Computational Studies on the Non-Covalent Interactions of Asphaltene Model Compounds and Related Systems

by

Nathanael King

A thesis submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Department of Chemistry

University of Alberta

Abstract

In this thesis, the non-covalent interactions of asphaltene model compounds are explored computationally using a combination of semi-empirical methods and density functional theory. New methods and workflows are developed with which to make more accurate predictions for reduced computational expense. In Chapter 2, a new synthetic procedure for pyrene-4,5-dione is disclosed. In developing this procedure, unusual purification difficulties were encountered, which led to the computational study in Chapter 3. Here, we used the Grimme group's conformer-rotamer ensemble sampling tool (CREST) to generate ensembles of the low-energy homodimers of pyrene, pyrene-4,5-dione, and pyrene-4,5,9,10-tetraone, and of the heterodimer of pyrene with pyrene-4,5,9,10-tetraone. These ensembles were then further refined using density functional theory (DFT) to give high-quality geometries and energies. We found that the difficulty in purification in Chapter 2 likely originated in the relatively strong interactions between pyrene and pyrene-4,5,9,10-tetraone, which could cause the formation of a cocrystal and interfere with the purification of pyrene-4,5-dione. When the same computational methodology was extended to realistic model asphaltene systems, it became apparent that CREST was providing insufficient sampling of conformational space for those systems, so Chapter 4 details the development and implementation of a new algorithm for non-covalent complexes of flexible monomers, including those involving microhydration. This new algorithm is still based on GFN2-xTB metadynamics, like CREST, but it uses modified settings, with a somewhat weaker biasing, intended to preserve important directed non-covalent interactions for longer, and starts each cycle from a diversity of structures found in the previous cycle, rather than just from the single lowest-energy structure that CREST uses. Thus, we have termed the new algorithm a Low-Energy Diversity-Enhanced variant on CREST (LEDE-CREST). Chapter 5 applies LEDE-CREST to a realistic system of asphaltene model compounds for which there is experimental data available (*Org. Biomol. Chem.*, **2015**, *13*, 6984-6991). Various different stoichiometries of clusters of model compounds were tested. LEDE-CREST was used to

explore the possible geometries for each stoichiometry, and the lowest-energy structure found by LEDE-CREST was then reoptimized using DFT. Unfortunately, due to computational cost of using DFT for systems of this size, we were unable to reoptimize even a portion of the ensembles. The results give insight into motifs in model compound aggregation, the relative importance of different interaction modes (π-π stacking, hydrogen bonding), and shed some doubt on the importance of microhydration. Chapter 6 discusses the results of the entire thesis, and presents directions for future work.

Preface

The candidate confirms that the work submitted is his own, except where work that has formed part of jointly-authored publications is concerned. The candidate created this thesis under the scientific and editorial guidance of Prof. Alex Brown.

Chapter 1 is adapted from the introduction to a paper submitted to *Energy & Fuels*: "A Computational Study of the Non-Covalent Interactions of a Pair of Asphaltene Model Compounds, With and Without Microhydration" by Nathanael J. King and Alex Brown. This introduction was written by the candidate, with subsequent edits suggested by Prof. Brown.

Chapter 2 is an adaptation of a paper which was prepared for submission to *Organic Syntheses*, but never submitted for publication due to purification problems. This work was conceived by the candidate, Prof. Jeffrey Stryker, and Dr. Robin Hamilton, and experiments were performed by the candidate. The candidate wrote the initial manuscript, with subsequent edits suggested by Dr. Robin Hamilton, Dr. Orain Brown, Prof. Stryker, and Prof. Brown.

Chapter 3 is adapted from a paper originally published as King, N.; Brown, A. "Intermolecular Interactions of Pyrene and Its Oxides in Toluene Solution." *J. Phys. Chem. A,* **2022,** *126,* 4931-4940. This work was conceived by the candidate, and computations were carried out by the candidate under the scientific guidance of Prof. Brown. The first draft of the manuscript was prepared by the candidate, with subsequent edits suggested by Prof. Brown, Prof. Stryker, and Dr. Robin Hamilton.

Chapter 4 is an adaptation of a manuscript under review at *J. Comput. Chem.*: "A Variation on the CREST Algorithm for Non-Covalent Clusters of Flexible Molecules." This project was conceived by the candidate and Prof. Brown, and computations were carried out under the scientific guidance of Prof. Brown by the candidate. The coding of the computer script described in this work was carried out by the candidate and Ian LeBlanc. The first draft of the manuscript was written by the candidate, with subsequent edits suggested by Prof. Brown.

Chapter 5 is an adaptation of the body of a manuscript under review at *Energy & Fuels*: "Computational Investigations Complement Experiment for a System of Non-Covalently Bound Asphaltene Model Compounds" by Nathanael J. King and Alex Brown. This project was conceived by Prof. Jeffrey Stryker, Dr. Robin Hamilton, Prof. Alex Brown, and the candidate, and all computations were performed by the candidate under the scientific guidance of Prof. Brown. The first draft of the manuscript was prepared by the candidate, with subsequent edits suggested by Prof. Brown.

Chapter 6 summarizes Chapters 2-5, draws overall conclusions, and proposes future work. This chapter was written by the candidate, with subsequent edits suggested by Prof. Brown. No part of this chapter has been published elsewhere.

The computations performed as part of the work described in this thesis were performed on computing clusters managed by the Digital Research Alliance of Canada, specifically the Graham, Cedar, and Narval clusters.

Acknowledgements

I would like to thank everyone who has helped me along the way, and in doing so, has made this thesis possible. I thank God for making me who I am, for gifting me with the personality He did, for making me His child, and for constantly working on me to improve my heart, my character, and my behaviour. I thank my parents for raising me, supporting me, teaching me, and training me, for encouraging my bents, abilities, and desires, and for helping me to decide on a direction for life. I thank my wonderful wife, Emily, for all of her support, encouragement, help, and brainstorming over the last years. Thank you, Emily, for being my wife, my companion, for putting in more than your fair share at home, to help me make faster progress. I appreciate everything about you from the bottom of my heart. I thank Dr. Steve Twa for getting me hooked on Chemistry. I thank all the professors and TAs who taught my undergrad and graduate courses for helping me to learn. I thank Prof. Jeff Stryker for taking me on and giving my PhD program a start. I thank Prof. Alex Brown for taking me on partway through my program, as I switched from synthesis to computation, and for his support, wisdom, guidance, and encouragement, along with so much high-quality editing advice. I thank Dr. Robin Hamilton for believing in me, and for giving me the encouragement and inspiration to make the switch. I thank Dr. David Scott, Dr. Orain Brown, Mark Aloisio, Dr. Shyam Parshotam, Ryan Johannson, Minh Tang Duc Hoang, Dr. Antoine Aerts, and the various other members of the Stryker and Brown research groups for their companionship, discussions, advice, and other help over the last six years. I thank my supervisory committee, candidacy committee, and defense committee for their oversight, insight, and suggestions. I thank Dr. Hayley Wan and Connor Part for an excellent experience as an organic TA. I thank Anita Weiler for being an excellent administrator, source of information, and problem solver. I thank Mark Miskolzie and Ryan McKay for support and patient explanations in NMR. I thank Dr. Mariusz Klobukowski, Dr. Wolfgang Jaeger, and Dr. Yunjie Xu for encouragement and helpful discussions. I also thank the storeroom and support staff for all their various help over the years. Last of all, thanks to the National Science and Engineering Research Council for funding, and to the Digital Research Alliance of Canada, along with their predecessor, Compute Canada, and regional partners, for computational resources.

Table of Contents

List of Tables

List of Figures

List of Schemes

List of Abbreviations and Symbols

Å – Angstrom – unit of length equal to 10^{-10} m

AFM – Atomic Force Microscopy

BSIE – Basis Set Incompleteness Error

BSSE – Basis Set Superposition Error

CCSD(T) - Coupled Cluster with Singles, Doubles, and perturbative Triples

CI – Configuration Interaction

CPCM – Conductor-like Polarizable Continuum Model of solvation

CREST – Conformer-Rotamer Ensemble Sampling Tool

Da – Dalton – Mass unit equal to one twelfth the mass of carbon-12

DCM - DiChloroMethane

DFT – Density Functional Theory

DFTB – Density Functional Tight-Binding

DLPNO – Domain-based Local Pair Natural Orbital theory

DNA – DeoxyriboNucleic Acid

EDA – Energy Decomposition Analysis – a method for decomposing interaction energies into components

E_h – Hartree – unit of energy equal to 2625.5 kJ/mol

ESI FT-ICR MS – ElectroSpray Ionization Fourier-Transform Inductively Coupled Resonance Mass Spectroscopy

FD – Fluorescence Depolarization

GC – Genetic Crossing

GPC – Gel Permeation Chromatography

HF – Hartree-Fock

HOMO – Highest Occupied Molecular Orbital

LED – Local Energy Decomposition – another term for EDA.

LDI-MS – Laser Desorption Ionisation Mass Spectroscopy

L^2MS – Laser desorption Laser ionization Mass Spectroscopy

LUMO – Lowest Unoccupied Molecular Orbital

MAD – Mean Absolute Deviation

MBD – Many-Body Dispersion

MD – Molecular Dynamics

MTD – MeTaDynamics

MUE – Mean Unsigned Error

NCI – Non-Covalent Interactions

NMR – Nuclear Magnetic Resonance spectroscopy

PES – Potential Energy Surface

RICO – Ruthenium Ion Catalyzed Oxidation

R_f – retention factor

RIJCOSX – Resolution of the Identity Chain Of Spheres density-fitting approximation

RMSD – Root-Mean-Squared Deviation

RRHO/qRRHO – (*quasi*-)Rigid-Rotor-Harmonic-Oscillator approximation

rt – room temperature

SANS – Small Angle Neutron Scattering

SAPT – Symmetry-Adapted Perturbation Theory

SCF – Self-Consistent Field, referring to the method for finding a solution to the many-body problem in the context of HF or DFT

SMD – Solvation Model based on Density

TLC – Thin Layer Chromatography

TMS – TetraMethylSilane

VCD – Vibrational Circular Dichroism

VPO – Vapour Pressure Osmometry

XDM – eXchange-hole Dipole Moment model of dispersion

xTB – extended Tight Binding, a variant of DFTB

Chapter 1

1. Asphaltene Model Compounds and Methods of Study

1.1 Introduction

1.1.1 Asphaltene Composition

Petroleum is important worldwide as an energy source and as feedstock for myriad chemical processes. Even with a global shift toward more renewable sources of energy and chemicals, petroleum use continues to grow. Despite a decline in petroleum use in 2020, usage is rebounding, and 2024 is projected to be a record year for petroleum demand.[1] The simplest reserves to recover and use are those of light, sweet oils, and this has lead to the depletion of such oils in many reservoirs.[2] The remaining heavy oils are much more challenging to exploit, due to factors such as higher viscosity, higher nitrogen and sulfur content, a greater tendency to form emulsions, and greater problems with coking and fouling.[2, 3] All of these issues originate, in part or in whole, from the high concentration of asphaltenes in bitumen and heavy oil.[2-4] In this context, it is important to continually develop understanding which can aid in reducing the environmental and economic impacts of petroleum recovery, processing, and consumption, especially for the case of heavy oils and their asphaltenes.

Asphaltenes are defined as the fraction of crude oils which dissolves in toluene, but not in light *n*-alkanes, although this definition has frequently been critiqued as unhelpful, imprecise, or impractical.[5] Due to their solubility characteristics, asphaltenes frequently precipitate from solution at inopportune times during petroleum extraction, transport, upgrading, or refining, causing constriction or plugging in pipelines and wellheads, fouling and coking on heaters and heat exchangers, and blocking or poisoning catalysts, among other problems, leading to enormous economic and environmental expense. However, the mechanisms behind the aggregation and precipitation of asphaltenes are still poorly understood. Much research has been devoted to elucidating the nature of asphaltene molecules.[6, 7] After many years, and much debate, some consensus has been gained on certain aspects of asphaltenes.

The bulk atomic composition of asphaltenes is easy to measure, and uncontroversial. While there is moderate variation in the elemental composition of asphaltenes between different deposits, they are typically 80-90 wt% C, 6-9 wt% H, 0.8-2.8 wt% N, 2-9 wt% S, and 0.4-5.4 wt% O, with up to a few hundred ppm of Fe, V, and Ni.[5, 8-11] Average molecular mass has been a far more contentious issue. Attempts to measure the molecular mass of asphaltenes have been made by many methods, including Vapour Pressure Osmometry (VPO),[12] Gel Permeation Chromatography (GPC),[13-17] various forms of

Laser Desorption Ionization Mass Spectroscopy (LDI-MS),[17-21] Fluorescence Depolarization,[22, 23] Small Angle Neutron Scattering (SANS),[24] NMR diffusion studies,[25] and Electrospray Ionization Fourier-Transform Ion Cyclotron Resonance Mass Spectroscopy (ESI FT-ICR MS).[26-30] A brief description of each method and its findings appears below. The current consensus among most authors is that most petroleum asphaltene molecules have a mass between 350 and 1200 Da.

VPO determines the concentration of a solution based on its vapour pressure.[12] From the concentration and the mass of solute used to prepare the solution, the molar mass can be determined. However, VPO requires concentration of at least 0.5 g L^{-1}, and asphaltenes are known to aggregate at such concentrations. This leads to artificially high molar masses from VPO, frequently 4000 Da or above.[12]

GPC is frequently used for the determination of polymer and protein molar masses, and involves passing a solution of analyte through a column packed with porous gel. The smaller the molecule, the more detours are possible through the pores of the gel, and the slower it elutes. However, GPC typically uses concentrations of 0.1 % to 0.2 % w/w, and asphaltenes are known to aggregate at such concentrations, again leading to artificially high results for molar mass. In addition, there are concerns that the 3D conformations of some asphaltene molecules may prevent them from entering the pores, leading to erroneously high predicted mass. GPC typically reports a bimodal distribution of molecular mass for asphaltenes, with one peak in the $10^5 – 10^7$ Da range, and another in the 700 – 10 000 Da range.[13-16]

Laser desorption methods must be carefully tuned in order to give useful results. Depending on the exact conditions used, asphaltene molecules may aggregate in the plume or fragment. Either of these undesirable outcomes will negatively impact the quality of the results obtained. One of the more successful laser desorption methods for asphaltenes is the Laser-Desorption Laser-Ionization Mass Spectrometry method (L^2MS), developed by Pomerantz *et al*.[20, 21] L^2MS uses two separate laser pulses: one in the mid-infrared, to desorb without ionizing, followed by a mid-ultraviolet pulse to ionize the desorbed plume. This procedure has been demonstrated to give results that are consistent over a wider range of conditions, and gives a mass distribution for asphaltenes that peaks around 550 Da.[19-21]

Fluorescence Depolarization does not directly measure molecular mass, but instead rotational diffusion coefficients. Put simplistically, the rotational diffusion of molecules is limited by their size, and causes a change in polarization of fluoresced light.[22] These coefficients give approximate molecular

sizes, which can, in turn, be related to approximate molecular weights. Fluorescence depolarization indicates that asphaltene molecules have molar masses of approximately 500-1000 Da.[23]

SANS involves observing neutron diffraction patterns upon passing a neutron beam through a solution, and has been used to measure the size distribution of particles in colloids. In the case of asphaltenes, SANS has been used to track the change in size of molecular clusters with changes in solvent or temperature. The lowest reported correlation length (effectively equivalent to average particle diameter) was 24 Å.[24] Depending on whether there was still clustering under those conditions, the average asphaltene molecule from this sample could be that size or smaller.

NMR diffusion studies use a Pulsed Field Gradient Spin Echo technique (PFG-SE NMR). With increasing magnetic gradient, or alternatively, a longer delay between pulses, the echo signal decays. The rate of decay is related to the diffusion of nuclei parallel to the magnetic gradient. Rather than molecular mass, diffusion can be related to molecular size.[25] Experiments show that apparent molecular size increases with the addition of napthenic acids, which indicates that PFG-SE NMR is measuring the diffusion of aggregates, rather than of individual molecules.

ESI FT-ICR-MS uses electrospray ionization: a solution is pumped through a charged nozzle into a vacuum chamber, and Coulomb repulsion causes the droplets to fragment until each is singly charged. As the solvent evaporates, any solute present acquires the charge which resided in its droplet, leading to soft, single ionization of molecules. FT-ICR-MS is the most sensitive and highest-resolution mass spectroscopy technique yet devised, and is able to resolve peaks of the same nominal mass into their individual atomic compositions. As an illustrative example, $CH_2D_2^+$, with a nominal mass of 18, could be resolved from NH_4^+ and H_2O^+, which also have nominal masses of 18 (although H_2O^+ would be unlikely to form under normal ESI conditions - H_3O^+ would be far more likely). One of the main drawbacks to ESI FT-ICR-MS is that the electrospray ionization is selective, preferentially ionizing bases over neutrals and acids (at least in the case of positively charged ESI – negatively charged ESI prioritizes acids instead), and continental-type molecules over archipelago-type ones (*vide infra* for more on these molecule types). However, careful fractionation before ionization is able to overcome this challenge, and opens the door to the possibility of being able to enumerate *every* molecular composition in an asphaltene sample. ESI FT-ICR-MS reports asphaltene masses mostly in the 400-800 Da range,[26, 27, 30] and has demonstrated the existence of both continental and archipelago asphaltene molecules.[29]

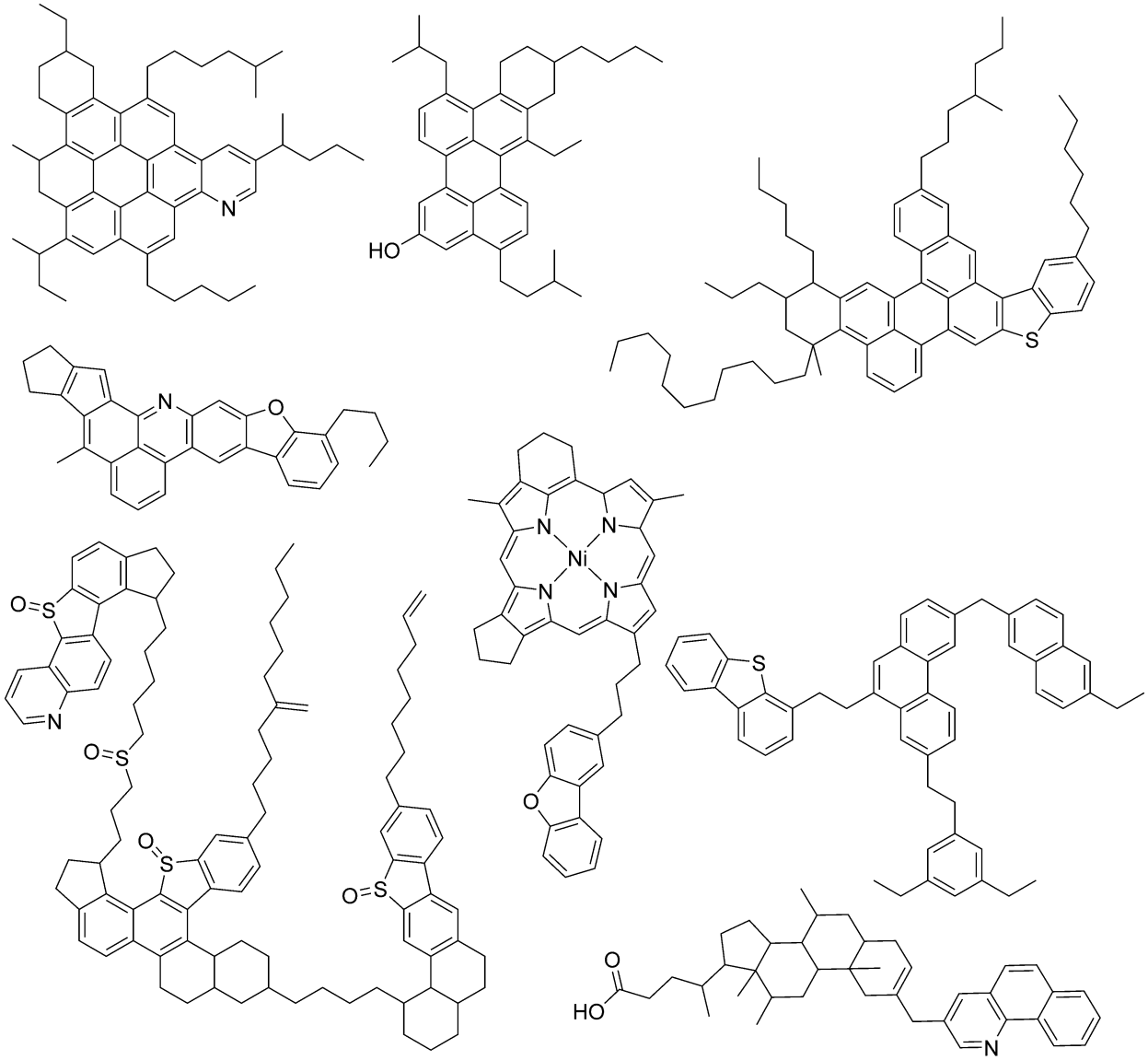
1.1.2 Asphaltene Model Compounds

Equally controversial is the molecular structure of asphaltene molecules. The molecular structure of asphaltenes is universally acknowledged to be very diverse and complex, but the two primary models are the continental model and the archipelago model.[7] In the continental model, all the aromatic and cycloalkyl rings are fused into a single large nucleus, surrounded by alkyl chains. On the other hand, the archipelago model has compounds with multiple, smaller polycyclic aromatic cores, connected by alkyl tethers, and sometimes fused with cycloalkyl rings, which are again surrounded by alkyl chains. The continental model generally considers asphaltene aggregation to be a process dominated by π–π stacking, while proponents of the archipelago model consider π–π stacking, hydrogen bonding, ligand coordination to open sites on porphyrins, electrostatic attraction between permanent charges, and water-mediated aggregation to all be important.[31-34] Both models posit that most of the heteroatoms are incorporated into the aromatic cores, although there are some ether, ketone, and sulfide linkages, and some carboxylic acid functional groups. Various degradation studies including by mild pyrolysis,[35, 36] or selective oxidation[37] or reduction,[38] appear to give evidence for the archipelago model, while Fluorescence Depolarization[22, 23] and Atomic Force Microscopy (AFM)[39-41] appear to support the continental model. Recently, fractionation followed by ESI FT-ICR MS has shown that both types of molecules exist, in varying amounts, depending on the deposit.[29]

Given the overwhelming complexity of asphaltenes, top-down methods (meaning detailed analyses of complete asphaltene samples) are extremely difficult and do not always yield useful information. Some of the best information from top-down analyses has come from ESI FT-ICR MS, which with its gentle ionization and extreme sensitivity and precision, can yield exact molecular formulae for thousands of constituents. However, despite the wealth of previously unobtainable information ESI FT-ICR MS provides, it can only give molecular formulae, not structures, and thus can give only limited insight into intermolecular interactions. In addition, solubility differences and preferential ionization can sometimes prevent some components of asphaltenes from being detected, thus giving only a partial picture. Another method that has come to the fore in recent years is Atomic Force Microscopy (AFM), which *does* give partial structures, but many molecules thus imaged have large regions of ambiguity, possibly resulting from non-flat structures.[39] It can also be difficult to differentiate heteroatoms.[40]

In view of these challenges to top-down analyses, some researchers instead are pursuing realistic model compounds which mimic what we know of asphaltene structure and behaviour.[7, 42, 43]

Scheme 1.1 Structures which have been suggested to be representative of asphaltenes in the literature,[5, 23, 29, 42, 43, 33] with continental-type molecules at the top, and archipelago-type molecules at the bottom.



Model compounds are attractive in that they can be used to test models and hypotheses in a controlled fashion. For example, model compounds possessing certain features believed to be representative of asphaltenes can be tested to see if their properties are in line with those of asphaltenes. If the model compound properties do not line up with those of asphaltenes, then they must deviate from asphaltenes in important ways, and the model can be further refined. For example, polyalkylated hexabenzocoronene has sometimes been considered as a model compound for asphaltenes, but its aggregation behaviour is not similar to that of asphaltenes, so the aggregation behaviour of asphaltenes

must be driven by more than just the π-π stacking of large polycyclic aromatic cores.[31] Alternatively, model compounds have been used to calibrate VPO and SANS,[44] or sulfur K-edge X-ray absorption spectroscopy,[45] to validate L^2MS,[46] or to discredit certain solvation theories,[47] among many other uses. Model compounds are also useful to gain a deeper understanding of how asphaltene-like molecules behave, interact, or react. In the simplified context of a model system, details usually emerge which are indiscernible in the incredibly complex context of natural asphaltenes.[32, 48-50] Several excellent reviews have been published recently on what is known of the nature of asphaltenes[5, 6] and on the use of model compounds in studying them.[7]

1.1.3 Computational Studies Using Asphaltene Model Compounds

Another important use of model compounds is in computational studies.[51-53] While computations must always be interpreted with caution, they can often give insights which are difficult or impossible to obtain experimentally. Molecular conformational preferences, for example, can sometimes be determined experimentally, but this is much simpler to do computationally. Binding energies or free energies of association can be determined either computationally or experimentally, but only computational methods can give a breakdown of the different contributions to such energies. As an example, a computation of Gibbs free energy of dimerization in solution can be easily broken down into changes in solvation, changes in enthalpy, and entropic changes ($T\Delta S$). The enthalpy term can be further decomposed into a binding energy and other enthalpy changes. The other enthalpy changes primarily arise from changes in zero-point energy upon dimerization. All these quantities could be derived from rigorous thermodynamic experiments, but are arrived at very simply from computational output. Certain computational techniques, such as Symmetry Adapted Perturbation Theory (SAPT)[54] or Domain-based Local Pair Natural Orbital approximation to Coupled-Cluster theory with Singles, Doubles, and iterative Triples (DLPNO-CCSD(T))[55] can utilize a method known as Energy Decomposition Analysis (EDA) to further break down the binding energy into components deriving from dispersion, Pauli exchange, electrostatics, and induction, if desired.[56, 57] These results are unique to computations, and cannot be obtained by experiment. It should be noted that the partitioning of the interaction energy in EDA is necessarily somewhat arbitrary, but comparisons between different systems, or between different geometries of the same system, using the same methods, can be used to highlight changes in relative contributions from case to case. Computational methods also allow for analysis of model compounds without going through the arduous process of synthesizing them. Some recent examples of the use of asphaltene model compounds in computational studies are detailed below:

A 2012 study by da Costa *et al*.[33] examined the non-covalent interactions (NCIs) of a series of bipyridyl-based archipelago-style model compounds, using dispersion-corrected density functional theory (DFT). They examined the effects of changing the size of the "islands" in the archipelago, and of the addition of water to complexes. This study reported increasing binding energy, and more stable Gibbs free energy for dimers of compounds with larger islands, and reported that water formed hydrogen-bond "bridges" between molecules, further stabilizing their interactions.

A 2017 study by Wang *et al*.[53] and a 2021 study by Ekramipooya *et al*.[51] each investigated the impact of heteroatom substitution on asphaltene model compounds. The studies used a combination of Molecular Dynamics (MD) and DFT, and tracked changes in propensity for aggregation, hydrogen bonding, electrostatic hardness, HOMO-LUMO gap, surface electrostatic potential, and binding energy. The Wang *et al*. study also used NCI plots to examine the details of the non-covalent interactions. Ekramipooya *et al.* found that incorporation of nitrogen and oxygen increased electrostatic interactions, especially by allowing hydrogen bonding, and that incorporation of sulfur increased dispersive interactions. On the other hand, Wang *et al.* found that in their chosen model system, incorporating a nitrogen into the ring system increased binding energy, but incorporating a sulfur into an alkyl chain decreased it, as did a partial hydrogenation of the ring system.

In 2020, Zhang *et al*. suggested[58] that pancake bonding may contribute to asphaltene aggregation. Pancake bonding is the phenomenon whereby two delocalized radicals form a multi-center, two-electron bond with each other. Because radicals are present but rare in asphaltene, true pancake bonding could only account for a small amount of aggregation. However, Zhang *et al*. also suggested that half-pancake bonding might occur. That is, a delocalized radical might interact with a closed π system, or with a stack of such systems. In 2021, Janesko and Brothers[52] examined this hypothesis by means of a computational study of model compounds. These were not strictly asphaltene model compounds, but model compounds which allowed for an examination of half-pancake bonding *versus* π-π stacking. Using the phenalenyl radical as prototypical of carbon-based delocalized radicals, and pyrene as typical of closed-shell aromatic systems, they compared the stacking interactions of each with a series of other aromatic compounds. They found that, except in cases where there were heteroatoms incorporated internal to the aromatic system, stacking with phenalenyl was always slightly weaker than stacking with pyrene. They concluded that half-pancake bonding could only be adding a small amount of stabilization relative to that due to π-π stacking, but allowed that incorporation of heteroatoms into asphaltene molecules might change this assessment somewhat.

Each of these studies demonstrates the utility of using computational methods to unravel the mechanisms by which asphaltene molecules interact with one another, thus yielding a better understanding of the asphaltene problem. An improved understanding of how asphaltene model compounds interact on the molecular level should enable researchers to better understand and manipulate asphaltene behaviour at larger scales, eventually leading to improved handling and decreased environmental and economic impacts from processing asphaltene-rich crudes.

1.2 Thesis Objectives and Outline

The purpose of this thesis is to give insight into the details of non-covalent interactions (NCIs) in asphaltene model compounds. A better understanding of how asphaltene model compounds interact on the molecular level should enable later researchers to better understand and manipulate asphaltene behaviour at larger scales, leading to improved handling and decreased environmental and economic impacts from processing asphaltene-rich crudes. The structure of the thesis is laid out below.

Chapter 2 of this thesis details a new procedure for the ruthenium-ion-catalyzed-oxidation (RICO) synthesis of pyrene-4,5-dione, a synthetic building block for asphaltene model compounds, optoelectronic materials, and other compounds.[59] Previously published syntheses either used expensive catalysts and/or reagents, or had low yields with difficult workups. For the most used synthesis, that of Harris and co-workers,[60] or the improvement to that procedure by Bodwell and co-workers,[59] the workup involved separating two opaque black layers in a filter flask, followed by a slow filtration step to remove a very fine precipitate of ruthenium salts. Purification was by column chromatography, but due to the low solubility of pyrene-4,5-dione in low-polarity solvents, large silica-to-analyte and solvent-to-analyte ratios were required for a clean separation. Overall, the procedure, workup, and purification takes about a week to perform, with only a moderate yield of 40-60%. Our intention in the work described in Chapter 2 was to develop a new procedure, with a simplified workup and purification which would be more efficient both in terms of materials and time. We succeeded in finding a simpler, less time intensive procedure which could be completed in two days, circumventing the separatory funnel and slow filtration steps. However, we failed to improve the yield or find a more efficient purification procedure. In particular, all solubility-based purifications we tried failed to remove a small residual amount of pyrene-4,5,9,10-tetraone and a small amount of pyrene equimolar to the tetraone. Due to the purification difficulties, the work remains unpublished.

Chapter 3 consists of a computational analysis of the NCIs of pyrene, pyrene-4,5-dione, and pyrene-4,5,9,10-tetraone. This project was undertaken to help probe the difficulties in the purification

of pyrene-4,5-dione, and to gain understanding on how heteroatom substitution affects the non-covalent interactions of polycyclic aromatic systems. The work demonstrated that heteroatom substitution increased the strength of non-covalent interactions between two substituted molecules, and even more so between one substituted and one unsubstituted molecule. This result suggested that the difficulty in purification of pyrene-4,5-dione may have been due to complications arising from the presence of a 1:1 complex of pyrene and pyrene-4,5,9,10-tetraone.

Chapter 4 develops a new variant on the Conformer-Rotamer Ensemble Sampling Tool (CREST) algorithm[61] which improves our results with non-covalent clusters of flexible molecules. We had found CREST to be insufficient when exploring the conformational space of NCI clusters of flexible molecules. CREST results depended heavily on starting structure, frequently finding "best" conformers separated by more than 12 kJ/mol for the same system, depending on the starting structure. To get consistent results, we had to both adjust metadynamics settings and call for restarts from a diverse set of conformers, rather than only the best conformer found so far, as in the CREST algorithm.

Chapter 5 applies the geometry exploration algorithm developed and explored in Chapter 4, and part of the methodology of Chapter 3, to the reinterpretation of experimental results obtained on the interaction of two asphaltene model compounds in solution.[31] We demonstrated that some of the interpretive tools used in the experimental paper were unreliable due to underappreciated complexities of the model system. This work showed that a wide variety of NCI complexes were possible, depending on the circumstances.

Chapter 6 draws overall conclusions from Chapters 2-5, and outlines future directions for further research.

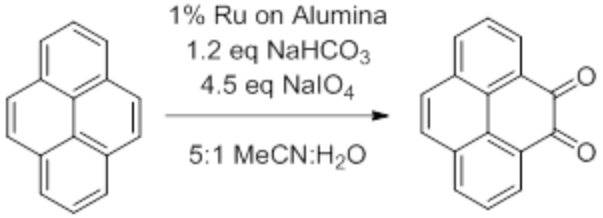
Chapter 2

2. A Procedure for the Synthesis of Pyrene-4,5-dione with an Alumina-Supported Ruthenium Catalyst

2.1 Introduction

Any synthesis of model asphaltene compounds will necessarily be a multistep process, with several intermediates. One attractive approach, especially for archipelago-type model compounds, is to functionalize simple polycyclic aromatics.[7, 62] This approach gives them synthetic "handles" which can then be further modified, cross-coupled, etc. One of the most common functionalizations for polycyclic heteroaromatics is halogenation, which primes them toward metal-catalyzed cross-coupling.[63-65] Oxidation can be a good complement to halogenation, as it can activate different regions of a molecule than halogenation, allows different subsequent modifications, and can even direct further halogenation toward previously unreactive sites.[66-68] Oxidized sites (generally phenolic groups or ketones) can be reactive to nucleophilic (ketones) or electrophilic (phenols) attack. Nucleophilic attack in particular is valuable for its ability to give new carbon-carbon bonds. Thus, the oxidation of polycyclic aromatics is an important avenue toward preparing asphaltene model compounds.

Scheme 2.1 Synthesis conditions for pyrene-4,5-dione in this work



Pyrene-4,5-dione has historically been difficult to synthesize, but remains a very useful starting point for preparing photoactive compounds and novel organic materials.[69-73] The earliest synthesis, published in 1937,[74] suffered from very low yields (< 5%). In 1959, Oberender and Dixon published the first ruthenium catalyzed procedure, but the yield was still only 11%.[75] The most widely used modern procedure is the ruthenium-catalyzed oxidation developed by Hu *et al.*[60] or the improvement thereon by Walsh *et al.*[59] The method of Walsh *et al.* uses an *N*-methylimidazole additive, and has a somewhat easier workup than that of Hu *et al*. Unfortunately, for both cases, the reaction work-up requires separating two opaque black layers, which can easily form emulsions. The work-up is further complicated by the presence of very small particles of insoluble ruthenium salts which make filtration very slow and often ineffective. Furthermore, column chromatography is required to obtain pure 4,5-

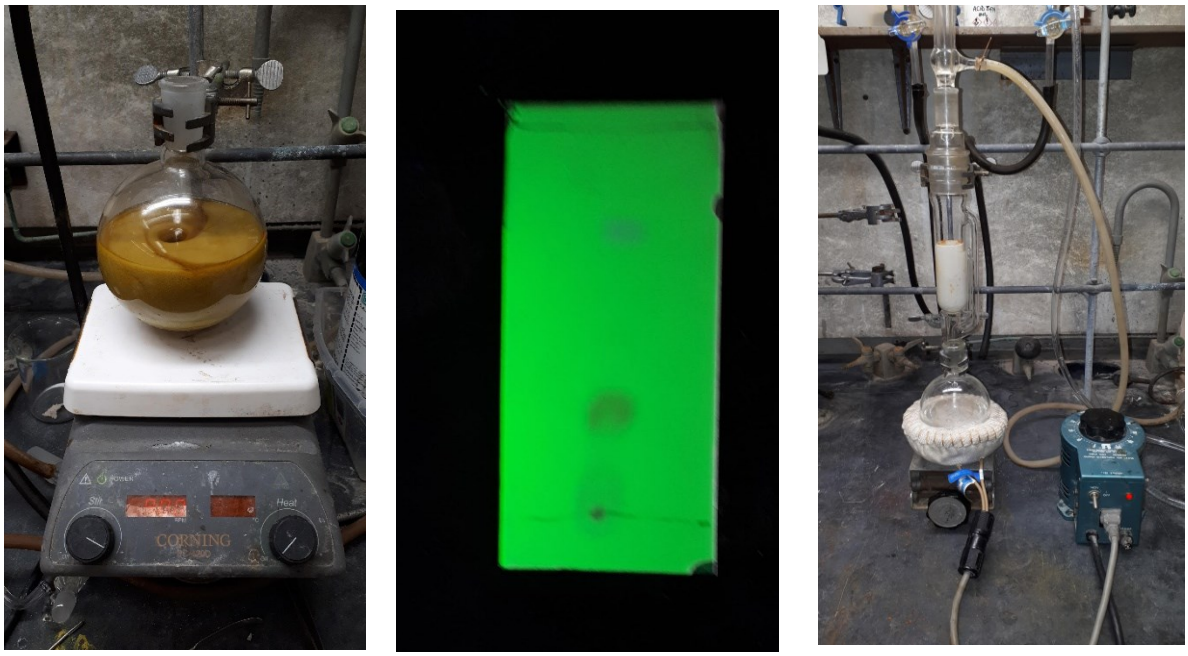
pyrenedione. These challenges limit the yield, with the highest reported yield for the procedure of Hu *et al.* being 60%.[76] In our hands, 40% was found to be the upper limit. Much of the difficulty in the workup is associated with removing the ruthenium colloids which form during the reaction. Our proposed solution was to disperse the ruthenium on a solid support. Alumina proved to be a suitable support, so long as both water and a weak base were present in the solvent. After the reaction is complete, the supported catalyst is easily removed by filtration, greatly simplifying the workup procedures. We have developed a new process that has a comparable yield to existing protocols, but with a simplified, chromatography-free procedure.

2.2 Procedure

2.2.1 Catalyst Preparation

A 1000-mL round-bottom single-necked flask is equipped with a 3/4-inch-long, egg-shaped stir bar, and charged with 150 mg ruthenium (III) chloride trihydrate (0.57 mmol), 30 g adsorption alumina, and 5 g sodium bicarbonate (60 mmol). 500 mL acetonitrile and 100 mL water are added. The mixture is stirred, open to air, at room temperarure for 30 minutes, until the liquid phase is clear and colourless, and solids are a uniform grey-brown, then placed in an ice bath and stirred for a further 30 minutes.

Figure 2.1: Panel a (left) shows the reaction in progress, panel b (middle) shows reaction progress being monitored by TLC, and panel c (right) shows the Sohxlet extraction setup.



2.2.2 Reaction Procedure

10 g Pyrene (50 mmol) is added to the cooled flask, followed by 45.4 g Sodium metaperiodate (212 mmol), in approximately 0.5 g portions. Dissolution of $NaIO_4$ in the $MeCN/H_2O$ solvent is

exothermic, and the extra heat can cause side reactions and decreased yield. Thus, the reaction flask is cooled in an ice bath before adding periodate, and it must be added slowly enough not to raise the temperature in the flask above 30 °C (About 10 to 15 minutes). By the end of the addition, the mixture becomes yellow-orange. The ice bath is then removed, and the flask stirred, open to air, at room temperature for 4 hours, see Figure 2.1a. As the reaction progresses, the colour changes from orange to a reddish-brown. The progress of the reaction can be monitored using thin layer chromatography (TLC) on silica with 3:1 hexane:ethyl acetate as an eluent, see Figure 2.1b, although it will not attain completion. The reaction mixture must be diluted prior to spotting on the TLC plate, or it will streak during plate development. Pyrene has an R_f of 0.72, while the product has an R_f of 0.25. Regardless of how much oxidant is added, some pyrene remains.

2.2.3 Workup

After 4 hours, the reaction mixture is diluted with 350 mL water and stirred vigorously for 5 minutes. This extra water dissolves the salt byproducts, so that they do not interfere with filtration. If it is added during the reaction instead, the yield is decreased. The reaction mixture is then filtered using a Büchner funnel and Whatman grade 1 filter paper. The filter cake is washed with 250 mL of 1:1 water and acetonitrile, then 250 mL of 1:3 water and acetonitrile, and finally with 500 mL of pure acetonitrile. A total of 20 g of sodium thiosulfate is added to the filtrate to quench the remaining oxidant and reduce residual ruthenium to Ru(III). The filter cake is then further washed with 750 mL of DCM. The water/acetonitrile filtrate is extracted with 2 x 750 mL DCM, and the organic layers are combined with the DCM wash. Rotational evaporation removes DCM and acetonitrile, affording a red-brown solid. The solid is made up of pyrene, 4,5-pyrenedione, and 4,5,9,10-pyrenetetraone (*vide infra*), as assessed by nuclear magnetic resonance spectroscopy (NMR). The solid mixture is placed in a cellulose extraction thimble, and subjected to Sohxlet extraction, see Figure 2.1c, with 250 mL of hexane, to remove most of the pyrene and pyrene tetraone. The remaining solids are dissolved in 400 mL hot toluene (110 °C), gravity filtered while hot, cooled to room temperature over 1 h, then cooled a further 1 h on ice. The crystals are collected by vacuum filtration using a Büchner funnel, and washed with ice cold toluene. At this point, small amounts of pyrene and 4,5,9,10-pyrene tetraone remain as impurities. See Appendix A for an NMR spectrum showing the peaks for the product and the two impurities. A second recrystallization can be performed to reduce the amounts of these impurities to <1%, but this will result in decreased yield. If the next synthetic step will tolerate the presence of pyrene and the tetraone, then one recrystallization is sufficient. The impurities become much easier to separate after the dione has been derivatized, for example, by bromination. To remove residual toluene, the crystalline solid is dried

under high vacuum (10^{-5} Torr) for two days. In total, 2.65 g of red-orange needles were isolated. 3.8 g of unreacted pyrene is recovered by evaporating the mother liquor from recrystallization together with the hexane from the Sohxlet extraction. Overall yield, after recovery of starting material, is 37 %.

2.3 Conclusion

We have developed alternate conditions for the RICO synthesis of pyrene-4,5-dione from pyrene. While the procedure is simpler and faster than previous methods, and can be done on a multigram scale, the yield is not improved over previous methods. We have identified a recrystallization method which removes most impurities, but a small, refractory percentage of pyrene and pyrene-4,5,9,10-tetraone remain. Column chromatography, as performed in previous published syntheses,[59, 60] would work, in theory, but due to the low solubility of pyrene-4,5-dione in typical normal-phase chromatography solvents, concentrations must be kept low to avoid streaking. This requires massive amounts of silica and solvent for effective chromatography, and so is not an attractive solution. Further work is needed to achieve an effective purification. See Chapter 3 for a computational study of the reasons why purification is so difficult.

Chapter 3

3. The Intermolecular Interactions of Pyrene and its Oxides in Toluene Solution

3.1 Introduction

The non-covalent interactions of extended π systems are one of the more interesting and important aspects of the chemistry of condensed phases, yet they are still not well understood. Non-covalent interactions primarily consist of dispersion, which is always attractive, exchange repulsion, and electrostatics. Electrostatics and exchange can be handled well by mean-field theories such as Hartree-Fock (HF) or Density Functional Theory (DFT), but dispersion requires advanced wavefunction methods such as Configuration Interaction (CI) or coupled-cluster theory to describe it robustly. However, in the last two decades, remarkable progress has been made with dispersion corrections to DFT, such as the DFT-D series of corrections by Grimme,[77-79] the MBD (Many-body Dispersion) correction of Tkatchenko *et al*.,[80] the XDM (exchange-hole dipole moment) correction of Becke and Johnson,[81, 82] or the VV10 correction of Vydrov and Van Voorhis.[83] Each of these approaches built upon DFT provide a computationally less expensive, though also less rigorous, means of investigating dispersion than wavefunction theory allows.

The non-covalent interaction of extended π systems, both aromatic and non-aromatic, can be considered a special case of dispersion-dominated van der Waals interactions, generally referred to as π–π stacking. The molecular planarity allows for a close approach between molecules, which enhances dispersion, while the delocalization of the π clouds softens repulsive exchange.[84, 85] This stable interaction is reinforced by electrostatic interactions – primarily charge interpenetration and quadrupole-quadrupole effects.[86, 87] There has been much research done on the interactions between aromatic hydrocarbons, especially the acenes. The gas-phase interaction in the most stable conformer of the benzene dimer is similar in strength to that in the cyclohexane dimer, and that in the naphthalene dimer is similar to that in the decalin dimer.[85, 87] However, at sizes larger than two fused rings, the interaction of acenes becomes stronger than that of perhydroacenes.[85, 87, 88] Non-linear acenes, such as pyrene, perylene, and coronene, have intermolecular interactions in line with those of similarly-sized linear acenes.[87]

Virtually all published computations on π −π stacking of hydrocarbons were computed in the gas phase. However, many real-life processes involving π−π stacking, such as DNA base pairing and asphaltene nanoaggregation, occur in solution phase instead. It is important to know the effects of solvation on π−π stacking if we want to understand these processes.

A rigorous treatment of solvation would involve adding many molecules of explicit solvent, and running molecular dynamics over nanoseconds to show the interactions of the molecule with solvent, but for *ab initio* methods, the cost is absolutely prohibitive. Instead, various approximations have been developed that allow an approximate determination of solvation effects. Most such approximations belong to the family of continuum solvation methods, which approximate the solvent as a dielectric continuum which reacts to and stabilizes accumulations of charge on the surface of the solute. Continuum solvation models do a poor job on directed interactions of solvent with solute, such as solute-solvent hydrogen bonds, but usually give reasonable results for large, neutral species without specific solute-solvent interactions.

While substantial work has been done toward understanding π−π stacking in hydrocarbons,[84-87] there has been very little work on the π−π stacking of hydrocarbons with heteroatom substitutions.[89-91] Whether the heteroatoms are incorporated into heterocycles or as substituents on a ring, they affect the electrostatic properties of a π system, especially its dipole and/or quadrupole moments. The effects of heteroatom substitution on π−π stacking are of great interest in a number of chemical fields, including DNA base pairing[92] and asphaltene nanoaggregation.[93] It is also of interest to determine the changes in π−π stacking in solution phase *vs.* gas phase.

Asphaltenes are a class of compounds in heavy oil that are defined as the fraction that is soluble in toluene, but not in pentane.[94, 95] They are characterized by high C/H ratios, large amounts of nitrogen, sulfur, and oxygen relative to bulk crude, high aromaticity, and molecular masses over 500 Da.[94, 95] Due to their solubility characteristics, asphaltenes commonly precipitate from solution during processing, plugging or fouling wellheads, pipelines, heat exchangers, and other machinery.[24, 94] The shutdowns and maintenance required to remove precipitate, and the efficiency losses due to asphaltene fouling, have immense environmental and economic impacts. Much research is currently underway to understand the mechanism of asphaltene aggregation.[7, 96, 97] Further understanding of the intermolecular interactions between asphaltene molecules may lead to an improved understanding of the aggregation process, and may offer ways to prevent such aggregation, thus reducing the environmental and economic impacts.

Pyrene is a hydrocarbon well suited to the theoretical exploration of π–π stacking. It is large enough that the unique effects of large, multi-ring π systems come into play,[85, 87] but small enough that computations at a reasonable level of theory are still feasible. Pyrene is also a good test case on which to examine the results of aromatic oxidation on π–π stacking. In 2005, Harris and co-workers developed a ruthenium-catalyzed oxidation to produce either pyrene-4,5-dione or pyrene-4,5,9,10-tetraone, depending on conditions.[60] In 2016, Bodwell and co-workers published an improvement to the procedure which improves the yield and convenience of the reaction,[59] though the workup is still tedious. Pyrene-4,5-dione has a large dipole moment, and pyrene-4,5,9,10-tetraone has a large quadrupole moment which is opposite in sign to that in pyrene. While pyrene and its oxides are not large enough molecules to be good model compounds for asphaltenes, the effects of oxidation on both the orientation and energetics of dimerization should yield valuable insights into intermolecular interactions of larger, asphaltene-like molecules having carbonyls in the ring system.

Another point of interest is to see how pyrene-4,5,9,10-tetraone interacts with pyrene. Their opposed quadrupole moments can be expected to lead to a stronger than normal intermolecular interaction. In fact, this study was partially motivated by unpublished experimental work on pyrene-4,5-dione synthesis where we found that such interactions seemed to complicate the purification of pyrene-4,5-dione (see Chapter 2). Specifically, the crude pyrenedione product of our procedure (as well as those of Harris and Bodwell) is contaminated by a substantial amount of leftover pyrene and a small amount of pyrene-4,5,9,10-tetraone. While both Harris and Bodwell purify by chromatography, we were aiming for a simpler purification by recrystallization. However, upon recrystallization from toluene, it was found that the product was still contaminated by the tetraone, as well as an equimolar (to tetraone) amount of pyrene. Sohxlet extraction with hot hexane also failed to remove the tetraone and the equimolar amount of pyrene. One possible conclusion was that the pyrene and pyrenetetraone had formed a cocrystal with similar solubility characteristics to pyrenedione crystals.

The crystal structures of pyrene,[98] pyrenedione,[99] and pyrenetetraone[100] are known. In each, π–π stacking between molecules is a dominant motif. To a first approximation, the stability of a crystal can be related to the strength of the π–π stacking interaction between molecules. Indeed, in forming a crystal from solution, the first step must be the formation of a dimer, which could then add further monomers or coalesce with other dimers or larger structures. This approximation neglects edge-to-π stacking and other interactions between molecules in the crystal, but the π–π stacking is expected to be the dominant interaction, and its strength should be a good indicator of the stability of a crystal.

The dimerization of pyrene has been studied computationally.[54, 87, 100-112] Much of the published work uses force fields, paired with molecular dynamics or metadynamics.[101, 106, 107, 113, 114] Such work has mostly focused on kinetics or on binding energies, without considering free energies. There is a single example reporting the Helmholz free energy.[114] Most published *ab initio* work considers only one conformer of the dimer, usually SP-L (slipped parallel, displaced along the long axis), and only the gas-phase binding energy. The binding energy is reported to be in the range of 41.4 to 62.3 kJ/mol.[54, 100, 102, 103, 110] Only one study used *ab initio* methods to determine a free energy of dimerization for pyrene.[100] This study, by Zreid *et al*., determined both the binding energy, and the free energy of dimerization for the dimers of pyrene, pyrene-4,5-dione, and pyrene-4,5,9,10-tetraone, using the nearest-neighbour structures extracted from the crystal structures. They computed both binding and Gibbs free energies using ωB97X-D/6-311++G(2d,p). This resulted in a binding energy of 60.63 kJ/mol and a free energy of dimerization of -4.33 kJ/mol for gas-phase pyrene dimer, but at the geometry of nearest neighbours in the crystal, rather than a gas-phase optimized geometry. This paper also reported the binding energy and free energy of dimerization for pyrene-4,5-dione to be 79.53 kJ/mol and -17.50 kJ/mol, respectively, while those for pyrene-4,5,9,10-tetraone were 69.54 kJ/mol and -5.09 kJ/mol. As before, these reported energies are for the nearest neighbours in the crystal structure, but computed in the gas phase. To date, there are no computational studies reporting on the dimerization of pyrene, pyrene-4,5-dione, or pyrene-4,5,9,10-tetraone in solution, and none on the heterodimerization of pyrene-4,5,9,10-tetraone with pyrene in the gas phase or in solution.

Only a few reports consider pyrene dimer geometries beyond SP-L.[54, 87, 102, 115] The designations for the various conformers vary somewhat from report to report, so in the following paragraph, our designation for a conformer will be indicated in parentheses, where applicable, or a description will be given if our methods did not reproduce the conformer. Our designations for conformers are given and illustrated in the first part of the Section 3.3. In 2003, Gonzales and Lim found four structures, using HFD/6-31G: SP-L (0SP-L), with a binding energy of 51.00 kJ/mol, G (0G), with a binding energy of 50.09 kJ/mol, SP-S (0SP-S), with a binding energy of 50.08 kJ/mol, and X (crossed axes, at approximately 60 degrees, with centre points superimposed), with a binding energy of 48.95 kJ/mol.[102] In 2008, Podeszwa and Szalewicz reported a similar collection of four structures, using SAPT(DFT)/aug-cc-pVDZ based upon PBE0: SP-L (0SP-L), with a binding energy of 41.42 kJ/mol, Gr (0G), with a binding energy of 41.30 kJ/mol, SP-S (0SP-S), with a binding energy of 40.79 kJ/mol, and X (90X), with a binding energy of 39.46 kJ/mol.[54] The main differences between these two ensembles is that the X conformer is perpendicular in SAPT(DFT), but not in HFD, and for the other conformers, the SAPT(DFT) binding

energies are smaller than the HFD ones. In 2009, Rapacioli *et al*., using c-DFTB-D, found seven conformers of the pyrene dimer: SE (all atoms superimposed), with a binding energy of 37.36 kJ/mol, ST (90X) at 41.59 kJ/mol, S3 (60G1) at 43.26 kJ/mol, PD1 (0PD-L) at 41.92 kJ/mol, PD2 (0PD-S) at 42.76 kJ/mol, T1 (T-shaped, short axis perpendicular) at 24.94 kJ/mol, and T2 (T-shaped, long axis perpendicular) at 16.69 kJ/mol.[115] This study has made the most systematic search for low-energy pyrene dimer conformations of any reports, by scanning along several intermolecular coordinates. In 2018, Cabaliero-Lago and Rodríguez-Otero reported the binding energies for the SP-L (0SP-L), G (0G), X (90X), and sandwich (all atoms superimposed) geometries as 53.60 kJ/mol, 54.64 kJ/mol, 50.50 kJ/mol, and 37.91 kJ/mol, respectively, using the MP2.X//TPSS-D3BJ/def2-TZVPP method.[87] While these four studies give the binding energies for a range of conformations of the dimer, none of them appear to have systematically explored the conformational space in such a way as to determine whether there are any more low-lying minima.

In this work, CREST 2.0.1[61] (The Grimme group's Conformer-Rotamer Ensemble Sampling Tool) is applied to explore the conformational ensemble of each of the selected dimers. The ensembles are then refined using ωB97X-D4[116]/def2-TZVP,[117] both in the gas phase and using CPCM[118] (during optimization) and SMD[119] (at CPCM minima) to approximate toluene solution. Free energies are available through Grimme's quasi-rigid-rotor-harmonic-oscillator[120] (RRHO) approximation. The Gibbs free energies of dimerization of these complexes should provide a good proxy for the relative stability of the four crystal phases *vs*. dissolved monomers in solution. Single point computations of selected conformers are performed using DLPNO-CCSD(T)[55]/cc-pVTZ,[121] to obtain high quality interaction energies, and to further validate the DFT results. Thus, we achieve our primary goals of investigating the nature and strength of interaction in solution for each of the dimers investigated, thus gaining insight into crystal stabilities, and also fulfil our secondary goal of demonstrating the reliability of our approach of using CREST followed by ωB97X-D4 reoptimization for problems involving the interactions of polycyclic aromatic systems in solution.

3.2 Methods

The free energy of dimerization of molecules in solution (ΔG_{dim}) is usually determined using a thermodynamic cycle,[122] such that

$$\Delta G_{dim} = \Delta E_{gas,dim} + \Delta G_{rrho,dim} + \Delta\Delta G_{solv.dim}. \quad (3.1)$$

The terms on the right hand side are given by

$$\Delta E_{gas,dim} = E_{gas,dim} - E_{gas,mon1} - E_{gas,mon2}, \quad (3.2)$$

$$\Delta G_{rrho.dim} = G_{rrho,dim} - G_{rrho,mon1} - G_{rrho,mon2}, \quad (3.3)$$

and

$$\Delta\Delta G_{solv.dim} = \Delta G_{solv,dim} - \Delta G_{solv,mon1} - \Delta G_{solv,mon2}. \quad (3.4)$$

Here, the $E_{gas,x}$ terms are the electronic energies of the dimer (dim) and the two monomers (mon1 and mon2). The $G_{rrho,x}$ terms are the free energy corrections from Grimme's quasi-rigid-rotor-harmonic-oscillator approximation.[120] G_{rrho} terms are commonly computed in the gas phase, but in cases where the minimum energy geometry changes appreciably upon solvation, these should be determined in solution instead.[123] The $\Delta G_{solv,x}$ terms are the Gibbs free energies of solvation for each species, usually obtained using a continuum solvation model, such that

$$\Delta G_{solv.x} = E_{solv.x} - E_{gas,x} + \Delta G_{conc,x}. \quad (3.5)$$

In Eq. 3.5, $E_{solv.x}$ is the energy of species x in solution, at its solution-optimized geometry, and using continuum solvation, while $E_{gas,x}$ is the gas-phase energy for species x, determined at its gas-phase optimized geometry. Electronic structure programs typically use 1 atm as the standard state, but computations in solution should use a 1M standard state. This change in concentration, and thus in free energy, requires adding a correction term:[124]

$$\Delta G_{conc} = -RT \ln \frac{V_f}{V_i}. \quad (3.6)$$

Here, V_f and V_i are the final and initial molar volumes, respectively. When compressing from 1 atm to 1 M at 298.15 K, ΔG_{conc} has the approximate value of 7.9 kJ/mol. This value needs to be added manually to the value produced by the electronic structure program to arrive at a correct value for ΔG_{solv} . It is also convenient to define the term $\Delta E_{solv,dim}$:

$$\Delta E_{solv.dim} = E_{solv.dim} - E_{solv.mon1} - E_{solv,mon2}. \quad (3.7)$$

Since electronic structure programs typically report E_{solv} as the result of a computation in solution, and two of the ΔG_{solv} terms cancel, Eq. (1) can be expanded to:

$$\Delta G_{dim} = E_{solv.dim} - E_{solv,mon1} - E_{solv.mon2} + RT \ln \frac{V_f}{V_i} + G_{rrho,dim} - G_{rrho,mon1} - G_{rrho,mon2}. \quad (3.8)$$

In Eq. 3.8, all the terms for computing ΔG_{dim} are directly available from the output of quantum chemistry geometry optimizations and frequency determinations. For ensembles in the gas phase, we will use the term $\Delta G_{gas,dim}$, where

$$\Delta G_{gas,dim} = \Delta E_{gas,dim} + \Delta G_{rrho,dim}$$

$$= E_{gas,dim} - E_{gas,mon1} - E_{gas,mon2} + G_{rrho,dim} - G_{rrho,mon1} - G_{rrho,mon2}. \quad (3.9)$$

The final effect that needs to be considered to obtain correct values for dimerization energies is that there are frequently several different conformations that a dimer (or its monomers) may take in solution. If more than one such conformation is thermodynamically accessible, then the correct value to take as the free energy of dimerization is the ensemble-averaged free energy of dimerization:

$$\Delta G_{dim,ens} = \sum_i \left(p_i \times G_{dim,i}\right) - T \times S_{conf}. \quad (3.10)$$

In Eq. 3.10, $\Delta G_{dim,i}$ is the free energy of dimerization of the i-th conformer, p_i is the probability for the i-th conformer, T is the absolute temperature, and S_{conf} is the conformational entropy, that is, the entropy which results from populating different conformations. S_{conf} and p_i are given by[61]

$$p_i = \frac{e^{-\Delta G_{dim,i}/RT}}{\sum_i e^{-\Delta G_{dim,i}/RT}} \quad (3.11)$$

and

$$S_{conf} = -R \times p_i \times \ln p_i. \quad (3.12)$$

In this work, DLPNO-CCSD(T)[55, 125-127]/cc-pVTZ[121, 128, 129]/SMD(Toluene)[119, 130] computations at the ωB97X-D4[116]/def2-TZVP[117, 129]/CPCM(Toluene)[118] geometries are used to obtain the most reliable values, using the *TightPNO*[131] cutoff setting, in the ORCA 5.0.1 code.[132] The default resolution of the identity-chain of spheres density fitting approximation (RIJCOSX) was used for both DFT and DLPNO-CCSD(T).[133] However, because of the computational expense, DLPNO-CCSD(T) single point calculations have only been performed in solution and in the gas phase for the most interesting conformers of each dimer (*i.e.*, both the lowest energy conformer, and the one that best corresponds to the crystal structure, where the two are not the same). Except where otherwise specified, energies reported in this paper were computed using density functional theory (ωB97X-D4), rather than DLPNO-CCSD(T). Even where DLPNO-CCSD(T) is indicated, the electronic energy and solvation energy have been determined at this level, but the G_{rrho} term and all properties relating to the whole ensemble ($S_{conf}, p_i,\ \Delta G_{dim,ens}$) were determined using the DFT energies.

Initial geometry explorations were performed using CREST v. 2.10.1,[61] which uses the GFN2-xTB[134] density tight-binding semi-empirical functional. CREST requires a constraining potential for

supramolecular assemblies to prevent the components from dissociating during metadynamics. Rather than using the constraining potential built into the NCI mode of CREST, a spherical *logfermi* constraining potential was added to the standard workflow of CREST. This was done because test computations appeared to show that important conformers were being missed in NCI mode. The radius of the constraining potential was chosen to be 17 Bohr, which was found to be large enough to allow the two molecules to move past each other, but without allowing them to dissociate. The standard workflow in CREST includes a genetic crossing (GC) algorithm, which can increase the diversity of the ensemble, but when used on multiple molecules, this algorithm can cause CREST to crash,[135] so we turned it off using the *-nocross* keyword.

All conformers within 0.5 kcal/mol (2.1 kJ/mol) of the most stable conformer in the CREST output were then reoptimized using ORCA 5.0.1, at the ωB97X-D4/def2-TZVP level of theory, with CPCM(Toluene) solvation. Additionally, the remaining conformers from the CREST output were visually examined for unique conformers, and all unique conformers found were also reoptimized in ORCA. The solvation model based on density (SMD)[119] gives a more nuanced representation of solvation than the conductor-like polarizable continuum solvation model (CPCM), but analytical frequencies are not available in ORCA for SMD. Therefore, geometries and the G_{rrho} term were computed using analytical frequencies and CPCM, followed by a single point calculation at the CPCM geometry using SMD. Thus, E_{solv} is taken from the SMD computation, and G_{rrho} is taken from the CPCM computation.

Geometry optimizations were performed using the *TightSCF* convergence level, along with *TightOpt* geometry convergence. For each conformer optimized, all frequencies were examined. When any were found to be imaginary, geometry convergence thresholds were tightened, and the conformer was reoptimized. Where this was still insufficient, the imaginary mode was followed to a new starting structure, which optimized to a true minimum.

There is reason for concern that the optimal geometry using SMD may be different from the optimal geometry using CPCM solvation. Ideally, SMD would have been used for all computations in solution, but the lack of an analytical Hessian when using SMD made the calculation of frequencies impractical. Instead, single point energies with SMD at the CPCM optimal geometry were used, along with frequencies calculated using CPCM. To make sure that any differences in geometry and energy between SMD and CPCM would be minimal, 15 conformers, selected from all four dimers, were selected, and reoptimized using SMD. Among these 15 conformers, the average RMSD between the CPCM and SMD geometries was 0.057 Å, with a maximum RMSD of 0.143 Å. Similarly, the difference in

energy between the SMD single point calculations at the CPCM minimum and the SMD minimum averaged -0.46 kJ/mol, with a maximum difference of -1.22 kJ/mol. Thus, it can be concluded that the differences between the SMD and CPCM geometries are small, and the difference in energies is well within the expected errors.

Due to the symmetry of the monomers, there are many ways to put two molecules together to arrive at each conformer of each dimer. As an example, for the 60G1 conformer of $(pyrene)_2$, there are 64 distinct ways two pyrene molecules can stack together to make the conformer: We can define the nearer molecule (bolded in Figure 3.1) to be molecule A, and the further molecule (coloured blue in Figure 3.1) to be molecule B. Molecule A and molecule B can each be considered to have an arbitrarily-defined top face and bottom face. Because pyrene is symmetric, either the top or bottom face on each molecule is free to be the face which approaches the other molecule, leading to four stacking arrangements. For each of these stacking arrangements, molecule A has one end ring that does not overlap with the rings of molecule B, and one end ring which is centred on one of the quaternary carbons of molecule B. The long axes of the molecules are oriented at 60° to each other, either in a clockwise or counter-clockwise direction. Since molecule A has two end rings, and molecule B has two quaternary carbons, and there are two options for axial rotation, there are 2 × 2 × 2 = 8 arrangements possible. Finally, since A and B occupy inequivalent positions in the dimer, one can double the number of possible conformations by exchanging them for each other. This makes 4 × 8 × 2 = 64 total distinct ways of arranging two pyrene molecules into a 60G2 dimer, for a 64-fold degeneracy of this dimer. Boltzmann weights for the various conformers of each ensemble were calculated using ΔG_{dim} for each dimer, and p_i and S_{conf} were determined taking the manually counted degeneracy into account.

3.3 Results and Discussion

The geometries and energies of the conformers that make up the ensemble for the pyrene homodimer in toluene solution are first described and discussed. Subsequent discussions focus on the ensembles in toluene solution for the pyrene-4,5-dione homodimer, the pyrene-4,5,9,10-tetraone homodimer, and the pyrene/pyrene-4,5,9,10-tetraone heterodimer.

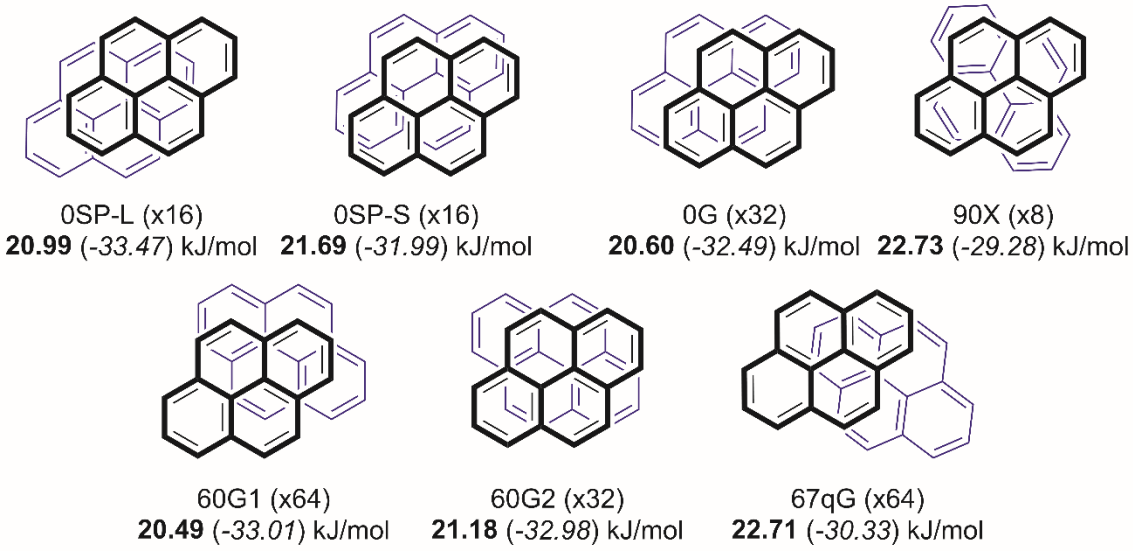
3.3.1 Pyrene Homodimer

For $(pyrene)_2$, the CREST output gives 73 geometries within a 25 kJ/mol energy range. Twenty-five of these geometries occur in the lower 4.6 kJ/mol, while the rest occur in the range of 13-25 kJ/mol above the most stable conformation. This result leaves a curious gap, with no conformations reported in the energy range 4.6-13 kJ/mol above the most stable conformer. When compared with subsequent DFT

computations, these results, while they do an excellent job of exploring the different possible geometries, do not order the energies particularly well, and also find many spurious local minima. Additional discussion of the relationship between CREST and DFT structures, including energetics, is provided in Appendix B. Because of these differences, the ensembles from CREST need to be reoptimized using DFT to obtain more accurate results. Rather than reoptimize the entire ensemble, we reoptimized all the structures in the lowest 2.1 kJ/mol range, and any other structures from the CREST ensemble that appeared to be unique upon visual inspection.

For $(pyrene)_2$, seven unique conformers were found, see Figure 3.1. Five of these conformers (0SP-L, 0SP-S, 0G, 60G2 and 90X) have been previously reported,[54, 87, 102, 115] but the CREST routine allows a very thorough search of the conformational space to be performed, which has resulted in two new low-lying conformers being identified. Both new conformers have an angle between the long axes of approximately 60 degrees. In contrast to Rapacioli *et al*.,[115] all the T-shaped structures generated by

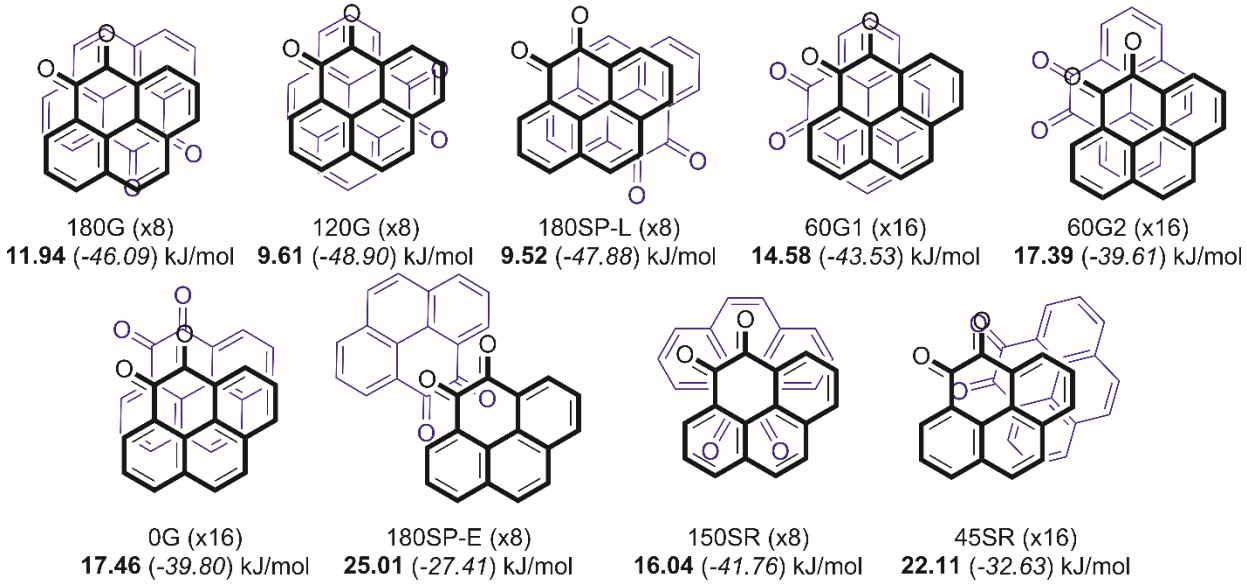
Figure 3.1 Geometries of the $(pyrene)_2$ ensemble, determined at the ωB97X-D4/def2-TZVP/CPCM(Toluene) level of theory. Degeneracies are given in parentheses, with **Gibbs free energies of dimerization** ($\Delta G_{solv,dim}$) in bold, and *binding energies* ($\Delta E_{solv,dim}$) in italics. Degeneracies were determined as discussed in the methods section.



CREST relaxed to one of the seven cofacial structures when using DFT. To accommodate these new structures, we have expanded the naming system to include the angle between the long axes. Thus, the previously published[54, 102] global minimum structure, SP-L (slipped-parallel along the long axis),

becomes 0SP-L. The nearly isoenergetic SP-S (slipped-parallel along the short axis) and G (graphitic stacking) conformers become 0SP-S and 0G, while the somewhat higher energy X (crossed axes) conformer becomes 90X. The 60G1 conformer (C_1 symmetry) and the 60G2 conformer (C_2 symmetry) are nearly isoenergetic with the 0PD-L conformer, being only 0.46 kJ/mol and 0.49 kJ/mol higher in energy, respectively. The 67qG (C_1 symmetry) conformer employs *quasi*-graphitic stacking, with the molecules slightly twisted away from an optimal graphitic position. It sits at 3.14 kJ/mol above 0PD-L, and slightly below 90X. All geometries tested from the CREST output converged to one of these seven geometries at the ωB97X-D4/def2-TZVP/CPCM(Toluene) level of theory. Of these seven conformers, the 0PD-L conformer (0° between the long axes, parallel-displaced along the long axis) corresponds best to the geometry of nearest neighbours in the crystal structure,[98] and has the lowest energy.

Figure 3.2 Geometries of the (pyrene-4,5-dione)$_2$ ensemble, determined at the ωB97X-D4/def2-TZVP/CPCM(Toluene) level of theory. Degeneracies are given in parentheses, with **Gibbs free energies of dimerization** ($\Delta G_{solv,dim}$) in bold, and *binding energies* ($\Delta E_{solv,dim}$) in italics. Degeneracies were determined as discussed in the methods section.



Taking the whole ensemble together, the conformational entropy, S_{conf} , has the value of 13.93 J/(mol K), and $\Delta G_{dim,ens}$ is then 16.94 kJ/mol. See Appendix B for details. This positive value indicates that the dimer is unstable in solution, relative to the monomers, despite the significant stabilizing binding energy. This instability can be attributed to both a loss in surface area available for interactions with solvent molecules, and a loss of translational entropy upon dimerization. Thus dimerization is

unfavourable, and, at least under the first-order approximation of taking the dimerization energy as a proxy for lattice energy, pyrene should completely dissolve in toluene at room temperature, which is borne out by the fact that pyrene does dissolve in toluene, up to a mole fraction of 0.067.[136]

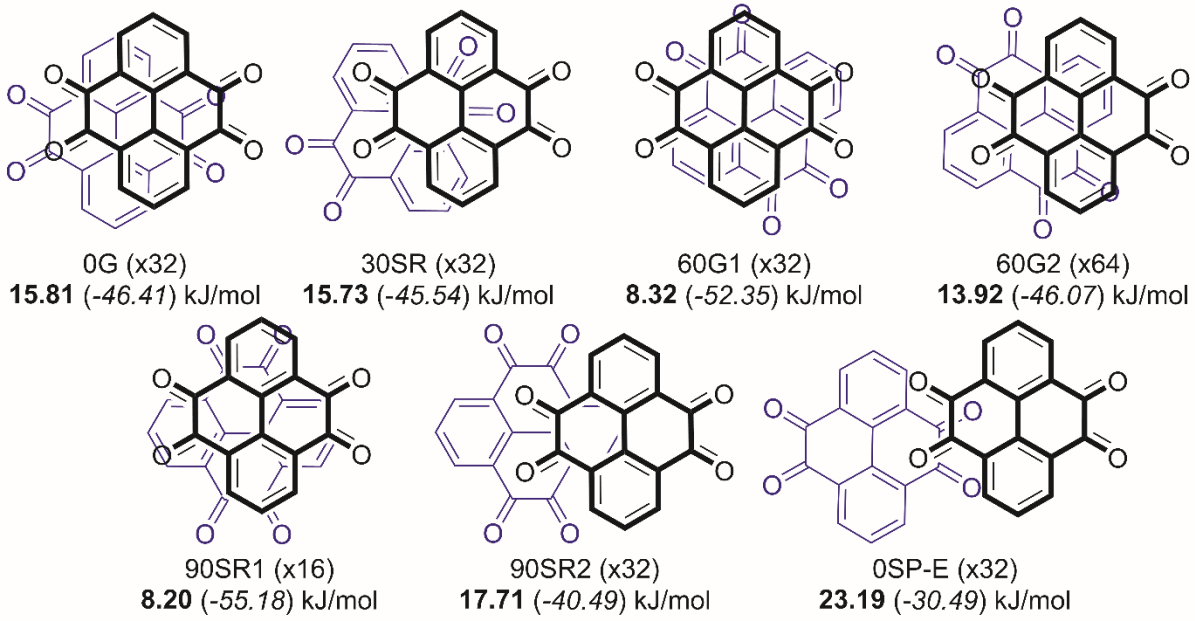
The gas phase ensemble was also determined, primarily for the purpose of comparison to existing literature. Details are in Appendix B. Our results agree well (within 2 kJ/mol) with previously published results.[54, 102, 115]

3.3.2 Pyrene-4,5-dione Homodimer

For (pyrene-4,5-dione)$_2$, DFT reoptimization of the low-lying portion of the CREST ensemble yields nine conformers, see Figure 3.2. Two conformers, 180SP-E (slipped-parallel-extended) and 45SR (slipped and rotated) occur in the low-lying part of the CREST ensemble (3.93 and 3.45 kJ/mol above the minimum, respectively), but after reoptimization using ωB97X-D4/def2-TZVP/CPCM(Toluene), they are much higher in relative energy, at 21.49 and 16.27 kJ/mol, respectively. This is a good demonstration of how the CREST potential energy surface is not always parallel to that of DFT, and underlines why reoptimization is necessary. The 180G conformer (C_i symmetry) corresponds best to the crystal structure,[99] but is not the lowest-energy conformer at the ωB97X-D4 level of theory, as 120G (C_2 symmetry) is 2.81 kJ/mol lower in energy. The 180SP-L conformer (C_i symmetry) is also favourable, at 1.79 kJ/mol lower in energy than 180G. There are two conformers with the long axes at 60 degrees to each other. 60G1 (C_1 symmetry) is 2.56 kJ/mol above 180G, while 60G2 (C_1 symmetry), which has a carbonyl oxygen on one molecule interacting with a carbonyl carbon on the other, is 6.48 kJ/mol above. The 0G conformer (C_1 symmetry) also has a carbonyl oxygen on one molecule interacting with a carbonyl carbon on the other, and is 6.29 kJ/mol above 180G. The 180SP-E conformer (C_i symmetry) has its long axes antiparallel like 180G, minimizing the total dipole of the dimer, but has the carbonyls stacked on top of each other, which prevents the main bodies of the ring systems from interacting with each other. This loss of π–π stacking opportunities results in this being the highest energy conformer by far, at 18.68 kJ/mol higher than 180G. 150SR (C_2 symmetry) has the carbonyl-bearing rings directly superimposed on each other, with dipoles nearly opposed, but with a 30 degree twist away from perfectly opposed, in a manner that minimizes close approaches between atoms, for an energy 4.33 kJ/mol above 180G. Finally, 45SR (C_1 symmetry) has only moderate overlap between ring systems, and nearly aligned dipoles, for a fairly high energy: 13.46 kJ/mol above 180G.

As mentioned above, the 180G conformer best corresponds to nearest neighbours in the crystal structure,[99] but 120G is the lowest energy conformer in solution. Inclusion of the rovibrational and concentration terms give Gibbs free energies of dimerization of 11.94 kJ/mol for 180G, and 9.61 kJ/mol for 120G. Taking the Boltzmann-weighted average free energy of all conformers, and adding the conformational entropy term, gives a free energy of dimerization of 8.39 kJ/mol. Gas phase results are available in Appendix B.

Figure 3.3 Geometries of the (pyrene-4,5,9,10-tetraone)$_2$ ensemble, determined at the ωB97X-D4/def2-TZVP/CPCM(Toluene) level of theory. Degeneracies are given in parentheses, with **Gibbs free energies of dimerization** ($\Delta G_{solv,dim}$) in bold, and *binding energies* ($\Delta E_{solv,dim}$) in italics. Degeneracies were determined as discussed in the methods section.



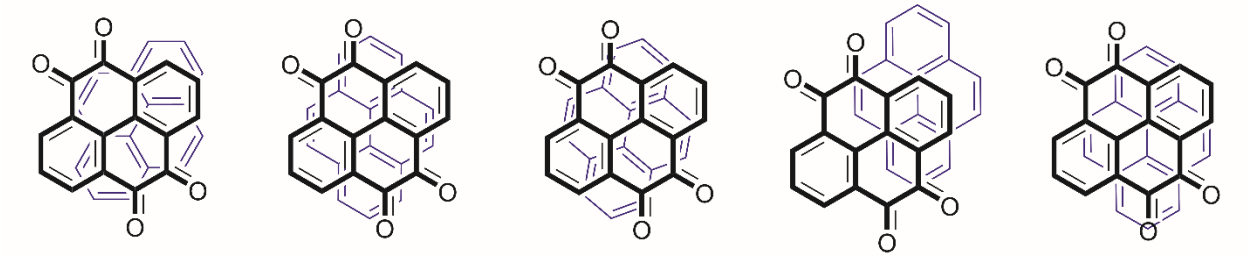
3.4.3 Pyrene-4,5,9,10-tetraone Homodimer

For (pyrene-4,5,9,10-tetraone)$_2$, reoptimization of the low-lying portion of the CREST ensemble yields seven conformers, see Figure 3.3. The 0G conformer (C_i symmetry) best approximates the crystal structure,[100] but is not the lowest in energy. Due to the extra carbonyls in pyrenetetraone relative to pyrenedione, this conformer can be considered analogous to both 180G and 0G of pyrene-4,5-dione. The 30SR conformer (C_2 symmetry) is analogous to the 150SR conformer of pyrene-4,5-dione, and is nearly isoenergetic with 0G, being only 0.87 kJ/mol higher in energy. There are two conformers with long axes at a 60 degree angle. 60G1 (C_2 symmetry) is analogous to both 60G2 and 120G of pyrene-4,5-

dione, and is 5.94 kJ/mol lower in energy than 0G, while 60G2 (C_1 symmetry) is analogous to 60G1 of pyrene-4,5-dione, and is only 0.34 kJ/mol above 0G. The 90SR1 conformer (C_1 symmetry) turns out to be the global minimum, at 8.77 kJ/mol below 0G, while the 90SR2 conformer (C_s symmetry) is significantly higher in energy, at 5.92 kJ/mol above 0G. Finally, the 0SP-E conformer (C_i symmetry) is analogous to 180SP-E of the dione dimer, and is by far the highest energy conformer, at 15.92 kJ/mol above 0G.

The 0G conformer, which best corresponds to the crystal structure,[100] and the 90SR1 conformer, which has the lowest energy in solution, have dimerization energies of -46.41 kJ/mol and -55.18 kJ/mol, respectively. Including the rovibrational and concentration terms, the ΔG_{dim} is 15.81 kJ/mol for 0G, and 8.20 kJ/mol for 90SR1. Ensemble averaging and conformational entropy bring the total free energy of dimerization for pyrene-4,5,9,10-tetraone to 6.49 kJ/mol. Gas phase results are available in Appendix B.

Figure 3.4 Geometries of the heterodimer ensemble, determined at the ωB97X-D4/def2-TZVP/CPCM(Toluene) level of theory. Degeneracies are given in parentheses, with **Gibbs free energies of dimerization** ($\Delta G_{solv,dim}$) in bold, and *binding energies* ($\Delta E_{solv,dim}$) in italics. Degeneracies were determined as discussed in the methods section.



3.3.4 Pyrene/Pyrene-4,5,9,10-tetraone Heterodimer

For the heterodimer of pyrene and pyrene-4,5,9,10-tetraone, reoptimization of the CREST ensemble produces only five conformers, see Figure 3.4. This ensemble is different from the others in that the angle between the long axes is always somewhere between 30 and 60 degrees. There is no known crystal structure for the cocrystal, but the lowest energy dimer in solution is the 30SR conformer. The long axes are rotated 30 degrees relative to each other, and the pyrene molecule is translated slightly to one side, such that its centre point lies directly under one of the quaternary carbons of the pyrenetetraone. The 60X and 50X conformers are nearly isoenergetic with each other and with the minimum energy geometry, being only 0.83 kJ/mol and 1.01 kJ/mol above 30SR, respectively. The 60G1

and 60G2 conformers are slightly higher in energy, at 1.98 kJ/mol and 2.57 kJ/mol above 30SR, respectively, but this is still a very small energy window, barely half the size of the 4 kJ/mol generally considered to be “chemical accuracy.” Thus, we have a small ensemble of relatively similar structures, within a very small energy window for the heterodimer.

The 30SR conformer is the most stable, and has an interaction energy of -53.01 kJ/mol. Adding in the rovibrational and concentration terms gives a free energy of dimerization of 6.84 kJ/mol, and ensemble averaging plus the conformational entropy bring the corrected free energy of dimerization to 4.13 kJ/mol, which is the most stable of any of the dimers investigated. Gas phase results are available in Appendix B.

A breakdown of the components contributing to the dimerization free energy is available in Appendix B. These results appear to indicate that the gains in interaction energy that occur with oxidation are primarily due to the $\Delta E_{SCF,gas,dim}$ term (the electronic energy), with a small amount of additional stabilization for the heterodimer coming from a less unfavourable $\Delta\Delta G_{solv,dim}$ (change in solvation free energy).

Given that all investigated dimers are unstable relative to their monomers, one would expect pyrenedione and pyrenetetraone, as well as the putative pyrene/pyrenetetraone complex, to dissolve in room temperature toluene, but this is not observed (see Chapter 2). Instead, we were able to recrystallize pyrenedione from toluene, with only minimal amounts remaining in solution, and the crystals thus formed were contaminated with equal amounts of pyrene and pyrenetetraone. From this, it appears that the pyrene/pyrenetetraone complex is also (nearly) insoluble in room temperature toluene, and pyrenetetraone may be as well. To reconcile these observations with the computational results, we need to revisit our simplifying approximation. While π–π stacking between nearest neighbours should be the primary interaction between molecules in a crystal, it will not be the only interaction, and the additional interactions in three dimensions must be enough to overcome the (weakly) endergonic nature of dimerization. It is also important to remember that continuum solvation models are also approximate, and do not take specific solute-solvent interactions, such as the π–π interaction between toluene and pyrene molecules, into account. Unfortunately, computations that would properly take those additional intermolecular interactions into account, both within the crystal, and in solution, are out of reach using the methods in this paper. However, the relative stability of the different dimers should still give an approximate ranking of the relative stability of the various crystals.

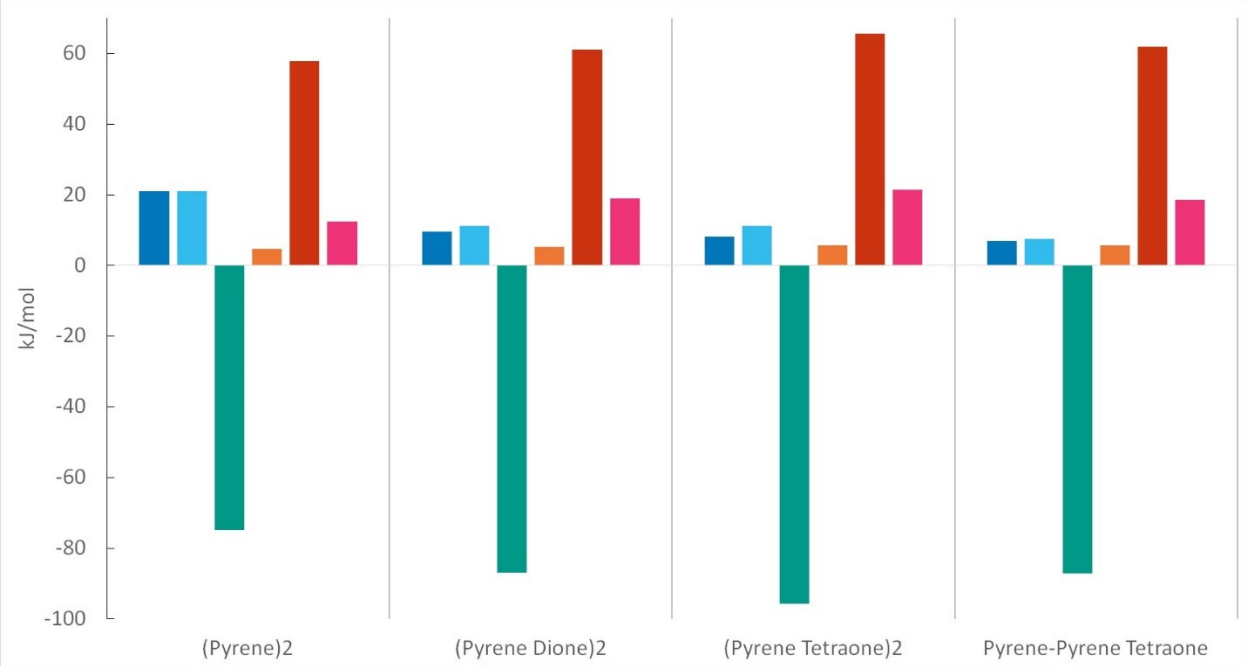
Table 3.1 compares the results using DLPNO-CCSD(T) to those using ωB97X-D4. Note that the differences between the methods are small, with a mean error of 1.47 kJ/mol (the MUE was also 1.47 kJ/mol, as all DFT energies were higher than DLPNO-CCSD(T) ones), and a maximum error of 2.19 kJ/mol, well within chemical accuracy. This indicates that ωB97X-D4 is an appropriate and accurate method for these kinds of systems, though with what appears to be a small systematic error. LED of the DLPNO-CCSD(T) interaction energy was performed, but no meaningful trends were found. The LED results are available in Appendix B.

Table 3.1 ωB97X-D4/def2-TZVP/SMD(Toluene) and DLPNO-CCSD(T)/cc-pVTZ/SMD(Toluene) energies of dimerization of selected conformers. Difference is computed as $E_{DFT} - E_{DLPNO}$. %Difference is computed as $[(E_{DLPNO} - E_{DFT})/E_{DLPNO}]*100\%$.

	$\Delta E_{solv,dim}$ (DFT) (kJ/mol)	$\Delta E_{solv,dim}$ (DLPNO) (kJ/mol)	Difference (kJ/mol)	% Difference
$(pyrene)_2$ 0SP-L	-35.00	-33.47	1.53	4.6%
$(pyrenedione)_2$ 180G	-47.94	-46.09	1.85	4.0%
$(pyrenedione)_2$ 120G	-50.17	-48.90	1.27	2.6%
$(pyrenetetraone)_2$ 0G	-48.27	-46.41	1.86	4.0%
$(pyrenetetraone)_2$ 90SR	-57.37	-55.18	2.19	4.0%
pyrene/pyrenetetraone 30SR	-53.14	-53.01	0.13	0.2%

When computations are done at less than the basis set limit, basis set incompleteness error (BSIE) will affect the calculations. One of the more problematic facets of BSIE is that when computing a dimerization, the monomers can ”borrow” basis functions from adjacent monomers to help resolve their own BSIE. This effect is known as basis set superposition error (BSSE). The larger the basis set used, the smaller both BSIE and BSSE will be. One common solution to BSSE is the counterpoise correction, where an atom’s nucleus and electrons are removed, but its basis sets remain, to gauge the level of ”borrowing,” however, counterpoise tends to overestimate BSSE. Interestingly, local correlation methods, such as DLPNO, do not suffer nearly as much from BSSE, and the counterpoise correction is unnecessary, and even counterproductive for such methods.[137] We did not expect BSSE to be a significant problem when using def2-TZVP, but at the request of a reviewer, we performed counterpoise corrections on the same subset of six conformers. The counterpoise correction ranged from 2.1 to 3.7 kJ/mol, with the larger corrections in conformers with larger binding energies. In all cases, this overcorrected for the difference between ωB97X-D4 and DLPNO-CCSD(T), leading to a slightly worse MAD of 1.92 kJ/mol, so counterpoise corrections were not further pursued for other conformers.

Figure 3.5 Components of ΔG_{dim}. From left to right, total ΔG_{dim} is shown in blue, $\Delta E_{SCF,gas,dim}$ is shown in cyan, $\Delta E_{disp,gas,dim}$ in teal, $\Delta G_{rrho,H,dim}$ in orange, $\Delta G_{rrho,S,dim}$ in red, and $\Delta\Delta G_{solv,dim}$ in magenta.



3.3.5 Energy Comparison and Breakdowns

The free energy of interaction, as described in the methods section, can be broken down into the changes in electronic energy, solvation energy, and quasi-rigid-rotor-harmonic-oscillator (rrho) energy. The rrho energy changes can be further subdivided into those due to changes in enthalpy, and those due to changes in entropy. Entropy will always decrease significantly upon dimerization, due to the loss of translational degrees of freedom, leading to a positive change in free energy. For a dispersion-corrected functional, such as ωB97X-D4, the electronic energy can be further subdivided into the SCF energy and the dispersion energy. Thus, Equation 3.1 can be expanded to

$$\Delta G_{dim} = \Delta E_{SCF,gas,dim} + \Delta E_{disp,gas,dim} + \Delta G_{rrho,H,dim} + \Delta G_{rrho,S,dim} + \Delta\Delta G_{solv.dim}. \quad (3.13)$$

Figure 3.5 shows the breakdown in energy for the lowest-energy conformer of each of the four dimers. Numerical results are given in Appendix B. The leftmost (blue) column shows Gibbs free energy of dimerization. It is clear that dimerization for the homodimers becomes less unstable with oxidation, and that the heterodimer is even more stable. It is also immediately apparent that the most stabilizing component for all dimers is dispersion, and the most destabilizing is entropy, as should be expected. The pyrene dimer has the most positive ΔG_{dim}, which appears to be largely due to an increase in

$\Delta E_{gas,SCF,dim}$ relative to the other dimers – that is, the electronic energy is less favourable. Dispersion is also weaker here, but that is mostly offset by slightly less unfavourable entropy and solvation terms. The pyrene-pyrene tetraone heterodimer has the most favourable (though still positive) ΔG_{dim}, and this appears to be due primarily to slight improvements in $\Delta E_{gas,SCF,dim}$ and $\Delta\Delta G_{solv,dim}$, relative to the pyrene dione and pyrene tetraone dimers.

3.4 Conclusion

In this work, we have generated and analyzed the low-lying portion of the conformational ensembles of the $(\text{pyrene})_2$, $(\text{pyrenedione})_2$, $(\text{pyrenetetraone})_2$, and pyrene/pyrenetetraone dimers, both in the gas phase and in solution. Our results show that the CREST program[61] and the associated GFN2-xtb[134] approach are useful tools for exploring potential energy surfaces of weakly bound dimers. However, care must be taken in interpreting the results, as the geometries CREST generates for dimers bound by π –π stacking are of uncertain quality, spurious minima are common, and errors in relative energies can be as great as 30 kJ/mol. On the other hand, we have found that, for these systems, the agreement between ωB97X-D4[116]/def2-TZVP[117] and DLPNO-CCSD(T)[55]/cc-pVTZ[121] energies is excellent, though with a tendency to slightly underbind with ωB97X-D4. We have found two conformers of the $(\text{pyrene})_2$ dimer that had not been previously reported, and reported the ensembles of three other dimers which have not been previously studied. In comparing the stacking interactions of pyrene *vs.* those of its oxides, it appears that oxidation makes available a wider range of geometries, and increases the strength of intermolecular interactions. Our final reported Gibbs free energies of dimerization are 16.94 kJ/mol for $(\text{pyrene})_2$, 8.39 kJ/mol for $(\text{pyrene-4,5-dione})_2$, 6.49 kJ/mol for $(\text{pyrene-4,5,9,10-tetraone})_2$, and 4.13 kJ/mol for the pyrene/pyrene-4,5,9,10-tetraone heterodimer. This indicates that none of the dimers are stable in toluene solution, but the energies are small enough that it is likely inclusion of secondary intermolecular interactions in the crystal form would be enough to make crystals stable. Our findings give weight to the suggestion that pyrene and pyrenetetraone can form a cocrystal, potentially explaining the difficulty in separating them from pyrenedione.

Chapter 4

4. A Variant on the CREST Algorithm for Non-covalent Clusters of Flexible Molecules

4.1 Introduction

The search for low-energy conformers of flexible molecules or clusters of molecules is a central problem in computational chemistry.[61, 138-142] A geometry optimization of a flexible molecule from a randomly generated initial structure will not usually find the global minimum. Even when the global minimum can be found, the full ensemble is often important for many applications, including, but not limited to, ligand docking,[138] the determination of conformational free energy,[61] and predicting spectra of flexible molecules, particularly for vibrational circular dichroism (VCD)[143-146] and NMR.[147] Some applications, such as ligand docking, require only that the ensemble is diverse and somewhat representative of the true ensemble in solution. Others, such as predicting vibrational circular dichroism spectra[143, 144] or free energy of association,[61, 147] require a diverse ensemble which is correctly weighted by energy, and includes the global minimum conformer.

Many algorithms exist for conformational exploration, divided into two main classes: (i) systematic and (ii) stochastic. Systematic algorithms explore all possible combinations of dihedral torsions in a molecule, usually in set increments, *e.g.*, every 15°. This approach makes them deterministic and thorough, but unless the conformers thus generated are reoptimized, will only give approximations to the true minima. Also, systematic algorithms can only be used for systems with a small number of rotatable bonds, as the number of possible conformers increases extremely quickly with increasing rotatable bond count.[138, 140] Stochastic algorithms attempt to generate a representative ensemble by selecting conformations in a probabilistic fashion. Two methods to accomplish this goal are knowledge-based algorithms and simulations. In a knowledge-based algorithm, structures are generated by randomly assigning values to dihedral torsions that match known low-energy conformations of similar dihedrals in other systems. There are a great variety of implemented knowledge-based algorithms for conformer ensemble generation, but some examples include ConfGen,[148] OMEGA,[149] Conformator,[150] and iCon.[151] For simulations, Monte Carlo methods,[152, 153] molecular dynamics,[154, 155] or metadynamics[61] are used to sample conformations of a system. Stochastic methods can be applied

to larger systems than systematic ones can, but because they rely on inherently random processes, they are not deterministic, and may miss important conformers entirely.

Conformer exploration for a cluster of weakly interacting flexible molecules is an even more daunting task than for isolated flexible molecules. There are significantly more rotatable bonds in such a cluster than in an individual molecule, and thus exponentially more conformational space to explore. In addition, the molecules are also free to rotate and translate relative to each other, so one must ensure that they remain close enough to each other to interact. A further problem arises if there are directed interactions between the molecules, such as intermolecular hydrogen bonds. In such cases, the low-energy conformations frequently reside in a series of narrow potential wells, which can be quite difficult for an automated procedure to find. Given these challenges, even the best algorithms can struggle to produce a reasonable ensemble.

One of the most robust[156] stochastic methods for exploring conformational ensembles is the Conformer-Rotamer Ensemble Sampling Tool (CREST) program,[61] developed by the Grimme group. CREST relies heavily on the root mean squared deviation (RMSD) of geometry between conformers. The RMSD is defined as the square root of the mean of the difference in coordinates for each atom in the molecule or cluster, relative to a given reference geometry – which is effectively the distance between the conformers in a 3N-6 dimensional Cartesian conformational space. CREST uses an RMSD-biased form of metadynamics (MTD) and a GFNx-xtb (x=0,1,2) density functional tight-binding approach[134] to explore widely the conformational space of systems. The RMSD bias causes the MTD runs to constantly explore spaces dissimilar to those which they have already visited, and then snapshots from the trajectory are screened for low-energy conformers. CREST gives excellent results for single molecules, or for clusters of small or rigid molecules, and has been used as a method of choice for conformational exploration of systems,[157, 158] partly because it frequently finds many more low-energy conformers than other stochastic methods.[156] However, recent work[159] (see Chapter 5) using CREST to explore dimers, trimers, and larger n-mers containing flexible monomers demonstrated the need for algorithmic improvements within CREST to locate the minimum energy structures.

4.2 CREST Technical Details.

To help understand the need for modifications to CREST for the problems of interest, it is important to understand the mechanics of CREST and the options available within CREST for the exploration of the conformational space of dimers and n-mers. The RMSD biasing within CREST is controlled by two parameters, *kpush* and *alpha*, which control the energetics of the biasing and the

width of the Gaussian bias potential, respectively. Units are E_h (Hartree) for *kpush* and $Bohr^{-2}$ for *alpha*. A default run of CREST performs 14 MTD runs in the first cycle, and 12 in each subsequent cycle. Within each cycle, each MTD run has its own combination of *kpush* and *alpha*. Once the MTD is complete, the generated conformers from all 12 or 14 runs are collected and then screened and optimized. The screening is done in an iterative fashion, starting from very loose convergence thresholds, and proceeding to tighter ones, discarding the highest-energy conformers along the way. When one such cycle has been completed, CREST starts a new cycle from the best (*i.e.*, lowest energy) conformer found in the previous cycle. All conformers found in any cycle are retained and used in the final ensemble, as long as they are within the low-energy window (default: 6 kcal/mol above best conformer). When there is no improvement from one cycle to the next, CREST runs molecular dynamics (MD) starting from the four lowest-energy conformers and runs a genetic crossing algorithm to search for more diverse conformers. Genetic crossing exchanges motifs found in parts of one conformer with those from another conformer, potentially increasing the volume available to sampling, but as currently available in CREST, it only works as intended for single molecules, not for clusters. If a lower energy conformer is found at this point, CREST reverts to a new set of MTD runs. Otherwise, CREST terminates and gives the final ensemble.

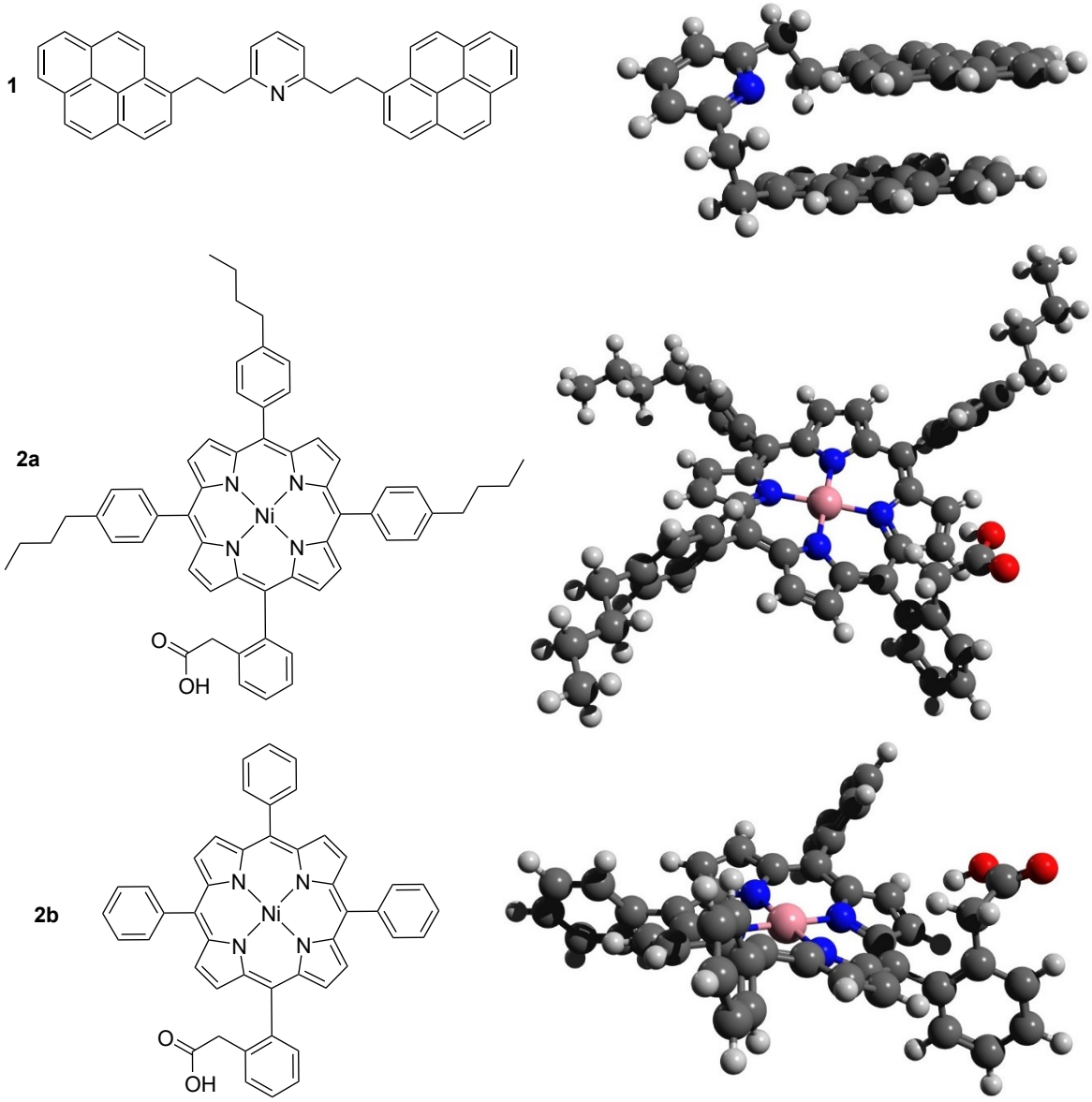
For non-covalently bound clusters of molecules, the RMSD bias has a tendency to cause clusters to dissociate, so there is a specialized non-covalent interactions (NCI) mode which uses weaker RMSD biasing (smaller *kpush*) and adds an ellipsoid constraining potential which forces the molecules to remain near each other.[61] CREST also has options for runtypes that only use part of its functionality. Two examples are the *screen* function, which will take an already generated ensemble or trajectory and sort and optimize the individual structures, while discarding those of high energy, and the *rmsd* function, which will simply compute the RMSD between two structures. Unfortunately, CREST does not contain options to adjust the number of MTD runs per cycle, nor the values of *kpush* or *alpha*. It also does not contain options to start new cycles from more than one conformer.

4.3 Results and Discussion

In a recent paper,[159] (see Chapter 5) we reported on the structure and energetics of several dimers and trimers composed of asphaltene model compounds **1**, **2a**, and **2b**, with and without microhydration. These are large, flexible molecules, with each molecule having 70-120 atoms and 4-14 rotatable bonds, and having both hydrogen bonding and π−π stacking sites, see Figure 4.1. In the course of that work, it became evident that initial geometry explorations using CREST provided an insufficient

sampling of the low-energy conformers of these systems. For example, an attempt to double-check the energy difference between two binding motifs of the trimer composed of two units of **1** and one of **2b** (the $\mathbf{1}_2\mathbf{2b}$ trimer) demonstrated that the best conformer found by CREST depended strongly on the starting structure. That is, when two different conformers of this complex were used as input for a CREST run, the output geometries and energies were significantly different. For reference, these runs were performed using *gbsa* benzene solvation, with a spherical constraining potential (size automatically generated), and MTD length of 50 ps. In the double-checking case mentioned above,

Figure 4.1 Structures of monomers. On the left are the molecular structures, and on the right are 3D representations of the DFT-optimized geometries (ωB97X-D4/def2-SVP/CPCM(Toluene) level of theory). Carbon is grey, hydrogen is white, nitrogen is blue, oxygen is red, and nickel is pink.

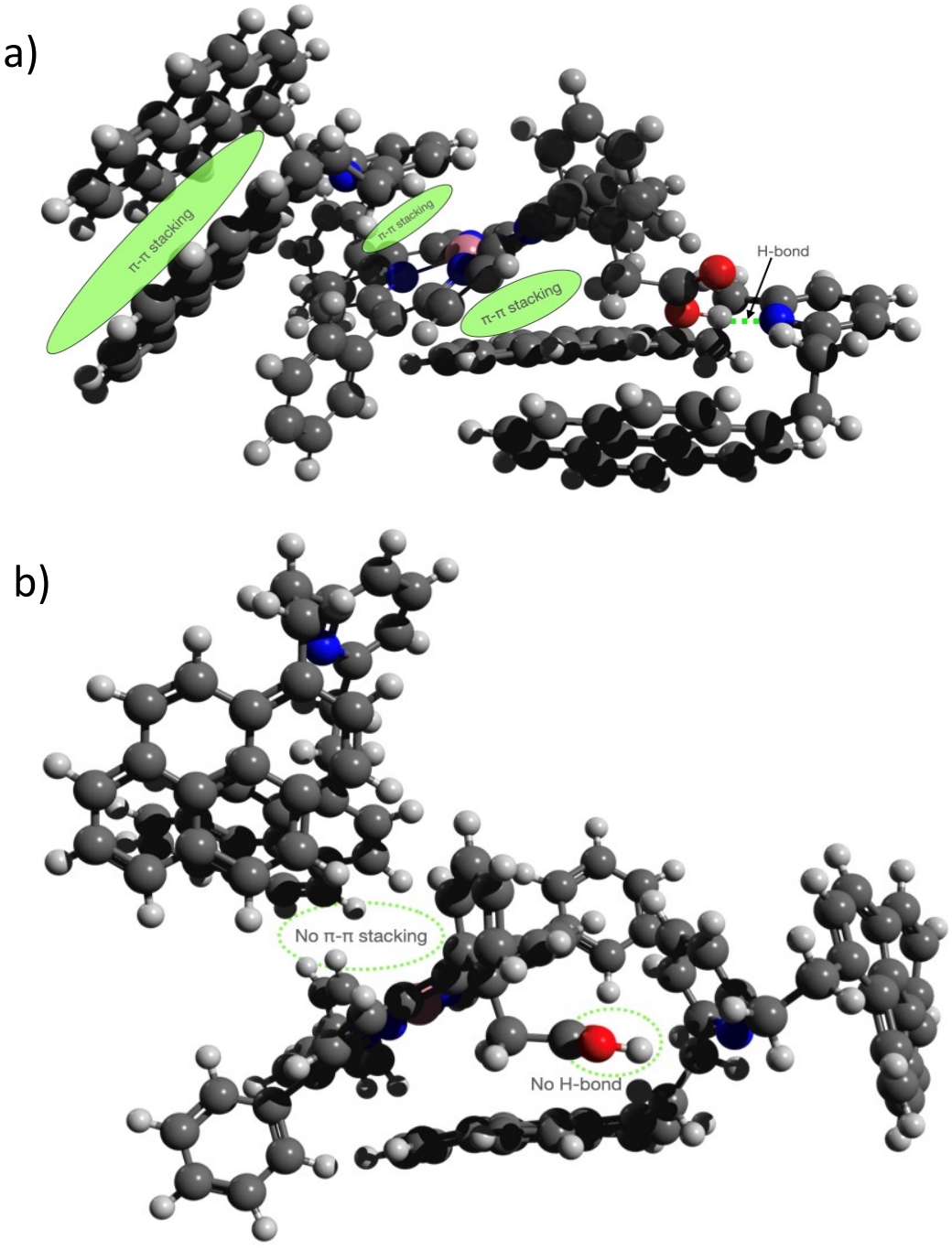


when started from a new initial structure, CREST found a conformer which was 50 kJ/mol lower in energy than anything that had been found to that point. Even starting two different CREST sampling runs from the same initial structure sometimes gave very different final results, with different lowest-energy structures having markedly different energies being found. Further investigations showed that for most of the molecular clusters under investigation, CREST had missed geometries substantially lower in energy than the lowest-energy structure it reported, demonstrating that a more thorough search algorithm was required. This is not to say that there are problems with CREST, *per se* – it consistently gives excellent results for single molecules or for complexes of small or rigid molecules – but, in the special case at hand, with complexes of highly flexible molecules (including with possible microhydration) issues with the algorithm at the heart of CREST arise. This problem was not solved by increasing the MTD length to 100 ps, by switching between NCI mode and regular CREST with a manually added constraining potential, nor by decreasing the sampling interval from 100 fs to 25 fs.

In an effort to understand the reasons for these discrepancies, the MTD trajectories were inspected. In inspecting these trajectories, it was noted that the intermolecular hydrogen bonds and π–π stacking were disrupted early in each run. See Figure 4.2 for one example of such disruption. Once disrupted, these weak interactions had trouble forming again. Thus, most of the conformers produced were of substantially higher energy than the starting conformer (greater than 12 kcal/mol difference), and were thus screened out early, and excluded from the final ensemble. Not only is this wasteful of compute time, but this preferential exploration of the higher energy space means that increasing the simulation length does not improve the results. In addition, the preferential exploration of this high-energy space during the MTD run means that the low-energy space with hydrogen bonds and π–π stacking intact is poorly explored and thus CREST cannot reliably produce better conformers. Under these circumstances, the best conformer found is usually very similar to the starting conformer. Because CREST uses only the lowest-energy conformer from a cycle as the starting point for the MTD in the next cycle, it becomes rare for CREST to find a conformer that is significantly lower in energy than the starting conformer before terminating. In some cases, better conformers are found in the MD step, but this is both unreliable and incredibly inefficient, since the MTD is then wasted, and MD is slow to explore conformers.

The twofold solution would appear to be (i) to make the RMSD biasing even less energetic than in NCI mode, and (ii) to perform restarts from a more diverse set of conformers than just the lowest-energy one. As CREST does not contain options to adjust the RMSD biasing, nor to select additional

Figure 4.2 Two snapshots from an MTD trajectory from a standard CREST run (not NCI mode, but with an autosized spherical constraining potential manually added, using *gbsa* benzene solvation). Panel a) is the starting geometry, with intact hydrogen bonding and π–π stacking highlighted to guide the eye. Panel b) shows hydrogen bonding and most π–π stacking disrupted after only 4.2 ps of simulation time. Carbon is grey, hydrogen is white, nitrogen is blue, oxygen is red, and nickel is pink.

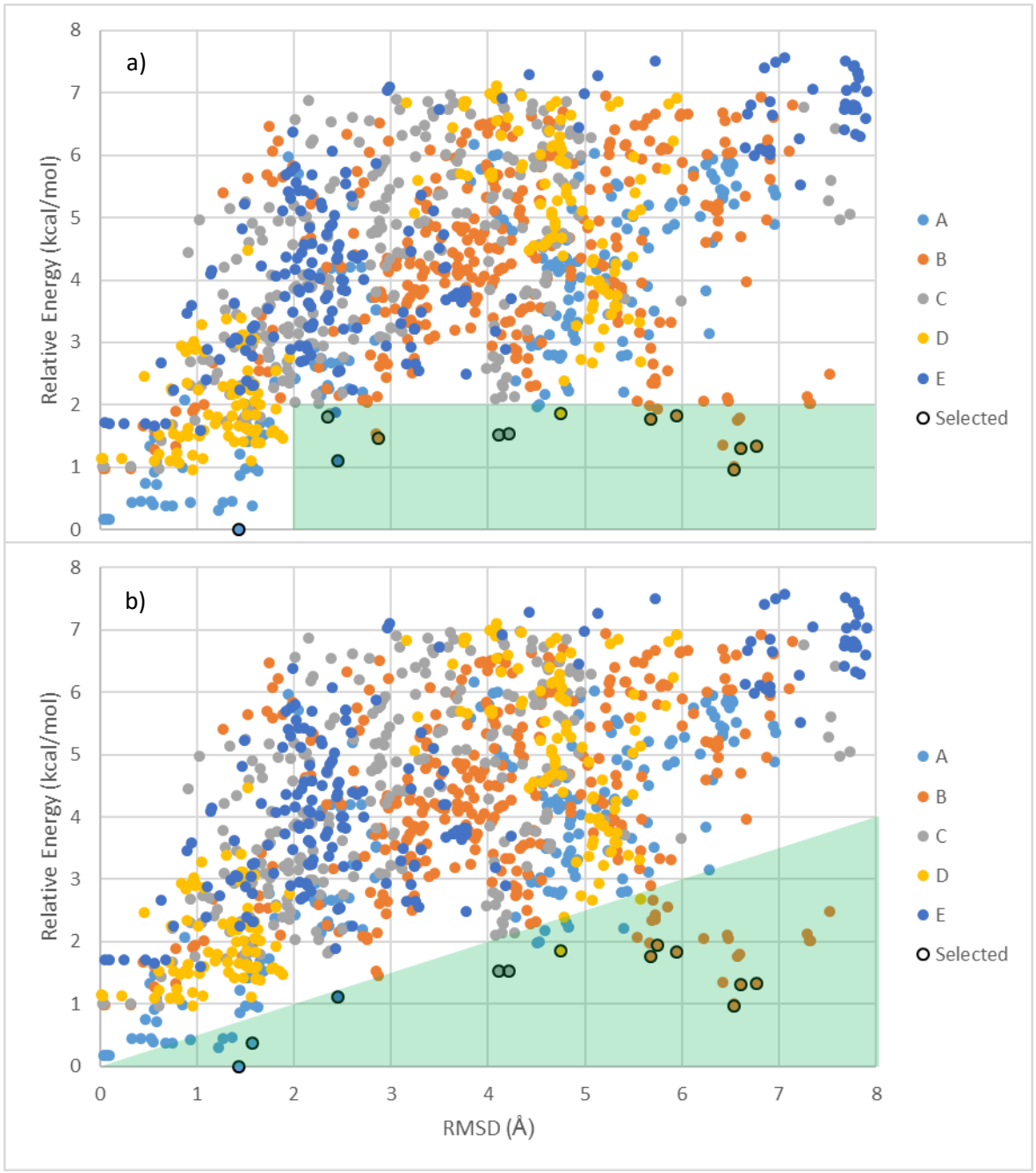


conformers to start new cycles from, a variation on the CREST algorithm which would allow the implementation of these solutions needed to be developed.

The first attempts to improve sampling consisted of running a series of MTD runs in xtb[134] with different values of the *kpush* and *alpha* settings, then processing the ensembles with the *screen* function of CREST. These runs were initiated from the lowest-energy structure provided by the initial CREST computation. CREST uses *kpush* between 0.1688 E_h and 1.125 E_h in standard runs, or between 0.1125 E_h and 0.225 E_h for NCI mode and an *alpha* between 0.1 $bohr^{-2}$ and 1.3 $bohr^{-2}$ for standard runs, or between 0.25 $bohr^{-2}$ and 1.0 $bohr^{-2}$ for NCI mode. It was already clear from the examination of trajectories that the NCI mode biasing was more energetic than desired, as evidenced by the rapid disruption of hydrogen bonds and π − π stacking. Thus, a range of *kpush* values, between 0.015 and 0.1 E_h, and a range of *alpha* values, between 0.8 and 3.1 $bohr^{-2}$ were tested. It was found that *kpush* values of 0.05 or 0.015 E_h and alpha values of 1.3 or 3.1 $bohr^{-2}$ gave the greatest number of low-energy conformers for our molecular clusters in a 30 ps run, for the case of the $\mathbf{1}_2\mathbf{2a}$ trimer. These values were then adopted for all subsequent runs. All test runs improved on the energy of the CREST structure by 11.9 kcal/mol (50 kJ/mol) or more. This result was very encouraging, but attempts to continue improving the energy by starting a new cycle from the best energy conformer found were unsuccessful. Although lower energy conformers had been found, the global minimum had not yet been located, since by starting CREST from a variety of other starting structures, at least one minimum 1.0 kcal/mol (4.3 kJ/mol) lower in energy had been located previously.

Evidently, in the rugged energy landscape of NCI, reducing the energetics of the RMSD biasing is not sufficient to reliably find the lowest-energy conformers. In addition, a diversity of starting structures are required to reliably find better minima. Starting from a collection of known low-energy structures would intuitively be preferable to random or high-energy structures, as the low-energy structures can be presumed to be more similar to the global minimum. However, it is not immediately obvious how to select the starting structures. For example, it would make little sense to choose a set of very similar structures, as they would likely cover similar areas of the potential energy surface (PES) in their simulations, and thus defeat the purpose of diversity. In the common case that the structures lowest in energy were all similar to each other, it would then be counterproductive to simply choose starting conformers based on low energy. One useful measurement in selecting conformers is the pairwise RMSD. This quantity is easily computed using CREST, and gives a good metric for how similar or

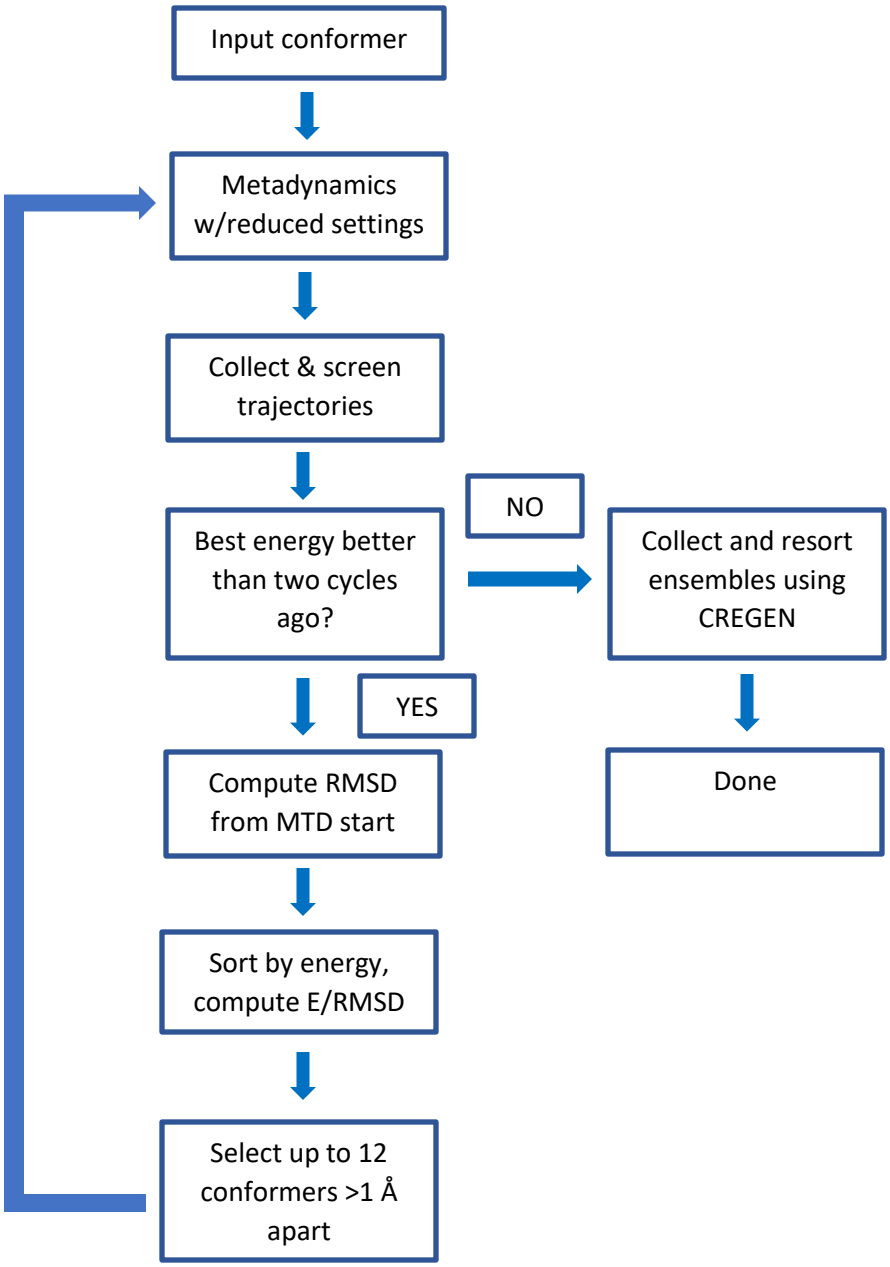
Figure 4.3 Here, the two algorithms for conformer selection are compared. Panel a illustrates selection based on energy and RMSD thresholds, and panel b illustrates selection based on the quotient of energy divided by RMSD. In these scatterplots, the relative energy of the conformers generated in one cycle is plotted against their RMSD from the structure with which their particular MTD run started. Conformers are colour-coded by the starting conformer of their MTD run. Selection regions are indicated by the shaded green regions. Points circled in black indicate conformers which were selected as starting points for the next cycle.



dissimilar two conformers are. As such, an additional constraint that any conformers chosen would have to be at least 1 Å apart, as measured by RMSD, was chosen. Further, a large RMSD relative to the starting conformer of a run indicates a higher probability that a simulation had crossed an energy barrier into a previously unexplored region of the PES. Thus, large RMSD relative to the starting conformer should be prioritized in picking structures from which to start new MTD runs. There are a number of methods one could choose to pick structures with low energy and high RMSD. One approach is to select structures above a certain RMSD cutoff (usually 1 or 2 Å), and below an energy cutoff (usually 2 kcal/mol above the lowest-energy conformer found so far), but this method proved to be unsatisfactory. Notably, such criteria often selected structures with, for example, an RMSD of 2.1 Å, and a relative energy of 1.9 kcal/mol. Here, the RMSD is barely above the threshold, while the energy is barely below it, indicating a conformer unlikely to be useful in searching for better conformers. At the same time, these selection criteria can lead to the exclusion of structures with very high RMSD, but slightly above the energy cutoff. To capture such structures, a better selection rule is needed. The quotient of relative energy divided by RMSD provides a metric that can select for high RMSD and low energy in a less arbitrary fashion. Using this metric, up to 11 conformers with a quotient below 0.5 kcal mol^{-1} $Å^{-1}$ were selected, plus the lowest-energy conformer found, for a total of up to 12 new conformers per cycle. Of course, the number of new conformers one wants to generate per cycle can depend on the exact nature of the problem of interest, however, 12 proved suitable for a diversity of dimers and trimers, with and without microhydration.[159] To avoid repeatedly exploring the same region of the PES, these conformers are selected subject to the additional constraint that none of the 11 can have an RMSD less than 1 Å relative to each other, the lowest-energy conformer, or any conformer selected in a previous cycle. Figure 4.3 shows the difference between these two selection rules, and the conformers which would be selected using each. The conformers presented in this example come from the third cycle of the geometry search for the heterodimer of model compounds **1** and **2a**. Panel a shows the threshold selection criteria as the green shaded area, while panel b shows the quotient selection criterion in the same way. Black circles indicate conformers selected to use as starting points for the next cycle. Those conformers within the selection area which are not circled were either too similar to an already-selected conformer, or else there were 11 conformers lower in energy already selected.

Figure 4.4 illustrates the improved algorithm for ensemble generation for clusters of highly flexible molecules. Four MTD runs are performed at each geometry, one with each combination of *kpush* and *alpha* of the reduced settings (*vide supra*). This weaker biasing still encourages the exploration of unvisited conformations but is more likely to retain the important non-covalent interactions while it

Figure 4.4 Schematic of new algorithm.

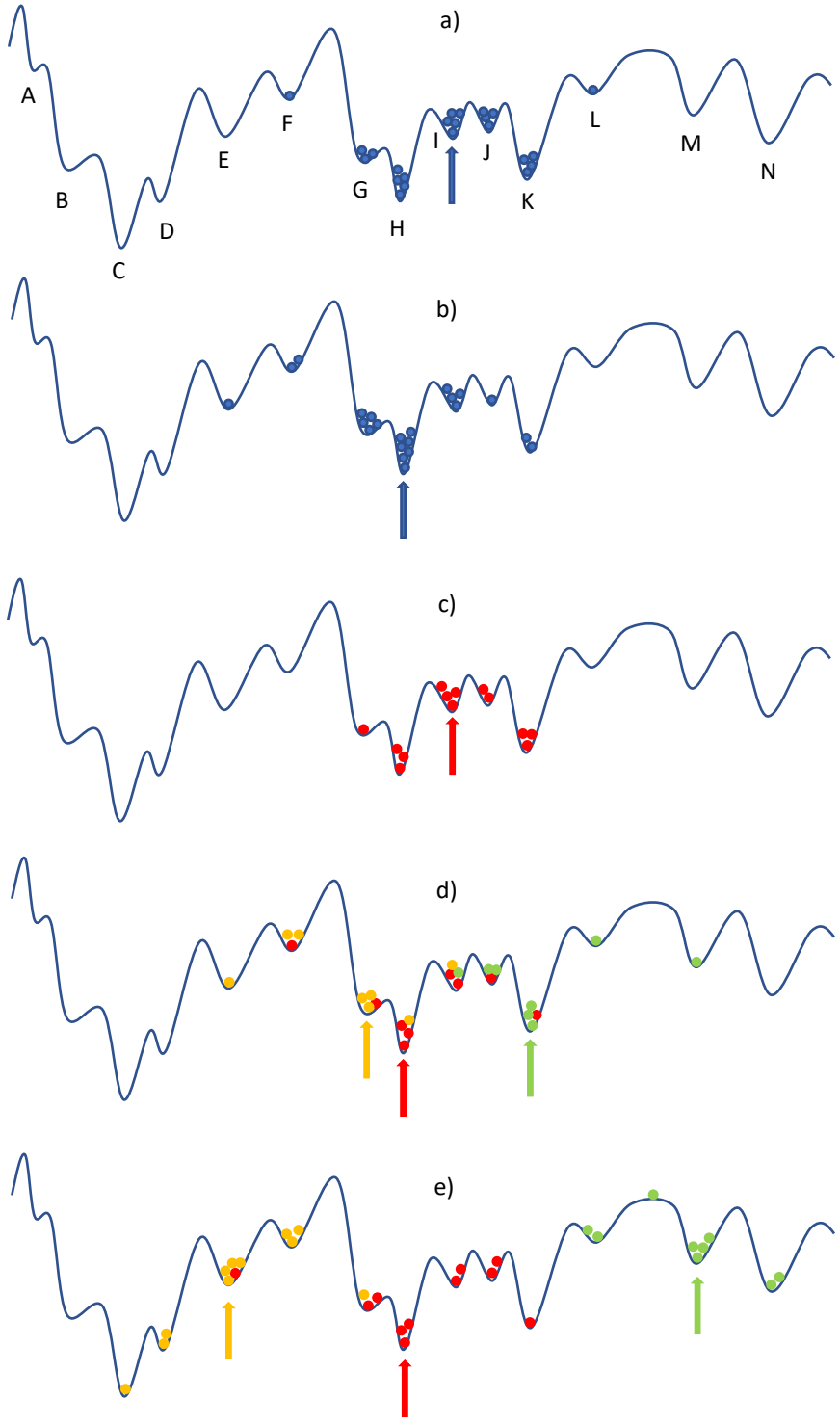


does so. Once the MTD runs are complete, the four trajectories for each structure are combined and screened using the *screen* function of CREST. Each conformer in the resulting ensembles is then compared to the starting structure of the MTD run from which it originated using the *rmsd* function of CREST. For each conformer, the ratio of relative energy (in kcal/mol) to RMSD (in Å) is calculated. To select conformers for the next cycle, all conformers from a cycle are ranked by energy, relative to the lowest energy found. The lowest energy conformer is carried forward, along with up to 11 conformers for which this ratio is less than 0.5 kcal mol^{-1} $Å^{-1}$, subject to the requirement that they are at least 1 Å in

RMSD from each other and from all structures which have been used as starting points in previous cycles. If a conformer does overlap with a previously selected conformer (as judged by RMSD), then the next highest in energy is selected, until 11 have been selected, or until the list of conformers with an energy to RMSD ratio less than 0.5 kcal mol^{-1} A^{-1} is exhausted. This approach ensures that a wide range of starting points are used, thus exploring over a much greater portion of configuration space than if only the lowest-energy conformer was used, while still prioritizing low energy starting points. If there is no improvement in the lowest-energy conformer for two successive cycles, then the algorithm terminates. All conformers found in all cycles are then collected and resorted using the *cregen* function of CREST to produce a final ensemble.

Figure 4.5 schematically illustrates, compares and contrasts a search using CREST with one using the present algorithm, on a hypothetical slice of the PES for a cluster of flexible molecules. Panels a and b illustrate CREST behaviour. In panel a, the initial starting structure is indicated by a blue arrow, at potential well I. Blue dots indicate conformers found in the first cycle of CREST. Due to the highly energetic biasing, CREST is unable to sample the low-energy region of the potential energy surface thoroughly before wandering off into much higher-energy conformers in directions orthogonal to the slice shown here. Panel b shows the second cycle of CREST. This cycle starts from the lowest-energy conformer found in the first cycle, which is located in potential well H. This cycle does not find a better conformer than that in potential well H, so CREST terminates. Panels c through e illustrate the modified algorithm. In panel c, the run is initiated from the same potential well, I, as in the CREST run. While the weaker biasing keeps the simulation on the low-energy portion of the potential energy surface longer, fewer MTD runs (four vs fourteen) per starting point are used, and the biasing doesn't "push as hard." Usually, this results in substantially better exploration of the low-energy potential energy surface than CREST gives, but here the performance is illustrated as slightly inferior, to be conservative. Panel d shows the second cycle of the modified algorithm. In this cycle, MTD runs are started from three conformers, in the G, H, and K potential wells. Conformers from I are not used as starting points, as they are too similar to the already used conformer from the first cycle, and those from J are not used, as the ratio of their energy (relative to the best conformer found so far, in well H) to RMSD (from the starting point of cycle 1) is too large. Usually, the algorithm will run 12 conformers per cycle, but in this illustration, at least within this slice of the PES, only up to three can be chosen within the constraints. In this panel, conformer dots are colour-coded to match the arrow indicating the starting position of the MTD run that produced them. Panel e shows the third cycle of the new algorithm. Here, the starting conformers are in wells E, H, and M. Conformers from well H would normally be discarded for being too

Figure 4.5 Comparison of CREST vs the present algorithm on a hypothetical potential energy surface. Panels a and b illustrate the first and second cycles of the CREST algorithm, respectively. Panels c through e illustrate the first through third cycles of the modified algorithm. Coloured dots indicate conformers found during the cycle, and match the colour of the arrow indicating the starting position(s) of each cycle



close to a previous starting position, but this is still the best energy found so far, so this conformer is retained. As an aside, this step illustrates why it is important to wait for two cycles without improvement before terminating the algorithm rather than the single cycle without improvement that CREST uses. It is common, especially after a cycle which has resulted in a significant improvement in energy, that few of the conformers generated in that cycle are low enough in energy to be selected as starting points, and so the next cycle may not find a lower-energy conformer, even as it does find conformers that will lead to better energies. The other wells with acceptable energy to RMSD ratios and sufficient distance from previous starts in this cycle are E and M. The third cycle finds the global minimum in well C, and after two additional cycles, the algorithm will terminate, having failed to find a lower-energy conformer.

When applied to the original problem of clusters of flexible molecules,[159] this algorithm gave substantially improved results over CREST. For example, for the $\mathbf{1}_2\mathbf{2b}$ trimer where the problem with sensitivity to the starting conformation was first noticed, the improved algorithm gave a structure which was lower in energy by 55.5 kJ/mol (13.3 kcal/mol) over the CREST result. This is a large enough difference that essentially none of the conformers found by CREST would exist at equilibrium, with a Boltzmann population ratio of less than 2×10^{-10}. Having developed the algorithm, we applied it to the other 16 supramolecular complexes under study (Complexes **3** – **19** in Chapter 5), and in every case, obtained superior results (lower energy) than when using CREST.

4.4 Conclusion

The generation of conformer ensembles and conformational analysis is an essential component of computational chemistry. While CREST performs these functions very well for single molecules and for clusters of rigid molecules, it is not always up to the task for clusters of flexible molecules. Thus, we have developed a new algorithm which extends CREST to such systems. We have termed this new algorithm LEDE-CREST, *i.e.*, a Low-Energy, Diversity-Enhanced variant of CREST. While, in principle, this algorithm could be implemented within CREST, for the present, we have written a script to automate the implementation of our algorithm. This script is designed to run on Unix computing clusters using Slurm schedulers, with CREST[61] (v. 2.12) and xtb[134] (v. 6.5.0) installed, but could be adapted to function on other architectures, as desired. The script, along with instructions, is freely available on Github at https://github.com/nking1/LEDE-CREST. The script is written in such a way that the user is free to adjust many parameters, such as the values of *kpush* and *alpha* used, the number of new conformers to use as starting points in each cycle, the value of the ratio for conformer selection, and simulation length,

among others. As noted, the numerical parameters discussed here were used successfully for a diversity of model asphaltene dimers and trimers and microhydrated versions thereof,[159] but one may have to do some testing and adjustment for the specific flexible monomers of interest.

Chapter 5

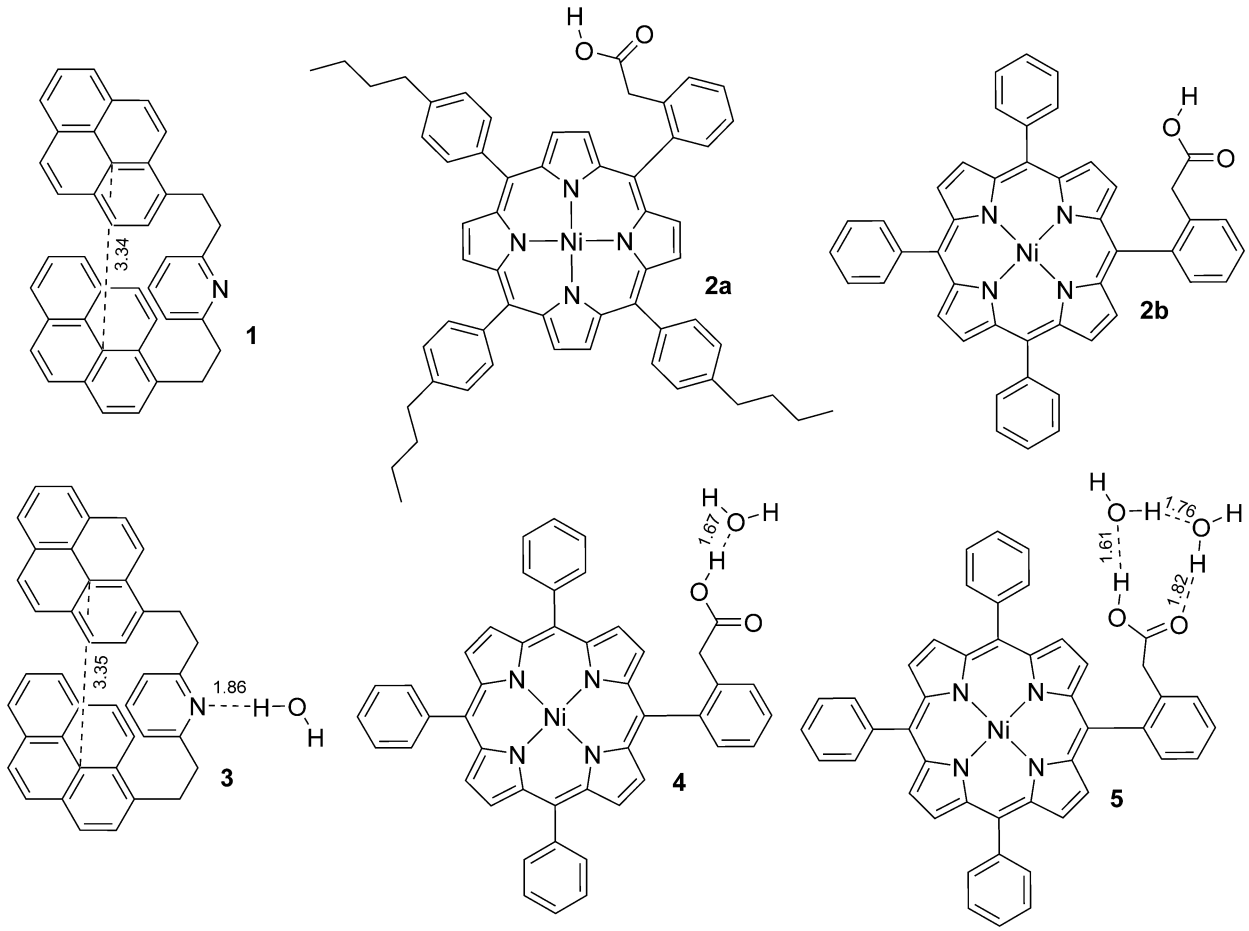
5. Computational Investigations Complement Experiment for a System of Non-Covalently Bound Asphaltene Model Compounds

5.1 Introduction

As discussed in Chapter 1, the NCIs of asphaltene molecules are still poorly understood, but gaining further understanding would lead to major economic and environmental benefits in the handling and processing of heavy oils. In 2015, Schulze *et al.* published a study on two model compounds for asphaltenes, and measured their interactions in solution using NMR.[4] The model compounds chosen were an archipelago-style compound with three islands - a central pyridine, with two pyrene moieties on tethers (**1**) - and a nickel porphyrin with an acidic side chain (**2a**), see Scheme 5.1. The study was designed to probe the role of acid-base hydrogen bonding in asphaltene aggregation, and began with an investigation of the base-pairing of phenylacetic acid with pyridine, in benzene solution, followed by **1** with phenylacetic acid, and **2a** with pyridine. Finally, the aggregation of **1** with **2a** was studied. Schulze *et al.* used a Job plot of the change in NMR resonance of select protons near the acid and base functionalities to gauge aggregation and to determine association constants. The evidence was equivocal as to whether the model compounds formed a 1:1 dimer or a 2:1 trimer, but they calculated the association constant for either case. For the case with a 1:1 dimer, the association constant was found to be 316 M^{-1}. On the other hand, for the case where the complex formed is a trimer, the association constant was found to be 1.23×10^6 M^{-2}. The equivalent Gibbs free energies of formation at 298 K are -14.3 kJ/mol and -34.7 kJ/mol, respectively, indicating moderately strong association. **2a** was also found to homodimerize, with an association constant of 390 M^{-1}, which corresponds to a Gibbs free energy of formation of -14.8 kJ/mol.

While the Schulze paper gives valuable insight into the role of acid-base hydrogen bonding in asphaltene aggregation, it leaves several unanswered questions which computations can address. The first, and most obvious, question is whether a 1:1 or 2:1 complex of **1** and **2a** is formed. Related to the above, the optimal structures of these complexes, and thus the nature of their interactions, is unknown. While it is natural to expect that hydrogen bonding is involved, and the changes in NMR chemical shift near hydrogen bonding sites seem to support this idea, there is nothing in the data which requires this to be the case, nor to indicate whether π–π stacking plays a role, nor how significant that role may be.

Scheme 5.1 Molecular structures of the asphaltene model compounds used in the Schulze *et al*. paper.[4] Dashed lines indicate non-covalent contacts, and numbers in fine print indicate corresponding distances, in Å. **2b** is a truncated version of **2a**, which we have used in most of our computations for reduced computational cost. **3** - **5** are microhydrated versions used in our study.



A second related question pertains to the change in chemical shift. Why does aggregation cause the protons alpha to the acid group to shift downfield, while those on the tethers of **1** shift upfield? Thirdly, the Schulze paper indicated that **2a** formed a homodimer in solution, which complicated their results, and precluded a definite conclusion on the stoichiometry of the interaction with **1**. They were able to calculate the strength of this homodimerization, with an equilibrium constant of 390 M^{-1}, but again the minimum-energy structure and the nature of the interactions are unknown. Fourthly, the authors allude to a paper by Tan *et al*. where water was demonstrated to enhance the aggregation of a model compound similar to **1**.[32] They suggest that the acid group of **2a** may be able to fulfil a similar role, but they do not address the possible role of water in promoting aggregation of their model compounds. It is of interest to know whether this can occur, and whether it was interfering with their measurements. Finally, the primary evidence in favour of the 2:1 complex was the Job plot of chemical shift. Given the

scrutiny that Job plots have attracted in recent years,[160] it is reasonable to ask whether this is an appropriate tool to use. Answers to many of these questions should yield insight that is useful toward developing a better understanding of aggregation in asphaltenes in general. For example, if water is important in inducing aggregation both here and in the case of Tan *et al*.,[32] then this may be a general result. Similarly, if hydrogen bonding turns out to be similar or greater in importance as π–π stacking, this bears implications for our general understanding of asphaltene aggregation.

5.2 Computational Methods

Given the number of rotatable bonds in each of the monomer compounds **1** and **2a**, and to a lesser extent in **2b**, finding the lowest-energy conformers is not a trivial task. In this work, initial geometry searches were performed using CREST v. 2.10.1 (Conformer-Rotamer Ensemble Sampling Tool)[61] for single molecules and LEDE-CREST (Low-Energy Diversity-Enhanced variant of CREST), which we developed in Chapter 4,[161] for clusters of flexible molecules. All DFT computations were performed using ORCA 5.0.4.[132] One of the best DFT functionals[162] for non-covalent interactions is ωB97X-V,[163] but due to the complexity of the -V dispersion, gradients are difficult to compute, resulting in long computation times for optimizations. Computations using the closely related ωB97X-D4, which also performs well,[116] proceed much faster. While benchmarks of the performance of ωB97X-D4 for geometries are not available, we presume that since it handles non-covalent energies well, it should perform well for non-covalent geometries, too. For each species studied, the lowest-energy structure obtained from CREST or LEDE-CREST was reoptimized using ωB97X-D4/def2-SVP/CPCM(Benzene),[79, 116, 117, 129] and thermodynamic computations were also carried out at this level, using Grimme's quasi-rigid-rotor-harmonic-oscillator approximation,[120] as implemented in ORCA. Frequencies were used as computed, and not scaled. High quality single point energies at the optimized geometries were obtained using ωB97X-V/def2-QZVPP/SMD(Benzene).[117, 129, 163] NMR shieldings are very sensitive to electron density around the nucleus, so specialized basis sets should be used.[164] Thus, NMR computations were performed using TPSS/pcSseg-2/SMD(Benzene)[119, 164, 165]//ωB97X-D4/def2-SVP/CPCM(Benzene). TPSS was chosen for NMR computations, as it has been shown to give good results for NMR shieldings, for a very moderate computational cost.[166] NMR chemical shifts are given as the difference between the computed shielding of tetramethylsilane (at the same level of theory as above) and the computed shielding of the proton of interest.

Free energies were determined according to the equations detailed in Chapter 3[167] and references cited therein.[122, 124] Specifically, we compute ΔG_{assoc}, the Gibbs free energy of association, using

$$\Delta G_{assoc} = E_{solv,comp} - \sum_i E_{solv,i} + G_{rrho,comp} - \sum_i G_{rrho,i} + (n-1)RT \ln\left(\frac{V_f}{V_i}\right). \quad (5.1)$$

Here, $E_{solv,comp}$ is the electronic energy of the complex, as given in the output of a single point computation in SMD solution, $E_{solv,i}$ is the electronic energy in solution for the *i*-th monomer in the complex, $G_{rrho,comp}$ is the Gibbs free energy correction for the complex, as given by Grimme's quasi-rigid-rotor-harmonic-oscillator approximation, and $G_{rrho,i}$ is the Gibbs free energy correction for the *i*-th monomer in the complex. The last term in this equation is the concentration correction, which accounts for the fact that electronic structure programs such as ORCA report free energies in the gas phase, where the standard concentration is 1 atm, but the standard concentration in solution is 1M. Here, n is the number of molecules in the complex, R is the ideal gas constant, T is the temperature, V_f is the molar volume at the solution standard state of 1M (which is simply 1L), and V_i is the molar volume at the gas standard state of 1 atm (22.4 L at 298K). Similarly, we define binding energy, E_b, as

$$E_b = \sum_i E_{solv,i} - E_{solv,comp}. \quad (5.2)$$

This definition ensures that stable complexes have positive binding energies, as per convention.

In Chapter 3,[167] we used ensembles to obtain the conformational entropy, for a more accurate overall ΔG_{assoc}. Conformational entropy reflects the improvement in free energy that results from the ability to populate multiple conformers. Both flexible monomers and flexible complexes benefit from conformational entropy, but the conformational entropy of a complex is usually less than the sum of the conformational entropies of the monomers, at least in cases where the monomers are flexible. This is due to the typical restriction of flexibility to meet the demands of forming an energetically favourable complex. However, computations of conformational entropy have not been attempted in the present work for several reasons. First, the computational cost to perform the geometry optimizations and subsequent harmonic frequencies computations using DFT on the many possible conformers generated by CREST would have been immense. Second, visual inspection of the structures for the lowest 1 kcal/mol of the CREST ensemble of complex **6** showed very little change in the core structure, with most of the changes between structures occurring in rotation and folding of butyl chains, or the outer pyrene moiety sliding back and forth. Similar results were observed for the other complexes, although their ensembles were much smaller, due to the exclusion of butyl chains. Thirdly, there was a moderate

amount of rearrangement upon DFT reoptimization, leading us to believe that many of the low-lying CREST conformers would converge to a single structure upon reoptimization, as we observed in other cases.[167]

It should be noted that CREST and LEDE-CREST compute conformational entropies at the GFN2-xTB level by default. The CREST-computed conformational free energies for **1**, **2a** and **2b** are -6.3, -18.2, and -11.0 kJ/mol, respectively, while the LEDE-CREST computed conformational free energies of complexes **6** to **19** range from -1.8 to -17.1 kJ/mol. This gives changes in conformational free energy ranging from -3.2 kJ/mol (complex **8**) to 20.2 kJ/mol (complex **13**). A full accounting of the CREST and LEDE-CREST computed conformational free energies is available in Appendix C. However, in our experience, these xTB-based ensembles are poorly ordered, energetically speaking, and contain many spurious minima. In our experience in smaller systems, reoptimization of a CREST or LEDE-CREST ensemble using DFT can reduce the number of conformers by up to an order of magnitude.[167] Thus, we do not trust these values for conformational entropy/free energy, and have not applied them in this study. A better solution to both the conformational entropy problem and to ensuring we find the lowest-energy conformer is the subject of ongoing work.

5.3 Results and Discussion

5.3.1 Monomers

5.3.1.1 Anhydrous Monomers

The first matter of interest is the monomer geometries. The molecular structures are given in Scheme 5.1. Geometry exploration as described in the methods section yielded the structures shown in Figure 5.1. Monomer **1** exhibits intramolecular π − π stacking, with an average interplane distance of 3.34 Å, while monomers **2a** and **2b** show no significant intramolecular contacts.

5.3.1.2 Hydrated Monomers

The geometries of the monomers were also determined for the case of microhydration. The cluster of **1** with one water molecule (complex **3**) and the clusters of **2b** with one or two water molecules (complexes **4** and **5**, respectively) were studied. For each, the water molecules were manually added to likely hydrogen bonding positions on the lowest-energy structures from the CREST runs for **1** and **2b**. These hydrated structures were optimized at the GFN2-xTB[134] level, and then used as starting points for LEDE-CREST[161] runs. The lowest-energy structure from each LEDE-CREST run was then reoptimized using ωB97X-D4/def2-SVP/CPCM(Benzene). The resulting structures are also shown in Figure 5.1. For hydrated monomer **3**, the π − π stacking of **1** is retained (average interplane distance of 3.35 Å), and the water molecule forms a hydrogen bond with the nitrogen of **1** (OHN angle 172°, N-H

distance 1.86 Å). For hydrated monomer **4**, an acid-water hydrogen bond is formed (OHO angle 163°, O-H distance 1.67 Å), while one of the water hydrogens points toward one of the porphyrin nitrogens. For hydrated monomer **5**, the two water molecules form a closed cycle of hydrogen bonds with the acid (OHO angles 176°, 160°, and 166°, O-H distances 1.61 Å, 1.76 Å, and 1.82 Å). The water molecules lie directly above the porphyrin system, but do not appear to be directly interacting with any specific atoms in it.

Figure 5.1 DFT-optimized geometries for the monomers (**1**, **2a**, and **2b**) and hydrated monomers (**3**, **4**, and **5**), at the ωB97X-D4/def2-SVP/CPCM(Benzene) level of theory. Carbon is grey, hydrogen is white, oxygen is red, nitrogen is blue, and nickel is pink.

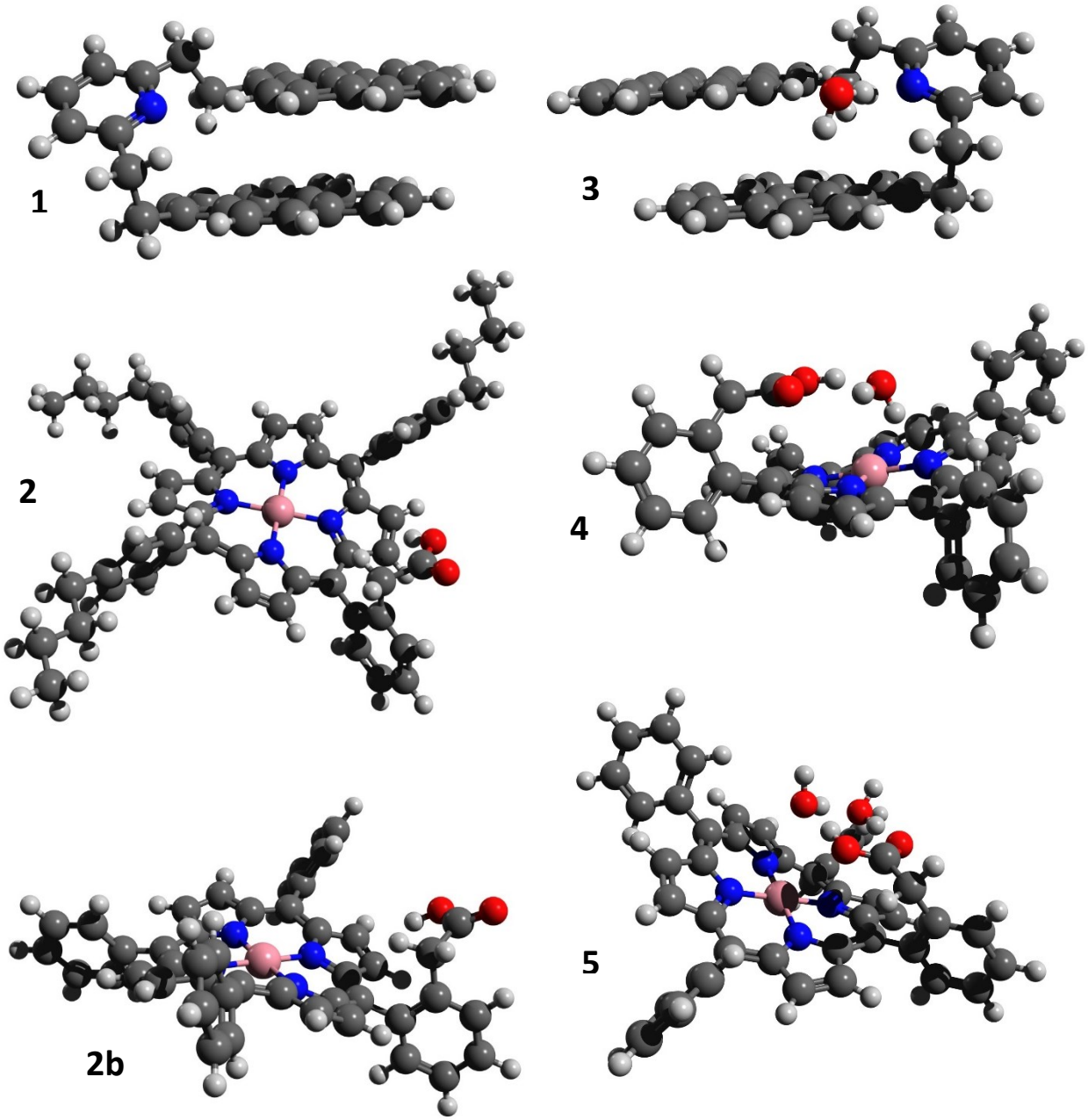
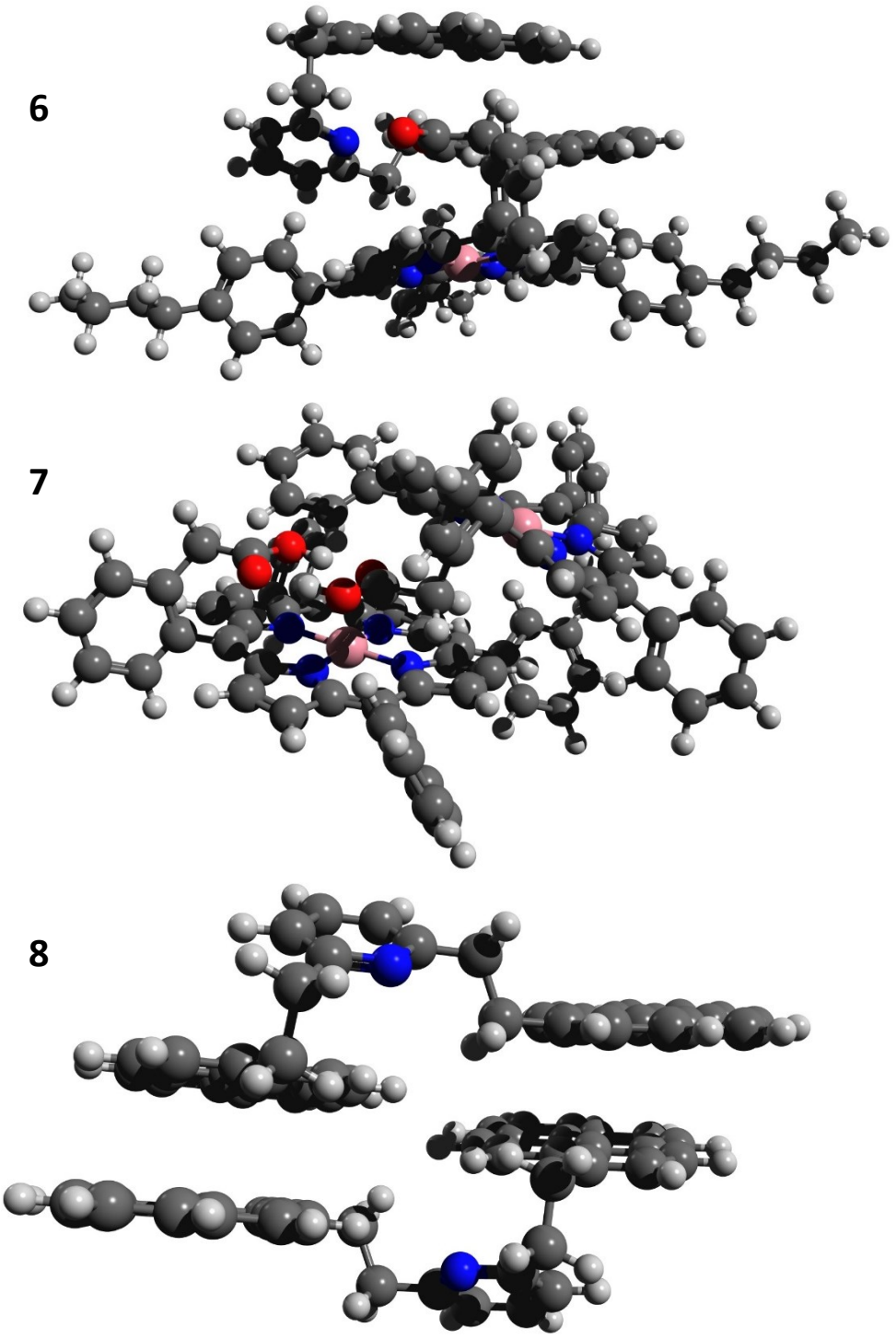
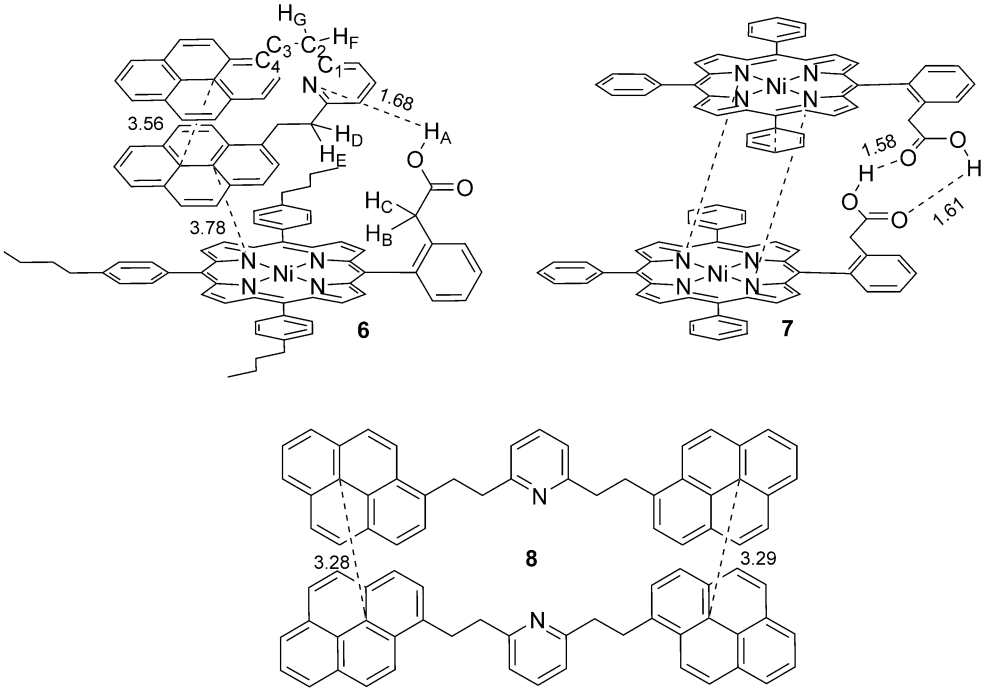


Figure 5.2 DFT-optimized geometries for the anhydrous dimers, at the ωB97X-D4/def2-SVP/CPCM(Benzene) level of theory. Carbon is grey, hydrogen is white, oxygen is red, nitrogen is blue, and nickel is pink. **6** is the heterodimer of **1** and **2a**, while **7** is the homodimer of **2b**, and **8** is the homodimer of 1.



Scheme 5.2 Schematic representation of the anhydrous dimers, see Figure 5.2. Dashed lines indicate non-covalent contacts, and numbers in fine print indicate corresponding distances, in Å. Labeled hydrogens (H_i, i = A-G) indicate those selected for NMR predictions. Labeled carbons (C_n, n = 1-4) indicate the carbons which define the dihedral angle described in section 5.6 "NMR."



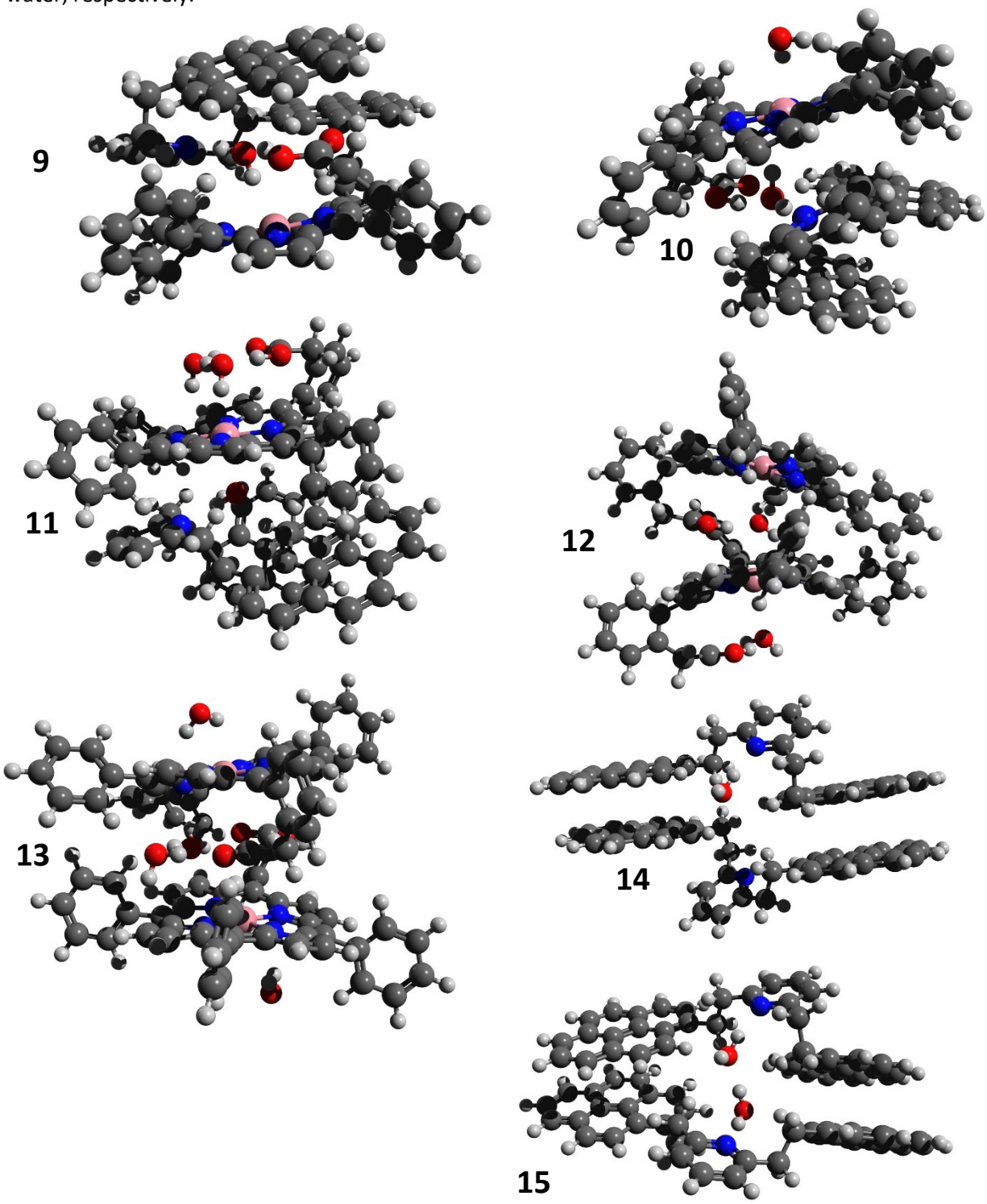
5.3.2 Dimers

For all dimer complexes, anhydrous or hydrated monomer structures, as appropriate, were selected and manually placed near each other. These structures were then optimized at the GFN2-xTB[134] level, and then used as starting points for LEDE-CREST[161] runs. The lowest-energy structure from each LEDE-CREST run was then reoptimized using DFT, as with the monomers. A schematic showing the primary interactions of the dimer complexes is given in Schemes 5.2 and 5.3, and the optimized geometries are shown in Figures 5.2 and 5.3. In looking at the structure for complex **6**, it does not appear that the butyl groups play any role in bonding. This is unsurprising, as the primary reason they were incorporated experimentally was to increase solubility.[4] To reduce computational expense, complexes **7**-**19** were computed using **2b** instead of **2a**. On **2b**, the *n*-butyl groups have been replaced by H, eliminating a total of 36 atoms, and reducing computational cost.

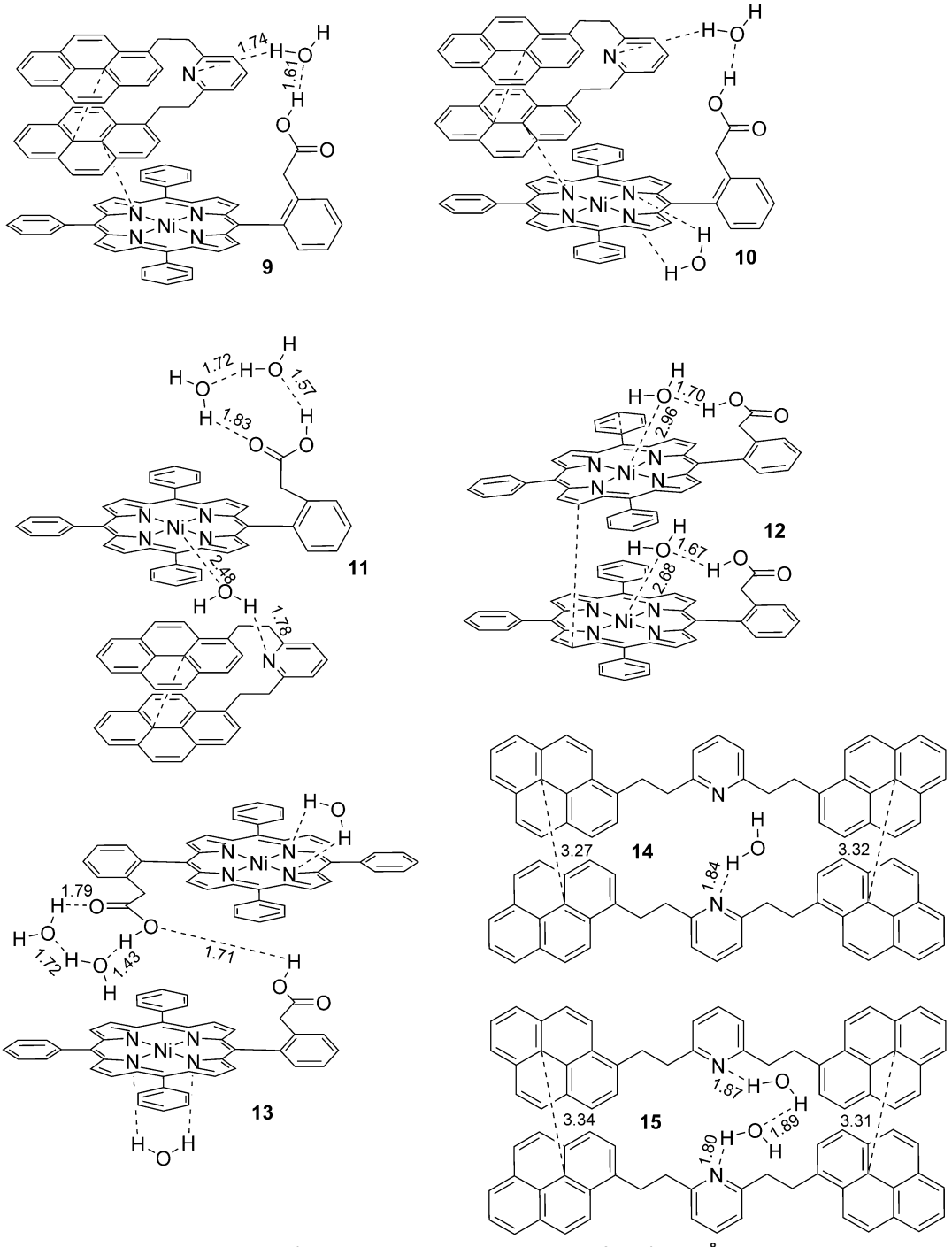
5.3.2.1 Anhydrous Dimers

In complex **6**, the pyridinic nitrogen of **1** is hydrogen-bonded to the proton of the acid group of **2a** (N-H distance 1.68 Å, NHO angle 167°), while the pyrene moieties of **1** are π-stacked to each other

Figure 5.3 DFT-optimized geometries for the hydrated dimers, at the ωB97X-D4/def2-SVP/CPCM(Benzene) level of theory. Carbon is grey, hydrogen is white, oxygen is red, nitrogen is blue, and nickel is pink. **9** to **11** are the heterodimers of **1** and **2b**, with one, two, or three molecules of water, respectively. **12** and **13** are the hydrated homodimers of **2b**, with two or four molecules of water, respectively. **14** and **15** are the hydrated homodimers of **1**, with one or two molecules of water, respectively.



Scheme 5.3 Schematic representation of the anhydrous dimers, see Figure 5.3. Dashed lines indicate non-covalent contacts, and numbers in fine print indicate corresponding distances, in Å.



(3.56 Å average interplane distance) and to the porphyrin core of **2a** (3.75 Å average interplane distance). In addition, there are close contacts between the outer pyrene and the CH_2COOH group, and between the inner pyrene and the pendant benzene rings. Complex **7** is the hydrogen-bonded homodimer of **2b** (O-H distances 1.58 Å and 1.61 Å, OHO angles 179° and 177°), with some additional

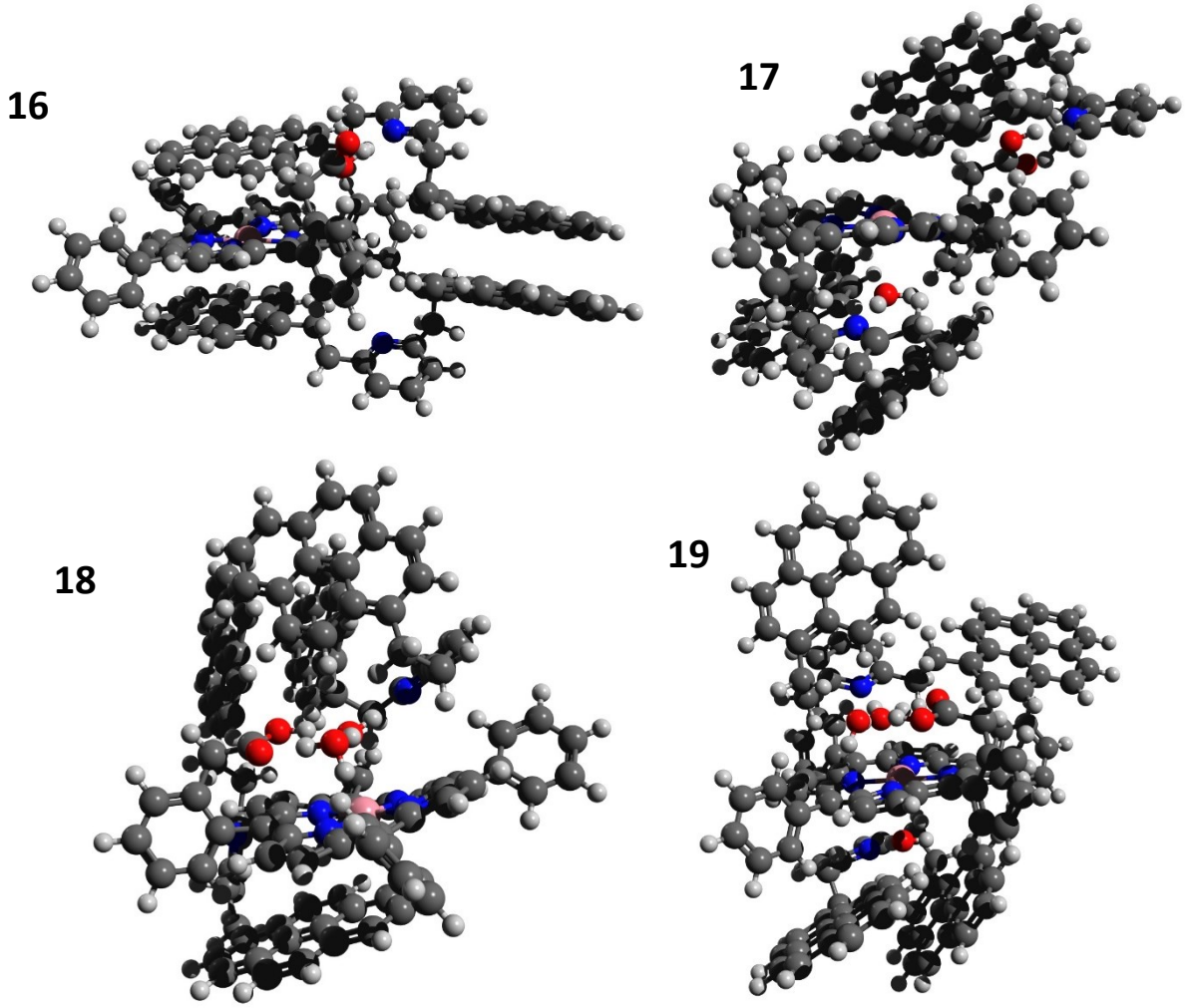
close contacts between parts of the porphyrin cores, and between pendant benzene rings. Our computations for complex **7** are consistent with the findings of Schulze *et al*. that **2a** partially dimerizes in solution,[4] see section 5.3.4 "Energetics of Binding" for details. Complex **8** is the homodimer of **1**. No hydrogen bonds are possible, and only π – π stacking holds the dimer together. Average interplane distances are 3.29 Å and 3.28 Å.

5.3.2.2 Hydrated Dimers

Seven possible hydrated dimers have been investigated: the heterodimers of **1** and **2b**, with one, two, or three water molecules (complexes **9**-**11**), the homodimers of **2b**, with two or four water molecules (complexes **12** and **13**), and the homodimers of **1**, with one or two molecules of water (complexes **14** and **15**). For complex **9**, the lowest-energy structure found has the water molecule accepting a hydrogen bond from the acid group (O-H distance 1.61 Å, OHO angle 161°) and donating a hydrogen bond to the pyridine nitrogen (N-H distance 1.74 Å, OHN angle 170°). As in complex **6**, the pyrene moieties exhibit nonbonded contacts with the porphyrin core, pendant benzene rings, and the acid side chain. In complex **10**, the lowest-energy structure is very similar to complex **9**, but with a second water molecule interacting with the open face of the porphyrin. The two protons on the second molecule of water are each pointing at a nitrogen in the porphyrin. For complex **11**, the situation is quite different. Here, two molecules of water interact with the acid side chain, forming a closed cycle of hydrogen bonds (O-H distances 1.57 Å, 1.72 Å, and 1.83 Å, OHO angles 174°, 162°, and 166°) while the third molecule of water is coordinated to the other face of the porphyrin (O-Ni distance 2.48 Å, O-Ni-N angles between 87° and 93°) while donating a hydrogen bond to the pyridine nitrogen (N-H distance 1.78 Å, OHN angle 170°). The pyrene moieties exhibit nonbonded contacts reminiscent of π – π stacking with each other and the porphyrin core, except that they are not parallel to each other, nor to the porphyrin core. Complex **12**, the dihydrated porphyrin homodimer, has each water molecule hydrogen bonded with an acid group (OH distances 1.67 Å and 1.70 Å, OHO angles 164° and 158°) and coordinated to the corresponding nickel centre (O-Ni distances 2.96 Å and 2.68 Å, O-Ni-N angles between 79° and 99°). Rather than having the hydrated acid groups interact with each other in the space between the porphyrin cores, one acid/water moiety is sandwiched between the cores, while the other is exposed to surrounding solvent. On the other hand, in complex **13**, the tetrahydrated version, the acid groups *do* interact with each other, and are sandwiched between the porphyrin cores. One acid group has a closed hydrogen bond cycle with two water molecules, like in complex **10** (OH distances 1.43 Å, 1.72 Å, and 1.79 Å, OHO angles 174°, 157°, and 165°), while also accepting a hydrogen bond from the other acid via the OH oxygen (OH distance 1.71 Å, OHO angle 166°). The two remaining water molecules interact with

the outer faces of the porphyrin cores, with each hydrogen atom pointing at a porphyrin nitrogen. In complex **14**, the monohydrated homodimer of **1**, the pyridine moieties are π – π stacked much like those in **8**, with average interplane distances of 3.27 Å and 3.32 Å. The water molecule donates a hydrogen bond to one nitrogen (N-H distance 1.84 Å, OHN angle 171°). Complex **15** is similar, with average interplane distances of 3.34 Å and 3.31 Å. Each water molecule donates a hydrogen bond to a nitrogen (N-H distances 1.80 Å and 1.87 Å, OHN angles 166° and 171°), but the water molecules also form a hydrogen bond between them (O-H distance 1.89 Å, OHO angle 171°).

Figure 5.4 DFT-optimized geometries for the trimers, at the ωB97X-D4/def2-SVP/CPCM(Benzene) level of theory. Carbon is grey, hydrogen is white, oxygen is red, nitrogen is blue, and nickel is pink. Complexes **16** through **19** consist of one molecule of **2b**, and two molecules of **1**, plus zero, one, two, or three molecules of water, respectively.



5.3.3 Trimers

Given that the Schulze paper[4] was ambiguous as to whether a dimer or trimer was formed, we explored the possibilities for trimerization, with and without microhydration. The methodology was the same as that used for the dimers. That is, optimised monomer structures were placed near each other, optimized using GFN2-xTB, and the resulting optimized structures were used as input for LEDE-CREST. The lowest-energy structure from LEDE-CREST was then recomputed at the ωB97X-V/def2-QZVPP/SMD(Benzene)//ωB97X-D4/def2-SVP/CPCM(Benzene) level. DFT-optimized geometries are shown in Figure 5.4, and a schematic view of their non-covalent interactions is shown in Scheme 5.4.

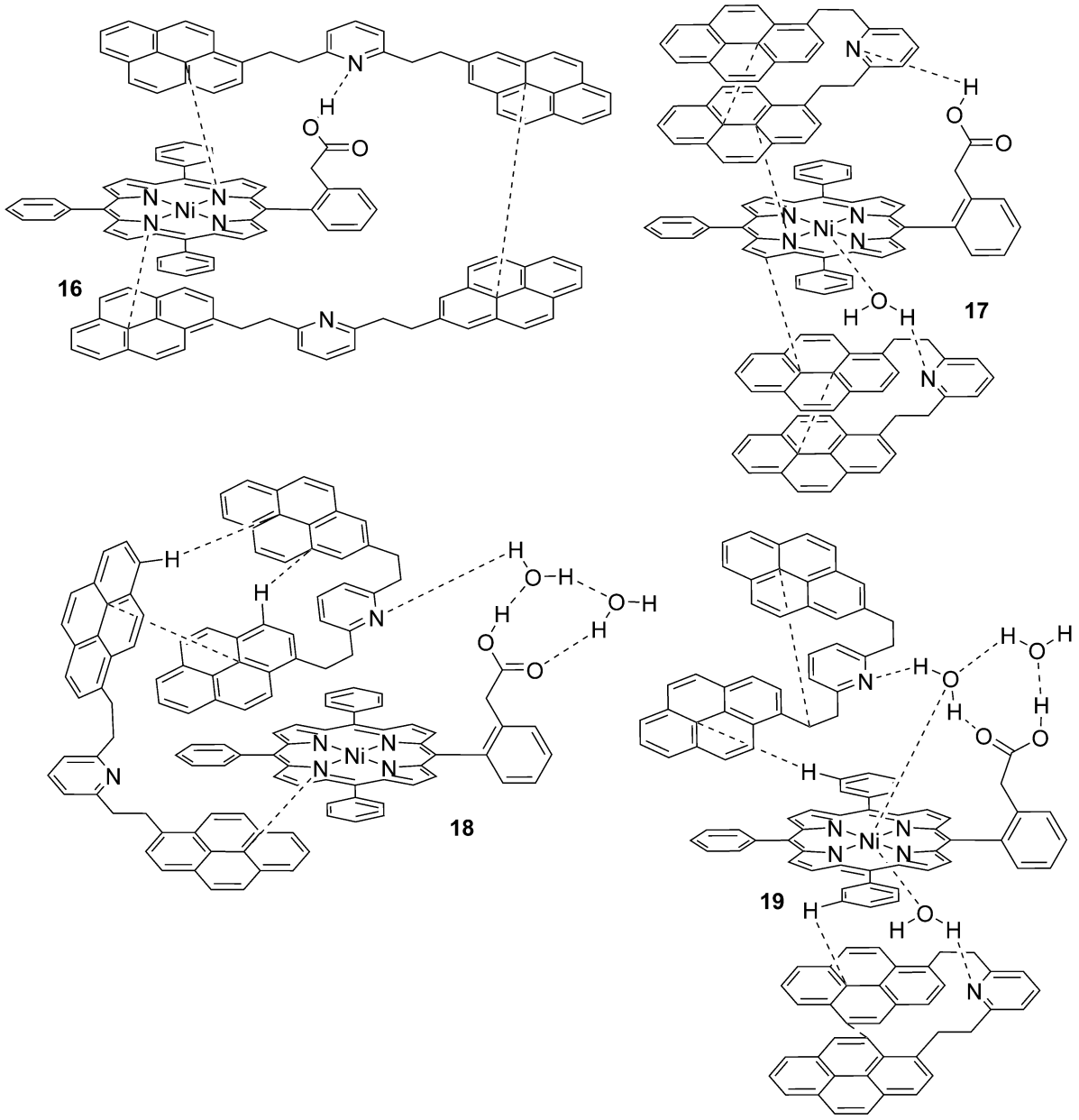
5.3.3.1 Anhydrous Trimer

Complex **16** has one unit of **1** hydrogen bonded to the acid of **2b** (N-H distance 1.60 Å, OHN angle 171°) and with one pyrene moiety π–π stacked to the porphyrin core (interplane distance 3.49 Å). The second unit of **1** has one pyrene moiety π–π stacked with the other face of the porphyrin core (interplane distance 3.31 Å). The remaining pyrene moieties (one from each unit of **1**) are π–π stacked with each other (interplane distance 3.30 Å).

5.3.3.2 Hydrated Trimers

The interactions of the first molecule of **1** with **2b** in complex **17** is very similar to those in complex **6**. The acid donates a hydrogen bond to the pyridine nitrogen (N-H distance 1.63 Å, NHO angle 169°) and one pyrene unit approximately stacks with the porphyrin core (not fully parallel, but average contact distance is about 3.4 Å) while the second pyrene unit has close non-covalent contacts with both the first pyrene unit and the acid side chain of the porphyrin. On the other face of the porphyrin core, the water molecule is coordinated to nickel (Ni-O distance 2.64 Å, N-Ni-O angles 85° to 94°) and donates a hydrogen bond to the second molecule of **1** (N-H distance 1.81 Å, OHN angle 170°). The pyrene moieties of this second molecule of **1** exhibit close non-covalent contacts with each other, the porphyrin core, and the pendant benzene rings. In complex **18**, two water molecules form a closed ring of hydrogen bonds with the acid group of the porphyrin, as in complex **11** (O-H distances 1.52 Å, 1.75 Å, and 1.82 Å, OHO angles 172°, 159°, and 169°), but one of the water molecules also donates a hydrogen bond to the nitrogen of one unit of **1** (N-H distance 1.75 Å, OHN angle 172°) and coordinates to the nickel centre (O-Ni distance 2.83 Å, O-Ni-N angles 82° to 98°). The unit of **1** without a hydrogen bond has one pyrene moiety partially stacked with the open face of the porphyrin core (not fully parallel, but nearest approach is 3.03 Å) and the other pyrene moiety wraps around to the other face of the porphyrin to stack with one of the pyrenes on the hydrogen-bonded unit of **1** (interplane distance 3.16 Å). The remaining pyrene moiety on the hydrogen-bonded unit of **1** is engaged in edge-to-π stacking

Scheme 5.4 Schematic representation of the trimers, see Figure 5.4. Dashed lines indicate non-covalent contacts.



with the two stacked pyrenes. Complex **19** exhibits the same ring of hydrogen bonds as complexes **11** and **18** (O-H distances 1.59 Å, 1.65 Å, and 1.84 Å, OHO angles 171°, 165°, and 156°), while one of those water molecules donates a hydrogen bond to one unit of **1** (N-H distance 1.74 Å, OHN angle 173°) and coordinates to the nickel centre (Ni-O distance 2.74 Å, O-Ni-N angles between 80° and 100°). On the other face of the porphyrin, the third molecule of water also coordinates to the nickel (Ni-O distance 2.58 Å, O-Ni-N angles between 86° and 94°) and donates a hydrogen bond to the other unit of **1** (N-H

distance 1.82 Å, OHN angle 173°). The pyrene moieties are not π−π stacked, but exhibit numerous close non-covalent contacts with each other and with the porphyrin core and pendant benzene rings.

5.3.4 Energetics of Binding

Binding energies and Gibbs free energies for the complexes are given in Table 5.1. Note that, by convention, binding energy for a bound complex is reported as a positive value, but for Gibbs free energies, a negative value indicates a stable complex.

As expected, as the complexes grow to contain more molecules, the Gibbs free energy trends toward positive (unstable), even as the binding energies become larger (more stable). If one takes the sum of the Gibbs free energy of formation and the binding energy (equivalent to the difference in stabilization as determined by the two metrics), the resulting quantity can be considered the penalty for constraining molecules to be close to each other. This penalty is largely entropic in nature, although there are smaller effects included, such as changes in zero-point energy. As complexes are progressively microhydrated, the increase in the penalty is fairly consistent, ranging from 34.3 kJ/mol to 50.9 kJ/mol per water molecule. We anticipate that this is a general result, and that for water to participate in aggregation mechanisms in benzene or similar solvents, it must contribute at least 34-50 kJ/mol to the binding energy to improve Gibbs free energy of association.

Dimers **6** and **7**, the anhydrous heterodimer and homodimer, respectively, are similar in binding energy, at 127 kJ/mol and 140 kJ/mol, and in Gibbs free energy, at -46.0 and -50.7 kJ/mol, which agrees qualitatively with the experimental results showing them to have similar equilibrium constants of formation.[4] However, the computed equilibrium constants of formation are 10^5 to 10^6 times larger than those reported experimentally. In terms of free energy, the computed Gibbs free energies of formation are 24.0 kJ/mol and 28.2 kJ/mol more stable than those reported experimentally. For the trimer, the computed equilibrium constant of formation is 7000 times larger, and 21.9 kJ/mol more stable.

There are a number of possible reasons why computation and experiment may differ in this case. It is tempting to simply jump to the conclusion that the computation is in error, but there are reasons against doing so in this case. Benchmarking against the GMTNK55 set showed ωB97X-V to have a weighted mean absolute deviation (WTMAD-1) of only 1.81 kcal/mol (7.57 kJ/mol) for non-covalent interactions,[162] and SMD has an MAD of about 0.7 kcal/mol (2.9 kJ/mol) for non-aqueous solvents.[119] Even taken together, these margins of error are much smaller than the difference in question. The molecules under consideration here are somewhat larger than those in the referenced test sets, so it

Table 5.1 Binding Energy, Gibbs free energy, "penalty" (sum of Gibbs free energy and binding energy), and equilibrium constant of formation for each complex.

Complex	# of water molecules	E_b (kJ/mol)	ΔG_{assoc} (kJ/mol)	$E_b + \Delta G_{assoc}$ (kJ/mol)	K
Hydrated Monomers					
3	1	34.3	8.0	42.2	0.040
4	1	55.3	-10.7	44.5	76.8
5	2	82.2	3.7	86.0	0.22
Anhydrous Dimers					
6	0	126.6	-46.0	80.5	1.2×10^{8}
7	0	139.8	-50.7	89.0	7.9×10^{8}
8	0	57.7	4.0	61.7	0.20
Hydrated Dimers					
9	1	130.2	-5.6	124.6	9.6
10	2	150.4	8.5	158.9	0.033
11	3	208.4	-4.5	203.9	6.1
12	2	156.7	6.1	162.7	0.087
13	4	229.5	10.0	239.5	0.017
14	1	80.2	20.4	100.6	2.6×10^{-4}
15	2	120.9	27.1	148.0	1.8×10^{-5}
Anhydrous Trimer					
16	0	201.8	-56.6	145.2	8.6×10^{9}
Hydrated Trimers					
17	1	228.4	-46.2	182.2	1.3×10^{8}
18	2	247.7	-14.6	233.1	360
19	3	285.7	-16.9	268.8	920

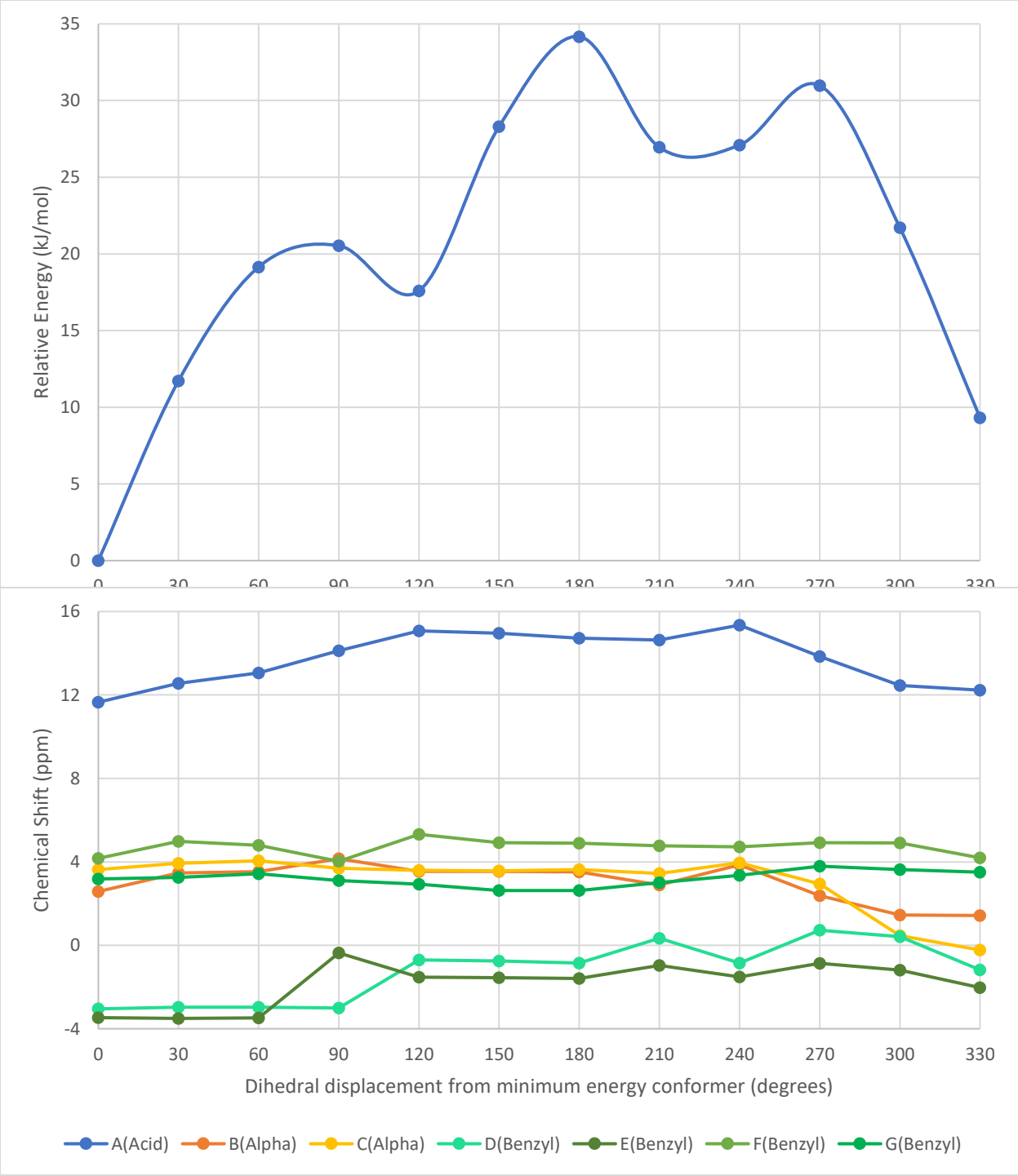
could be argued that our absolute errors should be expected to be larger, but even so, it seems unlikely that the errors would become three to four times larger, as they would need to be to explain the 21-28 kJ/mol discrepancy. If our conformational search were insufficient, and the conformers we used are not the global minimum energy structures, this could be a systematic source of error, but this would make our computed complexes less stable, rather than the present case where we report values which are more stable than the experimentally reported values. A more likely source of error actually lies within the experiment itself. Formation of dimers was measured exclusively through changes in chemical shift, as interpreted by a Job plot. However, Job plots are only appropriate when only one complex is formed from a set of monomers.[160] We have demonstrated that there are several different complexes which can form in competition with each other, which confounds any analysis based on a Job plot, and introduces possibly catastrophic errors in the reported equilibrium constants.

It should be mentioned that it is possible that part of the difference in association free energy between computation and experiment is due to a failure of the solvation model to properly account for directed interactions. It is known that continuum solvation approaches struggle with directed interactions.[168] Benzene solvent is able to π–π stack with the various aromatic regions of the monomers and complexes, and because these aromatic regions are partially blocked by other monomers in the complexes, such interactions would tend to stabilize monomers more than complexes. Thus, any missed stabilization from π–π stacking would tend to cause computed free energies of association to be too negative, potentially explaining part of the difference between computation and experiment. However, we do not believe that this effect is significant in this case for three reasons. First, all published examples we found where directed interactions caused problems for continuum solvation involved hydrogen bonds, rather than π–π stacking. Second, SMD includes a term to account for dispersion between solute and solvent, which should capture much or all of this effect.[119] Third, any stabilization from π–π stacking with solvent is likely to be offset by an entropy penalty for confining solvent near the solute. A full exploration of the effects of directed interactions could have been attempted using cluster-continuum modeling,[168] but this was deemed too computationally demanding.

The only complexes with large (>1000) equilibrium constants of formation (equivalent to Gibbs free energies of formation below -17 kJ/mol) are **6**, **7**, **16**, and **17**. Given that an equilibrium constant is defined as the concentration(s) of the product(s) of a reaction divided by the concentration(s) of its reactant(s), we can set up a system of equations to find the concentration of each species in solution. The experimental conditions varied from 1.25 mM to 10 mM for the initial concentration of each of 1 and 2a.[4] If the initial concentration of each is set to 2.5 mM, and the concentration of water is also assumed to be 2.5 mM, then solving the system of equations using Matlab[169] gives final concentrations of 2.5 mM for water, 87.3 µM for **1**, 234 nM for **2b**, 2.41 mM for **6**, 43.6 µM for **7**, 41.1 µM for **16**, and 1.52 pM for **17**. Full details are given in Appendix C.

One could argue, given the dominance of complex **6**, that a Job plot *is* appropriate here. However, it is impossible to predict how great a change in signal (NMR shift) each of the complexes will produce. If complex **7** produced a change in chemical shift that was, say, ten times larger than that produced by **6**, then it could still have a significant impact on the Job plot, even if it only accounts for 1.8% of complexes in solution, as in this case. In addition, the initial concentrations given above are only the midrange of those used in experiment. Measurements were made at many initial concentrations including up to 8.3 mM of **2**, and 1.7 mM of **1**. At these initial concentrations, the equilibrium

Figure 5.5 Energies (at the ωB97X-V/def2-QZVPP/SMD(Benzene)//ωB97X-D4/def2-SVP/CPCM(Benzene) level of theory) and NMR chemical shifts (at the TPSS/pcSseg-2/SMD(Benzene)//ωB97X-D4/def2-SVP/CPCM(Benzene) level of theory) for the relaxed surface scan of **6**. Panel a shows energies relative to the minimum energy conformer, in kJ/mol. Panel b shows the computed chemical shifts (relative to tetramethylsilane, computed at the same level of theory) of the protons indicated by letters in Scheme 5.2



concentrations change to 2.5 mM for water, 7.0 µM for **1**, 2.0 µM for **2b**, 1.7 mM for **6**, 3.3 mM for **7**, 870 nM for **16**, and 32 nM for **17**, for a situation where **7** dominates. Alternatively, measurements were also made at initial concentrations of 9.15 mM for **1** and 0.85 mM for **2**, which results in concentrations of 2.5 mM for water, 8.0 mM for **1**, 570 pM for **2b**, 540 µM for **6**, 260 pM for **7**, 310 µM for **16**, and 12 nM for **17**, giving a situation where **6** and **16** are present in nearly equivalent amounts. Clearly, the situation is far more nuanced than the single-complex ideal for Job plots.

5.3.5 NMR

Attempts were made to simulate NMR spectra to compare with experimental spectra, but this was quickly determined to be a futile endeavour. There are many possible conformers for each complex only slightly higher in energy than the minimum, and each would have to be reoptimized and reweighted using DFT to obtain an ensemble of high enough quality to make a useful NMR simulation. Even more problematic is the fact that individual NMR chemical shifts are very sensitive to atom position relative to the porphyrin and pyrene moieties. Thus, even with a complete ensemble, and even with small errors in energy, and thus in weighting, there will be a large error in the computed chemical shifts.

To illustrate this point, a relaxed surface scan of complex **6** was performed along a single degree of freedom - the dihedral angle defined by the carbon atom in the 2 position of the pyridine ring, the two carbon atoms of the C_2H_4 tether connecting the pyridine ring to one of the pyrene moieties, and the carbon atom in the 1 position on the pyrene moiety. These four carbon atoms are signified by the labels C_1, C_2, C_3, and C_4 in Scheme 5.2. This dihedral angle controls the rotation of the pyrene moiety relative to the rest of the complex. The dihedral was scanned in 30-degree increments from the minimum geometry, and all other degrees of freedom were optimized. High quality energies (ωB97X-V/def2-QZVPP/SMD(Benzene)) and NMR shieldings (TPSS/pcSseg-2/SMD(Benzene)) were then computed for each structure.

The protons selected for NMR predictions are indicated by letters in Scheme 5.2. Figure 5.5 shows the change in chemical shift for the acid proton and the protons at the alpha position relative to the acid. Numerical results are available in Appendix C. Note that the chemical shift can change by as much as 4.3 ppm, based only on the position of one pyrene moiety.

Table 5.2 Computed Gibbs free energies of association for complexes **20** – **24**, at the ωB97X-V/def2-QZVPP/SMD(Benzene)//ωB97X-D4/def2-SVP/CPCM(Benzene) level of theory, with NMR chemical shifts for selected protons in those complexes and in their monomers, computed at the TPSS/pcSseg-2/SMD(Benzene)//ωB97X-D4/def2-SVP/CPCM(Benzene) level of theory. Where there are chemically equivalent protons in the monomer, the average of their computed chemical shifts is reported, even if those positions are not necessarily chemically equivalent in the dimer. The protons investigated are the acid proton, the protons in the alpha position relative to the acid functional group, and the protons which are *ortho* to nitrogen in the pyridine moiety, or the protons on the methyl/methylene group in that position, whichever is applicable to the specific case. Also reported is the change in chemical shift upon dimerization.

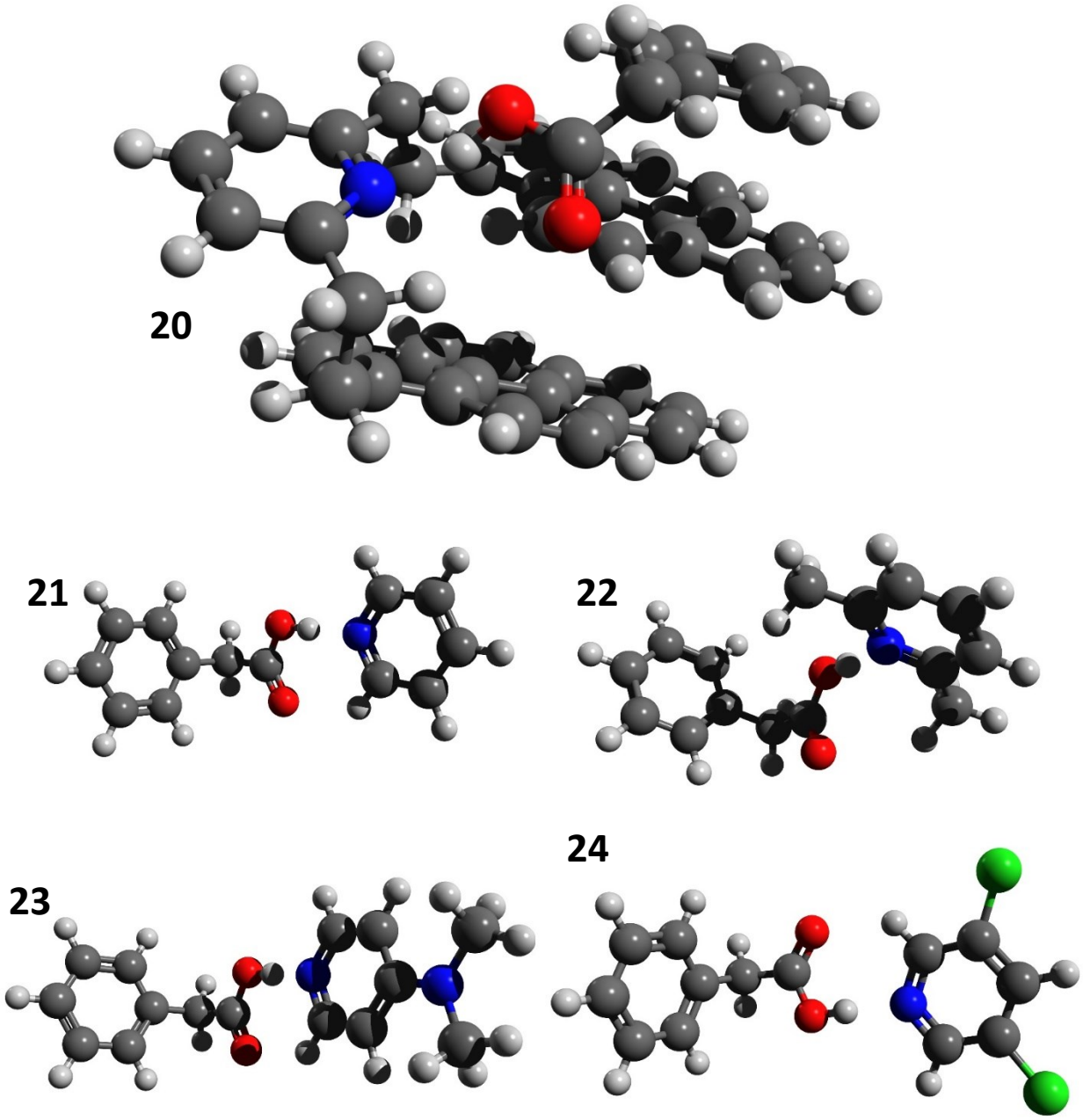
Complex	ΔG_{assoc} (kJ/mol)	Monomer Chemical Shift (ppm)			Dimer Chemical Shift (ppm)			Change in Chemical Shift (ppm)		
		Acid	Alpha	Ortho	Acid	Alpha	Ortho	Acid	Alpha	Ortho
20	11.4	5.73	3.73	3.39	12.34	3.26	3.58	6.61	-0.47	0.19
21	2.6	5.73	3.73	8.80	15.09	3.71	9.07	9.36	-0.02	0.28
22	-4.8	5.73	3.73	2.51	15.62	3.37	2.13	9.89	-0.37	-0.38
23	-1.4	5.73	3.73	8.30	16.29	3.60	8.62	10.56	-0.13	0.33
24	9.6	5.73	3.73	8.50	13.44	3.72	8.87	7.71	-0.01	0.37

5.3.6 Hydrogen Bond Strength

One interesting point raised in the experimental paper[4] is the weaker binding of **1** with phenylacetic acid than that of pyridine with phenylacetic acid. Specifically, the K_{assoc} of the pyridine-phenylacetic acid heterodimer was 123 M^{-1}, for a Gibbs free energy of -11.9 kJ/mol at 25 °C, while that of the heterodimer composed of **1** and phenylacetic acid was only 24 M^{-1}, for a Gibbs free energy of -7.9 kJ/mol. To investigate the reasons behind this difference, a series of smaller dimers with phenylacetic acid were considered. Specifically, the dimers of phenylacetic acid with **1** (complex **20**), with pyridine (complex **21**), with 2,6-lutidine (complex **22**), with *N,N*-dimethyl-4-aminopyridine (complex **23**), and with 3,5-dichloropyridine (complex **24**), were modeled. Optimized structures are given in Figure 5.6. Gibbs free energies of dimerization are given in Table 5.2. Unsurprisingly, the trend shows that pyridine units with electron withdrawing groups bind less strongly to phenylacetic acid, while those with electron donating groups bind more strongly. The exception is the dimer of 2,6-lutidine with phenylacetic acid, which is slightly less strongly bound than that involving pyridine, despite having two electron donating groups on lutidine. This appears to be partially a steric effect. The methyl groups in the 2 and 6 positions preclude the formation of a CH-O secondary hydrogen bond, and force the pyridine ring out of the plane of the carboxylic acid functional group, leading to an overall less stable dimer.

The same effect comes into play for the complex of **1** and phenylacetic acid, but here there is the additional effect where the pyrene moieties stack in such a position that the acid cannot get into

Figure 5.6 DFT-optimized geometries for complexes **20** - **24**, at the ωB97X-D4/def2-SVP/CPCM(Benzene) level of theory. Carbon is grey, hydrogen is white, oxygen is red, nitrogen is blue, and nickel is pink.



position to form a linear hydrogen bond. While the resulting energy penalty is partially offset by the ability of the phenyl ring to stack with a pyrene moiety, the overall result is an unstable Gibbs free energy of dimerization of 11.4 kJ/mol, which is only 8.8 kJ/mol less stable than the pyridine-phenylacetic acid complex. While the experimental results show stronger interactions for both than our results do, our results show the reason that the complex of **1** and phenylacetic acid is less stable than might be expected.

Table 5.2 also shows chemical shifts for select protons in complexes **20**-**24** and for the free monomers. No trend is evident in the chemical shift of the alpha or *ortho* protons, but the acid proton becomes more deshielded in complexes with electron donating groups (**22** and **23**) *vs*. the pyridine/phenylacetic acid complex (**21**), which is again more deshielded that the complex with electron withdrawing groups (**24**). Complex **20**, even though it has electron donating groups, is sterically hindered from achieving a normal hydrogen bond, and the acid proton resonates upfield of even complex **24**. All the complexes, however, are well downfield from isolated free acid. It appears that hydrogen bonding is very deshielding, and that the more electron density on nitrogen, the stronger the hydrogen bond, and thus the more deshielded the proton will be. Effects on the chemical shifts of the alpha and ortho protons are small, and do not correlate well with electron-donating strength, but we do see agreement with the experimental data that hydrogen bonding is deshielding for the *ortho* protons in most cases, and shielding for the alpha protons.

5.4 Conclusion

In conclusion, we have demonstrated that there are several possible complexes which can be formed using asphaltene model compounds **1** and **2**. Gibbs free energies of formation for the complexes range from strongly stabilized at -56.6 kJ/mol to moderately unstable at 10.0 kJ/mol, with complexes **6**, **7**, **16**, and **17** being the most stable. At the low concentrations attainable experimentally, **6** is the dominant complex, although **7** and **16** also form in noticeable quantities, especially if there is an excess of **1** or **2**, respectively. Both hydrogen bonding and π−π stacking figure prominently in all complexes studied, while coordination to the porphyrin and water-aided aggregation were less important. Further experimental and computational studies involving dimers or oligomers of other asphaltene model compounds should investigate whether these are general results. These results show that rather than the ideal situation of a single stable complex in solution, this system of model compounds has at least three different complexes in solution, and thus Job plot analysis, as performed in the original paper, is

necessarily unreliable. Alternative methods should be used for experimental studies hoping to measure concentrations of complexes and/or equilibrium constants of formation.

Chapter 6

6. Conclusions and Future Directions

6.1 Conclusions

The stated purpose of this thesis is to give insight into the details of both intermolecular and intramolecular NCIs in asphaltene model compounds. The research described herein has laid important groundwork and has granted insight. Importantly, existing methods have been tested (CREST, DFT), new tools have been developed (LEDE-CREST), and workflows have been created to assist in gaining additional insights in the future. As such, frameworks for further studies have been established.

In Chapter 3, we showed that computations can give insight into the mechanism behind problems encountered during the attempted synthesis of model compounds related in Chapter 2. This chapter also began to explore the impact of heteroatom substitution on π-π stacking. We used the Conformer-Rotamer Ensemble Sampling Tool (CREST) with the underlying semi-empirical GFN2-xtb method for automated geometry exploration of the homodimers of pyrene, pyrene-4,5-dione, and pyrene-4,5,9,10-tetraone, along with the heterodimer of pyrene and pyrene-4,5,9,10-tetraone. Geometries and energies of the dimers were further refined at the ωB97X-D4/def2-TZVP level of theory, both in gas phase, and in toluene solution. Computations in solution were handled using the CPCM (Conductor-like Polarizible Continuum Model) and SMD (Solvation Model based on Density) models. Two previously unidentified pyrene-homodimer conformations were identified, and the effects of oxidation on the geometries and energies of dimerization were explored; in general, oxidation leads to stronger intermolecular interactions, and decreased solubility in toluene. For selected dimers, DLPNO-CCSD(T)/cc-pVTZ/SMD(Toluene) energies were determined at the DFT geometries, which illustrated the accuracy of the ωB97X-D4 approach, with an MAD of 1.47 kJ/mol. We showed that CREST followed by DFT is an effective way to explore the ensemble of ways in which molecules can interact in solution.

To leverage computations to gain insight into the nature of NCIs in asphaltenes, sound computational approaches are needed. Among the most difficult of the computational challenges is the conformer problem – finding the conformer or ensemble of conformers which most contribute to the behaviour of molecules and aggregates. In Chapter 4, we developed the LEDE-CREST algorithm and scripts as a variant on CREST to determine structural ensembles and energetics of non-covalent clusters of flexible molecules (including with microhydration) at a semiempirical level. As with CREST, the energies are evaluated using the semiempirical GFN2-xTB[134] extended tight binding approach. The

utility of the algorithm was highlighted using dimers and trimers of model asphaltene compounds. However, further work is needed in this direction. In particular, GFN2-xTB,[134] the functional on which both CREST[61] and LEDE-CREST are based, gives insufficiently accurate geometries and energies for ensembles, and the computational cost to reoptimize the ensemble using DFT is simply too great to be feasible.

In Chapter 5, we applied the knowledge gained and workflows established in Chapters 2-4 to examine a system of two asphaltene model compounds which had previously been studied experimentally.[4] The model compounds investigated include a porphyrin with an acidic side chain, and a three-island archipelago compound with pyridine as the central island and pyrene for the outer islands. The possible stoichiometries and conformations for complexes were explored and compared to the experimental results. Our computational results showed that there are four possible complexes involving these two model compounds with large (K>1000) equilibrium constants of formation, which will exist in competition with each other. We found that both hydrogen bonding and $\pi - \pi$ stacking are important to this aggregation. On the other hand, neither water-mediated aggregation nor coordination to open porphyrin sites were found to be significant for this system, in contrast to some previous suggestions of their importance.[31-34] The multiple possible stoichiometries of complexes confound some of the analysis done in the experimental paper, as Job plots assume that only one complex is present. Gibbs free energies of association were determined for various complexes, with and without microhydration, at the ωB97X-V/def2-QZVPP//ωB97XD4/def2-SVP level of theory, and using SMD configured for benzene solvent. We also briefly explored some of the factors influencing the change in NMR chemical shift for select nuclei reported in the experimental paper.

6.2 Future Directions

As noted above, the computational cost to reoptimize entire ensembles (or even moderately-sized subsets thereof) using DFT quickly becomes prohibitive when the complexes consist of more than about 100 atoms. However, this reoptimization has been shown to be crucial, due to inaccuracies in the underlying GFN2-xTB method. Therefore, one approach to explore is the use of DFT-3c methods as an intermediate optimization and screening step for processing CREST and LEDE-CREST ensembles before DFT optimization. DFT-3c methods are DFT-based semiempirical methods which use specially-tailored minimal basis sets, along with some empirical corrections to approximate the accuracy of full DFT.[170-172] They are computationally much more intensive than GFN2-xTB, but much less than DFT with a large basis set. At the same time, their accuracy for energies and geometries is much better than GFN2-xtb,

but not as good as full DFT. The 3c methods in general, but especially B97-3c, were designed specifically to give reliable results for structural details and non-covalent interaction energies of large systems, without incurring too high a computational cost.[171] Preliminary results[173] indicate that screening and reoptimizing GFN2-xTB-derived ensembles with B97-3c[171] is effective for screening out high-energy conformers, eliminating spurious minima, and getting close to the DFT geometries of the true minima, for a greatly reduced computational effort. This approach, as an intermediate step, may make it feasible to perform DFT optimizations on a low-lying portion of the ensemble, rather than just for the single best-energy conformer, even for fairly large complexes. A B97-3c-refined ensemble should also give a lot more confidence when selecting the lowest-energy conformer.

Currently, two projects are underway using the methodology developed in this thesis, with the addition of an intermediate ensemble refinement using B97-3c. In the first, we are exploring the effect of heteroatom substitution (nitrogen, oxygen, sulfur, or in one case, both sulfur and oxygen) in small polycylic aromatic systems. As asphaltenes contain significant amounts of all these heteroatoms, it should be useful to identify motifs that are common in the interactions of systems containing these heteroatoms. The knowledge gained should then be useful as context when we proceed to look at the interactions of realistic archipelago-type asphaltene model compounds containing such polycyclic heteroaromatic systems.

The second project currently underway is another study of a system consisting of an acidic porphyrin and a pyridine-based archipelago compound. This study was initiated due to unexplained experimental results of NMR chemical shifts, as measured by Tykwinski and co-workers (unpublished). The system is similar to, but distinct from, that studied in Chapter 5, and refining the ensembles using B97-3c has already taught us some interesting things regarding how small changes in monomer structure can have large consequences in aggregate structures. While this project is still in progress, it is already demonstrating the utility of the B97-3c refinement step.

When these projects are complete, a study of the NCIs in a series of archipelago-type model compounds prepared by Dr. David Scott,[174] for which there is experimental solubility data,[49] and which contain small, heteroatom-substituted polycyclic aromatics tethered together by alkyl chains is planned. Preliminary results show that the length of the linking tethers may make a significant difference in the energetics of binding. Another interesting avenue of study would be to examine cation-π stacking, especially for the case of the partially-delocalized cations that form when nitrogen is at the junction of multiple rings in a polycyclic aromatic system.[175]

This thesis has opened up multiple avenues for the computational study of asphaltene model compounds. Previous work in this direction has mostly been using MD,[176-179] which, due to its use of force fields, can only give very approximate results. The few studies which have used DFT[33, 51, 53] suffer from an insufficient sampling of conformer space. The multi-tiered approach we have developed allows for a more rigorous and accurate treatment of model asphaltene systems, which should open the door to increasing understanding of these very complex molecules.

Bibliography

[1] U.S. Energy Information Administration - EIA - independent statistics and analysis. 2023; https://www.eia.gov/outlooks/steo/report/-global_oil.php#:~:text=Global%20oil%20demand%3A%20Global%20liquids,and%20other%20non%2DOECD%20countries., date accessed: Feb 11, 2023.

[2] Marafi, A.; Albazzaz, H.; Rana, M. S. Hydroprocessing of heavy residual oil: Opportunities and challenges. *Catalysis Today* **2019**, *329*, 125–134, III Scientific-Technological Symposium CATALYTIC HYDROPROCESSING IN OIL REFINING.

[3] Al-Attas, T. A.; Ali, S. A.; Zahir, M. H.; Xiong, Q.; Al-Bogami, S. A.; Malaibari, Z. O.; Razzak, S. A.; Hossain, M. M. Recent Advances in Heavy Oil Upgrading Using Dispersed Catalysts. *Energy & Fuels* **2019**, *33*, 7917–7949.

[4] Schulze, M.; Lechner, M. P.; Stryker, J. M.; Tykwinski, R. R. Aggregation of asphaltene model compounds using a porphyrin tethered to a carboxylic acid11Electronic supplementary information (ESI) available: NMR spectra of compounds 3 10, 2D NMR spectra of compound 4, as well as experimental details and NMR data of the aggregation studies. See DOI: 10.1039/c5ob00836k. *Organic & Biomolecular Chemistry* **2015**, *13*, 6984–6991.

[5] Alshareef, A. H. Asphaltenes: Definition, Properties, and Reactions of Model Compounds. *Energy & Fuels* **2020**, *34*, 16–30.

[6] Gray, M. R.; Yarranton, H. W.; Chac n Pati o, M. L.; Rodgers, R. P.; Bouyssiere, B.; Giusti, P. Distributed Properties of Asphaltene Nanoaggregates in Crude Oils: A Review. *Energy & Fuels* **2021**, *35*, 18078–18103.

[7] Scott, D. E.; Schulze, M.; Stryker, J. M.; Tykwinski, R. R. Deciphering structure and aggregation in asphaltenes: hypothesis-driven design and development of synthetic model compounds. *Chem. Soc. Rev.* **2021**, *50*, 9202–9239.

[8] Ansari, S.; Bahmaninia, H.; Mohammadi, M.-R.; Ostadhassan, M.; Norouzi-Apourvari, S.; Schaffie, M.; Ranjbar, M.; Hemmati-Sarapardeh, A. On the evaluation of asphaltene adsorption onto dolomite surface: The roles of flow condition, composition of asphaltene, and dolomite size. *Alexandria Engineering Journal* **2022**, *61*, 9411–9425.

[9] Tharanivasan, A. K.; Yarranton, H. W.; Taylor, S. D. Asphaltene Precipitation from Crude Oils in the Presence of Emulsified Water. *Energy & Fuels* **2012**, *26*, 6869–6875.

[10] Guida, P.; Colombo, E.; Colleoni, E.; Saxena, S.; Frassoldati, A.; Roberts, W. L.; Faravelli, T. Chemical Kinetics of Asphaltene Pyrolysis. *Energy & Fuels* **2021**, *35*, 8672–8684.

[11] Ghloum, E. F.; Al-Qahtani, M.; Al-Rashid, A. Effect of inhibitors on asphaltene precipitation for Marrat Kuwaiti reservoirs. *Journal of Petroleum Science and Engineering* **2010**, *70*, 99–106.

[12] Acevedo, S. c.; Guzman, K.; Ocanto, O. Determination of the Number Average Molecular Mass of Asphaltenes (Mn) Using Their Soluble A2 Fraction and the Vapor Pressure Osmometry (VPO) Technique. *Energy & Fuels* **2010**, *24*, 1809–1812.

[13] Herod, A. A.; Bartle, K. D.; Kandiyoti, R. Characterization of Heavy Hydrocarbons by Chromatographic and Mass Spectrometric Methods:? An Overview. *Energy & Fuels* **2007**, *21*, 2176–2203.

[14] Calemma, V.; Iwanski, P.; Nali, M.; Scotti, R.; Montanari, L. Structural Characterization of Asphaltenes of Different Origins. *Energy & Fuels* **1995**, *9*, 225–230.

[15] Sato, S.; Takanohashi, T.; Tanaka, R. Molecular Weight Calibration of Asphaltenes Using Gel Permeation Chromatography/Mass Spectrometry. *Energy & Fuels* **2005**, *19*, 1991–1994.

[16] Karaca, F.; Islas, C. A.; Millan, M.; Behrouzi, M.; Morgan, T. J.; Herod, A. A.; Kandiyoti, R. The Calibration of Size Exclusion Chromatography Columns:? Molecular Mass Distributions of Heavy Hydrocarbon Liquids. *Energy & Fuels* **2004**, *18*, 778–788.

[17] Acevedo, S. c.; Castro, A.; Negrin, J. G.; Fern ndez, A.; Escobar, G. n.; Piscitelli, V.; Delolme, F.; Dessalces, G. Relations between Asphaltene Structures and Their Physical and Chemical Properties:? The Rosary-Type Structure. *Energy & Fuels* **2007**, *21*, 2165–2175.

[18] Mart nez Haya, B.; Hortal, A. R.; Hurtado, P.; Lobato, M. D.; Pedrosa, J. M. Laser desorption/ionization determination of molecular weight distributions of polyaromatic carbonaceous compounds and their aggregates. *Journal of Mass Spectrometry* **2007**, *42*, 701–713.

[19] Wu, Q.; Seifert, D. J.; Pomerantz, A. E.; Mullins, O. C.; Zare, R. N. Constant Asphaltene Molecular and Nanoaggregate Mass in a Gravitationally Segregated Reservoir. *Energy & Fuels* **2014**, *28*, 3010–3015.

[20] Pomerantz, A. E.; Hammond, M. R.; Morrow, A. L.; Mullins, O. C.; Zare, R. N. Two-Step Laser Mass Spectrometry of Asphaltenes. *Journal of the American Chemical Society* **2008**, *130*, 7216–7217.

[21] Pomerantz, A. E.; Hammond, M. R.; Morrow, A. L.; Mullins, O. C.; Zare, R. N. Asphaltene Molecular-Mass Distribution Determined by Two-Step Laser Mass Spectrometry. *Energy & Fuels* **2009**, *23*, 1162–1168.

[22] Groenzin, H.; Mullins, O. C. Asphaltene Molecular Size and Structure. *The Journal of Physical Chemistry A* **1999**, *103*, 11237–11245.

[23] Groenzin, H.; Mullins, O. C. Molecular Size and Structure of Asphaltenes from Various Sources. *Energy & Fuels* **2000**, *14*, 677–684.

[24] Spiecker, P.; Gawrys, K. L.; Kilpatrick, P. K. Aggregation and solubility behavior of asphaltenes and their subfractions. *Journal of Colloid and Interface Science* **2003**, *267*, 178–193.

[25] stlund, J.-A.; Nyd n, M.; Auflem, I. H.; Sj blom, J. Interactions between Asphaltenes and Naphthenic Acids. *Energy & Fuels* **2003**, *17*, 113–119.

[26] Klein, G. C.; Kim, S.; Rodgers, R. P.; Marshall, A. G.; Yen, A.; Asomaning, S. Mass Spectral Analysis of Asphaltenes. I. Compositional Differences between Pressure-Drop and Solvent-Drop Asphaltenes Determined by Electrospray Ionization Fourier Transform Ion Cyclotron Resonance Mass Spectrometry. *Energy & Fuels* **2006**, *20*, 1965–1972.

[27] Rowland, S. M.; Robbins, W. K.; Corilo, Y. E.; Marshall, A. G.; Rodgers, R. P. Solid-Phase Extraction Fractionation To Extend the Characterization of Naphthenic Acids in Crude Oil by Electrospray Ionization Fourier Transform Ion Cyclotron Resonance Mass Spectrometry. *Energy & Fuels* **2014**, *28*, 5043–5048.

[28] McKenna, A. M.; Williams, J. T.; Putman, J. C.; Aeppli, C.; Reddy, C. M.; Valentine, D. L.; Lemkau, K. L.; Kellermann, M. Y.; Savory, J. J.; Kaiser, N. K.; Marshall, A. G.; Rodgers, R. P. Unprecedented Ultrahigh Resolution FT-ICR Mass Spectrometry and Parts-Per-Billion Mass Accuracy Enable Direct Characterization of Nickel and Vanadyl Porphyrins in Petroleum from Natural Seeps. *Energy & Fuels* **2014**, *28*, 2454–2464.

[29] Chac n Pati o, M. L.; Rowland, S. M.; Rodgers, R. P. Advances in Asphaltene Petroleomics. Part 1: Asphaltenes Are Composed of Abundant Island and Archipelago Structural Motifs. *Energy & Fuels* **2017**, *31*, 13509–13518.

[30] Rowland, S. M.; Smith, D. F.; Blakney, G. T.; Corilo, Y. E.; Hendrickson, C. L.; Rodgers, R. P. Online Coupling of Liquid Chromatography with Fourier Transform Ion Cyclotron Resonance Mass Spectrometry at 21 T Provides Fast and Unique Insight into Crude Oil Composition. *Analytical Chemistry* **2021**, *93*, 13749–13754.

[31] Gray, M. R.; Tykwinski, R. R.; Stryker, J. M.; Tan, X. Supramolecular Assembly Model for Aggregation of Petroleum Asphaltenes. *Energy & Fuels* **2011**, *25*, 3125–3134.

[32] Tan, X.; Fenniri, H.; Gray, M. R. Water Enhances the Aggregation of Model Asphaltenes in Solution via Hydrogen Bonding. *Energy & Fuels* **2009**, *23*, 3687–3693.

[33] da Costa, L. M.; Stoyanov, S. R.; Gusarov, S.; Tan, X.; Gray, M. R.; Stryker, J. M.; Tykwinski, R.; de M. Carneiro, J. W.; Seidl, P. R.; Kovalenko, A. Density Functional Theory Investigation of the Contributions of p p Stacking and Hydrogen-Bonding Interactions to the Aggregation of Model Asphaltene Compounds. *Energy & Fuels* **2012**, *26*, 2727–2735.

[34] Yin, C.-X.; Tan, X.; M llen, K.; Stryker, J. M.; Gray, M. R. Associative π-π Interactions of Condensed Aromatic Compounds with Vanadyl or Nickel Porphyrin Complexes Are Not Observed in the Organic Phase. *Energy & Fuels* **2008**, *22*, 2465–2469.

[35] Rubinstein, I.; Spyckerelle, C.; Strausz, O. Pyrolysis of asphaltenes: a source of geochemical information. *Geochimica et Cosmochimica Acta* **1979**, *43*, 1–6.

[36] Karimi, A.; Qian, K.; Olmstead, W. N.; Freund, H.; Yung, C.; Gray, M. R. Quantitative Evidence for Bridged Structures in Asphaltenes by Thin Film Pyrolysis. *Energy & Fuels* **2011**, *25*, 3581–3589.

[37] Strausz, O. P.; Mojelsky, T. W.; Lown, E. M. The molecular structure of asphaltene: an unfolding story. *Fuel* **1992**, *71*, 1355–1363.

[38] Ignasiak, T.; Kemp-Jones, A. V.; Strausz, O. P. The molecular structure of Athabasca asphaltene. Cleavage of the carbon-sulfur bonds by radical ion electron transfer reactions. *The Journal of Organic Chemistry* **1977**, *42*, 312–320.

[39] Schuler, B. et al. Heavy Oil Based Mixtures of Different Origins and Treatments Studied by Atomic Force Microscopy. *Energy & Fuels* **2017**, *31*, 6856–6861.

[40] Schuler, B.; Meyer, G.; Pe a, D.; Mullins, O. C.; Gross, L. Unraveling the Molecular Structures of Asphaltenes by Atomic Force Microscopy. *Journal of the American Chemical Society* **2015**, *137*, 9870–9876.

[41] Ruiz-Morales, Y. Application of the Y-Rule and Theoretical Study to Understand the Topological and Electronic Structures of Polycyclic Aromatic Hydrocarbons from Atomic Force Microscopy Images of Soot, Coal Asphaltenes, and Petroleum Asphaltenes. *Energy & Fuels* **2022**, *36*, 8725–8748.

[42] Mane, J. Y.; Stoyanov, S. R. Molecular dynamics investigation of the asphaltene kaolinite interactions in water, toluene, and water toluene mixtures. *Phys. Chem. Chem. Phys.* **2023**, *25*, 5638–5647.

[43] Gray, M. R.; Chac n Pati o, M. L.; Rodgers, R. P. Structure Reactivity Relationships for Petroleum Asphaltenes. *Energy & Fuels* **2022**, *36*, 4370–4380.

[44] Akbarzadeh, K.; Bressler, D. C.; Wang, J.; Gawrys, K. L.; Gray, M. R.; Kilpatrick, P. K.; Yarranton, H. W. Association Behavior of Pyrene Compounds as Models for Asphaltenes. *Energy & Fuels* **2005**, *19*, 1268–1271.

[45] George, G. N.; Gorbaty, M. L. Sulfur K-edge x-ray absorption spectroscopy of petroleum asphaltenes and model compounds. *Journal of the American Chemical Society* **1989**, *111*, 3182–3186.

[46] Sabbah, H.; Morrow, A. L.; Pomerantz, A. E.; Zare, R. N. Evidence for Island Structures as the Dominant Architecture of Asphaltenes. *Energy & Fuels* **2011**, *25*, 1597–1604.

[47] Dechaine, G. P.; Maham, Y.; Tan, X.; Gray, M. R. Regular Solution Theories Are Not Appropriate for Model Compounds for Petroleum Asphaltenes. *Energy & Fuels* **2011**, *25*, 737–746.

[48] Cardozo, S. D.; Schulze, M.; Tykwinski, R. R.; Gray, M. R. Addition Reactions of Olefins to Asphaltene Model Compounds. *Energy & Fuels* **2015**, *29*, 1494–1502.

[49] Morimoto, M.; Fukatsu, N.; Tanaka, R.; Takanohashi, T.; Kumagai, H.; Morita, T.; Tykwinski, R. R.; Scott, D. E.; Stryker, J. M.; Gray, M. R.; Sato, T.; Yamamoto, H. Determination of Hansen Solubility Parameters of Asphaltene Model Compounds. *Energy & Fuels* **2018**, *32*, 11296–11303.

[50] Tan, X.; Fenniri, H.; Gray, M. R. Pyrene Derivatives of 2,2'-Bipyridine as Models for Asphaltenes: Synthesis, Characterization, and Supramolecular Organization. *Energy & Fuels* **2008**, *22*, 715–720.

[51] Ekramipooya, A.; Valadi, F. M.; Farisabadi, A.; Gholami, M. R. Effect of the heteroatom presence in different positions of the model asphaltene structure on the self-aggregation: MD and DFT study. *Journal of Molecular Liquids* **2021**, *334*, 116109.

[52] Janesko, B. G.; Brothers, E. N. Half-Pancake Bonding in Asphaltenes. *Energy & Fuels* **2021**, *35*, 15657–15662.

[53] Wang, H.; Xu, H.; Jia, W.; Liu, J.; Ren, S. Revealing the Intermolecular Interactions of Asphaltene Dimers by Quantum Chemical Calculations. *Energy & Fuels* **2017**, *31*, 2488–2495.

[54] Podeszwa, R.; Szalewicz, K. Physical origins of interactions in dimers of polycyclic aromatic hydrocarbons. *Phys. Chem. Chem. Phys.* **2008**, *10*, 2735–2746.

[55] Guo, Y.; Riplinger, C.; Becker, U.; Liakos, D. G.; Minenkov, Y.; Cavallo, L.; Neese, F. Communication: An improved linear scaling perturbative triples correction for the domain based local pair-natural orbital based singles and doubles coupled cluster method [DLPNO-CCSD(T)]. *The Journal of Chemical Physics* **2018**, *148*, 011101.

[56] Altun, A.; Neese, F.; Bistoni, G. Effect of Electron Correlation on Intermolecular Interactions: A Pair Natural Orbitals Coupled Cluster Based Local Energy Decomposition Study. *J. Chem. Theory Comput.* **2019**, *15*, 215–228.

[57] Podeszwa, R.; Bukowski, R.; Szalewicz, K. Potential Energy Surface for the Benzene Dimer and Perturbational Analysis of π-π Interactions. *J. Phys. Chem. A* **2006**, *110*, 10345–10354.

[58] Zhang, Y.; Siskin, M.; Gray, M. R.; Walters, C. C.; Rodgers, R. P. Mechanisms of Asphaltene Aggregation: Puzzles and a New Hypothesis. *Energy & Fuels* **2020**, *34*, 9094–9107.

[59] Walsh, J. C.; Williams, K.-L. M.; Lungerich, D.; Bodwell, G. J. Synthesis of Pyrene-4,5-dione on a 15 g Scale. *European Journal of Organic Chemistry* **2016**, *2016*, 5933–5936.

[60] Hu, J.; Zhang, D.; Harris, F. W. Ruthenium(III) Chloride Catalyzed Oxidation of Pyrene and 2,7-Disubstitued Pyrenes: An Efficient, One-Step Synthesis of Pyrene-4,5-diones and Pyrene-4,5,9,10-tetraones. *The Journal of Organic Chemistry* **2005**, *70*, 707–708.

[61] Pracht, P.; Bohle, F.; Grimme, S. Automated exploration of the low-energy chemical space with fast quantum chemical methods. *Phys. Chem. Chem. Phys.* **2020**, *22*, 7169–7192.

[62] Diner, C.; Scott, D. E.; Tykwinski, R. R.; Gray, M. R.; Stryker, J. M. Scalable, Chromatography-Free Synthesis of Alkyl-Tethered Pyrene-Based Materials. Application to First-Generation "Archipelago Model" Asphaltene Compounds. *The Journal of Organic Chemistry* **2015**, *80*, 1719–1726.

[63] Cheung, K. Y.; Segawa, Y.; Itami, K. Synthetic Strategies of Carbon Nanobelts and Related Belt-Shaped Polycyclic Aromatic Hydrocarbons. *Chemistry - A European Journal 26*, 14791–14801.

[64] Sarmah, M.; Sharma, A.; Gogoi, P. Exploration of Kobayashi's Aryne Precursor: A Potent Reactive Platform for the Synthesis of Polycyclic Aromatic Hydrocarbons. *Org. Biomol. Chem.* **2021**, *19*, 722–737.

[65] Patern, G. M.; Goudappagouda; Chen, Q.; Lanzani, G.; Scotognella, F.; Narita, A. Large Polycyclic Aromatic Hydrocarbons as Graphene Quantom Dots: from Synthesis to Spectroscopy and Photonics. *Advanced Optical Materials* **2021**, *9*, 2100508.

[66] He, M.; Yu, X.; Wang, Y.; Bao, M. Tunable Redox Potential Photocatalyst: Aggregates of 2,3-Dicyanopyrazino Phenanthrene Derivatives for the Visible-Light-Induced a-Allylation of Amines. *The Journal of Organic Chemistry* **2021**, *86*, 14720–14731.

[67] Brunner, K.; van Dijken, A.; B rner, H.; Bastiaansen, J. J. A. M.; Kiggen, N. M. M.; Langeveld, B. M. W. Carbazole Compounds as Host Materials for Triplet Emitters in Organic Light-Emitting Diodes:? Tuning the HOMO Level without Influencing the Triplet Energy in Small Molecules. *Journal of the American Chemical Society* **2004**, *126*, 6035–6042.

[68] Z phel, L.; Beckmann, D.; Enkelmann, V.; Chercka, D.; Rieger, R.; M llen, K. Asymmetric pyrene derivatives for organic field-effect transistors. *Chem. Commun.* **2011**, *47*, 6960–6962.

[69] Salonen, L. M.; Medina, D. D.; Carb Argibay, E.; Goesten, M. G.; Mafra, L. s.; Guldris, N.; Rotter, J. M.; Stroppa, D. G.; Rodr guez Abreu, C. A supramolecular strategy based on molecular dipole moments for high-quality covalent organic frameworks. *Chem. Commun.* **2016**, *52*, 7986–7989.

[70] Li, Y.; Zhang, C.; Li, Z.; Gu, P.; Wang, Z.; Li, H.; Lu, J.; Zhang, Q. Controlled deposition of large-area and highly-ordered thin films: effect of dip-coating-induced morphological evolution on resistive memory performance. *J. Mater. Chem. C* **2019**, *7*, 3512–3521.

[71] Yuan, Y.; Lo, K.-C.; Szeto, L.; Chan, W.-K. Synthesis of Pyrazinopyrazine-Fused Azaacenes through Direct Condensation Reactions between Quinoxalinediamine and Diketones. *The Journal of Organic Chemistry* **2020**, *85*, 6372–6379.

[72] Yang, H.; Li, Y.; Zhao, Y.; Yu, S.; Ma, H.; Qian, L.; Wang, R.; Yu, T.; Su, W. Four new bipolar Indolo[3,2-b]carbazole derivatives for blue OLEDs. *Dyes and Pigments* **2021**, *187*, 109096.

[73] Yi, X.; Yang, P.; Huang, D.; Zhao, J. Visible light-harvesting cyclometalated Ir(III) complexes with pyreno[4,5-d]imidazole C^N ligands as triplet photosensitizers for triplet triplet annihilation upconversion. *Dyes and Pigments* **2013**, *96*, 104–115.

[74] Vollman, H.; Becker, H.; Corell, M.; Streek, H. Beitrage zur Kenntnis des Pyrens und seiner Derivate. *Justus Liebigs Annalen Der Chemie* **1937**, *531*, 1–159.

[75] Oberender, F. G.; Dixon, J. A. Osmium and ruthenium tetroxide-catalyzed oxidations of pyrene1. *The Journal of Organic Chemistry* **1959**, *24*, 1226–1229.

[76] Li, Y.; Zhang, C.; Li, Z.; Gu, P.; Wang, Z.; Li, H.; Lu, J.; Zhang, Q. Controlled deposition of large-area and highly-ordered thin films: effect of dip-coating-induced morphological evolution on resistive memory performance. *J. Mater. Chem. C* **2019**, *7*, 3512–3521.

[77] Grimme, S. Accurate description of van der Waals complexes by density functional theory including empirical corrections. *Journal of Computational Chemistry* **2004**, *25*, 1463–1473.

[78] Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *The Journal of Chemical Physics* **2010**, *132*, 154104.

[79] Caldeweyher, E.; Ehlert, S.; Hansen, A.; Neugebauer, H.; Spicher, S.; Bannwarth, C.; Grimme, S. A generally applicable atomic-charge dependent London dispersion correction. *The Journal of Chemical Physics* **2019**, *150*, 154122.

[80] Tkatchenko, A.; DiStasio, R. A.; Car, R.; Scheffler, M. Accurate and Efficient Method for Many-Body van der Waals Interactions. *Phys. Rev. Lett.* **2012**, *108*, 236402.

[81] Becke, A. D.; Johnson, E. R. Exchange-hole dipole moment and the dispersion interaction. *The Journal of Chemical Physics* **2005**, *122*, 154104.

[82] Becke, A. D.; Johnson, E. R. Exchange-hole dipole moment and the dispersion interaction revisited. *The Journal of Chemical Physics* **2007**, *127*, 154108.

[83] Vydrov, O. A.; Van Voorhis, T. Nonlocal van der Waals density functional: The simpler the better. *The Journal of Chemical Physics* **2010**, *133*, 244103.

[84] Bloom, J. W. G.; Wheeler, S. E. Taking the Aromaticity out of Aromatic Interactions. *Angewandte Chemie International Edition* **2011**, *50*, 7847–7849.

[85] Grimme, S. Do Special Noncovalent π-π Stacking Interactions Really Exist? *Angewandte Chemie International Edition* **2008**, *47*, 3430–3434.

[86] Carter-Fenk, K.; Herbert, J. M. Reinterpreting π-stacking. *Phys. Chem. Chem. Phys.* **2020**, *22*, 24870–24886.

[87] Cabaleiro-Lago, E. M.; Rodríguez-Otero, J. On the Nature of σ-σ, σ-π, and π-π Stacking in Extended Systems. *ACS Omega* **2018**, *3*, 9348–9359.

[88] Janowski, T.; Pulay, P. A Benchmark Comparison of σ/σ and π/π Dispersion: the Dimers of Naphthalene and Decalin, and Coronene and Perhydrocoronene. *Journal of the American Chemical Society* **2012**, *134*, 17520–17525.

[89] Hutchison, G. R.; Ratner, M. A.; Marks, T. J. Intermolecular Charge Transfer between Heterocyclic Oligomers. Effects of Heteroatom and Molecular Packing on Hopping Transport in Organic Semiconductors. *Journal of the American Chemical Society* **2005**, *127*, 16866–16881.

[90] Loeffler, J. R.; Schauperl, M.; Liedl, K. R. Hydration of Aromatic Heterocycles as an Adversary of Ï€-Stacking. *Journal of Chemical Information and Modeling* **2019**, *59*, 4209–4219.

[91] Gryn'ova, G.; Corminboeuf, C. Implications of Charge Penetration for Heteroatom-Containing Organic Semiconductors. *The Journal of Physical Chemistry Letters* **2016**, *7*, 5198–5204.

[92] Jurecka, P.; poner, J.; Cern, J.; Hobza, P. Benchmark database of accurate (MP2 and CCSD(T) complete basis set limit) interaction energies of small model complexes, DNA base pairs, and amino acid pairs. *Phys. Chem. Chem. Phys.* **2006**, *8*, 1985–1993.

[93] Silva, H. S.; Sodero, A. C. R.; Bouyssiere, B.; Carrier, H.; Korb, J.-P.; Alfarra, A.; Vallverdu, G.; B gu, D.; Baraille, I. Molecular Dynamics Study of Nanoaggregation in Asphaltene Mixtures: Effects of the N, O, and S Heteroatoms. *Energy & Fuels* **2016**, *30*, 5656–5664.

[94] Sheu, E. Y. Petroleum AsphalteneProperties, Characterization, and Issues. *Energy & Fuels* **2002**, *16*, 74–82.

[95] Gray, M. R. *Upgrading Oilsands Bitumen and Heavy Oil*; University of Alberta Press: Edmonton, 2015.

[96] Haji-Akbari, N.; Teeraphapkul, P.; Fogler, H. S. Effect of Asphaltene Concentration on the Aggregation and Precipitation Tendency of Asphaltenes. *Energy & Fuels* **2014**, *28*, 909–919.

[97] Subramanian, S.; bastien Simon, S.; blom, J. S. Asphaltene Precipitation Models: A Review. *Journal of Dispersion Science and Technology* **2016**, *37*, 1027–1049.

[98] Robertson, J. M.; White, J. G. 72. The crystal structure of pyrene. A quantitative X-ray investigation. *J. Chem. Soc.* **1947**, 358–368.

[99] Wang, Z.; Enkelmann, V.; Negri, F.; M llen, K. Rational Design of Helical Columnar Packing in Single Crystals. *Angewandte Chemie International Edition* **2004**, *43*, 1972–1975.

[100] Zreid, M.; Tabasi, Z. A.; Zhao, Y. Comparative studies of the noncovalent interactions in the single-crystal packing of pyrene, pyrene-4,5-dione, and pyrene-4,5,9,10-tetraone. *Journal of Physical Organic Chemistry* **2021**, *34*, e4192.

[101] Schuetz, C. A.; Frenklach, M. Nucleation of soot: Molecular dynamics simulations of pyrene dimerization. *Proceedings of the Combustion Institute* **2002**, *29*, 2307–2314.

[102] Gonzalez, C.; Lim, E. C. Evaluation of the Hartree-Fock Dispersion (HFD) Model as a Practical Tool for Probing Intermolecular Potentials of Small Aromatic Clusters: Comparison of the HFD and MP2 Intermolecular Potentials. *The Journal of Physical Chemistry A* **2003**, *107*, 10105–10110.

[103] Grimme, S. Accurate description of van der Waals complexes by density functional theory including empirical corrections. *Journal of Computational Chemistry* **2004**, *25*, 1463–1473.

[104] Lee, N. K.; Kim, S. K. Ab initio-based intermolecular carbon carbon pair potentials for polycyclic aromatic hydrocarbon clusters. *The Journal of Chemical Physics* **2005**, *122*, 031102.

[105] von Lilienfeld, O. A.; Andrienko, D. Coarse-grained interaction potentials for polyaromatic hydrocarbons. *The Journal of Chemical Physics* **2006**, *124*, 054307.

[106] Herdman, J. D.; Miller, J. H. Intermolecular Potential Calculations for Polynuclear Aromatic Hydrocarbon Clusters. *The Journal of Physical Chemistry A* **2008**, *112*, 6249–6256.

[107] Totton, T. S.; Misquitta, A. J.; Kraft, M. A First Principles Development of a General Anisotropic Potential for Polycyclic Aromatic Hydrocarbons. *Journal of Chemical Theory and Computation* **2010**, *6*, 683–695.

[108] Totton, T. S.; Misquitta, A. J.; Kraft, M. A quantitative study of the clustering of polycyclic aromatic hydrocarbons at high temperatures. *Phys. Chem. Chem. Phys.* **2012**, *14*, 4081–4094.

[109] Elvati, P.; Violi, A. Thermodynamics of poly-aromatic hydrocarbon clustering and the effects of substituted aliphatic chains. *Proceedings of the Combustion Institute* **2013**, *34*, 1837–1843.

[110] Silva, N. J.; Machado, F. B. C.; Lischka, H.; Aquino, A. J. A. π-π stacking between polyaromatic hydrocarbon sheets beyond dispersion interactions. *Phys. Chem. Chem. Phys.* **2016**, *18*, 22300–22310.

[111] Takeuchi, H. Structures, stability, and growth sequence patterns of small homoclusters of naphthalene, anthracene, phenanthrene, phenalene, naphthacene, and pyrene. *Computational and Theoretical Chemistry* **2013**, *1021*, 84–90, Clusters: From Dimers to Nanoparticles.

[112] Hoche, J.; Schmitt, H.-C.; Humeniuk, A.; Fischer, I.; Mitric, R.; R hr, M. I. S. The mechanism of excimer formation: an experimental and theoretical study on the pyrene dimer. *Phys. Chem. Chem. Phys.* **2017**, *19*, 25002–25015.

[113] Chakraborty, D.; Lischka, H.; Hase, W. L. Dynamics of Pyrene-Dimer Association and Ensuing Pyrene-Dimer Dissociation. *The Journal of Physical Chemistry A* **2020**, *124*, 8907–8917.

[114] Elvati, P.; Violi, A. Thermodynamics of poly-aromatic hydrocarbon clustering and the effects of substituted aliphatic chains. *Proceedings of the Combustion Institute* **2013**, *34*, 1837–1843.

[115] Rapacioli, M.; Spiegelman, F.; Talbi, D.; Mineva, T.; Goursot, A.; Heine, T.; Seifert, G. Correction for dispersion and Coulombic interactions in molecular clusters with density functional derived methods: Application to polycyclic aromatic hydrocarbon clusters. *The Journal of Chemical Physics* **2009**, *130*, 244304.

[116] Najibi, A.; Goerigk, L. DFT-D4 counterparts of leading meta-generalized-gradient approximation and hybrid density functionals for energetics and geometries. *Journal of Computational Chemistry* **2020**, *41*, 2562–2572.

[117] Weigend, F.; Ahlrichs, R. Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: Design and assessment of accuracy. *Phys. Chem. Chem. Phys.* **2005**, *7*, 3297–3305.

[118] Barone, V.; Cossi, M. Quantum Calculation of Molecular Energies and Energy Gradients in Solution by a Conductor Solvent Model. *The Journal of Physical Chemistry A* **1998**, *102*, 1995–2001.

[119] Marenich, A. V.; Cramer, C. J.; Truhlar, D. G. Universal Solvation Model Based on Solute Electron Density and on a Continuum Model of the Solvent Defined by the Bulk Dielectric Constant and Atomic Surface Tensions. *The Journal of Physical Chemistry B* **2009**, *113*, 6378–6396.

[120] Grimme, S. Supramolecular Binding Thermodynamics by Dispersion-Corrected Density Functional Theory. *Chemsitry - A European Journal* **2012**, *18*, 9955–9964.

[121] Dunning, T. H. Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen. *The Journal of Chemical Physics* **1989**, *90*, 1007–1023.

[122] Bryantsev, V. S.; Diallo, M. S.; Goddard III, W. A. Calculation of Solvation Free Energies of Charged Solutes Using Mixed Cluster/Continuum Models. *The Journal of Physical Chemistry B* **2008**, *112*, 9709–9719.

[123] Ribeiro, R. F.; Marenich, A. V.; Cramer, C. J.; Truhlar, D. G. Use of Solution-Phase Vibrational Frequencies in Continuum Models for the Free Energy of Solvation. *The Journal of Physical Chemistry B* **2011**, *115*, 14556–14562.

[124] Patel, D. H.; East, A. L. L. Semicontinuum (Cluster-Continuum) Modeling of Acid-Catalyzed Aqueous Reactions: Alkene Hydration. *The Journal of Physical Chemistry A* **2020**, *124*, 9088–9104.

[125] Riplinger, C.; Neese, F. An efficient and near linear scaling pair natural orbital based local coupled cluster method. *The Journal of Chemical Physics* **2013**, *138*, 034106.

[126] Riplinger, C.; Sandhoefer, B.; Hansen, A.; Neese, F. Natural triple excitations in local coupled cluster calculations with pair natural orbitals. *The Journal of Chemical Physics* **2013**, *139*, 134101.

[127] Riplinger, C.; Pinski, P.; Becker, U.; Valeev, E. F.; Neese, F. Sparse maps A systematic infrastructure for reduced-scaling electronic structure methods. II. Linear scaling domain based pair natural orbital coupled cluster theory. *The Journal of Chemical Physics* **2016**, *144*, 024109.

[128] Weigend, F.; K hn, A.; H ttig, C. Efficient use of the correlation consistent basis sets in resolution of the identity MP2 calculations. *The Journal of Chemical Physics* **2002**, *116*, 3175–3183.

[129] Weigend, F. Accurate Coulomb-fitting basis sets for H to Rn. *Phys. Chem. Chem. Phys.* **2006**, *8*, 1057–1065.

[130] Garcia-Rat s, M.; Becker, U.; Neese, F. Implicit solvation in domain based pair natural orbital coupled cluster (DLPNO-CCSD) theory. *Journal of Computational Chemistry* **2021**, *42*, 1959–1973.

[131] Liakos, D. G.; Sparta, M.; Kesharwani, M. K.; Martin, J. M. L.; Neese, F. Exploring the Accuracy Limits of Local Pair Natural Orbital Coupled-Cluster Theory. *Journal of Chemical Theory and Computation* **2015**, *11*, 1525–1539.

[132] Neese, F. Software update: the ORCA program system, version 4.0. *WIREs Computational Molecular Science* **2018**, *8*, e1327.

[133] Helmich-Paris, B.; de Souza, B.; Neese, F.; Izs k, R. b. An improved chain of spheres for exchange algorithm. *The Journal of Chemical Physics* **2021**, *155*, 104109.

[134] Grimme, S. Exploration of Chemical Compound, Conformer, and Reaction Space with Meta-Dynamics Simulations Based on Tight-Binding Quantum Chemical Calculations. *Journal of Chemical Theory and Computation* **2019**, *15*, 2847–2862.

[135] Pracht, P. Private Communication, 2021.

[136] Powell, J. R.; Voisinet, D.; Salazar, A.; Jr., W. E. A. Solubility of Pyrene in Organic Nonelectrolyte Solvents. Comparison of Observed Versus Predicted Values Based Upon Mobile Order Theory. *Physics and Chemistry of Liquids* **1994**, *28*, 269–276.

[137] Sch tz, M.; Rauhut, G.; Werner, H.-J. Local Treatment of Electron Correlation in Molecular Clusters: Structures and Stabilities of H2O(n), n = 2-4. *The Journal of Physical Chemistry A* **1998**, *102*, 5997–6003.

[138] Schwab, C. H. Conformations and 3D pharmacophore searching. *Drug Discovery Today: Technologies* **2010**, *7*, e245–e253, 3D Pharmacophore Elucidation and Virtual Screening.

[139] Hawkins, P. C. D. Conformation Generation: The State of the Art. *Journal of Chemical Information and Modeling* **2017**, *57*, 1747–1756.

[140] Riniker, S.; Landrum, G. A. Better Informed Distance Geometry: Using What We Know To Improve Conformation Generation. *Journal of Chemical Information and Modeling* **2015**, *55*, 2562–2574.

[141] Supady, A.; Blum, V.; Baldauf, C. First-Principles Molecular Structure Search with a Genetic Algorithm. *Journal of Chemical Information and Modeling* **2015**, *55*, 2338–2348.

[142] Ajagekar, A.; Humble, T.; You, F. Quantum computing based hybrid solution strategies for large-scale discrete-continuous optimization problems. *Computers & Chemical Engineering* **2020**, *132*, 106630.

[143] Batista Jr., J. o. M.; Blanch, E. W.; Bolzani, V. d. S. Recent advances in the use of vibrational chiroptical spectroscopic methods for stereochemical characterization of natural products. *Nat. Prod. Rep.* **2015**, *32*, 1280–1302.

[144] Demarque, D. P.; Kemper, M.; Merten, C. VCD spectroscopy reveals that a water molecule determines the conformation of azithromycin in solution. *Chem. Commun.* **2021**, *57*, 4031–4034.

[145] Yang, Q.; Carlson, C. D.; J ger, W.; Xu, Y. Conformational Landscape of the Hydrogen-Bonded 1-Phenyl-2,2,2-Trilfuoroethanol 1,4-Dioxane Complex: Dispersion Interactions and Conformational Conversion. *The Journal of Physical Chemistry A* **2022**, *126*, 2942–2949.

[146] Carlson, C. D.; Hazrah, A. S.; Mason, D.; Yang, Q.; Seifert, N. A.; Xu, Y. Alternating 1-Phenyl-2,2,2-Trifluoroethanol Conformational Landscape With the Addition of One Water: Conformations and Large Amplitude Motions. *The Journal of Physical Chemistry A* **2022**, *126*, 7250–7260.

[147] Grimme, S.; Bohle, F.; Hansen, A.; Pracht, P.; Spicher, S.; Stahn, M. Efficient Quantum Chemical Calculation of Structure Ensembles and Free Energies for Nonrigid Molecules. *The Journal of Physical Chemistry A* **2021**, *125*, 4039–4054.

[148] Watts, K. S.; Dalal, P.; Murphy, R. B.; Sherman, W.; Friesner, R. A.; Shelley, J. C. ConfGen: A Conformational Search Method for Efficient Generation of Bioactive Conformers. *Journal of Chemical Information and Modeling* **2010**, *50*, 534–546.

[149] Hawkins, P. C. D.; Skillman, A. G.; Warren, G. L.; Ellingson, B. A.; Stahl, M. T. Conformer Generation with OMEGA: Algorithm and Validation Using High Quality Structures from the Protein Databank and Cambridge Structural Database. *Journal of Chemical Information and Modeling* **2010**, *50*, 572–584.

[150] Friedrich, N.-O.; Flachsenberg, F.; Meyder, A.; Sommer, K.; Kirchmair, J.; Rarey, M. Conformator: A Novel Method for the Generation of Conformer Ensembles. *Journal of Chemical Information and Modeling* **2019**, *59*, 731–742.

[151] Poli, G.; Seidel, T.; Langer, T. Conformational Sampling of Small Molecules With iCon: Performance Assessment in Comparison With OMEGA. *Frontiers in Chemistry* **2018**, *6*.

[152] Sperandio, O.; Souaille, M.; ois Delfaud, F.; Miteva, M. A.; Villoutreix, B. O. MED-3DMC: A new tool to generate 3D conformation ensembles of small molecules with a Monte Carlo sampling of the conformational space. *European Journal of Medicinal Chemistry* **2009**, *44*, 1405–1409.

[153] Chang, G.; Guida, W. C.; Still, W. C. An internal-coordinate Monte Carlo method for searching conformational space. *Journal of the American Chemical Society* **1989**, *111*, 4379–4386.

[154] Tsujishita, H.; Hirono, S. Camdas: An Automated Conformational Analysis System Using Molecular Dynamics. *Journal of Computer-Aided Molecular Design* **1997**, *11*, 305–315.

[155] Malloci, G.; Serra, G.; Bosin, A.; Vargiu, A. V. Extracting Conformational Ensembles of Small Molecules from Molecular Dynamics Simulations: Ampicillin as a Test Case. *Computation* **2016**, *4*.

[156] Iribarren, I. i.; Trujillo, C. Efficiency and Suitability when Exploring the Conformational Space of Phase-Transfer Catalysts. *Journal of Chemical Information and Modeling* **2022**, *62*, 5568–5580.

[157] Axelrod, S.; Gomez-Bombarelli, R. GEOM, energy-annotated molecular conformations for property predictions and molecular generation. *Scientific Data* **2022**, *9*, 185.

[158] Eik s, K. D. R.; Beerepoot, M. T. P.; Ruud, K. A Computational Protocol for Vibrational Circular Dichroism Spectra of Cyclic Oligopeptides. *The Journal of Physical Chemistry A* **2022**, *126*, 5458–5471.

[159] King, N. J.; Brown, A. Computational Investigations Complement Experiment for a System of Non-Covalently Bound Asphaltene Model Compounds. *Energy & Fuels* **2023**, *37*, 12796–12810.

[160] Brynn Hibbert, D.; Thordarson, P. The death of the Job plot, transparency, open science and online tools, uncertainty estimation methods and other developments in supramolecular chemistry data analysis. *Chem. Commun.* **2016**, *52*, 12792–12805.

[161] King, N. J.; LeBlanc, I. D.; Brown, A. A Variant on the CREST Algorithm for Non-Covalent Clusters of Flexible Molecules. 2023; https://chemrxiv.org/engage/chemrxiv/article-details/-646b83b4b3dd6a65308e7595, This content is a preprint and has not been peer-reviewed.

[162] Goerigk, L.; Hansen, A.; Bauer, C.; Ehrlich, S.; Najibi, A.; Grimme, S. A look at the density functional theory zoo with the advanced GMTKN55 database for general main group thermochemistry, kinetics and noncovalent interactions. *Phys. Chem. Chem. Phys.* **2017**, *19*, 32184–32215.

[163] Mardirossian, N.; Head-Gordon, M. ?B97X-V: A 10-parameter, range-separated hybrid, generalized gradient approximation density functional with nonlocal correlation, designed by a survival-of-the-fittest strategy. *Phys. Chem. Chem. Phys.* **2014**, *16*, 9904–9924.

[164] Jensen, F. Segmented Contracted Basis Sets Optimized for Nuclear Magnetic Shielding. *Journal of Chemical Theory and Computation* **2015**, *11*, 132–138.

[165] Tao, J.; Perdew, J. P.; Staroverov, V. N.; Scuseria, G. E. Climbing the Density Functional Ladder: Nonempirical Meta–Generalized Gradient Approximation Designed for Molecules and Solids. *Phys. Rev. Lett.* **2003**, *91*, 146401.

[166] Stoychev, G. L.; Auer, A. A.; Neese, F. Efficient and Accurate Prediction of Nuclear Magnetic Resonance Shielding Tensors with Double-Hybrid Density Functional Theory. *Journal of Chemical Theory and Computation* **2018**, *14*, 4756–4771.

[167] King, N. J.; Brown, A. Intermolecular Interactions of Pyrene and Its Oxides in Toluene Solution. *The Journal of Physical Chemistry A* **2022**, *126*, 4931–4940.

[168] Simm, G. N.; T rtscher, P. L.; Reiher, M. Systematic microsolvation approach with a cluster-continuum scheme and conformational sampling. *J. Comput. Chem.* **2020**, *41*, 1144–1155.

[169] Inc., T. M. MATLAB version: 9.13.0 (R2022b). 2022; https://www.mathworks.com.

[170] Sure, R.; Grimme, S. Corrected small basis set Hartree-Fock method for large systems. *Journal of Computational Chemistry* **2013**, *34*, 1672–1685.

[171] Brandenburg, J. G.; Bannwarth, C.; Hansen, A.; Grimme, S. B97-3c: A revised low-cost variant of the B97-D density functional method. *The Journal of Chemical Physics* **2018**, *148*, 064104.

[172] Grimme, S.; Hansen, A.; Ehlert, S.; Mewes, J.-M. r2SCAN-3c: A Swiss army knife composite electronic-structure method. *The Journal of Chemical Physics* **2021**, *154*, 064103.

[173] King, N. J.; Ortlieb, J.; Brown, A. unpublished.

[174] Scott, D. E.; Aloisio, M. D.; Rodriguez, J. F.; Morimoto, M.; Hamilton, R. J.; Brown, O.; Tykwinski, R. R.; Stryker, J. M. Optimizing the Iodide/Iodonium/O2 Oxidation Cycle Enhances the Scope, Selectivity, and Yields of Hydroiodic Acid-Catalyzed Multicomponent Cyclocondensation Reactions. *Advanced Synthesis & Catalysis* **2021**, *363*, 4720–4727.

[175] Guo, Y. Archipelago Model Compound Synthesis by Rhodium-catalyzed Annulations of Island-tethered Alkynes with Aromatic Substrates. Ph.D. thesis, University of Alberta, 2022.

[176] Celia-Silva, L. G.; Martins, R. N.; Carvalho, A. J. P.; Ramalho, J. P. P.; Morgado, P.; Filipe, E. J.; Martins, L. F. Influence of Ionic Liquids on the Aggregation and Pre-aggregation Phenomena of Asphaltenes in Model Solvent Mixtures by Molecular Dynamics Simulations and Quantum Mechanical Calculations. *Energy & Fuels* **2022**, *36*, 9048–9065.

[177] Ali, A.; Cole, D. R.; Striolo, A. Understanding the Aggregation of Model Island and Archipelago Asphaltene Molecules near Kaolinite Surfaces using Molecular Dynamics. *Energy & Fuels* **2023**, *37*, 11662–11674.

[178] Moncayo-Riascos, I.; Rojas-Ruiz, F. A.; Orrego-Ruiz, J. A.; Cundar, C.; Torres, R. G.; CaÃ±as-MarÃ-n, W. Reconstruction of a Synthetic Crude Oil Using Petroleomics and Molecular Dynamics Simulations: A Multistructural Approach to Understanding Asphaltene Aggregation Behavior. *Energy & Fuels* **2022**, *36*, 837–850.

[179] Ying, J.; Zhao, H.; Wang, Z.; An, K.; Cao, Q.; Li, C.; Jia, J.; Zhang, Z.; Liu, X. Effect of asphaltene structure characteristics on asphaltene accumulation at oil-water interface: An MD simulation study. *Colloids and Surfaces A: Physicochemical and Engineering Aspects* **2023**, *675*, 132014.

[180] Schneider, W. B.; Bistoni, G.; Sparta, M.; Saitow, M.; Riplinger, C.; Auer, A. A.; Neese, F. Decomposition of Intermolecular Interaction Energies within the Local Pair Natural Orbital Coupled Cluster Framework. *Journal of Chemical Theory and Computation* **2016**, *12*, 4778–4792.

Appendix A

Appendix A. Supporting Information for Chapter 2

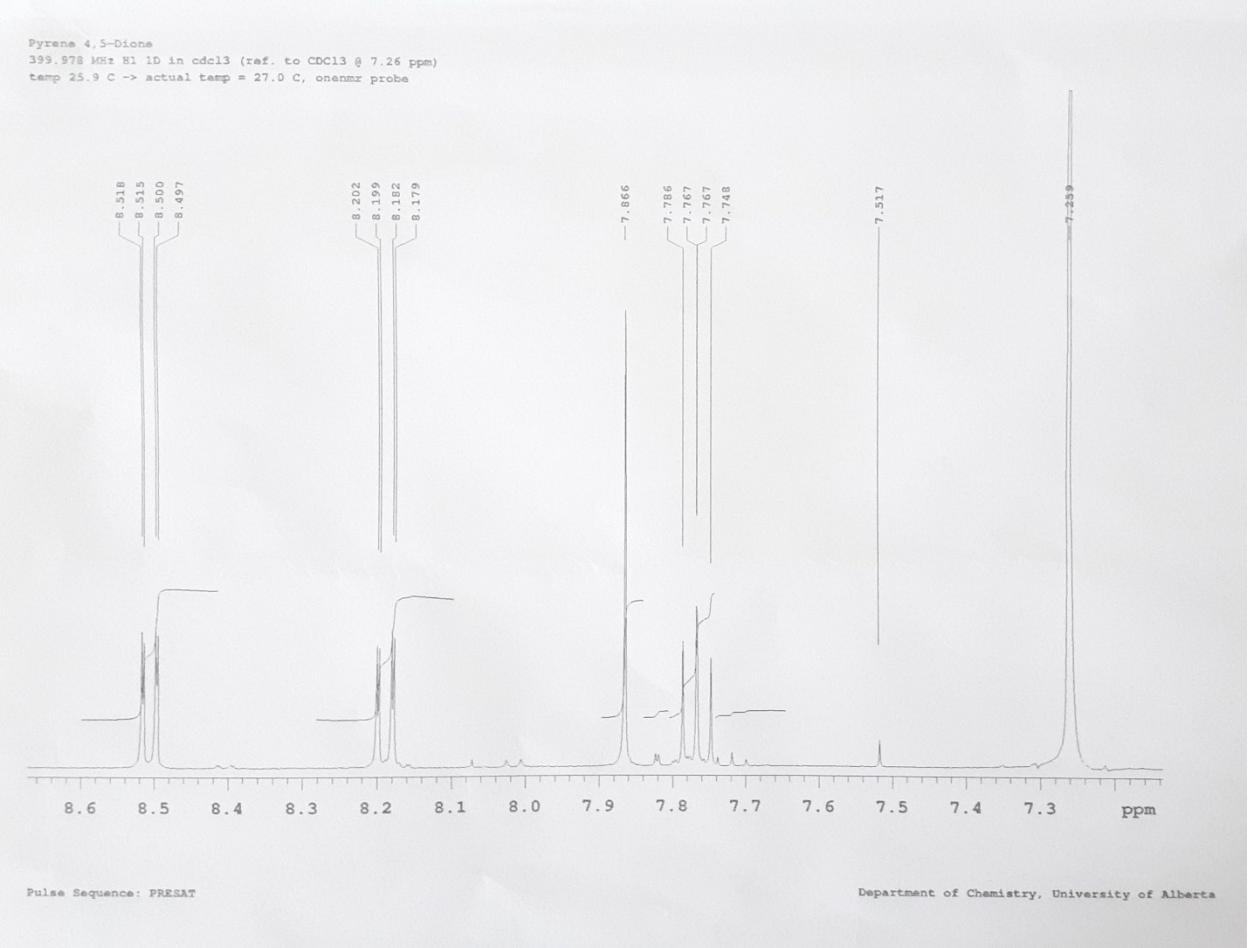


Figure A.1 ^{1}H NMR spectrum of pyrene-4,5-dione in $CDCl_3$, as produced by the procedure in Chapter 2. Pyrene-4,5-dione peaks occur at 8.51 ppm (dd, J = 7.2 Hz, 1.2 Hz), 8.19 ppm (dd, J = 8.0 Hz, 1.2 Hz), 7.87 ppm (s), and 7.77 ppm (dd, J = 7.2 Hz, 8.0 Hz). Peaks corresponding to pyrene occur at 8.17 ppm (dd) (partly obscured by pyrene-4,5-dione signal), 8.07 ppm (s), and 8.01 ppm (t), while those corresponding to pyrene-4,5,9,10-tetraone occur at 8.40 ppm (dd) and 7.72 ppm (t). Peaks at 7.82 ppm and 7.52 ppm are not assigned.

Appendix B

Appendix B. Supporting Information for Chapter 3

B.1 Gas Phase Ensembles

Previous studies of the $(pyrene)_2$ ensemble have been done in the gas phase.[54, 87, 100-112] While the ensemble that exists in toluene solution is of primary interest, the gas phase energies are necessary to isolate the solvation energy, and are useful for comparison with previous results. Reoptimization of the solution-phase ensemble using ωB97X-D4/def2-TZVP (gas phase calculation) gives the gas-phase ensemble. Most of the conformers do not appear to be qualitatively different, but there is a small amount of geometric relaxation in each. The exception is 67qG, which rotates even further away from ideal graphitic stacking in the gas phase, becoming 71qG. Excluding 71qG, the average RMSD between gas-phase and solution-phase (CPCM) geometries is 0.012 Å, and for 71qG, it is 0.309 Å. The energies of dimerization for the gas-phase ensemble are given in Table B.1. For the 5 previously identified conformers (0SP-L, 0SP-S, 0G, 60G1, and 90X), our results agree well (within 2 kJ/mol) with previously published results.[54, 102, 115]

Table B.1 Ensemble for $(pyrene)_2$, in the gas phase, with energies determined at the ωB97X-D4/def2-TZVP level of theory, and corresponding probabilities (p_i), see Eq. 3.11. Asterisk indicates the structure which best corresponds to crystal structure.

$(Pyrene)_2$	$\Delta E_{gas,dim}$ (kJ/mol)	$\Delta G_{gas,dim}$ (kJ/mol)	$e^{(-\Delta G/RT)}$	degen.	p_i
0SP-L*	-53.92	9.78	0.0193	16	0.085
60G1	-53.17	9.43	0.0177	32	0.392
60G2	-53.38	9.99	0.0223	64	0.156
90X	-48.47	13.20	0.0049	8	0.011
0SP-S	-52.32	10.49	0.0145	16	0.064
0G	-52.49	9.71	0.0199	32	0.175
71qG	-49.76	12.43	0.0066	64	0.117

We find an $S_{conf,gas}$ of 13.69 J/(mol K), and a $\Delta G_{gas,dim,ens}$ of 5.97 kJ/mol, indicating that dimerization of pyrene at 1 atm and 298K is also mildly unfavourable in the gas phase.

The $(pyrene\ dione)_2$ gas phase ensemble (see Table B.2) is similar to that in solution, but the loss of stabilization of the monomers that solution phase provided allows the dimerization free energies of some conformers to become negative. Thus, dimerization is favourable at 1 atm and 298K. After

ensemble averaging and conformational entropy are taken into account, $\Delta G_{gas,dim,ens}$ for (pyrene dione)$_2$ is -10.61 kJ/mol.

Table B.2 Ensemble for (pyrene-4,5-dione)$_2$, in the gas phase, with energies determined at the ωB97X-D4/def2-TZVP level of theory, and corresponding probabilities (p_i), see Eq. 3.11. Asterisk indicates the structure which best corresponds to crystal structure.

(Pyrene dione)$_2$	$\Delta E_{gas,dim}$ (kJ/mol)	$\Delta G_{gas,dim}$ (kJ/mol)	$e^{(-\Delta G/RT)}$	degen.	p_i
180G*	-72.62	-6.35	12.95	8	0.172
120G	-75.83	-8.98	37.48	8	0.499
180SP-L	-73.68	-7.33	19.29	8	0.258
60G1	-67.53	-1.11	1.56	16	0.042
60G2	-62.23	2.80	0.323	16	0.009
150SR	-67.31	-0.73	1.34	8	0.018
0G	-59.96	4.94	0.136	16	0.004
45SR	-51.50	11.97	0.008	16	0.000
180SP-E	-42.88	17.43	0.001	8	0.000

Table B.3 Ensemble for (pyrene-4,5,9,10-tetraone)$_2$, in the gas phase, with energies determined at the ωB97X-D4/def2-TZVP level of theory, and corresponding probabilities (p_i), see Eq. 3.11. Asterisk indicates the structure which best corresponds to crystal structure.

(Pyrene tetraone)$_2$	$\Delta E_{gas,dim}$ (kJ/mol)	$\Delta G_{gas,dim}$ (kJ/mol)	$e^{(-\Delta G/RT)}$	degen.	p_i
0G*	-69.09	-2.00	2.24	32	0.003
90SR1	-84.59	-15.65	554.5	16	0.376
90SR2	-65.77	-2.33	2.56	32	0.003
60G1	-80.81	-15.14	449.9	32	0.610
30SR	-71.39	-4.41	5.93	32	0.008
0SP-E	-48.14	10.97	0.0119	32	0.000

Table B.4 Ensemble for pyrene/pyrene-4,5,9,10-tetraone heterodimer, in the gas phase, with energies determined at the ωB97X-D4/def2-TZVP level of theory, and corresponding probabilities (p_i), see Eq. 3.11.

Pyrene/pyrene tetraone	$\Delta E_{gas,dim}$ (kJ/mol)	$\Delta G_{gas,dim}$ (kJ/mol)	$e^{(-\Delta G/RT)}$	degen.	p_i
30SR	-79.48	-12.26	140.6	32	0.443
50X	-78.84	-11.73	113.9	16	0.189
60X	-79.31	-11.85	119.6	16	0.188
60G2	-76.61	-10.14	60.0	32	0.180

Similar trends hold for (pyrene tetraone)$_2$ (see Table B.3) and pyrene/pyrene tetraone (see Table B.4), but for each ensemble, one solution phase conformer is not a minimum on the gas phase potential energy surface. For (pyrene tetraone)$_2$, the 60G2 conformer collapses to 90SR2, and for pyrene/pyrene tetraone, the 60G1 conformer collapses to 50X. $\Delta G_{gas,dim,ens}$ for (pyrene tetraone)$_2$ is -17.01 kJ/mol, and for pyrene/pyrene tetraone it is -14.90 kJ/mol.

Table B.5 Ensemble for (pyrene)$_2$, in toluene solution, with interplane distances (d) in Angstrom, energies determined at the ωB97X-D4/def2-TZVP level of theory, and corresponding probabilities (p_i), see Eq. 3.11. Asterisk indicates the structure which best corresponds to crystal structure.

(Pyrene)$_2$	d (Å)	$\Delta E_{solv,dim}$ (kJ/mol)	$\Delta G_{solv,dim}$ (kJ/mol)	$e^{(-\Delta G/RT)}$	degen.	p_i
0SP-L*	3.34	-33.47	20.99	2.09×10^{-4}	16	0.076
60G1	3.35	-32.98	21.18	1.94×10^{-4}	32	0.142
60G2	3.33	-33.01	20.49	2.56×10^{-4}	64	0.374
90X	3.45	-29.28	22.73	1.04×10^{-4}	8	0.019
0SP-S	3.34	-31.99	21.69	1.57×10^{-4}	16	0.057
0G	3.76	-32.49	20.60	2.45×10^{-4}	32	0.179
67qG	3.82	-30.33	22.71	1.05×10^{-4}	64	0.153

Table B.6 Ensemble for (pyrene-4,5-dione)$_2$, in toluene solution, with interplane distances (d) in Angstrom, energies determined at the ωB97X-D4/def2-TZVP level of theory, and corresponding probabilities (p_i), see Eq. 3.11. Asterisk indicates the structure which best corresponds to crystal structure.

(Pyrene dione)$_2$	d (Å)	$\Delta E_{solv,dim}$ (kJ/mol)	$\Delta G_{solv,dim}$ (kJ/mol)	$e^{(-\Delta G/RT)}$	degen.	p_i
180G*	3.28	-46.09	11.94	8.07×10^{-3}	8	0.131
120G	3.30	-48.90	9.61	2.07×10^{-2}	8	0.339
180SP-L	3.29	-47.88	9.52	2.14×10^{-2}	8	0.350
60G1	3.31	-43.53	14.58	2.78×10^{-3}	16	0.091
60G2	3.27	-39.61	17.39	8.93×10^{-4}	16	0.029
150SR	3.24	-41.76	16.04	1.55×10^{-3}	8	0.025
0G	3.31	-39.80	17.46	8.71×10^{-4}	16	0.028
45SR	3.26	-32.63	22.11	1.33×10^{-4}	16	0.004
180SP-E	2.96	-27.41	25.01	4.13×10^{-5}	8	0.001

B.2 Solution Phase Ensembles

The conformational ensemble for (pyrene)$_2$ is presented in Table B.5,while those for the (pyrene-4,5-dione)$_2$, (pyrene-4,5,9,10-tetraone)$_2$, and pyrene-pyrene-4,5,9,10-tetraone dimers are given in Tables B.6, B.7, and B.8, respectively. The intermolecular separation (distance between best-fit planes), binding energy ($\Delta E_{solv,dim}$) and Gibbs free energy of dimerization (ΔG_{dim}) for each conformer is

given. Also provided is the normalized probability for each conformer, as determined by Eq. 3.11. The monomers remain quite planar in all dimers, with pyrene molecules having an RMSD from best plane of 0.06 Å or better in all cases. The pyrenedione and pyrenetetraone monomers are somewhat less planar, being 0.10 Å or better and 0.11 Å or better, respectively, but this is not surprising, as the carbonyls are known to deviate somewhat from the best plane.[100] Ignoring the oxygen atoms improves the RMSD from best plane to 0.05 Å or better for pyrenedione, and 0.06 Å or better for pyrenetetraone.

Table B.7 Ensemble for (pyrene-4,5,9,10-tetraone)$_2$, in toluene solution, with interplane distances (d) in Angstrom, energies determined at the ωB97X-D4/def2-TZVP level of theory, and corresponding probabilities (p_i), see Eq. 3.11. Asterisk indicates the structure which best corresponds to crystal structure.

(Pyrene tetraone)$_2$	d (Å)	$\Delta E_{solv,dim}$ (kJ/mol)	$\Delta G_{solv,dim}$ (kJ/mol)	$e^{(-\Delta G/RT)}$	degen.	p_i
0G*	3.22	-46.41	15.81	1.70×10^{-3}	32	0.026
90SR1	3.20	-55.18	8.20	3.65×10^{-2}	16	0.282
90SR2	3.01	-40.49	17.71	7.85×10^{-4}	32	0.012
60G1	3.26	-52.35	8.32	3.49×10^{-2}	32	0.539
60G2	3.21	-46.07	13.92	3.63×10^{-3}	64	0.112
30SR	3.21	-45.54	15.73	1.75×10^{-3}	32	0.027
0SP-E	2.95	-30.49	23.19	8.61×10^{-5}	32	0.001

Table B.8 Ensemble for the pyrene/pyrene-4,5,9,10-tetraone heterodimer, in toluene solution, with interplane distances (d) in Angstrom, energies determined at the ωB97X-D4/def2-TZVP level of theory, and corresponding probabilities (p_i), see Eq. 3.11.

Pyrene/pyrene tetraone	d (Å)	$\Delta E_{solv,dim}$ (kJ/mol)	$\Delta G_{solv,dim}$ (kJ/mol)	$e^{(-\Delta G/RT)}$	degen.	p_i
30SR	3.23	-53.01	6.84	0.0632	32	0.395
50X	3.27	-52.00	8.05	0.0388	16	0.121
60X	3.26	-52.18	8.21	0.0364	16	0.114
60G1	3.27	-51.03	8.48	0.0267	32	0.121
60G2	3.14	-50.44	8.98	0.0327	32	0.166

B.3 Local Energy Decomposition

Local Energy Decomposition (LED) is a technique which breaks down an interaction energy into components. In ORCA, the default algorithm is the Foster-Boys algorithm,[180] which decomposes an interaction energy into $\Delta E_{int,elst}$ (electrostatics), $\Delta E_{int,exch}$ (exchange), $\Delta E_{int,disp}$ (dispersion), $\Delta E_{int,corr}$ (non-dispersive correlation), and $\Delta E_{int,elprep}$ (electronic preparation). $\Delta\Delta G_{solv,dim}$ is calculated via Equation 3.4, and $\Delta E_{int,elprep}$ is determined by subtraction:

$$\Delta E_{int,elprep} = \Delta E_{solv,dim} - \Delta\Delta G_{solv,dim} - \Delta E_{int,elst} - \Delta E_{int,exch} - \Delta E_{int,disp} - \Delta E_{int,corr} \quad \text{(B.1)}$$

Table B.9 gives the energy breakdowns for the six conformers studied using DLPNO-CCSD(T).

Table B.9 LED of selected conformers.

Conformer	$\Delta E_{solv,dim}$ (kJ/mol)	$\Delta\Delta G_{solv,dim}$ (kJ/mol)	$\Delta E_{int,elst}$ (kJ/mol)	$\Delta E_{int,exch}$ (kJ/mol)	$\Delta E_{int,disp}$ (kJ/mol)	$\Delta E_{int,corr}$ (kJ/mol)	$\Delta E_{int,elprep}$ (kJ/mol)
$(pyrene)_2$ 0PD-L	-35.00	19.90	-138.26	-69.61	-99.57	-47.10	319.54
$(pyrene\ dione)_2$ 180G	-47.94	25.09	-198.79	-79.51	-109.01	-50.16	389.52
$(pyrene\ dione)_2$ 120G	-50.17	25.47	-221.20	-89.13	-114.43	-55.07	429.66
$(pyrene\ tetraone)_2$ 0G	-48.27	22.29	-228.45	-92.05	-115.44	-55.64	443.31
$(pyrene\ tetraone)_2$ 90SR	-57.37	26.43	-270.37	-101.13	-121.50	-60.04	497.67
pyrene/pyrene tetraone	-53.14	26.67	-251.59	-97.56	-115.30	-59.60	470.91

B.4 CREST vs. DFT Ensemble Comparison

Table B.10 shows the conformers of the DFT ensemble of $(pyrene)_2$, and the closest matches (by RMSD) from the CREST ensemble.

Table B.10 DFT vs. CREST Ensemble Comparison. Each DFT conformer is paired with the most similar CREST conformer, and an RMSD is calculated. Energy differences are calculated as DFT Energy - CREST Energy.

Conformer	DFT Energy (kJ/mol)	CREST Conformer	CREST Energy (kJ/mol)	Difference (kJ/mol)	RMSD (Å)
0PD-L	-49.37	2	-45.23	4.14	0.11
60G1	-48.76	1	-45.53	3.22	0.09
60G2	-48.51	5	-44.33	4.19	0.23
90X	-44.17	7	-44.29	-0.11	0.07
0PD-S	-47.62	9	-44.12	3.50	0.25
0G	-47.89	27	-31.74	16.15	0.27
67qG	-45.11	31	-31.22	13.89	0.21

Table B.11 ΔG_{dim} and components for lowest energy conformer of each dimer. All values are in kJ/mol.

Conformer	ΔG_{dim}	$\Delta E_{SCF,gas,dim}$	$\Delta E_{disp,gas,dim}$	$\Delta G_{rrho,H,dim}$	$\Delta G_{rrho,S,dim}$	$\Delta\Delta G_{solv,dim}$
$(pyrene)_2$ 0PD-L	20.99	20.94	-74.86	4.59	57.79	12.52
$(pyrene\ dione)_2$ 120G	9.61	11.12	-86.96	5.29	61.15	19.01
$(pyrene\ tetraone)_2$ 90SR	8.20	11.14	-95.68	5.60	65.70	21.43
pyrene/pyrene tetraone 30SR	6.84	7.57	-87.05	5.80	61.99	18.54

Table B.11 shows the numerical data behind figure 3.5. From these four complexes, six conformers were selected for closer examination using DLPNO-CCSD(T)/cc-pVTZ/SMD(Toluene) at the ωB97X-D4/def2-TZVP/CPCM(Toluene) geometries, and using DLPNO-CCSD(T)/cc-pVTZ (with no solvation) at the gas phase ωB97X-D4/def2-TZVP geometries. DLPNO-CCSD(T) is among the most

accurate tractable methods for systems of this size. The conformers selected for further scrutiny were the 0SP-L of $(pyrene)_2$, the 180G of $(pyrene\text{-}4,5\text{-}dione)_2$, and the 0G of $(pyrene\text{-}4,5,9,10\text{-}tetraone)_2$ (best matches to crystal structures), along with the 120G of (pyrene-4,5-dione)2, the 90SR of (pyrene-4,5,9,10-tetraone)2, and the 30SR of pyrene/pyrene-tetraone (lowest energy in solution). For $(pyrene)_2$, 0SP-L is also the lowest-energy conformer, and for pyrene/pyrene-tetraone, there is no known experimental crystal structure.

While the errors in ωB97X-D4 energies are small, slightly more accurate numbers can be obtained by using the DLPNO-CCSD(T) energies for $\Delta E_{gas,dim}$ and $\Delta\Delta G_{solv,dim}$. Taking an ensemble average of numbers that were calculated using different methods would be meaningless, so for these calculations, the DFT-derived energies and Boltzmann weights of each conformer were used to calculate the ensemble average free energy, but were then adjusted by a correction equal to the difference in ΔG_{dim} between DLPNO-CCSD(T) and DFT for the conformer of interest. For the dimers where two conformers were evaluated using DLPNO-CCSD(T), this leads to two different sets of ensemble-corrected values, thanks to different sizes of errors between DLPNO-CCSD(T) and ωB97X-D4. One could conceivably use either value for the overall $\Delta G_{dim,ens}$, but an average is probably the best choice. The results are in Table B.12.

Table B.12 DLPNO-CCSD(T) energies of dimerization of selected conformers, broken into components. The $\Delta G_{rrho,dim}$ correction is from DFT, and the $\Delta G_{dim,ens}$ values are DFT results, corrected by the difference between ΔG_{dim} computed using DFT and ΔG_{dim} computed using DLPNO-CCSD(T).

Conformer	$\Delta E_{gas,dim}$ (kJ/mol)	$\Delta\Delta G_{solv.dim}$ (kJ/mol)	$\Delta G_{rrho,dim}$ (kJ/mol)	ΔG_{dim} (kJ/mol)	$\Delta G_{dim,ens}$ (kJ/mol)
$(pyrene)_2$ 0PD-L	-54.90	4.05	62.38	19.46	15.42
$(pyrene\ dione)_2$ 180G	-73.03	17.16	65.60	9.73	6.19
$(pyrene\ dione)_2$ 120G	-75.63	17.46	66.44	8.34	7.36
$(pyrene\ tetraone)_2$ 0G	-70.57	14.36	70.15	13.95	4.63
$(pyrene\ tetraone)_2$ 90SR	-83.79	18.50	71.31	6.01	4.30
pyrene/pyrene tetraone	-79.80	18.74	67.78	6.72	4.01

Appendix C

Appendix C. Supporting Information for Chapter 5

C.1 Details for Computations of Equilibrium Concentrations using MatLab

The equilibrium constant is defined as the concentration(s) of the product(s) divided by the concentration(s) of the reactant(s):

$$K = \frac{[product]}{[reactants]}. \quad \text{(C.1)}$$

If we set [**6**] equal to a, [**7**] equal to b, [**13**] equal to c, and [**14**] equal to d, we can substitute in each of our product and reactant concentrations, as follows:

$$K_6 = \frac{a}{[\mathbf{1}][\mathbf{2}\boldsymbol{a}]} \quad \text{(C.2)}$$

$$K_7 = \frac{b}{[\mathbf{2}\boldsymbol{b}]^2} \quad \text{(C.3)}$$

$$K_{13} = \frac{c}{[\mathbf{1}]^2[\mathbf{2}\boldsymbol{b}]} \quad \text{(C.4)}$$

$$K_{14} = \frac{d}{[\mathbf{1}]^2[\mathbf{2}\boldsymbol{b}][H_2O]}. \quad \text{(C.5)}$$

The concentrations of **1**, **2a**, and water can be set equal to their initial concentrations minus any amounts used to form complexes:

$$[\mathbf{1}] = 0.0025 - a - 2 \times c - 2 \times d \quad \text{(C.6)}$$

$$[\mathbf{2}\boldsymbol{a}] = 0.0025 - a - 2 \times b - c - d \quad \text{(C.7)}$$

$$[H_2O] = 0.0025 - d \quad \text{(C.8)}$$

Rearranging the equilibrium constant equations (C.2 – C.5), and substituting in the concentration equations (C.6 – C.8) and the values of the equilibrium constants, we get the following system of equations:

$$a = 117,800,000\ M^{-1} \times (0.0025\ M - a - 2c - 2d) \times (0.0025\ M - a - 2b - c - d) \quad \text{(C.9)}$$

$$b = 793,200,000\ M^{-1} \times (0.0025\ M - a - 2b - c - d)^2 \quad \text{(C.10)}$$

$$c = 8,566,000,000\ M^{-2} \times (0.0025\ M - a - 2c - 2d)^2 \times (0.0025\ M - a - 2b - c - d) \quad \text{(C.11)}$$

$$d = 126{,}700{,}000\, M^{-3} \times (0.0025\, M - a - 2c - 2d)^2 \times (0.0025\, M - a - 2b - c - d) \times (0.0025\, M - d) \quad \text{(C.12)}$$

Solving the system of equations using Matlab[169] gives ten solutions, but only one solution has real, positive numbers for all four concentrations. Using this solution, we get final concentrations of 2.41 mM for **6**, 43.6 µM for **7**, 41.1 µM for **13**, and 1.52 pM for **14**.

Subtracting the amounts needed to form these complexes from the initial concentrations, we can also find the equilibrium monomer concentrations to be 2.5 mM for water, 87.3 µM for **1**, and 234 nM for **2b**.

To repurpose the equations to other starting concentrations, all that needs to be changed is to substitute the new initial concentration for the 0.0025 M in the equation.

C.2 Numerical Results for Energy and NMR Chemical Shifts for the Relaxed Surface Scan of Complex 6

Table C.1 Computed Energies, at the ωB97X-V/def2-QZVPP/SMD(Benzene)//ωB97X-D4/def2-SVP/CPCM(Benzene) level of theory, and NMR chemical shifts, at the TPSS/pcSseg-2/SMD(Benzene)//ωB97X-D4/def2-SVP/CPCM(Benzene) level of theory, for selected protons in complex 6, at each 30-degree step of a relaxed surface scan. See Scheme 5.2 for atom assignment. Angles indicate change in dihedral from the minimum energy conformer.

Angle (degrees)	Energy (kJ/mol)	A (acid)	B (alpha)	C (alpha)	D (benz)	E (benz)	F (benz)	G (benz)
0	0	11.66	2.57	3.63	-3.05	-3.46	4.17	3.18
30	11.7	12.56	3.47	3.93	-2.97	-3.50	4.98	3.26
60	19.1	13.06	3.53	4.05	-2.97	-3.48	4.80	3.43
90	20.5	14.12	4.15	3.70	-3.00	-0.37	4.03	3.11
120	17.6	15.08	3.54	3.59	-0.70	-1.53	5.32	2.93
150	28.3	14.96	3.54	3.57	-0.75	-1.55	4.92	2.63
180	34.2	14.73	3.52	3.63	-0.86	-1.59	4.90	2.63
210	27.0	14.64	2.89	3.44	0.34	-0.96	4.77	3.01
240	27.1	15.35	3.86	3.96	-0.85	-1.51	4.72	3.36
270	31.0	13.85	2.38	2.94	0.72	-0.86	4.92	3.79
300	21.7	12.46	1.46	0.47	0.42	-1.19	4.91	3.63
330	9.32	12.23	1.43	-0.23	-1.17	-2.03	4.20	3.51

All chemical shifts are reported as the difference between the computed shielding for TMS (31.507 ppm) and the shielding of the proton of interest.

C.3 Semi-Empirically Derived Conformational Free Energies

Table C.2 gives the conformational free energies (G_{conf}) for monomers **1**, **2a**, and **2b**, as computed by CREST, and those of **6**-**19**, as computed by LEDE-CREST. The table also lists the change in conformational

free energy (ΔG_{conf}) for complexes **6**-**19**, computed as the difference between their conformational free energy, and those of their flexible monomers. Note that water has no rotatable bonds, and thus no conformational entropy, and is considered a rigid monomer.

Table C.2 Conformational free energies (G_{conf}) for the anhydrous monomers and all the complexes, plus the change in conformational free energy (ΔG_{conf}) for all the complexes. Values are computed by CREST (for monomers) and LEDE-CREST (for complexes).

	G_{conf} (kJ/mol)	ΔG_{conf} (kJ/mol)
1	-6.34	
2a	-18.23	
2b	-11.00	
6	-17.07	7.50
7	-10.48	11.53
8	-15.90	-3.22
9	-13.57	3.77
10	-7.79	9.56
11	-9.12	8.23
12	-8.74	13.27
13	-1.83	20.18
14	-11.74	0.94
15	-6.73	5.96
16	-8.75	3.57
17	-9.85	13.83
18	-7.68	16.01
19	-13.45	10.24