UNIVERSITÉ DE STRASBOURG

ÉCOLE DOCTORALE DES SCIENCES CHIMIQUES

Institut de Science et d'Ingénierie Supramoléculaires (ISIS)

THÈSE présentée par :

Simone CONTI

soutenue le : **30 septembre 2016**

pour obtenir le grade de : **Docteur de l'Université de Strasbourg**

Discipline/Spécialité : Chimie

Études computationnelles de l'Auto-Assemblage Moléculaire au Niveau des Surfaces: de la Conception Rationnelle à la Fonction

THÈSE dirigée par :

M. CECCHINI Marco	Maître de Conférences, Université de Strasbourg
M. SAMORì Paolo	Professeur, Université de Strasbourg

RAPPORTEURS :

M. TROISI Alessandro	Professeur, University of Warwick
M. BELJONNE David	Professeur, Université de Mons

AUTRES MEMBRES DU JURY :

M. LACKINGER Markus	Professeur, Technische Universität München

COMPUTATIONAL STUDIES OF MOLECULAR SELF-ASSEMBLY AT SURFACES: FROM RATIONAL DESIGN TO FUNCTION

SIMONE CONTI

Institut de Science et d'Ingénierie Supramoléculaires
École doctorale des Sciences Chimiques
Université de Strasbourg

Thesis Director: Dr. Marco Cecchini
Thesis Director: Prof. Paolo Samorì
30th September 2016

Dedicated to my parents
Franco and Rosa

ACKNOWLEDGEMENTS

First and foremost, I would like to thank Marco Cecchini, my supervisor and mentor during these last three years. Only thanks to him this doctorate became a very productive and stimulating experience. This would not have been possible without the support of the whole IFM group. Even if working on very different research fields, all discussions and experiences with past and present members of the group have been of extraordinary importance to make these years enjoyable and unforgettable.

For a computational and theoretical work as mine, it is of uttermost importance to have a continuous experimental feedback to benchmark all predictions and conclusions. For this, I must thanks Paolo Samorì and the Nanochemistry Laboratory, in particular Artur Ciesielski who introduced me to all the experimental aspects of self-assembly, STM, and graphite exfoliation.

I have also to thank the Initiative d'Excellence (IdEx, grant Nr. 16141), the University of Strasbourg, and the Institut de Science et d'Ingénierie Supramoléculaires (ISIS) who gave me financial support and stimulating environments.

Last, I would like to thank my father, my mother, my brothers and the whole my family, who never stopped to support me in all my decisions, even if from far away from my home.

PUBLICATIONS

Some ideas and figures appeared previously in the following publications:

S. Conti and M. Cecchini. "Accurate and efficient calculation of the desorption energy of small molecules from graphene". *The Journal of Physical Chemistry C*, 119(4), pp. 1867–1879, **2015**.

S. Haar, A. Ciesielski, J. Clough, H. Yang, R. Mazzaro, F. Richard, S. Conti, N. Merstorf, M. Cecchini, V. Morandi, C. Casiraghi, and P. Samorì. "A supramolecular strategy to leverage the liquid-phase exfoliation of graphene in presence of surfactants: unraveling the role of the length of fatty acids". *Small*, 11, pp. 1691–1702, **2015**.

S. Bonacchi, M. El Garah, A. Ciesielski, M. Herder, S. Conti, M. Cecchini, S. Hecht, and P. Samorì. "Surface-induced selection during in situ photoswitching at the solid/liquid interface". *Angewandte Chemie International Edition*, 54, pp. 4865–4869, **2015**.

S. Conti, M. G. del Rosso, A. Ciesielski, J. Weippert, A. Böttcher, Y. Shin, G. Melinte, O. Ersen, C. Casiraghi, X. Feng, K. Müllen, M. M. Kappes, P. Samorì, and M. Cecchini. "Perchlorination of coronene enhances its propensity to self-assembly on graphene". *ChemPhysChem*, 17, pp. 352–357, **2016**.

S. Conti and M. Cecchini. "Predicting molecular self-assembly at surfaces: a statistical thermodynamics and modeling approach". *Submitted*.

RÉSUMÉ DE THÈSE

Études computationnelles de l'Auto-Assemblage Moléculaire au Niveau des Surfaces : de la Conception Rationnelle à la Fonction

Certains des phénomènes les plus fascinants observés dans la Nature sont le résultat de l'auto-organisation, lorsque des interactions locales entre blocs élémentaires provoquent l'apparition d'ordre macroscopique. On peut citer comme exemples le repliement des protéines, la formation de cristaux moléculaires, les réactions d'oscillation ou encore les organismes biologiques. Alors que les deux premiers exemples sont dominés par un gradient d'énergie libre, dans le cas des deux suivants, un apport continue d'énergie ou de matière est nécessaire. Ainsi, il est possible de classer les deux premiers exemples comme le résultat d'un processus d'auto-assemblage, tandis que les deux seconds sont les résultats d'une auto-organisation. Au niveau moléculaire, il existe deux cas limites de l'auto-assemblage. Dans le premier scénario, les monomères s'auto-organisent en objets de taille finie, comme des polyèdres supramoléculaires ou des micelles. Dans le second scénario, les monomères s'associent dans des structures périodiques infinies avec différentes dimensionnalités : polymères supramoléculaires linéaires (1D), monocouches auto-assemblées (2D), ou cristaux (3D) ; voir Fig. 1.

Le sujet de ma Thèse est l'étude de l'auto-assemblage moléculaire au niveau des surfaces et des interfaces. Ce processus consiste en l'association spontanée de monomères diffusant librement en solution et formant une architecture cristalline bidimensionnelle physisorbée à l'interface solide-liquide (monocouche auto-assemblée or *Self-Assembled Monolayer*, SAM). Mon travail a été dédié à la compréhension fondamentale des facteurs qui contrôlent la probabilité d'auto-assemblage à l'équilibre chimique. Ma thèse est divisée en trois parties. Dans la première partie, je présente le cadre théorique que j'ai développé pour l'étude de l'auto-assemblage moléculaire en 2D. Cette approche est basée sur la détermination numérique des potentiels chimiques des monomères et de la SAM en utilisant les relations données par la thermodynamique statistique. De plus, dans cette partie, je discute les approches déjà publiées dans la littérature pour étudier l'auto-assemblage moléculaire au niveau des surfaces. Dans la deuxième partie, quatre applications sont présentées. D'abord la façon de modéliser la chimie sur le graphène sera discutée, en portant une attention particulière à l'estimation précise et rapide de l'énergie d'adsorption des blocs de construction moléculaires sur graphite/graphène. Ensuite, trois exemples qui impliquent l'auto-assemblage seront présentés : 1) l'emploi des acides gras utilisés comme surfactants pour l'exfoliation du graphite en phase liquide ; 2) l'auto-assemblage de trois isomères d'un diaryléthène (DAE) photoswitcher ; et 3) le perchlorocoronène utilisé comme une molécule conçue pour avoir une meilleure propension à l'auto-assemblage que le coronène. La troisième et dernière partie décrit les efforts mis en œuvre pour aller au-delà des approximations introduites précédemment. Ceci implique principalement une étude approfondie de l'adsorption moléculaire à l'interface solide-liquide. Le manuscrit de thèse se termine par un dernier chapitre de conclusions et perspectives.

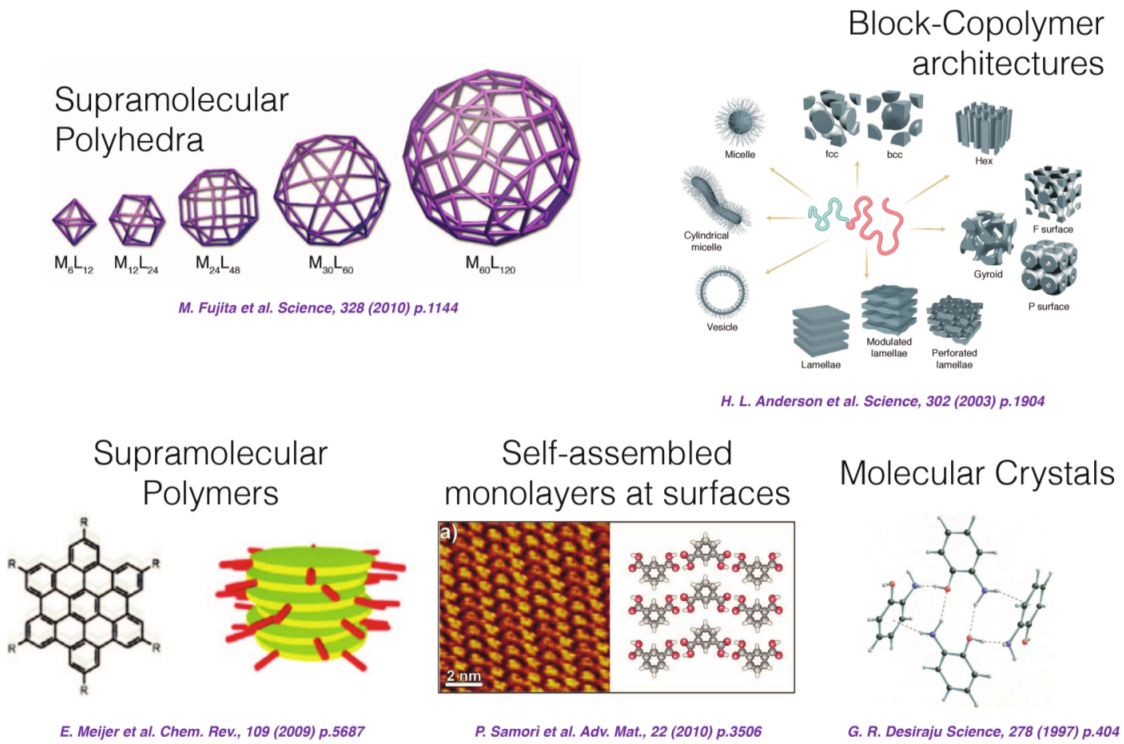


FIGURE 1 – Exemples d'auto-assemblage moléculaire de taille fini et infini.

Développement Théorique

L'objectif principal de mon travail est l'étude de l'équilibre chimique entre les états monomère et auto-assemblé. Selon les lois de la thermodynamique, cet équilibre est dicté par la différence de potentiels chimiques standards $\Delta\mu^{\ominus}$ de la réaction d'auto-assemblage. En utilisant une approche de thermodynamique statistique, en l'occurrence dans les limites des approximations du gaz idéal, de l'oscillateur harmonique et du rotateur rigide, le $\Delta\mu^{\ominus}$ peut être directement calculé à partir des propriétés moléculaires des états finaux. Ces approximations importantes simplifient grandement le calcul des potentiels chimiques, en fournissant un accord qualitatif avec les expériences. Lorsque l'auto-assemblage de taille finie est d'intérêt, toutes les propriétés moléculaires nécessaires pour évaluer le $\Delta\mu^{\ominus}$ peuvent être facilement obtenues en utilisant un modèle atomistique de l'interaction, par exemple un champ de force classique. Ceci permet

notamment d'étudier la variation de la probabilité d'auto-assemblage en fonction de la concentration initiale de monomères et de déterminer la concentration initiale de monomères correspondant à l'équiprobabilité des états auto-assemblées et désassemblés à l'équilibre. Cependant, lorsque l'auto-assemblage mène à une architecture infinie, l'approche théorique doit être différente puisque les propriétés moléculaires extensives (comme le nombre de molécules impliquées) de la SAM infinie sont mal définies.

Pour surmonter ce problème, j'ai trouvé que, dans la limite d'une architecture infiniment grande, le potentiel chimique de l'état auto-assemblée 2D peut être exprimé comme une fonction linéaire du potentiel chimique de sa maille élémentaire (μ'_{uc}). Ainsi, si la réaction d'auto-assemblage est décrite par

$$\alpha A \rightleftharpoons B \tag{1}$$

avec A le monomère, B la SAM, et α le nombre de molécules qui forment la couche auto-assemblée, le potentiel chimique de la SAM peut être écrit

$$\mu_B = \alpha \mu'_{uc} \tag{2}$$

Un des résultats théoriques majeurs de mon travail de thèse a été de montrer que ce potentiel chimique est une observable thermodynamique de la SAM et qu'il est accessible par modélisation. Ce résultat permet d'écrire la différence de potentiel chimique pour la réaction d'auto-assemblage $\Delta\mu$ d'une manière simplifiée

$$\Delta\mu = \alpha \Delta\mu_{AB} = \alpha \left(\mu'_{uc} - \mu_A\right) \tag{3}$$

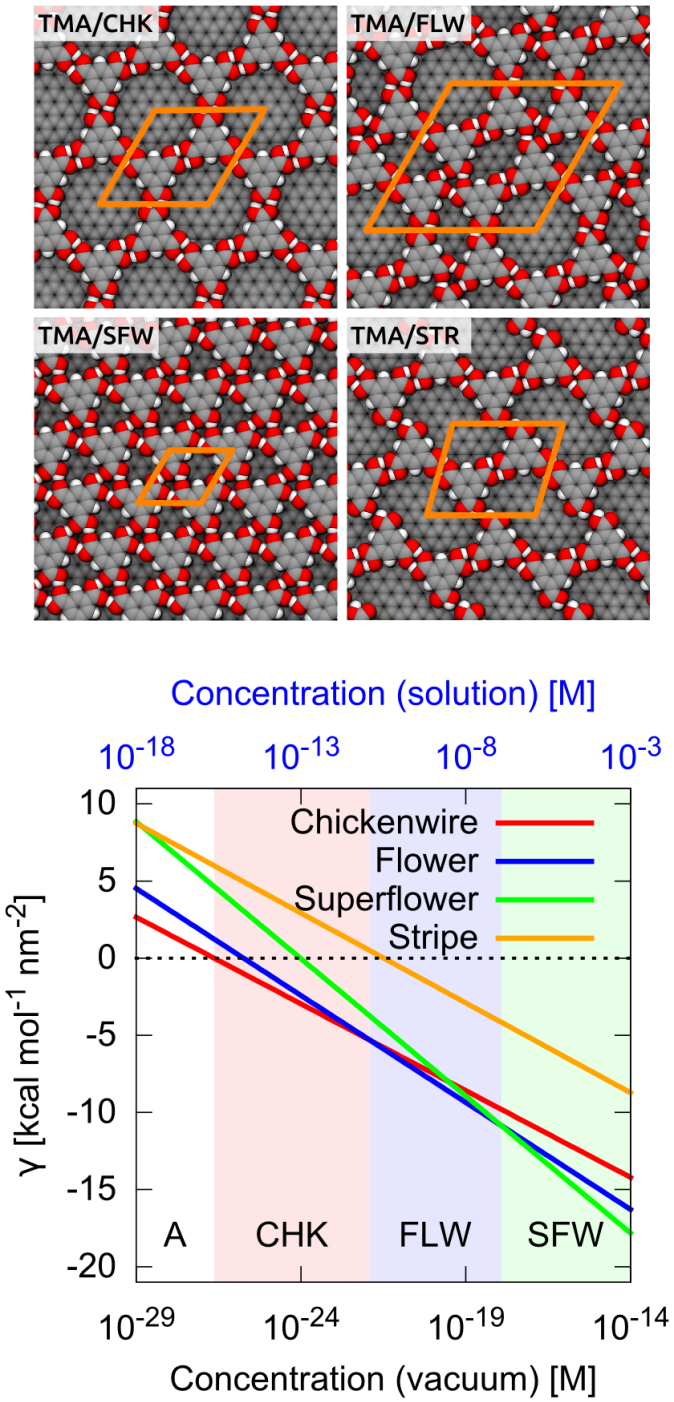


FIGURE 2 – Énergie libre de surface en fonction de la concentration de monomères pour l'auto-assemblage de l'acide trimésique (TMA) dans les trois architectures observées expérimentalement - chickenwire (CHK), fleur (FLW), et superfleur (SFW) - et l'architecture hypothétique stripe (STR). La dépendance à la concentration est évidente, montrant comment, à faible concentration, l'état monomère A est favorisé, alors que à concentration plus élevée les trois auto-assemblages CHK, FLW, et SFW sont visibles dans cet ordre. La valeur de γ pour l'architecture STR est, à toutes les concentrations, plus élevée que celle des autres. Cela signifie que cet architecture n'est jamais visible à l'équilibre thermodynamique.

où la différence totale de potentiel chimique est proportionnelle à la taille de l'architecture (α) et à le coût/gain pour déplacer une molécule à partir de la solution sur l'architecture auto-assemblée. Toutefois, cette quantité dépend toujours de manière linéaire du nombre de molécules, quantité que n'est pas définie pour une architecture infinie. Ce problème est résolu en exprimant α comme le rapport entre la surface totale couverte par la SAM (A_{sam}) et la zone couverte par une cellule unitaire (A'_{uc}). Ceci fournit une expression utile de l'énergie libre de surface de la SAM γ

$$\gamma = \frac{\Delta\mu}{A_{sam}} = \frac{1}{A'_{uc}} \left(\mu'_{uc} - \mu_A\right) \tag{4}$$

qui correspond au travail réversible par unité de surface pour recouvrir le substrat par un auto-assemblage moléculaire à partir d'une solution de concentration en monomères donnée. En outre, Eq. 4 implique que si, à une concentration donnée, γ est positif, l'état monomère est thermodynamiquement favorisé et l'auto-assemblage ne se produira pas. En revanche, si γ est négatif, la SAM sera l'état favorisée. De même, si deux ou plusieurs SAMs peuvent être produites par les mêmes blocs de construction, celle avec l'énergie libre de surface la plus faible est celle qui sera observée. En introduisant la notion d'état standard, l'énergie libre de surface peut être exprimée comme

$$\gamma = \gamma^{\ominus} - \frac{1}{A'_{uc}} kT \ln \frac{C_A}{C^{\ominus}} \tag{5}$$

où $\gamma^{\ominus}$ est l'énergie libre de surface calculée à une concentration standard $C^{\ominus}$ donnée, et C_A est la concentration du monomère. Par ailleurs, l'équation 5 permet d'étudier le polymorphisme sur la surface en fonction de la concentration en monomères dans la solution. En effet, en calculant γ pour chaque SAM en fonction de la concentration, les domaines de dominance des différents polymorphes peuvent être obtenus ; voir Fig. 2.

D'autre part, γ dépend naturellement de la densité de l'architecture et montre comment celle-ci peut influencer la probabilité d'auto-assemblage. Enfin, de simples manipulations de l'équation 5 donnent une définition de la concentration critique d'agrégation (C_{cac}), qui correspond à la plus petite concentration à laquelle la SAM est l'état favorisé :

$$C_{cac} = C^{\ominus} \exp\left(\frac{\gamma^{\ominus} A'_{uc}}{kT}\right) \tag{6}$$

Étant donné que plus la concentration critique est basse, plus la SAM est thermodynamiquement stable, la valeur de la C_{cac} définit une échelle absolue de la propension d'auto-assemblage 2D. Parallèlement, les concentrations de commutation entre polymorphes 2D d'une même molécule peuvent aussi être considérées comme des concentrations critiques caractéristiques. Ces concentrations critiques sont d'un grand intérêt pour l'étude de l'auto-assemblage au niveau des surfaces car elles peuvent être utilisées pour comparer des SAM chimiquement différentes et elles permettent de relier directement les expériences et les calculs.

Applications

Bien qu'une approche théorique solide soit de la plus grande importance, les prédictions ainsi faites peuvent être inexactes si elles sont basées sur une modélisation simplifiée des interactions. Ceci est particulièrement vrai dans l'évaluation de l'énergie par maille élémentaire E'_{uc}, laquelle peut être décomposée en une contribution de l'interaction de la SAM et de la surface sous-jacente, et une contribution des interactions molécule-molécule dans la SAM. Pour faire face à ce problème, différents modèles de calcul, allant des champs de forces empiriques (*Force Field*, FF) à la théorie de la fonctionnelle de la densité (DFT) en passant par la mécanique quantique semi empirique (SQM), ont été validés sur leur capacité à reproduire de manière précise et efficace l'énergie

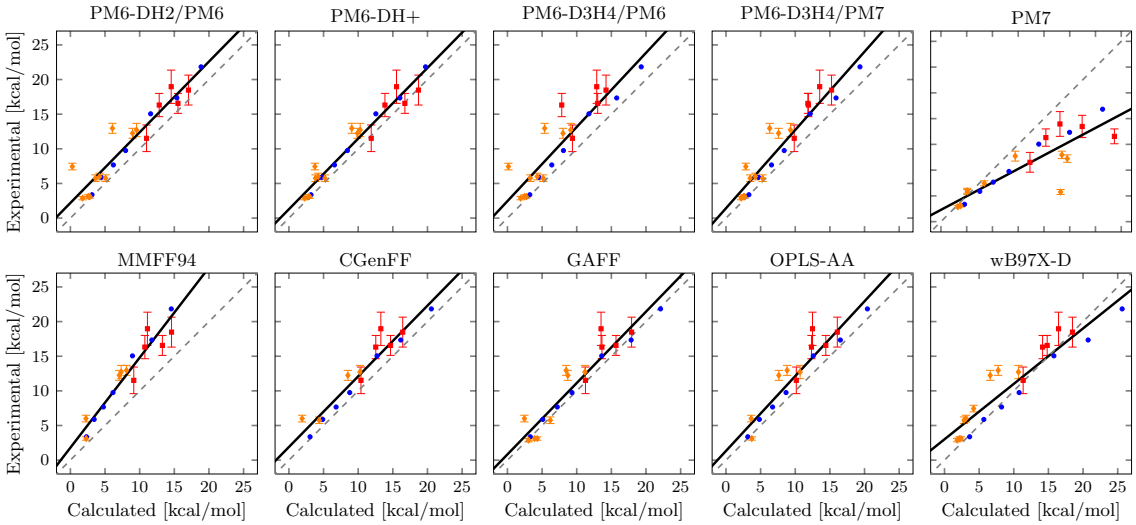


FIGURE 3 – Graphes de corrélation entre les énergies de désorption expérimentales et calculées pour cinq méthodes semi-empiriques (PM6-DH2/PM6, PM6-DH+, PM6-D3H4/PM6, PM6-D3H4/PM7 et PM7), quatre champs de force classiques (MMFF94, CGenFF, GAFF et OPLS-AA) et une méthode DFT (ωB97XD/6-31G(d)). L'ajustement idéal est représenté par la ligne en pointillé gris, ce qui correspond à une égalité parfaite entre les énergies de désorption calculées et expérimentales. L'ajustement obtenu est la ligne noire solide. Les barres d'erreur correspondent à l'incertitude dans la détermination expérimentale de l'énergie de désorption.

de désorption de petites molécules depuis le graphite. Cette dernière est mesurée par des expériences de désorption en température programmée (TPD) dans le vide ; voir Fig. 3. J'ai montré que l'évaluation de l'énergie de désorption est précise même si on utilise des méthodes computationnelles peu intensives telles que la SQM et les champs de forces empiriques avec un RMSE absolu de ~1 kcal mol^{-1} sur de petits fragments moléculaires. Ceci permet un criblage rapide pour trouver des molécules avec une énergie d'adsorption optimale ce qui a nous conduits à identifier la chloration comme une fonctionnalisation clé pour augmenter l'adsorption sur le graphite (voir ci-après). La précision de ces approches de calcul de l'interaction inter-moléculaire en deux dimensions est en cours d'évaluation.

Trois applications de l'approche théorique présentée ci-dessus sont exposées. Comme premier exemple, l'énergie libre de surface γ a été utilisée pour comprendre l'exfoliation en phase liquide (*liquid-phase exfoliation*, LPE) du graphite sous assistance molécu-

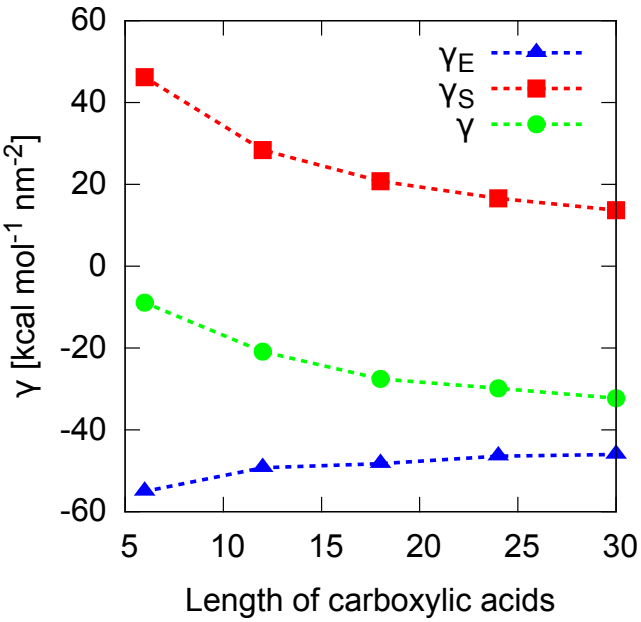


FIGURE 4 – Différences d'énergie (bleu), entropie (rouge) et d'énergie libre (vert) normalisées par la surface pour l'auto-assemblage des molécules étudiées en fonction de la longueur de la chaîne.

laire. Cette technique se démarque comme une approche polyvalente et potentiellement adaptable à la production de graphène de haute qualité en grande quantité. Récemment, il a été démontré que l'addition de petites molécules formant des monocouches auto-assemblées compactes sur le graphène exfolié peut augmenter le rendement de l'exfoliation. Sur cette base, cinq acides gras avec des chaînes alkyles de longueurs différentes, en partant de l'acide hexanoïque (C6) jusqu'à l'acide triacontanoïque (C30), ont été testés comme surfactants pour le graphène. En collaboration avec le laboratoire de Nanochimie à l'ISIS, il a été montré que le rendement de l'exfoliation peut être amélioré en augmentant la longueur de la chaîne alkylique. Étant donné que les plus courtes chaînes permettent une plus grande densité de liaisons hydrogène sur la surface (données bleues, Fig. 4), ce résultat contraste avec le "dogme" selon lequel plus fortes sont les interactions dans la SAM, meilleur est le rendement de l'exfoliation. En évaluant γ pour les SAMs générées par les cinq acides gras et en isolant la contribution entropique,

j'ai montré que l'auto-assemblage des molécules plus courtes est fortement défavorisé en raison du coût entropique associé au confinement sur la surface. En effet, plus courte est la molécule, plus élevé est le nombre de blocs de construction nécessaires pour couvrir une surface donnée (données rouges, Fig. 4). Cependant, c'est uniquement lorsque les deux contributions entropique et énergique sont correctement prises en compte, que la tendance expérimentale est reproduite (données vertes, Fig. 4).

La seconde application de γ est l'étude de l'auto-assemblage 2D d'un diaryléthène (DAE) photoswitcher sur du graphite. Dans ce travail, il a été montré par STM qu'il est possible de visualiser la commutation réversible in situ sur graphite entre SAMs formées par les isomères ouvert et fermé par irradiation de lumière visible et UV. Sous une irradiation de lumière UV-prolongée, un troisième photo-isomère irréversible a également été observé et considéré comme un sous-produit de la réaction. La formation de ce sous-produit favorise le passage à une troisième architecture 2D sur le graphite, ce même quand le sous-produit est présent en faible pourcentage en solution. En évaluant l'énergie libre de surface pour les trois architectures modèles, construites en se basant sur les images STM (Fig. 5), il a été constaté que le sous-produit d'auto-assemblage correspond à la plus forte augmentation de l'énergie libre par unité de surface due à une plus grande densité sur la surface et de plus fortes interactions entre les molécules dans la monocouche. Au contraire des acides carboxyliques, ici l'auto-assemblage est uniquement dominé par les contributions énergétiques, l'entropie joue un rôle négligeable. Ces résultats permettent de comprendre les observations expérimentales.

La troisième application consiste en une combinaison de la modélisation et de la thermodynamique statistique. Cette approche a été exploitée pour rationaliser l'effet de la fonctionnalisation chimique sur l'énergie libre d'auto-assemblage. Par l'étude d'un hydrocarbure polyaromatique prototypique, nous avons montré que par perchloration

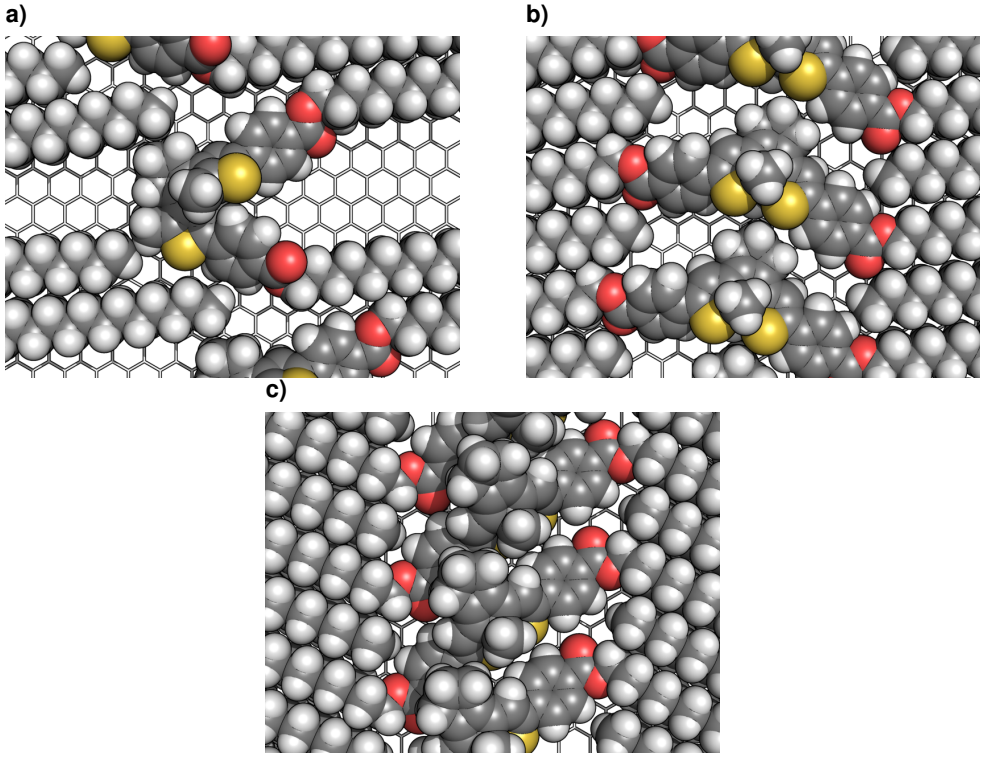


FIGURE 5 – Modèles de couches auto-assemblées générées par les formes a) ouvertes, b) fermées, et c) sous-produit du diaryléthène photoswitcher.

l'affinité de liaison pour le graphite est significativement améliorée. En comparant les barrières de désorption du coronène et du coronène perchloriné mesurées par désorption à température programmée (TPD, en collaboration avec M. M. Kappes au KIT), j'ai pu quantifier l'augmentation de l'énergie d'adsorption par molécule ; Fig. 6a. Ensuite, les SAMs formées par les deux molécules ont été modélisées (Fig. 6b) et l'énergie libre de surface a été évaluée. Les résultats montrent que la perchloration diminue l'énergie libre d'auto-assemblage non seulement energetiquement (en augmentant la force de liaison à la surface), mais aussi de manière entropique (en diminuant la concentration sur la surface). L'avantage fonctionnel de cet auto-assemblage modulé chimiquement a été démontré dans le cadre de la LPE du graphite sous assistance moléculaire (en collaboration avec le laboratoire de Nanochimie).

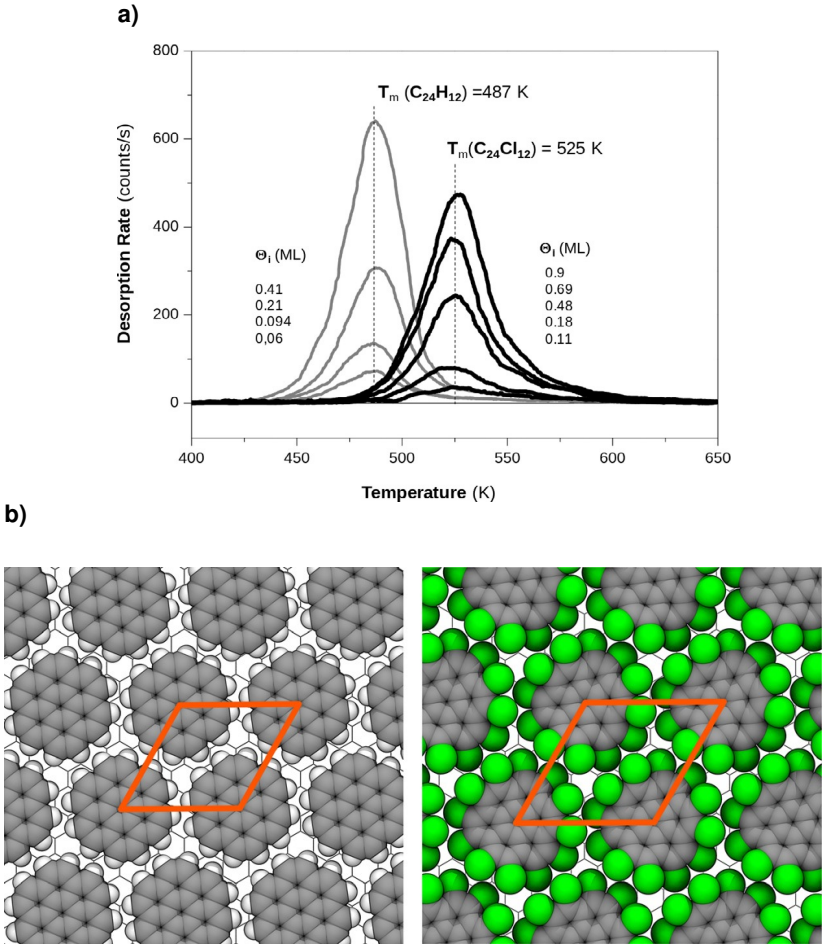


FIGURE 6 – a) Données obtenues par désorption à température programmée (TPD) pour le coronène et le coronène perchloriné. L'augmentation de la température de désorption après perchloration est clairement visible. b) monocouches auto-assemblées modélisées pour le coronène et le perchlorocoronene. Les deux s'assemblent en un réseau hexagonal, la seconde avec une maille élémentaire 30% plus grande que la première.

Au delà des approximations

La théorie développée s'est montrée utile pour comprendre et prédire l'auto-assemblage moléculaire au niveau des surfaces. Cependant, un certain nombre d'approximations ont été introduites, ce qui en limite le spectre d'application. L'une de ces approximations consiste à considérer l'auto-assemblage comme un processus à deux états : le monomère et l'architecture périodique 2D. Ceci est évidemment incorrect puisque lorsqu'une solution est mise en contact avec une surface, un certain nombre d'équilibres chimiques sont établis, comme l'adsorption du soluté sur la surface ou les étapes in-

termédiaires d'oligomérisation. Bien que cela ne réfute pas les résultats des équations simplifiées ci-dessus, puisque l'énergie libre est une fonction d'état, des caractéristiques intéressantes sont cachées ou perdues. Pour explorer ce monde caché, l'équilibre d'adsorption a été étudié de manière approfondie. Après une étude théorique de l'équilibre d'adsorption pour des molécules simples à l'état dilué (ni l'auto-assemblage ni l'agrégation ne sont considérés), j'ai dérivé l'expression suivante pour la probabilité d'adsorption en 2D

$$P_{2D} = \frac{K}{K + \eta/\eta^{\circ}} \tag{7}$$

où K est la constante d'équilibre pour le processus d'adsorption et le paramètre $\eta = V/S$ est le rapport d'extension entre le volume accessible et la surface accessible. L'Eq. 7 montre que la probabilité d'absorption ne dépend pas de la concentration initiale en solution, mais elle est uniquement fonction du rapport d'aspect ; c'est-à-dire le volume disponible sur la surface disponible. Ce résultat théorique a été confirmé par des calculs de dynamique moléculaire de l'adsorption sur du graphite pour cinq petites molécules, où le paramètre η a été systématiquement changé. Bien que la validation expérimentale soit en cours, ce résultat met en évidence le rapport volume/surface en tant que paramètre de contrôle nouveau et totalement inexploré de l'auto-assemblage moléculaire au niveau des surfaces.

En conclusion, j'ai développé un cadre théorique basé sur la modélisation et la thermodynamique statistique pour étudier l'auto-assemblage moléculaire au niveau des surfaces permettant une interprétation *ab initio* de l'équilibre thermodynamique. Bien que nous ayons montré que le cadre proposé est en mesure de rationaliser et reproduire les résultats expérimentaux dans trois cas d'études, des travaux supplémentaires sont

nécessaires pour améliorer la précision des prévisions. En particulier, nous anticipons une amélioration significative des résultats une fois introduits les effets du solvant, manquants dans le traitement actuel. En outre, l'analyse des états oligomériques intermédiaires pour comprendre la cinétique et les voies d'auto-assemblage est laissée à l'avenir. Grâce aux bases théoriques claires et aux approximations introduites de manière transparente, le cadre développé peut être utilisé comme une base pour des approches plus précises ou spécialisées où les approximations sont progressivement supprimées. Ceci est impossible par la théorie seule ; un développement expérimental parallèle est nécessaire pour permettre l'identification et l'évaluation de quantités thermodynamiques accessibles à la fois expérimentalement et par modélisation, comme la concentration critique d'agrégation C_{cac}, afin de faire le lien entre ces deux disciplines.

CONTENTS

1

INTRODUCTION

Most fascinating phenomena observed in Nature are the product of self-organization, where macroscopic order arises from the local interactions between smaller building blocks[1]. At the nanoscale, of wide interest is molecular self-assembly, not only because it is the basis of life[2], but also because it provides the path of lowest energy consumption to the fabrication of nano-objects with controlled morphologies and properties[3,4].

In self-assembly preexisting components form ordered structures through simple local interactions. Conditions to self-assembly are the ability of the components to diffuse one respect to the others, and establish reversible interactions which can be formed and broken. Most importantly, self-assembly is defined by the properties of the building blocks, which implies that the outcome can be in principle modulated by proper design of the molecular building blocks. With this definition, self-assembled processes can be divided in two classes. In the first, the system evolves to a free energy minimum and is stable at thermodynamic equilibrium. Examples are molecular crystals, protein folding, or the assembly of virus capsizes. In the second class, an influx of energy or material is constantly dissipated to maintain the order of the self-assembled structure. Examples are oscillating reactions, biological organism, or the Rayleigh-Bénard convection cells. A common convention[5] is to call the first class of processes as the outcome of *self-assembly*, while the seconds as the outcome of *self-organization*. Focusing on the first class, at the molecular level two limiting cases exist. In the first, the monomers self-assemble into finite-size objects, such as supramolecular polyhedra[6] or micellae[7]. In the second case, the monomers associate into infinite periodic structures with different dimensionality: linear supramolecular polymers (1D)[8], self-assembled monolayers (2D)[9], or crystals (3D)[10]; see Fig. 1.1. The focus of this thesis is molecular self-assembly at surfaces and interfaces, which is a prominent example of the 2D infinite case.

Molecular self-assembly at surfaces consists in the spontaneous association of freely diffusing monomers in solution into a self-assembled monolayer (SAM), which is an ordered 2D crystalline architecture physisorbed at the solid-liquid interface. What makes it particularly fascinating is the ability to watch live how the molecules organize on the surface at an atomic level of detail. This is possible thanks to scanning probe techniques, in particular the Scanning Tunneling Miscroscopy (STM). First introduced by Binnig and Rohrer[13,14], for which they were awarded the Nobel prize in physics in 1986, the STM exploits the tunneling electron current between the molecular orbitals of the substrate and of a metallic tip when a potential bias is applied between the two[15–17]. If the tip is moved across the surface and the intensity of the tunneling current is measured, it is possible to obtain an imaging of the orbitals of the substrate. Thanks to its sub-

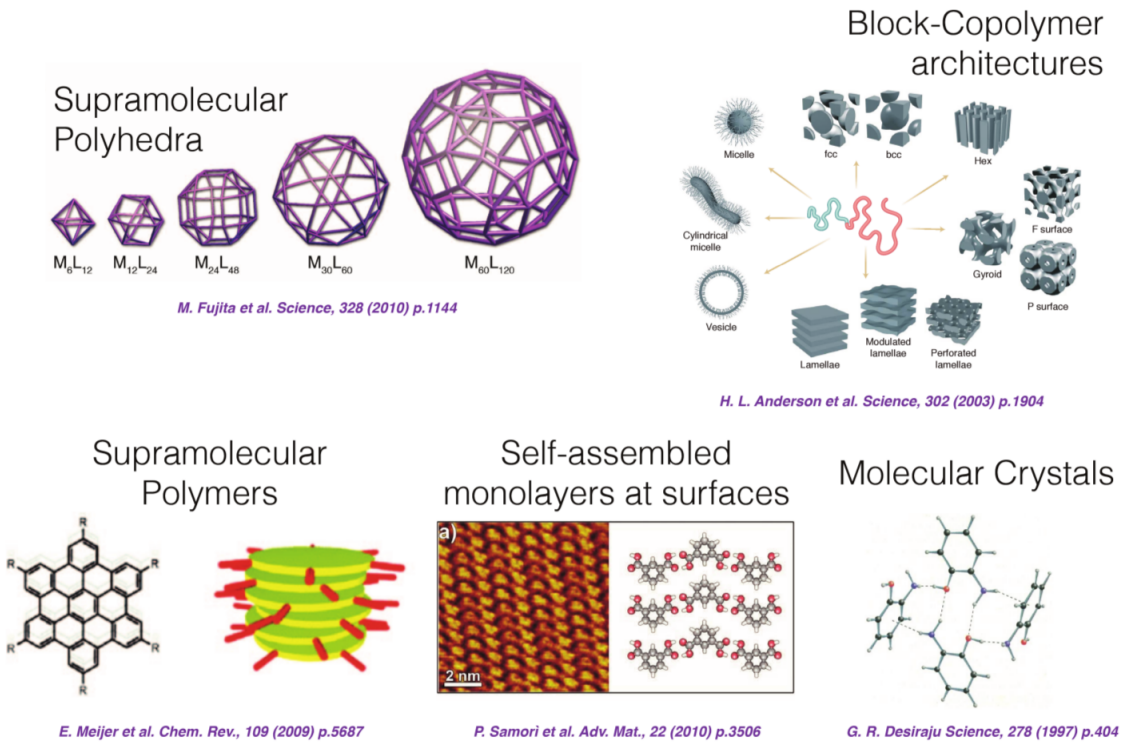


Figure 1.1: Examples of finite and infinite-size molecular self-assembly. Adapted with permission from ref. 6 (copyright 2010 American Association for the Advancement of Science), ref. 7 (copyright 2003 American Association for the Advancement of Science), ref. 11 (copyright 2009 American Chemical Society), ref. 12 (copyright 2004 American Chemical Society), and ref. 10 (copyright 1997 American Association for the Advancement of Science).

nanometer spatial resolution, the STM is *de facto* the gold standard to investigate physical and chemical processes involving molecules at surfaces[9,18–22].

Among its facets[4,23], molecular self-assembly at surfaces stands out as a prominent example for technological applications. Achieving control over the forces that steer 2D self-assembly by chemical functionalization of the building blocks is key to boost the development of bottom-up approaches in fields as diverse as the nanofabrication of electronic circuits[24], organic electronics[25], sensing[26,27], catalysis[28], energy storage[29] and conversion[30,31]. Despite this, only a few are the attempts to formally refine predictions of self-assembly in material science. In particular, there exists no general framework to assess the thermodynamic stability of the self-assembled monolayer (SAM), predict its polymorphism as a function of temperature[32], solvent[33], or concentration of monomers[34], and/or rationalize competitive self-assembly at surfaces and interfaces[35,36]. One major obstacle in this endeavor is that the thermodynamic equilibrium structure of a SAM is the result of a subtle interplay between the energy gained upon molecular adsorption and association and the entropy lost upon surface confinement, whose accurate evaluation is computationally very challenging. The primary intent of this dissertation is to address these problems. At first, the computational methods avail-

able to study the graphene chemistry will be presented, followed by an introduction to the concepts of chemical equilibrium and statistical thermodynamics. Then, the currently available theoretical approaches to model 2D self-assembly will be reviewed. Last, I will present the theoretical framework I developed to study 2D self-assembly, which focuses on the fundamental understanding of the factors that control the probability of self-assembly at chemical equilibrium.

The substrate used in all works reported in this thesis is Highly Oriented Pyrolytic Graphite (HOPG). This substrate is optimal for STM experiments, and is a perfect model to study molecular self-assembly on graphene, which recently emerged as an exciting new material holding potential impact in many areas of science and technology[37]. However, the rise of graphene strongly relies on the development of efficient strategies to produce quality material in high quantities[38]. This two-dimensional, atomically thick carbon-based material can be obtained using both bottom-up approaches[39–43], which allow the production of very high quality layers but in limited amounts, and top-down methods[44–52], with which it is possible to produce graphene in much more abundant quantities, with the flaw of possibly producing lower quality layers. Among these seconds, sonication-assisted liquid-phase exfoliation (LPE) of graphite stands out as a promising approach[52,53]. Graphite LPE generally involves three steps: the dispersion of graphite powder in solvents having a surface tension of 30–40 $mJ\,m^{-2}$ [53–56], the exfoliation of the 2D-layered material assisted by sonication, which causes the mechanical breaking of the dispersive interactions between graphene layers, and the purification of the exfoliated product. Better exfoliation yields can be obtained if small molecules are added to the solution, as they can physisorb on the exfoliated graphene layers, preventing their restacking[45,57–66]. An interesting direction in this field is the use as dispersion-stabilizing agents molecules that can form self-assembled monolayers on graphene[67]. This possibility will be explored in the second part of this manuscript which contains four technological applications involving the self-assembly on graphitic surfaces. The first will be focused on the design of optimal graphene binders. The second will exploit the propensity of fatty acids to form SAMs to leverage the graphite LPE. Third, the self-assembly of three isomers of a diarylethene (DAE) photoswitcher is studied focusing on the role of the substrate in the selection of the preferred isomer. Fourth, the design of perchlorocoronene is described, as a rationally designed molecule with chemically enhanced self-assembly propensity relative to coronene. In the last part of this this thesis, the efforts to go beyond the limits of the developed framework are reported, which mainly involve an in depth study of the molecular adsorption at the solid-liquid interface.

Part I

Theoretical Methods to Self-Assembly at Surfaces

2

COMPUTATIONAL METHODS OVERVIEW

The general problem tackled in this chapter is how to model a system which involves the adsorption and association of many small molecules on the top of a graphene substrate. The objective is twofold: obtaining the geometry of the studied complex together with its energy. This implies being able to first assign an energy to a molecular geometry, and then to optimize the geometry to minimize its energy. The obtained geometry would then be the optimal structure describing the complex in study. The general ability to assign an energy to a given geometry is of main interest, since it gives access to the most different information, like the chemical equilibrium for a chemical reaction (see next chapters), or the modeling of the time evolution of a complex molecular system.

The most important theoretical result in this respect is the Schrödinger equation, which, given the Hamiltonian operator $\hat{H}$, assigns an energy value E to the wave function Ψ of the system trough the solution of an eigenvalue problem

$$\hat{H}\Psi = E\Psi \tag{2.1}$$

Although "pure", this result can not be applied straightforwardly, and several approximations need to be introduced. In this line, the first step is the application of the Born-Oppenheimer approximation, that states the dynamics of nuclei is not usually entangled to the dynamics of the electrons, since the relaxation time of the latter is much faster than the movements of the formers. In this approximation, the wave function Ψ can be generally decomposed as product of a nuclei component $\Psi_N(R)$, which depends only on the position of the nuclei (R), and an electronic one $\Psi_e(R : r)$, which depends on the position of the electrons (r) and, parametrically, on the position of the nuclei

$$\Psi = \Psi_N(R)\Psi_e(R : r) \tag{2.2}$$

Similarly, the total Hamiltonian operator can be decomposed into a nuclear Hamiltonian $H_N(R)$ and an electronic one $H_e(R : r)$, the first including the kinetic energy of the protons and their pairwise coulomb interaction, the second the kinetic energy of the electrons and their pairwise electron-electron and electron-nuclei coulomb interactions. In this way, the initial Schrödinger equation can be decomposed into a nuclear and electronic problems, and the total energy of the system can be simply written as sum of a nuclear E_N and electronic E_e energy

$$E_{tot}(R, r) = E_N(R) + E_e(R : r) \tag{2.3}$$

Given a set of coordinates R for the atoms, the nuclear energy can be easily calculated as the coulomb interaction between the nuclei. For the electronic terms instead, a Schrödinger-like equation has to be solved, that only depends on the degrees of freedom of the electrons

$$\hat{H}_e(R)\Psi_e(R:r) = E_e\Psi_e(R:r) \tag{2.4}$$

Assuming that one is able to solve this last expression, given a set of atomic coordinates R the energy of the system can be calculated. Even, the initial coordinates can be systematically changed to build the full Potential Energy Surface (PES), or evolved in time to simulate the dynamics of the system.

In the next sections, different approaches are briefly reviewed, which allow to solve Eq. 2.4 and thus calculate the energy of the system. A more in depth discussion of most of the methods here presented can be find in introductory molecular modeling books[68–70].

2.1 WAVE FUNCTION THEORY (WFT)

The most rigorous methods to evaluate the electronic energy use directly the wave function Ψ. In fact, the energy associated to a wave function can be calculated directly as the expectation value of the electronic Hamiltonian operator $\hat{H}_e$

$$E_e = \frac{\langle\Psi|\hat{H}_e|\Psi\rangle}{\langle\Psi|\Psi\rangle} \tag{2.5}$$

For this reason, these methods are generally referred to as *ab initio* or *Wave Function Theory* (WFT) approaches.

The problem is how to write the wave function Ψ. First, it is necessary to remember that two identical electrons in the same system cannot occupy the same quantum state. This statement, known generally as the Pauli exclusion principle, is applied here in its most rigorous form as the wave function of the system needs to be antisymmetric, i.e. the wave function has to change sign when two electrons are exchanged. A general way to build an antisymmetric wave function is through the use of Slater determinants, where the wave function is written as combination of spin orbitals $\chi(r_i)$, each populated by one electron. The problem now is how to solve Eq. 2.4 to find the correct wave function of the system, which can then be used in Eq. 2.5 to obtain the electronic energy. The biggest problem is that the Hamilton operator contains, among others, one term which accounts for the potential energy between electrons, and thus depends on the relative positions of each pair of electrons. This implies that the Hamilton operator cannot be written as sum of one-electron operators, making the electronic Schrodinger equation unsolvable analytically.

To overcome this problem, one can make use of the variational principle, which states that for any trial wave function, the associated energy will be surely higher than the exact energy, and is equal to it only if the wave function is the exact ground state wave function of the system. Thus, it is possible to optimize variationally the wave function Ψ minimizing the electronic energy evaluated by Eq. 2.5. This is equivalent to solve the eigenvalue problem

$$f_i\chi(r_i) = \epsilon_i\chi(r_i) \tag{2.6}$$

where $\chi(r_i)$ and ϵ_i are the spin orbitals composing the total wave function and their energies, respectively, and

$$f(i) = \frac{1}{2}\nabla_i^2 - \sum_j^{nuclei} \frac{Z_j}{r_{iA}} + v^{HF}(i) \tag{2.7}$$

is the Fock operator for the ith electron, which depends on both its kinetic and potential energies, and on the average potential v_i^{HF} due to the presence of all other electrons. Interestingly, $f(i)$ is a one electron operator, which allows to calculate the spin orbitals of each electron independently and the total wave function as a product of all spin orbitals. However, $f(i)$ depends on v_i^{HF}, which is a global quantity that depends on the spin orbitals of all electrons. As such, the system composed by all f_i is not linear and has to be solved iteratively, which is usually referred to as the self-consistent field (SCF) method, that is at the heart of the Hartree-Fock (HF) approach.

A useful formulation of the HF method is the one proposed by Roothaan[71], which makes use of matrix operations instead of manipulating integrals on the wave function as was originally formulated by Hartree

$$FC = SC\epsilon \tag{2.8}$$

Here F is the Fock matrix, C are the expansion coefficient to build the wave function, S the overlap matrix, and ϵ the matrix containing the energies associated to each spin orbital. As said before, the Fock matrix depends on the state of all the electrons of the system, and so it depends on the coefficient matrix C. In this framework, the SCF procedure begins with a guess of the C matrix, it solvs the Hartree-Fock-Roothan equations, which produces a new C matrix and it iterates until convergence.

To increase the precision of the calculated energies in the Hartree-Fock-Roothan framework, one may increase the number of basis functions to define the molecular spin orbitals. For a system of N electrons, at least N basis functions are used, which generates $2N$ spin orbitals. If an arbitrary big number K of orbitals is used, the first N will be populated by the electrons, while the remaining $2K - N$ will be left empty and define virtual orbitals. Since the Hartree-Fock-Roothan equations access the electronic energy

variationally, the larger is the set of basis functions, the more freedom is left to the electrons to populate the optimal state, and the better the electronic energy will be. The main problem of this approach is how the correlation between electrons is modeled. In HF, in fact, all electrons move in the average field generated by all the electrons, which neglects the correct pairwise repulsion energy. Thus, even the use of an infinite basis set will not converge to the correct electronic energy of the system, but to a certain HF limit. The difference between the HF limit and the correct energy is defined as the electronic correlation energy.

The most straightforward approach to account for the electronic correlation energy is the Configuration Interaction (CI) method[72]. In HF, K basis functions are used to accommodate N electrons, and a Slater determinant to calculate the electronic energy is built using the N lowest populated orbitals. The Slater determinant formed by the lowest energy orbitals is just one possible way of populating the K orbitals with the N electrons. It is possible to classify all other possible electronic configurations by how much they differ from the HF determinant. If only one electron is moved from one occupied to one virtual orbital, single excited determinants are formed, if two, double excited, and so on. The "correct" wave function can thus be written as a linear combination of all these determinants. If all possible excited states are used, a full CI calculation is performed, which recovers all the correlation energy in the limit of the basis set in use. Although correct, this method is impractical even for small molecules due to huge number of determinants to be considered. To account for most of the correlation energy with the minimal computational cost, it is possible to truncate the CI expansion by considering only the single excited determinants, double excited (CISD), triply excited (CISDT), and so on. If a CISD (Configuration Interactions with Singles and Doubles) is used, 80-90% of the total correlation energy can be recovered. Although variationally correct, the CI method is not size consistent, i.e. the energy of two non interacting particle is not two times the energy of one particle.

One possible way to solve this problem is provided by the Many Body Perturbation Theory (MBPT), where the actual Hamiltonian is written as sum of a Hamiltonian for which the exact solution is known, along with a perturbation Hamiltonian, which are coupled using the parameter λ. Using this perturbation scheme the energy can be written as a Taylor expansion over the coupling parameter, which can be truncated at the zeroth, first, second, third order, or higher. If the unperturbed Hamiltonian is written as sum of Fock operators, the Moller-Plesset (MP) perturbation theory is obtained[73], which, although not variationally correct, is size extensive. In this way, if the expansion is truncated at the first order, the method MP1 is obtained, which is equivalent to HF. If higher orders are considered, more and more excited Slater determinants will be included similarly to the CI method, thus recovering more and more correlation energy. When stopped at the second order, the MP2 is obtained which is slightly more expensive than HF, but recovers 80-90% of the correlation energy. The MP4 method, which is

obtained by truncating the expansion at the fourth order, is similar in cost to CISD, but recovers 95-98% of the correlation energy.

Last popular method that aims at an accurate and efficient recovery of the correlation energy is the Coupled Cluster (CC) approach[74]. Here, the wave function is obtained by applying a "cluster" operator $\exp(\hat{T})$ to the HF Slater determinant. This operator is written in such a way that it generates all excited state determinants starting from the HF one. The cluster operator can thus be written as sum of different $\hat{T}_i$ operators, where each one generates all the i-th excited determinants: $\hat{T}_1$ generates all singly-excited determinants, $\hat{T}_2$ the doubly-excited, and so on. Again as for CI and MP, the sum can be truncated at any level, generating different methods at different levels of accuracy: truncating at $\hat{T}_2$ the CCSD method is obtained, which include single and double excited determinants, at $\hat{T}_3$ the CCSDT, at $\hat{T}_4$ CCSDTQ. The explicit inclusion of all these terms is computationally very expensive. To decrease the computational effort, the effect of the excited determinants can be added via a perturbative approach. Using this strategy, the popular CCSD(T) is obtained, where the contributions from triply excited states is calculated using MBPT. This method is commonly considered as the "Gold Standard" of quantum chemistry, and is used as reference for intermolecular interactions[75].

Note that although these methods theoretically converge to the correct answer, they are computationally very intensive, in particular when corrections above the double excited states are included. While an MP2 calculation is usually affordable for a molecular system in isolation, the cost of MP4 or CCSD(T) is so high only very small systems can be studied.

2.2 DENSITY FUNCTIONAL THEORY (DFT)

Although Hartree-Fock and corresponding wave function theory (WFT) methods derived from it can be used to solve the problem of calculating the electronic energy of a molecular or supramolecular system, they do not provide one that is at the same time accurate and efficient. A common problem of the *ab initio* approaches is that they all aim at the wave function Ψ of the system, which depends on the x, y, and z coordinates of all N electrons that belong to the system. This makes the WFT methods poorly scalable with the size of the system in study when accurate results are required.

A major result was the demonstration by Hohenberg and Kohn that the electronic energy of a system is uniquely defined by its electron density $\rho(x, y, z)$[76]. This quantity, which is defined as the integral over the position of all $N - 1$ electrons of the square of the wave function, depends only on the cartesian coordinates of the last electron; it essentially measures the probability of one electron to be in at a specific position in space. The Hohenberg-Kohn theorem states that there exists a functional $F[\rho(x, y, z)]$ that, when applied to the electron density, returns the electron energy of the system. In principle, this allows to reduce the complexity of the calculation from a problem depending on

3N coordinates (the full wave function Ψ) to a problem depending on 3 (the electron density ρ). However, the main problem is that no analytical expression of the functional $F[\rho]$ has been demonstrated and approximated expressions must be used to calculate the electronic energy of a molecular system.

Similar to what was done in HF, the functional $F[\rho]$ can be decomposed into a kinetic energy term $T[\rho]$ of the electrons, the nuclei-electrons interaction $E_{ne}[\rho]$, and the electron-electron interaction $E_{ee}[\rho]$. By separating this last term into a Coulomb $J[\rho]$ and an electron exchange $E_x[\rho]$ contribution, a general expression for the density functional is obtained

$$F[\rho] = T[\rho] + E_{ne}[\rho] + J[\rho] + E_x[\rho] \tag{2.9}$$

For both the nuclei-electrons ($E_{ne}[\rho]$) and the Coulomb ($J[\rho]$) terms analytical expressions exist and can be evaluated given the electron density. However, the kinetic ($T[\rho]$) and the exchange ($E_x[\rho]$) contributions are unknown. If a non-interacting uniform electron gas is assumed, these last two terms can be evaluated analytically, but the resulting model is so approximated that even bonding cannot be reproduced. The main problem is the kinetic energy contribution, which is poorly represented. If the electrons are modeled as non-interacting particles, as done in HF, the kinetic energy can be written as the sum of two terms: a first term (T_S) that includes most of the kinetic energy, and a second one that accounts for the correlation between electrons (E_c). The first contribution, T_S, can be actually calculated analytically from a Slater determinant of the molecular orbitals. The second is usually summed to the exchange term, yielding

$$F[\rho] = T_S[\rho] + E_{ne}[\rho] + J[\rho] + E_{xc}[\rho] \tag{2.10}$$

where only the exchange-correlation term (E_{xc}) is not known. To access the exchange-correlation energy it is necessary to reintroduce the molecular orbitals as the fundamental quantity, and the electronic density is calculated from them. Once again, the orbitals can be obtained by minimizing variationally the electronic energy, but, since $E_{ne}[\rho]$ and $J[\rho]$ depend on the orbitals, an iterative SCF procedure is required. This is the key idea underlying the Kohn-Sham equations[77].

Now the only problem left is how to evaluate the correlation-exchange energy accurately, which is not known analytically. To a first approximation, the electron density can be treated pointwise as an uniform electron gas and thus the functional only depends on the value of the electron density at the coordinates where it is calculated. In this approach named Local Density Approximation (LDA)[78], the exchange-correlation energy is local. To avoid assuming that the electron density is uniform in the whole space, information on the gradient can be added, yielding the Generalized Gradient Approximation (GGA) approach[79]. Since the gradient is included, the electron density is better described in the neighborhoods of the point where the functional is evaluated, giving a

better description of the system. A natural improvement are the meta-GGA functionals, which include not only the gradient (first derivative of the electron density) but also the laplacian (second derivative of ρ). A third family of methods, includes part of the exact exchange energy directly from HF, generating hybrid functionals[80], which are generally more accurate but also more expensive due to the HF step.

A tricky point on most DFT functionals is that the need for optimal exchange-correlation functional has led to the inclusion of empirical weights into the expression of the functional, which are usually obtained by fitting to theoretical or experimental data. The use of these parameters improves the accuracy of the results, at the cost of adding a paremetrization step.

One last but important note, all functionals described up to now are based on the local description of the system, and miss the non-local interactions, such as dispersion. In fact, the long range C_6/r^6 dependence of dispersive interactions is not correctly reproduced by any method presented above[81]. Theoretical attempts to include dispersion into DFT can be divided into three groups. A first attempt is to reparametrize existing functionals so that they can describe dispersion properly, including into the parametrization procedure dispersion-dominated complexes. The main pitfall is that if no special corrections are included into the form of the functional, these methods generally fail. Perhaps some appositely parametrized functional can give good results for small systems, probably due to error cancellation, or perform in a good way for a special class of interaction, e.g. they can predict gas dimer interactions, but non staking between aromatics rings. A second approach consists in adding to the SCF energy a non-local correlation energy trough the integration over a kernel function $\phi(r, r')$, which depends on the relative position of all pairs of electrons. Various methods have been developed following this approach, which differ in the form of the kernel function. This approach was first introduced in the vdW-DF[82] method, which was later updated (vdW-DF2[83]), and further developed and generalized (VV09[84], VV10[85]). These are theoretically founded approaches with little empiricism, e.g. the vdW-DF2 method used only one global empirical parameter, while the VV10 uses two. The third class of methods is based on the direct addition to the DFT energy of an empirical C_6/r^6 correction

$$E_{tot} = E_{DFT} + E_{dis} \tag{2.11}$$

$$E_{dis} = -s_6 \sum_{ij} f_{damp}(r_{ij}, R^0_{ij}) \frac{C_{6ij}}{r^6_{ij}} \tag{2.12}$$

where f_{damp} is a dumping function which brings the correction to zero at small distances R_{ij} since the DFT functional already describe well that area, and s_6 is an empirical parameter which is fitted based on the DFT functional used. Different methods can be derived from this formalism, each characterized by a different choice of the damping function, and how the C_{6ij}, s_6 and others parameters inside the damping function are

obtained. Among others[86,87], the corrections proposed by Grimme DFT-D1[88], DFT-D2[89], and the latest and most used DFT-D3[90] are popular. Although more heavily parametrized, these empirical approaches are much faster than the theoretical more justified non-local methods. When the accuracy of the two were compared, similar results were obtained[91,92].

2.3 BASIS SET SUPERIMPOSITION ERROR (BSSE)

Both *ab-initio* and DFT methods present quite serious problems when used to calculate intermolecular interaction energies, e.g. in a dimer. In both methods in fact, the accuracy and speed of the calculation depend on the size of the basis set, e.g. the number of basis function used to describe the wave function or the electron density of the system. The larger the size, the more accurate the predictions, but also the more expensive the calculations will be. Thus, quite often small and far from being complete basis sets are used to study larger, and interesting, systems.

Consider, for example, the calculation of the interaction energy for a dimerization reaction. Three energy calculations usually done are: one for the full dimer, and one for each of the two monomers. Looking at the calculation for the dimer, each of its monomers can use all the basis functions placed on each of their atoms in order to build the wave function, but for the atoms placed near the other monomer, i.e. the atoms involved in the interaction, they can see also the function placed on the other monomer and maybe use them to build a better wave function. This means that each monomer can effectively use a bigger basis set borrowing some functions from the other monomer. So if the interaction energy is calculated simply as the difference between the energy of the dimer minus the energy of the two separated monomers, an error occurs since the dimer is actually calculated with a bigger basis set then for the monomer. This is know as Basis Set Superimposition Error (BSSE). Of curse this problem disappears in the limit of an infinite basis set, but it is more and more pronounced when small and incomplete basis sets are used.

To fix this problem, a possible correction is the counterpoise method[93]. The idea is to calculate the energy of the monomers using their own basis functions, but in the presence of all the functions of the other monomer in conformation of the dimer, without filling them with electrons. In this way the monomer can effectively use all the "ghost function" of the other monomer as if it was in the dimer. Thus, performing a counterpoise-corrected DFT or WFT calculation, five different energies values are evaluated: the energy of the dimer with its basis set E_{AB}^{AB}, the energy of the first monomer with the dimer basis set E_{A}^{AB}, the energy of the second monomer with the dimer basis set E_{B}^{AB}, the energy of the first monomer with its basis set E_{A}^{A}, and the energy of the second

monomer with its basis set E_B^B. Then, the usual non corrected interaction is calculated as

$$E_{int} = E_{AB}^{AB} - E_A^A - E_B^B \tag{2.13}$$

while the basis set superimposition error is given by

$$BSSE = E_A^A - E_A^{AB} + E_B^B - E_B^{AB} \tag{2.14}$$

such that the corrected interaction energy is

$$E_{int}^{BSSE} = E_{int} + BSSE = E_{AB}^{AB} - E_A^{AB} - E_B^{AB} \tag{2.15}$$

This approach partially corrects for the basis set incompleteness, allowing to model large molecular systems with small basis sets.

2.4 SEMIEMPIRICAL QUANTUM MECHANICS (SQM)

Previously, it was shown that to increase the accuracy of HF, more accurate *ab initio* methods were developed by including more Slater determinants in the calculation. An alternative and computationally less intensive method goes in the opposite direction[94]. First, it was noted that among all electrons, only the valence electrons (the most external) are involved in bonding, while core electrons are generally less important. Second, to simplify the calculation, only a minimal basis set is considered, which provides the minimum number of orbitals to accommodate the valence electrons. Finally, the product between basis of the same electron on different atoms are neglected, as done also for one-electron/three-centers integrals, and all three and four centers integrals. To compensate for all these approximations, the remaining integrals are parametrized, generating approaches that are often referred to as Semiemprical Quantum Mechanics (SQM). The idea presented above forms the basis of the Zero Differential Overlap (ZDO) approximation, which is at the core of all SQM methods. Some variations include the Neglect of Diatomic Differential Overlap (NDDO)[95], the Intermediate Neglect of Diatomic Differential Overlap (INDO)[96], the Complete Neglect of Diatomic Differential Overlap (CNDO)[95], and the Modified Intermediate Neglect of Diatomic Differential Overlap (MINDO)[97] approximations. These approaches provide very efficient estimates of the electronic energy at the compromise that the accuracy of the results is subjected to the parametrization procedure.

A particularly successful approximation is the Modified NDDO (MNDO) family of methods[98], which includes the Austin Model 1 (AM1)[99] and the Parametric Method 3 (PM3)[100,101], which have been after re-parametrized into the Recife Method 1 (RM1)[102] and the PM6[103], to fix issues in the original methodology and include higher quality *ab*

initio and experimental data. In the last years, major concerns about SQM methods arose about their ability to reproduce hydrogen bonding and accounting for dispersive interactions. To solve this problem, empirical corrections were developed and introduced as *a posteriori* corrections, leading to the development of the AM1-D and PM3-D[104], PM6-DH[105], AM1-FS1[106], PM6-DH2[107], and the PM6-DH+[108]. All these methods share essentially the same dispersion correction scheme and the main development involved the hydrogen bonding correction. These corrections are based on geometrical criteria, such as distances and angles, and partial charges. The main problem of the DH and DH2 variants is that it is not possible to perform geometry optimization, since the expression of the gradient of the Hamiltonian used for the hydrogen bond correction present discontinuities[109]. The DH+ method prevents this problem, also introducing a better description of hydrogen transfer between groups. The most recent corrections are the D3H4[110] and the D3H+[111]. Both of them included the Grimme D3 dispersion correction[90] developed for DFT, but differ in the hydrogen bonding correction. The last SQM method released is the PM7[112], which results from the parametrization of a larger sets of data, and includes native corrections for both dispersion and hydrogen bonding interactions.

2.5 CLASSICAL FORCE FIELDS (FF)

The last family of methods to evaluate the energy of a system is the one formed by classical, or empirical, force fields. This is the family that includes the strongest approximations, effectively neglecting all electrons, and modeling the molecule as "beads" (the atoms) connected by "springs" (the bonds). The key point is that, at the end, all we want is a function that correlates the atomic coordinates with the energy of the system, conceptually equivalent to what Eq. 2.3 does by solving the time-independent Schrodinger equation. In this class of methods, the electrons are completely removed, and their average effect is replaced by an effective analytic function of the nuclei position. If this function is a good representation of the "correct" potential energy surface, the energy of a molecular system can be obtained from the position all the atoms. What is important to remember is that by doing so the electronic properties of the system can not be modeled and the accuracy of the force field totally depends on the reliability of the parametrization of the energy function.

In this framework, the energy of the molecule can be written as a sum of additive contributions which have, more or less, a physical interpretation. The first term is a contribution that keeps pairs of atoms connected by a bond, oscillating around an equilibrium distance. This is modeled by a harmonic potential, in analogy to the Hooke law. It follows that if a force constant k, which describes the strength of the covalent

bond, and an equilibrium distance R^{eq} are associated to each bond, the corresponding contribution to the energy of the system is

$$E_{bond} = \sum_{i}^{bonds} \frac{1}{2} k_i \left(R_i - R_i^{eq}\right)^2 \tag{2.16}$$

Obviously, this term alone is not sufficient to reproduce the geometry of a molecule. It is important, for instance, to constrain the angle between three atoms involved in consecutive bonds, e.g. enforcing the tetrahedral geometry of a sp^3 carbon atom. Also in this case, the simplest way to model this constrain is to introduce harmonic springs on the each covalent angle

$$E_{angle} = \sum_{i}^{angles} \frac{1}{2} k_i \left(\theta_i - \theta_i^{eq}\right)^2 \tag{2.17}$$

where θ is the angle between three atoms, θ^{eq} the reference equilibrium angle, and k the force constant. Using these two energy terms only, however, it is not possible to obtain a correct torsional profile for e.g. a butane molecule. To correct for this a dihedral energy term is introduced. Since this term must be periodic with the dihedral angle, a different expression is used

$$E_{dihedral} = \sum_{i}^{dihedrals} \sum_{n} \frac{1}{2} k_{i,n} \left[1 + \cos\left(n\phi + \phi_0\right)\right] \tag{2.18}$$

where a cosine series (generally up to order $n = 6$) is used for each dihedral. To correctly reproduce the planar configuration of an aromatic ring, additional corrections, commonly referred to as improper dihedrals, need to be introduced, with and energy definition similar to the proper dihedral.

These four energy contributions are called bonded terms, since they consider only atoms covalently bonded one to another. Non-bonded terms also exist, which account for the energetic interaction between atoms that are far in connectivity. The two most important contributions are the Coulomb term, to account for the charge-charge electrostatic interaction, and the van der Waals term, to account for dispersion and core-core repulsion. The former arises because atoms in a molecule are not neutral, but own a finite (partial) charge. To account for this a Coulomb energy term is added

$$E_{coulomb} = \sum_{i}^{atoms} \sum_{j}^{atoms} \frac{q_i q_j}{R_{ij}^2} \tag{2.19}$$

with q_i the partial charge on each atom and R_{ij} the distance between each pair of atoms. Finally, to avoid superimposition between colliding atoms and account for weak dispersive interaction a Lennard-Jones potential is included

$$E_{LJ} = \sum_{i}^{atoms} \sum_{j}^{atoms} \left(\frac{C_{12}}{R_{ij}^{12}} - \frac{C_6}{R_{ij}^6} \right) \tag{2.20}$$

Thus, the total potential energy per molecule is given by the sum of six terms

$$E = E_{bond} + E_{angle} + E_{dihedral} + E_{improper} + E_{coulomb} + E_{LJ} \tag{2.21}$$

If properly parametrized, this expression is able to reproduce the potential energy surface of both small molecules in isolation and large macromolecules, like proteins and DNA in solution.

The obvious question is how to assign all the necessary parameters to an arbitrary molecule. To simplify this task, it is possible to recognize that, for example, the bond distance between two carbon atoms in an alkane is almost always the same, as also the force constant to reproduce their vibrations. Thus, it is useful to assign to each atom in the molecule one type, and assign the parameters for bonds, angles, etc., to each atom type. The entire list of atom types and parameters for bonds, angles, dihedrals, van der Waals radii, etc., constitutes a classical force field. Depending of the number of atoms types and the strategies to obtain the empirical parameters, different force fields can be created. Historically, the first and most used force fields used are the ones to model proteins. Some examples are the OPLS[113], AMBER[114] and CHARMM[115] force fields.

In the last years, these methods has grown significantly both in accuracy and performances in particular for biological molecules. Unfortunately, they are not able to simulate a generic organic compounds, such as a drug that binds to a protein, or the small building blocks used in molecular self-assembly. To solve this problem, a great effort has been put forward in recent years to develop generalized molecular force fields, which are able to model almost any organic molecule. In this endeavor, OPLS was gradually extended into the OPLS-AA[116–122] force field that includes more chemotypes, and, similarly, AMBER and CHARMM were extended to the General Amber Force Field (GAFF)[123,124] and the CHARMM General Force Field (CGenFF)[125], respectively. Some other force fields have been specialized to deal with small molecules in great accuracy. These include the MMx family of methods (MM2[126], MM3[127], and MM4[128]) and the Merck Molecular Force Field (MMFF94)[129–133].

2.6 TAKE HOME MESSAGES

In previous sections a lot of different methods have been briefly reviewed. These are all summarized in Fig. 2.1, where the WFT methods are colored in yellow, DFT in green,

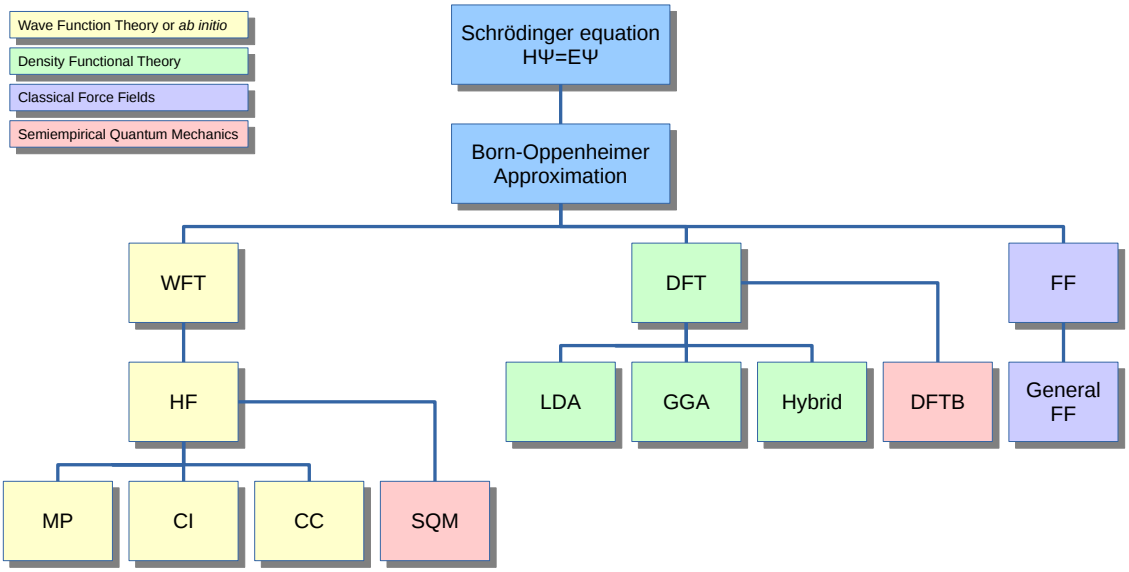


Figure 2.1: Overview of the computational methods briefly presented in this Chapter. The common root is the Schrödinger equation followed by the Born-Oppenheimer approximation. For this, three class of methods can be generated. First, the Wave Function Theory (WFT), which includes Hartree-Fock (HF), and the correlation-corrected Moller-Plesset (MP), Configuration Interaction (CI), and Coupled Cluster (CC). Second the Density Functional Theory (DFT) in its formulations: Local Density Approximation (LDA), Generalized Gradient Approximation (GGA), and Hybrid DFT. Third, the empirical Force Fields (FF), and their generalized extension. Last, the Semiempirical Quantum Mechanics (SQM) obtained from HF, and also from DFT (DFTB, not discussed in this work).

SQM in red and FF in blue. What is missing up to now is a comparison of all these methods in terms of both the applicability to different problems and the actual size of the system studied. In a very broad discussion, the WFT methods are considered to be the most accurate and correct. This is generally true only if the most elaborate methods with the biggest basis set are considered. This limits considerably the applicability of WFT method to very small systems up to tens of atoms. Goes without saying that this constrain excludes the possibility of using these methods to study the chemistry on graphene, as to model correctly the substrate, an extended graphene layer needs to be modeled. Valuable alternative are DFT methods. These scale much better with the system size – N^3 of DFT against N^7 of CCSD(T) – and usually they give accurate results. Still, DFT can be used up to a few hundreds atoms, so they are suitable for small molecule/graphene complexes, but it remains unpractical to model a self-assembled monolayer, as will be also discussed later in this thesis. A great unknown are the semiempirical methods. Only a few works are available in the literature, in particular when applied to material science[134]. In theory, these methods should be able to scale up to a few thousand of atoms, making them interesting to study self-assembly at surfaces. Another class of methods are the empirical force fields. These have the great advantage of being much less computational intensive, allowing the simulation of hundreds of thousands

of atoms. The only problem is their accuracy, as they have been originally developed mainly for protein simulation and their use in material sciences is still at the embryonic stage. For a more detailed overview, applications of all these computational methods have been recently reviewed [135].

3

CHEMICAL EQUILIBRIUM AND STATISTICAL THERMODYNAMICS

In the previous chapter some computational methods have been presented that can be used to model graphene. This chapter is instead used as introduction to the concepts and methodologies related to the study of chemical equilibrium and the use of statistical thermodynamics to evaluate the equilibrium constant for a chemical process. This is of main relevance to better follow the next two chapters dedicated to model the 2D self-assembly: the first one reviewing the literature approaches, and the second presenting the framework developed in this thesis.

3.1 FROM THE SECOND LAW TO CHEMICAL EQUILIBRIUM

Chemical Equilibrium

The aim of this paragraph is to derive the basic relations which define the study of chemical equilibrium. Fundamental point to start with is the second law of thermodynamics as formulated by Max Planck: *Every physical or chemical process in nature takes place in such a way as to increase the sum of the entropies of all the bodies taking any part in the process*[136]. If an isolated system undergoes any process, the total variation in entropy during the process (dS_{proc}) is the variation in entropy of the system (dS_{sys}), plus the variation in entropy of the environment that surrounds the system (dS_{env}), and this sum is said to be positive from the second law

$$dS_{proc} = dS_{env} + dS_{sys} > 0 \tag{3.1}$$

The only way a closed system has to affect the entropy of the environment is through a flux of heat dq across the system boundaries. Working at constant volume, the change in heat for the system corresponds to its internal energy variation (dU_{sys}), and so the change in heat of the environment is $-dU_{sys}$, since all heat lost/acquired from the system is acquired/lost by the environment. By definition, the entropy is the ratio between the exchanged heat and the temperature, which allows to write the difference in entropy for the environment as

$$dS_{env} = \frac{dq_{env}}{T} = \frac{-dU_{sys}}{T} \tag{3.2}$$

This allows to rewrite Eq. 3.1 as

$$dF_{sys} = dH_{sys} - TdS_{sys} < 0 \tag{3.3}$$

which recasts the second law of thermodynamics in terms of only properties of the system, and defines the difference in free energy dF as the thermodynamic quantity which describes the chemical equilibrium of closed systems at constant volume. Thus, a chemical process would move towards equilibrium decreasing the free energy of the system.

Consider now a generic chemical reaction where dN_i molecules of one or more reactants are converted in dN_j molecules of products. The process towards equilibrium can be conveniently expressed as a function of a generic extent of reaction ξ, such that the variation dN_i in the quantity of each specie can be written as

$$dN_i = \nu_i d\xi \tag{3.4}$$

where ν_i are the stoichiometric coefficients of the chemical reaction and are taken as positive for products and negative for reactants. The associated variation in free energy at constant volume and temperature is

$$dF = \sum_i \frac{\partial F}{\partial N_i} dN_i = \sum_i \mu_i dN_i = \sum_i \mu_i \nu_i d\xi \tag{3.5}$$

where μ_i is the chemical potential of the i-th specie and represents the free energy contribution brought by each molecule. As said before, the reaction will proceed until a decrease in free energy is possible, thus until the minimum of the free energy is reached. Since the process is described by the coordinate ξ, the minimum of the free energy is obtained when

$$\frac{\partial F}{\partial \xi} = 0 \tag{3.6}$$

such that the condition for chemical equilibrium is

$$\Delta\mu = \sum_i \mu_i \nu_i = 0 \tag{3.7}$$

which implies that the difference in chemical potentialbetween reactants and products is zero at chemical equilibrium. To better exploit this important result, assume that the chemical potential of a specie can be expressed as a function of its concentration as

$$\mu_i = \mu_i^{\ominus} + RT \ln\left(\frac{C_i}{C^{\ominus}}\right) \tag{3.8}$$

where the chemical potential is decomposed for convenience in a standard chemical potential $\mu_i^{\ominus}$ which is concentration or volume independent, and a concentration-dependent contribution. This allows to define an equilibrium constant K_{eq} for the reaction

$$K_{eq} = \exp\left(-\frac{\Delta\mu^{\ominus}}{RT}\right) = \prod_i \left(\frac{C_i}{C^{\ominus}}\right)^{\nu_i} \tag{3.9}$$

which relates the equilibrium concentrations of all reactive species. The $C^{\ominus}$ factor remembers that the standard chemical potentials, and thus the equilibrium constant, refer to a particular reference state and that the concentrations C_i are expressed in terms of $C^{\ominus}$.

In the following, the chemical equilibrium of a few prototypical reactions is studied, with the ultimate goal of showing how the equilibrium probabilities of the system components depend on the experimental conditions, like the initial concentration of reactants.

Conformational Equilibrium

Consider the simple reaction $A \rightleftharpoons B$, which corresponds, e.g., to a conformational change of a flexible molecule. The aim is to analyze how the population of the states A and B at equilibrium depends on the initial concentration $C_{A,0}$ of A. This means to monitor the molar fraction χ of each conformer as a function of $C_{A,0}$, which expresses the probability in solution. This problem can be solved by writing the equilibrium constant for the reaction above under the constrain of mass balance

$$\begin{cases} K_{eq} = \dfrac{C_B/C^{\ominus}}{C_A/C^{\ominus}} = \dfrac{C_B}{C_A} \\ C_A + C_B = C_{A,0} \end{cases} \tag{3.10}$$

This system can be solved by finding the zeros of

$$C_A + K_{eq} C_A - C_{A,0} = 0 \tag{3.11}$$

whose solution is analytical

$$C_A = \frac{1}{1+K_{eq}} C_{A,0} \tag{3.12}$$

$$C_B = \frac{K_{eq}}{1+K_{eq}} C_{A,0} \tag{3.13}$$

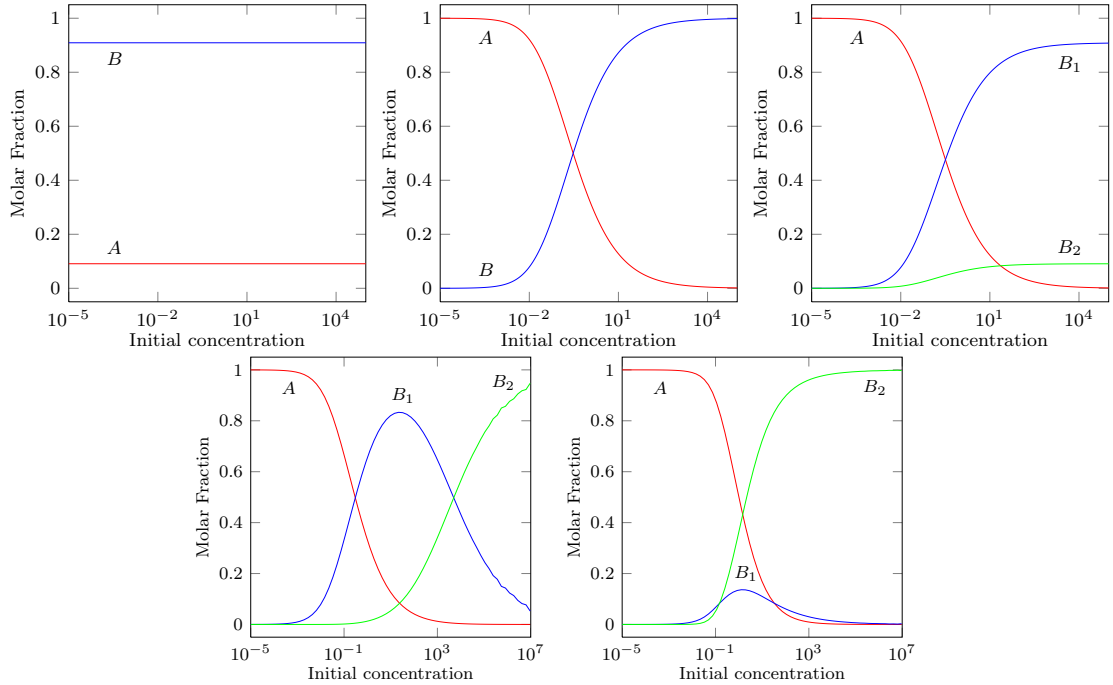


Figure 3.1: Molar fractions as a function of the initial concentration of A for different kind of equilibria.

Since the reaction preserves the total number of molecules, which is $N_{tot} = N_A + N_B = N_{A,0}$, the molar fractions of A and B are given by

$$\chi_A = \frac{1}{1 + K_{eq}} \tag{3.14}$$

$$\chi_B = \frac{K_{eq}}{1 + K_{eq}} \tag{3.15}$$

While the treatment of this reaction is very simple, it shows the interesting feature that the equilibrium probability of the two states A and B *only* depends on the value of the equilibrium constant K_{eq}, which means that whatever the initial concentration of A or B is, the equilibrium probabilities will not vary, and the outcome of the reaction is not concentration dependent. For instance, an equilibrium constant near one gives equiprobability to observe the two states, whereas a bigger value would favor the state B and *vice versa*. An example plot is reported in Fig. 3.1 where $K_{eq} = 10$.

Binding Reaction

Consider now a more complex chemical reaction that involves the association of two molecules of A to produce one molecule of B, like a dimerization $2A \rightleftharpoons B$. As before,

the molar fractions of A and B are studied as a function of the initial concentration $C_{A,0}$. At chemical equilibrium, and imposing the mass balance one obtains

$$\begin{cases} K_{eq} = \dfrac{C_B/C^{\circ}}{(C_A/C^{\circ})^2} = \dfrac{C_B}{C_A^2}C^{\circ} \\ C_A + 2C_B = C_{A,0} \end{cases} \tag{3.16}$$

which is solved by finding the zeros of

$$C_A + 2\frac{K_{eq}}{C^{\circ}}C_A^2 - C_{A,0} = 0 \tag{3.17}$$

Interestingly, as the reaction does not preserve the total number of molecules, the value of the equilibrium constant depends on the standard concentration. In the one-to-one reaction this dependence fortuitously cancels out. However, in all other reactions the standard state dependence does not cancel out making the numerical value of the equilibrium constant (as the one of the standard chemical potential difference) dependent on the arbitrary standard conditions.

By solving the previous equation and computing the molar fractions one obtains

$$\chi_A = \frac{4}{z+3} \tag{3.18}$$

$$\chi_B = \frac{z-1}{z+3} \tag{3.19}$$

with

$$z = \sqrt{8K_{eq}\frac{C_{A,0}}{C^{\circ}} + 1} \tag{3.20}$$

Clearly, the molar fractions in this case are concentration dependent, since the factor z depends on the initial concentration $C_{A,0}$. Fig. 3.1 reports the molar fraction of A and B as a function of $C_{A,0}$ for $K_{eq} = 10$, which shows a sigmoidal behavior. A characteristic point of the curve is $C_{A,1/2}$, that corresponds to the concentration for which the states A and B are equiprobable, i.e. $\chi_A = \chi_B = 1/2$. It is straightforward to show that $\chi_A = \chi_B$ when $C_{A,1/2} = 3C^{\circ}K_{eq}^{-1}$. As such, an increase of the equilibrium constant will correspond to a smaller $C_{A,1/2}$, i.e. a smaller initial concentration of A will be required to stabilize the associated state B. Also, this implies that even if the equilibrium constant is extremely small, at some perhaps unphysical concentration of A, B will be most favored. This is coherent with the idea that a reaction is driven by the interplay between enthalpic and entropic contributions. A dimerization reaction is driven by an enthalpy gain associated with the newly established interactions, and by an entropy cost, in particular associated with the loss of translational and rotational degrees of freedom on binding. At low concentration this entropic cost may be very high, since the volume accessible to each monomer is very high and the enthalpy contribution may be not

large enough to stabilize the associated state. At higher concentrations, the accessible volume to each molecule is smaller and the entropy cost of dimerization may be negligible. Stronger interactions means an increase in the enthalpic gain, so a decrease in the chemical potential of B, and thus reducing the standard chemical potential difference, which produces an increase of the equilibrium constant. Finally, as already observed, an increase in the equilibrium constant makes the value of $C_{A,1/2}$ lower. To complete the discussion, the dimer can be also preferred over the monomer due to entropic contributions, e.g. when the vibrations of the complex are "softer" than the vibration of the two separated monomers. Although not enough alone to overcome the translational and rotational cost, this vibrational entropy contribution can make the formation of the complex more favored[137].

Competitive Binding

Consider a second dimerization process, where two monomers A can react to give two distinct products B_1 and B_2

$$2A \rightleftharpoons B_1 \tag{3.21}$$

$$2A \rightleftharpoons B_2 \tag{3.22}$$

A characteristic standard chemical potential difference $\Delta\mu_1^\ominus$ and $\Delta\mu_2^\ominus$ is associated to each reaction, along with an equilibrium constant $K_{eq,1}$ and $K_{eq,2}$. The corresponding system to solve is thus

$$\begin{cases} K_{eq,1} = \dfrac{C_{B_1}/C^\ominus}{(C_A/C^\ominus)^2} = \dfrac{C_{B_1}}{C_A^2}C^\ominus \\ K_{eq,2} = \dfrac{C_{B_2}/C^\ominus}{(C_A/C^\ominus)^2} = \dfrac{C_{B_2}}{C_A^2}C^\ominus \\ C_A + 2C_{B_1} + 2C_{B_2} = C_{A,0} \end{cases} \tag{3.23}$$

which can be solved by finding the zeros of

$$C_A + 2\frac{K_{eq,1}}{C^\ominus}C_A^2 + 2\frac{K_{eq,2}}{C^\ominus}C_A^2 - C_{A,0} = 0 \tag{3.24}$$

It is still possible to solve this equation analytically

$$\chi_A = \frac{4}{z+3} \tag{3.25}$$

$$\chi_{B_1} = \frac{z-1}{z+3} \cdot \frac{K_{eq,1}}{K_{eq,1} + K_{eq,2}} \tag{3.26}$$

$$\chi_{B_2} = \frac{z-1}{z+3} \cdot \frac{K_{eq,2}}{K_{eq,1} + K_{eq,2}} \tag{3.27}$$

where

$$z = \sqrt{8\left(K_{eq,1} + K_{eq,2}\right)\frac{C_{A,0}}{C^{\ominus}} + 1} \tag{3.28}$$

Interestingly, one can see how this example uses the results obtained from both the conformational and the dimerization equilibria. First, the quilibrium probabilities are concentration dependent, since z depends on the initial concentration $C_{A,0}$; see Fig. 3.1. At low concentrations χ_A tends to one, and so only the state A is visible. Increasing the concentration, the population of A decreases while the ones of B_1 and B_2 increases. Second, in the limit of high initial concentrations of monomers, the concentration-dependence for B_1 and B_2 vanishes, and the probability of populating them depends only on the ratio of the $K_{eq,1}$ and $K_{eq,2}$ equilibrium constants. This is due to the molecularity of the two reactions. Both involve two molecules of A forming one molecule of B_1 or B_2. While the formation of one of the B states is preferred at high concentrations due to the smaller unfavorable translational entropy, the formation of B_1 or B_2 is mostly driven by their relative stability, so on the ratio of their equilibrium constants.

Thus, if the two dimerization reactions have similar equilibrium constants, both state will be observed at equilibrium, otherwise one phase will prevail the other. In Fig. 3.1 an example graph is plotted using $K_{eq,1} = 10$ and $K_{eq,2} = 1$.

Competitive Equilibria

Consider two association reactions with different molecularity: the same monomer A that forms a dimer (molecularity equal to two) or a trimer (molecularity equal to three)

$$2A \rightleftharpoons B_1 \tag{3.29}$$

$$3A \rightleftharpoons B_2 \tag{3.30}$$

At chemical equilibrium, the equation to be solved is now

$$C_A + 2\frac{K_{eq,1}}{C^{\ominus}}C_A^2 + 3\frac{K_{eq,2}}{C^{\ominus 2}}C_A^3 - C_{A_0} = 0 \tag{3.31}$$

Unfortunately, no analytical solutions can be found in this case, and the equation are to be solved numerically. Example plots are shown in Fig. 3.1. The interesting behavior is that in an intermediate range of concentrations, the state B_1 with molecularity equal to two dominates, but then disappears with increasing the concentration of monomers until the system is dominated by the state B_2 with a molecularity equal to three. This behavior is general and is not influenced by the value of the equilibrium constants: the state with the higher molecularity will dominate at high concentrations, even if its equilibrium constant is significantly smaller. The actual ratio between $K_{eq,1}$ and $K_{eq,2}$ plays

only a role in determining if, and at which range of concentrations, the intermediate state B_1 will be observed.

As a take home message, molecular assemblies with high molecularity are more stable at higher concentrations of monomers. This is important in 2D self-assembly, where as we shall see the molecularity of the reaction is associated with the surface packing of the SAM.

General Self-Assembly Process

Extending, the previous results to the generic self-assembly reaction $\alpha A \rightleftharpoons B$, at chemical equilibrium it yields

$$C_A + \alpha \frac{K_{eq}}{(C^{\ominus})^{\alpha-1}} C_A^{\alpha} - C_{A,0} = 0 \tag{3.32}$$

which implies finding the zeros of a polynomial of degree α. Unfortunately, analytical solutions for this kind of polynomial equations are not available, such that numerical approaches need to be used. Once a solution for C_A is found, the molar fractions of A and B can be found as before. It is also interesting to study when the two species A and B are equiprobable. From what discussed until here, it should be clear that looking at $\Delta\mu^{\ominus} = 0$ or $K_{eq} = 1$ is not useful since these quantities depends on the chosen standard state and, moreover, will describe the situation when $C_B = (C_A)^{\alpha}$. More informative is to determine the initial concentration of A that allows for equipartition of the A and B in solution, or $\chi_A = \chi_B$. The latter can be evaluated using the definition of the equilibrium constant

$$K_{eq} = \frac{(C_B/C^{\ominus})}{(C_A/C^{\ominus})^{\alpha}} \tag{3.33}$$

Imposing the equiprobability

$$C_A = C_B = C \tag{3.34}$$

one obtains

$$K_{eq} = \frac{(C/C^{\ominus})}{(C/C^{\ominus})^{\alpha}} = \left(\frac{C}{C^{\ominus}}\right)^{1-\alpha} \tag{3.35}$$

which can be re-elaborated as

$$C = C^{\ominus} K_{eq}^{1/(1-\alpha)} \tag{3.36}$$

Finally, the initial concentration of A corresponding to the equal probability of A and B can be obtained by the mass balance of the reaction

$$C_{A,\frac{1}{2}} = C_A + \alpha C_B = (\alpha + 1)\, C^{\ominus} K_{eq}^{-1/(\alpha-1)} \tag{3.37}$$

This equation reduces to $C_{A,\frac{1}{2}} = 3C^{\ominus}K^{-1}$ for $\alpha = 2$. Interestingly, it is standard state independent, since the equilibrium constant is correctly weighted by the standard state concentration.

In conclusion, starting from the balanced reaction of a chemical equilibrium, an equilibrium constant can be evaluated. Since a reaction is generally not one-to-one, the numerical value of the equilibrium constant is not very meaningful since it depends on the definition of the standard state and the equilibrium probabilities of A and B depend on the initial concentration of monomers. Only the analysis of the probability of the different states (or the molar fractions) through the value of K_{eq} as a function of the concentration provides a meaningful chemical description of the system at equilibrium.

3.2 STATISTICAL MECHANICS APPROACH TO THE CHEMICAL POTENTIAL

Classical Partition Function

In the previous section, it has been shown that the equilibrium probabilities of the different states of the system can be accessed from the knowledge of the equilibrium constant for a given reaction and the initial concentration of monomers. This section focuses on the numerical evaluation of the equilibrium constant by classical statistical thermodynamics.

The fundamental quantity to evaluate is the canonical partition function $Q(N, V, T)$, which depends on the number of molecules N, the volume V, and the temperature T of the system. It can be shown[138] that thermodynamic observables like the free energy F, the internal energy U, or the entropy S can be written as a function of the partition function Q

$$F = -k_B T \ln Q \tag{3.38}$$

$$U = k_B T^2 \frac{\partial}{\partial T} \ln Q \tag{3.39}$$

$$S = \frac{\partial}{\partial T} (k_B T \ln Q) \tag{3.40}$$

Remembering the beginning of this chapter, the chemical potential μ, or also molar free energy F_m, is defined as the intensive free energy over the number of molecules

$$\mu = F_m = \frac{\partial F}{\partial N} = -k_B T \frac{\partial \ln Q}{\partial N} \tag{3.41}$$

This definition can also be extended to the molar entropy S_m and the molar internal energy U_m, allowing to decompose the chemical potential in energetic and entropic contributions:

$$\mu = U_m - TS_m \tag{3.42}$$

Given the link between the partition function and the chemical potential, a way to evaluate the partition function is needed. For a system N molecules each one described by f degrees of freedom, the classical partition function can be expressed as an integral over all degrees of freedom of the Hamiltonian H of the system

$$Q = \frac{1}{h^{Nf}} \int dp^{Nf} \int dq^{Nf} \exp\left[-\frac{H(p,q)}{k_B T}\right] \tag{3.43}$$

where h is the Planck constant, k_B the Boltzmann constant, T the absolute temperature and $H(p,q)$ is the Hamiltonian as a function of the coordinates q and their associated momenta p. Unfortunately, the Hamiltonian of a polyatomic molecule is a complex function of the Nf momenta and coordinates and the integral cannot be solved without introducing approximations.

The first approximation is to consider all molecules in the system as identical and independent, which means the energy of one molecule is not influenced by any other molecule in the system. In other way, we consider the system to be an ideal gas, where all molecules are independent and indistinguishable. This allows to write the Hamiltonian of the whole system (H_{sys}) as sum of the Hamiltonian of the single molecules (H_{mol}), which in turn are all the same

$$H_{sys}(p,q) = \sum_i^N H_{mol,i}(p,q) = NH_{mol}(p,q) \tag{3.44}$$

The partition function can thus be decomposed in terms which only depends on the molecular Hamiltonian, which means the *canonical* partition function Q can be written as product of N *molecular* partition functions q:

$$\begin{aligned} Q &= \frac{1}{h^{Nf}} \int dp^{Nf} \int dq^{Nf} \exp\left[-\frac{1}{k_B T} N H_{mol}(p,q)\right] = \\ &= \left(\frac{1}{h^f} \int dp^f \int dq^f \exp\left[-\frac{1}{k_B T} H_{mol}(p,q)\right]\right)^N = q^N \end{aligned} \tag{3.45}$$

This last expression does not consider the indistinguishability of the N molecules, and thus brings to an overestimation of the canonical partition function due to over count-

ing of identical states. To correct for this, the partition function is divided by the total number of identical arrangement of the N molecules, giving

$$Q(N, V, T) = \frac{q(V, T)^N}{N!} \tag{3.46}$$

This is a fundamental step, since now the problem of the evaluation of the canonical partition function is recast into the evaluation of the partition function of one molecule, which depends only on the volume of the system, its temperature, and a set of molecular properties.

Polyatomic Molecule

Consider now a generic polyatomic molecule composed of N atoms that is free to move in a box of volume V. Each atom of the molecule can move in all three Cartesian directions, giving a total of $3N$ degrees of freedom to describe the full phase space of the molecule. Among these, three degrees are associated with the free translations of the center of mass, and three to the rotations. The remaining $3N - 6$ are considered as internal vibrations. In order to evaluate the partition function of this system, a number of approximations needs to be introduced. The first is to consider that the molecule moves as a rigid body. This implies that the rotations and translations are not coupled with the vibrations. Thus, if the energy depends only on the local conformation of the molecule, the total Hamiltonian can be decomposed into four independent terms – translational, rotational, vibrational and electronic

$$H(p, q) = H_{tr}(p, q) + H_{rot}(p, q) + H_{vib}(p, q) + H_{elec} \tag{3.47}$$

If so, the molecular partition function can be also decomposed as a product of terms

$$q = q_{tr}\, q_{rot}\, q_{vib}\, q_{elec} \tag{3.48}$$

With this definition of the molecular partition function is it also possible to decompose the free energy and the other thermodynamic quantities into sums of contributions; e.g. for the free energy

$$F = -k_B T \ln Q = -k_B T \ln \frac{q^N}{N!} \tag{3.49}$$

and using Stirling approximation for $\ln N!$ it yields

$$\ln Q = N \ln q - N \ln N + N = N \ln \frac{qe}{N} \tag{3.50}$$

and

$$\begin{aligned} F &= -Nk_B T \ln \frac{qe}{N} = \\ &= -Nk_B T \ln \frac{q_{tr} e}{N} - Nk_B T \ln q_{rot} - Nk_B T \ln q_{vib} - Nk_B T \ln q_{elec} = \\ &= F_{tr} + F_{rot} + F_{vib} + F_{elec} \end{aligned} \tag{3.51}$$

where the free energy F has been decomposed as a sum of a translational F_{tr}, rotational F_{rot}, vibrational F_{vib}, and electronic F_{elec} term. The same decomposition can be done for the internal energy, entropy, or the chemical potential. The same can be done for the chemical potential, obtaining

$$\begin{aligned} \mu &= -k_B T \ln \frac{q}{N} = \\ &= -k_B T \ln \frac{q_{tr}}{N} - k_B T \ln q_{rot} - k_B T \ln q_{vib} - k_B T \ln q_{elec} = \\ &= \mu_{tr} + \mu_{rot} + \mu_{vib} + \mu_{elec} \end{aligned} \tag{3.52}$$

Now the problem is to find expressions for the translational, rotational, vibrational, and electronic partition functions, which is done in the next subsections using simplified test cases.

Free translating particle

At first, the translational degrees of freedom are studied. Consider a particle of mass m free to move along a string of length L. The particle can move along the string in both directions, and at whichever velocity. Moreover, the motion is free, so that the potential energy is zero in all available space L. Given this, the Hamiltonian of this system is just the kinetic energy of the particle, so it is only a function of the momentum p and the mass m of the particle as

$$H(p) = \frac{p^2}{2m} \tag{3.53}$$

The partition function for this system can be easily integrated. Since only one degree of freedom is present

$$q(T, L) = \frac{1}{h} \int_{-\infty}^{+\infty} dp \int_0^L dq \exp\left[-\frac{1}{k_B T}\left(\frac{p^2}{2m}\right)\right] \tag{3.54}$$

where the momentum is integrated over the whole unlimited space, while the position over the available length L. Since the kinetic term is only a function of the momentum p, and not of the position q, the integral over q can be separated

$$q(T, L) = \frac{1}{h}\int_{-\infty}^{+\infty} dp \exp\left[-\frac{1}{k_B T}\left(\frac{p^2}{2m}\right)\right] \cdot \int_0^L dq \tag{3.55}$$

The integral over the momentum is a Gaussian integral, so solvable analytically, while the integral over the position is just the dimension of the available space. So the partition function can be written in closed form as

$$q(T, L) = \sqrt{\frac{2\pi m k_B T}{h^2}} L \tag{3.56}$$

Enhance now the degrees of freedom of the particle, allowing it to move in a three dimensional orthorombic box of dimension L_x, L_y, and L_z. As before, the potential energy is zero in all points of the space, but now the system is defined by three degrees of freedom, which correspond to the motion in the three cartesian coordinates x, y, and z, and the motion of the particle can be described by its momentum projected on the three coordinates, obtaining three independent momenta p_x, p_y, and p_x. The partition function can thus be written as

$$\begin{aligned} q(T, L_x, L_y, L_z) = &\frac{1}{h}\int_{-\infty}^{+\infty} dp_x \exp\left[-\frac{1}{k_B T}\left(\frac{p_x^2}{2m}\right)\right] \cdot \int_0^{L_x} dq_x \cdot \\ &\cdot\frac{1}{h}\int_{-\infty}^{+\infty} dp_y \exp\left[-\frac{1}{k_B T}\left(\frac{p_y^2}{2m}\right)\right] \cdot \int_0^{L_y} dq_y \cdot \\ &\cdot\frac{1}{h}\int_{-\infty}^{+\infty} dp_z \exp\left[-\frac{1}{k_B T}\left(\frac{p_z^2}{2m}\right)\right] \cdot \int_0^{L_z} dq_z \end{aligned} \tag{3.57}$$

where all six integrals over the three momenta and coordinates have been split due to non-correlation between them. Clearly, each line of this partition function resembles the partition function of the particle moving in just one dimension, so it possible to write

$$q(T, L_x, L_y, L_z) = q_x(T, L_x) \cdot q_y(T, L_y) \cdot q_z(T, L_z) \tag{3.58}$$

which can be written as

$$q(T, L_x, L_y, L_z) = \sqrt{\frac{2\pi m k_B T}{h^2}} L_x \cdot \sqrt{\frac{2\pi m k_B T}{h^2}} L_y \cdot \sqrt{\frac{2\pi m k_B T}{h^2}} L_z \tag{3.59}$$

or, in a more compact way as

$$q_{tr}(V, T) = \left(\frac{2\pi m k_B T}{h^2}\right)^{\frac{3}{2}} V \tag{3.60}$$

noting that $L_x \cdot L_y \cdot L_z = V$ is the volume of the box. This is the general partition function of a particle of mass m moving in a box of volume V at a temperature T. From this last definition, the thermodynamic quantities associated to the translational degrees of freedom can be easily derived

$$F_{tr} = -Nk_B T \ln \left[\left(\frac{2\pi m k_B T}{h^2} \right)^{\frac{3}{2}} \frac{Ve}{N} \right] \tag{3.61}$$

$$S_{tr} = Nk_B \left[\ln \left(\frac{2\pi m k_B T}{h^2} \right)^{\frac{3}{2}} + \frac{5}{2} - \ln \frac{N}{V} \right] \tag{3.62}$$

$$U_{tr} = \frac{3}{2} Nk_B T \tag{3.63}$$

where the equation for the entropy is the well known Sakur-Tetrode equation. The respective molar quantities can be obtained deriving the previous expression by the number of molecules N, which yields

$$\mu_{tr} = -k_B T \ln \left[\left(\frac{2\pi m k_B T}{h^2} \right)^{\frac{3}{2}} \frac{V}{N} \right] \tag{3.64}$$

$$S_{tr,m} = k_B \left[\ln \left(\frac{2\pi m k_B T}{h^2} \right)^{\frac{3}{2}} + \frac{3}{2} - \ln \frac{N}{V} \right] \tag{3.65}$$

$$U_{tr,m} = \frac{3}{2} k_B T \tag{3.66}$$

Notably, the translational partition function, and thus also the free energy, entropy and internal energy, can be evaluated simply by knowing the temperature T of the system, the ratio N/V, which is the concentration, and the mass m of the particle.

Rigid Rotor

Consider now a rigid object which can freely rotate around its center of mass. Its orientation in space is best described in spherical coordinates by three Eulerian angles θ, ϕ, and ψ. If the center of gravity is fixed, which means no net translation, the Hamiltonian is composed by just the kinetic energy term, written as a function of these angles, and their conjugate momenta

$$\begin{aligned} H = &\frac{\sin^2 \psi}{2I_A} \left\{ p_\theta - \frac{\cos \psi}{\sin \theta \sin \psi} \left(p_\phi - \cos \theta p_\psi \right) \right\}^2 + \\ &+ \frac{\cos^2 \psi}{2I_B} \left\{ p_\theta - \frac{\sin \psi}{\sin \theta \cos \psi} \left(p_\phi - \cos \theta p_\psi \right) \right\}^2 + \frac{1}{2I_C} p_\psi^2 \end{aligned} \tag{3.67}$$

In this expression I_A, I_B and I_C are the moments of inertia of the molecule, which are defined as the mass property of a rigid body that determines the torque required to

accelerate the rotation around the three principal axes. The moments of inertia depend on the shape of the body and may be different around different axis of rotation: a larger moment of inertia around a given axis requires more torque to increase, or stop, the rotation of the molecule about that axis. Following the derivation of Mayer & Mayer[139], the Hamiltonian can be rewritten for easier manipulation as

$$\begin{aligned} H = \frac{1}{2}\left(\frac{\sin^2\psi}{I_A} + \frac{\cos^2\psi}{I_B}\right) \cdot \\ \cdot \left\{ p_\theta + \left(\frac{1}{I_B} - \frac{1}{I_A}\right) \frac{\sin\psi\cos\psi}{\sin\theta\left(\frac{\sin^2\psi}{I_A} + \frac{\cos^2\psi}{I_B}\right)} \left(p_\phi - \cos\theta p_\psi\right) \right\}^2 + \\ + \frac{1}{2I_A I_B \sin^2\theta\left(\frac{\sin^2\psi}{I_A} + \frac{\cos^2\psi}{I_B}\right)} \left(p_\phi - \cos\theta p_\psi\right)^2 + \frac{1}{2I_C} p_\psi^2 \end{aligned} \tag{3.68}$$

The partition function is thus defined as

$$q = \frac{1}{h^3} \int_{-\infty}^{+\infty} dp_\theta \int_{-\infty}^{+\infty} dp_\phi \int_{-\infty}^{+\infty} dp_\psi \int_0^\pi d\theta \int_0^{2\pi} d\phi \int_0^{2\pi} d\psi \exp\left(\frac{-H}{k_B T}\right) \tag{3.69}$$

which seems to lead to a formidable integral. Even so, it can be actually analytically solved without difficulties resorting to the well know Gaussian integral. First, integrating over p_θ leads to the factor

$$(2\pi k_B T)^{1/2} \left(\frac{\sin^2\psi}{I_A} + \frac{\cos^2\psi}{I_B}\right)^{-1/2} \tag{3.70}$$

Subsequent integration over p_ϕ yields

$$(2\pi k_B T I_A I_B)^{1/2} \sin\theta \left(\frac{\sin^2\psi}{I_A} + \frac{\cos^2\psi}{I_B}\right)^{1/2} \tag{3.71}$$

which partially cancels out with the term obtained in the first integration. Integration over the third momentum p_ψ yields

$$(2\pi k_B T I_C)^{1/2} \tag{3.72}$$

and, finally, integration of $\sin\theta d\theta$ from 0 to π gives 2, and the other two angles give each a factor 2π, so that

$$q = \pi^{1/2} \left(\frac{8\pi^2 k_B T}{h^2}\right)^{3/2} (I_A I_B I_C)^{1/2} \tag{3.73}$$

This is the closed-form expression of the rotational partition function for an asymmetric rigid rotor, which depends only on the three moments of inertia I_A, I_B and I_C, and

Point Group	Symmetry Number σ
$C_1, C_i, C_s, C_{\infty v}$	1
$D_{\infty h}$	2
C_n, C_{nv}, C_{nh}	n
D_n, D_{nh}, C_{nd}	$2n$
T, T_d, T_h	12
S_n	$n/2$
O, O_h	24
I, I_h	60

Table 3.1: Symmetry numbers as a function of the point group of a molecule.

the temperature T. If the molecule is symmetric, a number of conformations are identical upon rotation, and these indistinguishable states are overcounted. As the number of identical states upon rotation is identified by the symmetry number σ, the correct partition function is σ-times smaller, thus yielding

$$q_{rot}(T) = \frac{\pi^{1/2}}{\sigma}\left(\frac{8\pi^2 k_B T}{h^2}\right)^{3/2}(I_A I_B I_C)^{1/2} \tag{3.74}$$

From this definition, the free energy, entropy and internal energy can be derived

$$F_{rot} = -Nk_B T\left[\ln\left(\frac{8\pi^2 k_B T}{h^2}\right)^{\frac{3}{2}} + \ln\left(\frac{\pi}{\sigma^2} I_A I_B I_C\right)^{\frac{1}{2}}\right] \tag{3.75}$$

$$S_{rot} = Nk_B\left[\ln\left(\frac{8\pi^2 k_B T}{h^2}\right)^{\frac{3}{2}} + \ln\left(\frac{\pi}{\sigma^2} I_A I_B I_C\right)^{\frac{1}{2}} + \frac{3}{2}\right] \tag{3.76}$$

$$U_{rot} = \frac{3}{2}Nk_B T \tag{3.77}$$

with the molar analogous obtained by dividing these expressions for the number of molecules N in the system.

These quantities depend only on the three moments of inertia and the symmetry number σ. Since a rotation in 3D is described by a three-dimensional inertia tensor, the principal axes of rotation and the inertia moments can be obtained from the eigenvectors and eigenvalues of the inertia tensor. Thus, only the conformation of the molecule at the minimum of the potential energy surface is necessary, and the three moments of inertia are trivially determined. If the principal axes of rotation are not needed, the diagonalization of the inertia tensor is not strictly required, since the product of the inertia moments can be obtained as the determinant of the inertia tensor (the diagonalization of a square matrix does not change its determinant). Also, the symmetry number can be easily obtained from the optimized geometry, searching for its symmetry point group and the corresponding symmetry number; see Table 3.1.

Classical Harmonic Oscillator

The third term to evaluate is related to the internal vibrations. As a toy system, consider two masses m_1 and m_2 connected by a spring of force constant k. If $\mu = m_1 m_2/(m_1 + m_2)$ is the reduced mass, the Hamiltonian of the harmonic oscillator can be written as

$$H = \frac{p^2}{2\mu} + \frac{1}{2}kq^2 \tag{3.78}$$

where p and q are the momentum and the position describing the vibrational motion. The corresponding partition function is thus

$$q = \frac{1}{h}\int_{-\infty}^{+\infty} dp \int_{-\infty}^{+\infty} dq \exp\left[-\frac{1}{k_B T}\left(\frac{p^2}{2\mu} + \frac{1}{2}kq^2\right)\right] \tag{3.79}$$

which can be separated in two Gaussian integrals, each solvable analytically

$$\begin{aligned} q &= \frac{1}{h}\int_{-\infty}^{+\infty} dp \exp\left(-\frac{p^2}{2\mu k_B T}\right)\int_{-\infty}^{+\infty} dq \exp\left(-\frac{kq^2}{2k_B T}\right) = \\ &= \frac{1}{h}\sqrt{2\pi\mu k_B T}\sqrt{\frac{2\pi k_B T}{k}} = \frac{2\pi k_B T}{h}\sqrt{\frac{\mu}{k}} = \frac{k_B T}{h\nu} \end{aligned} \tag{3.80}$$

where

$$\nu = \frac{1}{2\pi}\sqrt{\frac{k}{\mu}} \tag{3.81}$$

is the frequency of the oscillator, and solely determines the value of the vibrational partition function.

This result can be used to study the vibrational partition function of a polyatomic molecule. The potential energy function V of a molecule is generally a complicated function of all the f internal coordinates q_i. Assuming that the flexibility of the molecule is entirely described by small amplitude vibrations around the position $q_{0,i}$ of the energy minimum, the potential energy can be approximated as a Taylor expansion around the minimum

$$\begin{aligned} V(q) \simeq \sum_i^f V(q_{0,i}) + \sum_i^f \frac{\partial V(q_{0,i})}{\partial q_i}(q_i - q_{0,i}) + \\ + \frac{1}{2}\sum_i^f\sum_j^f \frac{\partial^2 V(q_{0,i})}{\partial q_i \partial q_j}(q_i - q_{0,i})(q_j - q_{0,j}) + \dots \end{aligned} \tag{3.82}$$

The first term of the expansion is the energy of the molecule in the minimum. This is a constant for a given conformation and is exactly the electronic energy, which cor-

responds to the last term of molecular Hamiltonian. The second term vanishes, since the first partial derivatives of the potential energy around a minimum are by definition equal to zero. What remains is the third term and higher orders contributions, which are negligible in the harmonic approximation. The problem now is that the second order term is not written as a sum of independent contributions arising from each coordinate q_i, but it contains cross terms, which depend at the same time on the coordinates q_i and q_j. By choosing a new set of coordinates $\tilde{q}_i$ that form an orthonormal coordinates set and whose value is zero at the position of the minimum ($\tilde{q}_{0,i} = 0$), Eq. 3.82 becomes

$$V(\tilde{q}) \simeq \frac{1}{2} \sum_i^f \frac{\partial^2 V(\tilde{q}_{0,i})}{\partial^2 \tilde{q}_i} \tilde{q}_i^2 = \sum_i^f \frac{1}{2} k_i \tilde{q}_i \tag{3.83}$$

where the curvatures of the potential energy is indicated by the force constants k_i. With this definition of the potential energy, the full Hamiltonian is obtained as a sum of the kinetic energy associated to the new coordinates and their potential energy just derived

$$H(p, \tilde{q}) \simeq \sum_i^f \frac{p_i^2}{2\mu_i} + \frac{1}{2} \sum_i^f k \tilde{q}_i^2 \tag{3.84}$$

where p_i, μ_i, and k_i are the momentum, reduced mass and force constant associated with the coordinate $\tilde{q}_i$. From this formulation, it is now straightforward to write the Hamiltonian as sum of f independent terms

$$H(p, \tilde{q}) \simeq \sum_i^f \left(\frac{p_i^2}{2\mu_i} + \frac{1}{2} k_i \tilde{q}_i^2 \right) = \sum_i^f H_i(p_i, \tilde{q}_i) \tag{3.85}$$

and the total vibrational partition function as a product of independent contributions as

$$q_{vib,cl} = \prod_i^f q_{vib,i} = \prod_i^f \frac{k_B T}{h\nu_i} \tag{3.86}$$

From this expression, the thermodynamic quantities related to the vibrational degrees of freedom can be derived

$$F_{vib,cl} = -N k_B T \sum_{i=1}^f \ln \frac{k_B T}{h\nu_i} \tag{3.87}$$

$$S_{vib,cl} = N k_B \left(f + \sum_{i=1}^f \ln \frac{k_B T}{h\nu_i} \right) \tag{3.88}$$

$$U_{vib,cl} = N k_B T f \tag{3.89}$$

with their molar analogous obtained dividing by N.

To evaluate these vibrational contributions, the required molecular properties are the vibrational frequencies of the $3N-6$ internal modes. As shown above, these quantities depend on the second derivative of the energy with respect to the internal coordinates, that is the Hessian calculated at the minimum of the potential energy. Usually, the Hessian is computed in cartesian coordinates and is diagonalized to obtain a set of independent vibrational coordinates $\tilde{q}_i$. In practice, it is even more useful to diagonalize the mass-weighted hessian: in this way the eigenvalues obtained are directly the frequencies associated to the normal modes of the molecule.

Electronic partition function

The forth and last term to the molecular partition function is the electronic contribution. As saw in the previous subsection, the electronic energy for a molecule corresponds to its ground state energy E, that is the potential energy of the conformation at the minimum. Since the system considered here is made covalently bonded atoms at room temperature (i.e. a molecule) the excited electronic states are usually not accessible. This implies the partition function relative to the ground state energy is just

$$q_{elec} = \exp\left(-\frac{E}{k_B T}\right) \tag{3.90}$$

From this definition, the electronic contributions to the thermodynamic observables are

$$F_{elec} = -Nk_B T \ln q_{elec} = NE \tag{3.91}$$

$$U_{elec} = Nk_B T^2 \frac{\partial}{\partial T} \ln q_{elec} = NE \tag{3.92}$$

$$S_{elec} = Nk_B \ln q_{elec} + Nk_B T \frac{\partial}{\partial T} \ln q_{elec} = 0 \tag{3.93}$$

and their molar analogous are obtained dividing by N. One should note that the energy of the system is not interesting per se. What is more interesting is the difference in energy between two states. Based on this, the electronic energy of one state is usually defined as the zero of the energy scale and all other energies are defined relative to it.

Quantum Correction to the Vibrational Partition Function

All expressions derived up to now refer to a classical description of system. While this is reasonably true for the rotations and translations at room temperature, it can be significantly inaccurate for the vibrations.

Using a quantum description of the harmonic oscillator, the energy of the quantum vibrational levels are

$$E_n = \left(n + \frac{1}{2}\right) h\nu \tag{3.94}$$

where n is the quantum index of the n-excited level, and ν is the frequency of the vibration. The partition function associated to the quantum oscillator is obtained by summing up over all discrete energy levels

$$\begin{aligned} q &= \sum_n \exp\left(-\frac{E_n}{k_B T}\right) = \exp\left(-\frac{h\nu}{2k_B T}\right) \sum_{n=0}^{+\infty} \exp\left(-\frac{nh\nu}{k_B T}\right) \\ &= \frac{\exp\left(-\frac{h\nu}{2k_B T}\right)}{1 - \exp\left(-\frac{h\nu}{k_B T}\right)} = \left[2\sinh\left(\frac{h\nu}{2k_B T}\right)\right]^{-1} \end{aligned} \tag{3.95}$$

While this expression is valid for one harmonic oscillator, it can be extended to the $3N-6$ vibrations of a polyatomic molecule exactly as done for the classical oscillator yielding

$$q_{vib,qm} = \prod_{i=1}^{f} \frac{\exp\frac{h\nu}{2k_B T}}{1 - \exp\frac{h\nu}{k_B T}} = \prod_{i=1}^{f} \left[2\sinh\left(\frac{h\nu}{2k_B T}\right)\right]^{-1} \tag{3.96}$$

and, the free energy, entropy and internal energy associated to this partition function as

$$F_{vib,qm} = Nk_B T \sum_i^f \ln\left[2\sinh(x)\right] \tag{3.97}$$

$$U_{vib,qm} = Nk_B T \sum_i^f \frac{x}{\tanh(x)} \tag{3.98}$$

$$S_{vib,qm} = Nk_B \sum_i^f \left[\frac{x}{\tanh(x)} - \ln\left[2\sinh(x)\right]\right] \tag{3.99}$$

with

$$x = \frac{h\nu}{2k_B T} \tag{3.100}$$

While the solution obtained for the classical harmonic oscillator is valid only in the limit that the excited vibrational level are not populates, these new expressions provide a correct quantum solution, which is valid in the entire range of frequency and temperature. Since a closed form for the vibrational partition function is available, it is possible to cor-

rect the results obtained using a classical description of the vibrations, and introducing a quantum free energy correction in the form

$$F_{qm\,corr} = F_{vib,qm} - F_{vib,cl} \tag{3.101}$$

where $F_{vib,qm}$ and $F_{vib,cl}$ are the quantum and classical vibrational free energy, respectively. Since both classical and quantum expressions cam be evaluated numerically using the vibrational frequencies ν_i, the correction is straightforward. This quantum correction was shown to be as large as about 1 kcal mol^{-1} for a small protein[140].

Standard State

As shown at the beginning of the chapter, it is useful to make the concentration dependence of the chemical potential explicit. Since the translational contribution is the only term that depends on the number of molecules and the volume, it is possible to write an expression for it that separates out the concentration dependent from the concentration independent contributions as

$$\begin{aligned}
\mu_{tr} &= -k_B T \ln \left[\left(\frac{2\pi m k_B T}{h^2} \right)^{\frac{3}{2}} \frac{V}{N} \right] = \\
&= -k_B T \ln \left[\left(\frac{2\pi m k_B T}{h^2} \right)^{\frac{3}{2}} \frac{V^\ominus}{N^\ominus} \right] + k_B T \ln \frac{N}{V} \frac{V^\ominus}{N^\ominus} = \\
&= -k_B T \ln \left[\left(\frac{2\pi m k_B T}{h^2} \right)^{\frac{3}{2}} \frac{1}{C^\ominus} \right] + k_B T \ln \frac{C}{C^\ominus} = \\
&= \mu_{tr}^\ominus + k_B T \ln \frac{C}{C^\ominus}
\end{aligned} \tag{3.102}$$

To do so it was necessary to introduce a reference concentration $C^\ominus$. This implies that the chemical potential depends on $\mu_{tr}^\ominus$ which is independent of the actual concentration of the experiment. Using this result, the chemical potential of the state A is

$$\mu_A = \mu_A^\ominus + k_B T \ln \frac{C_A}{C^\ominus} \tag{3.103}$$

which expresses the dependence on the concentration as a "perturbation" from the standard chemical potential. This allows the use of tables of standard chemical potentials, which all refer to a given standard concentration. A second consequence of this manipulation is that it sets the concentration scale: the concentration of the system is now expressed in units of $C^\ominus$. Changing the standard state, which is arbitrary defined, results in different numerical value for the standard chemical potential $\mu^\ominus$, as well as the value of the equilibrium constant, but the equilibrium probabilities of reactants and products will not change.

Summing Up Everything

In conclusion, a statistical mechanics approach to the chemical potential of a molecular system has been presented. Upon the introduction of a number of approximations, i.e. the ideal gas and the rigid-rotor harmonic oscillator approximations, the chemical potential can be decomposed in four terms, for which closed-form expressions exist. In the limit of the approximations, the molecular partition function is

$$q(V,T) = \left[\left(\frac{2\pi m k_B T}{h^2}\right)^{\frac{3}{2}} V\right]_{tr} \cdot \left[\frac{\sqrt{\pi}}{\sigma}\left(\frac{8\pi^2 k_B T}{h^2}\right)^{\frac{3}{2}} \sqrt{I_A I_B I_C}\right]_{rot} \cdot \left[\prod_i^{\kappa} 2\sinh\left(\frac{h\nu_i}{2k_B T}\right)\right]_{vib} \cdot \left[e^{E/k_B T}\right]_{elec} \tag{3.104}$$

whose numerical value at a given temperature and concentration solely depends on molecular properties, such as the mass m of the molecule, the moments of inertia I_A, I_B, and I_C, the symmetry number σ, the vibrational frequencies ν_i, and the ground state energy E. In this expression the classical partition function has been substituted with the its quantum analogous to account for quantum effects otherwise neglected. Using this definition for the molecular partition function, the chemical potential is

$$\begin{aligned}\mu(V,T) = &- kT \ln\left[\left(\frac{2\pi m k T}{h^2}\right)^{\frac{3}{2}} \frac{V}{N}\right] + \\ &- kT \ln\left[\frac{\sqrt{\pi}}{\sigma}\left(\frac{8\pi^2 kT}{h^2}\right)^{\frac{3}{2}} \sqrt{I_X I_Y I_Z}\right] + \\ &- kT \sum_{j=1}^{3n-6} \ln\left[2\sinh\left(\frac{h\nu_j}{2kT}\right)\right] + \\ &- E\end{aligned} \tag{3.105}$$

which can be used to explore chemical reactions at equilibrium.

3.3 MODELING SUPRAMOLECULAR HEXAMERIZATION

In this section, the theory above is applied to a simple example. In the case study, the Janus-type molecule[141] in Fig. 3.2 is modeled to form a supramolecular macrocycle mediated by hydrogen bonds[142]. The reaction involve the aggregation of six monomers A to form one hexameric structure B

$$6A \rightleftharpoons B \tag{3.106}$$

To evaluate the equilibrium constant for this reaction, the difference in standard chemical potentials is calculated using Eq. 3.105 derived in the previous section. Thus, the required molecular quantities for the monomer and the hexamer are: the mass m, the three moments of inertia I_A, I_B, I_C, the symmetry number σ, the electronic energy E, and the vibrational modes ν_i. While the mass can be easily obtained as the molecular weight of the two species, for all other quantities a molecular model is necessary.

A fast way to create a molecular model for the monomer and the hexamer is using a classical force field (FF). Here, the General Amber Force Field (GAFF)[123,124] is used, with automatic topologies created by the ANTECHAMBER software[124] and the ACPYPE script[143]. For the monomer, a single molecule is simulated in vacuum. Since the molecule is rigid, the optimal conformation is obtained by a simple energy minimization using the adopted basis Newton Raphson (ABNR) energy-minimization algorithm. For the hexamer, six molecules have been disposed on the plane to mimic the expected supramolecular architecture. As for the monomer, the optimal geometry was obtained by energy minimization. Interestingly, the optimized geometry for the hexamer is not planar, but resembled to a spherical cap with mean and gaussian curvatures of 0.033 Å^{-1} and 0.0011 Å^{-2}, respectively (comparable to the curvature of a sphere of radius 30 Å); see Fig. 3.3.

From the optimized geometries, the moments of inertia were calculated by diagonalizing the inertia tensor (see Tab. 3.2) and the symmetry number was obtained from the symmetry point group of the molecule. Since the monomer is asymmetric, its symmetry number is one, whereas for the haxamer is six, due to the six-fold rotation axis passing through the center of the supramolecule. The vibrational modes were obtained by normal mode analysis. The six lowest internal modes are reported in Tab. 3.2. The striking difference between the monomer and the hexamer is that the vibrational modes in the hemaxer are much softer (lower frequencies), which indicates the presence of large-scale collective motions, involving many atoms. Last, the electronic energy was obtained from the force-field energy at the minimum. Both the minimization and the normal mode analysis were performed using the CHARMM software[144].

From the calculated molecular quantities (see Tab. 3.2), and the definition of the ensemble conditions of temperature (300 K) and concentration (1 M), the chemical potentials of the monomer and hexamer were calculated using Eq. 3.105. Since these calculations are error prone, mostly due to the conversions between units of measure, a C program called *Thermo*, was developed[145].

The calculated chemical potentials for the monomer and hexamer and the chemical potential difference for the hexamerization reaction are given in Tab. 3.3, along with the decomposition in translational, rotational, vibrational and electronic contributions. As expected, the hexamerization reaction is driven by a very large energy gain ($-169.4\ \text{kcal}\,\text{mol}^{-1}$) due to the 18 hydrogen bonds formed between neighbor molecules. Nonetheless, it involves a large entropy cost ($\Delta\mu^{\circ}_{tr} = 44.4\ \text{kcal}\,\text{mol}^{-1}$ and $\Delta\mu_{rot} =$

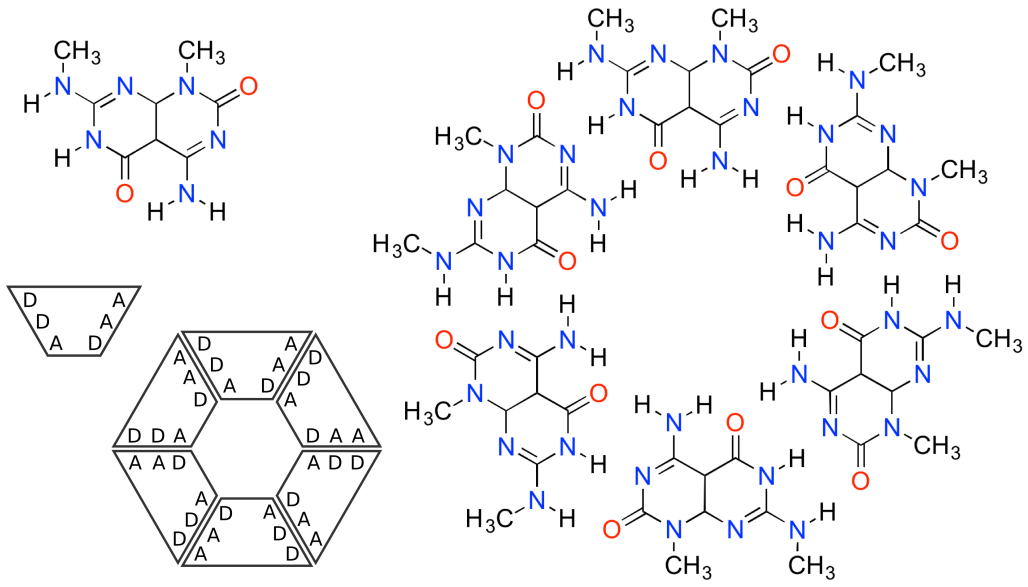


Figure 3.2: Chemical structures of the monomer and hexamer studied in this section. The monomer show two complementary hydrogen pattern (AAD and DDA) on the two sides of the molecule, allowing the self-assembly in the hexamer.

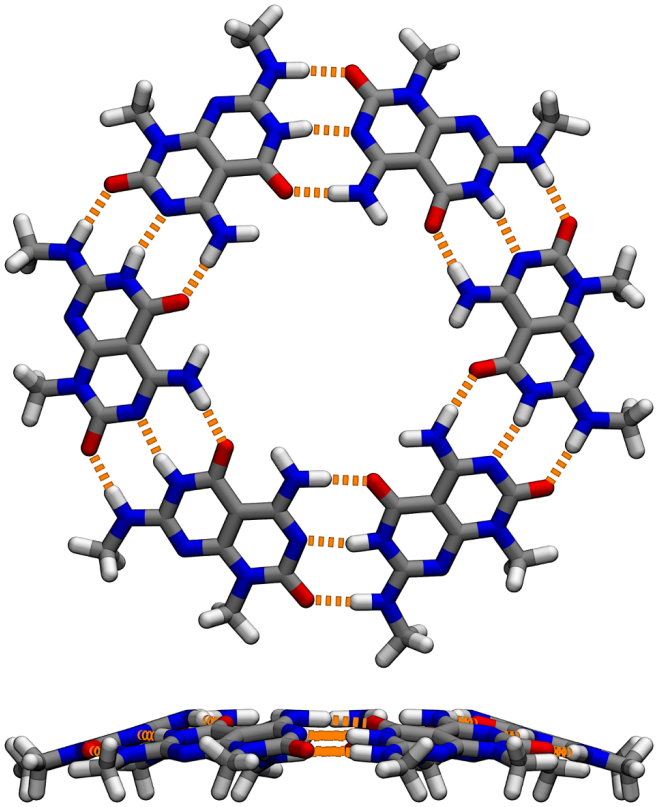


Figure 3.3: Top and side view of the optimized model of the hexamer. In the side view the curvature is evident as all monomers are not flat on the surface, but are slightly tilted upwards by about 20°.

	Monomer	Hexamer
Number of atoms	26	156
Electronic energy [kcal mol^{-1}]	-391.9	-2520.8
Mass m [g mol^{-1}]	222.2	1333.2
Inertia moments [g mol^{-1} Å^{-2}]	636.6	46215.2
	1059.4	46215.4
	1689.3	91748.8
Number of vibrations	72	462
Lowest vibrational frequencies [cm^{-1}]	46.2	5.0
	58.6	5.0
	92.6	10.4
	99.0	14.3
	114.4	15.2
	148.8	15.2
Symmetry number	1	6

Table 3.2: Molecular properties of the monomer and the hexamer used to calculate their chemical potentials.

41.5 kcal mol^{-1}), which is due the loss of a number of translational and rotational degrees of freedom on self-assembly. Since the total number of degrees of freedom is conserved, all the lost translations and rotations are converted into vibrations in the hemaxer. This increased "vibrational freedom" is reflected in a vibrational entropy gain (-19.6 kcal mol^{-1}), which partially counterbalances the rotational and translational cost. This is a common feature for most aggregation reactions, as it was shown insulin dimerization[137]. Overall the difference in chemical potential at standard state is -103.1 kcal mol^{-1}, showing that the hexamer is highly favored with respect to the monomer at standard conditions.

Since the translational contribution to the chemical potentials is strongly concentration dependent, the chemical potential difference as well as the outcome of the hexamerization reaction will be strongly dependent on the experimental conditions. As shown at the beginning of this chapter, for a chemical reaction in the form $\alpha A \rightleftharpoons B$ the characteristic concentration $C_{A,\frac{1}{2}}$, which corresponds to equiprobability of the monomeric and the self-assembled forms, can be calculated analytically. Using Eq. 3.37), the value of $C_{A,\frac{1}{2}}$ for the hexamerization reaction $\sim 5 \cdot 10^{-15}$ M. Fig 3.4 shows the molar fractions of the monomer and hexamer at equilibrium as a function of the initial concentration of monomers evaluated numerically. Strikingly, at low concentration the monomeric state is preferred, whereas at higher concentration the hexameric state is most favored. The concentration range can thus be divided decomposed in two domains of dominance of the monomer/hexamer at low/high concentration. The boundary between the two domains is $C_{A,\frac{1}{2}}$, which is represented as dashed gray line in the plot. This result is analogous to the dimerization equilibrium studied previously and all considerations done previously hold. What is most striking is the predicted concentration range, which goes from 10^{-18} to 10^{-10} M. This range is clearly non physical in solution and is due to

	μ_{elec}	μ_{tr}	μ_{rot}	$\mu_{vib,cl}$	μ_{tot}
Monomer (A)	-391.9	-9.2	-8.8	58.1	-351.8
Hexamer (B)	-2520.8	-10.8	-11.3	329.0	-2213.9
$\Delta = B - 6A$	-169.4	44.4	41.5	-19.6	-103.1

Table 3.3: Calculated chemical potential for the monomer, the hexamer and for the self-assembly reaction. All the electronic, translational, rotational and vibrational contributions are shown. All values are reported in kcal mol^{-1}.

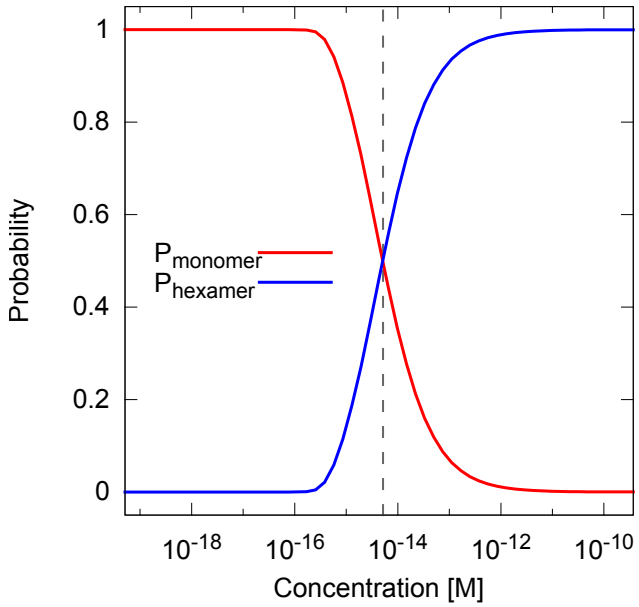


Figure 3.4: Molar fraction of the monomer and hexamer as a function of the initial concentration of monomers. The dashed gray line indicates the $C_{A,\frac{1}{2}}$, which delimits the domains of dominance of the monomer (at low concentrations) and of the hexamer (at high concentrations).

the fact that interactions with the solvent is completely neglected. This results in an over stabilization of the self-assembled state, which lowers the value of $C_{A,\frac{1}{2}}$ to non-physical concentrations. Unfortunately, including all enthalpic and entropic contributions from the solvent is computationally challenging and significantly more expensive than the energy minimization used for the evaluations of the chemical potentials.

As a last note, the vibrational quantum correction was evaluated. The calculated difference in vibrational chemical potential for the hexamerization reaction as calculated using the classical expression is -19.6 kcal mol^{-1}, as reported in Tab. 3.3. If the same frequencies are used in the quantum expression, the vibrational contribution drops to -15.1 kcal mol^{-1}, corresponding to a vibrational quantum correction of 4.5 kcal mol^{-1}. This value is extremely large as compared to the maximum of 1 kcal mol^{-1} presented in the original paper[140]. If this correction is included in the calculation of the $C_{A,\frac{1}{2}}$, a

new value of $\sim 2 \cdot 10^{-14}$ is calculated, which is four times larger than the value obtained with no quantum correction.

3.4 TAKE HOME MESSAGES

Two important results were presented in this chapter. First the equilibrium properties of a chemical reaction are defined by its stoichiometry. Moving from a one-to-one reaction like a conformational equilibrium, to a α-to-one reaction like self-assembly, the behavior of the system changes significantly. This is particularly evident when the effect of the concentration is analyzed: all α-to-one reactions are concentration dependent, while the one-to-one equilibrium is the only that is concentration independent. Second, the value of the equilibrium constant for a chemical reaction can be calculated from first principles using a statistical thermodynamics approach. However, this is done in the limit of rather strong approximations, among which the complete neglect of solvent effects. This has a major impact on the predicted concentrations, which are mostly non-physical. Nonetheless, the statistical thermodynamics framework presented here allows to correlate the value of the equilibrium constant with the molecular properties of the system components, opening in principle the rational design of the building blocks for self-assembly.

4

MODELING MOLECULAR SELF-ASSEMBLY AT SURFACES

Given the very appealing and promising features of molecular self-assembly in 2D described in the introduction of this thesis, the work presented here is not the only one aiming at a theoretical understanding of self-assembly. The most important problem first to be defined is what are the rules of the game. In particular, to study a chemical process it is necessary to understand if it is dominated by thermodynamics or by kinetics. Is a self-assembled monolayer a thermodynamic product, or a metastable state kinetically trapped? This question has been already asked in the literature[146,147]: some SAMs were shown to be at thermodynamic equilibrium with the supernatant solution, e.g. those obtained by dynamic covalent chemistry of bisimines[36], others were found as kinetically trapped states whose formation depends on both the experimental protocol and/or the temperature annealing. Due to the countless complications of studying kinetically trapped systems, the treatment is here limited to self-assembly under thermodynamic control, and thus no kinetic effects will be considered.

As we shall see in the next chapter, chemical reactions at equilibrium minimize the system free energy. Rephrasing the second law of thermodynamics, the total free energy of the system G can be written as sum of contributions, each coming from one system component. These contributions can be generally expressed as the amount of that component, like the number of moles N, multiplied by its molar free energy, or chemical potential μ_i. This brings to the following definition of the total free energy of the system

$$G = \sum_i N_i \mu_i \tag{4.1}$$

where the index i loops over all system components. Trying to rewrite this expression for a self-assembly process, it is necessary to list all components involved. First, considering a generic self-assembly involving a number of different monomers in solution, each of them will bring a contribution $N_i \mu_i$, as shown by Eq. 4.1. Second, the outcome of the self-assembly on a surface is a 2D layer. In this case it is more convenient to express the amount of product not by the number of molecules involved, but instead on the surface area A_j covered. The free energy contribution introduced by the SAM will be thus $A_j \gamma_j$, where γ is a free energy per surface area, which will be referred to as its surface free energy. Last, if the different formed SAMs do not cover the whole available surface, it is possible to expect a free energy term proportional to the free surface $(A_{tot} - \sum_j A_j)$

and a surface free energy γ_{surf} of the free surface. This decomposition brings to the definition of the total free energy of the system[35,148]

$$G = \sum_i \mu_i N_i + \sum_j \gamma_j A_j + \gamma_{surf} \left(A_{tot} - \sum_j A_j \right) \tag{4.2}$$

This result states that at chemical equilibrium molecular self-assembly at surfaces is simply defined by the combination of N_i and A_j parameters that minimize G. Unfortunately, solving Eq. 4.2 in a correct and general way is far from trivial, as demonstrated by the lack of studies using this equation directly.

One of these work, from Palma, Samorì *et al.*[148], used Eq. 4.2 to study the self-assembly of bicomponent systems by melamine and different ditopic imidic linkers. There, two limiting cases were drawn. First, at high concentration one SAM covers the whole surface, and the free energy is minimized by increasing γ as much as possible, that is the number of interaction per unit of area; thus highly packed SAM are favored. Second case, at low concentration of monomers, a few molecules adsorb on the surface and the covered area is a fraction of the total available surface. In this case the SAM that optimizes the interactions per molecule will be favored, thus allowing the formation of porous architectures.

A similar expression was used by Kampschulte, Lakinger *et al.*[35] to study the phase diagram of a mixture 1,3,5-benzenetribenzoic acid (BTB) and 1,3,5-tricarboxybenzene (trimesic acid, TMA). In that work, different SAMs were observed to form when different concentrations of TMA and BTB were used or by varying their relative concentration ratio. Interestingly, the experimental phase diagram could be correctly reproduced using an analogous of Eq. 4.2, and using the chemical potentials as free fitting parameters.

In the following, a number of different approaches to study the thermodynamics of self-assembly at surfaces are presented. As it will be shown, most of them are based on the calculation of a "free energy per surface area", which corresponds to the surface free energy γ.

4.1 EXPLOITING THE SURFACE FREE ENERGY

The most intuitive way to introduce the concept of surface free energy is to first consider a perfect 3D crystal. If a shear force is applied to it, the crystal will eventually cut along a plane creating two new non-interacting crystals, so that the result of the applied force is the formation of new exposed surface area. The reversible work done by the force can be expressed as the product between the amount of created surface area dA

and a certain quantity γ, which is characteristic of the crystal and the cutting plane. The variation in internal energy for the process can thus be expressed as

$$dU = TdS - pdV + \gamma dA \tag{4.3}$$

where TdS measures the exchanged heat and $-pdV$ the work done by the expansion/-compression of the system. This expression can be introduced into the definition of the Gibbs free energy ($G = U - TS + PV$) obtaining

$$\begin{aligned} dG &= dU - TdS - SdT + Vdp + PdV \\ &= \gamma dA - SdT + Vdp \end{aligned} \tag{4.4}$$

which, at constant pressure and temperature, simplifies to

$$dG = \gamma dA \tag{4.5}$$

It is thus possible to define the surface free energy of a crystal γ as the free energy required, per unit of surface area, to split the crystal in two along a plane[149]. Coming back to the original example, γ for a given crystal can be defined as the variation in free energy ΔG when a new surface area A is created

$$\gamma = \frac{\Delta G}{A} \tag{4.6}$$

Interestingly, the same approach can be used to evaluate the work to create an interface between, e.g. a liquid and a gas or two immiscible liquids[150,151]. In this case, dA is the amount of interfacial area produced, and γ is the surface (or interface) tension of the liquid. Following the derivation above, the surface tension γ can be defined in different ways, each associated with some partial derivative of the Gibbs free energy G in the isobaric-isothermal ensemble, the Helmholtz free energy F in the canonical ensemble, or the gran-potential Ω in the grand-canonical ensemble

$$\gamma = \left(\frac{\partial G}{\partial A}\right)_{NPT} = \left(\frac{\partial F}{\partial A}\right)_{NVT} = \left(\frac{\partial \Omega}{\partial A}\right)_{\mu VT} \tag{4.7}$$

All these definitions are analogous, and the use of one or the other only depends on which formulation makes it easier to treat the problem.

Ab-initio Thermodynamics of Chemisorbed Species

One of the first attempts towards the study of the functionalization of a surface has been the work of Reuter and Sheffler[152]. They modified the definition of the surface free energy to study the oxydation/reduction of the surface of a metal oxide MO_x as

a function of the partial pressure of oxygen. In this approach, sometimes referred to as *ab-initio* thermodynamics, the surface free energy is written as a function of the chemical potential of the system components as

$$\gamma(T,p) = \frac{\Omega}{A} = \frac{1}{A}\left[G_{slab} - N_M\mu_M - N_O\mu_O\right] \tag{4.8}$$

where G_{slab} is the free energy of the simulated surface, which can be differently oxidized, N_M and N_O are the number of metal and oxygen atoms in the system and μ_M and μ_O are the chemical potentials of the metal and of oxygen, respectively. The next step is to note that the chemical potentials of the bulk oxide, the metal and oxygen are not independent. Indicating the free energy per chemical formula of the bulk oxide as g_{MO_x}, it is possible to write $g_{MO_x} = \mu_M + x\mu_O$, allowing to recast the previous expression

$$\gamma(T,p) = \frac{1}{A}\left[G_{slab} - N_M g_{MO_x} + (xN_M - N_O)\mu_O\right] \tag{4.9}$$

which now depends only on the chemical potential of free oxygen and the chemical potential of the bulk oxide, which is a constant of the problem. Last, the chemical potential of the oxygen atom can be related to the chemical potential of molecular oxygen, which in turn depends on the oxygen pressure

$$\mu_O(T,p) = \mu_O(T,p^{\ominus}) + \frac{1}{2}kT\ln\frac{p}{p^{\ominus}} \tag{4.10}$$

where p is the pressure of the molecular oxygen. This last step allows to correlate directly the surface free energy with the pressure of the oxygen, which is the experimental quantity under control. Thus, Eq. 4.9 can be solved by evaluating the free energy of the slab G_{slab}, the one of the bulk oxide g_{MO_x} and the standard chemical potential of the oxygen atom $\mu_O(T,p^{\ominus})$. The first two can be approximated as the energy of one cell simulated at DFT level using periodic boundary conditions: for the first a cell of the oxidized surface is used, while for the second the bulk oxide is modeled. This was possible only after the evaluation of the vibrational contribution to the free energy, which was estimated to be negligible compared to the energy. For $\mu_O(T,p^{\ominus})$, this is obtained evaluating the energy of one oxygen molecule in vacuum at DFT level, and reporting it at the desired temperature and pressure thanks to the use of thermodynamic tables. The characteristic result of this treatment is a graph where the surface free energy for different oxidation states is plotted as a function of the oxygen chemical potential; see Fig. 4.1. This analysis produces straight lines whose slopes depend on the unit cell area and on the number of metal and oxygen atoms in the simulation cell, while the intercept depends essentially on the energy of the slab. This graph allows to identify in a very straightforward way which phase corresponds to the lowest surface free energy, that is the most favored structure. This approach have been used in a number of applications, ranging from the study of the surface properties of hematite[153], aluminum oxide[154],

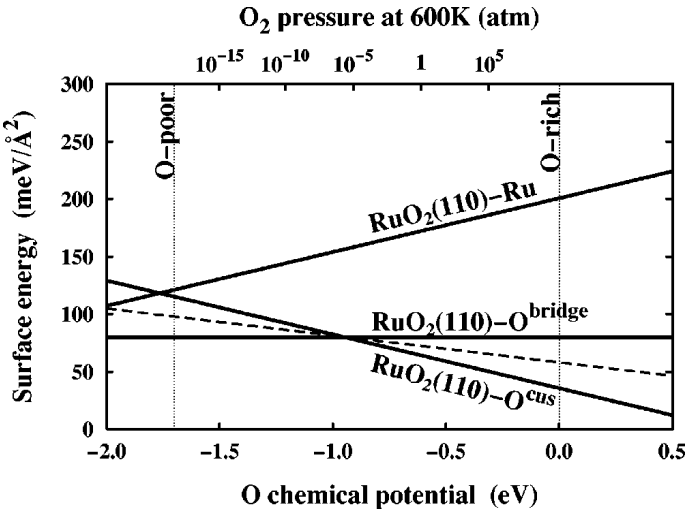


Figure 4.1: Example graph showing how the surface free energy changes as a function of the oxygen chemical potential for three different terminations of a RuO_2 (110) surface. Reproduce with permission from ref. 152.

ruthenium oxide[152], clean and Au/Pd covered magnetite[155], the catalytic properties of heterogeneous catalysts[156], and many more (the original paper by Reuter and Sheffler has been cited more than one thousand times).

Trying to use this approach to study the formation of a SAM composed by one molecule A per unit cell, Eq. 4.8 can be recast as

$$\gamma = \frac{1}{A_{sam}} [G_{sam} - \mu_A] \tag{4.11}$$

Making the assumption that the vibrational contribution is negligible ($G_{sam,vib} \simeq 0$), and considering only properties of the minimum unit cell, this expression can be further simplified as

$$\gamma = \frac{1}{A_{uc}} [E_{uc} - \mu_A] \tag{4.12}$$

Remarkably, this is the definition of γ used by Kučera and Gross[157] as well as the definition of the free energy of self-assembly in Meier, Ziener *et al.*[11] (with $\rho = 1/A'_{uc}$). In the first of these two works, Eq. 4.12 was used to study the chemisorption of 4-mercaptopyridine (Mpy) on Au(111). As chemisorption was involved, the unit cell energy was approximated as the adsorption energy of one molecule on gold calculated by DFT, thus neglecting all molecule-molecule interaction in the layer. Moreover, the chemical potential of the adsorbate could not be related to the concentration or pressure. In the second work, the same Eq. 4.12 was used to study the self-assembly of an oligopyridine (BTP) on Highly Ordered Pyrolytic Graphite (HOPG). In this case, the SAM was phisisorbed, making the contribution from the molecule-molecule interaction in the monolayer as important as the molecule-substrate interaction. Because of

the larger size of the BTP molecule, and the non commensurability of the SAM unit cell with the HOPG lattice, DFT could not be used to evaluate the unit cell energy, and a classical force field models for the interactions was used. As before, the surface free energy was studied as a function of the chemical potential of the monomer in solution, and no explicit relationship was made with the concentration of free monomers in solution, that is the experimental quantity under control. Moreover, while in Reuther and Sheffler the vibrational contribution to the free energy was shown to be negligible, this is not generally true, implying both graphs in Fig. 6 of Kučera and Gross and Fig. 9 of Meier, Ziener *et al.* can be significantly influenced by this approximation.

One example that goes beyond the above limits is the study of the work of Loffreda, Delbecq and Sautet[158] on the chemisorption of acrolein on Pt(111).. Few interesting modifications to the previous approach are here introduced. First, a zero point energy (ZPE) contribution is considered explicitly, even if the coupling with the surface phonons is neglected in the adsorbed state. Second, the chemical potential of acrolein in the gas phase was written as the product of a rotational (Z_{rot}) and a translational (Z_{tr}) partition function, which allows a direct connection with its partial pressure. These modifications yield the following expression for the surface free energy of the SAM

$$\begin{aligned}\gamma = &\frac{\theta}{A}\Big[\Delta E_{ads} - 3kT + \Delta ZPE + \\ &- N_{acro}kT \cdot \ln\left(\frac{1}{Z_{rot}Z_{tr}}\right) - N_{acro}kT \ln\frac{p}{p^{\ominus}}\Big]\end{aligned} \tag{4.13}$$

where Θ is the surface coverage, 3kT is the internal energy contribution from the rotations and the translations of the monomer and

$$\Delta E_{ads} = E_{surf+acro} - E_{surf} - E_{acro} \tag{4.14}$$

Interestingly, these modifications make the approach more complete, since now all the translational, rotational and vibrational degrees of freedom of the adsorbed molecule both in the gas phase and in the adsorbed states are explicitly considered. Surprisingly, the vibrational contribution was considered only as a zero point energy, and no vibrational entropy was included.

Physisorbed SAMs at Solid-Liquid Interface

Similar expressions for the surface free energy of the SAM have been used by Gutzler, Lackinger *et al.*[159] and Dienstmaier, Lackinger *et al.*[160] in the study of the temperature dependent opening and closing of nanopores, and the designing of nanoporous networks in tricarboxylic acids self-assembly. In these works, the self-assembly process is defined as the result of the interplay of an enthalpy gain due to adsorption on the surface and interaction among neighbors physisorbed molecules, and an entropy cost

due to the reduced number of degrees of freedom in the adsorbed state. This is made explicit when the free energy difference per unit of area on self-assembly is expressed as

$$\Delta g = \frac{\Delta G}{A} = \frac{\Delta H}{A} - \frac{T\Delta S}{A} \tag{4.15}$$

with A the surface area covered by one molecule in the SAM, ΔH is the energetic gain per molecule, and ΔS the entropy cost per molecule. Clearly, the free energy gain per surface area in Eq. 4.15 is equivalent to γ of Reuter and Sheffler.

Since these SAMs are too large to be studied at the DFT level of theory, the energy of adsorption and interaction were evaluated empirically using both experimental interaction energies and classical force field calculations. Interestingly, the entropy variation was directly evaluated using statistical thermodynamics. First, the total entropy difference was written as a sum of contributions arising from the lost translations and rotations (S_{tr} and S_{rot}), the decreased conformational freedom (S_{conf}) and a vibration term (S_{vib}).

$$\Delta S = \Delta S_{tr} + \Delta S_{rot} + \Delta S_{vib} + \Delta S_{conf} \tag{4.16}$$

Then, assuming that the molecules forming the SAM are rigid and that they completely lose all translational and rotational degrees of freedom on self-assembly, Eq. 4.16 yields

$$\Delta S = -\left(S_{tr} + S_{rot}\right) \tag{4.17}$$

where both S_{tr} and S_{rot} were evaluated analytically using standard statistical thermodynamics formulas[138]. To correct for the overestimation of the translational entropy due to the presence of solvent in the liquid phase, the concentration of the molecule was estimated based on the free volume of the solvent, which can be evaluated through the hard cube approximation[161].

Introducing the results of Eq. 4.17 into Eq. 4.15 the surface free energy of the SAM becomes

$$\gamma = \frac{1}{A}\left[\Delta H + T\left(S_{A,tr} + S_{A,rot}\right)\right] \tag{4.18}$$

Interestingly, this expression is analogous to Eq. 4.12, with the clear advantage that the chemical potential of the monomer is function of its translational and rotational entropy, which are directly related to the molecular properties and the monomer concentration in solution. In neither works ,γ was used to study 2D polymorphism as a function of the concentration, but was used to evaluate the surface free energy at the experimental conditions.

4.2 CHEMICAL EQUILIBRIUM BETWEEN A DENSE AND A POROUS ASSEMBLY

A different approach is the one introduced by de Feyter *et al.* to study the concentration and temperature dependent self-assembly of alkylated dehydrobenzo[12]annulene (DBA)[32,162] and alkoxy substituted 1,2,5-tristyrylbenzene (TSB)[163] at the liquid-HOPG interface. Both molecules were shown to associate in two different 2D architectures B and C, which could be selected by changing the concentration and/or the temperature. Considering the phase B to be more dense than C, the equilibrium between the two phases can be written as

$$bB \rightleftharpoons cC + (b-c)A \tag{4.19}$$

where b and c are the number of molecules of B and C per unit area. Eq. 4.19 states that for every b molecules of B converted in c molecules of C, $b-c$ molecules are released in solution. The stoichiometric coefficients b and c are proportional to the surface area covered by the SAM per molecule. At chemical equilibrium, the difference in chemical potential for the above reaction is zero, which yields the equilibrium condition

$$m\mu_B = \mu_C + (m-1)\mu_A \tag{4.20}$$

where m is b/c and μ_A, μ_B and μ_C are the chemical potentials of one molecule in the monomeric state A, or the self-assembled monolayers B, and C. To study this equilibrium, the chemical potential of the solution state (A) was written as a function of the monomer concentration

$$\mu_A = \mu_A^{\circ} + kT \ln C_A \tag{4.21}$$

and, by analogy, the chemical potentials of B and C were assumed to be proportional to the log of their surface coverage Y_i as

$$\mu_B = \mu_B^{\circ} + kT \ln Y_B \tag{4.22}$$

$$\mu_C = \mu_C^{\circ} + kT \ln Y_C \tag{4.23}$$

If so, Eq. 4.20 yields

$$\frac{Y_C}{Y_B^m} = K C_A^{(1-m)} \tag{4.24}$$

with

$$K = \exp\left[\frac{m\mu_B^{\circ} - \mu_C^{\circ} - (m-1)\mu_A^{\circ}}{kT}\right] \tag{4.25}$$

Assuming monolayer coverage ($Y_B + Y_C = 1$), the equilibrium constant K and the ratio m can be obtained by measuring Y_C at different initial concentrations C_A and fitting the linear expression

$$\ln \frac{Y_C}{C_A} = m \cdot \ln \frac{1 - Y_C}{C_A} + \ln K \tag{4.26}$$

Working with TSB, Bellec, Charra *et al.*[163] observed that this set of expressions does not work properly, as the predicted switching between the two phases B and C as a function of the concentration was slower than the experimental trend, which indicated the existence of a critical concentration C_0 at which a sharp phase transition occurs. To recover the correct behavior, it was noted that Eq. 4.20 does not consider the interconversion of large domains, but just the transfer of a single molecule, overestimating the entropy of the SAM. To correct for this, Eq. 4.20 was rewritten as

$$m\mu_{B,N} = \mu_{C,N} + (m-1)N\mu_A \tag{4.27}$$

where the formation of domains of size N was considered (as an approximation, only domains of *exactly* size N were considered), which yields

$$\left(\frac{Y_C}{Y_B^m}\right)^{1/N} = K_N C_A^{(1-m)} \tag{4.28}$$

with

$$K_N = \exp\left[\frac{(m/N)\mu^{\circ}_{B,N} - (1/N)\mu^{\circ}_{C,N} - (m-1)\mu^{\circ}_A}{kT}\right] \tag{4.29}$$

When these expressions are used to predict the surface coverage of the two assemblies as a function of the monomer concentration by varying the parameter N from one to one hundred, the curves become increasingly steeper, and reproduce the experimental results; see Fig. 4.2.

In addition, Bellec, Charra *et al.* developed an interesting expression for the critical concentration C_0, which corresponds to the switching concentration between the porous and the dense SAM, i.e. $Y_B = Y_C = 0.5$

$$C_0 = \exp\left(-\frac{S_0}{k}\right) \exp\left[-\frac{1}{kT}\left(H_B - \frac{H_C - H_B}{m-1}\right)\right] \tag{4.30}$$

where S_0 is the entropy cost at the critical concentration, and H_B and H_C are the enthalpy gain to form the B and C SAM.

A similar expression for the critical concentration which explicitly includes the entropy associated with solvent coadsorption was derived by Blunt, de Feyter *et al.*[32]. Characteristic of this work is the study of the equilibrium between a dense and a porous SAM as a function of the temperature, which provided definitions for both the critical

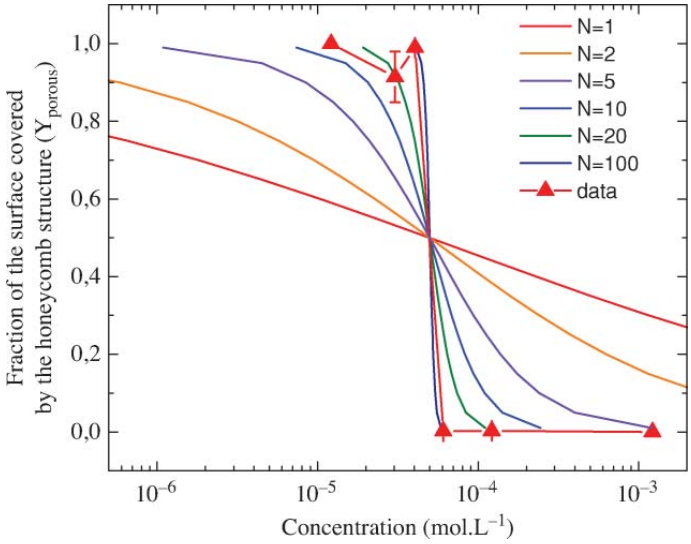


Figure 4.2: Surface coverage as a function of the concentration. The red points are the experimental values, while the different curves are the predicted surface coverages with different values for the domain size parameter N. Reproduced with permission from ref. 163 (Copyright 2011 AIP Publishing LLC).

concentration C_0 and temperature T_0. Plotting $\ln C_0$ as a function of $1/T_0$ at different initial concentrations, Eq. 4.30 was linearized, obtaining the enthalpy and the entropy components of the self-assembly free energy from the slope and the intercept, respectively.

These three works by Lei[162], Bellec[163] and Blunt[32] show how it is possible to obtain quantitative information for 2D self-assembly at thermodynamic equilibrium. The drawback is that this formulation is limited to the study of two self-assembled phases, and the extension to surface polymorphism is not immediate.

4.3 INTRODUCING THE EFFECTS OF THE SOLVENT

Thermodynamic Cycles to Study Chemisorption

A method that gives very similar results to the *ab initio* thermodynamic approach of Reuter and Sheffler, butis based on a completely different derivation, is the one described by Costa *et al.*, which was used to study the chemisorption of small molecules, like hydroxyl groups on γ-alumina surfaces[164], water on Cr_2O_3[165], halides on NiO_2[166], glycine on Cr_2O_3[167], diaminoethane on TiO_2[168] or glutamic acid on silver (100)[169].

Here, the free energy of adsorption at a given temperature and pressure/concentration is obtained via the thermodynamic cycle

$$\begin{array}{ccccc} S_{(sol)} & + & nM_{(sol)} & \xrightarrow{\Delta G_{liq}} & SAM_{(sol)} \\ {\scriptstyle \Delta G_1}\downarrow & & {\scriptstyle \Delta G_3}\downarrow & & {\scriptstyle \Delta G_5}\downarrow \\ S_{(T)} & + & nM_{(T,p)} & \xrightarrow{\Delta G_{gas}} & SAM_{(T)} \\ {\scriptstyle \Delta G_2}\downarrow & & {\scriptstyle \Delta G_4}\downarrow & & {\scriptstyle \Delta G_6}\downarrow \\ S_{(0K)} & + & nM_{(0K,vac)} & \xrightarrow{\Delta G_{ads}} & SAM_{(0K)} \end{array} \tag{4.31}$$

In this scheme, S indicates the surface, M the monomer and SAM the monolayer. The subscripts sol, T and $0K$ indicate that the specie is solvated, at a given temperature but in absence of solvent, and at zero kelvin, respectively. For the free molecule, the additional subscripts p and vac indicate the molecule at a given pressure/concentration, or alone in vacuum.

Reading the scheme from bottom to top, the first reaction is the chemisorption of one molecule in vacuum at zero kelvin to produce a SAM also at zero kelvin. The meaning of the corresponding ΔG_{ads}, is thus the free energy of adsorption of the molecule, but since it is calculated at zero kelvin, it is essentially the energy of adsorption ΔE_{ads} plus the zero point energy contribution. This quantity can be obtained straightforwardly by a DFT calculation in the optimized geometries of the three species. The middle reaction corresponds to the adsorption of the free molecule at a given pressure when the system is heated at a given temperature. To calculate this difference in free energy ΔG_{gas}, the value obtained for ΔG_{ads} can be used, together with ΔG_2, ΔG_4, and ΔG_6 which correspond to the thermal corrections for the zero-temperature values. These "corrections" are essentially a vibrational contribution for ΔG_2 and ΔG_6, while ΔG_4 contains both a vibrational contributions and a translational and rotational one for the free molecule. The top line includes the effect of solvation on the process. Unfortunately, this quantity is not easily accessible due to the high computational cost, and is determined assuming $\Delta G_1 = \Delta G_5$ and evaluating ΔG_3 using highly approximated expressions[168]. Finally, ΔG_{liq} is normalized by the surface area, providing a surface free energy.

Comparing this approach to the *ab-initio* thermodynamics of Reuther and Sheffler discussed previously, the main difference is how the self-assembly reaction is written here explicitly, and the final ΔG is obtained solving the corresponding thermodynamic cycle. Nevertheless, these two approaches finally give essentially the same expressions if the same approximations are used.

Assessing the Enthalpic Driving Force via Born-Haber Cycle

A remarkable use of thermodynamic cycles to study the formation of a 2D self-assembled monolayer phisysorbed on graphite was proposed by Song, Lackinger *et al.*[170], which used a Born-Haber cycle to access the total enthalpic force (ΔH) of 2D self-assembly. The typical thermodynamic cycle is reproduced in Fig. 4.3. The quantity of interest is $\Delta H_{sol \to monolayer}$ (the red arrow in the figure), which corresponds to the enthalpy change on moving one molecule from the solution to the SAM. This quantity cannot be accessed experimentally, or computationally, and is obtained by a four-steps cycle, whose ΔH can be measured and/or computed.

The first step is the dissolution enthalpy, $\Delta H_{sol \to crystal}$, which corresponds to the transfer of monomers from the crystalline state to the solvent used for self-assembly. This quantity can be determined experimentally by measuring the solubility of the molecule as a function of the temperature monitored e.g. by UV-Vis absorption spectroscopy. The second step is the sublimation enthalpy $\Delta H_{crystal \to vacuum}$, which corresponds to the enthalpy of removing one molecule from its crystalline packing. This enthalpy can be also determined experimentally, e.g. by measuring the effusion rate from a Knudsen cell in high vacuum conditions as a function of temperature. The effect of these two steps together is actually the desolvation of a molecule, moving it from the solution to the vacuum: $\Delta H_{sol \to vacuum} = \Delta H_{sol \to crystal} + \Delta H_{crystal \to vacuum}$. This term can also be determined computationally by evaluating the average interaction energy between one molecule and the surrounding solvent. Unfortunately, to best evaluate the solvation enthalpy, a time average of the molecule-solvent interaction energy has to be computed, which needs the sampling of a lot of configurations to converge to an accurate value. What is much more easy to evaluate by modeling is just the sublimation enthalpy, which is the lattice energy of crystal. In fact, knowing the crystal structure, it is possible to evaluate the lattice energy by single point calculation on the optimized crystal geometry, thus avoid expensive and slow converging molecular dynamic simulations and being limited only by the accuracy of the energy model used (a classical force field in this case). Doing so, it was observed the experimentally and computationally derived $\Delta H_{crystal \to vacuum}$ were in quantitative agreement, while the $\Delta H_{sol \to crystal}$, obtained by modeling as $\Delta H_{sol \to vacuum} - \Delta H_{crystal \to vacuum}$, was only in good qualitative agreement, due to high uncertainties in the evaluation of the solvation enthalpy.

The third and fourth steps are the binding in the monolayer $\Delta H_{monolayer \to vacuum}$ and the enthalpy of dewetting ΔH_{dewet}. The former corresponds to the energy required to decompose the SAM into molecular components in vacuum, and involves contributions from the molecule-molecule and the molecule-substrate interactions in the SAM, which can be accessed experimentally by TPD of the SAM in ultra high vacuum conditions, or, if high-resolution STM images of the SAM are available, they can be evaluated by modeling and single point energy calculations. The latter is the dewetting enthalpy, which

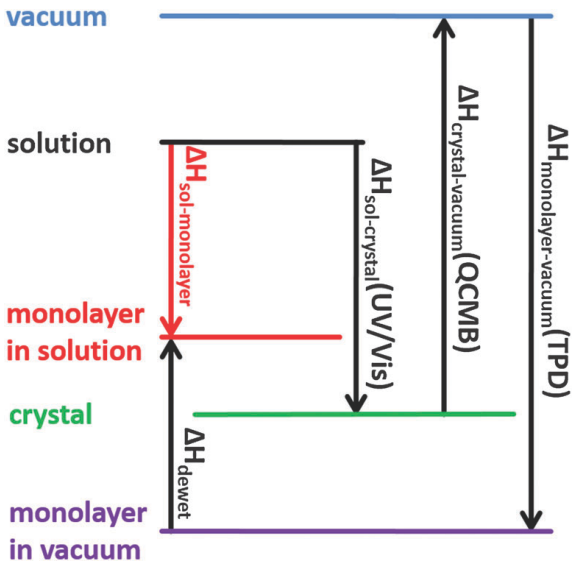


Figure 4.3: Born-Haber thermodynamic cycle used to access the enthalpic driving force for self-assembly at surfaces. Reproduced with permission from ref. 171.

includes both the desolvation of the substrate and the solvation energy of the SAM and is not experimentally accessible. Thus, this term is evaluated by modeling under two assumptions. First, in the absence of the SAM the solvent forms a self-assembled monolayer on the substrate. Second, in the presence of the SAM, the solvent forms a monolayer on the top of the SAM. If so, the dewetting enthalpy can be calculated as the difference between the cost of removing the solvent from the substrate ($\Delta H^{solv}_{graphite\rightarrow vacuum}$) and the gain of wetting the SAM on the top of the substrate ($\Delta H^{solv}_{monolayer\rightarrow vacuum}$) as

$$\Delta H_{dewet} = \left(\Delta H^{solv}_{graphite\rightarrow vacuum} - \Delta H^{solv}_{monolayer\rightarrow vacuum}\right) \frac{A_{monolayer}}{A_{solv}} \tag{4.32}$$

Both $\Delta H^{solv}_{graphite\rightarrow vacuum}$ and $\Delta H^{solv}_{monolayer\rightarrow vacuum}$ are accessed by modeling and are calculated using as reference one solvent molecules. Thus, the $A_{monolayer}/A_{solv}$ reweighing factor is used to move from a solvent-molecule to a solute-molecule property. This is an highly approximated expression for the dewetting term, but simple enough to be straightforwardly evaluated by modeling.

Overall, the enthalpic driving force of self-assembly is given by the sum of the four terms

$$\begin{aligned}\Delta H_{sol\rightarrow monolayer} = \Delta H_{sol\rightarrow crystal} + \Delta H_{crystal\rightarrow vacuum} \\ + \Delta H_{vacuum\rightarrow monolayer} + \Delta H_{dewet}\end{aligned} \tag{4.33}$$

At the critical concentration C_0, the monomeric and self-assembled states are in equilibrium, which means $\Delta G = 0$. This implies the entropy cost at C_0 can be evaluated by

$$\Delta S = \frac{\Delta H}{T} \tag{4.34}$$

where T is the temperature in kelvin and ΔH is known from the Born-Haber cycle. This methods has been successfully used to study the self-assembly of 1,4-benzenedicarboxylic acid (TPA)[170], 4,4'-stilbenedicarboxylic acid (SDA)[171], and hexabromotriphenylene (HBTP)[172] all three physisorbed on graphite.

Implicit Solvation Models

Typically, the solvation free energy of a molecule can be estimated using two different representations for the solvent: explicitly, like what described in the previous section describing the solvent in a full atomic detail, or implicitly, avoiding an atomistic representation of the solvent, and modeling its effect as a continuum. This second route is the one followed by Reimers, Crossley *et al.*[173,174] for the study of the self-assembly of tetraalkylporphyrins on graphite.

In these works, the images obtained by STM at different concentrations were analyzed to obtain the experimental free energy of self-assembly for the studied molecules. This is done writing the self-assembly equilibrium as

$$A + G \rightleftharpoons SAM \tag{4.35}$$

where A is the molecule that self-assembles, G the number of carbon atoms covered by the 2D architecture, which is proportional to the covered surface area per molecule, and SAM is the produced monolayer. The equilibrium constant is

$$K_{eq} = \frac{[SAM]}{[G][A]} = \frac{\kappa}{[A]} \tag{4.36}$$

where κ is the ratio between the covered and the total surface area per molecule, approximated to be around 10^2–10^3 looking at the percentage of disordered areas at the STM. From this expression, the free energy necessary to form a SAM, can be evaluated as

$$\Delta G_1^{\circ} = RT \ln \frac{[A]}{\kappa C^{\circ}} \tag{4.37}$$

with [A] being the minimal concentration promoting 2D self-assembly and C° the standard concentration (1 M). In case of 2D polymorphism, the standard free energy of self-assembly in a second distinct architecture can be obtained as a function of $\Delta G_1^s tst$.

In fact, assuming that chemical equilibrium is reached when the free energy differences normalized by the surface area ($\Delta G/A$) equalize

$$\Delta G_2^{\circ} = \frac{A_2}{A_1}\Delta G_1^{\circ} + \left(1 - \frac{A_2}{A_1}\right) RT \ln \frac{[A]}{C^{\circ}} \tag{4.38}$$

where $[A]$ is the experimental concentration of monomers at which both polymorphs are equally represented. This expression is analogous to the one derived by Bellec, Charra *et al.* by Eq. 4.30. The outcome of this analysis is the possibility to extract through Eq. 4.37 and 4.38 the ΔG of self-assembly from the experimental data.

Having an experimental estimation of the free energy difference, computational methods are used to reproduce the ΔG of self-assembly and understand the polymorphism of tetraalkylporphyrins as a function of the length of the alkyl chains. The free energy of self-assembly was calculated as sum of three terms: the energy of the SAM, a free energy correction, and a solvation free energy. For the first, the self-assembled monolayers were modeled using DFT or QM/MM to get the total molecule-molecule and molecule-substrate interaction. One particularity of these energy calculations is the recognition that the periodicity of the SAM is not commensurate with the periodicity of the underlying substrate. This complicates the DFT calculation which relies on a periodic simulation cell. The solution found was to simulate large supercells, up to the modeling of 12 to 15 cells in order to minimize the strain imposed by the forced symmetry[175]. The second contribution is a thermal free energy correction, in particular a vibrational free energy term coming from the phonon of the SAM. Last, an implicit solvent model based on the solvent-accessible surface arwa was used to account for the solvation free energy, the latter evaluated using the method of Floris *et al.*[176,177]

$$\begin{aligned} \Delta G_{solv} &= \Delta G_{solv,SAM} - \Delta G_{solv,P} - \Delta G_{solv,HOPG} \\ &= -\alpha\,(A_{SAM} - A_P - A_{HOPG}) \end{aligned} \tag{4.39}$$

where the proportionality parameter α was set to 0.0866 kcal mol^{-1} Å^{-2}. By comparing the experimental and the computed free energy differences, evaluated using different DFT functional and dispersion corrections, only a qualitative agreement was obtained.

4.4 TAKE HOME MESSAGES

In conclusion, a plethora of different theoretical methods to study self-assembly at surfaces exists in the literature. Some approaches are mainly devoted to the quantification of the self-assembly process, identifying and processing experimental quantities to access the relative stability of the formed SAMs. Clear examples are the works of Lei, Bellec, or Blunt. Other approaches try to study the self-assembly from a thermodynamic point of view decomposing the self-assembly free energy in energetic and entropic con-

tributions, and evaluating them using approximated analytical expressions. Examples in this class were derived from the *ab initio* thermodynamics approach of Reuther and Sheffler. The result of the intersection between these two approaches is the identification of few shared points. The most important is the reduction from SAM properties to "one molecule" properties, which implies that the stability of the SAM is related to the thermodynamic stability of one molecule inside the SAM. Although this may seem obvious, it is a theoretical results that is not clearly discussed in the literature. Second, most methods agree on the use of the surface free energy γ, even if the normalization of the free energy difference by the surface area is rarely justified. The third point is the existence of critical temperatures and concentrations, which define the stability boundaries between pairs of SAMs or with the monomeric state. Finally, although most methods were developed to provide a quantitative understanding of self-assembly, all of them may archive at best a qualitative agreement with the experiments. This major drawback motivates the development of more quantitative theoretical and modeling approaches.

5

A STATISTICAL THERMODYNAMICS APPROACH TO 2D SELF-ASSEMBLY

In the previous chapters it has been shown how it is possible to study the chemical equilibrium for a self-assembly reaction using a combination of modeling and statistical thermodynamics. In the following, that general framework is extended to provide a theoretical interpretation of molecular self-assembly in 2D. By introducing the concept of chemical potential per unit cell, a useful expression for the surface free energy of the SAM is derived, which can be numerically evaluated in the limit of the ideal gas approximation. This result provides an interpretation of the 2D polymorphism evidenced by scanning probe techniques and allows to quantify the critical aggregation concentration in 2D self-assembly. In the limit of the approximations introduced, this approach sets the ground for a first principles interpretation of molecular self-assembly at surfaces and interfaces.

5.1 THEORY

Finite-Size 2D Self-Assembly

Consider the spontaneous association (or self-assembly) of freely diffusing molecules in solution into a finite-size supramolecular object at the solid-liquid interface; see Fig. 5.1a. If we restrict ourselves to monocomponent systems at chemical equilibrium, this process is described by

$$\alpha A \rightleftharpoons B \tag{5.1}$$

where A is the monomeric specie, B the 2D self-assembled architecture, and α the molecularity of the reaction. Under chemical equilibrium conditions, the sum of the chemical potentials of the products equalizes that of the reactants (multiplied by their stoichiometric coefficients) such that for Eq. 5.1

$$\Delta\mu_{AB} = \mu_B - \alpha\mu_A = 0 \tag{5.2}$$

By expressing all chemical potentials as a function of their concentration in solution

$$\mu_i(V,T) = \mu_i^{\ominus}(T) + kT\ln\left(\frac{C_i}{C^{\ominus}}\right) \tag{5.3}$$

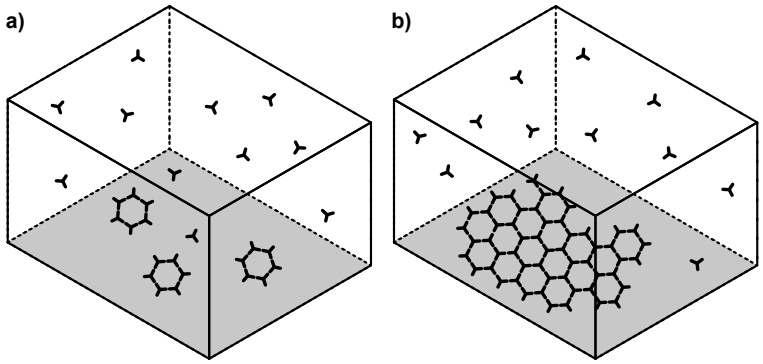


Figure 5.1: Schematic representation of the self-assembly at surfaces process. In a) the free molecules in solution are in equilibrium with finite-size object physisorbed at the solid-liquid interface. In b) the monomer in solution are in equilibrium with an extended periodic structure.

with T being the temperature, $C^{\ominus}$ the standard concentration, and $\mu_i^{\ominus}$ the standard chemical potential of the i-th component, Eq. 5.2 leads to the definition of the equilibrium constant K for the self-assembly reaction

$$K_{eq} = \frac{C_B \, (C^{\ominus})^{\alpha-1}}{C_A^{\alpha}} = \exp\left(-\frac{\Delta\mu_{AB}^{\ominus}}{kT}\right) \tag{5.4}$$

which provides a constraint on the equilibrium concentration of reactants and products at a given temperature. In the limit of idealized solution behavior, i.e. the particle independence ansatz, and at constant temperature T and volume V, the chemical potential can be expressed as

$$\mu_i(V,T) = -kT \ln \frac{q_i(V,T)}{N_i} \tag{5.5}$$

with q_i and N_i being the molecular partition function and the number of molecules of the i-th component, respectively[138]. By introducing the rigid-rotor, harmonic-oscillator (RRHO) approximation and the Born-Oppenheimer hypothesis, the partition function

of a freely diffusing monomer in vacuum can be written in closed form, which provides an approximated expression for the chemical potential of the monomers

$$\begin{aligned}\mu_{3D}(V,T) = &- kT \ln\left[\left(\frac{2\pi mkT}{h^2}\right)^{\frac{3}{2}} \frac{V}{N}\right] + \\ &- kT \ln\left[\frac{\sqrt{\pi}}{\sigma}\left(\frac{8\pi^2 kT}{h^2}\right)^{\frac{3}{2}} \sqrt{I_X I_Y I_Z}\right] + \\ &- kT \sum_{j=1}^{3n-6} \ln\left[2\sinh\left(\frac{h\nu_j}{2kT}\right)\right] + \\ &- E \end{aligned} \tag{5.6}$$

with n the number of atoms, m the molecular mass, I_X, I_Y, I_Z the principal moments of inertia, σ the symmetry number, ν_j the vibrational frequencies, and E the electronic energy of the ground state. Similarly, the chemical potential of a finite-size supramolecular object physisorbed on a surface is

$$\begin{aligned}\mu_{2D}(S,T) = &- kT \ln\left[\left(\frac{2\pi mkT}{h^2}\right) \frac{S}{N}\right] + \\ &- kT \ln\left[\frac{\sqrt{\pi}}{\sigma}\left(\frac{8\pi^2 kT}{h^2}\right)^{\frac{1}{2}} \sqrt{I_Z}\right] + \\ &- kT \sum_{j=1}^{3n-3} \ln\left[2\sinh\left(\frac{h\nu_j}{2kT}\right)\right] + \\ &- E \end{aligned} \tag{5.7}$$

where the volume V has turned into a surface area S, I_Z is the principal moment of inertia of the object about the axis perpendicular to the surface, the number of vibrational modes has increased to $3n-3$ as one translational and two rotational degrees of freedom are effectively converted into internal vibrations upon 2D confinement, and E includes the interaction energy with the surface. In the limit of the approximations above, Eqs. 5.6 and 5.7 provide numerical access to the standard chemical potentials difference for a finite-size 2D self-assembly.

Unit-Cell Chemical Potential

The approach presented above is generalized here to self-assembly into infinite, defect-free supramolecular architectures at the solid-liquid interface, which can still be described by Eq. 5.1. In this case and due to the non finite nature of B, major concern arises from the molecularity of the reaction, α, which is not defined. Exploiting the periodicity of the SAM, this problem is tackled by expressing the chemical potential of B

as a function of the free energy per unit cell, which, as we shall see, is a thermodynamic observable in the limit of infinite 2D architectures. Also, since the number of molecules per unit cell (n_{uc}) is architecture dependent, it is useful to consider the unit-cell free energy per molecule

$$\mu'_{uc} = \frac{\mu_{uc}}{n_{uc}} = \lim_{\alpha\to\infty} \frac{\mu_B}{\alpha} \tag{5.8}$$

Hereafter the superscript $'$ is used to indicate a quantity that is normalized by n_{uc}; note that given an architecture of n_{cells} unit cells, the relation $n_{uc} \cdot n_{cells} = \alpha$ holds. By deconvoluting the chemical potential of B into translational, rotational, vibrational and electronic contributions, the unit-cell free energy per molecule is analyzed in the limit of α going to infinity. Because the translational contribution grows logarithmically with the mass of the supramolecule (Eq. 5.7), which is a linear function of α as $m_B = \alpha m_A$, its contribution to the unit-cell chemical potential varies as $\ln(\alpha)/\alpha$ and goes to zero with increasing α. Thus,

$$\mu'_{uc,tr} = \lim_{\alpha\to\infty} \frac{\mu_{B,tr}}{\alpha} = 0 \tag{5.9}$$

A similar conclusion can be drawn for the rotational contribution, which increases logarithmically with the square root of I_Z (Eq. 5.7). In fact, approximating the 2D supramolecule with a solid disk of radius r_B, its moment of inertia is $I_Z = m_B r_B^2 = m_B A_B/\pi$ with m_B and A_B its mass and surface area. Since the surface area can be expressed as $A_B = \alpha A_A$ with A_A the surface area per monomer, the moment of inertia can be approximated as $I_Z = \alpha^2 m_A A_A/\pi$, which grows quadratically with α. If so, the rotational contribution to the unit cell chemical potential varies as $\ln(\alpha)/\alpha$, which goes to zero with increasing α

$$\mu'_{uc,rot} = \lim_{\alpha\to\infty} \frac{\mu_{B,rot}}{\alpha} = 0 \tag{5.10}$$

More involved is the evaluation of the vibrational contribution to the chemical potential of B. As mentioned above, the internal dynamics of a freely diffusing molecule (or supramolecule) in 2D is described by $3n-3$ vibrational modes, with n the number of atoms. Assuming that the internal vibrations of the monomers do not change upon self-assembly, i.e. the rigid-body ansatz, for each monomer addition to the SAM, two translational and one rotational degrees of freedom will be effectively transformed into three internal vibrations of the architecture. Thus

$$\begin{aligned}\mu_{B,vib} = -kT\alpha \sum_i^{3n-3} \ln\left[2\sinh\left(\frac{h\nu_i}{2kT}\right)\right] + \\ - kT(\alpha-1)\sum_j^{3} \ln\left[2\sinh\left(\frac{h\nu_j}{2kT}\right)\right]\end{aligned} \tag{5.11}$$

with the indexes i and j running over the internal and the external vibrations of the monomer in the self-assembled architecture, respectively; the latter term corresponding to the rigid-body oscillations and rocking. Clearly, Eq. 5.11 shows that the vibrational contribution to the unit-cell free energy does not vanish with increasing α. Rather, it converges to a finite value that depends on the chemical nature of the monomer and the architecture. Thus,

$$\mu'_{uc,vib} = \lim_{\alpha\to\infty} \frac{\mu_{B,vib}}{\alpha} \tag{5.12}$$

Finally, using a classical force field representation of interactions, which is pairwise additive, the electronic energy of B is $E_B = n_{cells} E_{uc} = \alpha E'_{uc}$, such that

$$\mu'_{uc,elec} = \lim_{\alpha\to\infty} \frac{\mu_{B,elec}}{\alpha} = E'_{uc} \tag{5.13}$$

which is independent of α. Moreover, since the energy per unit cell includes contributions from the interaction with the substrate (ads), the interaction between monomers in the monolayer (sam) and the strain energy upon 2D confinement ($strain$), E'_{uc} can be further decomposed as

$$E'_{uc} = E'_{ads} + E'_{sam} + E'_{strain} \tag{5.14}$$

Taken together, the results of Eqs. 5.9–5.13 show that the chemical potential per unit cell is

$$\mu'_{uc} = E'_{uc} + \mu'_{uc,vib} \tag{5.15}$$

whose value is finite and independent of α. Importantly, Eq. 5.15 can be numerically evaluated using an explicit molecular model of the SAM and the substrate. In the limit of sufficiently large supramolecular architectures ($\alpha \to +\infty$), Eq. 5.8 yields

$$\mu_B = \alpha \mu'_{uc} \tag{5.16}$$

which shows that the chemical potential of the SAM is a linear function of the chemical potential of one unit cell. Because the translational contribution to the unit-cell free energy vanishes with increasing α (Eq. 5.9), the chemical potential of an infinite-size 2D architecture is volume and thus concentration independent, which implies that $\mu'_{uc} = \mu^{\ominus\prime}_{uc}$.

Surface Free Energy of the SAM

The result of Eq. 5.16 enables one to express the chemical potential difference on self-assembly as a linear function of α. In fact, introducing the result of Eq. 5.16 into Eq. 5.2 yields

$$\Delta\mu_{AB} = \alpha\left(\mu'_{uc} - \mu_A\right) \tag{5.17}$$

which shows that the chemical potential difference on self-assembly is proportional to the reversible work of transferring one freely diffusing monomer in solution to the SAM. Although this does not solve the problem of the molecularity for an infinite-size self-assembly, it has important consequences. In fact, since α is related to the surface area A covered by the SAM by

$$\alpha = \frac{A}{A'_{uc}} \tag{5.18}$$

with A'_{uc} being the area of the unit cell per molecule inside the cell, Eq. 5.17 can be usefully recast in

$$\gamma = \frac{\Delta\mu_{AB}}{A} = \frac{1}{A'_{uc}}(\mu'_{uc} - \mu_A) \tag{5.19}$$

which is now independent of α. This fundamental result yields a numerically accessible expression for the surface free energy of the SAM, γ, which provides access to the thermodynamic stability of 2D self-assembly. Importantly, the use of γ addresses the molecularity issue by expressing the free energy of self-assembly per unit of surface area covered by the SAM.

Finally, in the limit of idealized solution behavior and the RRHO approximation, the result of Eq. 5.19 can be separated into energetic and entropic contributions, which provides fundamental insights on the key factors affecting the stability of the SAM. This can be obtained remembering that the chemical potential of the unit cell can be written as $\mu'_{uc} = E'_{uc} + \mu'_{uc,vib}$, while for the monomer in solution μ_A can be decomposed in translational, rotational and vibrational contributions $\mu_A = \mu_{A,tr} + \mu_{A,rot} + \mu_{A,vib}$. No energetic contribution for the molecule in solution is present, since, from the definition of the unit cell energy, the energy of the molecule in solution is the zero of the energy scale. Substituting these expressions into the definition of γ one obtains

$$\gamma = \frac{1}{A'_{uc}}\left(E'_{uc} + \mu'_{uc,vib} - \mu_{A,tr} - \mu_{A,rot} - \mu_{A,vib}\right) \tag{5.20}$$

where all chemical potential contributions can be expressed as sum of energetic and entropic terms

$$\gamma = \frac{1}{A'_{uc}}\left[\left(E'_{uc} + E'_{uc,vib} - E_{A,tr} - E_{A,rot} - E_{A,vib}\right) + \right.$$
$$\left. - T\left(S'_{uc,vib} - S_{A,tr} - S_{A,rot} - S_{A,vib}\right) \right] \tag{5.21}$$

As shown in chapter 3, the energy associated with each translational and rotational degree of freedom is RT/2. This allows to express γ as

$$\gamma = \frac{1}{A'_{uc}}\left[\left(E'_{uc} + \Delta E'_{vib} - 3RT\right) - T\left(\Delta S'_{vib} - S_{A,tr} - S_{A,rot}\right)\right] \tag{5.22}$$

which can be separated in energetic (γ_E) and entropic contributions (γ_S) by defining

$$\gamma_E = \frac{1}{A'_{uc}}\left(E'_{uc} + \Delta E'_{vib} - 3RT\right) \tag{5.23}$$
$$\gamma_S = \frac{1}{A'_{uc}}\left[-T\left(\Delta S'_{vib} - S_{A,tr} - S_{A,rot}\right)\right] \tag{5.24}$$

Last, if the vibrational contributions in Eq. 5.22 are neglected, a simplified expression for γ is obtained

$$\gamma = \frac{1}{A'_{uc}}\left[E'_{uc} - 3RT + T\left(S_{tr} + S_{rot}\right)\right] \tag{5.25}$$

which will be extensively used in the following chapters.

Since the chemical potential of the unit cell is concentration independent ($\mu'_{uc} = \mu^{\ominus\prime}_{uc}$), introducing the result of Eq. 5.3 into Eq. 5.19 yields a second expression for γ

$$\gamma = \gamma^{\ominus} - \frac{1}{A'_{uc}} kT \ln \frac{C_A}{C^{\ominus}} \tag{5.26}$$

which makes the dependence of γ on the concentration of monomers in solution as well as the density of surface packing (i.e. the inverse of the unit-cell area) explicit. Finally, when chemical equilibrium between the molecules adsorbed on the surface and those dissolved in the supernatant solution is established, $\gamma = 0$ and Eq. 5.26 yields

$$C_{cac} = C^{\ominus} \exp\left(\frac{A'_{uc}\gamma^{\ominus}}{kT}\right) \tag{5.27}$$

which indicates that in the limit of sufficiently large architectures there exists a critical concentration of monomers above which the self-assembled state is preferred.

5.2 THE METHOD OF THE SUPRAMOLECULES

The numerical evaluation of the surface free energy was carried out using a fully atomistic model of the monomer and the SAM. Due to the large number of atoms involved, up to 32000, geometry optimization was based on the classical General Amber Force Field (GAFF)[123], which was shown to accurately reproduce the binding energy of small molecules on graphite[178]. Although higher levels of theory, e.g. semiempirical quantum mechanics (SQM) or density functional theory (DFT), would be desirable, their use is hindered by the too intense computational cost for non-commensurate physisorbed monolayers[11]. The evaluation of γ is based on the numerical calculation of three quantities: the chemical potential of the monomer (μ_A), the chemical potential of the unit cell of the SAM (μ'_{uc}), and the area of the unit cell (A'_{uc}).

First, the chemical potential of the monomeric state (μ_A) was determined solving Eq. 5.6 on a fully optimized geometry of one monomer in vacuum. For this purpose, the vibrational frequencies ν_i were obtained by normal mode analysis, while the electronic energy D_e was set to be the zero of the energy scale. Because the result of Eq. 5.6 is concentration dependent, i.e. it depends on the volume V and the number of molecules N, the value of μ_A was calculated at a standard concentration of $C^{\ominus} = 1\,M$ and then evaluated at any other concentration via Eq. 5.3.

The chemical potential of the SAM (μ'_{uc}) was obtained from the numerical evaluation of Eq. 5.8 using the method of the supramolecules. This involves the modeling a series of increasingly larger finite-size architectures, whose chemical potentials in vacuum can be determined using Eq. 5.7, and evaluating Eq. 5.8 numerically in the limit of α going to infinity. For each value of α an atomistic model of the SAM was generated based on high resolution STM images and geometry optimized on top of a graphene layer to resolve steric clashes and find a potential energy minimum in proximity to the initial guess. Vibrational frequencies were obtained by a reduced basis normal mode analysis[179], where the atomic coordinates of the substrate are fixed in space and the normal modes determined in the mean field of the immobilized substrate. For the evaluation of the electronic contribution (E'_{uc}), it was found that Eq. 5.13 converges very slowly with the size of the supramolecule due to border effects, i.e. the large number of molecules sitting at the periphery of the model SAM. Actually, a significantly more efficient approach to E'_{uc} is provided by averaging the result of Eq. 5.14 over the inner (non peripheral) cells of the 2D architecture. As the internal energy of the monolayer (E_{sam}) can be decomposed into energy of interaction between the molecules inside the unit cell (E_{intra}) and energy of interaction with the surrounding cells (E_{inter})

$$E_{sam} = E_{intra} + \frac{1}{2}E_{inter} \tag{5.28}$$

with the factor 1/2 introduced to avoid double-counting[180], E'_{uc} can be straightforwardly obtained from a single optimization of a large SAM model (up to 200 molecules) followed by a series of single-point energy evaluations using e.g. the BLOCK facility in CHARMM[144], which switches on only the required interactions. Finally, the area of the unit cell (A'_{uc}) was obtained by averaging over the inner cells of the geometry-optimized 2D architecture.

As a last note, since the chemical potentials of SAM and monomers are evaluated in the limit of the ideal gas approximation, the calculated surface free energies are systematically affected by the absence of solvent, which leads to a non-realistic overestimation of the chemical potential of the monomer. As we shall see, the introduction of solvent corrections based on an approximated implicit solvent model (see Appendix B) demonstrates the impact of the solvent on the thermodynamics of 2D self-assembly.

5.3 RESULTS AND DISCUSSION

Chemical Potential of the Unit Cell

The first theoretical result to prove is the existence of a chemical potential per unit cell in 2D self-assembly. This goals implies showing the existence of the limit in Eq. 5.8 or equivalently that the translational, rotational, vibrational, and electronic contributions converge to finite values in the limit of infinite crystalline architectures.

For illustration, the self-assembly of trimesic acid (TMA, see Fig. 5.2A) on graphite in the characteristic chickenwire architecture (CHK)[33] was analyzed. As shown by Fig. 5.2B, this SAM is characterized by a hexagonal unit cell, which contains two TMA molecules paired by double hydrogen bonding interactions. These molecules, in turn, interact with the surrounding by forming eight additional hydrogen bonds, which are equally shared among the four neighboring molecules. Both the directionality and the strength of the intermolecular interactions make this SAM an ideal system to study 2D self-assembly.

Using the method of the supramolecules, larger and larger architectures were modeled to evaluate the chemical potential per unit cell of the SAM; see Fig. 5.2C. The various contributions to μ'_{uc} are shown in Fig. 5.2D with increasing α. As predicted theoretically, the translational and rotational contributions rapidly go to zero. The convergence of the vibrational contribution is somewhat slower but its variation with α has a total amplitude that is less than $1\,\mathrm{kcal\,mol^{-1}}$. Concerning the electronic contribution, two trends are shown in Fig. 5.2D, which correspond to Eq. 5.13 (red) and Eq. 5.14 (blue), respectively. As discussed previously, the evaluation of Eq. 5.13 is suboptimal and converges very slowly with increasing α. In sharp contrast, the result of Eq. 5.14 averaged over the inner unit cells of the 2D architecture (blue points) converges rapidly, although it does so to a significantly lower value. Interestingly, by fitting the red points in Fig. 5.2D

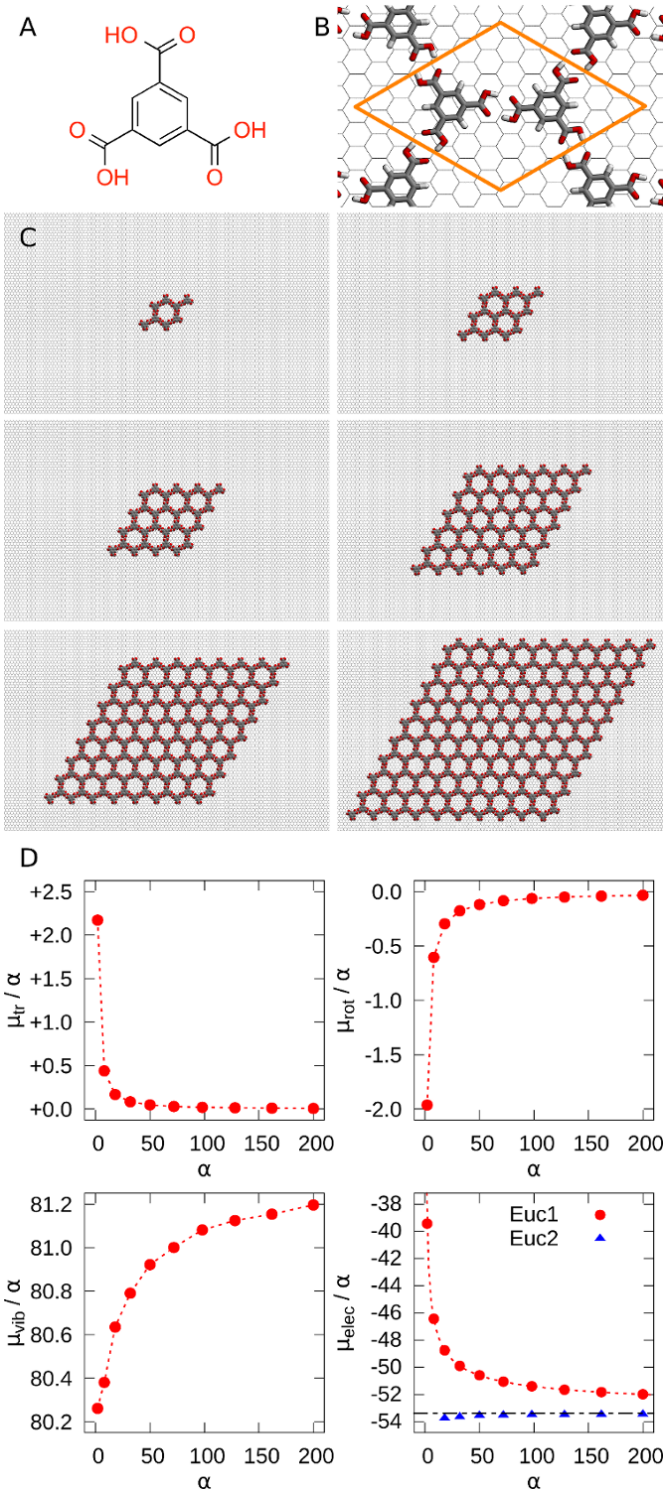


Figure 5.2: A) Chemical structure of trimesic acid (TMA, benzene-1,3,5-tricarboxylic). B) Unit cell of TMA in the chickenwire architecture. C) Six example supramolecules used to evaluate the limit of the chemical potential at increasing α (8, 18, 32, 72, 128, and 200). D) Translational, rotational, vibrational and electronic contributions to the total chemical potential for supramolecule formed by α TMA molecules when normalized by α. All chemical potentials are reported in kcal mol^{-1}.

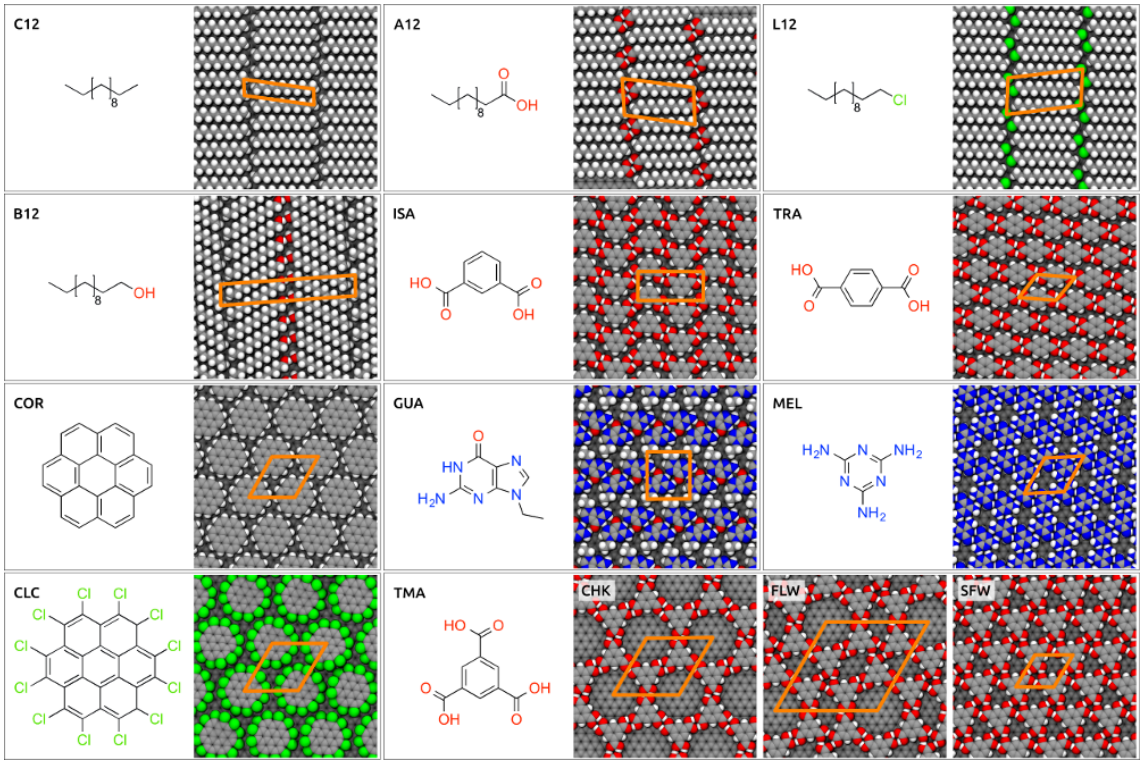


Figure 5.3: Modeled self-assembled monolayers and their chemical structures.

with $f(\alpha) = E'_{uc,limit} - b/\sqrt{(\alpha/n_{uc})}$, the value of the unit-cell energy in the limit of infinite architectures is equivalent to that provided by Eq. 5.14; see the Appendix A for a derivation of $f(\alpha)$.

Based on these results, the calculated energy per unit cell of the chickenwire SAM is $-53.5\,\mathrm{kcal\,mol^{-1}}$. By decomposing this energy in adsorption ($E'_{ads} = -25.7\,\mathrm{kcal\,mol^{-1}}$), association ($E'_{sam} = -29.1\,\mathrm{kcal\,mol^{-1}}$) and strain ($E'_{strain} = 1.3\,\mathrm{kcal\,mol^{-1}}$) contribution as done in Eq. 5.14, one finds that adsorption and recognition in the SAM introduce similar contributions, whereas molecular strain is negligible. Perhaps surprisingly, this indicates that although TMA is a relatively small molecule exposing strongly interacting recognition groups, the contribution from physisorption is of the same magnitude as the stabilization arising from molecular recognition. Finally, including the vibrational contribution per unit cell ($\mu'_{uc,vib} = 81.2\,\mathrm{kcal\,mol^{-1}}$) Eq. 5.15 yields a μ'_{uc} of $27.7\,\mathrm{kcal\,mol^{-1}}$.

To demonstrate the generality of the approach, the convergence of Eq. 5.8 was explored in thirteen chemically diverse SAMs on graphite. The choice of the substrate was just a matter of convenience as both the theoretical results and the computational methodologies are general and valid for any architecture physisorbed on a periodic substrate. The investigated SAMs include:

1. strongly interacting benzenecarboxylic acids, such as isophthalic acid (ISA)[12], terephthalic acid (TRA)[12], and trimesic acid (TMA)[33,181];

2. two basis, N^9-ethyl guanine (GUA)[182] and melamine (MEL)[183] whose assembly is steered by multiple hydrogen bonds;

3. a series of linear alkanes with chain length of twelve that are substituted at one terminus with one carboxylic group (A12)[184], one hydroxyl group (B12)[185], one chlorine group (L12)[186], or nothing (C12)[185–187];

4. two large poly-aromatic hydrocarbons, coronene (COR)[188,189] and perchlorocoronene (CLC)[189].

An illustration of the modeled SAMs, which span a wide range of molecular size, density of packing and energy of interaction, is given in Fig. 5.3. In all cases, the translational and rotational contributions to μ'_{uc} go to zero with increasing α, whereas the vibrational and electronic contributions converge to finite values; see Appendix D. These results demonstrate that in the limit of sufficiently large 2D architectures the chemical potential per unit cell is a thermodynamic observable, which exists independently of the chemical nature and the strength of molecular recognition in the SAM.

We note in passing that whereas the vibrational contribution to the chemical potential per unit cell is significant, its contribution to the chemical potential difference is often small; compare μ'_{vib} vs. $\Delta\mu'_{vib}$ in Tab. 5.2. Although this is not always true, e.g. for coronene and perchlorocoronene where it accounts for more than 10%, these calculations suggest that the vibrational contribution to γ is generally small even in physisorbed SAMs.

Concentration Dependent Self-Assembly

The existence of a chemical potential per unit cell (Eq. 5.15) provides straightforward access to the thermodynamic stability of the SAM through the evaluation of its surface free energy (Eq. 5.19). In fact, because γ is proportional to $\Delta\mu_{AB}$, 2D self-assembly will occur if and only if γ is lower than zero. In addition, since γ is architecture-dependent through the strength of molecular recognition in the monolayer (E'_{sam}) and the density of packing (A'_{uc}), its numerical evaluation allows one to compare the relative stability of different SAMs, which is useful to explore competitive equilibria at surfaces[189,190] and rationalize the concentration dependent 2D polymorphism[160,162,191].

To illustrate this aspect, the concentration dependence of trimesic acid (TMA) self-assembly was analyzed, which was shown by STM to form three distinct monolayers: (1) the porous hexagonal or chickenwire (CHK); (2) the slightly more dense flower (FLW); and (3) the densely packed superflower (SFW); see Figure 5.3. In all cases, 2D self-assembly is mediated by strong recognition events based on multiple H-bonding interactions, which promote the formation of linear dimers to form the essential unit of CHK, or trigonal trimers that are found both in FLW and SFW. Moreover, it was shown by STM on gold that under ultra-high vacuum conditions TMA self-assembly is strongly

	A'_{uc} [Å^2]	E'_{uc} [kcal mol^{-1}]	$\Delta\mu^{\ominus\prime}$ (1M) [kcal mol^{-1}]	γ [kcal mol^{-1} nm^{-2}] 1M	10^{-20}M	10^{-25}M
CHK	121.1	-53.5	-36.3	-30.0	-7.5	-1.8
FLW	98.2	-52.4	-35.1	-35.7	-8.0	-1.0
SFW	76.6	-50.1	-32.7	-42.7	-7.1	1.8
STR	117.0	-45.2	-29.3	-25.0	-1.7	4.1

Table 5.1: Stability of the three experimental (CHK, FLW, SFW) and one hypothetical (STR) assemblies of TMA. The first three columns contain the unit cell area, unit cell energy and the difference in chemical potential per molecule, respectively. In the last three columns the value of the surface free energy is reported at three different monomer concentrations.

Arch.	Monomer in solution				Self-Assembled Monolayers						Difference						
	$\mu^{\ominus}_{tr}$	μ_{rot}	μ_{vib}	$\mu^{\ominus}_{tot}$	E'_{sam}	E'_{ads}	E'_{strain}	E'_{uc}	μ'_{vib}	μ'_{uc}	$\Delta\mu'_{vib}$	$\Delta\mu^{\ominus\prime}$	A'_{uc}	$\gamma^{\ominus}$	γ_E	$\gamma^{\ominus}_S$	$\log C_{cac}$
ISA	-9.0	-8.4	75.7	58.3	-18.5	-20.8	1.0	-38.3	74.2	35.9	-1.5	-22.4	59.9	-37.4	-59.1	21.7	-16.3
TRA	-9.0	-7.5	75.7	59.2	-19.5	-20.8	0.9	-39.4	74.3	34.9	-1.4	-24.3	59.0	-41.2	-61.7	20.5	-17.7
GUA	-9.0	-8.4	101.1	83.7	-20.9	-23.5	1.5	-42.9	99.0	56.1	-2.1	-27.6	65.9	-41.9	-61.3	19.4	-20.1
MEL	-8.7	-7.7	66.0	49.6	-19.8	-16.3	2.3	-33.8	64.6	30.8	-1.4	-18.8	44.8	-42.0	-68.5	26.5	-13.7
COR	-9.5	-7.8	181.7	164.4	-1.4	-38.9	0.0	-40.3	176.1	135.8	-5.6	-28.6	114.2	-25.0	-33.5	8.5	-20.8
CLC	-10.3	-10.1	99.5	79.1	-6.6	-58.2	7.1	-57.7	94.4	36.7	-5.1	-42.4	163.9	-25.9	-33.7	7.8	-30.9
C12	-9.0	-8.3	214.4	197.1	-9.5	-25.2	0.1	-34.6	214.6	180.0	0.2	-17.1	77.5	-22.1	-41.3	19.2	-12.4
B12	-9.1	-8.9	215.2	197.2	-11.8	-26.6	0.2	-38.2	216.0	177.8	0.8	-19.4	81.2	-23.9	-43.6	19.7	-14.1
A12	-9.2	-9.0	202.8	184.6	-16.4	-28.0	0.9	-43.5	204.3	160.8	1.5	-23.8	81.8	-29.1	-49.6	20.5	-17.3
L12	-9.2	-9.0	207.7	189.5	-10.1	-27.7	0.1	-37.7	207.8	170.1	0.1	-19.4	82.8	-23.4	-42.4	19.0	-14.1
TMA/CHK	-9.2	-8.2	81.4	64.0	-22.4	-25.7	1.3	-46.8	81.2	34.4	-0.2	-29.6	121.1	-24.4	-35.8	11.4	-21.6
TMA/CHK*	-9.2	-8.2	81.4	64.0	-29.1	-25.7	1.3	-53.5	81.2	27.7	-0.2	-36.3	121.1	-30.0	-41.4	11.4	-26.5
TMA/FLW	-9.2	-8.2	81.4	64.0	-23.5	-25.7	1.3	-47.9	81.3	33.4	-0.1	-30.6	98.2	-31.2	-45.3	14.1	-22.3
TMA/FLW*	-9.2	-8.2	81.4	64.0	-28.0	-25.7	1.3	-52.4	81.3	28.9	-0.1	-35.1	98.2	-35.7	-49.8	14.1	-25.6
TMA/SFW	-9.2	-8.2	81.4	64.0	-25.8	-25.7	1.3	-50.2	81.4	31.2	0.0	-32.8	76.6	-42.8	-61.1	18.3	-23.9
TMA/STR	-9.2	-8.2	81.4	64.0	-15.9	-25.7	0.9	-40.7	79.9	39.2	-1.5	-24.8	117.0	-21.2	-32.3	11.1	-18.1

Table 5.2: summarizing all the modeled architectures and the calculated thermodynamic quantities. All energies and chemical potentials are reported in kcal mol^{-1}, the unit cell area in Å^2, the values of γ in kcal mol^{-1} nm^{-2}, the $\log C_{cac}$ is the decimal logarithm of the molar concentration. The asterisk (*) indicates that the interaction energy in the SAM is corrected to fit quantum data (see previous section).

dependent on the degree of surface coverage, with the porous CHK observed at low coverage, whereas the denser FLW and SFW stabilized at progressively increased coverage[181]. To explore the concentration-dependent polymorphism of TMA self-assembly, the three SAMs in Fig. 5.4 were modeled based on the corresponding STM images and their γ determined using Eq. 5.19 at 1 M concentration of monomers. Due to the apparent inability of the GAFF force field to reproduce the dimerization energy of TMA in vacuum relative to MP2 calculations, both the energy and the chemical potential of the unit cell per architecture were corrected *a posteriori* based on *ab initio* calculations of smaller molecular fragments (see Appendix C). The results (Table 5.1) show that the energetics actually favor CHK versus FLW and SFW by 1 and 3 kcal mol^{-1} respectively, due to the stronger interaction energy per molecule associated with the formation of linear versus trigonal recognition events. As the monomeric state of the three association reactions is the same, this trend is also reflected by the standard chemical potential difference per molecule, $\Delta\mu^{\ominus\prime}$, which favors CHK. However, when the density of packing is taken into consideration (Eq. 5.19), the SFW architecture corresponds to the lowest γ and is correctly predicted as the thermodynamic product. Interestingly, this indicates that at 1 M concentration of monomers the higher density of packing of SFW, which allows to dispose more monomers on a given surface area, provides the largest energy gain on self-assembly, despite this architecture is characterized by weaker recognition events. The situation is different at a lower concentration of monomers, where the formation of more porous architectures such as FLW, which involves less monomers per unit of area, is favored by the lower entropy cost of surface confinement. Indeed, at a monomer concentration of 10^{-20} M the calculations indicate that the FLW architecture is more stable by 0.5 and 0.9 kcal mol^{-1} nm^{-2} relative to CHK and SFW, respectively (Table 5.1). Analogously, at a monomer concentration of 10^{-25} M the porous CHK is predicted as the stable state. These results support the idea that the concentration-dependent 2D polymorphism arises from an energy/entropy compensation that favors porous architectures at low concentrations of monomers and densely-packed monolayers at saturating conditions.

The entropy/enthalpy compensation invoked above is quantitatively expressed by Eq. 5.26, which makes the concentration dependence of the free energy of self-assembly explicit. In fact, Eq. 5.26 shows that γ is a linear function of $\log C_A$ with a negative slope that is proportional to the density of packing ($1/A'_{uc}$). This implies that the thermodynamic stability of the SAM (relative to the disassembled state) increases with increasing the monomer concentration in the supernatant solution but also that more packed architectures are more sensitive to changes in the monomer concentration. Thus, Eq. 5.26 can be effectively used to compare the stability of different SAMs in the full concentration range and predict the domains of dominance of the various morphs. The results for TMA in vacuum are shown in Fig. 5.4. The data indicate that at high concentrations of monomers SFW corresponds to the lowest γ. However, because γ_{SFW} grows steeper than

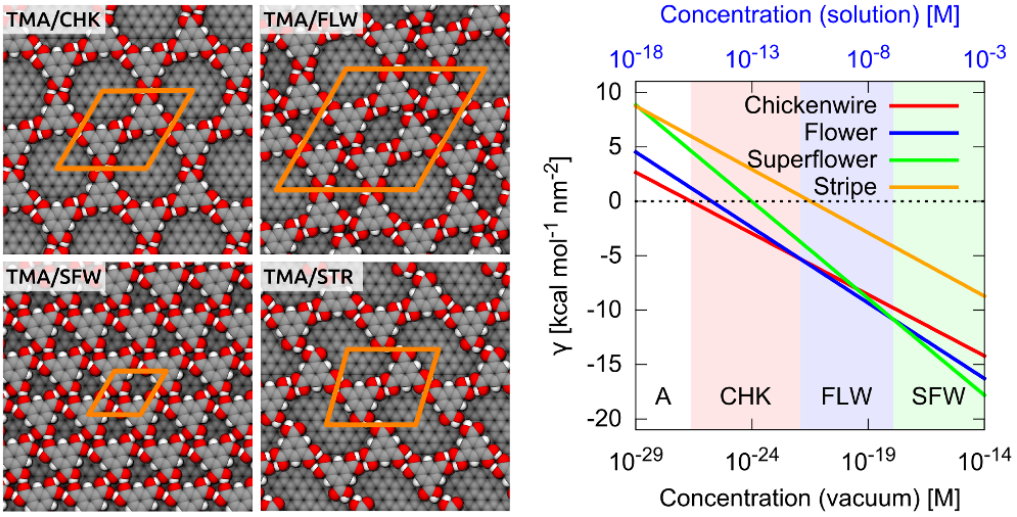


Figure 5.4: Surface free energy excess γ as a function of the monomer concentration for trimesic acid (TMA) self-assembly in the three different experimentally observed architectures – chickenwire (CHK), flower (FLW), and superflower (SFW) – and the hypothetical stripe (STR) architecture. The concentration dependency is evident, showing how at low concentration the disassembled state A is the favored, while at higher concentration the three assemblies CHK, FLW, and SFW are visible in order. The value of γ for the STR architecture is at all concentrations higher than the one of the others, indicating that architecture is never visible. The lower x axis (in black) indicates the concentration in vacuum, while the upper x axis (in blue) includes an approximation of the effect of the solvent (see Appendix B).

γ_{FLW}, there exists a critical concentration ($10^{-17.9}$ M) below which the less dense FLW is thermodynamically more favored with the SAM undergoing a 2D phase transition. Similarly, the existence of a second critical point corresponding to the phase transition to the porous CHK is predicted at a significantly lower concentration of monomers ($10^{-21.9}$ M). Finally, at extremely low concentrations ($10^{-26.5}$ M), there exists a third critical point below which the entropically stabilized monomeric state dominates and self-assembly does not occur. Since the critical concentrations predicted by our analysis are unphysically low, which is due to the absence of solvent in the calculations, a correction based on the solvent accessible surface area was introduced to account for the solvent effects (see Appendix B). Interestingly, the inclusion of an approximated solvation free energy contribution to the surface free energy of the SAM upshifts the critical concentrations by ten order of magnitudes thus yielding physically more reasonable results; see the upper x axis (in blue) in Fig. 5.4. Finally, by imposing chemical equilibrium between any two observed architectures B and C (i.e. $\gamma_B = \gamma_C$) Eq. 5.26 yields

$$C_{A,BC} = C^{\ominus} \exp\left[\frac{1}{kT}\left(\gamma_B^{\ominus} - \gamma_C^{\ominus}\right)\frac{A'_C A'_B}{A'_C - A'_B}\right] \tag{5.29}$$

which shows that the value of the switching concentration solely depends only on the unit-cell area and the standard surface free energy of the two SAMs. Remarkably, this

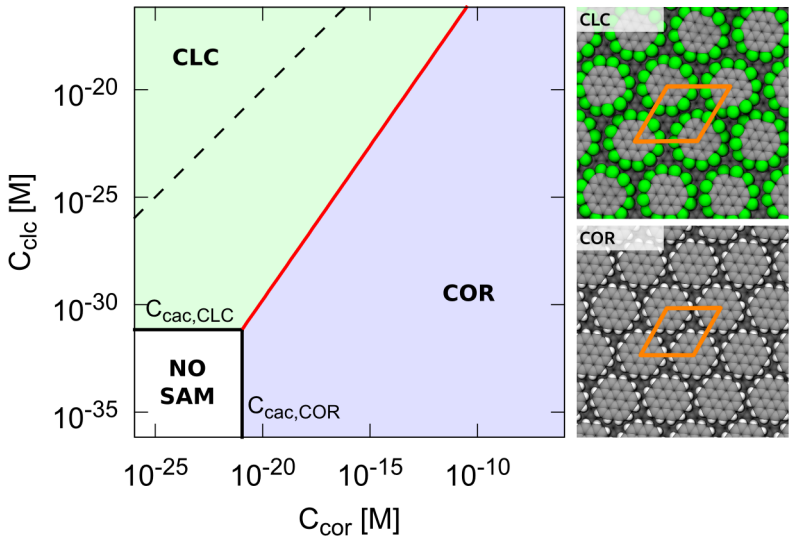


Figure 5.5: Phase diagram for the competitive self-assembly of coronene/perchlorocoronene mixtures. The two black solid lines indicate the critical aggregation concentrations, below which no self-assembly occurs. The red line represents the boundary where $\gamma_{COR} = \gamma_{CLC}$, which implies equal probability to observe one or the other SAM. Below the red line the preferred SAM is the one formed by coronene; over it, the preferred is perchlorocoronene. The dashed line indicate the locus where the concentration of coronene and perchlorocoronene are equal.

result is strikingly similar to the expression obtained by Bellec *et al.* using a completely different approach.

Competitive Equilibria at Surfaces

In addition to 2D polymorphism, Eq. 5.26 can be effectively used to rationalize competitive self-assembly equilibria at surfaces. To illustrate this aspect, the 2D self-assembly of coronene (COR) and perchlorocoronene (CLC) were compared. As will be discussed in chapter 9, a perchloro functionalization of coronene was found to enhance the 2D self-assembly propensity both energetically and entropically. Using Eq. 5.26, a phase diagram for the competitive self-assembly of these two molecules on graphite can be built by comparing the surface free energy of the corresponding SAMs as a function of the concentration of COR (x axis) and CLC (y axis) in solution; see Fig. 5.5. Using the condition $\gamma_{COR} = \gamma_{CLC}$ to define the phase boundary, which is a straight line a in log-log plot, the domains of dominance of coronene ($\gamma_{COR} < \gamma_{CLC}$) versus perchlorocoronene ($\gamma_{CLC} < \gamma_{COR}$) self-assembly can be identified. In addition, Fig. 5.5 shows that at very low concentration of monomers, neither γ_{COR} nor γ_{CLC} are negative and no SAM is formed. Interestingly, these results indicate that despite perchlorocoronene self-assembly is preferred at equal concentration of coronene and perchlorocoronene in solution, coronene self-assembly can be observed by changing the relative concentrations in solution. Consistently, several examples of how deterministic control over the

SAM can be achieved by playing with relative concentrations in mixtures have been reported in the literature[35,36].

Critical Aggregation Concentration

The last theoretical result we want to discuss is the existence of a critical aggregation concentration (C_{cac}) in self-assembly at surfaces and interfaces. As shown by Eq. 5.27, the theory predicts that at chemical equilibrium there exists a critical concentration of monomers above which self-assembly will occur. Because $A'_{uc}\gamma^{\circ}$ in Eq. 5.27 is equal to $\Delta\mu^{\circ\prime}$, which corresponds to the reversible work for transferring one monomer from the solution to the 2D architecture, the theory predicts that such a critical aggregation concentration: (1) is architecture-dependent through the strength of the interactions in the SAM (E'_{sam} in Eq. 5.14); (2) it is substrate-dependent through the strength of physisorption (E'_{ads} in Eq. 5.14); (3) it depends on the chemical nature of the solvent, whose interaction with the monomers modulates the value of μ°_{A}; and finally, (4) it is surface-packing independent. Interestingly, since the lower the C_{cac}, the higher the thermodynamic stability of the SAM relative to the monomeric state at standard conditions, C_{cac} sets an absolute scale of 2D self-assembly propensity, which can be used to compare chemically distinct and apparently unrelated events.

To illustrate this aspect, the critical aggregation concentration was predicted for the thirteen 2D architectures in Fig. 5.3. To this aim, the value of C_{cac} was calculated using Eq. 5.27 by modeling the SAM on a single-layer graphene and evaluating $\Delta\mu^{\circ\prime}$ based on GAFF force-field calculations in vacuum. The results in Fig. 5.6 show that the calculated C_{cac} span twenty orders of magnitude. Although the absolute values have little physical meaning, which is due to the absence of solvent in the calculations, the ranking indicates that the architectures corresponding to the lowest C_{cac} are those formed by perchlorocoronene (CLC) and trimesic acid (TMA), which share little chemical homology. In fact, while the former is a large polyaromatic hydrocarbon with highly polarizable groups that significantly enhance its physisorption on graphite[189], the latter is a small planar compound that steers the formation of a 2D SAM by strong H-bonding interactions in the monolayer. Similarly, coronene (COR) and N^9-ethyl guanine (GUA) are predicted to have similar C_{cac} despite being chemically diverse. Thus, these results indicate that similar 2D self-assembly propensities can be achieved by orthogonal chemical strategies.

An interesting comparison is provided by dodecane (C12) and derivatives obtained by chemical functionalization at one terminus via carboxy- (A12), alcohol- (B12), or chlorine- (L12) substitutions. Because these molecules produce compact architectures with comparable unit-cell areas, this analysis illustrates how the thermodynamic stability of the SAM may be modulated by chemical design. As shown by Fig. 5.6, the introduction of an alcohol or a chlorine group lowers the C_{cac} by two orders of magnitude relative to dodecane (C12), whereas chemical functionalization by carboxylic acid sub-

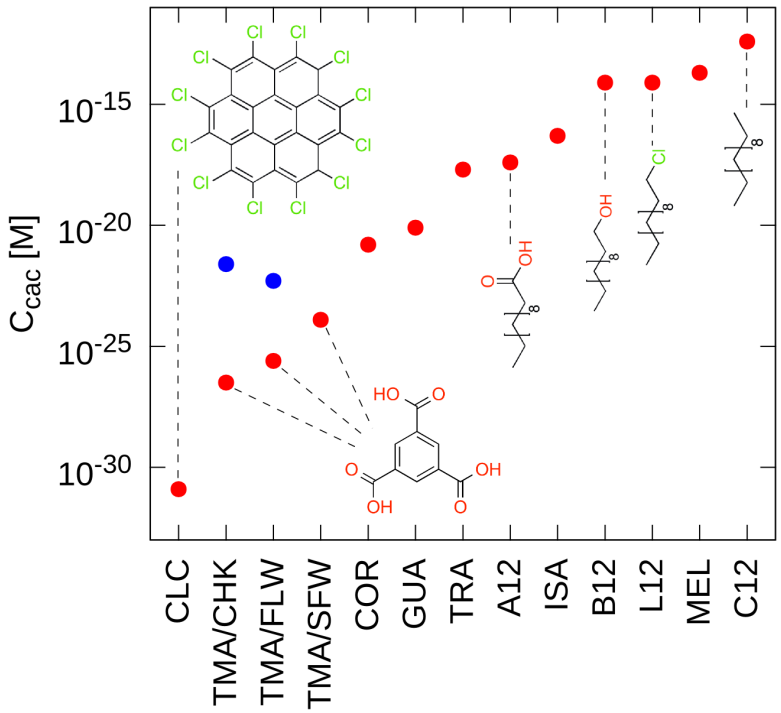


Figure 5.6: Critical aggregation concentration for the thirteen studied self-assembled architectures. For TMA, both the hydrogen bond uncorrected (blue) and corrected (red) values are reported.

stitution does so by five orders of magnitude. Also, the results indicate that the effect of chlorine substitution (L12), which enhances surface adsorption energy to C12, versus alcohol functionalization (B12), which strengthen molecular recognition in the SAM by H-bonding, produce similar effects. Taken together and in agreement with a recent report by us[189], this analysis suggests that chemical tailoring of the molecular components is a potentially useful strategy to modulate molecular self-assembly at surfaces and interfaces.

Towards Supramolecular Engineering

To fully understand the polymorphism observed in competitive equilibria at surfaces in interfaces, it is important to rationalize why alternative and energetically possible supramolecular arrangements are actually not observed. For this purpose, a previously unreported SAM by TMA was model by arranging monomers in a stripe-like (STR) fashion; see Fig. 5.4. In this architecture TMA is involved in two linear recognition events along the stripe and one weaker dipolar interaction across the stripes, which effectively corresponds to an out-of-register version of CHK. As shown by Table 5.1, this model SAM is energetically favorable and slightly more dense than CHK. Analysis of γ in Fig. 5.4 shows that unlike CHK, FLW and SFW, there exists no concentration of monomers for which γ_{STR} is lowest. Thus, despite the intrinsic energetic stability, there is

no thermodynamic reason to observe STR at chemical equilibrium. This interesting observation raises the fundamental question on the requirements for a given architecture to be observed in a concentration dependent STM experiment, i.e. where the concentration of monomers is progressively increased and self-assembly is imaged at the solid-liquid interface[160,162,191]. The results in Fig. 5.4 are quite instructive in this respect. In fact, since the slope of γ is proportional to the density of packing ($1/A'_{uc}$), they explain why the most dense SFW architecture is observed at high concentration of monomers. In addition, because the strength of the per-molecule interaction in the SAM dictates the value of the critical aggregation concentration, i.e. the intercept on the x axis, Fig. 5.4 shows that the emergence of more porous SAMs (like CHK or FLW) is observed at more diluted conditions only when self-assembly involves the formation of stronger recognition events, e.g. the substitution of trigonal (weak) with linear (strong) hydrogen bonds in TMA self-assembly. Finally, Fig. 5.4 shows that supramolecular patterns like STR, which are characterized by weaker interactions per molecule and intermediate densities of packing cannot be observed. Based on these observations, it is possible to conclude that supramolecular polymorphism at surfaces can be in principle predicted by comparing the energetics and the density of packing of different model architectures. By ranking the SAMs according to their standard chemical potential difference ($\Delta\mu^{\ominus\prime}$) in ascending order, the sequence of 2D phases can be predicted by searching for less and less energetically favorable but densely more packed architectures until the most dense is found. Interestingly, this implies that independently of the energetics, the densest architecture will always be observed at high concentrations of monomers, provided that this concentration is lower than the solubility in a given solvent. By contrast, porous architectures characterized by weakly interacting building blocks will be hardly imaged at surfaces.

5.4 TAKE HOME MESSAGES

Molecular self-assembly at surfaces and interfaces is a prominent example of self-organization of matter with outstanding technological applications. Achieving control over the equilibrium structure of the self-assembled monolayer (SAM) from first principle is key to the development of *bottom up* strategies in a number of fields. Following up the seminal work of Reuter and Scheffler[152], I have generalized their theoretical approach and presented a self-consistent framework based on modeling and statistical mechanics for a first-principle interpretation of 2D self-assembly. In the limit of the ideal gas approximation and the accuracy of the model of energetics in use (here molecular mechanics), this approach provides straightforward access to the thermodynamic stability of the SAM, and opens to a quantitative interpretation of the concentration dependence of 2D self-assembly, competitive equilibria at surfaces and interfaces, and 2D

polymorphism. The theoretical results emerging from this development are summarized below along with a discussion of their significance.

First, I demonstrated the existence of a chemical potential per unit cell of the SAM in the limit of infinite and defect-free architectures. This fundamental result, which is usually taken for granted[11,159,162], shows that the chemical potential per unit cell (and not per molecule) is a thermodynamic observable, which is dominated by the strength of molecular adsorption and recognition but it also includes a sizable vibrational contribution, which is usually neglected[11,159,161]. The existence of a chemical potential per unit cell provides a simple expression for the surface free energy of the SAM (Eq. 5.19), which provides numerical access to the thermodynamic stability of the SAM in the limit of the ideal gas and RRHO approximations. This expression of γ is the most important result of this study and provides a theoretical foundation to the *per molecule*, *per unit of area* approach of Lackinger and coworkers[159,160]. Moreover, since the vibrational contribution to the chemical potentials of SAM and monomers are of the same order of magnitude, it justifies the common assumption that the contribution of the vibrations to the free energy of self-assembly can be safely ignored[11,159]; note however that in the case of coronene or perchlorocoronene the vibrational contribution to γ is $>10\%$. Finally, in the limit of the introduced approximations, the expression of γ can be straightforwardly separated in energy and entropy contributions, which provides fundamental insight on the key factor controlling the stability of the SAM. In this context, we note that our derivation of γ provides correct expressions for both the energy and entropy components including e.g. the internal energy contribution of translations, rotations and vibrations, which was missing in previous work[159]. Finally, the newly introduced decomposition of the unit-cell energy (Eq. 5.14) provides a general strategy for its numerical evaluation, which is valid even when the molecules within the unit cell are not energetically equivalent.

Simple manipulations of Eq. 5.19 yielded a second important result, which makes the concentration dependence of the surface free energy of the SAM explicit (Eq. 5.26). In the limit of the model of energetics, this result enabled to study the competition between different SAMs formed by the same building blocks and rationalize the 2D polymorphism evidenced by scanning probe techniques with varying the initial concentration of monomers in the supernatant solution. Our analysis of trimesic acid self-assembly (Fig. 5.4) illustrates that the evaluation of γ per architecture is useful to predict the domains of dominance of the various polymorphs, to quantify the critical concentrations corresponding to 2D phase transitions, and rationalize why alternative and theoretically possible SAMs are not observed experimentally. The fundamental result in Eq. 5.26 shows that the thermodynamic stability of the SAM is strongly dependent on the density of surface packing, i.e. the number of molecules per unit of area, indicating that the higher the density, the stronger the sensitivity of the SAM to the concentration of monomers in solution. This result supports the idea that 2D polymorphism is the

manifestation of an enthalpy/entropy compensation, which favors (energy-stabilized) densely-packed architectures at high concentrations of monomers and (entropically-stabilized) porous architectures at diluted conditions[32,162]. Importantly, our analysis makes it clear that despite the densest SAM will be always observed at high concentration of monomers (provided that these concentrations are physical and lower than the monomer solubility) more porous architectures will form at diluted conditions if and only if their energetics per molecule are more favorable, i.e. if the SAM involves stronger recognition events. Thus, we conclude that by ranking a series of model SAMs based on their energy gain on self-assembly, the sequence of 2D morphs in a concentration-dependent STM experiment can be predicted by starting from the energetically more favorable and searching for increasingly denser molecular arrangements.

Finally, the theory predicts that at chemical equilibrium there exists a critical concentration of monomers (C_{cac}) above which 2D self-assembly is spontaneous. Our development provides a quantitative expression of C_{cac}, which is found to be architecture-, substrate-, and solvent-dependent. Since the lower the critical concentration, the higher the thermodynamic stability of the SAM relative to the monomeric phase, the value of C_{cac} sets an absolute scale of 2D self-assembly propensity, which may be used to compare chemically distinct and apparently unrelated events. Intriguingly, the existence of a critical aggregation concentration indicates that 2D self-assembly can be regarded as a "precipitation" in a solubility equilibrium, which does not stop until such a critical concentration of monomers is attained in solution. Thus, if the volume of the supernatant solution is small enough, the critical concentration will be reached before the entire surface is covered and self-assembly should stop at submonolayer coverage. Providing experimental evidence of a concentration-dependent degree of surface coverage, e.g. by STM imaging[160,162,191], would prove the existence of a C_{cac} in 2D self-assembly. More generally, the values of the C_{cac} and/or the switching concentrations in polymorphic 2D self-assembly, which are both computationally and experimentally accessible, open up to quantitative analysis of the phenomenon beyond imaging. Experimental determinations of these quantities will provide stringent benchmarks for available thermodynamic models and guide the development of future theoretical and modeling approaches.

A major shortcoming of current approaches is the lack of a proper treatment of the solvent, which results in non-physical values of the critical aggregation concentrations in 2D self-assembly. A pioneering attempt in this direction has been recently reported by Lackinger and coworkers, which provides experimental access to the enthalpy contribution to the solvation free energy by solving a modified Born-Haber cycle[170]. In the same spirit, approximated corrections to include solvent effects have been introduced based on implicit solvent models[173] or Molecular Dynamics simulations with an explicit treatment of the solvent[170]. Here, the inclusion of solvent correction based on a solvent accessible surface area model (see Appendix B) was shown to have dramatic effects on the predictions, i.e. up-shifting the critical concentrations for TMA self-assembly

by ten order of magnitude. This and previous[33] results support a pivotal role of solvent on the thermodynamics of self-assembly. The development of more quantitative models to predict 2D self-assembly beyond the ideal gas approximation is an important step forward, which is left for the future.

In conclusion, in this chapter it was presented a self-consistent framework for the theoretical interpretation of molecular self-assembly at surfaces and interfaces based on modeling and statistical thermodynamics. In the limit of the approximations introduced, this approach provides numerical access to the thermodynamic stability of the SAM from first principles, which is expected to aid the design of bottom-up approaches in a number of technological applications.

Part II

Applications

6

MODELING THE CHEMISTRY ON GRAPHENE: BENCHMARKING THE DESORPTION ENERGY

In chapter 2 at the beginning of this thesis, a lot of computational methods and approaches have been presented, each based on different approximations and modeling strategies. Due to this diversity, it is not possible to know *a priori* which method is most suited to study the chemistry on graphene, giving the best balance between accuracy of the results and efficiency of the calculations. The problem used for this benchmark is the calculation of the desorption energy of small molecules from a graphene layer. This is a very significant question, since adsorption at surfaces is the first step of a variety of processes, like molecular self-assembly at surfaces[9], catalysis[28], gas sensing[192,193], hydrogen storage[194], extraction of pollutants from chemical wastes[195,196], or the liquid-phase exfoliation of graphite[52].

Despite all these applications, the accurate determination of the interaction energy of a small molecule with a graphene layer is still a challenging problem, both experimentally or computationally. From the experimental point of view, the desorption energy can be obtained by Temperature Programmed Desorption (TPD) experiments in Ultra-High Vacuum (UHV) conditions[197]. Assuming a first order desorption kinetics, the interaction energy with the subtrate can be obtained by the analysis of different TPD spectra obtained at differ heating rates[198], or different initial substrate coverages[199]. Although these methods can give accurate results for small and weakly interacting molecules, this approach is indirect, time consuming, and special care has to be taken to consider the existence of non equivalent binding sites on the surface or the presence of adsorbate-adsorbate interactions[200].

All these problems can be solved by modeling, where the adsorption of a single molecule from a perfectly periodic and defect-free surface can be calculated. Ideally, most accurate results will be obtained by *ab initio* methods, among which the most expensive CCSD(T) methods can be possibly used. Unfortunately, due to the size of the molecule-graphene complex, the WFT methods are generally not useful. Only a few exceptions appeared so far in the literature, where small gases, like methane, have been modeled at these levels of theory[201,202]. Attractive in this respect is DFT, which generally allows accurate calculation at a much better computational time. Successful application of DFT have been recently reported[202–206], but the results are strongly dependent on the exchange-correlation functional[202] and the type of the dispersion correction[90,207]. Very recently, also the dispersion-corrected SQM methods were shown to be useful to study molecular adsorption on graphene[208]. These methods, in fact, scale significantly better than DFT and form a valuable class of methods. Last class of meth-

ods are the classical force fields. These have not been commonly used, probably due to the widespread skepticism on their accuracy, but they are suitable to explore systems involving thousands of atoms, making them unique in this respect.

In the following, several methods belonging to the three classes DFT, SQM, and FF, are tested and the calculated desorption energies from graphite in vacuum compared with TPD results for a set of small molecules. Most of the results presented have been published in The Journal of Physical Chemistry C[178].

6.1 REFERENCE DESORPTION ENERGY VALUES

To benchmark the performances of different computational methods in molecular adsorption problems, reference values for the molecule-surface interaction is needed. Among the computational methods, the reference must be looked for in very accurate *ab initio* methods, such as the most expensive coupled cluster methods with the largest basis sets. As mentioned previously, the problem is that if an explicit graphene layer is simulated, the number of atoms, and then the number of basis functions, to consider is too large. Thus, the only possibility is to look for accurate experimental data. In this work two sets of experimental desorption energies from Highly-Oriented Pyrolitic Graphite (HOPG) were used, which were obtained by TPD experiments in UHV conditions. The data include seven linear alkanes from methane to decane, some aromatic compounds and a few inert gases and solvents; see Table 6.1.

The first set was obtained from the work of Tait *et al.*[200], who measured the desorption energy of linear alkanes from graphite by TPD. To evaluate the desorption energies, for each molecule different TPD experiments were carried out at different initial coverage, and the obtained spectra were analyzed by the inversion-optimization method[199]. Data for longer alkanes (more than decane) are also available in the literature[198], but the desorption energies of these molecules are affected by the conformational entropy change at the desorption temperature, and are thus not considered in this work.

The molecules included in a second set introduce more chemical diversity in the benchmark. This set were measured by Ulbricht *et al.*[209] using a Redhead maximum-peak analysis[197], and include 23 compounds among which there are polyaromatic hydrocarbons, inert gases, solvents, and hydrogen bonding molecules. Not all molecules in this set were used in the benchmark. Methane were removed as already present in the first set, coronene and ovalene were removed due to the inconsistency between independent determinations in the literature[188,209,210], and also nitrogen dioxide was removed to avoid complications with the treatment of radicals. In total, 18 molecules were considered in the benchmark, whose experimental desorption energies are reported in Tab. 6.1.

Molecule	E_{des} [kcal mol^{-1}]	Ref.
	Linear Alkanes	
Methane	3.37	Tait *et al.* [200]
Ethane	5.88	Tait *et al.* [200]
Propane	7.67	Tait *et al.* [200]
Butane	9.75	Tait *et al.* [200]
Hexane	15.05	Tait *et al.* [200]
Octane	17.34	Tait *et al.* [200]
Decane	21.83	Tait *et al.* [200]
	Aromatic Compounds	
Benzene	11.52 ± 1.92	Ulbricht *et al.* [209]
Naphtalene	18.48 ± 2.16	Ulbricht *et al.* [209]
Toluene	16.32 ± 1.68	Ulbricht *et al.* [209]
Ethylbenzene	18.96 ± 2.4	Ulbricht *et al.* [209]
o-dichlorobenzene	16.56 ± 1.44	Ulbricht *et al.* [209]
	Gas and Solvents	
Methanol	11.52 ± 0.72	Ulbricht *et al.* [209]
1,1-dichloroethane	12.24 ± 0.72	Ulbricht *et al.* [209]
N,N-dimethylformamide (DMF)	12.72 ± 0.96	Ulbricht *et al.* [209]
Ethanol	12.00 ± 0.72	Ulbricht *et al.* [209]
Water	11.04 ± 0.72	Ulbricht *et al.* [209]
Ammonia (NH_3)	6.00 ± 0.48	Ulbricht *et al.* [209]
Trichloromethane ($CHCl_3$)	12.96 ± 0.72	Ulbricht *et al.* [209]
Nitrogen (N_2)	3.12 ± 0.24	Ulbricht *et al.* [209]
Carbon dioxide (CO_2)	5.76 ± 0.48	Ulbricht *et al.* [209]
Carbon Monoxide (CO)	3.12 ± 0.24	Ulbricht *et al.* [209]
Oxygen (O_2)	2.88 ± 0.24	Ulbricht *et al.* [209]
Sulphur hexafluoride (SF_6)	7.44 ± 0.48	Ulbricht *et al.* [209]
Xenon (Xe)	5.76 ± 0.48	Ulbricht *et al.* [209]

Table 6.1: Experimental desorption energy values used as benchmark in this work. All data were obtained by TPD in UHV condition from a graphite substrate.

6.2 COMPUTATIONAL METHODS

The desorption energy for all molecules in Tab. 6.1 were computed using DFT, SQM, and FF methods. In general, the desorption was modeled as a barrier-less process, which provides the value of the desorption energy by subtracting the energy of both the single molecule (E_a) and the graphene substrate (E_s) from the energy of the complex ($E_{a/s}$) as

$$E_{des} = -\left[E_{a/s} - (E_a + E_s)\right] \tag{6.1}$$

The graphene substrate was modeled as a large polyaromatic hydrocarbon of different size depending on the method used; see Fig. 6.1. In practice, the substrate model was first allowed to relax in the xy plane by fixing the position of the z coordinates to zero. Then, the adsorbate was placed at the center of the substrate and deeply energy minimized to find the best adsorption geometry. The z coordinate of all graphene atoms were always fixed to zero. The energy obtained is the energy of the complex $E_{a/s}$.

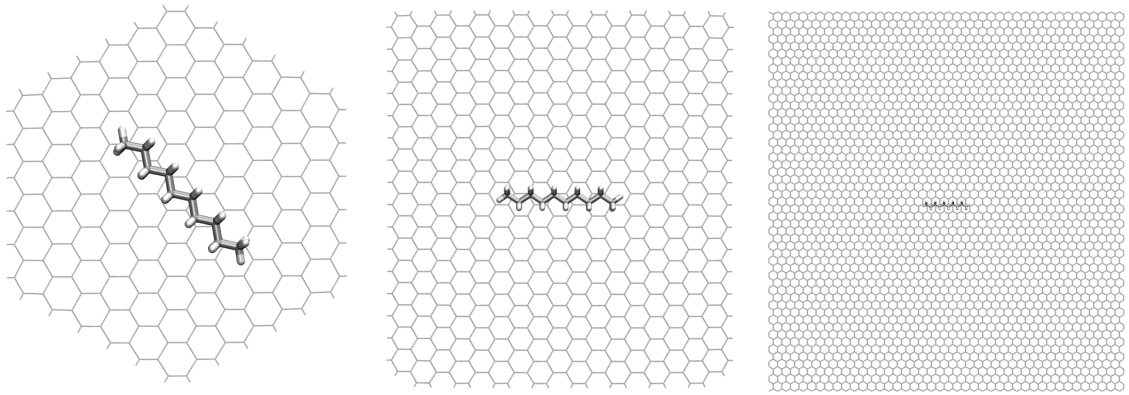


Figure 6.1: Models for the graphite substrate used in this work. On the left, the 216 hexagonal model used for the SQM and DFT methods. On the center, the 610 carbon atoms model used for the CGenFF force field. On the right, the 103 x 110 Å periodic graphene layer used for the GAFF and OPLS-AA force fields. The model used for the MMFF94 force field is similar to this last, but with double size (206 x 216 Å). As reference, the biggest molecule used in this work (decane) is represented at the center of each model.

Then, the adsorbate was separated from the substrate, and their energies (E_a and E_s) determined by a single point energy calculation (no geometry optimization involved) in vacuum. Having these three values, the desorption energy was evaluated using Eq. 6.1. To keep the procedure as general as possible, no particular care was used to find the best possible adsorbed geometry of the molecule. Due to the small molecular size of the adsorbate, the error introduced by this approximation was estimated to be smaller than experimental uncertainty.

From the calculated values, the accuracy of each method was evaluated analyzing the correlation with the experimental desorption energies. Specifically, the values of the slope, the correlation coefficient R^2, and the intercept of the linear correlation can be used to quantify the accuracy, the precision, and the systematic error of the method. To rank the methods based on their predictivity, the Root Mean Square Error (RMSE) was calculated

$$RMSE = \sqrt{\frac{1}{N}\sum_{i}^{N}\left(E_i^{calc} - E_i^{exp}\right)^2} \tag{6.2}$$

where N is the number of calculated desorption energies, and E_i^{exp} and E_i^{calc} are the experimental value and the computational prediction for the i-th compound, respectively.

In the DFT family, the ωB97X-D dispersion corrected functional[87] was used in combination with the 6-31G(d) basis set, which was shown to accurately describe dispersion-dominated systems[81]. Because DFT calculations on these systems are computationally very expensive, no other DFT functional or basis sets were tested. To model the graphite substrate, a hexagonal graphene layer of 216 carbon atoms with dangling hydrogens was

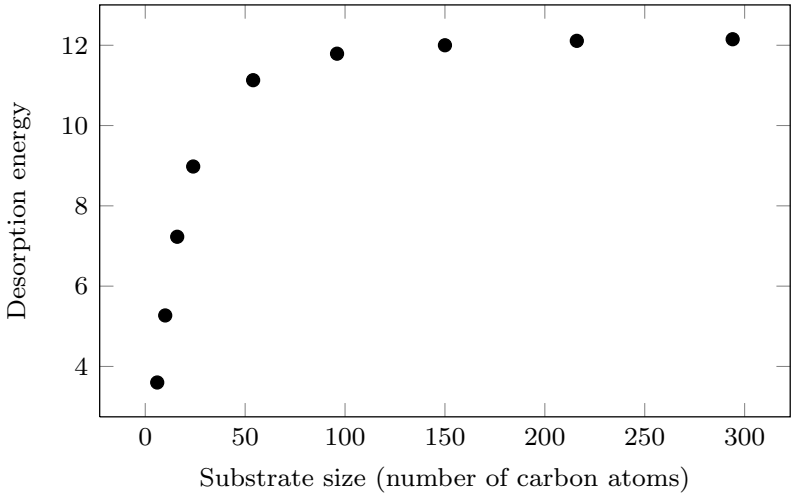


Figure 6.2: Desorption energy for benzene as a function of the size (measured as number of carbon atoms) of the graphene substrate calculated using the PM6-DH+ semiempirical method. The desorption energy essentially converge after 150 carbon atoms, while the used model contains 216 atoms.

used; see Fig. 6.1. Finally, to correct for the BSSE, the counterpoise method was used[93]. All calculations were performed using Gaussian 09[211].

In the SQM family, ten methods were tested: AM1, AM1-D3H4/AM1, RM1, RM1-D3H4/RM1, PM6, PM6-DH2/PM6, PM6-DH+, PM6-D3H4/PM6, PM6-D3H4/PM7, and PM7. All these methods are available in MOPAC2012[212], except for the D3H4 correction, where the Rezac H4[110] and the Grimme D3[90] corrections were applied *a posteriori*; note that methodA/methodB indicates that the geometry optimization was performed using methodB, whereas the final energy was evaluated with methodA. This is used for the DH2 correction, which presents discontinuities in the energy gradient and can not be used for geometry optimization[108], and for the D3H4 correction, which was not available on MOPAC2012 (newer MOPAC2012 versions now include natively the D3H4 correction). Similar to DFT, the substrate was modeled as a graphene layer of 216 carbon atoms; see Fig. 6.1. To verify that the size of the substrate is enough to provide accurate estimates, the desorption energy of benzene from graphene models of different size was calculated. As shown in Fig. 6.2, the interaction energy converges with a substrate of 150 carbon atoms.

The last class of methods tested is the empirical force field. For these calculations, automatic procedures were used to generate the topology of the adsorbates, thus neglecting molecule-specific parametrizations. Four FF were used: MMFF94, CGenFF, GAFF, and OPLS-AA. For MMFF94, the atom typing was perfomed directly by CHARMM (version 36b2)[144] reading a Sybyl mol2 file. The model substrate used was a 206 x 216 Å graphene layer with periodic boundaries, where each carbon atom was parametrized as a type 2 (generic sp2 carbon) atom with zero charge. The adsorbate/substrate complex was optimized using 1000 steps of steepest descent (SD) followed by an adopted

basis Newton Raphson (ABNR) optimization until convergence to an energy gradient of 10^{-8} kcal mol^{-1} Å^{-1}. For CGenFF, atom typing was performed by the ParamChem webserver (version 0.9.7)[213,214], which generates parameters for CGenFF version 2b8. The substrate was modeled as a graphene layer of 610 carbon atoms with dangling hydrogens, where carbons were parametrized using the atom type CG2R61 (aromatic carbon) with zero charge if internal and -0.115 e.u. at the substrate edges, and hydrogens were parametrized with the atom type HGR61 (aromatic hydrogen) with a partial charge of +0.115 e.u.; see Fig. 6.1. The calculation of the desorption energy was performed as for MMFF94 using CHARMM. Atom typing for GAFF was performed using the ANTECHAMBER[124] software distributed with AmberTools aided by the ACPYPE[143] script, while the desorption energy was calculated using GROMACS 4.5.5[215–217] compiled in double precision. The substrate was modeled as a 2D 103 x 110 Å periodic graphene layer, see Fig. 6.1, with carbon atoms modeled with the atom type CA (sp2 aromatic carbon). The adsorbate/substrate complex was optimized using the l-bfgs method (a quasi-Newtonian algorithm based on the low-memory Broyden-Fletcher-Goldfarb-Shanno approach)[218,219] until convergence to an energy gradient of 10^{-8} kcal mol^{-1} Å^{-1}. Finally, for OPLS-AA the atom typing was performed using the MKTOP script[220], which produces topologies in GROMACS format. The graphene substrate was a 103 x 110 Å layer with periodic boundaries modeled with the carbon atom type opls_147 (naphtalene fusion carbon); see Fig. 6.1. The desorption energy was evaluated as done for GAFF.

6.3 BENCHMARKING

Using the described computational methods, the desorption energy of all compounds in Tab. 6.1 were calculated and compared to the experimental values. Starting with the SQM methods, the oldest AM1 and RM1 methods gave poor results, with equilibrium distances between the molecules and the substrate between 4.5 and 5.0 Å, and interaction energies close to zero; see Tab. 6.2. Clearly, these methods are not able to account for the dispersive interaction that dominate the formation of an adsorbate/substrate complex. Even when the D3H4 correction was applied *a posteriori*, no significant improvement was shown; see Tab. 6.2. These methods were thus no further considered. Quite different were the results of the PM6 method. Although the absolute desorption energies were one order of magnitude off, the predicted complex were much more realistic with an average adsorbate/substrate distance of about 3.4 Å; see Tab 6.2. Thanks to the good geometries, the *a posteriori* DH2 and D3H4 corrections to PM6 provided accurate results, in quantitative agreement with the experimental data, as much as the native PM6-DH+ method; see Tab. 6.3. Consistent with what was previously reported[221], the use of the last generation PM7 yielded inaccurate results. Although the optimized geometries were correct, the predicted interaction energies are clearly overestimated. This

	AM1	AM1-D3H4/AM1	RM1	RM1-D3H4/RM1	PM6	Exp.
			ALKANES			
Methane	0.02	0.53	0.03	0.76	0.45	3.37
Ethane	0.04	0.79	0.04	1.20	0.70	5.88
Propane	0.05	1.07	0.06	1.66	1.04	7.67
Butane	0.07	1.43	0.08	2.13	1.20	9.75
Hexane	0.09	2.08	0.11	3.06	1.77	15.05
Octane	0.12	2.72	0.15	4.07	2.55	17.34
Decane	0.15	3.31	0.18	5.02	3.07	21.83
			AROMATIC COMPOUNDS			
Benzene	0.15	2.68	0.18	3.52	1.44	11.52 ± 1.92
Naphtalene	0.16	4.04	0.23	5.37	2.07	18.48 ± 2.16
Toluene	0.15	2.94	0.19	3.99	1.65	16.32 ± 1.68
Ethylbenzene	0.12	2.49	0.16	3.49	1.93	18.96 ± 2.4
o-dichlorobenzene	0.32	3.53	0.37	5.44	2.50	16.56 ± 1.44
			GAS AND SOLVENT MOLECULES			
Methanol	0.30	0.90	0.30	1.32	1.46	11.52 ± 0.72
1,1-dichloroethane	0.34	0.73	0.21	6.65	1.98	12.24 ± 0.72
DMF	0.61	1.70	0.56	1.97	2.05	12.72 ± 0.96
Ethanol	0.31	1.15	0.19	6.10	1.72	12.00 ± 0.72
Water	0.34	0.65	0.28	3.45	1.60	11.04 ± 0.72
NH_3	0.37	0.91	0.34	1.36	1.36	6.00 ± 0.48
$CHCl_3$	0.33	2.23	4.38	10.35	1.95	12.96 ± 0.72
N_2	0.02	0.53	0.04	0.72	0.52	3.12 ± 0.24
CO_2	0.00	0.32	0.04	0.83	0.68	5.76 ± 0.48
CO	-0.02	0.38	0.05	0.99	0.94	3.12 ± 0.24
O_2	0.03	0.38	0.03	0.49	0.34	2.88 ± 0.24
SF_6	-0.24	0.01	0.27	0.33	0.59	7.44 ± 0.48
Xe	0.67	5.72			3.03	5.76 ± 0.48

Table 6.2: Desorption energies calculated by different semiempirical methods. All values are reported in kcal mol^{-1}. Reference TPD values are in the last column.

problem disappears when the PM6-D3H4 method was used to recalculate the desorption energies on the geometries optimized with the PM7 method; see Tab. 6.3. Strikingly, methanol, ethanol and water were always predicted to have an interaction energy with the substrate much lower than the experimental ones. This inconsistency can be rationalize considering the chemical nature of these compounds, which all expose hydrogen bond donor and acceptor groups that may mediate the formation of 2D supramolecular complexes on the surface connected by hydrogen bonds. Since the strength of the hydrogen bond is comparable with the desorption energy of these molecules, what is actually measured by TPD can be either the desorption of dimers and trimers, or the breaking of the hydrogen bond network. Due to this uncertainty, these molecule were not further considered.

Tab. 6.4 reports the desorption energies calculated by four force fields: MMFF94, CGenFF, GAFF and OPLS-AA. Perhaps surprisingly, all methods provide interaction energies with the surface in good agreement with the experimental data, even if it they seem to underestimate it. In analogy to SQM, the desorption energy of methanol, ethanol, and water are highly underestimated by the FF methods. The data missing in the tables correspond to cases where the automatic typing procedure failed due to faults

	PM6-DH2 /PM6	PM6-DH+	PM6-D3H4 /PM6	PM7	PM6-D3H4 /PM7	Exp.
			Alkanes			
Methane	3.12	3.18	3.23	4.58	3.31	3.37
Ethane	4.44	4.73	4.54	7.54	4.71	5.88
Propane	6.20	6.58	6.40	10.15	6.57	7.67
Butane	7.97	8.44	8.08	13.20	8.40	9.75
Hexane	11.58	12.55	11.76	18.99	12.16	15.05
Octane	15.37	16.02	15.77	24.97	15.89	17.34
Decane	18.86	19.72	19.31	31.38	19.34	21.83
			Aromatic compounds			
Benzene	10.99	11.88	9.38	17.28	9.85	11.52 ± 1.92
Naphtalene	17.04	18.70	14.23	27.41	15.24	18.48 ± 2.16
Toluene	12.82	13.86	7.84	20.37	11.89	16.32 ± 1.68
Ethylbenzene	14.55	15.53	12.89	23.10	13.53	18.96 ± 2.4
o-dichlorobenzene	15.53	16.74	13.03	33.67	11.83	16.56 ± 1.44
			Gas and Solvent Molecules			
Methanol	4.91	4.74	5.08	6.33	4.29	11.52 ± 0.72
1,1-dichloroethane	8.97	9.89	8.02	24.52	7.59	12.24 ± 0.72
DMF	9.55	10.29	9.13	14.42	9.31	12.72 ± 0.96
Ethanol	6.47	6.54	6.57	9.30	6.14	12.00 ± 0.72
Water	3.63	3.64	3.89	4.19	3.84	11.04 ± 0.72
NH_3	4.21	4.23	4.30	5.11	4.25	6.00 ± 0.48
$CHCl_3$	6.07	9.07	5.37	23.47	6.31	12.96 ± 0.72
N_2	2.83	2.97	2.71	3.80	2.70	3.12 ± 0.24
CO_2	3.61	3.83	3.30	5.11	3.49	5.76 ± 0.48
CO	2.56	2.92	2.48	3.79	2.52	3.12 ± 0.24
O_2	1.76	2.26	1.91	3.28	2.19	2.88 ± 0.24
SF_6	0.28	3.75	0.14	8.41	2.86	7.44 ± 0.48
Xe	5.08	5.20	5.16	23.23	5.29	5.76 ± 0.48

Table 6.3: Desorption energies calculated by different semiempirical methods. All values are reported in kcal mol^{-1}. Reference TPD values are in the last column.

in the typing software, or missing atom type in the force field. This can be a problem when large libraries of compounds are screened, as it can limit the number of molecules effectively to be tested. In this respect, GAFF is most robust. In the same table the results for the ωB97X-D DFT functional are also reported. Interestingly, this method showed highly accurate results for some molecules, in particular the shortest alkanes and the aromatic compounds, but overestimated the interaction for the two longer alkanes, and underestimated it for gas and solvent molecules. Thus, it appears that the ωB97X-D predictions are unreliable, perhaps due to the small basis set used (6-31G(d)), or a bias of the functional towards some specific class of molecules. Not enough data are currently available to conclude on either hypothesis.

6.4 COMPARING THE ACCURACY

To quantify and compare the accuracy of the different methods, the linear correlation between computed and experimental desorption energies were calculated. What was expected was a one to one correlation, thus having a slope of one, an intercept of zero,

	MMFF	CGenFF	GAFF	OPLS-AA	ωB97X-D	Exp.
			Linear Alkanes			
Methane	2.32	3.05	3.30	3.15	3.63	3.37
Ethane	3.43	4.88	5.10	4.82	5.68	5.88
Propane	4.76	6.81	7.19	6.73	8.24	7.67
Butane	6.15	8.78	9.32	8.67	10.77	9.75
Hexane	8.96	12.72	13.61	12.60	15.79	15.05
Octane	11.76	16.14	17.85	16.51	20.73	17.34
Decane	14.56	20.56	22.10	20.41	25.64	21.83
			Aromatic compounds			
Benzene	9.15	10.40	11.26	10.15	11.35	11.52 ± 1.92
Naphtalene	14.60	16.43	17.92	16.07	18.47	18.48 ± 2.16
Toluene	10.75	12.54	13.64	12.34	14.16	16.32 ± 1.68
Ethylbenzene	11.12	13.27	13.51	12.49	16.46	18.96 ± 2.4
o-dichlorobenzene	13.28	14.67	15.67	14.41	14.82	16.56 ± 1.44
			Gas and Solvent Molecules			
Methanol	3.15	4.51	4.85	4.40	4.53	11.52 ± 0.72
1,1-dichloroethane	7.06	8.51	8.71	7.63	6.60	12.24 ± 0.72
DMF	7.33	10.25	11.19	10.70	10.67	12.72 ± 0.96
Ethanol	4.18	6.23	6.72	6.05	6.32	12.00 ± 0.72
Water	1.85				2.31	11.04 ± 0.72
NH_3	2.25	1.91	2.45	3.64	3.10	6.00 ± 0.48
$CHCl_3$	8.10		8.48	8.82	7.75	12.96 ± 0.72
N_2	2.28		3.94	3.73	2.38	3.12 ± 0.24
CO_2		4.37	6.19		2.82	5.76 ± 0.48
CO			4.31		2.15	3.12 ± 0.24
O_2			3.05		1.82	2.88 ± 0.24
SF_6					4.20	7.44 ± 0.48
Xe						5.76 ± 0.48

Table 6.4: Desorption energies calculated by different force field and DFT methods. All values are reported in kcal mol^{-1}. Reference TPD values are in the last column. Empty values in the force fields correspond to molecules non supported in the force fields

and correlation coefficient R^2 as close as possible to one. All the plots obtained for the different methods are reported in Fig. 6.3, with the values of slope, intercept and correlation coefficient given in Tab. 6.5.

By comparing at first the semiempirical methods, the PM6-DH2/PM6, PM6-DH+, PM6-D3H4/PM6 and PM6-D3H4/PM7 are all in good agreement with the experimental results. Among them, the most accurate is PM6-DH+, showing an error on the slope of 2.10% and an error on the R^2 of 4.68%. Similarly, the PM6-DH2 gave good results, but with an error on the R^2 of 10.40% indicating a more pronounced spread of the data. For the PM6-D3H4, both the use of PM6 or PM7 geometries was able to reproduce nicely the experimental desorption energies: using the PM6 geometry a better slope was obtained but a worse spread, while using the PM7 geometry, the spread decreased, increasing the correlation at the expense of a poorer slope. Last, the PM7 method yielded the worst results, completely overestimating the desorption energy (slope of about 0.5) and having the largest spread (23%).

On the FF methods, all of them showed a very good correlation with the experiments, with errors in R^2 less than 10%. The most accurate results were obtained with

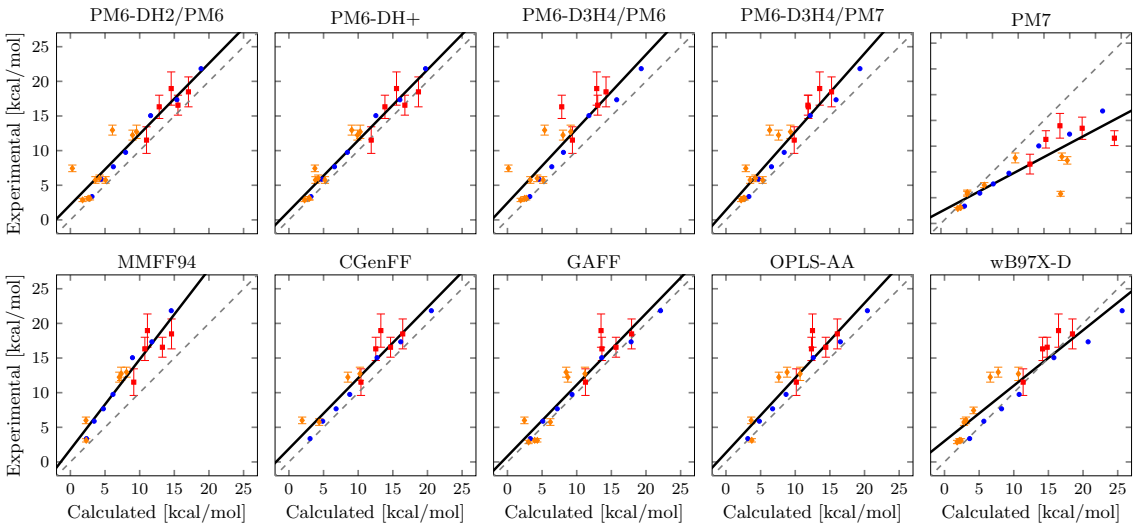


Figure 6.3: Correlation plots between the calculated and the experimental desorption energies. Five semiempirical methods (PM6-DH2/PM6, PM6-DH+, PM6-D3H4/PM6, PM6-D3H4/PM7 and PM7), four force fields (MMFF94, CGenFF, GAFF and OPLS-AA), and one DFT functional (ωB97XD/6-31G(d)) are reported. The color of the dots represents the class of the molecule: blue alkanes, red aromatic compounds, and orange small molecules and solvents. The error bars represent the experimental uncertainties in the TPD data. The dashed gray line indicates the ideal fit, compared to the real linear fit (solid black line).

Method	Level of theory	Slope	Intercept	R^2	Error % Slope	Error % R^2	RMSE
PM6-DH+	SQM	1.02	1.28	0.95	2.10	4.68	1.94
PM6-DH2/PM6	SQM	1.02	2.19	0.90	1.70	10.40	2.99
PM6-D3H4/PM7	SQM	1.13	1.42	0.91	13.48	8.69	3.12
PM6-D3H4/PM6	SQM	1.07	2.45	0.83	7.08	16.64	3.83
PM7	SQM	0.53	2.29	0.77	47.01	23.20	7.38
GAFF	FF	1.03	0.83	0.90	2.55	9.71	2.14
CGenFF	FF	1.02	1.90	0.93	2.28	6.88	2.57
OPLS-AA	FF	1.08	1.33	0.91	8.30	9.23	2.77
MMFF94	FF	1.30	1.80	0.93	30.01	7.43	4.65
ωB97XD/6-31G(d)	DFT	0.80	3.01	0.88	19.77	12.01	2.62

Table 6.5: Results of the linear correlation between the computed and experimental desorption energy values. Reported values are: the slope, the intercept and the determination coefficient (R^2) of the linear regression, the deviation from the ideal behavior of the slope and the determination coefficient, and finally the Root Mean Square Error (RMSE) of the computed values.

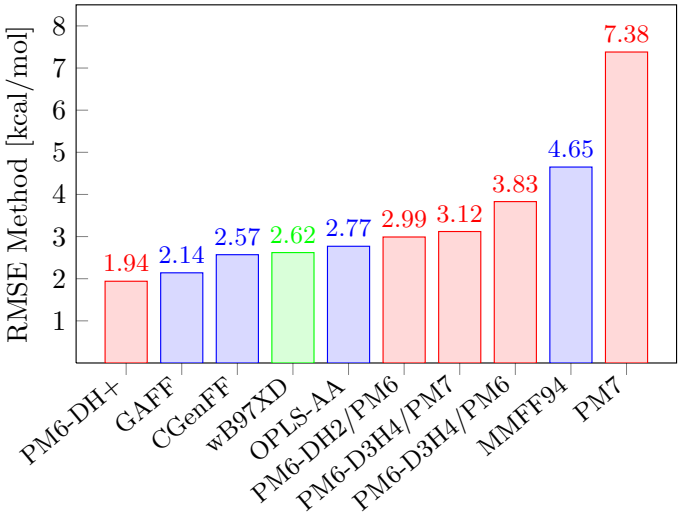


Figure 6.4: Root Mean Square Error (RMSE) of the tested computational methods relative to the experimental TPD data. All values are reported in kcal mol^{-1}. SQM are in red, force field in blue and DFT in green. Methods are ranked by RMSE.

the CGenFF and GAFF force fields, which accurately reproduced the experimental values with errors on the slope of 2.28% and 2.55%, respectively. Remarkably, these results are comparable with the best semiempirical methods, while the computational cost is significantly lower (see next paragraph). Good performances were also obtained with the last two force fields (MMFF94 and OPLS-AA), but these showed a slope larger than one, indicating a systematic underestimation of the interaction energy with the substrate. Finally, the results of the ωB97X-D functional in combination with the 6-31G(d) basis set poorly describe the molecule/substrate interaction energy in this set of data, with sizable errors both on the slope and the determination coefficient.

In order to better rank all methods, their predictability was measured as the Root Mean Square Error (RMSE) of the computed desorption energies versus the experimental values. The RMSE are shown in Fig. 6.4. Based on this analysis, the semiempirical PM6-DH+ is the best method with an average error of 1.9 kcal mol^{-1}. Very good is also the GAFF force field with a RMSE of 2.1 kcal mol^{-1}. All the other methods showed errors larger than 2.5 kcal mol^{-1}.

6.5 EFFICIENCY & CPU TIME

Beside the accuracy, the efficiency of the calculations is of great importance since only an accurate and efficient method is suitable for exploring the chemical space. Thus, to evaluate the efficiency of the various methods, the CPU time to calculate the desorption energy per molecule was evaluated as the average time per molecule to analyze the experimental data set. These times are reported in Fig. 6.5: force field methods generally need a few minutes, SQM few hours, and DFT months of CPU time. Importantly, since

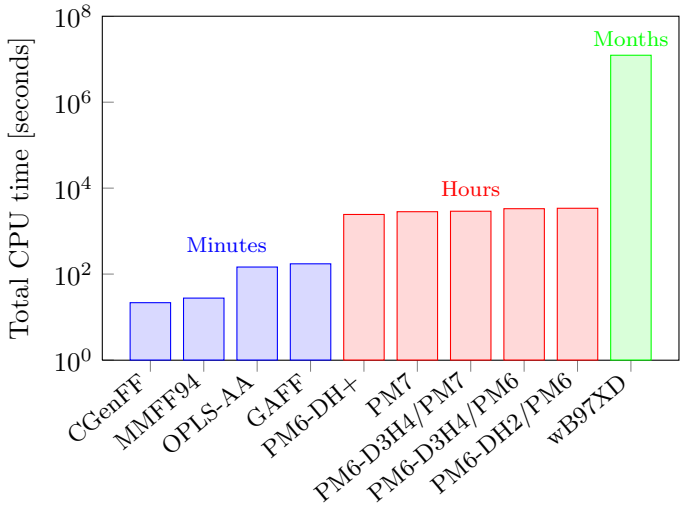


Figure 6.5: Average CPU time needed to calculate the desorption energy of one molecule using the different computational methods here benchmarked. This corresponds to the time on a single processor: SQM and DFT calculation are actually run in parallel decreasing significantly the user time.

SQM and DFT can be run in parallel on multiple processors, the actual wall clock time is obtained by dividing the toal CPU time by the number of processors used. This means that the DFT calculations are actually performed in about one week of user time. Still, DFT is too expensive for routine calculations, while FF and SQM methods provide a valuable alternative, which is able to grant an efficient translation of the computational predictions into chemical information.

6.6 EFFECT OF A SECOND AND THIRD LAYER OF GRAPHENE

Up to this point, the graphite substrate used in the TPD experiments has been always modeled as a graphene monolayer. In this last section the effect of adding a second and third layer is studied. As before, the semiempirical PM6-DH+ is used as reference method for this job. Since the calculation with two and three graphene layers is one order of magnitude computationally more intensive, only the desorption energy of benzene was studied. Adding a second layer in an AB stacking resulted in a significant increase in the desorption energy for benzene of 0.76 kcal mol^{-1}. Adding a third layer resulted instead in a very small change of only 0.04 kcal mol^{-1}. Thus, it was worth recalculating the desorption energies of all the molecules in the dataset using a two layer graphene model, and re-analyzing the correlation with the experimental data. Effectively, it was observed that while the correlation coefficient R^2 and the intercept did not change significantly, the slope moved from 1.02 to 0.95, which means from a slight understimation to a slight overestimation. Most importantly, using a two-layer surface model the RMSD decreased from 1.94 kcal mol^{-1} to 1.44 kcal mol^{-1}. This implied the use of a second

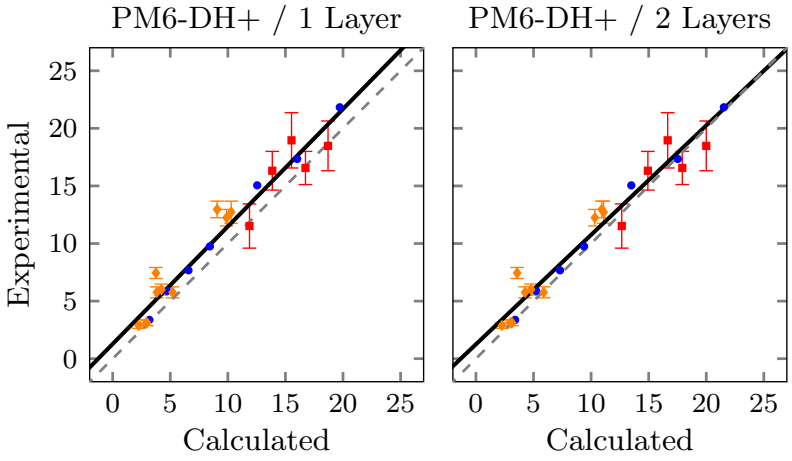


Figure 6.6: Correlation between the experimental and the calculated desorption energies evaluated using the PM6-DH+ semiempirical method using one (left) or two (right) graphene layers to model the graphite substrate.

layer is important to model the interaction with a graphite substrate. This at the cost of a tenfold increase in the computational time. The correlations between the experimental and calculated desorption energies with the one and two layers graphite model are reported in Fig. 6.6.

6.7 EXPLORING THE CHEMICAL SPACE

At this point, it was possible to identify in the semiempirical PM6-DH+ a method able to give accurate results in a reasonable amount of time per molecule. Here, this semiempirical method is used to explore the chemical space, looking for punctual chemical functionalizations which can improve the binding energy to graphene. In this direction, a benzene scaffold was functionalized by introducing heteroatoms inside the aromatic ring (nitrogen, oxygen, and sulfur), or by a 1,3,5 tri-substitution with halogen, hydroxyl, amino, nitro, methyl, and carboxyl groups; see Fig. 6.7. The goal of this analysis was to see how the desorption energy would change when a molecule is decorated with polarizable, electron donating, or electron withdrawing groups. The computed desorption energies are reported in Tab. 6.6.

The first comparison is between benzene and hexane, its saturated counterpart. Surprisingly, the calculation showed that hexane binds stronger by 0.7 kcal mol^{-1}, which is confirmed by the experimental data in Tab. 6.1. This suggests that linear alkanes have a better affinity for graphene relative to polycyclic aromatic hydrocarbons; see below.

A second comparison, is based on the substitution of carbon atoms in a benzene ring by nitrogen or oxygen, which results in a consistent decrease of the desorption energy. Looking at the series of pyridine, 1,3-pyrimidine, and 1,3,5-triazine in Tab. 6.6, which correspond to a mono, di, and tri substituted ring by nitrogen, the desorption energy

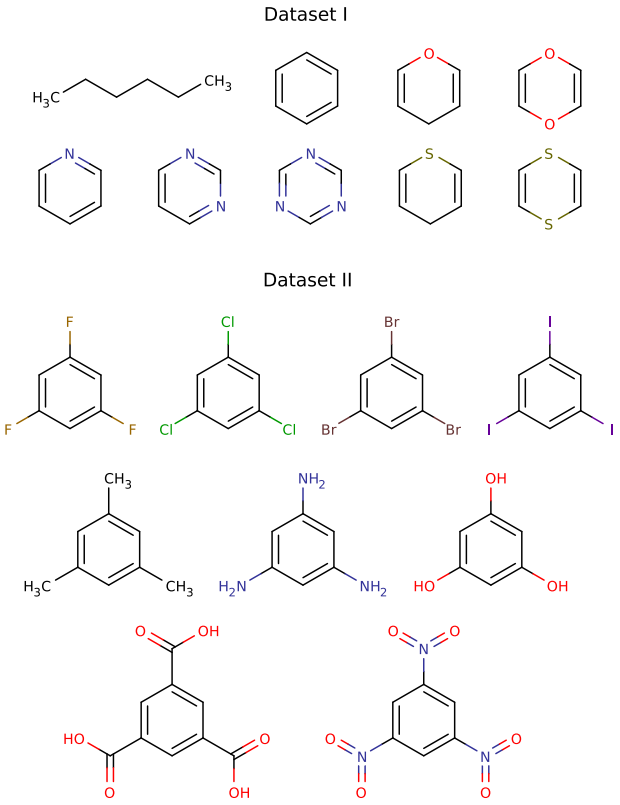


Figure 6.7: Molecular structures of the two data sets used to explore the effect of chemical functionalization on the desorption energy from graphene. In the first set, a benzene scaffold is modified by in-ring substitution with heteroatoms; in the second set, the benzene is 1,3,5 tri-substituted by different polarizable, electron donating, or electron withdrawing groups.

First Dataset		Second Dataset			
Molecule	E_{des}	Molecule	E_{des}	E_{grp}	$\Delta E_{SQM/FF}$
Hexane	12.55	Methyl	18.42	1.96	0.01
Benzene	11.88	Fluoro	13.76	0.40	0.28
Pyridine	11.59	Chloro	19.35	2.27	-0.38
1,3-pyrimidine	11.24	Bromo	24.68	4.04	-1.57
1,3,5-triazine	10.88	Iodio	27.55	5.00	-1.86
4H-pyran	11.25	Hydroxy	16.39	1.28	0.05
1,4-dioxin	10.32	Carboxy	25.23	4.26	-0.21
4H-thiopyran	12.69	Amino	19.90	2.45	-0.83
1,4-dithiine	12.79	Nitro	23.03	3.49	0.66

Table 6.6: Desorption energies calculated using the PM6-DH+ method for the compounds represented in Fig. 6.7. In the second dataset, the desorption energy per group was calculated as $E_{grp} = (E_{molecule} - E_{benzene})/3$, while the difference with the CGenFF results ($\Delta E_{SQM/FF}$) as as $\Delta E_{SQM/FF} = E_{grp}^{SQM} - E_{grp}^{FF}$. All values are given in kcal mol^{-1}.

decreases by about 0.33 kcal mol^{-1} for each nitrogen atom introduced. Similarly, substitution with oxygen reduces the desorption energy, although in a more marked way: 0.63 kcal mol^{-1} for monosubstitution (4H-pyran) and 1.56 kcal mol^{-1} for the disubstitution (1,4-dioxin). Essentially, the substitution of two oxygen atoms has the same effect as the substitution of three nitrogens. In the opposite direction goes the substitution with sulfur. The introduction of one sulfur atom (4H-thiopyran) has the effect of increasing the desorption energy by 0.81 kcal mol^{-1}, whereas the introduction of a second sulfur atom (1,4-dithiine) further increases the affinity by only 0.10 kcal mol^{-1}. This can be understood considering that the introduction of the first sulfur does not change the total number of atoms in the molecule, but the introduction of the second sulfur causes the net loss of two hydrogens, which counterbalance the effect of the substitution with sulfur. It is thus possible to conclude that the substitution with heteroatoms across the period (nitrogen and oxygen) reduces the interaction energy with the surface, while moving down the group the interaction is enhanced. This is in agreement with the fact that dispersive interactions, the most important contribution to physisorption, are defined by the polarizability of the molecule[222]: nitrogen and oxygen have a smaller polarizability than carbon, whereas moving to sulfur the polarizability increases.

The effect of the polarizability was further explored by comparing the desorption energy of trisubtituted benzene in 1, 3, and 5 position with halogens from fluorine to iodine; see Fig. 6.7. The effect of each halogen substitution was quantified as $E_{grp} = (E_{molecule} - E_{benzene})/3$, which removes the contribution from the common benzene core. As shown in Tab. 6.6, halogen substitution always enhances the binding energy to graphene with a magnitude that increases down the group: 0.4 kcal mol^{-1} per fluorine atom, 2.27 for chlorine, 4.04 for bromine, and 5.00 kcal mol^{-1} for iodine. Noteworthy, the substitution with bromine and iodine almost double the desorption energy of benzene.

As a last comparison, benzene was substituted with electron donating (methyl, hydroxyl and amino) and electron withdrawing (carboxyl and nitro) groups to analyze inductive effects on the adsorption energy on graphite. From Tab. 6.6 it appears that electron withdrawing groups increase the desorption energy more than electron donating groups: more than 4.0 kcal mol^{-1} per group with carboxyl and nitro, while less than 2.2 kcal mol^{-1} with methyl, hydroxyl, or amino substituents. In order to understand if this difference is due to genuine inductive effects, the results of the PM6-DH+ calculations were compared with the desorption energies calculated using the CGenFF classical force field, which does not capture inductive effects by definition. The last column in Tab. 6.6 gives the difference between the E_{grp} calculated using PM6-DH+ and CGenFF, $\Delta E_{SQM/FF}$, which approximates the effect of inductive effects. As shown in Tab. 6.6 (last column), these contributions are quite small, and, most importantly, they are not correlated with the electron donating/withdrawing character of the substituent. Thus, it is possible to conclude that inductive effect, if present, are generally small and a more important descriptor for the desorption energy is the polarizability of the molecule.

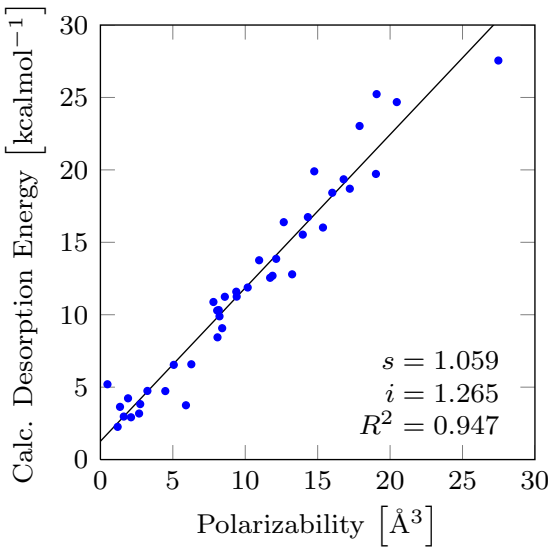


Figure 6.8: Correlation plot between the polarizability of a molecule and its desorption energy from graphene both calculated using the PM6-DH+ semiempirical method.

To further confirm this, the correlation between the desorption energy and the isotropic polarizability of all molecules studied in this work, evaluated at the PM6-DH+ level using the finite-field method implemented in MOPAC2012[223], was calculated. Fig. 6.8 shows a very good correlation, with a determination coefficient $R^2 = 0.947$.

6.8 SMART DESIGN OF NOVEL GRAPHENE SURFACTANTS

In the introduction chapter of this thesis it was described how the molecules with a good affinity for the graphene layers can be used as surfactants to enhance the liquid phase exfoliation (LPE) of graphite. In this section, the PM6-DH+ semiempirical method is used to identify and design molecules with enhanced adsorption energy on graphene, which may be used as graphene surfactants. To do so, four series of compounds were studied, comparing their calculated desorption energy as a function of their size. The four classes of molecules are: polycyclic aromatic hydrocarbons (PAH), linear alkanes, fully fluorinated alkanes (FFA), and fully chlorinated alkanes (FCA). The PAH group includes benzene, naphthalene, coronene and ovalene. These are used as reference, since extrapolating their desorption energy to an infinitely extended PAH, the interlayer cohesive energy of graphite is obtained. Calculating the desorption energy of these molecules and correlating them with the number of carbon atoms, a perfect linear correlation is obtained with a slope of 1.65 kcal mol^{-1} per carbon atom. This means that for a PAH with n carbon atoms, its desorption energy can be evaluated as $1.65 \cdot n$ kcal mol^{-1}. Thus, if it is possible to find a class of molecules with a stronger adsorption energy on graphite, i.e. with a larger slope with the molecular size, it is possible to think that these molecules can be possibly used as graphene surfactants to enhance the LPE of graphite, since

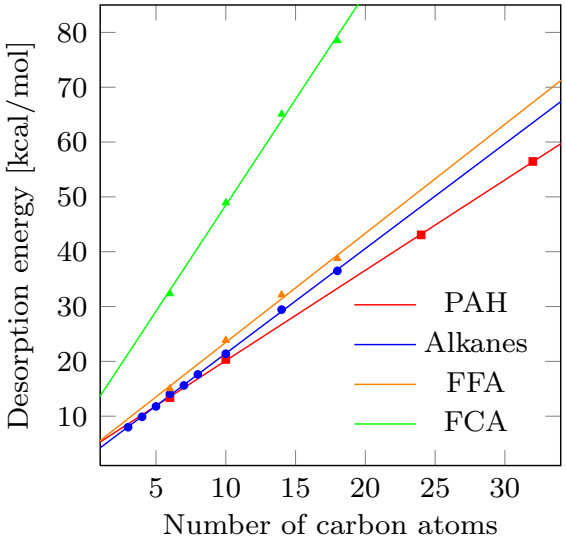


Figure 6.9: Desorption energy of the four series of compounds studied as a function of the number of carbon atoms; polycyclic aromatic hydrocarbons (PAH) are represented in blue, linear alkanes in red, fully fluorinated alkanes (FFA) in yellow, and fully chlorinated alkanes (FCA) in green.

their interaction energy with the graphene is better then the interaction energy between two graphene layers. The linear correlations for the PAH and the other three class of molecules (discussed below) are shown in Fig. 6.9.

The second class of molecules is composed by linear alkanes from propane (C_3) to C18. As shown before comparing benzene to hexane, these molecules have a better adsorption energy on graphite than the corresponding PAHs. This is reflected here as the linear correlation with the number of carbon atoms is 1.91, which is about 16% higher than the one of the PAHs. Interestingly, recently linear alkanes have been successfully used as gaphene surfactants to enhance the liquid-phase exfoliation of graphite[67]. Moving from linear alkanes to fully fluorinated alkanes (FFA), no great improvements is observed, with the slope that marginally increases to 1.99. Strikingly, when the alkyl chain is fully chlorinated (FCA) the results change dramatically. The slope now is 3.87, which implies the desorption energy of a FCA is double that of a PAH with the same number of carbon atoms.

While perchlorination of alkanes seems to be a viable route to enhance the LPE of graphite, these molecules are actually not synthetically accessible. In fact, perchlorinated alkanes are generally unstable and tend to spontaneously decompose eliminating chlorine[224], and they have been synthesized only up to butane[225]. By contrast, many partially chlorinated linear alkanes (PCA) are commercially available[226] as they seem to be more synthetically accessible[227]. It is thus possible to use the PM6-DH+ method to explore these PCA, in order to design in a smart way compounds that optimize the interaction with the surface, but still are synthetically accessible. Along this line, five PCA were studied. All are based on the same skeleton (a undecane) and are partially chlori-

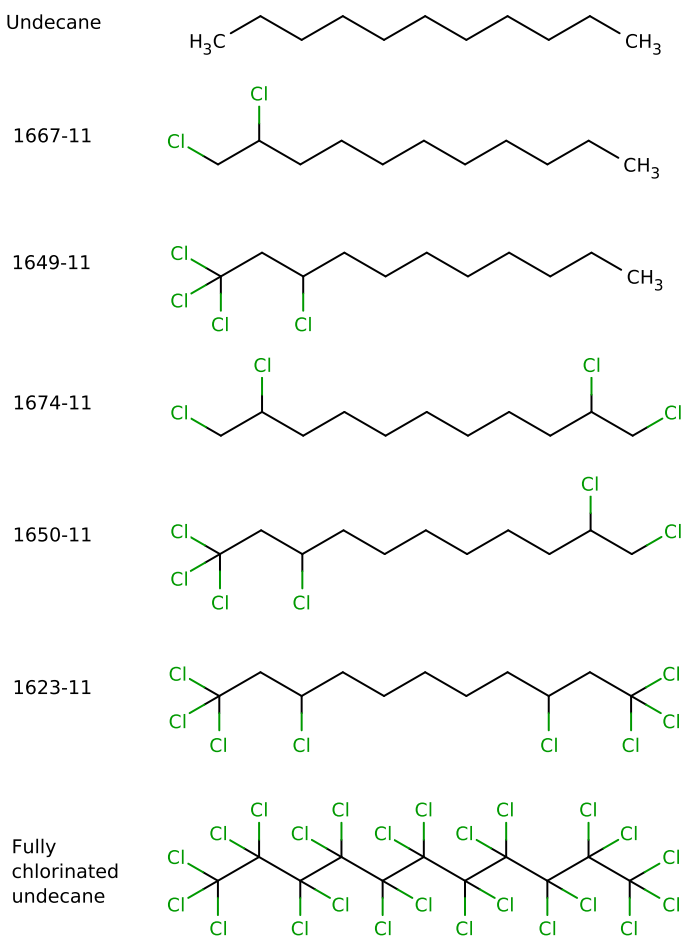


Figure 6.10: Molecular structured of the commercially available partially chlorinated undecane molecules studied in this work.

nated by two, four, six, and eight chlorine atoms; see Fig. 6.10. The desorption energies of these molecule were computed and compared with the ones of undecane and its fully perchlorinated analogous to monitor the increment in the adsorption energy as a function of the number of chlorine atoms; see Tab. 6.6. Although perchlorination essentially doubles the desorption energy of undecane, the PCA 1623-11, which has only 8 chlorine atoms with respect to the 24 of the fully perchlorinated one, has a desorption energy that is 59% higher then udecane. If so, partial chlorination with one third of the chlorine atoms can recover more than half of the maximum possible desorption energy. This is due to the non-linear increase of the desorption energy as a function of the number of substituted chlorine atoms. That means fully chlorination is not only synthetically non accessible, but is also not necessary to optimize the enhancement of the binding affinity.

To conclude, the availability of a computational method fast enough for routine calculations, but still accurate in the predictions, allows for the design of molecules with optimized adsorption energies on graphite that are still synthetically accessible.

Molecule	Number of chlorines	E_{des} [kcal mol^{-1}]	Ratio to undecane	
Undecane	0	23.86	1.00	
PCA 1667-11	2	27.81	1.17	+17%
PCA 1649-11	4	31.02	1.30	+30%
PCA 1674-11	4	31.79	1.33	+33%
PCA 1650-11	6	33.51	1.40	+40%
PCA 1623-11	8	38.01	1.59	+59%
Fully chlorinated undecane	24	48.10	2.02	+102%

Table 6.7: Desorption energies calculated for the partially chlorinated alkanes. The undecane is set as reference, and the desorption energies of all PCA is reported as percentage increase with respect to undecane. The fully perchlorinated undecane is also reported as upper bound.

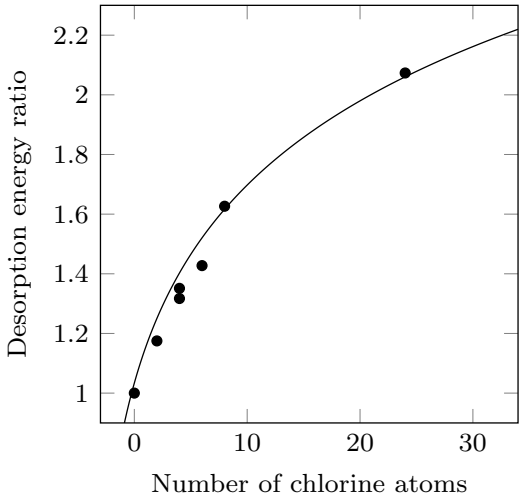


Figure 6.11: Relative desorption energies of the PCA with respect to undecane plotted as a function of the number of chlorine atoms present in the molecule. As main feature, the increase in the desorption energy is not linear with the chlorine content.

6.9 TAKE HOME MESSAGES

In this chapter, different methods ranging from classical Force Field (FF), Semiempirical Quantum Mechanics (SQM), and Density Functional Theory (DFT) have been tested on their ability to accurately predict the desorption energy of 25 chemically diverse compounds from graphene. This benchmark showed that the semiempirical PM6-DH+ was the most accurate among the 16 computational methods tested, still being efficient enough for routine calculations (a few hours per molecule). Also, the empirical GAFF and CGenFF force fields showed surprising accuracies only slightly worse than PM6-DH+, with the big advantage of requiring a few minutes of computational time. The discovery of fast and accurate methods allowed the exploration of the chemical space to identify possible chemical functionalizations to enhance the binding affinity for graphene. Interestingly, the use of highly polarizable compounds, e.g. by halogen substitution, was shown to increase the adsorption energy quite significantly. In this line,

the use of partial chlorination of linear alkanes was explored to design compounds with increased adsorption energy for graphene and synthetically accessible.

7

LEVERAGING THE LIQUID-PHASE EXFOLIATION OF GRAPHITE THROUGH SELF-ASSEMBLY

7.1 SURFACTANTS-ASSISTED LIQUID-PHASE EXFOLIATION OF GRAPHITE

Already in the introduction of this thesis, all potential technological advantages of graphene have been described, together with all the difficulties in its production. Among all available methods to produce high quality graphene flakes[39–47], very promising is the Liquid Phase Exfoliation (LPE) of graphite[48–52]. In this technique the starting graphite powder is first dispersed into a solvent. Then the application of a mechanical force, typically ultrasounds, causes the actual exfoliation by breaking the van der Waals interaction that keep the graphene sheets stacked together. Finally, the exfoliated material is purified, removing the remaining graphitic materials[52]. Obviously, the critical step is the second, where sonication causes the exfoliation. In this phase of the process it is very important to setup the environmental conditions in such a way that the exfoliated graphene layers can be stabilized in the solution by preventing their restacking. It has been recently shown that this can be actually obtained by a careful choice of the solvent, such as N-methyl-2-pyrrolidone (NMP)[53], ortho-dichlorobenzene (o-DCB)[54], 1,2,4-trichlorobenzene (TCB), N,N-dimethylformamide (DMF)[55], or pentafluorobenzonitrile (PFBN)[56]. Common characteristic of these solvents is to have a surface tension in the order of $30 - 40$ mJ m^{-2}, which is predicted to be optimal for the LPE[53]. A parallel strategy to enhance the stability of the exfoliated graphene flakes consists in the addition of some surfactants, which can stabilize the graphene sheets in solution; see Fig. 7.1. These can be the most diverse molecules, and are here generally referred as dispersion-stabilizing agents. However, it is not clear how to rationalize the effect of these molecules on the exfoliation yield, which has prevented the design of optimized exfoliators.

One possible way to engineer optimal graphene surfactants is through the use of molecules which can generate self-assembled monolayers on the graphene surface. The idea is that the formation of a SAM on the top of the exfoliated graphene can greatly stabilize the graphene flakes, preventing their restacking. First result in this direction was the demonstration that arachidic acid (C20), which is known to produce very stable SAMs on graphite[228], can be used to enhance the exfoliation yield in NMP[67]. The question that naturally opens at this point is how to engineer a molecule that can self-assemble on the exfoliated graphene layers to increase as much as possible their stability. To answer this question, the arachidic acid was modified by changing the length of the aliphatic chain to see the effect on the exfoliation yield. Five linear fatty acids were

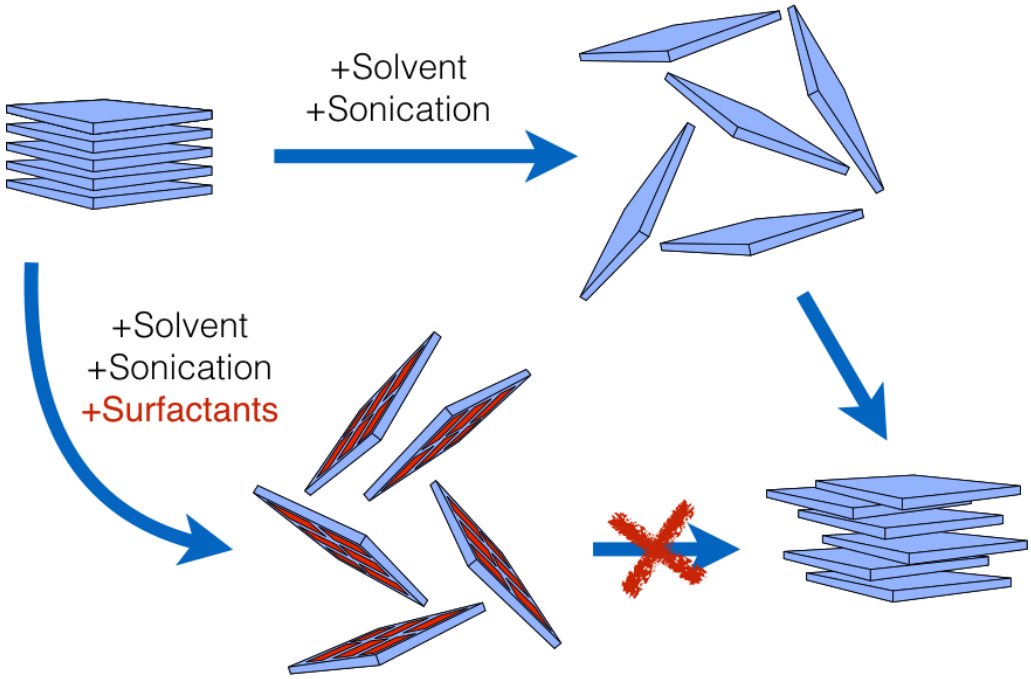


Figure 7.1: Schematic representation of the liquid-phase exfoliation by sonication in the absence (top) and presence (bottom) of dispersion stabilizing agents.

studied: hexanoic acid (C6), lauric acid (C12), stearic acid (C18), lignoceric acid (C24) and melissic acid (C30) (see Fig. 7.2), each of them in four different solvents: N-methyl-2-pyrrolidone (NMP), ortho-dichlorobenzene (o-DCB), 1,2,4-trichlorobenzene (TCB), and N,N-dimethylformamide (DMF). First, their ability to self-assemble on graphite was studied by STM. Then, the exfoliation yield was measured, and the results interpreted. To this aim, we developed a thermodynamic model to assess the relative stability of the SAMs generated by the five molecules and to disentangle the energetic versus the entropic contributions in order to provide with a full thermodynamic picture of the process. This was done based on the model derived in chapter 5 of this thesis. In this endeavor we were able to identify the molecular determinants that influence the stability of the SAM, which allows to tune the exfoliation yield by chemical design. Last, the quality of the exfoliated material was analyzed by Transmission Electron Microscopy (TEM) and micro Raman experiments.

This work is the result of the cooperative efforts of four different research groups: S. Haar, A. Ciesielski, F. Richard, and P. Samorì (Nanochemistry Laboratory, ISIS & icFRC, Université de Strasbourg & CNRS, Strasbourg, France) designed the experiment and performed the liquid-phase exfoliation of graphite, N. Merstorf, M. Cecchini and myself generated the molecular models of the SAMs and accessed their thermodynamic stability, R. Mazzaro and V. Morandi (IMM-CNR Sezione di Bologna, Bologna, Italy) did the TEM analysis, while J. Clough, H. Yang, and C. Casiraghi (School of Chemistry, University of Manchester, Manchester, UK) performed the micro Raman experiments. Most of the results presented here were published in *Small*[184].

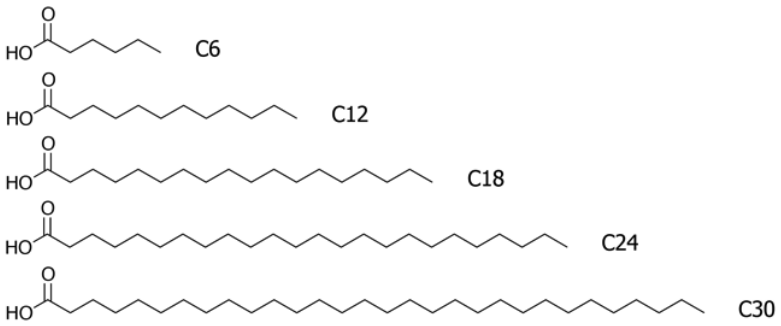


Figure 7.2: Chemical structure of the fatty acids studied, from hexane (C6) to triacontane (C30).

7.2 STM STUDY OF THE SELF-ASSEMBLED MONOLAYERS

The self-assembly of the five carboxylic acids in Fig. 7.2 on graphite was analyzed by STM. As the solvents used in LPE are not suitable for STM imaging, as they have a too high dielectric constant, 1-phenyloctane was used and STM images were recorded applying a 4 µL drop of 20 mM solution on a freshly cleaved HOPG substrate. The obtained crystalline patterns were analyzed extracting the unit cell parameters. The length of the two translational vectors a and b, the angle γ among the two, the unit cell area A_{uc}, the number of molecules per unit cell n_{uc}, and the unit cell area normalized per number of molecules A'_{uc} are reported in Tab. 7.1.

The self-assembly of hexanoic acid (C6) was studied first. As expected, C6 did not form any stable SAM on the surface. In fact, C6 is commonly used as solvent in different self-assembly studies[18,33,229,230]. Lauric acid (C12) did self-assemble on the surface producing the characteristic lamellar structures; see Fig. 7.3a. In this SAM, the alkyl tails of the acid are adsorbed in an in-plane zig-zag, and all molecules form dimers through the formation of strong hydrogen bonds between the acid heads. The dimers are interdigitated, which makes two molecules per unit cell and the unit cell vectors corresponding to the thickness of two molecules (vector a) and the length of one molecule (vector b). The cell is almost rectangular, with an angle $\gamma = (90 \pm 2)°$. The obtained area per unit cell per molecule is $A'_{uc} = 0.86 \pm 0.14$ nm^2.

All longer derivatives showed exactly the same pattern, with two molecules forming dimers through the carboxylic groups, and the alkyl chains producing interdigitated lamellar structures; see Fig. 7.3. Comparing the unit cell parameters in Tab. 7.1, all molecules have exactly the same cell vector a and the same angle γ. The only difference is the cell vector b, which is proportional to the length of the molecule. This is reflected in the A'_{uc}, which is proportional to the chain length. In fact, the difference between $A'_{uc}(C12)$ and $A'_{uc}(C18)$ is exactly the same as the difference between $A'_{uc}(C18)$ and $A'_{uc}(C24)$ and $A'_{uc}(C24)$ and $A'_{uc}(C30)$, which corresponds to the area occupied by a C6 fragment. Noteworthy, the unit cell parameters in Tab. 7.1 are in perfect agreement with previous works[231], and SAMs grown on CVD graphene[67].

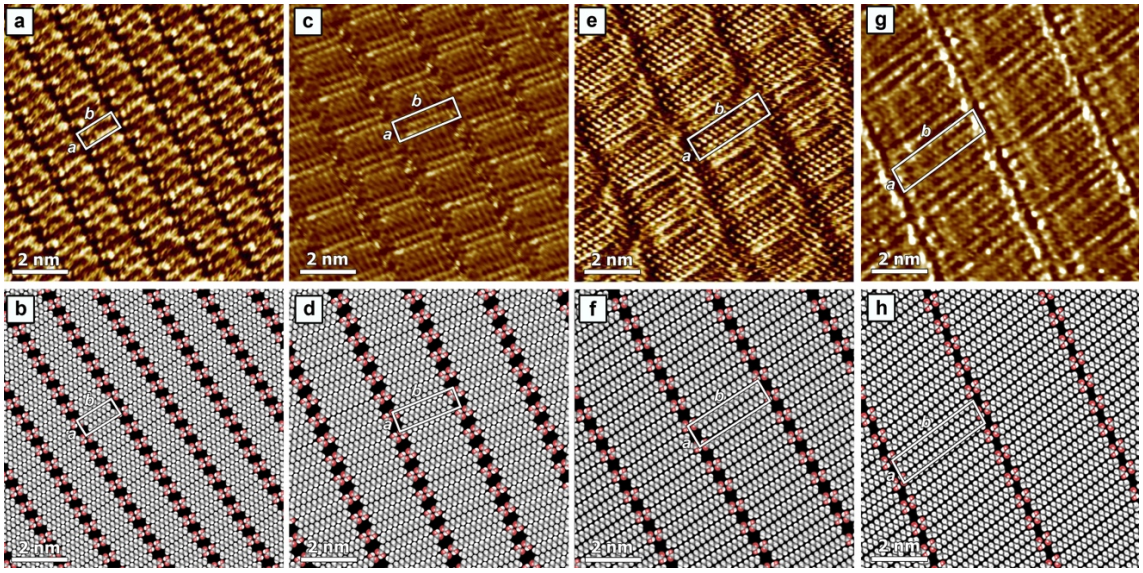


Figure 7.3: STM images (top row) and molecular models (bottom row) of the SAMs formed by the four molecules from C12 (left), to C30 (right). The shorted C6 does not show ordered SAMs.

Molecule		a [nm]	b [nm]	γ [°]	A'_{uc} [nm^2]
C6	*Exp.*	//	//	//	//
	Model	0.89	1.14	87	0.51
C12	*Exp.*	0.95 ± 0.2	1.81 ± 0.2	90 ± 2	0.86 ± 0.14
	Model	0.84	1.92	88	0.81
C18	*Exp.*	0.94 ± 0.2	2.61 ± 0.2	88 ± 2	1.23 ± 0.09
	Model	0.88	2.78	90	1.22
C24	*Exp.*	0.95 ± 0.2	3.36 ± 0.2	90 ± 2	1.59 ± 0.07
	Model	0.91	3.31	86	1.50
C30	*Exp.*	0.95 ± 0.2	4.15 ± 0.2	92 ± 2	1.97 ± 0.05
	Model	0.91	4.09	86	1.85

Table 7.1: Experimental and modeled unit cell parameters for the five molecule studied in this work. No experimental parameters for C6 are present as no stable STM images have been observed for this molecule.

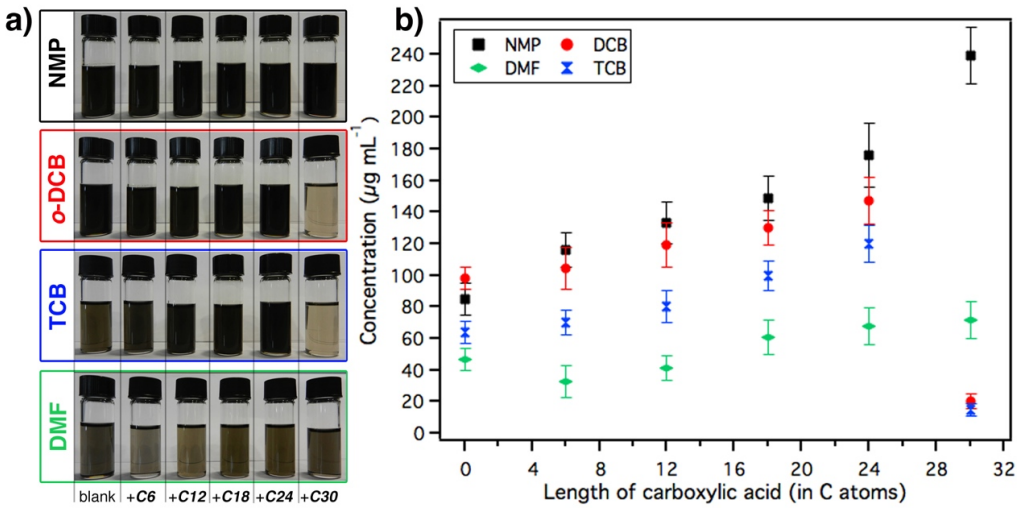


Figure 7.4: a) Photographs of the graphene dispersion obtained by LPE of graphite flakes in the four studied solvents and in the present of the five fatty acids. b) Concentration of the graphene dispersion; the error bars correspond to the standard deviation from 15 independent experiments.

7.3 FATTY-ACID ENHANCED LIQUID-PHASE EXFOLIATION

The ability of the five derivatives in Fig. 7.2 to aid in the liquid-phase exfoliation of graphite was tested. The experimental setup involved the dissolution of the carboxylic acid in NMP, o-DCB, TCB or DMF, the addition of the graphite powder, and sonication for six hours at the constant temperature of $40 \pm 2)°C$. By this procedure, all samples led to grey solutions composed of two phases: a homogenous solution and a large volume of solid aggregates. These latter are composed of non-exfoliated graphitic materials and can be removed by ultra centrifugation, yielding a homogenous dark dispersion of graphene flakes; see Fig. 7.4. To compare the efficacy of the five carboxylic acids in modulating the exfoliation yield, blank experiments were also performed, where only the pure solvent was used.

Following centrifugation, the yield of the exfoliation was calculated as follow. Most probably the graphene dispersions contain a large amount of the carboxylic acids, and these need to be removed to accurately measure the mass of the exfoliated material. This was done by mixing the graphene dispersions with chloroform and heating the sample at 50°C for half an hour. In this way, the carboxylic acids desorb from the graphene sheets and, hopefully, move in solution. Then, the solution was filtered using a polytetrafluoroethylene (PTFE) membrane (pore size 100 nm), and the resulting material washed several times using diethyl ether and chloroform to remove all traces of solvent and carboxylic acids. Finally, the filtered mass was weighted on a microbalance, yielding the total mass of exfoliated material, and thus the concentration of the graphene dispersion. For each carboxylic acid / solvent combination, 15 independent exfoliation experiments were performed.

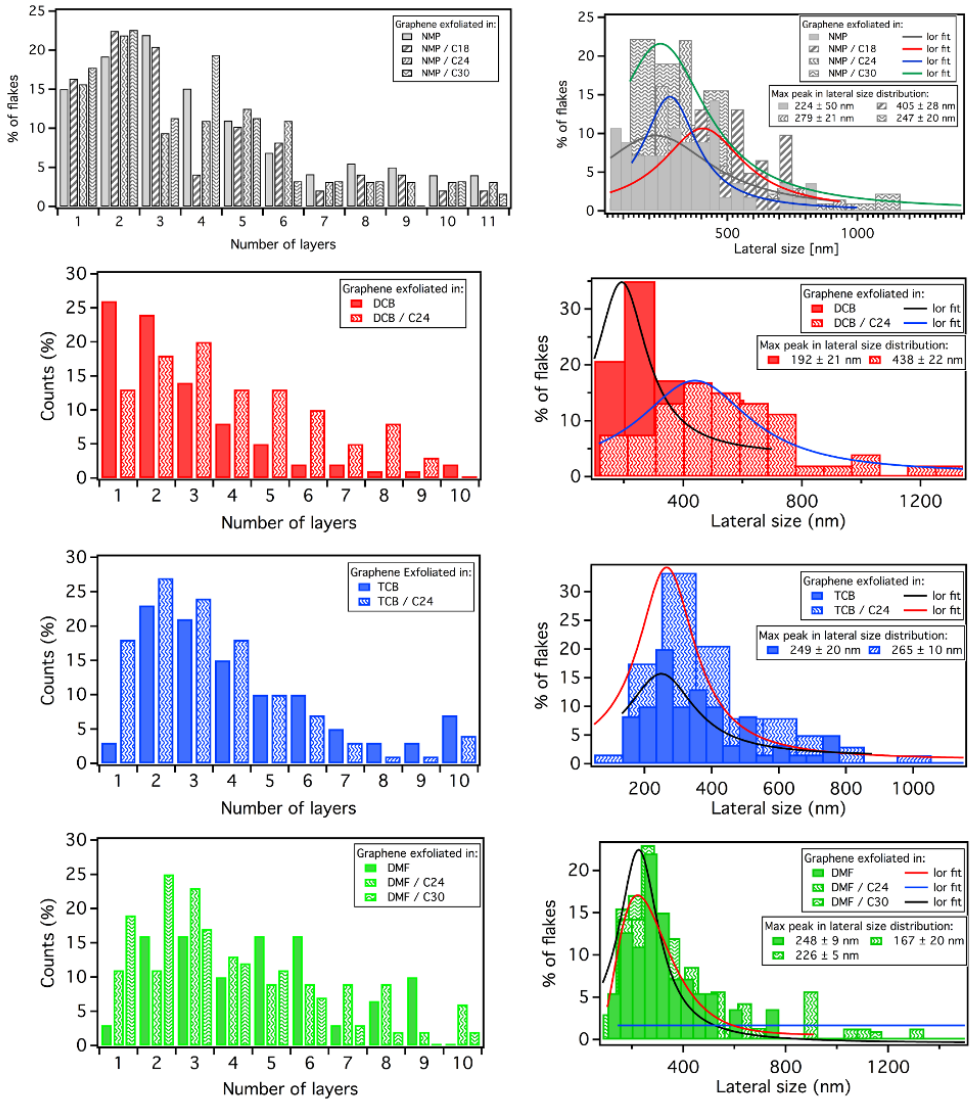


Figure 7.5: Results of the TEM analysis for the four solvents (from top to bottom: NMP, DCB, TCB and DMF) in the presence of the five acids. The left column reports the histograms of the number of layers counted in each analyzed graphene flake. The right column reports the distribution of the lateral sized for the graphene flakes.

In pure solvents (blank experiments) the largest graphene concentrations were obtained using NMP (86 ± 10 μg mL^{-1}) and o-DCB (98 ± 9 μg mL^{-1}), while both TCB and DMF produced dispersions of significantly lower concentration (64 ± 7 μg mL^{-1} and (47 ± 9 μg mL^{-1}, respectively). In the presence of the carboxylic acids the graphene concentration was found to increase almost linearly with the length of the aliphatic chain, except when the longest derivative, C30, was used; see Fig. 7.4. Only exception was DMF, in which the addition of carboxylic acid was essentially negligible and no increase in the graphene concentration was observed. In sharp contrast, when NMP was used as solvent, the graphene concentration increased significantly using longer and longer fatty acids reaching an almost three-fold increase in the exfoliation yield with C30 as

compared to the pure solvent. Using the chlorinated solvents, i.e. o-DCB and TCB, the maximum concentration was obtained with C24, with an increase of about 50% and 90%, respectively. The use of C30 in these solvents produced essentially a zero graphene concentration, which is due to the tendency of long fatty acids to form gel-like dispersions in chlorinated solvents at room temperature[232]. Despite these particular cases, all solvents show a marked increase in the graphene concentration using longer and longer fatty acids; see Fig. 7.4. These results demonstrate that it is possible to enhance the yield of graphite LPE by using molecules that form tightly packed SAMs on graphene.

The increase in graphene concentration is not the only parameter to measure. In fact, one should also check the quality of the exfoliated material and verify whether the dispersion contains single-sheet or multilayer flakes. For this purpose, two experimental techniques were used: Transmission Electron Microscopy (TEM) and micro Raman spectroscopy. In the first, drops of the exfoliated solution were deposited on a carbon grid, the solvent was evaporated and the obtained flakes were imaged. Thanks to the high resolution of the TEM, it is possible to isolate the single flakes and count the number of graphene layers in the flake. Also, it is possible to measure the dimension of the flakes, providing the average later size of the obtained flakes. By micro Raman spectroscopy, the quality of the exfoliated material is assessed by measuring the intensity ratio between the D and the G peaks, and fitting the 2D peak with a Lorentzian lineshape. These two quantities provide qualitative information on both the number of layers in the flake and the presence of defects. Analysis of the graphene dispersion obtained by LPE in the presence of the fatty acids shows that both in the absence and in the presence of surfactants the graphene flakes contain a high percentage of single and double layers graphene sheets. Also, it appears also that the shorter acids tend to reduce the average lateral size of the flakes, while with longer acids the quality improves; see Fig. 7.5.

7.4 PREDICTING THE BINDING AFFINITY FOR GRAPHENE

To provide an interpretation of the fatty acid chain length dependence of the exfoliation yield, the thermodynamic analysis based on the surface free energy γ developed in chapter 5 was used. This implies the evaluation of the free energy difference involved in the confinement of the carboxylic acids from the solution to the self-assembled monolayer on the surface. Thus, two are the forces involved: the energetic gain upon self-assembly due to the new interaction among the molecules in the SAM and with the surface, and the entropic cost due to the confinement of freely-diffusing monomers in solution to fixed positions in the SAM.

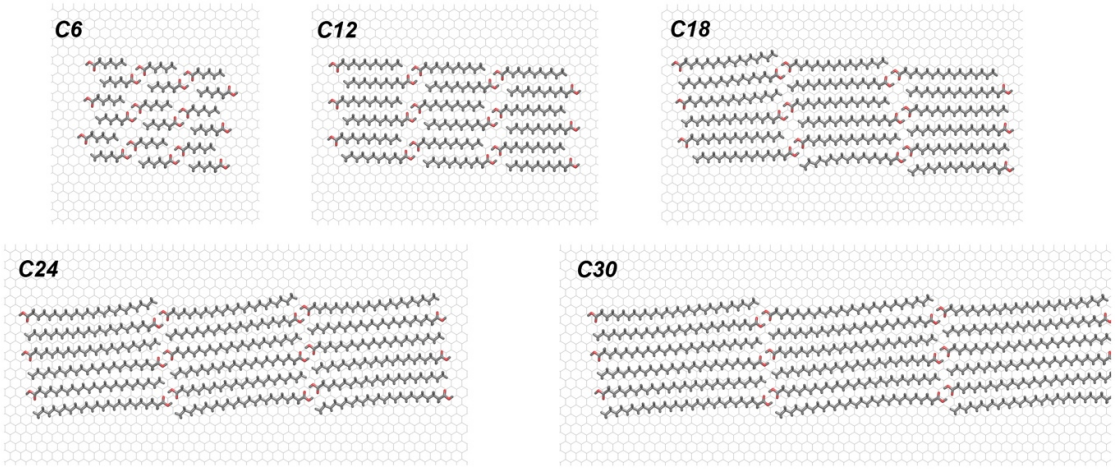


Figure 7.6: Molecular models for the five carboxylic acids.

To evaluate these two contributions Eq. 5.23 and 5.24 were used. This allows to evaluate the energetic (γ_E) and entropic (γ_S) contributions to the surface free energy of the SAM as

$$\gamma_E = \frac{1}{A'_{uc}} \left(E'_{uc} - 3kT\right) \tag{7.1}$$

$$\gamma_S = \frac{1}{A'_{uc}} T \left(S_{A,tr} + S_{A,rot}\right) \tag{7.2}$$

where A'_{uc} and E'_{uc} are the unit cell area and the unit cell energy per molecule, $S_{A,tr}$ and $S_{A,rot}$ the translational and rotational entropy of one molecule in solution. To evaluate these terms, atomistic models of the five SAMs were generated. This was done by modeling 3x3 cells (i.e. 18 molecules) SAMs, physisorbed on a graphene layer and using the CHARMM General Force Field[125] to model the interactions; see Fig. 7.6. For the calculation of the unit cell energy, the central cell was used as reference. For the graphene layer, a square 176.83 x 147.47 Å polyaromatic molecule with dangling hydrogens formed by 10874 carbons atoms was used, where carbon atoms were parametrized using the CG2R61 (aromatic carbon) atom type and the hydrogens with the HGR61 atom type (aromatic hydrogen). All atoms were set to neutral charge, apart from the hydrogens which bring a partially positive charge of 0.115 e.u., and the carbons attached to them a −0.115 e.u. partial charge. The atom typing and parametrization of the carboxylic acids were performed using the ParamChem webserver (version 0.9.7)[213,214], which generates automatically parameters for CGenFF (version 2b8). To build the 3x3 cells model SAMs, an initial guess was built displacing the 18 molecules on the graphene layer and using the experimental unit cell parameters as translational vectors. Then, the models were relaxed by energy minimization using 1000 steps of steepest descent (SD) followed by the adopted basis Newton Raphson method (ABNR) until convergence to an energy gradient of $1 \cdot 10^{-8}$ kcal mol^{-1} Å^{-1}. The CHARMM software was used for all

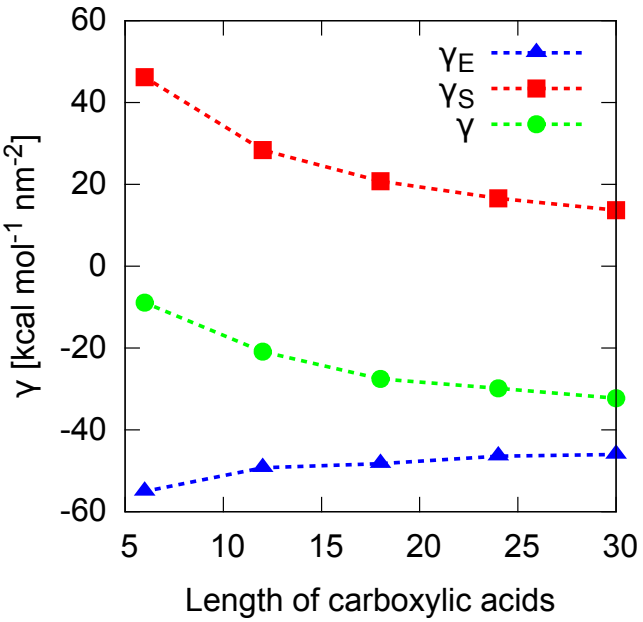


Figure 7.7: Surface free energy decomposed in energetic and entropic contributions evaluated for the five SAMs from C6 to C30.

	E_{ads}	E_{intra}	E_{inter}	E'_{uc}	$-TS_{tr}$	$-TS_{rot}$	A'_{uc}	γ_E	γ_S	γ
C6	-28.24	-3.05	-37.89	-25.12	-13.49	-9.14	0.51	-55.07	46.17	-8.90
C12	-50.31	-7.44	-46.47	-40.50	-13.99	-10.40	0.81	-49.26	28.37	-20.89
C18	-74.80	-11.85	-56.79	-57.52	-14.32	-11.19	1.22	-48.28	20.74	-27.54
C24	-94.56	-15.76	-67.15	-71.95	-14.56	-11.76	1.50	-46.42	16.55	-29.87
C30	-119.13	-20.01	-76.65	-88.73	-14.75	-12.20	1.85	-45.99	13.68	-32.30

Table 7.2: Thermodynamic analysis for the five self-assembled monolayers from C6 to C30. The energetic values are reported in kcal mol^{-1}, the surface area in nm^2, the surface free energy in kcal mol^{-1} nm^{-2}.

calculations, using the BLOCK facility to evaluate the three components of the unit cell energy (Eq. 5.14). Finally, the optimized models were confronted with the experimental STM results, in particular comparing the unit cell parameters. As shown in Tab. 7.1, the experimental and modeled unit cell parameters are in perfect agreement with the former, validating the latter.

To evaluate the entropy contribution to γ we assumed idealized solution behavior and measured the translatinal and rotational entropy loss analytically using the expression derived in chapter 3, and applied to the self-assembly process in chapter 5.

Combining the energetic (γ_E) and entropic (γ_S) contributions, the thermodynamic stability of the different SAMs can be evaluated and compared. The numerical results for the five carboxylic acids are reported in Tab. 7.2 and shown graphically in Fig. 7.7. Perhaps surprisingly, the results show that the energetically most stable SAMs, i.e. the one formed by the shortest molecules, are not the thermodynamically preferred at the experimental conditions. How the shortest molecules allows for the best energy gain

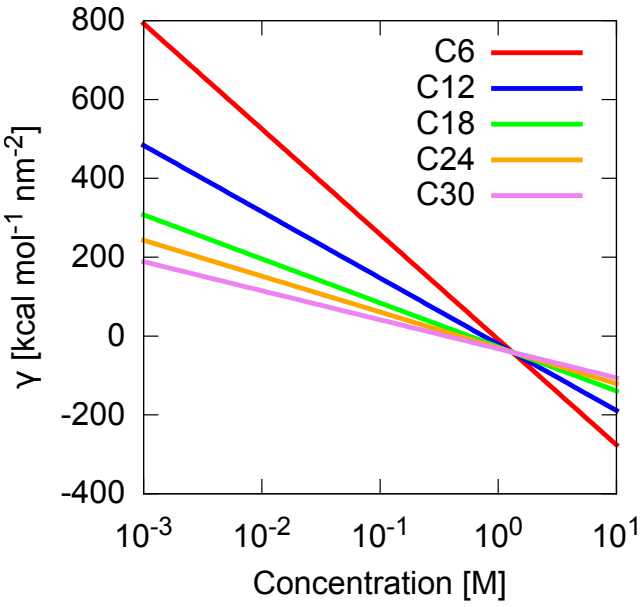


Figure 7.8: Effect of the concentration of the molecules in solution to the stability of the self-assembled monolayers. Since the stability of the shortest molecules is more strongly influenced by the entropy cost to self-assembly, due to higher number of molecules to constrain on the surface, γ changes more steeply, allowing an inversion of the stability ranking at high concentrations.

can be easily understood considering the packing of the SAMs. At a given surface area covered by the five SAMs, the shortest molecules allow a large density of hydrogen bond interactions, each one bringing a high energetic stabilization. This thus explains the observed trend in the γ_E and leaves the problem on how to justify the higher exfoliation yield for the longer molecules.

While the SAM formed by the shortest molecules allows for a higher density of hydrogen bonds, they also need to confine on the surface a higher number of molecules per unit of area. This clearly involves a high entropic cost, which is inversely proportional to the unit cell area of the SAM. Thus, short acids self-assembly involves a higher entropic cost compared to the longer ones. This is quantified by γ_S, as shown in Fig. 7.7. Summing up the energetic and entropic contributions, the thermodynamic stability of the SAMs can be accessed. Interestingly, the trend of γ in Fig. 7.7 is consistent with the exfoliation yields in Fig. 7.4, demostrating that it is crucial to capture correctly both the energetic and entropic contributions to rationalize the self-assembly. Also, these results show that methods based on purely energetic analysis of the SAM are inadequate, as the supramolecular self-assembly process is inherently dominated by an energy/entropy compensation.

As a last note, it is worth studying how γ changes as a function of the concentration of monomers in solution. In the data shown, a concentration of 10 mM was used,

which was close to the experimental one. If the same experiment were performed at higher concentrations of monomers, the entropic contribution would be expected to be less important as the translational entropy is strongly concentration dependent and is proportional to the available volume per molecule. Our thermodynamic model predicts that at higher concentrations of monomers the entropic contribution to the surface free energy of the SAM becomes negligible, making the exfoliation with the shortest acid most effective; see Fig. 7.8. In fact, the qualitative plot in Fig. 7.8 shows that at very high concentrations, γ for the short acid self-assembly is more favorable (more negative) than the one for the longest acids.

7.5 TAKE HOME MESSAGES

In this chapter is has been shown that it is possible to tune the yield of graphite LPE by mastering a supramolecular approach involving the formation of self-assembled monolayers on the top of the exfoliated graphene. Five aliphatic carboxylic acids of different length, which form self-assembled monolayers on graphite, were used as surfactants in four different solvents. In all solvents, it was shown that an increase in the chain length produces a net increase in the exfoliation yield. Noticeably, the use of C30 with NMP increased the yield by nearly 200%, obtaining a graphene concentration in solution of 240 $\mu g\,mL^{-1}$ (compared to the 85 $\mu g\,mL^{-1}$ in pure NMP). Moreover, the addition of surfactants did not reduce the quality of the exfoliated material. To rationalize these results, the stability of the SAMs formed by the five different molecules was studied by using the thermodynamic model developed in this thesis. Our modeling approach highlighted the interplay between entropy, which favors the longest molecules at the experimental conditions, and energy, which favors the shortest ones. Due to this energy/entropy interplay, it was possible to tune the relative stability of the SAMs not only changing the chemical nature of the building blocks, but also just by changing the experimental conditions, such as the concentration of the monomers in solutions. This suggests that thermodynamic model developed in chapter 5 can be effectively used to predict the stability of the SAMs and guide the chemical design of new building blocks for self-assembly.

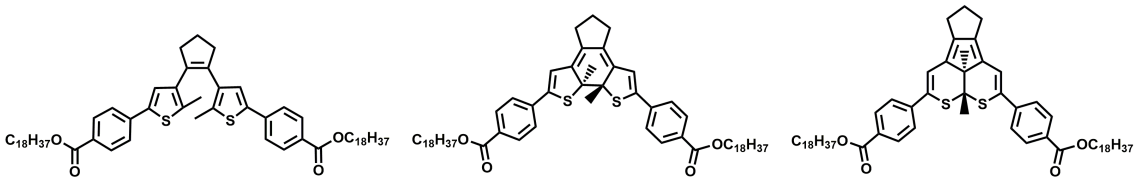
8

PHOTOSWITCHING AT THE SOLID/LIQUID INTERFACE

8.1 MOLECULAR SWITCHES AT SURFACE

In the last twenty years the fabrication of various self-assembled systems and materials possessing tunable chemical and physical properties was made possible by the increased ability in the design of sophisticated molecular building blocks. Recently, numerous efforts have been put towards the construction of functional supramolecular architectures with sub-nanometer control over their mechanical movements, aiming at the development of molecular switches and machines. On surfaces, cleverly designed molecular switches can be fueled by light, as it is the most abundant energy source in Nature. This is possible thanks to small organic photochromic molecules, which can switch, when light-irradiated, among two or more stable isomers characterized by significantly different physical properties. Among the existing photochromic systems, diarylethenes have attracted particular attention, mostly thanks to the thermodynamic stability of both their open and closed isomers, and to the very fast cyclization reaction which takes place on the picoseconds timescale[233,234]. Moreover, the photoisomerization from the open to the closed isomer have two major effects. First, their electronic properties, like the orbital energies and the excitation electronic transitions, change dramatically, making these molecules optimal for electronic devices, such as memories[235] and switches[236]. Second, the molecular structure also changes dramatically upon photoisomerization, providing an optimal playground for their reversible molecular ordering when physisorbed on surfaces.

While self-assembled monolayers (SAM) based on diarylethene molecules have already been studied by STM[237–240], the *in situ* reversible interconversion between the open and closed states has never been resolved with high-resolution imaging. In this chapter, the joint experimental and computational effort to image and understand the *in situ* photoswitching and self-assembly of a diarylethene derivative at the solid/liquid interface is presented. The focus is on the 1,2-bis(2-methyl-5-(4-octadecyloxycar-



DAE-o DAE-c DAE-b

Figure 8.1: Chemical structure of the three form of the investigated diarylethene. From the left: the open (DAE-o), closed (DAE-c), and byproduct (DAE-b).

bonylphenyl)thien-3-yl)cyclopent-1-ene diarylethene – DAE, see the chemical structure in Fig. 8.1. Thanks to the two long aliphatic chains attached in the para position of the two terminal phenil rings, the molecule is expected to have a high affinity for HOPG. As the molecule/substrate interaction is of dispersive nature, a dynamic scenario of adsorption/desorption events is expected, allowing fast reaching of chemical equilibrium conditions.

This work is the outcome of the collaboration between two experimental and one computational groups. The synthesis of the molecule was done by M. Herder and S. Hecht at the Department of Chemistry of the Humboldt-Universität zu Berlin (Germany). The photocromic characterization and STM imaging were performed by S. Bonacchi and M. El Garah, respectively, under the supervision of A. Ciesielski and P. Samorì at the Nanochemistry Laboratory, Institut de Science et d'Ingénierie Supramoléculaires (ISIS), Universitè de Strasbourg (France). I performed the theoretical and modeling analysis under the supervision of M. Cecchini. Most of the results presented in this chapter were published in *Angewandte Chemie*[190].

8.2 PHOTOSWITCHING OF DIARYLETHENES

The intrinsic photoswitching ability of the DAE molecule was verified *ex situ* using a 0.5 mM solution in cyclohexane:1-Phenyloctane (1:5). Upon UV irradiation at 310 nm, the colorless open form of DAE (DAE-o) quantitatively converted to the violet closed isomer (DAE-c). Subsequent visible-light irradiation (530 nm) converted it back to the open form in a perfectly reversible way. The switching process was monitored by bright-field microscopy, where one drop of DAE solution (4 µL, diameter ~ 6 mm) was deposited on a quartz substrate and subsequently irradiated by UV (310 nm, 1.7 mW cm^{-2}) and visible light (530 nm, 5.6 mW cm^{-2}) 2 cm over the sample; see Fig. 8.2a. To quantitative study the kinetic process, the intensity of the green RGB channel was tracked upon irradiation cycles; see Fig. 8.2b. This graph shows how the process is effectively reversible, with the DAE-c isomer formed upon short UV irradiation, and converted back to the DAE-o by visible light.

Interestingly, upon prolonged UV irradiation the violet DAE-c isomer is converted into a pinkish third isomer. This is in line with previous reports[241,242], where prolonged UV irradiation caused the annulation of the diarylethene core, inducing the formation of a photoirreversible byproduct; see the structure of DAE-b in Fig. 8.1[239,240]. As shown by both Fig. 8.2, the formation of the DAE-c is irreversible, and subsequent visible light irradiation is not able to convert it back to the DAE-o isomer.

The same process has been followed by UV/Vis adsorption spectroscopy. As shown in Fig. 8.3a, starting from the DAE-o (black solid line) upon UV irradiation the molecule converts to the DAE-c (violet solid line). Then, upon visible light irradiation, the DAE-c converts back to the DAE-o (dashed spectra). Notably, as the isosbestic point at 363 nm

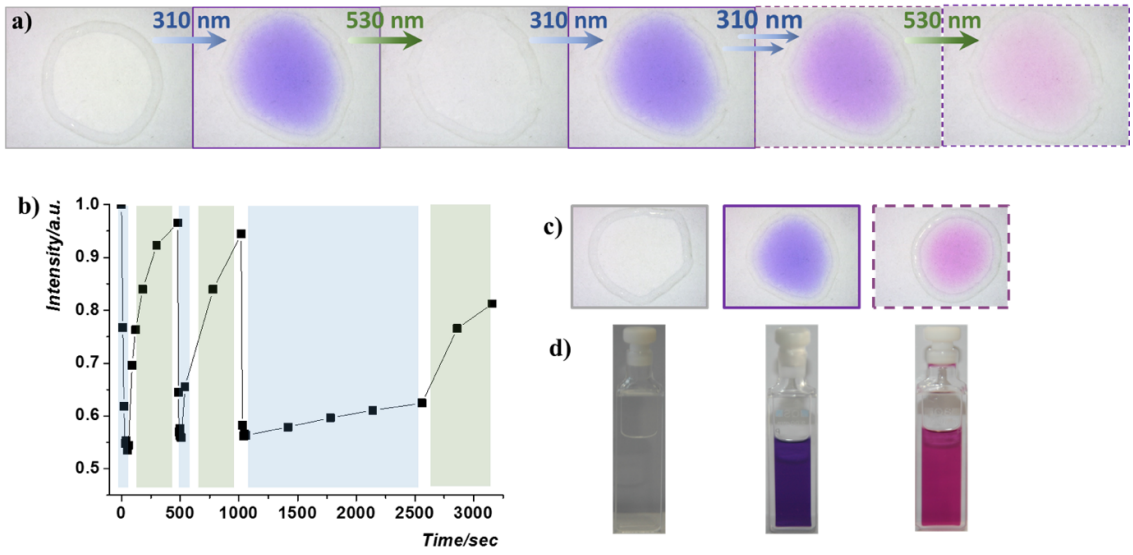


Figure 8.2: a) Bright-field images of a DAE drop on a quartz substrate during two UV/Visible irradiation cycles, with the formation of the DAE-b upon prolonged UV irradiation. b) Normalized intensity of the green RGB channel of the DAE drop during the photoconversion cycles in a). c) Bright-field images of DAE-o, DAE-c, DAE-b drops. d) Photos of DAE-o, DAE-c, DAE-b solutions in 2 mm quartz cuvettes.

remains constant during the whole irradiation cycle, the appearance of side products in the short (several minutes) experiment time is ruled out. In Fig. 8.3, an analogous study is reported for the irreversible conversion to the DAE-b. Starting from the DAE-o (black line), a first UV irradiation brings it to the DAE-c isomer (violet line). Subsequent long UV irradiation (~ 10 hours) leads to the formation of the byproduct DAE-b (pink line). The new isosbestic point at 470 nm is stable during this second reaction.

8.3 IN SITU SWITCHING IMAGED BY STM

The switching capabilities of the DAE molecule was studied at the solid/liquid interface (*in situ*). This was done by imaging the self-assembled monolayers formed by the DAE isomers at the HOPG/1-phenyloctane interface, and analyzing the changes in the SAMs induced by UV or visible light irradiation. This was done by depositing one drop of DAE solution on HOPG under the STM and irradiating the DAE using a led positioned over the solution drop.

Initially, the self-assembly of DAE-o was investigated. The STM image of the corresponding SAM (Fig. 8.4a) showed a lamellar crystalline architecture characterized by unit cell parameters $a = 1.5 \pm 0.1$ nm, $b = 3.4 \pm 0.1$ nm, and $\alpha = 79 \pm 2°$, leading to an unit cell area of $A = 5.0 \pm 0.1$ nm^2, with one molecule per unit cell. Due to the steric hindrance between the two thienyl groups, the cores of the DAE-o are only partially physisorbed on the surface, and the molecule adopts a U-shape conformation having

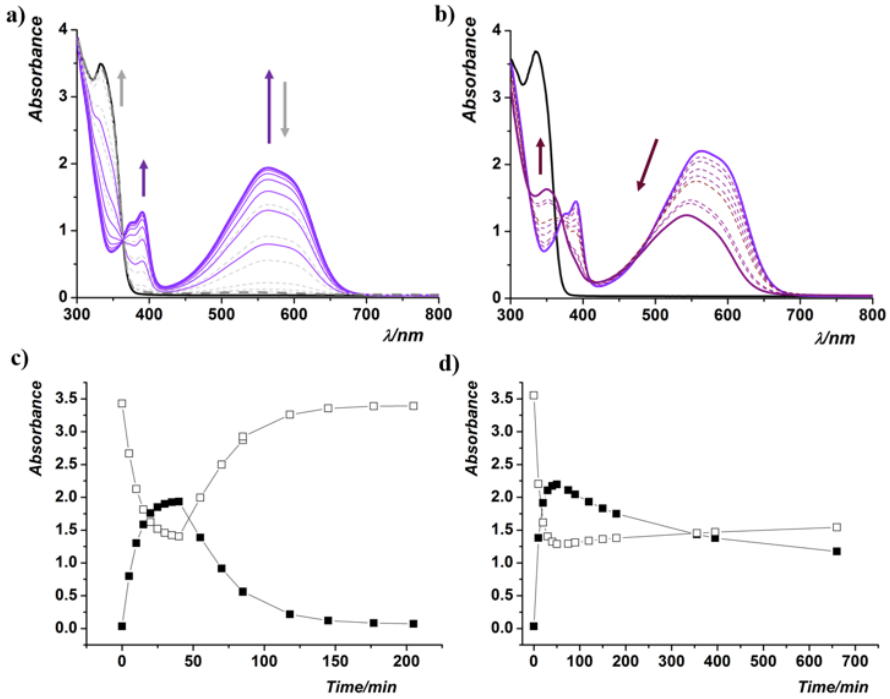


Figure 8.3: Photochromic behavior of DAE in solution followed by UV/Vis absorption spectrocscopy. In a) reversible isomerization of the DAE-o (solid black) to the DAE-c (solide violet) by UV light, and back to the DAE-o (dashed black) by visible light. In b) irreversible conversion by UV light of the DAE-o (solid black) to the DAE-c (solid violet) and then to the DAE-b (solid pink). In c) and d) absorption changes at 330 nm and 560 nm (white and black squares, respectively) as a function of the irradiation time in the a) and b) studies.

the octadecyl side chains pointing parallel in the same direction, although they are not clearly resolved with the STM.

Then, the switching to the DAE-c was tested by irradiating with UV light (310 nm) for 5 min. Since only the antiparallel conformation of the DAE is photoactive[243,244], ring-closure cannot happen in the monolayer, such that the molecule needs to desorb, photoisomerize in solution, and reassemble on the surface. UV irradiation produced a second SAM exhibiting unit cell parameters $a = 1.0 \pm 0.1$ nm, $b = 4.1 \pm 0.1$ nm, and $alpha = 79 \pm 2°$, leading to an unit cell area $A = 4.0 \pm 0.1$ nm^2 with one molecule per unit cell. STM imaging of this second SAM (Fig. 8.4b) shows the occurrence of three bright spots in line, perpendicular to the lamellar axis. These can be assigned to the two phenyl rings, and the center one to the core of the DAE-c, which is brighter due to the extended π-conjugation. Importantly, if this second SAM is irradiated by visible light (530 nm), the first SAM characteristic of the DAE-o is observed, indicating that the DAE-c is able, upon visible light irradiation, to desorb from the HOPG, isomerize back to the DAE-o and re-adsorb into the DAE-o monolayer. Noteworthy, if the switching is performed *ex situ* and one drop of the DAE-o or DAE-c are deposited on the HOPG surface, the exactly same patterns as in the *in situ* experiment are observed; see Tab. 8.1.

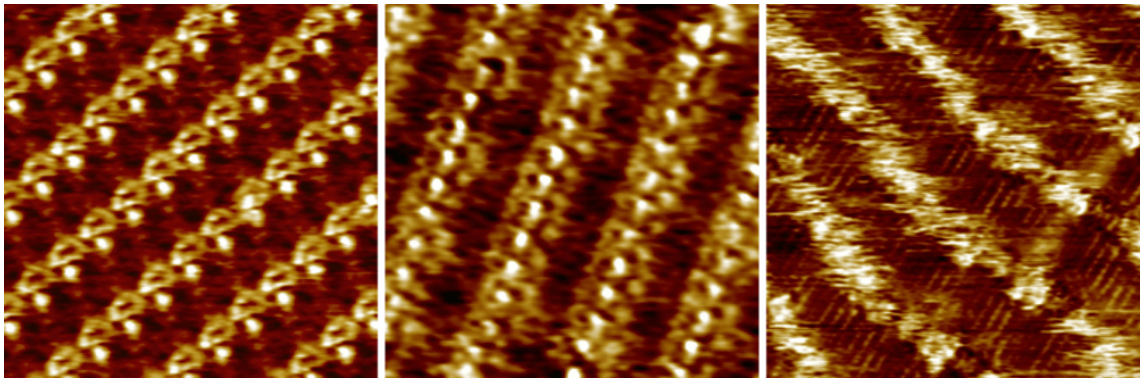


Figure 8.4: Height STM images obtained for the self-assembled monolayer of the DAE-o (left), DAE-c (center), and DAE-b (right) at the HOPG/1-phenyloctane interface.

Structures		a [nm]	b [nm]	α [°]	A [nm^2]
DAE-o	*in situ*	1.5 ± 0.1	3.4 ± 0.1	78 ± 2	4.9 ± 0.1
	ex situ	1.5 ± 0.1	3.4 ± 0.1	79 ± 2	5.1 ± 0.1
	model	1.5	3.4	76	4.9
DAE-c	*in situ*	1.0 ± 0.1	4.1 ± 0.1	78 ± 2	4.1 ± 0.1
	ex situ	1.1 ± 0.1	4.1 ± 0.1	78 ± 2	4.4 ± 0.1
	model	0.9	4.2	78	3.8
DAE-b	*in situ*	0.9 ± 0.3	4.0 ± 0.5	79 ± 2	3.5 ± 0.5
	ex situ	0.8 ± 0.3	4.0 ± 0.5	78 ± 2	3.1 ± 0.4
	model	0.9	3.9	78	3.4

Table 8.1: Unit cell parameters for the three DAE isomers measured experimentally *in situ* and *ex situ*, and obtained by modeling.

After prolonged UV irradiation of either the DAE-o or DAE-c monolayers, a third different 2D architecture was observed, exhibiting unit cell parameters $a = 0.9 \pm 0.3$ nm, $b = 4.0 \pm 0.5$ nm, $\alpha = 79 \pm 2°$ and one molecule per cell, leading to an unit cell area $A = 3.5 \pm 0.5$ nm^2; see Fig. 8.4. This SAM can be assigned to the byproduct of the photoisomeration reaction (DAE-b). Noteworthy, the interdigitated alkyl chains are clearly visible in this SAM, along with brighter areas that can be assigned to the polycyclic core of the molecule. Markedly different from the previous SAMs, this image shows a fuzzy contrast, suggesting that the core of the DAE-b does not interact strongly with the surface, but it is most likely positioned out of the plane of the alkyl chains. Irradiation of this third SAM by either UV or visible light does not alter the self-assembled architecture, confirming that this SAM is formed by the DAE-b isomers and cannot be switched back to the open or closed monolayers. As last test, the SAM obtained by the *in-situ* irradiation was compared with the one obtained by using a drop of the solution produced by *ex-situ* photoisomerization. As shown in Tab. 8.1, the two SAMs match perfectly.

The most interesting feature of this experiment is that preparing *ex-situ* a solution containing the byproduct in very small quantities (DAE-c : DAE-b = 92 : 8), and imaging it at the STM, only one SAM corresponding to the byproduct was observed. This indicates a high propensity to self-assembly of the DAE-b isomer with respect to the DAE-c.

8.4 MODELING THE SELF-ASSEMBLY

In order to better understand why the SAM produced by the byproduct was so stable relative to the closed or the open forms, we performed a thermodynamic analysis of the monolayers produced by the DAE-o, DAE-c and DAE-b by evaluating their surface free energy γ as described in chapter 5.

To this aim, molecular models for the three SAMs were produced using an atomistic description of the graphite substrate and the molecules, and the classical General Amber Force Fields (GAFF)[123] to model the interactions. For each DAE photoisomer, a self-assembled architecture of 4x4 cells was generated arranging the molecules to reproduce as close as possible the experimental STM images. As a first step, the configuration of the molecule inside the unit cell was modeled based on the position of the brights spots in the STM images. The collected information include:

- the conformation of the core of the molecule: linear in the DAE-b and DAE-c, U-shaped in the DAE-o;
- the alignment and conformation of the octadecyl chains: parallel in-plane zig-zag in the DAE-o, anti-parallel in-plane zig-zag in the DAE-c, anti-parallel out-of-plane zig-zag in the DAE-b;
- the correlation of the unit cell parameters obtained experimentally and in the model.

Once a reasonable configuration of the unit cell was obtained, a 4x4 supramolecular architecture was built by replicating the unit cell onto the xy plane and using the experimental unit cell parameters as translational vectors. Then, the structures were energy minimized by 1000 steps of steepest descent (SD) followed by an adopted basis Newton Raphson (ABNR) optimization until convergence to an energy gradient of 10^{-8} kcal mol^{-1} Å^{-1}. The resulting architectures were then compared with the experimental STM images, focusing on the modeled unit cell parameters. If the model was in broad disagreement with the STM image, the initial configuration of the unit cell was manually modified and used to start a new cycle of modeling until a qualitative agreement with the experiments was found. All these manipulations and energy minimizations were performed using the CHARMM software[144]. Remarkably, the model SAMs in Fig. 8.5 are in very good agreement with the experimental unit cell parameters; see Tab. 8.1. Starting with the optimized self-assembled monolayers shown in Fig. 8.5, the unit cell energy E_{uc} was calculated as a sum of the molecule-molecule interaction in the layer $E_{inter}/2$ and the adsorption energy E_{ads}. To minimize border effects, the unit cell energy was averaged over the four internal cells. Then, the surface free energy γ of the three SAMs was evaluated using the simplified expressions (see chapter 5

$$\gamma = \frac{1}{A_{uc}} \left[E_{uc} - 3RT + T \left(S_{tr} + S_{rot} \right) \right] \tag{8.1}$$

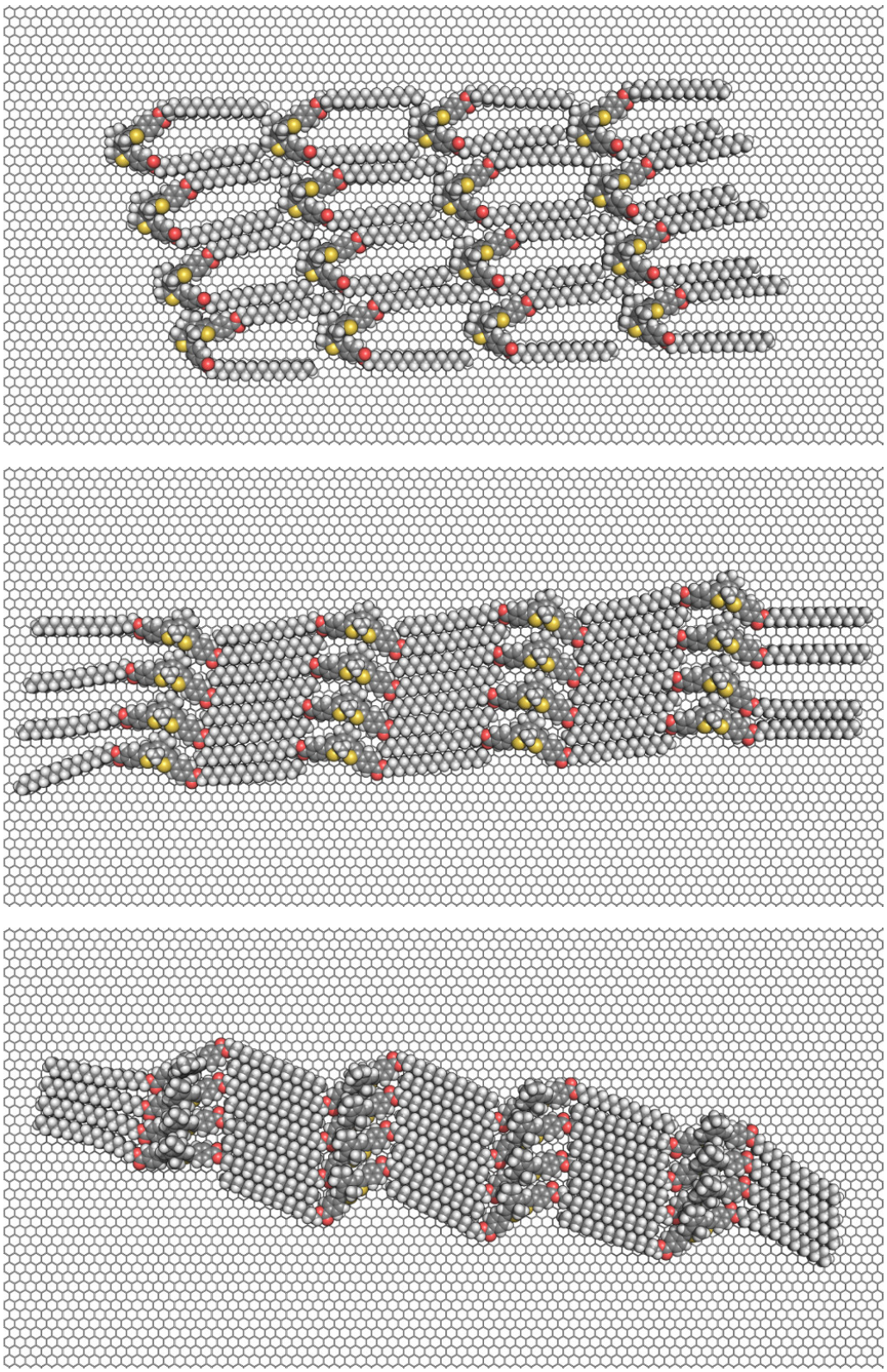


Figure 8.5: Optimized models of the 4x4 cells self-assembled architectures for the open (top), closed (center), and byproduct (bottom) isomers of DAE.

	E_{ads}	$E_{inter}/2$	γ_E	γ_S	γ
DAE-o	-119.1	-15.7	-26.38	5.34	-21.04
DAE-c	-112.7	-32.6	-38.14	7.31	-30.83
DAE-b	-101.3	-46.1	-42.60	7.99	-34.61

Table 8.2: Thermodynamic analysis of the self-assembled monolayers generated by the three isomers of the DAE. The adsorption (E_{ads}) and interaction ($E_{inter}/2$) energies are reported in kcal mol^{-1}, while the value of the surface free energy (γ), and its decomposition in energetic (γ_E) and entropic (γ_S) terms, are reported in kcal mol^{-1} nm^{-2}.

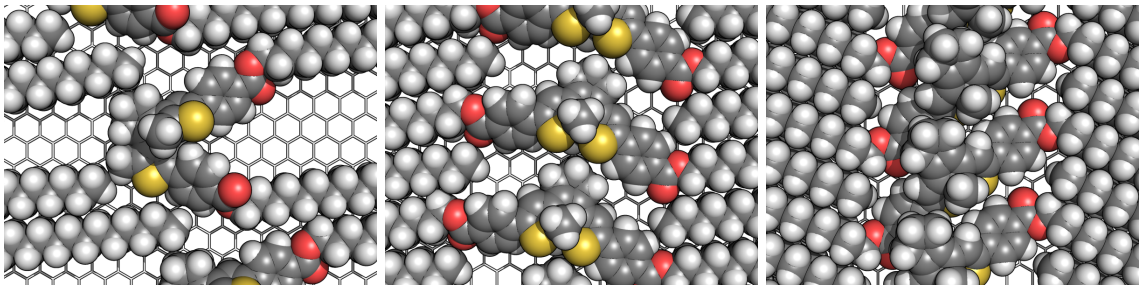


Figure 8.6: Magnification of the molecular models generated for the DAE-o (left), DAE-c (center), and DAE-b (right). Note how the alkyl chains are parallel in the DAE-o and antiparallel in the DAE-c and DAE-b, how the DAE-b is more dense than the DAE-c thanks to the different alkyl chain conformations, and the overlap id the cores of the molecules which is present only in the DAE-b.

which neglects the vibrational contribution The translational and rotational entropies were evaluated analytically in the limit of the ideal gas approximation (see chapter 3) at a temperature of 300 K and a concentration of 0.5 mM, which are consistent with the experimental conditions. The calculated values of the unit cell energy, the entropy and the surface free energy for the three SAMs are reported in Tab. 8.2.

The results show that the DAE-b architecture is the most stable in the experimental conditions with $\gamma = -34.61$ kcal mol^{-1} nm^{-2}, followed by the DAE-c and DAE-o in this order; see Tab. 8.2. Trying to understand why, the DAE-o isomer self-assembles in a porous architecture where the cores of the molecule are face-on and are not interacting each others. The two alkyl chains only partially interact through weak dispersive interactions, preserving a certain mobility and preventing an accurate imaging at the STM. Overall, the total interaction energy between molecules in the SAM, $E_{inter}/2$ is only -15.7 kcal mol^{-1}, that is less then half of the interaction energy for the DAE-c and DAE-b. Moreover, the DAE-o is about 30% less compact than the other two isomers, resulting in an overall small energetic gain distributed on a big surface area. This makes the DAE-o assembly the less favored of the three SAMs.

Comparing the DAE-c and DAE-b assemblies, both are dense. The byproduct is actually the preferred one as result of the compensation of two opposing contributions. Looking just at the unit cell energy, the byproduct is disfavored by 11.4 kcal mol^{-1} *per molecule* if compared to the DAE-c. At the same time, the byproduct is slightly more dense than the closed isomer, which implies a more favorable interaction energy *per unit*

area of 4.5 kcal mol^{-1} nm^{-2}. As shown in Fig. 8.6, two are the main differences between the byproduct and open SAM. First, the core of the molecule is face-on in the DAE-c assembly, whereas it is edge-on in the DAE-b. Second, the alkyl chains are physisorbed in an in-plane zig-zag fashion in the former, while they adopt an out-of-plane conformation in the latter, which involves a rotation by 90°. This out-of-plane zig-zag conformation of the alkyl chains in the DAE-b is energetically less favored, and it is thus the cause of the lower unit cell energy, but allows a more compact packing of the chains, decreasing the unit cell area by about 10%. Moreover, the entropic cost of self-assembly for the DAE-b and DAE-c is essentially the same ($\Delta\gamma_S < 1$ kcal mol^{-1} nm^{-2}), leaving the DAE-b as the thermodynamically preferred SAM by about 3.8 kcal mol^{-1} nm^{-2}. Note that when the alkyl chains are artificially rotated in either of the two SAMs, models can be created, but when their surface free energy is calculated higher values are obtained making them less stable then when the "natural" alkyl chain conformation is used.

Finally, the atomistic molecular models generated for the DAE SAMs and the used theoretical framework are here shown to be able to rationalize all STM observations and to provide new insights in the nature of the selection and amplification of the byproduct induced by the self-assembly on the graphite surface.

8.5 SURFACE MOBILITY

Intrigued by the higher mobility of the DAE-b SAM with respect to the DAE-c, as highlighted by the fuzziness of the STM images in Fig. 8.4, a more depth study of the surface diffusion properties of the DAE-c and DAE-b was done, focusing in particular on the energy barriers for the diffusion on a model graphene surface. As the surface mobility is determined by the strength of the interaction between the molecule and the substrate, and this is dominated here by the adsorption energy of the two octadecyl chains, the 2D diffusion properties of a C18 chain adsorbed in the in-plane (DAE-c) and out-of-plane (DAE-b) zig-zag configurations were investigated. Fig. 8.7 shows two energy-minimized conformations, whereas Fig. 8.8 (left) reports the adiabatic energy profiles obtained by displacing the molecule along the x and y direction on the surface. These profiles show that the adsorption of the in-plane zig-zag configuration is more favorable, having a deeper energy minimum of about 6 kcal mol^{-1}. Moreover, the energy barriers to diffuse to neighbor minima along x and y are 2.34 and 2.80 kcal mol^{-1}, respectively. These barriers are four and two times higher those corresponding to 2D diffusion in the out-of-plane zig-zag configuration, i.e. 0.62 and 1.22 kcal mol^{-1}. Interestingly, these value are as small as the thermal energy at 300 K ($RT = 0.56$ kcal mol^{-1}), suggesting that the diffusion of one molecule in the out-of-plane conformation is free, while for the in-plane conformation it is more hindered.

To model the surface diffusion of a monolayer, the height of the 2D barriers were estimated using supramolecular assemblies made of two, five and ten molecules arranged

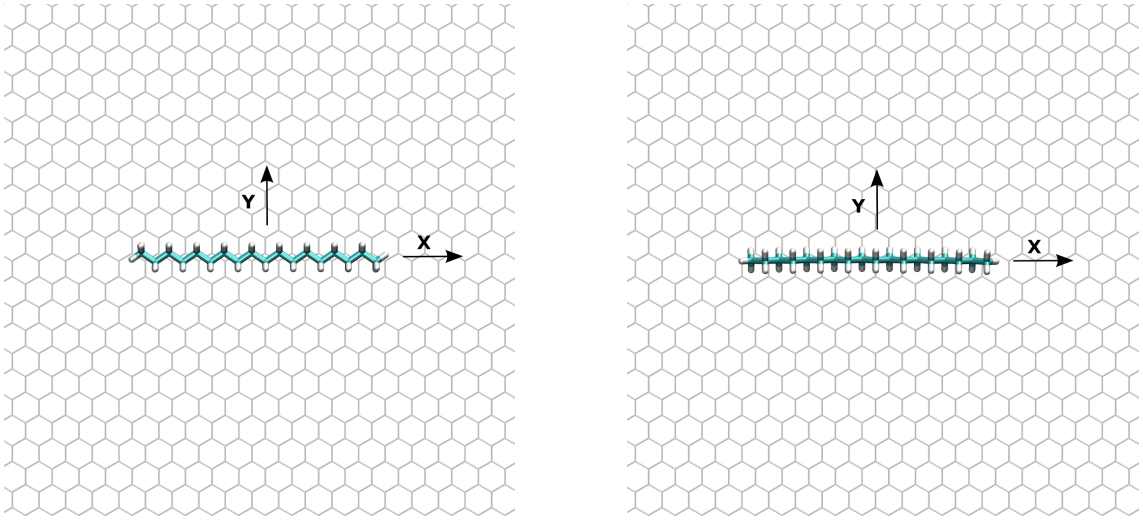


Figure 8.7: Minimized in-plane and out-of-plane conformation of octadecyl chain on graphene. The arrows show the direction of the displacements used to build the energy profiles and evaluate the energy barriers.

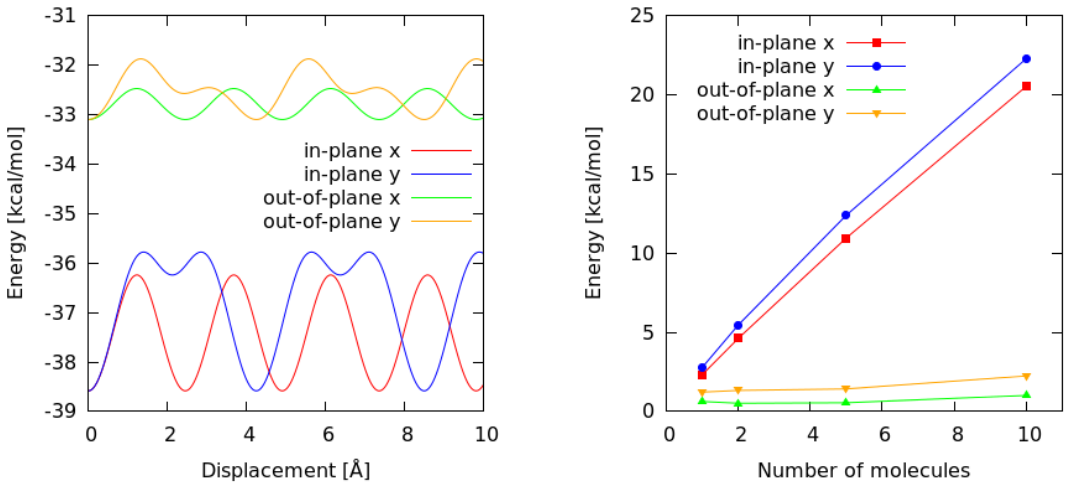


Figure 8.8: Left) Energy profiles obtained by displacing a C18 alkyl chain on a graphene surface along the x and y directions when adsorbed in the in-plane and out-of-plane zig-zag conformations. Right) Height of the energy barriers involved in the displaced in the x and y direction of a cluster of 1, 2, 5, and 10 molecules adsorbed in a in-plane and out-of-plane zig-zag conformation.

side by side. The corresponding energy profiles (see Fig. 8.8 right), clearly show that the energy barriers increase almost linearly with the number of molecules for the in-plane zig-zag conformation, and they are as high as 22 kcal mol^{-1} for a decamer. In sharp contrast, the diffusion barriers in the out-of-plane configuration appears to be essentially independent on the number of packed molecules, and they remain comparable with the thermal energy at 300 K. Thus, we conlude that the surface mobility of multimeric assemblies of alkyl chains in a out-of-plane zig-zag configuration (Fig. 8.7, right) is independent on the cluster size and the barriers are always comparable to the thermal fluctuations at room temperature.

To verify this hypothesis two Langevin dynamics (20 ns, 300 K, friction coefficient 1 ps^{-1}) were performed on a cluster of five molecules in the in-plane and out-of-plane zig-zag conformation physisorbed on a graphene surface, and constrained to keep the same conformation by the application of an harmonic RMSD constrain on the geometry of the minimum (force constant 10 kcal mol^{-1} $Å^{-2}$). The molecular dynamics trajectories demonstrate that the surface mobility of the out-of-plane ziz-zag cluster is significantly higher than the one of the in-plane zig-zag cluster. In fact, whereas the out-of-plane cluster is freely diffusing on the surface, the motion of the in-plane cluster is strongly hindered and "jumps" between one minimum to another are unlikely (only two events were observed in the 20 ns simulation).

This surprising behavior can be explained by comparing the periodicity of the in-plane and out-of-plane conformations with the one of the underlying graphene layer. The periodicity of the graphene layer can be represented as an orthorhombic cell with translational vectors of 2.46 Å on the x direction and 4.25 Å on the y direction. While the distance between two CH_2 groups along the x direction is 2.55 Å in both the in-plane and out-of-plane conformation, the space occupied by two alkyl chains along the y direction is 8.7 Å for the former and 7.85 Å for the latter, i.e. the out-of-plane conformation is more compact. Interestingly, the periodicity in the y direction for the in-plane conformation is approximately twice the periodicity of graphene. In sharp contrast, the dimension of the out-of-plane conformation in the y direction is not an integer multiple of the graphene periodicity. This implies that the periodicity of graphene and a monolayer of alkyl chains in the out-of-plane zig-zag configuration are not commensurate. As a result, the almost perfect additivity of the energy barriers in the in-plane conformation (see Fig, 8.8 right), since all molecules can reside at the same time in the ideal energy minimum, disappears in the out-of-plane conformation, causing cancellation of the difusion barrier in 2D by a dynamic compensation of favorable and non-favorable interactions with the surface.

Thus, the fuzziness observed in the STM image of the DAE-b and not observed for the DAE-c and DAE-o isomers can be explained in terms of commensurability between the adsorbed monolayer and the underlying surface, and the corresponding barrier addition/cancellation. This analysis provides new insight on the understanding of molecular mobility at surfaces telling that the adsorption energy is only one factor, but most im-

portant is the commensurability of the unit cells. This can lead to a rational design of 2D monolayers with tunable friction.

8.6 TAKE HOME MESSAGES

To conclude, in this chapter the photocromic and self-assembly properties of a newly designed and synthesized diarylethene derivative have been shown. The DAE derivative here used exhibits high propensity to self-assembly on the graphite surface in all its isomers, thus providing evidence for the reversible *in situ* photoswitching at the solid/liquid interface between an open and closed forms observed by high definition STM imaging for the first time. Prolonged irradiation of both the solution or the "wet" film has been shown to bring to the irreversible formation of a third isomer, considered to be a byproduct of the reaction, which shows highly affinity for the underlying graphite substrate. The application of the theory developed in this thesis, exemplified by the calculation of the surface free energy for the SAMs generated by the three DAE isomers, provided evidence for the ability of the graphite surface of acting as a selector, amplifying on the surface the presence of a constituent (the DAE-b) from mixtures containing it only in small percentages. Most interesting from a theoretical point of view, the DAE self-assembly is dominated by the energy component of the surface free energy. The entropy differences are instead essentially negligible, which is due to the small differences in the unit cell areas for the three isomers.

9

ENHANCING THE SELF-ASSEMBLY PROPENSITY BY PERCHLORINATION

9.1 TUNING DISPERSIVE INTERACTIONS

Dispersive interactions play a critical role in 2D self-assembly physisorbed at the solid-liquid interface. Full understanding of these relative weak and additive intermolecular forces remains an outstanding challenge[245], even if the first derivation of their C_6/R^6 nature was derived by Fritz London almost a century ago[246]. In the meantime, dispersive interactions are ubiquitous in Nature, where they play critical roles at all scales, from the structure of the DNA double helix[247] to a gecko's walk[248].

On the road towards rational design of 2D self-assembly, the ability to harness dispersive interactions by chemical tailoring of the molecular components is of major importance. Milestone contributions in this direction primarily came from Temperature Programmed Desorption (TPD) experiments, where the binding energy of a molecule on a surface can be measured quantitatively. Early studies involved the desorption of linear alkanes of different lengths. These works showed that the binding energy increases almost linearly with the length of the molecules up to decane[249], where each methylene group contributes by about 2 kcal mol^{-1} [200]. When longer alkanes were studied it was found that at the desorption temperature significant contributions from the internal degrees of freedom of the adsorbed state actually reduce the binding energy per methylene, resulting in a non-linear dependence of the desorption barrier on the length of the alkane[198]. When the same analysis was performed on Polycyclic Aromatic Hydrocarbons (PAH) of different size, like benzene, naphthalene, coronene and ovalene, it was observed a linear dependence of the desorption barrier up to more than thirty carbon atoms, thanks to the rigidity of the adsorbate, albeit the interaction energy per carbon atom decreases to 1.2 kcal mol^{-1} [210]. Finally, TPD of linear alkanes substituted by a carboxylic group or a bromine atom demonstrated that the desorption energy can be modulated by chemical functionalization, resulting in a significant increase of the surface binding affinity[250].

Apart from these experimental studies, computational approaches have also been used to model dispersive interactions on nanostructured carbon materials[135]. Recently, both dispersion-corrected density functional theory (DFT)[207,251], semiempirical quantum mechanics[208] and force field calculations[178] were shown to accurately reproduce the desorption energy of small molecules from graphene, in quantitative agreement with TPD data.

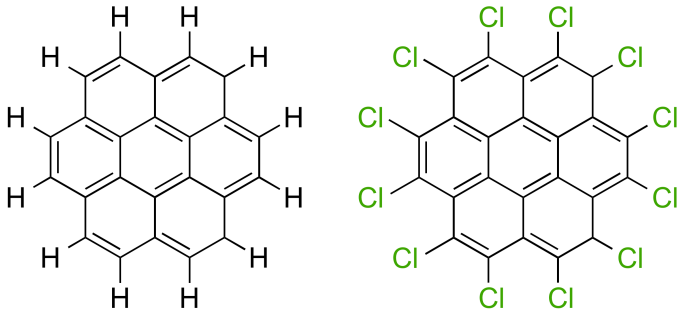


Figure 9.1: Molecular structures of coronene ($C_{24}H_{12}$) on the left and perchlorocoronene ($C_{24}Cl_{12}$) on the right.

In the following I show how both the adsorption affinity and the 2D self-assembly propensity of a prototypical PAH, coronene, can be enhanced by chemical functionalization, in particular edge perchlorination. First, the binding energy of the hydrogenated ($C_{24}H_{12}$) and perchlorinated ($C_{24}Cl_{12}$) coronene on graphite (see Fig. 9.1) were calculated by molecular modeling and compared with the desorption barriers obtained experimentally by TPD. Then, the self-assembled monolayers of these two molecules were modeled, and their surface free energies compared. Last, the improved self-assembly propensity of perchlorocoronene was exploited in the liquid phase exfoliation of graphite, where both coronene and perchlorocoronene were used as surfactants to increase the exfoliation yield.

This work results from the synergy of six groups. M. Cecchini and myself performed the theoretical and computational analysis. X. Feng and K. Müllen from the Max Plank Institute for Polymer Research (Mainz, Germany) synthesized the perchlorocoronene. The experimental determination of the desorption energies from graphite by TPD was performed by J. Weippert, A. Böttcher, and M. M. Kappes at the Institute of Physical Chemistry, Karlsruhe Institute of Technology (Karlsruhe, Germany). M. G. del Rosso, A. Ciesielski, and P. Samorì at the Nanochemistry Laboratory (ISIS & icFRC, Université de Strasbourg & CNRS, Strasbourg, France) carried out the liquid-phase exfoliation experiments. G. Melinte and O. Ersen at Institut de Physique et Chimie des Matériaux de Strasbourg (IPCMS, Strasbourg, France) characterized the exfoliated material by HR-TEM, while Y. Shin and C. Casiraghi at the School of Chemistry (University of Manchester, Manchester, UK) used micro Raman spectroscopy. This was was published in *ChemPhysChem*[189].

9.2 CALCULATING THE ADSORPTION ENERGY ON GRAPHENE

First, the binding energy to graphite was evaluated by modeling at thesingle-molecule level. Following the same methodology described in chapter 6, the desorption energies of $C_{24}H_{12}$ and $C_{24}Cl_{12}$ were computed using a force field (FF) and a semiempirical quantum mechanics (SQM) models of interactions. The theoretical methods were the

CHARMM General Force Field (CGenFF)[125] and the dispersion-corrected PM6-DH+ semiempirical method[108], which were shown to accurately reproduce the desorption energy to graphite when compared to TPD data; see chapter 6.

For the force field calculations, the atom typing was performed *via* the ParamChem web server (v. 1.0.0)[213,214]. The graphite substrate was modeled as a square polyaromatic molecule of 610 carbon atoms with dangling hydrogens at the edges. The carbon atoms of the substrate were modeled using the atom type CG2R61 (aromatic carbon) with zero charge when in the middle of the substrate and -0.115 e.u. at the edges. Hydrogen atoms were parameterized with the atom type HGR61 (aromatic hydrogen) with a partial charge of $+0.115$ e.u.To calculate the desorption energy, the adsorbate in complex with the graphene model was energy minimized by 1000 steps of Steepest Descent (SD) followed by an Adopted Basis Newton-Raphson (ABNR) optimization until convergence to an energy gradient of 10^{-8} kcal mol^{-1} Å^{-1}. The CHARMM software[144] was used for all force fields calculations. At the semiempirical level, the MOPAC2012 software[212] was used, which implements the dispersion-corrected PM6-DH+ method. In this case, the substrate was modeled by a hexagonal graphene of 216 carbon atoms with dangling hydrogens. In both cases, the desorption energy was evaluated by subtracting the energy of the adsorbate and the substrate in isolation from that of the complex upon geometry optimization. At this point, computational methods based on higher levels of theory were not considered, as the size of substrate/adsorbate complex is too big for *ab initio* methods, and the accuracy of DFT, even if dispersion-corrections are used, may be dependent on the functional used[207].

The computational results are reported in Tab. 9.1. Both methods (CGenFF and PM6-DH+) consistently predicted a significant enhancement of the adsorption energy upon perchlorination, with an increase of the desorption barrier of 28.0 kcal mol^{-1} using PM6-DH+, and of 9.1 kcal mol^{-1} with CGenFF. This striking discrepancy can be explained by the significantly different geometry of perchlorocoronene on graphene obtained with the two methods: while it adopts a fully planar configuration at the semiempirical level, a configuration with pairs of peri-related chlorine atoms flipped up and down relative to the plane of the molecule is predicted by the force field; see Fig. 9.2. The non-planar conformation predicted by CGenFF is consistent with the X-ray structure of perchlorocoronene[252] and other extended perchlorinated PAHs[253]. Noteworthy, as perchlorocoronene in the bound state assumes a non-planar conformation, edge perchlorination gives rise to a competition among the strength of interaction with the surface and the distortion of the molecule to accommodate the more bulky substituents. Consistently, when $C_{24}Cl_{12}$ is optimized in isolation, an unfavorable strain energy of 2.1 kcal mol^{-1} is calculated.

The discrepancy between the computational results at the FF and SQM level of the desorption energy is associated to the inability of the SQM method to correctly quantify the strain energy involved in the planarization of perchlorocoronene in interaction with the

Method	E_{des} $C_{24}H_{12}$	E_{des} $C_{24}Cl_{12}$	Δ
PM6-DH+ (SQM)	41.4	69.4	+28.0
CGenFF (FF)	36.1	45.2	+9.1
TPD (Exp.)	36.9	41.5	+4.6

Table 9.1: Calculated and experimental values of the desorption energy E_{des} for coronene and perchlorocoronene. All values are expressed in kcal mol^{-1}.

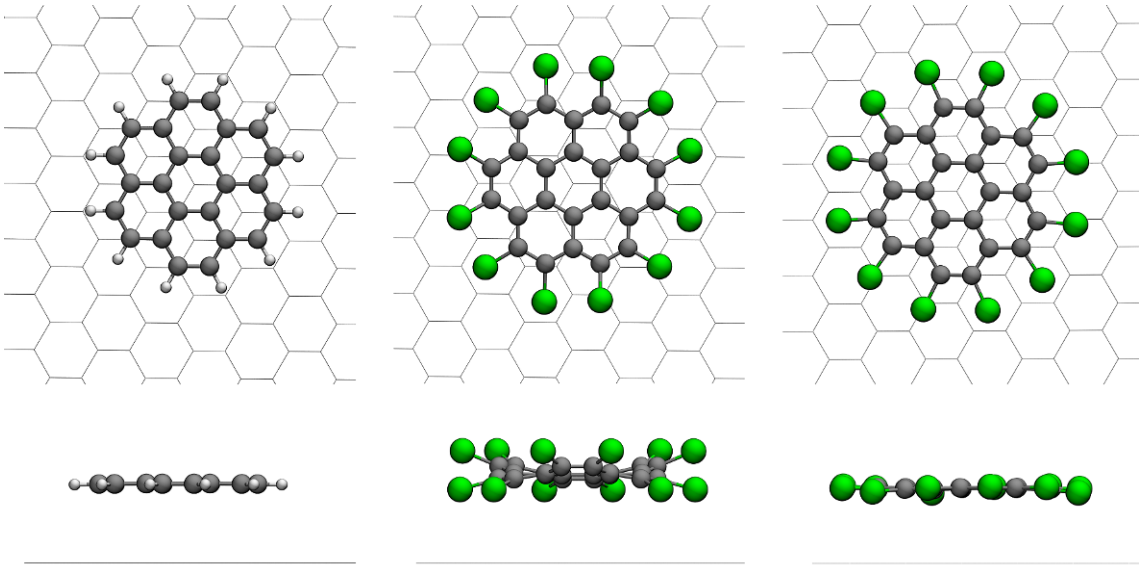


Figure 9.2: Top and side views of the optimized structures when adsorbed on graphene of coronene (left) and perchlorocoronene (center and right). For coronene only one structure is shown, that corresponds to the geometry obtained by both the CGenFF force field and the OM6-DH+ semiempirical method. For the perchlorocoronene the structures optimized at the two levels of theory are significantly different with the force field (center) producing a very distorted structure, while in the semiempirical (right) it is much more planar.

surface. To show this, adiabatic energy profile for planarization were computed at different levels of theory (i.e. FF and SQM), including DFT (wB97X-D[87] and B3LYP[254–257], cc-pVDZ basis set[258], run with Gaussian09[211]). To simulate the planarization process, the change of six pseudo-dihedral angles between the twelve chlorine atoms and the aromatic core was used as a reaction coordinate; see Fig. 9.3a. The energy profiles, see Fig. 9.3, clearly show that the PM6-DH+ underestimate the cost of planarization, with a total energy barrier of about 5 kcal mol^{-1}, versus the 25 and 30 kcal mol^{-1} predicted by DFT. When the PM6 and PM6-DH+ results are compared, it appears that the source of error is encoded in the PM6 core and not in the modeling of the dispersive interaction. Also, the optimal dihedral angle is significantly different moving from 25° for PM6 and PM6-DH+, to 42° for DFT. Noteworthy, the energy profile obtained with CGenFF is in better agreement with DFT, with a realistic barrier of 31 kcal mol^{-1} and a position for the energy minimum at 40°. Thus, we conclude that CGenFF results are more reliable in this particular case, and predict that edge perchlorination of coronene introduces an enhancement of the binding energy to graphite of about 25%.

Since perchlorocoronene is perfectly symmetrical with a C_3 rotational axis, no net molecular dipole is present. Thus, the observed increase in the adsorption energy can be solely attributed to the larger polarizability of the chlorine atoms as compared to the corresponding hydrogens in coronene. The isotropic polarizability for coronene and perchlorocoronene calculated at the PM6-DH+ level using the optimized geometries in vacuum and the finite-field method implemented in MOPAC2012 (POLAR module)[223], shows a two-fold increase for the PAH under perchlorination; 42.02 Å^3 versus 73.81 Å^3 for coronene and perchlorocoronene, respectively.

9.3 TEMPERATURE PROGRAMMED DESORPTION EXPERIMENTS

To benchmark the theoretical prediction, the desorption energy of both $C_{24}H_{12}$ and $C_{24}Cl_{12}$ to HOPG were measured by Temperature Programmed Desorption (TPD) experiments. For this purpose, samples of both molecules were prepared growing films on clean HOPG surface under ultrahigh vacuum (UHV) conditions. The TPD spectra were obtained heating the sample at constant heating rate

$$T(t) = T_0 + \beta t \tag{9.1}$$

with T_0 the initial low temperature and β the heating rate. The number of desorbed molecules as a function of time produces the desorption kinetic profile $R(t)$, which depends on the initial coverage Θ and the heating rate β. By fitting the $R(t)$ profiles to established desorption models[259], both the lateral molecule-molecule and the vertical molecule-substrate interactions can be determined. In the typical desorption profiles obtained for coronene (Fig. 9.4), two main peaks are distinguishable. The first one at

a temperature $T_B \simeq 410$ K mirrors the sublimation from the bulk and its intensity is proportional to the thickness of the deposed film when the initial coverage Θ_0 is greater than one. Molecular desorption from the monolayer corresponds to the second peak at a temperature $T_s \simeq 480$ K. All samples at initial sub-monolayer coverage ($\Theta_0 < 1$) feature only this second peak.

To determine the desorption mechanism, first it is necessary to study the dependence of the desorption profile $R(t)$ on the initial coverage Θ_i at the constant heating rate β [259]. For both coronene and perchlorocoronene, the temperature of the monolayer desorption (T_m) does not depend on Θ_i, but it stays constant over the whole range $0 < \Theta_i < 1$; see Fig. 9.5. This shows that the desorption process is dominated by the molecule-graphene interaction, with the molecule-molecule interaction playing a negligible role. Moreover, the spectra evidence first-order desorption kinetics, which can be expressed by the the Polanyi–Wigner equation

$$R(t) = \frac{d\Theta}{dt} = \nu\,\Theta \exp\left(-\frac{E_{des}}{k_B T}\right) \tag{9.2}$$

where both the preexponential factor ν and the desorption barrier E_{des} are not dependent on the initial coverage. This enables the use of the Redhead equation, which correlates the energy of desorption E_{des} with the temperature of the sub-monolayer desorption peak T_m as [197]

$$E_{des} = k_B T_m \left[\ln \frac{\nu T_m}{\beta} - 3.46\right] \tag{9.3}$$

For all initial coverages it was found that $T_m\,(C_{24}Cl_{12}) > T_m\,(C_{24}H_{12})$ as evidenced by Fig. 9.5, which indicates that the desorption energy for $C_{24}Cl_{12}$ is significantly higher than that of $C_{24}H_{12}$.

Strictly, this result holds only if the desorption kinetics of both molecules are characterized by the same preexponential factor ν. To verify this, an established method based on measurements of the $R(t)$ profiles at different heating rates β (here in the range between 0.5 and 17 Ks^{-1}) was used [198] to determine the velue of ν. By plotting $\ln(\beta/T_m^2)$ versus $1/k_B T_m$, straight lines were obtained, whose slopes and intercepts provide values of E_{des} and ν; see Fig. 9.6. As a result, the obtained preexponential factors differ significantly: $\nu(C_{24}H_{12}) = 1.03 \cdot 10^{16}$ s^{-1} and $\nu(C_{24}Cl_{12}) = 2.8 \cdot 10^{16}$ s^{-1}. The corresponding activation barriers are $E_{des}(C_{24}H_{12}) = 36.9 \pm 1.6$ kcal mol^{-1} and $E_{des}(C_{24}Cl_{12}) = 42.5 \pm 2.1$ kcal mol^{-1}, with an increase of 4.6 kcal mol^{-1} upon perchlorination. Overall, these experiments confirm the modeling predictions, and support the conclusion that perchlorination significantly enhances the desorption energy from graphite.

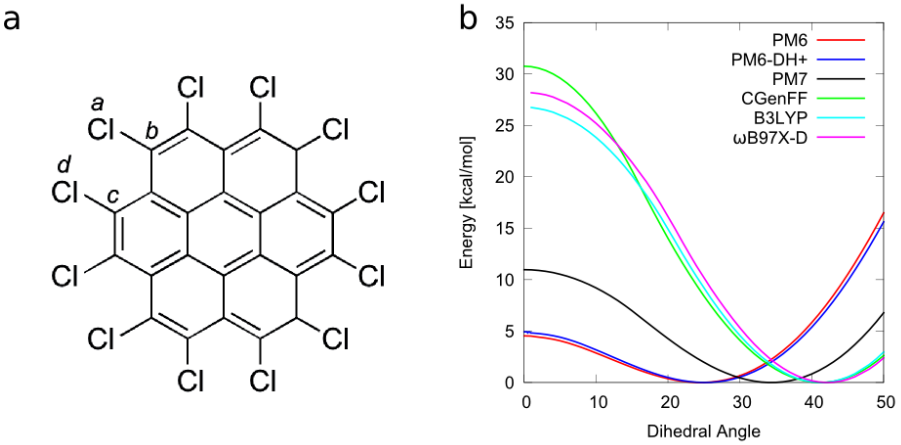


Figure 9.3: a) Chemical structure of perchlorocoronene. The four labeled atoms form the pseudo-dihedral used to evaluate the planarization cost; six of these dihedrals are present. b) Adiabatic energy profiles for the planarization of perchlorocoronene. In the minimum energy the molecule is not planar but presents a pseudo-dihedral angle of about 41° when modeled with the reference DFT. When the pseudo-dihedral angle is forced to zero, the molecule is fully planar and the energy value represents the cost of planarization.

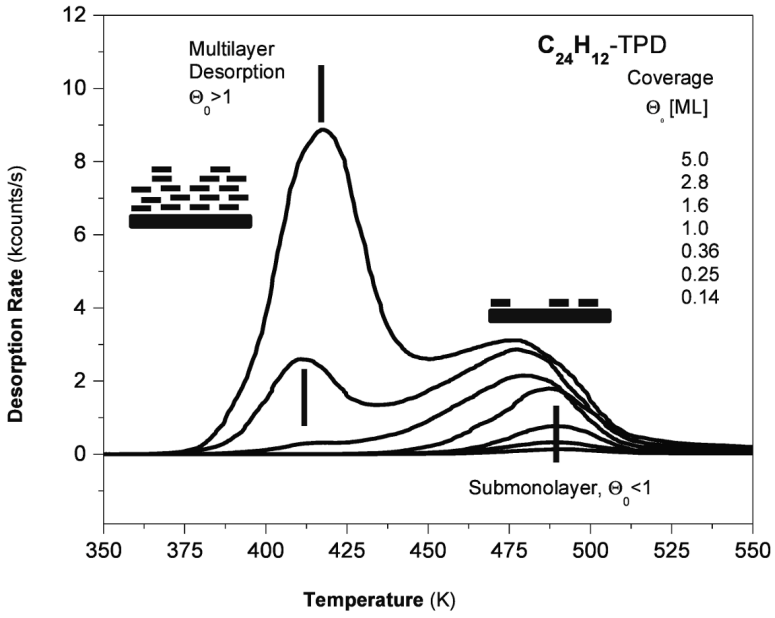


Figure 9.4: TPD spectra for the desorption of coronene obtained at different initial coverage Θ_0. The thermal desorption proceeds via two distinguishable peaks. The first at $T_B \simeq 410$ K reveals the sublimation of the multilayers. The second, peaked at $T_S \simeq 480$ K, represents desorption from the monolayer, where the molecule–surface interaction raises the binding energy by about 70 K.

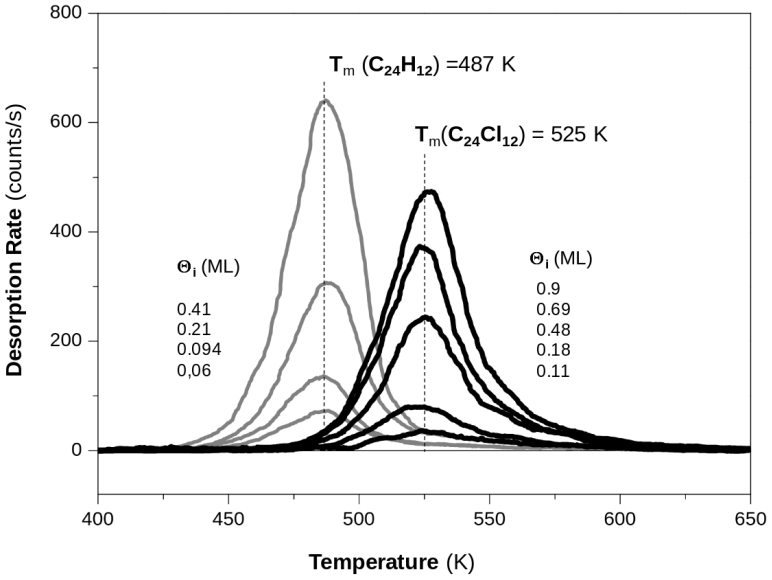


Figure 9.5: TPD spectra obtained at constant heating rate $\beta = 2.5\ Ks^{-1}$ for several $C_{24}H_{12}$ and $C_{24}Cl_{12}$ samples at different submonolayer initial coverage Θ_0. The increase in the desorption temperature T_m from 487 K to 525 K indicates that perchlorination raises the binding energy.

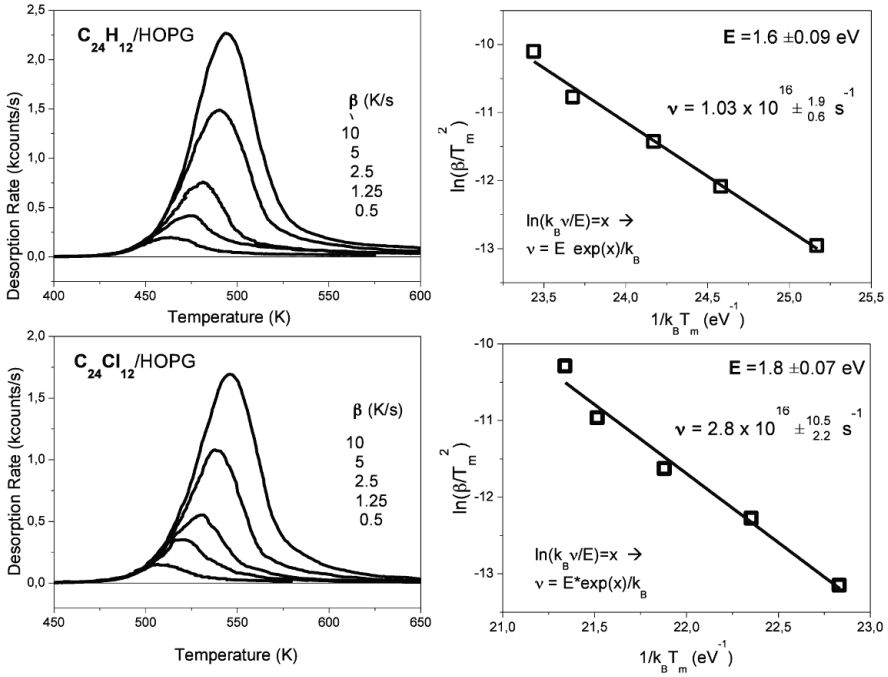


Figure 9.6: On the left, $R(t)$ profiles at different heating rate β and, on the right, the linear plot obtained by graphing $\ln(\beta/T_m^2)$ versus $1/k_B T_m$. The top row refers to coronene with at initial coverage of 0.75 ML. The lower row refers to perchlorocoronene.

9.4 MODELING THE SELF-ASSEMBLED MONOLAYER

The 2D self-assembly of coronene and perchlorocoronene were analyzed and compared. First, model architectures of the two SAMs were built using the CGenFF force field (version 3.0.1), based on the STM images obtained for coronene in UHV[188]. For the substrate, a periodic graphene layer of 304.544 x 204.192 $Å^2$ was used, with each atom assigned the atom type CG2R61 (aromatic carbon) and zero charge. The coronene SAM was built based on the hexagonal lattice found by STM[188], which has unit cell parameters $a = b = 11.28$ Å, $\alpha = 60°$, $A_{uc} = 110$ $Å^2$, and one molecule per unit cell. Since no experimental information was available on the orientation of the molecules inside the cell, a small lattice of 3x3 cells was built, with arbitrary initial orientations. Upon energy minimization optimization, the optimum orientation was used to built a full 10x10 cells layer, which was further optimized to resolve possible steric clashes. The same procedure was applied to model the SAM of perchlorocoronene. As no experimental data was available for the latter, the coronene hexagonal lattice was used as initial guess. The optimized SAMs are shown in Fig. 9.7. Both display a perfect hexagonal lattice with $\alpha = 60°$, one molecule per unit cell, and cell vector 11.31 Å for coronene and 12.88 Å for perchlorocoronene. The unit cell areas are 110.9 $Å^2$ and 143.8 $Å^2$ respectively, corresponding to an increase of 30% upon perchlorination, which is caused by the bulky chlorine substituents at the periphery of the molecule. Remarkably, the modeled and experimental unit cell parameters for coronene self-assembly are in perfect agreement.

The atomistic models of the two SAMs were used to evaluate the thermodynamic stability by evaluating their surface free energy γ. The energetic contribution to γ was accessed using the CGenFF force field, by evaluating the molecule-molecule (E_{inter}) and molecule-surface (E_{ads}) interactions, and the conformational strain energy upon adsorption (E_{strain}); see Tab. 9.2. In agreement with the previous results, the adsorption energy is greatly in favor of the perchlorocoronene by 8.5 kcal mol^{-1}. Interestingly, also the molecule-molecule interaction in the layer appears to be more favorable in the perchlorinated SAM, contributing for -10.3 kcal mol^{-1} compared to the -3.3 kcal mol^{-1} for coronene. This energetic gain for perchlorocoronene is partially counterbalanced by the strain energy, which is essentially zero for coronene and about 1.5 kcal mol^{-1} for perchlorocoronene. By summing all these contributions, the unit cell energy of perchlorocoronene is 14.0 kcal mol^{-1} more favorable than coronene.

The entropic contribution to γ was evaluated using the statistical thermodynamics expressions derived in chapter 3 at the temperature of 300 K and a 1 M concentration of monomers. Due to the bigger mass (+137%) and the lower symmetry (from D_{6h} to D_{3d}) of perchlorocoronene, the surface confinement of coronene is easier. This introduce a small translational and rotational entropy difference for the two molecules, which slightly counterbalances the energy gain by perchlorination to -11.4 kcal mol^{-1} per molecule; see Tab. 9.2.

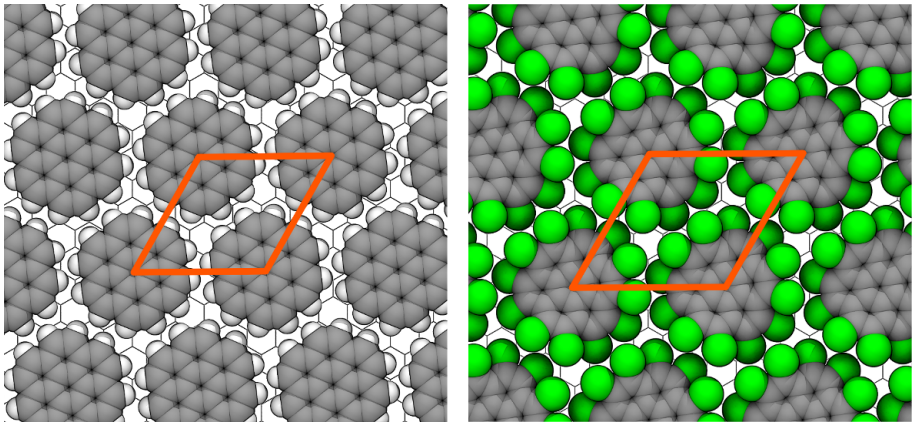


Figure 9.7: SAMs modeled for coronene (left) and perchlorocoronene (right) on graphene. Both architectures adopt a hexagonal unit cell of size of 11.32 Å and 12.89 Å, respectively, and an unit cell area of 110.88 $Å^2$ and 143.81 $Å^2$.

	E_{ads}	$E_{inter}/2$	E_{strain}	E_{uc}	TS_{tr}	TS_{rot}	$\Delta\mu$	A_{uc}	γ	γ_E	γ_S
$C_{24}H_{12}$	-35.7	-3.3	0.1	-38.9	10.4	8.7	-21.5	110.9	-19.4	-36.6	17.2
$C_{24}Cl_{12}$	-44.1	-10.3	1.5	-52.9	11.2	10.6	-32.9	143.8	-22.9	-38.0	15.1
Δ	-8.5	-7.0	1.4	-14.0	0.8	1.9	-11.4	+32.9	-3.5	-1.4	-2.1

Table 9.2: Thermodynamic quantities calculated for the SAMs formed by coronene and perchlorocoronene, and the difference between the two to highlight the gains/losses upon perchlorination. In order from the left: adsorption energy, interaction between molecules in the layer, strain energy upon adsorption, energy per unit cell, translational and rotational free energy costs, total difference in chemical potential per unit cell, area of the unit cell, surface free energy, and decomposition into energetic and entropic terms of γ. All energies are expressed in kcal mol^{-1}, the unit cell area in $Å^2$, and gamma and its decomposition in kcal mol^{-1} nm^{-2}.

As the vibrational contribution to self-assembly is neglected, the surface free energy can be calculated using the simplified expression

$$\gamma = \frac{\Delta\mu}{A_{uc}} \left[E_{uc} - 3RT + T\left(S_{tr} + S_{rot}\right)\right] \tag{9.4}$$

where A_{uc} and E_{uc} are the area and energy per unit cell and S_{tr} and S_{rot} the translational and rotational entropies of the monomer. The results of this analysis indicates that γ is 3.5 kcal mol^{-1} nm^{-2} more favorable for $C_{24}Cl_{12}$; see Tab. 9.2. Notably, the decomposition of γ into energy and entropy contributions indicates that they are equally important and both favor the self-assembly of $C_{24}Cl_{12}$ by 1.4 and 2.1 kcal mol^{-1} nm^{-2}, respectively (Tab. 9.2).

These results show that the formation of the perchlorinated SAM is favored by both energetic and entropic reasons. The first result, comes directly from the observation that the adsorption energy is enhanced upon perchlorination, as shown by TPD, but also that the molecule-molecule interactions in the layer are improved by the presence of polarizable chlorine atoms. Perhaps surprisingly, the self-assembly of the perchlorocoronene appears to be more stabilized for entropic reasons. This comes from the observation that

the subsitution with the bigger chlorine atoms at the periphery of the molecule increases the unit-cell area. The latter effectively reduces the number of molecules necessary to cover a given surface area, thus lowering the overall translational and rotational entropy loss per unit of surface area.

9.5 GRAPHENE EXFOLIATION

Finally, the functional advantages of the perchloro substitution were investigated in the molecule-assisted liquid phase exfoliation (LPE) of graphite. To analyze the effect of coronene versus perchlorocoronene on the exfoliation yield, graphene dispersions were prepared adding 100 mg of graphite power in 10 ml of orto-dichlorobenzene (o-DCB, solvent) and sonicating the solution for 6 hours in the present of the same number of molecules of $C_{24}H_{12}$ or $C_{24}Cl_{12}$ (3.2 ± 0.1 mg of $C_{24}H_{12}$ and 7.5 ± 0.1 mg of $C_{24}Cl_{12}$). The dispersions were allowed to settle for 15 minutes and 90% of the surnatant solution was centrifuged for one hour at 10000 rpm. 70% of the centrifuged dispersion was removed and characterized by UV/Vis spectroscopy, High Resolution Transmission Electron Microscopy (HR-TEM), and micro Raman analysis.

UV/Vis Spectroscopy

The yield of the exfoliation was quantified in two ways. First, the dispersions were passed through polytetrafluoroethylene (PTFE) membrane filters (pore size 0.1 μm). Weighting the mass of the filtered material on a microbalance provides a measure of the graphene concentration in the produced dispersions. By this measure, perchlorocoronene increases the yield of exfoliation by 34%; see Tab. 9.3. As a second way, the graphene dispersions were characterized by UV/Vis absorption spectroscopy. For both coronene and perchlorocoronene, the final dispersions contained a large excess of surfactant, as shown by the UV/Vis absorbance spectra which present superposition of the coronene and perchlorocoronene absorption bands and the uniform decaying light dispersion of graphene layers; see Fig. 9.8. The graphene concentration in the dispersion was then determined from the adsorption at 660 nm, using an extinction coefficient of 2460 $mL\,mg^{-1}\,m^{-1}$ [53], where only the graphene is active and no bands from the surfactant is present. So, each dispersion was diluted several times and for each dilution the adsorbance ot 660 nm was recorded. By plotting the absorbance normalized by the cell pathlength versus the relative concentration to the initial solution, both coronene and perchlorocoronene dispersions followed the Lambert-Beer behavior, which allows for an accurate determination of the concentration from the absorbance; see Fig. 9.9. From the determination of the concentration of the exfoliated graphene, the yield Y_W was determined as the ratio between the weight of the exfoliated material and the starting graphite mass. This measure indicates that the use of perchlorocoronene results in a considerable

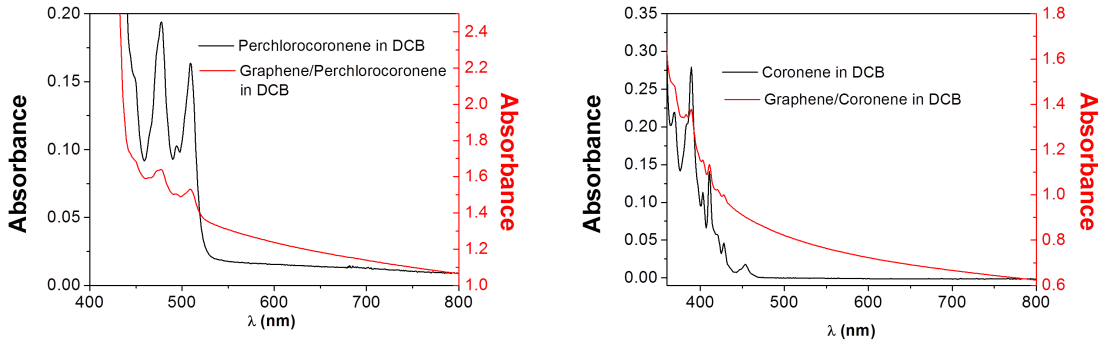


Figure 9.8: Comparison between the adsorbance spectra of the graphene dispersion obtained by the use of a surfactant (red curve), and the absorbance of the surfactants alone (black curve). The plot on the left refers to perchlorocoronene, on the right coronene.

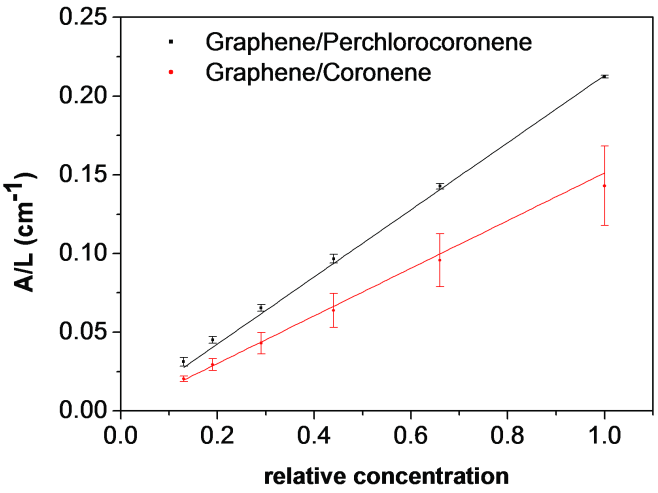


Figure 9.9: Absorbance at 660 nm upon dilution of the graphene dispersions obtained by coronene (red) and perchlorocoronene (black). From the slope of the linear interpolation, the concentration of graphene is determined.

improvement increasing the yield from 0.33% to 0.57%; see Tab. 9.3. Combining this data with the result of the HR-TEM (see below), the yield Y_{WM} can be quantified, which measures the fraction of single and few layer graphene (SLG and FLG) produced per unit of initial graphitic material. The results in Tab. 9.3 show that perchlorocoronene doubles the Y_{WM} yield, indicating an augmented production of high-quality graphene (i.e. SLG). These results indicate an overall improvement between 30% and 50% of the exfoliation yield upon exchanging coronene with perchlorocoronene.

Characterization via HR-TEM

To complement the UV/Visible spectroscopy analysis, the exfoliated material was characterized counting the number of planes of the exfoliated layers by visual inspections of micrographs obtained by High-Resolution Transmission Electron Microscopy

Graphene Yield	$C_{24}H_{12}$	$C_{24}Cl_{12}$	Δ
Conc. [μg/mL] (mass)	95±6	128±2	33±6
Conc. [μg/mL] (UV)	114±20	171±1	57±20
Y_W [%] (mass)	0.95	1.28	0.33
Y_W [%] (UV)	1.14	1.71	0.57
Y_{WM} [%] (mass+TEM)	0.09	0.17	0.08
Y_{WM} [%] (UV+TEM)	0.10	0.22	0.12

Table 9.3: Determinations of the exfoliation yield in the presence of coronene and perchlorocorone.

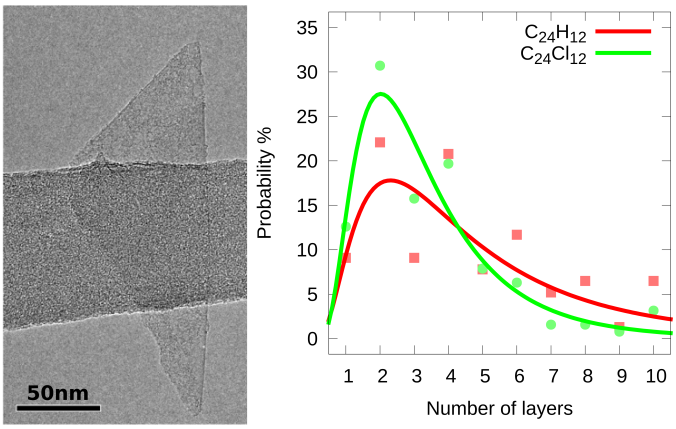


Figure 9.10: On the left, HR-TEM image of an exfoliated graphene bilayer. On the right, statistical analysis on the number of layers for the exfoliated graphene dispersion in the presence of coronene (red) or perchlorocoronene (green).

(HR-TEM). The samples were prepared by deposing a drop of graphene dispersion on a TEM grid which was then kept under UHV conditions overnight to remove the solvent. Upon analyzing hundreds of flakes, the results show that the use of perchlorocoronene as dispersion stabilizing agent increases the production of single and few layer graphene (SLG and FLG) by approximately 4% and 14% respectively relative to coronene.

Characterization via Micro Raman

Finally, the quality of the exfoliated material was assessed by Micro Raman analysis. In this analysis the observables of interest are 1) the shape of the 2D peak, measured as the determination coefficient R^2 for the fitting of the 2D peak with a Lorentzian function, and 2) the intensity ratio between the D and G peaks ($I(D)/I(G)$). The first allows to discriminate single layer graphene from few layers, and graphitic material, since the 2D peak is completely symmetric in the first case, asymmetric in the second, and shows two peaks in the third. Representative Raman spectra for the three cases are showed in Fig. 9.11. This analysis provides qualitative information on the thickness distribution of the exfoliated material. The second parameter, the intensity ratio $I(D)/I(G)$, gives information on the amount of defects in the sample. Note that the liquid phase exfoliation of

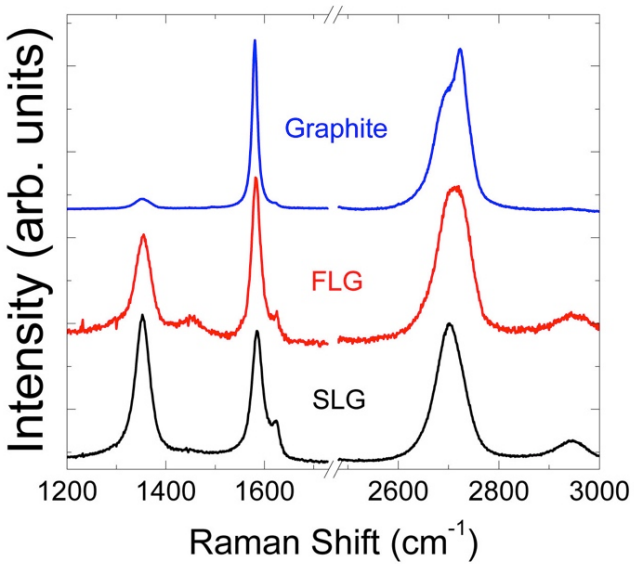


Figure 9.11: Typical Raman spectra obtained for graphene of different thickness produced by liquid phase exfoliation.

	SLG	FLG	Graphite
No stabilizer	31%	61%	8%
$C_{24}H_{12}$	36%	58%	6%
$C_{24}Cl_{12}$	51%	45%	4%

Table 9.4: Percentages of single layers (SLG), few layers (FLG), and graphitic materials obtained by Raman analysis.

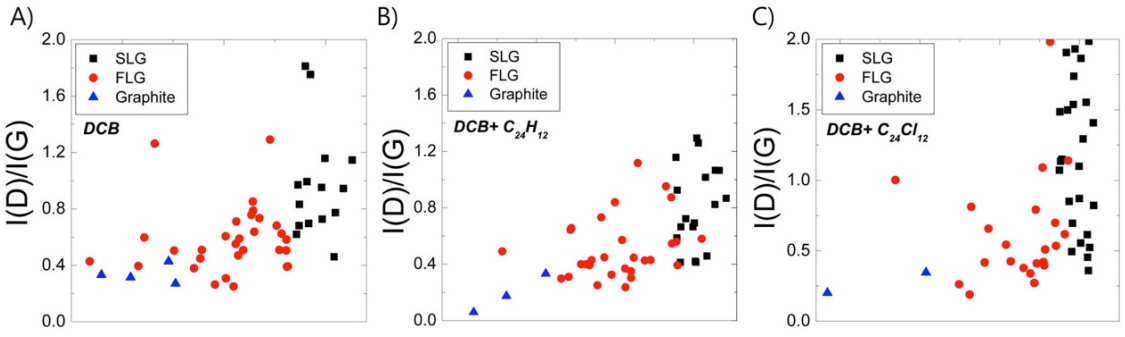


Figure 9.12: I(D)/I(G) ratio analysis of graphene dispersions in a) pure solvent (o-DCB), b) coronene, and c) perchlorocoronene.

graphite always produces small graphene flakes. The D peak is thus always present, as it is activated by the broken translational symmetry due to the edge of the flakes. Therefore, only qualitative comparisons can be done in this case using this Raman observable.

Percentages of SLG, FLG, and graphite obtained by LPE using the pure solvent and solutions of coronene and perchlorocoronene are reported in Tab. 9.4. These show that the addition of surfactant improves the quantity of single layer graphene by 5% (coronene) and 20% (perchlorocoronene), confirming the HR-TEM result that perchlorocoronene is more effective dispersion stabilizing agent. Concerning the intensity ratio $I(D)/I(G)$, a general increase is observed with decrease in the thickness, which is a typical observation in graphite LPE; see Fig. 9.12. In fact, typically the thinner the flake, the smaller the size will be. The average $I(D)/I(G)$ ratio measured for the pure solvent, coronene and perchlorocoronene solutions are 0.96, 0.80 and 1.15, respectively, which shows no appreciable difference as the typical error on the Raman intensity ratio is about 15%.

In conclusion, both the HR-TEM and Raman results consistently support that the use of perchlorocoronene over coronene improves both quality and the quantity of the exfoliated material.

9.6 DISCUSSION

The results presented in this chapter show that the edge perchlorination of coronene enhances its self-assembly propensity, which increases its ability to act as a surfactant in the liquid phase exfoliation of graphene. A thermodynamic analysis of the corresponding monolayers provides and interpretation at the molecular level.

First, it has been showed by TPD that perchlorination of coronene increases the desorption energy from graphite by 12.5%. The calculations performed at the single molecule level – one molecule physisorbed on the surface – indicate that this enhancement in the binding energy mainly arises from the substitution of the small hydrogen atoms with highly polarizable groups, which allow for stronger dispersive interactions with the graphite substrate. Interestingly, the calculations show that the obtained enhancement is somehow suboptimal, since the bulky chlorine atoms cause a strong distortion of the molecular geometry, which introduces an intermolecular strain upon adsorption equivalent to 23% of the total increase in the desorption energy. Thus it is possible that a regioselective substitution of only some of the hydrogens with chlorine atoms, e.g. yielding the hypothetical 1,2,5,6,9,10-hexachlorocoronene, that is planar on both the surface and in isolation, and would be beneficial as there would be no strain energy upon adsorption. Calculation of the desorption energy of this partially chlorinated derivative on graphene predicts adsorption energy that is 3 $\mathrm{kcal\,mol^{-1}}$ stronger than the perchlorocorone; see Fig. 9.13. Thus, cleverly designed regioselective chlorine substitutions may further enhance the affinity for the graphite substrate.

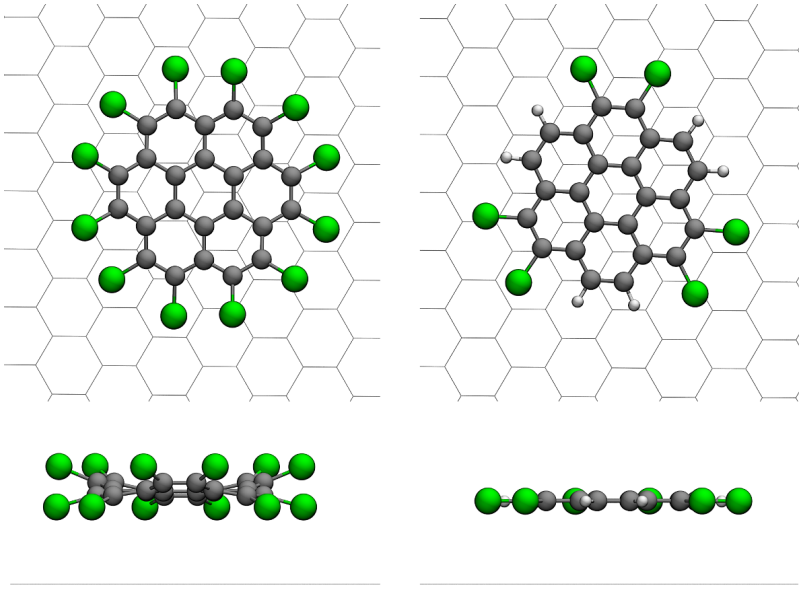


Figure 9.13: Comparison of the geometries of perchlorocoronene (left) and the 1,2,5,6,9,10--hexachlorocoronene (right). While the chlorine atoms are highly out of plane in the former due to the high steric repulsions, the second is completely flat.

Second, the results of the liquid phase exfoliation of graphite show that both the quantity and quality of the exfoliated graphene are increased when perchlorocoronene is used instead of coronene as graphene surfactant. Importantly, the thermodynamic analysis, based on the approach presented in chapter 5, predicts that the substitution of the peripheral hydrogens with chlorines favors the self-assembly by both increasing the strength of the interaction with the surface and within the SAM, and reducing the surface concentration in the SAM, i.e. the number of molecules per unit of area. The latter, which is due to the presence of sterically more demanding substituents at the periphery, causes a significant expansion of the unit cell, that corresponds to a decreased translational and rotational entropy cost upon self-assembly. Based on these results, we conclude that perchloro substitution improves the stability of the SAM not only energetically, but also entropically, which was unexpected from the initial chemical design.

9.7 TAKE HOME MESSAGES

In conclusion, the results presented in this chapter demonstrate that the thermodynamic approach presented in chapter 5 is useful to rationalize the effect of the chemical functionalization (perchlorination) of coronene on the stability of the corresponding SAM. The developed theoretical analysis based on modeling and statistical thermodynamics enabled to disentangle the energetic from the entropic contributions to the self-assembly free energy, which provided fundamental insight onto the forces underlying molecular self-assembly at surfaces and interfaces. From the comparison between coronene and perchlorocoronene it was possible to observe that perchloro functionaliza-

tion improves the thermodynamic stability of the generated SAM not only energetically by increasing the unit cell energy, but also entropically by reducing the surface concentration. The functional advantages of the perchloro substitution were exemplified in the context of the liquid-phase exfoliation of graphite for the production of high-quality graphene flakes.

Part III

Beyond Current Approximations

10

ADSORPTION EQUILIBRIUM ON GRAPHENE: EFFECT OF THE VOLUME/SURFACE RATIO

10.1 MOLECULAR ADSORPTION ON SURFACES

One of the first steps when a gas or a solution is put in contact with a solid surface is the interaction and adsorption of the fluid components onto the solid. This basic process has outstanding importance in a wide set of applications, ranging from the cleaning of aqueous solutions from pollutants[260,261] and the physisorption of proteins on biomedical implants[262,263], to the physi/chemisorption of chemicals onto nanomaterials. Among these, remarkable is molecular self-assembly at surfaces, which can be viewed as a two steps process: first, the molecules adsorb to the surface, and second they create self-organized 2D periodic structures.

In the first part of this dissertation, a theoretical framework to rationalize and predict the propensity to 2D self-assembly has been presented. One of the approximations introduced was to consider self-assembly as a two-states process: the disassembled state in solution and the self-assembled architecture physisorbed on the surface. In this chapter, we investigate in more details the first step of the self-assembly reaction, which corresponds to the adsorption equilibrium. In particular, the focus is on the special case of physisorption from a closed system to a flat surface, like the weak van der Waals interaction obtained when a molecule adsorbs on graphite. While the nature and strength of the interaction between the molecule and the graphene surface have been already studied[178,207,208], the question is how a given compound is partitioned between the solution and the surface at chemical equilibrium, and what are the environmental conditions that influence the ratio of the adsorbed versus the desorbed state. To this aim, this problem is first tackled theoretically, showing, on the basis of simple thermodynamic considerations, that the adsorption probability solely depends on the ratio between the available volume and accessible surface. Second, we aim at finding out how to efficiently evaluate the adsorption probability for an arbitrary selected molecule. This implies finding a clever approach to evaluate the difference in the standard chemical potentials between adsorbed and desorbed states, which means evaluating the equilibrium constant for the adsorption process. To this goal, equilibrium sampling simulations results on the adsorption of five molecules on graphene (Fig. 10.1) were used as reference to test the accuracy of both the analytical integration of the partition functions[138], and the numerical sampling of the phase space via the confinement method[264,265]. The theoretical results are discussed in the context of molecular self-assembly at surfaces, highlighting

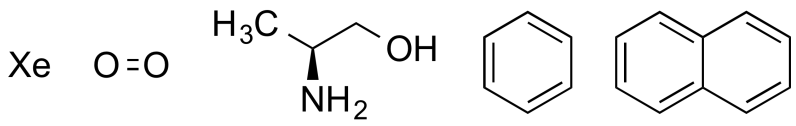


Figure 10.1: Chemical structures of the five studied molecules in this work. From left to right: xenon, oxygen, alaninol, benzene, and naphthalene.

the existence of two possible pathways, one of them how involving the formation of an intermediate disordered state in 2D.

10.2 THEORY

Consider a closed system composed of N_0 molecules of A that can be dispersed in the solution phase (A_{3D}), or physisorbed on a flat surface (A_{2D}). When the two states are in equilibrium, i.e. each molecule can freely adsorb and desorb, the following equation holds

$$A_{3D} \rightleftharpoons A_{2D} \tag{10.1}$$

To study how the N_0 initial molecules are distributed in the two states at equilibrium, their chemical potentials are needed, which, in the limit of idealized conditions, can be written as

$$\mu_{3D} = \mu^{\ominus}_{3D} + kT \ln \frac{C_{3D}}{C^{\ominus}_{3D}} \tag{10.2}$$

$$\mu_{2D} = \mu^{\ominus}_{2D} + kT \ln \frac{C_{2D}}{C^{\ominus}_{2D}} \tag{10.3}$$

where $\mu^{\ominus}_{3D}$ and $\mu^{\ominus}_{2D}$ are the standard chemical potentials at the $C^{\ominus}_{3D}$ and $C^{\ominus}_{2D}$ standard concentrations, and C_{3D} and C_{2D} are the concentrations of the two states. At chemical equilibrium, the difference between the chemical potentials equals zero ($\Delta\mu = \mu_{2D} - \mu_{3D} = 0$), giving an expression for the equilibrium constant of the reaction

$$K_{eq} = \exp\left(-\frac{\Delta\mu^{\ominus}}{kT}\right) = \frac{C_{2D}/C^{\ominus}_{2D}}{C_{3D}/C^{\ominus}_{3D}} \tag{10.4}$$

Defining the concentration in 2D (3D) as the ratio between the number of molecules N_{2D} (N_{3D}) and the accessible surface S (volume V), the equilibrium constant can be rewritten as

$$K_{eq} = \frac{N_{2D}}{N_{3D}} \frac{V/S}{V^{\ominus}/S^{\ominus}} = \frac{N_{2D}}{N_{3D}} \frac{\eta}{\eta^{\ominus}} \tag{10.5}$$

where it is assumed $N^{\ominus}_{3D} = N^{\ominus}_{2D}$ and the parameter $\eta = V/S$ defines the extension ratio between the accessible volume and the accessible surface.

Coupling the definition of the equilibrium constant with the constrain of mass balance ($N_{2D} + N_{3D} = N_0$), the probabilities of the adsorbed and desorbed states are

$$P_{3D} = \frac{\eta/\eta^{\circ}}{K_{eq} + \eta/\eta^{\circ}} \tag{10.6}$$

$$P_{2D} = \frac{K_{eq}}{K_{eq} + \eta/\eta^{\circ}} \tag{10.7}$$

These relations show how the probability of populating the two states depends solely on the value of the equilibrium constant K_{eq}, which is a function of the standard chemical potential difference for the adsorption reaction, and the value of the extension ratio η. Interestingly, the probabilities are not a function of the initial number of molecules N_0, thus the equilibrium in Eq. 10.1 does not depend on the initial concentration of the system. This implies that, in the limits of the model above, two systems having the same concentration but different η, i.e. different box "shapes", will have different adsorption probabilities. Also, systems with very different concentrations but same extension parameter η will have the same adsorption probability.

As a remark, no excluded volume is considered for the adsorbed state, which implies that the adsorbed molecules are points in space that occupy no surface area, so that an infinite amount of molecules can absorb on the surface. Moreover, the adsorbate is physisorbed and free to translate on the surface, i.e. the translation barrier is lower than the thermal energy. These conditions preclude the recovering of classic isotherms that were derived to study molecular adsorption at surfaces and porous materials, like the Langmuir[266,267], BET[268], or other isotherms[269–271]. The results presented here can thus be considered as an ideal gas solution, so valid only at very low concentrations or pressure of the adsorbate and low surface area coverage.

10.3 COMPUTATIONAL METHODS

The theoretical results above are investigated by modeling. This is done by simulating the adsorption equilibrium of small molecules on graphene in vacuum. Five molecules were studied, which differ in size and symmetry: xenon as a van der Waals gas, oxygen as simple diatomic gas, alaninol (2-amino-1-propanol) as small asymmetric molecule, and benzene and naphthalene as rigid symmetric species; see Fig. 10.1. At first, equilibrium simulations were performed, to monitor the dependence of the adsorption probability on η and its independence on the initial concentration. Then, results from two different free energy methods are presented, which aim at the evaluation of the difference in standard chemical potentials between the adsorbed and gas phases in an accurate and efficient way. All simulations were performed using the CHARMM software[144] and the General Amber Force Field (GAFF)[123,124].

Equilibrium Sampling

Molecular dynamics simulations were performed to sample the equilibrium probability of populating the adsorbed and desorbed states. As default setup, 25 or 49 molecules of adsorbate were positioned on a graphene surface and the system simulated using periodic boundary conditions; see Fig. 10.2 for an example of the simulation cell. The system was heated for 50 ps, equilibrated for 1 ns, and then a production run of 10 ns was performed. The dynamics were performed via the Langevin integrator with a timestep of 1 fs (no bond constrain was used) and a friction coefficient of 0.1 ps^{-1}, which allows for fast desorption kinetics. For each molecule, 16 independent runs were carried out, using a small (122.8 x 106.35 $Å^2$) or a big (245.6 x 212.7 $Å^2$) graphene surface, a high (49) or low (25) number of molecules, and four different values for the dimension along the z coordinate (50, 100, 200, or 300 Å), which correlates directly with the value of η.

From the equilibrium dynamics, the probabilities to populate the desorbed and the adsorbed states were calculated as follows. First, the basins describing the two states were defined. For the desorption process, a reaction coordinate is obtained using the distance of each molecule from the surface, and the desorbed probability measured by counting all molecules with a distance from the surface less than a cutoff. For each simulation, also the value of η can be calculated. Due to the use of the periodic boundary conditions, the accessible surface is twice the surface of the graphene layer, so, by defining as x and y the dimension of the rectangular graphene, the accessible surface is $S = 2xy$. The volume is instead the whole box volume xyz, minus the volume used to define the adsorbed state which is $2xyc$, where z is the dimension of the box along the z coordinate and c the cutoff to define the adsorbed state. Thus, η is

$$\eta = \frac{V}{S} = \frac{xyz - 2xyc}{2xy} = \frac{z}{2} - c \tag{10.8}$$

Having the value of η and the probabilities of the adsorbed and gas states, the standard difference in chemical potential $\Delta\mu^{\ominus}$ can be obtained by inverting of Eq. 10.4

$$\Delta\mu^{\ominus} = -kT \ln\left(\frac{\eta}{\eta^{\ominus}} \cdot \frac{P_{2D}}{P_{3D}}\right) \tag{10.9}$$

The difference in chemical potential and its uncertainty were then obtained averaging the results of Eq. 10.9 over the 16 independent simulations.

Ideal Gas Approximation

The standard difference in chemical potential can be obtained analytically if the adsorbed and gas states are modeled as ideal gases in the limits of the rigid-rotor harmonic-

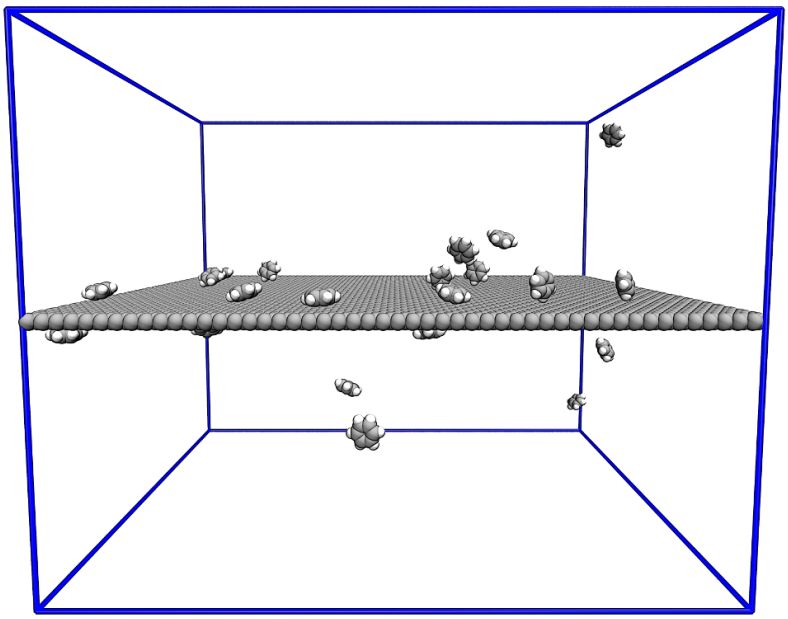


Figure 10.2: Typical simulation box used in the equilibrium sampling simulations. The graphene layer defines the xy plane, while the z dimension is changed systematically to simulate different values of the η parameter. Due to the periodic boundary conditions, the molecules can diffuse through the z dimension and adsorb on the graphene layer on both sides.

oscillator approximation (RRHO). In these approximations, the chemical potentials can be obtained from Eq. 5.6 and 5.7, as presented in chapter 5.

The limitations of this approach are essentially two. First, the molecule is considered to be harmonic. This means that any anharmonicity of any degree of freedom is neglected, which is generally not appropriate for torsions and internal rotations. Second, the exact nature of the three degrees of freedom gained upon adsorption, in particular the two rotations, is unclear, i.e. if the molecule is small enough, these rotations are not prohibited and should be treated more properly as hindered rotations[272,273].

Confinement Free Energy

To overcome the two problems above, the confinement free energy method[264,265] was used. This involves the gradual restrain of the molecule in the harmonic state, whose chemical potential is exact. According to the developments, the best procedure is to confine the molecule using a best-fit harmonic restrain, which only confines the vibrations of the molecule but not the translations and rotations[265]. This is very useful, since these two last terms are know analytically, and only the anharmonicity associated to the vibrations is evaluated.

Here, the confinement was used to evaluate the standard chemical potential of the molecule physisorbed on the surface, to account in particular for the anharmonicity of

the three vibrations that are gained upon physisorption (the lost translation and two rotations). For this reason the best-fit approach cannot be used, since it would remove these three degrees of freedom. Thus, an absolute confinement was used, where the molecule is confined using as reference the molecule in the adsorbed conformation. This introduces a problem associated with the two free translations on the surface. Although analytical, these are effectively confined. Moreover, to have a good estimation of these terms, the full translational space has to be explored during the simulation, which means collecting very long simulation trajectories. This can be done using a small simulation box ($34.384 \times 34.032\,\text{Å}^2$), for which the exploration of the translational space is affordable. Finally, in the confinement approach the atomic fluctuations around a reference structure need to be evaluated, which is usually done by measuring the root mean square deviation (RMSD) from the reference structure. This is a problem for symmetric molecules, for which the RMSD algorithm is not able to assign zero fluctuations to symmetry-related conformations, and effectively introduces an artificial free energy cost related to the labeling of the atoms, which accounts for the loss of symmetry in the RMSD computation.

The calculation of the standard chemical potential on the surface can thus be view as a three steps process

$$A^{\ominus}_{2D} \rightarrow A_{2D} \rightarrow A^{\dagger}_{2D} \rightarrow A^{*}_{2D} \tag{10.10}$$

The first step reduces the translational space from the molecule in its standard state ($A^{\ominus}_{2D}$) to a limited-box system (A_{2D}) which can be effectively simulated. The difference in chemical potential associated to this step is due to the reduction on the translational partition function, obtaining

$$\Delta\mu_{A^{\ominus}_{2D} \rightarrow A_{2D}} = kT \ln \frac{C_{2D}}{C^{\ominus}} \tag{10.11}$$

The second step removes the symmetry of the molecule by labeling the atoms ($A^{\dagger}_{2D}$). The $\Delta\mu$ in this second step can be simply evaluated as

$$\Delta\mu_{A_{2D} \rightarrow A^{\dagger}_{2D}} = -kT \ln \sigma \tag{10.12}$$

where σ is the symmetry number for the molecule, that is the number of symmetry-related equivalent configurations. The third and last step confines the molecule to the adsorbed conformation (A^{*}_{2D}) using an absolute harmonic restrain. This free energy cost is the confinement free energy $\Delta\mu_{conf}$, which includes the loss in the two translational degrees of freedom on the surface, the rotational along z and all anharmonicity associ-

ated to the vibrational degrees of freedom. Finally, the chemical potential of A^*_{2D} can be evaluated analytically as

$$\mu^*_{2D} = \mu_{elec} + \mu^*_{vib} \tag{10.13}$$

where μ_{elec} and μ^*_{vib} are the electronic and vibrational chemical potentials in the confined state. All together, the standard state chemical potential of the adsorbed state can be evaluated as

$$\begin{aligned}\mu^{\circ}_{2D,CF} &= \mu^*_{2D} - \Delta\mu_{A^\dagger_{2D}\to A^*_{2D}} - \Delta\mu_{A_{2D}\to A^\dagger_{2D}} - \Delta\mu_{A^\circ_{2D}\to A_{2D}} \\ &= \mu_{elec} + \mu^*_{vib} - \Delta\mu_{conf} + kT\ln\sigma - kT\ln\frac{C}{C^\circ}\end{aligned} \tag{10.14}$$

On the opposite, the standard state chemical potential of the molecule in 3D ($\mu^{\circ}_{3D,CF}$) can be evaluated using the standard best-fit confinement approach in isolation, as

$$\mu^{\circ}_{3D,CF} = \mu^{\circ}_{tr} + \mu_{rot} + \mu^*_{vib} - \Delta\mu_{conf} + \mu_{elec} \tag{10.15}$$

For both the adsorbed and desorbed states, the confinement free energy is obtained by numerical integration of the RMSD fluctuations sampled by molecular dynamics at different restraining force constants[264,265]. As the force constant increases, a problem arises from the simulation time step dt, since in the presence of strong restraints the frequencies of the internal vibrations increase, which requires smaller and smaller time steps to preserve the physical meaning of the simulations. To account for this problem an adaptive time step approach was implemented, where the time step is decreased as the force constant is increased; see the Appendix E. For each confinement, 36 Langevin simulations spanning uniformly (in logarithmic scale) the force constant range from 0.00001 to 100 kcal mol^{-1} Å^{-2} were run. Each molecular dynamics simulation was integrated for a total of $2 \cdot 10^7$ steps, which is equivalent to 20 ns of dynamics at 1 fs time step. A value of 10 ps^{-1} was used as friction coefficient for the Langevin integrator, and, for the adsorbed state, the fluctuations were calculated only on the frames where the distance of the center of mass of the molecule from the surface was less then the cutoff used to define the desorbed state.

10.4 RESULTS AND DISCUSSIONS

Equilibrium Sampling

The result obtained from simulations at equilibrium are presented first. Five molecules were studied: xenon, oxygen, alaninol, benzene, and naphthalene. Since a great number of desorption events are needed to sample the adsorbed and desorbed probabilities,

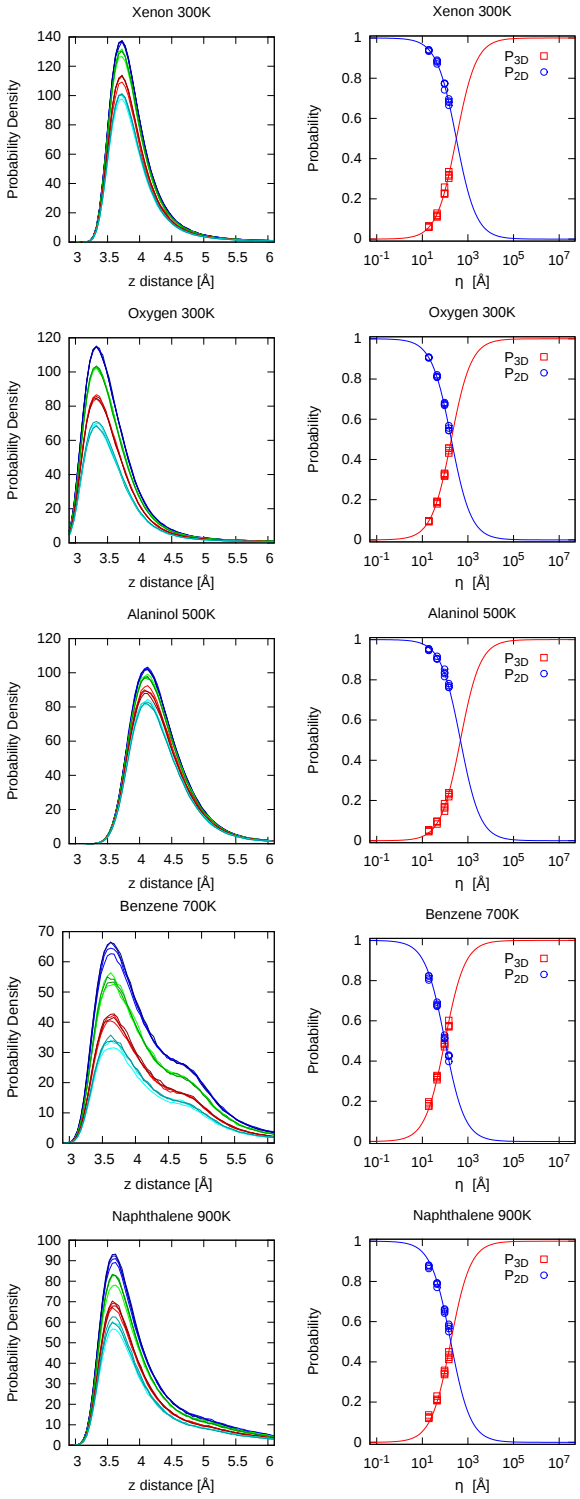


Figure 10.3: Equilibrium sampling results for the five studied systems. Left: probability density as a function of the distance from the surface; the blue, green, red and cyan curves indicate the value of η (19, 44, 94, 144 Å). Right: probability of the desorbed and adsorbed states as a function of the extension parameter η. Each of the four points corresponds to the superimposition of the four simulations at constant η value.

η [Å]	Box size [Å^3]	N_0	Xenon, 300K			Oxygen, 300K			Alaninol, 500K		
			P_{2D}	P_{3D}	$\Delta\mu^{\ominus}$	P_{2D}	P_{3D}	$\Delta\mu^{\ominus}$	P_{2D}	P_{3D}	$\Delta\mu^{\ominus}$
19	122.8 x 106.35 x 50	25	94.1	5.9	8.95	90.9	9.1	9.23	95.1	4.9	14.7
	245.6 x 212.7 x 50	25	93.4	6.6	9.02	90.6	9.4	9.25	94.6	5.4	14.8
	122.8 x 106.35 x 50	49	93.9	6.1	8.97	90.8	9.2	9.23	95.5	4.5	14.6
	245.6 x 212.7 x 50	49	94.1	5.9	8.95	90.5	9.5	9.25	94.8	5.2	14.8
44	122.8 x 106.35 x 100	25	87.1	12.9	8.96	81.1	18.9	9.23	90.6	9.4	14.6
	245.6 x 212.7 x 100	25	88.1	11.9	8.90	82.1	17.9	9.19	90.3	9.7	14.6
	122.8 x 106.35 x 100	49	88.5	11.5	8.88	81.4	18.6	9.22	91.7	8.3	14.4
	245.6 x 212.7 x 100	49	89.0	11.0	8.85	80.8	19.2	9.24	90.5	9.5	14.6
94	122.8 x 106.35 x 200	25	74.2	25.8	9.01	67.3	32.7	9.22	83.7	16.3	14.4
	245.6 x 212.7 x 200	25	77.2	22.8	8.92	67.8	32.2	9.20	81.6	18.4	14.6
	122.8 x 106.35 x 200	49	77.5	22.4	8.90	68.3	31.7	9.19	85.3	14.7	14.3
	245.6 x 212.7 x 200	49	77.4	22.6	8.91	69.9	30.1	9.22	83.1	16.9	14.5
144	122.8 x 106.35 x 300	25	68.0	32.0	8.94	54.4	45.6	9.28	76.2	23.8	14.5
	245.6 x 212.7 x 300	25	69.7	30.3	8.89	54.4	45.6	9.28	77.1	22.9	14.4
	122.8 x 106.35 x 300	49	66.5	33.5	8.98	55.6	44.4	9.26	78.1	21.9	14.4
	245.6 x 212.7 x 300	49	68.6	31.4	8.92	56.8	43.2	9.23	76.3	23.7	14.5

η [Å]	Box size [Å^3]	N_0	Benzene, 700K			Naphthalene, 900K		
			P_{2D}	P_{3D}	$\Delta\mu^{\ominus}$	P_{2D}	P_{3D}	$\Delta\mu^{\ominus}$
19	122.8 x 106.35 x 50	25	81.2	18.8	22.7	87.2	12.6	28.3
	245.6 x 212.7 x 50	25	82.3	17.7	22.6	88.1	11.9	28.2
	122.8 x 106.35 x 50	49	80.3	19.7	22.8	86.4	13.6	28.5
	245.6 x 212.7 x 50	49	82.4	17.6	22.6	87.9	12.1	27.8
44	122.8 x 106.35 x 100	25	69.3	30.7	22.4	78.6	21.4	28.0
	245.6 x 212.7 x 100	25	67.7	32.3	22.5	79.1	20.9	27.9
	122.8 x 106.35 x 100	49	67.2	32.8	22.6	77.0	23.0	28.1
	245.6 x 212.7 x 100	49	68.3	31.7	22.5	79.3	20.7	27.7
94	122.8 x 106.35 x 200	25	51.4	48.6	22.4	65.8	34.2	27.7
	245.6 x 212.7 x 200	25	52.9	47.1	22.3	66.3	33.7	27.7
	122.8 x 106.35 x 200	49	51.0	49.0	22.4	64.2	35.8	27.9
	245.6 x 212.7 x 200	49	51.8	48.2	22.4	65.0	35.0	27.8
144	122.8 x 106.35 x 300	25	42.4	57.6	22.3	57.7	42.3	27.6
	245.6 x 212.7 x 300	25	43.0	57.0	22.3	58.7	41.2	27.5
	122.8 x 106.35 x 300	49	38.8	60.2	22.5	55.0	45.0	27.8
	245.6 x 212.7 x 300	49	42.6	57.4	22.3	56.6	43.4	27.7

Table 10.1: Equilibrium sampling results for the five studied molecules in the 16 different setups. The equilibrium probability of the desorbed and adsorbed states are reported, together with the standard chemical potential difference obtained from them.

arbitrary high simulation temperatures were used. For xenon and oxygen a temperature of 300 K is enough, while for the other systems higher temperatures were used, namely 500 K for alaninol, 700 K for benzene, and 900 K for naphthalene. This setup, together with the small used friction coefficient (0.1 ps^{-1}) in the Langevin integrator, increases the probability to populate the desorbed state and allows for fast desorption kinetics.

For each of the five molecules, 16 simulations were carried out, which differ in the size of the graphene substrate (a small or a big layer), the number of molecules in the simulation box (25 or 49), and the size of the box along the z dimension (50, 100, 200, or 300 Å). For each simulation the probability densityas a function of the distance from the surface was monitored; see Fig. 10.3 left column. All systems show an almost gaussian peak centered between 3 and 4 Å from the surface, which corresponds to the adsorption distance. The only exception is benzene, which clearly shows a pick at about 3.6 Å, but also a prominent shoulder at about 4.8 Å, probably corresponding to a T-shaped adsorbed conformation.

For all five adsorbates these peaks are quite sharp, and decay to a small value, corresponding to the probability density of the desorbed state, 6 Å away from the surface. Based on this observation, a molecule was assigned to the adsorbed state if its separation from the surface was lower than 6 Å, otherwise to the desorbed state. This allows the calculation of the adsorbed and desorbed probability at equilibrium counting the frequency of populating the adsorbed and desorbed states, which provides access to the standard difference in chemical potential thanks to Eq. 10.9.

The obtained probabilities and standard difference in chemical potentials are reported in Tab. 10.1. The first important observation, which confirms the theory, is that at constant extension ratio η the adsorption probability is constant with the initial number of molecules, or the size of the graphene layer used; see both Tab. 10.1 and Fig. 10.3 right column. This is also apparent from the fact that from the 16 simulation runs, only four different equilibrium probabilities were measured, corresponding to the different values of the extension parameter η. The second observation is that the adsorbed probability is a function of η, in particular the larger is the simulation box along the z direction, the higher the probability to populate the desorbed state. Third, the difference in chemical potential in all cases converges to the same standard value, if normalized by η. Averaging the standard chemical potentials obtained from the 16 simulations, a reference value is obtained, which will be used to compare the analytical ideal gas and enhanced sampling results in the next sections. These are reported in Tab. 10.4, along with the statistical errors evaluated from the standard deviation over the 16 simulations. These errors are very small, and range from 0.03 $kcal\,mol^{-1}$ for oxygen to 0.3 $kcal\,mol^{-1}$ for naphthalene.

Molecule	Atoms	3D Tran.	3D Rot.	3D Vibr.	2D Tran.	2D Rot.	2D Vibr.
Xenon	1	3	0	0	2	0	1
Oxygen	2	3	2	1	2	1	3
Alaninol	14	3	3	36	2	1	39
Benzene	12	3	3	30	2	1	33
Naphthalene	18	3	3	48	2	1	51

Table 10.2: Subdivision of the 3N degrees of freedom for one molecule in the gas and adsorbed states used to evaluate analytically the difference in free energy upon adsorption.

Ideal Gas Approximation

A straightforward approach to estimate of the standard chemical potential difference is trough Eqs. 10.2 and 10.3, for which only the minimized structures of the molecule both in vacuum and on the surface are required, which are accessible in a risible computer time. For a molecule of N atoms, 3N degrees of freedom need to be analytically integrated to evaluate its chemical potential. While the subdivision in 3 translations, 3 rotations, and $3N-6$ vibrations is standard for a non-linear molecule in vacuum, the assignment of the degrees of freedom in the adsorbed state is more arbitrary. In this work, the adsorbed molecule is assumed to freely translate on the surface and freely rotate along the axis perpendicular to the surface. This implies that 2 translations and 1 rotation are assigned to the translation and rotation of the molecule, with the remaining $3N-3$ degrees of freedom considered as vibrations. The partition scheme used to decompose the degrees of freedom is reported in Tab. 10.2.

The calculated chemical potentials, decomposed in electronic, translational, rotational, and vibrational contributions, are reported in Tab. 10.3. From a first overview, it is possible to see that the largest cost of adsorption comes from the translational and rotational degrees of freedom. This is natural, since confinement on the surface causes a net loss of these degrees of freedom. This is partially compensated by the three gained vibrations, for which the associated chemical potential can be an important stabilizing contribution, since it is of comparable magnitude with the adsorption energies. The high translational entropy cost upon adsorption also justifies the strong dependence of the adsorbed probability with the extension ratio η.

Finally, it is possible to compare how good is this analytical approach if compared with the previous standard chemical potential differences obtained via equilibrium simulations; see Tab. 10.4. Not surprisingly, the analytical solution for the monoatomic xenon is in perfect agreement with the equilibrium results. Although expected, the ideal gas approximation for the adsorbed state involves two assumptions: first the gained vibrational mode along the z direction is harmonic, second the surface is homogeneous, which means all points on the surface have the same adsorption energy. Obviously, both assumptions are incorrect, but the good agreement with the equilibrium sampling result

	μ_{elec}	$\mu^{\ominus}_{tr}$	μ_{rot}	μ_{vib}	$\mu^{\ominus}_{tot}$
		Xenon, 300K			
3D	0.00	-8.77	0.00	0.00	-9.11
2D	-3.78	5.04	0.00	-1.13	-0.21
3D $\rightarrow$ 2D	-3.78	13.81	0.00	-1.13	8.90
		Oxygen, 300K			
3D	0.00	-7.51	-3.09	0.97	-9.63
2D	-3.98	5.88	-1.51	-0.23	0.16
3D $\rightarrow$ 2D	-3.98	13.39	1.58	-1.20	9.79
		Alaninol, 500K			
3D	0.00	-14.54	-12.24	41.32	14.54
2D	-8.87	8.44	-4.15	36.25	31.67
3D $\rightarrow$ 2D	-8.87	22.98	8.09	-5.07	17.13
		Benzene, 700K			
3D	0.00	-21.14	-14.20	40.27	4.93
2D	-10.96	11.29	-4.24	32.48	28.57
3D $\rightarrow$ 2D	-10.96	32.43	9.96	-7.79	23.64
		Naphthalene, 900K			
3D	0.00	-29.19	-23.80	51.64	-1.35
2D	-17.44	13.19	-8.68	40.45	27.52
3D $\rightarrow$ 2D	-17.44	42.38	15.12	-11.19	28.87

Table 10.3: Analytical evaluation of the standard chemical potentials. All the electronic, translational, rotational, and vibrational contributions are reported in unit of kcal mol^{-1}.

suggests that these errors are very small. Goes without saying, the good agreement can be due also to a possible errors compensation.

Regarding the other four molecules, the ideal gas approximation is also good, with errors on the order of 1 kcal mol^{-1}. The only exception is alaninol, which has a discrepancy with the equilibrium sampling of about 2.6 kcal mol^{-1}. This can be explained considering that this molecule has more than one adsorbed conformations, which provides a larger conformational space that is neglected by the analytical expressions.

Confinement Free Energy

The evaluation of the standard chemical potential difference by both equilibrium sampling and analytical statistical mechanics was shown in the previous sections. Both methods have strengths and weaknesses. Equilibrium sampling can give the correct answer, but at the cost of running the simulation at a temperature where desorption is favored. Analytical expressions in the ideal gas approximation give qualitatively correct answer at a ridiculous computational cost, but the performances worsen as the complexity (name flexibility) of the molecule increases.

To overcome these limitations enhanced sampling techniques, like the confinement free energy method[264,265], can be used. In this method, the molecule on the surface is gradually restrained to the harmonic state for which the chemical potential is known analytically. The confinement allows to evaluate the anharmonicity associated to the

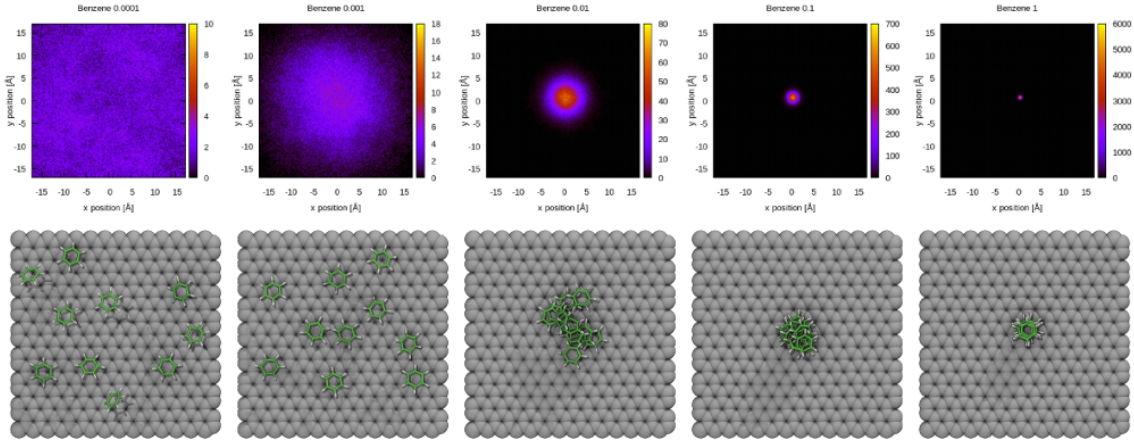


Figure 10.4: Top row: projection of the position of the center of mass of benzene on the graphene surface along the confinement simulations. Bottom row: ten superimposed frames from the confinement dynamics at each force constant. At low restraining force constants the whole surface is explored, while at higher forces the available surface is gradually confined.

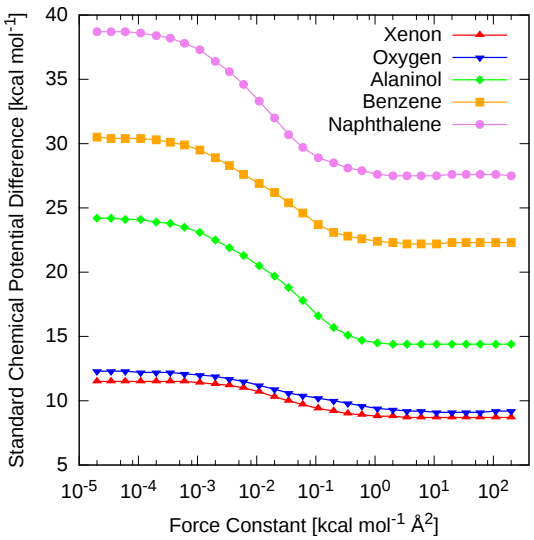


Figure 10.5: Standard chemical potential differences for the five studied systems as a function of the restraining force constant in the confinement simulations. All curves correctly converge for forces higher than 10 kcal mol^{-1} Å^{-2}.

	ES	IG	CF
Xenon, 300K	8.93 ± 0.05	8.9	8.7
Oxygen, 300K	9.23 ± 0.03	9.8	9.5
Alaninol, 500K	14.5 ± 0.2	17.1	14.4
Benzene, 700K	22.5 ± 0.1	23.6	22.3
Naphthalene, 800K	27.9 ± 0.3	28.9	27.6

Table 10.4: Comparison between the standard differences in the chemical potential obtained by equilibrium sampling (ES), ideal gas approximation (IG), and confinement (CF). All values are reported in kcal mol^{-1}.

hindered rotation of the molecule on the surface, the movement along the z axis, the roughness of the surface, and all the internal degrees of freedom, such as torsions or internal rotations. Moreover, there is no need to sample desorption events, since the chemical potential of the adsorbed state is analytical and that of the adsorbed state can be evaluated in the 6 Å layer from the surface.

For the small molecules used in this work, the major problem of the confinement approach is sampling the translational degrees of freedom on the surface. Although a small surface area (34.384 x 34.032 $Å^2$) was used in the simulations, runs of 20 ns are necessary to fully sample the translational phase space at low restraining force constants. When the force increases, the space to sample decreases, which reduces the required simulation time quite significantly, even though in this work the same number of steps has been taken in all simulations. Fig. 10.4 shows the position of the center of mass of a benzene molecule along the dynamics for five representative force constants. While at the two lowest restrain strengths the whole surface is equally explored, at restrain strength larger than 1 kcal mol^{-1} $Å^{-2}$ the space is confined, to a very small region close to the reference minimum. A second problem arising from the numerical integration of the translational space, is that the resulting free energy of confinement is "dirty" including the very different contributions coming from the translational degree of freedom and the anharmonicity associated to the internal torsions of the molecule, the roughness of the surface, and the hindered rotations. Nevertheless, the use of a small surface and long simulation time allows for an accurate determination of the standard difference in chemical potentials by confinement simulation. This is plotted in Fig. 10.5 as a function of the applied force constant. The profiles for all five molecules smoothly converge at force constants higher than 10 kcal mol^{-1} $Å^{-2}$. Looking at the obtained values, summarized in Tab. 10.4, the confinement results are in perfect agreement with equilibrium sampling for all five molecules, with a maximum error of 0.3 kcal mol^{-1}. This shows the ability of the method to recover the equilibrium sampling result using a clever, but sometimes tricky, combination of numerical and analytical integrations of the partition function of the adsorbed and desorbed states.

Application to 2D Molecular Self-Assembly

The results obtained in this work are discussed in the context of molecular self-assembly at the solid-liquid interface. Self-assembly at surfaces can be exemplified as the process where free molecules in solutions adsorb on a substrate, and aggregate to form a 2D periodic monolayer (SAM). Thus, there are at least three states in equilibrium: the monomer in solution (A_{3D}), the free molecule on the surface (A_{2D}), and the SAM

$$A_{3D} \rightleftharpoons A_{2D} \rightleftharpoons \text{SAM} \tag{10.16}$$

The aim of this section is to study how the probabilities of the three states change as a function of the initial concentration in solution.

As modeled in this work, the chemical potentials of A_{3D} and A_{2D} can be written as a function of the concentration via Eq. 10.2 and 10.3, respectively. Here it is useful here to rewrite Eq. 10.3 as function of the concentration in solution, writing explicitly the dependence from η as

$$\mu_{2D} = \mu_{2D}^{\ominus} + kT \ln \frac{\eta}{\eta^{\ominus}} + kT \ln \frac{C_{3D}}{C_{3D}^{\ominus}} \tag{10.17}$$

In this expression, analogous to Eq. 10.3, η has been used to convert a surface concentration to a volume concentration. Last, the chemical potential of the SAM is reduced to the chemical potential of the unit cell μ_{uc}, which is not concentration dependent in the limit of large architectures (see chapter 5).

Equating the chemical potential of the monomer in solution and that of the unit cell, it brings to the definition of a critical concentration of self-assembly

$$C_{cac}^{3D} = C_{3D}^{\ominus} \exp\left(\frac{\mu_{uc} - \mu_{3D}^{\ominus}}{kT}\right) \tag{10.18}$$

which, as discussed previously, sets the boundary between the monomeric and self-assembled state: at concentrations lower than C_{cac}^{3D}, the monomer in solutions is preferred, whereas the SAM prevails at higher concentrations. In the same way it is possible to define a critical concentration equating the chemical potentials of A_{2D} and of the SAM

$$C_{cac}^{2D} = C_{3D}^{\ominus} \frac{\eta^{\ominus}}{\eta} \exp\left(\frac{\mu_{uc} - \mu_{2D}^{\ominus}}{kT}\right) \tag{10.19}$$

Finally, equating the chemical potentials of A_{3D} and A_{2D} does not result in the definition of a critical concentration since, as shown above, this equilibrium is not concentration dependent, and the probabilities of the two states depend only on η.

Based on these results, it is thus possible to study the equilibrium in Eq. 10.16 as a function of the initial concentration of monomers. First, at high concentrations ($\gg C_{cac}^{3D}$ and C_{cac}^{2D}), the SAM is thermodynamically preferred with a probability close to unity; see Fig. 10.6. At very low concentrations ($\ll C_{cac}^{3D}$ and C_{cac}^{2D}), the SAM is not favored and only the states A_{3D} and A_{2D} can be populated. As shown by Eq. 10.7, their relative probability depends on η and the adsorption equilibrium constant, but is constant for any initial concentration. Two scenarios can be drawn. If $\eta < K_{eq}$, the chemical potential of A_{2D} is lower and A_{2D} is preferred; left column in Fig. 10.6. Conversely, if $\eta > K_{eq}$, A_{3D} is preferred at low concentrations of monomers; right column in Fig. 10.6. What is left are the middle-range concentrations, in particular between C_{cac}^{3D} and C_{cac}^{2D}, where the actual phase transition between the free molecule and the SAM takes place. Unfor-

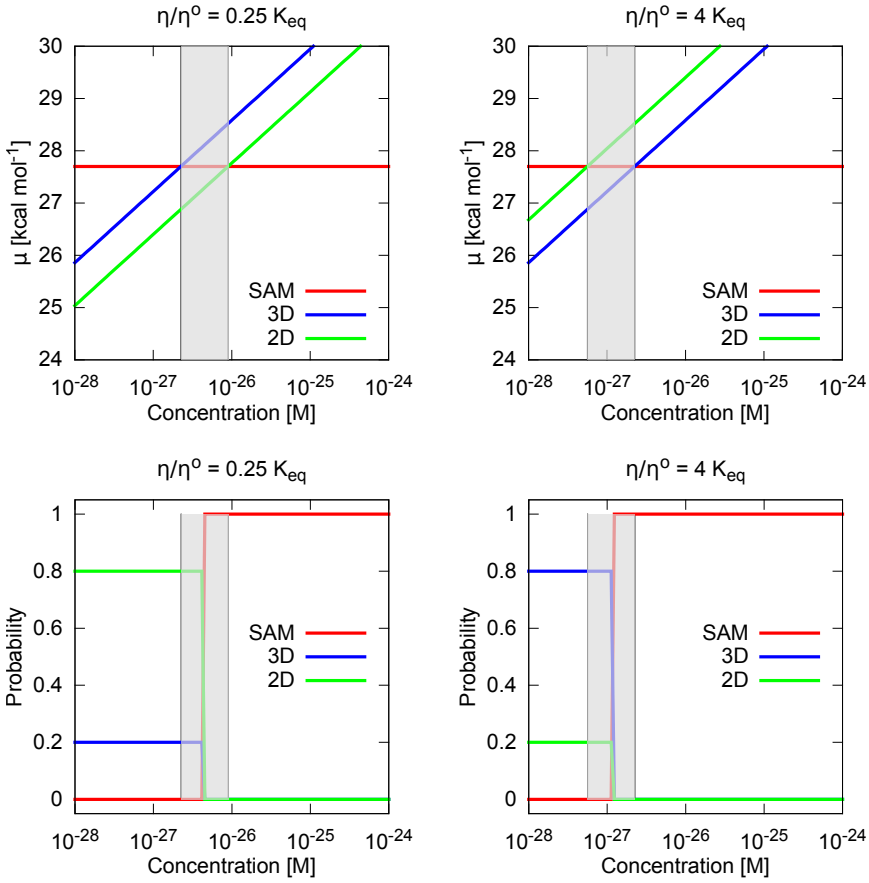


Figure 10.6: Chemical potential (top row) and equilibrium probabilities (bottom row) of A_{2D}, A_{3D} and the SAM as a function of the initial concentration. Two examples are depicted: $\eta < K_{eq}$ (left) and $\eta > K_{eq}$ (right). These examples refer to the adsorption equilibrium and self-assembly of trimesic acid (TMA) into the chickenwire (CHK) architecture. The chemical potentials are: $\mu^{\circ}_{3D} = 64.0$ kcal mol^{-1}, $\mu^{\circ}_{2D} = 56.8$ kcal mol^{-1} (calculated using Eq. 10.2 and 10.3) and $\mu_{uc} = 27.7$ kcal mol^{-1} as calculated in chapter 5. The non-physical concentrations are due to the leaking of a proper treatment of the solvent effects.

tunately, there exists no straightforward analytical solution for this concentration range. However, it is interesting to note is that two distinct self-assembly mechanisms can be drawn, following the two scenarios above. When $\eta \ll K_{eq}$, A_{2D} is the preferred state at low concentrations while the A_{3D} state is not populated. Once C^{2D}_{cac} is reached, the molecules on the surface start to assemble forming the SAM. This is the most expected scenario with molecules that first adsorb on the surface and associate only at higher concentrations. However, when $\eta \gg K_{eq}$, A_{3D} is the only populated state at low concentration, which leads to the curious situation where, once the C^{3D}_{cac} concentration is reached, the SAM forms without populating the disordered state A_{2D},which can thus be considered a kinetic intermediate.

The same conclusions can be reached playing with the relative magnitude of the three chemical potentials μ°_{3D}, μ°_{2D}, and μ_{uc}. For a big molecule with high adsorption energy

the A_{2D} state will be preferred as its chemical potential μ°_{2D} will be lower than the chemical potential in solution μ°_{3D}. Such molecule would thus follow the first more classical mechanism where disordered adsorption is followed by self-assembly. A small molecule with low adsorption energy but strong interactions in the SAM would instead follow the second scenario, where the entropic cost of 2D confinement on the surface cannot be overcome by the interaction with the surface only, and the adsorbed state will be promoted by the formation of the SAM. In this case, the A_{2D} state will not be populated at any concentration and the molecule would transfer directly from A_{3D} to the SAM.

10.5 TAKE HOME MESSAGES

In conclusion, this work sheds new light onto the theoretical understanding and modeling of molecular adsorption at equilibrium on a flat, periodic surface in the limit of diluted solutions. Perhaps surprisingly, the available volume/surface aspect ratio was shown to be the only parameter dictating the adsorption probability, while the solute concentration has no effect on it. Based on these theoretical results, the adsorption of five simple molecules on graphene was analyzed by equilibrium sampling simulations with the goal of confirming the theory, and providing reference values of the equilibrium constant for the adsorption process. In this respect, the confinement method, which includes anharmonic contributions, was shown to accurately reproduce the equilibrium sampling results. Finally, the effects of these results on self-assembly at surfaces have been discussed, with the main result that two pathways to self-assembly can be possibly identified depending on the environmental conditions (the parameter η) and the chemical potential differences to adsorb and self-assemble. This last result is potentially very interesting, but more theoretical and experimental works is required to understand its impact.

11

PERSPECTIVES AND CONCLUSIONS

In this dissertation an original theoretical framework to study molecular self-assembly at surfaces and interfaces has been presented. As stated in the introduction, the main goal was to identify and understand the factors that control the probability of self-assembly at chemical equilibrium. To tackle this problem, the chemical potential of the unit cell μ_{uc} was shown to be the quantity of interest to access the thermodynamic stability of a self-assembled monolayer (SAM). While approximated expressions of this quantity have been used in the literature, I demonstrated how the reduction of the free energy of the SAM to the chemical potential of the unit cell is justified in the limit of extended periodic architectures. This allowed to study the chemical equilibrium for the general self-assembly reaction $\alpha A \rightleftharpoons B$, which has led to the definition of the surface free energy γ, that is the second derived key quantity. Analyzing the dependence of γ on the monomer concentration it is possible to study how the stability of the SAM changes, allowing the comparison of the stability of different monolayers made of the same building block (surface polymorphism) or of different molecules (competitive self-assembly). The third derived key quantity is the critical concentration C_{cac} which, setting the boundary between the domain of dominance of the monomers and the SAM, is the perfect candidate to set an absolute scale for the stability of different SAMs.

The application of this theory to the case studies presented in the second part of this thesis showed a great diversity of examples for which the most important driving forces to self-assembly changed dramatically with the molecular building blocks. The first example that was showed, the self-assembly of carboxylic acids, revealed a strong energy/entropy interplay, where entropy actually dominated the self-assembly. In the self-assembly of the DAE isomers instead, the self-assembly was shown to be fully described by energetic considerations only, where entropy played essentially no role. Last perchlorocoronene was shown to have a better self-assembly propensity than coronene thanks to both energetic and entropic considerations. The theoretical framework developed in this thesis granted the possibility to study this plurality of behaviors in a consistent manner, highlighting how the surface free energy allows to understand and predict self-assembly at surfaces.

The main advantage of this approach with respect to literature methods is in its foundations. As highlighted in the first part of this thesis, all the theoretical results are here obtained by the application of basic chemical equilibrium and statistical thermodynamics relations. Thanks to this, the *ab initio* approach developed in this thesis allows to identify the advantages and drawbacks of the derived relations, suggesting the points to improve or the approximations to remove to get a more quantitative description of

self-assembly. In fact, although the developed method was shown to be adequate to tackle real technological questions, as demonstrated in the second part of this thesis, all the results have only a qualitative agreement with the experiments, reproducing and predicting nicely the experimental trends, but failing at more strict comparisons, e.g. for the critical concentrations. The strongest approximation introduced in this work is obviously the lack of a proper description of the solvent effects. While these can be included using both implicit or explicit solvent models, their accurate estimation is particularly critical due to both the nature of the solvent-solute interaction, which is the result of the cancellation of large entropic and enthalpic contributions and is thus very difficult to account for, and the actual solvents studied, which are not commonly used in molecular modeling and thus need a proper (non trivial) parametrization. Other approximations that can be improved are, for example, the treatment of the vibrational free energy (here modeled as simple normal modes), the inclusion of anharmonicity and conformational entropy (here neglected), or the use of a better model for the interactions (here a simple classical force field).

Making a step back, it is possible to view the current work in perspective of the more general problem of predicting from scratch the self-assembled monolayers formed by a given arbitrary molecule. The full process should involve at least two steps in which at first a series of putative assemblies which are able to tile the surface are generated, and then the stability of each of them is studied as a function of the environmental conditions, such as temperature or monomer concentration, to access their domains of stability. The work put forward in this dissertation is focused only on the second step of this ambitious project. Moreover, to better study the chemical equilibrium between the disassembled and self-assembled state is actually important to define these two states properly. While the definition of the second is straightforward, the first is more ambiguous. The actual question to answer here is if the monomeric state in solution is the real disassembled state or if it is in competition with e.g. a disassembled state physisorbed on the surface or any other oligomeric state. Partial answer to this question was given in the last chapter, where the adsorption equilibrium was studied, and required the introduction of other environmental quantities (like the η parameter) to proper describe it.

Part IV

Appendix

A

ENERGY PER UNIT CELL LIMIT

Given a layer of $n \times n$ cells it will have $(n-2)^2$ internal cells and the remaining $4(n-1)$ will be border cells with missing interactions. The energy of such assembly assembly can be approximated as

$$E(n) \simeq (n-2)^2 E_{uc} + 4(n-1)E_{border}$$

Given

$$\alpha = n_{cells} \cdot n_{uc}$$

and

$$n_{cells} = n^2$$

the previous expression can be written as function of α as

$$\begin{aligned} E(\alpha) &\simeq \left(\sqrt{\frac{\alpha}{n_{uc}}} - 2\right)^2 E_{uc} + 4\left(\sqrt{\frac{\alpha}{n_{uc}}} - 1\right) E_{border} = \\ &= \alpha E'_{uc} + 4\left(E_{uc} - E_{border}\right)\left(1 - \sqrt{\frac{\alpha}{n_{uc}}}\right) \end{aligned}$$

When the energy of the self-assembly divided by α is studied:

$$\frac{E(\alpha)}{\alpha} \simeq E'_{uc} - \frac{4\left(E'_{uc} - E'_{border}\right)}{\sqrt{\alpha/n_{uc}}} + \frac{4\left(E'_{uc} - E'_{border}\right)}{\alpha/n_{uc}}$$

Discarding the third term, which is negligible if compared with the second, the equation reported in the Main Text is obtained

$$\frac{E(\alpha)}{\alpha} \simeq E'_{uc} - \frac{b}{\sqrt{\alpha/n_{uc}}}$$

where $b = 4\left(E'_{uc} - E'_{border}\right)$.

B

SOLVATION FREE ENERGY

Solvent effects on the probability of 2D self-assembly can be directly included in the definition of the surface free energy as

$$\gamma = \gamma_E + \gamma_S + \gamma_{solv} \tag{B.1}$$

where γ_E and γ_S correspond to the energetic and entropic contributions to the surface free energy (see above), and γ_{solv} is the contribution of the solvent which is defined as

$$\gamma_{solv} = \frac{1}{A'_{uc}} \Delta\mu'_{solv} \tag{B.2}$$

with $\Delta\mu'_{solv}$ being the solvation free energy change per molecule upon 2D self-assembly, that is

$$\Delta\mu'_{solv} = \mu'_{solv,sam} - \mu_{solv,A} - \mu'_{solv,sub} \tag{B.3}$$

where the three term in r.h.s. are the per-molecule contribution to the solvation free energy of the SAM, the monomer, and the portion of substrate covered by the SAM. A rigorous evaluation of $\Delta\mu'_{solv}$ is computationally challenging and would require intensive free energy calculations. Based on previous work[173,274], the solvation free energy is approximated here as a linear function of the solvent accessible surface area (SASA)

$$\mu_{i,solv} = \alpha \cdot SASA_i \tag{B.4}$$

with α being an empirical parameter which depends on the solvent. Introducing this result in Eq. B.3 for a SAM of N molecules and assuming the same α for the both SAM and the substrate, it yields

$$\Delta\mu'_{solv} = \frac{\alpha}{N} [SASA_{sam} - N \cdot SASA_A - SASA_{sub}] = \alpha \Delta SASA' \tag{B.5}$$

which provides numerical access to the solvent contribution to the surface free energy (Eq. B.2) from the only knowledge of the SASA of the SAM, monomer and substrate, and the value of α.

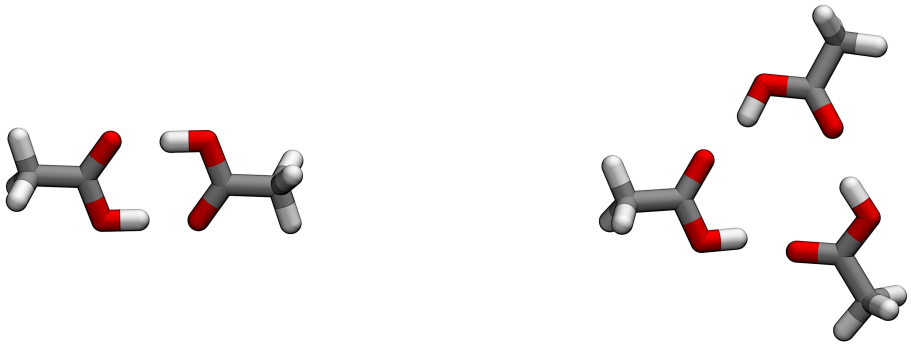
In this work, the solvation correction in Eq. B.5 was included in the analysis of TMA self-assembly to quantify its impact on the critical concentrations predicted by the theory. In the calculations, the solvent accessible surface area was evaluated using the APBS software[275] with van der Waals radii taken from the GAFF force field[123,124] and a probe radius of 1.4 Å. $\Delta SASA'$ was evaluated using model architectures of about 200

molecules (200 for CHK, 216 for FLW and 196 for SFW). The value of α was obtained from the solvation free energy of model compounds and their SASA using Eq. B.4. To this aim, the solvation free energy of TMA (the molecule studied) and coronene (as a model of graphene) have been determined by free energy perturbation (FEP) / molecular dynamics (MD) calculations in the apolar solvent toluene. For FEP, 10 and 21 windows of 1 ns each were used to evaluate the polar (electrostatic) and the non polar (van der Waals) contributions to the solvation free energy, respectively. All simulations were carried out using GROMACS 5.1.2[276], and the FEP analysis was performed using the Alchemical Analysis tool[277]. The μ_{solv} calculated for TMA and coronene were $-17.12\ \mathrm{kcal\,mol^{-1}}$ and $-19.35\ \mathrm{kcal\,mol^{-1}}$, respectively. Their SASA were 3.90 $\mathrm{nm^2}$ and 5.11 $\mathrm{nm^2}$, respectively, which provides values for the parameter α of $-4.39\ \mathrm{kcal\,mol^{-1}\,nm^{-2}}$ and $-3.79\ \mathrm{kcal\,mol^{-1}\,nm^{-2}}$. Hence, an approximated value of $-4\ \mathrm{kcal\,mol^{-1}\,nm^{-2}}$ for α appears to be reasonable for modeling the solvation free energy of TMA self-assembly using Eq. B.5. With this value of α and the $\Delta\mathrm{SASA}'$ for the CHK, FLW and SFW architectures, i.e. -3.66, -3.72, and $-3.85\ \mathrm{nm^2}$, respectively, the solvent contribution to the free energy of self-assembly per molecule (Eq. B.3) were 14.64, 14.88, and 15.40 $\mathrm{kcal\,mol^{-1}}$, for the different architectures. Introducing these results in Eq. B.2 yields surface free energy corrections of 12.09, 15.15, and 20.10 $\mathrm{kcal\,mol^{-1}\,nm^{-2}}$ for CHK, FLW, and SFW. As shown in the Main Text, this causes an almost systematic shift of the critical concentrations of about ten orders of magnitude.

C

HYDROGEN BOND CORRECTION FOR TRIMESIC ACID

From the force field energies obtained for the trimesic acid in the three assemblies (chickenwire, flower and superflower) it is clear that the GAFF force field is not adequate to describe the energetics involved in the carboxylic acid interaction. In fact, the superflower architecture is predicted to be the energetically most favored, while experimentally the chickenwire and the flower architectures are clearly more stable[33]. To understand the problem, the interaction energy for the two typical patterns of hydrogen bonds present in the chickenwire, flower and superflower architectures, see figure, were benchmarked. The first kind of interaction involves linear dimers of carboxylic acid groups and it is know to be highly stabilized by resonance, as a prototypical example of resonance-assisted hydrogen bond (RAHB)[278]. High level of theory *ab-initio* calculations on this dimer are available from the S66 database, where the dimerization energy of acetic acid dimer is -19.41 kcal mol^{-1} [279]. The second hydrogen bond pattern involves three carboxylic groups and, as opposite to the first pattern, no reference energy value was found present in literature.



Sketch representations of the linear (left) and trigonal (right) hydrogen bond patterns of carboxylic acid self-assembly.

Since no reference values for the trimer patter are present in literature, *ab-initio* calculation have been performed on both systems. The chosen level of theory is a density-fitted (DF) MP2 with an extended basis sets (the aug-cc-pVTZ and the cc-pVQZ have been used). The system was geometry optimized and then the interaction energy calculated using a counterpoise correction (CP) to minimize the basis set superimposition error (BSSE). The resulting dimerization energies are reported in the following table. When comparing the energy for the acetic acid dimer with the CCSD(T)/CBS literature value, the DF-MP2/cc-pVQZ gives clearly optimal results with an error of only 0.26 kcal mol^{-1}.

If the reference energy value of -16.41 kcal mol^{-1} is compared with the GAFF value of -14.80 kcal mol^{-1}, it is evident that this force field underestimates the interaction energy by -4.5 kcal mol^{-1}. In sharp contrast, for the trigonal pattern the recognition strength by GAFF matches the value predicted by DF-MP2/cc-pVQZ. A possible reason for this could be that while the nature of the interaction in the two cases are completely different (resonance effect in the first, mainly electrostatic in the second), the force field models both in a pure electrostatic way.

	Acetic acid dimer	Acetic acid trimer
df-MP2/aug-cc-pVTZ	-18.81	-24.47
df-MP2/cc-pVQZ	-19.15	-24.68
CCSD(T)/CBS[279]	-19.41	//
GAFF	-14.80	-24.78
Empirical Correction	-4.5	//

Interaction energies for the two patterns of hydrogen bonds evaluated at ab-initio (DF-MP2) and force field (GAFF) levels of theory. All values are reported in kcal mol^{-1}.

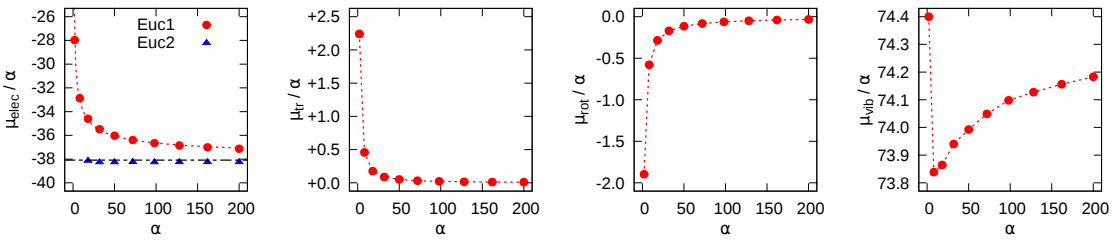
Given these results, the energy per unit cell for the three architectures of trimesic acid have been *a posteriori* corrected: for all hydrogen bond patterns of the first type found in the three architectures, a stabilization energy of -4.5 kcal mol^{-1} has been added, while for the second type of hydrogen bond pattern no correction is added. Focusing on just the first pattern of hydrogen bonds, the chickenwire architecture has 1 occurrence inside the unit cell, plus 4 occurrences shared with the surrounding cells. It means a total of 3 occurrences per unit cell (1 intra plus 4/2 inter), which correspond to a total energy correction of 3 times the -4.5 kcal mol^{-1}. Once normalized per the number of molecules inside the unit cell (2 for the chickenwire architecture) a final correction on E'_{uc} of -6.75 kcal mol^{-1} is calculated. For the flower architecture the hydrogen bond pattern count give 2 occurrences inside the cell and 8 between neighbors, which give a total correction on E'_{uc} of -4.5 kcal mol^{-1}. In the superflower architecture, no occurrences of the first pattern are present, so no correction on the energy per unit cell is applied. In Table 5.2, both the uncorrected and corrected values, marked with an asterisk (*), are reported.

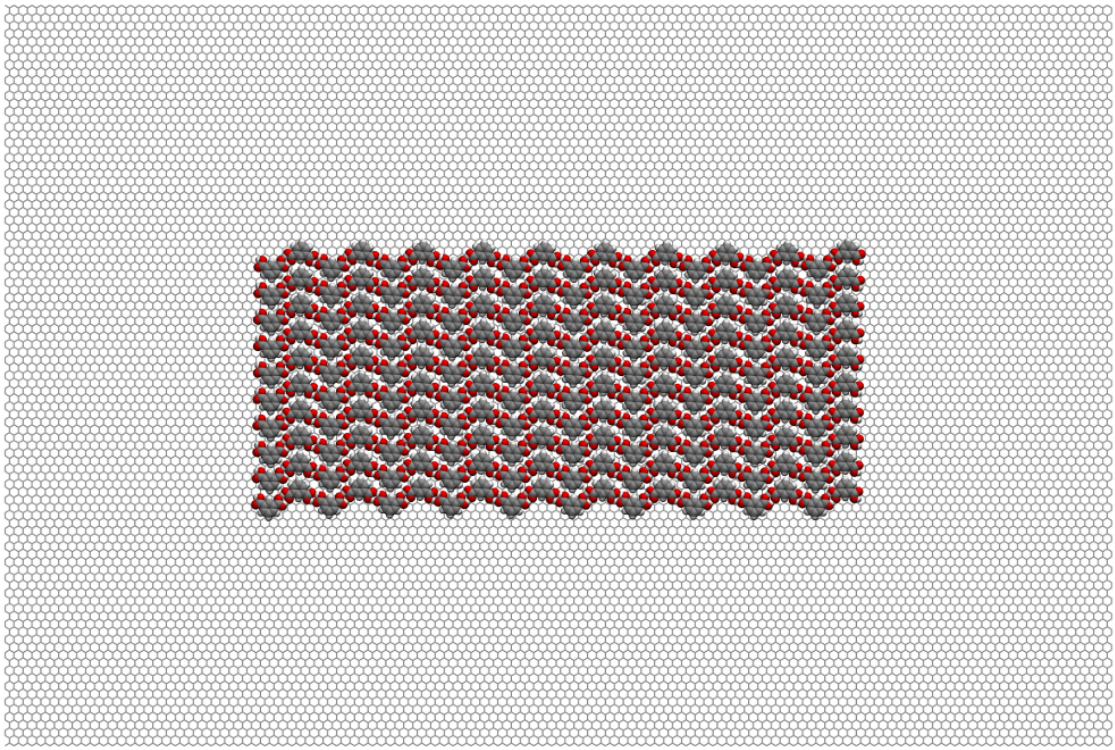
D

MODELED SELF-ASSEMBLED MONOLAYERS

In the following pages, all modeled architectures are shown, with the plots of the convergence of the unit cell chemical potential.

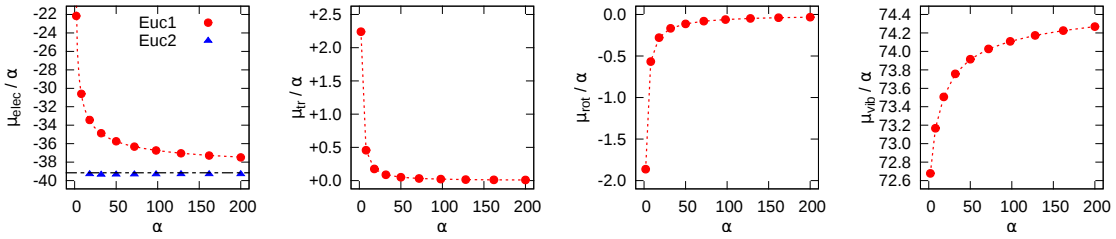
Isophthalic Acid

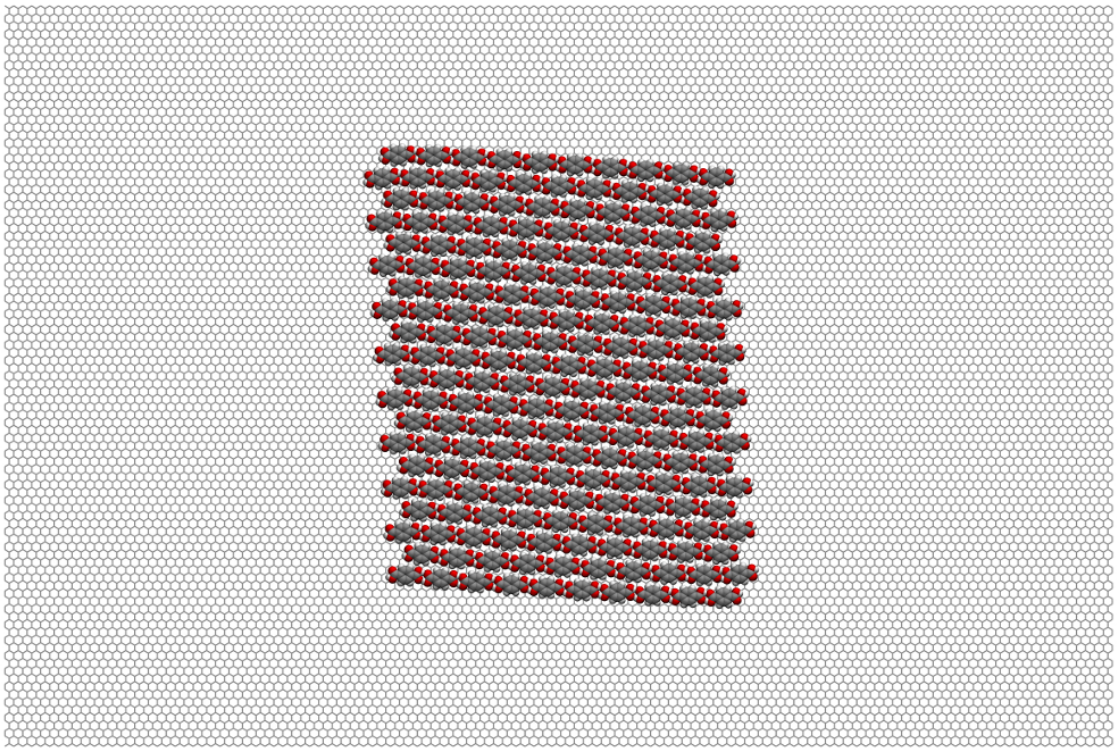




$\mu_{A,tr} =$	-9.0	$E'_{ads} =$	-20.8	$\mu'_{vib,uc} =$	74.2	$A'_{uc} =$	59.9
$\mu_{A,rot} =$	-8.4	$E'_{sam} =$	-18.5	$\mu'_{uc} =$	35.9	$\gamma =$	-37.4
$\mu_{A,vib} =$	75.7	$E'_{strain} =$	1.0	$\Delta\mu'_{vib} =$	-1.5	$\gamma_E =$	-59.1
$\mu_{A,tot} =$	58.3	$E'_{uc} =$	-38.3	$\Delta\mu' =$	-22.4	$\gamma_S =$	21.7
						$C_{cac} =$	$10^{-16.3}$

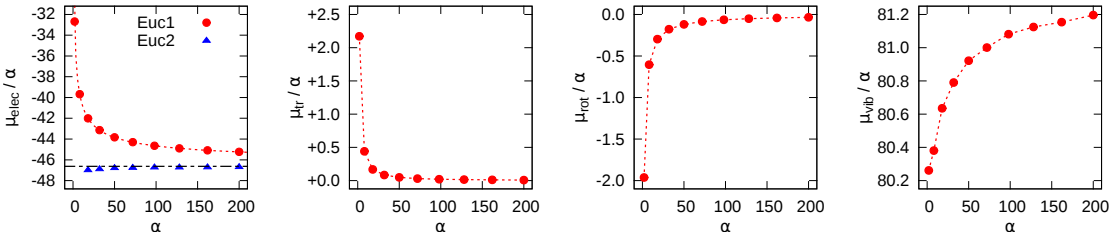
Terephthalic Acid

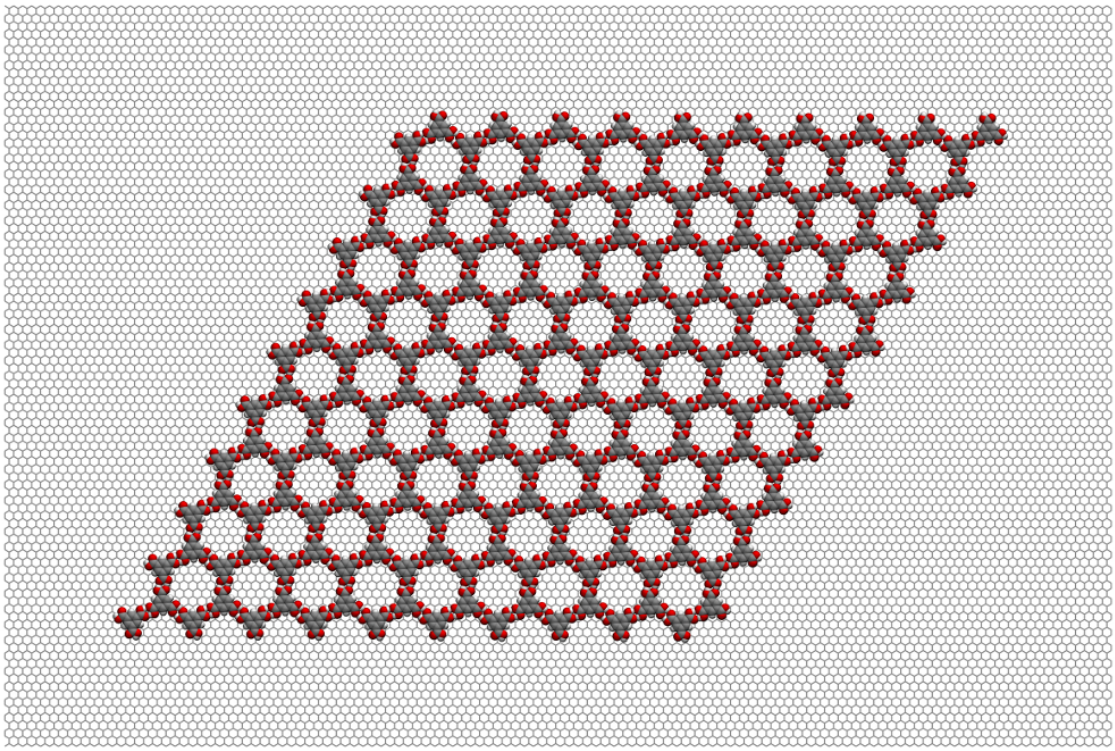




$\mu_{A,tr}$ =	-9.0	E'_{ads} =	-20.8	$\mu'_{vib,uc}$ =	74.3	A'_{uc} =	59.0
$\mu_{A,rot}$ =	-7.5	E'_{sam} =	-19.5	μ'_{uc} =	34.9	γ =	-41.2
$\mu_{A,vib}$ =	75.7	E'_{strain} =	0.9	$\Delta\mu'_{vib}$ =	-1.4	γ_E =	-61.7
$\mu_{A,tot}$ =	59.2	E'_{uc} =	-39.4	$\Delta\mu'$ =	-24.3	γ_S =	20.5
						C_{cac} =	$10^{-17.7}$

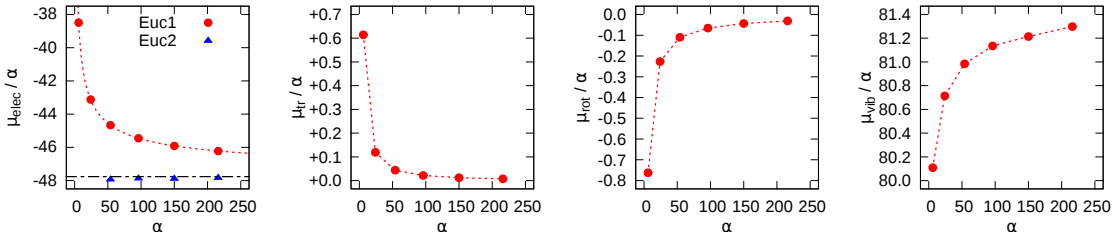
Trimesic Acid – Chickenwire Architecture

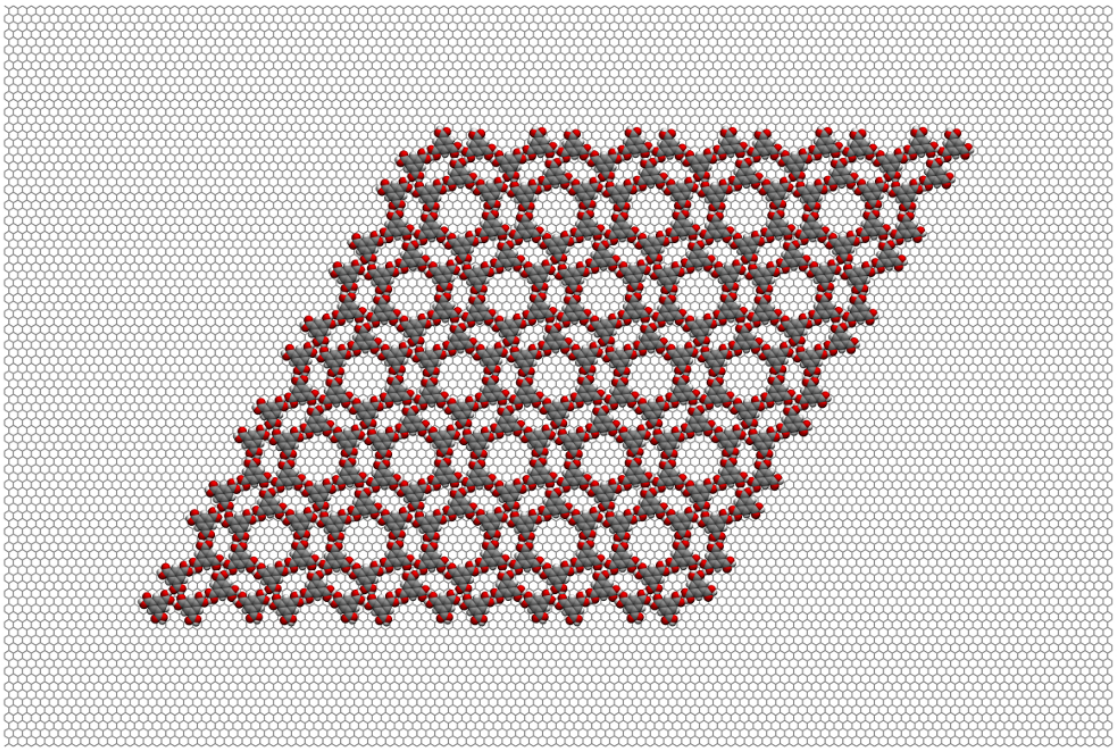




$\mu_{A,tr}$ =	-9.2	E'_{ads} =	-25.7	$\mu'_{vib,uc}$ =	81.2	A'_{uc} =	121.1
$\mu_{A,rot}$ =	-8.2	E'_{sam} =	-22.4	μ'_{uc} =	34.4	γ =	-24.4
$\mu_{A,vib}$ =	81.4	E'_{strain} =	1.3	$\Delta\mu'_{vib}$ =	-0.2	γ_E =	-35.8
$\mu_{A,tot}$ =	64.0	E'_{uc} =	-46.7	$\Delta\mu'$ =	-29.6	γ_S =	11.4
						C_{cac} =	$10^{-21.6}$

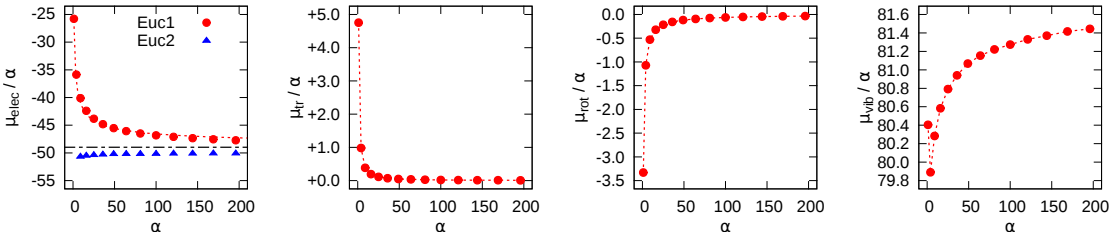
Trimesic Acid – Flower Architecture

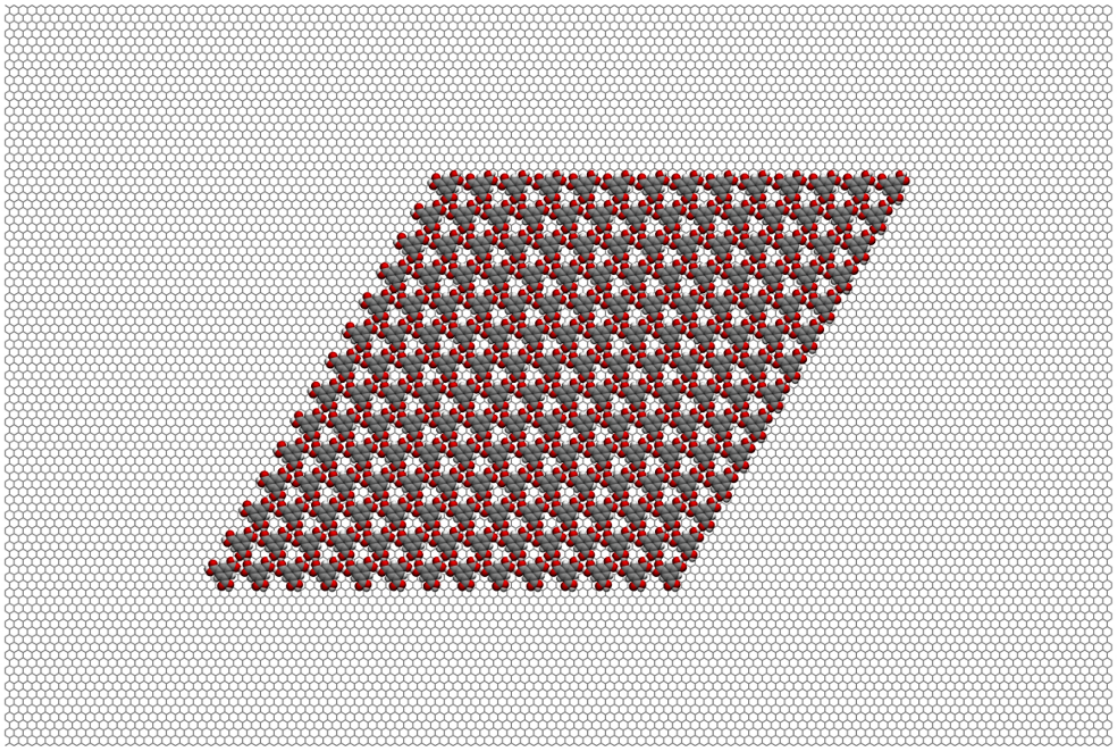




$\mu_{A,tr}$ =	-9.2	E'_{ads} =	-25.7	$\mu'_{vib,uc}$ =	81.3	A'_{uc} =	98.2
$\mu_{A,rot}$ =	-8.2	E'_{sam} =	-23.5	μ'_{uc} =	33.4	γ =	-31.2
$\mu_{A,vib}$ =	81.4	E'_{strain} =	1.3	$\Delta\mu'_{vib}$ =	-0.1	γ_E =	-45.3
$\mu_{A,tot}$ =	64.0	E'_{uc} =	-47.9	$\Delta\mu'$ =	-30.6	γ_S =	14.1
						C_{cac} =	$10^{-22.3}$

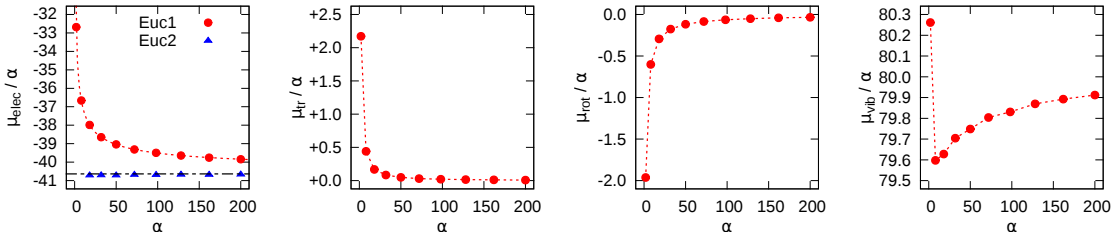
Trimesic Acid – Superflower Architecture

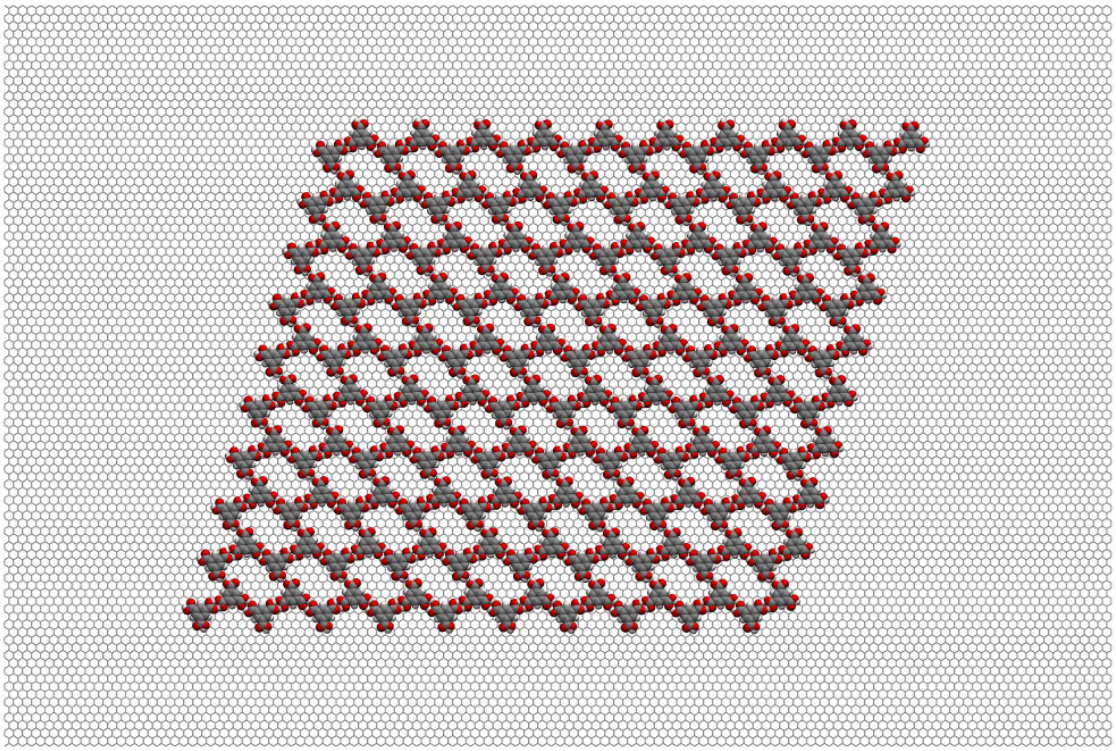




$\mu_{A,tr}$ =	-9.2	E'_{ads} =	-25.8	$\mu'_{vib,uc}$ =	81.4	A'_{uc} =	76.6
$\mu_{A,rot}$ =	-8.2	E'_{sam} =	-25.7	μ'_{uc} =	31.2	γ =	-42.8
$\mu_{A,vib}$ =	81.4	E'_{strain} =	1.3	$\Delta\mu'_{vib}$ =	0.0	γ_E =	-61.1
$\mu_{A,tot}$ =	64.0	E'_{uc} =	-50.2	$\Delta\mu'$ =	-32.8	γ_S =	18.3
						C_{cac} =	$10^{-23.9}$

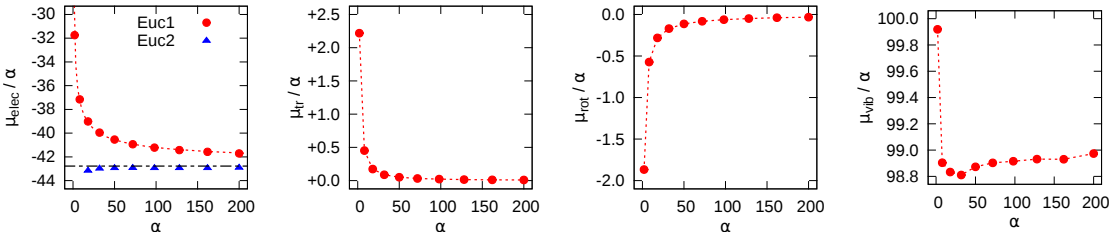
Trimesic Acid – Stripe (hypotetical) Architecture

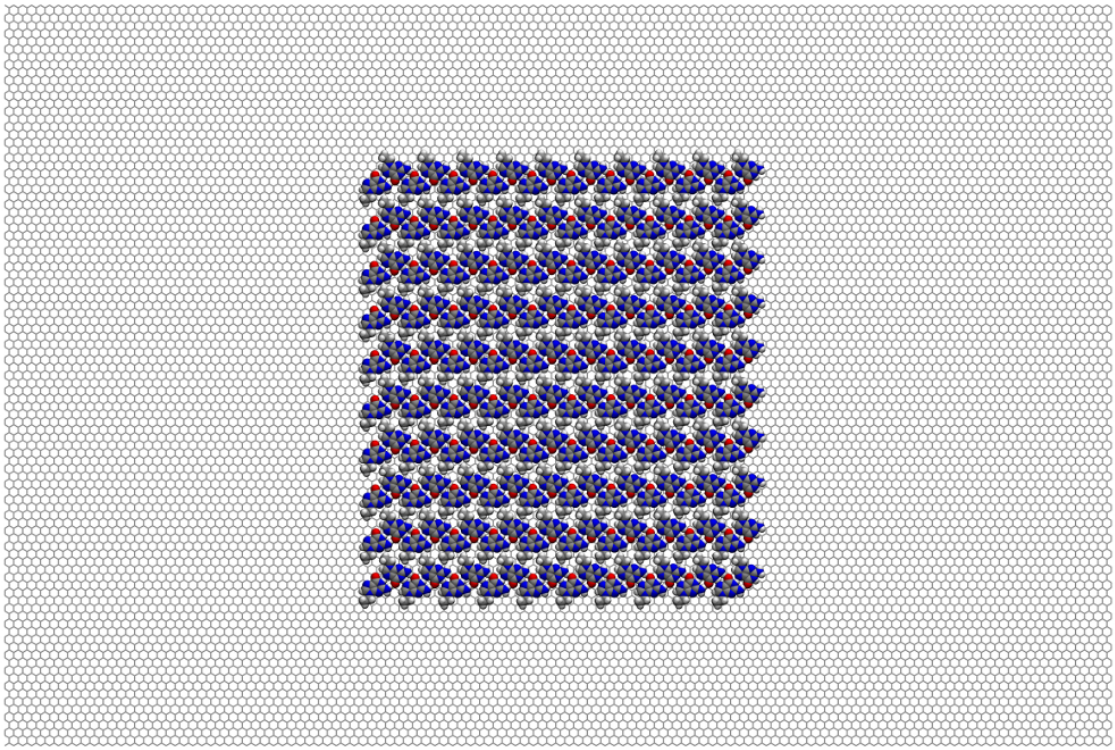




$\mu_{A,tr}$ =	-9.2	E'_{ads} =	-25.7	$\mu'_{vib,uc}$ =	79.9	A'_{uc} =	117.0
$\mu_{A,rot}$ =	-8.2	E'_{sam} =	-15.9	μ'_{uc} =	39.2	γ =	-21.2
$\mu_{A,vib}$ =	81.4	E'_{strain} =	0.9	$\Delta\mu'_{vib}$ =	-1.5	γ_E =	-32.3
$\mu_{A,tot}$ =	64.0	E'_{uc} =	-40.7	$\Delta\mu'$ =	-24.8	γ_S =	11.1
						C_{cac} =	$10^{-18.1}$

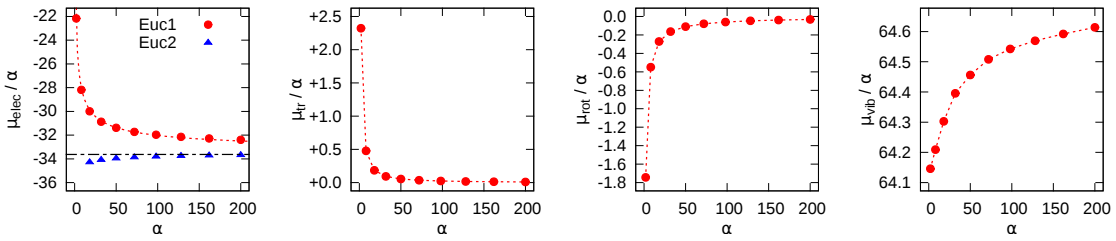
Ethyl Guanine

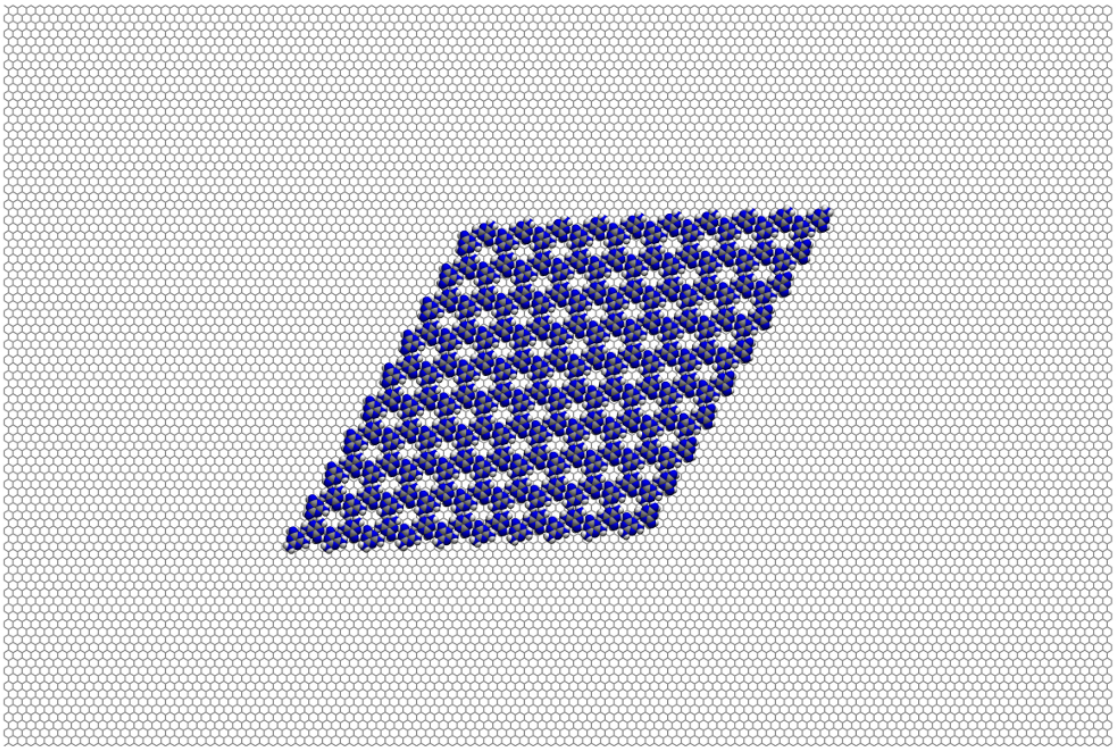




$\mu_{A,tr} =$	-9.0	$E'_{ads} =$	-23.5	$\mu'_{vib,uc} =$	99.0	$A'_{uc} =$	65.9
$\mu_{A,rot} =$	-8.4	$E'_{sam} =$	-20.9	$\mu'_{uc} =$	56.1	$\gamma =$	-41.9
$\mu_{A,vib} =$	101.1	$E'_{strain} =$	1.5	$\Delta\mu'_{vib} =$	-2.1	$\gamma_E =$	-61.3
$\mu_{A,tot} =$	83.7	$E'_{uc} =$	-42.9	$\Delta\mu' =$	-27.6	$\gamma_S =$	19.4
						$C_{cac} =$	$10^{-20.1}$

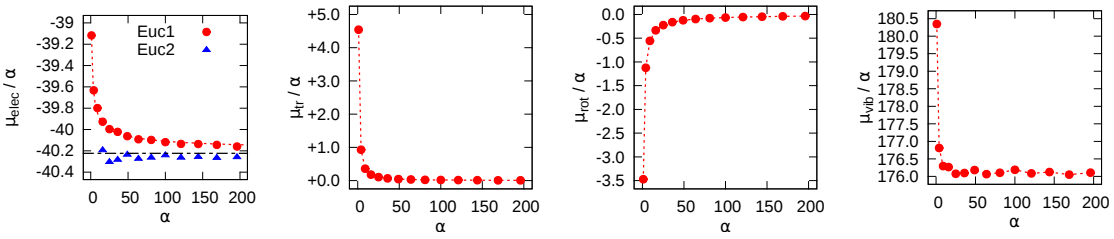
Melamine

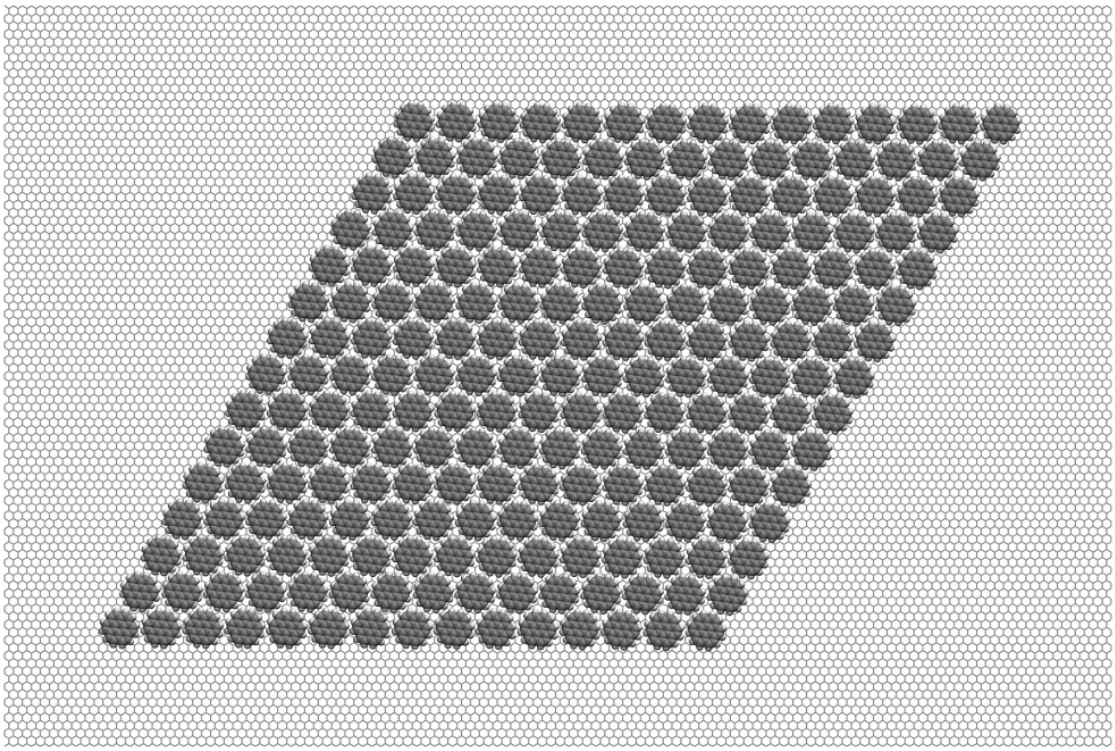




$\mu_{A,tr}$ =	-8.7	E'_{ads} =	-16.3	$\mu'_{vib,uc}$ =	64.6	A'_{uc} =	44.8
$\mu_{A,rot}$ =	-7.7	E'_{sam} =	-19.8	μ'_{uc} =	30.8	γ =	-42.0
$\mu_{A,vib}$ =	66.0	E'_{strain} =	2.3	$\Delta\mu'_{vib}$ =	-1.4	γ_E =	-68.5
$\mu_{A,tot}$ =	49.6	E'_{uc} =	-33.8	$\Delta\mu'$ =	-18.8	γ_S =	26.5
						C_{cac} =	$10^{-13.7}$

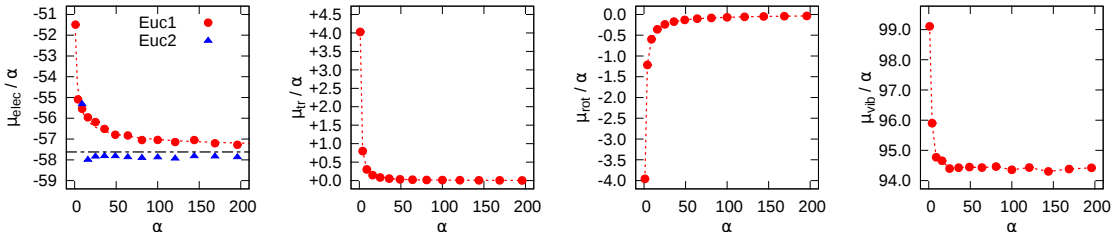
Coronene

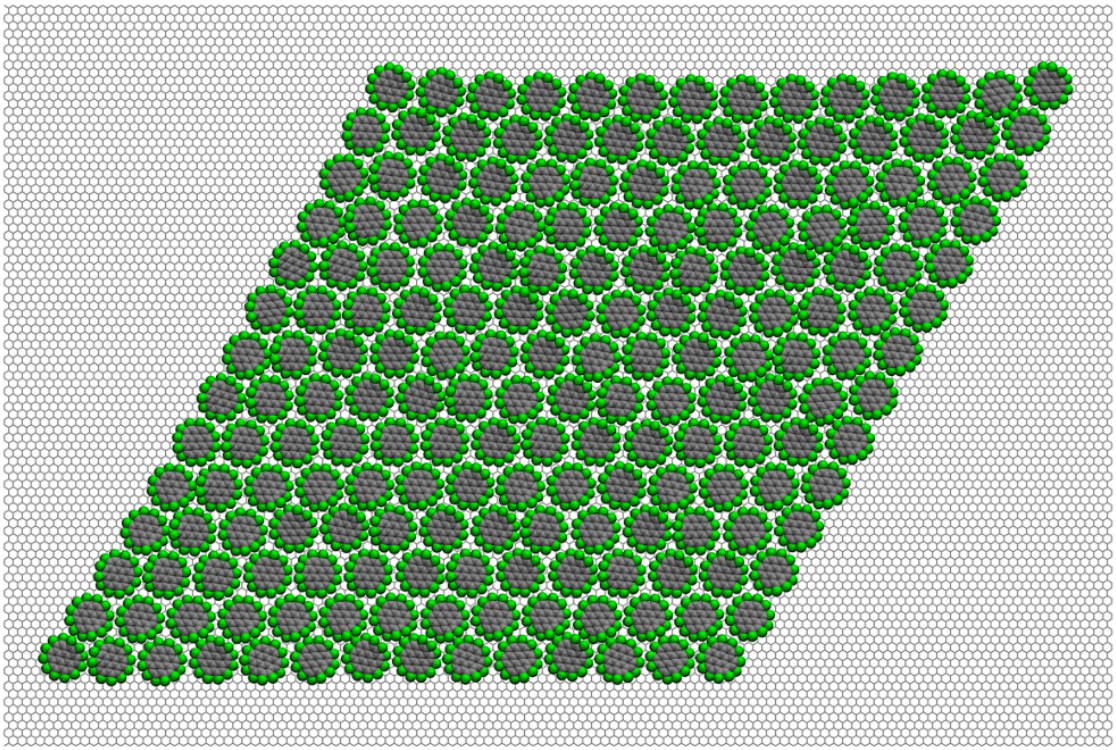




$\mu_{A,tr}$ =	-9.5	E'_{ads} =	-38.9	$\mu'_{vib,uc}$ =	176.1	A'_{uc} =	114.2
$\mu_{A,rot}$ =	-7.8	E'_{sam} =	-1.4	μ'_{uc} =	135.8	γ =	-25.0
$\mu_{A,vib}$ =	181.7	E'_{strain} =	0.0	$\Delta\mu'_{vib}$ =	-5.6	γ_E =	-33.5
$\mu_{A,tot}$ =	164.4	E'_{uc} =	-40.3	$\Delta\mu'$ =	-28.6	γ_S =	8.5
						C_{cac} =	$10^{-20.8}$

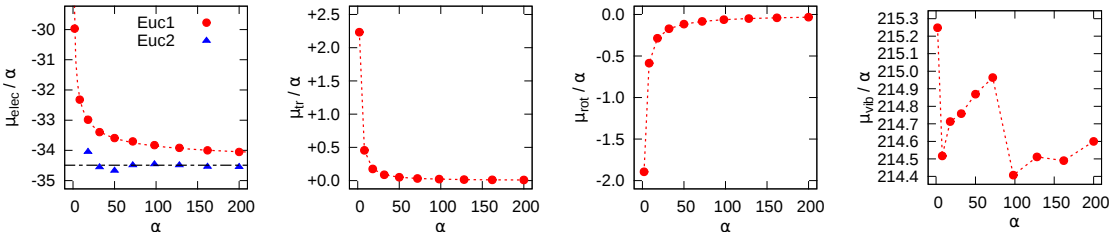
Perchlorocoronene

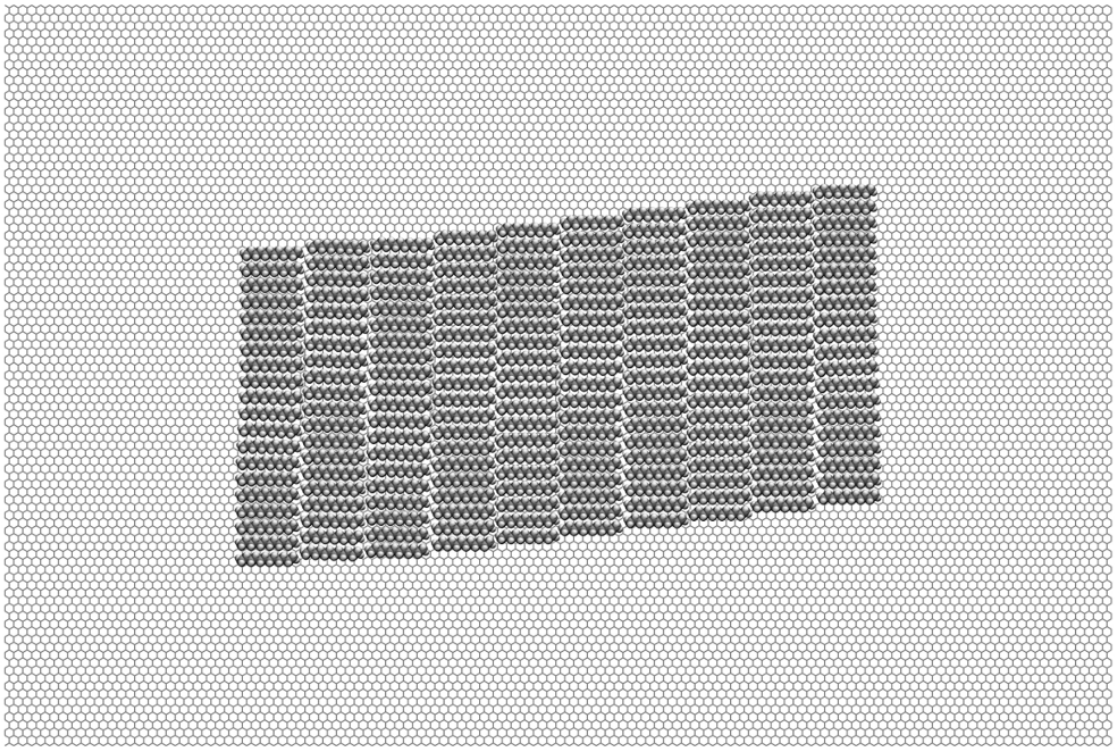




$\mu_{A,tr}$ =	-10.3	E'_{ads} =	-58.2	$\mu'_{vib,uc}$ =	94.4	A'_{uc} =	163.9
$\mu_{A,rot}$ =	-10.1	E'_{sam} =	-6.6	μ'_{uc} =	36.7	γ =	-25.8
$\mu_{A,vib}$ =	99.5	E'_{strain} =	7.1	$\Delta\mu'_{vib}$ =	-5.1	γ_E =	-33.7
$\mu_{A,tot}$ =	79.1	E'_{uc} =	-57.7	$\Delta\mu'$ =	-42.4	γ_S =	7.8
						C_{cac} =	$10^{-30.9}$

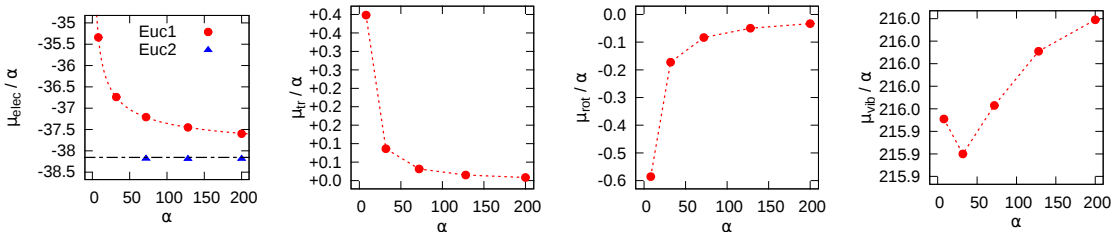
Dodecane

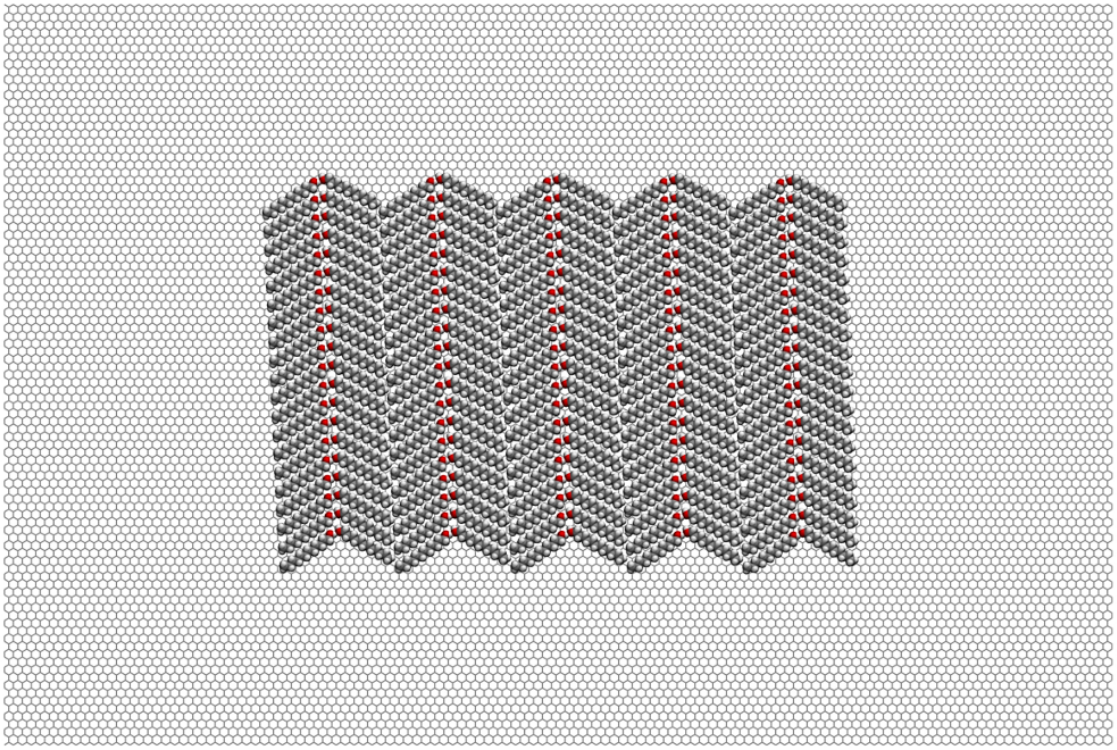




$\mu_{A,tr}$ =	-9.0	E'_{ads} =	-25.2	$\mu'_{vib,uc}$ =	214.6	A'_{uc} =	77.5
$\mu_{A,rot}$ =	-8.3	E'_{sam} =	-9.5	μ'_{uc} =	180.0	γ =	-22.1
$\mu_{A,vib}$ =	214.4	E'_{strain} =	0.1	$\Delta\mu'_{vib}$ =	0.2	γ_E =	-41.3
$\mu_{A,tot}$ =	197.1	E'_{uc} =	-34.6	$\Delta\mu'$ =	-17.1	γ_S =	19.2
						C_{cac} =	$10^{-12.4}$

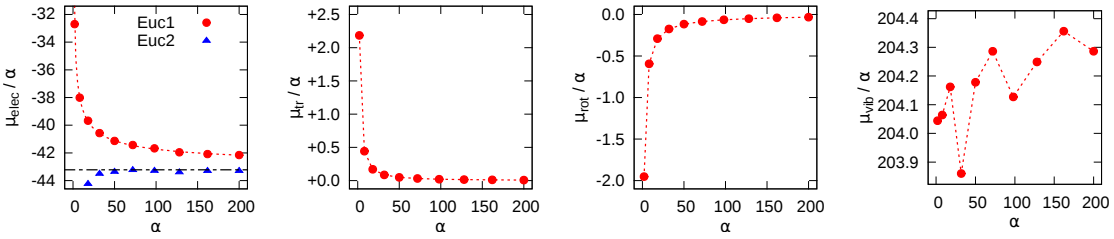
1-dodecanole

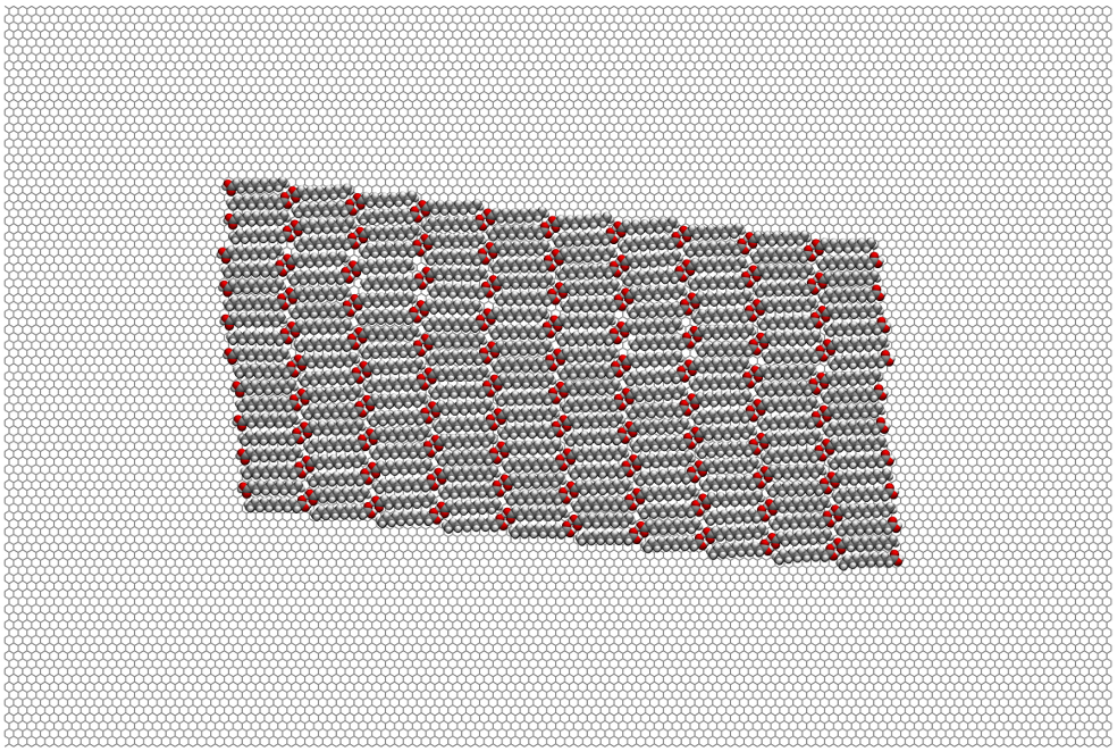




$\mu_{A,tr}$ =	-9.1	E'_{ads} =	-26.6	$\mu'_{vib,uc}$ =	216.0	A'_{uc} =	81.2
$\mu_{A,rot}$ =	-8.9	E'_{sam} =	-11.8	μ'_{uc} =	177.8	γ =	-23.9
$\mu_{A,vib}$ =	215.2	E'_{strain} =	0.2	$\Delta\mu'_{vib}$ =	0.8	γ_E =	-43.6
$\mu_{A,tot}$ =	197.2	E'_{uc} =	-38.2	$\Delta\mu'$ =	-19.4	γ_S =	19.7
						C_{cac} =	$10^{-14.1}$

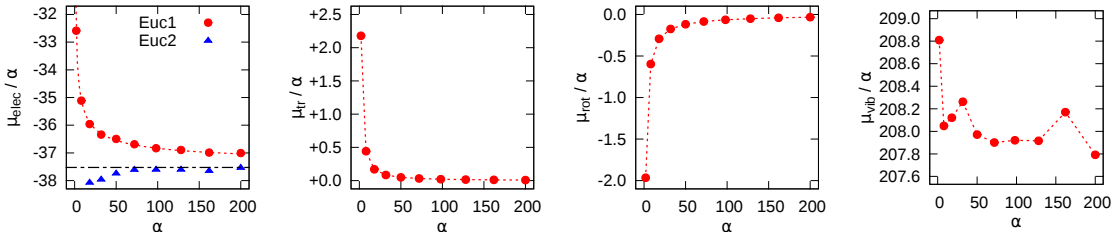
Dodecanoic Acid

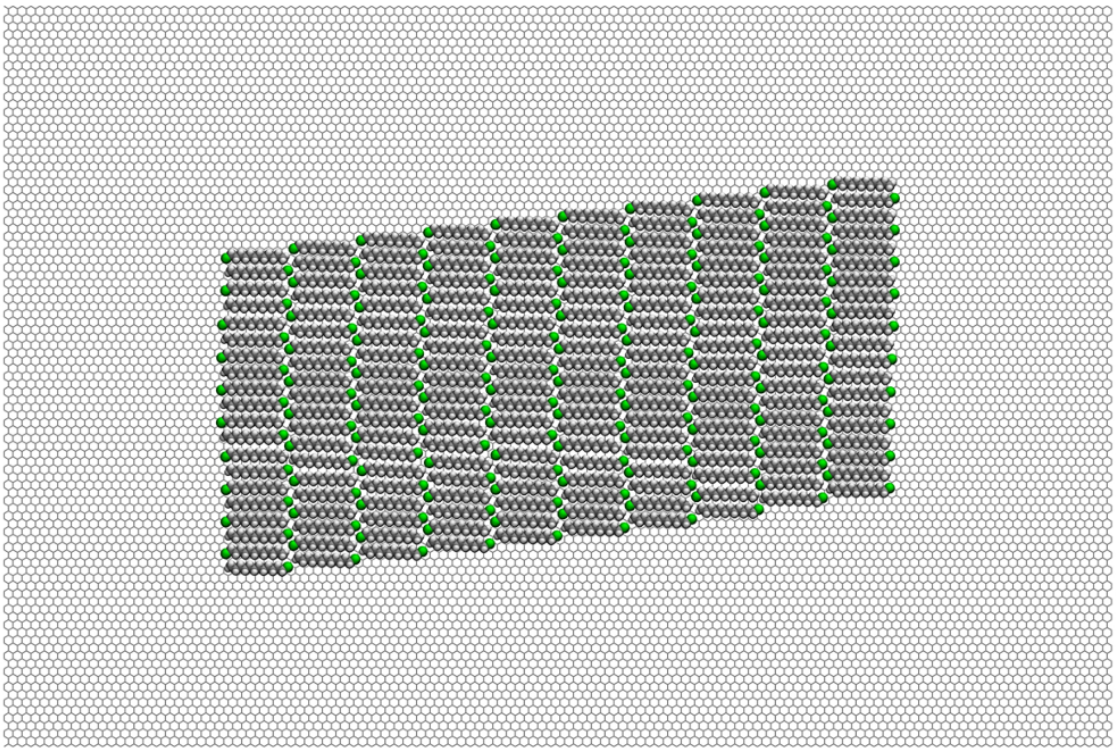




$\mu_{A,tr}$ =	-9.2	E'_{ads} =	-28.0	$\mu'_{vib,uc}$ =	204.3	A'_{uc} =	81.8
$\mu_{A,rot}$ =	-9.0	E'_{sam} =	-16.4	μ'_{uc} =	160.8	γ =	-29.1
$\mu_{A,vib}$ =	202.8	E'_{strain} =	0.9	$\Delta\mu'_{vib}$ =	1.5	γ_E =	-49.6
$\mu_{A,tot}$ =	184.6	E'_{uc} =	-43.5	$\Delta\mu'$ =	-23.8	γ_S =	20.5
						C_{cac} =	$10^{-17.3}$

1-chlorododecane





$\mu_{A,tr}$ =	-9.2	E'_{ads} =	-27.7	$\mu'_{vib,uc}$ =	207.8	A'_{uc} =	82.8
$\mu_{A,rot}$ =	-9.0	E'_{sam} =	-10.1	μ'_{uc} =	170.1	γ =	-23.4
$\mu_{A,vib}$ =	207.7	E'_{strain} =	0.1	$\Delta\mu'_{vib}$ =	0.1	γ_E =	-42.4
$\mu_{A,tot}$ =	189.5	E'_{uc} =	-37.7	$\Delta\mu'$ =	-19.4	γ_S =	19.0
						C_{cac} =	$10^{-14.1}$

E

ADAPTIVE TIME STEP

During confinement, an RMSD or absolute harmonic restrain is applied to atom coordinates to gradually confine the molecule in its reference state. The application of the restraining potential causes a gradually increase of the frequencies associated to the normal modes of the molecule. At low force constants, the confinement acts principally on the low frequency modes, increasing gradually their frequencies consuming their anharmonicity. For high force constants the harmonic restrain starts acting also on the high frequency modes, making the classical 1 fs time step of the simulation inadequate. This is the reason to introduce an adaptive time step approach, where the time step of the simulation is chosen as a function of the applied force constant in the harmonic restrain.

To derive a theoretically-justified expression for the time step, consider the Hamiltonian H of an harmonic oscillator

$$H_n(p,q) = \frac{1}{2}k_n q^2 + \frac{p^2}{2\mu} \tag{E.1}$$

where q is the coordinate associated to the movement along the vibrational mode, p is the associated momentum, μ the associated reduced mass, and k_n the natural resting force. The associated natural harmonic vibration is thus

$$\nu_n = \frac{1}{2\pi}\sqrt{\frac{k_n}{\mu}} \tag{E.2}$$

If an harmonic restrain of strength k is applied to the q coordinate, the Hamiltonian can be written as

$$\begin{aligned} H(p,q) &= \frac{1}{2}k_n q^2 + \frac{1}{2}kq^2 + \frac{p^2}{2\mu} \\ &= \frac{1}{2}(k_n + k)\, q^2 + \frac{p^2}{2\mu} \end{aligned} \tag{E.3}$$

for which the harmonic oscillator frequency is

$$\nu = \frac{1}{2\pi}\sqrt{\frac{k_n + k}{\mu}} \tag{E.4}$$

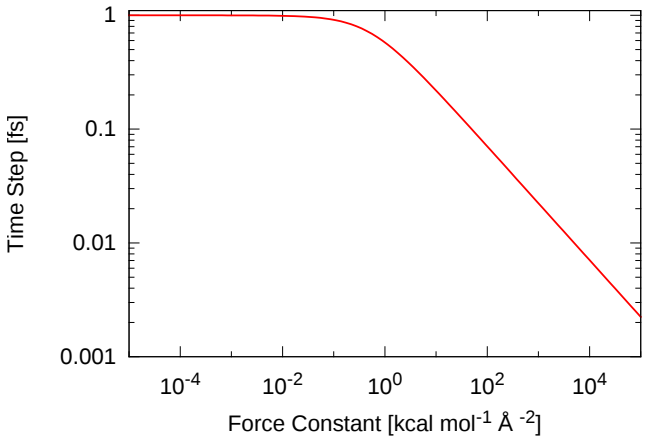


Figure E.1: Time step as a function of the force constant of the applied harmonic restrain.

If the reduced mass does not change with the applied constrain, which is generally not true for more complex systems, the harmonic frequency can be rewritten as

$$\nu = \sqrt{\nu_n^2 + \frac{1}{4\pi^2}\frac{k}{\mu}} \tag{E.5}$$

which correlates the frequency of an oscillator with its natural frequency and the force of the applied constrain.

If the verlet algorithm is considered for integrating the dynamics, the time step has to satisfy the relation

$$dt < (2\pi\nu_{max})^{-1} \tag{E.6}$$

where ν_{max} is the highest frequency in the system. Substituting the obtained expression for the frequency in the present of a harmonic restrain

$$dt < \left[(2\pi\nu_{n,max}) + \frac{k}{\mu}\right]^{-1/2} \tag{E.7}$$

which can be simplified as

$$dt < (1 + ak)^{-1/2} \tag{E.8}$$

which gives the value of the time step in femtoseconds if the force constant is expressed in kcal mol^{-1} Å^{-2}. The a parameter can be considered as empirical, which here has been given the value of 2. This expression says that for small force constants (less than 0.1 kcal mol^{-1} Å^{-2}) the time step equals 1 fs, as expected. At higher force constants instead, the time step decreases linearly in a double log plot, which means the time step is divided by 10 if the force constant is multiplied by 100; see Fig. E.1.

This continuous function correlate easily the time step of the simulation with the force constant of the applied constrain, making the confinement simulations at even the highest force constant very stable. The reduced effective time that can be simulated at such small time step is even not a problem, since there most of the anharmonicity is already confined, and the fluctuations are very small and easy to sample.

BIBLIOGRAPHY

[1] G. M. Whitesides and B. Grzybowski. "Self-assembly at all scales". *Science*, 295(5564), pp. 2418–2421, 2002. (see p. 1.)

[2] D. J. Kushner. "Self-assembly of biological structures." *Bacteriological reviews*, 33(2), p. 302, 1969. (see p. 1.)

[3] A. V. Pinheiro, D. Han, W. M. Shih, and H. Yan. "Challenges and opportunities for structural DNA nanotechnology". *Nature Nanotechnology*, 6(12), pp. 763–772, 2011. (see p. 1.)

[4] C.-A. Palma, M. Cecchini, and P. Samorì. "Predicting self-assembly: from empirism to determinism". *Chemical Society Reviews*, 41(10), pp. 3713–3730, 2012. (see pp. 1 and 2.)

[5] J. Halley and D. A. Winkler. "Consistent concepts of self-organization and self-assembly". *Complexity*, 14(2), pp. 10–17, 2008. (see p. 1.)

[6] Q.-F. Sun, J. Iwasa, D. Ogawa, Y. Ishido, S. Sato, T. Ozeki, Y. Sei, K. Yamaguchi, and M. Fujita. "Self-assembled $\{M\}_{24}\{L\}_{48}$ polyhedra and their sharp structural switch upon subtle ligand variation". *Science*, 328(5982), pp. 1144–1147, 2010. (see pp. 1 and 2.)

[7] D. G. Bucknall and H. L. Anderson. "Polymers get organized". *Science*, 302(5652), p. 1904, 2003. (see pp. 1 and 2.)

[8] T. F. A. De Greef, M. M. J. Smulders, M. Wolffs, A. P. H. J. Schenning, R. P. Sijbesma, and E. W. Meijer. "Supramolecular polymerization". *Chemical Reviews*, 109(11), pp. 5687–5754, 2009. (see p. 1.)

[9] A. Ciesielski, C.-A. Palma, M. Bonini, and P. Samorì. "Towards supramolecular engineering of functional nanomaterials: pre-programming multi-component 2D self-assembly at solid-liquid interfaces". *Advanced Materials*, 22(32), pp. 3506–3520, 2010. (see pp. 1, 2, and 91.)

[10] G. R. Desiraju. "Crystal gazing: structure prediction and polymorphism". *Science*, 278(5337), p. 404, 1997. (see pp. 1 and 2.)

[11] C. Meier, M. Roos, D. Künzel, A. Breitruck, H. E. Hoster, K. Landfester, A. Gross, R. J. Behm, and U. Ziener. "Concentration and coverage dependent adlayer structures: from two-dimensional networks to rotation in a bearing". *The Journal of Physical Chemistry C*, 114(2), pp. 1268–1277, 2010. (see pp. 2, 53, 72, and 85.)

[12] M. Lackinger, S. Griessl, T. Markert, F. Jamitzky, and W. M. Heckl. "Self-assembly of benzene-dicarboxylic acid isomers at the liquid solid interface: Steric aspects of hydrogen bonding". *The Journal of Physical Chemistry B*, 108(36), pp. 13652–13655, 2004. (see pp. 2 and 75.)

[13] G. Binnig and C. Rohr. "Scanning tunneling microscopy". *IBM Journal of Research and Development*, 30(4), 1986. (see p. 1.)

[14] G. Binnig, H. Rohrer, C. Gerber, and E. Weibel. "Surface studies by scanning tunneling microscopy". *Physical Review Letters*, 49(1), p. 57, 1982. (see p. 1.)

[15] J. Bardeen. "Tunnelling from a many-particle point of view". *Physical Review Letters*, 6(2), p. 57, 1961. (see p. 1.)

[16] J. Tersoff and D. R. Hamann. "Theory and application for the scanning tunneling microscope". *Physical Review Letters*, 50(25), p. 1998, 1983.

[17] J. Tersoff and D. R. Hamann. "Theory of the scanning tunneling microscope". *Physical Review B*, 31(2), pp. 59–67, 1985. (see p. 1.)

[18] S. De Feyter and F. C. De Schryver. "Two-dimensional supramolecular self-assembly probed by scanning tunneling microscopy". *Chemical Society Reviews*, 32(3), pp. 139–150, 2003. (see pp. 2 and 113.)

[19] F. Rosei, M. Schunack, Y. Naitoh, P. Jiang, A. Gourdon, E. Laegsgaard, I. Stensgaard, C. Joachim, and F. Besenbacher. "Properties of large organic molecules on metal surfaces". *Progress in Surface Science*, 71(5–8), pp. 95–146, 2003.

[20] J. V. Barth, G. Costantini, and K. Kern. "Engineering atomic and molecular nanostructures at surfaces". *Nature*, 437, pp. 671–679, 2005.

[21] L. Grill. "Functionalized molecules studied by STM: motion, switching and reactivity". *Journal of Physics: Condensed Matter*, 20(5), p. 053001, 2008.

[22] K. Müllen and J. P. Rabe. "Nanographenes as active components of single-molecule electronics and how a scanning tunneling microscope puts them to work". *Accounts of Chemical Research*, 41(4), pp. 511–520, 2008. (see p. 2.)

[23] J. M. MacLeod and F. Rosei. "Molecular SelfAssembly on Graphene". *Small*, 10(6), pp. 1038–1049, 2014. (see p. 2.)

[24] J. Puigmartí-Luis, A. Minoia, H. Uji-i, C. Rovira, J. Cornil, S. De Feyter, R. Lazzaroni, and D. B. Amabilino. "Non-covalent control for bottom-up assembly of functional supramolecular wires". *Journal of the American Chemical Society*, 128(39), pp. 12602–12603, 2006. (see p. 2.)

[25] R. M. Westervelt. "Graphene nanoelectronics". *Science*, 320(5874), pp. 324–325, 2008. (see p. 2.)

[26] Q. H. Wang and M. C. Hersam. "Room-temperature molecular-resolution characterization of self-assembled organic monolayers on epitaxial graphene". *Nature Chemistry*, 1(3), pp. 206–211, 2009. (see p. 2.)

[27] T. Zhang, Z. Cheng, Y. Wang, Z. Li, C. Wang, Y. Li, and Y. Fang. "Self-assembled 1-octadecanethiol monolayers on graphene for mercury detection". *Nano Letters*, 10(11), pp. 4738–4741, 2010. (see p. 2.)

[28] G. M. Scheuermann, L. Rumi, P. Steurer, W. Bannwarth, and R. Mülhaupt. "Palladium nanoparticles on graphite oxide and its functionalized graphene derivatives as highly active catalysts for the Suzuki-Miyaura coupling reaction". *Journal of the American Chemical Society*, 131(23), pp. 8262–8270, 2009. (see pp. 2 and 91.)

[29] M. Pumera. "Graphene-based nanomaterials for energy storage". *Energy & Environmental Science*, 4(3), pp. 668–674, 2011. (see p. 2.)

[30] P. V. Kamat. "Graphene-based nanoassemblies for energy conversion". *The Journal of Physical Chemistry Letters*, 2(3), pp. 242–251, 2011. (see p. 2.)

[31] V. Ferri, M. Elbing, G. Pace, M. D. Dickey, M. Zharnikov, P. Samorì, M. Mayor, and M. A. Rampi. "Light-Powered electrical switch based on cargo-lifting azobenzene monolayers". *Angewandte Chemie*, 120(18), pp. 3455–3457, 2008. (see p. 2.)

[32] M. O. Blunt, J. Adisoejoso, K. Tahara, K. Katayama, M. Van der Auweraer, Y. Tobe, and S. De Feyter. "Temperature-induced structural phase transitions in a two-dimensional self-assembled network". *Journal of the American Chemical Society*, 135(32), pp. 12068–12075, 2013. (see pp. 2, 56, 57, 58, and 86.)

[33] M. Lackinger, S. Griessl, W. M. Heckl, M. Hietschold, and G. W. Flynn. "Self-assembly of trimesic acid at the liquid-solid interface: a study of solvent-induced polymorphism". *Langmuir*, 21(11), pp. 4984–4988, 2005. (see pp. 2, 73, 75, 87, 113, and 180.)

[34] A. Ciesielski, P. J. Szabelski, W. Rżysko, A. Cadeddu, T. R. Cook, P. J. Stang, and P. Samorì. "Concentration-dependent supramolecular engineering of hydrogen-bonded nanostructures at surfaces: predicting self-assembly in 2D". *Journal of the American Chemical Society*, 135(18), pp. 6942–6950, 2013. (see p. 2.)

[35] L. Kampschulte, T. L. Werblowsky, R. S. Kishore, M. Schmittel, W. M. Heckl, and M. Lackinger. "Thermodynamical equilibrium of binary supramolecular networks at the liquid- solid Interface". *Journal of the American Chemical Society*, 130(26), pp. 8502–8507, 2008. (see pp. 2, 50, and 82.)

[36] A. Ciesielski, M. El Garah, S. Haar, P. Kovaříček, J.-M. Lehn, and P. Samorì. "Dynamic covalent chemistry of bisimines at the solid/liquid interface monitored by scanning tunnelling microscopy". *Nature Chemistry*, 6(11), pp. 1017–1023, 2014. (see pp. 2, 49, and 82.)

[37] A. K. Geim. "Graphene: status and prospects". *Science*, 324(5934), pp. 1530–1534, 2009. (see p. 3.)

[38] K. S. Novoselov, V. I. Fal'ko, L. Colombo, P. R. Gellert, M. G. Schwab, and K. Kim. "A roadmap for graphene". *Nature*, 490(7419), pp. 192–200, 2012. (see p. 3.)

[39] K. S. Kim, Y. Zhao, H. Jang, S. Y. Lee, J. M. Kim, K. S. Kim, J.-H. Ahn, P. Kim, J.-Y. Choi, and B. H. Hong. "Large-scale pattern growth of graphene films for stretchable transparent electrodes". *Nature*, 457(7230), pp. 706–710, 2009. (see pp. 3 and 111.)

[40] X. Li, W. Cai, J. An, S. Kim, J. Nah, D. Yang, R. Piner, A. Velamakanni, I. Jung, E. Tutuc, S. K. Banerjee, L. Colombo, and R. S. Ruoff. "Large-area synthesis of high-quality and uniform graphene films on copper foils". *Science*, 324(5932), pp. 1312–1314, 2009.

[41] C. Berger, Z. Song, X. Li, X. Wu, N. Brown, C. Naud, D. Mayou, T. Li, J. Hass, A. N. Marchenkov, and others. "Electronic confinement and coherence in patterned epitaxial graphene". *Science*, 312(5777), pp. 1191–1196, 2006.

[42] L. Chen, Y. Hernandez, X. Feng, and K. Müllen. "From nanographene and graphene nanoribbons to graphene sheets: chemical synthesis". *Angewandte Chemie International Edition*, 51(31), pp. 7640–7654, 2012.

[43] A. Narita, X. Feng, Y. Hernandez, S. A. Jensen, M. Bonn, H. Yang, I. A. Verzhbitskiy, C. Casiraghi, M. R. Hansen, A. H. R. Koch, G. Fytas, O. Ivasenko, B. Li, K. S. Mali, T. Balandina, S. Mahesh, S. De Feyter, and K. Mullen. "Synthesis of structurally well-defined and liquid-phase-processable graphene nanoribbons". *Nature Chemistry*, 6(2), pp. 126–132, 2014. (see p. 3.)

[44] K. Novoselov, D. Jiang, F. Schedin, T. Booth, V. Khotkevich, S. Morozov, and A. Geim. "Two-dimensional atomic crystals". *Proceedings of the National Academy of Sciences*, 102(30), pp. 10451–10453, 2005. (see p. 3.)

[45] V. León, M. Quintana, M. A. Herrero, J. L. Fierro, A. de la Hoz, M. Prato, and E. Vazquez. "Few-layer graphenes from ball-milling of graphite with melamine". *Chemical Communications*, 47(39), pp. 10936–10938, 2011. (see p. 3.)

[46] Z. Y. Xia, S. Pezzini, E. Treossi, G. Giambastiani, F. Corticelli, V. Morandi, A. Zanelli, V. Bellani, and V. Palermo. "The exfoliation of graphene in liquids by electrochemical, chemical, and sonication-assisted techniques: a nanoscale study". *Advanced Functional Materials*, 23(37), pp. 4684–4693, 2013.

[47] K. Parvez, Z.-S. Wu, R. Li, X. Liu, R. Graf, X. Feng, and K. Müllen. "Exfoliation of graphite into graphene in aqueous solutions of inorganic salts". *Journal of the American Chemical Society*, 136(16), pp. 6083–6091, 2014. (see p. 111.)

[48] G. Cravotto and P. Cintas. "Sonication-assisted fabrication and post-synthetic modifications of graphene-like materials". *Chemistry - A European Journal*, 16(18), pp. 5246–5259, 2010. (see p. 111.)

[49] X. Cui, C. Zhang, R. Hao, and Y. Hou. "Liquid-phase exfoliation, functionalization and applications of graphene". *Nanoscale*, 3(5), pp. 2118–2126, 2011.

[50] F. Torrisi, T. Hasan, W. Wu, Z. Sun, A. Lombardo, T. S. Kulmala, G.-W. Hsieh, S. Jung, F. Bonaccorso, P. J. Paul, and others. "Inkjet-printed graphene electronics". *ACS Nano*, 6(4), pp. 2992–3006, 2012.

[51] J. N. Coleman. "Liquid exfoliation of defect-free graphene". *Accounts of Chemical Research*, 46(1), pp. 14–22, 2012.

[52] A. Ciesielski and P. Samorì. "Graphene via sonication assisted liquid-phase exfoliation". *Chemical Society Reviews*, 43(1), pp. 381–398, 2014. (see pp. 3, 91, and 111.)

[53] Y. Hernandez, V. Nicolosi, M. Lotya, F. M. Blighe, Z. Sun, S. De, I. T. McGovern, B. Holland, M. Byrne, Y. K. Gun'Ko, J. J. Boland, P. Nitraj, G. Duesberg, S. Krishnamurthy, R. Goodhue, J. Hutchison, V. Scardaci, A. C. Ferrari, and J. N. Coleman. "High-yield production of graphene by liquid-phase exfoliation of graphite". *Nature Nanotechnology*, 3(9), pp. 563–568, 2008. (see pp. 3, 111, and 145.)

[54] C. E. Hamilton, J. R. Lomeda, Z. Sun, J. M. Tour, and A. R. Barron. "High-yield organic dispersions of unfunctionalized graphene". *Nano Letters*, 9(10), pp. 3460–3462, 2009. (see p. 111.)

[55] X. Li, G. Zhang, X. Bai, X. Sun, X. Wang, E. Wang, and H. Dai. "Highly conducting graphene sheets and Langmuir–Blodgett films". *Nature Nanotechnology*, 3(9), pp. 538–542, 2008. (see p. 111.)

[56] A. B. Bourlinos, V. Georgakilas, R. Zboril, T. A. Steriotis, and A. K. Stubos. "Liquid-phase exfoliation of graphite towards solubilized graphenes". *Small*, 5(16), pp. 1841–1845, 2009. (see pp. 3 and 111.)

[57] A. A. Green and M. C. Hersam. "Solution phase production of graphene with controlled thickness via density differentiation". *Nano Letters*, 9(12), pp. 4031–4036, 2009. (see p. 3.)

[58] M. Lotya, Y. Hernandez, P. J. King, R. J. Smith, V. Nicolosi, L. S. Karlsson, F. M. Blighe, S. De, Z. Wang, I. T. McGovern, G. S. Duesberg, and J. N. Coleman. "Liquid phase production of graphene by exfoliation of graphite in surfactant/water solutions". *Journal of the American Chemical Society*, 131(10), pp. 3611–3620, 2009.

[59] X. An, T. Simmons, R. Shah, C. Wolfe, K. M. Lewis, M. Washington, S. K. Nayak, S. Talapatra, and S. Kar. "Stable aqueous dispersions of noncovalently functionalized graphene from graphite and their multifunctional high-performance applications". *Nano Letters*, 10(11), pp. 4295–4301, 2010.

[60] A. Schlierf, H. Yang, E. Gebremedhn, E. Treossi, L. Ortolani, L. Chen, A. Minoia, V. Morandi, P. Samorì, C. Casiraghi, D. Beljonne, and V. Palermo. "Nanoscale insight into the exfoliation mechanism of graphene with organic dyes: effect of charge, dipole and molecular structure". *Nanoscale*, 5(10), pp. 4205–4216, 2013.

[61] H. Yang, Y. Hernandez, A. Schlierf, A. Felten, A. Eckmann, S. Johal, P. Louette, J.-J. Pireaux, X. Feng, K. Mullen, V. Palermo, and C. Casiraghi. "A simple method for graphene production based on exfoliation of graphite in water using 1-pyrenesulfonic acid sodium salt". *Carbon*, 53, pp. 357–365, 2013.

[62] S. Sampath, A. N. Basuray, K. J. Hartlieb, T. Aytun, S. I. Stupp, and J. F. Stoddart. "Direct exfoliation of graphite to graphene in aqueous media with diazaperopyrenium dications". *Advanced Materials*, 25(19), pp. 2740–2745, 2013.

[63] N. Behabtu, J. R. Lomeda, M. J. Green, A. L. Higginbotham, A. Sinitskii, D. V. Kosynkin, D. Tsentalovich, A. N. G. Parra-Vasquez, J. Schmidt, E. Kesselman, Y. Cohen, Y. Talmon, J. M. Tour, and M. Pasquali. "Spontaneous high-concentration dispersions and liquid crystals of graphene". *Nature Nanotechnology*, 5(6), pp. 406–411, 2010.

[64] A. M. Dimiev, G. Ceriotti, N. Behabtu, D. Zakhidov, M. Pasquali, R. Saito, and J. M. Tour. "Direct real-time monitoring of stage transitions in graphite intercalation compounds". *ACS Nano*, 7(3), pp. 2773–2780, 2013.

[65] A. B. Bourlinos, V. Georgakilas, R. Zboril, T. A. Steriotis, A. K. Stubos, and C. Trapalis. "Aqueous-phase exfoliation of graphite in the presence of polyvinylpyrrolidone for the production of water-soluble graphenes". *Solid State Communications*, 149(47), pp. 2172–2176, 2009.

[66] F. Liu, J. Y. Choi, and T. S. Seo. "DNA mediated water-dispersible graphene fabrication and gold nanoparticle-graphene hybrid". *Chemical Communications*, 46(16), pp. 2844–2846, 2010. (see p. 3.)

[67] A. Ciesielski, S. Haar, M. El Gemayel, H. Yang, J. Clough, G. Melinte, M. Gobbi, E. Orgiu, M. V. Nardi, G. Ligorio, V. Palermo, N. Koch, O. Ersen, C. Casiraghi, and P. Samorì. "Harnessing the liquid-phase exfoliation of graphene using aliphatic compounds: a supramolecular approach". *Angewandte Chemie*, 126(39), pp. 10523–10529, 2014. (see pp. 3, 107, 111, and 113.)

[68] F. Jensen. *Introduction to computational chemistry*. John Wiley & Sons, 2013. (see p. 8.)

[69] A. R. Leach. *Molecular modelling: principles and applications*. Pearson Education, 2001.

[70] A. Szabo and N. S. Ostlund. *Modern quantum chemistry: introduction to advanced electronic structure theory*. Dover Publications, Inc., 1996. (see p. 8.)

[71] C. C. J. Roothaan. "New developments in molecular orbital theory". *Reviews of Modern Physics*, 23(2), pp. 69–89, 1951. (see p. 9.)

[72] C. D. Sherrill and H. F. Schaefer III. "The configuration interaction method: advances in highly correlated approaches". In M. C. Z. Per-Olov Löwdin and J. R. S. Erkki Brändas, editors, "Advances in Quantum Chemistry", vol. 34, pp. 143–269. Academic Press, 1999. (see p. 10.)

[73] C. Møller and M. S. Plesset. "Note on an approximation treatment for many-electron systems". *Physical Review*, 46(7), pp. 618–622, 1934. (see p. 10.)

[74] G. D. Purvis III and R. J. Bartlett. "A full coupled-cluster singles and doubles model: the inclusion of disconnected triples". *The Journal of Chemical Physics*, 76(4), pp. 1910–1918, 1982. (see p. 11.)

[75] K. E. Riley, M. Pitoňák, P. Jurečka, and P. Hobza. "Stabilization and structure calculations for noncovalent interactions in extended molecular systems based on wave function and density functional theories". *Chemical Reviews*, 110(9), pp. 5023–5063, 2010. (see p. 11.)

[76] P. Hohenberg and W. Kohn. "Inhomogeneous electron gas". *Physical Review*, 136(3B), p. B864, 1964. (see p. 11.)

[77] W. Kohn and L. J. Sham. "Self-consistent equations including exchange and correlation effects". *Physical Review*, 140(4A), p. A1133, 1965. (see p. 12.)

[78] R. G. Parr. "Density functional theory of atoms and molecules". In K. Fukui and B. Pullman, editors, "Horizons of Quantum Chemistry", No. 3 in Académie Internationale Des Sciences Moléculaires Quantiques / International Academy of Quantum Molecular Science, pp. 5–15. Springer Netherlands, 1980. (see p. 12.)

[79] D. C. Langreth and M. J. Mehl. "Beyond the local-density approximation in calculations of ground-state electronic properties". *Physical Review B*, 28(4), pp. 1809–1834, 1983. (see p. 12.)

[80] A. D. Becke. "A new mixing of Hartree-Fock and local densityfunctional theories". *The Journal of Chemical Physics*, 98(2), pp. 1372–1377, 1993. (see p. 13.)

[81] S. Grimme. "Density functional theory with London dispersion corrections". *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 1(2), pp. 211–228, 2011. (see pp. 13 and 94.)

[82] M. Dion, H. Rydberg, E. Schröder, D. C. Langreth, and B. I. Lundqvist. "Van der Waals density functional for general geometries". *Physical Review Letters*, 92(24), p. 246401, 2004. (see p. 13.)

[83] K. Lee, É. D. Murray, L. Kong, B. I. Lundqvist, and D. C. Langreth. "Higher-accuracy van der Waals density functional". *Physical Review B*, 82(8), p. 081101, 2010. (see p. 13.)

[84] O. A. Vydrov and T. Van Voorhis. "Nonlocal van der Waals density functional made simple". *Physical Review Letters*, 103(6), p. 063004, 2009. (see p. 13.)

[85] O. A. Vydrov and T. Van Voorhis. "Nonlocal van der Waals density functional: The simpler the better". *The Journal of Chemical Physics*, 133(24), p. 244103, 2010. (see p. 13.)

[86] P. Jurečka, J. Černý, P. Hobza, and D. R. Salahub. "Density functional theory augmented with an empirical dispersion term. Interaction energies and geometries of 80 noncovalent complexes compared with ab initio quantum mechanics calculations". *Journal of Computational Chemistry*, 28(2), pp. 555–569, 2007. (see p. 14.)

[87] J.-D. Chai and M. Head-Gordon. "Long-range corrected hybrid density functionals with damped atom–atom dispersion corrections". *Physical Chemistry Chemical Physics*, 10(44), pp. 6615–6620, 2008. (see pp. 14, 94, and 139.)

[88] S. Grimme. "Accurate description of van der Waals complexes by density functional theory including empirical corrections". *Journal of Computational Chemistry*, 25(12), pp. 1463–1473, 2004. (see p. 14.)

[89] S. Grimme. "Semiempirical GGA-type density functional constructed with a long-range dispersion correction". *Journal of Computational Chemistry*, 27(15), pp. 1787–1799, 2006. (see p. 14.)

[90] S. Grimme, J. Antony, S. Ehrlich, and H. Krieg. "A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H – Pu". *The Journal of Chemical Physics*, 132(15), p. 154104, 2010. (see pp. 14, 16, 91, and 95.)

[91] O. A. Vydrov and T. Van Voorhis. "Benchmark assessment of the accuracy of several van der Waals density functionals". *The Journal of Chemical Theory and Computation*, 8(6), pp. 1929–1934, 2012. (see p. 14.)

[92] W. Hujo and S. Grimme. "Performance of the van der Waals density functional VV10 and (hybrid) GGA variants for thermochemistry and noncovalent interactions". *The Journal of Chemical Theory and Computation*, 7(12), pp. 3866–3871, 2011. (see p. 14.)

[93] S. F. Boys and F. Bernardi. "The calculation of small molecular interactions by the differences of separate total energies. Some procedures with reduced errors". *Molecular Physics*, 19(4), pp. 553–566, 1970. (see pp. 14 and 95.)

[94] W. Thiel. "Semiempirical quantum chemical methods". *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 4(2), pp. 145–157, 2014. (see p. 15.)

[95] J. A. Pople, D. P. Santry, and G. A. Segal. "Approximate self-consistent molecular orbital theory. I. Invariant procedures". *The Journal of Chemical Physics*, 43(10), pp. S129–S135, 1965. (see p. 15.)

[96] J. A. Pople, D. L. Beveridge, and P. A. Dobosh. "Approximate self-consistent molecular-orbital theory. V. Intermediate neglect of differential overlap". *The Journal of Chemical Physics*, 47(6), pp. 2026–2033, 1967. (see p. 15.)

[97] N. C. Baird and M. J. S. Dewar. "Ground states of σ-bonded molecules. IV. The MINDO method and its application to hydrocarbons". *The Journal of Chemical Physics*, 50(3), pp. 1262–1274, 1969. (see p. 15.)

[98] M. J. S. Dewar and W. Thiel. "Ground states of molecules. 38. The MNDO method. Approximations and parameters". *Journal of the American Chemical Society*, 99(15), pp. 4899–4907, 1977. (see p. 15.)

[99] M. J. S. Dewar, E. G. Zoebisch, E. F. Healy, and J. J. P. Stewart. "Development and use of quantum mechanical molecular models. 76. AM1: A new general purpose quantum mechanical molecular model". *Journal of the American Chemical Society*, 107(13), pp. 3902–3909, 1985. (see p. 15.)

[100] J. J. P. Stewart. "Optimization of parameters for semiempirical methods I. Method". *Journal of Computational Chemistry*, 10(2), pp. 209–220, 1989. (see p. 15.)

[101] J. J. P. Stewart. "Optimization of parameters for semiempirical methods II. Applications". *Journal of Computational Chemistry*, 10(2), pp. 221–264, 1989. (see p. 15.)

[102] G. B. Rocha, R. O. Freire, A. M. Simas, and J. J. P. Stewart. "RM1: A reparameterization of AM1 for H, C, N, O, P, S, F, Cl, Br, and I". *Journal of Computational Chemistry*, 27(10), pp. 1101–1111, 2006. (see p. 15.)

[103] J. J. P. Stewart. "Optimization of parameters for semiempirical methods V: Modification of NDDO approximations and application to 70 elements". *Journal of Molecular Modeling*, 13(12), pp. 1173–1213, 2007. (see p. 15.)

[104] J. P. McNamara and I. H. Hillier. "Semi-empirical molecular orbital methods including dispersion corrections for the accurate prediction of the full range of intermolecular interactions in biomolecules". *Physical Chemistry Chemical Physics*, 9(19), pp. 2362–2370, 2007. (see p. 16.)

[105] J. Řezáč, J. Fanfrĺik, D. Salahub, and P. Hobza. "Semiempirical quantum chemical PM6 method augmented by dispersion and H–bonding correction terms reliably describes various types of noncovalent complexes". *Journal of Chemical Theory and Computation*, 5(7), pp. 1749–1760, 2009. (see p. 16.)

[106] M. E. Foster and K. Sohlberg. "A new empirical correction to the AM1 method for macromolecular complexes". *Journal of Chemical Theory and Computation*, 6(7), pp. 2153–2166, 2010. (see p. 16.)

[107] M. Korth, M. Pitonak, J. Řezáč, and P. Hobza. "A transferable H-bonding correction for semiempirical quantum-chemical methods". *Journal of Chemical Theory and Computation*, 6(1), pp. 344–352, 2009. (see p. 16.)

[108] M. Korth. "Third-generation hydrogen-bonding corrections for semiempirical QM methods and force fields". *Journal of Chemical Theory and Computation*, 6(12), pp. 3808–3816, 2010. (see pp. 16, 95, and 137.)

[109] M. Korth. "Empirical hydrogen-bond potential functions: an old hat reconditioned". *ChemPhysChem*, 12(17), pp. 3131–3142, 2011. (see p. 16.)

[110] J. Řezáč and P. Hobza. "Advanced corrections of hydrogen bonding and dispersion for semiempirical quantum mechanical methods". *Journal of Chemical Theory and Computation*, 8(1), pp. 141–151, 2011. (see pp. 16 and 95.)

[111] J. C. Kromann, A. S. Christensen, C. Steinmann, M. Korth, and J. H. Jensen. "A third-generation dispersion and third-generation hydrogen bonding corrected PM6 method: PM6-D3H+". *PeerJ*, 2, p. e449, 2014. (see p. 16.)

[112] J. J. P. Stewart. "Optimization of parameters for semiempirical methods VI: More modifications to the NDDO approximations and re-optimization of parameters". *Journal of Molecular Modeling*, 19(1), pp. 1–32, 2013. (see p. 16.)

[113] W. L. Jorgensen and J. Tirado-Rives. "The OPLS potential functions for proteins: energy minimizations for crystals of cyclic peptides and crambin". *Journal of the American Chemical Society*, 110(6), pp. 1657–1666, 1988. (see p. 18.)

[114] J. W. Ponder and D. A. Case. "Force fields for protein simulations". *Advances in Protein Chemistry*, 66, pp. 27–86, 2003. (see p. 18.)

[115] A. D. MacKerell, M. Feig, and C. L. Brooks. "Extending the treatment of backbone energetics in protein force fields: Limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations". *Journal of Computational Chemistry*, 25(11), pp. 1400–1415, 2004. (see p. 18.)

[116] W. L. Jorgensen, D. S. Maxwell, and J. Tirado-Rives. "Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids". *Journal of the American Chemical Society*, 118(45), pp. 11225–11236, 1996. (see p. 18.)

[117] W. L. Jorgensen and N. A. McDonald. "Development of an all-atom force field for heterocycles. Properties of liquid pyridine and diazenes". *Journal of Molecular Structure: THEOCHEM*, 424(1), pp. 145–155, 1998.

[118] N. A. McDonald and W. L. Jorgensen. "Development of an all-atom force field for heterocycles. Properties of liquid pyrrole, furan, diazoles, and oxazoles". *The Journal of Physical Chemistry B*, 102(41), pp. 8049–8059, 1998.

[119] R. C. Rizzo and W. L. Jorgensen. "OPLS all-atom model for amines: resolution of the amine hydration problem". *Journal of the American Chemical Society*, 121(20), pp. 4827–4836, 1999.

[120] M. L. P. Price, D. Ostrovsky, and W. L. Jorgensen. "Gas-phase and liquid-state properties of esters, nitriles, and nitro compounds with the OPLS-AA force field". *Journal of Computational Chemistry*, 22(13), pp. 1340–1352, 2001.

[121] E. K. Watkins and W. L. Jorgensen. "Perfluoroalkanes: Conformational analysis and liquid-state properties from ab initio and Monte Carlo calculations". *The Journal of Physical Chemistry A*, 105(16), pp. 4118–4125, 2001.

[122] G. A. Kaminski, R. A. Friesner, J. Tirado-Rives, and W. L. Jorgensen. "Evaluation and reparametrization of the OPLS-AA force field for proteins via comparison with accurate quantum chemical calculations on peptides". *The Journal of Physical Chemistry B*, 105(28), pp. 6474–6487, 2001. (see p. 18.)

[123] J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman, and D. A. Case. "Development and testing of a general amber force field". *Journal of Computational Chemistry*, 25(9), pp. 1157–1174, 2004. (see pp. 18, 43, 72, 128, 157, and 178.)

[124] J. Wang, W. Wang, P. A. Kollman, and D. A. Case. "Automatic atom type and bond type perception in molecular mechanical calculations". *Journal of Molecular Graphics and Modelling*, 25(2), pp. 247–260, 2006. (see pp. 18, 43, 96, 157, and 178.)

[125] K. Vanommeslaeghe, E. Hatcher, C. Acharya, S. Kundu, S. Zhong, J. Shim, E. Darian, O. Guvench, P. Lopes, I. Vorobyov, and A. D. Mackerell Jr. "CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields". *Journal of Computational Chemistry*, 31(4), pp. 671–690, 2010. (see pp. 18, 118, and 137.)

[126] N. L. Allinger. "Conformational analysis. 130. MM2. A hydrocarbon force field utilizing V1 and V2 torsional terms". *Journal of the American Chemical Society*, 99(25), pp. 8127–8134, 1977. (see p. 18.)

[127] N. L. Allinger, Y. H. Yuh, and J. H. Lii. "Molecular mechanics. The MM3 force field for hydrocarbons. 1". *Journal of the American Chemical Society*, 111(23), pp. 8551–8566, 1989. (see p. 18.)

[128] N. L. Allinger, K.-H. Chen, J.-H. Lii, and K. A. Durkin. "Alcohols, ethers, carbohydrates, and related compounds. I. The MM4 force field for simple compounds". *Journal of Computational Chemistry*, 24(12), pp. 1447–1472, 2003. (see p. 18.)

[129] T. A. Halgren. "Merck molecular force field. I. Basis, form, scope, parameterization, and performance of MMFF94". *Journal of Computational Chemistry*, 17(5-6), pp. 490–519, 1996. (see p. 18.)

[130] T. A. Halgren. "Merck molecular force field. II. MMFF94 van der Waals and electrostatic parameters for intermolecular interactions". *Journal of Computational Chemistry*, 17(5-6), pp. 520–552, 1996.

[131] T. A. Halgren. "Merck molecular force field. III. Molecular geometries and vibrational frequencies for MMFF94". *Journal of Computational Chemistry*, 17(5-6), pp. 553–586, 1996.

[132] T. A. Halgren and R. B. Nachbar. "Merck molecular force field. IV. Conformational energies and geometries for MMFF94". *Journal of Computational Chemistry*, 17(5-6), pp. 587–615, 1996.

[133] T. A. Halgren. "Merck molecular force field. V. Extension of MMFF94 using experimental data, additional computational data, and empirical rules". *Journal of Computational Chemistry*, 17(5-6), pp. 616–641, 1996. (see p. 18.)

[134] N. D. Yilmazer and M. Korth. "Enhanced semiempirical QM methods for biomolecular interactions". *Computational and Structural Biotechnology Journal*, 13, pp. 169–175, 2015. (see p. 19.)

[135] M. Pykal, P. Jurečka, František Karlický, and M. Otyepka. "Modelling of graphene functionalization". *Physical Chemistry Chemical Physics*, 18, pp. 6351–6372, 2016. (see pp. 20 and 135.)

[136] M. Planck. *Treatise on Thermodynamics*. Longmans, Greens, and Co., 1903. (see p. 21.)

[137] B. Tidor and M. Karplus. "The contribution of vibrational entropy to molecular association: the dimerization of insulin". *Journal of Molecular Biology*, 238(3), pp. 405–414, 1994. (see pp. 26 and 45.)

[138] D. A. McQuarrie. *Statistical Mechanics*. University Science Books, 2000. (see pp. 29, 55, 66, and 155.)

[139] J. E. Mayer and M. G. Mayer. *Statistical Mechanics*. John Wiley & Sons, Inc, 1940. (see p. 35.)

[140] M. Cecchini. "Quantum correction to the free energy difference between peptides and protein conformers". *Journal of Chemical Theory and Computation*, 11, pp. 4011–4022, 2015. (see pp. 41 and 46.)

[141] A. Marsh, E. G. Nolen, K. M. Gardinier, and J.-M. Lehn. "Janus molecules: synthesis of doubleheaded heterocycles containing two identical hydrogen bonding arrays". *Tetrahedron Letters*, 35(3), pp. 397–400, 1994. (see p. 42.)

[142] A. Marsh, M. Silvestri, and J.-M. Lehn. "Self-complementary hydrogen bonding heterocycles designed for the enforced self-assembly into supramolecular macrocycles". *Chemical Communications*, pp. 1527–1528, 1996. (see p. 42.)

[143] A. W. S. da Silva and W. F. Vranken. "ACPYPE – Antechamber python parser interface". *BMC research notes*, 5(1), p. 367, 2012. (see pp. 43 and 96.)

[144] B. R. Brooks, C. L. Brooks, A. D. Mackerell, L. Nilsson, R. J. Petrella, B. Roux, Y. Won, G. Archontis, C. Bartels, S. Boresch, A. Caflisch, L. Caves, Q. Cui, A. R. Dinner, M. Feig, S. Fischer, J. Gao, M. Hodoscek, W. Im, K. Kuczera, T. Lazaridis, J. Ma, V. Ovchinnikov, E. Paci, R. W. Pastor, C. B. Post, J. Z. Pu, M. Schaefer, B. Tidor, R. M. Venable, H. L. Woodcock, X. Wu, W. Yang, D. M. York, and M. Karplus. "CHARMM: the biomolecular simulation program". *Journal of Computational Chemistry*, 30(10), pp. 1545–1614, 2009. (see pp. 43, 73, 95, 128, 137, and 157.)

[145] S. Conti. "Thermo", 2014–2016. (see p. 43.)

[146] R. Gutzler, L. Cardenas, and F. Rosei. "Kinetics and thermodynamics in surface-confined molecular self-assembly". *Chemical Science*, 2(12), pp. 2290–2300, 2011. (see p. 49.)

[147] U. Mazur and K. W. Hiips. "Kinetic and thermodynamic processes of organic species at the solutionsolid interface: the view through an STM". *Chemical Communications*, 51, pp. 4737–4749, 2015. (see p. 49.)

[148] C.-A. Palma, J. Bjork, M. Bonini, M. S. Dyer, A. Llanes-Pallas, D. Bonifazi, M. Persson, and P. Samorì. "Tailoring bicomponent supramolecular nanoporous networks: phase segregation, polymorphism, and glasses at the solid- liquid interface". *Journal of the American Chemical Society*, 131(36), pp. 13062–13071, 2009. (see p. 50.)

[149] N. D. Lang and W. Kohn. "Theory of metal surfaces: charge density and surface energy". *Physical Review B*, 1(12), pp. 4555–4568, 1970. (see p. 51.)

[150] J. S. Rowlinson and B. Widom. *Molecular theory of capillarity*. Dover Publications, Inc., 2002. (see p. 51.)

[151] P. De Gennes, F. Brochard-Wyart, and D. Quéré. *Capillarity and wetting phenomena: drops, bubbles, pearls, waves*. Springer Science & Business Media, 2004. (see p. 51.)

[152] K. Reuter and M. Scheffler. "Composition, structure, and stability of RuO_2 (110) as a function of oxygen pressure". *Physical Review B*, 65(3), p. 035406, 2001. (see pp. 51, 53, and 84.)

[153] X.-G. Wang, W. Weiss, S. K. Shaikhutdinov, M. Ritter, M. Petersen, F. Wagner, R. Schlögl, and M. Scheffler. "The hematite (α-Fe_2O_3)(0001) surface: evidence for domains of distinct chemistry". *Physical Review Letters*, 81(5), p. 1038, 1998. (see p. 52.)

[154] X.-G. Wang, A. Chaka, and M. Scheffler. "Effect of the environment on α-Al_2O_3 (0001) surface structures". *Physical Review Letters*, 84(16), p. 3650, 2000. (see p. 52.)

[155] A. Kiejna, T. Ossowski, and T. Pabisiak. "Surface properties of the clean and Au/Pd covered Fe_3O_4(111): DFT and DFT+U study". *Physical Review B*, 85, p. 125414, 2012. (see p. 53.)

[156] K. Reuter and M. Scheffler. "First-principles atomistic thermodynamics for oxidation catalysis: surface phase diagrams and catalytically interesting regions". *Physical Review Letters*, 90(4), p. 046103, 2003. (see p. 53.)

[157] J. Kucera and A. Gross. "Adsorption of 4mercaptopyridine on Au (111): a periodic DFT study". *Langmuir*, 24(24), pp. 13985–13992, 2008. (see p. 53.)

[158] D. Loffreda, F. Delbecq, and P. Sautet. "Adsorption thermodynamics of acrolein on Pt (111) in realistic temperature and pressure from first-principle calculations". *Chemical Physics Letters*, 405(4), pp. 434–439, 2005. (see p. 54.)

[159] R. Gutzler, T. Sirtl, J. F. Dienstmaier, K. Mahata, W. M. Heckl, M. Schmittel, and M. Lackinger. "Reversible phase transitions in self-assembled monolayers at the liquid-solid interface: temperature-controlled opening and closing of nanopores". *Journal of the American Chemical Society*, 132(14), pp. 5084–5090, 2010. (see pp. 54 and 85.)

[160] J. F. Dienstmaier, K. Mahata, H. Walch, W. M. Heckl, M. Schmittel, and M. Lackinger. "On the scalability of supramolecular networks - High packing density vs optimized hydrogen bonds in tricarboxylic acid monolayers". *Langmuir*, 26(13), pp. 10708–10716, 2010. (see pp. 54, 76, 84, 85, and 86.)

[161] M. Mammen, E. I. Shakhnovich, J. M. Deutch, and G. M. Whitesides. "Estimating the entropic cost of self-assembly of multiparticle hydrogen-bonded aggregates based on the cyanuric acid - melamine lattice". *Journal of Organic Chemistry*, 63(12), pp. 3821–3830, 1998. (see pp. 55 and 85.)

[162] S. Lei, K. Tahara, F. C. De Schryver, M. Van der Auweraer, Y. Tobe, and S. De Feyter. "One building block, two different supramolecular surface-confined patterns: concentration in control at the solid-liquid interface". *Angewandte Chemie*, 120(16), pp. 3006–3010, 2008. (see pp. 56, 58, 76, 84, 85, and 86.)

[163] A. Bellec, C. Arrigoni, G. Schull, L. Douillard, C. Fiorini-Debuisschert, F. Mathevet, D. Kreher, A.-J. Attias, and F. Charra. "Solution-growth kinetics and thermodynamics of nanoporous self-assembled molecular monolayers". *Journal of Chemical Physics*, 134(12), p. 124702, 2011. (see pp. 56, 57, and 58.)

[164] M. Digne, P. Sautet, P. Raybaud, P. Euzen, and H. Toulhoat. "Hydroxyl Groups on γ-Alumina Surfaces: A DFT Study". *Journal of Catalysis*, 211, pp. 1–5, 2002. (see p. 58.)

[165] D. Costa, K. Sharkas, M. M. Islam, and P. Marcus. "Ab initio study of the chemical states of water on Cr_2o_3(0001): From the isolated molecule to saturation coverage". *Surface Science*, 603, pp. 2484–2493, 2009. (see p. 58.)

[166] A. Bouzoubaa, D. Costa, B. Diawara, N. Audiffren, and P. Marcus. "Insight of DFT and atomistic thermodynamics on the adsorption and insertion of halides onto the hydroxylated NiO (111) surface". *Corrosion Science*, 52, pp. 2643–2652, 2010. (see p. 58.)

[167] D. Costa, P.-A. Garrain, B. Diawara, and P. Marcus. "Biomolecule-Biomaterial Interaction: A DFT-D Study of Glycine Adsorption and Self-Assembly on Hydroxylated Cr_2o_3 Surfaces". *Langmuir*, 27, pp. 2747–2760, 2011. (see p. 58.)

[168] A. Hémeryck, A. Motta, J. Swiatowska, C. Pereira-Nabais, P. Marcusa, and D. Costa. "Diaminoethane adsorption and water substitution on hydrated TiO_2: a thermochemical study based on first-principles calculations". *Physical Chemistry Chemical Physics*, 15, pp. 10824–10834, 2013. (see pp. 58 and 59.)

[169] D. Costa, M. Smerieri, I. Tranca, L. Savio, L. Vattuone, and F. Tielens. "DFT atomistic thermodynamics applied to elucidate the driving force behind glutamic acid self-assemblies on silver (100) surface". *The Journal of Physical Chemistry C*, 118, pp. 29874–29879, 2014. (see p. 58.)

[170] W. Song, N. Martsinovich, W. M. Heckl, and M. Lackinger. "Born-Haber cycle for the monolayer self-assembly at the liquid-solid interface: assessing the enthalpic driving force". *Journal of the American Chemical Society*, 135(39), pp. 14854–14562, 2013. (see pp. 60, 62, and 86.)

[171] W. Song, N. Martsinovich, W. M. Heckl, and M. Lackinger. "Thermodynamics of 4,4'-stilbenedicarboxylic acid monolayer self-assembly at the nonanoic acid-graphite interface". *Physical Chemistry Chemical Physics*, 16, pp. 13239–13247, 2014. (see pp. 61 and 62.)

[172] W. Song, N. Martsinovich, W. M. Heckl, and M. Lackinger. "Thermodynamics of halogen bonded monolayer self-assembly at the liquid-solid interface". *Chemical Communications*, 50, pp. 13465–13468, 2014. (see p. 62.)

[173] J. R. Reimers, D. Panduwinata, J. Visser, Y. Chin, C. Tang, L. Goerigk, M. J. Ford, M. Sintic, T. Sum, M. J. J. Coenen, B. L. M. Hendriksen, J. A. A. W. Elemans, N. S. Hush, and M. J. Crossley. "A priori calculations of the free energy of formation from solution of polymorphic self-assembled monolayers". *Proceedings of the National Academy of Sciences*, 112(45), pp. E6101–E6110, 2015. (see pp. 62, 86, and 178.)

[174] J. R. Reimers, D. Panduwinata, J. Visser, Y. Chin, C. Tang, L. Goerigk, M. J. Ford, M. Baker, T. J. Sum, M. J. J. Coenen, B. L. M. Hendriksen, J. A. A. W. Elemans, N. S. Hush, and M. J. Crossley. "From chaos to order: chain-length dependence of the free energy of formation of meso-tetraalkylporphyrin self-assembled monolayer polymorphs". *The Journal of Physical Chemistry C*, 120(3), pp. 1739–1748, 2016. (see p. 62.)

[175] J. R. Reimers, M. J. Ford, and L. Goerigk. "Problems, successes and challenges for the application of dispersion-corrected density-functional theory combined with dispersion-based implicit solvent models to large-scale hydrophobic self-assembly and polymorphism". *Molecular Simulation*, 42(6–7), pp. 494–510, 2016. (see p. 63.)

[176] F. Floris and J. Tomasi. "Evaluation of the dispersion contribution to the solvation energy. A simple computational model in the continuum approximation". *Journal of Computational Chemistry*, 10(5), pp. 616–627, 1989. (see p. 63.)

[177] F. M. Floris, J. Tomasi, and J. L. Pascual Ahuir. "Dispersion and repulsion contributions to the solvation energy: refinements to a simple computational model in the continuum approximation". *Journal of Computational Chemistry*, 12(7), pp. 784–791, 1991. (see p. 63.)

[178] S. Conti and M. Cecchini. "Accurate and efficient calculation of the desorption energy of small molecules from graphene". *The Journal of Physical Chemistry C*, 119(4), pp. 1867–1879, 2015. (see pp. 72, 92, 135, and 155.)

[179] B. R. Brooks, Dušanka Janežič, and M. Karplus. "Harmonic analysis of large systems. I. Methodology". *Journal of Computational Chemistry*, 16(12), pp. 1522–1542, 1995. (see p. 72.)

[180] A. Warshel and S. Lifson. "Consistent force field calculations. II. Crystal structures, sublimation energies, molecular and lattice vibrations, molecular conformations, and enthalpies of alkanes". *The Journal of Chemical Physics*, 53(2), pp. 582–594, 1970. (see p. 73.)

[181] Y. Ye, S. Wei, Y. Wang, X. Shao, X. Xu, F. Cheng, L. Jianlong, and K. Wu. "A unified model: self-assembly of trimesic acid on gold". *The Journal of Physical Chemistry C*, 111, pp. 10138–10141, 2007. (see pp. 75 and 79.)

[182] A. Ciesielski, R. Perone, S. Pieraccini, G. P. Spada, and P. Samorì. "Nanopatterning the surface with ordered supramolecular architectures of N^9-alkylated guanines: STM reveals". *Chemical Communications*, 46(25), pp. 4493–4495, 2010. (see p. 76.)

[183] A. Ciesielski, S. Haar, G. Paragi, Z. Kupihár, Z. Kele, S. Masiero, C. F. Guerra, F. M. Bickelhaupt, G. P. Spada, L. Kovács, and others. "Supramolecular H-bonded porous networks at surfaces: exploiting primary and secondary interactions in a bi-component melamine–xanthine system". *Physical Chemistry Chemical Physics*, 15(30), pp. 12442–12446, 2013. (see p. 76.)

[184] S. Haar, A. Ciesielski, J. Clough, H. Yang, R. Mazzaro, F. Richard, S. Conti, N. Merstorf, M. Cecchini, V. Morandi, C. Casiraghi, and P. Samorì. "A supramolecular strategy to leverage the liquid-phase exfoliation of graphene in presence of surfactants: unraveling the role of the length of fatty acids". *Small*, 11, pp. 1691–1702, 2015. (see pp. 76 and 112.)

[185] C. L. Claypool, F. Faglioni, W. A. Goddard, H. B. Gray, N. S. Lewis, and R. A. Marcus. "Source of image contrast in STM images of functionalized alkanes on graphite: A systematic functional group approach". *The Journal of Physical Chemistry B*, 101(31), pp. 5978–5995, 1997. (see p. 76.)

[186] Q. Chen, H.-J. Yan, C.-J. Yan, G.-B. Pan, L.-J. Wan, G.-Y. Wen, and D.-Q. Zhang. "STM investigation of the dependence of alkane and alkane ($C_{18}h_{38}$, $C_{19}h_{40}$) derivatives self-assembly on molecular chemical structure on HOPG surface". *Surface Science*, 602(6), pp. 1256–1266, 2008. (see p. 76.)

[187] L. C. Giancarlo and G. W. Flynn. "Raising flags: applications of chemical marker groups to study self-assembly, chirality, and orientation of interfacial films by scanning tunneling microscopy". *Accounts of Chemical Research*, 33(7), pp. 491–501, 2000. (see p. 76.)

[188] J. Thrower, E. E. Friis, A. L. Skov, L. Nilsson, M. Andersen, L. Ferrighi, B. Jørgensen, S. Baouche, R. Balog, and B. Hammer. "Interaction between coronene and graphite form temperature-programmed desorption and DFT-vdW calculations: Importance of entropic effects and insights into graphite interlayer binding." *The Journal of*

Physical Chemistry C, 117(26), pp. 13520–13529, 2013. (see pp. 76, 92, and 143.)

[189] S. Conti, M. G. del Rosso, A. Ciesielski, J. Weippert, A. Böttcher, Y. Shin, G. Melinte, O. Ersen, C. Casiraghi, X. Feng, K. Müllen, M. M. Kappes, P. Samorì, and M. Cecchini. "Perchlorination of coronene enhances its propensity to self-assembly on graphene". *ChemPhysChem*, 17, pp. 352–357, 2016. (see pp. 76, 82, 83, and 136.)

[190] S. Bonacchi, M. El Garah, A. Ciesielski, M. Herder, S. Conti, M. Cecchini, S. Hecht, and P. Samorì. "Surface-induced selection during in situ photoswitching at the solid/liquid interface". *Angewandte Chemie International Edition*, 54, pp. 4865–4869, 2015. (see pp. 76 and 124.)

[191] X. Shen, X. Wei, P. Tan, Y. Yu, B. Yang, Z. Gong, H. Zhang, H. Lin, Y. Li, Q. Li, Y. Xie, and L. Chi. "Concentrationcontrolled reversible phase transitions in selfassembled monolayers on HOPG surfaces". *Small*, 11(19), pp. 2284–2290, 2015. (see pp. 76, 84, and 86.)

[192] F. Schedin, A. K. Geim, S. V. Morozov, E. W. Hill, P. Blake, M. I. Katsnelson, and K. S. Novoselov. "Detection of individual gas molecules adsorbed on graphene". *Nature Materials*, 6(9), pp. 652–655, 2007. (see p. 91.)

[193] V. Dua, S. P. Surwade, S. Ammu, S. R. Agnihotra, S. Jain, K. E. Roberts, S. Park, R. S. Ruoff, and S. K. Manohar. "All-organic vapor sensor using inkjet-printed reduced graphene oxide". *Angewandte Chemie International Edition*, 49(12), pp. 2154–2157, 2010. (see p. 91.)

[194] G. K. Dimitrakakis, E. Tylianakis, and G. E. Froudakis. "Pillared graphene: a new 3D network nanostructure for enhanced hydrogen storage". *Nano Letters*, 8(10), pp. 3166–3170, 2008. (see p. 91.)

[195] W. Chen, L. Duan, and D. Zhu. "Adsorption of polar and nonpolar organic chemicals to carbon nanotubes". *Environmental science & technology*, 41(24), pp. 8295–8300, 2007. (see p. 91.)

[196] D. Lin and B. Xing. "Adsorption of phenolic compounds by carbon nanotubes: role of aromaticity and substitution of hydroxyl groups". *Environmental science & technology*, 42(19), pp. 7254–7259, 2008. (see p. 91.)

[197] P. A. Redhead. "Thermal desorption of gases". *Vacuum*, 12(4), pp. 203–211, 1962. (see pp. 91, 92, and 140.)

[198] K. R. Paserba and A. J. Gellman. "Effects of conformational isomerism on the desorption kinetics of n-alkanes from graphite". *Journal of Chemical Physics*, 115(14), pp. 6737–6751, 2001. (see pp. 91, 92, 135, and 140.)

[199] S. L. Tait, Z. Dohnálek, C. T. Campbell, and B. D. Kay. "n–alkanes on MgO (100). II. Chain length dependence of kinetic desorption parameters for small n–alkanes". *The Journal of Chemical Physics*, 122(16), p. 164708, 2005. (see pp. 91 and 92.)

[200] S. L. Tait, Z. Dohnálek, C. T. Campbell, and B. D. Kay. "n-alkanes on Pt (111) and on C (0001)/Pt (111): Chain length dependence of kinetic desorption parameters". *The Journal of Chemical Physics*, 125(23), p. 234308, 2006. (see pp. 91, 92, 93, and 135.)

[201] J. Kysilka, M. Rubeš, L. Grajciar, P. Nachtigall, and O. Bludský. "Accurate description of argon and water adsorption on surfaces of graphene-based carbon allotropes". *The Journal of Physical Chemistry A*, 115(41), pp. 11387–11393, 2011. (see p. 91.)

[202] C. Thierfelder, M. Witte, S. Blankenburg, E. Rauls, and W. G. Schmidt. "Methane adsorption on graphene from first principles including dispersion interaction". *Surface Science*, 605(7), pp. 746–749, 2011. (see p. 91.)

[203] M. Rubeš, J. Kysilka, P. Nachtigall, and O. Bludský. "DFT/CC investigation of physical adsorption on a graphite (0001) surface". *Physical Chemistry Chemical Physics*, 12(24), pp. 6438–6444, 2010.

[204] O. Leenaerts, B. Partoens, and F. M. Peeters. "Adsorption of H_2o, NH_3, CO, NO_2, and NO on graphene: A first-principles study". *Physical Review B*, 77(12), p. 125416, 2008.

[205] K. T. Chan, J. B. Neaton, and M. L. Cohen. "First-principles study of metal adatom adsorption on graphene". *Physical Review B*, 77(23), p. 235430, 2008.

[206] O. V. Ershova, T. C. Lillestolen, and E. Bichoutskaia. "Study of polycyclic aromatic hydrocarbons adsorbed on graphene using density functional theory with empirical dispersion correction". *Physical Chemistry Chemical Physics*, 12(24), pp. 6483–6491, 2010. (see p. 91.)

[207] J. Björk, F. Hanke, C.-A. Palma, P. Samorì, M. Cecchini, and M. Persson. "Adsorption of aromatic and anti-aromatic systems on graphene through π-π stacking". *The Journal of Physical Chemistry Letters*, 1(23), pp. 3407–3412, 2010. (see pp. 91, 135, 137, and 155.)

[208] E. G. Gordeev, M. V. Polynski, and V. P. Ananikov. "Fast and accurate computational modeling of adsorption on graphene: a dispersion interaction challenge". *Physical Chemistry Chemical Physics*, 15(43), pp. 18815–18821, 2013. (see pp. 91, 135, and 155.)

[209] H. Ulbricht, R. Zacharia, N. Cindir, and T. Hertel. "Thermal desorption of gases and solvents from graphite and carbon nanotube surfaces". *Carbon*, 44(14), pp. 2931–2942, 2006. (see pp. 92 and 93.)

[210] R. Zacharia, H. Ulbricht, and T. Hertel. "Interlayer cohesive energy of graphite from thermal desorption of polyaromatic hydrocarbons". *Physical Review B*, 69(15), p. 155406, 2004. (see pp. 92 and 135.)

[211] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, Ö. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski, and D. J. Fox. "Gaussian 09 Revision A.02", 2009. (see pp. 95 and 139.)

[212] J. J. P. Stewart. "MOPAC 2012", 2012. (see pp. 95 and 137.)

[213] K. Vanommeslaeghe and A. D. MacKerell Jr. "Automation of the CHARMM General Force Field (CGenFF) I: Bond perception and atom typing". *Journal of Chemical Information and Modeling*, 52(12), pp. 3144–3154, 2012. (see pp. 96, 118, and 137.)

[214] K. Vanommeslaeghe, E. P. Raman, and A. D. MacKerell Jr. "Automation of the CHARMM General Force Field (CGenFF) II: Assignment of bonded parameters and partial atomic charges". *Journal of Chemical Information and Modeling*, 52(12), pp. 3155–3168, 2012. (see pp. 96, 118, and 137.)

[215] D. Van Der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark, and H. J. C. Berendsen. "GROMACS: fast, flexible, and free". *Journal of Computational Chemistry*, 26(16), pp. 1701–1718, 2005. (see p. 96.)

[216] H. J. C. Berendsen, D. van der Spoel, and R. van Drunen. "GROMACS: A message-passing parallel molecular dynamics implementation". *Computer Physics Communications*, 91(1), pp. 43–56, 1995.

[217] B. Hess, C. Kutzner, D. van der Spoel, and E. Lindahl. "GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation". *Journal of Chemical Theory and Computation*, 4(3), pp. 435–447, 2008.

(see p. 96.)

[218] R. H. Byrd, P. Lu, J. Nocedal, and C. Zhu. "A limited memory algorithm for bound constrained optimization". *SIAM Journal on Scientific Computing*, 16(5), pp. 1190–1208, 1995. (see p. 96.)

[219] C. Zhu, R. H. Byrd, P. Lu, and J. Nocedal. "Algorithm 778: L-BFGS-B: Fortran subroutines for large-scale bound-constrained optimization". *ACM Transactions on Mathematical Software (TOMS)*, 23(4), pp. 550–560, 1997. (see p. 96.)

[220] A. A. S. T. Ribeiro, B. A. C. Horta, and R. B. de Alencastro. "MKTOP: A program for automatic construction of molecular topologies". *Journal of the Brazilian Chemical Society*, 19(7), pp. 1433–1435, 2008. (see p. 96.)

[221] J. Hostaš, J. Řezáč, and P. Hobza. "On the performance of the semiempirical quantum mechanical PM6 and PM7 methods for noncovalent interactions". *Chemical Physics Letters*, 568, pp. 161–166, 2013. (see p. 96.)

[222] J. N. Israelachvili. *Intermolecular and surface forces*. Academic press, third ed., 2011. (see p. 105.)

[223] H. A. Kurtz, J. J. P. Stewart, and K. M. Dieter. "Calculation of the nonlinear optical properties of molecules". *Journal of Computational Chemistry*, 11(1), pp. 82–87, 1990. (see pp. 106 and 139.)

[224] J. Clayden, N. Greeves, and S. Warren. *Organic Chemistry*. Oxford University Press, 2012. (see p. 107.)

[225] J. Cioslowski, G. Liu, and D. Moncrieff. "Theoretical thermochemistry of homolytic C–C and C–Cl bond dissociations in unbranched perchloroalkanes". *The Journal of Physical Chemistry A*, 102(48), pp. 9965–9969, 1998. (see p. 107.)

[226] "CHIRON: The Biomarker Catalogue". 2008. Chiron AS, Stiklestadveien 1, NO-7041 Trondheim, Norway. (see p. 107.)

[227] V. Zitko. "Chlorinated Paraffins". In O. Hutzinger, editor, "The Handbook of Environmental Chemistry: Anthropogenie Compounds", Springer, 1980. (see p. 107.)

[228] J. P. Rabe and S. Buchholz. "Commensurability and mobility in two-dimensional molecular patterns on graphite". *Science*, 253(5018), pp. 424–427, 1991. (see p. 111.)

[229] B. J. Gyarfas, B. Wiggins, M. Zosel, and K. Hipps. "Supramolecular structures of coronene and alkane acids at the Au (111)-solution interface: A scanning tunneling microscopy study". *Langmuir*, 21(3), pp. 919–923, 2005. (see p. 113.)

[230] L. Kampschulte, M. Lackinger, A.-K. Maier, R. S. Kishore, S. Griessl, M. Schmittel, and W. M. Heckl. "Solvent induced polymorphism in supramolecular 1, 3, 5-benzenetribenzoic acid monolayers". *The Journal of Physical Chemistry B*, 110(22), pp. 10829–10836, 2006. (see p. 113.)

[231] M. Hibino, A. Sumi, H. Tsuchiya, and I. Hatta. "Microscopic origin of the odd-even effect in monolayer of fatty acids formed on a graphite surface by scanning tunneling microscopy". *The Journal of Physical Chemistry B*, 102(23), pp. 4544–4547, 1998. (see p. 113.)

[232] R. G. Weiss and P. Terech. "Molecular gels". *Materials with Self-Assembled Fibrillar Networks*, 2006. (see p. 117.)

[233] H. Miyasaka, T. Nobuto, A. Itaya, N. Tamai, and M. Irie. "Picosecond laser photolysis studies on a photochromic dithienylethene in solution and in crystalline phases". *Chemical Physics Letters*, 269(3), pp. 281–285, 1997. (see p. 123.)

[234] M. Irie, T. Fukaminato, K. Matsuda, and S. Kobatake. "Photochromism of diarylethene molecules and crystals: memories, switches, and actuators". *Chemical Reviews*, 114(24), pp. 12174–12277, 2014. (see p. 123.)

[235] R. C. Shallcross, P. Zacharias, A. Köhnen, P. O. Körner, E. Maibach, and K. Meerholz. "Photochromic transduction layers in organic memory elements". *Advanced Materials*, 25(3), pp. 469–476, 2013. (see p. 123.)

[236] E. Orgiu, N. Crivillers, M. Herder, L. Grubert, M. Pätzel, J. Frisch, E. Pavlica, D. T. Duong, G. Bratina, A. Salleo, N. Koch, S. Hecht, and P. Samorì. "Optically switchable transistor via energylevel phototuning in a bicomponent organic semiconductor". *Nature chemistry*, 4(8), pp. 675–679, 2012. (see p. 123.)

[237] N. Katsonis, A. Minoia, T. Kudernac, T. Mutai, H. Xu, H. Uji-i, R. Lazzaroni, S. De Feyter, and B. L. Feringa. "Locking of helicity and shape complementarity in diarylethene dimers on graphite". *Journal of the American Chemical Society*, 130(2), pp. 386–387, 2008. (see p. 123.)

[238] R. Arai, S. Uemura, M. Irie, and K. Matsuda. "Reversible photoinduced change in molecular ordering of diarylethene derivatives at a solution-HOPG interface". *Journal of the American Chemical Society*, 130(29), pp. 9371–9379, 2008.

[239] T. Sakano, Y. Imaizumi, T. Hirose, and K. Matsuda. "Formation of two-dimensionally ordered diarylethene annulated isomer at the liquid/HOPG Interface upon in situ UV irradiation". *Chemistry Letters*, 42(12), pp. 1537–1539, 2013. (see p. 124.)

[240] S. Yokoyama, T. Hirose, and K. Matsuda. "Phototriggered formation and disappearance of surface-confined self-assembly composed of photochromic 2-thienyl-type diarylethene: a cooperative model at the liquid/solid interface". *Chemical Communications*, 50(45), pp. 5964–5966, 2014. (see pp. 123 and 124.)

[241] M. Irie, T. Lifka, K. Uchida, S. Kobatake, and Y. Shindo. "Fatigue resistant properties of photochromic dithienylethenes: byproduct formation". *Chemical Communications*, 8, pp. 747–750, 1999. (see p. 124.)

[242] M. Herder, B. M. Schmidt, L. Grubert, M. Pätzel, J. Schwarz, and S. Hecht. "Improving the fatigue resistance of diarylethene switches". *Journal of the American Chemical Society*, 137(7), pp. 2738–2747, 2015. (see p. 124.)

[243] K. Uchida, E. Tsuchida, Y. Aoi, S. Nakamura, and M. Irie. "Substitution effect on the coloration quantum yield of a photochromic bisbenzothienylethene." *Chemistry Letters*, 28(1), pp. 63–64, 1999. (see p. 126.)

[244] W. Li, C. Jiao, X. Li, Y. Xie, K. Nakatani, H. Tian, and W. Zhu. "Separation of photoactive conformers based on hindered diarylethenes: efficient modulation in photocyclization quantum yields". *Angewandte Chemie International Edition*, 53(18), pp. 4603–4607, 2014. (see p. 126.)

[245] S. Grimme. "Do special noncovalent π–π stacking interactions really exist?" *Angewandte Chemie International Edition*, 47(18), pp. 3430–3434, 2008. (see p. 135.)

[246] F. London. "Zur theorie und systematik der molekularkräfte". *Zeitschrift für Physik*, 63(3), pp. 245–279, 1930. (see p. 135.)

[247] J. Černý, M. Kabeláč, and P. Hobza. "Doublehelical $\rightarrow$ ladder structural transition in the B-DNA is induced by a loss of dispersion energy". *Journal of the American Chemical Society*, 130(47), pp. 16055–16059, 2008. (see p. 135.)

[248] K. Autumn, M. Sitti, Y. A. Liang, A. M. Peattie, W. R. Hansen, S. Sponberg, T. W. Kenny, R. Fearing, J. N. Israelachvili, and R. J. Full. "Evidence for van der Waals adhesion in gecko setae". *Proceedings of the National Academy of Sciences*, 99(19), pp. 12252–12256, 2002. (see p. 135.)

[249] K. R. Paserba and A. J. Gellman. "Kinetics and energetics of oligomer desorption from surfaces". *Physical Review Letters*, 86(19), p. 4338, 2001. (see p. 135.)

[250] T. Müller, G. W. Flynn, A. T. Mathauser, and A. V. Teplyakov. "Temperature-programmed desorption studies of n-alkane derivatives on graphite: Desorption energetics and the influence of functional groups on adsorbate self-assembly". *Langmuir*, 19(7), pp. 2812–2821, 2003. (see p. 135.)

[251] R. Erni, M. D. Rossell, M.-T. Nguyen, S. Blankenburg, D. Passerone, P. Hartel, N. Alem, K. Erickson, W. Gannett, and A. Zettl. "Stability and dynamics of small molecules trapped on graphene". *Physical Review B*, 82(16), p. 165443, 2010. (see p. 135.)

[252] T. Baird, J. H. Gall, D. D. MacNicol, P. R. Mallinson, and C. R. Michie. "Perchlorocoronene: a novel host precursor". *Journal of the Chemical Society, Chemical Communications*, 22, pp. 1471–1473, 1988. (see p. 137.)

[253] Y.-Z. Tan, B. Yang, K. Parvez, A. Narita, S. Osella, D. Beljonne, X. Feng, and K. Müllen. "Atomically precise edge chlorination of nanographenes and its application in graphene nanoribbons". *Nature communications*, 4, 2013. (see p. 137.)

[254] A. D. Becke. "Density functional thermochemistry. III. The role of exact exchange". *Journal of Chemical Physics*, 98(7), pp. 5648–5652, 1993. (see p. 139.)

[255] C. Lee, W. Yang, and R. G. Parr. "Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density". *Physical Review B*, 37(2), p. 785, 1988.

[256] S. H. Vosko, L. Wilk, and M. Nusair. "Accurate spin-dependent electron liquid correlation energies for local spin density calculations: a critical analysis". *Canadian Journal of physics*, 58(8), pp. 1200–1211, 1980.

[257] P. J. Stephens, F. J. Devlin, C. F. Chabalowski, and M. J. Frisch. "Ab initio calculation of vibrational absorption and circular dichroism spectra using density functional force fields". *The Journal of Physical Chemistry C*, 98(45), pp. 11623–11627, 1994. (see p. 139.)

[258] T. H. Dunning Jr. "Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen". *Journal of Chemical Physics*, 90(2), pp. 1007–1023, 1989. (see p. 139.)

[259] A. M. De Jong and J. W. Niemantsverdriet. "Thermal desorption analysis: Comparative test of ten commonly applied procedures". *Surface Science*, 233(3), pp. 355–365, 1990. (see pp. 139 and 140.)

[260] G. McKay, M. J. Bino, and A. R. Altamemi. "The adsorption of various pollutants from aqueous solutions on to activated carbon". *Water Research*, 19(4), pp. 491–495, 1985. (see p. 155.)

[261] A. R. Kahn, T. A. Al-Bahri, and A.-H. A. "Adsorption of pthenol based organic pollutants on activated carbon from multicomponent diluite aqueous solutions." *Water Research*, 31(8), pp. 2102–2112, 1997. (see p. 155.)

[262] S. J. Geelhood, T. A. Horbett, W. K. Ward, M. D. Wood, and M. J. Quinn. "Passivating protein coatings for implantable glucose sensors: Evaluation of protein retention". *Journal of Biomedical Materials Research Part B: Applied Biomaterials*, 81(1), pp. 251–260, 2007. (see p. 155.)

[263] J. M. Anderson, A. Rodriguez, and D. T. Chang. "Foreign body reaction to biomaterials". In "Seminars in immunology", vol. 20, pp. 86–100. 2008. (see p. 155.)

[264] M. D. Tyka, A. R. Clarke, and R. B. Sessions. "An efficient, path-independent method for free-energy calculations". *Journal of Physical Chemistry B*, 110, pp. 17212–17220, 2006. (see pp. 155, 159, 161, and 166.)

[265] M. Cecchini, S. V. Krivov, M. Spichty, and M. Karplus. "Calculation of free-energy differences by confinement simulations. Application to peptide conformers". *Journal of Physical Chemistry B*, 113, pp. 9728–9740, 2009. (see pp. 155, 159, 161, and 166.)

[266] I. Langmuir. "The constitution and fundamental properties of solids and liquids. Part I. Solids." *Journal of the American Chemical Society*, 38(11), pp. 2221–2295, 1916. (see p. 157.)

[267] I. Langmuir. "The adsorption of gases on plane surfaces of glass, mica and platinum." *Journal of the American Chemical Society*, 40(9), pp. 1361–1403, 1918. (see p. 157.)

[268] S. Brunauer, P. H. Emmett, and E. Teller. "Adsorption of gases in multimolecular layers". *Journal of the American Chemical Society*, 60(2), pp. 309–319, 1938. (see p. 157.)

[269] C. H. Giles, D. Smith, and A. Huitson. "A general treatment and classification of the solute adsorption isotherm. I. Theoretical". *Journal of Colloid and Interface Science*, 47(3), pp. 755–765, 1974. (see p. 157.)

[270] G. Limousin, J.-P. Gaudet, L. Charlet, S. Szenknect, V. Barthes, and M. Krimissa. "Sorption isotherms: a review on physical bases, modeling and measurement". *Applied Geochemistry*, 22(2), pp. 249–275, 2007.

[271] K. Y. Foo and B. H. Hameed. "Insights into the modeling of adsorption isotherm systems". *Chemical Engineering Journal*, 156(1), pp. 2–10, 2010. (see p. 157.)

[272] L. H. Sprowl, C. T. Campbell, and L. Árnadóttir. "Hindered translator and hindered rotor models for adsorbates: partition functions and entropies". *The Journal of Physical Chemistry C*, 120(18), pp. 9719–9731, 2016. (see p. 159.)

[273] C. T. Campbell, L. H. Sprowl, and L. Árnadóttir. "Equilibrium constants and rate constants for adsorbates: two-dimensional (2D) ideal gas, 2D ideal lattice gas, and ideal hindered translator models". *The Journal of Physical Chemistry C*, 120(19), pp. 10283–10297, 2016. (see p. 159.)

[274] B. Roux and T. Simonson. "Implicit solvent models". *Biophysical chemistry*, 78(1), pp. 1–20, 1999. (see p. 178.)

[275] N. A. Baker, D. Sept, S. Joseph, M. J. Holst, and J. A. McCammon. "Electrostatics of nanosystems: application to microtubules and the ribosome". *Proceedings of the National Academy of Sciences*, 98(18), pp. 10037–10041, 2001. (see p. 178.)

[276] M. J. Abraham, T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess, and E. Lindahl. "GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers". *SoftwareX*, 1–2, pp. 19–25, 2015. (see p. 179.)

[277] P. V. Klimovich, M. R. Shirts, and D. L. Mobley. "Guidelines for the analysis of free energy calculations". *Journal of Computer-Aided Molecular Design*, 29(5), pp. 397–411, 2015. (see p. 179.)

[278] G. Gilli and P. Gilli. "Towards an unified hydrogen-bond theory". *Journal of Molecular Structure*, 552, pp. 1–15, 2000. (see p. 180.)

[279] J. Řezáč, K. E. Riley, and P. Hobza. "Extensions of the S66 Data Set: More Accurate Interaction Energies and Angular-Displaced Nonequilibrium Geometries". *Journal of Chemical Theory and Computation*, 7, pp. 3466–3470, 2011. (see pp. 180 and 181.)

UNIVERSITÉ DE STRASBOURG

Simone CONTI

Études computationnelles de l'Auto-Assemblage Moléculaire au Niveau des Surfaces: de la Conception Rationnelle à la Fonction

Résumé

L'auto-assemblage moléculaire au niveau des surfaces est un exemple majeur de l'auto-organisation de la matière avec des applications technologiques d'intérêt. La capacité de prédire la structure de la monocouche auto-assemblée (SAM) formée à l'équilibre est d'une importance fondamentale. Dans cette thèse, je présente une théorie aux premiers principes pour l'interprétation de l'auto-assemblage 2D basée sur la modélisation et la thermodynamique statistique. La méthode mise au point permet d'accéder à la stabilité thermodynamique de la SAM et à sa dépendance à l'égard de la concentration. Cela permet d'étudier les équilibres concurrentiels au niveau des surfaces et de rationaliser le polymorphisme en 2D. De plus, cette théorie prédit l'existence d'une concentration critique de monomères au-dessus de laquelle l'auto-assemblage 2D est favorisé. Celle-ci est utilisée pour définir une échelle absolue définissant le propension à l'auto-assemblage en 2D. Enfin, quatre applications technologiques sont discutées, exposant les possibilités de la méthode développée.

Mots-clés: Auto-Assemblage Moléculaire au Niveau des Surfaces, Modélisation Moléculaire, Thermodynamique Statistique

Résumé en anglais

Molecular self-assembly at surfaces is a prominent example of self-organization of matter with outstanding technological applications. The ability to predict the structure of the self-assembled monolayer (SAM) formed at equilibrium is of great fundamental and technological importance. In this dissertation I present a self-consistent theory for a first-principle interpretation of 2D self-assembly based on modeling and statistical thermodynamics. The developed framework provides access to the thermodynamic stability of the SAM and to its concentration dependence. This allows to study competitive equilibria at surfaces and to rationalize the 2D polymorphism evidenced by scanning probe techniques. The theory predicts the existence of a critical concentration of monomers, which is used to set an absolute scale for the 2D self-assembly propensity. Last, four technological applications are discussed, showing the potentials of the developed framework.

Keywords: Molecular Self-Assembly at Surfaces, Molecular Modeling, Statistical Thermodynamics