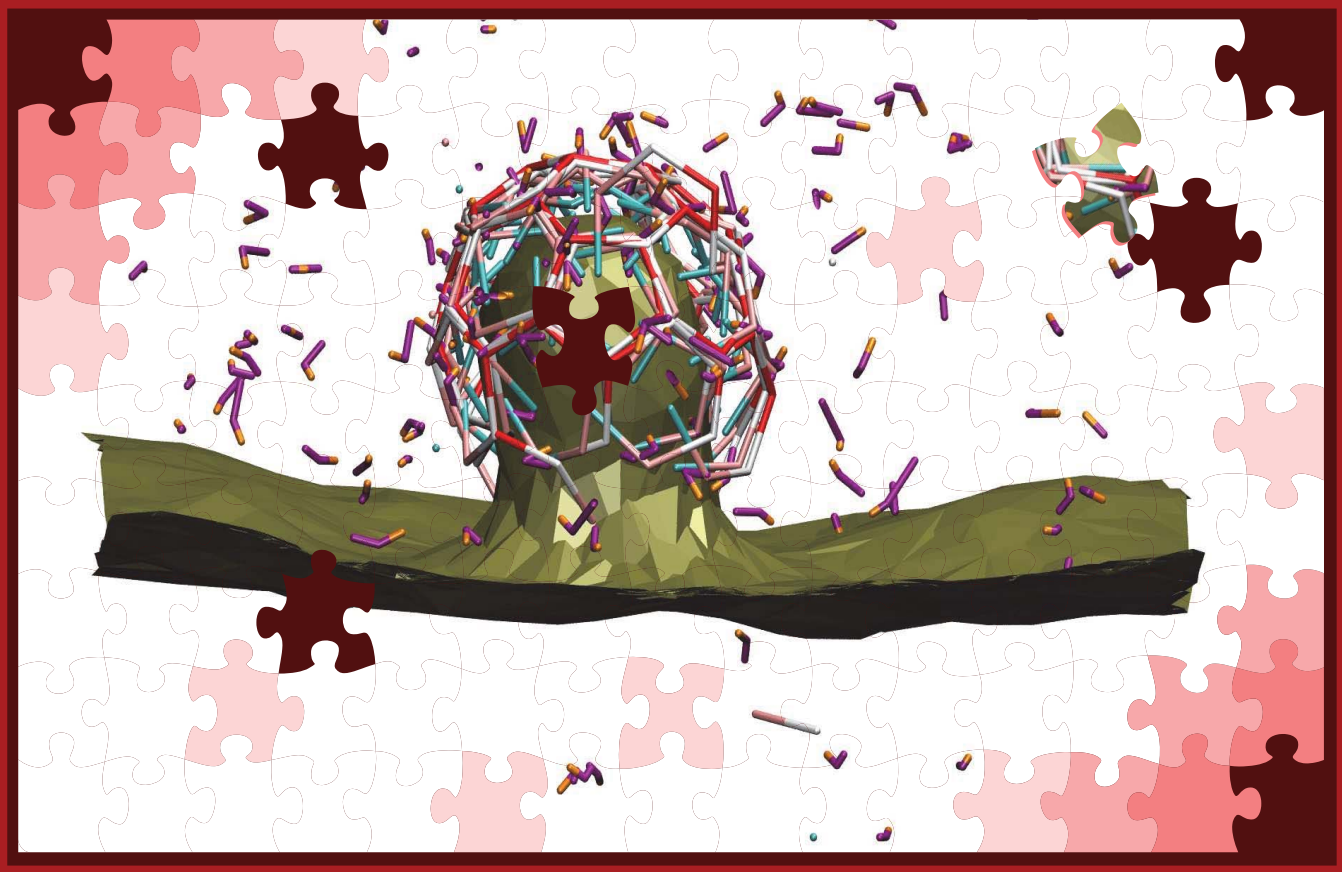
Matteo Giani



Modeling the self-assembly of clathrin coats

Modeling the self-assembly of clathrin coats

Dissertation

to obtain the degree of doctor at the University of Twente, on the authority of the rector magnificus, Prof. dr. T.T.M. Palstra, on account of the decision of the graduation committee, to be publicly defended on Wednesday, July the 19^{th}, 2017 at 16:45.

by

Matteo Giani

born on October the 1^{st} 1987, in Monza, Italy.

This doctoral dissertation is approved by:

Prof. dr. W. J. Briels	Promotor
Prof. dr. rer.-nat. S. Luding	Promotor
Dr. Ir. W.K. den Otter	Assistant Promotor

ISBN: 978-90-365-4348-4
DOI: 10.3990/1.9789036543484
Cover Art: A clathrin coated vesicle, rendered in VMD
Typeset by LaTeX

Members of the commitee:

Chairman:	Prof. dr. G. P. M. R. Dewulf	Universiteit Twente
Promotor:	Prof. dr. W. J. Briels	Universiteit Twente
Promotor:	Prof. dr. rer.-nat. S. Luding	Universiteit Twente
Ass. Promotor:	Dr. ir. W.K. den Otter	Universiteit Twente
Com. Members:	Prof. dr. ir. M.M.A.E. Claessens	Universiteit Twente
	Prof. dr. ir. J. Huskens	Universiteit Twente
	Prof. dr. F. Schmid	Johannes Gutenberg Universität
	Prof. dr. ir. N.F.A. van der Vegt	Technische Universität Darmstadt
	Prof. dr. P. Bassereau	Institut Curie, CNRS & Sorbonne Univ.

This work is part of the research programme 'Self-assembly of protein coats at membranes' (project nr. 711.012.004) which is financed by the Netherlands Organisation for Scientific Research (NWO). The research described in this thesis was performed using the computational resources of the Computational Biophysics (CBP) group within the MESA+ Institute for Nanotechnology of the University of Twente.

Summary

The assembly of clathrin coats in the presence of adaptor proteins was studied through computer simulations using coarse-grained models and through statistical mechanics. Adopting a reductionist approach based on recent experimental results, we aimed at reproducing and studying the minimal conditions that lead to the successful formation of aggregates, and at investigating the molecular properties and mechanisms required by the assembly process both in bulk conditions and at a membranous surface. In order to tackle this challenging task, coarse-grained models were used to describe all the assembly units involved in the simulations presented in this thesis. These models are based on the available structural data and are engineered to capture the key elements and behavior of the modeled proteins.

In Chapter **??** we introduce a coarse grained model of adaptor proteins, inspired by and representing the AP2 complex. The latter, the second most abundant component of endocytic coats after clathrin, is known to play a fundamental role in promoting and assisting the creation of coats at the cytosolic surface of the membrane. It is reported to be able to trigger polymerization of clathrin triskelia in physiological conditions of salt and pH, under which purified clathrin triskelia do not spontaneously self-assemble. The interaction between APs and clathrin were modeled throughout this thesis through a click potential, introduced for the first time in this chapter. The characteristics of the AP model, and of this interaction, have been tuned to reproduce the existing experimental assembly data of an AP2 and clathrin mixture. Our computer simulations provide novel insights into the role of AP2 in the self-assembly of clathrin cages and suggest that the mechanical properties of adaptor proteins are of fundamental importance. In the same chapter, we also developed a statistical mechanical theory that describes the equilibrium concentration of clathrin cages as a function of the other assembly variables and parameters, such as the protein concentrations and interaction strengths.

This theoretical model has been further developed in Chapter **??**, in order to explicitly take into account the effect of the flexibility of the clathrin triskelion, previously neglected. The main aim of the chapter is to investigate the equilibrium properties of clathrin cages resulting from the aggregation process, with emphasis on their size in the absence and in the presence of adaptor proteins. In order to perform this study, the essential features and characteristics of clathrin and AP2s are captured through a small number of effective parameters, and the number of allowed aggregates is determined on the base of geometrical considerations and arguments. The model is able to capture the key mechanisms determining the experimentally known ability of AP2s to influence the size of a clathrin cage, and thus to shape the resulting cage size distribution.

In Chapter **??**, we introduce a triangulated mesh model for an elastic membrane to investigate the formation of clathrin/AP2 coats at the cytosolic face of a cellular membrane. The model parameters are tuned to reproduce the typical properties of a biological membrane within the computational limits imposed by our simulations. In order to be able to extract the dynamical behavior associated to the aggregation process from the simulations, we make use of a compact Rotational Brownian Dynamics algorithm that uses quaternions to describe rotations, recently developed within the group. In the same spirit that led to the development of the statistical-

mechanical model accompanying the simulations of Chapter **??**, we developed a Langmuir-like adsorption model for the clathrin/AP complex at the membrane. Through the combination of simulations and theory, we characterize the mechanisms by which an initial nucleation point constituted by a small number of assembly units is stabilized through a cooperative effect between APs and clathrin at the membrane surface. We furthermore describe and predict the conditions under which this nucleation point is able to grow into a hemispherical clathrin coat.

In all the simulations performed in Chapter **??**, the growth of the clathrin coat halts upon reaching a hemispherical configuration, hinting towards the existence of an activation barrier in the free energy profile associated with the assembly of a clathrin coat at the membrane. Chapter **??** is devoted to investigating and computing the free energy profile by means of constrained Brownian Dynamics simulations. The free energy is here expressed and computed as a function of a reaction coordinate by integrating the average constraint force. Our results confirm the existence of a free energy barrier, implying the action of other endocytic components, possibly other membrane-bending proteins, at a specific step of the assembly process.

MODELING THE SELF-ASSEMBLY OF CLATHRIN COATS

Dissertation

to obtain the degree of doctor at the University of Twente, on the authority of the rector magnificus, Prof. dr. T.T.M. Palstra, on account of the decision of the graduation committee, to be publicly defended on Wednesday, July the 19^{th}, 2017 at 16:45.

by

Matteo Giani

born on October the 1^{st} 1987, in Monza, Italy.

This dissertation is approved by:

Prof. dr. W. J. Briels	Supervisor
Prof. dr. rer.-nat. S. Luding	Supervisor
Dr. Ir. W.K. den Otter	Co-Supervisor

ISBN: 978-90-365-4348-4
DOI: 10.3990/1.9789036543484
Cover Art: A clathrin coated vesicle, rendered in VMD
Typeset by LaTeX

Graduation committee:

Chairman:	Prof. dr. G. P. M. R. Dewulf	Universiteit Twente
Supervisor:	Prof. dr. W. J. Briels	Universiteit Twente
Supervisor:	Prof. dr. rer.-nat. S. Luding	Universiteit Twente
Co-Supervisor:	Dr. ir. W.K. den Otter	Universiteit Twente
Members:	Prof. dr. ir. M.M.A.E. Claessens	Universiteit Twente
	Prof. dr. ir. J. Huskens	Universiteit Twente
	Prof. dr. F. Schmid	Johannes Gutenberg Universität
	Prof. dr. ir. N.F.A. van der Vegt	Technische Universität Darmstadt
	Prof. dr. P. Bassereau	Institut Curie, CNRS & Sorbonne Univ.

This work is part of the research programme 'Self-assembly of protein coats at membranes' (project nr. 711.012.004) which is financed by the Netherlands Organisation for Scientific Research (NWO). The research described in this thesis was performed using the computational resources of the Computational Biophysics (CBP) group within the MESA+ Institute for Nanotechnology of the University of Twente.

Contents

Chapter 1

Introduction

Clathrin-mediated endocytosis is a process by which eukaryotic cells internalize cargo molecules through the formation of clathrin-coated lipid vesicles. The assembly of this protein coat in a polyhedral lattice on the cytosolic face of the plasma membrane requires the interplay between clathrin, the major component of the coat, and a number of other protein complexes, among which adaptor proteins, that mediate the interaction with the lipid membrane and cargo molecules. The aim of this chapter is to provide a brief overview of the fundamental biological aspects that inspired our models and of the computational methods used throughout this thesis to study the mechanisms of the assembly process.

1.1 Clathrin Mediated Endocytosis

1.1.1 Introduction

In all living cells, the lipid bilayer constituting the plasma membrane defines, separates, and preserves the intracellular cytoplasm from the extracellular environment, by forming a continuous barrier surrounding the entire cell volume. Membranes are also observed surrounding separate cellular compartments and organelles - a distinctive feature of eukaryotic cells.

The lipids [1] forming the membrane bilayer (phospholipids, glycolipids, and cholesterol) are amphiphilic and are organized in the bilayer in such a way to point their polar ends towards the exterior. Thanks to this structure, hydrophilic molecules cannot easily cross the membrane hydrophobic bilayer core. Membranes thus act as an effective barrier, able to maintain a concentration gradient between two separate compartments, while at the same time measuring only a few nanometers in width.

Several transport processes across the membrane exist, characterized by different types of mechanisms and molecules involved. The transport rate of small molecules across the membrane is dictated by the organization and composition of the membrane lipids [1], and by the proteins embedded in the lipid matrix that regulate the permeability of the cell membrane. Some notable examples of transport proteins are ion channels, able to create pores in the membrane and thus regulate the flow of ions across the membrane, or ion transporters, proteins able to move ions against their concentration gradient. Larger molecules such as proteins, that cannot cross the membrane through passive means, are typically collected, sorted, and internalized in lipid vesicles, used as transportation devices between membrane-defined cellular compartments.

A particularly important example is the endocytic pathway. Endocytosis is a fundamental internalization process of cargo molecules and lipids at the plasma membrane, observed in all eukaryotic cells. Lipid vesicles bud from the cytosolic side of the donor membrane, finally detaching from it to form a free vesicle, able to diffuse in the cytosol (possibly moved by molecular

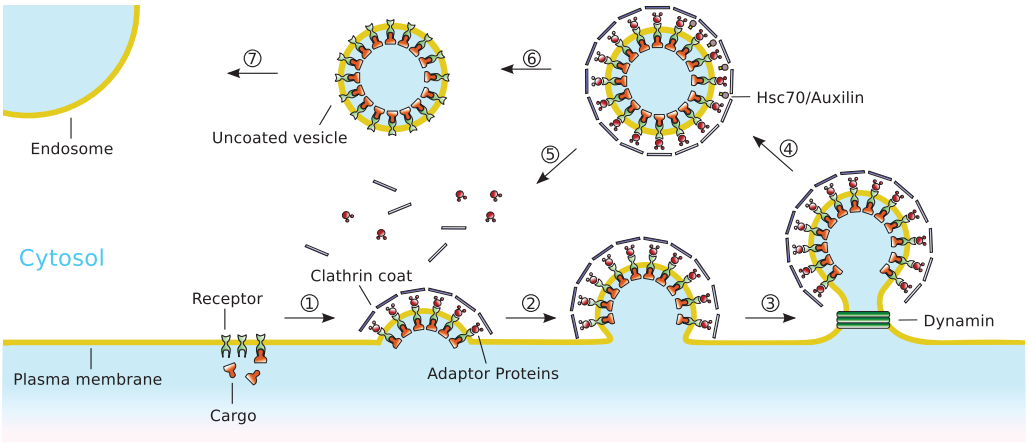


Figure 1.1: An overview of the main steps in the *nucleation*, *budding*, *scission*, and *uncoating* of a clathrin coated vesicle during endocytosis at the cellular membrane, shown counterclockwise in chronological order. An initial nucleation point is formed on the cytosolic side of the plasma membrane through a cooperative interplay involving cargo molecules, able to influence adaptor localization. ① Adaptors recruit clathrin triskelia that are diffusing in the cytosol. ②, ③ The protein lattice grows by further clathrin polymerization and bends the underlying membrane, causing cargo sequestration and sorting. ④ Membrane scission is mediated by dynamin rings around the neck of the clathrin coated vesicle, resulting in a free clathrin coated vesicle. ⑤ Clathrin triskelia are uncoated through the action of Hsc70 and auxilin in an ATP dependent process, followed by adaptors dissociation from the membrane in a second separate step. Adaptors in the cytosol are phosphorylated, to make them unable to trigger clathrin polymerization in the bulk phase. ⑥ Free clathrin and APs are recycled in the next endocytic event. ⑦ The uncoated vesicle will diffuse in the cytosol and finally fuse with an acceptor endosome.

motors, or along tracts of microtubules). One of the elements involved in targeting specific membranes is a class of protein known as SNAREs. Vesicles are finally able to deliver their engulfed cargo by fusing with a target receiving membrane, often through a protein complex [1]. The engulfed cargo is typically delivered to endosomes, where sorting occurs, and then can be redirected to late endosomes and lysosomes for degradation, to the trans-Golgi network (TGN), or to recycling carriers that bring the cargo and lipids back to the plasma membrane (this reverse process is called *exocytosis*). The endocytic pathway also provides a way to control and regulate the composition of the plasma membrane, an important factor for processes like cytokinesis, cell adhesion and morphogenesis, or cell fusion. It is in fact estimated that some cells internalize their cell surface equivalent one to five times per hour [2].

In general, the different endocytic pathways [3] can be grouped in two main groups, namely clathrin-dependent and clathrin-independent endocytosis. The latter group, which existence has been proven only in relatively recent years [4], includes caveolae-dependent endocytosis [5] and macropinocytosis [6]. In the following sections we will focus on clathrin-mediated endocytosis, the most important and perhaps best understood endocytic route and the object of study of this thesis.

[1] The 2013 Nobel prize in Physiology or Medicine was assigned to James E. Rothman, Randy W. Schekman and Thomas C. Südhof for their studies on the organization of the cellular transport system.

1.1.2 The life cycle of a clathrin coated pit

The initial observation of 'bristle-coated' pits [7] in 1964, and the recognition [8] of their role in clathrin-mediated endocytosis (CME) by Barbara Pearse in 1976, created an evolving paradigm for the analysis of cellular trafficking, and founded an entire new interdisciplinary research field. In CME [9–12], cargo molecules such as transferrin, immunoglobulins, lipoproteins, hormones, and signaling receptors are transported by vesicles surrounded by a protein coat, which most abundant component is the clathrin protein [13]; for this reason, they are referred to as *clathrin coated vesicles* (CCVs). Clathrin triskelia do not directly bind the lipid membrane or cargo proteins. These interactions are mediated by a family of protein complexes called adaptor proteins (APs), that also constitute a binding hub for other endocytic proteins and complexes [14, 15]. The entire life cycle of a clathrin-coated vesicle, summarized in Fig. 1.1, involves a sequence of highly regulated events at a time scale of seconds to minutes. An initial, coated pit assembles at the cytosolic face of a cellular membrane, see Fig. 1.2, by growth of a nucleation point, composed of a small clathrin lattice anchored by adaptor proteins to the underlying membrane [16]. This initial nucleation point grows by additional clathrin polymerization, deforming the underlying membrane, and finally leading to a coated vesicle, attached to the plasma membrane by a narrow neck. The fully formed vesicle pinches off to form a free coated vesicle through the action of dynamin. The latter forms rings or spirals around the neck of the vesicle, causing it to eventually break and release the vesicle. Experiments *in vivo*, using among others live-cell fluorescent microscopy through total internal reflection fluorescence (TIRF) or confocal microscopy, revealed a number of proteins associated with the nascent coat at different stages of maturation with different functional tasks. Eps15, epsin, FCHo1 and FCHo2 form a complex required at a very initial stage for the initiation of the coated pit [17–21]. Amphiphysin and endophilin are thought to contribute to the maturation of a clathrin coated vesicle by means of their membrane bending BAR domain [22]. Actin, indirectly recruited by clathrin through binding with Hip1R, is implicated in the process of coat maturation [20, 23]. Other components are necessary for binding to particular cargo molecules [24]. The ATP-dependent disassembly of the outer layer is triggered by auxilin, that in turn recruits Hsc70 to direct uncoating [25–27].

Despite this high number of proteins and complexes directly or indirectly involved in the creation and maturation of CCVs, recent experiments *in vitro* showed that purified clathrin, in the sole presence of a disordered polypeptide acting as an adaptor protein, is capable of generating spherical buds. In the presence of dynamin, these buds are released from the membrane, suggesting that these three endocytic components possess the key elements leading CCV assembly [28]. In this thesis, we adopted a similar reductionist approach to describe the physical mechanisms and conditions leading to the assembly process of a CCV, by identifying the fundamental characteristics and functional roles of each of the main endocytic players.

Coat components

A clathrin protein possesses a characteristic shape that resembles that of a pinwheel or triskelion, visible in Fig. 1.3. This triskelion represents the basic lattice unit for a vesicular coat [29–32]. Each of the three legs is composed of a 190 kDA clathrin heavy chain (CHC) subunit, and posses a contour length of 45 nm. The three legs are joined at a central hub near their C-terminus. Each leg is subdivided in segments of uniform thickness, referred to as 'proximal', 'distal', and 'terminal domain', connected and smoothly bent at a 'knee', 'ankle', and 'linker'. The globular N-terminal domain consists of a seven-bladed β propeller structure, while the remaining domains are characterized by 8 copies of a large repeating motif of about 145 residues in five helical zigzags, known as the clathrin heavy-chain repeat (CHCR) [33]. Associated with

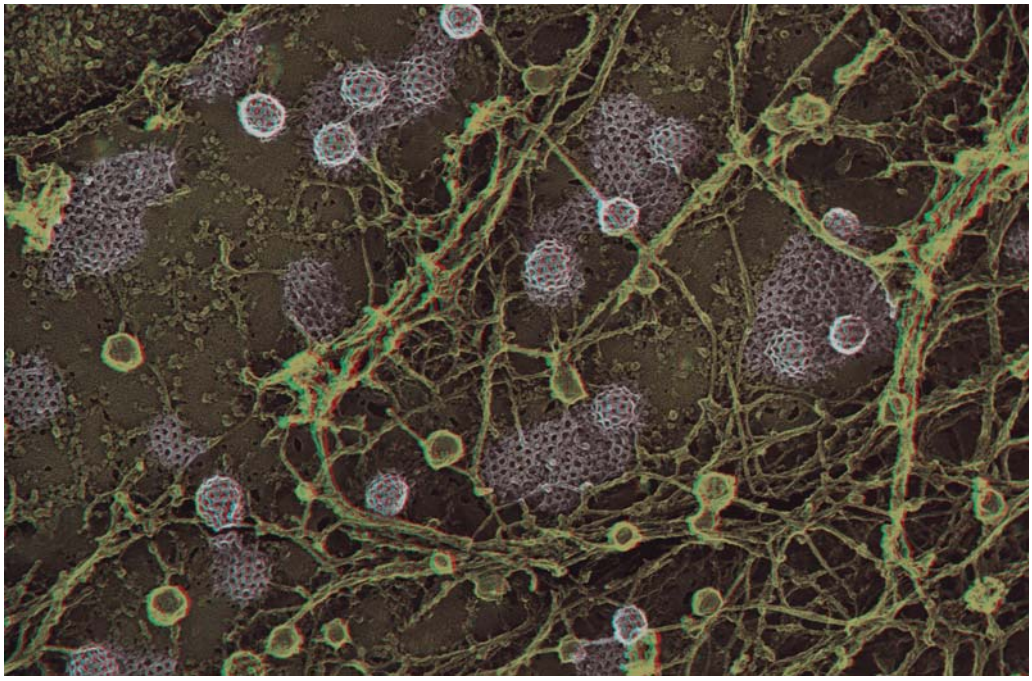


Figure 1.2: Clathrin pits of different sizes, formed at the cytoplasmic face of the plasma membrane. These pits are regions of the donor membrane where the assembly of the vesicle coat takes place, and they bud to form a coated vesicle. Electron micrograph courtesy of Dr. John Heuser.

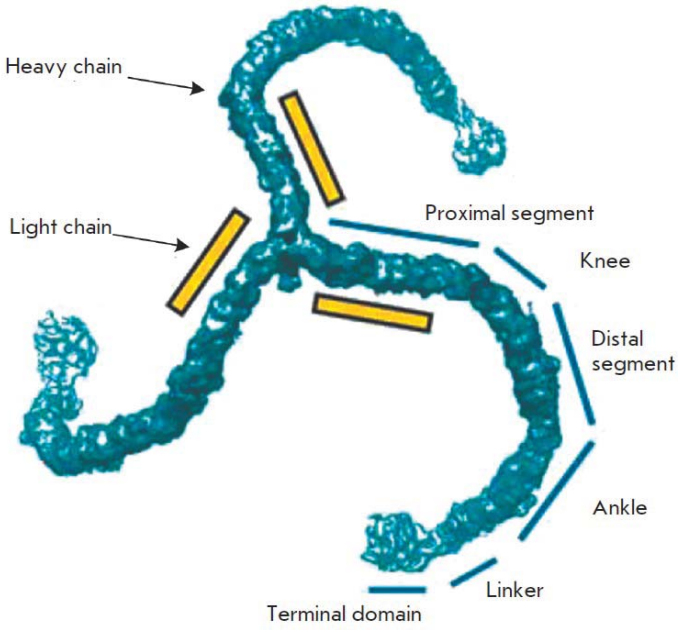


Figure 1.3: The clathrin triskelion is formed by three identical clathrin heavy chains forming the backbone of a very elongated and curled structure with a contour length of about 45 nm, subdivided in the segments labeled in the figure. The legs are connected at their C-terminus near the hub, while the N-terminus of the chain is the terminal domain. The position of the three clathrin light chains associated to each leg is indicated schematically. Reproduced with permission from Fotin *et al.* [33].

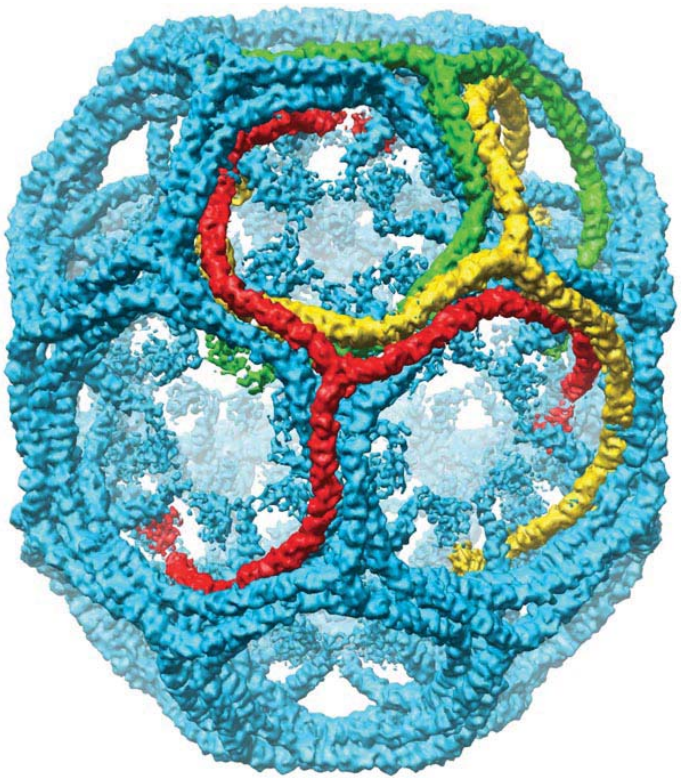


Figure 1.4: Hexagonal clathrin barrel obtained at subnanometric resolution. Only the heavy chains of clathrin are indicated. The structure is constituted by 36 clathrin and has D6 symmetry. Reproduced with permission from Fotin *et al.* [33].

every CHC proximal segment is a smaller 23 kDA clathrin light chain (CLC) subunit with a simpler α-helical heptad repeat, which is thought to regulate the clathrin assembly process [34] at physiological conditions. Without CLC, triskelia of CHC polymerize into a lattice at physiological pH [35]. The overall clathrin structure is characterized by an intrinsic curvature, that plays an important role in the coat ability of bending the membrane.

The AP2 complex belongs to the heterotrameric family of AP1-4, and probably represents the most studied adaptor protein. AP2 mediates endocytosis at the plasma membrane, it is the most abundant endocytic clathrin adaptor, and it covers a fundamental role in coat assembly under physiological conditions [36]. AP2 consists of the α_2 and β_2 large subunits of 100-110 kDa together with a medium sized 50 kDa μ_2 subunit and a small 20 kDa σ_2 subunit. AP2 is characterized by a core that binds to a component of the cell membrane (PtdIns(4,5)P_2) and peptide motifs present in the cytosolic tail of membrane-embedded cargo receptors [37]. The C-terminal appendages of α_2 and β_2, connected to the core by flexible linkers [38], possess binding motifs for many other endocytic proteins with various accessory and regulatory functions [39].

Purified clathrin triskelia are able *in vitro* to self-assemble into polyhedral cages in slightly acidic conditions [38], without further assistance. The geometry and topology of these structures closely resemble those of the lattices constituting the coats around membranous vesicles. The protein coat surrounding the lipid vesicle forms a remarkably well defined polyhedral structure, see Fig. 1.5, featuring hexagonal and pentagonal facets, [2] first described by Kanaseki and Kadota

[2]Curiously, identical geometrical structures are found on different scales in nature. Examples include fullerenes

[40] in 1969. Cryo-EM maps of a polyhedral clathrin lattice, assembled with the adaptor protein complex AP2, were presented in 1998 and 2004 by Smith [41] and Fotin [33] at 2.1 nm and subnanometer resolution, respectively. The latter is visible in Fig. 1.4. In these structures, a clathrin hub is found at each lattice vertex, and each edge consists of two intertwined antiparallel proximal domains beneath which lie two anti parallel distal domains, all leg sections coming from different triskelia. The length of an edge is approximately 17 nm. The clathrin terminal domains project inwards, towards the center of the structure.

As a final remark, we observe that clathrin is not the only protein able to form coats of a transport vesicle. Other important examples are the COating Protein (COP) I and II [42–44], involved in cargo transport from the endoplasmatic reticulum to, and within, the Golgi apparatus. The characteristics and functions of the polyhedrical protein coat closely resemble those of a clathrin coat [45], but their assembly does not require APs.

1.2 Computer Simulations

1.2.1 Introduction

Computer simulations represent an increasingly popular and effective method for both research and industrial applications, and are a very convenient way to study and predict the statistical properties and dynamical behavior of complex macromolecules. They present themselves as a flexible, complementary method bridging the gap between experiments and analytical models, providing an adaptable framework both for proving theoretical predictions and for simulating experiments by allowing complete control over the details of the system under investigation. A subset of biological applications includes protein folding, structural predictions, ligand docking, viral assembly, or DNA supercoiling. Simulation results are usually re-expressed and interpreted through the language of statistical mechanics, that provides the rigorous mathematical expressions that allow to explore and extract the macroscopic properties of interest from the microscopic details of the system. These quantities are obtained through averages over all the frames in the trajectory that, for very long simulations, replaces averages over an equivalent statistical ensemble. In formulas,

$$\langle A \rangle = \int A(\{\mathbf{p}^i, \mathbf{r}^i\})\rho(\{\mathbf{p}^i, \mathbf{r}^i\})d\{\mathbf{r}^i, \mathbf{p}^i\} = \lim_{\tau \to +\infty} \frac{1}{\tau} \int_0^\tau A(\{\mathbf{p}^i(t), \mathbf{r}^i(t)\})dt = \bar{A}, \tag{1.1}$$

where A is the macroscopic property, $\{\mathbf{p}^i, \mathbf{r}^i\}$ is the complete set of coordinates and conjugate momenta for all the particles in the system, $\{\mathbf{p}^i(t), \mathbf{r}^i(t)\}$ a particular realization in a trajectory at time t, and ρ represents the probability density of the statistical ensemble. Equation (1.1) is known as the ergodic theorem for a random process.

It is beyond the scope of this introduction to provide a description of the theoretical aspects of computer simulations; the interested reader can refer to the excellent existing literature [46–48] for further details. In the following sections, the main aspects and simulation techniques used in this thesis are briefly introduced.

1.2.2 From molecules to computational models

Computer simulations are based on mathematical models that capture the behavior of the modeled system, often based on the ever-increasing number of available structural data for the macromolecule [3]. The level of accuracy of these models is limited by the available computational

or protozoa skeletons, 100 times smaller and 1000 times larger than clathrin cages.

[3] 126006 total structures at the Protein Data Bank at the end of 2016.

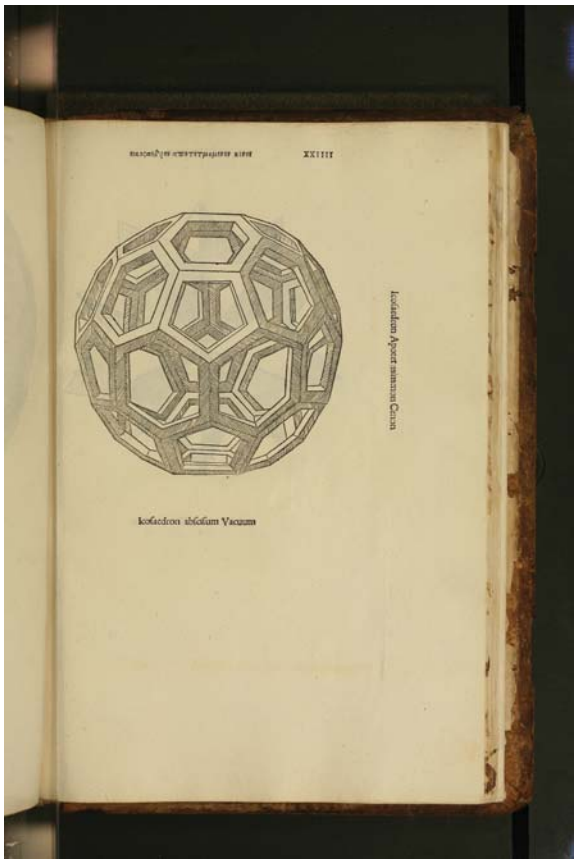


Figure 1.5: One of Leonardo da Vinci's illustrations for Luca Pacioli's 1509 book *De Divina Proportione* (Leonardo da Vinci, public domain, via Wikimedia Commons). The term *Icosaedron absisum* means truncated icosahedron, and the term *Vacuum* refers to the hollow faces. This 'soccer ball' structure, a common result of clathrin assembly, features 60 vertices forming 12 pentagonal and 20 hexagonal facets.

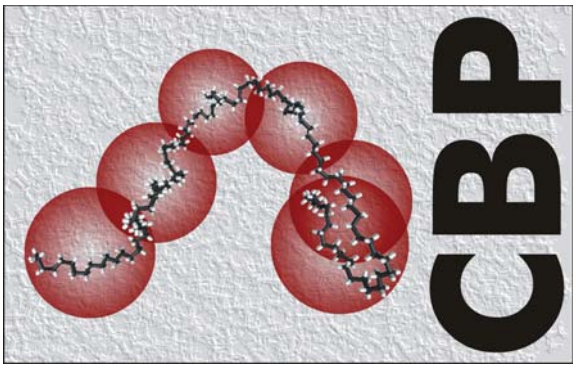


Figure 1.6: The logo of the Computational BioPhysics group, outlining the coarse graining procedure. Individual atoms in the polymer (white and black beads) are grouped together in the coarse grained model (red spheres), thus greatly reducing the total number of degrees of freedom. These new groups are the fundamental units of the coarse-grained simulations. The procedure allows to simulate much longer time and larger space scales.

resources. In fact a low-resolution, coarse-grained (CG) description of biomolecules is often the only way to simulate processes on a large time or length scale, that would be inaccessible to all-atom simulations.

In CG models, an effective interaction site or 'bead' represents a group of atoms, and the overall number of degrees of freedom is drastically reduced, as illustrated in Fig. 1.6. Thereby a vast amount of local patterns of motion and their short relaxation timescales can be excluded from consideration, making the simulation of a CG system significantly less resource and time demanding. For this reason, during the last decade multiscale CG simulations have gained an increasingly important role in the studies of long time or large scale phenomena, and are presently often used as components of multiscale modeling protocols in combination with, or in substitution of, atomistic resolution models [4].

All the proteins and macromolecules involved in the simulations presented in this thesis are described using highly coarse grained models. In particular, the clathrin model is directly derived on the basis of the model developed by den Otter and Briels [49–52]. The details of the model are presented in the following chapters.

1.2.3 Simulation algorithms

In recent years, thanks to the increased availability and interest for computational methods and tools, simulations found new applications in several fields and the number of simulation techniques has consequently greatly increased. In the realm of biological macromolecules, many specialized techniques for specific problems have been developed, including multiple scales simulations and mixed quantum-classical simulations. All the computer simulations in this thesis are performed under the assumption that, given the time and length scales associated to the biological system at study, classical mechanics can be used to describe its motion, interactions, and properties.

The methods employed in this thesis, and briefly introduced in the following sections, belong to the family of Monte Carlo (MC) computational methods and Brownian Dynamics (BD) simulations. The basic idea, in both cases, is to generate a discrete trajectory by evolving a specified

[4] In 2013, the Nobel Prize in Chemistry was awarded to M. Levitt, A. Warshel, and M. Karplus for the development of multiscale models for complex chemical systems.

initial state through an algorithm, and to compute quantities of interest through Eq. (1.1). A typical Monte Carlo or Brownian Dynamic simulation is performed in the canonical (NVT) ensemble; in order to sample a different statistical ensemble, the BD simulations in Chapters 4, 5 were performed by coupling the system to an external MC chemostat and barostat. The simulated system is confined in a cubic box. In order to avoid boundary problems in finite simulation boxes, we chose boundary conditions that mimic the presence of an infinite bulk around the simulated particles by using periodic boundary conditions (PBC). The box is surrounded by a number of identical copies, and the computations make use of the Minimum Image Convention, that states that each particle can only interact with the closest copy of every other particle. Finally, the box size was chosen to avoid unphysical artifacts linked to the periodicity of the images or to the isotropy of the system.

Monte Carlo

The Monte Carlo algorithms used in this thesis belong to a much larger class of computational statistical methods aimed at estimating numerically quantities that, even having an analytic expression, are in practice very difficult to obtain. They are as such employed in the fields of numeric integration or optimization, and become particularly useful when the number of dimensions involved in the problem is very high [53]. Another physical application, used in this thesis, is the sampling of the Boltzmann equilibrium probability distribution for a system characterized by a high number of coupled degrees of freedom:

$$\begin{aligned} P(\epsilon) &= \frac{1}{Z} g(\epsilon) e^{-\beta\epsilon} \\ Z &= \sum_{\epsilon} g(\epsilon) e^{-\beta\epsilon}. \end{aligned} \tag{1.2}$$

In these equations, $P(\epsilon)$ is the probability that the system is in a state with energy ϵ, $g(\epsilon)$ is the number of those states, $\beta = 1/k_B T$ with k_B Boltzmann constant and T the system temperature, and Z is known as the partition function. For large macromolecules, the energy of a configuration is a complex function of the set of coordinates, and the number of configurations is extremely high, thus making the direct calculation of Z extremely inconvenient, if not impossible.

The MC simulations used in this thesis consist in generating a series of independent configurations of the system, distributed according to this desired distribution, known as a Markov Chain. Individual frames are then used as samples of the distribution, and quantities can be computed accordingly. Even within the subfield of Markov Chain Monte Carlo (MCMC), there exists a number of general or specialized algorithms. In this thesis, we make use of the Metropolis iterative scheme. The system is initially prepared in a configuration $\mathbf{x}_i$. The algorithm consists then of the following steps:

- a new configuration $\mathbf{x}_{i+1}$ is chosen according to a (symmetrical) trial probability distribution $P^{trial}_{\mathbf{x}_i \to \mathbf{x}_{i+1}}$;
- the energy of the new configuration, $E(\mathbf{x}_{i+1})$, and the ratio $\alpha = e^{-\beta E(\mathbf{x}_{i+1})}/e^{-\beta E(\mathbf{x}_i)} = e^{-\beta \Delta E}$ are computed;
- the probability of accepting the move is given by $P^{acc}_{\mathbf{x}_i \to \mathbf{x}_{i+1}} = \min(1, \alpha)$.

The procedure is then repeated starting from the configuration selected in the last step. The total number of iterations determines the accuracy by which the Boltzmann distribution is sampled. The exact details of the trial move are discussed in the method sections of their

respective chapters. Here, we note that the 'dynamics' observed in MCMC simulations is a mere consequence of the particular choice of the trial move distribution, that in turn is often chosen to make the sampling efficient. As such, it usually does not reproduce the actual, real dynamics and it is not straightforward to recover time scales from the trajectories. In this context, a common problem observed in MC simulations is the sudden drop of the diffusion coefficient ('freezing') of particles when aggregates are formed. A possible solution to this problem is the introduction of cluster moves, introduced in Chapter 2 for the AP/clathrin complex.

The Langevin equation and the Brownian Dynamics algorithm

The core component of a Brownian dynamics simulation is the Langevin equation of motion [54], probably the best known example of a stochastic differential equation. The equation is integrated forward in time to create discrete trajectories over a 'small' timestep δt, which range will be specified shortly. The Langevin equation is used to simulate the irregular motion of small particles in a solution, named *Brownian motion* after the botanist Robert Brown, who first described it in 1827. As an illustrative example, in this section we will consider the translational motion of a particle, described in terms of the coordinates $\mathbf{x}$ of the particle's center of mass. In its motion through the solvent, the particle continuously experiences collisions with the liquid molecules. The latter ones, of no special interest in this context, are not explicitly considered; rather, the collective average effect of the collisions is represented by a friction term and the fluctuations around the average by a random contribution to the equation of motion of the particle,

$$m\frac{d\mathbf{v}}{dt} = \mathbf{f} - \gamma\mathbf{v} + \mathbf{\Theta}, \tag{1.3}$$

where m is the particle mass, γ is its friction constant, often estimated through Stokes' equation, $\mathbf{f}$ is the force acting on the particle, and $\mathbf{\Theta}$ is a stochastic force that accounts for the random scattering with the solvent. The time scale associated to these collisions is assumed to be much shorter than that describing the motion of the particle. Thus, at every instant, the particle experiences many scatterings and, upon assuming each scattering event to be independent, the resulting distribution of $\mathbf{\Theta}$ is taken to be Gaussian. By the fluctuation dissipation theorem, there exists a relation between the strength of stochastic force fluctuations and friction. Assuming isotropy, the average and variance of the distribution are given by:

$$\begin{aligned} \langle\mathbf{\Theta}\rangle &= 0, \\ \langle\mathbf{\Theta}(t)\mathbf{\Theta}(t')\rangle &= 2\gamma k_B T\delta(t-t'), \end{aligned} \tag{1.4}$$

where k_B is Boltzmann constant and T is the system temperature Upon solving Eq. (1.3), one easily finds that average and square average velocity of a particle not subject to an external force $\mathbf{f}$, *i.e.* a Brownian particle, become exponentially independent of its initial velocity with a time constant given by γ/m. On a much bigger time scale, *i.e.* in the *overdamped* case, accelerations can thus be neglected and inertia plays no role. The equation of motion becomes in this limit particularly simple:

$$\mathbf{v} = \frac{d\mathbf{x}}{dt} = \frac{\mathbf{f}+\mathbf{\Theta}}{\gamma} = \frac{\mathbf{f}D}{k_B T} + \mathbf{\Theta}', \tag{1.5}$$

where D is the diffusion coefficient of the particle. Unlike the original Langevin equation, it is now a first-order equation that does not depend anymore on the mass of the particle. Brownian Dynamics algorithms consist of a numerical scheme to integrate this overdamped Langevin equation of motion for a large system of N particles. Only the positions are propagated in time, often enabling the use of large time steps limited by the condition that the conservative forces

f acting on the particle must vary only slightly over the course of a single step. The numerical integrator is:

$$\Delta \mathbf{x} = \mathbf{f}\frac{D}{k_B T}\delta t + S(\delta t), \tag{1.6}$$

where $S(\delta t)$ has a variance of $2D\delta t$. The displacement of a particle undergoing pure Brownian motion can in fact be obtained by solving the equivalent Fokker-Planck or Smoluchowski diffusion equation, under appropriate boundary conditions. The displacement is then found to vary with the square root of the time. In this introduction, only the translational degrees of freedom of the particle have been considered, and the particle orientation has been neglected. Traditional Rotational Brownian Dynamics (RBD) algorithms are characterized by both theoretical and numerical difficulties, associated with the use of coordinate systems such as Euler angles. In chapter Chapter 4 and Chapter 5, we make extensive use of a remarkably stable and compact RBD algorithm based on quaternions [52], developed by the CBP group to simulate anisotropic particles. The algorithm structure, presented in the appropriate chapters, resembles that of Equation (1.6).

1.2.4 Simulation code

A quite substantial, yet perhaps not much evident, part of this thesis has been the implementation and testing of the CPU intensive simulation code. The framework of the Monte Carlo and Brownian Dynamics simulations was designed and implemented from scratch in C++, making use of several libraries from the Boost collection [5]. Besides the standard algorithms and numerical recipes described in the literature [46, 48], the OpenMP library [6] was used to optionally parallelize CPU demanding, suitable calculations loops over several processors. The code for the statistical mechanical calculations was written in FORTRAN [55].

The simulations were run on the computational clusters 'MrFox' and 'Snowwhite' consisting respectively one 'master' and 14 and 18 'computational' nodes, for a grand total of 468 cores. The latter cluster was installed and administered by the author.

1.2.5 Previous results

The highly ordered geometrical structure of clathrin coats makes their assembly process highly suitable to be studied through theories borrowed from other self-assembling systems, *e.g.* micelles, and of course computer simulations.

Matthews and Likos modeled clathrin as a flexible triskelion composed of 13 bead particles with anisotropic interactions [56], and studied their self-assembly into cages and their ability to create membrane invaginations. Spakowitz and collaborators modeled clathrin as a particle that forms anisotropic harmonic bonds with three neighbors, to study the mechanical properties and fluidization of lattices against a membrane [57, 58]. Muthukumar and Nossal developed a micelle-like thermodynamic model to investigate the conditions under which clathrin triskelia form polyhedral baskets [59]. Roux and collaborators studied the role of membrane tension on the ability of clathrin to deform the membrane [60].

In earlier works, den Otter and Briels developed a highly coarse-grained patchy particle model representing a rigid clathrin triskelion, that constitutes the basis of the work done in this thesis. They studied the dependence of the assembly behavior of clathrin on the intrinsic curvature of the triskelion and on the environmental conditions, and showed that anisotropic leg-leg interactions are the key to self-assembly [49–51]. Simulations with this model predicted

[5] `http://www.boost.org/`
[6] `http://www.openmp.org/`

a binding energy of about 23 k_BT per clathrin in a cage. More recent simulations of the same model yielded an assembly dynamics compatible with experiments [52].

An explicit description of adaptor proteins was omitted in all these simulations, and their role was modeled by enabling clathrin particles to bind directly to the membrane.

1.3 Thesis Outline

Despite the enormous progress made since the first discovery of clathrin mediated endocytosis, a number of fundamental questions remain unresolved both on a basic and on a conceptual level. For instance, how does the clathrin coat assemble, and which of the many proteins involved in the endocytic cycle are essential for cargo internalization? Is the curvature built into the lattice from the very beginning of its assembly process, or is a reorganization of the lattice induced in a later stage? Is clathrin driving membrane curvature *in vivo*, or does it rather exploit the shaping mechanisms of other proteins?
In this thesis, we present the result of our study of the self-assembly process of clathrin coats through novel Brownian Dynamics and Monte Carlo simulations, and statistical physics, representing our contribution towards answering some of to the previous questions.

In Chapter 2 we study the assembly process of clathrin *in vitro* in polyhedral cages, aided and directed by AP2. The clathrin model by den Otter and Briels is here extended to include a leg segment involved in binding with AP2s. The AP subunits involved in clathrin binding are represented through two binding sites connected by a flexible linker, modeled as a random polymer. Based on the simulation results, we derive a statistical mechanical model describing the assembly process. Simulations and theory are found in good agreement with the available experimental data, and provide a novel framework able to explain the mechanisms leading to clathrin assembly.

In Chapter 3 the theoretical model developed in Chapter 2 is extended in order to account for clathrin coats of different compositions, taking into consideration the intrinsic curvature and flexibility of the clathrin triskelion. The model is used to fit the experimental coat size distributions in the presence and absence of APs, and provides an insight on the mechanisms enabling APs to influence the size of a clathrin cage.

In Chapter 4 we introduce a coarse grained model for a lipid membrane, based on a viscoelastic triangulated network, to study the essential conditions leading to the formation of a clathrin coated pit able to induce curvature on the cellular membrane. In line with the results from Chapter 2, the mechanical properties of AP2s are found to be of paramount importance for the assembly process and appear crucial to determine the correct orientation of the curved clathrin lattice relative to the membrane. We also develop a Langmuir-like absorption model to investigate the stability of a clathrin-AP complex at the membrane, that relies on a cooperative binding effect.

In Chapter 5 we prove the presence of a free energy activation barrier, that halts the spontaneous polymerization of clathrin coats past the invagination step studied in the previous chapter. We characterize the free energy profile relative to this process as a function of suitably chosen reaction coordinate through the use of the potential of mean constraint force (PMCF). In the last section of this chapter we briefly summarize possible future developments of this work.

At the end of the thesis, the main results are summarized in English and Dutch.

Bibliography

[1] G. Van Meer, D. R. Voelker, G. W. Feigenson, *Nature reviews Molecular cell biology* **2008**, *9*, 112–124.

[2] R. M. Steinman, I. S. Mellman, W. A. Muller, Z. A. Cohn, *J cell biol* **1983**, *96*, 1–27.

[3] S. Mayor, R. E. Pagano, *Nature reviews Molecular cell biology* **2007**, *8*, 603–612.

[4] K. Sandvig, B Van Deurs, *Journal of Biological Chemistry* **1990**, *265*, 6382–6388.

[5] I. R. Nabi, P. U. Le, *The Journal of cell biology* **2003**, *161*, 673–677.

[6] J. P. Lim, P. A. Gleeson, *Immunology and cell biology* **2011**, *89*, 836–843.

[7] T. F. Roth, K. R. Porter, *The Journal of cell biology* **1964**, *20*, 313–332.

[8] B. Pearse, *Proceedings of the National Academy of Sciences* **1976**, *73*, 1255–1259.

[9] M. J. Taylor, D Perrais, C. J. Merrifield, *PLoS Biol.* **2011**, *9*, e1000604.

[10] J Rappoport, *Biochem. J* **2008**, *412*, 415–423.

[11] T. Kirchhausen, D. Owen, S. C. Harrison, *Cold Spring Harbor perspectives in biology* **2014**, *6*, a016725.

[12] A. Young in Seminars in cell & developmental biology, *Vol. 18*, Elsevier, **2007**, pp. 448–458.

[13] B. M. Pearse, *Journal of molecular biology* **1975**, *97*, 93IN1197–96IN1398.

[14] L. M. Traub, J. S. Bonifacino, *Cold Spring Harbor perspectives in biology* **2013**, *5*, a016790.

[15] N. Popova, I. Deyev, A. Petrenko, *Acta naturae* **2013**, *5*, 62.

[16] M. A. Edeling, C. Smith, D. Owen, *Nature Reviews Molecular Cell Biology* **2006**, *7*, 32–44.

[17] W. M. Henne, H. M. Kent, M. G. Ford, B. G. Hegde, O. Daumke, P. J. G. Butler, R. Mittal, R. Langen, P. R. Evans, H. T. McMahon, *Structure* **2007**, *15*, 839–852.

[18] G Hollopeter, J. J. Lange, Y Zhang, T. N. Vu, M Gu, M Ailion, E. J. Lambie, B. D. Slaughter, J. R. Unruh, L Florens, E. M. Jorgensen, *Elife* **2014**, *3*, e03648.

[19] M. G. J. Ford, I. G. Mills, B. J. Peter, Y Vallis, G. J. K. Praefcke, P. R. Evans, H. T. McMahon, *Nature* **2002**, *419*, 361–366.

[20] S. Saffarian, E. Cocucci, T. Kirchhausen, *PLoS Biol* **2009**, *7*, e1000191.

[21] L Ma, P. K. Umasankar, A. G. Wrobel, A Lymar, A. J. McCoy, S. S. Holkar, A Jha, T Pradhan-Sundd, S. C. Watkins, D. J. Owen, L. M. Traub, *Developmental cell* **2016**, *37*, 428–443.

[22] H. T. McMahon, P. Wigge, C. Smith, *FEBS letters* **1997**, *413*, 319–322.

[23] S. Ferguson, A. Raimondi, S. Paradise, H. Shen, K. Mesaki, A. Ferguson, O. Destaing, G. Ko, J. Takasaki, O. Cremona, et al., *Developmental cell* **2009**, *17*, 811–822.

[24] S. E. Miller, D. A. Sahlender, S. C. Graham, S. Höning, M. S. Robinson, A. A. Peden, D. J. Owen, *Cell* **2011**, *147*, 1118–1131.

[25] E. Eisenberg, L. E. Greene, *Traffic* **2007**, *8*, 640–646.

[26] Y. Xing, T. Böcking, M. Wolf, N. Grigorieff, T. Kirchhausen, S. C. Harrison, *The EMBO journal* **2010**, *29*, 655–665.

[27] D. Owen, Y Vallis, B. Pearse, H. McMahon, P. Evans, *The EMBO journal* **2000**, *19*, 4216–4227.

[28] P. N. Dannhauser, E. J. Ungewickell, *Nat. Cell Biol.* **2012**, *14*, 634–639.

[29] F. M. Brodsky, *Annual review of cell and developmental biology* **2012**, *28*, 309.

[30] E. Ungewickell, D. Branton, *Nature Publishing Group* **1981**.

[31] J. D. Wilbur, P. K. Hwang, F. M. Brodsky, *Traffic* **2005**, *6*, 346–350.

[32] S. J. Royle, *Cellular and molecular life sciences* **2006**, *63*, 1823–1832.

[33] A. Fotin, Y. Cheng, P. Sliz, N. Grigorieff, S. C. Harrison, T. Kirchhausen, T. Walz, *Nature* **2004**, *432*, 573–579.

[34] E Ungewickell, H. Ungewickell, *Journal of Biological Chemistry* **1991**, *266*, 12710–12714.

[35] J. A. Ybe, B. Greene, S.-H. Liu, U. Pley, P. Parham, F. M. Brodsky, *The EMBO journal* **1998**, *17*, 1297–1303.

[36] K. H. Wrighton, *Nature Reviews Molecular Cell Biology* **2014**, *15*, 560–561.

[37] L. P. Jackson, B. T. Kelly, A. J. McCoy, T. Gaffry, L. C. James, B. M. Collins, S. Höning, P. R. Evans, D. J. Owen, *Cell* **2010**, *141*, 1220–1229.

[38] S. Zaremba, J. H. Keen, *The Journal of cell biology* **1983**, 1339–1347.

[39] E. M. Schmid, H. T. McMahon, *Nature* **2007**, *448*, 883.

[40] T. Kanaseki, K. Kadota, *The Journal of cell biology* **1969**, *42*, 202.

[41] C. J. Smith, N. Grigorieff, B. M. Pearse, *The EMBO Journal* **1998**, *17*, 4943–4953.

[42] R Beck, M Ravet, F. Wieland, D Cassel, *FEBS letters* **2009**, *583*, 2701–2709.

[43] J Bethune, F Wieland, J Moelleken, *The Journal of membrane biology* **2006**, *211*, 65–79.

[44] H. Hughes, D. J. Stephens, *Histochemistry and cell biology* **2008**, *129*, 129–151.

[45] H. T. McMahon, I. G. Mills, *Current opinion in cell biology* **2004**, *16*, 379–391.

[46] M. P. Allen, D. J. Tildesley, *Computer Simulation of Liquids*, Clarendon Press, New York, NY, USA, **1989**.

[47] D. A. McQuarrie, *Statistical thermodynamics*, HarperCollins Publishers, **1973**.

[48] D. Frenkel, B. Smit, *Understanding Molecular Simulation*, 2nd, Academic Press Inc., Orlando, FL, U.S.A., **2001**.

[49] W. K. den Otter, M. R. Renes, W. J. Briels, *J. Phys. Condens. Matt.* **2010**, *22*, 104103.

[50] W. K. den Otter, M. R. Renes, W. J. Briels, *Biophys. J.* **2010**, *99*, 1231–1238.

[51] W. K. den Otter, W. J. Briels, *Traffic* **2011**, *12*, 1407–1416.

[52] I. M. Ilie, W. K. den Otter, W. J. Briels, *J. Chem. Phys.* **2014**, *141*, 065101.

[53] D. P. Landau, K. Binder, *A Guide to Monte Carlo Simulations in Statistical Physics*, Cambridge University Press, Cambridge, U.K., **2000**.

[54] D. A. McQuarrie, *Statistical Mechanics*, Harper & Row Publishers, New York, NY, USA, **1976**.

[55] Numerical recipes in FORTRAN (Cambridge).

[56] R. Matthews, C. N. Likos, *Soft Matter* **2013**, *9*, 5794–5806.

[57] N. Cordella, T. J. Lampo, N. Melosh, A. J. Spakowitz, *Soft matter* **2015**, *11*, 439–448.

[58] J. J. VanDersarl, S. Mehraeen, A. P. Schoen, S. C. Heilshorn, A. J. Spakowitz, N. A. Melosh, *Soft matter* **2014**, *10*, 6219–6227.

[59] M Muthukumar, R. Nossal, *The Journal of chemical physics* **2013**, *139*, 121928.

[60] M. Saleem, S. Morlot, A. Hohendahl, J. Manzi, M. Lenz, A. Roux, *Nature communications* **2015**, *6*, 6249.

Chapter 2

Clathrin assembly regulated by adaptor proteins in coarse-grained simulations

The assembly of clathrin triskelia into polyhedral cages during endocytosis is regulated by adaptor proteins (APs). We explore how APs achieve this by developing coarse-grained models for clathrin and AP2, employing a Monte-Carlo click interaction, to simulate their collective aggregation behaviour. The phase diagrams indicate that a crucial role is played by the mechanical properties of the disordered linker segment of AP. We also present a statistical-mechanical theory for the assembly behaviour of clathrin, yielding good agreement with our simulations and experimental data from the literature. Adaptor proteins are found to regulate the formation of clathrin coats under certain conditions, but can also suppress the formation of cages. [1]

2.1 Introduction

In eukaryotic cells, clathrin-mediated endocytosis is a major pathway for the internalization of cargo molecules like hormones, receptors, transferrin, membrane lipids, and the occasional virus [1–6]. The cargo molecules are collected and sorted in a clathrin coated pit (CCP), which subsequently evolves into an encapsulating clathrin coated vesicle (CCV). These coats arise through a self-assembly or polymerization process of clathrin proteins against the cytoplasmic face of cellular membranes. The clathrin protein has a peculiar shape with three long curved legs, see Fig. 2.1, that allows it to bind with many partners into a wide range of polyhedral cages, as well as to bind accessory proteins that assist at various stages of the endocytosis process [7–12]. Although clathrin is a major component and the namesake of CCPs and CCVs, it does not bind directly to either the membrane or the cargo. These are the tasks of so-called adaptor proteins, which often are active only at specific membranes in the cell [13–17]. The members of the AP-family, AP1 through AP5, are tetrameric complexes consisting of two large and two small subunits. A second family of adaptor proteins is formed by the clathrin-associated sorting proteins (CLASP), a collection of monomeric proteins including AP180, epsin and Eps15 [17, 18]. The global structure of the members of both families is very similar: they consist of a neatly folded section that binds to the membrane and a long disordered segment with clathrin binding motifs. Members of the AP-family posses a second long disordered segment, to attract assisting proteins. Of all adaptor proteins (henceforth abbreviated as AP, irrespective of family), probably

[1] This chapter has been published as **M. Giani**, W. K. den Otter, and W. J. Briels, 'Clathrin assembly regulated by adaptor proteins in coarse-grained models' Biophysical Journal, vol. 111, no. 1, pp. 222-235, 2016.

the most studied adaptor protein is the AP2 complex regulating endocytosis, which will also be the reference point in this study [8, 19–21].

In addition to linking clathrin to membrane and cargo, a main function of APs is to regulate the assembly of clathrin cages by binding to multiple triskelia simultaneously. A series of *in vitro* experiments established that clathrin proteins in solution can be induced to self-assemble by adding APS [7, 13]. Recent structural studies revealed that AP2 can adopt two configurations, *i.e.* a closed state with part of the linker blocked from interacting with clathrin and an open state where AP2 can bind two triskelia [22, 23]. With AP2 adapting the open state only when bound to a membrane, the formation of clathrin cages in a cell is effectively limited to the membrane. This mechanism may also explain why the *in vitro* assembly behaviour of clathrin varies with the preparation state of the adaptor proteins, with well-cleaned adaptors inducing less activity [24]. Our objective in this study is the little explored question: beyond the ability to bind two triskelia simultaneously, what else is required of APs to induce the formation of clathrin cages in solution?

The presence of an AP binding site at the end of each clathrin leg, a location henceforth informally referred to as the 'toes' by following the common analogy of the clathrin leg with the human leg, see Fig. 2.1, is well established. Experiments with recombinant clathrin fragments indicate that this binding site is crucial to the inducement by AP2 of cage formation [25]. At least one additional binding site, also required for cage formation, resides higher up each leg. Experiments with 'clipped' triskelia point at a location on the trimer hub [26], *i.e.* in the region extending from the 'hip' to just beyond the 'knee', see Fig. 2.1. Pull-down experiments identified a binding site near the 'ankle' [27]. Both options will be explored here.

Besides *in vivo* and *in vitro* experiments, the assembly behaviour of clathrin has also been explored by *in silico* studies. In earlier work, two of us developed a highly coarse-grained patchy particle model of clathrin as a rigid triskelion with either straight or bend legs, and showed that anisotropic leg-leg interactions are the key to self-assembly [28, 29]. Simulations with this model predicted a binding energy of about $23k_BT$ per clathrin in a cage, suggested a novel scenario for the transition from flat plaque to curved coat and yielded an assembly time scale in reasonable agreement with experiments [30, 31]. Matthews and Likos modeled clathrin as a collection of 13 bead patchy particles, endowed with anisotropic interactions, and showed how these triskelia deformed a lipid membrane into a bud [32]. Spakowitz and collaborators modeled clathrin as a spherical particle with anisotropic interactions accounting for three straight legs, and studied, among other properties, how a membrane influences an adjacent clathrin lattice [33, 34]. Adaptor proteins, which are crucial in bringing triskelia together under *in vivo* conditions, have been omitted in all clathrin simulations to date.

To address our research question, we apply coarse-grained simulations and statistical-mechanical theory to explore the ability of APs to induce the assembly of triskelia cages in solution. Because the AP model is based on the aforementioned key features, it is to be expected that other adaptor proteins can be modeled in a similar way. This paper is organized as follows: In Section 2.2 the clathrin simulation model is briefly discussed, the matching AP simulation model is introduced, and the implementation of click-interactions in Monte Carlo simulations is described. The results on simulations of mixtures of triskelia and APs are presented and interpreted in Section 2.3. The deduced qualitative understanding is translated into a fairly simple quantitative theory in Section 2.4, obtaining remarkably good agreement with simulations and experiments. We end with a summary of the main conclusions.

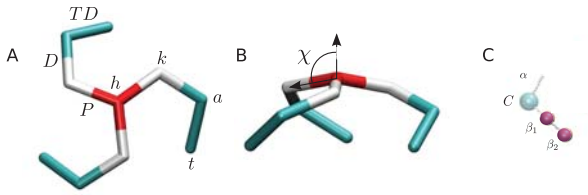


Figure 2.1: The highly coarse-grained simulation models of (A and B) clathrin and (C) AP2 on the same scale. In the rigid clathrin model, three proximal leg segments (P) radiate from a central hip (h) to the knees (k), at a pucker angle χ relative to the symmetry axis, followed by distal leg segments (D) running to ankles (a) and terminal domains (TDs) ending at the toes (t). The AP model features two binding sites for clathrin, β_1 and β_2, connected by a flexible linker. In the full AP2 protein, the β linker connects to a folded core (c) and a flexible α linker; these are omitted in the simulations because they do not play a role in the *in vitro* assembly process.

2.2 Model and method

In several preceding studies [28–31], we modeled clathrin as rigid patchy particles with three identical curved legs, see Fig. 2.1. The three legs are connected at a central 'hub', at a 'pucker' angle χ relative to the threefold rotational symmetry axis of the particle, reflecting clathrin's intrinsic non-zero curvature. We here select a pucker angle, $\chi = 101°$, typical of soccer-ball cages containing 60 triskelia, which is the most common cage size in *in vitro* experiments in the presence of AP [35].

Each leg consists of two segments (the proximal and distal sections; the terminal domains were not included because of their expected small contribution to the clathrin-clathrin binding interaction) connected at the 'knee' under a fixed angle and ending at the 'ankle.' All leg segments are straight and of identical length, $\sigma = 17\,\mathrm{nm}$. The orientation of the distal segments relative to the proximal segments was chosen to allow maximum overlap between a particle and a secondary particle whose hub is situated at a knee of the primary particle. In a completed cage, a hub is located at every vertex, on top of three knees and three ankles of neighboring and next-nearest triskelia, respectively. A lattice edge is thus composed of two proximal and two distal segments, where the amino acid sequences in both pairs of like segments run in opposite directions (*i.e.* anti-parallel). The attractive interaction between any pair of segments, which for clathrin is believed to result from a multitude of weak interaction sites along the legs [36–38], is modeled by a four-site potential based on the distances between the end-points of the two segments, with a minimum value of $-\epsilon$ for two perfectly aligned segments, as described in detail in the Supplementary Material. The interaction is anisotropic under rotations around the long axes of the segments, to reflect that the binding sites are most likely concentrated on one side of the segment, to wit, the side that in a cage edge faces the three adjacent segments. Simulations revealed that this anisotropy of the attractive potential is crucial for the spontaneous self-assembly of triskelia into polyhedral cages [28, 29]. Excluded volume interactions between triskelia were omitted for computational reasons: this requires a more complex particle shape with non-linear proximal and distal segments, as well as demands some flexibility of the legs, in order for the particles to pack together into cages with four legs interweaving along each edge, while the simulation step has to be reduced to prevent the relatively thin legs from crossing each other. Excluded volume interactions are important to prevent triskelia from binding to a cage edge in a 'slot' that is already occupied by another triskelion; this property is incorporated in the simulation model by a repulsive potential between parallel like segments. The moderate

flexibility of the clathrin protein extends its interaction range beyond that of a rigidified protein; this effect is to some extend accounted for by the enlarged range of the inter-segmental potential. The terminal domains (TD) at the ends of the legs, see Fig. 2.1, were not included in our previous simulations, but they are required in the current study as binding sites for APs. The length and orientation of the TD with respect to the proximal and distal segments were estimated using the structural information file 1XI4 for a clathrin cage [2, 36], available at the Protein Data Bank (PDB). Since the TD is about equally long as the proximal and distal segments, they are all assigned the same length σ in the model. The TD is attached to the ankle at an angle of 114° relative to the distal domain, with the three segments of a leg forming a dihedral angle of 28°. The clathrin-clathrin interactions are kept identical to those in the previous model; the TDs do not contribute to these interactions.

Continuing in this reductionist approach, we here introduce a matching simulation model of an AP, see Fig. 2.1. The model comprises the part of the AP2 protein that is involved in clathrin binding, *i.e.* the C-terminal region of the β linker comprising the clathrin-box LLNLD of residues 631 through 635, the clathrin-binding appendage domain formed by residues 705 through 937, and the flexible linker connecting these two interaction sites [19]. Our coarse-grained representation of this AP2 fragment consists of two point particles, embodying the two binding sites, connected by a tether. Since the remainder of the AP2 tetramer does not partake in clathrin binding and assuming that AP2s do not bind to each other, the omission of the majority of the protein is of no further consequence to the cage assembly process studied here. Excluded volume interactions are again omitted for reasons of computational efficiency; we note that the interior volume of a cage is far larger than the collective volume of the APs bound to a cage. The short range of the clathrin-AP binding interaction is inconvenient from a numerical point of view (see below). In stead, we developed a potential in which the α^{th} binding site on the i^{th} triskelion and the β^{th} particle of the j^{th} AP dimer are bound with a fixed energy $-\zeta$ and are limited to a maximum separation ρ in the 'clicked' state ($b_{i\alpha,j\beta} = 1$), while there are no interactions between these sites in the 'unclicked' state ($b_{i\alpha,j\beta} = 0$). As a function of the distance $r_{i\alpha,j\beta}$, the interaction potential then reads as

$$\phi_{\text{click}}(r_{i\alpha,j\beta}, b_{i\alpha,j\beta}) = \begin{cases} 0 & \text{for} \quad b_{i\alpha,j\beta} = 0 \\ \begin{cases} -\zeta & \text{for} \quad r_{i\alpha,j\beta} < \rho \\ \infty & \text{for} \quad r_{i\alpha,j\beta} \geq \rho \end{cases} & \text{for} \quad b_{i\alpha,j\beta} = 1, \end{cases} \tag{2.1}$$

where $\zeta > 0$, as illustrated in Fig. S4. Because excluded volume interactions between AP2 tetramers ensure that a binding site on a clathrin can host at most one AP site, the clicks in the simulation model are constructed to bemutually exclusive: a site can partake in one click only. The clicks are also specific: the β_1 AP bead solely binds to the end of the TD, *i.e.* at the toes, while the β_2 bead clicks only to a site higher up a triskelion's leg.

The two clathrin binding sites of AP2 are connected by an essentially structure-less sequence of about 70 residues [19]. According to polymer theory, this flexible linker will effectively act as an entropic spring with a spring constant k and a maximum length L [39, 40]. This behaviour is modeled here by the finite extensible nonlinear elastic (FENE) potential [41],

$$\phi_{\text{linker}}(l_j) = \begin{cases} -\frac{1}{2}kL^2 \ln\left[1 - (l_j/L)^2\right] & \text{for} \quad r < L \\ \infty & \text{for} \quad r \geq L, \end{cases} \tag{2.2}$$

where l_j denotes the length of the j^{th} AP dimer. The spring constant of an entropic spring is given by [40]

$$k = \frac{3k_BT}{2Ll_p}, \tag{2.3}$$

where l_p is the persistence length. Given an average residue length of 0.37 nm, the linker of 70 residues connecting the two clathrin binding sites has a contour length of $L \approx 26\,\mathrm{nm} \approx 1.5\sigma$. Combination with the experimental value $l_p \approx 0.6\,\mathrm{nm}$ for disordered proteins then yields $k \approx 30 k_B T/\sigma^2$ for the linker.

The assembly characteristics of the combined models were simulated by the Monte Carlo (MC) method, *i.e.* by the weighted acceptance of randomly generated changes of the system configuration [42, 43]. Suppose that, by a sequence of steps, the system arrives in state m. In the MC technique, the transition probability from the current state m to a new state n is expressed as

$$P_{m\to n} = P^{\mathrm{trial}}_{m\to n} P^{\mathrm{acc}}_{m\to n}, \tag{2.4}$$

where $P^{\mathrm{trial}}_{m\to n}$ denotes the probability of generating the trial configuration n from state m, and $P^{\mathrm{acc}}_{m\to n}$ is the probability of accepting n as the next state in the sequence of states; if the move is rejected, the system remains in the current state and m is added (again) to the sequence of sampled states. For a symmetric trial move generator, $P^{\mathrm{trial}}_{m\to n} = P^{\mathrm{trial}}_{n\to m}$, the acceptance probability

$$P^{acc}_{m\to n} = \min\left(1, \exp\left\{-\frac{\Phi(m) - \Phi(n)}{k_B T}\right\}\right), \tag{2.5}$$

with $\Phi(m)$ denoting the potential energy of state m, will produce a sequence of states in agreement with the equilibrium Boltzmann distribution.

The algorithm employed in the current study applies two different types of trial moves, namely trial moves that alter the positions and orientations of particles and trial moves that alter the connectivity between particles. The type of move is selected at random in every MC step, with positional moves selected f times as often as connectivity moves. Positional trial moves start by randomly selecting a protein. If a clathrin is selected, its center of mass is displaced along all 3 Cartesian directions by random values in the range $[-\frac{1}{4}\sigma, \frac{1}{4}\sigma]$, and the particle is rotated around a random axis through the center of mass over a random angle in the range $[-\frac{1}{2}, \frac{1}{2}]$ rad. A known complication in MC simulations is the drastic reduction of the mobility of particles interacting with neighbours, relative to the mobility of non-interacting particles, as can be seen clearly in movies of MC simulations [29]. This is a minor issue in the assembly of cages from a solution containing clathrin only, as the free triskelia readily diffuse to a nearly immobile cage fragment. In simulations of mixtures of clathrin and AP, however, the binding of APs to triskelia will slow down their combined diffusion and hence significantly delay their attachment to cage fragments, especially if the AP-clathrin bond is strong and short-ranged. The solution adopted here is to apply cluster moves [42, 44], *i.e.* the AP beads clicked to the selected triskelion move together with this clathrin, maintaining the statuses $b_{i\alpha,j\beta}$ and distances $r_{i\alpha,j\beta}$ of all clicks. Consider an AP with a bead clicked to the selected triskelion. If its other bead is unclicked or clicked to the same triskelion, the entire AP is moved with the clathrin as if they formed a rigid unit. If the AP's other bead is clicked to another clathrin, then this second bead is excluded from the trial move and, consequently, the length of the AP changes in the trial move. Next, the move is accepted or rejected following Eq. (5.15). If in a positional trial move an AP is selected, its two beads will be displaced independently. An unclicked bead is displaced in all 3 Cartesian directions by random values in the range $[-\frac{1}{4}\sigma, \frac{1}{4}\sigma]$, while a clicked bead is moved to a random position within a sphere of radius ρ centered around the clathrin's matching clicking site. Next, the move is accepted or rejected following Eq. (5.15). Again, the statuses $b_{i\alpha,j\beta}$ of all clicks are conserved by these trial moves. In a clicking trial move, an AP bead is selected at random. The neigbourhood of radius ρ around this particle is scanned for matching clicking sites on triskelia; for a bead that is already clicked, its current partner will be among the K detected sites. The unclicked state is included as the zeroth option. In stead of

the above selection and acceptance steps, we directly accept one of the $K+1$ trial states as the next state. The probability of selecting the k^{th} option is given by

$$P_k = \frac{\exp(-\beta\Delta\phi_k^{\text{click}})}{\sum_{k'=0}^{K}\exp(-\beta\Delta\phi_{k'}^{\text{click}})}, \tag{2.6}$$

where the energy change $\Delta\phi_k^{\text{click}}$ between the current state and the k^{th} trial state can only yield the values $\Delta\phi_k^{\text{click}} = \zeta$ for an unclicking trial move, $\Delta\phi_k^{\text{click}} = -\zeta$ for a clicking trial move and $\Delta\phi_k^{\text{click}} = 0$ if the connection remains (un)clicked.

A number of simulations were run to verify that the unconventional click-potential and click-dependent MC cluster moves sample the correct equilibrium distribution. Simulations with 1,000 clathrins and 3,000 APs in a cubic box of volume $10^6\sigma^3$ were used to determine the equilibrium constants of the reactions between triskelia and AP, defined as

$$K_{n,m}^{\text{tri}} = \frac{\left[\text{CA}'_n\text{A}''_m\right]_0}{\left[\text{C}\right]_0\left[\text{A}\right]_0^{n+m}}, \tag{2.7}$$

where $\left[\text{C}\right]_0$, $\left[\text{A}\right]_0$ and $\left[\text{CA}'_n\text{A}''_m\right]_0$ denote, respectively, the concentrations of unbound triskelia, unbound APs, and triskelia complexes with n single-bound and m double-bound adaptor proteins, in molars, see Appendix 2.6.1. To improve the sampling efficiency, we reduced the number of distinct reactions products to three by reducing the number of clicking sites per triskelion from six to two – at the toes and ankle of the same leg – and reduced the entropic spring constant to $k = 1k_BT/\sigma^2$, while retaining the maximum extensibility of 1.5σ. Furthermore, to enable comparison with exact analytical solutions, the adaptor proteins were not allowed to click to two clathrin particles simultaneously and the interactions between triskelia were turned off. Figure S5 shows the equilibrium constants for triskelia that click once with an AP, $K_{1,0}^{\text{tri}}$, and for triskelia that bind two APs, $K_{2,0}^{\text{tri}}$, as functions of the clicking energy. Excellent quantitative agreement is observed with the statistical mechanical reaction equilibrium theory presented in Appendix 2.6.1, which is shown in the graph as straight lines. Additional simulations confirm that the equilibrium constants scale with the clicking radius ρ conform the power-law dependence derived in Appendix 2.6.1 (data not shown). The graph also shows the equilibrium constants for APs that double-click to a clathrin leg, $K_{0,1}^{tri}$, *i.e.* both sites of the AP are bound to the same triskelion leg. This occurs because the estimated maximum extensibility of the AP linker, $L \approx 26\,\text{nm}$, well exceeds the length of the TD, $\sigma \approx 17\,\text{nm}$, though the considerable elongation of the AP linker makes this double-click unfavourable. Again, the equilibrium constant is in good agreement with the theory. Several simulations were run with smaller systems to verify that the translation versus click attempt ratio does not affect the results presented in this paper; we settled on a value of $f = 10$ for reasons of computational efficiency.

The production simulations were all run with 1,000 triskelia confined to a cubic box of volume $10^6\sigma^3$ with periodic boundary conditions. The number density of 1 triskelion per $10^3\sigma^3$ corresponds to an *in vitro* condition of about $\sim$0.2 mg/ml. Self-assembly in the absence of APs is observed *in vitro* for a slightly acidic solution (pH 6.2, 20 mM $MgCl_2$), with a critical assembly concentration (CAC) of $\sim$0.1 mg/ml [45], *i.e.* the overall concentration where the fractions of bound and unbound triskelia are equal. In an earlier simulation study, we established that this concentration is the CAC of coarse-grained triskelia that gain $E_c \approx 23k_BT$ upon binding to a cage, which is realized for a segment-segment interaction parameter $\epsilon \approx 6k_BT$ [30]. There we also showed that concepts borrowed from the thermodynamics of micelles allow a theoretic derivation of the binding energy from the measured CAC. Muthukumar and Nossal extended these ideas with energetic contributions reflecting the curvature of the clathrin coat and applied them to

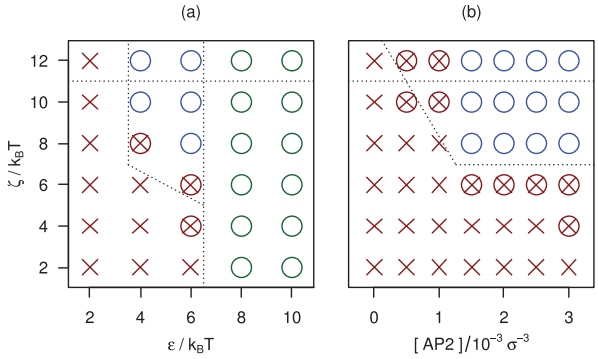


Figure 2.2: Cage assembly diagrams for clathrin, for 1,000 triskelia at a concentration of $10^{-3}\sigma^{-3}$, combined with model APs clicking to the ends of the TDs and the ankles of clathrin, (a) as a function of the clathrin-clathrin binding strength ϵ and the clathrin-AP clicking strength ζ, at an AP-to-clathrin ratio of 3, and (b) as a function of the AP concentration (for AP-to-clathrin ratios from 0 to 3) and the clathrin-AP clicking strength, at a clathrin-clathrin binding strength of $\epsilon = 6k_BT$. The markers denote parameter combinations that result in the self-assembly of cages (a green circle if cages are also formed in the absence of APs, a blue circle if assembly only proceeds in the presence of APs), combinations that do not yield cages (red crosses), and conditions where cages do not assemble spontaneously but pre-assembled cages appear stable (red cross in red circle). The dashed lines indicate the approximate locations of phase boundaries, as discussed in more detail in the main text.

analyse cages grown in the presence of AP2, even though the adaptor molecules themselves were not included in the theoretical model [46]. A novel statistical mechanical derivation linking the binding energy to the CAC, by considering a cage as a collection of p rigid triskelia with highly restricted translational and rotational freedom, is presented in Appendix 2.6.2. For the assembly reaction $p\mathrm{C} \rightleftharpoons \mathrm{C}_p$, we obtain a standard state free difference of

$$\Delta G^0_p = \mu^0_{\mathrm{C}_p} - p\mu^0_{\mathrm{C}} \approx p\Delta\mu^0_{\mathrm{C}}, \tag{2.8}$$

with μ^0_{X} the standard reference chemical potential of component X and $\Delta\mu^0_{\mathrm{C}} \approx -16.4k_BT$. Applied to the simulation model, this translates into a binding energy $E_c \approx 27k_BT$, in good agreement with the simulations. Recent experiments on the mechanical properties of clathrin coats adjacent to membranes confirm the binding (free) energies predicted by simulations and theory [47].

2.3 Results I. Simulations

The effect of model APs on the self-assembly behaviour of model triskelia is studied by systematically varying the clathrin-clathrin interaction ϵ, the AP-clathrin clicking strength ζ and the AP-to-clathrin ratio. Figure 2.2 shows the assembly behaviour on two cross-sections of this three-dimensional parameter space, for the AP model clicking to the ankles and toes of clathrin. Every marker represents five independent simulations of 10^{10} MC steps, requiring about a week each on a desktop computer. Red crosses mark conditions where no spontaneous self-assembly of sizable cage fragments is observed. Green and blue circles indicate the self-assembly of at least one complete cage across the five simulations. For the green circles, $\epsilon > 6k_BT$, cages already self-assemble in the absence of APs. The blue circles highlight conditions where triskelia do not self-assemble in the absence of APs but do form cages in their presence – this is the region of

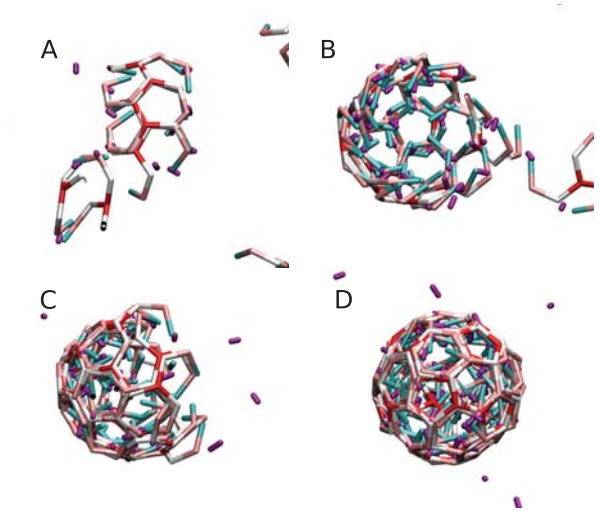


Figure 2.3: A sequence of snapshot of triskelia assembling into a cage in the presence of APs, for $\epsilon = 6k_BT$, $\zeta = 8k_BT$, and a AP-to-clathrin ratio of 3, at intervals of 10^9 MC steps. The coloring of the particles is the same as in Fig. 2.1.

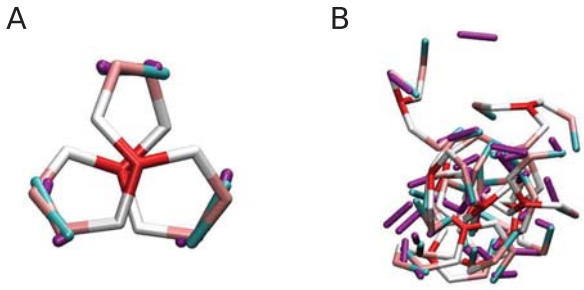


Figure 2.4: Adaptor proteins will bring triskelia together without regard for the relative positioning and orientation of these triskelia. A common aggregate (a) comprising two clathrins bonded by six APs (purple), saturating all clicking sites of the cluster. When the cluster is small and the interactions are weak, there are many opportunities to break the AP bonds and reshuffle the triskelia into a more favourable configuration. At high AP-clathrin clicking strengths, large disordered clusters develop rapidly (b); these will only very slowly acquire more order.

parameter space where APs induce and control the formation of clathrin cages. The assembly of cages in the green and blue regions proceeds by a nucleation and growth process, just like in clathrin-only simulations [28, 31]. Small clusters of a few triskelia and APs, see Figs 2.3(a) and 2.4(a), are formed and destroyed continuously. Occasionally, one of these small aggregates crosses the nucleation barrier and grows into a cage, as illustrated by the snapshots in Fig. 2.3. Because of the rigidity of the clathrin model, these cages are all of about the same size, containing ∼60 triskelia in near-spherical polyhedra with 12 pentagonal and ∼20 hexagonal faces. The average cage diameter of about 4.5σ (∼75 nm) agrees with that for cages grown *in vitro* in the presence of APs [35], which motivated our choice of a 101° pucker angle. Cages grown in simulations with and without AP particles are of the same size. In *in vitro* experiments, however, a size difference is observed between cages grown with AP and cages grown without AP [35]. It is unclear whether this difference is caused by the presence of APs, or by the pH reduction to induce cage formation in the absence of APs. We note that the cage size is very sensitive to the pucker; a decrease from 101° to 100° increases the average cage size by about 10 particles [28]. Almost all self-assembled cages are 'complete', *i.e.* triskelion hubs reside at every vertex. Only rarely do one or two vertices of a nearly complete cage remain unoccupied, presumably because the remaining vacancies are less favourable binding sites than the occupied slots. The high prevalence of completed cages indicates that all vertices in these cages are of approximately equal binding affinity, which appears to confirm the 'probable roads' hypothesis by Schein *et al.* [48]. For low attachment rates at the edge of a growing fragment, particles binding in an unfavourable way have a high probability of being released again before the defect becomes permanently incorporated in the lattice through the attachment of subsequent particles. Aggregation becomes frustrated when the binding energies are too strong. For inter-segmental interactions exceeding $\sim 10 k_B T$, the triskelia easily stick together and thereby quickly form a multitude of small aggregates which only very slowly merge into larger clusters. This evolution is reminiscent of that observed *in vitro* below pH 5.8 [49]. A clicking energy exceeding $\sim 11 k_B T$ makes the APs eager to click to triskelia, thereby rapidly forming disordered clusters like that in Fig. 2.4(b), which only very slowly develop into cage fragments and ultimately cages.

The rarity of nucleation necessitates excessively long simulations to accurately locate phase boundaries or to determine equilibrium cage concentrations (these will be obtained below by other means). The expedient used in the simulation phase diagrams of this section is the binary detection of self-assembled cages: green or blue circles if cages are formed, and red crosses otherwise. For phase points close to a phase boundary, additional simulations were initiated with configurations containing several half-spherical coats, to explore whether these aggregates grow into complete cages or disintegrate into monomers. In this context we note that the disassembly of an unstable coat fragment typically proceeds much faster than the completion of a stable fragment. The results of these simulations are included as green or blue circles or as red crosses in all simulation phase diagrams. For Fig. 2.2 only, a further refinement of the phase boundaries was obtained by running an additional set of simulations initiated with fully assembled cages stabilized by nearly three APs per clathrin (obtained from simulations at another phase point). The surviving cages are marked in Fig. 2.2 by red circles, superposed on the red cross indicating 'no spontaneous assembly'. If two simulations with the same parameter settings but 'opposite' starting configurations converge to the same final state, it is very likely that this final state is the equilibrium state. If their final states differ, then either the stability difference between these states is small or (at least) one of the simulations is trapped in a local minimum of the free energy landscape.

The dashed lines in Fig. 2.2 indicate the estimated phase boundaries, where the boundary slightly to the right of $\epsilon = 6 k_B T$ was established previously and with greater accuracy [30] than the other boundaries. One sees in Fig. 2.2(a) that, at the prevailing concentrations, the APs

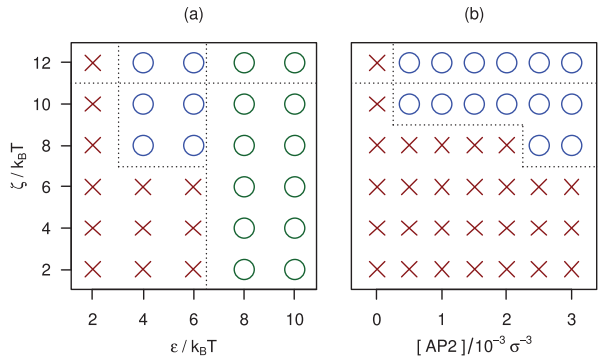


Figure 2.5: Assembly diagrams for model APs clicking to the ends of the TDs and the knees of clathrin, with all other conditions identical to those in Fig. 2.2. The blue circles again highlight the parameter space where cage formation is controlled by APs.

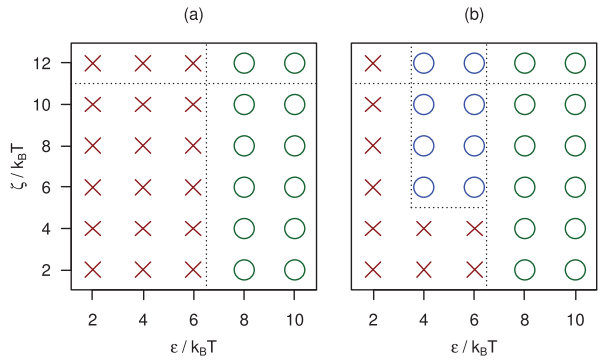


Figure 2.6: Assembly diagrams for model APs with a reduced (entropic) spring constant, $k = 1\epsilon/\sigma^2$, all other parameters identical to those in Fig. 2.2. APs clicking to the ankles and TDs of clathrin (a) are no longer able to regulate the formation of cages, while APs clicking to the knees and TDs of clathrin (b) are still operational.

are able to regulate the emergence of cages for $4k_BT \lesssim \epsilon \lesssim 6k_BT$ and $\zeta \gtrsim 7k_BT$ (*i.e.* the blue region). A cross section of this region, by varying the AP concentration at fixed $\epsilon = 6k_BT$, is presented in Fig. 2.2(b). This plot shows that AP-induced cage assembly requires a clicking energy $\zeta \gtrsim 7k_BT$ as well as an AP concentration equal to or exceeding the clathrin concentration.

Besides the AP model discussed above, simulations were run with a number of alternative models to explore the conditions conducive to adaptor-induced cage formation. APs clicking at the knees and toes yield the assembly diagrams presented in Fig. 2.5. The graph on the left is similar to its counterpart in Fig. 2.2, and shows that APs binding at the knees are equally capable of regulating the assembly of cages as APs binding at the ankles. The graph on the right shows an interesting difference between the two cases: self-assembly continuous down to much smaller AP concentrations. Lowering the effective spring constant of the linker between the AP beads to $k = 10k_BT/\sigma^2$ has little impact on the assembly diagrams of either adaptor model (data not shown). Upon a further reduction to $k = 1k_BT/\sigma^2$, see Fig. 2.6, the AP clicking to the knees and toes remains operational (with a slight shift in the smallest ζ inducing cage formation), while the AP clicking to the ankles and toes ceases to function.

To understand the results reported above, we now turn to unraveling the mechanism by which APs induce the aggregation of triskelia.

The discussion presented here is qualitative in nature; a quantitative analysis of the insights gained is presented in the next section. Consider first the AP model that binds to the toes and the knees. It is clearly energetically favourable for an AP to click to triskelia. The largest gain in energy is obtained when the adaptor clicks twice, which is only achieved – note that the toe-knee distance in a clathrin is longer than the maximum extensibility of the linker – if the AP binds to two distinct triskelia. Bringing two triskelia together strongly enhances their chances of adopting the correct relative positions and orientations, and hence promotes successful binding. Adaptor proteins may thus contribute to both the stability of clathrin aggregates and the rate at which they are formed. Note that this line of thought assumes that the energetic gain upon binding outweighs the accompanying entropic loss in translational freedom (and in rotational freedom for clathrin-clathrin binding) and thereby lowers the overall Helmholtz free energy of the system. Hence, whether the AP plays an supporting role in cage formation depends on the clicking strength as well as on the AP and clathrin concentrations.

For the adaptor clicking at the toes and ankle, the energetic gain upon double clicking to one clathrin is identical to that of clicking to two triskelia. This partially invalidates the mechanism proposed above, by providing the APs with an alternative binding option that does not contribute toward cage assembly. Yet, the simulations of Fig. 2.2 indicate that these adaptors are able to induce cage formation. Inspection of the length distribution of the linkers (data not shown) reveals that i) most APs bound to a cage are bridging between pairs of triskelia, and ii) the nearest toe-ankle distance in a cage is shorter than the toe-ankle distance of 1σ along a clathrin leg. This suggests that the shorter linker length in a cage, and between triskelia in the process of coming together, results in a lower elastic energy and hence a higher Boltzmann factor, and thereby favours APs connecting between sites on distinct triskelia over APs connecting to two sites on the same clathrin. The reader might note that the distribution of end-to-end distances of the real linker is determined by entropic effects, while this distribution is modeled here as an energetic effect, see Eq. (2.2), but this does not present any conceptual problem as both yield the same dependence of the free-energy on the inter-bead distance.

In support of the above considerations, we recall the impact on the assembly behaviour of reducing the linker spring constant at constant maximum extensibility, see Fig. 2.6. For the model AP clicking at toes and knees, the reduction of the spring constant was of little consequence, in agreement with the mechanism where an adaptor clicking twice always establishes a link between two distinct triskelia. For the model AP clicking to toes and ankles, however, lowering the spring constant reduces the difference in internal energy between AP double-clicked to one clathrin (with the linker stretched to 1σ) and AP clicked to two triskelia (with a shorter linker length). With this reduction, the preference for inter-clathrin over intra-clathrin bonds diminishes and at $k = 1k_BT/\sigma^2$ the number of APs links holding triskelia together becomes too low to stabilize a cage.

2.4 Results II. Theory

A statistical mechanical theory of AP-induced cage assembly, built on the concepts deduced above, is derived in Appendix 2.6.3. The theory predicts the equilibrium constant $K^{\text{cage}}_{p,n,m}$ relating the concentrations of unbound triskelia and unbound APs to the concentration of cages of p triskelia 'decorated' with n single clicked APs and m inter-triskelion double clicked APs. Suppose one knows the average binding energy of a triskelion in a cage devoid of APs, E_c, the clathrin-AP interaction strength ζ, and the total concentrations of clathrin and AP in a system, $\left[\mathrm{C}\right]_{\mathrm{t}}$ and $\left[\mathrm{A}\right]_{\mathrm{t}}$, respectively, it is now possible to compute the equilibrium concentrations of all decorated cages in that system, $\left[\mathrm{C}_p\mathrm{A}'_n\mathrm{A}''_m\right]$, by the iterative procedure outlined in Appendix 2.6.3; the

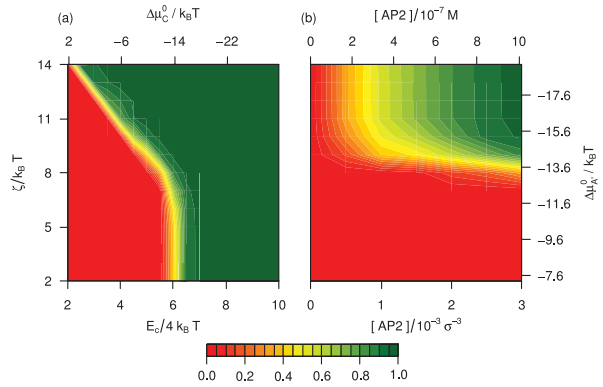


Figure 2.7: Assembly diagrams calculated using the theory derived in Appendix 2.6.3, showing the fraction of clathrin bound in cages, for APs clicking to the ends of the TDs and the ankles of triskelia, (a) as a function of the binding energy per clathrin in an AP-free cage, E_c, and the clathrin-AP clicking strength, ζ, at an AP-to-clathrin ratio of 3, and (b) as a function of the AP concentration (for AP-to-clathrin ratios from 0 to 3) and the clathrin-AP clicking strength, at $E_c = 22k_BT$. The two graphs refer to the same total clathrin concentration, $\left[\mathrm{C}\right]_\mathrm{t} = 10^{-3}\sigma^{-3} \approx 3.4 \cdot 10^{-7}\,\mathrm{M}$, and similar interaction energies, as their counterparts in Fig. 2.2. For comparison purposes, the horizontal axis of the left plot is scaled by the simulation-based ratio $E_c/\epsilon \approx 4$ (see main text). The alternative axes to the graphs are labeled with the standard chemical free energy differences of AP single-clicking to clathrin, see Eq. (2.23), and of clathrin assembling into AP-free cages, see Eq. (2.33), and with total AP concentrations in molars.

overall cage concentration then follows by a summation over all decorated cages, *i.e.* all values of p, n and m. Since the simulations predominantly produced cages of 60 triskelia, we restrict the theoretical calculations to one cage size, $p = 60$. The phase diagrams calculated for the ankle-binding AP model are shown in Fig. 2.7. To facilitate the comparison with the simulation results in Fig. 2.2, the plots are based on the same total clathrin concentration, $\left[\mathrm{C}\right]_\mathrm{t} = 10^{-3}\sigma^{-3}$, and similar inter-clathrin binding energies. In theory, the maximum binding energy due to inter-clathrin interactions amounts to $E_c = 6\epsilon$ per triskelion in a cage. In practice, due to thermal vibrations and the inevitable alignment mismatches in cages formed by rigid identical particles, the average potential energy in the simulations is given by $E_c \approx 4\epsilon$ [30]. The latter relation has been used to rescale the horizontal axes of several phase diagrams in this section for ease of comparison with simulation results. For increasing binding strengths at constant AP concentration, Fig. 2.7(a) shows a narrow transition region (yellow) between virtually no cage formation (dark red) and almost all triskelia absorbed in cages (dark green). A more gradual transition with increasing AP concentration is observed in Fig. 2.7(b). Considering the relative simplicity of the theory, the good agreement between Figs 2.2 and 2.7 is very satisfactory. The theory does not reproduce two properties observed in the simulations: there are no disordered aggregates at high clicking energy, because this transient intermediate state is not included in the theory, and the self-assembly for $\zeta = 10k_BT$ continues down to low AP concentrations. The latter confirms our earlier suspicions that the self-assembly simulations have not reached equilibrium, and agrees with the observation that pre-assembled cages appear stable under these conditions, see the red crossed circles in the top-left of Fig. 2.2(b).

Calculated phase diagrams for the AP model binding to knee and toes are presented in Fig. 2.8, and compare well with the diagram deduced from the simulations, see Fig. 2.5. The striking resemblance between the calculated phase diagrams, compare Figs 2.7 and 2.8, suggests

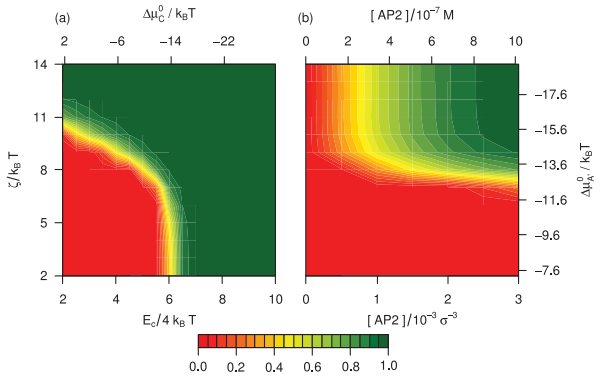


Figure 2.8: Calculated fraction of clathrin bound in cages, for APs clicking to the ends of the TDs and the knees of triskelia, with all other conditions equal to those in Fig. 2.7. These graphs are the theoretical counterparts to the simulation results in Fig. 2.5.

that the sole difference between the two calculations, *i.e.* an AP model that can double click to a single clathrin versus an AP model that can not, is of little consequence to the equilibrium behaviour. The main difference, the slope of the yellow phase boundary in the plots on the left, results from APs double-clicking to triskelia. For APs binding to the knee, intra-clathrin double clicks are impossible. Double clicks are unlikely at moderate click strengths for APs binding to the ankle, because of the free energy penalty in stretching the AP linker, but they become important at high click strengths. The phase diagrams calculated for a reduced linker spring constant of $k = 1\epsilon/\sigma^2$ also agree well with the simulations: the model APs binding to the ankle do not induce cage assembly, while the model APs binding to the knee continue to function (data not shown). Collectively, these results provide strong support for the theory and the underlying concepts on the mechanism of cage stabilization by APs.

Under experimental conditions, the binding strengths E_c and ζ are typically unknown constants, whose values are co-determined by the acidity and salt conditions of the solvent, while the concentrations are readily varied. Four assembly phase diagrams pertaining to various binding strengths are presented in Fig. S6.

To facilitate comparison with experiments, the data are presented in terms of the standard chemical potential difference of AP single-clicking to clathrin, $\Delta\mu^0_{A'}$, as defined in Eq. (2.23), and the standard chemical potential difference of the formation of AP-free cages, $\Delta\mu^0_C$ as defined in Eq. (2.33). At $\Delta\mu^0_C = -13.8k_BT$, see Fig S6(a), the triskelia readily aggregate in the absence of APs at the higher end of the clathrin concentration range; adding APs with $\Delta\mu^0_{A'} = -15.3k_BT$ enhances the cage concentration, but the effect quickly saturates. For the slightly weaker binding triskelia at $\Delta\mu^0_C = -15.8k_BT$, the assistance of APs is crucial to cage formation, with APs binding at $\Delta\mu^0_{A'} = -14.3k_BT$ yielding significantly more cages than APs clicking at $\Delta\mu^0_{A'} = -13.3k_BT$, compare Figs. S6(b) and S6(c). An interesting feature is observed at even weaker clathrin bounding, $\Delta\mu^0_C = -7.8k_BT$, in combination with $\Delta\mu^0_{A'} = -15.3k_BT$, see Fig. S6(c), where for a constant overall clathrin concentration, of say $1.7 \cdot 10^{-7}$ M, the concentration of cages at first increases with the overall AP concentration, passes through a maximum, and then decreases with increasing AP concentration. This cross-section is highlighted in Fig. 2.9, along with three profiles at lower and higher clathrin concentrations. A similarly shaped profile was obtained by the *in vitro* assembly experiments of Zaremba and Keen [35], but there the assembled protein mass fraction is plotted; curves of this type are also included in Fig. 2.9. These authors explain the local maximum as a saturation effect, with clathrin becoming the

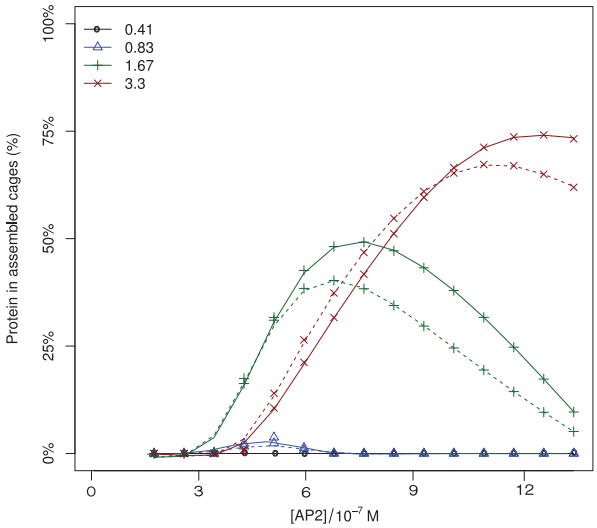


Figure 2.9: Calculated number fraction of clathrin (solid lines) and weight fraction of protein (dashed lines) in self-assembled cages, as a function of the total AP concentration. The total clathrin concentration is indicated in the legend, in units of 10^{-7} molar, the APs bind to the ankle and TD of clathrin, $\Delta\mu^0_{\mathrm{C}} = -11.8 k_B T$ and $\Delta\mu^0_{\mathrm{A}'} = -15.3 k_B T$. Note that the fractions bound in cages do not increase monotonically but pass through a maximum, for reasons explained in the main text.

limiting component upon increasing the AP concentration. This effect is visible in the curves for $\left[\mathrm{C}\right]_{\mathrm{t}} = 3.3 \cdot 10^{-7}\,\mathrm{M}$, which saturates in the fraction of bound clathrin but decays in the fraction of bound protein. Our calculations provide an additional explanation for a maximum in the assembled mass fraction: the number of cages decreases beyond an optimum AP concentration. The underlying mechanism is the replacement of double-clicked APs with two single-clicked APs each, thereby weakening the integrity of cages. Hence increasing the AP concentration beyond its optimum results in a reduction of the cage concentration, as can be clearly seen for $\left[\mathrm{C}\right]_{\mathrm{t}} = 1.7 \cdot 10^{-7}\,\mathrm{M}$.

Plots of the number of APs bound to cages, normalized by the number of triskelia in a cage, are presented in Fig. 2.10. Since the cages are nearly saturated with double-clicked APs for the phase point explored in Fig. 2.9, we opted to present results for the chemical potential difference combinations in Fig. S6(b) and (c). Markers are plotted for cage concentrations exceeding $3 \cdot 10^{-10}\,\mathrm{M}$, which corresponds to one cage in the simulated system. At this threshold, the average number of double-clicked APs per encaged clathrin equals about one, *i.e.* a clathrin is clicked to two cross-linking APs on average, while the average number of single-clicked APs is substantially lower. With increasing AP concentration, the number of double-clicked APs rises with about the same slope as the number of single-clicked APs for $\Delta\mu^0_{\mathrm{A}'} = -13.3 k_B T$, while for higher click strengths the number of double-clicked increases more than the number of single-clicked. In addition to the growing number of APs per cage, the number of cages also rises over the range of AP concentrations. For $\Delta\mu^0_{\mathrm{A}'} = -15.3 k_B T$, the number of single-clicked only starts to deviate from zero when the number of double-clicked APs levels off, at about 2.7 AP per triskelion. These turning points coincide with the number of cages leveling off to a broad maximum, akin to those in Fig. 2.9.

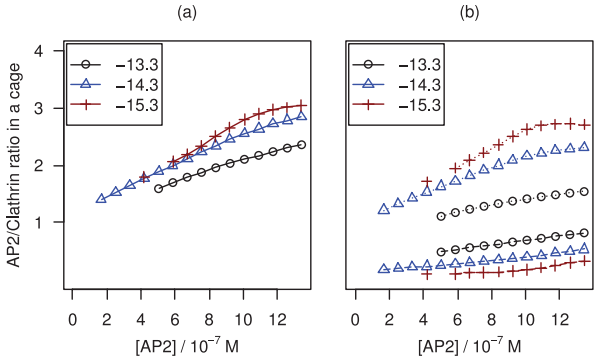


Figure 2.10: Calculated average AP-to-clathrin ratios for cages (top) and subdivision (bottom) into single-clicked (dashed lines) and double-clicked (solid lines), as functions of the total AP concentration. The clicking standard chemical free energy difference is indicated in the legend, in units of k_BT, the APs bind to the ankle and TD of clathrin, $\Delta\mu^0_{\mathrm{C}} = -11.8k_BT$ and $\left[\mathrm{C}\right]_{\mathrm{t}} = 10^{-3}\sigma^{-3}$.

2.5 Conclusions

A coarse-grained simulation model and a theory were developed to study the AP-induced self-assembly of triskelia into cages. The results of both approaches are in line with the experimental data, and provide a better understanding of how APs regulate the assembly of cages. This study reveals a number of restrictions on functional APs. Clearly, APs must bind clathrin sufficiently strong to bring two triskelia together, but cage formation is frustrated when APs bind too strongly. The flexible linker between the two binding sites of an AP must be long enough for inter-triskelion connections in cages, but the linker should not be too long to avoid intra-triskelion bonding. On a related note, the effective spring constant of the linker must be weak enough to allow inter-triskelion connections in cages, but not too weak too suppress intra-triskelion bonding. And the AP to clathrin ratio must be high enough, though not too high. While the numerical values used in the model and theory are based on AP2, we expect the results to apply to all types of APs. For the advancement of simulation models and theories, as well as for an improved understanding of the thermodynamics of coat and vesicle formation during endocytosis, it would be useful to obtain experimental values of all binding constants involved, as well as of the mechanical properties of the AP linker. One way of measuring these parameters is proposed in Appendix 2.6.3.

2.6 Appendices

2.6.1 Clathrin - AP complexes

In these appendices, expressions are derived for the reaction equilibrium constants of AP binding to a triskelion, clathrin self-assembly into cages and AP-induced cage assembly, respectively. We start by considering a mixture of clathrin (C) and adaptor (A) proteins in equilibrium with their supra-molecular aggregates by reactions of the type

$$\mathrm{C} + (n+m)\mathrm{A} \rightleftharpoons \mathrm{CA}'_n\mathrm{A}''_m, \tag{2.9}$$

where the primes represent the number of clicks binding an AP to the clathrin, in this case n single clicked and m double clicked APs. Because clathrin has 6 binding spots for AP, each capable of binding at most one AP, it follows that $n \geq 0$, $m \geq 0$ and $n + 2m \leq 6$. For simplicity, we assume these six sites to have identical binding properties. Likewise, the two clicking sites of AP are assumed to have identical properties, except for their specificity to either the TD or the ankle/knee binding site of clathrin. The equilibrium constant of the above reaction can be defined as [50]

$$K^{\mathrm{tri}}_{n,m} = \frac{\left(\left[\mathrm{CA}'_n\mathrm{A}''_m\right]/c_0\right)}{\left(\left[\mathrm{C}\right]/c_0\right)\left(\left[\mathrm{A}\right]/c_0\right)^{n+m}}, \tag{2.10}$$

with the square brackets denoting concentrations, *i.e.* particles per unit of volume, and c_0 a reference concentration typically taken to be 1 molar. From the statistical mechanics of reaction equilibria in ideal mixtures [50–52] follows that

$$K^{\mathrm{tri}}_{n,m} = \frac{(q_{n,m}/V)}{(q_{\mathrm{C}}/V)\,(q_{\mathrm{A}}/V)^{n+m}} c_0^{n+m} = e^{-\beta\Delta G^0_{n,m}} \tag{2.11}$$

where q_{C}, q_{A} and $q_{n,m}$ are the molecular partition functions of unbound clathrin, unbound AP and the $\mathrm{CA}'_n\mathrm{A}''_m$ supra-molecule, respectively, and $\Delta G^0_{n,m}$ is the standard state free energy change of the reaction.

The semi-classical partition function of a rigid clathrin particle in an infinitely dilute solution, *i.e.* in the limit that non-bonded interactions can be ignored, is given by

$$q_{\mathrm{C}} = \frac{1}{\Delta_{\mathrm{C}}} \iint e^{-\beta\Phi} d\mathbf{r} d\boldsymbol{\varphi} \approx \frac{8\pi^2 V}{\Delta_{\mathrm{C}}} e^{-\beta\Phi_{\mathrm{C}}}, \tag{2.12}$$

with Φ the interaction potential and Φ_{C} the average solvatation free energy of a clathrin. The position integrals run over the volume V of the system and the 3D orientation angles run over their entire range, *e.g.* for the Euler angles $\varphi_1 \in [0, 2\pi)$, $\varphi_2 \in [0, \pi)$, $\varphi_3 \in [0, 2\pi)$ and $d\boldsymbol{\varphi} = \sin\varphi_2 d\varphi_1 d\varphi_2 d\varphi_3$. The elementary volume element Δ_{C} follows from

$$\Delta_{\mathrm{C}^{-1}} = \left(\frac{2\pi k_B T}{h^2}\right)^3 m_{\mathrm{C}}^{3/2} \left|\mathbf{I}_{\mathrm{C}}\right|^{1/2} \frac{1}{\sigma_{\mathrm{C}}}, \tag{2.13}$$

with h denoting Planck's constant, m_{C} and $\mathbf{I}_{\mathrm{C}}$ the mass and inertia tensor of a triskelion, respectively, the brackets $|\ldots|$ denote a determinant, and where the symmetry number σ_{C} has the value 3 for a particle with a three-fold rotational axis. Note that h, m_{C} and $\mathbf{I}_{\mathrm{C}}$ do not enter the MC simulations, hence the theoretical and simulated equilibrium constants will only agree if these factors can be made to cancel out in the final expression. Treating an AP protein as two point particles of type a, one obtains at infinite dilution

$$q_{\mathrm{A}} = \frac{1}{\Delta_{\mathrm{a}}^2} \iint e^{-\beta\Phi} d\mathbf{r}_1 d\mathbf{r}_2 \approx \frac{1}{\Delta_{\mathrm{a}}^2} V q_{\mathrm{s}}\, e^{-\beta\Phi_{\mathrm{A}}}, \tag{2.14}$$

with $\Delta_{\rm a} = h^3/(2\pi m_{\rm a} k_B T)^{3/2}$ the elementary volume element per particle, $\Phi_{\rm A}$ the average solvation free energy of an AP, and

$$q_{\rm s} = 4\pi \int_0^\infty e^{-\beta\psi(r_{12})} r_{12}^2 dr_{12} \tag{2.15}$$

the contribution of the internal spring, with potential energy $\psi(r_{12})$ at elongation r_{12}, to the partition function. The integral is readily solved for a Hookean spring with spring constant k, yielding $q_{\rm s} = (2\pi k_B T/k)^{3/2}$. Next, the partition function of a clathrin adorned with one single-clicked AP takes the form

$$\begin{aligned} q_{1,0} &= \frac{1}{\Delta_{\rm C}\Delta_{\rm a}^2} \iiiint e^{-\beta\Phi} d\mathbf{r} d\boldsymbol{\varphi} d\mathbf{r}_1 d\mathbf{r}_2 \\ &\approx \mu_{1,0} \frac{8\pi^2 V}{\Delta_{\rm C}} \frac{4\pi\rho^3 q_{\rm s}}{3\Delta_{\rm a}^2} e^{-\beta(\Phi_{\rm C}+\Phi_{\rm A}-\zeta)}, \end{aligned} \tag{2.16}$$

where in the last step it has been used that either site of the AP dimer must be within a sphere of radius ρ centered around a clicking site on the triskelion, and ζ denotes the strength of the click. The number of clicking combinations will be denoted by $\mu_{n,m}$, and in the current case has the value $\mu_{1,0} = 6$ because a triskelion offers six binding spots. Note that the AP-clathrin complex is not treated as a single molecule, but as a combination of two molecules with reduced rotational and translational freedom [53, 54]. By combining the above equations, one arrives at the equilibrium constant

$$K_{1,0}^{\rm tri} = \mu_{1,0} \frac{4}{3}\pi\rho^3 e^{\beta\zeta} c_0 \tag{2.17}$$

where the elementary volumes have indeed canceled out. The approach is readily extended to several single-clicked APs per triskelion, with at most one AP per triskelion binding site, under the assumption that other interactions between these APs may be ignored. Figure S5 shows that the theory is in good agreement with the simulations.

The partition function of a triskelion adorned with one double-clicked AP is given by an integral similar to that in Eq. (2.16), with the restriction that now both sites of the AP must be clicked to their counter-part sites on the triskelion. In view of the estimated maximum extensibility of the AP linker, $L \approx 1.5\sigma$, a double clicked AP will bind to two triskelion sites on the same leg and hence their interstitial distance is fixed, $d_t = 1\sigma$. We then arrive at

$$q_{0,1} \approx \mu_{0,1} \frac{8\pi^2 V}{\Delta_{\rm C}} \frac{q_{\rm s}''(d_t)}{\Delta_{\rm a}^2} e^{-\beta(\Phi_{\rm C}+\Phi_{\rm A}-2\zeta)}, \tag{2.18}$$

where $\mu_{0,1} = 3$ and the contribution of AP's internal spring reads as

$$q_{\rm s}''(d) = \int_{v_1}\int_{v_2} e^{-\beta\psi(|\mathbf{r}_1-\mathbf{r}_2|)} d\mathbf{r}_1 d\mathbf{r}_2, \tag{2.19}$$

with v_1 and v_2 denoting the spherical volumes of radius ρ of two clicking sites at centre-to-centre distance d. For the actual proteins, the range of the click interaction is short compared to the distance between clicking sites, $\rho \ll d$, and the integral may be approximated as

$$q_{\rm s}''(d) \approx \left(\frac{4}{3}\pi\rho^3\right)^2 e^{-\beta\psi(d)} \tag{2.20}$$

in the limit of $\beta k d\rho \ll 1$. Figure S5 shows that the resulting equilibrium constant $K_{0,1}^{\rm tri}$, is in good agreement with the simulations, for the low $k = 1 k_B T/\sigma^2$ value used in that plot. The

combination of spring constant $k = 30k_BT/\sigma^2$ and click radius $\rho = 0.25\sigma$ used in the production simulations exceeds this limit and it proved necessary to evaluate the integral of Eq. (2.19) numerically, yielding a value $q_{\mathrm{s}}''(\sigma) \approx 9.6 \cdot 10^{-8}\sigma^6$ about two orders larger than the estimate $1.0 \cdot 10^{-11}\sigma^6$ by Eq. (2.20), to obtain a good agreement between theoretical and simulation phase diagrams.

The above results can be combined to obtain the equilibrium constants for all reactions of the type expressed in Eq. (2.9), in the dilute limit. Upon neglecting interactions between APs bound to the same triskelion, except for the mutual exclusivity of the clathrin-AP clicks, we arrive at

$$K_{n,m}^{\mathrm{tri}} = \mu_{n,m} \left(\frac{4}{3}\pi\rho^3\right)^n \left(\frac{q_{\mathrm{s}}''(d_t)}{q_{\mathrm{s}}}\right)^m e^{\beta(n+2m)\zeta} c_0^{n+m}. \tag{2.21}$$

The multiplicity $\mu_{n,m}$ is readily established by counting the number of ways of attaching n single clicked and m double clicked APs to a single triskelion, but in practice this number proves of little consequence because the other factors in the above equation are much larger. Upon neglecting this factor, the standard state free energy differences [50, 52] for the 15 possible (n, m) combinations with $n + m > 0$ reduce to

$$\Delta G_{n,m}^0 = \mu_{n,m}^0 - \mu_{\mathrm{C}}^0 - (n+m)\mu_{\mathrm{A}}^0 \tag{2.22}$$

$$\approx n\Delta\mu_{\mathrm{A}'}^0 + m\Delta\mu_{\mathrm{A}''}^0(d_t), \tag{2.23}$$

with μ_i^0 the reference chemical potential of compound i at the reference concentration c_0, and where the reference chemical potential differences follow from Eq. (2.21) as

$$\Delta\mu_{\mathrm{A}'}^0 = -k_BT \ln\left(\frac{4}{3}\pi\rho^3 e^{\beta\zeta} c_0\right), \tag{2.24}$$

$$\Delta\mu_{\mathrm{A}''}^0(d) = -k_BT \ln\left(\frac{q_{\mathrm{s}}''(d)}{q_{\mathrm{s}}} e^{\beta 2\zeta} c_0\right). \tag{2.25}$$

Inserting the parameters of the simulation model into the former difference yields

$$\Delta\mu_{\mathrm{A}'}^0 \approx -5.3k_BT - \zeta, \tag{2.26}$$

while in combination with the approximation in Eq. (2.20), the latter difference can be rewritten as

$$\Delta\mu_{\mathrm{A}''}^0(d) = 2\Delta\mu_{\mathrm{A}'}^0 + \psi(d) + k_BT \ln\left[c_0 \left(2\pi k_BT/k\right)^{3/2}\right], \tag{2.27}$$

while with the numerical evaluation of $q_{\mathrm{s}}''(d_t)$ we find for the simulation model

$$\Delta\mu_{\mathrm{A}''}^0(d_t) = 2\Delta\mu_{\mathrm{A}'}^0 + 16.4k_BT. \tag{2.28}$$

These expressions are readily extended to include site-dependent clicking strengths, *i.e.* ζ_1 for binding to the feet and ζ_2 for binding to the ankle or knee.

2.6.2 Clathrin cages

The partition function of a clathrin cage of p triskelia is obtained by integrating over the positions $\mathbf{r}$ and orientations $\boldsymbol{\varphi}$ of all p triskelia, subject to the condition that the particles remain sufficiently close and properly oriented – relative to each other – to qualify as a cage. The overall translational and rotational freedom of a triskelion – amounting to V and $8\pi^2$, respectively, for

a particle in solution, see Eq. (2.12) – are effectively reduced by these binding restrictions to v_t and ω_r, respectively, for a triskelion wobbling around a fixed location in a cage. The partition function of a cage of p triskelia can therefore be approximated as

$$q_p = \frac{1}{p!\Delta_{\mathrm{C}}^p} \iint \cdots \iint e^{-\beta\Phi} d\mathbf{r}_1 d\boldsymbol{\varphi}_1 \cdots d\mathbf{r}_p d\boldsymbol{\varphi}_p \tag{2.29}$$

$$\approx \mu_p \frac{8\pi^2 V}{\Delta_{\mathrm{C}}^p} (v_t \omega_r)^{p-1} e^{-\beta p(\Phi_{\mathrm{C}} - E_c)}, \tag{2.30}$$

where E_c denotes the average binding energy of a clathrin in a cage, and $p!$ corrects for the indistinguishability of the p triskelions forming the cage. In evaluating the integral, one particle has retained the full factor $8\pi^2 V$ to account for the translational and rotational freedom that a rigid coat will sample, while the remaining $(p-1)$ particles each contribute a factor $v_t\omega_r$ reflecting thermal fluctuations around this rigid shape. The multiplicity μ_p denotes the degeneracy of the ground state. Cages with pentagonal and hexagonal facets require an even p; there exists one cage structure for $p = 20$, none for $p = 22$ and multiple cage structures for $p \geq 24$. Schein *et al.* [48, 55] argued that, for $20 \leq p \leq 60$, there typically exists just one preferred cage structure for every even value of p, since all other cages incorporate one or more edges with an unfavourably high torsional energy. This theory is confirmed by the cages spontaneously grown in our simulations. We note that the 'exclusion of head-to-tail dihedral angle discrepancies' rule proposed by Schein *et al.* can be expressed much more concisely as the 'excluded 5566' rule: an unfavourable torsion arises when a facet has among its neighbouring facets a sequence of two pentagons followed by two hexagons, regardless of clockwise or counter-clockwise order. The 'isolated pentagon rule' applies for $p > 60$, and there typically exist multiple favourable cages for $p \geq 70$. [48, 55] Since the multiplicity is a small integer for the $p \approx 60$ cages grown in the simulations, the exact value of μ_p proves to be of little consequence to the results of the calculations.

Combining the above results, the equilibrium constant for the cage formation reaction

$$p\mathrm{C} \rightleftharpoons \mathrm{C}_p \tag{2.31}$$

is found to be given by

$$K_p^{\mathrm{cage}} = \mu_p \left(\frac{v_t\omega_r}{8\pi^2}\right)^{p-1} e^{\beta p E_c} c_0^{p-1}. \tag{2.32}$$

The corresponding standard free energy difference can be expressed as

$$\Delta G_p^0 = -k_B T \ln K_p^{\mathrm{cage}} \approx p\Delta\mu_{\mathrm{C}}^0 \tag{2.33}$$

$$\Delta\mu_{\mathrm{C}}^0 \approx -k_B T \ln\left(\frac{v_t\omega_r}{8\pi^2} e^{\beta E_c} c_0\right), \tag{2.34}$$

for $p \gg 1$. Assuming that the simulated triskelia bound in a cage experience an estimated translational freedom of 0.1σ along every Cartesian direction and an estimated rotational freedom of 0.1 rad ($\sim 6°$) around every Cartesian axis,

$$\Delta\mu_{\mathrm{C}}^0 \approx 10.2 k_B T - E_c. \tag{2.35}$$

The resulting fraction of clathrin bound in cages, $f = p[\mathrm{C}_p]/[\mathrm{C}]_{\mathrm{t}}$, is plotted in Fig. 2.11 as a function of the total clathrin concentration, $[\mathrm{C}]_{\mathrm{t}} = [\mathrm{C}] + p[\mathrm{C}_p]$. This fraction reaches a value of 50%, *i.e.* the number of bound triskelia equals the number of unbound triskelia, when the total concentration equals the critical assembly concentration (CAC); the first cages appear at about half this overall concentration. The above equilibrium constant can be related to the CAC, and

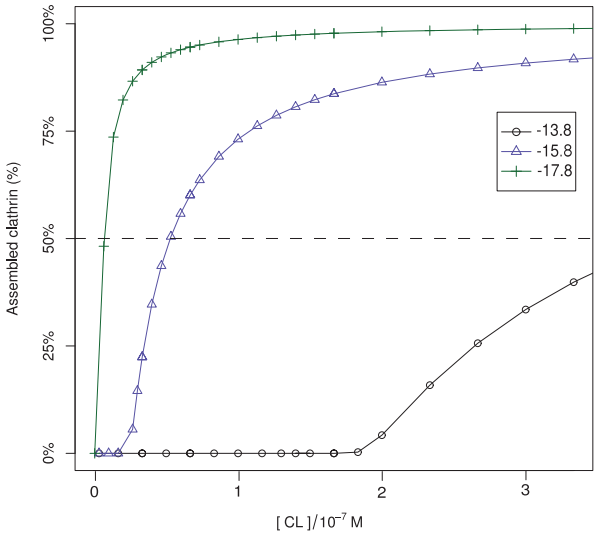


Figure 2.11: Theoretical concentration dependence of the fraction of particles bound in cages, for several values of the standard chemical potential difference $\Delta\mu^0_{\rm C}$, indicated in units of k_BT in the legend, in the absence of APs. At the critical assembly concentration (CAC), which varies with the interaction strength, the concentrations of free and bound clathrin are equal (dashed line). In this calculation, all cages are assumed of identical size, $p = 60$. Note the strong resemblance to the experimental data on *in vitro* assembly of clathrin cages, see Fig. 9 in Ref. [56].

thence to experimental data on clathrin. At the CAC, the number density of free triskelia reads as $\left[\mathrm{C}\right] = c_{\rm CAC}/2$ and that of cages as $\left[\mathrm{C}_p\right] = c_{\rm CAC}/(2p)$, hence

$$K_p^{\rm cage} = \frac{1}{p}\left(\frac{c_{\rm CAC}}{2c_0}\right)^{1-p}, \tag{2.36}$$

$$\Delta\mu^0_{\rm C} \approx k_BT \ln\frac{c_{\rm CAC}}{2c_0}, \tag{2.37}$$

for $p \gg 1$. The experimental CAC of $100\,\mu$g/ml [45] then translates into $\Delta\mu^0_{\rm C} \approx -16.4k_BT$, and this value is reproduced by the simulation model for $E_c \approx 27k_BT$. In simulations with an overall triskelion density close to the experimental CAC, the numbers of bound and unbound particles were about equal when using a leg-leg interaction strength $\epsilon \approx 6k_BT$; the resulting average potential energy of $\sim 23k_BT$ per clathrin [30] is in good agreement with the above theoretical estimate. We note that the elementary volume elements $\Delta_{\rm C}$ have again cancelled out in the statistical mechanical expression for the equilibrium constant. This was not the case in our earlier derivation, which consequently overestimated the binding energy [30]. Muthukumar and Nossal [46] presented a derivation based on mole fractions, following the common practice in micelle theory [57], to arrive at an enthalpic energy $E_c \approx k_BT \ln c_{\rm s}/c_{\rm CAC} \approx 21k_BT$, with the subscript s referring to the solvent. There is no compelling physical reason to use mole fractions, and one now sees that the method works because the volume per solvent molecule, $v_{\rm s} = 1/c_{\rm s}$, provides a reasonable order of magnitude estimate for the libration volume of a clathrin bound in a cage, $v_t\omega_r$.

2.6.3 Decorated clathrin cages

Finally, we consider the formation of a cage decorated with n single-clicked APs and m double-clicked APs,

$$p\mathrm{C} + (n+m)\mathrm{A} \rightleftharpoons \mathrm{C}_p\mathrm{A}'_n\mathrm{A}''_m. \tag{2.38}$$

To keep the derivation manageable, it is assumed that for every clicking site on a clathrin in a cage there is one nearest clicking site on an adjacent clathrin in that cage, such that the two sites – and hence the two triskelia – can be linked by an AP. Distance measurement reveal that the separation between two nearest sites on differing triskelia in a cage, d_c, is shorter than the distance d_t between two nearest sites on the same triskelion. Because of the functional forms of q''_s and ψ, a small reduction of the elongation of the entropic spring results in a pronounced increase of q''_s – we may therefore ignore intra-clathrin double-clicked APs. Combining the results from the two preceding appendices, we then arrive at the equilibrium constant

$$K^{\mathrm{cage}}_{p,n,m} = \mu_{p,n,m}\left(\frac{v_t\omega_r}{8\pi^2}\right)^{p-1}\left(\frac{4}{3}\pi\rho^3\right)^n\left(\frac{q''_\mathrm{s}(d_c)}{q_\mathrm{s}}\right)^m e^{\beta[pE_c+(n+2m)\zeta]}c_0^{p+n+m-1}, \tag{2.39}$$

where $n \geq 0$, $m \geq 0$ and $n + 2m \leq 6p$. Again, the elementary volume elements have cancelled out. The multiplicity is estimated as

$$\mu_{p,n,m} \approx \mu_p \frac{(3p)!}{m!(3p-m)!}\frac{(6p-2m)!}{n!(6p-2m-n)!}, \tag{2.40}$$

where the first factor, accounting for the cage structure, has been discussed before, $\mu_p \approx 1$, the second factor counts the permitted distributions of m double-clicked APs over $3p$ pairs of nearest unlike click sites in a cage, and the third factor represents the permitted distributions of n single-clicked APs over the remaining $6p - 2m$ free clicking sites of the cage. The standard free energy difference of the reaction can be expressed as

$$\Delta G^0_{p,n,m} \approx p\Delta\mu^0_\mathrm{C} + n\Delta\mu^0_{\mathrm{A}'} + m\Delta\mu^0_{\mathrm{A}''}(d_c) - k_BT\ln\mu_{p,n,m}, \tag{2.41}$$

where the multiplicity is no longer negligibly small. The extension to site-dependent clicking strengths is again straight forward.

To obtain the number of cages at every point in the assembly diagrams of Figs 2.7 through 2.10, we consider a closed system of volume V with given total clathrin concentration $\left[\mathrm{C}\right]_\mathrm{t}$ and AP concentration $\left[\mathrm{A}\right]_\mathrm{t}$. For simplicity, we again consider only one cage size, $p = 60$. We denote the estimated concentrations of free, *i.e.* unbound, triskelia and AP as $\left[\mathrm{C}\right]_\mathrm{f}$ and $\left[\mathrm{A}\right]_\mathrm{f}$, respectively. The concentrations of decorated triskelia and decorated cages then follow by using the equilibrium constants derived above. A weighted sum over all species yields the sum concentrations of triskelia and APs present in the box,

$$\begin{aligned}\left[\mathrm{C}\right]_\mathrm{s} &= \left[\mathrm{C}\right]_\mathrm{f} + \sum_{n,m} K^{\mathrm{tri}}_{n,m}\left[\mathrm{C}\right]_\mathrm{f}\left[\mathrm{A}\right]_\mathrm{f}^{n+m} \\ &\quad + p\sum_{n,m} K^{\mathrm{cage}}_{p,n,m}\left[\mathrm{C}\right]_\mathrm{f}^p\left[\mathrm{A}\right]_\mathrm{f}^{n+m},\end{aligned} \tag{2.42}$$

$$\begin{aligned}\left[\mathrm{A}\right]_\mathrm{s} &= \left[\mathrm{A}\right]_\mathrm{f} + \sum_{n,m} (n+m)\, K^{\mathrm{tri}}_{n,m}\left[\mathrm{C}\right]_\mathrm{f}\left[\mathrm{A}\right]_\mathrm{f}^{n+m} \\ &\quad + \sum_{n,m} (n+m)\, K^{\mathrm{cage}}_{p,n,m}\left[\mathrm{C}\right]_\mathrm{f}^p\left[\mathrm{A}\right]_\mathrm{f}^{n+m}.\end{aligned} \tag{2.43}$$

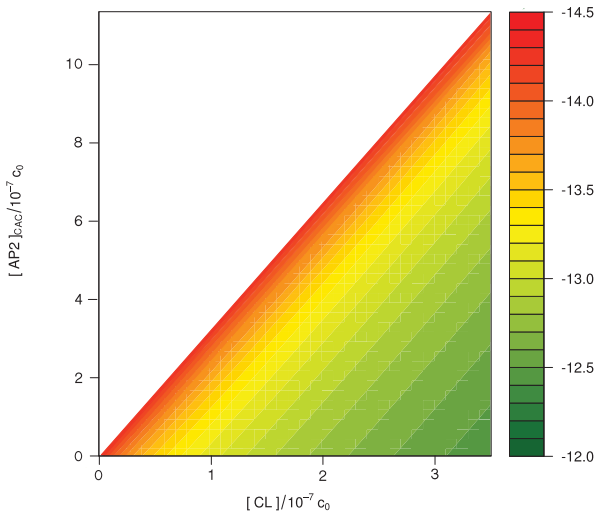


Figure 2.12: For any given total clathrin concentration (horizontal axis), there exists a total AP concentration (vertical axis) for which the fraction of bound AP equals the fraction of unbound AP. The plot relates both concentrations to the corresponding standard chemical potential difference for single-clicked APs, $\Delta\mu^0_{\mathrm{A}'}$ in units of k_BT. The underlying calculation is based on $\Delta\mu^0_{\mathrm{C}} = -11.8k_BT$ and assumes $\Delta\mu^0_{\mathrm{A}''}$ and $\Delta\mu^0_{\mathrm{A}'}$ to be related by Eqs. (2.28) and (2.44). The diagonal represents the maximum attainable AP CAC for a given clathrin concentration.

We now solve the equations $\left[\mathrm{C}\right]_{\mathrm{s}} = \left[\mathrm{C}\right]_{\mathrm{t}}$ and $\left[\mathrm{A}\right]_{\mathrm{s}} = \left[\mathrm{C}\right]_{\mathrm{t}}$ by varying $\left[\mathrm{C}\right]_{\mathrm{f}}$ and $\left[\mathrm{A}\right]_{\mathrm{f}}$. This is achieved by minimizing the sum-squared differences between the imposed and calculated overall concentrations, followed by a Newton-Raphson zero-point solver [58]. The second summation on the r.h.s. of Eq. (2.42) then yields the concentration of cages in the system.

The presented calculations assume $d_c = 0.6\sigma$; since the limit underlying Eq. (2.20) does not hold true for the simulation model, the integral in Eq. (2.19) is solved numerically to yield $q''_{\mathrm{s}}(d_c) \approx 7.8 \cdot 10^{-5}\sigma^6$. The standard chemical potential difference of single and double clicked APs are then related by

$$\Delta\mu^0_{\mathrm{A}''}(d_c) = 2\Delta\mu^0_{\mathrm{A}'} + 9.7k_BT, \tag{2.44}$$

while Eq. (2.27) predicts a slightly higher offset of $11.5k_BT$.

Measuring the hundreds of reaction equilibrium constants condensed into Eqs. (2.39) and (2.41) offers a daunting challenge, the more so as this requires identifying the numbers of triskelia, single clicked and double clicked APs of every cage in the reaction mixture. To facilitate comparison of theory with experiment, we revert to a CAC. The condition where half the APs in a solution are bound to triskelia – both free triskelia and those in cages – is a complex function of all equilibrium constants of the mixture, as well as of the overall clathrin concentration, but it can be measured without detailed resolution of the compositions of aggregates. By measuring this CAC at a number of clathrin concentrations, it is in principle possible to extract all relevant standard chemical potential differences. An illustration hereof is provided in Fig. 2.12, where it is assumed that $\Delta\mu^0_{\mathrm{C}}$ and the relation between $\Delta\mu^0_{\mathrm{A}''}(d)$ and $\Delta\mu^0_{\mathrm{A}'}$ are known in advance; by measuring AP's CAC at a given overall clathrin concentration, one can read the corresponding $\Delta\mu^0_{\mathrm{A}'}$ from the graph.

Acknowledgments

This work is part of the research programme 'self-assembly of protein coats at membranes' (project nr. 711.012.004) which is financed by the Netherlands Organisation for Scientific Research (NWO).

References

[1] S. J. Royle, *Cellular and molecular life sciences* **2006**, *63*, 1823–1832.

[2] A. Young in Seminars in cell & developmental biology, *Vol. 18*, Elsevier, **2007**, pp. 448–458.

[3] H. T. McMahon, E Boucrot, *Nat. Rev. Mol. Cell Biol.* **2011**, *12*, 517–533.

[4] F. M. Brodsky, *Annual review of cell and developmental biology* **2012**, *28*, 309.

[5] T. Kirchhausen, D. Owen, S. C. Harrison, *Cold Spring Harbor perspectives in biology* **2014**, *6*, a016725.

[6] M. S. Robinson, *Traffic* **2015**, *16*, 1210–1238.

[7] E. Ungewickell, D. Branton, *Nature Publishing Group* **1981**.

[8] B. M. Pearse, C. J. Smith, D. J. Owen, *Current opinion in structural biology* **2000**, *10*, 220–228.

[9] L. Maldonado-Báez, B. Wendland, *Trends in cell biology* **2006**, *16*, 505–513.

[10] E. M. Schmid, H. T. McMahon, *Nature* **2007**, *448*, 883.

[11] Y. Xing, T. Böcking, M. Wolf, N. Grigorieff, T. Kirchhausen, S. C. Harrison, *The EMBO journal* **2010**, *29*, 655–665.

[12] L. M. Traub, J. S. Bonifacino, *Cold Spring Harbor perspectives in biology* **2013**, *5*, a016790.

[13] J. H. Keen, *The Journal of cell biology* **1987**, *105*, 1989–98.

[14] T. Kirchhausen, *Cell* **2002**, *109*, 413–416.

[15] D. J. Owen, B. M. Collins, P. R. Evans, *Annual Review of Cell and Developmental Biology* **2004**, *20*, 153–191.

[16] M. S. Robinson, *Trends in cell biology* **2004**, *14*, 167–174.

[17] M. A. Edeling, C. Smith, D. Owen, *Nature Reviews Molecular Cell Biology* **2006**, *7*, 32–44.

[18] N. Popova, I. Deyev, A. Petrenko, *Acta naturae* **2013**, *5*, 62.

[19] D. Owen, Y Vallis, B. Pearse, H. McMahon, P. Evans, *The EMBO journal* **2000**, *19*, 4216–4227.

[20] J Rappoport, *Biochem. J* **2008**, *412*, 415–423.

[21] B. M. Collins, A. J. McCoy, H. M. Kent, P. R. Evans, D. J. Owen, *Cell* **2002**, *109*, 523–535.

[22] L. P. Jackson, B. T. Kelly, A. J. McCoy, T. Gaffry, L. C. James, B. M. Collins, S. Höning, P. R. Evans, D. J. Owen, *Cell* **2010**, *141*, 1220–1229.

[23] B. Kelly, S. Graham, N Liska, *Science* **2014**, *345*, 459.

[24] R. Lindner, *Interaktionen der Hüllstrukturproteine von Clathrin-coated vesicles.* Technische Universität München, **1992**.

[25] B Greene, S. Liu, A Wilde, F. Brodsky, *Traffic* **2000**.

[26] J. H. Keen, K. A. Beck, T Kirchhausen, T Jarrett, *J. Biol. Chem.* **1991**, *266*, 7950–7956.

[27] C Knuehl, C. Y. Chen, V Manalo, P. K. Hwang, N Ota, F. M. Brodsky, *Traffic* **2006**, *7*, 1688–1700.

[28] W. K. den Otter, M. R. Renes, W. J. Briels, *J. Phys. Condens. Matt.* **2010**, *22*, 104103.

[29] W. K. den Otter, M. R. Renes, W. J. Briels, *Biophys. J.* **2010**, *99*, 1231–1238.

[30] W. K. den Otter, W. J. Briels, *Traffic* **2011**, *12*, 1407–1416.

[31] I. M. Ilie, W. K. den Otter, W. J. Briels, *J. Chem. Phys.* **2014**, *141*, 065101.

[32] R. Matthews, C. N. Likos, *Soft Matter* **2013**, *9*, 5794–5806.

[33] N. Cordella, T. J. Lampo, S. Mehraeen, A. J. Spakowitz, *Biophysical journal* **2014**, *106*, 1476–1488.

[34] J. J. VanDersarl, S. Mehraeen, A. P. Schoen, S. C. Heilshorn, A. J. Spakowitz, N. A. Melosh, *Soft matter* **2014**, *10*, 6219–6227.

[35] S. Zaremba, J. H. Keen, *The Journal of cell biology* **1983**, 1339–1347.

[36] A. Fotin, Y. Cheng, P. Sliz, N. Grigorieff, S. C. Harrison, T. Kirchhausen, T. Walz, *Nature* **2004**, *432*, 573–579.

[37] J. D. Wilbur, P. K. Hwang, F. M. Brodsky, *Traffic* **2005**, *6*, 346–350.

[38] T. Böcking, F. Aguet, I. Rapoport, M. Banzhaf, A. Yu, J. C. Zeeh, T. Kirchhausen, *Structure* **2014**, *22*, 819–829.

[39] P. Tompa, *Structure and Function of Intrinsically Disordered Proteins*, Chapman and Hall/CRC, Boca Raton, FL, USA., **2009**.

[40] D. Boal, *Mechanics of the Cell*, Cambridge University Press, Cambridge, U.K., **2002**.

[41] K. Kremer, G. S. Grest, *The Journal of Chemical Physics* **1990**, *92*, 5057–5086.

[42] D. Frenkel, B. Smit, *Understanding Molecular Simulation*, 2nd, Academic Press Inc., Orlando, FL, U.S.A., **2001**.

[43] D. P. Landau, K. Binder, *A guide to Monte Carlo simulations in statistical physics*, Cambridge university press, **2014**.

[44] R. H. Swendsen, J.-S. Wang, *Physical review letters* **1987**, *58*, 86.

[45] R. Crowther, B. Pearse, *The Journal of cell biology* **1981**, *91*, 790–797.

[46] M Muthukumar, R. Nossal, *The Journal of chemical physics* **2013**, *139*, 121928.

[47] M. Saleem, S. Morlot, A. Hohendahl, J. Manzi, M. Lenz, A. Roux, *Nature communications* **2015**, *6*, 6249.

[48] S. Schein, M. Sands-Kidner, *Biophysical journal* **2008**, *94*, 958–976.

[49] A. P. Schoen, N. Cordella, S. Mehraeen, M. A. Arunagirinathan, A. J. Spakowitz, S. C. Heilshorn, *Soft Matter* **2013**, *9*, 9137–9145.

[50] D. Nelson, T. Piran, S. Weinberg, *Statistical Mechanics of Membranes and Surfaces*, 2nd ed., World Scientific Publishing, Singapore, **2004**.

[51] D. A. McQuarrie, *Statistical thermodynamics*, HarperCollins Publishers, **1973**.

[52] R. Phillips, J. Kondev, J. Theriot, H. Garcia, *Physical biology of the cell*, Garland Science, **2012**.

[53] M. F. Hagan, D. Chandler, *Biophysical journal* **2006**, *91*, 42–54.

[54] N. Ben-Tal, B. Honig, C. K. Bagdassarian, A. Ben-Shaul, *Biophysical journal* **2000**, *79*, 1180–1187.

[55] S. Schein, M. Sands-Kidner, T. Friedrich, *Biophysical journal* **2008**, *94*, 938–957.

[56] E Ungewickell, H. Ungewickell, *Journal of Biological Chemistry* **1991**, *266*, 12710–12714.

[57] *Micelles, Membranes, Microemulsions and Monolayers*, (Eds.: W. M. Gelbart, A. Ben-Shaul, D. Roux), Springer Verlag, Berlin, Germany, **1994**.

[58] Numerical recipes in FORTRAN (Cambridge).

Chapter 3

Adaptor proteins shape the size distribution of clathrin cages

Clathrin triskelia, the building block of polyhedral coats during endocytosis, are also able to self-assemble in vitro *under appropriate acidity and salt conditions as well as in the presence of adaptor proteins. It is well known that the latter results in markedly smaller cages, but the origins of this difference are still unclear. We present an equilibrium statistical mechanical model of clathrin cages including the flexibility and intrinsic pucker of the triskelion, geometrical considerations on the allowed cage structures and the interaction with single and double bound adaptor proteins. The model is used to analyse the cage size distributions measured by Zaremba and Keen* [J. Cell Biol. **97**, 1339–1347 (1983)] *and thereby reveals that the mechanical properties of adaptor proteins can be explain the different cage size distributions.* [1]

3.1 Introduction

The three identical curved legs connected to a central hub endow the clathrin protein with its characteristic shape and a slight intrinsic curvature, see Fig. 3.1(A) and (B). Adaptor proteins, like the adaptor protein complex 2 (AP2) partaking to clathrin mediated endocytosis (CME), are capable of tethering clathrin triskelia to a membrane and cargo. The aggregated triskelia subsequently polymerize into curved polyhedral lattices with pentagonal and hexagonal facets, the membrane is bent to form a clathrin coated pit (CCP) and eventually the cargo is fully enveloped by a clathrin coated vesicle (CCV) that is detached from the mother membrane by scission proteins [1–6]. Early explorations revealed that AP2 is capable of inducing *in vitro* clathrin assembly in preparations containing only these two proteins [7–9]. Clathrin self-assembly *in vitro*, without the assistance of adaptor proteins, may even be induced under conditions of low salt and slight acidity, [10–12] with highly deformed tiny cages being formed in extreme conditions [13, 14]. Interestingly, the clathrin cages assembled with AP2 are smaller and more uniform than the cages produced in their absence, all other conditions being equal [9]. Similar observations have been reported for the adaptor protein complex 3 [15] (AP3), and for the assembly protein AP180 [16] – note that the hetero-tetramers AP2 and AP3 share a common structure, while both differ markedly from the monomeric AP180 [17, 18]. Adaptor proteins appear to also influence cage dimensions *in vivo*, as the vesicles produced in a squid giant synapse increase in size upon the injection of AP180-inhibitors. The objective of the current study will be to provide a physical explanation for the impact of adaptor proteins on the cage size distribution.

[1] This chapter is being prepared for a publication in a journal.

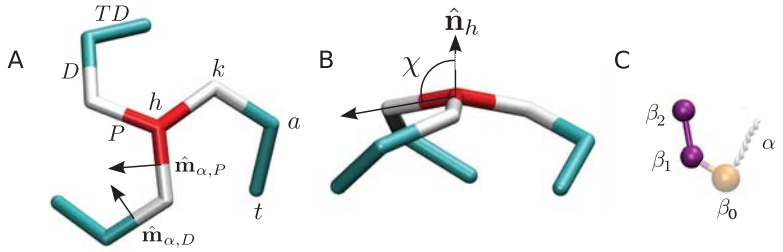


Figure 3.1: Coarse grained models representing (A and B) the clathrin triskelion and (C) the adaptor protein on the same scale, as used in previous work to simulate their collective assembly behavior [19]. The three proximal leg segments (P) of clathrin radiate from a central hip (h) to the knees (k), at a pucker angle χ relative to the symmetry axis $\hat{\mathbf{n}}_h$, followed by distal leg segments (D) running to ankles (a) and terminal domains (TDs) ending at the toes (t). The adaptor protein features two clathrin motifs, with β_1 binding to the toes and β_2 to the ankles, connected by a flexible linker. The full AP2 protein also possesses a core with lipid and cargo motifs and a flexible α linker, which are omitted in the current study as they not partake in clathrin binding.

The intrinsic pucker and the flexibility of the clathrin triskelion, be this flexibility concentrated in the joints or more evenly distributed over the entire length of the legs, evidently plays an important role in determining cage sizes. The mechanical properties of clathrin monomers have been studied by cryo-electron microscopy, [20–22] light scattering, [23] small-angle neutron scattering (SANS), [24] and atomic force microscopy (AFM) [25]. Subtle complications arise when applying these mechanical properties to triskelia in a cage, as the mechanical properties will be affected by the interactions with neighbouring triskelia and the geometrical restrictions imposed by a multitude of ring closures (discussed in more detail below).

The size distributions of clathrin cage have been explored in a number of theoretical studies. Nossal and co-workers explained the distributions in terms of the intrinsic curvature and bending rigidity of clathrin triskelia. In simulations of triskelia with legs truncated at the knees, den Otter *et al.* observed a strong correlation between the rigid pucker of the triskelia and the cage size [26]. The adaptor proteins were ignored in these studies, as well as in most other theoretical and simulation studies of clathrin assembly [27–30]. We recently presented the first simulations exploring the regulatory effect of AP2 on the assembly of (rigid) triskelia, thereby revealing the importance of the mechanical properties of the disordered linker between the two clathrin motifs of AP2 [19]. A statistical mechanical theory based on these insights, assuming a fixed cage size of 60 triskelia, yielded good quantitative agreement with the simulation results, including the ability of AP to induce cage assembly when the inter-clathrin interactions are too weak for unassisted self-assembly [19]. In the current study, this theory is extended to include the intrinsic curvature and flexibility of triskelia as well as geometrical considerations on the permitted cage structures [19, 31, 32] and the distance to be covered by AP2s bound to two triskelia. The values of the independent parameters entering the theory have been determined by fitting the data of Zaremba and Keen. The optimal binding strength of a triskelion in a cage is determined separately, based on the critical assembly concentration (CAC).

Below we present a theoretical method to calculate the equilibrium distribution of clathrin cages of various sizes as a function of the total clathrin and AP concentrations. In cases when only triskelia are available, our approach will be to calculate the equilibrium constants of the

reactions

$$p\mathrm{C} \rightleftharpoons \mathrm{C}_p, \tag{3.1}$$

and to combine these with mass balance equations to obtain the concentrations in equilibrium. To achieve this goal we will calculate the chemical potentials

$$\mu_{\mathrm{C}} = \mu_{\mathrm{C}}^0 + k_B T \ln \frac{[\mathrm{C}]}{c_0} \tag{3.2}$$

$$\mu_p = \mu_p^0 + k_B T \ln \frac{[\mathrm{C}_p]}{c_0}, \tag{3.3}$$

where the square brackets denote number concentrations, *i.e.* number of molecules per unit of volume, and c_0 is a reference concentration typically taken to be 1 molar. We refrain from taking a factor of p out of μ_p^0 because it has no meaning in the generalized situation treated later. Notice that we have assumed ideal mixing behavior for both clathrin monomers and clathrin cages. Applying the equilibrium condition $p\mu_{\mathrm{C}} - \mu_p = 0$, we obtain the equilibrium constant

$$K_p = \frac{[\mathrm{C}_p]/c_0}{\big([\mathrm{C}]/c_0\big)^p} = e^{-\beta \Delta G_p^0}, \tag{3.4}$$

where

$$\Delta G_p^0 = \mu_p^0 - p\mu_{\mathrm{C}}^0 \tag{3.5}$$

and $\beta = (k_B T)^{-1}$. The changes needed when APs are present will be given below.

This paper is organized as follows. In Section 3.2 we present the statistical physical model used to predict chemical potentials, and hence cage size distributions, both in the absence and presence of APs. We clearly indicate the parameters that determine the theoretical prediction. Section 3.3 briefly describes the numerical methods used to solve the final set of equations. The numerical results are presented and discussed in Section 3.4. We end with a short summary.

3.2 Theory

3.2.1 Clathrin

According to statistical physics, the chemical potential of component i in an ideal mixture of several components may be calculated according to $\mu_i = -k_B T \ln (q_i/N_i)$, where q_i is the single particle partition function of component i and N_i is the number of molecules of that component in the volume V [33–35]. From this we infer the reference chemical potentials to be

$$\mu_{\mathrm{C}}^0 = -k_B T \ln \frac{q_{\mathrm{C}}}{V c_0}, \tag{3.6}$$

$$\mu_p^0 = -k_B T \ln \frac{q_p}{V c_0}. \tag{3.7}$$

Notice that $V c_0 = V/v_0$, where v_0 is the volume available to one clathrin or one cage at the refence concentration. We now set forth to calculate the relevant single particle partition functions.

Free clathrin triskelia

Adopting a reductionist approach, a clathrin triskelion is represented using a restricted number of effective parameters that allow a simple description of their complex flexible structure. A

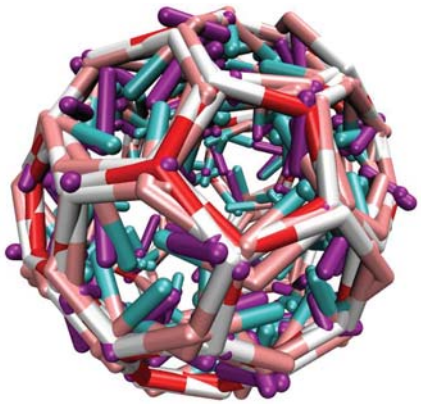


Figure 3.2: Snapshot of a cage of $p = 60$ triskelia and ∼120 APs, obtained using computer simulations and visualized with Visual Molecular Dynamics [36] (VMD), using the same colour scheme as in Fig. 3.1.

clathrin triskelion is idealized to be a symmetric molecule with a central three-fold axis rotating its three legs into oneanother. Each of the three legs is oriented with respect to this axis by the same planar pucker angle of χ, see Fig. 3.1. The pucker angle is allowed to fluctuate around some equilibrium value χ_f^0, within a range dictated by a bending rigidity κ_f. In the harmonic approximation, *i.e.* for small fluctuations, the energy penalty for bending the structure away from its equilibrium structure is proportional to the square of the deviation, so

$$\phi_{\mathrm{C}}(\chi) = \Phi_{\mathrm{C}} + \frac{1}{2}\kappa_f(\chi - \chi_f^0)^2, \tag{3.8}$$

where Φ_{C} is the free energy of an equilibrium triskelion in solution, which arises from interactions with and changes induced in the solvent. The partition function of a triskelion then reads as

$$q_{\mathrm{C}} = \frac{8\pi^2 V}{\Delta_{\mathrm{C}}} \int e^{-\beta\phi_C} d\chi \approx \frac{8\pi^2 V}{\Delta_{\mathrm{C}}} \sqrt{\frac{2\pi k_B T}{\kappa_f}} e^{-\beta\Phi_{\mathrm{C}}}. \tag{3.9}$$

Here $8\pi^2 V$ results from integrations over all possible orientations and positions, while the elementary volume element Δ_{C} results from integrations over their conjugate momenta. The elementary volume element also takes care of Planck's constant and the symmetry of the molecule; its exact structure is given in Appendix 3.6.

Clathrin cages

Clathrin cages are well defined, closed, convex polyhedral structures. The vertices of the polyhedrons have three neighbors located approximately at the same distance, and they form the corners of polygonal facets, see Fig. 3.2. In a clathrin cage, a clathrin hub lies at every vertex of the structure and a cage edge is composed of two pairs of intertwined leg segments, where in each pair two like segments run in opposite directions, *i.e.* antiparallel.

In order to calculate the partition function of cages containing p triskelia, we adopt a mean field approach according to which each particle moves uncorrelated to the other particles in the cage, feeling only an average potential produced by all the others. As a further approximation, we assume that each particle in the cage is restricted to freely move around in limited translational and rotational volumes, v_{C}^t and ω_{C}^r, respectively, and that the potential in this volume is constant.

When a clathrin is placed at a vertex of a cage, its hub will roughly sit on three knees from neighboring triskelia and three ankles from next-nearest neighbors. As a result, we expect that its bending rigidity and equilibrium pucker angle, κ_v and χ_v^0, will be different from those in solution, κ_f and χ_f^0, respectively. The remaining part of the interaction between the triskelia in a cage is concentrated along the edges of the cage, being of van der Waals type or hydrophobic. These contributions may be expected to be optimal when the pucker angle dictated by the geometry of the cage, χ_p, equals the preferred value χ_v^0. For puckers deviating from the geometrically imposed value, which we assume to be identical for all vertices in the cage, stresses will occur that lead to an increase of the interaction energy; we include this energy by a harmonic term with a spring constant κ_g. Combining the two contributions, the energy of a clathrin with pucker χ, confined to a rotational and translational free volume in a cage of p triskela, reads as

$$\begin{aligned}\phi_p(\chi) &= \Phi_{\mathrm{C}} + E_{\mathrm{C}} + \frac{1}{2}\kappa_v(\chi-\chi_v^0)^2 + \frac{1}{2}\kappa_g(\chi-\chi_p)^2 \\ &\approx \Phi_{\mathrm{C}} + E_{\mathrm{C}} + \frac{1}{2}\kappa_v(\chi_p-\chi_v^0)^2 + \frac{1}{2}\kappa_g(\chi-\chi_p)^2.\end{aligned} \tag{3.10}$$

The last line is been obtained by combining the squares and assuming that κ_g is much larger than κ_v. The value of E_{C} has been estimated by simulations and theory [19, 27] and by experimental means [37] to be around $E_{\mathrm{C}} = 24k_BT$. To complete the description of our model, we mention that χ_p is solved from [26]

$$p\left[2\pi - 3\arccos\left(\frac{3\cos^2\chi_p - 1}{2}\right)\right] = 4\pi. \tag{3.11}$$

This equation arises from an application of Descartes' equation for the total angle deficit in a polyhedron [38]. A plot of χ_p is presented in Figure 3.3.

We are now in the position to calculate the partition function q_p of a cage containing p triskelia. Using the mean field potential described above, the total partition function of the cage is just the product of p identical individual partition functions of a single clathrin triskelion in a cage, apart from the fact that of all translational degrees of freedom one may run through the total volume V and similarly of all orientational degrees of freedom one may run through the full range of $8\pi^2$,

$$q_p = m_p \frac{8\pi^2 V}{v_{\mathrm{C}}^t \omega_{\mathrm{C}}^r}\left[\frac{v_{\mathrm{C}}^t\omega_{\mathrm{C}}^r}{\Delta_{\mathrm{C}}}\sqrt{\frac{2\pi k_BT}{\kappa_g}}e^{-\beta\Phi_p}\right]^p, \tag{3.12}$$

where

$$\Phi_p = \Phi_{\mathrm{C}} + E_{\mathrm{C}} + \frac{1}{2}\kappa_v\left(\chi_p - \chi_v^0\right)^2. \tag{3.13}$$

The multiplicity factor m_p will be explained below. Notice that there is no factor of $1/p!$ in q_p accounting for the indistinguishability of the p triskelia, since it is canceled by a factor of $p!$ obtained by permuting all triskelia in the cage. The latter factor takes care of the fact that in the given expression only a limited part of the total phase space is incorporated. since each triskelion is restricted to a unique (relative) position in the cage. This situation is analogous to that for an Einstein crystal. Finally, in evaluating the partition function we have again assumed that κ_g is much larger than κ_v.

Cage multiplicity

Let us now briefly discuss the multiplicity factor m_p introduced above in the partition function for cages containing p triskelia, see Eq. (3.12). We start by noticing that the well-known Euler

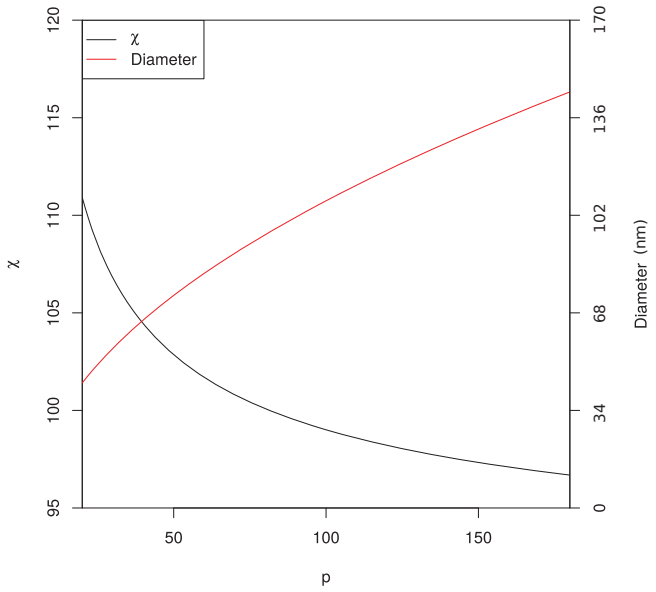


Figure 3.3: The pucker angle χ_p (black solid line, leftaxis) and the diameter D (red solid line, right axis) of clathrin cages as functions of the number of triskelia p

relation, [38] $V - E + F = 2$, applying to all structures topologically equivalent to a sphere, restricts the number of possible cages structures severely. Here, V is the number of vertices, E is the number of edges and F the number of facets of the cage; in cages composed of p clathrin proteins, $V = p$ and $E = 3p/2$. Assuming that only pentagons and hexagons are allowed, it can be shown that the number of pentagons must be exactly 12 in all cases. The number of hexagonal facets is then found to be $(p - 20)/2$. The smallest cage is thus a dodecahedron, corresponding to $p = 20$; all further possible cages have even values of p, and only $p = 22$ is geometrically forbidden.

Previous computer simulations with a completely rigid model revealed that despite the high number of geometrically possible configurations for any given p, *i.e.* distinct distributions of the pentagons and hexagons, typically just one preferred cage structure self-assembles for any even value of $p \lesssim 70$ [26]. The other structures, although geometrically possible, are rarely formed because their internal stresses result in unfavourably high energies, and for some values of p all possible structures are under adverse tensions. Schein *et al.* have analyzed all possible structures and suggested a geometric rule to determine which structures are favourable and which are not [31, 32]. On the basis of the same data, we formulated the concise 'excluded 5566' rule which says that structures containing one or more facets whose neighboring facets form a sequence of two pentagons followed by two hexagons, either clockwise or anti-clockwise, are unfavourable [39]. For large cages, $p > 60$, these rules are equivalent to the more familiar isolated pentagon rule (IPR). The number of geometrically allowed cage structures and the cage multiplicities m_p, *i.e.* the number of cages satisfying the above requirements for given number of triskelia p, are shown in Figure 3.4. For of values $p > 100$ we estimated the numbers of cages based on exponential extrapolations of the data for $p \leq 100$.

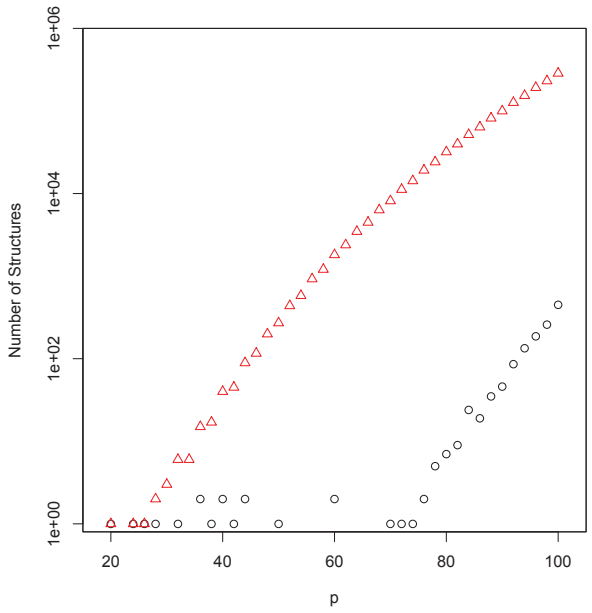


Figure 3.4: The numbers of possible cage structures (red triangles) and favorable cage structures (black circles, m_p) as functions of the number of triskelia per cage, p. The absence of circular markers for certain even values of p reflects a lack of favourable cages of these sizes, $m_p = 0$. Values copied from Schein *et al.* [32].

Equilibrium

As mentioned in the Introduction, equilibrium is governed by Eq. (3.5). With the above information we now calculate

$$\Delta G_p^0 = p\Delta\mu_{\rm C}^0 + \frac{1}{2}p\kappa_v\left(\chi_p - \chi_v^0\right)^2 - k_BT\ln m_p + k_BT\ln\left(\frac{v_{\rm C}^t\omega_{\rm C}^r}{8\pi^2 v_0}\right), \tag{3.14}$$

where

$$\Delta\mu_{\rm C}^0 = E_{\rm C} + k_BT\ln\left(\sqrt{\frac{\kappa_g}{\kappa_f}}\frac{8\pi^2 v_0}{v_{\rm C}^t\omega_{\rm C}^r}\right). \tag{3.15}$$

Note that the elementary volume elements $\Delta_{\rm C}$ have canceled in the final result. The first term in Eq. (3.14) depends linearly on p. The next two terms depend non-linearly on p, the first one through χ_p – which saturates at $\chi_p = \pi/2$ for large p – and the second one through m_p. As mentioned above, for small p the multiplicity m_p differs only slightly from unity, depending rather irregularly on p, while for large values of p one sees in Fig. 3.4 that $\ln(m_p)$ depends roughly linearly on p. The last term in Eq. (3.14) is the only term not to depend on the cage size, and therefore its relevance deminishes with increasing p. Rather than ignoring this term altogether, we estimated the translation and rotational volumes as $v_{\rm C}^t = 5\,{\rm nm}^3$ and $v_{\rm C}^t = 10^{-3}\,{\rm rad}^3$, respectively.

3.2.2 Adaptor proteins

In this section we present the necessary ingredients to describe clathrin cage formation in the presence of adaptor proteins. Our description is based on the reaction

$$p{\rm C} + n{\rm A} \rightleftharpoons {\rm C}_p{\rm A}_n, \tag{3.16}$$

where ${\rm C}_p{\rm A}_n$ is a cage consisting of p triskelia with n attached APs. The AP2 complex possesses two binding motif for clathrin, connected by a flexible linker, see Fig. 3.1(C). Each binding site on the AP binds to one unique site, out of two distinct sites, on each of the three clathrin legs. The first of these sites is located at the end of the terminal domain (TD), see Fig. 3.1, while the second is located at the ankle. The length of its linker allows APs to bind to a cage in two ways, either by binding only one site to one triskelion or by binding both sites to two distinct triskelia. We assume that binding of an AP to a triskelion does not alter the three-dimensional structure of the latter in any substantial manner. Since we are only interested in the distribution of cage sizes over different values of p, independently of the way the APs are bound to the cages, we do not discriminate between cages having the same total number of bound APs but different numbers of singly and doubly bound APs. Generalization of our methods to these cases are obvious from our treatment below. Following the coarse-grained approach in our recent simulation study of AP-induced cage assembly, [19] we represent the β-linker of the AP2 complex by two point particles, for convenience both taken to be identical, each with the ability to bind clathrin at a specific site. The remainder of the AP2 complex does not bind to clathrin and does therefore not contribute to the equilibrium constants derived below.

Chemical potentials are defined as before, including now

$$\mu_{\rm A} = \mu_{\rm A}^0 + k_BT\ln\frac{[{\rm A}]}{c_0}, \tag{3.17}$$

$$\mu_{p,n} = \mu_{p,n}^0 + k_BT\ln\frac{[{\rm C}_p{\rm A}_n]}{c_0}, \tag{3.18}$$

where $\left[\mathrm{C}_p\mathrm{A}_n\right]$ and $\left[\mathrm{A}\right]$ are the concentrations of clathrin-AP cages and unbound APs, respectively. At equilibrium the reaction is described by the equilibrium constant

$$K_{p,n} = \frac{\left[\mathrm{C}_p\mathrm{A}_n\right]/c_0}{\left(\left[\mathrm{C}\right]/c_0\right)^p \left(\left[\mathrm{A}\right]/c_0\right)^n} = e^{-\beta\Delta G^0_{p,n}}, \tag{3.19}$$

where

$$\Delta G^0_{p,n} = \mu^0_{p,n} - p\mu^0_{\mathrm{C}} - n\mu^0_{\mathrm{A}}. \tag{3.20}$$

The reference chemical potentials are obtained from expressions similar to those used before.

Free adaptor proteins

The internal potential energy of the AP complex is assumed to depend solely on the distance r_{12} between its two clathrin binding sites. Furthermore, we assume that the linker between the two binding sites behaves as a spring limited to have a finite extension L,

$$\phi_{\mathrm{A}}(r_{12}) = \Phi_{\mathrm{A}} + \phi_l(r_{12}), \tag{3.21}$$

where Φ_{A} is the average solvation free energy of an AP and the energy of the linker reads as [40]

$$\phi_l(r_{12}) = -\frac{1}{2}kL^2 \ln\left[1 - (r_{12}/L)^2\right] \tag{3.22}$$

for $r_{12} < L$ and infinity otherwise. The flexible linker connecting the two clathrin binding sites is composed of about 70 residues, and hence an approximate contour length of $L = 26\,\mathrm{nm}$. The spring constant is estimated using polymer theory [41] as $k = 3k_BT/(2Ll_p)$; combination with the experimental persistence length of disordered proteins, $l_p \approx 0.6\,\mathrm{nm}$, then yields $k = 0.1k_BT/\mathrm{nm}^2$.

With these assumptions, the partition function of an AP at infinite dilution reads as

$$q_{\mathrm{A}} = \frac{1}{\Delta^2_{\mathrm{A}}} \iint e^{-\beta\phi_{\mathrm{A}}} d\mathbf{r}_1 d\mathbf{r}_2 = \frac{V q_l}{\Delta^2_{\mathrm{A}}} e^{-\beta\Phi_{\mathrm{A}}}, \tag{3.23}$$

with $\Delta_{\mathrm{A}} = h^3/(2\pi m_{\mathrm{A}} k_B T)^{3/2}$ the elementary volume element per AP particle, and

$$q_l = 4\pi \int_0^\infty e^{-\beta\phi_l(r_{12})} r^2_{12}\, dr_{12}. \tag{3.24}$$

In case the potential ϕ_l is approximated by a harmonic spring with spring constant k, the integral is readily solved yielding $q_l = (2\pi k_B T/k)^{3/2}$.

Clathrin-AP cages

As mentioned before, both AP's binding sites have a corresponding binding site on each of the three clathrin legs. In principle therefore, there are two distinct binding energies associated with the two binding processes. To the best of our knowledge, these binding energies have not been determined so far. In the present model, we will assume them to be equal and denoted by E_{A}. The exact value of this parameter, like that of E_{C}, will depend on the acidity and salt conditions of the solvent. Once the binding has taken place, the translational freedom of the AP binding site is reduced to a small volume v^t_{A}. The energy of an AP bound to a cage by a single binding site only then reads as

$$\phi^s_{\mathrm{A}}(r_{12}) = \Phi_{\mathrm{A}} + E_{\mathrm{A}} + \phi_l(r_{12}), \tag{3.25}$$

and in case of double binding to a cage,

$$\phi_{\mathrm{A}}^{d}(r_{12}) = \Phi_{\mathrm{A}} + 2E_{\mathrm{A}} + \phi_l(d_p) \tag{3.26}$$

with d_p the distance between the two binding locations. The position-dependent part of the AP's partition function now gives rise to volume factors $v_{\mathrm{A}}^t q_l$ for the single clicked molecule and $(v_{\mathrm{A}}^t)^2 \exp[-\beta\phi_l(d_p)]$ for the double clicked molecule.

The partition function of a cage containg p triskelia and n APs then takes the form

$$q_{p,n} = q_p \, q_{p,n}^{\mathrm{A}}, \tag{3.27}$$

with q_p given by Eq. (3.12). It is not difficult to see that the adaptors yield

$$q_{p,n}^{\mathrm{A}} = \Omega_{p,n} \left[\frac{q_l v_{\mathrm{A}}^t}{\Delta_{\mathrm{A}}^2} e^{-\beta(\Phi_{\mathrm{A}}+E_{\mathrm{A}})} \right]^n, \tag{3.28}$$

with

$$\Omega_{p,n}(d_p) = \sum_{m=0}^{\min(n,3p-n)} \Omega_{p,n,m} \left[\frac{v_{\mathrm{A}}^t}{q_l} e^{-\beta[E_{\mathrm{A}}+\phi_l(d_p)]} \right]^m. \tag{3.29}$$

The second factor in Eq. (3.28) is the partition function in case all n APs are bound to the cage by one bond only. The first factor, $\Omega_{p,n}(d_p)$, takes care of the m APs doubly bounded to the cage. Since the binding energy E_A is negative and outweighs the linker energy $\phi_l(d_p)$, the factor $\Omega_{p,n}(d_p)$ will be dominated by terms with m as large as possible. The combinatorial factor $\Omega_{p,n,m}$ is a rather complicated object. We will assume that it exists for every $n > 0$, $m \geq 0$ and $n + m \leq 3p$, and that it is approximately given by

$$\Omega_{p,n,m} \approx \frac{(3p)!}{m!(3p-m)!} \frac{(6p-2m)!}{(n-m)!(6p-n-m)!}. \tag{3.30}$$

This expression is obtained by first distributing m doubly clicked APs over $3p$ possible positions in the cage; here we assume that $3p$ positions are available for one AP site to click to an ankle, and that this AP's second site binds to a uniquely defined nearest TD. Once this is done, the remaining $n-m$ singly clicked APs are distributed over the $6p-2m$ clathrin sites left available. Equation (3.29) can also be applied for freely floating triskelia, $p = 1$, provided the sum in Eq. (3.30) is restricted to $m = 0$ only.

The distance d_p between nearest unlike binding sites for APs on a cage containing p triskelia is estimated at $d_{60} = 12\,\mathrm{nm}$ for the most abundant cage size in *in vitro* experiments in the presence of APs. It is, however, evident that the distance between an ankle and the nearest TD is affected by the curvature of the cage, and therefore d_p varies with p. Without making further geometrical assumptions, we model the distance as

$$d_p = d_{60} + l\,(\chi_p - \chi_{60})\,, \tag{3.31}$$

where l is an unknown Taylor expansion parameter, which will be used as a fit parameter to obtain the cage size distribution in the presence of APs, starting from that in the absence of APs. Notice that we aim not only for a shift of the distribution, but also for a change of shape. The dependence of d_p on p via χ_p ensures that d_p converges to a constant value for large p, as expected for nearest neighbour distances in polyhedrons.

Equilibrium

Combining the partition functions obtained so far, we arrive at

$$\Delta G^0_{p,n} = \Delta G^0_p + n\Delta\mu^0_{\mathrm{A}} - k_B T \ln \Omega_{p,n}, \tag{3.32}$$

where

$$\Delta\mu^0_{\mathrm{A}} = E_{\mathrm{A}} - k_B T \ln \frac{v^t_{\mathrm{A}}}{v_0}. \tag{3.33}$$

It is clear from Eq. (3.32) that the standard reaction free energy for the formation of cages in the presence of APs, besides depending on the clathrin related parameters identified in the previous section, through $\Omega_{p,n}$ depends critically on the numbers of double clicked APs, which in turn are sensitive to the elongation distances d_p of the linkers. This agrees with our earlier observation in coarse-grained simulations, where the mechanical properties of the AP-linker proved crucial to the APs ability to regulate the assembly of cages [19]. With this definition of the standard chemical potential difference, the factors $v^t_{\mathrm{A}} \exp(-\beta E_{\mathrm{A}})$ in Eqs. (3.28) and (3.29) reduce to $v_0 \exp(-\beta\Delta\mu^0_{\mathrm{A}})$.

3.3 Numerical methods

In this section the numerical methods used to calculated concentrations of interest are briefly described. We present the methods only for the general case when APs are present; the adaptations to obtain results in the absence of APs are immediate.

To calculate the distribution of clathrin cages over their possible sizes as functions of the total concentrations of triskelia and APs, $\big[\mathrm{C}\big]_{\mathrm{t}}$ and $\big[\mathrm{A}\big]_{\mathrm{t}}$ respectively, we make use of the mass balances for both components:

$$\begin{cases} \big[\mathrm{C}\big]_{\mathrm{t}} = \displaystyle\sum_{p=1}^{\infty}\sum_{n=0}^{6p} pK_{p,n}\big[\mathrm{C}\big]^p\big[\mathrm{A}\big]^n c_0^{1-p-n}, \\ \big[\mathrm{A}\big]_{\mathrm{t}} = \displaystyle\sum_{p=1}^{\infty}\sum_{n=0}^{6p} nK_{p,n}\big[\mathrm{C}\big]^p\big[\mathrm{A}\big]^n c_0^{1-p-n} + \big[\mathrm{A}\big]_{\mathrm{f}}. \end{cases} \tag{3.34}$$

The right hand sides of this equation simply represent the total concentrations of triskelia and APs, respectively, expressed as weighted sums over all equilibrium cage concentrations $\big[\mathrm{C}_p\mathrm{A}_n\big]$ at given unbound monomer concentrations $\big[\mathrm{C}\big]$ and $\big[\mathrm{A}\big]$. Notice that the terms with $p = 1$ represents single triskelia, possibly dressed with single clicked APs. The next value of p that contributes to the sums is $p = 20$, followed by all even $p \geq 24$. The equilibrium constants are obtained by combining Eq. (3.19) with Eq. (3.32).

When applied to an experimental system containing known total concentrations, the above relations constitute two equations with two unknowns, *i.e.* the concentrations of unbound monomers. Once these have been solved, all concentrations $\big[\mathrm{C}_p, \mathrm{A}_n\big]$ may be calculated from the corresponding formation reactions. Although solving a set of two non-linear equations for two unknowns usually is not a very challenging task, things are different in the present case. The main problem is caused by the presence of very large powers, both in the mass balance equations and in the expressions for the equilibrium constants. This may be partly solved by performing products by adding their logarithms. Even then, the usual recursive methods to solve the set of equations turned out to be rather unstable. We have been successful by using a combination of a

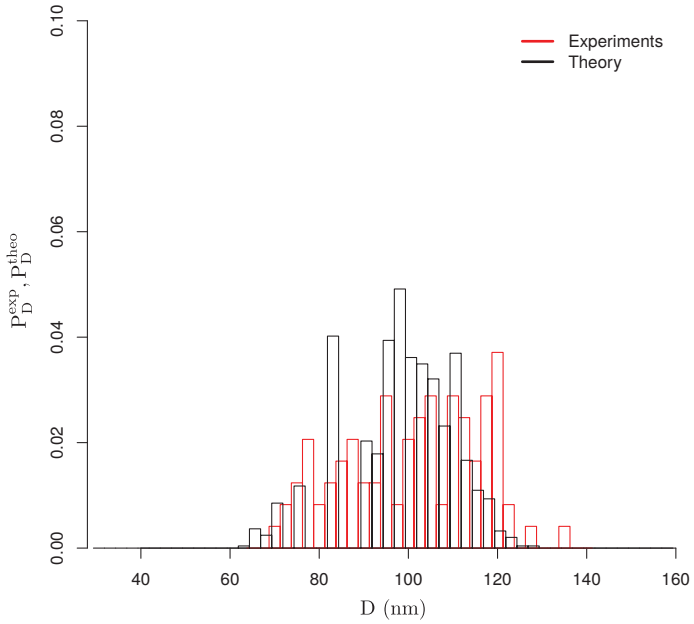


Figure 3.5: The experimental (red) and theoretical (black) cage size distributions for clathrin self-assembly in the absence of adaptor proteins. Measurement data are copied from Zaremba and Keen, [9] who used a pH 6.2 buffer with a mM salt concentration. The theoretical curve, which is slightly shifted along the horizontal axis for presentation purposes, is obtained with $\chi_v^0 = 98°$ and $\kappa_v = 80 k_B T/\mathrm{rad}^2$.

Newton-Raphson first-order method and a modified version of Powell's method [42–44]. Powell's method was used to iteratively minimize the norm of the concentrations difference vector

$$\Delta\mathbf{c} = \left(\left[\mathrm{C}\right]_\mathrm{s} - \left[\mathrm{C}\right]_\mathrm{t}, \left[\mathrm{A}\right]_\mathrm{s} - \left[\mathrm{A}\right]_\mathrm{t}\right) \tag{3.35}$$

where $\left[\mathrm{C}\right]_\mathrm{s}$ and $\left[\mathrm{A}\right]_\mathrm{s}$ represent the numerical values of the right hand sides of Eq. (3.34). This method does not require derivatives and is less sensitive to the chosen initial values than the Newton-Raphson method. The latter was used to iteratively zero the difference vector $\Delta\mathbf{c}$, employing the readily obtained analytical derivatives of $\left[\mathrm{C}\right]_\mathrm{s}$ and $\left[\mathrm{A}\right]_\mathrm{s}$ with respect to $\left[\mathrm{C}\right]$ and $\left[\mathrm{A}\right]$.

In the following section we will calculate probability distribution functions (PDFs) of the various possible cage sizes and compare them with experimental results. The calculations will be done by binning equilibrium cage concentrations according to their diameter in bins of constant widths, to conform with the published experimental data, see Fig. 3.5. Cage sizes p are readily converted into cage diameters D_p by means of $\pi D_p^2 = p a_v$, with the average area per vertex $a_v \approx 350\,\mathrm{nm}^2$ for cages with edges of $\sigma = 17\,\mathrm{nm}$. A plot of D_p is presented in Fig. 3.3. To quantify the difference between the normalized experimental and theoretical PDFs, P_D^{exp} and P_D^{theo} respectively, we define a scoring function

$$S(\mathbf{x}) = \sum_D \left[\frac{P_D^{\mathrm{exp}} - P_D^{\mathrm{theo}}(\mathbf{x})}{\sigma_D}\right]^2 \tag{3.36}$$

where the diameter D runs over a set of discrete values and $\mathbf{x}$ denotes the set of parameter values entering the theory; the best set is the one with the lowest S. The weight factors σ_D have

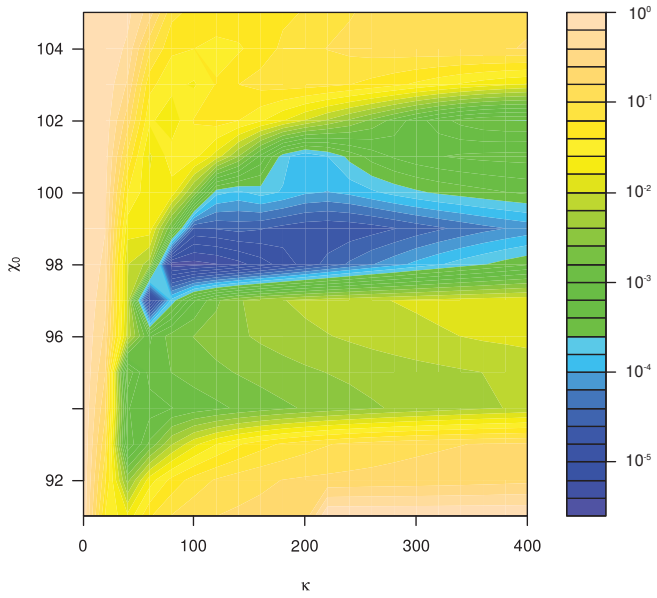


Figure 3.6: The similarity score, comparing experimental and theoretical cage size distributions, as a function of the χ_v^0 and κ_v. Note that the colour scheme employs a logarithmic scale, with dark blue representing the best fit.

been introduced to take into account that the accuracy of the experimental data may vary with D. We experimented with several definitions of the weight and found that the best set is rather insensitive to the choice made, hence we here stick to $\sigma_D = 1$.

3.4 Results

To the best of our knowledge, the literature contains only a few quantitative experimental studies on the impact of adaptor proteins on the size distribution of clathrin cages grown in *in vitro*. Here we will focus on the distributions measured for AP2 by Zaremba and Keen, [9] which have also been the topic of earlier theoretical studies on the cage size distribition [45–47]. The size distributions obtained with AP3 and AP180 are quantitatively very similar, indicative of a single underlying mechanism [15, 16]. We start with cages that self-assemble in the absence of adaptor proteins, which Zaremba and Keen realized in a buffer containing 5 mM 2-[*N*-morpholino]ethane sulfonic acid (NaMES) and 2 mM $CaCl_2$, pH 6.2 buffer [9]. The parameters to be extracted from the experimental data are $\Delta\mu_C^0$, κ_v and χ_v^0. Next are the cages grown in the presence of adaptor proteins. Zaremba and Keen present a size distribution for a different buffer than for the calthrin-only experiment, but state that 'similar size coats' were formed in the aforementioned buffer. We will interpret 'similar' as 'identical', in which case the three aforementioned parameter retain their values. This leaves $\Delta\mu_A^0$ and l to be extracted.

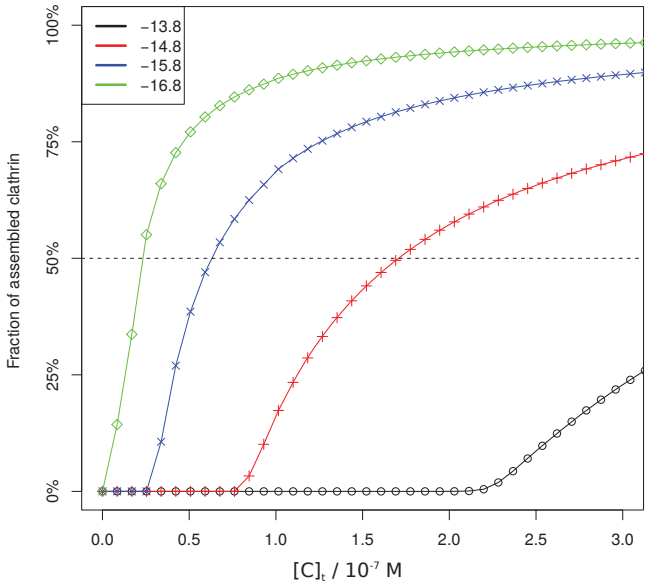


Figure 3.7: The fraction of clathrin assembled into cages for several values of $\Delta\mu_{\mathrm{C}}^0$, expressed in units of k_BT in the legend, at $\chi_v^0 = 98°$ and $\kappa_v = 80k_BT$, in the absence of APs. The concentrations of free and bound triskelia are equal (dashed line) at the critical assembly concentration (CAC).

3.4.1 Clathrin Cages

For the fit of the theory to the experimental data, we selected a total clathrin concentration of $\left[\mathrm{C}\right]_{\mathrm{t}} = \frac{1}{3} \cdot 10^{-6}\,\mathrm{molar}^{-3} = 0.21\,\mathrm{mg/ml}$, while lies close to the concentration of 0.16 mg/ml used in the experiments with AP and a factor 3 below that used in the experiments without AP. Our previous calculations at this concentration (which ignored the flexibility of the triskelion) show a sharp transition with $\Delta\mu_{\mathrm{C}}^0$ between solutions containing solely monomers and solutions with virtually all triskelia bound in cages, reminiscent of a critical assembly concentration (CAC) [19]. We now find that the cage size distribution under strong assembly conditions is insensitive to $\Delta\mu_{\mathrm{C}}^0$ (data not shown), thus ruling out an extraction of that parameter. This leaves κ_v and χ_v^0 as the parameters that determine the size distributions. Upon selecting $\Delta\mu_{\mathrm{C}}^0 = -20k_BT$, which lies $\sim 6k_BT$ into the assembly regime, we find that χ_v^0 and κ_v set the mean value and with of the size distribution, respectively, in agreement with expectations. Figure 3.6 shows the similarity score as a function of these two parameters. The best agreement with the experimental data is obtained for $\chi_v^0 = 98°$ and $\kappa_v = 80k_BT/\mathrm{rad}^2$. The corresponding distribution is plotted in Fig. 3.5, showing a satisfactory agreement with the experimental distribution. The differences between the histograms may result from several sources, including the approximations made in the theory and the low number (96) of analyzed experimental cages. Interestingly, replacing the number of favoured low-stress cage structures, *i.e.* the multiplicity m_p in Eq. (3.14), by the far larger number of geometrically allowed cage structures, see Fig. 3.4, results in a deterioration of the scoring function.

The sole remaining unknown parameter is $\Delta\mu_{\mathrm{C}}^0$. Its value can be related to the CAC by combining the mass balance equation with the condition that the number of bound triskelia

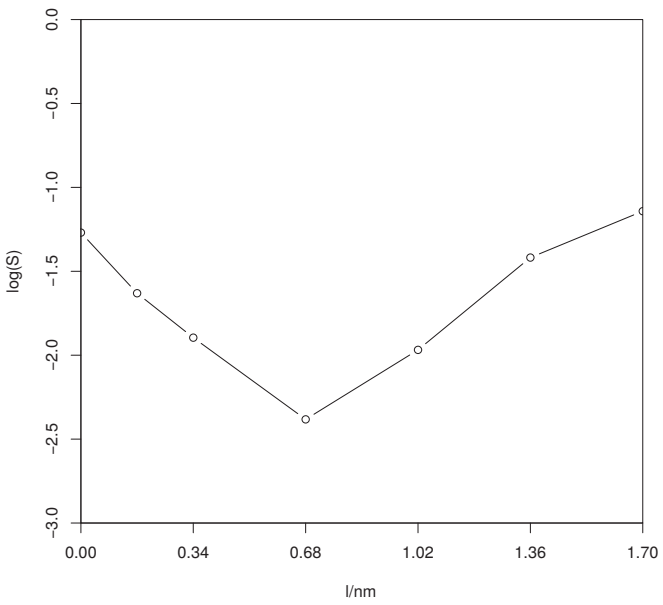


Figure 3.8: The similarity score, between experimental and calculated cage size distributions in clathrin-AP mixtures, as a function of the parameter l relating the ankle-TD distance, *i.e.* the elongation of the linker in a double bound AP, to the pucker χ_v^0, see Eq. (3.31). The clathrin parameters where kept at their best values obtained from Fig. 3.6, and $\Delta\mu_{\rm A}^0 = -15k_BT$.

equals the number of free monomers, and hence $\big[\mathrm{C}\big]_{\rm t} = 2\big[\mathrm{C}\big]$, at the critical concentration, $\big[\mathrm{C}\big]_{\rm t} = c_{\rm CAC}$. One readily derives an analytic relation between $\Delta\mu_{\rm C}^0$ and $c_{\rm CAC}$ for monodisperse cages, [19]

$$\Delta\mu_{\rm C}^0 \approx k_BT \ln \frac{c_{\rm CAC}}{c_0}, \tag{3.37}$$

but a numerical approach is required for the current wide distributions. From the assembly curves calculated at fixed $\Delta\mu_{\rm C}^0$, see Fig. 3.7, we can easily read of the matching CAC. The experimental CAC of 100 μg/ml reported by Crowther and Pearse, [48] unfortunately at slightly different buffer and salt conditions, then corresponds to $\Delta\mu_{\rm C}^0 = -14.8k_BT$, which lies close to the value established earlier [19, 28, 37].

3.4.2 Cages with APs

Our theoretical model has been based on the assumption that the roles of clathrin and AP in mixed cages can be described by two independent additive contributions to the potential energy of the cages. We therefore retain the above established values of χ_v^0, κ_v and $\Delta\mu_{\rm C}^0$ in the analysis of the system with APs; note that these three parameter will vary between solvents. The two 'new' parameters to be determined from the experimental size distribution [9] then are the standard chemical potential difference of AP binding to clathrin, $\Delta\mu_{\rm A}^0$, and the variation with the cage size of the elongation of a double-bound AP, *i.e.* the parameter l introduced in Eq. (3.31).

Strong attachment of APs to the triskelia gives rise to a size distribution that is insensitive to $\Delta\mu_{\rm A}^0$, similar to the observations in the clathrin-only case. So we select $\Delta\mu_{\rm A}^0 - 15k_BT$ and restrict our attention to the l. For $l = 0$, *i.e.* assuming that the distances between ankles and nearest TDs

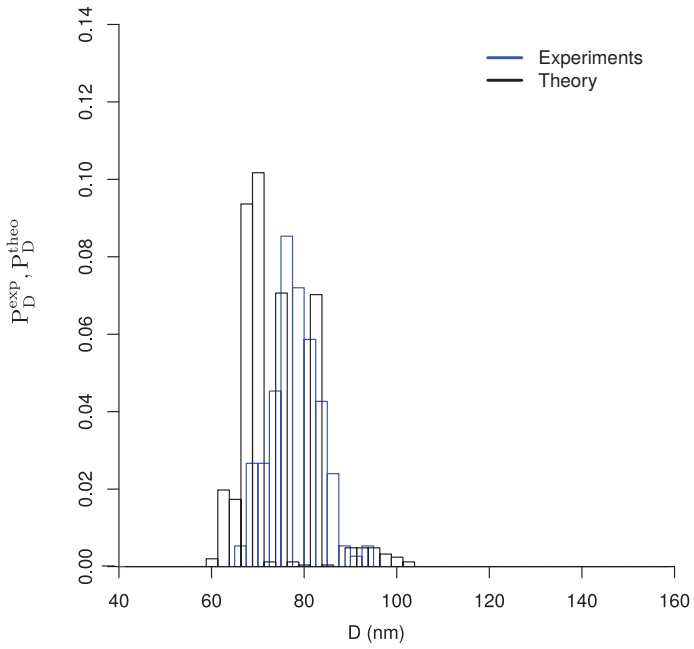


Figure 3.9: The experimental (blue) and theoretical (black) cage size distributions for caged assembled in the presence of adaptor proteins. Measurement data are copied from Zaremba and Keen, [9] who used a pH 6.2 buffer with a mM salt concentration. The theoretical curve, which is slightly shifted along the horizontal axis for presentation purposes, is obtained with the best fit parameters deduced from Fig. 3.6 and Fig. 3.8, $\chi_v^0 = 98°$, $\kappa_v = 80 k_B T/\,\mathrm{rad}$, and $l = 0.7\,\mathrm{nm/rad}$, in combination with strong interactions, $\Delta\mu_{\mathrm{C}}^0 = -20 k_B T$ and $\Delta\mu_{\mathrm{A}}^0 = -15 k_B T$.

in cages does not depend on the cage size p, we recover the size distributions of the clathrin-only case. The scores of the theoretical distributions are plotted in Fig. 3.8 as a function of l. The size distribution obtained for the best fit, at $l = 0.7\,\mathrm{nm/rad}$, shown in Fig. 3.9, is in satisfactory agreement with the experimental distribution. Comparison with Fig. 3.5 shows that the impact of AP is to shift the distribution to smaller cages and to reduce its width. We conclude that these marked changes arise from a rather subtle effect, the slightly longer (shorted) elongation of double clicked APs connecting a knee to the nearest TD in bigger (smaller) cages. Cages with diameters typical for the two distributions, $D = 75\,\mathrm{nm}$ and $D = 100\,\mathrm{nm}$ contain 51 and 88 triskelia, respectively, and consequently have puckers of $\chi_{51} = 102°$ and $\chi_{88} = 99°$, respectively. The resulting difference in the pucker elongation is a fraction of a nm, whereas the employed reference distance $d_{60} = 12\,\mathrm{nm}$. The small energy change upon a slight elongation of an AP linker, when multiplied with the number of double-bound APs present in a cage, readily gives rise to an energy term in the k_BT range for the entire cage, and thereby makes a substantial contribution to the values of the equilibrium constants.

Finally, let us briefly discuss the mechanism that leads to the assistance of APs in clathrin cage formation. Consider a situation in which clathrin alone do not assemble, *i.e.* a situation when the concentration of clathrin is below its CAC, or similarly when the binding energy is too small for the given concentration of clathrin to lead to cage formation. Adding small amounts of APs will not alter the situation, since for entropic reasons they will remain dissolved. On increasing the concentration of APs, the entropy loss of binding APs to clathrin will decrease and become comparable to the decrease of energy on binding to clathrin. Since double clicking of APs to clathrin in cages leads to the same binding energy as clicking two APs to one or two free clathrin, the lower loss of AP-entropy in the first case will favor doubly clicking APs to (fragments of) clathrin cages. With increasing concentrations of AP, therefore, the clathrin will finally form cages, even if their concentration is (slightly) below the CAC for calthrins only. With these AP concentrations, APs will mainly occur doubly clickes to cage clathrin. If we increase the concentration of APs even more, so much that all clathrin binding sites are saturated with APs, the gain of AP entropy on double clicking decreases and the efficiency of assisting clathrin to form cages goes down again. This is consistent with our findings in Fig. 2.9, see Chapter 2.

3.5 Discussion and conclusions

Adaptor proteins are well known to affect the size distributions of clathrin cages grown *in vitro*, but the underlying mechanism remained unresolved for several decades. The statistical mechanical model introduced in the current study reveals a subtle origin, *i.e.* the slight variation with cage size of the distance between a knee and its nearest terminal domain in cages results in a slight variation of the free energy of the disordered linker of an AP connecting these two sites. Since a cage contains many of these double bound APs, between one and three per triskelion, [19] the cumulative contribution to the reaction free energy is substantial. The slightly shorter (longer) distances in small (large) cages causes the cage size distribution to shift to smaller cages, and hence to a narrower distribution, when AP is added to an *in vitro* solution of clathrin. We note that this mechanism also applies *in vivo*, *e.g.* in the increase of the average vesicle size in a squid giant synapse induced by AP180-inhibitors. The model also explains how double clicked APs stabilise cages under conditions where these are unstable in the absence of AP [19].

Recent experiments by Kelly *et al.* [49] indicate that AP2, and therefore probably also AP3, posses only a single binding site for clathrin, thereby drawing into question the experimental cage size distribution discussed here. They suggest that these early experiments were contaminated with traces of AP180, which does double-bind to clathrin [50]. Cage size distributions reported

for AP180, [16] which is a four times more potent inducer of cage assembly than AP2, [51] are quantitatively similar to those studied here. This does not discredite our theory, but indicates that the adaptor protein being analyzed was not AP2 but AP180.

The sole main parameter not determined in this study is $\Delta\mu^0_{\mathrm{A}}$. We furthermore recall the assumed simplification of both sites of the adaptor protein binding equally strong to clathrin, see Section 3.2.2. This assumption is readily removed, at the expanse of more elaborate equations. The two binding constants can be measured by binding assays using clathrin fragments, thus ensuring that only one type of bond occurs. A viable alternative is to explore the CAC of AP-clathrin mixtures, specifically the AP concentrations required to induce half the triskelia in solutions of various concentrations to assemble into cages, which can then be combined with the theory to extract the binding constants.

It would be desirable to have more quantitative experimental data on the mechanical and binding properties of clathrin and adaptor proteins, and – in view of the above critique – to remeasure the cage size distributions. This would allow a validation of the theory and a more realiable determination of its main parameters, so it can be applied to further explore the intricacies of CME from a statistical mechanical perspective. Experiments over a range of temperatures would enable splitting the reaction free energy differences, and thereby parameters such as $\Delta\mu^0_{\mathrm{C}}$ and $\Delta\mu^0_{\mathrm{A}}$, into their enthalpic and entropic contributions.

3.6 Appendix: Momentum of the pucker

The partition function of a triskelion described by a position $\mathbf{r}_\mathrm{C}$, an orientation $\boldsymbol{\varphi}_\mathrm{C}$, a pucker χ and their conjugate momenta p_X, is given by

$$q_\mathrm{C} = \frac{1}{\sigma_\mathrm{C} h^7} \iint e^{-\beta(\Phi_\mathrm{C}+K)} d\mathbf{r}_\mathrm{C} d\boldsymbol{\varphi}_\mathrm{C} d\chi d\mathbf{p}_{\mathbf{r}_\mathrm{C}} d\mathbf{p}_{\boldsymbol{\varphi}_\mathrm{C}} dp_\chi, \tag{3.38}$$

with the potential energy Φ_C as in Eq. (3.8) and $\sigma_\mathrm{C} = 3$ the symmetry number of a triskelion [33]. In the kinetic energy

$$K = \frac{\mathbf{p}_{\mathbf{r}_\mathrm{C}}^2}{2M_\mathrm{C}} + \frac{1}{2}\mathbf{p}_{\boldsymbol{\varphi}_\mathrm{C}}^\mathrm{T} \mathbf{I}_\mathrm{C}^{-1} \mathbf{p}_{\boldsymbol{\varphi}_\mathrm{C}} + \frac{p_\chi^2}{2\mu_\chi}. \tag{3.39}$$

the inertia tensor $\mathbf{I}_\mathrm{C}$ and the effective mass of the pucker,

$$\mu_\chi = \left[\sum_i \frac{1}{m_i} \left(\frac{\partial \chi}{\partial \mathbf{r}_i} \right)^2 \right]^{-1}, \tag{3.40}$$

are in principle functions of χ. If the thermally induced variations of these inertia are small, *i.e.* if they depend only weakly on χ and/or the fluctuations in χ are small, then

$$q_\mathrm{C} = \frac{1}{\sigma_\mathrm{C} h^7} \sqrt{2\pi k_B T}^7 M_\mathrm{C}^{3/2} |\mathbf{I}_\mathrm{C}|^{1/2} \mu_\chi^{1/2} \iint e^{-\beta\Phi_\mathrm{C}} d\mathbf{r}_\mathrm{C} d\boldsymbol{\varphi}_\mathrm{C} d\chi \tag{3.41}$$

$$= \frac{1}{\Delta_\mathrm{C}} \iint e^{-\beta\Phi_\mathrm{C}} d\mathbf{r}_\mathrm{C} d\boldsymbol{\varphi}_\mathrm{C} d\chi, \tag{3.42}$$

where the last step defines the elementary volume Δ_C.

References

[1] B. M. Pearse, *Journal of molecular biology* **1975**, *97*, 93IN1197–96IN1398.

[2] B. Pearse, *Proceedings of the National Academy of Sciences* **1976**, *73*, 1255–1259.

[3] H. T. McMahon, E Boucrot, *Nat. Rev. Mol. Cell Biol.* **2011**, *12*, 517–533.

[4] F. M. Brodsky, *Annual review of cell and developmental biology* **2012**, *28*, 309.

[5] T. Kirchhausen, D. Owen, S. C. Harrison, *Cold Spring Harbor perspectives in biology* **2014**, *6*, a016725.

[6] M. S. Robinson, *Traffic* **2015**, *16*, 1210–1238.

[7] J. H. Keen, M. C. Willingham, I. H. Pastan, *cell* **1979**, *16*, 303–312.

[8] J. Keen, M. Willingham, I Pastan, *Journal of Biological Chemistry* **1981**, *256*, 2538–2544.

[9] S. Zaremba, J. H. Keen, *The Journal of cell biology* **1983**, 1339–1347.

[10] M. P. Woodward, T. F. Roth, *Proceedings of the National Academy of Sciences* **1978**, *75*, 4394–4398.

[11] W. Schook, S. Puszkin, W. Bloom, C. Ores, S. Kochwa, *Proceedings of the National Academy of Sciences* **1979**, *76*, 116–120.

[12] S. C. Harrison, T. Kirchhausen, *Cell* **1983**, *33*, 650–652.

[13] P. Sorger, R. Crowther, J. Finch, B. Pearse, *J. Cell Biol* **1986**, *103*, 1213–1219.

[14] J. E. Heuser, J. H. Keen, L. M. Amende, R. E. Rippoldt, K Prased, *J. Cell Biol.* **1987**, *105*, 1999–2009.

[15] W Ye, E. Lafer, *Journal of neuroscience research* **1995**, *41*, 15–26.

[16] S. Ahle, E. Ungewickell, *The EMBO Journal* **1986**, *5*, 3143.

[17] D. J. Owen, B. M. Collins, P. R. Evans, *Annual Review of Cell and Developmental Biology* **2004**, *20*, 153–191.

[18] M. A. Edeling, C. Smith, D. Owen, *Nature Reviews Molecular Cell Biology* **2006**, *7*, 32–44.

[19] M Giani, W. K. den Otter, W. J. Briels, *Biophys. J.* **2016**, *111*, 222–235.

[20] J Heuser, T Kirchhausen, *Journal of ultrastructure research* **1985**, *92*, 1–27.

[21] E. Kocsis, B. L. Trus, C. J. Steer, M. E. Bisher, A. C. Steven, *Journal of structural biology* **1991**, *107*, 6–14.

[22] A. J. Jin, R. Nossal, *Biophysical Journal* **2000**, *78*, 1183–1194.

[23] T. Yoshimura, K. Kameyama, S. Maezawa, T. Takagi, *Biochemistry* **1991**, *30*, 4528–4534.

[24] M. L. Ferguson, K. Prasad, H. Boukari, D. L. Sackett, S. Krueger, E. M. Lafer, R. Nossal, *Biophysical journal* **2008**, *95*, 1945–1955.

[25] S. Kotova, K. Prasad, P. D. Smith, E. M. Lafer, R. Nossal, A. J. Jin, *FEBS letters* **2010**, *584*, 44–48.

[26] W. K. den Otter, M. R. Renes, W. J. Briels, *J. Phys. Condens. Matt.* **2010**, *22*, 104103.

[27] W. K. den Otter, M. R. Renes, W. J. Briels, *Biophys. J.* **2010**, *99*, 1231–1238.

[28] W. K. den Otter, W. J. Briels, *Traffic* **2011**, *12*, 1407–1416.

[29] R. Matthews, C. N. Likos, *Soft Matter* **2013**, *9*, 5794–5806.

[30] I. M. Ilie, W. K. den Otter, W. J. Briels, *Biophysical Journal* **2015**, *108*, 227a.

[31] S. Schein, M. Sands-Kidner, T. Friedrich, *Biophysical journal* **2008**, *94*, 938–957.

[32] S. Schein, M. Sands-Kidner, *Biophysical journal* **2008**, *94*, 958–976.

[33] D. A. McQuarrie, *Statistical Mechanics*, Harper & Row Publishers, New York, NY, USA, **1976**.

[34] D. Nelson, T. Piran, S. Weinberg, *Statistical Mechanics of Membranes and Surfaces*, 2nd ed., World Scientific Publishing, Singapore, **2004**.

[35] R. Phillips, J. Kondev, J. Theriot, H. Garcia, *Physical biology of the cell*, Garland Science, **2012**.

[36] W. Humphrey, A. Dalke, K. Schulten, *Journal of molecular graphics* **1996**, *14*, 33–38.

[37] M. Saleem, S. Morlot, A. Hohendahl, J. Manzi, M. Lenz, A. Roux, *Nature communications* **2015**, *6*, 6249.

[38] D. S. Richeson, *Euler's gem: the polyhedron formula and the birth of topology*, Princeton University Press, **2012**.

[39] M. Giani, W. K. den Otter, W. J. Briels, *The Journal of Chemical Physics* **2017**, *146*, 155102.

[40] K. Kremer, G. S. Grest, *The Journal of Chemical Physics* **1990**, *92*, 5057–5086.

[41] D. Boal, *Mechanics of the Cell*, Cambridge University Press, Cambridge, U.K., **2002**.

[42] M. J. Powell, *The computer journal* **1964**, *7*, 155–162.

[43] Numerical recipes in FORTRAN (Cambridge).

[44] R. P. Brent, *Algorithms for minimization without derivatives*, Courier Corporation, **2013**.

[45] R. Nossal, *Traffic* **2001**, *2*, 138–147.

[46] R. Nossal in Macromolecular Symposia, *Vol. 219*, Wiley Online Library, **2005**, pp. 1–8.

[47] M Muthukumar, R. Nossal, *The Journal of chemical physics* **2013**, *139*, 121928.

[48] R. Crowther, B. Pearse, *The Journal of cell biology* **1981**, *91*, 790–797.

[49] B. Kelly, S. Graham, N Liska, *Science* **2014**, *345*, 459.

[50] L. Moshkanbaryans, J. Xue, J. R. Wark, P. J. Robinson, M. E. Graham, *Protein Sci.* **2016**, *11*, e0162050.

[51] R. Lindner, *Interaktionen der Hüllstrukturproteine von Clathrin-coated vesicles.* Technische Universität München, **1992**.

Chapter 4

Early stages of clathrin aggregation at a membrane

The self-assembly process of clathrin coated pits during endocytosis has been simulated by combining and extending coarse grained models of the clathrin triskelion, the adaptor protein AP2 and a flexible network membrane. The AP2's core, upon binding to membrane and cargo, releases a motif that can bind clathrin. In conditions where the core-membrane-cargo binding is weak, the binding of this motif to clathrin can result in a stable complex. We characterize the conditions and mechanisms resulting in the formation of clathrin lattices that curve the membrane, i.e. clathrin coated pits. The mechanical properties of the AP2 β linker appear crucial to the orientation of the curved clathrin lattice relative to the membrane, with wild-type short linkers giving rise to the inward curving buds enabling endocytosis while long linkers produce upside-down cages and outward curving bulges. [1]

4.1 Introduction

Eukaryotic cells possess the remarkable ability to collect, sort and internalize a variety of membrane components and external cargo molecules, by a process known as clathrin-mediated endocytosis (CME) [1–3]. The two main proteins involved in CME are clathrin and the AP2 adaptor protein complex, assisted by a series of accessory proteins [4]. Clathrin is a three-legged protein, see Fig. 5.1, with the ability to self-assemble into a variety of polyhedral cages *in vivo* and *in vitro* [5, 6]. In cages, a clathrin triskelion is centered at every vertex, with each of its legs running along two edges before bending inward at the leg's terminal domain (TD) [7–9]. Cages grown *in vivo* enclose lipid vesicles, with the clathrin triskelia tethered to the membrane by multiple AP2 complexes. The latter carries motifs for the phosphatidylinositol 4,5-bisphosphate [PtdIns(4,5)P_2] lipids specific to the plasma membrane [10] and motifs binding cargo molecules, all located in the folded core of the protein, as well as two binding sites for clathrin on a long flexible linker [11–13]. Crystallography experiments indicate that the mid-linker motif binding a clathrin TD [9] is only released by the AP2 core when that core has bound PtdIns(4,5)P_2 and cargo, [14, 15] while earlier binding essays suggest that the permanently available 'ear' or 'appendage' site at the free end of the linker binds a site higher up the clathrin leg [16]. In the highly coordinated process of endocytosis, [17] AP2 intermediates by bringing cargo and triskelia together in a clathrin coated pit (CCP), *i.e.* a membrane invagination coated by a

[1] This chapter has been published as **M. Giani**, W. K. den Otter, and W. J. Briels, 'Early stages of clathrin aggregation at a membrane in coarse-grained simulations', The Journal of Chemical Physics, vol. 146, no. 15, pp. 155102, 2017.

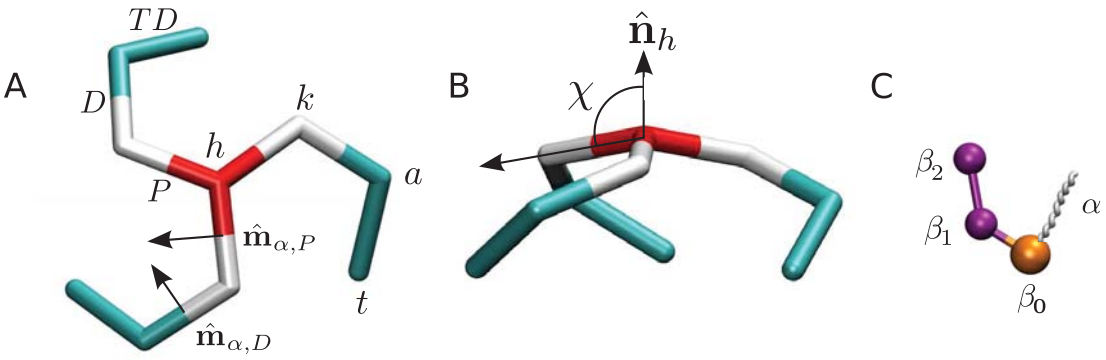


Figure 4.1: The highly coarse-grained models used for simulating clathrin (A and B) and AP2 (C), plotted on the same scale. In the rigid clathrin triskelion three proximal leg segments (P) radiate from a central hip (h) to the knees (k), at a pucker angle χ relative to the symmetry axis $\hat{\mathbf{n}}_h$, followed by distal leg segments (D) running to ankles (a) and terminal domains (TDs) ending at the toes (t). The directionality of leg-leg interactions, represented by the polarity vectors $\hat{\mathbf{m}}$, is elaborated on in Fig. 5.2. The AP model features three beads connected by two flexible linkers: the β_1 and β_2 beads can bind to the toes and ankles of clathrin, respectively, while the core β_0 can bind to a membrane bead. The full AP2 protein also possesses a flexible α linker, which is omitted in the simulations as it does not bind to membranes nor to clathrin.

clathrin lattice. As the lattice continues to grow and curve, the membrane is wrapped until the above mentioned clathrin coated vesicle (CCV) is pinched off from the membrane. Finally, the proteins decorating the vesicle are released for the next cycle while the vesicle carries the cargo to an organelle for further processing. Clathrin also produces transport vesicles at other sites in the cell, in collaboration with a wide range of membrane-specific and/or cargo-specific adaptor proteins [11, 18, 19].

Current experimental techniques do not permit examination at the molecular level of CCP formation in living cells. Cryo-electron microscopy produces high resolution stills of CCPs at various stages of maturation, thereby providing insights into the structure of these CCPs [20] as well as revealing the presence of large flat clathrin lattices [21, 22]. Labeling selected proteins with fluorescent labels enables imaging endocytosis in living cells [23]. Fluorescence recovery after photobleaching (FRAP) reveals that triskelia in pits are readily exchanged, unlike those in cages grown *in vitro* [24–26]. Total internal reflection fluorescence microscopy (TIRF) unravels the sequence of arrival and departure of various proteins – including cargo, clathrin, adaptor proteins and the membrane scission protein dynamin – over the course of approximately one minute between the initiation and successful conclusion of an endocytic event [27–30]. Besides initiations that mature to completion, these experiments also detect aborted events [31, 32]. In two recent studies, single particle tracking is combined with photo-activation localization microscopy (PALM) to follow the internalization by CME of nano-particles in unparalleled resolution [33] and fluorescence intensity tracing is used to shift the focus from the plasma membrane - coverglas interface to CCP formation at unrestrained membranes [34]. Despite the wealth of information obtained by these and other optical techniques, a number of fundamental questions on CCP dynamics are still hotly debated. These include the mechanical contribution of clathrin toward bending the membrane,[35–38] and whether CCPs are formed by nucleation and growth or by gradually increasing the curvature of a pre-existing planar lattice[20, 26, 33, 34, 39–43].

Computer simulations provide a powerful tool to explore the statistical-mechanical feasibility of hypothesized self-assembly pathways of CCPs, based on experimentally deduced and/or estimated interactions between the key proteins. Matthews and Likos modelled clathrin as a triskelion composed of 13 bead particles with interaction patches on their surfaces, showing that these triskelia can self-assemble into cages as well as create deep membrane invaginations [44]. Spakowitz and collaborators modelled clathrin as a particle that forms harmonic bonds with three neighbours, to study the mechanical properties of lattices against a flat membrane and the fluidization of these lattices when the membrane is nano-indented [45, 46]. Adaptor proteins were omitted in both types of simulations, by enabling the clathrin particles to bind directly to the membrane. The indirect coupling of trisklia to the membrane in cells – note that the disordered linkers of adaptor proteins transmit pulling forces but not pushing forces – is likely to strongly affect the deformation of the membrane by the coat. We have developed a simulation model of clathrin as a rigid particle with kinked legs, see Fig. 5.1, and with this model showed that asymmetric leg-leg interactions hold the key to self-assembly, related the binding energy to the critical assembly concentration (CAC), established a time-scale for self-assembly in solution, and observed how flat lattices release early CCPs when forced to curve [47–50]. In a recent study, we developed a simulation model for AP2 [51]. Several combinations of AP-binding locations on the clathrin leg were explored, to separate combinations that enable cage assembly from those that do not, thereby arriving at the insight that the entropic spring linking the two clathrin motifs of AP2 is crucial to the functioning of this adaptor protein. The combination of both binding sites residing at the TD, which was not explicitly mentioned in that paper, did not prove able to promote cage assembly in the bulk. This may explain the recent observation by Moshkanbaryans *et al.* [52] that a newly discovered 'site 1' on AP180 binds weaker to clathrin than eight previously identified TD-binding motifs yet proves crucial to AP-induced cage formation, by the hypothesis that site 1 is the only motif that binds clathrin at a location higher up the leg.

Modeling any process, especially one as complex as CME, involves making assumptions – based on the available experimental data – regarding the features to be included in the model and the details that can be ignored. In the current study, we explore the early stages of CCP formation for the simplest experimentally functional combination of components: clathrin, membrane and an adaptor protein [15, 37]. The clathrin protein is modelled as a rigid triskelion with smooth leg-leg interactions. Based on 10^6 triskelia in a cellular volume of $10^3 \mu\text{m}^3$, the clathrin concentration is estimated at 10^{-6} molar [53–55]. A comparable concentration is used in the simulations, while the critical assembly concentration (CAC) of the model triskelion is about an order of magnitude higher [49, 51]. The model for AP2 is equipped with two clathrin binding sites to enable the protein to bring two triskelia together and thereby induce cage assembly, as observed in *in vitro* experiments [5]. Our simulations of this process indicated that adaptor proteins – and, by implication, dimers of adaptor proteins [56–58] – are ineffectual when binding to terminal domains only, because of the relatively large distance between TDs in a cage. Hence we included binding of the β-linker appendix to the ankle of a triskelion. Because of the modest experimental evidence supporting this second site, [15, 16, 59] we also present simulations without this site.

Spontaneous AP2-induced assembly of cages in solution is prevented by incorporating the switching mechanism proposed by Owen and co-workers, [14, 15] whereby a clathrin binding site on the β-linker is released only when the AP2's core is bound to membrane, PtdIns(4,5)P_2 and cargo. Alternative mechanisms exist, as for instance in neurons where the influx of Ca^{2+} ions triggers the temporary unlocking by dephosphorylation of a collection of adaptor and accessory proteins crucial to CME, collectively known as dephosphins [60]. Recent studies indicate that FCHo and Eps15 form a complex with AP2 at the membrane, at least in the initial stages of CME, and thereby induce AP2's conformational change that releases the clathrin motif on the

β linker [61, 62]. The common feature of these mechanisms, *i.e.* AP2's ability to bind clathrin changes when AP2 binds to the membrane and cargo, is incorporated in our simulation model. For reasons of simplicity, we do not include auxiliary proteins vying with AP2 for the binding sites on the triskelia [9, 63]. Nor do we include proteins that promote membrane bending, like BAR-domains and CALM, [64–66] or proteins that resist membrane bending, like crowding effects by bulky cargo proteins [38, 67]. The model we have build, guided by the experimental data on a diverse set of proteins with complex interactions, is a highly coarse-grained idealization involving only clathrin, AP2 and membrane beads, to gain a deeper understanding by statistical mechanical simulations of the minimum requirements for the formation of clathrin coated pits.

The simulation models for clathrin, AP2 and a lipid membrane are introduced in Section 4.2. In Section 4.3 the simulation results are presented. The main conclusions and their biophysical implications are discussed in Section 4.4.

4.2 Model

In this section, the simulation models for clathrin, AP2 and the membrane are described successively. A discussion of the assumptions made to arrive at the model was presented at the end of the previous section. The model parameters for the force field and the Brownian Dynamics (BD) propagator are summarized in Tables 4.1 and 4.2, the rates of the Monte Carlo (MC) steps are collected in Table 4.3. Readers less interested in the fine details of the model are advised to proceed to the presentation of the results in Section 4.3, where the features of the models will be briefly reiterated, and the discussion of the biological implications in Section 4.4.

4.2.1 Clathrin

The clathrin model

In several preceding studies, we modelled clathrin as curved, rigid patchy particles to study the *in vitro* self-assembly of clathrin cages in bulk through Monte Carlo and Brownian Dynamics simulations [47–51]. We here provide a brief description of the model and refer the reader to our previous publications for a more detailed discussion. The model triskelion consists of three identical legs stemming from a central 'hub' (h) at a 'pucker' angle χ relative to the normal vector $\hat{\mathbf{n}}_h$ along the threefold rotational symmetry axis of the particle, see Fig. 5.1. Each leg consists of three segments: the proximal (P) and distal (D) sections and the terminal domain (TD), connected by the knee (k) and ankle (a) respectively. All three leg segments are straight and of identical length, $\sigma = 17\,\mathrm{nm}$. The orientation of the distal segments relative to the proximal segments is chosen to allow maximum overlap between the legs of a triskelion and those of a secondary triskelion whose hub is situated at a knee of the primary triskelion. According to the structural information file 1XI4, [8, 68] available at the Protein Data Bank (PDB), the terminal domain forms an angle of $\sim$114° with the adjacent distal segment and a dihedral angle of $\sim$28° relative to the distal and proximal segments of the same leg. We select a pucker angle of $\chi = 101°$, as this value corresponds to soccer-ball cages containing 60 triskelia, the most commonly observed cage size in *in vitro* experiments in the presence of AP [5].

In a completed clathrin cage, a hub is located at every vertex, on top of three knees and three ankles of neighbouring and next-nearest triskelia, respectively, see Fig. 5.2. A lattice edge is thus composed of two proximal and two distal segments, where the amino acid sequences in both pairs of like segments run in opposite directions (*i.e.* anti-parallel). In our model, the interaction between two triskelia is described by a sum of inter-segmental interactions inspired by the segmental pairings observed in experimental cage edges, and thus attractive interactions

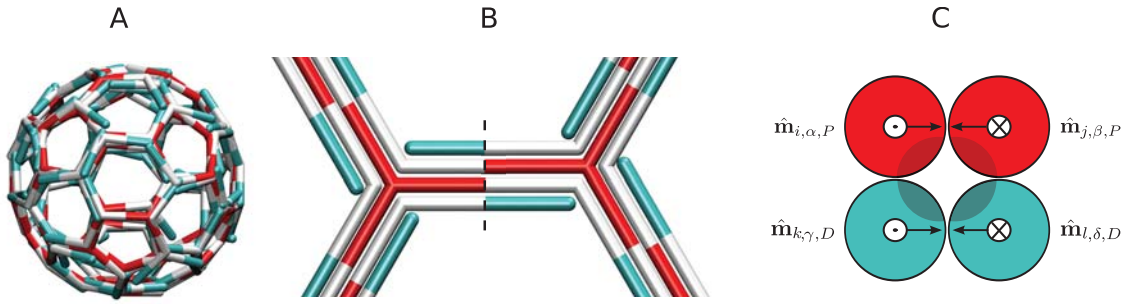


Figure 4.2: Cartoons of the cage structure, using the same colouring scheme as in Fig. 5.1. The terminal domains, curving inward from the ankle toward the center of the cage, have been omitted for clarity. (A) Snapshot of a self-assembled cage in solution. (B) Edges are composed of two proximal and two distal leg segments, with like segments oriented anti-parallel; vertices are meeting points of one hub (red), three knees (white) and three ankles (blue). For clarity, the leg segments are drawn next to each other. (C) The cross section of an experimental cage edge, along the dashed line in (B), with the two distal segments below (*i.e.* inside the cage) the two proximal segments. The markers $\odot$ and $\otimes$ denote segments pointing (from hip to knee or from knee to ankle) in to and out of the plane of the picture, respectively. Shaded areas highlight the hypothesized locations of binding sites; their asymmetric distribution along the leg's circumference is included in the simulation model by means of a torsion potential acting on the polarity vectors $\hat{\mathbf{m}}$.

are introduced between aligned pairs of two anti-parallel proximal segments, of two anti-parallel distal segments, and between any aligned pair of one proximal and one distal segment – the contribution of the TDs is considered negligible. The attractive interaction between any pair of segments is modelled by a four-site potential based on the distances between the end-points and the orientations of the two segments. All these attractive interactions are anisotropic under rotations around the long axes of the leg segments, to mimic that the many weak interaction sites are predominantly located at that side of the leg that faces the neighbouring legs along the same cage edge, see Fig. 5.2. Our previous simulations indicate that this asymmetry or 'polarity' holds the key to spontaneous self-assembly of cages [47, 48].

As an example, consider the interaction between the proximal segment of the α^{th} leg of particle i and the proximal segment of the β^{th} leg of particle j. When the two segments are properly aligned, their respective ends are close to each other. The two average distances between the four ends of these segments are

$$r_{j\beta,kh}^{i\alpha,hk} = \frac{1}{2}\left|\mathbf{x}_{i,h} - \mathbf{x}_{j\beta,k}\right| + \frac{1}{2}\left|\mathbf{x}_{i\alpha,k} - \mathbf{x}_{j,h}\right|, \tag{4.1}$$

$$r_{j\beta,hk}^{i\alpha,hk} = \frac{1}{2}\left|\mathbf{x}_{i,h} - \mathbf{x}_{j,h}\right| + \frac{1}{2}\left|\mathbf{x}_{i\alpha,k} - \mathbf{x}_{j\beta,k}\right|, \tag{4.2}$$

with $\mathbf{x}$ denoting the position of the specific joint indicated in the subscript. The distance on the first line is small if the hub of i is close to the β^{th} knee of j and the α^{th} knee of i is close to the hub of j (*i.e.* aligned and anti-parallel), while the distance on the second line is small if the hub of i is close to the hub of j and the α^{th} knee of i is close to the β^{th} knee of j (*i.e.* aligned and parallel). The former combination occurs in clathrin cages, hence an attractive interaction is assigned:

$$\phi_{j\beta,kh}^{i\alpha,hk} = -\varepsilon_{kh}^{hk} \cdot f\left(r_{j\beta,kh}^{i\alpha,hk}\right) \cdot g\left(\hat{\mathbf{m}}_{i\alpha,P} \cdot \hat{\mathbf{m}}_{j\beta,P}\right), \tag{4.3}$$

with the positive parameter ε_{kh}^{hk} denoting the (absolute) maximum inter-segmental binding energy. The distance dependence smoothly decreases from unity for coinciding end points to zero at the cut-off distance r_{cut}, following

$$f(r) = \frac{1}{2}\left[1 - \frac{\tanh[A(r - r_{\mathrm{cut}}/2)]}{\tanh[Ar_{\mathrm{cut}}/2]}\right], \tag{4.4}$$

where A determines the steepness of the potential. The numerical values of these parameters are provided in Table 4.1. Although this function favours the proper (anti)parallel alignment of leg segments, it leaves both segments free to rotate around their long axes. The cross section of a cage edge in Fig. 5.2(C), however, suggest that binding sites are located at one side of the segment and that the segments in a cage present this side to their neighbours. To model this, we associate with every proximal segment a polarity vector, defined for the α^{th} leg as

$$\hat{\mathbf{m}}_{\alpha,P} = \frac{\hat{\mathbf{n}}_h \times (\mathbf{x}_{\alpha,k} - \mathbf{x}_h)}{|\hat{\mathbf{n}}_h \times (\mathbf{x}_{\alpha,k} - \mathbf{x}_h)|}, \tag{4.5}$$

see Figs 5.1(A) and 5.2. Alignment of the polarities of the two leg segments in Eq. (4.3) is imposed through

$$g(x) = \begin{cases} -x & \text{for} \quad x < 0 \\ 0 & \text{for} \quad x \geq 0. \end{cases} \tag{4.6}$$

All other attractive interactions between segment pairs, see the tabulated list of combinations, are constructed along the same lines. Polarity vectors to distal segment $\hat{\mathbf{m}}_{\alpha,d}$ are defined as in Eq. (4.5), based on the end points of that segment and the normal at the knee, which is obtained by mirroring the normal vector at the hub in a plane perpendicular to the proximal domain, running through the center of that segment, see Figs 5.1 and 5.2.

Excluded volume interactions between triskelia are omitted for computational reasons. Their introduction would require a more complex particle shape and some flexibility of the legs to enable the particles to interweave four legs along each cage edge. Excluded volume interactions are important in preventing a triskelion from binding to a cage in a position and orientation already occupied by another triskelion. This effect is reproduced by a repulsion between aligned parallel segments of the same type. For two proximal segments the potential takes the form

$$\phi_{j\beta,hk}^{i\alpha,hk} = -\varepsilon_{hk}^{hk} \cdot f\left(r_{j\beta,hk}^{i\alpha,hk}\right), \tag{4.7}$$

and likewise for two distal segments, with the parameters listed in the Table 4.1 [48].

Propagator

The trajectories of the triskelia are calculated using first order Brownian Dynamics equations of motion for translation and rotation [69–71]. Since a clathrin triskelion is modelled as a rigid particle, the position and orientation of all leg segments are fully described by the coordinates of its hydrodynamic center [72] $\mathbf{x}$ in the laboratory (l) coordinate system and a rotation matrix $\mathbf{A}$ relative to the laboratory frame. In the laboratory frame, the coordinates of a site α on the i^{th} particle are then given by

$$\mathbf{x}_{i\alpha}^{(l)} = \mathbf{x}_i + \mathbf{A}_i \mathbf{x}_{\alpha}^{(b)} \tag{4.8}$$

where $\mathbf{x}_{\alpha}^{(b)}$ represents the constant coordinates of the α^{th} site in the body-fixed (b) coordinate system. The latter is chosen as follows: the origin coincides with the hydrodynamic center, the third coordinate axis runs parallel to the threefold symmetry axis of the triskelion, the first and third coordinate axes define the plane that includes the proximal segment of the first leg of the

$i\alpha - j\beta$	ε	A/σ^{-1}	r_{cut}/σ	x
attractive				
$hk - kh$	ϵ	4	0.4	$\hat{\mathbf{m}}_P \cdot \hat{\mathbf{m}}_P$
$ka - ak$	ϵ	4	0.4	$\hat{\mathbf{m}}_D \cdot \hat{\mathbf{m}}_D$
$hk - ka$	$\epsilon/2$	4	0.4	$-\hat{\mathbf{m}}_P \cdot \hat{\mathbf{m}}_D$
$hk - ak$	$\epsilon/2$	4	0.4	$\hat{\mathbf{m}}_P \cdot \hat{\mathbf{m}}_D$
repulsive				
$hk - hk$	-10ϵ	0.8	0.8	-1
$ka - ka$	-10ϵ	0.8	0.8	-1

Table 4.1: Interaction parameters of the six distinct clathrin leg segment pairings. In the first column, the letters refer to the hub (h), knee (k) and ankle (a) of legs α and β of particles i and j, respectively. Note that the order is important: the two proximal-proximal pairings, *i.e.* the first attractive combination and the first repulsive combination, refer to Eqs. (4.1) and (4.2), respectively. The elements in the last column represent the arguments x to the polarity function $g(x)$, where the first polarity vector in the dot products refers to a segment of the α leg of particle i and the second polarity vector to a segment of the β leg of particle j, and where $g(-1) = 1$.

triskelion, and the second coordinate axis lies perpendicular to this plane, in such a way as to form a right-handed coordinate system. In Brownian Dynamics, the effect of numerous collisions with solvent molecules is translated into a friction and a random term. The translational motion of the hydrodynamic center over a time step δt then reads as

$$\begin{aligned}\mathbf{x}_i(t+\delta t) = \mathbf{x}_i(t) &+ \boldsymbol{\mu}^{t(l)}(t)\mathbf{F}_i^{(l)}(t)\delta t \\ &+ \left[\boldsymbol{\mu}_i^{t(l)}(t)\right]^{1/2} \boldsymbol{\Theta}_i^t(t)\sqrt{2k_BT\delta t},\end{aligned} \tag{4.9}$$

where $\boldsymbol{\mu}_i^{t(l)}$ denotes the laboratory-based translational mobility tensor, $\mathbf{F}_i^{(l)}$ is the sum of all conservative forces acting on the triskelion and $\boldsymbol{\Theta}_i^t$ is a random Markovian contribution distributed according to a standard normal distribution. The random displacements are related to the mobility tensor by the fluctuation-dissipation theorem, incorporated by the last term of Eq. (4.9), where k_B is Boltzmann's constant and T temperature. The translational mobility tensor varies with the orientation of the particle,

$$\boldsymbol{\mu}_i^{t(l)}(t) = \mathbf{A}_i(t)\boldsymbol{\mu}^{t(b)}\mathbf{A}_i^{\mathrm{T}}(t), \tag{4.10}$$

where $\boldsymbol{\mu}_i^{t(b)}$ represents the constant mobility tensor in the body-fixed coordinate system. The square-rooted matrix, which obeys $[\boldsymbol{\mu}_i^{t(l)}]^{1/2}[\boldsymbol{\mu}_i^{t(l)}]^{1/2} = \boldsymbol{\mu}_i^{t(l)}$, rotates likewise and hence the matrix square root has to be evaluated only once in the entire simulation.

We have recently introduced a Rotational Brownian Dynamics (RBD) algorithm [50] to simulate the rotational dynamics of rigid anisotropic bodies by expressing their orientation in space through unit quaternions, *i.e.* a set of four coordinates q_γ, with $\gamma \in \{0,1,2,3\}$, and a constraint of unit length $|\mathbf{q}| = 1$. The rotation matrix $\mathbf{A}$ is then expressed as

$$\mathbf{A}(\mathbf{q}) = \begin{pmatrix} q_0^2+q_1^2-q_2^2-q_3^2 & 2(q_1q_2-q_0q_3) & 2(q_1q_3+q_0q_2) \\ 2(q_1q_2+q_0q_3) & q_0^2-q_1^2+q_2^2-q_3^2 & 2(q_2q_3-q_0q_1) \\ 2(q_1q_3-q_0q_2) & 2(q_2q_3+q_0q_1) & q_0^2-q_1^2-q_2^2+q_3^2 \end{pmatrix}. \tag{4.11}$$

This description avoids the well-known singularities encountered when using three angular coordinates, *e.g.* Euler angles. The resulting algorithm, in the Itô representation, [70] is numerically stable and remarkably compact, with the i^{th} particle updated by

$$\begin{aligned}\mathbf{q}_i(t+\delta t) = \mathbf{q}_i(t) &+ \mathbf{B}_i(t)\boldsymbol{\mu}_i^{r(b)}\mathbf{A}_i^{\mathrm{T}}(t)\mathbf{T}_i^{(l)}(t)\delta t \\ &+ \mathbf{B}_i(t)\left[\boldsymbol{\mu}_i^{r(b)}\right]^{1/2}\boldsymbol{\Theta}_i^r(t)\sqrt{2k_BT\delta t} \\ &+ \lambda_i(t)\mathbf{q}_i(t),\end{aligned} \tag{4.12}$$

where $\boldsymbol{\mu}_i^{r(b)}$ is the constant rotational mobility matrix in the body fixed frame. The second term on the r.h.s. represents the contribution due to the torque $\mathbf{T}_i^{(l)}$ acting on the triskelion, and the third term represents the Brownian random contribution, where the components of the vector $\boldsymbol{\Theta}_i^r$ are again Markovian and distributed according to a standard Gaussian. The transformation matrix $\mathbf{B}_i$ converts angular displacements in the body frame into quaternion displacements [69, 73],

$$\mathbf{B}(\mathbf{q}) = \frac{1}{2}\begin{pmatrix} -q_1 & -q_2 & -q_3 \\ q_0 & -q_3 & q_2 \\ q_3 & q_0 & -q_1 \\ -q_2 & q_1 & q_0 \end{pmatrix}. \tag{4.13}$$

In the last contribution to Eq. (4.12), λ_i is a Lagrange multiplier used to constrain the modulus of the quaternion to be unitary. Its value is obtained by solving the quadratic equation

$$|\mathbf{q}_i^u(t+\delta t) + \lambda_i(t)\mathbf{q}_i(\delta t)|^2 = 1 \tag{4.14}$$

in every step, where $\mathbf{q}_i^u(t+\delta t)$ are the unconstrained quaternions obtained when $\lambda_i = 0$. The constraint also conveniently eliminates the calculation of a metric tensor correction and a drift term resulting from the $\mathbf{q}$-dependent mobility, as both turn out to be parallel to the constraint direction [50]. The translational and rotational mobility tensors are related to their diffusivity counterparts by $\mathbf{D} = k_BT\boldsymbol{\mu}$. The latter are determined using the HYDRO++ package [72] by modeling clathrin as an array of 52 spheres arranged to reflect the overall shape of the protein, using a viscosity of $\eta = 10^{-3}\,\mathrm{Pa\,s}$ for water at room temperature. In the body-fixed frame discussed above, the body-fixed translational and rotational diffusivity tensors are both simultaneously diagonal, with $D_1^t = D_2^t = 1.29\cdot10^{-7}\,\mathrm{cm^2\,s^{-1}}$ and $D_3^t = 1.07\cdot10^{-7}\,\mathrm{cm^2\,s^{-1}}$ for the translational diffusivity tensor, and with $D_1^r = D_2^r = 1.64\cdot10^4\,\mathrm{s^{-1}}$ and $D_3^r = 1.02\cdot10^4\,\mathrm{s^{-1}}$ for the rotational diffusivity tensor. The square roots of the body-based mobility tensors, as needed in the evaluation of the random terms, are then readily obtained by taking the square roots of the diagonal elements. The time step is set at $10^{-8}\,\mathrm{s}$, which is the limit dictated by the rotational Brownian motion of the triskelia.

4.2.2 Adaptor Protein 2

The AP2 model

In a previous study, [51] we introduced a coarse-grained model for adaptor proteins inspired by the AP2 complex, and used it to study their role in the assembly process of clathrin triskelia in bulk through Monte Carlo simulations. Only those parts of the AP2 complex involved in clathrin binding were represented in the model. Here the model is extended to include the core section of the protein, which was excluded in the previous model as it plays no role in the clathrin binding process in bulk. The two clathrin binding sites present in the C-terminal region of the

β linker, *i.e.* residues 631–635, and the β appendage domain formed by residues 705–937, are modelled by point particles, β_1 and β_2 respectively, see Fig. 5.1. The flexible, structureless linker connecting the two binding sites is composed of about 70 residues, with an estimated contour length of $L_{12} \approx 26\,\text{nm}$. The linker, which acts as an entropic spring, [74] is modelled through a finite extensible nonlinear elastic (FENE) potential, [75]

$$U_{12}(r) = \begin{cases} -\frac{1}{2}k_{12}L_{12}^2 \ln\left[1 - \left(\frac{r}{L_{12}}\right)^2\right] & \text{for} \quad r < L_{12} \\ \infty & \text{for} \quad r \geq L_{12}, \end{cases} \tag{4.15}$$

where r denotes the distance between the two beads. The spring constant k_{12} is estimated using an expression from polymer physics, [74]

$$k_{12} = \frac{3k_BT}{2L_{12}l_p}, \tag{4.16}$$

where l_p denotes the persistence length. With an experimental value of $l_p \approx 0.6\,\text{nm}$ for disordered proteins, the resulting spring constant for the linker becomes $k_{12} = 30k_BT/\sigma^2$.

The core section of the AP protein is responsible for binding with the membrane and cargo. Following the above reductionist approach, it is represented by an additional point particle, β_0 , and connected by a flexible tether to the β_1 point particle, see Fig. 5.1. This section of the linker, representing the initial section of the β linker and composed of about 40 residues, has a contour length of $L_{01} \approx 15\,\text{nm}$ and is also modelled through a FENE potential. The spring constant is estimated through Eq. (5.2) as $k_{01} = 50k_BT/\sigma^2$. The α linker and α appendage of AP2 are not represented in the model as they do not bind with clathrin nor with the membrane.

Interaction with clathrin

APs possess two different binding sites for clathrin that specifically bind to two matching sites on a triskelion leg. The presence of a binding site at the end of the TD matching the β_1 site on the AP2 linker is well established [12]. The AP2 β_2 site binds clathrin near the ankle [16] and possibly also at a site higher up along the leg [59]. In previous work, we explored the effect of sites located at the ankle and at the knee on cage assembly in solution, establishing that both possibilities permit AP-regulated clathrin aggregation [51]. Our study revealed that the mechanical properties of the AP linker region play a crucial role in regulating the assembly mechanism. Here, the β_2 bead binds to the clathrin ankle only, as this is the accepted site in the current literature.

To overcome the numerically inconvenient short range of the AP-clathrin site-site interaction, we developed a clicking potential to describe the AP binding interactions with clathrin [51]. Consider the interaction between the α^{th} binding site on the i^{th} triskelion and the β^{th} bead on the j^{th} AP. Said interaction is described by a flag $b_{i\alpha,j\beta}$ that assumes two values reflecting the discrete state of the interaction. In the unclicked state, $b_{i\alpha,j\beta} = 0$, there is no interaction. In the clicked state, $b_{i\alpha,j\beta} = 1$, the interaction is characterized by a fixed interaction energy ϵ_{CA} and a maximum distance ρ_{CA} between the two binding sites. The interaction potential then reads as

$$\phi\text{click}(r_{i\alpha,j\beta}, b_{i\alpha,j\beta}) = \begin{cases} 0 & \text{for} \quad b_{i\alpha,j\beta} = 0 \\ \begin{cases} -\epsilon_{\text{CA}} & \text{for} \quad r_{i\alpha,j\beta} < \rho_{\text{CA}} \\ \infty & \text{for} \quad r_{i\alpha,j\beta} \geq \rho_{\text{CA}} \end{cases} & \text{for} \quad b_{i\alpha,j\beta} = 1. \end{cases} \tag{4.17}$$

The two distinct clathrin-AP bounds, β_1 binding toes and β_2 binding ankles, are probably characterized by different interaction parameters. Since neither interaction free energy has been

measured, to the best of our knowledge, we here assume that both bonds are equally strong and explore a range of values for ϵ_{CA} in the simulations. The same radius $\rho_{\mathrm{CA}} = 0.25\sigma$ is used to describe both interactions.

The click interactions are turned on and off by an MC procedure. Attempts to modify the interaction status of a randomly selected AP binding site are made at a rate of r_{CA} attempts per second and per AP bead. A list is made of all K possible binding sites within the interaction radius ρ_{CA} of that bead, including the current site if applicable. The unclicked state is also considered, as the 0^{th} option. For each of these $K+1$ trial moves, the accompanying energy change $\Delta\phi_k^{\mathrm{click}}$ is calculated, which can only assume the values 0 and $\pm\epsilon_{\mathrm{CA}}$, and one of the states is selected with probability

$$P_k = \frac{\exp(-\beta\Delta\phi_k^{\mathrm{click}})}{\sum_{k'=0}^{K}\exp(-\beta\Delta\phi_{k'}^{\mathrm{click}})}. \tag{4.18}$$

Note that the coordinates of the particles are left untouched. Excluded volume interactions between AP beads are omitted for reasons of computational efficiency. We note, however, that excluded volume interactions play an important role in preventing multiple AP beads binding to the same clathrin site: this is effectively inhibited in the model by making the clicks mutually exclusive, allowing the AP and clathrin sites to partake in only one interaction at a time. We furthermore note that the total volume of all APs bound to a cage is significantly smaller than the interior volume of the cage.

Propagator

The positions of the three particles composing an AP are updated separately, in a manner that depends on their clicking status. An unclicked bead follows a translational BD equation of motion, see Eq. (4.9), with an isotropic mobility tensor $\boldsymbol{\mu}_\beta^t = \mu_\beta^t\mathbf{I}$, where $\mathbf{I}$ is the identity matrix. The mobility μ_β^t is estimated through the Stokes-Einstein equation for isolated Brownian spheres,

$$\mu_\beta^t = \frac{1}{6\pi\eta R_\beta}, \tag{4.19}$$

where η is the viscosity of the suspending fluid and R_β the sphere's radius. Assuming a diameter [76] of $\sim$9 nm for the 200 kDa AP2 core and a diameter of $\sim$4 nm for the appendage domain, the mobilities of the β_0 and β_2 beads become $\mu_0^t = 1.2\cdot10^{10}\,\mathrm{Ns/m}$ and $\mu_2^t = 2.6\cdot10^{10}\,\mathrm{Ns/m}$, respectively, in water at room temperature. Unlike those two bulky sites, the motion of the four-residue β_1 site will be dominated by the dynamics of the linker. For reasons of simplicity, the β_1 bead is simulated as a Brownian particle with the same properties as the β_2 bead. The beads do not rotate.

We now turn to clicked beads. All AP beads clicked to a triskelion move with that triskelion; the aggregate of a triskelion and one or several attached AP beads moves as a single rigid body, obeying Eqs. (4.9) and (4.12), subject to the sum of all the forces and torques acting on the aggregate. This preserves the clicking status of the AP beads involved, as well as their relative positions in the body frame of the triskelion. The translational and rotational mobilities of the aggregate are assumed to be identical to those of a free triskelion. On top of this motion, the AP beads explore the small clicking volume by Monte Carlo trial moves taking them to random positions withing this volume, at a rate of r_v attempts per second and per bead. These trial moves also preserve the clicking status, but alter the potential energy. Trial moves are accepted or rejected with a probability given by the Metropolis scheme [77, 78]

$$P_{o\to n}^{\mathrm{acc}} = \min\left\{1, e^{-\beta\Delta\Phi_{o\to n}}\right\}, \tag{4.20}$$

where $\Delta\Phi_{o\to n} = \Phi(n) - \Phi(o)$ is the potential energy change between the old configuration o and the new configuration n.

Chemostat

In the relatively small volume of the simulation box, the adhesion of APs and triskelia to the membrane will strongly deplete their concentrations in the dissolved phase. To counter this effect, a chemostat is employed to maintain a constant chemical potential, and hence constant concentration in the bulk phase, by the insertion and removal of particles, thus mimicking exchange with a large ideal reservoir. A Monte Carlo algorithm is applied to both solute molecule types independently; in the following we refer to clathrin as an illustrative example. Trial moves to exchange triskelia with the reservoir are attempted at a rate of r_c attempts per second. The trial move consists in adding or removing a triskelion with equal likelihood. Consider a system with N_C triskelia in a volume V. The probability of acceptance of an insertion move is given by [69, 78]

$$P^{\mathrm{acc}}_{N_\mathrm{C}\to N_\mathrm{C}+1} = \min\left\{1, \frac{[\mathrm{C}]_r V}{N_\mathrm{C}+1} e^{-\beta\Delta\Phi_{N_\mathrm{C}\to N_\mathrm{C}+1}}\right\}, \tag{4.21}$$

where $[\mathrm{C}]_r$ refers to the clathrin concentration in the reservoir and $\Delta\Phi_{N_\mathrm{C}\to N_\mathrm{C}+1}$ is the potential energy change accompanying the insertion of a triskelion at a random position and with a random orientation. Similarly, the acceptance probability for the removal of a randomly selected triskelion is given by

$$P^{\mathrm{acc}}_{N_\mathrm{C}\to N_\mathrm{C}-1} = \min\left\{1, \frac{N_\mathrm{C}}{[\mathrm{C}]_r V} e^{-\beta\Delta\Phi_{N_\mathrm{C}\to N_\mathrm{C}-1}}\right\}. \tag{4.22}$$

Similar equations apply for the AP2 chemostat.

The chemostat algorithm causes the numbers of triskelia and APs to fluctuate throughout a simulation. For an ideal gas at constant chemical potential, volume and temperature, the equilibrium distribution of the number of molecules follows a Poisson distribution. A test simulation was run using reference concentrations of $[\mathrm{C}]_r = 0.01\sigma^{-3}$ and $[\mathrm{A}]_r = 0.01\sigma^{-3}$, respectively, with all inter-molecular interactions turned off. Under these conditions, the system should behave like a mixture of two ideal gases featuring different internal degrees of freedom. Good numerical agreement was observed between the averages and variances of the number of molecules present and their theoretical values (data not shown).

4.2.3 Membrane

The membrane model

The free energy of a membrane is well described by the Helfrich expression, [74, 79]

$$F = 2\kappa \int_A H^2 da + \frac{1}{2} K_A A_0 \left(\frac{A - A_0}{A_0}\right)^2, \tag{4.23}$$

where the integral runs over the area A of the membrane, H denotes the local mean curvature, κ the bending rigidity, K_A is the elasticity modulus, and A_0 is the equilibrium surface area. In this form, the equation applies to membranes with no intrinsic curvature, fixed topology and without edges. The coarse grained membrane model aims at reproducing Helfrich's free energy.

particle	parameter	symbol	value
clathrin	segmental length	σ	17 nm
	pucker	χ	101°
	intersegment energy[2]	ϵ	$6k_BT$[3]
	transl. diff. coeff.	D_1^t, D_2^t	$4.46 \cdot 10^4 \sigma^2/\mathrm{s}$
		D_3^t	$3.70 \cdot 10^4 \sigma^2/\mathrm{s}$
	rotat. diff. coeff.	D_1^r, D_2^r	$1.64 \cdot 10^4/\mathrm{s}$
		D_3^r	$1.02 \cdot 10^4/\mathrm{s}$
	bulk concentration[4]	$[\mathrm{C}]_r$	$10^{-3}\sigma^{-3}$
AP	linker length	L_{01}	0.9σ
		L_{12}	1.5σ
	spring constant	k_{01}	$50k_BT/\sigma^2$
		k_{12}	$30k_BT/\sigma^2$
	transl. diff. coeff.	D_0^t	$2 \cdot 10^4 \sigma^2/\mathrm{s}$
		D_1^t, D_2^t	$1 \cdot 10^5 \sigma^2/\mathrm{s}$
	bulk concentration	$[\mathrm{A}]_r$	$\left[10^{-3}{-}10^{-2}\right]\sigma^{-3}$
membrane	bending rigidity	k_κ	$\left[10{-}30\right]k_BT$
	spring constant	k_m	$100k_BT$
	equilibrium length	r_m^0	0.5σ
	maximum length	L_m	0.8σ
	excluded volume	$\sigma_\mathrm{m}^\mathrm{ev}$	0.35σ
		$\epsilon_\mathrm{m}^\mathrm{ev}$	$10k_BT$
	transl. diff. coeff.	D_m^t	$10^2\sigma^2/\mathrm{s}$
AP – CL	click strength	ϵ_CA	$\left[6{-}14\right]k_BT$
	click radius	ρ_CA	0.25σ
AP – mb	click strength	ϵ_Am	$\left[6{-}14\right]k_BT$
	click radius	ρ_Am	0.1σ
mb – X[5]	repulsion	ϵ_m	$10^3 k_BT/\sigma^4$
	vertical range	b	0.1σ
	horizontal range	d	1.0σ

Table 4.2: Summary of the applied simulation parameters for clathrin, AP and membrane particles, as well as their interaction parameters.

We simulate the membrane through a mesh composed of N_m beads, connected by $3N_m$ flexible tethers forming a multi-faceted surface composed by $3N_m/2$ triangles. The elasticity is accounted for by modeling the flexible tethers through a FENE potential,

$$U_{\mathrm{m}}^{\mathrm{el}}(r) = \begin{cases} -\frac{1}{2}k_{\mathrm{m}}L_m^2 \ln\left[1-\left(\frac{r-r_{\mathrm{m}}^0}{L_{\mathrm{m}}}\right)^2\right] & \text{for} \quad |r-r_{\mathrm{m}}^0| < L_{\mathrm{m}} \\ \infty & \text{for} \quad |r-r_{\mathrm{m}}^0| \geq L_{\mathrm{m}}, \end{cases} \tag{4.24}$$

where k_m denotes the spring constant, and L_m the maximum deviation from the equilibrium length r_m^0. Self-avoidance of the membrane is guaranteed by introducing excluded volume interactions between beads, modelled through a Weeks-Chandler-Andersen (WCA) potential,

$$U_{\mathrm{m}}^{\mathrm{ev}}(r) = \begin{cases} 4\epsilon_{\mathrm{m}}^{\mathrm{ev}}\left[\left(\frac{\sigma_{\mathrm{m}}^{\mathrm{ev}}}{r}\right)^{12} - \left(\frac{\sigma_{\mathrm{m}}^{\mathrm{ev}}}{r}\right)^{6} + \frac{1}{4}\right] & \text{for} \quad r < 2^{1/6}\sigma_{\mathrm{m}}^{\mathrm{ev}} \\ 0 & \text{for} \quad r \geq 2^{1/6}\sigma_{\mathrm{m}}^{\mathrm{ev}}. \end{cases} \tag{4.25}$$

The radius $\sigma_{\mathrm{m}}^{\mathrm{ev}} = 0.35\sigma$ and strength $\epsilon_{\mathrm{m}}^{\mathrm{ev}} = 10k_BT$ are chosen such that membrane beads are prevented from crossing the surface of any membrane triangle.

A bending potential is introduced between neighbouring triangles. Several potentials have been proposed in the literature in which the mean curvature is calculated either through the angles formed by adjacent triangles [79–81] or using a lattice dual to the membrane triangulation [79, 81–83]. The latter approach, unlike the former, yields a bending rigidity that does not vary with the membrane topology [81, 84] and is therefore followed here. For computational convenience, we use a simplified version of this potential, [82]

$$\Phi = \frac{1}{2}k_\kappa \sum_i \phi i, \tag{4.26}$$

where k_κ denotes the strength of the potential. The summation runs over all membrane beads, with

$$\phi i = \frac{1}{\Omega_i}\Big[\sum_{j(i)}(\mathbf{x}_i - \mathbf{x}_j)\Big]^2, \tag{4.27}$$

where the sum runs over all nodes j connected to i and Ω_i represents the collective area of all triangles joining at site i.

Propagator

The motion of individual membrane nodes is described by a translational BD equation of motion, see Eq. (4.9), with a constant isotropic mobility tensor $\boldsymbol{\mu}_{\mathrm{m}}^t = \mu_{\mathrm{m}}^t\mathbf{I}$. Membrane thermal undulations with wavelengths comparable to the clathrin size relax much faster than the time required for clathrin to diffuse over its own length [85]. Since we want to simulate both processes, we use the expedient of slowing down the membrane motion by judiciously choosing the highest mobility scalar μ_{m}^t that maintains a numerically stable membrane when using the maximum time step suitable for the triskelia, thus arriving at a diffusion coefficient of $D_{\mathrm{m}}^t = \mu_{\mathrm{m}}^t k_BT = 100\sigma^2/\mathrm{s}$.

Networks with fixed connectivity behave as elastic struts and prove unable to form pit-like invaginations under appropriate forces. To model the fluidity and deformability of the membrane, Monte Carlo moves are used to modify the network connectivity [79]. A random bond is selected and flipped to connect the tops of the two triangles that hitherto shared this bond, as illustrated in Fig. 4.3. This move leaves the positions of the nodes untouched, but nevertheless it is likely to result in a change of the potential energy of the flipping bond and of the bending energies

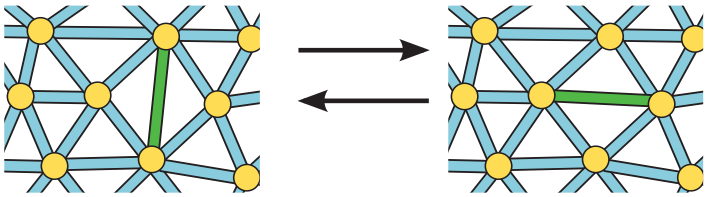


Figure 4.3: Cartoon of the Monte Carlo bond flipping move, employed in the simulations to ensure membrane fluidity and enable large amplitude deformations. A randomly selected bond, here coloured in green, is flipped from one diagonal to the other diagonal of the quadrangle formed by its two adjacent triangles.

of the neighbouring triangles. The acceptance probability of the flipping move is again given by the Metropolis scheme in Eq. (5.15). Note that the total number of triangles is conserved, while the connectivity numbers of the four nodes involved in the bond swap increase or decrease by one. In order to conserve the geometrical properties of the triangulated network, any attempt that reduces the connectivity of a node below five is rejected [79]. Bond flips are attempted at a rate of r_f attempts per bond and per second.

Interaction with other particles

The membrane – with a main orientation parallel to the xy plane – floats roughly in the middle of the simulation box and thereby divides the box into an 'interior' region above the membrane and an 'exterior' region below the membrane (where we, for the sake of argument, temporarily ignore the periodic boundary conditions along the z axis). The presence of particles only in the interior region will result in an osmotic pressure that pushes the membrane down; this effect is removed by employing periodic boundary conditions in the z direction, because this allows the particles to diffuse freely between the two regions and thereby equalizes pressures and chemical potentials on both sides of the membrane. A flag l_i is assigned to every triskelion and AP molecule to distinguish between those in the interior, $l_i = +1$, and those in the exterior, $l_i = -1$. When a particle i crosses the periodic boundaries at $z = \pm\frac{1}{2}L_z$, its flag l_i reverses sign.

The interaction of AP beads with the membrane comprises two parts, a repulsion that prevents the β_1 and β_2 beads from crossing the membrane and a click potential that enables the core, *i.e.* β_0, to bind to a membrane bead. The repulsive bead-membrane potential should allow particles to come close to the membrane but not to cross the membrane permanently, a combination that can not be realized in BD simulations by merely introducing a repulsive potential between beads and membrane nodes. In stead, we introduce an interaction between particle i and membrane triangles t of the form

$$\phi_i^{\mathrm{m}} = \epsilon_{\mathrm{m}} \sum_t f(l_i r_{it}^{\perp}) g(r_{it}^{||}), \tag{4.28}$$

where ϵ_{m} is a positive strength parameter. Using $\hat{\mathbf{n}}_t$ as the normal to the triangle, pointing to the interior, one may calculate the height of the particle above the membrane, $r_{it}^{\perp} = \mathbf{r}_{it} \cdot \hat{\mathbf{n}}_t$, and its lateral displacement along the membrane, $r_{it}^{||} = |\mathbf{r}_{it} - r_{it}^{\perp}\hat{\mathbf{n}}_t|$, with $\mathbf{r}_{it}$ being the vector pointing from the center of mass of the triangle to the bead. The above introduced flag l_i is included in the argument of f to ensure that the resulting repulsive force always points in the

correct direction, *i.e.* to the interior for $l_i = +1$ and to the exterior for $l_i = -1$, even if the bead has slightly penetrated the membrane.

The penalty function f is zero for particles beyond a height b, is quadratic just above the membrane triangle, and increases linearly for particles penetrating the membrane,

$$f(x) = \begin{cases} 0 & \text{for} \quad x \geq b \\ (x-b)^2 & \text{for} \quad 0 < x < b \\ b^2 - 2bx & \text{for} \quad 0 \leq x. \end{cases} \tag{4.29}$$

The function g distributes the interaction over several adjacent triangles, reflecting the size of the protein fragment interacting with the membrane,

$$g(x) = \begin{cases} (x-d)^2 & \text{for} \quad 0 < x < d \\ 0 & \text{for} \quad x \geq d, \end{cases} \tag{4.30}$$

with d setting the decay length of the spread function. In the simulations, $\epsilon_{\mathrm{m}} = 10^3 k_B T/\sigma^4$, $b = 0.1\sigma$ and $d = 1\sigma$.

The above repulsion applies to the β_1 and β_2 beads of AP, as well as to the hubs, knees and ankles of triskelia. The core particle of AP is excluded to permit binding of this particle to the membrane, as discussed below. A modest flexibility of the clathrin protein, allowing a leg to bend when pressed against a membrane, is mimicked in the simulations by a reduction by 90% of the repulsion strength ϵ_{m} for the extremal parts of the triskelion, *i.e.* the ends of the TDs.

The short-ranged binding interaction between the AP core, β_0 , and a membrane node is described, like the equally short-ranged AP-clathrin binding interaction, by the clicking potential introduced in section 4.2.2. Attempts are made, at a rate of r_{Am} attempts per AP and per second, to click a free core with any available membrane node within a clicking radius of $\rho_{\mathrm{Am}} = 0.1\sigma$ or to unclick a bound core. Upon clicking (unclicking), the system energy is lowered (raised) by the click energy ϵ_{Am}. Clicks are again mutually exclusive, to prevent two APs from binding to the same membrane bead. Structural studies have revealed that the AP2 protein undergoes a large conformational change from a 'locked' state in the cytosol to an 'open' state when bound to the membrane [14, 15]. In the locked state, the β_1 site is buried within the core and only the β_2 site is available for clathrin binding, thus inhibiting APs from bridging two triskelia and thereby effectively suppressing the cage assembly process in the cytosol. Upon binding to the membrane, the β_1 site is released and the AP2 complex can bind two triskelia. This mechanism is mimicked in the simulations by allowing a β_1 bead to click to the end of a TD only when the matching core bead is clicked to a membrane node. Since a membrane-bound AP2 complex can diffuse relative to the membrane, in the simulation MC moves are included that enable the AP core bead to jump from its membrane node to a randomly selected neighbouring node, at a rate of r_j attempts per membrane bound AP and per second. The acceptance probability of trial moves to unoccupied nodes is again given by Eq. (5.15), while trial jumps to occupied nodes are rejected.

Monte Carlo Barostat

Together with the bending rigidity, tension is a key factor determining the membrane deformability. A change of the in-plane tension can be induced by an affine rescaling of the simulation box, thus stretching or compressing all tethers forming the triangulated mesh. A vanishing tension is imposed on the system through Monte Carlo rescaling moves affecting the two in-plane directions x and y independently, while preserving the size of the box in the z direction. Attempts are made at a rate of r_b per second to change the length of a selected box edge from the

function	symbol	value (s^{-1})
chemostats, CL and AP	r_c	10^5
barostat	r_b	10^6
membrane bond flipping	r_f	10^4
AP-clathrin click	r_{CA}	10^6
AP-membrane click	r_{Am}	10^6
reshuffle clicked AP	r_v	10^4
AP-membrane jumps	r_j	10^4

Table 4.3: Attempt rates of all Monte Carlo moves discussed in the main text. The five values in the second half of the table denote rates per particle, to eliminate system size dependence.

old value L_o to a new value $L_n = L_o + \Delta L$, where ΔL is a random number uniformly distributed over the interval $[-0.1, 0.1]\,\sigma$. Rescaling causes a volume change of the simulation box from V_o to V_n. The affine transformation is applied to all membrane beads, the hydrodynamic centers of the triskelia and all unclicked AP beads; clicked AP beads move with the bead they are connected to, in order to preserve their click statuses. The probability of accepting the trial move as the new configuration is given by [78]

$$P^{\mathrm{acc}}_{o\to n} = \min\left\{1, \left(\frac{L_n}{L_o}\right)^N e^{-\beta[\Delta\phi o\to n + P(V_n - V_0)]}\right\}, \tag{4.31}$$

where N is the total number of molecules in the system – including the membrane, counted as one molecule – and P the imposed pressure. To obtain a tensionless membrane, P is set to zero. Since the two dimensions are scaled independently, the initial shape of the box is not conserved throughout the simulation.

Model validation

The mechanical properties of the membrane were tested using a system containing 1188 membrane nodes, initially prepared in a regular hexagonal lattice. The equilibrium length of the tethers was set at $r_m^0 = 0.5\sigma$ and the maximum extensibility at $L_m = 0.8\sigma$. Typical experimental values for giant unilamellar vesicles of fluid-phase double-tailed phosphatidylcholine lipids at room temperature were selected for the bending rigidity, $\kappa = 0.4 \cdot 10^{-19}\mathrm{J} = 10 k_B T$, and the elastic modulus, $K_A = 250\,\mathrm{mJ/m^2}$ [86]; the former is realised in the simulation model by $k_\kappa = \kappa$, the latter through the formula $k_m = 2K_A/\sqrt{3}$ obtained by relating area expansion to tether elongation in a hexagonal lattice. During simulations with this membrane, the MC steps to swap bonds were never accepted and consequently the membrane displayed anisotropic behaviour. Moreover, when initiating simulations with a membrane containing a hand-made pit-like dome, which required the introduction of beads with 5 or 7 neighbours, the pit was never observed to collapse spontaneously. Since the main properties of a membrane relevant for the endocytic process are its non-permeability and bending stiffness, we found it expedient to lower the elastic modulus to $1\,\mathrm{mJ/m^2} \approx 70 k_B T/\sigma^2$. The resulting model membrane permits bond swaps, shows isotropic behaviour, and artificially introduced domes readily collapse.

The elastic modulus of the model membrane was determined by a series of runs uniformly stretching both in-plane box dimensions from $L_\parallel = 220\,\sigma$ to $230\,\sigma$, with the barostat turned off. The tension on the membrane $\tau_\parallel$ was calculated from the diagonal elements of the stress tensor, $P_{\alpha\beta}$, where the latter was obtained by the virial expression [69, 78]. When the elastic

contribution dominates the Helfrich free energy, the tension is related to the relative elongation of the membrane by [87, 88]

$$\begin{aligned}\tau_{\parallel} &= L_z \left[P_{zz} - \frac{1}{2}(P_{xx} + P_{yy}) \right] \\ &= K_{A\parallel 0} \frac{A_{\parallel} - A_{\parallel 0}}{A_{\parallel 0}},\end{aligned} \tag{4.32}$$

where the membrane area is approximated by the ground plane area of the simulation box, $A_{\parallel} = L_{\parallel}^2$. The stress-strain curve of the membrane is shown in Fig. 4.4. From the slope of the tangent at the zero-tension intercept, at $A_{\parallel 0} \approx 224.4\sigma^2$, we obtain an effective elastic modulus $K_{A\parallel 0} \approx 58 k_B T/\sigma^2$. The curve is not the straight line predicted by Eq. (4.32), and the elastic modulus is less then expected, because the contributions of the membrane undulations to both the free energy and the membrane area are ignored in that equation. Both can be corrected for, yielding [86, 88, 89]

$$K_{A\parallel 0}^{-1} = K_A^{-1} + \frac{k_B T}{32\pi^3 \kappa^2} A_{\parallel 0}. \tag{4.33}$$

Upon inserting the bending rigidity (to be discussed next) we obtain $K_A = 67 k_B T/\sigma^2$, in good agreement with the input value. This value is also approached by the slope of the curve in the large area limit of Fig. 4.4, where the high tension has effectively suppressed the undulations. Upon turning the barostat on, the areas of the simulation boxes were observed to converge to $A_{\parallel 0}$.

The undulations of the membrane are commonly expressed by a superposition of two-dimensional tranverse waves. Taylor expanding the Helfrich free energy to second order in the complex amplitudes $c_{\mathbf{q}}$ of a Fourier series with wave vectors $\mathbf{q}$, followed by applying the equipartition theorem, yields the structure factors [87, 90]

$$S(\mathbf{q}) = \langle |c_{\mathbf{q}}|^2 \rangle = \frac{k_B T}{A_{\parallel}(\kappa q^4 + \tau q^2)}, \tag{4.34}$$

where τ is the tension on the membrane. Figure 4.5 shows the results for two barostatted membranes with $k_\kappa = 10 k_B T$ and $20 k_B T$, respectively. From the slopes of the fitted straight lines, we extract effective bending rigidities of $\kappa = 11 k_B T$ and $18 k_B T$, respectively, in good agreement with the input values. The small non-zero intercept indicates that there is a small residual tension on the membrane, which arises from a subtle difference between the tension τ measured in this plot and the tension $\tau_{\parallel}$ nullified by the barostat [88, 91].

4.3 Results

4.3.1 Membrane coverage

The beads of our membrane model perform three roles. Firstly, the positions of the beads describe the configuration of the membrane and the interactions between the beads endow the membrane with an equilibrium area, an elastic modulus and a bending rigidity. Secondly, the multi-faceted plane formed by the beads act as a barrier that prevents molecules from crossing the membrane. Thirdly, the membrane beads act as binding sites for APs, in the same way that the combination of PtdIns(4,5)P_2 and cargo functions as the membrane-bound binding partner for the AP2 complex *in vivo*. The latter role is the focus of the current subsection. AP2 is modelled as a chain of three beads connected by springs, see Fig. 5.1. The first bead, β_0 ,

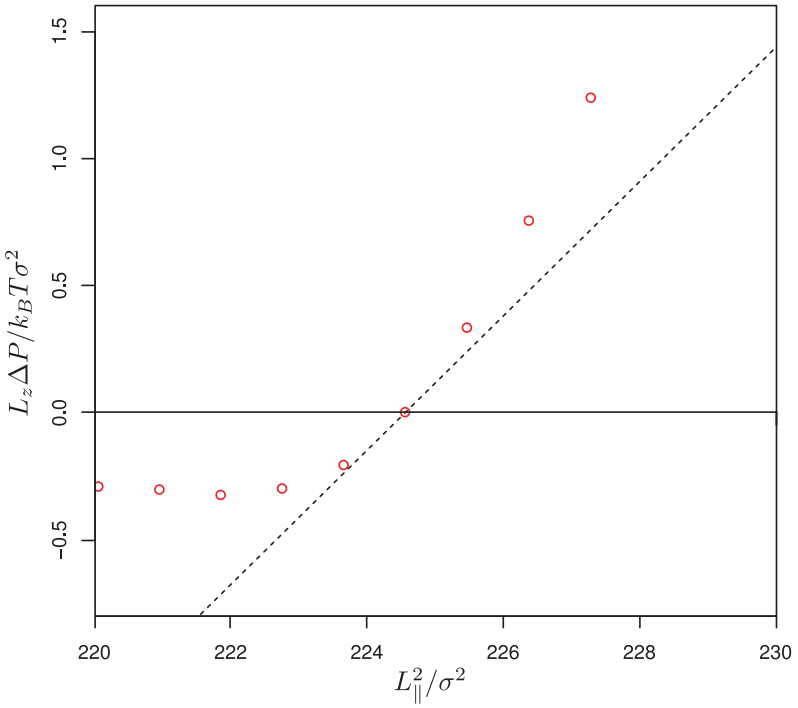


Figure 4.4: The tension on the membrane as a function of the area of the ground plane of the simulation box. The non-linearity of the curve reflects the presence of membrane undulations: undulations are suppressed under elongation and promoted under compression, causing the membrane to buckle for the smallest areas. The slope of the dotted line, *i.e.* the tangent at the tensionless state, yields the effective elastic modulus $K_{A\|0}$.

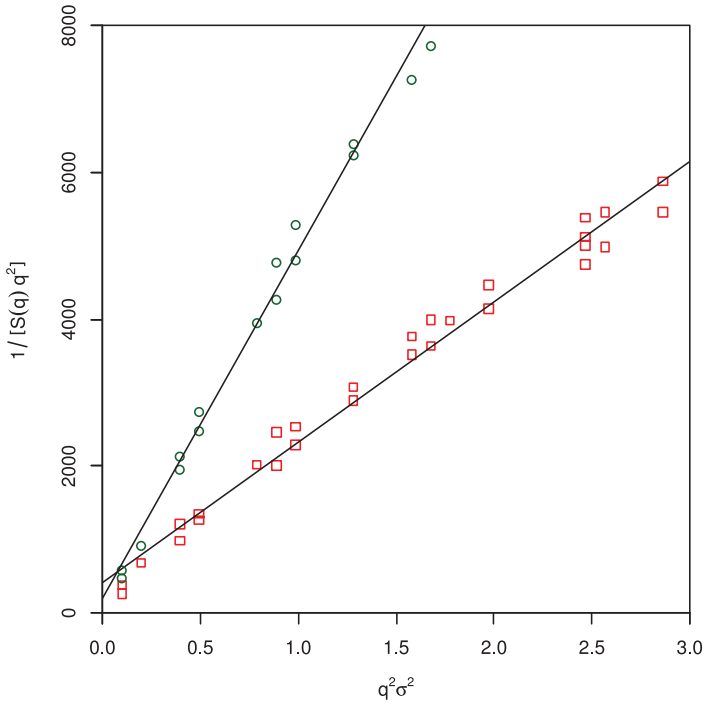


Figure 4.5: Structure factors of thermal membrane undulations S versus wave number q, plotted in a format that linearizes their theoretical relationship by Eq. (4.34). The green circles and red squares mark simulations at two differing input values of the bending rigidity k_κ.

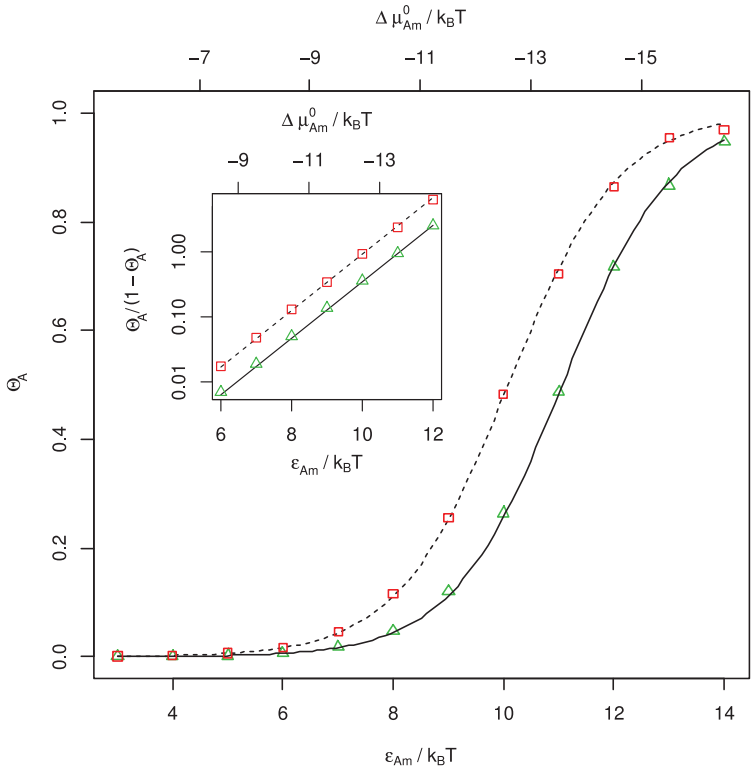


Figure 4.6: The fraction θA of membrane beads occupied by AP as a function of the AP-membrane binding strength ϵ_{Am}, for a membrane exposed to an AP solution of $\left[\mathrm{A}\right] = 10^{-2}\sigma^{-3} \approx 3.4 \cdot 10^{-6}$ molar, with the restriction potential preventing AP's linker beads β_1 and β_2 from crossing the membrane turned off (red squares, dashed line) and turned on (green triangles, solid line). Markers represent averages over simulations of 10 s each, the lines are obtained by the statistical-mechanical theory outlined in Appendix 4.5.1. The inset shows a linear representation of the same data.

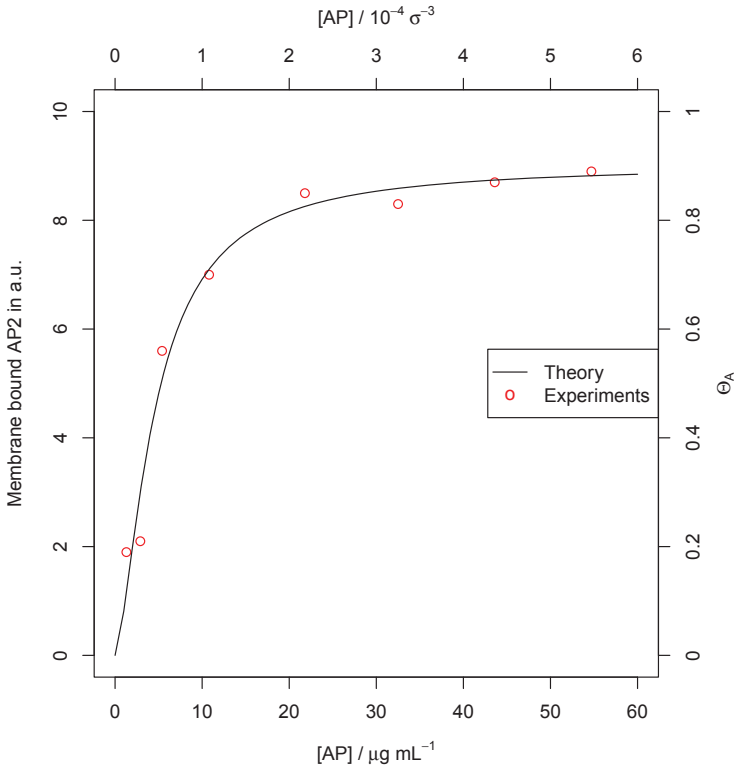


Figure 4.7: The coverage of a membrane by AP2 as a function of the AP2 concentration in the bulk, in arbitrary units (a.u.). Markers show experimental data by Chang *et al.*, [92] the solid line is a fit with the Langmuir theory, see Eq. (4.35).

represents the core and can bind to a membrane bead. Beads β_1 and β_2 along the β linker are capable of binding to the clathrin terminal domain and knee, respectively.

When the membrane is exposed to a solution of APs, a number of membrane beads will become occupied by one AP-core bead each. The fractional occupancy θA is expected to be a function of the bulk AP concentration, $[\mathrm{A}]$, and the strength of the binding interaction, ϵ_{Am}. This relation was explored by varying the bond strength in a series of simulations using a membrane of 2064 nodes. A chemostat simulates the exchange of APs with a hypothetical ideal AP reservoir and thereby maintains a bulk AP concentration of $0.01\sigma^{-3}$, irrespective of the number of APs attached to the membrane. To minimize interference with the processes at the membrane and to avoid possible superpositions of inserted particles with the membrane surface, the chemostat is applied only to a section of the box well below the membrane, *i.e.* in the 'exterior' to the cell, covering ∼25% of the total box volume. Diffusion of molecules between the 'exterior' and 'interior', via the periodic boundary conditions along the z direction, equalizes the chemical potential and osmotic pressure throughout the box.

To accelerate the realization of thermodynamic equilibrium and to improve sampling, in this subsection only, the membrane is configured as an immobile ideal hexagonal lattice, the barostat is turned off and APs are not permitted to jump diffusively between membrane beads – note that these expedients do not affect the equilibrium occupancy. We remind the reader that the interaction between membrane-clicked APs is designed to prevent multiple APs from binding to the same membrane node, thus mimicking explicit excluded volume interactions that are omitted for reasons of computational efficiency. Clathrin triskelia are absent in this initial set of simulations.

The simulation results collected in Fig. 4.6 show a smooth transition, from a low occupancy of the membrane beads to their near saturation with APs, in a narrow region of clicking strengths ϵ_{Am}. A statistical-mechanical derivation of this Langmuir-like behaviour [93] is presented in Appendix 4.5.1, yielding

$$\frac{\theta\mathrm{A}}{1-\theta\mathrm{A}} = \frac{[\mathrm{A}]}{c_0} e^{-\beta\Delta G^0_{\mathrm{Am}}}, \tag{4.35}$$

where c_0 denotes the standard reference concentration of 1 molar, and ΔG^0_{Am} is the change in the standard state free energy accompanying the AP-membrane binding reaction. That is,

$$\Delta G^0_{\mathrm{Am}} = \mu^0_{\mathrm{Am}} - \mu^0_{\mathrm{A}} - \mu^0_{\mathrm{m}} = \Delta\mu^0_{\mathrm{Am}}, \tag{4.36}$$

with μ^0_{X} the reference chemical potential of component X at the standard concentration and $\Delta\mu^0_{\mathrm{Am}}$ the standard chemical potential difference associated with the reaction. The agreement between theory and simulations is very good, see Fig. 4.6, both in the absence and presence of the potential that prevents the two AP linker beads β_1 and β_2 from crossing the membrane. The reduced configurational freedom imposed by this restraint causes a shift of the equilibrium fractional coverage function to slightly higher binding energies. Chang *et al.* [92] measured the saturation binding of AP2 on plasma membrane fragments prepared by freeze-thaw lysis of cells, which were stripped of their endogenous coat proteins before exposure for 30 minutes to AP2 solutions of various concentrations. Their experimental data are well described by the Langmuir theory, see Fig. 4.7, yielding $\Delta\mu^0_{\mathrm{Am}} = -17.8 k_B T$.

Let us now consider a mixture of clathrin and APs, both subject to a chemostat. As in our preceding simulation studies, [48, 49] clathrin is modelled as a rigid triskelion with kinked legs following the characteristic shape of the protein, see Fig. 5.1. Attractions (repulsions) are introduced between pairings of straight leg segments that are (are not) present in experimental cage edges. The interactions were originally tuned to simulate *in vitro* self-assembly in a slightly

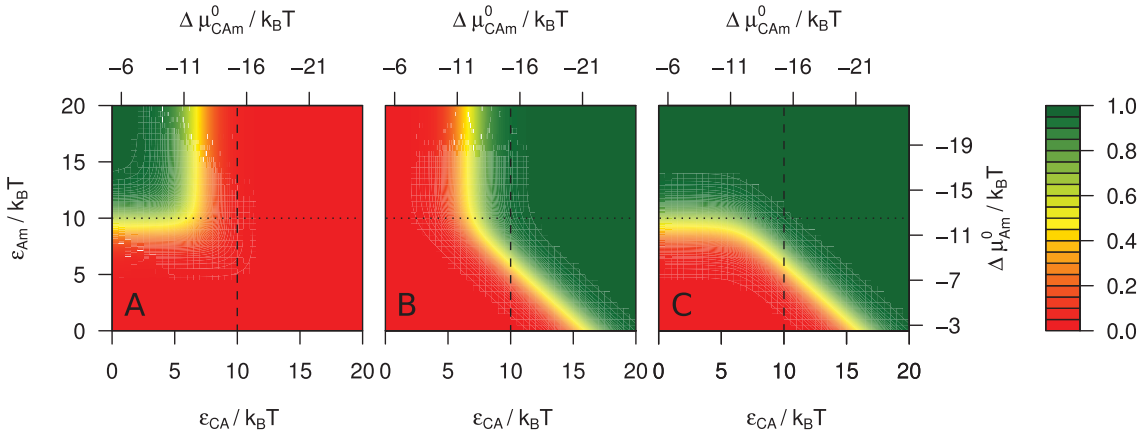


Figure 4.8: The theoretical fractional occupancies θ of membrane beads (A) by AP, θA, (B) by AP-clathrin complexes, θCA, and (C) the sum of these two occupancies. In order to make this theoretical calculation feasible, AP has been curtailed to one clathrin binding site, which is activated only when AP is bound to the membrane, the triskelia have only one site that clicks to AP, and the penetration barrier of the membrane is turned off. The green triangular area for $\Delta\mu^0_{\mathrm{Am}} \gtrsim -12k_BT$ in the two rightmost plots, while the leftmost plot is red at these low membrane-AP binding strengths, highlights a cooperative effect with APs binding to the membrane only in the presence of triskelia. The concentrations of free monomers in the solution are fixed at $\left[\mathrm{A}\right] = 10^{-2}\sigma^{-3}$ and $\left[\mathrm{C}\right] = 10^{-3}\sigma^{-3}$. Simulation results along the dashed and dotted lines are presented in Fig. 4.9.

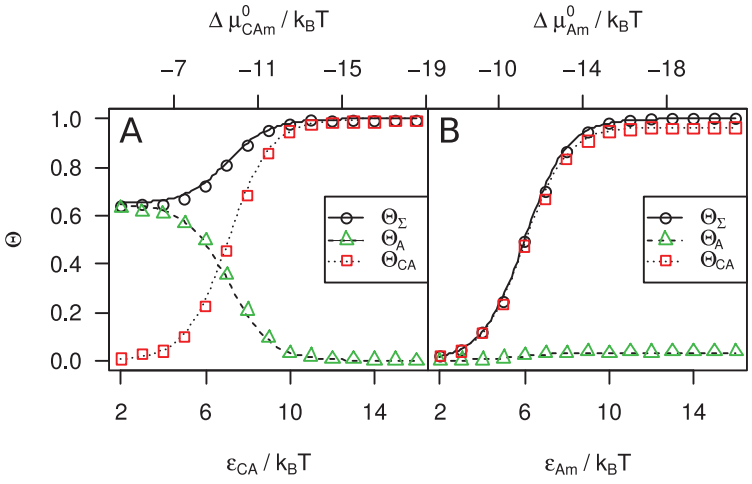


Figure 4.9: The simulated fractional occupancies θ of membrane beads by AP (θA, green triangles), by AP-clathrin complexes (θCA, red squares) and their sum (black circles), for (A) an AP-membrane binding strength $\epsilon_{\mathrm{Am}} = 10k_BT$, and (B) an AP-clathrin binding strength $\epsilon_{\mathrm{CA}} = 10k_BT$, The lines represent the corresponding theoretical curves *i.e.* the cross sections of Fig. 4.8(A through C) along the dotted and dashed lines, respectively.

acidic solution; upon weakening these interactions, the model reproduces the *in vitro* regulation of cage assembly by AP2 [51]. The AP2 complex carries a mid-linker clathrin binding site that is available only when AP2 binds to PtdIns(4,5)P_2 and cargo at a membrane [15]. This feature is mimicked in the AP simulation model by activating the clicking potential of the β_1 site of an AP molecule only when its core bead is clicked to a membrane bead. With the membrane bead representing a patch of membrane including PtdIns(4,5)P_2 and cargo, the β_1 sites of all membrane-bound APs are activated. The proper functioning of this feature is tested by temporarily simplifying the properties of the AP model to the point where the membrane coverage can be solved analytically. In this simplified model, the β_2 bead does not bind clathrin, the number of AP binding sites per clathrin is reduced to one, at the end of one leg, and the interactions among clathrin triskelia are turned off.

A statistical mechanical derivation of the fractional occupations of membrane sites by APs and AP-clathrin complexes is presented in Appendix 4.5.1, and the resulting phase diagrams are depicted in Fig. 4.8. In agreement with the simulations discussed above, Fig. 4.8(A) shows that APs saturate the membrane beads when the standard chemical potential difference associated with AP clicking to the membrane, $\Delta\mu^0_{\mathrm{Am}}$, is lower (more negative) than about $-12k_BT$. Below this value, *i.e.* for the high binding strengths ϵ_{Am} at the top of the plots, with increasing AP-clathrin binding strength, ϵ_{CA}, a transition from predominantly unoccupied APs to APs occupied by triskelia is observed in Fig. 4.8(A) and (B) when the standard chemical potential difference associated with clathrin clicking to a membrane-bound AP reaches $\Delta\mu^0_{\mathrm{CAm}} \approx -12k_BT$. Simulations with the simplified coarse-grained models, employing the acceleration expedients listed above, yield good agreement with the theory, see Fig. 4.9(A). Because this standard chemical potential difference is closely related to that of clathrin binding a (hypothetical) active β_1-bead of an AP in solution, $\Delta\mu^0_{\mathrm{CAm}} \approx \Delta\mu^0_{\mathrm{CA}}$, the onset of AP-assisted clathrin adsorption at the membrane approximately coincides with the onset of AP-clathrin complexation in the bulk if the latter is not prevented by AP adopting the closed conformation in solution. Interestingly, the triangular green region for $\Delta\mu^0_{\mathrm{Am}} \gtrsim -12k_BT$ in Figs 4.8(B) and (C), while Fig. 4.6 and Fig. 4.8(A) show that $\theta\mathrm{A} \approx 0$ for for $\Delta\mu^0_{\mathrm{Am}} \gtrsim -12k_BT$, indicates that the binding of AP to the membrane can be strongly affected by clathrin: for the combination of binding strengths corresponding to the green triangular area, the AP-membrane interaction is too weak to induce AP adsorption, but the (prospective) release of the AP-clathrin binding free energy – which requires AP binding to the membrane – makes the APs click to the membrane nevertheless. This cooperative behaviour is confirmed by the simulation results presented in Fig. 4.9(B). If endocytosis is initiated by AP2 tethering clathrin to the plasma membrane in response to the detection of cargo molecules, then this sequence of events will occur in the current simplified model if the combination of AP-membrane and AP-clathrin interactions corresponds to a point in the red area of Fig. 4.8(B) in the absence of cargo and shifts upwards – increasing the strength of the AP-membrane bond – into the green region of that figure in the presence of cargo.

ubsectionFormation of coated pits

The aggregation of triskelia at the membrane was studied using the full simulation model described in Section 4.2. A chemostat stabilized the bulk concentrations of clathrin and AP at $\big[\mathrm{C}\big] = 10^{-3}\sigma^{-3}$ and $\big[\mathrm{A}\big] = 10^{-2}\sigma^{-3}$, respectively. The interaction strength between the triskelion leg segments was set at $\epsilon = 6k_BT$, which lies slightly below the threshold for spontaneous cage formation in the bulk at the chosen clathrin concentration [47]. Since AP's mid-linker clathrin binding site is not activated in the bulk, cage formation is strictly limited to a small area adjacent to the membrane, as was confirmed by the simulations. The CCP assembly phase diagram in Fig. 4.10 was obtained by varying both the AP-membrane click strength ϵ_{Am} and the AP-clathrin click strength ϵ_{CA} over ranges from 4 to 14 k_BT. Each run lasted 1 s, requiring about 10 days of CPU time. All simulation boxes were visually inspected for the presence of aggregates

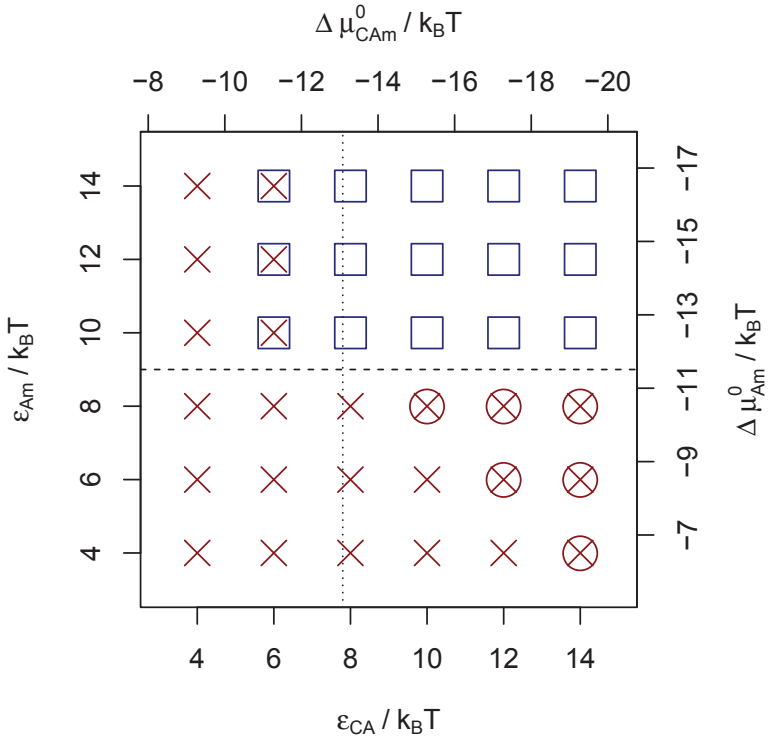


Figure 4.10: Phase diagram of clathrin-coat assembly and disassembly at the membrane, as a function of the clathrin-AP clicking strength ϵ_{CA} and the membrane-AP clicking strength ϵ_{Am}, The clathrin and AP bulk concentrations are maintained at $[\mathrm{C}] = 10^{-3}\sigma^{-3}$ and $[\mathrm{A}] = 10^{-2}\sigma^{-3}$, respectively. Spontaneous assembly within 10^8 time steps (corresponding to 1 s) of at least one clathrin pentamer or hexamer ring from solution is marked with empty blue squares, a lack hereof by red crosses. For simulations initiated with a partial coat attached to the membrane, this coat either appears stable (blue squares and blue squares overlaying red crosses), detaches from the membrane and gradually disintegrates (red circles overlaying red crosses) or directly disintegrates (red crosses). Above the dashed line, more than half the membrane beads is occupied by APs in the absence of clathrin, see Fig. 4.8. To the right of the dotted line, triskelia in solution are bound to at least one AP on average, by β_2 clicking to the ankles.

with at least one closed ring of triskelia, *i.e.* a pentamer or hexamer facet of an incipient coat. Every marker in the figure represents results collected from two independent simulations at the same phase point. Clathrin lattices nucleate and grow at the membrane, see the empty blue squares in Fig. 4.10, when the conditions $\epsilon_{\mathrm{CA}} \gtrsim 7k_BT$ ($\Delta\mu^0_{\mathrm{CAm}} \lesssim -13k_BT$) and $\epsilon_{\mathrm{Am}} \gtrsim 9k_BT$ ($\Delta\mu^0_{\mathrm{Am}} \lesssim -12k_BT$) are both met, while no spontaneous assembly was observed outside this region, see the red crosses. Due to the rarity of coat nucleation events at the membrane, mapping a phase diagram by observing spontaneous assembly within a restricted time scale provides an upper limit on the locations of the phase boundaries. A lower limit was obtained by observing the fate of simulations initiated with coated pit structures (grown under conducive conditions); the employed lattice comprised 24 triskelia, forming four hexagonal and one pentagonal facet, connected by about 20 APs to a mildly curved membrane. For most parameter combinations, seeded and unseeded systems showed similar final states. The biggest difference between the two sets of simulations occurred for the blue squares inscribed by a red cross in Fig. 4.10, marking conditions where coats appeared stable and/or growing but their nucleation from solution was too rare to detect. Red circles denote lattices that partly or fully detached from the membrane and subsequently slowly disintegrated. Interestingly, spontaneous coat assembly (empty blue squares) is only observed in simulations where clathrin binds AP2 in solution (to the right of the dotted line), in agreement with experimental data suggesting that clathrin arrives at the membrane with two attached AP2s [30].

A typical sequence of events during the formation of a coated pit is illustrated in Fig. 4.11 (Multimedia view). Initially, triskelia diffusing into the neighbourhood of the membrane are tethered to the membrane by APs. Next, triskelia diffusing along the membrane meet and form aggregates. Even small diffusing aggregates have repeatedly been observed to merge into larger aggregates, a process that is rare (and more difficult to spot) in solution [50]. The first completed ring of triskelia is always a hexagon, probably because the resulting construct – an annulus with sprawling unpaired legs – is flatter than a similar construct centered around a pentamer. At this time, the membrane underneath the ring is still essentially flat, save for the perpetual thermal undulations. Small mobile aggregates have a clear preference to reside at membrane sections that match their curvature, as this reduces the elongation of the APs linkers and thereby lowers the free energy. This effect becomes more pronounced when imposing artificial undulations on the membrane shape to mimick the potential effect of *e.g.* BAR domains – bending the membrane inward attracts clathrin aggregates – or mutually repulsive cargo molecules – bending the membrane outward repels clathrin aggregates – (data not shown). A similar disposition of coats to curved membranes is observed in experiments [94]. With increasing time, aggregates continue to grow by binding additional APs and triskelia that form additional rings adjacent to previously formed facets. Most lattice patches remain fairly circular; elongated structures arise occasionally, but they often cease to grow and have a tendency to break up into smaller fragments that can grow again. As a patch grows, its extremal segments start to poke into the membrane and consequently the central part of the patch rises. The multiple APs linking the central triskelia to the membrane are stretched and eventually lift the membrane below the center of the patch relative to the membrane at the edge of the patch, thus inducing the membrane to smoothly curve toward the lattice and into the cell interior, see Fig. 4.11(B). With the continued growth of the AP-clathrin assembly, a clathrin coated pit is formed, see Fig. 4.11(C) and (D) and the supplemental movie.

The bending resistance by the membrane puts the lattice patch under considerable stress. An additional set of simulations was run to study the stability of the lattice for various bending rigidities exceeding the $10k_BT$ of giant unilamellar vesicles of fluid-phase double-tailed phosphatidylcholine lipids, as expected for biological membranes of more complex lipid compositions and attached proteins. In simulations seeded with the coated pit structures described before, the

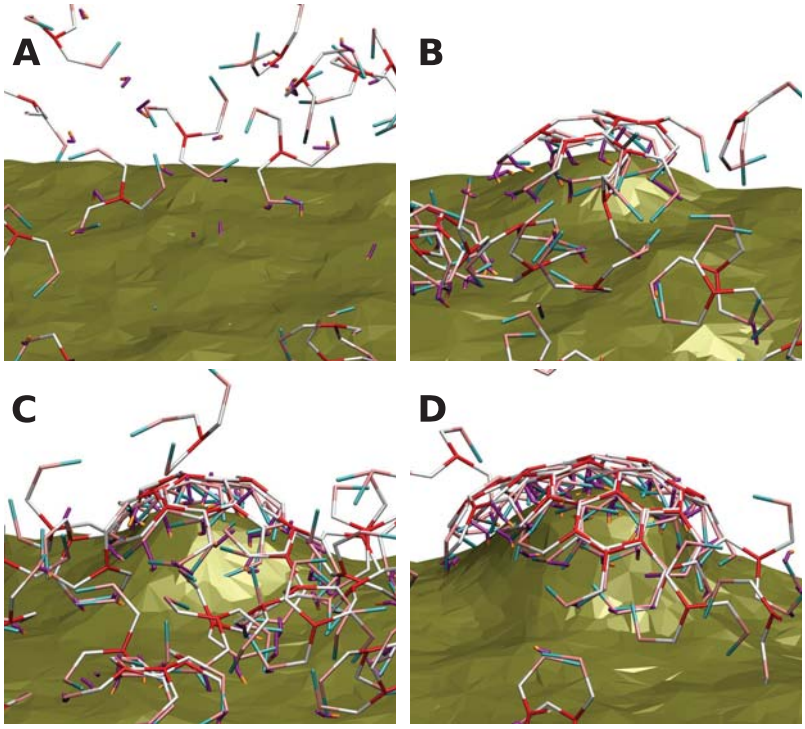


Figure 4.11: Snapshots of clathrin triskelia and adaptor proteins spontaneously assembling into a coat at the membrane surface, at time intervals of 0.1 s. The colour scheme is the same as in Fig. 5.1, with the membrane represented by darkish yellow triangles; the images are taken from a vantage point inside the cell, looking down at the membrane. The interaction strengths are $\epsilon_{\mathrm{CA}} = 10k_BT$ and $\epsilon_{\mathrm{Am}} = 10k_BT$. A movie of another self-assembly event is available (multimedia view). For presentation purposes, the snapshots and movie were generated with the numbers of triskelia and AP fixed at 60 and 100, respectively, rather than using chemostats, as in the phase diagram of Fig. 4.10, to stabilize the chemical potentials of these proteins.

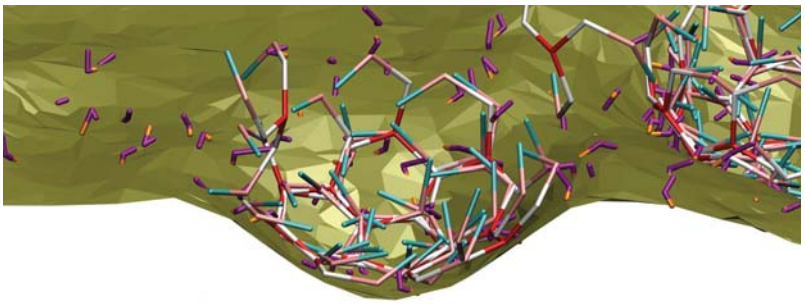


Figure 4.12: Clathrin coats formed by mutant APs that bind clathrin only by the appendage β_2 bead clicking to an ankle. The images are again taken from a vantage point inside the cell, looking down at the membrane. While coats formed in the presence of default APs always bend the membrane inward, see Fig. 4.11, the coats compiled by these partially defunctionalized APs systematically cause the membrane to bulge outward.

lattice was stable and/or continued to grow for $k_\kappa \lesssim 30 k_B T$. At higher values of the bending rigidity, the initially curved membrane largely detached from the lattice and flattened, remaining in contact with the lattice only at its rim; subsequently, these lattices tended to disassemble slowly. On these stiff membranes, the spontaneous assembly of clathrin aggregates ended with the completion of a single hexagonal facet. The unfavourable elongation of the AP tethers connecting the lattice to the unyielding membrane prevented these structures from growing any larger or fusing together. Clearly, this limiting size depends on the pucker angle at the clathrin hub, which we recall to be fixed at 101° in these simulations, and on the properties of the AP linkers. The stresses exerted by the membrane's bending resistance on the tethered lattice patch are also evidenced by the structure of that lattice. When rigid triskelia with a pucker of 101° are induced to self-assemble in solution, either by increasing the leg-leg interaction ϵ or through the help of APs with a permanently active β_1 bead, they form stress-minimized near spherical cages of about 60 triskelia with a homogeneous distribution of the pentagonal facets [47, 51] – the assembly appears to be guided by the 'exclusion of head-to-tail dihedral angle discrepancies' [95, 96] or 'excluded 5566' rule [51]. The clathrin patches growing against the model membrane, however, typically have developed into a hexagonal facet surrounded by a ring of six hexagonal facets before the first pentagonal facet forms at the edge of the patch. Such a configuration of hexagons is not observed for cages grown in solution at the same χ, but results from the stresses imposed by the tethering to a membrane that resists curvature. Because the initial patch is flattened and denuded in pentagons, it is to be expected that the final cage surrounding a vesicle will be larger than the cages grown in solution at the same χ, the twelve pentagonal facets will be unevenly distributed over the surface of the cage, and hence the cage will be less spherical. Cryo-electron tomography images do indeed show that clathrin cages grown during endocytosis deviate more from a symmetrical shape than empty clathrin cages [6].

4.3.2 Mechanism

The role of APs in the aggregation and assembly process is explored by disabling either of their clathrin clicking sites. Since these modified APs can bind one clathrin at most, they can not gather triskelia in solution and consequently will not induce cage formation in the bulk. But they can still collect triskelia at the membrane and thus potentially raise the local clathrin density beyond the critical concentration to spontaneous lattice formation. For APs binding

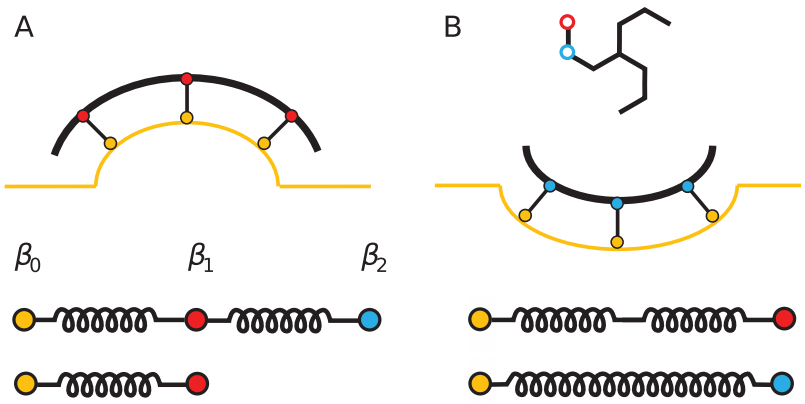


Figure 4.13: Cartoons of cross-sections (top) of clathrin coats (black line) attached to the cytosolic side of membranes (yellow line) and the adaptor proteins (bottom) giving rise to these coats. Right way up coats (A) are formed when the terminal domain of clathrin (red ring) is bound by an AP β_1 site (red circle) that is closely tethered to the AP core (orange), irrespective of the AP β_2 site (blue circle). Upside down coats (B) are formed by mutant APs whose β_1 site binding the TD is connected to the AP core by a long linker and by mutant APs that only bind with their β_2 site to the clathrin ankle (blue ring).

clathrin only through the mid-linker β_1 site clicking to the end of a TD, the coat assembly process still proceeds as described above. Mutant APs binding clathrin only by the appendage β_2 bead clicking to an ankle are also capable of assembling coats adjacent to the membrane, for $\epsilon_{\mathrm{Am}} \gtrsim 10k_BT$ and $\epsilon_{\mathrm{CA}} \gtrsim 10k_BT$. Surprisingly, these coats are systematically oriented upside down: the TDs now point toward the cell interior, rather than to the exterior, and these coats induce *e*vagination of the membrane, rather than the *in*vagination required for endocytosis, as illustrated in Fig. 4.12. The upside down coats incorporate pentagonal facets at an earlier stage than the functional coats, indicating that the former are under less stress. Upside down patches are also observed in the presence of the default APs when conditions permit lattice assembly to commence in solution. When these fragments land upside-down on a membrane, they typically are anchored by a small number of APs, near the point of closest approach, and consequently hardly bend the membrane on the simulation time scale. Fragments landing the right way up become anchored by APs at the lattice edge, but form few membrane connections in the center of the patch and consequently are also inefficient in bending the membrane. We also observed upside-down coats for mutant APs with a functioning β_1 bead, irrespective of the β_2 bead, whose entropic spring constant between β_0 and β_1 has been reduced from $k_{01} = 50k_BT/\sigma^2$ to $1k_BT/\sigma^2$, which approximately triples the linker's average end-to-end length. A visual summary of the AP2-clathrin combinations giving rise to curved membranes is provided in Fig. 4.13.

4.4 Discussion and conclusions

The nucleation of clathrin coats tethered to a membrane by adaptor proteins has been studied by simulations. This was made possible by combining, adapting and tuning coarse grained models for the two proteins and the membrane, using both soft interactions and click potentials, simulated by alternating Brownian Dynamics and Monte Carlo steps.

Simulations and theory show that the coverage of a membrane by non-aggregating AP2s follows Langmuir-like behaviour. The experimental data by Chang *et al.* [92] are also well described

by this theory, see Fig. 4.7 suggesting that aggregation of AP2 plays a minor role. The release of the β_1 clathrin binding site by the AP2 core upon binding of the latter to cargo and membrane can lead to conditions wherein an AP2 core does not bind membrane and cargo in the absence of clathrin but does bind them in the presence of clathrin, see Fig. 4.8. Along the same line, a small change in a living cell – like the arrival of a cargo molecule – may initiate a sequence of events – like the formation of a clathrin coat – by tipping the thermodynamic stability balance.

The simulations show that the total free energy released by the numerous binding processes during coat formation can be sufficient to bend a membrane into a coated pit. The stress exerted by the membrane on the coat is significant, however. Lattice fragments grown against the membrane are flatter than equally sized fragments in solution, and the formation of pentagonal facets is postponed till a later stage in the growth process. With increasing stiffness of the membrane, the growth of a coat may even cease after a single hexagonal ring has formed, because the attachment of additional triskelia results in strong unfavourable steric interactions between these triskelia and the stiff membrane. Of course, a more detailed assessment of whether this process can proceed likewise *in vivo* requires additional quantitative experimental information on the binding free energies of the proteins involved, on the bending rigidity of clathrin, both as monomers in solution and intertwined in cages, and on the bending strengths and bending resistances of the auxiliary proteins and transmembrane cargo proteins involved.

The rigid clathrin model is based on the triskelion configuration and inter-segmental contacts in cages, and therefore potentially biased to the nucleation-and-growth mechanism of coated pit formation observed in the simulations. Besides experimental data supporting this mechanism, the literature also contains studies reporting coated pit formation by the gradual curling up of pre-existing flat hexagonal clathrin lattices [20, 26, 33, 34, 39–42]. The modifications of the clathrin model required to enable this second mechanism, like the inclusion of flexibility and the reduction of inter segmental binding energies to enable frequent clathrin exchange and lattice reshuffles, are a topic of ongoing research.

The modelled second clathrin motif at the appendix of AP2's β-linker proves crucial to induce cage assembly in solution, but the simulations indicate that this motif is not required for coat assembly at the membrane. This suggests two alternative assembly mechanisms: i) in solution the adaptor protein itself brings two triskelia together, while ii) at the membrane an AP binds to a single triskelion and diffusion along the membrane brings the triskelia together. Since the latter mechanism appears to suffice, it is unclear why AP2s are capable of inducing non-functional cage assembly in solution. We speculate that the β_2 site serves to accelerate the assembly of coats at the membrane. omparing the AP models that give rise to normal and upside-down coats, it appears that the competition between steric clathrin-membrane interactions and the elongation of the linker determines the orientation of a coat. For the wild-type short core-β_1 linkers, a triskelion tethered to the membrane by all three legs will experience a significant steric repulsion when oriented the wrong side up, *i.e.* with its hub sticking into the membrane, and hence there is a strong preference to orient the right way up. Upon binding to neighbouring triskelia, the resulting lattice patch will bend the membrane inward. For mutant APs with longer core-β_1 linkers, and for mutant APs that bind clathrin through β_2 only, this steric repulsion between a triple-tethered upside down triskelion and the membrane is much smaller, and consequently the probability of binding in the wrong orientation is much larger. The simulations suggest that small aggregates formed by these upside down triskelia are under less stress than regular lattices – they show a higher curvature at the same patch size, which indicates they spend less free energy on bending the membrane – and thereby further enhances their stability relative to the functional orientation. To the best of our knowledge, upside-down cages – which do not contribute to endocytosis – have never been observed in *in vivo* or *in vitro* experiments. It will be interesting to see whether these upside-down cages can be realised in experiments with mutant

AP2s.

Acknowledgments

We thank Prof. Stefan Luding for stimulating discussions on the membrane model. This work is part of the research programme 'self-assembly of protein coats at membranes' (project nr. 711.012.004) which is financed by the Netherlands Organisation for Scientific Research (NWO).

4.5 Appendix

4.5.1 Adsorption at the membrane

In this appendix, statistical mechanical concepts [97] are used to derive expressions for the numbers of APs bound to triskelia and the numbers of APs and triskelia adsorbed at the membrane. For the reaction of clathrin (C) binding n APs (A),

$$\mathrm{C} + n\mathrm{A} \rightleftharpoons \mathrm{CA}_n, \tag{4.37}$$

the equilibrium constant can be defined as

$$K_{\mathrm{CA}_n} = \frac{\left([\mathrm{CA}_n]_{3\mathrm{D}}/c^0_{3\mathrm{D}}\right)}{\left([\mathrm{C}]_{3\mathrm{D}}/c^0_{3\mathrm{D}}\right)\left([\mathrm{A}]_{3\mathrm{D}}/c^0_{3\mathrm{D}}\right)^n} = e^{-\beta\Delta G^0_{\mathrm{CA}_n}}, \tag{4.38}$$

where $[\mathrm{X}]_{3\mathrm{D}}$ denotes the three-dimensional number density of component X, $c^0_{3\mathrm{D}}$ is a reference concentration typically taken to be 1 molar, and $\Delta G^0_{\mathrm{CA}_n}$ is the standard free energy change of the reaction. In statistical mechanics, assuming ideal solutions, the above equilibrium constant is calculated as [97–99]

$$K_{\mathrm{CA}_n} = \frac{(q_{\mathrm{CA}_n}/V)}{(q_{\mathrm{C}}/V)\,(q_{\mathrm{A}}/V)^n}\left(c^0_{3\mathrm{D}}\right)^n, \tag{4.39}$$

where q_{X} denotes the molecular partition function of a molecule X and V the volume of the system. In terms of the chemical potentials of the reaction participants in ideal solutions at the reference concentration,

$$\mu^0_{\mathrm{X}} = -k_B T \ln \frac{(q_{\mathrm{X}}/V)}{c^0_{3\mathrm{D}}}, \tag{4.40}$$

the standard free energy change of the reaction reads as

$$\Delta G^0_{\mathrm{CA}_n} = \mu^0_{\mathrm{CA}_n} - \mu^0_{\mathrm{C}} - n\mu^0_{\mathrm{A}} = \Delta\mu^0_{\mathrm{CA}_n}, \tag{4.41}$$

with $\Delta\mu^0_{\mathrm{CA}_n}$ the standard chemical potential change of the reaction.

The partition function of a rigid triskelion in solution follows from the integral of the Boltzmann factor over all particle positions $\mathbf{r}_{\mathrm{C}}$ and orientations $\boldsymbol{\varphi}_{\mathrm{C}}$,

$$q_{\mathrm{C}} = \frac{1}{\Delta_{\mathrm{C}}}\iint e^{-\beta\Phi} d\mathbf{r}_{\mathrm{C}} d\boldsymbol{\varphi}_{\mathrm{C}} \approx \frac{1}{\Delta_{\mathrm{C}}} 8\pi^2 V e^{-\beta\Phi_{\mathrm{C}}}, \tag{4.42}$$

where Δ_{C} denotes the elementary volume of the particle (a combination of elementary constants, the temperature, the mass and the inertia tensor of the triskelion; the exact value is of no consequence as it will drop out in all major results derived below), with Φ the potential energy, Φ_{C} the averaged solvation free energy of clathrin and where use was made of the Euler angles with $\varphi 1 \in [0, 2\pi)$, $\varphi 2 \in [0, \pi)$, $\varphi 3 \in [0, 2\pi)$, and $d\boldsymbol{\varphi}_{\mathrm{C}} = \sin\varphi 2 d\varphi 1 d\varphi 2 d\varphi 3$.

The AP model described in the main text consists of three beads connected by non-linear springs. The partition function of this molecule reads as

$$q_{\mathrm{A}} = \frac{1}{\Delta_{\mathrm{a}}^3} \iiint e^{-\beta\Phi} d\mathbf{r}_0 d\mathbf{r}_1 d\mathbf{r}_2 \tag{4.43}$$

$$\approx \frac{1}{\Delta_{\mathrm{a}}^3} e^{-\beta\Phi_{\mathrm{A}}} V q_{\mathrm{s}}01 q_{\mathrm{s}}12, \tag{4.44}$$

with Δ_{a} the elementary volume per bead (for simplicitiy taken to be identical for all beads; again, this factor will cancel out in the final expressions), Φ the potential energy as a function of the particle positions $\mathbf{r}$, and Φ_{A} the average solvation free energy of AP. Integrating over $\mathbf{r}_2$ at fixed $\mathbf{r}_1$ yields a factor

$$q_{\mathrm{s}}12 = \int e^{-\beta\psi_{12}(x12)} d\mathbf{r}_{12} \approx \left(\frac{2\pi k_B T}{k_{12}}\right)^{3/2}, \tag{4.45}$$

where the linker potential ψ_{12} has been approximated by a Hookean spring with spring constant k_{12}. One obtains a similar factor when integrating over $\mathbf{r}_1$ at fixed $\mathbf{r}_0$, and the final integration over $\mathbf{r}_0$ yields the volume V, collectively arriving at Eq. (4.44).

For a dissolved protein complex of a triskelion bound to an AP by the β_2 bead, the partition function reads as

$$q_{\mathrm{CA}} = \frac{1}{\Delta_{\mathrm{C}}\Delta_{\mathrm{a}}^3} \int \cdots \int e^{-\beta\Phi} d\mathbf{r}_{\mathrm{C}} d\boldsymbol{\varphi}_{\mathrm{C}} d\mathbf{r}_0 d\mathbf{r}_1 d\mathbf{r}_2 \tag{4.46}$$

$$\approx \frac{q_{\mathrm{C}} q_{\mathrm{A}}}{V} \mu \frac{4}{3}\pi\rho_{\mathrm{CA}}^3 e^{\beta\epsilon_{\mathrm{CA}}}, \tag{4.47}$$

where ρ_{CA} and ϵ_{CA} denote the radius and strength, respectively, of the click interaction between the β_2 bead and any of the $\mu = 3$ matching sites on the triskelion. This remarkably compact result is a consequence of the click potential, which reduces the translational freedom of one of the partners from V to the volume of the click interaction and contributes a Boltzmann factor related to the clicking strength [51]. This insight is readily extended to triskelia decorated with multiple APs; insertion in Eq. (4.39) then yields the equilibrium constants of these clathrin-AP complexes.

If the membrane beads, in their role as binding sites for APs, are both free to diffuse over the entire membrane area A and sufficiently thinly distributed so as to ignore interactions, their partition function will read as

$$q_{\mathrm{m}} = \frac{1}{\Delta_{\mathrm{m}}} \int e^{-\beta\Phi} d\mathbf{r}_{\mathrm{m}} \approx \frac{1}{\Delta'_{\mathrm{m}}} A e^{-\beta\Phi_{\mathrm{m}}}, \tag{4.48}$$

with Φ_{m} the average free energy of a bead, and where the elementary area, Δ'_{m}, represents the ratio of the elementary volume for a membrane bead, Δ_{m}, to the mean transverse displacement of a bead confined to a membrane.

The partition function of the complex of an AP molecule clicked by its β_0 bead to a membrane bead is given by

$$q_{\mathrm{Am}} = \frac{1}{\Delta_{\mathrm{a}}^3\Delta_{\mathrm{m}}} \iiiint e^{-\beta\Phi} d\mathbf{r}_0 d\mathbf{r}_1 d\mathbf{r}_2 d\mathbf{r}_{\mathrm{m}}. \tag{4.49}$$

Upon ignoring the non-bonding AP-membrane interactions, *i.e.* the potential that prevents the β_1 and β_2 beads from crossing the membrane, one readily obtains the approximation, denoted by a tilde,

$$\tilde{q}_{\mathrm{Am}} \approx q_{\mathrm{m}} \frac{q_{\mathrm{A}}}{V} \frac{4}{3}\pi\rho_{\mathrm{Am}}^3 e^{\beta\epsilon_{\mathrm{Am}}}, \tag{4.50}$$

where ρ_{Am} and ϵ_{Am} denote the radius and strength, respectively, of the click interaction between the membrane bead and the AP's core. By analogy with the above equilibrium constant in bulk solution, the equilibrium constant of the reaction

$$\mathrm{A} + \mathrm{m} \rightleftharpoons \mathrm{Am} \tag{4.51}$$

can be defined as

$$K_{\mathrm{Am}} = \frac{\left([\mathrm{Am}]_{2\mathrm{D}}/c^0_{2\mathrm{D}}\right)}{\left([\mathrm{A}]_{3\mathrm{D}}/c^0_{3\mathrm{D}}\right)\left([\mathrm{m}]_{2\mathrm{D}}/c^0_{2\mathrm{D}}\right)} \tag{4.52}$$

$$= \frac{(q_{\mathrm{Am}}/A)}{(q_{\mathrm{A}}/V)\,(q_{\mathrm{m}}/A)}c^0_{3\mathrm{D}}, \tag{4.53}$$

where the membrane beads and their complexes are bound to the two-dimensional membrane surface, and $c^0_{2\mathrm{D}}$ denotes a two-dimensional reference concentration. By inserting the preceeding expressions for the molecular partition functions, one arrives at

$$\frac{[\mathrm{Am}]_{2\mathrm{D}}}{[\mathrm{m}]_{2\mathrm{D}}} = \frac{[\mathrm{A}]_{3\mathrm{D}}}{c^0_{3\mathrm{D}}}e^{-\beta\Delta G^0_{\mathrm{Am}}} \tag{4.54}$$

with the free energy change accompanying the reaction given by

$$\begin{aligned}\Delta G^0_{\mathrm{Am}} &= \mu^0_{\mathrm{Am}} - \mu^0_{\mathrm{m}} - \mu^0_{\mathrm{A}} = \Delta\mu^0_{\mathrm{Am}} \\ &= -k_B T \ln\left(\frac{4}{3}\pi\rho^3_{\mathrm{Am}}e^{\beta\epsilon_{\mathrm{Am}}}c^0_{3\mathrm{D}}\right),\end{aligned} \tag{4.55}$$

where the reference chemical potentials of membrane-bound particles and complexes follow from

$$\mu^0_{\mathrm{X}} = -k_B T \ln\frac{(q_{\mathrm{X}}/A)}{c^0_{2\mathrm{D}}}. \tag{4.56}$$

Upon noting that $[\mathrm{Am}]_{2\mathrm{D}}$ and $[\mathrm{m}]_{2\mathrm{D}}$ are proportional to the fractions of occupied and unoccupied membrane beads, $\theta\mathrm{A}$ and $1-\theta\mathrm{A}$ respectively, one recognizes in Eq. (4.54) the Langmuir expression of Eq. (4.35).

Returning to the partition function of Eq. (4.49), the non-bonded interactions of the AP's β_1 and β_2 beads with a locally nearly flat membrane, taken to be the plane $z = 0$, can be approximated by the conditions $z_1 \geq 0$ and $z_2 \geq 0$. Evaluating the partition function part by part, the integral over $\mathbf{r}_2$ at given $\mathbf{r}_1$ then yields

$$\hat{q}_{s,2}(z_1) = \frac{1}{2}\left(\frac{2\pi k_B T}{k_{12}}\right)^{3/2}\left[1 + \mathrm{erf}\left(\sqrt{\tfrac{1}{2}\beta k_{12}}z_1\right)\right], \tag{4.57}$$

where erf denotes the error function. Assuming $z_0 \approx 0$ and $k_{01} \approx k_{12}$, the integral over $\mathbf{r}_1$ at constant $\mathbf{r}_0$ becomes

$$\hat{q}_{s,12} = \frac{1}{4}\left(\frac{2\pi k_B T}{k_{12}}\right)^3\left(1 + \int_0^\infty \mathrm{erf}(s)\; e^{-s^2}ds\right), \tag{4.58}$$

which is readily solved by noting that the Gaussian is the derivative of the error function. The integral over $\mathbf{r}_0$ proceeds as before, and by collecting factors we arrive at the remarkably simple result

$$\hat{q}_{\mathrm{Am}} = \frac{3}{8}\tilde{q}_{\mathrm{Am}}, \tag{4.59}$$

in excellent agreement with the simulation results in Fig. 4.6.

In the AP2 complex and in the simulation model, the mid-linker binding site for clathrin, β_1 becomes actived when the core, β_0, binds to the membrane. This feature of the model is tested in Section 4.3.1, using an analytically solvable simplified model system with only one clathrin binding site per AP, *i.e.* β_1, and only one matching binding site per triskelion. The partition function of a triskelion bound to an AP which in turn is bound to a membrane bead can be constructed by combining the previous results in this appendix. Partition functions of molecules are typically calculated by integrating the internal and rotational motions relative to the molecule's center of mass, followed by an integration over the position of that reference point [97]. We note that the partition function is independent of the location of the reference point relative to the molecule. In the case under study, it proves advantageous to calculate the partition function of the triskelion with respect to its single AP-binding site. The partition function of an clathrin-AP complex bound to a membrane bead, when ignoring the repulsive non-penetration potential of the membrane, then becomes

$$\tilde{q}_{\mathrm{CAm}} = \tilde{q}_{\mathrm{Am}} \frac{q_{\mathrm{C}}}{V} \frac{4}{3} \pi \rho_{\mathrm{CA}}^3 e^{\beta \epsilon_{\mathrm{CA}}}. \tag{4.60}$$

The standard free energy change accompanying the reaction

$$\mathrm{C} + \mathrm{Am} \rightleftharpoons \mathrm{CAm} \tag{4.61}$$

then reads as

$$\begin{aligned} \Delta G^0_{\mathrm{CAm}} &= \mu^0_{\mathrm{CAm}} - \mu^0_{\mathrm{Am}} - \mu^0_{\mathrm{C}} = \Delta \mu^0_{\mathrm{CAm}} \\ &= -k_B T \ln \left(\frac{4}{3} \pi \rho_{\mathrm{CA}}^3 e^{\beta \epsilon_{\mathrm{CA}}} c^0_{\mathrm{3D}} \right). \end{aligned} \tag{4.62}$$

The equilibrium of the two coupled reactions, Eqs. (4.51) and (4.61), follows by solving the fractional occupancies of membrane beads by APs and AP-clathrin complexes, θCA and θA respectively, from the coupled equations

$$\frac{[\mathrm{CAm}]_{\mathrm{2D}}}{[\mathrm{Am}]_{\mathrm{2D}}} = \frac{\theta \mathrm{CA}}{\theta \mathrm{A}} = \frac{[\mathrm{C}]_{\mathrm{3D}}}{c^0_{\mathrm{3D}}} e^{-\beta \Delta G^0_{\mathrm{CAm}}}, \tag{4.63}$$

$$\frac{[\mathrm{Am}]_{\mathrm{2D}}}{[\mathrm{m}]_{\mathrm{2D}}} = \frac{\theta \mathrm{A}}{1 - \theta \mathrm{A} - \theta \mathrm{CA}} = \frac{[\mathrm{A}]_{\mathrm{3D}}}{c^0_{\mathrm{3D}}} e^{-\beta \Delta G^0_{\mathrm{Am}}}, \tag{4.64}$$

where in the second line it was used that the fractions of empty and occupied membrane beads add up to unity.

References

[1] F. M. Brodsky, *Annual review of cell and developmental biology* **2012**, *28*, 309.

[2] T. Kirchhausen, D. Owen, S. C. Harrison, *Cold Spring Harbor perspectives in biology* **2014**, *6*, a016725.

[3] M. S. Robinson, *Traffic* **2015**, *16*, 1210–1238.

[4] E. M. Schmid, H. T. McMahon, *Nature* **2007**, *448*, 883.

[5] S. Zaremba, J. H. Keen, *The Journal of cell biology* **1983**, 1339–1347.

[6] Y. Cheng, W. Boll, T. Kirchhausen, S. C. Harrison, T. Walz, *Journal of molecular biology* **2007**, *365*, 892–899.

[7] R. Crowther, B. Pearse, *The Journal of cell biology* **1981**, *91*, 790–797.

[8] A. Fotin, Y. Cheng, P. Sliz, N. Grigorieff, S. C. Harrison, T. Kirchhausen, T. Walz, *Nature* **2004**, *432*, 573–579.

[9] S. K. Lemmon, L. M. Traub, *Traffic* **2012**, *13*, 511–519.

[10] M. P. Czech, *Cell* **2000**, *100*, 603–606.

[11] D. J. Owen, B. M. Collins, P. R. Evans, *Annual Review of Cell and Developmental Biology* **2004**, *20*, 153–191.

[12] L. M. Traub, *Nat. Rev. Mol. Cell Biol.* **2009**, *10*, 583–596.

[13] B. J. Canagarajah, X. Ren, J. S. Bonifacino, J. H. Hurley, *Protein Science* **2013**, *22*, 517–529.

[14] L. P. Jackson, B. T. Kelly, A. J. McCoy, T. Gaffry, L. C. James, B. M. Collins, S. Höning, P. R. Evans, D. J. Owen, *Cell* **2010**, *141*, 1220–1229.

[15] B. Kelly, S. Graham, N Liska, *Science* **2014**, *345*, 459.

[16] C Knuehl, C. Y. Chen, V Manalo, P. K. Hwang, N Ota, F. M. Brodsky, *Traffic* **2006**, *7*, 1688–1700.

[17] H. S. Moskowitz, C. T. Yokoyama, T. A. Ryan, *Mol. Biol. Cell* **2005**, *16*, 1769–1776.

[18] L. Maldonado-Báez, B. Wendland, *Trends in cell biology* **2006**, *16*, 505–513.

[19] H. T. McMahon, E Boucrot, *Nat. Rev. Mol. Cell Biol.* **2011**, *12*, 517–533.

[20] J. Heuser, L. Evans, *J. Cell Biol.* **1980**, *84*, 560–583.

[21] J. E. Heuser, J. H. Keen, L. M. Amende, R. E. Rippoldt, K Prased, *J. Cell Biol.* **1987**, *105*, 1999–2009.

[22] P. N. Dannhauser, M Platen, H Böning, H Ungewickell, I. A. T. Schaap, E. J. Ungewickell, *Traffic* **2015**, *16*, 519–533.

[23] I Gaidarov, F Santini, R. A. Warren, J. H. Keen, *Nat. Cell Biol.* **1999**, *1*, 1–7.

[24] X Wu, X Zhao, L Baylor, S Kaushal, E Eisenberg, L. E. Greene, *J. Cell Biol.* **2001**, *155*, 291–300.

[25] D Loerke, M Wienisch, O Kochubey, J Klingaud, *Traffic* **2006**, *6*, 918–929.

[26] O Avinoam, M Schorb, C Beese, J. A. G. Briggs, M Kaksonen, *Science* **2015**, *348*, 1369–1372.

[27] C. J. Merrifield, D Perrais, D Zenisek, *Cell* **2005**, *121*, 593–606.

[28] D Loerke, M Mettlen, S. L. Schmid, G Danuser, *Traffic* **2011**, *12*, 815–825.

[29] M. J. Taylor, D Perrais, C. J. Merrifield, *PLoS Biol.* **2011**, *9*, e1000604.

[30] E Cocucci, F Aguet, S Boulant, T Kirchhausen, *Cell* **2012**, *150*, 495–507.

[31] D. Loerke, M. Mettlen, D. Yarar, K. Jaqaman, H. Jaqaman, G. Danuser, S. L. Schmid, *PLoS Biol* **2009**, *7*, e1000057.

[32] S. H. Hong, C. L. Cortesio, D. G. Drubin, *Cell Rep.* **2015**, *12*, 2121–2130.

[33] Y Li, L Shang, G. U. Nienhaus, *J. Cell Biol.* **2001**, *155*, 291–300.

[34] J. P. Ferguson, N. M. Willy, S. P. Heidotting, S. D. Huber, M. J. Webber, C Kural, *J. Cell Biol.* **2016**, *214*, 347–358.

[35] M. G. J. Ford, I. G. Mills, B. J. Peter, Y Vallis, G. J. K. Praefcke, P. R. Evans, H. T. McMahon, *Nature* **2002**, *419*, 361–366.

[36] W. M. Henne, H. M. Kent, M. G. Ford, B. G. Hegde, O. Daumke, P. J. G. Butler, R. Mittal, R. Langen, P. R. Evans, H. T. McMahon, *Structure* **2007**, *15*, 839–852.

[37] P. N. Dannhauser, E. J. Ungewickell, *Nat. Cell Biol.* **2012**, *14*, 634–639.

[38] J. C. Stachowiak, F. M. Brodsky, E. A. Miller, *Nat. Cell Biol.* **2013**, *15*, 1019–1027.

[39] M Ehrlich, W Boll, A van Oijen, R Hariharan, K Chandran, M. L. Nibert, T Kirchhausen, *Cell* **2004**, *118*, 591–605.

[40] S. Saffarian, E. Cocucci, T. Kirchhausen, *PLoS Biol* **2009**, *7*, e1000191.

[41] W. Kukulski, M. Schorb, M. Kaksonen, J. A. G. Briggs, *Cell* **2012**, *150*, 508–520.

[42] D. J. M. J. Grove and, A. E. Knight, S. T. Wavre-Shapton, T. Suna, E. D. Protonotarios, L. D. Griffin, J. Lippincott-Schwartze, M. Marsh, *Mol. Biol. Cell* **2014**, *25*, 3581–3594.

[43] M. Lampe, S. Vassilopoulos, C. Merrifield, *Journal of structural biology* **2016**, *196*, 48–56.

[44] R. Matthews, C. N. Likos, *Soft Matter* **2013**, *9*, 5794–5806.

[45] N. Cordella, T. J. Lampo, N. Melosh, A. J. Spakowitz, *Soft matter* **2015**, *11*, 439–448.

[46] J. J. VanDersarl, S. Mehraeen, A. P. Schoen, S. C. Heilshorn, A. J. Spakowitz, N. A. Melosh, *Soft matter* **2014**, *10*, 6219–6227.

[47] W. K. den Otter, M. R. Renes, W. J. Briels, *J. Phys. Condens. Matt.* **2010**, *22*, 104103.

[48] W. K. den Otter, M. R. Renes, W. J. Briels, *Biophys. J.* **2010**, *99*, 1231–1238.

[49] W. K. den Otter, W. J. Briels, *Traffic* **2011**, *12*, 1407–1416.

[50] I. M. Ilie, W. K. den Otter, W. J. Briels, *J. Chem. Phys.* **2014**, *141*, 065101.

[51] M Giani, W. K. den Otter, W. J. Briels, *Biophys. J.* **2016**, *111*, 222–235.

[52] L. Moshkanbaryans, J. Xue, J. R. Wark, P. J. Robinson, M. E. Graham, *Protein Sci.* **2016**, *11*, e0162050.

[53] B Goud, C Huet, D Louvard, *J. Cell Biol.* **1985**, *100*, 521–527.

[54] S. J. Doxsey, F. M. Brodsky, G. S. Blank, A Helenius, *Cell* **1987**, *50*, 453–463.

[55] F. J. López-Murcia, S. J. Royle, A. Llobet, *Journal of Neuroscience* **2014**, *34*, 8618–8629.

[56] K. A. Beck, J. H. Keen, *J. Biol. Chem.* **1991**, *266*, 4437–4441.

[57] W Hao, Z Luo, L Zheng, K Prasad, E. M. Lafer, *Journal of Biological Chemistry* **1999**, *274*, 22785–22794.

[58] M Skruzny, A Desfosses, S Prinz, S. O. Dodonova, A Gieras, C Uetrecht, A. J. Jakobi, M Abella, W. J. H. Hagen, J Schulz, R Meijers, V Rybin, J. A. G. Briggs, C Sachse, M Kaksonen, *Developmental cell* **2015**, *33*, 150–162.

[59] J. H. Keen, K. A. Beck, T Kirchhausen, T Jarrett, *J. Biol. Chem.* **1991**, *266*, 7950–7956.

[60] M. A. Cousin, P. J. Robinson, *Trends in neurosciences* **2001**, *24*, 659–665.

[61] G Hollopeter, J. J. Lange, Y Zhang, T. N. Vu, M Gu, M Ailion, E. J. Lambie, B. D. Slaughter, J. R. Unruh, L Florens, E. M. Jorgensen, *Elife* **2014**, *3*, e03648.

[62] L Ma, P. K. Umasankar, A. G. Wrobel, A Lymar, A. J. McCoy, S. S. Holkar, A Jha, T Pradhan-Sundd, S. C. Watkins, D. J. Owen, L. M. Traub, *Developmental cell* **2016**, *37*, 428–443.

[63] L von Kleist, W Stahlschmidt, H Bulut, K Gromova, D Puchkov, M. J. Robertson, K. A. MacGregor, N Tomilin, A Pechstein, N Chau, M Chircop, J Sakoff, J. P. von Kries, W Saenger, H.-G. Krausslich, O Shupliakov, P. J. Robinson, A McCluskey, V Haucke, *Cell* **2011**, *146*, 471–484.

[64] J. A. L. J C Dawson, L. M. Machesky, *Trends Cell Biol.* **2006**, *16*, 493–498.

[65] H. T. McMahon, E Boucrot, *J. Cell Sci.* **2015**, *128*, 1065–1070.

[66] S. E. Miller, S Mathiasen, N. A. Bright, F Pierre, B. T. Kelly, N Kladt, A Schauss, C. J. Merrifield, D Stamou, S Honing, D. J. Owen, *Developmental cell* **2015**, *33*, 163–175.

[67] J. C. Stachowiak, E. M. Schmid, C. J. Ryan, H. S. Ann, D. Y. Sasaki, M. B. Sherman, P. L. Geissler, D. A. Fletcher, C. C. Hayden, *Nat. Cell Biol.* **2012**, *14*, 944–949.

[68] A. Young in Seminars in cell & developmental biology, *Vol. 18*, Elsevier, **2007**, pp. 448–458.

[69] M. P. Allen, D. J. Tildesley, *Computer Simulation of Liquids*, Clarendon Press, New York, NY, USA, **1989**.

[70] C. W. Gardiner, *Stochastic Methods: A Handbook for the Natural and Social Sciences*, Springer, Berlin, Germany, **2009**.

[71] I. M. Ilie, W. J. Briels, W. K. den Otter, *J. Chem. Phys.* **2015**, *142*, 114103.

[72] J Garcia de la Torre, G del Rio Echenique, A Ortega, *J. Phys. Chem. B* **2007**, *111*, 955–961.

[73] D. J. Evans, S. Murad, *Molecular physics* **1977**, *34*, 327–331.

[74] D. Boal, *Mechanics of the Cell*, Cambridge University Press, Cambridge, U.K., **2002**.

[75] K. Kremer, G. S. Grest, *The Journal of Chemical Physics* **1990**, *92*, 5057–5086.

[76] B. M. Collins, A. J. McCoy, H. M. Kent, P. R. Evans, D. J. Owen, *Cell* **2002**, *109*, 523–535.

[77] D. P. Landau, K. Binder, *A Guide to Monte Carlo Simulations in Statistical Physics*, Cambridge University Press, Cambridge, U.K., **2000**.

[78] D. Frenkel, B. Smit, *Understanding Molecular Simulation*, 2nd, Academic Press Inc., Orlando, FL, U.S.A., **2001**.

[79] D. Nelson, T. Piran, S. Weinberg, *Statistical Mechanics of Membranes and Surfaces*, 2nd ed., World Scientific Publishing, Singapore, **2004**.

[80] D. H. Boal, M. Rao, *Physical Review A* **1992**, *46*, 3037.

[81] G Gompper, D. M. Kroll, *J. Phys. Condens. Matt.* **1997**, *9*, 8795–8834.

[82] D Espriu, *Phys. Lett. B* **1987**, *194*, 271–276.

[83] S.-J. Zhao, J. T. Kindt, *EPL* **2005**, *69*, 839–845.

[84] G. Gompper, D. M. Kroll, *J. Phys. I (France)* **1996**, *6*, 1305–1320.

[85] S. A. Shkulipa, W. K. den Otter, W. J. Briels, *Phys. Rev. Lett.* **2006**, *96*, 178302.

[86] W Rawicz, K. C. Olbrich, T McIntosh, D Needham, E Evans, *Biophys. J.* **2000**, *79*, 328–339.

[87] R Goetz, G Gompper, R Lipowsky, *Phys. Rev. Lett.* **1999**, *82*, 221–224.

[88] W. K. den Otter, *J. Chem. Phys.* **2005**, *123*, 214906.

[89] E. Evans, W. Rawicz, *Phys. Rev. Lett.* **1990**, *64*, 20942097.

[90] F Brochard, J. F. Lennon, *J. Phys. (France)* **1975**, *36*, 1035–1047.

[91] A Imparato, *J. Chem. Phys.* **2006**, *124*, 154714.

[92] M. P. Chang, W. G. Mallet, K. E. Mostov, F. M. Brodsky, *The EMBO journal* **1993**, *12*, 2169.

[93] P. Atkins, J. de Paula, *Physical Chemistry*, 9th ed., Oxford University Press, Oxford, U.K., **2010**.

[94] T. J. Pucadyil, S. S. Holkar, *Mol. Biol. Cell* **2016**, *27*, 3145–3163.

[95] S. Schein, M. Sands-Kidner, *Biophysical journal* **2008**, *94*, 958–976.

[96] S. Schein, M. Sands-Kidner, T. Friedrich, *Biophysical journal* **2008**, *94*, 938–957.

[97] D. A. McQuarrie, *Statistical Mechanics*, Harper & Row Publishers, New York, NY, USA, **1976**.

[98] P. C. Nelson, *Biological Physics*, updated first edition, W.H. Freeman and Company, New York, NY, U.S.A., **2008**.

[99] R. Phillips, J. Kondev, J. Theriot, *Physical Biology of the Cell*, Garlan Science, New York, NY, U.S.A., **2009**.

Chapter 5

Simulations of clathrin mediated endocytosis reveal a midway activation barrier

The self-assembly process of clathrin coated membrane buds during clathrin mediated endocytosis (CME) is explored by coarse-grained simulations of clathrin triskelia, adaptor proteins and a flexible network membrane. Preceding simulations indicated that the spontaneous formation of a clathrin coated pit becomes very slow, if not arrested, when the clathrin coat has grown into an approximately hemi-spherical dome. The free energy calculations presented here, using the potential of mean constraint force (PMCF), reveal the presence of an activation barrier of about $30\,k_BT$*, followed by a spontaneous maturation into a nearly completed CCV. This suggests that surmounting this barrier, which may act as a decision event in CME, requires the assistance of auxilliary proteins.*[1]

5.1 Introduction

Clathrin-mediated endocytosis is a critical process involved in the cellular uptake of nutrients, growth factors and receptors, that relies on the regulated assembly of clathrin-coated vesicles (CCVs) at the plasmatic surface of the cell membrane [1–3]. The fundamental assembly unit, and namesake of clathrin-coated vesicles, is the clathrin triskelion, a three-legged protein with the ability to self-assemble *in vitro*, without assistance, into polyhedrical cages that closely resemble the coats surrounding the *in vivo* vesicles [4]. After clathrin, the adaptor protein 2 (AP2) is the second most abundant protein in clathrin coats formed during endocysis [5]. AP2s are tetrameric complexes that carry binding motifs for clathrin, cargo molecules and lipids specific to the plasma membrane, with the fundamental function of tethering clathrin triskelia to the plasma membrane following the external arrival of cargo. The formation of a CCV begins with a nucleation event, yielding an aggregate that grows at a fairly constant rate [2] through the recruitment of individual clathrin triskelia and adaptor proteins from the cytosol. This assembly deforms the underlying membrane as it grows in size, creating an invagination called a clathrin coated pit (CCP) and eventually the coat fully envelopes the cargo [6, 7]. Finally, the CCV is released by cutting the thin neck still connecting the membrane inside the coat to the plasma membrane, the coat is disassembled and the coat units are recycled in the next endocytic event. The time elapsed between nucleation and release of a CCV ranges from 30 to 120 seconds [8].

[1] This chapter is being prepared for publication in a journal.

The assembly dynamics of CCVs continues to attract the interest of research groups using a wide variety of experimental techniques. Cryo-electron microscopy (cryo-EM) yields detailed images of the 3D structure of CCPs and CCVs, as well as stills of flat hexagonal lattices [9, 10]. Labelling selected proteins with fluorescent labels has unleashed the ability to follow the dynamics of proteins during endocytosis, using a steadily growing collection of optical techniques [11, 12]. Combining single particle tracking with photoactivation localization microscopy (PALM) captures the endocytosis dynamics at near-molecular resolution [13]. Fluorescence intensity traces of hundreds of simultaneous clathrin aggregates in a cell provide access to their formation and dissolution rates [14]. The exchange of triskelia between CCPs and the cytosol was revealed by fluorescence recovery after photobleaching (FRAP) [8, 15]. Combining cryo-EM with live-cell imaging relates high resolution structural information to specific stages in the evolution of CCPs [16]. Total internal reflection fluorescence microscopy (TIR-FM) has been used to determine the sequence and numbers of proteins arriving at, and leaving from, the membrane over the course of many endocytic events. Interestingly, these experiments detected a large number of short-lived clathrin structures, referred to as 'abortive' pits, whose lifetime is less than 20 seconds, well below that of 'productive' CCVs [8, 17, 18]. Despite the wealth of experimental information obtained by a range of supplementary techniques, a number of fundamental questions on CCP dynamics are still hotly debated. These include the mechanical contribution of clathrin toward bending the membrane,[19–22] and whether CCPs are formed by nucleation and growth or by gradually increasing the curvature of a pre-existing planar lattice[9, 13, 14, 16, 17, 23–25].

Coarse-grained simulation models, reducing the protein models to their bare essentials based on interpretations of the experimentally available data on the proteins's structures and their interactions, offer a potent method to improve our current understanding of CME. If the model does not deliver, this suggests that an essential component was missing; simulations are also known to find unexpected results which in hindsight appear obvious [26]. Matthews and Likos simulated clathrin as a triskelion composed of 13 bead particles endowed with interaction patches on their surfaces, showing that these triskelia can self-assemble into cages and that they are able to create deep membrane invaginations [27]. Spakowitz and collaborators modeled clathrin as a particle that forms harmonic bonds with three neighbors, to study the mechanical properties of lattices against a flat membrane and the fluidization of these lattices when the membrane is nano-indented [28, 29]. In a recent study, we combined our coarse-grained clathrin [30] and AP models [31] with an elastic membrane model, to simulate their combined aggregation process by Brownian Dynamics [32]. The presence of cargo and phosphatidylinositol 4,5-bisphosphate [PtdIns(4,5)P_2] lipids specific to the plasma membrane, both crucial to the binding of AP2 to the membrane and the release by AP2 of a clathrin motif, were mimicked by allowing AP to bind to a fraction of the membrane beads Other proteins that potentially playing a role in endocytosis, like membrane bending proteins, were not included to keep the model as basic as possible, in line with experiments indicating that clathrin plus adaptor protein suffice to creat CCPs [21]. The simulations spontaneously follow the sequence of events described above, up to midway the vagination process. First, individual triskelia are tethered to the membrane by adaptor proteins and aggregate to form a small lattice. The latter grows by incorporating additional triskelia and APs from the solution, and progressively deforms the underlying membrane as it increases in size. Surprisingly, the growth process discontinues when the clahtrin coat is about to reach a hemi-spherical configuration; not a single CCP was observed to grow beyond this point.

Here we will explore why the simulated CCPs do not grow to completion. At the midway point, the process has either become very slow, and merely appears arrested, or the spontaneous growth has reached a local free energy minimum. In the latter case, the spontaneous growth may continue once the system has surmounted an activation barrier, which constitutes a rare event on the simulation time scale. To distinguish between the two cases, the free energy profile of a coated

bud will be calculated as a function of the bud's area. Several methods exist to calculate free energies along a reaction coordinate, including umbrella sampling, thermodynamic integration, metadynamics and the potential of mean constraint force (PMCF) [33–36]. The latter method is used here, *i.e.* we perform simulations with fixed values of the bud area and recover the free energy profile form the averaged constraint forces required to keep the bud areas constant. As will be discussed in detail below, the hemi-spherical CCP constitutes a local minimum along the reaction path. The activation barrier to escape from this minimum amounts to $\sim 30 k_B T$, and is followed by spontaneous growth into a nearly completed CCV. Speculations on the intriguing ability that this barrier might be related to a go / no-go decision, and on how *in vivo* endocytic events overcome this barrier, will be postponed till the discussion.

This paper is structured as follows. The simulation models for clathrin, AP2 and the lipid membrane are described in Section 5.2, together with the simulation algorithm and the PMCF method. In Section 5.3 we present the simulation results. Finally, we discuss the results and their implications for the endocytic pathway in Section 5.4.

5.2 Models and Methods

In the first part of this section we briefly discuss the simulation models for clathrin, AP2 and the membrane, successively. These models were used in the Chapters 2 and 4 to model the assembly processes of purified clathrin triskelia, clathrin with AP2, and recently clathrin with AP2 and a membrane; more detailed descriptions of the models can be found in these references. In the second part of this section, we describe the Brownian equations of motion, the PMCF method, and introduce the reaction coordinate.

5.2.1 Models

Clathrin

Clathrin triskelia are modeled here as rigid patchy particles with three identical legs radiating from a common 'hub', see Fig. 5.1. Each leg is subdivided into three leg segments of identical length $\sigma = 17$ nm, referred to as the proximal and distal segments and the terminal domain (TD). The proximal legs are at an angle $\chi = 101^\circ$ relative to the three-fold rotational symmetry axis; this 'pucker' is typical of triskelia forming soccer-ball structures containing 60 triskelia, a commonly observed cage in *in vitro* self-assembly experiments in the presence of adaptor proteins (AP2). The orientation of distal leg segments was chosen to maximize the overlap between the segments of two triskelia, each with its hub adjacent to a knee of the other triskelion. The relative position and orientation of TDs was calculated using the structural data file 1XI4, [4, 37] available at the protein data bank (PDB); the TD forms an angle of $\sim 114^\circ$ with the adjacent distal segment and a dihedral angle of $\sim 28^\circ$ relative to the distal and proximal segments of the same leg. As a result, TDs are oriented to point inward, towards the center of the clathrin cage. In a cage, a clathrin hub is found at every vertex and each edge is composed of two proximal and two distal leg segments, with the like segments aligned in antiparallel fashion, see Fig. 5.2.

The binding of clathrin legs in cages is thought to result from a multitude of weak interactions distributed along the entire lenght of the leg. Our inter-segmental interaction potential introduces attractions between all segmental pairings observed in experimental cages, see Fig. 5.2(B). The strength of the interaction depends on the collective distance between the two ends of the two segments and is anisotropic under rotations of these segments around their long axis. The latter reflects a concentration of binding sites along one side of the leg, see Fig. 5.2(C), and proved crucial to the self-assembly of cages in previous simulations [30, 38]. Excluded volume

particle	parameter	symbol	value
clathrin	segmental length	σ	17 nm
	pucker	χ	101°
	intersegment energy[2]	ϵ	$6k_BT$[3]
	transl. diff. coeff.	D_1^t, D_2^t	$4.46 \cdot 10^4\sigma^2/\mathrm{s}$
		D_3^t	$3.70 \cdot 10^4\sigma^2/\mathrm{s}$
	rotat. diff. coeff.	D_1^r, D_2^r	$1.64 \cdot 10^4/\mathrm{s}$
		D_3^r	$1.02 \cdot 10^4/\mathrm{s}$
	bulk concentration[4]	$[\mathrm{C}]_r$	$10^{-3}\sigma^{-3}$
AP	linker length	L_{01}	0.9σ
		L_{12}	1.5σ
	spring constant	k_{01}	$50k_BT/\sigma^2$
		k_{12}	$30k_BT/\sigma^2$
	transl. diff. coeff.	D_0^t	$2 \cdot 10^4\sigma^2/\mathrm{s}$
		D_1^t, D_2^t	$1 \cdot 10^5\sigma^2/\mathrm{s}$
	bulk concentration	$[\mathrm{A}]_r$	$\left[10^{-3}{-}10^{-2}\right]\sigma^{-3}$
membrane	bending rigidity	k_κ	$10k_BT$
	spring constant	k_m	$100k_BT$
	equilibrium length	r_m^0	0.5σ
	maximum length	L_m	0.8σ
	excluded volume	$\sigma_\mathrm{m}^\mathrm{ev}$	0.35σ
		$\epsilon_\mathrm{m}^\mathrm{ev}$	$10k_BT$
	transl. diff. coeff.	D_m^t	$10^2\sigma^2/\mathrm{s}$
AP – CL	click strength	ϵ_CA	$\left[6{-}14\right]k_BT$
	click radius	ρ_CA	0.25σ
AP – mb	click strength	ϵ_Am	$\left[6{-}14\right]k_BT$
	click radius	ρ_Am	0.1σ
mb – clathrin or AP.	repulsion	ϵ_m	$10^3k_BT/\sigma^4$
	vertical range	b	0.1σ
	horizontal range	d	1.0σ

Table 5.1: Summary of the applied simulation parameters for clathrin, AP and membrane particles, as well as their interaction parameters.

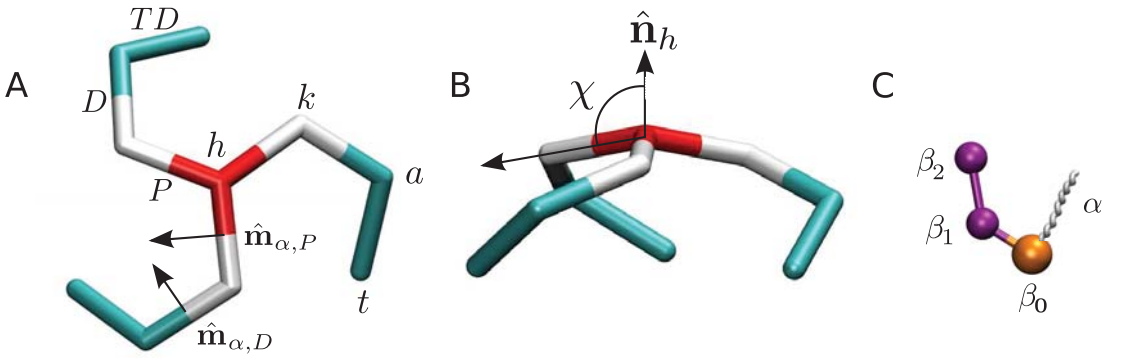


Figure 5.1: The coarse-grained simulation models for the clathrin triskelion (A and B) and AP2 (C), on the same scale. The rigid clathrin triskelion is constituted by three identical legs, each composed by a proximal leg segment (P) radiating from a central hip (h) to the knees (k), followed by distal leg segments (D) running to ankles (a) and terminal domains (TDs) ending at the toes (t). The intrinsically curved structure of the triskelion is characterized by a pucker angle χ relative to the symmetry axis $\hat{\mathbf{n}}_h$. The directionality of leg-leg interactions is represented by the polarity vectors $\hat{\mathbf{m}}$ attached to each leg segment. The AP model features three beads: the β_1 and β_2 beads can bind to the toes and ankles of clathrin, respectively, while the core β_0 can bind to a membrane bead. The beads are connected by two flexible linkers. The full AP2 protein also possesses a flexible α linker, which is omitted in the simulations as it does not bind to membranes nor to clathrin.

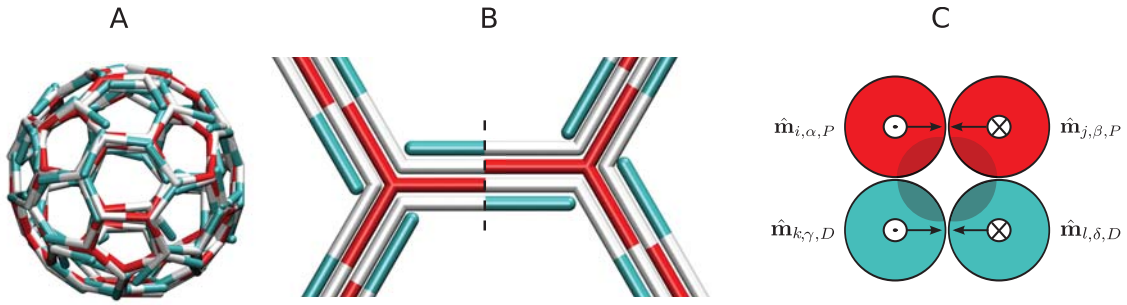


Figure 5.2: Representations of the clathrin cage structure. (A) Snapshot of a self-assembled cage in solution. (B) Edges are composed of two proximal and two distal leg segments, with like segments oriented anti-parallel; vertices are meeting points of one hub (red), three knees (white) and three ankles (blue). For clarity, the leg segments are drawn next to each other. (C) The cross section of an experimental cage edge, along the dashed line in (B), with the two distal segments below (*i.e.* inside the cage) the two proximal segments. The markers $\odot$ and $\otimes$ denote segments pointing (from hip to knee or from knee to ankle) in to and out of the plane of the picture, respectively. Shaded areas highlight the hypothesized locations of binding sites; their asymmetric distribution along the leg's circumference is included in the simulation model by means of a torsion potential acting on the segmental polarity vectors $\hat{\mathbf{m}}$. For clarity, the terminal domains, curving inward from the ankle toward the center of the cage, have been omitted in these illustrations.

interactions, which prevent two triskelia from binding at the same vertex, are not explicitly modeled for reasons of computational efficiency but are mimicked by a repulsive potential between parallel like segments. The TDs are thought to play no role in the inter-clathrin interactions, and merely serve as binding sites for APs.

AP2

The AP2 model consists of three point particles, representing the β_1 and β_2 sites on the β linker binding to the clathrin TD and knee, respectively, and the core, β_0, with a membrane binding site, see Fig. 5.1. The interactions of these sites with their respective partners are described by click potentials, discussed below. The three sites are connected by two structureless flexible linkers, modeled as finitely extensible non-linear elastic (FENE) elements, [39]

$$U(r) = \begin{cases} -\frac{1}{2}kL^2 \ln\left[1 - \left(\frac{r}{L}\right)^2\right] & \text{for} \quad r < L \\ \infty & \text{for} \quad r \geq L, \end{cases} \tag{5.1}$$

where r denotes the elongation of the linker and L its maximum lenghth, estimated based on the numbers of amino acids in the linkers as 0.8σ and 1.5σ for the $\beta_0 - \beta_1$ and $\beta_1 - \beta_2$ linkers, respectively. The spring constant k is estimated using the expression for an entropic spring in polymer physics, [40]

$$k = \frac{3k_BT}{2Ll_p}, \tag{5.2}$$

where l_p denotes the persistence length. With an experimental value of $l_p \approx 0.6\,\text{nm}$ for disordered proteins, the resulting spring constants for the linkers become $30k_BT/\sigma^2$ and $50k_BT/\sigma^2$, respectively. The α linker and α appendage are not represented in the model, as they do not directly interact with clathrin nor with the membrane.

Structural studies revealed that the AP2 protein undergoes a large conformational change [41, 42]. The protein is in the 'locked' state when dissolved in the cytosol, with the β_1 site buried within the core to suppress the formation clathrin cages. Once AP binds to the cell membrane and cargo, it transforms to the 'open' state and releases the β_1 site to promote cage assembly. This mechanism is reproduced in the simulations by allowing a β_1 bead to click to a TD only when the matching core bead has clicked to a membrane node [32].

The membrane model

The membrane bilayer is modeled as an elastic triangulated mesh with bending rigidity. This mesh is a computationally attractive structure, formed by N_m vertices connected by $3N_m$ edges forming $3N_m/2$ triangular facets. A detailed description of the model, the model parameters and the validation, is provided in Chapter 4. The free energy of a membrane bilayer with no intrinsic curvature, fixed topology and without edges, is described by the Helfrich expression, [40, 43, 44]

$$F = 2\kappa \int_A H^2 da + \frac{1}{2}K_A A_0 \left(\frac{A - A_0}{A_0}\right)^2, \tag{5.3}$$

where the integral runs over the entire area A of the membrane, H denotes the local mean curvature, κ the bending rigidity, K_A is the elasticity modulus, and A_0 is the equilibrium surface area. The coarse grained membrane model aims at reproducing this free energy, and in particular the bending rigidity because of its relevance to bud formation.

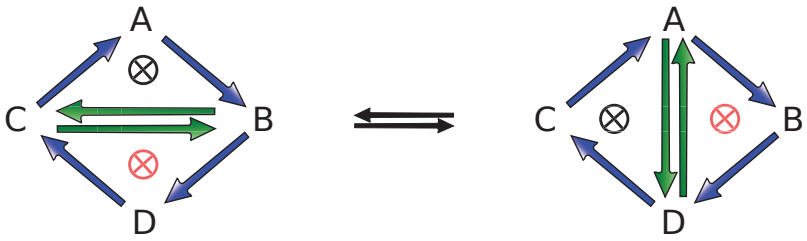


Figure 5.3: Representation of the Monte Carlo bond flipping move employed in the simulations to ensure membrane fluidity and enable large amplitude deformations. A randomly selected bond, here represented by the two green arrows, is flipped from one diagonal to the other diagonal of the quadrangle formed by its two adjacent triangles. Each triangle in the network is equipped with a normal vector, represented in the picture by the arrows on its edges and following the right-hand convention, allowing identification of the inner and outer spaces separated by the membrane. The orientation of the flipped edge is assigned so as to preserve the definitions of these normals.

The elasticity of the membrane is modeled by FENE springs along all edges of the mesh, with a spring constant k_{m} and finite extensibility L_{m}. As a first approximation, we note that for a purely hexagonal lattice $k_m = 2K_A/\sqrt{3}$. An equilibrium area is introduced by making the potential non-Hookean, substituting the argument r in Eq. (5.1) by $r - r_{\mathrm{m}}^0$. The bending rigidity of the membrane is modeled by

$$\Phi = \frac{1}{2} k_\kappa \sum_i \phi_i, \tag{5.4}$$

with k_κ denoting the strength of the potential. The summation runs over all membrane beads, with

$$\phi_i = \frac{1}{\Omega_i} \Big[\sum_{j(i)} (\mathbf{x}_i - \mathbf{x}_j) \Big]^2, \tag{5.5}$$

where the sum runs over all nodes j connected to i and Ω_i represents the collective area of all triangles joining at site i. The typical bending rigidity of lipid bilayers, [45] $\kappa = 0.4 \times 10^{-19}$ J, is reproduced by selecting $k_\kappa = \kappa$. Finally, vertices are prevented from crossing edges and facets, and self-avoidance of the membrane is imposed, by introducing Weeks-Chandler-Andersen (WCA) excluded volume interactions between all membrane beads, using a strength $\epsilon_{\mathrm{m}}^{\mathrm{ev}}$ and a diameter $\sigma_{\mathrm{m}}^{\mathrm{ev}}$.

Previous simulations revealed that triangulated networks with a fixed connectivity behave as purely elastic surfaces and are unable to form or relax pit-like invaginations [32]. In order to create a deformable fluid membrane, Monte Carlo moves are used to modify the network connectivity and thereby enable long-range diffusion of membrane beads. The MC move consists of selecting a random bond composing the common edge of two triangles in the lattice, and flipping it to become the other diagonal of the quadrilateral, [43] as illustrated in Fig 5.3. This move conserves the total number of triangles and the position of the nodes involved, but it is likely to change the stretching energy of the flipped bond and the bending energies of the two triangles and their neighbours. Hence, a Metropolis Monte Carlo scheme is used to calculate the acceptance probability of every attempted move. The connectivity numbers of the nodes involved in the move increase or decrease by one. In order to conserve the geometrical properties of the triangulated network, any attempt move that reduces the connectivity of a node below five is rejected [43]. Bond flips are attempted at a rate of r_f attempts per bond and per unit of time; see Table 5.2 for an overview of all MC rates. For later use, we note that a unit vector

function	symbol	value (s^{-1})
chemostats, CL and AP	r_c	10^5
membrane bond flipping	r_f	10^4
AP-clathrin click	r_{CA}	10^6
AP-membrane click	r_{Am}	10^6
reshuffle clicked AP	r_v	10^4
AP-membrane jumps	r_j	10^4

Table 5.2: Attempt rates of all Monte Carlo moves discussed in the main text. The five values in the second half of the table denote rates per particle, to eliminate system size dependence.

normal is attached to every facet to distinguish 'inside the cell' from 'outside'. Previously, the orientations of the normals of the two newly generated triangles were determined by selecting the orientations with the highest dot product relative to the normals of the two disappearing triangles. This procedure proved unreliable for the high curvatures encountered in the current study, and was replaced be determining the orientation compatible with applying the right-hand convention to the orientations of the conserved non-diagonal triangle edges.

Interactions

The short ranges, relative to the protein dimensions, of the interactions between the AP sites β_0, β_1 and β_2 binding to a membrane bead, a clathrin TD and knee, respectively, are numerically inconvenient. In stead, we apply click potentials [31, 32]. Consider, as an example, the interaction between the α^{th} site of the i^{th} triskelion and the β^{th} site of the j^{th} AP. The status of this interaction is described by a flag $b_{i\alpha,j\beta}$ that assumes two values, reflecting the discrete state of the interaction. The interaction potential then reads as

$$\phi_{\mathrm{click}}(r_{i\alpha,j\beta}, b_{i\alpha,j\beta}) = \begin{cases} 0 & \text{for} \quad b_{i\alpha,j\beta} = 0 \\ \begin{cases} -\epsilon_{\mathrm{CA}} & \text{for} \quad r_{i\alpha,j\beta} < \rho_{\mathrm{CA}} \\ \infty & \text{for} \quad r_{i\alpha,j\beta} \geq \rho_{\mathrm{CA}} \end{cases} & \text{for} \quad b_{i\alpha,j\beta} = 1, \end{cases} \tag{5.6}$$

with $\epsilon_{\mathrm{CA}} > 0$ the strength of the click and ρ_{CA} the maximum distance between two clicked sites.

The click interactions are turned on and off by MC trial moves, at a rate of r_{CA} attempts per second and per AP bead. The MC move consists of several steps, starting with the random selection of an AP bead. A list is made of all K possible binding sites within the interaction radius of that bead, including the current site if the bead is already clicked, and the unclicked state is added as the 0th option. For each of these $K+1$ trial moves, the accompanying energy change is calculated. One of these states is then selected, with probability

$$P_k = \frac{\exp(-\beta \Delta \phi_k^{\mathrm{click}})}{\sum_{k'=0}^{K} \exp(-\beta \Delta \phi_{k'}^{\mathrm{click}})}. \tag{5.7}$$

Note that the coordinates of the particles are left untouched.

Excluded volume interactions between AP beads have been omitted for reasons of computational efficiency. The total volume of all APs bound to a cage is small in comparison to the available volume, *i.e.* the spherical shell bounded by the clathrin cage and the enclosed vesicle. We note, however, that excluded volume interactions play an important role in preventing multiple AP beads binding to the same clathrin or membrane site. This is effectively inhibited in the model by making the clicks mutually exclusive, allowing the AP, clathrin and membrane

sites to partake in only one interaction at a time. We furthermore note that the two distinct clathrin-AP bounds, β_1 binding toes and β_2 binding ankles, are probably characterized by different interaction parameters. Since neither interaction free energy has been measured, to the best of our knowledge, for reasons of convenience we assume here that both bonds are identical.

A repulsive interaction prevents the triskelia and APs from crossing the membrane floating roughly in the middle of the simulation box, dividing it into an 'interior' region above the membrane and an 'exterior' region below it. The use of 3D periodic boundary conditions allows the particles to diffuse freely between these two regions, and thereby eliminates an osmotic pressure difference that would cause the membrane to drift downward. The membrane offers a repulsive boundary to beads arriving from both sides; the membrane is inert, except for AP β_0 beads approaching from the interior. Because of the random displacement term in the BD algorithm of Equation (5.13), a thin repulsive boundary is not easily realized by merely introducing a repulsive potential between the particles and the membrane nodes. To ensure that particles are pushed back in the right direction, including those that have slightly overstepped the membrane, each particle is labelled with a flag l_i that takes the values $+1$ and -1 for particles in the interior and exterior, respectively, and retains that value until the particle crosses the periodic boundaries. The interaction between the i^{th} particle and the membrane then reads as

$$\phi_i^{\text{m}} = \epsilon_{\text{m}} \sum_t f(l_i r_{it}^{\perp}) g(r_{it}^{||}), \tag{5.8}$$

where ϵ_{m} is a positive strength parameter and the sum runs over all membrane triangles. With $\hat{\mathbf{n}}_t$ the inward pointing normal to the t^{th}membrane triangle, the vector $\mathbf{r}_{it}$ from the center of this triangle to a bead is decomposed into the height of the particle relative to the triangle $r_{it}^{\perp} = \mathbf{r}_{it} \cdot \hat{\mathbf{n}}_t$ and and its lateral displacement $r_{it}^{||} = |\mathbf{r}_{it} - r_{it}^{\perp} \hat{\mathbf{n}}_t|$. Multiplication of the height with the particle's flag l_i, as in the argument to the penalty function f, yields a positive (negative) value for particles at the permitted (forbidden) side of the membrane triangle. This function is defined as zero for particles beyond a height b, is quadratic just above the membrane triangle, and increases linearly for particles penetrating the membrane,

$$f(x) = \begin{cases} 0 & \text{for} \quad x \geq b \\ (x-b)^2 & \text{for} \quad 0 < x < b \\ b^2 - 2bx & \text{for} \quad 0 \leq x. \end{cases} \tag{5.9}$$

The function g distributes the interaction over several adjacent triangles, reflecting the size of the protein fragment interacting with the membrane,

$$g(x) = \begin{cases} (x-d)^2 & \text{for} \quad 0 < x < d \\ 0 & \text{for} \quad x \geq d, \end{cases} \tag{5.10}$$

with d setting the decay length of the spread function. The repulsion by the membrane applies to the β_1 and β_2 beads of AP, as well as to the hubs, knees, ankles and toes of triskelia. The core particle of AP is excluded to allow it to 'click' to the membrane, as discussed above. The modest flexibility of the clathrin protein, allowing a leg to bend when pressed against a membrane, is mimicked in the simulations with a 90% reduction of the repulsion strength ϵ_{m} at the extreme ends of the legs, *i.e.* the toes.

Chemostat

In the modestly sized simulation box, the accumulation of triskelia and APs in a nascent coat results in a significant depletion of the cytosolic concentrations of these proteins. We therefore

simulate the exchange of clathrin triskelia and APs with ideal large reservoirs, by Monte Carlo algorithms maintaining constant chemical potentials, and hence constant bulk phase concentrations, for both solute molecule types independently. Consider a system with N_{C} triskelia in a volume V. Trial moves to add or remove a triskelion, with equal likelihood, are attempted at a rate of r_c attempts per second. The probability of accepting an insertion move is given by [33, 34]

$$P^{\mathrm{acc}}_{N_{\mathrm{C}}\to N_{\mathrm{C}}+1} = \min\left\{1, \frac{[\mathrm{C}]_r V}{N_{\mathrm{C}}+1} e^{-\beta\Delta\Phi_{N_{\mathrm{C}}\to N_{\mathrm{C}}+1}}\right\}, \tag{5.11}$$

where $[\mathrm{C}]_r$ refers to the clathrin concentration in the reservoir and $\Delta\Phi_{N_{\mathrm{C}}\to N_{\mathrm{C}}+1}$ is the potential energy change accompanying the insertion of a triskelion at a random position and a random orientation. Similarly, the acceptance probability for the removal of a randomly selected triskelion is given by

$$P^{\mathrm{acc}}_{N_{\mathrm{C}}\to N_{\mathrm{C}}-1} = \min\left\{1, \frac{N_{\mathrm{C}}}{[\mathrm{C}]_r V} e^{-\beta\Delta\Phi_{N_{\mathrm{C}}\to N_{\mathrm{C}}-1}}\right\}. \tag{5.12}$$

The reference bulk concentrations and the attempt rates are listed in Tables 5.1 and 5.2, respectively.

5.2.2 Methods

Constrained Brownian Dynamics Simulations

The trajectories of the simulated clathrin triskelia, AP and membrane beads, are calculated by first order Brownian Dynamics equations of motion with superimposed Monte Carlo moves for the clicked beads. In Brownian Dynamics, the solvent particles are not explicitly simulated, but the effect of the numerous collisions with solvent molecules is translated into friction and random terms in the equations of motion of the simulated particles, related by the fluctuation-dissipation theorem [33]. Here we present the Brownian Dynamics equation of motion in a generalized form, for a full set of generalized coordinates $\mathbf{Q}$ representing the positions and orientations of the particles. We refer the reader to previous publications [32, 46] for the equivalent expanded equations, discussed in detail in Chapter 4.

The generalized Brownian dynamics equation of motion reads as

$$\mathbf{Q}(t+\delta t) - \mathbf{Q}(t) = -\boldsymbol{\mu}^Q \frac{\partial F}{\partial \mathbf{Q}}\delta t + k_B T \frac{\partial \boldsymbol{\mu}^Q}{\partial \mathbf{Q}}\delta t + (\boldsymbol{\mu}^Q)^{1/2}\boldsymbol{\Theta}^Q(t)\sqrt{2k_B T \delta t}. \tag{5.13}$$

The first term on the r.h.s. describes the displacement over a time step δt resulting from a generalized force, with F representing the free energy as a function of the generalized coordinates, being opposed by friction from the solvent, with $\boldsymbol{\mu}^Q$ representing a generalized mobility matrix depending on the coordinates. The second term on the r.h.s. accounts for the inhomogeneity of the mobility tensor, which is required in the Itô representation to recover the correct Boltzmann equilibrium distribution. The last contribution on the r.h.s. represents the Brownian stochastic displacements due to the collisions with the solvent. The numerical values of the parameters are reported in Table 5.1.

Because the clathrin triskelion is modeled as a rigid particle, the positions and orientations of its leg segments are described by the Cartesian coordinates of its center and a rotation matrix accounting for the overall orientation. The latter is conveniently expressed in terms of unit quaternions, *i.e.* four coordinates q_γ with a constraint of unit length $|\mathbf{q}| = 1$, and propagated by Eq. (5.13). The constant body-based translational and rotational mobility tensors were determined using the HYDRO++ package [47] by modeling clathrin as an array of 52 spheres arranged

to reflect the overall shape of the protein [48]; these are readily converted by means of the rotation matrix to the space-based mobility tensors entering the equation of motion.

The motion of individual membrane beads is described by a translational BD equation for their positions. A constant diffusion coefficient was chosen for simplicity. Because the membrane is the fastest moving component of the simulated system, for reasons of computational efficiency we selected the smallest possible mobility coefficient that yields a numerically stable membrane for the maximum time step permitted by the triskelia, see Chapter 4.

The positions of the three beads composing an AP are updated separately, in a fashion that depends on their clicking status with clathrin and/or membrane. All unclicked beads are propagated by translational Brownian Dynamics assuming an isotropic mobility tensor; the AP beads do not rotate. The mobility coefficient μ_β^t is estimated through the Stokes-Einstein equation for isolated Brownian spheres,

$$\mu_\beta^t = \frac{1}{6\pi\eta R_\beta}, \tag{5.14}$$

where η is the viscosity of the suspending fluid and R_β the sphere's radius. Assuming diameters of $\sim$9 nm for the 200 kDa AP2 core and $\sim$4 nm for the appendage domain, [49] the mobilities of the β_0 and β_2 beads become $\mu_0^t = 1.2 \cdot 10^{10}\,\mathrm{Ns/m}$ and $\mu_2^t = 2.6 \cdot 10^{10}\,\mathrm{Ns/m}$, respectively, in water at room temperature. For reasons of simplicity, the mid-linker β_1 bead is equipped with the same dynamic properties as the β_2 bead.

All clicked AP beads move with the clathrin triskelion or membrane bead they are clicked to, as a single rigid body following the equation of motion of this second particle. The aggregate is subject to the sum of all forces and torques acting on the composing units. This motion preserves the clicking status of the AP beads involved. The translational and rotational mobilities of the aggregate are assumed to be identical to those of a free triskelion or membrane node. On top of this motion, the AP beads explore the small clicking volumes by Monte Carlo trial moves taking them to random positions within these volumes, at a rate of r_{CA} attempts per second and per bead. These trial moves also preserve the clicking status, but may alter the potential energy and therefore trial moves are accepted or rejected with a probability given by the Metropolis scheme [33, 34]:

$$P_{o\to n}^{\mathrm{acc}} = \min\left\{1, e^{-\beta\Delta\Phi_{o\to n}}\right\}, \tag{5.15}$$

where $\Delta\Phi_{o\to n} = \Phi(n) - \Phi(o)$ is the potential energy change between the old configuration o and the new configuration n.

Since a membrane-bound AP2 complex can diffuse relative to the membrane, in the simulation MC moves are included that enable a clicked AP core bead to jump from its membrane node to a randomly selected neighboring node, at a rate of r_j attempts per membrane bound AP and per second. The acceptance probability of trial moves to unoccupied nodes is again given by Eq. (5.33), while trial jumps to occupied nodes are rejected.

The PMCF method

In the introduction we speculated that the spontaneous growth from a hemispherical lattice to a fully enveloping clathrin coat requires surmounting a free energy barrier. If this proves to be correct, this region of the free energy landscape can only be sampled efficiently through the use of biased simulations, as commonly used to study 'rare events' like reactions characterized by the presence of a high activation barrier. A reaction coordinate $\xi(\{\mathbf{r}_i\})$ is defined, chosen to map the entire coordinate space onto a single coordinate describing the progression of the reaction. The free energy profile along this coordinate is then calculated by biased sampling methods

that improve the sampling efficiency at those values of the reaction coordinate that are rarely sampled by regular equilibrium simulations, *e.g.* through the use of a constraint or an umbrella potential, taking care to correct the simulation results for the applied bias [33, 34].

The potential of mean constraint force (PMCF) method traditionally uses a series of constrained Molecular Dynamics (MD) simulations for a range of values ξ^* of the reaction coordinate. Following the SHAKE routine of MD, we add to the BD equation of motion Eq. (5.13) a constraint force of the form $\mathbf{F} = -\lambda_\xi \nabla \xi$, where the Lagrange multiplier λ_ξ is computed iteratively from $\xi(\{\mathbf{r}_i(t+\delta t)\}) = \xi^*$. By varying ξ^*, one forces the system to sample hyperplanes of coordinate-space, including planes that would be infrequently sampled in unconstrained simulations. In MD simulations the free energy derivative can then be calculated from the average constraint force, including minor corrections related to the mass-metric tensor [35, 50, 51]. A similar expression is expected to hold true in Brownian Dynamics simulations, [46]

$$\frac{dF(\xi^*)}{d\xi} \simeq \langle \lambda_\xi \rangle_{\xi^*}, \tag{5.16}$$

where $\langle \ldots \rangle_{\xi^*}$ denotes the average over a constrained BD simulation at $\xi = \xi^*$. Upon integrating the latter expression, one obtains the free energy function

$$F(\xi) = \int_{\xi'}^{\xi} \langle \lambda_\xi \rangle_{\xi^*} \, d\xi^* + F_{\xi'}, \tag{5.17}$$

where the last term is an arbitrary reference constant.

The reaction coordinate

To assist clathrin in bending the membrane, we introduce a reaction coordinate based on the number of membrane beads in the proximity of the center of the membrane bud, $\mathbf{r}_c$. The selected definition is a sum of one-particle contributions,

$$\widetilde{\xi}\big(\{\mathbf{r}_i\}, \mathbf{r}_c\big) = \sum_i \bar{\xi}(|\mathbf{r}_i - \mathbf{r}_c|_\parallel) \tag{5.18}$$

where the sum runs over all membrane nodes in the simulation box, and $|\mathbf{r}_i - \mathbf{r}_c|_\parallel$ measures the distance between the i^{th} particle and the vertical axis through $\mathbf{r}_c$, *i.e.* the projected distance parallel to the groundplane of the box. The smoothed one-particle function is defined by

$$\bar{\xi}(r) = \frac{1}{2}\left(1 - \tanh\frac{r - r_0}{\alpha}\right), \tag{5.19}$$

where r_0 sets a reference distance and α the width of transition range. In the limit of $\alpha \to 0$ this function reduces to a step function with $\bar{\xi} = 1$ within a cylinder of radius r_0 around $\mathbf{r}_c$ and $\bar{\xi} = 0$ outside this cylinder, and consequently the value of $\widetilde{\xi}$ reduces to the number of membrane beads inside the cylinder. This definition is inspired by the reaction coordinate used by Tolpekina *et al.* [52] to measure the free energy of pores in lipid membranes.

To create a reaction coordinate ξ whose value remains unchanged under rigid body translations of the entire membrane, one must make $\mathbf{r}_c$ a function of the bead positions. Following earlier work on transmembrane pores, [53] we define $\mathbf{r}_c$ as the point that minimizes $\widetilde{\xi}$. Then

$$\frac{\partial \widetilde{\xi}(\{\mathbf{r}_i\}, \mathbf{r}_c)}{\partial \mathbf{r}_c} = -\sum_i \frac{\partial \bar{\xi}(|\mathbf{r}_i - \mathbf{r}_c|_\parallel)}{\partial \mathbf{r}_i} = 0, \tag{5.20}$$

where the first equivalence follows from Eq. (5.19). With this choice for the center, the reaction coordinate

$$\xi(\{\mathbf{r}_i\}) = \widetilde{\xi}\big(\{\mathbf{r}_i\}, \mathbf{r}_c(\{\mathbf{r}_i\})\big) \tag{5.21}$$

is an internal coordinate and consequently the total force exerted by the constraint equals zero. The position of the center can not be solved analytically but is easily determined numerically by a number of algorithms. Since the analytical expressions for the first and second derivatives of $\widetilde{\xi}$ with respect to $\mathbf{r}_c$ are available, see Eq. (5.20) and its derivative

$$\frac{\partial^2 \widetilde{\xi}}{\partial \mathbf{r}_c \partial \mathbf{r}_c} = \sum_i \frac{\partial^2 \bar{\xi}(|\mathbf{r}_i - \mathbf{r}_c|_{\|})}{\partial \mathbf{r}_i \partial \mathbf{r}_i}, \tag{5.22}$$

the center can be located efficiently by the Newton-Raphson method. The total derivative of the reaction coordinate with respect to the position of particle i, to be used in the SHAKE algorithm, is given by

$$\frac{\partial \xi}{\partial \mathbf{r}_i} = \frac{\partial \widetilde{\xi}}{\partial \mathbf{r}_i} + \frac{\partial \widetilde{\xi}}{\partial \mathbf{r}_c} \cdot \frac{\partial \mathbf{r}_c}{\partial \mathbf{r}_i} = \frac{\partial \bar{\xi}}{\partial \mathbf{r}_i}, \tag{5.23}$$

where the last step follows from the definitions of $\mathbf{r}_c$ and $\widetilde{\xi}$.

5.3 Results

5.3.1 The assembly of clathrin pits

In our previous paper, all membrane beads were available for binding with the core bead of APs [32]. To simulate the presence of a low number of cargo molecules, here only a fraction of the membrane beads can bind AP. In an initial set of simulations, these 'active' beads were uniformly distributed over the membrane and only diffused through the motion of these membrane beads. At low fractions of active beads, the membrane bound triskelia rarely meet one another and consequently no coat assembly was observed within feasible simulation time scales (~ 1 second). Since membrane-bound cargo will diffuse substantially more than an entire membrane patch, *i.e.* a membrane bead, we allow the 'activity' to jump to a neighbouring bead. A Monte Carlo move is implemented producing jumps of the activity from one unclicked membrane node to the next at a rate of r_s attempts per activated particle per unit of time. For clicked active membrane beads, the activity jumps with the occupying AP core. These moves conserve both the total number of 'activated' membrane beads and the number of clicks. With this procedure, clathrin and AP are again able to assemble into coated pits. These coats and their formation process closely resemble those observed in our earlier simulations with all membrane beads in the active state, as described in the previous chapter. A relatively small fraction of active sites, around 5% corresponding to a density of $74\,/\,\mathrm{nm}^2$ suffices for the spontaneous formation of nucleation points on the simulation time scale. These nuclei are then able to recruit additional triskelia and APs from the solution and grow into coated pits.

5.3.2 Free energy calculations

The constraint method and related free energy calculation are validated by simulating a bare membrane, in the absence of clathrin and AP. This simple scenario allows for a comparison of the simulation results against an analytical expression based on Helfrich's theory. Since the constraint controls the number of membrane nodes in a cylindrical 'inner' region of radius r_0 around $\mathbf{r}_c$ but does not restrict the shape of the membrane in any other way, the membrane will

Figure 5.4: The tubular protrusion induced in a membrane when the number of membrane nodes within r_0 from the tube's long axis is constrained to $\xi^* = 300$.

minimize its free energy by forming a tubular protrusion of radius $\sim r_0$ centered around $\mathbf{r}_c$, see Fig. 5.4. The reduction of the number of nodes in the 'outer' region will result in the build-up of a tension in that region, which is reduced by selecting a ground plane area approximately 10 times larger than the cross section of the tube. Note that this tension does not 'enter' the tube, as the constraint exerts forces in the intermediate region, *i.e.* near the bottom of the tube. The resulting simulation box with $A_\parallel = 412\sigma^2$ contains a membrane of $N_m = 2064$ beads. In the absence of clathrin, the time step is increased to $\delta t = 10^{-7}$ s. A set of simulations are performed, systematically varying the value of the constrained reaction coordinate ξ^* from 190 to 300, while keeping all other characteristics of the simulation box fixed at their initial values. In particular, $A_\parallel$ is constant.

For an equilibrium thermally fluctuating membrane, $\xi \approx 100$. When the reaction coordinate is increased beyond this value by the constraint, the membrane in the inner region passes through contorted and buckled shapes until at sufficiently high ξ^* the tubular protrusion emerges, which grows in length with increasing ξ. The free energy of this tube, as calculated by Eq. (5.17), is shown in Fig. 5.5. Markers represent averages over two independent simulations of about 7 s, each requiring about a week of CPU time on a single core. We observe a good quantitative agreement between the simulation results and our analytical expression for $\xi^* > 150$, The theoretical predictions also shown in that graph are obtained by applying Helfrich's theory to the cylindrical and flat sections of the membrane, is then approximated as

$$F(\xi) = 2\frac{\kappa}{r_0^2}A_{\rm cyl} + \frac{1}{2}K_{A0\parallel}A_{0\parallel}\left(\frac{\left[A_\parallel - \pi r_0^2\right] - A_{0\parallel}(\xi)}{A_{0\parallel}(\xi)}\right)^2. \tag{5.24}$$

The area of the cylinder may be assumed to be given by $A_{\rm cyl} = \xi a_{0\parallel}$, with $a_{0\parallel} = 0.18\sigma^2$ the projected area per node for a flat tensionless membrane. For the outer region, the term between square brackets represents the constant projected area covered by the membrane, while the equilibrium area $A_{0\parallel}(\xi) = a_{0\parallel}\xi$ decreases as more and more nodes are pulled into the tube. Contributions resulting from the tip and rim of the tube are assumed to be independent of ξ and therefore irrelevant for the comparison with simulations. The good quantitative agreement between simulations and theory for $\xi > 230$ indicates that the constraint is working properly; the small deviations may be attributed to the approximations in the theory.

We now turn to the simulations of a clathrin coated pit. In our preceding simulations of triskelia and APs near a membrane, the bulk concentrations (in a volume well away from the

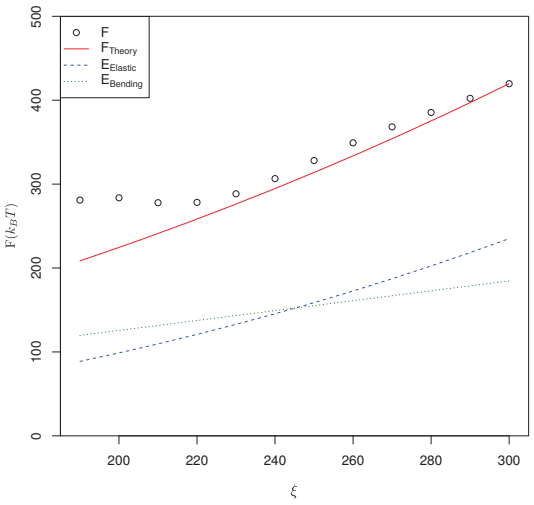


Figure 5.5: The free energy of the tubular membrane protrusion, see Fig. 5.4, formed by imposing a constraint on the number ξ of membrane nodes within r_0 from a vertical axis. Markers indicate simulation results and lines the theoretical results of Eq. (5.24) for the total free energy (solid), the elastic component (dashed) and the bending component (dotted). The theoretical values are treated as absolute values, the simulation results where shifted to match the theoretical value at $\xi = 250$.

membrane) of both proteins were stabilized at $[\mathrm{C}]_r = 0.001\sigma^{-3}$ and $[\mathrm{A}]_r = 0.003\sigma^{-3}$, respectively, by a chemostat mimicking coupling to ideal reservoirs at these concentrations, while the tension on the membrane was effectively removed by a barostat scaling the two in-plane box lengths independently. The constrained simulations, using a time step of $\delta t = 10^{-8}$ s, are initiated with a system from these earlier simulations, which had reached the point where the single coated pit in this system ceased to grow. To prevent the formation of a second pit, all APs outside the pit were removed and the ability of APs to (un)click was turned off, freezing them all in their initial click state. After equilibrating the system, the barostat was also turned off; the production runs are at constant ground plane area $A_\parallel$ and at constant $[\mathrm{C}]_r$.

Constrained simulations were run for values of the reaction coordinate ranging from 150 to 300. Snapshots of these simulations are shown in Fig. 5.6, corresponding to three values of the imposed reaction coordinate. Note that the constraint effectively blocks the growth and shrinkage of the pit area – each run samples an equilibrium at a different stage of CCP maturation. A sequence of images at successive values of ξ then looks like a pit growing to completion, but the simulations are not following the actual path that a single maturating pit would follow in time. It is clear from the images that with the constraint we can generate coats that are well beyond the hemi-spherical stage, *i.e.* beyond the stable end point reached in the preceding conventional equilibrium simulations. At high values of ξ, when a clear neck starts to form connecting the membrane inside the coat to the mother membrane, we start to run into the limitations of the current coarse-grained model. The free energy profile is again extracted from the constraint force. We have to take into account that the flat area of the membrane is exposed to increasing tension as the bud grows bigger; the correction for this contribution to the free energy is identical to that applied when growing the tubular protrusion. The calculated free energy profile is presented

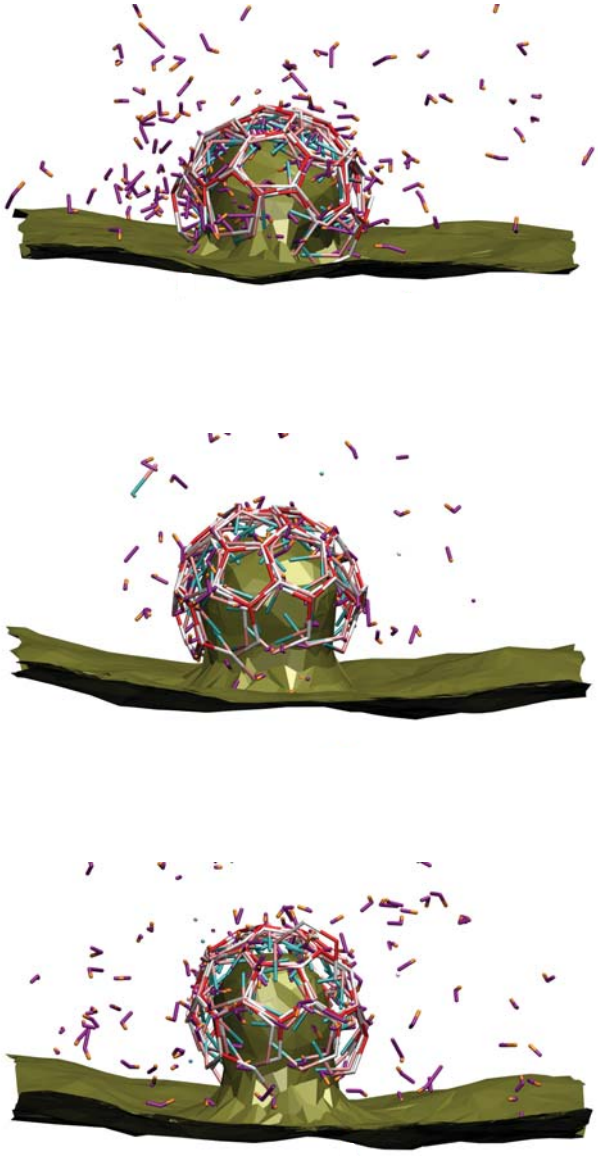


Figure 5.6: Snapshots of clathrin coated pits from three simulations with the reaction coordinate, obtained for a pucker angle of $\chi = 103°$ *i.e.* the bud area, constrained to $\xi = 200$, 250 and 300, from top to bottom. The pictures are made with Visual Molecular Dynamics [54] (VMD), using the same colour scheme as in Fig. 5.1.

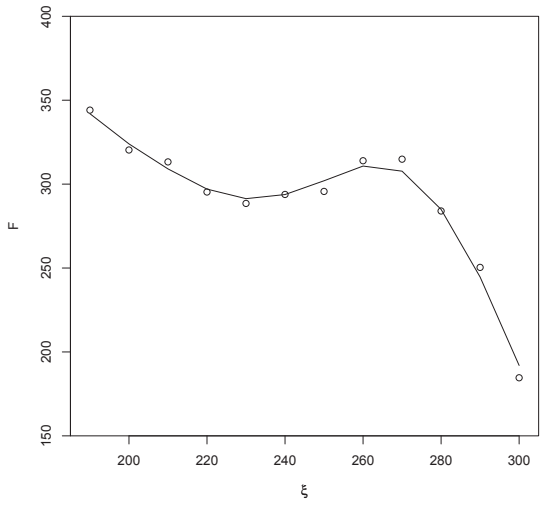


Figure 5.7: The free energy profile of a clathrin coated pit as a function of the reaction coordinate ξ. The markers represent the simulation results, the line is a guide to the eye. Note the existence of a local minimum and an activation barrier.

in Fig. 5.7, with each marker representing a simulation of 1 s (requiring about two weeks on a single core). One readily observes a local minimum, followed by a reaction barrier of $\sim 30 k_B T$. This explains why the spontaneously formed buds in our earlier simulations did not mature but became arrested midway. Beyond the activation barrier lies a steep decline to lower free energies, indicating that this part of the maturation path will proceed spontaneously. Unconstrained simulations initiated with configurations extracted from constrained simulations with $\xi \geq 260$ do indeed show a steady growth of the coat.

5.4 Discussion and Conclusions

The assembly of a clathrin coated vesicle mediated by AP2 at a membrane was studied by means of Brownian dynamics simulations. Both these proteins and the membrane are represented by coarse-grained models, largely derived from the available experimental data on their structure, mechanical properties and interaction characteristics. All the clathrin pits observed in our previous simulations [32] grow by polymerization of single clathrin triskelia or adaptor proteins, and are able to bend the underlying membrane, but fail to incorporate additional units upon reaching a configuration corresponding to about half the size of a complete vesicle, prior to the scission through dynamin. This behavior, observed in all simulations, can be explained by the existence of a free energy barrier, effectively halting the spontaneous transition to a fully grown clathrin vesicle.

We made use of the potential of mean constraint force method to explore the nature of the barrier, by expressing the free energy profile as a function of a reaction coordinate ξ through constrained Brownian Dynamics simulations. The reaction coordinate was chosen to allow a direct control on the density of membrane beads, that represent in our coarse-grained model patches of membrane lipids, within a given distance from the center of the coat. We validated the method by pulling a membrane tube from a flat membrane, enabling a direct comparison between

simulation results and a theoretical prediction. Upon constraining the reaction coordinate to high values, the membrane forms an elongated structure by balancing the bending and elastic energy contributions to the free energy. The configuration is well described by means of the Helfrich free energy expression. The presence of clathrin, in conjunction with AP2s, shapes the membrane in a structure resembling an Ω, typical of clathrin coated vesicle. The free energy profile of the assembly process on the tensionless plasma membrane of a cell is then calculated by subtracting the free energy related to the tension induced by the constraint from the overall reconstructed profile. This calculation revealed the presence of a barrier of about $30k_BT$, that makes the spontaneous completion of a clathrin cage an extremely unlikely event, in line with our previous observations from earlier simulations.

The height of the barrier suggests that other endocytic proteins, not required for the initial stages of the nucleation and growth process, are rather implicated in surmounting the activation barrier. We note that this result is relative only to the transition from a partial to a full coated pit; a possible method for calculating the complete energy profile associated with the assembly process of clathrin and AP units from the solution is briefly discussed in the outlook section.

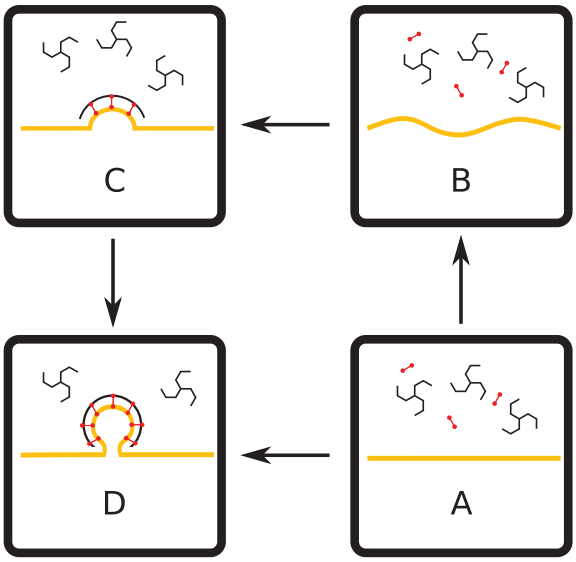


Figure 5.8: The proposed procedure to calculate the total mechanical work F_{AD} performed by clathrin triskelia and AP, initially in the bulk phase at a tensionless membrane (A), in bending the membrane into a clathrin coated membrane invagination during endocytosis (D). In Chapter 5, we characterize the transition from a system with a partially to a fully built coat (C $\rightarrow$ D). In the main text we describe the calculations required to compute the energy released upon building the partial coat from a system with a membrane with a compressive tension resulting from keeping the surface area constant during the process (B $\rightarrow$ C), and to calculate the energy of the compression (A $\rightarrow$ B), and thus to complete the cycle.

5.5 Outlook and future developments

In this section we briefly discuss how to calculate the total mechanical work performed by clathrin triskelia in bending the membrane during endocytosis into a clathrin coated membrane invagination, based on the results presented in this thesis. The complete process is divided in a sequence of steps, pictured in Fig. 5.8, where the arrows represent the analytical or computational calculations for the free energy difference between two steps. In this chapter we established the free energy profile of the activation barrier to completely encoated membrane buds. *i.e.* step $C \rightarrow D$ in Fig. 5.8.

The free energy released during the formation of the hemi-spherical initial bud, *i.e.* step $B \rightarrow C$, can be established by performing simulations using the coupling parameter approach, whereby an interaction term in the force field, namely the AP-clathrin click interaction, is stepwise turned on and off through multiplication by a coupling parameter Λ, with $0 \leq \Lambda \leq 1$ [33, 34]. In these simulations, the conditions must be kept identical to those in the constrained simulations described in this chapter, *i.e.* same constant ground plane area, with a chemostat stabilizing the chemical potential of the triskelia in solution and the APs remaining attached to the membrane.

Using thermodynamic integration, the free energy difference between the state with the potential term at conventional strength and the state with vanishing AP-clathrin click interaction follows from

$$F(1) - F(0) = \int_0^1 \frac{dF(\Lambda)}{d\Lambda} d\Lambda = \int_0^1 \langle \phi_{clAP}(\Lambda \epsilon_{\mathrm{CA}}) \rangle_\Lambda d\Lambda. \tag{5.25}$$

where ϵ_{CA} denotes the total energy of the AP-clathrin clicks.

Several minor steps are required to convert the calculated free energy differences into free energy differences relevant in a biological context. Assuming that the cell membrane is tensionless throughout the endocytic process, we have already corrected the path $C \rightarrow D$ for the tension resulting from the constraint. The free energy of the stable hemi-spherical bud, state C in Fig. 5.8, can be calculated from the constrained simulations as

$$F_C = \int e^{-\beta F(\xi)} d\xi, \tag{5.26}$$

where the integral runs over the entire potential well corresponding to state C in Fig. 5.7. A similar equation applies to the stable near-spherical bud in state D and, by combining the two free energies obtained from the single free energy curve, we arrive at the free energy difference F_{DC}. We furthermore note that the free energy of state C corresponds to $F(1)$ in the coupling parameter approach, and that of state B to $F(0)$. Since the ground plane of the box was kept constant, at a value that nullified the tension for the hemi-spherical bud, the membrane in state B experiences a compressive tension. The free energy change upon zeroing this tension can be calculated from the mechanical properties of the membrane. Combining the above partial steps then yields F_{DA}, which represents the desired free energy difference.

References

[1] F. M. Brodsky, *Annual review of cell and developmental biology* **2012**, *28*, 309.

[2] T. Kirchhausen, D. Owen, S. C. Harrison, *Cold Spring Harbor perspectives in biology* **2014**, *6*, a016725.

[3] M. S. Robinson, *Traffic* **2015**, *16*, 1210–1238.

[4] A. Fotin, Y. Cheng, P. Sliz, N. Grigorieff, S. C. Harrison, T. Kirchhausen, T. Walz, *Nature* **2004**, *432*, 573–579.

[5] E. M. Schmid, H. T. McMahon, *Nature* **2007**, *448*, 883.

[6] L. Hinrichsen, A. Meyerholz, S. Groos, E. J. Ungewickell, *Proceedings of the National Academy of Sciences* **2006**, *103*, 8715–8720.

[7] H. T. McMahon, J. L. Gallop, *Nature* **2005**, *438*, 590–596.

[8] D. Loerke, M. Mettlen, D. Yarar, K. Jaqaman, H. Jaqaman, G. Danuser, S. L. Schmid, *PLoS Biol* **2009**, *7*, e1000057.

[9] J. Heuser, L. Evans, *J. Cell Biol.* **1980**, *84*, 560–583.

[10] J. E. Heuser, J. H. Keen, L. M. Amende, R. E. Rippoldt, K Prased, *J. Cell Biol.* **1987**, *105*, 1999–2009.

[11] I Gaidarov, F Santini, R. A. Warren, J. H. Keen, *Nat. Cell Biol.* **1999**, *1*, 1–7.

[12] T. Kirchhausen, *Trends in cell biology* **2009**, *19*, 596–605.

[13] Y Li, L Shang, G. U. Nienhaus, *J. Cell Biol.* **2001**, *155*, 291–300.

[14] J. P. Ferguson, N. M. Willy, S. P. Heidotting, S. D. Huber, M. J. Webber, C Kural, *J. Cell Biol.* **2016**, *214*, 347–358.

[15] X Wu, X Zhao, L Baylor, S Kaushal, E Eisenberg, L. E. Greene, *J. Cell Biol.* **2001**, *155*, 291–300.

[16] W. Kukulski, M. Schorb, M. Kaksonen, J. A. G. Briggs, *Cell* **2012**, *150*, 508–520.

[17] M Ehrlich, W Boll, A van Oijen, R Hariharan, K Chandran, M. L. Nibert, T Kirchhausen, *Cell* **2004**, *118*, 591–605.

[18] D Loerke, M Mettlen, S. L. Schmid, G Danuser, *Traffic* **2011**, *12*, 815–825.

[19] M. G. J. Ford, I. G. Mills, B. J. Peter, Y Vallis, G. J. K. Praefcke, P. R. Evans, H. T. McMahon, *Nature* **2002**, *419*, 361–366.

[20] W. M. Henne, H. M. Kent, M. G. Ford, B. G. Hegde, O. Daumke, P. J. G. Butler, R. Mittal, R. Langen, P. R. Evans, H. T. McMahon, *Structure* **2007**, *15*, 839–852.

[21] P. N. Dannhauser, E. J. Ungewickell, *Nat. Cell Biol.* **2012**, *14*, 634–639.

[22] J. C. Stachowiak, F. M. Brodsky, E. A. Miller, *Nat. Cell Biol.* **2013**, *15*, 1019–1027.

[23] S. Saffarian, E. Cocucci, T. Kirchhausen, *PLoS Biol* **2009**, *7*, e1000191.

[24] D. J. M. J. Grove and, A. E. Knight, S. T. Wavre-Shapton, T. Suna, E. D. Protonotarios, L. D. Griffin, J. Lippincott-Schwartze, M. Marsh, *Mol. Biol. Cell* **2014**, *25*, 3581–3594.

[25] O Avinoam, M Schorb, C Beese, J. A. G. Briggs, M Kaksonen, *Science* **2015**, *348*, 1369–1372.

[26] W. K. den Otter, W. J. Briels, *Traffic* **2011**, *12*, 1407–1416.

[27] R. Matthews, C. N. Likos, *Soft Matter* **2013**, *9*, 5794–5806.

[28] N. Cordella, T. J. Lampo, S. Mehraeen, A. J. Spakowitz, *Biophysical journal* **2014**, *106*, 1476–1488.

[29] N. Cordella, T. J. Lampo, N. Melosh, A. J. Spakowitz, *Soft matter* **2015**, *11*, 439–448.

[30] W. K. den Otter, M. R. Renes, W. J. Briels, *Biophys. J.* **2010**, *99*, 1231–1238.

[31] M Giani, W. K. den Otter, W. J. Briels, *Biophys. J.* **2016**, *111*, 222–235.

[32] M. Giani, W. K. den Otter, W. J. Briels, *The Journal of Chemical Physics* **2017**, *146*, 155102.

[33] M. P. Allen, D. J. Tildesley, *Computer Simulation of Liquids*, Clarendon Press, New York, NY, USA, **1989**.

[34] D. Frenkel, B. Smit, *Understanding Molecular Simulation*, 2nd, Academic Press Inc., Orlando, FL, U.S.A., **2001**.

[35] W. Den Otter, *The Journal of Chemical Physics* **2000**, *112*, 7283–7292.

[36] A. Laio, M. Parrinello, *Proceedings of the National Academy of Sciences* **2002**, *99*, 12562–12566.

[37] A. Young in Seminars in cell & developmental biology, *Vol. 18*, Elsevier, **2007**, pp. 448–458.

[38] W. K. den Otter, M. R. Renes, W. J. Briels, *J. Phys. Condens. Matt.* **2010**, *22*, 104103.

[39] K. Kremer, G. S. Grest, *The Journal of Chemical Physics* **1990**, *92*, 5057–5086.

[40] D. Boal, *Mechanics of the Cell*, Cambridge University Press, Cambridge, U.K., **2002**.

[41] L. P. Jackson, B. T. Kelly, A. J. McCoy, T. Gaffry, L. C. James, B. M. Collins, S. Höning, P. R. Evans, D. J. Owen, *Cell* **2010**, *141*, 1220–1229.

[42] B. Kelly, S. Graham, N Liska, *Science* **2014**, *345*, 459.

[43] D. Nelson, T. Piran, S. Weinberg, *Statistical Mechanics of Membranes and Surfaces*, 2nd ed., World Scientific Publishing, Singapore, **2004**.

[44] W. Helfrich, *Zeitschrift für Naturforschung C* **1973**, *28*, 693–703.

[45] W Rawicz, K. C. Olbrich, T McIntosh, D Needham, E Evans, *Biophys. J.* **2000**, *79*, 328–339.

[46] I. M. Ilie, W. J. Briels, W. K. den Otter, *J. Chem. Phys.* **2015**, *142*, 114103.

[47] J Garcia de la Torre, G del Rio Echenique, A Ortega, *J. Phys. Chem. B* **2007**, *111*, 955–961.

[48] I. M. Ilie, W. K. den Otter, W. J. Briels, *J. Chem. Phys.* **2014**, *141*, 065101.

[49] B. M. Collins, A. J. McCoy, H. M. Kent, P. R. Evans, D. J. Owen, *Cell* **2002**, *109*, 523–535.

[50] W. Den Otter, W. Briels, *The Journal of chemical physics* **1998**, *109*, 4139–4146.

[51] W. Den Otter, W. Briels, *Molecular Physics* **2000**, *98*, 773–781.

[52] T. Tolpekina, W. Den Otter, W. Briels, *The Journal of chemical physics* **2004**, *121*, 12060–12066.

[53] W. K. den Otter, *The Journal of chemical physics* **2009**, *131*, 11B614.

[54] W. Humphrey, A. Dalke, K. Schulten, *Journal of molecular graphics* **1996**, *14*, 33–38.

Summary

The assembly of clathrin coats in the presence of adaptor proteins was studied through computer simulations using coarse-grained models and through statistical mechanics. Adopting a reductionist approach based on recent experimental results, we aimed at reproducing and studying the minimal conditions that lead to the successful formation of aggregates, and at investigating the molecular properties and mechanisms required by the assembly process both in bulk conditions and at a membranous surface. In order to tackle this challenging task, coarse-grained models were used to describe all the assembly units involved in the simulations presented in this thesis. These models are based on the available structural data and are engineered to capture the key elements and behavior of the modeled proteins.

In Chapter 2 we introduce a coarse grained model of adaptor proteins, inspired by and representing the AP2 complex. The latter, the second most abundant component of endocytic coats after clathrin, is known to play a fundamental role in promoting and assisting the creation of coats at the cytosolic surface of the membrane. It is reported to be able to trigger polymerization of clathrin triskelia in physiological conditions of salt and pH, under which purified clathrin triskelia do not spontaneously self-assemble. The interaction between APs and clathrin were modeled throughout this thesis through a click potential, introduced for the first time in this chapter. The characteristics of the AP model, and of this interaction, have been tuned to reproduce the existing experimental assembly data of an AP2 and clathrin mixture. Our computer simulations provide novel insights into the role of AP2 in the self-assembly of clathrin cages and suggest that the mechanical properties of adaptor proteins are of fundamental importance. In the same chapter, we also developed a statistical mechanical theory that describes the equilibrium concentration of clathrin cages as a function of the other assembly variables and parameters, such as the protein concentrations and interaction strengths.

This theoretical model has been further developed in Chapter 3, in order to explicitly take into account the effect of the flexibility of the clathrin triskelion, previously neglected. The main aim of the chapter is to investigate the equilibrium properties of clathrin cages resulting from the aggregation process, with emphasis on their size in the absence and in the presence of adaptor proteins. In order to perform this study, the essential features and characteristics of clathrin and AP2s are captured through a small number of effective parameters, and the number of allowed aggregates is determined on the base of geometrical considerations and arguments. The model is able to capture the key mechanisms determining the experimentally known ability of AP2s to influence the size of a clathrin cage, and thus to shape the resulting cage size distribution.

In Chapter 4, we introduce a triangulated mesh model for an elastic membrane to investigate the formation of clathrin/AP2 coats at the cytosolic face of a cellular membrane. The model parameters are tuned to reproduce the typical properties of a biological membrane within the computational limits imposed by our simulations. In order to be able to extract the dynamical behavior associated to the aggregation process from the simulations, we make use of a compact Rotational Brownian Dynamics algorithm that uses quaternions to describe rotations, recently developed within the group. In the same spirit that led to the development of the statistical-

mechanical model accompanying the simulations of Chapter 2, we developed a Langmuir-like adsorption model for the clathrin/AP complex at the membrane. Through the combination of simulations and theory, we characterize the mechanisms by which an initial nucleation point constituted by a small number of assembly units is stabilized through a cooperative effect between APs and clathrin at the membrane surface. We furthermore describe and predict the conditions under which this nucleation point is able to grow into a hemispherical clathrin coat.

In all the simulations performed in Chapter 4, the growth of the clathrin coat halts upon reaching a hemispherical configuration, hinting towards the existence of an activation barrier in the free energy profile associated with the assembly of a clathrin coat at the membrane. Chapter 5 is devoted to investigating and computing the free energy profile by means of constrained Brownian Dynamics simulations. The free energy is here expressed and computed as a function of a reaction coordinate by integrating the average constraint force. Our results confirm the existence of a free energy barrier, implying the action of other endocytic components, possibly other membrane-bending proteins, at a specific step of the assembly process.

Samenvatting

De assemblage van clatrine-mantels in de aanwezigheid van adapter-eiwitten is bestudeerd middels computersimulaties met grof-korrelige modellen en middels statistische mechanica. Met een reductionistische aanpak gebaseerd op recente experimentele resultaten richten we ons op het reproduceren en bestuderen van de minimale condities die leiden tot de succesvolle formatie van aggregaten, en het onderzoeken van de moleculaire eigenschappen en mechanismes benodigd voor het assemblageproces zowel in bulkcondities en aan een membraanoppervlak. Om deze uitdagende taak aan te pakken worden grof-korrelige modellen gebruikt voor de beschrijving van alle assemblage eenheden betrokken bij de simulaties die gepresenteerd worden in dit proefschrift. Deze modellen zijn gebaseerd op de beschikbare structurele data en zijn ontworpen om de essentiële elementen en gedrag van de gemodeleerde eiwitten te vangen.

In Hoofdstuk 2 introduceren we een grofkorrelig model van adaptereiwitten, geïnspireed op en voorstellende het AP2 complex. Van deze laatste, op clathrine na de meestvoorkomende component in endocytische mantels, is bekend dat zij een fundamentele rol spelen in het aanjagen en assisteren van de creatie van mantels aan het cytosol oppervlak van het membraan. Er wordt gerapporteerd dat het in staat zou zijn de polymerisatie van clathrine driepoten in gang te zetten in fysiologische zout en zuurgraad condities, waarbij gezuiverde clathrine driepoten niet spontaan zelf-assembleren. De interactie tussen APs en clathrine werd in het gehele proefschrift gemodeleerd door een klikpotentiaal, voor het eerst geïntroduceerd in dit hoofdstuk. De karakteristieken van het AP model, en van zijn interacties, zijn afgestemd op het reproduceren van de bestaande experimentele assemblage data van een AP2 en clathrine mengsel. Onze computersimulaties leveren nieuwe inzichten in de rol van AP2 in de zelf-assemblage van clathrine kooien en suggereert dat de mechanische eigenschappen van adaptereiwitten van fundamenteel belang zijn. In hetzelfde hoofdstuk ontwikkelen we ook een statisch-mechanische theorie die de evenwichtconcentraties van clathrine kooien beschrijft als functie van de andere assemblage variabelen en parameters, zoals de eiwitconcentraties en interactiesterktes.

Dit theoretische model wordt verder uitgewerkt in Hoofdstuk 3, om daarmee het effect van de flexibiliteit van clathrine driepoten, voorheen genegeerd, expliciet in rekening te brengen. Het belangrijkste doel van het hoofdstuk is het onderzoeken van de evenwichtseigenschappen van clathrine kooien zoals die uit het aggregatieproces volgen, met nadruk op hun grootte in de afwezigheid en aanwezigheid van adaptereiwitten. Om deze studie uit te voeren zijn de essentiële eigenschappen en karakteristieken van clathrine en AP2 gevangen in een klein aantal effectieve parameters, en het aantal toegestane aggregaten is bepaald aan de hand van geometrische overwegingen en argumenten. Het model is in staat het sleutelmechanisme te vangen dat leidt tot de experimenteel bekende competentie van AP2 om de grootte van clathrine kooien te beïnvloeden, en daarmee vorm geeft aan de resulterende verdeling van kooigroottes

In Hoofdstuk 4 introduceren we een driehoekig netwerk model om de formatie van clathrine/AP2-mantels aan de cytosol-zijde van cellulaire membranen te onderzoeken. De modelparamaters zijn afgestemd op het reproduceren van typische eigenschappen van biologische membranen binnen de computationele limieten opgelegd door onze simulaties. Om in staat te zijn het dynamische

gedrag geassocieerd met het aggregatieproces uit de simulaties af te leiden, maken we gebruik van een compact Rotationeel Brownse Dynamica algoritme dat quaternionen gebruikt om rotaties te beschrijven, dat recentelijk ontwikkeld is in de groep. In dezelfde geest die leide tot de ontwikkeling van een statistisch-mechanisch model ter begeleiding van de simulaties in Hoofdstuk 2, ontwikkelen we een Langmuir-achtig adsorptiemodel voor het clathrine/AP-complex aan het membraan. Door het combineren van deze twee methodes karakteriseren we het mechanisme waardoor een initieel nucleatiepunt opgebouwd uit een klein aantal assemblage-eenheden wordt gestabiliseerd foor een coöperatief effect tussen APs en clathrine aan het membraanoppervlak. Verder beschrijven en voorspellen we de omstandigheden waaronder dit nucleatiepunt in staat is uit te groeien tot een hemisferische clathrine-mantel.

In alle simulaties uitgevoerd in Hoofdstuk 4 stopt de groei van de clahrine-mantels als de hemisferisch configuratie wordt bereikt, hetgeen suggereert dat er een activeringsbarrière bestaat in het vrije-energie profiel geassocieerd met de assemblage van een clathrine mantel aan het membraan. Hoofdstuk 5 is gewijd aan het onderzoeken en berekenen van het vrij-energie profiel door middel van geforceerde Brownse Dynamica simulaties. De vrije energie wordt hier uitgedrukt en berekend als een functie van een reactiecoördinaat door integratie van de gemiddelde opgelegde kracht. Onze resultaten bevestigen het bestaan van een vrije-energie barrière, en impliceert daarmee de actie van andere endocytische componenten, mogelijk membraan-buigende eiwitten, tijdens een specifieke stap in het assemblage proces.

Scientific Output

Publications

- R. Cagliani, U. Pozzoli, D. Forni, A. Cassinotti, M. Fumagalli, **M. Giani**, M. Fichera
  Crohns disease loci are common targets of protozoa-driven selection,
  Molecular biology and evolution, p. mst020, 2013.

- **M. Giani**, W. K. den Otter, and W. J. Briels
  Clathrin assembly regulated by adaptor proteins in coarse-grained models
  Biophysical Journal, vol. 111, no. 1, pp. 222-235, 2016 (featured on the journal website).

- **M. Giani**, W.K. den Otter, W. J. Briels
  Early stages of clathrin aggregation at a membrane
  The Journal of Chemical Physics, 146(15), 155102.

- **M. Giani**, W.K. den Otter
  Adaptor proteins shape the size distribution of clathrin cages
  in preparation

- **M. Giani**, W.K. den Otter, W. J. Briels
  Constrained simulations of Clathrin coated vesicles
  in preparation

Talks

- **M. Giani**, W.K. den Otter, W. Briels,
  Simulating the self-assembly of clathrin cages
  Gordon Research Seminar, Andover (NH), USA June 2014 (invited speaker).

- **M. Giani**, W.K. den Otter, W. Briels,
  Coarse-grained simulations of clathrin assembly initiated by adaptor proteins
  Amsterdam, NL. MMBS,November 2014.

- **M. Giani**, W.K. den Otter, W. Briels,
  Clathrin assembly regulated by adaptor proteins in coarse-grained models
  Amsterdam, NL. DMDD, February 2015.

- **M. Giani**, W.K. den Otter, W. Briels,
  Clathrin assembly regulated by adaptor proteins in coarse-grained models
  Cambridge, UK. IOP, April 2016.

Posters

- *Simulating the self-assembly of clathrin cages*
  Cargse, Corsica, France, June 2013.
- *Simulating the self-assembly of clathrin cages*
  Villars, Switzerland, September 2013.
- *Simulating the self-assembly of clathrin cages*
  Veldhoven, Netherlands, December 2013.
- *Simulating the self-assembly of clathrin cages*
  Veldhoven, Netherlands, February 2014.
- *Simulating the self-assembly of clathrin cages*
  Leiden, Netherlands, February 2014.
- *Simulating the self-assembly of clathrin cages*
  Andover (NH), USA, June 2014.
- *Coarse-grained simulations of clathrin assembly initiated by adaptor proteins*
  Julich GER, November 2014, .
- *Coarse-grained simulations of clathrin assembly initiated by adaptor proteins*
  Bad Honnef, GER, November 2014.

Acknowledgments

The experience of this PhD has truly been life changing, and it could not have been possible without the support I received throughout all these years from many people that I had the fortune to meet, inside and outside the University.

I would like to express my gratitude and appreciation to my supervisor, Wouter, that supported and encouraged me throughout all these years and from whom I learned so much. I want to thank him for his guidance and for allowing me to learn and grow; without his persistent help this dissertation would not have been possible. I sincerely admire his attention to details and, most of all, the sharpness and rigorousness of his analysis and thinking. I surely appreciated the enthusiasm and energy that my promotor, Wim, demonstrated in these years, that constitute the base of his passion and scientific career. Thanks to his experience, interests and knowledge of different areas he was able to widen the perspective of our discussions. His long-term vision helped me overcome the daily issues that inevitably emerged during these years. I want to thank Stefan, my second promotor, for our discussions and his availability. I appreciate his honesty towards me and the interest and time he dedicated to me and my project.

I am very thankful to all present and past members and friends of the CBP group for our frequent and quite long coffee breaks, (I miss them already!) some of which are surely something to be remembered. I want to thank Vishal (the first) for our discussions, through which we became friends: not only did I came to discover and appreciate the similarities and differences between our cultures, religions and traditions but I came to understand more about my own culture. I could have used the very same words to thank Vishal (the second), to whom I am grateful also for the interesting chats about movies and music. I wish both of them the best, and hope to meet them soon in Bangalore (or wherever they will be). I am glad I finally convinced my officemate Ioana to come and play volleyball with me, after just some years of insisting! I wish her good luck with her future academic career. Our already mentioned coffee breaks were not the same without Csaba. All of a sudden nobody dared to discuss or propose 'the very popular internet argument of who would win in a deathmatch between 1000 duck-sized horses or 1 horse-sized duck' [5]. It is really a pity that my stay in the CBP group, and the one of Tom did not overlap more. I am grateful for his kind advices and suggestions during my first (and last!) months. A special thanks to the ex-postdocs of our group, now scattered around the globe: Gabriel, Li, Igor, Barry. I wish them the best with their future careers! I want to thank Claudia for the kind words she had when I needed them. I wish I had Riccardo's dedication and passion for swimming and sports! I wish Monika and Habib good luck with completing their PhD, even though I know already they won't need it. A special thanks to Els and Sylvia, the secretaries of the CBP and MSM groups respectively, especially for this last stretch!

I am glad I took part in P-NUT (the PhD Network of the University of Twente) - I only regret having joined the association so late in my PhD! Good luck to Koen, Guido, Federica, Koen2, Davood, Ines, Imke, Sara and all the other board and regular members of P-NUT!

[5] just an example, and by far not the weirdest.

Some may say I have been, well, a bit annoying at one point of my PhD. I guess it's fair to say that everybody that came even remotely close to me in the last 4 years has been invited at least once to play volleyball. I am not at all sorry about that - in fact, I am proud of being one of the hardcore members and organizer of the Flying Nerds, one of my best memories of that happened during these years. I want to thank the handful of people - me, Bojan, Zarko, Andrej, Igors - constituting the very initial group. I joined them in the spring/summer of 2013 - I figured that would be a nice way of meeting people, and have fun. Some months later, the Flying Nerds VolleyBall Team was born. Since then, I have been playing with literally hundreds of people from all over the world. It has been simply amazing. Trying to mention them all would be an impossible task, but I want at least to thank our trainers, Daniel, Jozef, Denisa, Valentina, Kirill, for their efforts, and everybody who was involved in organizing our sport and social events.

A special thanks to Lucie, Michel - the dinners and lunches I had the pleasure of joining were simply amazing, and I'll make sure to be there for the next ones! I would like to thank all of my friends, still in Enschede (and surroundings) or scattered around the globe among which Sandro, Aliki, Andrea and everybody else who I did not mentioned here, but that still supported my journey, and that in many cases completed, or are about to complete, a similar one. I can simply say that I lived some of the best moments of my PhD with Effie and Andrea. Moments of pure freedom - of foolishness. I am smiling as I type this in, knowing they know what I am thinking about. Moo! My trip to Belgrade with Bojan, one of the kindest persons I know, was a true milestone in my PhD. I am writing this on the first birthday of his, and Anja's, daughter - I will be always grateful to them for having given me the possibility to be part of their family, as Jozef once said, and to share their joy. I have always admired Jozef's dedication to his job, and most importantly to his friends, and I'm happy to be one of them.

I was surely an unexpected surprise, after having studied physics in different universities (in a time so far away he still had hair) to end up in the same country as Luca for our PhD! Thanks for the nice moments together.

A special thanks goes of course to my family - without all of you, I simply would not be here. First I would like to thank my grandmother - she has taken care of me throughout all my life.

Thanks to my father and his collection, as a child I became a science fiction books addict, with my favorite book (still) being *The Skylark of Space* (*L'allodola dello spazio*) by Edward E. 'Doc' Smith. I believe those books had a major impact on my decision to study physics. Even in this difficult time, it's impossible not to smile thinking about his enthusiasm when, with simplicity, he first introduced me to the mysteries of science. And thanks to my mother, for always believing in me. A special thanks to my wonderful little sister, also scattered around Europe (and beyond?). I am glad to share this moment with all of them.

And finally, thanks to Lucia - I would never have been able to be here without her. One could almost reconstruct our story through all of our theses, of which this is (possibly!) the last one - from our Bachelors to our PhDs, and at the same time from Milano, through Trieste, to Enschede. The story of a mathematician and a physicist that fell in love long time ago and grew together sharing the same curiosity and passion. I wish for us never to lose the enthusiasm, the courage that led us here, and that will guide our lives together.

Matteo Giani was born in 1987 in Monza (Italy). In 2006 he started studying physics at the University of Milan, where he developed soon a passion for biophysics and for computational problems. After an experience in the bioinformatics group of the scientific institute IRCSS Medea (Bosisio Parini, Italy), he moved to the International School for Advanced Studies in Trieste (Italy), where he wrote his master thesis about the cyclization rate of polymers. In December 2012 he started his PhD work in a collaboration between the Multi Scale Mechanics and the Computational Biophysics group under the supervision of Dr. W. K. den Otter, Prof. W. J. Briels and Prof. S. Luding within the framework of a project founded by NWO. Since May 2016, he works in Delft as a scientific software engineer.