Universiteit Utrecht

Institute for Theoretical Physics
Faculty of Science
Utrecht University

Pathway Complexity of the Self-Assembly of Linear Viruses and Virus-Like Particles

Bachelor Thesis

Sander C. Kuipers

Physics and Astronomy

Supervisor:

Prof. Dr. Ir. P.P.A.M. van der Schoot
Institute for Theoretical Physics

June 14, 2017

Abstract

The capsid of linear virus-like particles and viruses are created by a self-assembly process in a solution containing the coat proteins and genome. The self-assembly may take place via different, molecular pathways, which are influenced by the associated binding free energies. Measurements of the self-assembly of virus-like particles show an overshoot in the capsid assembly and an increase of the total protein density may result in a smaller percentage of fully encapsulated particles. These observations are possibly due to the assembly of micelles in the solution. In this thesis a model, which is a combination of three known models, is proposed for describing the assembly-pathways. This model considers a cooperative and non-cooperative capsid assembly, the assembly of micelles in the solution, and a nucleation barrier. The equilibrium properties and dynamics of this model show a high dependence on the relation between the critical density of each binding mode, the overall density of the proteins and the stoichiometry. The occurrence of an overshoot in the capsid assembly seen experimentally follows from a slow micelle assembly. However, according to this model, an increase of the total protein concentration can not result in a lower probability of a fully encapsulated particle.

Contents

1 Introduction

The capability of altering a persons DNA in a contaminated part or adding a missing gene that is needed for reproduction can help the lives of many people. This technique, called gene therapy, is therefore a very interesting and promising treatment. This treatment would enable us to prevent or fight certain diseases, such as cancer or HIV infection, by altering the DNA in specific cells of the human body. To change the DNA, a gene is delivered by a carrier, called a vector. The most commonly used vectors in gene therapy are viruses or virus-like particles, because viruses have evolved in the perfect mechanisms for entering a target cell and implementing a gene in its DNA [1]. Although they are one of the simplest living creatures, scientist have great difficulty interpreting all the steps of the assembly and structure of viruses. Scientists have already created artificial viruses with the help of genetic manipulation, however, synthetic virology is still in its infancy and there is a lot of room for improvement and further research.

1.1 Virus structure

Viruses have a wide variety of structures and sizes. A virus always consist of a single- or double-stranded nucleic acid genome. The nucleic acid genome is very vulnerable and should thus stay in a latent state when the virus is outside the target cell. This means they are not well protected against for example a high PH-value or enzymes. A protective layer, called the capsid, which consists of many proteins, protects the nucleic acid genome outside the target cell. The capsid plays an important role in the architecture of the virus. The capsid proteins (CPs) fold the nucleic acid genome in a symmetrical structure. The basic types of the structure of the virus are the helical and icosahedral structure [1–3]. This thesis focusses on one of the best studied virus and AVs similar to this virus: the tobacco mosaic virus (TMV). TMV is a linear helical virus with single strand nucleic acid genome. TMV has a rigid, rod-like structure which is 300 nm long, with a diameter of 18 nm and a 4 nm hole in the middle. The helix of a helical virus is defined by its pitch (P), which is the distance covered by each turn of the helix. The helix of TMV has a pitch of 23Å[4].
The infectious cycle (or the viral life cycle) differs for each type of virus, however a few general steps can be

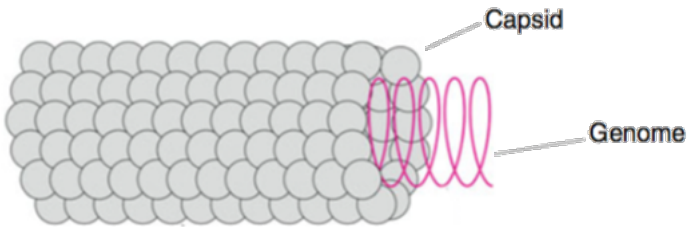


Figure 1.1: Schematic illustration of a virus with a helical symmetry. An example of such a virus is TMV. Taken and modified from [3].

recognized. Each viral life cycle consists of 1) recognition and entry of the cell, 2) viral genome uncoating, 3) viral gene expression and replication, 4) morphogenesis, and 5) virion release out of the cell. In this thesis we focus on the morphogenesis of the virus and in particular the assembly and disassembly of the CPs. By using *in vitro* assembly systems combined with theoretical research, it has been possible to find the fundamental principles of virus capsid self-assembly [1]. The next section gives a more detailed description of some aspects of the viral self-assembly.

1.2 Capsid Self-Assembly

Virion morphogenesis is a complicated process and for most viruses it is poorly understood, however, TMV is one of the best researched viruses and its structure is well known. The stages of the virion morphogenesis are the capsid assembly, nucleic acid packaging and the virus particle maturation. In this thesis we focus on the capsid assembly. There are three general ways for the capsid assembly. The capsid of TMV is built

by self-assembly, which only requires CPs subunits. These combine themselves under the right conditions to capsid building blocks without the help of a scaffolding protein or the viral nucleic acid. After the capsid building blocks are formed by CPs, stable capsid building blocks can be seen as the starting substance for producing the capsid (and in our model we consider them as stable) [1].
The capsid assembly of TMV is a hierarchical self-assembly. This means that every next step in the self-assembly process costs less energy. A 20S coat protein is required for the nucleation of the capsid of TMV. The precise structure of this protein is still uncertain. It is possibly shaped as a small helix or a disk. The nucleation of the 20S coat protein requires a conformational switch to bind at the origin-of-assembly sequence (OAS). Consecutively, bidirectional assembly occurs from the OAS. One end is probably closed with single CPs, while the other may be using discs, small helices, or single CPs [5, 6].

1.3 Research Background

In earlier research scientists have done experiments on the self-assembly of TMV-like artificial viruses. The particles consisted of double stranded DNA and had a rod-shaped structure. The capsid building blocks were biosynthesized single capsid proteins. As a consequence of the results scientists proposed a simplified model for the self-assembly of rod-shaped AVs and TMV-like particles. This model is an adjustment of a model, which was initially designed for TMV, the kinetic zipper model [6, 7]. Note that there are some differences between the artificial virus in these experiments and TMV. TMV has single stranded RNA and its capsid building blocks consists of multiple capsid proteins. For this reason the precise self-assembly steps differ on many points. However, this model only considers some basic processes used in the self-assembly of many viruses. For our model it is not important to know the precise steps of the self-assembly of each different virus. All the deviations are described by different binding free energies.
The zipper model gives us insight how the self-assembly of virus particles behaves. It describes why viruses prefer to be either fully encapsulated or to have no capsid at all. The model considers the nucleic acid genome to be an one dimensional rod-particle with a number of binding sites. It considers an allosteric site, which allows a protein to bind, resulting in a conformational change. After the nucleation more capsid building blocks can bind to the RNA adjacent to each other. In this model it is not allowed for capsid building blocks to bind to random binding sites, only adjacent each other or at one end of the rod [6].
M. Punter *et al.* has continued the work on the kinetic zipper model and has answered some question that still arose with the model [8, 9]. An assembly signal has been added to the model and it has been shown how the assembly behaves with a finite supply of capsid proteins. But there were still a few questions about and issues with the new model. Experiments have shown that, besides to the zipper-type of binding, the capsid building blocks can also randomly bind to the RNA template. The capsid building blocks possibly first bind randomly to the nucleic acid genome following the rules of the Langmuir adsorption model [10]. These bound capsid building blocks can freely move along the rod. Then the randomly (Langmuir) bound capsid building blocks may bind according to the zipper model as shown in the schematic illustration in Figure 1.2 [7, 11]. We expect these random binding to compete with the zipper-type binding and to have a negative effect on the percentage of fully encapsulated viruses in equilibrium. So, one of the questions that has come up is: **How does the Langmuir binding affect the capsid of the artificial viruses in equilibrium**.
Also combining Langmuir with the zipper model changes the differential equations found by M. Punter. It is not possible for capsid building blocks to bind directly as an zipper protein, but instead should first bind through the Langmuir binding. This gives us the next research question: **What is the influence of the non-cooperative (Langmuir) binding to the rate of the binding process?** Also one of the things that scientists discovered in the experiments was that the capsid building blocks would form micelles when the solution had a high concentration of capsid building blocks, possibly due to their hydrophilic and -phobic part. A schematic illustration of this is shown in Figure 1.2. They showed that increasing the total protein density may result in a lower fraction of fully encapsulated viruses [11]. Having a high concentration of capsid building blocks in the solution may result in a negative outcome of the encapsulation and this is possibly because the micelles would have a more preferable assembly free energy when there is a high density of capsid building blocks. So the last question in this thesis is: **What is the influence of the assembly of micelles in the solution on the assembly of the capsids?**
In the next section we outline when these question are addressed in this thesis.

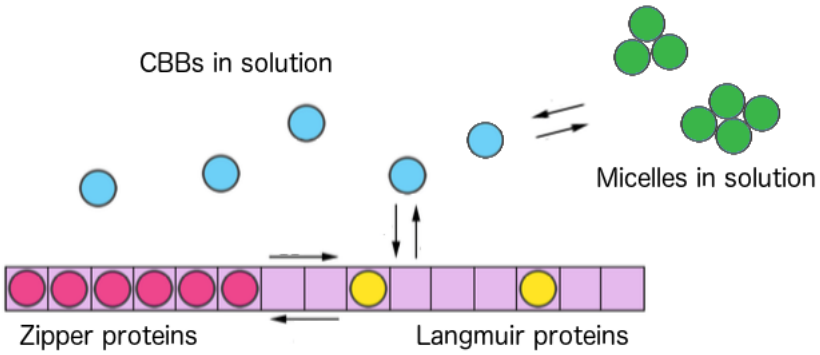


Figure 1.2: Schematic illustration of the reaction pathways of the capsid- and micelle assembly. It shows the four states of which a capsid building block (CBB) could be in: 1) Free in the solution, 2) part of a micelle, 3) Non-cooperatively and randomly adsorbed on a RNA template or 4) cooperatively adsorbed on a RNA template. Taken and modified from [11].

1.4 Outline

In Chapter 2 we discuss the models under the conditions of thermodynamic equilibrium. First, we introduce the free energy of the system. By minimizing the free energy for the template and free protein density, we obtain an expression for these two densities. Using the semi-grand partition function we also derive an expression for the fraction of occupied sites on the templates. With these expression we first introduce the Langmuir adsorption model and discuss its characteristics. Next we describe the aspects of the zipper model. After the specific characterizations of, and differences between the two models are discussed, we combine the two models and focus on the competition between the two assembly models. We show when a transition between the two assembly methods happens by presenting a phase diagram and calculating the heat capacity. Subsequently, we introduce the Debye model for micelle assembly and try to understand its behaviour under the conditions of thermodynamic equilibrium [12]. This is done without, and afterwards with the occurrence of capsid assembly. We focus on how changing the binding free energies of the three binding methods, changes the fraction of occupied sites.
In Chapter 3, we derive the dynamical equations for the combination of the two capsid assembly methods and the micelle assembly. We do not study the dynamics of the models separately. The comparison between the differential equations in equilibrium and the values obtained by the statistical evaluation in Chapter 2 are used to simplify the differential equations. We numerically analyse the differential equations and this shows how the assembly signal delays the assembly and the different ways the micelles can effect the capsid assembly.
Finally in chapter 4 we summarize the results, formulate a conclusion and make some recommendations for future research.

2 Statics

In this Chapter we consider our model to be under the conditions of thermodynamic equilibrium and all the quantities we use here are under the conditions of thermodynamic equilibrium. First we describe the model and assumptions we use in this thesis. In the next sections we describe the random adsorption of proteins on the templates with the Langmuir adsorption model. Then we introduce the zipper model for the cooperative binding as proposed by Kraft *et al.* [6] and we study the competition between the Langmuir- and the zipper-types of adsorption modes. In the last section of this chapter we introduce the Debye model for micelle assembly and study its characteristics.
Our model concerns a solution with a volume V and a temperature T. RNA templates and capsid building blocks are dissolved in this solution. In the rest of the thesis we call the capsid building blocks proteins for convenience. The number density of proteins is ρ_p and the number density of RNA templates ρ_t. The viruses are linear viruses, thus we consider them to be one dimensional in our model. Each RNA template has q binding sites, to which a protein can bind. A fully encapsulated RNA template has q proteins bound. $\rho_t(n)$ is the number density of dissolved RNA templates in the solution with n adsorbed proteins ($0 \leq n \leq q$). The total number of proteins in the solution is conserved, consequently, the number density of free proteins in the solution ρ_{fp} is given by ρ_{fp}=ρ_p-$\sum_{n=0}^{q} n\rho_t(n)$. Because the volume and temperature are constant, we can write the dimensionless Helmholtz free energy for this configuration

$$\frac{\beta F}{V} = \sum_{n=0}^{q} \rho_t(n)\Big(\ln(\rho_t(n)\upsilon_{t(n)}) - 1 - \beta E_{int}(n)\Big) + \rho_{fp}\ln(\rho_{fp}\upsilon_{fp}), \tag{2.1}$$

Here, $\beta = \frac{1}{k_B T}$, with k_B the Boltzmann constant and T the temperature, υ_{fp} is the typical volume scale of a free protein and $\upsilon_{t(n)}$ the typical volume scale of a template with n bound proteins. The typical volume scale depends on multiple variables like the density and the radius of the solvent and the potential used to describe the internal partition function [13, 14]. In this thesis we are not interested in what this quantity should be and we take $\rho_t(n)$ and ρ_{fp} to be $\rho_t(n)\upsilon_{t(n)} \to \rho_t(n)$ and $\rho_{fp}\upsilon_{fp} \to \rho_{fp}$. $E_{int}(n)$ is the internal energy of a template with n occupied sites on the template.
In equilibrium the Helmholtz free energy is minimal for the two densities. However, we should consider that we have a fixed number of proteins and templates in the solution. To enforce these restrictions, we use two Lagrange multipliers $\lambda_1(\sum_{n=0}^{q}\rho_t(n) - \rho_t)$ and $\lambda_2(\rho_{fp} + \sum_{n=0}^{q} n\rho_t(n) - \rho_p)$. The Lagrange multipliers are the $\beta\mu'_{fp}d\rho_{fp}$ and $\beta\mu'_{fp}d\rho_{fp}$ terms of the change of the Helmholtz free energy, with μ'_{fp} the chemical potential of the proteins and μ'_t the chemical potential of the templates, and $d\rho_{fp}$ and $d\rho_t$ the change in the the free protein- and template number density, which is zero in equilibrium. For this reason $\lambda_1 = \beta\mu'_t$ and $\lambda_2 = \beta\mu'_{fp}$. For convenience we define $\beta\mu'_{fp} \to \mu_{fp}$ and $\beta\mu'_t \to \mu_t$. By minimising the Helmholtz free energy, we can derive an expression for the density of the templates with n absorbed proteins in equilibrium

$$\rho_{t,eq}(n) = \exp(-\beta E_{int}(n) + n\mu_{fp} + \mu_t), \tag{2.2}$$

and an expression for the density of free proteins in the solution in equilibrium

$$\rho_{fp,eq} = \exp(\mu_{fp}). \tag{2.3}$$

We are interested in the number of occupied binding sites on the templates. We therefore define the probability of the template having n occupied binding sites in equilibrium ($0 \leq n \leq q$)

$$P_{eq}(n) = \frac{\rho_t(n)}{\sum_{n=0}^{q}\rho_t(n)}, \tag{2.4}$$

and we define the fraction of occupied binding sites of the templates $\langle\theta\rangle_{eq}$, i.e.

$$\langle\theta\rangle_{eq} = \sum_{n=0}^{q}\frac{n}{q}P_{eq}(n) = \frac{1}{q}\frac{\sum_{n=0}^{q} n\exp(-E_{int}(n) + n\mu_{fp} + \mu_t)}{\sum_{n=0}^{q}\exp(-E_{int}(n) + n\mu_{fp} + \mu_t)}. \tag{2.5}$$

Due to this function we introduce the semi-grand partition function. In a semi-grand ensemble the number of particles is fixed, but the composition can change. For this reason the semi-grand partition function is a

convenient way to describe this mixture. This function is given by

$$\Xi = \sum_{n=0}^{q} \exp(-E_{int}(n) + n\mu_{fp} + \mu_t), \tag{2.6}$$

with $E_{int}(n)$ the dimensionless internal energy and μ_{fp} and μ_t the dimensionless chemical potentials of the free proteins and the templates. Consequently, the fraction of occupied binding sites $\langle\theta\rangle_{eq}$ can be written as a function of the semi-grand partition function

$$\langle\theta\rangle_{eq} = \frac{1}{q}\frac{\partial \ln \Xi}{\partial \mu_{fp}}. \tag{2.7}$$

Thus by acquiring the semi-grand partition of the binding model as a function of μ_{fp} and the internal energy, the fraction of occupied sites and the probability of a template being fully encapsulated can be derived. In the next section we use this method to describe these two functions for the Langmuir model.

2.1 The Langmuir Model

The Langmuir model is a fairly simple adsorption model. It considers all binding sites of the templates to liberate the same amount of free energy if a protein is absorbed. All binding sites are equivalent, and there is no reaction between bound proteins. Because the number of configurations differs per number of proteins bound, the entropy has an important role in this model,
From Equation 2.5 it follows that knowing the internal energy is a crucial step for determining the fraction of occupied sites on the templates. We assume the system to gain a free energy $\delta < 0$ for every filled binding site. This adsorption energy is negative, because there is an attractive interaction between the proteins and the templates. This can be a result or combination of multiple kinds of attractive interaction, such as a van der Waals- or electrostatic attraction. The internal energy of the model is thus defined as

$$E_{int,L}(m) = m\delta, \tag{2.8}$$

with m the number of proteins bound on the template. The system has the lowest internal energy when all binding sites are occupied, however, this macrostate has only one configuration. The macrostate of a template with m of the sites occupied by a protein has $\binom{q}{m}$ ways to distribute the proteins over the binding sites, with q the number of binding sites on a template. For this reason a template has only a small probability of having q occupied binding sites for small protein densities and free energy per adsorbed protein. The semi-grand partition function of this system is given by

$$\Xi_L = \sum_{m=0}^{q} \binom{q}{m} \exp(-m\delta + \mu_{fp}\delta + \mu_t) = (1 + \exp(\mu_{fp} - \delta))^q \tag{2.9}$$

We define the critical Langmuir density $\phi_L = \exp(\delta)$ and a parameter that compares the number density of proteins in equilibrium with the critical Langmuir density $a_{eq} = \frac{\rho_{fp}}{\phi_L}$. Using the equations 2.6 and 2.7 the fraction of occupied sites of the templates can be written as

$$\langle\theta_L\rangle = \frac{\exp(\mu_{fp} - \delta)}{1 + \exp(\mu_{fp} - \delta)} = \frac{a_{eq}}{1 + a_{eq}}. \tag{2.10}$$

Note that with this model the fraction of occupied sites has no dependence on the number of binding sites q. We have calculated the fraction of occupied sites as a function of a_{eq}. We are however more interested in the fraction of occupied sites as a function of the total protein number density ρ_p. The total protein number density is conserved, consequently, this is given by $\rho_p = \rho_{fp} + \sum_{n=0}^{q} n\rho_t(n)$. We define two new parameters, $A = \frac{\rho_p}{\phi_L}$ and $L_L = \frac{q\rho_t}{\phi_L}$. The parameter A scales the total protein density with the critical Langmuir density and L_L the total number of binding sites with the critical Langmuir density. The mass conservation can be rewritten as

$$A = a_{eq} + L_L\langle\theta_L\rangle_{eq}. \tag{2.11}$$

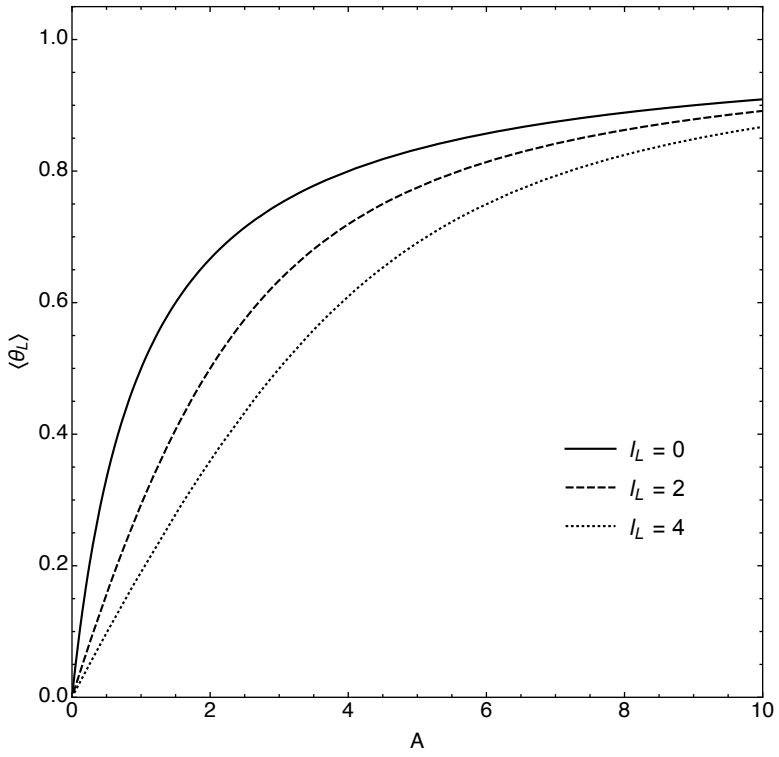


Figure 2.1: The fraction of occupied sites of the templates $\langle\theta_L\rangle_{eq}$ according to the Langmuir model as a function of the total protein number density scaled with the critical Langmuir density A for different values of the total number of binding sites scaled with the critical density L_L= 0, 2 or 4. For a low free protein concentration some sites are occupied. To obtain a high fraction of occupied the sites the total protein number density has to be very large.

Both A and $\langle\theta_L\rangle_{eq}$ depend on a_{eq}. Using this quantity, we are able to plot $\langle\theta_L\rangle_{eq}$ as a function of the total protein density scaled by the critical Langmuir density A.

In Figure 2.1 the fraction of occupied sites is shown as a function of the total protein density scaled by the critical Langmuir density A. This is done for three different values of L_L. It shows clearly that the random adsorption occurs for any $\rho_p > 0$, so even with a low concentration of proteins in the solution. However, to get a high fraction of occupied sites, a high concentration of proteins is needed in the solution. As mentioned earlier this is due to the fact that the number of configurations is given by $\binom{q}{m}$, so a very high occupation has only a few configurations.

Almost fully encapsulated viruses are still very latent outside the target cell. So, the RNA is not well protected and is easily disassembled. For this reason we are also interested in the percentage of templates with all binding sites filled. We calculate the probability of having q binding sites filled in equilibrium. This function is given by

$$P_{eq}(q) = \frac{\rho_t(q)}{\sum_{m=0}^{q} \rho_t(m)} = \frac{a_{eq}^q}{\Xi}, \tag{2.12}$$

with Ξ the semi-grand partition function of the model. In Figure 2.2 the probability of a template being fully encapsulated $P_{eq}(q)$ is shown as a function of A. From this figure follows that if the templates are small a high probability of fully encapsulated templates is obtained with a low concentration of proteins. However, when there is a large number of binding sites per template, the protein concentration has to be large to obtain a high percentage of fully encapsulated templates. When a large number of fully encapsulated templates is gained, many proteins stay unused in the solution. For this reason viruses also use another binding method, a cooperative binding method. In the next section we focus on this binding model.

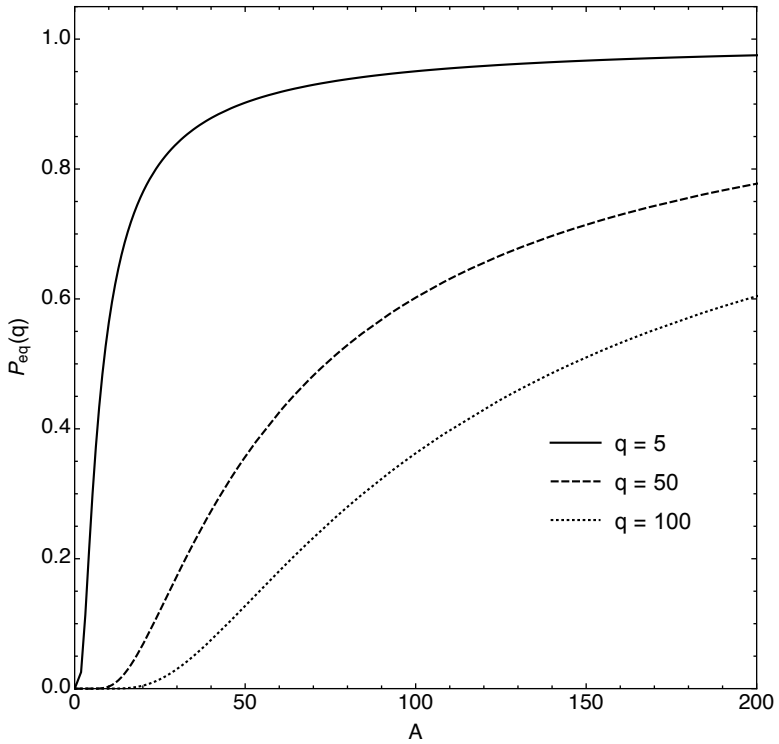


Figure 2.2: The probability of having a template being fully encapsulated in equilibrium $P_{eq}(q)$ according to the Langmuir model, with the total number of binding sites scaled with the critical Langmuir density $L_L = 2$, for three values of the total number of binding sites q. This is shown as a function of the total protein density scaled with the critical Langmuir density A. For large L_L the probability of a template being fully encapsulated is very small.

2.2 The Zipper Model

The zipper model has a very unique and complicated energy model, so we are going to explain this model in a few steps.

First we assume that all binding sites are identical and that when a protein binds to a binding site, the system gains a free energy $h' \gg 0$. This is an energy barrier that makes the nucleation difficult. The system also gains a free energy $g < 0$, because there is still some attraction between the protein and the template; similar to the Langmuir model. Again a free energy $h' + g$ is gained, if another protein is adsorbed. Except when it binds on a site adjacent to the first protein. If a protein binds adjacently to another bound protein, a free energy $g + \epsilon$ is gained, with $\epsilon < 0$. The free energy ϵ is gained due to of the protein-protein interaction. This is a cooperative binding method and this process can be repeated for all the subsequent proteins that bind. Due to this favourable energy, one would expect a sequential binding to happen. An illustration of this binding model is given in figure 2.3 and 2.4.

The next assumption that should be made to describe the zipper model is assuming that the binding site at one end of the template differs from the rest of the binding sites. This is the origin-of-assembly binding site on the template from where the binding starts. The rest of the sites remains identical to each other. If a protein binds at the assembly signal, again a free energy g is gained, with $g < 0$. Next to that, an energy h is gained, with $h' \gg h > 0$.

The last step is to make sure that the sequential binding in this model starts from one side of the template. Thus the probability of random adsorption is so small, it can be neglected. We can neglect this, provided the difference between the two energy barriers h and h' is sufficient large. The probability of the first protein to bind to the assembly signal is given by $P = \frac{\exp(-h)}{(q-1)\exp(-h')+\exp(-h)}$. Consequently, from this formula follows that the entropic effect of binding to a random site can be neglected if $h' - h \gg \ln(q-1)$, and it can be assumed that the first protein always binds at the assembly signal [6].

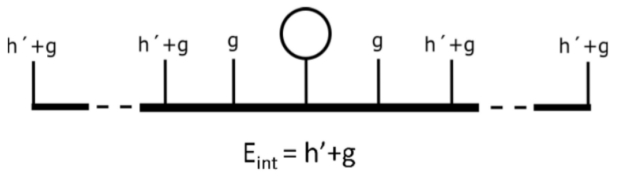


Figure 2.3: Schematic illustration of a template with one protein bound according to the zipper model. The internal energy of the system is $h' + g$ due to the adsorption of one protein. The next protein could bind cooperatively and give an energy $g + \epsilon$ or could bind non-cooperative and give $h' + g$. Taken from [8]

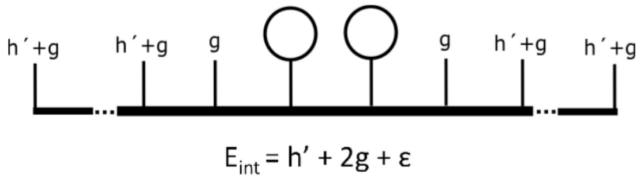


Figure 2.4: Schematic illustration of a template with two proteins bound adjacently. The system gained a free energy $h' + g$, because of the binding of the first protein, g for the binding of the second protein and ϵ for the cooperative interaction between the proteins. The third protein can also bind adjacently to a protein, then again a free energy $g + \epsilon$ is gained, or it can bind non-adjacently and a free energy $h' + g$ is gained. Taken from [8]

In this limit the (dimensionless) internal energy for this model is defined as

$$E_{int,Z}(n) = \begin{cases} 0, & \text{for} \quad n = 0, \\ n(g+\epsilon) + h - \epsilon, & \text{for} \quad 1 \leq n \leq q. \end{cases} \tag{2.13}$$

With this internal energy the semi-grand partition function can be defined in terms of the (dimensionless) chemical potential of the free proteins μ_{fp}. Note that the zipper model rules out the entropy, unlike the Langmuir model. For every n number of proteins bound to a template, there is only one configuration. We can write the semi-grand partition function Ξ as

$$\Xi = 1 + \exp(\mu_{fp} - g - h) \sum_{n=0}^{q-1} \exp\left(n(\mu_{fp} - g - \epsilon)\right) = 1 + \sigma \sum_{n=1}^{q} s_{eq}^n, \tag{2.14}$$

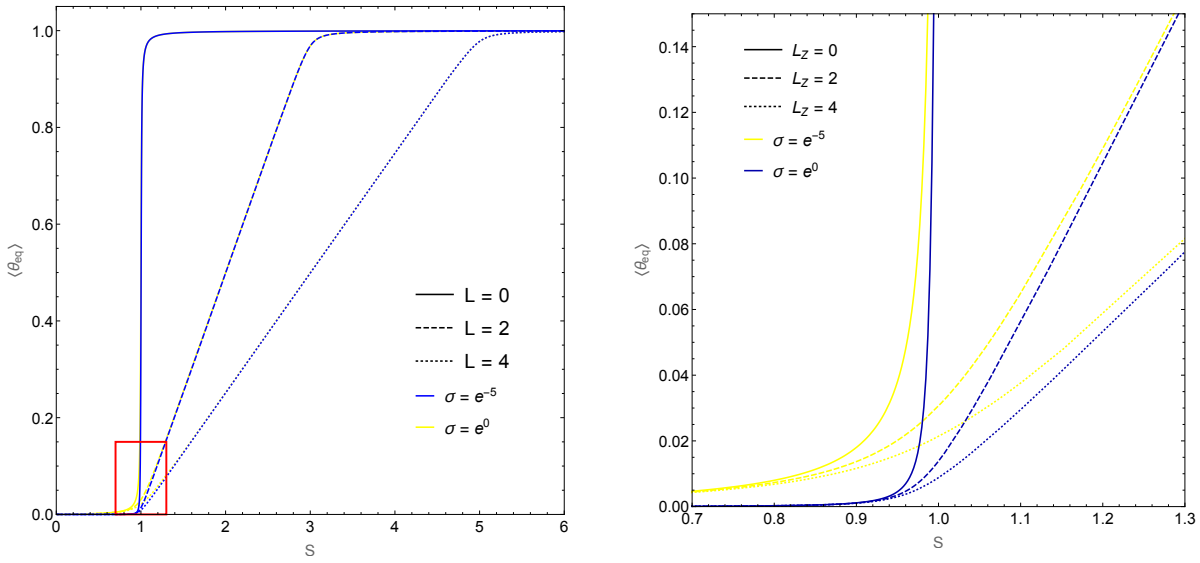
with $s_{eq} = \mathrm{e}^{\mu_{fp} - g - \epsilon} = \frac{\rho_{fp}}{\phi_Z}$ and $\sigma = \mathrm{e}^{\epsilon - h}$. The variable σ is a measure for the energy barrier for the nucleation and the variable s_{eq} scales the free protein number density in equilibrium ρ_{fp} with the critical zipper density ϕ_Z. Now combining the results from equations 2.7 and 2.14, the fraction of occupied sites of the templates can be written as

$$\langle\theta\rangle_{eq} = \frac{\sigma s_{eq}}{q(1 - s_{eq})} \frac{1 - (q+1)s_{eq}^q + q s_{eq}^{q+1}}{1 - s_{eq} + \sigma s_{eq}(1 - s_{eq}^q)}. \tag{2.15}$$

From analytical evaluation follows that a high $s_{eq} \gg 1$ results in high fraction of occupied sites $\langle\theta\rangle_{eq} \to 1$ and a small $s_{eq} \ll 1$ in a low fraction $\langle\theta\rangle_{eq} \to 0$ Because of mass conservation it is required that $\rho_p = \rho_{fp} + \sum_{n=0}^{q} n\rho_t(n)$. This can also be written as

$$S = s_{eq} + L_Z \langle\theta\rangle_{eq}, \tag{2.16}$$

with $S = \frac{\rho_p}{\phi_Z}$, $s_{eq} = \frac{\rho_{fp}}{\phi_Z}$ and $L_Z = \frac{\rho_t q}{\phi_Z}$. The variable S scales the total protein number density with the critical zipper density and L_Z scales the total number of binding sites with the critical zipper density.



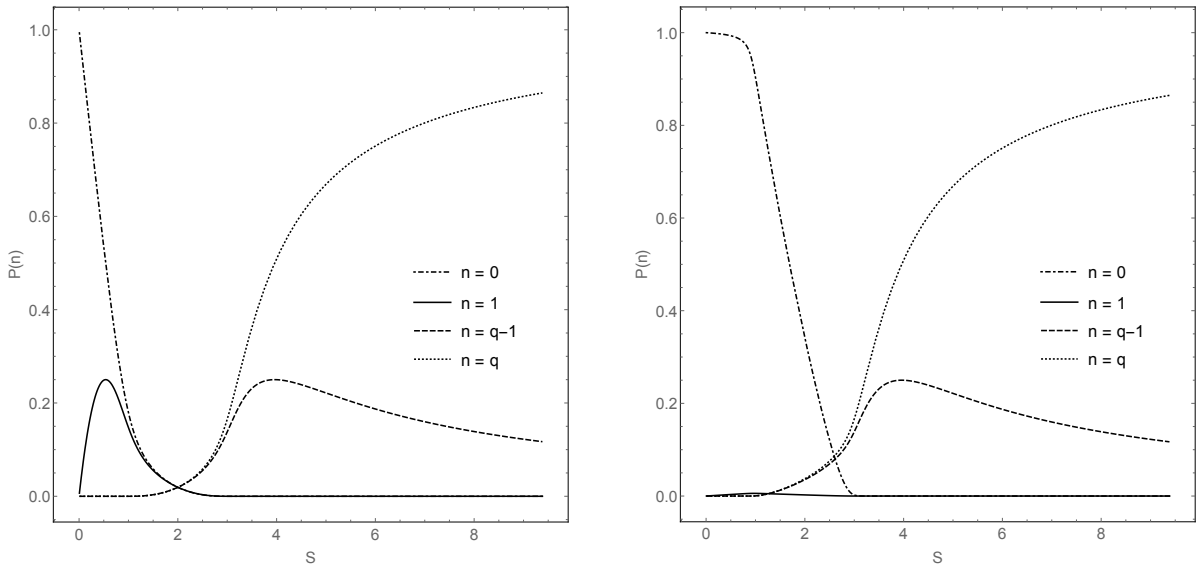
(a) (b)

Figure 2.5: The fraction of occupied sites according to the zipper model $\langle\theta_Z\rangle_{eq}$ as a function of the total protein density scaled with critical zipper density S for different values of $L = \frac{q\rho_t}{\phi_Z}$ and the energy barrier σ. Figure 2.5b is zoomed into the red box of figure 2.5a to get a better view of the difference that σ makes. σ only has an influence if $\phi_Z \approx \rho_p$, because this is the density for which the nucleation of the templates occurs.

In Figure 2.5 the fraction of occupied binding sites $\langle\theta\rangle_{eq}$ is shown as a function of S for different values for the energy barrier σ and the number of binding sites L_Z. We observe that the assembly starts when the total protein number density ϕ_p is approximately the critical zipper density ϕ_Z, so when $S \approx 1$. The slope of the function from $S \approx 1$ until the template is almost fully encapsulated is equal to $\frac{1}{L}$. From this follows that from then on every extra protein that is put in the solution binds to a template until all the templates are almost all fully encapsulated. This is due to the fact that for $s_{eq} > 1$ the energy that the system gains when a protein binds is lower than the chemical potential.
The parameter σ changes the energy cost of the nucleation. For $\sigma = 1$ there is no energy barrier for the nucleation. The energy gained from the first binding is the same as for all the other bindings. For small $\sigma \ll 1$ nucleation is hard and only occurs when $S \approx 1$. The variable σ does not affect the binding any further after the first protein is adsorbed on the template. For this reason changing σ has a small influence in equilibrium, only around $S \approx 1$. However, the nucleation barrier has a great influence on the dynamics of this system, which will be clarified in Chapter 3.
If there is no nucleation barrier, the nucleation already occurs for $S < 1$. What is not visible in Figure 2.5 is whether there are already some templates that are fully encapsulated for $\sigma = 1$ and $S < 1$. To study this we describe the probability of finding a template having n adsorbed proteins for the zipper model: $P_{eq}(n) = \frac{\rho_t(n)}{\sum_{n=0}^{q}\rho_t(n)}$. Filling in the internal energy for this system gives $P_{eq}(n)$ for this model, i.e.

$$P_{eq}(n) = \begin{cases} \frac{1}{\Xi}, & \text{for,} \quad n = 0 \\ \frac{\sigma s_{eq}^n}{\Xi}, & \text{for,} \quad 1 \leq n \leq q \end{cases} \tag{2.17}$$

In Figure 2.6 the probability of a template having n occupied binding sites in equilibrium $P_{eq}(n)$ is shown for different values of n and the energy barrier σ. The energy barrier σ has only a small effect for $S > 1$ on $P(n)$ for any n and on the probability of a template being fully encapsulated $P(q)$. For $S < 1$ the probability of a template being fully encapsulated $P(q)$ is very small for any σ. However, for small n the energy barrier σ has an important role for $P(n)$. With a nucleation barrier, like in figure 2.6b, almost no proteins bind, $P(0) \approx 1$ for values of the total protein density below the critical zipper density, $S < 1$, but for $\sigma = \exp(-5)$ the probability of a template having no proteins bound, $P(0)$, rapidly decreases if the concentration of proteins increases.



(a) (b)

Figure 2.6: The probability of a template having n proteins bound for 0, 1, q-1 and q proteins bound. With q the total number of binding sites per template, $q = 51$, and the total number of binding sites scaled with the critical zipper density $L = 2$. In figure 2.6a there is no nucleation energy barrier: $\sigma = 1$. In figure 2.6b the nucleation barrier is given by $\sigma = \exp(-5)$. For a high nucleation barrier, almost no proteins bind to a template when the total protein density is smaller than the critical zipper density. For large S the nucleation has no influence on the chance distribution.

2.3 The Langmuir- and Zipper Model Combined

In this section we combine the two adsorption models that are already outlined in the previous two sections. We consider the proteins to first bind randomly and non-cooperatively (Langmuir model) on a template. These randomly adsorbed proteins can move freely on the template. They can then start cooperative (zipper) binding at one end of the template. This is also what we try to accomplish in our model. The pathways of the binding however, are not important in this Chapter, because here we study only the the equilibrium properties.

We begin by having a closer look at the competition between Langmuir and zipper binding. We again define the number of zipper proteins bound as n, the number of Langmuir proteins bound as m and the total number of binding sites on the template by q (see Figure 2.7).

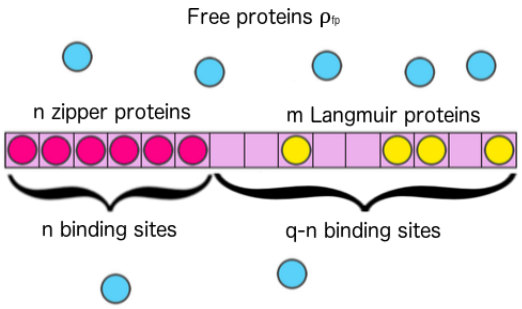


Figure 2.7: Schematic illustration of the zipper model combined with the Langmuir model.

Entropy has an important role in this model, because the number of configurations differs per number of cooperatively- and non-cooperatively bound proteins. The number of configurations is given by $\binom{q-n}{m}$, with n the number of cooperatively bound proteins bound and m of non-cooperatively bound proteins. Fewer cooperative proteins results thus in more configurations. As a consequence this has a negative effect on the cooperatively binding.
Again we describe the semi-grand partition function. This function can be written as a combination of the semi-grand partition function of the both models

$$\Xi = \sum_{n=0}^{q} \sum_{m=0}^{q-n} \binom{q-n}{m} \exp\Big(- E_{int,Z}(n) - E_{int,L}(m) + (n+m)\mu_{fp} \Big), \tag{2.18}$$

$$= \Big(1 + \exp(\mu_{fp} - \delta)\Big)^q + \exp\Big(- q(g + \epsilon - \mu_{fp}) - h + \epsilon \Big) \frac{1 - \Big(\exp(g + \epsilon - \mu_{fp}) \Big(1 + \exp(\mu_{fp} - \delta)\Big) \Big)^q}{1 - \exp(g + \epsilon - \mu_{fp}) \Big(1 + \exp(\mu_{fp} - \delta)\Big)}. \tag{2.19}$$

Here, $E_{int,Z}(n)$ and $E_{int,L}(m)$ denote the internal energy of the zipper- and Langmuir model with n cooperative- and m non-cooperative proteins bound. With this partition function, we are able to find the mean occupation number per template. For $\langle m \rangle_{eq}$ we need to take the derivative of $\ln \Xi$ with respect to δ and for $\langle n \rangle_{eq}$ with respect to g. We can check this by verifying whether $\frac{\partial \ln \Xi}{\partial \mu_{fp}} = -\frac{\partial \ln \Xi}{\partial \delta} - \frac{\partial \ln \Xi}{\partial g}$ is true. This relation can also be derived by analysing the semi-grand partition function.
Solving these equations gives a complicated formula. In Appendix A the derivation of this formula is done. The formula depends on the same parameters as those of the fraction of occupied sites of the zipper model, which are s_{eq}, q and σ. There is one extra parameter, namely $\phi_c = \exp(g + \epsilon - \delta) = \frac{\phi_Z}{\phi_L}$ with ϕ_L the critical Langmuir density and ϕ_Z the critical zipper density. This parameter denotes the energy difference between a cooperatively and non-cooperatively bound protein and it has an important role in determining which of the two binding methods is preferred.

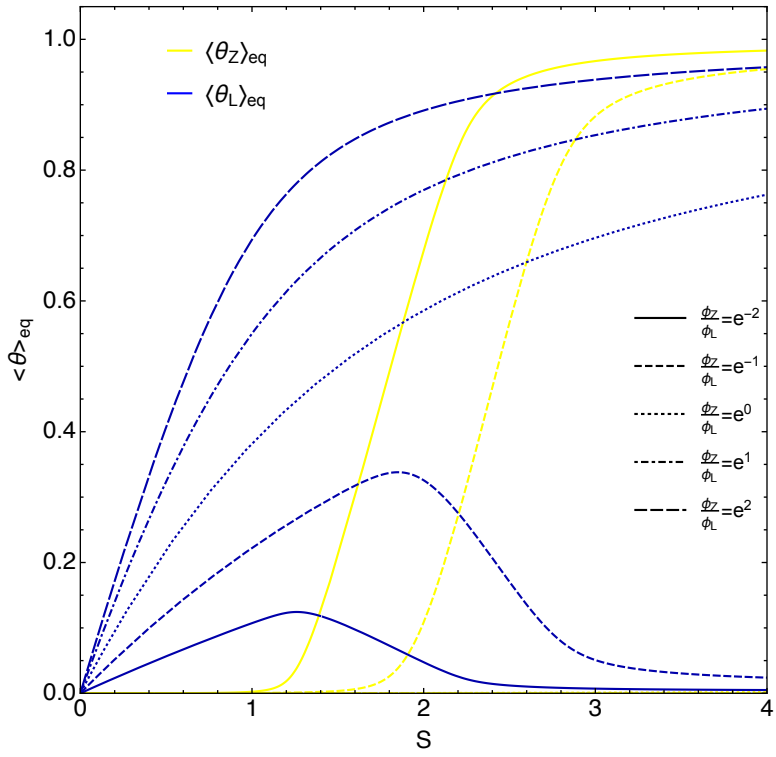


Figure 2.8: The fraction of occupied sites by cooperatively bound proteins $\langle\theta_Z\rangle_{eq}$ (yellow) and non-cooperatively bound proteins $\langle\theta_L\rangle_{eq}$ (blue) as a function of S for different values of $\phi_c = \frac{\phi_Z}{\phi_L}$. The energy barrier is given by $\sigma = \exp(-5)$ and the total binding sites scaled with the critical zipper density is $L_Z = 1$. ϕ_c has an important role in which binding type is preferred. When the critical zipper density is larger than the critical Langmuir density, $\phi_c \geq 1$, no cooperative binding occurs.

Figure 2.8 shows that Langmuir proteins already start to bind for small protein number densities, $S < 1$, and, depending on ϕ_c, the cooperative capsid assembly can start for $S \geq 1$. If the critical Langmuir density is smaller than or equal to the critical zipper density $\phi_c \geq 1$ we recognise the same kind of binding as in figure 2.1. So if the Langmuir energy is smaller than the zipper energy, the zipper model has no contribution and only non-cooperative binding occurs. Note that for $\phi_c = 1$, Langmuir-type of binding is also preferred. This is due to an entropic effect, namely the non-cooperative binding generate more configurations than the cooperative binding.
If the cooperative binding free energy is more favourable than the free energy gained by non-cooperatively binding, the zipper proteins start to "push" the Langmuir proteins off the template above some values of S. Note that, similar to the binding according to only the zipper model in Chapter 2.2, if the template is nucleated the slope of the fraction of occupied sites of the cooperatively bound proteins is $\frac{1}{L}$, so after the first zipper protein is absorbed, almost every protein that is put in the solution binds cooperatively to a template.

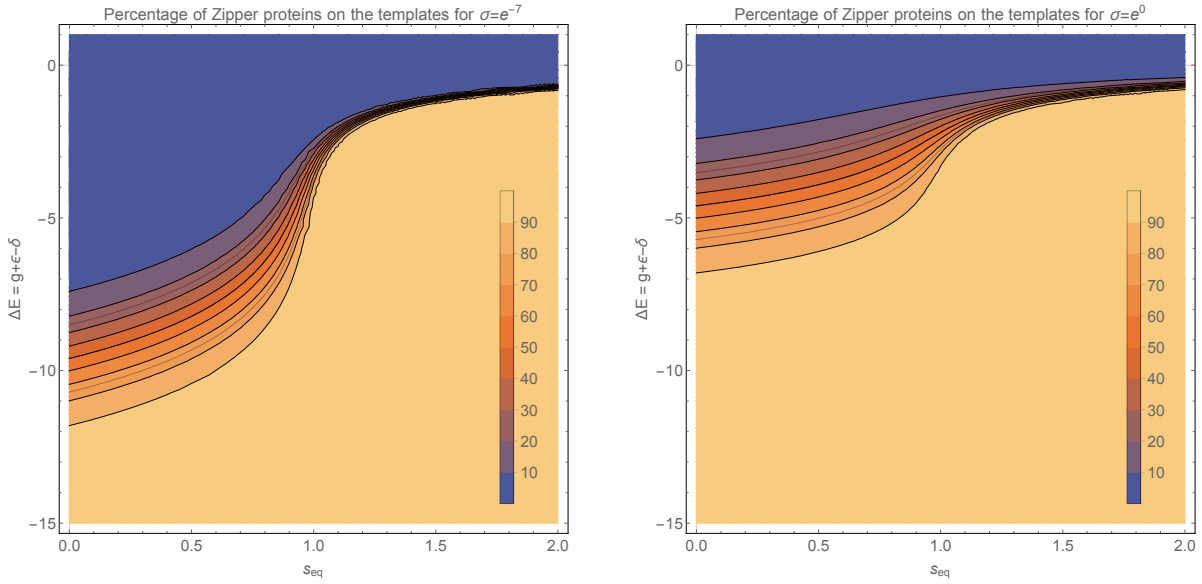
2.4 Phase Transition

To get a better view of the competition between these two models and of how all these components change the equilibrium, we define a new quantity, namely, the fraction of cooperatively adsorbed proteins on the templates,

$$f_Z = \frac{\langle\theta\rangle_{Z,eq}}{\langle\theta\rangle_{tot,eq}} = \frac{\langle\theta\rangle_{Z,eq}}{\langle\theta\rangle_{Z,eq} + \langle\theta\rangle_{L,eq}}, \tag{2.20}$$

This quantity does not give us insight of the fraction of occupied sites on the templates, however, it gives us a phase diagram and shows us whether zipper- or Langmuir-type of binding is preferred. We can more

clearly study when the transition happens. In Figure 2.9 this parameter is shown as a function of the energy difference between cooperative binding and non-cooperative binding $\Delta E = g + \epsilon - \delta = \ln \phi_c$ and s_{eq}. This figure shows that if $\rho_{fp} > \phi_Z$ (so $s_{eq} > 1$) the fraction of cooperatively bound proteins only depends on whether its free energy is more favourable than the binding free energy of Langmuir. So, whether ϕ_c is larger or smaller than 1. This is because in the zipper model the template is almost fully encapsulated for $s_{eq} > 1$. So, the unfavourable nucleation energy has almost no impact on the total internal energy of this energy. If $s_{eq} < 1$ the Langmuir-type of binding is almost always preferred. Only if the adsorption free energy of the zipper-type of binding is much more favourable than the adsorption free energy of the Langmuir-type of binding, the cooperative binding gets the overhand. This is a result of the cooperative binding not preferring to bind for $s_{eq} < 1$, consequently the non-cooperative free energy has to be very small to have a larger fraction of cooperatively bound proteins than the fraction of non-cooperatively bound proteins. In addition, if the total protein number density is smaller than the critical zipper density, $S < 1$, thus $s_{eq} < 1$, the nucleation energy has a great impact as shown in Figure 2.6. In the limit of the free protein density in equilibrium to infinity, $\rho_{fp} \to \infty$, the line of $f_Z = 0.5$ approaches $\Delta E = 0$ asymptotically. When $\Delta E = 0$ the non-cooperative binding is preferred, because the non-cooperative binding generates more configurations than the zipper model, consequently the occurrence of non-cooperative binding has a larger probability.



(a) (b)

Figure 2.9: A diagram of the percentage of the proteins being cooperatively bound on the templates as a function of ΔE and s_{eq}. The nucleation barrier only has an influence on the binding for $s_{eq} \leq 1$. When s_{eq} is large the line of $f_Z = 0.5$ goes asymptotically to $\Delta E = 0$. In Figure **??** a nucleation barrier is present, $\sigma = \exp -9$. In Figure **??** no nucleation barrier is present, $\sigma = \exp 0$.

e However, we can not conclude from studying this function f_Z if a phase transition occurs.
In Appendix B we describe the heat capacity for the zipper- and Langmuir model combined to study whether a phase transition occurs. The heat capacity shows a discontinuity if the phase transition occurs in the mean-field theory [15]. The heat capacity per binding site of the Langmuir- and zipper model combined has a maximum when the transition from non-cooperative to cooperative binding occurs. The maximum scales with the total binding sites per template q and the peak becomes narrower for larger q. For the infinite limit of q and the energy barrier h the heat capacity becomes a delta-peak. So, for finite q the transition appears to be a phase transition, but studying the heat capacity shows it is not.

2.5 Micelles

Experiments show that increasing the protein number density in the solution may result in a lower fraction of occupied sites on the template and that the creation of micelles occur when there is a high concentration of proteins in the solution. The question is if a higher protein density results in a lower fraction of occupied sites on the templates due to the micelle assembly [11]. To describe the aggregation of micelles we consider the Debye model for micelles [10].
Define the average number of proteins of which a micelle consists $d = \langle N \rangle$. This model assumes that for the infinite limit of the volume, $V \to \infty$, the probability of a micelle to consist of the average number of proteins is one, $P(d) \to 1$. In this limit we can consider all the micelles to consist of d proteins. This is given by the following reaction

$$dP \rightleftharpoons P_d. \tag{2.21}$$

Here, P is a free protein in the solution and P_d is a micelle consisting of P proteins. So, we assume every micelle consists of d proteins. From this reaction follows that in equilibrium $\exp(d\mu_{fp}) = \exp(\mu_m)$, with μ_m the (dimensionless) chemical potential of a micelle and μ_{fp} the dimensionless chemical potential of the free proteins. By minimizing the dimensionless Helmholtz free energy for this system we derive that the number density of micelles in the solution is given by $\rho_m = \exp(-E_{int}(d) + \mu_m)$. Here $E_{int}(d)$ is the dimensionless internal free energy for a micelle consisting of d proteins. We consider the internal free energy to be given by dQ, with $Q < 0$ the dimensionless energy per protein in the micelle. The free energy Q is due to attractive interactions between the micelles. The density of micelles in equilibrium in the solution can thus be written as

$$\rho_m = \exp\Big(- d(Q - \mu_{fp})\Big) = \left(\frac{\rho_{fp}}{\phi_M}\right)^d,$$

with ϕ_M the critical density for the assembly of micelles and ρ_{fp} the free protein density in equilibrium scaled to the thermal energy β. We consider the total protein density to be conserved. Due to mass conservation we can describe the total protein number density as $\rho_p = \rho_{fp} + d\rho_m + \sum_{n=0}^{q} n\rho_t(n)$. Here is ρ_t the template number density with n adsorbed proteins.
First, we consider there to be no templates in the solution $\rho_t = 0$. In Figure 2.10, the density of free proteins scaled with the critical micelle density is shown as a function of the total protein number density scaled with the critical micelle density. Here $\rho_t = 0$, so due to mass conservation every protein that is not a free protein is part of a micelle. For large micelles $d \gg 2$ there is no micelle aggregation when the total protein density is smaller than the critical micelle density $\rho_p < \phi_M$. However, for $\phi_M < \rho_p$ every next protein put in the solution will be part of a micelle, consequently the free protein number density is never larger than the critical micelle density for $d \gg 2$.
If the micelles are small the assembly of micelles already occurs for densities smaller than the critical micelle density, and for large total protein concentrations the free protein density is larger than the critical micelle assembly.
For the capsid- and micelle assembly combined we are interested in for what relation between the critical densities and the total densities the capsid assemblies occurs. We experienced in Chapter 2.3 that the cooperative capsid assembly only occurs if the critical zipper density is smaller than the critical Langmuir density and the total protein number density, $\phi_Z < \phi_L$ and $\phi_Z < \rho_p$. For the reason that experiments show the occurrence of the cooperative assembly, we assume $\phi_Z < \phi_L$ and $\phi_Z < \rho_p$ in this chapter.

If the critical micelle density is smaller than the critical zipper density $\phi_M < \phi_Z$, the cooperative binding only occurs if $\rho_{fp,eq} > \phi_Z$ but this is only true for very large protein densities $\rho_p \gg \phi_Z$ and d has to be small, with d the number of proteins in a micelle, for the reason that for large d, $\rho_{fp,eq} < \phi_M$ for any ρ_p (see Figure 2.10).
If $\phi_Z < \phi_M$ the situation becomes more complicated. We still consider $\phi_Z < \phi_L$ and $\phi_Z < \rho_p$. If $\phi_Z \ll \phi_M$ and $\phi_Z \ll \rho_p$, the cooperative assembly always occurs and has a high average fraction of occupied binding sites. For the infinite limit of d the cooperative binding occurs as long as $\phi_Z < \phi_M$ and the micelle assembly stays absent.
If the total protein number density ρ_p is a small amount larger than the critical zipper density, cooperative

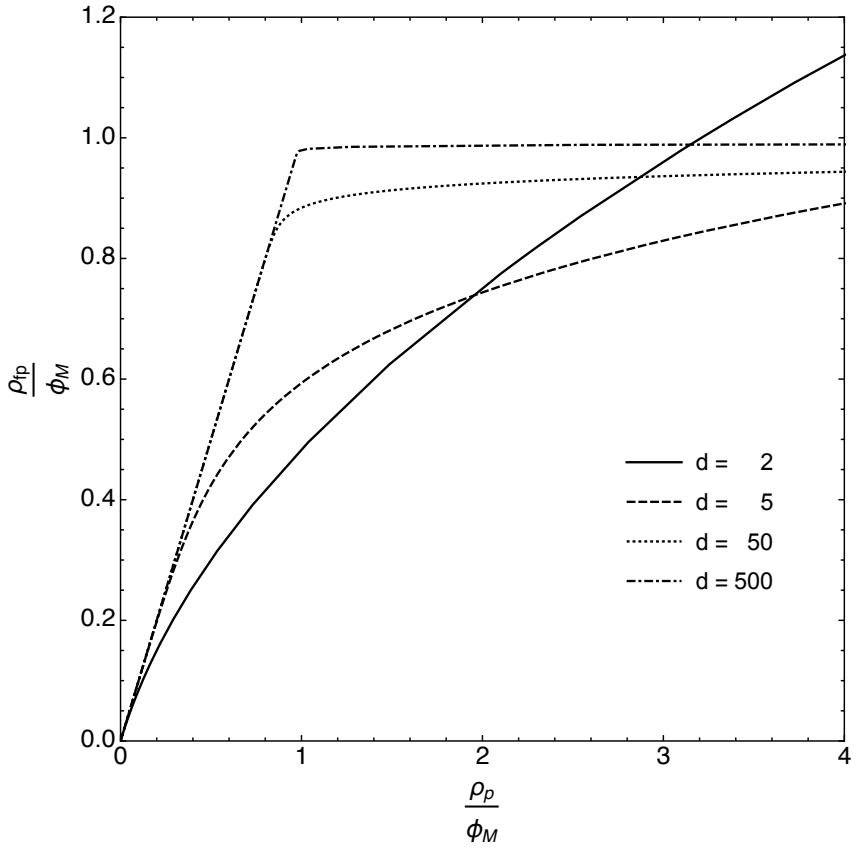


Figure 2.10: The density of free proteins in the solution in equilibrium scaled with the critical micelle density as a function of the density of the total proteins in the solution scaled with the critical micelle density for $d = 2, 5, 50, 500$, with d the number of proteins per micelle. For large micelles $d \gg 2$ the micelle assembly starts when $\rho_p = \phi_M$ and the free protein number density is never larger than the critical micelle density.

binding only occurs provided $\phi_Z \ll \phi_L$ and (for small d) $\phi_Z \ll \phi_M$. Otherwise the random adsorption or the micelle assembly starts and consequently the free protein density drops below the the critical zipper density and the cooperative binding stays absent. For $\phi_L < \phi_M$, the non-cooperative binding still occurs, however, the probability of finding a fully encapsulated template is very small, due to the entropic effect on the non-cooperative binding.

3 Dynamics

In the previous chapter we studied the fraction of occupied binding sites in equilibrium. However, we are also interested in the behaviour of the fraction of occupied binding sites as a function of the time. To analyse the time evolution of the assemblies, we need to define the pathways of the assembly. As outlined in Chapter 1.3, the proteins first bind non-cooperatively to the RNA templates. These proteins can move along the template and a protein could nucleate on one end of the template. The cooperative binding can start from there after the nucleation. [7]. Figure 1.2 shows a schematic illustration of these pathways. In this chapter we describe the differential equations of the binding models and engage in a numerical analysis of these equations.
Firstly, we focus on the differential equations of the zipper model, because the cooperative assembly only depends on one reaction. Secondly, we describe the differential equations of the Langmuir model. Thirdly, we add the micelles, which changes the equation of the total protein number density, due to the mass conservation. And finally, we numerically analyse the influence of the parameters of the model. We only study all the model combined and do not study the characteristics of the single models.

3.1 Zipper Model

The zipper model assembly only depends on one reaction, namely the reaction of a non-cooperatively (Langmuir) bound protein binding cooperatively (zipper) and contrariwise. The reaction can be described as

$$p_L + T_n \underset{Z_-(n)}{\overset{Z_+(n)}{\rightleftharpoons}} T_{n+1}, \tag{3.1}$$

for $0 \leq n \leq q-1$. Here, T_n is a template with n proteins cooperatively bound and p_L is a non-cooperatively bound protein on the same template. Z_- and Z_+ are the assembly constants which denote the rate of the reaction. The reaction depends on the presence of a Langmuir-type of protein adjacent to the cooperative proteins. The probability of this is defined as the occupation number of the non-cooperative proteins divided by the number of sites that is not occupied by the cooperative binding (see Figure 2.7). This can be written as

$$\langle\theta_L\rangle^*(t) = \frac{m}{q-n} = \frac{\langle\theta_L\rangle(t)}{1-\langle\theta_Z\rangle(t)}, \tag{3.2}$$

with m the number of non-cooperative proteins bound to the template and n the number of cooperative proteins, and q the number of binding sites per template. This function denotes the probability of a non-cooperative protein being adjacent to a cooperative protein. The assembly rate of the cooperative binding highly depends on this fraction. In the next section we discuss how we can describe this function, however, first we only describe the cooperative binding.
Similar to the previous Chapter, we want to describe the fraction of occupied binding sites of the cooperative bound proteins, however, now as a function of time t, $\langle\theta_Z\rangle(t) = \frac{1}{q}\sum_{n=0}^{q} nP(n,t)$, with $P(n,t)$ the probability of a template having n cooperatively bound proteins at time t. The time dependence of the fraction of occupied binding sites of the zipper model is given by $\frac{\partial\langle\theta_Z\rangle(t)}{\partial t} = \frac{1}{q}\sum_{n=0}^{q} n\frac{\partial P(n,t)}{\partial t}$. For this reason the partial derivatives of $P(n,t)$ can be used to describe the derivative of the fraction of occupied binding sites. We define the reaction $v_n(t) = Z_+(n)\langle\theta_L\rangle^*(t)P(n-1,t) - Z_-(n)P(n,t)$ for $1 \leq n \leq q$. This describes the rate of the reaction of a template with $n-1$ to n adsorbed cooperative proteins. Considering the reaction in Equation 3.1, the differential equation of the probability of a template being empty can be written as

$$\frac{\partial P(0,\tau)}{\partial \tau} = -v_1(t). \tag{3.3}$$

For the derivative of the probability of a template having n proteins, with $2 \leq n \leq q-1$, we obtain

$$\frac{\partial P(n,\tau)}{\partial \tau} = v_n(t) - v_{n+1}(t), \tag{3.4}$$

and for fully encapsulated templates the time derivative is given by

$$\frac{\partial P(q,\tau)}{\partial \tau} = v_q(t). \tag{3.5}$$

These equations can be compared to the statistical properties when applying the infinite limit of the time t $\lim_{t\to\infty}\langle\theta_L\rangle^*(t) = \langle\theta_L\rangle^*_{eq}$, $\lim_{t\to\infty} P(n,t) = P_{eq}(n)$ and $\lim_{t\to\infty}\frac{\partial P(n,t)}{\partial t} = 0$, for $0 \leq n \leq q$. As a result $\lim_{t\to\infty} v_n(t) = 0$ for $1 \leq n \leq q$. From the statistical physics for the Langmuir- and zipper model combined, follows that the relation between $P_{eq}(n)$ and $P_{eq}(0)$ is given by $P_{eq}(n) = \sigma(\langle\theta_L\rangle^*_{eq}\phi_c)^n P_{eq}(0)$. Here, $\phi_c = \frac{\phi_Z}{\phi_L}$, with ϕ_Z and ϕ_L the critical zipper- and Langmuir density, and $\sigma = e^{\epsilon - h}$, denoting the nucleation barrier. This gives us for the relation of $Z_+(n)$ and $Z_-(n)$

$$\frac{Z_-(n)}{Z_+(n)} = \begin{cases} \frac{\sigma}{\phi_c}, & \text{for} \quad n = 1, \\ \frac{1}{\phi_c}, & \text{for} \quad 2 \leq n \leq q. \end{cases} \tag{3.6}$$

Next we simplify the model by defining $Z_+(n) = Z_+$, for $2 < n < q$, and $Z_+(1) = \kappa Z_+$. The parameter κ denotes the difficulty of the nucleation. We furthermore introduce $\tau = Z_- t$. Applying all the defined quantities, we can write down four differential equation that describe the intra-chain binding of the zipper model. The differential equation for the probability of the template having no cooperatively bound proteins is given by

$$\frac{\partial P(0,\tau)}{\partial \tau} = -\kappa\frac{\langle\theta_L\rangle^*(\tau)}{\phi_c}P(0,\tau) + \frac{\kappa}{\sigma}P(1,\tau). \tag{3.7}$$

We also need to define the differential equation of the probability of one cooperatively bound protein separately, because it depends on the nucleation barrier, i.e.

$$\frac{\partial P(1,\tau)}{\partial \tau} = \frac{\langle\theta_L\rangle^*(\tau)}{\phi_c}\Big(\kappa P(0,\tau) - P(1,\tau)\Big) + P(2,\tau) - \frac{\kappa}{\sigma}P(1,\tau). \tag{3.8}$$

The time derivative of a template having n cooperatively bound proteins, for $2 \leq n \leq q-1$ is given by

$$\frac{\partial P(n,\tau)}{\partial \tau} = \frac{\langle\theta_L\rangle^*(\tau)}{\phi_c}\Big(P(n-1,\tau) - P(n,\tau)\Big) + P(n+1,\tau) - P(n,\tau), \tag{3.9}$$

And the time dependence of the probability of a template being fully encapsulated is written down as

$$\frac{\partial P(q,\tau)}{\partial \tau} = \frac{\langle\theta_L\rangle^*(\tau)}{\phi_c}P(q-1,\tau) - P(q,\tau). \tag{3.10}$$

This is the form of the differential equations of the cooperative binding, which we will be applying in our numerical analysis. In the next section we focus on the differential equation of the non-cooperative binding.

3.2 Langmuir Model

The assembly of the random adsorption depends on two reaction, namely the assembly of a protein in the solution binding non-cooperatively to a binding site on a template and the reaction of a non-cooperatively bound protein to cooperative binding. The latter is described in the previous section. Due to mass conservation the equation for the Langmuir model can be written as

$$\frac{\partial\langle\theta_L\rangle(t)}{\partial t} = -\frac{\partial\langle\theta_Z\rangle(t)}{\partial t} + F(t). \tag{3.11}$$

Here, $F(t)$ is a function dependent on the time t as a result of the random adsorption of proteins from the solution on the templates. The reaction of a free protein binding randomly on a template can be written down as follows

$$p_s + T_m \underset{K_-}{\overset{K_+}{\rightleftharpoons}} T_{m+1}, \tag{3.12}$$

for $0 \leq m \leq q-1$. Here, p_s is a free protein in the solution, T_m a template with m filled binding sites by non-cooperative bound proteins and K_- and K_+ are assembly constants which denote the rate of the reaction. The disassembly rate is proportional to the fraction of non-cooperative bound proteins and the

assembly rate with the fraction of free binding sites and the free protein density. The time derivative of the fraction of non-cooperative bound proteins is given by

$$\frac{\partial\langle\theta_L\rangle(t)}{\partial t} = L_+\rho_{fp}(t)\Big(1-\langle\theta_L\rangle(t)\Big) - L_-\langle\theta_L\rangle(t) - \frac{\partial\langle\theta_Z\rangle(t)}{\partial t}. \tag{3.13}$$

Similarly to the previous section, we compare the differential equation with the statics by considering the infinite limit of the time t of the differential equations $\lim_{t\to\infty}\langle\theta_L\rangle(t) = \langle\theta_L\rangle_{eq}$, $\lim_{t\to\infty}\frac{\partial\langle\theta_L\rangle(t)}{\partial t} = 0$ and $\lim_{t\to\infty}\frac{\partial\langle\theta_Z\rangle(t)}{\partial t} = 0$. This gives the ratio between the assembly constants of the non-cooperative binding, L_- and L_+,

$$\frac{L_+}{L_-} = \phi_L, \tag{3.14}$$

with ϕ_L the critical Langmuir density. Next we define the quantity $C_{LZ} = \frac{L_-}{Z_-}$, $s(t) = \frac{\rho_{fp}(t)}{\phi_Z}$, with $\rho_{fp}(t)$ the free protein density at time t, and $f_{fp}(t) = \frac{s(t)}{S}$, with $S = \frac{\rho_p}{\phi_Z}$. The difference between the rate of the disassembly of a non-cooperatively bound protein and of cooperatively bound protein is given by C_{LZ}. f_{fp} describes the fraction of proteins that is free in the solution. From this the following equation can be derived

$$\frac{\partial\langle\theta_L\rangle(\tau)}{\partial\tau} = C_{LZ}\phi_c S f_{fp}(\tau)\Big(1-\langle\theta_L\rangle(\tau)\Big) - C_{LZ}\langle\theta_L\rangle(\tau) - \frac{\partial\langle\theta_Z\rangle(\tau)}{\partial\tau}. \tag{3.15}$$

However, we are interested in the probability of a non-cooperatively bound protein being adjacent to the cooperative proteins $\langle\theta_L\rangle^*(\tau)$, defined in Equation 3.2, because the differential equations of the cooperative binding depend on this function. The time derivative of this function is given by $\frac{\partial\langle\theta_L\rangle^*(\tau)}{\partial\tau} = \frac{1}{1-\langle\theta_Z\rangle(\tau)}\Big(\frac{\partial\langle\theta_L\rangle(\tau)}{\partial\tau} + \frac{\partial\langle\theta_Z\rangle(\tau)}{\partial\tau}\langle\theta_L\rangle^*(t)\Big)$. Combining this with equation 3.15, we obtain the following differential equation

$$\frac{\partial\langle\theta_L\rangle^*(\tau)}{\partial\tau} = \frac{\partial\langle\theta_Z\rangle(\tau)}{\partial\tau}\frac{\langle\theta_L\rangle^*(\tau)-1}{1-\langle\theta_Z\rangle(\tau)} + C_{LZ}\phi_c S f_{fp}(\tau)\Big(1-\langle\theta_L\rangle^*(\tau)\Big) - C_{LZ}\langle\theta_L\rangle^*(\tau). \tag{3.16}$$

This equation has a dependence on the same parameters as those of Equations 3.7 to 3.10. It introduces one extra parameter, C_{LZ}, which is the ratio of the assembly rate of the non-cooperative and that of the cooperative proteins. The influence of all these parameters is studied in the numerical analysis in section 3.4. In the next section we describe the differential equation for micelle assembly and the formula for the mass conservation.

3.3 Micelle and Free Protein Density

The next step is to consider the reaction of the assembly for the micelles. The micelle assembly is dependent on one equation. This equation is given by

$$dP \underset{M_-}{\overset{M_+}{\rightleftharpoons}} P_d, \tag{3.17}$$

with $d \geq 2$ the number of proteins a micelle consists of and P a free protein in the solution. The assembly has a dependence on the free protein density $\rho_{fp}(t)$ and the disassembly on the density of micelles $\rho_m(t)$, i.e.

$$\frac{\partial\rho_m(t)}{\partial t} = M_+(\rho_{fp}(t))^d - M_-\rho_m(t). \tag{3.18}$$

We follow the same steps as in the previous sections. Firstly, we study the differential equation of the assembly in the infinite limit of the time t and compare the result with the results of the statics of the micelle assembly $\lim_{t\to\infty}\frac{\partial\rho_m(t)}{\partial t} = 0$, $\lim_{t\to\infty}\rho_m(t) = \rho_{m,eq}$ and $\lim_{t\to\infty}\rho_{fp}(t) = \rho_{fp,eq}$. From this follows a relation between the assembly constants of the micelle assembly

$$\frac{M_+}{M_-} = \frac{\rho_{m,eq}}{(\rho_{fp,eq})^d}. \tag{3.19}$$

Secondly, we define $C_{MZ} = \frac{M_-}{Z_-}$, which is the ratio of the disassembly rate of the micelles to that of the cooperative binding, and we define the fraction of proteins being part of a micelle $f_m(t) = \frac{\rho_m(t)}{\rho_p}$, with ρ_p the total protein number density. The differential equation of the micelle assembly can be rewritten as

$$\frac{\partial f_m(\tau)}{\partial \tau} = \frac{C_{MZ}}{\phi_M}\left(\frac{S}{\phi_{MZ}}\right)^{d-1} f_{fp}^d(\tau) - C_{MZ} f_m(\tau), \tag{3.20}$$

with ϕ_M the critical micelle density, $\phi_{MZ} = \frac{\phi_M}{\phi_Z}$ and $\tau = Z_- t$
Finally, we are interested in the fraction of free proteins in the solution $f_{fp}(t)$. A constraint for this function can be derived using the conservation of mass. In Chapter 2.5 we defined a constraint in equilibrium due to the conservation of mass. However, this should hold for any time τ, so, we can write the total protein density as $\rho_p = \rho_{fp}(\tau) + d\rho_m(\tau) + q\rho_t\langle\theta_{tot}\rangle(\tau)$, with ρ_t the density of templates in the solution. This can be rewritten in the form

$$f_{fp}(\tau) = 1 - df_m(\tau) - \lambda\Big(\langle\theta_L\rangle^*(\tau) + \langle\theta_Z\rangle(\tau) - \langle\theta_L\rangle^*(\tau)\langle\theta_Z\rangle(\tau)\Big), \tag{3.21}$$

with $\lambda = \frac{q\rho_t}{\rho_p}$. Hence, λ is a constant and denotes the stoichiometry of the reaction. If $\lambda > 1$ there is an excess in binding sites and not all the templates can be fully encapsulated. The maximum fraction of occupied binding sites is then given by $\frac{1}{\lambda}$.

3.4 Numerical Analysis

We defined the time derivatives of the fraction of occupied sites of cooperatively and non-cooperatively bound proteins, $\langle\theta_Z\rangle(\tau)$ and $\langle\theta_L\rangle^*(\tau)$ and the fraction of proteins that are part of a micelle $f_m(\tau)$. Furthermore, we defined the fraction of proteins that is free in the solution $f_{fp}(\tau)$. We assume that the starting conditions are given by $\langle\theta_Z\rangle(0) = 0$, $\langle\theta_L\rangle^*(0) = 0$, $f_m(0) = 0$ and $f_{fp}(0) = 1$.
First we consider there to be no micelle assembly taking place in the solution $\rho_m(\tau) = 0$ for any τ. For small τ proteins bind non-cooperatively to the binding sites on the templates. If $\phi_Z < \phi_L$ and ρ_p is larger than the critical zipper density ϕ_Z, the cooperative binding starts when the random adsorption is close to equilibrium. This is shown in Figure 3.1.
For small τ, if the total protein density ρ_p is smaller than the critical zipper density ϕ_Z or if the critical Langmuir density is smaller than the critical zipper density, the time-derivative of the fraction of occupied sites of the zipper model is negligible $\frac{\partial\langle\theta_Z\rangle(\tau)}{\partial\tau} \ll 1$. Assuming $\frac{\partial\langle\theta_Z\rangle(\tau)}{\partial\tau} = 0$, the differential equation of the Langmuir model can be solved analytically. The analytical solution of the Equation 3.16 with $\frac{\partial\langle\theta_Z\rangle(\tau)}{\partial\tau} = 0$ is given by

$$\langle\theta_L\rangle^*(\tau) = \langle\theta_L\rangle_{eq}^*\Big(1 - e^{\frac{-C_{LZ}S\phi_c}{\langle\theta_L\rangle_{eq}^*}\tau}\Big), \tag{3.22}$$

with C_{LZ} the ratio between the assembly constants of the cooperative and non-cooperative binding, S the total protein density scaled with the critical zipper density and ϕ_c the critical zipper density scaled with the critical Langmuir density. From this formula follows that for $\tau = 0$ the rate of the non-cooperative binding scales with C_{LZ}, S, and ϕ_c. This formula provides a good fit for small τ, for $\phi_L < \phi_Z$ or for $\phi_Z > \rho_p$.
However, we are interested in the occurrence of the cooperative binding, for the reason that this is what is observed in the experiments [7], so we assume $\phi_Z < \phi_L$ and $\phi_Z < \rho_p$, because only under these conditions the cooperative assembly occurs. If one of these two conditions is not true, the assembly can be described by Equation 3.22. The energy barrier σ denotes the difficulty of the nucleation for the cooperative binding. In Figure 3.1 the effect of the presence of a nucleation barrier is shown. A larger nucleation barrier $\sigma < 1$ results in a delay of the assembly of the cooperative binding. The parameter κ also denotes the energy barrier. From Equation 3.7 and Equation 3.8 follows that decreasing κ, makes the nucleation of a template more difficult, but there are a few differences between the parameter σ and κ. A small κ, so $\kappa \ll 1$, implies that $Z_+(1) \ll Z_+(n)$, however, it also implies that $Z_-(1) \ll Z_-(n)$. Consequently, the lag time of decreasing κ is much smaller than that of σ.

The number of differential equations of the zipper model scales with the number of binding sites per protein q, therefore the lag time for the cooperative adsorption also scales with q. The number of binding sites per

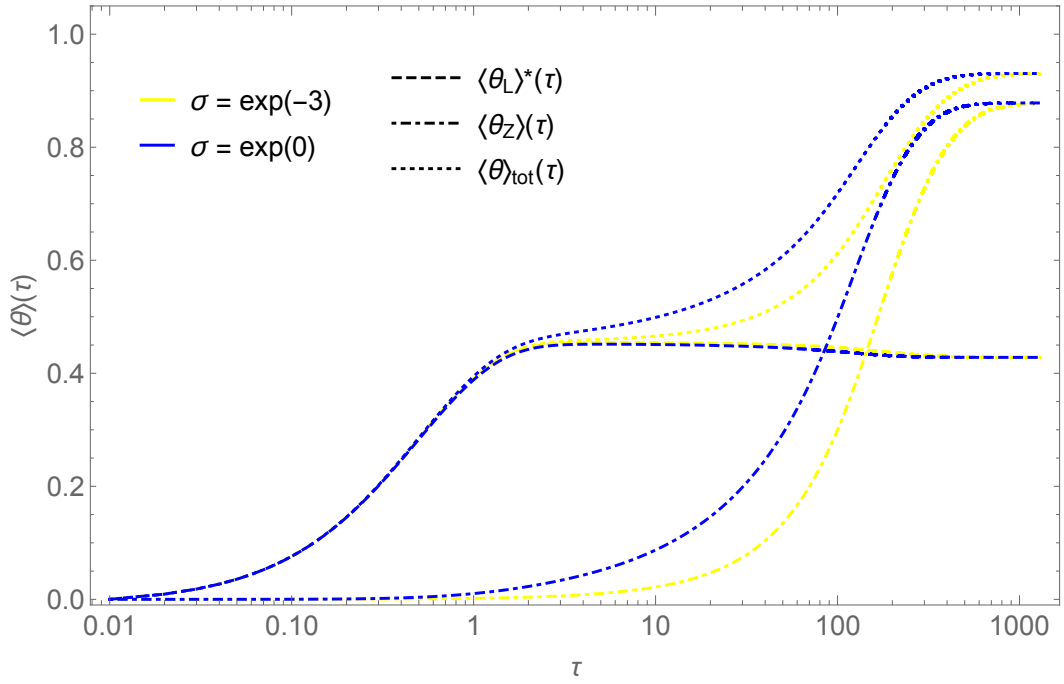


Figure 3.1: The fraction of occupied binding sites on the templates by cooperatively bound proteins $\langle\theta_Z\rangle(\tau)$, the total fraction of occupied binding sites $\langle\theta_{tot}\rangle(\tau)$, and the probability of a non-cooperatively bound protein being adjacent to the cooperative proteins $\langle\theta_L\rangle^*(\tau)$ as a function of the time τ with and without a nucleation barrier; $\sigma = \exp(-3)$ (yellow lines) and $\sigma = 1$ (blue lines). The parameters are the ratio of the the critical density of the Langmuir and the zipper density $\frac{\phi_Z}{\phi_L} = e^{-1}$, the number of binding sites per template $q = 50$, the comparison between the disassembly constant of the zipper model and Langmuir model $C_{LZ} = 1$, and the ratio of the total number of proteins and the critical zipper density equals $S = 2.5$. The stoichiometry equals, $\lambda = \frac{1}{5}$, so there is an excess of proteins in the solution. Increasing the energy barrier results in a delay of the cooperative binding. It has almost no influence on $\langle\theta_L\rangle^*(\tau)$

protein q does not have any influence on the binding rate of the non-cooperative binding.
The parameter that changes the rate of the random adsorption is C_{LZ}, which denotes the ratio of the disassembly constant of the zipper model to that of the Langmuir model. Increasing C_{LZ} increases the disassembly constant of the random adsorption L_-, however, the critical Langmuir density $\phi_L = \frac{L_+}{L_-}$ does not change. Consequently, the equilibrium values do not change. So, increasing C_{LZ} results also in an increase of L_+ and thus the random adsorption rate increases. The lag time of the random adsorption scales with $\frac{1}{C_{LZ}}$. It also has an influence on the assembly of the cooperative binding. This is because it changes the rate of the random adsorption, and the cooperative binding has a dependence on the fraction of occupied sites of the random adsorption (see Equations 3.7 to 3.10).
Now consider micelle assembly to take place. The micelle assembly brings four new parameters into the model: C_{MZ}, which compares the disassembly constant of the zipper adsorption model with that of the micelle assembly model, d, which denotes how many proteins a micelle is built up of, ϕ_M, the critical micelle concentration given by $\exp(Q)$ with $Q < 0$ the free energy per protein in a micelle, and $\phi_{MZ} = \frac{\phi_M}{\phi_Z}$, which is the ratio of the critical micelle density to the critical zipper density.
The ratio of the critical density of the micelle assembly to the zipper model ϕ_{MZ} is an important parameter for the equilibrium as shown in Chapter 2.5. Remember that for $\phi_M < \phi_Z$ the cooperative binding can only occur if $\rho_{fp,eq} > \phi_Z$, but this is for very large protein densities $\rho_p \gg \phi_Z$ and small d. An increase of ϕ_{MZ} with a factor a, results in an increase of the assembly rate at time $\tau = 0$ with a^{d-1}. Here, d is the number of proteins per micelle. (see Equation 3.20).
From analysing Equation 3.20 it appears that the assembly rate increases if the number of proteins per micelle d increases. This feels contradictory. This is however not what actually happens. It appears as such as a

result of the internal energy increasing per micelle with $Q < 0$; $E_{m,int} = dQ$, with $E_{m,int}$ the internal energy of a micelle. However, the assembly constants of the micelle assembly M_- and M_+ are also dependent of d, so C_{MZ} is dependent of d. Increasing d results in a decrease of the assembly constants M_- and M_+ and thus C_{MZ} decreases.
C_{MZ} denotes the assembly rate of the micelles. For a large C_{MZ} the micelle assembly has a higher assembly rate than the capsid assembly. If the fraction of occupied sites in equilibrium is large, it results in an overshoot of the number of micelles in the solution, so $\rho_m(t) > \rho_{m,eq}$ for some time τ (see Figure 3.2).

A small C_{MZ} has an opposite effect (see Figure 3.2 and Figure 3.3). For very small $C_{MZ} \ll 1$ the capsid

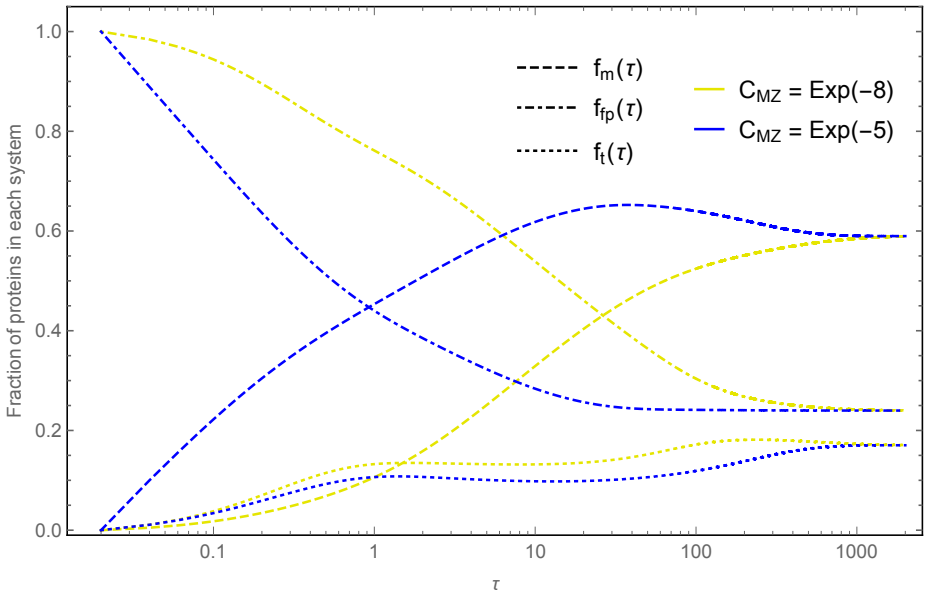


Figure 3.2: The distribution of the proteins as a function of time. The proteins are either freely in the solution $f_{fp}(\tau)$, in a micelle $f_m(\tau)$ or bound on a template $f_t(\tau)$. This is done for two values of C_{MZ}: $C_{MZ} = \exp(-5)$ (the blue lines), and $C_{MZ} = \exp(-8)$ (the yellow lines). $C_{MZ} = \exp(-5)$ results in an overshoot for the micelles. From $C_{MZ} = \exp(-8)$ follows an overshoot in the fraction of occupied sites on the templates. The critical densities are $\phi_Z = \exp(-12)$, $\phi_L = \exp(-11)$, $\phi_M = \exp(-9)$. And the total number of proteins is given by $S = \exp(2)$. For the additional parameters we have used the same values as in Figure 3.1

assembly has an overshoot as a result of the micelle assembly going slowly to its equilibrium. At first the free protein density at time t, $\rho_{fp}(t)$, is larger than the critical Langmuir and zipper density ϕ_L and ϕ_Z, this results in the capsid assembly. However, micelles form slowly and this brings the density of free proteins to a value for which $\langle\theta_{tot}\rangle < 0$. For high protein densities ρ_p, a very small C_{MZ} and $\rho_{fp,eq} < \phi_Z$ this would result in a very high average fraction of occupied binding sites $\langle\theta_{tot}\rangle(\tau) \approx 1$ for some time τ, but in equilibrium the average fraction of binding sites would be small $\langle\theta_{tot}\rangle_{eq} \ll 1$.

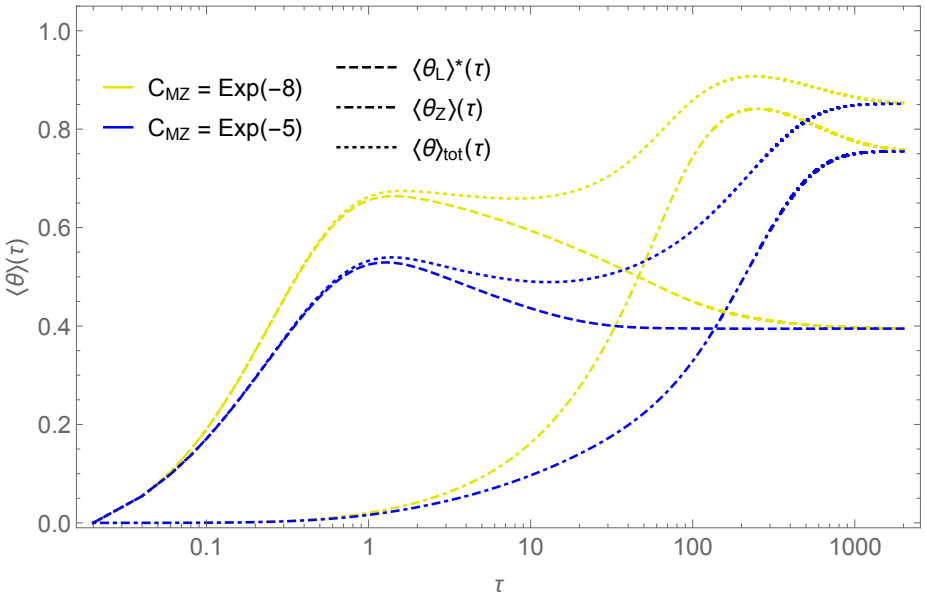


Figure 3.3: The fraction of occupied sites on the templates as a function of the time τ for two different values of $C_{MZ} = \exp(-5)$ (blue lines) and $C_{MZ} = \exp(-8)$ (yellow lines). A small C_{MZ} results in an overshoot for the self-assembly. C_{MZ} does not influence the equilibrium. Additionally, the same values for all the parameters are used as in Figure 3.2

4 Discussion and Conclusion

The measurements of H. Cingil *et al.* on the self-assembly of virus-like particles show an overshoot in the fraction of occupied sites on the templates. An overshoot of the capsid assembly means that the fraction of occupied sites on the templates is exceeds his equilibrium value. It also shows that an increase in the total protein concentration may result in a decrease of the fraction of occupied sites in equilibrium [11]. Both these observations are possibly a result of the micelle assembly in the solution.
In this thesis we have tried to understand these results of the experiments of H. Cingil *et al.* by modelling the self-assembly of virus-like particles. We have combined the Langmuir adsorption model and the kinetic zipper model for our model and added an nucleation barrier [6, 10]. Furthermore, the Debye adsorption model for micelles is applied to describe the micelle assembly, which is observed in the same measurements [12] These three models have been combined and we discussed the equilibrium properties and the dynamics of this combination.
In this Chapter the results are discussed and follow-up research is suggested, additionally the research questions are answered and a conclusion is drawn.

Discussion

Comparing the zipper- and the Langmuir model under the conditions of thermodynamic equilibrium showed that the cooperative binding has a higher probability of a template being fully encapsulated than the non-cooperative binding if the protein densities is larger than the critical density of the binding models. This is due to the cooperative binding ruling out the entropy. The cooperative binding does not occur for smaller total protein densities than the critical zipper density. Combining the two adsorption model showed an occurrence of non-cooperative binding for small protein density, while the cooperative binding stays absent. If the critical zipper density is smaller than that of the Langmuir adsorption, the cooperative binding occurs for a protein concentration that is greater than the critical zipper density and "pushes" the non-cooperative proteins off the template. The critical densities, and especially the relative magnitude between each other, mostly determines what binding method is preferred. If the critical zipper density is larger than that of the Langmuir adsorption or larger than the total protein density the cooperative binding does not occur. From studying the heat capacity of the combination of the zipper- and Langmuir model follows that the shift of non-cooperative binding to the cooperative binding is actually not a phase transition for finite binding sites per template. In the infinite limit of the number of binding sites per template and the nucleation barrier, the transition is indeed a phase transition.
Adding the Debye model for micelles makes the model more complicated. The micelle assembly always results in a lower fraction of occupied sites on the templates in comparison with no micelle assembly. For large micelles the free protein density never exceeds the critical micelle density, however, for protein densities smaller than the critical micelle density, there is no occurrence of the micelle assembly in the solution. If large micelles are preferred and the critical micelle density is smaller than the critical zipper density, no cooperative capsid assembly is observed, due to the predominance of the micelle assembly in that case. This results in the absence of the cooperative binding. However, with this model we do not detect that increasing the protein density S results in a decrease of the fraction of occupied sites in equilibrium $\langle\theta_{tot}\rangle_{eq}$, as may have been the case for the self-assembly of the virus-like particles used in the experiments of H. Cingel *et al.* [11].
From the analysis of the differential equations of the binding methods follow some surprising results. The binding starts with the non-cooperatively bound proteins. If the critical zipper density is smaller than the other critical densities and the total protein density, the cooperative binding starts after some time. The time it takes for the cooperative binding to start is highly dependent on the nucleation barrier. A high nucleation barrier results in a lag time for the cooperative binding. The nucleation barrier has no effect on the non-cooperative binding. The rate of the cooperative assembly scales with total number of binding sites per template. The lag time of the non-cooperative binding is given by the relation of the binding constants of the cooperative binding and non-cooperative binding. This parameter does not influence the cooperative adsorption.
A high rate of the micelle assembly results in an overshoot of the micelles, so the micelle density exceeds for a period of time its value in equilibrium. A small rate results in the opposite, namely an overshoot in the

capsid assembly. The latter is observed for large protein densities in the measurements of H. Cingil *et al.*. The system in these experiments was probably not in equilibrium when the last measurement was done. It is a possibility that the rate of the micelle assembly is smaller at lower protein density. This can explain why only an overshoot was observed at large protein densities. The lower densities do also exceed their equilibrium value, but this is not yet observed. A different

Conclusion

If the critical zipper density is not much smaller than the critical Langmuir density, the fraction of occupied binding sites in equilibrium is affected by the non-cooperative binding. The critical zipper-, Langmuir and micelle density are extremely important in this model and have a great influence on both the equilibrium and the dynamics.
The binding pathways proposed by Cingil *et al.* show that the self-assembly begins with the random, non-cooperative adsorption followed by the cooperative binding. The parameters of the non-cooperative binding have a great influence on the fraction of occupied sites by the cooperative proteins. A small fraction of non-cooperatively bound proteins results in a slow assembly rate of the cooperative.
In the measurements is observed that the capsid assembly can have an overshoot, so the number of occupied sites for some time exceeds the value in equilibrium. A similar result is obtained with the model, however, increasing the total protein density does not result in a lower fraction of occupied sites, as observed in these measurements [11]. To have a better understanding of the experimental finding we propose a measurement of the self-assembly over a longer period of time, such that we are certain we learn the equilibrium values, for the reason that much of the understanding and fitting is done using the equilibrium values of the fraction of occupied sites, the free protein density and the micelle density. The next recommendation is fitting this model with the experimental data of the self-assembly of the virus-like particles and finding the critical density of each model. Finally, we recommend analytically deriving the equations of the assemblies from the differential equation. These equations can assist in the insight of the presence of an under- or overshoot of the micelle or capsid assembly.

A Fraction of occupied binding sites for the Zipper- and Langmuir Model

In Chapter 2.2 we outlined the steps to calculate the fraction of occupied binding sites of the non-cooperatively and cooperatively bound proteins under the conditions of the thermodynamic equilibrium. These formulas are complicated yet very important. For this reason we give a derivation of these functions in this appendix. To determine the fraction of occupied binding sites of the templates, we write down the semi-grand partition function, because this function considers a mass conservation while the composition of the system can change. This function is already introduced in Chapter 2.3 (see Equation 2.18).

$$\Xi = \sum_{n=0}^{q} \sum_{m=0}^{q-n} \binom{q-n}{m} \exp\left(-E_Z(n) - E_L(m) + (n+m)\mu_p\right), \tag{A.1}$$

$$= (1+a_{eq})^q + s_{eq}\sigma \frac{1-(\frac{1}{s_{eq}}(1+a_{eq}))^q}{1-(\frac{1}{s_{eq}}(1+a_{eq}))}, \tag{A.2}$$

with $E_Z(n)$ and $E_L(m)$ the internal free energy of a template with n cooperatively and m non-cooperatively bound proteins, $a_{eq} = \exp(\mu_{fp} - \delta)$, $\sigma = \exp(\epsilon - h)$ and $s_{eq} = \exp(\mu_{fp} - g - \epsilon)$. Here, μ_{fp} is the (dimensionless) chemical potential of the free proteins in the solution. $\delta < 0$ is the free energy per non-cooperatively bound protein on a template, $g + \epsilon < 0$ is the free energy per cooperatively bound protein per template and $h - \epsilon \geq 0$ denotes the energy barrier for the nucleation of the cooperative binding. All free energies are scaled to the thermal energy $k_B T$ and thus dimensionless.
As shown in Chapter 2, to calculate the fraction of occupied binding sites we need to take the derivative of $\ln \Xi$ with respect to μ_{fp}, and divide it by the number of binding sites per template q. This is also given by

$$\langle \theta_{tot} \rangle_{eq} = \frac{1}{q} \frac{\partial \ln \Xi}{\partial \mu_{fp}}, \tag{A.3}$$

$$= \frac{1}{q} \frac{\partial a_{eq}}{\partial \mu_{fp}} \frac{\partial \ln \Xi}{\partial a_{eq}} + \frac{1}{q} \frac{\partial s_{eq}}{\partial \mu_{fp}} \frac{\partial \ln \Xi}{\partial s_{eq}}, \tag{A.4}$$

$$= -\frac{1}{q} \frac{\partial a_{eq}}{\partial \delta} \frac{\partial \ln \Xi}{\partial a_{eq}} - \frac{1}{q} \frac{\partial s_{eq}}{\partial g} \frac{\partial \ln \Xi}{\partial s_{eq}}, \tag{A.5}$$

$$= \langle \theta_L \rangle_{eq} + \langle \theta_Z \rangle_{eq}. \tag{A.6}$$

From this follows that the fraction of occupied binding sites can be written as the sum of two terms, namely the fraction of occupied binding sites with the non-cooperatively bound proteins and that of the cooperatively bound proteins. These derivatives can be calculated. Note that $a_{eq} = \phi_c s_{eq}$, with $\phi_c = \frac{\phi_Z}{\phi_L}$. The fraction of occupied binding sites in equilibrium of the zipper model is given by

$$\langle \theta_Z \rangle_{eq} = \frac{1}{\Xi} \frac{\sigma s_{eq}^{q+1} \Big(q(-1 + s_{eq} - s_{eq}\phi_c) + (1 + s_{eq}\phi_c)\Big((\frac{1}{s_{eq}} + \phi_c)^q - 1 \Big) \Big)}{q(1 + s_{eq}(\phi_c - 1))^2}, \tag{A.7}$$

and the fraction of occupied binding sites of the Langmuir model

$$\begin{aligned} \langle \theta_L \rangle_{eq} = \frac{1}{\Xi} \frac{s_{eq}\phi_c}{q(1 + s_{eq}(\phi_c - 1))^2} \Big(& \sigma q s_{eq}^q (1 + s_{eq}(\phi_c - 1))(\frac{1}{s_{eq}} + \phi_c)^{q-1} \\ & + q(1 + s_{eq}(\phi_c - 1))^2 (1 + s_{eq}\phi_c)^{q-1} - s_{eq}^{q+1} \sigma ((\frac{1}{s_{eq}} + \phi_c)^q - 1) \Big). \end{aligned} \tag{A.8}$$

These are complicated formulas, however, it is determined that for $s_{eq} < 1$ and $q \gg 1$ the fraction of occupied binding sites of the cooperative binding $\langle \theta_L \rangle_{eq} \ll 1$, due to the pre-factor s_{eq}^{q+1}. Further analytical evaluation is not straightforward, so we rely on a numerical analysis.

B Heat Capacity

In Chapter 2.4 the transition from cooperative to non-cooperative adsorption is studied. In this Appendix we make use of the heat capacity to clarify if this transition is indeed a phase transition. We derive the heat capacity and try to understand the behaviour of this function. If a phase transition takes place, the heat capacity shows a discontinuity in mean-field theory [15].
The heat capacity is given by $C_V = \frac{\partial \langle E \rangle}{\partial T}$, with $\langle E \rangle$ the average free energy of the system and T the absolute temperature. To describe the heat capacity, we first describe the average internal energy in equilibrium. This can be written as

$$\frac{\langle E \rangle}{q} = \delta' \langle \theta_L \rangle + \frac{h' - \epsilon'}{q}(1 - P(0,m)) + (g' + \epsilon')\langle \theta_Z \rangle, \tag{B.1}$$

with $P(n,m)$ the probability that a template has n cooperatively- and m non-cooperatively bound proteins in equilibrium, $g' + \epsilon'$ the binding free energy of a cooperatively bound protein and δ' of a non-cooperatively bound one, q the number of binding sites per template, $h' - \epsilon'$ the free energy gained by the nucleation and $\langle \theta_L \rangle_{eq}$ and $\langle \theta_Z \rangle_{eq}$ the fraction of occupied binding sites due to the Langmuir- and zippertype of binding. The functions $\langle \theta_L \rangle_{eq}$, $\langle \theta_Z \rangle_{eq}$ and $P(0,m)$ depend on T. The temperature dependence is hidden in the dimensionless free binding energies, for the reason that we defined $\beta\epsilon' = \epsilon$, $\beta g' = g$, $\beta h' = h$ and $\beta\delta' = \delta$, with $\beta = \frac{1}{k_B T}$. Here, k_B is the Boltzmann constant.
These three functions also depend on $\mu_{fp} = \beta\mu'_{fp}$, with μ'_{fp} the chemical potential of the free proteins. However, we do not examine this temperature dependence, because μ_{fp} is a parameter which describes the free protein number density in equilibrium $\rho_{fp,eq} = \exp(\mu_{fp})$ and we consider that in this model we can change the total number of proteins in the solution, and thus $\rho_{fp,eq}$.
From this follows that the heat capacity is given by

$$C_V = \frac{\partial \langle E \rangle}{\partial \epsilon}\frac{\partial \epsilon}{\partial T} + \frac{\partial \langle E \rangle}{\partial \delta}\frac{\partial \delta}{\partial T} + \frac{\partial \langle E \rangle}{\partial h}\frac{\partial h}{\partial T} + \frac{\partial \langle E \rangle}{\partial g}\frac{\partial g}{\partial T} = -\frac{1}{T}\left(\epsilon\frac{\partial \langle E \rangle}{\partial \epsilon} + \delta\frac{\partial \langle E \rangle}{\partial \delta} + h\frac{\partial \langle E \rangle}{\partial h} + g\frac{\partial \langle E \rangle}{\partial g}\right). \tag{B.2}$$

Calculating this function gives a complicated formula. We observe, from studying the heat capacity as a function of the total protein density divided by the critical zipper density S, that the heat capacity has a maximum if the non-cooperative binding is transitioning to the cooperative binding (see Figure B.1). The heat capacity does not show a discontinuity, however, increasing the number of binding sites q shows a higher and smaller peak for the heat capacity for the transition from non-cooperative to cooperative binding. For the limit of the number of binding sites $q \to \infty$ and the energy barrier $h \to \infty$, the transition can indeed be considered as a phase transition. Increasing the nucleation energy barrier makes the transition sharper, but it also makes the average free energy larger, for this reason, with finite q the transition from non-cooperative binding to cooperative binding does not occur for an infinite h.
The transition from non-cooperative binding to cooperative binding appears to be a phase transition, however, the heat capacity shows it formally is not a phase transition. Only for the infinite limit of q and h the transition is a phase transition.

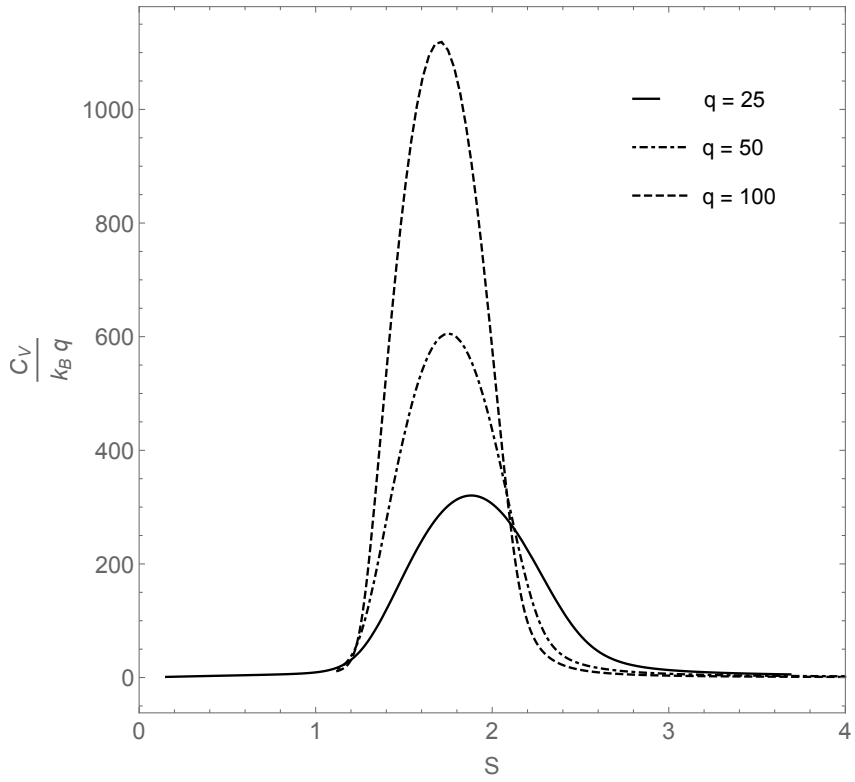


Figure B.1: The heat capacity per binding site for the Langmuir- and zipper model combined as a function of S scaled with the Boltzmann constant k_B for three values of the number of binding sites q. The critical zipper- and Langmuir density equal $\phi_Z = \exp(-10)$ and $\phi_L = \exp(-8)$. The nucleation barrier is given by $\sigma = \exp(-5)$ and the total number of binding sites scaled with the critical zipper density is $L_Z = 1$, so, for $S > 1$, there is an excess of proteins in the solution. Increasing the number of binding sites q results in a higher maximum of the heat capacity per binding site. The heat capacity has a maximum for the value of S when the non-cooperative binding changes to cooperative binding.

References

[1] M. Mateu, "The structural basis of virus function," in *Structure and Physics of Viruses* (M. Mateu, ed.), New York Dordrecht Heidelberg London: Springer, 2013.

[2] D. Caspar and A. Klug, "Physical principles in the construction of regular viruses," *Cold Spring Harbor Symposia On Quantitative Biology*, vol. 27, pp. 1–24, 1962.

[3] S. Modrow, "How are viruses structured?," in *Molecular Virology*, New York Dordrecht Heidelberg London: Springer, 2010.

[4] J. Castón and J. Carrascosa, "The basic architecture of viruses," in *Structure and Physics of Viruses* (M. Mateu, ed.), New York Dordrecht Heidelberg London: Springer, 2013.

[5] A. Bittner, J. Alonso, C. Wege, and M.Ł.Górzny, "Hierarchical assembly into complex structures," in *Structure and Physics of Viruses* (M. Mateu, ed.), pp. 681–682, New York Dordrecht Heidelberg London: Springer, 2013.

[6] D. Kraft, W. Kegel, and P. van der Schoot, "A kinetic zipper model and the assembly of tobacco mosaic virus," *Biophysics Journal*, vol. 102, no. 12, pp. 2845–2855, 2012.

[7] A. Hernandez-Garcia *et al.*, "Design and self-assembly of simple coat proteins for artificial viruses," *Nature Nanotechnology*, vol. 9, no. 2, pp. 698–702, 2013.

[8] M. Punter, "The role of assembly signals in the self-assembly of linear viruses," Master's thesis, Utrecht University, May 2015.

[9] M. Punter *et al.*, "Self-assembly dynamics of linear virus-like particles: Theory and experiment," *The Journal of Physical Chemistry*, vol. 120, no. 26, pp. 6286–6297.

[10] I. Langmuir, "The adsorption of gases on plane surfaces of glass, mica and platinum.," *Journal of the American Chemical Society*, vol. 40, no. 9, pp. 1361–1403, 1918.

[11] H. Cingil *et al.*, "Illuminating the reaction pathways of viromimetic assembly," vol. 139, no. 13, pp. 4962–4968.

[12] P. Debye, "Light scattering in soap solutions.," *The Journal of Physical and Colloid Chemistry*, vol. 53, no. 1, pp. 1–8, 1949.

[13] D. J. Kraft, *Model Systems for Self-Assembly.* PhD thesis, Utrecht University, November 2010.

[14] D. J. Kraft *et al.*, "Surface roughness directed self-assembly of patchy particles into colloidal micelles," *Proceedings of the National Academy of Sciences*, vol. 109, no. 27, pp. 10787–10792, 2012.

[15] S. Blundell and K. M. Blundell, *Concepts in Thermal Physics.*, vol. Second edition. OUP Oxford, 2010.