

# Illuminating the Reaction Pathways of Viromimetic Assembly

Hande E. Cingil,<sup>†</sup> Emre B. Boz,<sup>†</sup> Giovanni Biondaro,<sup>‡</sup> Renko de Vries,<sup>†</sup> Martien A. Cohen Stuart,<sup>†</sup> Daniela J. Kraft,<sup>‡</sup> Paul van der Schoot,<sup>§,||</sup> and Joris Sprakel<sup>\*,†,||</sup>

<sup>†</sup>Physical Chemistry and Soft Matter, Wageningen University & Research, Stippeneng 4, 6708 WE Wageningen, The Netherlands

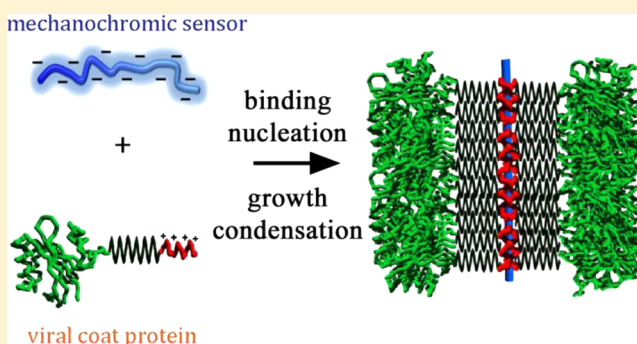
<sup>‡</sup>Soft Matter Physics, Huygens-Kamerling Onnes Laboratory, Leiden University, PO Box 9504, 2300 RA Leiden, The Netherlands

<sup>§</sup>Theory of Polymers and Soft Matter, Eindhoven University of Technology, PO Box 513, 5600 MB Eindhoven, The Netherlands

<sup>||</sup>Institute for Theoretical Physics, Utrecht University, Leuvenlaan 4, 3584 CE Utrecht, The Netherlands

## Supporting Information

**ABSTRACT:** The coassembly of well-defined biological nanostructures relies on a delicate balance between attractive and repulsive interactions between biomolecular building blocks. Viral capsids are a prototypical example, where coat proteins exhibit not only self-interactions but also interact with the cargo they encapsulate. In nature, the balance between antagonistic and synergistic interactions has evolved to avoid kinetic trapping and polymorphism. To date, it has remained a major challenge to experimentally disentangle the complex kinetic reaction pathways that underlie successful coassembly of biomolecular building blocks in a noninvasive approach with high temporal resolution. Here we show how macromolecular force sensors, acting as a genome proxy, allow us to probe the pathways through which a viromimetic protein forms capsids. We uncover the complex multistage process of capsid assembly, which involves recruitment and complexation, followed by allosteric growth of the proteinaceous coat. Under certain conditions, the single-genome particles condense into capsids containing multiple copies of the template. Finally, we derive a theoretical model that quantitatively describes the kinetics of recruitment and growth. These results shed new light on the origins of the pathway complexity in biomolecular coassembly.



## INTRODUCTION

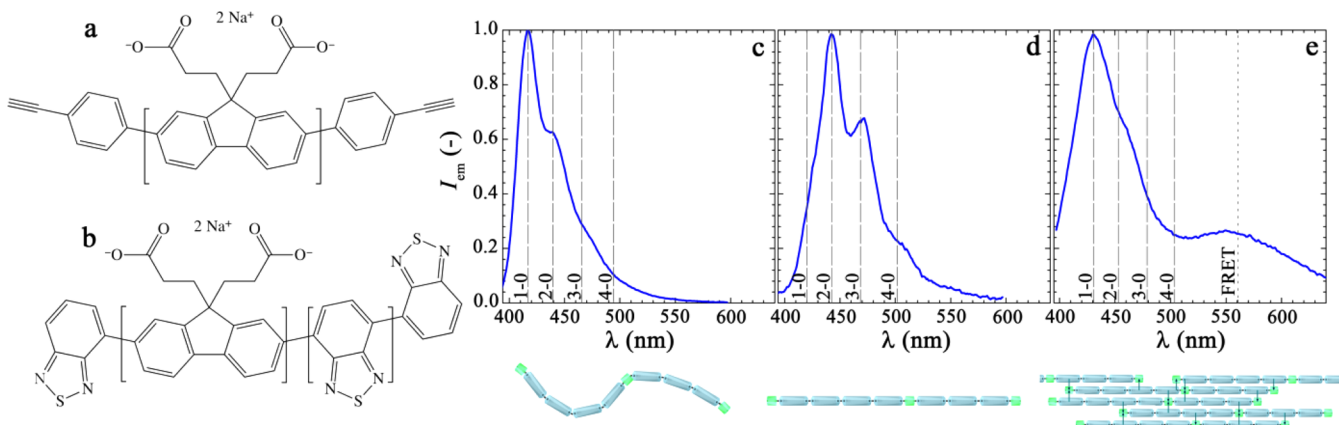
Supramolecular structures in nature derive their functionality from a precisely defined architecture, which in turn is formed by the assembly of biomolecular precursors.<sup>1</sup> Assembling these structures spontaneously, without intervention of the biochemical machinery of the cell, requires a finely tuned balance of repulsive and attractive interactions acting between multiple constituents. Controlling this delicate balance is often achieved through allostery, a highly cooperative process regulated through conformational switching.<sup>2</sup> Moreover, allosteric action also controls the kinetic pathways of assembly; this is crucial to obtain well-defined structures with a high degree of fidelity. A case in point is the spontaneous assembly of simple viruses, in which binding of the coat protein to its genetic cargo sets in motion a cascade of events that leads to successful coassembly.<sup>2c,3</sup> This route may involve a multitude of competing pathways,<sup>3b,4</sup> each of which in turn consists of a large number of elementary docking and folding steps between individual molecules. This gives rise to significant pathway complexity, a topic of intense study in the past decades, also in supramolecular chemistry.<sup>5</sup> Our understanding of what pathways dominate and why, and how they give rise to structural polymorphism and kinetic trapping, is incomplete, not the least because experimental methods to probe them at the relevant

length and time scales are scarce.<sup>6</sup> Radiation scattering methods, while offering excellent statistics, can be difficult to interpret due to the reciprocal space inversion problem. By contrast, imaging methods, such as time-resolved atomic force microscopy or electron microscopy, provide real-space insight into capsid formation but may suffer from poor statistics and time resolution.

Here, we present a new approach to resolve these challenges and probe capsid assembly kinetics in detail. We employ morphology-sensitive luminescent polymers as genome proxies, which act as optical sensors of their coassembly into linear virus-like particles with a recombinant viromimetic protein. We find that capsid formation is initiated by the random binding of coat proteins onto the template, after which a concerted capsid growth ensues, caused by conformational switching of the protein. Near conditions of charge compensation, the single-genome assemblies condense into viruslike particles which carry multiple copies of the template. The binding and reorganization of proteins on the template is captured by a simple model in which aspecific unimolecular binding competes with cooperative multimolecular reorganization. These results shed new light

Received: February 15, 2017

Published: March 22, 2017



**Figure 1.** Chemical structure of the mechano-optical molecular sensors (a) SP1 ( $M_w = 16.7$  kg/mol) and (b) SP2 ( $M_w = 12.0$  kg/mol). (c) Encapsulation of the polymers in a capsid can be detected spectrally; in absence of encapsulating protein, the luminescence spectrum of a relaxed chain features three visible vibronic bands, in which the 1–0 dominates ( $[SP1] = 0.06 \mu\text{M}$ ,  $f_+ = 0$ ). (d) Upon encapsulation in a capsid, the sensor polymer is stretched and planarized which changes the vibronic fine structure, with the 1–0 band decaying and the higher energy bands growing in intensity ( $[SP1] = 0.06 \mu\text{M}$ ,  $f_+ = 0.50$ ). (e) Addition of a benzothiadiazole acceptor-moiety within the chain leads to conformation-dependent Förster resonance energy transfer, introducing an optical read-out to chain bundling and condensation ( $[SP2] = 0.08 \mu\text{M}$ ,  $f_+ = 0.70$ ).

on the origins of the pathway complexity that result from competing interactions, stoichiometry, and the action of allostery in assembling biomolecular systems.

## RESULTS & DISCUSSION

**Coassembling Species.** We study the formation of capsids from a recombinant coat protein inspired by the structure of the tobacco mosaic virus (TMV), whose design and production is described in detail elsewhere<sup>7</sup> and in the [Supporting Information](#). Coat proteins of the TMV feature three distinct functionalities:<sup>8</sup> (I) a hydrophilic domain that protects the capsid and its cargo against aggregation, misfolding, and enzymatic attack, (II) a binding domain with high affinity for the nucleic acids, and (III) a domain that, when folded, provides a specific attraction between neighboring capsid proteins. Our recombinant protein ( $M_w = 45$  kDa), produced biosynthetically in *Pichia pastoris* hosts, features a cationic binding block *B* composed of 12 lysine residues (domain II), a silk-inspired association domain *S*<sub>10</sub> (domain III),<sup>9</sup> and a gelatin-like random coil motif *C* (domain I).<sup>10</sup> These proteins have been shown to form stable viromimetic capsids through templated coassembly with DNA;<sup>7,11</sup> however, the kinetic pathways through which these rod-shaped viral capsids form remain elusive.

To probe its assembly kinetics, we replace the nucleic acid polymer with an anionic conjugated polymer that acts simultaneously as a proxy for the templating genome and as a molecular sensor for the assembly process. We use two different sensor polymers: to study the early stages of capsid formation, in which we expect single template chains to become planarized upon encapsulation, we use a sensor polymer (SP1) that undergoes distinct changes in its luminescence spectrum upon supramolecular stretching.<sup>12</sup> The second sensor polymer<sup>13</sup> (SP2) is used to evaluate if capsids contain a single copy of the template or whether multiple copies become incorporated at some stage along the assembly pathway.

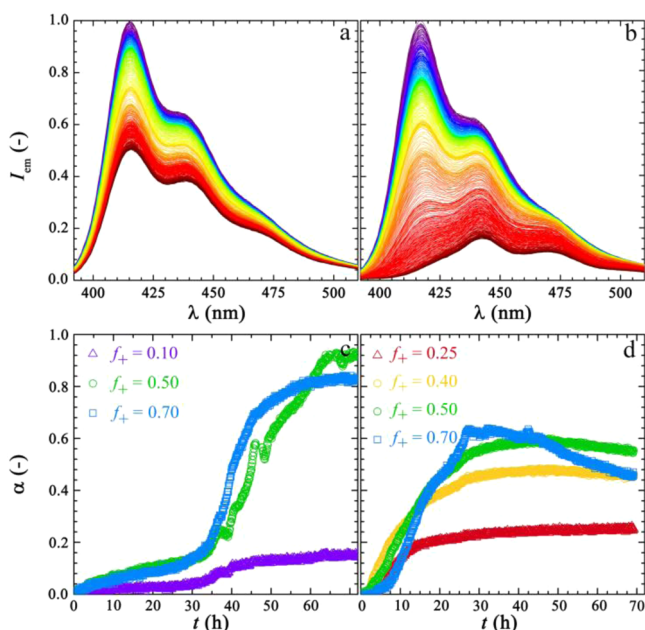
The proxy genome and sensor SP1 ([Figure 1a](#)) exhibits an optomechanical coupling between the conformation of the polymeric backbone and its photoluminescence (PL). In a relaxed state, its PL spectrum features three distinct vibronic bands: the 1–0 transition at  $\lambda = 418$  nm dominates the

luminescence, while minor transitions are visible as the lower energy 2–0 and 3–0 bands ([Figure 1c](#)). In this “naked” state, the conjugation length and delocalized electronic structure along the backbone are limited by rotations between the monomers and the conformational flexibility of the chain.<sup>12,13</sup>

Upon application of a stretching force (e.g., by encapsulation in our viruslike particles), the conformational degrees of freedom are reduced and the chain planarizes into a ribbonlike structure.<sup>12</sup> This results in a distinct change in the vibronic transitions: the highest energy band decays and vanishes upon full planarization, while the intensity of lower-energy transitions grows ([Figure 1d](#)). Here, we exploit this optomechanical coupling to detect stress-induced conformational changes of the polymer during coassembly as a function of time. The second sensor polymer SP2 ([Figure 1](#), panels b and e) that allows us to probe bundling of template chains<sup>13a</sup> will be discussed below.

**Capsid Assembly Kinetics.** We initiate the coassembly by mixing protein with the proxy genome. We express the mixing stoichiometry of the two species as  $f_+ = [ + ] / ([ + ] + [ - ])$ , with  $[ + ]$  and  $[ - ]$  the molar concentrations of cationic charges on the oligolysine binding block and the anionic charges on the template, respectively. Polyionic charge equality is reached if  $f_+ = 0.50$ . Directly after mixing, we begin recording PL spectra every 6 min for  $\sim 3$  days, during which the vibronic fine-structure of the ensemble-averaged luminescence spectra gradually shifts, signaling the progression of the coassembly process.

Initially, an intense peak at the 1–0 vibronic band is observed, which diminishes in time, while the second and third band grow ([Figure 2](#), panels a and b). We previously confirmed that electrostatic complexation of SP1 with a poly(lysine) homopolymer, in the absence of conformational changes in the sensor polymer, does not give rise to these distinct optical signatures.<sup>12</sup> Thus, they are a direct result of the changes in template conformation due to their encapsulation in a proteinaceous coat. As time progresses, coat proteins bind to the template and subsequently condense to form a rigid, partially complete, capsid. The time evolution of the spectra reveals that within the capsid, the template chain is forced into a planarized and stretched conformation.



**Figure 2.** Time-evolution of the photoluminescence spectra of molecular sensor SP1 ( $[\text{SP1}] = 0.06 \mu\text{M}$ ) during encapsulation by the viromimetic protein C-S<sub>10</sub>-B at charge stoichiometries of (a)  $f_+ = 0.10$  and (b)  $0.50$ , with time progressing from purple ( $t = 0$ ) to dark red ( $t = 72$  h). Fraction  $\alpha$  of encapsulated template chain as a function of time at different charge stoichiometries for (c)  $[\text{SP1}] = 0.06$  and (d)  $0.6 \mu\text{M}$ .

The DNA-templated growth of viruslike particles from the same biosynthetic protein has revealed that the assembly of rigid capsids is strongly cooperative.<sup>7,11</sup> Rather than forming a homogeneous coating that densifies gradually, a phase-separated structure emerges on the template, where sections of naked template coexist with segments of condensed capsid.<sup>7</sup> Ultimately, this cooperative coassembly leads to a population inversion, where most genome proxies are completely covered and some remain (almost) completely naked. This implies that in our experiments certain segments of the sensor polymer must be planarized, in a taut state, while the remainder exhibits the emission of a polymer in its relaxed, slack, state. The PL spectra we record are thus an ensemble-averaged convolution of both states.

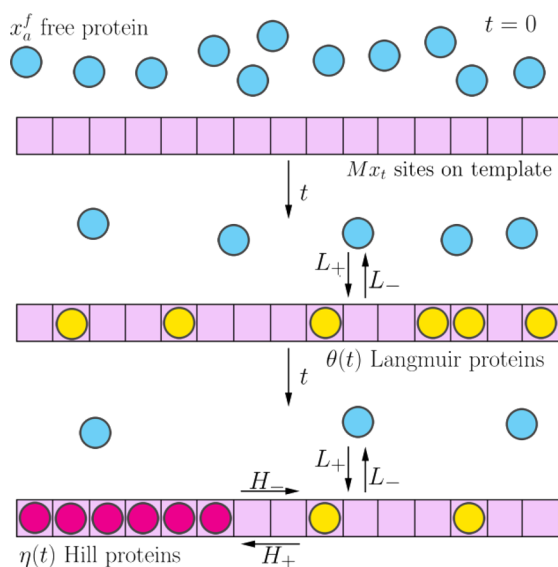
Assuming such a two-state scenario, the intensity of the  $i$ th vibronic band can be written as  $I^i = (1 - \alpha)I_n^i + \alpha I_c^i$ , in which  $I_n^i$  and  $I_c^i$  are the normalized intensities of naked (slack) and coated (taut) states, respectively. The quantity of interest is  $\alpha$ , the fraction of template chain that is encapsulated and stretched. We assume here that each capsid contains a single template chain; we will demonstrate below that is valid only for small values of  $f_+$ , the charge stoichiometry.

Our most sensitive measure of the encapsulation-induced planarization is the intensity ratio between the 1–0 and 2–0 bands  $r_{12} = I^{i=1}/I^{i=2}$ .<sup>12,13</sup> From this quantity, within the two-state approximation, we can directly obtain the value of  $\alpha$  from our experiments, as  $\alpha = (r_{12}I_n^2 - I_n^1)/[(r_{12}(I_n^2 - I_c^2) - I_n^1 + I_c^1)]$ . The reference intensity  $I_n^i$  ( $i = 1$  and  $2$ ) is a constant determined from a spectrum for naked genome proxy chains in the absence of protein. From previous experiments,<sup>12</sup> in which we stretched the sensor polymers to their contour length, we determine the reference intensity for the taut conformation  $I_c^i$ .

The fraction of taut chains  $\alpha$  that we extract from our experiments, allows us to quantify the coassembly kinetics. For a dilute solution of components, with  $[\text{SP1}] = 0.06 \mu\text{M}$ , we observe a three-stage assembly process (Figure 2c). Initially, the degree of encapsulation increases weakly to a plateau at  $\alpha = 0.05$ , which persists for approximately 30 h. After this time lag, the nucleation of dense capsids commences and reaches completion over the course of a few hours as marked by a steep growth in  $\alpha$  up to a mixing-ratio-dependent plateau. For  $f_+ = 0.10$ , we observe only partial encapsulation as insufficient protein is available to neutralize the available charges on the template chains in the solution (Figure 2c), which is consistent with theoretical predictions.<sup>11</sup> At perfect stoichiometry ( $f_+ = 0.50$ ), almost complete coverage is achieved. Interestingly, further increasing the protein content decreases the encapsulation efficiency. We could hypothesize this to be caused by the self-assembly of empty shells at high enough protein concentrations,<sup>7</sup> but as we will show below, it in fact signals the emergence of a third stage in the pathway toward forming complete capsids.

Upon increasing the overall concentration of coassembling species, the nucleation lag time decreases significantly by approximately a factor of 10 (Figure 2d), as to be expected for a nucleation-limited process. Again we observe that in the presence of an excess of coat protein, full coating of the template chains is not accomplished. Note that for a reaction equilibrium, increasing the overall amounts of reactants, should push the equilibrium to the right (product) side. If we compare the data for  $f_+ = 0.50$  at low (Figure 2c) and high (Figure 2d) concentrations, we observe exactly the opposite; the plateau value of  $\alpha$  decreases with the overall concentration. We could presume that this counterintuitive behavior is due to the competing self-assembly process in the solution referred to above. However, as we shall show below using a different proxy genome, that in fact the observation signals the transformation of the single-genome particles formed initially into particles that contain multiple copies of the template that need not be taut to accommodate interactions with the coat proteins. For single-genome particles, stretching the template chain increases the probability that anionic and cationic charges can form a tight electrostatic complex and is thus driven by the Coulombic interactions between genome-binding domain on the protein and template. Once the condensation occurs, each cationic charge can opt to bond to several templates, which thus reduces the necessity for the template stretching, which in itself is unfavorable as it decreases the conformational entropy of the templating polymer. Thus, in the condensed capsid state, the coacervate core of the capsid allows the template to relax to some extent.

**Reaction Pathway Model.** To demonstrate that the assembly process involves more pathways than just binding and cooperative condensation, we quantify these two assembly steps in a kinetic reaction model, illustrated schematically in Figure 3. First, free proteins randomly and reversibly bind to the template following first-order kinetics, which are described by a Langmuir adsorption model.<sup>14</sup> Subsequently, the adsorbed proteins associate and condense along the template to form a rigid capsid, a process we model by cooperative  $n$ th-order Hill-type kinetics.<sup>15</sup> We presume that cooperatively bound molecules can only leave the template by first transitioning to the Langmuir state and following that are able to desorb. Conversely, only adsorbed molecules can undergo the transition to the co-operatively bound Hill state. The fact



**Figure 3.** Schematic illustration of the reaction pathway model that describes capsid assembly as a combination of Langmuir adsorption of free proteins to a template, and the subsequent Hill-type co-operative reorganization of bound proteins into a dense capsid.

that template binding is a required intermediate between free proteins and the final capsids they form is the essence of our model. A complete derivation of our theory for the reaction pathways of viral capsid assembly is provided in the [Supporting Information](#). We note that the sequence of binding followed by lateral assembly has also been considered in previous theoretical efforts.<sup>16</sup>

In brief, we consider a solution of template chains, at mole fraction  $x_t$ , and proteins with mole fraction  $x_a$ . Each template chain has  $M$  binding sites, with a total number of available binding site  $Mx_t$ . A fraction  $\eta(t)$  of these sites is occupied by cooperatively associated proteins in a capsid, whereas  $\theta(t)$  denotes the fraction of the remaining  $1 - \eta(t)$  sites filled by bound but unassociated molecules, in a Langmuir state. At  $t = 0$ , all proteins are free in solution, and thus  $\eta(t) = \theta(t) = 0$  and the fraction of free proteins  $x_a^f(t=0) = x_a$ . During protein binding, their total number  $x_a = x_a^f + \{\eta(t) + [1 - \eta(t)]\theta(t)\}Mx_t$  is conserved. We describe the dynamics of aspecific binding by a Langmuir process:<sup>14</sup>

$$\frac{d\theta}{dt} = L_+[1 - \theta(t)] - L_-\theta(t) - \frac{d\eta}{dt} \quad (1)$$

where  $L_+$  and  $L_-$  are the time-dependent Langmuir adsorption and desorption rates, respectively. In equilibrium, we have

$$\lim_{t \rightarrow \infty} \theta(t) = \lim_{t \rightarrow \infty} \frac{K_L x_a^f(t)}{1 + K_L x_a^f(t)} \quad (2)$$

with  $K_L = L_-/L_+$  the dimensionless Langmuir binding constant. After binding, proteins can dock together by forming  $\beta$ -rolls along the template.<sup>7,9</sup> We describe this process of folding and cooperative self-assembly with the Hill equation:<sup>15</sup>

$$\frac{d\eta}{dt} = H_+[1 - \eta(t)] - H_-\eta(t) \quad (3)$$

in which the association and dissociation rates  $H_+$  and  $H_-$  of the Hill process are time-dependent. The Hill equation-of-state yields the fraction of coverage of dense and associated capsids at equilibrium:

$$\lim_{t \rightarrow \infty} \eta(t) = \lim_{t \rightarrow \infty} \frac{[K_H \theta(t)]^n}{1 + [K_H \theta(t)]^n} \quad (4)$$

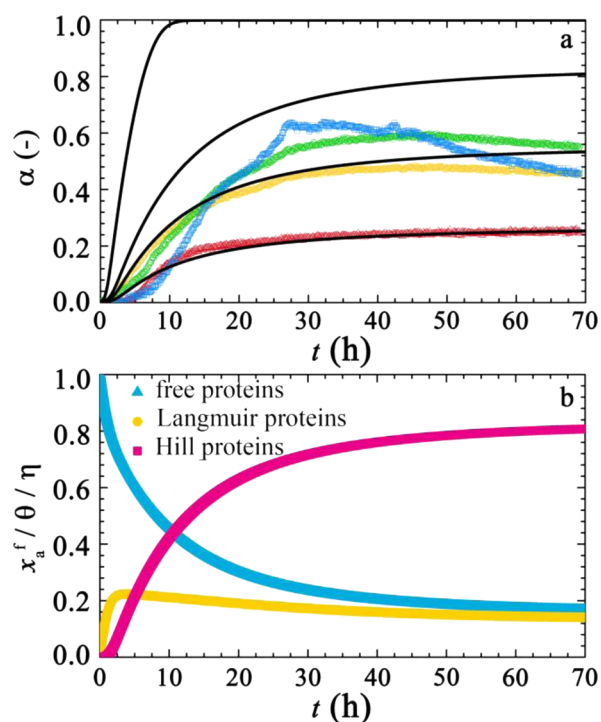
where  $K_H^n = H_+/H_-\theta(t)^n$  is the dimensionless Hill constant associated with cooperative binding and  $n$  the Hill coefficient that controls the degree of cooperativity of the system.<sup>15</sup> We solve these equations numerically using the Runge–Kutta approach<sup>17</sup> (see the [Supporting Information](#)) and compare the results to the experimental data for  $[SP1] = 0.6 \mu\text{M}$ . Each site on the template can exist in three states: (i) naked, (ii) a loosely bound, and (iii) a strongly bound capsid state. The fraction of the template sites coated with loosely bound proteins is  $\theta(t)$  and the fraction of the remaining sites encapsulated in a dense capsid is  $\eta(t)$ . Hence, the overall fraction of occupied sites on the template is

$$F(t) \equiv \eta(t) + [1 - \eta(t)]\theta(t) \quad (5)$$

Adsorbed proteins engaged in Langmuir- and Hill-type adsorption produce different signals in a measurement that probes the occupied fraction of sites. Thus, we weight them with a weighting factor  $w$ :  $F_w(t) \sim \alpha = w\eta(t) + (1 - w)[1 - \eta(t)]\theta(t)$ . The experimental data shows how initial Langmuir binding is followed by a low plateau value of  $\alpha = 0.05$ , after which a second rise increases  $\alpha$  to its final equilibrium value. We interpret this initial plateau at  $\alpha = 0.05$  as the system existing solely in the Langmuir state, where proteins are randomly bound to the template but not yet folded into the beta-sheet structure required for capsid formation. Thus, we take the value of  $w = 0.05$  as an approximate measure for the signal intensity resulting from pure Langmuir bound proteins.

We find quantitative agreement between model and experiment under stoichiometric ratios sufficiently far from charge compensation ( $f_+ = 0.25$  and  $0.40$  in [Figure 4a](#)). Decomposing the signal  $F_w(t)$  into its separate contributions illustrates how free proteins first bind onto the template, followed by their cooperative reorganization into a dense capsid ([Figure 4b](#)). The fraction of Langmuir sites initially increases until it reaches a maximum after which it decreases again. The maximum is located at the end of the nucleation lag time, where the Hill process takes over. Interestingly, the Hill coefficient  $n = 5$  that we need to describe our data indicates that five coat proteins are required to form a critical capsid nucleus. The agreement between model and experiment shows that our presumption of the two-step aggregation process holds at least under certain conditions. Our experiments feature a distinct nucleation lag time. This effect is much weaker in our theoretical model. This is because the model does not feature an energy barrier for forming a nucleus, while this is most likely the case in the experiments. As such, the experiments will feature thermally activated nucleation, while our model does not account for this. Since introducing this feature would increase the number of adjustable parameters, we choose here, for the sake of simplicity, not to introduce this additional term.

At higher values of  $f_+$  (i.e.,  $0.50$  and  $0.70$  in [Figure 4a](#)), as charge compensation is approached, the predicted degree of encapsulation exceeds the value determined experimentally. Clearly, the coassembly pathways are more complex under stoichiometric and superstoichiometric circumstances. A clue to what is happening is provided by earlier experiments on electrostatic complexes of synthetic polymers in the context of what is known as complex coacervation. Indeed, we have very recently observed a phase transition from small and soluble complexes under substoichiometric conditions, to multi-



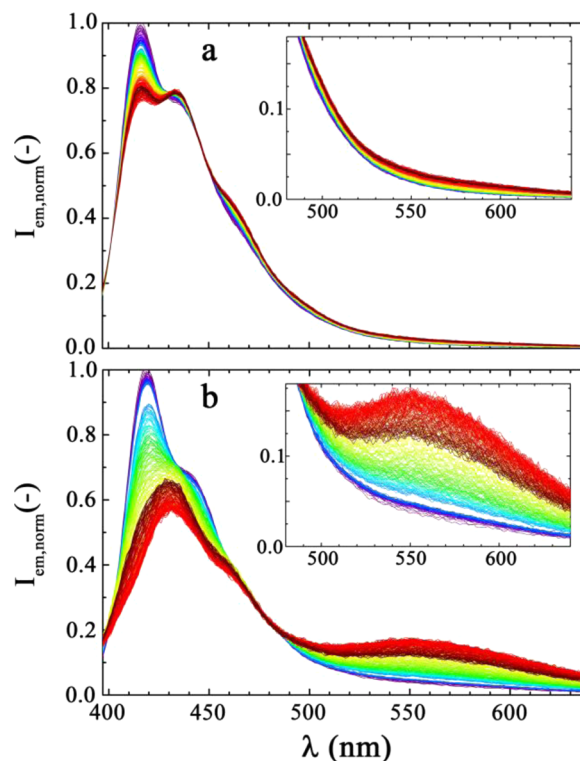
**Figure 4.** (a) Comparison of the experimentally determined value of the coating fraction  $\alpha$  (symbols are experimental data with a legend as in Figure 2d) and predictions by the model as described in the text (solid lines) for  $f_+ = 0.25, 0.40, 0.50,$  and  $0.70$  (from bottom to top). (b) Time-evolution of the fraction of free  $x_a^f$ , Langmuir  $\theta$ , and Hill proteins  $\eta$  for  $f_+ = 0.50$ .

molecular liquidlike structures upon approaching charge neutralization at  $f_+ = 0.5$ .<sup>13a</sup> If this translates to our system then one would expect a similar transition between a capsid containing a single, stretched, genome proxy to one where the capsid contains multiple copies of the template that no longer need to stretch to accommodate the coat proteins and interact electrostatically. We speculate that the gain in configurational entropy of the template offsets the Coulombic penalty of bringing multiple templates into close proximity. Moreover, the latter are screened anyway by the presence of the positively charged tails of the coat proteins.

**Capsid Condensation.** To probe if such bundling indeed occurs and explains the observed deviations from our theoretical predictions, we employ a second molecular sensor, coded SP2. This genome proxy is randomly doped with a small fraction of benzothiadiazole (BT), an acceptor for Förster resonance energy transfer (Figure 1b).<sup>13a,18</sup> In isolation, the fluorescence spectra of this polymer exhibits the same response as its homopolymer equivalent SP1 with no appreciable energy transfer between the fluorene donors and BT acceptors due to its low-doping degree. However, when these sensor polymers condense and bundle or fold, a significant increase in luminescence at 550–600 nm is observed due to intermolecular energy transfer between the fluorene and BT units.<sup>13a,18</sup> Encapsulation of SP2 with our coat protein indeed shows an energy transfer signal emerging under (super)stoichiometric conditions (Figure 1e).

In order to test our coacervation hypothesis, we repeat the capsid assembly studies with the molecular sensor SP2. The normalized PL spectra show the same change in vibronic structure as observed for SP1 but with an additional energy

transfer band due to the BT acceptor, as shown in Figure 5; this is particularly pronounced for high charge stoichiometries, as can be seen in the inset in Figure 5b.

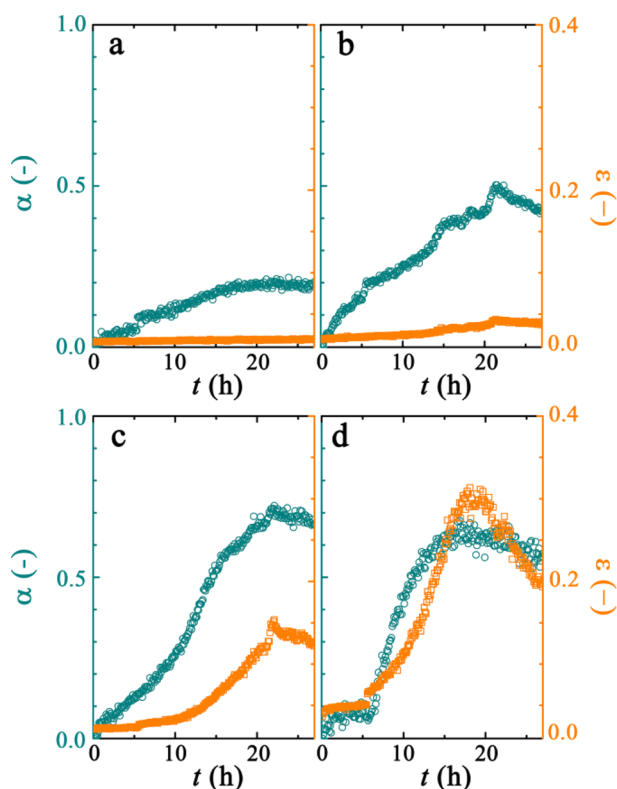


**Figure 5.** Time-evolution of emission spectra, normalized to the total emitted intensity, of molecular sensor SP2 ( $[SP2] = 0.08 \mu\text{M}$ ) during coassembly with C-S<sub>10</sub>-B at  $f_+ =$  (a) 0.3 and (b) 0.5. Insets show a close-up around the emission peak of the energy transfer acceptor benzothiadiazole (BT).

From the ratio of the vibronic bands 1–0 and 2–0, as discussed above for SP1, we find the highest fraction of coated templates is reached at  $\alpha \sim 0.7$ , as shown for different stoichiometries in Figure 6. This is slightly lower than what we find for the homopolymer sensor SP1 under the same conditions. We attribute this small change to a reduction in linear charge density upon introducing uncharged BT moieties at the expense of dicarboxylated fluorene units.

As the point of charge compensation is approached and exceeded, we observe a significant increase in the intensity of the energy transfer band (inset of Figure 5b and Figure 6, panels c and d); this signals the electrostatic condensation transition of the coassembled objects into multitemplate capsids.<sup>13a</sup> We determine the energy transfer efficiency  $\epsilon$  as the ratio of the BT acceptor intensity to that of the fluorene donor. We observe a significant bundling-induced energy transfer as multiple template chains become encapsulated in a single capsid, which becomes increasingly pronounced as the mixing ratio of the template and the protein is increased (Figure 6, panels a–d).

These data suggest that the void of the viral capsid is occupied by more than one template chain at charge-neutral stoichiometry, while each viruslike particle contains a single template at low values of  $f_+$ . Interestingly, the spectroscopic signal for template condensation does not appear until the final kinetic stages of capsid formation. This can be most clearly seen



**Figure 6.** Fraction of encapsulated template chains  $\alpha$  and corresponding energy-transfer ratio  $\epsilon$  between donor (fluorene) and acceptor (BT) for  $f_a =$  (a) 0.10, (b) 0.30, (c) 0.50, and (d) 0.70.

in Figure 6c, where condensation does not occur until about half of the template chains have been encapsulated. This indicates that a third kinetic phase emerges in assembly pathways, during which nucleated capsids around a single template undergo a condensation transition to multigenome objects.

## CONCLUSIONS

Our results reveal the pathway complexity of the coassembly of a viromimetic protein and a genome proxy, hinting that similar pathway complexity underlies the formation of natural viruses. Our data and theoretical analysis suggests that for our experimental model for linear viruses, capsid assembly follows a distinct sequence of steps: (i) coat proteins bind randomly to the template and (ii) bound proteins reorganize and fold until a multiprotein nucleus is formed and a dense capsid grows. Interestingly, a similar sequence of events has been evidenced in vitro for spherical viruses.<sup>19</sup> (iii) Under conditions of (super)stoichiometry, a third stage emerges in which single-genome viruslike particles condense to form objects that contain multiple copies of the template. Whether this is relevant in the biological context is unclear, but it would be interesting to test this using the experimental strategy we follow. Our approach relies on the planarization of the sensor molecule in a linear capsid cavity. Engineering the molecular design of the sensor polymer (e.g., by introducing donor–acceptor moieties or biomolecule-specific binding sites) could open up possibilities to explore the assembly pathways of more complex and naturally occurring viral structures such as the tobacco mosaic virus.

## ASSOCIATED CONTENT

### Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/jacs.7b01401.

Figure S1 details synthesis protocols and characterization of the materials, Figures S2 and S3 provide additional excitation–emission spectra of the sensor polymers and lifetime measurements, Figures S4 and S5 give the protein sequence and MALDI-TOF measurements, and additional information on the methods is provided (PDF)

## AUTHOR INFORMATION

### Corresponding Author

\* joris.sprakel@wur.nl

### ORCID

Daniela J. Kraft: 0000-0002-2221-6473

Joris Sprakel: 0000-0001-6532-4514

### Author Contributions

The manuscript was written through contributions of all authors, and all authors have given approval to the final version of the manuscript.

### Notes

The authors declare no competing financial interest.

## ACKNOWLEDGMENTS

This work was financially supported by a European Research Council Advanced Grant (ERC-267254) and The Netherlands Organisation for Scientific Research (NWO) through projects 712.012.007 and 680-47-431.

## REFERENCES

- (1) (a) Philp, D.; Stoddart, J. F. *Angew. Chem., Int. Ed. Engl.* **1996**, *35*, 1154. (b) Whitesides, G. M.; Grzybowski, B. *Science* **2002**, *295*, 2418.
- (2) (a) Packianathan, C.; Katen, S. P.; Dann, C. E.; Zlotnick, A. *J. Virol.* **2010**, *84*, 1607. (b) Stockley, P. G.; Rolfsson, O.; Thompson, G. S.; Basnak, G.; Francese, S.; Stonehouse, N. J.; Homans, S. W.; Ashcroft, A. E. *J. Mol. Biol.* **2007**, *369*, 541. (c) Zlotnick, A. *J. Mol. Biol.* **1994**, *241*, 59. (d) Zlotnick, A.; Mukhopadhyay, S. *Trends Microbiol.* **2011**, *19*, 14.
- (3) (a) Bruinsma, R. F.; Gelbart, W. M.; Reguera, D.; Rudnick, J.; Zandi, R. *Phys. Rev. Lett.* **2003**, *90*, 248101. (b) Misra, N.; Lees, D.; Zhang, T.; Schwartz, R. *Comput. Math Methods Med.* **2008**, *9*, 277. (c) Schwartz, R.; Shor, P. W.; Prevelige, P. E.; Berger, B. *Biophys. J.* **1998**, *75*, 2626. (d) Zandi, R.; van der Schoot, P.; Reguera, D.; Kegel, W.; Reiss, H. *Biophys. J.* **2006**, *90*, 1939.
- (4) Hagan, M. F.; Chandler, D. *Biophys. J.* **2006**, *91*, 42.
- (5) (a) Korevaar, P. A.; George, S. J.; Markvoort, A. J.; Smulders, M. M.; Hilbers, P. A.; Schenning, A. P.; De Greef, T. F.; Meijer, E. *Nature* **2012**, *481*, 492. (b) Korevaar, P. A.; de Greef, T. F. A.; Meijer, E. W. *Chem. Mater.* **2014**, *26*, 576. (c) Besenius, P. *J. Polym. Sci., Part A: Polym. Chem.* **2017**, *55*, 34.
- (6) (a) Uetrecht, C.; Barbu, I. M.; Shoemaker, G. K.; Van Duijn, E.; Heck, A. J. *Nat. Chem.* **2011**, *3*, 126. (b) Zlotnick, A.; Johnson, J. M.; Wingfield, P. W.; Stahl, S. J.; Endres, D. *Biochemistry* **1999**, *38*, 14644.
- (7) Hernandez-Garcia, A.; Kraft, D. J.; Janssen, A. F.; Bomans, P. H.; Sommerdijk, N. A.; Thies-Weesie, D. M.; Favretto, M. E.; Brock, R.; De Wolf, F. A.; Werten, M. W.; van der Schoot, P.; Stuart, M. C.; de Vries, R. *Nat. Nanotechnol.* **2014**, *9*, 698.
- (8) (a) Caspar, D. L.; Namba, K. *Adv. Biophys.* **1990**, *26*, 157. (b) Klug, A. *Philos. Trans. R. Soc., B* **1999**, *354*, 531. (c) Namba, K.; Pattanayek, R.; Stubbs, G. *J. Mol. Biol.* **1989**, *208*, 307.
- (9) Zhao, B.; Stuart, M. A. C.; Hall, C. K. *Soft Matter* **2016**, *12*, 3721.

- (10) Werten, M. W.; Wisselink, W. H.; Jansen-van den Bosch, T. J.; de Bruin, E. C.; de Wolf, F. A. *Protein Eng., Des. Sel.* **2001**, *14*, 447.
- (11) Punter, M. T.; Hernandez-Garcia, A.; Kraft, D. J.; de Vries, R.; van der Schoot, P. *J. Phys. Chem. B* **2016**, *120*, 6286.
- (12) Cingil, H. E.; Storm, I. M.; Yorulmaz, Y.; te Brake, D. W.; de Vries, R.; Cohen Stuart, M. A.; Sprakel, J. *J. Am. Chem. Soc.* **2015**, *137*, 9800.
- (13) (a) Cingil, H. E.; Boz, E. B.; Wang, J.; Stuart, M. A. C.; Sprakel, J. *Adv. Funct. Mater.* **2016**, *26*, 1420. (b) Zhang, Y.; Liu, B.; Cao, Y. *Chem. - Asian J.* **2008**, *3*, 739.
- (14) Langmuir, I. *J. Am. Chem. Soc.* **1918**, *40*, 1361.
- (15) Hill, T. L. *J. Chem. Phys.* **1946**, *14*, 263.
- (16) Hu, T.; Shklovskii, B. I. *Phys. Rev. E* **2007**, *75*, 051901.
- (17) Butcher, J. C. *J. Austral. Math. Soc.* **1963**, *3*, 185.
- (18) Liu, B.; Bazan, G. C. *Chem. Mater.* **2004**, *16*, 4467.
- (19) Garmann, R. F.; Comas-Garcia, M.; Gopal, A.; Knobler, C. M.; Gelbart, W. M. *J. Mol. Biol.* **2014**, *426*, 1050.