Coarse-Grained Simulations of the Assembly of Empty Viral Capsids

Author: Miguel Garrido Zornoza
Facultat de Física, Universitat de Barcelona, Diagonal 645, 08028 Barcelona, Spain.

Advisor: David Reguera López

Abstract: The morphogenesis of a virus is one of the most important steps in its life cycle, since a well-formed structure is vital to survive in hostile environments and to infect hosts. Viral self-assembly resembles a nucleation process where individual proteins stochastically form aggregates driven by a favourable free-energy of binding. The kinetics of the process is strongly regulated by the existence of a free-energy barrier between the viral constituents in solution and the fully-formed virus, in analogy with a first-order phase transition. In this work a coarse-grained model of viral assembly units is implemented in a Brownian Dynamics simulation and used to characterize viral nucleation kinetics. From simulations we evaluate the steady-state rate of capsid formation and reconstruct the free-energy landscape of the process, from which the critical capsid size and the nucleation barrier are obtained. Finally, a brief comparison of the simulations results with the predictions of classical nucleation theory (CNT) is discussed.

I. INTRODUCTION

Viruses are biological entities with no metabolism, constituted by a protein shell (or *capsid*) filled with genetic material (DNA or RNA). In its life cycle, a virus travels through hostile environments to find new host cells and deliver genetic material, used by the cell to create new copies of the viral constituents that self-assemble into a new virus. Viruses have the essential ability to self-assemble from its constituents into a wide, but defined, variety of nanometer-sized structures. The most common viral arquitecture has icosahedral symmetry, and can be classified by a discrete sequence of triangulation numbers T = 1,3,4,7,... The Caspar and Klug model [1] describes the way to construct a spherical viral shell with icosahedrical symmetry using two morphological units: hexamers and pentamers. These units are aggregates of 6 or 5 proteins respectively, that constitute the main capsid building blocks.

The self-assembly of a virus is a fundamental step in its life cycle since it is where the hollow protein structure, or capsid, that protects the genetic material inside, is constituted. Hence, to understand this process, and the key parameters that rule it, is of great interest from both a fundamental and a practical point of view, since its characterisation can be used, for example, to stablish possible targets for the development of vaccines.

In recent years there have been some *in vitro* experiments of the assembly of empty viral capsids [3] showing a series of features suggesting nucleation as the underlying mechanism: a strong concentration dependence; the lack of stable intermediates (i.e., partially formed capsids); and the existence of hysteresis in the conditions under which self-assembly and disassembly take place. All these experimental findings can be attributed to the presence of an energetic barrier between fully-formed capsids and its subunits in solution in analogy with a classical first-order phase transition. From a physical point of view, viral self-assembly is a spontaneous process that can be understood in terms of a favourable binding free-energy between subunits, due to hydrophobic interactions, that is able to overcome the entropic penalty of removing the subunits from the solution when the capsid is formed and the energetic penalty due to the fraction of subunits in the rim of a partially formed shell [2]. This energetic competition is responsible for the existence of a free-energy barrier and the activated kinetics characterized by a lag time before completed capsids are made. This has indeed been observed in experiments [3].

In this work, we will study the self-assembly kinetics of T = 3 empty viral capsids by using simulations of a coarse-grained model. From the simulations we are able to reconstruct the free-energy landscape of the nucleation process, obtaining the critical size, the free-energy barrier height and the steady-state rate of capsid formation, all as a function of subunits concentration. We compare the simulation results with the predictions from CNT applied to this particular 2 - dimensional case.

II. METHODS

A. CNT for viral assembly

Here we summarize the main ideas and key expressions of CNT applied to describe the assembly of empty viral capsids [4].

The free-energy of formation of a partial capsid of n subunits (see Fig. 1a) equals [4]

$$\Delta G(n) = n\Delta\mu + \gamma l\,, \tag{1}$$

where the first term is the bulk energy driving the assembly and the second term represents the energetic cost of the interface. In the previous expression, $\Delta\mu = -k_B T ln(c/c^*)$ is the difference of chemical potentials between a subunit in the capsid and in the solution, where c is the subunit concentration and c^* is the critical concentration (discussed later). The second term of Eq.(1) implements the energetic cost of having an incomplete

capsid, where γ is the line tension (i.e., the energetic cost per unit lenght) and l is the total contour of the boundary (see Fig. 1a). Taking a quasi-continuum aproximation we can calculate l for a spherical shell of fixed radius R, and rewrite Eq.(1) as [4]

$$\Delta G(n) = n\Delta\mu + a\sqrt{n(q-n)}\,, \tag{2}$$

being q the total number of subunits in the complete capsid, equal to $10T+2=32$ for a T = 3 icosahedrical shell [5], and $a = 4\pi R\gamma/q$. The barrier height and the critical size (i.e., the size of a partially formed capsid where the barrier has a maximum value) can be calculated by imposing the condition $\delta\Delta G\ /\delta n = 0$, yielding

$$\Delta G^* \equiv \Delta G(n^*) = \frac{q}{2}(\sqrt{\Delta\mu^2 + a^2} + \Delta\mu) \tag{3}$$

for the barrier heigth, and

$$n^* = \frac{q}{2}(1 + \frac{\Delta\mu}{\sqrt{\Delta\mu^2 + a^2}}) \tag{4}$$

for the critical size. Considering the process as a reaction cascade where only one subunit can be incorporated into the partial capsid at a time, one can describe the kinetics of viral assembly in terms of a master equation from which the steady-state rate of assembly can be calculated, obtaining [2]

$$J = c\beta^* Z e^{-\Delta G^*/kT}. \tag{5}$$

In the previous expression, c is the subunit concentration, β^* the rate of attachment of subunits to the critical capsid and

$$Z = \sqrt{\frac{|\Delta G''(n^*)|}{2\pi kT}} = \sqrt{\frac{a}{kTq\pi}}\left(1 + \left(\frac{\Delta\mu}{a}\right)^2\right)^{3/4} \tag{6}$$

is the so-called Zeldovich factor.

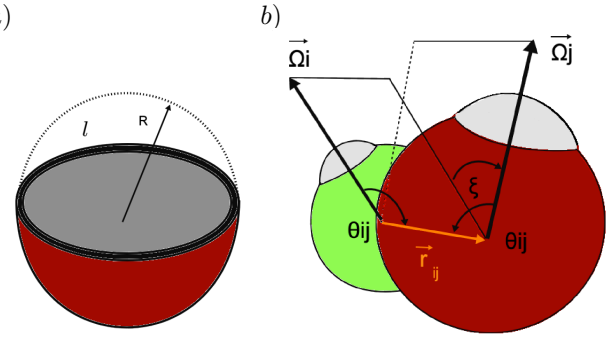

Figure 1: *a*) Representation of a partially formed capsid of radius R in the continuum aproximation, showing the rim of lenght l. *b*) Graphic representation of two capsomers, modeled as spheres, and its characterizing positional and orientational parameters, $(\vec{\Omega}_i, \theta_{ij}, \hat{r}_{ij})$.

B. Coarse-Grained Model

For this study a low-resolution model of the capsid building blocks, which has proved to be successful to mimic capsid assembly [6], has been used. In this particular coarse-grained model capsomers (i.e., hexamers and pentamers) are modeled as spheres of different size, and constitute the fundamental assembly subunits. Their interactions are described by a phenomenological anisotropic pair potential, built by three main ingredients. The first one is a standard Lennard-Jones potential accounting for short-range repulsion and mid-range attraction between subunits due to steric and hidrophobic effects. The second one is a bending contribution, which implements a preferred angle for subunit interactions by introducing an energetic penalty for angles that differ from the preferred one. This term will define the size of the shell and prevent the formation of compact clusters instead of hollow structures. The last ingredient is a torsion contribution, which introduces an additional energetic penalty keeping the subunits from assembling into non-closed surfaces of different concavity. The total potential is given by:

$$V(\mathbf{r}_{ij}, \vec{\Omega}_i, \vec{\Omega}_j) = \begin{cases} V_{LJ}, & r_{ij} \leq 2^{-1/6}\sigma_{ij} \\ V_{LJ}V_{ang}V_{tor} & r_{ij} > 2^{-1/6}\sigma_{ij} \end{cases} \tag{7}$$

where $r_{ij} = 2^{-1/6}\sigma_{ij}$ is the distance at which the Lennard-Jones potencial vanishes. This potential has the form of

$$V_{LJ}(r_{ij}) = \varepsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - 2\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right], \tag{8}$$

being ε_{ij} the binding energy between capsomers and σ_{ij} the equilibrium distance between subunits. The bending factor is implemented by the following expression:

$$V_{ang}(\hat{r}_{ij}, \vec{\Omega}_i, \vec{\Omega}_j) = e^{-(\theta_{ij}-\nu)^2/2\alpha^2} e^{-(\theta_{ji}-\nu)^2/2\alpha^2}. \tag{9}$$

Here ν controls the preferred angle and α implements a local bending stiffness. Finally, the torsion is modeled by:

$$V_{tor}(\vec{\Omega}_i, \vec{\Omega}_j) = e^{-k_t(1-cos\xi)/2} \tag{10}$$

where k_t is the torsion constant and ξ is the angle between two planes constructed with the unitary vector that links the position of the two interacting structures, (i, j), and its respective orientation vectors $(\vec{\Omega}_i, \vec{\Omega}_j)$ (see Fig. 1b).

C. Brownian Dynamics Simulations

The coarse-grained model described in the previous section has been implemented in a Brownian Dynamics simulation. Solvent presence has implicitly been taken into account through effective frictional plus random

forces exerted on each of the particles, representing random collisions of the subunits with the solvent. Given the typical size of the protein subunits, on the order of a few nanometers, inertial effects can be neglected, leading to the overdamping regime. The resulting Langevin equation has been integrated with a standard stochastic Euler algorithm. We have worked using reduced units, taking $\varepsilon_{ij} = \varepsilon_0 \equiv 1$, $T^* \equiv k_B T/\varepsilon_0$, the diffusion coefficient $D \equiv 1$ and $\sigma_{hh} \equiv \sigma \equiv 1$ (i.e., the equilibrium distance for hexamer-hexamer interaction is taken as unit lengh). Based on [6], we chose the torsion and preferred angle parameters to be $k_t = 1.5$ and $\nu = 1.9$. A reduced temperature $T^* = 0.1$ was selected, since self-assembly is assured to happen into a stable icosahedrical $T = 3$ structure, as previous studies indicate [5]. The Lennard-Jones potential is truncated at a distance of 5σ for the sake of computational efficiency. All simulations have been performed with only one type of subunit (hexamers) and a total of $N_h = 300$.

In order to study the concentration dependence of the assembly, simulations were performed in the range of $[0.007-0.1]\,\sigma^{-3}$. At high concentrations the formation of the capsid was not controlled by nucleation, so we focused on small concentrations. 100 repetitions were launched for every concentration in the interval $[0.007-0.013]\,\sigma^{-3}$ in steps of $0.001\,\sigma^{-3}$. Every simulation starts from a random configuration of positions and orientations. The cluster-size distribution is then analyzed every 1000 steps using a Stillinger criterion with $r_s = 1.2\sigma$ (i.e., if two subunits are closer than $r_s = 1.2\sigma$ they are considered to be part of the same cluster), until the first complete capsid with $n = 32$ subunits is formed. Fig. 2 shows a snapshot of the initial configuration and a complete capsid.

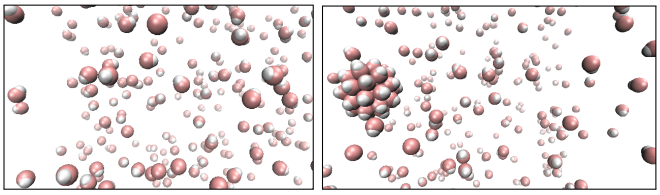

Figure 2: Snapshots showing the initial configuration and the formation of a complete capsid of $n = 32$ subunits.

D. Data analysis

1. *Kinetic Analysis*

A mean first-passage time (MFPT) analysis [7] was implemented to obtain kinetic information of the nucleation process. For this, the size of the largest cluster is tracked and saved every 1000 steps. The time where the size n has been reached for the first time, $t_i^{larg}(n)$, is stored for each simulation and averaged over all repetitions, yielding the MFPT, $\tau_{larg}(n) = \Sigma_i t_i^{larg}(n)/100$. The resulting MFPTs were fitted onto the error function [7]

$$\tau_{larg}(n) = \frac{\tau_J^{larg}}{2}[1 + erf(Z\sqrt{\pi}(n - n^*))] \qquad (11)$$

in order to obtain relevant kinetic parameters of the nucleation process (see Fig. 3). In particular, the steady-state nucleation rate is $J^{st} = 1/\tau_J^{larg} V$, n^* is the critical cluster size and Z is the Zeldovich factor.

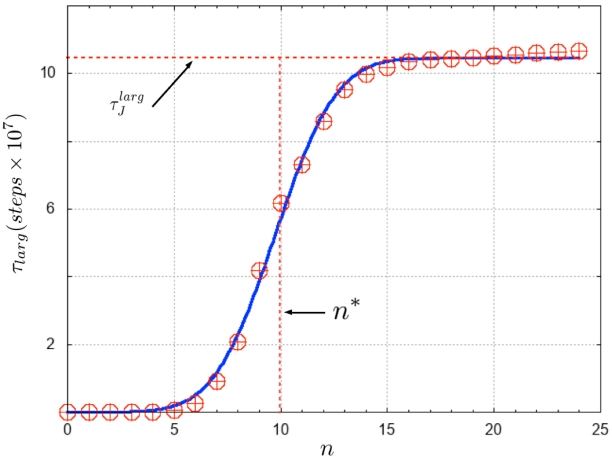

Figure 3: Sigmoidal curve of the MFPT for the nucleation process obtained in the simulation for the concentration $c = 0.008\,\sigma^{-3}$. Once the energetic barrier has been overcome the system reaches a time-independent rate of nucleation, given by the inverse of the MFPT plateau value multiplied by the volume of the system. The inflection point indicates the critical cluster size n^*.

2. *Reconstruction of the Free-Energy Landscape*

Along with the kinetic information one can perform a different type of analysis to obtain the free-energy landscape of the nucleation process [8]. For this we construct a histogram of the cluster-size distribution by counting the number of times a certain size of the cluster has been the largest over all performed simulations for a given concentration value. By normalizing the histogram one obtains a steady-state probability distribution of sizes $P_{larg}^{st}(n)$. Using this distribution and the MFPT, $\tau_{larg}(n)$, one can reconstruct the free-energy barrier as [8]

$$\beta\Delta G_{larg}(n) = \beta\Delta G_{larg}(n_i) + ln\left(\frac{B(n)}{B(n_i)}\right) - \int_{n_i}^{n} \frac{dn'}{B(n')}, \qquad (12)$$

where

$$B(n) = -\frac{1}{P_{larg}^{st}(n)}\left[\int_n^b P_{larg}^{st}(n')dn' - \frac{\tau_{larg}(b) - \tau_{larg}(n)}{\tau_{larg}(b)}\right], \qquad (13)$$

b is an absorbing boundary condition set to $n = 32$ (this is, once the largest cluster has reached this size the simulation stops) and $n_i = 3$ represents the minimum usable size, since $n = 1$ can't be generally used as $P_{larg}^{st}(1) = 0$ for most of the processes. Both integrals (12),(13) are

calculated by discretizing them into sumatorys, given the discrete nature of our computational data. From the free-energy of the "largest" cluster one can obtain the free-energy of any cluster as [9]:

$$\beta\Delta G(n) = \beta\Delta G_{larg}(n) + lnN_h \tag{14}$$

Fig. 4 shows an example of the resulting free-energy landscape.

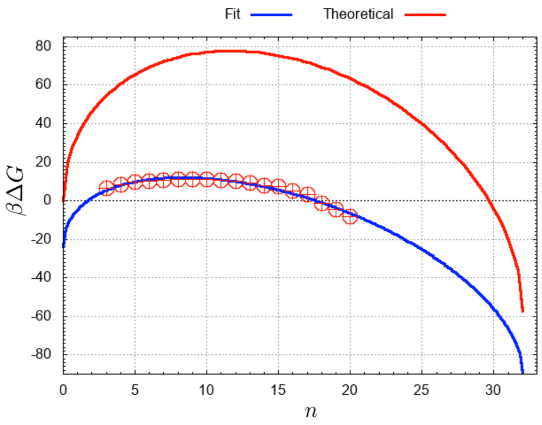

Figure 4: Free-energy landscape reconstructed from the simulations with $c = 0.008\,\sigma^{-3}$ (red circles). The red line represents the CNT theoretical prediction and the blue line is a fit to Eq.(2) plus a constant (see section *IV*).

III. RESULTS

1. *Critical concentration*

Several simulations were launched to obtain the so-called critical concentration c^*, which is one of the main ingredients of CNT. Since $\Delta\mu = -k_B T ln(c/c^*)$, at the critical concentration $\Delta\mu = 0$, and the critical size becomes $n^*(\Delta\mu = 0) = q/2$, as we can see from Eq.(4). Thus, at c^* a half-size cluster has 50% chance of dissassembling and 50% chance of becoming a full-size capsid. To determine c^* another type of simulations were performed. In these simulations a half-size cluster was placed at the start of the simulations surrounded by $N_h - 16$ randomly located particles, keeping all other parameters unchanged. A hundred repetitions were run at each concentration for $2 \cdot 10^7$ timesteps, until finding the concentration at which 50% of the simulations ended in a full cluster and 50% in a disgregated state. The value found was $c^* = 0.001325 \pm 0.000025$ (σ^{-3}).

2. *Kinetic and Thermodynamic results*

The simulations results are summarized in Figs. 5 and 6. The top graphic in Fig. 5 shows the critical cluster size obtained for different concentrations. It seems that the system has reached a concentration independent value in the region where simulations were launched. This could account for a possible magic number structure corresponding to a quarter of the full capsid. The bottom graphic in Fig. 5 corresponds to the steady-state nucleation rate as a function of capsid concentration. Fig. 6

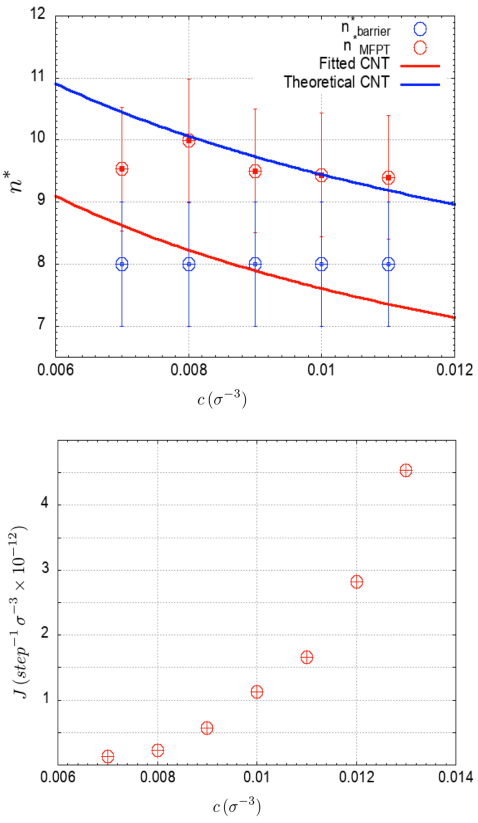

Figure 5: Critical cluster size (top) and steady-state nucleation rate (bottom) obtained from the BD simulations as a function of subunit concentration. The critical size has been estimated by fitting the sigmoidal curve (red) as well as directly from the barrier reconstruction (blue). The red curve shows the CNT prediction from Eq.(4) using the values for a and c^* obtained from the fit to Eq.(2) with $c = 0.008\,\sigma^{-3}$. The blue curve shows the theoretical CNT prediction discussed in section *IV*.

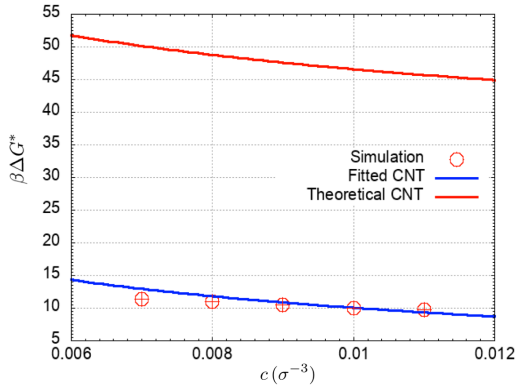

Figure 6: Barrier height as a function of capsid concentration, estimated from the free-energy landscape reconstruction. The analitical curves show the CNT prediction from Eq.(3) using the values for a and c^* obtained from: the fit to Eq.(2) with $c = 0.008\,\sigma^{-3}$ (blue); the theoretical estimation (red).

shows the barrier height obtained from the free-energy landscape reconstruction.

IV. DISCUSSION

At the highest concentrations, $c = (0.012, 0.013)$, the process is no longer activated, as signaled by a non-sigmoidal MFPT. In that case, the formation rate is evaluated as $1/(\tau(32)V)$. As expected: the nucleation rate exponentially drops when capsid concentration decreases, following the Arrhenius-like behaviour predicted by CNT in Eq.(5); and lower concentration values have higher energy barriers. The study has not pursued higher concentration values since the phase transition would no longer be regulated by nucleation, progressively entering into a spinodal process. The qualitative behaviour of the self-assembly observed in the simulations is clearly the one expected in a nucleation process.

To stablish the quantitative comparison with CNT we performed a 3-parameter fit (a, c^*, S) to Eq.(2) with the simulations results from $c = 0.008\,\sigma^{-3}$. Here S represents an additive term accounting for a possible shift between CNT and the computational estimation. Fig 4 shows that the fit is excellent whit $a = 0.37$, $c^* = 0.001023\,\sigma^{-3}$ and $S = -23$. To establish a theoretical comparison for a, we geometrically estimated its value by considering a partial capsid cut along its "zig-zag" direction [10], obtaining $a_{geo} = 0.45$. This represents a 17% discrepancy with respect to the one resulting from the fit. A 23% discrepancy between the critical concentrations is also found. We can use the fitted parameters to compare the barrier height and the critical size obtained from simulations with the ones predicted by CNT for $c = 0.008\,\sigma^{-3}$, by evaluating Eqs. (3) and (4), respectively, using $(a, c^*, S)_{fit}$. Fig.5 shows the critical size comparison, reflecting a partial failure of CNT both quantitatively and qualitatively. This could be accounting for a local magic number regarding viral structure that has not been taken into account in the classical theory. Fig. 6 shows the barrier height comparison, where, in the case of the fitted parameters, a close behaviour is observed.

This discrepancy $(a, c^*, S)_{theo} = (0.45, 0.001325, 0) \neq (0.37, 0.001023, -23) = (a, c^*, S)_{fit}$ could suggest either an oversimplification of the continuum theoretical description and its consequent failure to provide a precise quantification of the main parameters controlling the evolution of the system, or the possible existence of underlying mechanisms that have not been taken into account in CNT. These two options do not necessarily exclude each other. The successful fit to CNT suggests that the main physical ingredients regulating nucleation have been taken into account, but the value of the parameters involved is different.

V. CONCLUSION

In this work a coarse-grained model was used to study the kinetics of viral self-assembly. The steady-state nucleation rate as well as the critical capsid size and their concentration dependence were obtained. Besides kinetic information, the free-energy landscape of the nucleation process was reconstructed in order to estimate the free-energy barrier height. A quantitative comparison with CNT yields a large numerical discrepancy between predicted values and those obtained in the fit.

Although the CNT description seems qualitatively correct, the origin of the quantification failure stays open, and further inquiries will have to be done in future studies.

Acknowledgments

I would like to thank my parents for their support and David Reguera, for his patience and guidance at all times.

[1] D. L. D. Caspar and A. Klug, *Physical principles in the construction of regular viruses*, Cold Spring Harbor Symp. Quant. Biol., **27**, 1-24, 1962.

[2] A. Luque, *Structure, Mechanical Properties, and Self-Assembly of Viral Capsids*, PhD Thesis, Universitat de Barcelona, 2011.

[3] G. L. Casini, D. Graham, D. Heine, R. L. Garcea, D. T. Wu, *In vitro papillomavirus capsid assembly analyzed by light scattering*, Virology, **325**, 320-327, 2004.

[4] R. Zandi, P. van der Schoot, D. Reguera, W. Kegel, H. Reiss, *Classical Nucletion Theory of Virus Capsids*, Biophys. J., **90**, 1939-1948, 2006.

[5] M. Aznar, *Coarse-Grained Modeling of the Assembly and Mechanical Properties of Viruses*, PhD Thesis, Universitat de Barcelona, 2013.

[6] M. Aznar, D. Reguera, *Physical Ingredients Controlling Stability and Structural Selection of Empty Viral Capsids*, J. Phys. Chem. B, **120**, 6147-6159, 2016.

[7] J. Wedekind, R. Strey, D. Reguera, *New method to analyze simulations of activated processes*, J. Chem. Phys., **126**, 134103, 2007.

[8] J. Wedekind, D. Reguera, *Kinetic Reconstruction of the Free-Energy Landscape*, J. Phys. Chem. B, **112**, 11060-11063, 2008.

[9] J. Wedekind, G. Chkonia, J. Wölk, R. Strey, D. Reguera, *Crossover from nucleation to spinodal decomposition in a condensing vapor*, J. Chem. Phys. , **131**, 114506, 2009.

[10] A. Luque, D. Reguera, A. Morozov, J. Rudnick, R. Bruinsma, *Physics of shell assembly : Line tension, hole implosion, and closure catastrophe*, J. Chem. Phys. , **136**, 184507, 2016.