

## Molecular Modeling of Supramolecular Structures

Piskorz, Tomasz

**DOI**

[10.4233/uuid:2c876b61-a850-4ae1-b47d-38a60a576006](https://doi.org/10.4233/uuid:2c876b61-a850-4ae1-b47d-38a60a576006)

**Publication date**

2019

**Document Version**

Final published version

**Citation (APA)**

Piskorz, T. (2019). *Molecular Modeling of Supramolecular Structures*. [Dissertation (TU Delft), Delft University of Technology]. <https://doi.org/10.4233/uuid:2c876b61-a850-4ae1-b47d-38a60a576006>

**Important note**

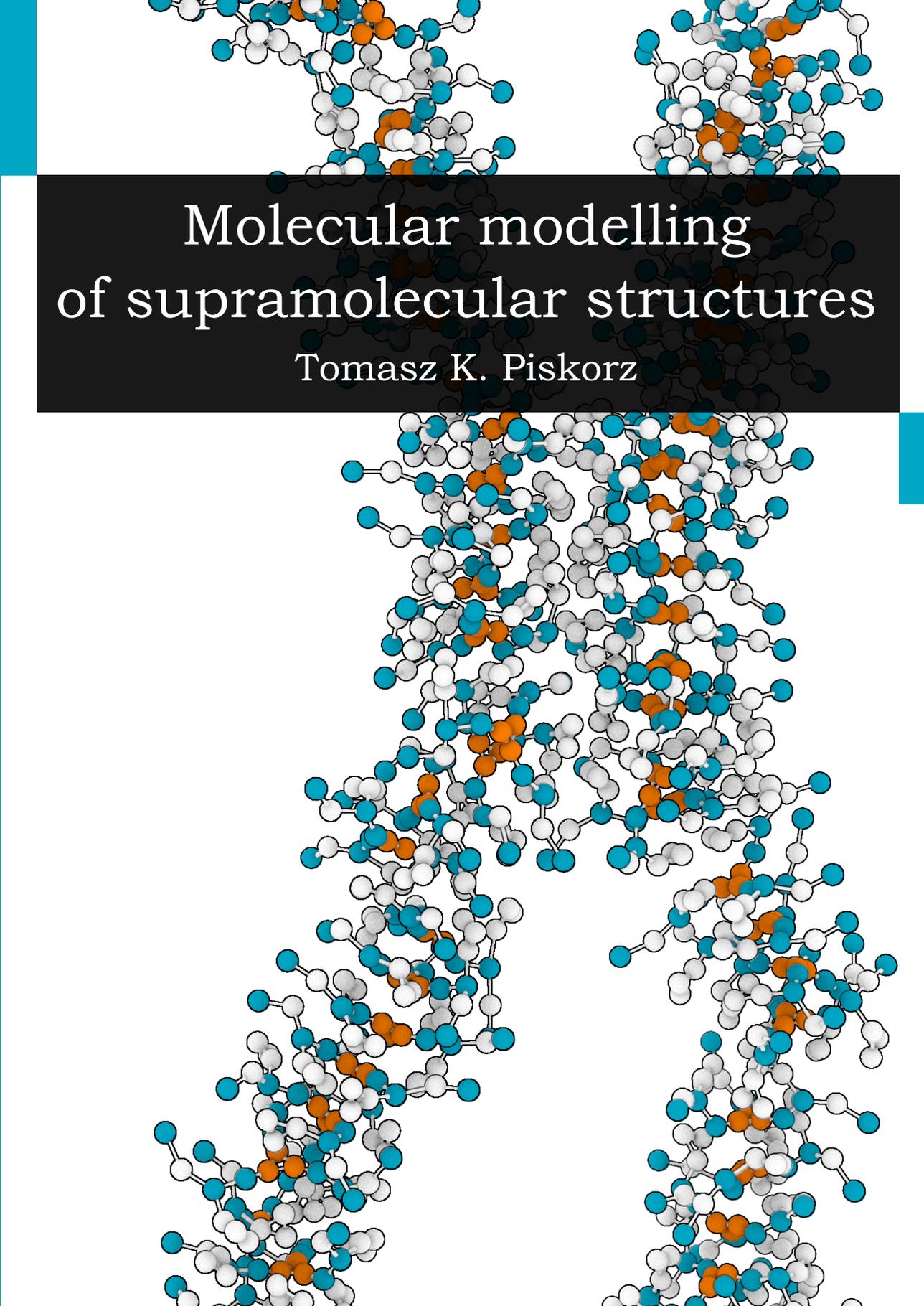
To cite this publication, please use the final published version (if applicable).  
Please check the document version above.

**Copyright**

Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

**Takedown policy**

Please contact us and provide details if you believe this document breaches copyrights.  
We will remove access to the work immediately and investigate your claim.

The image features a central black rectangular area containing white text. Above and below this area are several ball-and-stick molecular models. The models consist of interconnected spheres in white, light blue, and orange, representing atoms and their bonds. The structures are complex and three-dimensional, showing various orientations and interactions. The background is white with a light blue vertical bar on the left and a dark blue vertical bar on the right.

# Molecular modelling of supramolecular structures

Tomasz K. Piskorz



# Molecular Modeling of Supramolecular Structures

Dissertation

for the purpose of obtaining the degree of doctor

at Delft University of Technology

by the authority of the Rector Magnificus, prof. dr. ir. T. H. J. J. van der Hagen,

chair of the Board for Doctorates

to be defended publicly on

Thursday, 5<sup>th</sup> of September 2019 at 12:30 o'clock

by

Tomasz Krzysztof PISKORZ

Master of Science in Chemistry,

University of Warsaw, Poland

born in Kraków, Poland

This dissertation has been approved by the  
promoters: Prof. dr. J. H. van Esch  
copromotor: Dr. A. H. de Vries

Composition of the doctoral committee:

Rector Magnificus	chairman
Prof. dr. J. H. van Esch	Delft University of Technology
Dr. A. H. de Vries	University of Groningen

Independent members:

Prof. dr. R. Oda	CNRS, France
Prof. dr. X.Y. Liu	Xiamen University, China
Prof. dr. ir. T.J.H. Vlugt	Delft University of Technology
Prof. dr. S.J. Picken	Delft University of Technology
Prof. dr. E.J.R. Sudhölter	Delft University of Technology, reserve member

Other members:

Prof. dr. T. Tuttle	University of Strathclyde, United Kingdom
---------------------	---

The work described in this thesis was carried out in the Advanced Soft Matter group at Delft University of Technology. This research was funded by EC 7<sup>th</sup> Framework Programme Marie Curie Actions via the European ITN SMARTNET No. 316656.

©Tomasz K. Piskorz, 2019

ISBN 978-94-028-1665-5

Printed by: Ipskamp Printing

All rights reserved. The author encourages the communication of scientific contents and explicitly allows reproduction for scientific purposes with proper citation of the source. Parts of this thesis have been published in scientific journals and copyright is subject of different terms and conditions.

*“... universities are truly storehouses of knowledge: students arrive from school confident that they know nearly everything, and they leave years later certain that they know practically nothing. Where did the knowledge go in the meantime? Into the university, of course, where it is carefully dried and stored.”*

Terry Pratchett, *The Science of Discworld*





# Table of Contents

Chapter 1 Introduction.....	1
General introduction to self-assembly .....	2
Self-assembly .....	2
Research aims and outline of the thesis .....	8
Chapter 2 Lessons learned from molecular dynamics of trifold hydrogen bonded fibers with different force-fields .....	11
Introduction .....	12
Results .....	13
Conclusions .....	21
Methods .....	21
Supporting Information .....	23
Chapter 3 Fiber formation seen through the high resolution computational microscope.....	31
Introduction .....	32
Approach .....	33
Results and discussion .....	35
Summary and discussion of the mechanism of supramolecular self-assembly .....	45
Conclusions and Outlook .....	46
Supporting information .....	48
Chapter 4 Nucleation mechanisms of self-assembled physisorbed monolayers on graphite .....	69
Introduction .....	70
Results and discussion .....	71
Conclusions .....	86
Methods .....	86
Supporting information .....	90

Chapter 5 Mechanism of Ostwald Ripening in 2D physisorbed assemblies at molecular time- and length-scales by Molecular Dynamics simulations .....	123
Introduction .....	124
Results and discussion .....	124
Conclusions .....	132
Methods .....	133
Supporting Information .....	135
Summary .....	147
Samenvatting .....	150
References .....	154
Acknowledgements .....	163
About the author .....	167
List of publications .....	167



# Chapter 1

## Introduction

## General introduction to self-assembly

Self-assembly is a method of engineering structures by spontaneous organization of small molecular blocks<sup>1-4</sup>. The resulting structures, called supramolecular, have unique properties, such as rich and dynamic behavior. There are many examples of supramolecular systems, such as micelles<sup>5</sup>, vesicles<sup>6,7</sup>, fibers<sup>8-10</sup>, molecular crystals<sup>11,12</sup> and many more. Moreover, Nature exploits this process for many vital functions (e.g., for cellular integrity), but also for abnormal processes such as degenerative diseases (for example Alzheimer's or prion diseases)<sup>13-15</sup>. In the last decades, self-assembly has drawn attention as a promising alternative to “top-down” techniques used by industry to produce well-defined nano- to microscale structures<sup>16</sup>, since it can be seen as the next step to organic synthesis, i.e., synthesis of relatively large structures from smaller building blocks by reversible, non-covalent interactions<sup>1</sup>. The strategies for fabrication of structures linked to their length-scales is shown in Figure 1.1.

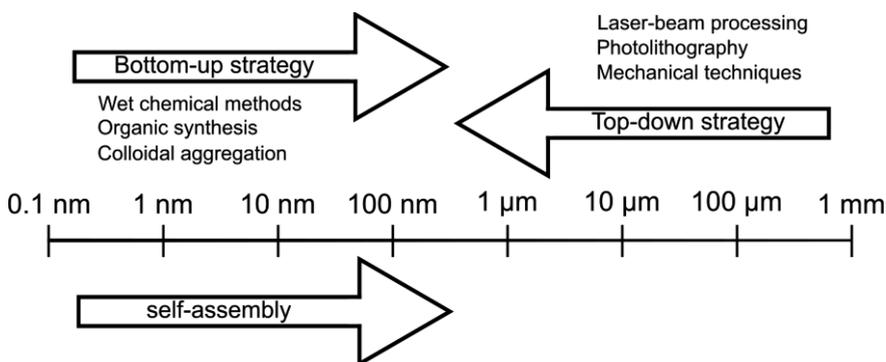


Figure 1.1. Bottom-up and top-down approaches. Reproduced from the reference<sup>17</sup>.

## Self-assembly

Probably the most fascinating fact about self-assembly is that the information about supramolecular structure is encoded in a single molecular block<sup>4</sup>. The complementary interactions between molecular blocks dictate the final structure of the assembly. Therefore, one of the critical points of understanding self-assembly is understanding the forces by which molecules are interacting. Molecular blocks interact with each other by noncovalent interactions, namely: electrostatic, hydrogen bonds, Van der Waals, and hydrophobic interactions<sup>4,18</sup>. These interactions are summarized in Figure 1.2. In general, non-covalent interactions are weaker (2-300 kJ/mol and only

electrostatic interactions are larger than 100 kJ/mol) than covalent interactions (150-450 kJ/mol). Although a single noncovalent interaction is weak, the ensemble of them can result in strong bonding.

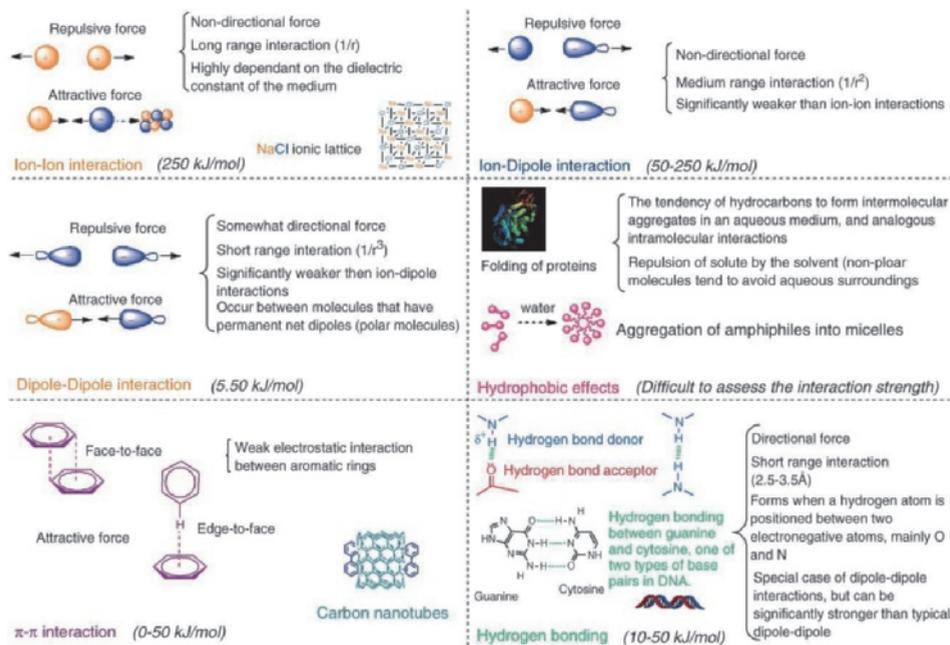
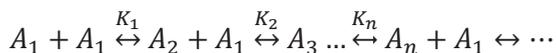


Figure 1.2. Noncovalent interactions involved in supramolecular chemistry. Reproduced from the reference<sup>4</sup>.

The desire of direct and robust translation of the molecular structure to self-assembled systems has motivated many researchers to formulate design rules, which would enable the design of molecular structures that result in the desired supramolecular assembly. Over the past decades, several popular molecular design rules have been established for self-assembling systems. Probably the most famous design rule is surfactant packing parameter proposed by Israelachvili et al.<sup>19</sup>, which from the shape of a single surfactant molecule predicts the overall shape of the formed aggregate, a micelle, bilayer, or inverted micelle. Another notable design rule is developed by Hanabusa et al.<sup>20</sup> for supramolecular gelators. According to this rule the molecule which creates a gel has to have three properties: (a) strong directional interactions inducing a directional aggregate, (b) some factor preventing crystallization, and (c) ability to cross-link between aggregates for network formation. There are also more subtle rules which emerge as "crystal engineering" in the context of 2D crystals

formed on the solid-liquid interface<sup>21,22</sup>. Although all these rules help to design molecules which would form self-assembled structure many unexpected results can be found. Probably the best example is the gelators, for which although there are design rules, most of them are still found by serendipity<sup>9,23,24</sup>. The reason for that can be, of course, that design rules are rules-of-thumb and only roughly show how to design molecules, but it can also be because these rules are based only on the final assembled structure and do not take into account the process which leads to the assembly.

Therefore, there is a need for understanding the mechanism of self-assembly, because this unexpected behavior could arise from the way how the supramolecular structures are formed and be related to metastable states that occur during the process<sup>24</sup>. In general, self-assembly can be described as a series of associations of monomers to aggregates:



Since the measurement of individual equilibrium constants is not feasible, the model is simplified by making some assumptions, from which three are especially often used in supramolecular community<sup>19,25-27</sup>:

- (a) open association/isodesmic model, in which equilibrium constants are constant with each step,  $K_1 = \dots = K_N = \dots$ . As a result, every addition of a successive monomer results in a decrease of free energy. Such mechanism is often associated with the formation of linear one-dimensional supramolecular polymers<sup>28</sup>.
- (b) Closed association model, in which one value of equilibrium constants is dominant,  $K_N \gg K_{i \neq N}$ . Such a model describes monodisperse assemblies and most often, it is used to describe the formation of micelles, but it can be also extended to the bilayer, vesicle, and other micellar phases<sup>19</sup>.
- (c) cooperative model, for which equilibrium constants have two values: small for aggregates smaller than a certain size,  $N$ , and large for aggregates above that size,  $K_1 = \dots = K_{N-1} < K_N = \dots$ . As a result, assemblies below size  $N$  are unstable, and only after they reach that size they grow rapidly. Change of the equilibrium constant has its origin in a cooperative effect, which most often is associated with structural changes. The model describes formation of quasi one-dimensional

supramolecular polymers (such as microtubules)<sup>28</sup>. Cooperative effect can also have an origin in hydrophobic and electronic effects and some strictly one-dimensional supramolecular polymers can also undergo cooperative formation<sup>27</sup> (see Chapter 2 and 3 of this thesis).

Thus, the mechanism describing formation of assemblies depends on the selected model. For (a) it follows aggregation, because every association results in the decrease of energy. On the other hand, (b) and (c) follow nucleation-growth mechanism, because first stages are less preferable than the later ones. As a result, formation of a small stable aggregate (called nucleus) requires activation energy. In the cooperative model, after the nucleus is formed it grows by associating with free monomers. (Note that the terms “open/closed association” are used in the context of surfactants, and “isodesmic/cooperative” in the context of supramolecular polymers).

In summary, many experimental techniques try to give insights into self-assembly, but they face many challenges. They are especially useful to provide insights into the final structures; however, insights into the process are more challenging to obtain. According to Frederix et al.<sup>29</sup> there are three main limitations of experimental techniques: (a) experiment measures ensemble of states, (b) raw data is interpreted by models aimed at describing macroscopic systems, (c) self-assembly often takes place on a very short timescale which cannot be reached by most of the experimental techniques.

Due to the experimental limitations, many researchers focus on computational methods which could give insights into early stages of self-assembly (we refer readers more interested in the topic to recent reviews of Bochicchio<sup>30</sup> and Frederix<sup>29</sup>). Depending on the timescale and the size of the systems different levels of accuracy and different computational methods can be applied (see Figure 1.3a). The most accurate quantum mechanics calculations are limited to just a few molecules and find application in a study of the conformational flexibility of the single molecule building blocks and the stability of small parts of self-assembled structures. Less accurate Molecular Dynamics calculations allow studying larger systems on a larger timescale. A number of such studies of the process of self-assembly have been presented in the literature. However, the computational cost of the method still limits most of the application to the study of the dynamics of the final assembled structure. To tackle the computational cost of this method two strategies are widely used: (a) coarse-grained molecular dynamics (CG MD), and (b) rare-event sampling

methods (RES), see Figure 1.3b. CG techniques substantially speed up simulation by treating several atoms as one bead (to represent functionality of the chemical groups, rather than individual atoms). This approach speeds up simulations due to fewer calculations (fewer atoms), and also due to smoothing of the energy landscape<sup>31</sup>. Therefore, it can be applied to systems on large temporal and spatial scale, which resulted in many applications in the study of self-assembly. However, the loss of the resolution and some of the interactions may lead to nonphysical results. On another hand, RES methods retain all atomistic details while allowing them to study them on long time scales. In comparison to coarse-grained techniques they, unfortunately, do not allow to study systems on large length scales, and only study of small systems is possible, which limits the application of these methods for this field. Moreover, their efficiency can suffer from the non-optimal choice of the reaction coordinate used to describe the system. Some examples of the usage of different computational techniques for the study of self-assembling systems are presented in Table 1-1.

Table 1-1. Example of application of different computational techniques for study supramolecular systems.

Method		Pros	Cons	Example works	
Molecular Dynamics	QM	High accuracy	Only very small structures can be studied	Cyclic peptides <sup>32</sup>	
	AA	Validation of the model, insights into dynamics of stable structure	<i>A priori</i> knowledge about the structure required, no insights into self-assembly process	Peptide amphiphiles <sup>33-35</sup> Dye aggregates <sup>36</sup> Oligopeptides <sup>37,38</sup> Supramolecular polymers <sup>39,40</sup>	
MD	AA	Conventional	Simplicity	Only systems on a small temporal and spatial scale can be simulated	Self-assembly of peptide amphiphiles <sup>41,42</sup>
		RES (Umbrella sampling, Metadynamics)	Assurance of reaching the final state	Results might depend on the choice of reaction coordinate and alter dynamics	Peptide amphiphiles <sup>34</sup> Crystallization <sup>43-48</sup> Supramolecular polymer <sup>49</sup>
		Enhanced/adaptive (Transition Path Sampling, Adaptive Sampling)	Correct reproduction of dynamics	No guarantee of reaching the final state	Crystallization <sup>50</sup> Virus capsid <sup>51</sup> Oligopeptides <sup>52</sup>
	Coarse-grained	Able to study large systems on a large timescale	Loss of resolution	Oligopeptides <sup>14,53-55</sup> Peptides amphiphiles <sup>56-58</sup> Self-assembly BTA <sup>59</sup> Bilayer formation <sup>60</sup> Organogel <sup>61</sup>	

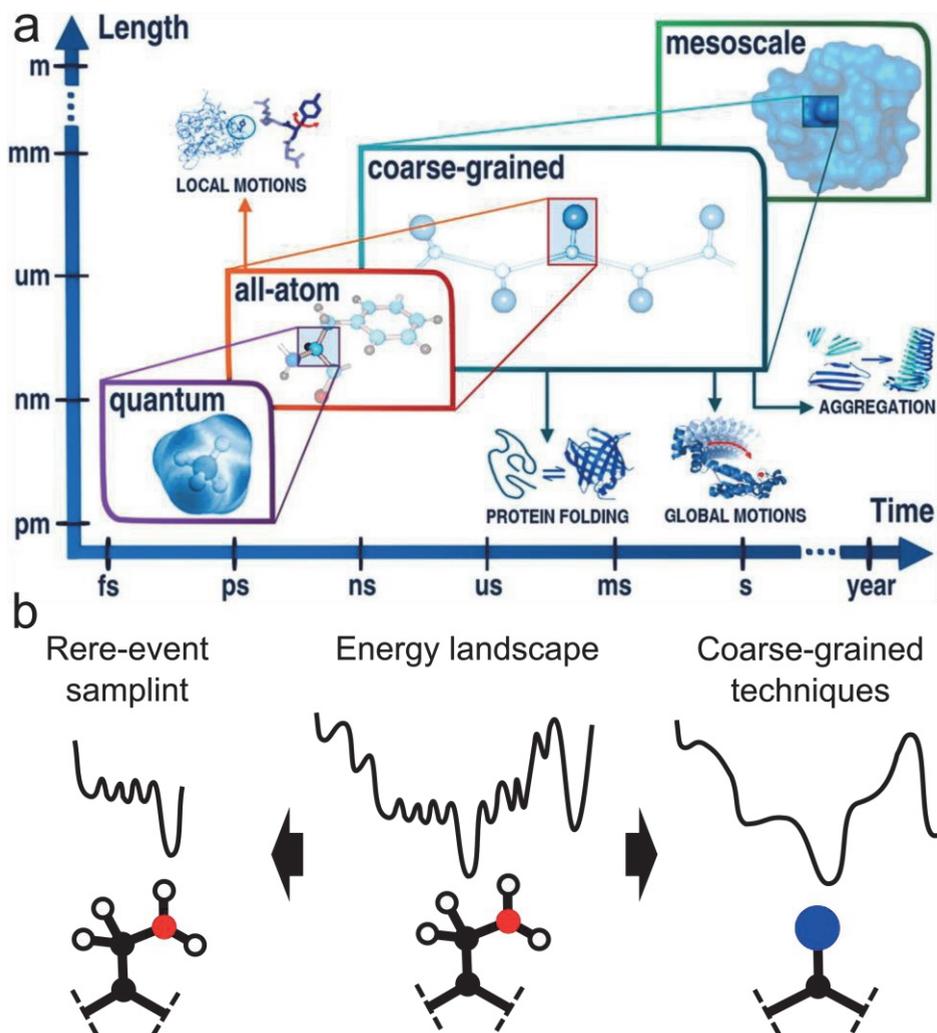


Figure 1.3. (a) Overview of computational methods on different temporal and spatial scale. Reproduced from reference<sup>31</sup>. (b) Schematic comparison between rare-event sampling methods and coarse-graining techniques. RES may give detailed all-atomistic insights into a small part of the energy landscape. CG techniques, on the other hand, allow to study a large part of the energy landscape, but at the expense of resolution.

Although a vast palette of the techniques is available, simulations of self-assembly are still challenging due to their large spatial and temporal scale<sup>29,31,62–65</sup>. Challenge of timescale arises due to the fact that self-assembly occurs on times ranging from nanosecond to weeks<sup>29</sup>. Although conventional all-

atomistic methods allow simulations on hundreds of nanoseconds timescale, this is still often shorter than the experimental timescale<sup>64</sup> (especially for processes involving nucleation, which happens on much longer timescale). Challenge of spatial scale arises from the size of the self-assembling systems, which often requires simulating hundreds of monomers<sup>62</sup>. Moreover, most of these systems contain a large amount of solvent, which also needs to be modeled. As a result, simulations of self-assembling systems are often beyond capabilities of conventional simulations.

This thesis focuses on the application of coarse-grained and rare-event sampling techniques to address the temporal and spatial challenges of self-assembling systems.

## **Research aims and outline of the thesis**

The aim of this thesis is to provide insights into the mechanism of self-assembly processes by showing the essential steps leading to the ordered structure. Due to the transient nature of the early stages of self-assembly, they are beyond the capabilities of experimental techniques. Therefore, in this thesis, we explore strategies of simulation of various self-assembling systems, which could provide insights into the process. In particular, in chapters 2 and 3, we study self-assembly of supramolecular fiber formation and in chapters 4 and 5, we describe our study on the self-assembly of long functionalized alkanes on a graphite surface.

In chapter 2, we explore the most common techniques used to give insight into self-assembling systems, i.e., (a) simulations of reorganization of randomly distributed molecules in a solvent, and (b) simulation of the assembly from the experimentally known structure. As a model example, we used derivatives 1,3,5-triamidocyclohexane (CTA), which belongs to a class of molecules creating supramolecular fibers by trifold hydrogen bonds, which recently are subject of many studies. The mechanism of formation of these fibers follows nucleation-growth. Therefore, the early stages of the self-assembly are challenging to study experimentally due to their transient nature. In this chapter, we run simulations with different force-fields to see differences between them. The results show two major issues which researchers can encounter during self-assembly simulations: (a) time scale and (b) choice of force-field. Early stages of self-assembly seem to be on a timescale far beyond the access of standard simulation. Thus, to successfully simulate these processes

use of methods which allow access to longer timescales, like coarse-grained techniques or rare-event sampling methods, might be necessary. Moreover, the researcher should be careful when choosing force-field and parameters, since self-assembling systems can be sensitive to the choice of the parameters.

In chapter 3, we tackle the time scale problem by developing adaptive sampling allowing the study of processes on an implied timescale, far beyond the access of conventional techniques. We use the CTA system with the force-field which showed the best reproducibility of stable fiber structure (see chapter 2). The method gives unique insights into the formation of the self-assembled fiber. Not only, have we been able to successfully simulate the process, but also we provide insights into pathway complexity and kinetics of the nucleation process. The method also turned out to give insights into processes which might follow the initial formation of a fiber. Although the techniques provide a level of detail which has not been accessible by other methods, at present, they are still limited to study small systems.

In chapter 4 we use coarse-grained simulations to give insights into self-assembly of long functionalized alkanes on graphite into a physisorbed monolayer. Although detailed insights into the final structure have been obtained experimentally, there is little known about the mechanism of formation due to the diffusive nature of the adsorbent. These systems are also challenging to study computationally due to temporal and spatial scale. In this work, we use coarse-grained techniques, which at the expense of a loss of atomistic detail, allow accessing temporal and spatial scale far beyond standard all-atomistic simulations. Using this technique we were able to show the mechanism of the nucleation. In summary, the mechanism of formation is temperature dependent. At low temperature, adsorbed molecules form independent, small, and stable ordered domains. At high temperature, adsorbed molecules form an unordered, liquid-like phase, which upon increase of surface coverage to a certain level, rapidly transforms into an ordered structure. Moreover, we were able to reproduce experimental structures of the simulated assembly, which shows the predictive power of the model.

In chapter 5 we give insights into the final stage of the self-assembly process of long functionalized alkanes on graphite, Ostwald Ripening. This process plays a vital role in improving the quality of long-range order. However, it happens on a computationally long timescale. Therefore to give insights we employ coarse-grained molecular dynamics. Our results show that

an essential part of the process is partial desorption from the surface which allows other molecules to rearrange. Moreover, the rate of partial desorption can be controlled by temperature, and therefore, the speed of the ripening.

Overall, this thesis contributes to a better understanding of self-assembling systems by using computational methods. The methods presented here give insights into the process on a temporal and spatial scale which is not accessible by experimental methods. This thesis shows the great potential of molecular modeling methods to gain insights into self-assembly, and therefore, gain better control over the process.

## Chapter 2

# Lessons learned from molecular dynamics of trifold hydrogen bonded fibers with different force-fields

In recent years, computational methods become an essential element of studies focusing on the self-assembly process. Although they provide unique insights, they face challenges, from which two are the most mentioned in literature: the temporal and spatial scale of the self-assembly. A less often mention issue, but not less important is the choice of the force-field. The repetitive nature of supramolecular structure results in many similar interactions. In consequence, even a small deviation in these interactions can lead to significant energy differences in a whole structure. However, studies comparing different force-fields for self-assembling systems are scarce. In this paper, we compare molecular dynamics simulations for trifold hydrogen bonded fibers with different force-field, namely GROMOS, CHARMM, CHARMM Drude, Martini and polarized Martini. Our results raise awareness of the importance of validation of the force field for self-assembling systems.

## Introduction

Self-assembled fibers have recently drawn attention, because of their rich, dynamic behavior, which is similar to that of many materials occurring in biological systems. Moreover, they often exhibit features which conventional polymers, connected by covalent interactions, do not have<sup>66</sup>. However, there is still little known about the mechanism of the formation of supramolecular fibers. Such knowledge would allow improving control over their structure and function<sup>67</sup>. Therefore, substantial effort is directed towards understanding the self-assembly process as a whole and steps in the self-assembly mechanism. However, self-assembly steps often occur on a temporal and spatial scale which is beyond the reach of experimental techniques.

Consequently, for processes on short time and length scales, many computational studies are devoted to supramolecular polymers, as can be seen in recent reviews by Frederix et al.<sup>29</sup> and Bochicchio et al.<sup>30</sup>. The main challenges of simulation of supramolecular systems are connecting the small computational systems to the large spatial and temporal scales of the experiment<sup>62</sup>. Most of the time, they involve large systems which form structures on timescales spanning from nanoseconds to weeks<sup>29</sup>. Such a timescale is often far beyond capabilities of current computational methods. These spatial and temporal challenges of supramolecular systems are often mentioned in literature in the context of molecular simulation<sup>29,31,62-65,68</sup> and are the focus of many studies<sup>29,30,69,70</sup>. However, the nature of supramolecular systems (i.e. the molecules are non-covalently connected by the same type of interactions) is such that small errors in a model describing an interaction are amplified by the number of molecules, as has been pointed out in the context of protein modeling<sup>71</sup>. Despite this fact, studies on how different force-fields influence the supramolecular structures are scarce<sup>29</sup>.

In this work, we study two currently standard approaches that are employed to give insight into self-assembly: simulation of spontaneous self-assembly starting from randomly distributed molecules<sup>41,49,53,72</sup>, and simulation starting from a pre-built model structure of the proposed final assembly<sup>33,34,38,73,74</sup>. As a model example, we use a derivative of 1,3,5-trisamidocyclohexane (CTA), which is known to create long ordered fibers upon self-assembly and for which there is a crystal structure of its analog<sup>75</sup>. CTA belongs to a large class of supramolecular molecular blocks which form fibers via trifold hydrogen bonding and recently are a subject of intensive

studies<sup>39,74,76–78</sup>. Here, we simulate self-assembly and stability of CTA fibers using different force-fields (namely: GROMOS<sup>79</sup>, CHARMM<sup>80–82</sup>, CHARMM Drude<sup>83–86</sup>, Martini<sup>87,88</sup>, and polarized Martini<sup>89</sup>). This work draws awareness of the common issues which occur during simulations of self-assembly.

## Results

**Attempts on spontaneous self-assembly.** The most common way of using Molecular Dynamics (MD) simulations is to run a simulation for as long as possible. In general, this approach should work because a system that starts from a non-equilibrium situation progresses on an energy landscape and explores it, finding local minima, which ideally correspond to the experimentally stable structures. For many systems, like proteins, it is hypothesized that energy landscape has the shape of a funnel<sup>90</sup>; therefore, on long enough simulation, exploring it will lead to the most stable structure being visited most often. In practice, however, one might end up in a local minimum and not escape from it. We have attempted simulation of supramolecular self-assembly by performing 100 ns simulations of 8 molecules in a small simulation box. Such length simulations have been successful in self-assembly simulations of amphiphiles in water<sup>53,55</sup>. We have simulated systems using different force-fields at different levels of resolution: the all-atom model including electronic polarizability CHARMM Drude model<sup>83–86</sup>, the parent all-atom CHARMM model<sup>80–82</sup>, the CHARMM model with modified charges obtained by mapping effective charges from the CHARMM Drude model (CHARMM mod.), the united-atom GROMOS model<sup>79</sup>, the coarse-grained (CG) polarized Martini (MartiniP) model<sup>89</sup>, and its parent coarse-grained model Martini<sup>87,88</sup>. The final snapshots of these simulations are shown in Figure 2.1a-f. Performance is the best for Martini and MartiniP, than two orders of magnitude slower for CHARMM and GROMOS, and three orders of magnitude slower for CHARMM Drude (see SI). In all simulations, molecules aggregate into a cluster as can be observed by visual inspection and from the solvent accessible surface area (SASA) (see Figure 2.1h). However, in none of them we have observed formation of long-range ordered structures, for which we measured the number of hydrogen bonds between amides of the CTA, as shown in Figure 1g. The most ordered structures were observed in MartiniP, where we could observe small ordered fragments (dimers and trimers). The variations in both number of hydrogen bonds and SASA as a function of simulation time informs about the dynamics of the system; well-defined SASA corresponds to well-defined structure. It can be seen that for GROMOS a compact and stable structure is

obtained, which is reflected in a low value of and small variation in SASA. CHARMM and CHARMM Drude have large variations and reflect rather large dynamics in the structure. Since the volume of the bead for CG models is different it is difficult to compare the SASA for them with atomistic force-fields, but similar trends are observed (that is, the Martini models form a fairly compact structure).

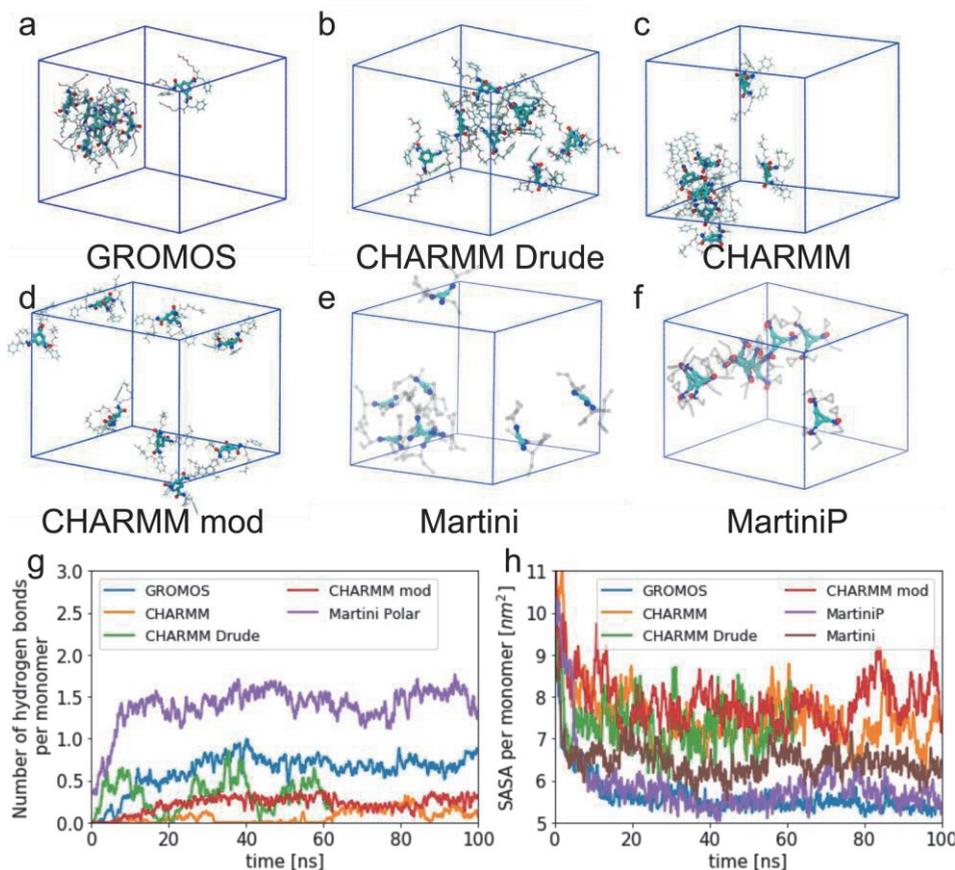


Figure 2.1. Simulations of eight molecules in small simulation box for simulations up to 100 ns for different force-field. Final snapshots of the simulations for: (a) GROMOS, (b) CHARMM Drude, (c) CHARMM, (d) CHARMM with charges obtained from mapping CHARMM Drude (see text and SI), (e) Martini, and (f) MartiniP. In none of the simulation we have observed formation of long-range ordered structures as can be seen. (g) Progression of number of amide-amide hydrogen bonds. (h) Progression of solvent accessible surface area (SASA).

**Fiber stability.** Success of conventional MD simulations relies on the assumption that the experimentally observed structure is in the global minimum

on the energy landscape obtained from the simulation. For a force-field that models interactions perfectly, simulations would lead to the experimentally stable structure for long enough simulation. However, self-assembled systems are challenging to model: even small errors in parameterization of the force-fields are important, since they are multiplied by number of molecular blocks. Therefore, one of the reasons why our spontaneous self-assembly simulations might not lead to experimentally observed final structure might be that fibers are not the global minimum for the chosen force-field. It is the second strategy commonly used in supramolecular structure modeling: assess the stability of a proposed architecture<sup>33,34,38,73,74</sup>. In order to check the stability of the supramolecular fiber, we have created a stack of 24 CTA molecules (see Figure 2.2a) from a known crystal structure of its analogue<sup>75</sup>. Then we have simulated the structures for ~100 ns using the same force-fields as in the previous paragraph. The final snapshots of ~100 ns simulations are presented in Figure 2b-g. To quantitatively measure the stability of the structure we analyzed the trajectories by calculating the number of hydrogen bonds between CTA amides (Figure 2.2h) and solvent accessible surface area (SASA; Figure 2.2i).

The fiber in GROMOS, CHARMM Drude and polarized Martini force-field stays in the ordered structure during the course of the simulations, as can be observed by visual inspection and from the constant number of hydrogen bonds (Figure 2h) and solvent SASA (Figure 2i). With the standard CHARMM force-field, the fiber collapses: most of the hydrogen bonds are immediately broken, and the structure rearranges into an unordered, compact agglomerate (see SASA in Figure 2.2i). Although, CHARMM Drude is able to sustain stable structure, it is computationally the most expensive force-field. Therefore, we checked if we can use the standard CHARMM force-field with modified charges obtained from the CHARMM Drude force-field that reflect the average polarization of the chemical groups in the stable assembly (see SI for details of backmapping of the charges). Although the fiber structure using this modified force-field with effective charges (CHARMM mod.) is more stable than in standard CHARMM, the fiber collapses in the course of the simulation. For coarse-grain force-fields, it seems that reproduction of directional interactions of amide groups is necessary to model a stable fiber: for Martini it collapses, whereas for MartiniP it stays stable. This contrasts with results of Boichichio et al. for BTA molecule, for which Martini and MartiniP yield similar results<sup>59</sup>. Simulations for NPT qualitatively lead to the same conclusions and are presented in SI.

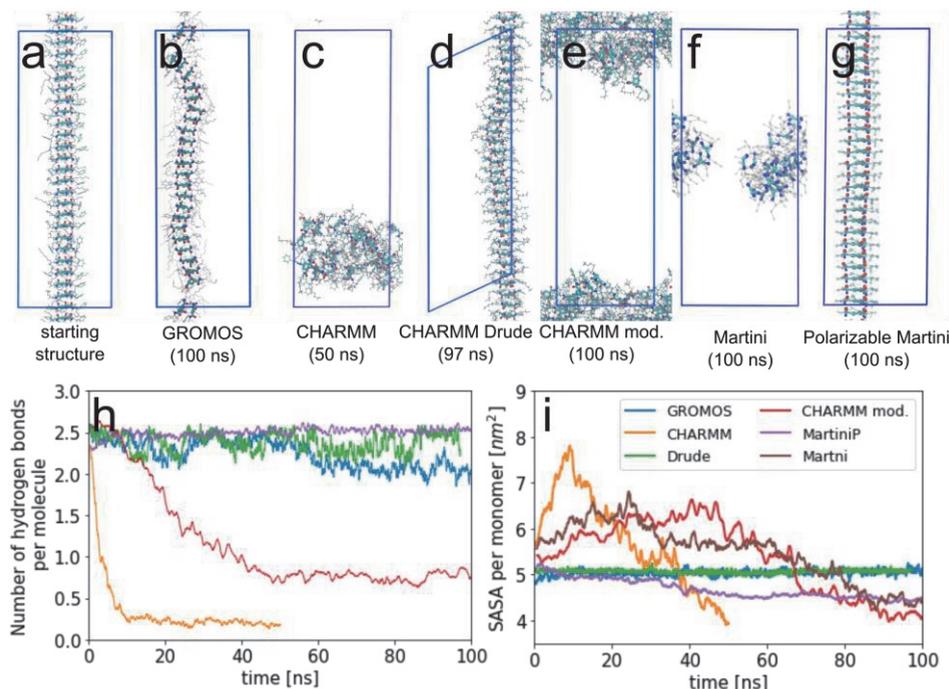


Figure 2.2. Simulation of the supramolecular fiber of CTA obtained from the crystal structure. (a) Starting structure. Single simulations (from 50 ns up to 100ns) of the fiber in different force-field: (b) GROMOS, (c) CHARMM, (d) CHARMM Drude, (e) CHARMM with charges obtained from a mapping from CHARMM Drude FF (see main text), (f) Martini, and (g) MartiniP. The most stable structures were obtained for GROMOS, CHARMM Drude and MartiniP, which can also be seen on graphs of the number of hydrogen bonds per molecule (h), and solvent accessible surface area (SASA) (i). For stable fibers, the number of hydrogen bonds and SASA are constant.

**Formation of the fiber in vacuum.** In order to investigate what interactions are responsible for stability of the fiber, we analyzed addition of molecules to a growing small stack and analyze the energy gain. The enthalpy of creation of dimer, trimer, etc. from monomers in vacuum is presented in Figure 2.3a. All force-fields show a strong cooperative effect. The gain in the energy per molecule upon addition of a further monomer is large for dimers until tetramers, and substantially slows down for pentamers and longer stacks. Although these trends hold for all force-fields, the differences between enthalpies for different force-fields also grows with size of the stack. The differences are significant, and for addition of a monomer to an 11-mer reaches  $\sim 70$  kJ/mol between MartiniP and CHARMM Mod. (and  $\sim 30$  kJ/mol for all-atomistic force-fields, i.e., between CHARMM Mod. and CHARMM Drude). Then, we analyzed the contributions of Van der Waals, electrostatic and bonded interactions to the

binding energy; these are presented in Figure 2.3b for the CHARMM Drude and GROMOS models. Analysis for the specific interaction shows that a cooperative effect is present in both coulomb and van der Waals interactions. Interestingly, the most important contribution for the GROMOS force-field comes from van der Waals interaction (Lennard-Jones, L-J), but for the CHARMM Drude force-field it comes from electrostatic interactions. Other atomistic non-polarizable force-fields give similar results (see SI). For coarse-grain force-fields the main contribution comes from L-J interactions. It is important to note, that Lennard-Jones interactions are short range (they decay with  $r^{-6}$ , where  $r$  is the distance between atoms) and therefore can only be weakly directional, but Coulomb interactions are long-range interactions (they decay with  $r^{-1}$ ) and therefore a combination of different partial charges can make these interactions highly directional. As a result, the driving force for self-assembly in simulation depends on the choice of force-field: for the force-fields studied here, for the coarse-grained ones the strongest interactions are L-J, for non-polarizable all-atomistic ones mainly L-J, with an important contribution of electrostatic interactions, and for the polarizable all-atomistic one the opposite is the case, i.e., mainly electrostatic with important contribution of L-J interactions. The choice of the force-field could be guided by the type of interactions which dominate in self-assembled structure.

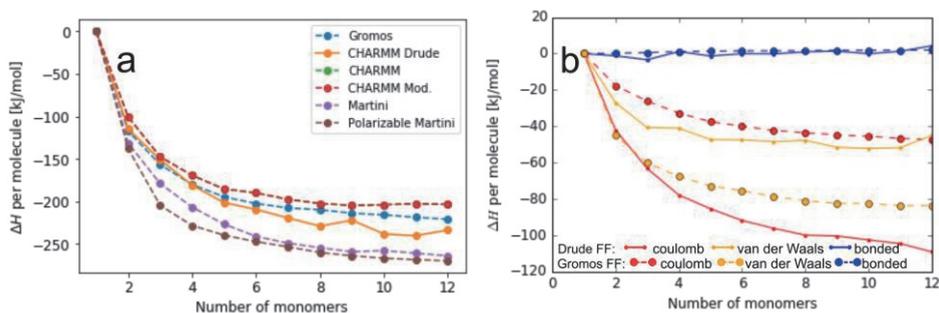


Figure 2.3. (a) Enthalpy of creation of dimer, trimer etc. per molecule. In all force-fields studied here molecules in stack show cooperative effect. (b) Decomposition of enthalpy of creation dimers, trimers etc. for GROMOS (large circles) and CHARMM Drude (small circles) force-fields (see others in SI). Although, the overall enthalpies are similar the contributions from coulomb (red) and van der Waals (orange) interactions are reversed. In Drude FF coulomb interactions are the strongest, whereas in GROMOS the van der Waals are strongest. Bonded interactions (blue) do not contribute to the binding.

**Formation of fiber.** We have shown that in some force-fields the fiber structure is stable; however, simulation from randomly distributed molecules did not lead to formation of small ordered structures. To investigate why simulations do not lead to stable structure we have performed simulations of the creation of a dimer from two free molecules and of a pentamer (small ordered fiber) from a free monomer and a tetramer. Division of self-assembly process into these single steps allows studying it on a much shorter timescale, hopefully accessible by conventional MD. To simulate dimerization, we have run multiple independent 10 ns simulations of two free molecules (see Figure 2.4a-d). To simulate pentamerization we have run multiple independent 10 ns simulations of a tetramer kept stable by position restraints on the cyclohexane core atoms with one additional free molecule (without restraints) (see Figure 2.4e-h). Simulations were performed using different force-fields: GROMOS and CHARMM Drude FF, CHARMM Mod. and MartiniP. We have calculated a 2D histogram of the distribution of the position of a single molecule with respect to the other molecule or the tetramer, measured over the final 1 ns. The position is characterized by the collective variables that reflect the distance between the centers of the two entities and the coordinate of the monomer along the stacking direction (defined as the z-direction). The simulations for different force-fields gave different outcomes. For CHARMM Drude two molecules do not create a stable dimer; however, addition of one molecule to a tetramer results in preferential attachment of the monomer to the end of the stack. For GROMOS, we observed a stable dimer; however, a free molecule attaches to the side of a tetramer rather than to its end. CHARMM mod. forms neither dimer nor pentamer. MartiniP formed both dimers and pentamers. These results show that one has to be careful when drawing conclusions about the expected outcome of self-assembly simulations on the basis of the outcome of monomer-monomer interactions. The most interesting outcomes are for GROMOS and CHARMM Drude. GROMOS shows that the tendency of a force-field to form stacked dimers from monomers does not indicate success in self-assembly simulation of multiple molecules into a small fiber. On growth of a small ordered stack there is also a chance of creation of new adsorption site. In GROMOS, the adsorption on the side of a small stack is apparently more favorable than at its end. In contrast, CHARMM Drude shows that dimers can be unstable, and only upon growth of the stack, a highly ordered structure becomes the more stable one due to increasing dipole polarization.

For long simulations which sample all states histograms such as shown in Figure 4a-h would be equivalent to a free energy landscape. However, the limited simulation time (10 ns) could mean that these histograms show local minima rather than global ones. To validate them, we calculated the difference in free energy of adsorption of an unbound monomer to the side of the fiber and to the end of the fiber using umbrella sampling. Due to lack of support of umbrella sampling for CHARMM Drude, we do not have results for this force-field. All force-fields prefer adsorption to the end of the fiber, and the difference between adsorption to the side and to the end of the fiber is similar for all force-fields, being at the level of  $\sim$ -20 kJ/mol. However, the difference in free energy between bound and unbound states is different for the different force fields. The largest difference is found for GROMOS, for which adsorption to the side and to the end is  $\sim$ -50 kJ/mol and -70 kJ/mol, respectively. During the self-assembly process, the accessible surface area for newly arriving molecules is larger at the side of the tetramer than at its ends. Therefore, we can anticipate that molecules initially adsorb on the side of the fiber and then migrate to its end. However, for GROMOS the strong adsorption to the side of the fiber (50 kJ/mol) might prevent an adsorbed molecule from desorbing or from moving along the side of fiber. Indeed, the 2D histograms presented before confirm that molecules adsorb preferentially on a side of the fiber, and during 10 ns simulations rarely desorb (see Figure 2.4f). This can also be seen from the mean square displacement of the monomer over last 1 ns of the tetramer-monomer simulations (see Figure 2.4j). From this calculation we can see that molecules that adsorb to the fiber do not move anymore for the GROMOS model. For other force-fields monomers have still some mobility. We can approximate the experimental value of the free energy of elongation by the pseudo-phase approximation<sup>19</sup>, which results in  $\sim$ 30 kJ/mol (see SI). Surprisingly, the force-fields which are the closest to this value, viz. CHARMM Mod. and Martini, do not lead to elongation of tetramer.

These results show how different behavior can be observed with different force-fields. Surprisingly, dimerization simulation gave us a little information of the outcome of spontaneous self-assembly in comparison to the formation of a pentamer from tetramer and monomer, and can therefore yield misleading insights into self-assembly. Formation of the pentamer gives important insights into formation of fibers; however, it requires *a priori* knowledge about the final self-assembled structure, that is the way in which the molecules stack. All force-fields give the same adsorption side for the formation of the pentamer. However, the levels of free energy of adsorption to the end of

the fiber and to its side vary between the force-fields. Some force-fields have deep local minima, from which escape is difficult if not impossible. Therefore, for successful simulation of self-assembly in some force-fields the usage of special techniques like rare-event sampling methods can be crucial.

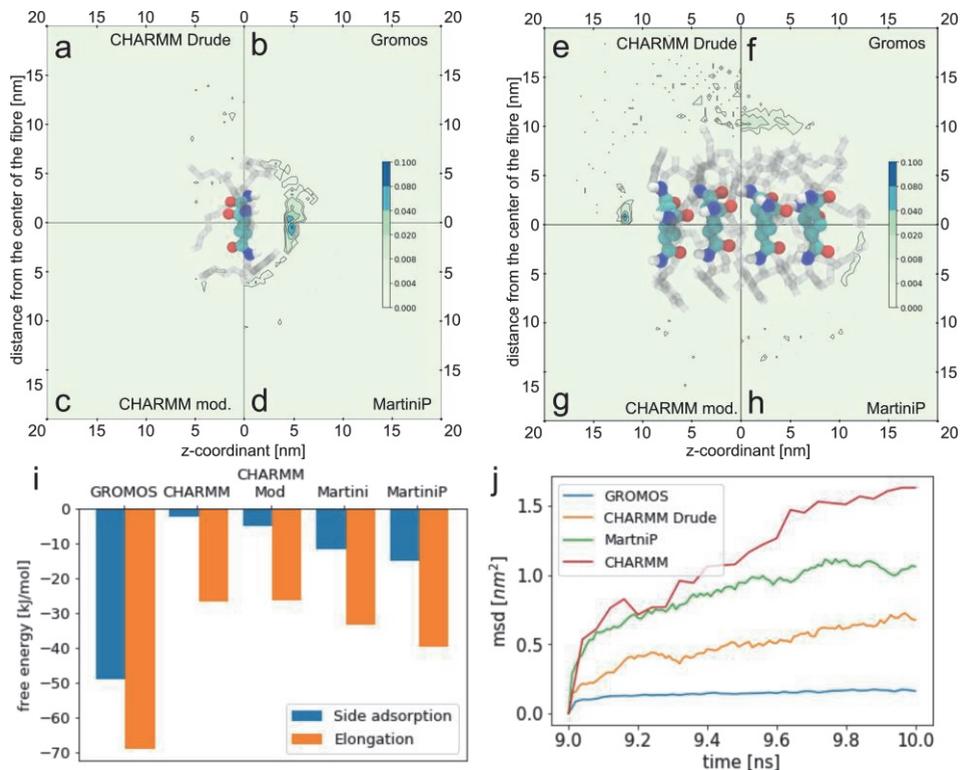


Figure 2.4. Simulation of monomer-monomer and monomer-tetramer pairs.(a-h) Histograms of positions in the last 1 ns of simulation of the added molecule to the system. The central molecule/tetramer shows a snapshot of the structure from simulations. Only the core of the molecules are shown, and sidechains are shown semi-transparent. (a-d) Histogram of distribution of two molecules around each other for different force-fields. In GROMOS and MartiniP molecules prefer to form dimers. For CHARMM Drude and CHARMM mod. (see main text) molecules actually do not form dimers. (e-h) Histogram of distribution of addition of one molecule to system with a tetramer (the tetramer is stabilized by position restrain atoms of the cyclohexane rings). For CHARMM Drude and MartiniP there is a preference to attach to end of the fiber (with much stronger preference for CHARMM Drude). For GROMOS, the monomer tends to attach to the side of the fiber. For CHARMM mod. there is no preferential attachment. It is important to note that (a-h) show results for short simulation (10 ns), and therefore show local minima rather than the global one. (i) Free energy of attachment of unbounded molecule s to the end of the fiber (orange) and to the side of the fiber (blue). The difference between energy levels is similar (~20kJ/mol), however, the energy levels for different force-fields vary substantially (up

to 40kJ/mol). There are no results for CHARMM Drude due of lack of support of umbrella sampling for this force-field. (j) Mean square displacement of molecule added to tetramer. For all force-fields except GROMOS, the molecules stay mobile.

## Conclusions

The work presented here provides insights into the reasons behind the success or failure of MD simulations of self-assembly systems. In particular, we study self-assembly and stability of CTA fibers using MD simulations with force-fields with different level of detail, namely coarse-grained Martini and MartiniP, all-atomistic GROMOS and CHARMM, and polarizable all-atomistic CHARMM Drude. This work shows crucial aspects which have to be considered when simulating self-assembly systems. In line with other research done in the field, the most challenging issue remains the timescale; even when the force-field reproduces anticipated behavior of the final structure (here GROMOS and CHARMM Drude), it does not guarantee the success of self-assembly during MD simulation. Moreover, *a priori* knowledge about the final structure might be crucial for tuning the force-field. It seems a good practice to validate force-field by simulating stability of the final structure. To our surprise, the study of dimerization of two molecules provides little information about the expected success of simulation. Here, four tested force-fields gave all possible outcomes of dimerization and elongation: upon formation of dimer, we observed elongation and its lack (for MartiniP and GROMOS, respectively); upon lack of formation of dimer, we observed also elongation and its lack (for CHARMM Drude and Martini, respectively). The driving force for self-assembly depends on the force-field. For Martini and MartiniP it L-J interactions (although the small electrostatic contribution in MartiniP is essential for stable, ordered structure). For all-atomistic force-fields it is the combination of L-J and electrostatic interactions, with L-J stronger for non-polarizable force-fields and electrostatic stronger for polarizable force-field. The awareness which interactions dominate in self-assembled structure might be a crucial criterion of choice of the force-field. However, often this choice is made by personal preference. Here we show that results of the simulations depend on this choice. This work demonstrates the importance of force-field validation for self-assembling process.

## Methods

Simulations were done with GROMACS, for non-polarizable force-fields version 5.1.2<sup>91</sup> and for polarizable CHARMM Drude modified version of

GROMACS allowing simulate using extended Lagrangian dynamics with a dual Nose–Hoover thermostat<sup>86</sup>. For different simulations we use different thermostats and barostats (see SI).

**Parameters for GROMACS.** Parameters of CTA molecules were obtained for GROMACS for different force-fields:

**GROMOS.** The parameters for GROMOS 53A6 force-field<sup>79</sup> were obtained using Automatic Topology Builder<sup>92</sup>.

**CHARMM.** The parameters for CHARMM General Force-field were obtained using `cgenff_charmm2gmx.py` script.

**CHARMM Drude.** The parameters for CHARMM Drude polarizable force-field were obtained on basis of existing parameterizations of small molecules. The final parameters for the CTA are included in SI.

**Martini and MartiniP.** The parameters for Martini and MartiniP were obtained according to the official parameterization tutorial available on the Martini FF website <http://cgmartini.nl>. See details in SI. In MartiniP, the amide bead was treated similarly to water bead: one bead with L-J potential, connected to two beads carrying charges. The charges were kept on opposite sides to connected bead by distance constrains.

**Hydrogen bonds.** We counted hydrogen bonds between amides groups using VMD and *HBonds* plugin. For all-atomistic force-fields we counted hydrogen bond if the distance between hydrogen donor and acceptor was below 0.33 nm and angle of donor-hydrogen-acceptor was below of 40°. For Martini FF we counted hydrogen bond if donor-acceptor distance was below 0.4 nm and the donor-hydrogen-acceptor angle of 40°. For more details see SI.

**SASA.** Solvent accessible surface area (SASA) has been calculated *gmx sasa* with probe radius 0.14 nm for all-atomistic force-fields and 0.265 nm for Martini-based force-fields.

# Supporting Information

## Methods

## Analysis

### Hydrogen bond analysis

Hydrogen bond analysis has been done using VMD and plugin HBonds. To set the criterion for hydrogen bonds we have calculated distribution of distance between hydrogen donor and acceptor and angle between donor-hydrogen-acceptor. We calculated the distributions for first 1 ns of fiber simulation for fibers, which visually seemed stable, i.e. CHARMM Drude, GROMOS and MartiniP (see Figure S2.1). From the distribution we established that for all-atomistic force-field we use the distance below 0.33 nm and angle below 40° and for MartiniP distance below 0.4 nm and angle below 40°.

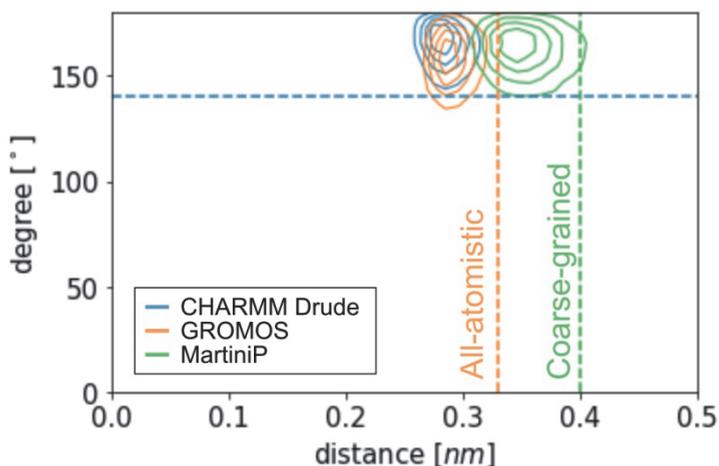


Figure S2.1. Distribution of distance between hydrogen acceptor and donor and angle of donor-hydrogen-acceptor for 1 ns of fiber simulation in CHARMM Drude, GROMOS and MartiniP force-fields.

## Parameterization

### Martini

CTA molecule (Figure S2.2a) has been parameterized according to official tutorial available on Martini force-field website. The bonded parameters has been parameterized on a basis of GROMOS simulation of a single molecule. We converted all-atomistic trajectory to coarse-grained by calculating center of mass of atoms creating beads. From such created coarse-grained trajectories we have extracted bonded parameters, that is bonds, angles and torsion angles used for coarse-grained force-field. The bead types were chosen on the basis of already parameterized molecules. We have tested several combinations of beads from which the best

result yielded topologies presented in Figure S2.2b for Martini and Figure S2.2c for MartiniP. The topology file is attached at the end of this document.

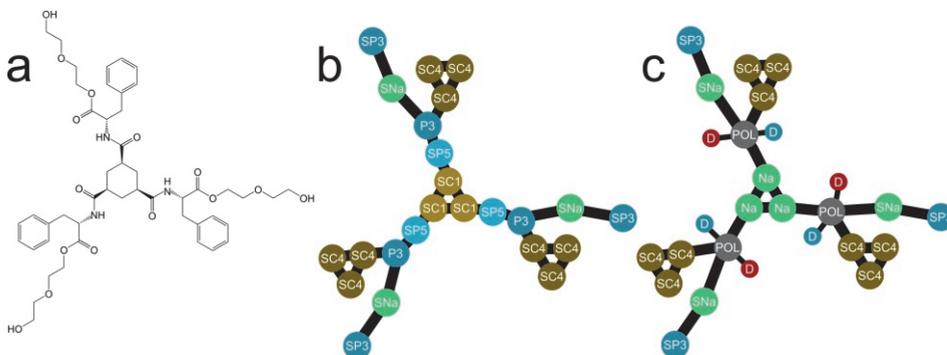


Figure S2.2. (a) Structure of CTA. (b) Martini representation of CTA. (c) MartiniP representation of CTA.

### CHARMM Mod.

Since CHARMM Drude showed that the structure of the pre-formed fibers were stable, we were interested to see if we can use the effective charges and find similar stability in simulations with the otherwise standard CHARMM force-field. We have constructed effective charges for the CHARMM model by redistributing the induced dipoles in the CHARMM Drude model over the atomic centers in a manner that preserves the local dipole moments optimally, akin to the Dipole Preserving Charge (DPC) introduced by Thole and van Duijnen<sup>93</sup>. The analysis as implemented here, calculates the induced dipole for each Drude atom as the charge of the Drude particle times the displacement of the Drude charge from its parent atom. The analysis aims to recreate this induced dipole by placing partial charges on all parent atoms in the non-polarizable model. The partial charges obey the restraints (common in many charge-fitting schemes such as RESP<sup>94</sup>) that the total charge is conserved and that the total dipole moment is conserved, but in addition, weight factors are introduced that favor placing the partial charges on the atoms nearest to the Drude charge that generates the local dipole. This idea is taken from the DPC analysis; the weighting scheme is exponential, i.e. the weight decreases exponentially with the distance of the atom to the Drude charge. The scheme is implemented in a Mathematica NOTEBOOK<sup>95</sup> as a matrix inversion problem. The effective charges were obtained by averaging over three snapshots of a stack of eight CTA molecules taken from an MD simulation in the CHARMM Drude model, and averaging over all equivalent atoms (three in each CTA) and snapshots (three). Table 1 reports the atoms whose charges were modified compared to the standard CHARMM. For these charges we show the base charge (atom charge in Drude model + Drude (mobile) charge of that atom, i.e. sum of the charges when they are on top of each other), the polarized charge (the change in the charge of the atom: the modified charge used is base charge + polarized charge), and the standard deviation in the polarized charge, calculated from the pooled measurements over CTA molecules in the stack and multiple snapshots.

Table S2.1. Mapped charge distribution for CHARMM Drude force-field for selected atoms.

	Element	Base charge	Polarized charge	St. dev. of polarized charge
Hydroxyl group	O	-0.46	-0.16	0.06
	H	0.36	0.07	0.03
Amide group	C	0.50	-0.04	0.04
	N	-0.38	0.03	0.11
	H	0.27	0.06	0.05
Ester group	C	0.70	0.04	0.04
	=O	-0.34	0.02	0.03
	-O-	-0.61	-0.01	0.03

Table S2.1 shows some insights into charge and polarization from CHARMM Drude for selected atoms. The most polarized atoms are the hydroxyl groups at the end of the side chains. Their polarization probably comes from interaction with water molecules. Amide groups, which we expected to have the strongest polarization due to the creation of an array of aligned hydrogen bond, also are polarized, but the polarization level is not close to that of the hydroxyl groups.

Our motivation of use of CHARMM Drude force-field in this work comes from work of Albuquerque et al., which have shown that amide polarization could be a crucial factor into stabilization of CTA type fibers. We have also calculated the interactions and dipole moments for the molecules from their works and results are presented in Figure S2.3. We have stacked molecules on top of each other (translated by 0.47 nm as in reference<sup>78</sup>), minimized the energy and calculated average dipole moment of the structure. The results for R1 are presented in Figure S2.3c. The dipole moment of molecules in the stack which were simulated using GROMOS force-field is constant, which contrast with results presented by Albuquerque, which showed that with increasing number of monomers also the dipole moment increases. The increased dipole moment of molecules in the stack could result in increased electrostatic interactions between monomer. Therefore, we have tested a polarizable CHARMM Drude force-field, which can take into account the polarizability of atoms. It can be seen in Figure S2.3c that for this force-field that upon increasing the number of monomers the dipole moment also increases and it reproduces similar trend as for the results reported in the literature.

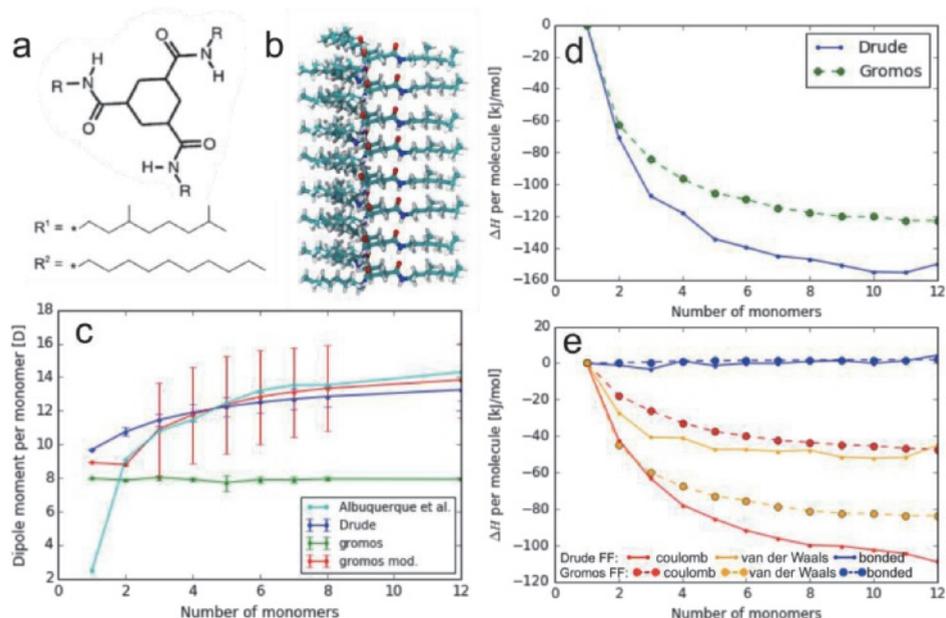


Figure S2.3. Analysis of fiber formation in vacuum for a molecule from work of Albuquerque et al., which shown that amide polarization could be crucial for stability of CTA type molecules.

## Simulations

### Protocol

Simulations in this work have been performed using GROMACS. The simulation procedure for different force-fields followed the scheme: (a) energy minimization, (b) short equilibration in NVT ensemble, (c) short equilibration in NPT ensemble, and (d) produce run (depending on the process in NVT or NPT ensemble). CHARMM Drude produce runs were performed only in NVT (simulations in NPT are not efficient). We used different barostat and thermostats, time step and number of steps for different force-fields. The parameters are presented in Table S2.2, Table S2.3, and Table S2.4.

Table S2.2. **Simulations of spontaneous self-assembly.**

		GROMOS	CHARMM	CHARM M Drude	CHARM M Mod	Martini	Martini P
Energy minimalization	Algorithm	Steepest	steepest	steepest	Steepest	Steepest	steepest
	Steps	5000	50000		50000	10000	10000
NVT ensemble	thermostat	berendse	berendse		berendse	v-rescale	v-rescale
	Steps	5000	5000		1000	5000	5000
	dt	0.0005	0.0005		0.0005	0.01	0.03 ps
NPT	Thermostat	berendse	V-rescale	V-rescale	V-rescale	v-	v-

ensemble	at	n				rescale	rescale
	barostat	berendse n	Parrinell o- Rahman	Parrinell o- Rahman	Parrinell o- Rahman	berends en	berends en
	Steps	5000	5000	10000	5000	1000	5000
	dt	0.0005	0.002	0.001	0.002	0.01	0.03 ps
RUN	thermostat	berendse n	V-rescale	Nose- Hoover	V-rescale	v- rescale	v- rescale
	Barostat	berendse n	Parrinell o- Rahman	None (NVT ensemble )	Parrinell o- Rahman	berends en	berends en
	Steps	5000000 0	5000000 0	1000000 00	5000000 0	333333 4	333333 4
	dt	0.002	0.001	0.001	0.002	0.03	0.03

Table S2.3. Simulations of the fiber

		GROMO S	CHARM M	CHARM M Drude	CHARM M Mod	Martini	Martini P
Energy minimalizati on	Algorith m	Steepest	steepest	steepest	Steepest	Steepest	steepest
	Steps	5000	50000		50000	10000	1000
NVT ensemble	thermost at	berendse n	berendse n		berendse n	v- rescale	v- rescale
	Steps	10000	5000		5000	5000	5000
	dt	0.0005	0.0005		0.0005	0.01	0.03 ps
NPT ensemble	Thermos tat	berendse n	V- rescale	V- rescale	V- rescale	v- rescale	v- rescale
	barostat	berendse n	Parrinell o- Rahman	Parrinell o- Rahman	Parrinell o- Rahman	berends en	berends en
	Steps	10000	5000	10000	500000	v- rescale	5000
	dt	0.0005	0.002	0.001	0.002	0.01	0.03 ps
NVT ensemble	thermost at	berendse n	Berendse n	Nose- Hoover	V- rescale	v- rescale	v- rescale
	Barostat	none	none	none	none	none	none
	Steps	1000000 00	5000000 0	1000000 00	5000000 0	333333 4	333333 4
	dt	0.001	0.001	0.001	0.002	0.03	0.03

Table S2.4. Simulations of dimers and pentamer

		GROMO S	CHARM M	CHARM M Drude	CHARM M Mod	Martini	Martini P
Energy minimalizati on	Algorith m	Steepest	steepest	steepest	Steepest	Steepest	steepest
	Steps	5000	5000		50000	10000	20000

NVT ensemble	thermostat	berendse n	berendse n		berendse n	v- rescale	v- rescale
	Steps	5000	5000		5000	5000	5000
	dt	0.0005	0.0005		0.0005	0.01	0.03 ps
NPT ensemble	Thermostat	V- rescale	V-rescale	V-rescale	V-rescale	v- rescale	v- rescale
	barostat	Parrinell o- Rahman	Parrinell o- Rahman	Parrinell o- Rahman	Parrinell o- Rahman	berends en	berends en
	Steps	5000	5000	5000	5000	1000	5000
	dt	0.002	0.002	0.001	0.002	0.01	0.03 ps
RUN	thermostat	V- rescale	V-rescale	Nose- Hoover	V-rescale	v- rescale	v- rescale
	Barostat	Parrinell o- Rahman	Parrinell o- Rahman	none	Parrinell o- Rahman	berends en	berends en
	Steps	5000000	5000000	1000000 0	5000000	333334	100000 0
	dt	0.002	0.002	0.001	0.002	0.03	0.01

### Position restraints

The simulations of the formation of pentamers have been done by adding one molecule to tetramer system. The tetramer has been created from the crystal structure. The tetramer has been kept together by putting position restraints on three atoms of cyclohexane rings (the strength of the restraints is 100 kJ/mol/nm<sup>2</sup>).

### Performance

Table S2.5 shows performance of the simulation for different force-fields for simulation of spontaneous self-assembly.

Table S2.5. Performance of MD for simulation of self-assembly of eight molecules for different force-fields measured for 10 ps simulations of fiber on single core of Intel® Core™ i7-5600U CPU @ 2.60 GHz.

Force-field	Performance [hours/ns]
CHARMM Drude	28.088
CHARMM	8.274
GROMOS	7.660
MartiniP	0.053

### Additional results

According to the pseudo-phase approximation, the difference in standard chemical potential between monomer and aggregate of size N, depends on the critical aggregation concentration (CAC) and can be expressed as<sup>19</sup>:

$$\Delta G = (\mu_1^0 - \mu_N^0) = RT \ln(X_{CAC}) \quad (\text{S2.1})$$

We approximate CAC by critical gelation concentration (CGC), which for CTA molecules is  $0.36 \text{ mM}^{75}$ . Therefore, the difference of the potentials  $\Delta G \approx 30 \frac{\text{kJ}}{\text{mol}}$ . This potential shows the difference of chemical potential between monomer and molecule in the aggregate, therefore it can be interpreted as free energy of elongation studied in the main text.

### Simulation in NPT

We have performed simulations of fiber stability also in the NPT ensemble (with semi-isotropic pressure coupling). The results are presented in Figure S2.4 and quantitatively are similar to results for the NVT ensemble. There are no results for CHARMM Drude due to the lack of an efficient way of solving the equations of motion in NPT ensemble. Similarly as in NVT ensemble the fiber is stable in GROMOS and MartiniP, and it disassembles in CHARMM, CHARMM Mod. and Martini as it can be seen from the progression of the number of hydrogen bonds (see Figure S2.4g). The solvent accessible surface area is less informative for trajectories in NPT (see Figure S2.4h).

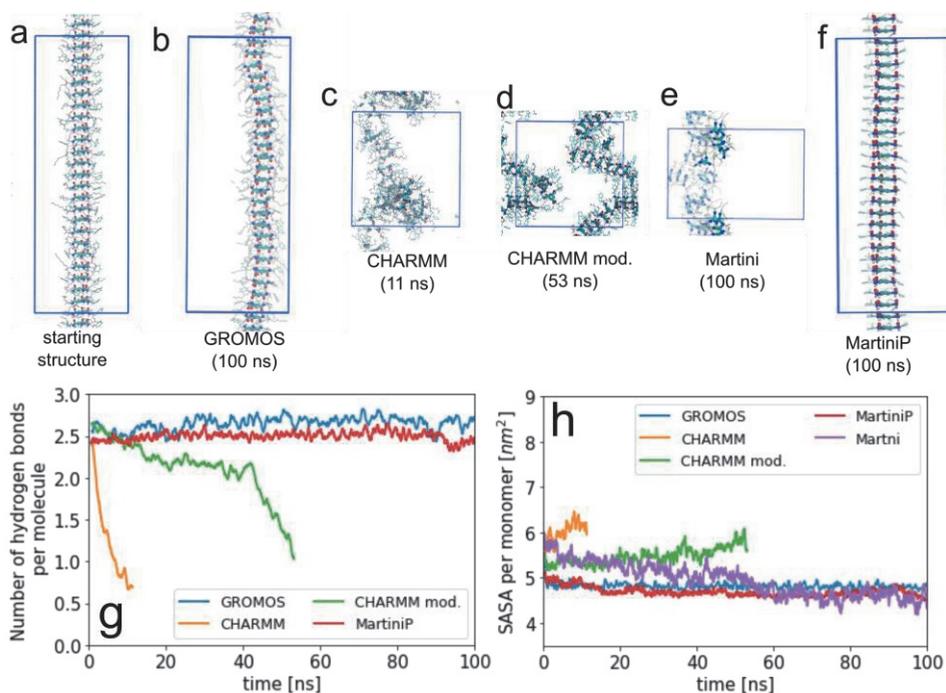


Figure S2.4. Simulations of fibers stability in NPT ensemble with semi-isotropic coupling. (a) Starting configuration. Final snapshot from simulations for (b) GROMOS, (c) CHARMM, (d) CHARMM Mod., (e) Martini, (f) MartiniP. (g) Progression of a number of hydrogen bonds between amide groups. (h) Progression of solvent accessible surface area.



## Chapter 3

# Fiber formation seen through the high resolution computational microscope

Supramolecular fibers draw attention because they can create complex materials exhibiting rich and dynamic behavior. They are widely present in biological systems, where they are responsible for normal cellular life (for example cellular integrity), but also for abnormal processes such as degenerative diseases (for example Alzheimer or prion diseases). One of the most remarkable facts of the supramolecular chemistry is that the information about the supramolecular structure is encoded in the single molecular block. However, a complete understanding on how this information translates into supramolecular arrangement requires not only a better insight into the local and global minima at the energy landscape of supramolecular structures, but also the possible routes across this landscape. Unfortunately, the route taken by the self-assembly process is governed by the kinetics of self-assembly, which is notoriously difficult to assess by experimental methods. An alternative approach to gain insight into the energy landscape and the possible routes across is to explore the energy landscape by molecular modelling. Here, we study the formation of 1,3,5-cyclohexanetricarboxamide fibers using a variation of Adaptive Markov Modeling with polarizable CHARMM Drude force-field. From our results, we were able to study primary nucleation, elongation, secondary nucleation and bundling occurring during the formation of the fiber. Our results demonstrate that with our method can provide unique insights into the kinetic pathways of fiber formation.

## Introduction

Early stages of self-assembly recently gained much interest due to their influence on the final self-assembled structure<sup>23,96</sup>. However, insights into these stages are scarce<sup>97</sup>, and just a few limited experimental results are present<sup>98</sup>. The most accepted mechanisms of self-assembly are isodesmic and nucleation-growth formation<sup>25-27</sup>, which give the generic pathway of the formation. More detailed insights, especially on a short timescale and molecular length scale, such as how such a nucleus is formed or what processes follow after a stable nucleus is established, are scarce because of transient nature of the process and the experimental limitations<sup>29</sup>, which most of the time do not allow study single processes, but their ensemble. Therefore, recently, many computational studies have been devoted to these systems<sup>29,30</sup>. The majority of these studies employ coarse-grained molecular dynamics (CG MD) models, which allow studying this process on long enough simulation timescales, at the cost of molecular resolution. The limited resolution of CG models is usually subject to the criticism that the models do not adequately describe the subtle interactions that lead to the fibers and therefore may not faithfully represent the stability and structural features of the fibers under study.

This work aims to give insights into the early stages of self-assembly of supramolecular fibers. Here, we study an atomistic and polarizable model of 1,3,5-cyclohexanetricarboxamide (CTA, see Figure 3.2a for the structure formula) molecules in water, which create columnar stacks via trifold hydrogen bonding between neighboring molecules<sup>75</sup>. Molecules forming such a trifold hydrogen bonding stacks (especially derivatives of 1,3,5-benzenetricarboxamide (BTA)) have recently become the focus of many experimental<sup>76,77,99</sup> and computational studies<sup>39,49,59,74,78,100</sup>. Both types of studies showed that the BTA supramolecular polymer is formed in a cooperative fashion via a nucleation and growth mechanism. Recent work of Bocchacio et al.<sup>59</sup> on BTA molecules on coarse-grained simulations showed that the molecules initially form the unordered aggregate which slowly reorganizes to more ordered structures. Since the formation of nuclei when using a model at atomistic resolution happens on a timescale beyond that of standard computational techniques, we use adaptive sampling and Markov state model (MSM) to model the process. MSM allows giving insights into processes beyond timescale on conventional molecular dynamics<sup>101-103</sup>. It also has been used to study self-assembly<sup>51,52</sup>. Using this method we were able not only to give insights into early stages of self-

assembly, but also about later stages such as elongation and secondary nucleation.

## Approach

Simulations of formation of supramolecular fiber in water by all-atomistic molecular dynamics (AA MD) are still challenging, because of the significant computational cost<sup>74</sup>. Moreover, in our previous work<sup>104</sup>, we have observed that simulations of self-assembly with the standard force-fields, such as CHARMM and GROMOS, leads to very stable, unordered aggregate. It seems that more expensive CHARMM Drude force field, which explicitly models atom polarizability, exhibits more dynamics behavior which could lead to the fiber formation. Study process of formation of a fibril consisting of several fibers requires simulations on large spatial and temporal scale, which seems to be not feasible for standard AA MD. To tackle a large size of the system, we have separated process into several smaller elementary processes and to address the long timescale we have used an adaptive sampling.

To tackle a large spatial scale, we have separated the whole process into several distinct phases, inspired by protein fiber aggregation<sup>13,105</sup>. In general, the formation of protein aggregates follows complex pathway consisting of many different transitions occurring simultaneously<sup>106</sup>. However, the elementary processes making up the pathway to fibrils are established (see Figure 3.1a-e). Here, we briefly summarize these results<sup>13,105,106</sup>. The self-assembly starts with the creation of a nucleus. Formed nucleus grows mainly by elongation, in which new molecules arrive at the end of the fiber by diffusion motion, but in rare cases, it can grow by fibril-fibril association. Reverse processes, dissociation, and fragmentation, are possible, but due to strong interactions between proteins are negligible<sup>106</sup>. Fragmentation typically is induced by thermal or mechanical forces. Additionally, the new nuclei can be formed by nucleation catalyzed by the existing surface of the fiber, called here secondary nucleation, which leads to the formation of a bundle of fibers<sup>106</sup>. In addition, bundles can be created by the aggregation of fibers with different shapes and organizations<sup>10</sup>. We would like to stress that although we have studied these processes separately, in the real system, all of these processes happen simultaneously creating a very complex pathway. Separating the process into distinct phases saves computational cost and allowed us to study processes such as elongation and bundling. However, even for small system nucleation was not observed upon long (100 ns) conventional MD simulation<sup>104</sup>.

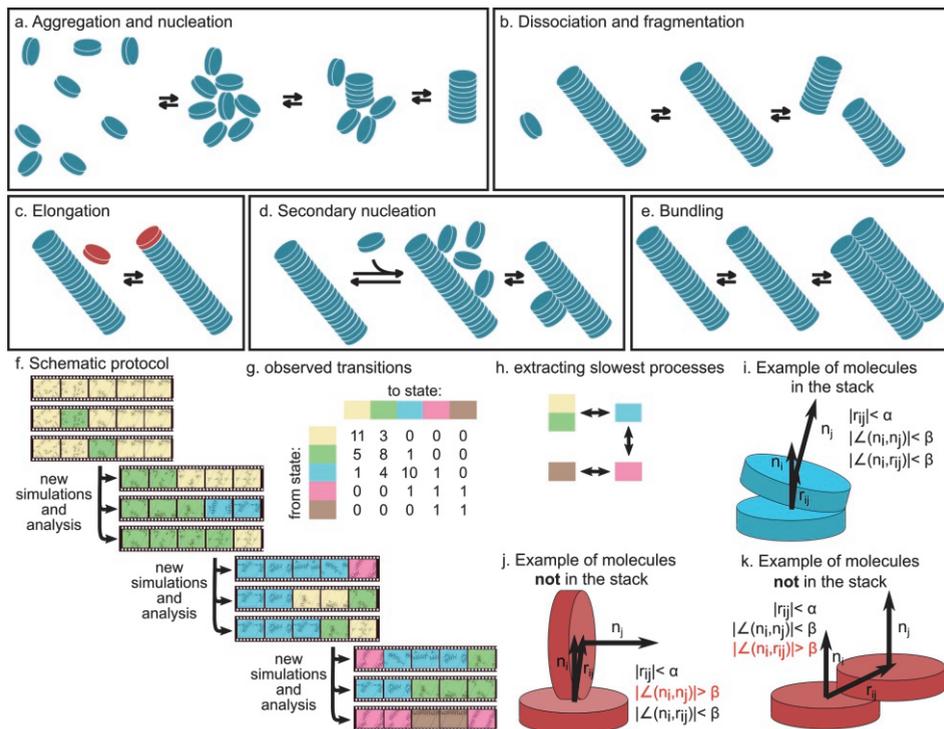


Figure 3.1. (a-e) Schemes summarizing processes studied in this work: (a) aggregation and nucleation, (b) dissociation and fragmentation, (c) elongation, (d) secondary nucleation, and (e) bundling. (f-j) Graphical explanation of simulation protocol used in this work. (f) Consecutive sets of simulations are run, and snapshots are analyzed to assign each frame as belonging to a distinguishable state, here represented by coloring each frame. In this work states are characterized by the number of molecules in an ordered stack (details in the SI). Each new set of simulations starts from snapshots taken from the collection obtained thus far, biasing the choice in favor of members of less visited states. (g) Transitions between observed states are counted and collected in a transition matrix, which leads to a Markov State Model that reveals the slowest kinetic pathways between states, grouping states that rapidly interconvert (here yellow and green) into one Markov state. (i-k) Two molecules are in the stack if the distance, the angle between normal vectors  $\vec{n}_i$  and  $\vec{n}_j$ , and the angle between the normal vector  $\vec{n}_i$  and the vector connecting centers of two rings  $\vec{r}_{ij}$  are in certain range (see SI for details). Examples of the molecules in the cluster (i) and not in the cluster, because of the angle criterium (j-k).

To simulate processes occurring on a long timescale (i.e., nucleation), we have applied adaptive sampling which is based on sampling the least visited states (see SI for more detailed description; the graphical explanation is present in Figure 3.1f). The method is similar to Adaptive Markov Modeling<sup>101-103</sup>. In principle, in our method, we run many parallel simulations starting from the

least visited state, which allows us to explore less visited parts of the energy landscape. Eventually, new states are visited. By repeating this procedure many times, we hope to sample the complete available conformational space. For the nucleation process, we have measured the size of the largest stack to distinguish different states (see Figure 3.1i-k). As a result, we obtained an ensemble of simulations consisting of all possible states and transitions between them. Most importantly, the ensemble also includes a completed fiber consisting of all molecules organized in an ordered array. Moreover, the combination of all trajectories allows us to study the pathways of the self-assembly by Markov State Modeling (MSM).

## Results and discussion

**Primary nucleation.** We have simulated formation of small CTA stacks from randomly distributed molecules in water with total simulation time  $6.7 \mu\text{s}$  using adaptive sampling and analyzed by MSM described above. The analysis showed six slow processes and the results are presented in Figure 3.2. Initially, molecules aggregate into a cluster, in which every molecule is in direct contact with four to five neighbors on average, see Figure 3.2k). Molecules in the aggregate are unordered, as can be seen from the low dipole moment and the low number of hydrogen bonds present (see Figure 3.2h), which suggests that molecules interact mostly by non-directional interactions. Thus, Van der Waals and hydrophobic interactions drive the first part of the formation process. Although most molecules stay in the aggregate throughout the majority of the simulations, they still have the flexibility to rearrange within it. Such rearrangements occasionally lead to a small ordered stack, whose lifetime depends on the length of the stack (see below). The ordered stack grows by a stepwise mechanism, i.e., monomers form dimers, a dimer and a monomer form a trimer, etc. Just a few other events that lengthen the ordered stack (e.g. dimer-dimer association into a tetramer) were observed, but these are not statistically relevant. As the ordered stack grows, the dipole moment and the number of CTA-CTA amide hydrogen bonds steadily grow, whereas the number of molecules in direct contact decreases, reaching two for the largest ordered fiber that can be reached in the simulation. The analysis and a video of the fastest possible process are presented in the SI.

The MSM analysis of the trajectories gives some insights into the kinetic pathway of the creation of a small ordered stack. Unfortunately, the model depends on the choice of the time step (see SI); therefore, the exchange

rates can be interpreted only semi-quantitatively and are shown here only to illustrate the power of the method rather than give accurate values. Moreover, for the MSM analysis, we have excluded all trajectories which led to the formation of an infinite fiber (i.e., a fiber that crosses the periodic boundary conditions connecting to itself, see Figure 3.2g), which on the timescale of the simulations performed here is irreversible (we have not observed disassembly of such an infinite fiber). We used the transition matrix of the MSM analysis to generate sequences from the least ordered state to the ordered stack of eight molecules, which on average takes  $58 \mu\text{s}$ , showing that the process is slow (see Figure 2b) on the MD timescale. The probability that in 100 ns of conventional MD simulation we observe a process of formation of an eight-molecule fiber is 0.0014, i.e., from 100 ns simulations, only 1 out of 714 yields an eight-molecule fiber.

The distribution of the stationary states shows the unordered agglomerate as the predominant state (reflected by the size of the circles representing the Markov states in Figure 3.2). However, from experimental work, we know that the fibers are stable. We can suggest a possible reason for why the unordered state is the most populated one in the MSM analysis. The favorable disassembly of the ordered stack could be a result of the very small number of molecules in the simulation; although simulation starts with a high concentration of unordered molecules, upon formation of the ordered stack, the concentration of unordered molecules decreases, reaching zero for the fiber, providing a driving force for its disassembly. In the macroscopic reality, the concentration of monomers can be assumed constant at equilibrium.

To further study the stability and critical size of formed nuclei, we have performed a series of simulations of small ordered stacks of different size in water. For every stack size, we run a series of 10 ns simulations (see SI for details) and measure the time of dissociation of a molecule from the stack. Since systems contain only the stack and water, in all systems equilibrium is shifted towards dissociation. Using the Kaplan-Meier estimator, we calculate the survival rate<sup>107</sup>; results are presented in Figure 3.2j. All dimers and most trimers dissociated within the 10 ns of simulation. More than half the simulations of ordered stacks with five and more molecules stayed intact during this time, and the survival fraction for them is similar, showing that the cooperativity effect reaches a maximum for such sizes. Thus we can anticipate the smallest stable nucleus (i.e., critical nucleation size) is around four-five molecules. In most of

the simulations, if dissociation takes place, it occurs at the end of the fiber (Figure 3.2o-n). However, we once observed fragmentation of fiber into two small fibers (see Figure 3.2o-p). Fragmentation is probably unlikely because the cooperative binding is strongest in the middle of the fiber. Although fragmentation is rare (the one event was observed in a total of 115 simulations of stacks with more than 5 molecules), it is also a way to generate new nuclei (Figure 3.1b).

**Elongation.** After the successful application of the adaptive Markov state method to the formation of critical nuclei, we wanted to give insight into the next step of fiber formation: elongation. Although in a sense growth of the ordered stack is observed in the simulations with eight molecules, an unambiguous elongation process is difficult to define and distinguish from forming a stable nucleus. We, therefore, started by adding one molecule to a system containing a 16-mer fiber with two free ends and simulate it in a series of 21 simulations of 10 ns each (Figure 3.3a). In most of them, the free molecule does not reach the vicinity of the end of the fiber. We observed only one elongation event of the added molecule on one of the ends of the fiber. However, interestingly, in most of the simulations, the added molecule does adsorb to the side of the fiber, and tends to stay near its surface and to laterally diffuse along it (see Figure 3.3a-e). To quantify this diffusional process, we ran ten independent 15 ns simulations of an infinite fiber and one free molecule. In all cases, the initially free molecule adsorbs on the surface of the fiber within 6 ns (Figure 3.3f). With seven simulations the absorbed molecule stays in the vicinity of the surface until the end of the simulation, while we also observed three events of desorption (see SI).

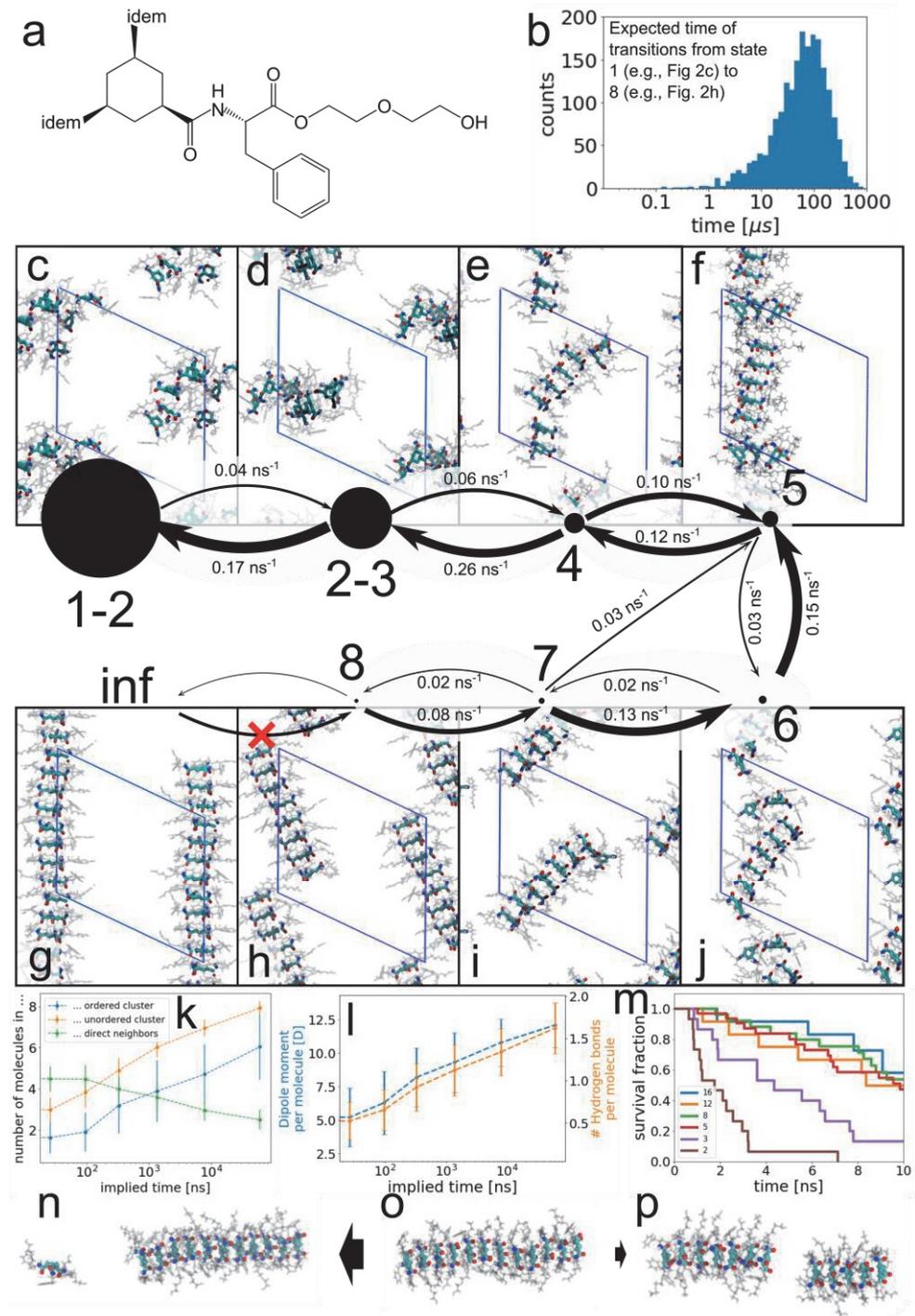


Figure 3.2. Primary nucleation. (a) Chemical structure of 1,3,5-cyclohexanetricarboxamide (CTA). (b) Histogram of expected simulation time for reaching eight molecules in an ordered stack, starting from randomly distributed molecules. One thousand trajectories were generated from the transition matrix obtained from MSM analysis of MD trajectories. The average time is 58  $\mu$ s. (c-j) Markov State Model. Different states are indicated by filled circles, whose size is proportional to the population of the state in the equilibrium ensemble. The arrows between the states represent rates of transitions between them, with the numerical values written above or below them. The numeral(s) by a circle represent the number of molecules in the ordered stack characterizing the Markov state. Although not included in the MSM analysis we show here the infinite fiber (indicated by *inf*). The pictures show one example snapshot for each state. For clarity, the trisamidocyclohexane core is shown in thick lines, side chains are semi-transparent, and water is not shown at all. Boundaries of the simulation box are indicated by blue lines. (k) Progress of the size of ordered (blue) and unordered (orange) cluster, and of the average number of neighbors of a CTA molecule (green). (l) Progress of the magnitude of the total dipole moment of all CTA molecules (blue) and number of CTA-CTA amide H-bonds per CTA molecule. (m) Survival analysis of series of simulations starting from ordered stacks of 2, 3, 5, 8, 12 and 16 molecules. During the simulations all dimers and most of the trimers disassemble. Stability of the ordered stack is similar for five and more molecules. (n-p) Example snapshots from simulations starting from ordered stacks of size 12: (n) if disassembly occurs most of the time it happens via dissociation of one molecule from the full stack (o); however, we once observed fragmentation into two parts (p).

We also found that the binding on the surface of the fiber is not strong enough to immobilize the adsorbed molecule. Analysis of the diffusion showed that a molecule on the fiber covers the similar distance in the mean-squared sense in the direction parallel to the main axis of the fiber as a free molecule in solution (see Figure 3.3g). The mean-squared displacement in the direction perpendicular to the fiber levels off, reflecting the fact that the molecule remains bound to the surface and movement resembles diffusion in confinement. We anticipate that the elongation progresses as follows: free molecules randomly diffuse in solution and eventually they encounter a fiber on which they adsorb. Then, they diffuse along the surface of the fiber, eventually reaching the end of the fiber, where they can adsorb to a free end. Adsorption anywhere on the fiber, followed by diffusion along the fiber is a reasonable proposition to explain the formation of fibers on the timescales observed. The alternative, in which monomers diffuse freely in the solution until they find a free end of a fiber and adsorb there, would take a very long time (especially for gelators which critical gelation concentrations are extremely low), especially in the later stages of gelation when few free molecules are left. The quantitative comparison of these two models can be found in the SI. Probably the diffusion along the fiber is crucial for low-molecular gelators to gelate in a reasonable time (such that an existing fiber does not dissolve before the new molecule

arrives) or it might even be the process which distinguishes low-molecular-weight gelator from other types of gelators. It is noteworthy that movement along the fiber consisting of large oligopeptides was recently also observed by Frederix et al.<sup>108</sup>.

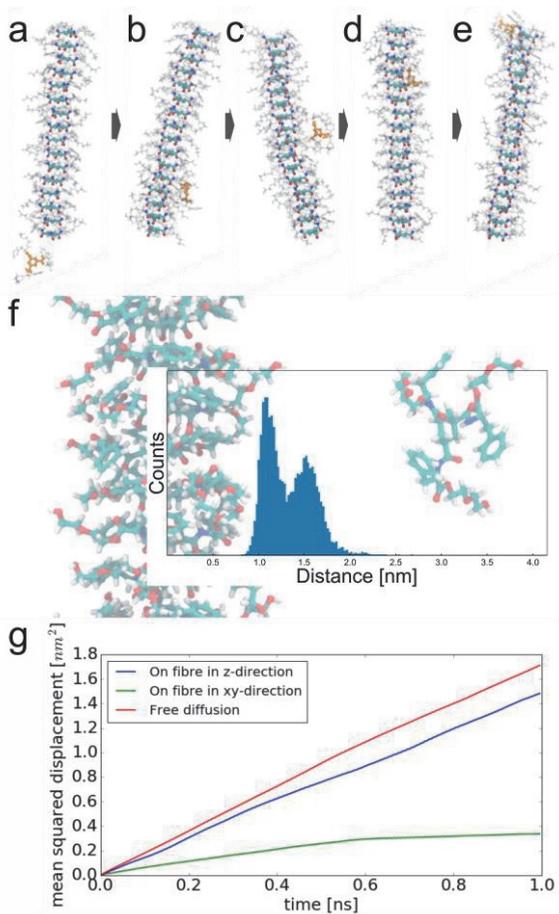


Figure 3.3. Elongation. (a-e) Representative snapshots of simulations of 16-mer and free molecule; a molecule adsorbs on the fiber and moves on its surface, eventually reaching one of the ends of the fiber. (f) Distribution of the distance of the center of the single molecule to the center of the infinite fiber (analyzed over the final 2.5 ns of 15ns simulations) (g) First 1 ns of mean squared displacement (MSD) for a molecule in the proximity of fiber along the main axis of the fiber (blue), in the plane of the fiber (green), and for free molecule (red). The curves are scaled for dimensionality (i.e. free diffusion by factor of 3, diffusion in xy-plane by factor of 2, and diffusion in z-direction by factor of 1). The in-plane movement is similar to diffusion in confinement, and the movement in direction of fiber resembles free diffusion.

**Secondary Nucleation.** If molecules tend to adsorb on the surface of the fiber and diffuse along the fiber axis, they can also interact with each other on the surface, and possibly create a nucleus which grows on the side of existing fiber. To see if such a process is feasible we have performed similar adaptive sampling simulations as before, however this time of eight molecules in the presence of an infinite fiber consisting of eight molecules (see Figure 3.4a). Similarly as before we present here the transitions between states obtained from MSM analysis of many short simulations (cumulative time 0.8  $\mu$ s) in Figure 3.4a-g (see SI for video of the fastest trajectory). Initially, molecules adsorb on the surface of the fiber reaching a high number of neighboring molecules ( $\sim$ 6, see Figure 4h, orange line). In contrast to the aggregates formed in the absence of fiber, molecules on the fiber interact not only with each other in an unordered cluster but also with the molecules forming the fiber (green and red lines, respectively, in Figure 4h). In contrast to primary nucleation, this leads to more dispersed and smaller aggregates on the surface of the fiber. Ordered stacks can form within the aggregates, and grow by further rearrangement or by new molecules joining. The newly formed fibril was oriented with its macrodipole (mostly due to aligned amide dipoles) parallel to that of the original fibril, suggesting that there is no strong preferential orientation of the fiber via macrodipole stabilization<sup>109</sup>. We have confirmed the existence of bundles of fibers by performing cryo-TEM imaging (Figure 3.4i, see SI for more pictures). From these pictures, we can see that fibers create very well ordered bundles with a consistent structure within the bundles (see histograms in SI): most of the time they have a Gaussian distribution of the fiber width within one picture. However, when all histograms are merged into one (see Figure 3.4j) the width is scattered which is indicative of a complex arrangement of fibers within the bundles.

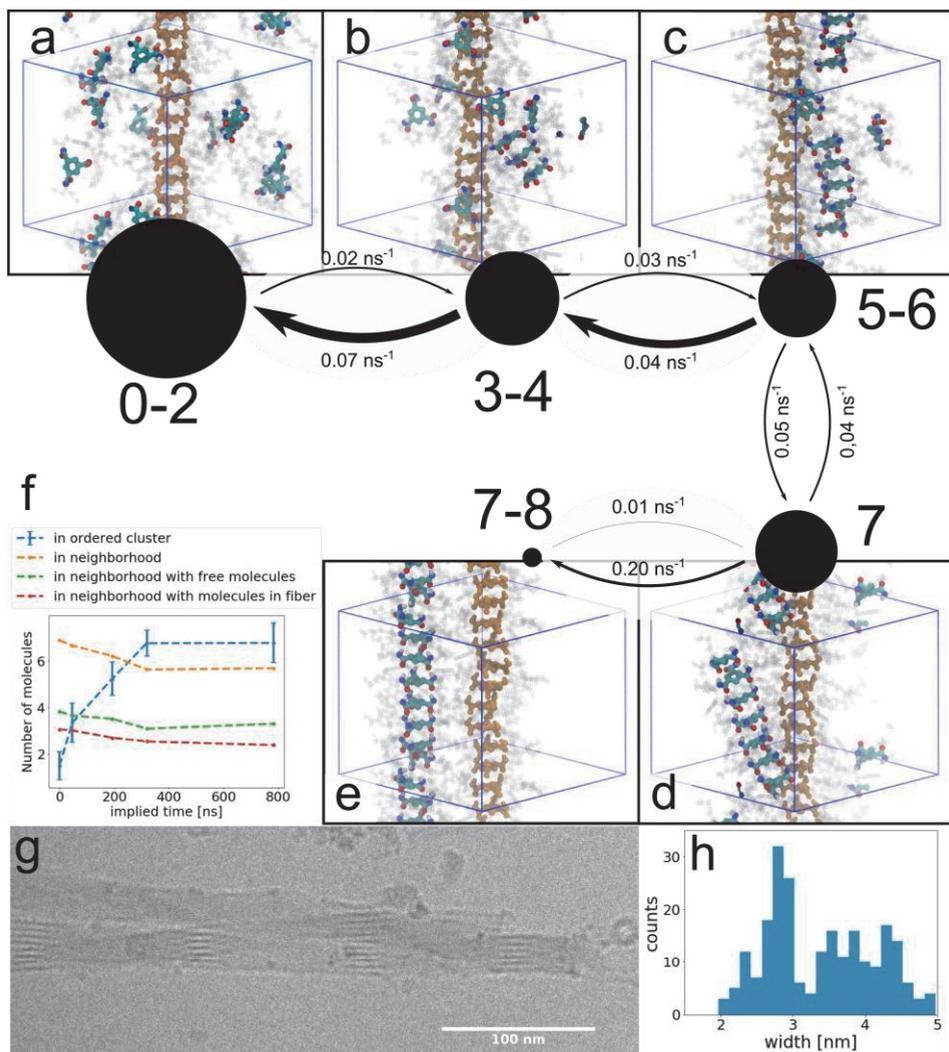


Figure 3.4. Secondary nucleation. (a-g) MSM analysis of simulations of randomly distributed molecules (indicated by cyan color) on the side of an existing fiber (indicated by orange color, and not included in the MSM). Representation of states, populations, transitions, and molecules analogous to Figure 2. (h) Number of molecules in the ordered stack (blue), and number of direct neighbors (orange) of the non-fiber CTA molecules. The number of the neighbors is decomposed in neighbors only with free molecules (green) and with molecules of the fiber (red). (i) Cryo-TEM image of the self-assembled bundle of fibers. (j) Histogram of widths of single fibers in bundles gathered from eight cryo-TEM images (see SI).

**Bundling.** Bundles can be formed by secondary nucleation, but also by individual fibers interacting with each other. Intertwining interactions are crucial for network formation and therefore gelation<sup>20</sup>. During simulations, we have not observed branching of the single stack. Therefore, one way how these single stack fibers can create a network could be by strongly interacting with each other. To check that, we have performed twenty independent 10 ns simulations of two infinite fibers (Figure 3.5a-c), and in all of them, the two fibers merge and in this form stay until the end of simulations (the histogram of the distance between the centers of fibers is presented in Figure 3.5d). Fibers randomly diffuse through the simulation box until they come close to each other and interact. Initially, when the hydroxyl side chains are in contact (and the distance between centers is larger than 1.9 nm), they can still dissociate. However, after some time the hydroxyl groups of side chains move aside to the edges of the interface between the fibers (the distribution of the hydroxyl groups in the final bundle is presented in Figure 3.5e), making the phenyl rings of the neighboring fibrils accessible for interaction. The fibers in the bundle interact with each other strongly by hydrophobic interaction (see Figure 3.5c) and the complex does not dissociate on timescales of the simulations. We have not noticed a difference between bundling of parallel and antiparallel oriented fibers (see SI). If we add one more fiber to the two- fiber bundle we observe that fiber to become part of the bundle. The angle between the three fibers shows a multimodal distribution around 60, 110 and 180 degrees, which could explain why in cryo-TEM pictures we observe such a variety of structures.

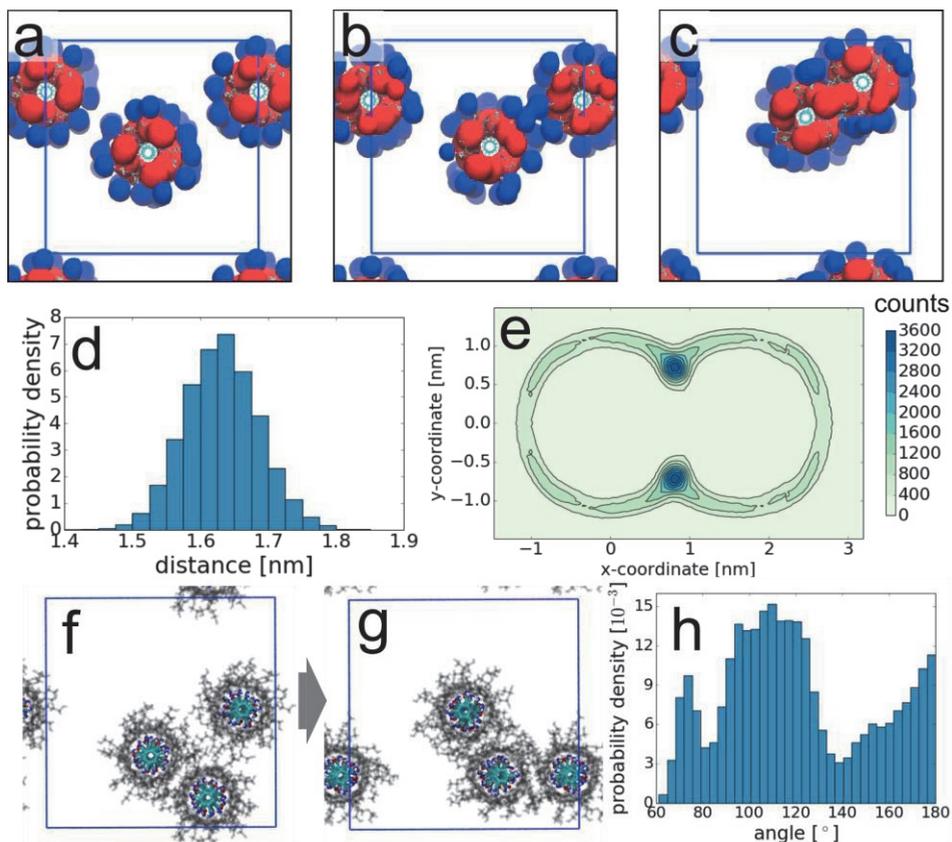


Figure 3.5. Bundling. (a-c) Snapshots of selected simulation of bundling of two fibers, red beads indicate the benzene rings and blue the end hydroxyl groups. Fibers merge upon the strong interaction of phenyl rings with each other. (d) Histogram of distances between cyclohexane ring of two fibers in the final state. (e) 2D histogram of end hydroxyl groups (-OH) in the final structure; it can be seen that hydroxyl groups mainly stay on the edge of the fiber interface, making a space for hydrophobic interactions of phenyl groups between fibers. (f-g) Snapshots of selected simulation of bundling of three fibers. (h) Histogram of the angle created by three fibers.

## **Summary and discussion of the mechanism of supramolecular self-assembly**

In this work, we study the mechanism of fiber and bundle formation of CTA. Our results not only align with existing research but also gives unique insights. We study the mechanism by investigating several distinct phases schematically summarized in Figure 3.1. Firstly, molecules aggregate into an unordered cluster by non-directional hydrophobic interactions. Then, they rearrange within the aggregate and create a small ordered stack by hydrogen bonding. When the stack consists of more than 4-5 molecules its stability no longer increases, reaching limiting value; such stacks may be seen as the critical nuclei. Nuclei grow into fibers by further rearrangement within smaller aggregates. We show that longer fibers are most likely to grow by new molecules adsorbing on their surface and diffusing along the surface until reaching one of the ends, where it can then lead to elongation. Given the large surface of the fiber adsorption and diffusion is more likely than a new molecule encountering one of the ends straight from solution. We speculate that the balance of interactions that allows both adsorption on (high enough affinity) and diffusion along (not too strong binding) the fiber could be the property which distinguishes gelators with low critical gel concentration from other gelators. If sufficient molecules are adsorbed on the surface of a fiber, they can also interact with other adsorbed molecules and result in the creation of a new nucleus on the side of the fiber, rather than diffuse to the ends and lead to elongation. Therefore, elongation and secondary nucleation are competitive processes. Although not studied here, we could imagine that a high concentration of molecules promotes secondary nucleation: high concentration could lead to many molecules being present on the surface of existing fibers, easily exceeding the critical nucleus size. On the other hand, low concentration could lead to a small number of molecules on the side of existing fiber, such that locally the nucleation size is not exceeded leading only to elongation of the existing fiber. However, very low concentration could lead to the main fiber slowly dissolving. We have observed fragmentation of a somewhat longer fiber into two fragments should be feasible, which can act as separate nuclei. Moreover, fibers can interact with each other by hydrophobic interactions of phenyl groups in the CTA side chains. This results in the creation of bundles of fibers, which we have observed experimentally. Such interaction could also happen between parts of long fibers resulting in their crosslinking, which if happened multiple times could result in the creation of a gel network.

We have analyzed the trajectories using Markov State Modeling, which gives insight into kinetic pathways of the process. Although the exchange rates obtained here can be interpreted only semi-quantitatively they give a new perspective on the formation of the fibers. Upon improvement of this method, it will be possible to give a reliable relation between specific pathways and the responsible interactions (implying a relation to the molecular structure). This will allow creating new design rules based on the kinetic pathways.

We would like to highlight that our results in many places align with existing reports in this research area. For example, similar to the computational work of Pavan et al.<sup>39,59</sup>, in which they studied BTA, we observed that the fast creation of an unordered cluster precedes slow formation of ordered fiber. We also found an increase of stability when an ordered stack consists of 4-5 molecules. It seems that there is no significant difference between interaction of antiparallel and parallel oriented fibers, which is in line with observations Pereira Oliviera et al.<sup>109</sup>. We found many similarities to research on protein fiber formation. Although protein fibers consist of molecules which are much larger than the molecules studied here and the nature of their interactions is different (mostly they interact by local interactions, without cooperative effect coming from electronic changes), similar steps have been distinguished: nucleation, elongation, dissociation, secondary nucleation and fragmentation<sup>13,105</sup>. Probably the biggest difference between these systems lies in the interaction strengths between molecular blocks, which in case of proteins is almost irreversible, allowing the neglect of dissociation in a model.

## **Conclusions and Outlook**

Overall, we show a detailed mechanism of supramolecular fiber formation with atomistic resolution, by distinguishing and studying several different processes occurring: aggregation and nucleation, elongation, secondary nucleation and bundling, schematically shown in Figure 3.6, showing the complexity of the network created by different processes involved in the self-assembly. In our system, each of these processes involves specific interactions. Aggregation and nucleation starts with hydrophobic interactions strong enough to hold molecules together, but weak enough to leave flexibility for the molecules in the aggregate to rearrange and become aligned into well-ordered stacks in which directional H-bonding between amide groups near the core stabilizes the fiber. Elongation and secondary nucleation rely on the Van der Waals and hydrophobic interactions of monomers with the surface of the fiber. Finally, bundling

involves the initial association of fibers by their surface hydroxyl groups, followed by stabilization of the bundle through interactions between phenyl groups in the side chains of the neighboring fibers. This knowledge promises improved control over this system by tuning certain interactions. We would like to stress that for development of design rules not only interactions crucial for the stability of the final structure should be taken into account, but also the ones involved in intermediate stages. We believe that a quantitative understanding the relation between molecular (sub)structure and kinetic pathways is within reach if the strengths of experimental and modeling techniques are combined with quantitative relations between the structure of the monomer and the final assembly.

## Supporting information

### Computational methods

#### Molecular dynamics simulation

Single simulations are done using modified version of GROMACS which allows simulations with CHARMM Drude force-field (GROMACS version 2016-dev-20170105-c53d212, which is accessible through git repository [git://git.gromacs.org/gromacs.git](https://git.gromacs.org/gromacs.git)). We use the CHARMM Drude force-field which allowed us to simulate a CTA fiber that remained stable when its initial structure was based on the crystal structure of its analogue. The CHARMM Drude force-field explicitly models electronic polarizability by the inclusion of Drude oscillators, which are small charge-carrying particles connected to atoms. Polarizable force fields are computationally demanding, however, recent implementation of extended Lagrangian dynamics with a dual Nose-Hoover thermostat allows one to perform simulation efficiently<sup>86</sup>. Since GROMACS has implemented only a thermostat and no barostat for this efficient use of the force-field, most of the simulations are run at constant volume, see next paragraph.

Systems were set up in triclinic or cubic simulation boxes under periodic boundary conditions. Before running the production simulation, the energy was minimised using the steepest descent algorithm. Then a short simulation (10,000 steps with a timestep of 1fs) in the NPT ensemble is performed using the V-rescale thermostat in 298.15 K (with coupling time 0.1) and isotropic Parinello-Rahman barostat of 1.0 bar (with coupling time 1.0 ps and a compressibility 4.5e-5 bar-1) with the self-consistent field treatment in which the positions of the Drude oscillators are relaxed to the potential energy minimum at each simulation step. For Drude particles we used standard parameters describe in<sup>86</sup>. After this equilibration, the volume was kept constant. The production simulations run were done with a dual Nose-Hoover thermostat developed by Lemkul et al.<sup>86</sup>, with a coupling constant of 0.005 ps for Drude particles and 0.1 ps for all other particles. All simulations were run at 298.15 K. The equations of motion were solved numerically using extended Lagrangian dynamics with a timestep of 1 fs.

#### The framework

A single simulation starting from randomly dispersed CTA molecules in solution, does not lead to the formation of a fiber, even a short one, within 100 ns, indicating that nucleation might be relatively slow on the simulation time scale. That is why we have developed a framework which would facilitate the formation of an ordered aggregate within reasonable simulation time. We were inspired by adaptive Markov state modeling (AMSM)<sup>101</sup>. In the AMSM approach many short simulations are run, which are analyzed to distinguish several different states by a similarity of conformations. Most of the time, the similarity of the conformations is measured by common analysis methods used in molecular dynamics like distances, angles, number of hydrogen bonds, RMSD etc. Similar conformations are then grouped into states. This results in a landscape consisting of many states. The number of visited states depends on the criterion set for grouping conformations into a state and on the total simulation time spent in combination with the dynamics of the system in visiting the different states, which depends on the starting structures and the presence of barriers between states. The next generation of simulations is run from states which are least visited. In this manner, a large part of conformational space can

be explored. The main advantage of such a framework is that all simulations are unbiased (in the sense that there is no additional term added to the Hamiltonian). That gives a big advantage: at several occasions during the procedure, we realized that our criteria for distinguishing states or for defining the fiber state is not precise enough. Since all the trajectories are unbiased, in the next iteration we simply recalculate all states by the modified criteria, and therefore we could still use all previously run simulations.

Here we describe in detail the procedure. Firstly we run the first generation of simulation from randomly distributed gelator molecules in water using gromacs (as described above). After the first run, we run several scripts which work in parallel. Each script takes all trajectories of simulations finished thus far and for every frame measures the size of largest ordered cluster (see Analysis). The histogram of these sizes is measured. Most often we would like to start the simulation with a state which is the least visited, therefore we chose the state for the next simulation with a probability inversely proportional to a number of counts for that state in the histogram. Consequently the least visited state is the most probable as the start of the next simulation, but it still leaves a chance to start from another state. We start running the new simulation starting from a randomly chosen conformation belonging to the selected state. After the simulation is finished, the scripts again analyze all trajectories (also the ones created by other scripts: this is the point where all simulations collaborate with each other) and searches for the least visited state. The entire procedure is visually presented in Figure S3.1.

We ran 12 jobs running in parallel, from which 8 were performing 10 ns simulations, and 4 were performing 1 ns simulations.

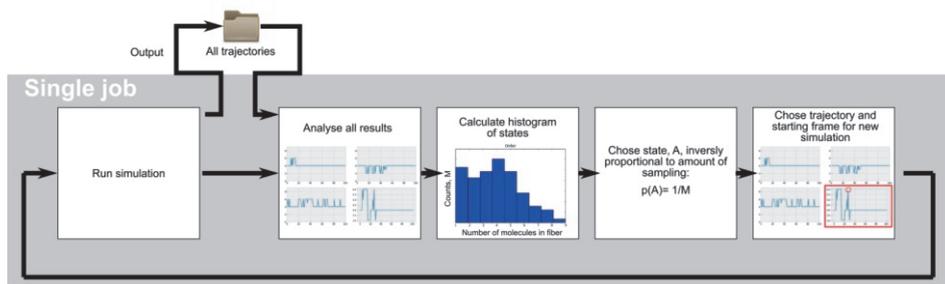


Figure S3.1. Single script from the framework follows the scheme: (i) run simulation, (ii) combine and analyze all trajectories present in the folder (created also by other scripts), (iii) calculate distribution of states, (iv) choose rarely visited state (with probability inversely proportional to the occurrence of the state), (v) choose randomly trajectory in which this state occurs and from this trajectory choose randomly starting frame for simulation with the chosen state, (vi) go to point (i).

## Analysis

**Cluster analysis.** Cluster analysis was done by creating a matrix  $A$  representing adjacent molecules (i.e. molecules in neighborhood). Two molecules are considered in neighborhood, according to the function:

$$A_{ij} = \begin{cases} 1, & \sigma(|\vec{r}_{ij}|)K(\angle(\vec{n}_i, \vec{n}_j))K(\angle(\vec{n}_i, \vec{r}_{ij})) > 0.5 \\ 0, & \sigma(|\vec{r}_{ij}|)K(\angle(\vec{n}_i, \vec{n}_j))K(\angle(\vec{n}_i, \vec{r}_{ij})) < 0.5 \end{cases} \quad (\text{S3.1})$$

Where  $\vec{r}_{ij}$  is the vector connecting the centers of the cyclohexane rings of molecules  $i$  and  $j$ ,  $\vec{n}_i$  is a normal vector to the plane created by the cyclohexane ring of molecule  $i$ .  $\sigma(r), K(\theta)$  are the switching functions for distance and angle, respectively. These functions are defined as:

$$\sigma(x) = \frac{1 - (\frac{x - d_0}{r_0})^6}{1 - (\frac{x - d_0}{r_0})^{12}} \quad (\text{S3.2})$$

$$K(\theta) = \begin{cases} 1, & \theta < b \text{ or } \theta > b \\ 2 - \left| \frac{\theta}{b} \right|, & 0 < 2 - \left| \frac{\theta}{b} \right| < 1 \\ 0, & \left| \frac{\theta}{b} \right| < 2 \text{ and } \left| \frac{\theta - \pi}{b} \right| < 2 \\ 2 - \left| \frac{\theta - \pi}{b} \right|, & 0 < 2 - \left| \frac{\theta - \pi}{b} \right| < 1 \end{cases} \quad (\text{S3.3})$$

Examples of these functions are presented in Figure S3.2 along with data from a trajectory of a long ordered fiber and of an unorganized cluster. We calibrated both functions to give a positive answer for a long ordered fiber, for which we know that all molecules are in neighborhood. This resulted in parameters  $r_0 = 0.2 \text{ nm}$ ,  $d_0 = 0.48 \text{ nm}$  and  $b = 0.35$ . Graphical explanation of this function is presented in Figure S3.3. As a result of the clustering algorithm we obtain the adjacency matrix  $A$ , from which we can measure the largest connected cluster.

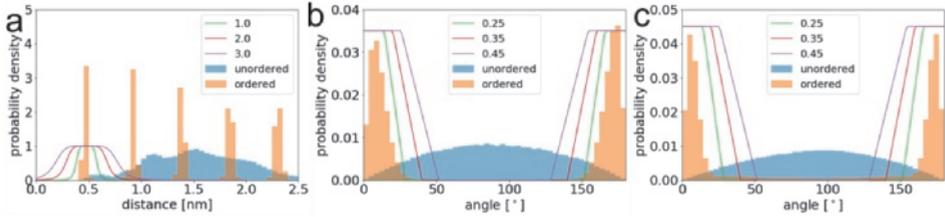


Figure S3.2. (a) Plot shows switching function,  $\sigma(r)$ , for different parameters  $r_0$ , histograms show the data for long fibre (green) and unordered system (blue). (b) Different angle switching function for different parameter  $b$  (normalized to 0.035), the histograms show the data for the angle between the normal vectors of the cyclohexane rings  $\angle(\vec{n}_i, \vec{n}_j)$ . (c) Different angle switching function (normalized to 0.045) for different parameter  $b$ , the histograms show the data for the angle between the normal vector of one cyclohexane ring and the vector connecting centers of two cyclohexane rings  $\angle(\vec{n}_i, \vec{r}_{ij})$ .

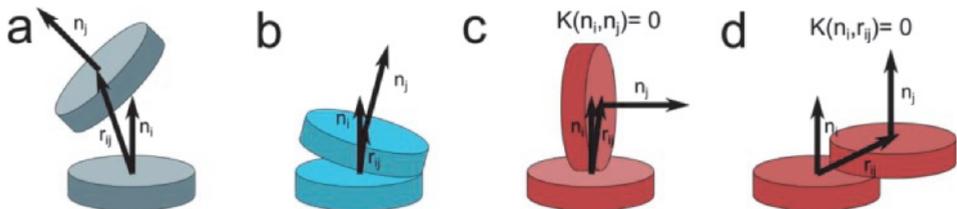


Figure S3.3. Schematic representation of possible conformations of two molecules. (a) Two molecules are considered in neighborhood if the distance between them is in a certain range, and the angle between normal vectors  $\vec{n}_i$  and  $\vec{n}_j$  is in certain range, and the angle between the normal vector  $\vec{n}_i$  and the vector connecting centers of two rings  $\vec{r}_{ij}$  is in a certain range. (b) Example of two molecules in neighborhood, the distance between centers,  $\vec{r}_{ij}$ , is small, and angles  $\angle(\vec{n}_i, \vec{n}_j)$  and  $\angle(\vec{n}_i, \vec{r}_{ij})$  are small. (c) Example of two molecules not in neighborhood, although molecules might be in the appropriate distance range, the angle  $\angle(\vec{n}_i, \vec{n}_j)$  is far from  $0^\circ$  or  $180^\circ$ . (d) Example of two molecule not in neighborhood, although the distance and angle between two normal vectors of two molecules might be in range, the angle between the normal and the vector connecting two centers  $\angle(\vec{n}_i, \vec{r}_{ij})$  is far from  $0^\circ$  or  $180^\circ$ .

**Unordered clusters.** In a manner similar to determining the largest ordered or connected cluster, the largest unordered cluster is measured. In this case only the distance switching function,  $\sigma(r)$ , is used (none of the angles between two molecules are taken into account). The parameters  $d_0$  and  $r_0$  used in the switching function are the same.

**Dipole moment.** Dipole moment was calculated using *gmx dipoles*, which is part of gromacs package.

**Hydrogen bonds.** Hydrogen bonds were calculated using *hbonds* packages from VMD. Only hydrogen bonds between amide groups were calculated and standard parameters were used, i.e. the distance between oxygen and nitrogen must be less than 0.3 nm, and the angle oxygen-hydrogen-nitrogen must be less than  $20^\circ$ .

**Surface accessible solvent area (SASA).** SASA was calculated using *gmx sasa*, which is part of gromacs package.

## Markov State Modeling

We have analyzed results using by Markov State Model. Although for the adaptive sampling used in this work, number of ordered molecules has been sufficient measure to sample formation of the fiber, it turned out to be not sufficient for Markov State Model, leading to transitions which are depended on history of transitions (that is depending on previous transitions), see Figure S3.4. Therefore, we have used measure which describes system in more detail. We have measured a vector containing eight elements  $v = (a_1, \dots, a_8)$ , which  $i$ -th element describes number of its ordered neighbors, that is  $a_i = \sum_j B_{ij}$ , where  $B_{ij} = \sigma(|\vec{r}_{ij}|)K(\angle(\vec{n}_i, \vec{n}_j))K(\angle(\vec{n}_i, \vec{r}_{ij}))$  (see equation (S3.1)). In result, every value  $a_i$  can have value 0,1 or 2 (more neighbors because of the special confinement is not possible). However,

such vector would depend on the numbering of the molecules and for two exact systems with different numbering of molecules would be described by different vector. A simple trick to ensure that obtained vector is invariant on numbering is to order of elements of the vector from smallest to the largest. Such trick, however, only works for vector consisted of scalars and more complex measure taking into account more parameters is challenging.

Using this measure we were able to construct Markov State Model. For primary nucleation we have divided space by K-mean clustering algorithm into 2000 clusters. We have calculated transitions between states for lag time of 0.8 ns. Upon coarse graining, we were able to distinguish 6 slow events, resulting in 7 states. However, since also infinite fiber (crossing periodic boundaries condition) was part of the model, the system is not ergodic, since infinite fiber never disassembles. See Figure S3.5. In result, we have excluded all trajectories leading to infinite fiber in MSM analysis. Figure S3.6 shows the result of MSM analysis. Figure S3.6c is in the main text, with removed low probable transitions. The properties of the states are present on Figure S3.7.

Similar analysis have been done for secondary nucleation. The chosen lag time was 0.4 ns and the 4 slow transition could be distinguish, resulting in 5 states.

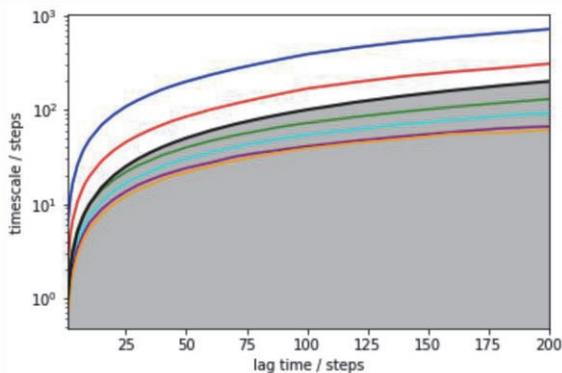


Figure S3.4. Implied timescales for the model based only on number of ordered molecules. Implied timescales do not coverage to constant value (i.e. that the process is independent from the choice of lag time).

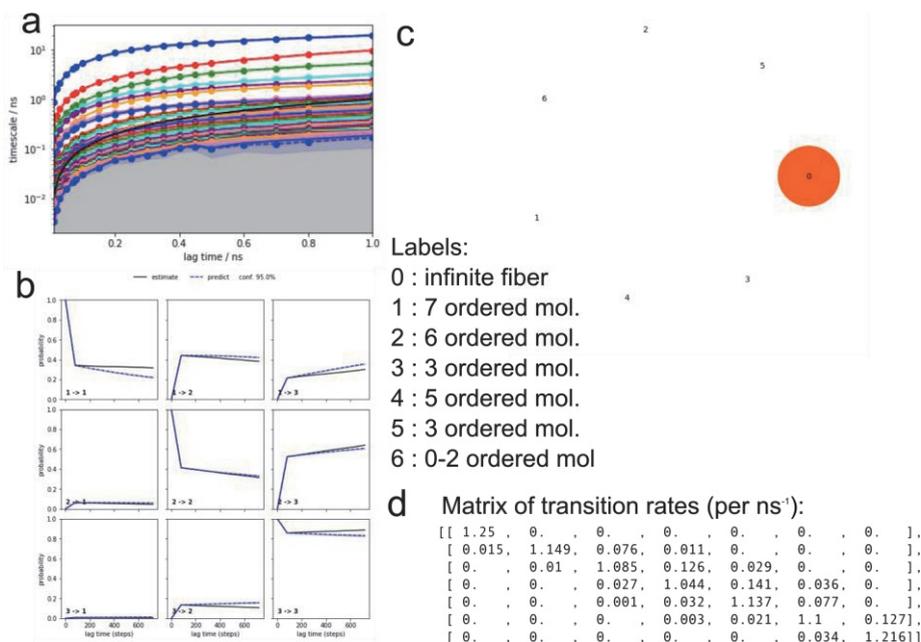


Figure S3.5. Markov State model of all primary nucleation including formation of infinite fiber. (a) Implied timescales. (b) Results of Chapman-Kolmogorov test. (c) Graph of the pathways between the states. (d) Transition matrix.

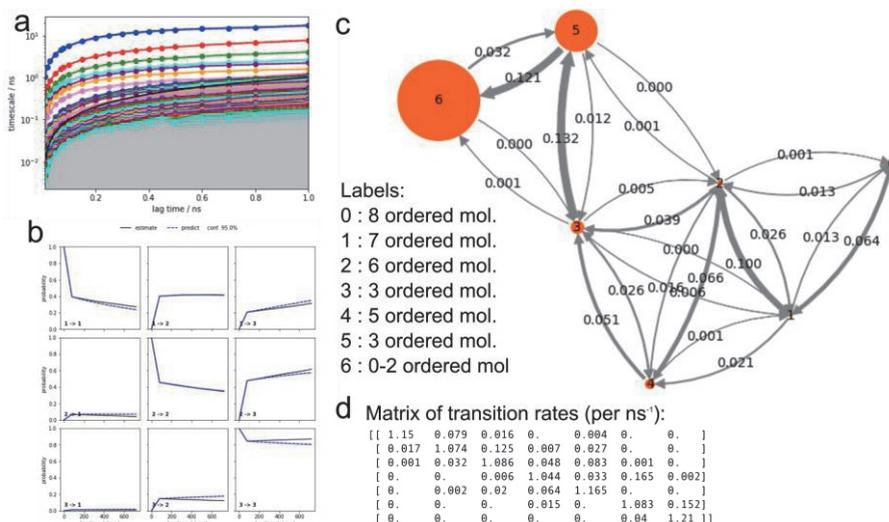


Figure S3.6. Markov State Model of primary nucleation simulations except the one in which infinite fiber is formed. This data is presented in the main text. (a) Implied timescales, (b) Results of Chapman-Kolmogorov test. (c) Graph of the pathways between the states. (d) Transition matrix.

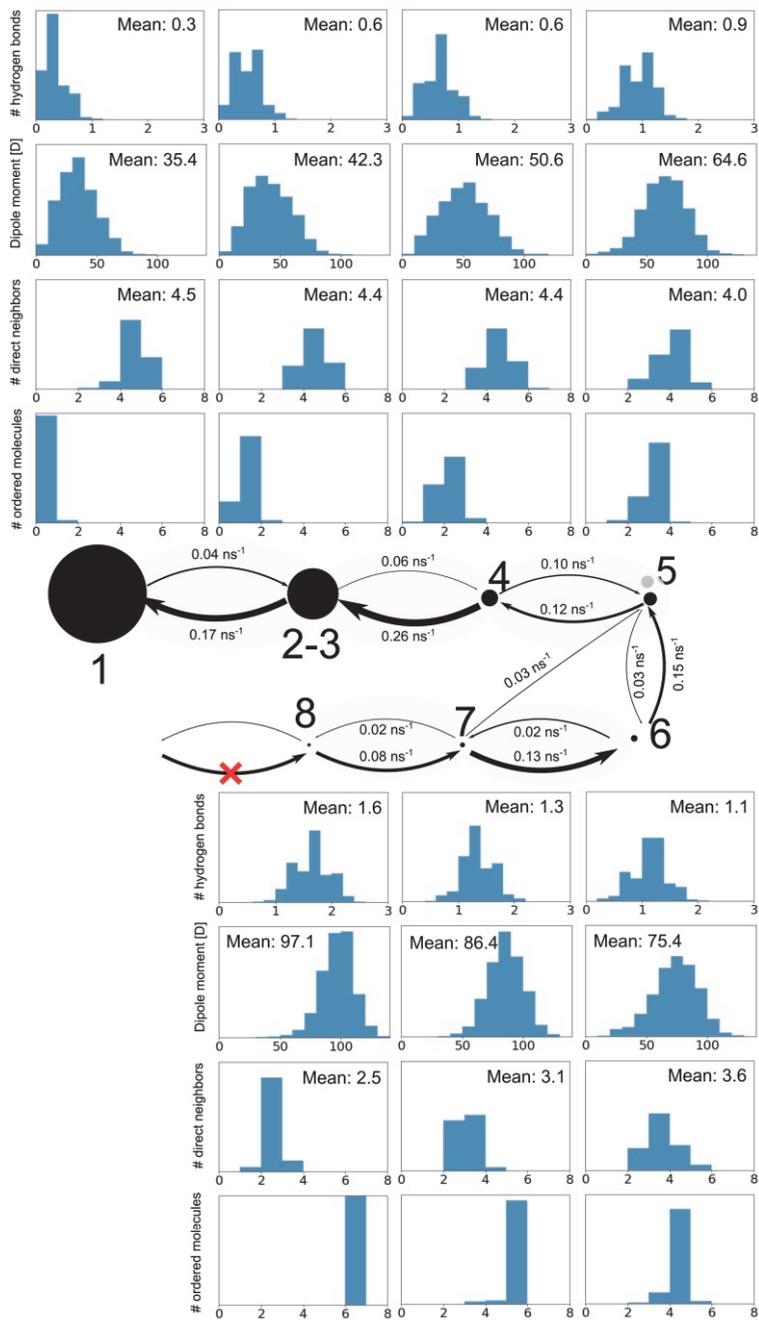


Figure S3.7. Distributions of number of hydrogen bonds (per molecule), dipole moment of CTA molecules, numbers of neighbors and number of ordered molecules in coarse-grained states from MSM analysis of primary nucleation.

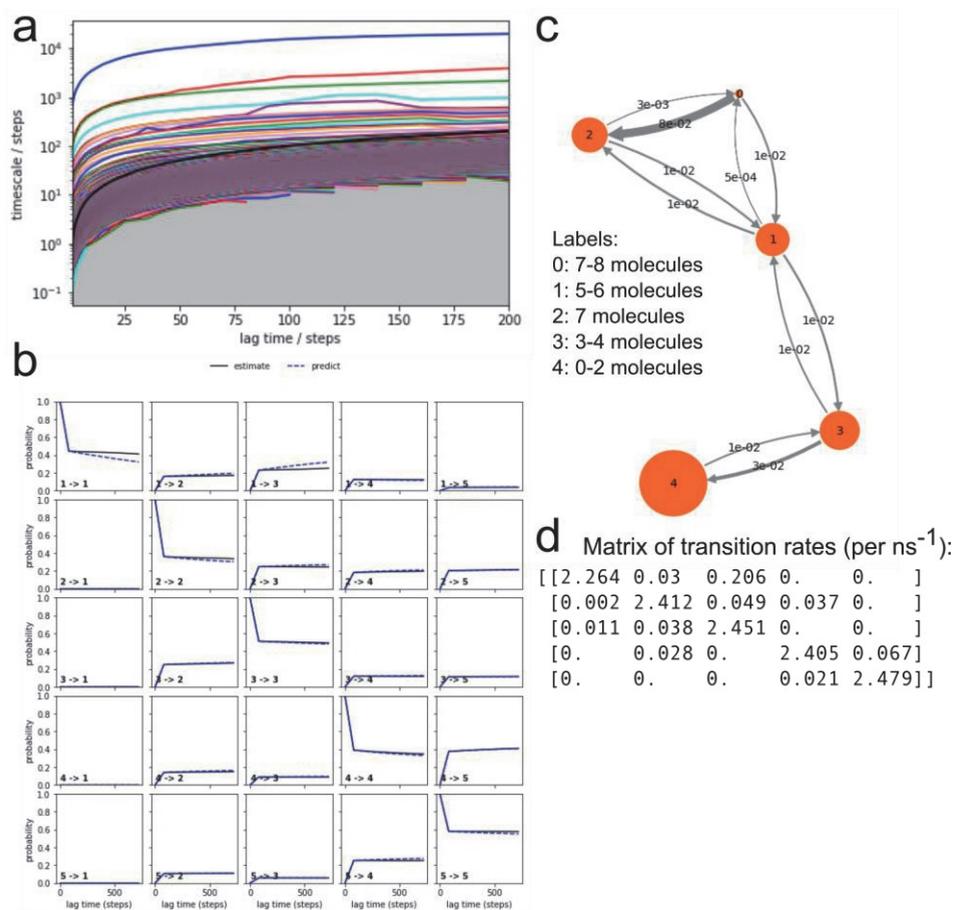


Figure S3.8. Markov State Model of secondary nucleation simulations. This data is presented in the main text. (a) Implied timescales, (b) Results of Chapman-Kolmogorov test. (c) Graph of the pathways between the states. (d) Transition matrix.

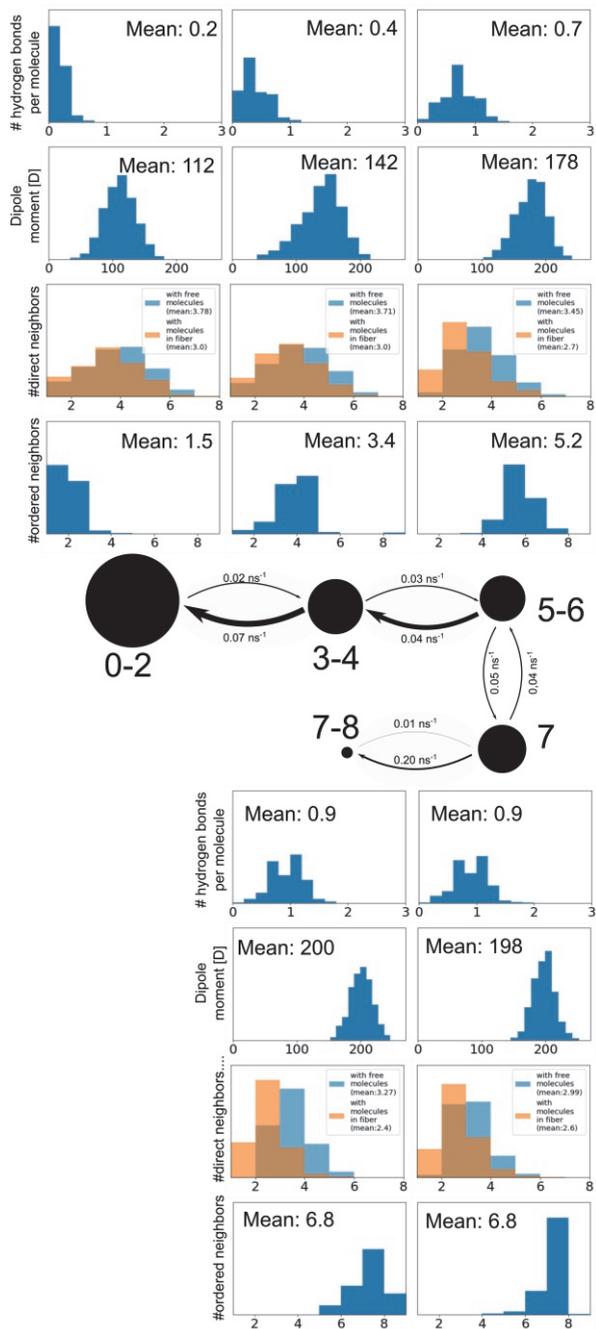


Figure S3.9. Distributions of number of hydrogen bonds, dipole moment, numbers of neighbors and number of ordered molecules in coarse-grained states from MSM analysis of secondary nucleation

## Additional results

### Nucleation

Figure S3.10 presents the number of hydrogen bonds (between gelator molecules) and solvent accessible surface area (SASA) for the trajectory discussed in the main text, i.e. formation of fiber from randomly distributed gelator molecules in water. The number of hydrogen bonds grows linearly, in contrast to the dipole moment (shown in Figure 3.2 in the main text), which has a rapid increase once a small cluster of 3-4 molecules is formed. At the beginning of the simulation, randomly distributed molecules rapidly merge into an unorganized cluster which reflects in large drop in SASA. The unorganized cluster slowly rearranges to a more compact ordered structure, which results in a slow decrease of SASA.

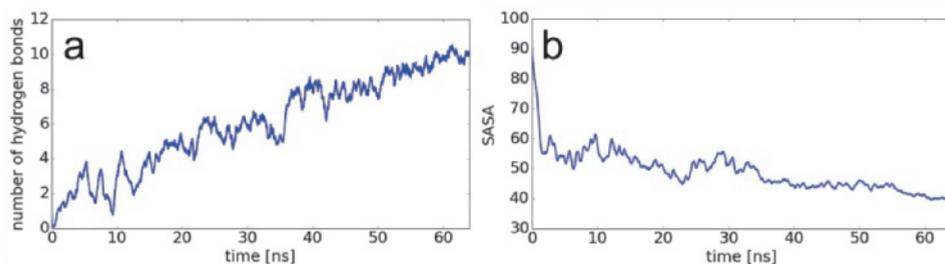


Figure S3.10. Additional analysis of the trajectory of formation of the fiber. (a) number of hydrogen bonds, (b) solvent accessible surface area.

### Stability of the ordered stack.

To study the stability of the ordered stack we have performed series of simulations of ordered stacks of size 2, 3, 5, 8, 12 and 16 in water (the list of simulations is present in Table S3.1). Example snapshot of dodecamer is present in Figure S3.11b). Most of the simulation lead to dissociation of one molecule (Figure S3.11a), but we have also observe one fragmentation (Figure S3.11c).

Table S3.1. List of simulation of ordered stack in water.

Simulation	Number of independent simulations	Time
Dimer	15	10 ns
Trimer	15	10 ns
Pentamer	33	10 ns
Octamer	58	10 ns
Dodecamer	12	10 ns
Sextamer	12	10 ns

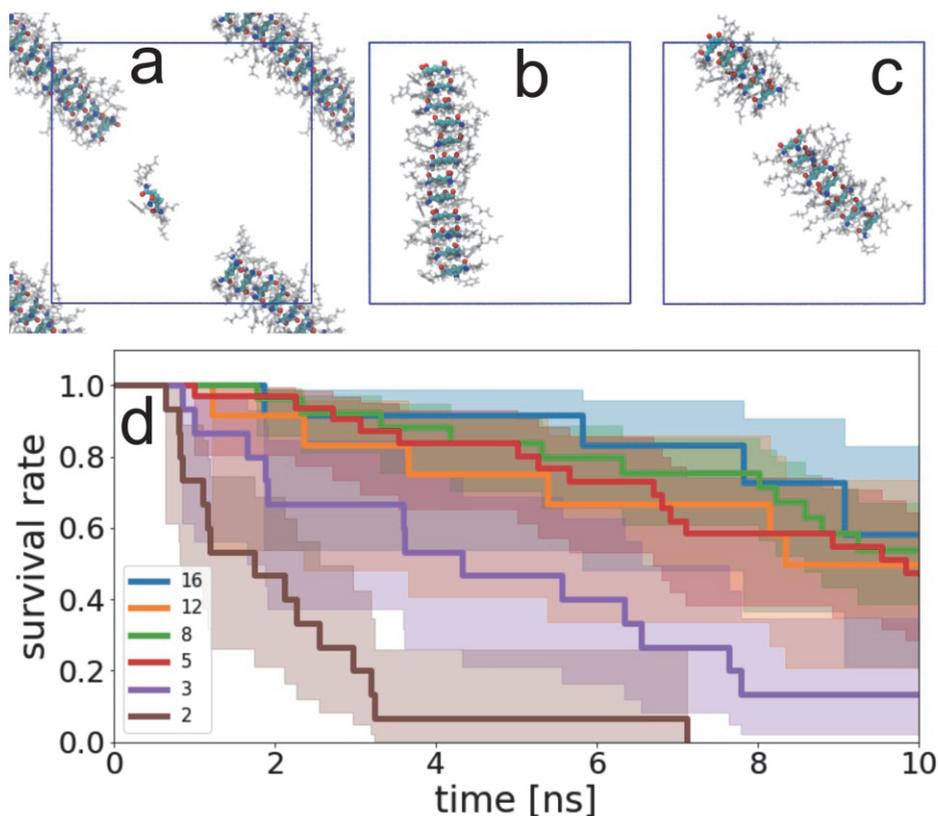


Figure S3.11. Example snapshots of simulation of ordered stack (b), most of the simulations lead to dissociation of one the molecules (a), but we have also observe one fragmentation (c). The same snapshots (but without simulation box) are present in the main text. (d) Survival rate of the ordered cluster with standard deviation error. The same figure (but without standard deviation error) is present in the main text.

## Elongation

**Adsorption and desorption on the fiber.** During the course of ten 15 ns simulations all molecules adsorb on the surface of the fiber in the first 7.5 ns. Later, three processes of desorption can be observed (see Figure S3.12a).

**Directionality of diffusion on the fiber.** By analyzing the diffusion in the direction of the main axis ( $z$ ) of the fiber we have not observed a preferential direction of the movement. Figure S3.12b shows the progress of  $z$ -coordinate in 10 independent simulations/molecules analyzed from the moment when molecule adsorb on the surface of the fiber. We also calculated autocorrelation function of these coordinates, and it is presented in Figure S3.12c. The autocorrelation shows the average change of the  $z$ -coordinate after a particular time, called lag-time. It shows that on average molecules diffuse more into one direction then other, but the difference is just a result of

low statistics (see the standard deviation error). . We were wondering if molecules orient during the diffusion. Therefore we have calculated directions of dipole moments. However, we have not observed that; the dipole moment of the fiber is always pointing in the direction of z-axis (see Figure S3.12d), whereas dipole moment of the free molecule is pointing in a random direction (see Figure S3.12e).

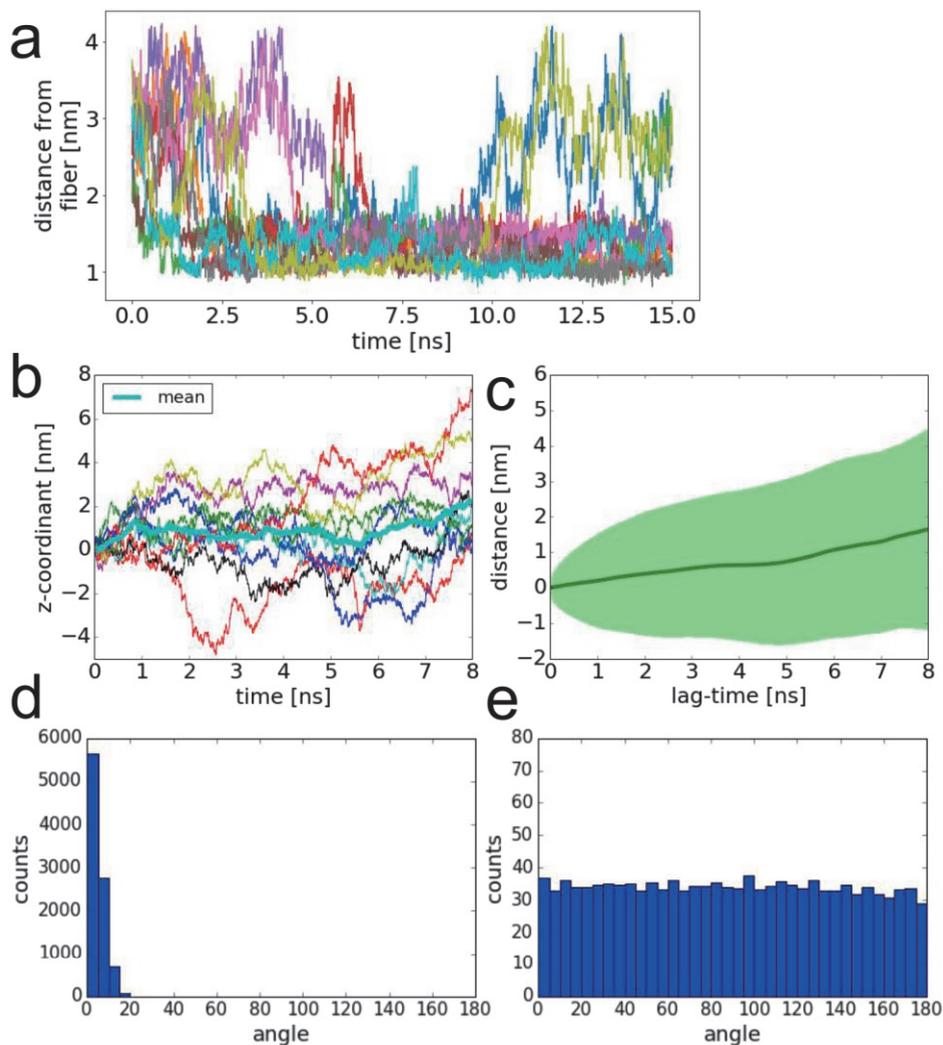


Figure S3.12. (a) Distance of the free molecule from the center of the fiber (different color represent traces for different simulation). During first 7.5 ns all molecules adsorb on the surface. Later three desorption can be observed. (b-d) Directionality of diffusion of free molecule in direction parallel to the axis of the fiber. (a) Progress of z-coordinate for ten different trajectories and its (b) mean autocorrelation function. (c) Histogram of the angle between vector of dipole moment of molecules creating fiber and z-axis (corrected for the geometrical  $\sin\phi$  factor<sup>110</sup>). (d)

Histogram of the angle between vector of dipole moment of free molecule and z-axis (corrected for the geometrical  $\sin\phi$  factor<sup>110</sup>).

## Elongation: simple models

Let's create simple models of two proposed models of elongation: one for finding the end of a fiber and a second for finding the side of a fiber. Let's imagine that we have a system with one free molecule and  $n - 1$  molecules in a fiber and the concentration is equal to  $c$ ; the volume of such a system is then equal to  $V = n/c$ . Let's assume that a single molecule is a disc and has a volume  $V_1 = R_1 \cdot h$ .

**Model without crawling.** The fiber has two ends, therefore the probability of finding the end of the fiber by the molecule is equal to:

$$p = \frac{2V_1}{V} = \frac{2V_1 \cdot c}{n}. \quad (\text{S3.4})$$

This implies that for large fibers, that is  $n \rightarrow \infty$ , we see that  $p \rightarrow 0$ .

**Model with crawling.** If a molecule can adsorb on a surface then the probability of finding the surface is equal to  $h(n - 1)(R_2 - R_1)$ , where  $h$  is the height of one molecule, and  $R_2$  and  $R_1$  are the outer and inner radii of a cylinder surrounding the fiber, from which the adsorbed molecule cannot escape (which can be calculated from the histogram shown Figure 2c in the main text). Therefore, the probability of a molecule to attach to the side of the fiber is equal to:

$$p = \frac{h(n - 1)(R_2 - R_1)}{V} = \frac{h(n - 1)(R_2 - R_1)c}{n}. \quad (\text{S3.5})$$

This implies that for large fibers, that is  $n \rightarrow \infty$ , we see that  $p \rightarrow \frac{h(R_2 - R_1)}{c}$  (*const.*).

## Secondary nucleation

Figure S3.13 presents an additional analysis of the trajectory of secondary nucleation: the number of hydrogen bonds and the solvent accessible surface area (in this case calculated for additional, free molecules, and also for the free molecules and molecules which create fiber).

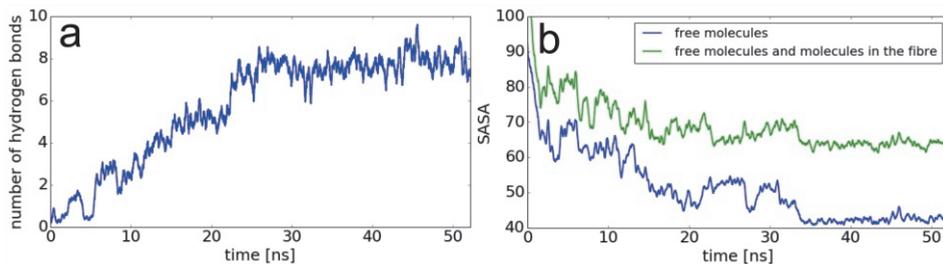


Figure S3.13. Additional analysis of the trajectory of secondary nucleation on the side of the fibre. (a) a number of hydrogen bonds, (b) solvent accessible surface area calculated only for additional free molecules (blue) or for those and the molecules in the fibre (green).

### **cryo-TEM imaging**

Diluted gel suspensions were first vitrified using Leica Vitrobot. About 4 $\mu$ L of sample was casted onto a plasma treated Cu 200 mesh Quantifoil™ grid and blotted against filter paper for 3s. The grid was then plunged into liquid Ethane maintained at -185°C using liquid Nitrogen to achieve sample vitrification. The grid was then transferred to a Cryo-storage container from where they were loaded into JEOL-1400 electron microscope via a Gatan single tilt Cryo holder. Images were recorded at an acceleration voltage of 120 kV at low dosage conditions.

For high resolution images, vitrified samples were loaded into a JEOL JEM3200-FSC microscope and images were recorded at an acceleration voltage of 300kV a low dosage conditions.

Sample pictures of the fibers are present in Figure S3.14 -Figure S3.18. Histograms of widths of bundles were calculated using the imgegeJ. Pictures were rotated in a way that they align with the y-axis. Then the profile was obtained by summing the grey values in rows from a rectangular area. The widths of the fibers were calculated as distances between consecutive minima. The process is presented in Figure S3.19.

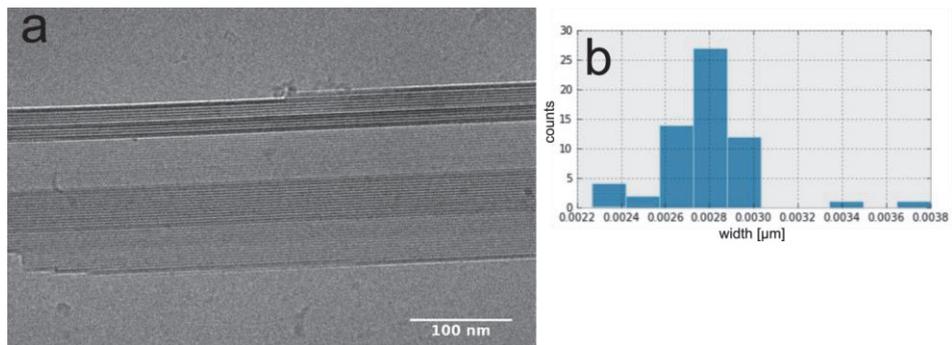


Figure S3.14. (a) Sample cryo-TEM picture of a bundle and (b) histogram of the widths.

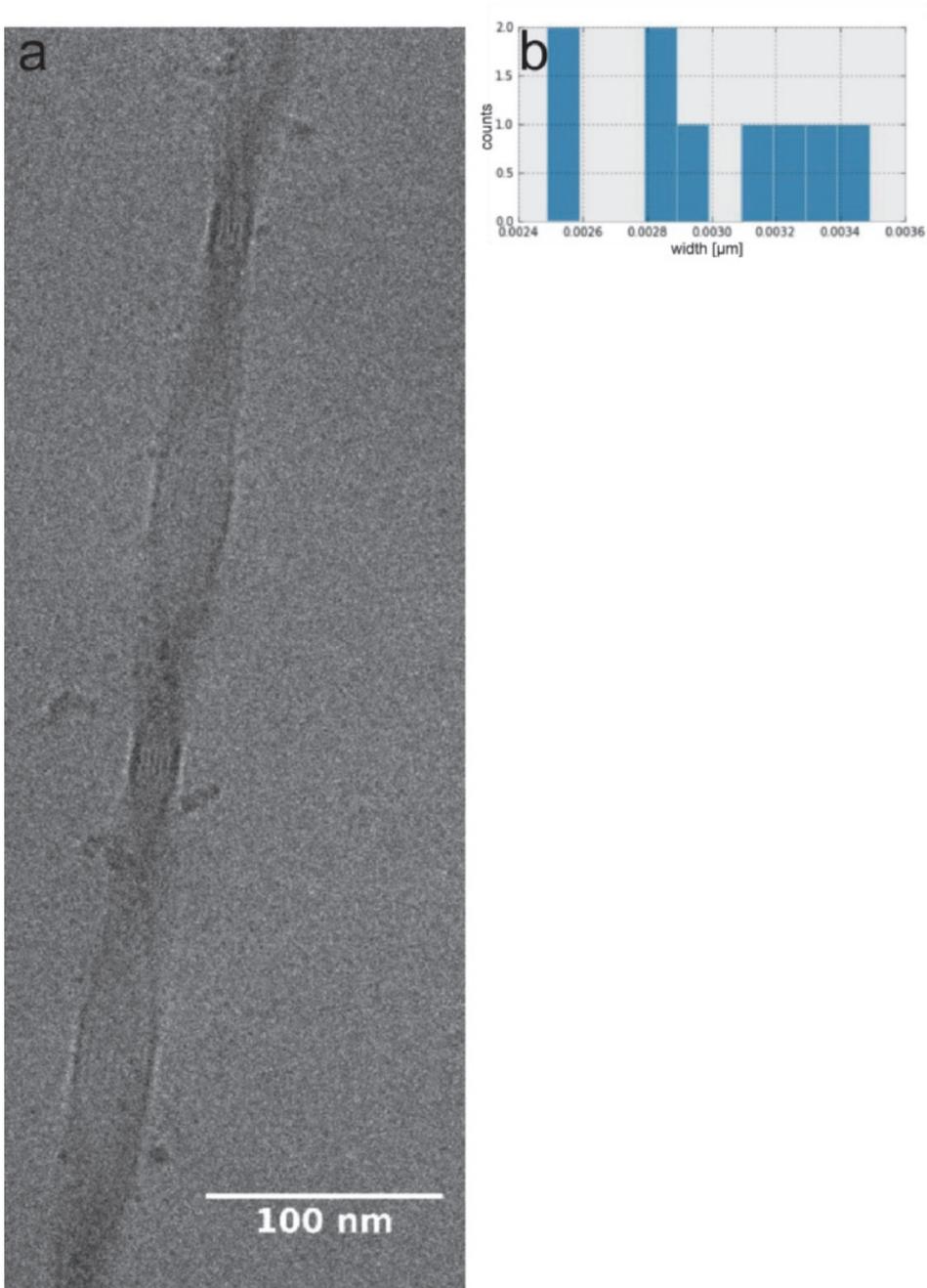


Figure S3.15. (a) Sample cryo-TEM picture of a bundle and (b) histogram of the widths.

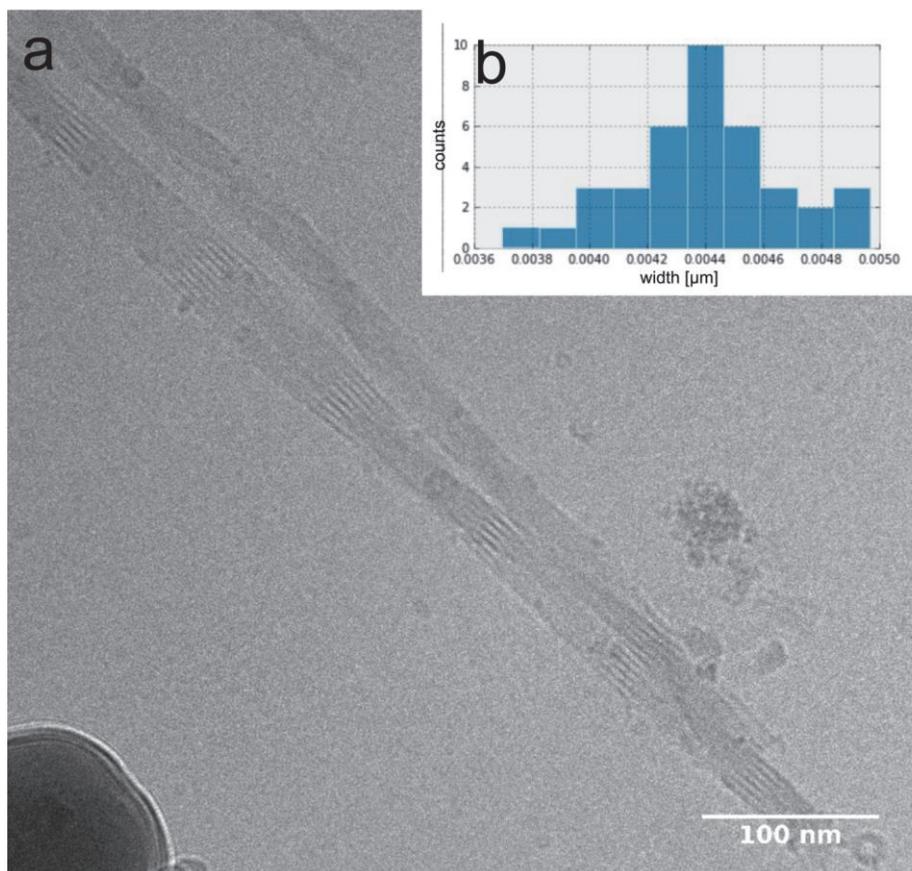


Figure S3.16. (a) Sample cryo-TEM picture of a bundle and (b) histogram of the widths.

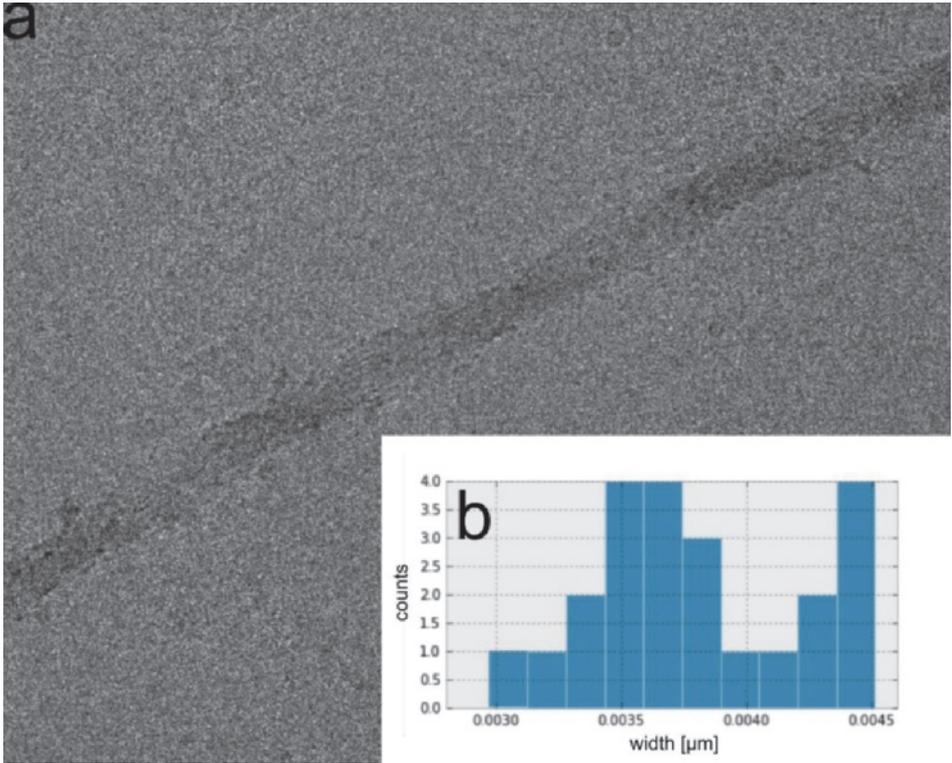


Figure S3.17. (a) Sample cryo-TEM picture of a bundle and (b) histogram of the widths.

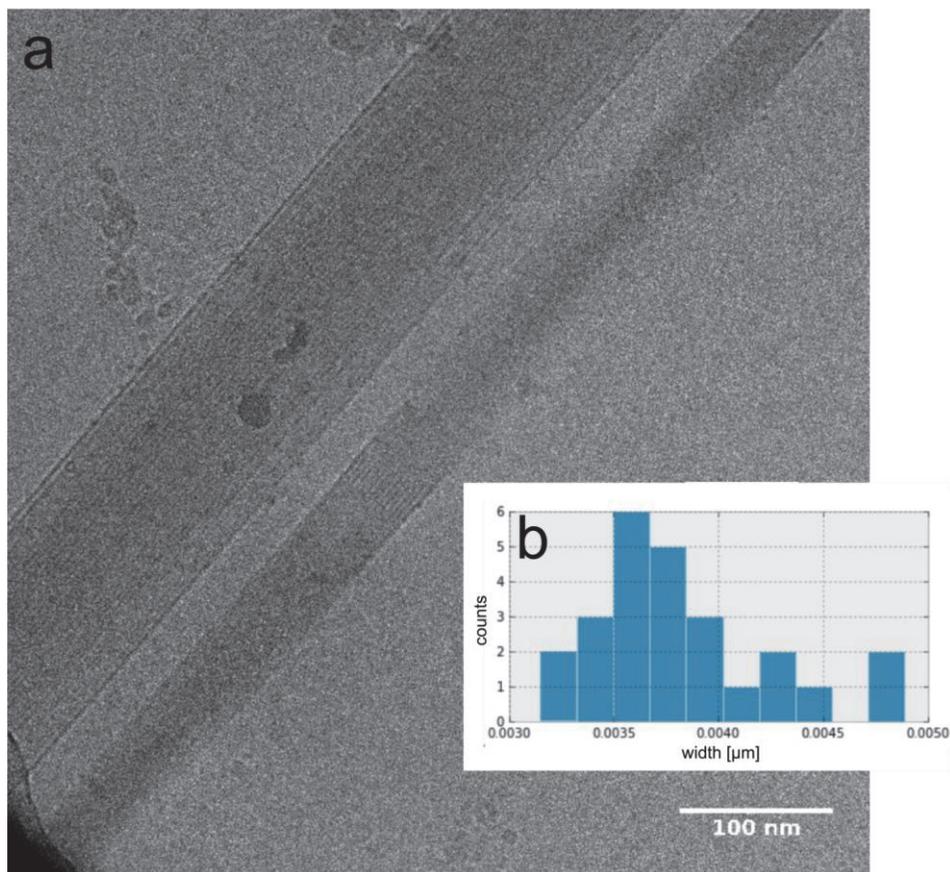


Figure S3.18. (a) Sample cryo-TEM picture of a bundle and (b) histogram of the widths.

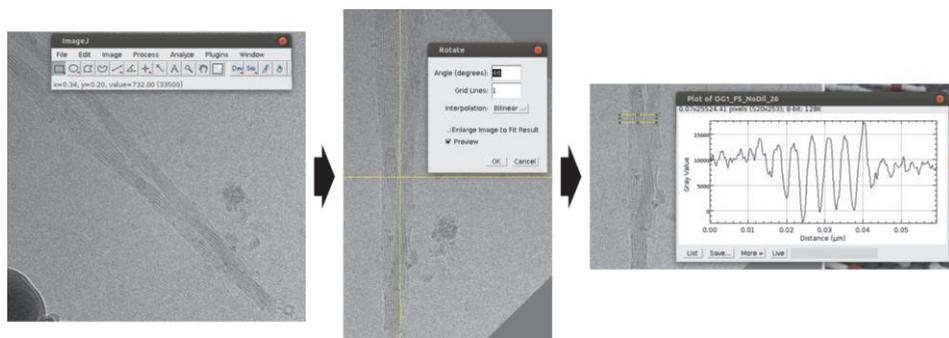


Figure S3.19. Image is processed using ImageJ. Image is rotated that the axis of the fibre is in y direction. Then the gray value is summed in rows resulting in a profile of the fibre.

## Bundling

### Bundling of parallel and antiparallel fibers

To check the bundling of two fibers we have run 20 independent simulation: 16 starting from parallel configuration of fibers and 4 from antiparallel. Although our statistics is rather poor, it seems that there is no difference between bundling of parallel or antiparallel fiber (see individual traces of distance between cores of fibers and their final distribution on Figure S3.20).

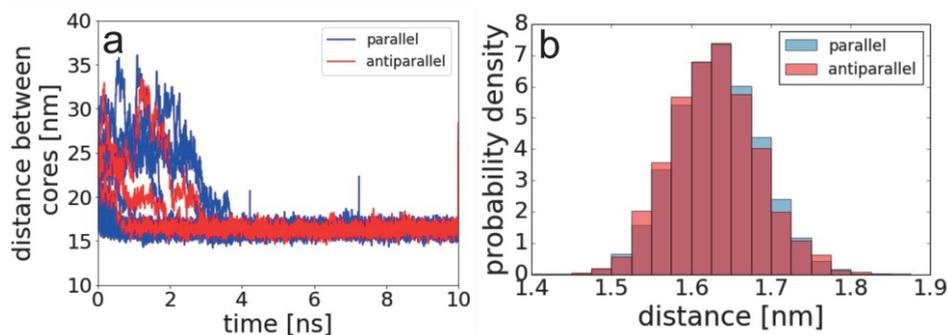


Figure S3.20. Distances of cores of two fibres oriented parallel or antiparallel at the beginning of the simulations) for 20 independent simulations. (a) Progress of the distances over time. (b) Final distribution of the distances (starting from 1ns).

### Four fibers bundling

We have also performed simulations of four fibers bundling. We have run 34 simulations of 10 ns starting from fibers located in the plane of a square (see Figure S3.21a). However, the size of the simulation box was too small to see bundling of just the four fibers, in the sense that many final structures crossed periodic boundary conditions and are therefore representing infinitely long bundles (see Figure S3.21c). Therefore the result here can be treated only semi-quantitatively. In all simulations, we observed strong affinity between fibers. Most importantly, we observed that in all simulations bundles create linear structures and not cyclic ones (which would have a rhombic structure). This could mean two things: bundles have a shape of a ribbon, or their structures consists of more than four fibers. If the fibers tend to form cyclic structures, it would mean that they create tubes with water molecules trapped in it.

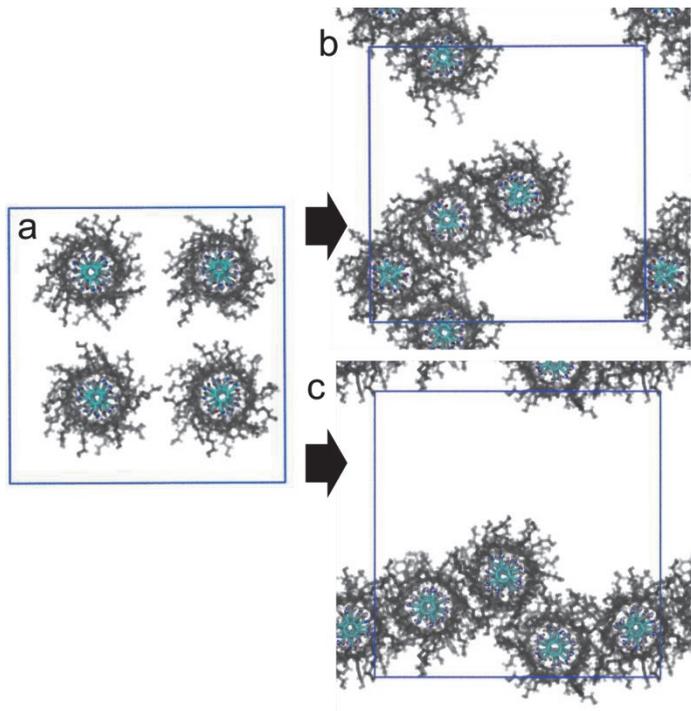


Figure S3.21. Bundling of four fibers. (a) Snapshot of starting configuration. (b) A final bundle which does not cross PBC. (c) A final bundle which crosses PBC.

### Caption to the videos

**SI\_video\_nucleation.mov:** The video shows the trajectory of formation of fiber from randomly distributed molecules, which snapshots are present in Figure 1b-g in the main text. Snapshots are saved every 50 ps and trajectory smoothing is applied for clarity. Cores (cyclohexane ring and amide groups) are indicated by cyan color, and the side branches are indicated by semi-transparent grey.

**SI\_video\_secondary\_nucleation.mov:** The video shows the trajectory of formation of fiber from randomly distributed molecules with a presence of existing fiber. Snapshots of this video are present in Figure 3a-f. Snapshots are saved every 50 ps and trajectory smoothing is applied for clarity. Cores (cyclohexane ring and amide groups) of free additional molecules are indicated by cyan color, cores of existing fiber by orange color, and all the side branches are indicated by semi-transparent grey.

## Chapter 4

# Nucleation mechanisms of self-assembled physisorbed monolayers on graphite

Coarse-grained Molecular Dynamics simulations are employed to obtain a detailed view of the formation of long-range ordered lamellar structures of physisorbed self-assembling long functionalized alkanes on graphite. During the self-assembly, two processes take place: Langmuir preferential adsorption and rearrangement on the surface. The rearrangement starts with nucleation, in which molecules create an ordered domain. The nucleation mechanism is temperature dependent. At lower temperature independent, small and stable nuclei seed the emergence of long-range ordered domains. In contrast, at a higher temperature, molecules adsorb on the surface, and only when a certain level of surface coverage by the adsorbent is reached, the whole structure undergoes a transition from a liquid-like structure to an ordered structure. After this step, relatively slow corrections of the structure take place by Ostwald ripening.

This chapter is published as:

T.K. Piskorz, C. Gobbo, S.J. Marrink, S. De Feyter, A.H. De Vries, J.H. van Esch. Nucleation Mechanisms of Self-Assembled Physisorbed Monolayers on Graphite. *J. Phys. Chem. C*, 2019, 123, 28, 17510-17520.

## Introduction

Self-assembly on a surface has emerged as a promising method to fabricate two-dimensional structures on the nanoscale. In recent years considerable improvement has been achieved in this field as evidenced by the growing number and complexity of obtained structures<sup>22</sup>. Here, we focus on the formation of physisorbed self-assembled monolayers (often referred to as self-assembled molecular networks, SAMNs), which could be an alternative approach to obtaining small two-dimensional devices on the single-molecular scale<sup>111-114</sup>. Such monolayers could have application in industry, e.g. in nanoelectronics, in which conventional methods have almost reached their limits<sup>115</sup>. Despite many efforts it is still challenging to predict the final assembled morphology from the molecular structure of the adsorbent. Therefore, understanding underlying mechanisms of self-assembly is crucial to tackle this problem<sup>116</sup>.

There are many studies devoted to self-assembled molecular networks and many different techniques are used, of which Scanning Tunneling Microscopy (STM) has given the most significant insights<sup>22,117</sup>. STM allows observing the actual distribution of atoms on an atomistic scale. It does not require periodic structure, and therefore it can observe defects, grain boundaries and different polymorphs on the surface. Besides studies of the final assembly, STM has been used to study thermodynamic aspects of the mechanism of formation of SAMNs. A well-accepted model to describe the formation of SAMNs is based on nucleation and growth<sup>118</sup>. Matsuda et al.<sup>119-124</sup> have proposed that similar formation mechanisms as in supramolecular polymerization are present for SAMNs: the isodesmic mechanism, in which molecules as they adsorb to the surface join ordered domains, and the cooperative mechanism, in which a certain concentration of molecules has to be present on the surface before ordered domains can be formed. Depending on the chemical nature of the molecular building block the formation of the assembly can follow one of these paths<sup>120</sup>. However, the kinetics of processes involved in the mechanisms, such as nucleation and growth are too fast to be observed by STM<sup>116</sup>, and the mechanisms that operate often cannot be determined with certainty. To our knowledge, experimental studies at the liquid-solid interface show only how already formed lamellar structures exchange adsorbent molecules with those in solution and rearrange to heal defects<sup>125</sup>. One way to address this problem is to use computational methods which are suitable to study dynamic processes at the molecular length scale and at high time-

resolution. Several simulation studies of self-assembly on a surface have been reported. Most of them are based on two-dimensional (2D) Monte Carlo (MC) and molecular dynamics (MD) simulations<sup>126–133</sup> and study rearrangement of the molecules on the surface. Also the complete process, including adsorption and rearrangement on a surface, was studied for surfactants on a graphite flake<sup>134–137</sup>, however long-range ordered structure formation was not the subject of those studies.

In this paper, we applied coarse-grained Molecular Dynamics simulations to give insights into the complete process of self-assembly of SAMNs on graphite starting from adsorbent molecules in solution and ending with the ordered assembly on the graphite surface. As archetypical molecular layers we selected the widely studied lamellar assemblies of long-chain alkanes and alkane derivatives on highly oriented pyrolytic graphite (HOPG)<sup>138–141</sup>. In these assemblies, linear molecules in solution adsorb onto a substrate, and adopt a straight-chain conformation on the surface, giving rise to parallel lamellae, as has been studied in detail by e.g. STM<sup>139</sup>. Due to the weak forces involved, these structures display high conformational freedom and moreover, the functionalization of alkanes resulted in observation of different arrangements (see Figure 4.1), allowing the investigation of chemical structure differences on monolayer formation. The model described here can predict final structures of the assembly and gives insights into the dynamics of the self-assembly processes, on a time scale that is not yet accessible by other techniques.

## Results and discussion

The recently developed coarse-grained (CG) Martini model for functionalized long-chain alkanes adsorbing onto a graphite surface<sup>142</sup> was used to study the formation and structure of domains of six compounds that differ in substitution pattern and length (Figure 4.1). In the first part, we report the final structures formed on the microsecond simulation time scale and compare them with experimental results. Next, we focus on the mechanism by which these structures are formed, singling out the system (AM25) that forms the structure of highest quality in terms of long-range ordering and domain size.

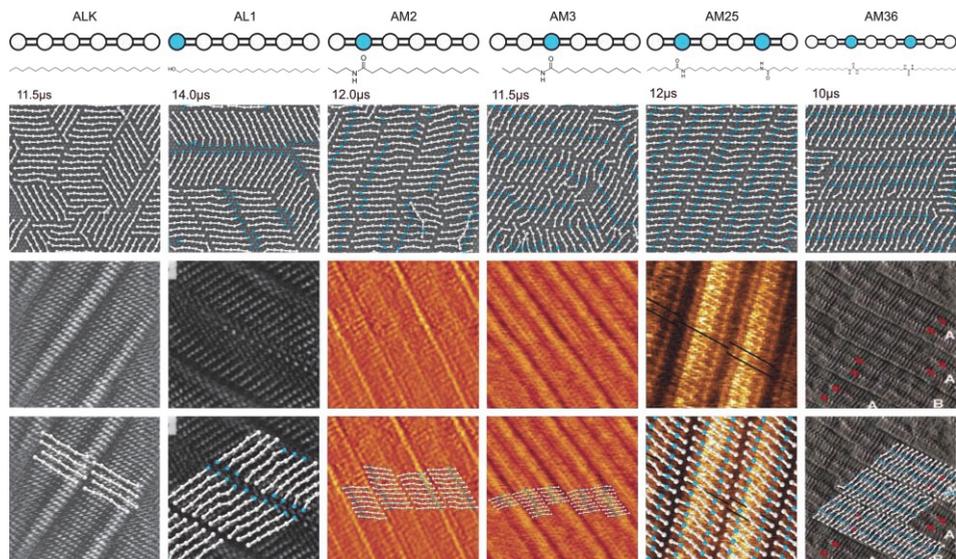


Figure 4.1. Martini model of long-chain functionalized alkanes and final structures of assemblies with comparison to experimental results. The top row shows the Martini models used for the adsorbent molecules labelled ALK, AL1, AM2, AM3, AM25, and AM36. White circles denote alkyl-type beads (C1S and C1E); blue beads denote polar beads (P1). Molecular structures at atomistic resolution are shown below for molecules the coarse-grained model may represent. The second row shows the final self-assembled structures of these molecules on a graphite surface obtained from the simulations and the simulation time at which they were obtained. The third row shows STM/AFM images of the molecules the coarse-grained model may represent (see main text). The bottom row shows experimental results of self-assembled structures with fragments of the simulated structures superimposed. For the molecules represented by structure of ALK, AL1 and AM25, structures were available in literature for the molecules as suggested by the standard mapping of Martini<sup>143–145</sup>. For AM36, AM2, and AM3 we compare with results for the most similar molecules that we found in literature<sup>141,146</sup>. Although lamellar structures for molecules represented by AM36 with amides as the polar groups were reported in literature (which are more distorted in comparison with AM25, which we also observe), the structure was not shown<sup>145</sup>. Therefore here we present results for molecule represented by AM36 with urea groups instead of amide (the urea group prevents molecules from tilting, therefore a different angle for simulation and experiment can be observed).

### Adsorbents form lamellae with polar substituents determining packing efficiency

Figure 4.1 shows a schematic representation of the Martini model of the six adsorbent molecules studied, the final structures of the six systems obtained after at least 10  $\mu\text{s}$  at 298 K, together with experimentally determined STM/AFM images of representative molecules, and a superposition of the CG structure on the experimental images. All of the adsorbent molecules form

lamellar structures. The quality of the structures in terms of alignment of molecules increases with the number of polar substituent groups: it is strongest for molecules with two polar groups (AM25, AM36), less strong for molecules with one polar group (AL1, AM2, AM3) and least for the apolar molecules (ALK). This tendency agrees with experimental results, in which amide groups impose strong directionality in assemblies<sup>147</sup>. Molecules in the lamellae tend to align with the underlying graphite structure, of which the Martini beads are hexagonally packed, leading to three equivalent orientations. The main axis of the molecules aligns with a basal line of the graphite structure. Therefore, almost all adsorbed molecules are parallel or rotated by multiples of 60 degrees with respect to each other (further quantification is given in the SI). Functionalized alkanes (i.e. AL1, AM2, AM3 and AM25) form assemblies which well reproduce experimental results. An exception is the long-chain alkane (ALK) for which molecules in simulation are also oriented with respect to each other in multiples of 60 degrees, whereas in the experiment they form lamellae whose axis is perpendicular to the molecular vector. This shows a limitation of our model as a result of coarse-graining; understanding this limitation may enable development of a CG model that reflects the long-chain alkane packing (see SI for figures).

In the simulations, molecules with two symmetrically placed polar groups, AM25 and AM36, formed lamellae of the best quality in terms of alignment of the molecules and the size of the domains, at least on the time scale of the simulations. The quality of all final structures is further assessed in the SI. The excellent alignment of molecules with two symmetrically placed polar groups can also be seen in experimentally determined structures (Figure 4.1). Molecules with one polar bead, AL1, AM2 and AM3, due to their asymmetry, can orient in two distinct directions inside lamella, i.e. parallel and anti-parallel. AL1 and AM2 neighbouring molecules are predominantly parallel, driven by the favorable polar-polar and apolar-apolar interactions. Moreover, neighboring AL1 lamellae tend to face each other with their polar groups, and their final assembly features many defects, in which lamellae show kinks or are oriented by 120 degrees with respect to each other. The abundance of these defects agrees with experimental results for long-chain alcohols<sup>144</sup>, and is further shown in the SI. Comparing AM3 and AM2 assemblies, the lamellae of AM3 can be seen to be more disordered than those of AM2, see Figure 4.1. Two neighboring molecules within these lamellae can adopt a number of conformations, characterized by being parallel or antiparallel on the one hand,

and showing an offset on the other hand. The minimum offset in our model is half a bead in perfectly aligned lamellae (this leads to domain boundaries that are oriented at multiples of 60 degrees). For both AM2 and AM3, the parallel-minimum offset conformation is the most frequent. In this conformation, the polar beads are next to each other and multiple lamellae can be stacked parallel to each other without defects or domain boundaries. A common relatively small defect is a so-called twin boundary, in which the offset changes from positive to negative. This results in a kink in the lamella. Another twin boundary with a negative to positive offset further down the lamella then causes a zig-zag appearance; these types of structures have been reported for long-chain molecules<sup>148</sup>. This is a relatively cheap defect as long as parallel lamellae all follow the same zig-zag pattern. Minimum offset but antiparallel conformations are costly: the polar beads are no longer neighbors, and the arrangements breaks multiple relatively favorable polar-polar and nonpolar-nonpolar interactions in return for relatively unfavorable polar-nonpolar bead interactions. The difference between AM2 and AM3 is, that an offset of one bead (in the right direction of course) can restore the polar-polar and nonpolar-nonpolar interactions for the AM3 molecule, but not for the AM2 molecule. The larger offset causes some misalignment between neighboring lamellae, but is apparently not prohibitively costly. AM2 would need an offset of two beads to restore favorable polar-polar and nonpolar-nonpolar interactions, which apparently leads to such a severe packing defect that it is only seen near the edges of the ordered assemblies.

In closing this section, it should be noted that coarse-graining of the system allows extending simulations to microsecond scale, but it is still several orders of magnitude shorter than the experimental scale (minutes). Furthermore, in coarse-graining, some details of the chemical structure are lost, and a comparison to the experimental systems can only be interpreted semi-quantitatively. In that context, it should also be noted that we do not have a close experimental match for AM36 model to compare with (the image on Figure 4.1 shows a bisurea functionalized molecule - the closest molecule which we found in the literature to AM36 which represents a bisamide molecule).

## Self-assembly mechanism depends on temperature

Having shown that the Martini model is capable of yielding well-ordered lamellar structures of adsorbent molecules on graphite, the model is used to study the mechanism by which such structures are formed during the self-assembly process. We observed two processes occurring simultaneously throughout self-assembly: adsorption of long-chained molecules to and rearrangement of these molecules on the surface. In the latter process we could distinguish two phases: first a nucleation phase, in which an initial ordered structure is created, followed by an Ostwald ripening phase<sup>149–151</sup>, where the final ordered structure emerges. We also observed that depending on the temperature, the mechanism of nucleation changes. At higher temperatures (in the simulations that is 298 K) the adsorbed molecules initially form a liquid-like phase, which changes to the well-ordered structure at higher surface coverage. In contrast, at lower temperatures (258K), adsorbed molecules immediately associate with each other creating semi-ordered clusters, which grow and eventually cover the graphite surface. To study the self-assembly process in more detail we focused on AM25, which yields the highest quality long-range ordered final structure on the simulation time scale.

### Cooperative nucleation mechanism at high temperatures

At higher temperature, AM25 molecules adsorb on the graphite surface forming a liquid-like phase, which after a while transforms to an ordered aligned structure. Figure 4.2a-h show representative snapshots of this process (for the other molecules see SI). Starting from a random solution (Figure 4.2a), the molecules rapidly adsorb onto the surface displacing the solvent phenyloctane (Figure 4.2b-c). They form highly dynamic clusters with parallel oriented molecules (Figure 4.2c-d), and diffuse on the surface as individuals and as clusters. The formed structure resembles a two-dimensional liquid and stays in this phase until no more molecules can adsorb on the surface (approximately 70% of surface coverage) unless reorganization occurs (Figure 4.2e-f). When 90% of the surface is covered by adsorbent molecules large lamellar domains can be seen (Figure 4.2f), which further rearrange into relatively large ordered structures (Figure 4.2g). After this stage, the assembly undergoes the slow process of Ostwald ripening: large clusters grow at the expense of small clusters, leading to a single domain that covers almost the entire graphite flake (Figure 4.2h). Most of the final structure is stable and just a small fraction, on the edges of the flake, remains dynamic. On much longer time-scales, remaining

defects heal; the mechanisms of the healing process are the subject of a recent study<sup>152</sup> and are not discussed here.

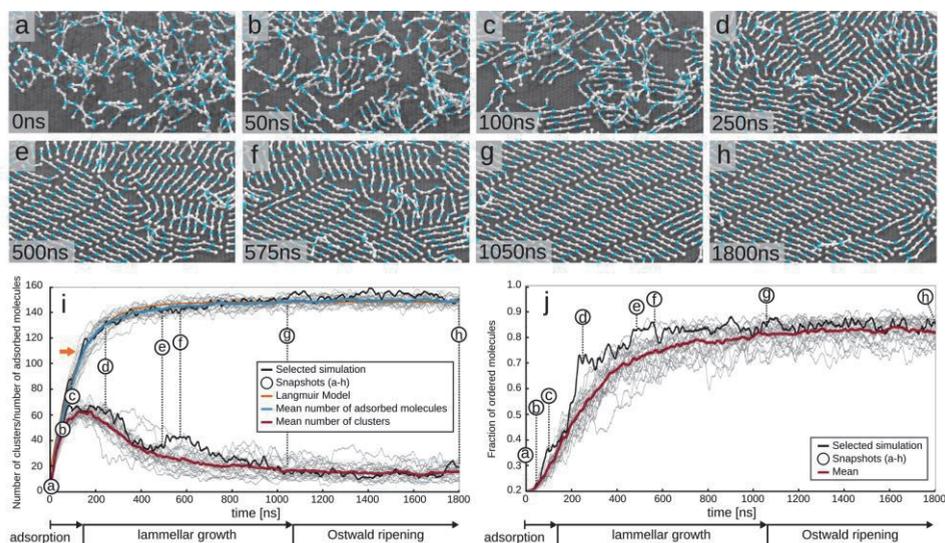


Figure 4.2. Snapshots from a representative simulation of AM25 at 298 K. (a-b) adsorption phase, (c-g) lamellar growth phase, (h) structure after 1.8  $\mu$ s of simulation. Solvent (phenyloctane) is not shown for clarity. The process of self-assembly on the surface can be described by monitoring the number of adsorbed molecules on the surface and the number of clusters (i), as well as the fraction of ordered molecules (j). In (i) and (j), gray lines indicate results for ten independent runs (i.e., twenty measurements: for every simulation there are two surfaces on which self-assembly process takes place) at 298 K, the thick black line results for an example trajectory (black dots correspond to snapshots a-g). In (i), the orange arrow indicates 70% surface coverage. The mean of the number of adsorbed molecules is shown as the thick blue line, and the mean of the number of clusters as the thick red line. A fit of the mean number of adsorbed molecules to the Langmuir absorption rate law is shown as the thick orange line. In (j), the average fraction of ordered molecules is shown as a thick red line.

The self-assembly process is stochastic in nature – it is difficult to create a clear image of the process from a single trajectory, but after several independent self-assembly simulations, clear patterns characterizing the mechanism can be observed. The first 1.8  $\mu$ s of ten runs were analyzed by monitoring, as a function of time, the number of adsorbed molecules, the number of clusters, the fraction of ordered molecules, the surface coverage of ordered molecules, and the area of the largest cluster. A detailed description of the metrics can be found in the SI, but briefly, a cluster is defined by a combination of distance and orientation criteria between neighboring adsorbent molecules which are designed to detect groups of molecules that form a lamellar structure. Results of

selected metrics for ten independent simulations are shown in Figure 4.2i-j as thin gray lines, together with the average metrics in thicker lines, and provide a clear picture on how surface coverage develops over time. We observed that the rate of adsorption follows the simple Langmuir growth law<sup>153,154</sup>: the rate of adsorption is proportional to the number of free spaces on the surface (Figure 4.2i, orange line). Thus, the number of adsorbed molecules can be expressed as a function of time by the relation:

$$n_{ads}(t) = n_{max}(1 - e^{-k_a t}) \quad (4-1)$$

where  $k_a$  is the adsorption rate and  $n_{max}$  is the maximal number of adsorbed molecules. In the beginning, adsorbed molecules do not create clusters: the number of adsorbed molecules (Figure 4.2i, blue line) is almost the same as the number of clusters (Figure 4.2, red line), i.e. nearby molecules are not identified as forming well-defined lamellae and single free molecules are present on the surface. After 50-100 ns there is not enough space to adsorb new free molecules, which leads to rearrangement of adsorbed molecules resulting in the appearance of larger clusters. After the initial adsorption phase, the number of clusters reaches a maximum and from this point on, the number of clusters steadily decreases and the structures rearrange to lamellae with longer-ranged order (Figure 4.2i).

The growth of long-range ordered structure can be described by the Kolmogorov–Johnson–Mehl–Avrami (KJMA) law<sup>150</sup>:

$$A(t) = A_{lim}(1 - \exp[-(kt)^n]) \quad (4-2)$$

where  $A(t)$  denotes the size of the structure of interest at time  $t$ ,  $A_{lim}$  is the final size of the cluster, and  $k$  and  $n$  are constants describing the rate of growth. The simplicity of the KJMA law has led to successful application in many phase transformation processes<sup>150</sup>. In most applications, all of the organized structure is monitored; due to the relatively small size of our surface, we were limited to measuring the size of the largest cluster instead. Figure 4.3a shows the best fit (orange line; fit parameters are given in the SI) of the size of the largest cluster (averaged over 10 independent runs, red line) to the KJMA law. Although the KJMA law perfectly describes the first stages of growth of the largest cluster, it fails to describe later stages of cluster growth in our simulations. In our

simulations, the size of largest cluster does not stabilize, but it slowly continues to grow. This continuous growth is caused by Ostwald ripening (OR); large clusters are growing at the expense of small ones. The correction for this phenomenon can be expressed by an additional term in the KJMA function<sup>150,155</sup>:

$$A(t) = A_{lim}(1 - \exp[-(kt)^n]) + \left( \frac{t - \tau_{OR}}{1 + \exp[-2\omega(t - \tau_{OR})]} \right) k_{OR} \quad (4-3)$$

where  $k_{OR}$  describes the rate of Ostwald ripening and  $\tau_{OR}$  describes the beginning of this process. The expression  $\exp[-2\omega(t - \tau_{OR})]$  is a switching-on function (for  $t \ll \tau_{OR}$  the Ostwald ripening part is absent) and the  $\omega$  parameter describes the smoothness of the switching (width of the turning-on period). The KJMA+OR function as observed from our multiple independent simulations describes the mean value of the size of the largest cluster as a function of time well, and the best fit is shown in Figure 4.3a as the blue line. However, the fitted rates of Ostwald ripening appear to be rather large in comparison to available experimentally observed rates<sup>156</sup>, indicating that these results should be regarded only semi-quantitatively (see SI for detailed discussion). It should be noted that the experimental data are obtained for larger and more complex molecules than the AM25 studied here; other possible reasons for this discrepancy are discussed further in the SI. The noise in single measurements is too large to fit this function reliably for the individual traces. This shows that the KJMA law is valid as a description of collective macroscopic quantities of the sample rather than of a single measurement.

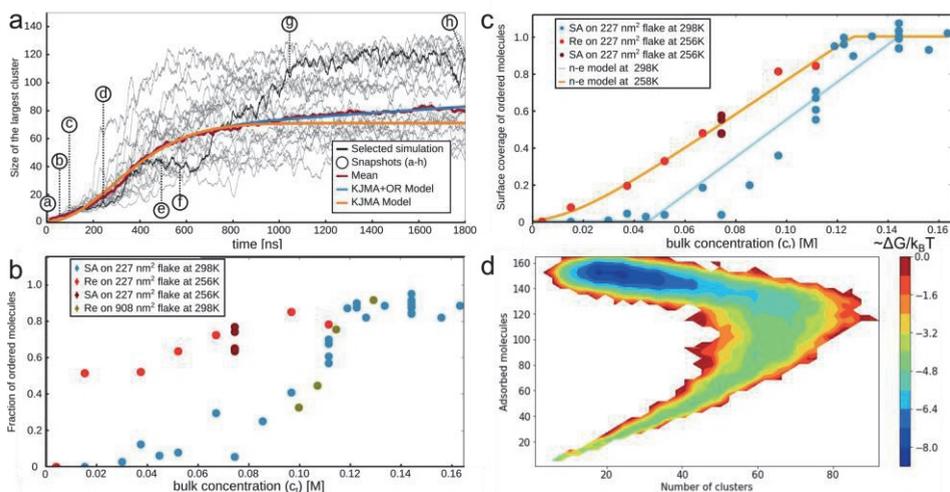


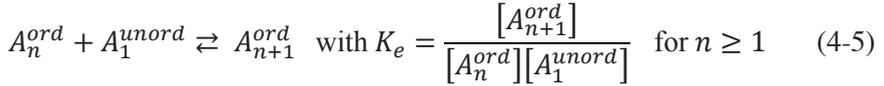
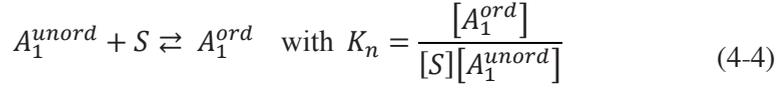
Figure 4.3. Size of the largest cluster as a function of time, fraction of ordered molecules as a function of concentration, and fraction of surface coverage of ordered molecules as a function of concentration, at different temperatures. (a) Size of largest cluster as a function of time. Results from ten independent simulations at 298 K are shown as thin grey lines. The mean value from the simulations is shown as a thick red line. A fit of the mean value to the KJMA function with correction of Ostwald ripening (KMJA+OR, see text) is shown as a thick blue line. For comparison a fit to the KJMA function without Ostwald ripening (KJMA) is also shown (thick orange line). (b) Fraction of ordered molecules as a function of bulk (i.e. total) concentration of adsorbent molecules in the system. Results are shown for two different graphite flake sizes and two different temperatures. At 258 K the ordering of the adsorbent molecules is higher than at 298 K at the lower bulk concentrations. At higher temperature, a relatively high concentration is required to obtain an ordered structure. (c) Fraction of surface covered by ordered molecules as a function of adsorbent bulk concentration. A fit of the nucleation-elongation model (n-e model) is shown as a thick light blue line at 298 K (fitted values:  $\sigma = 6 \cdot 10^{-3}$ ,  $K_e = 19 M^{-1}$ ,  $\alpha = 0.091 M$ ), and as a thick orange line at 258 K ( $\alpha$  is assumed to be the same as for 298 K (0.091 M), fitted values:  $\sigma = 4 \cdot 10^{-6}$ ,  $K_e = 33 M^{-1}$ ). For (b-c) results are presented for final frames of simulations of self-assembly (SA) from randomly distributed molecules in solvent, or for a simulation of the rearrangement (Re) of adsorbed molecules that were already on the surface. Simulations were performed on graphite flakes with sizes of 227 nm<sup>2</sup> and 908 nm<sup>2</sup>. (d) Free energy landscape obtained from distribution of states using metrics of number of adsorbed molecules and number of clusters. The color scale reflects the free energy proportional to the negative logarithm of the probability. Projections with fraction of ordered molecules see SI.

Interestingly, during the self-assembly process, we have not observed critical nuclei, i.e. stable clusters of minimal size. We suspected that the absence of stable nuclei is due to the highly dynamic nature of the formed clusters, but (1) the time scale in which the self-assembly is complete may also be too fast for the formation of thermodynamically stable clusters, and (2) the system is

relatively small, which may mean that the simulation model still thermodynamically favors well-defined and stable nuclei. We conducted simulations with fewer adsorbent molecules to better sample the early stages of the process and to investigate the intrinsic equilibrium dynamics of the molecules at fixed surface coverages below 90%. Depending on the computational cost of a simulation we performed self-assembly simulations (i.e. starting from randomly distributed molecules in solvent; denoted ‘SA’), or cheaper rearrangement simulations (i.e. the initial structures already contain molecules adsorbed on the surface, to avoid waiting for adsorbent molecules to reach the surface; denoted ‘Re’). The results are presented in Figure 4.3b (blue symbols), which shows the fraction of ordered molecules as a function of bulk concentration at equilibrium. At low concentrations most of the adsorbent molecules form an unordered phase on the graphite surface. With increasing concentration, at first small (but still dynamic) ordered clusters are observed. Further increased concentration changes the appearance of the system from a liquid phase to an ordered phase and the number of ordered molecules is increasing. Long-ranged stable domains are only found above ~80% coverage, as evidenced by a high fraction of ordered molecules. Although the fraction of ordered molecules increases with the concentration we have not observed small stable assemblies, which could act as nucleus. This behavior was independent of the system size: even if the size of the graphite flake (and the number of adsorbent molecules) was increased by a factor of four, almost full coverage of the surface was necessary to observe a long-ranged ordered domain (Figure 4.3b, olive symbols).

Despite the fact that clear nucleation events have not been observed in our simulations, the transition from an unordered to an ordered assembly on the surface can be described using the nucleation-elongation model for 2D self-assemblies at liquid/solid interface, developed by Matsuda et al.<sup>119–124</sup> (Figure 4.3c). Our modified version of this model (Eqs. (4-4) and (4-5)) assumes, that the assembly process can be described by two equilibrium constants: (a) a nucleation constant  $K_n$  between an adsorbent molecule in solution or on the surface in an unordered state (denoted  $A_1^{unord}$ ) and a free surface adsorption site ( $S$ ) on the one hand, and an ordered state ( $A_1^{ord}$ ) on the other hand (a molecule is in an ordered state when it has two close neighbors with the same orientation, by which we mean it can be parallel or anti-parallel, see SI for details), which acts as a nucleus on the surface; and (b) an elongation constant  $K_e$  between an adsorbent molecule in an unordered state and ordered clusters of any size ( $A_n^{ord}$ )

adsorbed on the surface (note that in the original model of Matsuda et al.<sup>13</sup> this equilibrium is only between monomers in solvent and clusters on the surface):



Using the steady-state approximation, an expression for the surface coverage of ordered molecules,  $\theta$ , as a function of bulk concentration of adsorbent,  $c_t$ , can be derived (a detailed derivation can be found in SI):

$$\theta = (1 - \theta) \frac{\sigma K_e (c_t - \alpha \theta)}{\{1 - K_e (c_t - \alpha \theta)\}^2} \quad (4-6)$$

where  $\sigma$  is the degree of cooperativity defined as the ratio between the two equilibrium constants for nucleation and elongation,  $\sigma = K_n/K_e$ , and  $\alpha$  is the maximum concentration of ordered molecules on the graphite surface. Parameter  $\sigma$  describes the tendency of formation of ordered structures depending on the number of molecules already present. When  $\sigma$  is close to one, adsorbed molecules have a high tendency to align with other molecules on the surface, even if just a few molecules are present (referred to in literature as the isodesmic mechanism<sup>13</sup>). When  $\sigma$  is close to zero this tendency is lower; molecules only tend to order when many molecules are in their vicinity (referred to in literature as the cooperative mechanism<sup>13</sup>). The data obtained for the surface coverage of ordered molecules from the equilibrium simulations at fixed concentration was fitted to the model, using  $K_e$ ,  $\sigma$ , and  $\alpha$  as variables (see Figure 4.3c). The fit shows that  $\sigma$  is smaller than one, which means that a specific concentration of unordered molecules is needed to be present on the surface before elongation sets in. Indeed, molecules hardly order (i.e., interact only weakly with neighbors) until a certain concentration is reached (see Figure 4.3c; ordering starts at around  $\sim \frac{1}{K_e}$ , which corresponds to  $c_t \sim 0.04 M$ , see also SI). After this concentration is reached, the system undergoes a phase transition from a liquid-like phase to an ordered phase. Complete ordering of the assembly on graphite starts with bulk concentration above 0.14 M. Since a similar trend can be found for the larger graphite flake, the ordering starts with a specific

concentration of molecules on the surface, rather than with a specific number of molecules on the surface (the latter would point to the existence of a critical nucleus of specific size). We also tried to develop a model which explicitly incorporates unordered adsorbed molecules as a separate state; however, the simpler nucleation-elongation model gives the most accurate description of our data (for the other model see SI).

An increasingly popular analysis of simulation data is Markov state modeling<sup>101–103</sup>, which has been applied to relatively simple self-assembling systems with some success<sup>51,157</sup>. We investigated two Markov State Models (MSMs), one using the global metrics used thus far to describe the self-assembly process and one using the molecule-based metric that determines which molecules can be classified as neighbors in an ordered cluster. The latter has been advocated as superior in self-assembling systems that form ordered structures<sup>158</sup>. Details can be found in the Supporting Information. Both MSMs indicated similar implied time scales for the slowest processes, suggesting analysis in terms of three hidden states. The three states can be interpreted as one consisting of largely unordered molecules (free in solution or adsorbed on the surface), and two interconverting states that differ in a subtle manner in their metrics, probably indicating more extensive and less extensive ordered domains. No metastable states are detected in this analysis. A free energy landscape in terms of the global metrics displays only a single minimum, the final assembled, ordered state (see Figure 4.3d).

### **Isodesmic nucleation mechanism at low temperatures**

In contrast to the cooperative nucleation observed at higher temperature, we found that at the lower temperature the nucleation mechanism resembles isodesmic nucleation. At high bulk concentration at low temperature, the presence of multiple nuclei in combination with rapid adsorption leads to relatively many domains or defects (particularly zig-zag patterns) that do not heal on the time scales simulated here and also does not allow clear distinction of the mechanism (see SI). Therefore, for simulations at low temperature we used the concentration of AM25 almost twice as low as for the simulation at the higher temperature of 298 K (approximately 0.07 M), to avoid rapid saturation of the surface. In fact, we made sure that the number of adsorbent molecules is insufficient to saturate the surface to be able to better characterize the assembly process. Representative snapshots are shown in Figure 4.4a-h. In contrast, in simulations for such concentration at higher temperature the final structure does

not display any ordering of the assembly as discussed in the previous section (see Figure 4.3c). At low temperature, randomly distributed molecules in phenyloctane (Figure 4.4a) adsorb on the surface and quickly aggregate creating small, thermodynamically stable, clusters (Figure 4.4b-c). These structures act as nucleation sites and grow to larger ordered structures over time as new molecules arrive on the surface and join the existing clusters (Figure 4.4d-f). It can be seen that the arrival of molecules is slower than at higher temperature, due to slower diffusion in the solution and to lower bulk concentration. The clusters on the surface largely maintain their orientation, and are not seen to undergo large reorganization on the time-scale of the simulation of 5  $\mu$ s (cf. Figure 4.4f-h, and a movie in the SI). This is also due to the slower diffusion of molecules and clusters on the surface.

The details of the adsorption and nucleation stages were investigated by running four independent simulations of 5  $\mu$ s, monitoring the same quantities as at higher temperature and averaging them. The results are shown in Figure 4.4i-j. In contrast to the self-assembly process at higher temperature, at the lower temperature, right from the start, the number of adsorbed molecules (Figure 4.4i, blue line) is larger than the number of clusters (Figure 4.4i, red line), which means that many molecules are part of clusters shortly after they adsorb to the surface. The Langmuir adsorption rate law (a fit is shown as the orange line in Figure 4.4i) is valid over the length of the simulation. The entire structure consists of many, small, ordered domains at different orientations. This is consistent with experimental evidence showing that upon surface saturation, the number of domains formed at low temperature is larger than at high temperature<sup>159</sup>. Mechanisms to heal such defects are discussed elsewhere<sup>152</sup>. At the lower temperature, there is no transition from an unorganized liquid-like phase to a solid phase, and most of the molecules are ordered (Figure 4.3b). In fitting the nucleation-elongation model (shown in Figure 4.3c), we assumed that  $\alpha$  is the same as for self-assembly at high temperature, i.e. the maximum amount of molecules which can adsorb does not change with temperature. From the fit it was found that at low temperature  $\sigma$  is close to 1, which confirms that molecules which adsorb on the surface tend to create ordered clusters even for low bulk concentration. Although potentially interesting, MSM was not possible at the lower temperature due to too limited sampling; many more repeats of the simulations and/or adaptive sampling techniques would be required. In closing this section, it should be noted that although at the lower temperature the clusters preserve the overall structure and

act as independent and stable nuclei, to some extent they do demonstrate some desorption-adsorption dynamics, i.e., molecules detach from and attach to the clusters during the simulation; more information is given in the SI.

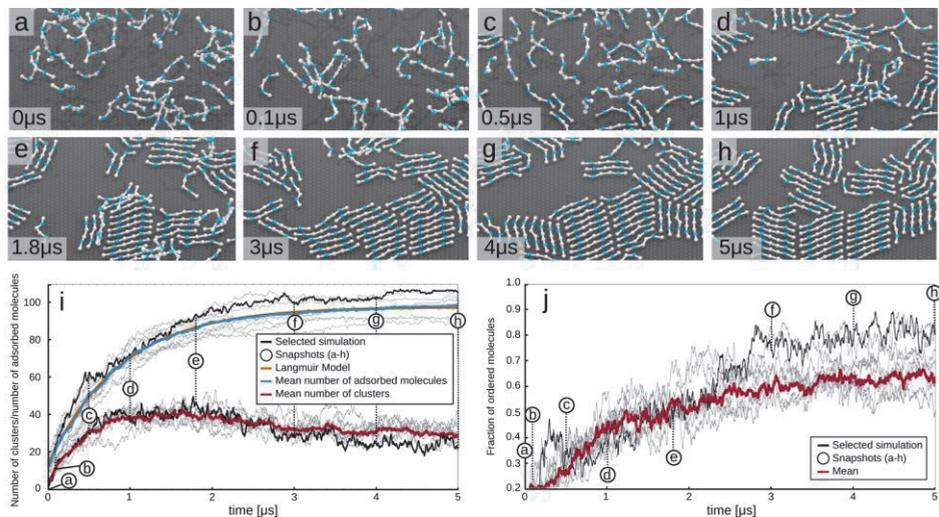


Figure 4.4. Snapshots from a representative simulation of AM25 at 258 K. (a-h) Adsorption and nucleation phase. (i) and (j) information as for Figure 4.2, see caption to Figure 4.2.

### Similarities to other processes

We noticed that processes described here for physisorption are similar to those occurring in crystallization. Two established mechanisms are distinguished for crystal nucleation: the single step nucleation mechanism and the multistep nucleation mechanism<sup>11,12</sup>. In the single step nucleation mechanism, which is based on classical nucleation theory (CNT), molecules strongly interact with each other. Once a critical number of molecules form an assembly, the interactions are holding the molecules together in a stable cluster, called the nucleus. The nucleus grows by acquiring additional molecules, eventually creating a large long-range ordered structure. For our system a similar mechanism takes place at lower temperature. Molecules strongly interact with each other forming a nucleus (in our case a molecule strongly interacts already with a single neighbor). Alternatively, in the multistep nucleation mechanism, the interaction between molecules is weaker and the system stays in a liquid state: molecules diffuse, interact with each other, but do not create stable structures. With growing concentration, eventually local oversaturation occurs, in which ordered structure starts to appear, eventually expanding to form long-

range ordered structure. This kind of mechanism applies for formation of SAMNs at higher temperature. Molecules firstly form a liquid-like phase, from which upon saturation forms the ordered structure.

The self-assembly of functionalized alkanes on graphite also resembles the formation of self-assembled monolayers (SAMs) on other surfaces. One of the most studied SAMs systems is thiols on a gold surface. The process can be described as a special form of chemisorption, in which the adsorbent interacts with the substrate by covalent, but reversible interaction. The covalent binding and unbinding allows the adsorbent to retain some lateral diffusivity on the gold surface, which is, however, sufficiently slow to make an observation of the processes of adsorption and of rearrangements on the partially covered surface possible by experimental techniques<sup>160</sup>. Similar observations are made here for physisorbed alkane assembly on graphite as described for SAMs: the mechanism is temperature dependent<sup>160</sup>. Below a certain temperature molecules start to form small ordered domains, which grow and eventually cover the whole surface. Above this point, molecules do not coalesce and create a condensed phase (liquid-like or ordered structures with lower density). Upon increasing the concentration of adsorbent on the surface the structure reorganizes into the ordered assembly. Although the underlying nature of the interaction between adsorbent and surface is different, both behaviors are similar, and the most significant difference is their time-scale, which in case of the graphite surface is so short that it is challenging to study the process by experimental techniques.

From a molecular simulation perspective, the slowness due to the covalent binding and unbinding equilibrium warrants methods that predict morphology of SAMs by adding molecules one by one, running short steered MD simulations to test if a molecule wants to adsorb to an increasingly occupied surface, as recently done by Dietrich et al.<sup>161</sup> They found that with that procedure, maximum occupation of the surface as known from experiment is not reached in the simulations. The final patterning was therefore later imposed by allowing binding only at certain sites. A better treatment would be to include sampling of the unbinding of the adsorbents, breaking the chemical bonds and possibly full desorption. The technique would need time-scales for the slow processes as input. Previous work by us on the physisorbed monolayers revealed that partial desorption of the alkanes is closely linked to the

rearrangement of molecules on a saturated surface and that the time-scales for these rearrangements are accessible by our model<sup>152</sup>.

## Conclusions

In this work, we present unprecedented insights concerning the formation of long-range ordered lamellar structures of physisorbed self-assembling long functionalized alkanes on graphite. The simulations show that the overall mechanism consists of two simultaneous processes, being adsorption to and rearrangement on the surface. The rearrangement process starts with nucleation (formation of long-range ordered structures) followed by Ostwald Ripening. The mechanism of nucleation is temperature dependent. At lower temperature the mechanism is more similar to isodesmic nucleation: adsorbed molecules rapidly form independent, small, ordered domains, which subsequently grow by addition of new arrivals. The different orientations of the domains lead to many domain boundaries and the overall order remains low. At higher temperature, the nucleation process can be described by cooperative nucleation. Here, molecules adsorb on the surface forming a liquid-like phase, and only when a certain level of surface coverage by the adsorbent is reached, the whole structure undergoes a phase transition to an ordered structure. In both cases, after the nucleation step, relatively slow corrections of the structure can be described by Ostwald ripening.

Taken together, in this work we have provided insights into the nucleation mechanism of physisorbed alkanes which is not accessible by other techniques such as STM. Moreover, our method can be used to predict the structure of the final assembly.

## Methods

**Model.** A recently developed modified Martini coarse-grained model was used to study the adsorption, structure, and dynamics of functionalized long-chain alkanes on a graphite surface<sup>87,142</sup>. The model treats groups of roughly four atoms (not counting hydrogens) as the basic building blocks, called beads. For aromatic moieties, a two-to-one correspondence (mapping) is used to preserve the ring structure. The mapping of the long-chain functionalized molecules used in this study is shown in Figure 4.1. The Martini model was originally developed to study the self-assembly, structure and dynamics of lipids<sup>72</sup>, but was successfully extended to study a variety of other self-assembly processes such as bulk heterojunctions<sup>162</sup>, supramolecular assemblies<sup>163</sup>, and oil-water

emulsions<sup>164</sup>. Briefly, bonded interactions include harmonic bonds between neighboring beads and cosine-type harmonic angle potentials between triplets of beads. Non-bonded interactions are modeled using a shifted form of the Lennard-Jones interaction, ensuring that both potential and force smoothly approach zero at the cut-off distance of 1.2 nm, starting the shifting from 0.9 nm. None of the adsorbents used in this study carry partial charges and there are no Coulomb forces between any of the beads. Beads are classified as hydrophobic (strongly partitioning into alkane solvent over water, C-type), and polar (strongly partitioning into water, P-type). The non-bonded interactions are parameterized based on partitioning free energies between different types of solvent of small model compounds, each represented by a single bead. The modifications of the model to make it suitable for adsorbents on a graphite surface are based on adsorption enthalpies of a range of compounds, and on preferential adsorption of long-chain alkanes over short-chain alkanes from a mixed solution. The parameterization involved defining a new bead type for the underlying graphite surface (SG4), optimizing the non-bonded interaction parameters of the beads with the surface and fine-tuning adsorbent-adsorbent interactions and the bonded parameters describing molecular geometry (bonds, angles, dihedrals), with the aim to semi-quantitatively reproduce packing on the surface. The present model uses the alkane (C1S and C1E) and alkanol/amide beads (P1) for the adsorbents and alkane (C1) and aromatic (SC4) beads for the solvent phenyloctane, as described by Gobbo et al.<sup>142</sup> and detailed in the Supporting Information (SI). Validation of the adsorbents is reported in Gobbo et al.<sup>142</sup>; validation of phenyloctane is described in the SI.

**Simulation details.** Simulations were done using the Gromacs package<sup>91,165</sup>, using several installations of versions 4 and 5 on different hardware platforms. The equations of motion were solved numerically using a time-step of 30 fs. A triclinic unit cell (simulation box) was used, with the lateral dimensions (x, y) fixed, so that the basal plane is commensurate with a hexagonal lattice of the graphite beads. The perpendicular dimension (z) was kept orthogonal to the basal plane, but not fixed in length. Periodic boundary conditions were applied in all directions. Pressure coupling to a pressure of 1.0 bar was achieved through the Berendsen barostat<sup>166</sup> with a coupling time of 3.0 ps and a compressibility in the z-direction of  $3.0 \cdot 10^{-5} \text{ bar}^{-1}$ . Temperature was maintained by coupling to a bath through a Berendsen thermostat<sup>166</sup> with coupling constant 0.3 ps. Different adsorbents were coupled to separate temperature coupling baths. The graphite beads were always frozen. Thus, there is no dynamics in the

beads making up the surface. Simulations were run at 298 K and at 258 K. The neighbor list update was done every 10 steps.

**Analysis and visualization.** Results were analyzed by MDAnalysis<sup>167,168</sup>. Visualizations has been done by VMD<sup>169</sup>.

**Systems.** An initial set of simulations was performed with systems consisting of 18,000 graphite beads, 388 adsorbent molecules, and 10,476 phenyloctane molecules, resulting in a bulk concentration of approximately 0.14 M. The graphite surface contains five layers of beads with an area of 227 nm<sup>2</sup>, and is a small free-standing flake, i.e. the surface is not connected through periodic boundary conditions; the idea is to avoid possible packing defects of the adsorbents due to imposed periodicity (see SI). The systems were prepared by first filling the simulation volume with the graphite surface already present with randomly placed adsorbent molecules, and subsequently adding randomly placed phenyloctane solvent molecules. The adsorbent molecules were 6-bead and 8-bead linear chains with zero, one, or two polar beads. Representations of the adsorbent molecules are shown in the top panels of Figure 4.1. In the Martini model these molecules can be taken as representative of tetracontane (C<sub>24</sub>H<sub>50</sub>, ALK), tetracosanol (C<sub>24</sub>H<sub>49</sub>OH, AL1), N-heptadecanobutanamide (C<sub>4</sub>H<sub>9</sub>-CO-NH-C<sub>17</sub>H<sub>35</sub>, AM2), N-tridecanooctanamide (C<sub>8</sub>H<sub>11</sub>-CO-NH-C<sub>13</sub>H<sub>27</sub>, AM3), N,N'-decanomethylenebis-pentamide (C<sub>4</sub>H<sub>9</sub>-CO-NH-(CH<sub>2</sub>)<sub>10</sub>-NH-CO-C<sub>4</sub>H<sub>9</sub>, AM25), N,N'-decanomethylenebisnonamide (C<sub>8</sub>H<sub>17</sub>-CO-NH-(CH<sub>2</sub>)<sub>10</sub>-NH-CO-C<sub>8</sub>H<sub>17</sub>, AM36), but may also be interpreted as chains that are one to two atoms longer or shorter and have the amide groups shifted up or down the chain by one or two positions. The adsorbents will be referred to by their polar substituent pattern: ALK: alkane, AL1: alcohol group at bead 1, AM2: amide at bead 2, AM3: amide at bead 3, AM25: amide groups at beads 2 and 5, AM36: amide groups at beads 3 and 6. Except AM36 which consist of 8 beads, all molecules are represented by a 6-bead chain. The number of adsorbent molecules in these simulations is such that it allows full coverage of both surfaces by the adsorbent molecules, with some excess molecules remaining in solution. The initial set of simulations was run for at least 10  $\mu$ s each. It is worth noting that coarse-graining smoothens the potential, which results in a speed-up of a process with respect to all-atomistic simulations. For water and lipid systems, the Martini model shows a speed-up of about four times compared to atomistic simulations<sup>72,87</sup> (i.e. 10  $\mu$ s coarse-grained simulation time is roughly equivalent to 40  $\mu$ s of all-atom simulation time), but the scaling may be

different for the present system, and was not investigated. All times reported are coarse-grained simulation times. Simulations for selected systems were repeated with different random starting structures to gather more statistics on the early stages of monolayer formation.

In order to test for nucleation of adsorbent domains on the surface and to study domain size and dynamics at partial surface coverage, additional simulations were performed with a smaller number of adsorbent molecules. For some simulations at 258 K, the initial structures already contain molecules adsorbed on the surface, to avoid simulating the slow process of diffusion of adsorbent molecules onto the surface.

Simulations were also performed on larger systems, multiplying the initial size of the graphite flake by a factor of 4 (2x2 in lateral directions) to study formation, organization, and dynamics of larger domains for the molecule showing the most ordered surface structures, AM25. A complete list of simulations is presented in the SI.

# Supporting information

## Methods

### Analysis

#### Measuring/Characterizing self-assembled structures and self-assembly process

To study the adsorption process in more detail several quantities were monitored as a function of time.

1. **The fraction of the surface covered by adsorbent.** For this purpose, the molecules adsorbed to one of the surfaces were selected. A molecule was considered adsorbed when both its end beads were within 5 Å of the graphite surface in the Z direction (the substrate plane is the XY plane). Counting molecules then allows calculation of the fraction of the surface covered by adsorbent with the highest value of observed adsorbed molecules as denominator.

2. **The nematic order parameter.** The nematic order parameter was calculated by considering all molecules adsorbed to one particular surface. First, the so-called director was calculated, by diagonalizing the matrix  $Q_{ab} = \frac{1}{2} \sum_j (3d_{ja}d_{jb} - 1)$ , where  $d_j$  is the vector connecting the end beads of molecule  $j$ , and  $a, b$  denote  $\{x, y, z\}$ :

$$Q = \begin{pmatrix} \frac{1}{2} \sum_j 3d_{jx}d_{jx} - 1 & \frac{1}{2} \sum_j 3d_{jx}d_{jy} & \frac{1}{2} \sum_j 3d_{jx}d_{jz} \\ \frac{1}{2} \sum_j 3d_{jy}d_{jx} & \frac{1}{2} \sum_j 3d_{jy}d_{jy} - 1 & \frac{1}{2} \sum_j 3d_{jy}d_{jz} \\ \frac{1}{2} \sum_j 3d_{jz}d_{jx} & \frac{1}{2} \sum_j 3d_{jz}d_{jy} & \frac{1}{2} \sum_j 3d_{jz}d_{jz} - 1 \end{pmatrix} \quad (S4.1)$$

The eigenvector  $\mathbf{D}$  with the largest eigenvalue is the director. It characterizes a preferred orientation of the adsorbents, irrespective of directionality, i.e. rotating any molecule over 180 degrees leaves the director unchanged. The nematic order parameter is then calculated from the dot products of the molecular vectors and the director as  $S = \langle 3(\mathbf{d}_j \cdot \mathbf{D})^2 - 1 \rangle / 2$ , where  $\langle \rangle$  denotes averaging:

$$S = \frac{1}{2N} \sum_j \left( 3(d_{jx}D_x + d_{jy}D_y + d_{jz}D_z)^2 - 1 \right) \quad (S4.2)$$

Low values of the nematic order parameter indicate little ordering or alignment, high values indicate high ordering. Note that the nematic order parameter may have a low numerical value while most of the molecules are part of a locally ordered cluster or domain. This is then due to different orientations of the ordered domains, as seen in Figure 4.4f-h in the main text, for example.

**3. Cluster analysis.** The organization on the surface is further characterized using a clustering algorithm. Clusters were built by identifying neighbors amongst the molecules on a surface. The analysis was performed separately for each surface. Molecules were considered neighbors if they are close to each other and showed similar orientation (not considering their direction) on the surface. Two molecules were considered neighbors if they were both stretched (their end-to-end distance being at least 21 Å) and (i) (a) their center-to-center distance was less than 10 Å and (b) the absolute value of the dot product between their molecular vectors (a molecular vector is the vector connecting the two end beads of the molecule) was larger than 0.98 (the angle is then smaller than 11.5 degrees), or (ii) (a) their center-to-center distance was less than 26 Å and (b) the absolute value of the dot product between one molecular vector and the vector connecting the potential neighbors was at least 0.98. In this manner both molecules lying parallel or antiparallel to each other within a lamella (first criterion) and in neighboring lamellae (second criterion) are included in the same cluster. Again, the directionality (parallel or antiparallel) was not considered by taking the absolute value of the dot product between molecular vectors. Illustrations of the criteria are shown in Figure S4.1. Molecules 1 and 2 (cyan text and dashed lines) are neighbors based on criterion (i) because their orientation is similar (the angle  $\theta_{12}$  between the dashed line is small) and the distance between their midpoints is small. Molecules 3 and 4 (yellow text and dashed lines) are neighbors based on criterion (ii) because they lie head-tail (the angle  $\varphi_{34}$  between the dashed line and drawn arrowed line) and the distance between their midpoints is not much larger than the length of a stretched molecule. Molecules 5 and 6 are not neighbors: although the distance between their midpoints is less than that of the length of a stretched molecule, meeting criterion (iia), the angle between the vector connecting their midpoints and the vector connecting the head and tail of one of the molecules is large, not meeting criterion (iib).

After assigning all neighbor pairs, one molecule is picked as the first member of the first cluster. Its neighbors are added to the cluster, and their neighbors are added to the cluster, etc., until no more neighbors are found. One of the remaining molecules is then picked as the first member of the second cluster, its neighbors are added, etc. Figure S4.1 shows the clusters that were assigned for this particular snapshot by color. The molecules with two colors are single-molecule clusters.

Performing cluster analysis on different snapshots allows a monitoring of the self-assembly process by gathering statistics on the clusters. The number of clusters is a measure of the extent of self-assembly into ordered structures on the surface. The average area of clusters is a measure of domain size on the surface. The standard deviation in the average area of the clusters is a measure for the homogeneity of the domain sizes. The area of the largest cluster is a measure of the extent to which molecules tend to form long-range ordered domains.

**4. Fraction of ordered molecules.** The order of formed structures is measured by calculating the fraction of adsorbed molecules which are ordered, for at least half of the time during the final 10 ns of simulation (i.e. longer than 5 ns). A molecule is considered ordered when it is in a lamella, that means when it has two neighbors (the center-to-center distance is less than 10 Å), and the dot product of their molecular vectors is larger than 0.98 (the angle is then smaller than 11.5 degrees; the molecules are parallel). The fraction of ordered molecules is the number of ordered molecules divided by the the total number of adsorbed molecules.

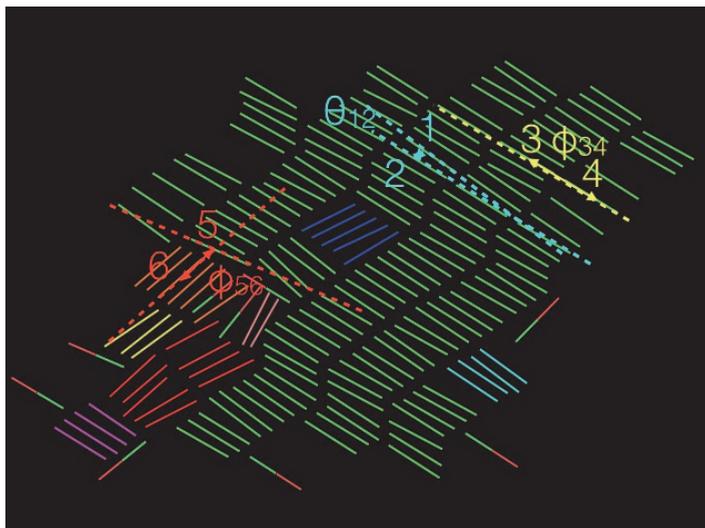


Figure S4.1. Cluster analysis. Molecules belonging to the same cluster are indicated by the same color. Three pairs of molecules illustrate the definitions of the angles  $\theta$  and  $\phi$  and the criteria for assigning neighbors for the clustering algorithm (see text for details).

5.  **$\theta$ - $\phi$  plot.** Scatter plots were made of all  $\theta$ - $\phi$  combinations between pairs of molecules on the surface. Note that all molecule pairs are considered, not just the ones that are assigned to be neighbors. A single dot is shown at coordinate  $\{\cos\theta, \cos\phi\}$ . The angles  $\theta$  and  $\phi$  are defined in the description of the clustering procedure and illustrated in Figure S4.1. The pattern informs about longer-range ordering. Well-defined stripes in the  $\theta$ - $\phi$  plot mean that molecules are oriented similarly over large distances at well-defined angles to each other. Within a lamella, the  $\theta$  and  $\phi$  angles are approximately perpendicular to each other. Between neighboring lamella, the  $\theta$  and  $\phi$  angles are close to each other. A domain with a different orientation shows up by similar  $\theta$  and  $\phi$  angles between pairs of molecules, one from each domain.

6. **Miscellaneous.** The distance between the first and last bead of an adsorbent molecule indicates how stretched the molecule is. In a lamellar arrangement, molecules must be (almost) straight to pack well. The average distance between first and last beads of molecules can serve as an indicator for domain formation on the premise that domains contain mostly stretched molecules. Surface coverage was measured as the fraction of adsorbent molecules on the surface, full coverage being 1. This metric informs about kinetics of adsorption but may also be used to study correlation between surface coverage and domain formation.

### Plotting

For clarity all plots are plotted using rolling mean (sometimes referred as running mean or moving average). Rolling mean creates series of averages of the subset of the full data. In this paper we use windows size 50 on which all data is presented. For comparison here we present a plot with and without rolling mean for the selected simulation (whose snapshots are presented in in the main text).

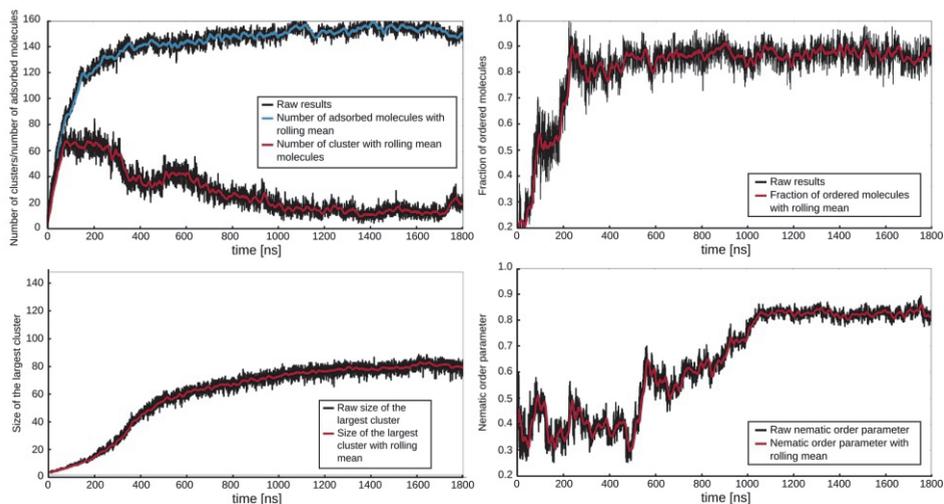


Figure S4.2. Plot in this paper are presented with rolling mean (window size 50). Figure presents comparison of raw results (black lines) and the results with rolling mean. Considerable noise can be observed in raw data.

## Solver

The functions (adsorption and KJMA+OR model) are fitted with *curve\_fit* from SciPy python package<sup>170</sup>. The method uses the Levenberg-Marquardt algorithm to minimize the sum of squares of the residuals.

## Force-field

In this work we used exactly the same force-field parameters as presented in paper by Gobbo et al.<sup>142</sup>. The summary of Lennard-Jones interactions is shown in Table S4.1, and the summary of bonded interactions are presented in Table S2.

Table S4.1. Interaction matrix for used beads in the recommended model. Roman numerals indicate the Lennard-Jones type potential interaction level of the bead-bead interaction as given in the Martini model 2 (sigma = 0.47 nm; epsilon values (kJmol<sup>-1</sup>) II 4.5; III 4.0; IV 3.5; V 3.1; VI 2.7). Prefixes denote a scaling of the standard C6 and C12 parameters (75 means scale by 0.75, 375 by 0.375). Postfixes denote using a sigma other than 0.47 nm (43 means 0.43 nm, 24 0.24 nm). The OO interaction level has epsilon = 102 kJmol<sup>-1</sup>). C1: alkane bead for short alkanes; C1S: alkane bead for long alkanes in the middle of a chain; C1E:alkane bead for terminal groups of long alkanes; SG4: graphite bead; SC4: benzene/phenyl bead; Polar: alcohol/amide bead.

Bead	C1	C1S	C1E	SG4	SC4	Polar
C1	IV	IV	IV	75VI43	V	VI

C1S	IV	III43	V43	75V43	V	VI
C1E	IV	V43	III43	75V43	V	VI
SG4	75V43	75V43	75V43	OO24	375V43	75V43
SC4	V	V	V	375V43	75IV43	IV
Polar	VI	VI	VI	75V43	IV	II

Table S4.2. Bonded potentials used in this study. C1X = C1S or C1E, Pol is the polar bead representing bead containing an alcohol or amide functional group. SC4 is a benzene/phenyl bead.

<i>Bond</i>	$V(b)=1/2 K_b (b- b_0)^2$		<i>Angle</i>	$V(\square)=1/2K_\square(\cos\square- \cos\square_0)^2$	
Beads	$b_0$ (nm)	$K_b$ (kJmol <sup>-1</sup> nm <sup>-2</sup> )	Beads	$\square_0$ (deg)	$K_\square$ (kJmol <sup>-1</sup> )
C1X-C1X	0.51	1250	C1X-C1X-C1X	180	30
Pol-C1X	0.51	1250	C1X-Pol-C1X	180	30
			Pol-C1X-C1X	180	30
C1-C1	0.51	1250	C1-C1-SC4	180	25
C1-SC4	0.37	2500	C1-SC4-SC4	150	50
SC4-SC4	0.27	constraint			

### List of all simulations

Lists of all simulations are presented in Table S4.3.

Table S4.3. List of all simulations performed for this work.

Type of molecule	Temperature [K]	Time of simulation [ $\mu$ s]	Number of molecules	Number of solvent molecules	Concentration [mM]	Surface area [nm <sup>2</sup> ]
ALK	298	11.5	388	10476	136.4	227
AL1		14.0				
AM2		12.0				

AM3		11.5				
AM36		10.0				
AM25		12.0				
		1.8				
		1.8				
		1.8				
		1.8				
		1.8				
		1.8				
		1.8				
		1.8				
		1.8				
		1.8				
		1.8				
		1.8				
		1.8				
	258	5				
		5				
		5				
		5				
Nucleation; Influence of number of adsorbed molecules on order of lamella:						
Type of molecule	Temperature [K]	Time of simulation [ $\mu$ s]	Number of molecules	Number of solvent molecules	Concentration [mM]	Surface area [ $\text{nm}^2$ ]
AM25	298	2.0	10	10476	3.7	227
		2.0	40	10476	14.9	
		1.8	80	10476	29.0	
		2.0	100	10476	36.8	
		2.0	120	10476	44.1	
		2.0	140	10476	51.2	
		2.0	180	10476	65.5	
		2.0	200	10476	72.6	

		1.8	230	10476	83.2
		1.8	260	10476	93.6
		1.5	300	10476	107.4
		1.5	300	10476	107.4
		1.5	300	10476	107.4
		1.5	300	10476	107.4
		1.5	300	10476	107.4
		1.5	320	10476	114.3
		2.0	330	10476	117.6
		2.0	330	10476	117.6
		1.5	340	10476	121.1
		1.5	360	10476	127.8
		1.5	420	10476	147.8
		1.5	440	10476	154.3
	258 (*)	5.0	10	10476	3.7
		5.0	40	10476	14.9
		5.0	100	10476	36.8
		5.0	140	10476	51.2
		5.0	180	10476	65.5
		5.0	200	10476	72.6
		5.0	260	10476	93.6
		2.5	300	10476	107.4

Type of molecule	Temperature [K]	Time of simulation [ $\mu$ s]	Number of molecules	Number of solvent molecules	Concentration [mM]	Surface area [ $\text{nm}^2$ ]
Reorganization on large (2x2) graphite flake:						
AM25		0.72	1072	20176	177.3	908
		0.72	1152	20176	189.3	
		0.72	1232	20176	197.7	
		0.72	1392	20176	215.2	
Different condition (increased concentration, doubled volume):						
AM25	298	1.5	776	10476	271.1	227
		1.0	776	20952	140.4	
Different shapes of graphite flake, periodic boundary conditions:						
AM25	298	0.9	388	5044	254.2	227 (free-standing squared graphite flake, see Figure S3d)
		0.9	388	5044	253.3	227 (infinite graphite flake in cubic simulation box, see Figure S3b)
		10.0	388	5044	255.6	227 (infinite graphite flake in rhombic simulation box, see Figure S3a)

(\*) To check if nucleation happens at lower temperature we took the last frame from a trajectory of simulation at 298K and used it as the initial frame for 258 K. Simulation of the entire process

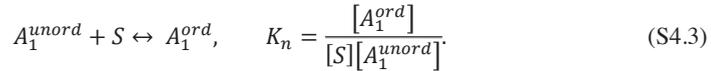
was computationally too expensive at 258K, since molecules diffuse considerably more slowly to the surface at 258 K than at 298 K.

## Nucleation-elongation models

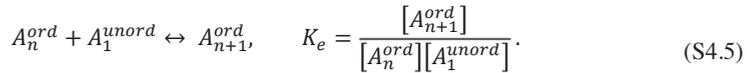
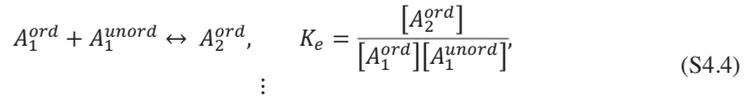
### Modified Matsuda's nucleation-elongation model

Here, we recall the derivation of the nucleation-elongation model for 2D self-assembly on an interface developed by Matsuda et al.<sup>119</sup>. The model takes inspiration from cooperative and isodesmic mechanisms of self-assembly of supramolecular polymers. It distinguishes two regimes of the process: nucleation and elongation. An unordered molecule in the solvent or on the surface,  $A_1^{unord}$ , adsorbs on the surface,  $S$ , creating a nucleus,  $A_1^{ord}$ . Then, the nuclei grow by coalescence with other molecules from solution, forming dimers, trimers, etc.

The first step, the nucleation, can be described as an equation, and is associated with an equilibrium constant,  $K_n$ :



The growth or elongation of the nucleus can be described as a series of reactions with associated equilibrium constants,  $K_e$ , which for simplicity are assumed to be the same for all species:



The amount of free graphite surface (the part of the surface on which molecules can adsorb) can be expressed in terms of a free surface concentration,  $[S]$ :

$$[S] = (1 - \theta) \frac{A_{surface}}{A_{single}} \frac{1}{N_a V} \quad (S4.6)$$

where  $\theta$  is the fraction of occupied surface,  $A_{surface}$  is the total surface (in our case of the graphite flake),  $A_{single}$  is the surface occupied by a single adsorbed molecule,  $N_a$  is Avogadro number, and  $V$  is the total volume of the solution. We simplify the notation of the functions by introducing the maximum concentration of ordered molecules on the graphite surface,  $\alpha$ :

$$\alpha = \frac{A_{surface}}{A_{single}} \frac{1}{N_a V}. \quad (S4.7)$$

Therefore, the free graphite surface concentration can be expressed as:

$$[S] = \alpha(1 - \theta). \quad (\text{S4.8})$$

At equilibrium - or steady state - we can now express the concentration of ordered clusters of size  $n$  adsorbents as:

$$\begin{aligned} [A_n^{ord}] &= K_e [A_{n-1}^{ord}] [A_1^{unord}] = (K_e [A_1^{unord}])^{n-1} [A_1^{ord}] \\ &= (K_e [A_1^{sunord}])^{n-1} K_n [S] [A_1^{unord}] \\ &= \sigma (K_e [A_1^{unord}])^n (1 - \theta) \alpha \end{aligned} \quad (\text{S4.9})$$

Where  $\sigma = (K_n K_e^{-1})$ .

Then we can calculate the total concentration of ordered molecules:

$$\begin{aligned} c_{ord} &= [A_1^{ord}] + 2[A_2^{ord}] + \dots + n[A_n^{ord}] + \dots = \sum_{m=1}^{\infty} m [A_m^{ord}] \\ &= \sum_{m=1}^{\infty} m \sigma (K_e [A_1^{unord}])^m \alpha (1 - \theta) \\ &= \sigma \alpha (1 - \theta) \sum_{m=1}^{\infty} m (K_e [A_1^{unord}])^m \\ &= \sigma \alpha (1 - \theta) \frac{K_e [A_1^{unord}]}{(1 - K_e [A_1^{unord}])^2}, \end{aligned} \quad (\text{S4.10})$$

because  $\sum_{m=1}^{\infty} m a^m = \frac{a}{1-a^2}$ , for  $a < 1$ .

Since

$$[A_1^{unord}] = c_t - c_{ord} = c_t - \alpha \theta, \quad (\text{S4.11})$$

where  $c_t$  is the bulk (i.e., total) concentration of adsorbent molecules in the system, we can write:

$$\theta = \sigma (1 - \theta) \frac{K_e (c_t - \alpha \theta)}{(1 - K_e (c_t - \alpha \theta))^2} \quad (\text{S4.12})$$

This equation can be expanded to polynomial form, giving a cubic equation relating  $\theta$  and  $c_t$ , and therefore, it can be solved analytically. We used the analytical form to fit the data. The solution has a spacious and complex form, and we don't show it here.

### **Intuitive example: fully adsorption controlled process**

To better understand the Matsuda function, let's can consider special case, for a system which is completely adsorption controlled, that means  $\sigma = 0$ . The above equation then can be transformed to the form:

$$(1 - K_e(c_t - \alpha\theta))^2\theta = \sigma(1 - \theta)K_e(c_t - \alpha\theta) = 0 \quad (\text{S4.13})$$

Therefore, for  $\theta \neq 0$ ,  $\theta \neq 1$  and  $(c_t - \alpha\theta) \neq 0$ :

$$1 - K_e(c_t - \alpha\theta) = 0 \quad (\text{S4.14})$$

And we can express the surface coverage of ordered molecules as a linear function in the total concentration of adsorbents:

$$\theta(c_t) = \frac{1}{\alpha}c_t - \frac{1}{\alpha K_e} \quad (\text{S4.15})$$

Therefore, surface coverage for the fully adsorption controlled process is equal to zero for  $c_t \leq \frac{1}{K_e}$ ; then it has the form of a linear function with a slope  $\frac{1}{\alpha}$ , for  $\frac{1}{K_e} < c_t < \frac{1}{K_e} + \alpha$ , and finally for  $c_t > \frac{1}{K_e} + \alpha$  is equal to 1. The boundaries of the function, come from the assumption that  $0 < \theta < 1$ .

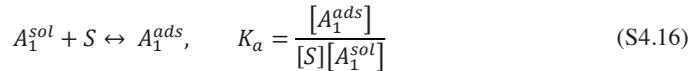
### Other models

Although, the elongation-nucleation model gives a good fit to our data (Figure S4.3 and Figure 4.3c in the main text), it does not include in its derivation a species of molecules that are adsorbed on the surface but that are not part of an ordered structure. In our simulations, we observe that during the initial stages of self-assembly molecules adsorb on the surface and create unordered clusters. When the concentration is large enough the clusters start to create ordered structures. However, in all simulation, even in the final stages, part of the surface is covered by these unordered molecules (see Figure S4.3, green points). We wanted to develop a model that explicitly includes this species.

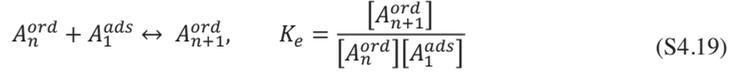
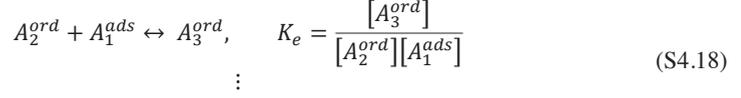
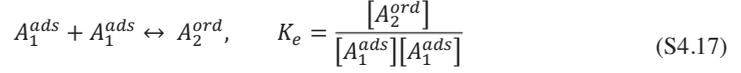
Here we show two attempts of development of such a model. However, the model of elongation-nucleation developed by Matsuda gives a better fit (see Figure S4.3).

### Model 1: Rearrangement on the surface model

In this model, in the first stage molecules adsorb on a surface,  $A_1^{ads}$ , but these do not create nuclei:



The adsorbed molecules diffuse on the surface, eventually merging into bigger clusters that are considered to contain ordered molecules:



The concentration of a cluster of size  $n$  can be expressed by:

$$\begin{aligned} [A_n^{ord}] &= K_e [A_{n-1}^{ord}] [A_1^{ads}] = (K_e [A_1^{ads}])^{n-1} [A_1^{ads}] \\ &= (K_e K_a [S] [A_1^{sol}])^{n-1} K_a [S] [A_1^{sol}] = \\ &= \sigma_2 \left( K [A_1^{sol}] \alpha (1 - \theta) \right)^n \end{aligned} \quad (\text{S4.20})$$

Where  $K = K_e K_a$  and  $\sigma_2 = K_e^{-1}$

Then the total concentration of ordered molecules is:

$$\begin{aligned} c_{ads} &= \sum_{m=1}^{\infty} m [A_m^{ord}] = \sigma_2 \sum_{m=1}^{\infty} m \left( K [A_1^{sol}] \alpha (1 - \theta) \right)^m \\ &= \sigma_2 \frac{K [A_1^{sol}] \alpha (1 - \theta)}{(1 - K [A_1^{sol}] \alpha (1 - \theta))^2} \end{aligned} \quad (\text{S4.21})$$

$$\theta = \sigma_2 \frac{K (c_t - \alpha \theta) (1 - \theta)}{(1 - K (c_t - \alpha \theta) \alpha (1 - \theta))^2} \quad (\text{S4.22})$$

The resulting fit of our data to this model is presented in Figure S4.3 as model 1. The solution of this equation is unstable numerically, and strongly depends on the choice of the starting point. Therefore, we fitted it by scanning the variables and searching for the best solution as the one with the lowest mean squared error between data points and fit.

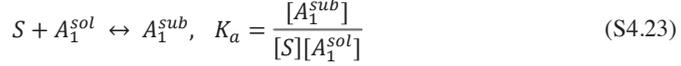
## Model 2: Rearrangement on the surface with gain of free surface

Model 1 described above takes into account the rearrangement of molecules on the surface and the existence of an unordered phase, but it does depend only weakly on the available free graphite surface. In our simulations, we observed that at higher concentrations, when there is little or no

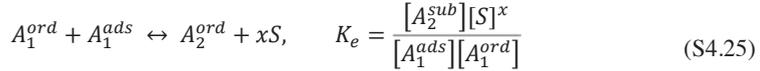
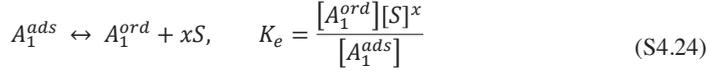
free space on the graphite surface, the equilibrium is strongly shifted toward ordered structures (see green points on Figure S4.3 which indicate the fraction of unordered molecules).

In the second model we attempt to incorporate the gain of free surface area upon ordering of the molecules (ordered molecules are better packed than unordered ones).

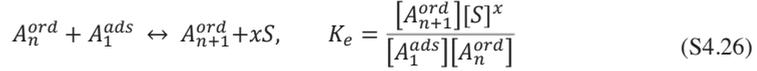
The first step is the same as in model 1:



Then, when molecules order, they create a better packed structure, resulting in a small gain of free graphite space:



⋮



Where  $x$  is the fraction of surface which is released upon packing. Manipulation similar to the one shown for the nucleation-elongation model and for model 1 at this stage, in this case leads to:

$$\begin{aligned} [A_n^{ord}] &= K_e [A_{n-1}^{ord}] [A_1^{ord}] [S]^{-x} = (K_e [A_1^{ord}] [S]^{-x})^n [A_1^{ord}] \\ &= (K_e [A_1^{ord}] [S]^{-x})^n = (K_e K_a [A_1^{sol}] [S] [S]^{-x})^n \\ &= \sigma (K [A_1^{sol}] [S]^{1-x})^n \end{aligned} \quad (S4.27)$$

$$\begin{aligned} c_{ads} &= \sum_{m=1}^{\infty} m (K [A_1^{sol}] [S]^{1-x})^m = \sigma \sum_{m=1}^{\infty} m (K [A_1^{sol}] (\alpha(1-\theta))^{1-x})^m \\ &= \sigma \frac{K [A_1^{sol}] (\alpha(1-\theta))^{1-x}}{(1 - K [A_1^{sol}] (\alpha(1-\theta))^{1-x})^2} \end{aligned} \quad (S4.28)$$

$$\alpha\theta = \sigma \frac{(c_t - \alpha\theta) K (\alpha(1-\theta))^{1-x}}{(1 - K (c_t - \alpha\theta) (\alpha(1-\theta))^{1-x})^2} \quad (S4.29)$$

The resulting fit of model 2 to our data is also presented in Figure S4.3. The solution of this equation is unstable numerically, and strongly depends on a choice of the starting point. Therefore, we fitted it by scanning the variables and searching for the best solution as the one with the lowest mean squared error between data points and fit.

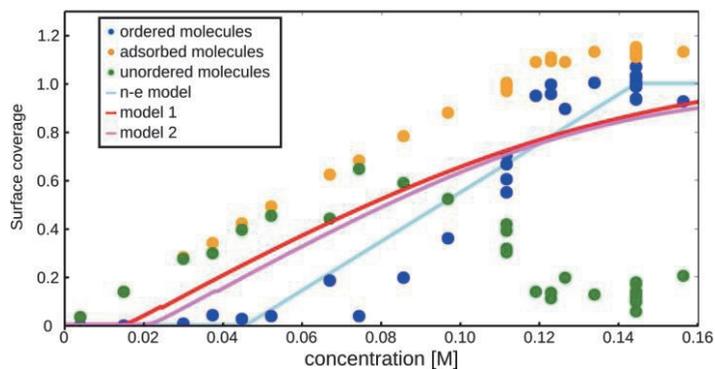


Figure S4.3. Surface coverage as a function of concentration for assemblies in equilibrium. Surface coverage is normalised to ordered molecules from the nucleation-elongation model (the y-axis is then the same as in Figure 4.3 in the main text). Blue dots represent ordered molecules, orange all adsorbed molecules (not necessarily ordered), and the green dots their difference, i.e. unordered but adsorbed molecules. It can be seen that from one point (around 0.09 M) the coverage by unordered molecules decreases. This is the result of limited available space on the graphite surface, and better packing allows more molecules to adsorb. The nucleation-elongation model with rearrangement on the surface is presented as model 1 (red line). The nucleation-elongation model with rearrangement and gain of the free surface upon ordering is shown as model 2 (purple line). None of these models reproduces trend of the ordered molecules as well as the nucleation-elongation model developed by Matsuda (light blue line).

## Discussion

The resulting trends of both developed models (model 1 and model 2) are similar to each other. This indicates that the gain of the surface upon ordering, if it is part of the mechanism, has little influence on the final trend. The inferior fit of these models compared to the Matsuda model indicates that the unordered part of the graphite flake, at least in equilibrium, might not be an important part of the model. This could indicate that the presence of edges of graphite flake (where molecules can easily desorb and are more mobile) is the reason for the disordering of molecules.

## Additional results

### Underlying graphite structure

The simulations presented in this work were done for self-assembling systems on a finite graphite flakes, in contrast to the model published in the original paper by Gobbo et al.<sup>142</sup> that connects the surface beads across periodic boundary conditions. For systems in a rhombic simulation box with an infinite graphite layer spanning through periodic boundary condition (Figure S4.4a), we were not able to reproduce correct alignment (for example<sup>22</sup>) of molecules to the graphite structure, whose hexagonal packing directions are indicated in the bottom right of the panels in Figure S4.4. When the simulation box was changed to cubic we observed a large number of kinks and defects (Figure S4.4b). This is due to minimization of surface tension between two different domains.

Although kinks are observed experimentally, in this case, the number of kinks is believed to be too large. To avoid the artifacts due to periodic boundary conditions, we use a finite graphite flake. For this system it was possible to reproduce parallel alignment of the main axis of molecules on the basal plane of graphite and long stable lamellae are observed (sometimes with a small number of kinks). Therefore if not stated differently, in all simulations mentioned here we are using the rhombic flake (Figure S4.4c). We also tested self-assembly on a squared free standing type of flakes (Figure S4.4d). In this case, we also reproduced the angle between molecules and basal vectors of graphite. This proves that the orientation of molecules is not driven by edges of the flake. We believe that this type of flake better represents reality as the lateral dimensions of all graphene layers are finite.

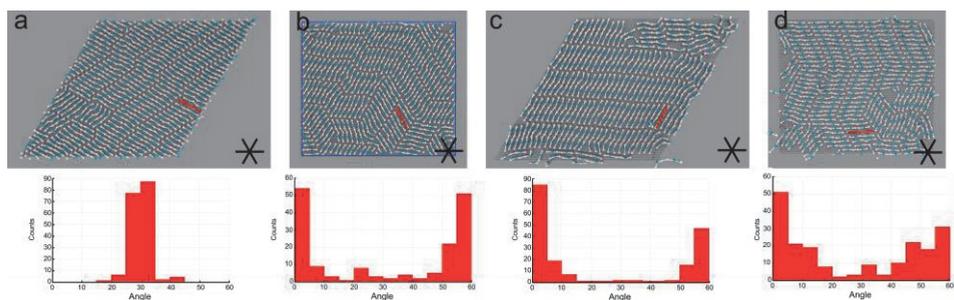


Figure S4.4. Snapshots of final structures (top row) and distribution of the angle between the molecular vectors (examples as red bars in the top row) and the main symmetry axes of graphite (black) for different shapes of graphite layer (bottom row). (a) infinite graphite flake in a rhombic box (b) infinite graphite layer in a cubic box (c) free standing rhombic flake (d) free standing rectangular flake. The basal vectors of the underlying graphite surface are indicated in black in the bottom-right corner of each panel.

### Parameterization and validation of phenyloctane

The solvent phenyloctane was mapped and described according to the standard Martini model version 2<sup>87</sup>. The mapping can be seen in Figure S4.5 in which the phenyl ring is described by three SC4 beads and its alkyl fragment by two highly apolar C1 beads. Bulk properties representative of the molecular interactions in the pure phenyloctane liquid model were used to validate the model. Although quantitative agreement is not expected due to the coarse grained nature of the model, simple comparison of the density of the liquid at 25 °C, enthalpy of vaporization and surface tension at 20 °C (see Table S4.4) shows that the standard Martini model can reproduce density and surface tension of phenyloctane reasonably well, but that the enthalpy of vaporization is too large. The latter is a common observation for the Martini model. Remember, however, that the Martini model is primarily parameterized to reproduce partitioning, i.e. free energy rather than enthalpy data.

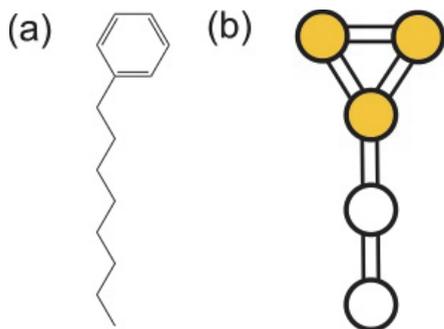


Figure S4.5. (a) Phenyloctane (b) and its representation in Martini model. Yellow circles denote small alkyl beads (SC4); white circles denote alkyl-type beads (C1).

Table S4.4. Comparison of experimental results with simulated values of density, enthalpy of vaporization and surface tension of pure phenyloctane.

Physical data	Empirical Value	Model value
Density of the liquid (25 °C)	0.854 g cm <sup>-3</sup> [171]	0.786 g cm <sup>-3</sup>
Enthalpy of vaporization	63 kJ mol <sup>-1</sup> [172]	87 kJ mol <sup>-1</sup>
Surface tension (20 °C)	30 mN m <sup>-1</sup> [173]	35 mN m <sup>-1</sup>

### Final structures

Figure S4.6 shows the final structures of the different adsorbent molecules studied at 298 K.

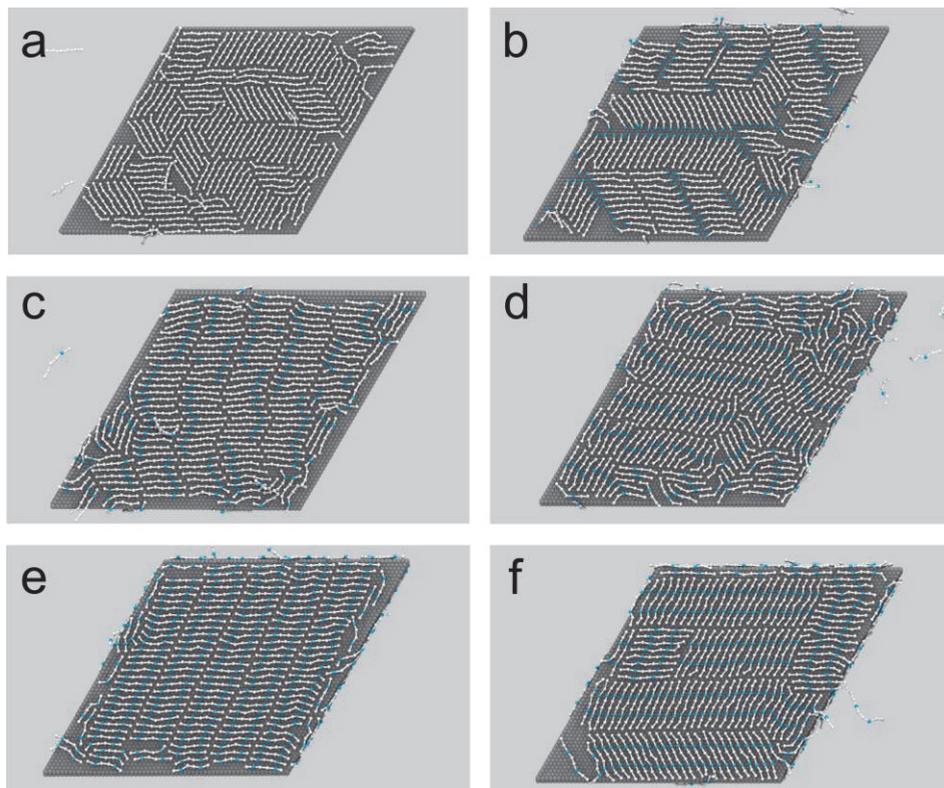


Figure S4.6. Final assemblies on graphite surface of molecules studied in this work: (a) ALK, (b) AL1, (c) AM2, (d) AM3, (e) AM25, and (f) AM36. This figure is an extension of Figure 4.1 in the main text.

Defects of the kind seen for molecules represented by coarse-grained structure AL1 (Figure S4.6b) were reported in the literature<sup>144</sup>. In Figure S4.7 we present a comparison of such a defect with experimental results.

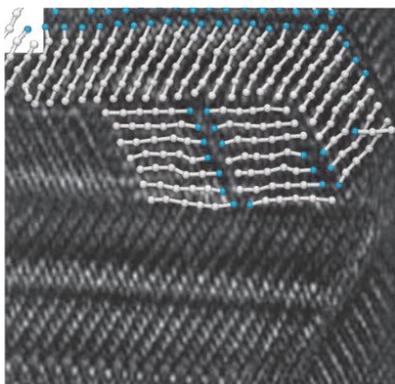


Figure S4.7. Comparison of defect obtained by simulation and experiment from literature<sup>144</sup>.

The additional final structures were analyzed as described in the Analysis section 5 of the SI in terms of the  $\theta$ - $\phi$  plot, which measures the quality of the final structure. The results are presented in Figure S4.8.

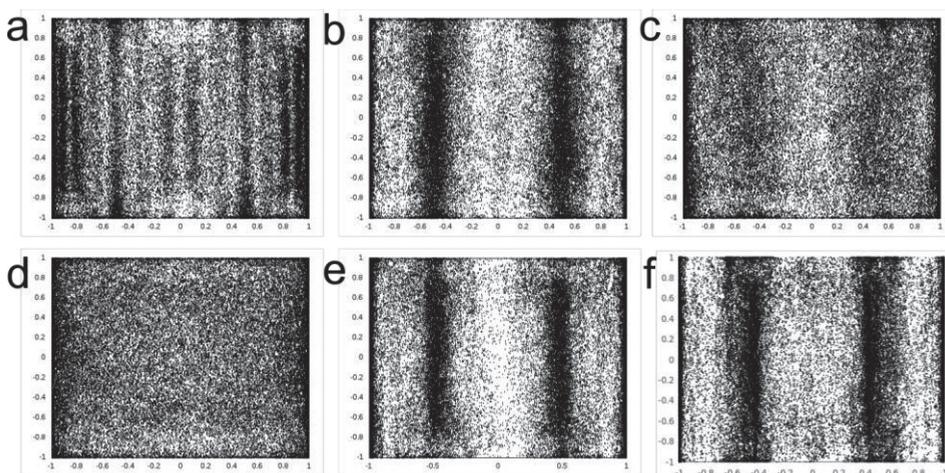


Figure S4.8.  $\theta$ - $\phi$  plot of all studied structures in this work: (a) ALK, (b) AL1, (c) AM2, (d) AM3, (e) AM25, and (f) AM36. The pattern informs about longer-range ordering. Well-defined stripes mean that molecules are oriented similarly over large distances at well-defined angles to each other. For all the plots the strongest stripes are close to  $\cos(\theta)=-1$  and  $\cos(\theta)=1$ , which represents a parallel arrangement of molecules in structure. For AL1, AM25 and AM36 two strong stripes around  $\cos(\theta)=-0.5$  and  $\cos(\theta)=0.5$ , indicate many molecules are arranged by  $60^\circ$  in respect to each other, which is a result of the presence of domains rotated by  $60^\circ$  with respect to each other. Many stripes for ALK show that the final structure is well defined, however, it has many different arrangements of molecules.

### Comparison of experimental and simulated assembly of ALK

Simulated assemblies of functionalized alkanes (i.e. AL1, AM2, AM3 and AM25) reproduce experimental results by creating assemblies where lamellae and molecules are oriented at an angle of 60 degrees with respect to each other. However, this is not the case for alkane (ALK) assembly, where experimentally lamellae and molecules are oriented by 90 degrees with respect to each other. This is a result of coarse-graining of the molecules and of the graphite surface (see Figure S4.9). If the adsorbents were arranged such that the lamella's central axis is at 90 degrees with respect to the long axis of the adsorbents, the beads would be at a distance  $\sim 0.47$  nm (see middle scheme in Figure S4.9), which is the same distance as the sigma parameter of LJ interaction of beads (see Force-field section of this SI). Therefore, at this distance repulsion and attraction are in the balance. However, the energy minimalisation would result in a shift of the molecules with respect to each other, which would result in attractive interaction between beads (see bottom scheme in Figure S4.9): their distance is now closer to the minimum of the potential; moreover, most beads interact in this manner with two other beads from the same neighbor. In our simulations this shift is always favorable. It also plays a role in the observed difference in ordering of the AM2 and AM3 beads. The shift as shown in bottom panel of Figure S4.9 is referred to in the main text as offset by half a bead. Parallel arrangement as shown in the middle scheme of Figure S4.9 probably requires tuning the interaction both between the adsorbent beads and between the adsorbent beads and the graphite beads, but we did not attempt to do this.

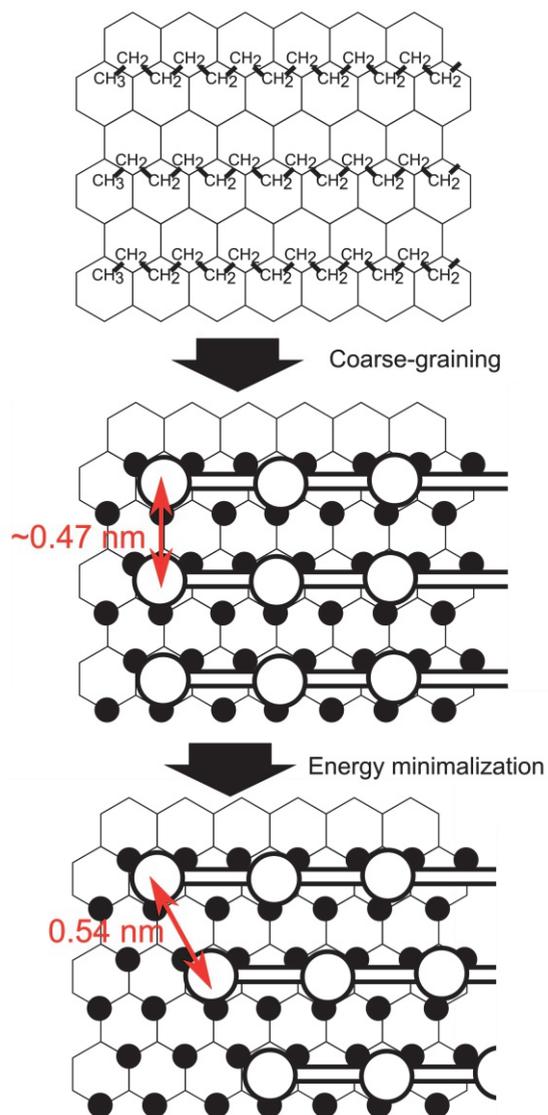


Figure S4.9. The figure presents the origin of the mismatch between experimental and simulated assembly. As a result of coarse-graining, beads in the assembly where lamellae and molecules are arranged by 90 degrees with respect to each other are not in an optimal configuration. Therefore while simulating (or energy minimalizing) the assembly rearranges into an assembly where lamellae and molecules create 60 degrees angle with respect to each other.

## Nematic order parameter

The nematic order parameter can be used as a measure of the order of the assembly as a whole. Note that although the fraction of ordered molecules can be high, the number of clusters can still be high: molecules can be well ordered in many differently oriented domains. Therefore, the fraction of ordered molecules cannot be used as a measure of the overall orientational ordering of the entire assembly. Overall long-range ordering can be measured by the nematic order parameter, which is a measure of the orientation of all molecules in one direction. (For calculation, see Analysis section 2 in the SI.) It gives additional information about the mechanism.

We observed that at higher temperature, often, the order of the assembly is gained over short time intervals, which can be observed as a relatively large jump of the nematic order parameter, occurring at different times in individual runs (Figure S4.10a, exemplified by panels 2f and 2g of Figure 4.2 in the main text). On average, the nematic order parameter follows a sigmoidal shape as a function of time, consistent with a stochastic process underlying the formation of long-ranged structure (red line in Figure S4.10a).

The nematic order parameter at lower temperature shown in Figure S4.10b does not have one large rapid transition for any individual simulation, again in contrast to the observation at the higher temperature. At the lower temperature, there is no transition from an unorganized liquid-like phase to a solid phase, and the increase in the average nematic order parameter is slow and steady as opposed to the sigmoidal shape at the higher temperature. The entire structure consists of many, small, ordered domains at different orientations, which is reflected in a low value of the nematic order parameter. Although not pursued here, it may be expected that further adsorption of molecules to reach saturation of the surface leads to more domains than observed at high temperature, separated by multiple domain boundaries.

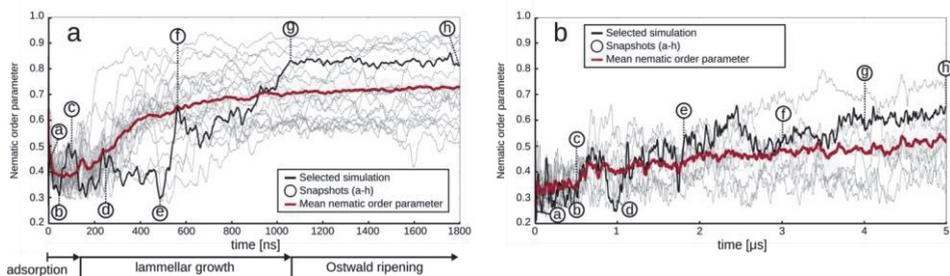


Figure S4.10. Nematic order parameter of self-assembly of AM25. Gray lines indicate results for (a) ten independent runs at 298 K and (b) four independent runs at 258 K. The thick black lines show the results for the representative trajectories discussed in the main text (black dots correspond to selected snapshots shown in Figures 2 and 4 in main text for higher and lower temperatures, respectively). The average nematic order parameter is shown as a thick red line. The nematic order parameter is a measure of a global ordering of molecules on a surface. (a) At high temperature order is gained over short time intervals, which can be observed as a relatively large jump of the nematic order parameter, occurring at different times in individual runs. On average, the nematic order parameter follows a sigmoidal shape as a function of time, consistent with a stochastic process underlying the formation of long-ranged structure. (b) At lower

temperature the nematic order parameter does not have one large rapid transition for any individual simulation, and on average increases slowly.

#### Trajectory for ALK, AL1, AM2, AM3, AM36.

Comparison of snapshots of representative trajectories for different molecules are shown in Figure S4.11.

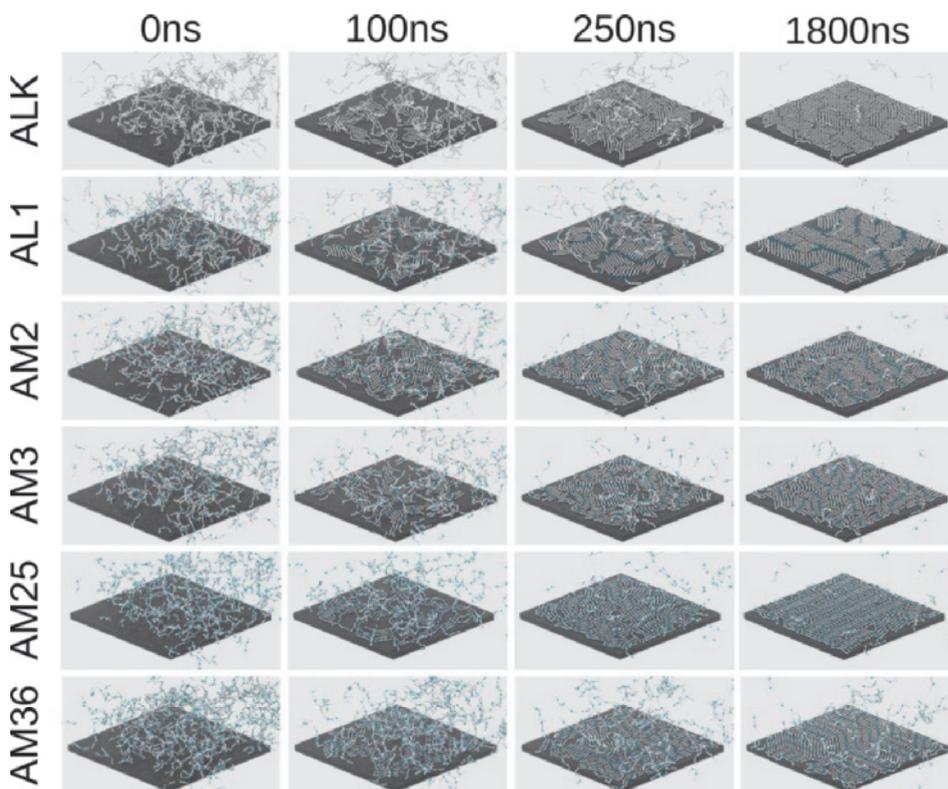


Figure S4.11. Snapshots of trajectory of formation of monolayers for studied molecules at 298 K.

The parameters describing self-assembly process: number of adsorbed molecules, number of clusters and nematic order parameter for all molecules are presented in Figure S4.12. Figure S4.12a can be compared to Figure 4.2i in the main text; Figure S4.12b to Figure S4.10a.

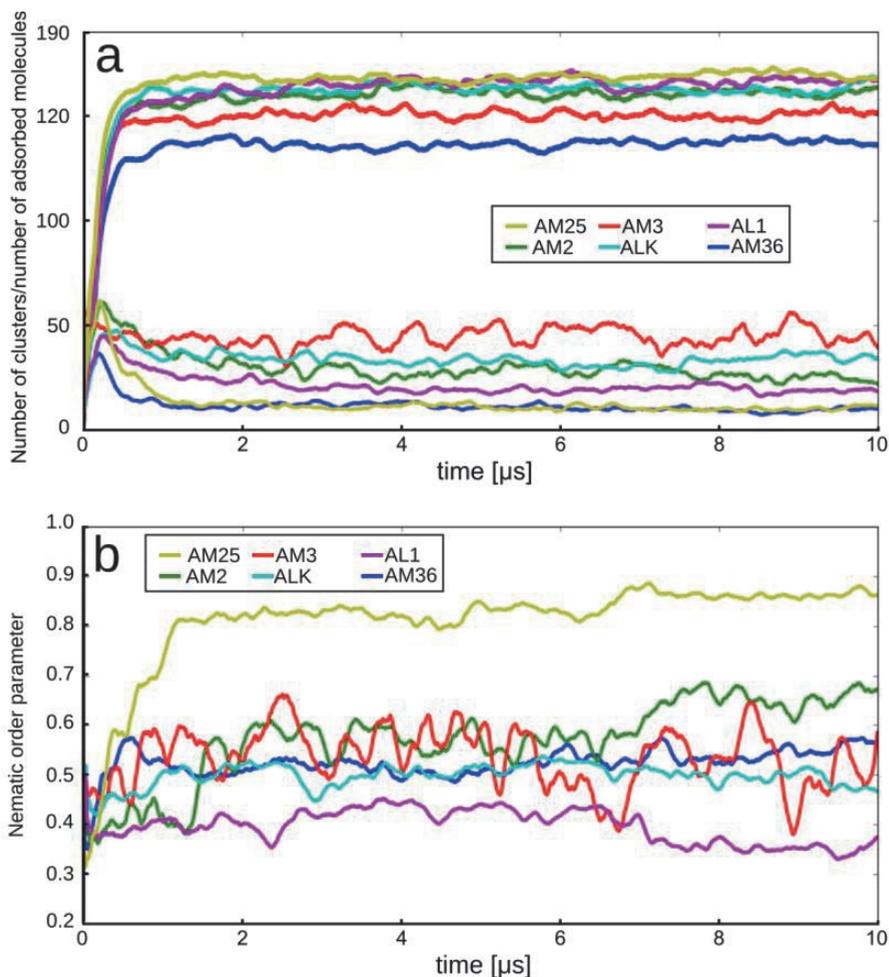


Figure S4.12. Parameters describing self-assembly of different molecules on graphite surface at 298 K.

### Markov State Model

The apparent absence of clear nucleation events was confirmed by investigating two Markov State Models (MSMs), one using the global metrics used thus far to describe the self-assembly process and one using the molecule-based metric that determines which molecules can be classified as neighbors in an ordered cluster. The latter has been advocated as superior in self-assembling systems that form ordered structures. Both MSMs indicated similar implied time scales for the slowest processes, suggesting analysis in terms of three hidden states. The three states can be interpreted as one consisting of largely unordered molecules (free in solution or adsorbed on the surface), and two interconverting states that differ in a subtle manner in their metrics, probably indicating more extensive and less extensive ordered domains. The global

metrics MSM allows visualization of a free energy landscape, which displays only a single minimum, the final assembled, ordered state. No metastable states are detected in this analysis.

Here, we present results of MSM analysis which has been done using pyEMMA<sup>102</sup> for self-assembling in high temperature. For low temperature there is insufficient sampling to perform this analysis.

### Markov State model with global metrics

We analyze trajectories using the metrics presented in the manuscript (i.e., the number of adsorbed molecules, number of clusters and fraction of ordered molecules). The used metrics allow studying the system on a three-dimensional energy landscape. Its projections on two-dimensional spaces are presented in Figure S4.13. From the pictures, it can be seen that the energy landscape is shaped as a slope leading towards the structure with a large number of adsorbed molecules and significant fraction of ordered molecules (i.e., towards a stable self-assembled structure - from visual inspection this is a lamellar arrangement). Except for the stable state, it appears that the energy landscape does not have any local minima, which would reflect metastable states.

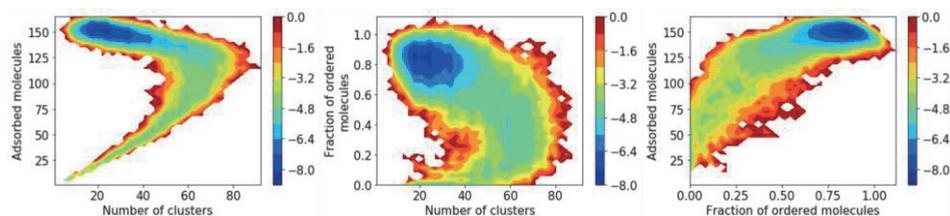


Figure S4.13. Two-dimensional projections of the energy landscape from MSM using the metrics number of adsorbed molecules, number of clusters, and fraction of ordered molecules. The color scale reflects the free energy proportional to the negative logarithm of the probability.

After discretization of the space (by K-means cluster analysis) and calculation of implied timescales, we have chosen a lag time 100 frames, which gave the best results for Chapman-Kolmogorov test (see Figure S4.14). However, there is still a small difference between predicted probabilities and the one from the trajectories, which unfortunately means, that the exact values of the result can be interpreted only semi-quantitatively.

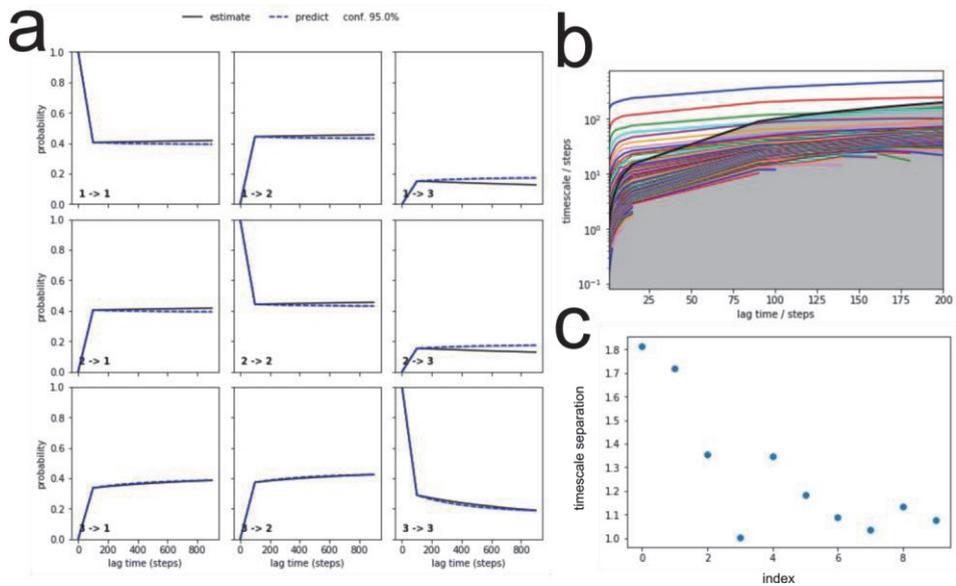


Figure S4.14. (a) Chapman-Kolmogorov test. (b) Implied timescale. (c) Separation between timescales.

We have constructed the Hidden Markov Model with three discrete states, which is presented in Figure S4.15. From this picture, it can be seen that states below  $\sim 80$  adsorbed molecules are not included in the model (because of the lack of the ergodicity). The model shows the reorganization of two states on the surface (in Figure S4.15 denoted as 0 and 1). Unfortunately, the division between these states is somewhat arbitrary, and both correspond to states in which most of the molecules are ordered on the surface.

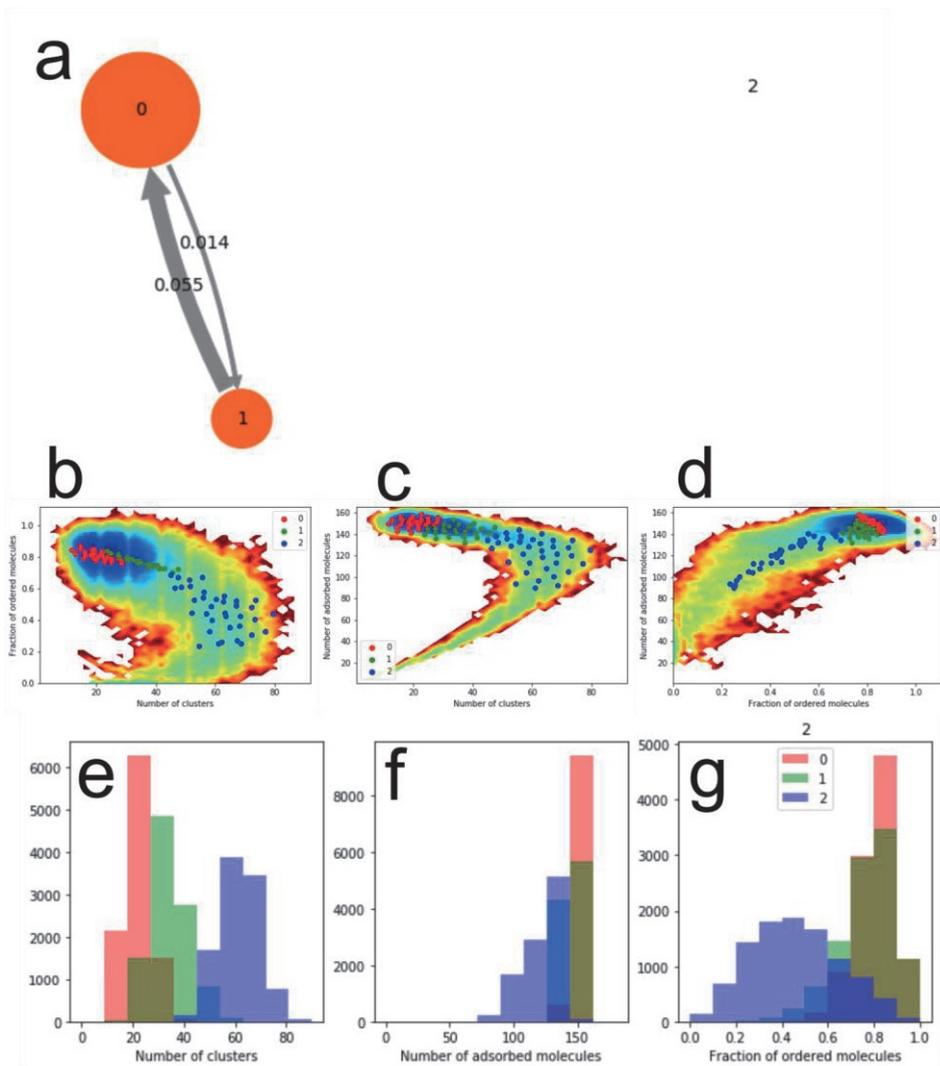


Figure S4.15. Hidden Markov state model constructed with three macrostates. (a) Transitions between three macrostates, (b-d) Two-dimensional projections of energy landscape with centers of the clusters and its membership to one of the macrostates. (e-g) Distribution of number of clusters, number of adsorbed molecules and fraction of ordered molecules for different macrostates.

### Markov State model with molecule-based metrics

However, measures used above describe the system as a whole and does not make a distinction between molecules and their particular state. We could think about a measure which would describe the system by describing the state of every single molecule. Such a system would then be described as a vector  $v = (a_1, a_2, \dots, a_{388})$ , where  $a$  describes the state of a single molecule.

However, one of the challenges of systems with many identical molecules is that such vector would depend on the numbering of the molecules<sup>174</sup>, and two exact systems with the different numbering of molecules would be described by a different vector. A simple trick to ensure that such vector is invariant on numbering, is to sort its values from smallest to largest. Such trick ensures invariance on numbering, however, it limits elements only to scalars. Therefore, as a measure of the a single element we chose the coordination number, which describes a molecule by number of its neighbors (in this work is used to calculate the fraction of ordered molecules, see SI). In such a vector every element has values  $a_i \in (0,1,2,3,4)$  describing the number of neighbors. To understand it intuitively, let's consider an example of a system with 5 molecules: the stat with 5 unordered molecules is described by  $v_1 = (0, \dots, 0)$ , the state with two monomers and a cluster of three molecules is described by  $v_2 = (0, 0, 1, 1, 2)$ , and the state of a single strand lamella containing all 5 molecules is described by  $v_3 = (1, 1, 2, 2, 2)$ . The distances between the states can be easily calculated in Euclidean metrics:  $d(v_1, v_2) = \sqrt{6}$ ,  $d(v_1, v_3) = \sqrt{14}$ , and  $d(v_2, v_3) = \sqrt{4}$ . As it can be seen from these examples, we can compare systems simply by calculating the Euclidean distance between the vectors. The more similar systems are, the shorter the distance is, and the more different they are, the longer the distance is.

Having such a detailed measure we were able to construct a Markov State Model again. We have again clustered the states by the K-Means algorithm. The implied timescales converged better than for the previous description; it seems that for even short lag-times some processes are invariant on its choice (Figure S4.16a). We have chosen a lag-time of 80 frames, which gave the best result for the Chapman-Kolmogorov test (Figure S4.16b). We could distinguish two slow processes. Therefore we have performed coarse-graining with three states. Unfortunately, results of coarse-graining are similar as for the previous measure (Figure S4.16c) and provide little additional insight. There are three states, but exchange occurs only between two. Moreover, since a long vector describes the system, it is hard to represent states in 2-dimensional space. However, we can reduce such a vector by calculating the distribution of coordination number in the states (see histograms in Figure S4.16c). As a result, we learn that the exchange between states is between two states that both have the majority of their molecules in ordered clusters, but which differ in the degree of 'orderedness', as quantified by the difference in the distributions of coordination numbers of the molecules in ordered clusters, which can be interpreted in terms of a difference in the number of ordered domains and/or size of the largest domain.

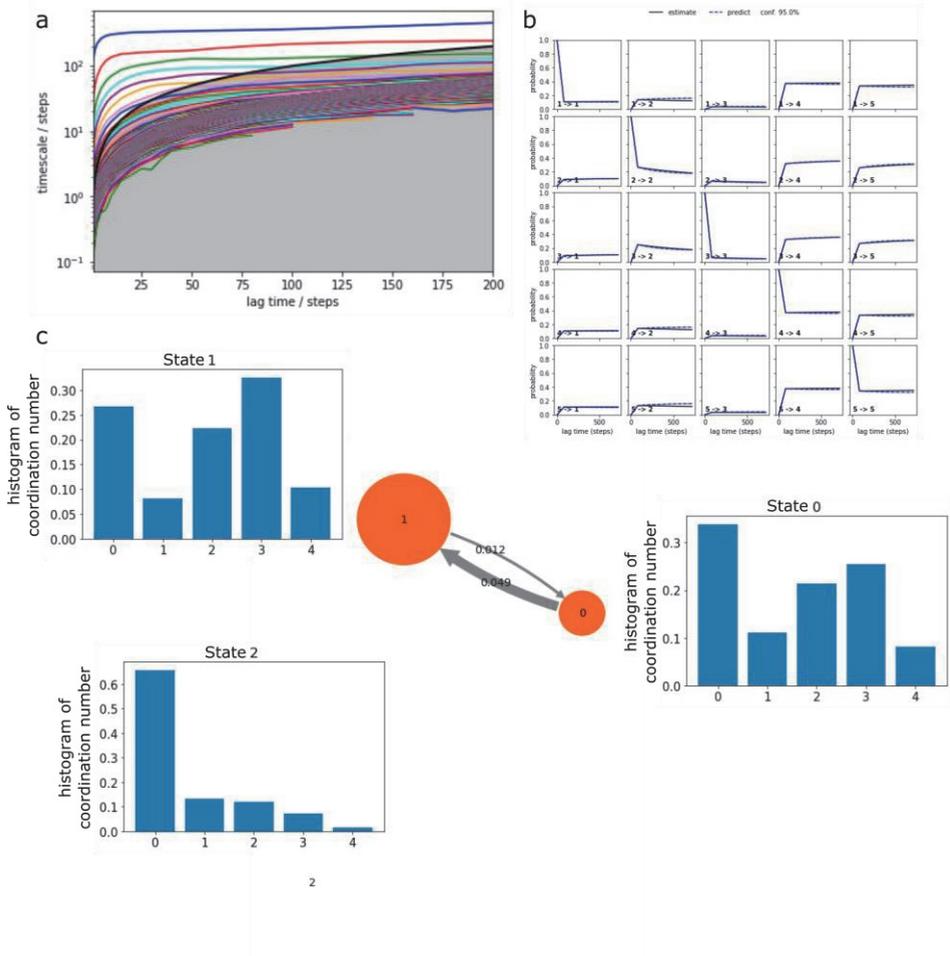


Figure S4.16. (a) Implied timescales. (b) Chapman-Kolmogorov test. (c) Hidden Markov Model and transition between states. States can be summarized by the histogram of coordination number: State 0 and 1 correspond to ordered structure (with state 0 more ordered), because most of the states have coordination number larger than two, and state 2 corresponds to less ordered structure, because most of the coordination numbers is below three.

### KJMA – fitting

The Kolmogorov–Johnson–Mehl–Avrami equation (see also main text):

$$A(t) = A_{\text{lim}}(1 - \exp[-(kt)^n]) + \left( \frac{t - \tau_{OR}}{1 + \exp[-2\omega(t - \tau_{OR})]} \right) k_{OR} \quad (\text{S4.30})$$

is fitted utilizing the Python function `curve_fit`. The results of fitting the average size of the largest cluster of AM25 molecules are presented in Table S4.5. Although the function fits the average size of the clusters well, the values themselves appear to be much larger than commonly reported experimentally. Stabel et al.<sup>151</sup> have investigated Ostwald ripening of 2-hexadecyl-anthraquinone and tetradodecyl-octathiophen on highly oriented pyrolytic graphite. They have shown that a 3500 nm<sup>2</sup> defect of 2-hexadecyl-anthraquinone is coarsened in 30 s and a 1500 nm<sup>2</sup> defect of tetradodecyl-octathiophen is coarsened in 45 s, giving the rates of Ostwald ripening,  $k_{OR}$ , of  $\sim 120 \frac{\text{nm}^2}{\text{s}}$  and  $\sim 30 \frac{\text{nm}^2}{\text{s}}$ , respectively. The fitted value of Ostwald ripening for our system is 5-6 orders of magnitude larger. Many factors could contribute to this speed up: the molecules studied here are smaller (and therefore more dynamic), therefore we would anticipate speed up of coarsening; a further likely cause is usage of coarse-grained force-field which in general speeds up dynamics<sup>72,87</sup>; finally, the small size of the graphite flake (and closeness of graphite edges) speeds up coarsening<sup>104</sup>. Thus, the numerical results shown here should be treated only semi-quantitatively.

Table S4.5. Fit parameters for AM25 self-assembly runs to KJMA+OR expression.

	Estimated value
$k$	$0.00265 \pm 0.00001 \frac{1}{\text{ns}}$
$n$	$2.185 \pm 0.016$
$A_{\text{lim}}$	$90.58 \pm 0.31 \text{ molecules} = 126.8 \pm 0.4 \text{ nm}^2$
$k_{OR}$	$0.0118 \pm 0.002 \frac{\text{molecules}}{\text{ns}} = (1.5 \pm 0.025) \cdot 10^7 \frac{\text{nm}^2}{\text{s}}$
$\tau_{OR}$	$712 \pm 26 \text{ ns}$
$\omega$	$91.0 \pm 0.0 \frac{1}{\text{ns}}$

### Double concentration and double volume

The influence of double concentration and excess of adsorbent was checked for AM25.

Doubling the initial concentration of adsorbent, but keeping the volume of the simulation box the same as in the original set of simulations, results in shortening of the first phase of the self-assembly process: molecules diffuse arrive more quickly at the surface through diffusion and adsorb on the surface. However, the quality of the final structure is decreased, and it takes longer for the molecules to rearrange into aligned lamellae. This is illustrated in Figure S4.17a-c.

An excess of adsorbent was achieved by doubling the volume of the initial supernatant solution (including the adsorbent) of the simulation box while keeping the same size of the graphite flake (see Figure S4.17d). Therefore when the AM25 molecules cover the entire surface, a relatively large excess of molecules is present in the solvent (see Figure S4.17e,f). The adsorption phase follows exactly the same trend as for normal volume. However, after this phase the excess of molecules in the solvent slows down alignment of molecules on the surface, as indicated by the metrics shown in Figure S4.17g and Figure S4.17h. This trend is also observed experimentally<sup>119</sup>.

In both cases an excess of molecules in the solvent is present. However, in neither of them, we have observed the formation of a second layer on the surface.

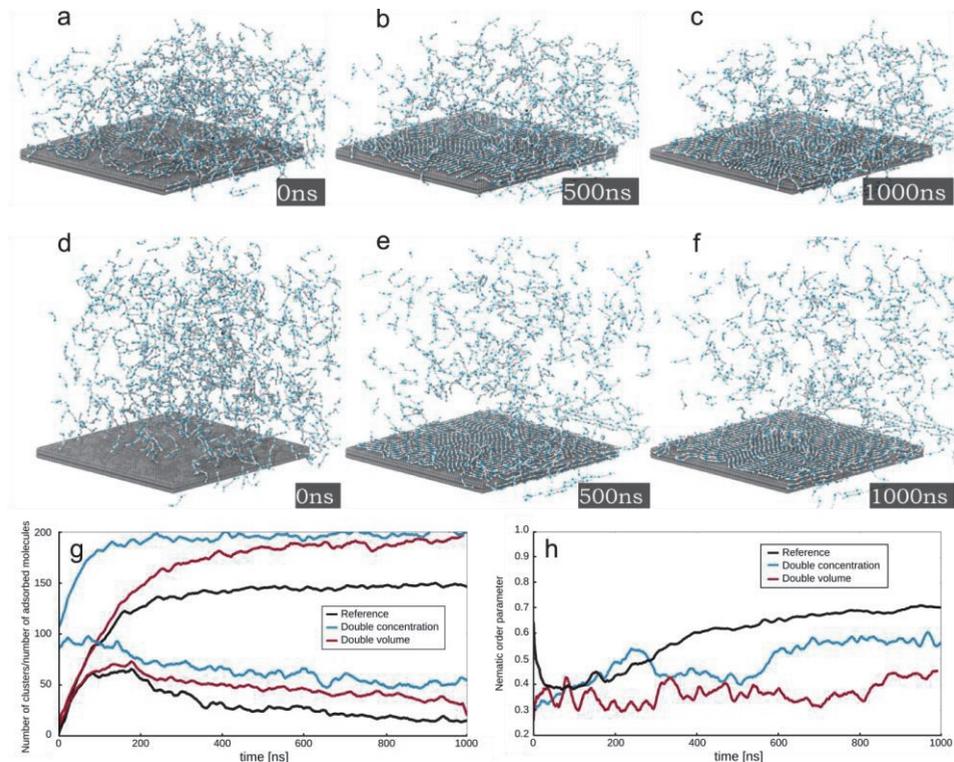


Figure S4.17. Influence of concentration on self-assembly of AM25 monolayers. (a-c) Snapshots of simulation of self-assembly with doubled initial concentration of adsorbed molecules (volume kept the same as for simulation in main text), (d-f) snapshots of simulation with doubled volume of simulation box (concentration kept the same as for simulation in main text), (g) number of clusters/adsorbed molecules as a function of time, (h) nematic order parameter (see methods). The black line is given as a reference to simulation present in the main text (see Figure 4.2).

### Self-assembly at high bulk concentration at low temperature

Figure S4.18 show the assembly after 4.3  $\mu$ s from high concentration ( $\sim 0.14$  M) at low temperature. It can be seen that assembly consist of many domains. The defects do not heal on time scale of the simulation.

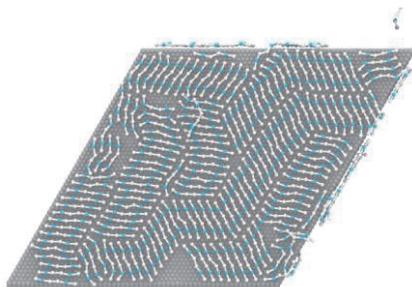


Figure S4.18. Snapshot of assembly formed from high concentration ( $\sim 0.14$  M) at low temperature (258 K).

### Stability of clusters at lower temperature

From analysis of the number of clusters and the number of adsorbed molecules, it can be seen that stable clusters are formed. However, observation of specific clusters revealed that they are stable only to certain extent and after long simulation runs they can rearrange, in particular that molecules exchange between well-formed lamellae over relatively long times. Figure S4.19 shows how the number and size of lamellae remain similar after 2 and 5 microseconds of simulation, but by coloring the members of the clusters that exist after 2 microseconds, it can be seen that clusters acquire new members and that members from different clusters are exchanged between clusters.

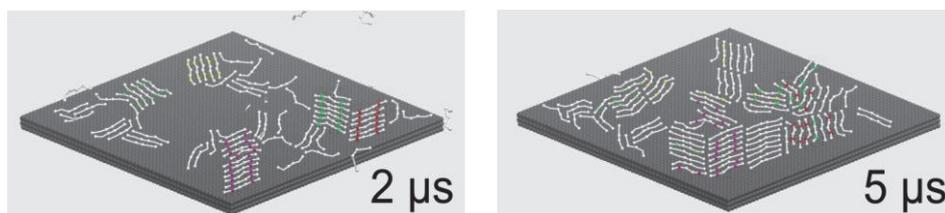


Figure S4.19. Simulation of self-assembly of 200 AM25 molecules on graphite flake at 258K after  $2 \mu\text{s}$  and  $5 \mu\text{s}$ . Although in both cases clusters can be observed, their internal structure changes over time.

### Description to supplemented video

Trajectories were visualized using VMD<sup>169</sup>. To all the presented videos we applied trajectory smoothing to make videos more clear (for every video window of trajectory smoothing is 4). The solvent (phenyloctane) is not shown for clarity.

SI\_self-assembly\_388mol\_298K.mov: video presents the simulation of self-assembly of 388 AM25 molecules on graphite surface at 298 K. The video shows a progression of self-assembly from the initial configuration at  $0 \mu\text{s}$  until the formation of the stable self-assembled monolayer at

1.8  $\mu$ s. The snapshots of this trajectory are shown in Figure 4.2 in the main text. Frames are saved every 1.2 ns.

SI\_self-assembly\_200mol\_298K.mov: video presents the simulation of self-assembly of 200 AM25 molecules on graphite surface at 298 K. The video shows a progression of self-assembly from the initial configuration at 0  $\mu$ s until complete adsorption and formation of unstable clusters at 2  $\mu$ s. Frames are saved every 1.2 ns.

SI\_self-assembly\_200mol\_258K.mov: video presents the simulation of self-assembly of 200 AM25 molecules on graphite surface at 258K. The video shows a progression of self-assembly from the initial configuration at 0  $\mu$ s until the formation of large domain at 12.9  $\mu$ s. Frames are saved every 9.0 ns. The snapshots of this trajectory are shown in Figure 4.4 in the main text.



## Chapter 5

# Mechanism of Ostwald Ripening in 2D physisorbed assemblies at molecular time- and length-scales by Molecular Dynamics simulations

Ostwald ripening can improve the long-range order of self-assembled monolayers by the growth of large domains and disassembly of smaller ones. Here, coarse-grained molecular dynamics simulations are used to study dynamics of the stable assembly and the coarsening of defects of physisorbed monolayers of long-chain functionalized alkanes. Our results show that the partial desorption from the surface of one or more adsorbent molecules is the essential process that allows other adsorbent molecules to rearrange on the surface and thereby improve alignment. We also show that the ripening process is faster at higher temperature, because the rate of partial desorption is higher.

This chapter is published as:

T.K. Piskorz, A.H. de Vries, S. De Feyter, J.H. van Esch, Mechanism of Ostwald Ripening in 2D Physisorbed Assemblies at Molecular Time and Length Scale by Molecular Dynamics Simulations, *J. Phys. Chem. C*, 2018, 122, 42, 24380-24385.

## Introduction

Long-range ordered physisorbed self-assembled monolayers (SAMs) on a substrate surface are promising for building controlled nanopatterned structures as an alternative to conventional top-down methods currently used in industry<sup>111-114</sup>. However, obtaining perfectly aligned structures is still a challenging task<sup>175</sup>. The alignment of the final structure can be improved by Ostwald ripening, in which large domains grow at the expense of small ones<sup>151,156,176</sup>. A better understanding of the process could improve control over the alignment of the assemblies.

The Ostwald ripening of physisorbed self-assembled monolayers has been studied experimentally using scanning tunneling microscopy (STM) by Stabel et al.<sup>151</sup>. They have shown that the process occurs while the adsorbents remain on the surface and is reaction-controlled, i.e. the limiting step is desorption from one domain and adsorption to another. Such a reordering on the surface was directly observed using STM<sup>125</sup>. This explanation assumes that a liquid-like zone exists between domains in which a molecule can rearrange and adsorb to a domain. However, more stable domains with tightly packed interfaces<sup>176</sup> also can undergo coarsening. For example, for an alkylated porphyrin on graphite, the growth of one domain at the expense of another was shown to occur only at a defect<sup>125</sup>. Note that the term 'coarsening' in the context of domain formation and growth refers to the appearance of large, long-range ordered domains. Experimental studies of the molecular details of the coarsening process by STM are limited by its spatial and temporal resolution, as well as by potentially strong interactions between the probe and the adsorbent<sup>123</sup>.

In this work, we provide a molecular level description of the coarsening process. Since the process happens on a short timescale, we choose to use molecular dynamics simulations to study these systems, which allow high resolution in time and space.

## Results and discussion

**Self-assembly on graphite.** We utilize a recently developed variant of the Martini model to simulate adsorption from a solution and subsequent self-assembly of molecules on a graphite surface at near atomistic resolution<sup>142</sup>. Molecules of a six-bead adsorbent, representing a long-chain alkane functionalized with amide groups (Figure 5.1a), were randomly placed in a

simulation box containing a graphite flake (Figure 5.1b), and solvent, phenyloctane, was added. In repeated simulation runs at 298 K, the adsorbent preferentially adsorbs on the surface, and forms an ordered structure after about  $2 \mu\text{s}$ , usually consisting of several large domains (Figure 5.1c). The mechanism of this process has been described elsewhere<sup>177</sup>. Simulation surrounded by a small number of molecules near the edge of the flake that have a different orientation, likely due a mismatch between the width of the lamellae and the size of the flake.

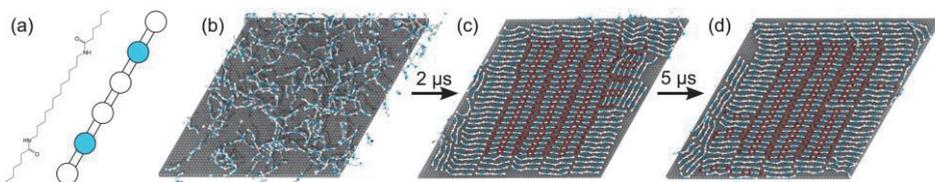


Figure 5.1. (a) Molecule studied in this work and its coarse-grained representation. (b) Snapshot of initial system: graphite flake and randomly distributed adsorbent in 1-phenyloctane (not shown for clarity), (c) snapshot of an initial self-assembled monolayer after  $2 \mu\text{s}$  of simulation, (d) snapshot of a well-ordered self-assembled monolayer after additional  $5 \mu\text{s}$  of simulation. Inner lamellae (indicated by red; 2 nm from the edge of the graphite flake) do not undergo substantial structural changes. In contrast, molecules on the edge of graphite flake stay dynamic.

**Dynamics of the self-assembled monolayer.** In order to better understand the dynamics of the defects, we first analyzed dynamics of the molecules in a single, stable, well-ordered domain, as illustrated in Figure 5.2d. We measured the rates of full and partial vertical desorption from and readsorption to the surface during a period of  $5 \mu\text{s}$ . As an example, Figure 5.1c and Figure 5.1d illustrate the initial and final structures of such a run. In partial vertical desorption, two or more beads lift off the surface, but at least one bead remains on the surface (see Methods). To give a better impression of the types of desorption molecules display, Figure 5.2a shows some examples of partially desorbed molecules pointed out in snapshots taken from the simulations (and for clarity Figure 5.2b shows its schematic representation). It can be seen that one or more beads from a molecule can be lifted from the surface at different positions within the chain. The classification of different states and processes is schematically represented on Figure 5.2c. Figure 5.2f summarizes the results (additional results and study of different types of molecules can be found in Figure S5.1 and Table S5.1). The rates are represented by frequencies how often (on average) an adsorbed molecule vertically fully desorbs, or partially desorbs. The process of full vertical desorption is observed rarely, and mostly at the

edges of the graphite flake. The inner part of the assembly is especially stable (compare Figure 5.1c and Figure 5.1d). The full desorption of these molecules is  $\sim 10$  times less frequent than desorption for molecules close to the edge of the graphite flake.

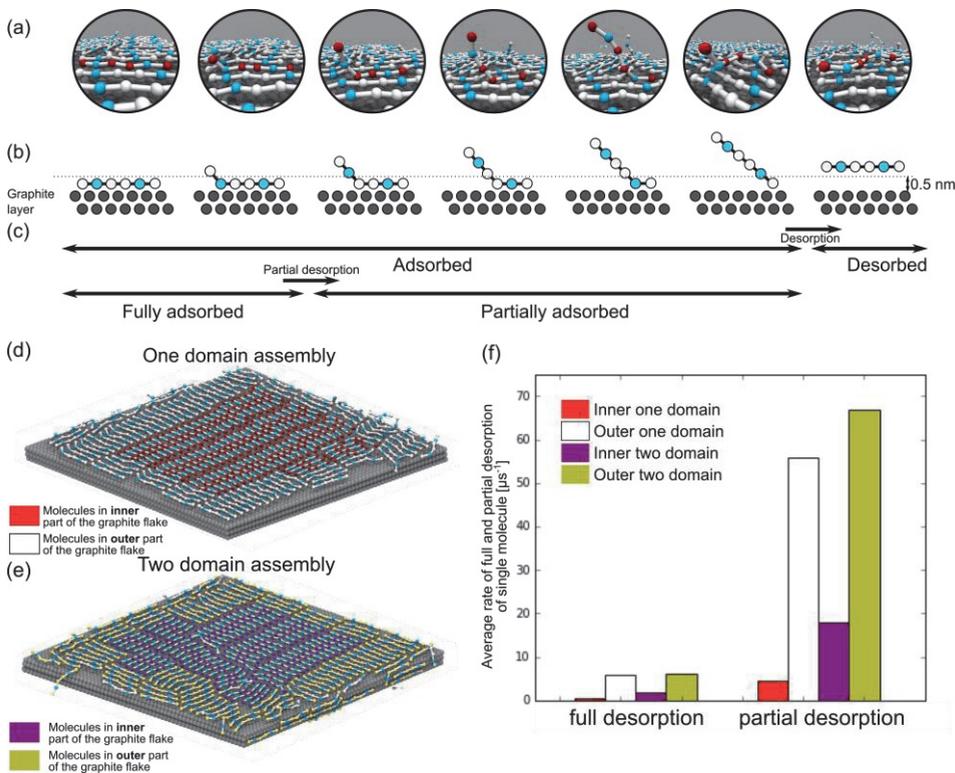


Figure 5.2. (a) Snapshots and (b) schematic representation of examples of partially desorbed molecules. (c) Classification of different states and processes analyzed in this work. Snapshots of initial (d) one domain and (e) two domain structures of assemblies on graphite. For analysis, two parts of the assembly were distinguished: inner (molecules further than 2 nm from the graphite edge; indicated by red for one domain assembly and purple for two domain assembly) and outer (molecules within 2 nm from the graphite edge; indicated by white for one domain assembly and yellow for two domain assembly). (f) Rates of full and partial desorption of molecules from a self-assembled monolayer on graphite flake during 5  $\mu\text{s}$  of simulations.

Although it appears that the inner part of the assembly is static, a closer look reveals that partial desorption is frequent (the rate is  $\sim 10$ -20 times larger than that of full desorption). The lifetime of this partially desorbed unstable state is short (in all simulations  $\sim 1$  ns, see Table S5.1), and molecules tend to reabsorb quickly on the surface.

It is noteworthy that the dynamics of the assemblies depend on temperature and the chemical nature of the adsorbents. By increasing the temperature, the monolayer becomes more dynamic, which can be observed as the increased rate of full and partial desorptions (Figure 5.2f).

The dynamics of the molecules in the single domain were contrasted with that in assemblies composed of two domains separated by a domain boundary (Figure 5.2e), which shows that the partial desorption in the latter case is more frequent (see Figure 5.2f). This finding can explain “fuzziness” of domain boundaries observed on STM images<sup>156</sup>. At the domain interface, molecules have different orientations, which hinders stabilization by alignment of the functional groups, resulting in more frequent partial desorption. Increased rate of desorption/readsorption of molecules near defects was also observed experimentally for systems of long-chain functionalized alkanes<sup>148</sup>.

**Ostwald ripening.** Ostwald ripening (in the context of physisorbed self-assembled monolayers sometimes referred to as self-healing or domain coarsening) is a process in which large domains grow at the expense of small ones<sup>151</sup>. The thermodynamic force driving this process is the decrease of interfacial energy generated by different domain boundaries. In our earlier work, most of the independent repeated self-assembly simulations led to well-ordered lamellae in an equilibrated state, but we observed several types of defect, with long lifetimes. Here, we elaborate on the domain coarsening of two types of defect: (a) two large domains which both occupy a large area of graphite and (b) a small defect surrounded by a large domain.

Firstly, we analyzed coarsening of two domains with different orientations which span on the entire graphite flake. Figure 3a shows a zoomed-in image on the domain boundary, the full graphite surface can be found in Figure S5. Molecules are colored on the basis of their initial membership of one of the two domains. As mentioned before, the structure is most dynamic on the interface of two domains. We observed that in both temperature regimes one of the domain orientations starts to dominate. Figure 5.3b-c show the changes in organization at lower temperature, Figure 5.3d-f at higher temperature for one out of six independent simulations of this process. At low temperature the coarsening process occurs in several distinct stages. Local fluctuation causes one of the lamellae of one domain to grow at the expense of the second domain. One by one, lamellae start to extend by rearrangement of the second domain, until the entire surface is covered by a single domain. The increased alignment can be

seen as small jumps in nematic order parameter plot (see Figure 5.3g, blue line; see Methods). In the higher temperature regime this transformation is more dramatic. The instability on the domain boundary progresses through the major part of one of the domains, resulting in disassembly of the domain into partially ordered clusters. These clusters restructure and align with the second domain, covering the entire flake. This can be seen in the nematic order parameter as a small decrease (disassembly of one of the domains) followed by a sudden increase which leads to ordered structure (see Figure 5.3g, red line). At both temperatures, kinks occur in lamellae structure, forming “zig-zag” patterns. Such structures have been observed amongst others in docosylether physisorbed on graphite, and were characterized as the effect of a pair of twin boundaries between lamellae. In the study by Padowitz et al.<sup>148</sup>, motion of the pair of twin boundaries was observed. A twin boundary is the result of the change in displacement or shift between two neighboring molecules within a lamella, in which the angle between the lamellar axis and the molecular axis is not  $90^\circ$ , but  $60^\circ$ . In our model, there is such a displacement between neighboring molecules, because of the coarse-grained nature of the beads. The lateral interaction between neighbors is more favorable if the neighboring molecules are displaced by half a bond length, because a bead interacts with two neighboring beads instead of one. More discussion can be found in SI.

Although both assemblies show increased order, the composition of the final structure is different. At the lower temperature, the reorganization is local, and molecules rearrange on the surface, i.e. without full desorption. In the final structure, it can be clearly seen that the assembly was formed from different domains by reorientation of the molecules in one of the domains. In contrast, the rearrangement in the higher temperature regime leads to extensive mixing of the molecules in the final structure. This indicates that significant rearrangement of the molecules was involved. By looking at the trajectories, several types of molecular rearrangements can be observed. It is difficult to quantitatively classify these events; we feel the nematic order parameter is a useful measure to monitor the increased ordering. Movies of several rearrangement events are provided in SI.

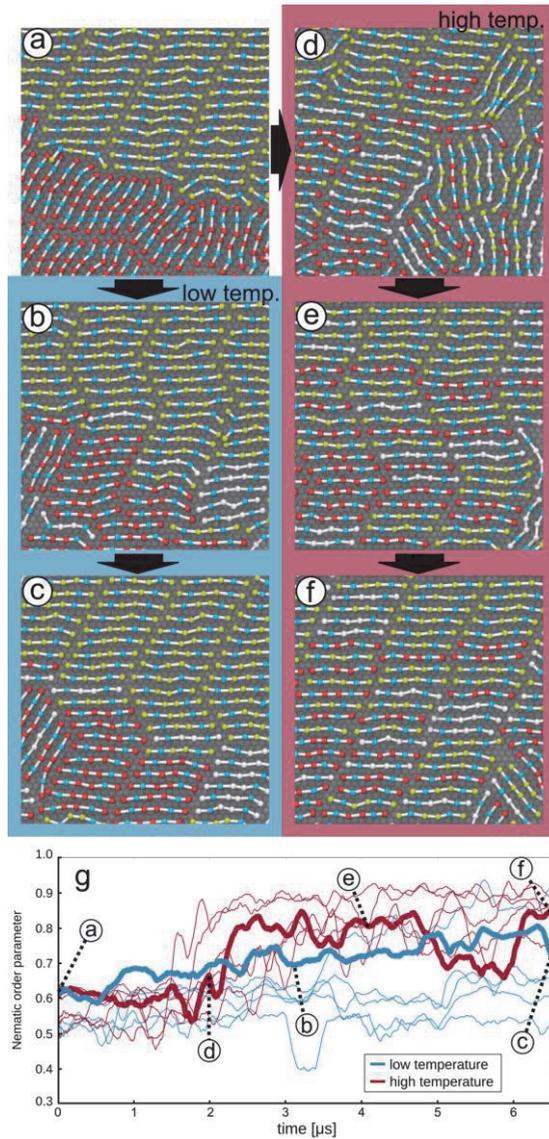


Figure 5.3. Reorganization of two domains with a different orientation. (a-f) Snapshots of two representative trajectories. The color of the molecules (red and yellow) indicates initial membership to one of the original domains, whereas white color indicates molecules which are not a part of the original domains (they originate in the supernatant or from the edges of the flake). At both temperatures the domains rearrange, but at lower temperature (b-c), they do not mix extensively, whereas at higher temperature (d-f) they do. (g) Nematic order parameter for the process at room temperature (blue) and high-temperature regime (red) for six independent simulations. The bold lines represent the nematic order parameter for the simulations from which the snapshots were taken. At higher temperature molecules align faster.

The second type of commonly observed defect is one in which in the middle of the well-ordered structure, a small defect is incorporated, as shown in Figure 5.4a (snapshots of the whole system can be found in SI). We constructed a single large domain with one such defect in order to be able to study only the rearrangement of the defect. This way we avoid all the issues present in the two-domain study such as edge effect, imperfect packing etc. Defects that are close to an edge of the graphite flake rapidly disappear due to increased mobility on the edge of the flake, see SI. We ran several instances of simulations and observed different behavior at different temperatures (Figure 5.4b-e and Figure 5.4f-i). Similar as for coarsening of two domains, the process is slower at the lower temperature (Figure 5.4k). At the lower temperature reorganization occurs on the surface of the graphite flake; molecules partially vertically desorb resulting in free space on the graphite flake for other molecules to laterally desorb from one domain and re-adsorb on the other. At higher temperature the partial vertical desorption occurs more often. This results in higher lateral mobility of the molecules on the surface, effectively speeding up the healing. It is noteworthy to mention that vacancies on the graphite flake were not observed; the area freed by (partially) desorbing molecules is occupied by beads of adsorbent or solvent molecules within the time between snapshots (which is 30 ps).

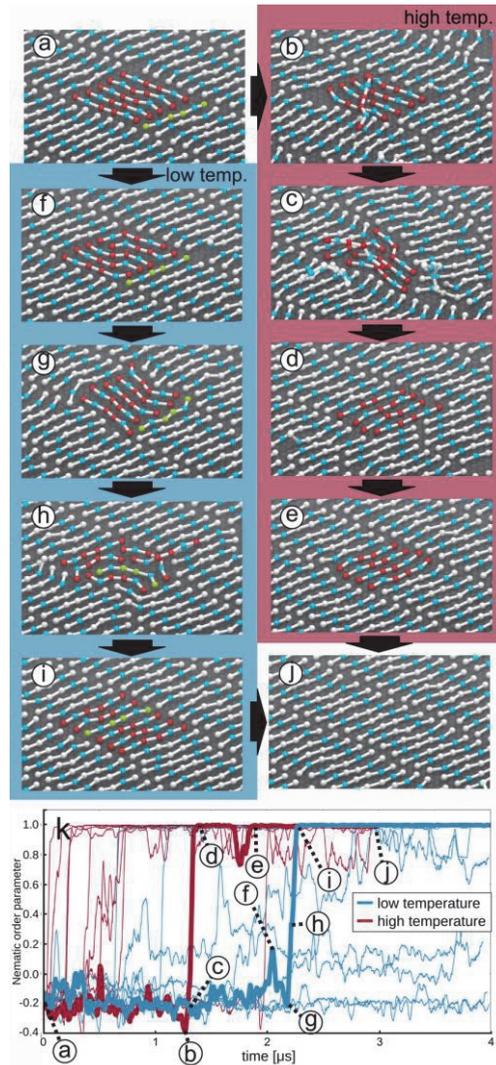


Figure 5.4. Reorganization of a small defect surrounded by a large domain. Selected beads of the molecules that constitute the defect are colored red. (a) Snapshot of starting small structure defect. (b-e) Representative snapshots of Ostwald ripening at higher temperature and (f-i) at lower temperature (the molecule indicated by yellow color takes part in final healed structure, although at the beginning it was part of the well-ordered lamella), (j) representative snapshot of healed structure. (k) Nematic order parameter of the defect molecules with respect to the large-domain director for the process at lower temperature (blue) and higher temperature (red). The bold lines represent the trajectories from which snapshots were taken; the snapshots are indicated by the letters. At high temperature, the process starts early, and it is fast, whereas at room temperature it takes longer to begin the process, and when initiated it takes a longer time to complete.

Finally, simulations of an asymmetrically functionalized molecule (see SI) showed that the rearrangement process of the isolated domain occurs in similar fashion as for the symmetric molecule described above. Although partial desorption rates for different molecules differ, we do not have enough statistics to make statements about how these affect the rate of the rearrangement of the isolated domain.

**Mobility.** Our simulations at different temperatures suggest that the rate of coarsening depends on the partial vertical desorption of molecules from the surface, which is enhanced near the edge of the graphite flake, at domain boundaries, and at higher temperature. Partial vertical desorption is expected to increase the lateral mobility of molecules on the surface because packing is less tight. We previously also observed that when the surface is not completely covered, the lateral mobility of adsorbent on the surface is high<sup>177</sup>. To investigate the influence of the partial desorption on the surface mobility, we imposed forces in the z-direction (the normal vector to the graphite plane) restraining the adsorbed molecules to the surface, thereby suppressing (partial and complete) desorption. For a single molecule, the restraining force can be related to adsorption energy. We randomly removed ~10% molecules from the system and observed changes in the mobility. Increasing the restraining force leads to a drop in mobility (see Table S5.2 for quantification). Thus, the lateral mobility of the molecules on the surface strongly depends on the level of partial desorption. Our simulations are consistent with the view that molecules do not need to fully desorb from the surface to enable coarsening; the coarsening process is, however, slowed down considerably if the molecules are more tightly bound to the surface.

## Conclusions

In this work, we systematically studied the mechanism and the parameters influencing the Ostwald ripening process of self-assembled monolayers. We observed that molecules are more mobile on the interface between different domains, and they tend to partially vertically desorb from the surface. Partial vertical desorption allows Ostwald ripening to take place: molecules can laterally desorb from one domain, laterally diffuse and adsorb on another. Since the rate of partial desorption is higher on the interface, in the long term, the defects promote coarsening. In all of the simulations, increasing the temperature resulted in speeding up the healing of structure. This is caused by increased desorption from the substrate. However, higher temperature also results in a

substantial mixing of the adsorbed molecules on the surface. If the assembly consists of identical molecules, as in this work, this effect is irrelevant; however, it can be crucial for assemblies consisting of various molecules and must be taken into account. Although we did not follow the effect of different solvents here, we expect it to be similar to the temperature effect: a different solvent could influence partial desorption from the substrate, thus speeding up or hindering the healing. We believe that all of the presented results give a clear picture of the mechanism of Ostwald ripening.

## Methods

**Coarse-grained model.** Simulations were performed using Gromacs 5 software<sup>91</sup> with the Martini coarse-grained force-field<sup>87</sup>. The Martini force-field, initially developed for lipids, was adapted rapidly for other biomolecules<sup>87,88</sup>, and recently to adsorption and self-assembly of small organic molecules on graphite<sup>142</sup>. As a rule of thumb, the force field represents the system by treating four non-hydrogen atoms as one bead (except for atoms inside a ring, where mapping is 2:1). Here, we studied a six-bead molecule with two polar beads, which can be viewed as a coarse-grained representation of N,N'-dodecanomethylenebis(hex)anoamide. Results for other adsorbents can be found in the SI. Self-assembly on a graphite surface of these molecules has been studied in the past<sup>145</sup>. The molecules were randomly distributed in the system and solvated in phenyloctane. The molecules adsorb on the surface and form long-range ordered lamellar structures, which are aligned with the underlying graphite layer (see SI). After 2  $\mu$ s of simulations an ordered structure formed on a graphite flake. Most often a single, well-ordered domain is formed. However, in some instances, we observed several domains. We used them to study their coarsening by further simulation up to 6.5  $\mu$ s. For self-healing of the small defect, a well aligned long-range ordered structure was used. Five molecules in the middle of the structure were rotated by 60°. Then the simulations were carried out for another 4  $\mu$ s. All simulations were run at two temperatures – 298 K referred here as lower temperature, and 308 K, referred as higher temperature. We intentionally avoid referring to the absolute temperatures because in general it is difficult to interpret temperature scales in coarse-grained models<sup>178</sup>. We also performed simulations at 318 K, at which long-range ordered structure completely disappears (see SI).

**Analysis of monolayer.** Vertical desorption was analyzed as two separate telegraph processes<sup>179</sup> describing desorption and partial desorption, respectively.

For analysis of desorption, molecules were assigned to one of two states: adsorbed (all beads are adsorbed) or desorbed (all beads are desorbed). A single bead is considered adsorbed if it is within 0.5 nm from the graphite surface. The residence time of an adsorbed state was calculated, and from it, rate of desorption, i.e. the change from adsorbed to desorbed state, was calculated. Similar calculations were performed for the partial desorption process, in which molecules were assigned to one of three states: fully adsorbed (five or six beads adsorbed), partially adsorbed (one to four beads adsorbed) and fully desorbed (all beads desorbed). The residence time of a fully adsorbed state was calculated and from it the rate of partial desorption, i.e. change from fully adsorbed state to partial desorbed state, was calculated. Details of this analysis can be found in SI.

**Nematic order parameter** quantifies alignment of a structure. To calculate the parameter, first the preferred orientation (director) has to be found. This is achieved by finding eigenvectors of the matrix  $Q = \left[ \frac{1}{2} \sum_j (3d_{j,a}d_{j,b} - 1) \right]_{a,b \in (x,y,z)}$ , where  $d$  denotes the molecular vector (vector connecting the two ends of molecule) of molecule  $j$ , and the sum runs over all adsorbent molecules. The eigenvector with the largest eigenvalue is called the director,  $D$ , and it is the preferred orientation of the molecules. The nematic order parameter is calculated as a sum  $S = \frac{1}{2N} \sum_j (3\cos^2\alpha_j - 1)$ , where  $\alpha_j$  is angle between the director,  $D$ , and the molecular vector of  $j$ -th molecule,  $d_j$ . Low values mean that the molecules are oriented in random directions, a value close to one means that all molecules are aligned in the same direction. In case of self-healing of small defects, the director was calculated with respect only to the large well-ordered structure.

## Supporting Information

### Different types of molecules

We analyzed dynamics of assemblies consisted of different types of six-beads molecules (Figure S5.1a): AM25 (two polar in position 2 and 5, this molecule is the focus of the article), AM2 (the polar bead in position 2), AM3 (the polar bead in position 3), ALK (no polar beads), AL1 (the polar beads in position 1). The stability of the structures was highest for AM25 molecules (the focus of the article) due to the presence of two polar groups, which stabilize the lamellae. Although the stability of the assemblies varies across different groups, the same conclusions can be drawn: (a) inner parts of assembly are much more stable than parts close to the edge of the graphite flake, (b) the frequency of partial desorption is  $\sim 10$  times higher than the full desorption (compare timescale of partial and full desorption). It is also noteworthy that when a molecule desorbs, reabsorption occurs on a similar time scale ( $\sim 10$  ns) across all types of molecules. All results are summarized in Table S1.

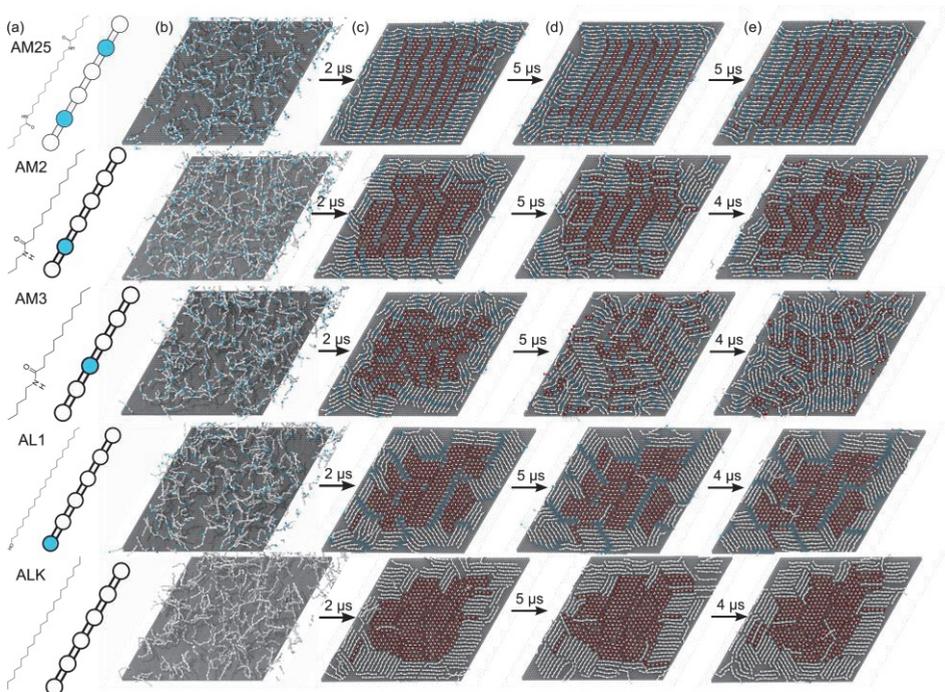


Figure S5.1. (a) Studied molecules and their coarse-grained representation. Snapshots of self-assembly of these molecules: (b) initial system, (c) self-assembled systems after 2  $\mu\text{s}$  (the molecules in the inner part of the assembly,  $\sim 2$  nm from the edge, are indicated in red), (d) the same assemblies after additional 5  $\mu\text{s}$  of simulation; these 5  $\mu\text{s}$  of simulation were used for analysis, (e) further 4  $\mu\text{s}$  (in case of AM25 5  $\mu\text{s}$ ) of simulations, not analyzed in this work.

### Underlying graphite structure

The distribution of angles between molecular vector and the main symmetry axes of graphite (Figure S5.2a) in self-assembled monolayers are close to 0 and 60 (Figure S5.2b-d), which indicates that molecules which are part of the ordered domains align with underlying graphite structure.

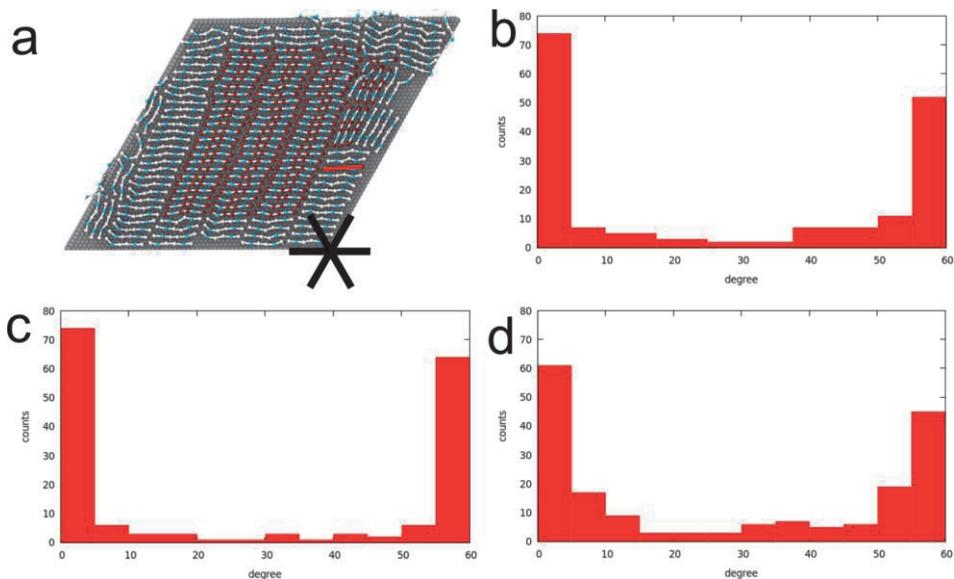


Figure S5.2. (a) Snapshot of self-assembled monolayer after  $2 \mu\text{s}$  of simulations (the same as Figure 5.1c in main text), red line exemplifies the molecular vector of adsorbent, and the black lines show main symmetry axes of graphite. (b-f) Distribution of the angle between the molecular vectors and the main symmetry axes of graphite for (b) self-assembled monolayer after  $2 \mu\text{s}$  of simulation (Figure 5.1c in main text; Figure 5.2a), (c) self-assembled monolayer after additional  $5 \mu\text{s}$  of simulation (Figure 5.1d in main text), (d) two-domain self-assembled monolayer (Figure 5.3a in the main text; Figure S5.5a).

### Rates of partial and full desorption

Vertical desorption was analyzed as two separate telegraph processes<sup>179</sup> (Figure S5.3 and Figure 5.2 in the main text). The telegraph process is a description of a process as a sequence of numbers, representing different states (Figure S5.3b schematically represents different states and Figure S5.3a show example snapshots). To describe full desorption, we assign to every molecule one of two states (Figure S5.3c): adsorbed, when all beads are adsorbed, or desorbed when all beads are desorbed. A single bead is considered desorbed if it is at least 0.5 nm from the graphite surface. Then, we analyze the trajectory for a single molecule by measuring the time spent in one of these states. We average these times over all molecules present in the system, obtaining an average residence time of adsorbed state,  $\tau_a$ , and desorbed state,  $\tau_d$ . Since the only change from the adsorbed state is to the desorbed state, we can calculate the rate of leaving the adsorbed state, that means the rate of desorption as  $f_d = \frac{1}{\tau_a}$ . Similarly, we can calculate the rate of readsorption as  $f_{ra} = \frac{1}{\tau_d}$ . The process of partial desorption is calculated in the same manner. In this case, we

distinguish three states (Figure S5.3d): fully adsorbed, when five or six beads are adsorbed, partially adsorbed, when one to four beads are adsorbed, and desorbed when all beads are desorbed. The relaxation time of adsorption-desorption of only one bead from a molecule is too short to be resolved; therefore, the molecule with one desorbed bead is considered as adsorbed. Similarly, we calculate residence times of fully adsorbed state,  $\tau_{fa}$ , partially adsorbed state,  $\tau_{pa}$ , and desorbed state,  $\tau_d$ . The rate of partial desorption can be then calculated as  $f_{pd} = \frac{1}{\tau_{fa}}$ , and similarly rate of desorption as  $f_{ra} = \frac{1}{\tau_d}$ . Since molecule in partially desorbed state can change state to two different states (fully adsorbed or desorbed), the rate  $\frac{1}{\tau_{fa}}$  is hard to interpret. The rates of desorption,  $f_d$ , and partial desorption (for both cases is the same),  $f_{pd}$  are represented as a bar chart in Figure 5.2f in the main text. All rates  $f_d$ ,  $f_{pd}$ , and  $f_{ra}$ , as well residence time of partially adsorbed state,  $\tau_{pa}$ , are shown in Table S5.1. For asymmetric molecules (i.e. AM2, AM3, AL1) we counted which side of the molecule (with or without polar bead) partially desorb more often. In these type of molecules, apolar part of the molecules desorbs more often, which probably is a result of strong interaction between polar beads in lamellae.

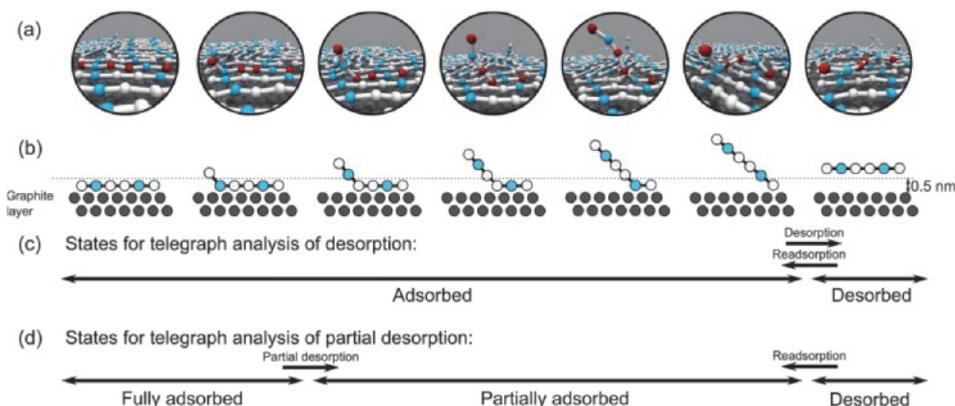


Figure S5.3. Scheme representing different states of molecules. (a-b) Molecules are classified to different states depending on the number of adsorbed beads. Different states are presented on (a) as a representative snapshots of simulation or on (b) as schematic representation. A single bead is considered desorbed if it is at least 0.5 nm from the graphite surface. (c) States used for telegraph analysis of desorption: adsorbed (one to six beads adsorbed) and desorbed (all beads desorbed). (d) States used for telegraph analysis of partial desorption: fully adsorbed (five or six beads adsorbed), partially desorbed/adsorbed (two to six beads are desorbed) and fully desorbed (six beads are desorbed).

Table S5.1. Rates of partial and full desorption and residence time of the partial desorbed state of molecules from an assembly on a graphite flake for five different functionalized long alkanes, see Figure S5.1. The calculation of the rates and residence time is explained in the text accompanying Figure S5.3. For partial desorption of asymmetric molecules, we counted which side of the molecule (with or without polar bead) desorbs more often, and the percentage of desorption of polar side is shown in brackets.

Molecule	Part of flake		Rate of desorption, $f_d$ [ $\mu\text{s}^{-1}$ ]	Rate of partial desorption, $f_{pd}$ [ $\mu\text{s}^{-1}$ ]	Residence time of partial desorbed state, $\tau_{pa}$ [ns]	Rate of readsorption, $f_{ra}$ [ $\mu\text{s}^{-1}$ ]
AM25	Inner	298K	0.49	4.46	0.66	84.59
		308K	2.43	22.18	0.58	116.95
	Outer	298K	5.77	55.89	0.63	85.77
		308K	8.55	71.43	0.58	105.40
	2 domains Inner	298K	1.79	17.85	0.68	93.76
		308K	2.76	24.80	0.62	96.82
	2 domains Outer	298K	6.00	66.94	0.62	93.71
		308K	8.22	69.67	0.61	99.39
AM2	Inner		1.23	20.73 (polar side: 33%)	0.66	93.30
	Outer		6.47	79.11 (polar side: )	0.67	85.88

			41%)		
<b>AM3</b>	Inner	5.54	94.37 (polar 42%)	side: 0.54	107.51
	Outer	6.78	117.19 (polar 45%)	side: 0.56	118.89
<b>AL1</b>	Inner	0.26	7.24 (polar side:42%)	0.74	733.94
	Outer	4.86	72.46 (polar 38%)	side: 0.78	88.50
ALK	Inner	0.64	9.41	1.66	65.79
	Outer	5.33	74.65	0.77	126.79

### Ostwald ripening close to the graphite flake edge

We have observed that defects, which are close to the edge of the graphite flake coarsen faster due to increased mobility of molecules in this region (see Table S1). See Figure S5.4.

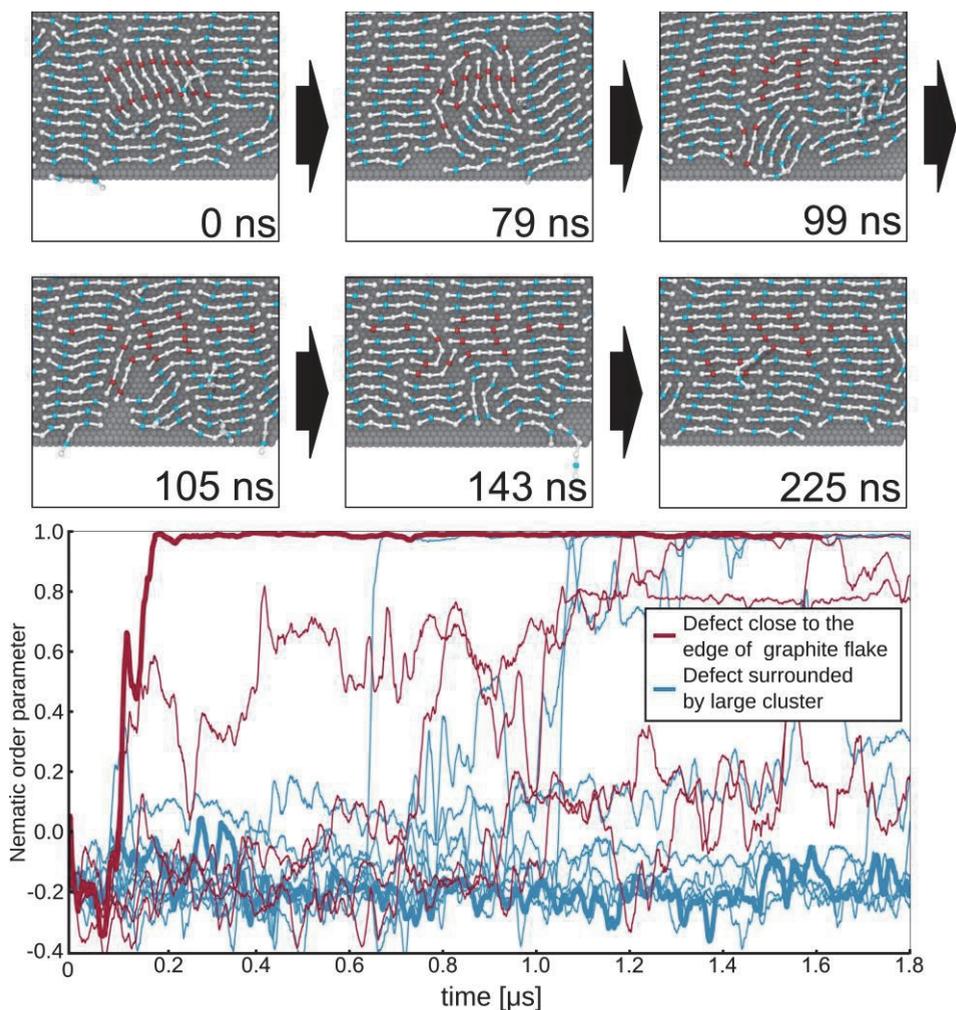


Figure S5.4. Self-healing of an isolated defect close to the graphite edge and surrounded by a larger cluster; compare to Figure 5.4 in the main text. Five independent simulations of the process were performed at 298 K.

### Diffusion coefficients

Influence of partial desorption on the mobility of molecules was checked by imposing harmonic restraining forces in the  $z$ -direction (the normal vector to the graphite plane). This additional force was applied to all beads of all adsorbent molecules and not to beads of solvent molecules. 10% of adsorbent molecules were removed from the system to avoid very low diffusion due to tight lateral packing of long-range ordered domains. We observed that diffusion dropped with increasing force constant (Table S5.2a).

Table S5.2. Lateral diffusion coefficient and its dependence on partial desorption.

Force constant [kJ/mol/nm <sup>2</sup> ]	Lateral diffusion coefficient [m <sup>2</sup> /s]
0	$3.3 \cdot 10^{-10}$
10	$6.3 \cdot 10^{-12}$
500	$2.5 \cdot 10^{-12}$

### Ostwald ripening of two-domain system (whole)

Figure S5.5 presents the full view of the Ostwald ripening of two domains of self-assembled monolayers oriented in different directions. Zoom-ins are shown in Figure 5.3 in the main text.

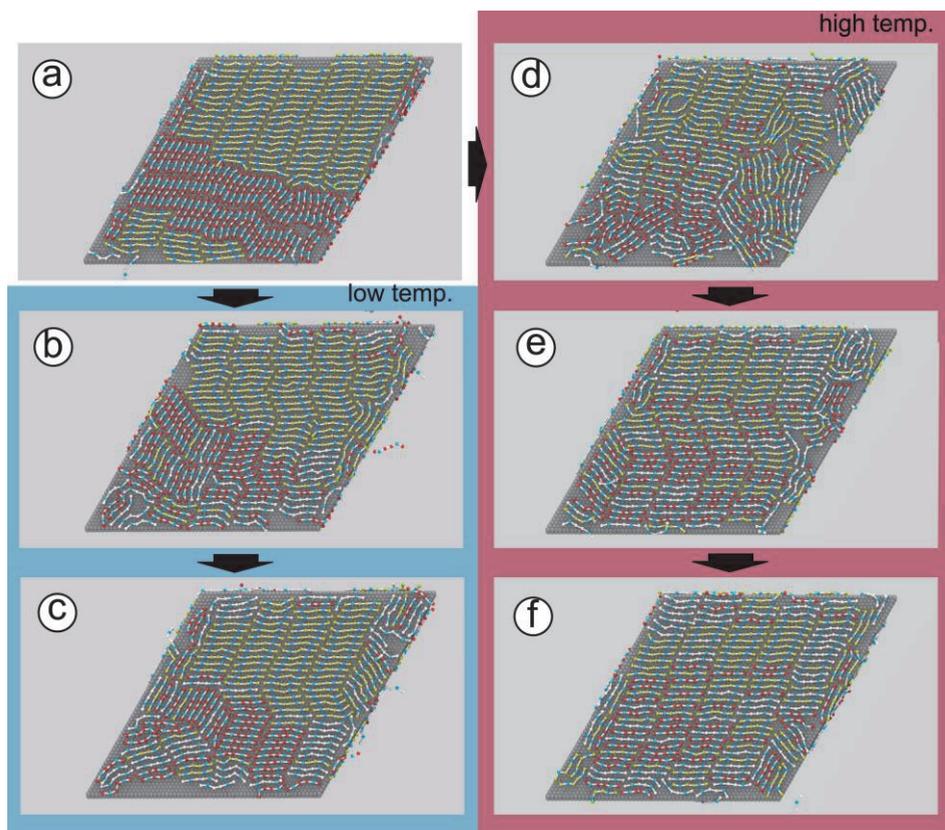


Figure S5.5. Ostwald-ripening of self-assembled monolayer consisted of two domains at low (b-c) and at high (d-f) temperature. Figure show reorganization of the whole assembly.

### Zig-zag patterns

Zig-zag patterns are observed in almost all our simulations of these long-chained molecules. As explained by Padowitz et al. (their Figures 5 and 7), their occurrence is a reflection of one or more

twin boundaries, in which the displacement or shift between two neighboring molecules within a lamella changes. In this case, the angle between the lamellar axis and the molecular axis is not  $90^\circ$ , but  $60^\circ$ . Whether or not this occurs depends on the nature of the adsorbent molecules. In our model, there is such a displacement between neighboring molecules in general, because of the coarse-grained nature of the beads. The lateral interaction between neighbors is more favorable if the neighboring molecules are displaced by half a bond length, because a bead interacts with two neighboring beads instead of one.

In our simulations, we observe 2-3 kinks in long lamellae, consisting of 30-40 molecules, which are rather dynamic and often disappear during long simulation, but may also appear. We therefore view them as defects that can heal.

In our view, the free energy difference between a straight lamella and one showing a zig-zag pattern is probably relatively small. The zig-zags can easily accommodate defects or domains at a different orientation, as can be seen in Figure 5.3b-c in the bottom left-hand corner. While maintaining an overall favorable lamellar structure, in particular enabling parallel lamellae to remain in favorable contact, the zig-zags sacrifice a little free energy by kinking. It would be interesting to substantiate this belief by explicit calculation of the free energy difference. We expect this is in principle possible, but also non-trivial to implement correctly in simulations, and are not able to complete this now.

In support of the view that the zig-zags can be seen as defect, note also that the zig-zags are observed at both high and low temperatures, as can be seen in Figure 5.3c, and Figure 5.3e, respectively. The difference is that the zig-zag defect heals at high temperature on the simulated time scale, whereas it does not at low temperature, which is consistent with the slower healing process at lower temperatures.

The dynamics of a zig-zag pattern was observed by Padowitz et al.<sup>148</sup>. They suggest that the motion is driven by the labile site at the domain boundaries.

### **Ostwald ripening of small defect system (whole)**

Figure S5.6 presents the full view of the Ostwald ripening of a small defect in self-assembled monolayer on a graphite flake. The zoomed-in images are shown in Figure 5.4 in the main text.

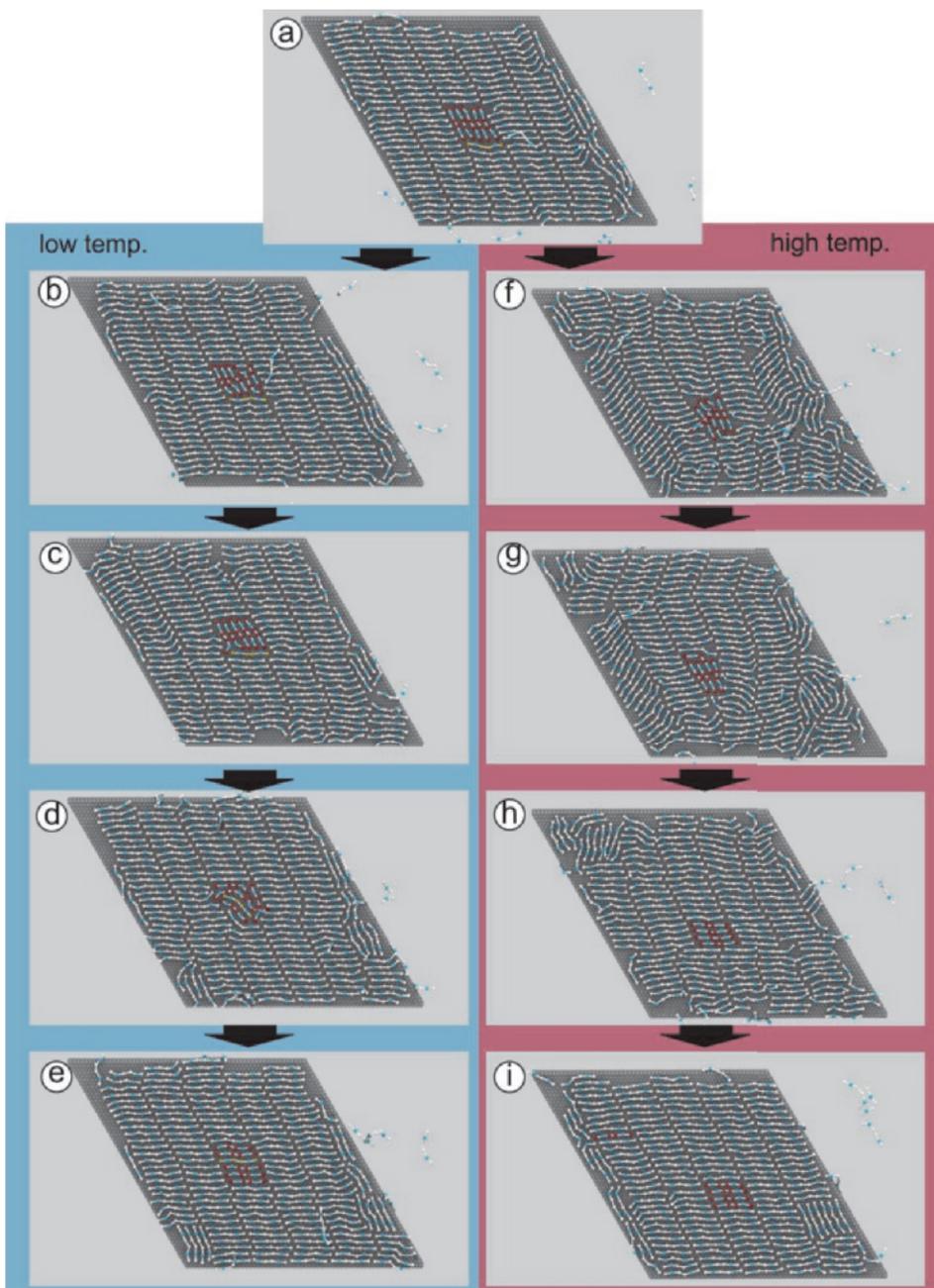


Figure S5.6. Ostwald ripening of small defect in self-assembled monolayer (a) at low (b-e) and at high (f-i) temperature. Figure shows reorganization of the whole assembly.

### Ostwald ripening of a small defect for AM2 type of molecules

Figure S5.7 shows rearrangement of a small isolated defect for asymmetric molecules with one polar bead. We have run three independent simulations (i.e. simulated six defects) and observed three Ostwald ripening processes, from which two coarsened to a parallel configuration and one with an antiparallel orientation to surrounding molecules (the snapshots from this particular simulation are shown in Figure S5.7). The overall reorganization is similar to the symmetric molecules: molecules rearrange on the surface of the graphite flake facilitated by partial desorption. Although partial desorption rates for different molecules differ, we do not have enough statistics to make statements about how these affect the rate of the rearrangement of the isolated domain.

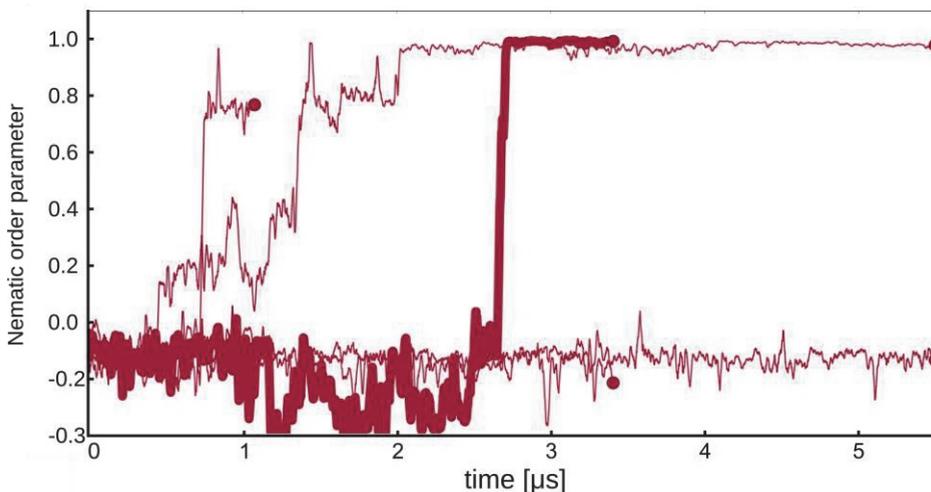
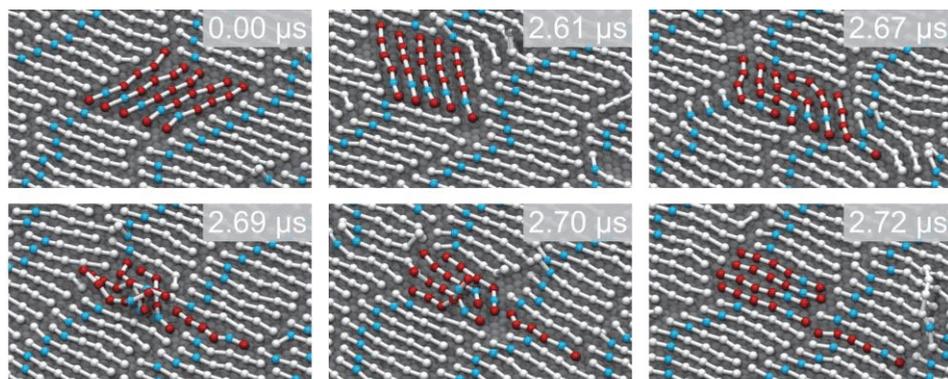


Figure S5.7. Coarsening of a small isolated defect of AM2 type molecules; compare to Figure 5.4 in the main text. Three independent simulations (with different simulation time indicated by red circles) of the process were performed at 298 K.

### Ostwald ripening at 318 K

Simulations presented in this work were performed at 298 K and 308 K, referred in this paper as low and high temperature respectively (see Methods). We observed that additional increase of temperature to 318 K results in disassembly of the structure. (Figure S5.8a-f). This can also be seen as a significant drop in nematic order parameter of whole structure (Figure S5.8g).

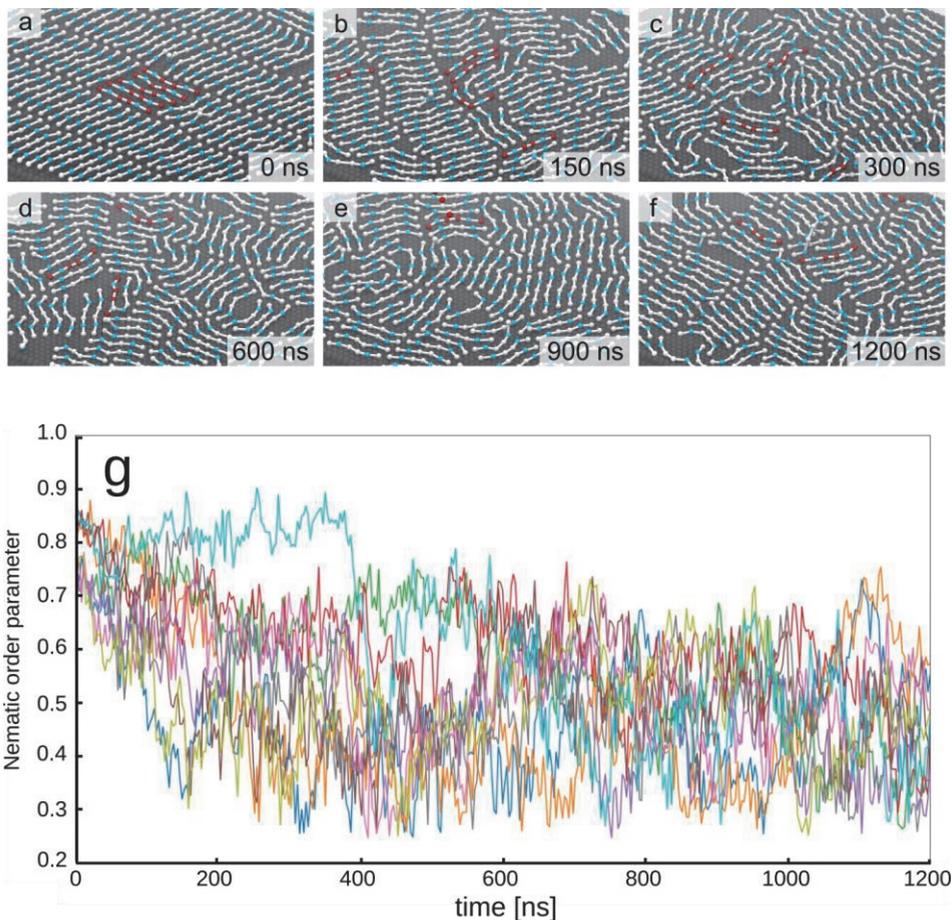


Figure S5.8. Disassembly at 318 K. At 318 K the whole structure disassembles a-b, which can be seen on individual traces of the nematic order parameter of the whole structure (g). Different traces represent independent simulations, all starting from a well-ordered structure.

### Plotting

For clarity, all plots are treated by rolling mean. Rolling mean is calculated by taking average of subsets of full data. It is used to smooth plots, which consists of many short-term fluctuations. In this paper we use window size of 50, which is equivalent to 15 ns of simulation.

## Description to supplemented videos

All supplemented videos are done with VMD<sup>169</sup>. For every video, we applied trajectory smoothing to make them more clear (therefore, on some of them it can be seen that molecules act unphysical (i.g. overlap) – it means that in between frames molecules changed conformation extensively). In all videos solvent (phenylooctane) is not shown.

SI\_2domain\_healing\_298.mp4: Video presents coarsening of a two-domain system (initial membership to one of the domains is indicated by the color codes red and yellow) at lower temperature (298 K). White color indicates molecules which are not part of the original domains (they originate in the supernatant or from the edges of the flake). The video presents representative simulation from an initial structure at 0.0  $\mu\text{s}$  until its final stage at 6.5  $\mu\text{s}$ . Snapshots are taken every 6 ns, and trajectory smoothing with a window of 5 frames is used. From right to left it can be seen that a defected is healed lamella after lamella, eventually covering most of the surface with one domain. Snapshots of this video are presented in the main text on Figure 5.3b-c.

SI\_2domain\_healing\_308.mp4: Video presents coarsening of a two-domain system (initial membership to one of the domains is indicated by the color codes red and yellow) at higher temperature (308 K). White color indicates molecules which are not part of the original domains (they originate in the supernatant or from the edges of the flake). The video presents representative simulation from an initial structure at 0.0  $\mu\text{s}$  until 6.5  $\mu\text{s}$ . Snapshots are taken every 6 ns, and trajectory smoothing with a window of 7 frames is used. In comparison to the lower temperature, the whole system is more dynamic, many changes in the structure during the course of simulation can be observed. Around 2.3  $\mu\text{s}$  one domain disassembles into an unordered structure. After some rearrangement, around 2.7  $\mu\text{s}$ , it reassembles to existing domain creating one ordered structure. Snapshots of this video are presented in the main text on Figure 5.3d-f.

SI\_defect\_healing\_298.mp4: Video presents coarsening of a small defect (indicated by red) surrounded by one domain in lower temperature (298K). Molecule indicated by yellow does not take part in the defect, but at the end of the simulation, it is part of the healed defect. The video presents representative simulation from 2.0  $\mu\text{s}$  until 2.3  $\mu\text{s}$ . Snapshots are taken every 0.9 ns, and trajectory smoothing with a window of 3 frames is used. During the rearrangement, molecules move only on the surface. This movement is possible because of the partial desorption of molecules. Snapshots of this video are presented in the main text on Figure 5.4f-i.

SI\_defect\_healing\_308.mp4: Video presents coarsening of a small defect (indicated by red) surrounded by one domain in higher temperature (308K). The video presents representative simulation from 1.2  $\mu\text{s}$  until 1.7  $\mu\text{s}$ . Snapshots are taken every 0.9 ns, and trajectory smoothing with a window of 3 frames is used. During the simulation, one of the molecules from the defect desorbs (around 1.27  $\mu\text{s}$ ). Shortly after, rest of the molecules rearrange creating ordered structure. Snapshots of this video are presented in the main text on Figure 5.4b-e.

# Summary

## General

In this thesis, we explore the application of various molecular simulations techniques to give insights into the self-assembly of supramolecular systems. Chapter 1 explains the importance of molecular simulation to study the self-assembly process. In chapter 2 we study self-assembly of a derivative of 1,3,5-triamidocyclohexane (CTA) using common techniques: simulations of self-assembly from randomly distributed molecules and simulations of the final structure. The results show the importance of the choice of force-field and limitations of conventional molecular dynamics to give insights into processes which occur on a long timescale. In Chapter 3 we tackle the timescale issue by using an adaptive sampling method. The results provide a unique insight into the kinetic pathways of the self-assembly process. Moreover, we were able to provide insights into the next stages of the self-assembly. Although, the method provides insight at a level of detail hardly accessible by any other technique it is limited to rather small systems. In Chapter 4 we study self-assembly of long functionalized alkanes on a graphite flake. We use coarse-grained molecular dynamics to tackle both temporal and spatial scales instead of the high resolution of the all-atomistic model. These results give insights into the mechanism of self-assembly of monolayers on graphite. Chapter 5 is an extension to Chapter 4. Here, we study the last stage of the self-assembly process, Ostwald ripening, responsible for correction of the structure, which leads to high quality long-range ordered assemblies.

The results presented in this thesis have two major outcomes: (a) methodology to simulate of self-assembly, and (b) insights into self-assembly process.

## Simulations of self-assembling systems

The results of this thesis show challenges, which the researcher can encounter during the simulations of self-assembly of supramolecular systems. The three most significant challenges are:

- (a) Force-field quality,
- (b) Long time scale,
- (c) Large spatial scale.

Chapter 1 shows that depending on the choice of the force-field simulations can lead to different results. This chapter shows the importance of force-field validation. The choice of force-field strongly depends on the system and the level of details of interest. For the self-assembly of long functionalized alkanes we used a coarse-grained force-field, which only roughly estimates interactions between molecules, but still gives results aligning with experiments. The supramolecular fiber formation from CTA molecules turned out to be more challenging and usage of a high quality, computationally expensive force-field (polarizable CHARMM Drude) turned out to be necessary to simulate the self-assembly process.

To tackle issues of temporal and spatial scales we used different techniques allowing access to scales beyond conventional molecular simulations. An adaptive sampling method, a rare-event method which we developed in this thesis, turned out to be a great method to provide high quality and resolution insights into small self-assembling systems. We were able to show the kinetic pathways leading to the formation of the final structure. Although the level of detail provided by this method is excellent, it is limited only to small systems. To study systems with large spatial and temporal scale, such as self-assembly on graphite, we used a coarse-grained technique. Although this technique loses fine details of the process, it still provides unique insights into the process.

### **Insights into self-assembly**

The results presented in this thesis provide new insights into self-assembly processes of (a) supramolecular fiber and (b) physisorbed monolayer on a graphite surface.

In Chapter 3 we give insights into the mechanisms of formation of a supramolecular fiber by studying several, distinct phases: aggregation and nucleation, elongation, secondary nucleation, and bundling. We show that the process starts with aggregation of molecules into an unordered cluster, which slowly by stepwise reorganization changes into an ordered structure. The formed fiber elongates most likely by attachment of new molecules on the side of the fiber followed by its diffusion towards one of its ends, where it absorbs. Moreover, the surface of the fiber catalyzes the formation of new nuclei, whose growth on the side of the existing fiber results in the creation of a bundle of fibers. Lastly, a bundle can be also created by interaction of already existing

fibers. Results presented here give detailed insights into processes leading to the formation of supramolecular fiber.

In Chapters 4 and 5 we give insights into the formation of long-ranged ordered lamellar structures by long functionalized alkanes on a graphite surface. The results show that the mechanism follows two processes happening simultaneously: adsorption and rearrangement on the surface. The reorganization starts with nucleation (Chapter 4), followed by Ostwald ripening (Chapter 5). We show that the mechanism of nucleation is temperature dependent. At lower temperature adsorbed molecules interact with each other, which results in small ordered domains. At higher temperature, adsorbed molecules do not create ordered structures, but a liquid-like, unordered phase, and only upon reaching a certain level of surface coverage of molecules the system undergo phase transition leading to an ordered structure. The resulting structure may still have some defects, which can be repaired by Ostwald ripening, i.e., a process in which large ordered structures grow at the expense of smaller ones. We show that in Ostwald ripening the partial desorption is an essential process, which allows rearrangement on the surface. The Ostwald ripening can be controlled by changing temperature due to its influence on partial desorption.

Although the nature of these systems is different, they exhibit similarities. Both self-assembly processes start with separation into an unordered phase by non-directional interaction (for CTA in the unordered aggregate by van der Waals and hydrophobic interactions, and for long functionalized alkanes in molecules adsorbed on graphite by van der Waals interactions). The newly separated, unordered phase undergoes rearrangement, which leads to the formation of the ordered structure via strong directional interactions (in both cases by hydrogen bonds).

## **Overall conclusions**

In conclusion, this thesis contributes to understanding self-assembly process by molecular simulations. The results presented here show that molecular simulations can be used as a molecular microscope for self-assembling systems, which provides insights into the process on a spatial and temporal scale not accessible by experimental techniques. This thesis shows the great potential of molecular simulation into providing insights into the self-assembly of supramolecular systems.

# Samenvatting

## Algemeen

In dit proefschrift verkennen wij de toepassing van verscheidene moleculaire simulatietechnieken om meer inzicht te krijgen in de zelf-assemblage van supramoleculaire systemen. Hoofdstuk 1 legt uit hoe belangrijk moleculaire simulatie is voor het bestuderen van het zelf-assemblageproces. In hoofdstuk 2 bestuderen we de zelf-assemblage van een 1,3,5-triamidocyclohexaan (CTA) derivaat met gebruikelijke technieken: simulatie van zelf-assemblage van willekeurig verdeelde moleculen en simulaties van de eindstructuur. De resultaten laten zien hoe belangrijk de keuze in krachtveld en de beperkingen van conventionele moleculaire dynamica zijn om inzicht te krijgen in de processen op lange tijdschaal. In hoofdstuk 3 pakken we het probleem van tijdschaal aan door een ‘adaptive sampling’ methode te gebruiken. De resultaten verschaffen een uniek inzicht in de kinetische paden van het zelf-assemblageproces. Bovendien kunnen we inzicht verkrijgen in de volgende fasen van zelf-assemblage. Hoewel de methode inzicht verschaft op een niveau van detail dat nauwelijks bereikbaar is met andere technieken, is het gelimiteerd tot nogal kleine systemen. In hoofdstuk 4 bestuderen we de zelf-assemblage van lange gefunctionaliseerde alkanen op een grafietschilfer. We gebruiken ‘course-grained’ moleculaire dynamica om de tijd- en ruimtelijke schaal te onderzoeken in plaats van de hoge resolutie van ‘all-atomistic’ modellen te gebruiken. Deze resultaten geven inzicht in het mechanisme van zelf-assemblage van monolagen op grafiet. Hoofdstuk 5 is een voorbereiding van hoofdstuk 4. Hierin bestuderen we het laatste stadium van het zelf-assemblageproces, ostwaldrijping, verantwoordelijk voor het corrigeren van de structuur, dat leidt tot hoge kwaliteit lange-afstand geordende assemblages.

De gepresenteerde resultaten in dit proefschrift hebben twee grote uitkomsten: (a) de methodologie om zelf-assemblage te simuleren, en (b) het inzicht in het zelf-assemblageproces.

## Simulaties van zelf-assemblagesystemen

De resultaten in dit proefschrift laten de uitdagingen zien die de onderzoeker tegen kan komen tijdens de simulaties van zelf-assemblages van supramoleculaire systemen. De drie belangrijkste uitdagingen zijn:

- (a) Kwaliteit in krachtveld,

- (b) Lange tijdschaal,
- (c) Grote ruimtelijke schaal.

Hoofdstuk 1 laat zien dat afhankelijk van de keuze in de krachtveldsimulatie kan leiden tot verschillende resultaten. Dit hoofdstuk laat zien hoe belangrijk de validatie van het krachtveld is. De keuze in krachtveld is sterk afhankelijk van het systeem en het niveau van gewenst detail. Voor de zelf-assemblage van lange gefunctionaliseerde alkanen hebben we de ‘course-grained’ krachtveld gebruikt dat alleen een ruwe schatting maakt van de interacties tussen de moleculen, maar nog steeds resultaten geeft die in overeenstemming zijn met experimenten. De supramoleculaire fibervorming van CTA moleculen bleek uitdagender te zijn en het gebruik van een hoge kwaliteit, rekenkundig duur krachtveld (polariseerbare CHARMM Drude) bleek nodig te zijn om het zelf-assemblageproces te simuleren.

Om de problemen van tijd- en ruimtelijke schalen aan te pakken hebben we verschillende technieken gebruikt die toegang bieden tot schalen buiten de conventionele moleculaire simulaties. Een ‘adaptive sampling’ methode, een zeldzaam-gebeurtenis methode die we hebben ontwikkeld in dit proefschrift bleek een geweldige methode te zijn om hoge kwaliteit en resolutie inzichten in kleine zelf-assemblagesystemen te verkrijgen. We kunnen de kinetische paden laten zien die leiden tot de vorming van de eindstructuur. Hoewel het niveau van details verstrekt door de methode uitstekend is, is het gelimiteerd tot kleine systemen. Om systemen met een grote ruimtelijke en tijdschaal te bestuderen, zoals de zelf-assemblage of grafiet, hebben we een ‘coarse-grained’ techniek gebruikt. Hoewel deze techniek fijne details verliest, verschaft het nog steeds unieke inzichten in het proces.

### **Inzichten in zelf-assemblage**

De resultaten gepresenteerd in dit proefschrift verschaffen nieuwe inzichten in zelf-assemblageprocessen van (a) supramoleculaire fiber en (b) fysisch geadsorbeerde monolaag op een grafietoppervlak.

In Hoofdstuk 3 geven we inzicht in de mechanismen van de vorming van een supramoleculaire fiber door verschillende, onderscheidende fases: aggregatie en nucleatie, groei, tweede nucleatie, en bundeling. We laten zien dat het proces start met de aggregatie van moleculen in een ongeordende cluster, die langzaam door stapsgewijze reorganisatie verandert in een geordende structuur. De

gevormde fibers groeien waarschijnlijk door de aanhechting van nieuwe moleculen aan de zijkant van de fiber gevolgd door zijn diffusie richting één van de einden, waar het absorbeert. Bovendien katalyseert de zijkant van de fiber de vorming van nieuwe kernen, die groeien op de zijkant van de bestaande fiber wat resulteert in de creatie van een bundel fibers. Tenslotte kan een bundel ook gecreëerd worden door de interactie van al bestaande fibers. De hier gepresenteerde resultaten geven gedetailleerde inzichten in de processen die leiden tot de vorming van supramoleculaire fibers.

In de Hoofdstukken 4 en 5 geven we inzicht in de vorming van lange-afstand geordende gelaagde structuren door lange gefunctionaliseerde alkanen op grafiet. De resultaten laten zien dat het mechanisme uit twee processen bestaat die tegelijkertijd gebeuren: adsorptie en herschikking van het oppervlak. De reorganisatie begint met nucleatie (Hoofdstuk 4), gevolgd door ostwaldrijping (Hoofdstuk 5). We laten zien dat het mechanisme van nucleatie temperatuur afhankelijk is. Op lagere temperatuur hebben de geadsorbeerde moleculen interactie met elkaar, wat resulteert in kleine geordende domeinen. Op hogere temperaturen creëren de geadsorbeerde moleculen geen geordende structuren, maar een vloeibaarachtige, ongeordende fase, en alleen door een bepaalde moleculaire bedekking van het oppervlak te bereiken, ondergaat het systeem een faseovergang dat leidt tot een geordende structuur. De resulterende structuur kan nog steeds wat gebreken hebben, welke hersteld kunnen worden door ostwaldrijping, i.e., een proces waarbij grote geordende structuren groeien ten koste van kleinere. We laten zien dat tijdens ostwaldrijping de gedeeltelijke desorptie een essentieel proces is, dat de herschikking van het oppervlak mogelijk maakt. De ostwaldrijping kan gecontroleerd worden door de temperatuur te veranderen vanwege zijn invloed op de gedeeltelijke desorptie.

Hoewel de aard van deze systemen verschillend is, vertonen ze gelijkenis. Beide zelf-assemblageprocessen starten met de scheiding in een ongeordende fase door niet-directionele interactie (voor CTA in het ongeordende aggregaat door Van der Waals en hydrofobe interacties, en voor de lange gefunctionaliseerde alkanen in moleculen geadsorbeerd op grafiet door Van der Waals interacties). De nieuwe gescheiden, ongeordende fase ondergaat herschikking, dat leidt tot de vorming van de geordende structuur via sterke directionele interacties (in beide gevallen door waterstofbindingen).

## **Algemene conclusies**

Ter conclusie, dit proefschrift draagt bij tot het begrijpen van zelf-assemblageprocessen door moleculaire simulaties. De hier gepresenteerde resultaten laten zien dat moleculaire simulaties gebruikt kunnen worden als een moleculaire microscoop voor zelf-assemblagesystemen, die inzicht geven in het proces op ruimtelijke en tijdschaal, niet toegankelijk voor experimentele technieken. Dit proefschrift laat de grote mogelijkheid zien van moleculaire simulaties voor het verkrijgen van inzicht in de zelf-assemblage van supramoleculaire systemen.

# References

- (1) George, M.; John, P.; Christopher, T. Molecular Self-Assembly and Nanochemistry: A Chemical Strategy for the Synthesis of Nanostructures. *Science* **1991**, *254*, 1312.
- (2) Whitesides, G. M.; Grzybowski, B.; Philip, D.; Stoddart, J. F.; Singh, R.; Maru, V. M.; Moharir, P. S.; Jakubith, S.; Rotermund, H. H.; Engel, W.; et al. Self-Assembly at All Scales. *Science* **2002**, *295* (5564), 2418–2421.
- (3) Meena, R. S. Possession and Transfer of Copyrights of a Cinematograph Film. *J. Intellect. Prop. Rights* **2016**, *21* (5–6), 283–287.
- (4) Mendes, A. C.; Baran, E. T.; Reis, R. L.; Azevedo, H. S. Self-Assembly in Nature: Using the Principles of Nature to Create Complex Nanobiomaterials. *Wiley Interdiscip. Rev. Nanomedicine Nanobiotechnology* **2013**, *5* (6), 582–612.
- (5) Israelachvil, J. N.; Mitchell, D. J.; Ninham, B. W.; Israelachvili, J. N.; Mitchell, D. J.; Ninham, B. W. Theory of Self-Assembly of Hydrocarbon Amphiphiles into Micelles and Bilayers View Online. *J. Chem. Soc., Faraday Trans. 2* **1976**, *72*, 1525–1568.
- (6) Antonietti, M.; Förster, S. Vesicles and Liposomes: A Self-Assembly Principle Beyond Lipids. *Adv. Mater.* **2003**, *15* (16), 1323–1333.
- (7) Israelachvili, J. N.; Mitchell, D. J.; Ninham, B. W. Theory of Self-Assembly of Lipids Bilayers and Vesicles. *Biochim. Biophys. Acta* **1977**, *470*, 185–201.
- (8) Aida, T.; Meijer, E. W.; Stupp, S. I. Functional Supramolecular Polymers. *Science* **2012**, *335* (6070), 813–817.
- (9) Weiss, R. G. The Past, Present, and Future of Molecular Gels. What Is the Status of the Field, and Where Is It Going? *J. Am. Chem. Soc.* **2014**, *136* (21), 7519–7530.
- (10) Aggeli, A.; Nyrkova, I. A.; Bell, M.; Harding, R.; Carrick, L.; McLeish, T. C. B.; Semenov, A. N.; Boden, N. Hierarchical Self-Assembly of Chiral Rod-like Molecules as a Model for Peptide  $\alpha$ -Sheet Tapes, Ribbons, Fibrils, and Fibers. *Proc. Natl. Acad. Sci.* **2001**, *98* (21), 11857–11862.
- (11) Anwar, J.; Zahn, D. Uncovering Molecular Processes in Crystal Nucleation and Growth by Using Molecular Simulation. *Angew. Chemie - Int. Ed.* **2011**, *50* (9), 1996–2013.
- (12) Davey, R. J.; Schroeder, S. L. M.; Ter Horst, J. H. Nucleation of Organic Crystals - A Molecular Perspective. *Angew. Chemie - Int. Ed.* **2013**, *52* (8), 2167–2179.
- (13) Wei, G.; Su, Z.; Reynolds, N. P.; Arosio, P.; Hamley, I. W.; Gazit, E.; Mezzenga, R. Self-Assembling Peptide and Protein Amyloids: From Structure to Tailored Function in Nanotechnology. *Chem. Soc. Rev.* **2017**, *46*, 4661–4708.
- (14) Slyngborg, M.; Fojan, P. A Computational Study of the Self-Assembly of the RFFFR Peptide. *Phys. Chem. Chem. Phys.* **2015**, *17*, 30023–30036.
- (15) Görbitz, C. H. The Structure of Nanotubes Formed by Diphenylalanine, the Core Recognition Motif of Alzheimer's  $\beta$ -Amyloid Polypeptide. *Chem. Commun.* **2006**, No. 22, 2332–2334.
- (16) Drexler, K. E. Molecular Engineering: An Approach to the Development of General Capabilities for Molecular Manipulation. *Proc. Natl. Acad. Sci. U. S. A.* **1981**, *78* (9), 5275–5278.
- (17) Adams, F. C.; Barbante, C. Nanoscience, Nanotechnology and Spectrometry. *Spectrochim. Acta - Part B* **2013**, *86*, 3–13.
- (18) Steed, J. W.; Atwood, J. L. *Chemistry Supramolecular Chemistry Second Edition*; 2009.
- (19) Israelachvili, J. N. *Intermolecular and Surface Forces*; 2010.
- (20) Hanabusa, K.; Yamada, M.; Kimura, M.; Shirai, H. Prominent Gelation and Chiral Aggregation of Alkylamides Derived Fromtrans-1,2-Diaminocyclohexane. *Angew. Chemie Int. Ed. English* **1996**, *35* (17), 1949–1951.
- (21) Furukawa, S.; Feyter, S. de. Two-Dimensional Crystal Engineering at the Liquid-Solid Interface. *Top. Curr. Chem.* **2009**, *287*, 87–133.
- (22) Elemans, J. A. A. W.; Lei, S.; De Feyter, S. Molecular and Supramolecular Networks on

- Surfaces: From Two-Dimensional Crystal Engineering to Reactivity. *Angew. Chemie - Int. Ed.* **2009**, *48* (40), 7298–7333.
- (23) Dastidar, P. Supramolecular Gelling Agents: Can They Be Designed? *Chem. Soc. Rev.* **2008**, *37* (12), 2699–2715.
- (24) Van Esch, J. H. We Can Design Molecular Gelators, but Do We Understand Them? *Langmuir* **2009**, *25* (15), 8392–8394.
- (25) Hunter, R. J. *Foundations of Colloid Science*; Oxford: Oxford University Press, 2001.
- (26) Ciferri, A. *Supramolecular Polymers*; Taylor & Francis Group, 2005.
- (27) De Greef, T. F. a; Smulders, M. M. J.; Wolfs, M.; Schenning, A. P. H. J.; Sijbesma, R. P.; Meijer, E. W. Supramolecular Polymerization. *Chem. Rev.* **2009**, *109* (11), 5687–5754.
- (28) Zhao, D.; Moore, J. S. Nucleation-Elongation: A Mechanism for Cooperative Supramolecular Polymerization. *Org. Biomol. Chem.* **2003**, *1* (20), 3471–3491.
- (29) Frederix, P. W. J. M.; Patmanidis, I.; Marrink, S. J. Molecular Simulations of Self-Assembling Bio-Inspired Supramolecular Systems and Their Connection to Experiments. *Chemical Society Reviews*. Royal Society of Chemistry 2018, pp 3470–3489.
- (30) Bochicchio, D.; Pavan, G. M. Molecular Modelling of Supramolecular Polymers. *Adv. Phys. X* **2018**, *3* (1), 1436408.
- (31) Kmiecik, S.; Gront, D.; Kolinski, M.; Wieteska, L.; Dawid, A. E.; Kolinski, A.; Models, S.; Field, F. Coarse-Grained Protein Models and Their Applications. **2016**.
- (32) Li, L.; Zhan, H.; Duan, P.; Liao, J.; Quan, J.; Hu, Y.; Chen, Z.; Zhu, J.; Liu, M.; Wu, Y. D.; et al. Self-Assembling Nanotubes Consisting of Rigid Cyclic  $\gamma$ -Peptides. *Adv. Funct. Mater.* **2012**, *22* (14), 3051–3056.
- (33) Lee, O. S.; Stupp, S. I.; Schatz, G. C. Atomistic Molecular Dynamics Simulations of Peptide Amphiphile Self-Assembly into Cylindrical Nanofibers. *J. Am. Chem. Soc.* **2011**, *133* (10), 3677–3683.
- (34) Yu, T.; Schatz, G. C. Free Energy Profile and Mechanism of Self-Assembly of Peptide Amphiphiles Based on a Collective Assembly Coordinate. *J. Phys. Chem. B* **2013**, *117* (30), 9004–9013.
- (35) Sasselli, I. R.; Pappas, C. G.; Matthews, E.; Wang, T.; Hunt, N. T.; Ulijn, R. V.; Tuttle, T. Using Experimental and Computational Energy Equilibration to Understand Hierarchical Self-Assembly of Fmoc-Dipeptide Amphiphiles. *Soft Matter* **2016**, *12* (40), 8307–8315.
- (36) Friedl, C.; Renger, T.; Berlepsch, H.; Ludwig, K.; Schmidt, M.; Megow, J. Structure Prediction of Self-Assembled Dye Aggregates from Cryogenic Transmission Electron Microscopy, Molecular Mechanics, and Theory of Optical Spectra. **2016**.
- (37) Hamley, I. W.; Alema, C. Their Hierarchical Assembly †. **2016**, 1265–1278.
- (38) Marchesan, S.; Waddington, L.; Easton, C.; Winkler, D. A.; Goodall, L.; Forsyth, J.; Hartley, P. G. Unzipping the Role of Chirality in Nanoscale Self-Assembly of Tripeptide Hydrogels. *Nanoscale*, **2012**, *4*, 6752–6760.
- (39) Garzoni, M.; Baker, M. B.; Leenders, C. M. A.; Voets, I. K.; Albertazzi, L.; Palmans, A. R. A.; Meijer, E. W.; Pavan, G. M. Effect of H-Bonding on Order Amplification in the Growth of a Supramolecular Polymer in Water. *J. Am. Chem. Soc.* **2016**, jacs.6b07530.
- (40) Bejagam, K. K.; Remsing, R. C.; Klein, M. L.; Balasubramanian, S. Understanding the Self-Assembly of Amino Ester-Based Benzene-1,3,5-Tricarboxamides Using Molecular Dynamics Simulations. *Phys. Chem. Chem. Phys.* **2016**, *19*, 258–266.
- (41) Ozkan, A. D.; Tekinay, A. B.; Guler, M. O.; Tekin, E. D. Effects of Temperature, pH and Counterions on the Stability of Peptide Amphiphile Nanofiber Structures. *RSC Adv.* **2016**, *6* (106), 104201–104214.
- (42) Ramos Sasselli, I.; Ulijn, R. V.; Tuttle, T. CHARMM Force Field Parameterization Protocol for Self-Assembling Peptide Amphiphiles: The Fmoc Moiety. *Phys. Chem. Chem. Phys.* **2016**, *18*, 4659–4667.
- (43) Giberti, F.; Salvalaglio, M.; Mazzotti, M.; Parrinello, M. Insight into the Nucleation of

- Urea Crystals from the Melt. *Chem. Eng. Sci.* **2015**, *121*, 51–59.
- (44) Salvalaglio, M.; Perego, C.; Giberti, F.; Mazzotti, M.; Parrinello, M. Molecular-Dynamics Simulations of Urea Nucleation from Aqueous Solution. *Proc. Natl. Acad. Sci.* **2015**, *112* (1), E6–E14.
- (45) Salvalaglio, M.; Mazzotti, M.; Parrinello, M. Urea Homogeneous Nucleation Mechanism Is Solvent Dependent. *Faraday Discuss.* **2015**, *179*, 291–307.
- (46) Giberti, F.; Tribello, G. a.; Parrinello, M. Transient Polymorphism in NaCl. *J. Chem. Theory Comput.* **2013**, *9* (6), 2526–2530.
- (47) Giberti, F.; Salvalaglio, M.; Parrinello, M. Metadynamics Studies of Crystal Nucleation. *IUCrJ* **2015**, *2* (2), 256–266.
- (48) Salvalaglio, M.; Giberti, F.; Parrinello, M. 1,3,5-Tris(4-Bromophenyl)benzene Prenucleation Clusters from Metadynamics. *Acta Crystallogr. Sect. C Struct. Chem.* **2014**, *70* (2), 132–136.
- (49) Bochicchio, D.; Salvalaglio, M.; Pavan, G. M. Into the Dynamics of a Supramolecular Polymer at Submolecular Resolution. *Nat. Commun.* **2017**, *8* (1), 1–11.
- (50) Shevchuk, R. Markov State Model of Ion Assembling Process. *J. Phys. Chem. A* **2016**, *120* (18), 2783–2788.
- (51) Perkett, M. R.; Hagana, M. F. Using Markov State Models to Study Self-Assembly. *J. Chem. Phys.* **2014**, *140* (21).
- (52) Kelley, N. W.; Vishal, V.; Krafft, G. A.; Pande, V. S. Simulating Oligomerization at Experimental Concentrations and Long Timescales: A Markov State Model Approach. *J. Chem. Phys.* **2008**, *129* (21).
- (53) Frederix, P. W. J. M.; Scott, G. G.; Abul-Haija, Y. M.; Kalafatovic, D.; Pappas, C. G.; Javid, N.; Hunt, N. T.; Ulijn, R. V.; Tuttle, T. Exploring the Sequence Space for (Tri-)Peptide Self-Assembly to Design and Discover New Hydrogels. *Nat. Chem.* **2014**, *7* (1), 30–37.
- (54) Moreira, I. P.; Piskorz, T. K.; Esch, J. H. Van; Tuttle, T.; Ulijn, R. V. Biocatalytic Self-Assembly of Tripeptide Gels and Emulsions. **2017**.
- (55) Frederix, P. W. J. M.; Ulijn, R. V.; Hunt, N. T.; Tuttle, T. Virtual Screening for Dipeptide Aggregation: Toward Predictive Tools for Peptide Self-Assembly. **2011**, 2380–2384.
- (56) Velichko, Y. S.; Stupp, S. I.; De La Cruz, M. O. Molecular Simulation Study of Peptide Amphiphile Self-Assembly. *J. Phys. Chem. B* **2008**, *112* (8), 2326–2334.
- (57) Yu, T.; Schatz, G. C. Free-Energy Landscape for Peptide Amphiphile Self-Assembly: Stepwise versus Continuous Assembly Mechanisms. **2013**.
- (58) Fu, I. W.; Markegard, C. B.; Nguyen, H. D. Solvent Effects on Kinetic Mechanisms of Self-Assembly by Peptide Amphiphiles via Molecular Dynamics Simulations. **2015**.
- (59) Bochicchio, D.; Pavan, G. M. From Cooperative Self-Assembly to Water-Soluble Supramolecular Polymers Using Coarse-Grained Simulations. *ACS Nano* **2017**, *11* (1), 1000–1011.
- (60) Marrink, S. J.; Vries, A. H. De; Mark, A. E. Coarse Grained Model for Semiquantitative Lipid Simulations. **2004**, 750–760.
- (61) Bejagam, K. K.; Balasubramanian, S. Supramolecular Polymerization: A Coarse Grained Molecular Dynamics Study. *J. Phys. Chem. B* **2015**, *119* (17), 5738–5746.
- (62) Klein, M. L.; Shinoda, W. Large-Scale Molecular Dynamics Simulations of Self-Assembling Systems. *Science* **2008**, *321*, 798–800.
- (63) Manandhar, A.; Kang, M.; Chakraborty, K.; Tang, P. K.; Loverde, S. M. Molecular Simulations of Peptide Amphiphiles. **2017**, 7993–8005.
- (64) Tuttle, T. Computational Approaches to Understanding the Self-Assembly of Peptide-Based Nanostructures. *Israel Journal of Chemistry*. 2015, pp 724–734.
- (65) Yuan, C.; Li, S.; Zou, Q.; Ren, Y.; Yan, X. Multiscale Simulations for Understanding the Evolution and Mechanism of Hierarchical Peptide Self-Assembly. **2017**, 23614–23631.
- (66) Cordier, P.; Tournilhac, F.; Soulié-Ziakovic, C.; Leibler, L. Self-Healing and Thermoreversible Rubber from Supramolecular Assembly. *Nature* **2008**, *451* (7181),

- 977–980.
- (67) Kulkarni, C.; Bejagam, K. K.; Senanayak, S. P.; Narayan, K. S.; Balasubramanian, S.; George, S. J. Dipole-Moment-Driven Cooperative Supramolecular Polymerization. *J. Am. Chem. Soc.* **2015**, *137* (11), 3924–3932.
- (68) Shinoda, W.; DeVane, R.; Klein, M. L. Coarse-Grained Molecular Modeling of Non-Ionic Surfactant Self-Assembly. *Soft Matter* **2008**, *4* (12), 2454.
- (69) Arefi, H. H.; Yamamoto, T. Communication: Self-Assembly of a Model Supramolecular Polymer Studied by Replica Exchange with Solute Tempering. *J. Chem. Phys.* **2017**, *147* (21).
- (70) Korlepara, D. B. Molecular Modelling of Supramolecular One Dimensional Polymers. **2018**, *1*, 22659–22669.
- (71) Van Gunsteren, W. F.; Bakowies, D.; Baron, R.; Chandrasekhar, I.; Christen, M.; Daura, X.; Gee, P.; Geerke, D. P.; Glättli, A.; Hünenberger, P. H.; et al. Biomolecular Modeling: Goals, Problems, Perspectives. *Angew. Chemie - Int. Ed.* **2006**, *45* (25), 4064–4092.
- (72) Marrink, S. J.; Vries, A. H. De; Mark, A. E. Coarse Grained Model for Semiquantitative Lipid Simulations. *J. Phys. Chem. B* **2004**, *108* (2), 750–760.
- (73) Sasselli, I. R.; Pappas, C. G.; Matthews, E.; Wang, T.; Hunt, N. T.; Ulijn, R. V.; Tuttle, T. Using Experimental and Computational Energy Equilibration to Understand Hierarchical Self-Assembly of Fmoc-Dipeptide Amphiphiles. *Soft Matter* **2016**, *12* (40), 8307–8315.
- (74) Baker, M. B.; Albertazzi, L.; Voets, I. K.; Leenders, C. M. a.; Palmans, A. R. a.; Pavan, G. M.; Meijer, E. W. Consequences of Chirality on the Dynamics of a Water-Soluble Supramolecular Polymer. *Nat. Commun.* **2015**, *6*, 6234.
- (75) van Bommel, K. J. C.; van der Pol, C.; Muizebelt, I.; Friggeri, A.; Heeres, A.; Meetsma, A.; Feringa, B. L.; Esch, J. van. Responsive Cyclohexane-Based Low-Molecular Weight Hydrogelators with Modular Architecture. *Angew. Chem. Int. Ed.* **2004**, *43*, 1663–1667.
- (76) Cantekin, S.; de Greef, T. F. A.; Palmans, A. R. A. Benzene-1,3,5-Tricarboxamide: A Versatile Ordering Moiety for Supramolecular Chemistry. *Chem. Soc. Rev.* **2012**, *41*, 6125–6137.
- (77) Boekhoven, J.; Poolman, J. M.; Maity, C.; Li, F.; van der Mee, L.; Minkenberg, C. B.; Mendes, E.; van Esch, J. H.; Eelkema, R. Catalytic Control over Supramolecular Gel Formation. *Nat. Chem.* **2013**, *5* (May 2013), 433–437.
- (78) Albuquerque, R. Q.; Timme, A.; Kress, R.; Senker, J.; Schmidt, H. W. Theoretical Investigation of Macrodipoles in Supramolecular Columnar Stackings. *Chem. - A Eur. J.* **2013**, *19* (5), 1647–1657.
- (79) Oostenbrink, C.; Villa, A.; Mark, A. E.; Van Gunsteren, W. F. A Biomolecular Force Field Based on the Free Enthalpy of Hydration and Solvation: The GROMOS Force-Field Parameter Sets 53A5 and 53A6. *J. Comput. Chem.* **2004**, *25* (13), 1656–1676.
- (80) Vanommeslaeghe, K.; Hatcher, E.; Acharya, C.; Kundu, S.; Zhong, S.; Shim, J.; Darian, E.; Guvench, O.; Lopes, P.; Vorobyov, I.; et al. CHARMM General Force Field: A Force Field for Drug-Like Molecules Compatible with the CHARMM All-Atom Additive Biological Force Fields. *J. Comput. Chem.* **2009**, *31* (4), 671–690.
- (81) Vanommeslaeghe, K.; MacKerell, A. D. Automation of the CHARMM General Force Field (CGenFF) I: Bond Perception and Atom Typing. *J. Chem. Inf. Model.* **2012**, *52* (12), 3144–3154.
- (82) Vanommeslaeghe, K.; Raman, E. P.; MacKerell, A. D. Automation of the CHARMM General Force Field (CGenFF) II: Assignment of Bonded Parameters and Partial Atomic Charges. *J. Chem. Inf. Model.* **2012**, *52* (12), 3155–3168.
- (83) Yu, W.; Lopes, P. E. M.; Roux, B.; MacKerell, A. D. Six-Site Polarizable Model of Water Based on the Classical Drude Oscillator. *J. Chem. Phys.* **2013**, *138* (3).
- (84) Lamoureux, G.; Harder, E.; Vorobyov, I. V.; Roux, B.; MacKerell, A. D. A Polarizable Model of Water for Molecular Dynamics Simulations of Biomolecules. *Chem. Phys. Lett.* **2006**, *418* (1–3), 245–249.

- (85) Lopes, P. E. M.; Huang, J.; Shim, J.; Luo, Y.; Li, H.; Roux, B.; Mackerell, A. D. Polarizable Force Field for Peptides and Proteins Based on the Classical Drude Oscillator. *J. Chem. Theory Comput.* **2013**, *9* (12), 5430–5449.
- (86) Lemkul, J. A.; Roux, B.; Van Der Spoel, D.; Mackerell, A. D. Implementation of Extended Lagrangian Dynamics in GROMACS for Polarizable Simulations Using the Classical Drude Oscillator Model. *J. Comput. Chem.* **2015**, *36* (19), 1473–1479.
- (87) Marrink, S. J.; Risselada, H. J.; Yefimov, S.; Tieleman, D. P.; de Vries, A. H. The MARTINI Force Field: Coarse Grained Model for Biomolecular Simulations. *J. Phys. Chem. B* **2007**, *111* (27), 7812–7824.
- (88) Marrink, S. J.; Tieleman, D. P. Perspective on the Martini Model. *Chem. Soc. Rev.* **2013**, *42* (16), 6801–6822.
- (89) De Jong, D. H.; Singh, G.; Bennett, W. F. D.; Arnarez, C.; Wassenaar, T. A.; Schäfer, L. V.; Periole, X.; Tieleman, D. P.; Marrink, S. J. Improved Parameters for the Martini Coarse-Grained Protein Force Field. *J. Chem. Theory Comput.* **2013**, *9*, 687–697.
- (90) Dill, K. a; Chan, H. S. From Levinthal to Pathways to Funnels. *Nat. Struct. Biol.* **1997**, *4* (1), 10–19.
- (91) Abraham, M. J.; Murtola, T.; Schulz, R.; Pall, S.; Smith, J. C.; Hess, B.; Lindahl, E. Gromacs: High Performance Molecular Simulations through Multi-Level Parallelism from Laptops to Supercomputers. *SoftwareX* **2015**, *1–2*, 19–25.
- (92) Malde, A. K.; Zuo, L.; Breeze, M.; Stroet, M.; Poger, D.; Nair, P. C.; Oostenbrink, C.; Mark, A. E. An Automated Force Field Topology Builder (ATB) and Repository: Version 1.0. *J. Chem. Theory Comput.* **2011**, *7* (12), 4026–4037.
- (93) Thole, B. T.; van Duijnen, P. T. A General Population Analysis Preserving the Dipole Moment. *Theor. Chim. Acta* **1983**, *63* (3), 209–221.
- (94) Bayly, C. I.; Cieplak, P.; Cornell, W. D.; Kollman, P. A. A Well-Behaved Electrostatic Potential Based Method Using Charge Restraints for Deriving Atomic Charges: The RESP Model. *J. Phys. Chem.* **1993**, *97* (40), 10269–10280.
- (95) Wolfram Research, I. Mathematica. Wolfram Research, Inc.: Champaign, Illinois.
- (96) Binder, W. H. The Past 40 Years of Macromolecular Sciences: Reflections on Challenges in Synthetic Polymer and Material Science. *Macromol. Rapid Commun.* **2018**, *1800610*, 1–7.
- (97) Li, J. L.; Liu, X. Y. *Soft Fibrillar Materials : Fabrication and Applications*; John Wiley & Sons, Incorporated, 2013.
- (98) De Greef, T. F. A.; Schenning, A. P. H. J.; George, S. J.; Korevaar, P. A.; Smulders, M. M. J.; Markvoort, A. J.; Meijer, E. W.; Hilbers, P. A. J. Pathway Complexity in Supramolecular Polymerization. *Nature* **2012**, *481* (7382), 492–496.
- (99) Poolman, J. M.; Maity, C.; Boekhoven, J.; Mee, L. Van Der; Sage, V. A. A.; Groenewold, G. J. M.; Kasteren, S. I. Van; Versluis, F.; Esch, J. H. Van; Eelkema, R. A. Toolbox for Controlling the Properties and Functionalisation of Hydrazone-Based Supramolecular Hydrogels. *J. Mater. Chem. B* **2016**, *4*, 852–858.
- (100) Albertazzi, L.; Zwaag, D. Van Der; Leenders, C. M. A.; Fitzner, R.; Hofstad, R. W. Van Der; Meijer, E. W. Probing Exchange Pathways in One-Dimensional Aggregates with Super-Resolution Microscopy. *Science* **2014**, *344* (May), 491–495.
- (101) Doerr, S.; Harvey, M. J.; Noé, F.; De Fabritiis, G. HTMD: High-Throughput Molecular Dynamics for Molecular Discovery. *J. Chem. Theory Comput.* **2016**, *12* (4), 1845–1852.
- (102) Scherer, M. K.; Trendelkamp-Schroer, B.; Paul, F.; Perez-Hernandez, G.; Hoffmann, M.; Plattner, N.; Wehmeyer, C.; Prinz, J. H.; Noe, F. PyEMMA 2: A Software Package for Estimation, Validation, and Analysis of Markov Models. *J. Chem. Theory Comput.* **2015**, *11* (11), 5525–5542.
- (103) Husic, B. E.; Pande, V. S. Markov State Models: From an Art to a Science. *J. Am. Chem. Soc.* **2018**, *140*, 2386–2396.
- (104) Piskorz, T. K.; Vries, A. H. de; Esch, J. H. van. Manuscript in Preparation.
- (105) Cohen, S. I. A.; Vendruscolo, M.; Welland, M. E.; Dobson, C. M.; Terentjev, E. M.;

- Knowles, T. P. J. Nucleated Polymerization with Secondary Pathways. I. Time Evolution of the Principal Moments. *J. Chem. Phys.* **2011**, *135* (6), 65105.
- (106) Cohen, S. I. A.; Vendruscolo, M.; Dobson, C. M.; Knowles, T. P. J. From Macroscopic Measurements to Microscopic Mechanisms of Protein Aggregation. *J. Mol. Biol.* **2012**, *421* (2–3), 160–171.
- (107) Kaplan, E. L.; Meier, P. Nonparametric Estimation from Incomplete Observations. *J. Am. Stat. Assoc.* **1958**, *53* (282), 4457–4481.
- (108) Frederix, P. W. J. M. Personal Communication.
- (109) Pereira Oliveira, M.; Schmidt, H. W.; Queiroz Albuquerque, R. Unveiling the Role of Macrodipolar Interactions in the Properties of Self-Assembled Supramolecular Materials. *Chem. - A Eur. J.* **2018**, *24*, 2609–2617.
- (110) Kroon, J.; Kanters, J. A.; van Duijneveldt-van De Rijdt, J. G. C. M.; van Duijneveldt, F. B.; Vliegthart, J. A. O-H · O Hydrogen Bonds in Molecular Crystals a Statistical and Quantum-Chemical Analysis. *J. Mol. Struct.* **1975**, *24* (1), 109–129.
- (111) Hecht, S. Welding, Organizing, and Planting Organic Molecules on Substrate Surfaces - Promising Approaches towards Nanoarchitectonics from the Bottom up. *Angew. Chemie - Int. Ed.* **2003**, *42* (1), 24–26.
- (112) Joachim, C.; Gimzewski, J. K.; Aviram, A. Electronics Using Hybrid-Molecular and Mono-Molecular Devices. *Nature* **2000**, *408* (6812), 541–548.
- (113) Kühnle, A. Self-Assembly of Organic Molecules at Metal Surfaces. *Curr. Opin. Colloid Interface Sci.* **2009**, *14* (2), 157–168.
- (114) Reed, M. A.; Zhou, C.; Muller, C. J.; Burgin, T. P.; Tour, J. M. Conductance of a Molecular Junction. *Science* **1997**, *278* (5336), 252–254.
- (115) Palermo, V.; Samorì, P. Molecular Self-Assembly across Multiple Length Scales. *Angew. Chemie - Int. Ed.* **2007**, *46* (24), 4428–4432.
- (116) Gutzler, R.; Cardenas, L.; Rosei, F. Kinetics and Thermodynamics in Surface-Confined Molecular Self-Assembly. *Chem. Sci.* **2011**, *2* (12), 2290–2300.
- (117) Goronzy, D.; Ebrahimi, M.; Rosei, F.; Arramel; Fang, Y.; Wee, A. T. S.; De Feyter, S.; Tait, S. L.; Wang, C.; Beton, P. H.; et al. Supramolecular Assemblies on Surfaces: Nanopatterning, Functionality, and Reactivity. *ACS Nano* **2018**, *12* (8), 7445–7481.
- (118) Barth, J. V.; Costantini, G.; Kern, K. Engineering Atomic and Molecular Nanostructures at Surfaces. *Nature* **2005**, *437* (7059), 671–679.
- (119) Yokoyama, S.; Hirose, T.; Matsuda, K. Phototriggered Formation and Disappearance of Surface-Confined Self-Assembly Composed of Photochromic 2-Thienyl-Type Diarylethene: A Cooperative Model at the Liquid/solid Interface. *Chem. Commun.* **2014**, *50* (45), 5964–5966.
- (120) Yokoyama, S.; Hirose, T.; Matsuda, K. Effects of Alkyl Chain Length and Hydrogen Bonds on the Cooperative Self-Assembly of 2-Thienyl-Type Diarylethenes at a Liquid/Highly Oriented Pyrolytic Graphite (HOPG) Interface. *Chem. - A Eur. J.* **2015**, *21* (39), 13569–13576.
- (121) Frath, D.; Sakano, T.; Imaizumi, Y.; Yokoyama, S.; Hirose, T.; Matsuda, K. Diarylethene Self-Assembled Monolayers: Cocrystallization and Mixing-Induced Cooperativity Highlighted by Scanning Tunneling Microscopy at the Liquid/Solid Interface. *Chem. - A Eur. J.* **2015**, *21* (32), 11350–11358.
- (122) Nishitani, N.; Hirose, T.; Matsuda, K. Investigation on the Surface-Confined Self-Assembly Stabilized by Hydrogen Bonds of Urea and Amide Groups: Quantitative Analysis of Concentration Dependence of Surface Coverage. *Chem. - An Asian J.* **2015**, *10* (9), 1926–1931.
- (123) Nishitani, N.; Hirose, T.; Matsuda, K. Influence of Multidirectional Interactions on Domain Size and Shape of 2-D Molecular Assemblies. *Langmuir* **2017**, *33* (36), 9151–9159.
- (124) Maeda, N.; Hirose, T.; Yokoyama, S.; Matsuda, K. Rational Design of Highly Photoresponsive Surface-Confined Self-Assembly of Diarylethenes: Reversible Three-

- State Photoswitching at the Liquid/solid Interface. *J. Phys. Chem. C* **2016**, *120* (17), 9317–9325.
- (125) Coenen, M. J. J.; Cremers, M.; Den Boer, D.; Van Den Bruele, F. J.; Khoury, T.; Santic, M.; Crossley, M. J.; Van Enkevort, W. J. P.; Hendriksen, B. L. M.; Elemans, J. A. A. W.; et al. Little Exchange at the Liquid/solid Interface: Defect-Mediated Equilibration of Physisorbed Porphyrin Monolayers. *Chem. Commun.* **2011**, *47* (34), 9666–9668.
- (126) Silly, F.; Weber, U. K.; Shaw, A. Q.; Burlakov, V. M.; Castell, M. R.; Briggs, G. A. D.; Pettifor, D. G. Deriving Molecular Bonding from a Macromolecular Self-Assembly Using Kinetic Monte Carlo Simulations. *Phys. Rev. B - Condens. Matter Mater. Phys.* **2008**, *77* (20), 2–5.
- (127) Weber, U. K.; Burlakov, V. M.; Perdigao, L. M. A.; Fawcett, R. H. J.; Beton, P. H.; Champness, N. R.; Jefferson, J. H.; Briggs, G. A. D.; Pettifor, D. G. Role of Interaction Anisotropy in the Formation and Stability of Molecular Templates. *Phys. Rev. Lett.* **2008**, *100* (15), 1–4.
- (128) Rohr, C.; Balbas Gamba, M.; Gruber, K.; Constable, E. C.; Frey, E.; Franosch, T.; Hermann, B. A. Molecular Jigsaw: Pattern Diversity Encoded by Elementary Geometrical Features. *Nano Lett.* **2010**, *10* (3), 833–837.
- (129) Lei, S.; Tahara, K.; Müllen, K.; Szabelski, P.; Tobe, Y.; De Feyter, S. Mixing Behavior of Alkoxyated Dehydrobenzo [12]annulenes at the Solid-Liquid Interface: Scanning Tunneling Microscopy and Monte Carlo Simulations. *ACS Nano* **2011**, *5* (5), 4145–4157.
- (130) Calmettes, B.; Estrampes, N.; Coudret, C.; Roussel, T. J.; Faraudo, J.; Coratger, R. Observation and Modeling of Conformational Molecular Structures Driving the Self-Assembly of Tri-Adamantyl Benzene on Ag(111). *Phys. Chem. Chem. Phys.* **2016**, *18*, 20281–20289.
- (131) Szabelski, P.; De Feyter, S. Chiral Occlusion in Two-Dimensional Binary Supramolecular Networks Studied by the Monte Carlo Method. *CrystEngComm* **2011**, *13*, 5542–5550.
- (132) Martsinovich, N.; Troisi, A. Modeling the Self-Assembly of Benzenedicarboxylic Acids Using Monte Carlo and Molecular Dynamics Simulations. *J. Phys. Chem. C* **2010**, *114* (10), 4376–4388.
- (133) Chen, J.; Zhu, E.; Liu, J.; Zhang, S.; Lin, Z.; Duan, X.; Heinz, H.; Huang, Y.; De Yoreo, J. J. Building Two-Dimensional Materials One Row at a Time: Avoiding the Nucleation Barrier. *Science* **2018**, *362* (6419), 1135–1139.
- (134) Srinivas, G.; Nielsen, S. O.; Moore, P. B.; Klein, M. L. Molecular Dynamics Simulations of Surfactant Self-Organization at a Solid-Liquid Interface. *J. Am. Chem. Soc.* **2006**, *128* (3), 848–853.
- (135) Sun, H.; Yang, X. Molecular Simulation of Self-Assembly Structure and Interfacial Interaction for SDBS Adsorption on Graphene. *Colloids Surfaces A Physicochem. Eng. Asp.* **2014**, *462*, 82–89.
- (136) Wu, D.; Yang, X. Coarse-Grained Molecular Simulation of Self-Assembly for Nonionic Surfactants on Graphene Nanostructures. *J. Phys. Chem. B* **2012**, *116*, 12048–12056.
- (137) Wu, B.; Yang, X. Molecular Simulation of Electrolyte-Induced Interfacial Interaction between SDS/Graphene Assemblies. *J. Phys. Chem. C* **2013**, *117* (44), 23216–23223.
- (138) De Feyter, S.; De Schryver, F. C. Two-Dimensional Supramolecular Self-Assembly Probed by Scanning Tunneling Microscopy. *Chem. Soc. Rev.* **2003**, *32* (3), 139–150.
- (139) Rabe, J. P.; Buchholz, S. Commensurability and Mobility in Two-Dimensional Molecular Patterns on Graphite. *Science* **1991**, *253* (5018), 424–427.
- (140) McGonigal, G. C.; Bernhardt, R. H.; Thomson, D. J. Imaging Alkane Layers at the Liquid/graphite Interface with the Scanning Tunneling Microscope. *Appl. Phys. Lett.* **1990**, *57* (1), 28–30.
- (141) De Feyter, S.; Grim, P. C. M.; Van Esch, J. H.; Kellogg, R. M.; Feringa, B. L.; De Schryver, F. C. Nontrivial Differentiation between Two Identical Functionalities within the Same Molecule Studied by STM. *J. Phys. Chem. B* **1998**, *102* (45), 8981–8987.

- (142) Gobbo, C.; Beurroies, I.; De Ridder, D.; Elkema, R.; Marrink, S. J.; De Feyter, S.; Van Esch, J. H.; De Vries, A. H. MARTINI Model for Physisorption of Organic Molecules on Graphite. *J. Phys. Chem. C* **2013**, *117* (30), 15623–15631.
- (143) Askadskaya, L.; Rabe, J. P. Anisotropic Molecular Dynamics in the Vicinity of Order-Disorder Transitions in Organic Monolayers. *Phys. Rev. Lett.* **1992**, *69* (9), 1395–1398.
- (144) Buchholz, S.; Rabe, J. P. Molecular Imaging of Alkanol Monolayers on Graphite. *Angew. Chem. Int. Ed.* **1992**, *31* (2), 189–191.
- (145) Lim, R.; Li, J.; Li, S. F. Y.; Feng, Z.; Valiyaveetil, S. Formation of Two-Dimensional Supramolecular Chiral Lamellae by Diamide Molecules at the Solution/graphite Interface: A Scanning Tunneling Microscopy Study. *Langmuir* **2000**, *16* (17), 7023–7030.
- (146) Zou, B.; Dreger, K.; Mück-Lichtenfeld, C.; Grimme, S.; Schäfer, H. J.; Fuchs, H.; Chi, L. Simple and Complex Lattices of N-Alkyl Fatty Acid Amides on a Highly Oriented Pyrolytic Graphite Surface. *Langmuir* **2005**, *21* (4), 1364–1370.
- (147) Ciesielski, A.; Palma, C. A.; Bonini, M.; Samori, P. Towards Supramolecular Engineering of Functional Nanomaterials: Pre-Programming Multi-Component 2D Self-Assembly at Solid-Liquid Interfaces. *Adv. Mater.* **2010**, *22* (32), 3506–3520.
- (148) Padowitz, D. F.; Sada, D. M.; Kemer, E. L.; Dougan, M. L.; Andrew Xue, W. Molecular Tracer Dynamics in Crystalline Organic Films at the Solid-Liquid Interface. *J. Phys. Chem. B* **2002**, *106* (3), 593–598.
- (149) Lackinger, M.; Griessel, S.; Kampschulte, L.; Jamitzky, F.; Heckl, W. M. Dynamics of Grain Boundaries in Two-Dimensional Hydrogen-Bonded Molecular Networks. *Small* **2005**, *1* (5), 532–539.
- (150) Wang, F.; Richards, V. N.; Shields, S. P.; Buhro, W. E. Kinetics and Mechanisms of Aggregative Nanocrystal Growth. *Chem. Mater.* **2014**, *26* (1), 5–21.
- (151) Stabel, A.; Heinz, R.; De Schryver, F. C.; Rabe, J. P. Ostwald Ripening of Two-Dimensional Crystals at the Solid-Liquid Interface. *J. Phys. Chem.* **1995**, *99* (2), 505–507.
- (152) Piskorz, T. K.; de Vries, A. H.; De Feyter, S.; van Esch, J. H. Mechanism of Ostwald Ripening in 2D Physisorbed Assemblies at Molecular Time and Length Scale by Molecular Dynamics Simulations. *J. Phys. Chem. C* **2018**, *122*, 24380–24385.
- (153) Dannenberger, O.; Buck, M.; Grunze, M. Self-Assembly of N-Alkanethiols: A Kinetic Study by Second Harmonic Generation. *J. Phys. Chem. B* **1999**, *103*, 2202–2213.
- (154) Schreiber, F. Structure and Growth of Self-Assembling Monolayers. *Prog. Surf. Sci.* **2000**, *65* (5–8), 151–256.
- (155) Richards, V. N.; Rath, N. P.; Buhro, W. E. Pathway from a Molecular Precursor to Silver Nanoparticles: The Prominent Role of Aggregative Growth. *Chem. Mater.* **2010**, *22* (11), 3556–3567.
- (156) Stabel, A.; Heinz, R.; Rabe, J. P.; Wegner, G.; De Schryver, F. C.; Corens, D.; Dehaen, W.; Sueling, C. STM Investigation of 2D Crystals of Anthrone Derivatives on Graphite: Analysis of Molecular Structure and Dynamics. *J. Phys. Chem.* **1995**, *99* (21), 8690–8697.
- (157) Sengupta, U.; Carballo-Pacheco, M.; Strodel, B. Automated Markov State Models for Molecular Dynamics Simulations of Aggregation and Self-Assembly. *J. Chem. Phys.* **2019**, *150*, 115101.
- (158) Mukhtyar, A. J.; Escobedo, F. A. Developing Local Order Parameters for Order – Disorder Transitions From Particles to Block Copolymers : Methodological Framework. *Macromolecules* **2018**, *51*, 9769–9780.
- (159) Fang, Y.; Ghijsens, E.; Ivasenko, O.; Cao, H.; Noguchi, A.; Mali, K. S.; Tahara, K.; Tobe, Y.; De Feyter, S. Dynamic Control over Supramolecular Handedness by Selecting Chiral Induction Pathways at the Solution-Solid Interface. *Nat. Chem.* **2016**, *8* (7), 711–717.
- (160) Schwartz, D. K. Mechanism and Kinetics of Self-Assembled Monolayer Formation.

- Annu. Rev. Phys. Chem.* **2001**, *52*, 107–137.
- (161) Dietrich, H.; Schmaltz, T.; Halik, M.; Zahn, D. Molecular Dynamics Simulations of Phosphonic Acid-Aluminum Oxide Self-Organization and Their Evolution into Ordered Monolayers. *Phys. Chem. Chem. Phys.* **2017**, *19* (7), 5137–5144.
- (162) Alessandri, R.; Uusitalo, J. J.; De Vries, A. H.; Havenith, R. W. A.; Marrink, S. J. Bulk Heterojunction Morphologies with Atomistic Resolution from Coarse-Grain Solvent Evaporation Simulations. *J. Am. Chem. Soc.* **2017**, *139* (10), 3697–3705.
- (163) Frederix, P. W. J. M.; Patmanidis, I.; Marrink, S. J. Molecular Simulations of Self-Assembling Bio-Inspired Supramolecular Systems and Their Connection to Experiments. *Chem. Soc. Rev.* **2018**, *47* (10), 3470–3489.
- (164) Couallier, E.; Riaublanc, A.; David Briand, E.; Rousseau, B. Molecular Simulation of the Water-Triolein-Oleic Acid Mixture: Local Structure and Thermodynamic Properties. *J. Chem. Phys.* **2018**, *148* (18).
- (165) Pronk, S.; Pall, S.; Schulz, R.; Larsson, P.; Bjelkmar, P.; Apostolov, R.; Shirts, M. R.; Smith, J. C.; Kasson, P. M.; Van Der Spoel, D.; et al. GROMACS 4.5: A High-Throughput and Highly Parallel Open Source Molecular Simulation Toolkit. *Bioinformatics* **2013**, *29* (7), 845–854.
- (166) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. Molecular Dynamics with Coupling to an External Bath. *J. Chem. Phys.* **1984**, *81* (8), 3684–3690.
- (167) Gowers, R. J.; Linke, M.; Barnoud, J.; Reddy, T. J. E.; Melo, M. N.; Seyler, S. L.; Domański, J.; Dotson, D. L.; Buchoux, S.; Kenney, I. M.; et al. MDAnalysis: A Python Package for the Rapid Analysis of Molecular Dynamics Simulations. In *Proceedings of the 15th Python in Science Conference*; Benthall, S., Rostrup, S., Eds.; 2016; pp 98–105.
- (168) Michaud-Agrawal, N.; Denning, E. J.; Woolf, T. B.; Beckstein, O. MDAnalysis: A Toolkit for the Analysis of Molecular Dynamics Simulations. *J. Comput. Chem.* **2011**, *32* (10), 2319–2327.
- (169) Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual Molecular Dynamics. *J. Mol. Graph.* **1996**, *14* (1), 33–38.
- (170) Jones, E.; Oliphant, T.; Peterson, P. {SciPy}: Open Source Scientific Tools for {Python}.
- (171) Gray, F. W.; Irving J. Krems. Distribution of Para and Ortho Isomers in Some Model Long Chain Alkylbenzenesulfonates. *J. Org. Chem.* **1961**, *26* (1), 209–212.
- (172) Schiessler, R. W.; Rytina, C. H.; Weisel, Fischl, F. L., JicLaughlin., R. L.; Keuhner, H. H. No Title. *Proc. Am. Pet. Inst.* **1946**, *26*.
- (173) Tschamler, H. Uber Binare Flussige Mischungen .1. Mischungswarmen, Volumseffekte Und Zustandsdiagramme von Chlorex Mit Benzol Und Normal-Alkylbenzolon. *Monatshefte fur chemie* **1948**, *79* (2), 162–177.
- (174) Hong, E.; Lee, K.; Wenzel, W. RMSD Computation for Clusters of Identical Particles. *Int. J. Biol. Biomed. Eng.* **2007**, *1* (2), 50–52.
- (175) Metzger, R. M. Unimolecular Electronics and Rectifiers. *Synth. Met.* **2009**, *159* (21–22), 2277–2281.
- (176) Samorí, P.; Müllen, K.; Rabe, J. P. Molecular-Scale Tracking of the Self-Healing of Polycrystalline Monolayers at the Solid-Liquid Interface. *Adv. Mater.* **2004**, *16* (19), 1761–1765.
- (177) Piskorz, T. K.; Gobbo, C.; Marrink, S.-J.; Feyter, S. de; Vries, A. H. de; Esch, J. H. van. Manuscript in Preparation.
- (178) Marrink, S.-J.; Periole, X. *Biomolecular Simulations*; Monticelli, L., Salonen, E., Eds.; Methods in Molecular Biology; Humana Press: Totowa, NJ, 2013; Vol. 924.
- (179) Kolesnik, A. D.; Ratanov, N. *Telegraph Processes and Option Pricing*; Springer, 2013.

# Acknowledgements

Every journey has an end. This one took over 5 years, countless hours in front of the screens and (if I estimated correctly) one thousand liters of coffee. This would not be possible if not for the help and support of supervisors, colleagues, friends, and family. I would like to thank all of the people who accompany me on this journey.

Firstly, I would like to thank Jan. Thank you for giving me the opportunity to join your lab and the SMARTNET project. Discussions with you were always inspiring, and I have learned a lot from you. Thank you for all the guidance. Alex, completing my Ph.D. would not be possible without you. Thank you for the countless hours which you spend on discussions and over the manuscripts. Many times during my journey, I doubted about this path, but I made it through because of your support. Ger, thank you for all the discussions which we had. I learned from you that one can explain the most complex problems just by basic thermodynamics laws. Eduardo, your teaching and support is something which a lot of people can learn from you. Steven Picken, thank you for the discussions which we had. Steven de Feyter thank you for all the help with self-assembly on graphite research. Rienk, although we did not have so much direct contact, your lectures were always inspiring.

My journey would not be possible if not for the help of my colleagues. Great science comes most of the time of discussions in offices, labs over the lunch a beer. Qian, I will miss the countless talks about science (and not only) which we had over our desk late in the evenings. I have learned a lot from you. I am sure you will soon become a professor. Serhii, my close friend from the very first day of my Ph.D., until the very end (and longer). I will miss all our trips and countless coffee breaks. Matija, Matiga, Matidza, we had a lot of fun adventures, from you chasing me in Bordeaux, expelling ghosts, or eating all all-you-can-eat diners. I am impressed by how well you are organized on many levels (personal and professional) of your life. Fahimeh, thank you for all the discussions which we had and all the “board-games” which always turned in delicious dinner parties. Susan, I feel so comfortable next to you, and I really like all the ridiculous discussions which we had. Thank you for your help in translating parts of my thesis. Hendrik thanks for all the trips, dinner and jokes which have done together. Bowen, my boardgame buddy, thanks for the great meetings. Good luck with learning the instrument. Jos, you always enlighten the

mood of the whole group. I miss all your bad jokes. Emma, my longest officemate. I loved all the discussions which we had. I am happy that in the new office we were a little bit separated and could not discuss so much: otherwise, we would have never finished our studies. Vasu, your ability to sell anything to anyone is fantastic (especially convincing me every year on 1st of April that you are getting married). Thanks for the help with cryo-TEM. Roman, I have so many memories with you during the first years of my Ph.D.: all parties which we had and the trip to Georgia. Alexandra, our office was so happy with you. Łukasz, dzięki za wszystkie dyskusje i powodzenia w karierze naukowej. Benjamin and Tobias, thank you for all the parties, and to taking over the borrel responsibility. I like that you managed to make our group great again. Mark, you are fantastic officemate, always brighten the mood in our office. Kai, thank you for being so cheerful and warm. Cansel, I believe I have never seen you without a smile. Thank you. Yiming, thanks for all the scientific diffusions which we had. Elmira, thank you for the friendship. Marcel, thank you for the countless discussions and the ice cream made with liquid nitrogen. Ben, conversations with you were always a pleasure. Good luck with playing on saxophone. Michelle, Wouter, Lars, Frank, Vincent, Karolis, Chandan, Sander, Peggy, Fan, Guotai, Reece, Saahil, Suellen, Irene and Piotr, it was really a pleasure to work with you. All the best! Veby, thank you for helping me with all administrative work, especially at the end of my journey.

Due to close collaboration with Alex, I have been a frequent visitor of the MD group at the University of Groningen. Siewert-Jan, thank you for all the feedback and discussions. Pim, thanks for the guidance, especially at the early stage of my Ph.D. Riccardo, thank you for organizing DMD day together with me. Manu, thanks for introducing me to MDAnalysis. Ilias, thanks for all discussions about simulations of self-assembly. Ignacio, Carsten, Sebastian, Paulo, Peter and Jonathan good luck, and thank you for making my visits so pleasant!

I would like to thank my colleagues from SMARTNET projects. All the progress meetings and conferences were genuinely inspiring. Ana Catarina, Jorge, Kars, Laura, Marco, Nishant, Philip, and Vânia thank you for that, and good luck in the future! Due to this project, I had the opportunity to work in Glasgow and Lyon. I have spent in Glasgow almost half a year. Tell, thank you for this opportunity and your support. I have learned a lot from you. Especially, I am impressed by how efficiently you lead the group and how some of the most

boring tasks you turn into exciting exercise. Iván, your enthusiasm, and curiosity are really amazing. I believe that one of the most important lessons during my Ph.D. (about the attitude towards science) I have learned from you. Inês, thank you for your warm welcome and that we were able to work together and share our experiences. I would like to thank also all the rest from Glasgow Family: Flo, Ana, Miguel, Mari, Dan, and Davide. It was a pleasure to spend this half a year with you. Anthony, thank you for giving me the opportunity to work with you in Solvey. Francesco, Aghate, Jessica, and Rita, thank you for being my colleague during my stay in Lyon.

Although Ph.D. life is mostly about science, it would be not possible for me to finish it if not for a few hobbies which kept my mind always in a good mood. When I started my Ph.D. journey, I have not imagined that I would run half-marathon, or be an enthusiast of beach volleyball, or (what is the most incredible) give a music concert.

Fanny, your innocent curiosity makes you one of the most inspiring people I know. All the sports, traveling and music which you have done during those years were inspiring. I am happy that we shared so many of those activities.

Steffen, thank you for all the music sessions which we had together. I really miss them. Angie, you were the first person with whom I played music, despite it wasn't the Greatest Song in the World (it was just a tribute). We also had other adventures, but I will, for sure, remember you for having many 'interesting' food combinations with peanut butter. Audrey, thank you for all the music sessions which we had (your ukulele skills are amazing) and for the most delicious crepe which I have eaten in my life. Mattijs, thank you for all the singing! I will always be amazed at how magically instruments appear in your hand. Jorik, despite all the stage fright we had, I am happy we managed to give our first concert. Robin, Claudia, Domas thank you for being part of our little band. Erik, thank you for all the music lessons.

Tim and Dominik, thank you for all the trips which we had together (and will have). In particular, the time which we had in Iceland (and probably the most expensive pasta with peso) and surfing. Fruzi, we had some many incredible runs together. When I bike to Rotterdam, I am amazed that we managed to run all this way (and back!) in one go. Even more, if someone told me that one day I will be learning cross-ski country in Slovakia with Hungarian

triathlon team, I would think that this is a joke. Giorgio, Nate, and Anirudh thanks for all the volleyball machetes which we had. Giorgio, your enthusiasm is so contagious! Elena, my sports adventures started (quite modestly) with you. Thank you! Xiaoli, thanks for being an amazing flat mate.

I would like to thank all board member of YoungDelft, which I have been part of. Anne, Clair, Hamid, Anna, Chris, Manas, Rishabh, Alessandro, Maedeh and Vincent thanks for everything and good luck in making life in Delft more exciting!

Na końcu chciałbym podziękować moim polskim przyjaciołom. KSAŚ, dzięki za wszystkie, niekończące się rozmowy i wsparcie. Krzysiu i Basiu, dziękuję, że pomimo naszej odległości nasza przyjaźń jest tylko mocniejsza. Adamie i Ulu, dziękuję za waszą przyjaźń. Anitko, dziękuję za wszystkie rozmowy i przygody. Krsztofie, dzięki za wszystkie spotkania.

A na samym końcu chciałbym podziękować mojej rodzinie. Mamo, tato, dziękuję za Waszą cierpliwość, że przez te wszystkie lata mnie wspieraliście. Babciu, babciu, dziadziusiu i ciociu dziękuję za wsparcie. Krzysiu, Gosiu i Adamie, dziękuję, że zawsze mogę na Was liczyć.

## About the author

Tomasz K. Piskorz was born 17th of June, 1989 in Kraków, Poland. He obtained his B.Sc. and M.Sc. degree in chemistry from the University of Warsaw in 2011 and 2013, respectively. He did his master project in Institute of Physical Chemistry of Polish Academy of Science under supervision of dr. Anna Ochab-Marcinek, where he studied diffusion in a crowded environment using simulations of fluorescent correlation spectroscopy. In 2014, he started Ph.D. in Advanced Soft Matter group at Delft University of Technology under supervision of prof. Jan van Esch and dr. Alex de Vries. During his Ph.D. he had the opportunity to go for secondments to the group of prof. Tuttle at the University of Strathclyde, where he worked on simulations of peptide-based supramolecular structures, and to Solvay, where he worked on multiscale simulations of polymers. The focus of his research was the application of computer simulations into providing insights into supramolecular structures. This research resulted in the thesis which you are holding.

## List of publications

T.K. Piskorz, C. Gobbo, S.J. Marrink, S. De Feyter, A.H. De Vries, J.H. van Esch. Nucleation Mechanisms of Self-Assembled Physisorbed Monolayers on Graphite. *J. Phys. Chem. C*, 2019, 123, 28, 17510-17520.

T.K. Piskorz, A.H. de Vries, S. De Feyter, J.H. van Esch, Mechanism of Ostwald Ripening in 2D Physisorbed Assemblies at Molecular Time and Length Scale by Molecular Dynamics Simulations, *J. Phys. Chem. C*, 2018, 122, 42, 24380-24385.

I.P. Moreira, T.K. Piskorz, J.H. van Esch, T. Tuttle, R.V. Ulijn, Biocatalytic Self-Assembly of Tripeptide Gels and Emulsions, *Langmuir*, 2017, 33, 20, 4986-4995.

T.K. Piskorz, A. Ochab-Marcinek, A Universal Model of Restricted Diffusion for Fluorescence Correlation Spectroscopy, *J. Phys. Chem. B*, 2014, 118, 18, 4906-4912.





