

# SCIENTIFIC REPORTS



OPEN

## Rule-based modelling provides an extendable framework for comparing candidate mechanisms underpinning clathrin polymerisation

Anatoly Sorokin<sup>1,2</sup>, Katharina F. Heil<sup>3</sup>, J. Douglas Armstrong<sup>3</sup> & Oksana Sorokina<sup>3</sup>

Polymerisation of clathrin is a key process that underlies clathrin-mediated endocytosis. Clathrin-coated vesicles are responsible for cell internalization of external substances required for normal homeostasis and life-sustaining activity. There are several hypotheses describing formation of closed clathrin structures. According to one of the proposed mechanisms cage formation may start from a flat lattice buildup on the cellular membrane, which is later transformed into a curved structure. Creation of the curved surface requires rearrangement of the lattice, induced by additional molecular mechanisms. Different potential mechanisms require a modeling framework that can be easily modified to compare between them. We created an extendable rule-based model that describes polymerisation of clathrin molecules and various scenarios of cage formation. Using Global Sensitivity Analysis (GSA) we obtained parameter sets describing clathrin pentagon closure and the emergence/production and closure of large-size clathrin cages/vesicles. We were able to demonstrate that the model can reproduce budding of the clathrin cage from an initial flat array.

Clathrin is the major protein component of clathrin-mediated endocytosis (CME)<sup>1,2</sup>. Due to its particular shape and (auto-) polymerization capacity, clathrin is believed to induce the cell membrane to adopt a vesicular shape. A range of different mechanisms have been proposed for this process<sup>3-5</sup>, from a few minimalistic ones propose that clathrin polymerization alone is sufficient to generate buds in a planar membrane<sup>6</sup> to the consensus that describe the orchestrated action of additional proteins and signaling cascades on the intracellular side of the membrane, so that ~30 proteins directly participate in the various steps of endocytosis<sup>1,7-9</sup>.

The structural properties of clathrin have been extensively investigated with respect to their role in vesicle formation. Usually a clathrin molecule is composed of one heavy (~190 kDa) as well as one light chain (~25 kD) and is about 475 Ångström (Å) in length<sup>10</sup>. Within the cell clathrin exists in a form of trimers (triskelia), consisting of three clathrin molecules (three heavy and three light chains respectively), where individual clathrin monomers are referred to as “legs”. Deviating from the normal 1:1 ratio between light and heavy chain several studies have also revealed the existence of triskelia with fewer light chains. Triskelia formation itself does not seem to be influenced by a loss of light chain molecules<sup>11</sup>, but regulatory control of vesicle formation and cargo selection have been proposed.

Due to its internal trimeric structure every single clathrin molecule in the triskelia complex can polymerize with another clathrin molecule from a different clathrin triskelia. Hence every triskelia is able to undergo interactions with three further triskelia. This leads to the formation of dimers and trimers, which can grow to construct large polymers. However, in a normal biological context, hexagonal and pentagonal shapes are among the most frequently observed<sup>12,13</sup>. Specific combinations of these shapes induce the formation of the typical vesicle closed spherical structure. Normally, closed structures contain 12 pentagonal faces and  $(N-20)/2$  hexagonal faces. The

<sup>1</sup>Institute of Cell Biophysics, RAS, Pushchino, 142290, Russia. <sup>2</sup>Moscow Institute of Physics and Technology, 141700, Dolgoprudnyi, Moscow Region, Russia. <sup>3</sup>School of Informatics, University of Edinburgh, Edinburgh, EH8 9AB, United Kingdom. Anatoly Sorokin and Katharina F. Heil contributed equally to this work. Correspondence and requests for materials should be addressed to O.S. (email: [Oksana.Sorokina@ed.ac.uk](mailto:Oksana.Sorokina@ed.ac.uk))

fixed relative numbers between pentagonal and hexagonal faces are based on geometric constraints, given the clathrin structure and minimal flexibility of the trimer legs. Based on the number (N) of triskelia different sphere sizes can emerge, three of which are well defined: The mini-coat, hexagonal-barrel and soccer ball<sup>13</sup>.

Since its discovery in 1975<sup>14</sup>, significant attention has been focused on the mechanism of clathrin polymerisation. It was highlighted in<sup>1</sup> that understanding CME is not possible without proper knowledge of its key process, the clathrin cage formation. Although it was experimentally shown that clathrin self-assembles following pH decrease from 8 to 6.5<sup>15</sup> or under bivalent cation administration<sup>16</sup>, to obtain biologically realistic vesicle shapes the participation of external regulatory proteins is likely critical<sup>1</sup>.

A range of computational models for clathrin self-assembly exists that describes the formation of clathrin cages<sup>12,13,17–19</sup>, or pits and vesicles<sup>13,15,20</sup>. Early models considered the association of 3-valent polymers with equi-reactive binding sites from the Flory's theory point of view with<sup>20</sup> or without<sup>21</sup> allowance for intramolecular loop formation. These studies dissected the dependence of the solution/gel phase transition linked to the critical concentration of the monomer on the equilibrium constants of different steps of the polymerisation process. In the early theoretical models of multivalent condensation, the term “gel” was used to describe the situation when the majority of agents participate in one global complex. There are two phases in such system: a solution consisting of many small complexes and monomers, and a gel, composed of one global complex and a few free monomers. The formation of the global complex is a key phase transition in the systems dynamics. Prior to gel formation, the dynamics of the system are driven by bi-molecular reactions (when two complexes form a bigger one, or a monomer attaches to the complex). After gel formation, the dynamics are driven by uni-molecular reactions within the complex. The key finding of Falk and Thomas<sup>20</sup> is that before the transition to the gel phase, uni-molecular reactions are negligible.

In particular, it was shown by Pastan and Willingham<sup>15</sup>, that the critical concentration of clathrin, sufficient for the phase transition was 30 mg/ml. Taking into account that the triskelia molecular mass is about 640 kDa, this value corresponds to the molar concentration of 46  $\mu\text{M}$ , or approximately 55000 triskelia per eukaryotic cell.

More recent studies examined the assembly of 5- and 6- member rings in parallel with investigation of how different physical triskelia characteristics might impact on cage formation. These characteristics include triskelia rigidity<sup>21</sup>, their asymmetry<sup>17</sup>, emergent tension during cage closure<sup>22</sup> and the effects of superficial membrane tension<sup>23</sup>. These studies provide approximations of binding energy between the chains of the neighbouring clathrin triskelia<sup>17</sup>.

The polymerisation process alone presents a significant challenge for mechanistic modeling, as the number of molecular species, which have to be described, grows exponentially with the number of available monomers. Rule-based modeling<sup>24–26</sup> provides a viable solution allowing a network-free simulation technique<sup>27–29</sup>. It uses ‘lumped’ reaction rules to concisely represent molecule interactions. One can assume the rules as implicit combinations of different reactions into classes, where all the members of the same class perform a common transformation. This modeling approach is generally exploited for large-scale biochemical systems to overcome combinatorial complexity and it has previously demonstrated its effectiveness in simulations of ligand-receptor complex polymerization<sup>25</sup>.

Here we present a suite of rule-based models of clathrin polymerisation with increasing complexity, starting from a very basic model where the molecule has three equally reactive binding sites to a more advanced model reproducing realistic triskelia clathrin structure. We examined the correspondence of each model's behavior with the existing theoretical models while sampling from a wide range of parameter values.

We found that although the basic model exactly reproduces Flory's findings, it is unable to provide the amounts of 5- and 6- member rings required for cage formation and, therefore, it fails to reproduce clathrin vesicle formation. A revised model with a more realistic clathrin structure that explicitly supports predominant closure of pentagons and hexagons allows 3D cage formation and permits the evolution of flat 2D clathrin patches into a 3D cage structures by shifting the ratio of the pentagon/hexagon dissociation constants.

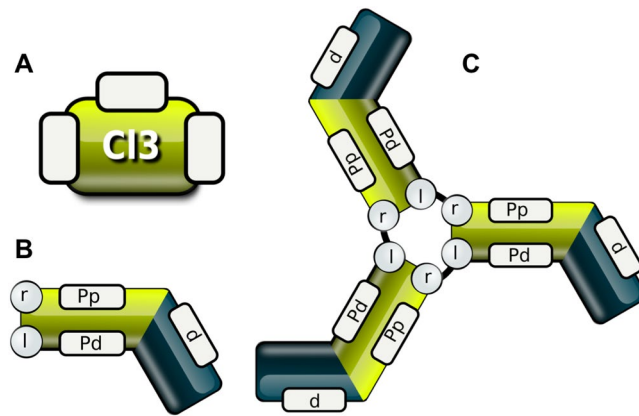
## Methods

**Models and simulation.** We used the Kappa language<sup>30</sup>, a member of the family of rule-based modeling languages, for building the models. All models were simulated by KaSim3.5 (<http://dev.executableknowledge.org/>). We used Kappa extensions where appropriate, e.g the MetaKappa (<https://github.com/kappamodeler/metakappa>) extension for building the first model to handle the combinatorial explosion caused by three equal binding sites (see Appendix for details). Also, we use the RKappa extension<sup>31</sup> for sampling the large parameter space, statistical analysis of simulation results, global sensitivity analysis (GSA) and visualization of the Kappa molecular structures as more comprehensible 2D and 3D graphs.

We first investigated the capability of rule-based models to reproduce clathrin cage structures based on random self-assembly processes. For this we assume that clathrin triskelia interact in 3D, in a well-mixed solution and all binding sites of the clathrin triskelia are assumed to be identical. Due to the combinatorial nature of the clathrin molecule association, the size of aggregates is unbounded and limited only by the amount of available substrate.

We started with a reduced model of triskelia monomers similar to Perelson and Goldstein's equilibrium and continuous model<sup>21</sup>, in which monomers carry three identical equally reactive binding sites. Two variants of this model were implemented in the rule-based Kappa language to investigate the polymerization of branched complexes from a single class of trivalent agents under ‘rings allowed’ and ‘rings forbidden’ conditions similar to that proposed by<sup>20</sup> (Model 1).

We then developed a more elaborate model, based on clathrin monomers, that considers triskelia as a predefined complex of three monomers. This model more accurately reproduces the structure of clathrin with distinct legs and binding sites along with specified defined steric and chirality constraints (Model 2). It also contains explicit rules describing formation of penta- and hexagonal rings and demonstrates the dynamics of closed cage

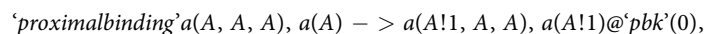


**Figure 1.** Structure of agents for Model 1 and Model 2. Three identical binding sites in a simple agent (A) interact with each other to form a lattice (Model1). Monomer (B) of detailed Model 2 has two sites to form the triskelia hub (l,r) and three sites to interact with other triskelia (d, Pd, Pp) (C).

structure formation. All the models presented here are kinetic and do not include notions of space. However these could be added by use of existing extensions like SpatialKappa<sup>26</sup> or Geometric Kappa<sup>32</sup> if required later.

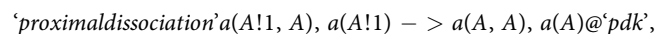
**Equireactive trivalent agent model.** In the first model (Model 1) we simplify the realistic triskelia structure of clathrin to the trivalent agent Cl3 with three identical binding sites. This is effectively a kinetic version of the model described by Perelson and Goldstein in 1985<sup>16,21</sup> (Fig. 1A, Supplementary Data). As clathrin is known to aggregate on the membrane, we assume that with complex growth its ability to diffuse would decrease. Thus, in our configuration complex growth happens preferentially via addition of new monomers rather than merging of existing complexes, in the same way as in Perelson and Goldstein.

The (kappa) rule looks as follows:



where 'pbk' is the rate of binding.

To ensure stability of the rings in clathrin complexes we make an assumption that molecules with three occupied binding sites cannot dissociate. Thus, dissociation is only possible at the periphery of the complex when at least one binding site is/remains free.



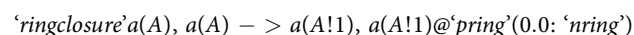
where 'pdk' is the rate of dissociation.

This rule partially contradicts the work of Perelson and Goldstein, where the dissociation is possible only at the monomer level. However, the rule includes the dissociation of terminal monomers as a special case.

We studied the random polymerisation of trivalent monomers under two traditional Flory-Stockmayer assumptions: 'ring forbidden' (Model 1 A) and 'rings allowed' (Model 1B).

In the case of Model 1 A ('rings forbidden'), the intramolecular bonds between the binding sites of the same polymers are not allowed as the only free agent (with all three sites non-occupied) can bind the polymer. The detailed models for the original Perelson's model and its two Kappa implementations: Model 1 A and Model 1B are presented in Supplementary Data.

In the case of Model 1B ('ring allowed') intra-molecular reactions are allowed, so that rings of different sizes may occur. As in<sup>20,21,33</sup> reactions occur with an equal probability for each of the free binding site to react until the reaction extent  $R_{\text{ext}} = 1$ , which means that all binding sites are fully occupied. Although cubical structures of clathrin were observed experimentally under special conditions<sup>34</sup>, the formation of rings of size 4 and less is not reported under conditions approximating intracellular environments. Hence we set a specific constraint on the polymer chain ability to make intramolecular bonds only when ring size (nring) exceeds 5 bonds in length.



In the rule above 'pring' is a rate of ring closure, while 'nring' refers to the minimal number of bonds in the ring (set to 5 in this case). The constraints enforce limitations on the condition of equal reactivity to be always fulfilled; yet the probability to close a short ring within a large complex is quite small. We also assume the equilibrium constants for initiation, elongation and branching are equal.

**Triskelia model.** To generate a more realistic model we next considered clathrin monomers and their structural properties. Each monomer consists of a proximal region ("P", light green in Fig. 1B), which contains a binding domain on its "right", long part ("r") and "left", short part ("l"), and the distal region ("d", dark green in Fig. 1B). Domains in the proximal region facilitate the internal binding of monomers to form trimers. The additional binding sites "Pp" and "Pd" in the proximal region allow binding amongst different triskelia. Binding rules

Structure	Max number
g531	2
g532	4
g541	16
g551	250
g643	2
g651	4
g661	36

**Table 1.** Ranges for parameter space exploration.

presume the ‘right’ part of one monomer can only bind to the “left” part of another, and so forth to make correct triskelia structures (Fig. 1C).

In kappa language this is expressed in the following way:

$$Cl(l!1, r!2), Cl(r!1, l!3), Cl(r!3, l!2)$$

“Cl” refers to a single clathrin molecule with proximal right (“r”) and left (“l”) binding site. All distal parts of the long legs are oriented in one direction, showing a clockwise drift/turn (Fig. 1C).

Once assembled, triskelia form the structural unit for the polymerisation process, which is governed by the interaction of domains localised on the right, long leg of each monomer. These are: a proximal (Pp), a distal “receiving” (Pd) and distal “giving” (d) domain. Based on the given clathrin triskelia structure, formation of one bond utilizes four triskelia simultaneously: two monomers bind with their proximal parts, and two form additional bonds with their distal parts (see Supplementary Data for triskelia binding code and a visualization). As was shown by den Otter *et al.*<sup>17</sup> and Fotin *et al.*<sup>35</sup>, the proper orientation of all four legs is vital for formation of closed structures. Initial polymerization steps along with the model rules are presented in detail in Supplementary Data.

In addition to the binding rule, a few specific rules enforce the closure/formation of pentagons and hexagons. Dissociation is implemented as follows. Closed rings cannot be reopened. At least one monomer needs to be unbound for dissociation to happen. Details can be seen in the model code in the Supplementary Data, which shows the rules used in the current model version.

**Data Availability.** All data generated or analysed during this study are included in this published article (and its Supplementary Information files).

## Results

We investigated the ability of rule-based models to reproduce the clathrin cage structures based on a random self-assembly process. Specifically, two traditional Flory-Stockmayer conditions: “rings forbidden” and “rings allowed” were applied separately, similar to<sup>20</sup>. All models were simulated 5000 times with parameter ranges shown in Table 1.

**Trivalent model.** In the first model we used a simplified triskelia structure of clathrin with a trivalent agent Cl3 containing three identical binding sites with equal reactivity, similar to the Perelson and Goldstain model in 1985<sup>21</sup>.

The key parameters that have been analyzed are (see also<sup>21</sup>):

$$R_{ext} = \frac{2 * N_{bond}}{3 * amount} \quad (1)$$

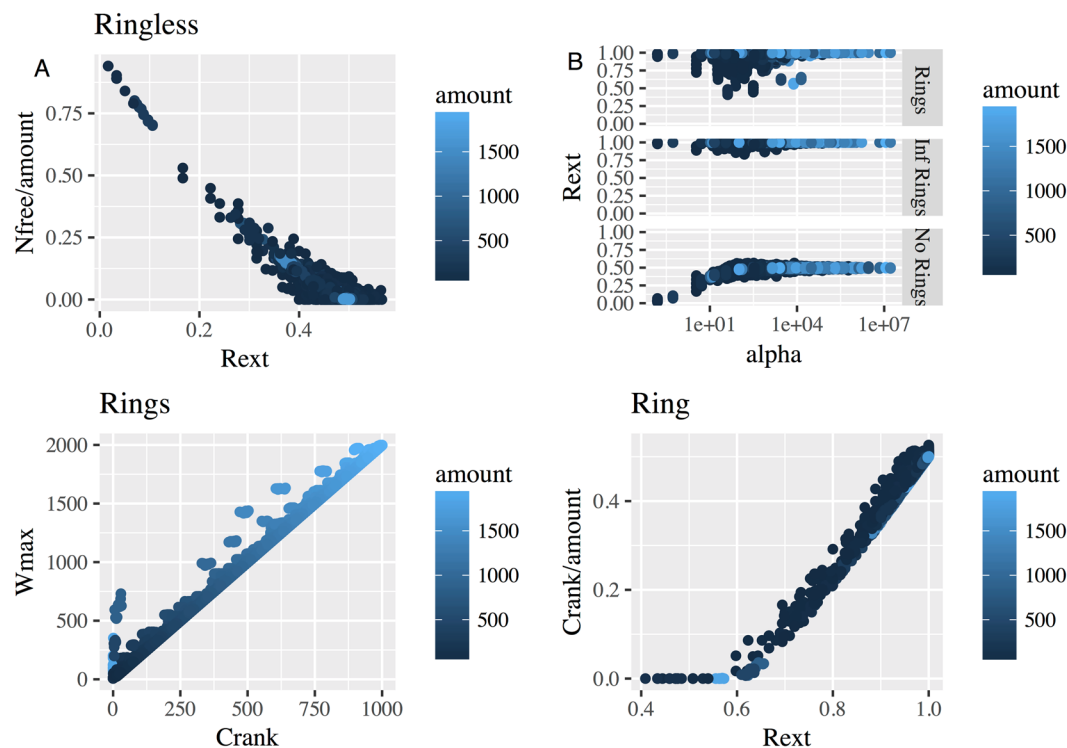
$$\alpha = 6 * K * C_r = 6 * \frac{pbk * (N_A * V)}{pdk} = 6 * N_r * \frac{pbk}{pdk} \quad (2)$$

where  $R_{ext}$  - reaction extent,  $\alpha$  - nondimensional equilibrium constant,  $N_{bond}$  - the number of bonds in the polymer, and  $K$  - the equilibrium constant.  $C_t$  and  $N_t$  describe the total concentration and total number of monomers (respectively),  $amount$  - amount of available triskelia.

We showed that in the “ring forbidden” setup, the distribution of free clathrin with dependence on  $R_{ext}$  exactly followed the prediction of Perelson’s theory (Fig. 2A and B). The vast majority of parameter sets in “ring forbidden” are grouped around  $R_{ext} = 0.5$ , and the dependency between  $R_{ext}$  and  $N_{bond}/N_t$  is linear. We found that  $R_{ext}$  never exceeded the theoretical limit of gel formation (Fig. 2B) while in most of the “ring allowed” instances, reactions stopped only when the available binding sites were saturated (Fig. 2B and D).

To explore the types of complexes our simulations produced, we calculated the size of the largest aggregate ( $W_{max}$ ) and the number of the rings in the system. The latter was estimated as the cyclomatic number of the clathrin graph, which is the number of bonds that need to be removed to form an acyclic graph:

$$C_{rank} = E_g - V_g + C_g \quad (3)$$



**Figure 2.** Simulation (5000 parameter sets) of the trivalent model with “ring forbidden” (A) and “ring allowed” (C,D) assumptions. (A) The number of free agents  $N_{\text{free}}$  decreases with  $R_{\text{ext}}$  and trends to 0 at  $R_{\text{ext}} = 0.5$ . (B) The relationship of  $\alpha$  and  $R_{\text{ext}}$  under different experimental conditions: “no ring”, “ring” and “infinite ring”. (C) The dependency between the size of the largest aggregate and cyclomatic number under “ring allowed” condition. (D) Relationship between reaction extent and loop structure under “ring allowed”.

with  $E_g$  number of edges and  $V_g$  number of nodes in the graph.  $C_g$  is the number of connected components in the graph. We found that the number of rings in the system  $C_{\text{rank}}$  almost always reached the theoretical limit (Fig. 2D), where the total number of monomers was equal to the size of the largest aggregate ( $W_{\text{max}}$ ) in agreement with analysis from Falk *et al.*<sup>20</sup>

In agreement with<sup>20</sup>, when intramolecular bonds are allowed (Model 1b) ring formation only starts after gel structure formation (Fig. 2D), when the reaction extent reaches the 0.5 threshold. This means that in the simple agent model closed cages would be formed only when 7/8 of the available clathrins form a large single complex.

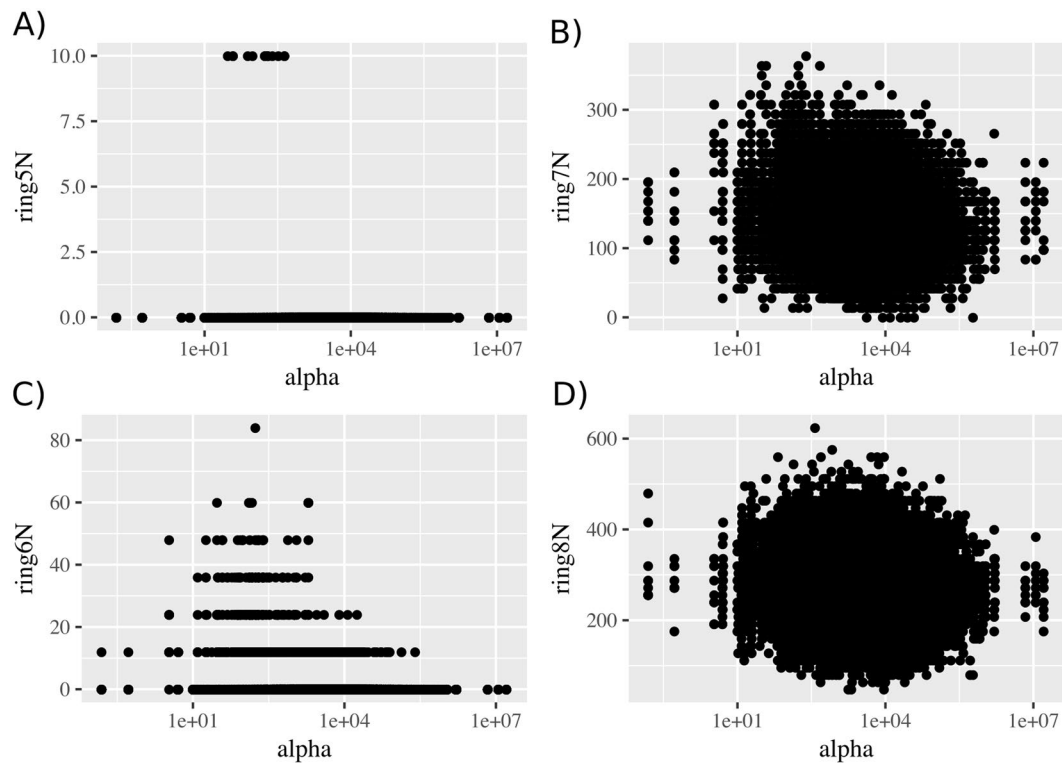
Further analysis (Supplementary Data) shows that probability of the ring closure grows with the size of the ring. Therefore, the number of short rings (pentagons and hexagons) is quite low even when we set the rate of the ring closure reactions to infinity (Supplementary Data and Fig. 3). Therefore we conclude that the simple model is not able to describe the closed cage structures, as the clathrin geometry provides the optimal mutual disposition of the monomers only when 5- and 6-membered rings are formed. To resolve this we developed a more plausible model as follows.

**Triskelia model.** Model 2 described above corresponds to a more realistic structure of clathrin with distinct regions within the monomer and respective binding sites that reflect the experimental literature<sup>10,12</sup>. We also introduced a specific rule for orientation of the monomers to ensure that the “right” site of one monomer binds the “left” side of another. This preserves the correct geometry of triskelia and chirality of the monomers. To ensure we obtain realistic clathrin complexes, 5- and 6- ring closure reactions were explicitly specified.

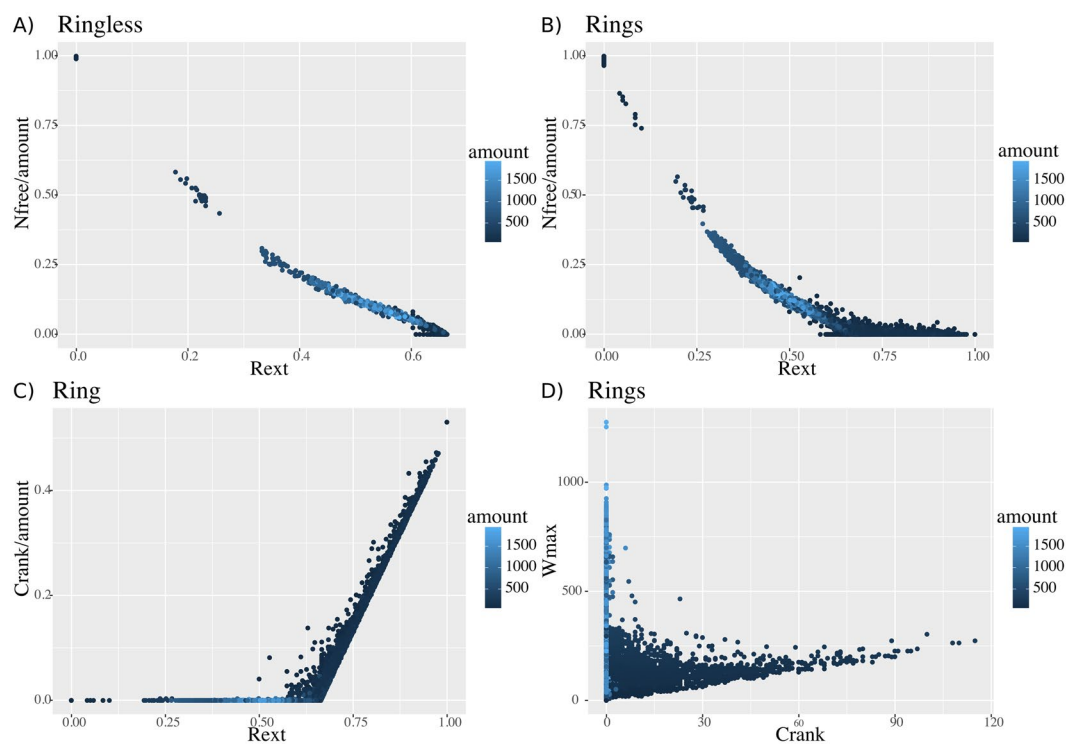
We started with parameter sampling for the model. To ensure comparability between simulations we used the same parameter sets as before by assigning the ring closure rate the value of “pring” to both hexagons and pentagons. Again, the two cases - “ring forbidden” and “ring allowed” were investigated.

The behavior of the “ring forbidden” version of Model 2 is clearly similar to the behavior of the Model 1 and theoretical predictions of Perelson (Fig. 4A). The number of free triskelia monotonically decreases towards zero at  $R_{\text{ext}} = 0.6$ . The difference between the theoretical prediction of 0.5 and the observed value is explained by the association rule in the Model 2, which does not prevent associations of clusters and therefore does not follow the monomer attachment mechanisms considered in Perelson<sup>21</sup>. Association between clusters results in a higher numbers of triskelia with all their legs involved in the complex formation, which in turn prevents their dissociation.

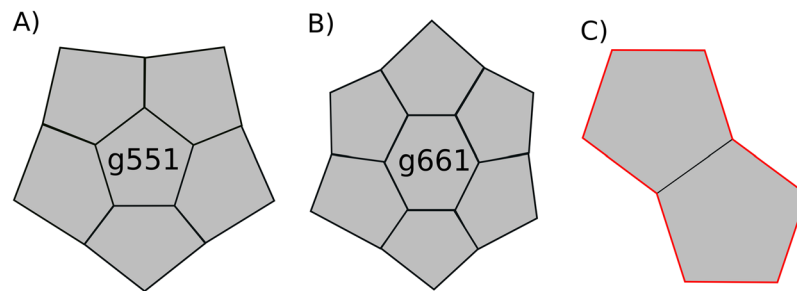
The “ring allowed” version of the Model 2 (Fig. 4B) follows the same scenario as Model 1 (Fig. 2D) and behaves as predicted by theory<sup>20</sup>. Ring formation starts only after solution to gel transition at  $R_{\text{ext}} = 0.6$ . Contrary to the Model 1, the number of rings does not grow linearly with the size of the complex (Fig. 2C). Instead, due to



**Figure 3.** Distribution of different cyclic structures obtained from 5000 simulations. (A) 5-membered rings. (B) 6-membered rings. (C) 7-membered rings. (D) 8-membered rings.



**Figure 4.** Results of simulation of “ring forbidden” and “ring allowed” models. (A) The number of free agents  $N_{free}$  decreases with  $R_{ext}$  towards 0 at  $R_{ext} = 0.6$  for “ring forbidden” model. (B) The number of rings (cyclomatic number of the graph) per triskelia in the “ring allowed” model.



**Figure 5.** Most populated structures obtained in 5000 simulations of the unconstrained model. **(A)** Most populated pentagon structure. **(B)** Most populated hexagon structure. **(C)** An 8-ring formed by two pentagons.

Parameter description	Parameter name	min	max
Association rate constant	<i>pbk</i>	10 E-6	1.00
Dissociation rate constant	<i>pdk</i>	10 E-6	1.00
Ring closure rate constant	<i>pring</i>	10 E-6	1.00
Amount of available triskelia	<i>amount</i>	10 E2	10 E4

**Table 2.** Number of pent-Rings (g5) and hex-Rings (g6) found in GSA of the “ring allowed” version of the Model 2.

the system not being allowed to form rings of arbitrary size, we obtain many small complexes with few or no rings (Supplementary Figure 2).

For simplicity of simulation and comparison with Model 1 we did not introduce separate kinetic constants for 5- and 6-ring closure. As a result, the vast majority of the rings in our simulations are pentagons. Nevertheless we observed a number of hexagons as well. The relatively high number of octagons observed is a consequence of high number of 5-rings, as hull of two adjacent pentagons can form an octagon (Fig. 5C).

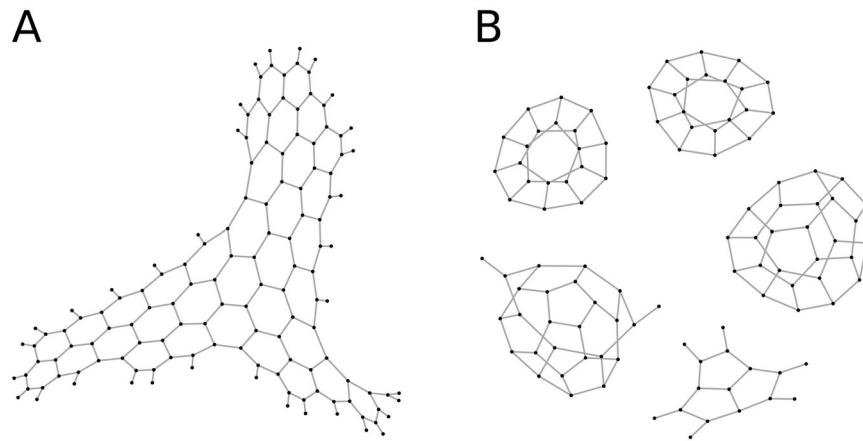
To explore the geometry of complexes, which contain 5- and 6-rings we used a set of all possible combinations of pentagons and hexagons as described in<sup>22</sup>. Table 2 shows that pentagons tend to form adjacent dodecahedron-like structures (see g551, Fig. 5A, Supplementary Figure 5), while hexagons are most often surrounded by pentagons as visualized in structure g661 (Fig. 5B). We found no clear distinction between ring forming and ring preventing values in parameter sets (Supplementary Figure 3). To further investigate which parameters influence the ring formation the most we performed GSA on Model 2 with the “ring allowed” condition (Supplementary Table 1). We thus concluded that Model 2 is able to produce various structures of different shapes (Fig. 5) without the initial constraints, but that they do not all necessarily end up being cage-like structures.

The type of clathrin cage formed *in vivo* is known to depend on the ratio of pentagons and hexagons<sup>3,22</sup>. Moreover, planar clathrin consists of just hexagons. As an example we tested to see whether our model could be reconciled with the invagination mechanism (e.g. described in Avinoam *et al.*<sup>5</sup>). Avinoam’s (2015) mechanism requires the presence of pentagons. To reproduce this we tuned the rate constants for pentagon and hexagon closure and changed the equilibrium of association and dissociation rates for them. First we simulated the model where only 6-rings were allowed by setting the 5-ring closure reaction to 0 to form a planar structure (Fig. 6A and Supplementary movie 1). When the reaction extent was close to 1, 5-ring closure was allowed by adjusting rate constant to non-zero value. With a rate of closure for 5- and 6-ring close to each other we observed invaginations, but they never reached the scissoring stage so that the completely closed structure never occurred (Supplementary movie 2). At this point we set the rate of closure for 5-rings to infinite and after 10<sup>4</sup> events we obtained the structures shown in Fig. 6B and Supplementary Movie 3.

To evaluate the influence of rates of pentagon and hexagon closure/disruption we performed GSA on the model starting with a flat hexagonal mesh (Supplementary Table 2). Here, *b* and *d* are the coefficients defining the extent to which pentagon closure is faster than hexagon closure (*b*), and hexagon compared to pentagon dissociation (*d*); *rng5* and *rng6* are the ratios of ring closure to ring disruption for pentagons and hexagons, respectively. For each parameter the significance level is calculated as described in<sup>36</sup>. The rate of the pentagon closure did not significantly influence any property of the system, while the rate of hexagon dissociation appeared important for the size of the most frequent complex (*wNmax*) and the presence of hexagon-containing subgraphs (g501, g511, g521, g522, g601, g611, g621, g622, g631, g632, g633, g641, g642, g643 in Supplementary Tables 1 and 2)<sup>22</sup>. During the course of a simulation we were able to obtain different numbers of closed cages in almost half of the parameter sets, which indicates that the formation of flat structures requires additional constraints, while cage formation happens spontaneously<sup>4</sup>.

## Discussion

Computational models describing formation of clathrin-coated vesicles (CCVs)<sup>2,17,23,37</sup> mostly focus on clathrin self-association or its association with the membrane. However, vesicle recycling is regulated via a large number



**Figure 6.** Results of model simulation with different  $K_d$  for 5 and 6-membered rings. **(A)** Only hexagons are allowed. **(B)** Rate closure for 5-rings was set to infinite.

of signalling processes<sup>2,38</sup>. Existing computational models struggle to incorporate these regulatory elements either because of high computational cost, which becomes prohibitive in case of incorporation of all involved protein types, or because the structure/type of the model can/does not include the reactions controlled by regulatory systems. For example, the equilibrium model<sup>21</sup> considered growth of pits as a linear set of reactions, assuming that all three legs of the new triskelia in the pit assemble using the best possible free sites in the net. As shown by simulations in<sup>17</sup> and confirmed in our Models 1 A and B this is not the case.

As was proposed in<sup>39</sup>, these signaling processes can be incorporated into models as a modification of clathrin association/dissociation rates. With these factors in mind we have developed a model capable of describing the formation of CCVs, avoiding the more resource expensive computational algorithms and using a modeling format familiar to the signal transduction modelling community.

Our first version of the model, which described clathrin as a trivalent agent demonstrated that formation of closed structures required an additional manual closure to achieve 5- and 6- rings. With the flexibility of the clathrin molecule and no evidence for energy differences between penta- and hexameric rings we saw no preferences towards either specific ring composition. Weak interactions, which have been proposed to have a major effect on the association of clathrin legs<sup>39</sup>, and comparatively low bending energy of the clathrin lattice suggest that when on the flat part of the membrane, clathrin will create a flat hexagonal lattice. That process was considered in<sup>3</sup>, where clathrin was modeled as hexagonal lattice with 5- and 7-sided rings occurring as defects, but the study only considered the equilibrium state, whereas in our analysis we were able to investigate the kinetics of the process. Although the “canonic” mechanism of clathrin pits formation proposes constant curvature growth as a function of clathrin polymerization<sup>40</sup>, the evolution of curved clathrin structures from flat plaque has also some supporting experimental evidence<sup>5,38</sup>. The recent study of Leyton-Puig *et al.*<sup>7</sup> reports the ability of clathrin plaques to act as hubs for CME and proposes actin polymerisation and actin-based adhesion are major regulating factors for their remodeling<sup>7</sup>.

Our model shows that switching pentagon ring formation on/off allows the process to switch between planar patches and closed cages. *In vivo*, this switching could be driven by changes in physical properties of the membrane or by additional regulatory mechanisms<sup>1,37,41</sup>.

In our model we assume the size and the shape of the clathrin lattice to be controlled by three processes: i) the association/dissociation of triskelia; ii) the 5-ring formation/dissociation and iii) the 6-ring formation/dissociation. Several other factors are known to influence the cage and coat formation and dissociation<sup>42,43</sup>. For example in<sup>44</sup>, the main difference in pentagon and hexagon closure is attributed to the stiffness of the underlying membrane, while in<sup>41</sup> the rigidity variation of the clathrin net itself is explained by binding to an adaptor protein (AP2, AP3, AP180)<sup>8,45</sup>. Their influence on clathrin coat formation has been studied in distinct experimental setups and binding to clathrin has been confirmed. Due to their influence on clathrin triskelia structure and hence their ability to influence coat formation it might be debatable if their main role is in maintaining a flat structure or “forcing”/inducing the formation of vesicles. This mechanism could be easily embedded into the model (see the example in Supplementary Data).

The clathrin light chain is an additional part of the triskelia, which connects to the heavy chain in the region extending from the self-association domain to the knee<sup>39</sup>. One of the possible conformations can force the knee to bend in a direction that inhibits cage formation. This inhibitory effect is thought to be regulated (inhibited) by interaction with Ca ions or by lowering the pH<sup>39</sup>. The light chain also influences the rigidity of the clathrin lattice and its ability to bend the lipid membrane at low temperature<sup>4</sup>. The light chain contains 19 serines that are potential kinase targets (GRK2) and phosphorylation of the light chain has been proposed as a discriminator for different cargo inclusion in the vesicle<sup>46</sup>. An example of how the model can be extended to incorporate the above mechanism is presented in Supplementary Data.

The rule-based approach we have used allows us to build and compare kinetic models that describe different possible mechanisms of clathrin cage formation, from direct assembly from monomers at the vesicle budding



site to the invagination of flat membrane plaque. More in depth functional details such as the role of N-WASP through Arp2/3<sup>7</sup> can help to expand models and gain deeper insights. Hence, our implementation is easily extendable allowing the future inclusion of more detailed mechanistic models of CME regulation.

## References

- McMahon, H. T. & Boucrot, E. Molecular mechanism and physiological functions of clathrin-mediated endocytosis. *Nat. Rev. Mol. Cell Biol.* **12**, 517–533 (2011).
- Jung, N. & Haucke, V. Clathrin-mediated endocytosis at synapses. *Traffic* **8**, 1129–1136 (2007).
- Jin, A. J. & Nossal, R. Topological mechanisms involved in the formation of clathrin-coated vesicles. *Biophys. J.* **65**, 1523–1537 (1993).
- Dannhauser, P. N. *et al.* Effect of clathrin light chains on the stiffness of clathrin lattices and membrane budding. *Traffic* **16**, 519–533 (2015).
- Avinou, O., Schorb, M., Beese, C. J., Briggs, J. A. G. & Kaksonen, M. Endocytic sites mature by continuous bending and remodeling of the clathrin coat. *Science* **348**, 1369–1372 (2015).
- Dannhauser, P. N. & Ungewickell, E. J. Reconstitution of clathrin-coated bud and vesicle formation with minimal components. *Nat. Cell Biol.* **14**, 634–639 (2012).
- Leyton-Puig, D. *et al.* Flat clathrin lattices are dynamic actin-controlled hubs for clathrin-mediated endocytosis and signalling of specific receptors. *Nat. Commun.* **8**, 16068 (2017).
- Smith, S. M., Baker, M., Halebian, M. & Smith, C. J. Weak Molecular Interactions in Clathrin-Mediated Endocytosis. *Front Mol Biosci* **4**, 72 (2017).
- Saheki, Y. & De Camilli, P. Synaptic vesicle endocytosis. *Cold Spring Harb. Perspect. Biol.* **4**, a005645 (2012).
- Fotin, A. *et al.* Structure of an auxilin-bound clathrin coat and its implications for the mechanism of uncoating. *Nature* **432**, 649–653 (2004).
- Girard, M., Allaire, P. D., McPherson, P. S. & Blondeau, F. Non-stoichiometric relationship between clathrin heavy and light chains revealed by quantitative comparative proteomics of clathrin-coated vesicles from brain and liver. *Mol. Cell. Proteomics* **4**, 1145–1154 (2005).
- Kirchhausen, T., Owen, D. & Harrison, S. C. Molecular structure, function, and dynamics of clathrin-mediated membrane traffic. *Cold Spring Harb. Perspect. Biol.* **6**, a016725 (2014).
- Fotin, A. *et al.* Structure determination of clathrin coats to subnanometer resolution by single particle cryo-electron microscopy. *J. Struct. Biol.* **156**, 453–460 (2006).
- Pearse, B. M. Coated vesicles from pig brain: purification and biochemical characterization. *J. Mol. Biol.* **97**, 93–98 (1975).
- Pastan, I. & Willingham, M. C. The pathway of endocytosis. *J. Mol. Biol.* **97**, 1–44 (1985).
- Pearse, B. M. & Crowther, R. A. Structure and assembly of coated vesicles. *Annu. Rev. Biophys. Biophys. Chem.* **16**, 49–68 (1987).
- den Otter, W. K., Renes, M. R. & Briels, W. J. Asymmetry as the key to clathrin cage assembly. *Biophys. J.* **99**, 1231–1238 (2010).
- den Otter, W. K. & Briels, W. J. The generation of curved clathrin coats from flat plaques. *Traffic* **12**, 1407–1416 (2011).
- Matthews, R. & Likos, C. N. Structures and pathways for clathrin self-assembly in the bulk and on membranes. *Soft Matter* **9**, 5794–5806 (2013).
- Falk, M. & Thomas, R. E. Molecular size distribution in random polyfunctional condensation with or without ring formation: computer simulation. *Can. J. Chem.* **52**, 3285–3295 (1974).
- Perelson, A. S. & Goldstein, B. The equilibrium aggregate size distribution of self-associating trivalent molecules. *Macromolecules* **18**, 1588–1597 (1985).
- Schein, S. & Sands-Kidner, M. A geometric principle may guide self-assembly of fullerene cages from clathrin triskelia and from carbon atoms. *Biophys. J.* **94**, 958–976 (2008).
- Banerjee, A., Berezhevskii, A. & Nossal, R. Stochastic model of clathrin-coated pit assembly. *Biophys. J.* **102**, 2725–2730 (2012).
- Danos, V., Feret, J., Fontana, W., Harmer, R. & Krivine, J. Rule-based modelling and model perturbation. *Transactions on Computational Systems Biology XI* 116–137 (2009).
- Monine, M. I., Posner, R. G., Savage, P. B., Faeder, J. R. & Hlavacek, W. S. Modeling multivalent ligand-receptor interactions with steric constraints on configurations of cell-surface receptor aggregates. *Biophys. J.* **98**, 48–56 (2010).
- Sorokina, O., Sorokin, A., Armstrong, J. D. & Danos, V. A simulator for spatially extended kappa models. *Bioinformatics* **29**, 3105 (2013).
- Danos, V., Feret, J., Fontana, W. & Krivine, J. Scalable simulation of cellular signaling networks. *Computational Methods In Systems Biology, Proceedings* 139–157 (2009).
- Colvin, J. *et al.* RuleMonkey: software for stochastic simulation of rule-based models. *BMC Bioinformatics* **11**, 404 (2010).
- Sneddon, M. W., Faeder, J. R. & Emonet, T. Efficient modeling, simulation and coarse-graining of biological complexity with NFsim. *Nat. Methods* **8**, 177–183 (2011).
- Danos, V., Feret, J., Fontana, W. & Krivine, J. Abstract interpretation of cellular signalling networks. *Verification, Model Checking, and Abstract Interpretation* 83–97 (2008).
- Sorokin, A., Sorokina, O. & Armstrong, J. D. RKappa: Statistical sampling suite for Kappa models. in *Hybrid Systems Biology* (eds Maler, O., Halasz, A. & Piazza, C.) 128–142 (Springer, 2015).
- Danos, V., Honorato-Zimmer, R., Jaramillo-Riveri, S. & Stucki, S. Rigid Geometric Constraints for Kappa Models. *Electron. Notes Theor. Comput. Sci.* **313**, 23–46 (2015).
- Goldstein, B. & Perelson, A. S. Equilibrium theory for the clustering of bivalent cell surface receptors by trivalent ligands. *Application to histamine release from basophils. Biophys. J.* **45**, 1109–1123 (1984).
- Sorger, P. K., Crowther, R. A., Finch, J. T. & Pearse, B. M. Clathrin cubes: an extreme variant of the normal cage. *J. Cell Biol.* **103**, 1213–1219 (1986).
- Fotin, A. *et al.* Molecular model for a complete clathrin lattice from electron cryomicroscopy. *Nature* **432**, 573–579 (2004).
- Marino, S., Hogue, I. B., Ray, C. J. & Kirschner, D. E. A methodology for performing global uncertainty and sensitivity analysis in systems biology. *J. Theor. Biol.* **254**, 178–196 (2008).
- Muthukumar, M. & Nossal, R. Micellization model for the polymerization of clathrin baskets. *J. Chem. Phys.* **139**, 121928 (2013).
- Ungewickell, E. J. & Hinrichsen, L. Endocytosis: clathrin-mediated membrane budding. *Curr. Opin. Cell Biol.* **19**, 417–425 (2007).
- Wilbur, J. D. *et al.* Conformation switching of clathrin light chain regulates clathrin lattice assembly. *Dev. Cell* **18**, 854–861 (2010).
- Lampe, M., Vassilopoulos, S. & Merrifield, C. Clathrin coated pits, plaques and adhesion. *J. Struct. Biol.* **196**, 48–56 (2016).
- Nossal, R. Energetics of clathrin basket assembly. *Traffic* **2**, 138–147 (2001).
- Böcking, T., Aguet, F., Harrison, S. C. & Kirchhausen, T. Single-molecule analysis of a molecular disassemblase reveals the mechanism of Hsc70-driven clathrin uncoating. *Nat. Struct. Mol. Biol.* **18**, 295–301 (2011).
- Doherty, G. J. & McMahon, H. T. Mechanisms of endocytosis. *Annu. Rev. Biochem.* **78**, 857–902 (2009).
- Shraiman, B. I. On the role of assembly kinetics in determining the structure of clathrin cages. *Biophys. J.* **72**, 953–957 (1997).
- Saleem, M. *et al.* A balance between membrane elasticity and polymerization energy sets the shape of spherical clathrin coats. *Nat. Commun.* **6**, 6249 (2015).
- Ferreira, F. *et al.* Endocytosis of G protein-coupled receptors is regulated by clathrin light chain phosphorylation. *Curr. Biol.* **22**, 1361–1370 (2012).

## Acknowledgements

This research has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No. 720270 (HBP SGA01).

## Author Contributions

A.S., O.S. model design; A.S., O.S., K.F.H. model simulation, G.S.A., analysis; A.S., K.F.H., J.D.A., writing; O.S. direction and writing.

## Additional Information

**Supplementary information** accompanies this paper at <https://doi.org/10.1038/s41598-018-23829-x>.

**Competing Interests:** The authors declare no competing interests.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2018