Yagai and G. Fernández, *Chem. Sci.*, 2019, **10**, 752-760.
13. X. Yan, T. R. Cook, J. B. Pollock, P. Wei, Y. Zhang, Y. Yu, F. Huang and P. J. Stang, *J. Am. Chem. Soc.*, 2014, **136**, 4460-4463.
14. K. Jayaramulu, P. Kanoo, S. J. George and T. K. Maji, *Chem. Com.*, 2010, **46**, 7906-7908.
15. P. L. Marek, H. Hahn and T. S. Balaban, *Energy Environ. Sci.*, 2011, **4**, 2366-2378.
16. C. M. Drain, I. Goldberg, I. Sylvain and A. Falber, in *Functional Molecular Nanostructures*, ed. A. D. Schluter, Springer-Verlag Berlin, Berlin, 2005, vol. 245, pp. 55-88.
17. J. D. Waals and J. S. Rowlinson, *J.D. Van Der Waals: On the Continuity of the Gaseous and Liquid States*, North-Holland, 1988.
18. J. L. Atwood and J. Steed, *Supramolecular Chemistry*, 2009.
19. J. M. Lehn, *Science*, 1993, **260**, 1762-1763.
20. J. M. Lehn, *Supramolecular Chemistry: Concepts and Perspectives*, 1995.
21. J. M. Lehn, *Angew. Chem. Int. Ed. Engl.*, 1988, **27**, 89-112.
22. T. F. A. De Greef and E. W. Meijer, *Nature*, 2008, **453**, 171-173.
23. J. F. Stoddart, *Angew. Chem. Int. Ed.*, 2017, **56**, 11094-11125.
24. F. García, P. M. Viruela, E. Matesanz, E. Ortí and L. Sánchez, *Chem. -Eur. J.*, 2011, **17**, 7755-7759.
25. N. K. Allampally, A. Florian, M. J. Mayoral, C. Rest, V. Stepanenko and G. Fernández, *Chem. -Eur. J.*, 2014, **20**, 10669-10678.
26. N. Caturello, Z. Csok, G. Fernandez and R. Q. Albuquerque, *Chem. -Eur. J.*, 2016, **22**, 17681-17689.
27. F. García and L. Sánchez, *J. Am. Chem. Soc.*, 2012, **134**, 734-742.
28. G. Fernández, F. García, F. Aparicio, E. Matesanz and L. Sánchez, *Chem. Commun.*, 2009, 7155-7157.
29. C. Zhou, Y. Liu and X. Zhao, *Inorg. Chim. Acta*, 2015, **425**, 11-16.
30. T. Birnbaum, T. Hahn, C. Martin, J. Kortus, M. Fronk, F. Lungwitz, D. R. T. Zahn and G. Salvan, *J. Phys. Condens. Matter*, 2014, **26**.

31. Y. Kobuke, in *Non-Covalent Multi-Porphyrin Assemblies Synthesis and Properties*, ed. E. Alessio, Springer-Verlag Berlin, Berlin, 2006, vol. 121, pp. 49-104.
32. S. Mohnani and D. Bonifazi, *Coord. Chem. Rev.*, 2010, **254**, 2342-2362.
33. M. Samperi, D. Limón, D. B. Amabilino and L. Pérez-García, *Cell Reports Physical Science*, 2020, **1**, 100030.
34. Q. Zhao, Y. Wang, Y. S. Xu, Y. Yan and J. B. Huang, *Sci. Rep.*, 2016, **6**, 31338.
35. M. Ruben, J.-M. Lehn and P. Müller, *Chem. Soc. Rev.*, 2006, **35**, 1056-1067.
36. H. M. Titi, B. K. Tripuramallu and I. Goldberg, *Crystengcomm*, 2016, **18**, 3318-3339.
37. M. Kasha, H. R. Rawls and M. A. El-Bayoumi, *Pure and Appl. Chem.*, 1965, **11**, 371-392.
38. J. Otsuki, S. Kawaguchi, T. Yamakawa, M. Asakawa and K. Miyake, *Langmuir*, 2006, **22**, 5708-5715.
39. P. C. Lo, X. Leng and D. K. P. Ng, *Coord. Chem. Rev.*, 2007, **251**, 2334-2353.
40. S. Mohnani and D. Bonifazi, *Coord. Chem. Rev.*, 2010, **254**, 2342-2362.
41. R. E. Banerjee, *Functional Supramolecular Materials*, 1st ed., **2017**.
42. K. S. Suslick, P. Bhyrappa, J. H. Chou, M. E. Kosal, S. Nakagaki, D. W. Smithenry and S. R. Wilson, *Acc. Chem. Res.*, 2005, **38**, 283-291.
43. W. Ji, B. Xue, S. Bera, S. Guerin, Y. Q. Liu, H. Yuan, Q. Li, C. Q. Yuan, L. J. W. Shimon, Q. Ma, E. Kiely, S. A. M. Tofail, M. S. Si, X. H. Yan, Y. Cao, W. Wang, R. S. Yang, D. Thompson, J. B. Li and E. Gazit, *ACS Nano*, 2020, **14**, 10704-10715.
44. S. Tanaka, M. Shirakawa, K. Kaneko, M. Takeuchi and S. Shinkai, *Langmuir*, 2005, **21**, 2163-2172.
45. C. S. Jin, J. F. Lovell, J. Chen and G. Zheng, *ACS Nano*, 2013, **7**, 2541-2550.
46. M. R. Detty, S. L. Gibson and S. J. Wagner, *J. Med. Chem.*, 2004, **47**, 3897-3915.
47. C. D. Ding, Y. Liu, M. D. Wang, T. Wang and J. J. Fu, *J. Mater. Chem. A*, 2016, **4**, 8041-8052.
48. H.-J. Schneider, *New J. Chem.*, 2019, **43**, 15498-15512.
49. A. M. Haruk and J. M. Mativetsky, *Int. J. Mol. Sci.*, 2015, **16**, 13381-13406.
50. J. Delaey, P. Dubruel and S. Van Vlierberghe, *Adv. Funct. Mater.*, **23**.
51. Q. Wei, C. Schlaich, S. Prévost, A. Schulz, C. Böttcher, M. Gradzielski, Z. Qi, R. Haag and C. A. Schalley, *Adv. Mater.*, 2014, **26**, 7358-7364.
52. L.-L. Li, C.-J. Yang, W.-H. Chen and K.-J. Lin, *Angew. Chem. Int. Ed.*, 2003, **42**, 1505-1508.
53. A. Ciferri, *Supramolecular polymers*, Taylor & Francis, Boca Raton, 2nd ed. edn., 2005.
54. M. J. Katz, K. Sakai and D. B. Leznoff, *Chem. Soc. Rev.*, 2008, **37**, 1884-1895.
55. J.-D. Zhou, W.-Q. Zhang, L.-L. Liu, Z.-Q. Xie and Y.-G. Ma, *Chinese Chem. Lett.*, 2016, **27**, 1350-1356.
56. I. A. Baburin, V. A. Blatov, L. Carlucci, G. Ciani and D. M. Proserpio, *Cryst. Growth Des.*, 2008, **8**, 519-539.
57. L. Brunsveld, B. J. B. Folmer, E. W. Meijer and R. P. Sijbesma, *Chem. Rev.*, 2001, **101**, 4071-4098.
58. A. Ciferri, *Supramolecular Polymers*, 2005.
59. E. R. T. Tiekink, *Crystals*, 2020, **10**, 19.
60. S. Ogi, T. Fukui, M. L. Jue, M. Takeuchi and K. Sugiyasu, *Angew. Chem. Int. Ed.*, 2014, **53**, 14363-14367.
61. R. Takahashi and Y. Kobuke, *J. Am. Chem. Soc.*, 2003, **125**, 2372-2373.
62. R. Purrello, L. Monsu' Scolaro, E. Bellacchio, S. Gurrieri and A. Romeo, *Inorg. Chem.*, 1998, **37**, 3647-3648.
63. Y. Arai and H. Segawa, *J. Phys. Chem. B*, 2011, **115**, 7773-7780.

64. M. Shirakawa, S.-i. Kawano, N. Fujita, K. Sada and S. Shinkai, *J. Org. Chem.*, 2003, **68**, 5037-5044.
65. M. Kasha, *Radiat. Res.*, 1963, **20**, 55-70.
66. R. Friedlein, F. von Kieseritzky, S. Braun, C. Linde, W. Osikowicz, J. Hellberg and W. R. Salaneck, *Chem. Commun.*, 2005, 1974-1976.
67. R. B. Martin, *Chem. Rev.*, 1996, **96**, 3043-3064.
68. L. K. S. von Krbek, C. A. Schalley and P. Thordarson, *Chem. Soc. Rev.*, 2017, **46**, 2622-2637.
69. C. Kulkarni, E. W. Meijer and A. R. A. Palmans, *Acc. Chem. Res.*, 2017, **50**, 1928-1936.
70. T. Riis-Johannessen, N. D. Favera, T. K. Todorova, S. M. Huber, L. Gagliardi and C. Piguet, *Chem. -Eur. J.*, 2009, **15**, 12702-12718.
71. F. Biedermann and H. J. Schneider, *Chem. Rev.*, 2016, **116**, 5216-5300.
72. J. D. Badjic, A. Nelson, S. J. Cantrill, W. B. Turnbull and J. F. Stoddart, *Acc. Chem. Res.*, 2005, **38**, 723-732.
73. J. Calbo, J. C. Sancho-Garcia, E. Orti and J. Arago, *Molecules*, 2018, **23**.
74. A. C. Laungani, M. Keller, J. M. Slattery, I. Krossing and B. Breit, *Chem. -Eur. J.*, 2009, **15**, 10405-10422.
75. J. Bojarska, M. Remko, A. Fruzinski and W. Maniukiewicz, *J. Mol. Struct.*, 2018, **1154**, 290-300.
76. L. T. Bergendahl and M. J. Paterson, *Comput. and Theor. Chem.*, 2014, **1040-1041**, 274-286.
77. L. T. Bergendahl and M. J. Paterson, *J. Phys. Chem. B*, 2012, **116**, 11818-11828.
78. D. S. Philips, K. K. Kartha, A. T. Politi, T. Kruger, R. Q. Albuquerque and G. Fernandez, *Angew. Chem. Int. Ed.*, 2019, **58**, 4732-4736.
79. R. Q. Albuquerque, A. Timme, R. Kress, J. Senker and H. W. Schmidt, *Chem. -Eur. J.*, 2013, **19**, 1647-1657.
80. J. Rezac and P. Hobza, *J. Chem. Theory Comput.*, 2012, **8**, 141-151.
81. J. Calbo, E. Ortí, J. C. Sancho-García, J. Aragón and A. D. David, *Annu. Rep. Comput. Chem.*, 2015, **30**, 876-890.
82. J. Calbo, R. Viruela, E. Ortí and J. Aragón, *ChemPhysChem*, 2016, **17**, 3881.
83. L. Moreira, J. Calbo, R. M. Krick Calderon, J. Santos, B. M. Illescas, J. Aragón, J. F. Nierengarten, D. M. Guldi, E. Ortí and N. Martín, *Chem. Sci.*, 2015, **6**, 4426.
84. J. Guilleme, M. J. Mayoral, J. Calbo, J. Aragón, P. M. Viruela, E. Ortí, T. Torres and D. González-Rodríguez, *Angew. Chem. Int. Ed.*, 2015, **54**, 2543-2547.
85. J. Calbo, *Supramolecular Chemistry*, 2018, **30**, 876-890.
86. W. M. C. Foulkes, L. Mitas, R. J. Needs and G. Rajagopal, *Rev. Mod. Phys.*, 2001, **73**, 33-83.
87. C. Riplinger and F. Neese, *J. Chem. Phys.*, 2013, **138**, 034106.
88. C. Riplinger, B. Sandhoefer, A. Hansen and F. Neese, *J. Chem. Phys.*, 2013, **139**, 134101.
89. E. Runge and E. K. U. Gross, *Phys. Rev. Lett.*, 1984, **52**, 997-1000.
90. M. C. Per and D. M. Cleland, *Nano Futures*, 2020, **4**, 23.
91. M. van Schilfgaarde, T. Kotani and S. Faleev, *Phys. Rev. Lett.*, 2006, **96**, 226402.
92. G. Onida, L. Reining and A. Rubio, *Rev. Mod. Phys.*, 2002, **74**, 601-659.
93. J. W. Hwang, P. Li and K. D. Shimizu, *Org. Biomol. Chem.*, 2017, **15**, 1554-1564.
94. J. Hao and A. J. Poë, *Transit. Met. Chem.*, 1998, **23**, 739-747.
95. H. J. Schneider, *New J. Chem.*, 2019, **43**, 15498-15512.
96. C. Puzzarini, L. Spada, S. Alessandrini and V. Barone, *J. Phys. Condens. Matter*, 2020, **32**, 25.

97. J. Noga and R. J. Bartlett, *J. Chem. Phys.*, 1987, **86**, 7041-7050.
98. J. Rezac and P. Hobza, *Chem. Rev.*, 2016, **116**, 5038-5071.
99. R. Sure, J. Antony and S. Grimme, *J. Phys. Chem. B*, 2014, **118**, 3431.
100. R. Sure and S. Grimme, *J. Comput. Chem.*, 2013, **34**, 1672-1685.
101. J. Moellmann and S. Grimme, *J. Phys. Chem. C*, 2014, **118**, 7615-7621.
102. L. Goerigk, H. Kruse and S. Grimme, *Chemphyschem*, 2011, **12**, 3421-3433.
103. J. Řezáč, J. Fanfrlík, D. Salahub and P. Hobza, *J. Chem. Theory Comput.*, 2009, **5**, 1749-1760.
104. E. Caldeweyher, S. Ehlert, A. Hansen, H. Neugebauer, S. Spicher, C. Bannwarth and S. Grimme, *J. Chem. Phys.*, 2019, **150**, 154122.
105. C. Bannwarth, S. Ehlert and S. Grimme, *J. Chem. Theory Comput.*, 2019, **15**, 1652-1671.
106. M. Bursch, E. Caldeweyher, A. Hansen, H. Neugebauer, S. Ehlert and S. Grimme, *Acc. Chem. Re.*, 2019, **52**, 258-266.
107. R. Sure and S. Grimme, *J. Chem. Theory Comput.*, 2015, **11**, 3785-3801.
108. B. Aradi, B. Hourahine and T. Frauenheim, *J. Phys. Chem. A*, 2007, **111**, 5678-5684.
109. S. Grimme, C. Bannwarth and P. Shushkov, *J. Chem. Theory Comput.*, 2017, **13**, 1989-2009.
110. E. Schrödinger, *Ann. Phys.*, 1926, **384**, 361-376.
111. M. Born and R. Oppenheimer, *Ann. Phys.*, 1927, **389**, 457-484.
112. J. C. Slater, *Phys. Rev.*, 1929, **34**, 1293-1322.
113. D. R. Hartree, *Math. Proc. Cam. Phil. Soc.*, 1928, **24**, 111-132.
114. W. Kohn and L. J. Sham, *Phys. Rev.*, 1965, **140**, A1133-A1138.
115. W. Kohn, *Rev. Mod. Phys.*, 1999, **71**, 1253.
116. J. C. Slater, *Quantum Theory of Molecules and Solids*, 1974.
117. S. H. Vosko, L. Wilk and M. Nusair, *Can. J. Phys.*, 1980, **58**, 1200.
118. A. D. Becke, *Phys. Rev. A*, 1988, **38**, 3098-3100.
119. R. G. Parr and W. Yang, *Density-Functional Theory of Atoms and Molecules*, Oxford University Press: Oxford, U.K., Oxford, 1989.
120. J. P. Perdew, K. Burke and M. Ernzerhof, *Phys. Rev. Lett.*, 1996, **77**, 3865-3868.
121. A. D. Becke, *Physical Review A*, 1988, **38**, 3098-3100.
122. A. D. Rabuck and G. E. Scuseria, *Theor. Chem. Acc.*, 2000, **104**, 439-444.
123. A. M. Teale, S. Coriani and T. Helgaker, *J. Chem. Phys.*, 2010, **132**, 164115.
124. A. D. Becke, *J. Chem. Phys.*, 1993, **98**, 5648.
125. P. J. Stephens, F. J. Devlin, C. F. Chabalowski and M. J. Frisch, *J. Phys. Chem.*, 1994, **98**, 11623-11627.
126. Y. Zhao, N. E. Schultz and D. G. Truhlar, *J. Chem. Theory Comput.*, 2006, **2**, 364-382.
127. Y. Zhao and D. G. Truhlar, *Theor. Chem. Acc.*, 2008, **120**, 215.
128. Y. Zhao and D. G. Truhlar, *J. Phys. Chem. A*, 2005, **109**, 5656-5667.
129. T. Leininger, H. Stoll, H.-J. Werner and A. Savin, *Chem. Phys. Lett.*, 1997, **275**, 151-160.
130. H. Iikura, T. Tsuneda, T. Yanai and K. Hirao, *J. Chem. Phys.*, 2001, **115**, 3540.
131. J. D. Chai and M. Head-Gordon, *J. Chem. Phys.*, 2008, **128**.
132. Y. S. Lin, G. D. Li, S. P. Mao and J. D. Chai, *J. Chem. Theory Comput.*, 2013, **9**, 263.
133. A. Görling and M. Levy, *Phys. Rev. B*, 1993, **47**, 13105.
134. J. P. Perdew, A. Ruzsinszky, J. Tao, V. N. Staroverov, G. E. Scuseria and G. I. Csonka, *J. Chem. Phys.*, 2005, **123**, 062201.
135. S. Grimme, *J. Comput. Chem.*, 2006, **27**, 1787-1799.
136. F. M. Raymo, M. D. Bartberger, K. N. Houk and J. F. Stoddart, *J. Am. Chem. Soc.*, 2001, **123**, 9264-9267.

137. J. J. P. Stewart, *J. Mol. Model.*, 2007, **13**, 1173.
138. J. J. P. Stewart, *J. Mol. Model.*, 2013, **19**, 1-32.
139. W. Thiel, *MNDO2005*, 2005.
140. M. J. S. Dewar, E. G. Zoebisch, E. F. Healy and J. J. P. Stewart, *J. Am. Chem. Soc.*, 1985, **107**, 3902-3909.
141. J. J. P. Stewart, *J. Comput. Chem.*, 1989, **10**, 209-220.
142. P. O. Dral, X. Wu, L. Sporkel, A. Koslowski, W. Weber, R. Steiger, M. Scholten and W. Thiel, *J. Chem Theory Comput.*, 2016, **12**, 1082-1096.
143. A. J. Stone, *The Theory of Intermolecular Forces*, 1997.
144. S. Grimme, *J. Comput. Chem.*, 2004, **25**, 1463-1473.
145. S. Grimme, J. Antony, S. Ehrlich and H. Krieg, *J. Chem. Phys.*, 2010, **132**, 154104.
146. S. Grimme, S. Ehrlich and L. Goerigk, *J. Comput. Chem.*, 2011, **32**, 1456.
147. H. B. G. Casimir and D. Polder, *Phys. Rev.*, 1948, **73**, 360-372.
148. J. Řezáč and P. Hobza, *J. Chem. Theory Comput.*, 2012, **8**, 141-151.
149. B. M. Axilrod and E. Teller, *J. Chem. Phys.*, 1943, **11**, 299.
150. Y. Muto, *Proc. Phys. Math. Soc. Jpn.*, 1943, **17**, 629.
151. P. Pracht, F. Bohle and S. Grimme, *Phys. Chem. Chem. Phys.*, 2020, **22**, 7169-7192.
152. S. Grimme, Introduction to CREST -xtb doc 6.2 documentation, <https://xtb-docs.readthedocs.io/en/latest/crest.html>, (accessed 28-08-2020, 2020).
153. J. Rezac, K. E. Riley and P. Hobza, *J. Chem. Theory Comput.*, 2011, **7**, 2427-2438.
154. A. S. Christensen, T. Kubař, Q. Cui and M. Elstner, *Chem. Rev.*, 2016, **116**, 5301-5337.
155. L. Goerigk and S. Grimme, *J. Chem. Theory Comput.*, 2010, **6**, 107-126.
156. M. Korth and W. Thiel, *J. Chem. Theory Comput.*, 2011, **7**, 2929-2936.
157. J. Hostas, J. Rezac and P. Hobza, *Chem. Phys. Lett.*, 2013, **568**, 161-166.
158. J. Řezáč, L. Šimová and P. Hobza, *J. Chem. Theory Comput.*, 2013, **9**, 364-369.
159. J. Antony, R. Sure and S. Grimme, *Chem. Commun.*, 2015, **51**, 1764-1774.
160. Y. Zhao, O. Tishchenko and D. G. Truhlar, *J. Phys. Chem. B*, 2005, **109**, 19046-19051.
161. S. Grimme, *Angew. Chem. Int. Ed.*, 2008, **47**, 3430-3434.
162. L. J. White, N. J. Wells, L. R. Blackholly, H. J. Shepherd, B. Wilson, G. P. Bustone, T. J. Runacres and J. R. Hiscock, *Chem. Sci.*, 2017, **8**, 7620-7630.
163. S. Chakraborty, S. Varghese and S. Ghosh, *Chem. –Eur. J.*, 2019, **25**, 16725-16731.
164. C. Kulkarni, S. K. Reddy, S. J. George and S. Balasubramanian, *Chem. Phys. Lett.*, 2011, **515**, 226-230.
165. D. B. Korlepara, K. K. Bejagam and S. Balasubramanian, *J. Phys. Chem. B*, 2017, **121**, 11492-11503.
166. M. Haring and D. D. Diaz, *Chem. Commun.*, 2016, **52**, 13068-13081.
167. K. K. Kartha, N. K. Allampally, A. T. Politi, D. D. Prabhu, H. Ouchi, R. Q. Albuquerque, S. Yagai and G. Fernandez, *Chem. Sci.*, 2019, **10**, 752-760.
168. S. M. Landge, E. Tkatchouk, D. Benítez, D. A. Lanfranchi, M. Elhabiri, W. A. Goddard and I. Aprahamian, *J. Am. Chem. Soc.*, 2011, **133**, 9812-9823.
169. F. A. Arroyave and P. Ballester, *J. Org. Chem.*, 2015, **80**, 10866-10873.
170. A. Kay and M. Grätzel, *J. Phys. Chem.*, 1993, **97**, 6272-6277.
171. J. Ridley and M. Zerner, *Theor. Chim. Acta*, 1973, **32**, 111-134.
172. Y. A. Mantz and R. L. Musselman, *Inorg. Chem.*, 2002, **41**, 5770-5777.
173. A.-R. Allouche, *J. Comput. Chem.*, 2011, **32**, 174-182.
174. Jmol: an open-source Java viewer for chemical structures in 3D, <http://www.jmol.org/>.
175. M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini,

- F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, T. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman, and D. J. Fox, (2009) Gaussian 09, *Revision C.01*, Gaussian, Inc., Wallingford CT.
176. C. Adamo and V. Barone, *J. Chem. Phys.*, 1999, **110**, 6158-6170.
 177. V. A. Rassolov, J. A. Pople, M. A. Ratner and T. L. Windus, *J. Chem. Phys.*, 1998, **109**, 1223-1229.
 178. P. J. Hay and W. R. Wadt, *J. Chem. Phys.*, 1985, **82**, 299-310.
 179. W. R. Wadt and P. J. Hay, *J. Chem. Phys.*, 1985, **82**, 284-298.
 180. P. J. Hay and W. R. Wadt, *J. Chem. Phys.*, 1985, **82**, 270-283.
 181. C.-H. Liu, M. R. Niazi and D. F. Perepichka, *Angew. Chem. Int. Ed.*, 2019, **58**, 17312-17321.
 182. D. Hashemi, X. Ma, R. Ansari, J. Kim and J. Kieffer, *Phys. Chem. Chem. Phys.*, 2019, **21**, 789-799.
 183. M. J. Mayoral, C. Rest, J. Schellheimer, V. Stepanenko and G. Fernández, *Chem. –Eur. J.*, 2012, **18**, 15607-15611.
 184. M. A. Castriciano, A. Romeo, G. De Luca, V. Villari, L. M. Scolaro and N. Micali, *J. Am. Chem. Soc.*, 2011, **133**, 765-767.
 185. K. Hosomizu, M. Oodoi, T. Umeyama, Y. Matano, K. Yoshida, S. Isoda, M. Isosomppi, N. V. Tkachenko, H. Lemmetyinen and H. Imahori, *J. Phys. Chem. B*, 2008, **112**, 16517-16524.
 186. V. Borovkov, *Symmetry*, 2014, **6**, 256-294.
 187. J. Van Gestel, P. Van Der Schoot and M. A. J. Michels, *Macromolecules*, 2003, **36**, 6668-6673.
 188. V. Villari, P. Mineo, E. Scamporrino and N. Micali, *RSC Adv.*, 2012, **2**, 12989-12998.
 189. T. Fournier, T. H. Tran-Thi, N. Herlet and C. Sanchez, *Chem. Phys. Lett.*, 1993, **208**, 101-105.
 190. M. Vatankhah-Varnoosfaderani, A. GhavamiNejad, S. Hashmi and F. J. Stadler, *Macromol. Rapid Commun.*, 2015, **36**, 447-452.
 191. L. Y. Tang, X. L. Chen, L. Wang and J. Q. Qu, *Polym. Chem.*, 2017, **8**, 4680-4687.
 192. J. K. Gupta, D. J. Adams and N. G. Berry, *Chem. Sci.*, 2016, **7**, 4713-4719.
 193. Y. Nakamura, N. Aratani and A. Osuka, *Chem. Soc. Rev.*, 2007, **36**, 831-845.
 194. H. Kano and T. Kobayashi, *The Journal of Chem. Phys.*, 2001, **116**, 184-195.
 195. E. Collini, C. Ferrante and R. Bozio, *J. Phys. Chem. B*, 2005, **109**, 2-5.
 196. J. Parkash, J. H. Robblee, J. Angew, E. Gibbs, P. Collings, R. F. Pasternack and J. C. de Paula, *Biophys. J.*, 1998, **74**, 2089-2099.
 197. T. Ogawa, E. Tokunaga and T. Kobayashi, *Chem. Phys. Lett.*, 2005, **410**, 18-23.
 198. X. Feng, X. Wang, H. Wang, H. Wu, Z. Liu, W. Zhou, Q. Lin and J. Jiang, *ACS Appl. Mater. Interfaces*, 2019, **11**, 45118-45125.
 199. R. F. Pasternack and P. J. Collings, *Science*, 1995, **269**, 935.
 200. I. Beletskaya, V. S. Tyurin, A. Y. Tsivadze, R. Guillard and C. Stern, *Chem. Rev.*, 2009, **109**, 1659-1713.
 201. S. Ghosh, X.-Q. Li, V. Stepanenko and F. Würthner, *Chem. –Eur. J.*, 2008, **14**, 11343-11357.

202. N. Micali, F. Mallamace, A. Romeo, R. Purrello and L. M. Scolaro, *J. Phys. Chem. B*, 2000, **104**, 5897-5904.
203. S. Yagai, T. Seki, T. Karatsu, A. Kitamura and F. Würthner, *Angew. Chem. Int. Ed.*, 2008, **47**, 3367-3371.
204. G. De Luca, A. Romeo and L. M. Scolaro, *J. Phys. Chem. B*, 2006, **110**, 14135-14141.
205. C.-M. Che, V. K.-Y. Lo, C.-Y. Zhou and J.-S. Huang, *Chem. Soc. Rev.*, 2011, **40**, 1950-1975.
206. Avogadro: *an open-source molecular builder and visualization tool. Version 1.02 ed.*, <http://avogadro.cc/>; Marcus D Hanwell, Donald E Curtis, David C Lonie, Tim Vandermeersch, Eva Zurek and Geoffrey R Hutchison; "Avogadro: An advanced semantic chemical editor, visualization, and analysis platform" *J. Cheminformatics*, 2012, 4:17.
207. M. J. Krische and J. M. Lehn, *Molecular Self-Assembly: Organic Versus Inorganic Approaches*, 2000, **96**, 3-29.
208. R. W. Saalfrank and I. Bernt, *Curr. Opin. Solid State Mater. Sci.*, 1998, **3**, 407-413.
209. C. Oliveras-González, F. Di Meo, A. González-Campo, D. Beljonne, P. Norman, M. Simón-Sorbed, M. Linares and D. B. Amabilino, *J. Am. Chem. Soc.*, 2015, **137**, 15795-15808.
210. C. Oliveras-Gonzalez, F. Di Meo, A. Gonzalez-Campo, D. Beljonne, P. Norman, M. Simon-Sorbed, M. Linares and D. B. Amabilino, *J. Am. Chem. Soc.*, 2015, **137**, 15795-15808.
211. L. L. Li and E. W. G. Diau, *Chem. Soc. Rev.*, 2013, **42**, 291-304.