

A-Priori Calculations of the Free Energy of Formation from Solution of Polymorphic Self-

Assembled Monolayers

Jeffrey R. Reimers,^{1,2} Dwi Panduwinata,³ Johan Visser,³ Yiing Chin,³ Chunguang Tang,³ Lars Goerigk,^{3,4} Michael J. Ford,² Maxine Sintic,³ Tze Jing Sum,³ Michiel J. J. Coenen,⁵ Bas L. M. Hendriksen,⁵ Johannes A. A. W. Elemans,⁵ Noel S. Hush,^{3,6*} and Maxwell J. Crossley^{3*}*

1. International Centre for Quantum and Molecular Structure, College of Sciences, Shanghai University, Shanghai 200444, China.
2. School of Mathematical and Physical Sciences, The University of Technology Sydney, Sydney NSW 2007 Australia.
3. School of Chemistry, The University of Sydney, NSW 2006, Australia
4. School of Chemistry, The University of Melbourne, VIC 3010, Australia
5. Radboud University Nijmegen, Institute for Molecules and Materials, Heyendaalseweg 135, 6525 AJ Nijmegen, The Netherlands
6. School of Biomolecular Sciences, The University of Sydney, NSW 2006, Australia.

***Corresponding authors, email:** reimers@shu.edu.acn, jeffrey.reimers@uts.edu.au, maxwell.crossley@sydney.edu.au, noel.hush@sydney.edu.au

Phone: Reimers: +8615618155341, Hush +61293514688, Crossley +61293512751

Classification: Physical Sciences, Chemistry

Keywords: self-assembled monolayers, density-functional theory, dispersion, solvent interaction, free energy, polymorphism

Significance statement: First principles free energy calculations, characterizing polymorphism of self-assembled monolayers (SAMs) of porphyrin molecules formed from solution onto graphite, are performed using efficient methods previously applied only to small-molecule reactivity. SAM structures are typically optimized in the absence of solvent using Density-Functional Theory embodying explicit dispersion corrections. Added then are dispersion-dominated implicit solvation energies and SAM formation entropies derived from both molecular and phonon vibration frequencies. Scanning-Tunnelling Microscopy (STM) images are measured, and polymorph formation free energies approximated. Close parallels between experiment and theory support the hypothesis that the first seconds of SAM formation are under thermodynamic control, despite formed SAMs being kinetically trapped. Polymorphism is associated with large opposing changes to entropy and substrate-molecule and solvent-molecule interaction energies.

ABSTRACT: Modern quantum-chemical electronic-structure methods typically applied to localized chemical bonding are developed to predict atomic structures and free energies for meso-tetraalkylporphyrin self-assembled monolayer (SAM) polymorph formation from organic solution on highly-ordered pyrolytic graphite (HOPG) surfaces. Large polymorph-dependent dispersion-induced substrate-molecule interactions (e.g., -100 to -150 kcal mol⁻¹ for tetrakisdecylporphyrin) are found to drive SAM formation, opposed nearly completely by large polymorph-dependent dispersion-induced solvent interactions (70 to 110 kcal mol⁻¹) and entropy effects (25 to 40 kcal mol⁻¹ at 298 K) favouring dissolution. Dielectric continuum models of the solvent are used, facilitating consideration of many possible SAM polymorphs, along with quantum-mechanical molecular-mechanical (QM/MM) and dispersion-corrected density functional theory (DFT) calculations. These predict and interpret newly measured and existing high-resolution scanning tunnelling microscopy (STM) images of SAM structure, rationalizing polymorph formation conditions. A wide range of new molecular condensed-matter properties at room temperature now appear suitable for prediction and analysis using electronic-structure calculations.

A-priori calculations of the free energies of chemical reactions using Density-Functional Theory (DFT) and/or ab initio methods are now well established for gas-phase processes (1, 2), gas-surface reactions (3), and, using continuum self-consistent reaction-field (SCRf) methods, for condensed-phase processes also (4). Over the last few years, a major advance in computational methods has occurred, however, allowing for rapid and accurate evaluation of intermolecular dispersion interactions. This makes feasible similar SCRf calculations for physisorption, macromolecule structuring, and other self-assembly processes in condensed phases. We present the first application of this type, considering the free-energy of formation of various polymorphs of tetraalkylporphyrin self-assembled monolayers (SAMs) at solid/liquid interfaces on highly-ordered pyrolytic graphite (HOPG) surfaces. Calculated structures and free energies are used to interpret new and existing high-resolution scanning tunnelling microscopy (STM) images, focusing on the critical roles played by the dispersion interaction in driving SAM formation and by entropy and dispersion-based desolvation effects that oppose it.

Calculating a priori free energies for SAM formation from solution is a significant challenge. Accurate representations of the substrate-molecule energies, the intermolecular energies, the solvent interaction energies, and the effects of solvent structure are required, a set of tasks that, with a few exceptions (see e.g., (5, 6)) has remained prohibitive a priori. Alternatively, model calculations have been useful for identifying key qualitative features (7-11), while some full molecular dynamics simulations of the free energy using empirical force fields have been revealing for SAMs (12-15) and also for macromolecules (16), and DFT molecular dynamics approaches are appearing (17). A-priori computational methods have significant advantages in that they do not require parameterization, treat interactions both generally and accurately, and, through use of implicit solvation models, can avoid extensive liquid-structure simulations. However, applications to self-assembly processes have historically been prevented by difficulties in treating discrete and continuum van der Waals dispersive forces both quantitatively and economically. Modern methods (18-25) can now describe discrete interactions well, and self-assembly processes such as molecular crystal formation (22) and host-guest

chemistry (26) have now been successfully modelled. Dispersive interactions with a solvent continuum (18, 27) are also now included in many codes including GAUSSIAN (28) and, very recently, VASP (29). Like discrete dispersion interactions, the best method for including continuum solvent dispersion is a topic of considerable current interest (30-32). In both cases many options are available, and consideration of SAM formation offers avenues for finding the most robust approaches. Herein we consider 3 possible approaches obtained by combining commonly used options for different aspects of the calculations, as described later. In overview, our computational approach is to:

- 1) Systematically search through conformational options to establish sets of possibly feasible SAM structures,
- 2) To optimize these SAM structures typically in the absence of solvent,
- 3) To determine SAM solvation energy contributions at these optimized structures for comparison with bare-surface and free-molecule results, and
- 4) To determine zero-point energy, entropy, and enthalpy contributions to the free energy of SAM formation, including evaluating both the molecular and phonon vibration frequencies in situ in the SAM as well as the vibration frequencies of the gas-phase monomer molecules.

To be useful, a new computational strategy must address relevant experimental challenges, and indeed the *measurement* of free energies for SAM formation is in itself a significant challenge facing modern nanotechnology (33). A critical experiment, the first measurement of the temperature dependence of composition of SAM components, has recently been reported (34), opening this field to future quantitative analysis. Indeed, in total only a few studies of SAM temperature dependence have ever been performed, as recently reviewed (12, 34, 35). An important conclusion drawn is that the nature of observed SAM composition can be controlled by either kinetic and/or thermodynamic effects (9, 12, 14, 15, 33-44). Slow-moving SAMs in thermodynamic equilibrium can have free energies determined directly from dynamics observed in the STM (17). More generally, while methods for

measuring enthalpies of SAM formation are becoming available (45), it was recently concluded that “The Achilles’ heel of [the current] approach is the semi-theoretical evaluation of the dewetting enthalpy which at this point necessarily relies on plausible assumptions” (46). A priori calculations can not only directly calculate the required quantities but they are also relevant to this field of endeavour as they can directly examine postulates concerning kinetic versus thermodynamic control. They can also reveal the factors that control observed free energies, providing insight into SAM synthetic strategies and functional properties.

Herein our techniques for a-priori SAM simulation are applied to understand the factors controlling two-dimensional (2D) polymorphism. This is attractive to study as it involves only one type of adsorbate molecule, simplifying analysis. We consider room temperature meso-tetraalkylporphyrin (M-C_mP) SAMs on HOPG as they show controllable polymorphism and synthetic diversity, considering (see Fig. 1): C₁₃P and Co-C₁₃P, presenting their syntheses and STM measurements, C₁₂P, reprocessing observed images (47, 48) using recently (49) developed internal calibration techniques to provide improved quantitative analysis, Zn-C₁₂P, (47) C₁₁P (50), and Cu-C₁₁P (50, 51), as well as briefly also C₁₉P (49).

Different 2D-polymorphs of Cu-C₁₁P have been produced by varying the porphyrin concentration in the supernatant solution, with higher coverage polymorphs forming at higher solution concentrations (51, 52). Increasing the porphyrin concentration once a SAM has formed brings about dynamic changes to the SAM only around defects, however, indicating that the SAMs are kinetically trapped, and this trapping takes only seconds to manifest after first application of the solution to the substrate. Similar results have recently been observed for porphyrin SAMs on Au(111) (34). This kinetic trapping makes thermodynamic quantities difficult to measure and sets a key challenge for future experimental work. As what happens in those first few seconds qualitatively reflects thermodynamics (high-coverage polymorphs develop at high porphyrin concentrations), we introduce the simplistic hypothesis that indeed thermodynamics completely controls this aspect.

Experimentally testing this hypothesis requires quantitative information (7, 8, 33), as has been measured say for co-deposition of single SAMs containing mixed porphyrins (40) to reveal that molecules inside SAM phases are *not* in equilibrium with those in solution. However, only limited data is available concerning relative polymorph propensities. Here, we test it by comparing observed structures and crudely estimated formation free energies to calculated ones. While the uncertainties in the experimentally derived free energies are large, they are obtained using new and existing data of the type commonly available from STM measurements. Our method therefor has widespread application, providing starting points for quantitative analyses of the factors driving SAM formation.

The computational studies also focus on the factors that control free energies of SAM formation and polymorphism. For meso-tetraalkylporphyrin SAMs, chain-length variation can be exploited to obtain understanding of the entropy/enthalpy balance (39) and we consider three different chain lengths, C₁₁ to C₁₃. Solvent dependence can also provide direct means of modulating free energies of polymorphism (43), but herein we do not dwell on this effect, considering only solvents that are likely to have similar critical dielectric and density parameters, ignoring the differences between them.

Results

a. STM images and their calibration. For Cu-C₁₁P, various polymorphs have been observed at the HOPG / 1-octanoic acid interface, and we consider only the low-density (L) and medium-density (M) forms (51). Figs. 2a and 2c show sample images, whilst Fig. 2b shows an interface region (I) containing both motifs. Fig. 2d shows the observed^{27,28} M polymorph for C₁₂P in *n*-tetradecane (no L polymorph has so far been observed in this system). Also shown are the first observed STM images of C₁₃P: in Fig. 2f an L polymorph for C₁₃P measured in 1-phenyloctane, and in Fig. 2e a corresponding M polymorph and also an L-M interface. All of these images are scaled and skewed to an orthogonal coordinate system, as determined by our internal calibration procedures (see Methods); the corresponding raw

images are shown in Supporting Information (SI) Sect. S2, where images for Co-C₁₃P are also shown. SAM geometrical properties, averaged over 10-20 similar images for Cu-C₁₁P, Co-C₁₃P, and C₁₃P are given in Table 1. Of particular note, the observed lattice properties of Co-C₁₃P and C₁₃P are statistically equivalent, a result also observed (48) for Zn-C₁₂P and C₁₂P and for (at least) the M polymorphs of Cu-C₁₁P and C₁₁P (50), simplifying analyses. No co-adsorption of solvent molecules is observed in any of these SAMs.

b. Choice of representative computational schemes. A wide range of methods is available for treating dispersion interactions in electronic-structure calculations (21, 23, 53) and we consider four alternatives. Three present different flavours of dispersion-corrected DFT and are labelled PBE-D3/PAW, PBE-DRSLL/DZP and PW91-D3/USPP. These combinations explore the choice of the density-functional (PBE (54) vs. PW91 (55)), the choice of dispersion correction (D3 (19, 20) vs. DRSLL (18)), and the choice of basis set and core potential (plane-waves with ultrasoft pseudopotentials (USPP) (56) or PAW (57) or the DZP atomic basis set with basis-set superposition error (BSSE) corrections applied to all calculated energies). In addition, an older QM/MM scheme is also used that combines (49) a traditional quality method for representing the *intramolecular* aspects of porphyrin macrocycle and kinked-chain structure, B3LYP/6-31G* (58, 59), with the well-established AMBER *intermolecular* force field (60) and a specialized HOPG-alkane force field that has been fitted (61) to experimental thermodynamic data. The advantages of this approach are in terms of both computational speed and agility as well as the conceptual one that each of its parts has a solid foundation so that quality results are expected. Of particular significance, this method easily treats incommensurate SAM and substrate lattices whereas the other methods demand that calculations be performed at only certain discrete geometries at which the lattices become commensurate (see Methods). Its disadvantage is the inclusion of an empirically derived aspect that would need to be expanded in order to treat many related systems.

c. Simulation of possible SAM polymorph structures. As meso-tetraalkylporphyrins are extremely conformationally flexible and must take on locally distorted “kinks” in order to form SAMs (49), very many atomic structures are feasible. These kinks involve conformation changes to the chain torsional angles specifying the locations of the 3rd, 4th, and 5th C atoms along the chain; 5 low-energy kink patterns leading to stable L polymorphs have been previously identified and related to experimental STM images (49). We consider only these 5 possibilities for the structures of C₁₁P and C₁₃P.

For the M polymorphs, alternate alkyl-chain conformations are required so as to direct chains either into solution or else say to lie on top of the base-porphyrin layer. Many seemingly independent molecular structural features combine to produce many different ways in which the alkyl chains can lift off the surface. While in principle molecular dynamics simulations can be performed to identify feasible structures, this is not feasible for kinetically trapped SAMs. Instead we use a systematic conformational search procedure, constructing all possible variations amongst a set of critical structural variables. Most of the structures so produced can be easily eliminated, however, as rotation of the porphyrin macrocycles around their centres always accompanies changes in the structures of the alkyl chains, and that only very restricted combinations of the identified structural elements lead to viable SAMs. As an example of what can be achieved using systematic conformational searching, for alkanethiol SAMs on Au(111), analogous methods identified over 10⁹ possible conformers of which all but a few hundred could be immediately eliminated (62, 63). For alkylporphyrin SAMs, 9 different conformations were chosen for the chains not lying flat on the surface and then these were combined with the set of 5 feasible kinked structures for the flat-lying chains identified for the L polymorphs. This is done by scanning all possible porphyrin ring angular orientations, eliminating structures with overlapping non-bonded atoms. As a result, as a set of 40 initial SAM structures was prepared for each of the C₁₁P, C₁₂P, and C₁₃P SAMs. Optimization of these initial structures resulted in the identification

for each molecule of at most 22 unique M SAM polymorphs. These structures were then assessed based on their calculated free energies of formation at 298 K and compared to the observed STM images.

d. Identification of SAM atomic structures by comparison of observed and calculated STM images. Comparison of the observed STM images and lattice parameters for both pure phases and for interfaces with those from the 27 unique optimized structures (5 for L, 22 for M) reveals that for each observed image only one optimized structures matches. The matching structure is always either the one of calculated lowest free-energy or else the second-lowest energy one, as discussed quantitatively later in Subsection f. These are overlaid on the observed images in Fig. 2, highlighted in Fig. 3, and specified in full in the SI Sects. S8 and S9, whilst calculated and observed lattice parameters are compared in Table 1. The simulated images are prepared using the Tersoff-Hamann approximation (64), displaying surfaces at a constant value of the electron density chosen such that resolved features parallel those resolved in the observed constant-current images. SI Sect. S5 compares the actual observed and calculated images, indicating that the calculations offer explanation for the key resolved internal structural features. This verifies that the feature assignments implied by the simplistic atomic-coordinate overlay shown in Fig. 2 are indeed valid.

Overviewing the results shown in Fig. 2, the structures for the L polymorph have all chains lying flat on the surface whilst the M polymorphs have two chains so positioned and two chains dangling into solution. Interface molecules between the two polymorphs typically have one chain in solution only. Across interfaces, the orientations of the porphyrin macrocycle rings and two of the alkyl chains do not change significantly, meaning that just changes in the conformations of the other two alkyl-chains to/from their flat-lying orientations distinguish the atomic structures of the molecules in the abutted L, I, and M regions. The same structural motif is preserved for the L polymorphs of $C_{11}P$ and $C_{13}P$ and corresponds to that observed (49) for $C_{19}P$, but the M polymorphs of $C_{11}P$, $C_{12}P$, and $C_{13}P$ all take on different forms. In particular, the chains not lying on the surface can orient in a variety of ways,

producing complex three-dimensional patterns above the surface, as highlighted in Fig. 3. These patterns are controlled equally by the interactions with the completed SAM layer, interactions between different dangling chains, conformational energy costs, and, most important, solvation effects.

Considering the observed STM images in more detail, we note that metalloporphyrins may generate significant STM signal from their metal centres, but otherwise images are dominated by the kinks in the alkyl chains at their β and γ positions (49). Sometimes atomically resolved tunnelling is also seen amongst the long flat-lying chain regions as well, dominated by the chain hydrogen atoms, (65, 66). Naively one hydrogen atom per methylene group is expected to be observed, and indeed sometimes is, but more often the images reveal only a hydrogen atom on every second group. This arises owing to the chains not sitting perfectly flat at room temperature and to consequences of the SAM lattice being incommensurate with that of HOPG (65). In all cases the alkyl chains are calculated by QM/MM and observed to lie parallel or nearly parallel to the HOPG substrate $\langle 11\bar{2}0 \rangle$ lattice vector, as has been observed for the SAMS of pure alkanes (65-68). These calculations and all measurements also indicate that the SAMs are incommensurate with the substrate lattice (49, 51), and Table 1 shows good agreement between the observed and calculated SAM lattice parameters.

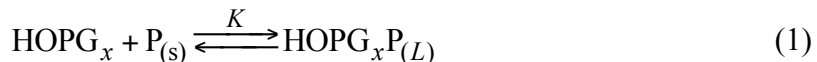
Considering the comparison (Table 1) between observed and calculated lattice parameters in more detail, we note that agreement is excellent except for two outliers. For the L polymorphs, differences of 1 Å are found for the parameters a and c (see Table 1 and Fig. 2), with corresponding deviations in the associated angles of a few degrees. However, the observed interface structures for both C₁₁P and C₁₃P (see Table 1) portray L lattice parameters that are slightly distorted and actually much closer to those calculated for the pure L polymorphs, indicating that the L polymorphs are quite pliable. Indeed, calculated electronic interaction energies between the surface and the SAM increase by only up to 5 kcal mol⁻¹ when the porphyrin rings are forcibly rotated about the surface normal by 15°, driving large changes to the optimized lattice parameters. Hence the difficulty that the calculations

perceive in predicting accurately these lattice parameters leads to an explanation of a significant unexpected observed property of the L SAMs. At first observation, the large variation observed between structures was believed sufficient to warrant the introduction of a second label “B” depicting interacting interface molecules (51), whereas we are now inclined to interpret “B” as being instead just a form of “L” distorted by the boundary conditions imposed on the interface.

The other outlier is for the M polymorph of C₁₃P and seems to be the *opposite* in nature: the observed structure appears to be very unique and quite strictly defined. For this, the observed STM images display some very unusual features and of all the optimized structures for this polymorph, only *one*, the structure presented, reproduces them; these features include the very short distance between chain kinks and the critical observation that the macrocycle ring has the same angular orientation in both the L and M polymorphs, with the chain kinks pointing directly in the *b'* direction (see Fig. 2f). The calculations therefore indicate that this structural motif is intrinsically one of high free energy and that its stabilization for C₁₃P is fortuitous. It is also an unintuitive structure as it fails to cover significant portions of the HOPG surface (black regions in Fig. 3), unlike almost all of the other optimized structures for C₁₃P. This highlights the general importance of the specific size-dependent effects on the structure and properties of medium-density polymorphs and the need for a thorough scan of plausible atomic structures when interpreting even high-resolution STM images such as those obtained.

e. Estimating free energies of formation from observed SAM properties. While formed SAMs are kinetically trapped (34, 52), our hypothesis explaining the formation of higher density polymorphs at higher solution porphyrin concentrations is that equilibrium thermodynamics controls the first few seconds of SAM production. We label the porphyrin molecules as either P_(L), P_(M), or P_(s) depending on whether they are in the L polymorph, M polymorph, or in solution, respectively, and represent the (non-integral) number of carbon atoms of HOPG per porphyrin molecule in a SAM as *x*. The raw (but

solvated) surface is therefore labelled as HOPG_x , the L polymorph as $\text{HOPG}_x\text{P}_{(L)}$, and the M polymorph as $\text{HOPG}_x\text{P}_{(M)}$ (note $x' \neq x$ as the coverages differ). Formation at low concentration of the SAM of the L polymorph is then expressed as:



and the equilibrium constant for this reaction is given by

$$K = \frac{[\text{HOPG}_x\text{P}_{(L)}]}{[\text{HOPG}_x][\text{P}_{(s)}]} = \frac{\kappa}{[\text{P}_{(s)}]} \quad (2)$$

where κ is the ratio of the surface areas of regions containing the SAM to regions of uncovered HOPG. The standard Gibbs free-energy change for SAM formation for n moles of adsorbed porphyrin is then given by

$$\Delta G_L = nRT \ln \frac{[\text{P}_{(s)}]}{\kappa C^0} \quad (3)$$

where C^0 specifies a standard state of 1 M. This relationship allows estimates of SAM free-energies of formation based on the SAM formation conditions provided that some (crude) estimate is available for κ . Under our experimental conditions, dense highly ordered SAMs cover most of the surface but disordered regions that are presumably also quite dense are observed as well. The fraction of seemingly uncovered surface was never observed to be greater than 1% ($\kappa = 10^2$) and here we take a rough estimate of $\kappa = 10^3$, but larger values are plausible. Estimated free energies of formation change by $-2.303RT \sim -1.4 \text{ kcal mol}^{-1}$ per decade increase in κ . While uncertainties of this order are large compared to what accurate experimental measurements can achieve, they are tiny compared to the magnitudes of the competitive forces controlling SAM production (see later) and so facilitate commencement of quantitative analysis.

For Cu-C₁₁P, stable SAMs of the L polymorph are observed (51) at $[P_{(s)}] = 10^{-7}$ M in 1-octanoic acid, implying a SAM formation free energy of $\Delta G_L / n = -13.7$ kcal mol⁻¹ assuming $\kappa = 10^3$. Here we find stable SAMs for the L polymorph of C₁₃P in 1-phenyloctane at $[P_{(s)}] = 10^{-5}$ M, implying $\Delta G_L / n = -12.3$ kcal mol⁻¹. These crude estimates are reported in Table 2.

Reporting free-energy changes in the form of the associated *chemical potential* changes $\Delta G / n$ is common for the consideration of the thermodynamics of solids, liquids, and gases, and we report such quantities so that SAM formation may be readily understood in terms of traditional chemical analyses. However, the total Gibbs free energy of a SAM is proportional to the available substrate surface area, making the thermodynamically most stable SAM that which has the most negative value of $\Delta G / A$, where A is the surface area covered by the SAM per molecule ($A = ab \sin \theta$, see Fig. 2 and Table 1) (3). These values are also reported in Table 2. Only $\Delta G / A$, *not* the chemical potential, is a thermodynamic state property whose value can be simply manipulated during SAM composition changes. So while the free-energy of production of the M polymorph can be described by analogy to Eqn. (1) using



equilibrium also with the L polymorph means that Eqn. (2) has no analogue. The chemical potential of the M polymorph can be expressed in terms of that L, the covered surface areas per molecule A_M and A_L , and the porphyrin solution concentration at which both polymorphs are equally represented as

$$\frac{\Delta G_M}{n} = \frac{A_M}{A_L} \frac{\Delta G_L}{n} + \left(1 - \frac{A_M}{A_L}\right) RT \ln \frac{[P_{(s)}]}{C^0} . \quad (5)$$

For Cu-C₁₁P and C₁₃P, the two phases are stable at porphyrin concentrations $[P_{(s)}]$ of 10^{-5} M and 10^{-4} M, respectively, therefore implying $\Delta G_M / n \approx -12.0$ and -10.9 kcal mol⁻¹. These values are listed in Table 2, as are the associated values of $N_A \Delta G / A$ (where N_A is Avagadros' number) which, as one

would intuitively expect, are more negative for M than for L. Improved estimates could be obtained by systematically varying the solution concentration and recording many images, but the estimates we obtain using only known information are accurate to of order 1 kcal mol⁻¹ and can be readily applied to interpret many previously observed images.

f. Comparison of observed and calculated free energies of SAM formation. Table 2 compares the observed (ΔG^{OBS}) and calculated ($\Delta G^{\text{QM/MM}}$, $\Delta G^{\text{PBE-D3/PAW}}$, $\Delta G^{\text{PBE-DRLSS/DZP}}$, and $\Delta G^{\text{PW91-D3/USPP}}$) free energies of SAM formation from solution. The computational methods predict $\Delta G/n$ values that differ by up to 40 kcal mol⁻¹ from each other, a large range that would not be expected if this computational field was mature. The a-priori methods used have been optimized for the treatment of small van der Waals dimers, not for condensed-matter properties such as SAM formation, and improved methods are likely to be developed in the future. Two of the computational methods, the QM/MM scheme and PBE-D3/PAW, agree qualitatively with experiment in that the variations in ΔG with alkane chain length and with polymorph variations are small; both methods underestimate the binding strengths by ~ 6 kcal mol⁻¹, however. The other two methods PBE/DZP-D and PW91-D3/PAW significantly overestimate the attraction and cannot account for the observed chemical variations; hence these methods are not preferred. The variable nature of the PBE-DRSLL/DZP results most likely stem from the use of the small double-zeta DZP basis set and the imprecision of BSSE corrections. However, D3 corrected PW91 also yields erratic results during the fitting of the two D3 parameters to the test data set (see SI. Sect S2). While this functional often gives results better than any other similar one when dispersion is excluded, including dispersion appears to expose its weaknesses.

For the conserved L polymorph of C₁₁P, C₁₃P, and C₁₉P, five possible structure types have been considered, as discussed previously (49) for C₁₉P; this set explores various ways in which the chains can lie on the surface (the plane through the carbon-carbon bonds can be parallel or perpendicular), and various ways of making the required chain kinks. The lowest-energy structure calculated for each of the

porphyrins is indeed the one that matches the experimental observations, with all other structures being calculated by QM/MM to be at least $\Delta G/n = 7 \text{ kcal mol}^{-1}$ (0.3 eV) higher in energy. For the M polymorphs of each porphyrin, ~ 4 structures had a calculated free energy at most 5 kcal mol^{-1} higher than the lowest energy structure, with either the lowest-energy structure or the second-lowest-energy one (typically 1 kcal mol^{-1} higher in energy) matching the experimental images. The calculations are therefore useful for *predicting* the polymorph atomic structures as well as for *identifying* them.

g. The factors controlling SAM binding and polymorphism. As the calculations are able to predict the geometry and free-energy of the SAMs and their chemical variation with chain length and polymorph, it is anticipated that they also provide a realistic description of the driving forces. While a detailed breakdown of the energy components calculated by all 4 computational methods is given in Table 2, key features are indicated in Fig. 4 for the L and M polymorphs of $C_{13}P$ evaluated using QM/MM and PBE-D3/PAW (an expanded figure showing similar trends predicted by the PBE-DRSLL/DZP and PW91-D3/USPP methods is provided in SI Sect. S6). The most striking feature of Fig. 4 is the energy scale, which ranges over $200 \text{ kcal mol}^{-1}$ (9 eV). Whilst individual interatomic dispersion interactions are weak, their collective strengths driving the solvent interactions, the intermolecular interactions and the molecule-substrate interactions are very large indeed.

Fig. 4a views the total free energies of SAM formation of $C_{13}P$ L and $C_{13}P$ M in terms of the solvation, thermal, and binding contributions listed in Table 2. The thermal contributions opposing SAM formation of $\Delta G_{\text{corr}}/n = 37$ and 25 kcal mol^{-1} for L and M, respectively, are dominated by the entropy contributions at 298 K of $-T\Delta S/n = 38$ and 26 kcal mol^{-1} , respectively. Also opposing SAM formation are differential solvation energy contributions of $\Delta G_s/n = 106$ and 70 kcal mol^{-1} , respectively. Driving SAM formation are the binding energy contributions of $\Delta E^{\text{QM/MM}}/n = -149$ and $-103 \text{ kcal mol}^{-1}$, respectively, calculated by QM/MM, or of $\Delta E^{\text{PBE-D3/PAW}}/n = -148$ and $-100 \text{ kcal mol}^{-1}$ calculated by

PBE-D3/PAW. Near-perfect cancellation of these opposing large effects results in very similar free energies for the L and M polymorphs (the maximum calculated L-M difference is just 2 kcal mol⁻¹). Table 2 show that this cancellation is modulated as the alkyl chain length changes from 11 to 13 carbon atoms to control structure.

The net solvation energy ΔG_s itself arises from the differences between the large solvation energies $\Delta G_{s,P}$, $\Delta G_{s,HOPG}$, and $\Delta G_{s,SAM}$ of the molecules in solution, the bare HOPG surfaces (which differ in area for L and M), and the SAMs on these surfaces, respectively. Details are given in Table 2 and depicted diagrammatically in Fig. 4b.

Similarly, Table 2 and Fig. 4d show that the QM/MM binding energies $\Delta G^{QM/MM}$ can be decomposed into: small distortion energies required to kink the alkyl chains ΔE_{dist} , significant alkyldipyrromethane to alkyldipyrromethane intermolecular interactions between neighbouring molecules in the SAMs ΔE_{P-P} , and strong attraction between the alkyldipyrromethanes and HOPG ΔE_{P-HOPG} , made up of roughly equal contributions coming from alkyl chain and central porphyrin ring interactions. While the distortion energies in absolute magnitude are quite high at 4-6 kcal mol⁻¹, this effect is dwarfed by the magnitudes of the other contributing terms, explaining how a wide range of chain conformers can be stabilized. Also, as the intermolecular interactions and substrate interactions are large, we see that subtle modulation effects could have a profound impact on relative polymorph free energies, explaining the diverse nature of observed SAM properties.

While dispersion is naively expected to be critical to binding, Fig. 4c shows an important result: the raw PBE density functional predicts that the molecules in the gas phase would be *repelled* by HOPG. Use of dispersion correction is essential to even the most qualitative description of the binding, contributing ca. -175 and -115 kcal mol⁻¹ for the L and M polymorphs of C₁₃P, respectively (Table 2). This effect is well-known for small-molecule interactions in gas-phase dimers and in molecular solids (18-24), and also for SAMs of small molecules (69, 70). However the magnitude of the effect we see

for SAMs of large molecules is unprecedented, not just in absolute value but also in terms of its variations between feasible polymorphs. This stresses the criticality of dispersion forces to the self-assembly processes - not only for oligoporphyrin molecules but also for natural and synthetic polymers as well.

Discussion

This work presents the first use of modern electronic-structure computational methods to determine a priori the free energy of formation of a SAM from solution. It shows that computational techniques standard for the interpretation of gas-phase chemistry and small-molecule solution chemistry also depict complex self-assembly processes of large molecular systems. Calculations can reveal and/or predict the atomic structure of partially resolved SAMs as well as expose the key chemical features controlling SAM production and polymorphism.

Our application is specifically for tetraalkylporphyrin SAMs on HOPG for which the interplay between kinetic and thermodynamic effects in controlling the properties of hydrophobic SAMs is of considerable current interest. As our calculated free energies parallel those implied by observed SAM compositions, support is given to the hypothesis that thermodynamics controls the critical first few seconds of SAM growth prior to the SAMs becoming kinetically trapped. This hypothesis therefore warrants the development of new techniques permitting further experimental and computational investigation.

However, our methods are of general use and can be applied to all types of self-assembly processes on surfaces from solution, as well as to general processes involving macromolecular self assembly. As an example of the general importance of the concepts discussed, one critical feature found here for hydrophobic SAMs, the importance of intermolecular dispersion interactions in controlling SAM structure (and chirality), has recently been demonstrated also for gold-sulfur SAMs with strong

head-group interactions (62, 63). In terms of specific details, in a fully general method Eqn. (6) would need to be replaced with one of the many available solvation-energy protocols embodying the use of arbitrary atoms and also electrostatic contributions. Such protocols may need slight modification to support the required 2-dimensional boundary conditions of the SAM and/or solvent-excluded surfaces.

Some of the computational and experimental techniques combined in this work will be essential to many subsequent studies. These include advances in theoretical methods to treat the entropy of a SAM as opposed to standard gas-phase entropies, and methods to treat incommensurate lattices (see Materials and Methods). Our advances in STM calibration and data analysis allow for accurate lattice parameters and molecular shapes to be determined, as we demonstrate using either commercial equipment (at U. Sydney), or purpose-built apparatus (at Nijmegen), or by re-processing already published images. These advances will all help drive state of the art modern experiments aimed at determining the thermodynamics and kinetics of SAM production, allowing for greatly enhanced control procedures of widespread applicability through nanotechnology and molecular electronics.

Many utilized experimental and computational techniques require significant improvement, however. Our method for crudely estimating free energies of SAM formation provides a means to engage in quantitative thermodynamics discussions as it can be used to reprocess a wide range of already published data. Its accuracy is below that needed in the future, however, and well below that already achieved in pioneering studies performed over the last few years. New and accurate experimental techniques need to be developed. The four computational methods used were chosen to explore very different ways of completing the calculations. If essentially the same results had been predicted independent of method then this would have provided a powerful statement of confidence in the methods, as would be expected in a mature research field. Instead two very different and yet reasoned highly applicable methods, QM/MM and PBE-D3/PAW, encouragingly gave very similar results whilst PBE-DRSLL/DZP and PW91-D3/USPP showed anticipatable flaws. From this we conclude that plane-wave basis sets are much more suited to SAM calculations than are atomic basis

sets, owing to the criticality of BSSE, and that methods focusing on correctly representing chemical interactions are preferred to those that can give good results by cancellation of internal errors. The PW91 method was for a long time the preferred method for calculations involving intermolecular interactions as its poor representation of covalent bonds gave more useful results than methods that described them better such as PBE. When dispersion is evaluated more rigorously, as in PBE-D3 and PW91-D3, only PBE-D3 gives realistic results, however. PBE-D3 is still somewhat crude given the wide range of improved hybrid, double-hybrid, and meta-GGA density-functionals now available, including those with correct behaviour of the long-range potential. Our study shows how important it is to do accurate calculations in which likely errors are reduced to below 1 kcal mol⁻¹. This is a difficult task for just the interaction energy, let alone the required additional solvation and thermal energies. The obtained results imply that the D3 explicit dispersion correction and the Floris-Tomassi-Pascual Ahuir solvation dispersion interaction are finely balanced. This is a profound result as these terms were determined independently and are usually applied independently. It requires further investigation.

Many challenges thus remain, yet the benefits are high. Calculations using implicit solvent require dramatically less resources than do molecular dynamics simulations, allow much higher quality methods to be used for evaluating energies, can be readily applied to consider chemical diversity (e.g., herein to ca. 150 self-assembly processes in solution), and directly reveal the thermodynamic quantities of greatest interest to the understanding of new phenomena.

Materials and Methods

C₁₃P and Co-C₁₃P were synthesised in gram amounts by the methods of Crossley et al. (71), while established STM imaging and calibration techniques were used, as described in SI Sects. S1 and S2, respectively.

The QM/MM calculations were performed using a method previously applied to meso-tetraalkylporphyrin SAMS (49). A limitation of this approach is that it explicitly treats the HOPG surface as being rigid. However it is justified as PBE-D3/PAW calculations indicate that substrate relaxation contributes less than 1 kcal mol⁻¹ to binding energies. The GAUSSIAN (28) EXTERNAL command is used to include the non-standard force-field elements into an ONIOM (72) QM/MM force field. Long-range intermolecular interactions are treated using a real-space sum over replicated images of the incommensurate HOPG and SAM lattices, with the unit-cell parameters optimized externally to GAUSSIAN using analytical derivatives. All atomic geometry optimizations are performed using initial analytical second derivatives for the Hessian matrix. All optimized structures are confirmed to be local minima by second-derivative analysis. Entropy is evaluated using the standard procedures applied by GAUSSIAN (73) except that gas-phase translational and rotational contributions are not included for the SAMs and are replaced by vibrational contributions derived from calculated phonon mode frequencies; the phonon modes and their interactions with the molecular vibrations are evaluated using a real-space sum over all vibrations of a 5 × 5 grid of SAM unit cells. As the HOPG surface is taken to be frozen, it does not contribute to the free energy. The net “thermal” correction to the free energy is therefore taken to be the sum of the entropy, enthalpy, and zero-point energy differences calculated for the SAM and for the adsorbate molecule in the gas phase. This is then combined with a solvation entropy correction, needed for conversion of the gas-phase standard state of 1 atm concentration to the solution conditions, of 1 M of -1.89 kcal mol⁻¹ (73, 74). Changes in entropy completely dominate the net thermal correction, reflecting the confinement that the molecules experience when the SAM forms. During these calculations, the geometry optimizations are always performed so as to minimize the binding energy only. Results from PBE-D3/PAW calculations also justify this approach as they have been performed optimizing both this binding energy and $\Delta G/A$, but no significant geometrical differences were found.

The PBE-D3/PAW calculations were performed using the VASP package (74, 75) to optimize the molecular geometries using 12-15 commensurate lattices with lattice vectors similar to the QM/MM optimized lattices. Periodic imaging of the dispersion interactions is not performed in the direction normal to the SAM surfaces, and not performed at all for “isolated” molecule calculations. All calculations are performed at the Γ -point of the Brillouin zone using the default precision (plane wave cutoff 300 eV); this basis set is sufficiently close to the complete-basis-set limit so that basis-set superposition error is not manifest in these calculations. Results are reported for only the commensurate lattice that produces the lowest value of $\Delta G / A$. The correct orientation of the alkyl chains with respect to the HOPG lattice is not preserved in this process and as a result of this and the use of discrete lattice points, calculated free energies are likely to be too positive by up to 2-3 kcal mol⁻¹. STM images are simulated using the Tersoff-Hamann approximation (64) at a bias voltage of 0.8 V.

The PW91-D3/USPP calculations were similarly formed using VASP while the PBE-DRSLL/DZP calculations were performed using SIESTA (76). In both cases only the VASP/PAW-D3 optimized incommensurate lattice was considered. For the PW91 density functional, the parameters in the D3 functional form (19) with Becke-Johnson damping (20) were determined to be $s_6=1$, $s_8=1.9598$, $a_1=0.6319$, and $a_2=4.5718$ by fitting data for 130 non-covalent energies, see SI Sect. S3, but the fit was unusually poor, indicative of the poor quality of the PW91 functional. Dispersion-corrected PBE is thought to be always preferable to dispersion-corrected PW91.

For the PBE-D3/PAW, PBE-DRSLL/DZP and PW91-D3/USPP calculations, QM/MM values of the thermal corrections are added to the native calculated binding energies. While in principle such terms can be evaluated directly (3), the size of the system treatable in this fashion remains small and the use of more approximate methods for thermochemical analysis is standard chemical practice (1, 2).

Solvation energies are evaluated at geometries calculated by each computational method. However, the variations in these quantities as a function of the associated geometry changes is found to be small, too small to be apparent in Fig. 3. In most cases, geometries are optimized using the surface-

adsorbate energy and forces only, as is often a common practice. Solvation can effect geometry, however (77), and we find that this applies to SAMs with tight internal cavities like the identified structure for C₁₂P M (Fig. 3). Only for this structural motif were the calculations performed alternatively, optimizing the energy in the presence of the solvent. Note that an implementation of this type of optimizations (without a solvent-excluded-surface option) has just become available in VASP (29).

To obtain the net contribution of solvation energy to the free energy of SAM formation, individual energies are evaluated for the adsorbate molecules in free solution, for the raw HOPG surface, and for the SAMs. This process fully includes the effects of wetting/dewetting surfaces and dangling chains and compliments the entropic contributions previously included within the “thermal” correction.

Each solvation energy is determined using an interpolating function. The parameters in this function were determined by calculating the dispersion-dominated solvation energies of gas-phase molecules by GAUSSIAN using a dielectric constant of 2.5, solvent density of toluene, and a refractive index of 1.49 to mimic all three solvents of interest 1-phenyloctane, 1-octanoic acid, and *n*-tetradecane. In this process, the method of Floris, Tomasi, and Pascual Ahuir (27) is used to depict the dispersive contributions to the solvation energy, a method that is in widespread current use that is reported to be a useful component in calculations that predict thermochemical properties of small molecules to chemical accuracy, ca. 2 kcal mol⁻¹ (78). This approach expresses the dispersion contribution to the solvation free energy as being proportional to the exposed molecular surface area (27, 29) and so we express the change in solvation energy on SAM formation as

$$\Delta E_s = \Delta E_{s,SAM} - \Delta E_{s,P} - \Delta E_{s,HOPG} = -\alpha(A_{SAM} - A_P - A_{HOPG}) \quad (6)$$

where $\Delta E_{s,SAM}$, $\Delta E_{s,P}$, and $\Delta E_{s,HOPG}$ are the solvation energies of the SAM, the porphyrin in solution, and a solvated bare HOPG surface, respectively, and A_{SAM} , A_P , and A_{HOPG} are the corresponding

exposed surface areas. The proportionality constant $\alpha = 0.0866 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$ is fitted to B3LYP/6-31G* calculated data for the optimized structures of C₁₁P and C₁₉P. All exposed surface areas are evaluated using the solvent-excluded surface obtained using van der Waals radii of 1.55 Å, 1.7 Å, and 1.2 Å for N, C, and H, respectively, while the solvent “radius” is taken to be 2.2 Å. This procedure ensures that only the SAM surface area is included next to which one of the alkyl chains from the 1-phenyloctane, 1-octanoic acid, or *n*-tetradecane solvent molecules can be placed, eliminating any small internal cavities from consideration. These parameters depict the exposed surface area for HOPG as being ca. 3.3% larger than the substrate-lattice surface area A , owing to the atomic structure of HOPG.

Acknowledgments

We thank the Australian Research Council for funding this research (grants LP0455238, DP12010259, and DE140100550) and National Computational Infrastructure (NCI) and INTERSECT for provision of computing resources. J.A.A.W.E. thanks the Council for the Chemical Sciences of the Netherlands Organization for Scientific Research (CW-NWO) for a Vidi grant (700.58.423), and the European Research Council for an ERC Starting Grant (NANOCAT-259064), and the Ministry of Education, Culture and Science (Gravity program 024.001.035).

References

1. Curtiss LA, Redfern PC, Raghavachari K (2011) G n theory. *WIREs Comput Mol Sci* 1:810-825.
2. Karton A, Rabinovich E, Martin JML, Ruscic B (2006) W4 theory for computational thermochemistry: In pursuit of confident sub-kJ/mol predictions. *J. Chem. Phys.* 125:144108.
3. Reuter K, Scheffler M (2001) Composition, structure, and stability of RuO₂ (110) as a function of oxygen pressure. *Physical Review B* 65:035406.
4. Tomasi J, Mennucci B, Cammi R (2005) Quantum Mechanical Continuum Solvation Models. *Chem. Rev.* 105:2999-3094.
5. Zhang J, Bilic A, Reimers JR, Hush NS, Ulstrup J (2005) Coexistence of Multiple Conformations in Cysteamine Monolayers on Au(111). *J. Phys. Chem. B* 109:15355-15367.
6. Rodziewicz P, Meyer B (2014) Interplay between molecule-molecule and molecule-substrate interactions: first-principles study of fluoroform aggregates on a hexagonal ice (0001) surface. *Phys. Chem. Chem. Phys.* 16:940-954.

7. Lei S, *et al.* (2008) One building block, two different supramolecular surface-confined patterns: concentration in control at the solid-liquid interface. *Angew. Chem., Int. Ed.* 47:2964-2968.
8. Bellec A, *et al.* (2011) Solution-growth kinetics and thermodynamics of nanoporous self-assembled molecular monolayers. *J. Chem. Phys.* 134:124702/124701-124707.
9. Gutzler R, *et al.* (2010) Reversible Phase Transitions in Self-Assembled Monolayers at the Liquid-Solid Interface: Temperature-Controlled Opening and Closing of Nanopores. *J. Am. Chem. Soc.* 132:5084-5090.
10. Mammen M, Shakhnovich EI, Whitesides GM (1998) Using a convenient, quantitative model for torsional entropy to establish qualitative trends for molecular processes that restrict conformational freedom. *J. Org. Chem.* 63:3168-3175.
11. Mammen M, Shakhnovich EI, Deutch JM, Whitesides GM (1998) Estimating the Entropic Cost of Self-Assembly of Multiparticle Hydrogen-Bonded Aggregates Based on the Cyanuric Acid-Melamine Lattice. *J. Org. Chem.* 63:3821-3830.
12. Blunt MO, *et al.* (2013) Temperature-Induced Structural Phase Transitions in a Two-Dimensional Self-Assembled Network. *J. Am. Chem. Soc.* 135:12068-12075.
13. Killian BJ, Kravitz JY, Gilson MK (2007) Extraction of configurational entropy from molecular simulations via an expansion approximation. *J. Chem. Phys.* 127:024107.
14. Künzel D, Groß A (2013) Influence of the solvent on the stability of bis(terpyridine) structures on graphite. *Beilstein Journal of Nanotechnology* 4:269-277.
15. Meier C, *et al.* (2010) Concentration and Coverage Dependent Adlayer Structures: From Two-Dimensional Networks to Rotation in a Bearing. *J. Phys. Chem. C* 114:1268-1277.
16. Hensen U, Lange OF, Grubmueller H (2010) Estimating Absolute Configurational Entropies of Macromolecules: The Minimally Coupled Subspace Approach. *Plos One* 5:e9179/9171-9178.
17. Nirmalraj PN, Thompson D, Riel HE (2015) Capturing the embryonic stages of self-assembly - design rules for molecular computation. *Sci. Rep.* 5.
18. Dion X, Rydberg H, Schröder E, Langreth DC, B. I. Lundqvist (2004) *Phys. Rev. Lett.* 92:246401.
19. Grimme S, Antony J, Ehrlich S, Krieg H (2010) A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *J. Chem. Phys.* 132:154104.
20. Grimme S, Ehrlich S, Goerigk L (2011) Effect of the Damping Function in Dispersion Corrected Density Functional Theory. *J. Comput. Chem* 32:1456-1465.
21. Risthaus T, Grimme S (2013) Benchmarking of London Dispersion-Accounting Density Functional Theory Methods on Very Large Molecular Complexes. *J. Chem. Theory Comput.* 9:1580-1591.
22. Moellmann J, Grimme S (2014) DFT-D3 Study of Some Molecular Crystals. *J. Phys. Chem. C* 118:7615-7621.
23. Goerigk L (2014) How Do DFT-DCP, DFT-NL, and DFT-D3 Compare for the Description of London-Dispersion Effects in Conformers and General Thermochemistry? *J. Chem. Theory Comput.* 10:968-980.
24. Klimeš J, Michaelides A (2012) Perspective: Advances and challenges in treating van der Waals dispersion forces in density functional theory. *J. Chem. Phys.* 137:120901.
25. Goerigk L, Grimme S (2011) A thorough benchmark of density functional methods for general main group thermochemistry, kinetics, and noncovalent interactions. *Phys. Chem. Chem. Phys.* 13:6670-6688.
26. Sure R, Antony J, Grimme S (2014) Blind Prediction of Binding Affinities for Charged Supramolecular Host-Guest Systems: Achievements and Shortcomings of DFT-D3. *J. Phys. Chem. B* 118:3431-3440.
27. Floris FM, Tomasi J, Pascual Ahuir JL (1991) Dispersion and Repulsion Contributions to the Solvation Energy: Refinements to a Simple Computational Model in the Continuum Approximation. *J. Computat. Chem.* 12:784-791.
28. Frisch MJ, Trucks GW, Schlegel HB, *et al* (2009) *Gaussian 09, Revision C.09* (Gaussian, Inc., Pittsburgh PA).
29. Mathew K, Sundararaman R, Letchworth-Weaver K, Arias TA, Hennig RG (2014) Implicit solvation model for density-functional study of nanocrystal surfaces and reaction pathways. *J. Chem. Phys.* 140:084106.
30. Mozgawa K, Mennucci B, Frediani L (2014) Solvation at Surfaces and Interfaces: A Quantum-Mechanical/ Continuum Approach Including Nonelectrostatic Contributions. *Journal of Physical Chemistry C* 118:4715-4725.

31. Amovilli C, Floris FM (2015) Study of Dispersion Forces with Quantum Monte Carlo: Toward a Continuum Model for Solvation. *J. Phys. Chem. A* 119:5327-5334.
32. Klamt A (2011) The COSMO and COSMO-RS solvation models. *Wiley Interdisciplinary Reviews: Computational Molecular Science* 1:699-709.
33. Mazur U, Hipps KW (2015) Kinetic and thermodynamic processes of organic species at the solution-solid interface: the view through an STM. *Chemical Communications* 51:4737-4749.
34. Bhattarai A, Mazur U, Hipps KW (2014) A Single Molecule Level Study of the Temperature-Dependent Kinetics for the Formation of Metal Porphyrin Monolayers on Au(111) from Solution. *J. Am. Chem. Soc.* 136:2142-2148.
35. Hu F, *et al.* (2014) Temperature-induced transitions of self-assembled phthalocyanine molecular nanoarrays at the solid-liquid interface: from randomness to order. *Nanoscale* 6:4243-4249.
36. Jahanbekam A, Vorpahl S, Mazur U, Hipps KW (2013) Temperature Stability of Three Commensurate Surface Structures of Coronene Adsorbed on Au(111) from Heptanoic Acid in the 0 to 60°C Range. *J. Phys. Chem. C* 117:2914-2919.
37. Friesen BA, Bhattarai A, Mazur U, Hipps KW (2012) Single Molecule Imaging of Oxygenation of Cobalt Octaethylporphyrin at the Solution/Solid Interface: Thermodynamics from Microscopy. *J. Am. Chem. Soc.* 134:14897-14904.
38. English WA, Hipps KW (2008) Stability of a Surface Adlayer at Elevated Temperature: Coronene and Heptanoic Acid on Au(111). *J. Phys. Chem. C* 112:2026-2031.
39. Miyake Y, *et al.* (2012) Entropy-Controlled 2D Supramolecular Structures of N,N'-Bis(n-alkyl)-naphthalenediimides on a HOPG Surface. *ACS Nano* 6:3876-3887.
40. Ikeda T, Asakawa M, Miyake K, Goto M, Shimizu T (2008) Scanning Tunneling Microscopy Observation of Self-Assembled Monolayers of Strapped Porphyrins. *Langmuir* 24:12877-12882.
41. Otsuki J, Kawaguchi S, Yamakawa T, Asakawa M, Miyake K (2006) Arrays of Double-Decker Porphyrins on Highly Oriented Pyrolytic Graphite. *Langmuir* 22:5708-5715.
42. Ikeda T, *et al.* (2004) STM Observation of Alkyl-Chain-Assisted Self-Assembled Monolayers of Pyridine-Coordinated Porphyrin Rhodium Chlorides. *Langmuir* 20:5454-5459.
43. Sirtl T, *et al.* (2013) Solvent-Dependent Stabilization of Metastable Monolayer Polymorphs at the Liquid-Solid Interface. *ACS Nano* 7:6711-6718.
44. Song W, Martsinovich N, Heckl WM, Lackinger M (2014) Thermodynamics of halogen bonded monolayer self-assembly at the liquid-solid interface. *Chemical Communications* 50:13465-13468.
45. Song W, Martsinovich N, Heckl WM, Lackinger M (2013) Born-Haber Cycle for Monolayer Self-Assembly at the Liquid-Solid Interface: Assessing the Enthalpic Driving Force. *J. Am. Chem. Soc.* 135:14854-14862.
46. Song W, Martsinovich N, Heckl WM, Lackinger M (2014) Thermodynamics of 4,4[prime or minute]-stilbenedicarboxylic acid monolayer self-assembly at the nonanoic acid-graphite interface. *Phys. Chem. Chem. Phys.* 16:13239-13247.
47. Visser J, Katsonis N, Vicario J, Feringa BL (2009) Two-Dimensional Molecular Patterning by Surface-Enhanced Zn-Porphyrin Coordination. *Langmuir* 25:5980-5985.
48. Katsonis N, *et al.* (2006) Self-Organized Monolayer of meso-Tetradodecylporphyrin Coordinated to Au(111). *J. Am. Chem. Soc.* 128:15537-15541.
49. Chin Y, *et al.* (2011) Atomic-Resolution Kinked Structure of an Alkylporphyrin on Highly Ordered Pyrolytic Graphite. *J. Phys. Chem. Lett.* 2:62-66.
50. Hulsken B, *et al.* (2006) Scanning tunneling microscopy and spectroscopy studies of porphyrins at solid-liquid interfaces. *Japanese Journal of Applied Physics, Part 1: Regular Papers and Short Notes and Review Papers* 45:1953-1955.
51. Coenen MJJ, *et al.* (2013) Polymorphism in porphyrin monolayers: the relation between adsorption configuration and molecular conformation. *Phys. Chem. Chem. Phys.* 15:12451-12458.
52. Coenen MJJ, *et al.* (2011) Little exchange at the liquid/solid interface: defect-mediated equilibration of physisorbed porphyrin monolayers. *Chem. Commun.* 47:9666-9668.
53. Goerigk L, Kruse H, Grimme S (2011) Benchmarking Density Functional Methods against the S66 and S66x8 Datasets for Non-Covalent Interactions. *ChemPhysChem* 12:3421-3433.

54. Perdew JP, Burke W, Ernzerhof M (1996) Generalized gradient approximation made simple. *Phys. Rev. Lett.* 77:3865-3868.
55. Perdew JP, Wang Y (1992) Accurate and simple analytic representation of the electron-gas correlation energy. *Phys. Rev. B* 45:13244-13249.
56. Kresse G, Hafner J (1994) Norm-conserving and ultrasoft pseudopotentials for first-row and transition elements *J. Phys. Condens. Matter* 6:8245-8257.
57. Kresse G, Joubert D (1999) From ultrasoft pseudopotentials to the projector augmented-wave method. *Phys. Rev. B* 59:1758.
58. Becke AD (1993) Density-functional thermochemistry. III. The role of exact exchange. *J. Chem. Phys.* 98:5648-5652.
59. Hehre WJ, Ditchfield R, Pople JA (1972) Self-consistent molecular orbital methods. XII. Further extensions of gaussian-type basis sets for use in molecular orbital studies of organic molecules *J. Chem. Phys.* 56:2257-2261.
60. Weiner SJ, *et al.* (1984) A New Force Field for Molecular Mechanical Simulation of Nucleic Acids and Proteins. *J. Am. Chem. Soc.* 106:765-784.
61. Battezzati L, Pisani C, Ricca F (1975) Equilibrium conformation and surface motion of hydrocarbon molecules physisorbed on graphite. *J. Chem. Soc., Faraday Trans. 2* 71:1629-1639.
62. Yan J, *et al.* (2014) Controlling the stereochemistry and regularity of butanethiol self-assembled monolayers on Au(111) *J. Amer. Chem. Soc.* 136:17087-17094.
63. Ouyang R, *et al.* (2015) Intermixed adatom and surface bound adsorbates in regular self-assembled monolayers of racemic 2-butanethiol on Au(111) *ChemPhysChem* 16:928-932.
64. Tersoff J, Hamann DR (1983) Theory and application for the scanning tunneling microscope *Phys. Rev. Lett.* 50:1998-2001.
65. Ilan B, Florio GM, Hybertsen MS, Berne BJ, Flynn GW (2008) Scanning Tunneling Microscopy Images of Alkane Derivatives on Graphite: Role of Electronic Effects. *Nano Lett.* 8:3160-3165.
66. Yang T, Berber S, Liu J-F, Miller GP, Tománek D (2008) Self-assembly of long chain alkanes and their derivatives on graphite. *J. Chem. Phys.* 128:124709.
67. Hentschke R, Schuermann BL, Rabe JP (1992) Molecular-dynamics simulations of ordered alkane chains physisorbed on graphite. *J. Chem. Phys.* 96:6213-6221.
68. Yin S, Wang C, Qiu X, Xu B, Bai C (2001) Theoretical study of the effects of intermolecular interactions in self-assembled long-chain alkanes adsorbed on graphite surface. *Surf. Interface Anal.* 32:248-252.
69. Cafe PF, *et al.* (2007) Chemisorbed and Physisorbed Structures for 1,10-Phenanthroline and Dipyrido[3,2-a:2',3'-c]phenazine on Au(111). *J. Phys. Chem. C* 111:17285-17296.
70. Bilic A, Reimers JR, Hoft RC, Ford MJ (2006) A priori modelling of the adsorption of benzene on surfaces of Group IB metals. *J. Chem. Theor. Comput.* 2:1093-1105.
71. Crossley MJ, Thordarson P, Bannerman JP, Maynard PJ (1998) A convenient procedure for moderate-scale Rothmund synthesis of lipophilic porphyrins: an alternative to the Adler-Longo and Lindsey methodologies. *J. Porphyrins Phthalocyanines* 2:511-516.
72. Dapprich S, Komaromi I, S. BK, Morokuma K, Frisch MJ (1999) A new ONIOM implementation in Gaussian98. Part I. The calculation of energies, gradients, vibrational frequencies and electric field derivatives. *J. Mol. Struct. (THEOCHEM)* 461-462:1-21.
73. Ochterski JW (Thermochemistry in GAUSSIAN. (GAUSSIAN Inc.).
74. Kresse G, Hafner J (1993) Ab initio molecular dynamics for liquid metals. *Phys. Rev. B* 47:558-561.
75. Kresse G, Furthmüller J (1996) Efficiency of ab-initio total energy calculations for metals and semiconductors using a plane-wave basis set. *Comput. Mat. Sci.* 6:15-50.
76. Soler JM, *et al.* (2002) The SIESTA method for ab initio order-N materials simulation. *J. Phys.: Condens. Matter* 14:2745.
77. Marenich AV, Cramer CJ, Truhlar DG (2009) 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* 113:6378-6396.
78. Ishikawa A, Nakai H (2015) Quantum chemical approach for condensed-phase thermochemistry (II): Applications to formation and combustion reactions of liquid organic molecules. *Chemical Physics Letters* 624:6-11.

Fig. 1. The meso-tetraalkylporphyrins considered.

Fig. 2. Observed STM images for tetraalkylporphyrin polymorphs overlaid by optimized (free-base) atomic structures: (a) Cu-C₁₁P L polymorph, (b) Cu-C₁₁P L-M interfaces, (c) Cu-C₁₁P M polymorph, (d) C₁₂P M polymorph, (e) C₁₃P L polymorph, and (f) C₁₃P L-M interfaces. The specified observed and calculated lattice parameters are given in Table 1; the white arrows indicate the direction of the underlying $\langle 11\bar{2}0 \rangle$ lattice vectors. For the interfaces, L and M indicate rows of porphyrins in (possibly distorted) polymorph lattices whilst the I rows indicate different types of interface structures containing mixed L and M features. The raw STM images utilized in (a)-(c) are reproduced from *Physical Chemistry Chemical Physics* (51) with permission of the *PCCP Owner Societies*, whilst that for (d) is reproduced from the *Journal of the American Chemical Society* (48) with permission of the *American Chemical Society* Copyright 2006.

Fig. 3. Two views each of the QM/MM optimized atomic SAM structures with lattice parameters and predicted STM images in good agreement with observed data for the M polymorphs of C₁₁P, C₁₂P, and C₁₃P highlighting (top) the two-dimensional surface plan and (bottom) solvent-sized cavities formed amidst the alkyl chains oriented into solution. One molecule per image is highlighted in green.

Fig. 4. Calculated contributions to the Gibbs free-energy of formation ΔG per mole of porphyrin of the L and M polymorphs of the SAM of C₁₃P on HOPG in solution, obtained using a continuum solvent model representing 1-phenyloctane and 1-octanoic acid. Colour code: grey- crudely estimated observed free energies, red- PBE-D3/PAW, blue- QM/MM (see SI Sect.S6 for PBE-DRSLL/DZP and PW91-D3/USPP). Entropy effects dominate the summed QM/MM zero-point energy, enthalpy, and entropy “thermal” contributions (these are evaluated only by the QM/MM method and the results universally applied). Solvation energy differences are evaluated for each method but the resulting differences are on average 0.1 ± 0.3 kcal mol⁻¹ and hence too small to be apparent in the figure. The zero in energy is taken to be that of the unsolvated surface, molecule, and non-interacting SAM.

