

AMYLOID NUCLEATION IN PRESENCE OF CROWDERS

Jonas Wessén



LUNDS
UNIVERSITET

2014

Master's Thesis

Computational Biology and Biological Physics
Department of Astronomy and Theoretical Physics
Lund University

Thesis advisor: *Anders Irbäck*

Preface

During the past year, I've been working on the mechanisms behind the formation of amyloid fibrils, a subject in which I, in my education in Theoretical Physics, had no experience before this work was initiated. However, the beauty of statistical mechanics, which has been the main tool in this thesis, is that it can, with the same fundamental principles, be applied to virtually any system that contains many microscopic degrees of freedom but where only macroscopic properties are of interest, and thereby aid in explaining e.g. why ferromagnets lose their magnetization at high temperatures, how to optimize a heat engine, why there are different phases of water, the colour of hot objects, how earthquakes work, abrupt changes on the stock market, the particle contents of the early universe, the elasticity of rubber bands, thermodynamics of black holes, or (as attempted in this thesis) how proteins aggregate. In addition to studying the crowding effects on amyloid nucleation with Monte Carlo simulations, much time was on my initiative devoted to analytically study the simulated model, using methods from statistical mechanics. Besides for the excellent guidance regarding the simulations, I would like to thank my supervisor Anders Irbäck for his engagement in these analytical explorations, which (after a few dead ends) actually led to something worth mentioning (Section 4 in this thesis). Thanks to his supervision, and fruitful discussions with Sigurður Ægir Jónsson and others at the Computational Biology and Biological Physics Group at the Department of Astronomy and Theoretical Physics, Lund University, the work leading up to this thesis has been a truly inspiring experience.

Part of the simulations were performed on resources provided by the Swedish National Infrastructure for Computing (SNIC) at the LUNARC facility.

Jonas Wessén, May 2014

Contents

| | | |
|----------|--|-----------|
| 1 | Introduction | 1 |
| 2 | Model description | 2 |
| 2.1 | Background | 2 |
| 2.2 | The stick model | 2 |
| 2.3 | Crowding particles | 5 |
| 3 | Methods | 5 |
| 3.1 | Thermodynamic simulations | 5 |
| 3.1.1 | Swendsen-Wang cluster updates | 7 |
| 3.1.2 | Wang-Landau sampling | 8 |
| 3.1.3 | Equilibrium measurements | 9 |
| 3.2 | Kinetic Simulations | 10 |
| 4 | Analytical considerations | 11 |
| 4.1 | 1D growth | 12 |
| 4.2 | 2D growth | 14 |
| 5 | Results | 16 |
| 5.1 | Analytical thermodynamics | 16 |
| 5.2 | Simulated thermodynamics | 20 |
| 5.2.1 | Without crowders | 20 |
| 5.2.2 | With crowders | 24 |
| 5.2.3 | Depletion forces between crowding particles | 27 |
| 5.3 | Simulated kinetics | 30 |
| 6 | Summary | 36 |
| 7 | Outlook | 37 |
| 8 | Appendix A: Detailed balance in Swendsen-Wang cluster updates | 38 |

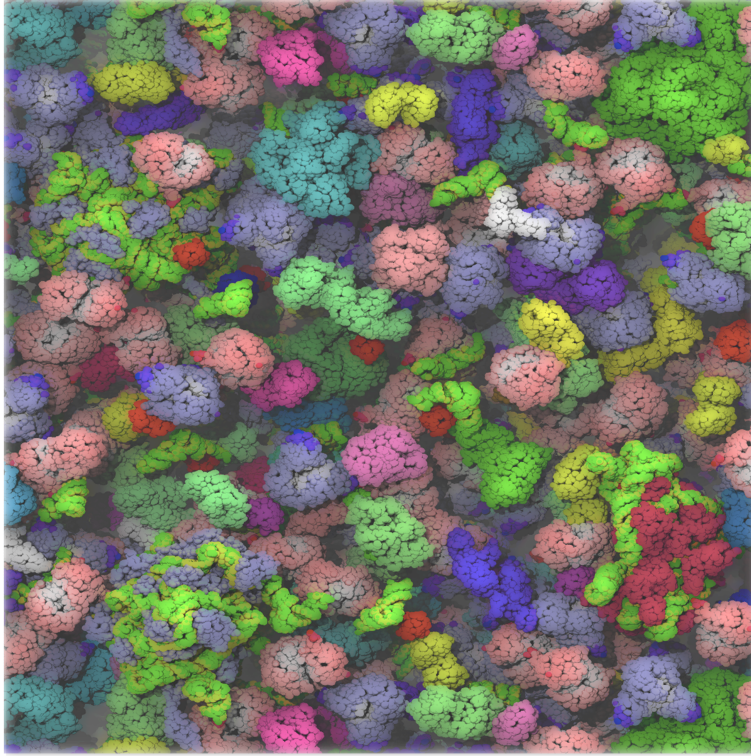


Figure 1: Illustration of the high degree of crowding in the cellular environment. The figure was taken from [7], in which the cytoplasm of an *E. Coli* bacterium is modelled on an atomically detailed level and simulated using Brownian Dynamics.

1 Introduction

During the last few years, crowding effects on the physics of proteins has become an increasingly popular topic of research [1]. This is because most biological processes involving proteins naturally take place in a crowded environment, e.g. in the cellular environment where macromolecules may occupy 30% of the volume (see Figure 1 for an illustration of the high degree of crowding in a cytoplasmic environment). One such biological process would be the formation of amyloid aggregates, which are cross- β -sheet rich protein structures that have been associated with e.g. Alzheimer's disease [2, 3]. In this work, we investigate the crowding effects on the formation of amyloid fibrils by adding neutral crowding particles (i.e. no explicit interaction except excluded volume effects) to a previously studied lattice model [4]. We hypothesise that the dominant effect of such crowding can be derived from the effective increase in the peptide density, which depends on the total volume occupied by the crowding particles ('crowders'), and not on the total surface area of the crowders, as have recently been discussed in the literature [5]. Any dominant surface effects, such as the observed dual effect on the aggregation kinetics of amyloid β fibrils [6], is likely due to an explicit interaction between the crowding particles and the peptides.

In addition, we develop an analytical approach that permits us to study the thermodynamics of the model without crowders. In this approach, we treat the collection of each type (i.e. of given length and width) of aggregates in the system as a collection of non-interacting objects in grand-canonical ensembles, with the over-all constraint of peptide number conservation. This method is used to study systems much larger than those we can simulate using Monte Carlo methods, and to compute an approximate phase diagram for the model.

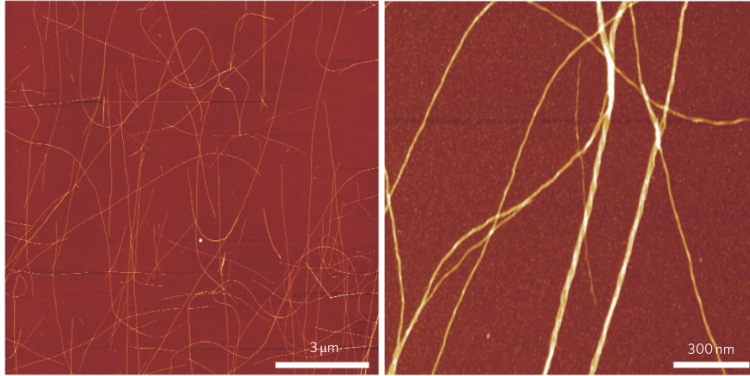


Figure 2: AFM picture of amyloid fibrils. The figure was taken from [8].

2 Model description

2.1 Background

When proteins, for some reason or another, become unfolded from their functional structure, they often acquire a strong tendency to clump together in an uncontrolled manner, forming protein aggregates. An example of this, familiar to everyone, is the boiling of an egg, where the high temperature causes the egg proteins to unfold and clump together. While these aggregates can be disordered (as for the boiled egg), they often have a strongly ordered shape even if the underlying protein (the ‘peptide’) is intrinsically disordered.

A very common ordered structure for protein aggregates is the ‘amyloid’ structure, which consists of many stacked β -sheets (a sheet-like secondary structure present in many proteins). Amyloid aggregates tend to have the shape of long fibers, known as ‘amyloid fibrils’, where the β -sheets are ordered with their strands perpendicular to the fibril axis [10] (this is known as the ‘cross- β structure’). In Figure 2, a collection of amyloid fibrils are shown in an AFM (Atomic Force Microscopy) image. The peptides constituting the β -sheets are bound together by hydrogen bonds, leaving the possibly hydrophobic side-chains sticking out, exposed to the environment. For this reason, layers of cross- β structures may be attached to each other, with their hydrophobic side-chains hiding between the layers from the aqueous surroundings. In Figure 3, the molecular structure of a four layer aggregate is shown.

A strong motivation to study the aggregation mechanism of amyloid fibrils is of course the relevance to many diseases, including Alzheimer’s, Parkinson’s, Type II Diabetes or Huntington’s disease. There are also cases in Nature in which amyloid structure has a functional purpose. Examples of functional amyloids include storage of protein hormones [11], or components of bacterial biofilm, cell adhesion in diploid fungi or suspending fibers for lacewing eggs [2]. Besides for its biological role, the amyloid fibril is interesting for engineering purposes due to its extraordinary mechanical properties combined with its ability to self-assemble [12]. The amyloid fibril has even been suggested as functional food ingredient, such as texture builder or foaming agent [13].

2.2 The stick model

In this work, we model the fibrillation of peptides which are assumed to be short (6-7 residues), such that internal dynamics happens on short time-scales and therefore has no significant effect on the aggregation. The peptides are modelled as rigid sticks and live on the sites of a cubic lattice with side L and periodic boundaries. The orientation of peptide i is described by two perpendicular unit vectors (Figure 4): the backbone vector $\hat{\mathbf{b}}_i$ pointing along the N-to-C direction of the protein, and the hydrogen bond vector $\hat{\mathbf{p}}_i$ representing the two directions ($\pm\hat{\mathbf{p}}_i$) in which hydrogen bonds with other peptides can form. $\hat{\mathbf{b}}_i$ and $\hat{\mathbf{p}}_i$ are confined to the 6 lattice directions, giving $6 \cdot 4$ possible orientations for the peptide. It is useful to define the side-chain vector $\hat{\mathbf{s}}_i = \hat{\mathbf{b}}_i \times \hat{\mathbf{p}}_i$ which points in the direction of the hydrophobic side of the peptide.

The peptides only interact via nearest neighbour interactions. Two neighbouring peptides, i

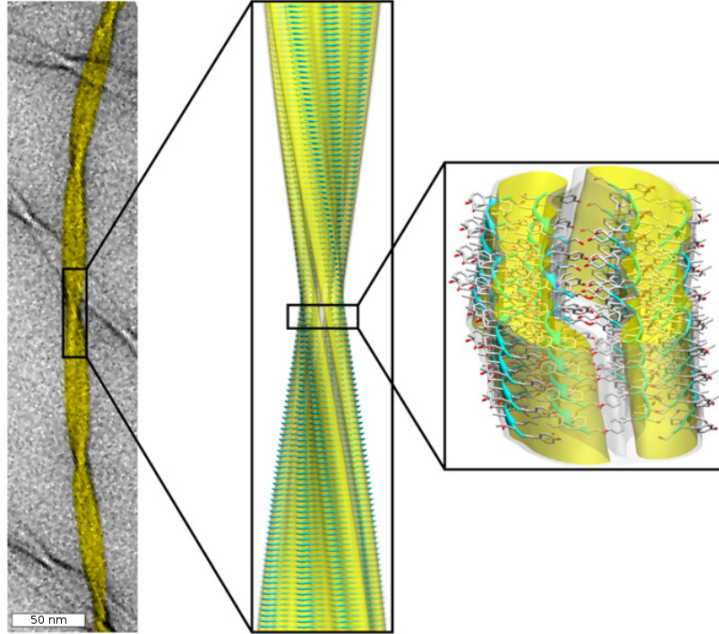


Figure 3: Example of the molecular structure of an amyloid fibril with four layers of cross- β structures. The yellow areas in the right most picture shows the hydrophobic regions. The figure was taken from [9].

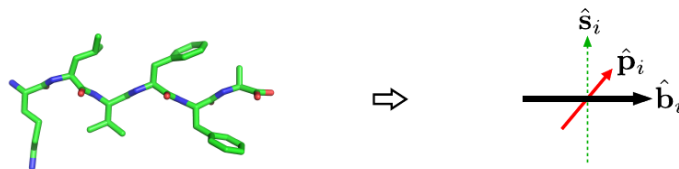


Figure 4: Abstraction of a peptide molecule. The figure was taken from [4].

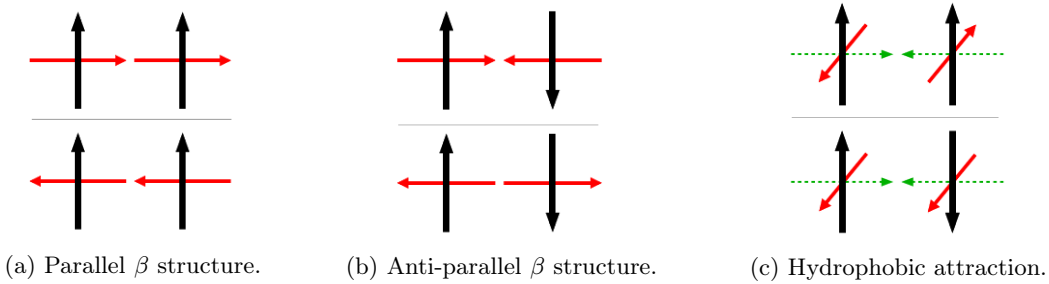


Figure 5: Illustrations of the interactions in the model. The figure was taken from [4].

and j , interact only if $\hat{\mathbf{b}}_i \cdot \hat{\mathbf{b}}_j = \pm 1$ (parallel/anti-parallel backbone vectors) and $\hat{\mathbf{b}}_i \cdot \mathbf{r}_{ij} = 0$ where $\mathbf{r}_{ij} = \mathbf{r}_j - \mathbf{r}_i$ is the vector connecting the positions of the two peptides. If these two criteria are met, an attractive interaction (from e.g. a van der Waals force) is assumed. The interaction is attributed an energy $\epsilon_{ij} = -1$. In the three cases shown in Figure 5, there are additional contributions to the interaction energy due to hydrogen bonds and hydrophobic attraction. In these cases, the interaction energy is

$$\epsilon_{ij} = \begin{cases} -(1 + a_p), & \text{parallel } \beta \\ -(1 + a_a), & \text{anti-parallel } \beta \\ -(1 + b), & \text{hydrophobic attraction} \end{cases}. \quad (2.2.1)$$

The interactions of the model are designed such that the peptides, at sufficiently low temperatures, form two-dimensional aggregates (the interactions do not support growth in the third dimension). In this work, we choose the same parameters as in [4], $a_p = 5$, $a_a = 3$ and $b = 1$. With these parameters, parallel β structures are favoured over anti-parallel β , and double layer aggregates form due to the hydrophobic interactions. The van der Waals interaction makes it possible to form aggregates with even more layers. In the width direction, the layers will be bound together, alternating between the hydrophobic attraction and the small residual attraction. The choice of asymmetric interactions around the peptide backbone vector is physically motivated, with the pairwise hydrophobic attraction mimicking the observed ‘steric zipper’ structure of the fibril spine [14]. The recent experimental fibril structure depicted in Figure 3 shows indeed a pairwise organization of the β -sheets.

In this work, the geometry of the aggregates are of interest. We define an aggregate as a set of peptides which form a connected network with their interactions. These aggregates will in general not have a simple rectangular form, and there is thus an ambiguity in defining their length l and width w . In this work, we define l and w from the eigenvalues λ of the inertia tensor \mathcal{I} corresponding to the aggregate.

Consider an aggregate consisting of n peptides with positions $\mathbf{r}^{(k)} = r_1^{(k)}\mathbf{e}_1 + r_2^{(k)}\mathbf{e}_2 + r_3^{(k)}\mathbf{e}_3$, $k = 1, \dots, n$, relative to the center of mass of the aggregate. The components of the inertia tensor are

$$I_{ij} = \int dV \rho(\mathbf{r}) (\delta_{ij} |\mathbf{r}|^2 - r_i r_j) \quad (2.2.2)$$

$$= m \sum_{k=1}^n [\delta_{ij} \mathbf{r}^{(k)} \cdot \mathbf{r}^{(k)} - r_i^{(k)} r_j^{(k)}] \quad (2.2.3)$$

where m is the mass of a peptide and δ_{ij} is the Kronecker delta. In the rest of this work, we assign each peptide unit mass. Assuming a coordinate system along the lattice directions in which the aggregate lies in the $\mathbf{e}_1\mathbf{e}_2$ -plane, the inertia tensor in this basis takes the form

$$\mathcal{I} \rightarrow \mathbf{l} = \begin{pmatrix} I_{11} & I_{12} & 0 \\ I_{12} & I_{22} & 0 \\ 0 & 0 & I_{33} \end{pmatrix}. \quad (2.2.4)$$

We define $\tilde{\Gamma}$ to be the 2×2 matrix in the upper left corner of Γ . The eigenvalues of Γ are

$$\lambda_{1,2} = \frac{\text{tr}(\tilde{\Gamma})}{2} \pm \sqrt{\left(\frac{\text{tr}(\tilde{\Gamma})}{2}\right)^2 - \det(\tilde{\Gamma})}, \quad (2.2.5)$$

$$\lambda_3 = I_{33}. \quad (2.2.6)$$

where we choose $\lambda_1 \geq \lambda_2$.

For a rectangular aggregate with length l along \mathbf{e}_1 and width w along \mathbf{e}_2 , one obtains

$$\begin{cases} I_{11} &= \frac{n}{12}(w^2 - 1), \\ I_{22} &= \frac{n}{12}(l^2 - 1), \\ I_{12} &= 0, \end{cases} \quad (2.2.7)$$

using Eq. (2.2.3), where $n = l \cdot w$ is the number of peptides in the rectangular aggregate. Consequently,

$$\lambda_{1,2} = \frac{I_{11} + I_{22} \mp |I_{11} - I_{22}|}{2}. \quad (2.2.8)$$

With $l \geq w$,

$$\begin{cases} \lambda_1 &= \frac{n}{12}(l^2 - 1), \\ \lambda_2 &= \frac{n}{12}(w^2 - 1), \end{cases} \quad (2.2.9)$$

or

$$\begin{cases} l &= \sqrt{\frac{12\lambda_1}{n} + 1}, \\ w &= \sqrt{\frac{12\lambda_2}{n} + 1}. \end{cases} \quad (2.2.10)$$

For aggregates with a more arbitrary shape, we *define* their length and width using Eqs. (2.2.10). Figure 6 shows all possible lengths and widths for aggregates containing 10 peptides or less. There are apparent discretization effects for aggregates with $w = 1$ (especially for small lengths), while otherwise there are plenty of states. The overall shape of Figure 6 is expected to persist when considering even larger aggregates.

Experimentally, the amyloid fibrils are seen to have twisted shapes (such as the fibril seen in Figure 3), with a varied handedness and periodicity [15]. A limitation of our model is that this feature cannot be studied, since the fibrils in the model by construction never acquire a twist.

2.3 Crowding particles

In this work, we model the crowding particles (‘crowders’) as cubes with side R_C . We include no explicit interaction (such as a van der Waals force or electrostatic attraction/repulsion) either between crowders themselves or between the crowders and the peptides. The choice of cubical crowders allows for fibrillar growth along the surface of the crowder (like the crowder surface, the growth of an aggregate is limited to the lattice directions). Since, in reality, the orientation of a peptide would be limited when adjacent to a surface, we only allow for states where the peptides that are nearest neighbours to crowders all have their backbone vectors $\hat{\mathbf{b}}_i$ parallel to the surfaces of the crowders. That is, if \mathbf{r}_{iC} is the vector between the peptide position and an adjacent lattice site occupied by a crowder ($|\mathbf{r}_{iC}| = 1$), then a state in which $\hat{\mathbf{b}}_i \cdot \mathbf{r}_{iC} \neq 0$ is counted as a steric clash and hence not allowed. This gives rise to an entropic force between crowders, which is discussed in more detail in the Section 5.2.3.

It should be noted that the crowding particles in experiments sometimes have a surface charge, causing them to strongly interact with the peptides. This is also the case for biological crowders, where the crowders themselves can be other proteins. Such an interaction is not included in our model.

3 Methods

3.1 Thermodynamic simulations

In this section, we quickly review basics of statistical mechanics and describe how to study the thermodynamics of the stick model using Monte Carlo simulations. In the end, we are interested

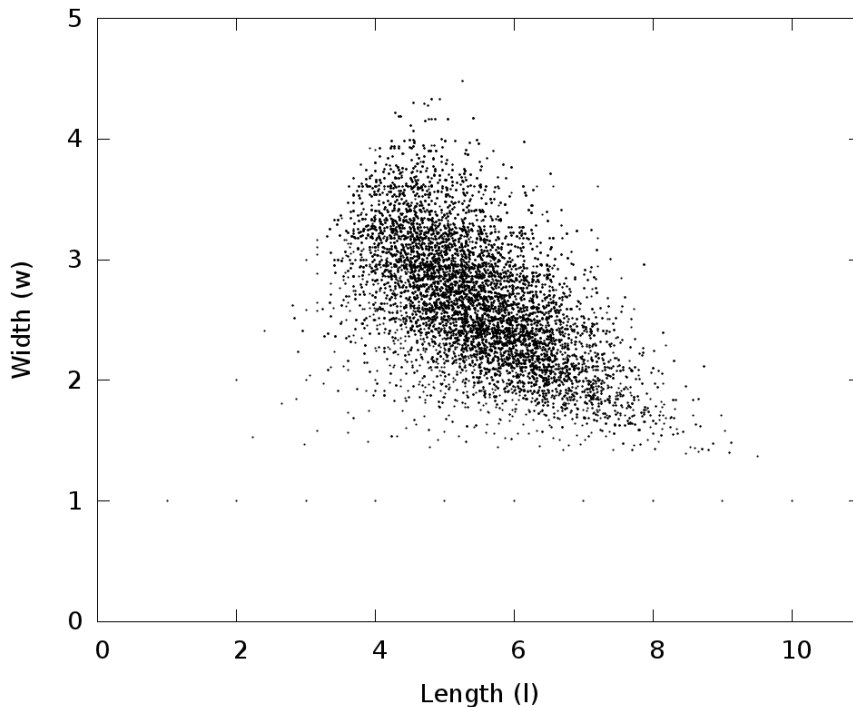


Figure 6: Possible geometries for aggregates containing 10 peptides or less.

in thermodynamic averages of various quantities (e.g. system energy, length and width of the aggregates etc.). The averages are to be taken w.r.t the canonical ensemble, i.e. we imagine that the system is in thermal contact and in equilibrium with a heat bath with a fixed temperature (“thermal contact” meaning that our system can exchange energy with the heat bath, but only while keeping the number of particles constant). Under these circumstances, the different states ν of the system can be shown to be distributed according to the Boltzmann distribution

$$P_\nu(\beta) = \frac{1}{Z(\beta)} e^{-\beta E_\nu}, \quad (3.1.1)$$

where $\beta = 1/k_B T$ is the inverse temperature of the heat bath and E_ν is the total energy of our system. We use units in which the Boltzmann constant $k_B = 1$ and hence express temperatures in relation to the interaction energies between peptides. The normalization constant

$$Z(\beta) = \sum_\nu e^{-\beta E_\nu} = \sum_E g(E) e^{-\beta E}, \quad (3.1.2)$$

is known as the canonical partition function, with $g(E)$ being the density of states (i.e. the number of microstates with total energy E).

Given the canonical ensemble $P_\nu \propto e^{-\beta E_\nu}$, it is instructive to ask what energy the system is most likely to have at a given temperature. The probability distribution over energies becomes

$$P(E) \propto g(E) e^{-\beta E} = e^{\log g(E) - \beta E} \quad (3.1.3)$$

and will be peaked around the energy (or energies) for which $\log g(E) - \beta E$ is maximized. By a trivial rearrangement, this becomes equivalent to minimizing the quantity

$$F = E - TS, \quad (3.1.4)$$

which is known as the Helmholtz free energy, with $S = \log g(E)$ being the entropy of the system at the given energy E . At low temperatures, where the entropic term in Eq. (3.1.4) can be neglected, we see that we are likely to find the system in a state with low energy. For high temperatures, we

can neglect the energy term in Eq. (3.1.4), and are hence likely to find the system at an energy corresponding to many microstates, i.e. with a high entropy. In the stick model, we obtain the lowest energy when all peptides form one big aggregate, and have the highest entropy when the system contains only free monomers.

The partition function is essential in describing the thermodynamics, and can be used to calculate e.g. the average system energy at different temperatures,

$$\langle E \rangle = \sum_{\nu} E_{\nu} P_{\nu}(\beta) = -\frac{\partial \log Z}{\partial \beta}, \quad (3.1.5)$$

or the variance in E ,

$$\langle E^2 \rangle - \langle E \rangle^2 = \frac{\partial^2 \log Z}{\partial \beta^2}, \quad (3.1.6)$$

which is directly related to the heat capacity

$$C \equiv \frac{\partial \langle E \rangle}{\partial T} = \beta^2 \frac{\partial^2 \log Z}{\partial \beta^2} = \beta^2 [\langle E^2 \rangle - \langle E \rangle^2]. \quad (3.1.7)$$

The latter is particularly useful when studying phase transitions. At the transition temperature, one expects diverging fluctuations in thermodynamic quantities such as $\langle E \rangle$ in the infinite system size limit. This means that a phase transition can show up as a spike in the heat capacity at the transition temperature.

Note that we are of course always limited to finite systems with the MC simulations. Even though we in this work will encounter abrupt changes in thermodynamic quantities, the discontinuities associated with true phase transitions will only happen in the infinite system size limit. This should be kept in mind since this distinction is not always explicitly written out in what follows.

In this work we use the Wang-Landau algorithm [16] to compute $g(E)$. This algorithm is very efficient when there are large free energy barriers the system has to cross (which is the case for a first order phase transition). To perform the MC simulation, we need to decide on what updates to use. The naïve choice would be to only to move or to rotate one peptide at a time. In principle, this is enough since all states of the system can be reached by these single peptide updates. However, we will also include translations and rotations of clusters of peptides, according to an algorithm invented by Swendsen and Wang [17], originally intended for spin systems. Besides the obvious advantage with cluster updates that many degrees of freedom are simultaneously updated, the Swendsen-Wang algorithm is capable of breaking up and merging aggregates. A translation of a large aggregate using only single peptide updates would be very unlikely since it means that the aggregate first has to be dissembled at the old location, then that all peptides in the aggregate individually are moved and finally that the aggregate is assembled at the new location.

3.1.1 Swendsen-Wang cluster updates

The cluster update is initiated by picking one peptide i in the system at random. Each neighbouring peptide j are identified and added to the cluster with probability

$$P_{\text{add}}(i, j) = 1 - e^{\beta_{\text{SW}} \varepsilon_{ij}} \quad (3.1.8)$$

where ε_{ij} is the interaction energy between peptide i and j . Recursively, the neighbours (that are not already members are clusters) of each added peptide are proposed as cluster members until no more peptides are added to the cluster.

When a cluster has been constructed, a rigid body translation or rotation of the cluster is proposed. A translation amounts to moving all cluster members one step in one of the 6 lattice directions (picked at random with equal probabilities). In the case of a rotation, one peptide in the cluster is randomly picked to serve as the pivot point. The plane of rotation is any of the 3 orthogonal planes along the lattice coordinates and the rotation angle is either $\pi/2$, π or $3\pi/2$ (picked from a flat probability distribution).

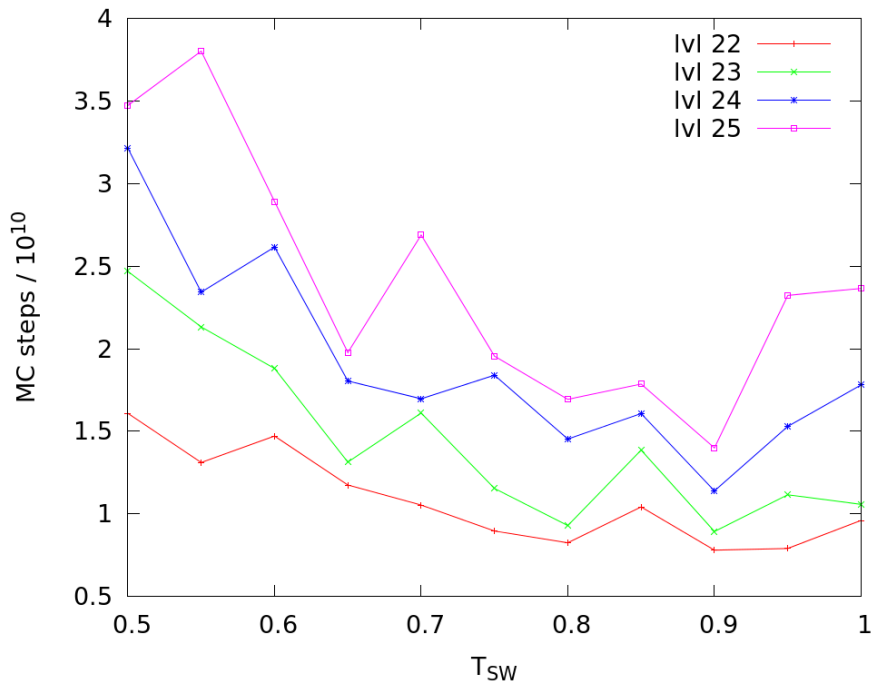


Figure 7: Number of MC steps (i.e. number of proposed updates) needed for the Wang-Landau algorithm to converge in 22 - 25 levels at different T_{SW} .

The update is directly rejected if there is a steric clash. Otherwise, the update is accepted with probability

$$P_{acc}(\nu \rightarrow \nu') = \min \left(1, \frac{g(E_\nu)e^{-\beta_{SW}E_\nu}}{g(E_{\nu'})e^{-\beta_{SW}E_{\nu'}}} \right). \quad (3.1.9)$$

Here, $\beta_{SW} = 1/T_{SW}$ is just a parameter which the final result will not depend on. In the simulations, we picked $T_{SW} = 0.8$ which is slightly above the transition temperature $T_c \approx 0.62$ for the system studied in this work. The choice of T_{SW} was based on preliminary runs where it was observed that the convergence time for the Wang-Landau algorithm (described in the next section) was fastest in this region (see Figure 7).

In Appendix A, it is shown that this procedure satisfies detailed balance with respect to the ensemble $P_\nu \propto 1/g(E_\nu)$ which is used in the Wang-Landau algorithm.

3.1.2 Wang-Landau sampling

To investigate the equilibrium properties of the system, we use *Wang-Landau sampling* [16] to estimate the density of states $g(E)$ (E is the total energy of the system).

If one were to make a MC simulation of a system by proposing updates and accepting them with a probability proportional to $1/g(E)$, the resulting histogram $h(E)$ over visited energies will be flat. The Wang-Landau method exploits this by initially assuming a constant density of states $g(E) = 1$ and then, during the simulation, systematically modifying $g(E)$ while sampling the energy until $h(E)$ is sufficiently flat.

As updates, we use crowder translation, and translation and rotation of single peptides as well as of Swendsen-Wang clusters (10% crowder translation, 30% single peptide translation, 30% single peptide rotation, 15% cluster translation and 15% cluster rotation). The crowder translations (which never change the system energy) are accepted whenever there are no steric clashes. For the single peptide updates, the direction of translation, or the angle and plane of rotation are chosen

symmetrically. A single peptide update from state ν to ν' is accepted with probability

$$P_{\text{acc}}(\nu \rightarrow \nu') = \min\left(1, \frac{g(E_{\nu'})}{g(E_{\nu})}\right). \quad (3.1.10)$$

After each step (accepted or rejected) the energy histogram and the density of states are updated as

$$h(E_f) \rightarrow h(E_f) + 1, \quad (3.1.11)$$

$$g(E_f) \rightarrow f \cdot g(E_f), \quad (3.1.12)$$

where E_f is the new energy (i.e. $E_f = E_{\nu'}$ if the update is accepted, and $E_f = E_{\nu}$ otherwise). The modification factor f initially has a large value $f = e^1$ which ensures that the system quickly samples all energies. Following every update, we check if the histogram is sufficiently flat. In this context, “sufficiently flat” conventionally means

$$\tilde{h}(E_i) \geq 0.7\langle h \rangle \quad \forall E_i, \quad (3.1.13)$$

where $\langle h \rangle$ is the arithmetic mean of $h(E)$, and

$$\tilde{h}(E_i) = \frac{1}{7} \sum_{k=-3}^3 h(E_{i+k}), \quad (3.1.14)$$

is a smoothed version of the histogram.

Once Eq. 3.1.13 is satisfied we reset the histogram, i.e. $h(E_i) = 0$ for all E_i , and update the modification factor f as

$$f \rightarrow \sqrt{f}. \quad (3.1.15)$$

The procedure is then repeated with the new f until the flatness condition for $h(E)$ is once again reached, which is followed by updating f and resetting $h(E)$.

This scheme is repeated 25 times where, during the final “Wang-Landau level”, the modification factor $f = e^{2^{-24}} \approx 1 + 6 \cdot 10^{-8}$ is very close to unity. The obtained $g(E)$ will differ from the actual density of states only by a multiplicative constant. The absolute density of states could in principle be found by considering the degeneracy of the ground state of the system. However, in all cases considered in this report, this multiplicative constant in front of $g(E)$ will trivially cancel out.

3.1.3 Equilibrium measurements

It should be noted that detailed balance is violated in the Wang-Landau algorithm since $g(E)$ is updated after each MC step (it is restored in the limit $f \rightarrow 1$). But the problem of the small detailed balance violation can be evaded by an additional simulation in the ensemble $P_{\nu} \propto 1/\tilde{g}(E_{\nu})$, but now at fixed \tilde{g} (\tilde{g} is the Wang-Landau estimate of the density of states). The system is sampled (i.e. measuring the total energy and what kinds of aggregates that are present) every 10^5 th MC step. The simulation is stopped when both the highest and lowest energy have been visited 15 times, ensuring that the whole energy interval has been thoroughly sampled. The collected energy histogram yields, after normalization, an observed energy distribution $\tilde{P}(E)$ which is approximately flat. Since the acceptance probabilities are weighted as $1/\tilde{g}(E)$, the observed energy distribution will approach

$$\tilde{P}(E) \rightarrow \frac{g_{\text{True}}(E)}{\tilde{g}(E)} \quad (3.1.16)$$

where $g_{\text{True}}(E)$ is the true density of states. Hence, by using,

$$g(E) = \tilde{P}(E)\tilde{g}(E) \quad (3.1.17)$$

we eliminate any bias coming from the detailed balance violation from the Wang-Landau algorithm (in the infinite sampling time limit).

Given $g(E)$, we can calculate canonical averages of observables from the generalized-ensemble simulation. If an observable $O(E)$ only depends on the system energy then

$$\langle O(E) \rangle(\beta) = \sum_{\text{states } \nu} O(E_\nu) P_\nu(\beta) \quad (3.1.18)$$

$$= \sum_{\nu} O(E_\nu) \frac{e^{-\beta E_\nu}}{Z(\beta)} \quad (3.1.19)$$

$$= \frac{\sum_{\nu} O(E_\nu) e^{-\beta E_\nu}}{\sum_{\nu} e^{-\beta E_\nu}} \quad (3.1.20)$$

$$= \frac{\sum_E O(E) g(E) e^{-\beta E}}{\sum_E g(E) e^{-\beta E}}. \quad (3.1.21)$$

In particular, we can calculate $\langle E \rangle$ and $\langle E^2 \rangle$ with Eq. (3.1.21), from which we obtain the heat capacity

$$C = \frac{\langle E^2 \rangle - \langle E \rangle^2}{T^2}. \quad (3.1.22)$$

In this work, we are often interested in observables describing the aggregate geometries (such as average width, average distribution of lengths, average number of peptides per aggregate etc.), which are not directly given by the system energy. Let $O(M)$ be a function of such a generic quantity M . Then

$$\langle O(M) \rangle(\beta) = \sum_{\text{states } \nu} O(M_\nu) P_\nu(\beta) \quad (3.1.23)$$

$$= \sum_E \sum_M g(E, M) O(M) \frac{e^{-\beta E}}{Z(\beta)} \quad (3.1.24)$$

$$= \sum_E \frac{\sum_M g(E, M) O(M)}{\sum_M g(E, M)} \sum_M g(E, M) \frac{e^{-\beta E}}{Z(\beta)} \quad (3.1.25)$$

$$= \sum_E \langle\langle O(M) \rangle\rangle(E) \frac{g(E) e^{-\beta E}}{Z(\beta)} \quad (3.1.26)$$

$$= \sum_{\nu} \langle\langle O(M) \rangle\rangle(E_\nu) P_\nu(\beta), \quad (3.1.27)$$

with

$$\langle\langle O(M) \rangle\rangle(E) = \frac{\sum_M g(E, M) O(M)}{\sum_M g(E, M)}, \quad (3.1.28)$$

being the microcanonical average of $O(M)$ at energy E . The microcanonical average is very convenient to compute during the simulation at fixed \tilde{g} , and does not require large 2-dimensional histograms $P(E, M)$.

3.2 Kinetic Simulations

When doing kinetic MC simulations (i.e. studying the time evolution) of a system, it can be difficult to disentangle the physical results from the unphysical effects originating from the particular kinds of updates used. In this work, we only use single peptide updates and crowder translations in the kinetic simulations, motivated by experimental indications that amyloid fibrils dominantly grow by monomer addition at the fibril ends [18]. However, one should keep in mind that it might be possible to construct updates that are more physically reasonable by including certain cluster translations and rotations. In particular, one would ideally like to allow for secondary nucleation mechanisms such as fragmentation or possibly the feedback loop mechanism that was recently discovered [19].

In reality, a large crowder will move less than a small crowder. We implement this idea in the ratio of the number of peptide to crowder updates. Stoke's law for viscous flow around a small

spherical object with radius R , gives the viscous friction coefficient ζ as $\zeta = 6\pi\eta R$ where η is the fluid viscosity. We assume that a similar dependence holds for the crowders in our system,

$$\zeta_{R_C} \propto R_C. \quad (3.2.1)$$

The Einstein-Smoluchowski relation, $D\zeta = k_B T$ relates ζ to the diffusion constant D . For a crowder with size R_C , we therefore have

$$D_{R_C} = \frac{D_{R_C=1}}{R_C}. \quad (3.2.2)$$

For simplicity, and to keep the number of parameters at a minimum, we assume that a crowder with $R_C = 1$ diffuses as fast as a peptide, i.e.

$$D_{R_C=1} = D_{\text{Pep.}}. \quad (3.2.3)$$

For a freely diffusing particle in d dimensions with diffusion constant D and position $\vec{r} = \vec{0}$ at $t = 0$, the average deviation from its initial position, at a later time t , is given by

$$\langle \vec{r}^2 \rangle = 2dDt. \quad (3.2.4)$$

Let $P^{(C)}$ be the probability that an update is a crowder translation, and let $P_{\text{Trans.}}^{(1)}$ and $P_{\text{Rot.}}^{(1)}$ be the probability that the update is a single peptide translation and rotation, respectively. For simplicity, we let

$$P_{\text{Trans.}}^{(1)} = P_{\text{Rot.}}^{(1)}. \quad (3.2.5)$$

We of course also have

$$P^{(C)} + P_{\text{Trans.}}^{(1)} + P_{\text{Rot.}}^{(1)} = 1. \quad (3.2.6)$$

The average number of updates one has to propose before an update becomes a crowder translation is just $1/P^{(C)}$. Suppose that we have a system with N peptides and N_C crowders of size R_C . Eq. (3.2.4) says that the update probabilities should be such that when we have had N_C crowder translations, we also should have had $N \cdot D_{\text{Pep.}}/D_{R_C} = NR_C$ peptide translations on average. This means that

$$\frac{N_C}{P^{(C)}} = R_C \frac{N}{P_{\text{Trans.}}^{(1)}}. \quad (3.2.7)$$

By combining Eqs. (3.2.5), (3.2.6) and (3.2.7) we get the update probabilities as

$$P_{\text{Trans.}}^{(1)} = P_{\text{Rot.}}^{(1)} = \frac{1}{2 + \frac{N_C}{R_C N}}, \quad (3.2.8)$$

and

$$P^{(C)} = \frac{1}{1 + \frac{2R_C N}{N_C}}. \quad (3.2.9)$$

4 Analytical considerations

In this section we attempt to give analytical estimates of the equilibrium behaviour of the stick model. An analytical treatment would be a good complement to the time-consuming MC simulations. Due to long computation times, equilibrium properties of only relatively small systems were studied using MC simulations, but with an analytical tool at hand, we could gain some insight in what would happen in the thermodynamic limit $L \rightarrow \infty$ which is inaccessible to MC simulations.

Our approach is to treat the collection of each type of fibril (i.e. of given length and width) as a collection of non-interacting objects in a grand canonical ensemble (i.e. an ideal gas with fluctuating particle number), and minimize the free energy subject to the constraint of conservation of the

number of the total number of peptides. This method has been used by Oosawa and Kasai in 1969 to study aggregation into helical polymers [20]¹.

In this work, most attention was given to the system with $L = 50$ and $N = 80$ (with number density $N/L^3 = 6.4 \cdot 10^{-4}$). In the following calculations, we assume parameters accordingly (except when otherwise noted) and investigate if this method manages to reproduce simulated equilibrium properties of the system. The ground state (i.e. the state of lowest energy) of this particular system is when all peptides sit in a rectangular aggregate of length 20 and width 4 (i.e. four β -sheets).

4.1 1D growth

To simplify the problem as much as possible, we start by assuming that the fibrils only grow in one dimension, and thus neglect the hydrophobic effect that makes it possible to form multi-layer aggregates. Let $\langle N_l \rangle$ label the average number of fibrils in a solution with length l (i.e. consisting of l peptides in a linear chain). A chemical potential μ_l for species l can be defined through the Helmholtz free energy, F ,

$$\mu_l \equiv \frac{\partial}{\partial \langle N_l \rangle} F(T, V, \langle N_1 \rangle, \langle N_2 \rangle, \dots, \langle N_{l_{\max}} \rangle). \quad (4.1.1)$$

Here, l_{\max} is the maximum length of a fibril and can in principle be arbitrarily large. However, when making comparisons to MC simulations, l_{\max} should be chosen such that it mimics both the finite size of the system and the finite number of available peptides. If e.g. $N < L$ then obviously $l_{\max} = N$. Otherwise, we can set $l_{\max} = L - 1$ since we do not want to count aggregates that are longer than side of the box.

The individual $\langle N_l \rangle$'s may vary with temperature, but we want the average total number of peptides in the solution, $\sum_l l \langle N_l \rangle = N$, to stay constant. Therefore, the function

$$\tilde{F} \equiv F + \lambda \left(\sum_{l=1}^{l_{\max}} l \langle N_l \rangle - N \right), \quad (4.1.2)$$

is minimized at equilibrium. Note that, in the simulations, we of course have an exact conservation of peptides, but that we in this analytical approach allow the total peptide number to fluctuate around N . The relative fluctuations should however become negligible for large systems.

Taking the partial derivatives of (4.1.2) w.r.t. $\langle N_l \rangle$ and the Lagrange multiplier λ , followed by eliminating λ gives the following condition for chemical equilibrium

$$\tilde{l} \mu_l = l \mu_{\tilde{l}} \quad l, \tilde{l} = 1, \dots, l_{\max}. \quad (4.1.3)$$

In the stick model, the energy of a fibril exclusively comes from the energy of the internal bonds. Given that the energy in a single bond between two peptides is $-\Delta$ and that the number of bonds in a fibril of length l is $l - 1$, the total energy of the fibril is

$$E_l = -(l - 1)\Delta. \quad (4.1.4)$$

The canonical partition function for a single fibril of length l is then

$$\begin{aligned} Z_l &= g \int d^3x e^{-\beta E_l} \\ &= gV e^{-\beta E_l} \end{aligned} \quad (4.1.5)$$

where β is the inverse temperature and $g = 24$ is the number of internal degrees of freedom corresponding to the $6 \cdot 4$ possible orientations of the fibril. Neglecting excluded volume effects (which is reasonable for the dilute systems considered in this work) the available volume (i.e. the number of lattice sites) to a fibril is just $V = L^3$.

¹Readers familiar with cosmology might recognize this approach, since it is used to calculate abundances in the equilibrium approximation of e.g. light elements during the Big Bang Nucleosynthesis epoch, or of electrons, protons and hydrogen atoms during Recombination. The resulting equation for the number density of the desired particle is in this context called the *Saha equation*.

The grand canonical partition function for the length l fibrils is

$$\mathcal{Z}_l = \sum_{N_l=0}^{\infty} \frac{z_l^{N_l}}{N_l!} Z_l^{N_l} \quad (4.1.6)$$

$$= \exp \left\{ gV e^{\beta(\mu_l - E_l)} \right\}, \quad (4.1.7)$$

where $z_l = e^{\beta\mu_l}$ is the fugacity, and the factor $1/N_l!$ has been inserted since fibrils of the same length are indistinguishable. This gives the average number of length l fibrils as

$$\langle N_l \rangle = \frac{1}{\beta} \frac{\partial \log \mathcal{Z}_l}{\partial \mu_l} \quad (4.1.8)$$

$$= gV e^{\beta(\mu_l - E_l)}. \quad (4.1.9)$$

In terms of the number densities $n_l \equiv \langle N_l \rangle / V$, this equation becomes

$$n_l = g e^{\beta(\mu_l - E_l)}. \quad (4.1.10)$$

By virtue of eq. (4.1.3), the chemical potentials can be eliminated by taking the ratio

$$\frac{n_l^{\tilde{l}}}{n_l^l} = g^{\tilde{l}-l} e^{-\beta(\tilde{l}E_l - lE_{\tilde{l}})}, \quad (4.1.11)$$

or

$$n_l = n_l^{l/\tilde{l}} g^{1-l/\tilde{l}} \exp \left\{ -\beta \left(E_l - \frac{l}{\tilde{l}} E_{\tilde{l}} \right) \right\} \quad (4.1.12)$$

$$= g e^{-\beta\Delta} \left(\frac{n_{\tilde{l}}}{g e^{-\beta\Delta}} \right)^{l/\tilde{l}} \quad (4.1.13)$$

$$\equiv g e^{-\beta\Delta} y^l \quad (4.1.14)$$

Since this equation holds for any \tilde{l} , the quantity $y(\beta) \equiv (n_{\tilde{l}}(\beta)/g e^{-\beta\Delta})^{1/\tilde{l}}$ is independent of \tilde{l} . With this equation, any number density $n_l(\beta)$ can be calculated once $y(\beta)$ is known². It is clear from the definition of y that y is always positive. From Eq. (4.1.14) we see that

$$n_1 > n_2 > n_3 > \dots > n_{l_{\max}} \quad , \quad y(\beta) < 1, \quad (4.1.15)$$

$$n_1 < n_2 < n_3 < \dots < n_{l_{\max}} \quad , \quad y(\beta) > 1, \quad (4.1.16)$$

and that, at some $T = T_{\text{eq}} = \beta_{\text{eq}}^{-1}$ where $y(\beta_{\text{eq}}) \equiv 1$, we have

$$n_1 = n_2 = n_3 = \dots = n_{l_{\max}} = g e^{-\beta_{\text{eq}}\Delta}. \quad (4.1.17)$$

Hence, for $T < T_{\text{eq}}$ the system is dominated by the longest possible fibrils, while for $T > T_{\text{eq}}$, the system contains mostly monomers. To get an equation for $y(\beta)$, we express the peptide number conservation in terms of the number densities

$$n_{\text{tot}} = \sum_l l n_l, \quad (4.1.18)$$

with $n_{\text{tot}} = N/V$ and substitute eq. (4.1.14) giving

$$n_{\text{tot}} = g e^{-\beta\Delta} \sum_{l=1}^{l_{\max}} l y^l. \quad (4.1.19)$$

For the special case $y(\beta_{\text{eq}}) = 1$, we obtain

$$n_{\text{tot}} = g e^{-\beta_{\text{eq}}\Delta} \frac{l_{\max}(l_{\max} + 1)}{2}, \quad (4.1.20)$$

²In particular, this equation relates the monomer number density n_1 to any other n_l by $n_l = a^{-1}(a n_1)^l$ with $a = e^{\beta\Delta}/g$. This is the *mass action law* which is the starting point in [20].

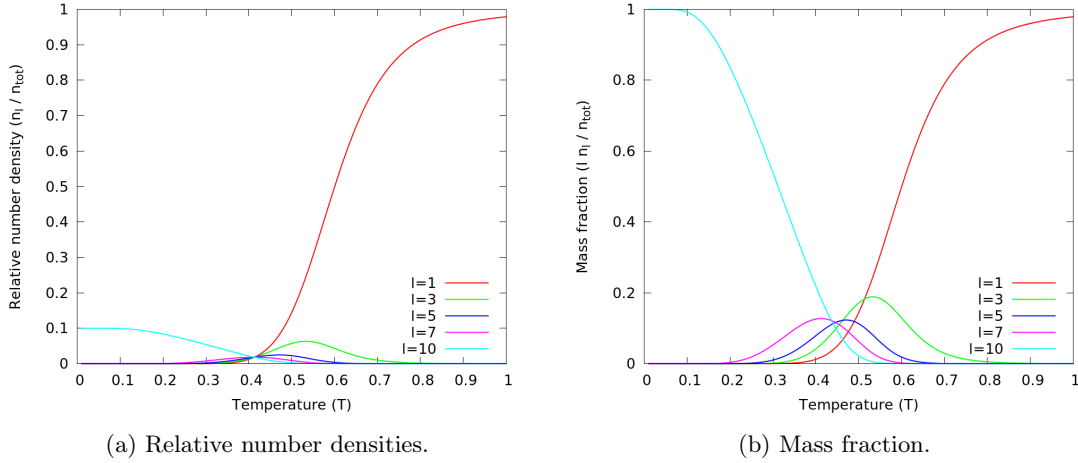


Figure 8: Analytical estimates of relative number densities $n_l(T)/n_{\text{tot}}$ and mass fractions $\rho_l = l \cdot n_l(T)/n_{\text{tot}}$ of fibrils with different lengths for 1D growth. The parameters used are $\Delta = 6$, $l_{\text{max}} = 10$ (corresponding to parallel β structure in the stick model) and $n_{\text{tot}} = 6.4 \cdot 10^{-4}$. Eq. (4.1.21) gives $T_{\text{eq}} \approx 0.413$.

which gives

$$T_{\text{eq}} = \frac{\Delta}{\log \left\{ \frac{g l_{\text{max}} (l_{\text{max}} + 1)}{2 n_{\text{tot}}} \right\}}. \quad (4.1.21)$$

Note that $T_{\text{eq}} \rightarrow 0$ for $l_{\text{max}} \rightarrow \infty$. If l_{max} is interpreted as the system size (i.e. nothing else than the available volume limits the fibril growth), then the system will be monomer dominated for all non-zero temperatures in the thermodynamic limit. In agreement with classical nucleation theory and [21], this shows that we have no critical size for a pure 1D growth.

For $y \neq 1$, the sum in Eq. (4.1.19) can be performed since it is essentially a derivative of an ordinary geometric series

$$n_{\text{tot}} = g e^{-\beta \Delta} y \sum_{l=1}^{l_{\text{max}}} l y^{l-1} \quad (4.1.22)$$

$$= g e^{-\beta \Delta} y \frac{\partial}{\partial y} \sum_{l=0}^{l_{\text{max}}} y^l \quad (4.1.23)$$

$$= g e^{-\beta \Delta} \left\{ \frac{y}{(1-y)^2} - y^{l_{\text{max}}} \left[\frac{l_{\text{max}}}{1-y} + \frac{y}{(1-y)^2} \right] \right\}. \quad (4.1.24)$$

The second term in the curly brackets corresponds to finite volume corrections since it disappears for any temperature $T > 0$ in the infinite volume limit where the fibrils have no maximum length ($l_{\text{max}} \rightarrow \infty$). This equation can be numerically solved for $y(\beta)$ from which one can calculate all number densities with Eq. (4.1.14).

Figure 8a illustrates examples of $n_l(T)/n_{\text{tot}}$ obtained by this method for a system with $l_{\text{max}} = 10$, and in Figure 8b the mass fractions for fibrils of varying lengths are shown as a function of temperature. The mass fraction of fibrils with length l is given by $\rho_l \equiv l \cdot n_l/n_{\text{tot}}$ (a fibril of length l consists of l peptides with the same masses). In this figure, for pedagogical reasons, we used the small $l_{\text{max}} = 10$, since the intermediate n_l 's becomes suppressed and hard to see for larger l_{max} .

4.2 2D growth

The arguments in the previous section are easily extended to account for growth in two dimensions, counting only rectangular aggregates. By noting that a rectangular aggregate with length l and width w consists of $w \cdot l$ peptides, the condition for equilibrium (constrained by the conservation

of the total number of peptides) becomes

$$\tilde{w}\tilde{l}\mu_{w\tilde{l}} = wl\mu_{\tilde{w}\tilde{l}}, \quad (4.2.1)$$

where w , l , \tilde{w} and \tilde{l} take on values allowed by the system size and the number of peptides. The grand canonical partition function for fibrils with length l and width w becomes

$$\mathcal{Z}_{wl} = \exp \left\{ gV e^{\beta[\mu_{wl} - E(w,l)]} \right\}, \quad (4.2.2)$$

where $E(w, l)$ is the total energy of the internal bonds in a fibril with length l and width w . The number densities $n_{wl} = \langle N_{wl} \rangle / V$ are

$$n_{wl} = \frac{1}{\beta V} \frac{\partial \log \mathcal{Z}_{wl}}{\partial \mu_{wl}} \quad (4.2.3)$$

$$= g e^{\beta[\mu_{wl} - E(w,l)]} \quad (4.2.4)$$

$$\equiv g e^{-\beta E(w,l)} z^{wl}, \quad (4.2.5)$$

where $z(\beta) \equiv e^{\beta \frac{\mu_{\tilde{w}\tilde{l}}}{\tilde{w}\tilde{l}}} = e^{\beta \mu_{11}}$ is the monomer fugacity. If $z(\beta)$ is known, this equation can be used to compute any $n_{wl}(\beta)$.

The equation for the conservation of the number of peptides,

$$0 = \sum_{w,l} wl n_{wl} - n_{\text{tot}}, \quad (4.2.6)$$

together with eq. (4.2.5) give a polynomial equation for $z(\beta)$:

$$0 = \sum_{k=0}^{k_{\text{max}}} a_k z^k, \quad (4.2.7)$$

where

$$a_0 = -n_{\text{tot}}/g, \quad (4.2.8)$$

$$a_k = \sum_{w \cdot l = k} wl e^{-\beta E(w,l)}, \quad k > 0, \quad (4.2.9)$$

and k_{max} is the maximal value the product $w \cdot l$ takes for the allowed values of w and l , (i.e. the number of peptides in the largest allowed aggregate).

To construct the energy function $E(w, l)$, assumptions need to be made about how the fibrils are structured. In the stick model, the most energetically favourable way to form a fibril with $w = 1$ is by parallel β -interactions (such a bond is attributed the energy $-\Delta = -6$ in the simulations). In this section we therefore assume that a fibril with $w > 1$ is made up out of (equally long) $w = 1$ fibrils, giving in total $N_{\Delta} = w(l - 1)$ parallel β -bindings. To mimic the asymmetric side chain interactions in the stick model, we assume that the bindings in the width direction are alternating between hydrophobic attractions (with energy $-\delta_{\text{h}} = -2$) and the weak bindings attributed to any two nearest-neighbour peptides with parallel back-bone vectors (the corresponding interaction energy is $-\delta_0 = -1$). Assuming that a fibril with $w = 2$ is built up only by parallel β -interactions and hydrophobic attractions, the energy function becomes

$$E(w, l) = -N_{\Delta}\Delta - N_{\text{h}}\delta_{\text{h}} - N_0\delta_0 \quad (4.2.10)$$

with

$$N_{\text{h}} = \frac{l}{2} \left[w - \frac{1 - (-1)^w}{2} \right] \quad (4.2.11)$$

being the total number of hydrophobic bonds, and

$$N_0 = \frac{l}{2} \left[w - \frac{3 + (-1)^w}{2} \right] \quad (4.2.12)$$

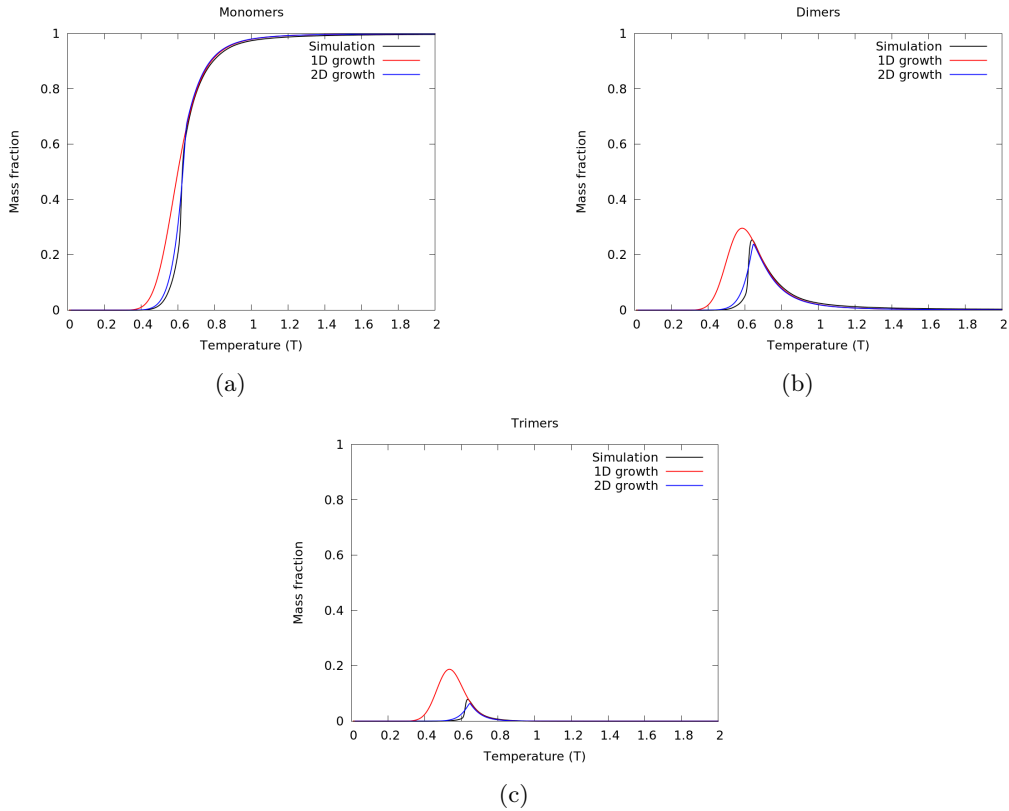


Figure 9: Comparison between analytical estimates of mass fractions and results from MC-simulations. The parameters are $N = 80$ and $L = 50$ (giving $n_{\text{tot}} = 6.4 \cdot 10^{-4}$).

being the total number of “residual” interactions.

We choose the limits on w and l such that they are in no contradiction with the system size and the total number of peptides. The maximum allowed l is $l_{\text{max}} = L - 1$ where L is the length of the sides of the cubic system. The width of a fibril is always bounded by the system size. However, we also do not want to include fibrils that contain more peptides than the entire system. Therefore, the maximum width of a fibril with length l can be expressed as

$$w_{\text{max}}(l) = \min \left[L - 1, \text{floor} \left(\frac{N}{l} \right) \right], \quad (4.2.13)$$

where $\text{floor}(x)$ is the integer part of x . Note that, with this choice of w_{max} , we have included aggregates with $w > l$. The corresponding configurations are possible in the MC simulations, but would be counted with their length and width interchanged. However, in the analytical model, these aggregates are strongly suppressed due to their high energy and hence have a negligible effect.

5 Results

5.1 Analytical thermodynamics

Before describing the results from the simulations, we use the analytical approach to explore the thermodynamics of the stick model in absence of crowders.

Figure 9 shows the mass fractions of the smallest aggregates ($w = 1$ and $l = 1, 2, 3$), obtained from the two analytical approaches and from simulation. Above the transition temperature $T_c \approx 0.62$, both analytical models agree very well with simulations, indicating that the solution is dominated by single-layer aggregates. At T_c , the three simulated mass fractions all have a “kink” which is also present in the 2D growth analytical model. However, this kink is not present in the

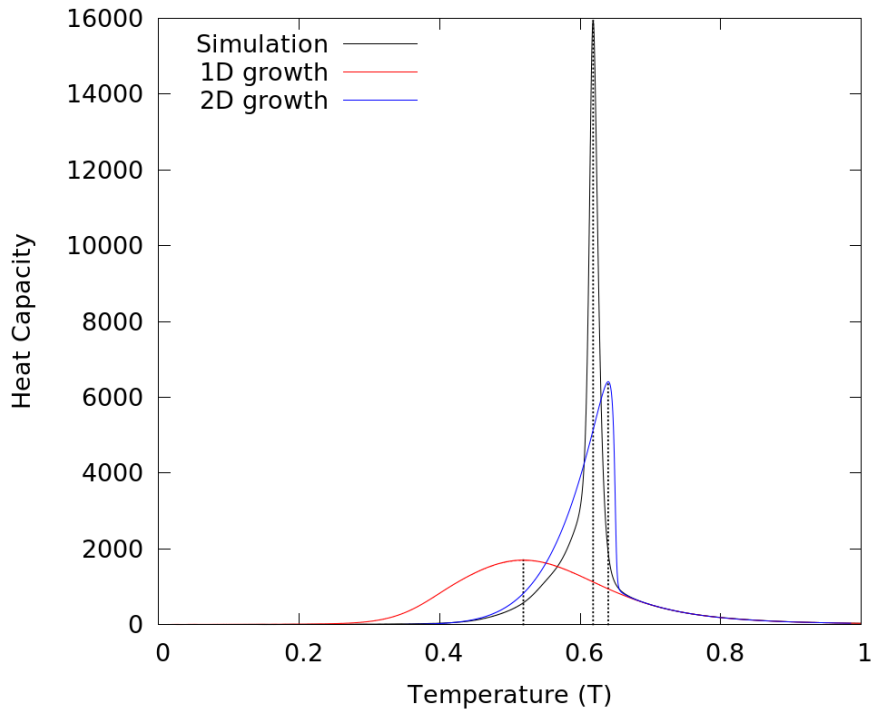


Figure 10: Heat capacities from MC simulation and the two analytical models. The parameters are $N = 80$ and $L = 50$.

1D growth case which is an indication that the transition separates a phase with multi-layer fibrils from a phase containing dominantly fibrils with $w = 1$. Below T_c , the analytical 2D growth model does not quite match the simulations and only qualitatively gives the right results. One reason for this is that the analytical model only counts rectangular aggregates whose principal axes are along the lattice directions, while the simulated aggregates will have more arbitrary shapes (e.g. out of all possible geometries in Figure 6, only 15 are taken into account in the 2D analytical model).

We define the transition temperature T_c such that it corresponds to the peak of the heat capacity $C = \partial E / \partial T$. The system energy is given by

$$E^{(1D)}(T) = \sum_l E_l n_l(T) \quad (5.1.1)$$

for 1D growth, and

$$E^{(2D)}(T) = \sum_{w,l} E(w,l) n_{wl}(T) \quad (5.1.2)$$

in the 2D growth case. The heat capacities can then be computed by numerically taking derivatives with respect to T . Figure 10 shows the simulated heat capacity compared to the heat capacities predicted by the two analytical models. The dashed vertical lines indicate T_c in the three cases. The 2D growth model predicts a transition temperature slightly above the simulated T_c , while the 1D model has a very different behaviour. It is striking that the two analytical models coincide at high temperatures, while there is a dramatic change around T_c where $C^{(2D)}$ suddenly spikes. This spike almost coincides with the spike in the simulated heat capacity, which again suggests that the high temperature phase and low temperature phase could be classified according to whether the system contains only single-layer fibrils, or if there are aggregates present with $w > 1$. It is also clear that spike in the simulated heat capacity is much sharper than in the analytical model.

It has previously been noted [4] that at T_c , the system contains mostly free monomers and short single-layer aggregates, but also a substantial amount of peptides in much larger aggregates. Let $p(m)$ be the probability that when picking a peptide at random, it belongs to an aggregate of

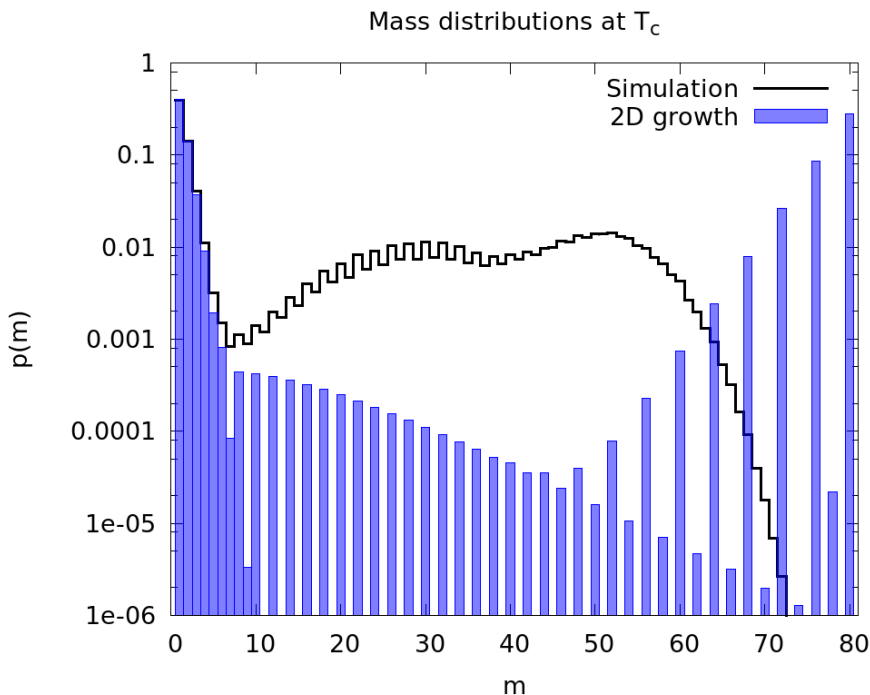


Figure 11: Mass distributions at $T = T_c \approx 0.618$ from simulations and from the analytical 2D growth model with parameters $N = 80$ and $L = 50$.

mass m (i.e. consists of m peptides). In the analytical 2D growth model, $p(m)$ is given by

$$p(m) = \frac{1}{n_{\text{tot}}} \sum_{wl=m} wl \cdot n_{wl}. \quad (5.1.3)$$

Figure 11 shows a comparison between the analytical and the simulated $p(m)$ at $T \approx 0.618$ which is the transition temperature for the simulated system. For the analytical $p(m)$, there is an odd-even effect for $m \gtrsim 8$. This is due to the double-layer aggregates which in the analytical model necessarily have an even number of peptides since both layers are equally long. Similarly, for the heaviest aggregates, the analytical $p(m)$ is large only when m is a multiple of 4, since these aggregates consists of four layers. The analytical model does a good job in describing the smaller (mainly single-layer) aggregates of $m \lesssim 8$, but fails for larger aggregates. There is an exponential suppression for these smaller aggregates in both the simulated and the analytical $p(m)$. However, at $20 \lesssim m \lesssim 60$, there are practically no aggregates in the analytical model, but plenty in the simulations. A reason for this is that the analytical approach doesn't take into account most of the allowed w, l for larger aggregates (those aggregates that are not perfect rectangles and aggregates in which the orientation of the peptides are not energetically optimal). The analytical approach instead favours aggregates of $m \gtrsim 72$, of which there are none in the simulations. Hence, both models predict a kind of coexistence between the smallest and the larger aggregates at $T = T_c$, but the geometry and mass of the larger aggregates are very different.

Using these analytical tools, we can study equilibrium properties of systems much larger than the ones we can study with MC simulations. Figure 12 shows C/V in the 2D growth model for the three system sizes $L = 50, 100$ and 200 . With increasing volume, the behaviour at T_c becomes sharper and sharper, suggesting a discontinuity in the infinite volume limit ($L \rightarrow \infty$).

To quantify the statement that the the system consists of linear aggregates in the high temperature phase, while aggregates with multiple layers start to take over in the low temperature phase, we look at the mass fraction of aggregates with $w > 1$, i.e.

$$\rho_{w>1} = \frac{1}{n_{\text{tot}}} \sum_{l=1}^{l_{\text{max}}} \sum_{w=2}^{w_{\text{max}}(l)} wl n_{wl}. \quad (5.1.4)$$

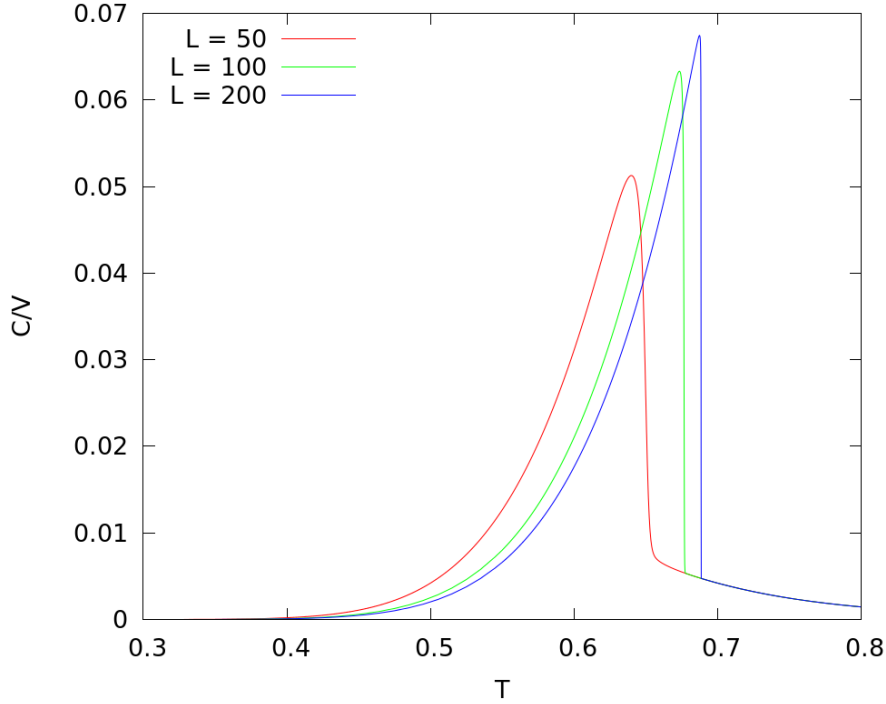


Figure 12: Heat capacities in the 2D growth model at different system sizes, but fixed $n_{\text{tot}} = 6.4 \cdot 10^{-4}$. In the limit $L \rightarrow \infty$ at T_c , there is a finite jump in the (specific) heat capacity.

as a function of temperature (shown in Figure 13). At T_c there is a kink in $\rho_{w>1}$ which becomes sharper and sharper with increasing system size. As the inset in Figure 13 shows, there is still a very small amount of aggregates with $w > 1$ at T_c but this small amount is negligible in comparison with the mass fraction of linear aggregates ($\rho_{w>1}/\rho_{w=1} \approx 10^{-3}$ at T_c).

The data points in Figure 14 show $C(T_c)/V \equiv C_{\text{max}}/V$ and T_c computed for system sizes ranging from $L = 50$ to $L = 1000$. For large volumes, the curves are well fitted by

$$\frac{C_{\text{max}}(V = \infty) - C_{\text{max}}(V)}{V} = \frac{A_1}{V^{\alpha_1}}, \quad (5.1.5)$$

$$T_c(V = \infty) - T_c(V) = \frac{A_2}{V^{\alpha_2}}, \quad (5.1.6)$$

with

$$C_{\text{max}}(V = \infty)/V \approx 0.0696, \quad (5.1.7)$$

$$\alpha_1 \approx 0.5106, \quad (5.1.8)$$

$$A_1 \approx 7.320, \quad (5.1.9)$$

and

$$T_c(V = \infty) \approx 0.6957, \quad (5.1.10)$$

$$\alpha_2 \approx 0.4720, \quad (5.1.11)$$

$$A_2 \approx 14.216. \quad (5.1.12)$$

Unlike T_{eq} in the 1D growth model, the transition temperature T_c in the 2D growth model converges to a non-zero value in the thermodynamic limit. At T_c there is a finite jump in the specific heat capacity, since $C_{\text{max}}(V \rightarrow \infty)/V$ does not diverge. For large V we seem to have the following

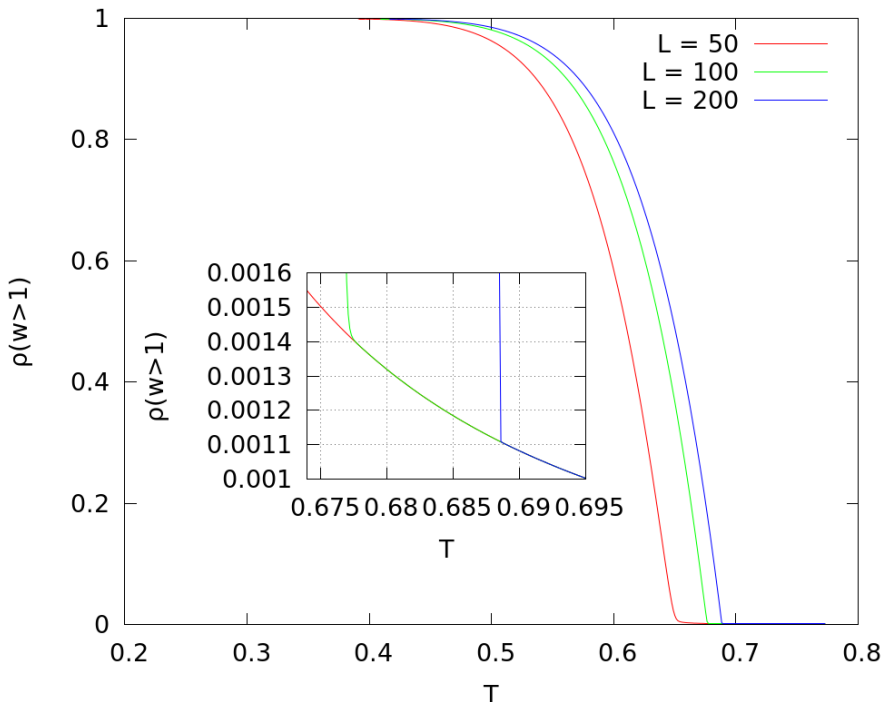


Figure 13: Mass fraction of multi-layered aggregates (i.e. with $w > 1$) in the 2D growth model.

relations

$$T_c(\infty) - T_c(V) \propto \frac{1}{\sqrt{V}}, \quad (5.1.13)$$

$$\frac{C_{\max}(\infty) - C_{\max}(V)}{V} \propto \frac{1}{\sqrt{V}}. \quad (5.1.14)$$

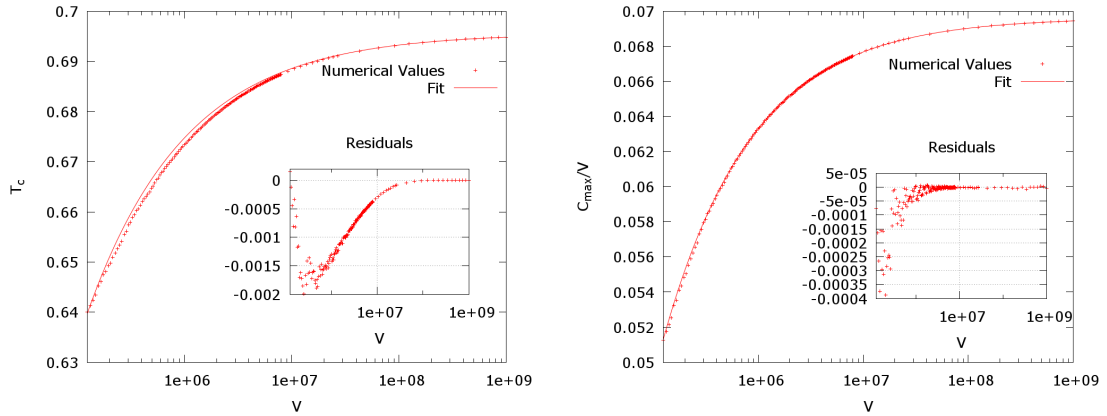
We can of course also compute T_c at different peptide concentrations n_{tot} . This gives a phase diagram which is shown in Figure 15. Here, T_c has been computed in the range $n_{\text{tot}} = 10^{-4} - 5 \cdot 10^{-3}$ for the three system sizes $L = 50, 100$ and 200 . The shape in Figure 15 qualitatively resembles similar phase diagrams computed for nucleating polypeptides [22]. Usually, one considers the critical concentration $n_c(T)$ (or solubility) at a given temperature T , rather than a critical temperature $T_c(n_{\text{tot}})$ at a given peptide concentration. In [22], the relation $n_c(T) = ae^{-b/T}$, where a and b are parameters. In Figure 15, it is not clear to what extent the $n_c(T)$ from the 2D growth follow this relation, due to the short range in n_{tot} for which T_c was computed.

5.2 Simulated thermodynamics

Due to the large number of degrees of freedom, it is computationally demanding to study thermodynamic properties of the stick model for large systems. We therefore choose to study a relatively small system with size $L = 50$ containing $N = 80$ peptides and crowders of varying number and size. With this system, we mainly study the crowding effects in the formation of a single fibrillar aggregate. But before we introduce crowding particles, we should establish what happens in the absence of crowders.

5.2.1 Without crowders

At some temperature T_c , the system undergoes an order-disorder transition. In the high temperature phase, the system contains many small aggregates which is entropically favourable, while in the low temperature phase, the system contains fewer but larger aggregates which is energetically



(a) Transition temperature as a function of system size. (b) Specific heat capacity at T_c as a function of system size.

Figure 14: The insets show the residuals (data points minus the fitted line). The parameters of the fitted lines were found such that the residuals become as small as possible for large volumes.

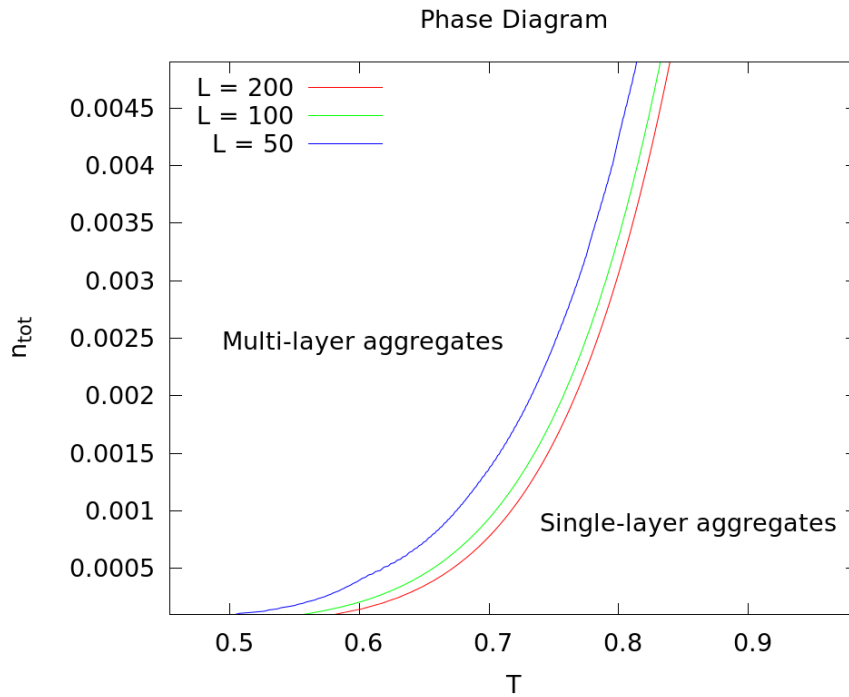


Figure 15: Phase diagram in the 2D growth model for the three system sizes $L = 50, 100$ and 200 . At high temperature, the system is dominated by linear aggregates (mostly monomers). Once the temperature is lowered below T_c , the mass fraction of aggregates with $w > 1$ grows rapidly as shown in Figure 13.

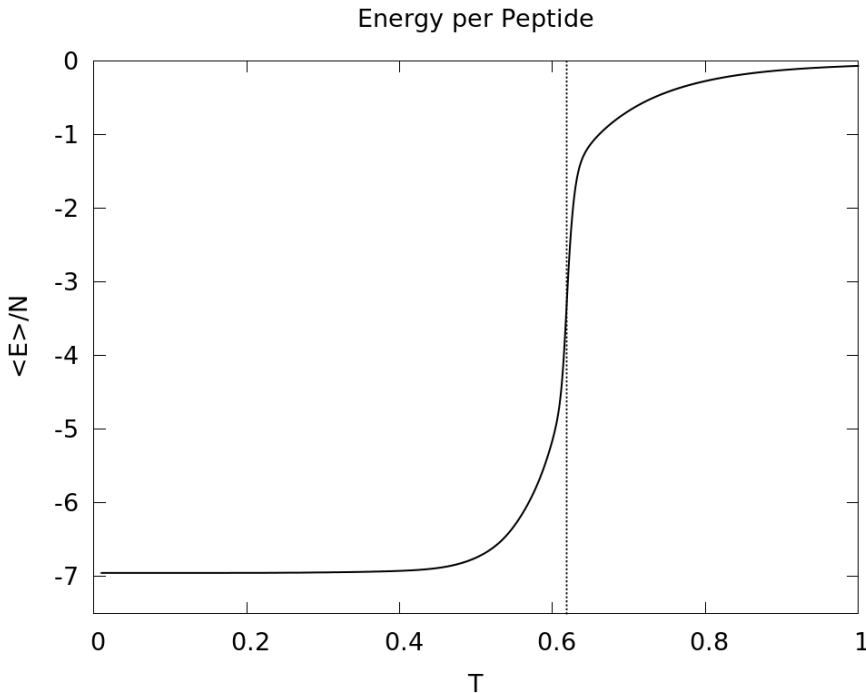


Figure 16: Average energy per peptide at different temperatures for a system with $N = 80$ and $L = 50$. The dashed line indicates the transition temperature.

cally favourable. We define³ the transition temperature T_c as the temperature at which the heat capacity peaks (see the black line in Fig. 10).

Figure 16 shows the average energy per peptide in the system at different temperatures. The transition temperature T_c is indicated by the dashed line. At high temperatures, $\langle E \rangle / N \sim 0$ indicating that we have mostly free monomers. For low temperatures, $\langle E \rangle / N \sim -7$ which is consistent with almost one parallel β bound (with energy $-(1 + a_p) = -6$) per peptide and the occasional hydrogen bond (with energy $-(1 + b) = -2$) or van der Waals bond (with energy -1).

Figures 17-19 considers the mass m_s , length l_s and width w_s of the largest aggregate in the system. Starting from a high temperature, Figure 17 shows that the mass m_s of the largest aggregate slowly increases with decreasing temperature until $T \sim T_c$ is reached. While crossing the transition temperature from above, m_s is increased dramatically from $\sim 5\%$ to $\sim 75\%$ of the total number of peptides in the system. By further decreasing the temperature, m_s continues to approach N but in a less dramatic fashion.

The behaviour of $\langle m_s \rangle(T)$ resembles that of $\langle l_s \rangle(T)$, shown in Figure 18. For temperatures above T_c , the length l_s of the largest aggregate slowly increases with decreasing temperature. While crossing $T = T_c$, $\langle l_s \rangle$ increases from ~ 4 to ~ 14 . A further lowering of the temperature causes l_s to, relatively slow, approach $l_s = 20$.

Figure 19 shows the average width of the largest aggregate. For $T > T_c$, $\langle w_s \rangle \approx 1$, meaning that the system is in the single layer phase. At $T \sim T_c$, there is a drastic jump in the width to $\langle w_s \rangle \approx 4$. This width is also the width of the aggregate in the state of lowest energy for these parameters (the lowest possible energy is $E = -556$ which is achieved when all $N = 80$ peptides form an aggregate with $w = 4$ and $l = 20$). The fact that $\langle w_s \rangle$ has an abrupt change at T_c and remains constant otherwise, points to a phase transition in which $\langle w_s \rangle$ (or some appropriate function⁴ of $\langle w_s \rangle$) conveniently serves as an order parameter. The fact that the abrupt behaviour happens first

³Due to the finite system size (or rather, the finite number of peptides), other phase characterizing thermodynamic quantities than the heat capacity may have their drastic change at temperatures only in the proximity of, but not necessarily equal to, T_c .

⁴For example, the function $(\langle w_s \rangle - 1) / (w_{gs} - 1)$ equals zero in the high temperature phase, and one in the low temperature phase where the ground state width w_{gs} has been reached.

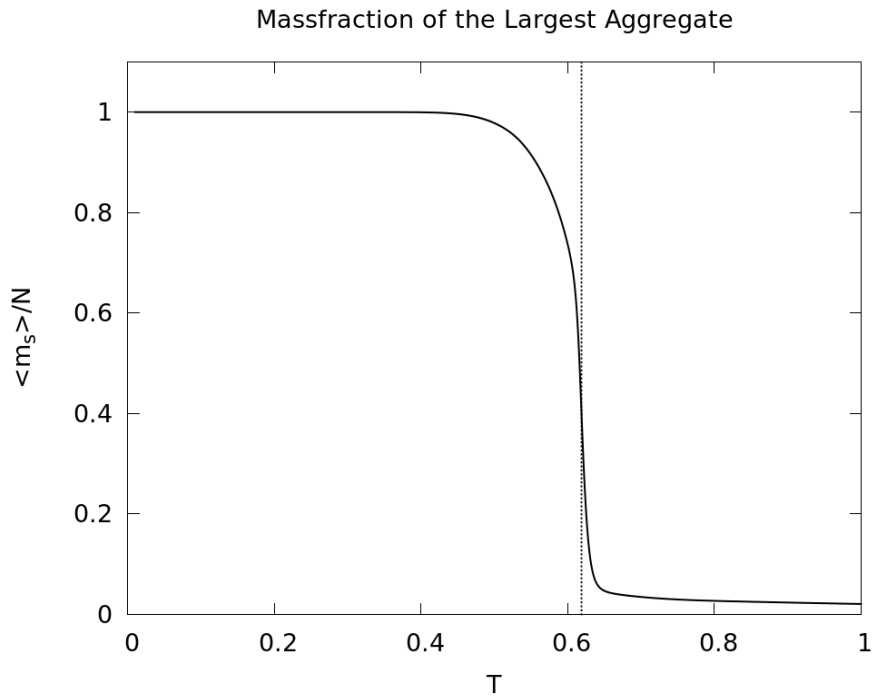


Figure 17: Average mass fraction of the largest aggregate in a system with $N = 80$ and $L = 50$. The dashed line indicates the transition temperature.

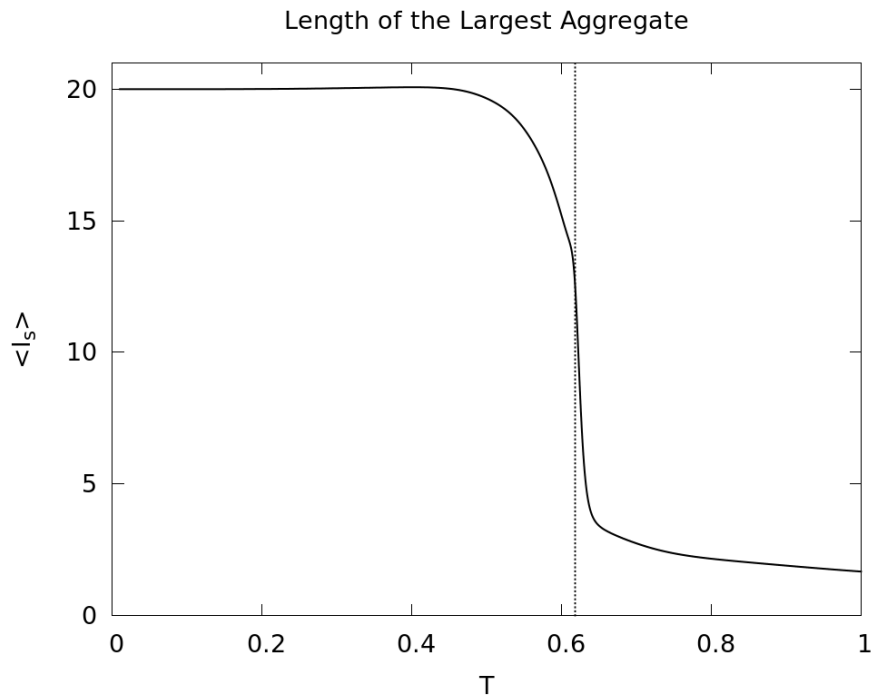


Figure 18: Average length of the largest aggregate in a system with $N = 80$ and $L = 50$. The dashed line indicates the transition temperature.

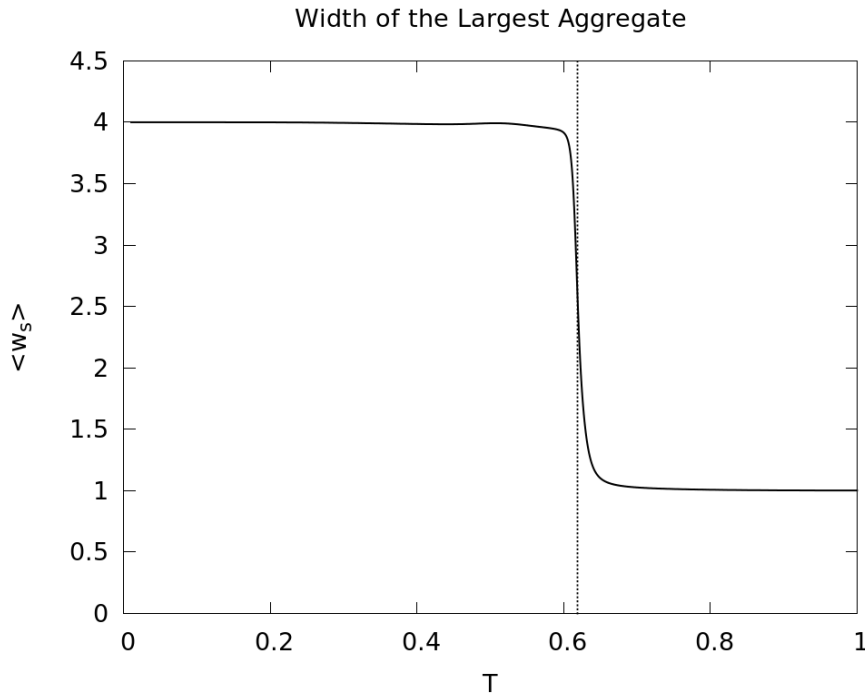


Figure 19: Average width of the largest aggregate in a system with $N = 80$ and $L = 50$. The dashed line indicates the transition temperature.

for $\langle w_s \rangle$ and not e.g. $\langle (w_s - \langle w_s \rangle)^2 \rangle$, or some higher moment, unveils a first-order like nature of the phase transition.

Growth in more than one dimension (i.e. the possibility to form multi-layer aggregates) is certainly necessary for the existence of the phase transition. However, it is not clear if it is a sufficient condition. In this model, the asymmetric side-chain interactions makes aggregates with an odd number of layers (excluding the $w = 1$ case) very unlikely. For our choice of interaction energies, given an even number layered aggregate, it will never be energetically favourable for new peptides to attach in the form of a new layer, rather than to attach at the fibril endpoints or form two new layers⁵. This causes a rather large free energy barrier between even number layered aggregates, which is consistent with the observation of a pairwise β -sheet organization [14]. This barrier would not be present with symmetric side-chain interactions and might in part be responsible for the coexistence at T_c and hence the (first-order like) transition. Further work is required to settle this question.

5.2.2 With crowders

We are now in position to study the effects of crowding particles. In our system of size $L = 50$ containing $N = 80$ peptides, we add $N_C = 24$ crowders of size $R_C = 8$. The volume fraction of the system that is occupied by crowders is then

$$\Phi = \frac{N_C R_C^3}{L^3} \approx 0.098, \quad (5.2.1)$$

i.e. about 10%. This effectively increases the peptide concentration by a factor $(1 - \Phi)^{-1} \approx 1.1$.

Figure 20 shows how the average peptide energy changes with temperature in presence and absence of the crowders. The crowders cause the transition to happen at a slightly higher temperature, which is qualitatively what one would expect just from an effective increase in the density

⁵It is however entropically favourable to increase w by 1 since each new layer can have two orientations w.r.t a neighbouring layer (parallel or anti-parallel hydrogen bond vectors). By adding another layer, the entropic part of the free energy of the aggregate is changed by $-T \log 2$ which is $\lesssim -0.6$ for all temperatures considered in this work. Since this is even less than the energy in a van der Waals bond, this entropic effect can safely be neglected.

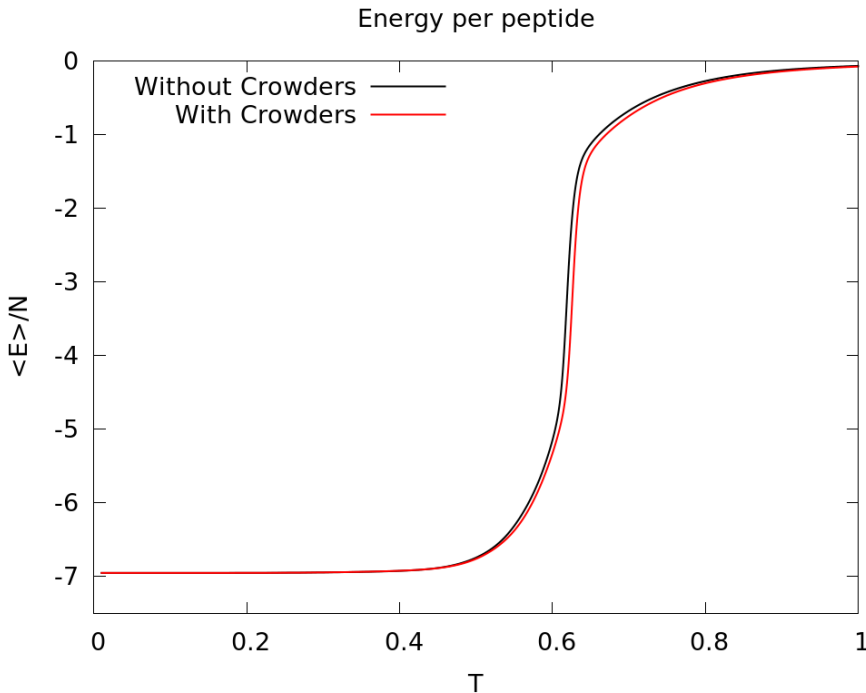


Figure 20: Average peptide energy for $N = 80$ and $L = 50$. The system corresponding to the red line also contains $N_C = 24$ crowders of size $R_C = 8$, which effectively increase the peptide concentration by a factor ~ 1.1 .

of peptides. The small shift in Figure 20 due to the crowders is about 1.2% of T_c . Assuming $T_c \sim 300\text{K}$ in reality, the transition temperature is increased by roughly 3–4K when the crowders occupy $\sim 10\%$ of the available volume. The same effect can be seen in $\langle w_s \rangle(T)$, shown in Figure 21.

We can also study the crowding effects on the aggregate content of the system. Let $p(w, l)\Delta l\Delta w$ be the probability that a peptide, picked at random, is part of an aggregate with length and width within a region $\Delta l\Delta w$ around (l, w) for small Δl and Δw . Figures 22 and 23 display $p(w, l)$ at $T = 0.618$ ($\approx T_c$ in absence of crowders) for the uncrowded and crowded scenarios. In Figure 22, the data points are more scattered than in Figure 23 due to a shorter simulation time. Both figures show that the system contains single-layer aggregates with density decreasing with length. There are also double-layer aggregates in a wide range of lengths. Aggregates with $w \sim 3$ are barely present due to their high energy. The neighbourhood of the ground state width $w = 4$ is well populated in both figures, with varying lengths. The crowding effects are not very dramatic. Mainly, the double-layer aggregates become suppressed in favour of the $w \sim 4$ aggregates which grow a bit longer.

It is of interest to know how the size of the crowders affects the thermodynamics. A question relevant to both experiments [6] and previous simulations [5] is whether it is the total crowder surface area $6N_C R_C^2$ or the total volume $N_C R_C^3$ that matters. For this reason we have, with the $R_C = 8$ and $N_C = 24$ simulation as a starting point, conducted simulations for $R_C = 4$ and $R_C = 16$ while adjusting N_C such that either $N_C R_C^2$ or $N_C R_C^3$ is fixed. Figures 24a and 24b show $\langle E \rangle(T)/N$ for fixed crowder area and volume for a short temperature range in the neighbourhood of T_c . For varying R_C and fixed $N_C R_C^2$, we get quite different results in the three cases. The largest deviation from the $N_C = 0$ case happens for $R_C = 16$ which also is the case where the total crowder volume is the largest. The $R_C = 4$ case, on the other hand, both has the smallest total volume and the smallest deviation from the $N_C = 0$ curve. In contrast, all coloured curves in Figure 24b, representing the fixed total volume cases, almost coincide.

There is however still a small variation among the curves in Figure 24b, where the $R_C = 4$ case yields the largest deviation from $N_C = 0$. This is also the situation among the three cases in

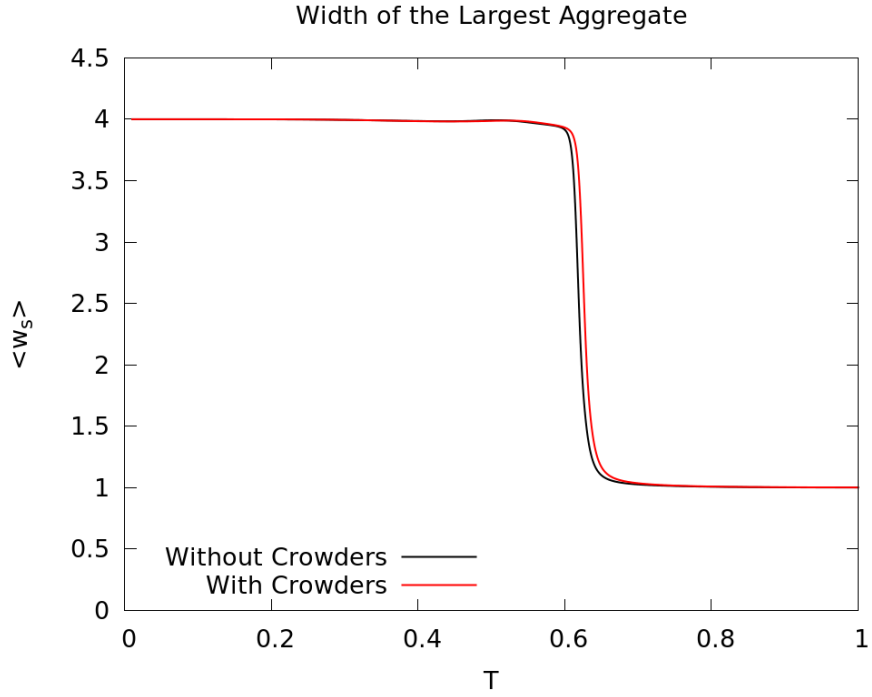


Figure 21: Average width of the largest aggregate in a system with $N = 80$ and $L = 50$. The system corresponding to the red line contains $N_C = 24$ crowders of size $R_C = 8$ in addition to the peptides.

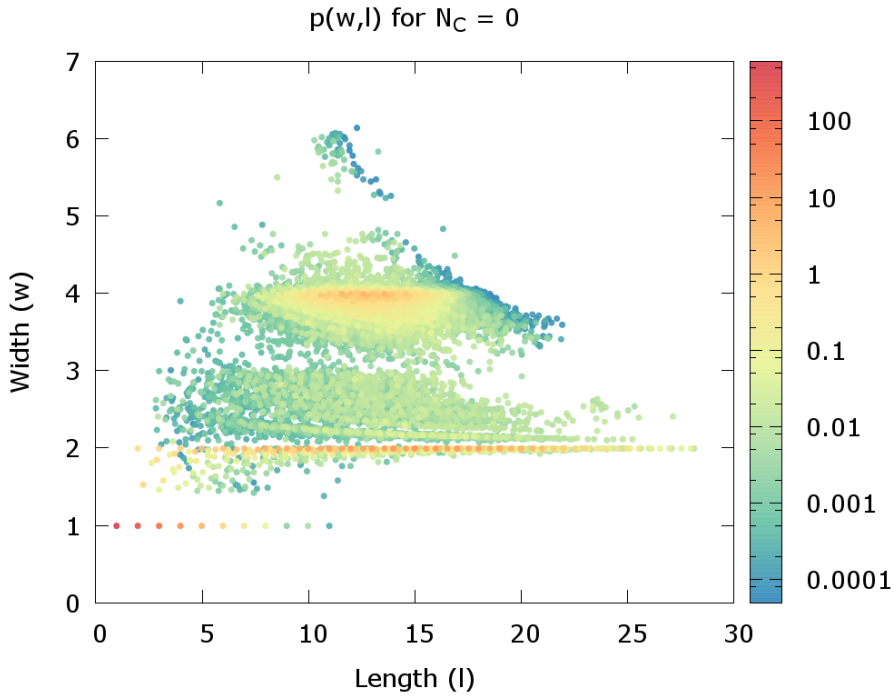


Figure 22: Aggregate mass distribution in absence of crowders at $T = T_c \approx 0.618$ for a system with $N = 80$ and $L = 50$.

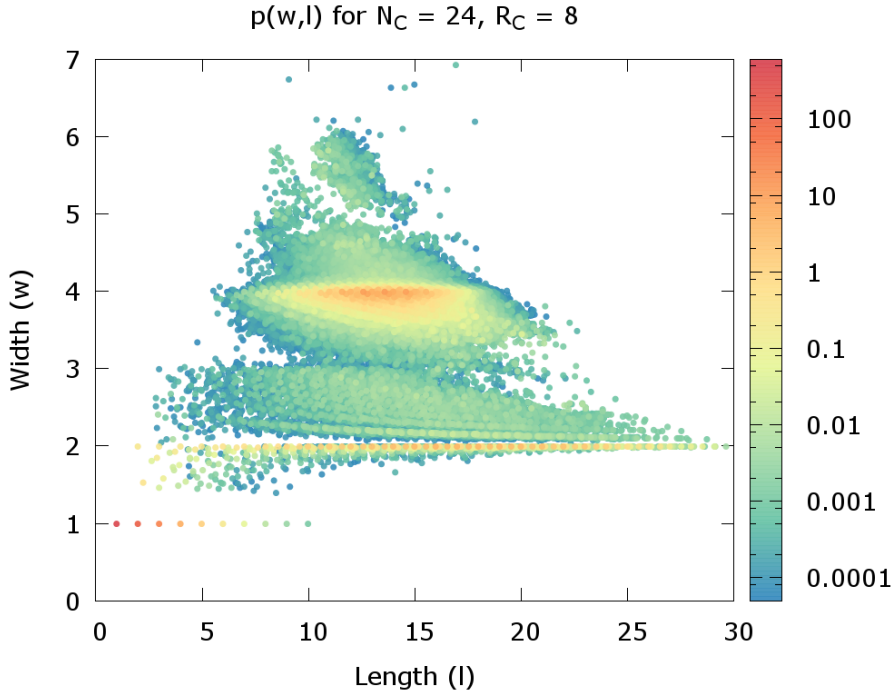


Figure 23: Aggregate mass distribution at $T \approx 0.618$ for a system with $N = 80$, $L = 50$, $N_C = 24$ and $R_C = 8$.

which the total crowder surface area is the largest. When modelling the crowdors, we only allow for states where a backbone vector of a peptide is parallel to the crowder surface when they are nearest neighbours. This removes 8 of the 24 possible peptide orientations at all sites adjacent to the crowdors. One can therefore expect that the effective peptide density is a bit higher for a larger total crowder surface area at fixed total crowder volume. This is likely the reason for the small variation among the three coloured curves in Figure 24b.

Rather than by adding crowdors, we can of course also increase the peptide density by adding more peptides to the system. By adding 9 more peptides to the $N_C = 0$ system, we reach about the same peptide density as the effective density in the systems corresponding to the coloured curves in Figure 24b. The resulting $\langle E \rangle(T)/N$ is indicated as a dashed line in Figure 24b. This curve coincides with the coloured curves to a good approximation, which confirms that it indeed is the total volume of the crowdors that matters.

We reach the same conclusion when considering $\langle w_s \rangle(T)$ which supposedly acts an order parameter for the transition. Figures 25a and 25b show $\langle w_s \rangle(T)$ for the same systems as in Figures 24a and 24b. The variation among the curves corresponding to fixed total surface area is much larger than among the curves corresponding to fixed total crowder volume, again indicating that it is the effective peptide density that matters.

5.2.3 Depletion forces between crowding particles

Even though we have not included any explicit interaction between the crowdors, there is still a possibility that there is a short range depletion force between two crowding particles when they are surrounded by small aggregates. The depletion force would be strongly related to the area of the crowdors, rather than their volume, which makes it a candidate responsible for any area effects. For this reason, we have made a MC simulation of the system with $L = 50$ and $N = 80$ as usual, and with only two crowdors of size $R_C = 16$. For simplicity, we give the two crowdors the same x and y coordinates, and only allow for crowder translations in the z direction. In this way, when the crowdors are adjacent to each other, the crowdors will always be “face to face”. We can then simply study the effective potential as a function of the distance between the crowdors. Let this

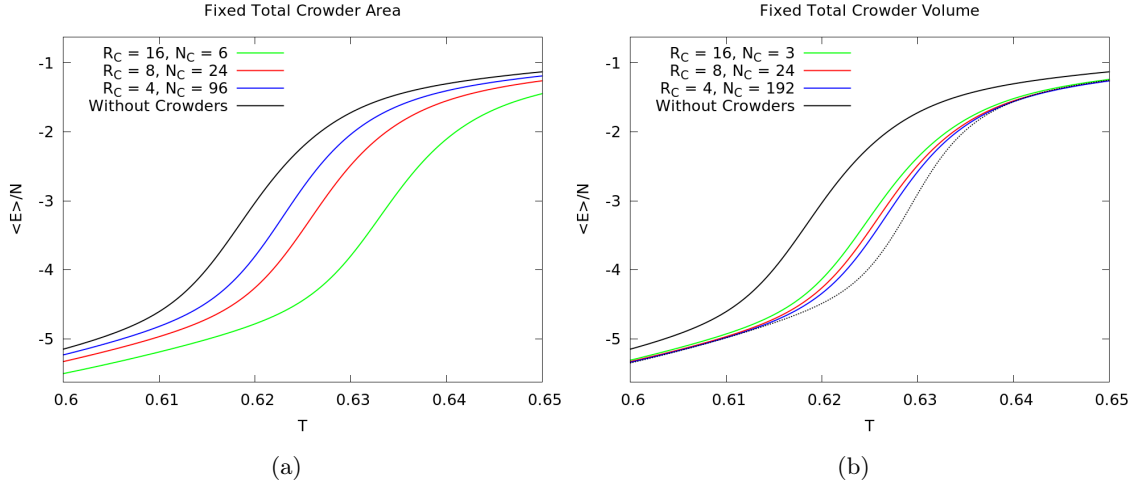


Figure 24: Average peptide energy in a system of size $L = 50$ containing $N = 80$ peptides and various number of crowders of different sizes. In (a), the coloured lines correspond to systems in which the total crowder surface area $6N_C R_C^2$ is held fixed, while in (b) the total crowder volume $N_C R_C^3$ is instead constant. The dashed line in (b) corresponds to a system with $N = 89$ and $N_C = 0$ which effectively gives the same peptide density as in the systems corresponding to the coloured lines.

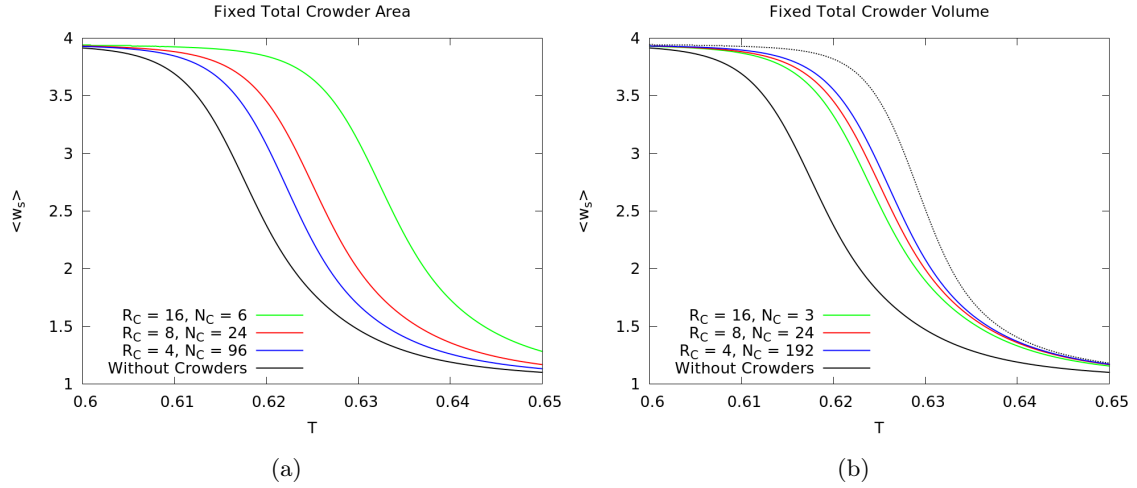


Figure 25: Average width of the largest aggregate in a system of size $L = 50$ containing $N = 80$ peptides and various number of crowders of different sizes. In (a), the coloured lines correspond to systems in which the total crowder surface area $6N_C R_C^2$ is held fixed, while in (b) the total crowder volume $N_C R_C^3$ is instead constant. The dashed line in (b) corresponds to a system with $N = 89$ and $N_C = 0$ which effectively gives the same peptide density as in the systems corresponding to the coloured lines.

distance be represented by d , being the number of lattice points between the two closest faces of the crowders.

The effective potential can be analytically estimated if we neglect the excluded volume effects from the peptides, and assume that the temperature is sufficiently high so that the system to a good approximation only contains free peptides. Let $\Omega(d > 1)$, $\Omega(d = 1)$ and $\Omega(d = 0)$ be the number of states for the peptides for the three relevant crowder separations. For $d > 1$ all $2 \cdot 6$ crowder faces are fully exposed to the peptides. At a lattice site adjacent to a crowder, a peptide is forbidden to have its backbone vector perpendicular to the crowder surface. This removes 8 of the $g = 24$ possible peptide orientations at these $12R_C^2$ sites. Out of the L^3 lattice sites, the $2 \cdot R_C^3$ sites already occupied by the two crowders are of course also inaccessible to the peptides. Thus, by neglecting the peptide excluded volume, we have

$$\Omega(d > 1) = \frac{[gV_{\text{eff}} - 8 \cdot 12R_C^2]^N}{N!}, \quad (5.2.2)$$

with $V_{\text{eff}} = L^3 - 2R_C^3$. For $d = 1$, we effectively have 11 exposed crowder faces, since the lattice sites between the crowders simultaneously are adjacent to both crowders. Hence,

$$\Omega(d = 1) = \frac{[gV_{\text{eff}} - 8 \cdot 11R_C^2]^N}{N!}. \quad (5.2.3)$$

In the case $d = 0$, only 10 crowder faces are exposed to the peptides which results in

$$\Omega(d = 0) = \frac{[gV_{\text{eff}} - 8 \cdot 10R_C^2]^N}{N!}. \quad (5.2.4)$$

The free energy differences for the different crowder separations are then

$$F(d = 1) - F(d > 1) = -T \log \left[\frac{\Omega(d = 1)}{\Omega(d > 1)} \right] \quad (5.2.5)$$

$$\approx -\frac{8R_C^2 N}{gV_{\text{eff}}} T \quad (5.2.6)$$

$$\approx -0.059T, \quad (5.2.7)$$

and similarly

$$F(d = 0) - F(d > 1) = -T \log \left[\frac{\Omega(d = 0)}{\Omega(d > 1)} \right] \quad (5.2.8)$$

$$\approx -\frac{16R_C^2 N}{gV_{\text{eff}}} T \quad (5.2.9)$$

$$\approx -0.118T. \quad (5.2.10)$$

Even at a high temperature, say $T = 0.85$, the entropic attraction between the two crowders only amounts to a free energy gain of ~ -0.1 , i.e. about 10% of the van der Waals force between two peptides. The depletion force in this model between two crowders should therefore not have any significant impact on the aggregation process.

Figure 26 shows the simulated effective potential (i.e. $-T \log P(d; T)$ with $P(d; T)$ being the probability distribution for the crowder separation d at temperature T) normalized to $d = 0$. Indeed, at large temperatures, there is a short range attraction in the same order of magnitude as predicted by Eq. (5.2.10). It is also clear that this attraction is not there when $T < T_c \approx 0.62$. This is reasonable since, at these temperatures, most peptides are in one big aggregate rather than in many small aggregates. One would also expect that the range of the depletion force is increased beyond $d = 1$ when the temperature slightly above T_c . This is because at these temperatures, the system contains many linear aggregates which are larger than free monomers. Hence, the depletion zone surrounding the crowders is increased which in turn increases the range of the depletion force. This effect can (admittedly with a bit of imagination) be seen in Figure 26.

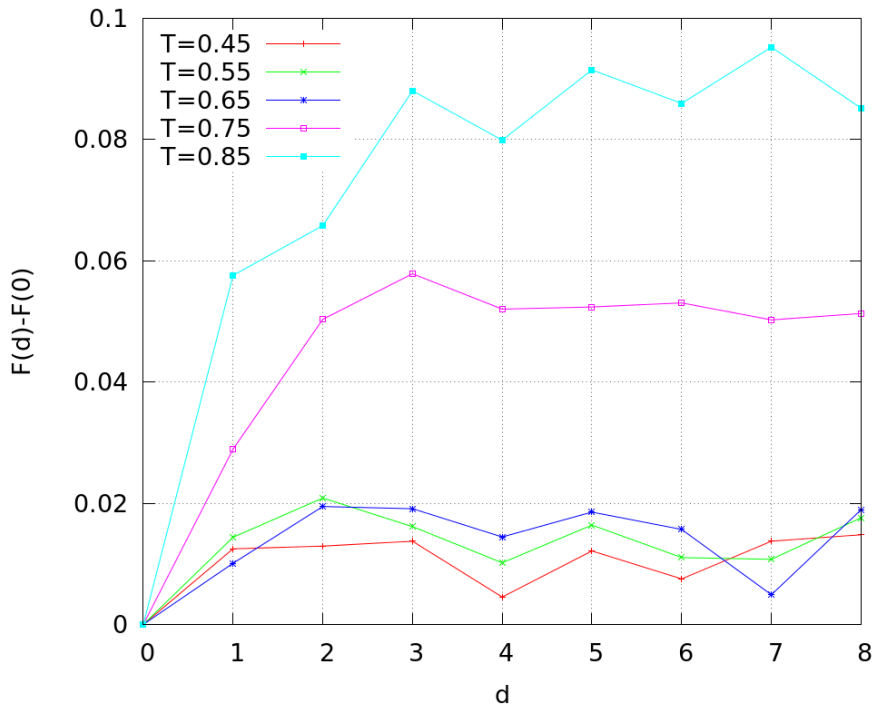


Figure 26: Effective potential between two crowders normalized to $d = 0$ in a system with $L = 50$, $N = 80$, $N_C = 2$ and $R_C = 16$. At high temperatures, there is a small attraction between the crowders at short distances due to the high number of small aggregates.

5.3 Simulated kinetics

Having studied the thermodynamics of the stick model, we now move on to the kinetics. We use a system with $L = 256$ and $N = 10737$ giving $n_{\text{tot}} = 6.4 \cdot 10^{-4}$, like the smaller system studied in the previous section. Thermodynamically, it is dominantly the volume fraction Φ of crowders that matters, and not the crowder size R_C . We therefore only study kinetics for the crowder size $R_C = 16$, and instead vary Φ and T . For each T and Φ , 5 independent runs were conducted, from which the results are averaged.

The system is initiated with random positions and orientations for its constituents. We then update the system using single-peptide updates and crowder translations (with relative frequency as described in Section 3.2). The updates are accepted with probability

$$P_{\text{acc}}(\nu \rightarrow \nu') = \min \left[1, e^{-\beta(E_{\nu'} - E_{\nu})} \right], \quad (5.3.1)$$

such that detailed balance is fulfilled in the canonical ensemble. This ensures that, when the system reaches an equilibrium state, $P_{\nu} \propto e^{-\beta E_{\nu}}$. We measure time in MC sweeps, with one MC sweep defined as N proposed single-peptide updates (accepted or rejected).

We start with a system at $T = 0.67$ and $N_C = 819$ such that $\Phi = 0.2$. The effective peptide density is $(1 - \Phi)^{-1} n_{\text{tot}} = 8 \cdot 10^{-4}$. At this temperature and density, the system turns out to be in the single-layer phase. We then cross the phase boundary by adding crowders such that $\Phi = 0.4$, which increases the effective peptide concentration to $(1 - \Phi)^{-1} n_{\text{tot}} = 1.1 \cdot 10^{-3}$. We further delve into the aggregated phase by considering systems with $\Phi = 0.6$ followed by $\Phi = 0.8$ (giving effective peptide densities of $1.6 \cdot 10^{-3}$ and $3.2 \cdot 10^{-3}$). Then, while keeping Φ fixed at 0.8, we set out to cross the phase boundary by increasing the temperature. The temperature is increased from $T = 0.67$ to $T = 0.81$ in steps of $\Delta T = 0.035$. In Figure 27, the 8 systems considered are indicated in a phase diagram as blue dots, connected with arrows displaying the order in which they are considered. The two red arrows indicate where the phase boundary has been crossed. The red curve is the analytical estimation of the phase boundary at this system size. As seen here and in Section 5.1, the analytical 2D growth model tends to over-estimate the transition temperature T_c at a given

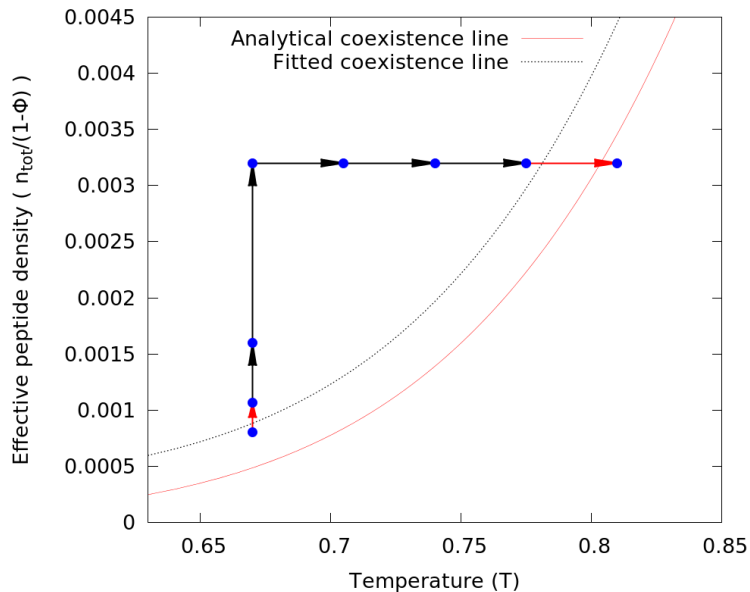


Figure 27: Phase diagram where the blue dots indicate the parameters for which the kinetics has been studied. The red line shows the coexistence line as predicted by the 2D analytical growth model. The red arrows indicate where the coexistence line has been crossed. The dashed curve is the same as the red curve, but shifted such that it crosses both red arrows, and hence is a more realistic estimation of the phase boundary.

peptide density. The dashed curve in Figure 27 is a more likely phase boundary, and was obtained by shifting the analytical curve such that it crosses both red arrows.

Figure 28 shows the time evolution of the aggregate content in the system with $\Phi = 0.2$ and $T = 0.67$. It reaches an equilibrium state extremely quick, where more than 99% of the peptides are in aggregates with $w < 1.1$ (i.e. roughly single-layer).

When increasing the crowder volume fraction to $\Phi = 0.4$, the situation becomes what is shown in Figure 29. Since a significant mass fraction of the aggregates have $w > 1$, we conclude that the system now is in the multi-layer phase. The simulation is aborted at $t \sim 12.5 \cdot 10^6$ MC Sweeps since in one of the five runs, an aggregate with length $l > L - 1 = 255$ was formed. Due to the periodic boundary conditions, these long aggregates can reach across the entire system and bind its endpoints together, which of course is an unphysical effect. We see that aggregates with a wide range of widths are formed relatively quickly. This indicates that we are close to the transition since, at the transition, we expect large fluctuations.

When $\Phi = 0.6$ (see Figure 30), the fraction of single-layer aggregates (mostly monomers) has further decreased. Unlike in the previous case, there are barely any aggregates present with $w \gtrsim 8$. Rather than quickly forming aggregates of many different widths, the aggregation process has a bit more structure. Initially, there is a rapid decrease in $w \sim 1$ aggregates in favour of the $w \sim 2$ aggregates. The fraction of $w > 2$ aggregates then grows mostly at the expense of the $w \sim 2$ aggregates but also of the single-layered. When the double-layer aggregates have depleted, the fraction of single-layer aggregates remains almost constant (slowly decreasing). At the same time, the remaining mass fraction is slowly shifted towards larger w 's.

This structure is further enhanced when $\Phi = 0.8$, shown in Figure 31. Initially, the fraction of double-layer aggregates grows larger at expense of the single-layer aggregates. When the double-layer aggregates are depleted, the fraction of single-layer aggregates stays roughly constant while the remaining aggregates grow wider.

We now keep Φ constant and approach the transition by increasing the temperature. The $T = 0.705$ case is shown in Figure 32. By increasing T , we have shortened the stages at which $w \sim 2$ aggregates and $w \sim 4$ are present. The fraction of single-layer aggregates (mainly monomers) is increased and the larger aggregates are formed faster.

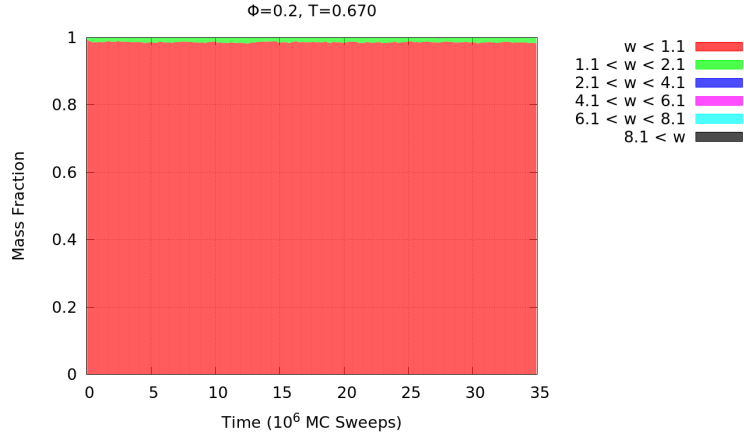


Figure 28: Time-evolution of the mass fractions of aggregates with a variety of widths, with a random initial condition, averaged over 5 runs. The parameters are $T = 0.670$, $L = 256$, $N = 10737$, $N_C = 819$ and $R_C = 16$, giving a crowder volume fraction of $\Phi = 0.2$. No aggregation of multi-layer aggregates happens, indicating that the system is in the single-layer phase.

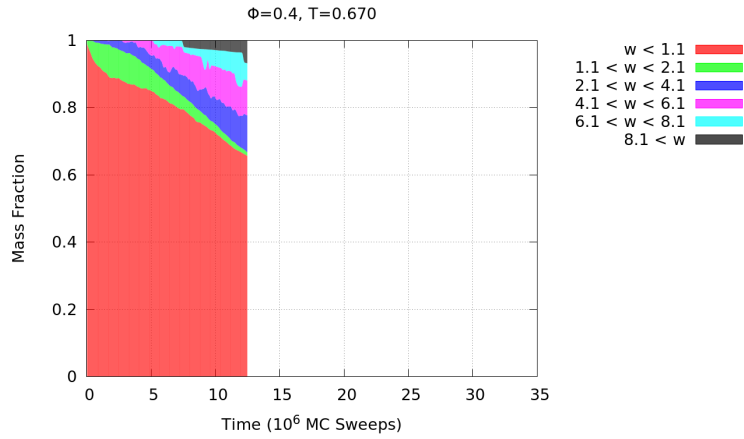


Figure 29: Time-evolution of the mass fractions of aggregates with a variety of widths, with a random initial condition, averaged over 5 runs. The parameters are $T = 0.670$, $L = 256$, $N = 10737$, $N_C = 1638$ and $R_C = 16$, giving a crowder volume fraction of $\Phi = 0.4$.

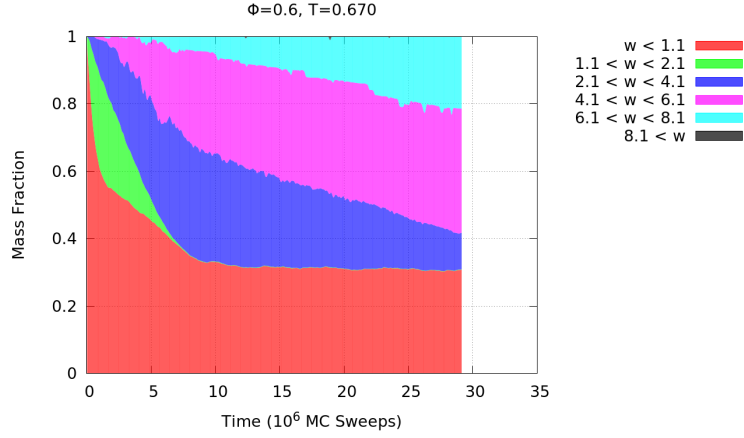


Figure 30: Time-evolution of the mass fractions of aggregates with a variety of widths, with a random initial condition, averaged over 5 runs. The parameters are $T = 0.670$, $L = 256$, $N = 10737$, $N_C = 2458$ and $R_C = 16$, giving a crowder volume fraction of $\Phi = 0.6$.

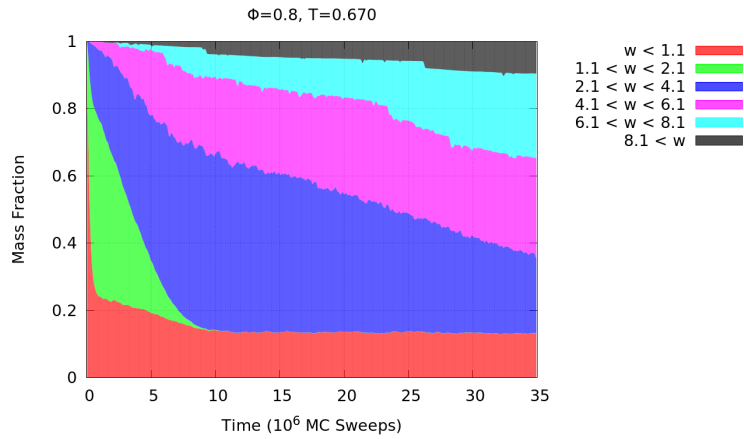


Figure 31: Time-evolution of the mass fractions of aggregates with a variety of widths, with a random initial condition, averaged over 5 runs. The parameters are $T = 0.670$, $L = 256$, $N = 10737$, $N_C = 3277$ and $R_C = 16$, giving a crowder volume fraction of $\Phi = 0.8$.

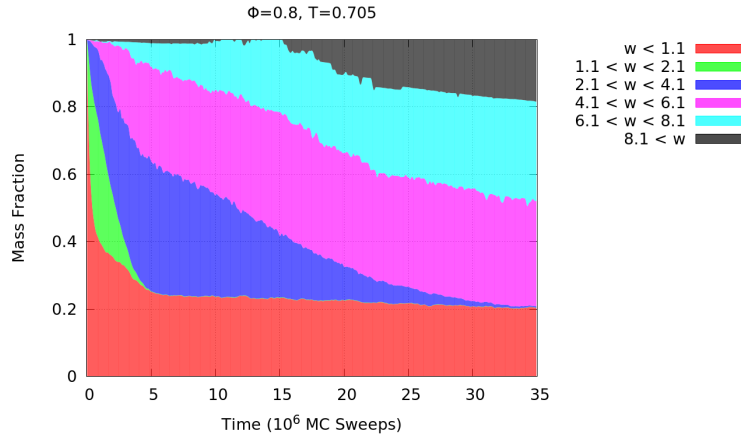


Figure 32: Time-evolution of the mass fractions of aggregates with a variety of widths, with a random initial condition, averaged over 5 runs. The parameters are $T = 0.705$, $L = 256$, $N = 10737$, $N_C = 3277$ and $R_C = 16$, giving a crowder volume fraction of $\Phi = 0.8$.

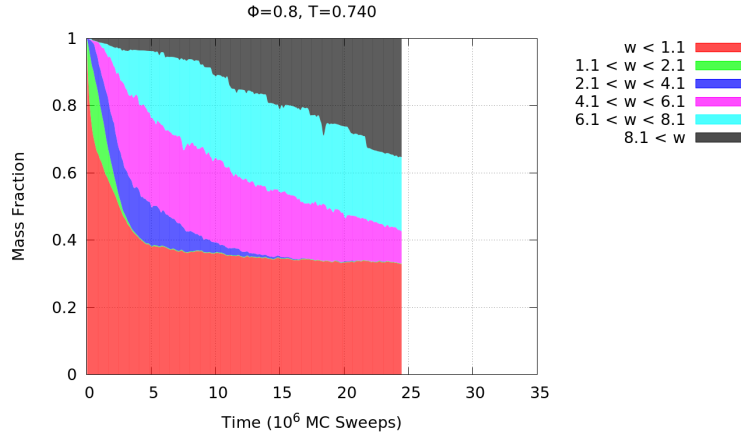


Figure 33: Time-evolution of the mass fractions of aggregates with a variety of widths, with a random initial condition, averaged over 5 runs. The parameters are $T = 0.740$, $L = 256$, $N = 10737$, $N_C = 3277$ and $R_C = 16$, giving a crowder volume fraction of $\Phi = 0.8$.

The trend continues in Figure 33 where the temperature has been increased to $T = 0.740$. While the fraction of single-layer aggregates has increased, the stage when $w \sim 2$ aggregates are present is almost vanishingly short. Similarly, the $w \sim 4$ aggregates are present under a relatively short time, and never in a large fraction. Instead, the wider aggregates are formed quite fast.

When $T = 0.775$, it is apparent from Figure 34 that we are close to the transition temperature. There is a large fraction of single-layer aggregates present, while larger aggregates are formed very quickly and in a very wide range of widths.

By a further increase in temperature to $T = 0.81$, Figure 35 shows that the system now is in the single-layer phase, where no formation of fibrillar aggregates happens.

One effect we can see from Figures 31-34 is that when we increase the temperature, the number of free monomers (these constitute the majority of the single-layer aggregates) is increasing. This is reasonable since they give the dominant contribution to the entropy of the system which should increase with rising temperature.

A less intuitive effect, seen in Figures 31-34, is that the formation of the larger aggregates is faster when increasing the temperature (see e.g. the $w > 8.1$ aggregates in Figures 31-34). This is reminiscent of what has been seen experimentally, where, for instance, it has been observed

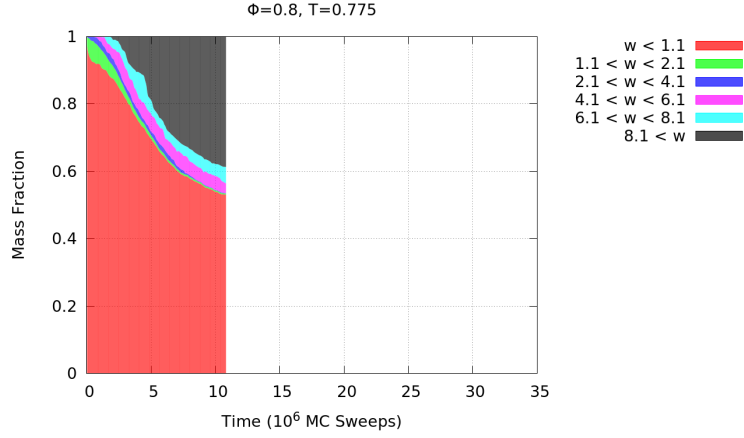


Figure 34: Time-evolution of the mass fractions of aggregates with a variety of widths, with a random initial condition, averaged over 5 runs. The parameters are $T = 0.775$, $L = 256$, $N = 10737$, $N_C = 3277$ and $R_C = 16$, giving a crowder volume fraction of $\Phi = 0.8$.

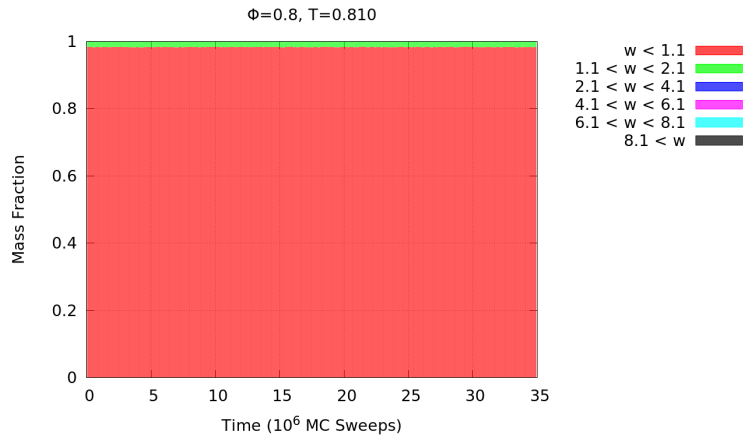


Figure 35: Time-evolution of the mass fractions of aggregates with a variety of widths, with a random initial condition, averaged over 5 runs. The parameters are $T = 0.810$, $L = 256$, $N = 10737$, $N_C = 3277$ and $R_C = 16$, giving a crowder volume fraction of $\Phi = 0.8$.

that the fibrillation of α -synuclein is strongly enhanced with an elevated temperature [23]. In the stick model, this might be related to the increasing number of free monomers that comes with a higher temperature. For an even number layered aggregate with width w to increase its width to $w + 2$, it first has to cross the unstable state with width $w + 1$ (i.e. an odd number of layers). The crossing of the corresponding free energy barrier is done by addition of single monomers, increasing the aggregate width. With the increasing number of free monomers that come with the elevated temperature, one would expect that the formation of aggregates with large widths happens faster. In contrast, at low temperatures the free monomers quickly form many small aggregates that are hard to break up due to the low temperature, and that have a more limited mobility (especially since we in the kinetic simulations only include single peptide updates).

6 Summary

In this work, we have employed MC simulations to study crowding effects on the formation of amyloid fibrils, by adding neutral crowding particles to a simplified lattice model. We have found that the presence of crowdors increases the transition temperature T_c , which separates a high entropy phase containing many short single-layer aggregates (mostly free monomers) from a low energy phase in which the peptides form large aggregates with multiple layers. The increase in T_c is largely consistent with the effective increase in peptide concentration due to the excluded volume of the crowdors. However, a small effect depending on the total crowder area was observed, that likely originates from the constrained orientation of the peptides when adjacent to a crowder surface.

We have also studied the model in absence of crowdors, and discovered an analytical approach to yield approximate results on the thermodynamics of the model. In this approach, non-rectangular aggregates are neglected and the collection of each type of aggregate (i.e. of a given length and width) is treated as a collection of non-interacting objects in the grand-canonical ensemble, with the over-all constraint of the conservation of the total number of peptides. The method predicts a critical-like behaviour similar to what is seen in the MC simulations. It gives an accurate description of the single-layer aggregates, but only qualitatively reproduces the simulated results of the multi-layer aggregates. Given a peptide density, the method is relatively successful in predicting the transition temperature T_c of the system. By computing T_c in a range of peptide concentrations, we have constructed a phase diagram for the model (which would be extremely time-consuming using only MC simulations). The analytical approach has the advantage that systems much larger than in the MC simulations can be studied. By computing T_c and the specific heat capacity $C(T_c)/V$ for increasing system sizes, we can fit a curve the data points which in turn predicts what happens in the limit $V \rightarrow \infty$. This analysis strongly suggests that, in the analytical approach, T_c converges to a finite value in the thermodynamic limit, where the heat capacity seems to exhibit a finite jump.

It was found that the observed order-disorder transition is strongly related to the ability for the aggregates to grow in more than one dimension. Considering the simulated thermodynamics of the formation of a single fibril, a drastic jump in its width was observed at T_c while remaining approximately constant at other temperatures. This suggests that the width acts as an order parameter for a first-order like transition. In the analytical approach, we also saw the necessity of growth in more than one dimension for the transition to happen, since no critical-like behaviour was observed when considering only one-dimensional growth.

By using random initial conditions, and letting the system equilibrate by single peptide updates and crowder translations, we have also studied the crowding effects on the kinetics of the model. We have demonstrated the effect of the crowdors by first considering a system that is in the single-layer phase in equilibrium, and then successively systems with an increasing volume fraction of crowdors. By adding more crowdors, we effectively increase the peptide density which in turn takes us into the multi-layer phase. We then consider systems with increasing temperatures, until we get back to the single-layer phase. With an elevated temperature, we observe that the formation of large fibrils happens faster, and that the mass fraction of free monomers is increased.

7 Outlook

A situation which has not been systematically studied in this work is the case $R_C = 1$. There is reason to believe that this crowder size yields non-trivial crowding effects, both on thermodynamics and kinetics. Small crowders have previously been reported [5] to have an inhibitory effect on aggregation, opposite of what one would expect just from an effective increase in peptide density. The reason for why this happens is however not clear, but if a similar inhibition would be seen in this model for $R_C = 1$, the underlying mechanism might become clearer due to the simplicity of our model.

As commented in Section 3.2, it is not clear what kind of updates that should be used in the kinetic simulations. Even though monomer addition (taken into account in our simulations by the single peptide updates) is physically important, secondary processes such as fragmentation certainly matter in reality. To be able to get more quantitative information about kinetics, one would ideally like to include corresponding updates. Updating Swendsen-Wang clusters would provide a possibility for fragmentation and in addition allow for translations and rotations of whole aggregates (which of course also happen in reality). To make the updates more physically reasonable, one might want to include that large clusters should move less than small ones, and that the rotations should depend on the moment of inertia about the rotation axis. To avoid ineffective simulations and problems with detailed balance violation, this idea should probably be implemented in how large fraction of an MC Sweep the cluster update is counted as, and not e.g. by additional accept-reject steps.

Regarding crowding effects, a next step could be to add an attraction between the peptides and the crowders (and possibly some interaction between the crowders themselves). This could cluster the peptides along the crowder surface and thus enhancing the fibrillation. However, for a large crowder concentration, the attraction could also deplete the monomers which would instead undermine the aggregation. This is the suggested mechanism behind the observed dual effect on the aggregation kinetics [6], and it would be interesting to see if the same effect could be reproduced with this model.

The analytical approach presented in this work was designed to study the stick model, but could be modified in an attempt to resemble amyloid nucleation in reality. The main difference would be an integration over linear and angular momentum degrees of freedom, and over the continuum of peptide orientations in the one-particle canonical partition functions. In these integrals, the mass and moments of inertia of the aggregates will matter, which in turn depend on the aggregate length and width.

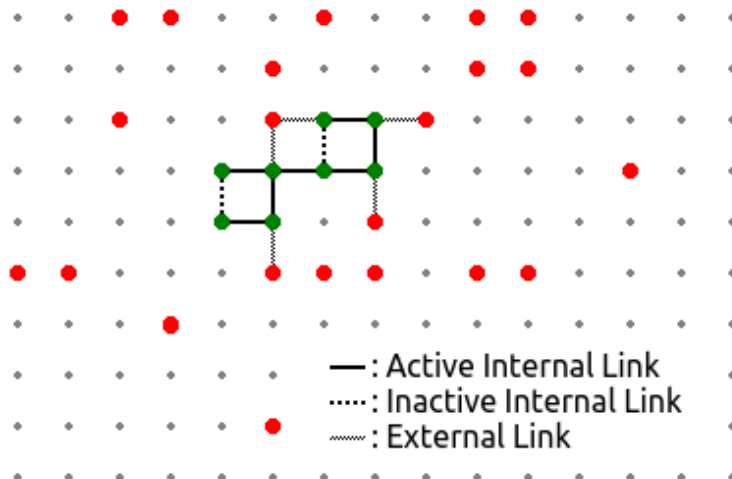


Figure 36: Example of a link configuration in a Swendsen-Wang cluster. The green dots represent peptides that are members of the cluster and the red dots indicate non-members.

8 Appendix A: Detailed balance in Swendsen-Wang cluster updates

In what follows, we show that the Swendsen-Wang cluster updates, together with the additional acceptance criteria in Eq. (3.1.9), satisfies detailed balance with respect to the ensemble $P_\nu \propto 1/g(E_\nu)$ which is used in the thermodynamics simulations. This is done by first showing that the cluster updates by themselves satisfies detailed balance with respect to the canonical ensemble $P_\nu \propto e^{-\beta_{\text{sw}} E_\nu}$. We then show that, by including the additional accept-reject step, the detailed balance condition is instead satisfied in the ensemble $P_\nu \propto 1/g(E_\nu)$.

A cluster is formed by recursively accepting or rejecting bonds that link neighbouring peptides, with the probability P_{add} given in Eq. (3.1.8). We classify the relevant links in three categories:

- Active internal links (accepted bonds between two cluster members).
- Inactive internal links (rejected bonds between two cluster members).
- External links (rejected bonds between a cluster member and a peptide not part of the cluster that will be broken if the cluster move is accepted).

In Figure 36, an example of a cluster is illustrated. Note that the same cluster may be formed by several different internal link configurations, which we denote by ξ , but always with the same external links. Let $\nu \xrightarrow{\mathcal{R}(X)} \nu'$ be the transition from state ν to ν' through a rigid body update \mathcal{R} of cluster X (X is the set of peptides connected by active links). If there are no additional accept-reject steps and if $\mathcal{R}(X)$ causes no steric clashes, the probability for $\nu \xrightarrow{\mathcal{R}(X)} \nu'$ to happen is

$$P\left(\nu \xrightarrow{\mathcal{R}(X)} \nu'\right) = \frac{N_X}{N} \sum_{\xi} P_{\text{Form}}(X; \xi) \cdot P(\mathcal{R}) \quad (8.0.2)$$

where the first factor accounts for the probability to pick any of the N_X cluster members as the initial peptide. The sum goes through all the possible internal link configurations ξ , i.e. the possible ways to form X . $P_{\text{Form}}(X; \xi)$ denotes the probability to reject the external bonds and to accept/reject internal links corresponding to ξ , and $P(\mathcal{R})$ is the probability for choosing the rigid body update \mathcal{R} . Let $\langle ab \rangle$ denote a pair of neighbouring peptides a and b connected with an active link ($a, b \in X$), and similarly let $\langle \bar{a}\bar{b} \rangle$ denote two neighbouring peptides \bar{a} and \bar{b} with an intermediate inactive link ($\bar{a}, \bar{b} \in X$). Also let $\langle \bar{a}\tilde{b} \rangle$ be the pairs of neighbouring peptides with an

intermediate external link ($\tilde{a} \in X$ but $\tilde{b} \notin X$ or vice versa). On writing out $P_{\text{Form}}(X; \xi)$ explicitly we obtain

$$P\left(\nu \xrightarrow{\mathcal{R}(X)} \nu'\right) = \frac{N_X}{N} \sum_{\xi} \prod_{\langle ab \rangle} P_{\text{add}}(a, b) \prod_{\langle \tilde{a}\tilde{b} \rangle} [1 - P_{\text{add}}(\tilde{a}, \tilde{b})] \prod_{\langle \tilde{a}\tilde{b} \rangle} [1 - P_{\text{add}}(\tilde{a}, \tilde{b})] P(\mathcal{R}) \quad (8.0.3)$$

$$= \frac{N_X}{N} P(\mathcal{R}) \prod_{\langle \tilde{a}\tilde{b} \rangle} [1 - P_{\text{add}}(\tilde{a}, \tilde{b})] \sum_{\xi} \prod_{\langle ab \rangle} P_{\text{add}}(a, b) \prod_{\langle \tilde{a}\tilde{b} \rangle} [1 - P_{\text{add}}(\tilde{a}, \tilde{b})], \quad (8.0.4)$$

where the product over external links factor out of the sum over ξ since the external links are independent of the internal link configuration. Similarly, we have for the inverted reaction

$$P\left(\nu' \xrightarrow{-\mathcal{R}(X)} \nu\right) = \frac{N_X}{N} P(-\mathcal{R}) \prod_{\langle \tilde{a}'\tilde{b}' \rangle} [1 - P_{\text{add}}(\tilde{a}', \tilde{b}')] \quad (8.0.5)$$

$$\times \sum_{\xi'} \prod_{\langle a'b' \rangle} P_{\text{add}}(a', b') \prod_{\langle \tilde{a}'\tilde{b}' \rangle} [1 - P_{\text{add}}(\tilde{a}', \tilde{b}')] \quad (8.0.6)$$

$$= \frac{N_X}{N} P(\mathcal{R}) \prod_{\langle \tilde{a}'\tilde{b}' \rangle} [1 - P_{\text{add}}(\tilde{a}', \tilde{b}')] \quad (8.0.7)$$

$$\times \sum_{\xi} \prod_{\langle ab \rangle} P_{\text{add}}(a, b) \prod_{\langle \tilde{a}\tilde{b} \rangle} [1 - P_{\text{add}}(\tilde{a}, \tilde{b})], \quad (8.0.8)$$

where $P(-\mathcal{R}) = P(\mathcal{R})$ since the rigid body updates are drawn symmetrically, and the internal link configurations are the same when forming the cluster X in the states ν and ν' .

Let

$$w\left(\nu \xrightarrow{\mathcal{R}(X)} \nu'\right) = P\left(\nu \xrightarrow{\mathcal{R}(X)} \nu'\right) P_{\nu}, \quad (8.0.9)$$

be the rate of the transition. To check the detailed balance condition, we compute

$$\frac{w\left(\nu \xrightarrow{\mathcal{R}(X)} \nu'\right)}{w\left(\nu' \xrightarrow{-\mathcal{R}(X)} \nu\right)} = \frac{\prod_{\langle \tilde{a}\tilde{b} \rangle} [1 - P_{\text{add}}(\tilde{a}, \tilde{b})] P_{\nu}}{\prod_{\langle \tilde{a}'\tilde{b}' \rangle} [1 - P_{\text{add}}(\tilde{a}', \tilde{b}')] P_{\nu'}}. \quad (8.0.10)$$

In the canonical ensemble, we have $P_{\nu} \propto e^{-\beta_{\text{sw}} E_{\nu}}$, where the system energy E_{ν} can be divided into the external links for the cluster X and everything else:

$$E_{\nu} = \sum_{\langle \tilde{a}\tilde{b} \rangle} \varepsilon_{\tilde{a}\tilde{b}} + E_{\nu}^{(\text{rest})}. \quad (8.0.11)$$

Similarly,

$$E_{\nu'} = \sum_{\langle \tilde{a}'\tilde{b}' \rangle} \varepsilon_{\tilde{a}'\tilde{b}'} + E_{\nu'}^{(\text{rest})} \quad (8.0.12)$$

$$= \sum_{\langle \tilde{a}'\tilde{b}' \rangle} \varepsilon_{\tilde{a}'\tilde{b}'} + E_{\nu}^{(\text{rest})}, \quad (8.0.13)$$

where the last equality holds since change in energy during the transition only comes from the breaking or formation of external links to the cluster. Eq. (3.1.8) gives

$$\prod_{\langle \tilde{a}\tilde{b} \rangle} [1 - P_{\text{add}}(\tilde{a}, \tilde{b})] = e^{\beta_{\text{sw}} \sum_{\langle \tilde{a}\tilde{b} \rangle} \varepsilon_{\tilde{a}\tilde{b}}} \quad (8.0.14)$$

from which we obtain

$$\begin{aligned} \frac{w\left(\nu \xrightarrow{\mathcal{R}(X)} \nu'\right)}{w\left(\nu' \xrightarrow{-\mathcal{R}(X)} \nu\right)} &= \frac{e^{\beta_{\text{sw}} \sum_{\langle \tilde{a}\tilde{b} \rangle} \varepsilon_{\tilde{a}\tilde{b}}} e^{-\beta_{\text{sw}} (\sum_{\langle \tilde{a}\tilde{b} \rangle} \varepsilon_{\tilde{a}\tilde{b}} + E_{\nu}^{(\text{rest})})}}{e^{\beta_{\text{sw}} \sum_{\langle \tilde{a}'\tilde{b}' \rangle} \varepsilon_{\tilde{a}'\tilde{b}'}}} e^{-\beta_{\text{sw}} (\sum_{\langle \tilde{a}'\tilde{b}' \rangle} \varepsilon_{\tilde{a}'\tilde{b}'} + E_{\nu}^{(\text{rest})})}} \\ &= 1 \end{aligned} \quad (8.0.15)$$

i.e. detailed balance is satisfied.

If we instead want to use the ensemble $P_\nu \propto 1/g(E_\nu)$, we can add an additional accept/reject step P_{acc} . Detailed balance then requires

$$P\left(\nu \xrightarrow{\mathcal{R}(X)} \nu'\right) P_{\text{acc}}(\nu \rightarrow \nu') \frac{1}{g(E_\nu)} = P\left(\nu' \xrightarrow{-\mathcal{R}(X)} \nu\right) P_{\text{acc}}(\nu' \rightarrow \nu) \frac{1}{g(E_{\nu'})}. \quad (8.0.16)$$

Using Eq. (8.0.15), we have

$$P\left(\nu' \xrightarrow{-\mathcal{R}(X)} \nu\right) = P\left(\nu \xrightarrow{\mathcal{R}(X)} \nu'\right) e^{\beta_{\text{sw}}(E_{\nu'} - E_\nu)}, \quad (8.0.17)$$

meaning that P_{acc} has to satisfy

$$P_{\text{acc}}(\nu \rightarrow \nu') \frac{1}{g(E_\nu) e^{-\beta_{\text{sw}} E_\nu}} = P_{\text{acc}}(\nu' \rightarrow \nu) \frac{1}{g(E_{\nu'}) e^{-\beta_{\text{sw}} E_{\nu'}}}. \quad (8.0.18)$$

When P_{acc} is given by Eq. (3.1.9), Eq. (8.0.18) indeed holds.

References

- [1] Zhou, H.-X., Rivas, G. & Minton, A. P. Macromolecular crowding and confinement: biochemical, biophysical, and potential physiological consequences. *Annual Review of Biophysics* **37**, 375 (2008).
- [2] Härd, T. Amyloid fibrils: formation, polymorphism, and inhibition. *The Journal of Physical Chemistry Letters* **5**, 607–614 (2014).
- [3] Chiti, F. & Dobson, C. M. Protein misfolding, functional amyloid, and human disease. *Annual Review of Biochemistry* **75**, 333–366 (2006).
- [4] Irback, A., Jónsson, S. Æ., Linnemann, N., Linse, B. & Wallin, S. Aggregate geometry in amyloid fibril nucleation. *Physical Review Letters* **110**, 058101 (2013).
- [5] Co, N. T., Hu, C.-K. & Li, M. S. Dual effect of crowders on fibrillation kinetics of polypeptide chains revealed by lattice models. *The Journal of Chemical Physics* **138**, 185101 (2013).
- [6] Cabaleiro-Lago, C., Quinlan-Pluck, F., Lynch, I., Dawson, A. D. & Linse, S. Dual effect of amino modified polystyrene nanoparticles on amyloid β protein fibrillation. *ACS Chemical Neuroscience* **1**, 279–287 (2010).
- [7] McGuffee, S. R. & Elcock, A. H. Diffusion, crowding & protein stability in a dynamic molecular model of the bacterial cytoplasm. *PLoS Computational Biology* **6**, e1000694 (2010).
- [8] Adamcik, J. *et al.* Understanding amyloid aggregation by statistical analysis of atomic force microscopy images. *Nature Nano* **5**, 423–428 (2010).
- [9] Fitzpatrick, A. W. *et al.* Atomic structure and hierarchical assembly of a cross- β amyloid fibril. *Proceedings of the National Academy of Sciences, USA* **110**, 5468–5473 (2013).
- [10] Eanes, E. D. & Glenner, G. G. X-ray diffraction studies on amyloid filaments. *Journal of Histochemistry & Cytochemistry* **16**, 673–677 (1968).
- [11] Maji, S. K. *et al.* Functional amyloids as natural storage of peptide hormones in pituitary secretory granules. *Science* **325**, 328–332 (2009).
- [12] Volpatti, L. R. & Knowles, T. P. J. Polymer physics inspired approaches for the study of the mechanical properties of amyloid fibrils. *Journal of Polymer Science Part B: Polymer Physics* **52**, 281–292 (2014).
- [13] Nicolai, T., Britten, M. & Schmitt, C. β -lactoglobulin and WPI aggregates: Formation, structure and applications. *Food Hydrocolloids* **25**, 1945–1962 (2011).

- [14] Nelson, R. *et al.* Structure of the cross- β spine of amyloid-like fibrils. *Nature* **435**, 773–778 (2005).
- [15] Usov, I., Adamcik, J. & Mezzenga, R. Polymorphism complexity and handedness inversion in serum albumin amyloid fibrils. *ACS Nano* **7**, 10465–10474 (2013).
- [16] Wang, F. & Landau, D. P. Efficient, multiple-range random walk algorithm to calculate the density of states. *Physical Review Letters* **86**, 2050–2053 (2001).
- [17] Swendsen, R. H. & Wang, J.-S. Nonuniversal critical dynamics in Monte Carlo simulations. *Physical Review Letters* **58**, 86–88 (1987).
- [18] Collins, S. R., Douglass, A., Vale, R. D. & Weissman, J. S. Mechanism of prion propagation: Amyloid growth occurs by monomer addition. *PLoS Biology* **2**, e321 (2004).
- [19] Cohen, S. I. A. *et al.* Proliferation of amyloid- β 42 aggregates occurs through a secondary nucleation mechanism. *Proceedings of the National Academy of Sciences, USA* **110**, 9758–9763 (2013).
- [20] Oosawa, F. & Kasai, M. A theory of linear and helical aggregations of macromolecules. *Journal of Molecular Biology* **4**, 10–21 (1962).
- [21] Zhang, J. & Muthukumar, M. Simulations of nucleation and elongation of amyloid fibrils. *The Journal of Chemical Physics* **130**, 035102 (2009).
- [22] Auer, S. Phase diagram of polypeptide chains. *The Journal of Chemical Physics* **135** (2011).
- [23] Uversky, V. N., Li, J. & Fink, A. L. Evidence for a partially folded intermediate in α -synuclein fibril formation. *Journal of Biological Chemistry* **276**, 10737–10744 (2001).