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 density functional tight-binding (DFTB) and density functional theory (DFT), in a manner that revisits an experimental study from 2015 by Schulze, Lechner, Stryker, and Tykwinski. 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 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 micro-hydration, at the \( \omega \text{B97X-V/def2-QZVPP//}\omega \text{B97X-D4/def2-SVP} \) level of theory. 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.

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 Spectroscopy (LDI-MS),\(^12-15\) Fluorescence Depolarization,\(^16,17\) NMR dif-
fusion studies,\textsuperscript{18} and Electrospray Ionization Fourier-Transform Ion Cyclotron Resonance Mass Spectroscopy (ESI FT-ICR MS).\textsuperscript{19–23} The current consensus among most authors 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.\textsuperscript{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 considers 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.\textsuperscript{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,\textsuperscript{28,29} or selective oxidation\textsuperscript{30} or reduction,\textsuperscript{31} appear to give evidence for the archipelago model, while Fluorescence Depolarization\textsuperscript{16,17} and Atomic Force Microscopy (AFM)\textsuperscript{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.\textsuperscript{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 and 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.\textsuperscript{32} It can also be difficult to differentiate heteroatoms.\textsuperscript{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.\textsuperscript{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.\textsuperscript{24} Alternatively, model compounds can be used to validate or discredit theories or analytical techniques.\textsuperscript{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.\textsuperscript{25,39–41} Several excellent reviews have been published recently on what is known of the nature of asphaltenes\textsuperscript{1,2} and on the use of model compounds in studying them.\textsuperscript{3}

Another important use of model compounds is in computational studies.\textsuperscript{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 com-
putationally. 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∆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. These results are unique to computations, and cannot be obtained by experiment. 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. The model compounds chosen were an archipelago-style model 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 $1.23 \times 10^6 \text{ M}^{-2}$. The equivalent Gibbs free energies of formation 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.

Scheme 1: Molecular structures of the asphaltene model compounds used in the Schulze et al. paper. 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.

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, of these 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, 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. 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, 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., 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)\textsuperscript{47} for single molecules and LEDE-CREST (Low-Energy Diversity-Enhanced variant of CREST), which we developed in a recent paper,\textsuperscript{48} for clusters of flexible molecules. All DFT computations were performed using ORCA 5.0.4. In each case, the lowest-energy structure obtained from CREST or LEDE-CREST was reoptimized using \(\omega\)B97X-D4/def2-SVP/CPCM(Benzene),\textsuperscript{50–53} and thermodynamic computations were also carried out at this level, using Grimme’s quasi-rigid-rotor-harmonic-oscillator approximation,\textsuperscript{54} as implemented in ORCA. Final single point computations were performed using \(\omega\)B97X-V/def2-QZVPP/SMD(Benzene).\textsuperscript{51–53,55,56} NMR shieldings are very sensitive to electron density around the nucleus, so specialized basis sets should be used.\textsuperscript{57} Thus, NMR computations were performed using TPSS/pcSseg-2/SMD(Benzene)\textsuperscript{56–58} //\(\omega\)B97X-D4/def2-SVP/CPCM(Benzene). NMR chemical shifts are given as the difference between the computed shielding of tetramethylsilane and the computed shielding of the proton of interest.

Free energies were determined according to the equations detailed in our previous work\textsuperscript{59} and references cited therein.\textsuperscript{60,61} Specifically, we compute \(\Delta G_{assoc}\), the Gibbs free energy of association, using

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

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

\[
E_b = \sum_i E_{\text{solv},i} - E_{\text{solv,comp}}. \tag{2}
\]

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

Our previous work\textsuperscript{59} used ensembles to obtain the conformational entropy, for a more accurate overall \( \Delta G_{\text{assoc}} \), but this has 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 3 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.\textsuperscript{59}
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.

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 level, and then used as starting points for LEDE-CREST 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.

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 level, and then used as starting points for LEDE-CREST 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 Scheme 2, and the optimized geometries are shown in Figure 2. 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
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.
solubility. To reduce computational expense, complexes 6-16 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.

Anhydrous Dimers

Complex 6 is the simplest of the complexes studied. 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$_2$COOH 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, see subsection ”Energetics of Binding” for details.

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 8, 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 9, the lowest-energy structure is very similar to complex 8, but with a second water molecule interacting with the open face of the porphyrin. The two protons on the
Figure 2: DFT-optimized geometries for the dimers and hydrated dimers, at the \( \omega \text{B97X-D4/def2-SVP/CPCM(Benzene)} \) 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 - 10 are the heterodimers of 1 and 2b, with one, two, and three water molecules added, respectively. 11 and 12 are the hydrated homodimers of 2b, with two and four water molecules added, respectively.
Scheme 2: Schematic representation of the dimers and hydrated dimers, with dashed lines indicating significant non-covalent contacts. Labeled hydrogens indicate those selected for NMR predictions.
second molecule of water are each pointing at a nitrogen in the porphyrin.

For complex 10, 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 11, 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 12, 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.

Trimers

Given that the Schulze paper 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 struc-
tures 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 \( \omega B97X-V/\text{def2-QZVPP}/\text{SMD}({\text{Benzene}})/\omega B97X-D4/\text{def2-SVP}/\text{CPCM}({\text{Benzene}}) \) level. DFT-optimized geometries are shown in Figure 3, and a schematic view of their non-covalent interactions is shown in Scheme 3.

Scheme 3: Schematic representation of the trimer and hydrated trimers, with dashed lines indicating significant non-covalent contacts.
Figure 3: 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 13 - 16 each consist of one molecule of 2b and two molecules of 1, with zero, one, two, and three molecules of water, respectively.
Anhydrous Trimer

Complex 13 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 units (one from each unit of 1) are π−π stacked with each other (interplane distance 3.30 Å).

Hydrated Trimers

The interactions of the first molecule of 1 with 2b in complex 14 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 15, two water molecules form a closed ring of hydrogen bonds with the acid group of the porphyrin, as in complex 10 (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 16 exhibits the same ring of hydrogen bonds as complexes 10 and 15 (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.

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.

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
Table 1: Binding energy, Gibbs free energy, "penalty" (sum of Gibbs free energy and binding energy), and equilibrium constant of formation for each complex.

<table>
<thead>
<tr>
<th>Complex</th>
<th># of water molecules</th>
<th>$E_b$ (kJ/mol)</th>
<th>$\Delta G_{assoc}$ (kJ/mol)</th>
<th>$E_b + \Delta G_{assoc}$ (kJ/mol)</th>
<th>$K$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Hydrated Monomers</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>34.3</td>
<td>8.0</td>
<td>42.2</td>
<td>0.040</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>55.3</td>
<td>-10.7</td>
<td>44.5</td>
<td>76.8</td>
</tr>
<tr>
<td>5</td>
<td>2</td>
<td>82.2</td>
<td>3.7</td>
<td>86.0</td>
<td>0.22</td>
</tr>
<tr>
<td><strong>Anhydrous Dimers</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>126.6</td>
<td>-46.0</td>
<td>80.5</td>
<td>$1.2 \times 10^8$</td>
</tr>
<tr>
<td>7</td>
<td>0</td>
<td>139.8</td>
<td>-50.7</td>
<td>89.0</td>
<td>$7.9 \times 10^8$</td>
</tr>
<tr>
<td><strong>Hydrated Dimers</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>1</td>
<td>130.2</td>
<td>-5.6</td>
<td>124.6</td>
<td>9.6</td>
</tr>
<tr>
<td>9</td>
<td>2</td>
<td>150.4</td>
<td>8.5</td>
<td>158.9</td>
<td>0.033</td>
</tr>
<tr>
<td>10</td>
<td>3</td>
<td>208.4</td>
<td>-4.5</td>
<td>203.9</td>
<td>6.1</td>
</tr>
<tr>
<td>11</td>
<td>2</td>
<td>156.7</td>
<td>6.1</td>
<td>162.7</td>
<td>0.087</td>
</tr>
<tr>
<td>12</td>
<td>4</td>
<td>229.5</td>
<td>10.0</td>
<td>239.5</td>
<td>0.017</td>
</tr>
<tr>
<td><strong>Anhydrous Trimer</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>0</td>
<td>201.8</td>
<td>-56.6</td>
<td>145.2</td>
<td>$8.6 \times 10^9$</td>
</tr>
<tr>
<td><strong>Hydrated Trimers</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>1</td>
<td>228.4</td>
<td>-46.2</td>
<td>182.2</td>
<td>$1.3 \times 10^8$</td>
</tr>
<tr>
<td>15</td>
<td>2</td>
<td>247.7</td>
<td>-14.6</td>
<td>233.1</td>
<td>360</td>
</tr>
<tr>
<td>16</td>
<td>3</td>
<td>285.7</td>
<td>-16.9</td>
<td>268.8</td>
<td>920</td>
</tr>
</tbody>
</table>
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.\textsuperscript{45} 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 $\omega$B97X-V to have a WTMAD-1 of only 1.81 kcal/mol (7.57 kJ/mol) for non-covalent
interactions,\textsuperscript{63} and SMD has an MAD of about 0.7 kcal/mol (2.9 kJ/mol) for non-aqueous
solvents.\textsuperscript{56} Together, this gives an expected deviation of only 1.9 kcal/mol (7.9 kJ/mol) for
this combination of methods. 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.\textsuperscript{46} 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.
The only complexes with large (>1000) equilibrium constants of formation (equivalent to Gibbs free energies of formation below -17 kJ/mol) are 6, 7, 13, and 14. 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. 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 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 13, and 1.52 pM for 14. 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 13, and 32 nM for 14, 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 13, and 12 nM for 14, giving a situation where 6 and 13 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$_2$H$_4$ tether connecting the pyridine ring to one of the pyrene moieties, and the carbon atom in the 1 position on the pyrene moiety. 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 ($\omega$B97X-V/def2-QZVPP/SMD(Benzene)) and NMR shieldings (TPSS/pCseg-2/SMD(Benzene)) were then computed for each structure.

The protons selected for NMR predictions are indicated by letters in Scheme 2. Figure 4 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\textsuperscript{45} is the weaker binding of 1 with phenylacetic acid than that of pyridine with phenylacetic acid. Specifically, the $K_{\text{assoc}}$ of the pyridine-phenylacetic acid heterodimer was 123 M$^{-1}$, for a Gibbs free energy of -11.9
Figure 4: Energies (at the $\omega$B97X-V/def2-QZVPP/SMD(Benzene)//$\omega$B97X-D4/def2-SVP/PCPM(Benzene) level of theory) and NMR chemical shifts (at the TPSS/pcSseg-2/SMD(Benzene)//$\omega$B97X-D4/def2-SVP/PCPM(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.
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 17), with pyridine (complex 18), with 2,6-lutidine (complex 19), with N,N-dimethyl-4-aminopyridine (complex 20), and with 3,5-dichloropyridine (complex 21), were modeled. Optimized structures are given in Figure 5. 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 17 - 21, at the $\omega$B97X-V/def2-QZVPP/SMD(Benzene)//$\omega$B97X-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)//$\omega$B97X-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.

<table>
<thead>
<tr>
<th>Complex</th>
<th>$\Delta G_{assoc}$ (kJ/mol)</th>
<th>Monomer Chemical Shift (ppm)</th>
<th>Dimer Chemical Shift (ppm)</th>
<th>Change in Chemical Shift (ppm)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Acid</td>
<td>Alpha</td>
<td>Ortho</td>
</tr>
<tr>
<td>17</td>
<td>11.4</td>
<td>5.73</td>
<td>3.73</td>
<td>3.39</td>
</tr>
<tr>
<td>18</td>
<td>2.6</td>
<td>5.73</td>
<td>3.73</td>
<td>8.80</td>
</tr>
<tr>
<td>19</td>
<td>-4.8</td>
<td>5.73</td>
<td>3.73</td>
<td>2.51</td>
</tr>
<tr>
<td>20</td>
<td>-1.4</td>
<td>5.73</td>
<td>3.73</td>
<td>8.30</td>
</tr>
<tr>
<td>21</td>
<td>9.6</td>
<td>5.73</td>
<td>3.73</td>
<td>8.50</td>
</tr>
</tbody>
</table>

The same effect comes into play for the complex of 1 and phenylacetic acid, but here
Figure 5: Minimum energy structures for complexes 17 - 21, at the \( \omega \text{B97X-D4/def2-SVP/CPCM}(\text{Benzene}) \) level of theory. Carbon is grey, hydrogen is white, oxygen is red, nitrogen is blue, and chlorine is green.
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 17-21 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 (19 and 20) vs. the pyridine/phenylacetic acid complex (18), which is again more deshielded that the complex with electron withdrawing groups (21). Complex 17, 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 21. 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, 13, and 14 being the most stable. At the low con-
centrations attainable experimentally, 6 is the dominant complex, although 7 and 13 also form in noticeable quantities. 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.

Acknowledgement

The authors thank the National Science 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.

Supporting Information Available

Details on system of equations in MatLab, numerical results for relaxed surface scan and NMR of 6, and xyz coordinates for all monomers and complexes.

References


(2) Gray, M. R.; Yarranton, H. W.; Chacón-Patiño, M. L.; Rodgers, R. P.; Bouyssiere, B.;


(33) Schuler, B.; Meyer, G.; Peña, D.; Mullins, O. C.; Gross, L. Unraveling the Molecular


(41) Tan, X.; Fenniri, H.; Gray, M. R. Pyrene Derivatives of 2,2'-Bipyridine as Models for


(48) 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/](https://chemrxiv.org/engage/)

34


(56) 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


TOC Graphic