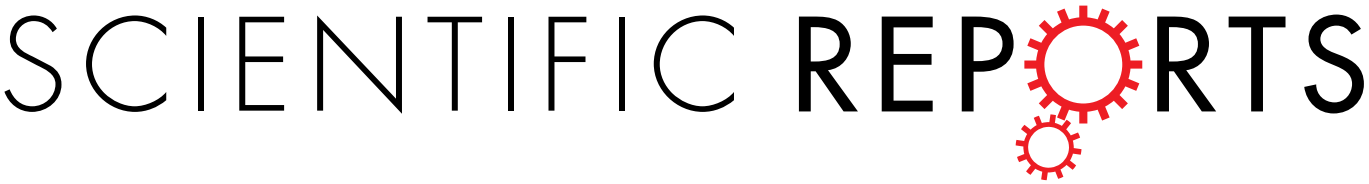
Kinetic constraints on self-assembly into closed supramolecular structures



OPEN

Kinetic constraints on self-assembly into closed supramolecular structures

Received: 11 May 2017
Accepted: 4 August 2017
Published online: 25 September 2017

Thomas C. T. Michaels[1,2], Mathias M. J. Bellaiche[1,3], Michael F. Hagan[4] & Tuomas P. J. Knowles[1,5]

Many biological and synthetic systems exploit self-assembly to generate highly intricate closed supramolecular architectures, ranging from self-assembling cages to viral capsids. The fundamental design principles that control the structural determinants of the resulting assemblies are increasingly well-understood, but much less is known about the kinetics of such assembly phenomena and it remains a key challenge to elucidate how these systems can be engineered to assemble in an efficient manner and avoid kinetic trapping. We show here that simple scaling laws emerge from a set of kinetic equations describing the self-assembly of identical building blocks into closed supramolecular structures and that this scaling behavior provides general rules that determine efficient assembly in these systems. Using this framework, we uncover the existence of a narrow range of parameter space that supports efficient self-assembly and reveal that nature capitalizes on this behavior to direct the reliable assembly of viral capsids on biologically relevant timescales.

The spontaneous formation of nanoscale materials with specific chemical and physical characteristics from basic molecular building blocks is a key process for the functioning of living systems and provides a bottom-up strategy for constructing novel nanomaterials for various applications[1]. A particularly important class of such molecular self-assembly processes is the formation of closed supramolecular structures, with examples including clathrin assemblies[2], self-assembling cages[3,4], micellar-like structures[5], small polyhedra[6–8] or icosahedral viral capsids[9–11]. Many assembly processes of this type underlie key events in normal biology[12], but are also implicated in the onset of diseases of humans, animals and plants. Moreover, the construction of such molecular topologies offers great potential as biomimetic nanocontainers for encapsulation, delivery and release of small molecules.

Elegant physical principles have emerged that determine the geometric and equilibrium constraints governing the shapes of the resulting assembly structures in these systems, motivating the question of whether or not analogous principles can be defined for their assembly kinetics. Probing molecular reaction mechanisms in complex systems represents a fundamental challenge through the Chemical Sciences; in this context, chemical kinetics has proven to be an extremely effective tool for testing mechanistic hypothesis in areas ranging from small molecule chemistry to enzyme kinetics. Recent advances have extended the applicability of this chemical kinetics approach to the study of filamentous protein assembly phenomena, such as amyloid formation[13,14], providing fundamental insights into the nature of the microscopic steps in the aggregation process[15–18]. These advances have been made possible by the discovery of integrated rate laws that allow relating experimental measurements to the underlying microscopic mechanisms and hence studying the self-assembly into open-ended fibrillar structures at a highly detailed level[15–18]. It has however remained challenging to exploit the full power of the chemical kinetics approach beyond fibril formation to probe the molecular-level mechanisms of the more complex phenomenon of self-assembly into closed supramolecular structures, a difficultly originating in large part from the absence of integrated rate laws describing such processes. Here, we make a step forward in this direction by deriving a closed-form solution to a set of rate equations describing the assembly kinetics of molecular building blocks into closed target structures[19,20], and show how the availability of this integrated rate law uncovers, from a kinetic

[1]Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge, CB2 1EW, UK. [2]Paulson School of Engineering and Applied Sciences, Harvard University, Cambridge, MA, 02138, USA. [3]Laboratory of Chemical Physics, National Institute of Digestive and Diabetes and Kidney Diseases, National Institutes of Health, Bethesda, MD, 20892, USA. [4]Department of Physics, Brandeis University, Waltham, MA, 02454, USA. [5]Cavendish Laboratory, Department of Physics, University of Cambridge, J J Thomson Avenue, Cambridge, CB3 1HE, United Kingdom. Correspondence and requests for materials should be addressed to T.P.J.K. (email: tpjk2@cam.ac.uk)

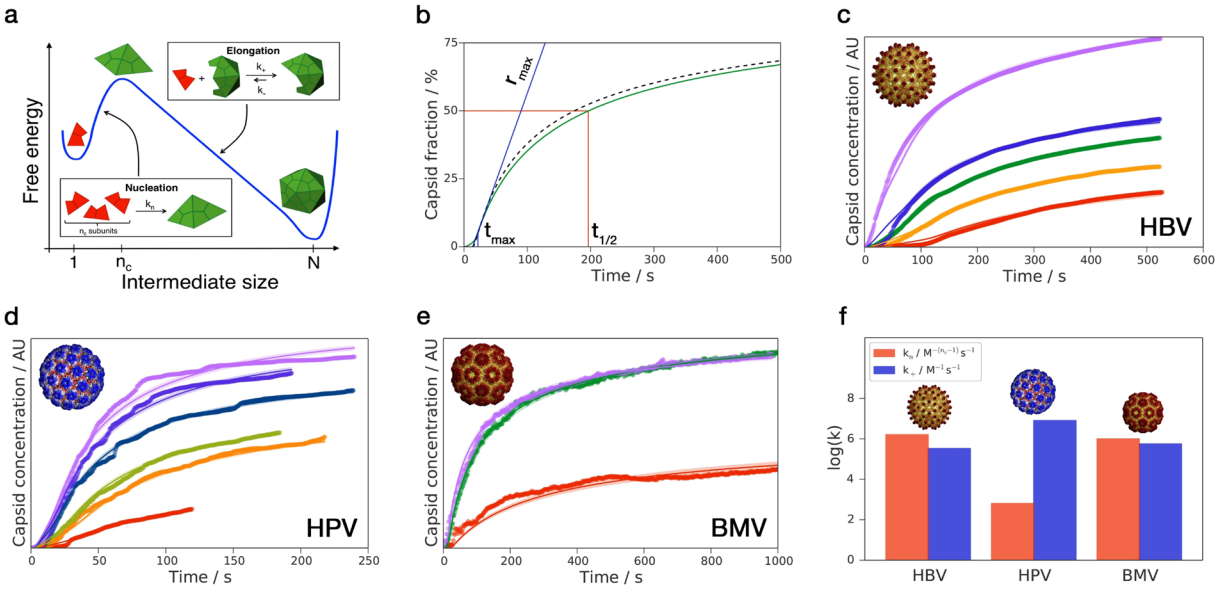


Figure 1. (**a**) Schematic representation of assembly line model: subunits nucleate first and then proceed downhill through elongation reactions to the final structure. Structures in the scheme exemplify assembly with $n_c = 3$ and $N = 30$. (**b**) Comparison of numerical solution to Eqs (1) and (2) (dashed black) with Eq. (9) (solid blue) and definitions of characteristic times t_{max} and $t_{1/2}$. Calculation parameters: $N = 90$, $n_c = 3$, $k_n = 1 \times 10^6$ $M^{-2}\,s^{-1}$ $k_+ = 5.6 \times 10^5\,M^{-1}\,s^{-1}$ and $m(0) = 10\,\mu M$. (**c**–**e**) Global fits of various virus kinetics. (**c**) Hepatitis B Virus with $m(0) = 3.8$, 5.4, 6.4, 8.2 and 10.8 μM. Data from[19]. (**d**) Human Papillomavirus with $m(0) = 0.40$, 0.41, 0.53, 0.72, 0.74, and 0.80 μM. Data from[21]. (**e**) Brome Mosaic Virus with $m(0) = 6.2$, 11.1 and 14.0 μM. Data from[22]. (**f**) Extracted elongation and nucleation rate constants for all viral systems considered. Note that all experimental data analyzed in this work were obtained using purified proteins. Viral images reproduced from[23] with permission.

analysis of experimental data, general dynamic constraints on the microscopic rate constants that control efficient supramolecular self-assembly in such systems.

Results and Discussion

Fundamental kinetic equations. The self-assembly of molecular building blocks into closed target structures may be captured by the following set of kinetic equations for the concentration $f(t, j)$ of intermediates of size j, known as the assembly line model (Fig. 1(a))[19,20]:

$$\begin{aligned}\frac{\partial f(t,j)}{\partial t} &= k_+ m(t) f(t, j-1) - k_+ m(t) f(t, j) + k_n m(t)^{n_c} \delta_{j,n_c} \\ \frac{\partial f(t,N)}{\partial t} &= k_+ m(t) f(t, N-1),\end{aligned} \tag{1}$$

where N is the number of subunits in the target structure and $m(t)$ is the concentration of free subunits in solution, as determined by conservation of the total subunit concentration

$$\frac{dm(t)}{dt} = -\frac{d}{dt}\sum_{j=n_c}^{N} j f\ (t, j). \tag{2}$$

The terms on the first line of Eq. (1) describe the growth of assembly intermediates through the addition of individual subunits with rate constant k_+. The term $k_n m(t)^{n_c}$ describes the initial nucleation step[24] as the spontaneous formation of the smallest growth-competent intermediate from the interaction of n_c subunits with rate constant k_n. Thus, the parameter n_c corresponds to the reaction order of the nucleation step and, in the simplest scenario, can be thought of as the size of the smallest stable assembly intermediate (n_c is the analogous quantity to the critical nucleus size in classical nucleation theory); intermediates with size $j < n_c$ are unstable and quickly dissociate back to free subunits, such that their concentration can be assumed to be negligible. Finally, the last equation of (1) describes the end step of the assembly line as the closure of an intermediate of size $N-1$ into the final structure. Note that in the limit of infinite N, Eq. (1) recover the kinetic equations commonly used to describe filamentous protein assembly processes[14–16]. Note that, in Eq. (1) we have assumed size-independent rate constants; this assumption was primarily made for minimizing the number of model parameters to avoid over-fitting in the analysis of kinetic data, but our framework can be extended straightforwardly to take this effect into account. Moreover, Eq. (1) are deterministic and neglect therefore the potential effect of statistical number fluctuations. Such fluctuations are often negligible in reactions in bulk, but can become dominant in reactions

under volume confinement[25]. We also note that Eq. (1) assume that assembly results in a single closed capsid geometry. Application to reactions that yield multiple capsid morphologies, such as the CCMV systems studied in refs[26,27], require a more complex master equation.

Integrated rate law for assembly kinetics. To obtain an integrated rate law to Eqs (1) and (2), we make use of the perturbative renormalization group (RG)[28], a general mathematical technique for constructing approximative solutions to nonlinear differential equations. In our case, the applicability of this method relies on the observation that the dimensionless ratio $\varepsilon \equiv k_n m(0)^{n_c-2}/k_+ \ll 1$ is a small parameter, where $m(0)$ is the initial subunit concentration. In order to take advantage formally of the smallness of ε, it is convenient to rewrite Eq. (1) in dimensionless form

$$\begin{aligned}\frac{\partial\phi(\tau, j)}{\partial\tau} &= \mu(\tau)\phi(\tau, j-1) - \mu(\tau)\phi(\tau, j) \\ &\quad +\varepsilon\mu(\tau)^{n_c}\delta_{j,n_c},\end{aligned} \tag{3}$$

where $\tau = k_+ m(0)t$, $\mu(\tau) = m(\tau)/m(0)$ and $\phi(\tau, j) = f(\tau, j)/m(0)$. Solving Eq. (3) perturbatively yields, after some algebra, the following result for the time-varying subunit concentration:

$$\begin{aligned}\mu(\tau) &= \rho - \varepsilon N\rho^{n_c}\tau \\ &\quad +\varepsilon\rho^{n_c-1}\left[\Omega_0 - e^{-\rho\tau}\sum_{k=0}^{N-n_c-1}\Omega_k\frac{\rho^k\tau^k}{k!}\right] + \cdots,\end{aligned} \tag{4}$$

where $\rho = 1$ and $\Omega_k = (N - n_c - k)(N - n_c - k + 1)/2$. While Eq. (4) is accurate for short times, we observe the emergence of a divergent term $(-\varepsilon N\tau)$ at later times, which prevents this linearized early-time solution from being valid over the full time course of the reaction. At a fundamental level, this divergence emerges due to our ignorance about the system's behavior in the future; in fact, while the information about the initial concentration of subunits is sufficient for describing the system dynamics over short timescales, at later times the lack of information about the way in which ρ varies with changing timescale is what causes Eq. (4) to depart from the true solution. Perturbative RG provides a systematic method for dealing with this undesired divergence and hence obtain a global approximation valid for the duration of the whole reaction. Note that this procedure mirrors very closely the conventional RG approaches of quantum field theory and condensed matter physics. In these theories, we are interested in describing how a certain quantity of interest, such as the charge or the mass of an electron, is renormalized as we vary the observation scale (e.g. momentum or energy scale in quantum field theory). The missing information about the large-scale (e.g. high-energy) behavior of the system is packed into so-called counter terms, which are constructed in order to cancel the divergencies in the theory. In our case, the analogous quantity to the electron charge or mass of quantum field theory is the initial concentration of monomers and the RG procedure should yield renormalized values for this quantity at different time scales. Following the conventional work-flow of RG, we start by introducing an arbitrary time scale σ which we will vary between the initial time 0 and the observation time τ and then allow for a σ-dependence of the initial subunit concentration by writing $\rho = \rho(\sigma) + \varepsilon\delta\rho(\sigma)$, where $\rho(\sigma)$ is the renormalized subunit concentration (at scale σ) and $\delta\rho(\sigma)$ is a counter term. The counter term $\delta\rho(\sigma) = N\rho^{n_c}\sigma$ removes the divergent term in Eq. (4) and so we arrive at the following renormalized expansion

$$\mu(\tau) = \rho(\sigma) - \varepsilon N\rho^{n_c}(\tau - \sigma) + \mathcal{R}, \tag{5}$$

where $\mathcal{R}$ stands for regular terms. As a next step in the RG framework, we require $\partial\mu/\partial\sigma = 0$ since σ is arbitrary. Doing so, we arrive at the following RG equation

$$\frac{\partial\rho}{\partial\sigma} = -\varepsilon N\rho(\sigma)^{n_c}. \tag{6}$$

By solving Eq. (6) and substituting in Eq. (5) as $\sigma \to \tau$ we obtain the uniformly valid solutions

$$\mu(\tau) = \rho(\tau) + \varepsilon\rho(\tau)^{n_c-1}\left[\Omega_0 - e^{-\rho(\tau)\tau}\sum_{k=0}^{N-n_c-1}\Omega_k\frac{\rho(\tau)^k\tau^k}{k!}\right], \tag{7}$$

$$\phi(\tau, j) = \varepsilon\rho(\tau)^{n_c-1}\left[1 - e^{-\rho(\tau)\tau}\sum_{k=0}^{j-n_c}\frac{\rho(\tau)^k\tau^k}{k!}\right]. \tag{8}$$

Finally, using conservation of total subunit concentration and transforming back to real time t we arrive at the following integrated rate law for the concentration of closed target structures:

$$f(t,N) = \frac{m(0)-\rho(t)}{N} - \frac{(N-n_c)k_n\rho(t)^{n_c-1}}{k_+} \times\left[1-e^{-k_+\rho(t)t}\sum_{k=0}^{N-n_c-1}\frac{(N-n_c-k)}{(N-n_c)}\frac{[k_+\rho(t)t]^k}{k!}\right], \quad (9)$$

where

$$\rho(t) = \frac{m(0)}{[1+N(n_c-1)k_n m(0)^{n_c-1}t]^{1/(n_c-1)}}. \quad (10)$$

This solution shows overall good agreement with the numerical evaluation of Eqs (1) and (2) (Fig. 1(b) and see Supplementary Material for a discussion on the accuracy of Eq. (9) as a function of ε). Moreover, we note that within the first-order RG approximation discussed here the kinetic trace for capsid formation is systematically underestimated by the analytical solution. This is because the function $\rho(t)$ obtained by solving the first-order RG equation decays faster than the true solution. These errors can be reduced by applying the RG method to higher orders in ε.

General characteristics of assembly kinetics. Using the integrated rate law, Eq. (9), we are now in the position to derive, from first principles, a number of relationships characterizing the time course of the assembly reaction. According to Eq. (9), the time evolution of the concentration of target structures demonstrates the characteristic sigmoidal shape defined by an initial lag phase followed by a phase of rapid growth and final asymptotic approach to the plateau[20]. A defining feature of the early time behaviour is the presence of a point of inflection t_{max} at which the growth rate $r = df(t,N)/dt$ is maximal. Solving the equation $dr/dt|_{t_{max}} = 0$ yields the position of the inflection point as

$$t_{max} = \frac{N-n_c}{k_+ m(0)}. \quad (11)$$

The time of inflection is determined completely by the characteristic elongation timescale $(k_+ m(0))^{-1}$. The physical interpretation of Eq. (11) is that of the time required for $N-n_c$ elongation steps to occur. This result is consistent with the idea that the lag phase of the reaction corresponds to a waiting period during which the assembly line is set up and all intermediate states are populated[20].

The maximal growth rate $r_{max} = df(t,N)/dt|_{t_{max}}$ is computed from Eq. (9) as

$$r_{max} = \frac{k_n m(0)^{n_c}(N-n_c)^{N-n_c}}{(N-n_c)!}e^{-(N-n_c)}. \quad (12)$$

Note that r_{max} is given by the product of the rate of rate-limiting nucleation step and the Poissonian probability of observing the minimal number $N-n_c$ of elongation steps required to complete the assembly structure. A key prediction of Eq. (12) is the emergence of a power-law scaling of the maximal growth rate with initial subunit concentration $r_{max} \sim m(0)^{\gamma}$. Because the scaling exponent γ solely depends on the nature of the nucleation step, $\gamma = n_c$, the critical nucleus size can be determined from the slope of a log-log plot of r_{max} vs $m(0)$. Thus, as in many other areas of science[29,30], scaling laws emerge in the context of supramolecular assembly as a general property that connects macroscopic data with the physical nature of the underlying microscopic processes through the value of the scaling exponent.

Equation (9) implies that the median assembly time $t_{1/2}$, defined by the condition $f(t_{1/2},N) = m(0)/(2N)$, is given by $t_{1/2} = 2t_{max} + t_{nuc}$, where

$$t_{nuc} = \frac{2^{n_c}-1}{N(n_c-1)k_n m(0)^{n_c-1}} \quad (13)$$

is the time needed to consume half the free subunits after the assembly line is set up, assuming that each nucleation event leads to the target structure through the eventual consumption of N subunits (see Supplementary Information). This result shows that the $t_{1/2}$ is given as a sum of two distinct contributions, one originating from t_{max}, the time necessary to form a quasi-steady state of intermediates, and the other from $t_{nuc}+t_{max}$, the time to nucleate a sufficient amount of intermediates that mature into the final structure through the the chain reactions of the assembly line. The former contribution to $t_{1/2}$ depends only on the efficiency of the elongation reactions in the assembly line, while the latter is governed by nucleation events. Crucially, the relative importance of these two contributions to the median assembly time is determined by the parameter $\varepsilon = k_n m(0)^{n_c-2}/k_+$. This quantity–which measures the ratio of the rates of nucleation and elongation–naturally emerges from our theoretical framework as the key parameter controlling the assembly kinetics. In general, large values of ε correspond to a kinetic trap, whereby subunits are significantly depleted by nucleating too quickly, leaving less material to complete the assembly of target structures. By contrast, when ε is small, few nuclei are formed and the assembly yield is low for relevant time scales. The crossover between these two regimes occurs when $t_{nuc} = 2t_{max}$. Using the results above, this criterion can be formulated as a condition on the parameter ε as:

$$\varepsilon_c = \frac{2^{n_c-1} - 1}{2N(N - n_c)(n_c - 1)}. \quad (14)$$

When $\varepsilon > \varepsilon_c$, the system is susceptible to kinetic traps, whereas when $\varepsilon < \varepsilon_c$ the assembly is inefficient. According to this criterion, successful assembly is the result of a delicate balance between the necessity of forming appreciable amounts of target structures and the danger of being kinetically trapped. Controlling the relative importance of nucleation and elongation processes provides therefore a high degree of intrinsic regulation of self-assembly[20]. We note that ε_c decreases with increasing size N of the target geometry as N^{-2}. This behavior follows intuition because larger target structures impose stronger constraints on the time available for nucleation, $t_{nuc} \sim 1/N$, while the time required for producing the quasi-steady state assembly line, $t_{\max} \sim N$, is inevitably longer for larger N.

Kinetic analysis of experimental data. Through the analysis of experimental kinetic data, we now demonstrate that the theoretical framework provided by Eq. (9) is capable of describing macroscopic features of supra-molecular self-assembly into closed topologies in terms of microscopic rate constants. We took a representative example and considered kinetic data of the formation of icosahedral viral capsids. Since the current version of our theory only considers empty capsid assembly, we limit our comparison to *in vitro* experiments on the assembly of purified capsid proteins; i.e., the systems do not include viral genomes, other viral proteins, or host factors. Previous studies modeling viral capsid assembly kinetics using master equations[19,20,31–36], continuum models[37–39] or molecular dynamics simulations[40–49] have led to important insights into the system characteristics, yet it remains a key challenge to elucidate the general physical principles underlying capsid assembly. First, we consider the assembly kinetics of Human Hepatitis B Virus (HBV)[19], a representative icosahedral virus comprised predominantly of $N = 120$ subunits. Figure 1(c) shows the time evolution of HBV capsid concentration, as monitored by light scattering intensity, fit globally to the integrated rate law Eq. (9) with fixed $n_c = 3$ (as determined from the scaling of maximal growth rate, Fig. 2(b)), yielding rate constants of $k_+ = 3.32 \pm 0.15 \times 10^5\ \mathrm{M^{-1}\,s^{-1}}$ and $k_n = 1.6 \pm 0.9 \times 10^6\ \mathrm{M^{-2}\,s^{-1}}$. The global nature of the fit demonstrates the consistent agreement between Eq. (9) and the full time courses observed in the experiment over a wide range of initial subunit concentrations, including the characteristic sigmoidal shape of kinetic traces. We note that the entire data set could be fitted to Eq. (9) using just two global rate constants and one concentration-dependent plateau parameter for each kinetic curve that accounts for the constant of proportionality between the measured light scattering signal and the capsid concentration. In the SI, we provide also fitting to HBV assembly data under reducing conditions at 37 °C and pH 7.5 from ref.[50]. The fits to these higher temperature data, however, are less accurate, which could arise due to late stage intermediates[52] and protein interconversion between assembly-active and assembly-inactive conformations[53,54]. Next, we consider kinetic data for the formation of Human Papillomavirus (HPV, Fig. 1(d))[21] and Brome Mosaic Virus (BMV, Fig. 1(e)) capsids[22]. Using $n_c = 2$, $N = 72$ (HPV) and $n_c = 3$, $N = 90$ (BMV), global fits of experimental data to Eq. (9) with $k_+ = 8.0 \pm 0.5 \times 10^6\ \mathrm{M^{-1}\,s^{-1}}$, $k_n = 6.4 \pm 0.4 \times 10^2\ \mathrm{M^{-1}\,s^{-1}}$ (HPV) and $k_+ = 5.6 \pm 1.0 \times 10^5\ \mathrm{M^{-1}\,s^{-1}}$, $k_n = 9.9 \pm 0.8 \times 10^6\ \mathrm{M^{-2}\,s^{-1}}$ (BMV) are again able to describe the full time course of the assembly reactions. Furthermore, from the analysis of the experimental kinetic data, we can also directly verify the scaling predictions that have resulted from our analytical treatment of the master equation (1) for the three virus systems discussed here. Figure 2(a) shows a double logarithmic plot of the measured inflection times, t_{max}, against initial subunit concentration for the three systems considered in Fig. 1 together with the predicted scaling law, $t_{max} \sim m(0)^{-1}$. The scatter in the data for the inflection time is due to increased experimental noise in the kinetic profiles close to the initial point of the reaction. Moreover, Fig. 2(b) illustrates how the relevant value for the reaction order for the nucleation step, n_c, can be determined from the analysis of the maximal growth rate as a function of total subunit concentration. Through the analysis of the maximal growth rate, it is therefore possible to fix the value of n_c necessary for fitting kinetic traces. We note that the value of n_c for BMV was set to 3 as reported previously in the literature[22]; this was done because the corresponding dataset has too few points for confident fitting. We also note that similar scaling laws have been previously obtained approximately by assuming an ad hoc separation between nucleation and growth processes[39,55].

The availability of microscopic rate constants enabled by the present analysis allows mechanistic comparisons to be made between the assembly of different virus capsid systems. Interestingly, while the absolute values of the rate constants obtained from the fitting of experimental data vary over several orders of magnitude (Fig. 1(f)), we observe that the parameter ε takes similar values across the different data sets: $\varepsilon = 5.0 \pm 0.4 \times 10^{-5}$ ($m(0) = 10\,\mu$M) for HBV, $\varepsilon = 8.0 \pm 0.7 \times 10^{-5}$ ($m(0) = 1\,\mu$M) for HPV and $\varepsilon = 1.7 \pm 0.3 \times 10^{-5}$ ($m(0) = 10\,\mu$M) for BMV. Moreover, these values fall in the same order of magnitude as the theoretical predictions for ε_c: 5.5×10^{-5} (HBV), 1.0×10^{-4} (HPV) and 9.5×10^{-5} (BMV). This illustrates how the apparently distinct viral systems studied in this work are characterized by a similar balance of the relative rates of elongation and nucleation to achieve successful assembly. By contrast, for filamentous protein self-assembly[14–16] the long-time average length of aggregates $\langle L \rangle$ becomes $\langle L \rangle \sim 1/\sqrt{\varepsilon_c}$. As linear systems such as actin[51] are required by their biological function to be long, they should have low measured ε so as to maximize efficiency. This prediction is in agreement with what is observed in Fig. 2(c).

Conclusions

In conclusion, although it is of both fundamental and practical interest to identify and characterize the kinetic constraints governing supra-molecular self-assembly into closed target structures, this understanding has proved challenging to achieve in practice. Here, we have demonstrated how the availability of integrated rate laws to

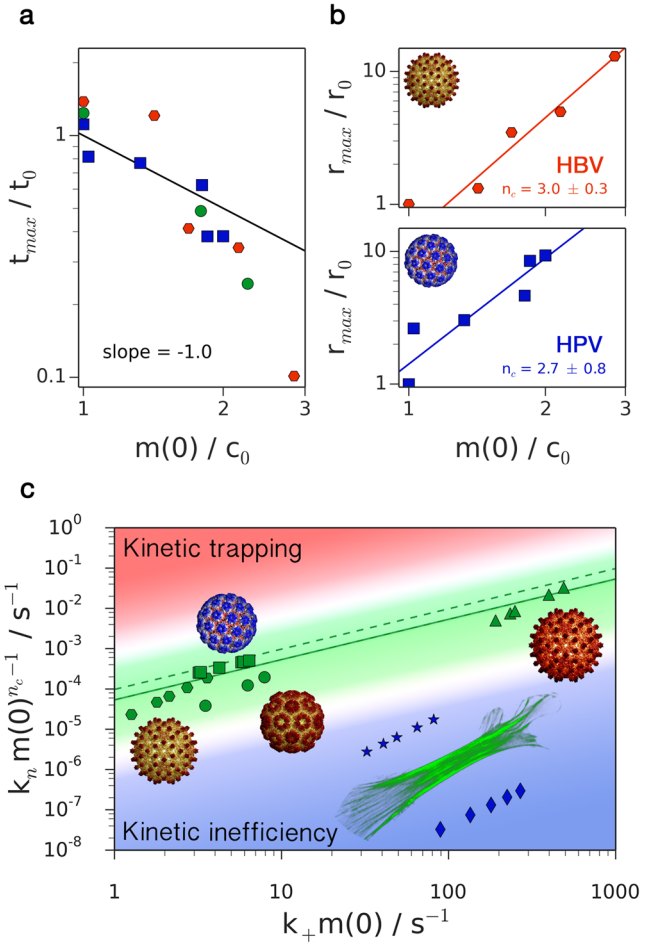


Figure 2. Scaling behavior of viral capsid assembly properties and assembly efficiency. (**a**) Scaling of t_{max} with $m(0)$ as predicted by Eq. (11) for various viral systems with data shown as circles (BMV), squares (HPV) and hexagons (HBV). (**b**) The reaction order for nucleation, n_c, is obtained from the scaling behavior of r_{max}. The data are for HBV and HPV. Note that discontinuities in the experimental kinetic traces for HPV assembly are responsible for inaccuracies in determining r_{max} for data at higher initial concentrations. (**c**) Balance between elongation and nucleation in the viral systems studied in this work. Green solid line corresponds to $\varepsilon = 4.9 \times 10^{-5}$, green dashed lines correspond to ε_c (Eq. (14)) for the various viruses. BMV data denoted by circles, HPV by squares, HBV by hexagons and HBV assembly data obtained at 37 °C and pH 7.5 from[50] (see Supplementary Information) by triangles. Blue data are from actin polymerization measurements in magnesium (stars) or in calcium (diamonds) from[51]. Viral images reproduced from[23] with permission.

the underlying kinetic equations illuminates the dynamic design criteria that characterize the efficiency of such processes. We showed that efficient assembly only occurs in a narrow range of parameter space. By applying this kinetic analysis to experimental data of icosahedral viral capsid assembly we demonstrated that these structures occupy this narrow region of parameter space corresponding to efficient assembly.

References

1. Whitesides, G. M. & Grzybowski, B. Self-assembly at all scales. *Science* **295**, 2418–2421 (2002).
2. Edeling, M. A., Smith, C. & Owen, D. Life of a clathrin coat: insights from clathrin and AP structures. *Nat. Revs. Mol. Cell Biol.* **7**, 32–44 (2006).
3. Fletcher, J. M. *et al.* Self-assembling cages from coiled-coil peptide modules. *Science* **340**, 595–599 (2013).
4. Fath, S., Mancias, J. D., Bi, X. & Goldberg, J. Structure and organization of coat proteins in the COPII cage. *Cell* **129**, 1325–1336 (2007).
5. Boato, F. *et al.* Synthetic virus-like particles from self-assembling coiled-coil lipopeptides and their use in antigen display to the immune system. *Angew. Chem. Int. Ed.* **46**, 9015–9018 (2007).
6. Wörsdörfer, B., Woycechowsky, K. J. & Hilvert, D. *Science* **331**, 589–592 (2011).
7. King, N. P. *et al.* Computational design of self-assembling protein nanomaterials with atomic level accuracy. *Science* **336**, 1171–1174 (2012).
8. Lai, Y.-T., Cascio, D. & Yeates, T. O. Structure of a 16-nm cage designed by using protein oligomers. *Science* **336**, 1129 (2012).
9. Caspar, D. L. & Klug, A. Physical principles in the construction of regular viruses. *Cold Spring Harb Symp Quant Biol* **27**, 1–24 (1962).

10. Bancroft, J. B., Hills, G. J. & Markham, R. A study of the self-assembly process in a small spherical virus. *Formation of organized structures from protein subunits in vitro. Virology* **31**, 354–379 (1967).
11. Crick, F. H. C. & Watson, J. D. Structure of small viruses. *Nature* **177**, 473–475 (1956).
12. Fasching, L. *et al*. TRIM28 represses transcription of endogenous retroviruses in neural progenitor cells. *Cell Rep* **10**, 20–28 (2015).
13. Knowles, T. P. J., Vendruscolo, M. & Dobson, C. M. The amyloid state and its association with protein misfolding diseases. *Nat. Rev. Mol. Cell Biol.* **15**, 384–96 (2014).
14. Oosawa, F. & Asakura, S. *Thermodynamics of the polymerisation of proteins* (Academic Press, 1975).
15. Knowles, T. P. J. *et al*. An analytical solution to the kinetics of breakable filament assembly. *Science* **326**, 1533–1537 (2009).
16. Cohen, S. I. A. *et al*. Proliferation of amyloid-β42 aggregates occurs through a secondary nucleation mechanism. *Proc. Natl. Acad. Sci. USA* **110**, 9758–9763 (2013).
17. Meisl, G. *et al*. Molecular mechanisms of protein aggregation from global fitting of kinetic models. *Nat. Protoc.* **11**, 252–272 (2016).
18. Thomas C. T. Michaels *et al*. Hamiltonian Dynamics of Protein Filament Formation. *Phys Rev Lett* **116**, 038101 (2016).
19. Zlotnick, A., Johnson, J. M., Wingfield, P. W., Stahl, S. J. & Endres, D. A theoretical model successfully identifies features of Hepatitis B virus capsid assembly. *Biochemistry* **38**, 14644–14652 (1999).
20. Endres, D. & Zlotnick, A. Model-based analysis of assembly kinetics for virus capsids or other spherical polymers. *Biophys J* **83**, 1217–1230 (2002).
21. Casini, G. L., Graham, D., Heine, D., Garcea, R. L. & Wu, D. T. *In vitro* papillomavirus capsid assembly analyzed by light scattering. *Virology* **325**, 320–327 (2004).
22. Chen, C., Kao, C. C. & Dragnea, B. Self-assembly of Brome Mosaic virus capsids: Insights from shorter time-scale experiments. *J Phys Chem A* **112**, 9405–9412 (2008).
23. Fauquet, C. M., Mayo, M. A., Maniloff, J., Desselberger, U. & Ball, L. A. Eds *Virus Taxonomy: VIIIth Report of the International Committee on Taxonomy of Viruses*, (Elsevier/Academic Press, 2005).
24. Prevelige, P. E. Jr., Thomas, D. & King, J. Nucleation and growth phases in the polymerization of coat and scaffolding subunits into icosahedral procapsid shells. *Biophys J* **64**, 824–835 (1993).
25. Michaels, T. C. T. *et al*. *Phys Rev Lett* **116**, 258103 (2016).
26. Zlotnick, A., Aldrich, R., Johnson, J. M., Ceres, P. & Young, M. J. *Virology* **277**, 450 (2000).
27. Johnson, J. M. *et al*. *Nano Lett.* **5**, 765–770 (2005).
28. Chen, L.-Y., Goldenfeld, N. & Oono, Y. Renormalization group theory for global asymptotic analysis. *Phys Rev Lett* **73**, 1311–1315 (1994).
29. Barabasi, A. L. & Albert, R. Emergence of scaling in random networks. *Science* **286**, 509–512 (1999).
30. Skotheim, J. M. & Mahadevan, L. Physical limits and design principles for plant and fungal movements. *Science* **308**, 1308–1310 (2005).
31. Zhang, T. & Schwartz, R. Simulation study of the contribution of oligomer/oligomer binding to capsid assembly kinetics. *Biophys J* **90**, 57–64 (2006).
32. Keef, T., Micheletti, C. & Twarock, R. Master equation approach to the assembly of viral capsids. *J Theor Biol* **242**, 713–721 (2006).
33. Hemberg, M., Yaliraki, S. N. & Barahona, M. Stochastic kinetics of viral capsid assembly based on detailed protein structures. *Biophys J* **90**, 3029–3042 (2006).
34. Smith, G. R., Xie, L., Lee, B. & Schwartz, R. Applying molecular crowding models to simulations of virus capsid assembly *in vitro*. *Biophys J* **106**, 310–320 (2014).
35. Dykeman, E. C., Stockley, P. G. & Twarock, R. Building a viral capsid in the presence of genomic RNA. *Phys Rev E* **87**(022717), 1–12 (2013).
36. Hagan, M. F. Modeling viral capsid assembly. *Adv Chem Phys* **155**, 1–67 (2014).
37. Zandi, R., van der Schoot, P., Reguera, D., Kegel, W. & Reiss, H. Classical nucleation theory of virus capsids. *Biophys J* **90**, 1939–1948 (2006).
38. van der Schoot, P. & Zandi, R. Kinetic theory of virus capsid assembly. *Phys Biol* **4**, 296–304 (2007).
39. Morozov, A. Y., Bruinsma, R. F. & Rudnick, J. Assembly of viruses and the pseudo-law of mass action. *J Chem Phys* **131**(155101), 1–17 (2009).
40. Hagan, M. F. & Chandler, D. Dynamic pathways for viral capsid assembly. *Biophys J* **91**, 42–54 (2006).
41. Matthews, R. & Likos, C. N. Dynamics of self-assembly of model viral capsids in the presence of a fluctuating membrane. *J Phys Chem B* **117**, 8283–8292 (2013).
42. Johnston, I. G., Louis, A. A. & Doye, J. P. K. Modelling the self-assembly of virus capsids. *J Phys: Condens Matter* **22**(104101), 1–9 (2010).
43. Hicks, S. D. & Henley, C. L. Irreversible growth model for virus capsid assembly. *Phys Rev E* **74**(031912), 1–17 (2006).
44. Nguyen, H. D., Reddy, V. S. & Brooks, C. L. Deciphering the kinetic mechanism of spontaneous self-assembly of icosahedral capsids. *Nano Lett* **7**, 338–344 (2007).
45. Nguyen, H. D., Reddy, V. S. & Brooks, C. L. Invariant polymorphism in virus capsid assembly. *J Am Chem Soc* **131**, 2606–2614 (2009).
46. Schwartz, R., Shor, P. W., Prevelige, P. E. & Berger, B. Local rules simulation of the kinetics of virus capsid self-assembly. *Biophys J* **75**, 2626–2636 (1998).
47. Wilber, A. W. *et al*. Reversible self-assembly of patchy particles into monodisperse icosahedral clusters. *J Chem Phys* **127**(085106), 1–11 (2007).
48. Rapaport, D. C. Molecular dynamics simulation of reversibly self-assembling shells in solution using trapezoidal particles. *Phys Rev E* **86**(051917), 1–7 (2012).
49. Perlmutter, J. D. & Hagan, M. F. Mechanisms of virus assembly. *Annu Rev Phys Chem* **66**, 217–239 (2015).
50. Selzer, L., Katen, S. P. & Zlotnick, A. The Hepatitis B virus core protein intradimer interface modulates capsid assembly and stability. *Biochemistry* **53**, 5496–5504 (2014).
51. Cooper, J. A., Buhle, E. L. Jr., Walker, S. B., Tsong, T. Y. & Pollard, T. D. Kinetic evidence for a monomer activation step in actin polymerization. *Biochemistry* **22**, 2193–2202 (1983).
52. Pierson, E. E. *et al*. Detection of late intermediates in virus capsid assembly by charge detection mass spectrometry. *J. Am. Chem. Soc.* **136**, 3536–41 (2014).
53. Packianathan, C., Katen, S. P., Dann, C. E. 3rd & Zlotnick, A. Conformational changes in the hepatitis B virus core protein are consistent with a role for allostery in virus assembly. *J. Virol.* **84**, 1607–15 (2010).
54. Lazaro, G. R. & Hagan, M. F. Allosteric Control of Icosahedral Capsid Assembly. *J. Phys. Chem. B* **120**, 6306–18 (2016).
55. Hagan, M. F. & Elrad, O. M. Understanding the concentration dependence of viral capsid assembly kinetics-the origin of the lag time and identifying the critical nucleus size. *Biophys J* **98**, 1065–1074 (2010).

Acknowledgements

We acknowledge support from the Swiss National Science Foundation (TCTM), Peterhouse, Cambridge (TCTM), the Cambridge Commonwealth, European and International Trust (MMJB), the NIH-Oxford/Cambridge Scholars Program (MMJB), Award Number R01GM108021 from the National Institute of General

Medical Sciences (MFH), ERC (TCTM, TPJK), BBSRC (TPJK) and the Newman foundation (TPJK). This work was in part supported by the Intramural Research Program of the National Institute of Digestive and Diabetes and Kidney Diseases at the National Institutes of Health (MMJB). We thank C. Chen for providing the data in ref.[22]. We thank A. Zlotnick and L. Seltzer for kindly providing the data in refs[19,50] and for fruitful discussions. We thank R.B. Best, R.L. Jack, G. Meisl and C.M. Dobson for useful comments and discussions.

Author Contributions

All authors conceived and designed the project; T.C.T.M. developed the theory; T.C.T.M. and M.M.J.B. analysed the experimental data; all authors wrote the paper.

Additional Information

Supplementary information accompanies this paper at https://doi.org/10.1038/s41598-017-12528-8.

Competing Interests: The authors declare that they have no competing interests.

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