# Geometrically Frustrated Self-Assembly on a Husimi Tree

Lukas Kalvoda



# Geometrically Frustrated Self-Assembly on a Husimi Tree

Master's Thesis

submitted to the

Faculty of Physics Ludwig-Maximilians-Universität München

and the

TUM School of Natural Sciences Technische Universität München

for the degree of

# Master of Science

by

# Lukas Kalvoda

born in Dresden, Germany



LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN



First referee: Prof. Dr. Martin Lenz Second referee: Prof. Dr. Erwin Frey Day of oral examination: July 5, 2024

# Zusammenfassung

Wenn sich einzelne Komponenten autonom zu größeren Strukturen zusammensetzen, spricht man von Selbstassemblierung. Dieser Prozess ist in der Natur allgegenwärtig. In lebenden Zellen führt die Selbstassemblierung von Proteinen beispielsweise zur Bildung großer funktionaler Komplexe wie dem Ribosom. Im Gegensatz dazu geht ein Mangel an Qualitätskontrolle von Proteinen mit schlanken, pathologischen Fibern in zahlreichen Krankheiten wie Alzheimer, Typ-2-Diabetes und Sichelzellenanämie einher. Es ist weitgehend unbekannt, warum komplexe Komponenten über verschiedene molekulare Kontexte hinweg wohldefinierte Aggregate formen.

Geometrische Frustration könnte ein physikalisches Prinzip der Selbstassemblierung sein, das schlanke Aggregate begünstigt. Sie tritt auf, wenn anisotrope Teilchen aufgrund geometrischer Zwänge nicht gleichzeitig günstige Wechselwirkungen mit all ihren Bindungspartnern eingehen können. Eine frühere numerische Studie deutet darauf hin, dass die Rolle der geometrischen Frustration auf der kleinskaligen Teilchenebene am bedeutendsten sein könnte. Bislang gibt es keine analytische Grundlage für diese Vermutung. Daher untersuchen wir in dieser Arbeit die kleinskaligen Auswirkungen geometrischer Frustration. Dafür entwerfen wir ein analytisches Gittergasmodell auf einem baumartigen Graphen, dem Husimi-Baum aus Dreiecken. Auf diesem Gitter interagieren Teilchen anisotrop mit ihren Nachbarn und ahmen so die Komplexität und die geometrischen Zwänge der Selbstassemblierung von Proteinen nach. Wir machen uns die baumartige Struktur des Gitters zunutze, um exakte Rekursionsbeziehungen für die großkanonische Zustandssumme einzelner Aggregate abzuleiten. In unserem Modell treten verschiedene Formen von Aggregaten auf, darunter Fibern und Kristalle. Diese resultieren aus der Konkurrenz vieler verschiedener Wechselwirkungen. Um zu beurteilen, ob Aggregate häufig wohldefinierte Formen aufweisen, formalisieren wir ihre strukturellen Merkmale. Dazu nutzen wir ihre Dicke, wie verzweigt sie sind und ob sie geordnet sind oder nicht. Trotz der komplexen Wechselwirkungen unserer anisotropen Teilchen entstehen mit zunehmender Anisotropie immer mehr wohldefinierte geordnete Aggregate. Im Regime großer Anisotropie begünstigt geometrische Frustration schlankere Aggregate wie Fibern und Fibernetzwerke, während ihr Ausbleiben für geordnete, raumfüllende Strukturen wie Kristalle charakteristisch ist.

Diese Masterarbeit unterstreicht geometrische Frustration auf der Teilchenebene als potenzielles Prinzip der Selbstassemblierung proteinartiger Teilchen, welches die auftretenden Formen kontextübergreifend beeinflusst. Darüber hinaus könnte unser Formalismus das Design von synthetischen Komponenten für die Selbstassemblierung gewünschter Formen erleichtern.

# Abstract

Self-assembly, the process by which individual components autonomously arrange into larger-scale structures, is ubiquitous. For instance, living cells rely on protein self-assembly to form large functional complexes like the Ribosome. By contrast, in numerous diseases like Alzheimer's, type 2 diabetes, and sickle cell anemia, a lack of protein quality control gives rise to slim, pathological fibers. Why complex particles like proteins robustly form well-defined morphologies across diverse molecular settings is largely unknown.

Geometrical frustration was proposed as a possible underlying physical principle of self-assembly that favors slim aggregates. It arises when complex anisotropic particles cannot simultaneously establish favorable interactions with all nearest neighbors in assembly due to geometric constraints. A previous numerical study suggests that the role of geometrical frustration may be most significant on the scale of individual self-assembling components. To date, no analytical self-assembly framework exists to corroborate this idea. Therefore, this thesis investigates the small-scale effects of geometrical frustration by designing an analytical lattice gas model on a treelike graph, namely the Husimi tree of triangles. Particles interact anisotropically with their nearest neighbors on this lattice, mimicking protein self-assembly's complexity and geometric constraints. We exploit the lattice's treelike structure to derive exact recursion relations for the grand canonical partition function of individual aggregates. Our model exhibits diverse aggregate morphologies, including fibers and crystals, that result from a nontrivial competition of many distinct interactions. To assess whether well-defined aggregate morphologies occur robustly, we formalize structural features to classify aggregates, namely their bulkiness, how often they branch, and whether they exhibit order. We find that, despite the complex interaction landscape, well-defined ordered aggregates using only a few distinct interactions emerge and dominate the self-assembly behavior, as particle anisotropy increases. In this regime, geometrical frustration strongly favors slimmer aggregates like fibers and fiber networks, while its absence is characteristic of ordered space-filling morphologies like crystals.

This master thesis underscores geometrical frustration on the single-particle scale as a principle of morphology selection in the self-assembly of anisotropic particles with complex interactions. Moreover, our partition function formalism for individual aggregates could facilitate the design of components for desired self-assembly outcomes.

# Contents

1	INTRODUCTION		
2	<ul> <li>MODEL</li> <li>2.1 Interacting Anisotropic Particles on the Husimi Tree</li></ul>	7 7 10 13 19	
3	<ul> <li>EXPLORING MORPHOLOGICAL DIVERSITY</li> <li>3.1 Aggregate Size Is Sensitive to Particle Interactions and Fugacity</li> <li>3.2 Aggregate Morphologies Are Diverse at the Gel Point</li> <li>3.3 Observables of Aggregate Morphology</li></ul>	$25 \\ 26 \\ 29 \\ 37$	
4	<ul> <li>COMPLEX PARTICLES ROBUSTLY FORM WELL-DEFINED MORPHOLOGIES</li> <li>4.1 Aggregate Morphologies Fall Into Stereotypical Categories</li> <li>4.2 Particle Anisotropy Promotes Aggregate Order</li></ul>	$\begin{array}{c} 45\\ 46\\ 52 \end{array}$	
5	<ul> <li>GEOMETRICAL FRUSTRATION</li> <li>5.1 Full Occupancy Case of Husimi Tree and Bethe Lattice</li></ul>	55 56 60	
6	CONCLUSION           6.1         Discussion         .	67 67 70	
A	SUPPLEMENTARY INFORMATIONA.1 Choice of Vanishing Interactions for Empty SitesA.2 Formal Properties of the Gel Point	73 73 75	
В	NUMERICAL METHODSB.1 Mathematica: Solving Systems of EquationsB.2 Python: Sampling and Drawing Aggregates	79 79 84	
Bı	BLIOGRAPHY	93	

# 1

# Introduction

Self-assembly is a ubiquitous process in nature [1], where individual constituents autonomously arrange into larger-scale structures or ordered patterns based only on their local interactions. Examples range from the self-assembly of viruses [2] and bacterial flagella [3] all the way to bacterial colonies [4], snowflakes [5], spider silk [6], and even galaxies [7]. In living cells, protein self-assembly is a key mechanism by which large functional complexes are formed [8]. For instance, RNA not only codes for proteins but may fold and self-assemble with other proteins to form Ribosomes [9], the cell's protein factories. In 2002, Whitesides and Grzybowski commented, "living cells self-assemble, and understanding life will therefore require understanding selfassembly" [1]. Due to its fundamental importance to life, self-assembly has attracted considerable interest in a large variety of scientific disciplines ranging from medicine [10] and chemistry [11, 12] to physics [13, 14]. Further, thanks to technological advances in supramolecular [15], polymeric [16] and colloidal self-assembly [17, 18] as well as DNA origami [19–21], self-assembly is increasingly utilized in nano- and microengineering [22], materials science [23], and bionanotechnology [24, 25]. Although still in its infancy, the concept of "Materials by Design" aims to integrate Machine Learning (ML) [26] to design constituent particles for self-assembled functional materials. Furthermore, scientists have successfully synthesized a ribosomal subunit on a chip, reconstructing this essential component of the cell's self-replication mechanism outside the cell for the first time [27]. This achievement only scratches the surface, as living cells are far more complex, combining intricate molecular structures with sophisticated regulatory mechanisms [28]. Reproducing life-like functions or even a full "synthetic cell" will require individual self-assembling components that can code for this complexity [29]. Adopting a broader physics perspective could aid in unifying the phenomenology of diverse self-assembly systems, thereby narrowing down the search for suitable self-assembling components. However, the general physical laws of how complex building blocks determine the resulting large-scale structure in self-assembly remain elusive.

Simple, well-fitting particles like identical cubes may form dense space-filling aggregates. What happens if ill-fitting irregular particles or, more generally, particles with generic anisotropic interactions are considered? The intracellular environment



Figure 1.1: Geometrical frustration could be prevalent in protein self-assembly. a, In sickle cell anemia, normally soluble hemoglobin aggregates to form pathological fibers due to mutation of a single amino acid (glutamate 6 to valine; in green). Adapted from [39]. b, Proteins are comparable to ill-fitting puzzle pieces. As such, proteins are subject to geometrical frustration if they cannot simultaneously establish favorable interactions with multiple nearest neighbors in assembly.

is crowded, which may lead to unspecific interactions between "ill-fitting" proteins, resulting in uncontrolled aggregation, potentially causing cell death [30, 31]. Healthy living cells rely on protein quality control via molecular chaperones to regulate protein folding, and hence their interactions [32, 33]. However, this safety mechanism can fail. In neurodegenerative diseases like Alzheimer's or Parkinson's, as well as in type 2 diabetes and sickle cell anemia, typically soluble proteins aggregate into pathological fibers [34–36]. Polymerization of sickle-cell hemoglobin (see Fig. 1.1a) is one of the best-understood instances of protein self-assembly in terms of the molecular mechanisms involved [37, 38].

However, why complex and anisotropic components can robustly form well-defined morphologies is largely unknown. Besides, the frequent occurrence of fibers in this context contrasts sharply with three-dimensional assemblies of isotropic particles like spherical colloids. This may suggest the existence of an underlying physical principle for complex anisotropic building blocks that favors aggregates with reduced dimensionality, meaning they do not grow in every spatial direction.

Geometrical frustration has recently emerged as a potential physical principle of self-assembly to explain this [40–42]. Complex anisotropic particles may be incapable of simultaneously establishing favorable interactions with multiple nearest neighbors in assembly due to geometric constraints. This phenomenon is broadly called "frustration". To visualize it, think of irregularly shaped puzzle pieces (see Fig. 1.1b). Due to their geometry, tiling the plane would require unfavorable binding, making the assembly geometrically frustrated. As a consequence, aggregates of these puzzle pieces only grow in one spatial direction, resulting in a reduced dimensionality that is reminiscent of fibers.

While geometrical frustration may arise very locally, like in the above example,

it can also progressively build up and prevent aggregates from growing indefinitely. It was first shown that geometrical frustration generically gives rise to self-limiting and fibrous aggregates using a minimal model of weakly deformable, ill-fitting, sticky particles [43]. Recently, this result has been reaffirmed and analytically extended using a continuum description of a broader class of elastic particles [44]. Proteins are indeed somewhat flexible and can change their conformation upon binding [45]. However, protein-protein interactions are not only influenced by their shape but are also dictated by their surface amino acids. Their interactions may stem from hydrophobichydrophilic effects or electrostatic forces, among others [46, 47]. As a result, interaction energies can independently vary in magnitude. Based on this observation, a numerical model of anisotropic hexagonal lattice particles was introduced [48]. The relative orientation of anisotropic particles toward each other determines whether binding is favorable or not. This leads to a competition between many distinct interaction energies. Despite the inherent complexity of their interactions, these particles only form a small number of qualitatively different aggregate morphologies that often exhibit a well-defined structure, including fibers and crystals. The results also strongly indicate that identical complex particles tend to form slim or porous aggregates to avoid geometrical frustration. Its effect seems strongest on the single-particle scale. This is corroborated by the fact that "propagability", the particles' ability to form periodic structures, as defined by the authors, is a reliable predictor of aggregate morphology [48].

In this Master's thesis, we analytically investigate the local effects of geometrical frustration by singling them out in a minimal model. Our goal is to gain a broad understanding of which aggregate morphologies generically emerge in frustrated self-assembly rather than designing particles to assemble into specific morphologies. Specifically, we show that increasing particle anisotropy induces the emergence of geometrical frustration as a morphology selection mechanism within our model. While geometrical frustration favors fiber formation, its absence often leads to ordered spacefilling structures like crystals.

Our discussion of geometrical frustration hinges on particles with anisotropic interactions. To study geometrical frustration, we thus adopt a particle-based description of self-assembly. We exclusively probe equilibrium self-assembly for simplicity. In the *lattice gas model*, particles are placed on the vertices of a lattice and interact only with their nearest neighbors [49–51]. This framework is suitable for describing particles with anisotropic interactions because their relative orientations toward each other can be formalized. Observables of interest can then be obtained by calculating an appropriate partition function. Lattice gas models can be based on two-dimensional lattices like a triangular lattice [48]. However, their analytical treatment becomes very complicated and involved even without considering the anisotropy of particles. This is mainly because an aggregate can loop back to itself, thereby introducing effective long-range interactions between particles of that aggregate. There are numerous ways in which this looping can happen, making it impossible to keep track of all inter-



Figure 1.2: Treelike graphs are a suitable framework to study geometrical frustration in a minimal setting analytically. a, Bethe lattice of coordination f = 4, a tree graph without loops. b, Husimi tree of triangles. While the treelike structure is preserved, the triangles constitute small-scale loops. c, Triangles represent a minimal setting of geometrical frustration (also see Fig. 1.1c).

actions and how they affect aggregate morphology. We make the approximation that large-scale loops do not qualitatively change the self-assembly behavior by utilizing treelike graphs (see Fig. 1.2). Akin to triangular or square lattices, a graph is defined as a collection of vertices at which particles may sit, joined by edges, corresponding to nearest-neighbor interactions. The characteristic feature of tree graphs is that there are no loops at all, while on treelike graphs, there are only small-scale loops (cf. Fig. 1.2). This allows a recursive approach to obtaining the partition function exactly. For our purposes, we will study interacting particles on a Husimi tree of triangles (see Fig. 1.2b), simply referred to as Husimi tree in the following. In a triangle, it might be impossible to realize favorable interactions between three nearest neighbor particles simultaneously (see Fig. 1.2c). Therefore, triangles represent a minimal setting of geometrical frustration that enables us to study its role on a local scale.

We define our model of interacting complex particles in Chapter 2. Building on this description, we introduce recursion equations for the grand canonical partition function of aggregates, quantifying the likelihood of different aggregates to be formed. In Chapter 3, we characterize aggregate morphologies based on their structural features: their bulkiness, how branchy they are, and whether they are ordered or not. On our treelike graph, these properties are sufficient to distinguish between all different morphologies. In particular, we can differentiate fibers of varying thickness from crystals and these two morphologies from disordered ones like liquids. With the self-assembly of pathological fibers in contrast to space-filling crystals in mind, we focus our analysis exclusively on large aggregates, where morphological differences are most apparent. Our model allows for a plethora of binding motifs. In Chapter 4, nevertheless, only a few categories of well-defined aggregate morphologies emerge if we increase particle anisotropy, including fibers and crystals. Finally, in Chapter 5, we put forward geometrical frustration as a governing principle of morphology selection for anisotropic particles with complex interactions by comparing self-assembled morphologies in our model system to those on an unfrustrated loopless tree graph (see

Fig. 1.2a). Geometrical frustration strongly favors slimmer aggregates like fibers and fiber networks, while its absence is characteristic of ordered space-filling morphologies, including crystals. Chapter 6 discusses our findings and provides an outlook for further research.

# 2 Model

Our objective is to design a minimal model of geometrical frustration for complex interacting particles. We choose the Husimi tree as our underlying lattice, as triangles constitute a minimal setting of frustration (cf. Fig. 1.2). At the same time, the lattice's treelike structure, obtained by simply merging triangles at their corners, enables the analytical study of the model as we will see later.

How do we model interacting particles on this lattice? With proteins as ill-fitting puzzle pieces in mind, they should be anisotropic such that they can be geometrically frustrated in the first place. Simultaneously, they should exhibit some degree of complexity reminiscent of proteins that is suitable to the lattice we place them on and still analytically tractable. Thus, we use identical copies of a particle with four different binding sites, since each lattice vertex also has four nearest neighbors. That way, interactions between neighboring particles vary depending on their relative orientation, without making the model needlessly complicated.

In this chapter, we first define how anisotropic particles interact on our lattice and introduce the system's Hamiltonian in Section 2.1. Then, we have a look at the associated binding motifs in Section 2.2, which represent the intermediate building blocks of aggregates. Ultimately, we are interested in the morphology of single aggregates. Thus, in Section 2.3 we set up the model's formal basis by deriving a recursion relation for the partition function of aggregates. We complement this perspective on single aggregates with a description of the statistics of the whole lattice in Section 2.4. Thereby, we show how these two perspectives relate to each other. Together, they formally cover all aggregate phenomenology of our model. In later chapters, we build on this formal framework to statistically examine and analyze aggregate morphologies.

## 2.1 Interacting Anisotropic Particles on the Husimi Tree

For the purpose of this section, picture anisotropic particles as irregular puzzle pieces that sit on vertices of the Husimi tree (see Fig. 2.1). Whenever two particles sit



Figure 2.1: Anisotropic particles on neighboring vertices of the Husimi tree constitute an aggregate. a, Anisotropic particles can be visualized as irregular puzzle pieces that are placed on the Husimi tree. A particle's orientation dictates which of its four faces (labeled 1 to 4) points to which neighboring vertice. For illustrative purposes, we only show a small part of the lattice throughout this thesis. b, If particles sit on neighboring vertices they bind together, here indicated by a bold black edge. A cluster of particles, that are interconnected through these bonds, is called an "aggregate". Note that in order to visualize the Husimi tree, we draw edges smaller the further away they are from the root of the treelike graph. From a formal viewpoint all vertices are equivalent, the lattice is translationally invariant.

on neighboring vertices, they interact and form a bond. A cluster of particles interconnected through bonds is then called an "aggregate". In this section we define anisotropic nearest neighbor interactions between particles by means of an interaction matrix. Subsequently, we formulate the associated overall energy of an aggregate.

Whether inter-particle interactions are favorable depends on which faces of these particles bind together. This in turn is dictated by the particles' relative orientation towards each other. For two particles with four faces, there are a priori  $4 \cdot 4 = 16$  different ways to bind (see Fig. 2.2). However, since our particles are identical the number of different interactions is reduced to 10 due to symmetry. For instance, face 1 of the first particle binding to face 2 of the second particle corresponds to the same interaction as face 2 of the first binding to face 1 of the second. This implies that the so-called interaction matrix is also symmetric:

$$\mathbf{E} = \begin{pmatrix} E_{11} & E_{12} & E_{13} & E_{14} \\ \cdots & E_{22} & E_{23} & E_{24} \\ \cdots & \cdots & E_{33} & E_{34} \\ \cdots & \cdots & \cdots & E_{44} \end{pmatrix}$$
(2.1)

In the above matrix, the two indices denote which pair of faces (labeled 1 to 4) is



Figure 2.2: Due to symmetry, two identical anisotropic particles with four faces exhibit up to 10 distinct interactions. Shown is a visualization of all 16 possible relative arrangements of these particles, with redundant arrangements in gray.

binding together. Each possible bond is quantified by an interaction energy E, with negative energies corresponding to a favorable interaction.

Formally, the Hamiltonian (or overall energy) of a given aggregate is

$$\mathcal{H} = \sum_{\varphi \le \nu} n_{\varphi \nu} E_{\varphi \nu} \,, \tag{2.2}$$

where  $n_{\varphi\nu}$  is the number of contacts between face  $\varphi$  and face  $\nu$  in that aggregate. Note that we sum over  $1 \leq \varphi \leq \nu \leq 4$  to avoid double-counting face pairs. Moreover, we assert that empty vertices do neither interact with particles nor with each other. Intuitively, this corresponds to the case that the solvent (in biological systems usually water) does not interact with assembly components, and hence the surface energy of the aggregate is zero. In Section A.1 we show that this assumption does not change the phenomenology of our model, since it only corresponds to an energy shift of interparticle interactions.

The set of contacts in an aggregate,  $\{n_{\varphi\nu}\}$ , is constrained. For example, the number of particles n in the aggregate places an upper bound, since each particle contributes only one face of each kind,  $n_{\omega\nu} \leq n$ . Moreover, each particle in an aggregate forms at least one bond, i.e.  $\sum_{\varphi \leq \nu} n_{\varphi\nu} \geq n/2$ . In addition to particle geometry, the lattice itself constrains the combinations of interactions that can be realized simultaneously in an aggregate. In Section 2.3 we introduce a formalism that naturally includes all physical aggregates and respects their structural constraints. However, first we want to get a qualitative idea of what self-assembly behavior to expect from our model.

## 2.2 SIMPLE BINDING MOTIFS AS BUILDING BLOCKS FOR AGGREGATE MORPHOLOGY

In contrast to tree graphs, our lattice exhibits small-scale loops, namely triangles. In addition to two-body interactions, these triangles allow particles to enter threebody interactions, possibly leading to frustration as discussed in the introduction (cf. Chapter 1, particularly Fig. 1.1 and Fig. 1.2). We call these three-body interactions "triangle interactions", because the three participating particles constitute a triangle on our lattice. Each two-body and three-body interaction, that can be realized by our anisotropic particles, represents a local binding motif. On the scale of a single lattice triangle, a single binding motif is realized based on their competition, to be formally introduced in the next section. However, aggregates may in principle be build up from any combination of distinct binding motifs. In that sense, these local binding motifs can be viewed as intermediate building blocks of self-assembled structures in our model system. By first isolating the assembly behavior associated to each individual binding motif, we can better understand which aggregate morphologies arise when they compete in our model later. In this section, we thus ask: What aggregate morphologies can be built from copies of only a single binding motif? Subsection 2.2.1 illustrates the structures that are connected to two-particle interactions, while Subsection 2.2.2 is concerned with three-particle interactions.

#### 2.2.1 Two-Particle Binding Motifs Yield Fibers or Dimers

Taking a closer look at all possible two-particle interactions in our system, it becomes clear that there are some qualitative differences between them (see Fig. 2.3). For instance, the interaction between faces 1 and 3, which we denote by the tuple (1,3), involves two particles that are both pointing in the same direction. This is an example of a so-called head-to-tail interaction. On the other hand, two particles have to face each other to realize a (1, 1) head-to-head interaction.

Think of a system, where only (1, 1) is allowed. Apart from monomers, only dimers (i.e. two-particle aggregates) would occur (see Fig. 2.4a). Conversely, (1, 3) does allow the formation of fibers by placing particles side by side indefinitely (see Fig. 2.4b). As it seems, not all interactions are created equal! In fact, different interactions may lead to qualitatively different aggregate morphologies. We similarly consider every other possible two-particle interaction, and determine the morphologies that can be built from copies of it. The results are summarized in Table 2.1. There is a third



Figure 2.3: **Two-body interactions in our system can be grouped into three qualitatively different cases.** Throughout this thesis, interactions are accordingly color-coded. The diagonal of the interaction matrix represents symmetric dimer interactions, also called "head-to-head" interactions (purple). Interactions on the first and third off-diagonal are called "cyclic", because particles can form closed loops through them, namely triangles (yellow). Lastly, "head-to-tail" interactions are capable of constituting straight fibers (blue-green). Also refer to Fig. 2.4.

type of two-body interaction, copies of which can either yield dimers or curved fibers (see Fig. 2.4). All 10 distinct two-body interactions in our model fall into one of the above three broad categories.

Table 2.1: If only a specific two-particle interaction is allowed in the system, particles self-assemble into either dimers, or straight fibers or both dimers and curved fibers, respectively (cf. Fig. 2.4).

Binding motif	Aggregate morphology
(1,1) $(2,2)$ $(3,3)$ $(4,4)$	Dimers
(1,3) $(2,4)$	Straight fibers
(1,2) $(2,3)$ $(3,4)$ $(1,4)$	Dimers and curved fibers

#### 2.2.2 Three-Particle Binding Motifs Lead to Oligomers, Fibers Composed of Triangles or Bulks

Three particles can constitute a three-body interaction in a lattice triangle, which we also call triangle interaction. For combinatoric reasons, there are 24 distinct triangle interactions in our model. To give an example, consider three particles connected through three (1, 2) bonds, which we denote by  $\{(1, 2), (1, 2), (1, 2)\}$ . This binding motif yields a threefold symmetric triangular aggregate or oligomer (see Fig. 2.5a).



Figure 2.4: Two-body interactions are associated with one of three types of basic aggregate morphologies: dimers, fibers and curved fibers. Shown are visualizations of aggregates with bonds colored according to Fig. 2.3, and a few puzzle pieces are added to illustrate particle orientation. **a**, Dimers stereotypically form through head-to-head interactions, e.g. (1, 1). **b**, Head-to-tail interactions, e.g. (1, 3), yield straight fibers. **c**, Systems with cyclic two-body interactions exhibit both dimers and curved fibers, depending on the orientation of particles on the lattice, as exemplified here for the (1, 2) interaction.

Aggregates that are composed of only a few particles (but more than one) are generally called oligomers. Throughout this thesis, we reserve this name to triangular aggregates. Therefore, (1, 2) is also called a cyclic interaction (cf. Fig. 2.3). The "heads" of all three particles are pointing towards the center of the triangle, which is why none of them can take part in yet another copy of the same binding motif. Correspondingly the aggregate cannot be grown further by only using the same binding motif. Not all oligomers are symmetric, as exemplified in Fig. 2.5b. In contrast to oligomers, some other triangle binding motifs lead to curved fibers that are composed of many triangles (see Fig. 2.5c) or bulks, i.e. aggregate morphologies that can tile the whole lattice (see Fig. 2.5d).



Figure 2.5: Copies of a single specific triangle-interaction can form one of three basic aggregate morphologies: oligomers, curved fibers of triangles or bulks. Here, examples for each morphology are visualized, with bonds colored according to Fig. 2.3. Some puzzle pieces are added to illustrate particle orientation. **a**, Symmetric oligomers are composed from three copies of the same cyclic interaction, here  $\{(1, 2), (1, 2), (1, 2)\}$ . **b**, Oligomers can also be asymmetric, e.g. for the  $\{(1, 4), (2, 4), (1, 1)\}$  triangle-interaction. **c**, A curved fiber of triangles can be formed through the combination of a cyclic interaction with dimer interactions. In this example, the  $\{(1, 1), (4, 4), (2, 3)\}$  triangle-interaction is shown. **d**, Bulks can be formed through triangle interactions of three different cyclic two-body interactions, as exemplified here for the binding motif  $\{(1, 2), (1, 4), (2, 3)\}$ .

# 2.3 Recursion Relations Yield Aggregate Partition Functions

Aggregates are like a "mosaic" that is uniquely defined by the different local binding motifs that make it up and their relative arrangement. In an infinite system at thermal equilibrium, all combinations of binding motifs that are permitted by constraints due to particle and lattice geometry do occur, and hence an infinite variety of aggregates. However, different aggregates are not equally as prevalent. Instead, the likelihood of encountering a specific aggregate is determined by its energetic favorability (quantified by its hamiltonian  $\mathcal{H}$ , see Eq. (2.2), together with the fixed interaction matrix  $\mathbf{E}$ , see Eq. (2.1)) as well as its entropic multiplicity. This notion of probability is formalized within a suitable statistical ensemble. While the total number of individual particles in our system may be fixed, its aggregates can accommodate a varying number of interacting particles. It is thus natural to employ a grand canonical ensemble of aggregates, which enables us to assign a probabilistic weight to each aggregate. To quantify probabilities of aggregates, we need a base distribution from which we sample uniformly. In our model, we sample aggregates according to their volume fraction, i.e. aggregates are weighted by the number of particles they contain. This amounts to choosing a random particle in the system (each with probability 1/N, where N is the total number of particles in the system) and examining the aggregate to which it belongs. That particle constitutes the root of the aggregate on our treelike lattice, which is why we term it the "root particle". In this section, we first derive a recursion relation of the entropic multiplicity of aggregates that contain a given number of particles n in Subsection 2.3.1. Subsequently, we define a grand canonical ensemble that includes aggregates of all sizes and employ a generating function approach in Subsection 2.3.2 in order to obtain the grand canonical partition function. Generalizing the aforementioned recursion relation to include given anisotropic interactions  $\mathbf{E}$  between particles (cf. Section 2.1), we obtain a self-consistency relation for the partition function. In doing so, we fully capture the probabilistic weights of all aggregates as a function of their interactions  $\mathbf{E}$  and the total number of particles in the system, which is quantified by a fugacity c. This allows us to later design and obtain observables of aggregate morphology, thereby enabling the statistical analysis of aggregate structure.

#### 2.3.1 Entropic Multiplicity of Aggregates

How many ways  $\Omega_n$  are there, to build an aggregate of a given size n (meaning it contains n particles) starting from the root particle? For one, the aggregate's multiplicity  $\Omega_n$  stems from the four different orientations each of its particles can have, a trivial contribution. In addition, a given fixed number of particles can be arranged on the lattice in a variety of ways to form different aggregates, which we now formalize.

At the vertex where the root particle sits, two triangles are joined at their corners (cf. Fig. 2.1). Thereby the root particle divides the aggregate to which it belongs into two subaggregates that emanate from it. These subaggregates accommodate  $k_1$  and  $k_2$  particles, respectively, so that the total particle constraint of the aggregate is satisfied,  $1 + k_1 + k_2 = n$ . Let  $\tilde{\Omega}_k$  be the multiplicity of such a subaggregate. Since the two subaggregates are independent of each other, the aggregate multiplicity  $\Omega_n$  is then given as the product of the multiplicities  $\tilde{\Omega}_{k_1}$  and  $\tilde{\Omega}_{k_2}$ :

$$\Omega_n = 4 \sum_{k_1+k_2=n-1} \tilde{\Omega}_{k_1} \tilde{\Omega}_{k_2}$$

$$(2.3)$$



Figure 2.6: The partial partition functions obey a self-consistent recursion relation thanks to the treelike structure of the Husimi tree. Each term in the recursion relation, Eq. (2.4), can be visualized in diagrammatic form. The left hand side (l.h.s.) of the diagrammatic equation corresponds to the partial partition function of a subaggregate emanating from a root particle with orientation  $\psi \in \{1, 2, 3, 4\}$ . We define the orientation  $\psi$  to be the label of the face pointing towards the top right triangle corner. As shown on the right hand side (r.h.s.), the root particle can either have zero (first term), one (second and third term) or two neighboring particles (fourth term). Bonds between particles are represented by bold black edges, while their absence is indicated by light gray edges.

Here, the prefactor 4 corresponds to the 4 orientations which the root particle itself can have. The sum runs over all allowed combinations of subaggregate sizes  $k_1$  and  $k_2$ . Determining the multiplicity of the subaggregates  $\tilde{\Omega}_k$  will give us the multiplicity of the whole aggregate  $\Omega_n$ . We formally set  $\tilde{\Omega}_k = 0$  for k < 0, because there cannot be aggregates of negative size. As visualized diagrammatically in Fig. 2.6, a subaggregate can be built up recursively by placing particles on the neighboring vertices of the root particle and subsequently considering the subaggregates that emanate from these particles in turn. Based on this,  $\tilde{\Omega}_k$  obeys the following recursion relation:

$$\tilde{\Omega}_{k} = \delta_{k,0} + 2 \cdot 4 \tilde{\Omega}_{k-1} + 4^{2} \sum_{k_{1}+k_{2}=k-2} \tilde{\Omega}_{k_{1}} \tilde{\Omega}_{k_{2}}$$
(2.4)

If there are no particles in the subaggregate, k = 0, it terminates and the recursion relation is closed via  $\tilde{\Omega}_1 = 1$ . This case is quantified by the Kronecker delta  $\delta_{k,0}$ , which is defined as

$$\delta_{ij} = \begin{cases} 1, & \text{if } i = j, \\ 0, & \text{if } i \neq j. \end{cases}$$

$$(2.5)$$

For k > 0, there can be either one or two particles on the neighboring vertices of the root particle in a lattice triangle. The prefactors  $2 \cdot 4$  and  $4^2$  of the second and third term in Eq. (2.4), respectively, correspond to their possible arrangements with respect to the root particle, where we take into account their orientational degree of freedom. Each of these particles constitutes another subaggregate, whose sizes respect the constraint on the total number of particles in the aggregate.

If all aggregates in the system were of size n, then each aggregate would be sampled with equal probability  $1/\Omega_n$  (for non-interacting particles). In the next subsection we introduce a grand canonical ensemble in order to consider aggregates of varying sizes that are composed of anisotropic interacting particles.

#### 2.3.2 Grand Canonical Partition Function via Generating Function Approach

The total number of particles N in the system influences the relative probabilities with which aggregates of different sizes are sampled. N is fixed through the particle density  $\rho$ , that relates the former to the system volume V,  $\rho = N/V$ . This density can be implicitly enforced in a suitable grand canonical ensemble. In this subsection, we first define a grand canonical ensemble of aggregates, that assigns a probabilistic weight to aggregates of different sizes. Subsequently, we employ a generating function approach to find a set of self-consistent equations for the grand canonical partition function, based on the recursion relations that we introduced in the last subsection (see Eq. (2.4)).

#### A GRAND CANONICAL ENSEMBLE OF AGGREGATES

Let  $Z_n$  be the canonical partition function of aggregates of size n.  $Z_n$  is a sum over all aggregates of that size, each weighted according to their energy  $\mathcal{H}$  (see Eq. (2.2)):

$$Z_n = \sum_{\{\text{aggregates of size } n\}} e^{-\beta \mathcal{H}(\text{aggregate})}$$
(2.6)

The grand canonical partition function of our system  $\mathcal{Z}$  can then be defined as a formal power series in a fugacity c:

$$\mathcal{Z} \coloneqq \sum_{n=1}^{\infty} Z_n c^n \,, \tag{2.7}$$

where each summand  $Z_n c^n$  quantifies the probability weight of aggregates of size n. Let  $\beta = 1/k_B T$  denote the inverse thermal energy. The fugacity  $c =: e^{\beta \mu}$  couples each aggregate to an infinite particle reservoir with chemical potential  $\mu$ . If  $\mu$  is large and positive, it is very favorable to introduce particles from the reservoir into the aggregate. Conversely, for negative  $\mu$  adding more particles is unfavorable. An aggregate of size n is thus weighted with a factor  $c^n$ . The particle density  $\rho$  fixes the fugacity c. Let  $\phi_n = nm_n/V$  be the volume fraction of aggregates of size n, where  $m_n$  is the total number of aggregates of that size in the system. Since  $\mathcal{Z}$  assigns a probabilistic weight to aggregates according to their volume fraction and  $Z_1 = 1$ , the fugacity is related to the (a priori unknown) volume fraction of monomers by

$$\frac{c}{\mathcal{Z}} = \frac{\phi_1}{\rho} \,, \tag{2.8}$$

where we used that the density is the combined volume fraction of all aggregates:

$$\rho = \frac{N}{V} = \sum_{n=1}^{\infty} \phi_n \tag{2.9}$$

Later, Subsection 2.4.2 also provides a relation between  $\rho$  and c in which we have eliminated the monomer fraction  $\phi_1$  (see Eq. (2.33)).

#### Self-Consistency Relations for Grand Canonical Partition Function

We now employ a generating function approach in order to obtain a set of selfconsistent equations for the grand canonical partition function  $\mathcal{Z}$ . In order to do this, we generalize the recursion relation for entropic multiplicity (see Eq. (2.4)) to incorporate interactions **E** between particles.

In the case of vanishing interactions,  $\mathbf{E} = \mathbf{0}$ , the partition function  $\mathcal{Z}(\mathbf{E} = \mathbf{0})$  is simply a generating function of the entropic multiplicities  $\Omega_n = |\{\text{aggregates of size } n\}|$ :

$$\mathcal{Z}(\mathbf{E} = \mathbf{0}) = \sum_{n=1}^{\infty} \Omega_n c^n \tag{2.10}$$

Using Eq. (2.3), we bring  $\mathcal{Z}(\mathbf{E} = \mathbf{0})$  into a form where we can apply the recursion relation Eq. (2.4):

$$\begin{split} \mathcal{Z}(\mathbf{E} = \mathbf{0}) &= 4c \sum_{n=1}^{\infty} \sum_{k_1+k_2=n-1} \tilde{\Omega}_{k_1} c^{k_1} \tilde{\Omega}_{k_2} c^{k_2} \\ &= 4c \sum_{n=0}^{\infty} \sum_{k=0}^{n} (\tilde{\Omega}_k c^k) (\tilde{\Omega}_{n-k} c^{n-k}) \\ &= 4c \left( \sum_{k_1=0}^{\infty} \tilde{\Omega}_{k_1} c^{k_1} \right) \left( \sum_{k_2=0}^{\infty} \tilde{\Omega}_{k_2} c^{k_2} \right) \equiv 4c G (\mathbf{E} = \mathbf{0})^2 \end{split}$$
(2.11)

In the second line, we performed an index shift and regrouped terms to apply the Cauchy product rule for infinite series in the third line. For two infinite series  $\sum_{i=0}^{\infty} a_i$  and  $\sum_{j=0}^{\infty} b_j$ , the Cauchy product rule is the following discrete convolution:

$$\left(\sum_{i=0}^{\infty} a_i\right) \cdot \left(\sum_{j=0}^{\infty} b_j\right) = \sum_{k=0}^{\infty} c_k \quad \text{with} \quad c_k = \sum_{l=0}^k a_l b_{k-l} \quad (2.12)$$

We also defined the partial (grand canonical) partition function for vanishing interactions

$$G(\mathbf{E} = \mathbf{0}) \coloneqq \sum_{k=0}^{\infty} \tilde{\Omega}_k c^k , \qquad (2.13)$$

which generates the multiplicities of subaggregates. Employing the recursion relation

for  $\tilde{\Omega}_k$  (see Eq. (2.4)), we find a self-consistency relation for  $G(\mathbf{E} = \mathbf{0})$ :

$$G(\mathbf{E} = \mathbf{0}) = \sum_{k=0}^{\infty} c^k \left( \delta_{k,0} + 2 \cdot 4\tilde{\Omega}_{k-1} + 4^2 \sum_{k_1 + k_2 = k-2} \tilde{\Omega}_{k_1} \tilde{\Omega}_{k_2} \right)$$
  
= 1 + 2 \cdot 4cG(\mathbf{E} = \mathbf{0}) + 4^2 c^2 G(\mathbf{E} = \mathbf{0})^2 (2.14)

We now show how this relation can be generalized to nonvanishing anisotropic interactions between particles. Qualitatively, the partial partition function of subaggregates G is of the form

$$G = \sum_{\{\text{subaggregates}\}} e^{-\beta \mathcal{H}(\text{subaggregate})} c^{\text{size of subaggregate}} \,.$$
(2.15)

Due to the lattice's treelike structure,  $e^{-\beta \mathcal{H}(subaggregate)}$  factorizes over individual lattice triangles: If E(triangle) is the sum of interaction energies in a given lattice triangle, then

$$e^{-\beta \mathcal{H}(subaggregate)} = \prod_{triangles} e^{-\beta E(triangle)},$$
 (2.16)

where the product runs over all lattice triangles that belong to the considered subaggregate. Equation (2.14) already self-consistently encodes the multiple ways in which particles in a triangle can be oriented towards each other (assuming a given root particle orientation). There are 8 ways in which a single particle can be arranged with respect to the root particle and 16 ways how this can be done for two particles. Each arrangement corresponds to a set of realized interactions between particles in that triangle. Thus, we have to decorate each particle arrangement in Eq. (2.14) with an interaction Boltzmann weight

$$y_{\varphi\nu} \coloneqq \mathrm{e}^{-\beta E_{\varphi\nu}} \tag{2.17}$$

for every pair of particles that bind together through faces  $\varphi$  and  $\nu$  ( $\varphi, \nu \in \{1, 2, 3, 4\}$ ). In order to formally do this, the orientation of each involved particle needs to be made explicit. We define the orientation  $\psi$  of the root particle as the face pointing towards its right neighbor in the considered triangle as visualized in Fig. 2.6. If this is the case for face 1, then we have  $\psi = 1$  and so on. The other particles in the triangle constitute new subaggregates and their orientations are defined correspondingly. Say, the root particle has orientation 1 and it binds to its right neighbor that has orientation 2. Then the realized interaction is  $(1, 2 + 3) \cong (1, 1)$ , where we use that two face labels are equivalent if they are the same up to addition of multiples of 4. Finally, if the root particle has orientation  $\psi \in \{1, 2, 3, 4\}$ , then the associated partial partition function  $G_{\psi}$  obeys the following self-consistency relation:

$$G_{\psi} = 1 + c \sum_{\varphi=1}^{4} (y_{\psi+1,\varphi+2} + y_{\psi,\varphi+3}) G_{\varphi} + c^2 \sum_{\varphi,\nu} y_{\psi+1,\varphi+2} y_{\psi,\nu+3} y_{\varphi+3,\nu+2} G_{\varphi} G_{\nu} \quad (2.18)$$

Equation (2.18) constitutes a system of four coupled quadratic equations in the partial partition functions. Lastly, the total partition function  $\mathcal{Z}$  includes one term for each possible orientation of the root particle. It is related to the partial partition functions by

$$\mathcal{Z} = c \sum_{\psi=1}^{4} G_{\psi} G_{\psi+2} \,. \tag{2.19}$$

In Eq. (2.18), morphologies of individual aggregates and their probabilistic weight are encoded, which is why we call this description "aggregate-centered".

We are now in the position to sample and illustrate aggregates for a given interaction matrix **E** and fugacity c by solving Eq. (2.18). Moreover, manipulating this self-consistent system of equations will enable the calculation of observables that characterize aggregate morphology in Chapter 3. However, our model toolbox is still incomplete.  $\mathcal{Z}$  is a Taylor series expansion in fugacity c which has a finite radius of convergence  $c_{\rm g}$  around c = 0.  $c = c_{\rm g}$  is the so-called gelation point, at which an infinite aggregate appears in the system (see Section 3.2). While the divergence of  $\mathcal{Z}$  ultimately allows us to purposefully examine the rich morphologies of large aggregates at gelation in the next chapter, the aggregate-centered formalism as described above is not applicable to self-assembly systems in which infinite aggregates comprise a nonzero fraction of all particles; this is the case for  $c > c_{\rm g}$ . The next section takes a different perspective which is better-suited to this regime.

#### 2.4 Relating Aggregate to Lattice Description

In order to be able to sample from the regime of infinite aggregates, we employ a "lattice-centered" framework, which considers the partition function of the whole lattice instead of the one of individual aggregates. Then, formal divergences can be eliminated from the model by appropriately rescaling the corresponding partial partition functions. In Subsection 2.4.1, we derive recursion relations for the lattice partition function  $\mathcal{Z}_{\text{lat}}$  similarly to before for single aggregates. We show how the aggregate-centered and lattice-centered descriptions are related to each other by providing a simple mapping in Subsection 2.4.2. Consequently, the partition function of single aggregates equivalently informs about properties of the whole lattice gas assembly, for instance the density  $\rho$  of particles on the lattice, as long as no infinite aggregates are present in the system. Combined, these two perspectives cover every possible self-assembly behaviour in our system and they enable the prediction of aggregate morphologies resulting from any anisotropic interactions and particle density.

#### 2.4.1 The Lattice Partition Function Follows From Similar Recursion Relations

We again invoke the treelike structure of our lattice to derive self-consistency relations for the partition function  $\mathcal{Z}_{lat}$  of the whole lattice. The main conceptual difference of

this perspective to the one in Section 2.3 is the following: In our aggregate-centered formalism, an empty vertex constitutes termination of the aggregate. Thus, the aggregate partition function is not concerned with what happens beyond this empty vertex. However, the lattice itself does not terminate at empty vertices but extends to infinity. So, the lattice partition function also needs to contain information on what happens beyond empty vertices.

In this perspective, the whole lattice is coupled to a particle reservoir via another fugacity  $z = e^{\beta \mu_{\text{lat}}}$ . Let  $Z_N$  be the canonical partition function of N interacting particles on the lattice:

$$Z_N \coloneqq \sum_{N \text{ particle configurations of the lattice}} e^{-\beta \mathcal{H}(\text{lattice})} , \qquad (2.20)$$

where  $\mathcal{H}(\text{lattice})$  is the total interaction energy of a given lattice configuration. The grand canonical lattice partition function is defined as

$$\mathcal{Z}_{\text{lat}} = \sum_{N=0}^{\infty} Z_N z^N \,. \tag{2.21}$$

Each summand in  $\mathcal{Z}_{\text{lat}}$  is the probability weight of the corresponding lattice configuration. In the last section, we used that the root particle of an aggregate divides it into two subaggregates (see Subsection 2.3.1). Instead, now each lattice vertex (either empty or accommodating a particle) separates the whole lattice into two infinite sublattices. Hence, the lattice partition function  $\mathcal{Z}_{\text{lat}}$  can be written in terms of partial partition functions of the sublattices. Let  $g_0$  be the partial partition function of an infinite sublattice that originates from an empty vertex, and let  $g_{\psi}$  denote the partial partition function corresponding to a nonempty vertex with a particle of orientation  $\psi$ . The partition function of the whole lattice  $\mathcal{Z}_{\text{lat}}$  is then given by (cf. Eq. (2.19))

$$\mathcal{Z}_{\text{lat}} = g_0^2 + z \sum_{\psi=1}^4 g_{\psi} g_{\psi+2} \,. \tag{2.22}$$

Similarly to before (cf. Eq. (2.18)) we obtain the partial partition functions via recursion:

$$g_{0} = g_{0}^{2} + 2zg_{0}\sum_{\varphi=1}^{4}g_{\varphi} + z^{2}\sum_{\varphi,\nu}y_{\varphi+3,\nu+2}g_{\varphi}g_{\nu}$$

$$g_{\psi} = g_{0}^{2} + zg_{0}\sum_{\varphi=1}^{4}(y_{\psi+1,\varphi+2} + y_{\psi,\varphi+3})g_{\varphi} + z^{2}\sum_{\varphi,\nu}y_{\psi+1,\varphi+2}y_{\psi,\nu+3}y_{\varphi+3,\nu+2}g_{\varphi}g_{\nu}$$
(2.23)

Note that these partial partition functions generally diverge to infinity: for vanishing fugacity z = 0, i.e. for a lattice without particles, we have  $g_0 = g_0^2$ . The physical solution to this equation is  $g_0 = 1$ , since  $\mathcal{Z}_{\text{lat}}(z = 0) = 1$ . However, this entails

that  $g_0 > 1$  for nonzero density of particles because z > 0. The formal solution to Equation (2.23) is then  $g_0 = \infty = g_{\psi}$ . To be able to numerically solve for the partition functions in order to sample and draw aggregates later (cf. Subsection B.1.1), we rescale all partial partition functions with  $g_0$ :

$$x_{\psi} \coloneqq \frac{g_{\psi}}{g_0} \tag{2.24}$$

A global prefactor in  $\mathcal{Z}_{lat}$  does not change the relative probability weights of lattice configurations and hence does not alter physical predictions. Thus, we can also redefine the lattice partition function

$$\mathcal{Z}_{\text{lat}} \to \frac{\mathcal{Z}_{\text{lat}}}{g_0^2} = 1 + z \sum_{\psi=1}^4 x_\psi x_{\psi+2} \,.$$
 (2.25)

Employing Eq. (2.23), we arrive at a self-consistency relation for  $x_{\psi}$ :

$$x_{\psi} = \frac{1 + z \sum_{\varphi=1}^{4} (y_{\psi+1,\varphi+2} + y_{\psi,\varphi+3}) x_{\varphi} + z^2 \sum_{\varphi,\nu} y_{\psi+1,\varphi+2} y_{\psi,\nu+3} y_{\varphi+3,\nu+2} x_{\varphi} x_{\nu}}{1 + 2z \sum_{\varphi=1}^{4} x_{\varphi} + z^2 \sum_{\varphi,\nu} y_{\varphi+3,\nu+2} x_{\varphi} x_{\nu}},$$
(2.26)

where we plugged the equations for  $g_0$  and  $g_{\psi}$  into  $x_{\psi} = g_{\psi}/g_0$ , respectively, and then divided both the numerator and denominator by  $g_0^2$  to eliminate all divergences. The lattice partition function  $\mathcal{Z}_{\text{lat}}$  can thus be obtained by solving Eq. (2.26), a system of four cubic equations. While  $\mathcal{Z}$  can be used to sample individual aggregates according to their probability weight,  $\mathcal{Z}_{\text{lat}}$  permits sampling of a whole lattice configuration. Both descriptions enable us to later visualize snapshots of equilibrium self-assembly.

#### 2.4.2 A MAPPING AND PARTICLE DENSITY

We now show how the statistical description of single aggregates from Section 2.3 is related to the one of the whole lattice via a simple mapping. This implies that they represent equivalent mathematical perspectives on our physical model. Moreover, we derive two expressions that relate the density  $\rho$  of particles to the lattice fugacity z and the aggregate fugacity c, respectively. This formally solidifies our choice of grand canonical ensemble, a posteriori.

The mapping has to link both self-consistency relations (2.18) and (2.23). Hence, the rationale is to divide (2.23) by  $g_0^2$  in order to turn the  $g_0^2$  term on the r.h.s. into

a "1". This motivates the ansatz  $G_{\psi} \equiv \frac{g_{\psi}}{g_0^2}$ , which leads to

$$G_{\psi} \equiv \frac{g_0^2}{g_0^2} + zg_0 \sum_{\varphi=1}^4 (y_{\psi+1,\varphi+2} + y_{\psi,\varphi+3}) \frac{g_{\varphi}}{g_0^2} + z^2 \sum_{\varphi,\nu} y_{\psi+1,\varphi+2} y_{\psi,\nu+3} y_{\varphi+3,\nu+2} \frac{g_{\varphi}g_{\nu}}{g_0^2}$$
$$\stackrel{!}{=} 1 + c \sum_{\varphi=1}^4 (y_{\psi+1,\varphi+2} + y_{\psi,\varphi+3}) G_{\varphi} + c^2 \sum_{\varphi,\nu} y_{\psi+1,\varphi+2} y_{\psi,\nu+3} y_{\varphi+3,\nu+2} G_{\varphi} G_{\nu} .$$
(2.27)

Comparing the two expressions we obtain the following mapping:

$$G_{\psi} \equiv \frac{g_{\psi}}{g_0^2}$$
 and  $c \equiv zg_0$  (2.28)

The first identification can also be expressed as

$$cG_{\psi} \equiv zx_{\psi} \,. \tag{2.29}$$

In different terms, the probability to encounter a particle of orientation  $\psi$  in an aggregate is intimately connected to the prevalence of such particles on the whole lattice. This is to be expected, because in our aggregate-centered framework we sample aggregates according to their volume fraction, which amounts to uniformly choosing a random particle on the lattice and analyzing the aggregate to which it belongs. The above mapping shows that the statistics of individual aggregates informs about the statistics of the whole lattice and vice versa.

As a consequence, there exists a unique relation between the density of particles on the whole lattice  $\rho$  and aggregate morphologies quantified by  $G_{\psi}$  and c, for instance. In the lattice-centered description with partition function  $\mathcal{Z}_{\text{lat}} = 1 + z \sum_{\psi} x_{\psi} x_{\psi+2}$ , the first term encodes the probability weight of a given lattice vertex to be empty while the second term corresponds to a nonempty vertex. The probability to find a particle at any given lattice vertex is also given by the volume fraction of particles  $\rho = N/V$ . Formalizing this idea yields

$$\rho = \frac{z \sum_{\psi} x_{\psi} x_{\psi+2}}{1 + z \sum_{\psi} x_{\psi} x_{\psi+2}}.$$
(2.30)

Applying Eq. (2.28) and Eq. (2.29) to z and  $x_{\psi},$  we obtain

$$\rho = \frac{c \sum G_{\psi} G_{\psi+2}}{\frac{z}{c} + c \sum G_{\psi} G_{\psi+2}} = \frac{c \sum G_{\psi} G_{\psi+2}}{1/g_0 + c \sum G_{\psi} G_{\psi+2}}.$$
(2.31)

Dividing the self-consistency condition for  $g_0$ , Eq. (2.23), by  $g_0^2$  and using the mapping again we find

$$1/g_0 = 1 + 2c \sum_{\varphi=1}^4 G_{\varphi} + c^2 \sum_{\varphi,\nu} y_{\varphi+3,\nu+2} G_{\varphi} G_{\nu}.$$
(2.32)

Plugging this back into  $\rho$  finally yields

$$\rho = \frac{c \sum G_{\psi} G_{\psi+2}}{1 + 2c \sum_{\varphi} G_{\varphi} + c^2 \sum_{\varphi,\nu} y_{\varphi+3,\nu+2} G_{\varphi} G_{\nu} + c \sum G_{\psi} G_{\psi+2}}$$

$$= \frac{\mathcal{Z}}{1 + 2c \sum_{\varphi} G_{\varphi} + c^2 \sum_{\varphi,\nu} y_{\varphi+3,\nu+2} G_{\varphi} G_{\nu} + \mathcal{Z}}.$$
(2.33)

This equation provides an implicit definition of the fugacity c in terms of the particle density  $\rho$ .

We now have two equivalent mathematical descriptions for our model in place. As we will see in the following chapter, the aggregate-centered framework from Section 2.3 is best-suited for sampling and visualizing single aggregates in the system and also enables the formal study of the morphology of large aggregates. However, it formally breaks down when infinite aggregates are present in the system. The lattice-centered framework from this section still applies in the latter regime and allows us to also consider infinite aggregates that span the whole system.

# 3

# **Exploring Morphological Diversity**

With the model in place, we want to sample aggregates, probe their morphology and classify them into multiple classes of morphologies accordingly. This chapter sets up the descriptive tools for the statistical analysis of aggregate morphologies. Our goal is to enable the relation of particle properties, namely their interactions, to large scale aggregate morphologies in subsequent chapters.

Aggregate morphologies will at least partially depend on how many particles are put into the system (see Fig. 3.1): If there are almost no particles, then all aggregates will be very small, corresponding to a dilute gas of mostly single-particle-aggregates or monomers. On the contrary, if particles are densely packed on the lattice, we obtain a single aggregate that accommodates all particles. In both cases, aggregate structure only weakly depends on particle interactions.



Figure 3.1: Aggregate morphologies depend on the density of particles on the lattice. Shown are qualitative illustrations of aggregates for two extreme cases: a lattice with almost no particles, exclusively yielding monomers (left), and a densely packed one (right). The bonds of the aggregate for the latter case are colored according to which type of interaction is realized (see Fig. 2.3).

To exhaust the self-assembly phenomenology our model has to offer, it is necessary to identify and quantify the intermediate density regime where nontrivial aggregates can arise. In our model, aggregation behavior is dictated by the interplay of particle density and particle interactions. When there are only a few particles in the system, then they are unlikely to meet and form aggregates. If this entropic cost of adding a particle to an aggregate is counterbalanced by the energetic gain due to favorable interactions, then aggregates become infinite in size. In the associated crossover density regime, the aggregates may range in size all the way from monomers to very large aggregates. The particle density  $\rho_{\rm g}$ , where the first infinitely large aggregate appears in the system, is called the gel point and depends on the particle interactions. At their gel point, systems with mostly repulsive interactions also yield large aggregates (instead of only monomers), whose structure could differ qualitatively from assembled morphologies of attractive interactions. We want to broadly understand how particle interactions dictate aggregate morphology in this nontrivial intermediate density regime.

In Section 3.1, we first visualize a few characteristic aggregates that arise in systems with different interactions. Thereby, it becomes clear that the sizes of aggregates are indeed very sensitive to the particle interactions and the fugacity, which controls the density. Nontrivial morphologies only arise at intermediate particle densities. To systematically investigate this regime, we formally introduce the gel point in Section 3.2. Based on this, we obtain an implicit equation involving the fugacity and the partial partition functions that has to be satisfied at the gel point. By solving this equation and the self-consistency relations from Eq. (2.18) simultaneously, we can probe the self-assembly behavior at the gel point for arbitrary interaction matrices. Thereby, we identify stereotypical categories of aggregate morphology, including fibers and crystals. In Section 3.3, we finally implement formal tools and observables to distinguish between different aggregate morphologies. This enables the automatic classification of aggregates according to their morphology in the following chapter.

### 3.1 Aggregate Size Is Sensitive to Particle Interactions and Fugacity

We want to get a first impression of how particle interactions determine aggregate structure. In this section, we consider a few qualitatively different interaction matrices  $\mathbf{E}$  (cf. Eq. (2.1)) and observe which aggregates they yield in self-assembly, respectively. As formalized in the previous chapter, there are two main quantities that influence the statistical weight of aggregate morphologies. These are their overall energy or Hamiltonian (cf. Section 2.1) and their entropy (cf. Subsection 2.3.1). Thus, we examine systems with weak interactions, which are mainly governed by entropy, as well as ones with strong interactions, whose aggregates are mostly due to the energetic competition of binding motifs.

In order to sample and draw aggregates, we employ a numerical procedure combining Mathematica and Python which we now briefly outline. Our objective is to
visualize a stereotypical aggregate that arises for a given interaction matrix  $\mathbf{E}$  at a fixed lattice fugacity z. For this purpose, we use the lattice-centered formal framework from Section 2.4, because it applies for arbitrary fugacities. For the given example system quantified by  $(\mathbf{E}, z)$ , we solve Eq. (2.26) numerically using Mathematica in order to obtain the numerical value of the partial partition functions  $G_{\psi}$  (see Subsection B.1.1 for further details). Together with the fugacity c and interaction energies  $\mathbf{E}$ , these fix the statistical weight of each term in Eq. (2.18). These self-consistency relations quantify the relative likelihoods for each binding motif to be realized in a lattice triangle, conditioned on the fact that there already is one particle in it. This allows us to recursively sample and draw a whole aggregate. As discussed in Section 2.3, we sample aggregates according to their volume fraction by considering a random particle as the root of an aggregate. This root particle splits its aggregate into two subaggregates. For each of these, we randomly sample a binding motif according to its statistical weight and draw it. Each neighboring particle of the root particle constitutes a new subaggregate in turn. Thus, the same sampling and drawing process can be repeated. When the "1"-term in Eq. (2.18) is sampled, then the corresponding subaggregate terminates. We implement this sampling and drawing procedure in Python, as detailed in Appendix B.2. Therewith, we let Python sample and draw multiple aggregates for the considered example system  $(\mathbf{E}, z)$  randomly. Subsequently, we pick a single aggregate that is qualitatively representative of the system's self-assembly behavior for illustrative purposes.

In Fig. 3.2, we consider three different example systems that exhibit weak interactions between particles. In all cases, the typical size of aggregates increases if more particles are put into the system, i.e. when the fugacity z is increased. Differences between the three systems are most pronounced at fugacity  $z = e^{\beta\mu} = 1$ . There, aggregates are larger the more attractive the interactions are. Moreover, the differently colored bonds show that aggregates are broadly composed of a random combination of binding motifs. This is indicative of the fact that entropy mostly dominates in this regime. However, in the case where particles are weakly repulsive with only a few favorable interactions as in Fig. 3.2b, a slight trend towards fewer distinct binding motifs can be observed. If the energetic malus of unfavorable interactions or conversely the energetic gain due to a favorable interaction is large, the system might exhibit only those few morphologies that exclusively use the energetically most favorable binding motifs.

We study examples of systems with strong interactions in Fig. 3.3. Differences in aggregate morphology are now very pronounced between distinct interaction matrices. For very repulsive particles as in Fig. 3.3b, even at high fugacities the system is dominated by monomers and other small aggregates. On the other hand, in Fig. 3.3c strongly attractive particles form dense space-filling aggregates even at low particle densities. Nontrivial behavior occurs for particles that exhibit both strongly repulsive and attractive interactions as in Fig. 3.3a. In this example, the particles only use a few distinct binding motifs as hypothesized above. In fact, the system strongly



Figure 3.2: Characteristic aggregates for three systems with weak interactions. Each interaction matrix is color coded with blue denoting favorable and red denoting unfavorable interactions. Example aggregates for three different fugacities  $z = e^{\beta\mu}$  with  $\mu \in \{-4, 0, 4\}$  are visualized from top to bottom, corresponding to an increase in the total number of particles in each system. The bond colors correspond to the three different types of binding motifs (see Fig. 2.3). **a**, No interactions, i.e. entropy governs the assembly. **b**, Weakly repulsive interactions. The system exhibits more copies of the only favorable binding motif. **c**, Weakly attractive interactions. Typical aggregates are larger than in the two other example systems.

favors cyclic binding motifs (signified by yellow bonds) while the single most favorable interaction is actually a head-to-tail motif, which is colored in blue-green (refer to Fig. 2.3 for details on the color code). This exemplifies that aggregate structure is not trivially related to the interaction matrix  $\mathbf{E}$ .

From the six example systems discussed above, we learnt that systems with repulsive interactions or few particles exhibit mostly small aggregates like monomers, while systems with attractive interactions or many particles yield dense space-filling aggregates. However, there are also hints at the fact that for "intermediate" fugacities, a non-trivial competition between individual binding motifs determines aggregate morphology. Hence, we expect aggregate morphology to be richer in this regime. This notion of "intermediate" fugacity depends on the interactions between particles. In the following, we formally identify and quantify it through the gel point.

## 3.2 Aggregate Morphologies Are Diverse at the Gel Point

Precisely at the gelation point,  $\rho = \rho_{\rm g}$ , an aggregate that percolates to infinity arises, the so-called "gel". At and below the gel point, the volume fraction  $\phi_{\infty}$  of the gel is zero and only begins to continuously rise once the particle density  $\rho$  in the system is increased beyond the gelation density  $\rho_{g}$ . Thus, the gel point marks a continuous second-order phase transition of the system as a function of  $\rho$ . Gelated solutions are mostly fluid, meaning almost all aggregates have a finite size, and that size can strongly vary. Therefore, arising aggregate morphologies are potentially nontrivial, i.e. neither monomers nor aggregates which fill the whole lattice. The chemists Paul Flory and Walter Stockmayer were the first to predict the onset of gelation in a theory of crosslinked polymers in the 1940s, termed Flory-Stockmayer theory [52, 53]. In Subsection 3.2.1, we introduce their definition of the gel point, which is constituted by the divergence of mean aggregate size. Aggregate size is accessible within our aggregate-centered formalism (cf. Section 2.3, which is therefore ideally suited to translate the definition of the gel point into our model. We find a formal expression for the mean aggregate size. Subsequently, we rewrite the gelation condition in order to derive equivalent characterizations for the gel point in Subsection 3.2.2. Thereby, we obtain an equation that relates the aggregate fugacity at gelation  $c_{g}$  to the corresponding partial partition functions  $G_{\psi}$  (cf. Section 2.3). This equation complements the self-consistent system from Eq. (2.18). Aggregate morphologies at gelation are thus formalized within five coupled equations in the five unknowns  $c_{\sigma}$ and  $G_{\psi}$  ( $\psi \in \{1, 2, 3, 4\}$ ). Lastly, we probe the self-assembly behavior at gelation for different interactions in Subsection 3.2.3 by sampling and drawing random aggregates according to their statistical weight. We observe 8 different stereotypical categories of aggregates, including morphologies with a well-defined structure, for instance, fibers and crystals.



Figure 3.3: Stereotypical aggregates for three systems with strong interactions. The color codes for the interaction matrices and the bonds are the same as in Fig. 3.2. Again, example aggregates for three different fugacities  $z = e^{\beta\mu}$  with  $\mu \in \{-4, 0, 4\}$  are visualized from top to bottom. **a**, Particles have both strongly repulsive and attractive interactions. Dense space-filling aggregates of mostly cyclic binding motifs emerge (in yellow). **b**, Strongly repulsive interactions. Even at high fugacity, the aggregates are small. **c**, Strongly attractive interactions. For all considered fugacities, the aggregates tile the lattice and are composed of many different binding motifs.

### 3.2.1 Formal Definition of the Gel Point

Let  $\phi_n$  be the volume fraction of aggregates of size n. Then  $\rho = \sum_{n=1}^{\infty} \phi_n$  and the mean aggregate size is defined as

$$\langle n \rangle \coloneqq \sum_{n=1}^{\infty} n \frac{\phi_n}{\rho} \,. \tag{3.1}$$

Note that this definition corresponds to a mass average instead of a number average, because the aggregate size n is weighted by a volume (or mass) fraction  $\phi_n$ . At the gel point, a transition from a system of only finitely sized aggregates to one with an infinite gel occurs. Thus, the gelation transition mathematically corresponds to the divergence of the mean aggregate size,

$$\langle n \rangle \to \infty$$
. (3.2)

We now derive a formal expression for  $\langle n \rangle$  in the language of our aggregate-centered framework (cf. Section 2.3). Since we are going to use a similar approach to obtain other observables in Section 3.3 as well, we guide through the detailed calculation below.

The volume fractions  $\phi_n$  are proportional to the corresponding statistical weight  $Z_n c^n$  in the partition function  $\mathcal{Z} = \sum_{n=1}^{\infty} Z_n c^n$ . Here,  $Z_n$  is the canonical partition function of aggregates of fixed size n (cf. Section 2.3). So,  $\langle n \rangle$  is given by

$$\langle n \rangle = \sum_{n=1}^{\infty} n \frac{Z_n c^n}{\mathcal{Z}} = \sum_{n=1}^{\infty} \frac{Z_n c \frac{\mathrm{d}}{\mathrm{d}c} c^n}{\mathcal{Z}} = \frac{c}{\mathcal{Z}} \frac{\mathrm{d}}{\mathrm{d}c} \mathcal{Z} = \frac{\mathrm{d}\ln\mathcal{Z}}{\mathrm{d}\ln c} \,. \tag{3.3}$$

Using  $\mathcal{Z} = c \sum_{\psi} G_{\psi} G_{\psi+2}$  (see Eq. (2.19)) and the Leibniz product rule, we can also write  $\langle n \rangle$  as

$$\langle n \rangle = 1 + 2c \frac{\sum_{\psi} \frac{\mathrm{d}G_{\psi}}{\mathrm{d}c} G_{\psi+2}}{\sum_{\psi} G_{\psi} G_{\psi+2}} \,. \tag{3.4}$$

To proceed, note that  $G_{\psi}$  is given by self-consistency relations that we have derived in Section 2.3 (see Eq. (2.18)):

$$G_{\psi} = 1 + c \sum_{\varphi=1}^{4} (y_{\psi+1,\varphi+2} + y_{\psi,\varphi+3}) G_{\varphi} + c^2 \sum_{\varphi,\nu} y_{\psi+1,\varphi+2} y_{\psi,\nu+3} y_{\varphi+3,\nu+2} G_{\varphi} G_{\nu} \quad (3.5)$$

Accordingly, the derivative  $\frac{dG_{\psi}}{dc}$  can be obtained in terms of  $G_{\psi}$  via implicit differentiation of these self-consistency relations. Let **G** denote the vector of partition functions

$$\boldsymbol{G} \coloneqq (G_1, G_2, G_3, G_4)^\top$$
 (3.6)

A short calculation shows that  $\frac{d\mathbf{G}}{dc}$  obeys the following inhomogeneous linear matrix equation:

$$\mathbf{A} \cdot \frac{\mathrm{d}\boldsymbol{G}}{\mathrm{d}\boldsymbol{c}} = \boldsymbol{b}^{(c)}, \qquad (3.7)$$

where  $\cdot$  denotes the matrix product throughout this thesis. The entries of the matrix  $\mathbf{A} = (a_{\psi\varphi})$  are

$$a_{\psi\varphi} = \delta_{\psi\varphi} - c(y_{\psi+1,\varphi+2} + y_{\psi,\varphi+3}) - c^2 \sum_{\nu} (y_{\psi+1,\varphi+2} y_{\psi,\nu+3} y_{\varphi+3,\nu+2} + y_{\psi+1,\nu+2} y_{\psi,\varphi+3} y_{\nu+3,\varphi+2}).$$
(3.8)

and the inhomogeneous term on the r.h.s.  $\boldsymbol{b}^{(c)}$  is given by

$$b_{\psi}^{(c)} = \sum_{\varphi} (y_{\psi+1,\varphi+2} + y_{\psi,\varphi+3}) G_{\varphi} + 2c \sum_{\varphi,\nu} y_{\psi+1,\varphi+2} y_{\psi,\nu+3} y_{\varphi+3,\nu+2} G_{\varphi} G_{\nu} \,. \tag{3.9}$$

Linear systems as in Eq. (3.7) are solved by

$$\frac{\mathrm{d}\boldsymbol{G}}{\mathrm{d}\boldsymbol{c}} = \mathbf{A}^{-1} \cdot \boldsymbol{b}^{(c)} = \frac{\mathrm{adj}(\mathbf{A}) \cdot \boldsymbol{b}^{(c)}}{\mathrm{det}\,\mathbf{A}}, \qquad (3.10)$$

where  $\operatorname{adj}(\mathbf{A})$  is the adjugate matrix of  $\mathbf{A}$ . Using this expression of  $\frac{\mathrm{d}\mathbf{G}}{\mathrm{d}c}$  in terms of the partial partition functions  $G_{\psi}$ , we now write the mean aggregate size  $\langle n \rangle$  without derivatives. Introducing the following notation

$$\boldsymbol{G}^{(+2)} \coloneqq (G_3, G_4, G_1, G_2) , \qquad (3.11)$$

we can use the matrix product to rewrite the sums, e.g.

$$\sum_{\psi} G_{\psi} G_{\psi+2} = \boldsymbol{G}^{(+2)} \cdot \boldsymbol{G} \,. \tag{3.12}$$

Summarizing, we finally arrive at the following characterization of the gel point in our system:

$$\langle n \rangle = 1 + \frac{2c}{\det \mathbf{A}} \frac{\mathbf{G}^{(+2)} \cdot \left( \operatorname{adj}(\mathbf{A}) \cdot \mathbf{b}^{(c)} \right)}{\mathbf{G}^{(+2)} \cdot \mathbf{G}} \to \infty$$
(3.13)

Since the self-consistency relations, Eq. (2.18), can only be solved for  $G_{\psi}$  numerically, this form is convenient, because it does not involve derivatives of a priori unknown quantities. In the above equation, the first term corresponds to a system of only monomers. The second term describes the contribution of all other aggregate sizes to the mean. In the next subsection, we show that the divergence of mean aggregate size is associated to det  $\mathbf{A} \to 0$ .

# 3.2.2 The Gelation Condition Augments the System's Governing Equations

In this subsection, we translate the divergence of  $\langle n \rangle$  at gelation into a condition for the corresponding fugacity  $c_{\rm g}$ .

First, gelation almost always occurs at a finite fugacity that is related to the *n*-dependence of the canonical partition functions  $Z_n$  (cf. Eq. (3.1)). In particular,  $c_{\rm g}$  is finite for particles that are not infinitely repulsive, i.e. the interaction matrix has no infinite entries ( $E \to \infty$  corresponds to  $y = e^{-\beta E} \to 0$ ). This holds for all biological self-assembly systems. The Hamiltonian and the total chemical potential of all particles in an aggregate,  $n\mu$ , are extensive in aggregate size n. Thus, there is a competition between the interactions **E** and the fugacity c that determines the aggregation behavior and, thereby, the gel point. Based on this, Appendix A.2.1 provides an upper bound on  $c_{\rm g}$ . In what follows we are going to frequently refer to the self-consistency relations between the partial partition functions  $G_{\psi}$  (see Eq. (3.5); recall Section 2.3 for the derivation). For each fugacity  $c \in [0, c_{\rm g}]$ , there exists a unique physical solution to these self-consistency relations: Differentiating Eq. (3.5) with respect to the fugacity c defines an initial value problem for  $c \in [0, c_{\rm g}]$  with initial condition  $G_{\psi} = 1$  for all  $\psi \in \{1, 2, 3, 4\}$  at c = 0. By the Picard-Lindelöf theorem for ordinary differential equations, the physical solution is then unique for all c. Thus, Eq. (3.5) represents a line of physical solutions in 5d space:

$$(c, G_1(c), G_2(c), G_3(c), G_4(c)) (3.14)$$

In biological systems, interaction energies are never actually infinite,  $-\infty < E < \infty$ . In particular, this implies that  $y = e^{-\beta E} > 0$ . Due to this, the  $G_{\psi}$ 's are all strictly monotonic in c and with respect to each other, because every term on the r.h.s. of the self-consistency relations is positive (cf. Eq. (3.5)). As a consequence, we have bijectivity between all 5 variables and we can equivalently interpret the fugacity c as a function of the partial partition functions

$$c \equiv c(G_{\psi}) \,. \tag{3.15}$$

This point of view enables us to find a condition on the gelation fugacity  $c_{\rm g}$  below.

To begin with,  $G_{\psi} \geq 1$  for all fugacities  $c = e^{\beta\mu} \geq 0$  due to monotonicity and  $G_{\psi}(c = 0) = 1$ . As a consequence, we find the following necessary condition for gelation using Eqs. (2.19) and (3.3):

$$\langle n \rangle = \frac{1}{\sum_{\psi} G_{\psi} G_{\psi+2}} \frac{\mathrm{d}}{\mathrm{d}c} \mathcal{Z} \le \frac{\mathrm{d}}{\mathrm{d}c} \mathcal{Z} \qquad \text{i.e.} \qquad \frac{\mathrm{d}}{\mathrm{d}c} \mathcal{Z} \to \infty \quad \text{at gelation}$$
(3.16)

This condition is also sufficient, which we show in Appendix A.2.2 based on the fact

that  $\mathcal{Z}$  is a sum of positive probability weights:

$$\langle n \rangle \to \infty \qquad \Leftrightarrow \qquad \frac{\mathrm{d}}{\mathrm{d}c} \mathcal{Z} \to \infty$$

$$(3.17)$$

Since all partial partition functions are strictly monotonic functions of each other, one can even conclude that (see Subsection A.2.2)

$$\langle n \rangle \to \infty \qquad \Leftrightarrow \qquad \frac{\mathrm{d}}{\mathrm{d}c} G_{\psi} \to \infty \quad \text{for all } \psi \in \{1, 2, 3, 4\}.$$
 (3.18)

In principle, we can use this property on the self-consistency relation (3.5), by differentiating it with respect to c. However, it is tedious to deal with the associated infinities. In particular, the divergence behavior  $\frac{d}{dc}G_{\psi} \to \infty$  at the gel point depends on orientation  $\psi$ . Instead, thanks to bijectivity of  $c \equiv c(G_{\psi})$  we also have:

$$\langle n \rangle \to \infty \qquad \Leftrightarrow \qquad \frac{\mathrm{d}c}{\mathrm{d}G_{\psi}} \to 0 \quad \text{for all } \psi \in \{1, 2, 3, 4\}.$$
 (3.19)

By means of this condition, below we obtain an implicit equation that relates  $c_g$  to **E** and the partial partition functions  $G_{\psi}$ . If we differentiate Eq. (3.5) with respect to  $G_1$  and employ Eq. (3.19), we get a homogeneous system of linear equations involving the same matrix **A** as in the previous subsection:

$$\mathbf{A} \cdot \frac{\mathrm{d}}{\mathrm{d}G_1} \boldsymbol{G} = 0 \tag{3.20}$$

All implicit derivatives of G are always related to this same matrix  $\mathbf{A}$ , because they are derived from the same set of self-consistency relations (also see Section 3.3 on our morphological observables). Homogeneous linear systems always allow the trivial zero solution,  $\frac{d}{dG_1}G = 0$ , which is also the only solution if and only if the matrix  $\mathbf{A}$ is non-singular, i.e. det  $\mathbf{A} \neq 0$ . This trivial solution is not physical, because  $\frac{dG_1}{dG_1} \stackrel{!}{=} 1$ . Consequently, singularity of  $\mathbf{A}$  is a necessary condition for gelation:

$$\langle n \rangle \to \infty \quad \Rightarrow \quad \det \mathbf{A} = 0$$
 (3.21)

Based on heuristic evidence, det  $\mathbf{A} = 0$  coupled to the self-consistency relations, Eq. (3.5), even seems to be equivalent to gelation. However, we were unable to show this generally. As an alternative, we can explicitly check that det  $\mathbf{A} = 0$  implies  $\langle n \rangle \to \infty$  by using Eq. (3.13).

For interaction matrices with finite entries  $\mathbf{E}$ , We are now able to sample and draw large aggregates that occur at the gel point by solving a coupled system of 5

polynomial equations for the aggregate fugacity c and the partial partition functions  $G_{\psi}$  (see Eqs. (3.5) and (3.21); cf. Appendices B.1.2 and B.2).

# 3.2.3 Stereotypical Categories of Morphologies at Gelation

Section 3.1 has illustrated how a given interaction matrix  $\mathbf{E}$  and a fixed lattice fugacity z broadly govern the size of aggregates, often trivially. In this subsection, we visualize large nontrivial aggregate morphologies that can arise for different interactions  $\mathbf{E}$  at the gel point. We thereby identify 8 stereotypical categories of aggregates in our model.

To sample aggregates at the gel point for given interactions **E**, we simultaneously solve the self-consistency relations for the partial partition functions and the fugacity at gelation (see Eqs. (2.18) and (3.21)) numerically using Mathematica as detailed in Appendix B.1.2. At the gel point, aggregate sizes vary strongly. For different interaction matrices, the resulting morphologies are most different for large aggregates. Hence, we sample multiple aggregates and, among these, arbitrarily choose a representative large aggregate. We repeat this procedure for 8 different interaction matrices and visualize the corresponding aggregates in Fig. 3.4. There are conspicuous morphological differences, both in the overall aggregate structures as well as in the binding motifs used. We accordingly assign each aggregate to one of 8 stereotypical categories of aggregate morphology. A "fiber" is defined as a slim, linear aggregate. Notably, fibers are naturally formed through one of the two head-to-tail interactions on the third off-diagonal of the interaction matrix (also see Fig. 2.3). "Fiber networks" are derived from shorter fibers that are joined at triangles, which may constitute branching points of the fiber network. We also identify a curved "fiber of triangles", which is one of the elementary aggregate morphologies that we have already seen in Section 2.2 and as such built from copies of a single triangle binding motif. However, if particle interactions allow it, then more triangle motifs can attach to these curved fibers, which yields a "decorated fiber". In addition to mostly linear fibrous aggregates, we also observe more bulky ones that contain more triangles. "Crystals" are periodic structures of two different triangle binding motifs and grow in every lattice direction. For example, the crystal in Fig. 3.4 arises from two cyclic triangle motifs only, namely  $\{(1,2), (1,2), (1,2)\}$  and  $\{(3,4), (3,4), (3,4)\}$ . In a "sponge" only every other binding motif is a triangle motif. On a 2d lattice, this aggregate would correspond to a crystal with holes. All six above classes represent well-defined morphologies that exhibit some bias in their binding motifs. On the other hand, "flocs" and "liquids" are disordered aggregates that are composed of a seemingly random variety of binding motifs. Liquids contain many triangle motifs, and flocs have fewer ones. These two categories are more ill-defined and broadly include all aggregates that do not fall into any of the six categories above.

With 8 stereotypical aggregate categories in place, we next aim to characterize ag-



Figure 3.4: Stereotypical categories of morphologies at gelation. Each interaction matrix  $\mathbf{E}$  is colorcoded with blue denoting favorable and red denoting unfavorable interactions (cf. Eq. (2.1)). The colors in the aggregate drawings correspond to the three different types of interactions (cf. Fig. 2.3): head-to-head bonds are colored in purple, head-to-tail bonds in blue-green, and cyclic bonds in yellow.

gregates using physical observables that we can measure in our formalism. Ultimately, we use them to quantitatively define the aggregate categories and to formally classify aggregates that result from given interactions  $\mathbf{E}$  into one of these categories.

## 3.3 Observables of Aggregate Morphology

To distinguish different aggregate morphologies, we now provide a suitable set of aggregate observables. While this section is quite technical, its results are going to be essential for the automatic classification of a large number of aggregates into one of the above categories (see Fig. 3.4) based only on numerical results for the fugacity  $c_{\rm g}$  and the partial partition functions  $G_{\psi}$  in Chapter 4.

Using the partition function formalism, we can calculate the average of any observable that can be made explicit within the self-consistent governing equations of aggregates, Eq. (2.18), via the use of some dummy variable. In Subsection 3.2.1, we calculated mean aggregate size  $\langle n \rangle$  using a derivative of the partition function with respect to the fugacity c, because each particle in an aggregate is made explicit with a factor of c. However, at gelation  $\langle n \rangle \to \infty$ , regardless of aggregate category. Thus, we have to rely on other descriptors of aggregate morphology, but we can still employ suitable derivatives of the partition function to probe other observables of large aggregates. In fact, the divergence of mean aggregate size at gelation, although aggregates of all sizes occur in the system, indicates that averages of observables at the gel point are dominated by large aggregates. Thus, statistical averages are well-suited to quantify the morphology of large aggregates.

Our choice of observables is illustrated in Fig. 3.5. The number of triangles in an aggregate is indicative of how bulky it is (see Fig. 3.5a). Therefore, it distinguishes slim, straight fibers from bulky, curved ones, for example. However, there are other morphologies that contain a lot of triangles, e.g. crystals. So, we use branching events as an indicator of how fibrous an aggregate is (see Fig. 3.5b). Curved fibers of triangles rarely branch and instead grow mainly linearly. Crystals, on the other hand, branch often and explore all lattice directions equally. Lastly, crystals and liquids both have a lot of triangles and many branchings. Crystals are made up of only a few different interactions that repeat periodically, while liquids use a random mixture of interactions. The relative abundances of all possible interactions  $E_{\psi\varphi}$  in an aggregate define a probability distribution. By referring to the associated Shannon entropy, we can distinguish ordered aggregates like crystals from disordered ones like liquids (see Fig. 3.5c). From a formal viewpoint, the average number of triangles, branchings, and bonds per aggregate goes to infinity at the gel point, so we consider ratios of observables instead.

In Subsection 3.3.1, we compute the triangle density  $\frac{\langle t \rangle}{\langle n \rangle}$ , i.e. the number of triangles per particle in large aggregates. Subsequently, we determine the branching ratio  $\frac{\langle b \rangle}{\langle t \rangle}$ 



Figure 3.5: **Observables of aggregate morphology. a**, Slim fibers (left) can be distinguished from more bulky curved fibers of triangles (right) via the number of triangles in each respective aggregate. **b**, The number of branchings discriminates between mostly linear aggregates like fibers of triangles (left) and space-filling crystals (right). **c**, Bond entropy can be used to discern ordered aggregates like crystals (left), which only use a few different interactions, from disordered ones like liquids (right), which are constituted by a wide variety of interactions.

in Subsection 3.3.2, which describes the probability of a triangle in a large aggregate constituting a branching point. Lastly, in Subsection 3.3.3, we calculate the bond entropy  $S_{\text{bonds}}$  of large aggregates based on the relative abundances of individual interactions  $E_{\psi\varphi}$ .

## 3.3.1 TRIANGLE DENSITY

To obtain the triangle density  $\frac{\langle t \rangle}{\langle n \rangle}$ , we must first compute the mean number of triangles  $\langle t \rangle$ . Generally, in statistical physics, observables can be computed by differentiating the free energy with respect to their conjugate variable. For instance, in the Ising model, the external magnetic field is the conjugate variable of the magnetization. Based on this analogy, we introduce a dummy factor T to each term in the self-consistency relation, Eq. (2.18), that corresponds to a triangle:

$$G_{\psi} = 1 + c \sum_{\varphi=1}^{4} (y_{\psi+1,\varphi+2} + y_{\psi,\varphi+3}) G_{\varphi} + Tc^2 \sum_{\varphi,\nu} y_{\psi+1,\varphi+2} y_{\psi,\nu+3} y_{\varphi+3,\nu+2} G_{\varphi} G_{\nu} \quad (3.22)$$

Similarly to the mean aggregate size (cf. Subsection 3.2.1), we compute the mean number of triangles  $\langle t \rangle$  by differentiating the partition function  $\mathcal{Z}$  with respect to T:

$$\langle t \rangle = \left. \frac{T}{\mathcal{Z}} \frac{\mathrm{d}\mathcal{Z}}{\mathrm{d}T} \right|_{T=1} = \left. \frac{\mathrm{d}\ln\mathcal{Z}}{\mathrm{d}\ln T} \right|_{T=1}$$
(3.23)

Due to  $\mathcal{Z} = c \sum_{\psi} G_{\psi} G_{\psi+2}$  (see Eq. (2.19)), we can write  $\langle t \rangle$  in terms of the partial partition functions

$$\langle t \rangle = \frac{2Tc}{\mathcal{Z}} \sum_{\psi} G_{\psi+2} \frac{\mathrm{d}}{\mathrm{d}T} G_{\psi} \bigg|_{T=1} .$$
 (3.24)

Implicit differentiation of Eq. (3.22) with respect to T yields a linear system of equations

$$\mathbf{A} \cdot \frac{\mathrm{d}\boldsymbol{G}}{\mathrm{d}T} = \boldsymbol{b}^{(T)} \tag{3.25}$$

with the same matrix  $\mathbf{A}$  as in Eq. (3.8)), but now there is another inhomogeneous term

$$b_{\psi}^{(T)} = c^2 \sum_{\varphi,\nu} y_{\psi+1,\varphi+2} y_{\psi,\nu+3} y_{\varphi+3,\nu+2} G_{\varphi} G_{\nu} \,. \tag{3.26}$$

Following the same formal steps as in Subsection 3.2.1, we obtain

$$\langle t \rangle = \frac{2}{\det \mathbf{A}} \frac{\mathbf{G}^{(+2)} \cdot \left( \operatorname{adj}(\mathbf{A}) \cdot \mathbf{b}^{(T)} \right)}{\mathbf{G}^{(+2)} \cdot \mathbf{G}}$$
(3.27)

Combining this expression with our result for mean aggregate size  $\langle n \rangle$  (see Eq. (3.13)), we finally find the triangle density:

$$\frac{\langle t \rangle}{\langle n \rangle} = \frac{1}{c} \frac{\boldsymbol{G}^{(+2)} \cdot \left( \operatorname{adj}(\mathbf{A}) \cdot \boldsymbol{b}^{(T)} \right)}{\boldsymbol{G}^{(+2)} \cdot \left( \operatorname{adj}(\mathbf{A}) \cdot \boldsymbol{b}^{(c)} \right)}$$
(3.28)

### 3.3.2 BRANCHING RATIO

Before deriving an expression for the branching ratio  $\frac{\langle b \rangle}{\langle t \rangle}$ , we define what constitutes a branching in an aggregate and introduce a counting mechanism.



Figure 3.6: The notion of a branching point. A triangle constitutes a branching point (green dots), if each of its particles also has neighboring particles outside of that triangle. If the root particle sits at the boundary of the aggregate (purple dot in the left panel), then its triangle does not represent a branching point (purple cross), according to our definition of a branching. This case is covered in Eq. (3.32).

As visualized in Fig. 3.6, only a triangle motif, each of whose particles has at least one more neighboring particle outside of that triangle, corresponds to a branching event of the aggregate.

To formalize the number of branchings in an aggregate, we have to modify Eq. (2.18) to be able to identify whether a triangle is also a branching point or not. For this purpose we define

$$G_{\psi} \coloneqq 1 + \tilde{G}_{\psi} \,. \tag{3.29}$$

 $\tilde{G}_{\psi}$  is the partial partition function  $G_{\psi}$  minus the contribution of subaggregate termination. In other words,  $\tilde{G}_{\psi}$  is the statistical weight for a subaggregate that accommodates at least one particle. Based on Eq. (2.18), we write a self-consistent system for  $\tilde{G}_{\psi}$  that associates the dummy variable B to a branching event:

$$\begin{split} \tilde{G}_{\psi} &= c \sum_{\varphi} (y_{\psi+1,\varphi+2} + y_{\psi,\varphi+3}) (1 + \tilde{G}_{\varphi}) \\ &+ c^2 \sum_{\varphi,\nu} y_{\psi+1,\varphi+2} y_{\psi,\nu+3} y_{\varphi+3,\nu+2} (1 + \tilde{G}_{\varphi} + \tilde{G}_{\nu} + B \tilde{G}_{\varphi} \tilde{G}_{\nu}) \end{split} \tag{3.30}$$

However, the root particle of the subaggregate described by  $\tilde{G}_{\psi}$  does not necessarily have more neighbors outside this subaggregate. Therefore, the partition function of

the whole aggregate  $\mathcal Z$  is not simply given by

$$\begin{aligned} \mathcal{Z} &\neq c \sum_{\psi} G_{\psi} G_{\psi+2} = c \sum_{\psi} (1 + \tilde{G}_{\psi}) (1 + \tilde{G}_{\psi} + 2) \\ &= c \sum_{\psi} (1 + \tilde{G}_{\psi} + \tilde{G}_{\psi+2} + \tilde{G}_{\psi} \tilde{G}_{\psi+2}) \,, \end{aligned} \tag{3.31}$$

because the terms  $\tilde{G}_{\psi}$  and  $\tilde{G}_{\psi+2}$  in the last equality do not actually correspond to a branching event. This is also visualized on the left of Fig. 3.6. To cover this case, we introduce another partial partition function  $\bar{G}_{\psi}$  that does not include the dummy variable B:

$$\begin{split} \bar{G}_{\psi} &\coloneqq c \sum_{\varphi} (y_{\psi+1,\varphi+2} + y_{\psi,\varphi+3}) (1 + \tilde{G}_{\varphi}) \\ &+ c^2 \sum_{\varphi,\nu} y_{\psi+1,\varphi+2} y_{\psi,\nu+3} y_{\varphi+3,\nu+2} (1 + \tilde{G}_{\varphi} + \tilde{G}_{\nu} + \tilde{G}_{\varphi} \tilde{G}_{\nu}) \end{split}$$
(3.32)

The correct expression for the partition function  $\mathcal{Z}$  is then

$$\mathcal{Z} = c \sum_{\psi} \left( 1 + \bar{G}_{\psi} + \bar{G}_{\psi+2} + \tilde{G}_{\psi} \tilde{G}_{\psi+2} \right) \,. \tag{3.33}$$

Similarly to the last subsection, the mean number of branchings in an aggregate  $\langle b \rangle$  can then be obtained via

$$\langle b \rangle = \frac{B}{\mathcal{Z}} \frac{\mathrm{d}\mathcal{Z}}{\mathrm{d}B} \Big|_{B=1} = \frac{\mathrm{d}\ln\mathcal{Z}}{\mathrm{d}\ln B} \Big|_{B=1}$$

$$= \frac{2Bc}{\mathcal{Z}} \sum_{\psi} \left( \frac{\mathrm{d}}{\mathrm{d}B} \bar{G}_{\psi} + \tilde{G}_{\psi+2} \frac{\mathrm{d}}{\mathrm{d}B} \tilde{G}_{\psi} \right) \Big|_{B=1} ,$$

$$(3.34)$$

where we have used Eq. (3.33) in the last step. Analogously to before, implicit differentiation of Eq. (3.30) yields an inhomogeneous linear matrix equation for  $\tilde{G}$ :

$$\mathbf{A} \cdot \frac{\mathrm{d}\tilde{\boldsymbol{G}}}{\mathrm{d}B} = \boldsymbol{b}^{(B)} \tag{3.35}$$

with  $\mathbf{A}$  as in Eq. (3.8) and

$$b_{\psi}^{(B)} = c^2 \sum_{\varphi,\nu} y_{\psi+1,\varphi+2} y_{\psi,\nu+3} y_{\varphi+3,\nu+2} \tilde{G}_{\varphi} \tilde{G}_{\nu} \,. \tag{3.36}$$

Moreover, we find

$$\frac{\mathrm{d}}{\mathrm{d}B}\bar{G}_{\psi} = \frac{\mathrm{d}}{\mathrm{d}B}\tilde{G}_{\psi} - b_{\psi}^{(B)}.$$
(3.37)

Putting everything together, the mean number of branchings can be brought into the

following form:

$$\langle b \rangle = \frac{2}{\det \mathbf{A}} \frac{\mathbf{G}^{(+2)} \cdot \left( \operatorname{adj}(\mathbf{A}) \cdot \mathbf{b}^{(B)} \right) - \det \mathbf{A} \sum_{\psi} b_{\psi}^{(B)}}{\mathbf{G}^{(+2)} \cdot \mathbf{G}}$$
(3.38)

Recalling the formula for the mean number of triangles  $\langle t \rangle$ , Eq. (3.27), and using that det  $\mathbf{A} = 0$  (cf. Eq. (3.21)), we finally obtain the branching ratio at gelation:

$$\frac{\langle b \rangle}{\langle t \rangle} = \frac{\boldsymbol{G}^{(+2)} \cdot \left( \operatorname{adj}(\mathbf{A}) \cdot \boldsymbol{b}^{(B)} \right)}{\boldsymbol{G}^{(+2)} \cdot \left( \operatorname{adj}(\mathbf{A}) \cdot \boldsymbol{b}^{(T)} \right)}$$
(3.39)

#### 3.3.3 Bond Entropy

We can characterize aggregates according to their global structure through the triangle density and the branching ratio. Lastly, we now consider one possible notion of aggregate order, which will be instrumental to distinguishing between well-defined ordered morphologies from random disordered ones in Chapter 4.

Figure 3.5c visualizes one typical distinction that we want to make, namely between crystals that use only a few different interactions and liquids that employ a wide variety of different interactions. The underlying theme is that the interactions between neighboring particles are determined by their orientation. Therefore, the realized interactions are closely related to the orientational order of particles in the aggregate. In a crystal, particle orientations follow periodic patterns, while in a liquid, particles are randomly oriented.

For the purpose of our discussion, we choose the relative abundance of different interactions in an aggregate as a descriptor of the orientational order of its particles. Consider a random bond in an aggregate. We are going to calculate the probability for each individual interaction to be realized at that bond and subsequently find the Shannon entropy of the associated probability distribution, which we term bond entropy  $S_{\text{bonds}}$ . For a crystal, the probability weights are concentrated in a few interactions corresponding to a low bond entropy, while in the case of a liquid, the distribution is closer to uniform, yielding a higher bond entropy.

Let  $n_{ab}$  denote the total number of bonds between faces a and b in an aggregate  $(a, b \in \{1, 2, 3, 4\})$ . The (a, b) interaction and the (b, a) interaction lead to the same bond due to the symmetry of the interaction matrix. We therefore consider bonds as unordered tuples  $\{a, b\}$ . The probability  $p_{ab}$  for the bond  $\{a, b\}$  to be realized at a randomly chosen bond of a large aggregate is then approximately equal to

$$p_{ab} \simeq \frac{\langle n_{ab} \rangle}{\sum_{i \le j} \langle n_{ij} \rangle} \,, \tag{3.40}$$

because  $\langle n_{ab} \rangle$  is dominated by large aggregates. Note that we only sum over  $i \leq j$  to

avoid overcounting of bonds. Hence, to obtain the following (Shannon) bond entropy

$$S_{\text{bonds}} = -\sum_{a \le b} p_{a,b} \ln p_{a,b} , \qquad (3.41)$$

we have to find  $\langle n_{ab} \rangle$ .

In our aggregate-centered formalism (cf. Section 2.3), each  $\{a, b\}$ -bond is associated to either the Boltzmann weight  $y_{ab} = e^{-\beta E_{ab}}$  or  $y_{ba}$ . Thus, we have

$$\langle n_{ab} \rangle = \frac{y_{ab}}{\mathcal{Z}} \frac{\mathrm{d}\mathcal{Z}}{\mathrm{d}y_{ab}} + (1 - \delta_{ab}) \frac{y_{ba}}{\mathcal{Z}} \frac{\mathrm{d}\mathcal{Z}}{\mathrm{d}y_{ba}} = \frac{\mathrm{d}\ln\mathcal{Z}}{\mathrm{d}\ln y_{ab}} + (1 - \delta_{ab}) \frac{\mathrm{d}\ln\mathcal{Z}}{\mathrm{d}\ln y_{ba}} \,. \tag{3.42}$$

Following the same procedure as before to compute the derivative of  $\mathcal{Z}$ , one finds

$$\langle n_{ab} \rangle = \frac{2y_{ab}}{\det \mathbf{A}} \frac{\mathbf{G}^{(+2)} \cdot \left( \operatorname{adj}(\mathbf{A}) \cdot \left( \mathbf{b}^{(n_{ab})} + (1 - \delta_{ab} \mathbf{b}^{(n_{ba})}) \right) \right)}{\mathbf{G}^{(+2)} \cdot \mathbf{G}} \,. \tag{3.43}$$

We obtain the inhomogeneous term  $b_{\psi}^{(n_{ab})}$  by using

$$\frac{\mathrm{d}}{\mathrm{d}y_{ab}}y_{\varphi\nu} = \delta_{a\varphi}\delta_{b\nu} \tag{3.44}$$

and the product rule of differentiation:

$$\begin{split} b_{\psi}^{(n_{ab})} &= c \sum_{\varphi} G_{\varphi} \frac{\mathrm{d}}{\mathrm{d}y_{ab}} (y_{\psi+1,\varphi+2} + y_{\psi,\varphi+3}) \\ &+ c^{2} \sum_{\varphi,\nu} G_{\varphi} G_{\nu} \frac{\mathrm{d}}{\mathrm{d}y_{ab}} (y_{\psi+1,\varphi+2} y_{\psi,\nu+3} y_{\varphi+3,\nu+2}) \\ &= c \sum_{\varphi} G_{\varphi} (\delta_{a,\psi+1} \delta_{b,\varphi+2} + \delta_{a,\psi} \delta_{b,\varphi+3}) \\ &+ c^{2} \sum_{\varphi,\nu} G_{\varphi} G_{\nu} (\delta_{a,\psi+1} \delta_{b,\varphi+2} y_{\psi,\nu+3} y_{\varphi+3,\nu+2} \\ &+ \delta_{a,\psi} \delta_{b,\nu+3} y_{\psi+1,\varphi+2} y_{\varphi+3,\nu+2} + \delta_{a,\varphi+3} \delta_{b,\nu+2} y_{\psi+1,\varphi+2} y_{\psi,\nu+3}) \end{split}$$
(3.45)

From the expression for  $\langle n_{ab} \rangle$ , Eq. (3.43), the bond probability  $p_{ab}$  and  $S_{\text{bonds}}$  then follow according to Eqs. (3.40) and (3.41).

We now have means to sample and draw aggregates at gelation (see Section 3.2) and quantitative observables that characterize them, namely the triangle density  $\frac{\langle t \rangle}{\langle n \rangle}$ , the branching ratio  $\frac{\langle b \rangle}{\langle t \rangle}$  and the bond entropy  $S_{\text{bonds}}$  (cf. Fig. 3.5). In the next chapter, we use them to quantitatively define the different aggregate categories.

# 4

# Complex Particles Robustly Form Well-Defined Morphologies

In this chapter, we will exploit that the triangle density, the branching ratio, and the bond entropy are sufficient to classify all possible aggregates, in fact. In the following, we investigate how different aggregate morphologies are related to particle properties, namely their interactions  $\mathbf{E}$  (see Eq. (2.1)). Proteins self-assemble in many different molecular contexts and protein-protein interactions are diverse [54]. Therefore, we will statistically explore our parameter space of interactions to assess which morphologies generically arise for protein-like complex particles. In particular, we will examine whether ordered aggregates like fibers or crystals emerge.

In Section 2.1, we drew an analogy between proteins and irregular puzzle pieces. This illustration suggests a lock-key mechanism that yields a single well-defined interaction per particle face pair. However, considering the aggregation of complex proteins, for instance, motivates a different perspective. When two proteins bind together, many microscopic interactions exist between their amino acid residues. It has been estimated that a protein-protein interface comprises  $57 \pm 22$  amino acid residues [55]. Each of these residues can have several bonded contacts on the order of ~ 10 [56]. The number of unbonded transient contacts is even higher. Thus, at least a few hundred individual interactions  $\beta \varepsilon_i$  determine the overall interaction  $\beta E$  between two proteins in assembly

$$\beta E = \beta \sum_{i} \varepsilon_{i} \,. \tag{4.1}$$

This many individual interactions could have a self-averaging effect. Moreover, amino acid residues often exhibit a high specificity [28]. Thus, we assume that the component interactions  $\beta \varepsilon_i$  are only weakly correlated. Motivated by the central limit theorem, we model the dimensionless interaction energies  $\beta E$  as a Gaussian distributed random variable with mean  $\mu$  and standard deviation  $\sigma$ .  $\mu$  can be interpreted as the average affinity or propensity of the considered complex particles, with a negative value representing sticky particles.  $\sigma$  quantifies particle anisotropy, where  $\sigma = 0$ would correspond to isotropic particles. Protein binding sites are typically made up of a combination of "hot regions", which are composed of multiple individual residues that contribute more dominantly to the interaction [57]. The distribution of these hot regions over the protein surface is not isotropic. Therefore, we assume that protein-like particles have a larger anisotropy  $\sigma \neq 0$ . The exact magnitude of these parameters may vary from one biological context to the other, for example due to different interacting proteins and binding sites. In conclusion, we choose to independently and identically sample the 10 independent components (once symmetry is taken into account) of the interaction matrix **E** from a Gaussian distribution  $\mathcal{N}(\mu, \sigma)$ . In our analysis, we will consider different pairs of affinity  $\mu$  and anisotropy  $\sigma$ .

In Section 4.1, we explore the parameter space of anisotropic interactions by sampling interaction matrices  $\mathbf{E}$  from a Gaussian distribution. Based on the resulting aggregate morphologies, we argue that all aggregates can broadly be classified into the stereotypical aggregate categories introduced in the last chapter (see Fig. 3.4). For given particle affinity and anisotropy, we then obtain the probability of each aggregate morphology in Section 4.2. In doing so, we demonstrate that increasing particle anisotropy promotes aggregate order. Moreover, slim fibrous morphologies are most common for repulsive anisotropic interactions, while bulky space-filling morphologies, including crystals, are most abundant for attractive ones.

## 4.1 Aggregate Morphologies Fall Into Stereotypical Categories

In order to probe the parameter space of complex interactions, we compile statistics for a particle affinity of  $\mu = 0$  and anisotropies  $\sigma \in \{1, 2, 3, 4, 5, 10\}$ . We quantify the state space of aggregate morphologies by the morphological observables from Section 3.3: triangle density, branching ratio and bond entropy. Thereby, we see that aggregates only fall into a few subregions of the state space. By visualizing the associated aggregate morphologies, we show that all aggregates broadly belong to one of the stereotypical categories of aggregate morphology from Fig. 3.4. Based on this, we provide a new quantitative definition for each of these categories by associating them to a region in the three-dimensional state space. This enables the classification of aggregates based on their morphological observables.

For each pair  $(\mu, \sigma)$ , we sample at least 10000 interaction matrices **E** by independently drawing each (dimensionless) interaction energy  $\beta E_{\varphi\nu}$  from the Gaussian distribution  $\mathcal{N}(\mu, \sigma)$ . Given **E**, we obtain the system's partition functions and fugacity at the gel point as detailed in Subsection B.1.2. From these, we calculate the triangle density, the branching ratio as well as the bond entropy of large aggregates using the expressions derived in Section 3.3. Each interaction matrix **E** is then associated to a point in the three-dimensional state space spanned by these three observables.

As an example, we visualize the state space for the pair  $(\mu, \sigma) = (0, 5)$  in the form of a scatter plot in Fig. 4.1. In order to grasp which aggregate morphologies are



Figure 4.1: Our three morphological observables span a state space of aggregate morphologies. Shown are 3d scatter plots for  $\mathbf{E} \sim \mathcal{N}(\mu = 0, \sigma = 5)$  as quantified by triangle density  $2\frac{\langle t \rangle}{\langle n \rangle} \in [0, 1]$  (factor 2 for illustrative purposes), branching ratio  $\frac{\langle b \rangle}{\langle t \rangle} \in [0, 1]$  and bond entropy  $S_{\text{bonds}} \in [0, \ln 10]$ . Each small blue point corresponds to one of 50000 sampled interaction matrices  $\mathbf{E}$ , while the 12 orange dots represent example aggregates that are pictured in Fig. 4.2. All subfigures show the same scatter plot but from different viewpoints: **a**, default, **b**, above, **c**, right, **d**, front.

associated to each region in the state space, we also illustrate a few example aggregates in Fig. 4.2 whose morphology corresponds to the orange dots in Fig. 4.1. Each of the stereotypical aggregate categories from Fig. 3.4 is represented at least once.

In the scatter plot of Fig. 4.1, we observe that the state space is not uniformly explored. This stems in part from the fact that we exclusively sample aggregates at the gel point. Fig. 4.1b illustrates well that there are no data points in the regime where the triangle density and the branching ratio are simultaneously high. A fully occupied lattice would achieve a theoretical maximum of  $\frac{\langle b \rangle}{\langle t \rangle} \times \frac{\langle t \rangle}{\langle n \rangle} = 2/3$ . Precisely at the gel point, however, an infinite aggregate first arises, and its volume fraction is zero, because gelation is a continuous second-order phase transition. In particular, the probability of a specific particle belonging to that infinite aggregate is zero. On a fully occupied lattice, each particle is in the bulk of the aggregate and belongs to two triangles. Also, each triangle constitutes a branching point. Large but finite aggregates necessarily have a large surface on our treelike graph. As a result, neither does



Figure 4.2: Examples from the state space of aggregates. The aggregates in the subpanels **a**–l correspond to the orange dots in Fig. 4.1 and are sampled from the following coordinates, respectively:  $(\frac{\langle b \rangle}{\langle t \rangle}, 2\frac{\langle t \rangle}{\langle n \rangle}, S_{\text{bonds}}) \in \{(0.99, 0.01, 0.06), (0.72, 0.20, 1.30), (0.02, 0.99, 1.11), (0.48, 0.99, 0.87), (0.26, 0.66, 1.24), (0.33, 0.87, 0.62), (0.37, 0.99, 0.64), (0.26, 0.99, 1.01), (0.25, 0.99, 1.11), (0.37, 0.52, 1.02), (0.44, 0.66, 1.48), (0.28, 0.85, 1.43)\}.$  According to our classification scheme these aggregates fall into the following categories (cf. Table 4.1 and Fig. 3.4): **a**, fiber, **b**, fiber network, **c**, fiber of triangles, **d**, decorated fiber, **e**, sponge, **f**–**i**, crystal, **j**, floc, **k** & **l**, liquid.

every particle belong to two triangles nor does every triangle correspond to a branching point. Thus, the number of branching points per particle in large aggregates, quantified by

$$2\frac{\langle b\rangle}{\langle n\rangle} = 2\frac{\langle t\rangle}{\langle n\rangle} \times \frac{\langle b\rangle}{\langle t\rangle}$$

$$(4.2)$$

in the figure, is bounded away from the theoretical maximum of  $2\frac{\langle b \rangle}{\langle n \rangle} = \frac{4}{3}$  for a fully occupied lattice. Fig. 4.1b visually suggests that the maximum at gelation is closer to  $\frac{1}{2}$ . This value is for example realized in decorated fibers (see Fig. 4.2d), which exist close to  $\frac{\langle b \rangle}{\langle t \rangle} = 0.5$  and  $2\frac{\langle t \rangle}{\langle n \rangle} = 1$ . The gel condition thus explains why many data points exhibit a trade-off between triangle density and branching ratio. Either systems exhibit many triangles, which are almost never branching points, or there are fewer triangles, but almost every one of them constitutes a branching point.

Data points seem to overwhelmingly fall into regions of the state space whose aggregates resemble one of the stereotypical aggregate categories we have introduced in Chapter 3 (cf. Figs. 4.2 and 3.4). For instance, Fig. 4.1c) illustrates that there are many data points at low bond entropy and low triangle density. In this regime, slim fibers live, which use a single type of head-to-tail interaction (cf. Fig. 2.3, Fig. 3.4) and see Fig. 4.2), such that  $S_{\rm bonds} \approx \ln 1 = 0$ . The minimum bond entropy of 0 never occurs for systems with high triangle density, i.e. for bulky aggregates. This is because no bulky aggregate morphologies can be built using copies of just a single interaction. Instead, they often exhibit a bond entropy close to  $S_{\rm bonds} = \ln 2 \approx 0.69$ or  $S_{\rm bonds} = \ln 3 \approx 1.10$ , which corresponds to the case of a crystal where two or three interactions are about equally as abundant, respectively. Many data points are concentrated in line-like and plane-like regions of the state space in Fig. 4.1. This suggests that sampled interaction matrices **E** either lead to one of a few stereotypical aggregate morphologies or to aggregates that interpolate between these. However, the scatter plot allows no quantification of the probability to sample from a specific region of the state space. Therefore, it is unclear whether our stereotypical categories of aggregate morphology actually dominate the state space or whether there are more well-defined categories.

We think of liquids and flocs as all-encompassing categories of ill-defined, disordered aggregates. Beyond, to decide whether more generic, well-defined morphologies than the six introduced before exist, We further examine the projection of the state space onto the 2d plane spanned by the branching ratio and the triangle density. These two observables specify the structure of aggregates without referencing the orientational order of individual particles.

Figure 4.3 shows 2d probability histograms for a collection of six pairs  $(\mu, \sigma)$  with  $\mu = 0$  and  $\sigma \in \{1, 2, 3, 4, 5, 10\}$ . For low particle anistropy  $\sigma \leq 1$ , all sampled interaction matrices lead to the formation of floc-like morphologies, which incorporate an intermediate amount of triangles and branchings (see Fig. 4.3a). Formally, this case is equivalent to a high temperature because the interactions only enter the model equations in their dimensionless form  $\beta E = \frac{E}{k_B T}$  (cf. (2.18)). Therefore, entropy dic-

4

tates the self-assembly behavior, which leads to both slim and bulky binding motifs being formed. As anisotropy of interactions is increased (or equivalently temperature lowered), a continuous flow of probability and then accumulation at multiple distinct regions in the state space can be observed (Figs. 4.3b–e). The case of  $\sigma = 10$  in Figure 4.3f suggests that in the large anisotropy limit, almost all aggregates exhibit one of only 6 different morphologies, corresponding to 6 accumulation points at the following coordinates:  $(\frac{\langle b \rangle}{\langle t \rangle}, 2\frac{\langle t \rangle}{\langle n \rangle}) \in \{(0, 1), (0.25, 1), (0.37, 1), (0.5, 1), (0.25, 0.66), (1, 0)\}$ . If two data points are close to each other in the state space, then they amount to morphologically very similar aggregates. Thus, accumulation points indicate the emergence of very well-defined aggregate morphologies, which are generically formed for many sampled interaction matrices **E**. This is remarkable because the parameter space associated with the 10 interactions in  $\mathbf{E}$  is vast. The 6 accumulation points in state space coincide with the following well-defined aggregate categories, respectively: fiber of triangles, crystal, crystal, decorated fiber, sponge, and (slim) fiber. The overall aggregate structure of the two types of crystals is exactly the same (cf. Fig. 4.2, for example). However, if one of the two triangle binding motifs realized in a crystal is much more favorable than the other, then it is also much more abundant in the system. This ultimately leads to a higher branching ratio of the resulting crystals. According to Fig. 4.3, the aggregate morphology, which we termed "floc", which is made up of a wide variety of interactions and exhibits an intermediate number of branchings and triangles, arises when most interactions are weak and, therefore, about equally as favorable. The 5 aforementioned more well-defined categories are stereotypical for the case of strong interactions between particles, where few binding motifs are much more favorable than the others. Before, we introduced two more aggregate categories, namely fiber networks and liquids (see Fig. 3.4). These are stereotypical for intermediately strong interactions. Indeed, in the state space, fiber networks interpolate between flocs and (slim) fibers, while liquids interpolate between flocs and sponges, crystals, fibers of triangles, or decorated fibers.

Based on these considerations, we define state boundaries between the 8 different categories of aggregate morphologies. Table 4.1 summarizes the regions in state space that are associated to each morphology.

In this section, we have illustrated that although our particles exhibit a 10-dimensional parameter space of anisotropic interactions, large aggregates fall into only 8 stereotypical categories of aggregate morphology (see Fig. 4.3), namely: fibers, fiber networks, fibers of triangles, decorated fibers, sponges, crystals, flocs, and liquids. We have seen first indications that particles with strong anisotropic interactions tend to form well-defined aggregates if these interactions are sampled from a Gaussian distribution with vanishing affinity  $\mu = 0$ . With a classification scheme for aggregates in place, we can further corroborate this concept and next quantify how likely each morphology is to occur for different values of affinity  $\mu$  and anisotropy  $\sigma$ .



Figure 4.3: Most aggregates fall into only a few restricted regions of the morphology state space when particle anisotropy is increased. Shown are 2*d* probability histograms of the triangle density  $2\frac{\langle t \rangle}{\langle n \rangle}$  and the branching ratio  $\frac{\langle b \rangle}{\langle t \rangle}$ . Each subfigure quantifies the probability with which distinct aggregate morphologies are formed based on interaction matrices randomly sampled from different underlying Gaussian distributions  $\mathcal{N}(\mu = 0, \sigma)$  with  $\sigma \in \{1, 2, 3, 4, 5, 10\}$ . Histogram bins have dimensions  $0.02 \times 0.02$ . The color bars quantify the probability for a system **E** to land in each bin, respectively. The histograms in the subpanels **a**–**e** are based on 50000 data points, while histogram f results from 4983 data points.

Table 4.1: Aggregates are classified into one of the below categories of aggregate morphology based on the following observables: branching ratio  $\frac{\langle b \rangle}{\langle t \rangle}$ , triangle density  $2\frac{\langle t \rangle}{\langle n \rangle}$  and bond entropy  $S_{\text{bonds}}$ . A dash denotes that there are no conditions on an observable, e.g. all aggregates with  $2\frac{\langle t \rangle}{\langle n \rangle} < 0.2$  are classified as (slim) fibers, regardless of their branching ratio and their bond entropy.

Morphology	Branching ratio	Triangle density	Bond entropy
	$rac{\langle b angle}{\langle t angle}$	$2rac{\langle t angle}{\langle n angle}$	$S_{ m bonds}$
(Slim) fiber	_	$0 \leq \cdots < 0.2$	_
Fiber network	_	$0.2 \leq \cdots < 0.35$	_
Fiber of trian- gles	$0 \leq \dots < 0.2$	$0.86 \leq \dots \leq 1$	$0 \leq \dots < \ln 4$
Decorated fiber	$0.45 \leq \cdots \leq 1$	$0.86 \leq \dots \leq 1$	$0 \leq \dots < \ln 4$
Sponge	_	$0.65 \leq \cdots < 0.86$	$0 \leq \dots < \ln 4$
Crystal	$0 \leq \dots < 0.45$	$0.86 \leq \dots \leq 1$	$0 \leq \dots < \ln 4$
Floc	_	$0.35 \leq \cdots < 0.65$	_
Liquid	_	$0.65 \leq \dots \leq 1$	$\ln 4 \leq \cdots \leq \ln 10$

## 4.2 Particle Anisotropy Promotes Aggregate Order

According to the last section, particles with anisotropic interactions may form welldefined morphologies, especially when these interactions are strong. This section highlights this conclusion by providing the probabilities of each aggregate morphology in our model (cf. Fig. 3.4). We sample interaction matrices from Gaussian distributions for 25 pairs of particle affinity and anisotropy. Thereby, we conclude that the previously observed trend towards well-defined and ordered aggregates persists. In fact, we find that the emergence of ordered aggregates is a direct consequence of anisotropic interactions because their prevalence is largely independent of the average affinity of particles. Increasing the anisotropy of interactions promotes both fibers and crystals. Fibers do generically self-assemble from anisotropic particles that are repulsive on average, while crystals arise if their interactions are mostly attractive.

We augment the discussion of the previous section by adding another dimension to the particle properties considered. Instead of keeping the average particle affinity  $\mu$ 

fixed at zero, we now survey all parameter pairs of  $\mu \in \{-2, -1, 0, 1, 2\}$  and particle anisotropy  $\sigma \in \{1, 2, 3, 4, 5\}$ . For each pair  $(\mu, \sigma)$ , we randomly sample between 10000 and 50000 interaction matrices from the Gaussian distribution  $\mathcal{N}(\mu, \sigma)$ . Then we characterize the morphologies of the resulting large aggregates at the gel point by computing their morphological observables using the tools from Chapter 3. Subsequently, we accordingly classify aggregates, employing the scheme from the previous section (see Table 4.1). Using pie charts, Fig. 4.4 illustrates the resulting probabilities to observe each aggregate morphology as a function of affinity  $\mu$  and anisotropy  $\sigma$ . Figure 4.4 exhibits multiple trends in self-assembly behaviour.

First, increasing the anisotropy of interactions promotes the occurrence of welldefined ordered aggregates: different types of fibers, as well as sponges and crystals, dominate the state space of anisotropic particles. In particular, this trend is monotonic for fibers and crystals, whose prevalence increases regardless of the given particle affinity. The fact that anisotropic particles enhance the order of self-assembled aggregates is a natural consequence of the energetic competition of the associated binding motifs. If we increase particle anisotropy  $\sigma$ , then all interactions get stronger due to  $|\beta E| \sim \sigma$ . Thus, an increase in anisotropy amplifies all interaction energies by the same factor. If distinct interactions are not equally as favorable, this enhances differences in the interaction energies. The few energetically most favorable binding motifs then dominate the assembled morphologies because entropy becomes negligible at low temperatures or large particle anisotropy. As a result, the bond entropy of aggregates successively decreases, and well-defined ordered aggregates are amorphous flocs.

Second, the average affinity  $\mu$  of particles is strongly correlated with aggregate bulkiness, i.e. their number of triangles. In Fig. 4.4, this is most apparent for the particle affinities  $\mu \in \{-2, 2\}$ . Bulky morphologies, namely Fibers of triangles, decorated fibers, liquids, and crystals, are much more likely to occur for particles with attractive interactions, while repulsive particles often yield slim fibers and fiber networks. However, this feature is not decoupled from the anisotropy of interactions  $\sigma$ . For small anisotropy  $\sigma \leq |\mu|$ , aggregates tend to be neither very bulky nor very slim. In this regime, the entropy gain related to realizing a wide variety of different interactions and binding motifs, both slim and bulky ones, outweighs their comparably weak energetic competition. Overall, a strong correlation between the triangle density and particle affinity is natural because the latter raises the energetic benefit of having more neighbors in assembly.

In this chapter, we have analyzed the statistics of large aggregates in frustrated self-assembly in detail, relying on results from Chapter 2 and Chapter 3. Thereby, we showed that only a few categories of well-defined ordered morphologies emerge dominant for anisotropic particles despite the inherent complexity of their interactions. In particular, Figure 4.4 demonstrates that within our model, fibers and crystals represent generic morphologies of aggregates in the self-assembly of complex particles.



Figure 4.4: Particle anisotropy promotes aggregate order. Shown are the probabilities to obtain each aggregate morphology when interaction matrices are randomly sampled from a Gaussian distribution with fixed affinity  $\mu \in \{-2, -1, 0, 1, 2\}$  and anisotropy  $\sigma \in \{1, 2, 3, 4, 5\}$ . Each pie chart contains information on aggregate morphologies of 10000 to 50000 interaction matrices.

With these conclusions in mind, we next investigate the role of geometrical frustration in selecting self-assembled morphologies.

# 5

# **Geometrical Frustration**

Equipped with a broad understanding of how particle properties influence the assembled morphologies, we can approach the following question: How does geometrical frustration impact self-assembly? According to our initial proposal, geometrical frustration is widespread for complex anisotropic particles. This is because interactions between them sensitively depend on their relative arrangement. In assembly, complex particles might therefore be unable to form a favorable interaction with one of their neighbors without being forced into a very unfavorable interaction with another neighbor. If geometrical frustration occurs in our model, we conjecture that it will favor slimmer aggregates like fibers, since it would force dense space-filling aggregates to incorporate energetically unfavorable binding motifs in lattice triangles. In this chapter, we put this idea to the test and analyze whether geometrical frustration on a local scale could be a main physical driver of robust fiber formation in the selfassembly of complex protein-like particles.

Our model on the Husimi tree is specifically designed towards the analysis of the local effects of geometrical frustration on the scale of a single lattice triangle. There are no loops besides the triangles, and interactions are exclusively realized between nearest neighbors on the lattice. Thus, triangles are the only source of frustration. To scrutinize the role of geometrical frustration, we contrast aggregate morphologies in our frustrated system with those that are formed in a comparable unfrustrated system. For that purpose, we eliminate geometrical frustration by removing the triangles from the lattice. Thus, the unfrustrated Bethe lattice of coordination f = 4 is a suitable reference point (see Fig. 5.1).

In Section 5.1, we consider the Husimi tree and the Bethe lattice with all sites occupied, which forces maximal frustration. To estimate the frustration associated with an interaction matrix, we investigate differences in particle arrangement between the two lattices. We formalize these in the contact map C of particles, which is a 4 by 4 matrix whose elements are the frequencies with which particle face a (rows of the matrix) is bound to face b (columns). In Section 5.2, we define our measure of frustration, which is larger the more different the contact maps are. Finally, we correlate geometrical frustration on the full lattices with aggregate morphology at the

gel point based on statistical data from many distinct random interaction matrices (cf. Chapter 4). We show that Geometrical frustration promotes fiber formation, while its absence favors crystals.



Figure 5.1: The Bethe lattice represents an unfrustrated alteration of our lattice, the Husimi tree of triangles. a, The Bethe lattice of coordination f = 4, a tree graph without loops. Particles only enter two-body interactions. Hence, there is no frustration because favorable bonds cannot come at the cost of a simultaneous forced unfavorable interaction. b, The Husimi tree of triangles, which we study in this thesis. The triangles constitute small-scale loops, which may give rise to geometrical frustration. c, Triangles constitute a minimal setting of geometrical frustration. Within triangles, a particle may form a favorable interaction with one of its neighbors at the cost of a very unfavorable one with its other neighbor.

## 5.1 Full Occupancy Case of Husimi Tree and Bethe Lattice

Our aggregate-centered framework from Chapter 2 is not well-suited to a fully occupied lattice because it breaks down beyond the gel point (cf. Section 3.2), where the gel fraction becomes nonzero. Instead, we employ a variant of the closely related lattice-centered framework in Subsection 5.1.1 (cf. Section 2.4), which is tailored towards the case that the whole lattice is covered by a single infinite aggregate. In Subsection 5.1.2, we derive a similar set of self-consistent equations for the partition function of the fully occupied Bethe lattice. For both lattices, we obtain the average contact map of each particle as a function of the interactions  $\mathbf{E}$ .

#### 5.1.1 HUSIMI TREE

In the grand canonical ensemble, full occupancy, i.e. particle density  $\rho = 1$ , is achieved for infinite fugacity  $z \to \infty$ . In that limit, the partial partition functions  $g_{\psi}$  of subaggregates are related via (cf. Eq. (2.23))

$$g_{\psi} = z^2 \sum_{\varphi,\nu} y_{\psi+1,\varphi+2} y_{\psi,\nu+3} y_{\varphi+3,\nu+2} g_{\varphi} g_{\nu} \,. \tag{5.1}$$

The corresponding grand canonical partition function of a single aggregate that spans the whole lattice reads

$$\mathcal{Z}_{\mathrm{H}}^{(\mathrm{full})} = \sum_{\psi} g_{\psi} g_{\psi+2} \,. \tag{5.2}$$

Equation (5.1) has no finite solutions for  $z \to \infty$ , which we need to draw and sample aggregates (cf. Appendix B.2). Thus, we solve for the ratio of partition functions

$$x_{\psi} \coloneqq \frac{g_{\psi}}{g_1} \,, \tag{5.3}$$

so that  $x_1 = 1$  by definition. For  $\psi \in \{2, 3, 4\}$  we find the following coupled set of cubic equations by plugging Eq. (5.1) into the definition of  $x_{\psi}$  and dividing both numerator and denominator by  $g_1^2$ :

$$x_{\psi} = \frac{\sum_{\varphi,\nu} y_{\psi+1,\varphi+2} y_{\psi,\nu+3} y_{\varphi+3,\nu+2} x_{\varphi} x_{\nu}}{\sum_{\varphi,\nu} y_{2,\varphi+2} y_{1,\nu+3} y_{\varphi+3,\nu+2} x_{\varphi} x_{\nu}}$$
(5.4)

Solving these equations numerically allows the sampling of aggregates for the fully occupied Husimi tree (see Subsection B.1.3).

Once we have computed the partition functions, we can calculate the contact map, for which we now derive an expression. Let  $C_{ab}$  be the frequency of contacts between face a of a randomly chosen particle with face b of its neighboring particle. The contact map **C** is formally defined as the following 4 by 4 matrix:

$$\mathbf{C} = (C_{ab})_{a,b \in \{1,2,3,4\}} \tag{5.5}$$

In this matrix, each row and each column sum to one due to normalization of the frequencies  $C_{ab}$ :

$$\sum_{a=1}^{4} C_{ab} = 1 = \sum_{b=1}^{4} C_{ab}$$
(5.6)

The frequency  $C_{ab}$  for an (a, b) contact of the root particle with one of its neighbors is given by the statistical weights of the associated terms in the partition function, i.e. qualitatively

$$C_{ab} \cong \frac{(a,b) \text{ terms in partition function}}{\text{partition function}} \,. \tag{5.7}$$

These weights are formally encoded in the two partial partition functions  $g_{\psi}$  and  $g_{\psi+2}$  of the two subaggregates that emerge from the root particle. We use Kronecker deltas  $\delta_{ij}$  to single them out. Let

$$h := \delta_{a,\psi+1} \delta_{b,\varphi+2} + \delta_{a,\psi} \delta_{b,\nu+3} + \delta_{a,\psi+3} \delta_{b,\eta+2} + \delta_{a,\psi+2} \delta_{b,\lambda+3} , \qquad (5.8)$$

where the four summands represent the four faces of the root particle  $\{\psi, \psi + 1, \psi + 2, \psi + 3\} = \{1, 2, 3, 4\}$  each of which could be face  $a \in \{1, 2, 3, 4\}$ . Using Eqs. (5.1)

and (5.2), we can write

$$C_{ab} = \frac{\sum_{\psi} \sum_{\varphi,\nu,\eta,\lambda} y_{\psi+1,\varphi+2} y_{\psi,\nu+3} y_{\varphi+3,\nu+2} g_{\varphi} g_{\nu} y_{\psi+3,\eta+2} y_{\psi+2,\lambda+3} y_{\eta+3,\lambda+2} g_{\eta} g_{\lambda} h}{\sum_{\psi} g_{\psi} g_{\psi+2}} \\ = \frac{\sum_{\psi} \sum_{\varphi,\nu,\eta,\lambda} y_{\psi+1,\varphi+2} y_{\psi,\nu+3} y_{\varphi+3,\nu+2} x_{\varphi} x_{\nu} y_{\psi+3,\eta+2} y_{\psi+2,\lambda+3} y_{\eta+3,\lambda+2} x_{\eta} x_{\lambda} h}{\sum_{\psi} \sum_{\varphi,\nu,\eta,\lambda} y_{\psi+1,\varphi+2} y_{\psi,\nu+3} y_{\varphi+3,\nu+2} x_{\varphi} x_{\nu} y_{\psi+3,\eta+2} y_{\psi+2,\lambda+3} y_{\eta+3,\lambda+2} x_{\eta} x_{\lambda}}.$$
(5.9)

Using the definition of h, it can be shown that this expression indeed satisfies the normalization constraint, Eq. (5.6).

We can now predict the self-assembly behavior on the fully occupied Husimi tree and obtain the corresponding contact map  $\mathbf{C}^{(\mathrm{H})}$  via Eqs. (5.5) and (5.9).

### 5.1.2 Bethe Lattice

In this subsection, we develop a set of self-consistent equations for the partition function  $\mathcal{Z}_{B}^{(\text{full})}$  of the Bethe lattice at full occupancy. Subsequently, this yields an expression for the contact map  $\mathbf{C}^{(B)}$ .

Four independent sublattices originate from each lattice vertex (see Fig. 5.2a). For  $\psi \in \{1, 2, 3, 4\}$ , let  $g_{\psi}$  be the partial partition function of the sublattice that emerges from face  $\psi$  of the corresponding root particle. As illustrated in Fig. 5.2a, the partition function  $\mathcal{Z}_{\mathrm{B}}^{(\mathrm{full})}$  of the whole lattice is given by a product of the partial partition functions,

$$\mathcal{Z}_{\rm B}^{\rm (full)} = \prod_{\psi=1}^{4} g_{\psi} \,.$$
 (5.10)

At full occupancy, there is a particle on each lattice vertex. Thus, the root particle binds to four neighboring particles. Beyond each neighboring particle, three equivalent sublattices emanate in turn. This setting is visualized in Fig. 5.2b. Taking into account the fact that each neighboring particle has four different possible orientations  $\varphi$ , the partial partition function  $g_{\psi}$  of a sublattice is self-consistently given by

$$g_{\psi} = \sum_{\varphi=1}^{4} y_{\psi,\varphi} \prod_{\nu \neq \varphi} g_{\nu} = \sum_{\varphi} \frac{y_{\psi,\varphi}}{g_{\varphi}} \mathcal{Z}_{\mathrm{B}}^{(\mathrm{full})} \,.$$
(5.11)

To compute the partial partition functions numerically, We again consider the ratio  $x_{\psi} := \frac{g_{\psi}}{g_1}$ , which obeys

$$x_{\psi} = \frac{\sum_{\varphi} y_{\psi,\varphi}/x_{\varphi}}{\sum_{\varphi} y_{1,\varphi}/x_{\varphi}}.$$
(5.12)



Figure 5.2: We exploit the tree structure of the Bethe lattice to obtain self-consistency relations for the partition function at full occupancy. a, Four equivalent sublattices originate from the root particle. Each subaggregate corresponds to a partial partition function  $g_{\psi}$ , where  $\psi \in \{1, 2, 3, 4\}$  is the face of the root particle from which the sublattice emerges. b, A subaggregate comprises a particle that binds to the root and three more subaggregates. Here, we show the subaggregate emerging from face 2 of the root particle. A (2, 3) interaction is realized, which formally corresponds to the Boltzmann weight  $y_{23} = e^{-\beta E_{23}}$ .

Notably, this self-consistency relation has a simpler structure than the one we found for the Husimi tree (cf. Eq. (5.4)) and is only quadratic in the partial partition functions. Similarly to the Husimi tree case, solving this equation for  $x_{\psi}$  permits the sampling and drawing of the full occupancy morphology on the Bethe lattice (see Appendix B.1.3).

We obtain the contact map  $\mathbf{C}^{(\mathrm{B})}$  of the fully occupied Bethe lattice using a similar ansatz as for the Husimi tree (cf. Eq. (5.7)). However, now only the subaggregate originating from face *a* of the root particle can yield an (a, b) contact. Therefore, the contact frequencies  $C_{ab}$  are given by

$$C_{ab} \widehat{=} \frac{(a,b) \text{ term in } g_a}{g_a} \,. \tag{5.13}$$

If we employ (5.11) for  $g_a$  as well as the definition  $x_{\psi} = g_{\psi}/g_1$ , then we find

$$\begin{split} C_{ab} &= \frac{\sum_{\varphi} y_{a,\varphi} \delta_{b,\varphi} \prod_{\nu \neq \varphi} g_{\nu}}{\sum_{\varphi} y_{a,\varphi} \prod_{\nu \neq \varphi} g_{\nu}} \\ &= \frac{y_{a,b} \prod_{\nu \neq b} g_{\nu}}{\sum_{\varphi} y_{a,\varphi} \prod_{\nu \neq \varphi} g_{\nu}} \\ &= \frac{y_{a,b}/x_b}{\sum_{\varphi} y_{a,\varphi}/x_{\varphi}} \,, \end{split}$$
(5.14)

where we divided both numerator and denominator by  $\prod_{\psi} g_{\psi}$  in the last line.

We now have the formal framework in place that allows us to consider the differences in contact maps and, hence, in relative particle arrangements between the Husimi tree and the Bethe lattice. The next and last section of this thesis builds on these formulas to determine the role of geometrical frustration in self-assembly within our model.

## 5.2 Geometrical Frustration Promotes Fiber Formation

In Chapter 4, we have compiled statistics on the probabilities of different aggregate morphologies to arise at the gelation point. Now, our objective is to assess how geometrical frustration influences morphology selection in this regime.

In the introduction, we first put forward the intuition that complex particles may robustly form fibers due to geometrical frustration on a very local level, namely the single-particle scale (cf. Chapter 1). For an aggregate to grow in each spatial direction, particles must simultaneously interact with many neighbors. Specifically in our model, particles would have to realize three-body interactions in a lattice triangle (cf. Section 2.2). This could entail geometrical frustration, meaning a very favorable interaction comes at the cost of an unfavorable one. This tradeoff of three-body interactions could make it energetically more favorable for particles to exclusively form two-particle binding motifs instead. This leads to a slim, linear aggregate extending only in one spatial direction. Thereby, geometrical frustration is mitigated because particles only have a minimal number of neighbors in assembly. We prevent this mitigation mechanism by considering fully occupied lattices. Consequently, a frustrated system may select an entirely different relative particle arrangement than an unfrustrated one to avoid strongly unfavorable conformations. Therefore, geometrical frustration is larger the more different the contact map of the fully occupied Husimi tree is compared to the one of the Bethe lattice.

Formally, we define geometrical frustration f as the distance between the contact maps  $\mathbf{C}^{(\mathrm{H})}$  and  $\mathbf{C}^{(\mathrm{B})}$  with respect to the 1-norm  $\|\cdot\|_1$ :

$$f \coloneqq \frac{1}{8} \left\| \mathbf{C}^{(\mathrm{H})} - \mathbf{C}^{(\mathrm{B})} \right\|_{1}$$
  
=  $\frac{1}{8} \sum_{a,b=1}^{4} |C_{ab}^{(\mathrm{H})} - C_{ab}^{(\mathrm{B})}|$  (5.15)

By using the 1-norm we define frustration as the frequency of contacts differing between the two lattices. If particles assume the exact same relative arrangement on both lattices, this amounts to vanishing frustration, f = 0. An entirely different arrangement, meaning that no contacts are the same, implies 100 % frustration, f = 1.  $f \in [0, 1]$  follows from the normalization of the rows and columns of the contact maps as well as the triangle inequality:

$$\left\| \mathbf{C}^{(\mathrm{H})} - \mathbf{C}^{(\mathrm{B})} \right\|_{1} \le \left\| \mathbf{C}^{(\mathrm{H})} \right\|_{1} + \left\| \mathbf{C}^{(\mathrm{B})} \right\|_{1} = 8$$
(5.16)

Using the expressions for  $C_{ab}^{(\mathrm{H})}$  and  $C_{ab}^{(\mathrm{B})}$  from the last section (see Eqs. (5.9) and (5.14)), we can calculate the geometrical frustration f for an arbitrary interaction matrix **E**.

In Chapter 4, we classified aggregates into only a few morphology categories. To correlate geometrical frustration to aggregate morphology, we revisit our statistical data for anisotropic particles whose interactions are sampled from the unbiased Gaussian distribution  $\mathcal{N}(\mu = 0, \sigma = 5)$ . For each of the corresponding 50000 interaction matrices, we solve the self-consistency equations of the partition function for the fully occupied Husimi tree and Bethe lattice (see Appendix Subsection B.1.3). Subsequently, we compute the geometrical frustration f, respectively. Thus, we obtain a mapping of each interaction matrix to its frustration value and the associated morphology:

$$\mathbf{E} \mapsto (f, \text{aggregate morphology}) \tag{5.17}$$

We combine this information for the 50000 random interaction matrices in the form of a stacked histogram (see Fig. 5.3). Most interaction matrices have a nonvanishing value of f, which indicates that geometrical frustration forces a change in conformation, at least in the full occupancy case. The histogram exhibits very prominent peaks at f = 0 and f = 1/3. Crystals are abundant for unfrustrated systems with  $f \simeq 0$ . On the other hand, interaction matrices that yield a comparably high value of geometrical frustration  $f \simeq 1/3$  disproportionately often lead to fibers. This value implies that 1/3 of the contacts differ between the Husimi tree and the Bethe lattice.

We first intuit why there are peaks at f = 0 and f = 1/3. These frustrations are stereotypical for crystals and fibers, respectively. For both morphologies, we illustrate stereotypical conformations at full occupancy on the Husimi tree and the Bethe lattice (see Fig. 5.4). As illustrated on the left of Fig. 5.4a, a crystal is already dense and space-filling. Its size is only limited by the number of available particles in the system. Indeed, at full occupancy, the same local particle arrangement is conserved for both the Husimi tree and the Bethe lattice, yielding a vanishing frustration f = 0. The fiber example in Fig. 5.4b is only built from very favorable head-to-tail interactions at gelation (here in blue-green; cf. Fig. 2.3 for the color code of interactions). This feature persists at full occupancy on the Bethe lattice. However, on the Husimi tree, it is impossible to build dense space-filling aggregates using only this type of binding motif due to geometric constraints enforced by the lattice triangles. The resulting equilibrium configuration at full occupancy is shown in the center of Fig. 5.4a. 1/3 of the interactions realized are less favorable cyclic ones (in yellow), yielding f = 1/3.



Figure 5.3: Geometrical frustration is widely spread in the self-assembly of complex particles and biases morphology selection. Shown is a probability density histogram with bin width 0.01 that is based on the geometrical frustration f of 50000 random interaction matrices and the associated morphologies (cf. Fig. 3.4 and Fig. 4.4). The interaction energies are all drawn from the same underlying Gaussian distribution  $\mathcal{N}(\mu = 0, \sigma = 5)$  (in units of  $k_B T$ ). Fibers are strongly correlated with high frustration and exhibit a prominent peak in frustration at around f = 1/3, while crystals are most commonly observed for unfrustrated systems with  $f \approx 0$ . Although  $f \in [0, 1]$  in principle, the highest observed value is  $f_{\text{max}} \approx 0.65$ . Despite differing geometrical constraints, assembly on both lattices is still dictated by the same interaction matrix, making an entirely different particle arrangement (f = 1) impossible. The maximum value is attained by a system that yields slim fibers.

There is a simple geometric argument for this particular value: If we undo one edge per lattice triangle on the Husimi tree, then we obtain the Bethe lattice. The difference between the two lattices thus essentially lies in the third interaction in the lattice triangles. This same explanation applies to a broader class of treelike lattices. For instance, if we considered a treelike graph defined by joining squares at their corners, then we could similarly undo one bond per square to obtain the same Bethe lattice of coordination f = 4. This would likely yield a peak of frustration at f = 1/4. For a treelike lattice of regular k-gons joined at their corners, we would expect f = 1/k. This also hints at the fact that frustration may decrease with the size of loops on the lattice, highlighting the potentially predominant role of geometrical frustration on small scales.

Lastly, we examine how the probability distribution of geometrical frustration f


Figure 5.4: Particles form slim fibers to avoid geometrical frustration, but full occupancy prevents this mitigation mechanism and forces a conformation change. In both subfigures, stereotypical examples of aggregate morphology at gelation and at full occupancy on the Husimi tree as well as the Bethe lattice are shown from left to right. Refer to Fig. 2.3 for the color code of the interactions. **a**, Crystals are typically unfrustrated. If the interactions allow the formation of crystals at the gel point, this self-assembly behavior is often conserved at full occupancy. Both lattices yield the same relative particle arrangement, corresponding to  $f \approx 0$ . **b**, Fiber systems are often associated with frustration. Interaction matrices that yield fibers at gelation stereotypically display systematic differences in relative particle arrangement between both lattices at full occupancy, as indicated by the different realized interactions. This example system corresponds to  $f \approx 1/3$ .

changes with aggregate morphology. For this, we plot separate probability histograms, see Fig. 5.5. Slimmer or porous morphologies, including fibers, fiber networks, flocs, and sponges, exhibit a peak at f = 1/3. This peak is most distinct for fibers and fiber networks, which hints at the fact that these morphologies arise to avoid geometrical frustration. Sponges and flocs are less constrained by the geometry of the lattice. Beyond that, there is a peak at f = 0 in the histograms for morphologies that are bulky and ordered, namely fibers of triangles, decorated fibers, and crystals. This shows that these morphologies are usually associated with the absence of frustration. Since frustration only comes in on the scale of lattice triangles, it is natural that

morphologies that contain many triangles tend to be unfrustrated. However, liquids do not feature such a peak at vanishing frustration. Liquids and flocs are ill-defined, disordered morphologies that use many distinct interactions. On the scale of lattice triangles, no single combination of binding motifs is selected and propagated periodically throughout the whole aggregate, as is the case for crystals, fibers of triangles, and decorated fibers. In particular, there is no clear winner in the energetic competition of distinct interactions. Thus, there is no systematic bias toward forming the same interactions on both lattices, entailing no peak at f = 0. Still, not all interactions are equally as favorable which could lead to small differences in conformation depending on lattice geometry. However, unlike what we observe for fibers (see Fig. 5.4), these differences do not systematically occur for only one of the three interactions realized in a triangle. Geometric constraints could have a stronger influence on aggregate morphology the fewer distinct interactions a system prefers.

In this chapter, we have first introduced a measure of geometrical frustration that is based on how different relative particle arrangements are at full occupancy on the Husimi tree compared to an unfrustrated reference system, namely the Bethe lattice. Geometrical frustration governs the morphology selection of complex particles, which we showed by correlating aggregate morphology to geometrical frustration. Fiber formation generically occurs to mitigate and avoid geometrical frustration, allowing particles to have a minimal number of nearest neighbors. Crystals, conversely, are often associated with unfrustrated systems, which allow the dense, space-filling propagation of periodic binding motifs.



Figure 5.5: Slim aggregates, especially fibers and fiber networks, are often formed to avoid geometrical frustration. This figure is based on the same data as Fig. 5.3. Each subfigure visualizes the probability density of aggregates as a function of geometrical frustration f. The bin width is 0.01. Among the 50000 interaction matrices that were sampled in total, **a**, 5635 yield (slim) fibers, **b**, 4425 fiber networks, **c**, 368 fibers of triangles, **d**, 306 decorated fibers, **e**, 10435 sponges, **f**, 8246 crystals, **g**, 15159 flocs, **h**, 5424 liquids.

# 6 Conclusion

This thesis establishes a theoretical framework for the lattice self-assembly of anisotropic particles with complex interactions, including geometrical frustration in a minimal setting, namely lattice triangles. The treelike structure of the Husimi tree enables the derivation of self-consistent recursion relations for the grand canonical partition functions of individual aggregates. Our approach is tailored towards the characterization of aggregate morphology in self-assembly, which is not analytically accessible in previously existing lattice gas models. Our results suggest that aggregates formed by anisotropic particles with complex interactions fall into only a few different categories of morphologies. Within the scope of our model, we identify geometrical frustration as a governing physical principle of morphology selection in self-assembly. Due to our simple minimalist approach, our model makes multiple approximations. To conclude this thesis, we discuss these modeling choices and their implications in Section 6.1. Thereby, we connect our results to existing research on frustrated self-assembly. In Section 6.2, we preview promising research directions that could further elucidate the physical laws of frustrated self-assembly.

# 6.1 DISCUSSION

Our recursion relations rely on the treelike structure of the Husimi tree. While this feature makes our self-assembly system analytically tractable, it comes with the assumption that aggregates do not contain loops on scales larger than lattice triangles. Consequently, geometrical frustration due to long-ranged interactions is not existent in our model, possibly underestimating geometrical frustration. Our approximation could be tested for validity by performing numerical simulations on the Kagome lattice, which exhibits the same two-dimensional local structure as the Husimi tree (see Fig. 6.1a). Most categories of morphologies we identified on the Husimi tree can also exist on the Kagome lattice, including fibers and crystals (cf. Fig. 3.4). We only expect a different phenomenology for the morphologies that we term fibers of triangles and decorated fibers on the Husimi tree. Due to their curved structure, such aggregates would form higher-order loops on the Kagome lattice, yielding morphologies reminiscent of micelles instead. Thus, most self-assembled morphologies on the Kagome lattice could be consistent with our predictions. This would suggest that the



Figure 6.1: Different aspects of the frustrated self-assembly of complex particles can be studied further on other lattices. a, The Kagome lattice is the two-dimensional counterpart of the Husimi tree. On this lattice, six triangles form a higher-order loop, which could be associated with additional geometrical frustration. b, A regular tetrahedron, the 3-simplex. A triangle corresponds to a 2simplex. Multiple d-Simplexes can be joined at their corners to form a treelike lattice that generalizes the Husimi tree to dimensions  $d \ge 3$ . Therewith, the infinity complexity limit of particles,  $d \to \infty$ , may be investigated. c, The Sierpinski triangle is a fractal lattice. As such, it is amenable to exact decimation methods from renormalization group theory, which could enable the study of hierarchical self-assembly.

effects of geometrical frustration on hard particles with nearest-neighbor interactions are indeed strongest on the single-particle scale.

Some other models of geometrical frustration consider elastic particles [41, 43, 44] that deform upon binding depending on their misfit. This may lead to a superextensive buildup of geometrical frustration that stabilizes aggregates at a finite size. This aspect of geometrical frustration is ignored in our model. Proteins exhibit complex interactions due to their anisotropy, and they are partially flexible [45, 54]. Thus, a unifying theory combining anisotropic interactions with particle elasticity could provide a more accurate description of protein self-assembly. We expect that elasticity would not qualitatively change the preferred binding motifs of small aggregates. Still, it could lead to secondary aggregation behavior [58], meaning some large aggregates change their conformation to prevent the super-extensive buildup of elastic frustration. Well-defined, periodically ordered aggregates like fibers or crystals could exhibit shape flattening [41], whereby elastic frustration is uniformly spread over the whole aggregate, making its contribution per particle size-independent.

In this thesis, we mainly focused on the gel point of intermediate particle density to acquire statistics of large aggregate morphologies, which is a consequential modeling choice. We motivated this point of view with the occurrence of large, well-defined aggregates in biological systems, including protein complexes in cells and pathological fibers in medical conditions like Alzheimer's and sickle cell anemia [8, 35, 37]. Consequently, our perspective is best suited for comparable self-assembly systems that exhibit large but finite aggregates. Besides, geometrical frustration may be particularly relevant in these systems: Particles frequently meet and interact but can still avoid it by lowering the number of their neighbors in assembly. We expect our conclusions to be robust with respect to the particle density. In our model, the overall energy of an aggregate is extensive in its size, meaning the local competition of interactions is largely unaffected by the total number of particles in the system. Dilute systems would yield smaller aggregates, but even short fibers can be considered as fibers, for instance. In dense systems, particles would mostly form the same morphologies unless different aggregates are forced so close to each other that additional unfavorable interactions must be realized. We observed this case in Fig. 5.4a, where geometrical frustration leads to a breakup of fibers at full occupancy.

We compiled statistics for many random interaction matrices to obtain a broad picture of the morphology state space of complex anisotropic particles. Our results align with the previous conclusion of Koehler et al. [48] that aggregates of anisotropic particles with complex interactions only fall into a few stereotypical categories, although the associated parameter space of interactions is vast.

We assume that the interaction energies of distinct particle face pairs can be independently and identically sampled from a Gaussian distribution. We motivate this procedure via the central limit theorem because the overall interactions between complex proteins result from many microscopic interactions between their surface amino acid residues that are likely only weakly correlated (see Chapter 4). Thereby, we also reference the concept of "hot regions" composed of multiple amino acid residues that contribute dominantly to the overall interaction [57]. If protein-protein interactions were dictated by only a small ratio of all amino acid residues, this could imply that the distribution of protein-protein interactions has a heavier tail than the Gaussian distribution. If anything, this would make protein interactions even more anisotropic. In line with our results, we could thus imagine that proteins predominantly selfassemble into well-defined ordered aggregates in vivo. A statistical analysis of protein aggregates observed in experiments could enable a quantitative comparison with our results.

Regarding the distribution of interaction energies, no statistical data is available for proteins, unfortunately. Binding energies can only be estimated from the prevalence of contacts in the Protein Data Bank [59]. However, in geometrically frustrated selfassembly, these two quantities are not strictly correlated, as numerous examples of aggregates in this thesis show. In light of this lack of information on protein-protein interaction energies, we cannot exclude the possibility that our results are biased by the Gaussian sampling procedure of interactions. Thus, it would be interesting to check whether our conclusions are robust with respect to our choice of random distribution of interactions.

Finally, it is nontrivial to define geometrical frustration for (dilute) self-assembly systems. For the comparably simple antiferromagnetic Ising model on a triangular lattice [60], geometrical frustration is hard-coded into the Hamiltonian, and every lattice vertex is occupied by a spin. Spins cannot satisfy the most favorable interaction with all their neighbors and are hence frustrated. By contrast, self-assembling particles in fibers may still be able to satisfy favorable interactions with all their neighbors, but only because they do have fewer neighbors. In this work, we based geometrical frustration on differences in the contact maps of the Husimi tree and the Bethe lattice for the case that both of them are fully occupied by particles. This definition provides a common baseline for all aggregate morphologies in our model. Thereby, we argued that anisotropic particles with complex interactions form fibers to avoid geometrical frustration. Our definition of frustration does not explicitly depend on the interaction energies. This represents a pragmatic perspective on geometrical frustration, which is focused on its effects on morphology and not on its source, namely the interactions themselves (although the two are related). Based on significant differences in the probability distributions of frustration for different morphologies, our measure seems to have captured at least some aspect of geometrical frustration. However, our definition is not generally applicable to self-assembly systems, especially to experimental ones that allow no reference to the full occupancy case. A broader definition that comprises many disparate self-assembly systems could allow the identification of common features across diverse settings in the future.

## 6.2 Outlook

A robust physical understanding of which features of the interaction matrix are most relevant in our model could help predict the outcome of self-assembly more generally. As pointed out in the introduction, there is an ever-increasing variety of techniques for synthesizing self-assembling components. Systematic data-driven approaches to designing building blocks for desired morphologies are still in their infancy [18]. Our model provides an easy way to probe the assembly outcomes of arbitrary nearestneighbor interactions between anisotropic particles without relying on computationally expensive methods. We chose affinity and anisotropy as the main descriptors of our interaction matrices. Although we observe unmistakable trends in self-assembly behavior (cf. Fig. 4.4), they are no reliable predictors of the self-assembly outcome. Thus, a deeper understanding of the relationship between our microscopic interactions and macroscopic aggregate morphology would be beneficial. The three qualitatively different types of interactions, which also motivated our choice of color code for aggregates could provide a starting point (cf. Fig. 2.3). There are head-to-tail interactions, which may naturally yield fibers; cyclic interactions, which are associated with crystals; and head-to-head interactions, which favor dimers. Simply based on energetics and geometric constraints, one could compute the most favorable combination of aggregate morphologies as a function of the total number of particles in the system. For instance, if only 3 particles are present, there is an energetic competition between short fibers using two head-to-tail interactions, oligomers formed by three cyclic interactions, and a head-to-head dimer together with an isolated monomer, among others. Another possible descriptor is the propagability introduced by Koehler et al. [48], i.e. the particles' ability to assemble into periodic structures.

In our simplified model, anisotropic particles interact with 4 nearest neighbors at most. However, protein complexes and protein-protein interaction networks may involve many more different binding partners per protein [61]. From a modeling viewpoint, the number of interactions could be increased by placing particles on a lattice of simplexes, which represent the generalization of triangles or tetrahedra to arbitrary dimensions (see Fig. 6.1b). If simplexes of dimension  $d \geq 3$  are joined at their corners, a treelike lattice reminiscent of the Husimi tree is obtained. Particles on a lattice of this type can have many more neighbors, which implies an increase in distinct interactions and a particle complexity reminiscent of proteins. It would be interesting to see, whether geometrical frustration dominates the self-assembly behavior in the infinite complexity limit. Building on our framework for the Husimi tree, recursion relations for the partition functions could be derived. It would be a challenge to introduce a formal language that can describe the plethora of possible relative particle orientations. As an alternative, interactions could be chosen to be isotropic and geometrical frustration made explicit by including a term in the Hamiltonian that grows nonlinearly with the number of neighbors of each particle in the assembly.

Hierarchical self-assembly is a promising approach to scaling up the self-assembly of microscopic particles to the macroscale [22, 62, 63]. This concept involves the formation of structures on successive levels or stages of organization, with each level hierarchically building upon the previous one. The Sierpiński triangle represents an example of a hierarchical lattice that could be utilized to theoretically study this type of self-assembly (see Fig. 6.1c). Lattices like the Sierpiński triangle are also amenable to exact renormalization group methods, which could elucidate prospective universality classes of self-assembled morphologies [64]. Interestingly, just recently, a natural protein that self-assembles into Sierpiński triangles was discovered [65].

The self-assembly of complex particles seems inseparably intertwined with the concept of geometrical frustration. However, to date, no common quantitative physical framework of geometrical frustration for disparate self-assembly systems exists. The prospective future research directions previewed here could guide future attempts to unify our physical understanding of geometrically frustrated self-assembly.

# A

# **Supplementary Information**

# A.1 Choice of Vanishing Interactions for Empty Sites

Throughout this thesis, we have only explicitly considered the case that the surface energy of aggregates vanishes, i.e. that empty sites neither interact with particles nor with each other. Generally, this does not have to be the case. We first show in Subsection A.1.1, that for fixed N our choice of vanishing interactions for empty sites only corresponds to a shift in the total energy of the system, which leaves its physics invariant. Thereby, we follow the presentation of Appendix B1 in [48]. Subsequently, we extend this conclusion to our grand canonical formalism via an appropriate simultaneous shift of the chemical potential  $\mu$  in Subsection A.1.2.

### A.1.1 Shift in the Total Energy Leaves Physics Invariant

Let  $E_{00}$  denote the energy of the interaction between two empty sites and  $E_{\varphi 0}$  that of face  $\varphi$  of a particle with an empty site ( $\varphi \in \{1, 2, 3, 4\}$ ). As in the main text,  $E_{\varphi \nu}$  is the energy of the interaction between faces  $\varphi$  and  $\nu$  of two different particles. We label the total number of each of the corresponding contacts in the system with  $N_{\cdot}$ , e.g.  $N_{00}$  for empty-empty contacts. In the general case of nonvanishing solvent interactions, the total energy of the system is

$$\mathcal{H}_{\rm tot} = N_{00} E_{00} + \sum_{\varphi=1}^{4} N_{\varphi 0} E_{\varphi 0} + \sum_{\varphi \le \nu} N_{\varphi \nu} E_{\varphi \nu} \,. \tag{A.1}$$

Adding a linear combination of conserved quantities to this energy does not change the physics of the system. In our case, the total number of bonds  $N_{\text{bonds}}$  (equivalently the number of lattice edges) and the total number of  $\varphi\text{-faces}~N_\varphi$  are conserved:

$$N_{\rm bonds} = N_{00} + \sum_{\varphi} N_{\varphi 0} + \sum_{\varphi \le \nu} N_{\varphi \nu} \tag{A.2}$$

$$N_{\varphi} = N_{\varphi 0} + \sum_{\nu \neq \varphi} N_{\varphi \nu} + 2N_{\varphi \varphi} = N \tag{A.3}$$

We shift the total energy as follows:

$$\begin{aligned} \overline{\mathcal{H}}_{\text{tot}} &= \mathcal{H}_{\text{tot}} - N_{\text{bonds}} E_{00} - N \sum_{\varphi} (E_{\varphi 0} - E_{00}) \\ &= \sum_{\varphi \leq \nu} N_{\varphi \nu} (E_{\varphi \nu} - E_{00} - (E_{\varphi 0} - E_{00}) - (E_{\nu 0} - E_{00})) \\ &=: \sum_{\varphi \leq \nu} N_{\varphi \nu} \overline{E}_{\varphi \nu} \end{aligned}$$
(A.4)

The interaction matrix

$$\overline{E}_{\varphi\nu} = E_{\varphi\nu} + E_{00} - E_{\varphi0} - E_{\nu0} \tag{A.5}$$

thus describes an equivalent system where empty sites do not interact, as in the main text.

#### A.1.2 EXTENSION TO THE GRAND CANONICAL ENSEMBLE

This restriction to systems with vanishing interactions of empty sites is also warranted, if the total number of particles N is only implicitly fixed through as in the grand canonical ensemble, which we now show.

The self-consistency relation of the aggregate partition functions, Eq. (2.18), can be generalized to the case that empty sites do interact:

$$G_{\psi} = y_{\psi 0} y_{\psi+1,0} y_{00} + c \sum_{\varphi=1}^{4} (y_{\psi+1,\varphi+2} y_{\psi 0} y_{\varphi+3,0} + y_{\psi,\varphi+3} y_{\psi+1,0} y_{\varphi+2,0}) G_{\varphi} + c^2 \sum_{\varphi,\nu} y_{\psi+1,\varphi+2} y_{\psi,\nu+3} y_{\varphi+3,\nu+2} G_{\varphi} G_{\nu} .$$
(A.6)

We can map these relations to the case that the interactions of empty sites vanish, which we considered in the main text. First, the Boltzmann weights are mapped according to Eq. (A.12):

$$\overline{y}_{\psi\varphi} \coloneqq \frac{y_{\psi\varphi}y_0}{y_{\psi0}y_{\varphi0}} \tag{A.7}$$

If we let

$$\overline{G}_{\psi} \coloneqq \frac{G_{\psi}}{y_{\psi 0} y_{\psi+1,0} y_0} \tag{A.8}$$

and

$$\overline{c} := c \frac{\prod_{\varphi} y_{\varphi 0}}{y_0} \tag{A.9}$$

then  $\overline{G}_{\psi}$  and  $\overline{c}$  satisfy the self-consistency relations from Eq. (2.18):

$$\overline{G}_{\psi} = 1 + \overline{c} \sum_{\varphi=1}^{4} (\overline{y}_{\psi+1,\varphi+2} + \overline{y}_{\psi,\varphi+3}) \overline{G}_{\varphi} + \overline{c}^{2} \sum_{\varphi,\nu} \overline{y}_{\psi+1,\varphi+2} \overline{y}_{\psi,\nu+3} \overline{y}_{\varphi+3,\nu+2} \overline{G}_{\varphi} \overline{G}_{\nu} \quad (A.10)$$

To summarize, by simultaneously shifting the chemical potential according to Eq. (A.9)

$$\overline{\mu} = \mu + E_{00} - \sum_{\varphi} E_{\varphi 0} \tag{A.11}$$

and the interaction matrix

$$\overline{E}_{\varphi\nu} = E_{\varphi\nu} + E_{00} - E_{\varphi0} - E_{\nu0} \tag{A.12}$$

the interactions of empty sites can be assumed to vanish without loss of generality.

# A.2 Formal Properties of the Gel Point

This appendix provides a collection of technical results that are complementary to the discussion of the gel point in Section 3.2.

#### A.2.1 Upper Bound for Gel Fugacity

We prove that the fugacity at gelation  $c_{\rm g}$  is finite. At the gel point,

$$\langle n \rangle = \sum_{n=1}^{\infty} n \frac{w_n c^n}{\mathcal{Z}} \tag{A.13}$$

diverges to infinity. Akin to the standard geometric series, this is the case if

$$w_n^{1/n} c \simeq 1 \qquad \Leftrightarrow \qquad c \simeq \frac{1}{w_n^{1/n}}$$
(A.14)

for large  $n \to \infty$ . If we can find a lower bound to  $w_n$ , then  $c_g$  is finite. So how does  $w_n$  scale with n?

 $\boldsymbol{w}_n$  quantifies the probability weight of aggregates of size n and we can bound it from below by

$$w_n \ge \Omega_n \mathrm{e}^{-\beta \mathcal{H}_{\max}(n)} \ge \Omega_n \left(\mathrm{e}^{-4\beta |E_{\max}|}\right)^n$$
 (A.15)

In the last inequality, we used that the maximum energy of an aggregate of size n is bounded from above  $\mathcal{H}_{\max}(n) \leq 4n|E_{\max}|$ , because each particle can at most have 4 interactions, each with "maximum unfavorability"  $|E_{\max}|$ .

We can find  $\Omega_n$  through the partition function in the absence of interactions,  $y_{\psi\varphi} \equiv$ 

y = 1 (see Eqs. (2.11) and (2.18)):

$$\mathcal{Z} = \sum_{n=1}^{\infty} \Omega_n c^n = 4cG^2 \tag{A.16}$$

and

$$G = 1 + 8cG + 16c^2G^2. (A.17)$$

We select the physical solution branch that satisfies  $G(z \rightarrow 0) = 1$ , and find

$$\mathcal{Z} = \frac{(1 - 8c - \sqrt{1 - 16c})^2}{256c^3} \,. \tag{A.18}$$

Taylor expanding around c = 0 yields

$$\Omega_n = 4^{n+1} \frac{(2n+1)!}{(n-1)!(n+3)!} \,. \tag{A.19}$$

Using the Stirling approximation in the large n limit, we find the scaling of  $\Omega_n$ :

$$\Omega_n^{1/n} \simeq 4 \frac{(2n+1)^2}{(n-1)(n+3)} \simeq 16 \quad \text{for } n \to \infty$$
 (A.20)

Summarizing, we have

$$w_n^{1/n} \gtrsim 16 \mathrm{e}^{-4\beta |E_{\max}|} \qquad \text{for } n \to \infty.$$
 (A.21)

To conclude, this implies:

$$c_{\rm g} \le \frac{\mathrm{e}^{4\beta |E_{\rm max}|}}{16} \tag{A.22}$$

# A.2.2 Gelation Is Equivalent to the Divergence of $\frac{dG_{\psi}}{dc}$

We first prove that

$$\langle n \rangle \to \infty \qquad \Leftrightarrow \qquad \frac{\mathrm{d}}{\mathrm{d}c} \mathcal{Z} \to \infty \,.$$
 (A.23)

In the main text, we have already argued that  $\frac{d}{dc}\mathcal{Z} \to \infty$  is necessary for gelation (cf. Eq. (3.16)). In our model (see Eq. (3.3)),

$$\langle n \rangle = \sum_{n=1}^{\infty} \frac{n w_n c^n}{\mathcal{Z}} = \frac{c}{\mathcal{Z}} \frac{\mathrm{d}}{\mathrm{d}c} \mathcal{Z}$$
 (A.24)

and  $c_{\rm g}$  is finite (see Subsection A.2.1). Thus, if  $\mathcal{Z}$  is finite, there is nothing to show. Without loss of generality, we assume that  $\mathcal{Z} = \sum w_n c^n$  diverges as well. Due to the positivity of each probability weight,  $w_n > 0$ ,  $\langle n \rangle$  is then greater than any natural number  $n^*$ :

$$\langle n \rangle > \sum_{n=n^*}^{\infty} \frac{n w_n c^n}{\mathcal{Z}} \ge n^* \frac{\sum_{n=n^*}^{\infty} w_n c^n}{\mathcal{Z}} \to n^*$$
(A.25)

This implies  $\langle n \rangle \to \infty$  and concludes the first step of the proof.

Moreover,

$$\frac{\mathrm{d}}{\mathrm{d}c}\mathcal{Z} = \sum_{\psi} G_{\psi}G_{\psi+2} + 2c\sum_{\psi} \frac{\mathrm{d}G_{\psi}}{\mathrm{d}c}G_{\psi+2} \to \infty \tag{A.26}$$

is equivalent to

$$G_{\psi^*} \to \infty \quad \text{or} \quad \frac{\mathrm{d}G_{\psi^*}}{\mathrm{d}c} \to \infty \quad \text{for some } \psi^* \{1, 2, 3, 4\}.$$
 (A.27)

Analogously to  $\mathcal{Z}$ , divergence of partial partition functions implies divergence of their derivatives with respect to c. In addition, since all the partial partition functions are strictly monotonic in each other, condition (A.27) equivalently holds for any  $\psi \in \{1, 2, 3, 4\}$ . Thus, we finally conclude

$$\langle n \rangle \to \infty \qquad \Leftrightarrow \qquad \frac{\mathrm{d}G_{\psi}}{\mathrm{d}c} \to \infty \quad \text{for all } \psi \{1, 2, 3, 4\}.$$
 (A.28)

# B

# **Numerical Methods**

This appendix outlines the numerical methods that we used throughout this thesis. Section B.1 first explains how we obtain the physical solutions for the partition function in different settings using Mathematica. Based on this, in Section B.2, we show how we randomly sample and draw aggregates in Python.

**Remark.** Note that the interaction energies  $E_{\varphi,\nu}$  always only enter the system's equations in the form of Boltzmann weights  $y_{\varphi,\nu} = e^{-\beta E_{\varphi,\nu}}$ . Thus, we consider the interaction energies as dimensionless parameters  $\beta E_{\varphi,\nu}$  throughout this appendix.

# B.1 MATHEMATICA: SOLVING SYSTEMS OF EQUATIONS

In the main text, we have derived exact self-consistency relations for the partition function, both for our aggregate-centered and the lattice-centered framework (see Eqs. (2.18) and (2.26)). While these relations allow us to obtain analytical expressions for observables in our system, we have to rely on numerical methods to solve these coupled polynomial equations.

What makes this nontrivial is the fact that systems of polynomial equations admit a large number of different solutions, generally. In our model, only one of them is physical (cf. Subsection 3.2.2), and we want to robustly find it numerically for any given parameter set of interest, i.e. the interactions **E** (see Eq. (2.1); sometimes complemented by a fixed fugacity z). If the physical solution is known analytically for a simple case, then this solution can be used as the initial condition of a system of implicit ordinary differential equations (ODEs). These ODEs describe how the physical solution changes upon variation of some suitable variable. In our case, this variable is either the fugacity  $z = e^{\beta\mu}$  or the inverse thermal energy  $\beta = 1/k_BT$ , because the physical solution can analytically be obtained for the formal initial conditions  $z_0 = 0$ and  $\beta_0 = 0$ . The appropriate ODEs can be derived from the self-consistency relations by differentiating with respect to the variable that should be varied formally. Since the ODEs are linear in the derivatives, they can, in principle, be brought into explicit ordinary differential equation form, as was done in the main text multiple times for derivatives of the partition functions (cf. Subsection 3.2.2 and Section 3.3). Therefore, the Picard-Lindelöf Theorem for ordinary differential equations applies to the ODEs we consider. Thus, the physical solution always stays unique (also cf. Subsection 3.2.2).

Using the general idea briefly outlined above, we obtain the physical solution for the partition function in the following three formal settings using Mathematica: the lattice-centered framework with fixed fugacity (Subsection B.1.1), the aggregatecentered framework at the gel point (Subsection B.1.2) and the case of full-occupancy on the Bethe lattice and Husimi tree (Subsection B.1.3).

#### B.1.1 HUSIMI TREE AT FIXED LATTICE FUGACITY

In Section 3.1, we have drawn example aggregates at a fixed fugacity z in the latticecentered framework, for a few example interaction matrices **E**. Here, we present the numerical procedure with which we found the numerical values of the corresponding partial partition functions. These are needed to recursively sample and draw an aggregate (cf. Section B.2).

Recall that the partition function of the lattice is given by (see Eq. (2.25))

$$\mathcal{Z}_{\text{lat}} = 1 + z \sum_{\psi=1}^{4} x_{\psi} x_{\psi+2} \,.$$
 (B.1)

Here, the (rescaled) partial partition functions  $x_{\psi}$  obey (see Eq. (2.26))

$$x_{\psi} = \frac{1 + z \sum_{\varphi=1}^{4} (y_{\psi+1,\varphi+2} + y_{\psi,\varphi+3}) x_{\varphi} + z^2 \sum_{\varphi,\nu} y_{\psi+1,\varphi+2} y_{\psi,\nu+3} y_{\varphi+3,\nu+2} x_{\varphi} x_{\nu}}{1 + 2z \sum_{\varphi=1}^{4} x_{\varphi} + z^2 \sum_{\varphi,\nu} y_{\varphi+3,\nu+2} x_{\varphi} x_{\nu}}$$
(B.2)

The interactions **E** enter the equations in the form of the Boltzmann weights  $y_{\varphi\nu} = e^{-\beta E_{\varphi\nu}}$ . Regardless of the interactions, however, for vanishing fugacity z = 0, this system only has one solution:

$$x_{\psi}(z=0) = 1$$
 (B.3)

This solution is physical and corresponds to the case that there is a vanishing number of particles in the system, meaning they never meet and never interact. If we differentiate Eq. (B.2) with respect to z, then we obtain an ODE of  $x_{\psi}$  and  $\frac{dx_{\psi}}{dz}$  in the variable z. In the main text, we considered the fugacities  $z = e^{\beta\mu}$  for  $\mu \in \{-4, 0, 4\}$ . In the Mathematica code in Fig. B.1, the function singleSol takes a chemical potential  $\mu$  and the 10 independent interaction energies  $(E_{11}, E_{12}, E_{13}, E_{14}, E_{22}, E_{23}, E_{24}, E_{33}, E_{34}, E_{44})$  as input. It then numerically computes the solution to the associated ODE by using the built-in Mathematica function NDSolve with the option Method -> {"EquationSimplification" -> "Residual"}, that is suitable for implicit ODEs. singleSol returns the

unique physical solution for the partial partition functions  $x_{\psi}$  at  $z = e^{\beta \mu}$ .

```
initial
condition = x_1[0] = x_2[0] = x_3[0] = x_4[0] = 1;
singleSol[chempot_, instance_] :=
       Block [{upperTriangle, Y, z, eqs, sol}, upperTriangle = SparseArray[Rule @@@ Thread@ {Flatten[Table[{i, j}, {i, 4}, {j, i, 4}], 1], instance}];
          Y = E^ (-β * (upperTriangle + Transpose[upperTriangle] - DiagonalMatrix[Diagonal[upperTriangle]])) // Simplify;
          eas =
               Evaluate
                Table [
                         \mathbf{X}_{\psi}[\mathbf{Z}] = (\mathbf{1} + \mathbf{Z} \star \mathsf{Sum}[(\mathbf{Y}[[\mathsf{Mod}[\psi + 1, 4, 1], \mathsf{Mod}[\phi + 2, 4, 1]]] + \mathbf{Y}[[\mathsf{Mod}[\psi, 4, 1], \mathsf{Mod}[\phi + 3, 4, 1]])) \\ \mathbf{X}_{\phi}[\mathbf{Z}], \{\phi, 1, 4\}] + \mathbf{Y}[[\mathsf{Mod}[\psi, 1, 4]], \mathsf{Mod}[\phi + 3, 4, 1]] + \mathbf{Y}[[\mathsf{Mod}[\psi, 1, 4]], \mathsf{Mod}[\phi + 3, 4, 1]]) \\ \mathbf{X}_{\phi}[\mathbf{Z}], \{\phi, 1, 4\}] + \mathbf{Y}[[\mathsf{Mod}[\psi, 1, 4]], \mathsf{Mod}[\phi + 3, 4]] + \mathbf{Y}[[\mathsf{Mod}[\psi, 1, 4]], \mathsf{Mod}[\phi + 3, 4]] 
                                       z^{2}*Sum[\forall [Mod[\psi+1,4,1], Mod[\psi+2,4,1]] * \forall [Mod[\psi,4,1], Mod[\vee+3,4,1]] * \forall [Mod[\psi+3,4,1], Mod[\vee+2,4,1]] * x_{\phi}[z] * x_{\phi}[z], \{\phi,1,4\}, \{\vee,1,4\}] / [z] * (z) * 
                                (1 + 2 z + Sum[x_{\phi}[z], \{\phi, 1, 4\}] + z^{2} + Sum[Y[Mod[\phi + 3, 4, 1], Mod[v + 2, 4, 1]] + x_{\phi}[z] x_{v}[z], \{\phi, 1, 4\}, \{v, 1, 4\}]), \{\psi, 1, 4\}] / . \beta \rightarrow 1];
          eqs = Map[Rationalize[#, 10<sup>-50</sup>] &, eqs, -1];
          sol = \{x_1[E^{(chempot)}], x_2[E^{(chempot)}], x_3[E^{(chempot)}], x_4[E^{(chempot)}]\} /.
                 NDSolve[{D[eqs, z], initialcondition}, {x<sub>1</sub>, x<sub>2</sub>, x<sub>3</sub>, x<sub>4</sub>}, {z, e, E^ (chempot)}, Method → {"EquationSimplification" → "Residual"}, InterpolationOrder → All,
                         AccuracyGoal → 12, PrecisionGoal → 12] [[1];
           Flatten[{sol, N[E^ (chempot)]}] (*return physical solution*)
       ];
```

Figure B.1: We use the built-in Mathematica function NDSolve can to numerically solve ODEs for physical solutions to our model equations. Here, we show a screenshot of the Mathematica code that calculates the physical solution to the (rescaled) partial partition functions  $x_{\psi}$  for fixed lattice fugacity z.

### B.1.2 Aggregates on the Husimi Tree at Gelation

In Section 3.2 as well as in Chapter 4, we draw and sample aggregates at the gel point using our aggregate-centered formalism. We follow a similar idea to the one discussed in the previous subsection. However, now we consider many random interaction matrices by independently and identically sampling the dimensionless energies  $\beta E_{\varphi,\nu}$  from a Gaussian distribution  $\mathcal{N}(\mu, \sigma)$  (see Fig. B.2).

Figure B.2: We sample many interaction matrices from a Gaussian distribution  $\mathcal{N}(\mu, \sigma)$  to compile a statistics of aggregate morphology. In this thesis, we considered all pairs of affinity  $\mu \in \{-2, -1, 0, 1, 2\}$  and anisotropy  $\sigma \in \{1, 2, 3, 4, 5\}$ . For  $\mu = 0$ , we sampled 50000 interaction matrices, and for the other affinities, we sampled 10000.

In this case, we rely on numerical solutions to the coupled system of 5 polynomial equations that relates the fugacity  $c_{\rm g}$  and the partial partition functions  $G_{\psi}$ . As a reminder, these equations are the self-consistency relations of our aggregate-centered

formalism (see Eq. (2.18))

$$G_{\psi} = 1 + c \sum_{\varphi=1}^{4} (y_{\psi+1,\varphi+2} + y_{\psi,\varphi+3}) G_{\varphi} + c^2 \sum_{\varphi,\nu} y_{\psi+1,\varphi+2} y_{\psi,\nu+3} y_{\varphi+3,\nu+2} G_{\varphi} G_{\nu}$$
(B.4)

and the necessary gelation condition (see Eq. (3.21))

$$\langle n \rangle \to \infty \quad \Rightarrow \quad \det \mathbf{A} = 0.$$
 (B.5)

We solve these equations for the physical solution at gelation  $(G_1, G_2, G_3, G4, c)|_g$ by considering a related system of ODEs obtained via differentiating the equations above with respect to  $\beta$ . For its initial condition, we consider the system at the gel point for  $\beta = 0$ , for which we can analytically calculate the solution.  $\beta = 1/k_BT = 0$ corresponds to the infinite temperature limit of our model, where interactions do not contribute and all the Boltzmann weights are identical

$$y_{\omega,\nu} = e^{-0 \times E_{\varphi,\nu}} = 1.$$
 (B.6)

As a consequence, we obtain the physical solution

$$\left. \left( G_1, G_2, G_3, G_4, c \right) \right|_{\rm g} \left( \beta = 0 \right) = \left( 4, 4, 4, 4, \frac{1}{16} \right). \tag{B.7}$$

When numerically solving the thus defined system of ODEs, we can ignore units by formally considering the dimensionless inverse thermal energy  $\overline{\beta} = \frac{k_B T_{\text{final}}}{k_B T}$  and energies  $\overline{E}_{\varphi\nu} = \frac{E_{\varphi\nu}}{k_B T_{\text{final}}}$ . Thus, if we vary the temperature from  $T = \infty$  to the temperature of interest  $T = T_{\text{final}}$ , this corresponds to varying  $\overline{\beta} = 0$  to  $\overline{\beta} = 1$ , where  $\beta E_{\varphi\nu} = \beta_{\text{final}} E_{\varphi\nu}$ . Using NDSolve from Mathematica, we solve the associated ODE to obtain the physical solution for the dimensionless energies of interest  $\beta_{\text{final}} E_{\varphi\nu} \sim \mathcal{N}(\mu, \sigma)$  (see Fig. B.3).

#### B.1.3 BETHE LATTICE AND HUSIMI TREE AT FULL OCCUPANCY

In Chapter 5, we compare the particle contact maps of the Husimi tree and the Bethe lattice at full occupancy to quantify frustration. Therewith, our objective is to correlate differences in conformation at full occupancy to the corresponding aggregate morphology we observe at the gel point. Hence, we consider the set of 50000 interaction matrices that we sampled from  $\mathcal{N}(\mu = 0, \sigma = 5)$  in Chapter 4 (also cf. previous subsection). To compute our measure of geometrical frustration for each of these (see Eq. (5.15)), we need the partial partition functions of the Husimi tree and the Bethe lattice at full occupancy. Here, we outline our numerical procedure for calculating them.

We follow the same steps as in the previous subsection. We consider a fixed set of dimensionless interaction energies  $\beta E_{\varphi\nu} = E_{\varphi\nu}/k_B T_{\text{final}}$ . Recall, the (rescaled)

```
initial
condition = G_1[0] = = G_2[0] = = G_3[0] = = G_4[0] = 4 & c[0] = 1/16;
singleSol[energies_] := Module[{upperTriangle, β, y, eqsG, matrixA, eqc, sol},
            upperTriangle = SparseArray[Rule @@@ Thread@{Flatten[Table[{i, j}, {i, 4}, {j, i, 4}], 1], energies}];
            y = E^ (-β* (upperTriangle + Transpose[upperTriangle] - DiagonalMatrix[Diagonal[upperTriangle]])) // Simplify;
             eqsG =
                Rationalize
                     \texttt{Evaluate[Table[G_{\psi}[\beta] = 1 + c[\beta] * Sum[(y[\texttt{Mod}[\psi + 1, 4, 1], \texttt{Mod}[\phi + 2, 4, 1]] + y[\texttt{Mod}[\psi, 4, 1], \texttt{Mod}[\phi + 3, 4, 1]]) G_{\phi}[\beta], \{\phi, 1, 4\}] + [\texttt{Mod}[\psi, 4, 1], \texttt{Mod}[\psi, 4, 1], \texttt{Mod}[\psi, 4, 1], \texttt{Mod}[\psi, 4, 1]] + [\texttt{Mod}[\psi, 4, 1], \texttt{Mod}[\psi, 4, 1]] + [\texttt{Mod}[\psi, 4, 1]] + [\texttt
                                     c\,[\beta\,]\,^{2}\,*\,Sum\,[y\,[Mod\,[\psi\,+\,1,\,4,\,1]\,,\,Mod\,[\psi\,+\,2,\,4,\,1]\,]\,\,*\,y\,[Mod\,[\psi\,+\,3,\,4,\,1]\,]\,\,*\,y\,[Mod\,[\psi\,+\,3,\,4,\,1]\,]\,\,*\,y\,[Mod\,[\psi\,+\,3,\,4,\,1]\,]\,\,*\,y\,[Mod\,[\psi\,+\,2,\,4,\,1]\,]\,\,*\,y\,[Mod\,[\psi\,+\,2,\,4,\,1]\,]\,\,*\,y\,[Mod\,[\psi\,+\,3,\,4,\,1]\,]\,\,*\,y\,[Mod\,[\psi\,+\,3,\,4,\,1]\,]\,\,*\,y\,[Mod\,[\psi\,+\,3,\,4,\,1]\,]\,\,*\,y\,[Mod\,[\psi\,+\,3,\,4,\,1]\,]\,\,*\,y\,[Mod\,[\psi\,+\,3,\,4,\,1]\,]\,\,*\,y\,[Mod\,[\psi\,+\,3,\,4,\,1]\,]\,\,*\,y\,[Mod\,[\psi\,+\,3,\,4,\,1]\,]\,\,*\,y\,[Mod\,[\psi\,+\,3,\,4,\,1]\,]\,\,*\,y\,[Mod\,[\psi\,+\,3,\,4,\,1]\,]\,\,*\,y\,[Mod\,[\psi\,+\,3,\,4,\,1]\,]\,\,*\,y\,[Mod\,[\psi\,+\,3,\,4,\,1]\,]\,\,*\,y\,[Mod\,[\psi\,+\,3,\,4,\,1]\,]\,\,*\,y\,[Mod\,[\psi\,+\,3,\,4,\,1]\,]\,,
                                                 G_{\phi} [\beta] G_{v} [\beta], {\phi, 1, 4}, {v, 1, 4}], {\psi, 1, 4}]], 10<sup>-50</sup>];
              eqc = Rationalize[Det[Evaluate[
                                 Table[KroneckerDelta[\psi, \phi] -
                                          (c[\beta] * (y[Mod[\psi + 1, 4, 1], Mod[\phi + 2, 4, 1]] + y[Mod[\psi, 4, 1], Mod[\phi + 3, 4, 1]]) +
                                                 c\,[\beta\,]\,^{2}\,*\,Sum\,[\,(y\,[Mod\,[\psi\,+\,1,\,4,\,1]\,,\,Mod\,[\phi\,+\,2,\,4,\,1]\,]\,\,*\,y\,[Mod\,[\psi\,+\,3,\,4,\,1]\,]\,\,*\,y\,[Mod\,[\phi\,+\,3,\,4,\,1]\,]\,\,*\,y\,[Mod\,[\psi\,+\,3,\,4,\,1]\,]\,\,*\,y\,[Mod\,[\psi\,+\,3,\,4,\,1]\,]\,\,*\,y\,[Mod\,[\psi\,+\,3,\,4,\,1]\,]\,\,*\,y\,[Mod\,[\psi\,+\,3,\,4,\,1]\,]\,\,*\,y\,[Mod\,[\psi\,+\,3,\,4,\,1]\,]\,,
                                                                     y[[Mod[\psi + 1, 4, 1], Mod[v + 2, 4, 1]] * y[[Mod[\psi, 4, 1], Mod[\phi + 3, 4, 1]] * y[[Mod[v + 3, 4, 1], Mod[\phi + 2, 4, 1]]) * G_v[\beta],
                                                          \{v, 4\}]), \{\psi, 4\}, \{\phi, 4\}]] == 0, 10<sup>-50</sup>];
             \texttt{sol} = \texttt{Quiet}@\texttt{Check}[\{\texttt{G}_1[1], \texttt{G}_2[1], \texttt{G}_3[1], \texttt{G}_4[1], \texttt{c}[1]\} /.
                              NDSolve[{D[eqsG, β], D[eqc, β], initialcondition}, {G1, G2, G3, G4, c}, {β, 0, 1}, Method → {"EquationSimplification" → "Residual"},
                                     InterpolationOrder → All, AccuracyGoal → 12, PrecisionGoal → 12] [1], $Failed] (*return physical solution or $Failed*)];
parallelSols[energiesList_] := ParallelMap[singleSol, energiesList];
```

Figure B.3: We numerically solve ODEs in the variable  $\overline{\beta} = k_B T_{\text{final}}/k_B T \in [0, 1]$  to obtain the physical solutions  $(G_1, G_2, G_3, G_4, c)|_{\text{g}}$  for the Bethe lattice and the Husimi tree at full occupancy, using the built-in Mathematica function NDSolve, as shown in the screenshot. We provide the known physical solution at  $\beta = 0$  as an initial condition, and then find the solution for the temperature of interest  $\overline{\beta} = k_B T_{\text{final}}/k_B T = 1$ . We efficiently solve the ODEs for many different (flattened) interaction matrices energies via a ParallelMap.

partial partition functions  $x_{\psi}$  of the full Husimi tree obey the following system of self-consistent equations (see Eq. (5.4)):

$$x_{\psi} = \frac{\sum_{\varphi,\nu} y_{\psi+1,\varphi+2} y_{\psi,\nu+3} y_{\varphi+3,\nu+2} x_{\varphi} x_{\nu}}{\sum_{\varphi,\nu} y_{2,\varphi+2} y_{1,\nu+3} y_{\varphi+3,\nu+2} x_{\varphi} x_{\nu}},$$
 (B.8)

where  $x_1 = 1$  by definition. The self-consistent equations for the Bethe lattice are (see Eq. (5.12))

$$x_{\psi} = \frac{\sum_{\varphi} y_{\psi,\varphi} / x_{\varphi}}{\sum_{\varphi} y_{1,\varphi} / x_{\varphi}}.$$
 (B.9)

At infinite temperature T, we have  $\overline{beta} = \frac{k_B T_{\text{final}}}{k_B T} = 0$ , meaning  $y_{\varphi\nu} = 1$ . In that case, the self-consistent equations are solved by

$$\binom{(*)}{2}, \binom{(*)}{3}, \binom{(*)}{4}\Big|_{\beta=0} = (1, 1, 1).$$
(B.10)

for  $* \in \{B, H\}$ . As in the previous subsection, we numerically solve the ODE obtained

```
X<sub>1</sub>[beta ] := 1;
initialcondition = x_2[0] = x_3[0] = x_4[0] = 4;
singleFullLattices[energies_] := Module[{upperTriangle, β, y, eqsxB, eqsxH, solB, solH},
                            upperTriangle = SparseArray[Rule @@@ Thread@{Flatten[Table[{i, j}, {i, 4}, {j, i, 4}], 1], energies}];
                         y = E^ (-β* (upperTriangle + Transpose[upperTriangle] - DiagonalMatrix[Diagonal[upperTriangle]])) // Simplify;
                            \mathsf{eqsxB} = \mathsf{Rationalize} \big[ \mathsf{Evaluate} [\mathsf{Table} [\mathsf{x}_{\psi} [\beta] =: \mathsf{Sum} [y [\![\psi, \phi]\!] / \mathsf{x}_{\phi}, \{\phi, \mathbf{1}, 4\}] / \mathsf{Sum} [y [\![\mathbf{1}, \phi]\!] / \mathsf{x}_{\phi}, \{\phi, \mathbf{1}, 4\}], \{\psi, \mathbf{2}, 4\}] \big], 10^{-50} \big];
                            eqsxH = Rationalize[Evaluate[Table[
                                                                               \mathbf{x}_{\psi}\left[\beta\right] = \operatorname{sum}\left[y\left[\operatorname{Mod}\left[\psi+1, 4, 1\right], \operatorname{Mod}\left[\phi+2, 4, 1\right]\right] + y\left[\operatorname{Mod}\left[\psi, 4, 1\right], \operatorname{Mod}\left[v+3, 4, 1\right]\right] + y\left[\operatorname{Mod}\left[\phi+3, 4, 1\right], \operatorname{Mod}\left[v+2, 4, 1\right]\right] + x_{\phi}\left[\beta\right] \times \left[\beta\right] \times \left
                                                                                                                        \{\phi, 1, 4\}, \{v, 1, 4\} ] / Sum[y[[Mod[2, 4, 1], Mod[\phi + 2, 4, 1]]] * y[[Mod[1, 4, 1], Mod[v + 3, 4, 1]]] * J[[Mod[v + 3, 4, 1
                                                                                                                                y [[Mod[\phi + 3, 4, 1], Mod[v + 2, 4, 1]] * x_{\phi} [\beta] x_{v} [\beta], \{\phi, 1, 4\}, \{v, 1, 4\}], \{\psi, 2, 4\}], 10^{-50}];
                            solB = Quiet@Check[{x<sub>2</sub>[1], x<sub>3</sub>[1], x<sub>4</sub>[1]} /.
                                                                               \mathsf{NDSolve}[\{\mathsf{D}[\mathsf{eqsxB},\beta],\mathsf{initial condition}\}, \{\mathsf{x}_2,\mathsf{x}_3,\mathsf{x}_4\}, \{\beta,0,1\},\mathsf{Method} \rightarrow \{\mathsf{"EquationSimplification"} \rightarrow \mathsf{"Residual"}\}, \{\mathsf{x}_3,\mathsf{x}_4\}, \{\beta,0,1\},\mathsf{Method} \rightarrow \{\mathsf{w}_1,\mathsf{w}_2,\mathsf{w}_3,\mathsf{w}_4\}, \{\mathsf{w}_3,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_4,\mathsf{w}_
                                                                                                         InterpolationOrder \rightarrow All, AccuracyGoal \rightarrow 12, PrecisionGoal \rightarrow 12] [1], $Failed];
                            solH = Quiet@Check[{x<sub>2</sub>[1], x<sub>3</sub>[1], x<sub>4</sub>[1]} /.
                                                                               \mathsf{NDSolve}[\{\mathsf{D}[\mathsf{eqsxH},\beta],\mathsf{initial condition}\}, \{\mathsf{x}_2,\mathsf{x}_3,\mathsf{x}_4\}, \{\beta,0,1\}, \mathsf{Method} \rightarrow \{\mathsf{"EquationSimplification"} \rightarrow \mathsf{"Residual"}\}, \{\mathsf{N}_2,\mathsf{N}_3,\mathsf{N}_4\}, \{\mathsf{N}_3,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4,\mathsf{N}_4
                                                                                                         {solB, solH} (*return physical solutions or $Failed*)];
parallelFullLattices[energiesList_] := ParallelMap[singleFullLattices, energiesList];
```

Figure B.4: We numerically solve ODEs in the variable  $\overline{\beta} = k_B T_{\text{final}}/k_B T \in [0, 1]$  to obtain the physical solutions  $(x_1 = 1, x_2, x_3, x_4)|_{\text{full}}$  for the Bethe lattice and the Husimi tree at full occupancy, using the built-in Mathematica function NDSolve. We provide the known physical solution at  $\overline{beta} = 0$  as an initial condition, and then find the solution for the temperature of interest  $\overline{\beta} = 1$ . We efficiently solve the ODEs for many different (flattened) interaction matrices energies via a ParallelMap.

by differentiating the above systems of equations with respect to  $\beta$  using NDSolve (see Fig. B.4). This yields the partial partition functions at the final temperature of interest  $T_{\text{final}}$ , corresponding to  $\overline{\beta} = 1$ .

# **B.2 Python:** SAMPLING AND DRAWING AGGREGATES

We use Python to randomly sample and draw aggregates throughout this thesis. Thereby, we consider three different settings: the Husimi tree at a fixed lattice fugacity (see Section 3.1), the Husimi tree at the gel point (see Subsection 3.2.3 and Chapter 4) and the Husimi tree as well as the Bethe lattice at full occupancy (see Chapter 5). Our sampling procedure only relies on the numerical solutions for the partial partition functions and the fugacity from the previous appendix on Mathematica. In this appendix, we detail each step of our sampling and drawing workflow in Python for the Husimi tree at the gel point. These steps apply analogously to the other settings above.

0. Import required packages (see Fig. B.5).

1. Provide the 10 independent interaction energies (after taking symmetry into account)

$$(E_{11}, E_{12}, E_{13}, E_{14}, E_{22}, E_{23}, E_{24}, E_{33}, E_{34}, E_{44})$$
(B.11)

and the corresponding physical solution for the partial partition functions  $G_{\psi}$ and the fugacity c (see Fig. B.6; cf. Subsection B.1.2). Therewith, we compute the statistical weight of each term on the r.h.s. of the self-consistency relations (see Eq. (2.18)).

2. Call the function aggregate() to randomly sample and draw an aggregate recursively based on these statistical weights (see Fig. B.7).

We present the inner workings of the function aggregate() in more detail in the following. First, it requires an input of the colors of bonds between particles. We choose these colors from the viridis colormap according to the color code in Fig. 2.3 (see Fig. B.8). As shown in Fig. B.9, aggregate() places the root particle of an aggregate with the orientation being randomly chosen according to the corresponding weights in the partition function

$$\mathcal{Z} = c \sum_{\psi=1}^{4} G_{\psi} G_{\psi+2} = 2c(G_1 G_3 + G_2 G_4).$$
(B.12)

Depending on the orientation of the root particle, we then call the functions orient1 and orient3 or orient2 and orient4 to initiate the two subaggregates that originate from it. These functions randomly choose a binding motif (here called branch) according to the statistical weights computed in step 1. Each binding motif can be drawn by using the functions triangle, leftline or rightline with appropriate parameters (see Fig. B.10). Correspondingly, we define a function for each binding motif that draws it, taking into account the correct placement of neighboring particles and the realization of color-coded bonds (see Fig. B.11). Each thus drawn neighboring particle again constitutes a subaggregate, according to its orientation, that is initiated by calling one of the functions orient1, orient2, orient3, or orient4. The Python code continues to draw new binding motifs until the maximum chosen recursion depth of the lattice is reached or all subaggregates have terminated. A subaggregate terminates when the term 1 is randomly chosen, corresponding to no new neighboring particles.

Figure B.5: Required Python packages for our sampling and drawing procedure of aggregates.

```
# provide interaction energies and associated partition functions and fugacity
energies = np.array([0,0,0,0,0,0,0,0,0])
G1, G2, G3, G4, c = 4, 4, 4, 4, 1/16
# define interaction Boltzmann weights (respect symmetry)
y11, y12, y13, y14, y22, y23, y24, y33, y34, y44 = np.exp(-energies) # beta=1
y21, y31, y32, y41, y42, y43 = y12, y13, y23, y14, y24, y34
# calculate statistical weight of each term in the self-consistency relations
s1 = np.
 -→array([1,c*y23*G1,c*y14*G1,c*y24*G2,c*y11*G2,c*y21*G3,c*y12*G3,c*y22*G4,c*y13*G
s2 = np.
 →array([1,c*y33*G1,c*y24*G1,c*y34*G2,c*y21*G2,c*y31*G3,c*y22*G3,c*y32*G4,c*y23*Ġ
s3 = np.
 →array([1,c*y43*G1,c*y34*G1,c*y44*G2,c*y31*G2,c*y41*G3,c*y32*G3,c*y42*G4,c*y33*Ġ
s4 = np.
 →array([1,c*y13*G1,c*y44*G1,c*y14*G2,c*y41*G2,c*y11*G3,c*y42*G3,c*y12*G4,c*y43*Ġ
p1 = s1/s1.sum()
p2 = s2/s2.sum()
p3 = s3/s3.sum()
p4 = s4/s4.sum()
```

Figure B.6: To sample and draw aggregates from a given example system, we provide the interaction energies as well as the partial partition functions and the fugacity.

```
scaling = 0.6 # draw successive levels of aggregates smaller by factor of 0.6_{\Box}
 ⇔for illustrative purposes
x0, y0 = 0, 0 # root particle sits at center of figure
margin = 2.1 # fix figure margins to exclude excessive white space
max_recursion_depth = 5 # only draw the center part of aggregates
minsize = 0 # optional: only draw aggregates of size n > minsize
num_aggregates = 3 \# randomly sample and draw a fixed number of aggregates from_u
 ⇔the system
fig counter = 1
for i in range(num_aggregates):
    # Dynamically create figure variable name
   fig_name = 'fig{}'.format(fig_counter)
   globals()[fig_name], ax = plt.subplots()
    # Sample and draw aggregate
   n = 0
   aggregate()
    # Layout of Axes
   ax.axis('equal')
   ax.axis('off')
   ax.set xlim(x0-margin, x0+margin)
   ax.set_ylim(y0-margin, y0+margin)
    # only show the large aggregates
   if n > minsize:
       plt.show()
       print(fig_name)
       fig_counter += 1 # Increment figure count for the next figure
    else:
        plt.close(globals()[fig_name])
```

Figure B.7: We draw multiple aggregates from the same system using the function aggregate (see Fig. B.9). The variable max\_recursion\_depth enforces that only the immediate vicinity of the root particle is drawn. If aggregates are infinite in size, this also prevents crashing of Python. Using dynamically created figure names, we can visually inspect different drawn aggregates to pick and save one or more examples.

c11 = mpl.colors.rgb2hex(cm.viridis(0.))#'#000000' c12 = mpl.colors.rgb2hex(cm.viridis(1.))#'#E3B5CC' c13 = mpl.colors.rgb2hex(cm.viridis(0.5))#'#E5D461' c14 = mpl.colors.rgb2hex(cm.viridis(1.))#'#D289AD' c21 = c12 c22 = mpl.colors.rgb2hex(cm.viridis(0.))#'#14141414' c23 = mpl.colors.rgb2hex(cm.viridis(1.))#'#C15C8F' c24 = mpl.colors.rgb2hex(cm.viridis(0.5))#'#D3BB22' c31 = c13 c32 = c23 c33 = mpl.colors.rgb2hex(cm.viridis(0.))#'#292929' c34 = mpl.colors.rgb2hex(cm.viridis(1.))#'#A33E70' c41 = c14 c42 = c24 c43 = c34

c44 = mpl.colors.rgb2hex(cm.viridis(0.))#'#3D3D3D'

Figure B.8: We choose the colors of bonds between particles from the viridis colormap of matplotlib based on the type of interaction they realize (cf. Fig. 2.3).

```
def aggregate(x0=0,y0=0,scale=1,depth=0): # draws the root particle and
 ⇔initiates the aggregate
    plt.plot([x0], [y0], marker='o', markersize=10, color="black")
    global n
    n += 1
    root = np.random.choice(np.array([1,2]), p = 1/(G1*G3+G2*G4)*np.
 →array([G1*G3,G2*G4]))
    if root == 1:
        orient1(x0,y0,pi/2,scale,depth)
        orient3(x0,y0,-pi/2,scale,depth)
    if root == 2:
        orient2(x0,y0,pi/2,scale,depth)
        orient4(x0,y0,-pi/2,scale,depth)
# Sample subaggregates depending on (root) particle orientation {1,2,3,4}
def orient1(x0,y0,angle,scale,depth):
   branch = np.random.choice(branches1, p = p1)
    # if branch=1 the subaggregate is terminated!
    if (not type(branch) is int) and (depth < max_recursion_depth):</pre>
        branch(x0,y0,angle,scale,depth+1)
def orient2(x0,y0,angle,scale,depth):
    branch = np.random.choice(branches2, p = p2)
    if (not type(branch) is int) and (depth < max_recursion_depth):</pre>
        branch(x0,y0,angle,scale,depth+1)
def orient3(x0,y0,angle,scale,depth):
    branch = np.random.choice(branches3, p = p3)
    if (not type(branch) is int) and (depth < max_recursion_depth):
        branch(x0,y0,angle,scale,depth+1)
def orient4(x0,y0,angle,scale,depth):
    branch = np.random.choice(branches4, p = p4)
    if (not type(branch) is int) and (depth < max recursion depth):
        branch(x0,y0,angle,scale,depth+1)
```

Figure B.9: The inner workings of the function aggregate. Based on the statistical weights of the different root particle orientations, it calls appropriate orient\* functions that constitute two subaggregates. Within these subaggregate functions, a random binding motif is chosen and the function branch that draws it is called (see Fig. B.10 and Fig. B.11).

B

```
# Define functions that draw: a triangle, a single bond to the left or the
 \rightarrowright,
# and call functions for subaggregates depending on particle orientation
def triangle(color1,color2,color3,lbranch,rbranch,x0,y0,angle,scale,depth):
   plt.plot([x0,x0+scale*cos(angle+pi/6)],[y0,y0+scale*sin(angle+pi/
 →6)],color1) # left bond
   plt.plot([x0,x0+scale*cos(angle-pi/6)],[y0,y0+scale*sin(angle-pi/
 →6)],color2) # right bond
   plt.plot([x0+scale*cos(angle+pi/6),x0+scale*cos(angle-pi/
 -6)],[y0+scale*sin(angle+pi/6),y0+scale*sin(angle-pi/6)],color3) # top bond
   global n
   n += 2
   lbranch(x0+scale*cos(angle+pi/6),y0+scale*sin(angle+pi/6),angle+pi/
 ⇔3,scale*scaling,depth) # left subaggregate
   rbranch(x0+scale*cos(angle-pi/6),y0+scale*sin(angle-pi/6),angle-pi/
 →3,scale*scaling,depth) # right subaggregate
def leftline(lcolor,lbranch,x0,y0,angle,scale,depth):
   plt.plot([x0,x0+scale*cos(angle+pi/6)],[y0,y0+scale*sin(angle+pi/
 (,6)],lcolor) # left bond
   global n
   n += 1
   lbranch(x0+scale*cos(angle+pi/6),y0+scale*sin(angle+pi/6),angle+pi/
 →3,scale*scaling,depth) # left subaggregate
def rightline(rcolor,rbranch,x0,y0,angle,scale,depth):
   plt.plot([x0,x0+scale*cos(angle-pi/6)],[y0,y0+scale*sin(angle-pi/
 →6)],rcolor) # right bond
   global n
   n += 1
   rbranch(x0+scale*cos(angle-pi/6),y0+scale*sin(angle-pi/6),angle-pi/
 ↔3,scale*scaling,depth) # right subaggregate
```

Figure B.10: We define functions that draw a triangle motif, a left bond motif or a right bond motif, coming from an existing particle, respectively. The functions lbranch and rbranch contain information on the new subaggregates that originate from this binding motif (also cf. Fig. B.11). The input parameters fix the colors of the thus formed bonds, the orientations of the neighboring particles (which are needed for the new subaggregates), and the coordinates of the desired drawings on the figure canvas.

```
# Function aliases for use in np.random.choice()
# subaggregates originating from G1
y23G1 = partial(leftline,c23,orient1)
G1y14 = partial(rightline,c14,orient1)
y24G2 = partial(leftline,c24,orient2)
G2y11 = partial(rightline,c11,orient2)
y21G3 = partial(leftline,c21,orient3)
G3y12 = partial(rightline,c12,orient3)
y22G4 = partial(leftline,c22,orient4)
G4y13 = partial(rightline,c13,orient4)
y23y14y43G1G1 = partial(triangle,c23,c14,c43,orient1,orient1)
y23y11y44G1G2 = partial(triangle,c23,c11,c44,orient1,orient2)
y23y12y41G1G3 = partial(triangle,c23,c12,c41,orient1,orient3)
y23y13y42G1G4 = partial(triangle,c23,c13,c42,orient1,orient4)
y24y14y13G2G1 = partial(triangle,c24,c14,c13,orient2,orient1)
y24y11y14G2G2 = partial(triangle,c24,c11,c14,orient2,orient2)
y24y12y11G2G3 = partial(triangle,c24,c12,c11,orient2,orient3)
y24y13y12G2G4 = partial(triangle,c24,c13,c12,orient2,orient4)
y21y14y23G3G1 = partial(triangle,c21,c14,c23,orient3,orient1)
y21y11y24G3G2 = partial(triangle,c21,c11,c24,orient3,orient2)
y21y12y21G3G3 = partial(triangle,c21,c12,c21,orient3,orient3)
y21y13y22G3G4 = partial(triangle,c21,c13,c22,orient3,orient4)
y22y14y33G4G1 = partial(triangle,c22,c14,c33,orient4,orient1)
y22y11y34G4G2 = partial(triangle,c22,c11,c34,orient4,orient2)
y22y12y31G4G3 = partial(triangle,c22,c12,c31,orient4,orient3)
y22y13y32G4G4 = partial(triangle,c22,c13,c32,orient4,orient4)
branches1 = np.
 array([1,y23G1,G1y14,y24G2,G2y11,y21G3,G3y12,y22G4,G4y13,y23y14y43G1G1,y23y11y4
```

Figure B.11: We define partial functions of the functions triangle, leftline or rightline, that preset parameters specific to each binding motif (cf. Fig. B.10). These partial functions can be called when the corresponding binding motif is randomly sampled.

# Bibliography

- G. M. Whitesides and B. Grzybowski. "Self-Assembly at All Scales". Science 295.5564 (Mar. 2002), pp. 2418–2421. ISSN: 1095-9203. DOI: 10.1126/science.1070821.
- P. J. G. Butler and A. Klug. "Assembly of the Particle of Tobacco Mosaic Virus from RNA and Disks of Protein". *Nature New Biology* 229.2 (Jan. 1971), pp. 47–50. ISSN: 2058-1092. DOI: 10.1038/newbio229047a0.
- K. Yonekura, S. Maki, et al. "The Bacterial Flagellar Cap as the Rotary Promoter of Flagellin Self-Assembly". *Science* 290.5499 (Dec. 2000), pp. 2148–2152. ISSN: 1095-9203. DOI: 10.1126/science.290.5499.2148.
- J. A. Shapiro. "Thinking About Bacterial Populations as Multicellular Organisms". Annual Review of Microbiology 52.1 (Oct. 1998), pp. 81–104. ISSN: 1545-3251. DOI: 10.1146/annurev.micro.52.1.81.
- K. G. Libbrecht. "The physics of snow crystals". Reports on Progress in Physics 68.4 (Mar. 2005), pp. 855–895. ISSN: 1361-6633. DOI: 10.1088/0034-4885/68/4/r03.
- [6] F. Vollrath and D. P. Knight. "Liquid crystalline spinning of spider silk". Nature 410.6828 (Mar. 2001), pp. 541–548. ISSN: 1476-4687. DOI: 10.1038/35069000.
- [7] V. Springel, S. D. M. White, et al. "Simulations of the formation, evolution and clustering of galaxies and quasars". *Nature* 435.7042 (June 2005), pp. 629–636. ISSN: 1476-4687. DOI: 10.1038/nature03597.
- [8] J. A. Marsh and S. A. Teichmann. "Structure, Dynamics, Assembly, and Evolution of Protein Complexes". Annual Review of Biochemistry 84.1 (June 2015), pp. 551–575. ISSN: 1545-4509. DOI: 10.1146/annurev-biochem-060614-034142.
- S. A. Woodson. "RNA Folding Pathways and the Self-Assembly of Ribosomes". Accounts of Chemical Research 44.12 (June 2011), pp. 1312–1319. ISSN: 1520-4898. DOI: 10.1021/ar2000474.
- Y. Gao, L. Wang, et al. "Advances in Self-Assembled Peptides as Drug Carriers". *Pharmaceutics* 15.2 (Feb. 2023), p. 482. ISSN: 1999-4923. DOI: 10.3390/pharmaceutics15020482.
- D. Philp and J. F. Stoddart. "Self-Assembly in Natural and Unnatural Systems". Angewandte Chemie International Edition in English 35.11 (June 1996), pp. 1154–1196. ISSN: 0570-0833. DOI: 10.1002/anie.199611541.

- [12] Y. Habibi, L. A. Lucia, and O. J. Rojas. "Cellulose Nanocrystals: Chemistry, Self-Assembly, and Applications". *Chemical Reviews* 110.6 (Mar. 2010), pp. 3479–3500. DOI: 10.1021/cr900339w.
- M. F. Hagan and G. M. Grason. "Equilibrium mechanisms of self-limiting assembly". *Reviews of Modern Physics* 93.2 (June 2021), p. 025008. DOI: 10.1103/revmodphys.93.025008.
- J. J. McManus, P. Charbonneau, E. Zaccarelli, and N. Asherie. "The physics of protein self-assembly". *Current Opinion in Colloid & Interface Science* 22 (Apr. 2016), pp. 73–79. DOI: 10.1016/j.cocis.2016.02.011.
- K. Ariga, J. P. Hill, et al. "Challenges and breakthroughs in recent research on self-assembly". *Science and Technology of Advanced Materials* 9.1 (Jan. 2008), p. 014109. ISSN: 1878-5514. DOI: 10.1088/1468-6996/9/1/014109.
- J. A. Champion, Y. K. Katare, and S. Mitragotri. "Making polymeric microand nanoparticles of complex shapes". *Proceedings of the National Academy of Sciences* 104.29 (July 2007), pp. 11901–11904. DOI: 10.1073/pnas.0705326104.
- Y. Wang, Y. Wang, et al. "Colloids with valence and specific directional bonding". Nature 491.7422 (Oct. 2012), pp. 51–55. DOI: 10.1038/nature11564.
- [18] M. Dijkstra and E. Luijten. "From predictive modelling to machine learning and reverse engineering of colloidal self-assembly". *Nature Materials* 20.6 (May 2021), pp. 762–773. ISSN: 1476-4660. DOI: 10.1038/s41563-021-01014-2.
- [19] N. C. Seeman. "DNA in a material world". Nature 421.6921 (Jan. 2003), pp. 427–431. ISSN: 1476-4687. DOI: 10.1038/nature01406.
- [20] I. Saaem and T. H. LaBean. "Overview of DNA origami for molecular self-assembly". WIREs Nanomedicine and Nanobiotechnology 5.2 (Jan. 2013), pp. 150–162. ISSN: 1939-0041. DOI: 10.1002/wnan.1204.
- [21] F. Zhang and H. Yan. "DNA self-assembly scaled up". Nature 552.7683 (Dec. 2017), pp. 34–35. ISSN: 1476-4687. DOI: 10.1038/d41586-017-07690-y.
- [22] H. Jin and H. D. Espinosa. "Mechanical Metamaterials Fabricated from Self-assembly: A Perspective" (2023). DOI: 10.48550/ARXIV.2311.06734.
- [23] K. Ariga, M. Nishikawa, et al. "Self-assembly as a key player for materials nanoarchitectonics". *Science and Technology of Advanced Materials* 20.1 (Jan. 2019), pp. 51–95. ISSN: 1878-5514. DOI: 10.1080/14686996.2018.1553108.
- [24] D. S. Goodsell. Bionanotechnology. Lessons from nature. Includes bibliographical references (p. 313-322) and index. Hoboken: Wiley-Liss, 2004. 1337 pp. ISBN: 9780471469582.
- [25] N. Stephanopoulos, J. H. Ortony, and S. I. Stupp. "Self-assembly for the synthesis of functional biomaterials". *Acta Materialia* 61.3 (Feb. 2013), pp. 912–930. ISSN: 1359-6454. DOI: 10.1016/j.actamat.2012.10.046.

- [26] E. M. King, C. X. Du, et al. Programmable patchy particles for materials design. 2023. DOI: 10.48550/ARXIV.2312.05360.
- M. Levy, R. Falkovich, S. S. Daube, and R. H. Bar-Ziv. "Autonomous synthesis and assembly of a ribosomal subunit on a chip". *Science Advances* 6.16 (Apr. 2020). ISSN: 2375-2548. DOI: 10.1126/sciadv.aaz6020.
- [28] B. Alberts, ed. Molecular biology of the cell. 4th ed. Includes bibliographical references and index. New York: Garland Science, 2002. 1146386 pp. ISBN: 0815340729.
- [29] K. P. Adamala, M. Dogterom, et al. "Present and future of synthetic cell development". *Nature Reviews Molecular Cell Biology* 25.3 (Dec. 2023), pp. 162–167. ISSN: 1471-0080. DOI: 10.1038/s41580-023-00686-9.
- [30] C. M. Dobson. "Protein folding and misfolding". Nature 426.6968 (Dec. 2003), pp. 884–890. ISSN: 1476-4687. DOI: 10.1038/nature02261.
- [31] M. E. Johnson and G. Hummer. "Nonspecific binding limits the number of proteins in a cell and shapes their interaction networks". *Proceedings of the National Academy of Sciences* 108.2 (Dec. 2010), pp. 603–608. ISSN: 1091-6490. DOI: 10.1073/pnas.1010954108.
- [32] H. Kampinga. "Chaperones in Preventing Protein Denaturation in Living Cells and Protecting Against Cellular Stress". *Handbook of Experimental Pharmacology*. Springer Berlin Heidelberg, 2006, pp. 1–42. ISBN: 9783540297178. DOI: 10.1007/3-540-29717-0\_1.
- [33] T. Sinnige, A. Yu, and R. I. Morimoto. "Challenging Proteostasis: Role of the Chaperone Network to Control Aggregation-Prone Proteins in Human Disease". HSF1 and Molecular Chaperones in Biology and Cancer. Springer International Publishing, 2020, pp. 53–68. ISBN: 9783030402044. DOI: 10.1007/978-3-030-40204-4\_4.
- [34] F. U. Hartl. "Protein Misfolding Diseases". Annual Review of Biochemistry 86.1 (June 2017), pp. 21–26. DOI: 10.1146/annurev-biochem-061516-044518.
- [35] C. M. Dobson. "Protein-misfolding diseases: Getting out of shape". Nature 418.6899 (Aug. 2002), pp. 729–730. DOI: 10.1038/418729a.
- [36] T. P. J. Knowles, M. Vendruscolo, and C. M. Dobson. "The amyloid state and its association with protein misfolding diseases". *Nature Reviews Molecular Cell Biology* 15.6 (May 2014), pp. 384–396. ISSN: 1471-0080. DOI: 10.1038/nrm3810.
- [37] W. A. Eaton and J. Hofrichter. "Sickle Cell Hemoglobin Polymerization". *Advances in Protein Chemistry*. Elsevier, 1990, pp. 63–279. DOI: 10.1016/s0065-3233(08)60287-9.
- [38] H. Garcia-Seisdedos, C. Empereur-Mot, N. Elad, and E. D. Levy. "Proteins evolve on the edge of supramolecular self-assembly". *Nature* 548.7666 (Aug. 2017), pp. 244–247. ISSN: 1476-4687. DOI: 10.1038/nature23320.

- [39] D. S. Goodsell. "Hemoglobin". *RCSB Protein Data Bank* (May 2003). DOI: 10.2210/rcsb\_pdb/mom\_2003\_5.
- [40] E. Sharon and H. Aharoni. "Frustrated shapes". Nature Materials 15.7 (June 2016), pp. 707–709. DOI: 10.1038/nmat4678.
- [41] G. M. Grason. "Perspective: Geometrically frustrated assemblies". *The Journal of Chemical Physics* 145.11 (Sept. 2016), p. 110901. DOI: 10.1063/1.4962629.
- [42] D. M. Hall, I. R. Bruss, J. R. Barone, and G. M. Grason. "Morphology selection via geometric frustration in chiral filament bundles". *Nature Materials* 15.7 (Mar. 2016), pp. 727–732. DOI: 10.1038/nmat4598.
- [43] M. Lenz and T. A. Witten. "Geometrical frustration yields fibre formation in self-assembly". *Nature Physics* 13.11 (July 2017), pp. 1100–1104. DOI: 10.1038/nphys4184.
- [44] H. L. Roy, M. M. Terzi, and M. lenz. Collective deformation modes promote fibrous self-assembly in protein-like particles. 2023. DOI: 10.48550/ARXIV.2308.04698.
- [45] K. Teilum, J. G. Olsen, and B. B. Kragelund. "Functional aspects of protein flexibility". *Cellular and Molecular Life Sciences* 66.14 (Mar. 2009), pp. 2231–2247. ISSN: 1420-9071. DOI: 10.1007/s00018-009-0014-6.
- [46] C. Chothia and J. Janin. "Principles of protein-protein recognition". Nature 256.5520 (Aug. 1975), pp. 705–708. ISSN: 1476-4687. DOI: 10.1038/256705a0.
- [47] A. J. McCoy, V. Chandana Epa, and P. M. Colman. "Electrostatic complementarity at protein/protein interfaces 1 1Edited by B. Honig". *Journal of Molecular Biology* 268.2 (May 1997), pp. 570–584. ISSN: 0022-2836. DOI: 10.1006/jmbi.1997.0987.
- [48] L. Koehler, P. Ronceray, and M. Lenz. *How do particles with complex interactions self-assemble?* 2023. DOI: 10.48550/ARXIV.2312.06350.
- [49] T. D. Lee and C. N. Yang. "Statistical Theory of Equations of State and Phase Transitions. II. Lattice Gas and Ising Model". *Physical Review* 87.3 (Aug. 1952), pp. 410–419. ISSN: 0031-899X. DOI: 10.1103/physrev.87.410.
- [50] N. Goldenfeld. Lectures on Phase Transitions and the Renormalization Group. CRC Press, Mar. 2018. ISBN: 9780429493492. DOI: 10.1201/9780429493492.
- [51] W. Nolting. Grundkurs Theoretische Physik 6: Statistische Physik. Springer Berlin Heidelberg, 2014. ISBN: 9783642253935. DOI: 10.1007/978-3-642-25393-5.
- [52] P. J. Flory. "Molecular Size Distribution in Three Dimensional Polymers. I. Gelation1". Journal of the American Chemical Society 63.11 (Nov. 1941), pp. 3083–3090. ISSN: 1520-5126. DOI: 10.1021/ja01856a061.
- [53] W. H. Stockmayer. "Theory of Molecular Size Distribution and Gel Formation in Branched-Chain Polymers". *The Journal of Chemical Physics* 11.2 (Feb. 1943), pp. 45–55. ISSN: 1089-7690. DOI: 10.1063/1.1723803.

- [54] I. M. Nooren. "New EMBO Member's Review: Diversity of protein-protein interactions". *The EMBO Journal* 22.14 (July 2003), pp. 3486–3492. ISSN: 1460-2075. DOI: 10.1093/emboj/cdg359.
- P. Chakrabarti and J. Janin. "Dissecting protein-protein recognition sites". *Proteins: Structure, Function, and Bioinformatics* 47.3 (Mar. 2002), pp. 334–343. ISSN: 1097-0134. DOI: 10.1002/prot.10085.
- [56] M. Gromiha and S. Selvaraj. "Inter-residue interactions in protein folding and stability". *Progress in Biophysics and Molecular Biology* 86.2 (Oct. 2004), pp. 235–277. ISSN: 0079-6107. DOI: 10.1016/j.pbiomolbio.2003.09.003.
- [57] O. Keskin, A. Gursoy, B. Ma, and R. Nussinov. "Principles of Protein—Protein Interactions: What are the Preferred Ways For Proteins To Interact?" *Chemical Reviews* 108.4 (Mar. 2008), pp. 1225–1244. ISSN: 1520-6890. DOI: 10.1021/cr040409x.
- [58] M. Wang and G. Grason. "Thermal stability and secondary aggregation of self-limiting, geometrically frustrated assemblies: Chain assembly of incommensurate polybricks". *Physical Review E* 109.1 (Jan. 2024), p. 014608.
   ISSN: 2470-0053. DOI: 10.1103/physreve.109.014608.
- [59] M. F. Lensink, G. Brysbaert, et al. "Blind prediction of homo- and hetero-protein complexes: The CASP13-CAPRI experiment". *Proteins: Structure, Function, and Bioinformatics* 87.12 (Oct. 2019), pp. 1200–1221.
   ISSN: 1097-0134. DOI: 10.1002/prot.25838.
- [60] G. H. Wannier. "Antiferromagnetism. The Triangular Ising Net". Physical Review 79.2 (July 1950), pp. 357–364. ISSN: 0031-899X. DOI: 10.1103/physrev.79.357.
- [61] A.-C. Gavin, P. Aloy, et al. "Proteome survey reveals modularity of the yeast cell machinery". *Nature* 440.7084 (Jan. 2006), pp. 631–636. ISSN: 1476-4687. DOI: 10.1038/nature04532.
- [62] Y. He, T. Ye, et al. "Hierarchical self-assembly of DNA into symmetric supramolecular polyhedra". *Nature* 452.7184 (Mar. 2008), pp. 198–201. ISSN: 1476-4687. DOI: 10.1038/nature06597.
- [63] A. Wang, J. Huang, and Y. Yan. "Hierarchical molecular self-assemblies: construction and advantages". Soft Matter 10.19 (2014), p. 3362. ISSN: 1744-6848. DOI: 10.1039/c3sm53214c.
- [64] K. G. Wilson. "The renormalization group: Critical phenomena and the Kondo problem". *Reviews of Modern Physics* 47.4 (Oct. 1975), pp. 773–840. ISSN: 0034-6861. DOI: 10.1103/revmodphys.47.773.
- [65] F. L. Sendker, Y. K. Lo, et al. "Emergence of fractal geometries in the evolution of a metabolic enzyme". *Nature* 628.8009 (Apr. 2024), pp. 894–900.
   ISSN: 1476-4687. DOI: 10.1038/s41586-024-07287-2.
## Acknowledgments

I am grateful to Martin Lenz for introducing me to frustrated self-assembly and for supervising me on this project. Our discussions were always fruitful, despite many of them taking place over Zoom between Paris and Munich. Thank you for guiding me into the world of academia. I also wish to express my gratitude to Erwin Frey, who welcomed me to his amazing group in Munich and agreed to be my TMP mentor and second referee. Thanks for making the last two years very special and fascinating for me, and for allowing me to work on mechano-chemical coupling on membranes as well.

I appreciate Martin, Tom, Simon, Andrey, Natan, and Lorenz, for providing valuable feedback on parts of this thesis, which helped me to greatly improve it. I also want to thank Tom, Simon, and Jonas, as well as the best frustrated and self-assembled group, Lara, Andrey, Vincent, Pawat, and Pietro, for great discussions on (mostly) serious research questions. I look forward to the upcoming three years with you!

A million thanks, Tom, for directly supervising me on my membrane side quest in Erwin's group and for always being available to discuss the inner workings of COMSOL and Mathematica. In this regard, I also want to thank you, Jan, and the other "LS Tom" members and COMSOL appreciators.

I truly worked in the best Master's office, lighting up the days with "coffee?", tea time, Pixie sessions, and hosting the greatest Karaoke night ever! Thank you for that, Antonia, Simon, Jonas, Linus, Nora, Yuhao, Jan, Natan, Carla, Divya, and Leon. A huge shoutout to the MusiCloud; we really rocked that Christmas party, Katrina, Antonia, Tom, Jan, Julian, Sevi, and Tobi! Has our album "last Antholz" received platinum status yet? Also, thank you to all the other amazing members of LS Frey, in particular Laeschkir, Davide, Bea, Ruslan, Henrik, Flo, and Daniel.

Beyond uni, I had the great pleasure of singing in the "UniChor München" which frequently pulled me out of that Tuesday evening research distress. Also, bouldering with you was awesome, Simon, Jonas, and Nora! I am sure you will join us next time, right, Antonia? I would also like to thank my amazing roommates for being there and spending great times together.

Last but not least, I can't thank my girlfriend, my parents, and my siblings enough for supporting me on my path. It means the world to me.



I herewith formally declare that I have composed the present thesis myself and without use of sources or tools other than those cited.

Munich, June 29, 2024

Lukas Kalvoda