

Supplementary Information: Identification of a major intermediate along the self-assembly pathway of an icosahedral viral capsid by using an analytical model of spherical patch

Didier Law-Hine,^a Mehdi Zeghal,^a Stéphane Bressanelli,^b Doru Constantin^{*a} and Guillaume Tresset^{*a}

Experiment reproducibility

The mixing achieved with the stopped-flow device was found to be of high quality and the TR-SAXS patterns were very reproducible (Fig. 1).

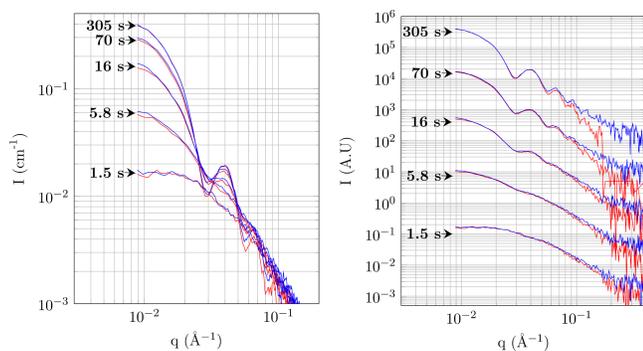


Fig. 1 Reproducibility of the TR-SAXS experiments. (Left) The two sets of kinetic data collected at a protein concentration of 1.49 mg/mL superimpose very well at low q -values. (Right) Same patterns shifted for clarity. One of the experiments ends up at 305 s.

Dimers

Figure 2 demonstrates that the objects at the beginning of the self-assembly process were dimers. The first pattern of the kinetics was compared to the form factor of dimers measured separately in static experiments and to the form factor of dimers calculated from their crystal structure. The three patterns are in excellent agreement.

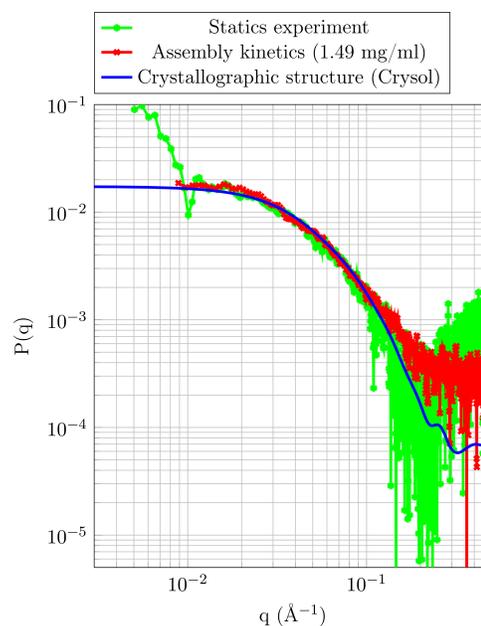


Fig. 2 Form factors $P(q)$ of CCMV capsid dimers. The red curve is the first pattern of the assembly kinetics at protein concentration of 1.49 mg/mL. The green curve is the form factor of dimers measured separately in a buffer solution at pH 7.5 and ionic strength 0.5 M. The blue curve is a form factor calculated by the CRY SOL package¹ from the crystal structure of a CCMV dimer (PDB 1ZA7). Error bars are removed for the sake of clarity.

^a Laboratoire de Physique des Solides, CNRS, Univ. Paris-Sud, Université Paris-Saclay, 91405 Orsay Cedex, France. Tel: +33 1 69 15 53 60; E-mail: doru.constantin@u-psud.fr and guillaume.tresset@u-psud.fr

^b Institute for Integrative Biology of the Cell (I2BC), CEA, CNRS, Univ. Paris-Sud, Université Paris-Saclay, 91198 Gif-sur-Yvette Cedex, France.

Model of empty capsid

Figure 3 shows a vesicle model used to describe the form factor of empty capsids. The volume of the capsid could be then calculated as well as the volume of the dimer. Since the capsid volume is 90 times higher than the dimer volume, the latter was found to be $V_{\text{dimer}} = 79\,100\ \text{\AA}^3$.

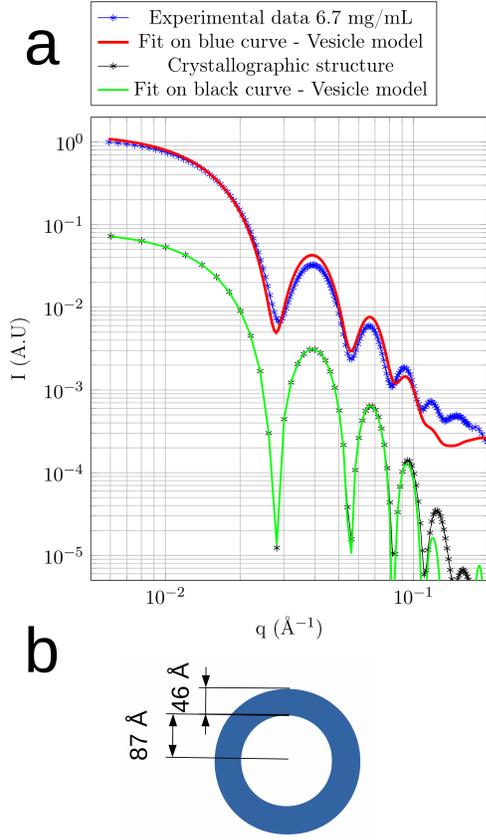


Fig. 3 Empty capsids modelled as vesicles. (a) The fit with a vesicle model is carried out on two sets of data: from the crystal structure (black) and from a static experiment (blue) at 6.7 mg/mL (pH 5.25, $I = 0.5$ M). The fit on crystallographic data (green) is obtained from a simple vesicle model and yields an inner radius of 89 Å and a thickness of 44 Å. The fit on static experimental data includes a Schulz distribution on the inner radius with constant outer-to-inner radii ratio. The inner radius is 87 Å, the thickness 46 Å and the polydispersity 0.08. (b) The corresponding vesicle model with inner radius 87 Å and thickness 46 Å.

Construction of the analytical model for the intermediate

Stereographic projection

Using the notations of Fig. 1 in the article, simple geometrical considerations lead to

$$\theta_c(\phi) = 2 \arctan \left(\frac{\rho(\phi)}{2R} \right)$$

where $\rho(\phi)$ denotes the radial distance in the 2D plane of the ellipse. For an ellipse of semi-axes a and b ($b > a$), we have

$$\left. \begin{aligned} \frac{x^2}{a^2} + \frac{y^2}{b^2} &= 1 \\ x &= \rho(\phi) \cos \phi, \quad y = \rho(\phi) \sin \phi \\ \rho(\phi) &= \frac{ab}{\sqrt{b^2 \cos^2 \phi + a^2 \sin^2 \phi}} \end{aligned} \right\}$$

Scattering intensity and form factor

Constantin showed that the scattering intensity can be expressed as²

$$I(q) = 4\pi \sum_{l=0}^{+\infty} \sum_{m=-l}^{+l} |c_{lm}|^2 |b_l(q)|^2 \quad (1)$$

The basis functions $b_l(q)$

$$\begin{aligned} b_l(q) &= \int_0^\infty a(r) r^2 j_l(qr) dr \\ &= \frac{1}{q^3} \int_0^\infty a(z) j_l(z) z^2 dz, \quad (z = qr) \\ &= \frac{1}{q^3} \int_{z_{min}}^{z_{max}} j_l(z) z^2 dz, \quad (z_{max} = qR_{max}, z_{min} = qR_{min}) \end{aligned} \quad (2)$$

The last integral admits an analytical expression yielding

$$b_l(q) = \frac{1}{q^3} \left[\frac{\sqrt{\pi}}{2^{l+2}} z^{l+3} \Gamma\left(\frac{l+3}{2}\right) {}_1\tilde{F}_2\left(\frac{l+3}{2}; \frac{2l+3}{2}, \frac{l+5}{2}; -\frac{z^2}{4}\right) \right]_{z_{min}}^{z_{max}}$$

where Γ is the Gamma function and ${}_1\tilde{F}_2$ is a regularized hypergeometric function.

Coefficients c_{lm}, c_l

The coefficients c_{lm} result from the projection of the ellipse of semi-axis (a, b) on the sphere of radius R and contain the information on the shape of the intermediate. They are defined in terms of the scalar product

$$\begin{aligned} c_{lm} &= \langle Y_{lm}(\Omega), F(\Omega) \rangle \\ &= \int_{\Omega} Y_{lm}^*(\Omega') F(\Omega') d\Omega' \end{aligned}$$

F is the characteristic function of the object of interest defined by:

$$F(\Omega) = F(\theta, \phi) = \begin{cases} 1, & \text{if } \theta \in [0, \theta_c(\phi)] \\ 0, & \text{otherwise} \end{cases}$$

Thus,

$$c_{lm} = \int_{\phi=0}^{2\pi} \int_{\theta=0}^{\theta_c(\phi)} \sqrt{\frac{2l+1}{4\pi}} \sqrt{\frac{(l-m)!}{(l+m)!}} P_l^m(\cos \theta) e^{-im\phi} \sin \theta d\theta d\phi \quad (3)$$

Since the basis functions do not depend on m , we can introduce the coefficients c_l defined by

$$c_l = \sum_{m=-l}^l |c_{lm}|^2$$

Calculation of $c_{l(-m)}$: From Eq. 3 and the properties of the associated Legendre polynomials,

$$c_{l(-m)} = (-1)^m c_{lm}^*$$

Hence,

$$c_l = \sum_{m=0}^l |c_{lm}|^2 + \sum_{m=-l}^0 |(-1)^m c_{lm}^*|^2 - |c_{l0}^2|$$

$$= 2 \sum_{m=0}^l |c_{lm}|^2 - |c_{l0}^2|$$

Odd values of m : Still from Eq.3, due to the term $e^{-im\phi}$ and because the object is symmetrical, every odd value of the index m leads to $c_{lm}=0$.

These two points reduce considerably the computation of the coefficients c_{lm} and thus c_l (by a factor of 4).

Form factor

The normalized form factor is defined as:

$$P(q) = \frac{\sum_l c_l |b_l(q)|^2}{\sum_l c_l |b_l(0)|^2}$$

From Eq. 1 and using $I(q) = V_{\text{int}}^2 P(q)$, where V_{int} is the volume of the intermediate, we have

$$V_{\text{int}}^2 = 4\pi \sum_l c_l |b_l(0)|^2$$

and $P(q)$ can be expressed by

$$P(q) = \frac{4\pi}{V_{\text{int}}^2} \sum_l c_l |b_l(q)|^2$$

Global fitting

The global fitting procedure used in this study was carried out as follows:

1. The stoichiometric coefficients to be tested, i.e., α , β and γ , were first set according to the kinetic model and the fit parameters to be calculated were $\{\xi_i\} = \{k_i^-, k_i^+, \mu_i, v\}$.
2. Several initial sets of parameters $\{\xi_i^0\}$ were randomly picked (uniform distribution) in their chosen range of values defined with lower and upper bounds $\{\xi_i^{\min}, \xi_i^{\max}\}$.
3. A loop over all the parameters $\{\xi_i\}$ was implemented using Sequential Quadratic Programming (SQP) and Interior-Point (IP) algorithms. The matrix of concentrations $\mathbf{c}_{\text{cat}}(\{k_i^-, k_i^+, \mu_i\})$ was calculated by solving the coupled differential equations. The matrix \mathbf{P}' was computed as a function of v . The residual χ^2 that made the relationship $\mathbf{I} \approx \mathbf{I}_{\text{cat}}^{\text{fit}}$ optimal was then stored as well as the corresponding R -factor.
4. Steps 2 and 3 were then repeated for other sets of stoichiometric coefficients.

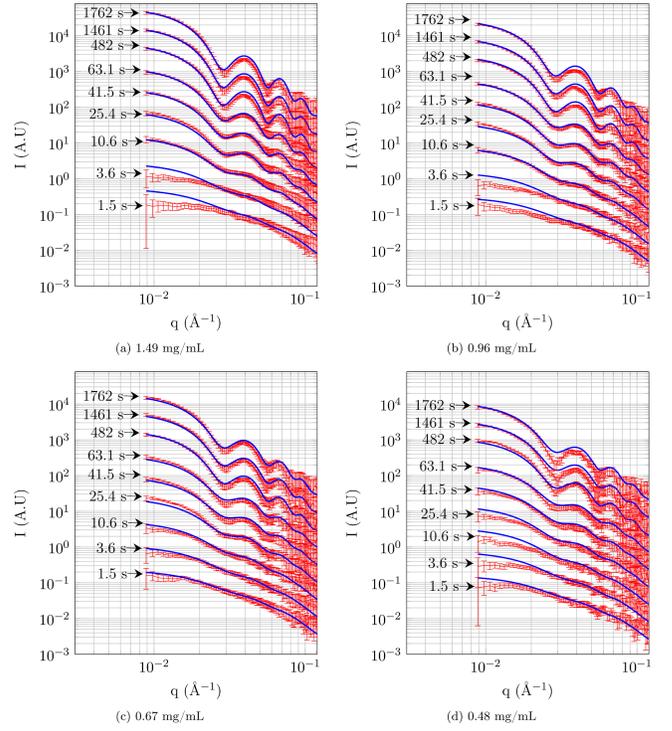


Fig. 4 Results of the global fitting for the three-state model with $\alpha = 37$ and $\beta = 1$. The experimental scattering intensities (red line) are compared to those reconstructed from the kinetic model (blue line) at four different protein concentrations. The residual is $\chi^2 = 1.67$.

Nucleation-elongation and three-state cooperative models

n	χ^2	R -factor (%)
2	28.75	29.9
5	23.51	61.7
10	20.08	56.0
18	16.30	49.5
24	13.68	44.2
30	11.41	40.5
36	8.79	32.9
40	7.34	30.9
44	6.02	27.1
48	5.42	24.5
54	3.81	19.9
60	3.20	20.1
66	2.23	14.9
70	3.11	20.3

Table 1 Residuals obtained from a global fitting of the data with a nucleation-elongation kinetic model for various intermediate size n .

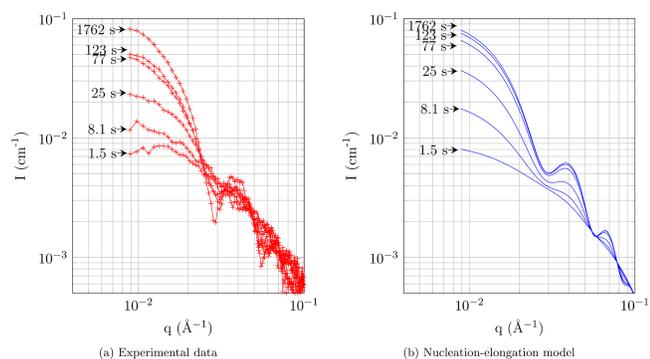


Fig. 5 Example of data fitting with the nucleation-elongation model. (a) Experimental data at 0.48 mg/ml. (b) Kinetic model that best fits the data with $n=66$ ($\chi^2=2.23$).

α	β	γ	χ^2	R-factor (%)
2	24	42	12.69	39.8
2	45	0	12.85	38.2
5	12	30	10.85	37.7
10	3	60	4.72	24.6
10	6	30	7.56	30.6
18	3	36	4.14	21.9
18	4	18	5.50	24.7
24	3	18	3.90	20.0
30	2	30	2.47	16.9
30	3	0	3.81	19.5
36	2	18	2.33	14.8
38	2	14	2.98	16.2
40	2	10	2.21	14.5
44	2	2	2.24	14.1

Table 2 Three-state cooperative model ($\beta > 1$). Residuals from global fitting for different combinations $\{\alpha, \beta, \gamma\}$.

References

- 1 D. I. Svergun, C. Barberato and M. H. J. Koch, *J. Appl. Crystallogr.*, 1995, **28**, 768–773.
- 2 D. Constantin, *J. Appl. Crystallogr.*, 2015, **48**, 1901–1906.