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

Nathanael J. King and Alex Brown*

Department of Chemistry, University of Alberta, Edmonton, AB, Canada, T6G 2G2

E-mail: alex.brown@ualberta.ca

Phone: +1-780-492-1854

Abstract

In this paper, the aggregation of asphaltene model compounds has been explored using a combination of an extended tight-binding method (GFN2-xtb) and density functional theory (DFT), in a manner that revisits an experimental study by Schulze, Lechner, Stryker, and Tykwinski (Org. Biomol. Chem. **2015**, *13*, 6984). 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 show 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 find 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 was found to be significant in this system, in contrast to some previous suggestions of their importance. 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//ωB97X-D4/def2-SVP level of theory, and using the Solvation Model based on Density (SMD) configured for benzene solvent. We also briefly explore some of the factors influencing the change in NMR chemical shift for select nuclei reported in the experimental paper.

Introduction

The nanoaggregation of asphaltenes is an important and poorly-understood field at the juncture of petrochemistry, analytical chemistry, and computational chemistry. Asphaltenes are defined as the fraction of a crude oil which dissolves in toluene, but not in light n-alkanes, although this definition has frequently been critiqued as unhelpful, imprecise, or impractical.[1] 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. 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.

Much research has been devoted to elucidating the nature of asphaltene molecules.[2,3] 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.[1,4–7] 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),[8] Gel Permeation Chromatography (GPC),[9–12] various forms of Laser Desorption Ionization Mass Spectrometry (LDI-MS),[12–15] Fluorescence Depolarization,[16,17] NMR diffusion studies,[18] and Electrospray Ionization Fourier-Transform Ion Cyclotron Resonance Mass Spectrometry (ESI FT-ICR MS).[19–23] The current consensus is that most petroleum asphaltene molecules have a mass between 350 and 1200 Da.

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.[3] 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 cosiders asphaltene aggregation to be a process dominated by $\pi - \pi$ stacking, while proponents of the archipelago model consider $\pi - \pi$ stacking, hydrogen bonding, ligand coordination to open sites on porphyrins, electrostatic attraction between permanent charges, and water-mediated aggregation to all be important.[24–27] 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,[28,29] or selective oxidation[30] or reduction,[31] appear to give evidence for the archipelago model, while Fluorescence Depolarization[16,17] and Atomic Force Microscopy (AFM)[32–34] 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.[22]

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 as well as extreme sensitivity and precision, can yield exact molecular formulae for thousands of constituents. However, despite the wealth of previously inaccessible 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 AFM, which does give partial structures, but many molecules thus imaged have large regions of ambiguity, possibly resulting from non-flat structures.[32] It can also be difficult to differentiate heteroatoms.[33]

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.[3] 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 $\pi - \pi$ stacking of large polycyclic aromatic cores.[24] Alternatively, model compounds can be used to validate or discredit theories or analytical techniques.[35–38] 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.[25,39–41] Several excellent

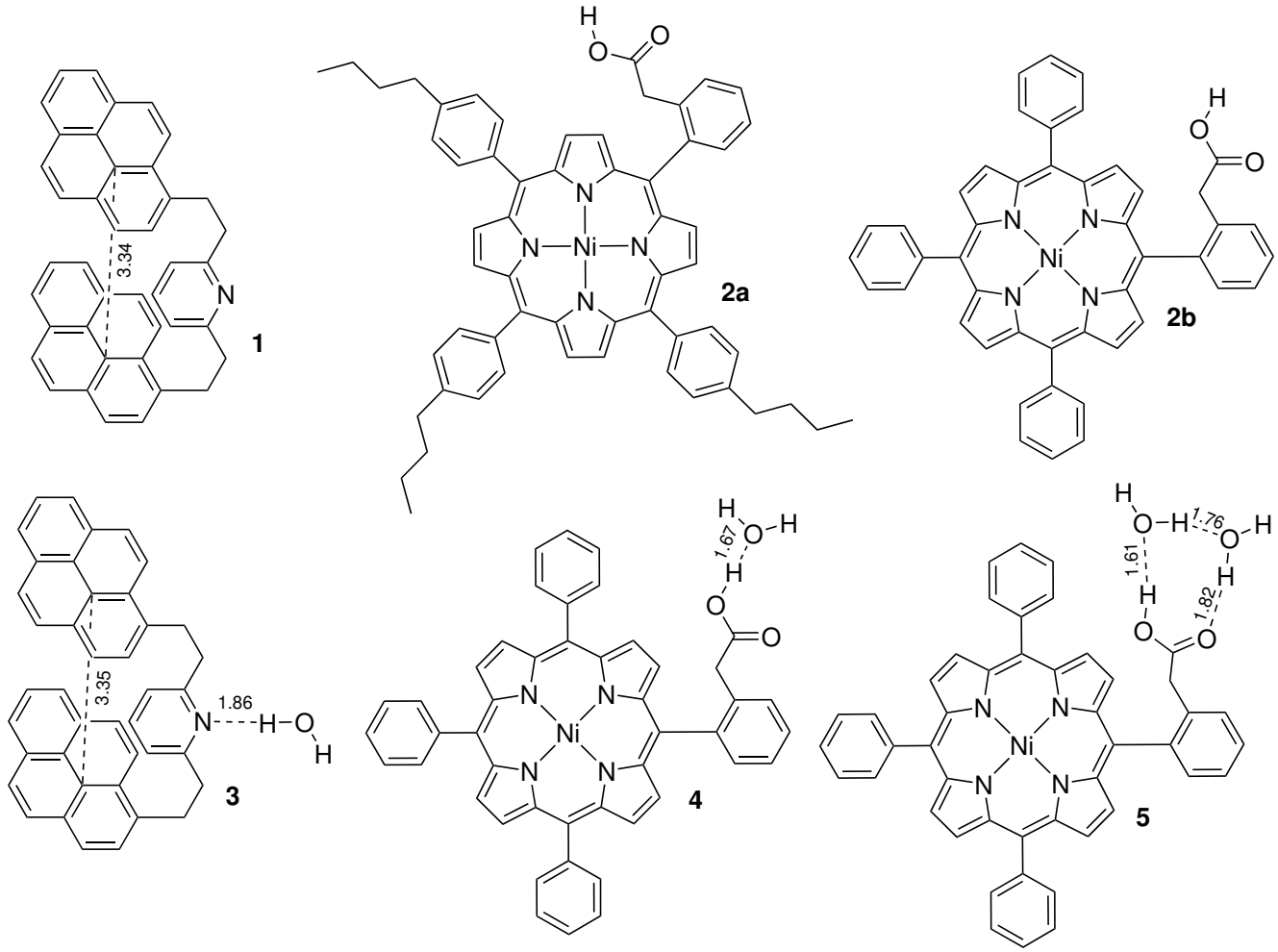
reviews have been published recently on what is known of the nature of asphaltenes[1,2] and on the use of model compounds in studying them.[3]

Another important use of model compounds is in computational studies.[42–44] 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) or DLPNO-CCSD(T)-based Energy Decomposition Analysis (EDA) can further break down the binding energy into components deriving from dispersion, Pauli exchange, electrostatics, and induction, if desired.[45,46] 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.

In 2015, Schulze et al. published a study on two model compounds for asphaltenes, and measured their interactions in solution using NMR.[47] 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 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 $\pi - \pi$ stacking plays a role, nor how significant that role may be. 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,



Scheme 1: Molecular structures of the asphaltene model compounds used in the Schulze et al. paper.[47] 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. Significant non-covalent interactions are indicated by dashed lines.

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**.[25] 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,[48] 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.,[25] then this may be a general result. Similarly, if hydrogen bonding turns out to be similar or greater in importance as $\pi - \pi$ stacking, this bears implications for our general understanding of asphaltene aggregation.

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)[49] for single molecules and LEDE-CREST (Low-Energy Diversity-Enhanced variant of CREST), which we developed in a recent paper,[50] for clusters of flexible molecules. All DFT computations were performed using ORCA 5.0.4.[51] One of the best DFT functionals for non-covalent interactions is ωB97X-V,[52,53] 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,[54] 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),[54–57] and thermodynamic computations were also carried out at this level, using Grimme's quasi-rigid-rotor-harmonic-oscillator approximation,[58] 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).[52,55–57,59] NMR shieldings are very sensitive to electron density around the nucleus, so specialized basis sets should be used.[60] Thus, NMR computations were performed using TPSS/pcSseg-2/SMD(Benzene)[59–61]//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.[62] 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 our previous work[63] and references cited therein.[64,65] Specifically, we compute ΔG_{assoc}, the Gibbs free energy of association, using

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

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

$$E_b = \sum_i E_{solv,i} - E_{solv,comp}. \tag{2}$$

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

Our previous work[63] 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.[63]

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 the ESI. 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.[63] 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.

Results and discussion

Monomers

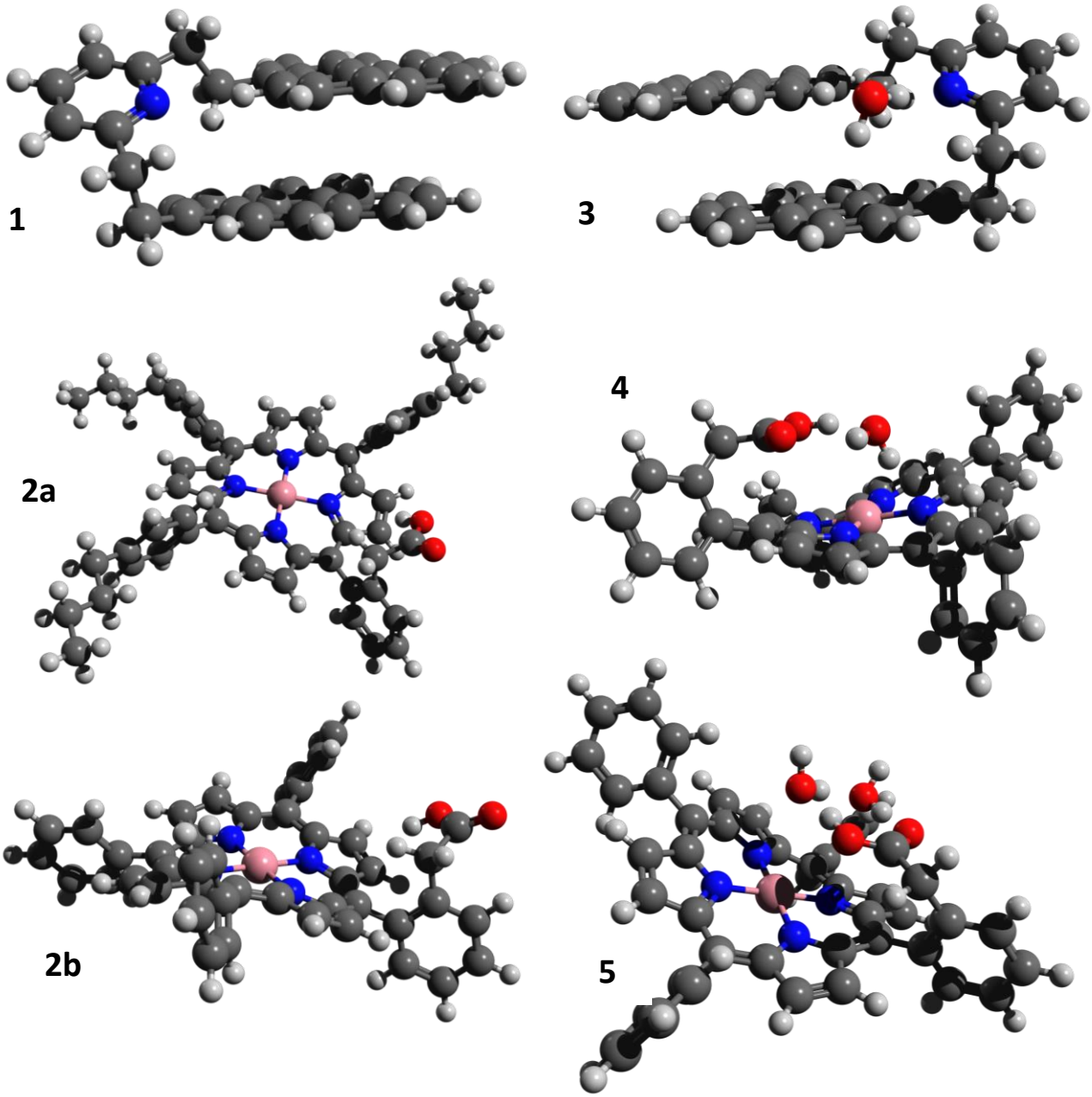
Anhydrous Monomers

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

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

Figure 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.



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[66] level, and then used as starting points for LEDE-CREST[50] 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 1. For hydrated monomer **3**, the $\pi - \pi$ 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.

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[66] level, and then used as starting points for LEDE-CREST[50] 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 2 and 3, and the optimized geometries are shown in Figures 2 and 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.[47] To reduce computational expense, complexes **6-19** 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.

Figure 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**. **8** is the hodimer of **1**.

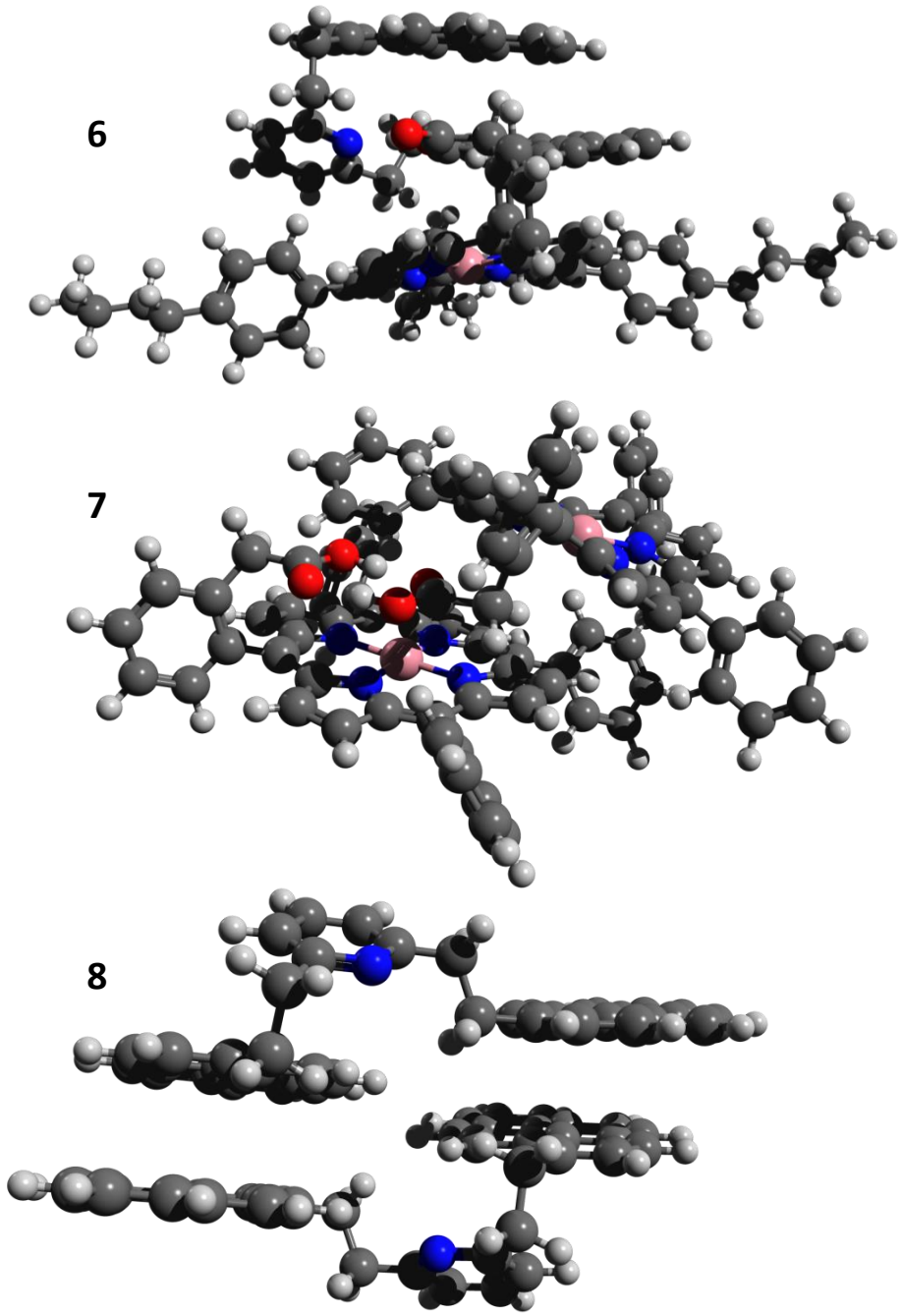
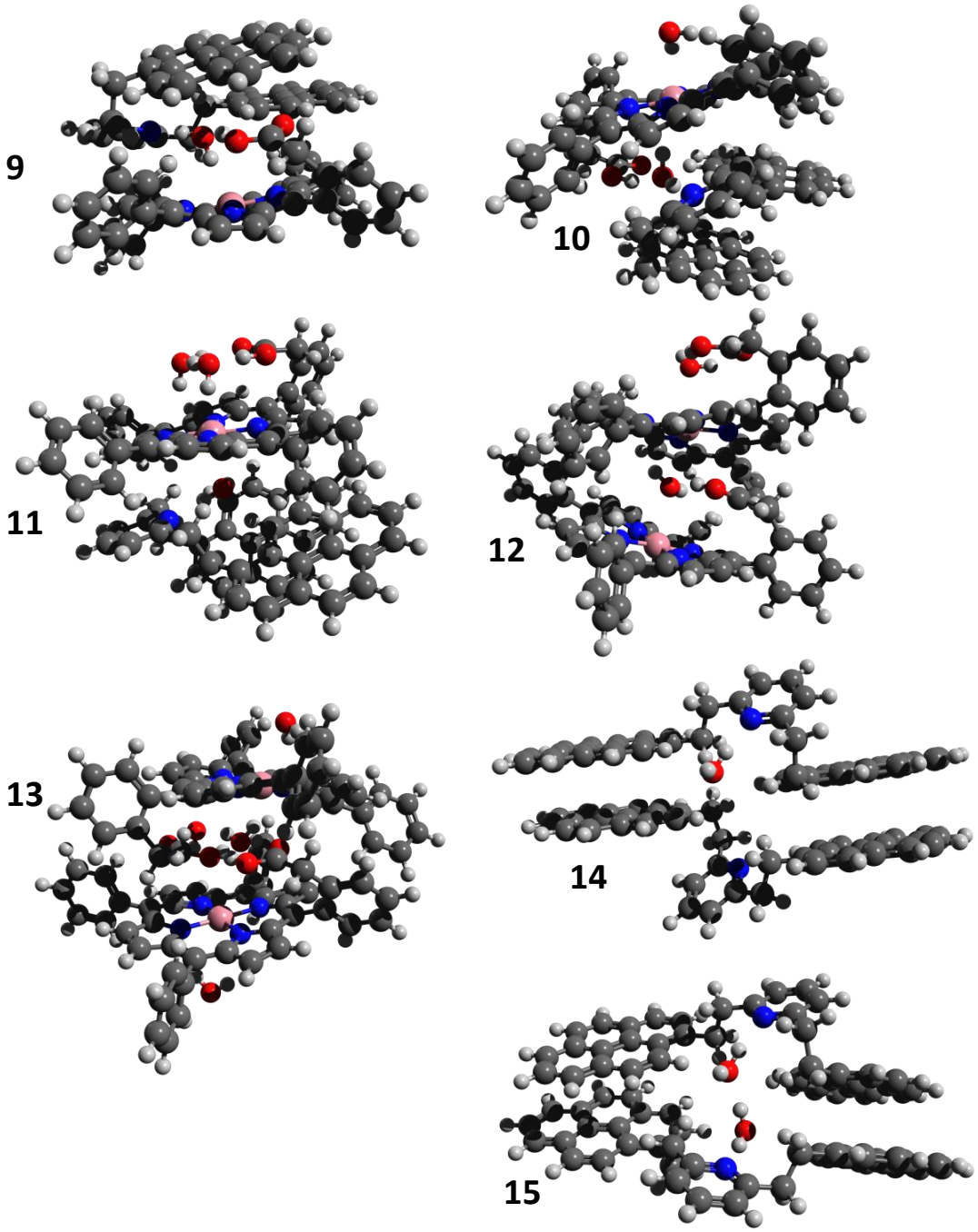
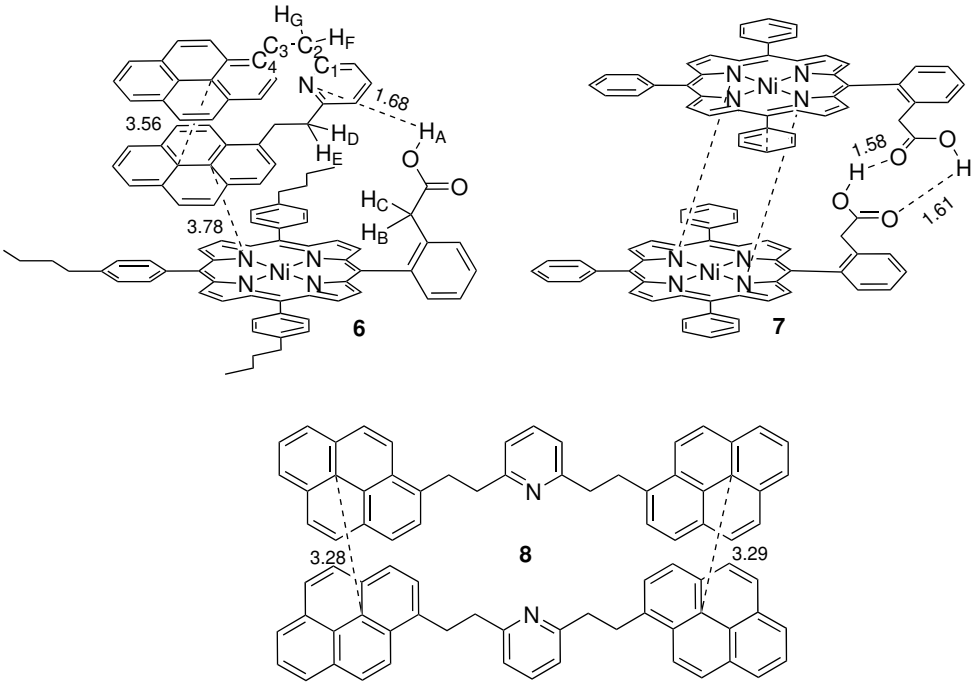
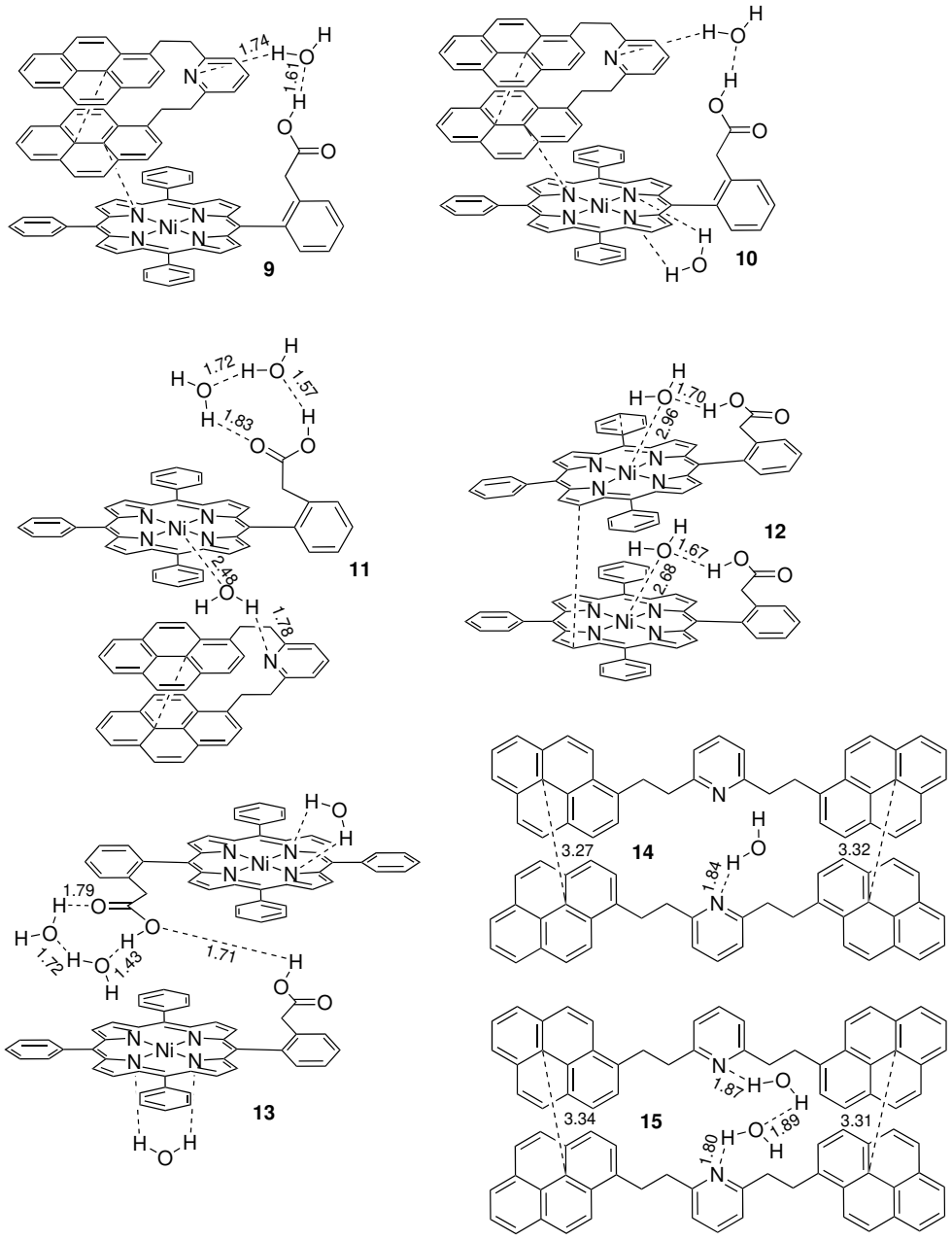


Figure 3: DFT-optimized geometries for the hydrated dimers, at the ωB97X-D4/def2-SVP/CPCM(Benzene) level of theory. Carbon is grey, hydrogen is white, oxygen is red, nitrogen is blue, and nickel is pink. **9 - 11** are the heterodimers of **1** and **2b**, with one, two, and three water molecules added, respectively. **12** and **13** are the hydrated homodimers of **2b**, with two and four water molecules added, respectively. **14** and **15** are the hydrated homodimers of **1**, with one and two molecules of water, respectively.





Scheme 2: Schematic representation of the anhydrous dimers, see Figure 2. Dashed lines indicate non-covalent contacts, and numbers in fine print indicate corresponding distances. Labeled hydrogens (H_i, i = A-G) indicate those selected for NMR predictions. Labeled carbons (C_i, i = 1-4) indicate the carbons which define the dihedral angle described in subsection "NMR."



Scheme 3: Schematic representation of the dimers and hydrated dimers, see Figure 3. Dashed lines indicate non-covalent contacts, and numbers in fine print indicate corresponding distances.

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 (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,[47] see subsection "Energetics of Binding" for details. Complex **8** is the homodimer of **1**. No hydrogen bonds are possible, and only $\pi - \pi$ stacking holds the dimer together. Average interplane distances are 3.29 Å and 3.28 Å.

Hydrated Dimers

Five possible hydrated dimers have been investigated: the heterodimers of **1** and **2b**, with one, two, or three water molecules (complexes **8-10**), and the homodimers of **2b**, with two or four water molecules (complexes **11** and **12**).

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 $\pi - \pi$ 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 $\pi - \pi$ 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°).

Trimers

Given that the Schulze paper[47] 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 4, and a schematic view of their non-covalent interactions is shown in Scheme 4.

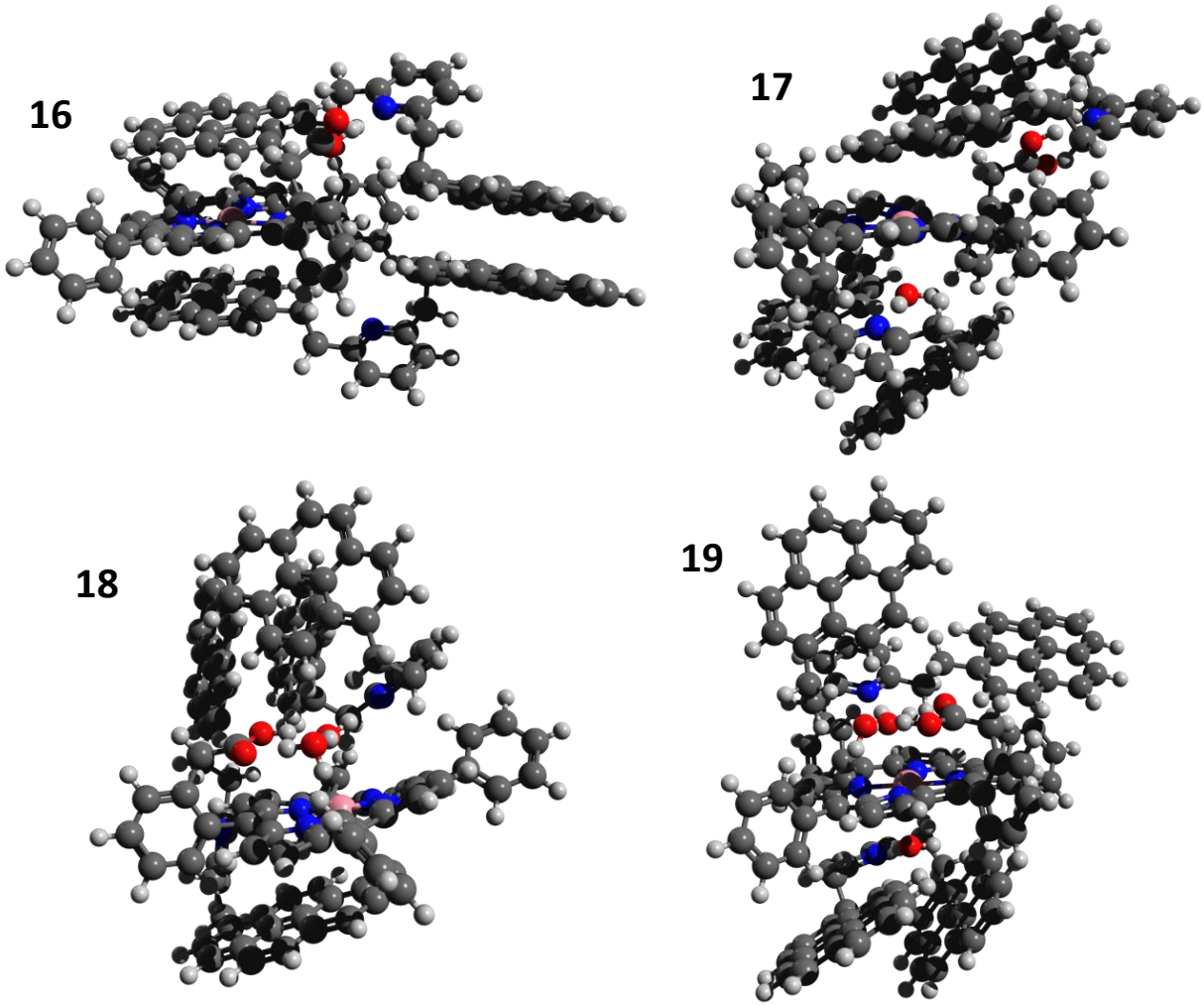
Anhydrous Trimer

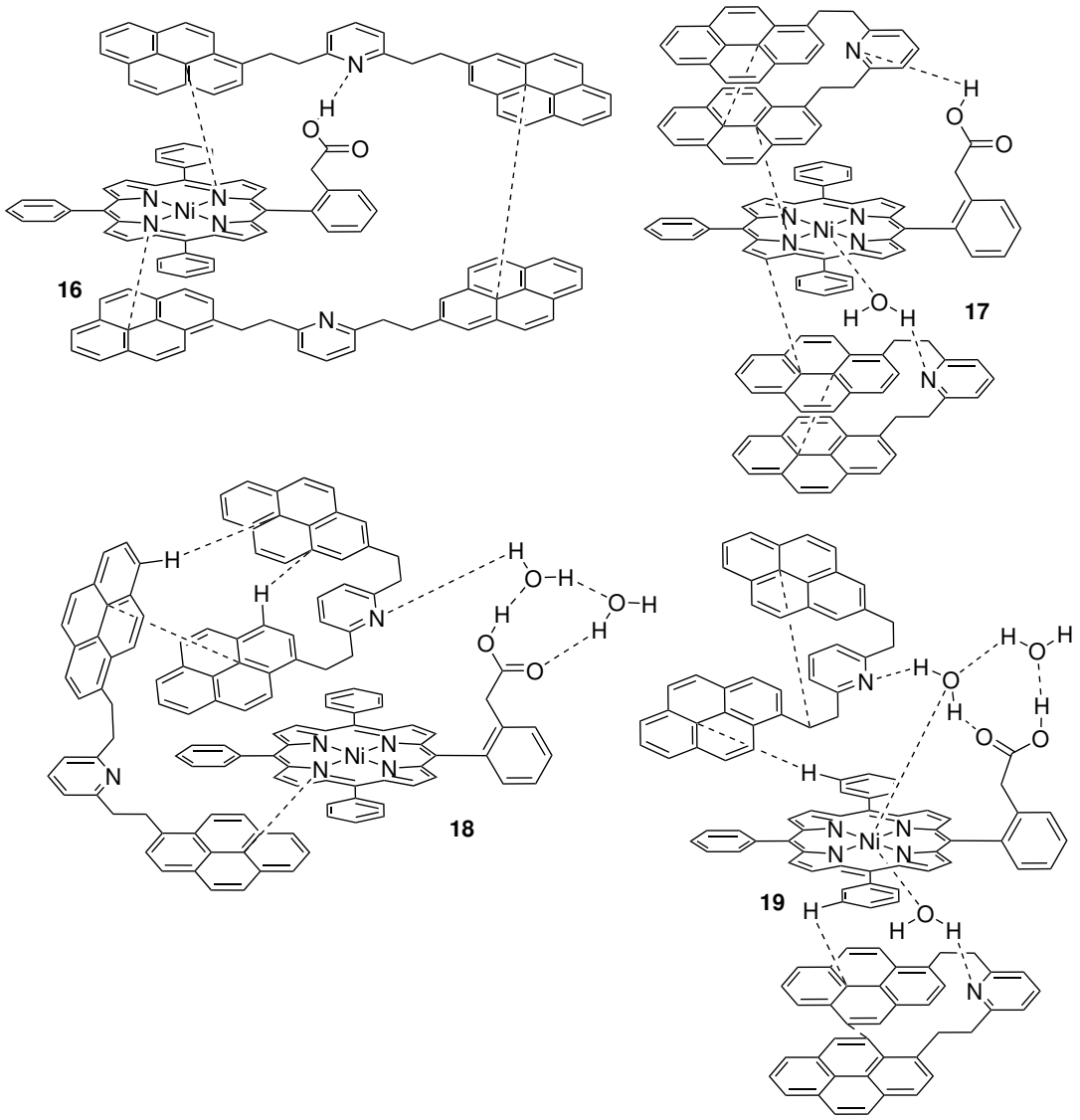
Complex **16** has one unit of **1** hydrogen bonded to the acid of **2b** (N-H distance 1.60 Å, OHN angle 171°) and with one pyrene moiety $\pi-\pi$ stacked to the porphyrin core (interplane distance 3.49 Å). The second unit of **1** has one pyrene moiety $\pi-\pi$ stacked with the other face of the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

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

Figure 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 - 19** each consist of one molecule of **2b** and two molecules of **1**, with zero, one, two, and three molecules of water, respectively.





Scheme 4: Schematic representation of the trimer and hydrated trimers, see Figure 4 with dashed lines indicating significant non-covalent contacts.

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 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 $\pi - \pi$ stacked, but exhibit numerous close non-covalent contacts with each other and with the porphyrin core and pendant benzene rings.

Energetics of Binding

Binding energies and Gibbs free energies for the complexes are given in Table 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.

Table 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

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.[47] 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,[53] and SMD has an MAD of about 0.7 kcal/mol (2.9 kJ/mol) for non-aqueous solvents.[59] 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 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.[48] 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.[67] Benzene solvent is able to $\pi - \pi$ 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 $\pi - \pi$ 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 $\pi - \pi$ stacking. Second, SMD includes a term to account for dispersion between solute and solvent, which should capture much or all of this effect. Third, any stabilization from $\pi - \pi$ 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,[67] 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**.[47] 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[68] 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 the ESI.

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 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 conentrations 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.

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.

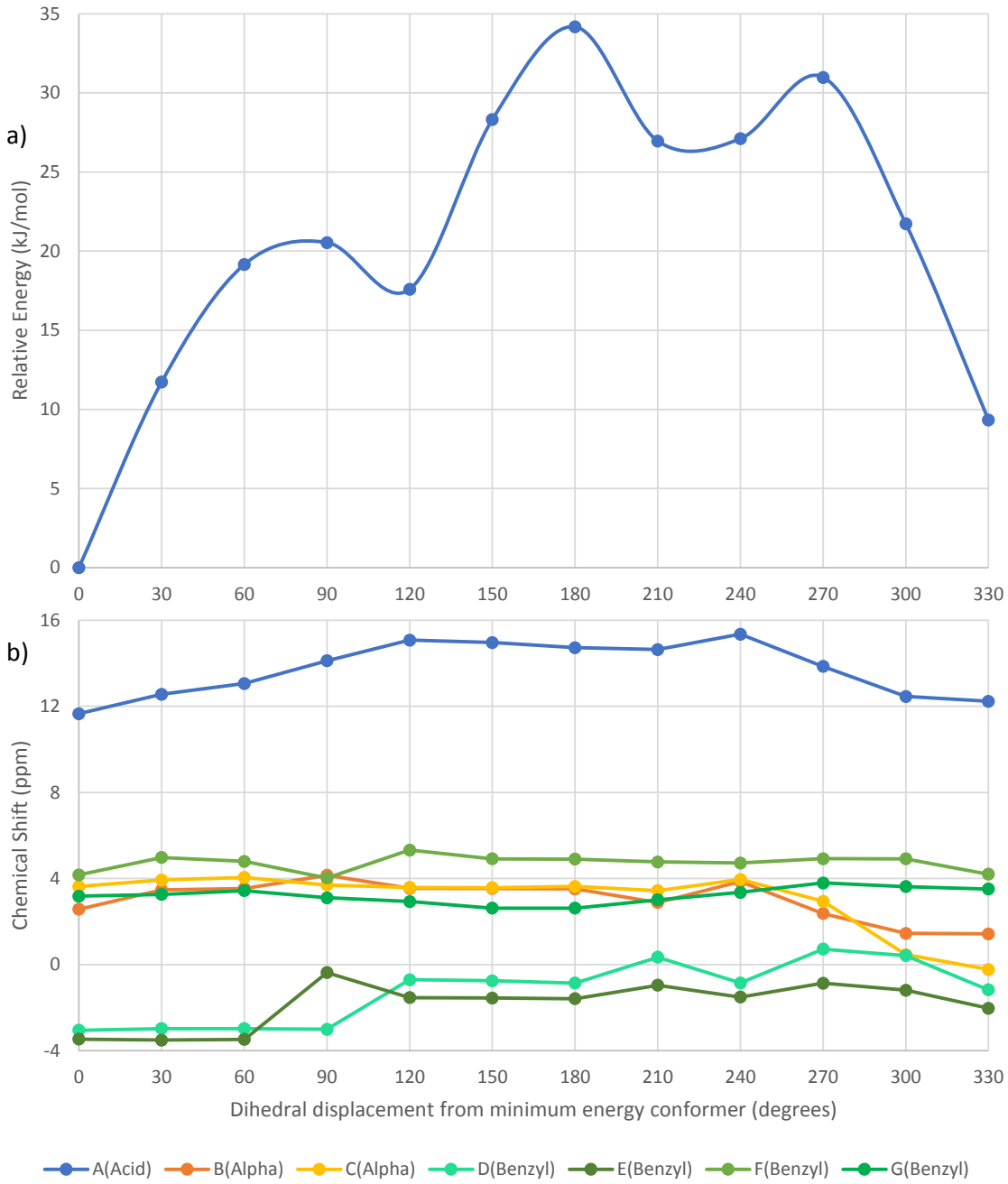
To illustrate this point, a relaxed surface scan of complex **6** was performed along a single degree of freedom - the dihedral angle defined by the carbon atom in the 2 position of the pyridine ring, the two carbon atoms of the C_2H_4 tether connecting the pyridine ring to one of the pyrene moieties, 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 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 2. Figure 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 the ESI. Note that the chemical shift can change by as much as 4.3 ppm, based only on the position of one pyrene moiety.

Hydrogen Bond Strength

One interesting point raised in the experimental paper[47] is the weaker binding of **1** with phenylacetic acid than that of pyridine with phenylacetic acid. Specifically, the K_{assoc} of

Figure 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 of the protons indicated by letters in Scheme 2.

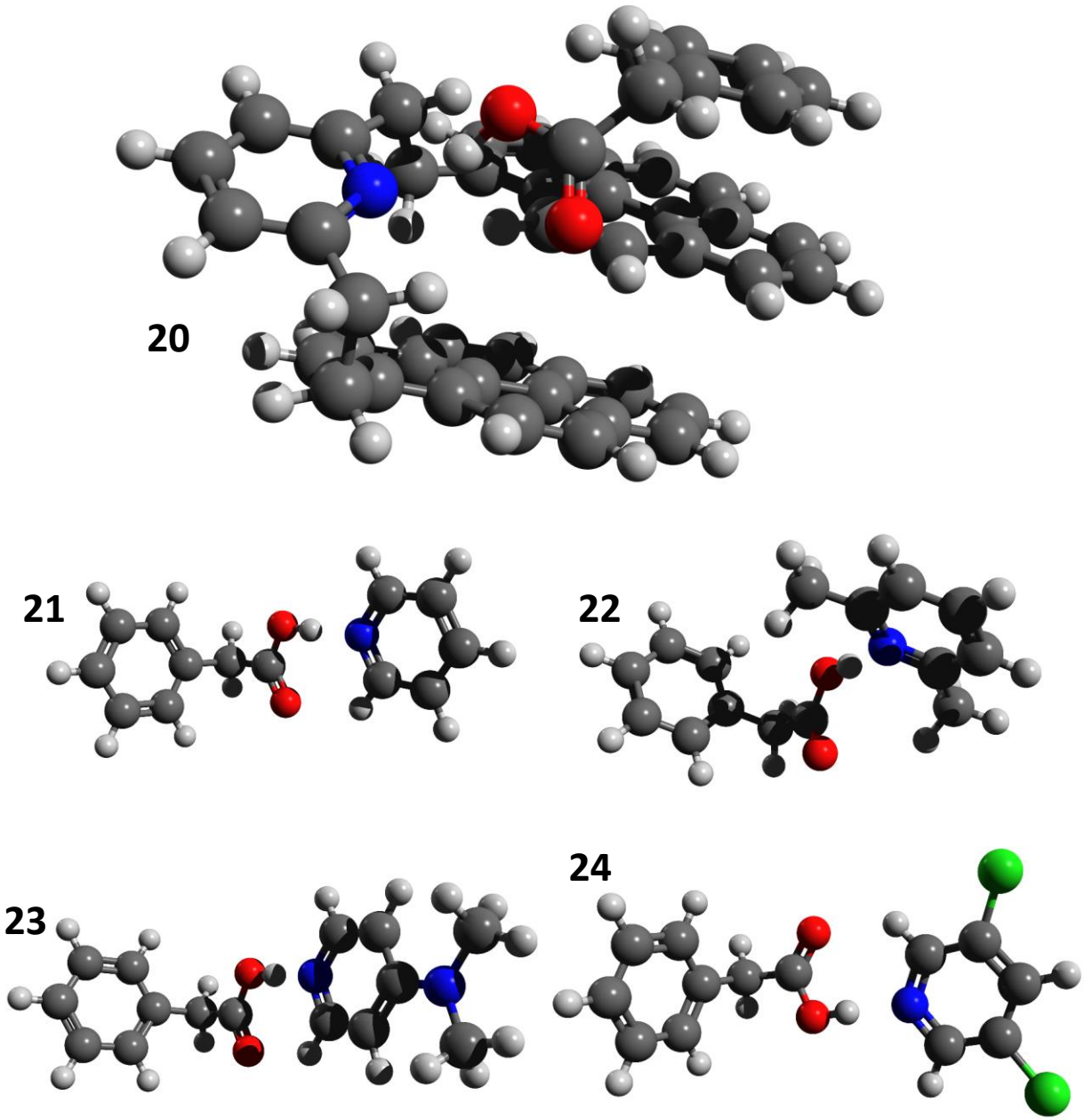


the pyridine-phenylacetic acid heterodimer was 123 M^{-1}, for a Gibbs free energy of -11.9 kJ/mol at 25 °C, while that of the heterodimer composed of **1** and phenylacetic acid was only 24 M^{-1}, for a Gibbs free energy of -7.9 kJ/mol. To investigate the reasons behind this difference, a series of 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 6. Gibbs free energies of dimerization are given in Table 2. Unsurprisingly, the trend shows that pyrene 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.

Table 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 the 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 group, and the protons which are *ortho* to nitrogen in the pyridine moiety, or the protons on the methyl/methylene groups in that *ortho* position, whichever is applicable to the specific case. Also reported is 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

Figure 6: Minimum energy structures 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 chlorine is green.



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 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 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.

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**. Both hydrogen bonding and $\pi - \pi$ 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.

Supporting Information Available

Details on system of equations in MatLab, numerical results for relaxed surface scan and NMR of **6**, semi-empirically-derived conformational free energies, and xyz coordinates for all monomers and complexes.

Acknowledgement

The authors thank the Natural Sciences and Engineering Research Council for funding (Discovery Grant to A. B.), the Digital Research Alliance of Canada for computational resources, and the Grimme group for developing CREST and making it freely available.

References

(1) Alshareef, A. H. Asphaltenes: Definition, Properties, and Reactions of Model Compounds. *Energy & Fuels* **2020**, *34*, 16–30.

(2) Gray, M. R.; Yarranton, H. W.; Chacón-Patiño, 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.

(3) 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.

(4) 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. *Alex. Eng. J.* **2022**, *61*, 9411–9425.

(5) 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.

(6) 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.

(7) Ghloum, E. F.; Al-Qahtani, M.; Al-Rashid, A. Effect of inhibitors on asphaltene precipitation for Marrat Kuwaiti reservoirs. *J. Pet. Sci. Eng.* **2010**, *70*, 99–106.

(8) Acevedo, S.; 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.

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

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

(11) 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.

(12) Acevedo, S.; Castro, A.; Negrin, J. G.; Fernández, A.; Escobar, G.; 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.

(13) Martínez-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. *J. Mass Spectrom.* **2007**, *42*, 701–713.

(14) 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.

(15) 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.

(16) Groenzin, H.; Mullins, O. C. Asphaltene Molecular Size and Structure. *J. Phys. Chem. A* **1999**, *103*, 11237–11245.

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

(18) Freed, D. E.; Lisitza, N. V.; Sen, P. N.; Song, Y.-Q. In *Asphaltenes, Heavy Oils, and Petroleomics*; Mullins, O. C., Sheu, E. Y., Hammami, A., Marshall, A. G., Eds.; Springer New York: New York, NY, 2007; pp 279–299.

(19) 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.

(20) 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.

(21) 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.

(22) Chacón-Patiño, 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.

(23) 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. *Anal. Chem.* **2021**, *93*, 13749–13754.

(24) 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.

(25) 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.

(26) 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 $\pi - \pi$ Stacking and Hydrogen-Bonding Interactions to the Aggregation of Model Asphaltene Compounds. *Energy & Fuels* **2012**, *26*, 2727–2735.

(27) Yin, C.-X.; Tan, X.; Müllen, K.; Stryker, J. M.; Gray, M. R. Associative $\pi - \pi$ Interactions of Condensed Aromatic Compounds with Vanadyl or Nickel Porphyrin Complexes Are Not Observed in the Organic Phase. *Energy & Fuels* **2008**, *22*, 2465–2469.

(28) Rubinstein, I.; Spyckerelle, C.; Strausz, O. Pyrolysis of asphaltenes: a source of geochemical information. *Geochim. Cosmochim. Acta* **1979**, *43*, 1–6.

(29) 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.

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

(31) 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. *J. Org. Chem.* **1977**, *42*, 312–320.

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

(33) Schuler, B.; Meyer, G.; Peña, D.; Mullins, O. C.; Gross, L. Unraveling the Molecular Structures of Asphaltenes by Atomic Force Microscopy. *J. Am. Chem. Soc.* **2015**, *137*, 9870–9876.

(34) 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.

(35) 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.

(36) 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.

(37) George, G. N.; Gorbaty, M. L. Sulfur K-edge x-ray absorption spectroscopy of petroleum asphaltenes and model compounds. *J. Am. Chem. Soc.* **1989**, *111*, 3182–3186.

(38) 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.

(39) 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.

(40) 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.

(41) 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.

(42) 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. *J. Mol. Liq.* **2021**, *334*, 116109.

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

(44) 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.

(45) 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.

(46) Podeszwa, R.; Bukowski, R.; Szalewicz, K. Potential Energy Surface for the Benzene Dimer and Perturbational Analysis of $\pi - \pi$ Interactions. *J. Phys. Chem. A* **2006**, *110*, 10345–10354, PMID: 16928128.

(47) 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. *Org. Biomol. Chem.* **2015**, *13*, 6984–6991.

(48) 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. (London)* **2016**, *52*, 12792–12805.

(49) 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.

(50) 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.

(51) Neese, F. Software update: the ORCA program system, version 4.0. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2018**, *8*, e1327.

(52) 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.

(53) 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.

(54) Najibi, A.; Goerigk, L. DFT-D4 counterparts of leading meta-generalized-gradient approximation and hybrid density functionals for energetics and geometries. *J. Comput. Chem.* **2020**, *41*, 2562–2572.

(55) Caldeweyher, E.; Ehlert, S.; Hansen, A.; Neugebauer, H.; Spicher, S.; Bannwarth, C.; Grimme, S. A generally applicable atomic-charge dependent London dispersion correction. *J. Chem. Phys.* **2019**, *150*, 154122.

(56) 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.

(57) Weigend, F. Accurate Coulomb-fitting basis sets for H to Rn. *Phys. Chem. Chem. Phys.* **2006**, *8*, 1057–1065.

(58) Grimme, S. Supramolecular Binding Thermodynamics by Dispersion-Corrected Density Functional Theory. *Chem. - Eur. J.* **2012**, *18*, 9955–9964.

(59) 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. *J. Phys. Chem. B* **2009**, *113*, 6378–6396.

(60) Jensen, F. Segmented Contracted Basis Sets Optimized for Nuclear Magnetic Shielding. *J. Chem. Theory Comput.* **2015**, *11*, 132–138.

(61) 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.

(62) 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, PMID: 30048136.

(63) King, N. J.; Brown, A. Intermolecular Interactions of Pyrene and Its Oxides in Toluene Solution. *J. Phys. Chem. A* **2022**, *126*, 4931–4940.

(64) Bryantsev, V. S.; Diallo, M. S.; Goddard III, W. A. Calculation of Solvation Free Energies of Charged Solutes Using Mixed Cluster/Continuum Models. *J. Phys. Chem. B* **2008**, *112*, 9709–9719.

(65) Patel, D. H.; East, A. L. L. Semicontinuum (Cluster-Continuum) Modeling of Acid-Catalyzed Aqueous Reactions: Alkene Hydration. *J. Phys. Chem. A* **2020**, *124*, 9088–9104.

(66) Grimme, S. Exploration of Chemical Compound, Conformer, and Reaction Space with Meta-Dynamics Simulations Based on Tight-Binding Quantum Chemical Calculations. *J. Chem. Theory Comput.* **2019**, *15*, 2847–2862.

(67) Simm, G. N.; Türtscher, P. L.; Reiher, M. Systematic microsolvation approach with a cluster-continuum scheme and conformational sampling. *J. Comput. Chem.* **2020**, *41*, 1144–1155.

(68) Inc., T. M. MATLAB version: 9.13.0 (R2022b). 2022; `https://www.mathworks.com`.

TOC Graphic