

# Theoretical Investigations of Molecular Self-Assembly on Symmetric Surfaces

by

Emilian Tuca

B.Sc., “Al. I. Cuza” University, 2009

M.Sc., “Al. I. Cuza” University, 2011

A Dissertation Submitted in Partial Fulfillment of the  
Requirements for the Degree of

DOCTOR OF PHILOSOPHY

in the Department of Chemistry

© Emilian Tuca, 2019  
University of Victoria

All rights reserved. This dissertation may not be reproduced in whole or in part, by photocopying or other means, without the permission of the author.

Theoretical Investigations of Molecular Self-Assembly on Symmetric Surfaces

by

Emilian Tuca

B.Sc., “Al. I. Cuza” University, 2009

M.Sc., “Al. I. Cuza” University, 2011

Supervisory Committee

---

Prof. Irina Paci, Supervisor  
(Department of Chemistry)

---

Prof. Cornelia Bohne, Departmental Member  
(Department of Chemistry)

---

Prof. Matthew Moffitt, Departmental Member  
(Department of Chemistry)

---

Prof. Byoung-Chul Choi, Outside Member  
(Department of Physics and Astronomy)

## Supervisory Committee

---

Prof. Irina Paci, Supervisor  
(Department of Chemistry)

---

Prof. Cornelia Bohne, Departmental Member  
(Department of Chemistry)

---

Prof. Matthew Moffitt, Departmental Member  
(Department of Chemistry)

---

Prof. Byoung-Chul Choi, Outside Member  
(Department of Physics and Astronomy)

---

## ABSTRACT

Surface self-assembly, the spontaneous aggregation of molecules into ordered, stable, noncovalently joined structures in the presence of a surface, is of great importance to the bottom-up manufacturing of materials with desired functionality. As a bulk phenomenon informed by molecular-level interactions, surface self-assembly involves coupled processes spanning multiple length scales. Consequently, a computational approach towards investigating surface self-assembled systems requires a combination of quantum-level electronic structure calculations and large-scale multi-body classical simulations. In this work we use a range of simulation approaches from quantum-based methods, to classical atomistic calculations, to mean-field approximations of bulk mixed phases, and explore the self-assembly strategies of simple dipoles and polyaromatic hydrocarbons on symmetric surfaces.

## ACKNOWLEDGEMENTS

I would like to express my gratitude to my advisor Prof. Irina Paci for her continuous support, bequeathed knowledge and unyielding patience. I could not have asked for a better advisor during all these years.

*According to a world-famous theoretical chemist, there are several dozens of quantum-chemical approaches available, and each of them is superior to all the others.*

Richard Dronskowski

## DEDICATION

I dedicate this work to Tania, my closest friend and colleague, and to Steven, my  
bedrock and complement

# Contents

<b>Supervisory Committee</b>	<b>ii</b>
<b>Abstract</b>	<b>iii</b>
<b>Acknowledgements</b>	<b>iv</b>
<b>Dedication</b>	<b>v</b>
<b>Table of Contents</b>	<b>vi</b>
<b>List of Tables</b>	<b>xi</b>
<b>List of Figures</b>	<b>xiii</b>
<b>1 Introduction</b>	<b>1</b>
1.1 The Importance of Surface Self-Assembly . . . . .	1
1.1.1 Complex Materials . . . . .	1
1.1.2 Self-Assembly . . . . .	2
1.1.3 Surface Self-Assembly . . . . .	4
1.2 Background on Surface Self-Assembly . . . . .	6
1.2.1 Self-Assembled Monolayers . . . . .	6
1.2.2 Aromatic Compounds on Coinage Metals . . . . .	7
1.2.3 Helicenes on Metals . . . . .	8
1.3 Theoretical Challenges for Surface Self-Assembly . . . . .	10
1.3.1 The Challenge of Multiple Scales . . . . .	10
1.3.2 The Challenge of Accurate, Large-Scale Simulations . . . . .	11
1.3.3 The Need for Theoretical Frameworks . . . . .	11
1.3.4 The Reductionist Approach . . . . .	12
1.4 What This Thesis Is About . . . . .	14

<b>2</b>	<b>Computational Methods for Studying Surface Self-Assembly</b>	<b>18</b>
2.1	Introduction . . . . .	18
2.2	Modelling Fundamentals . . . . .	20
2.2.1	Quantum Mechanics . . . . .	20
2.2.2	Statistical Mechanics and Classical Representations . . . . .	21
2.2.3	System Size and Representation . . . . .	25
2.3	Classical Methods . . . . .	26
2.3.1	Molecular Dynamics . . . . .	26
2.3.2	Monte Carlo Techniques . . . . .	28
2.3.3	Force Fields . . . . .	44
2.4	Quantum Methods . . . . .	50
2.4.1	The Hartree-Fock Method . . . . .	52
2.4.2	Density Functional Theory . . . . .	54
2.5	Treatment of Dispersion with DFT for Aromatic Compounds on Transition Metals . . . . .	58
2.5.1	Classification of Dispersion Corrections . . . . .	58
2.5.2	The Exchange-Hole Dipole Moment Dispersion Model . . . . .	62
2.5.3	Benzene On Metal Surfaces . . . . .	64
2.5.4	Larger Aromatic Compounds on Metal Surfaces . . . . .	66
2.6	Electronic Structure Calculations in Solid State Materials . . . . .	67
2.6.1	Wave Functions in Periodic Systems . . . . .	68
2.6.2	Basis Set Considerations . . . . .	71
2.6.3	Electronic Structure Analysis . . . . .	74
<b>3</b>	<b>Fundamental Aspects in Surface Self-Assembly: A Theoretical Study of Polarity and Shape<sup>s</sup></b>	<b>81</b>
3.1	Introduction . . . . .	81
3.1.1	Lennard-Jones Spheres on Surfaces . . . . .	82
3.1.2	Dipolar Models on Surfaces . . . . .	83
3.2	Methodology . . . . .	85
3.2.1	The Lennard-Jones-Electrostatic Potential . . . . .	85
3.2.2	Modelling Details . . . . .	85
3.2.3	Parallel Tempering Monte Carlo . . . . .	88
3.2.4	Order Parameters . . . . .	91
3.3	Series A: Diatomic Dipoles . . . . .	92

3.3.1	The Influence of Surface Potential . . . . .	93
3.3.2	The Influence of Dipole Strength . . . . .	95
3.3.3	Quantifying Order . . . . .	95
3.4	Series C, D, L: Linear Dipoles . . . . .	97
3.4.1	Equal Dipole Moments, Different Molecular Lengths . . . . .	97
3.4.2	Equal Partial Charges, Different Molecular Lengths . . . . .	101
3.4.3	Polymorphism . . . . .	104
3.4.4	Varying the Molecular Location of Charged Groups . . . . .	108
3.5	Series G: Asymmetric Dipoles . . . . .	109
3.6	Conclusion . . . . .	115
<b>4</b>	<b>Structural Analysis of Helicene Molecules Adsorbed on Symmetric Surfaces<sup>†</sup></b>	<b>117</b>
4.1	Introduction . . . . .	117
4.2	Helicene Properties . . . . .	118
4.2.1	Structure and Mechanical Properties . . . . .	119
4.2.2	Aromaticity and Opto-Electronic Properties . . . . .	120
4.2.3	Chirality . . . . .	122
4.3	Methodology . . . . .	122
4.3.1	Models . . . . .	122
4.3.2	Topology, Structure and Nomenclature . . . . .	124
4.3.3	Computational Details . . . . .	130
4.4	Coronene . . . . .	132
4.4.1	Planar Reference . . . . .	132
4.4.2	Monomers and Dimers . . . . .	133
4.4.3	Other Planar Aromatic Compounds . . . . .	135
4.5	Helicene Dimers in Gas Phase . . . . .	137
4.5.1	Steric Effects . . . . .	137
4.5.2	Size Effects . . . . .	139
4.6	Adsorption of Helicene Monomers . . . . .	140
4.6.1	General Trends . . . . .	140
4.6.2	Electronic Structure . . . . .	141
4.6.3	Steric Effects . . . . .	141
4.6.4	Size Effects . . . . .	143
4.6.5	Adsorption Energies . . . . .	143

4.7	Helicene Dimers on Surface . . . . .	144
4.7.1	Dimer Stability . . . . .	144
4.7.2	Geometry Changes in the Surface-Proximal Molecules . . . . .	146
4.8	Conclusion . . . . .	148
<b>5</b>	<b>Monte Carlo Simulations of Helicene Systems</b>	<b>150</b>
5.1	Introduction . . . . .	150
5.2	Methodology . . . . .	151
5.2.1	Model <i>D</i> . . . . .	151
5.2.2	Model <i>S</i> . . . . .	154
5.2.3	Simulation Details . . . . .	156
5.3	Results and Discussion . . . . .	157
5.3.1	Surface Potential vs. Intermolecular Interactions . . . . .	157
5.3.2	Beyond Competition: Multilayered Structures . . . . .	161
5.3.3	Chirality and Enantioselectivity . . . . .	166
5.3.4	Future Prospects . . . . .	168
5.4	Conclusion . . . . .	169
<b>6</b>	<b>Conclusions</b>	<b>171</b>
<b>A</b>	<b>Two-Component BSSE-Corrected Binding Energy Determination with SIESTA</b>	<b>177</b>
A.1	Introduction . . . . .	177
A.2	Binding Energy in SIESTA . . . . .	178
A.2.1	SIESTA Job Directory Details . . . . .	179
A.2.2	SIESTA Input Details . . . . .	180
A.2.3	Box Size and Geometry Constraints . . . . .	183
A.2.4	The Basics of a SIESTA Geometry Relaxation Run . . . . .	184
A.2.5	Submitting Jobs and Checking Output . . . . .	185
A.3	Interaction Energy and the Counterpoise Correction in SIESTA . . . . .	188
A.3.1	The Source of the Basis Set Superposition Error . . . . .	188
A.3.2	Seven Energy Terms . . . . .	190
A.3.3	From Optimization Step to Correction Step . . . . .	190
A.3.4	Input-Output Flow . . . . .	193
A.3.5	Checking Output . . . . .	195

<b>B Structural Details for Optimized Helicene Molecules</b>	<b>197</b>
<b>Bibliography</b>	<b>202</b>

# List of Tables

Table 3.1	Interaction parameters for the investigated models . . . . .	89
Table 3.2	First, second and fourth order parameters $S_1$ , $S_2$ and $S_4$ . . . . .	94
Table 3.3	Order Parameters for series G relative to the TAIL axis for temperatures around $T^*=1$ . . . . .	115
Table 4.1	Binding energy nomenclature. . . . .	132
Table 4.2	Binding energies and distances for [6]C systems. . . . .	133
Table 4.3	Binding energies in gas-phase helicene dimers. . . . .	138
Table 4.4	Angles between equivalent rings in gas-phase helicene dimers. . . . .	139
Table 4.5	Binding energies in surface-adsorbed helicene monomers and dimers. . . . .	140
Table 5.1	Carbon constitution for helicene series. . . . .	153
Table 5.2	Lennard-Jones parameters for models S and D . . . . .	154
Table 5.3	Numerical details for MC simulations of models <i>S</i> and <i>D</i> : number of surface atoms included, number of molecules present in the box, simulation box length and number of MC steps. For model <i>S</i> two system sizes were considered. . . . .	156
Table B.1	Numerical values for angles and distances between rings for penta-, hexa- and heptahelicene monomers and dimers optimized in the absence of a surface. Letters <b>L</b> , <b>U</b> , <b>b</b> and <b>t</b> stand for homochiral/“like”, heterochiral/“unlike”, “bottom” and “top” respectively. The letters <b>a</b> and <b>d</b> refer to “angle” and “distance” respectively. First half of the table represents angles in degrees, second half represents distances in Å. . . . .	198

- Table B.2 Angles and distances made by pentahelicene monomer, homochiral dimer and heterochiral dimer with the Ag, Au and Cu surfaces. Molecules are treated as collections of five ortho-fused rings, labelled **b1** – **b5**, **t1** – **t5**. Letters **b** and **t** refer to the “bottom” (closest to the surface) and “top” (furthest from the surface) molecule respectively. The letter **s** refers to the top surface plane, taken as a reference. Letters **M**, **L** and **U** refer to monomer, homochiral/“like” and heterochiral/“unlike”, respectively. The letters **a** and **d** refer to “angle” and “distance” respectively. First half of the table represents angles in degrees, second half represents distances in Å. . . . . 199
- Table B.3 Angles and distances made by hexahelicene monomer, homochiral dimer and heterochiral dimer with the Ag, Au and Cu surfaces. Molecules are treated as collections of six ortho-fused rings, labelled **b1** – **b6**, **t1** – **t6**. Letters **b** and **t** refer to the “bottom” (closest to the surface) and “top” (furthest from the surface) molecule respectively. The letter **s** refers to the top surface plane, taken as a reference. Letters **M**, **L** and **U** refer to monomer, homochiral/“like” and heterochiral/“unlike”, respectively. The letters **a** and **d** refer to “angle” and “distance” respectively. First half of the table represents angles in degrees, second half represents distances in Å. . . . . 200
- Table B.4 Angles and distances made by heptahelicene monomer, homochiral dimer and heterochiral dimer with the Ag, Au and Cu surfaces. Molecules are treated as collections of seven ortho-fused rings, labelled **b1** – **b7**, **t1** – **t7**. Letters **b** and **t** refer to the “bottom” (closest to the surface) and “top” (furthest from the surface) molecule respectively. The letter **s** refers to the top surface plane, taken as a reference. Letters **M**, **L** and **U** refer to monomer, homochiral/“like” and heterochiral/“unlike”, respectively. The letters **a** and **d** refer to “angle” and “distance” respectively. First half of the table represents angles in degrees, second half represents distances in Å. . . . . 201

# List of Figures

- Figure 2.1 Schematic representation of systems associated with the three ensembles: (a) microcanonical ensemble – the system is isolated from the rest of the universe; (b) canonical ensemble – the system is in thermal equilibrium with a large heat bath at constant temperature and is closed; the system experiences energy fluctuations; (c) grand canonical ensemble – the system is in chemical equilibrium with a large reservoir with a constant temperature and chemical potential; both energy and composition fluctuations occur. . . . . 25
- Figure 2.2 Representations of the most common configurational moves in Monte Carlo simulations: (a) basic translation move – random displacement in a random direction by a random magnitude of a particle randomly chosen; (b) basic rotation move – random rotation around a random axis by a random angle of a particle randomly chosen; care must be taken in the choice of rotation matrix as to not introduce any orientational bias, which would skew the sampling. . . . . 31
- Figure 2.3 Additional common moves used in Monte Carlo simulations: (a) basic molecular swap – the geometric center is swapped within a random pair of non-identical molecules; (b) spin flip move; analogously, chiral molecules may have their configuration inverted via the same procedure; (c) cluster moves – a collection of particles identified as a cluster are displaced (or rotate) together with the same random parameters; identifying clusters is difficult in general. . . . . 32

- Figure 2.4 The probability distribution  $P_E$  of finding the system at a given temperature  $T$  in a state with energy  $E$ . Individual low energy states are preferred, but higher energy states are more numerous and overall contribute to the average more. High energy states require large energy fluctuations between the system and the bath, which are exponentially less likely to occur than lower energy states and quantitatively depend on  $T$ . The outcome of these two energy constraints leads to a distribution resembling a bell curve whose peak ( $\langle E \rangle_T$ ) and width are  $T$ -dependent. The peak and width of the distribution increase with  $T$ . . . . . 33
- Figure 2.5 Representation of the configurational state space (S) as a two-dimensional surface contour map in potential energy (PES). Blue domains correspond to low energy configurations, red areas contain high energy configurations, with purple and pink domains representing intermediate energy. Two accepted MC moves translate the system across S via path (a–c) and are indicated on S by yellow arrows. Configurations (snapshots) (d–i) of varying energies are displayed and highlighted on S with green squares. Configurational state spaces of many-body systems have a large number of dimensions. . . . . 34
- Figure 2.6 Sampling of the configurational state space (S) with the Monte Carlo technique. Starting from a random initial configuration (a) the system reaches configuration (b) through a set number of Monte Carlo moves  $MC_{\text{steps}}^{\text{pre-eq}}$ . This is marked as the pre-equilibrium phase and from this step onward the sampling is assumed to be representative of equilibrium. A number of Monte Carlo moves  $MC_{\text{steps}}^{\text{sample}}$  samples the system for a set number of moves ending with configuration (c). The pre-equilibrium configurations are represented on S with yellow squares and the path taken through S is represented with yellow arrows. Green squares are the configurations constituting the sample  $S_{\text{sample}}$ , completed at the end of the run. The path through S after reaching equilibrium is omitted. . . . . 35

- Figure 2.7 Entrapment in local minima on the PES. Two runs [(a-c) vs (d-f)], are highlighted on S with different shades of yellow and green. Despite close initial configurations [(a) vs. (b)], the two runs rapidly diverge and reach different minima [(d) vs (e)], where they remain stuck until the end of the simulation [(c) vs (f)]. Neither reaches the true global minima, represented by snapshot (g). The probability of crossing the barriers required to reach (g) with classic Monte Carlo moves is very low at this temperature. . . . . 38
- Figure 2.8 The Parallel Tempering Monte Carlo Formalism. Four identical replicas  $R_1$ – $R_4$  sample the configurational state space S at different temperatures  $T_1$  through  $T_4$ , from lowest to highest respectively (top row). The sampling mobility of each replica is shown in the middle row. Replicas at low temperatures sample low energy domains and cannot pass between domains if the energetic barrier is too high. Green spirals represent the areas accessible by each replica in isolation. The bottom part displays the overlap of energy values between replicas at different temperatures. The probability of two replicas swapping configurations (and therefore energy levels) via a parallel tempering MC move is proportional to this overlap. . . . . 41

Figure 2.9 The relationship between temperature spacing, swapping probability and the mobility of low-temperature replicas on PES. Top example: temperatures  $T_1$  and  $T_2$  are so far apart that their energy overlap is negligible and the two replicas never swap configurations (top-left graph). The two replicas sample the configurational space in isolation, each with their assigned temperature (top-middle sketches). As a result, the low-temperature replica (blue) is locally trapped. The zero-swapping frequency is indicated in the top-right graph. Bottom example: in a different run a replica at intermediate temperature (purple) is introduced. The energy overlap between the low and intermediate temperature and the overlap between the intermediate and high temperature are both significantly greater than zero (bottom-left graph). As a result, the low-temperature replica gains access to the previously isolated minima (bottom-middle sketches). The colour of arrows linking different low-energy domains indicates the temperature required to overcome the energy barrier between them. The movement of different replicas across different temperatures via swaps is indicated in the bottom-right graph. . . . . . 42

Figure 2.10 Demonstration of how swap moves eliminate entrapment for low temperature replicas. Replica  $R_1$  (blue border) samples at low temperature  $T_1$  (blue area).  $R_2$  (purple border) samples at a higher temperature  $T_2$  (purple area). (a) –  $R_1$  is currently stuck in a local minima with two clusters that will not merge; (b) – following a successful swap, the configuration of  $R_1$  is transferred to the higher temperature; (c) – the two-cluster configuration has a chance to be relaxed by the increased mobility of a higher temperature; (d) – a subsequent swap brings the relaxed configuration back to the low temperature; (e) –  $R_1$  has now reached a different local minima. . . . . . 43

- Figure 2.11 Graphical representations of the Lennard-Jones energy as a function of the distance  $r_{i,j}$  between atomic pair  $(i, j)$  for two sets of parameters: (a)  $\sigma_{ij} = 2.0$ ,  $\epsilon_{ij} = 1.0$ , indicative of a softer potential between larger atoms; (b)  $\sigma_{ij} = 1.0$ ,  $\epsilon_{ij} = 3.0$ , indicative of a sharper potential between smaller atoms. The repulsive and attractive components of the potential are displayed with red-dashed and blue-dotted curves respectively. Their sum is shown as the continuous purple curve. All quantities presented are in reduced units. . . . . 49
- Figure 3.1 A complete listing of models used throughout this study. The atoms carrying partial charges highlighted in yellow and purple. Grey atoms carried no charge. Symbols are determined by the effect they help study. . . . . 87
- Figure 3.2 Snapshots of system C2- $l$  with different number of molecules: 50 (**panel a**) and 80 (**panel b**) reflecting similar SA behaviour for varying system sizes. Both snapshots are at  $T^* = 1.0$ . Regardless of size, the commensurate nature of phase and the antiparallel alignment of neighbouring molecules are apparent. **Panel c**: values for the order parameters (y axis) as a function of temperature,  $T^*$  (x axis).  $S_1$  – continuous line,  $S_2$  – dashed line,  $S_4$  – dashed-dotted line for small (50 molecules, red curve) and large (80 molecules, blue curve) systems. The similarity between curves suggests no significant system size effects. . . . . 90
- Figure 3.3 Effects of surface potential on order. Snapshots of A1 molecules adsorbed on a surface with  $\epsilon = 1$  (a) and  $\epsilon = 3$  (b). For a low surface potential (a) the alignment is weaker than for a high surface potential (b), as can be seen when comparing the S values in Table 3.2. . . . . 93

- Figure 3.4 Pattern formation in simple dipolar adsorbates. Snapshots of A-series molecules adsorbed on a surface with two different LJ energy parameters  $\epsilon_{ss}^*$  are presented. Panels (a) and (b) display condensed phase snapshots for A1 and A3, respectively, when surface interaction is relatively weak ( $\epsilon_{ss}^* = 1$ ). An insert in (b) shows a side view of the crystal. Panels (c) and (d) show snapshots for molecules A1 and A2, respectively, at  $\epsilon_{ss}^* = 3$ . Panels present replicas at  $T^*=0.1-0.2$ . . . . . 96
- Figure 3.5 Effects of molecular length on SA structure. Snapshots of dominant low-temperature phases corresponding to series D1 (panel a), D2 (panel b) and D3 (panel c), and an intermediate-temperature antiparallel phase for D4 (panel d) are presented. . . . . 99
- Figure 3.6 Temperature dependence of order parameters in systems D.  $S_1$ ,  $S_2$  and  $S_4$  are shown in panels a-c, respectively. The fraction of molecules in contact with the surface is presented in panel d. Black lines, red dashes, blue dots and green dot-dashed lines represent systems D1-4, respectively. . . . . 100
- Figure 3.7 Order parameters for **D4/C4-l** system. **Panel a:** averages for  $S_1$  (black line) and  $S_2$  (red dashed line) with associated standard deviations, as functions of reduced temperature,  $T^*$ .  $S_1$  is characterized by values close to zero and a low standard deviation, signifying the lack of nematic phase. In contrast,  $S_2$  has a high variation among snapshots in the region  $T^* = 0.5 - 3.5$ . **Panel b:** variation of  $S_1$  (black) and  $S_2$  (red) between a series of microstates for D4 at  $T^* = 2.9$ . The x axis represents the number of snapshots  $S_2$  fluctuates randomly between low (0.2) and high (0.8) values, signalling some variation in order among neighbours, compounded by a large variation in the range within which molecules adopt an antiparallel alignment. . . . . 101
- Figure 3.8 Snapshots of D4 system reflecting weaker (**panel a**, lower temperature) and stronger (**panel b**, higher temperature) antiparallel motifs. The stronger monolayer antiparallel alignment with increasing temperature in the region  $T^* = 0.5 - 3.5$  is attributed to the entropic effect of extended antiparallel, surface-adsorbed phase. . . . . 102

Figure 3.9 Strong alignment for long, high-charge molecules. A low-temperature snapshot C4-*h* is shown . . . . . 103

Figure 3.10 Fraction of molecules touching the surface as a function of  $T^*$ .  
**Panel a – series A:** black line – A1( $\epsilon_{ss}^* = 1$ ), red dashes – A1( $\epsilon_{ss}^* = 3$ ), blue dots – A2( $\epsilon_{ss}^* = 1$ ), green dot-dashed line – A2( $\epsilon_{ss}^* = 3$ ); higher surface potential leads to higher values; stronger charges lead to lower values, characteristic of three-dimensional crystalline state. **Panel b – series C-*h*:** black line – C1-*h*, red dashes – C2-*h*, blue dots – C3-*h*, green dot-dashed line – C4-*h*; fraction increases with charge separation, up to 0.6, characteristic of the crystalline state (charge-driven assembly). **Panel c – series D:** black line – D1, red dashes – D2, blue dots – D3, green dot-dashed line – D4; fraction increases with molecular length, tending towards unity (surface-driven assembly). **Panel d – series L:** black line – L1, red dashes – L2, blue dots – L3; transition from charge-driven to surface-driven assembly. **Panel e – series G-*h*:** black line – G1-*h*, red dashes – G2-*h*, blue dots – G3-*h*; charge-driven assembly; fraction unaffected by geometry for short tail length. . . . . 105

Figure 3.11 Complex phase diagrams in low-charge systems. In (a), the temperature dependence of  $S_2$  for systems C1-*l*, C2-*l*, C3-*l* and C4-*l* is shown with solid black, dashed red, dotted blue and dot-dashed green lines, respectively. A low temperature snapshot showing C2-*l* in antiparallel configuration is shown in (b). High ordering in antiparallel rows was observed for C3-*l* at intermediate temperatures -  $T^*=2.5$  is shown in (c). An example of the medium temperature nematic phase at  $T^*=2.5-3.5$  is given in (d) for C4-*l*/D4. . . . . 106

- Figure 3.12 Variation of order parameters between polymorphic states in early runs of **D4/C4**– $l$  system. The averages for both  $S_1$  (**panel a**) and  $S_2$  (**panel c**) exhibit high standard deviations at all temperatures, suggesting the adoption of more than one strategy for self-assembly. **Panels b** and **d**:  $S_1$  and  $S_2$  for individual snapshots at  $T^* = 1.45$ , as a function of simulation time. Both  $S_1$  and  $S_2$  fluctuate between high and low values. This indicates an initial coexistence of phases, including nematic (panel b). The long-run behaviour is that described in Figures 3.7 and 3.8. . . . . 107
- Figure 3.13 Order parameters for **G1<sub>2</sub>** –  $l$  system. **Panel a**: average  $S_2$  and associated standard deviation as a function of reduced temperature,  $T^*$ . The maximum of variation is around  $T^*=1.4$ , with lower temperature range and magnitude relative to **D4/C4**– $l$  (see Figure 3.7). **Panel b**: variation of  $S_1$  (black) and  $S_2$  (red) between a series of microstates for **D4** at  $T^*=1.4$ .  $S_1$  is close to zero and has a universally low variation, indicating the absence of nematic.  $S_2$  fluctuates randomly between a minimum and a maximum, suggesting the presence of a family of configurations with continuous variation in the long-range order. . . . . 108
- Figure 3.14 The effects of relative positioning of charged atoms within molecules. Snapshots of simulations corresponding to models **L1** at low to intermediate temperatures (a), **L3** in its antiparallel phase (b), and **L3** in a compact nematic - [panel (c)] configuration. . . . . 109
- Figure 3.15 The effect of asymmetry on self-assembled structures. Low-temperature snapshots from simulations corresponding to models **G1**– $h$ , **G1<sub>1</sub>** –  $l$  and **G3<sub>1</sub>**– $l$  are presented in panels a-c, respectively. The graph in panel d shows the fraction of molecules in contact with the surface ( $F_{mol}$ ) for the **G**– $h$  series (above) and the **G**– $l$  series (below), as a function of reduced temperature,  $T^*$ . Black solid lines, red dashes and blue dots correspond to systems **G1**, **G2** and **G3**, respectively. . . . . 111
- Figure 3.16 Tail-driven assembly in asymmetric molecules. Low-temperature snapshots from simulations corresponding to models **G1<sub>3</sub>** –  $l$  and **G3<sub>4</sub>** –  $l$  are presented in panels a and b, respectively. . . . . 112

- Figure 3.17 Effects of tail length for the SA of molecules  $G1-l$ : **Panels a and b**: averages  $S_1$  and  $S_2$  respectively, as a function of reduced temperature,  $T^*$  (x axis), when the tail is one (black line), two (red dashes), three (blue dots) and four (green dot-dashed lines) atoms long. Increasing tail length has no influence over  $S_1$  (panel a) and has a slight influence on  $S_2$  (panel b) by stabilizing the head-group driven antiparallel monolayer over longer distances, as can be observed in **panel c** for  $G1_2-l$ . Further tail increase causes layers to stack due to a high surface density. The head-group driven antiparallel assembly is preserved across layers (**panel d** -  $G1_3-l$ ) . . . . . 113
- Figure 3.18 Comparison of  $S_1$  (**panel a**) and  $S_2$  (**panel b**) between  $G2_1-l$  and  $G2_3-l$ . The x axis represents reduced temperature,  $T^*$ . G2 geometry is biaxial, leading to two versions of each parameter, one relative to the head group and one relative to the tail group:  $G2_1-l$  head - black line,  $G2_1-l$  tail - red dashes,  $G2_3-l$  head - blue dots,  $G2_3-l$  tail - green dot-dashed line. G2-l series is characterized by low order parameters in general. A longer tail size facilitates a stronger alignment among molecules, particularly along the tail axis. . . . . 114
- Figure 3.19 Comparison of  $S_1$  (**panel a**) and  $S_2$  (**panel b**) between  $G3_1-l$  and  $G3_4-l$ . The x axis represents reduced temperature,  $T^*$ . G3 geometry is biaxial, leading to two versions of each parameter, one relative to the head group and one relative to the tail group, with the two axes being perpendicular to each other:  $G3_1-l$  head - black line,  $G3_1-l$  tail - red dashes,  $G3_4-l$  head - blue dots,  $G3_4-l$  tail - green dot-dashed line.  $S_1$  and  $S_2$  for G3-l behave similar to those for G2-l. . . . . 114
- Figure 4.1 Optimized structures of PAHs series investigated. Top row from left to right: benzene ([1]H), naphthalene ([2]H), phenantrene ([3]H), tetrahelicene ([4]H). Bottom row from left to right: pentahelicene ([5]H), hexahelicene ([6]H), heptahelicene ([7]H), coronene ([6]C). . . . . 123

- Figure 4.2 Geometry of the three coinage metal surfaces optimized as three-layered 2D sheets. The bottom layer was kept frozen during optimization in order to enforce the [111] symmetry. Panel **(a)**: Cu[111]. Panel **(b)**: Ag[111]. Panel **(c)**: Au[111]. . . . . 124
- Figure 4.3 Optimized structures of [6]C and helicene gas phase monomers. Top projections of [6]C, (M)-[5]H, (M)-[6]H and (M)-[7]H are shown in panels **(a)**–**(d)**, respectively. Panels **(e)**–**(h)** show side projections of (M)-[5]H, (M)-[6]H, (P)-[6]H and (M)-[7]H, respectively. . . . . 125
- Figure 4.4 Gas phase optimized structures of [6]C and helicene stacked dimers. Panels **(a)** and **(b)** show the top and side projections of [6]C[6]C, respectively. Panels **(c)**–**(h)** show side projections of like-[5]H[5]H, unlike-[5]H[5]H, like-[6]H[6]H, unlike-[6]H[6]H, like-[7]H[7]H and unlike-[7]H[7]H, respectively. . . . . 126
- Figure 4.5 Ring ordering for [6]H on Cu. Panels **(a)** and **(b)** show the top and side view of adsorbed molecule, with the ring order displayed. Numbering begins from the ring closest to the surface, in ascending order. Panel **(b)** also shows diameters  $\mathbf{d}_l$  and  $\mathbf{d}_h$ , defined as the distance between diametrically opposed carbon atoms relative to the helix axis. Two diameters can be defined: one is the distance between the middle ring (Ring 4 in [6]H) and the ring closest to the surface (Ring 1), and the second is formed by the middle ring and the ring furthest from the surface (Ring 6 in [6]H). In the following pages, the projections of these diameters onto the surface plane are reported, referred to simply as “diameters”. . . . . 127

- Figure 4.6 Molecular twist profiles for helicenes: [5]H – red, [6]H – green, [7]H – blue. Ortho-fused rings are numbered from 1–7 ( $i$ ), starting at one end of the molecule. On a substrate, numbering begins with the ring closest to the underlying surface. The solid angle  $\Theta_i$  between ring  $i$  and the reference plane is displayed. The reference plane for all panels is Ring 1, with the exception of panel (c), where the surface plane is the reference. Panel (a) shows data for the gas phase monomer, whereas (b) and (c) display the monomer adsorbed on Cu, with the Ring 1 and the surface reference planes, respectively. . . . . 128
- Figure 4.7 Molecular twist profiles for helicenes: [5]H – red, [6]H – green, [7]H – blue. Ortho-fused rings are numbered from 1–7 ( $i$ ), starting at one end of the molecule. Numbering begins with the ring closest to the underlying surface. The solid angle  $\Theta_i$  between ring  $i$  and the reference plane is displayed. Panels (a) and (b) correspond to the surface-bound molecule for Cu-adsorbed like and unlike dimers, respectively. Panel (c) shows the surface-bound molecule for a like dimer adsorbed on Ag. . . . . 129
- Figure 4.8 [6]C on (111) coinage metal surfaces. Panels (a) and (b) show the [6]C monomer on Cu and Au, respectively. Panels (c) and (d) present dimer geometries on Cu and Au, respectively. . . . 134
- Figure 4.9 [6]C dimer on (111) coinage metal surfaces. Panels (a) and (b) present the geometries on Cu and Au, respectively. . . . . 135
- Figure 4.10 The behaviour of the projected density of states (PDOS) upon adsorption of coronene. Figures (a) and (b) correspond to the [6]C–Cu and [6]C–Ag adsorption, respectively. In top panels, gas phase  $2p_z$  states are presented for the monomer (black lines) and the dimer (red lines). The  $d$  (green lines) and  $sp$  (blue triangles) bands of the metal surfaces before adsorption are shown in the middle panels. Bottom panels include data for the adsorbed monomer complex: black lines are the  $2p_z$  states of the adsorbate, green lines are the  $d$  bands of the metals, and blue triangles show the  $sp$  bands of the metal. . . . . 136

- Figure 4.11 The behaviour of the projected density of states (PDOS) upon adsorption of [6]H on Cu. Top panel: gas phase  $2p_z$  states for the monomer (black lines) and the dimer (red lines), with states for the L dimer (red) and the U dimer (purple circles) given for the chiral [6]H molecules. Middle panel: the  $d$  (green lines) and  $sp$  (blue triangles) bands of the metal surfaces before adsorption. Bottom panel: data for the adsorbed monomer complex; black lines are the  $2p_z$  states of the adsorbate, green lines are the  $d$  bands of the metals, and blue triangles show the  $sp$  bands of the metal. . . . . 137
- Figure 4.12 Helicene monomers on the Cu(111) surface. [5]H is shown in panels (a)–(c), [6]H is shown in panels (d)–(f) and [7]H is shown in panels (g)–(i). Both top [panels (a), (d) and (g)] and side [panels (b), (e) and (h)] views are presented. The bottom row [panels (c), (f) and (i)] shows the superimposed structures of the respective molecule in the absence (blue) and the presence (gray) of the surface. . . . . 142
- Figure 4.13 Structures of pentahelicene stacked dimers on Cu(111). Panels (a) and (b) show the L–[5]H[5]H and U–[5]H[5]H dimers, respectively. Panels (c) and (d) show the superimposed structures of the surface-adsorbed molecule in the absence (light green) and the presence (gray) of the stacked molecule, for the respective systems. . . . . 145
- Figure 4.14 Structures of heptahelicene stacked dimers on Cu(111). Panels (a) and (b) show the L–[7]H[7]H and U–[7]H[7]H dimers, respectively. Panels (c) and (d) show the superimposed structures of the surface-adsorbed molecule in the absence (light green) and the presence (gray) of the stacked molecule, for the respective systems. . . . . 147

- Figure 4.15 Low ( $d_l$ ) and high ( $d_h$ ) diameters as a description of molecular footprint, relative to the coronene diameter. Panels (a) and (b) show low and high diameters for [5]H, respectively. The values for [6]H are shown in panels (c) and (d) and those for [7]H are shown in panels (e) and (f), respectively. Red, green and blue bars are used in each plot to denote a molecule in monomer, like dimer and unlike dimer structures, respectively. . . . . 148
- Figure 5.1 Visual representations of helicenes in model  $D$ , after the first simplification step. Each carbon atom in the original model is assigned one Lennard-Jones sphere, whose parameters include contributions from the outer hydrogen atoms. Each benzene ring is represented in this way by a hexagonally-arranged set of six Lennard-Jones spheres. [4]H: panels (a)–(d), [5]H: panels (e)–(h), [6]H: panels (i)–(l), [7]H: panels (m)–(p). For a better representation of the three-dimensional structure, each molecule is shown both in a top and a side view. Molecules coloured in purple represent the optical isomers corresponding to the yellow coloured counterparts. . . . . 152
- Figure 5.2 Visual representations of benzene ([1]H: (a),(b)), and coronene ([6]C: (c),(d)) in model  $D$ . Each carbon atom is assigned one Lennard-Jones sphere, whose parameters include the contributions from hydrogen atoms. Molecules are shown in a top and a side view. . . . . 153
- Figure 5.3 Visual representations of the molecular models within the  $S$  model. Each six-atom conjugated ring is represented by one LJ sphere. [4]H: panels (a)–(d), [5]H: panels (e)–(h), [6]H: panels (i)–(l), [7]H: panels (m)–(p). For a better representation of the three-dimensional structure, each molecule is shown both in a top and a side view. Molecules coloured in purple represent the optical isomers corresponding to the yellow coloured counterparts. . . . . 155

- Figure 5.4 Common examples of monolayer formation in [4]H – panel (a), [5]H – panel (b), [6]H – panel (c) and [7]H – panel (d). The display in the top-right corner of each panel can be used to estimate the molecular footprint of each homologue and the commensurability of the monolayer. All systems are enantio-pure. Colouring differences between molecules are intended to aid visualization and correspond to identical geometries. . . . . 158
- Figure 5.5 Monolayer adsorption of [4]H racemic (panels (a), (b)) and 3D aggregation of [7]H racemic (panels (c), (d)) on large surfaces. The planar character of [4]H molecules promotes flat adsorption, while the pronounced out of plane twist present in [7]H molecules facilitates interactions in three dimensions. Yellow and purple colours denote the two enantiomers. . . . . 160
- Figure 5.6 Comparison of adsorbed phases formed by highly-twisted helicenes in the presence of strong (left panels) versus weak (right panels) surface potential. Top panels: adsorbed racemic [6]H on a weak-binding (panel (a)) and a strong-binding (panel (b)) surface, respectively. Bottom panels: adsorbed enantiopure [7]H on a surface with weak (panel (c)) and strong (panel(d)) surface potential, respectively. The degree of orientational alignment between neighbouring molecules is slightly enhanced by a weaker potential. . . . . 162
- Figure 5.7 Close-up views of the second-layer nucleation centers for helicenes adsorbing beyond the saturated monolayer: panel (a) – racemic [4]H; panel (b) – enantiopure [5]H; panel (c) – racemic [6]H and panel (d) – enantiopure [7]H. The six-folded structure of the trimer is apparent in all homologues, irrespective of the chiral composition of the system. The degree of staggering between the first and second layer varies between homologues. The corresponding side views of these structures are displayed in Figure 5.8. . . . . 163

Figure 5.8 Side views of the second-layer nucleation centers, complementing the top views displayed in Figure 5.7: panel (a) – racemic [4]H; panel (b) – enantiopure [5]H; panel (c) – racemic [6]H and panel (d) – enantiopure [7]H. The strong layered character of the structure in all homologues is apparent. . . . .	164
Figure 5.9 Side views of second- and multi-layered formations built on top of the monolayer foundations by adsorbed helicenes: panel (a) – enantiopure [4]H; panel (b) – racemic [5]H; panel (c) – racemic [6]H and panel (d) – enantiopure [7]H. The strong layered characters appears to be preserved vertically across several layers when present. The presence of the layered structure appears independent of the racemic composition. . . . .	165
Figure 5.10 Adsorbed rows of enantiopure [7]H on surfaces with strong (left panels) and weak (right panels) potential. The bottom panels show a close-up view of the respective rows. Molecules within rows appear to be tilted and stabilized by inter-molecular end ring interactions. Rows adsorb along certain symmetry axes only, a surface effect which inhibits 3D aggregation. . . . .	167
Figure 5.11 Multi-layered formations of racemic [7]H on weak-binding surface. Molecules appear to stack in enantiopure columns, dispersed across the surface in an alternate or random manner. Panel (b) displays the close-up side view of the stacked columns in panel (a). . . . .	168
Figure A.1 Basic layout of a SIESTA job directory. . . . .	180
Figure A.2 Schematic representation of the workflow required to obtain the binding energy of benzene (1H) on gold with SIESTA. . . . .	187
Figure A.3 Schematic representation of the workflow required to obtain the counterpoise-corrected interaction energy of a two-component system with SIESTA. . . . .	191

# Chapter 1

## Introduction

### 1.1 The Importance of Surface Self-Assembly

#### 1.1.1 Complex Materials

The rapid development of materials and devices with specific, highly tunable and functional features is one fundamental metric of societal development. Whether their applications touch upon optimizing energy capturing and storage, increasing the accuracy of medical diagnosis and targeting in drug delivery, decreasing the cost of manufacturing industrial products or enhancing the efficacy of environmental cleanup, the development of materials with specific properties and controllable response to external stimuli is of central importance in industry and manufacturing. These highly desirable properties emerge as a consequence of the physicochemical complexity they encode. A large part of the current research effort in complex materials is dedicated to understanding and controlling this complexity at the molecular level. If highly complex materials are to be manufactured at large scale, in a sustainable and efficient manner, the advantages of control science will have to be materialized, and bottom-up methods of development, such as molecular self-assembly will have to be employed.

#### **Energy materials**

Although in recent times energy production has been dominated by the use of fossil fuels, energy technologies that are viewed as promising alternatives today were first developed as early as the nineteenth century. The first photovoltaic cell, the first

electric motor-driven car, the first rechargeable lead-acid battery were products of the mid-nineteenth century industrial revolution. Their realignment with the forefront of scientific research coincided both with society's need to decrease its reliance on fossil fuels, but also with the new experimental and computational capabilities in the development of functional materials. Functional materials for energy applications fall under a broad umbrella – from filtration materials for fossil fuel effluents, non-metallic catalysts for catalytic converters and fuel cells, to less expensive, higher-efficiency materials for solar cells.<sup>1;2</sup>

## **Metal-organic materials**

Organo-nanocomposites and nanostructured metal-organic materials, including thin organic films adsorbed onto metallic surfaces, are promising candidates for a variety of complex materials. The desirable properties of these mixed materials are a consequence of their hybrid nature: whereas inorganic materials (metals, metal oxides, ceramics) have highly specific and predictable opto-electronic and mechanical properties, organic-based materials can have low production costs, highly tunable response properties, flexibility and versatility. The behaviour of the mixed material is thus a complex combination of the individual components' properties, compounded by the interactions at the molecular and supramolecular level, between those components.

### **1.1.2 Self-Assembly**

#### **The need for theory**

Since heterogeneity is the key to the complex nano-features of these materials, two issues of great importance arise when one considers their manufacturing process: the synthesis of precursors and their subsequent assembly into final materials. Both stages of production require considerations of energy cost and sustainability, economic and environmental. Much scientific effort had been invested, over the past half century, into lowering these costs for molecular synthesis, with tremendous success. The process of assembly, however, is still in its infancy and much less optimized across all these metrics, often leading to considerable time, energy and financial costs during manufacturing, if manufacturing is at all feasible. Given this current state of development, research into the process of molecular assembly is needed. Current popular methods of manufacturing involve a high degree of top-down level control, which

is both energy-costly and hard to maintain as heterogeneity increases. Discovering and exploiting the principles by which targeted organization can be achieved reliably and in an energy-efficient manner is one of the holy grails of contemporary materials science.

### **Molecular self-assembly**

Among bottom-up fabrication methods, molecular self-assembly (SA) is the spontaneous aggregation of molecules or particles into structured, stable, noncovalently joined aggregates under equilibrium conditions.<sup>3</sup> Unlike the more general case of molecular organization, in which the the properties of the system depend on sustained forces applied by external sources, SA relies on the intrinsic information contained within the building blocks themselves to generate phases with stable properties over time. The manufacturing of materials with desired functionality, stable over time and with a low maintenance cost, is the principle interest driving the development of molecular SA.<sup>4;5</sup> From a research perspective molecular SA involves discovering of rules and principles by which molecules spontaneously organize according to their physicochemical features, developing design methods for the right building blocks and finding optimal assembly environments.

### **A problem spanning multiple scales**

SA is a broad and complex topic, breaking multiple bounds between classical disciplines such as chemistry, physics, biology, material science and even computational science. The knowledge and intuition of chemists about the interaction, reactivity and stability of matter can be put to great use in uncovering the principles governing SA. From a methodological point of view SA is a bulk process informed by molecular-level interactions, and thus requires taking into consideration processes spanning multiple length scales, as well as how such processes influence one another. Chemical principles and techniques can be used to determine how the geometry, configuration, reactivity and other chemical properties of molecules determine causally the ways in which they organize locally. From the other side of the scale spectrum, the reach and distribution of the mesoscopic phases are directly influenced by the order, structure and stability they offer over long ranges. In order to predict the outcome of SA with any practical level of detail, long-scale simulations with accurate local electronic interactions must be used, which, with the current level of computational technology is unfeasible.

## **The role of computational chemistry**

Given the computational constraint, a scale-reductionist approach is required, whereby the sequential focus on various scales leads to principles transferable to the others. Within this framework, the theoretical treatment of the relevant aspects of SA requires a combination of quantum-level calculations and large-scale classical simulations. Equally important is the development of tools for mesoscale exploration, connecting the limited quantum scale information with the multitude of outcomes accessible at the macroscale. At each scale the system evolves on a complex potential energy surface, with details from one level directing the evolution on the others. The scale-reductionist approach is effective in addressing the stiffness of the problem if the computational chemist succeeds in transferring data across scales without the loss of information crucial to an accurate representation of the system.

### **1.1.3 Surface Self-Assembly**

The subfield of surface self-assembly focuses on formation and stability of distinct clusters or layers in the presence of a solid surface. The presence of a surface changes an already complex energy landscape, by introducing at least two additional factors: the relative competition between particle-particle and particle-surface interactions, and the geometry of the surface, imposed through preferential adsorption sites. Both distort the clean process of SA in non-trivial ways, but they also restrict molecular SA to two dimensions and to patterns imposed by the surface symmetry. Thus, a surface acts as a template for certain phases and targets SA to two dimensions instead of three. By encouraging the formation of distinct stacked layers, a surface may lead to an anisotropic SA product, easier to be analyzed by parts and having more reliable responses to external stimuli than their three dimensional counterparts.<sup>6</sup>

## **Applications**

The behaviour of small to intermediate size clusters of particles on surfaces is relevant in many phenomena. Surface SA grew and matured as a field primarily due to interests in applications for surface functionalization.<sup>7-16</sup> In the case of surface protection<sup>17;18</sup> the ways in which monolayers self-assemble upon adsorption affects the quality of the coating. Pattern formation, size and reproducibility in dip-pen nanolithography,<sup>19;20</sup> are determined by the competition between kinetic and thermo-

dynamic factors influencing ink molecules attaching to the surface. Certain prochiral molecules form enantio-specific domains upon surface adsorption,<sup>21;22</sup> a phenomenon which is determined by surface SA principles of tremendous interest in chiral separation. Most proteins are known to change conformation upon adsorption, and an understanding of how the surface affects the optimal conformation helps minimize denaturation. The subfield had received further significant attention in the past two decades, with the growing interest in device miniaturization, fabrication of organic functional materials,<sup>23</sup> solar cells<sup>24</sup> biochemical sensing<sup>25</sup> and drug delivery.<sup>22</sup>

### **The role of chirality**

The mechanisms by which chiral molecules self-assemble on solid surfaces are of great importance in understanding how nanocrystals form superlattices fit for the development of devices with chiroptical and optoelectronic properties.<sup>26</sup> Understanding such mechanisms at the fundamental, molecular level opens the door for directed self-assembly and allows for the creation of devices with a high degree of reproducibility and low cost. Helicenes are at the heart of this pursuit as they possess both relatively stable molecular chirality and the ability to form highly ordered self-assembled monolayers on solid substrates, in particular on symmetric surfaces of transition metals. The role of stacking in the process of helicene binding to highly symmetric gold, silver and copper (coinage metals) surfaces is thus of great importance to SA.

### **Helicene-functionalized surfaces**

Surface functionalization with conjugated chiral aromatics, and particular helicenes, can lead to new materials with enhanced opto-electronic properties, as well as to advancements in chiral separation. The addition of functional groups onto the aromatic backbone can be used in surface science to direct self-assembly,<sup>27-30</sup> to modulate mobility on solid substrates<sup>27;31</sup> and to enhance or suppress enantiodiscrimination.<sup>32</sup> The virtually unending degrees of modulation and control of both intermolecular and surface-molecular interactions currently make chiral aromatics, including helicenes, a promising class of molecules in both the theoretical the and practical realm of surface functionalization.

## 1.2 Background on Surface Self-Assembly

### 1.2.1 Self-Assembled Monolayers

#### Alkanethiolates

The earliest and most well studied SA systems are those of self-assembled monolayers (SAMs) of alkanethiolates, which are not only historically relevant but representative of the scope and evolution of the field as a whole. Having been the focus of attention for many decades, SAMs offered an early glimpse into the potential of SA to manufacturing and molecular control, by demonstrating highly reproducible behaviour.<sup>33;34</sup> The patterns observed in SAMs have been linked to molecular properties of the building blocks, such as their tilt angles, side chains and chain length. Examples include the observation that longer chain lengths generally yield more stable SAMs, or the odd-even effect, where the addition of one carbon atom within the main chain of the molecule can dramatically alter SA.<sup>33;35</sup> These effects have been shown to directly influence properties of great interest for molecular engineering, such as molecular surface area, phase stability and level of molecular packing. For example, surface protection requires the formation of a highly stable, compact monolayer,<sup>17;18</sup> whereas in the manufacturing of molecular switches, a balance between molecular flexibility and structural stability has to be attained.<sup>36</sup> Similarly, in the case of surface functionalization, the stability of the monolayer has to be enhanced without a loss of functional activity.

#### Substituted alkanethiolates

Substitution of alkanethiol tails can influence and even direct SA,<sup>34;37;38</sup> to control molecular surface area, phase stability, as well as the chemical and physical behaviour of the surface.<sup>8;10;13;15;38-40</sup> This adds another design dimension to surface SA, as SAMs can be tailored to control for molecular surface area and phase stability, making it possible to immobilize desirable molecules with the right thiolate backbone in a way to enhance functionality.<sup>38</sup> One example is the fabrication of molecular switches,<sup>36;41;42</sup> where the molecule has the ability to adopt one distinct geometry out of two or more options available to it. In the presence of an external stimulus the assembly might adopt a different phase as the molecules switch between geometries (conformers). Whereas thiolate backbones can form stable SAMs by close packing, this can destabilize one of the two phases and inhibit the switching process.<sup>43</sup> In such cases,

the preparation of the monolayer in the less dense phase is required for the proper functioning of the switch.<sup>44</sup> On the other hand, weaker lateral interactions in the low-density monolayer can lead to enhanced configurational freedom and a higher degree of disorder in the molecular backbones.

### **Beyond alkanethiolates**

Building blocks more complex than alkanethiolates show an even greater potential for design, as is the case with chiral molecules conferring chirality and chiral recognition to the surface,<sup>22</sup> or with molecules capable of switching between conformers given the right stimulus. By unravelling the interplay between fundamental molecular aspects of building blocks leading to SAM formation, a more targeted approach towards maximally exploiting these properties can be achieved.

### **1.2.2 Aromatic Compounds on Coinage Metals**

Much of the appeal alkanethiolates have within SA research stems from their primarily linear geometry, which leads to relatively simple SA principles. Building blocks with a geometry defined across two, or even three, dimensions rank up the complexity of the principles governing organization but open the door to a vast array of patterns previously unexplored. Aromatic compounds, whether planar or not, are possibly the simplest example of such a class of building blocks. Much like alkanethiolates, poly-aromatic hydrocarbons (PAHs) serve as primary examples of SA precursors.

#### **Benzene**

The adsorption of aromatic compounds on coinage metals has been thoroughly studied theoretically. In the simplest case, that of the adsorption of benzene, there are four different high-symmetry sites (atop, bridge, face-centered cubic, hexagonal close-packed) each with two possible orientations ( $0^\circ$ ,  $30^\circ$ ) on [111] surfaces. No significant difference among these exists and there seems to be a lack of deep wells on the potential energy surface (PES) and a flat geometry.<sup>45</sup> Benzene molecules diffuse freely over high symmetry surfaces. Larger aromatic compounds, such as diindenoperylene, adopt a more or less flat geometry.<sup>46</sup> On Cu[111] however, larger asymmetric molecules display a bending of the aromatic plane in the longer direction, with the center of the molecule closest to the surface, characteristic of chemical bonding.

For Ag[111] and Au[111] adsorption Van der Waals attraction dominates<sup>47</sup> and the molecule retains the planar structure, similar to the case of benzene.

### **Prochiral and chiral molecules**

The notion of chiral symmetry breaking, central to modern day chemical and pharmaceutical research, has a special relevance in surface SA. Many molecules that are achiral in solution become chiral when deposited on a solid surface, and often undergo assembly into chiral supramolecular structures. Moreover, chiral molecules unable to undergo 3D enantiospecific crystallization may be able to enantio-segregate, by forming small homochiral clusters and even extended homochiral domains upon surface SA. Such assemblies have several applications in energy materials: as blueprints for the growth of helical organic nanotubes with elevated non-linear optical properties,<sup>48–50</sup> as sites or modifiers for heterogeneous chiral catalysis,<sup>51–54</sup> or as building blocks for molecular motors.<sup>55–57</sup> Restriction of the assembly process to two dimensions selects a part of the potential energy surface on which the recognition process occurs, thus chiral specificity is more easily attained. By understanding the complex effects leading to pattern formation and chirality transfer, a more targeted approach towards chiral structure formation may be achieved.

### **1.2.3 Helicenes on Metals**

#### **The relevance of helicenes**

Helicenes and their derivatives have been studied as components of polarized light emitting diodes,<sup>58</sup> fluorescent dyes with moderate quantum yields,<sup>59;60</sup> molecular junctions controlled through stretching and compressing,<sup>61</sup> chiroptical redox switches,<sup>62–64</sup> circularly polarized light detectors,<sup>65</sup> multifunctional pH-switchable chemical systems,<sup>66</sup> chemical detectors,<sup>67</sup> devices with size-dependent specific optical rotation,<sup>68</sup> helicene-metal semiconductor hybrids,<sup>69</sup> and Landauer conduction channels.<sup>69;70</sup> Their chiral polyaromatic skeleton makes them fit for carrying out asymmetric and enantioselective reactions<sup>71</sup>, including ones with biological relevance.<sup>72</sup> Helicenes are of high interest in supramolecular chemistry, both practically and theoretically, due to their ability to transfer chirality to supramolecular systems through directed self-assembly. They are good candidates for molecular building blocks when chiral control is a priority, be it one-dimensional,<sup>73;74</sup> two-dimensional,<sup>75</sup> or three-dimensional self-assembly.

## Self-assembly on coinage metals

The mechanism for surface-adsorbed layer formation and chirality transfer in helicene adsorption has been characterized extensively for various helicenes and surfaces with differing symmetries. Ernst and collaborators examined the self-assembly of helicenes on several metallic substrates. On Cu(111), racemic mixtures of [7]H resolve into units with six-fold (hexamer) or three-fold (trimer) symmetry,<sup>75;76</sup> which repeat along parallel rows arranged in mirror-image domains.<sup>77</sup> The enantiomorphism is due to these rows of racemic composition, having an oblique homochiral alignment with the surface, reflecting chiral discrimination on a mesoscopic scale.<sup>78</sup> Chirality is manifested throughout multiple layers, each growing quasi-epitaxially on top of the chiral monolayer, with small lattice mismatches.<sup>79</sup> The mismatch differs along the two adsorbate lattice vectors, transferring orientation, and therefore chirality, to higher thin films, vertically.

## Substrate influence

Chirality in adsorbed phases of helicenes is strongly influenced by the choice of metallic substrate. No enantiomorphism is observed on Ni(111) in spite of the metallic crystal's similarity with Cu(111). This is likely due to a lack of mobility of molecules on the surface, as Ni is stronger binding.<sup>80</sup> On the other hand, weaker-binding surfaces, such as Ag(111) and Au(111), exhibit locally repeating units of heptahelicene similar to those observed on Cu(111), but the long range enantiomorphism is absent.<sup>81</sup> The disruption of long range order reflects numerous discontinuities caused by the stress building up in the racemic row.

## Molecular size influence

The manifestation of chirality in multilayer deposition of racemic mixtures is highly dependent on the identity of the adsorbate. Whereas submonolayers of rac-[7]H form homochiral mirror-image domains on Cu(111),<sup>77</sup> several chirality transitions are observed at the deposition of further layers, from heterochiral zigzag M-P rows at  $1 < \theta_{rel} < 1.61$  (where  $\theta_{rel}$  is coverage relative to a monolayer), to homochiral alternating layers at higher coverages.<sup>82</sup> In contrast, [5]H organizes invariably into homochiral dimers, which then form patterns dependent on surface coverage: at  $\theta_{rel} < 0.7$  a racemic “checkerboard phase” is observed; the second layer forms at  $\theta_{rel} > 0.7$ , and the multilayer arrangement is a homochiral “honeycomb” pattern.<sup>83</sup>

## 1.3 Theoretical Challenges for Surface Self-Assembly

### 1.3.1 The Challenge of Multiple Scales

#### The drive towards molecular control

A sizable part of the current research effort in complex materials manufacturing is dedicated to understanding how to control matter at the molecular level. Observational science has brought us materials for solar cells with over 20% efficiency and fuel cell materials for first-generation vehicles that are in some respects not yet competitive with electric alternatives.<sup>1;84</sup> Advances in real-time observation of chemical reactions and the atomic-level layout of materials, as well as the ability to approach computationally length and time scales relevant to experiments, have given rise to a new approach to materials design: the creation and manipulation of building blocks designed to come together in predetermined ways, leading to the structural and property control of the resulting materials. Advanced theoretical and computational modelling of the self-assembly behaviour of molecular building blocks is essential to achieve this level of understanding and control over the relevant physical and chemical processes.

#### Towards bottom-up design

As materials become more complex at ever-smaller scales, classical top-down methods of manufacturing can become more unreliable and energy-inefficient. The alternative to top-down design and manufacturing are bottom-up development and self-assembly.<sup>4;26;85-87</sup> Relying on local intermolecular interactions to drive manufacturing processes instead of strong external stimuli leads to less energy use on the production side. Having self-assembled materials spontaneously adopting a desired state reduces the wear-and-tear factor and lowers maintenance cost. If highly complex materials are to be manufactured at large scale, in a sustainable and efficient manner, bottom-up methods of development of self-assembled materials will have to be established.

Locally, the structure of the emerging organized pattern is determined by a complex interplay between dipolar alignment, steric effects and electronic structure. Over large scales, multibody and entropic effects can also be important contributors in structure formation, in addition to pair interactions.

### 1.3.2 The Challenge of Accurate, Large-Scale Simulations

The short-scale features that drive local intermolecular interactions are generally well understood by chemists, material scientists and theoreticians alike. The long-range ordered phase formation of a rapidly increasing number of systems is carefully documented and analyzed by experimentalists. Despite this however, the principles that determine how the long-range organization of self-assembled systems results from their local interactions are yet to be determined. Computationally the challenge comes down to the difficulty of simulating systems complex and accurate enough to display, at least qualitatively, the relevant effects at different scales, simultaneously. Capturing long-scale phenomena with atomic or molecular resolution is prohibitively expensive, if the building blocks are to be treated quantum-mechanically. Many such crucial phenomena can only emerge in systems with a minimum of thousands, if not hundreds of thousands of particles. Setting the financial cost aside, the time required to run simulations on this scale with ab-initio techniques is astronomical, with little to no dent achieved when employing efficient use of parallel computing. Over the decades, many clever techniques have been developed to reduce the computational complexity of quantum mechanical systems and bring their associated simulation cost in the feasible realm. These developments, coupled with the ever-growing power of computing technology, have pushed the limit of molecular modelling to impressive scales, and the further development of such techniques is where the future progress likely resides.

#### **Balancing accuracy and simulation cost**

Approximations designed to lower the computational complexity inevitably come at a cost of accuracy of representation. If models become too simple, the crucial features that were sought for may be lost. Striking the right balance between accuracy and cost requires the employment of a great deal of chemical intuition and creativity on the part of theoretical chemists. This balance, in the context of the challenge of scale, is what makes any theoretical investigation into the nature of SA challenging.<sup>85</sup>

### 1.3.3 The Need for Theoretical Frameworks

Besides computational considerations, the field of surface self-assembly lacks a foundational base or principles of a mature and authoritative field. The multidisciplinary

nature requires scientists to juggle a variety of concepts and models adopted from multiple, often mutually confined fields, such as solid state physics, molecular chemistry, biochemistry, and material science, to name just a few. Bringing together many concepts within a unified base of theoretical frameworks is one of the hope within surface SA. Developing this foundation would enable scientists not only the ability to explain observed phenomenon with brilliant rigour, but also the ability to predict previously unknown forms of matter organization.

### **1.3.4 The Reductionist Approach**

The straightforward way to analyze a system undergoing surface SA is by considering the substrate – surface – and the self-assembly components – building-blocks – separately. This is warranted, as often times the physicochemical properties of the surface varies considerably from those of the building blocks. As a result, interactions among building blocks can easily be decoupled from the interactions with the surface, at least in theory.

#### **The challenge of building-block design**

Even for the simplest molecules, the energy landscapes of multi-body systems are complex, with multiple metastable states separated by high energy barriers. The subtle interplay of intermolecular forces causes such systems to give rise to unpredictable patterns.<sup>4;26;88-96</sup> Linking molecular properties of building blocks to the SA patterns produced is currently of great interest for the purpose of rational design of materials for energy applications. Attaining a balance between molecular flexibility and structural stability is a fundamental aspect of molecular engineering. One way to achieve this balance is by considering the fundamental forces that drive SA at the molecular level.

#### **The challenge of substrate choice**

When the SA process occurs at the solid-gas interface, the surface itself modifies a complex energy landscape, essentially reducing a three-dimensional phenomenon to a two-dimensional one. This is not to imply that either the complexity of the problem or the array of patterns observed is reduced with the addition of a surface. Additional variables come into play: a competition between interparticle and particle-

surface interactions, as well as the geometry of the surface, felt by the self-assembling molecules as preferential adsorption sites and configurations.

### **The challenge of pattern description**

The interparticle/particle-surface competition and the geometry of the surface influence the process of molecular SA in non-trivial ways. Strongly attractive solid substrates favour the formation of two-dimensional, monolayer SA patterns, with geometries strongly influenced or templated by surface symmetry. Experimental work revealed SA patterns on surfaces ranging from one-dimensional arrays<sup>90;97;98</sup> to cyclic networks,<sup>91</sup> to pentagonal and hexagonal aggregates.<sup>3;92;99-101</sup> The specifics differ from case to case, but a few factors appear to be essential in determining the outcomes of surface patterning: the size and shape of building blocks, the surface geometry and its geometric compatibility with the adsorbate, and intermolecular interactions. The latter are often a combination of dispersive effects from alkylthiol chains,  $\pi$ -stacking interactions and molecular dipole moments created by polar groups and heteroatoms.

One approach to tackling this complexity at a fundamental level is by examining the various contributions to the free energy of self-assembly, using statistical approaches.<sup>102-105</sup> Distinct formalisms have been developed for supramolecular and block copolymer bulk self-assembly,<sup>104</sup> as well as surface-bound molecules.<sup>105</sup> Thermodynamic and kinetic aspects of the process have been examined in an approximate or qualitative fashion, including the role of entropy,<sup>104;106</sup> polymorphism,<sup>106</sup> pair interactions<sup>105</sup> and cooperativity.<sup>107;108</sup> Molecular features were described using approximate partition functions and equilibrium constants,<sup>105;109</sup> and parallels were drawn to instances where designed self-assembly was achieved experimentally.<sup>104;106</sup> These studies have made significant inroads in the formal understanding of self-assembly processes, at a qualitative or semiquantitative level, but research still has to bridge the length and complexity scales between this level of theory and the variable space in which experiments take place.

### **The challenge of pattern prediction**

Given the complexity with which the many variables interact with one another, predicting what patterns one might observe in any particular system is a daunting task.

It is known that strongly polar diatomic molecules form three-dimensional crystals in the absence of a substrate. This tendency can be disrupted upon assembly on a solid

surface, and various other types of order may arise. Depending on the relative strength of the interactions, disordered phases, two-dimensional crystals commensurate to the surface, and unmodified crystals were observed upon adsorption of dipolar molecules. Steric features, in the form of a longer backbone or substituents external to the polar pair, lead to even richer phase diagrams. The formation of two-dimensional phases with nematic (parallel) or antiparallel alignment is accomplished by altering the polarity of the end groups on needle-like molecules, whereas embedded charged groups make two-dimensional structures unstable for even very long molecules. Some molecules prefer to self-assemble in long, often desorbed, molecular wires.

Given the subtle interplay of intermolecular forces, relatively simple systems, such as small organic molecules on metal surfaces, can give rise to unpredictable patterns.

### **The challenge of pattern control**

All these observations can potentially be used for the development of tools to direct the SA process, provided a thorough understanding of the underlying chemistry and physics is developed.

A change in one parameter of the system, such as the symmetry of the surface, the position of a functional group, or annealing temperature, can have a dramatic effect on SA. More weakly attractive substrates can template the formation of distinct stacked layers. These layered materials are often easier to characterize due to their solid support, and can have more reliable responses to external stimuli than their three dimensional counterparts, because of the template-induced order in the material.<sup>6;26;88;89;110;111</sup>

Despite significant effort<sup>6;91;104</sup> towards the development of guiding principles, both general and specific, in molecular and surface SA, much is still left to understand. Theoretical approaches have attempted to unravel this complex phenomenon by individually addressing distinct length and time scales, using different levels of theory.<sup>85</sup>

## **1.4 What This Thesis Is About**

The goal of this thesis is to bring together a variety of computational tools from disparate fields such as quantum molecular chemistry, material science and statistical mechanics with the aim at simulating heterogeneous systems, comprised of covalently-

joined building blocks self-assembling on solid substrates, within the current computational constraints. We use these techniques to explore fundamental theoretical principles – how competition between local interactions give rise to long-range order – and to answer targeted questions for specific systems – how helicene adsorption and stacking influences enantioselectivity.

As this chapter aimed to demonstrate, the rapidly blooming field of surface SA is of considerable interest to the chemistry community, material science and industry. The unprecedented speed at which surface self-assembled systems are discovered calls for the development of a strong theoretical foundation for the processes involved and the computational tools that can explore the consequences of these principles. This thesis presents several contributions to the development of a fundamental understanding of self-assembly processes, in particular self-assembly at the vacuum-solid interface, using computational and theoretical materials science. A combination of simulation approaches, from quantum-based methods, to classical atomistic calculations, to mean-field approximations of bulk mixed phases are used in an attempt to model the various length scales characteristic to self-assembled pattern formation for functional materials.

Chapter 2 offers a comprehensive description of the theoretical concepts required for describing surface SA processes, as well as a detailed background on the computational techniques available for modelling the systems in question. Quantum mechanical and statistical principles are briefly introduced, followed by a presentation of the two main computational techniques used to explore classical physico-chemical systems, molecular dynamics and Monte-Carlo (MC). Emphasis is placed on the latter as we demonstrate it is the method of choice for the task at hand. In particular, we demonstrate how the Parallel Tempering formalism, which builds upon the classical MC formalism, elegantly solves the issue of entrapment that the MC techniques are known to manifest. A section on force fields describes how molecular interactions are represented computationally in classical systems and why for large-scale many-body representations the Lennard-Jones-electrostatic potential offers the best compromise between accuracy and computational cost. The section on quantum methods describes in detail how molecular electronic structure is handled computationally, and why the method of choice at this scale is Density Functional Theory (DFT). To address the shortcoming of DFT in representing dispersion interactions, a number of dispersion correction strategies, with varying accuracy and cost, are presented. The available data on their performance for aromatic compounds on metal surfaces is also

reviewed here. Finally, a detailed description of electronic structure simulations for solid state materials closes Chapter 2. The strategies employed in representing periodic systems are contrasted with those used for molecules and subsequently reconciled to provide a basis for the quantum-mechanical simulation of hybrid systems.

In Chapter 3 we begin our foray into theoretical surface SA by investigating the competition between van der Waals interactions, dipole-dipole forces and steric effects on one hand, and the competition between intermolecular and surface-molecule interactions on the other. These interactions are in a sense a first-order approximation for surface SA, as we demonstrate that they are the minimum requirements for ordered phase formations. We use a series of simple model molecules with fine-tuned interactions to investigate their adsorption on symmetric surfaces. We use the parallel tempering Monte Carlo formalism coupled with the Lennard-Jones-electrostatic potential to sample the configurational space of multi-body systems with pure composition. By tuning the molecular and surface parameters of these two- to six-atom molecules and the substrate, we report on how the interplay of parameters gives rise to ordered and complex equilibrium patterns. Simple models give rise to either highly ordered crystalline or highly disordered phases, depending on the competition between intermolecular dipole-dipole and surface-molecule interactions. Models of greater complexity in shape and interactions disrupt this simple picture in favour of a variety of emergent phases, which often parallel experimental observations. We demonstrate that, in simple model systems, the outcome of SA can be traced back to the structural and energetic components of the building blocks. We also demonstrate how the gap between theoretical models and experimental results can be bridged by adding structural and geometric features to the molecular models.

In Chapter 4 we further our computational study by focusing on short-scale interactions in a specific class of experimentally popular systems: self-assembled helicenes on symmetric copper, silver and gold surfaces. There has been significant work seeking to understand the effects of surface selection and lateral molecular interactions in helicene assembly, with particular focus given to their enantioselective behaviour. We add to this body of knowledge by investigating at the local level the role of stacking in helicene adsorption. We use DFT to determine the electronic, structural and energetic properties of adsorbed monomers and stacked dimers of a series of helicenes on Cu(111), Ag(111) and Au(111) substrates. We discover that stacking has a strong enantioselective behaviour due to the greater ability of enantiopure dimers to stabilize through  $\pi$ -stacking, and that the surface enhances the enantioselectivity of adsorbed

helicenes through the deformation of the molecular helical twist. We further demonstrate that, in adsorbed dimers, the stacked molecule acts synergistically with the surface potential to further distort the sandwiched molecule.

In Chapter 5 we maintain our focus on metal adsorbed helicene systems, but shift away from local interactions, towards long scale assembly. Using the structural and energetic results acquired in the study presented in Chapter 4, we construct a series of coarse-grained helicene models and explore their multi-body self-assembly behaviour with the parallel tempering Monte Carlo technique examined in Chapter 3. The coarse-graining procedure is applied in two steps, from the DFT data to a classical detailed model, to a simple model where planar information about individual benzene rings is lost. For reasons having to do with computational limitations, the final helicene models are reduced to a level of detail comparable to that of molecules employed in Chapter 3. The adsorption strategies observed are discussed, including surface-driven monolayer formation, second layer nucleation, stacking and tilted rows. We also comment on the possibility of enantioselective behaviour within our systems and on the loss of information during model reduction.

In the final chapter we take a broad view of our results and the challenges faced by someone who approaches surface self-assembly with computer simulations. We demonstrate the tension between the need of accuracy and that of computational efficiency in our own example of adsorbed helicenes. We offer some possible research pursuits that build upon the research in this work.

## Chapter 2

# Computational Methods for Studying Surface Self-Assembly

### 2.1 Introduction

#### **Length scales and interactions in surface self-assembly**

Self-assembly at solid surfaces is a bulk process informed by molecular-level interactions and substrate-level order. Locally, molecule-molecule interactions are determined by physico-chemical molecular properties such as headgroup chemistry, backbone geometry and symmetry and molecule-substrate interactions depend on substrate affinity molecular footprint. On a long scale, substrate structure effects and layer-layer alignment define the self-assembly outcome, the types of phases that emerge, their periodicity, geometry, stability and subsequently the properties of self-assembled materials.

Rather independently of scale, a noteworthy feature of surface self-assembled systems is that they often include covalently-joined building blocks, such as molecules or ions, adsorbed on a metallic and highly crystalline substrate. Thus, solid-state self-assembled materials often include organic/inorganic parts, crystalline and molecular components and a generally heterogeneous character.

#### **Computational challenges in the study of surface self-assembly**

Because of the subtle, and sometimes unpredictable, ways in which local and long-range effects influence each other, the boundary between scales at which these effects matter crucially, slightly or not at all, is blurry at best and often times nonexistent.

Similarly, interactions between heterogeneous often times transcend the rules and heuristics that can be invoked on their homogeneous parts.

Thus, two main challenges emerge when one attempts to simulate surface self-assembly processes. The first is the challenge of scale. Methods that work best for short scale processes, such as density functional theory for the electronic structure determination and analysis, scale poorly with system size. Methods designed for statistical treatment of many-body systems, such as Monte Carlo techniques, permit only low resolution representations of building blocks. The second challenge is that of heterogeneity. Many computational techniques developed by chemists are optimized to treat accurately and efficiently covalently-joined systems, while posing difficulty in treating bulk, crystalline systems. Material scientists and crystallographers on their part have developed tools and techniques to simulate crystalline states specifically.

### **Computational treatment of surface self-assembly**

Under these circumstances, accurate modelling of surface self-assembly systems requires the concurrent use of a plurality of simulation techniques, with a thorough understanding of each's strengths and limitations. These have been developed by chemists, solid-state physicists and material scientists, often times in isolation of one another. Each of these tools was designed and optimized to solve targeted problems, characterizing very specific systems at certain, narrow length scales and targeted properties. From subatomic, quantum-level representations of electronic structures, to classical large-scale, statistically sampled ensembles, these conceptual and computational tools are presented in this chapter, with the hope that, taken together they would allow us to paint a comprehensive picture of surface self-assembly via the simulation route.

### **Chapter layout**

We begin this chapter by introducing the quantum mechanical and statistical principles required for a scientific treatment of physico-chemical systems in Section 2.2.

In Section 2.3 we delve into the simulation realm, by offering a presentation of the two main computational techniques used to explore classical physico-chemical systems, molecular dynamics and Monte-Carlo.

Here we pay significant attention to the Monte Carlo technique, as we show that it is the method of choice for sampling many-body systems at equilibrium. We de-

scribe the most common problem that Monte Carlo faces, that of configurational entrapment, and we demonstrate how the Parallel Tempering formalism, building upon Monte Carlo, elegantly solves this issue. Here we also describe how molecular interactions are represented computationally in classical systems and why for large-scale many-body representations the Lennard-Jones-electrostatic potential offers the best compromise between accuracy and computational cost.

Section 2.4 presents the various ways in which quantum techniques are utilized to simulate detailed molecular electronic structures computationally, and why the method of choice at this scale is Density Functional Theory.

Section 2.5 presents the various ways in which the treatment of dispersion interactions is approached within density functional theory, which by design excludes them. A number of dispersion correction strategies with varying degrees of accuracy and computational cost are presented. The available data on their performance for aromatic compounds on metal surfaces is also reviewed here.

In Section 2.6, the final section of this chapter we focus our attention on how electronic structure simulations are performed in the case of crystalline solid state materials. The strategies employed in representing periodic systems are contrasted with those used for molecules and subsequently reconciled to provide a basis for the quantum-mechanical simulation of hybrid systems.

## 2.2 Modelling Fundamentals

### 2.2.1 Quantum Mechanics

#### Schrödinger's equation

At the most fundamental level, any molecular or material system is subject to the principles of quantum mechanics. Any system, no matter the size, is characterized by its associated wave function,  $\psi(\mathbf{r}, t)$ , whose form is described by the solution to the time-dependent Schrödinger's equation:

$$i\hbar \frac{d}{dt} |\psi(t)\rangle = \hat{\mathcal{H}} |\psi(t)\rangle \quad (2.1)$$

where  $\hat{\mathcal{H}}$  is the Hamiltonian operator,  $\hbar$  is the reduced Planck constant equal to  $h/2\pi$ ,  $i$  is the imaginary unit and  $t$  represents time.  $\psi$  defines the state of the system at any moment  $t$  and point in space  $\mathbf{r}$ . If  $\psi$  is known any (well-defined) inquiry about

the system may in principle be answered with the right manipulation of the wave function.

### Stationary states

For bound systems in which the potential energy component of the Hamiltonian is time-independent, the wave function may be separated into a time-dependent and a space-dependent components. If we focus on stationary states only, that is, on states for which the observable properties of the system do not change over time, we restrict the set of wave functions to those satisfying the time-independent form of Schrödinger's equation:

$$\hat{\mathcal{H}}|\psi\rangle = E|\psi\rangle \quad (2.2)$$

If a wave function  $|\psi_i\rangle$  satisfies Equation (2.2), then  $|\psi_i\rangle$  is an eigenvector of  $\hat{\mathcal{H}}$ , and represents an eigenstate of the system. That is, if  $\hat{\mathcal{H}}|\psi_i\rangle = E_i|\psi_i\rangle$ , then  $|\psi_i\rangle$  corresponds to a unique quantum state in which the system has energy  $E_i$ . The value  $E_i$  is the eigenvalue of the energy eigenstate  $|\psi_i\rangle$ .

### The quantum representation of a system

The explicit form for the Hamiltonian  $\hat{\mathcal{H}}$  determines the nature of the quantum-mechanical system. Properties other than energy restrict the form of the wave function in an analogous manner. For an observable  $O$ , there is an operator  $\hat{O}$  which adopts discrete eigenstates  $|\psi_i\rangle$  and the associated eigenvalues  $O_i$  analogously to  $E$ , that is,  $\hat{O}|\psi_i\rangle = O_i|\psi_i\rangle$ . For an arbitrary state  $|\psi\rangle$  the expectation value of  $O$  is  $\langle\psi|\hat{O}|\psi\rangle$ . From this it follows that  $\psi$  completely defines the system. This formalism is general regardless of scale of the system, from a single particle, to macroscopic aggregates.

## 2.2.2 Statistical Mechanics and Classical Representations

### The power of averages over large numbers

Solving Schrödinger's equation and interpreting wave functions directly for the purpose of obtaining experimentally relevant information is a daunting task even for the simplest of systems. Fortunately, most experimental data that theoreticians aim to

reproduce or predict involves averages of multiple forms, over a large number of particles, length scales and/or over periods of time (relative to the size, length or time scales under which subatomic and molecular systems exist).

The role of statistical mechanics is to bridge the gap between the bewildering complexity, conceptual and computational, of the quantum realm, and the natural orderliness and often striking simplicity of thermodynamics. Macroscopic systems have a huge number of degrees of freedom and behave according to principles described by thermodynamics. Statistical mechanics has the role of abstracting away most quantum-mechanical details, keeping the bare minimum of assumptions and only adding a few more to reach thermodynamic concordance. Statistical principles are employed to determine the limit behaviour models, to establish what kind of averages correspond to experimentally relevant data, and to determine which approximations, if any, may be adopted without a loss of accuracy.

### The microcanonical ensemble

According to quantum-mechanical principles a system comprised of a large number of particles has a high degree of degeneracy for each energy value. By fixing the number of particles to  $N$  and the volume to  $V$ , the energy levels  $E_i$  are fixed as well. If  $\Omega(N, V, E_i)$  is the number of eigenstates  $|\psi_i\rangle$  with energy  $E_i$  within the specified conditions, then a high degeneracy means that  $\Omega$  is typically a very large number,<sup>112</sup>

$$\Omega(N, V, E_i) = \Omega_i(N, V) \approx 10^{10^{20}\ddagger}.$$

The collection of all states adding up to  $\Omega_i(N, V)$  is called the microcanonical ensemble for the  $(N, V, E_i)$  macrostate.

A fundamental axiom of statistical mechanics is *the postulate of equal a priori probability* which affirms that all states  $|\psi_i\rangle$  contribute an equal weight to the macrostate  $(N, V, E_i)$ . That is, an isolated system with  $N$  particles, of volume  $V$  and energy  $E_i$  is equally likely to be found in any given microstate  $|\psi_i\rangle$  out of the total of  $\Omega_i(N, V)$ . Certain macroscopic properties emerge naturally from these assumptions and behave analogously to their thermodynamic counterparts. For example, it can be demonstrated that the entropy of the macroscopic state  $(N, V, E_i)$  is linked to the number of microstates by

$$S(N, V, E_i) \sim \ln \Omega_i(V, N).$$

---

<sup>‡</sup>In decimal form this number is a one followed by one hundred billion billion zeroes.

Similarly, the thermodynamic temperature can also be linked to the energy-partitioned eigenstates:

$$\frac{1}{T} \sim \left( \frac{\partial \ln \Omega(N, V, E)}{\partial E} \right)_{N, V}. \quad (2.3)$$

### The canonical ensemble

For a system in contact with a large heat bath at a constant temperature  $T$ , macroscopic states with different energy levels are permitted. Given the postulate of equal a priori probability, each energy level determines a microcanonical ensemble, and the collection of all such ensembles over the range of possible energy levels constitutes the canonical ensemble. The probability to find the system in the state  $(N, V, E_i)$  is given by the well-known Boltzmann distribution:<sup>112</sup>

$$P(N, V, E_i) = \frac{e^{-\frac{E_i}{kT}}}{\sum_j e^{-\frac{E_j}{kT}}}, \quad (2.4)$$

where  $j$  covers all possible energy levels and  $k$  is the proportionality constant required for equivalence in Equation (2.3). The ratio of these probabilities corresponding to two states  $(N, V, E_i)$  and  $(N, V, E_j)$  is known as the Boltzmann factor,

$$\frac{P(N, V, E_i)}{P(N, V, E_j)} = \frac{e^{-\frac{E_i}{kT}}}{e^{-\frac{E_j}{kT}}} = e^{\frac{E_j - E_i}{kT}}, \quad (2.5)$$

and is a function of the energy difference between the two states. Knowing the Boltzmann distribution of energy levels for a given temperature enables one to determine the canonical average value of an observable at that temperature:

$$\langle O \rangle_T = \frac{\sum_i \langle \psi | \hat{O} | \psi \rangle_i e^{-\frac{E_i}{kT}}}{\sum_i e^{-\frac{E_i}{kT}}} \quad (2.6)$$

### The grand canonical ensemble and fluctuations

If the constraints on the system are further relaxed by allowing matter to traverse boundaries, a third ensemble can be defined at chemical equilibrium. In the grand canonical ensemble, the large bath surrounding the system has a fixed chemical potential, in addition to a fixed volume and temperature.

Each of the three ensembles define a relationship between a system at equilibrium and its environment, specifically what kind of fluctuations are allowed between the two and of what magnitude. Isolated systems are described by the microcanonical ensemble, where no exchange of energy or matter occur, and fluctuations of either kind are absent.

Closed systems, described by the canonical ensemble, exchange energy with their surroundings and therefore may, and will, experience fluctuations in total energy content. At different times, a system in the canonical ensemble will be observed in states with different energies. The relative frequency of such individual states depends exponentially on the energy difference as it is given by the Boltzmann factor. Qualitatively speaking, small fluctuations are exponentially more abundant than large ones.

The grand canonical ensemble aims to simulate open systems in equilibrium with their surroundings. Fluctuations in composition occur in addition to fluctuations in energy. At different times, a system in the grand canonical ensemble will be observed in states with different energy and will possess different numbers of particles.

The three ensembles are schematically shown and contrasted in Figure 2.1. Relative to the size of the system, the surroundings must be large enough to adopt a well defined thermodynamic state, unaffected by the fluctuations experienced by the comparatively small system.

## Ergodicity

A system is ergodic if, when isolated to  $N$  particles, fixed volume  $V$  and total energy  $E$ , it satisfies the postulate of equal a priori probability for all accessible microstates. From a dynamical perspective, ergodic systems spend an amount of time in a given region of their associated microcanonical state space proportional to that region, over sufficiently large time intervals.

For an ergodic system, the time-averaged value of an observable  $O$  equals the ensemble average, in the limit  $t \rightarrow \infty$ :<sup>112</sup>

$$\bar{O} = \lim_{t \rightarrow \infty} \frac{1}{t} \int_0^t O(t') dt' = \langle O \rangle. \quad (2.7)$$

This equivalence between time-averaged and ensemble-averaged properties represents an important bridge between dynamical and ensemble models, offering the justifi-

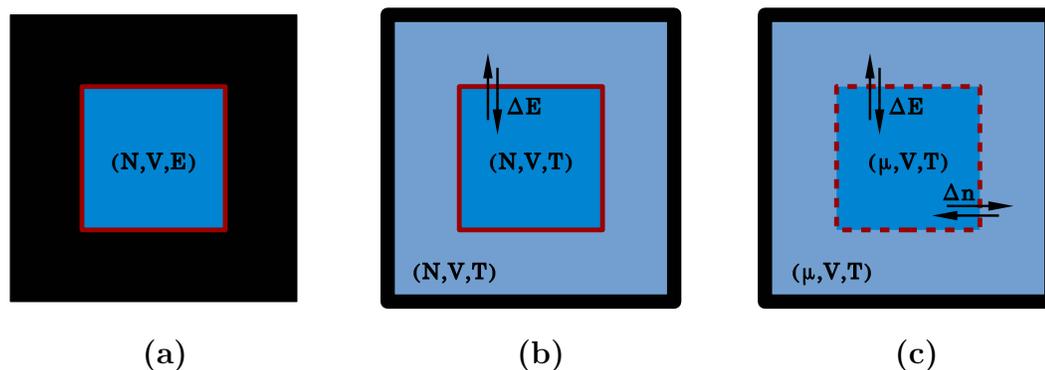


Figure 2.1: Schematic representation of systems associated with the three ensembles: (a) microcanonical ensemble – the system is isolated from the rest of the universe; (b) canonical ensemble – the system is in thermal equilibrium with a large heat bath at constant temperature and is closed; the system experiences energy fluctuations; (c) grand canonical ensemble – the system is in chemical equilibrium with a large reservoir with a constant temperature and chemical potential; both energy and composition fluctuations occur.

ation for the convergence of results from molecular dynamics simulations on one hand, and the statistical study of ensembles, on the other. Empirically, for real systems satisfying the ergodic hypothesis, theoretical ensemble studies produce results comparable to (existing or yet to be gathered) experimental data.

### 2.2.3 System Size and Representation

Within computational chemistry, the natural metric for determining the size of a system is not a length scale, but the total number of irreducible particles included in a simulation, be they electrons, atoms or molecules. Due to particle interaction, often the simulation time scales with the system size in a manner that is far from linear. Small scale systems, which tend to have a small number of degrees of freedom, may be represented quantum-mechanically within a reasonable simulation time. Large systems, however, have an astronomical number of degrees of freedom and are best represented by classical models. For short interactions, a quantum representation is required. Long-range interactions emerge statistically within classical representations. In surface SA both types of interactions contribute, therefore both classes of models must be employed for a complete picture.

## 2.3 Classical Methods

When simulating chemical systems using classical methods, the model is represented as a many-body system, for which the constituent parts obey the laws of classical mechanics. The keyword “classical” implies that the fundamental constituents have particle-like properties. The evolution of the system is governed by the interaction potential of the particles within the context of the system state. Equilibrium and/or dynamical properties of interest are computed as statistical averages of the model along the simulation path, with an acceptable sampling frequency and size.

Based on how the evolution of the system is undertaken, two main classes of simulation are used:

- (i) Molecular Dynamics (MD), for which a time-based evolution of the system is sought and
- (ii) Monte Carlo (MC), where one attempts to generate a random, albeit representative, sample of the state space of the system.

### 2.3.1 Molecular Dynamics

Conceptually, classical molecular dynamics is the most straightforward way to sample the configurational state space. The system is initialized by assigning each constituent particle adequate position and velocity. The initial configuration, together with the force field, uniquely define the forces acting on all particles. Ideally, integrating Newton’s equations of motion over a large enough time interval offers the evolution of the system. The complexity of the potential energy surface and the many-body problem renders the existence of exact analytical solutions impossible for all but the most trivial of cases.

#### The MD formalism

The time evolution of the system is determined via the numerical integration of Newton’s equations of motion. If  $\mathbf{r}_i(t)$  represents the position of particle  $i$  at time  $t$ , then the force  $\mathbf{F}_i(\mathbf{r}_i(t))$  acting on the particle can be derived from the force field:

$$\mathbf{F}_i(\mathbf{r}_i) = -\nabla U_i(\mathbf{r}_i) \quad (2.8)$$

where  $U(\mathbf{r})$  is the force field and  $\nabla$  represents the gradient operator. The computed forces are used to determine the position of particles after a given time step,  $\Delta t$ .

Using the Verlet algorithm<sup>112</sup> as an example, the new position for the particle is:

$$\mathbf{r}(t + \Delta t) = 2\mathbf{r}(t) - \mathbf{r}(t - \Delta t) + \frac{\mathbf{F}(t)}{m}\Delta t^2 \quad (2.9)$$

where  $m$  is the mass of the particle in question, and  $\mathbf{r}(t - \Delta t)$  is the particle's position at the prior step. Applying Equation 2.8 on the new coordinates  $\mathbf{r}(t + \Delta t)$  yields new forces, which are used to determine the positions at the next time step,  $t + 2\Delta t$ . For each new step the current state of the system is recorded. The process is iterated through until the system has traced and adequately long time interval and the recorded sampling size is statistically sufficient. During the post-processing phase any property of interest is computed for each recorded state, and the average over all iterated steps yields the true value.

### The limits of MD

Molecular dynamics can be used to trace the evolution of a chemical system in real time and accurately determine thermodynamic properties difficult to measure experimentally. There are, however, limits to the applicability of MD. Firstly, a sufficiently small time step  $\Delta t$  must be chosen in order to limit the truncation errors between steps and prevent them from pushing the system toward non-physical evolution. Lower time steps minimize errors, but come with a significant cost to the sampling speed. This trade-off is inherent in the MD formalism, and as a result, many processes are too slow to be wholly sampled with adequate time steps, within a realistic simulation time.

The processes in question often occur in systems with complex potential energy surfaces, where subregions of the configurational space are separated by barriers much greater than the thermal energy. Proteins, glasses and solid-state materials fall under this category. For these, “broken ergodicity” must be employed, where the average of interest is broken into sums of averages, each taken over a distinct subset of the phase space.<sup>113</sup> As the name implies, the ergodic hypothesis is “broken” by this approach, and with it, the guarantee that the resulting average of averages is representative of the state space average is lost.

Moreover, in many chemical systems, the problem of “nonergodicity” extends beyond that of “broken ergodicity”. Numerous simulations of chemical reaction dynamics have demonstrated that the assumptions of statistical theories are regularly violated. Impermeable or semipermeable barriers of many models strongly or com-

pletely limit the region of phase space a trajectory may explore, or the rate at which the exploration occurs, leading to predictions which greatly underestimate reaction rates.<sup>114–117</sup> Overestimations in the energy exchange caused by intramolecular vibrational relaxations during fast processes can also cause significant decreases in calculated reaction rates and equilibrium properties.<sup>118;119</sup>

In general, nonergodicity is common in chemical models, and when present, subverts the guarantee of equivalence between time averages and state space averages, thus limiting the reliability and applicability MD methods.

### 2.3.2 Monte Carlo Techniques

Molecular dynamics is inadequate in calculating equilibrium properties for all but the smallest of systems and the fastest of processes. The simulation time required to obtain data on a typical SA system, say a given long-range order parameter on a system with several hundreds of particles, is too long to be practical, even when a fast force field is used. There are ways of improving the sampling efficiency of the state space if one is willing to sacrifice the time-dependent information MD offers. The methods which focus on optimal sampling of the state space with disregard to the time evolution form a class of techniques broadly named Monte Carlo techniques. There are many Monte Carlo formalisms fit for a large variety of systems and all of them use randomness to dramatically increase the speed with which the state space is sampled.

#### The sampling challenge

The dimensionality of the state space is high even for a system with a relatively low number of components, as each state is determined by the position and momenta of all containing particles. Within the canonical ensemble  $(N, V, T)$ , the true value of observable  $O$  is given by the ensemble average.<sup>112</sup> In the continuous limit,

$$\langle O \rangle_T = \frac{\int O(\mathbf{r}^N, \mathbf{p}^N) e^{-\frac{\mathcal{H}(\mathbf{r}^N, \mathbf{p}^N)}{kT}} d\mathbf{p}^N d\mathbf{r}^N}{\int e^{-\frac{\mathcal{H}(\mathbf{r}^N, \mathbf{p}^N)}{kT}} d\mathbf{p}^N d\mathbf{r}^N},$$

where  $\mathbf{r}^N$  and  $\mathbf{p}^N$  represent the coordinates and momenta of all  $N$  particles. The

kinetic component of energy is a quadratic function of  $\mathbf{p}^N$  and this allows for an analytical solution of  $\mathbf{p}^N$ -dependent integrals,

$$\langle O \rangle_T = \frac{\int O(\mathbf{r}^N) e^{-\frac{\mathcal{U}(\mathbf{r}^N)}{kT}} d\mathbf{r}^N}{\int e^{-\frac{\mathcal{U}(\mathbf{r}^N)}{kT}} d\mathbf{r}^N}. \quad (2.10)$$

As a consequence, observables depending on momenta are typically easy to calculate. In contrast, the multidimensional integrals over particle coordinates in Equation (2.10) don't yield to analytical solutions in general, nor can they be tackled numerically in a straightforward manner. Due to the large dimensionality of even the most modest of many-body systems (on the order of, say, 100 particles), evaluating these integrals through equidistant or random sampling is an unfeasible task. Moreover, for most intermolecular potentials, including relatively simple ones, the Boltzmann factor is a rapidly varying function of the particle coordinates, creating a complex potential energy surface that requires a high sampling resolution.

### Weighted sampling

Fortunately, the overwhelming majority of state points have a vanishingly small Boltzmann factor, suggesting that it may be feasible to accurately sample the space with a small sampling size, through the use of importance sampling.<sup>112</sup> One could ideally sample the space nonuniformly, with the help of a weight function proportional to the Boltzmann factor. Constructing this function requires knowing the partition function for the system, which, if we would, defeats the purpose of sampling techniques in the first place.

The dimensionality and complexity of the state space rule out brute force Monte Carlo and the difficulty of determining partition functions rules out the straightforward importance sampling approach. However, Equation (2.10) suggests that, for the purposes of determining average properties, the exact form of the partition function is not required, and only the ratio of the two integrals is of importance. Crucially, while the absolute form of the partition function is unknown, the ratio of the Boltzmann factor between two known states is easily computable. A random walk through the state space with a hopping frequency proportional to such ratio yields to averages converging towards the canonical true value. This is the key insight being exploited

by the Metropolis approach.

### The Metropolis formalism

The Metropolis formalism<sup>112</sup> is an efficient way to sample systems in the canonical ensemble  $(N, V, T)$ . The system is generated as an initial configuration of  $N$  particles constrained to a box of volume  $V$ , bound by a force field yielding the energy  $U$ . The algorithm generates a Markov chain, by iterating through a fixed number of random walk trials. For each step the configuration is changed randomly from an old state  $o$  with energy  $U_o$ , to a new state  $n$  with energy  $U_n$  (see Figure 2.2). The values  $U_o$  and  $U_n$  are used to compute the relative ratio of the Boltzmann factors corresponding to the old and the new state:

$$r(o \rightarrow n) = \frac{e^{-U_n/kT}}{e^{-U_o/kT}} = e^{-\beta(U_n - U_o)}, \quad \beta = \frac{1}{kT}. \quad (2.11)$$

If the system moves to a more stable state then  $r(o \rightarrow n)$  is greater than one. In such a case, the new move is accepted to the sampling set with certainty, and a new trial is initiated. If, however, the state  $n$  is higher in energy than state  $o$ , then  $r(o \rightarrow n)$  is less than one. In the latter case  $r(o \rightarrow n)$  is used as the probability  $p_{\text{acc}}(o \rightarrow n)$  to accept the new state: a random number,  $r$ , between 0 and 1 is chosen from a uniform distribution and compared to  $r(o \rightarrow n)$ . If  $r < r(o \rightarrow n)$  then the trial is accepted and a new trial begins. Else, the new state is discarded, the old state is sampled, and another trial starting from the old state begins. In general, the acceptance probability for moving from state  $o$  to state  $n$  is

$$p_{\text{acc}}(o \rightarrow n) = \min(1, e^{-\beta(U_n - U_o)}). \quad (2.12)$$

The Metropolis sampling scheme is designed to generate an importance-weighted random walk through the region of the state space for which the contribution to the Boltzmann distribution is nonnegligible. Trial moves are accepted if they bring the system to a state with a higher Boltzmann weight, and are probably rejected if they bring the system to a state higher in energy.  $r(o \rightarrow n)$  rapidly converges to zero as the energy difference increases, and the rate of drop-off is influenced by  $T$ , as states lower in energy have relatively larger weights at lower temperatures. After an initial period of equilibration, in which all information about the arbitrary initial state is lost, the unweighted average over all accepted states yields a sample representative

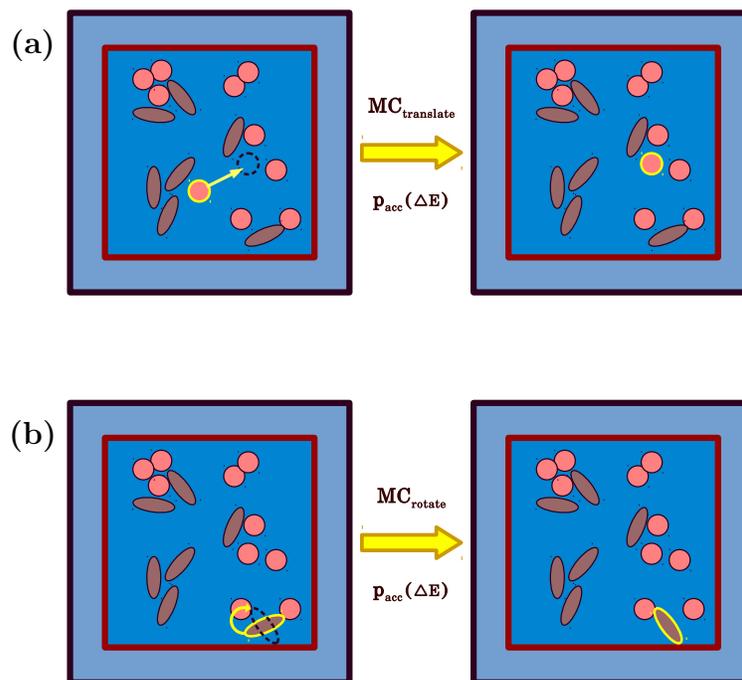


Figure 2.2: Representations of the most common configurational moves in Monte Carlo simulations: (a) basic translation move – random displacement in a random direction by a random magnitude of a particle randomly chosen; (b) basic rotation move – random rotation around a random axis by a random angle of a particle randomly chosen; care must be taken in the choice of rotation matrix as to not introduce any orientational bias, which would skew the sampling.

of the canonical ensemble at the given temperature.

Typical random walks through the state space include translational and rotational shifts of individual particles, conformational and configurational changes within flexible molecules, or swaps between non-identical particles (see Figure 2.2 and 2.3). A great deal of ingenuity can be used in designing efficient trial moves but care must be taken as to not introduce sampling biases through the accidental use of nonuniform random distributions or breaking the detailed balance condition.

### The physical interpretation of the Metropolis formalism

To better understand the Metropolis algorithm it is worth considering how the Monte Carlo scheme simulates the behaviour of a physical system, specifically in terms of energy fluctuations. A closed system in contact with a large bath at constant temper-

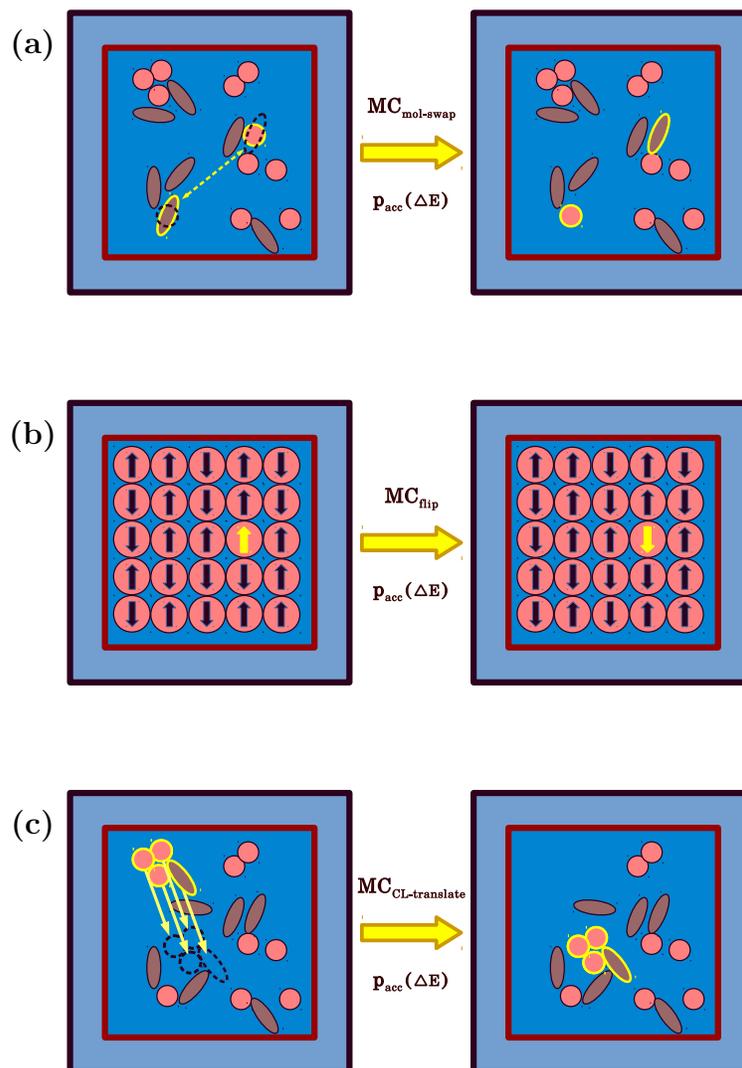


Figure 2.3: Additional common moves used in Monte Carlo simulations: (a) basic molecular swap – the geometric center is swapped within a random pair of non-identical molecules; (b) spin flip move; analogously, chiral molecules may have their configuration inverted via the same procedure; (c) cluster moves – a collection of particles identified as a cluster are displaced (or rotate) together with the same random parameters; identifying clusters is difficult in general.

ature eventually reaches thermal equilibrium with the bath (the canonical ensemble setup, see Figure 2.1).

Equilibrium fluctuations will center around an average energy value, dependent on the system size and temperature. Figure 2.4 sketches the probability distribution  $P_E$  of finding the system in a state with energy  $E$  at a set temperature. Values

higher than the average energy require a net flow from the heat bath to the system. Although the bath is much larger than the system and could provide the excess energy, the number of ways in which the excess energy can be partitioned across the degrees of freedom of the bath are much larger than the number of ways in which the same partitioning can be achieved within the system. The excess energy would lower the entropy of the system–bath couple and therefore the fluctuation is unstable.

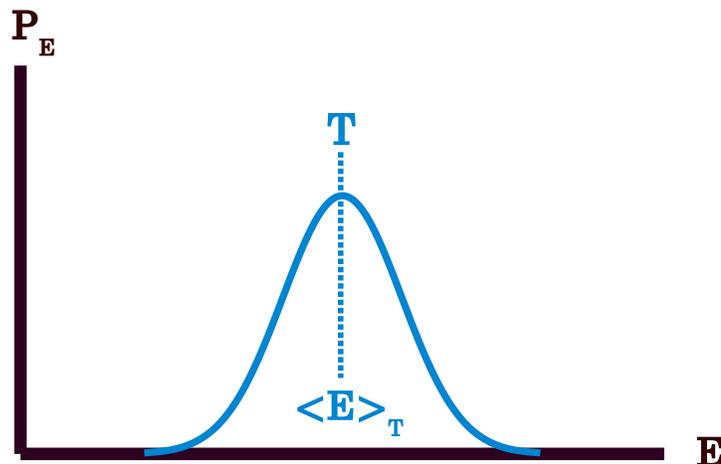


Figure 2.4: The probability distribution  $P_E$  of finding the system at a given temperature  $T$  in a state with energy  $E$ . Individual low energy states are preferred, but higher energy states are more numerous and overall contribute to the average more. High energy states require large energy fluctuations between the system and the bath, which are exponentially less likely to occur than lower energy states and quantitatively depend on  $T$ . The outcome of these two energy constraints leads to a distribution resembling a bell curve whose peak ( $\langle E \rangle_T$ ) and width are  $T$ -dependent. The peak and width of the distribution increase with  $T$ .

Values lower than the average lead to fluctuations which are not unstable, as the excess energy that the system must release in the surroundings can be redistributed within the large heat bath without an overall decrease in entropy. Lower energy states, however, are being outnumbered by higher energy states and therefore the system is more likely to be observed in one of the latter.

In essence, the scarcity of low energy states and the unavailability of energy trapped within the large degrees of freedom of the bath lead the system to adopt an intermediate average energy, with the frequency of deviations from the average dropping off rapidly with magnitude in both directions (see Figure 2.4).

Within the Metropolis scheme, the system “jumps” from configuration to config-

uration via Monte Carlo moves (see Figure 2.5), in a way that constrains the pattern of energetic fluctuations to that manifested by a system in the canonical ensemble. When the move results in a lowering of energy the move is accepted with certainty and the large bath readily absorbs the excess energy. At equilibrium the a priori probability of such a move to occur is quite low. If the system is set up in a high energy state far from equilibrium, such moves are common prior to reaching equilibrium and constitute the pre-equilibration phase. Figure 2.6 demonstrates how the Monte Carlo technique samples the configurational state space.

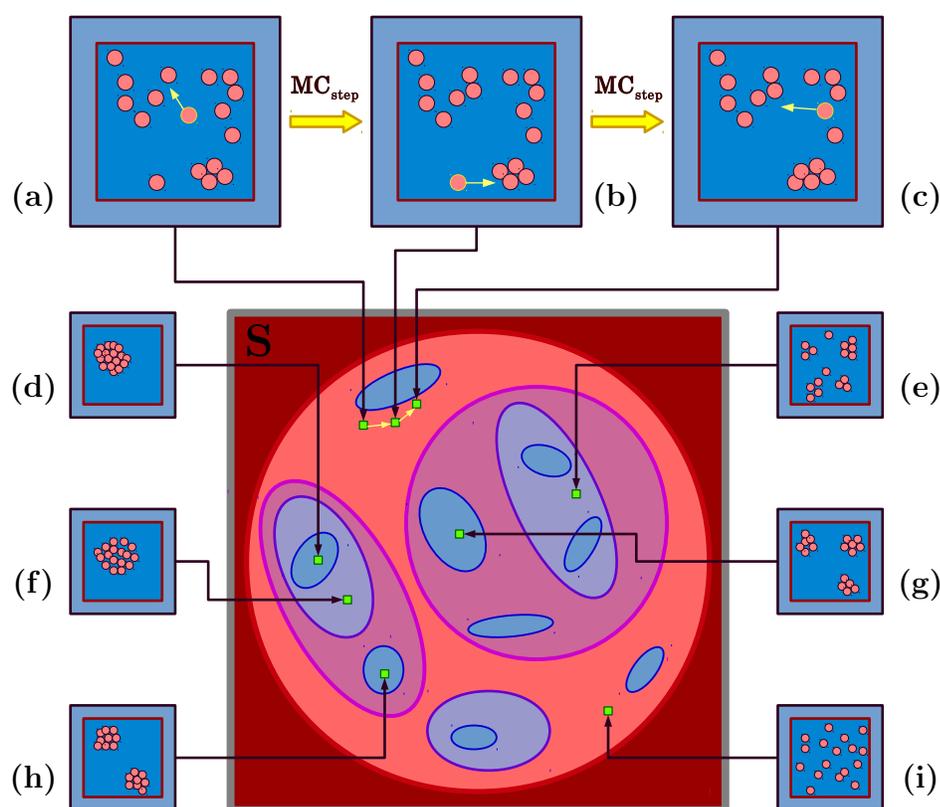


Figure 2.5: Representation of the configurational state space ( $S$ ) as a two-dimensional surface contour map in potential energy (PES). Blue domains correspond to low energy configurations, red areas contain high energy configurations, with purple and pink domains representing intermediate energy. Two accepted MC moves translate the system across  $S$  via path (a–c) and are indicated on  $S$  by yellow arrows. Configurations (snapshots) (d–i) of varying energies are displayed and highlighted on  $S$  with green squares. Configurational state spaces of many-body systems have a large number of dimensions.

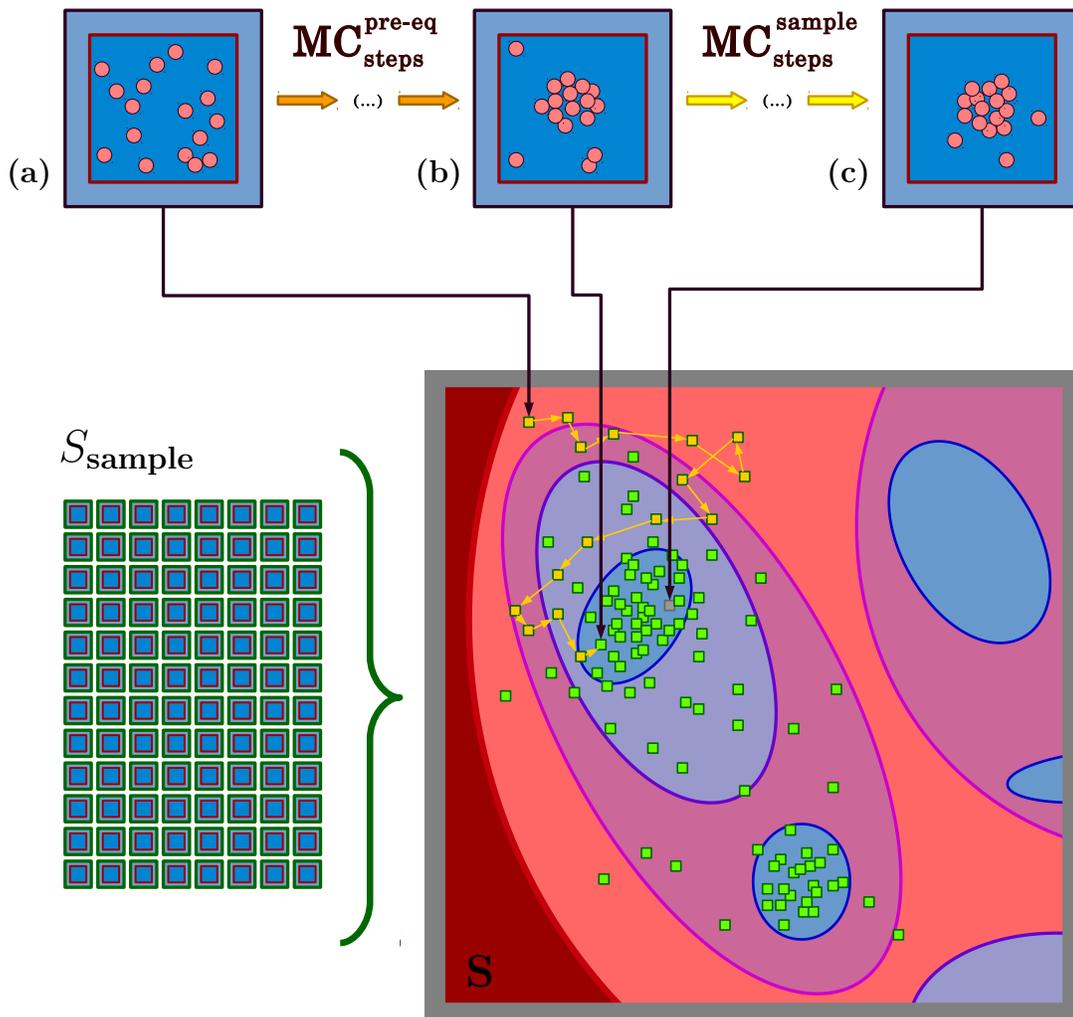


Figure 2.6: Sampling of the configurational state space ( $S$ ) with the Monte Carlo technique. Starting from a random initial configuration (a) the system reaches configuration (b) through a set number of Monte Carlo moves  $MC_{steps}^{pre-eq}$ . This is marked as the pre-equilibrium phase and from this step onward the sampling is assumed to be representative of equilibrium. A number of Monte Carlo moves  $MC_{steps}^{sample}$  samples the system for a set number of moves ending with configuration (c). The pre-equilibrium configurations are represented on  $S$  with yellow squares and the path taken through  $S$  is represented with yellow arrows. Green squares are the configurations constituting the sample  $S_{sample}$ , completed at the end of the run. The path through  $S$  after reaching equilibrium is omitted.

When the move results in an increase in energy the move may be rejected. The excess energy of the new state has to come from the heat bath, accompanied by the

corresponding lowering of overall entropy. The probability to accept such a higher energy state drops off exponentially with the amount of excess energy required from the heat bath, as determined by the Boltzmann factor. By picking a random number from the uniform distribution  $[0,1]$  and then accepting the move only if the number happens to be less than or equal to the Boltzmann factor, canonical ensemble fluctuations are modelled. The frequency with which this sampling procedure includes higher energy states mirrors that of energy fluctuations experienced by the system. The set of states thus sampled will be representative of the microstate distribution of the system in its given thermodynamic macrostate (provided that the sample size is large enough).

### **What the Metropolis scheme does and doesn't provide**

It is worth emphasizing that the sample sets collected via the Metropolis scheme are collections of microstates with relative frequencies representative of the macrostate. These can be used to compute macroscopic equilibrium quantities via averaging over each microstate in the sample. What sample sets do not contain is any information about the time evolution of the system, neither at nor away from equilibrium. The Metropolis scheme is entirely stochastic: what kinds of moves are attempted in what sequence is randomly determined and so are the parameters provided to a given move (which molecule to displace, which axis to rotate around and by what angle, etc). Moreover, the path by which the system reaches equilibrium, or moves between different configurational domains, is entirely determined by the choice of Monte Carlo steps utilized. These do not generally represent physical moves available to the system. Getting mechanistic information about these processes requires Molecular Dynamics simulations, where the relevant forces acting on the system drive the evolution through the state space.

### **The problem of entrapment**

The Metropolis scheme offers the potential to explore the state space of the system without the computationally expensive step of considering all of the forces acting on all of the particles at, each point in the evolution. Unbound by the constraints of physical evolution, much variety exists in the design choices for MC steps, and therefore a great deal of flexibility is inherent in the MC methodology. Properly designed MC steps allow the system to jump across the state space with great computational effi-

ciency, reach equilibrium structures much faster and sample the configurational state space more thoroughly than MD. The flip-side to this freedom of choice in movement through the configurational space is that, once the moves are set, the exploration is constrained to only those paths allowed by these moves.

To highlight how this can cause problems, consider the simple example presented in Figure 2.7. A number of mutually attractive spherical particles are equilibrated, starting from an arbitrary initial configuration. For the state space sampling only one MC move is chosen, the translation in a random direction by a random magnitude of a randomly chosen individual particle. A maximum cap for the displacement magnitude is set, since large translations are likely to result in configurations with strong potential overlap between the displaced particle and the rest. Such overlap leads to a strong repulsion and a large energy increase, making such a move very likely to be rejected. A high rejection rate makes the sampling inefficient.

It is easy to see that the equilibrium structure of such a system at low temperatures is a cluster of spherical symmetry, with inter-particle interactions maximized and a minimum surface edge [Figure 2.7, snapshot (h)]. Throughout the equilibration phase, however, it is entirely possible for the system to reach a configuration with two spherical clusters, separated by a substantial distance. Escaping this local energetic minimum structure could prove impossible. Tracing a path of the two clusters “approaching” one another and eventually “merging” via individual particle displacements may take an overwhelmingly large number of steps. If each cluster is comprised of a large number of particles, the number of simulation steps required may be impossible to reach in practice. Alternatively, individual particles may “break free” from one cluster and eventually reach the other. If the inter-particle potential is strong enough relative to the temperature at which the system is sampled, any set of moves guiding individual particles from one cluster to the other will be met with an overwhelmingly high rejection rate. To overcome this rejection barrier we are again forced to perform an absurdly high number of moves.

Either way, this scenario in practice leads to a false identification of the two-cluster local minima as the equilibrium structure. Worse still, starting from different initial configurations, different local minima could be reached, for example three or more isolated clusters. Not only is the equilibrium structure not reached, but the structures reached are crucially dependent on the arbitrary initial configurations. If reaching true equilibrium for a set of identical spherical particles with spherical interactions is not guaranteed, the problem is only exacerbated for particles of arbitrary shape and

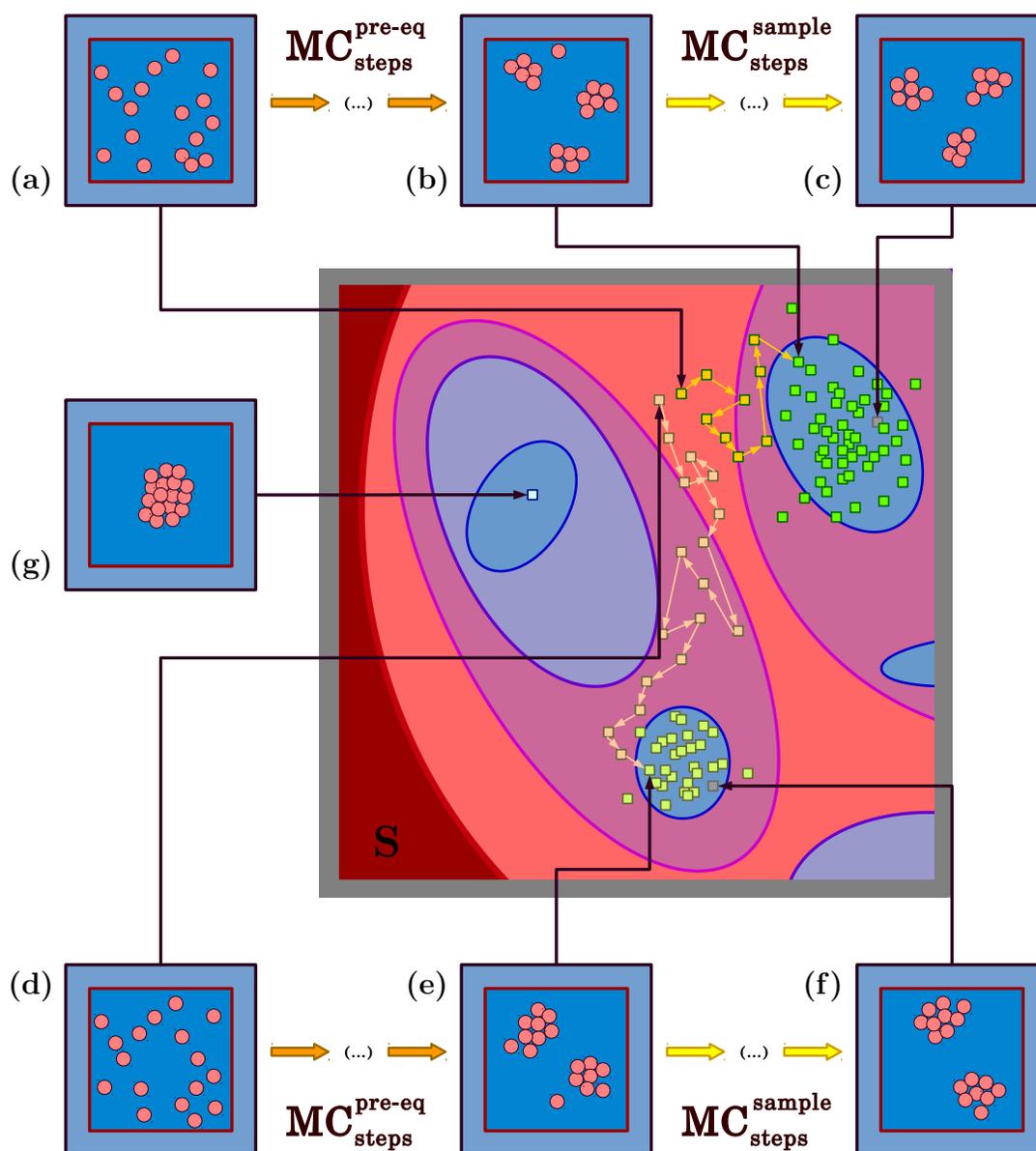


Figure 2.7: Entrapment in local minima on the PES. Two runs [(a-c) vs (d-f)], are highlighted on  $S$  with different shades of yellow and green. Despite close initial configurations [(a) vs. (b)], the two runs rapidly diverge and reach different minima [(d) vs (e)], where they remain stuck until the end of the simulation [(c) vs (f)]. Neither reaches the true global minima, represented by snapshot (g). The probability of crossing the barriers required to reach (g) with classic Monte Carlo moves is very low at this temperature.

potential or with systems with varying composition.

This example illustrates the famous problem of entrapment in local minima, which is common in MC simulations of all but the simplest of systems.

### **Avoiding entrapment**

Entrapment is a consequence of the rigidity embedded in the MC moves, in an attempt to reduce their computational complexity (which speeds up the sampling process) and avoid introducing configurational bias (which would skew the sample away from the equilibrium distribution). One solution is to design moves meeting both criteria while also increasing the flexibility with which the system can move through the state space. Examples include chiral inversion of random molecules in enantiomorphous systems, swaps of random particles of different type in heterogeneous systems or collective movement of particles. In the previous example, cluster displacement moves [translating collections of strongly-bound particles together, see Figure 2.3 (c)] can readily solve the issue of entrapment. While identifying such clusters can be done efficiently for simple systems, the problem becomes arbitrarily difficult to solve for general shapes and potential. Solving the problem of entrapment with the design of ingenious moves becomes a system-dependent pursuit.

Fortunately, a general solution to the entrapment problem exists in the form of Parallel Tempering Monte Carlo formalism.

### **The Parallel Tempering Monte Carlo formalism**

The Parallel Tempering Monte Carlo formalism (PTMC) is a smart sampling technique which has radically improved sampling for the Metropolis methodology.<sup>112;120–122</sup> The method offers a general solution for avoiding entrapment and requires the simultaneous simulation of several copies of the system (typically denoted as replicas), with each system being sampled in a different thermodynamic state (see Figure 2.8). Each individual replica is confined to the canonical ensemble (the method can be applied to other ensembles too), and the collection of replicas determines an extended ensemble. With different thermodynamic states, each replica samples the PES with different mobility and manifests a different energy average. Crucially, adjacent replicas must be close enough in thermodynamic state to have a non-negligible overlap of accessible configurations within the state space. Occasionally, a swap of configurations between two adjacent replicas is performed, allowing for configurations trapped in low-mobility

replicas to be broken apart and reorganized within higher-mobility replicas.

The central requirement to keep the thermodynamic states of the swapping replicas intact is achieved by making the swapping process stochastic, with an acceptance probability proportional to the overlap of energy levels between the two thermodynamic states. Swaps satisfying this criterion can be considered a type of MC move within the extended ensemble, subject to the same constraints that the individual replica moves must follow. They do not break detailed balance (each configuration is left intact), preserve the thermodynamic state of each replica and are computationally cheap (require no additional energy components beyond what each replica has already computed in past moves). With a good overlap between replicas, swap moves decrease the likelihood of any to get trapped in a metastable state for a large number simulation steps. This causes all replicas within the extended ensemble to reach the equilibrium configuration much faster than they would without tempering.

### Tempering over temperatures

Tempering can be done over any force or potential that influences the thermodynamic state of the system. Examples include chemical potential, density and temperature, with the latter being the most common.

Focusing on tempering over temperatures specifically, the PTMC formalism consists of the creation of a Markov chain with two classes of moves:

1. Standard MC moves: where the configuration of a given replica is changed randomly, with the move being accepted or not based on the Metropolis criterion:

$$p_{\text{acc}}^{\text{classic}}(o \rightarrow n) = \min(1, e^{-\beta(U_n - U_o)}), \quad (2.13)$$

where acc represents the acceptance probability from the  $o$  (old) to the  $n$  (new) configuration,  $U$  is the total potential energy and  $\beta$  is the inverse temperature ( $1/T$ ).

2. Swap moves between two neighbouring replicas of the system ( $i$  and  $j = i + 1$ ), where  $i$  is selected at random. The configurations of the two replicas are swapped with an acceptance probability given by

$$p_{\text{acc}}^{\text{swap}}([i, j] \rightarrow [j, i]) = \min(1, e^{-(\beta_i - \beta_j)(U_i - U_j)}). \quad (2.14)$$

The closer in temperature the two replicas  $i$  and  $j$  are, the higher the probability that the swaps be accepted, with the difference in energy between the two,  $U_i$  and  $U_j$ , determining the acceptance likelihood. The acceptance ratio can be predicted by

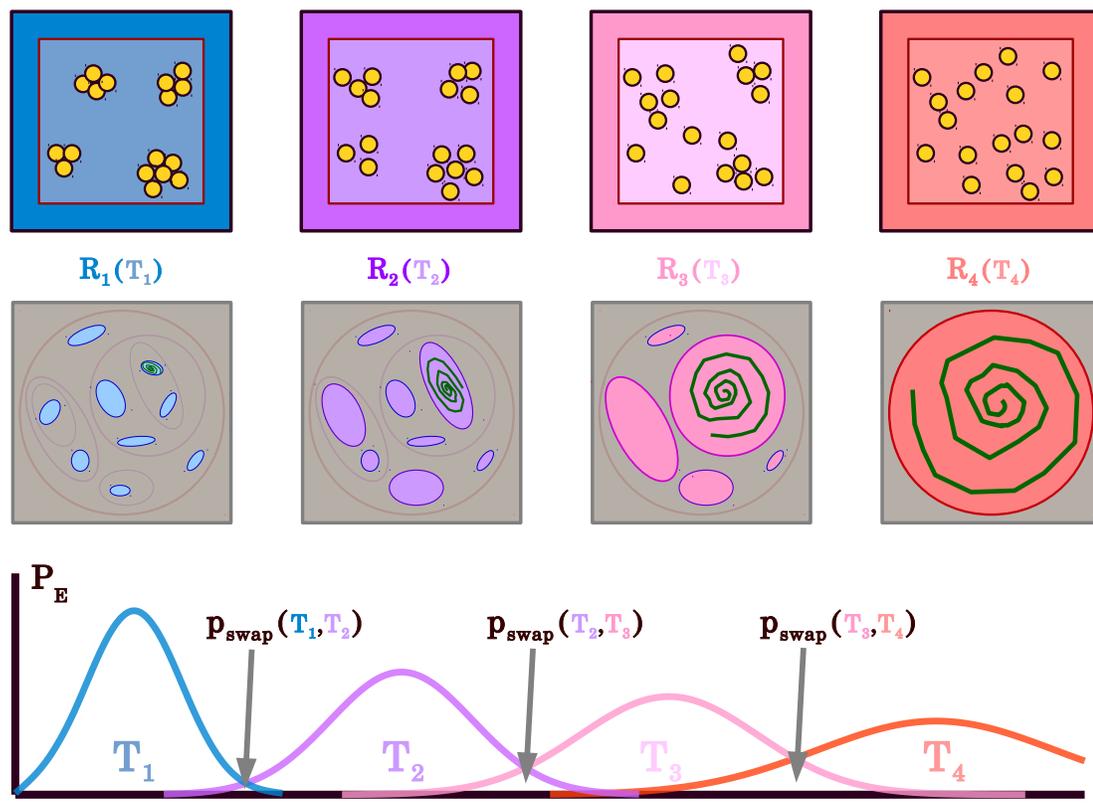


Figure 2.8: The Parallel Tempering Monte Carlo Formalism. Four identical replicas  $R_1$ – $R_4$  sample the configurational state space  $S$  at different temperatures  $T_1$  through  $T_4$ , from lowest to highest respectively (top row). The sampling mobility of each replica is shown in the middle row. Replicas at low temperatures sample low energy domains and cannot pass between domains if the energetic barrier is too high. Green spirals represent the areas accessible by each replica in isolation. The bottom part displays the overlap of energy values between replicas at different temperatures. The probability of two replicas swapping configurations (and therefore energy levels) via a parallel tempering MC move is proportional to this overlap.

the amount of overlap of the energy histograms corresponding to the two replicas. Replicas closer in temperature will display a larger overlap and will therefore swap with a higher probability than replicas further apart thermodynamically (see Figure 2.8, bottom graph and Figure 2.9).

Usually the temperature difference between neighbouring replicas should be chosen such that roughly 10% of the attempted moves be accepted. Acceptance values markedly higher than this indicate that different replicas needlessly sample the same portion of the PES, potentially leaving out other portions. Significantly lower values

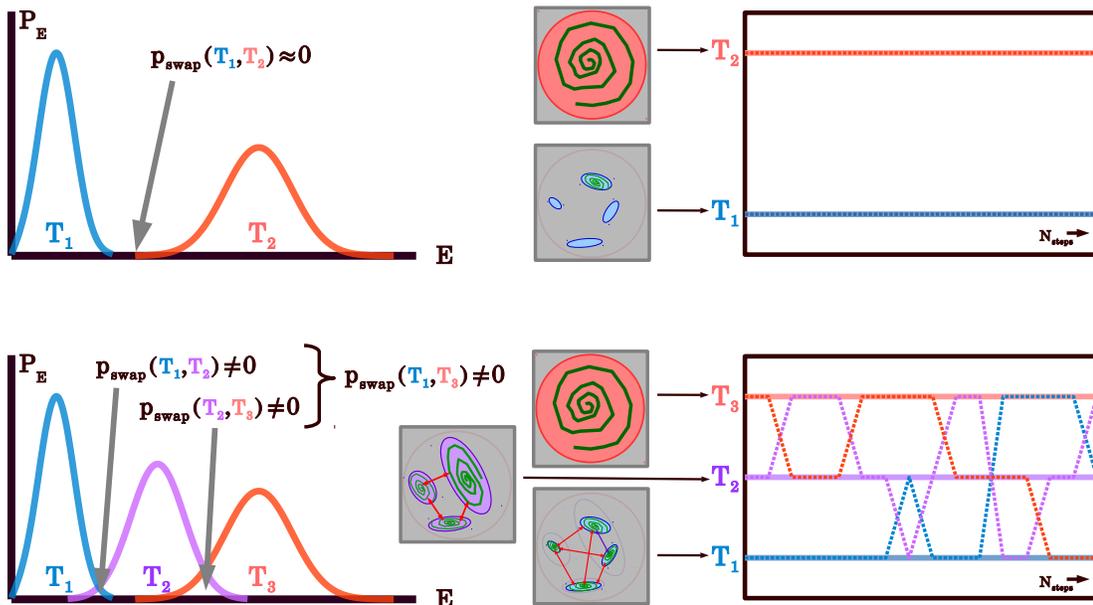


Figure 2.9: The relationship between temperature spacing, swapping probability and the mobility of low-temperature replicas on PES. Top example: temperatures  $T_1$  and  $T_2$  are so far apart that their energy overlap is negligible and the two replicas never swap configurations (top-left graph). The two replicas sample the configurational space in isolation, each with their assigned temperature (top-middle sketches). As a result, the low-temperature replica (blue) is locally trapped. The zero-swapping frequency is indicated in the top-right graph. Bottom example: in a different run a replica at intermediate temperature (purple) is introduced. The energy overlap between the low and intermediate temperature and the overlap between the intermediate and high temperature are both significantly greater than zero (bottom-left graph). As a result, the low-temperature replica gains access to the previously isolated minima (bottom-middle sketches). The colour of arrows linking different low-energy domains indicates the temperature required to overcome the energy barrier between them. The movement of different replicas across different temperatures via swaps is indicated in the bottom-right graph.

of acceptance indicate that the replicas are too far apart thermodynamically, essentially being sampled in isolation, defeating the purpose of parallel tempering. In order to ensure that the system of replicas is not trapped in a portion of the PES as a whole, the highest temperature replica  $i = n$  should be chosen such that it overcomes all energy barriers, i.e. is in the gas state. After an initial period of equilibration, in which the state of the system depends on the starting configuration, the sampling of each replica becomes representative of the equilibrium state.

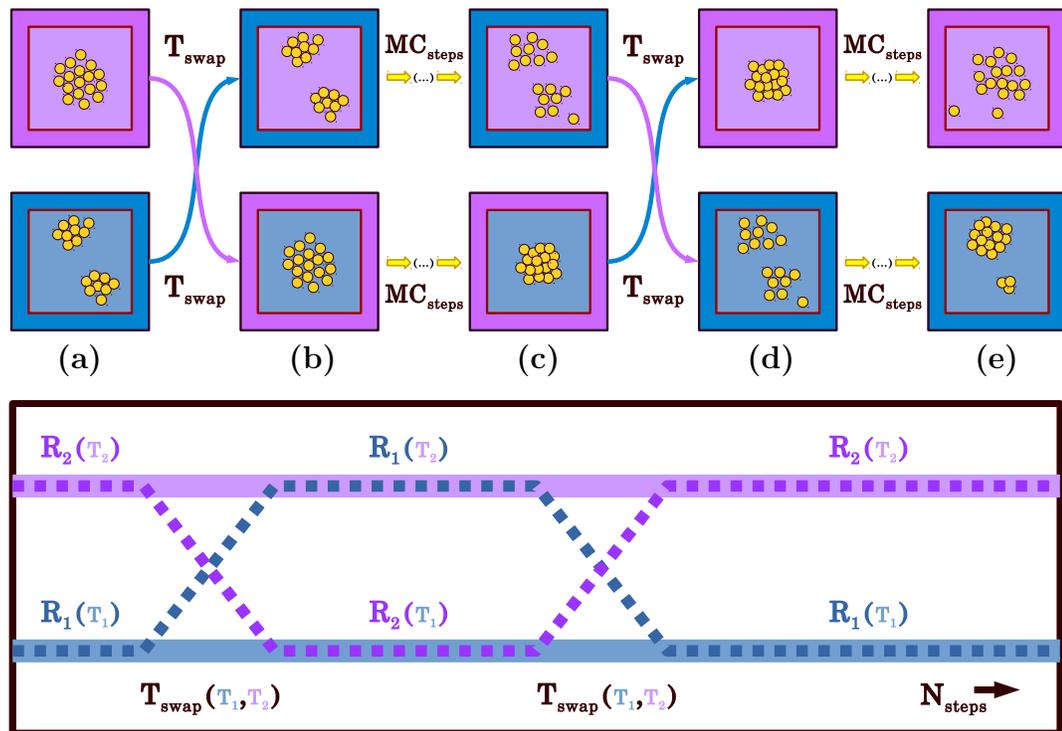


Figure 2.10: Demonstration of how swap moves eliminate entrapment for low temperature replicas. Replica  $R_1$  (blue border) samples at low temperature  $T_1$  (blue area).  $R_2$  (purple border) samples at a higher temperature  $T_2$  (purple area). (a) –  $R_1$  is currently stuck in a local minima with two clusters that will not merge; (b) – following a successful swap, the configuration of  $R_1$  is transferred to the higher temperature; (c) – the two-cluster configuration has a chance to be relaxed by the increased mobility of a higher temperature; (d) – a subsequent swap brings the relaxed configuration back to the low temperature; (e) –  $R_1$  has now reached a different local minima.

Figure 2.10 demonstrates graphically how tempering over temperatures solves the issue of entrapment for the example of spherical particles. Replica  $R_1$  is trapped at low temperature into a configuration with two clusters spaced apart, unable to merge via configurational moves. Following a successful swap move between  $R_1$  and the  $R_2$  with higher temperature  $T_2$ , the two clusters configuration is transferred to a higher temperature, where the chance of it being destabilized is higher. If this temperature is not high enough to allow for the sampling of configurations to overcome the cluster energy barrier, a possible subsequent move between  $R_2$  and  $R_3$  will bring the configuration to an ever higher temperature. Eventually the configuration will be at a high

enough temperature to overcome the energy barrier holding the clusters together, at which point the configuration has a chance to be relaxed. Stochastically, the relaxed configuration will have a chance to fall back to low energy replicas, where organization into the true minimum is now possible. For its part, the original configuration of R2 will be sampled at temperature T1 for some time. Within the extended ensemble, initial configurations from all replicas will have a chance to be sampled by all temperatures randomly and repeatedly, provided that the spacing between all pairs of neighbouring replicas is small enough to allow for swap acceptance.

The Parallel Tempering Monte Carlo methodology is an elegant and general solution to the problem of local entrapment, with the added bonus of simulating the system in different thermodynamic states concurrently. It does this by considering an extended ensemble in which replicas of the system exchange information without breaking the constraints imposed by the canonical ensemble. The increased computational cost created by the replication of the system can be mitigated through near perfect parallelization, as the communication between replicas is minimal by design. Coupled with an appropriate potential the PTMC formalism is an excellent tool for studying non-covalently bound rigid molecular models adsorbed on surfaces.

### 2.3.3 Force Fields

The most time consuming part of a simulation is almost always the calculation of the potential energy of various states (MC techniques), or the derivation of the force acting on the constituent particles (MD techniques). Both are heavily determined by the electronic energy, which is slow or impossible to calculate exactly for any system larger than a few atoms. Force field methods (FF) or molecular mechanics (MM) dramatically lowers the demand for energy computation by incorporating the electronic energy within the atomic potential, and treating the parametrized atoms as “building blocks” for the potential, instead of the bare electrons.<sup>123</sup> With an FF approach bonding information is provided explicitly, rather than by solving Schrödinger’s equation, as is any other kind of “special” interaction among atoms. The many-electron quantum-mechanical system is thus reduced to a many-atom Newtonian system. For time-independent phenomena, the computational task reduces to calculating the energy for a given geometric arrangement of atoms.

## Rationale behind the FF approach

A central observation in chemistry is that different molecules often tend to be composed of structurally and energetically similar units. Bonds of the same order between the same atoms have similar lengths and energies of formation in different molecules. This suggests that accurate models of molecules could be constructed from collections of “atom types”.<sup>123</sup> Each atom type is given a unique set of parameters which require proper calibration in the context of the types of molecules and systems being modelled. Calibration is performed by fitting a simple model to a set of available experimental data closest to what is studied.

## Force field energy

In FF methods the total energy of a multi-molecular system  $E_{\text{FF}}$  is considered as a sum of various terms, each contributing to either the covalent,  $E_{\text{bond}}$  or to the non-covalent,  $E_{\text{nonbond}}$ , component of energy:<sup>123</sup>

$$E_{\text{FF}} = E_{\text{bond}} + E_{\text{nonbond}} = E_{\text{str}} + E_{\text{bend}} + E_{\text{tors}} + E_{\text{vdw}} + E_{\text{el}}. \quad (2.15)$$

$E_{\text{bond}}$  includes contributions to energy due to interactions between covalently-bound atoms, such as bond stretching  $E_{\text{str}}$ , angle bending  $E_{\text{bend}}$  and rotation around a bond  $E_{\text{tors}}$ .  $E_{\text{nonbond}}$  covers contributions due to interactions between non-bonded atoms, such as van der Waals forces  $E_{\text{vdw}}$ , or electrostatic interactions between charged atoms,  $E_{\text{el}}$ .  $E_{\text{bond}}$  results exclusively from intramolecular interactions, while  $E_{\text{nonbond}}$  covers interactions between all non-covalently linked atoms, either belonging to the same, or to different molecules.

The explicit form for each of the terms in Equation (2.15) is a function of the atom types composing the molecules. With their parameters specified, the total energy becomes a function of the nuclear coordinates of all atoms. Identifying stable molecular geometries and supramolecular arrangements becomes a problem of finding local minima on the potential energy surface of the nuclear coordinates.

## Intra- and inter-molecular interactions

The principal distinction to the energy contribution between  $E_{\text{bond}}$  and  $E_{\text{nonbond}}$  in Equation (2.15) alludes to the targeted applicability of molecular mechanics. As chemical bonds among atoms are defined a priori and the atomic electronic structure

is reduced to an atomic electrostatic charge, force fields are ill-suited for studying chemical reactivity, processes involving the breaking or making of new bonds, or phenomena induced by detailed changes in the electronic structure. Rather, FF methods are best suited for simulating many-molecular systems of chemically stable molecules, where both intra- and inter-molecular interactions play significant roles in how the system behaves.<sup>123</sup> Typically, molecular self-assembly phenomena manifest in precisely these situations.

### Addressing long-range organization

The applicability of FF methods range from optimizing and characterizing the structures of individual molecules in gas-phase, all the way to the study of macroscopic phases of many molecules. For the latter cases, the contribution to energy from long-range, nonbonding components tends to dominate the total energy of the system. As the number of molecules grows, both the significance of  $E_{nonbond}$  and the computational cost of computing it increase. Most commonly the terms in  $E_{nonbond}$  are modelled as atomic pair-wise, additive contributions to the total energy, and consequently the computational cost for  $E_{nonbond}$  grows proportionally with the squared number of atoms. This sets a fast-reaching upper bound for the size of the model, limiting the use of FF methods in the long-range domain.

Several ways to improve the cost of computing  $E_{nonbond}$  are available, such as truncation of the potential or the use of periodic boundary conditions. These, however, are designed for models driven primarily by short-scale interactions, without much attention given to how they might influence the long-range behaviour. Truncation limits the computation of  $E_{nonbond}$  to only those terms for pairs of atoms within a certain relative range. For large systems, this cuts down both computational cost and some long-range resolution. Periodic boundary conditions may mimic bulk behaviour and remove edge effects, but restricts the long-range phenomena to only those commensurate with the box size and shape.

There is a way to increase the size limit without these limitations to the long-range freedom of the model, by sacrificing the short-scale resolution further. If molecules can be assumed to adopt largely discrete geometries with small variations around an optimal shape, then, over long ranges, the contributions due to local distortions cancel each other out. A model of many molecules with rigid optimal geometries will accurately replicate the long-range behaviour of such a system. By setting the

molecules to a rigid geometry, all contributions to energy due to intra-molecular interactions are thus fixed and included in the atomic parameters. The energy of the system is thus reduced down to

$$E_{\text{FF}}^{\text{rigid}} = E_{\text{nonbond}}^{\text{intermol}} = E_{\text{vdw}}^{\text{intermol}} + E_{\text{el}}^{\text{intermol}}. \quad (2.16)$$

As the superscripts in Equation (2.16) suggest, all contributions are from intermolecular atoms pairs.

### Van der Waals interactions

The van der Waals energy  $E_{\text{vdw}}(i, j)$  describes the interaction between two atoms  $i$  and  $j$  at the relative distance  $r_{ij}$ , not directly bonded covalently.  $E_{\text{vdw}}(i, j)$  is quantum mechanical in nature, originating from the overlap of the electron clouds between two atoms. At intermediate distances electron correlation induces dipole-dipole and higher multipole interactions, collectively referred to as “dispersion”, which results in a slight attraction. Because the leading, dipole-dipole, attraction term varies proportionally with  $r_{ij}^{-6}$ , the asymptotic behaviour of  $E_{\text{vdw}}(i, j)$  at long distances has a  $r_{ij}^{-6}$  dependence.  $E_{\text{vdw}}$  is asymptotically zero for large interatomic distances.<sup>123</sup> At short distances it is highly repulsive due to Coulomb repulsion and exchange interactions between electrons. Setting aside the Coulombic interactions for the moment, Pauli exclusion principle imposes a repulsive component that varies exponentially with distance,  $e^{-r_{i,j}}$ . This repulsion term combined with the attractive term constitutes the Buckingham potential,

$$E_{\text{Buckingham}}(i, j) = A_{i,j}e^{-B_{i,j}r} - \frac{C_{i,j}}{r_{ij}^6}. \quad (2.17)$$

The Buckingham potential describes the van der Waals forces accurately for short, intermediate and long distances, but can manifest non-physical attractive behaviour for very short distances as the exponential term converges while the second term does not. With a careful choice of parameters  $A_{i,j}$  and  $B_{i,j}$  the problematic domain can be made arbitrarily small but not eliminated. The risk of accessing the non-physical part of the potential is particularly pronounced in Monte Carlo simulations. Here, the generation of non-physical states is embedded into the recipe of the method. The formalism relies crucially on a correct form of the potential over the entire range of interactions. Violating this rule breaks the guarantee that non-physical states are

rejected.

### The Lennard-Jones potential

One way to eliminate this shortcoming of the Buckingham potential is to approximate its exponential term with a polynomial that diverges faster than the attractive part. This way arbitrarily short ranges would manifest ever-increasing repulsion. A popular choice of approximation is a 12-polynomial,

$$E_{\text{vdw}}(i, j) = A_{i,j} e^{-B_{i,j}r} - \frac{C_{i,j}^{\text{atr}}}{r_{ij}^6} \approx \frac{C_{i,j}^{\text{rep}}}{r_{ij}^{12}} - \frac{C_{i,j}^{\text{atr}}}{r_{ij}^6}, \quad (2.18)$$

where  $C_{i,j}^{\text{rep}}$  and  $C_{i,j}^{\text{atr}}$  are  $(i, j)$  atomic pair-dependent constants, for the repulsive and attractive part of the interaction, respectively. The approximate part of Equation (2.18) is the Lennard-Jones potential and can be rewritten as

$$E_{\text{vdw}}^{\text{LJ}}(i, j) = 4\epsilon_{ij} \left[ \left( \frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left( \frac{\sigma_{ij}}{r_{ij}} \right)^6 \right], \quad (2.19)$$

where  $\epsilon_{ij}$  and  $\sigma_{ij}$  are the energy depth and distance parameters for atoms  $i$  and  $j$  respectively. The  $r_{ij}^{-12}$  dependence of the repulsive term is justified by its adequate ability to approximate exponential behaviour over medium and short ranges, and by the ease with which it can be calculated. One great appeal of the Lennard-Jones potential is that it is several times faster to compute than the Buckingham potential, having replaced the need for an exponential computation with a squaring of the attractive term. Figure 2.11 displays the Lennard-Jones potential for two sets of  $(\sigma_{ij}, \epsilon_{ij})$  parameters.

### Electrostatic interactions

In order to represent ions and to account for the intra-molecular redistribution of electrons during bond formation, atom-based electrostatic interactions can be introduced via partial charges assigned to each atoms. For two atoms  $i$  and  $j$  separated by  $r_{ij}$ , with respective partial charges  $q_i$  and  $q_j$ , the contribution to  $E_{\text{nonbond}}$  due to electrostatic interactions is

$$E_{\text{el}}(i, j) = \frac{q_i q_j}{\epsilon r_{ij}}, \quad (2.20)$$

where  $\epsilon$  is the dielectric constant.

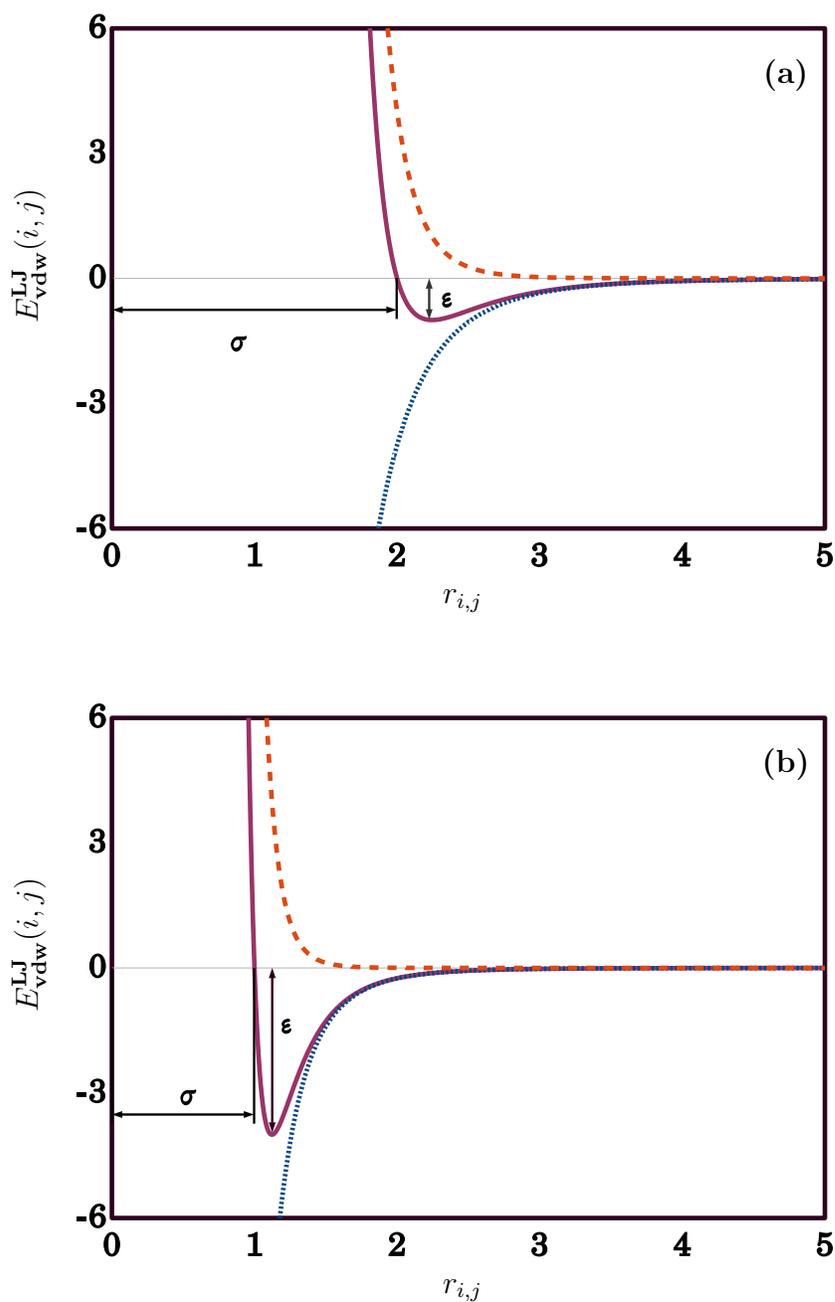


Figure 2.11: Graphical representations of the Lennard-Jones energy as a function of the distance  $r_{i,j}$  between atomic pair  $(i, j)$  for two sets of parameters: (a)  $\sigma_{ij} = 2.0$ ,  $\epsilon_{ij} = 1.0$ , indicative of a softer potential between larger atoms; (b)  $\sigma_{ij} = 1.0$ ,  $\epsilon_{ij} = 3.0$ , indicative of a sharper potential between smaller atoms. The repulsive and attractive components of the potential are displayed with red-dashed and blue-dotted curves respectively. Their sum is shown as the continuous purple curve. All quantities presented are in reduced units.

### The Lennard-Jones-electrostatic potential

By combining the contribution from van der Waals (Equation (2.19)) and electrostatic interactions (Equation (2.20)) for each non-bonded atomic pair, we derive the Lennard-Jones-electrostatic (LJe) potential for the system:

$$U_{nb}^{ab} = \sum_{\substack{i,j=1 \\ j < i}}^{n_{at}} \left[ 4\epsilon_{ij} \left( \left( \frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left( \frac{\sigma_{ij}}{r_{ij}} \right)^6 \right) + \frac{q_i q_j e^2}{4\pi\epsilon_0 r_{ij}} \right], \quad (2.21)$$

where  $a$  and  $b$  denote two interacting molecules,  $n_{at}$  is the number of atoms in a molecule,  $\epsilon_{ij}$  and  $\sigma_{ij}$  are the Lennard-Jones depth and distance parameters for atoms  $i$  and  $j$ ,  $r_{ij}$  is the distance between the centers of the two atoms,  $q_i$  is the partial charge on atom  $i$ ,  $e$  is the electronic charge and  $\epsilon_0$  is the dielectric constant of vacuum. Mixed Lennard-Jones parameters are normally obtained using the Lorentz-Berthelot mixing rules:  $\sigma_{ij} = (\sigma_{ii} + \sigma_{jj})/2$ ,  $\epsilon_{ij} = \sqrt{\epsilon_{ii}\epsilon_{jj}}$ .

Typically, with the use of the LJe potential, short range interactions are van der Waals driven while on longer ranges electrostatic interactions predominate, via the  $1/r_{i,j}$  dependence. A notable shortcoming of LJe potential is its inability to describe direction-dependent interactions, as it is spherically symmetric. Another is that it does not allow for coordination-dependent interactions, i.e. for the ability of atoms to change their reactivity based on their local environment. Despite these shortcomings, the LJe potential remains an excellent choice for describing interactions between non-covalently joined particles, where these shortcomings can be somewhat mitigated via the use of steric components and partial charges within particles of specifically tailored shapes and charges.

## 2.4 Quantum Methods

### Addressing short-scale complexity

A thorough study of self-assembly systems must include methods designed to accurately address long-range interactions. With a series of classical approximations, the dynamics of electron-electron interactions is collapsed down to a small set of static, atomistic parameters, rendering the exploration of long-range organization into a manageable challenge. The inevitable cost to this is the dramatic reduction in the resolution of the short-scale complexity coming out of the electronic structure,

within and between atoms. The details of electronic structure, however, crucially influence short-scale inter-molecular organization, and may only be accurately represented through quantum-mechanical models. It follows then that such models are a necessity tool in painting a complete picture of molecular self-assembly.

### Addressing chemically nonconventional systems

Historically, chemists have been using quantum-mechanical models to primarily study atomic spectra and electronic properties of simple molecules.<sup>124</sup> The molecular quantum chemistry bias is reflected in the techniques and approximations employed by many quantum-mechanical formalisms, such as approximating the system wave function as linear combinations of atomic orbitals (LCAO), or by approximating the shape of the atomic orbitals to Slater-type or Gaussian-type orbitals. Lately, methods that are better suited for the study of surface self-assembly or materials with a strong periodic character have been optimized.<sup>124</sup>

### Electronic structure methods and $\hat{\mathcal{H}}$

The general form for the Hamiltonian  $\hat{\mathcal{H}}$  in the Schrödinger equation (2.1) for a system containing  $n$  electrons (with mass  $m$  and charge  $-1$ ), and  $N$  nuclei, each with respective mass  $M_j$  and charge  $Z_j$ , can be written as<sup>125</sup>

$$\hat{\mathcal{H}} = - \sum_{i=1}^n \frac{\hbar^2}{2m} \nabla_i^2 - \sum_{j=1}^N \frac{\hbar^2}{2M_j} \nabla_j^2 - \frac{1}{4\pi\epsilon_0} \left( \sum_{i=1}^n \sum_{j=1}^N \frac{Z_j e^2}{r_{ij}} - \sum_{i=1}^n \sum_{j>i}^n \frac{e^2}{r_{ij}} - \sum_{i=1}^N \sum_{j>1}^N \frac{Z_i Z_j e^2}{r_{ij}} \right). \quad (2.22)$$

The first and second term in Equation (2.22) correspond to the kinetic energies of electrons and nuclei respectively, followed by the electrostatic attraction between electrons and nuclei (third term), electron-electron repulsion (fourth term) and repulsion between nuclei (fifth term).

The difficulty in solving Equation (2.1) comes from the explicit treatment of individual electrons and nuclei within the form of the Hamiltonian,  $\hat{\mathcal{H}}$ .<sup>125</sup> Firstly, the great difference between the masses of electrons and nuclei translates into vastly different time scales for the nuclear and electronic kinetics, rendering a stiff character

to Equation (2.1). This issue is solved with the Born-Oppenheimer approach, where the nuclear velocities, being so much smaller than those of electrons, are set to zero, removing the second term in Equation (2.22) and turning the fifth term into a constant:

$$\hat{\mathcal{H}} = - \sum_{i=1}^n \frac{1}{2} \nabla_i^2 - \sum_{i=1}^n \sum_{j=1}^N \frac{Z_j}{r_{ij}} + \sum_{i=1}^n \sum_{j>i}^n \frac{1}{r_{ij}}, \quad (2.23)$$

where Hartree atomic units were used:  $4\pi\epsilon_0 \equiv 1$ ,  $m \equiv 1$ ,  $\hbar \equiv 1$ . This renders the wave function dependent on electronic coordinates only. With this Hamiltonian expression, the solution to the Schrödinger equation represents the electronic structure for a given, static nuclear geometry.

Secondly, the last term in Equation (2.23), corresponding to electron-electron interactions, is in practice (and in principle) difficult to manage. The way in which this term is managed conceptually and numerically defines the degree of sophistication, computational cost and accuracy a given method provides.

## 2.4.1 The Hartree-Fock Method

### The independent-particle model

One way to deal with the difficulties posed by electron-electron interactions is to ignore them altogether. This approach is unacceptably crude, but it does serve as a basis for better techniques to build on. If the motion of each electron is considered independent of all others, then the third term in Equation (2.23) vanishes and  $\hat{\mathcal{H}}$  reduces to the sum of one-electron core Hamiltonians  $\hat{\mathbf{h}}_i$ . Equation (2.2) decomposes into  $n$  independent, one-electron wave equations, one for each electron  $i$ , with the corresponding solution  $|\psi_k(i)\rangle$ .<sup>125</sup>

$$\hat{\mathbf{h}}_i |\psi_k(i)\rangle = \epsilon_k |\psi_k(i)\rangle. \quad (2.24)$$

The system wavefunction  $|\psi_0\rangle$  is the product of all one-electron solutions  $|\psi_k\rangle$ :

$$\left( \sum_{i=1}^n \hat{\mathbf{h}}_i \right) |\psi_0\rangle = E_0^{\text{tot}} |\psi_\alpha(1)\psi_\beta(2)\dots\psi_\omega(n)\rangle. \quad (2.25)$$

Having to account for the spin of electrons, each orbital  $|\psi_k\rangle$  may be transformed into a spin-orbital  $|\phi_k\rangle$  by multiplication with an orthonormal spin function.<sup>125</sup> The

Pauli exclusion principle is accounted for by imposing an antisymmetric character to  $|\psi_0\rangle$  (flipping the sign when any two electrons are interchanged).  $|\psi_0\rangle$  can then be rewritten as a Slater determinant:

$$|\psi_0\rangle = \frac{1}{\sqrt{n!}} \det |\phi_\alpha(1)\phi_\beta(2)\dots\phi_\omega(n)| \quad (2.26)$$

### The mean field approximation

In the Hartree-Fock formalism the electron-electron repulsion is accounted for by having each electron interact with the average field of the remaining electrons.<sup>123;125</sup> The improved wavefunction  $|\psi\rangle$  has the Slater form of  $|\psi_0\rangle$ , for which the attraction between each electron and all nuclei is maximal and the repulsion between each electron and the average electron field of all others is minimal. Since the average field is influenced by the forms of the spin-orbitals, and the spin-orbital solutions themselves depend on the field, finding  $|\psi\rangle$  involves an iterative procedure in which the field converges to a self-consistent solution and the energies of the spin-orbitals converge to minimum values.

### The Fock, Coulomb and exchange operators

Towards this goal, a series of Hartree-Fock equations are constructed, one for each spin-orbital,<sup>125</sup>

$$\hat{\mathbf{f}}_i |\phi_k(i)\rangle = \epsilon_k |\phi_k(i)\rangle. \quad (2.27)$$

Equations (2.27) are analogous to (2.24), but whereas the Hamiltonian of the latter includes no electron interactions, the Fock operator  $\hat{\mathbf{f}}_i$  accounts for the mean electron field:

$$\hat{\mathbf{f}}_i = \hat{\mathbf{h}}_i + \sum_{u=\alpha}^{\omega} \left( \hat{\mathbf{J}}_u(i) - \hat{\mathbf{K}}_u(i) \right). \quad (2.28)$$

$\hat{\mathbf{J}}_u(i)$  is the Coulomb operator and  $\hat{\mathbf{K}}_u(i)$  the exchange operator, corresponding to the Coulombic repulsion among electrons, and the energy modification due to spin correlation effects, respectively. The sum in Equation (2.28) amounts to the average potential energy of electron  $i$  due to the presence of all other electrons. Individual spin-orbitals are determined by solving the  $n$  one-electron Schrödinger equations with a given set of Fock operators  $(\hat{\mathbf{f}}_i)_{i=1}^n$ . The shape of the new spin-orbitals modifies the mean electron field, which in turn changes the form of the operators, yielding a new

set of spin-orbital equations to be solved. The process is iterated in this manner until a self-consistent solution is found.

### **Basis sets for molecules and materials**

Like the Hamiltonian operator, the Fock operator is a well-defined Hermitian operator, adopting an infinite number of eigenfunctions as spin-orbital solutions, with associated energy  $\epsilon_k$ . With a complete basis set, the self-consistent field method converges towards the true wavefunction of the system. In practice, however, only a finite number of spin-orbitals can be considered in calculations. This way, the wavefunction  $|\psi\rangle$  has contributions from bases included in the selected, finite set only. In general, the larger the selected basis set, the better  $|\psi\rangle$  approximates the true form. The type or shape of basis functions is equally important, as the better the chosen orbitals are in reproducing the solution, the smaller the set size required for a given target accuracy.

An example highlighting this principle is the widespread use of atomic orbitals as the base for molecular orbitals. The linear combination of atomic orbitals approach (LCAO) is successful in approximating the electronic structures for small, localized molecules, because the formation of bonds is a perturbation on the atomic orbitals of the isolated atoms, in the presence of other atoms. For such classic chemistry cases, the molecular wave functions retain a large character of the base atomic orbitals, and only a small number of the latter are sufficient for reproducing the former with sufficient accuracy. This approach is not as efficient for solid-state materials.<sup>124</sup>

### **2.4.2 Density Functional Theory**

Using the Hartree-Fock approach to determine spin-orbitals as bases for molecular wavefunctions is limited by the computational challenges arising from calculations with large basis sets. Computational speed can be improved by using Gaussian-type orbitals (GTO) as basis sets, but the mathematical properties that make GTOs computationally preferable also cause them to inaccurately represent the behaviour of the wavefunction near the nucleus, and the drop-off rate far from the nucleus. Two or more GTOs may be combined for the construction of spin-orbitals with better shapes, a gain in accuracy which comes at the cost of dealing with more GTOs. Secondly, while HF methods account for electron-electron interactions, they do so in an average fashion, excluding instantaneous interactions and other quantum-induced effects, col-

lectively referred to as electron correlation. An alternative to HF is density functional theory (DFT), which is built around the concept of electron probability density  $\rho(\mathbf{r})$ , and the associated electronic energy functional  $E[\rho]$ .<sup>123–126</sup> DFT addresses electron correlation with significantly less computational effort than HF methods. While the wavefunction approach for an  $n$ -electron system involves  $4n$  variables, the electron density is independent of the system size. Finding the density-energy functional is not a straightforward task and represents the goal of a DFT method.

### The Hohenberg-Kohn theorem

Given an  $n$ -electrons system,  $\rho(\mathbf{r})$  is the electron density at point  $\mathbf{r}$ , integrating  $\rho$  over a volume determines the number of electrons over the volume, and  $\rho$  as a whole defines the electron distribution of the system across all space. The Hohenberg-Kohn theorem states that for a given  $\rho$ , the ground-state electronic energy  $E$  is uniquely defined.<sup>126</sup> In other words, there exists a functional  $E$ , which takes as argument a given electronic distribution in the form of  $\rho$ , and returns the ground-state energy of the system. Thus, the search for a lowest-energy wave function can be replaced with the search for the lowest-energy electron density, through a similar variational method. While the theorem proves the existence of such a functional, it does not offer a recipe for determining it. Attempts at expressing all energy components as explicit functionals of electron density yielded initial DFT methods with poor performance.

### The energy-density functional

The three main components defining the energy-density functional are the kinetic energy  $T[\rho]$ , the attraction between the nuclei and electrons  $E_{\text{ne}}[\rho]$  and the electronic interaction  $E_{\text{ee}}[\rho]$ :<sup>125</sup>

$$E[\rho] = T[\rho] + E_{\text{ne}}[\rho] + E_{\text{ee}}[\rho] = T[\rho] + E_{\text{ne}}[\rho] + J[\rho] + K[\rho], \quad (2.29)$$

where  $E_{\text{ee}}[\rho]$  can be further expressed as the Coulomb repulsion  $J[\rho]$  and exchange  $K[\rho]$  components, in line with the Hartree-Fock formalism. Out of the four density functionals in Equation (2.29) the kinetic  $T[\rho]$  and exchange  $J[\rho]$  components are difficult to express in general. In special circumstances the two functionals adopt proper forms, such as in the case of the Thomas-Fermi-Dirac (TFD) model, which assumes that electrons form a uniform electron gas. The TFD model is adequate

for valence electrons in certain periodic systems, such as metals, but is incapable of reproducing atomic shell structures, molecular bonding or the correct decay of electron density away from nuclei.

### The Kohn-Sham orbitals

Not having the ability to express the kinetic functional  $T[\rho]$  in an exact form is a serious shortcoming of DFT, as this term makes up a majority contribution to the total energy. In an attempt to address the issue, Kohn and Sham introduced back into DFT a set of reference orbitals  $|\psi_i^{\text{KS}}\rangle$ ,  $n_o$  in total, to be populated by  $n_{\text{el}}$  electrons, which together reproduce the density for the system. The kinetic energy for electrons takes the form<sup>125</sup>

$$T = \sum_{i=1}^{n_o} f_i \left\langle \psi_i^{\text{KS}} \left| -\frac{1}{2} \nabla^2 \right| \psi_i^{\text{KS}} \right\rangle, \quad (2.30)$$

and the ground-state electron density at location  $\mathbf{r}$  can be determined from the Kohn-Sham (KS) orbitals with

$$\rho(\mathbf{r}) = \sum_{i=1}^{n_o} f_i |\psi_i^{\text{KS}}(\mathbf{r})|^2. \quad (2.31)$$

The coefficients  $f_i$  are the orbital occupation numbers, eigenvalues of the density matrix. They correspond to the number of electrons in the respective spin-orbital and adopt a value between 0 and 1. For the total number of electrons to be conserved  $\sum_{i=1}^{n_o} f_i = n_{\text{el}}$ .

### The non-interacting kinetic energy, correlation and exchange

To understand the advantage of introducing KS orbitals, two limiting cases have to be considered.<sup>125</sup> First, if electrons are assumed not to interact, all occupation numbers  $f_i$  become 1 for  $i \leq n_{\text{el}}$  and 0 for  $n_{\text{el}} < i \leq n_o$ , leaving only the first  $n_{\text{el}}$  orbitals to contribute to the density and the kinetic functional, which now have an exact form, according to equations (2.30) and (2.31) respectively. Secondly, in the limit where we wish to know the exact form for  $T$  and  $\rho$ , an infinite number of reference orbitals, each with corresponding  $f_i$ , must be included, that is,  $T_0[\rho] \rightarrow T[\rho_{\text{exact}}]$  and  $\rho \rightarrow \rho_{\text{exact}}$  only for  $n_o \rightarrow \infty$ . Because in practice the non-interactive component of  $T$  is much larger than the correlation component, the occupation numbers are close to unity for the first  $n_{\text{el}}$  orbitals, and very close to zero otherwise. Because of this, approximating the kinetic energy with that given by  $n_{\text{el}}$  non-interacting electrons leads to a remarkably

accurate solution for  $T[\rho]$ . The small correlation component ignored this way can now be incorporated into a correlation-exchange term along with  $K$ , and Equation (2.29) becomes

$$E_{\text{DFT}}[\rho] = T_{\text{noninter}}[\rho] + E_{\text{ne}}[\rho] + J[\rho] + E_{\text{XC}}[\rho]. \quad (2.32)$$

The correlation-exchange term,

$$E_{\text{XC}}[\rho] = (T[\rho] - T_{\text{noninter}}[\rho]) + (E_{\text{ee}}[\rho] - J[\rho]) \quad (2.33)$$

includes the kinetic correlation energy within the first term, and the potential correlation and exchange energy, together represented by the second term.  $E_{\text{XC}}$  is considered the term for residual errors due to energy contributions which cannot be accounted for exactly and are better approximated collectively.

### The Kohn-Sham formalism

The Kohn-Sham formalism consists of finding the set of Kohn-Sham orbitals  $(\psi_i^{\text{KS}})_{i=1}^n$  by solving the Kohn-Sham equations:<sup>125</sup>

$$\left[ -\frac{1}{2}\nabla^2 + V_{\text{N}}(\mathbf{r}) + \int \frac{\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}' + V_{\text{XC}}(\mathbf{r}) \right] \psi_i(\mathbf{r}) = \epsilon_i \psi_i(\mathbf{r}), \quad (2.34)$$

where the first three terms in the expression of the Hamiltonian correspond to the kinetic energy for non-interacting electrons, the electrostatic attraction between electron density and the nuclei and the Hartree-style electrostatic repulsion. The last term  $V_{\text{XC}}$  is the exchange-correlation potential, the functional derivative of the exchange-correlation energy,

$$V_{\text{XC}}[\rho] = \frac{\partial E_{\text{XC}}[\rho]}{\partial \rho}, \quad (2.35)$$

In practice, the functional dependence of  $E_{\text{XC}}$  on density is given a fixed, approximate form, and  $V_{\text{XC}}[\rho]$  is readily obtained from (2.35) and used in Equation (2.34) when solving for KS orbitals.<sup>126</sup>

The Kohn-Sham equations are solved iteratively for a self-consistent solution. The initial form for  $\rho$  is approximated in some manner, for example as a superposition of atomic electron densities, and  $V_{\text{XC}}$  is determined from the fixed chosen functional dependence of  $E_{\text{XC}}$  on the density. Solving of KS equations (2.34) follows, and the

solutions are used to determine an improved density with Equation (2.31). The cycle is repeated with the new density form, and ends upon convergence of both  $\rho$  and  $E_{XC}$ . The KS orbitals may be determined numerically or expressed in terms of a set of basis functions.

## 2.5 Treatment of Dispersion with DFT for Aromatic Compounds on Transition Metals

DFT is a great achievement because it dramatically lowers the computational cost, relative to post HartreeFock techniques, and there are a wealth of examples for which it provides accurate and invaluable insight, all across chemical, physical and biological systems<sup>127</sup>. There are, however, many systems for which classic DFT fails to properly describe behaviour, most notably in condensed matter physics, certain biological macromolecules and surface adsorption of some organic compounds. One such example is the adsorption of benzene on metal surfaces, for which standard DFT predicts a binding energy one order of magnitude lower than what is experimentally observed.

DFT fails to reproduce accurately systems in which long-range dispersive interactions play an important role, specifically, it does not properly describe the asymptotic decay of dispersive forces. Often times this contribution is irrelevant. Binding interactions, classically considered the covalent and ionic bond, are typically of a higher strength than non-covalent interactions, and therefore the former were the focus of accuracy improvements. Dispersive interactions can nevertheless contribute significantly, sometimes in unpredictable circumstances. Strong-binding aromatic compounds on metal surfaces were believed to be dominated by covalent bonds. Dispersion corrections to DFT are currently a thriving topic of computational research with massive advances made over the past years.

### 2.5.1 Classification of Dispersion Corrections

The inaccurate depiction of dispersive interactions lead to the development of a variety of corrections. These vary in their approach, accuracy and computational cost. A useful classification based on the level of approximation used was proposed by Klimes and Michaelides.<sup>128</sup> They establish that a DFT method may be in one of the five stages of description, from methods with incorrect form for dispersive forces, to methods that go beyond pairwise additivity.

### **Class 1: methods with incorrect description of long-range dispersive forces**

Within this class we find methods for which dispersive interactions are not treated well, particularly the long-range behaviour. These methods are fit for systems in which dispersive interactions are known to be unimportant, but they notably underestimate the binding energy for the rest. The standard DFT approximations for exchange-correlation (XC), namely local density approximation (LDA) and general gradient approximation (GGA), fall in the lowest part of class 1, since they make no effort to correct for long-range dispersive forces.

Higher up within class 1, we find methods which attempt to incorporate weak interactions. Some functionals are fitted to reproduce weak interactions, and these show improvement for systems similar to those they have been fitted to. For example, the Minnesota functionals PWB6K and M05-2X provide good accuracy for many hydrogen-bonded and dispersion-dominated complexes<sup>129</sup> and are recommended for large biological systems.

Methods using specially adapted pseudopotentials, where a correction for weak interactions is added within an already existing pseudopotential, have also been developed. The dispersion corrected atom-centered potential approximates the long range correlation with a contribution from atom-electron terms depending on optimized nonlocal higher angular momentum.<sup>130</sup> These effective potentials offer a good correction for electron correlation across low-density regions but need to be properly optimized. Local atomic potential is based on an explicit decomposition of van der Waals potential into atomic contributions of various forms.<sup>131</sup> The accuracy of these methods is unreliable in systems different than those used in benchmarking because the form of dispersion interactions is fundamentally incorrect.

### **Class 2: methods with simple, but correct asymptotic dispersion**

The  $1/(r^6)$  asymptotic behaviour is necessary for a correct description of dispersive interactions. This is difficult to describe by local functionals for which the electron density decays exponentially. The simplest way to correct this is to add an atom-pair  $C_6$  contribution to the energy functional. Within class 2 the tabulated fitting parameters  $C_6$  are element dependent, unchanging with the environment or state of the atom, and isotropic (direction independent). These methods are generally termed DFT-D and they pay virtually no extra computational cost relative to classic DFT. Because a  $C_6$  correction diverges for small distances there is a significant double

counting of correlation effects for strong-binding complexes. To account for this, an arbitrary damping function is required, which requires careful fitting. Grimme proposed a method tested on a variety of molecular complexes<sup>132</sup> and found that a steep damping eliminates the correlation double counting. A global scaling factor for the atomic coefficients removes the functional dependence and makes them universal. Computing the coefficients requires experimental data specific to the systems studied, such as ionization potentials and polarizabilities. This limits the set of elements to those thoroughly studied experimentally in similar conditions.

The DFT-D2 method proposed by Grimme<sup>133</sup> is based on DFT-D and uses a more systematic and universal formula for  $C_6$  coefficients, by relying on properties of isolated atoms. Thus DFT-D2 is available for a wider range of elements. The arbitrary choice for the damping factor is still an issue and the  $C_6$  coefficients predicted for heavier atoms are not ideal.

### **Class 3: environment-dependent correct asymptotic dispersion**

Constant  $C_6$  coefficients cannot account for changes in the environment of atoms or hybridization state changes, which can have significant effects on polarizability, and dispersion. An improvement over class 2 methods is having  $C_6$  change dynamically throughout the simulation according to the local environment of atoms. These methods still rely on tabulated  $C_6$  values, occasionally updated to reflect local changes.

The DFT-D3 method proposed by Grimme<sup>134</sup> uses dynamically-computed coordination numbers (CN) to interpolate between dispersion coefficients of atoms in different chemical environments. The CN-dependent coefficients are computed from first principles and the method is relying on less fitting parameters, but the interpolation is somewhat ad-hoc. The method is more universal and gives better agreement with experimental results for a wide variety of systems, from small molecules with heavy atoms to organic molecules adsorbed on heavy metals.

Tkatchenko and Scheffler proposed the vdW-TS method<sup>135</sup> in which the environment dependent dispersion coefficients are calculated from the effective atomic volumes. These in turn are calculated with the Hirshfeld partitioning of electron density. The difference between effective volume and that of free atoms is used to scale  $C_6$ . The accuracy of this method is dependent on the correct scaling of coefficients with atomic volume.

The Becke-Johnson or the exchange-hole dipole moment dispersion model<sup>136;137</sup>

takes into account not only effective volume-based  $C_6$  modification, but also changes due to asymmetric electron density, which can cause, through an XC-hole, a non-zero dipole and higher-order electrostatic moments. There is no clear way to select the XC-hole and this is known to affect the results. The method produces accurate  $C_6$  coefficients but the computational cost is relatively high compared to other methods within this class.

#### **Class 4: long range density functionals**

This class consists of functionals that do not rely on predetermined values for polarizabilities, effective volumes, or other parameters required to compute  $C_6$ . Class 4 methods determine the dispersion interaction directly from the electron density and are therefore more general in nature. This is done by adding a non-local (long range) or semi-local correlation functional to the total energy. The correction is assumed to have a pairwise form and ignores the medium between interaction points.

The most notable example in class 4 is the van der Waals density functional (vdW-DF), developed by Dion.<sup>138</sup> The XC energy within this method has three contributions:

$$E_{xc} = E_x^{GGA} + E_c^{LDA} + E_c^{nl}, \quad (2.36)$$

where  $E_x^{GGA}$  is the exchange energy from GGA,  $E_c^{LDA}$  is the correlation energy from LDA and  $E_c^{nl}$  is the additional long-range correlation component. The original vdW-DF overestimates long-range dispersion and this error has been reduced in improved versions such as vdW-DF2 by changing the exchange and non-local correlation with better versions.<sup>139;140</sup> Another factor is the significant computational cost required for determining  $E_c^{nl}$ , which is a double space integral, though there were successful attempts to reduce it considerably.

Sato and Nakai proposed the local response dispersion approach,<sup>141</sup> which is based on the local responses approximation of Dobson and Dinte, the modified dielectric model of Vydrov and van Voorhis and Becke-type atomic partition function. The atomic polarizabilities are obtained only from the ground state electron density and these are used for the dispersion coefficients. This way the costly double integral within  $E_c^{nl}$  is replaced with a much cheaper, double atomic summation, without sacrificing accuracy. The method is applicable to any geometry, free from physical constraints (no vdW radii necessary for  $C_6$  calculation) and is computationally very efficient. It does however require a damping function dependent on the vdW

radii. The method overestimates dispersion.

### **Class 5: methods that go beyond pairwise additivity**

The methods in classes 2 to 4 assume a pairwise additive form for dispersion. Because of this they ignore the screening effect that the medium between interacting atoms might have, as well as collective excitations. These effects generally give opposite contributions and cancel out, but when they are important, like in condensed matter systems, the contribution is non-trivial. Methods within class 5 attempt to account for the non-additive contribution.

The Axilrod-Teller-Muto formula can be used to extend the atom-centered pairwise approaches to include three-body interactions explicitly.<sup>134;142</sup> Other methods to account for non-pairwise-additivity include the many-body dispersion method,<sup>143</sup> using a model of coupled dipoles and the random phase approximation, which calculates correlation energy from orbitals by using the adiabatic-connection fluctuation-dissipation theorem. For the latter method the vdW interactions are naturally included in the correlation energy, so no empirical data is needed. The method is very computationally demanding. The Lifshitz-Zaremba-Kohn theory for the nonlocal Coulomb screening within the bulk,  $vdW^{surf}$  proves particularly useful for hybrid inorganic-organic systems.<sup>144</sup> The collective many-body response (screening) of the substrate electrons are considered in the determination of the  $C_6$  coefficients and vdW radii. This method retains the low cost of pair-wise additivity, while considering multi-body corrections.

The above classification is in no way complete and leaves out many subtleties about the correctness and efficiency of dispersion-inclusive methods. Nevertheless, it provides a quick and easy metric to assess the level of detail with which dispersion is considered and how much confidence one should have in the accuracy of a method. As a general rule methods in a higher level class give results in better agreement with experimental and ab-initio data, and are overall more transferable.

### **2.5.2 The Exchange-Hole Dipole Moment Dispersion Model**

The exchange-hole dipole moment dispersion (XDM) model has recently been recognized as one of the more theoretically sound and practically versatile approaches for treating dispersion interactions in DFT methods.<sup>145</sup> In this section we briefly go over the history and development of XDM as its relevance to hybrid materials is

significant.

### **The origins of the XDM model**

In 2005 Becke and Johnson proposed a framework in which dispersion interactions are treated as effects of the instantaneous dipole moment of an exchange-hole.<sup>146</sup> The concept of an exchange hole was introduced to aid in visualizing the effects of self-interaction correlation and exchange. For an electron, at a given position, the hole measures to depletion in probability of finding another electron with the same spin in its vicinity, relative to the total electron density. The electron and the associated hole add up to zero charge overall, but the hole is not spherically symmetric around the electron in general. The instantaneous dipole moment resulting from the asphericity of the hole is assumed to be the cause of dispersion effects. Calculating the  $C_6$  dispersion coefficients with this method requires polarizability data for like-system interactions, and the authors take a heuristic approach to deriving the coefficients for the unlike systems. With no other correlation components added, and without time dependence, the model yielded remarkably accurate  $C_6$  coefficients for both atomic and molecular systems.

### **XDM adapted for DFT**

The authors soon reformulated the original model, which relied on occupied Hartree-Fock or Kohn-Sham orbitals for determining the exchange-hole dipole moment, into a density-functional version, which depends on the total electron density instead. The new “meta-GGA” model performs just as well as the original and has the obvious computational advantage accompanying the reduction in system size.<sup>147</sup> Accounting for the quadrupole, octopole and higher moments analogously to dipole offered comparable accuracy for derived higher dispersion coefficients  $C_8$  and  $C_{10}$ .<sup>148</sup>

### **The rigour and versatility of XDM**

In 2007 Becke and Johnson revisited the model with the goal of adding more theoretical rigour to the originally heuristic method. For this they employed a derivation based on a second-order perturbation theory with a semiclassical evaluation of the relevant integrals.<sup>149</sup> The ground-up derivation, starting from the nature of Pauli and Coulomb correlation, and how they affect the instantaneous dipole moments, lead to  $C_6$  and higher coefficients agreeing very well with ab initio reference data.

The XDM model is one of the most successful techniques for treating correlation within DFT. It is used on a variety of systems, both molecular and solid-state, such as molecular crystals, layered material and metal surfaces, without the need for reparametrization of the damping function in each case.<sup>137</sup> Recent tests against benchmark calculations have shown that the XDM model performs exceptionally well when describing forces and energies outside equilibrium for the benzene dimer,<sup>145</sup> making it difficult to attribute its accuracy to fortuitous cancellation of errors at equilibrium.

### 2.5.3 Benzene On Metal Surfaces

Aromatic compounds on metal surfaces are a popular class of hybrid inorganic/organic systems (HIOS) to study. Along with the classical areas of applied research in surface science - corrosion protection, lubrication, dye adhesion - these systems show promise as components in opto-electronic devices, organic photovoltaics and sensors. The weak nature of the interaction between aromatic compounds and many metal surfaces makes the experimental characterization challenging, and to date little is known about the details of the interactions. On the theoretical side the inability of standard DFT methods to treat dispersive interactions at medium and long ranges is a particular concern for aromatic systems. Because of this HIOS are often used as reference systems for standard tests in dispersion-correction methods.<sup>150</sup>

The most commonly studied HIOS is benzene on [111] transition metals (Bz/M). Based on the adsorption strength Bz/M can be categorized into weak-binding (Bz/Mw), on Cu, Ag, Au surfaces and strong-binding (Bz/Ms), on Pd, Pt, Rh, and Ir surfaces. Bz/Mw have a flat potential energy surface are stable below 280K: 225K for Cu[111],<sup>151</sup> 220K for Ag[111],<sup>152</sup> 239K for Au[111],<sup>153</sup> 280K for Cu[110].<sup>154</sup> Bz/Ms are stable at room temperature and decompose benzene upon heating. For low coverages the benzene molecule lies flat or mostly flat on both strong- and weak-binding metals. STM studies show that benzene can easily diffuse on Cu[111] and Au[111] surfaces.

DFT studies on Bz/M date back to more than a decade ago,<sup>155</sup> but because of the dispersion problem, their conclusions were questionable. Improvements in the treatment of dispersion at all levels of approximation lead to a few universal findings, such as the separation into weak- and strong-binding systems. All methods show significant differences in the overall adsorption energies between coinage metals and the rest. Quantitatively, the adsorption energy values vary significantly, within an

order of magnitude for physisorbed systems. The contributions to total energy are not always correct even if the value itself is close to that of experiment.

Concerning the adsorption geometry, the discrepancies between methods are smaller. All methods with a good description of covalent interaction can distinguish between physisorbed and chemisorbed systems. Both types of adsorptions prefer the benzene molecule to form a bridge configuration with an angle of  $30^\circ$  between the C-C and M-M bonds (bri $30^\circ$ ). On coinage metals this preference is negligible relative to other flat-lying configurations (flat PES). On strong-binding metals bri $30^\circ$  is preferred and benzene loses the planar configuration, with hydrogen atoms tilting upwards.<sup>47;156</sup>

As a general rule, PBE and other class 1 methods give considerably less binding, and methods from class 2, 3 and even 4 overestimate the binding, with methods from class 5 showing a lot of promise, even when relying on very different approximations.

Bilic<sup>157</sup> studied Bz/M on Cu[111], Ag[111] and Au[111], comparing the GGA functionals PW91 (Perdew-Wang) and PBE (Perdew-Burke-Ernzerhof) with results from CASPT2<sup>158</sup> (based on the Moller-Plesset perturbation theory). He found good agreement in terms of adsorption distance and flat geometry, but a dramatic underestimation of adsorption energy. The comparison of DFT with CASPT2 revealed the covalent binding as insignificant and a massive correlation correction. Functionals like VASP-DFT, which predict a total binding energy within 20% of the correct value, were accounting for correlation as part of perceived covalent bonding. In other words, these functionals were giving the right answer for the wrong reasons. Since then studies have systematically discovered that class 1 methods are inadequate for Bz/M systems and describing correlation.

Grimme<sup>134</sup> compared DFT-D class 2 methods with the newer class 3 method DFT-D3 on Bz/Ag[111]. He found significant overbinding for both, with DFT-D3 offering superior results due to better  $C_6$  coefficients and increased cutoff radius. Toyoda<sup>159</sup> investigated benzene on coinage metals with class 4 vdW-DF and produced good adsorption energies but overestimated binding distances. Ruiz proposed a class 5 functional, vdW<sup>surf</sup><sup>144</sup>, which overbinds benzene on Cu, Ag, Pt, Pd. Liu<sup>47</sup> investigated both physisorbed (Bz/Ag, Bz/Au) and chemisorbed (Pt, Pd, Ir, Rh) systems and compared optB88-vdW (class 4), PBE+vdW<sup>surf</sup> (class 5) and PBE-LDA (class 1). Both high-class methods give excellent agreement with one another and with experimental data. On Bz/Ms chemisorbed systems vdW-DF and vdW-DF2 give smaller adsorption energies than PBE. Nondispersive functionals like PBE-LDA underestimate adsorption energies for Bz/Mw, to a lesser extent than for Bz/Ms. The

vdW contribution to Bz/Ms was originally thought to be small and was subsequently shown to be larger than for Bz/Mw.

In a follow-up work Liu<sup>45</sup> compares PBE+vdW<sup>surf</sup> (class 5) and PBE+vdW (class 3), with a focus on the higher dispersion contribution in covalently bonded systems. He finds that much of the vdW contribution is short distance (3–4 Å), and is therefore higher for a shorter-distance adsorption, such as Bz/MS. It is also discovered that the PBEvdW<sup>surf</sup> method slightly overestimates interaction between metallic electrons in the bulk.

In a comprehensive study<sup>156</sup> Carrasco investigates Bz/Mw and Bz/Ms with a variety of methods: PBE (class 1), PBE+vdW (class 3), PBE+vdW<sup>surf</sup> (class 5), vdW-DF (class 4), vdW-DF2 (optPBE, optB88, optB86b, class 4). The general error trends for classes 1-3 methods are noted, as well as the dependence on binding-strength: destabilizing correction for Bz/Mw, stabilizing for Bz/Ms. The choice of the underlying functional for vdW-DF matters for accuracy. The methods tend to overestimate all values with respect to experimental data by being too repulsive at short interatomic distances. Optimized versions of functionals such as optB88-vdW elevate vdW-DF to a status equivalent with that of PBE-vdW<sup>surf</sup> in terms of delivered adsorption energies and equilibrium distances, despite the two methods having fundamentally different approaches. The methods agree well with experimental data. Version comparisons of vdW-DF (vdW-DF, vdW-DF2, vdW-DF2-Klimes) for Bz/Au reveal little differences among methods in terms of binding energies.<sup>160</sup> All three overestimate geometric parameters.

### 2.5.4 Larger Aromatic Compounds on Metal Surfaces

Bz/M systems contain the smallest aromatic molecule, yet cannot be properly investigated without a good description of dispersive forces. For larger aromatic molecules the problem of accuracy of dispersion is at least as important as it is for benzene. The discrepancy between experimental and DFT-calculated heats of adsorption increases with molecular size.<sup>155;161</sup> Liu tested DFTvdW<sup>surf</sup> on a variety of organic molecules, including naphthalene, and found the method to account for the binding energy difference.<sup>162</sup> Toyoda<sup>163</sup> compared PBE, DFT-D and vdW-DF for pentacene on Cu[111], Ag[111] and Au[111] and reached similar conclusions to those for Bz/M respective systems: virtually no binding for classic GGA, overbinding in DFT-D for heavy metals (for which  $C_6$  coefficients are difficult to deduce properly) and good

experimental agreement for vdW-DF. The molecule-metal distances in the case of DFT-D were only in excellent agreement for Cu.

These studies suggest that the current dispersion-corrected DFT methods scale well with the size of the aromatic compounds and that the quantitative accuracy that DFT has gained in the past years can be assumed for larger aromatic molecules. Methods within class 4 and 5, such as DFTvdW<sup>surf</sup>, vdW-DF and VdW-DF2 will likely give the same level of description and accuracy for large aromatic molecules as they do for benzene, at least as far as vdW interactions are concerned. As the experimental focus for HIOS shifts towards larger, multi-functional, chiral and physisorbed molecules, the theoretical breakthroughs in DFT methods continue to provide invaluable predictions and insights into the nature of these materials.

## 2.6 Electronic Structure Calculations in Solid State Materials

Since the inception of quantum theory, crystallographers and solid state physicists have developed conceptual and numerical techniques aimed at successfully applying it to the study of periodic crystals, while quantum chemists focused their efforts to modelling finite molecular systems. Both teams have independently developed a repertoire of ingenious techniques for simplifying the complexity of problems arising in predicting the properties of their respective systems of interest. As the range of crystallography grew to include larger unit cells with richer localized features and the range of molecular modelling shifted towards molecules and collection of molecules of increasing size, a convergence of interests occurred when solid-state materials captured the attention of both fields. The drive to accurately model and predict the electronic properties of systems with some periodic character and a strong local component has forced material scientists, crystallographers and computational chemists to combine their techniques and insights into highly versatile models, breaching the somewhat arbitrary barriers that separated the fields for decades.

In what follows we complement our picture of quantum chemistry with a series of concepts uniquely useful for solid-state materials. The successful methods used for crystals are generally compatible, and can be used in conjunction, with those specific to molecules, when investigating hybrid systems. Heterogeneity, however, requires the use of highly accurate and versatile methods for dealing with dispersion interactions,

the appropriate choice of basis sets, as well as techniques for minimizing errors in energy calculations. Finally, we require appropriate techniques for interpreting the electronic structure data generated in our calculations.

### 2.6.1 Wave Functions in Periodic Systems

Determining the electronic properties of bulk crystalline structures poses a markedly different problem than that of describing a molecular system. Solving Schrodinger's equation for a crystal extending periodically and indefinitely in three dimensions, with a classically molecular approach, is unfeasible on the basis of system size alone. Fortunately, by exploiting the translational symmetry of crystals, in a manner analogous to the use of point symmetry in molecules, the computational challenge is reduced drastically, and the crystal's wavefunction and electronic bands can be determined with remarkable accuracy.

In the context of periodicity, the problem of modelling an infinite (or very large) system in real space is transformed in a finite k-point sampling of a small zone of the reciprocal space. Metals need additional methods employed for determining the Fermi energy and Fermi surface. Physically interpreting the results of electronic structure calculations for periodic systems requires concepts like band structure, density of states and atom- and orbital-projected density of states. The methods developed for symmetric crystals can be transferred, with some reservations, to less symmetric systems with a periodic component.

#### The unit cell of a crystal

Given a periodic crystal, the unit cell is the smallest component that can fully describe the arrangement of atoms in the whole crystal. Given three vectors  $\mathbf{a}_1$ ,  $\mathbf{a}_2$  and  $\mathbf{a}_3$  representing the 3D volume of the unit cell, a lattice may reproduce the crystal if the points  $\mathbf{R}$  on the lattice satisfy the equation

$$\mathbf{R} = n_1\mathbf{a}_1 + n_2\mathbf{a}_2 + n_3\mathbf{a}_3, \quad (2.37)$$

where  $n_1$ ,  $n_2$  and  $n_3$  are some integers. The position of atoms within the unit cell represents the crystallographic basis. Unit cells must be space-filling, allowing no voids or overlaps in the periodic repetition of the cell. A lattice corresponding to a space-filling cell satisfying Equation (2.37) is called a Bravais lattice.

### Bloch's theorem

The action of a symmetry operation on the wave function  $\psi(\mathbf{r})$  describing the crystal can be represented with the aid of linear algebra on real-space vectors. If translation by a lattice vector is applied to  $\psi(\mathbf{r})$ , then there exist a vector  $\mathbf{k}$ , such that the transformation of  $\psi$  by  $\mathbf{T}_{\mathbf{R}_n}$  is equivalent to multiplying it by a phase factor,

$$\mathbf{T}_{\mathbf{R}_n}(\mathbf{r}) = \psi(\mathbf{r} + \mathbf{R}_n) = e^{(i\mathbf{k}\cdot\mathbf{R}_n)}\psi(\mathbf{r}) \quad (2.38)$$

In other words, the wave function of a periodic system must have the form of a product between a lattice-periodic function  $\mathbf{u}(\mathbf{r} + \mathbf{R}) = \mathbf{u}(\mathbf{r})$ , and a plane wave,<sup>164</sup>

$$\psi_{\mathbf{k}} = e^{i\mathbf{k}\mathbf{r}}\mathbf{u}_{\mathbf{k}}(\mathbf{r}), \quad (2.39)$$

where the vector  $\mathbf{k}$  may be considered a “quantum number”, determining how the wave function satisfies the periodicity of the crystal. The lattice-periodic function  $\mathbf{u}_{\mathbf{k}}(\mathbf{r})$  depends on  $\mathbf{k}$ , but the dependence is usually weak.

The notable consequence of Bloch's theorem is that the extended wave function, or crystal orbital, can be reduced to its repeating unit cell components and thus generated entirely from within the unit cell.

### Reciprocal lattice and the first Brillouin zone

The vector  $\mathbf{k}$  in Equation (2.39) has the dimensions of inverse length, indicating that the reduction of the wave function to the unit cell component resulting from Bloch's theorem is related to the reciprocal space. Given any lattice, a reciprocal counterpart may be derived by associating to each vector  $\mathbf{R}$  on the real lattice satisfying Equation (2.37) an inverse vector  $\mathbf{B}$  of components  $\mathbf{b}_1$ ,  $\mathbf{b}_2$  and  $\mathbf{b}_3$ , such that

$$\mathbf{a}_i\mathbf{b}_j = 2\pi\delta_{ij}. \quad (2.40)$$

Vectors  $\mathbf{R}$  and  $\mathbf{B}$  are relative inverses as  $e^{i\mathbf{R}\mathbf{B}} = 1$ . The unit cell of the reciprocal lattice defined by vectors of type  $\mathbf{B}$  is called a Brillouin zone, and its primitive unit cell is referred to as the first Brillouin zone. It follows from Bloch's theorem that a complete description of the electronic structure requires the description of the wave function for  $\mathbf{k}$  values from the first Brillouin zone only.<sup>164</sup>  $\mathbf{k}$  points outside this range do not provide any additional information, due to the invariance of  $\mathbf{u}_{\mathbf{k}}(\mathbf{r})$ .

## The $\mathbf{k}$ vector

The  $\mathbf{k}$  vector, which is sometimes termed the crystal momentum, is related to the energy of the associated wave function.  $\mathbf{k}$  may be thought of as a frequency factor, with high values indicating rapid oscillations. Its value is severely restricted by the size and symmetry of the unit cell, as only those oscillations that do not break the symmetry of the crystal are permitted. The highest energy  $\mathbf{k}$  vector considered in an electronic structure calculations determines both the number of wave functions included in the energy calculation and the maximum kinetic energy that can be represented.

## $\mathbf{k}$ -sampling

As the cell extends in three dimensions to form the crystal, “higher”  $\mathbf{k}$  values become closer together within the Brillouin zone. The problem reduces to the accurate sampling of the first Brillouin zone for various  $\mathbf{k}$  values, or  $\mathbf{k}$ -sampling. And since the lattice-periodic part of the wave function  $\mathbf{u}_{\mathbf{k}}(\mathbf{r})$  depends only weakly on  $\mathbf{k}$ , it is sufficient to sample the Brillouin zone at a finite, and rather small, number of points.

Nevertheless, depending on the cell size and the property of interest, very dense  $\mathbf{k}$ -point meshes may be required. For example, in calculations of phase diagrams with guaranteed accuracy for all phases with differently sized and shaped unit cells,  $\mathbf{k}$ -point densities as high as  $5,000 \text{ \AA}^{-3}$  must typically be employed.<sup>165</sup> Recently, adaptive schemes based on machine learning can automatically select a suitable sampling density and grid, based on the property of interest.<sup>166</sup>

## System size and $\mathbf{k}$ -sampling size

In general, getting an accurate total energy of the crystal, the values for  $\mathbf{k}$  used in the calculation of  $\psi_{\mathbf{k}}(\mathbf{r})$  should lie on a fine three-dimensional mesh within the Brillouin zone. A small unit cell in real space requires a small number of basis sets for the construction of Bloch functions, but corresponds to a large Brillouin zone, and a large number of  $\mathbf{k}$  points is needed for a smooth representation of the first Brillouin zone. Conversely, large unit cells require a large number of basis sets to represent local structures within the cell, but a very small number of  $\mathbf{k}$ -point sampling to accurately reproduce their associated small Brillouin zone.

For solid state materials the trade-off between the system/basis set size and the  $\mathbf{k}$ -point sampling size is notable, as they often include both important localized atomic

structures and a relevant periodic component<sup>124</sup> (this being the reason why they manifest interesting electronic properties). For band-structure calculations, the product of the number of atomic orbitals in the real unit cell and the number of  $\mathbf{k}$  points in the reciprocal space is an important quality criterion which controls the accuracy of results and is limited by computer resources.<sup>124</sup>

### Fast Fourier transform application

The concrete way to sample the Brillouin zone is based on the Fourier expansion of the lattice-periodic function  $F(\mathbf{k})$ , into a finite number of Fourier components,<sup>167</sup>

$$F(\mathbf{k}) = \sum_{j=1}^m F_j e^{i\mathbf{k}\mathbf{R}_j}, \quad (2.41)$$

where  $F(\mathbf{k})$  may contain implicit dependencies on the chosen wave functions  $\psi_{\mathbf{k}}(\mathbf{r})$  and eigenvalues  $\epsilon_i(\mathbf{k})$ . The function is expanded into Fourier components up to some  $\mathbf{R}_m = (x_m, y_m, z_m)$ . The integral of  $F(\mathbf{k})$  over the whole Brillouin zone is given by the lowest Fourier component  $F_0$ , and this integral is approximated by a finite sum over  $\mathbf{k}$ -points within the first Brillouin zone. The vanishing of Fourier components beyond  $\mathbf{R}_m$  can only be assumed to occur for semiconductors and insulators. Metals have high Fourier components in  $F(\mathbf{k})$ , because the occupation of the bands changes rapidly on the Fermi surface.<sup>164</sup>

Equation (2.41) is the extended analogue of the molecular LCAO method, where the crystal orbitals  $F_j$ , weighted by mixing coefficients given by the Bloch exponential factor, generate the extended function of the periodic crystal.

### 2.6.2 Basis Set Considerations

The right choice for the basis set used to expand the wave function enables accurate reproduction of the system function with only a small number of bases. The trade-offs present when making this choice are:

- (i) the larger the basis set, the better the representation of the wave function;
- (ii) improving the shape of the base functions lowers the required size of the base;
- (iii) “better”-shaped base functions tend to be more computationally demanding;
- (iv) computational efficiency and range of applicability are inversely correlated, of particular note for solid-state, heterogeneous materials.

In terms of shape, the two main types of localized basis used are STOs, preferred for their hydrogen-like atom radial dependence, and GTOs, widely popular because of their computational advantage. The gain in efficiency with the use of GTOs is partially traded for a larger number of functions required to overcome their inferior radial representation.<sup>123</sup>

One, two, three or multiple functions may be used for representing each AO, leading to minimal, double-zeta, triple-zeta or multiple-zeta basis sets respectively. Using multiple functions per AO allows for the inclusion of polarization, by adding a function with a higher angular momentum.<sup>123</sup>

The localized basis functions may be centered at the nuclei, at the center of bond or even between non-bonded atoms to improve calculations of van der Waals interactions.<sup>123</sup> Anions, systems with high electron density and/or strong polarizability are generally not modelled well by localized AOs, and in some cases this approach is particularly counterproductive. For example, the outer valence electrons in crystalline metals behave almost like free electrons in an electron gas. In these cases, using functions with an “infinite” range, aimed directly at the full system seems like a better choice.<sup>123</sup>

### **Plane wave basis sets**

In infinite periodic systems the energy spacing between molecular orbital levels vanishes and the result is an energy band. As Bloch’s theorem demonstrates, the electrons occupying these bands can be described by orbitals expanded in a basis set of plane waves. This set is constrained by the periodicity of the crystals’ lattice, as only the plane waves satisfying the crystal symmetry can contribute to the wave function. This constraint makes it easy to identify the candidates for the basis sets if the crystal lattice is known. Furthermore, the truncation of the basis set is determined by the kinetic energy constraint, predefined in advance. Only the reciprocal lattice vectors whose kinetic energy is lower than the cutoff are included in the expansion.

### **Challenges of plane wave basis sets**

Plane wave basis sets tend to be larger than typical GTOs sets, but their size depends only on the maximum cutoff energy and the size of the lattice, and not on the size of the system itself. This is contrasted with nuclear-centered GTO sets, whose size increases linearly with system size. For these reasons plane wave basis sets are

particularly favourable for describing large, electron-rich, periodic systems, with delocalized and slowly varying electron density, such as valence and conduction bands in metals.<sup>123</sup>

Describing core regions where electrons are strongly localized around the nuclei requires the incorporation of a large number of rapidly-oscillating functions in the wave expansion, and therefore a very large energy cutoff. In addition, it is impossible to describe the nucleus-electron singularity for core electrons with plane waves, a task which Gaussian-type functions can efficiently accomplish.

### **Pseudopotential methods**

Given the way in which localized AOs and plane waves complement each other in the ability to describe electronic structures, simultaneously describing both core and valence electrons in periodic systems requires using both types in a way that preserves their strengths and minimizes their weaknesses. The straightforward solution is to combine localized basis sets charged with the behaviour of core electrons, with plane wave basis sets for the description of valence electrons. This mixed basis set approach, collectively called *augmented plane wave methods*,<sup>168–170</sup> increases computational complexity as new integrals, involving different types of functions, are required. Additionally, the physical space must be partitioned into core and valence regions, a somewhat arbitrary choice which adds a continuity requirement at the boundary.

Alternatively, in the pseudopotential approximation method, the core energy levels from atoms are removed, and Pauli’s exclusion principle is enforced via a repulsive pseudopotential with orthogonal constraints. This removes the “wiggles” from AOs and allows for the efficient application of plane wave basis set expansion.

The pseudopotential method exploits the environmental independence of core region electrons and their negligible contributions to the chemical and electronic properties of a material. The pseudopotential replacing the Coulomb potential in the core region must be chosen to reproduce the atomic scattering properties and the Coulombic interaction from outside the core.<sup>171</sup> This approximation markedly improves the convergence of the Kohn-Sham states without a loss of accuracy, and is one of the most effective methods used to date in electronic structure calculations for solid state materials.

### 2.6.3 Electronic Structure Analysis

Analyzing the electronic structure data generated by DFT calculations is not a trivial task and the type of post-processing required depends heavily on the system and the sought-after information. One way to get an insight into electronic structure properties is to explore simple theoretical models, for which the electronic structure is easy to derive and the relationship to physico-chemical properties is straightforward. This approach is taken by Dronskowski, as he explores at length the consequences of increased complexity to the electronic structure.<sup>124</sup> As the electronic structure data becomes increasingly abstruse, he thoroughly introduces techniques designed to simplify the analysis.

The other approach is explored by Nørskov and Hammer, who focus their analysis on the change in electronic structure following a well-known phenomena.<sup>172</sup> Processes like surface adsorption are well suited for this type of analysis, and Nørskov and Hammer explore how atomic and molecular adsorption on a variety of surfaces yields meaningful electronic structure changes.

In the following section these two approaches are briefly presented. For a thorough exploration of either the reader is referred to the original materials by Dronskowski<sup>124</sup> and Hammer and Nørskov.<sup>172-176</sup>

#### Band structure

The band structure in a crystal is a consequence of how orbitals combine in the Bloch sum, with varying phase factors along the Brillouin zone. As a simple example put forward by Dronskowski,<sup>124</sup> in a one-dimensional system where both the lattice length  $a$  and  $k$  are scalars, the wavefunction  $\psi(k)$  is easy to evaluate from the zone center  $\Gamma$  to the zone edge  $\chi$ :

$$\psi(\mathbf{k}) = \sum_{n=1}^N e^{i\mathbf{k}n\mathbf{a}} \phi_n, \quad 0 \leq |k| \leq \frac{\pi}{a}. \quad (2.42)$$

In terms of interatomic overlap, at the two extremes  $\Gamma$  and  $\chi$  the wave functions result from an in-phase and out-of phase bonding of AOs with radial symmetry, leading to the strongest stabilization and anti-bonding interaction respectively. In the middle of the Brillouin zone the out of phase and equi-energetic  $\psi(\pi/2a)$  and  $\psi(-\pi/2a)$  linearly combine into crystal orbitals with different AO overlap, and distinct and maximal energies. With partial interference, the energy difference at other points within the Brillouin zone is intermediate. In the limit of an extended system,

these varying degrees of phase interference create a continuous band.

The dispersion of the band is increased by shorter interatomic distances and is proportional to the strength of the inter-atomic interaction between unit cells. Thus,  $s$  bands tend to be quite diffuse while  $d$  bands are notably contracted, the latter resembling the original AO shape even in compounds.

### Orbital symmetry and band topology

Systems with multiple types of orbitals manifest multiple bands, depending on the match or mismatch in symmetry and orientation between AOs and the crystal lattice. The topology of a band generally depends on the symmetry of the combining orbitals. In another example explored by Dronskowski, a one-dimensional chain system with a horizontal mirror plane symmetry element, he demonstrates that the radially symmetric  $s$  orbitals are in phase at the zone center and out of phase at the zone edge, with strong dispersion of energy between edges.<sup>124</sup>  $p_x$  and  $p_y$  are in phase at  $\Gamma$  forming  $\pi$ -type bonding and out of phase at  $\chi$ , where they form  $\pi^*$ -type antibonding. Because of the smaller overlap in  $\pi$ -type interactions, the dispersion of the resulting band is small. The band topology coming from  $p_z$  is reversed, and the dispersion is rather large compared to bands from  $p_x$  and  $p_y$ . In this scenario an  $ns$  and  $np_z$  orbitals are symmetric with respect with the mirror plane, their bands have identical symmetry properties and they interact throughout the entire zone from  $\Gamma$  to  $\chi$ , leading to  $s$ - $p$  mixing throughout the zone.

### Covalent bonds

With covalent bonds present in the lattice, the corresponding bonding and antibonding molecular orbitals are expanded into bands by the interaction with the cell unit.<sup>124</sup> The low-energy, bonding and high-energy, antibonding MOs have different symmetry properties, and so will their respective bands. For this reason, they belong to different point-group representations, cannot mix in principle and do not interact throughout the Brillouin zone. Generally speaking, the bandwidth of the antibonding band is greater.

Beyond the simple examples provided by Dronskowski, it is not immediately obvious from electron densities how to identify chemical bonding. The resulting electron densities of most real systems very closely resemble densities representative of their atomic constituents. Chemical bonding in general does strongly reshape the electron

density, this being the reason why the LCAO method is so successful in generating molecular and crystal orbitals.<sup>124</sup>

### Challenges in electronic structure analysis

Bloch's theorem applied to periodic materials reduces the problem of electronic activity in infinite systems to determining the right finite basis set of crystal orbitals. The  $\mathbf{k}$ -dependence of the crystal orbitals makes the electronic structure difficult to interpret. Unlike MOs, crystal orbitals adopt complex values and are not straightforward to combine and graphically represent as bands.<sup>124</sup>

Solutions to this problem require averaging formalisms that make the electronic structure transparent for the lattice in real space or comparative density analysis with energy-resolved indicators of chemical bonding.<sup>124</sup>

### Density of states

In the density of states (DOS) approach discussed at length by Dronskowski, all direction-related electronic information is discarded in favour of yielding an energy-resolved indicator of chemical bonding. The density of states  $P(E)$ , given by the reciprocal gradient of the band structure, represents the one-electron energy levels present in an infinitesimal energy interval,  $dE$ .<sup>124</sup> For a one-dimensional system  $P(E)$  has the form

$$P(E) \sim \left( \frac{dE}{dk} \right)^{-1} \sim v(k)^{-1}. \quad (2.43)$$

$P(E)$  is given by the slope of the band in question with steeper/flatter bands yielding lower/higher DOS.  $P(E)$  is also inversely proportional to  $v(k)$ , the velocity of an electron in a band. Electrons in a sharp DOS cannot easily move through the crystal and can be considered "slow", while electrons associated with a wide DOS are rapidly moving within the crystal.

$P(E)$  averages over the entire  $\mathbf{k}$ -space and manifests singularities at lowest and highest energies. Sharp DOS result from flat, atomic-like bands, from atomic orbitals which do not overlap much with neighbouring orbitals. Wide DOS indicate the presence of a steep band and strong interatomic coupling.

## Electron and energy partitioning

For molecular systems, the Mulliken population analysis is widely used for finding the electron distribution from LCAO-MO, by heuristically partitioning the available electrons into atom-centered (net populations) and bond-centered (overlap populations).<sup>124;177</sup> The method is limited, as the assignment of net and overlap populations is somewhat arbitrary, multi-centered bonds are ill-described and the overlap populations may take unreasonable (including negative) values. Population values are also strongly dependent on the choices of the basis set. These limitations stem from the fact that quantum mechanical systems are not, in general, decomposable into atoms and bonds.<sup>124</sup>

Mulliken population analysis is nevertheless valuable for crystals, most of which have unit cells with multiple AOs. Periodicity causes the mixing coefficients used in the population analysis to be  $\mathbf{k}$ -dependent, which is why they are converted to a density matrix. Subsequent integration in  $\mathbf{k}$ -space yields the desired average and the ability to calculate the populations. The partitioning of electrons is executed in an explicitly energy-dependent way, by differentiation of the density matrix, yielding the DOS matrix.<sup>124</sup>

## Projected density of states

The projected density of states (PDOS) can be generated from the DOS matrix through partitioning the number of electrons into atomic contributions (projecting the electrons onto atoms), in a manner similar to the case for molecules. Local or atom-projected DOS plots are generated by taking the sum of the elements of the DOS times the overlap matrix, corresponding to energy-dependent gross populations.<sup>124</sup> The projection can be performed on atomic groups or sets of well-defined group orbitals. Like DOS, PDOS represent  $\mathbf{k}$ -averages over the entire Brillouin zone and have all the electronic directional information discarded.

## Crystal orbital overlap population

Crystal orbital overlap population (COOP) analysis is similar to DOS analysis, additionally providing a quantum-chemical measure of the bond strength between selected neighbouring atoms. The advantage offered over the exclusive partitioning into atom-centered and between-atoms contributions, is that it enables the prediction of energetically advantageous electron filling, and therefore the reasonable positioning

of the Fermi level. COOP analysis is an important quantum-chemical tool for band structure calculations in solid-state materials chemistry, offering a better alternative to the previously over-used ionic model.<sup>124</sup>

### **Crystal orbital Hamilton population**

One overlap-independent derivative of COOP is the crystal orbital Hamilton population (COHP) analysis. The short-ranged basis sets used in parameter-free methods and first-principles DFT may be subject to various transformations changing the shapes of basis functions and introducing basis set dependence.<sup>124</sup> In COHP an alternative partitioning scheme based on the electronic energy is used and the overlap integrals are replaced with Hamiltonian matrix elements. COHP is useful for identifying bonds in complex materials simulations, as COOP calculations are performed as an average over bonds.

### **Transition metals surfaces**

The following four subsections highlight some of the findings by Nørskov and Hammer upon investigating adsorption phenomena on transition metals with electronic structure tools.<sup>172-176</sup> A one-electron state in atom or molecule outside a metal surface interacts with all the valence states of the surface atoms, the latter forming one or several electronic bands. Transition metals typically manifest a half-filled  $s$  band with a broad character and a much narrower  $d$  band. The interaction between the adsorbate and  $s$  electrons causes the adsorbate state to broaden and is responsible for the weak chemisorption phenomenon. While the  $s$  band is half-filled, the occupancy of  $d$  varies between metals. The  $d$  band is responsible for strong chemisorption and splits the state of the adsorbate into bonding and anti-bonding states.

### **Atomic adsorption**

The process of adsorption on transition metals may be separated into two steps. The electron in the adsorbed atom first interacts with the  $s$ - $p$  states of the metal, having the gas-phase state broaden and shifted down in energy upon adsorption. The adsorbate further interacts with the  $d$  narrow band of the metal, forming renormalized covalent bonding and anti-bonding states. The  $s$ - $p$  binding is predominant in noble metals.

In the case exemplified by chemisorbed atomic oxygen, Nørskov and Hammer demonstrate that the DOS has a similar structure between transition metals.<sup>172</sup> The two main peaks originate from: (i) the  $2p$  oxygen state interacting with the  $s$  electrons in metal, giving rise to a single resonance below the Fermi level and (ii) the strong interaction with the narrow  $d$  band, which splits the oxygen resonance into bonding and antibonding states. In the case of molecular adsorbates there exists several valence states that can interact with the surface, complicating the electronic structure changes.

### Surface periodic trends

Nørskov and Hammer have used electronic structure analysis to explain the differences in adsorptive properties between transition metals.<sup>172</sup> The contribution from the coupling to the metal  $s$  states is approximately the same between transition metals and the main periodic trends in chemisorption energy are determined by the coupling to the  $d$  electrons. The  $d$  bands move up in energy further left in the periodic table, and more antibonding adsorbate-metal  $d$  states become empty. In copper, silver and gold the antibonding states are completely occupied. Going further left the  $d$  bands are elevated and antibonding states become depopulated. Because the adsorption energy varies with the position of the  $d$  band relative to the Fermi level, and the filling and the width of  $d$  bands are strongly coupled, both the center of the  $d$  band and its width are good measures of reactivity in transition metal surfaces.<sup>172</sup>

Surface interaction weakens as we go down in the periodic table. This is a consequence of Pauli repulsion between adsorbate and the  $d$  states of the metal. The repulsion is dependent on the geometry of the adsorbate and on how extended the  $d$  orbitals are.  $5d$  orbitals are more extended than  $4d$  orbitals, which are more extended than  $3d$ . Consequently gold is more inert than silver, which is more inert than copper.

### Adsorbate-adsorbate interactions

Adsorbate-adsorbate interactions vary with coverage and can become more important than surface nature. Attractive interactions manifest experimentally as island formations at low coverages and temperatures, while strong coverage-dependent heats of adsorption indicate repulsive interactions with dispersed overlayers formations.<sup>172</sup>

Causes for adsorbate-adsorbate interactions include direct overlap of wavefunctions between neighbouring adsorbates, indirect changes in the electronic structure

or the lattice of the surface and nonlocal electrostatic effects such as dipole-dipole and higher dipole interactions.<sup>172</sup>

## Chapter 3

# Fundamental Aspects in Surface Self-Assembly: A Theoretical Study of Polarity and Shape<sup>§</sup>

### 3.1 Introduction

At the simplest level, surface self-assembly (SA) represents the outcome of the coupling between inter-molecular and molecule-surface interactions. In one limit, the surface acts as mere support, with SA being dictated by the three-dimensional crystalline structure adopted by molecules in the absence of the surface. In the other limit, the substrate-molecule interaction can be so strong as to strongly impact the geometry of the monolayer. Between these two extremes, entirely new phases with varying degrees of order may become energetically favoured.

In this Chapter we begin our foray into theoretical surface SA by investigating how the competition between van der Waals interactions, dipole-dipole forces and steric effects drive the formation of phases with varying degrees of order. We also wish to understand how the competition between intermolecular and surface-molecule interactions play a role in the same processes. As these interactions are a first-order approximation for surface SA, our goal is to demonstrate that these alone are sufficient components for simulating complex phase formations. We also wish to identify, where possible, causal links between molecular features and the structure of phases.

---

<sup>§</sup>This chapter is based on article “*Fundamental aspects in surface self-assembly: theoretical implications of molecular polarity and shape*” by E. Tuca and I. Paci published in *Phys. Chem. Chem. Phys.*, 2016, **18**, 6498-6508, DOI:10.1039/C5CP04479K.

We use a series of simple model molecules with fine-tuned interactions to investigate their adsorption on symmetric surfaces. We use the parallel tempering Monte Carlo formalism coupled with the Lennard-Jones-electrostatic potential to sample the configurational space of multi-body systems with pure composition. By tuning the molecular and surface parameters of these two- to six-atom molecules and the substrate, we report on how the interplay of parameters gives rise to ordered and complex equilibrium patterns resembling those seen in real systems, demonstrating that these interactions are sufficient foundational blocks for surface SA.

## Chapter layout

Section 3.2 provides an overview of the methodology used, including our potential of choice, the rationale behind our molecular models and naming, details of our sampling methodology and the statistical tools used to quantify order in our results.

Section 3.3 provides the results and discussion of diatomic dipoles in detail, followed by Section 3.4 which discusses linear models with variable length and van der Waals components. Asymmetric dipoles with variable length and angular shape are discussed further in Section 3.5.

Section 3.6 closes Chapter 3 with some conclusions. The rest of this section provides a context for our study and a high-level presentation of our results.

### 3.1.1 Lennard-Jones Spheres on Surfaces

The simplest models to investigate are those of spherical particles on symmetric surfaces. Whether these particles represent atoms, colloidal spheres, or proteins, the model isolates the interplay between intermolecular interactions and the energetic and geometric makeup of the surface in directing the outcome of the surface adsorption process. The assembly of atomic clusters has long been studied in vacuum conditions, in order to understand the formation and properties of nanoparticle assemblies. When assembly occurs at a solid surface, clusters of any size are distorted by the presence of the surface.<sup>120</sup> The tendency of Lennard-Jones atoms to form compact droplets competes with surface attraction, which favours a monolayer arrangement of the cluster atoms. In the limit of isolated clusters, atoms arrange to form compact structures of high symmetries. In the limit of strong surface-cluster interaction, surface adsorption at low temperatures leads to the formation of a monolayer.

## Surface symmetry

The symmetry of the adsorbed arrangement is also determined by an interplay of intermolecular and surface-driven tendencies. Spherical particles prefer to hexadic arrangements which are also achieved upon adsorption on a smooth or hexagonal surface. Surfaces with 4-fold symmetry impose a square geometry, with a different characteristic length. This tension is reflected with a loss of order in the adsorbed phase. Across temperature ranges, a succession of phase changes and rearrangements are observed, namely condensation to the surface at high temperature and the “freezing” of the cluster at a lower temperature. Between the two temperatures a liquid-like phase forms on the surface.

## Particle size

The size of the surface atoms relative to the cluster atom size is also relevant for the symmetry outcome of the assembly process, and, through density effects, for the overall attractive force acting on the cluster. In the limit when the surface is smooth, the cluster atoms determine the surface arrangement, with the formation of a hexadic layer. When the size of the surface atoms is comparable to that of the cluster atoms, the geometry of the surface has a strong impact on the symmetry of the arrangement. This effect is normally used in the templated growth of colloidal photonic crystals.<sup>178–180</sup>

### 3.1.2 Dipolar Models on Surfaces

#### Diatomic models

Diatomic, dipolar molecules are a first-approximation model for self-assembling systems. Using a series of partial charges at varying molecular locations to generate molecular dipoles, we simulated SA structures formed in conditions ranging between surface-driven and intermolecular interactions-driven adsorption. Strongly interacting, high-charge dipoles form multi-layer crystalline structures whereas weak intermolecular forces with low polarity lead to surface-directed patterning. In the intermediate regime, geometric mismatches between crystalline close-packing and surface-driven monolayer can lead to a loss of order.

At a given surface potential, simple diatomic dipolar molecules can be assembled in patterns ranging from partially ordered monolayers, to disordered multilayers, to

aggregates and ordered crystalline structures by varying the intensity of the molecular dipoles.

### **Linear models**

Molecules with identical dipole moments but increasing backbone lengths were illustrative of the competition between the tendency of dipolar molecules to form crystals, and the increasing surface attraction arising from a longer backbone. Simple two-atom molecules formed crystalline structures, with antiparallel or L-type pair alignment. Longer molecules experienced enhanced surface attraction and tended to form antiparallel monolayers. Molecules of intermediate lengths experienced frustration due to the combination of dipolar interactions, tending to multilayer crystalline structure, with insufficient dispersive molecule-surface and molecule-molecule interactions, to drive the formation of antiparallel monolayers. Overall stronger dipole moments lead to crystalline adsorbed structures even in molecules with long backbones, as intermolecular interactions gain in importance over molecule-surface attraction. On the other hand, a lower dipole strength relative to the surface field promotes the formation of the monolayer.

Examining the importance of charge location within linear molecules, when both charges are localized in the centre of the molecule, and flanked by the greasy backbone atoms, molecules exhibit amphiphilic behaviour, with the formation of clusters or chains. In these arrangements charges are grouped on the inside and bulky groups are pushed to the outside.

### **Bulky substituents**

The straightforward ways in which linear molecules can form closed-packed arrangements may be disrupted through the use of bulky substituents, a non-linear skeleton and asymmetric charge distribution. The presence of bulky groups impedes crystal growth in one or two dimensions, resulting in the the formation of crystalline wires, wherein bulky groups are pushed towards the outside of the chain. In cases where assembly is dominated by the surface attraction, structural features can be strongly impacted by the symmetry of the underlying surface. When the bulky part is a long molecular backbone, strong tail-tail and tail-surface interactions dominate, and tailgroup-directed assembly is observed. Such systems form compact, planar arrangements of tailgroups.

## 3.2 Methodology

### 3.2.1 The Lennard-Jones-Electrostatic Potential

Tens to hundreds of molecules are normally examined on an atomistic ideal surface, in a simulation box, with no periodic boundary conditions. Molecules are also described atomistically, with interactions built on the Lennard-Jones-electrostatic potential,

$$U_{nb}^{ab} = \sum_{\substack{i,j=1 \\ j < i}}^{n_{at}} \left[ 4\epsilon_{ij} \left( \left( \frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left( \frac{\sigma_{ij}}{r_{ij}} \right)^6 \right) + \frac{q_i q_j e^2}{4\pi\epsilon_0 r_{ij}} \right], \quad (3.1)$$

where  $a$  and  $b$  denote two interacting molecules,  $n_{at}$  is the number of atoms in a molecule,  $\epsilon_{ij}$  and  $\sigma_{ij}$  are the Lennard-Jones depth and distance parameters for atoms  $i$  and  $j$ ,  $r_{ij}$  is the distance between the centers of the two atoms,  $q_i$  is the partial charge on atom  $i$ ,  $e$  is the electronic charge and  $\epsilon_0$  is the dielectric constant of vacuum. Mixed Lennard-Jones parameters are normally obtained using the Lorentz-Berthelot mixing rules:  $\sigma_{ij} = (\sigma_{ii} + \sigma_{jj})/2$ ,  $\epsilon_{ij} = \sqrt{\epsilon_{ii}\epsilon_{jj}}$ .

### 3.2.2 Modelling Details

#### The rationale behind molecular shapes

Adsorption of Lennard-Jones spheres on surfaces offers valuable insight into how competition between inter-particle and surface-particle interactions directs SA. The spherical nature of particles, however, renders the models insufficient for describing systems where particles have detailed geometries. Tracing the outcome of SA process to its structural causes in these circumstances requires the ability to model particles with arbitrary shapes. Moreover, dipolar interactions are crucial in many SA processes, and their inclusion in simulation offers the option to explore additional layers of complexity that many systems possess. To explore the assembly behaviour of these elementary self-assembly systems, we considered models with an array of dipolar interactions and geometric effects and their adsorption on a solid surface (see Figure 3.1. By tuning the molecular and surface parameters in systems comprised of multiple two- to six-atom molecules and a substrate, we attempted to understand how the interplay of these parameters gives rise to complex equilibrium patterning.

The competition between intermolecular and molecule-surface interactions is of central importance in surface science. This study around this issue is designed to

provide insight into how such competition resolves, and what theoretical or experimental handles a researcher might use to control its outcome. By stripping off the complexities inherent in any experimental study, theoretical investigations have the ability to approach increasingly complex models, often isolating individual parameters for in-depth analysis, in order to establish fundamental links between SA strategies and molecular and surface features, such as size, shape, surface geometry and interaction potential.

### Molecular models and naming

Our molecular models were designed to interact through a Lennard-Jones-electrostatic (LJe) potential (see Equation 3.1), in a simulation box containing a surface. Mixed Lennard-Jones parameters were obtained using the Lorentz-Berthelot mixing rules. All molecules and the surface were rigid, i.e. all bond lengths, bond angles and torsion angles were held fixed, and no surface atom vibration or surface reconstruction were allowed. All models possessed a dipole moment created by a pair of atoms which had equal and opposite charges but were otherwise identical.

Series A (basic), a series of simple diatomic dipolar models, was used to investigate the competition between intermolecular interaction strengths and surface potentials. Steric effects were added in series C (constant Charge), comprised of molecules with charged end atoms, with constant charges and different molecular lengths (see Figure 3.1). Altering charge separation changes dipoles however, so series D (constant Dipole), with constant dipole moments but varying molecular lengths was additionally considered. In series D, the dipole moment was held constant by lowering atomic charges as charge separation increased. Molecules in series L (constant Length) also exhibited constant dipole moments, with varying dipolar lengths and charges, within overall constant molecular lengths. Finally, set  $G_N$  (for Geometric effects) was formed of molecules with two charged atoms and  $N$  neutral atoms, and varying bond angles.

Interaction parameters for the different model series are presented in Table 3.1. Note that reduced units are used throughout, and the results discussed below are valid for an infinite number of systems, due to the law of corresponding states. One such state, for the system with  $\sigma_{11} = 3.65 \text{ \AA}$  and  $\epsilon_{11} = 0.9 \text{ kJ/mol}$ , has  $T^* = 1$  corresponding to a temperature of 204 K. The interaction of a methyl group in an alkane with a graphite surface would correspond to a  $\sigma_{11} = 3.95 \text{ \AA}$  and a  $\sigma_{ss} = 3.55 \text{ \AA}$ .<sup>181</sup> Given the  $\epsilon_{11} \approx \epsilon_{ss} = 1.23 \text{ kJ/mol}$ ,  $T^* = 1$  in this system corresponds to a temperature of 148 K.

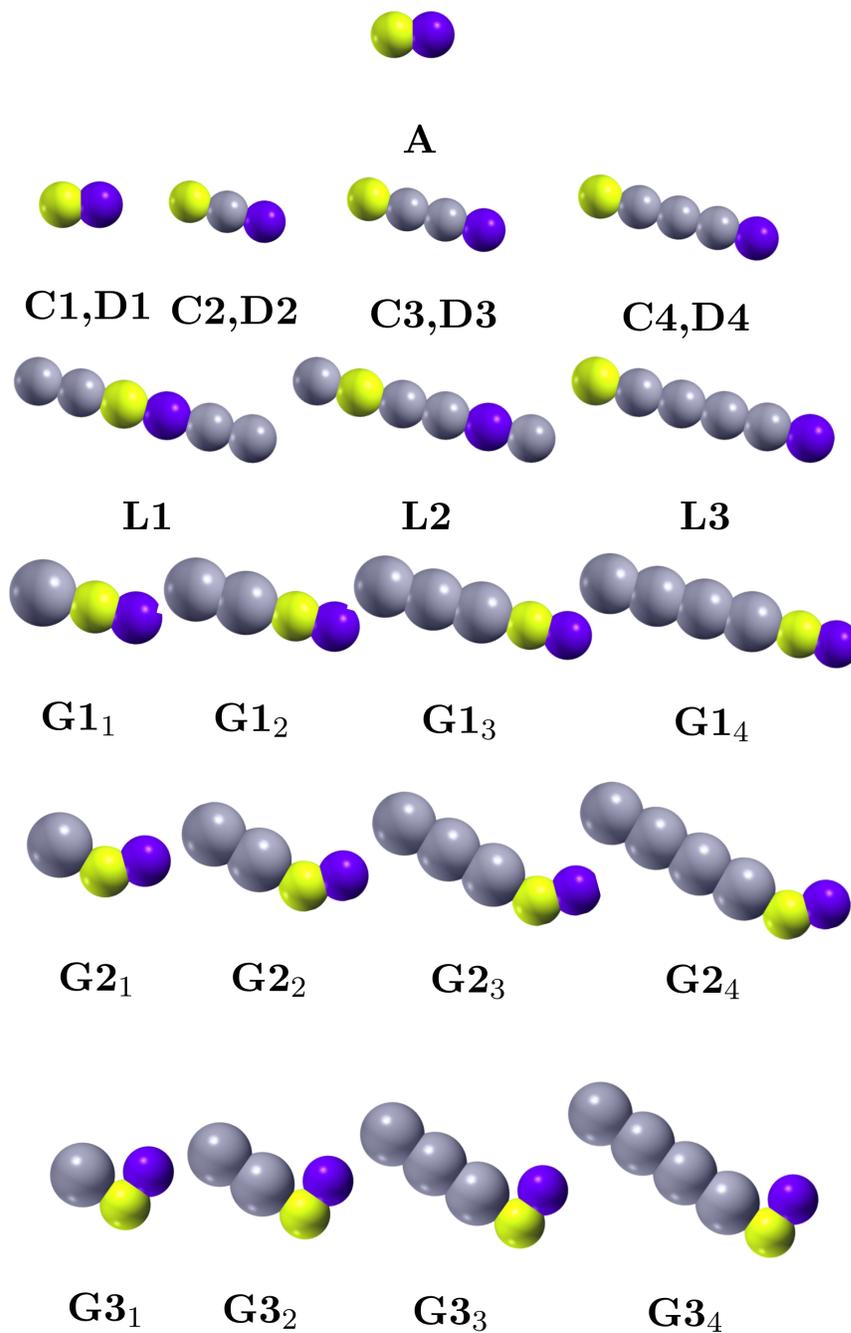


Figure 3.1: A complete listing of models used throughout this study. The atoms carrying partial charges highlighted in yellow and purple. Grey atoms carried no charge. Symbols are determined by the effect they help study.

On the other hand, for methanethiol/gold system, a back-of-the-envelope calculation shows the dipole-dipole contribution to lateral interactions to be roughly 5 times

smaller than the molecule-surface interaction. This corresponds in the present work to an A1-like model with  $\epsilon_{ss}=1$  and charges of 0.08 of an electron.

### Numerical details and surface model

Each simulation followed either 80 (in simulations of models C, D, L and G) or 100 (models A) molecules, using a Parallel Tempering Monte Carlo procedure over a range of temperatures. Preliminary calculations for models A and C with 30, 50, 80 and 100 molecules indicated a switch between edge-driven properties and bulk-driven properties around 50 molecules. Figure 3.2 illustrates the absence of size effects for the system C2- $l$ .

Two square surface layers were considered from the (100) facet of the FCC crystal, totalling 1156 atoms (for a simulation box length of 25 reduced units). Box sizes were chosen such that a low density two-dimensional phase was created, and condensed patterns were not affected by variation of box size. Periodic boundary conditions were found to have no impact on simulation outcomes in these circumstances, and were not employed in any of the reported results. Evaporative events (particles leaving the simulation box) were forbidden, by rejecting moves that lead to evaporation.

Preliminary calculations using slightly larger ( $\sigma_s = 1.2$ ) surface atoms and (111) facets produced results qualitatively similar to those reported here, with ( $\sigma_s = 1.2$ ) and (100) facet. The change in surface atom density, whether by larger surface atoms or by a more compact (111) layout, was reflected mainly in the increase of overall surface-molecule interaction strengths, and was qualitatively equivalent to an increase in the surface potential parameter. This mirrored our previous experience with purely dispersive substrates.<sup>120;182</sup> In consequence, the results presented below were obtained with a (100) substrate.

### 3.2.3 Parallel Tempering Monte Carlo

The large-scale simulations presented below were performed using an in-house Parallel Tempering Monte Carlo (PTMC) code, with simulation details available in the cited works.<sup>182;183</sup> Briefly, our PTMC calculations consist of a Markov chain of the following moves: (i) Standard MC molecular translations and rotations, with a Metropolis acceptance criterion for the canonical ensemble,<sup>112</sup> (ii) angular vibrations and torsional moves in flexible molecules, accomplished by bond rotations on a unit sphere,

Table 3.1: Interaction parameters for the investigated models

Model*	$n_{\text{at}}$	$q^\dagger$	$\sigma_{ii}^* (\epsilon_{ii}^*)^\ddagger$	$\sigma_{ss}^* (\epsilon_{ss}^*)^\ddagger$
A1	2	0.01	1 (1)	1 (1; 3 <sup>§</sup> )
A2	2	0.10	1 (1)	1 (1; 3 <sup>§</sup> )
A3	2	0.20	1 (1)	1 (1; 3 <sup>§</sup> )
C1- <i>l</i>	2	0.05	1 (1)	1 (1.5)
C2- <i>l</i>	3	0.05	1 (1)	1 (1.5)
C3- <i>l</i>	4	0.05	1 (1)	1 (1.5)
C4- <i>l</i>	5	0.05	1 (1)	1 (1.5)
C1- <i>h</i>	2	0.20	1 (1)	1 (1.5)
C2- <i>h</i>	3	0.20	1 (1)	1 (1.5)
C3- <i>h</i>	4	0.20	1 (1)	1 (1.5)
C4- <i>h</i>	5	0.20	1 (1)	1 (1.5)
D1	2	0.20	1 (1)	1 (1.5)
D2	3	0.10	1 (1)	1 (1.5)
D3	4	0.07	1 (1)	1 (1.5)
D4	5	0.05	1 (1)	1 (1.5)
L1	6	0.20	1 (1)	1 (1.5)
L2	6	0.07	1 (1)	1 (1.5)
L3	6	0.04	1 (1)	1 (1.5)
G1- <i>l</i>	3	0.05	1 <sup>¶</sup> (1)	1 (1.5)
G2- <i>l</i>	3	0.05	1 (1)	1 (1.5)
G3- <i>l</i>	3	0.05	1 (1)	1 (1.5)
G1- <i>h</i>	3	0.20	1 (1)	1 (1.5)
G2- <i>h</i>	3	0.20	1 (1)	1 (1.5)
G3- <i>h</i>	3	0.20	1 (1)	1 (1.5)

\*The *-h* and *-l* notation was used to denote the high charge and low charge versions of models C and G.

<sup>†</sup>In each molecule, one atom has  $+q$  charge and another atom has a  $-q$  charge (See Figure 3.1).  $q$  is given in fractions of one electronic charge.

<sup>‡</sup>Reduced LJ parameters are relative to the charged atom LJ parameters. The  $\epsilon$  parameter is given in brackets. The  $s$  subscript refers to surface parameters, while  $i$  refers to atoms in the adsorbate.

<sup>§</sup>Two different values of the reduced surface-molecule interaction parameters were used.

<sup>¶</sup>In the G-series molecules, all uncharged atoms had a larger  $\sigma^*$  ( $= 1.2$ ).

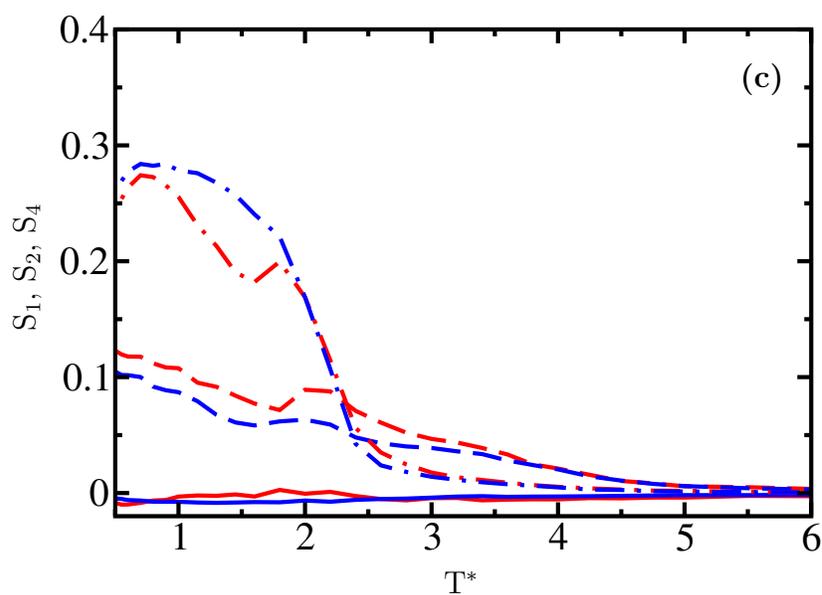
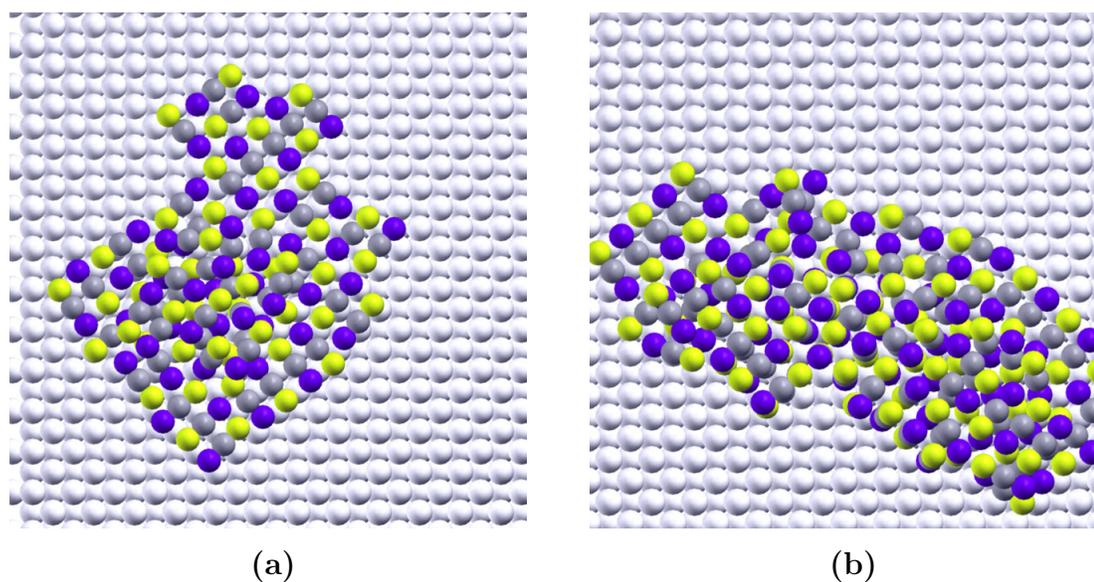


Figure 3.2: Snapshots of system C2- $l$  with different number of molecules: 50 (**panel a**) and 80 (**panel b**) reflecting similar SA behaviour for varying system sizes. Both snapshots are at  $T^* = 1.0$ . Regardless of size, the commensurate nature of phase and the antiparallel alignment of neighbouring molecules are apparent. **Panel c**: values for the order parameters (y axis) as a function of temperature,  $T^*$  (x axis).  $S_1$  – continuous line,  $S_2$  – dashed line,  $S_4$  – dashed-dotted line for small (50 molecules, red curve) and large (80 molecules, blue curve) systems. The similarity between curves suggests no significant system size effects.

also sampled with a Metropolis criterion, and (iii) swap moves between neighbouring-temperature replicas of the system.

Simulations were run for at least  $10^9$  MC steps, until structural and energy convergence was observed. Averages were collected post-equilibration over another  $10^9$  MC steps. The temperature range was, in all cases, large enough to sample the configurational space thoroughly, with the highest temperature always leading to gaseous phase. Between 35 and 70 replicas were necessary for the various systems, with attempted swaps performed in 5% of the moves. The target acceptance ratio for temperature swaps was 10%, and the number and spacing of replicas was adjusted to attempt to meet it. As a rule, systems with a sharper potential energy surface (created by stronger electrostatic potentials) required larger numbers of replicas to equilibrate.

### Snapshots, sample sets and averages

In this chapter we use snapshots to illustrate visually the phases formed and discussed. For each model, a collection of a large number of sample sets was generated via the PTMC formalism. Each sample set in this collection corresponds to a distinct temperature at which the sampling was performed within that set. The term “snapshot” refers to one configurational instance (with all simulated particles present) chosen from the sample set. The sampling events were spaced apart enough (in units of Markov chain steps) to allow for configurational decoupling between consecutively sampled snapshots. The illustrated snapshots were chosen to be representative of their corresponding sample sets. All statistical quantities presented in this chapter (order parameters, fraction of adsorbed molecules) were calculated per sample set.

### 3.2.4 Order Parameters

To investigate the degree of order in the systems, three order parameters,  $S_1$ ,  $S_2$  and  $S_4$  were chosen. These are statistical quantities given by:

$$\begin{aligned}
 S_1 &= \langle \cos \theta_{ij} \rangle \\
 S_2 &= \frac{\langle 3 \cos^2 \theta_{ij} - 1 \rangle}{2} \\
 S_4 &= \frac{\langle 35 \cos^4 \theta_{ij} - 30 \cos^2 \theta_{ij} + 3 \rangle}{8}
 \end{aligned} \tag{3.2}$$

where  $\theta_{ij}$  is the angle between the dipoles of molecule  $i$  and  $j$  and averaging is performed over all distinct pairs of molecules in the system, then over MC steps. The quantities  $S_1$ ,  $S_2$  and  $S_4$  can vary between 0 and 1 and are measures of long range alignment in the system. For systems with parallel alignment, i.e. all molecules having the same orientation and dipole moment direction, all three quantities converge to 1. In a perfectly disordered system all three are 0. A perfectly antiparallel dipolar alignment would also yield an  $S_1$  value of 0. Thus  $S_1$  alone is unable to discriminate between a disordered and an antiparallel state. On the other hand, a high value of  $S_2$  can be attributed to either parallel or antiparallel orientation. By using the two parameters together one can distinguish between disordered, parallel and antiparallel states effectively. The  $S_4$  order parameter can provide partial information about perpendicular ordering in the system:  $S_4$  close to 1 indicates, but does not discriminate between, parallel and antiparallel alignment. Instead,  $S_4$  tends towards a value of 0.6 in systems with both parallel and perpendicular alignment: note that in an extended system with full perpendicular organization, 50% of particles will lie in one direction (thus be parallel to each other), and the rest will lie along a perpendicular direction.

To quantify the degree to which molecules group in monolayer and multilayer structures, we calculated the fraction of molecules and the fraction of atoms which contact the surface. The molecule fraction is always lower than that of atoms, since at least one atom touches the surface when a molecule touches the surface, but is directly related to the definition of a monolayer or a multilayer. We report the fraction of molecules touching the surface in the following pages.

### 3.3 Series A: Diatomic Dipoles

Diatomic, dipolar molecules are a first-approximation model for self-assembling systems. The simplicity of the model allows the isolation of steric (excluded-volume) effects, important in close-packed environments, from the effect of Coulombic interactions. It is well known that alkylthiol self-assembly, for example, is greatly directed by the bulky alkyl group, from the organization of laying-down phases, to the tilt and phase behaviour of upright monolayers. Some of that complexity will be sought in subsequent models, but first, what would the headgroup, or polar substituents like to do, if they were allowed to adsorb alone on the substrate?

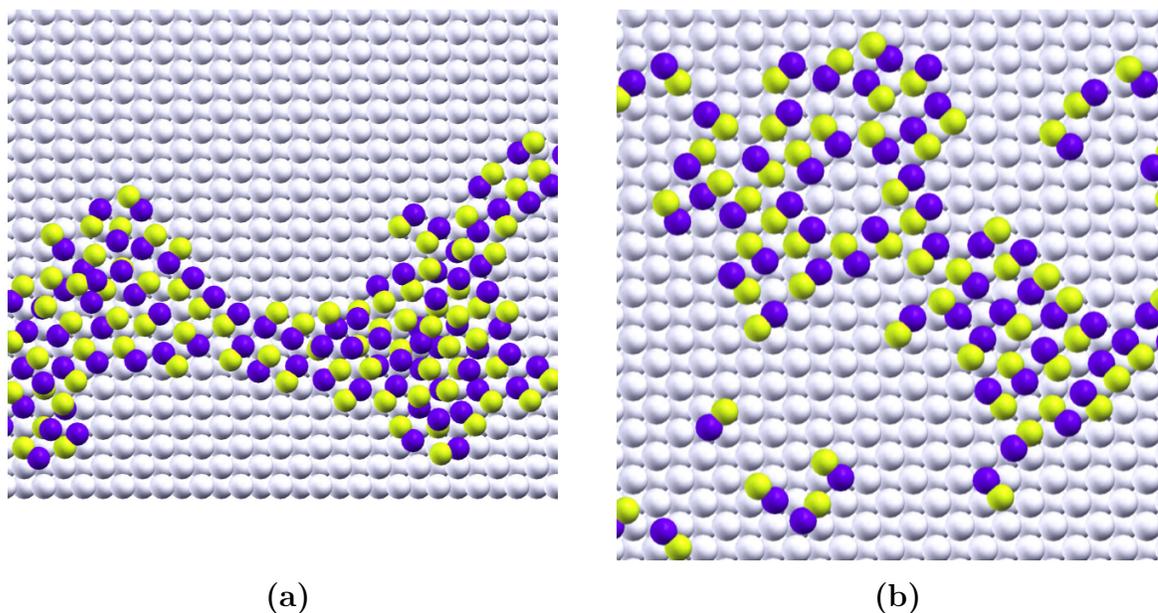


Figure 3.3: Effects of surface potential on order. Snapshots of A1 molecules adsorbed on a surface with  $\epsilon = 1$  (a) and  $\epsilon = 3$  (b). For a low surface potential (a) the alignment is weaker than for a high surface potential (b), as can be seen when comparing the  $S$  values in Table 3.2.

### 3.3.1 The Influence of Surface Potential

The series of models A1-A3 are diatomic molecules with partially charged atoms. The partial charges considered here range from almost neutral to high, and both strongly and weakly-attractive surface atoms (see Table 3.1). In these simple systems, the interplay between intermolecular and molecule–surface forces resulted in structures spanning a range between surface-driven and intermolecular interactions-driven adsorption. In almost all cases, a stronger surface interaction is correlated with a stronger supra-molecular order (see Figure 3.3).

In Table 3.2 Series A, model A1, the order parameters  $S_1$  and  $S_2$  are compared for a low and a high surface potential. A high surface potential induces more order when the dipole moment is kept constant. This may be attributed to the fact that a more attractive surface promotes the formation of a relatively stable monolayer, with molecules being more likely to align their dipole within monolayer than to desorb from it. A surface with high potential imposes its geometry on SA more than a low-potential surface does.

Table 3.2: First, second and fourth order parameters  $S_1$ ,  $S_2$  and  $S_4$ .

Molecule	$S_1$	$S_2$	$S_4$
A1*	0.00	0.05	0.10
A1 <sup>†</sup>	0.00	0.18	0.31
A2*	-0.01	0.01	0.26
A2 <sup>†</sup>	0.00	0.17	0.28
A3	-0.01	0.03	0.50
C1- <i>l</i>	0.00	0.15	0.33
C2- <i>l</i>	0.00	0.10	0.28
C3- <i>l</i>	0.01	0.20	0.26
C4- <i>l</i>	0.02	0.23	0.25
C1- <i>h</i>	-0.01	0.01	0.24
C2- <i>h</i>	-0.01	0.00	0.03
C3- <i>h</i>	0.64	0.43	0.10
C4- <i>h</i>	0.75	0.58	0.23
D1	-0.01	0.01	0.24
D2	-0.01	0.01	0.03
D3	-0.01	0.10	0.08
D4	0.00	0.19	0.22
L1	-0.01	0.03	0.03
L2	0.00	0.23	0.24
L3	0.00	0.23	0.27
G1 <sub>1</sub> - <i>l</i> <sup>‡</sup>	0.00	0.17	0.12
G1 <sub>2</sub> - <i>l</i>	-0.01	0.27	0.36
G1 <sub>3</sub> - <i>l</i>	0.06	0.31	0.37
G1 <sub>4</sub> - <i>l</i>	0.04	0.29	0.35
G2 <sub>1</sub> - <i>l</i>	0.00	0.12	0.03
G2 <sub>2</sub> - <i>l</i>	0.00	0.13	0.01
G2 <sub>3</sub> - <i>l</i>	0.17	0.18	0.07
G3 <sub>1</sub> - <i>l</i>	0.01	0.12	0.10
G3 <sub>2</sub> - <i>l</i>	0.04	0.08	0.15
G3 <sub>3</sub> - <i>l</i>	0.12	0.05	0.12
G3 <sub>4</sub> - <i>l</i>	0.13	0.05	0.11
G1 <sub>1</sub> - <i>h</i>	-0.01	0.11	0.21
G2 <sub>1</sub> - <i>h</i>	-0.01	0.01	0.04
G3 <sub>1</sub> - <i>h</i>	0.00	0.03	0.08

\*S values for  $\epsilon_{ss}^* = 1$  are shown here. <sup>†</sup>S values for  $\epsilon_{ss}^* = 3$ . <sup>‡</sup>Parameters for biaxial molecules reported for the correlation of dipole axes. Order parameters for backbone axes are presented in the ESI Table 1<sup>†</sup>.

### 3.3.2 The Influence of Dipole Strength

Strong lateral interactions dominated the self-assembly process, with small regard given to the surface attraction or geometry. In the case of molecule A3, for example, multi-layer crystalline structures were observed, similar to dipolar crystals formed in gas phase. At times, crystals formed above the surface without adsorption, particularly at higher temperatures. As shown in Figure 3.4(b), the multilayer structure exhibited sharp corners, indicative of a crystalline state.

At the other extreme, weak intermolecular forces in systems with low polarity led to surface-directed patterning, where molecules commensurated with the surface to a degree determined by the strength of the surface attraction. Figures 3.4(a) and (c), for example, show snapshots obtained for normal and strong surface attraction, respectively. In Figure 3.4(c), in particular, molecules moved apart from van der Waals distances in order to commensurate with surface pit sites. This led to molecular axes aligned on [110] or [-110] directions, along the lines of pit sites, with large intermolecular distances. In the case of molecule A1, polarity was too weak to even enforce intermolecular alignment at the larger intermolecular distances. These effects were reflected in the values of order parameters for these systems (see Table 3.2).

In the intermediate regime, a competition of the two driving forces ensued. As shown in Table 3.2 for molecule A2, lateral interactions of intermediate strength did not necessarily enhance ordering over weakly interacting molecules. Stronger interactions biased the structures towards close-packed arrangements, whereas surface-molecule interactions favoured a more sparsely-spaced commensurated structure. The mismatch led to frustration in conditions of comparable lateral interactions and molecule-surface interactions. When a stronger surface potential was imposed, molecules A2 formed a commensurate monolayer, with dipolar alignment enforced by the molecules' stronger charge [see Figure 3.4(c)].

### 3.3.3 Quantifying Order

One note must be made at this point regarding the relationship between dipolar ordering and the values of the order parameters  $S_1$  and  $S_2$ : crystalline order does not necessarily lead to order parameters that are close to unity. As revealed by close examination of the highly ordered snapshots in Figures 3.4(b) and (d), pairs of molecules arranged in two types of relative orientations: the well-known antiparallel arrangement, and a perpendicular, “L-shaped” arrangement that also optimizes Coulombic

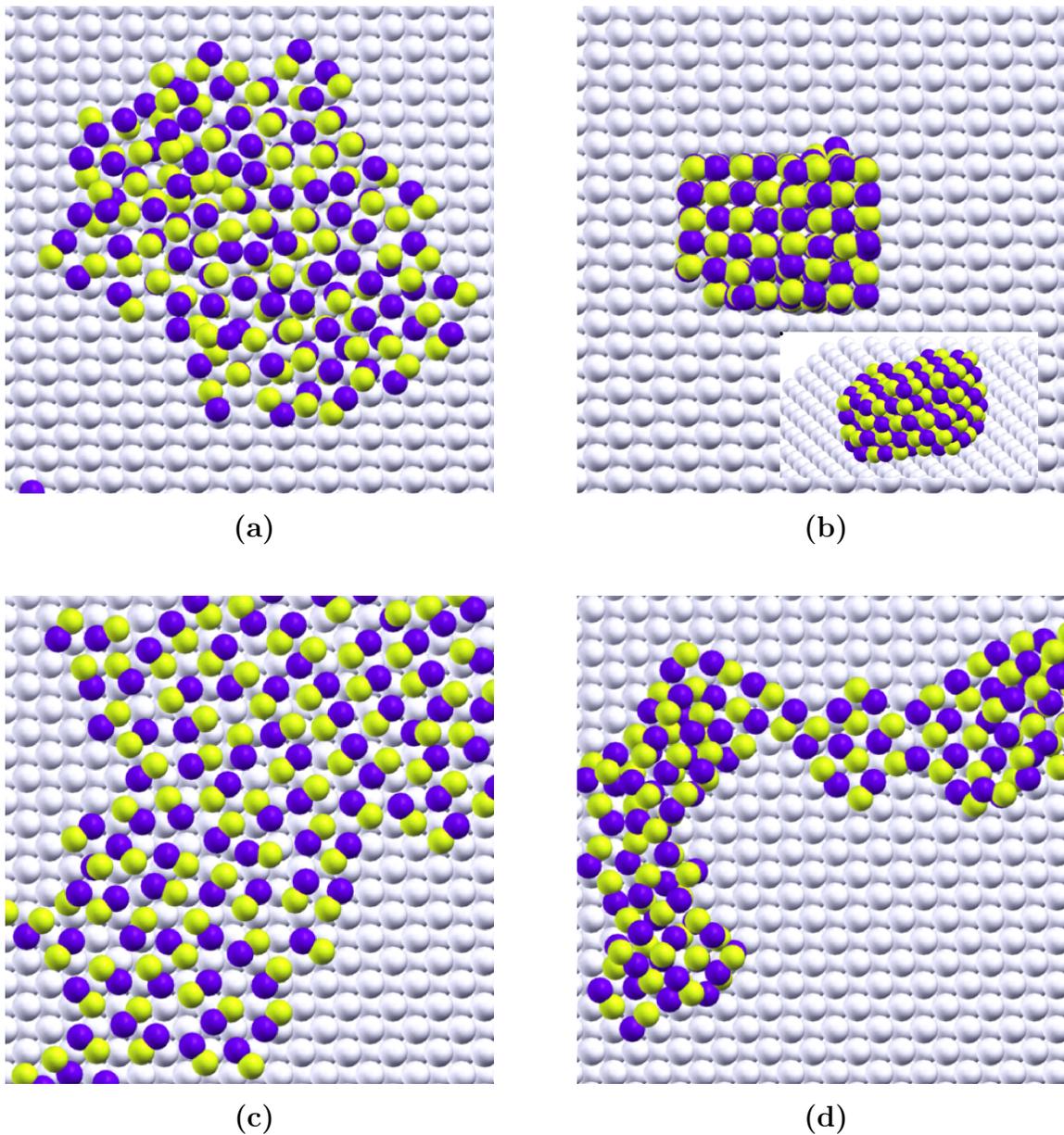


Figure 3.4: Pattern formation in simple dipolar adsorbates. Snapshots of A-series molecules adsorbed on a surface with two different LJ energy parameters  $\epsilon_{ss}^*$  are presented. Panels (a) and (b) display condensed phase snapshots for A1 and A3, respectively, when surface interaction is relatively weak ( $\epsilon_{ss}^* = 1$ ). An insert in (b) shows a side view of the crystal. Panels (c) and (d) show snapshots for molecules A1 and A2, respectively, at  $\epsilon_{ss}^* = 3$ . Panels present replicas at  $T^*=0.1-0.2$ .

interactions between charges systems with more than 2 molecules.

At a given surface potential, simple diatomic dipolar molecules can thus be as-

sembled in patterns ranging from partially ordered monolayers, to disordered multilayers, to aggregates and ordered crystalline structures by varying the intensity of the molecular dipoles (Figure 3.4). Many experimentally relevant systems have also been found somewhere between the two limiting cases of polarity-determined and surface-determined SA. Often times, organic molecules form SAMs by interacting strongly with the surface while having some geometric feature that directs how patterns are formed. For example, in two-dimensional arrays of 4-aminobenzoic acid on Cu(110), molecular positions were largely determined by the interaction between the surface and the benzene rings, and relative molecular orientations were determined by the molecular dipole moments.<sup>184</sup> Conversely, surprising close-packed structures due to strong lateral interactions have been observed even at low coverage adsorption. For example, 2,4- and 2,6- dinitrotoluene formed close-packed domains on Au(111), with molecules aligned in parallel, surface-normal orientations, as a result of strong dipole-dipole and  $\pi$ -stacking interactions.<sup>185</sup> In this case, the surface played only a secondary role, by immobilizing the layer. Beyond the simple dipole model, however, in most experimental SA studies, including the two mentioned here, steric (geometric) traits play important roles in determining SA outcomes. We explore below, with the use of series D, C, L and G, fundamental aspects of molecular geometry effects in surface SA.

### 3.4 Series C, D, L: Linear Dipoles

In general terms, longer molecules experience enhanced parallelism due to van der Waals interactions between constituent atoms, as well as excluded volume effects. Long molecules can more efficiently pack in parallel configurations, a fact well understood through decades of liquid crystal research and needlelike models.<sup>186–188</sup> The interplay between molecular length effects and dipole strength and length was examined here in some detail through series D, C and L.

#### 3.4.1 Equal Dipole Moments, Different Molecular Lengths

The effects of a weakly interacting backbone in molecules with similar dipole moments were examined in models D1-4. Here, dipole length was increased by addition of spacer atoms, while holding constant the overall dipole moment by modifying end group charges accordingly (see Figure 3.1 and Table 3.1). Whereas the diatomic

D1 molecules formed three-dimensional crystals [see Figure 3.5(a)], the behaviour of the longer D2-4 molecules changed systematically, as longer molecules were driven to align parallel to the surface. The spacer groups served to enhance surface-molecule interactions, but also inhibited L-shaped alignment within the condensed phase. As a result, fairly disordered multilayer structures were observed in condensed phases of molecules D2 and to some extent D3 [Figure 3.5(b) and (c), respectively]. System-wide order was only observed again in assemblies of D4, where stronger molecule-surface interactions led to monolayer formation. Within the monolayer, molecules acquired antiparallel orientations, with the formation of connected domains. Here, adjacent antiparallel domains organized along roughly perpendicular directions, as molecules aligned along the lines of surface hollow sites [e.g., Figure 3.5(d)].

These phase changes were reflected in the temperature dependence of the three order parameters: Figures 3.6(a)-(c) show decreased order at all temperatures in the shorter molecules D2 and D3, over D1 and D4. D4 molecules exhibited an apparent enhancement of their order parameter values around  $T^*=3$ . The increase paralleled an enhancement of monolayer character in the  $T^*=2.5-3.5$  temperature range, where thermal energies became competitive with electrostatic interactions [see Figure 3.5(d)]. At lower temperatures, D4 molecules satisfied their electrostatic interactions by multilayer structures, where multiple molecules could bring their charges into contact more effectively. This was reflected in a lower fraction of molecules in contact with the surface at low temperatures [see Figure 3.6(d)]. On the other hand, standard deviations of the  $S_2$  values for D4 were significant, as a result of the various packing arrangements of the antiparallel phase. This is illustrated in Figure 3.7, and the  $S_2$  maximum at  $T^*=3$  in Figure 3.6(b) is uncertain.

Figure 3.8 provides two snapshots at different temperatures for D4. At both temperatures the system manifests the antiparallel packing motif, with stronger alignment within rows at higher temperatures. Low temperature domains [Figure 3.8(a)] prefer stacking and electrostatic alignment between antiparallel rows to a greater extent compared to high temperature domains [Figure 3.8(b)]. The surface-extended domains allow for more ways for molecules to pack within an adsorbed antiparallel motif, offering an entropic advantage. Stacked domains allow for better electrostatic stabilization via interactions between layers, providing an energetic advantage. Higher temperatures can access the entropically favoured (more numerous) extended domains, leading to a positive trend between the size of monolayer antiparallel motifs and temperature. A metastable nematic phase, discussed in the following section,

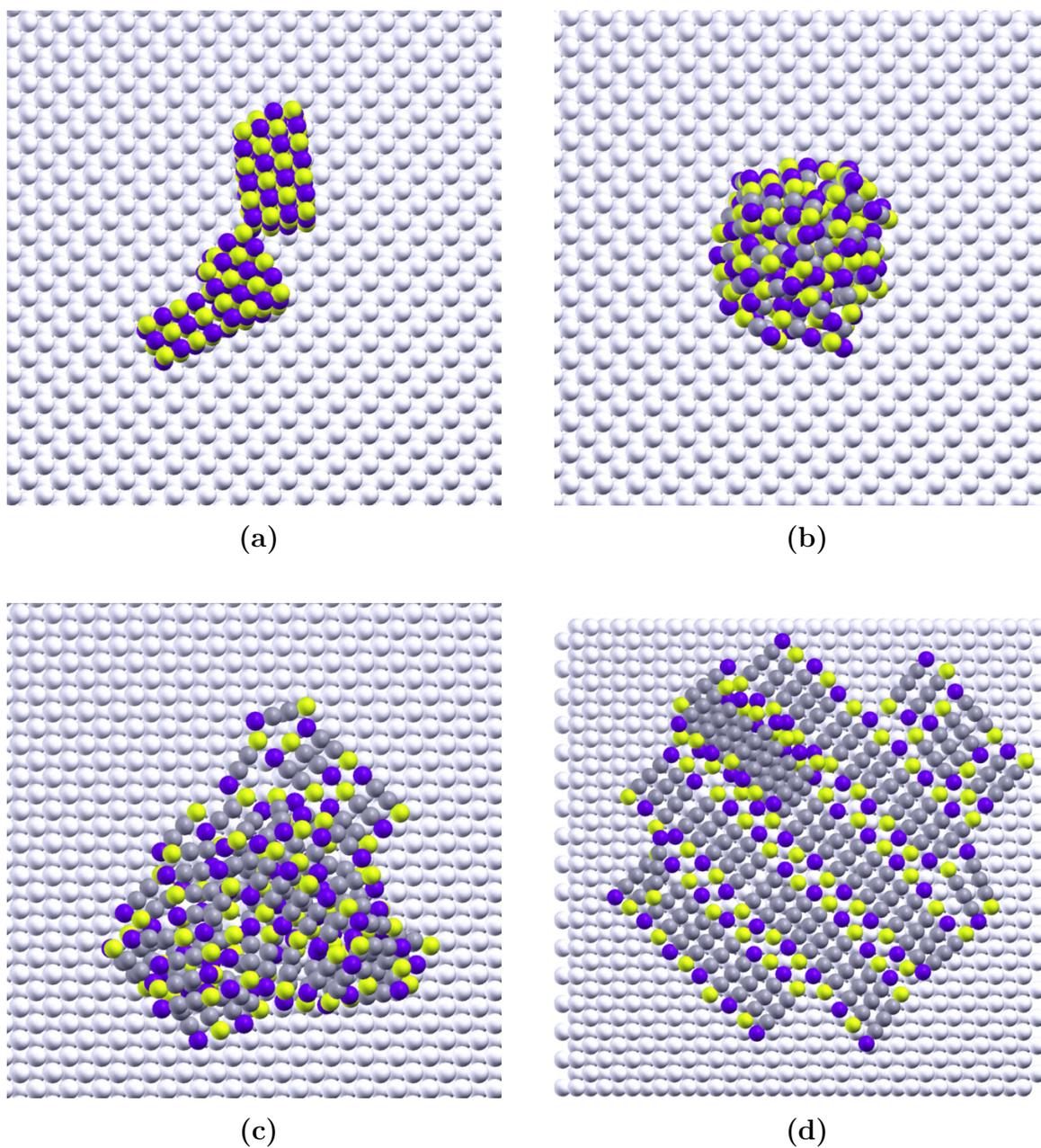


Figure 3.5: Effects of molecular length on SA structure. Snapshots of dominant low-temperature phases corresponding to series D1 (panel a), D2 (panel b) and D3 (panel c), and an intermediate-temperature antiparallel phase for D4 (panel d) are presented.

further complicates the picture in these systems.

In a nutshell, the preferred tail-to-tail antiparallel configurations of dipolar systems can be disfavoured in larger molecules, because of stronger surface-molecule

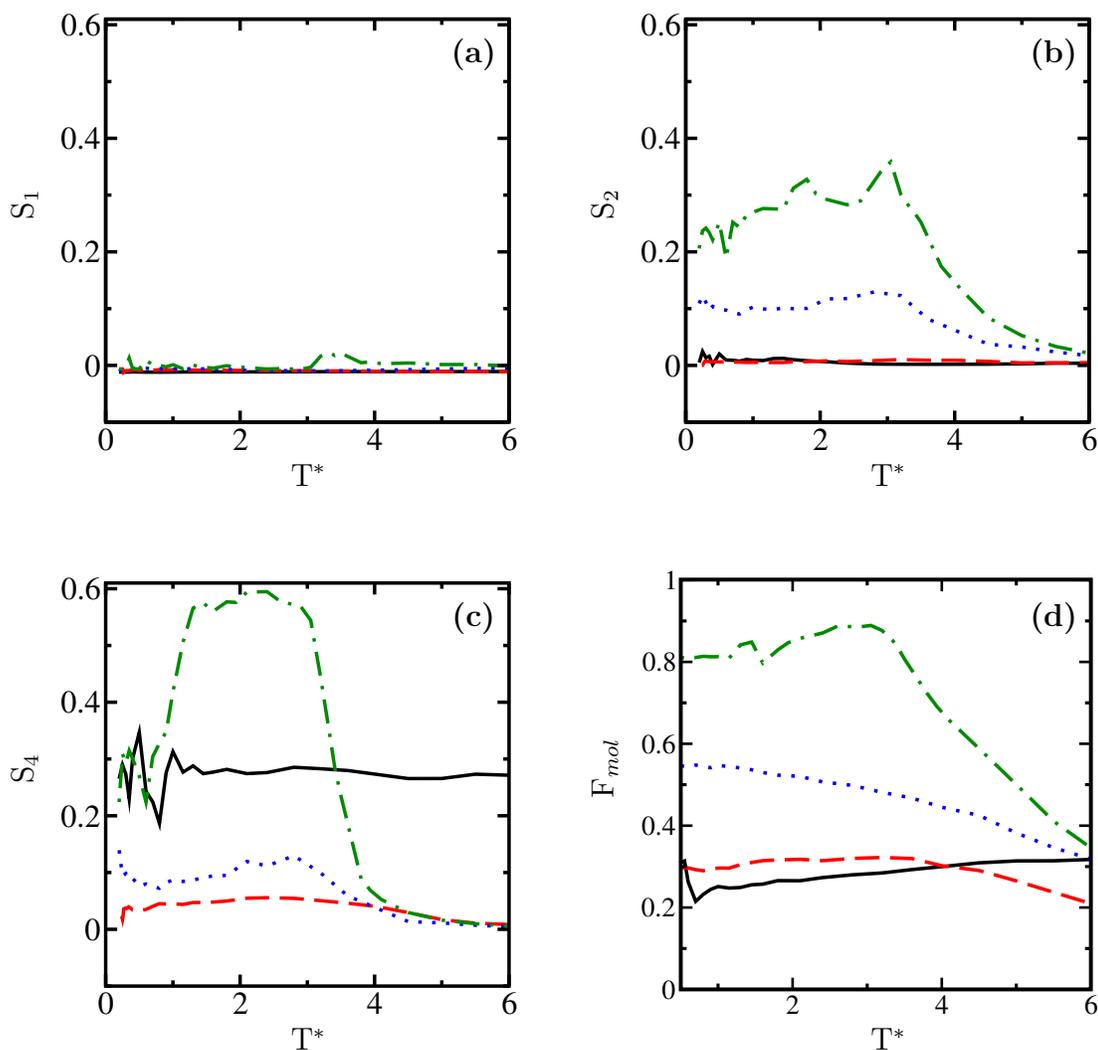


Figure 3.6: Temperature dependence of order parameters in systems D.  $S_1$ ,  $S_2$  and  $S_4$  are shown in panels a-c, respectively. The fraction of molecules in contact with the surface is presented in panel d. Black lines, red dashes, blue dots and green dot-dashed lines represent systems D1-4, respectively.

interactions, and the need to match substrate geometries. However, in many cases, any series of molecules employed in surface SA experiments would have ever more complex geometries with binding substituents designed for a specific type of interactions, and that do not change in nature as the molecular backbone is increased. Such molecules often form highly ordered, stable monolayers, with defect-free phases extended over tens and hundreds of nanometers. We employed series C to examine these effects.

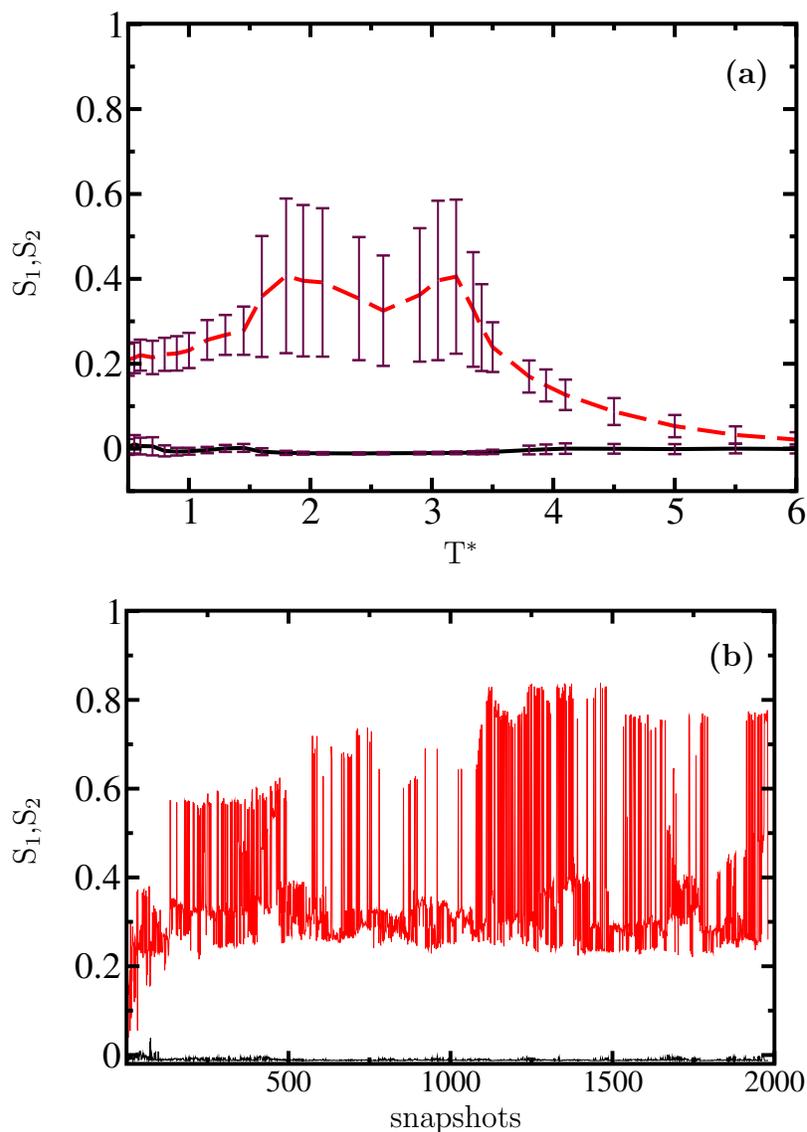
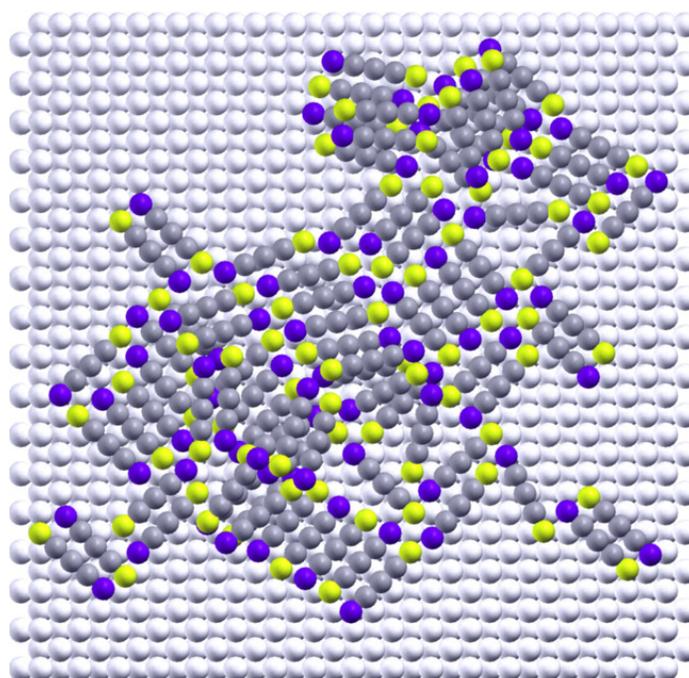


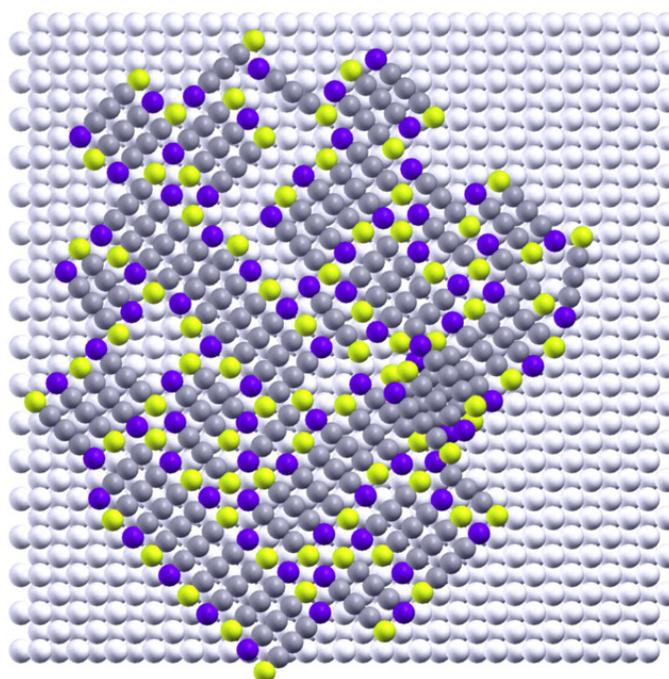
Figure 3.7: Order parameters for **D4/C4**– $l$  system. **Panel a**: averages for  $S_1$  (black line) and  $S_2$  (red dashed line) with associated standard deviations, as functions of reduced temperature,  $T^*$ .  $S_1$  is characterized by values close to zero and a low standard deviation, signifying the lack of nematic phase. In contrast,  $S_2$  has a high variation among snapshots in the region  $T^* = 0.5 - 3.5$ . **Panel b**: variation of  $S_1$  (black) and  $S_2$  (red) between a series of microstates for D4 at  $T^* = 2.9$ . The x axis represents the number of snapshots  $S_2$  fluctuates randomly between low (0.2) and high (0.8) values, signalling some variation in order among neighbours, compounded by a large variation in the range within which molecules adopt an antiparallel alignment.

### 3.4.2 Equal Partial Charges, Different Molecular Lengths

With series C, molecules of varying lengths were considered, with polar groups located in similar, accessible locations. Here, dipole length and dipole moment increased



(a)



(b)

Figure 3.8: Snapshots of D4 system reflecting weaker (**panel a**, lower temperature) and stronger (**panel b**, higher temperature) antiparallel motifs. The stronger monolayer antiparallel alignment with increasing temperature in the region  $T^* = 0.5 - 3.5$  is attributed to the entropic effect of extended antiparallel, surface-adsorbed phase.

proportionally, as charges on end groups were held constant. Strongly and weakly polar molecules were examined using two sets of charges: series  $C-h$  had high charges similar to D1, so that  $C1-h$  and D1 were identical, while series  $C-l$  had lower charges, so that  $C4-l$  and D4 were identical (see Table 3.1).

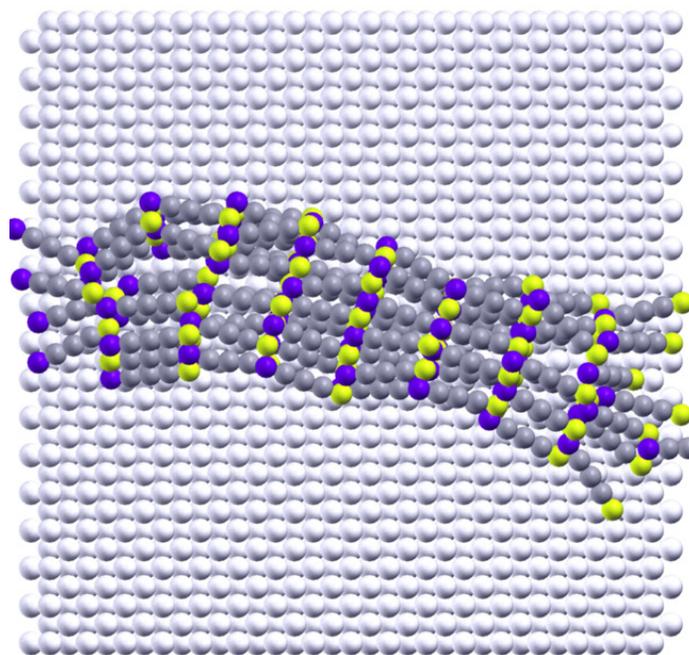


Figure 3.9: Strong alignment for long, high-charge molecules. A low-temperature snapshot  $C4-h$  is shown

### High charges

The more polar series  $C-h$  exhibited a strong tendency towards the formation of three-dimensional crystalline structures, as lateral interactions overwhelmed surface-molecule attraction.  $C-h$  molecules generally formed multilayer structures to satisfy this preference, although longer molecules were more likely to have an extensive surface layer. Figure 3.10 presents the fraction of molecules touching the surface as a function of temperature for a variety of systems, including series  $C$ .  $C3-h$  and  $C4-h$  molecules were mostly lying parallel to the surface, with charged groups engaged in crystal-like stacking (see, for example, Figure 3.9). Unlike their D counterparts, molecular backbones in the  $C3-h$  and  $C4-h$  condensed phase structures were not aligned in tail-to-tail rows, but rather stacked like logs in a rail fence, allowing charged groups to interact with several opposite charges simultaneously, to the detriment of

backbone dispersive interactions. Large values of  $S_1$  and  $S_2$  (Table 3.2), indicate that most molecular dipoles were pointing in the same direction in these systems.

### Low charges

The weakly-polar C- $l$  analogues exhibited dramatically different temperature-dependent behaviour (see Figure 3.11). Weaker electrostatic forces allowed for a mostly single-layer distribution of the molecules on the surface. Because of a greater balance between electrostatic and dispersive contributions to the potential energy surface, a more complex phase behaviour characterized systems C- $l$ , as illustrated by the temperature dependence of the  $S_2$  order parameter [Figure 3.11(a)]. Simple 2-D crystal ordering in the short C1- $l$  molecule gave place to antiparallel domain formation upon addition of spacer atoms in C2- $l$  and C3- $l$  [see Figure 3.11(b)]. At higher temperatures, mainly antiparallel rows were observed [Figure 3.11(c)].

### 3.4.3 Polymorphism

Entropic effects and lateral dispersive interactions in the long-backbone molecule C4- $l$ /D4 led to two polymorphs in this system: a metastable nematic phase,<sup>189–193</sup> found often at higher condensed-phase temperatures [Figure 3.11(d)], and the stable antiparallel-domains phase discussed in relation to Figure 3.5(d) above. Lengthy PTMC runs (over  $5 \times 10^{10}$  steps) eliminated completely the nematic phase, but its persistence throughout the simulation is worth noting. The presence of the two polymorphs (the nematic and antiparallel states), with distinct order parameters, led to a bimodal distribution for all of the order parameters for model C4- $l$ /D4, and ill-defined averages for these parameters. This is illustrated at some length in Figure 3.7 and in the contrast between Figures 3.12 and 3.13. The figures provide a comparison between the fully converged D4 model (Figure 3.8), the metastable C4- $l$ /D4 polymorph (Figure 3.12) and G1<sub>2</sub> -  $l$  (Figure 3.13). The latter did not exhibit polymorphic behaviour.

Whereas  $S_2$  for G1<sub>2</sub> -  $l$  showed a statistical spread around the 0.3 average at  $T^*=1.4$  and  $S_1$  for the same system was 0, both order parameters oscillated around two distinct values in the case of the metastable C4- $l$ /D4 (around 0 and 0.9 for  $S_1$ , and around 0.2 and 0.9 for  $S_2$ ). As the nematic state is not sampled at any temperature in the fully converged D4 system (see Figure 3.7), it appears that the state is kinetically trapped for the C4- $l$ /D4 simulation.

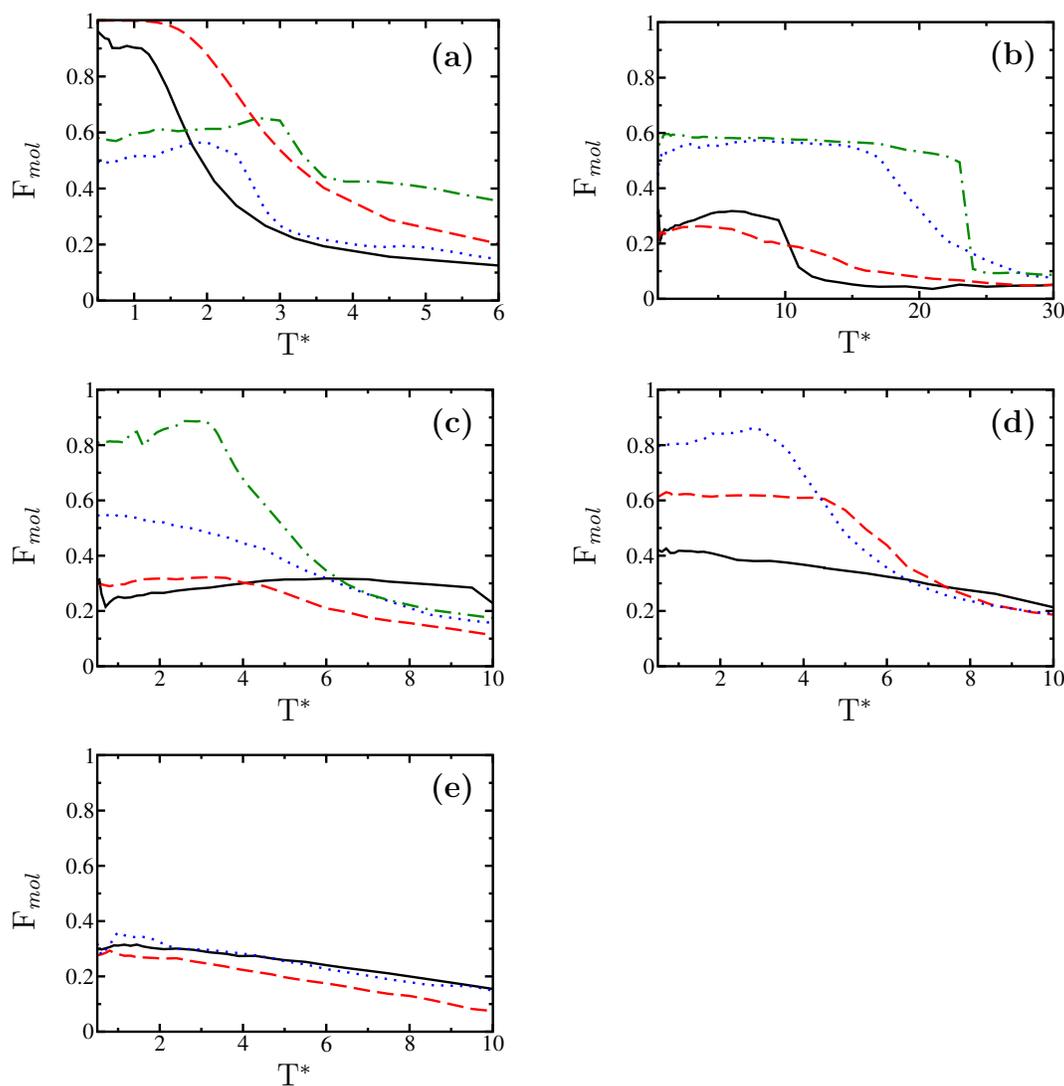
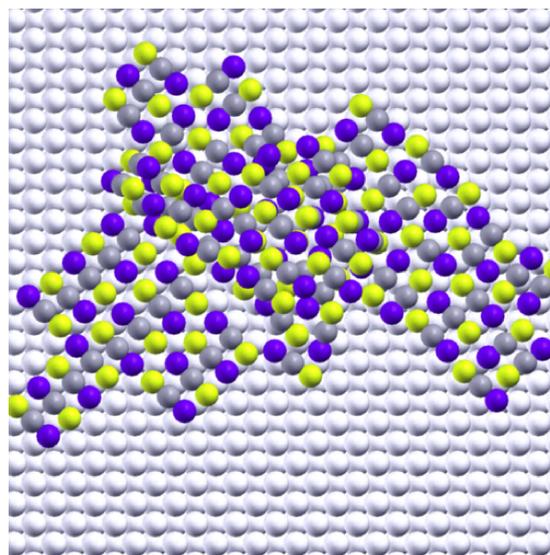
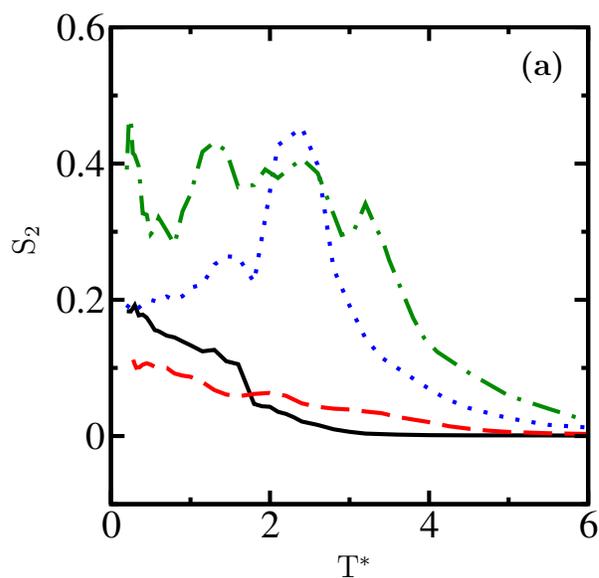
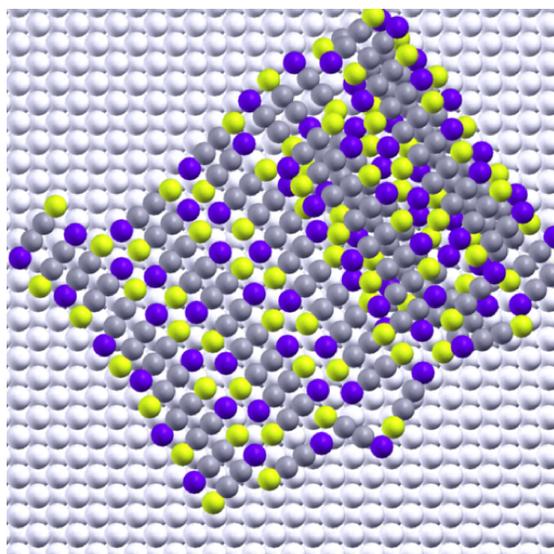


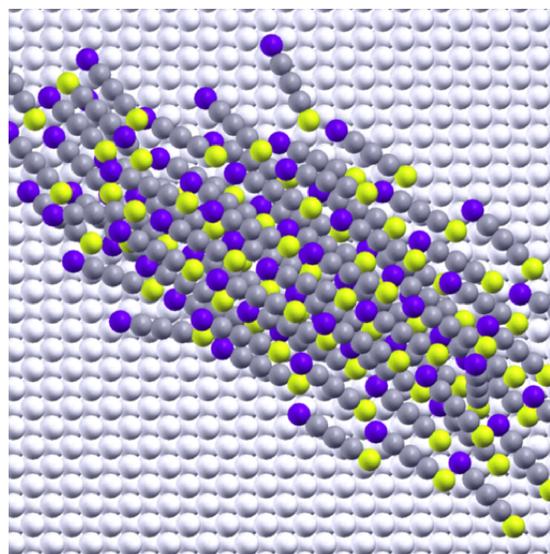
Figure 3.10: Fraction of molecules touching the surface as a function of  $T^*$ . **Panel a – series A**: black line –  $A1(\varepsilon_{ss}^* = 1)$ , red dashes –  $A1(\varepsilon_{ss}^* = 3)$ , blue dots –  $A2(\varepsilon_{ss}^* = 1)$ , green dot-dashed line –  $A2(\varepsilon_{ss}^* = 3)$ ; higher surface potential leads to higher values; stronger charges lead to lower values, characteristic of three-dimensional crystalline state. **Panel b – series C–h**: black line –  $C1-h$ , red dashes –  $C2-h$ , blue dots –  $C3-h$ , green dot-dashed line –  $C4-h$ ; fraction increases with charge separation, up to 0.6, characteristic of the crystalline state (charge-driven assembly). **Panel c – series D**: black line –  $D1$ , red dashes –  $D2$ , blue dots –  $D3$ , green dot-dashed line –  $D4$ ; fraction increases with molecular length, tending towards unity (surface-driven assembly). **Panel d – series L**: black line –  $L1$ , red dashes –  $L2$ , blue dots –  $L3$ ; transition from charge-driven to surface-driven assembly. **Panel e – series G–h**: black line –  $G1-h$ , red dashes –  $G2-h$ , blue dots –  $G3-h$ ; charge-driven assembly; fraction unaffected by geometry for short tail length.



(b)



(c)



(d)

Figure 3.11: Complex phase diagrams in low-charge systems. In (a), the temperature dependence of  $S_2$  for systems  $C1-l$ ,  $C2-l$ ,  $C3-l$  and  $C4-l$  is shown with solid black, dashed red, dotted blue and dot-dashed green lines, respectively. A low temperature snapshot showing  $C2-l$  in antiparallel configuration is shown in (b). High ordering in antiparallel rows was observed for  $C3-l$  at intermediate temperatures -  $T^*=2.5$  is shown in (c). An example of the medium temperature nematic phase at  $T^*=2.5-3.5$  is given in (d) for  $C4-l/D4$ .

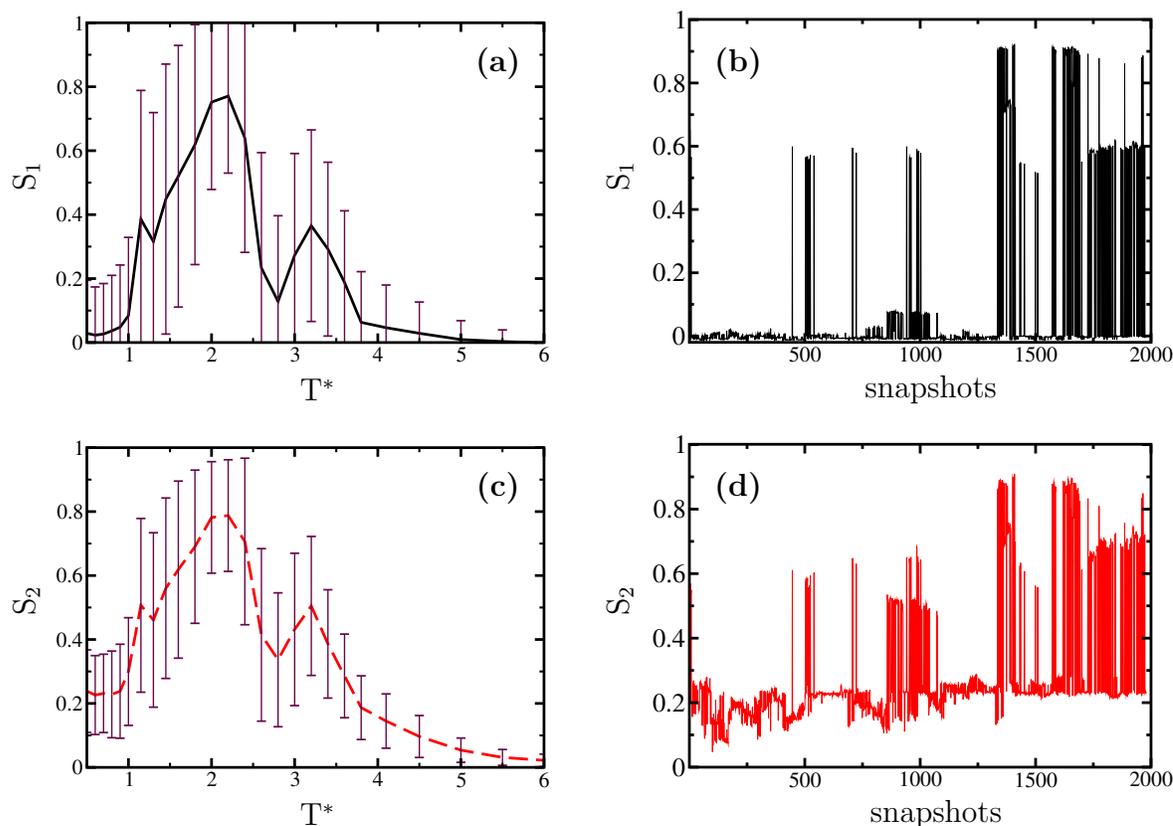


Figure 3.12: Variation of order parameters between polymorphic states in early runs of  $\mathbf{D4/C4-l}$  system. The averages for both  $S_1$  (**panel a**) and  $S_2$  (**panel c**) exhibit high standard deviations at all temperatures, suggesting the adoption of more than one strategy for self-assembly. **Panels b** and **d**:  $S_1$  and  $S_2$  for individual snapshots at  $T^* = 1.45$ , as a function of simulation time. Both  $S_1$  and  $S_2$  fluctuate between high and low values. This indicates an initial coexistence of phases, including nematic (**panel b**). The long-run behaviour is that described in Figures 3.7 and 3.8.

Applications of surface functionalization can be based on a variety of patterning motifs, from two dimensional arrays, to surface-supported clusters, to the formation of independent, self-contained domains on the surface. For example, sensing applications may require organized surface-supported clusters, heterogeneous catalysis or device applications need extended, well-organized two-dimensional arrays of specific densities, while nanofabrication of molecular wires requires sufficiently separated neighbouring wires.

In the latter case, molecules have to interact specifically in one direction while resisting interaction in the other directions. Positioning the partial charges on the outer atoms enables molecules to maximize dipole-dipole interaction by head-tail and

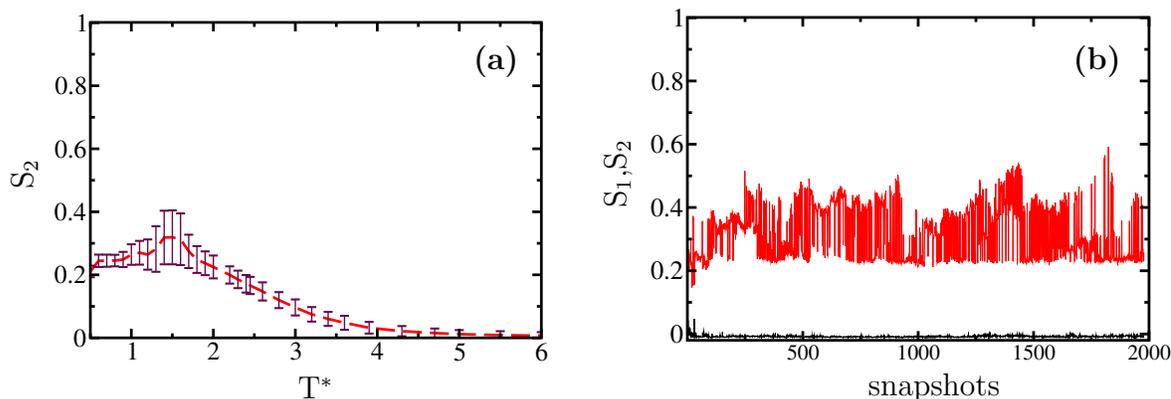


Figure 3.13: Order parameters for  $G1_2-l$  system. **Panel a:** average  $S_2$  and associated standard deviation as a function of reduced temperature,  $T^*$ . The maximum of variation is around  $T^*=1.4$ , with lower temperature range and magnitude relative to  $D4/C4-l$  (see Figure 3.7). **Panel b:** variation of  $S_1$  (black) and  $S_2$  (red) between a series of microstates for  $D4$  at  $T^*=1.4$ .  $S_1$  is close to zero and has a universally low variation, indicating the absence of nematic.  $S_2$  fluctuates randomly between a minimum and a maximum, suggesting the presence of a family of configurations with continuous variation in the long-range order.

lateral alignment. This ensures that a two-dimensional layer is preferred over one-dimensional domains. One way to partially restrict the dipole-dipole interaction to one direction is by embedding the partial charges within the molecule, with unreactive molecular ends.

### 3.4.4 Varying the Molecular Location of Charged Groups

Embedded charge models were represented in the current study by series L (see Figure 3.1). The series comprised three molecular models, with equal length, equal dipole moment, but varying charges and charge separation. In model L1, charges were located in the middle of the molecule in neighbouring positions, with relatively large dispersive groups on each side. Model L1 exhibited amphiphilic behaviour at low-to-medium temperatures: clusters or chains, with charges grouped on the inside and bulky groups pushed to the outside can be seen in Figure 3.14(a). As side substituents decreased and eventually disappeared, models L2 and L3 formed mainly antiparallel structures, as shown in Figure 3.14(b). Moreover, the long, end-group-polarized L3 exhibited polymorphism similar to that discussed for  $D4$ , but with molecular rows in the nematic phase arranged in a standard tail-to-tail distribution [Figure 3.14(b) and (c)].

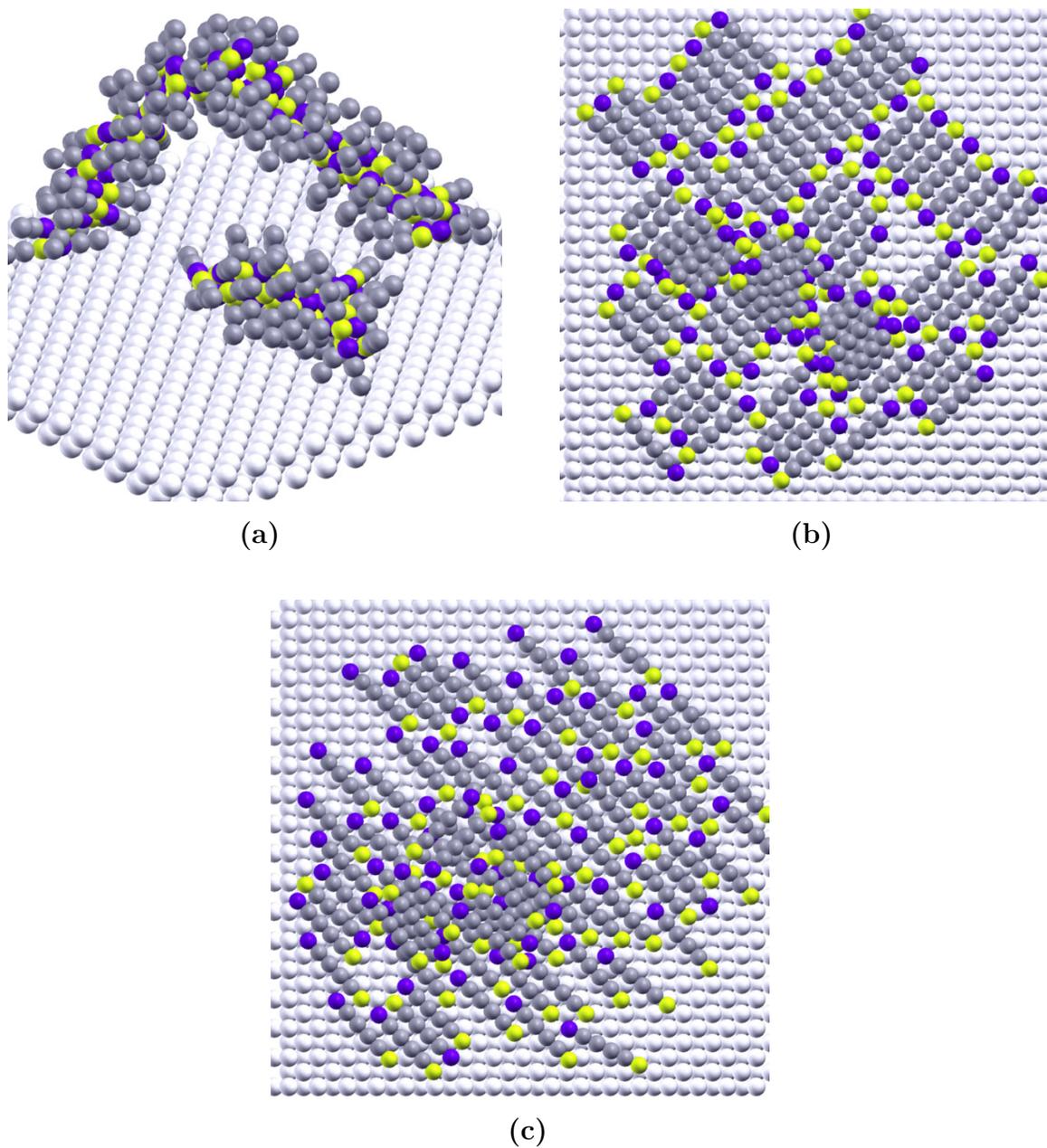


Figure 3.14: The effects of relative positioning of charged atoms within molecules. Snapshots of simulations corresponding to models L1 at low to intermediate temperatures (a), L3 in its antiparallel phase (b), and L3 in a compact nematic - [panel (c)] configuration.

### 3.5 Series G: Asymmetric Dipoles

Molecular asymmetry was introduced in series G by using bulky substituents and non-linearity. Both modifications were expected to interfere with the close-packed ar-

rangements discussed above for linear molecules. Models  $G-h$  have a pair of strong partial charges and one bulky substituent, arranged at different angles versus the dipole direction. A low-charge series,  $G-l$ , was also considered, as were longer tailgroups of the lower-charge series. Beside geometric asymmetry, these models consider an additional aspect of asymmetry of the charge distribution: whereas models C, D and L presented oppositely charged atoms placed symmetrically within the molecule, models G have their charge distribution concentrated on one side of the molecule (a headgroup), with the other side being uncharged and exclusively dispersive (a tailgroup). As observed for symmetric models with strong charges (A,  $C-h$ ), models  $G-h$  formed strongly crystalline structures. However, the presence of bulky groups in  $G-h$  impeded crystal growth in one or two directions, which resulted in the formation of crystalline wires, with bulky groups pushed towards the outside of the chain [see, for example, Figure 3.15(a)]. Similar results were obtained for  $G2-h$  and  $G3-h$ .

As charged groups were stacked in a variety of ways within these molecular wires, low values of the order parameters were observed, despite strong dipolar ordering in these systems. Sharper molecular angles were harder to accommodate in condensed phases, leading to a relative loss of dipolar alignment in condensed phases of  $G3-h$  and  $-l$ .

As before, weaker charges led to an increased dominance of surface-molecule interactions in the resulting structures. Chain-like structures in high-charge models became surface bound [see, for example, Figure 3.15(b) and (d)] in low-charge models, and monolayers exhibited atomic localization at pit surface sites, as observed previously for strongly-attractive surfaces. Figure 3.15(c) for example, shows how dipolar alignment and surface attraction can be reconciled in model  $G3_1-l$ , for example, by locating the bulky, uncharged atom above the dipolar plane.

In effect, the case presented in Figure 3.15(c) is one where the dipolar attraction overcame the surface-tailgroup interaction, leading to compact phases where the dipoles could interact with multiple neighbours. In contrast, molecule  $G1_1-l$  had no opportunity to push its tailgroup out of the way and form compact dipolar structures, leading to the formation of planar antiparallel chains.

A different situation arose when longer molecular backbones lead to stronger tail-tail and tail-surface interactions, and tailgroup-directed assembly. Systems  $G_N-l$  where  $N = 2, 3, 4$  provide illustrations of tail-driven assembly. Figure 3.16 presents two systems where compact, planar arrangements of the tailgroups were the dominant feature, with dipolar interactions determining the relative arrangement of molecular

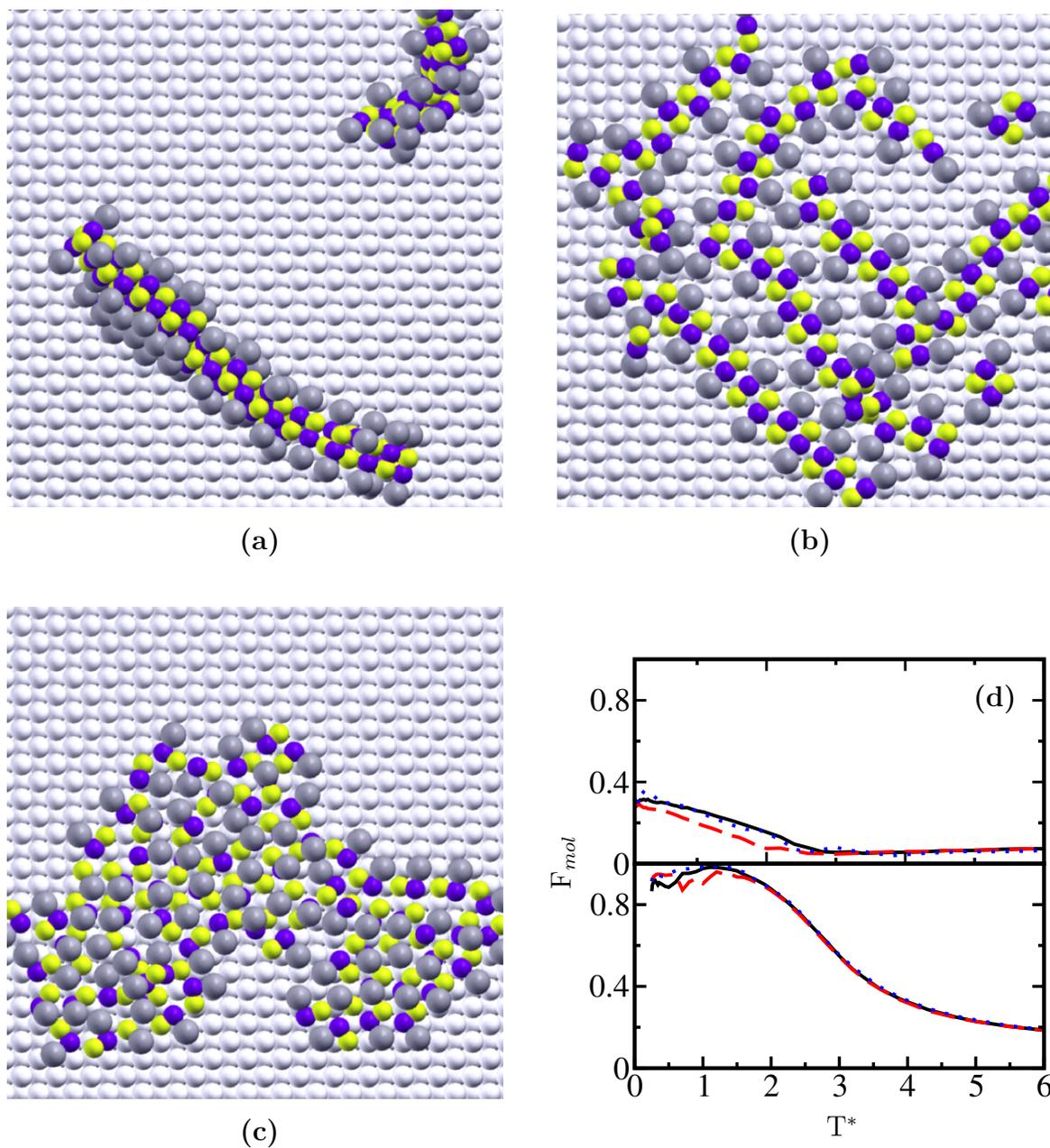


Figure 3.15: The effect of asymmetry on self-assembled structures. Low-temperature snapshots from simulations corresponding to models  $G1-h$ ,  $G1_l-l$  and  $G3_l-l$  are presented in panels a-c, respectively. The graph in panel d shows the fraction of molecules in contact with the surface ( $F_{mol}$ ) for the  $G-h$  series (above) and the  $G-l$  series (below), as a function of reduced temperature,  $T^*$ . Black solid lines, red dashes and blue dots correspond to systems G1, G2 and G3, respectively.

rows (aligned or antiparallel). In long-tail systems with two distinct interaction directions, it was useful to quantify the alignment of the second relevant direction. As

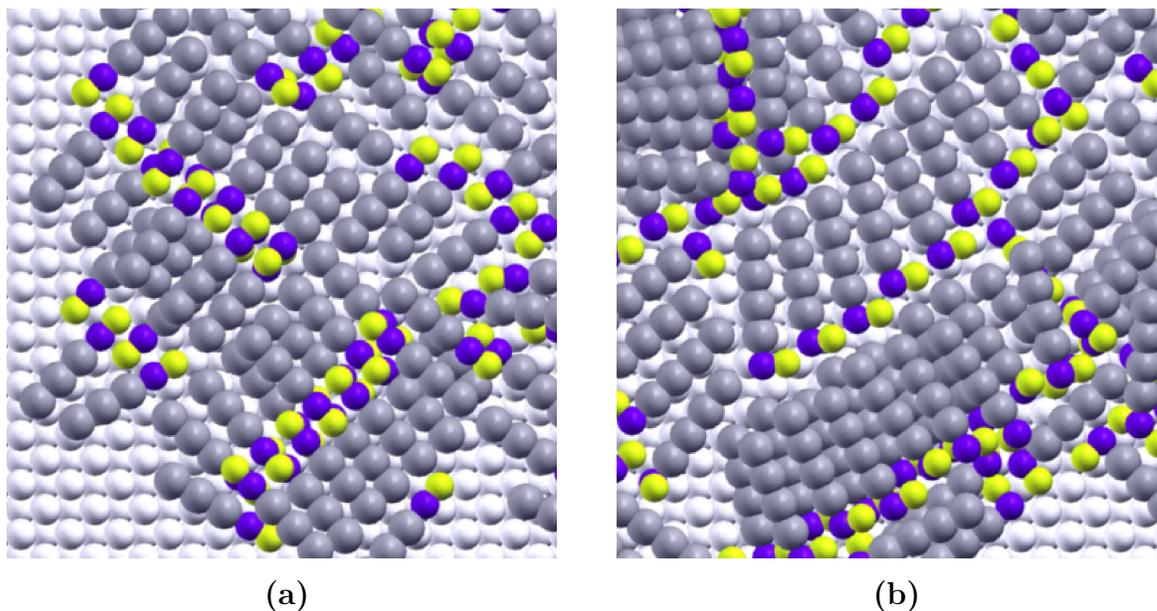


Figure 3.16: Tail-driven assembly in asymmetric molecules. Low-temperature snapshots from simulations corresponding to models  $G1_3 - l$  and  $G3_4 - l$  are presented in panels a and b, respectively.

shown in Figures 3.17 and 3.18, tailgroup alignment was more regular than dipolar alignment in systems with tailgroup-directed assembly (see also the dipole-based and tailgroup-based order parameter values presented Table 3.2 and Table 3.3, respectively).

Most molecules of interest to SA scientists are asymmetric in nature, often exhibiting distinct polar and apolar molecular regions. Despite more complex interaction patterns, similar effects have been observed in experimental studies seeking a fundamental understanding of the dipole/steric interplay in SA systems. For example, thiophenes substituted with long (C18) alkylamide groups adopted distinct alignment strategies depending on the binding location of the alkylamide group to the thiophene ring.<sup>90</sup> The long alkyl chains,  $\pi - \pi$  interactions, the availability of hydrogen bonds and the polarity of the amido group provided a strong blueprint for the formation of self-assembled structures on graphitic surfaces. However, the overall pattern changed significantly depending on the molecular polarity, modified by varying the binding location of the alkyl amide group on the thiophene ring. When the C18-amido group was bound at the position 2 of the thiophene ring, the alkyl backbone and the dipole moment were roughly perpendicular, as in our  $G3_3 - l$  and  $G3_4 - l$  models, whereas

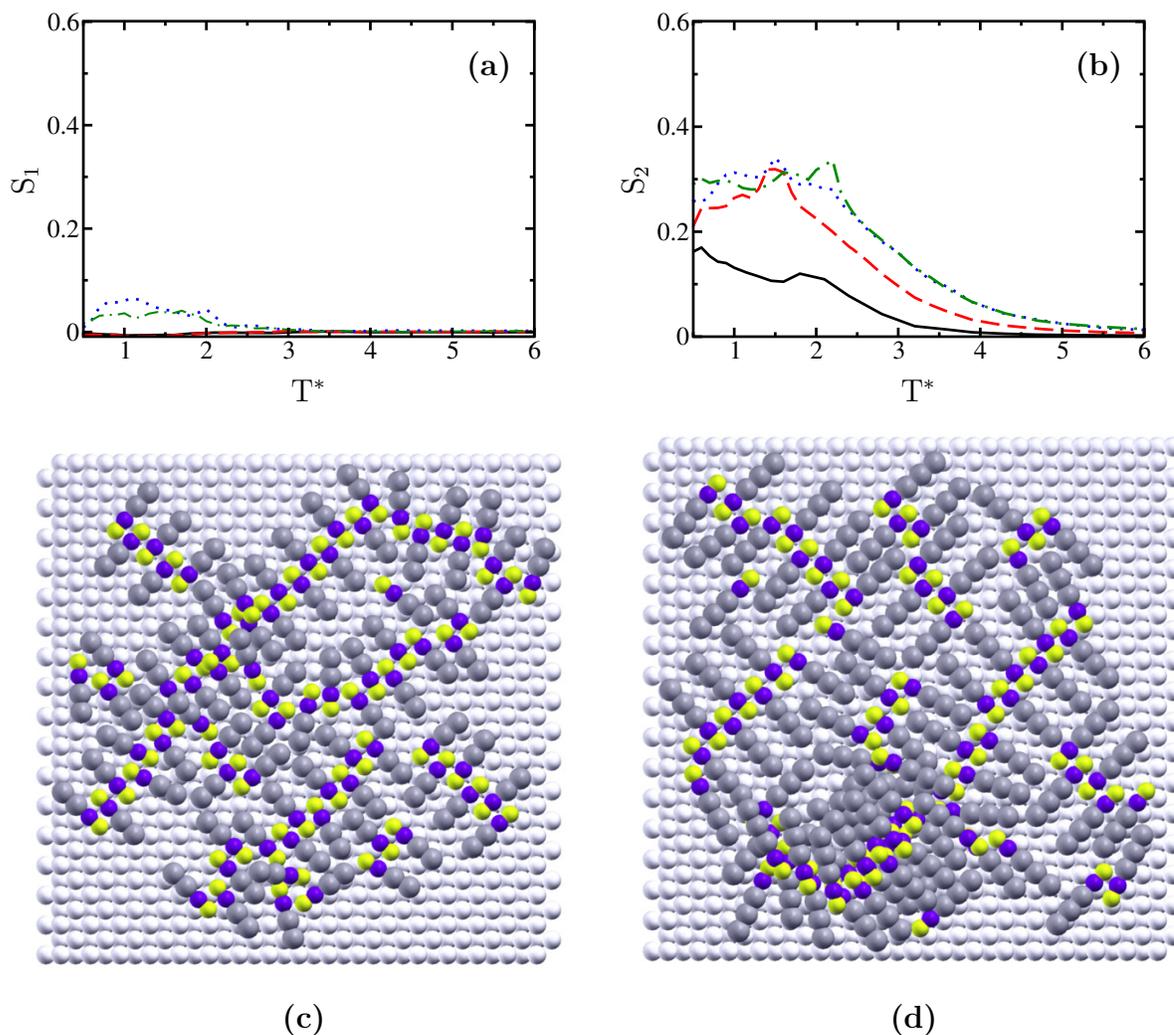


Figure 3.17: Effects of tail length for the SA of molecules  $G1-l$ : **Panels a** and **b**: averages  $S_1$  and  $S_2$  respectively, as a function of reduced temperature,  $T^*$  (x axis), when the tail is one (black line), two (red dashes), three (blue dots) and four (green dot-dashed lines) atoms long. Increasing tail length has no influence over  $S_1$  (panel a) and has a slight influence on  $S_2$  (panel b) by stabilizing the head-group driven antiparallel monolayer over longer distances, as can be observed in **panel c** for  $G1_2-l$ . Further tail increase causes layers to stack due to a high surface density. The head-group driven antiparallel assembly is preserved across layers (**panel d** -  $G1_3-l$ )

in the 3-substituted thiophene, the two entities were roughly collinear, in a fashion similar to the trends explored by our  $G1_3-l$  models. The overall effect was that the 2-substituted thiophenes assembled in a head to tail packing with a parallel alignment, whereas 3-substituted ones adopted head-to-head packing and an antiparallel

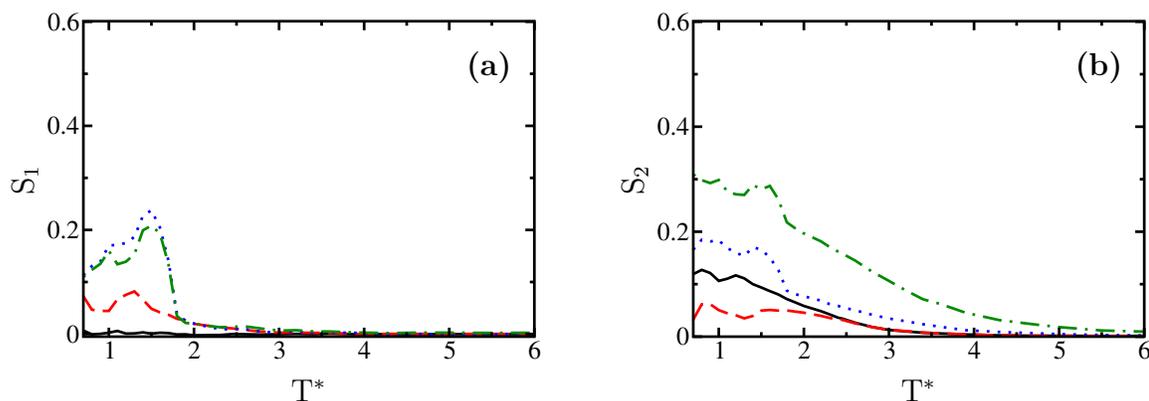


Figure 3.18: Comparison of  $S_1$  (**panel a**) and  $S_2$  (**panel b**) between  $G2_1 - l$  and  $G2_3 - l$ . The x axis represents reduced temperature,  $T^*$ .  $G2$  geometry is biaxial, leading to two versions of each parameter, one relative to the head group and one relative to the tail group:  $G2_1 - l$  head - black line,  $G2_1 - l$  tail - red dashes,  $G2_3 - l$  head - blue dots,  $G2_3 - l$  tail - green dot-dashed line.  $G2-l$  series is characterized by low order parameters in general. A longer tail size facilitates a stronger alignment among molecules, particularly along the tail axis.

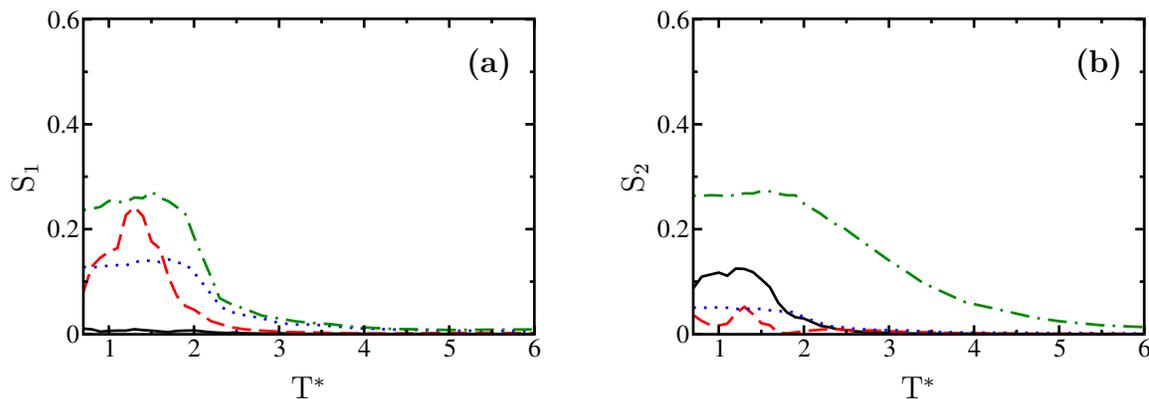


Figure 3.19: Comparison of  $S_1$  (**panel a**) and  $S_2$  (**panel b**) between  $G3_1 - l$  and  $G3_4 - l$ . The x axis represents reduced temperature,  $T^*$ .  $G3$  geometry is biaxial, leading to two versions of each parameter, one relative to the head group and one relative to the tail group, with the two axes being perpendicular to each other:  $G3_1 - l$  head - black line,  $G3_1 - l$  tail - red dashes,  $G3_4 - l$  head - blue dots,  $G3_4 - l$  tail - green dot-dashed line.  $S_1$  and  $S_2$  for  $G3-l$  behave similar to those for  $G2-l$ .

alignment, similar to those shown in Figure 3.16(b) and (a), respectively.

Table 3.3: Order Parameters for series G relative to the TAIL axis for temperatures around  $T^*=1$ .

Molecule	$S_1$	$S_2$	$S_4$
<b>G2-l</b>			
G2 <sub>1</sub> - l	0.04	0.06	0.02
G2 <sub>2</sub> - l	0.00	0.22	0.10
G2 <sub>3</sub> - l	0.15	0.03	0.23
<b>G3-l</b>			
G3 <sub>1</sub> - l	0.24	0.05	0.01
G3 <sub>2</sub> - l	0.04	0.07	0.02
G3 <sub>3</sub> - l	0.10	0.22	0.25
G3 <sub>4</sub> - l	0.25	0.26	0.25

### 3.6 Conclusion

In the present chapter, we examined some of the basic molecular variables controlling the emergence of order from disordered phases in surface self-assembly, by following several simple molecular models. We found that although simple dipoles formed the expected crystalline structures at low temperatures, modulation by the surface interaction and substituent effects significantly altered adsorbed structures. Longer molecules experienced stronger (cumulative) surface attraction and tended to form monolayers with lateral alignment in parallel (nematic) or antiparallel configurations. Stronger molecular dipoles led to multilayer aligned phases, in which many polar groups could be found in close vicinity. Limiting the availability of contact points by substituent effects, in addition to the steric effect of such apolar substituents further altered self-assembled structures, with the formation of molecular wires, stabilized on or off the surface by the relative strength of the interactions, in a number of cases.

Clearly, intermolecular interactions in experimental systems span a significantly broader range than those considered here, although, as shown in the current work, fundamental features can often be captured through simple models. Further investigations are desired, such as studying structure formation in experimentally-relevant molecules such as substituted polar acenes, using a combination of classical PTMC and density functional theory. However, at the base of all types of supramolecular interactions lay fundamental electrostatic and geometric effects.

In this chapter we explored the wealth of self-assembled phases generated by the interplay between local interactions and offered an understanding of how these effects can be exploited to alter molecular behaviour upon adsorption. In Chapter 4 we take

a step back from the long range view of surface self-assembly, and look more closely at local interactions upon adsorption. To this end, we choose a specific class of surface self-assembled systems and use electronic structure analysis techniques to explore the adsorption of individual molecules and dimers.

## Chapter 4

# Structural Analysis of Helicene Molecules Adsorbed on Symmetric Surfaces<sup>†</sup>

### 4.1 Introduction

In Chapter 3 we demonstrated how a small set of local interactions between molecules with simple shapes can give rise to complex phases with varying degrees of order upon substrate adsorption. Simple molecular features are sufficient to induce surface self-assembly resembling experimental behaviour. In real systems, however, the detailed electronic features of building blocks, how they couple between units and how they interact with the surface, crucially determine the properties of phases. Simple molecular features are not sufficient in predicting or explaining how real molecules self-assemble in specific ways. An understanding of real systems require a detailed understanding of the electronic structure of the building blocks.

In this chapter we focus on the role of short-scale interactions in surface self-assembly. Contrasting the Monte Carlo many-body simulations of Chapter 3, in this Chapter we perform density functional theory geometry optimizations of monomers and dimers in gas phase and adsorbed on a small size surface. We focus on a specific class of experimentally popular systems, those of helicenes on copper, silver and gold surfaces.

---

<sup>†</sup>This chapter is based on article “*Structural analysis of helicene molecules adsorbed on symmetric surfaces*” by E. Tuca and I. Paci published in *Phys. Chem. Chem. Phys.*, 2019, **21**, 9189-9199, DOI:10.1039/C9CP00284G.

There has been significant work seeking to understand the effects of surface selection and lateral molecular interactions in helicene assembly, with particular focus given to their enantioselective behaviour. We contribute to this with our investigation of the role of stacking in helicene adsorption. Our findings in this area are of significance to those interested in the enantioselective properties of adsorbed helicenes. Additionally, they demonstrate the crucial role local interactions play in self-assembly.

## Chapter layout

Section 4.2 provides some context for our study by detailing the structural, electronic and chiral properties of helicene molecules. In Section 4.3 we present the methodological details of our study, including the choice of models, the choice of analysis and the computational details.

In Section 4.4 we begin our electronic structure analysis with the planar reference molecule, Coronene. In Sections 4.5 – 4.7 we discuss helicene stacked dimers in gas phase, surface - adsorbed monomers and adsorbed dimers, respectively. We close this chapter with a concluding section.

## 4.2 Helicene Properties

The physical and chemical properties of helicenes are rooted in their aromaticity and their chirality. Helicenes are highly aromatic, despite their lack of planarity.<sup>194</sup> Because of the size of the conjugated system, large deviations from planarity in the overall molecule arise from small distortions in the individual rings.<sup>195</sup> Helicenes exhibit planar chirality (as opposed to the point chirality normally considered in basic discussions of stereochemistry), arising from the out-of-plane twist of the molecule. The sense of the helical twist may be clockwise(right-handed) or anti-clockwise (left-handed), with size-dependent specific optical rotation.<sup>68</sup> By convention right-handed helices are labelled *plus* (**P**) and left-handed ones are labeled *minus* (**M**). In general, right-handed compounds have a (+) dextrorotatory specific rotation and left-handed compounds have a (-) levorotatory specific rotation. Their intrinsic, continuously chiral polyaromatic skeleton crafts a chiral space which is responsible for subsequent enantioactivity.

### 4.2.1 Structure and Mechanical Properties

A  $[n]$ helicene represents the molecule resulting from the ortho-condensation of  $n$  benzene rings. Because of the inevitable intramolecular repulsion, helicenes with  $n > 3$  adopt a helical topology, which produces interesting chiral and electronic – disruption of aromaticity – effects. The intramolecular steric repulsion resulting from the condensed structure leads to a conformational distortion of the  $\pi$ -system. Unlike their planar polyaromatic counterparts, the 3-D nature of helicenes renders them chemically stable and soluble in common organic solvents and  $\pi$ -conjugated materials.

The general structure of helicenes has been investigated extensively, both in isolation and relative to their non-helical counterparts.

#### Pitch

The pitch of the helicenes can be defined by fitting the Cartesian coordinates of the corresponding atoms to a screw-type curve, with the standard deviation giving a measure of how much the atoms deviate from a regular helix. In general the pitch is small for helicenes as they form nearly regular helices. The pitch increases strongly from [4]H to [6]H, with only little further change.<sup>196</sup> Mean radius decreases with the same trend because the aromatic units avoid short distances by twisting.<sup>196</sup>

#### Dihedral angles

Another way to quantify the shape is through dihedral angles of consecutive carbon atoms in the inner, middle and outer parts of the molecules.<sup>58</sup> Within the series, the equilibrium structure has a general convergence towards a helical structure with interpitch distance of 3.75 Å, and a positive angle, indicating slightly more than six benzene rings in one helical turn.<sup>70</sup> The convergence is reached for  $n = 14$ . Before convergence, the dihedral angles increase from [4]H (18.6°) to 5H (30.5°), then decrease for 6H (28.9°), after which they remain approximately constant (28°).<sup>58</sup> For nonbonding  $C - C$  distances between overlapping rings, the shortest are 309 pm for [6]H and 304 pm for [7]H, close to twice the van der Waals radius of carbon.<sup>58</sup> This information suggest that the larger change in structure for helicenes, compared to their planar hydrocarbon equivalents, occurs between four and six aromatic rings.

## Stability

The stability of  $[n]$ helicenes is typically deduced from the extension of  $[n - 1]$ helicene to  $[n]$ helicene through addition of naphthalene, in order to conserve all bonds and types of each carbon atoms in the process. This gives an incremental energy of stabilization. Phenacene, the planar equivalents of helicenes, are slightly stabilized with the addition of another benzene ring, by a practically constant amount of energy, in agreement with the classical theory of Clar. This is in contrast with helicenes, which are destabilized by additional rings.<sup>70</sup> There is an anticipated stabilization upon the completion of the helical turn, from  $[6]$ helicene, through the effect of interaction between stacked benzene rings.<sup>197</sup> For higher helicenes, a general convergence of the incremental (de)stabilization energy occurs at about  $11kJ$ .

## Elasticity

The mechanical properties of helicenes are of great importance in nanosciences. The elasticity of the helical structures is decreasing linearly with size.<sup>70</sup> Normal mode analysis reveals that the lowest vibrational modes correspond to deformations of the helical skeleton, linear combination of symmetric and antisymmetric stretching modes along the  $z$  axis and twisting of the helical screw ( $xy$ -plane). Elasticity decreases with size. With respect to stretches/contractions along the  $z$  direction, the force constant has been found to be ten times higher than that for the DNA helix.

## 4.2.2 Aromaticity and Opto-Electronic Properties

### $\pi$ -system

$[n]$ acenes and  $[n]$ phenacenes are the linear and zig-zag equivalents of  $[n]$ helicenes, respectively. They share identical local connectivities, Kekule counts, conjugated circuits and other graph theoretical descriptors, suggesting similar aromaticity between the series. On the other hand, the lack of planarity in helicenes suggests a reduction of local aromaticity. Despite the lack of planarity inherent in helicenes, only a very slight loss of aromaticity is detected, compared to the planar equivalents - phenacenes.<sup>194</sup> The rings in helicenes are almost equally delocalized as those of phenacenes, and the corresponding loss of planarity is small, smaller than that which benzene can experience without substantial loss of electron conjugation.<sup>195</sup> Additionally, helicenes have in common with phenacens a difference in aromaticity between terminal and internal

rings. This intramolecular difference is higher than the difference between helicenes and corresponding phenacenes.<sup>70</sup>

### Magnetic and electronic properties

For larger systems, the NICS measure considers the terminal rings of helicenes more aromatic than those of phenacene equivalents.<sup>194</sup> This is attributed to magnetic couplings with neighbouring rings above. Magnetic susceptibilities of helicenes are found to vary linearly with  $n$ . While [16]phenacene shows convergence to the zigzag infinite polymer unit, [16]helicene suffers from magnetic shielding from nonadjacent rings, but does show some signs of convergence to the infinite helical unit.<sup>195</sup> The redox potential in the series converges to the value of  $-2.2V$  for reduction, and  $0.65V$  for oxidation, electron affinity to  $0.85eV$  and ionization potential to  $6.26eV$ .<sup>70</sup>

### Optical properties

Chiroptical properties are connected to the molecular structure of helicenes in a complicated way. The electronic excitations of  $C_2$ -symmetric helicenes are symmetric  $A$  transitions and non-symmetric  $B$  transitions. If the  $z$  axis is taken as the two-fold symmetry axis,  $A$  transitions are polarized along the  $z$  axis, while  $B$ -type transitions are orthogonal to  $z$ , leading to a polarization in the  $xy$  plane.  $A$  transitions tend to have positive rotatory strength in left-handed helicenes.<sup>58</sup> Excitation energies calculated by TD-DFT are in good agreement with experimental data for  $n = 4 - 7$ . This allows for some confidence in predicting trends for higher  $n$ . Excitation energies monotonically shift to longer wavelengths. Most intense peak is a  $B$  type transition. Computed excitation energy corresponds to a band gap of  $2.60eV$  for [14]helicene, close to the predicted value of  $2.49eV$  for  $[\infty]$ helicene.<sup>70</sup> For [6]helicene the left-handed enantiomer has a strong negative band at low wavelength (determining the rotatory dispersion in the visible) and a positive absorption for shorter wavelength. The low energy bands arise from transitions polarized in the  $xy$ -plane ( $B$ ) and positive high-energy bands are caused by  $A$ -type transitions along the  $z$  axis. The lowest excited state has  $B$  symmetry, with the large  $x$  and  $y$  contributions to the rotatory strength nearly cancelling out.

### 4.2.3 Chirality

The helical chirality of helicenes is of central importance to helicene chemistry. The sense of the helical twist may be clockwise(right-handed) or anticlockwise (left-handed) with size-dependent specific optical rotation.<sup>68</sup> As a general rule, right-handed compounds have a (+) dextrorotatory specific rotation and left-handed compounds have a (-) levorotatory specific rotation. Their intrinsic, continuously chiral polyaromatic skeleton crafts a chiral space which is responsible for subsequent enantioactivity.<sup>71</sup>

## 4.3 Methodology

The local interplay between intermolecular and molecule-surface interactions in adsorbed monomers and dimers was considered below using a DFT/PBE-vdW-TS level of theory in the SIESTA code (details in Ref.<sup>198</sup>). Periodic boundary conditions are used, with enough surface layers considered to allow convergence of the surface energy, and large confinement radii for pseudoatomic orbitals, so that binding energies are converged in calculations.<sup>199</sup>

### 4.3.1 Models

#### Molecules

We investigated the adsorption of the series of ortho-fused PAHs, from [1]H (which is benzene) to [7]H. The notation [n]H (where  $n$  is the number of fused rings: 1–7) is used for either individual molecules or systems containing them: Dimers of chiral molecules will be denoted as “like” (L) or “unlike” (U) [5]H[5]H, [6]H[6]H, [7]H[7]H, etc., depending on whether they are made up of molecules of the same, or opposite chirality, respectively. [4]H is achiral, so only one [4]H[4]H dimer will be considered.

#### Surfaces

As substrates, three highly symmetric (111) surfaces were selected: Ag(111), Au(111) and Cu(111). All systems are treated using periodic boundary conditions. Box sizes were chosen large enough to ensure no significant intermolecular interaction between neighbouring images, thus focusing our study to stacking effects. Three layers of surface atoms were used. The bottom-most surface layer was held frozen in the optimized geometry of the bulk crystal structure, while the top two layers were allowed

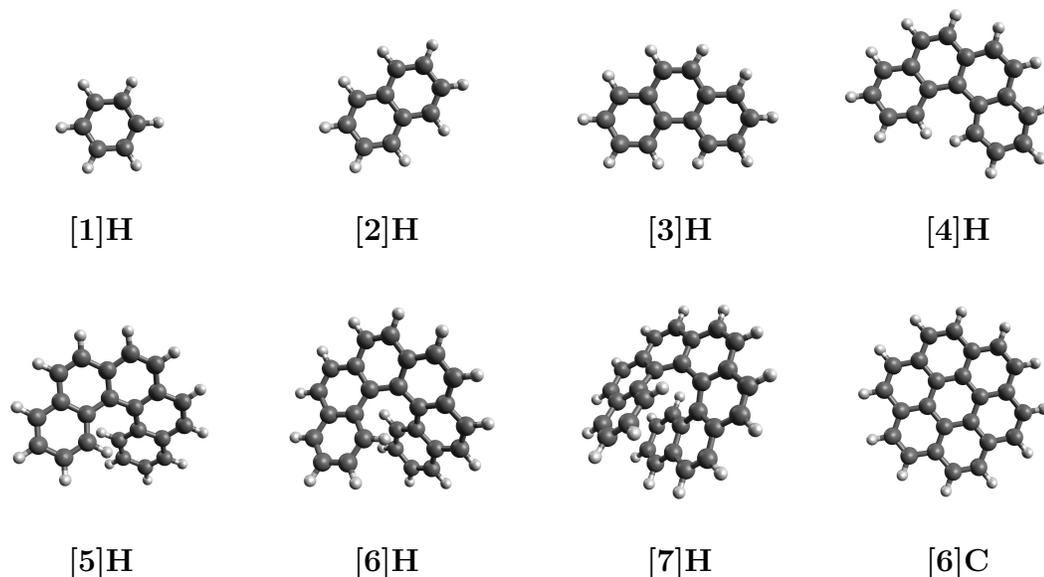


Figure 4.1: Optimized structures of PAHs series investigated. Top row from left to right: benzene ([1]H), naphthalene ([2]H), phenanthrene ([3]H), tetrahelicene ([4]H). Bottom row from left to right: pentahelicene ([5]H), hexahelicene ([6]H), heptahelicene ([7]H), coronene ([6]C).

to reorganize. Identical surface structures and simulation box sizes were used for all surface-bound systems. Figure 4.2 displays the geometry of the three surfaces following their optimization with 2D periodic boundary conditions.

### Helical turn

Figure 4.3 presents the structure of the molecules of focus investigated in this work. [5]H, [6]H, [7]H capture the effect of helicity on coinage metals. [6]H, with six condensed aromatic rings in its skeleton, has the closest geometry to that of [6]C, and differs from its planar homologue by the helical topology. [5]H, having one less condensed ortho-ring, contributes with a lower surface footprint, but may resemble [6]C more due to its less pronounced helical structure. Finally, [7]H has a strong helical character and intramolecular  $\pi$  stacking between end rings.

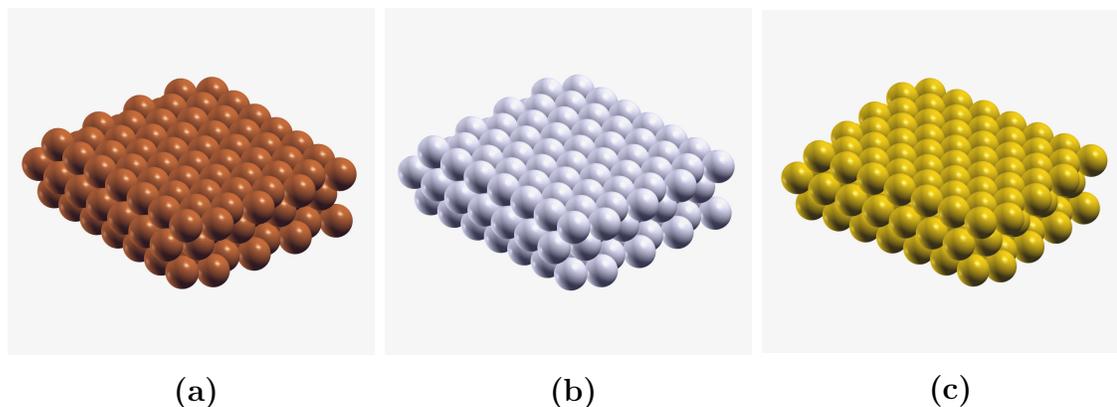


Figure 4.2: Geometry of the three coinage metal surfaces optimized as three-layered 2D sheets. The bottom layer was kept frozen during optimization in order to enforce the [111] symmetry. Panel (a): Cu[111]. Panel (b): Ag[111]. Panel (c): Au[111].

### System size

For each molecule, the structure of monomers and stacked homochiral and heterochiral dimers was considered, in both the gaseous phase and surface-bound environment (for example, the two enantiomers of [6]H are shown in Figures 4.3(f) and (g)). DFT-optimized gas phase dimer structures are presented in Figure 4.4.

### 4.3.2 Topology, Structure and Nomenclature

The topology of various rings in the helicene molecules is discussed further. Rings are numbered consecutively as shown in Figure 4.5(a), starting with the ring closest to the surface.

#### Solid angles between rings

We can quantify the helicity and geometric profile of helicenes by plotting the solid angle and distance of each ring relative to some reference plane. The presence of a surface offers the surface plane as a natural reference point. In the absence of a surface, or for comparisons, an end ring can be taken as the reference. To calculate these solid angles, an average plane is defined for each molecular ring, passing through three non-adjacent ring atoms.

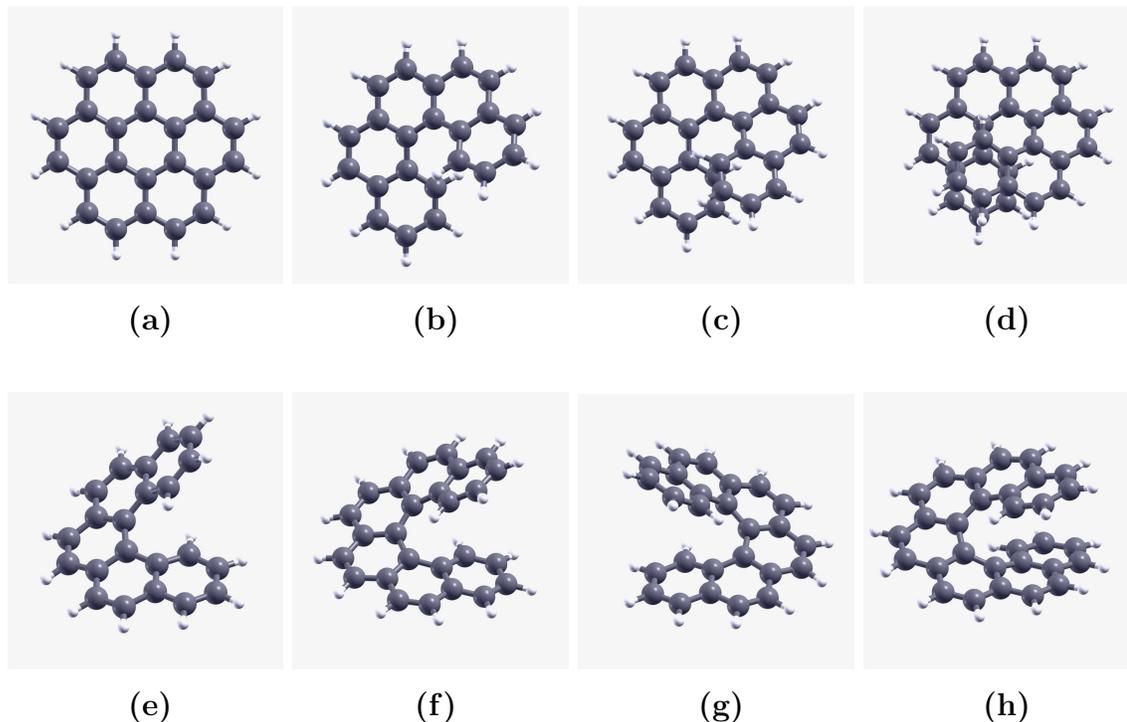


Figure 4.3: Optimized structures of [6]C and helicene gas phase monomers. Top projections of [6]C, (M)-[5]H, (M)-[6]H and (M)-[7]H are shown in panels (a)–(d), respectively. Panels (e)–(h) show side projections of (M)-[5]H, (M)-[6]H, (P)-[6]H and (M)-[7]H, respectively.

The angle made by each successive ring with the plane of Ring 1 will be reported in the following pages and denoted by  $\Theta_i$ , where  $i$  is the ring number. Similarly, the solid angles made by successive ring planes with the surface plane are denoted by  $\Theta_{i,S}$  in the following pages. As a reference, all solid angles are zero for gas phase or adsorbed [6]C. The collection of  $\Theta_i$  and  $\Theta_{i,S}$  for an individual adsorbed molecule forms its angular profile, which provides a molecular signature indicative of the orientation on the surface, and is particularly useful for comparing surface effects on molecular planarity. This angular profile is reflective of the interplay between the molecule-surface dispersive interactions, the repulsive forces responsible for the out-of-plane twist of the molecule, and intramolecular dispersion between end rings, in the longer molecules where this is significant.  $\Theta_i$  values for helicene molecules in the various environments relevant here are presented in Figures 4.6 and 4.7.

In gas phase monomers, ring angles increase monotonously for [5]H and [6]H,

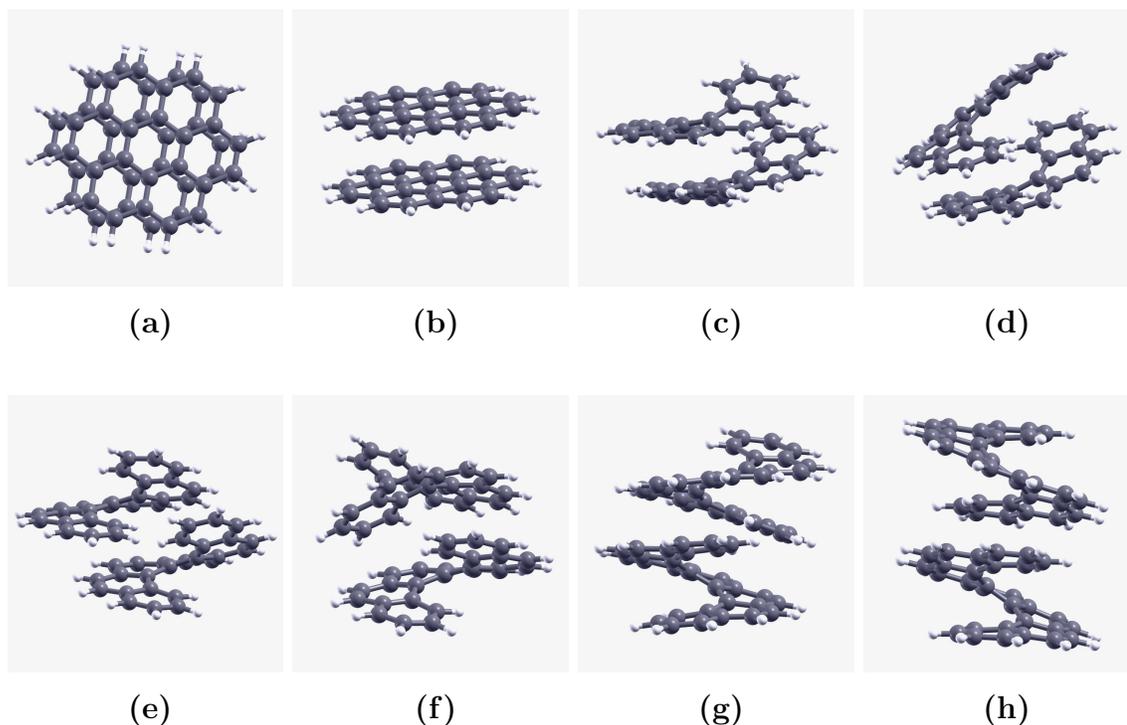


Figure 4.4: Gas phase optimized structures of [6]C and helicene stacked dimers. Panels (a) and (b) show the top and side projections of [6]C[6]C, respectively. Panels (c)–(h) show side projections of like-[5]H[5]H, unlike-[5]H[5]H, like-[6]H[6]H, unlike-[6]H[6]H, like-[7]H[7]H and unlike-[7]H[7]H, respectively.

whereas a decrease is seen after ring 4 for [7]H [see Figure 4.6(a)]. [6]H forms a full helical turn, and [7]H includes the beginning of the second turn, with the rings 1 and 7 dispersively stacked and pulling the molecule into a tighter twist.

### Equivalent rings

In the cases where helicene dimers are considered, the parallelism of equivalent rings in the two molecules can be used as one measure of stacking efficiency. The angles  $\phi_{i,i}$ , made by rings with the same number in the two molecules (also called here equivalent rings) are further reported. In planar, fully stacked dimers such as [6]C[6]C, these angles are zero for all pairs of rings.

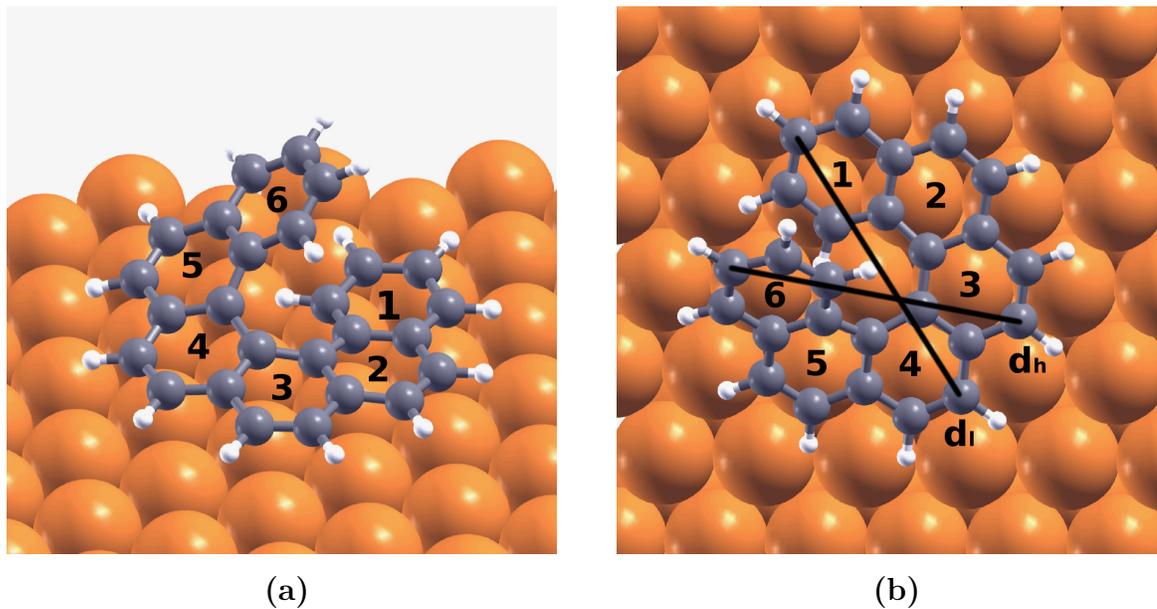


Figure 4.5: Ring ordering for [6]H on Cu. Panels (a) and (b) show the top and side view of adsorbed molecule, with the ring order displayed. Numbering begins from the ring closest to the surface, in ascending order. Panel (b) also shows diameters  $d_l$  and  $d_h$ , defined as the distance between diametrically opposed carbon atoms relative to the helix axis. Two diameters can be defined: one is the distance between the middle ring (Ring 4 in [6]H) and the ring closest to the surface (Ring 1), and the second is formed by the middle ring and the ring furthest from the surface (Ring 6 in [6]H). In the following pages, the projections of these diameters onto the surface plane are reported, referred to simply as “diameters”.

### Molecular footprint

In examining the effect of the surface identity on the molecular footprint, surface projections of the molecular diameters are also reported. The diameters were defined as distances between carbon atoms diametrically opposed relative to the helix axis. There are multiple diameters that could be defined in this manner within one molecule, and in the presence of the surface, two of them are noteworthy. One is defined relative to a carbon atom in ring 1, having the diametrically opposed atom in ring 4. This is the “low” diameter, defined by two atoms closest to the surface and we denote its surface projection as  $d_l$ . The second is defined relative to a carbon atom in the ring furthest from the surface, with the diametrically opposed atom three rings lower. This is the “high” diameter and its surface projection is denoted as  $d_h$ .

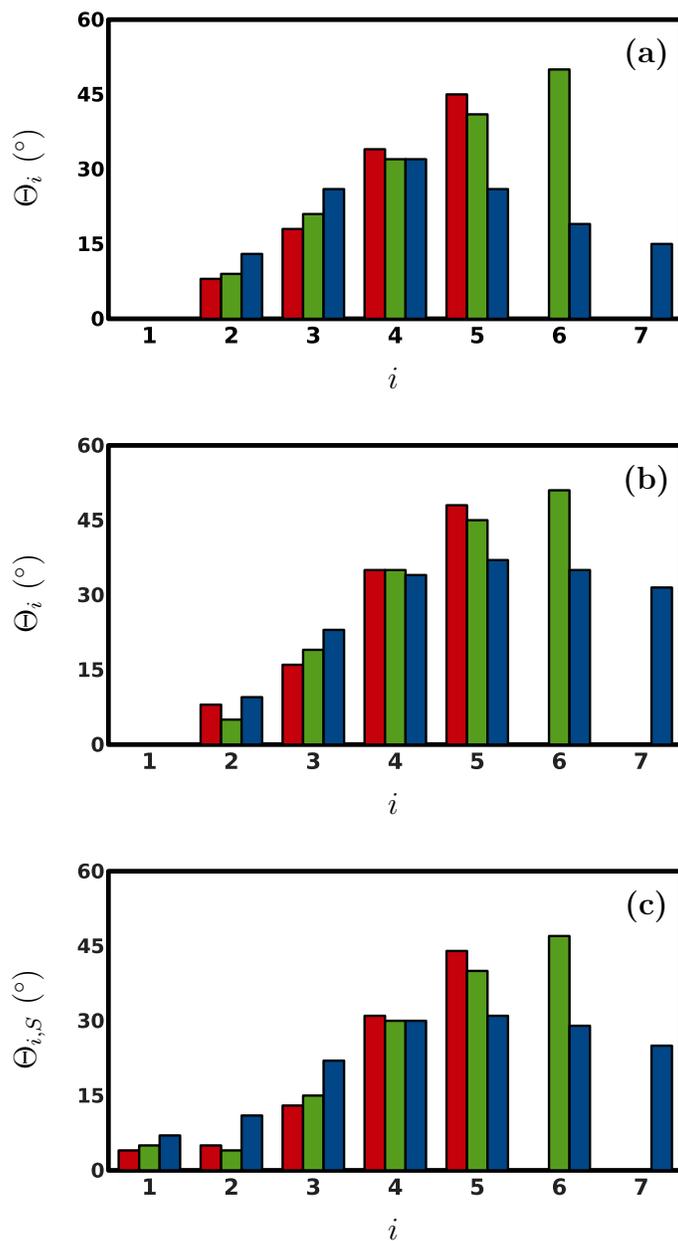


Figure 4.6: Molecular twist profiles for helicenes: [5]H – red, [6]H – green, [7]H – blue. Ortho-fused rings are numbered from 1–7 ( $i$ ), starting at one end of the molecule. On a substrate, numbering begins with the ring closest to the underlying surface. The solid angle  $\Theta_i$  between ring  $i$  and the reference plane is displayed. The reference plane for all panels is Ring 1, with the exception of panel (c), where the surface plane is the reference. Panel (a) shows data for the gas phase monomer, whereas (b) and (c) display the monomer adsorbed on Cu, with the Ring 1 and the surface reference planes, respectively.

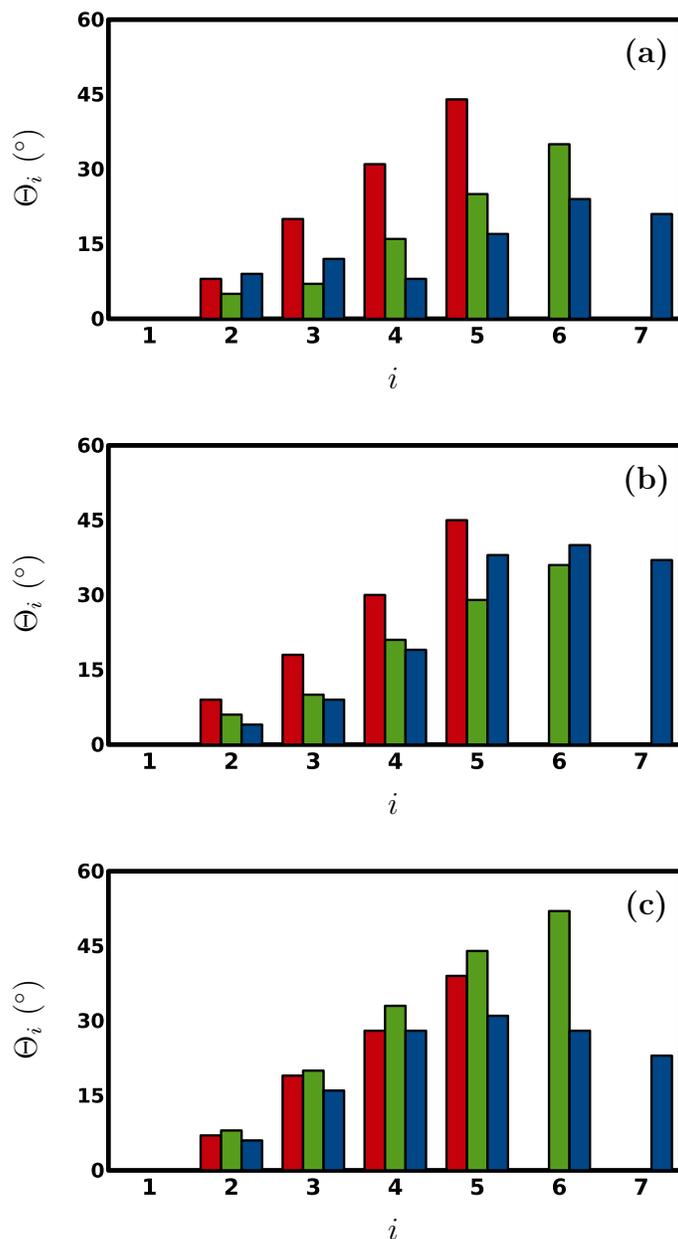


Figure 4.7: Molecular twist profiles for helicenes: [5]H – red, [6]H – green, [7]H – blue. Ortho-fused rings are numbered from 1–7 ( $i$ ), starting at one end of the molecule. Numbering begins with the ring closest to the underlying surface. The solid angle  $\Theta_i$  between ring  $i$  and the reference plane is displayed. Panels (a) and (b) correspond to the surface-bound molecule for Cu-adsorbed like and unlike dimers, respectively. Panel (c) shows the surface-bound molecule for a like dimer adsorbed on Ag.

Figure 4.5(b) presents  $\mathbf{d}_l$  and  $\mathbf{d}_h$  for [6]H.

### 4.3.3 Computational Details

Helicenes are relatively rigid large conjugated systems, with intermolecular interactions dominated by relatively strong dispersive forces. Both of these factors require some caution in approaching an appropriate computational methodology. Molecules with folded structures, characterized by intramolecular dispersion interactions are poorly reproduced by ab initio calculations such as MP2 because of the basis set superposition error.<sup>197</sup> This error may affect energy differences between conformers, intramolecular stability and even equilibrium structures. Dispersive corrections are commonly used to treat multimolecular systems, and their evaluation has been significantly improving in recent years. For large helicenes, the pitch of the helix was found to vary significantly depending on the method used for structure optimization: MP2 predicted a small pitch because of overstabilization (3.91 Å), whereas uncorrected DFT predicted the pitch to be too large, as a consequence of the lack of stabilization (4.93 Å). However, DFT-D methods predict a value in good agreement with experiment (4.64 Å vs 4.50 Å from crystallographic data).<sup>197</sup>

#### Adsorbed systems

Three important effects have to be considered in quantum calculations of adsorbed systems:

(i) The increased roughness of the potential energy surface due to the surface effects. Surfaces, in particular metallic ones, interact strongly with lone-pair-containing heteroatoms in organic molecules. This results in a particular "stickiness" of the surface, which, combined with the limited ability of quantum optimizations to explore configurational space, often leads to equilibration of the system in a local minimum, or a metastable state. This is not easily overcome in standard quantum calculations, which have no thermal fluctuations to aid in overcoming configurational barriers. To compensate, we normally sample more broadly the configurational space with the aid of classical calculations (either using Monte Carlo or molecular dynamics techniques), at finite temperatures, to generate a larger set of initial configurations for further analysis through quantum calculations. The approach usually provides significantly improved sampling, and lower energy structures, although it suffers from the fact that the classical potential energy surface is very different from the quantum one.

(ii) Basis set superposition errors (BSSE) may be important for physisorbed sys-

tems, and therefore they have to be estimated. We generally follow the standard counterpoise formalism,<sup>199;200</sup> to estimate BSSE corrections. In our experience, a hallmark of a well-converged pseudo-atomic orbital basis set cutoff (point (iii) below) is a small BSSE, but we perform these corrections regardless.

(iii) In pseudo-atomic orbital (PAO) codes such as SIESTA, one can essentially increase the size of the basis set by extending the confinement radius of the PAO's. We evaluated changes in binding energies for some of our chemisorbed systems with increasing confinement radius (implemented by decreasing the PAO confinement energy from its default value of 20 mRy down to 0.5 mRy).<sup>199;201</sup> We found that for many metal surface - organic adsorbate systems, a reliable confinement energy is 1 mRy, corresponding to a cutoff radius of the  $\zeta_1$  pseudo-orbitals for the carbon atom, of roughly 5.8 Å.

A 20-fold decrease of confinement energies is often regarded as an unaffordable precision in solid-state calculations. However, we found that this is not necessarily the case. A lower confinement energy cutoff leads to a more stable calculation throughout the SCF cycles and the geometry optimization steps. This often results into fewer convergence steps, in either or both cycles, and overall faster optimizations. The PAO cutoff handle is unfortunately often overlooked in solid state calculations, with serious consequences over the reliability of binding energy results.

## Optimization details

The optimized geometry of helicenes on metal surfaces was obtained with the DFT-vdW-TS method with Tkatchenko-Scheffler parameters.<sup>45;135</sup> Simulations were performed with the SIESTA<sup>202</sup> software, which uses pseudoatomic (PAO) orbitals.

Aromatic molecules adsorbed on metallic substrates are usually driven by non-covalent binding.  $\pi$ -stacking requires valence electrons to respond to the local chemical environment. Additionally, the effect on molecular electronic structure by the presence of metal's s and d bands cannot be neglected in systems with this degree of heterogeneity. For this reason we chose to represent valence electrons with standard GGA basis sets, double-zeta plus polarization.

Another concern for non-covalently bound systems modelled with pseudoatomic orbital (PAO) codes such as SIESTA is the basis set superposition error (BSSE), which is known to overestimate binding energies for physisorbed systems.<sup>197</sup> This error significantly affects the energy differences between conformers, intramolecular

stability and even equilibrium structures. A standard counterpoise correction was used here for BSSE,<sup>203</sup> in addition to an effective basis set extension through large cutoff radii.<sup>199</sup> Increasing the size of the basis sets through the extension of the confinement radius of the PAOs as described in Ref.<sup>199</sup>, has led to a roughly ten-fold reduction of the calculated BSSE, to  $\approx 10\%$  of the binding energy.

## Energy

Binding energies are reported below as the energy change corresponding to the adsorption reaction. As a result, bound systems have a negative binding energy. Several types of binding energies are reported, corresponding to different parts of the relevant binary or ternary complexes. They are defined in Table 4.1 based on the binding process they characterize:

Table 4.1: Binding energy nomenclature.

Process	Energy *	Comment
$[n]\text{H} + [n]\text{H} \rightarrow [n]\text{H}-[n]\text{H}$	$E_L, E_U$	Gas phase dimer
$[n]\text{H} + S \rightarrow [n]\text{H}-S$	$E^S$	Adsorbed monomer
$[n]\text{H}-[n]\text{H} + S \rightarrow [n]\text{H}-[n]\text{H}-S$	$E_L^S, E_U^S$	Adsorbed dimer
$[n]\text{H} + [n]\text{H}-S \rightarrow [n]\text{H}-[n]\text{H}-S$	$E_L^{H-S}, E_U^{H-S}$	Monomer-adsorbed mon.

\*Binding processes involving dimers can refer to homochiral ( $L$ ) or heterochiral ( $U$ ) dimers. Where molecules are achiral, the  $L$  subscript is used in the binding energy symbol.

## 4.4 Coronene

[6]C ( $\text{C}_{24}\text{H}_{12}$ , coronene) is a highly symmetric PAH constructed from seven perifused benzene rings (see Figure 4.3(a)). The molecule, as well as other large planar PAHs, serves often as a model for graphene sheets, which form building blocks for functionalized nanomaterials,<sup>204–206</sup> and has thus been studied extensively.

### 4.4.1 Planar Reference

[6]C also provides a good framework reference for the study of helicene adsorption on surfaces. The coverage area of [6]C is similar in size and shape to that of a helicene with a complete twist cycle in stacked orientation. [6]C has a high level of symmetry and is planar. Geometric deviations from [6]C behaviour in helicenes are thus due to their helical nature, lack of planarity and the associated steric effects.

The various structures of [6]C dimers have been well investigated in other works.<sup>207–209</sup> Stacked configurations were common stationary points on the potential energy surface (PES), preserving the molecular sixfold symmetry and keeping the molecular planes parallel. These include sandwich, twisted sandwich, parallel-displaced and staggered dimers. The staggered conformation offers the most stability.

#### 4.4.2 Monomers and Dimers

Gas-phase calculations in this work are consistent with these earlier results, indicating that a staggered dimer configuration is more stable than an eclipsed configuration (-1.13 eV vs -0.85 eV). The interplanar distance between the two molecules is 3.13 Å, in line with other theoretical results for PAHs,<sup>210;211</sup> though somewhat lower than the bulk graphite experimental and theoretical value of 3.35 Å,<sup>210;212</sup> and values for parallel-displace benzene, of 3.4–3.6 Å.<sup>213</sup> [see Figure 4.4 (a) and (b)]. Surface adsorption enhances binding between the two molecules, and is accompanied by a decrease in value of the intermolecular distance (see Table 4.2).

Table 4.2: Binding energies and distances for [6]C systems.

<b>S*</b>	$E_L$	$d_L^\dagger$	$E^S$	$E_L^S$	$E_L^{HS}$	$d^S$	$d_L^S$	$d_L^{HS}$
<b>Cu</b>	-1.13	3.1	-2.13	-2.33	-1.43	2.7	2.4	3.0
<b>Ag</b>			-1.99	-2.12	-1.31	2.8	2.7	3.0
<b>Au</b>			-1.63	-1.77	-1.31	3.1	3.0	3.1

\*Binding energy values,  $E$ , are presented in eV. Distances,  $d$ , are given in Å.  $^\dagger d$  values are distances between planes in the complexes, with indices representing the same type of interactions as described in Table 4.1

#### Binding on coinage metals

Table 4.2 provides binding energies and distances for gas phase and adsorbed [6]C, at the coinage metal substrates Cu, Ag and Au. The strongest binding substrate is Cu, and the weakest is Au, in agreement with other results of benzene and PAH binding on these surfaces.<sup>45;47</sup> Figure 4.8 presents some structures of adsorbed [6]C monomers. The adsorbate commensurates at the level of individual rings with the Cu(111) substrate (see Figure 4.8(a)). This behaviour explains the stronger affinity of [6]C for this substrate, and is the basis for using Cu(111) as a substrate for graphene CVD growth.<sup>214</sup> The other substrates considered here do not commensurate with the deposited [6]C (see Figure 4.8(b), for example).

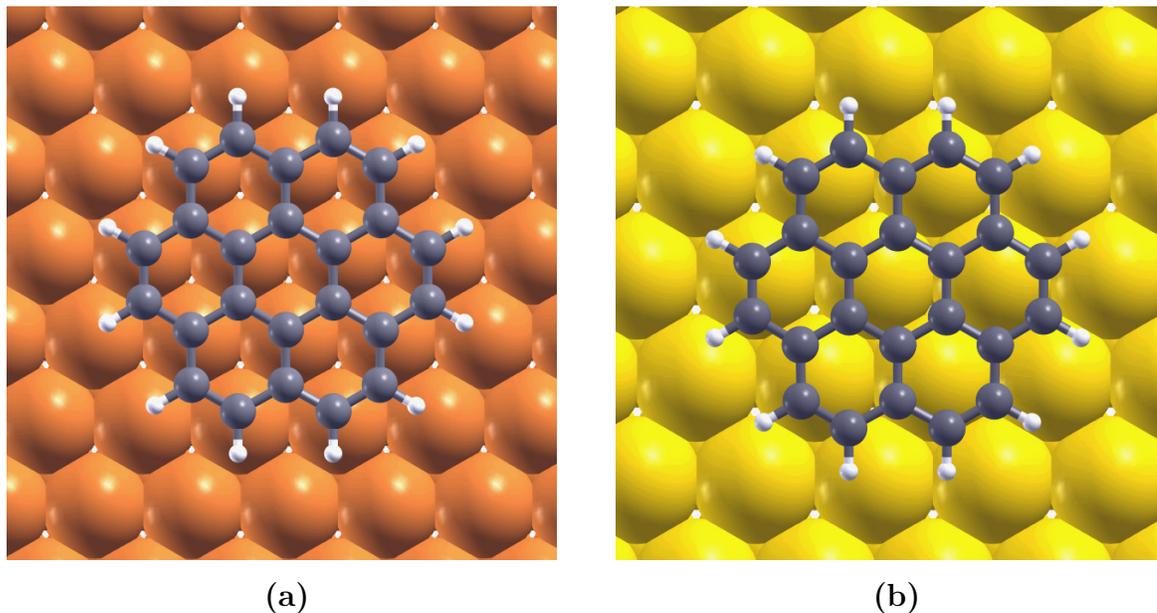


Figure 4.8: [6]C on (111) coinage metal surfaces. Panels (a) and (b) show the [6]C monomer on Cu and Au, respectively. Panels (c) and (d) present dimer geometries on Cu and Au, respectively.

### Projected density of states

The weak impact of the substrate on monomer and dimer geometries suggests a flat PES for [6]C on highly symmetric surfaces. This is supported by comparisons of the projected density of states (PDOS) for [6]C and the surface, in the free and adsorbed states [see Figure 4.10]. Upon adsorption, the interaction between the  $2p$  states of [6]C and the  $sp$  band of the metal is slightly stronger for Cu than for Ag, as the coronene states are more significantly broadened. No interaction between the adsorbate and the metal  $d$  band is observed. Antibonding  $2p$  states of the [6]C contribute to much of the interaction, as observed previously for ethene and other  $\pi$ -conjugated compounds,<sup>172;215</sup>. This is reflected in a slight elongation of the coronene double bonds upon adsorption.

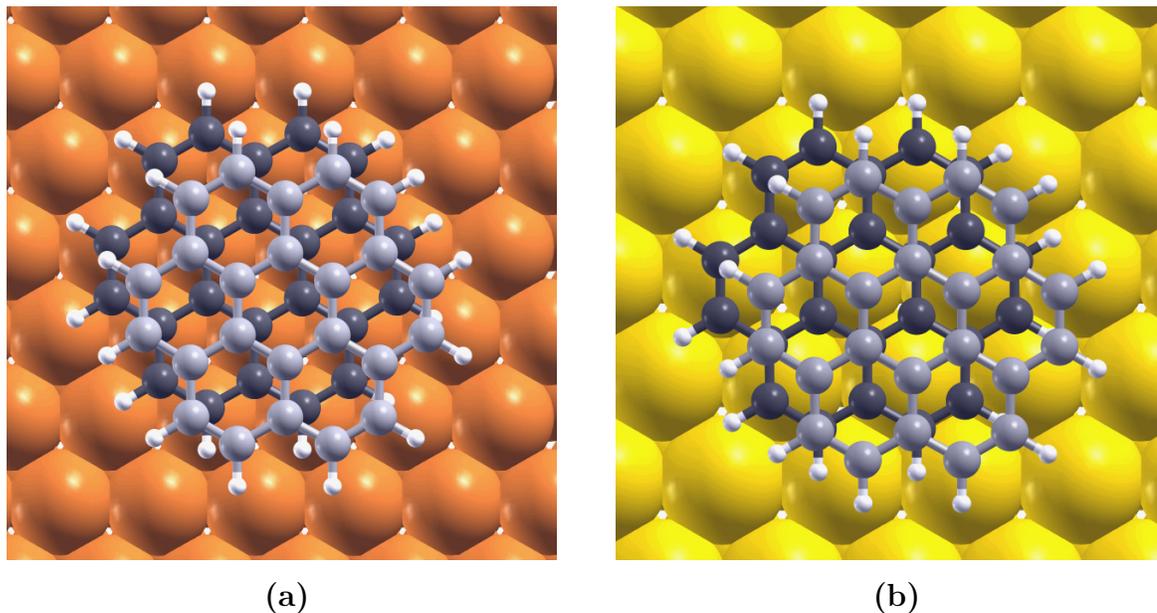


Figure 4.9: [6]C dimer on (111) coinage metal surfaces. Panels (a) and (b) present the geometries on Cu and Au, respectively.

### 4.4.3 Other Planar Aromatic Compounds

#### Benzene

This behaviour is similar to that previously observed in the adsorption of benzene monomers adsorbed on metal substrates:<sup>45</sup> Benzene was found to bind horizontally on the substrate, with the four distinct high-symmetry binding sites on the (111) surface (atop, bridge, face-centered cubic, hexagonal close-packed), presenting only minimal binding energy differences. The absence of deep wells on the PES allows benzene molecules to diffuse freely over the surface.

#### PAHs

Aromatic compounds larger than benzene have also been found to adopt surface-parallel geometries, such as in the case of diindenoperylene.<sup>46</sup> On Cu(111) however, larger, elongated molecules typically display a bending of the aromatic plane along its longest axis, with the center of the molecule closest to the surface, characteristic of chemical bonding. On Ag(111) and Au(111), van der Waals interactions dominate<sup>47</sup> and the molecule retains the planar structure, similar to the case of benzene.

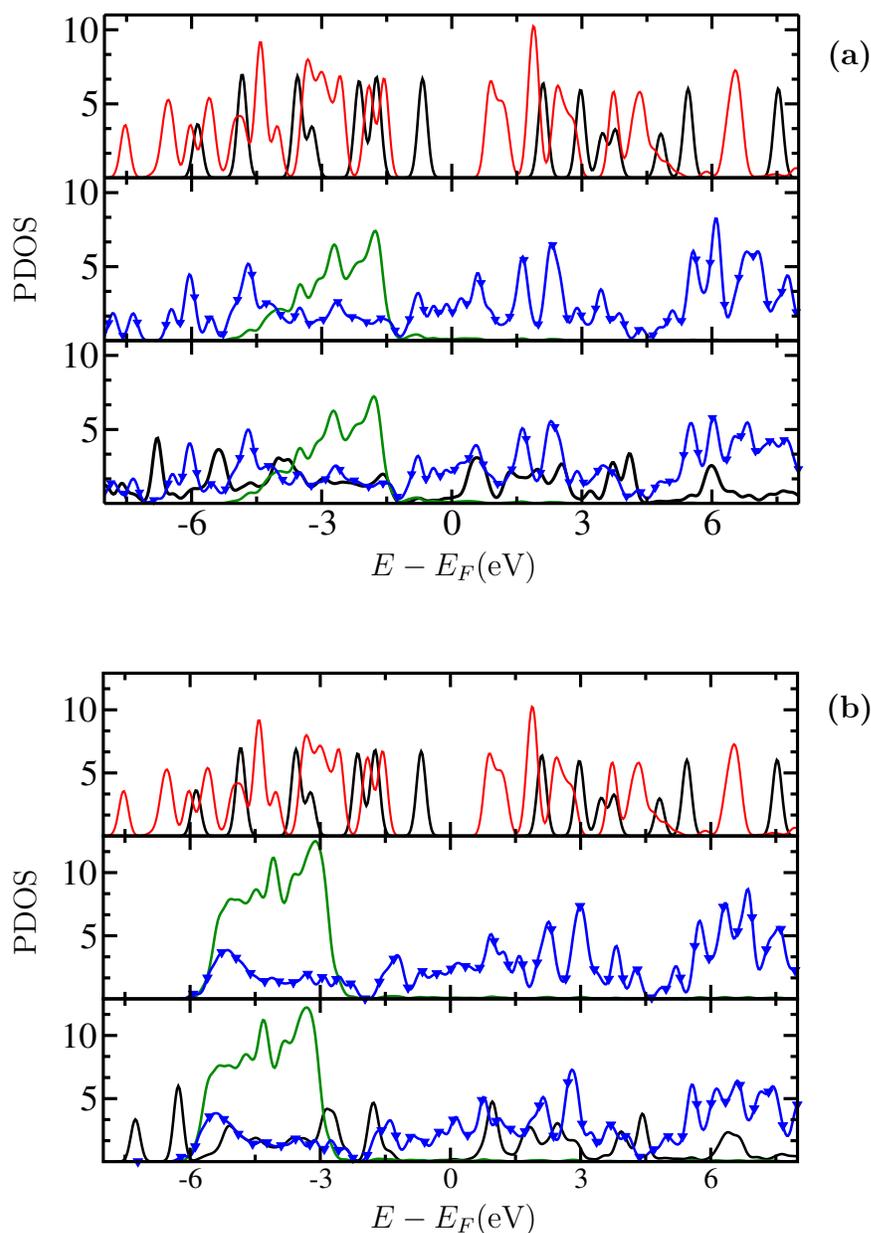


Figure 4.10: The behaviour of the projected density of states (PDOS) upon adsorption of coronene. Figures (a) and (b) correspond to the [6]C–Cu and [6]C–Ag adsorption, respectively. In top panels, gas phase  $2p_z$  states are presented for the monomer (black lines) and the dimer (red lines). The  $d$  (green lines) and  $sp$  (blue triangles) bands of the metal surfaces before adsorption are shown in the middle panels. Bottom panels include data for the adsorbed monomer complex: black lines are the  $2p_z$  states of the adsorbate, green lines are the  $d$  bands of the metals, and blue triangles show the  $sp$  bands of the metal.

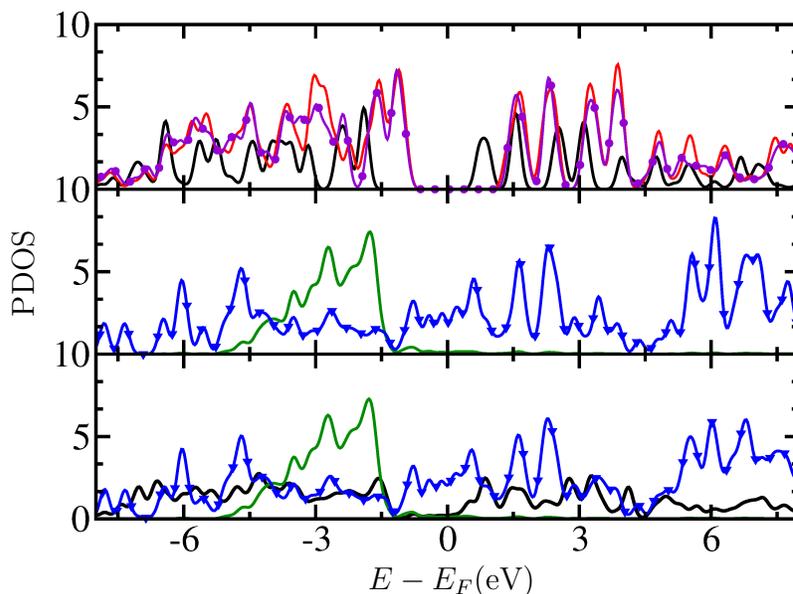


Figure 4.11: The behaviour of the projected density of states (PDOS) upon adsorption of [6]H on Cu. Top panel: gas phase  $2p_z$  states for the monomer (black lines) and the dimer (red lines), with states for the L dimer (red) and the U dimer (purple circles) given for the chiral [6]H molecules. Middle panel: the  $d$  (green lines) and  $sp$  (blue triangles) bands of the metal surfaces before adsorption. Bottom panel: data for the adsorbed monomer complex; black lines are the  $2p_z$  states of the adsorbate, green lines are the  $d$  bands of the metals, and blue triangles show the  $sp$  bands of the metal.

## 4.5 Helicene Dimers in Gas Phase

### 4.5.1 Steric Effects

#### Optimal stacking

In the absence of a surface, helicene dimers, like coronene dimers, interact mainly via dispersive forces, which are optimized in the most stable configurations. The  $\pi$  stacking interaction among rings acts as an aligning force among planes, both intra and intermolecularly. Additionally, each pair of rings has a distance for which the interaction is optimal, with deviations from this distance leading to less attraction. If this pair distance is maintained throughout the entire molecule, the dimer is optimally stacked. Both conditions are easily achieved by the coronene dimer. With no helical component acting as a steric disruptor, coronene dimer can adopt a purely parallel configuration between the planes of the two molecules, which in turn allows for optimal interplanar distance among all rings.

### Helix-induced deviations

Unlike for coronene, the out-of-plane twist of the molecular backbone in helicenes interferes with the usual tendency for stacked configurations in dimers of PAH molecules. The ortho-fused rings backbone leads to larger angles between some intermolecular planes and, inevitably, to some intra and intermolecular ring overlap, with the associated deviation from the optimal interaction distance. As a result the dimer energy of helicenes is less than that of coronene.

### Chirality and stacking

The steric effects will have distinct manifestations in homochiral (L) and heterochiral (U) dimers, and are therefore a direct consequence of molecular chirality. As shown in Figure 4.4, both L and U dimers can achieve stacked-staggered configurations in gas phase. In general, more rings can achieve optimal alignment in the L dimer, therefore achieving stronger intermolecular binding (see Table 4.3). Two homochiral helices can in principle be oriented such that each pair of equivalent rings is parallel, albeit with larger distances among equivalent rings than for the planar case.

In other words, like dimers can be thought to decouple the destabilizing effects of molecular non-planarity from those of equivalent ring orientation. Like dimers can retain the optimal helix angle of both molecules – similar to how coronene retains planarity – as well as the parallel ring orientation among equivalent rings. This is impossible for unlike dimers of non-planar molecules. The significantly greater angles among equivalent rings within unlike dimers are the result of a greater compromise between mismatching helices. Regardless of the ability of like helices to orient parallel, the planar distance between equivalent rings is on average larger than for coronene. Because of this, the most stable helicene dimers are still less efficiently bound than [6]C[6]C, by about 0.2 eV.

Table 4.3: Binding energies in gas-phase helicene dimers.

<b>mol</b>	<b>[1]H</b>	<b>[2]H</b>	<b>[3]H</b>	<b>[4]H</b>	<b>[5]H</b>	<b>[6]H</b>	<b>[7]H</b>
$E_L$	-0.16	-0.32	-0.54	-0.81	-0.91	-0.93	-0.90
$E_U$					-0.52	-0.86	-0.82

### 4.5.2 Size Effects

The out-of-plane twist becomes more pronounced for larger helicenes, impacting the ability of the dimer to efficiently stack. Thus, the binding energy of L-[7]H[7]H is similar to that of L-[5]H[5]H and L-[6]H[6]H, despite the larger size of the conjugated system in [7]H[7]H. The U dimer is bound even less efficiently, particularly for [5]H[5]H, where only two of the conjugated rings are able to stack. The longer [6]H and [7]H molecules have the ability to adapt their relative orientations in order to maximize the extent of intermolecular stacking, achieving heterochiral binding energies in the same range as those of homochiral dimers.

Table 4.4 displays the  $\phi_{i,i}$  for dimers of [4]H–[7]H. In general terms, L dimers can attempt to minimize these angles by parallel orientation of their helices. For L-[5]H[5]H and L-[6]H[6]H, all angles are less than  $20^\circ$ , reflecting well-stacked conformations. In L-[7]H[7]H, intramolecular stacking between the end rings inhibits intermolecular stacking by equivalent rings. Instead, the molecules are shifted so that one monomer's helix roughly continues the other's, so that the dimer geometry resembles that of a single [15]H molecule. In this case, intermolecular stacking is no longer correlated to the angles between equivalent rings (see Table 4.4).

Table 4.4: Angles between equivalent rings in gas-phase helicene dimers.

$i^*$	[4]H[4]H	[5]H[5]H		[6]H[6]H		[7]H[7]H	
	$\phi(i, i)$	$\phi_L(i, i)$	$\phi_U(i, i)$	$\phi_L(i, i)$	$\phi_U(i, i)$	$\phi_L(i, i)$	$\phi_U(i, i)$
1	9	4	33	6	61	28	20
2	6	12	19	8	42	23	23
3	3	14	7	10	22	24	26
4	7	15	25	9	19	29	10
5	–	13	44	12	21	27	19
6	–	–	–	12	20	26	31
7	–	–	–	–	–	31	25

\*Rings are considered to be equivalent if they belong to different molecules and have the same number, as defined in Figure 4.5. Values are presented in degrees.

The unlike dimers of these helicenes are unable to approach each other without a severe distortion of parallel stacking between some equivalent rings. They will adopt markedly different strategies of interaction, with partial stacking, leading to non-monotonous energetic effects with size, and large equivalent-ring angles for all molecules and rings (see Table 4.4).

## 4.6 Adsorption of Helicene Monomers

### 4.6.1 General Trends

The presence of a substrate distorts the molecular structure of an adsorbed monomer. Strong dispersive interactions characterize the adsorption of large conjugated molecules on metallic substrate (see for example Tables 4.2 and 4.5 and Refs.<sup>70;194</sup>). In helicenes, these interactions will affect all of the conjugated rings, but change subtly along the molecular backbone, due to the molecular helicity. Overall, surface binding in helicenes is weaker than in the planar, more efficiently-binding coronene, though binding trends in the coinage metal series (Cu > Ag > Au) are generally preserved.

Table 4.5: Binding energies in surface-adsorbed helicene monomers and dimers.

mol, $S^*$	$E^S$	$E_L^S$	$E_U^S$	$E_L^{H^S}$	$E_U^{H^S}$
[1]H, Cu	-0.82	-1.00			
[1]H, Ag	-0.66	-0.77			
[1]H, Au	-0.51	-0.59			
[2]H, Cu	-1.20	-1.37			
[2]H, Ag	-0.99	-1.10			
[2]H, Au	-0.81	-0.95			
[3]H, Cu	-1.60	-1.67			
[3]H, Ag	-1.35	-1.50			
[3]H, Au	-0.97	-1.09			
[4]H, Cu	-1.89	-2.34			
[4]H, Ag	-1.56	-1.78			
[4]H, Au	-1.30	-1.33			
[5]H, Cu	-1.58	-2.04	-1.94	-1.40	-1.01
[5]H, Ag	-1.50	-1.85	-1.89	-1.21	-0.96
[5]H, Au	-1.12	-1.46	-1.56	-1.24	-1.09
[6]H, Cu	-1.87	-2.11	-1.93	-1.40	-1.11
[6]H, Ag	-1.70	-1.92	-1.77	-1.02	-0.78
[6]H, Au	-1.41	-1.59	-1.41	-1.09	-0.72
[7]H, Cu	-1.68	-1.85	-1.84	-1.44	-1.03
[7]H, Ag	-1.77	-1.91	-1.78	-0.80	-0.70
[7]H, Au	-1.30	-1.48	-1.40	-1.12	-0.88

\* $S$  refers to the surface. Binding energy values are given in eV. Energy symbols defined in Table 4.1

## 4.6.2 Electronic Structure

Figure 4.12 displays optimized structures of [5]H, [6]H and [7]H adsorbed to the Cu substrate, and Figure 4.11 presents an example of electronic structure implications of adsorption - the modification of the PDOS upon adsorption of [6]H on Cu. Calculations show that surface adsorption affects the molecular twist, particularly for Cu adsorption - as in the case of [6]C binding, this is a reflection of the implication of virtual antibonding 2p states of the adsorbate in the intermolecular interaction [see Figure 4.11].

## 4.6.3 Steric Effects

The same structural description for the gas phase dimers can be used in the case of surface-adsorbed monomers, if the angle and distance of rings with the surface plane is considered instead of equivalent rings. Unlike the case for equivalent rings, however, the surface offers a fixed reference plane for all rings, and thus will have a general planar-inducing effect on the adsorbed molecule, regardless of its nature.

In terms of molecule-surface separations, in all cases Ring 1 is found closest to the surface, at a distance within 0.1 Å of the average [6]C-substrate distance on the corresponding metal. Subsequent helicene rings are found further and further away from the surface due to the molecular helicity.

### Head-group surface adsorption

Figures 4.6 (b) and (c) display  $\Theta_i$  and  $\Theta_{i,S}$ , respectively, for [5]H, [6]H and [7]H adsorbed on Cu. In all molecules, the first three rings make a low angle with the surface plane (roughly less than 20°). Structural comparisons between gas phase monomer structure and surface-adsorbed structure indicate surface-induced distortion of the molecule to enhance this relative planarity [See Figure 4.12(g-i)]. The values of average solid angles calculated relative to the surface are generally lower than solid angles calculated relative to Ring 1 (except of course for Ring 1). This is an indication that molecules tilt slightly when adsorbed, to maximize contact between Rings 2-4 and the surface, at the expense of the efficiency of the Ring 1-surface interaction.

These effects are consistent with previously reported experimental and theoretical results of surface-bound helicenes. For example X-ray photoelectron diffraction of [7]H monolayers on Cu(111) found that molecules bind to the surface via the termi-

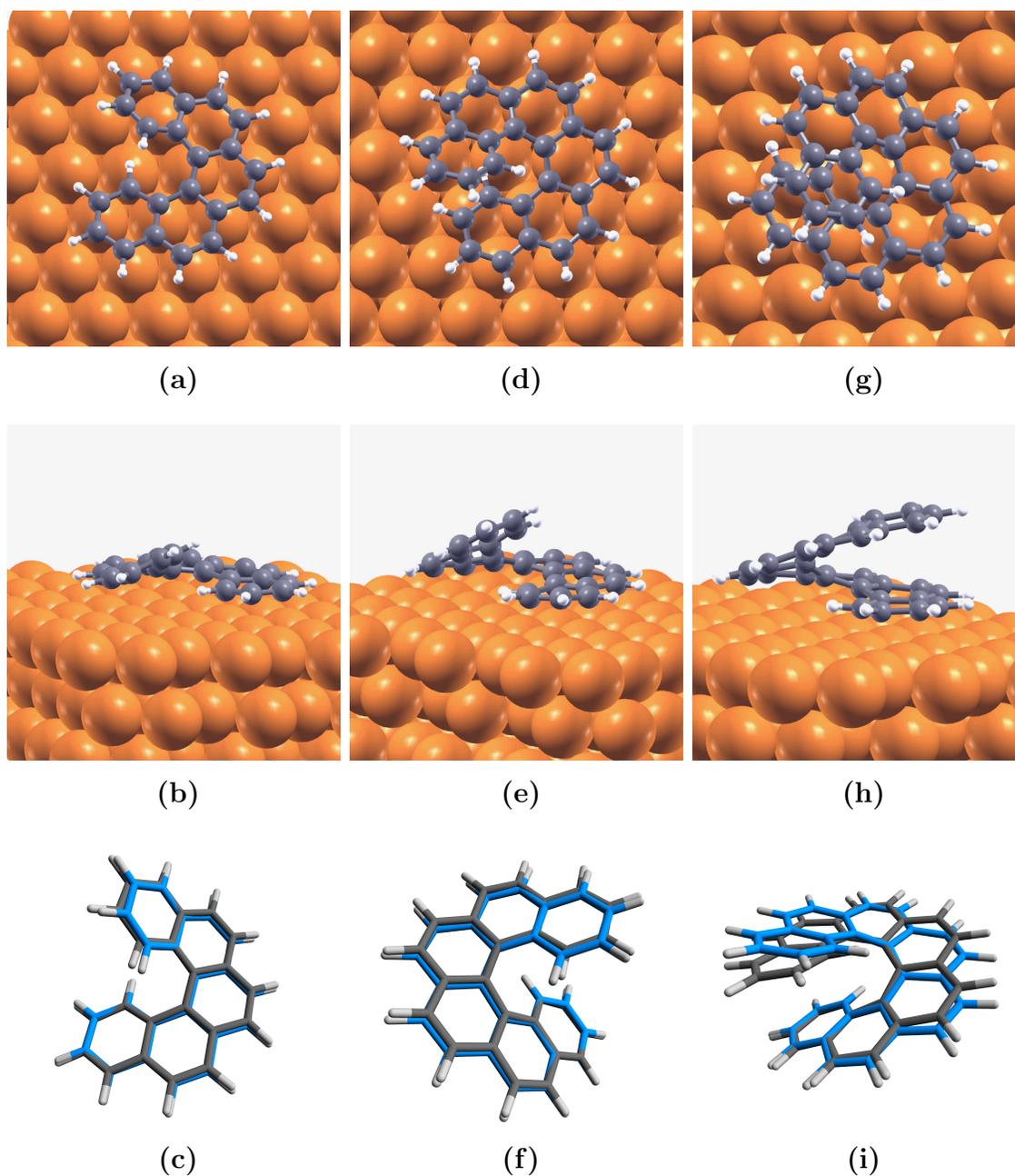


Figure 4.12: Helicene monomers on the Cu(111) surface. [5]H is shown in panels (a)–(c), [6]H is shown in panels (d)–(f) and [7]H is shown in panels (g)–(i). Both top [panels (a), (d) and (g)] and side [panels (b), (e) and (h)] views are presented. The bottom row [panels (c), (f) and (i)] shows the superimposed structures of the respective molecule in the absence (blue) and the presence (gray) of the surface.

nal phenanthrene group (rings 1-3), oriented parallel to the surface plane.<sup>76</sup> Similar theoretical adsorbed geometries of [5]H on Cu(111) were reported by Mairena et al.

from calculations employing plane-wave basis sets with a optB86-vdW functional.<sup>83</sup>

#### 4.6.4 Size Effects

The molecular size also impacts surface binding energies, as the increase in the size of the conjugated system enhances its dispersive interaction with the underlying surface. The dependence of binding energy on molecular length is thus monotonous in planar molecules [1]H–[4]H, with a striking departure from this trend accompanying the departure from molecular planarity in [5]H–[7]H.

##### Tail-group effects

A sharper departure from planarity starts generally at ring 4. The tail end of the molecule, starting with the fourth ring onward, spirals away from the surface, manifesting the helical property of the molecule. The enhanced planarity of the three rings proximal to the surface, and the associated optimization of molecule-surface interaction, comes at the cost of intramolecular interactions in [7]H: the end rings in the surface adsorbed molecule are no longer parallel [see Figure 4.6(a)-(b)], indicating a competition between the molecule-surface and intramolecular interactions in this case.

The frustration is directly manifested in binding energy trends (Table 4.5): the adsorption energy of [7]H is weaker than that of [6]H in general, and a reversal of the established trends between Cu and Ag surfaces is seen here.

#### 4.6.5 Adsorption Energies

The adsorption energies reported in Table 4.5 have the meaning of the energy associated with the reaction between the free molecule and the free surface, producing the complex between the two. There are two major contributors to this reaction energy: the attraction between the molecule and the surface, an overall stabilizing effect, and the deformation of the molecule (and to a lesser extent of the surface) in the product, an energetic cost with an overall destabilizing effect. Both of these contributors to the reaction energy are stronger on Cu than on Ag, resulting in a weaker adsorption energy for [7]H on Cu than on Ag, and also for [7]H on Cu than for [6]H on Cu.

## 4.7 Helicene Dimers on Surface

### 4.7.1 Dimer Stability

#### Stacking effects

The deposition of a second molecule atop an already adsorbed molecule changes the surface binding of the latter. For [6]C, the surface attraction of the remote molecule pushes the proximal molecule closer to the surface, particularly when the substrate is Cu. The overall adsorption energy of the dimer is slightly stronger than that of the monomer, but not equivalent to 2 monomer adsorption energies. This is due to:

- (i) the larger distance between the second molecule and the surface,
- (ii) the approach of the surface-proximal molecule beyond its optimum interaction distance, and
- (iii) a smaller intermolecular separation of the [6]C molecules in their adsorbed state than in gas phase(See Table 4.2).

These effects are weaker on Ag and Au, where the majority of the dimer-surface interaction can be attributed to the adsorption of the surface-proximal molecule.

For helicenes, the effects of an impinging second molecule are moderated by the helicity of the molecular structure. Thus, adsorbed dimer interactions depend on relative chirality and on the dimer's ability to stack, as pointed out for the gas-phase dimers discussed in Section 4.5. On a surface however, stacking is influenced by the surface-induced changes in molecular structure, particularly in the proximal molecule.

#### Chirality

As shown in Table 4.5, laying a homochiral second molecule atop an adsorbed monomer generally produces a more stable complex, with the binding between a like dimer and the surface stronger by between 8 and 30% than the monomer adsorption energy. The formation of unlike dimers also has a stabilizing effect on the adsorbed complex. Adsorption energies of U dimers are between 1 and 23% stronger than monomer adsorption energies, and generally weaker than L dimer adsorption.

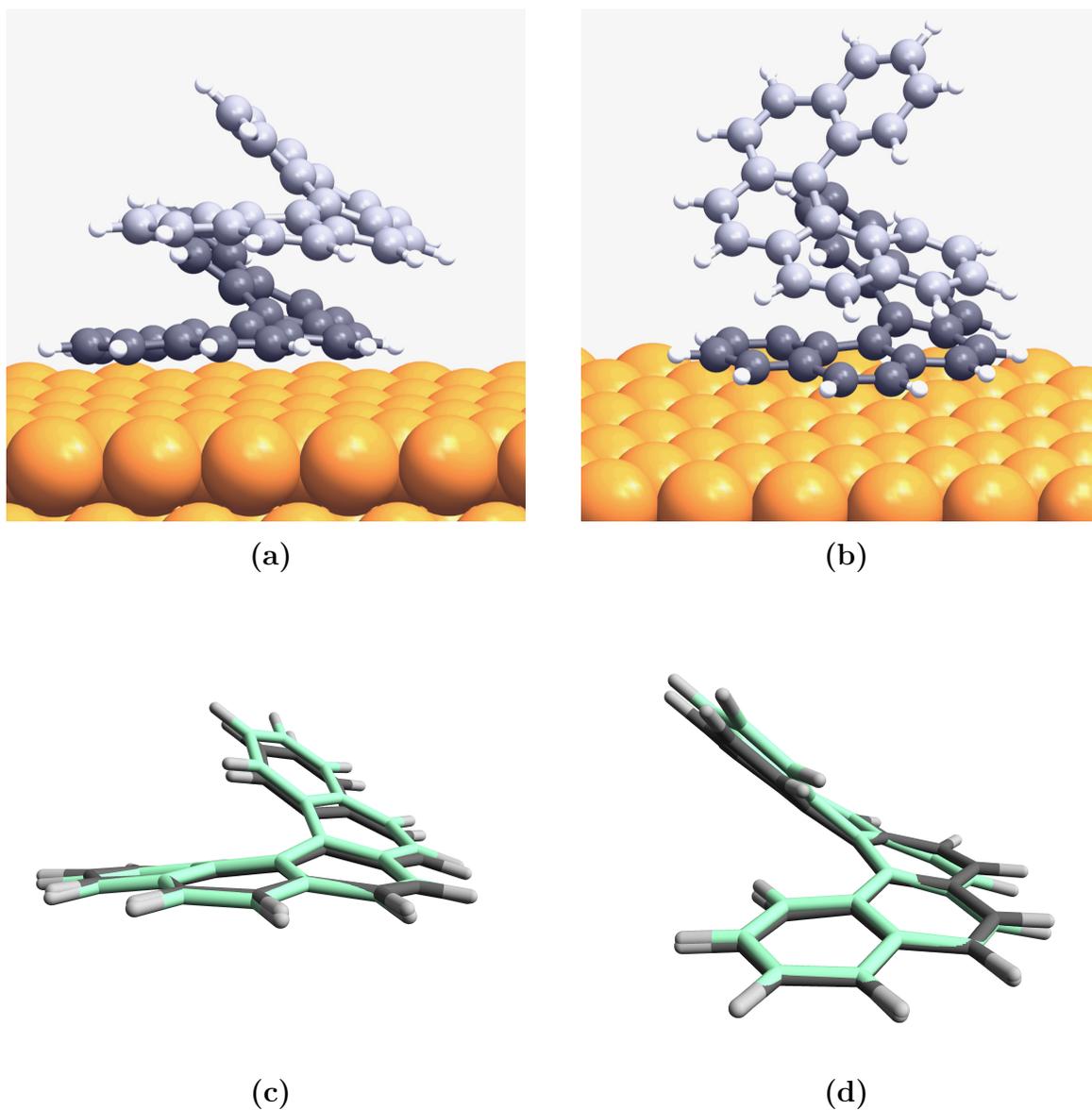


Figure 4.13: Structures of pentahelicene stacked dimers on Cu(111). Panels (a) and (b) show the L-[5]H[5]H and U-[5]H[5]H dimers, respectively. Panels (c) and (d) show the superimposed structures of the surface-adsorbed molecule in the absence (light green) and the presence (gray) of the stacked molecule, for the respective systems.

### Influence of surface-distal molecule

To examine whether surface adsorption has an effect on the chiral interactions of L and U dimers, the interaction strength between the approaching molecule and the adsorbed molecule-surface complex was estimated (See Table 4.5). This energy cor-

responds to the chiral interaction between the dimer molecules, but also includes a component of the dispersive interaction between the remote helicene and the substrate. Analysis of dimer energies in Tables 4.5 and 4.3 points to significantly enhanced on-surface dimer binding, and an enhanced preference for L dimer formation, particularly for [6]H and [7]H, which are non-discriminating in the gas phase. The effect is compounded by the surface attraction of the remote molecule, and also reflected in further structural changes in the proximal molecule (See Figures 4.13 and 4.14).

## 4.7.2 Geometry Changes in the Surface-Proximal Molecules

### The “sandwich” effect

The presence of the second molecule further alters the geometry of the adsorbed helicene. The surface-proximal molecule, sandwiched between the surface and its dimer partner, generally presents a lower profile in its dimer geometry than when adsorbed on its own. This effect is dependent on the effectiveness of the stacking interactions with the second helicene, and comprises the entire molecular backbone, not just its terminal rings. Panels (c) and (d) in Figure 4.13 illustrate changes in the molecular structure of the surface-proximal pentahelicene molecule that participates in a dimer, compared with the molecule when individually adsorbed. Similarly, panels (c) and (d) in Figure 4.14 illustrate the same effect for heptahelicene. As evidenced by the figures, and quantified by the solid angles presented in Figure 4.7, the “squashing” effect is stronger for L than U dimers, and again stronger on Cu than on Ag or Au.

### The influence of chirality

The stronger geometric effects in L dimer formation are a direct consequence of the relative efficiency of stacking in L and U dimers. L dimers interact by stacking their helices, throughout the molecular backbone. This results in the lifting of Rings 3-4 (sometimes including rings 2 and 5 as well), due to intermolecular dispersive attraction, and a flattening of the upper end rings, due to the need to accommodate the surface attraction of the second molecule. On the other hand, U dimers are only able to stack portions of their molecular backbone (see Figures 4.13, 4.14 and 4.4): those regions are affected by the intermolecular attraction, but not to the same extent as in the more efficiently stacked L dimers.

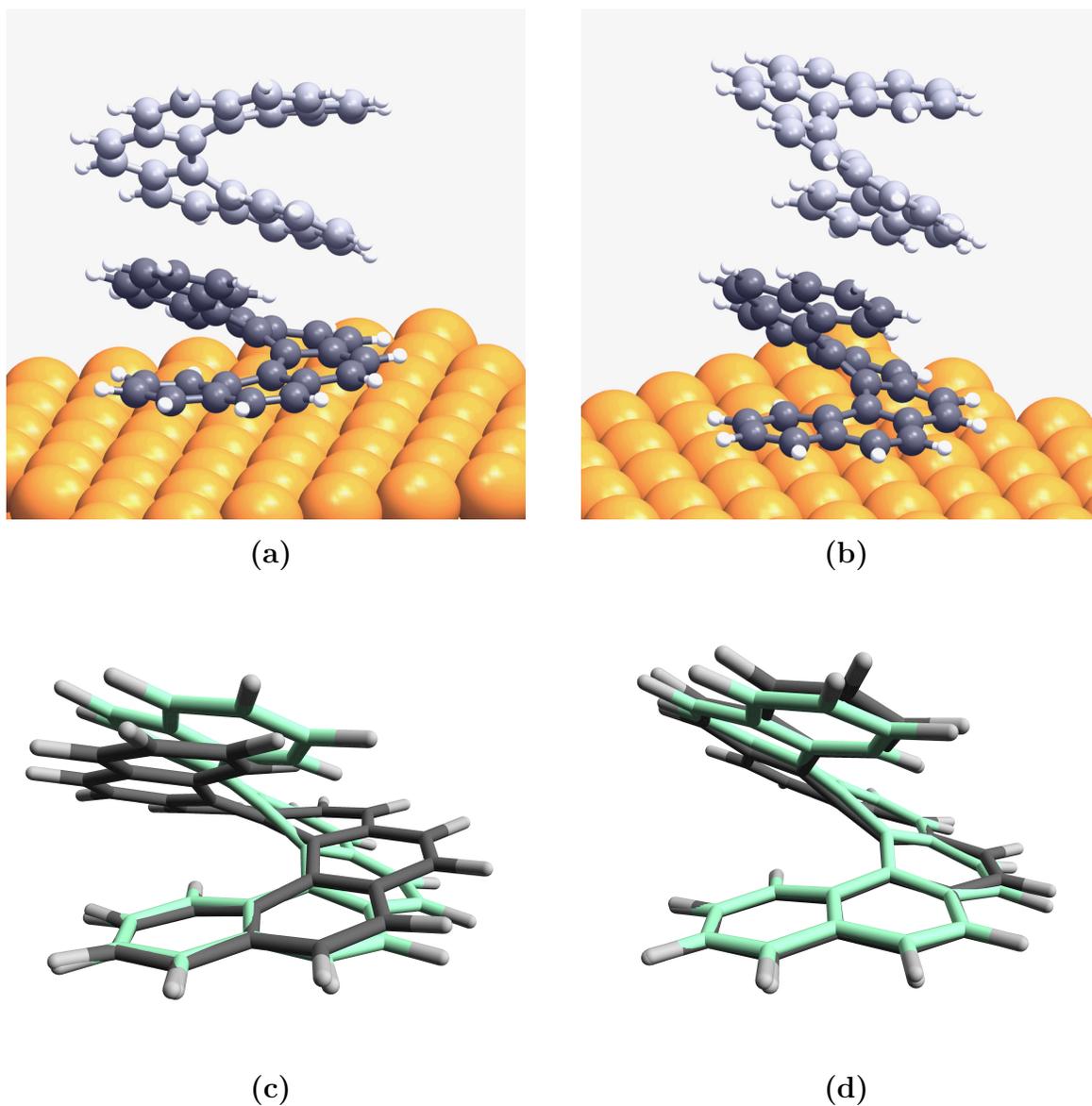


Figure 4.14: Structures of heptahelicene stacked dimers on Cu(111). Panels (a) and (b) show the L-[7]H[7]H and U-[7]H[7]H dimers, respectively. Panels (c) and (d) show the superimposed structures of the surface-adsorbed molecule in the absence (light green) and the presence (gray) of the stacked molecule, for the respective systems.

### Molecular footprint

One consequence of the “sandwich” effect on the surface-proximal helicene is a change in molecular footprint. A reduced molecular profile leads to the rings spreading out on the surface, distributing the strain throughout the conjugated backbone. The larger

footprint is somewhat visible in Figures 4.13(c)–(d), 4.14(c)–(d) and was quantified in the low and high diameters calculations, reported in Figure 4.15. While the effect appears minor on a single-molecule basis, minor footprint changes can be cumulative in high-coverage settings, as well as impact any commensuration effects such as those occurring on Cu substrates.

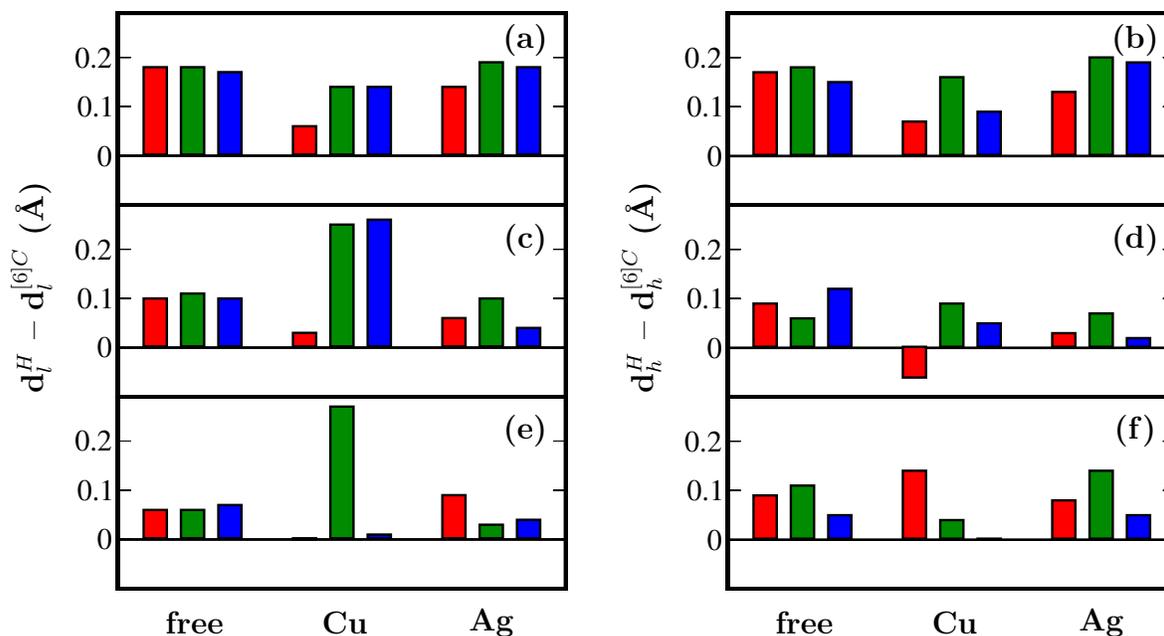


Figure 4.15: Low ( $d_l$ ) and high ( $d_h$ ) diameters as a description of molecular footprint, relative to the coronene diameter. Panels (a) and (b) show low and high diameters for [5]H, respectively. The values for [6]H are shown in panels (c) and (d) and those for [7]H are shown in panels (e) and (f), respectively. Red, green and blue bars are used in each plot to denote a molecule in monomer, like dimer and unlike dimer structures, respectively.

## 4.8 Conclusion

Helicenes adsorb on coinage metals via dispersive interactions between the conjugated system and the electrons in the metal's sp-band. For the adsorbate, the interaction induces a deformation of the helix: The phenantrenic head of the adsorbed helicene is flattened, while strong coiling within the rest of the molecule provides accommodation for the intramolecular steric effects.

In the systems considered in this chapter, surfaces did not generally facilitate

binding by way of opening of the molecular coil and increasing the molecular footprint. Footprint increases were however observed in the molecules closest to the substrate when the adsorption of stacked helicene dimers was considered. For the substrate considered here the surface-induced geometric asymmetry is proportional to the surface potential, which is strongest for Cu(111) and weakest for Au(111), with Ag(111) in between. Cu(111) bounds adsorbates strongly leading to significant molecular changes upon adsorption. Au(111) and Ag(111) are similar in their relatively smaller deformation effects on adsorbates. Thus when considering binding energies in conjunction with the associated geometric effects, the three substrates form two main groups, with Cu(111) representing the strongly-binding substrate, whereas both Ag(111) and Au(111) manifesting weakly-binding behaviour.

Within dimers, the relative handedness of molecules was a greater factor in dimer stability than the nature of the surface: like dimers were found to stack their entire backbones by either syncing or continuing their helices. Unlike dimers were only able to achieve stacking along a part of their molecular backbones, resulting in less effective interactions. The presence of the surface affected the stacking process through the flattening of the molecule proximal to the surface, and by exerting attractive forces on the furthest molecule. As a result, a “sandwich effect” was observed in the adsorbed dimers, in which the surface-proximal molecule was flattened and laterally expanded.

From the point of view of binding effectiveness within stacked adsorbed dimers, these simulations indicate that homochiral pairs are preferred to heterochiral ones. However, multibody interactions in monolayers and multilayers can change this preference, particularly since homochiral dimers exhibit larger footprints than adsorbed monomers. In fact, STM studies show that at coverages beyond a single monolayer, pentahelicene molecules prefer to form homochiral layers of alternating chirality upon deposition on metallic substrates.<sup>83</sup> The subsequent study of extended systems of helicene monolayers and bilayers, using classical Monte Carlo simulations, could prove useful for testing the suppositions made in this chapter, regarding the importance of footprint effects in these cases.

## Chapter 5

# Monte Carlo Simulations of Helicene Systems

### 5.1 Introduction

Theoretically investigating the self-assembly of helicenes on metal substrates over medium and long ranges requires simulations of models with correct physico-chemical properties over a large number of particles. In simulations of extended systems the computational cost per molecule must be kept low in order to access larger-size systems. In a simplified, yet effective model this trade-off does not come at the cost of accuracy in expressed interactions. Instead, the model successfully averages over the short-scale complexity, and reproduces the correct long-range behaviour.

In Chapter 4 we have established through DFT calculations that (i) adsorbed helicenes undergo distortions of the helix by flattening the phenantrenic head close to the surface and coiling the tail end of the molecule, a process enhanced by the presence of a stacked molecule and (ii) the relative handedness of helicenes plays a significant role in the adsorbed dimer stability, with “like” molecules stacking in a way that maximizes the  $\pi$ -interaction between inter-molecular rings. The limits on system size with DFT calculations prevented us from answering questions of long-range order or of changes in the local structure when molecules adsorb in the presence of other adsorbed molecules. We therefore don’t know how lateral interactions affect adsorption and stacking, and how the previous conclusions might be influenced by them.

In this study we use a series of simple, helical models to explore relationships

between the general topology of helicenes and their long-range adsorbed organization. Starting with the electronic structure data derived in our previous work, we apply a two-step particle-coarsening procedure and reduce the level of geometric detail to a coarse-grained model, designed to conserve the key structural properties of helicenes.

We perform parallel tempering Monte Carlo sampling of systems with 50 – 80 such molecules, in both enantiopure and racemic mixture, in a low density box, in the presence of a metallic surface.

We address questions of monolayer formation, stacking and enantioselectivity, as well as the tension between the accuracy of molecular representation and coarse-grained simplifications driven by computational limitations.

## Chapter layout

In Section 5.2 the methodology for this chapters' study is presented. This includes simulation details and details about the two-step coarse-graining procedure we used to transform the optimized structures from Chapter 4 into the coarse-grained models used in this chapter. Section 5.3 presents our results and the discussion around surface-driven and intermolecular interactions, multilayered structures and enantioselectivity. We close this chapter with concluding remarks in Section 5.4.

## 5.2 Methodology

### 5.2.1 Model $D$

As a first step to lowering the complexity of the helicene representation, we adopted model  $D$ , which is a standard united-atom treatment<sup>216;217</sup>. The input data used to parametrize model  $D$  was taken from our previous DFT calculations on adsorbed monomers and dimers. The united atoms in  $D$  are Lennard-Jones spheres, one for each carbon atom, with the hydrogen influence incorporated within all carbons equally. The size of  $D$  molecules is thus reduced to the size of their carbon backbone. Series  $D$  is represented graphically in Figure 5.1. Lennard-Jones parameters were fitted to reproduce the adsorption distances and energies from DFT calculations.

The carbon atoms contained in PAHs come in two distinct flavors: internal, bound only to other carbon atoms and external, bound covalently to a hydrogen. In model  $D$  we rendered all carbon atoms parametrically identical and evenly dispersed the influence of hydrogen atoms among all united atoms. The number of Lennard-Jones

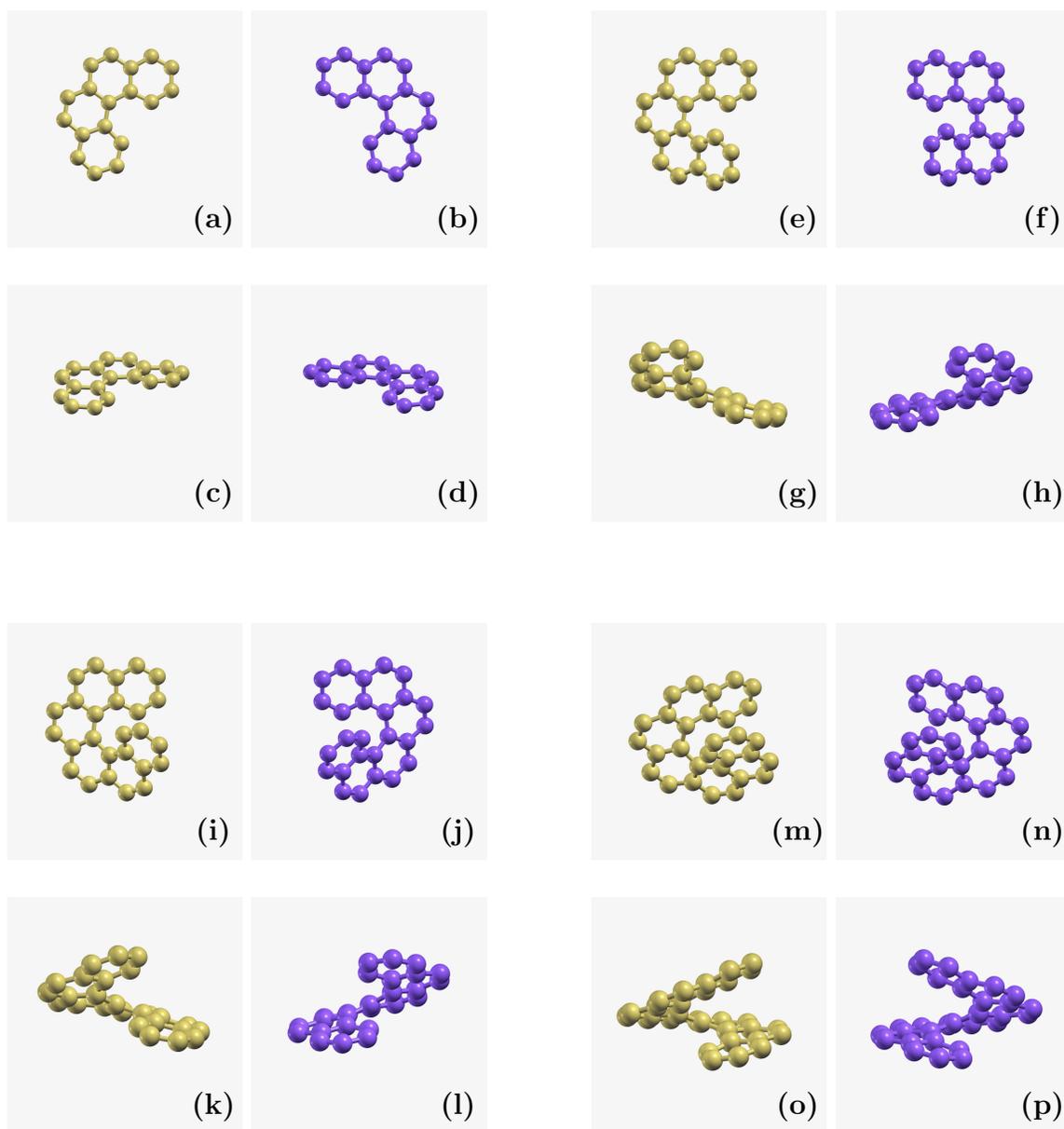


Figure 5.1: Visual representations of helicenes in model  $D$ , after the first simplification step. Each carbon atom in the original model is assigned one Lennard-Jones sphere, whose parameters include contributions from the outer hydrogen atoms. Each benzene ring is represented in this way by a hexagonally-arranged set of six Lennard-Jones spheres. [4]H: panels (a)–(d), [5]H: panels (e)–(h), [6]H: panels (i)–(l), [7]H: panels (m)–(p). For a better representation of the three-dimensional structure, each molecule is shown both in a top and a side view. Molecules coloured in purple represent the optical isomers corresponding to the yellow coloured counterparts.

Table 5.1: Carbon constitution for helicene series.

Molecule	$n_C^{\text{ext}}$	$n_C^{\text{int}}$	$n_C^{\text{tot}}$	$n_C^{\text{ext}}/n_C^{\text{tot}}$
[1]H	6	0	6	1.00
[4]H	12	6	18	0.67
[5]H	14	8	22	0.64
[6]H	16	10	26	0.62
[7]H	18	12	30	0.60
[6]C	12	12	24	0.50

species is two, one for the backbone carbons (denoted C), a second for the surface (denoted S) and the three distinct atomic pairs to consider are C–C, C–S and S–S. The resulting six Lennard-Jones parameters ( $\sigma_{CC}$ ,  $\sigma_{CS}$ ,  $\sigma_{SS}$ ,  $\epsilon_{CC}$ ,  $\epsilon_{CS}$  and  $\epsilon_{SS}$ ) were fitted to match the dimer intermolecular distance, dimer energy, surface binding distance and adsorption energy determined from electronic structure calculations.

The intermolecular and surface adsorption distance used were those for benzene ([1]H) and coronene ([6]C). The two reference molecules are planar, adsorb flat on the surface and provide two limit cases for averaging out hydrogen contributions over the united atoms. Figure 5.2 displays graphically [1]H and [6]C within model *D*. The fraction of external carbons is 100% for [1]H and 50% for [6]C (see Table 5.1). For helicenes this number is 50% in the limit of an arbitrarily large molecule, and between 60% and 70% for the molecules of interest. For this reason, the Lennard-Jones parameters adopted for helicenes are averages of [1]H and [6]C parameters. Interatomic distances do not change significantly with increasing molecular size, rendering the parameters determined from [1]H and [6]C transferable to helicenes.

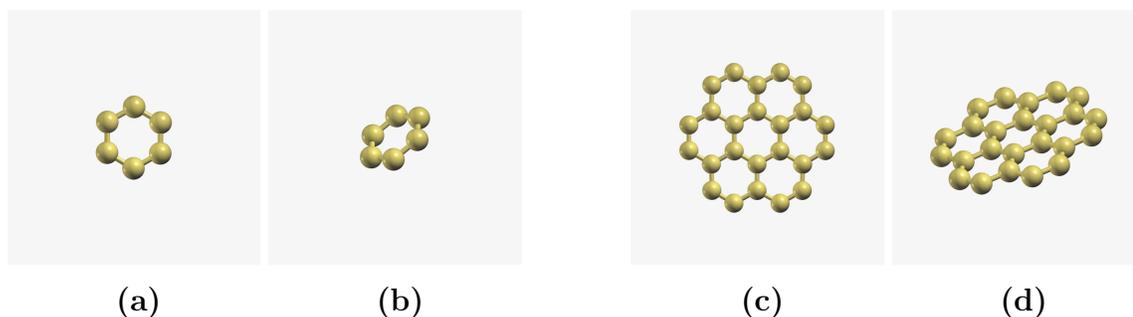


Figure 5.2: Visual representations of benzene ([1]H: (a),(b)), and coronene ([6]C: (c),(d)) in model *D*. Each carbon atom is assigned one Lennard-Jones sphere, whose parameters include the contributions from hydrogen atoms. Molecules are shown in a top and a side view.

## 5.2.2 Model $S$

Model  $D$  was further reduced to model  $S$ , a coarse-grained representation of the helicene as a sequence of Lennard-Jones spheres arranged in a three-dimensional helix. In  $S$  a sphere is positioned at the geometric center of each ring, and represents that ring. Figure 5.3 displays the helicene series represented in model  $S$ . The Lennard-Jones parameters for the coarse-grained model were derived from parameters in model  $D$ , by establishing the direct correspondence between the  $S$  united atoms and  $D$  rings. The Lennard-Jones parameters used in the subsequent systems are tabulated in Table 5.2.

The coarse-grained representation includes the minimum level of detail required for replicating a helix with Lennard-Jones spheres. Geometric details such as ring planarity and directionality are excluded. Deformations due to inter-molecular interactions and surface adsorption have also been excluded. The geometry of coarse-grained helicenes is mapped onto gas-phase optimized helicene monomers. This is achieved by replacing each of the ortho-fused benzene rings contained within the corresponding helicene molecule in model  $D$  with a Lennard-Jones sphere placed at the geometric center of that ring.

Table 5.2: Lennard-Jones parameters for models S and D

Model*	$n_C$	$\sigma_{CC}^*$	$\epsilon_{CC}^{\dagger}$	$\sigma_{SS}^*$	$\epsilon_{SS}^*$
[4]H- $S$	4	3.0	1.0	1.4	2.2 (1.0 $^{\ddagger}$ )
[5]H- $S$	5	3.0	1.0	1.4	2.2 (1.0 $^{\ddagger}$ )
[6]H- $S$	6	3.0	1.0	1.4	2.2 (1.0 $^{\ddagger}$ )
[7]H- $S$	7	3.0	1.0	1.4	2.2 (1.0 $^{\ddagger}$ )
[1]H- $D$	6	2.76	1.00	2.06 (2.57 $^{\ddagger}$ )	3.553 (3.101 $^{\ddagger}$ )
[4]H- $D$	18	2.76	1.00	2.06 (2.57 $^{\ddagger}$ )	3.553 (3.101 $^{\ddagger}$ )
[5]H- $D$	22	2.76	1.00	2.06 (2.57 $^{\ddagger}$ )	3.553 (3.101 $^{\ddagger}$ )
[6]C- $D$	24	2.76	1.00	2.06 (2.57 $^{\ddagger}$ )	3.553 (3.101 $^{\ddagger}$ )
[6]H- $D$	26	2.76	1.00	2.06 (2.57 $^{\ddagger}$ )	3.553 (3.101 $^{\ddagger}$ )
[7]H- $D$	30	2.76	1.00	2.06 (2.57 $^{\ddagger}$ )	3.553 (3.101 $^{\ddagger}$ )

\*The  $-S$  and  $-D$  notation was used to denote the simple and detailed versions of models respectively.

$^{\dagger}$ Reduced LJ parameters are relative to the potential of atoms in the adsorbate,  $\epsilon_{CC}$ , in the context of a strong-binding surface potential.

$^{\ddagger}$ Two different values for the LJ surface parameters were used. The value in parenthesis was used in the context of a weak-binding potential.

Three molecular characteristics vary significantly between the homologues chosen

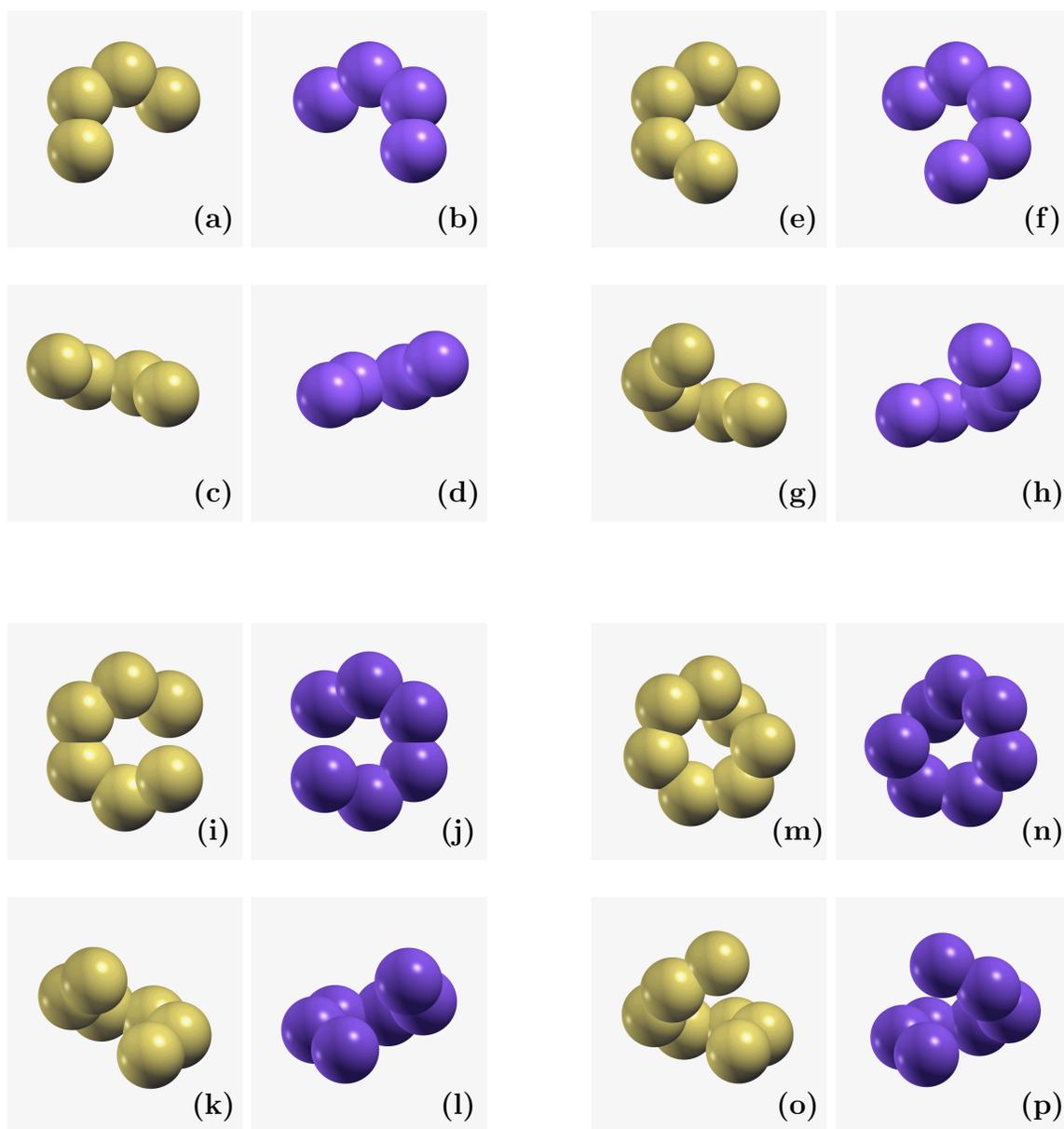


Figure 5.3: Visual representations of the molecular models within the  $S$  model. Each six-atom conjugated ring is represented by one LJ sphere. [4]H: panels (a)–(d), [5]H: panels (e)–(h), [6]H: panels (i)–(l), [7]H: panels (m)–(p). For a better representation of the three-dimensional structure, each molecule is shown both in a top and a side view. Molecules coloured in purple represent the optical isomers corresponding to the yellow coloured counterparts.

for simulations: footprint, planarity/twist and chirality. Lower homologues, in particular [4]H, have an incomplete helical turn between end rings and therefore a small

footprint, a strong planar geometry/small twist and weak chirality (see Figure 5.3 panels (a) – (d)). At the other end of the series, [7]H has a complete helical turn and some degree of overlap between end rings (see Figure 5.3 panels (m) – (p)), manifesting maximal footprint, helical twist and chirality. Model *S* captures these variations successfully and its simplicity allows us to isolate their influence on self-assembly.

## 5.2.3 Simulation Details

### Numerical details

Each simulation followed between 30 and 80 molecules, using a Parallel Tempering Monte Carlo procedure over a range of temperatures. Both enantiopure and racemic mixtures were considered for chiral molecules. We used two system sizes, standard and large, with the number of surface atoms, surface size and molecules varying between the two versions. The values for these numbers are presented in Table 5.3. Box sizes were chosen to allow for the formation of low density phases, with the possibility of monolayer formation in the vicinity of saturation. As monolayer density varies among homologues, this was only an approximate target. No periodic boundary conditions were employed.

Table 5.3: Numerical details for MC simulations of models *S* and *D*: number of surface atoms included, number of molecules present in the box, simulation box length and number of MC steps. For model *S* two system sizes were considered.

Model	$n_{\text{at}}^{\text{surf}}$	$n_{\text{mol}}^*$	$d_{\text{box}}^{\dagger}$	$n_{\text{steps}}^{\text{MC}}$
S	1560 (2368)	50 (80) <sup>‡</sup>	12 (15)	$> 10^8$
D	1674	30	20	$10^7$

\*For simulations of molecules with chirality, both enantiopure and enantiomeric mixtures were considered. In the case of mixtures, half the total molecules have inverted chirality. <sup>†</sup>The box length across an axis is presented in reduced units. <sup>‡</sup>The values in parenthesis correspond to larger systems.

### Surface models

Simulations were performed on rigid, two-layered surfaces with hexagonal symmetry, corresponding to the (111) facet of the FCC crystal, totalling approximately 1600 atoms for standard sizes, and over 2300 atoms for larger size simulations (see numerical details in Table 5.3). The surface potential corresponded to either a weakly-binding (gold) or a strongly-binding (copper) potential (see Table 5.2).

## Parallel Tempering Monte Carlo

Simulations were performed using an in-house Parallel Tempering Monte Carlo (PTMC) code, with details available in the cited works.<sup>182;183</sup> This consisted of a Markov chain with (i) standard MC molecular translations and rotations, with a Metropolis acceptance criterion for the canonical ensemble,<sup>112</sup> (ii) molecular swap moves between enantiomers and (iii) swap moves between neighbouring-temperature replicas. Evaporative effects (molecules leaving the simulation box) were forbidden. The number of MC steps post-equilibration varies with system size (see Table 5.3). Temperature swaps were attempted in 10% of the MC moves. Each system consisted of 30 replicas, with a temperature spacing adjusted to meet the target acceptance ratio of 10%.

## 5.3 Results and Discussion

### 5.3.1 Surface Potential vs. Intermolecular Interactions

Our choice in surface size and number of molecules targeted the formation of a low-density monolayer. Across the investigated series monolayer formation was confirmed as a dominant outcome. In Figure 5.4 we showcase a typical monolayer example for each homologue. The surface coverage per adsorbed molecule generally coincides with the molecular footprint. This implicates the surface potential as fundamental in self-assembly, directing the orientation of molecules upon adsorption and formation of intermolecular-driven structures. This is in line with experimental observations of helicenes on coinage metals, and suggests that the adsorption effects discovered in our previous electronic structure study are driving helicene surface self-assembly.

Alternatives to footprint-covered monolayers were also observed, specifically for higher homologues. In some circumstances [6]H and [7]H cluster into adsorbed 3D aggregates, with no detectable layered organization. Occasionally, hybrid aggregate structures with some degree of layered order are present, as well as distinctly multi-layered formations with vertical transfer of order from the monolayer upwards. While monolayers appear to be the most common self-assembly strategy, aggregates prefer larger surfaces, a strong molecular out of plane twist and stronger surface potentials.

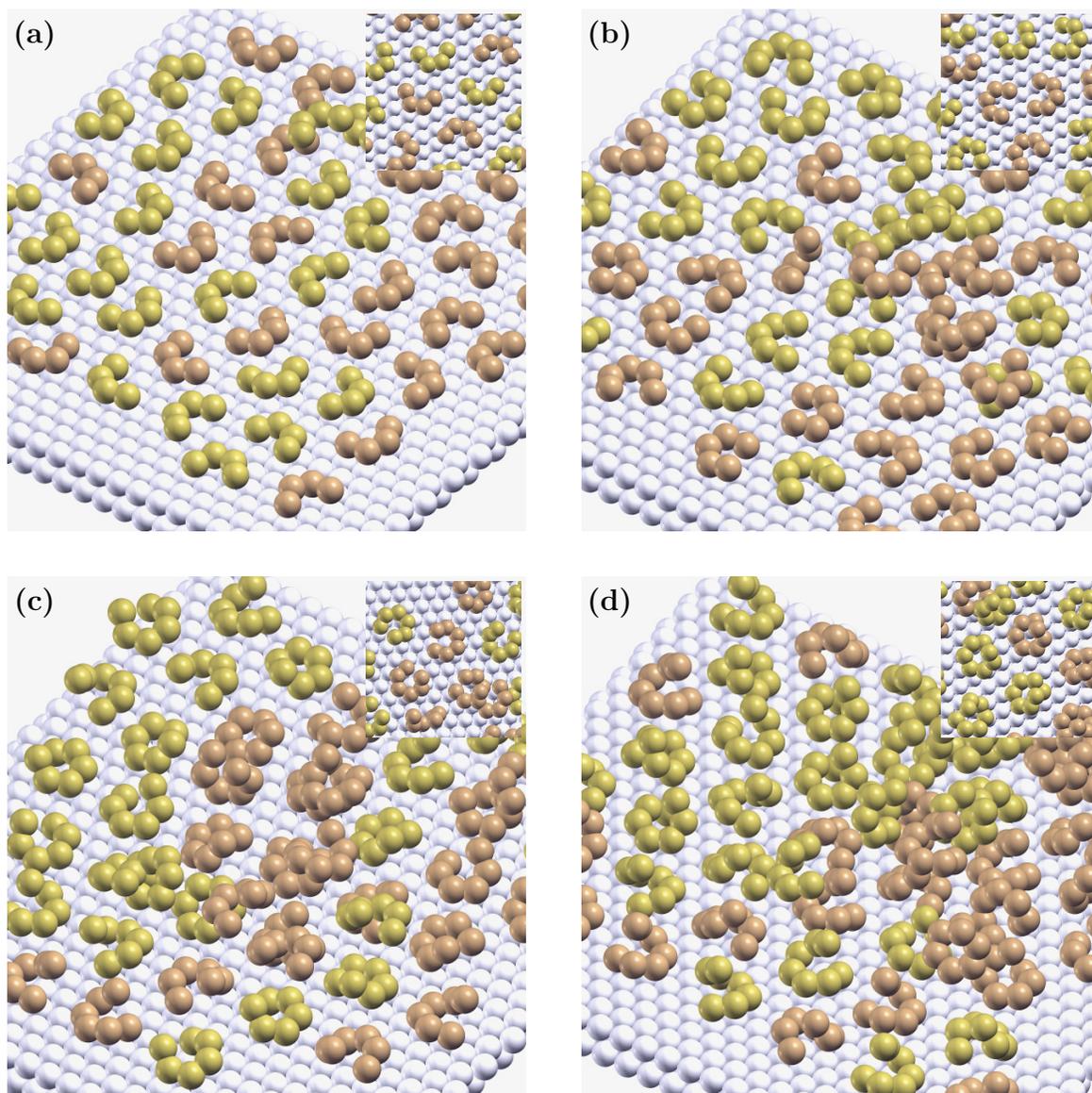


Figure 5.4: Common examples of monolayer formation in [4]H – panel (a), [5]H – panel (b), [6]H – panel (c) and [7]H – panel (d). The display in the top-right corner of each panel can be used to estimate the molecular footprint of each homologue and the commensurability of the monolayer. All systems are enantio-pure. Colouring differences between molecules are intended to aid visualization and correspond to identical geometries.

### Monolayers

In monolayers molecules adsorb flat, taking up a surface area determined by their molecular footprint. Planar molecules lie parallel to the surface, while higher homo-

logues adopt an orientation as close to parallel as possible, given their out of plane twist. This behaviour is compatible with our previous observations on electronic structure calculations, where we discovered that helicenes may undergo significant deformation to maintain the adsorbed phenantrenic group parallel to the surface.

A consequence of the flat adsorption is that the molecular density within the monolayer is inversely proportional to the molecular footprint. [4]H, having the smallest footprint of the series, packs more molecules per unit area of surface than other homologues, without sacrificing the maximal interaction between its rings and the surface (see Figure 5.4).

Monolayers retain the six-fold symmetry characteristic of the underlying surface. The transfer of symmetry from the surface to the monolayer is supported by the fact that molecules trapped within the monolayer adopt specific orientations relative to symmetry axes. In molecules with large footprint, there is evidence for the formation of a second layer, adsorbed on top of the monolayer, and some vertical transfer of order (see Figure 5.4, top-right inserts). There is a distinct lack of surface-specific features within the organization of the monolayer.

### **3D aggregates**

Conglomerates with no distinguishable order are observed in molecules with a strong twist, such as [7]H, on larger surfaces (see Figure 5.5, panels (c), (d)). Flat molecules like [4]H and [5]H show no tendency to form 3D structures, and instead prefer parallel surface adsorption whenever the density is below that of the saturated monolayer (see Figure 5.5, panels (a), (b)). This implicates the twist, or the deviation from planarity, as necessary precursors in the formation of 3D aggregates. The twist of helicenes is intimately coupled with their chiral character, making 3D aggregates relevant for enantioselectivity.

### **Surface-driven twist enhancement**

In the context of low density adsorption, rigid and planar PAHs lead exclusively to phases corresponding to flat adsorption, no intermolecular recognition and no enantioselectivity. The rigid nature of molecules excludes from our results the influence of surface potential on molecular twist. Previous electronic structure calculations have demonstrated that adsorption is affecting molecular twist, by flattening the phenantrenic head and coiling the tail end of the molecule. While this is more of a

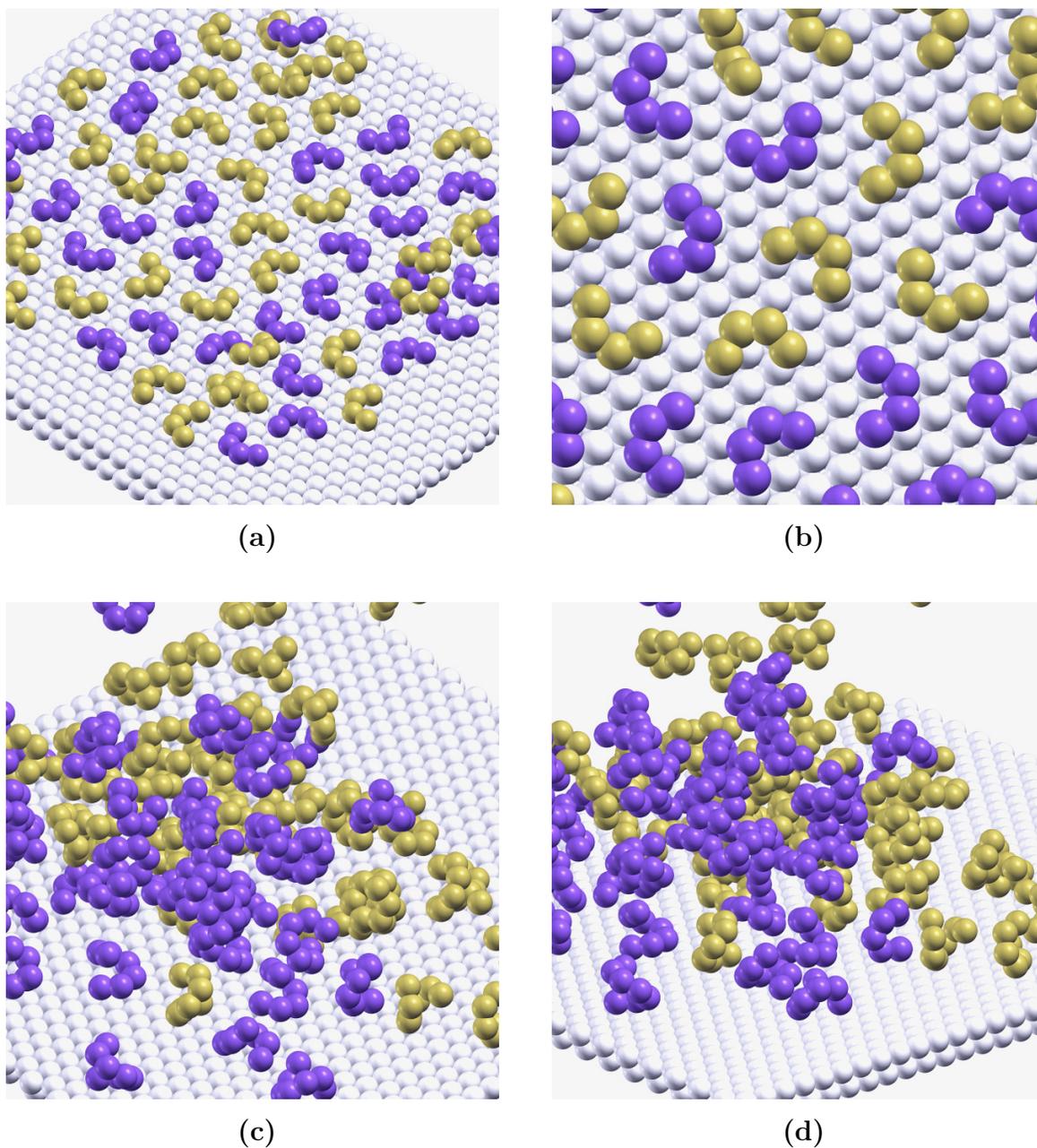


Figure 5.5: Monolayer adsorption of [4]H racemic (panels (a), (b)) and 3D aggregation of [7]H racemic (panels (c), (d)) on large surfaces. The planar character of [4]H molecules promotes flat adsorption, while the pronounced out of plane twist present in [7]H molecules facilitates interactions in three dimensions. Yellow and purple colours denote the two enantiomers.

transfer of twist towards the tail end of the molecule, it can be considered an effective twist-enhancing phenomenon, since only the tail is available for intermolecular

interactions following adsorption. In flexible systems this effect could drive the formations of 3D aggregates even with molecules as small and flat as [4]H. Further studies on helicenes with twist-enhanced geometries could offer more insight into the role of surface on molecular twist.

### **Twist-induced rotational alignment**

Helicenes with a complete turn and a strong overlap between end rings commonly form monolayers with a six-fold character. [6]H and especially [7]H manifest a high twist which can act as a trigger for orientational coupling between adsorbed neighbours. Overall, evidence for such order is absent. In the presence of a weak-binding surface, some evidence of a slight orientational coupling can be identified in some snapshots. Figure 5.6 compares the self-assembly outcome in the presence of a strong-binding and a weak-binding surface, in racemic of [6]H and enantiopure [7]H, respectively. The degree of orientational coupling is higher in the presence of a weak surface potential, in both systems. This effect is weak but promotes the view that the surface acts as support for self-assembly. The surface provides adsorbed molecules with a choice between six equally likely orientations, and molecular features, such as twist, will manifest discrimination between these orientations. A weaker surface potential maintains the six-fold symmetry in the monolayer and facilitates recognition between neighbouring molecules, acting synergistically to enhance phase order and stability. The lack of detailed structural features in our models are the most likely cause for the relative absence of orientational coupling between adsorbed molecules.

### **5.3.2 Beyond Competition: Multilayered Structures**

Between the extremes of footprint-determined monolayers and intermolecular-determined aggregates, we recorded a range of hybrid structures, broadly classified as multi-layers and adsorbed rows. Multi-layers are three-dimensional, vertically layered structures and adsorbed rows appear as aggregates, split apart along certain symmetry axes. Their frequency of sampling is remarkably rare, being vastly outnumbered by monolayers.

#### **Second layer nucleation**

With a density close to monolayer saturation, structures characteristic of second layer nucleation were identified. The simplest ordered structure forming above the

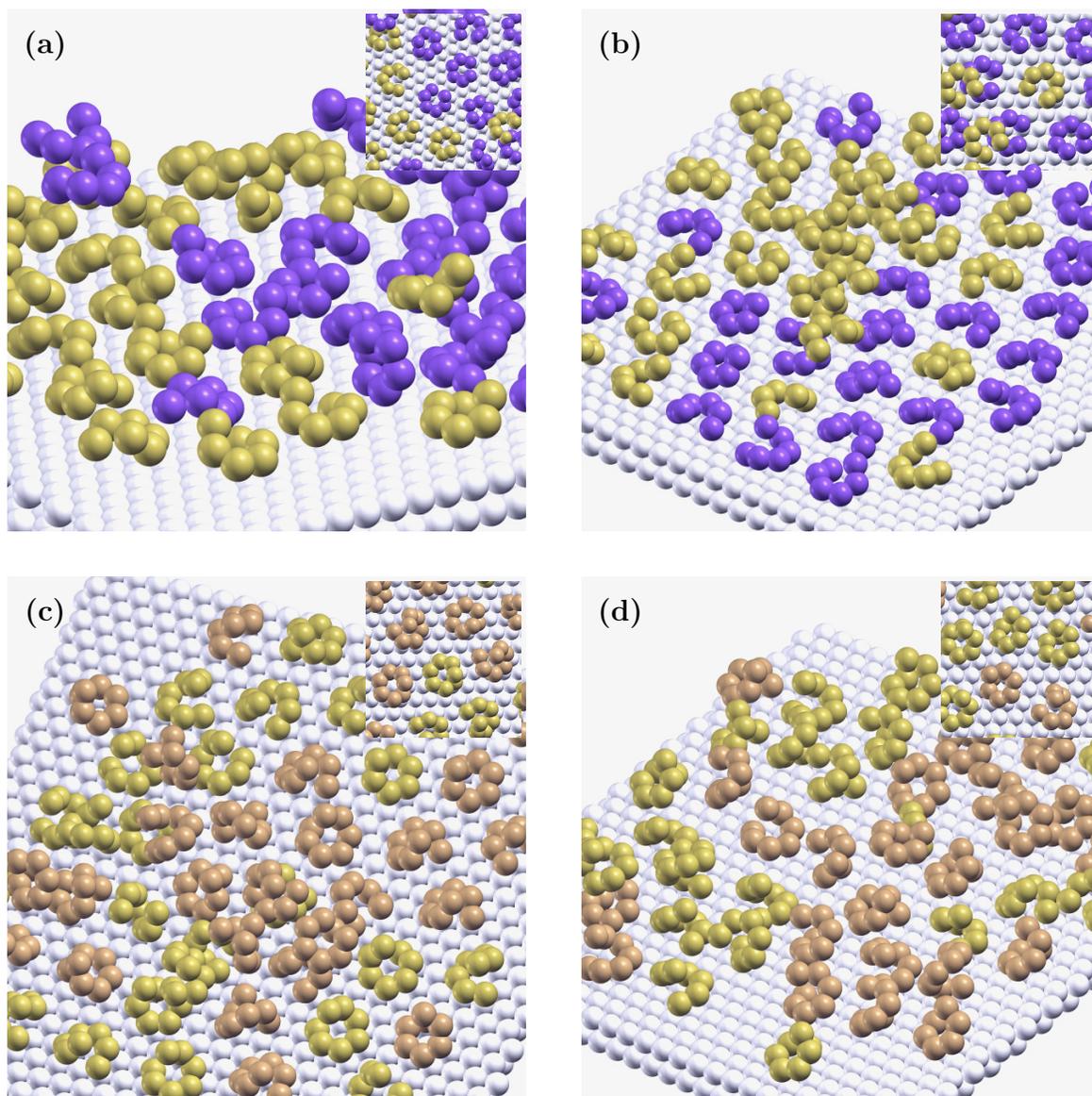


Figure 5.6: Comparison of adsorbed phases formed by highly-twisted helicenes in the presence of strong (left panels) versus weak (right panels) surface potential. Top panels: adsorbed racemic [6]H on a weak-binding (panel (a)) and a strong-binding (panel (b)) surface, respectively. Bottom panels: adsorbed enantiopure [7]H on a surface with weak (panel (c)) and strong (panel(d)) surface potential, respectively. The degree of orientational alignment between neighbouring molecules is slightly enhanced by a weaker potential.

monolayer is a trimer with molecules arranged in a ring-like manner and exhibiting orientational coupling within the trimer and with monolayer molecules. Figure 5.7 displays close-up top views of nucleation trimers for each of the four homologues in

the series.

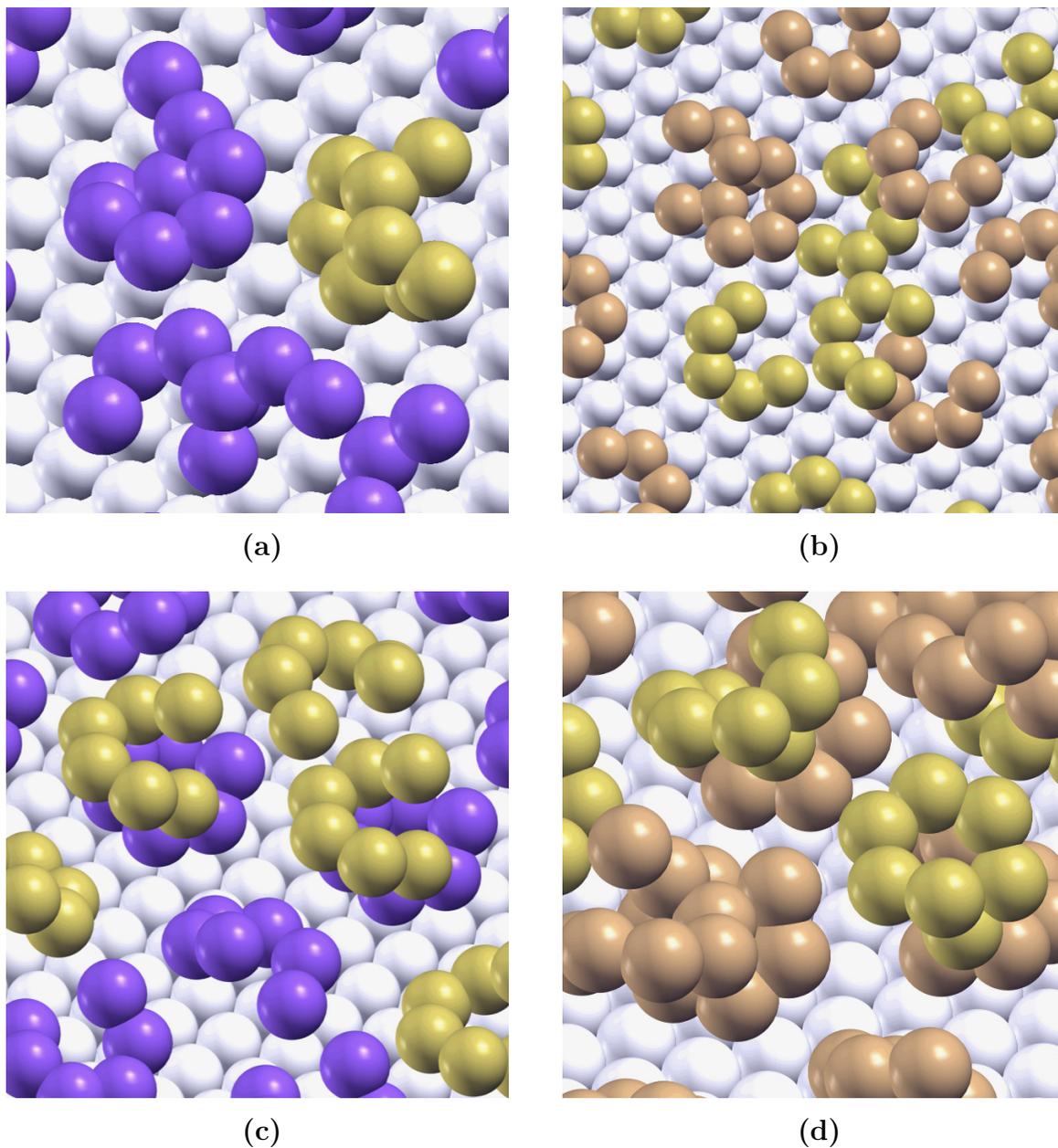


Figure 5.7: Close-up views of the second-layer nucleation centers for helicenes adsorbing beyond the saturated monolayer: panel (a) – racemic [4]H; panel (b) – enantiopure [5]H; panel (c) – racemic [6]H and panel (d) – enantiopure [7]H. The six-fold structure of the trimer is apparent in all homologues, irrespective of the chiral composition of the system. The degree of staggering between the first and second layer varies between homologues. The corresponding side views of these structures are displayed in Figure 5.8.

All homologues adopt the trimer structure with varying frequency. The orientational coupling is caused by the alignment of molecules along the twist axis, with variations in the relative orientation along the axis. Significant staggering between top and monolayer molecules is observed (see Figure 5.7). Common between homologues is the distinct layered structures of trimers, including those of molecules with a strong out of plane twist. This layered character is apparent in the side views of the trimers, which are displayed in Figure 5.8.

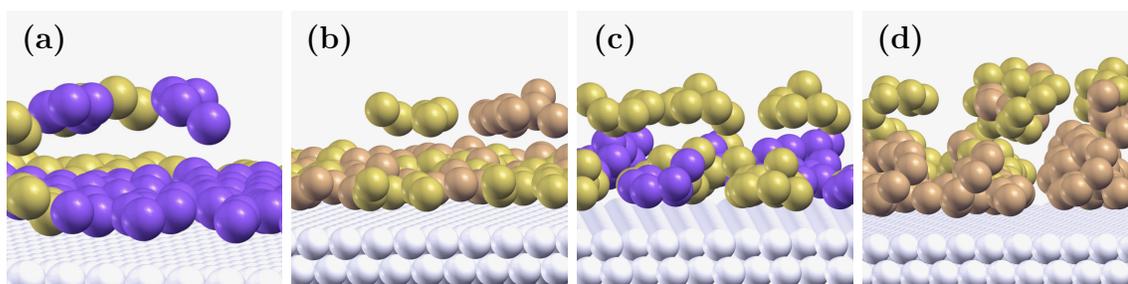


Figure 5.8: Side views of the second-layer nucleation centers, complementing the top views displayed in Figure 5.7: panel (a) – racemic [4]H; panel (b) – enantiopure [5]H; panel (c) – racemic [6]H and panel (d) – enantiopure [7]H. The strong layered character of the structure in all homologues is apparent.

The six-folded character of the surface is evidently responsible for molecular alignment along their twist axis and trimer formation, but the variation in the inter-layer staggering between homologues (see Figure 5.7 panel (a) versus panel (b)) indicates that the mechanism by which symmetry is transferred vertically depends on inter-molecular details.

What appears to be missing is any dependence of surface cell size or potential. This is in contrast with the results from Chapter 4 where the nature of the surface had a strong effect on the surface-molecule commensurability. We attribute this lack of surface selectivity to the lack of detail in models, the molecules being too coarse to respond to the subtle differences between surfaces.

### Multi-layers

All helicenes but higher homologues in particular, occasionally go beyond second layer nucleation and establish multi-layered domains. Such structures are displayed in Figure 5.9 for each of the homologues in the investigated series. They differ from

disordered aggregates by having a strong layered component maintained vertically. This suggests that inter-molecular interactions are important for their stability.

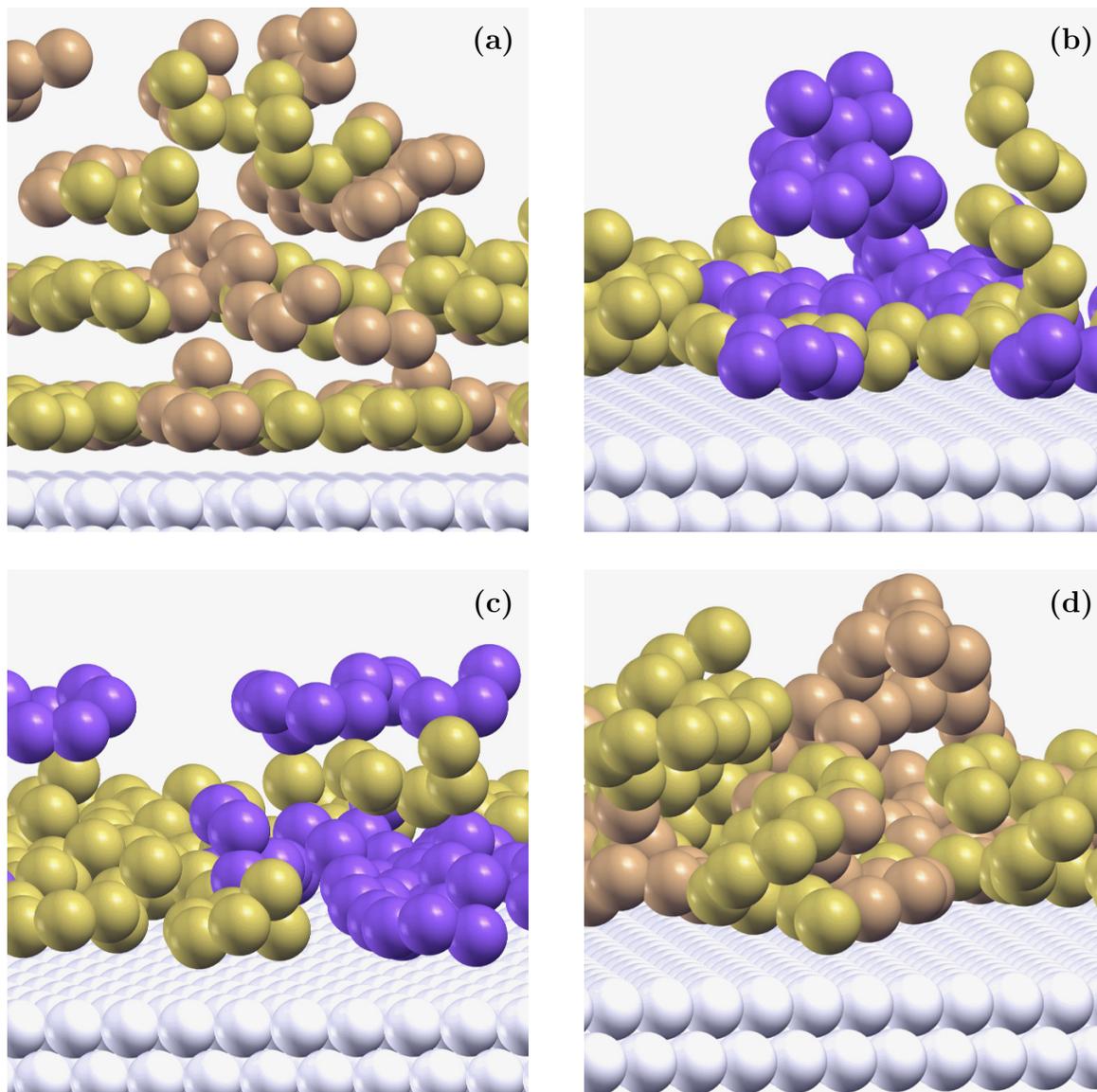


Figure 5.9: Side views of second- and multi-layered formations built on top of the monolayer foundations by adsorbed helicenes: panel (a) – enantiopure [4]H; panel (b) – racemic [5]H; panel (c) – racemic [6]H and panel (d) – enantiopure [7]H. The strong layered character appears to be preserved vertically across several layers when present. The presence of the layered structure appears independent of the racemic composition.

The topology of multi-layers is similar that of the second layer trimer, the latter

likely acting as a nucleation center for vertical growth. The geometric variations in trimers are thus transferred vertically to the extended domains, while the basic three-fold monolayer structure is preserved.

### Adsorbed rows

The other class of structures are horizontal rows of tilted molecules, held together by end-ring inter-molecular interactions. These formations appear in the self-assembly of [6]H and [7]H (Figure 5.10 displays them for enantiopure [7]H). Geometrically, rows appear as highly ordered, horizontally-adsorbed, intermolecular stacked aggregates (see Figure 5.10, panels (a), (b)). The tilted helicenes appear to be stable in the presence of other tilted neighbours, with which end-ring interactions become available.

Because adsorbed rows occur with very low frequency, they are likely unstable relative to the monolayer formation. Additional electronic structure optimizations designed to study tilted adsorption of helicene oligomers is needed. In addition to answering the question of tilted stability, such studies could also shed light on the issue of preferential adsorption along certain symmetry axes.

### 5.3.3 Chirality and Enantioselectivity

We assessed the role chirality plays in our simulations by contrasting and comparing the enantiopure and racemic versions of each simulated system. Most recorded phenomena demonstrate an independence from chirality.

#### Lateral enantioselectivity

With a majority of systems manifesting predominantly monolayer formations, the source of enantioselective behaviour can only be lateral interactions between neighbouring molecules. No conclusive evidence of lateral-driven enantioselectivity was observed. The distinct pockets of like and unlike domains match a distribution expected from random fluctuations. This indicates the dominance of surface potential over lateral interactions in low-density monolayer formations. This also demonstrates once more the lack of detail of our molecular models.

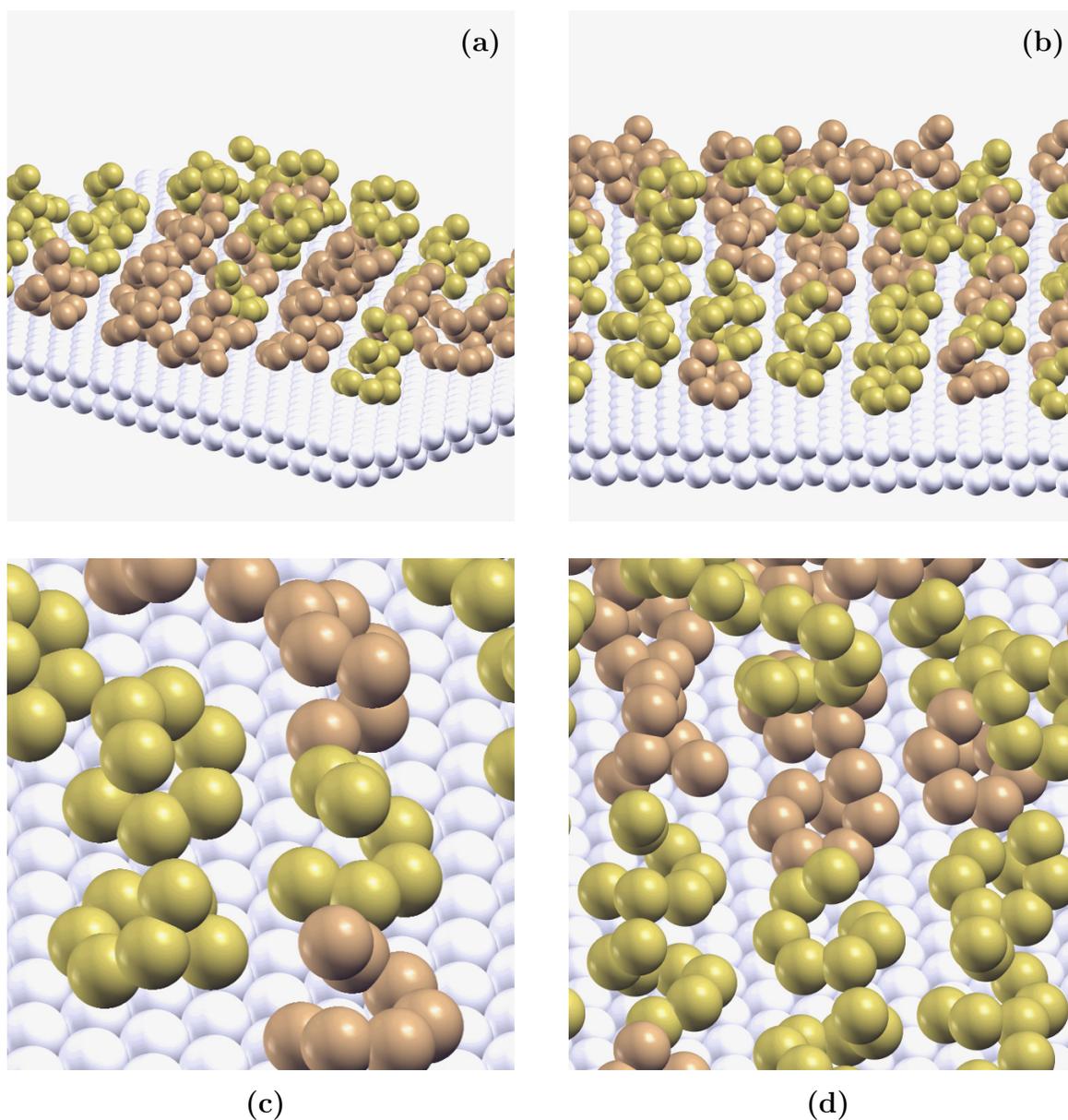


Figure 5.10: Adsorbed rows of enantiopure [7]H on surfaces with strong (left panels) and weak (right panels) potential. The bottom panels show a close-up view of the respective rows. Molecules within rows appear to be tilted and stabilized by intermolecular end ring interactions. Rows adsorb along certain symmetry axes only, a surface effect which inhibits 3D aggregation.

### Multi-layered enantioselectivity

Chiral recognition is likely to occur among stacked molecules. In this study, there is some observational evidence to support stacking-induced enantioselectivity. In

racemic mixtures with multi-layered structures we observed significant columnar segregation of the two enantiomers. Figure 5.11 displays such columns in [7]H racemic mixtures. Although molecules express significant vertical segregation when forming the columns, there is no evidence of horizontal enantioselectivity between columns, as their relative positioning is either random or possibly alternating. Despite their rarity, multi-layered structures are much more likely to display enantioselective characteristics than monolayers. Because of their rarity, we are unable to assess the question of enantioselectivity statistically.

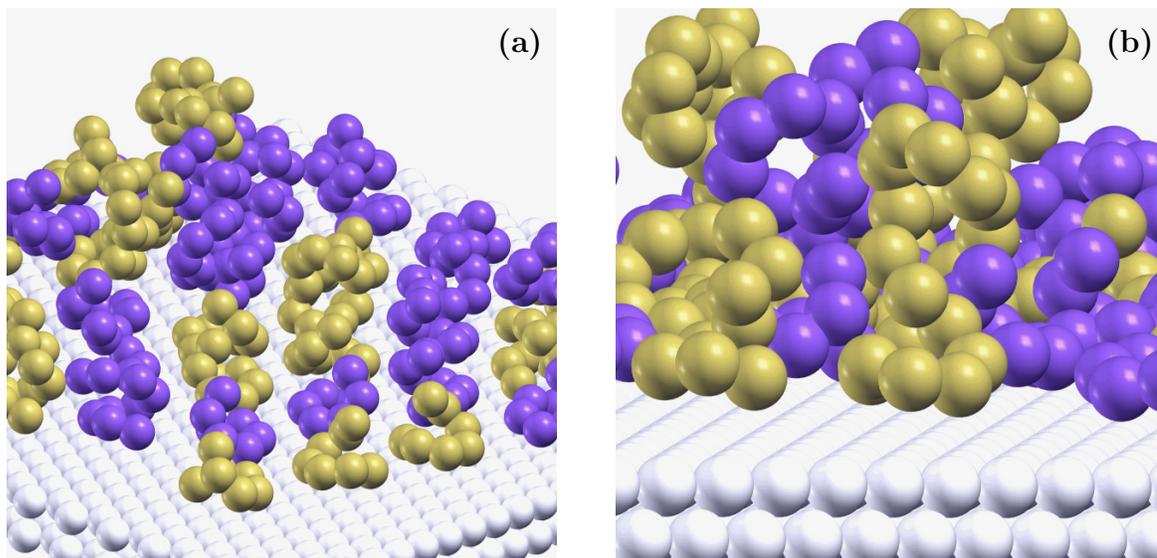


Figure 5.11: Multi-layered formations of racemic [7]H on weak-binding surface. Molecules appear to stack in enantiopure columns, dispersed across the surface in an alternate or random manner. Panel (b) displays the close-up side view of the stacked columns in panel (a).

### 5.3.4 Future Prospects

The study we conducted in this chapter targeted a small set of molecules and two kinds of surfaces, with the specific goal to extend the work on helicene dimer adsorption towards multi-molecular systems. While we elucidated some of the issues we had from that study, we naturally discovered new phenomena that warrant further investigations on their own.

In order to clearly determine the role enantioselectivity plays in the formation of the monolayer we require a series of Monte Carlo runs with larger size systems and

a weak binding surface potential. Additionally, lower density runs on larger surfaces could allow for a better description of unconstrained lateral interactions.

As we found evidence of tilted adsorption playing a role in the formation of surface-directed ordered domains, we need further electronic structure calculations of adsorbed trimers and possibly oligomers in order to test tilt stability in the presence of intermolecular end-ring interactions.

Lastly, in order to overcome some of the limitations inherent in rigid molecular models, we wish to conduct a series of Monte Carlo simulations where we include and exclude adsorption and stacking deformations withing molecular geometries, in otherwise identical systems. Comparing data across geometric models this way will help us identify the role molecular flexibility has in self-assembly.

## 5.4 Conclusion

In this chapter we conducted a first foray into the nature of multi-body self-assembly of helical molecules on symmetric surfaces. By applying a two-step procedure of molecular model reduction, we translated the knowledge and data gained from our previous study of monomer and dimer helicene adsorption, to a multi-molecular simulation procedure capable of capturing aspects of self-assembly beyond the range of local, pair-wise interactions. With this we explored a significant part of the self-assembly space of helical molecules.

In helicenes of any size we identified the surface-driven monolayer as a dominant scheme of ordered organization. When applicable, surface symmetry is transferred to higher layers, as evidenced by the presence of monolayer-adsorbed trimers with a ring-like, six-folded symmetry in the process of second layer nucleation. Intermolecular interactions, on the other hand, facilitate the formation of 3D, disordered aggregates, driven primarily by the molecular out of plane twist, evidenced by the absence of such formations in systems with planar molecules.

When the surface potential acts in conjunction with intermolecular interactions, formations with intermediate levels of organization and order are produced. Multilayered domains demonstrate that the surface is more than just support for 3D aggregates, as it actively transfers symmetry to structures vertically, as far as three to four layers high. In adsorbed rows, tilted helicenes are stabilized on the surface by inter-molecular stacking, by preferentially adsorbing along certain symmetry axes.

The lack of evidence for enantioselectivity in monolayers is likely due to the low

importance lateral interactions play in the low-density studies we conducted here. There is some evidence, however, of enantioselective behaviour, in formations where intermolecular stacking play a significant role.

Further work is required to elucidate the nature of hybrid structures observed, particularly electronic structure optimization studies of adsorbed oligomers, and Monte Carlo simulations at higher system sizes, higher densities, larger surfaces, and molecules with embedded deformations.

## Chapter 6

# Conclusions

### **The Blooming Field of Surface Self-Assembly**

Materials with complex properties and tunable responses to external stimuli have great potential of applicability to a wide variety of tasks. Their desirable properties are a consequence of their heterogeneous nature, a complex combination of highly specific and predictable inorganic components and flexible and cost-effective organic parts. As the complexity of interaction between components is manifested at ever-smaller scales, the role of molecular self-assembly in their manufacturing and maintenance becomes indispensable. For a cost-effective way of industrial production to be possible, bottom-up development methods must replace the classical, unreliable and energy inefficient top-down manufacturing methods.

Surface self-assembly is one promising development method, where a surface is used to direct a three-dimensional molecular self-assembly strategy onto a two-dimensional template, yielding layered materials which are easier to characterize and have reliable properties. The subtle interplay of intermolecular forces, coupled with the surface induced potential, causes such systems to self-assemble into a variety of unpredictable patterns, and consequently manifest unique opto-electronic properties. Understanding how these patterns are determined by properties of the constituent blocks and the nature and symmetry of the surface is currently of great interest for the purpose of rational design of solid-state materials.

### **The Increasing Role of Computer Simulations**

With the increasing availability and reducing cost of computational resources, larger systems can be modelled with increasing accuracy, at the same time as experimental

techniques probe ever-smaller scales with better resolution. Given this trend, the role of computer simulations in answering fundamental questions about the process of surface self-assembly, as well as answering detailed questions about specific systems becomes increasingly relevant. Simulations of simple models can direct us towards an understanding of how local interactions between building blocks give rise to long range ordered phases, in a way that experimental studies cannot provide. At the other end of the spectrum, detailed, accurate simulations of systems with practical value can help direct experimentalists' choices towards greater efficiency and away from trial and error. Within the computational constraints at the moment, there are two main roadblocks that limit the field of computational chemistry from achieving this role, the problem of scale and that of heterogeneity.

As self-assembly on solid surfaces is a process realized by various phenomena at multiple length scales, from intra-molecular interactions to the bulk phase formations, a comprehensive understanding in this area requires a combination of quantum-level information and large-scale classical simulations. Painting a complete picture of surface self-assembly requires the transfer of information from one level of analysis to another, reconciling the vastly different resolution scales at which the system evolves in the two approaches.

Solid-state self-assembled materials, with organic/inorganic parts, crystalline and molecular components, covalent and metallic traits, call for a variety of versatile methods to deal with interactions in a computationally cost-effective way. Various conceptual and numerical techniques borrowed from quantum theory, crystallography, and solid-state physics, can be used together to tackle the heterogeneity of solid-state assembled materials. In Chapter 2 we presented a comprehensive collection of such techniques.

### **Computational tools for studying self-assembled phases**

At one level of analysis, supramolecular interactions can be traced down to fundamental electrostatic and geometric effects. The basic molecular variables controlling phase formation in surface self-assembly can be isolated with the use of simple models and investigated in extended systems. Such systems must be capable of incorporating a large number of molecular building blocks in order to access the long range, supramolecular order. Monte Carlo techniques are uniquely suited to explore complex, multi-molecular phase spaces with great computational efficiency. Careful

choices in the design of Markov chain moves result in flexibility for sampling potential energy surfaces. The Parallel Tempering formalism elegantly solved the problem of entrapment in local minima and improving the efficiency of sampling. The Lennard-Jones-electrostatic potential provides the right balance between computational efficiency and detail of interaction at this level of analysis.

Investigating the degree of order in extended systems, statistical methods have to be employed in parsing through the large collection of output data from Monte Carlo simulations. Order parameters and other averages over the system and across the sampling set, can be linked to various classes of structure formations in order to establish a correspondence between experimental observations and simulation results.

### **Theoretical studies of surface self-assembled simple models**

In Chapter 3 we demonstrated the rich potential for the theoretical study of surface self-assembly with the design of simple molecular geometries with modulated features: dipole moment, steric effects, length molecular angle and relative interaction with the substrate. Our goal was to investigate at the fundamental level how the competition between van der Waals interactions, dipole-dipole forces and steric effects, as well as the competition between intermolecular and surface-molecule interactions, give rise to multi-molecular adsorbed phases. These interactions are a first-order approximation for surface SA, as we demonstrate that they are the minimum requirements for ordered phase formations.

Multi-body parallel tempering Monte Carlo simulations of our models revealed a rich array of phases being formed in response to the disruption to the strong crystalline motif. These often paralleled experimental observations, particularly when sufficient details were added. Long, linear molecules formed monolayers with nematic or antiparallel alignment, while stronger dipoles facilitated the formation of multilayer aligned phases. Substituents further altered the self-assembly process, by adding a steric component to the interaction and by limiting the availability of contact points, opening the self-assembly path towards molecular wires.

Overall, the self-assembly strategies available for even the the simplest of models are numerous and this number increases rapidly with added features to the model. In this way, we demonstrate that the gap between theoretical models and experimental results can be bridged by adding structural and geometric features to the molecular models.

Further model-based studies in this area can focus on the availability of multiple contact points, both between molecules and with the surface, as well as on solvent effects, bringing the complexity of the theoretical model closer to experimentally-relevant molecular systems.

### **Computational tools for studying local interactions**

In Chapter 4 we transitioned to the local scale, where electronic structure techniques are required to determine the details of adsorption and inter-molecular interactions between building blocks. Local interactions, even when small, can in aggregate change the nature of self-assembled structures qualitatively. Density functional theory coupled with a good treatment of dispersion interactions is the method choice for identifying and describing stable geometries of molecular systems on solid substrates. Protocols developed specifically for identifying stable geometries in solid state materials, such as comparative electronic structure analysis between monomers, dimers, in gas phase and on the surface, can provide insights into the nature of interactions between building blocks and with the substrate.

### **Helicene monomers and dimers adsorbed on metal surfaces**

In our short-scale interaction study we focused on a specific class of experimentally popular systems: self-assembled helicenes on metallic surfaces. Helicenes have a unique molecular structure, a pronounced enantioselective behaviour and are known to form a variety of phases via adsorption on metals. Much of the research in this field has been focused on the role of surface potential and lateral molecular interactions in helicene assembly. To this we added a study focusing on the role of stacking in the process of helicene adsorption. To this end we considered a series of helicenes and three metallic surfaces, Cu[111], Ag[111] and Au[111] as substrates. We used density functional theory to optimize and analyze the electronic, structural and energetic properties of adsorbed monomers and stacked dimers in gas phase and adsorbed. We considered both cases of enantiopure and racemic dimers in order to identify any enantioselective tendencies.

We discovered that stacking has a strong enantioselective behaviour due to the greater ability of enantiopure dimers to stabilize through  $\pi$ -stacking. On surfaces helicenes adsorbed via dispersive interactions between the conjugated system and the electrons in the metal's sp-band. For the adsorbate, we found that the interaction

induced asymmetry in the manifestation of the helix: the phenantrenic head of the adsorbed helicene was flattened, while strong coiling within the rest of the molecule provided accommodation for the intramolecular steric effects. We argued that the surface enhances the enantioselectivity of adsorbed helicenes through the deformation of the molecular helical twist and corroborated this with the finding that, in adsorbed dimers, the stacked molecule acts synergistically with the surface potential, to further distort the sandwiched molecule, and stabilize the adsorbed dimer.

### **Beyond adsorbed helicene dimers**

Beyond stacked helicene dimers, lateral and multi-body interactions in mono- and multi-layers have to be accounted for in self-assembly. Such scales require multi-molecular models with lower resolutions and a sampling methodology appropriate for extended systems, such as the parallel tempering Monte Carlo formalism. In extended systems density functional theory fails to cope with the rapidly surging computational complexity of the problem.

In Chapter 5 we used the structural and energetic data presented in Chapter 4 to generate a model compatible with our multi-body simulation tools. We applied a two-step model reduction procedure to helicene molecules, generating in the end a cork screw set of models, with all directional and planar information of constituent benzene rings lost. The coarse-grained models retained only the basic helical geometry and main structural and energetic features of helicenes, while discarding most of the short-scale details likely to be averaged out at the large scale.

We conducted parallel tempering Monte Carlo sampling of helicene coarse-grained multi-body systems in the presence of a surface. This revealed overwhelmingly the formation of a monolayer. At low densities, lateral interactions were not playing a significant role in self-assembly, as molecules could afford to be spaced apart and still make full use of their molecular footprint on the surface.

Despite the low density, we discovered some evidence of second layer and even higher layer nucleation, in the form of trimers with a six-folded symmetry transferred vertically from the surface. In the rare circumstances of absent monolayer, inter-molecular interactions promote disordered aggregates in three dimensions. Overall, the structural findings suggests that the surface directs helicene molecules towards a surface-determined self-assembly template, on which inter-molecular local interactions rarely manifest.

### **Balancing model accuracy with computational efficiency**

We attribute the lack of specific intermolecular interactions to the low density, but primarily to the loss of too much information during the coarse graining procedure, rendering our final models too simple to represent helicenes. A noteworthy fact suggesting this is the lack of enantioselective behaviour throughout our sample sets. Another line of evidence is the lack of surface-specific monolayer features, with all surfaces transferring identical three-fold symmetry to adsorbed molecules.

During the coarse-graining procedure we reduced the helicene models beyond the point of acceptable accuracy for a model representing a real system. The coarse-grained models presented in Chapter 5 belong closer to the simple models explored in Chapter 3, than with the system of helicenes it originated from in Chapter 4.

Future work in this area can be adding complexity to the coarse-grained models and determining the minimum required features that manifest enantioselectivity, surface-specific assembly and ordered adsorbed intermolecular domains. Additionally, development of techniques for sampling multi-body systems at higher density will provide access for computer simulations to a vast new area of exploration.

As the experimental and manufacturing techniques move towards materials of ever-increasing complexity with ever-smaller design scales, the value of computational techniques with predictive and descriptive power towards such materials is only going to increase. Material scientists and chemists have a wealth of tools from disparate fields to draw from, for modelling solid-state materials and deal with the inherent complexity of interactions and inevitable computational limitations. Whether it's the development of specific solutions for unique systems, the creation of algorithms to bypass current computational limitations or the integration and streamlining of previously separate techniques, molecular modelling will undoubtedly drive the progress of solid-state materials, and add knowledge to the field of surface self-assembly.

## Appendix A

# Two-Component BSSE-Corrected Binding Energy Determination with SIESTA

### A.1 Introduction

This document was written as a guide on how to use SIESTA for determining the energy of interaction of 2-component systems with the counterpoise correction and how to set up, modify and organize SIESTA run directories efficiently for this purpose. 2-component-dimer-on-surface-BSSE-corrected base directory is used as a concrete example. The layout has been proven to be useful in organizing the workflow when having to deal with multiple systems in a systematic manner. We start with a classic binding energy calculation in the first section, devoid of any corrections, and then add the necessary correction terms in the following section. In each we detail the input/output flow required for efficient determination of the interaction energy. Along the way we stop to consider in some detail the rationale behind the workflow.

The user should have some prior knowledge on how to navigate the Linux/UNIX environment, both locally and remotely, create, access and modify files and directories with basic instructions on the command line, and should be somewhat familiar with the vi text processing environment. No prior knowledge on how to run a SIESTA job is assumed. The first section goes through the basics of a SIESTA run pretty thoroughly, with the goal of guiding a new user to the software. Because of this, users familiar with the program might want to skim through this section, or skip it altogether, and

come back to it only if needed. That said, this is not meant to be a comprehensive guide on how to run SIESTA in general. Most of the parameters provided in SIESTA input files will not be referenced in this guide if they are not directly relevant. The user is encouraged to consult the SIESTA 4.0 manual, available online, for any issue falling outside the scope of this guide. The user is also encouraged to read David Sherrill’s concise paper, *Counterpoise Correction and Basis Set Superposition Error*, prior to using this guide. The source for BSSE and the rationale behind the correction is detailed in there, and will only partially be revisited in this guide.

The directory used with this guide can be found on the westgrid machine Cedar and can be copied by any westgrid user: The SIESTA submission scripts contained within directory are designed for use on Cedar.

```
ssh -X user@cedar.computecanada.ca
cp -r /home/etuca/BSSE_2comp_tmpl .
cd BSSE_2comp_tmpl/
ls
```

## A.2 Binding Energy in SIESTA

We start this tutorial with a classic calculation of the interaction energy on a 2-component system, namely a benzene molecule (1H) adsorbed on a gold surface (Au) periodic in the x and y directions:

```
cd 1_BE_1HAu_0base/
ls
```

To determine the interaction energy for this system, or in this case, the binding energy of 1H on Au, we need to run three separate SIESTA jobs, two for optimizing the geometries of the two isolated components, 1H and Au, and a third to optimize the geometry of the whole. The three directories 1\_1H, 2\_Au and 3\_1H\_Au correspond to these three SIESTA runs. Following all three optimization runs, the difference in the output energy of the system as a whole and the sum of energies of the two isolated components is what we consider the binding energy:

$$BE_{1H,Au} = E_{1H,Au} - (E_{1H} + E_{Au}) \quad (\text{A.1})$$

Let us now focus on how to run SIESTA optimizations on the two components

and the system and get the three terms in Equation (A.1).

### A.2.1 SIESTA Job Directory Details

Each SIESTA job directory must contain a number of files prior to execution. These are the SIESTA input file – file.fdf, a coordinate input file – coord.fdf, pseudopotential .psf files, and a script file to launch the job to the cluster scheduler, usually ending in .pbs. Provided that these are set up properly, SIESTA will write data in several other output files created and updated, during execution, inside the run directory.

Since we always run the risk of setting up the input files with erroneous parameters or with an improper input geometry, we should always preserve a clean version of the directory layout, that is, preserve the directory data in a state prior to SIESTA execution. We achieve this by having a clean subdirectory inside each run directory. The input files required for a run (and only those required) should be stored and set up in the clean subdirectory, and copied in the parent directory prior to submission. Assuming we are in the run directory, and have set up the input in the clean subdirectory, we can then copy the clean version and submit it to the cluster:

```
cp clean/* .
sbatch sc_script.pbs
<Job submitted>
```

This way we store the original input layout in the clean/ subdirectory, while SIESTA is writing data in the run directory. If we discover after the run that erroneous or improper input was submitted, we can go back to the clean version of the input, fix the issue and discard the erroneous output:

```
cd clean/
<make changes to input>
cd ../
rm *
cp clean/* .
sbatch sc_script.pbs
<Job submitted>
```

Had we not kept a clean version of the input, we would have had to manually delete the files outputted by SIESTA, being careful not to delete the input files in the process. Keeping output files from prior runs can cause SIESTA to import some

data from them instead of starting a job from scratch. This feature is useful in the case where prior jobs ended prematurely, before optimization, due to an error on the cluster or by reaching a maximum allotted run time, but it is undesirable when the input data itself is in error (recall the GIGO – garbage in, garbage out – principle). In the latter case, having the option to start out from scratch with a clean version and some minor tweaks to the input parameters can be a time saver. Figure A.1 shows the basic layout of a SIESTA job directory.

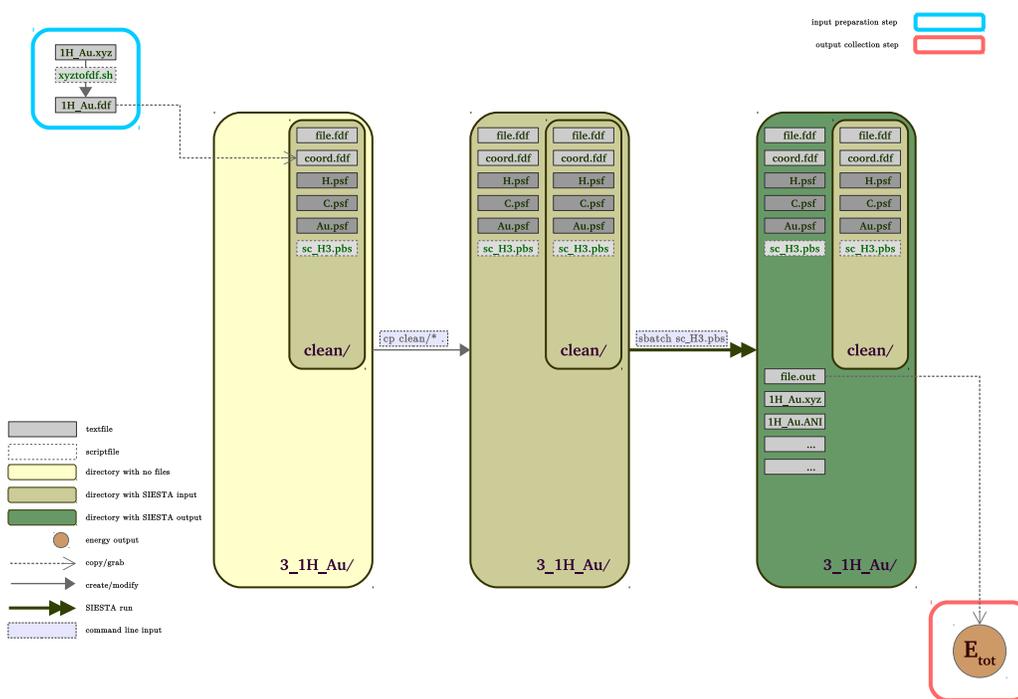


Figure A.1: Basic layout of a SIESTA job directory.

## A.2.2 SIESTA Input Details

The main input data read by SIESTA is stored in `file.fdf`. Here we specify what kind of simulation to run, what parameters should SIESTA use, and overwrite the default values for certain parameters, depending on our simulation goal. Most parameters take values in the form of a string, an integer, a real number, a real number followed by a unit or a boolean value in the form of “.true.” or “.false.”.

*SystemName* and *SystemLabel* should be given a descriptive name.

*NumberOfAtoms* must be set to the total number of atoms contained within the simulation box. *NumberOfSpecies* must be equal with the number of different atomic species in the system. For a benzene molecule adsorbed on a gold surface the number of species is 3: carbon, hydrogen and gold. Some simulations require two or more “kinds” of atoms of the same element, and the *NumberOfSpecies* value must reflect this: each kind of atom is a separate species, even if they might have the same atomic number. Later on, when we will correct for the basis set superposition error (BSSE), we might have both “real” and “ghost” atoms of one element within the same simulation box, and therefore having two different atomic species of the same element.

Parameter *MD.TypeOfRun* should be set to “CG” for a geometry relaxation run, and to “MD” for a molecular dynamics run. We will work exclusively with geometry relaxation runs, CG. We will have more to say about the difference between the two later on.

Other parameters are specified in *blocks*, such as the *%block ChemicalSpeciesLabel*, which defines the atomic species, or the *%block MM.Potentials*, which define the pair potential between the species. The input geometry too is read by default from *file.fdf*, from *%block AtomicCoordinatesAndAtomicSpecies*. We will, however, instruct SIESTA to read the input geometry from a separate file, *coord.fdf*, rather than from *file.fdf*. The *%block AtomicCoordinatesAndAtomicSpecies* is the section most likely to require improvement during the early stages of the setup. By keeping this block in a separate file we gain the flexibility to change the input geometry, without having to access *file.fdf* and risk making unwanted changes to the rest of the input.

Typically, we prepare the input geometry as an *.xyz* file, perhaps created in Avogadro, Molden, Crystal Maker, or some other geometry manipulation program. The *xyz* format stores the *xyz* coordinates of all atoms, one atom per line, beginning with the chemical identity of the atom. The order of atoms in the file does not usually matter to the *xyz* format. However, because the two components making up our system will be fed separately to jobs 1 and 2 (corresponding to the second and third terms in Equation (A.1) respectively), the input file must be partitioned based on these two components (atoms of 1H, followed by atoms of component Au). In general, if in our input geometry we put together more than two components that we might later want to isolate again, the same partitioning principle applies. As a best practice rule, in the input geometry, individual molecules should always be partitioned from one another,

even if the order of atoms within a molecule is not well established. Molecules should first be built or copied in their own individual xyz files, and then “fused” together in one final xyz file. The script “add\_xyz.sh” in 0\_util/ appends together xyz files, keeping the order of atoms the same as that of the file name arguments provided. If a surface is present, it should always be added as the last partition, and the surface itself, if possible, should further be partitioned into vertical layers (z position). This allows for easy block truncation of the component of interest, by knowing only the size of the components. SIESTA preserves the order of atoms from input to output.

In SIESTA the input geometry data must have the right format to be read prior to execution. Unlike an .xyz file, a coord.fdf file/block must have the atom identity represented by the numeric value defined in file.fdf, and must be placed in the fourth column, after the x, y and z positions. Transforming an .xyz file to the .fdf format can be done easily in vi. The visual block feature (Ctrl+V) allows the user to move the column with atom names from the first position to the last. Replacing the atom labels can be done with global commands. The following, for example: “:%s/Au/3/” replaces all occurrences of “Au” with “3” within the file. More systematically, a small sed or awk script can be written for this purpose. The script “xyztofdf.sh” in 0\_util/ is one such script which can be used to automatically construct .fdf files from .xyz files.

Next to the file.fdf and coord.fdf input files, the clean run directory should also contain the .psf (pseudopotential) files, with appropriate atomic labels, one file for each atomic species defined in file.fdf under the *%block ChemicalSpeciesLabel* (including separate copies for ghost atoms). Pseudopotential files for most elements can be downloaded from the SIESTA website.

Finally we need a .pbs script file to submit the job to the cluster. By convention the submission script file has a name starting in “sc\_” (“s” for “submit” and “c” for “cedar”). Here is where memory allocation, number of cores and maximum running time, i.e. machine dependent parameters, can be set and changed. If the maximum running time set in the script is reached, the job will be stopped. Most clusters have a maximum number of cores and a maximum amount of time allowed for one job. If the set values exceed the limits, the submitted jobs might stay in the queue indefinitely and never run. These limits can be found online on [www.westgrid.ca](http://www.westgrid.ca). Migrating SIESTA job directories from Cedar to another machine requires the replacement of the submission script file from each directory with a script file compatible with the new machine.

### A.2.3 Box Size and Geometry Constraints

SIESTA runs by default with periodic boundary conditions (PBC). For gas phase calculations, we can remove the interaction of the system with its adjacent images by making the box size large enough. The simulation box size is defined in `file.fdf` under *%block LatticeVectors* in vector format (having the values outside the main diagonal close to zero ensures us that we are dealing with a close to orthogonal box).

If the input parameter *MD.VariableCell* is set to `.true.` then SIESTA will relax the simulation box along with the input geometry. This is important to do when we prepare a system to have a periodic structure. In our example, the gold surface was optimized in advance to be periodic in the x and y directions, and the input lattice vectors for subsequent runs were set to the optimized box size. Any subsequent simulations with this surface that do not try to explicitly modify its symmetry or general shape should run with *MD.VariableCell* set to `.false.`, and with the set values for the x and y components of the box size (the z component can be modified, as we do not require periodicity in the z direction). Preparing a new xy-periodic box with another surface, whether of a different size, symmetry, or with new atoms altogether, will require a similar box size relaxation step, followed by a fixing of the x and y components of the box size in subsequent runs, once the shape, size and periodicity is confirmed. If we construct a new surface for the purpose of determining its binding energy with some molecule or complex, both the periodicity and the desired shape of the surface need to be confirmed in advance of binding. Moreover, the size of the surface should be large enough to make it impossible for the adsorbed molecule to interact with its image from adjacent cells, but not larger. The time required to complete a geometry relaxation grows fast with the number of atoms present in the simulation box.

Different jobs designed to produce comparable output data (such as the energy terms in Equation (A.1)) should have identical box sizes. If a surface is involved, the size of the periodic surface should set the simulation box size for all runs, at least in x and y. In our example, the gas phase simulation of 1H has the same box size as 1H adsorbed on Au, and both have inherited the values in *%block LatticeVectors* from Au.

SIESTA provides us with the option to fix certain atoms, preventing them from moving during CG steps. We specify which atoms we want fixed in `file.fdf` under the *%block GeometryConstraints*. The numbers specified here should match the order in

which the atoms appear in `coord.fdf` file. In our example, we make use of geometry constraints to maintain the bulk structure of the gold surface, while allowing for some flexibility in the top layers. The top two layers can respond structurally and reorganize under the influence of adsorbed benzene, while we fix the geometry of atoms in the bottom-most third layer. In effect, we tell SIESTA to mimic the rigidity of a bulk geometry, without having to impose absolute rigidity in the surface. Because the surface was partitioned in advance based on layers (the value of  $z$ ), we can easily provide this constraint request to SIESTA as a range of atoms. The surface consists of 168 atoms in total, the first 56 having the lowest  $z$  value, forming the bottom-most layer. In all SIESTA runs with a geometry, atoms 1 to 56 of the surface are kept fixed:

```
grep "position" 3_1H_Au/clean/file.fdf
position from 13 to 68
```

In the example above, the first 12 atoms represent benzene, which we do not wish to fix. The next 56 atoms, starting at 13 and ending at 68, are the bottom layer of the surface, which SIESTA will now know not to move.

#### A.2.4 The Basics of a SIESTA Geometry Relaxation Run

The parameter *MD.TypeOfRun* is set to “CG”, instructing SIESTA to perform a geometry relaxation of the system, by sampling the geometric space in search of a good minimum. This type of simulation is in contrast with the “MD”, molecular dynamics type, which would perform a time evolution of the input geometry. The biggest advantage of a CG run over an MD run is that the former achieves convergence much faster than the latter. Moreover, above a size of a few dozens of atoms MD becomes prohibitively slow, whereas CG relaxations can easily be performed on systems with hundreds of atoms. This is why we opt for CG runs, keeping in mind that the evolution of our system throughout the simulation does not reflect an actual time evolution (as would be the case with an MD run), but a search for a local energetic minimum in the vicinity of our starting geometry. With each CG step, a new atomic geometry is considered, by modifying and combining the geometries from previous CG steps (or input geometry) according to some criterion. Then the electronic structure of the new geometry is relaxed through scf cycles, leading to the minimum electronic energy for that geometry. CG steps are performed iteratively until some convergence criterion is met, at which point SIESTA decides to end the simulation.

*MD.NumCGsteps* sets a maximum limit to the number of CG steps performed. If convergence is achieved prior to reaching this number, SIESTA ends the simulation normally. If this number is reached, SIESTA ends the simulation even if convergence hasn't been achieved. Resubmitting the run will cause SIESTA to resume the relaxation, starting with the geometry from the last CG step, assuming that the output files haven't been removed and the parameter *UseSaveData* in the input file is set to `.true.` prior to resubmission. A properly created input geometry should lead to convergence within a couple of hundreds of CG steps. If the CG step exceeds 1000, the input geometry can likely be improved.

### A.2.5 Submitting Jobs and Checking Output

Now that we've gone through the basics of SIESTA input run and the geometry relaxation procedure, we return to Equation (A.1) and recall that we need to submit three jobs, one for each term in the equation:

```

cd 1_1H/
cp clean/* .
sbatch sc_H12.pbs
<Job submitted>
cd ../2_Au/
cp clean/* .
sbatch sc_H12.pbs
<Job submitted>
cd ../3_1H_Au/
cp clean/* .
sbatch sc_H3.pbs
<Job submitted>

```

We can check the status of jobs submitted:

```
squeue -u <username>
```

“R” in the ST column indicates that a job is currently running and “PD” means that the job is currently waiting in the queue. When a job ends it disappears from the queue list.

SIESTA writes output to various files as it runs, the main one being `file.out`. It also prints the output geometry in an `.xyz` file which can be readily grabbed to check

the relaxed geometry. If *LongOutput* is set to “.true.”, an animation file, .ANI will also be outputted. This is essentially the ordered collection of xyz coordinates for each CG step, all appended in one file, and can be used to visually track the geometry relaxation throughout the simulation steps. Recall that this does not represent a time evolution of the system for a CG run. The Molden visualization program supports the .ANI format.

If the SIESTA run is completed and there were no errors, the last line of file.out will show an “End of run” message:

```
tail -1 file.out
>> End of run:    8-MAY-2018   20:12:42
```

We get the energy of the relaxed system by searching for the last occurrence of “Etot”. In the vi environment we go to the end of the file by pressing Shift+g, then search backwards by typing ?Etot and Enter:

```
siesta: Program's energy decomposition (eV):
siesta: Ebs      =    -16435.722118
siesta: Eions    =    184821.183713
siesta: Ena      =     5519.144383
siesta: Ekin     =    109818.445951
siesta: Enl      =    -57901.434614
siesta: DEna     =     481.376197
siesta: DUscf    =     24.916584
siesta: DUext    =     0.000000
siesta: Exc      =    -36379.559107
siesta: eta*DQ   =     0.000000
siesta: Emadel   =     0.000000
siesta: Emeta    =     0.000000
siesta: Emolmec  =    -51.281390
siesta: Ekinion  =     0.000000
siesta: Eharris  =   -163309.575711
siesta: Etot     =   -163309.575711
siesta: FreeEng  =   -163309.731959
```

Here we see various energetic components and can check that all of them have reasonable values. If some components are out of bound, displaying “\*\*\*\*\*” instead of properly formatted numbers, that might be a sign of trouble with the input geometry

or parameters. The value we need for binding energy is on the “Etot” line. We could also grab this line directly from file, bypass vi, and print it to the prompt:

```
grep "Etot" file.out | tail -1
siesta: Etot      =    -163309.575711
```

From here we grab the energy value (sign included) and plug it in the appropriate term in Equation (A.1). We do this for all completed jobs to get the desired binding energy.

Figure A.2 summarizes schematically the workflow described for getting the binding energy with SIESTA jobs, from the input preparation step, all the way up to the gathering of output.

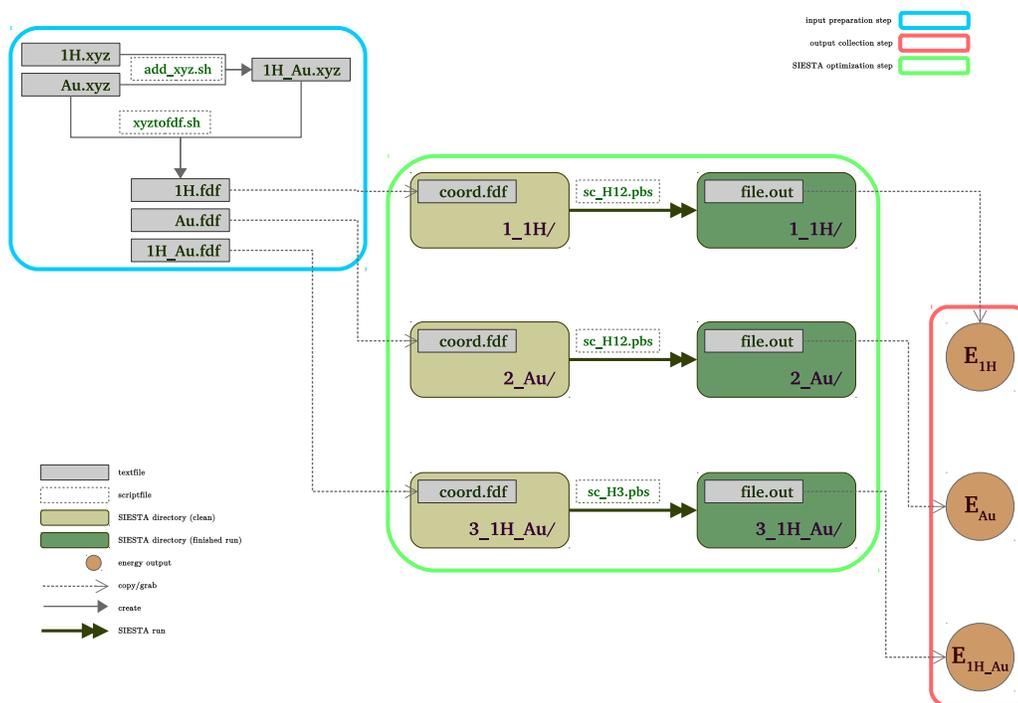


Figure A.2: Schematic representation of the workflow required to obtain the binding energy of benzene (1H) on gold with SIESTA.

## A.3 Interaction Energy and the Counterpoise Correction in SIESTA

### A.3.1 The Source of the Basis Set Superposition Error

The binding energy of benzene adsorbed on gold we got with Equation (A.1) is likely overestimated. In general, for a 2-component system,  $A$  and  $B$ , the interaction energy obtained as

$$\Delta\mathbf{E}(AB) = \mathbf{E}_{\text{geom}\langle AB \rangle}^{\text{base}\langle AB \rangle}(AB) - \left( \mathbf{E}_{\text{geom}\langle A \rangle}^{\text{base}\langle A \rangle}(A) + \mathbf{E}_{\text{geom}\langle B \rangle}^{\text{base}\langle B \rangle}(B) \right) \quad (\text{A.2})$$

will be artificially strengthened. In Equation (A.2) we specify, for each energy term, both the geometry and the basis set under which the system was relaxed.  $\mathbf{E}_{\text{geom}\langle AB \rangle}^{\text{base}\langle AB \rangle}(AB)$  is the energy of components  $A$  and  $B$ , having a relaxed geometry as a whole, in the basis set of both  $A$  and  $B$ .  $\mathbf{E}_{\text{geom}\langle B \rangle}^{\text{base}\langle B \rangle}(B)$  is the energy of component  $B$ , with its geometry relaxed in the absence of  $A$ , in the basis set of  $B$  only. The reason for specifying the geometry and basis set for each term will become apparent shortly.

According to Sherrill (see the Counterpoise Correction and Basis Set Superposition Error document), the cause for overestimation in the interaction energy as determined by Equation (A.2) is due to the inconsistent treatment of the basis set for each component. Taken together, each component can access the other component basis set to further stabilize itself, whereas in isolation, the additional basis set is absent. This leads to the first energy term in Equation (A.2) to be more negative, without a counteracting energy stabilization in subsequent terms. To account for this discrepancy, we need to find out how much each component is stabilized by the presence of the other component basis set:

$$\begin{aligned} \mathbf{E}_{\text{corr}}(A) &= \mathbf{E}_{\text{geom}\langle AB \rangle}^{\text{base}\langle AB \rangle}(A) - \mathbf{E}_{\text{geom}\langle A \rangle}^{\text{base}\langle A \rangle}(A) \\ \mathbf{E}_{\text{corr}}(B) &= \mathbf{E}_{\text{geom}\langle AB \rangle}^{\text{base}\langle AB \rangle}(B) - \mathbf{E}_{\text{geom}\langle B \rangle}^{\text{base}\langle B \rangle}(B) \end{aligned} \quad (\text{A.3})$$

The first right-handed term in each equation in (A.3) corresponds to the energy of each component, either  $A$  or  $B$ , with the geometry relaxed in the presence of the other component, and access to the basis functions of the other component. Therefore, the terms in each equation correspond to different geometries (that in the presence and

that in the absence of the other component, respectively), even if they represent the same component. This means that our attempt at correction captured not only the BSSE, but the corresponding deformation energy as well:

$$\begin{aligned}\mathbf{E}_{\text{corr}}(A) &= \mathbf{E}_{\text{BSSE}}(A) + \mathbf{E}_{\text{def}}(A) \\ \mathbf{E}_{\text{corr}}(B) &= \mathbf{E}_{\text{BSSE}}(B) + \mathbf{E}_{\text{def}}(B)\end{aligned}\tag{A.4}$$

Now we can isolate our sought after BSSE corrections by combining Equations (A.3) and (A.4):

$$\begin{aligned}\mathbf{E}_{\text{BSSE}}(A) &= \mathbf{E}_{\text{geom}\langle AB \rangle}^{\text{base}\langle AB \rangle}(A) - \mathbf{E}_{\text{geom}\langle A \rangle}^{\text{base}\langle A \rangle}(A) - \mathbf{E}_{\text{def}}(A) \\ \mathbf{E}_{\text{BSSE}}(B) &= \mathbf{E}_{\text{geom}\langle AB \rangle}^{\text{base}\langle AB \rangle}(B) - \mathbf{E}_{\text{geom}\langle B \rangle}^{\text{base}\langle B \rangle}(B) - \mathbf{E}_{\text{def}}(B)\end{aligned}\tag{A.5}$$

The deformation energy of  $A$  is the difference between the energy of  $A$ , having the geometry adopted in the presence of  $B$ , and the energy of  $A$ , with the geometry in isolation:

$$\begin{aligned}\mathbf{E}_{\text{def}}(A) &= \mathbf{E}_{\text{geom}\langle AB \rangle}^{\text{base}\langle A \rangle}(A) - \mathbf{E}_{\text{geom}\langle A \rangle}^{\text{base}\langle A \rangle}(A) \\ \mathbf{E}_{\text{def}}(B) &= \mathbf{E}_{\text{geom}\langle AB \rangle}^{\text{base}\langle B \rangle}(B) - \mathbf{E}_{\text{geom}\langle B \rangle}^{\text{base}\langle B \rangle}(B)\end{aligned}\tag{A.6}$$

In general, we cannot ignore the deformation of either component in the presence of the other, or assume that the the corresponding deformation energy is negligible. If we substitute Equations (A.6) in (A.5) and cancel out the terms we get our final form for the BSSE correction:

$$\begin{aligned}\mathbf{E}_{\text{BSSE}}(A) &= \mathbf{E}_{\text{geom}\langle AB \rangle}^{\text{base}\langle AB \rangle}(A) - \mathbf{E}_{\text{geom}\langle AB \rangle}^{\text{base}\langle A \rangle}(A) \\ \mathbf{E}_{\text{BSSE}}(B) &= \mathbf{E}_{\text{geom}\langle AB \rangle}^{\text{base}\langle AB \rangle}(B) - \mathbf{E}_{\text{geom}\langle AB \rangle}^{\text{base}\langle B \rangle}(B)\end{aligned}\tag{A.7}$$

We now subtract the BSSE error from our original estimation of the interaction energy:

$$\Delta\mathbf{E}_{\text{BSSE}}(AB) = \Delta\mathbf{E}(AB) - \mathbf{E}_{\text{BSSE}}(A) - \mathbf{E}_{\text{BSSE}}(B)\tag{A.8}$$

and expand (A.2) and (A.7) in (A.8) to get the corrected interaction energy, term by

term:

$$\begin{aligned} \Delta E_{\text{BSSE}}(AB) = & \mathbf{E}_{\text{geom}\langle AB \rangle}^{\text{base}\langle AB \rangle}(AB) - \left( \mathbf{E}_{\text{geom}\langle A \rangle}^{\text{base}\langle A \rangle}(A) + \mathbf{E}_{\text{geom}\langle B \rangle}^{\text{base}\langle B \rangle}(B) \right) \\ & - \mathbf{E}_{\text{geom}\langle AB \rangle}^{\text{base}\langle AB \rangle}(A) + \mathbf{E}_{\text{geom}\langle AB \rangle}^{\text{base}\langle A \rangle}(A) - \mathbf{E}_{\text{geom}\langle AB \rangle}^{\text{base}\langle AB \rangle}(B) + \mathbf{E}_{\text{geom}\langle AB \rangle}^{\text{base}\langle B \rangle}(B) \quad (\text{A.9}) \end{aligned}$$

Equation (A.9) will serve as the basis for how to set up SIESTA runs for calculating the interaction energy of a 2-component system with counterpoise correction.

### A.3.2 Seven Energy Terms

In the last section we have extended our classic definition of interaction energy from one containing only three terms to one with seven terms. Requiring seven simulations only for a 2-component system may seem like too much. The good news is that we can obtain each one of the seven terms in Equation (A.9) through a separate SIESTA run, in the tradition of the classic approach, where we had only three terms. The bad news is that the input flow of geometry data to and from SIESTA becomes somewhat more complicated. The reason for this is elucidated once we note that all four correction terms in Equation (A.9) have the qualifier  $\text{geom}\langle AB \rangle$ . This means that the input geometry for all four terms has to come from the geometry of the system relaxed as a whole, the output from the third SIESTA run, in our previous example. The first three terms in Equation (A.9) are obtained, as before, by following the workflow in Figure A.2. With the geometric data from the third run we now have to follow an additional step with four of SIESTA runs. The workflow is now updated in Figure A.3 to reflect this additional step. We will refer to the first step with the original three jobs as the SIESTA optimization (opt) step, and to the latter as the SIESTA correction (corr) step.

### A.3.3 From Optimization Step to Correction Step

The directory `2_BE_BSSE_1HAu_1opt` contains the updated version of our previous working directory for the purpose of determining the counterpoise corrected binding energy of benzene on gold surface. The seven subdirectories there are each a SIESTA run directory, one for each term in Equation (A.9). `1_A`, `2_B` and `3_A_B` correspond to the original three job directories, and to the original three terms of the equation. Following the job submission, `1_A` and `2_B` will output the energy of the isolated

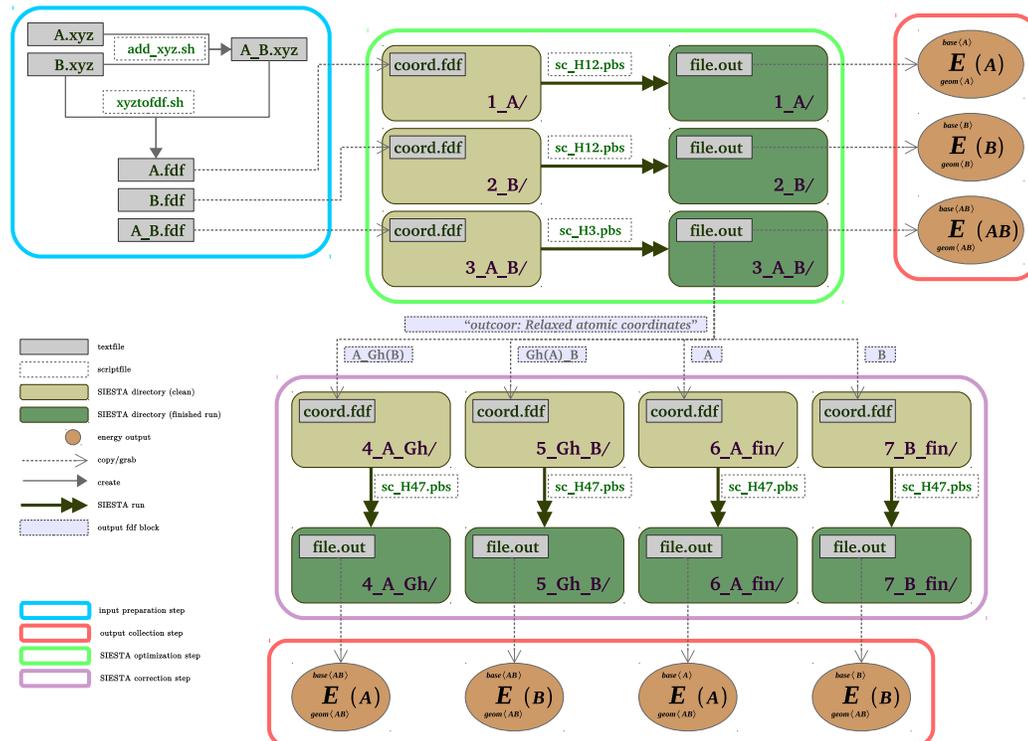


Figure A.3: Schematic representation of the workflow required to obtain the counterpoise-corrected interaction energy of a two-component system with SIESTA.

sub-components, and `3_A_B` will output the energy of the system as a whole. All three perform geometry relaxation of the structures considered prior to outputting the energy, which is why they require as input unconstrained geometries (except for any constraints required for preserving the structure of the surface). As they are, all three already contain the output from their respective SIESTA geometry relaxation run, their relaxed geometry and energy terms are already available in `file.out`. The following four run directories are all in a clean state, with no input geometry provided (`coord.fdf` files are empty). With this setup of `2_BE_BSSE.1HAu.1opt` we have therefore passed the SIESTA optimization step and have to set up the input for the SIESTA correction step.

Directories `4_A_Gh`, `5_Gh_B`, `6_A_fin` and `7_B_fin` correspond each to terms 4, 6, 5 and 7 of Equation (A.9). All four take as input the geometry of one component, as it was relaxed in `3_A_B`, and without modifying it further, calculate its energy, either in that's own component basis set, or in the basis set of the system as a whole.

Terms 4 and 5, namely  $\mathbf{E}_{\text{geom}\langle AB \rangle}^{\text{base}\langle AB \rangle}(A)$  and  $\mathbf{E}_{\text{geom}\langle AB \rangle}^{\text{base}\langle A \rangle}(A)$  both correspond to the energy of component  $A$  with bounded geometry, but the former is evaluated within the basis set of the system as a whole ( $\text{base}\langle AB \rangle$ ), whereas the latter is evaluated within the basis set of  $A$  only ( $\text{base}\langle A \rangle$ ). The same is true for terms 6 and 7, for  $B$ . To get terms 5 and 7, we isolate the respective component from the optimized output of the whole system, fix all atoms in place to prevent reoptimization and run a single-point calculation with SIESTA. The resulting energy is our desired term.

To get terms 4 and 6, a similar fixed-geometry single-point calculation is required, but with the addition of the basis set from the other component. For  $\mathbf{E}_{\text{geom}\langle AB \rangle}^{\text{base}\langle AB \rangle}(A)$ , we need to add to  $A$  all the basis functions of  $B$  on the atomic centers of  $B$ , while removing all electrons and nuclear charges of  $B$ . Similarly, we get  $\mathbf{E}_{\text{geom}\langle AB \rangle}^{\text{base}\langle AB \rangle}(B)$  with a single-point calculation of  $B$ , in the presence of the basis functions of  $A$ , positioned at the appropriate atomic centers of  $A$ , but stripped of all electrons and nuclear charges in  $A$ . The term for these basis functions stripped of all electrons and nuclei is “ghost atoms”. SIESTA accepts ghost atoms as input to `coord.fdf`, if they are defined in `%block ChemicalSpeciesLabel` as having a negative atomic numbers:

```
sed -n '/%block Chem/,/%endblock Chem/p' 5_Gh_B/clean/file.fdf
%block ChemicalSpeciesLabel
1  -1  H      # Species index, atomic number, species label
2  -6  C      # Species index, atomic number, species label
3   79 Au    # Species index, atomic number, species label
%endblock ChemicalSpeciesLabel
```

Here the hydrogen and carbon species are both defined as ghost atoms, and their appearance in the input geometry will only serve as basis functions for the real Au atoms to use, in addition to the basis functions of their own, during the electronic optimization at the fixed geometry. If both the real and ghost components share atomic species, each shared species must be defined twice, once as the real version, with a positive atomic number, and again as the ghost, with a negative atomic number:

```
%block ChemicalSpeciesLabel
1   1  H      # Species index, atomic number, species label
2   6  C      # Species index, atomic number, species label
3  -1  gH
4  -6  gC
```

```
%endblock ChemicalSpeciesLabel
```

Here the index for the real carbon atoms is 2, and that for the ghost carbon atoms is 4. The coord.fdf input should reflect this distinction. The run directory should contain four psf files: H.psf, gH.psf C.psf and gC.psf. Ghost psf files should be identical copies of their real counterparts.

To get terms 4 and 6 in Equation (A.9) we get the full optimized geometry of the system, fix all atoms in place, turn the other component to ghost, by changing their index accordingly, then run a single point calculation energy with SIESTA. The resulting energy is our desired term.

We have thus obtained all four correction terms and can plug them all in Equation (A.9) to get the counterpoise-corrected interaction energy.

### A.3.4 Input-Output Flow

Let us now reiterate the SIESTA correction step, with an emphasis on the workflow. Once runs 1\_A, 2\_B, 3\_A\_B have finished and the xyz output looks reasonable, the last "Relaxed atomic coordinates" block in file.out from 3\_A\_B serves as the input for runs 4 through 7 (in vi: "G"+"?Re" would do). These are the same coordinates printed in the output xyz file, but here they are already in fdf format and are easy to work with in vi (starting from the top-left corner of the block, in Visual Block mode, "45l1j" followed by "y" will select the benzene component and copy it to the clipboard).

4\_A\_Gh and 5\_Gh\_B require the entire block from 3\_A\_B in their coord.fdf input. They require that the ghost component (the B block for 4\_A\_Gh, the A block for 5\_Gh\_B) has its atom index matching the identity of ghost atoms, as they are defined in *%block ChemicalSpeciesLabel* in file.fdf. SIESTA treats a chemical species as ghost if a negative sign is preceding the atomic number. If *A* and *B* have atom species in common, then both real and ghost versions of those species have to be declared in file.fdf. Let's assume that the following block corresponds to a gas-phase benzene dimer, that is, components *A* and *B* are a benzene molecule, 12 atom size each:

```
%block ChemicalSpeciesLabel
1  1  H      # Species index, atomic number, species label
2  6  C      # Species index, atomic number, species label
3  -1 gH
4  -6 gC
```

```
%endblock ChemicalSpeciesLabel
```

If 4\_A\_Gh/coord.fdf is filled with the output from 3\_A\_B, then the vi command `":13,24s/2$/4/"` changes all carbon atoms to ghost carbon atoms within the second molecule, keeping the first molecule (lines 1 to 12) intact. `"13,24s/1$/3/"` does the same for H. For 5\_B\_Gh, the same input should be modified with `":1,12s/2$/4"`, `":1,12s/1$/3"`, turning the first component (lines 1 to 12) into ghosts, keeping the second unchanged.

```
cd 4_A_Gh/
cp clean/* .
sbatch sc_H47.pbs
<Job submitted>
cd ../5_Gh_B/
cp clean/* .
sbatch sc_H47.pbs
<Job submitted>
```

The runs in 4\_A\_Gh and 5\_Gh\_B provide the energy of the two sub-components A and B respectively, in the presence of ghost atoms of the other component.

Finally, 6\_A\_fin and 7\_B\_fin require as input only the truncated components from 3\_A\_B. Their NumberOfAtoms, and ChemicalSpeciesLabel should match exactly those of 1\_A and 2\_B respectively. It's easy to neglect a possible mismatch in ChemicalSpeciesLabel between 3\_A\_B and 6/7, which may lead in error.

```
cd ../6_A_fin/
cp clean/* .
sbatch sc_H47.pbs
<Job submitted>
cd ../7_B_fin/
cp clean/* .
sbatch sc_H47.pbs
<Job submitted>
```

The runs in 6\_A\_fin and 7\_B\_fin provide the energy of the two sub-components in isolation, but each constrained to the geometry of the system as a whole. The energetic difference between 1\_A and 6\_A\_fin, and that between 2\_B and 7\_B\_fin are measures of the deformation of components A and B, respectively, in the process of binding. In

all four correction runs, the geometry used should be that outputted by 3\_A\_B and all atoms should be constrained to prevent reoptimization.

### A.3.5 Checking Output

After the correction phase is complete, the directory containing the seven SIESTA job subdirectories should have error-free output:

```

for file in [1-7]*/file.out; do tail -1 $file; done
>> End of run:   8-MAY-2018   19:04:29
>> End of run:   8-MAY-2018   19:05:54
>> End of run:   8-MAY-2018   20:12:42
>> End of run:  17-MAY-2018    9:55:42
>> End of run:  17-MAY-2018    9:55:40
>> End of run:  17-MAY-2018    9:54:53
>> End of run:  17-MAY-2018    9:55:42

```

We can then grab all the energy terms:

```

for file in [1-7]*/file.out; do grep "Etot" $file | tail -1; done
siesta: Etot    =    -1068.897374
siesta: Etot    =   -162239.898212
siesta: Etot    =   -163309.575711
siesta: Etot    =   -1068.924323
siesta: Etot    =   -162240.129727
siesta: Etot    =   -1068.899206
siesta: Etot    =   -162239.882692

```

printed here in the numerical order, based on the names of our sub-directories.

To summarize this section, jobs 1 to 3 take the provided input geometry and perform optimizations of the two sub-components and the system as a whole, outputting both the final geometries and energies. The optimization process for these runs can span many CG steps, and the runs can take significant time to converge. The optimized geometry of 3\_A\_B is then manually, (or automatically with the help of a script) fed into jobs 4-7 as input, either as a whole (to 4\_A\_Gh, 5\_Gh\_B) or truncated to the appropriate subcomponent (to 6\_A\_fin, 7\_B\_fin). These last four jobs are each

a single point calculation, performed only to get the energy value of the fully constrained geometry. Because of their single-point nature, these jobs complete very fast once they make it through the queue.

## Appendix B

# Structural Details for Optimized Helicene Molecules

Table B.1: Numerical values for angles and distances between rings for penta-, hexa- and heptahelicene monomers and dimers optimized in the absence of a surface. Letters **L**, **U**, **b** and **t** stand for homochiral/“like”, heterochiral/“unlike”, “bottom” and “top” respectively. The letters **a** and **d** refer to “angle” and “distance” respectively. First half of the table represents angles in degrees, second half represents distances in Å.

<b>a</b>	<b>5H</b>	<b>L-5H</b>	<b>U-5H</b>	<b>6H</b>	<b>L-6H</b>	<b>U-6H</b>	<b>7H</b>	<b>L-7H</b>	<b>U-7H</b>
<b>b1</b>	0	0	0	0	0	0	0	0	0
<b>b2</b>	9	5	9	12	10	16	12	5	6
<b>b3</b>	18	14	23	27	18	22	24	16	18
<b>b4</b>	34	30	31	36	29	28	31	26	22
<b>b5</b>	44	40	33	42	40	36	26	26	24
<b>b6</b>	–	–	–	46	49	40	18	22	24
<b>b7</b>	–	–	–	–	–	–	15	16	19
<b>t1</b>	–	3	21	–	5	56	–	27	16
<b>t2</b>	–	17	26	–	13	49	–	25	23
<b>t3</b>	–	25	32	–	26	36	–	23	21
<b>t4</b>	–	32	29	–	34	38	–	19	25
<b>t5</b>	–	38	26	–	40	44	–	9	23
<b>t6</b>	–	–	–	–	44	50	–	9	14
<b>t7</b>	–	–	–	–	–	–	–	20	4
<b>d</b>	<b>5H</b>	<b>L-5H</b>	<b>U-5H</b>	<b>6H</b>	<b>L-6H</b>	<b>U-6H</b>	<b>7H</b>	<b>L-7H</b>	<b>U-7H</b>
<b>b1</b>	0	0	0	0	0	0	0	0	0
<b>b2</b>	2.5	2.5	2.5	2.5	2.5	2.5	2.5	2.5	2.5
<b>b3</b>	4.3	4.4	4.3	4.3	4.3	4.3	4.3	4.3	4.3
<b>b4</b>	5.2	5.3	5.2	5.2	5.2	5.2	5.1	5.1	5.1
<b>b5</b>	5.0	5.1	4.9	5.0	5.1	5.0	4.7	4.7	4.6
<b>b6</b>	–	–	–	4.3	4.4	4.0	3.6	3.6	3.6
<b>b7</b>	–	–	–	–	–	–	3.3	3.3	3.4
<b>t1</b>	–	3.8	5.5	–	4.7	4.8	–	7.4	6.6
<b>t2</b>	–	5.4	4.2	–	6.5	5.8	–	7.6	6.6
<b>t3</b>	–	6.6	4.0	–	7.4	6.5	–	6.9	6.7
<b>t4</b>	–	6.9	5.6	–	7.1	6.6	–	6.3	7.3
<b>t5</b>	–	6.6	7.2	–	6.4	7.0	–	6.9	8.4
<b>t6</b>	–	–	–	–	6.5	7.6	–	8.4	9.2
<b>t7</b>	–	–	–	–	–	–	–	9.7	9.6

Table B.2: Angles and distances made by pentahelicene monomer, homochiral dimer and heterochiral dimer with the Ag, Au and Cu surfaces. Molecules are treated as collections of five ortho-fused rings, labelled **b1** – **b5**, **t1** – **t5**. Letters **b** and **t** refer to the “bottom” (closest to the surface) and “top” (furthest from the surface) molecule respectively. The letter **s** refers to the top surface plane, taken as a reference. Letters **M**, **L** and **U** refer to monomer, homochiral/“like” and heterochiral/“unlike”, respectively. The letters **a** and **d** refer to “angle” and “distance” respectively. First half of the table represents angles in degrees, second half represents distances in Å.

<b>a</b>	<b>M-Ag</b>	<b>L-Ag</b>	<b>U-Ag</b>	<b>M-Au</b>	<b>L-Au</b>	<b>U-Au</b>	<b>M-Cu</b>	<b>L-Cu</b>	<b>U-Cu</b>
<b>s</b>	0	0	0	0	0	0	0	0	0
<b>b1</b>	8	3	4	6	5	4	6	1	5
<b>b2</b>	2	3	6	0	6	6	3	5	11
<b>b3</b>	4	11	13	5	15	12	1	11	20
<b>b4</b>	8	23	24	16	26	22	5	22	33
<b>b5</b>	30	34	35	34	38	30	24	35	52
<b>t1</b>	–	4	17	–	1	13	–	4	23
<b>t2</b>	–	13	23	–	14	20	–	12	33
<b>t3</b>	–	22	29	–	23	29	–	21	40
<b>t4</b>	–	29	30	–	31	33	–	29	46
<b>t5</b>	–	38	33	–	40	38	–	39	47
<b>d</b>	<b>M-Ag</b>	<b>L-Ag</b>	<b>U-Ag</b>	<b>M-Au</b>	<b>L-Au</b>	<b>U-Au</b>	<b>M-Cu</b>	<b>L-Cu</b>	<b>U-Cu</b>
<b>s</b>	0	0	0	0	0	0	0	0	0
<b>b1</b>	2.6	2.6	2.7	2.9	2.9	2.9	2.3	2.2	2.5
<b>b2</b>	2.7	2.7	2.7	3.0	2.9	2.9	2.4	2.2	2.5
<b>b3</b>	2.7	2.7	2.7	3.0	3.0	3.0	2.5	2.3	2.7
<b>b4</b>	2.8	3.0	3.0	3.2	3.4	3.2	2.5	2.7	3.5
<b>b5</b>	3.5	4.2	4.1	4.2	4.7	4.1	3.0	2.9	5.1
<b>t1</b>	–	5.6	5.6	–	5.9	5.9	–	5.1	5.4
<b>t2</b>	–	5.5	5.9	–	5.7	6.1	–	5.1	5.6
<b>t3</b>	–	5.3	6.7	–	5.5	6.6	–	5.0	6.6
<b>t4</b>	–	5.8	7.9	–	6.0	7.9	–	5.6	8.3
<b>t5</b>	–	7.0	9.0	–	7.4	9.2	–	6.9	9.5

Table B.3: Angles and distances made by hexahelicene monomer, homochiral dimer and heterochiral dimer with the Ag, Au and Cu surfaces. Molecules are treated as collections of six ortho-fused rings, labelled **b1** – **b6**, **t1** – **t6**. Letters **b** and **t** refer to the “bottom” (closest to the surface) and “top” (furthest from the surface) molecule respectively. The letter **s** refers to the top surface plane, taken as a reference. Letters **M**, **L** and **U** refer to monomer, homochiral/“like” and heterochiral/“unlike”, respectively. The letters **a** and **d** refer to “angle” and “distance” respectively. First half of the table represents angles in degrees, second half represents distances in Å.

<b>a</b>	<b>M-Ag</b>	<b>L-Ag</b>	<b>U-Ag</b>	<b>M-Au</b>	<b>L-Au</b>	<b>U-Au</b>	<b>M-Cu</b>	<b>L-Cu</b>	<b>U-Cu</b>
<b>s</b>	0	0	0	0	0	0	0	0	0
<b>b1</b>	3	4	4	5	4	5	2	0	4
<b>b2</b>	5	3	7	5	4	8	5	4	4
<b>b3</b>	12	10	14	12	12	17	5	8	6
<b>b4</b>	25	20	29	25	23	30	18	17	20
<b>b5</b>	38	31	38	37	34	38	29	27	31
<b>b6</b>	45	44	43	45	46	42	33	38	38
<b>t1</b>	–	2	57	–	2	60	–	7	54
<b>t2</b>	–	18	50	–	16	52	–	17	45
<b>t3</b>	–	28	38	–	26	39	–	23	35
<b>t4</b>	–	33	42	–	33	43	–	29	40
<b>t5</b>	–	39	49	–	39	50	–	36	45
<b>t6</b>	–	45	54	–	44	54	–	41	51
<b>d</b>	<b>M-Ag</b>	<b>L-Ag</b>	<b>U-Ag</b>	<b>M-Au</b>	<b>L-Au</b>	<b>U-Au</b>	<b>M-Cu</b>	<b>L-Cu</b>	<b>U-Cu</b>
<b>s</b>	0	0	0	0	0	0	0	0	0
<b>b1</b>	2.7	2.6	2.7	2.8	2.9	2.9	2.2	2.2	2.3
<b>b2</b>	2.7	2.7	2.8	3.0	3.0	3.0	2.4	2.3	2.4
<b>b3</b>	2.8	2.7	2.8	3.0	3.0	3.0	2.6	2.4	2.5
<b>b4</b>	3.1	2.9	3.2	3.2	3.2	3.5	2.6	2.5	2.5
<b>b5</b>	4.2	3.8	4.4	4.4	4.2	4.8	3.1	3.2	3.2
<b>b6</b>	5.8	5.3	5.9	5.9	5.7	6.2	4.4	4.6	4.7
<b>t1</b>	–	5.6	7.0	–	5.9	7.3	–	5.2	6.7
<b>t2</b>	–	5.3	8.8	–	5.6	9.2	–	5.0	8.2
<b>t3</b>	–	5.1	8.9	–	5.4	9.3	–	5.0	8.1
<b>t4</b>	–	5.6	7.4	–	5.9	7.8	–	5.5	6.7
<b>t5</b>	–	6.9	6.4	–	7.1	6.7	–	6.7	6.0
<b>t6</b>	–	8.5	7.3	–	8.7	7.5	–	8.2	7.0

Table B.4: Angles and distances made by heptahelicene monomer, homochiral dimer and heterochiral dimer with the Ag, Au and Cu surfaces. Molecules are treated as collections of seven ortho-fused rings, labelled **b1** – **b7**, **t1** – **t7**. Letters **b** and **t** refer to the “bottom” (closest to the surface) and “top” (furthest from the surface) molecule respectively. The letter **s** refers to the top surface plane, taken as a reference. Letters **M**, **L** and **U** refer to monomer, homochiral/“like” and heterochiral/“unlike”, respectively. The letters **a** and **d** refer to “angle” and “distance” respectively. First half of the table represents angles in degrees, second half represents distances in Å.

<b>a</b>	<b>M-Ag</b>	<b>L-Ag</b>	<b>U-Ag</b>	<b>M-Au</b>	<b>L-Au</b>	<b>U-Au</b>	<b>M-Cu</b>	<b>L-Cu</b>	<b>U-Cu</b>
<b>s</b>	0	0	0	0	0	0	0	0	0
<b>b1</b>	8	7	4	8	5	3	4	9	1
<b>b2</b>	2	4	1	1	0	2	2	5	3
<b>b3</b>	9	5	4	10	8	7	9	2	6
<b>b4</b>	23	16	15	24	23	20	22	5	18
<b>b5</b>	31	30	36	31	30	34	28	25	43
<b>b6</b>	26	27	39	26	28	38	21	33	48
<b>b7</b>	20	22	37	20	22	38	14	29	47
<b>t1</b>	–	27	37	–	28	36	–	34	45
<b>t2</b>	–	26	31	–	28	31	–	36	38
<b>t3</b>	–	29	25	–	28	29	–	39	37
<b>t4</b>	–	26	23	–	24	30	–	47	36
<b>t5</b>	–	16	17	–	12	21	–	43	26
<b>t6</b>	–	3	7	–	2	8	–	32	13
<b>t7</b>	–	8	7	–	12	8	–	22	7
<b>d</b>	<b>M-Ag</b>	<b>L-Ag</b>	<b>U-Ag</b>	<b>M-Au</b>	<b>L-Au</b>	<b>U-Au</b>	<b>M-Cu</b>	<b>L-Cu</b>	<b>U-Cu</b>
<b>s</b>	0	0	0	0	0	0	0	0	0
<b>b1</b>	2.6	2.6	2.7	2.8	2.9	2.9	2.2	2.2	2.5
<b>b2</b>	2.7	2.7	2.8	2.9	2.9	2.9	2.3	2.3	2.5
<b>b3</b>	2.7	2.7	2.8	3.0	3.0	3.0	2.5	2.4	2.5
<b>b4</b>	2.9	3.0	3.0	3.1	3.5	3.4	2.8	2.5	2.9
<b>b5</b>	3.7	3.9	4.1	3.9	4.7	4.5	3.5	3.1	4.1
<b>b6</b>	4.9	5.1	5.5	5.1	5.8	5.9	4.6	4.3	5.8
<b>b7</b>	5.7	5.7	6.3	6.0	6.1	6.6	5.2	5.2	6.4
<b>t1</b>	–	5.6	7.6	–	6.0	8.2	–	6.8	8.5
<b>t2</b>	–	6.1	8.4	–	6.7	8.9	–	6.3	9.2
<b>t3</b>	–	7.1	9.6	–	7.9	10.2	–	7.0	10.7
<b>t4</b>	–	8.3	10.6	–	8.9	11.4	–	8.7	12.1
<b>t5</b>	–	8.8	11.2	–	9.2	12.0	–	10.2	12.6
<b>t6</b>	–	8.9	11.3	–	9.2	11.8	–	10.5	12.3
<b>t7</b>	–	9.0	11.4	–	9.5	11.7	–	9.7	12.0

# Bibliography

- [1] New Science for a Secure and Sustainable Energy Future. Technical report, US Department of Energy Office of Science, 2008.
- [2] Computational Materials Science and Chemistry: Accelerating Discovery and Innovation through Simulation-based Engineering and Science. Technical report, US Department of Energy Office of Science, 2010.
- [3] G.M. Whitesides, J.P. Mathias, and C.T. Seto. Molecular self-assembly and nanochemistry: a chemical strategy for the synthesis of nanostructures. *Science*, 254:1312–1319, 1991.
- [4] Bartosz A. Grzybowski, Christopher E. Wilmer, Jiwon Kim, Kevin P. Browne, and Kyle J. M. Bishop. Self-assembly: from crystals to cells. *Soft Matter*, 5(6):1110, 2009.
- [5] Marek Grzelczak, Jan Vermant, Eric M. Furst, and Luis M. Liz-Marzan. Directed self-assembly of nanoparticles. *ACS Nano*, 4(7):3591–3605, jul 2010.
- [6] A. Ulman. Formation and structure of self-assembled monolayers. *Chem. Rev.*, 96:1533, 1996.
- [7] George M. Whitesides, Jennah K. Kriebel, and J. Christopher Love. Molecular engineering of surfaces using self-assembled monolayers. *Sci. Prog.*, 88(1):17–48, feb 2005.
- [8] Hsun-Yun Chang, Chih-Chieh Huang, Kang-Yi Lin, Wei-Lun Kao, Hua-Yang Liao, Yun-Wen You, Jiun-Hao Lin, Yu-Ting Kuo, Ding-Yuan Kuo, and Jing-Jong Shyue. Effect of surface potential on NIH3t3 cell adhesion and proliferation. *J. Phys. Chem. C*, 118(26):14464–14470, jul 2014.

- [9] Yusuke Arima and Hiroo Iwata. Effect of wettability and surface functional groups on protein adsorption and cell adhesion using well-defined mixed self-assembled monolayers. *Biomaterials*, 28(20):3074–3082, jul 2007.
- [10] George B. Sigal, Milan Mrksich, and George M. Whitesides. Effect of surface wettability on the adsorption of proteins and detergents. *J. Am. Chem. Soc.*, 120(14):3464–3473, apr 1998.
- [11] Nadia Afara, Sasha Omanovic, and Mehdi Asghari-Khiavi. Functionalization of a gold surface with fibronectin (fn) covalently bound to mixed alkanethiol self-assembled monolayers (sams): The influence of SAM composition on its physicochemical properties and fn surface secondary structure. *Thin Solid Films*, 522:381–389, nov 2012.
- [12] Fabian Meder, Supreet Kaur, Laura Treccani, and Kurosch Rezwan. Controlling mixed-protein adsorption layers on colloidal alumina particles by tailoring carboxyl and hydroxyl surface group densities. *Langmuir*, 29(40):12502–12510, oct 2013.
- [13] Nira Gozlan and Hossam Haick. Coverage effect of self-assembled polar molecules on the surface energetics of silicon. *J. Phys. Chem. C*, 112(33):12599–12601, aug 2008.
- [14] Che-Hung Kuo, Hsun-Yun Chang, Chi-Ping Liu, Szu-Hsian Lee, Yun-Wen You, and Jing-Jong Shyue. Effect of surface chemical composition on the surface potential and iso-electric point of silicon substrates modified with self-assembled monolayers. *Phys. Chem. Chem. Phys.*, 13(9):3649, 2011.
- [15] Wei-Chun Lin, Szu-Hsian Lee, Manuel Karakachian, Bang-Ying Yu, Ying-Yu Chen, Yu-Chin Lin, Che-Hung Kuo, and Jing-Jong Shyue. Tuning the surface potential of gold substrates arbitrarily with self-assembled monolayers with mixed functional groups. *Phys. Chem. Chem. Phys.*, 11(29):6199, 2009.
- [16] Eva Beurer, Nagaiyanallur V. Venkataraman, Marianne Sommer, and Nicholas D. Spencer. Protein and nanoparticle adsorption on orthogonal, charge-density-versus-net-charge surface-chemical gradients. *Langmuir*, 28(6):3159–3166, feb 2012.

- [17] Paul E. Laibinis and George M. Whitesides. Self-assembled monolayers of n-alkanethiolates on copper are barrier films that protect the metal against oxidation by air. *J. Am. Chem. Soc.*, 114:9022–9028, 1992.
- [18] I. Felhosi, E. Kalman, and P. Poczik. *Russian Journal of Electrochemistry*, 38(3):230–237, 2002.
- [19] P. Manandhar, J. Jang, G. C. Schatz, M. A. Ratner, and S. Hong. Anomalous surface diffusion in nanoscale direct deposition processes. *Phys. Rev. Lett.*, 90(11), mar 2003.
- [20] Keith A. Brown, Daniel J. Eichelsdoerfer, Xing Liao, Shu He, and Chad A. Mirkin. Material transport in dip-pen nanolithography. *Frontiers of Physics*, 9(3):385–397, sep 2013.
- [21] V Humblot, S.M Barlow, and R Raval. Two-dimensional organisational chirality through supramolecular assembly of molecules at metal surfaces. *Progress in Surface Science*, 76(1-2):1–19, sep 2004.
- [22] Li Wang, Huihui Kong, Xing Song, Xiaoqing Liu, and Hongming Wang. Chiral supramolecular self-assembly of rubrene. *Phys. Chem. Chem. Phys.*, 12(44):14682, 2010.
- [23] B.A. Parviz, D. Ryan, and G.M. Whitesides. Using self-assembly for the fabrication of nano-scale electronic and photonic devices. *IEEE Transactions on Advanced Packaging*, 26:233–241, 2003.
- [24] Jing Liu, Xinshuai Zhang, Gangqiang Dong, Yuanxun Liao, Bo Wang, Tianchong Zhang, and Futing Yi. Fabrication of silicon nanopillar arrays by cesium chloride self-assembly and wet electrochemical etching for solar cell. *Applied Surface Science*, 289:300–305, jan 2014.
- [25] Nirmalya K Chaki and K Vijayamohanan. Self-assembled monolayers as a tunable platform for biosensor applications. *Biosensors and Bioelectronics*, 17(1-2):1–12, jan 2002.
- [26] Michael A. Boles, Michael Engel, and Dmitri V. Talapin. Self-assembly of colloidal nanocrystals: From intricate structures to functional materials. *Chemical Reviews*, 116(18):11220–11289, sep 2016.

- [27] Jiri Rybacek, Gloria Huerta-Angeles, Adrian Kollarovic, Irena G. Stara, Ivo Stary, Philipp Rahe, Markus Nimmrich, and Angelika Kuhnle. Racemic and optically pure heptahelicene-2-carboxylic acid: Its synthesis and self-assembly into nanowire-like aggregates. *European Journal of Organic Chemistry*, 2011(5):853–860, dec 2010.
- [28] Hugo Ascolani, Maarten W. van der Meijden, Lucila J. Cristina, J. Esteban Gayone, Richard M. Kellogg, Javier D. Fuhr, and Magali Lingenfelder. Van der waals interactions in the self-assembly of 5-amino[6]helicene on cu(100) and au(111). *Chem. Commun.*, 50(90):13907–13909, sep 2014.
- [29] D. Abbasi-Perez, J. M. Recio, and L. Kantorovich. Building motifs during self-assembly of para-terphenyl-meta-dicarbonitrile on a metal surface: A gas-phase study. *The Journal of Physical Chemistry C*, 118(19):10358–10365, mar 2014.
- [30] David Abbasi-Perez, J. Manuel Recio, and Lev Kantorovich. The role of isomerization in the kinetics of self-assembly: p-terphenyl-m-dicarbonitrile on the ag(111) surface. *Physical Chemistry Chemical Physics*, 17(17):11182–11192, 2015.
- [31] Amina Bensalah-Ledoux, Delphine Pitrat, Thibault Reynaldo, Monika Srebro-Hooper, Barry Moore, Jochen Autschbach, Jeanne Crassous, Stephan Guy, and Laure Guy. Large-scale synthesis of helicene-like molecules for the design of enantiopure thin films with strong chiroptical activity. *Chemistry - A European Journal*, 22(10):3333–3346, jan 2016.
- [32] Meike Stohr, Serpil Boz, Michael Schar, Manh-Thuong Nguyen, Carlo A. Pignedoli, Daniele Passerone, W. Bernd Schweizer, Carlo Thilgen, Thomas A. Jung, and Francois Diederich. Self-assembly and two-dimensional spontaneous resolution of cyano-functionalized [7]helicenes on cu(111). *Angewandte Chemie International Edition*, 50(42):9982–9986, sep 2011.
- [33] Ramya Desikan, Sarah Armel, Harry M Meyer III, and Thomas Thundat. Effect of chain length on nanomechanics of alkanethiol self-assembly. *Nanotechnology*, 18(42):424028, sep 2007.
- [34] Madhavan Jaccob, Gopalan Rajaraman, and Federico Totti. On the kinetics and thermodynamics of s-x (x = h, CH<sub>3</sub>, SCH<sub>3</sub>, COCH<sub>3</sub>, and CN) cleavage in

- the formation of self-assembled monolayers of alkylthiols on au(111). *Theoretical Chemistry Accounts*, 131(3), feb 2012.
- [35] Piotr Cyganik, Manfred Buck, Thomas Strunskus, and Andrei Shaporenko. Competition as a design concept: Polymorphism in self-assembled monolayers of biphenyl-based thiols. *J. Am. Chem. Soc.*, 128:13868–13878, 2006.
- [36] Nils Heinemann, Jan Grunau, Till Leißner, Oleksiy Andreyev, Sonja Kuhn, Ulrich Jung, Dordaneh Zargarani, Rainer Herges, Olaf Magnussen, and Michael Bauer. Reversible switching in self-assembled monolayers of azobenzene thiolates on au (111) probed by threshold photoemission. *Chemical Physics*, 402:22–28, jun 2012.
- [37] J. Christopher Love, Lara A. Estroff, Jennah K. Kriebel, Ralph G. Nuzzo, and George M. Whitesides. Self-assembled monolayers of thiolates on metals as a form of nanotechnology. *Chemical Reviews*, 105(4):1103–1170, apr 2005.
- [38] Rachel K Smith, Penelope A Lewis, and Paul S Weiss. Patterning self-assembled monolayers. *Progress in Surface Science*, 75(1-2):1–68, jun 2004.
- [39] Che-Hung Kuo, Chi-Ping Liu, Szu-Hsian Lee, Hsun-Yun Chang, Wei-Chun Lin, Yun-Wen You, Hua-Yang Liao, and Jing-Jong Shyue. Effect of surface chemical composition on the work function of silicon substrates modified by binary self-assembled monolayers. *Phys. Chem. Chem. Phys.*, 13(33):15122, 2011.
- [40] Sreenivasa Reddy Puniredd, Ilia Platzman, Raymond T. Tung, and Hossam Haick. Bidirectional control of silicon’s surface potential by means of molecular coverage. *J. Phys. Chem. C*, 114(43):18674–18678, nov 2010.
- [41] Cornelius Gahl, Roland Schmidt, Daniel Brete, Erik R. McNellis, Wolfgang Freyer, Robert Carley, Karsten Reuter, and Martin Weinelt. Structure and excitonic coupling in self-assembled monolayers of azobenzene-functionalized alkanethiols. *J. Am. Chem. Soc.*, 132(6):1831–1838, feb 2010.
- [42] Wesley R. Browne and Ben L. Feringa. Light switching of molecules on surfaces. *Annual Review of Physical Chemistry*, 60(1):407–428, may 2009.
- [43] Tobias Weidner, Frauke Bretthauer, Nirmalya Ballav, Hubert Motschmann, Horst Orendi, Clemens Bruhn, Ulrich Siemeling, and Michael Zharnikov. Cor-

- relation between the molecular structure and photoresponse in aliphatic self-assembled monolayers with azobenzene tailgroups. *Langmuir*, 24(20):11691–11700, oct 2008.
- [44] G. Pace, V. Ferri, C. Grave, M. Elbing, C. von Hanisch, M. Zharnikov, M. Mayor, M. A. Rampi, and P. Samori. *Proc. Natl. Acad. Sci.*, 104:9937, 2007.
- [45] Wei Liu, Victor G Ruiz, Guo-Xu Zhang, Biswajit Santra, Xinguo Ren, Matthias Scheffler, and Alexandre Tkatchenko. Structure and energetics of benzene adsorbed on transition-metal surfaces: density-functional theory with van der waals interactions including collective substrate response. *New Journal of Physics*, 15(5):053046, May 2013.
- [46] Jonas Bjork and Sven Stafstrom. Adsorption of large hydrocarbons on coinage metals: A van der waals density functional study. *ChemPhysChem*, 15(13):2851–2858, jul 2014.
- [47] Wei Liu, Javier Carrasco, Biswajit Santra, Angelos Michaelides, Matthias Scheffler, and Alexandre Tkatchenko. Benzene adsorbed on metals: Concerted effect of covalency and van der waals bonding. *Physical Review B*, 86(24):6, dec 2012.
- [48] T. Verbiest, S. van Elshocht, M. Kauranen, L. Helleman, J. Snauwaert, C. Nuckolls, T. J. Katz, and A. Persoons. Strong enhancement of nonlinear optical properties through supramolecular chirality. *Science*, 282:913 – 915, 1998.
- [49] Monika Srebro, Emmanuel Anger, Barry Moore II, Nicolas Vanthuyne, Christian Roussel, Régis Réau, Jochen Autschbach, and Jeanne Crassous. Ruthenium-grafted vinylhelicenes: Chiroptical properties and redox switching. *Chemistry - A European Journal*, 21(47):17100–17115, oct 2015.
- [50] Minghua Liu, Li Zhang, and Tianyu Wang. Supramolecular chirality in self-assembled systems. *Chemical Reviews*, 115(15):7304–7397, aug 2015.
- [51] Elad Gross, Jack H. Liu, Selim Alayoglu, Matthew A. Marcus, Sirine C. Fakra, F. Dean Toste, and Gabor A. Somorjai. Asymmetric catalysis at the mesoscale:

- Gold nanoclusters embedded in chiral self-assembled monolayer as heterogeneous catalyst for asymmetric reactions. *J. Am. Chem. Soc.*, 135(10):3881–3886, mar 2013.
- [52] Xinjun Wu, Xuebing Ma, Yeling Ji, Qiang Wang, Xiao Jia, and Xiangkai Fu. Synthesis and characterization of a novel type of self-assembled chiral zirconium phosphonates and its application for heterogeneous asymmetric catalysis. *Journal of Molecular Catalysis A: Chemical*, 265(1-2):316–322, mar 2007.
- [53] Thittaya Yutthalekha, Chularat Wattanakit, Veronique Lapeyre, Somkiat Nokbin, Chompunuch Warakulwit, Jumras Limtrakul, and Alexander Kuhn. Asymmetric synthesis using chiral-encoded metal. *Nature Communications*, 7:12678, aug 2016.
- [54] Masazumi Tamura, Ryota Kishi, Yoshinao Nakagawa, and Keiichi Tomishige. Self-assembled hybrid metal oxide base catalysts prepared by simply mixing with organic modifiers. *Nature Communications*, 6:8580, oct 2015.
- [55] Ben L. Feringa, Richard A. van Delden, Nagatoshi Koumura, and Edzard M. Geertsema. Chiroptical molecular switches. *Chemical Reviews*, 100(5):1789–1816, may 2000.
- [56] Richard A. van Delden, Matthijs K. J. ter Wiel, Michael M. Pollard, Javier Vicario, Nagatoshi Koumura, and Ben L. Feringa. Unidirectional molecular motor on a gold surface. *Nature*, 437(7063):1337–1340, oct 2005.
- [57] Jos C. M. Kistemaker, Peter Štacko, Johan Visser, and Ben L. Feringa. Unidirectional rotary motion in achiral molecular motors. *Nature Chemistry*, 7(11):890–896, oct 2015.
- [58] Philipp Furche, Reinhart Ahlrichs, Claudia Wachsmann, Edwin Weber, Adam Sobanski, Fritz Vogtle, and Stefan Grimme. Circular dichroism of helicenes investigated by time-dependent density functional theory. *J. Am. Chem. Soc.*, 122(8):1717–1724, mar 2000.
- [59] Meng Li, Wei Yao, Jun-Dao Chen, Hai-Yan Lu, YongSheng Zhao, and Chuan-Feng Chen. Tetrahydro[5]helicene-based full-color emission dyes in both solution and solid states: synthesis, structures, photophysical properties and optical waveguide applications. *J. Mater. Chem. C*, 2(39):8373–8380, aug 2014.

- [60] I. Hernandez Delgado, S. Pascal, A. Wallabregue, R. Duwald, C. Besnard, L. Guenee, C. Nancoz, E. Vauthey, R. C. Tovar, J. L. Lunkley, G. Muller, and J. Lacour. Functionalized cationic [4]helicenes with unique tuning of absorption, fluorescence and chiroptical properties up to the far-red range. *Chem. Sci.*, 7:4685–4693, 2016.
- [61] Jaroslav Vacek, Jana Vacek Chocholousova, Irena G. Stara, Ivo Stary, and Yonatan Dubi. Mechanical tuning of conductance and thermopower in helicene molecular junctions. *Nanoscale*, 7(19):8793–8802, 2015.
- [62] Jun-ichi Nishida, Takanori Suzuki, Masakazu Ohkita, and Takashi Tsuji. A redox switch based on dihydro[5]helicene: Drastic chiroptical response induced by reversible cc bond making/breaking upon electron transfer. *Angewandte Chemie International Edition*, 40(17):3251–3254, 2001.
- [63] Colin Nuckolls, Renfan Shao, Won-Gun Jang, Noel A. Clark, David M. Walba, and Thomas J. Katz. Electro-optic switching by helicene liquid crystals. *Chemistry of Materials*, 14(2):773–776, feb 2002.
- [64] David Schweinfurth, Michal Zalibera, Michael Kathan, Chengshuo Shen, Marcella Mazzolini, Nils Trapp, Jeanne Crassous, Georg Gescheidt, and Francois Diederich. Helicene quinones: Redox-triggered chiroptical switching and chiral recognition of the semiquinone radical anion lithium salt by electron nuclear double resonance spectroscopy. *J. Am. Chem. Soc.*, 136(37):13045–13052, sep 2014.
- [65] Ying Yang, Rosenildo Correa da Costa, Matthew J. Fuchter, and Alasdair J. Campbell. Circularly polarized light detection by a chiral organic semiconductor transistor. *Nature Photonics*, 7(8):634–638, jul 2013.
- [66] Nidal Saleh, Barry Moore, Monika Srebro, Nicolas Vanthuyne, Loc Toupet, J. A. Gareth Williams, Christian Roussel, Kirandeep K. Deol, Gilles Muller, Jochen Autschbach, and Jeanne Crassous. Acid/base-triggered switching of circularly polarized luminescence and electronic circular dichroism in organic and organometallic helicenes. *Chemistry - A European Journal*, 21(4):1673–1681, nov 2014.
- [67] Jan Storch, Jaroslav Zadny, Tomas Strasak, Martin Kubala, Jan Sykora, Michal Dusek, Vladimir Cirkva, Pavel Matejka, Milos Krbal, and Jan Vacek. Synthesis

- and characterization of a helicene-based imidazolium salt and its application in organic molecular electronics. *Chemistry - A European Journal*, 21(6):2343–2347, dec 2014.
- [68] Marc Gingras, Guy Felix, and Romain Peresutti. One hundred years of helicene chemistry. part 2: stereoselective syntheses and chiral separations of carbohelicenes. *Chem. Soc. Rev.*, 42(3):1007–1050, 2013.
- [69] Gabin Treboux, Paul Lapstun, Zanghua Wu, and Kia Silverbrook. Electronic conductance of helicenes. *Chemical Physics Letters*, 301(5-6):493–497, mar 1999.
- [70] Lubomir Rulisek, Otto Exner, Lukasz Cwiklik, Pavel Jungwirth, Ivo Stary, Lubomir Pospisil, and Zdenek Havlas. On the convergence of the physicochemical properties of [n]helicenes. *J. Phys. Chem. C*, 111(41):14948–14955, oct 2007.
- [71] Maurice J. Narcis and Norito Takenaka. Helical-chiral small molecules in asymmetric catalysis. *European Journal of Organic Chemistry*, 2014(1):21–34, dec 2013.
- [72] Ken-ichi Shinohara, Yuta Sannohe, Shuji Kaieda, Ken-ichi Tanaka, Hideji Otsuga, Hidetoshi Tahara, Yan Xu, Takashi Kawase, Toshikazu Bando, and Hiroshi Sugiyama. A chiral wedge molecule inhibits telomerase activity. *Journal of the American Chemical Society*, 132(11):3778–3782, mar 2010.
- [73] Takahiro Kaseyama, Seiichi Furumi, Xuan Zhang, Ken Tanaka, and Masayuki Takeuchi. Hierarchical assembly of a phthalhydrazide-functionalized helicene. *Angewandte Chemie International Edition*, 50(16):3684–3687, mar 2011.
- [74] Takuji Hatakeyama, Sigma Hashimoto, Tsuyoshi Oba, and Masaharu Nakamura. Azaboradibenzo[6]helicene: Carrier inversion induced by helical homochirality. *J. Am. Chem. Soc.*, 134(48):19600–19603, dec 2012.
- [75] Roman Fasel, Manfred Parschau, and Karl-Heinz Ernst. Chirality transfer from single molecules into self-assembled monolayers. *Angewandte Chemie International Edition*, 42(42):5178–5181, nov 2003.

- [76] R. Fasel, A. Cossy, K.-H. Ernst, F. Baumberger, T. Greber, and J. Osterwalder. Orientation of chiral heptahelicene c<sub>30</sub>h<sub>18</sub> on copper surfaces: An x-ray photoelectron diffraction study. *The Journal of Chemical Physics*, 115(2):1020, 2001.
- [77] K.-H. Ernst, Y. Kuster, R. Fasel, M. Muller, and U. Ellerbeck. Two-dimensional separation of [7]helicene enantiomers on cu(111). *Chirality*, 13(10):675–678, 2001.
- [78] Roman Fasel, Manfred Parschau, and Karl-Heinz Ernst. Amplification of chirality in two-dimensional enantiomorphous lattices. *Nature*, 439(7075):449–452, jan 2006.
- [79] Manfred Parschau, Ursula Ellerbeck, and Karl-Heinz Ernst. Chirality transfer by epitaxial mismatch in multi-layered homochiral molecular films. *Colloids and Surfaces A: Physicochemical and Engineering Aspects*, 354(1-3):240–245, feb 2010.
- [80] K.-H Ernst, Y Kuster, R Fasel, C.F McFadden, and U Ellerbeck. Adsorption of helical aromatic molecules: heptahelicene on ni. *Surface Science*, 530(3):195–202, may 2003.
- [81] Johannes Seibel, Manfred Parschau, and Karl-Heinz Ernst. Two-dimensional crystallization of enantiopure and racemic heptahelicene on ag(111) and au(111). *J. Phys. Chem. C*, 118(50):29135–29141, dec 2014.
- [82] Manfred Parschau and Karl-Heinz Ernst. Disappearing enantiomorphs: Single handedness in racemate crystals. *Angewandte Chemie International Edition*, 54(48):14422–14426, oct 2015.
- [83] Anas Mairena, Laura Zoppi, Johannes Seibel, Alix F. Tröster, Konstantin Grenader, Manfred Parschau, Andreas Terfort, and Karl-Heinz Ernst. Heterochiral to homochiral transition in pentahelicene 2d crystallization induced by second-layer nucleation. *ACS Nano*, 11(1):865–871, jan 2017.
- [84] Dominic A. Notter, Katerina Kouravelou, Theodoros Karachalios, Maria K. Daletou, and Nara Tudela Haberland. Life cycle assessment of PEM FC applications: electric mobility and  $\mu$ -CHP. *Energy Environ. Sci.*, 8(7):1969–1985, 2015.

- [85] G. M. Whitesides. Self-assembly at all scales. *Science*, 295(5564):2418–2421, mar 2002.
- [86] Wai-Yip Lo, Na Zhang, Zhengxu Cai, Lianwei Li, and Luping Yu. Beyond molecular wires: Design molecular electronic functions based on dipolar effect. *Accounts of Chemical Research*, 49(9):1852–1863, sep 2016.
- [87] Vimalkumar Balasubramanian, Bárbara Herranz-Blanco, Patrick V. Almeida, Jouni Hirvonen, and Hélder A. Santos. Multifaceted polymersome platforms: Spanning from self-assembly to drug delivery and protocells. *Progress in Polymer Science*, 60:51–85, sep 2016.
- [88] Sabine H.L. Klapp. Collective dynamics of dipolar and multipolar colloids: From passive to active systems. *Current Opinion in Colloid & Interface Science*, 21:76–85, feb 2016.
- [89] Elizabeth Goiri, Patrizia Borghetti, Afaf El-Sayed, J. Enrique Ortega, and Dimas G. de Oteyza. Multi-component organic layers on metal substrates. *Adv. Mater.*, 28(7):1340–1368, dec 2015.
- [90] Laura E. Heller, Julianne Whitleigh, Danielle F. Roth, Elisabeth M. Oherlein, Felicia R. Lucci, Kristopher J. Kolonko, and Katherine E. Plass. Self-assembly of isomeric monofunctionalized thiophenes. *Langmuir*, 28(42):14855–14859, oct 2012.
- [91] Li Xu, Xinrui Miao, Xiao Ying, and Wenli Deng. Two-dimensional self-assembled molecular structures formed by the competition of van der waals forces and dipole–dipole interactions. *J. Phys. Chem. C*, 116(1):1061–1069, jan 2012.
- [92] Marina Pivetta, Marie-Christine Blum, Francois Patthey, and Wolf-Dieter Schneider. Two-dimensional tiling by rubrene molecules self-assembled in supramolecular pentagons, hexagons, and heptagons on a au(111) surface. *Angewandte Chemie International Edition*, 47(6):1076–1079, jan 2008.
- [93] Marina Pivetta, Marie-Christine Blum, Francois Patthey, and Wolf-Dieter Schneider. Coverage-dependent self-assembly of rubrene molecules on noble metal surfaces observed by scanning tunneling microscopy. *ChemPhysChem*, 11(7):1558–1569, may 2010.

- [94] Zhengyang Liu, Ruohai Guo, Guoxi Xu, Zihan Huang, and Li-Tang Yan. Entropy-mediated mechanical response of the interfacial nanoparticle patterning. *Nano Letters*, 14(12):6910–6916, dec 2014.
- [95] Kunlun Xu, Ruohai Guo, Bojun Dong, and Li-Tang Yan. Directed self-assembly of janus nanorods in binary polymer mixture: towards precise control of nanorod orientation relative to interface. *Soft Matter*, 8(37):9581, 2012.
- [96] Li-Tang Yan, Heiko G. Schoberth, and Alexander Böker. Lamellar microstructure and dynamic behavior of diblock copolymer/nanoparticle composites under electric fields. *Soft Matter*, 6(23):5956, 2010.
- [97] Josep Puigmarti-Luis, Andrea Minoia, Hiroshi Uji-i, Concepcio Rovira, Jerome Cornil, Steven De Feyter, Roberto Lazzaroni, and David B. Amabilino. Non-covalent control for bottom-up assembly of functional supramolecular wires. *J. Am. Chem. Soc.*, 128(39):12602–12603, oct 2006.
- [98] Q. Chen, C.C. Perry, B.G. Frederick, P.W. Murray, S. Haq, and N.V. Richardson. Structural aspects of the low-temperature deprotonation of benzoic acid on cu(110) surfaces. *Surface Science*, 446(1-2):63–75, feb 2000.
- [99] Giulia Tomba, Massimiliano Stengel, Wolf-Dieter Schneider, Alfonso Baldereschi, and Alessandro De Vita. Supramolecular self-assembly driven by electrostatic repulsion: The 1d aggregation of rubrene pentagons on au(111). *ACS Nano*, 4(12):7545–7551, dec 2010.
- [100] Fabio Cicoira, Jill A. Miwa, Dmitrii F. Perepichka, and Federico Rosei. Molecular assembly of rubrene on a metal/metal oxide nanotemplate †. *J. Phys. Chem. A*, 111(49):12674–12678, dec 2007.
- [101] M. Pivetta, M.-C. Blum, F. Patthey, and W.-D. Schneider. *J. Phys. Chem. C*, 113:4578, 2009.
- [102] Jean-Marie Lehn. Perspectives in chemistry-steps towards complex matter. *Angewandte Chemie International Edition*, 52(10):2836–2850, feb 2013.
- [103] R. Gutzler, L. Cardenas, and F. Rosei. *Adv. Mat.*, 25:449, 2013.
- [104] Carlos-Andres Palma, Marco Cecchini, and Paolo Samorì. Predicting self-assembly: from empirism to determinism. *Chem. Soc. Rev.*, 41(10):3713, 2012.

- [105] F. Zerbetto. *Adv. Mat.*, 25:449, 2013.
- [106] Rico Gutzler, Luis Cardenas, and Federico Rosei. Kinetics and thermodynamics in surface-confined molecular self-assembly. *Chemical Science*, 2(12):2290, 2011.
- [107] Gianfranco Ercolani. Assessment of cooperativity in self-assembly. *J. Am. Chem. Soc.*, 125(51):16097–16103, dec 2003.
- [108] Christopher A. Hunter and Harry L. Anderson. What is cooperativity? *Angewandte Chemie International Edition*, 48(41):7488–7499, sep 2009.
- [109] Chris Chapman and Irina Paci. Conformational behavior of chemisorbed azobenzene derivatives in external electric fields: A theoretical study †. *J. Phys. Chem. C*, 114(48):20556–20563, dec 2010.
- [110] Anat Hatzor, Tamar Moav, Hagai Cohen, Sophie Matlis, Jacqueline Libman, Alexander Vaskevich, Abraham Shanzer, and Israel Rubinstein. Coordination-controlled self-assembled multilayers on gold. *Journal of the American Chemical Society*, 120(51):13469–13477, dec 1998.
- [111] R. S. Krishnan, Michael E. Mackay, Phillip M. Duxbury, Alicia Pastor, Craig J. Hawker, Brooke Van Horn, Subashini Asokan, and Michael S. Wong. Self-assembled multilayers of nanocomponents. *Nano Letters*, 7(2):484–489, feb 2007.
- [112] D. Frenkel and B. Smit. *Understanding Molecular Simulation From Algorithms to Applications*. Academic Press, San Diego, California, 2002.
- [113] Ioan Andricioaei and John E. Straub. On monte carlo and molecular dynamics methods inspired by tsallis statistics: Methodology, optimization, and application to atomic clusters. *The Journal of Chemical Physics*, 107(21):9117–9124, December 1997.
- [114] A. J. Lichtenberg and M. A. Lieberman. *Regular and Stochastic Motion*. 03 1983.
- [115] William P. Reinhardt. Chaotic dynamics, semiclassical quantization, and mode-mode energy transfer: the boulder view. *The Journal of Physical Chemistry*, 86(12):2158–2165, June 1982.

- [116] R.S. Mackay, J.D. Meiss, and I.C. Percival. Transport in hamiltonian systems. *Physica D: Nonlinear Phenomena*, 13(1-2):55–81, August 1984.
- [117] Nelson De Leon and B. J. Berne. Intramolecular rate process: Isomerization dynamics and the transition to chaos. *The Journal of Chemical Physics*, 75(7):3495–3510, October 1981.
- [118] John E. Straub and Bruce J. Berne. Energy diffusion in many-dimensional markovian systems: The consequences of competition between inter- and intramolecular vibrational energy transfer. *The Journal of Chemical Physics*, 85(5):2999–3006, September 1986.
- [119] John E. Straub, Michal Borkovec, and Bruce J. Berne. Numerical simulation of rate constants for a two degree of freedom system in the weak collision limit. *The Journal of Chemical Physics*, 86(7):4296–4297, April 1987.
- [120] Irina Paci, Igal Szleifer, and Mark A. Ratner. Structural behavior and self-assembly of lennard-jones clusters on rigid surfaces. *J. Phys. Chem. B*, 109(26):12935–12945, jul 2005.
- [121] A. P. Lyubartsev, A. A. Martsinovski, S. V. Shevkunov, and P. N. Vorontsov-Velyaminov. New approach to monte carlo calculation of the free energy: Method of expanded ensembles. *J. Chem. Phys.*, 96(3):1776, 1992.
- [122] Ulrich H.E. Hansmann. Parallel tempering algorithm for conformational studies of biological molecules. *Chemical Physics Letters*, 281(1-3):140–150, dec 1997.
- [123] Frank Jensen. *Introduction to Computational Chemistry*. John Wiley & Sons, Incorporated, Somerset, 2016.
- [124] Richard Dronskowski. *Computational chemistry of solid state materials : a guide for materials scientists, chemists, physicists and others*. Wiley-VCH, Weinheim Germany, 2005.
- [125] P. W. Atkins and R. S. Friedman. *Molecular quantum mechanics*. Oxford University Press, Oxford New York, 1997.
- [126] Wolfram Koch and Max C. Holthausen. *A chemist's guide to density functional theory*. Wiley-VCH, Weinheim New York, 2001.

- [127] Kieron Burke. Perspective on density functional theory. *The Journal of Chemical Physics*, 136(15):150901, apr 2012.
- [128] Jiri Klimes and Angelos Michaelides. Perspective: Advances and challenges in treating van der waals dispersion forces in density functional theory. *The Journal of Chemical Physics*, 137(12):120901, 2012.
- [129] Yan Zhao and Donald G. Truhlar. Density functionals for noncovalent interaction energies of biological importance. *J. Chem. Theory Comput.*, 3(1):289–300, jan 2007.
- [130] O. Anatole von Lilienfeld, Ivano Tavernelli, Ursula Rothlisberger, and Daniel Sebastiani. Optimization of effective atom centered potentials for london dispersion forces in density functional theory. *Phys. Rev. Lett.*, 93(15), oct 2004.
- [131] Y. Y. Sun, Yong-Hyun Kim, Kyuho Lee, and S. B. Zhang. Accurate and efficient calculation of van der waals interactions within density functional theory by local atomic potential approach. *The Journal of Chemical Physics*, 129(15):154102, 2008.
- [132] Stefan Grimme. Accurate description of van der waals complexes by density functional theory including empirical corrections. *J. Comput. Chem.*, 25(12):1463–1473, 2004.
- [133] Stefan Grimme. Semiempirical gga-type density functional constructed with a long-range dispersion correction. *J. Comput. Chem.*, 27(15):1787–1799, 2006.
- [134] Stefan Grimme, Jens Antony, Stephan Ehrlich, and Helge 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):154104, 2010.
- [135] Alexandre Tkatchenko and Matthias Scheffler. Accurate molecular van der waals interactions from ground-state electron density and free-atom reference data. *Phys. Rev. Lett.*, 102(7), feb 2009.
- [136] Axel D. Becke. Real-space post-hartree–fock correlation models. *The Journal of Chemical Physics*, 122(6):064101, 2005.

- [137] Erin R. Johnson. The exchange-hole dipole moment dispersion model. In *Non-Covalent Interactions in Quantum Chemistry and Physics*, pages 169–194. Elsevier, 2017.
- [138] M. Dion, H. Rydberg, E. Schröder, D. C. Langreth, and B. I. Lundqvist. Van der waals density functional for general geometries. *Phys. Rev. Lett.*, 92(24), jun 2004.
- [139] Eamonn D. Murray, Kyuho Lee, and David C. Langreth. Investigation of exchange energy density functional accuracy for interacting molecules. *J. Chem. Theory Comput.*, 5(10):2754–2762, oct 2009.
- [140] Jiří Klimeš, David R Bowler, and Angelos Michaelides. Chemical accuracy for the van der waals density functional. *Journal of Physics: Condensed Matter*, 22(2):022201, dec 2009.
- [141] Takeshi Sato and Hiromi Nakai. Density functional method including weak interactions: Dispersion coefficients based on the local response approximation. *The Journal of Chemical Physics*, 131(22):224104, 2009.
- [142] O. Anatole von Lilienfeld and Alexandre Tkatchenko. Two- and three-body interatomic dispersion energy contributions to binding in molecules and solids. *The Journal of Chemical Physics*, 132(23):234109, 2010.
- [143] Alexandre Tkatchenko, Robert A. DiStasio, Roberto Car, and Matthias Scheffler. Accurate and efficient method for many-body van der waals interactions. *Phys. Rev. Lett.*, 108(23), jun 2012.
- [144] Victor G. Ruiz, Wei Liu, Egbert Zojer, Matthias Scheffler, and Alexandre Tkatchenko. Density-functional theory with screened van der waals interactions for the modeling of hybrid inorganic-organic systems. *Phys. Rev. Lett.*, 108(14), apr 2012.
- [145] Tim Gould, Erin R Johnson, and Sherif Abdulkader Tawfik. Are dispersion corrections accurate outside equilibrium? a case study on benzene. *Beilstein Journal of Organic Chemistry*, 14:1181–1191, May 2018.
- [146] Axel D. Becke and Erin R. Johnson. Exchange-hole dipole moment and the dispersion interaction. *The Journal of Chemical Physics*, 122(15):154104, April 2005.

- [147] Axel D. Becke and Erin R. Johnson. A density-functional model of the dispersion interaction. *The Journal of Chemical Physics*, 123(15):154101, October 2005.
- [148] Axel D. Becke and Erin R. Johnson. Exchange-hole dipole moment and the dispersion interaction: High-order dispersion coefficients. *The Journal of Chemical Physics*, 124(1):014104, January 2006.
- [149] Axel D. Becke and Erin R. Johnson. Exchange-hole dipole moment and the dispersion interaction revisited. *The Journal of Chemical Physics*, 127(15):154108, October 2007.
- [150] Petr Jurečka, Jiří Šponer, Jiří Černý, and Pavel Hobza. Benchmark database of accurate (MP2 and CCSD(t) complete basis set limit) interaction energies of small model complexes, DNA base pairs, and amino acid pairs. *Phys. Chem. Chem. Phys.*, 8(17):1985–1993, 2006.
- [151] Ming Xi, Michael X. Yang, Sam K. Jo, Brian E. Bent, and Paul Stevens. Benzene adsorption on cu(111): Formation of a stable bilayer. *The Journal of Chemical Physics*, 101(10):9122, 1994.
- [152] X.-L Zhou, M.E Castro, and J.M White. Interactions of UV photons and low energy electrons with chemisorbed benzene on ag(111). *Surface Science*, 238(1-3):215–225, nov 1990.
- [153] Denis Syomin, JooHo Kim, Bruce E. Koel, and G. Barney Ellison. Identification of adsorbed phenyl (c 6 h 5 ) groups on metal surfaces: electron-induced dissociation of benzene on au(111). *The Journal of Physical Chemistry B*, 105(35):8387–8394, sep 2001.
- [154] Julian R. Lomas, Christopher J. Baddeley, Mintcho S. Tikhov, and Richard M. Lambert. Ethyne cyclization to benzene over cu(110). *Langmuir*, 11(8):3048–3053, aug 1995.
- [155] S. J. Jenkins. Aromatic adsorption on metals via first-principles density functional theory. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 465(2110):2949–2976, jul 2009.

- [156] Javier Carrasco, Wei Liu, Angelos Michaelides, and Alexandre Tkatchenko. Insight into the description of van der waals forces for benzene adsorption on transition metal (111) surfaces. *The Journal of Chemical Physics*, 140(8):084704, feb 2014.
- [157] Ante Bilić, Jeffrey R. Reimers, Noel S. Hush, Rainer C. Hoft, and Michael J. Ford. Adsorption of benzene on copper, silver, and gold surfaces. *J. Chem. Theory Comput.*, 2(4):1093–1105, jul 2006.
- [158] Peter Pulay. A perspective on the CASPT2 method. *Int. J. Quantum Chem.*, 111(13):3273–3279, apr 2011.
- [159] Kenji Toyoda, Yosuke Nakano, Ikutaro Hamada, Kyuho Lee, Susumu Yanagisawa, and Yoshitada Morikawa. First-principles study of benzene on noble metal surfaces: Adsorption states and vacuum level shifts. *Surface Science*, 603(18):2912–2922, sep 2009.
- [160] Luiza Buimaga-Iarinca and Cristian Morari. Adsorption of small aromatic molecules on gold: a DFT localized basis set study including van der waals effects. *Theor Chem Acc*, 133(7), may 2014.
- [161] J. Michael Gottfried, Ebbe K. Vestergaard, Parthasarathi Bera, and Charles T. Campbell. Heat of adsorption of naphthalene on pt(111) measured by adsorption calorimetry. *The Journal of Physical Chemistry B*, 110(35):17539–17545, sep 2006.
- [162] Wei Liu, Alexandre Tkatchenko, and Matthias Scheffler. Modeling adsorption and reactions of organic molecules at metal surfaces. *Accounts of Chemical Research*, 47(11):3369–3377, nov 2014.
- [163] Kenji Toyoda, Ikutaro Hamada, Kyuho Lee, Susumu Yanagisawa, and Yoshitada Morikawa. Density functional theoretical study of pentacene/noble metal interfaces with van der waals corrections: Vacuum level shifts and electronic structures. *The Journal of Chemical Physics*, 132(13):134703, 2010.
- [164] Peter Kratzer and Jörg Neugebauer. The basics of electronic structure theory for periodic systems. *Frontiers in Chemistry*, 7, March 2019.

- [165] Konstantin Gubaev, Evgeny V. Podryabinkin, Gus L.W. Hart, and Alexander V. Shapeev. Accelerating high-throughput searches for new alloys with active learning of interatomic potentials. *Computational Materials Science*, 156:148–156, January 2019.
- [166] Kamal Choudhary and Francesca Tavazza. Convergence and machine learning predictions of monkhorst-pack k-points and plane-wave cut-off in high-throughput DFT calculations. *Computational Materials Science*, 161:300–308, April 2019.
- [167] Sverre Froyen. Brillouin-zone integration by fourier quadrature: Special points for superlattice and supercell calculations. *Physical Review B*, 39(5):3168–3172, February 1989.
- [168] Georg K. H. Madsen, Peter Blaha, Karlheinz Schwarz, Elisabeth Sjostedt, and Lars Nordstrom. Efficient linearization of the augmented plane-wave method. *Physical Review B*, 64(19):195134, October 2001.
- [169] Felix A. Pahl and Nicholas C. Handy. Plane waves and radial polynomials: a new mixed basis. *Molecular Physics*, 100(20):3199–3224, February 2002.
- [170] Laszlo Fusti-Molnar and Peter Pulay. Accurate molecular integrals and energies using combined plane wave and gaussian basis sets in molecular electronic structure theory. *J. Chem. Phys.*, 116(18):7795, April 2002.
- [171] P. J. Hasnip, K. Refson, M. I. J. Probert, J. R. Yates, S. J. Clark, and C. J. Pickard. Density functional theory in the solid state. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 372(2011):20130270–20130270, February 2014.
- [172] B. Hammer and J.K. Nørskov. Theoretical surface science and catalysis—calculations and concepts. In *Advances in Catalysis*, pages 71–129. Elsevier, 2000.
- [173] B. Hammer and J. K. Norskov. Why gold is the noblest of all the metals. *Nature*, 376(6537):238–240, July 1995.
- [174] B. Hammer, L. B. Hansen, and J. K. Nørskov. Improved adsorption energetics within density-functional theory using revised perdew-burke-ernzerhof functionals. *Physical Review B*, 59(11):7413–7421, March 1999.

- [175] J. K. Norskov, F. Abild-Pedersen, F. Studt, and T. Bligaard. Density functional theory in surface chemistry and catalysis. *Proceedings of the National Academy of Sciences*, 108(3):937–943, January 2011.
- [176] J. K. Nørskov, T. Bligaard, J. Rossmeisl, and C. H. Christensen. Towards the computational design of solid catalysts. *Nature Chemistry*, 1(1):37–46, April 2009.
- [177] R. S. Mulliken. Criteria for the construction of good self-consistent-field molecular orbital wave functions, and the significance of LCAO-MO population analysis. *The Journal of Chemical Physics*, 36(12):3428–3439, June 1962.
- [178] Alfons van Blaaderen, Rene Ruel, and Pierre Wiltzius. Template-directed colloidal crystallization. *Nature*, 385(6614):321–324, 01 1997.
- [179] Yurii A. Vlasov, Xiang-Zheng Bo, James C. Sturm, and David J. Norris. *Nature*, 414(6861):289–293, nov 2001.
- [180] Nina V. Dziomkina, Mark A. Hempenius, and G. Julius Vancso. Layer-by-layer templated growth of colloidal crystals with packing and pattern control. *Colloids and Surfaces A: Physicochemical and Engineering Aspects*, 342(13):8 – 15, 2009.
- [181] William L. Jorgensen, Jeffrey D. Madura, and Carol J. Swenson. Optimized intermolecular potential functions for liquid hydrocarbons. *J. Am. Chem. Soc.*, 106(22):6638–6646, oct 1984.
- [182] Tatiana Popa and Irina Paci. Surface effects in chiral adsorption. *Chemical Physics Letters*, 507(1-3):128–133, apr 2011.
- [183] Tatiana Popa and Irina Paci. Designing enantioselectivity in chiral self-assembly at a solid substrate: a theoretical study of competing interactions. *Soft Matter*, 9(33):7988, 2013.
- [184] Q. Chen, D. J. Frankel, and N. V. Richardson. Room-temperature surface structure of 4-aminobenzoic acid on cu(110) surfaces. *Surface and Interface Analysis*, 32(1):43–48, 2001.

- [185] H. Li, X. Jiang, and B. G. Willis. Prevailing intermolecular bonding for dinitrotoluene self-assembled monolayers on au(111). *J. Phys. Chem. C*, 118(26):14418–14426, jul 2014.
- [186] H. Kraack, B. M. Ocko, P. S. Pershan, E. Sloutskin, and M. Deutsch. Structure of a langmuir film on a liquid metal surface. *Science*, 298(5597):1404–1407, nov 2002.
- [187] K. Morishige, N. Kawai, and M. Shimizu. Occurrence of two-dimensional smectic liquid crystal in n-propanol adsorbed on graphite. *Phys. Rev. Lett.*, 70(25):3904–3906, jun 1993.
- [188] Julio C. Armas-Pérez and Jacqueline Quintana-H. Numerical evidence for nematic and smectic behavior of two-dimensional hard models. *Physical Review E*, 83(5), may 2011.
- [189] J M Kosterlitz and D J Thouless. Ordering, metastability and phase transitions in two-dimensional systems. *J. Phys. C: Solid State Phys.*, 6(7):1181–1203, apr 1973.
- [190] J. A. Martínez-González, J. C. Armas-Pérez, and J. Quintana-H. Phase behavior of bow-shaped hard particles in two dimensions. *Journal of Statistical Physics*, 150(3):559–571, oct 2013.
- [191] Szabolcs Varga, Peter Gurin, Julio C. Armas-Perez, and Jacqueline Quintana-H. Nematic and smectic ordering in a system of two-dimensional hard zigzag particles. *The Journal of Chemical Physics*, 131(18):184901, 2009.
- [192] Martin A. Bates and Daan Frenkel. Phase behavior of two-dimensional hard rod fluids. *The Journal of Chemical Physics*, 112(22):10034, 2000.
- [193] Raffaele Tavarone, Patrick Charbonneau, and Holger Stark. Phase ordering of zig-zag and bow-shaped hard needles in two dimensions. *The Journal of Chemical Physics*, 143(11):114505, sep 2015.
- [194] Guillem Portella, Jordi Poater, Josep M. Bofill, Pere Alemany, and Miquel Sola. Local aromaticity of [n]acenes, [n]phenacenes, and [n]helicenes ( n = 1-9). *The Journal of Organic Chemistry*, 70(7):2509–2521, apr 2005.

- [195] Jerome M. Schulman and Raymond L. Disch. Aromatic character of [n]helicenes and [n]phenacenes. *J. Phys. Chem. A*, 103(33):6669–6672, aug 1999.
- [196] S. Grimme and S.D. Peyerimhoff. Theoretical study of the structures and racemization barriers of [n]helicenes ( $n = 3-6, 8$ ). *Chemical Physics*, 204(2-3):411–417, apr 1996.
- [197] Haydee Valdes, Vojtech Klusak, Michal Pitonak, Otto Exner, Ivo Stary, Pavel Hobza, and Lubomir Rulisek. Evaluation of the intramolecular basis set superposition error in the calculations of larger molecules: [n]helicenes and phe-glyphe tripeptide. *J. Comput. Chem.*, 29(6):861–870, 2008.
- [198] Elvis C M Ting, Tatiana Popa, and Irina Paci. Surface-site reactivity in small-molecule adsorption: A theoretical study of thiol binding on multi-coordinated gold clusters. *Beilstein J. Nanotechnol.*, 7:53–61, jan 2016.
- [199] Chris R. L. Chapman, Elvis C.M. Ting, Ashley Kereszti, and Irina Paci. Self-assembly of cysteine dimers at the gold surface: A computational study of competing interactions. *J. Phys. Chem. C*, 117(38):19426–19435, sep 2013.
- [200] A good overview is given in D. Sherrill: Counterpoise Correction and Basis Set Superposition Error, <http://vergil.chemistry.gatech.edu/notes/>.
- [201] K. Lee, J. Yu, and Y. Morikawa. Comparison of localized basis and plane-wave basis for density-functional calculations of organic molecules on metals. *Phys. Rev. B*, 75:045402–1 – 5, 2007.
- [202] Jose M Soler, Emilio Artacho, Julian D Gale, Alberto Garcia, Javier Junquera, Pablo Ordejon, and Daniel Sanchez-Portal. The siesta method for ab initio order- n materials simulation. *Journal of Physics: Condensed Matter*, 14(11):2745, 2002.
- [203] C. David Sherrill. Counterpoise correction and basis set superposition error. Technical report, Georgia Institute of Technology, School of Chemistry and Biochemistry, 2010.
- [204] Veronica Barone, Oded Hod, and Gustavo E. Scuseria. Electronic structure and stability of semiconducting graphene nanoribbons. *Nano Letters*, 6(12):2748–2754, dec 2006.

- [205] Jonathan P. Hill, Wusong Jin, Atsuko Kosaka, Takanori Fukushima, Hideki Ichihara, Takeshi Shimomura, Kohzo Ito, Tomihiro Hashizume, Noriyuki Ishii, and Takuzo Aida. Self-assembled hexa-peri-hexabenzocoronene graphitic nanotube. *Science*, 304(5676):1481–1483, jun 2004.
- [206] Sasha Stankovich, Dmitriy A. Dikin, Geoffrey H. B. Dommett, Kevin M. Kohlhaas, Eric J. Zimney, Eric A. Stach, Richard D. Piner, SonBinh T. Nguyen, and Rodney S. Ruoff. Graphene-based composite materials. *Nature*, 442(7100):282–286, jul 2006.
- [207] M. Rapacioli, F. Calvo, F. Spiegelman, C. Joblin, and D. J. Wales. Stacked clusters of polycyclic aromatic hydrocarbon molecules. *The Journal of Physical Chemistry A*, 109(11):2487–2497, mar 2005.
- [208] Yan Zhao and Donald G. Truhlar. A prototype for graphene material simulation: Structures and interaction potentials of coronene dimers. *The Journal of Physical Chemistry C*, 112(11):4061–4067, mar 2008.
- [209] O. I. Obolensky, V. V. Semenikhina, A. V. Solovyov, and W. Greiner. Interplay of electrostatic and van der waals forces in coronene dimer. *International Journal of Quantum Chemistry*, 107(6):1335–1343, 2007.
- [210] Deborah J. Hankinson and Jan Almlöf. Cluster models for lithium intercalated graphite: electronic structures and energetics. *Journal of Molecular Structure: THEOCHEM*, 388:245–256, dec 1996.
- [211] Henna Ruuska and Tapani A. Pakkanen. Ab initio study of interlayer interaction of graphite: Benzene-coronene and coronene dimer two-layer models. *The Journal of Physical Chemistry B*, 105(39):9541–9547, oct 2001.
- [212] A. H. Castro Neto, F. Guinea, N. M. R. Peres, K. S. Novoselov, and A. K. Geim. The electronic properties of graphene. *Reviews of Modern Physics*, 81(1):109–162, jan 2009.
- [213] Mutasem Omar Sinnokrot, Edward F. Valeev, and C. David Sherrill. Estimates of the ab initio limit for  $\pi$ - $\pi$  interactions: the benzene dimer. *Journal of the American Chemical Society*, 124(36):10887–10893, sep 2002.

- [214] 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):1312–1314, may 2009.
- [215] Venkataraman Pallassana and Matthew Neurock. Electronic factors governing ethylene hydrogenation and dehydrogenation activity of pseudomorphic PdM-L/re(0001), PdML/ru(0001), pd(111), and PdML/au(111) surfaces. *Journal of Catalysis*, 191(2):301–317, apr 2000.
- [216] Allan D. Mackie, Bernard Tavitian, Anne Boutin, and Alain H. Fuchs. Vapour-liquid phase equilibria predictions of methane–alkane mixtures by monte carlo simulation. *Molecular Simulation*, 19(1):1–15, February 1997.
- [217] Xavier Daura, Alan E. Mark, and Wilfred F. Van Gunsteren. Parametrization of aliphatic CHn united atoms of GROMOS96 force field. *Journal of Computational Chemistry*, 19(5):535–547, April 1998.