

Programmable Polymorphism of a Virus-Like Particle

Artur Biela

Jagiellonian University <https://orcid.org/0000-0001-6733-242X>

Antonina Naskalska

Jagiellonian University

Farzad Fatehi

University of York

Reidun Twarock

University of York

Jonathan Heddle (✉ jonathan.heddle@uj.edu.pl)

Jagiellonian University <https://orcid.org/0000-0003-0994-9928>

Article

Keywords: MS2-VLP, bacteriophage, polymorphism, cryoEM, virus-like particles

Posted Date: September 28th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-668750/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Version of Record: A version of this preprint was published at Communications Materials on February 7th, 2022. See the published version at <https://doi.org/10.1038/s43246-022-00229-3>.

Programmable Polymorphism of a Virus-Like Particle

Artur P. Biela^{1‡}, Antonina Naskalska^{1‡}, Farzad Fatehi^{2,3}, Reidun Twarock^{2,3,4}, Jonathan G. Heddle^{1}*

¹ Malopolska Centre of Biotechnology, Jagiellonian University, Gronostajowa 7A, 30-392 Krakow, Poland

² Departments of Mathematics, University of York, York YO10 5DD, UK

³ York Cross-Disciplinary Centre for Systems Analysis, University of York, York YO10 5GE, UK

⁴ Department of Biology, University of York, York YO10 5DD, UK

‡ these authors contributed equally

*Email: jonathan.heddle@uj.edu.pl

KEYWORDS

MS2-VLP, bacteriophage, polymorphism, cryoEM, virus-like particles

ABSTRACT

Virus-like particles (VLPs) have significant potential as both artificial vaccines and drug delivery systems. The ability to control their size has wide ranging utility, but achieving such controlled polymorphism using a single protein subunit is challenging as it requires altering VLP geometry. Here we achieve size control of MS2 bacteriophage VLPs via insertion of amino acid sequences

in an external loop to shift its morphology to significantly larger forms. The resulting VLP size and geometry is controlled by altering the length and type of the insert used. Cryo-EM structures of the new VLPs, in combination with a kinetic model of their assembly, show that the abundance of wild type (T=3), T=4, D3 and D5 symmetrical VLPs can be controlled in this way. We propose a mechanism whereby the insert leads to a change in the dynamic behavior of the capsid protein dimer, affecting the interconversion between the symmetric and asymmetric conformers and thus determining VLP size and morphology.

INTRODUCTION

Protein cages, convex polyhedral protein containers self-assembled from multiple copies of identical protein subunits, are pillars of nanotechnology. They include naturally occurring virus capsids (virus-like particles, VLPs) as well as a plethora of particles with diverse symmetries. Given their structures, protein cages are obvious candidates for development as vaccines through addition of relevant antigens on their exterior surfaces) and as drug delivery systems (through encapsulation of relevant therapeutic molecules in their hollow cores)¹⁻³. Additionally, VLPs have an innate ability to encapsulate nucleic acids, which make them attractive as DNA/RNA delivery vehicles. However, to date, they have not been approved as therapeutic agents other than as vaccines. Two such vaccines (protecting against HBV and HPV infections) are currently approved for use in humans while many others are under clinical trial (reviewed in⁴).

The ability to control VLP particle size is attractive as it allows the same constituent protein components to be repurposed in VLPs having different morphologies and properties. Increasing of VLP size is particularly beneficial, resulting in both greater surface area for increased antigen

presentation and larger volume with increased cargo capacity. However, this is challenging to achieve given that VLPs are made from multiple proteins, each forming a number of distinct bonds with their neighbors. As a consequence, changing the number of subunits and the angle between them, as would be necessary to increase overall diameter, is difficult without disrupting the entire structure. Nevertheless, polymorphism is well established amongst naturally occurring viruses/VLPs^{5,6}. It has been reported, for example, that the VLP derived from bacteriophage P22 can expand or contract as a result of heating to 65 °C or following treatment with sodium dodecyl sulfate, resulting in the opening and closing of pores in the particle which can be used for cargo entry or egress⁷. The SV40 virus capsid, formed from major capsid protein VP1, are able to switch between $T=7d$, $T=1$ and tubular structures depending on salt concentration and pH^{8,9}. Lumazine synthase, a protein cage of bacterial origin, has also shown a similarly wide range of different forms, including cages of different sizes and even a protein tube depending on buffer conditions¹⁰. In all of these cases, the spectrum of variant morphologies, and how assembly can be biased to achieve desired structural outcomes, are unknown.

MS2 is an icosahedral RNA bacteriophage measuring 22–29 nm in diameter¹¹. Its $T=3$ capsid is built from 180 copies of a single coat protein (CP). The RNA genome (~3.6 kb) encodes a further three proteins: the maturation protein (A-protein), the replicase, and the lysis protein. When CP is expressed from a plasmid in a bacterial or yeast host, in the absence of other viral elements, it self-assembles into VLPs,¹² dimerizing in the process. An unstructured loop known as the FG-loop connects the F and G β -strands of each subunit. The 90 dimers forming the VLP can be classified into two structurally different groups: one group contains 60 asymmetric dimers (called A/B dimers) where the FG-loop of the B monomer adopts a different conformation from that in the A monomer. In the remaining 30 dimers (C/C dimers) the FG-loops of both adopt the same

conformation¹³ (**Figure 1a**). Binding of RNA stem-loops from within the viral genome can trigger the conformational change from the symmetric to the asymmetric dimer conformation via a displacement of the kinetic energy favoring conversion from the C/C to the A/B conformer^{14,15}. Multiple dispersed RNA stem-loops in the genome with a shared sequence motif, termed packaging signals, cooperatively ensure efficient virus assembly¹⁶⁻²². The differences in FG-loop conformation are a determinant of local symmetry of the MS2 capsid shell. A/B dimers are located in groups of five around the particle 5-fold axes and are responsible for pentamer formation. This is a consequence of the fact that in the B monomer the FG-loop is flipped (away from the cage lumen) which allows closer packing of coat proteins around the 5-fold axes, enabling the 5-fold symmetry arrangement of five B subunits. In the alternative monomer conformation, the FG-loop accommodates a rigid β hairpin-like structure and allows packing only in groups of six dimers around the particle 3-fold axes. C/C dimers are symmetrical, with both FG-loops in hairpin conformation, and participate in neighboring 6-fold clusters, thus acting like a staple between them. According to Euler's theorem, exactly 12 five-fold clusters are needed to build a closed shell, and therefore for the protein capsids we observe, both types of dimers are required to form a closed protein capsid.

MS2 VLPs have been used for a number of applications, including encapsulation of heterologous nucleic acids to serve as RNA "armors" in preparation of internal controls for qPCR²³⁻²⁵; as delivery vehicles^{3,26,27}; for incorporation of immunogenic epitopes displayed on the particle surface^{12,28-32}; and development of the so-called MS2-tagging system^{33,34}, which allows the study of RNA-protein interactions in cells. Recently MS2 VLPs carrying fragments of the SARS-CoV-2 genome have been developed as control standards for diagnostic tests³⁵.

In this work we show that inserting amino acid sequences into an externally facing loop of one monomer of the MS2 capsid dimer not only provides sites for attachment of arbitrary proteins of interest but, through its effect on the FG-loop conformation, alters the pentamer/hexamer equilibrium leading to a shift towards larger capsids being formed (**Figure 1b**). Moreover, we explain systematically, using a kinetic model, how different MS2 VLP geometries form (**Figure 1c-d**) and how changes engineered in the capsid protein structures impact on the abundance of the “wild type” form of the MS2 capsid and its variants (**Figure 1e**).

RESULTS

Design and production of modified MS2 VLPs

We designed MS2 VLPs to provide an externally displayed anchor for attachment of proteins/peptides of choice while simultaneously investigating effects on VLP morphology. To achieve this, we utilized the Spytag - Spycatcher system³⁶. This consists of a 13-residue peptide (called Spytag) and a 116-residue complementary domain (called Spycatcher). When the two are mixed, the holo Spy protein is spontaneously reconstituted. This technology has previously been used to create VLPs displaying immune epitopes or cellular receptor ligands for further development of vaccines or delivery vehicles, respectively³⁷. We designed MS2 CP constructs as tandem dimer fusions, which is known to confer tolerance to peptide insertions in one monomer without interfering with particle assembly³⁸. We inserted the Spytag sequence at the *KpnI* (144G) position (**Figure 1a**) as has been used for His-tag insertion³⁹. In order to optimize the Spytag exposure at the particle surface (and thus confer best availability for the Spycatcher attachment) we also created variants with longer flanking linkers. The designed constructs were generated by cloning (all primer sequences are listed in **Table S1**) and verified by sequencing (**Supplementary**

Data S1). Proteins were expressed in *E. coli* and the resulting VLPs were extracted and purified as described in Materials and Methods. These variants of MS2 VLPs were named “SpyTag”, “SpyTag4”, “SpyTag7” and “Random4”, reflecting the characteristics of the insert sequence as further described below.

Characterisation of MS2 VLPs

Purified MS2 VLPs were initially characterized using size exclusion chromatography (SEC), native PAGE and SDS-PAGE. For all VLP variants (MS2-SpyTag, MS2-SpyTag4, MS2-SpyTag7 and MS2-Random4) four peaks were typically present in SEC profiles (**Figure 2a**): peak 2 corresponds to properly assembled particles in the MDa range as confirmed by native-PAGE (**Figure 2b**). The other peaks (1, 3 and 4) are present at different levels depending on the variant (not shown) and likely consist of unassembled or partially assembled MS2 CP (~29 kDa, see SDS-PAGE, **Figure 2b**).

All variants of purified VLPs, collected from peak 2 fractions, were analyzed using DLS. We noted that for particles carrying longer insertions (MS2-SpyTag4, MS2-SpyTag7 and MS2-Random4) the mean hydrodynamic diameter was smaller than for the SpyTag variant (**Figure 2c**). Moreover, when size distributions are analyzed by intensity, the shape of the peaks indicates the possibility of more than one population of particles with similar but not identical hydrodynamic radii (**Figure 2c**). Further investigation using negative stain TEM confirmed the expected VLP structure for all produced proteins and also confirmed differences in diameter depending on the presence and length of the linker (**Figure 2d**), with diameter magnitudes observed in the order MS2-SpyTag4 \approx MS2-Random4 $>$ MS2-SpyTag7 $>$ MS2-SpyTag.

Attachment of external proteins

In order to confirm the presence of Spytag at the expected location, being both external and accessible, we attached the Spycatcher-mCherry fusion protein to the assembled VLP. MS2 Spytag VLP variants (MS2-SpyTag, MS2-SpyTag4, MS2-SpyTag7) were incubated with Spycatcher-mCherry protein and then analyzed by SDS-PAGE (**Figure 3a**). Bands corresponding to CP dimers conjugated to the Spycatcher-mCherry protein can be seen in the case of VLP variants where Spytag is flanked by linkers (MS2-SpyTag4, MS2-SpyTag7), but not in the case of MS2-SpyTag. As a next step, reaction mixtures containing both protein moieties were separated using size exclusion chromatography, allowing removal of the free Spycatcher-mCherry protein and providing additional confirmation that MS2 particles are efficiently decorated (**Figure 3b**). Finally, purified MS2 VLPs decorated with Spycatcher-mCherry protein were subjected to DLS measurement (**Figure 3c**) and inspected using transmission electron microscopy (TEM, **Figure 3d**) with results demonstrating the integrity of the particles as well as an increase in their volumes.

Cryo-EM structures of modified MS2 VLPs

All MS2 VLP variants underwent full cryoEM analysis (see Material and Methods for details, (**Figure S1-4**)). In total, 16 structures were identified and determined: three for MS2-SpyTag (**Figure S1 and S6**) with $T=3$, D_5 and $T=4$ symmetries, four for MS2-SpyTag4 (**Figure S2 and S7**) and MS2-SpyTag7 (**Figure S3 and S8**) with $T=3$, D_5 , D_3 ($D3-A$) and $T=4$ symmetries, and five for MS2-Random4 (**Figure S4 and S9**) with $T=3$, D_5 , D_3 ($D3-A$ and $D3-B$) and $T=4$ symmetries. The different symmetries imply that there are different ratios between the numbers of local 5- and 6-fold clusters in the corresponding particle morphologies. As the B monomers form the 5-fold clusters and the A and C monomers the 6-fold clusters in the MS2 capsid, this suggests

that the relative ratios between A/B and C/C dimers must be different, and that this ratio can be used to control particle morphology. All cryoEM reconstructions were deposited in EMDB – see **Table S2**.

Since the monomers in MS2 dimers either partake in a 5-fold clusters (in the case of the B monomer) or in a 6-fold cluster (in the case of the A and C monomers), any capsid geometry assembled from such building blocks must be geometrically identical to fullerene structures, with fullerene edges mapping onto MS2 coat proteins dimers. Analysis of the structures obtained in the experiment showed that in cases where there is more than one possible fullerene isomer, MS2 particles accommodate geometries related to the most energetically stable one (**Figure S5**) and with the highest possible symmetry.

Geometrical interpretation of the results

The five distinct types of particles observed must each have 12 pentagonal clusters according to Euler's theorem, and therefore precisely 60 A/B dimers. The numbers of C/C dimers, on the other hand, differ for distinct particle types (**Table S3**). The ensemble of particles observed for each SpyTag variant therefore also differs in its cumulative A/B:C/C dimer content (**Table S4**). However, the relative ratio of A/B:C/C dimers is not sufficient to explain the observed particle distribution, because the same ratio could also be used, from a purely stoichiometric perspective, to construct different combinations of particles. This is because the numbers of A/B and C/C dimers in two D5 particles are the same as for a $T=3$ plus a $T=4$ particle, and similarly two D3-A particles are equivalent to a D5 plus a D3-B particle (**Figure 1c-d**). For each SpyTag variant, we computed all possible ensembles consistent with the observed numbers of A/B and C/C dimers (**Table S4**), demonstrating that the scenarios seen in the experiment do not favor the higher

symmetry particles ($T=3$ plus $T=4$) over D5 particles. This implies that a thermodynamic argument alone is not sufficient to understand the experimental outcome.

Assembly kinetics reveals the origin of distinct particle morphologies

We constructed an assembly model based on reaction kinetics, that includes the interconversion between the C/C and A/B dimers, as well as the association of dimers according to a tree that indicated bifurcations in the assembly pathways resulting in the observed particle geometries (**Figure S10**). Under otherwise identical conditions, different interconversion rates between C/C and A/B dimers result in the distinct particle morphologies observed for SpyTag, Random4, and SpyTag7 (**Table S5**) suggesting that conversion from the C/C to the A/B conformer is indeed differentially affected by the Spytag insertions. SpyTag4 is the most affected with the lowest conversion rate, consistent with presenting the highest yield of $T=4$ particles for any of the cases. The rate consistent with obtaining mostly native $T=3$ particles is significantly higher than those obtained here, suggesting that Spytag insertions indeed have a major impact on the conversion from C/C to A/B compared with wild type virus (**Table S6**). Interestingly, the length of the SpyTag4 and Random4 inserts is identical, suggesting that insert size alone is not a predictor of the experimental outcome.

DISCUSSION

In this work we have demonstrated that through engineering of the CP protein, we are able to exert a level of control over the diameter and geometry of the resulting, assembled MS2 VLPs. Experimental validation was achieved by determining 16 cryo-EM structures from the designed protein variants, five of which are unique in terms of their overall geometry. We have been able to

identify and confirm a previously reported⁴⁰ MS2-like capsid with $T=4$ geometry. In addition, we were able to solve structures of three previously unreported MS2 capsid variants: one with D_5 , and two with D_3 symmetry (D3-A and D3-B).

In order to understand how the insertion of additional amino acid sequences in an external loop triggers this effect, we developed a kinetic model of particle assembly. This suggests that the difference in experimental outcome for distinct Spytag inserts is due to a delay in conversion from C/C to A/B, which could be mediated by a change in the dynamic behavior of the protein dimer as a result of the insertions. This would be consistent with our understanding that the dynamic properties of the C/C dimer affect the conversion to A/B, and that this shift can be triggered by binding to an RNA packaging signal¹¹. However, whilst packaging signals promote this shift, Spytag insertions appear to disfavor it. As a notable example, the largest difference between ideal and actual scenarios (**Table S8**) is seen for SpyTag and SpyTag4. Our model suggests that this is because they are slower in converting C/C into A/B dimers than SpyTag7 and Random4. Note that they both share the same AHIVMVDAYKPTK insertion, with the flanking GGGS in SpyTag4, apparently accounting for the slower conversion. In contrast, the longer flanks of SpyTag7 and randomisation of the inserted region have a less marked effect. In agreement with experimental results, our analysis suggests that production of desired particle morphologies can be achieved via the choice of SpyTag insert, affording unprecedented control over the production of the desired particle morphologies.

CONCLUSION

In this work we have shown that a VLP can be programmed to produce particles of increased diameters and altered symmetries. Given the proven and potential applications of MS2¹⁻³ and other

VLPs in virus nanotechnology the ability to produce containers of larger volume with enhanced carrying capacity is highly desirable. Further, understanding how the inserts alter assembly kinetics and the resulting VLP morphology gives a deeper control of VLP assembly overall. Taken together, these results are a first essential step towards production of bespoke VLPs with desired morphologies and assembly/disassembly properties.

METHODS

Cloning, expression and purification of MS2 VLPs.

The MS coat protein (CP) gene was designed as a tandem dimer, with the second part of the dimer bearing a *KpnI* restriction site allowing for foreign insertions, as reported previously^{38,39}. A synthetic gene coding for MS2 CP dimer with Spytag insertion (**Supplementary Material S1**) cloned into pET28a was purchased from BioCat GmbH (Germany). Variants with longer linkers flanking the Spytag insertion and the random peptide insertion were generated by PCR using appropriate primers (**Supplementary Material S2**) and then subcloned to pET28a harboring the MS2 CP dimer gene, using the *KpnI* restriction site. Obtained clones were verified by sequencing (**Supplementary Material S1**).

For expression, *E. coli* BL21(DE3) cells were transformed with appropriate plasmids and grown with shaking at 37°C until OD₆₀₀ = 0.6, induced with 1 M IPTG and then further shaken at 18°C for 16 hours. Cells were harvested by centrifugation, resuspended in 50 mM Tris-HCl, pH 7.9, 50 mM NaCl, 5 mM MgCl₂, 5 mM CaCl₂ and lysed by sonication at 4°C in the presence of protease inhibitors (Thermo Scientific). Lysates were clarified by centrifugation and Viscolase (10,000 U/1L culture; AA Biotechnology, Poland) was added to supernatant fraction, incubated for 20 min

at 37°C, followed by 10 min incubation at 50°C. The supernatant fraction was then centrifuged again and mixed 1:1 (V/V) with 3.7 M (NH₄)₂SO₄, and the reaction was incubated overnight at 4°C. Precipitated proteins were harvested by centrifugation for 10 min, 11,000 x g, at 4°C and resuspended in PBS. The solution was then filtered through 0.2 µm membrane filters (VWR) and passed through an Amicon MWCO 100 kDa (Millipore) filtering device, in order to remove residual (NH₄)₂SO₄ and low molecular mass proteins. Protein concentration was adjusted to 2.5 - 5 mg/mL, as measured by Nanodrop (A₂₈₀) and SEC – purified in PBS buffer, using Superose 6 Increase column (GE Healthcare) connected to an AKTA FPLC system.

SDS-PAGE and native PAGE

MS2 VLPs variants were analyzed by electrophoresis in both denaturing and native conditions. For SDS-PAGE, samples were separated on 12% gels Tris/Glycine gels using standard a Laemmli protocol, whereas for non-denaturing electrophoresis Bis-Tris gels 3-12% gradient gels were used (Life Technologies), following the manufacturer's recommendations. A Chemidoc detector (BioRad) was used for fluorescence detection with excitation at 546 nm. Gels were stained in InstantBlue (Expedeon).

Dynamic light scattering

Dynamic light scattering (DLS) was carried out using a Zetasizer Nano ZS (Malvern). Samples of purified MS2 VLPs were diluted to 0.05 mg/mL (A₂₈₀), 12045 × g centrifuged for 5 min and transferred to plastic/quartz cuvette (ZEN 2112). Measurements were performed in triplicates (15 runs for each measurement). Only measurements meeting Malvern software quality criteria were used for analysis.

Transmission electron microscopy

Samples of purified MS2 VLPs were diluted to 0.05 mg/mL, centrifuged at maximum speed for 15 min and additionally filtered through 0.1 µm membrane filter (VWR). Samples were then applied onto hydrophilized carbon-coated copper grids (STEM Co.), negatively stained with 1% uranyl acetate and visualized using a JEOL JEM-1230 transmission electron microscope (TEM) at 80 kV.

SpyCatcher- mCherry production and interaction with MS2 VLP

The His-tagged Spycatcher - mCherry construct was created by PCR amplification of the *mCherry* gene from a pACYC Duet plasmid (a kind gift from Yusuke Azuma) and its sub-cloning to pET28a harboring *His-tagged Spycatcher* fragment (synthetic construct, Biocat, Germany). The final construct was verified by sequencing.

E. coli BL21(DE3) cells were transformed with the above plasmid and protein expression and extraction were conducted as described above for MS2 CP. The protein was purified using Ni-NTA and following standard purification protocol. Briefly, the cell extract was incubated with agarose beads coupled with Ni²⁺-bound nitrilotriacetic acid (His-Pur Ni-NTA, Thermo Fisher Scientific) pre equilibrated in 50 mM Tris, pH 7.9, 150 mM NaCl, 20 mM imidazole (Buffer A). After three washes of the resin (with Buffer A) the protein was eluted with 50 mM Tris, pH 7.9, 150 mM NaCl, 300 mM imidazole (Buffer B). Fractions containing protein of interest were pooled and passed through Sephadex 25 (Millipore) columns in order to remove imidazole. Final protein concentration was measured by Nanodrop at 280 nm wavelength.

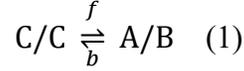
Attachment of purified Spycatcher-mCherry to MS2 VLP was conducted by mixing the two in a range of different molar ratios, followed by 90 min incubation at room temperature. The interaction efficiency was evaluated by SDS-PAGE whereas efficient particle decoration was confirmed using native gel electrophoresis with fluorescence detection and size exclusion chromatography (as described above).

Cryo-electron microscopy

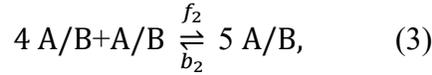
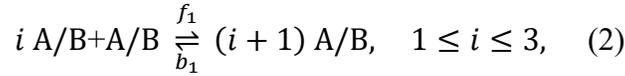
Purified samples of MS2 VLPs at ~1 mg/mL concentration were flash-frozen in liquid ethane using a FEI Vitrobot (sample volume 4 μ L, blot force 0, blot time 4 s) on previously glow-discharged copper grids (Quantifoli, Cu 1.2/1.3, mesh 400). All grids were imaged with a 300 kV acceleration voltage using a Titan Krios microscope armed with a Gatan K3 camera (0.86 A/px, 40 frames movies). Raw micrographs were motion corrected using WARP⁴¹ with all further steps carried out using the CryoSPARC v2.15.0 software package⁴². CTF values were calculated in patch mode using Patch CTF. Micrographs were accepted for particle picking when meeting a criterion of CTF fit better than 8 Å (CTF \leq Å). All reported resolution values are a result of independent half maps analysis with gold standard FSC criterion (FSC = 0.143). All figures containing cryoEM maps were prepared using either UCSF Chimera⁴³ or ChimeraX⁴⁴.

MS2 VLP assembly model

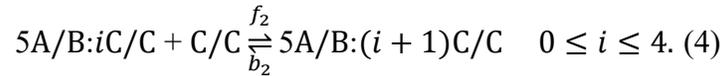
Particle assembly is modelled via reaction kinetics, that encode the interconversion between C/C and A/B dimers with forward rate f and backward rate b (equation 1):



Assembly starts with formation of particle 5-fold axes according to the following reactions (equations 2 and 3):

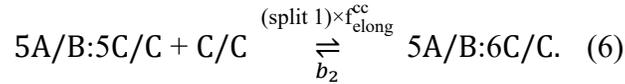
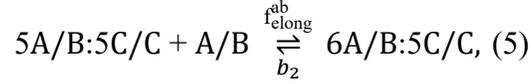


where $\frac{b_1}{f_1} = e^{\frac{\Delta G}{K_B T}}$ and K_B is the Boltzmann constant, T is temperature, and ΔG is the binding free energy which is -2.7 kcal M^{-1} ⁴⁵. For the last reaction the binding free energy is -5.4 kcal M^{-1} as there are two binding sites for the fifth A/B. This is followed by acquisition of five C/C dimers around the 5A/B complex (equation 4):



As this early assembly intermediate is shared by all particles, we assume that the first branching of the assembly pathways, resulting in the observed particle geometries, occurs at this point (cf. split1 in **Figure S10a,b**). At this stage, we assume that A/B and C/C dimers bind with rates $f_{\text{elong}}^{\text{ab}}$ and $f_{\text{elong}}^{\text{cc}}$, respectively ($f_1 = 10^3 \text{ M}^{-1}\text{S}^{-1}$, $f_2 = f_{\text{elong}}^{\text{ab}} = f_{\text{elong}}^{\text{cc}} = 10^6 \text{ M}^{-1}\text{S}^{-1}$ ^{46,47}), to the 5A/B:5C/C intermediate. These additions are based on a tree that indicates bifurcations in the assembly pathways whenever addition of an A/B or C/C dimer commits the intermediate to assembly of a distinct particle type (**Figure S10a**). To move towards the formation of $T=3$ particles, the intermediate 5A/B+5C/C must bind to an A/B dimer, whilst recruitment of a C/C

dimer will result in the formation of $T=4$ particles (**Figure S10b**). This has been modelled as follows (equations 5 and 6):



We assume that the forward rate of the second reaction is reduced by the factor *split 1* (**Table S5**) to reflect the fact that it is a bifurcation from the wild type ($T=3$) pathway. **Figure S10b** shows that in the assembly pathway of $T=3$ ($T=4$) particles, the intermediates $15A/B:8C/C$ ($15A/B:11C/C$) must acquire an A/B dimer to continue towards a $T=3$ ($T=4$) particle geometry. However, if they acquire a C/C dimer, they will continue towards the formation of D_3 particles (cf. **Figure S10b**). Thus, in the model we assume that $15A/B:8C/C$ ($15A/B:11C/C$) can bifurcate towards the formation of D_3 particles by binding to a C/C dimer. Similarly, the rate of this split is reduced by the factor *split 2* (*split 3*) as D_3 particles have a lower symmetry compared with $T=3$ ($T=4$) particles, and we model these splits as for *split 1*. The assembly pathways of $T=3$ and D_5 particles are similar until $30A/B:20C/C$ (**Figure S10c**), where recruitment of an A/B (C/C) dimer biases particle formation towards a $T=3$ (D_5) particle type (cf. *split 4* in **Figure S10a**). **Figure S10d** illustrates that the assembly pathways of D_3 -A and D_3 -B particles are similar until $44A/B:27C/C$, where recruitment of an A/B dimer results in formation of D_3 -A particles, and that of C/C dimers in the formation of D_3 -B particles (cf. *split 5* in **Figure S10a, d**).

In the absence of a split in the assembly tree, the transition from assembly intermediate $n_1A/B:m_1C/C$ to $n_2A/B:m_2C/C$ is modelled as the random binding of $(n_2 - n_1)A/B$ and $(m_2 -$

m_1)C/C dimers according to the following matrix, modelling the successive recruitment of individual A/B and C/C dimers in an n_2 A/B: m_2 C/C array:

$$\left(\begin{array}{cccccc} n_1 A/B:m_1 C/C & \xrightarrow{f_{\text{elong}}^{\text{cc}}} & n_1 A/B:(m_1 + 1)C/C & \xrightarrow{f_{\text{elong}}^{\text{cc}}} & \dots & \xrightarrow{f_{\text{elong}}^{\text{cc}}} & n_1 A/B:m_2 C/C \\ f_{\text{elong}}^{\text{ab}} \downarrow & & f_{\text{elong}}^{\text{ab}} \downarrow & & \dots & & f_{\text{elong}}^{\text{ab}} \downarrow \\ (n_1 + 1)A/B:m_1 C/C & \xrightarrow{f_{\text{elong}}^{\text{cc}}} & (n_1 + 1)A/B:(m_1 + 1)C/C & \xrightarrow{f_{\text{elong}}^{\text{cc}}} & \dots & \xrightarrow{f_{\text{elong}}^{\text{cc}}} & (n_1 + 1)A/B:m_2 C/C \\ f_{\text{elong}}^{\text{ab}} \downarrow & & f_{\text{elong}}^{\text{ab}} \downarrow & & \dots & & f_{\text{elong}}^{\text{ab}} \downarrow \\ \vdots & & \vdots & & \vdots & & \vdots \\ n_2 A/B:m_1 C/C & \xrightarrow{f_{\text{elong}}^{\text{cc}}} & n_2 A/B:(m_1 + 1)C/C & \xrightarrow{f_{\text{elong}}^{\text{cc}}} & \dots & \xrightarrow{f_{\text{elong}}^{\text{cc}}} & n_2 A/B:m_2 C/C \end{array} \right).$$

Parameter values

The default rate for the splits is first chosen for all but the SpyTag4 scenario, as the latter leads to a much higher yield in $T=4$ particles compared to the others. *Splits* rates are chosen to reflect the symmetry of the particles, as there are more equivalent contact points for particles with higher symmetry. Consistent with this, *split 2* is the lowest, as it leads to the formation of D_3 particles with the lowest symmetry on the $T=3$ (wild type) branch. *Split 3* is slightly larger, although it leads to D_3 particles, as it occurs on the $T=4$ branch of the assembly tree. *Split 1* is the largest as it occurs at the start of the assembly process and can lead to $T=4$ particles whose assembly intermediates offer the largest number of symmetry-equivalent positions for incoming subunits. *Split 4* is smaller than *split 1*, as D_5 is of lower symmetry than $T=4$. *Split 5* is slightly smaller than *split 1* because it occurs at the end of the assembly process, and since D_3 -A contains fewer C/C dimers than D_3 -B. Based on these arguments, values of *splits 1-5* were identified (Table S5) using the Gillespie algorithm⁴⁸ implemented in Fortran. At that point, only the conversion rate from the symmetric C/C to the asymmetric A/B dimer (f) remains a free parameter in the model,

and it is identified for each scenario based on the experimentally observed outcomes in relative particles numbers (**Table S7**). For the case of SpyTag4, for which the level of $T=4$ particles is much higher than for the other cases, variation of f alone is not sufficient to account for the data. We note that the best fit is obtained when f is smaller than in all other cases, implying that there is a resistance of C/C dimers to convert into A/B in this case. This is likely due to the dynamic properties of the dimer as a result of the SpyTag4 insert and may also affect the C/C binding rate to the assembly intermediates. We reflect this by reducing the value of the elongation rate of C/C dimers ($f_{\text{elong}}^{\text{cc}}$). This also implies that C/C dimers are more likely to occupy positions that require less dynamic flexibility, i.e., positions with lower curvature where C/C dimers need to bend less in order to attached. Consistent with this, C/C recruitment is higher at *split 1* and *split 4* (**Table S5**) as they lead to particles with lower curvatures, as $T=4$ is larger than $T=3$, and D_5 has a cylindrical shape.

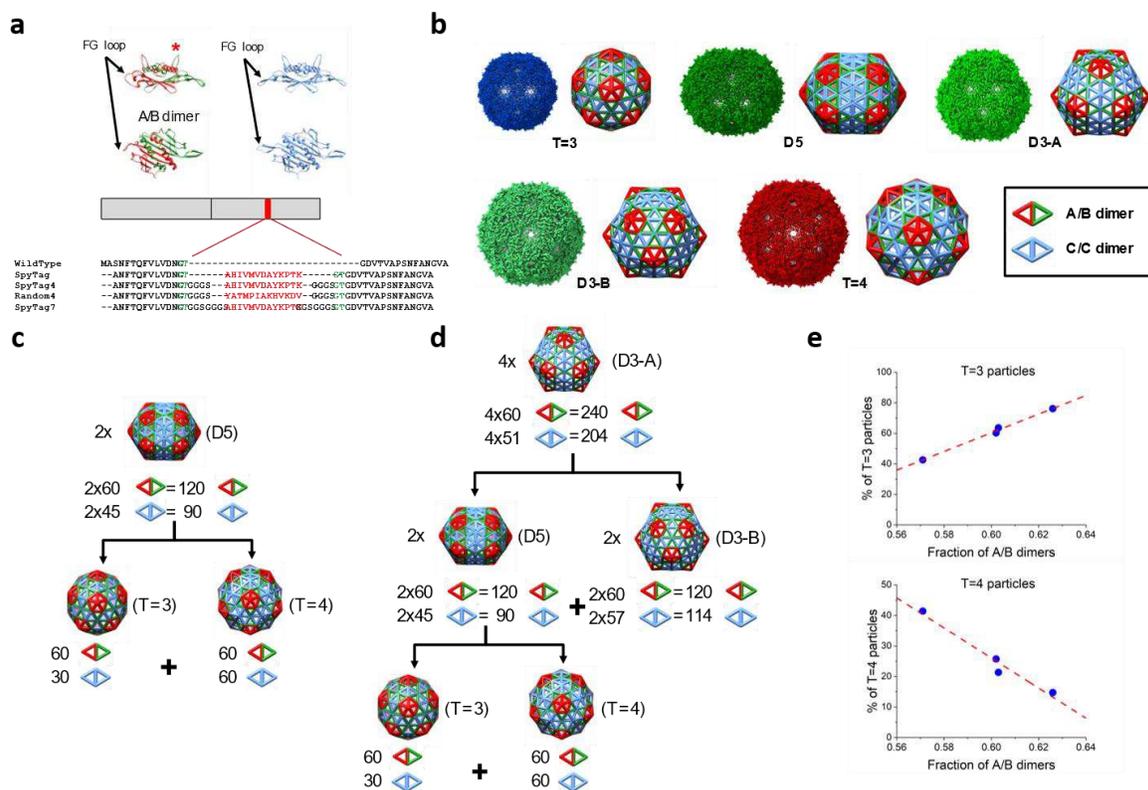


Figure 1. Virus like particles derived from bacteriophage MS2. **a** Orthogonal views of the A/B (top-left) and C/C (top-right) dimers. The FG-loop determining dimer conformation is indicated by arrows, and the insertion site by a red asterisk. The MS2 coat protein dimer construct, with the insert highlighted in red, is shown underneath, together with a sequence alignment to wild type protein (black), with the restriction site dependent amino acids in green and additional amino acids in bold black, respectively. **b** The reconstructed cryoEM densities of all identified MS2 VLP variants, with schematic representation of A/B and C/C dimer tilings (left and right, respectively); the symmetry of each structure is indicated below. **c-d** Examples of equivalent particle stoichiometries: **c**: two D-5 particles can be converted into one $T=3$ and one $T=4$ particle; and **d**: four D3-A particles can be converted into two D3-B and two D5 particles, which can be subsequently converted into one $T=3$ and one $T=4$ particle. **e** The percentages of $T=3$ and $T=4$ particles in the ensemble depend linearly on the fraction of asymmetric dimers; data points corresponding to, from left to right, SpyTag4; Random4; SpyTag7; and SpyTag, reveal a linear trend (dotted line) with respect to the fraction of A/B dimers.

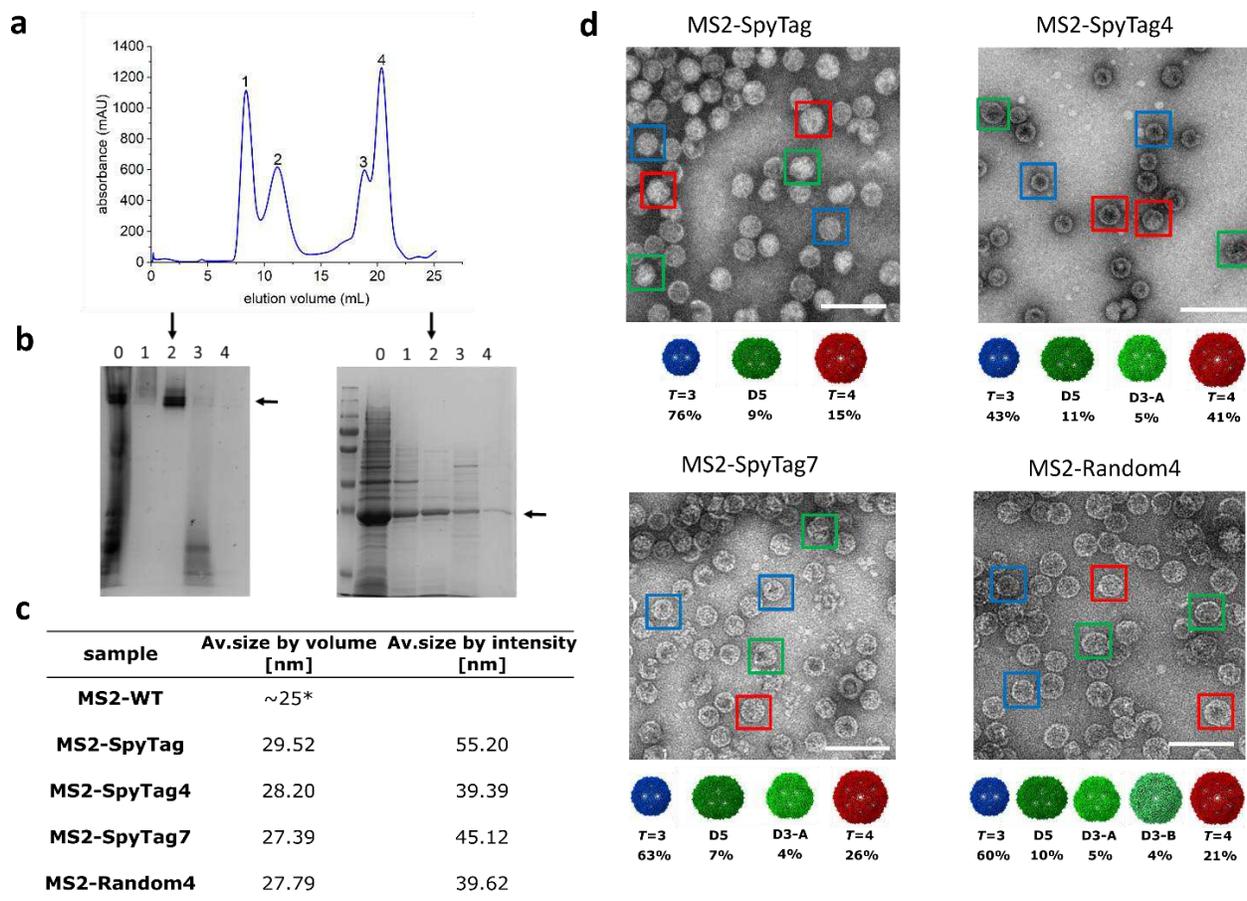


Figure 2. MS2 VLPs purification and analysis. **a** A representative size exclusion purification profile for an MS2 VLP; **b** native PAGE (left) and SDS PAGE (right) analysis of eluted fractions: lane numbers correspond to peak numbers on the chromatogram; “0” denotes sample before SEC. Vertical arrows indicate fractions containing purified VLP and used for further analysis. Horizontal arrows indicate bands corresponding to the expected molecular weight (of the coat protein dimer in SDS PAGE and the VLP in native PAGE); **c** Table showing representative DLS measurements of hydrodynamic diameter of the MS2-SpyTag particles present in fractions collected from peak 2 (on the chromatogram). Volume distribution (top left) in comparison with the intensity distribution (top right); together with diameter measurements for the entire ensemble of MS2 samples, *reference⁴⁹ **d**) Transmission electron microscopy images of MS2 VLP variants. Color frames show different morphologies of the assembled VLPs: $T=3$, $T=4$ and the lower symmetry variants (D5, D3-A and D3-B) in blue, red and green, respectively with overall percentage of each morphology identified in the analyzed samples; scale bar – 100 nm.

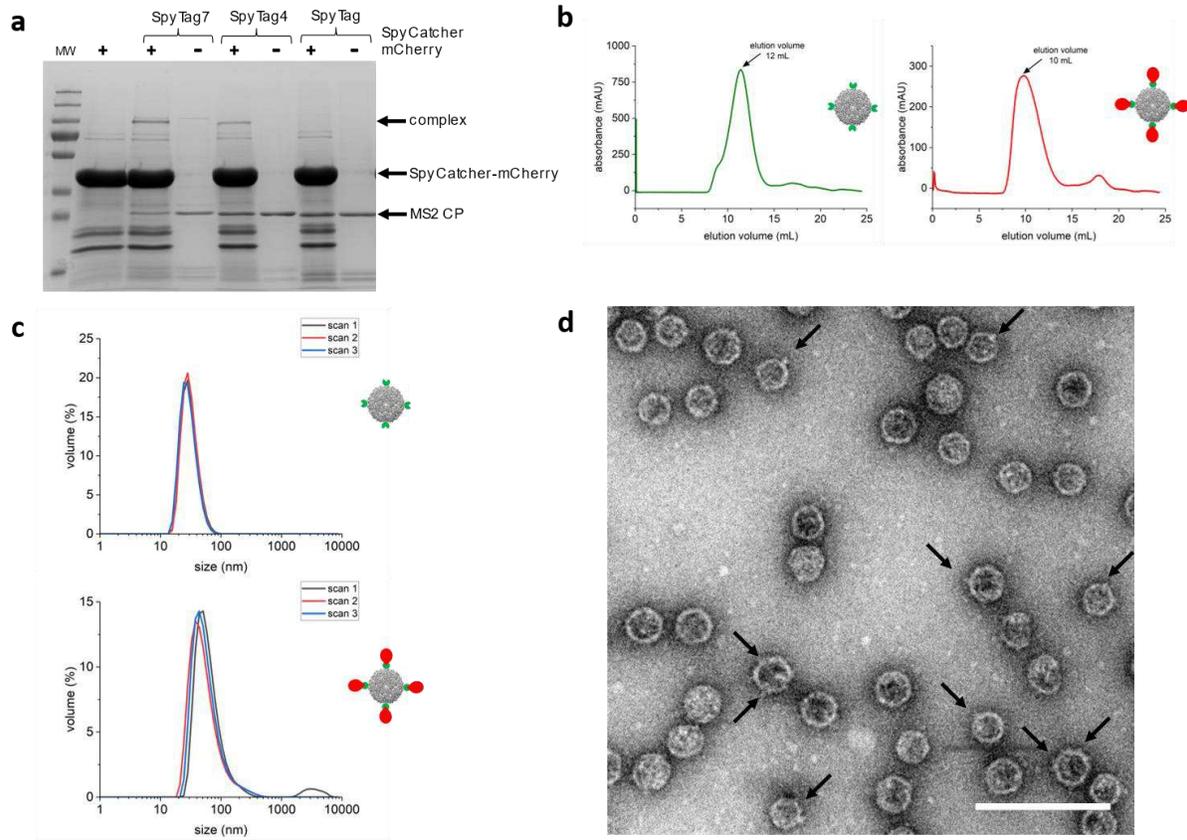


Figure 3. MS2 VLP decoration with the SpyCatcher-mCherry protein. **a** SDS-PAGE gel showing interaction of the Spycatcher mCherry protein with different versions of the MS2 coat protein fused with SpyTag variants; **b** Size exclusion elution profiles of naked MS2 VLPs (left/green) and Spycatcher-mCherry decorated MS2 VLPs (right/red). **c** DLS profiles (three independent scans of the same sample) of the SpyTag modified MS2 VLPs alone (top) and after addition of Spycatcher mCherry partner (bottom); **d** TEM image of Spycatcher-mCherry decorated MS2 VLPs, with black arrows pointing to the Spycatcher mCherry protein on the surface of MS2 VLPs (scale bar – 100 nm).

ACKNOWLEDGMENTS

Funding: A.N., A.P.B. and J.G.H. were supported by a National Science Centre (NCN, Poland) grant no. 2016/20/W/NZ1/00095 (Symfonia-4). J.G.H. and A.P.B were funded by a Polish National Science Centre Maestro grant no. 2019/34/A/NZ1/00196. RT acknowledges funding via an EPSRC Established Career Fellowship (EP/R023204/1), a Royal Society Wolfson Fellowship (RSWF/R1/180009). FF is funded via a Joint Wellcome Trust Investigator Award to Reidun Twarock and Peter Stockley from the University of Leeds (110145 & 110146).

We thank Olga Woznicka for TEM imaging. This research was supported in part by PL-Grid Infrastructure. We acknowledge the MCB Structural Biology Core Facility (supported by the TEAM TECH CORE FACILITY/2017-4/6 grant from the Foundation for Polish Science) for valuable support. Part of this research took place at SOLARIS National Synchrotron Radiation Centre, using the cryoEM infrastructure. Experiments were performed in collaboration with SOLARIS Staff. The open-access publication of this article was funded by the BioS Priority Research Area under the program “Excellence Initiative – Research University” at the Jagiellonian University in Krakow.

AUTHOR INFORMATION

Authors contribution

Conceptualization: APB, AN

Methodology: APB, AN

Investigation: APB, AN

Visualization: APB, AN, FF, RT

Supervision: RT, JGH

Writing—original draft: APB, AN, FF, RT, JGH

Writing—review & editing: APB, AN, FF, RT, JGH

The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript. ‡These authors contributed equally.

Corresponding Author

*Jonathan G. Heddle – email: jonathan.heddle@uj.edu.pl

Competing Interest: The authors declare the following competing financial interests J.G.H. is the founder of and holds equity in nCage Therapeutics LLC, which aims to commercialize protein cages for therapeutic applications.

ASSOCIATED CONTENT

Data Availability

The data that support the findings of this study are available from the corresponding author on reasonable request. The cryo-EM density maps have been deposited in the Electron Microscopy Data Bank under accession codes: support the findings of this study are available from the corresponding author on reasonable request. The cryo-EM density maps have been deposited in the Electron Microscopy Data Bank under accession codes: EMD-12778 (MS2-SpyTag, $T=3$), EMD-12779 (MS2-SpyTag, $T=4$), EMD-12780 (MS2-SpyTag, D5), EMD-12781 (MS2-SpyTag4, $T=3$), EMD-12782 (MS2-SpyTag4, $T=4$), EMD-12783 (MS2-SpyTag4, D5), EMD-12784 (MS2-SpyTag4, D3-A), EMD-12785 (MS2-Random4, $T=3$), EMD-12786 (MS2-Random4, $T=4$), EMD-12787 (MS2-Random4, D5), EMD-12788 (MS2-Random4, D3-A), EMD-12789 (MS2-Random4, D3-B), EMD-12790 (MS2-SpyTag7, $T=3$), EMD-12791 (MS2-SpyTag7, $T=4$), EMD-12792 (MS2-SpyTag7, D5), EMD-12793 (MS2-SpyTag7, D3-A).

Code Availability

Custom code used for modeling the assembly process is available for download via the authors'

GitHub page: <https://github.com/MathematicalComputationalVirology/MS2Nanotechnology>

Supplementary Information containing supplementary figures and tables necessary to fully understand the content of the article.

The following files are available free of charge.

Biela_Naskalska-Supplementary_Information.pdf

REFERENCES

- 1 Wu, M., Brown, W. L. & Stockley, P. G. Cell-specific delivery of bacteriophage-encapsidated ricin A chain. *Bioconjug. Chem.* **6**, 587-595, doi:10.1021/bc00035a013 (1995).
- 2 Wu, M., Sherwin, T., Brown, W. L. & Stockley, P. G. Delivery of antisense oligonucleotides to leukemia cells by RNA bacteriophage capsids. *Nanomedicine* **1**, 67-76, doi:10.1016/j.nano.2004.11.011 (2005).
- 3 Galaway, F. A. & Stockley, P. G. MS2 viruslike particles: a robust, semisynthetic targeted drug delivery platform. *Mol. Pharm.* **10**, 59-68 (2013).
- 4 Fietze, K. M., Peabody, D. S. & Chackerian, B. Engineering virus-like particles as vaccine platforms. *Curr Opin Virol* **18**, 44-49, doi:10.1016/j.coviro.2016.03.001 (2016).
- 5 Bond, K., Tsvetkova, I. B., Wang, J. C., Jarrold, M. F. & Dragnea, B. Virus Assembly Pathways: Straying Away but Not Too Far. *Small* **16**, e2004475, doi:10.1002/smll.202004475 (2020).
- 6 Zlotnick, A. & Mukhopadhyay, S. Virus assembly, allostery and antivirals. *Trends Microbiol* **19**, 14-23, doi:10.1016/j.tim.2010.11.003 (2011).
- 7 Selivanovitch, E., Koliyatt, R. & Douglas, T. Chemically Induced Morphogenesis of P22 Virus-like Particles by the Surfactant Sodium Dodecyl Sulfate. *Biomacromolecules* **20**, 389-400, doi:10.1021/acs.biomac.8b01357 (2019).
- 8 Xu, C. *et al.* Switch from Polymorphic to Homogenous Self-Assembly of Virus-Like Particles of Simian Virus 40 through Double-Cysteine Substitution. *Small* **16**, e2004484, doi:10.1002/smll.202004484 (2020).

- 9 Kanesashi, S. N. *et al.* Simian virus 40 VP1 capsid protein forms polymorphic assemblies in vitro. *J Gen Virol* **84**, 1899-1905, doi:10.1099/vir.0.19067-0 (2003).
- 10 Azuma, Y., Edwardson, T. G. W. & Hilvert, D. Tailoring lumazine synthase assemblies for bionanotechnology. *Chem Soc Rev* **47**, 3543-3557, doi:10.1039/c8cs00154e (2018).
- 11 Peabody, D. S. Translational repression by bacteriophage MS2 coat protein expressed from a plasmid. A system for genetic analysis of a protein-RNA interaction. *J Biol Chem* **265**, 5684-5689 (1990).
- 12 Mastico, R. A., Talbot, S. J. & Stockley, P. G. Multiple presentation of foreign peptides on the surface of an RNA-free spherical bacteriophage capsid. *J Gen Virol* **74 (Pt 4)**, 541-548, doi:10.1099/0022-1317-74-4-541 (1993).
- 13 Golmohammadi, R., Valegård, K., Fridborg, K. & Liljas, L. The refined structure of bacteriophage MS2 at 2.8 Å resolution. *J Mol Biol* **234**, 620-639, doi:10.1006/jmbi.1993.1616 (1993).
- 14 Dykeman, E. C., Stockley, P. G. & Twarock, R. Dynamic allostery controls coat protein conformer switching during MS2 phage assembly. *J Mol Biol* **395**, 916-923, doi:10.1016/j.jmb.2009.11.016 (2010).
- 15 Stockley, P. G. *et al.* A simple, RNA-mediated allosteric switch controls the pathway to formation of a T=3 viral capsid. *J. Mol. Biol.* **369**, 541-552 (2007).
- 16 Twarock, R. & Stockley, P. G. RNA-Mediated Virus Assembly: Mechanisms and Consequences for Viral Evolution and Therapy. *Annu Rev Biophys* **48**, 495-514, doi:10.1146/annurev-biophys-052118-115611 (2019).
- 17 Dykeman, E. C., Stockley, P. G. & Twarock, R. Solving a Levinthal's paradox for virus assembly identifies a unique antiviral strategy. *Proc Natl Acad Sci U S A* **111**, 5361-5366, doi:10.1073/pnas.1319479111 (2014).
- 18 Dykeman, E. C. *et al.* Simple rules for efficient assembly predict the layout of a packaged viral RNA. *J Mol Biol* **408**, 399-407, doi:10.1016/j.jmb.2011.02.039 (2011).
- 19 Borodavka, A., Tuma, R. & Stockley, P. G. Evidence that viral RNAs have evolved for efficient, two-stage packaging. *Proceedings of the National Academy of Sciences* **109**, 15769-15774 (2012).
- 20 Dykeman, E. C., Stockley, P. G. & Twarock, R. Packaging signals in two single-stranded RNA viruses imply a conserved assembly mechanism and geometry of the packaged genome. *J. Mol. Biol.* **425**, 3235-3249 (2013).
- 21 Rolfsson, Ó. *et al.* Direct evidence for packaging signal-mediated assembly of bacteriophage MS2. *J. Mol. Biol.* **428**, 431-448 (2016).
- 22 Twarock, R., Leonov, G. & Stockley, P. G. Hamiltonian path analysis of viral genomes. *Nature communications* **9**, 1-3 (2018).
- 23 Pasloske, B. L., Walkerpeach, C. R., Obermoeller, R. D., Winkler, M. & DuBois, D. B. Armored RNA technology for production of ribonuclease-resistant viral RNA controls and standards. *J Clin Microbiol* **36**, 3590-3594, doi:10.1128/JCM.36.12.3590-3594.1998 (1998).
- 24 Zhan, S. *et al.* Armored long RNA controls or standards for branched DNA assay for detection of human immunodeficiency virus type 1. *J Clin Microbiol* **47**, 2571-2576, doi:10.1128/JCM.00232-09 (2009).
- 25 Wang, S. *et al.* Preparation and evaluation of MS2 bacteriophage-like particles packaging hepatitis E virus RNA. *FEMS Microbiol Lett* **363**, doi:10.1093/femsle/fnw221 (2016).

- 26 Wu, M., Brown, W. L. & Stockley, P. G. Cell-specific delivery of bacteriophage-encapsidated ricin A chain. *Bioconjug. Chem.* **6**, 587-595 (1995).
- 27 Wu, M., Sherwin, T., Brown, W. L. & Stockley, P. G. Delivery of antisense oligonucleotides to leukemia cells by RNA bacteriophage capsids. *Nanomed. Nanotechnol. Biol. Med.* **1**, 67-76 (2005).
- 28 Peabody, D. S. *et al.* Immunogenic display of diverse peptides on virus-like particles of RNA phage MS2. *J Mol Biol* **380**, 252-263, doi:10.1016/j.jmb.2008.04.049 (2008).
- 29 Zhai, L. *et al.* A novel candidate HPV vaccine: MS2 phage VLP displaying a tandem HPV L2 peptide offers similar protection in mice to Gardasil-9. *Antiviral Res* **147**, 116-123, doi:10.1016/j.antiviral.2017.09.012 (2017).
- 30 Dong, Y. M., Zhang, G. G., Huang, X. J., Chen, L. & Chen, H. T. Promising MS2 mediated virus-like particle vaccine against foot-and-mouth disease. *Antiviral Res* **117**, 39-43, doi:10.1016/j.antiviral.2015.01.005 (2015).
- 31 Tumban, E., Peabody, J., Tyler, M., Peabody, D. S. & Chackerian, B. VLPs displaying a single L2 epitope induce broadly cross-neutralizing antibodies against human papillomavirus. *PLoS One* **7**, e49751, doi:10.1371/journal.pone.0049751 (2012).
- 32 Ord, R. L. *et al.* A malaria vaccine candidate based on an epitope of the Plasmodium falciparum RH5 protein. *Malar J* **13**, 326, doi:10.1186/1475-2875-13-326 (2014).
- 33 Haim, L., Zapor, G., Aronov, S. & Gerst, J. E. A genomic integration method to visualize localization of endogenous mRNAs in living yeast. *Nat Methods* **4**, 409-412, doi:10.1038/nmeth1040 (2007).
- 34 Yoon, J. H., Srikantan, S. & Gorospe, M. MS2-TRAP (MS2-tagged RNA affinity purification): tagging RNA to identify associated miRNAs. *Methods* **58**, 81-87, doi:10.1016/j.ymeth.2012.07.004 (2012).
- 35 Wang, Z. *et al.* External Quality Assessment for Molecular Detection of Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) in Clinical Laboratories. *J Mol Diagn* **23**, 19-28, doi:10.1016/j.jmoldx.2020.10.008 (2021).
- 36 Brune, K. D. *et al.* Plug-and-Display: decoration of Virus-Like Particles via isopeptide bonds for modular immunization. *Sci Rep* **6**, 19234, doi:10.1038/srep19234 (2016).
- 37 Brune, K. D. & Howarth, M. New Routes and Opportunities for Modular Construction of Particulate Vaccines: Stick, Click, and Glue. *Front Immunol* **9**, 1432, doi:10.3389/fimmu.2018.01432 (2018).
- 38 Peabody, D. S. & Lim, F. Complementation of RNA binding site mutations in MS2 coat protein heterodimers. *Nucleic Acids Res* **24**, 2352-2359 (1996).
- 39 Mikel, P., Vasickova, P. & Kralik, P. One-plasmid double-expression His-tag system for rapid production and easy purification of MS2 phage-like particles. *Sci Rep* **7**, 17501, doi:10.1038/s41598-017-17951-5 (2017).
- 40 de Martín Garrido, N. *et al.* Bacteriophage MS2 displays unreported capsid variability assembling T = 4 and mixed capsids. *Mol Microbiol* **113**, 143-152, doi:10.1111/mmi.14406 (2020).
- 41 Tegunov, D. & Cramer, P. Real-time cryo-electron microscopy data preprocessing with Warp. *Nat Methods* **16**, 1146-1152, doi:10.1038/s41592-019-0580-y (2019).
- 42 Punjani, A., Rubinstein, J. L., Fleet, D. J. & Brubaker, M. A. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat Methods* **14**, 290-296, doi:10.1038/nmeth.4169 (2017).

- 43 Pettersen, E. F. *et al.* UCSF Chimera--a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605-1612, doi:10.1002/jcc.20084 (2004).
- 44 Pettersen, E. F. *et al.* UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein Sci* **30**, 70-82, doi:10.1002/pro.3943 (2021).
- 45 Zlotnick, A. To build a virus capsid: an equilibrium model of the self assembly of polyhedral protein complexes. *J. Mol. Biol.* **241**, 59-67 (1994).
- 46 Endres, D. & Zlotnick, A. Model-based analysis of assembly kinetics for virus capsids or other spherical polymers. *Biophys. J.* **83**, 1217-1230 (2002).
- 47 Zlotnick, A. Distinguishing reversible from irreversible virus capsid assembly. *J. Mol. Biol.* **366**, 14-18 (2007).
- 48 Gillespie, D. T. Exact stochastic simulation of coupled chemical reactions. *The journal of physical chemistry* **81**, 2340-2361 (1977).
- 49 Robinson, S., Hartman, E., Ikwuagwu, B., Francis, M. & Tullman-Ereck, D. Engineering a Virus-like Particle to Display Peptide Insertions Using an Apparent Fitness Landscape. *Biomacromolecules* **21**, 4194-4204, doi:10.1021/acs.biomac.0c00987 (2020).

Supplementary Materials for

Programmable Polymorphism of a Virus-Like Particle

Artur P. Biela†, Antonina Naskalska†, Farzad Fatehi, Reidun Twarock, Jonathan G. Heddle*

*Email: jonathan.heddle@uj.edu.pl

This PDF file includes:

Supplementary Text

Figs. S1 to S10

Tables S1 to S8

Data S1

References (1 to 2)

Supplementary Text

Ensembles of particle morphologies compatible with a given dimer ratio.

Given a fixed number $N(A/B)$ of A/B dimers and $N(C/C)$ of C/C dimers, it is possible to construct distinct ensembles of particle morphologies from a geometrical point of view. Calling α , β , γ , δ , and ε the numbers of $T=3$, D5, $T=4$, D3-A, and D3-B particles in the ensemble, we obtain the relations $60(\alpha + \beta + \gamma + \delta + \varepsilon) = N(AB)$, and $30\alpha + 45\beta + 60\gamma + 51\delta + 57\varepsilon = N(CC)$. For SpyTag we have $\delta = \varepsilon = 0$. Using the experimentally determined values $N(A/B)=6\,883,200$ and $N(C/C)=4\,105,590$ (see **Table S4**), we obtain $\alpha = 92587 - \beta/2$, and $\gamma = 22,133 - \beta/2$. From a purely geometric point of view, any value β , i.e., the number of D5 particles in the ensemble, between 0 and 44,266 is possible. The value $\beta=0$ would favor the formation of icosahedrally symmetric $T=3$ and $T=4$ particles over lower-symmetry D5 particles. However, the number $\beta=10,456$ seen in the experiment implies that the symmetry of the final products is not the only determinant of particle morphology. A purely thermodynamic argument is therefore not sufficient, and aspects of assembly kinetics must also be considered.

In order to assess this effect qualitatively for different SpyTag options, we compare the maximal number of $T=3$ particles that could be built from the building blocks in the final ensemble by dissociating lower symmetry species and building higher symmetry particles from their constituent building blocks. The latter is a proxy for assembly given an idealised “instantaneous” conversion rate and can be used to benchmark against the particle numbers observed experimentally based on the actual conversion rate (**Figure 1e**). For SpyTag, this means comparing the $\beta = 0$ case with the $\beta = 10456$ scenario observed (**Figure 1c**). For all other scenarios, there are two reductions that need to be considered, first converting all D3-A particles into D5 and D3-B particles ($\delta = 0$) and

then converting all D5 particles into $T=3$ and $T=4$ particles ($\beta = 0$) (**Figure 1d**). As a result, the percentage of $T=3$ particles in the ensemble increases in this “ideal” scenario (**Table S8**), and the increase is an indicator of kinetic contributions to the assembly outcome. Our kinetic assembly model quantifies such contributions and demonstrates the importance of the conversion rate from the symmetric C/C to the asymmetric A/B dimer for the experimental outcome.

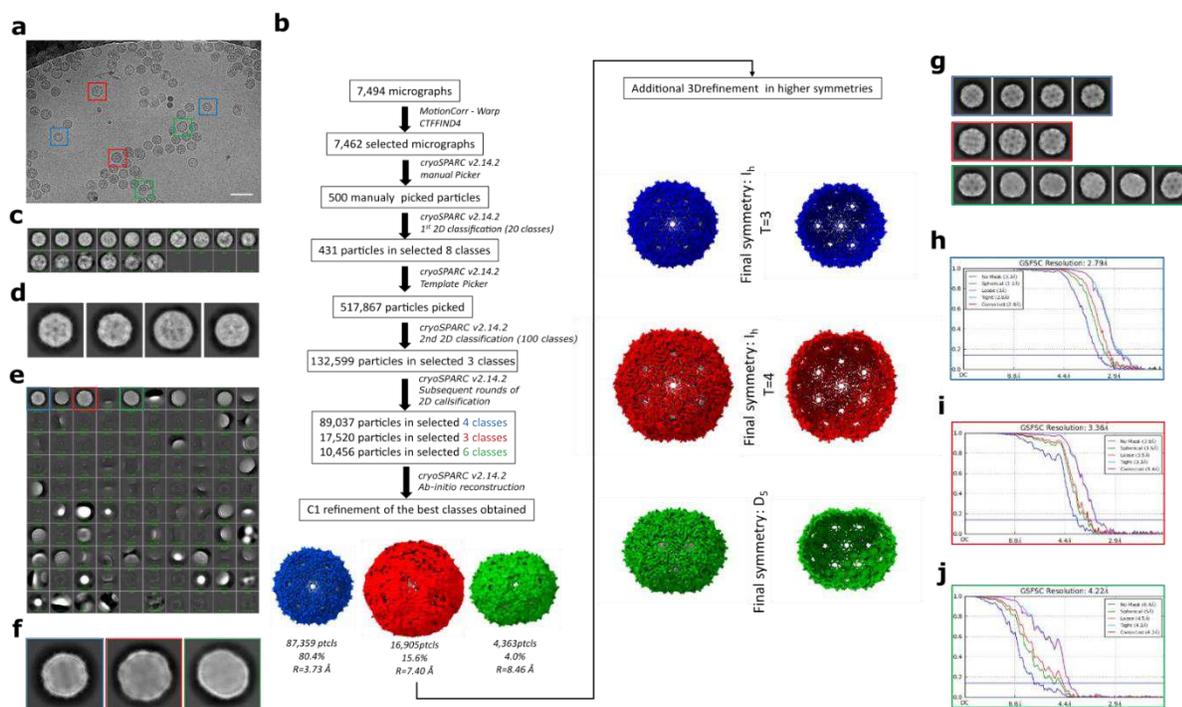


Figure S1: Procedure for cryo-EM reconstruction of MS2-SpyTag sample. **a**, representative micrograph with three different particles marked in boxes (f ~ 26 nm, f ~ 34 nm and non-spherical in blue, red and green respectively); Scale bar – 50 nm. **b**, Summary of the image processing procedure (see Methods). **c**, initial 2D class averages. **d**, selected 2D class averages used for template search. **e**, reference free 2D classification after template pick. **f**, selected 3 different 2D class averages. **g**, final 2D class averages for each of the identified particles. **h-j**, FSC correlation curves for each of the solved densities after applying highest possible symmetry.

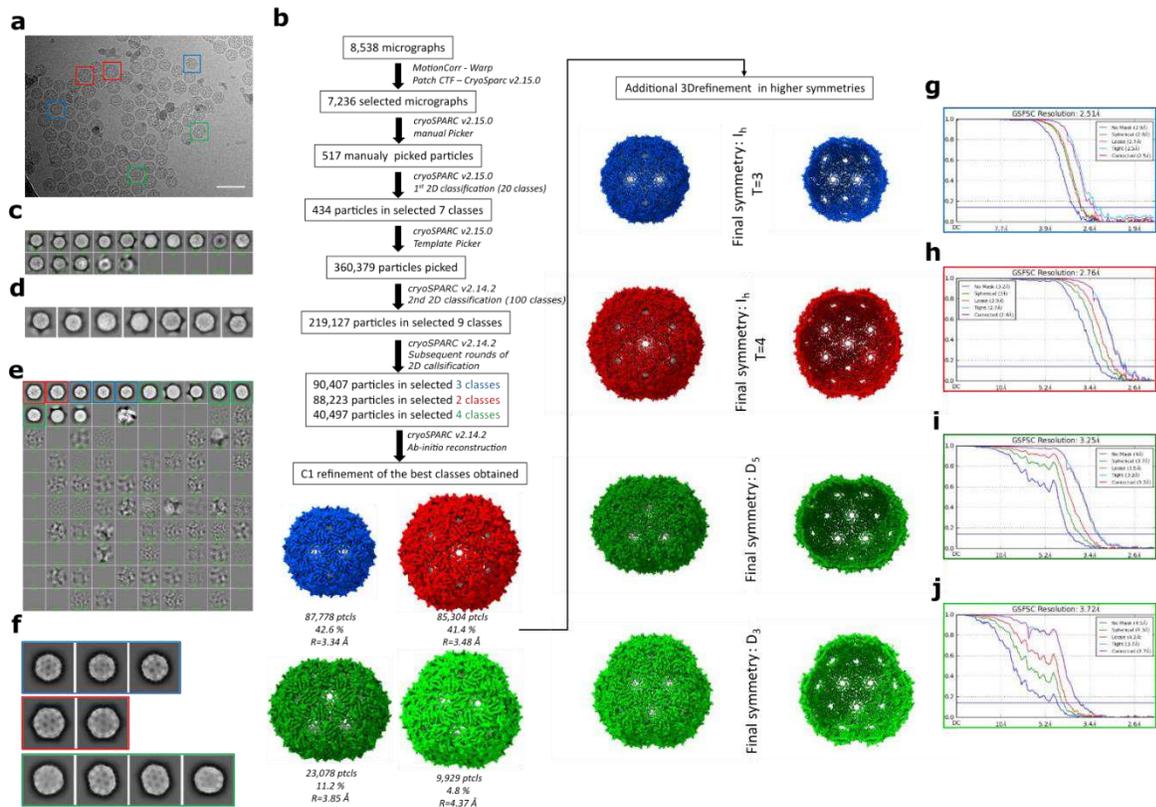


Figure S2: Procedure for cryo-EM reconstruction of MS2-SpyTag4 sample. **a**, representative micrograph with three different particles marked in boxes ($f \sim 26 \text{ nm}$, $f \sim 34 \text{ nm}$ and non-spherical in blue, red and green respectively); Scale bar – 50 nm. **b**, Summary of the image processing procedure (see Methods). **c**, initial 2D class averages. **d**, selected 2D class averages used for template search. **e**, reference free 2D classification after template pick. **f**, final 2D class averages for each of the identified particles. **g-j**, FSC correlation curves for each of the solved densities after applying highest possible symmetry.

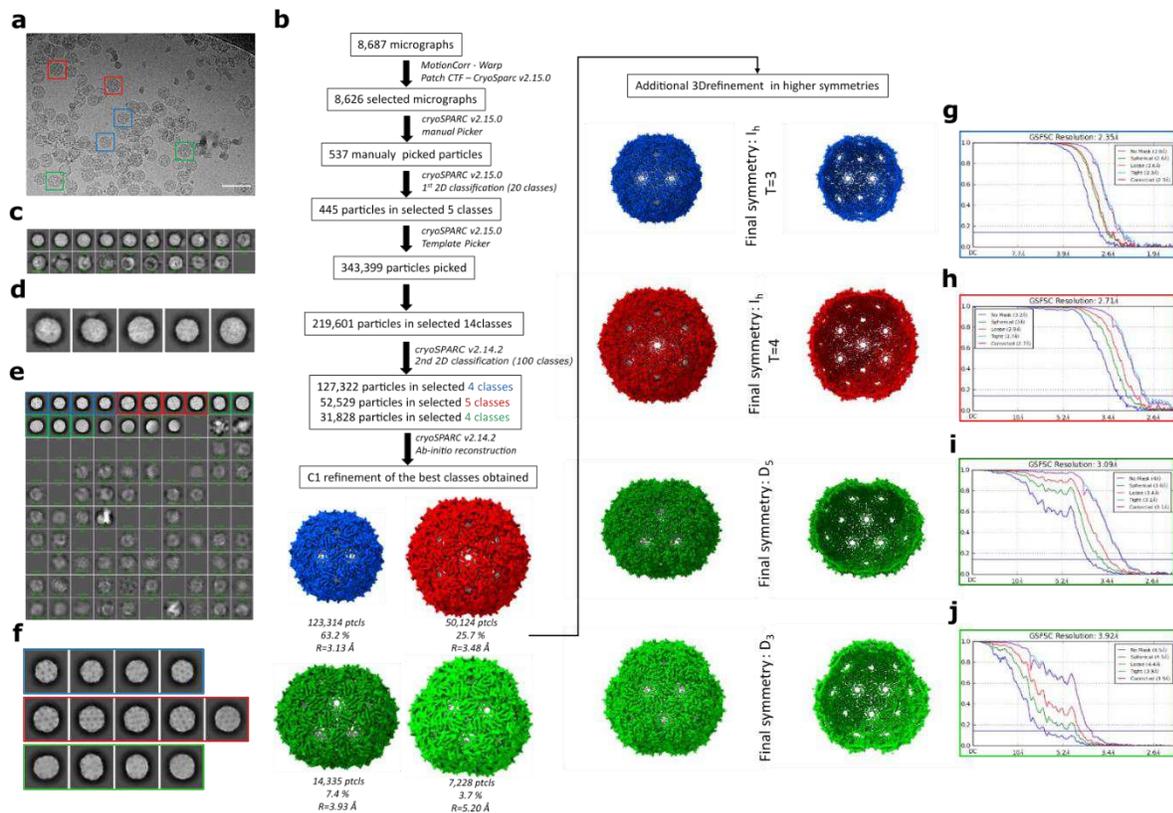


Figure S3: Procedure for cryo-EM reconstruction of MS2-SpyTag7 sample. **a**, representative micrograph with three different particles marked in boxes (f ~ 26 nm, f ~ 34 nm and non-spherical in blue, red and green respectively); Scale bar – 50 nm. **b**, Summary of the image processing procedure (see Methods). **c**, initial 2D class averages. **d**, selected 2D class averages used for template search. **e**, reference free 2D classification after template pick. **f**, final 2D class averages for each of the identified particles. **g-j**, FSC correlation curves for each of the solved densities after applying highest possible symmetry.

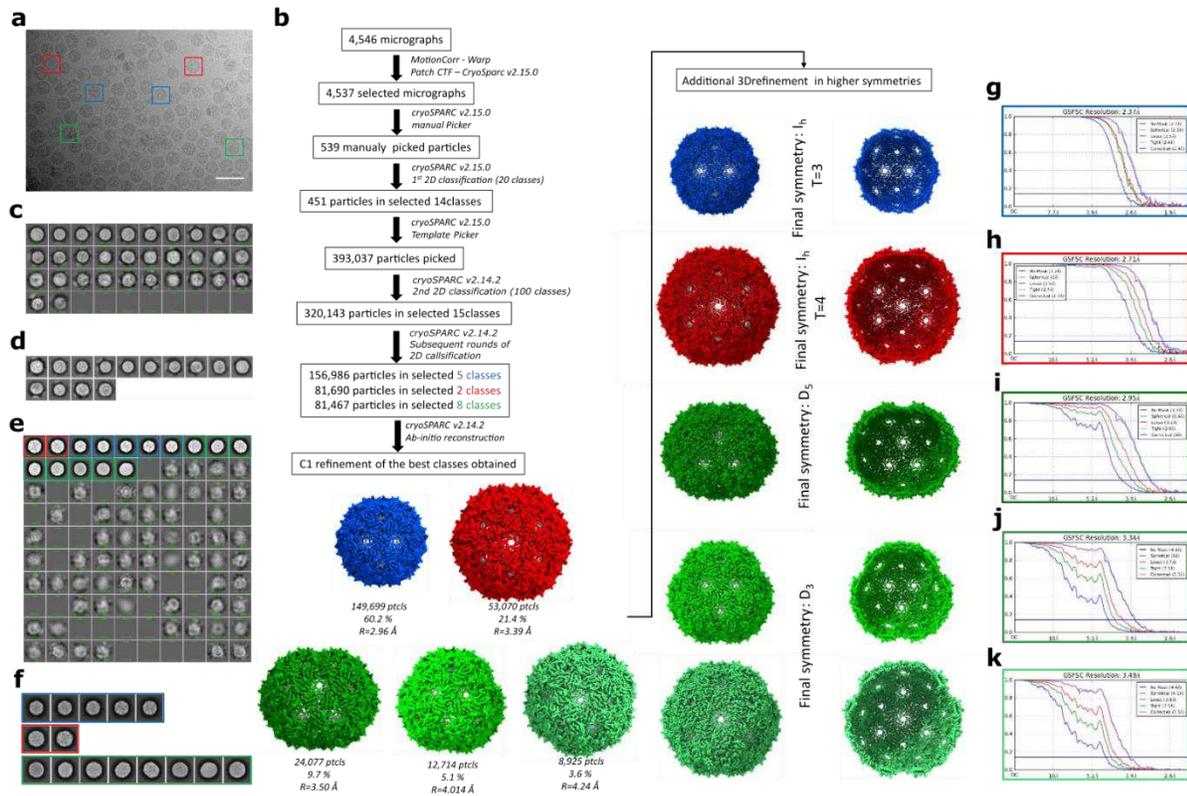


Figure S4: Procedure for cryo-EM reconstruction MS2-Random4 sample. **a**, representative micrograph with three different particles marked in boxes ($f \sim 26 \text{ nm}$, $f \sim 34 \text{ nm}$ and non-spherical in blue, red and green respectively); Scale bar – 50 nm. **b**, Summary of the image processing procedure (see Methods). **c**, initial 2D class averages. **d**, selected 2D class averages used for template search. **e**, reference free 2D classification after template pick. **f**, final 2D class averages for each of the identified particles. **g-k**, FSC correlation curves for each of the solved densities after applying highest possible symmetry.

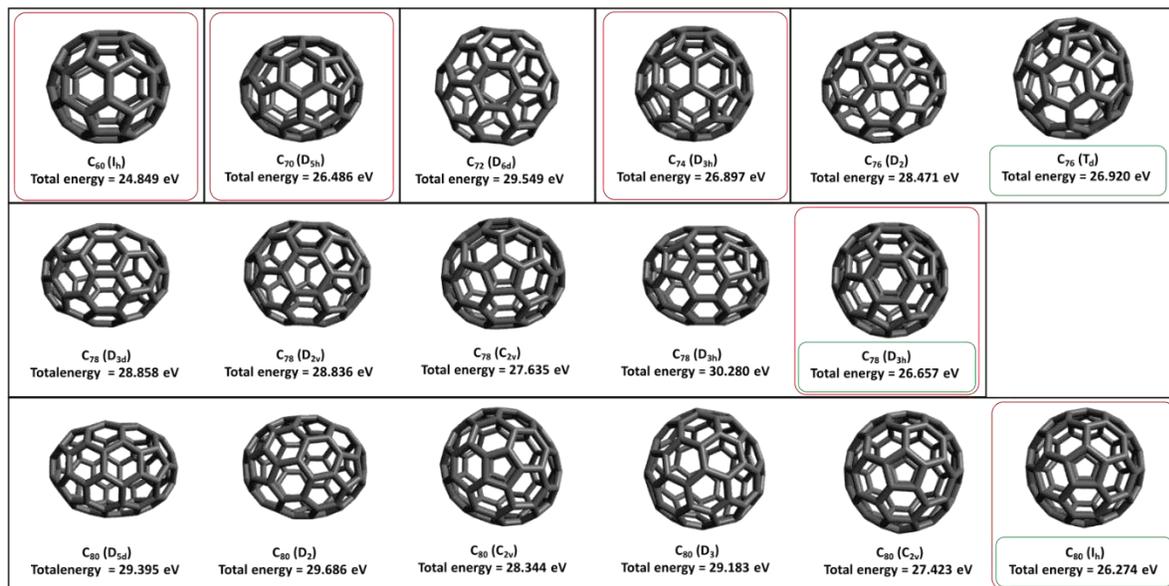


Figure S5: Structures of fullerenes with carbon atom numbers from 60 to 80 with associated symmetries and total strain energies¹: When isomers exist, those with lowest strain energies are marked with green frame; geometries identified in corresponding MS2 VLPs variants marked with red frames. Figure prepared using Avogadro².

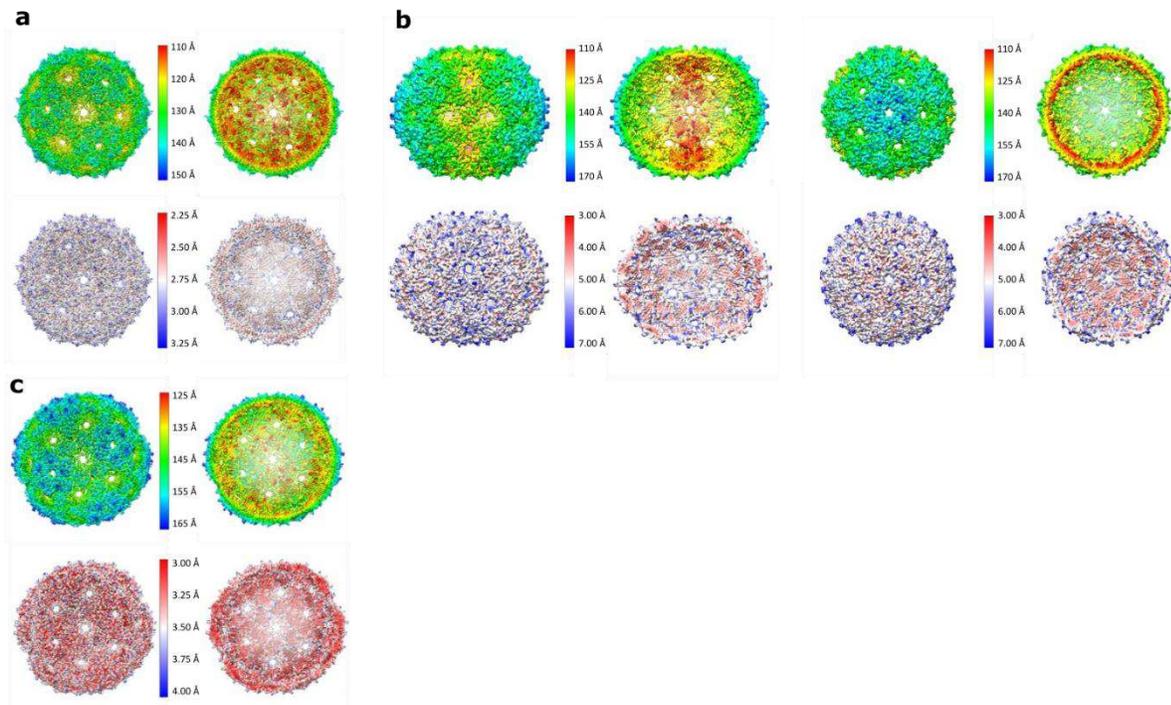


Figure S6: MS2-SpyTag VLPs showed in radius colouring scheme (top) and local resolution (bottom): a. $T=3$ particle, b. $D5$ particle; two orthogonal views are shown (left and right), c. $T=4$ particle. Every structure is shown in external surface view (left) as well as a cross section revealing inner surface of the VLPs (right).

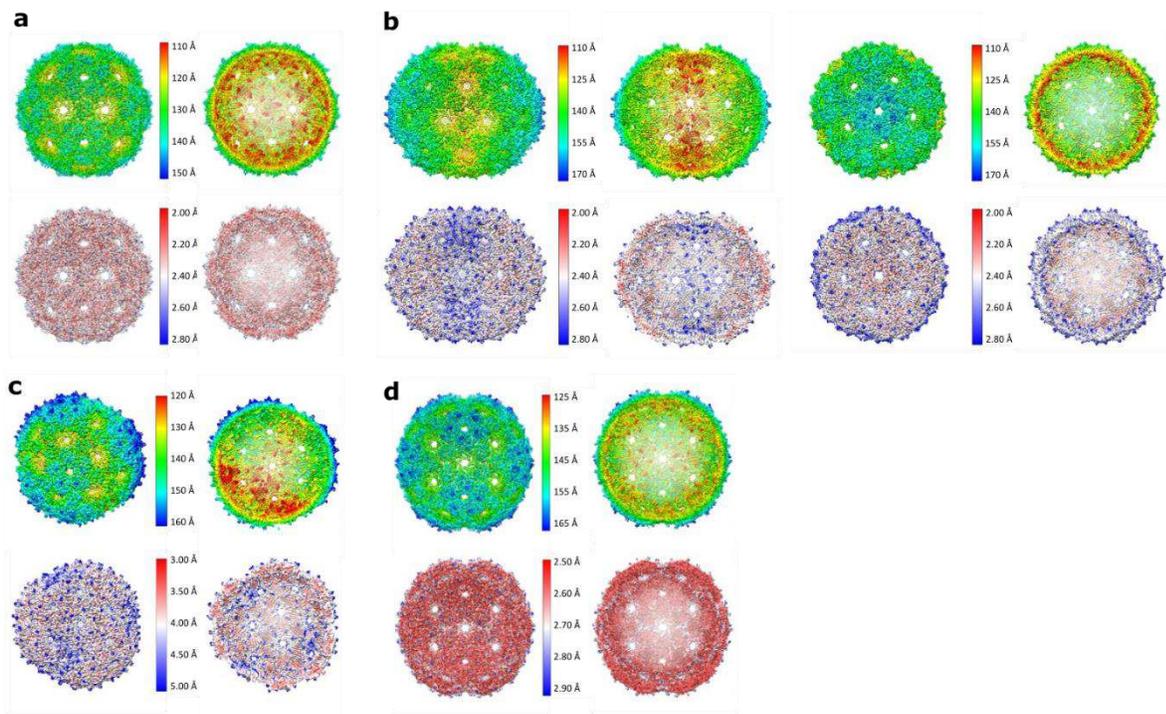


Figure S7: MS2-SpyTag4 VLPs showed in radius colouring scheme (top) and local resolution (bottom): a. $T=3$ particle, b. $D5$ particle; two orthogonal views are shown (left and right), c. $D3-A$ particle, d. $T=4$ particle. Every structure is shown in external surface view (left) as well as a cross section revealing inner surface of the VLPs (right).

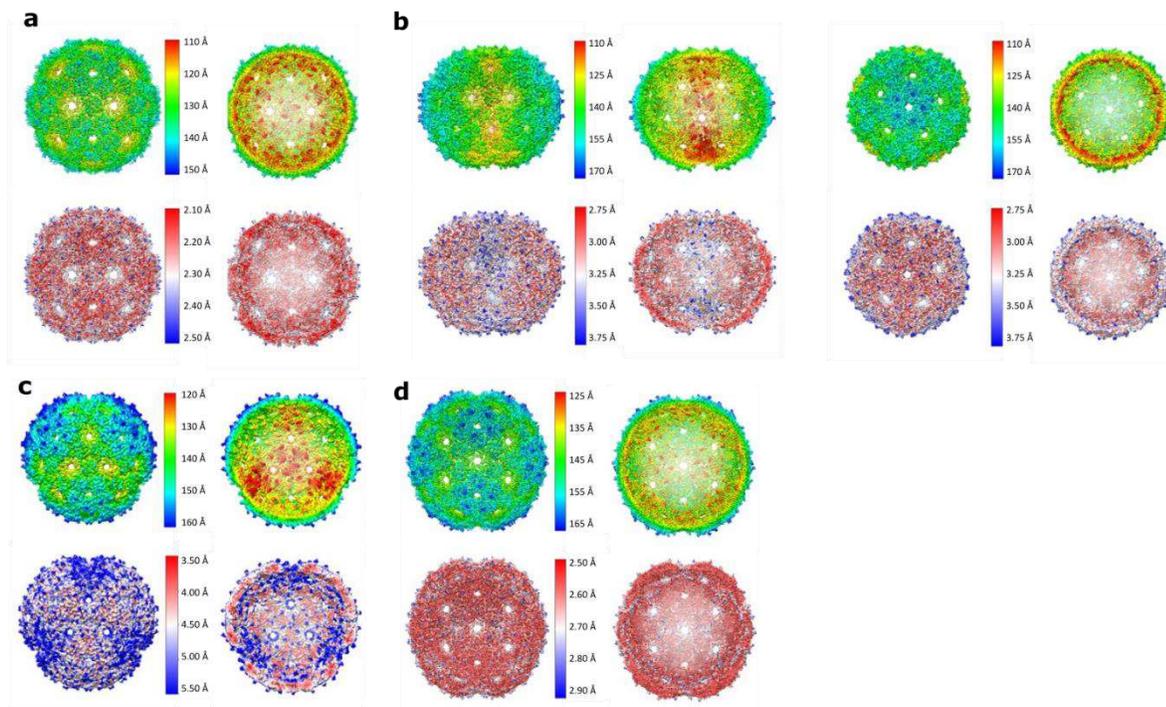


Figure S8: MS2-SpyTag7 VLPs showed in radius colouring scheme (top) and local resolution (bottom): a. $T=3$ particle, b. $D5$ particle; two orthogonal views are shown (left and right), c. $D3-A$ particle, d. $T=4$ particle. Every structure is shown in external surface view (left) as well as a cross section revealing inner surface of the VLPs (right).

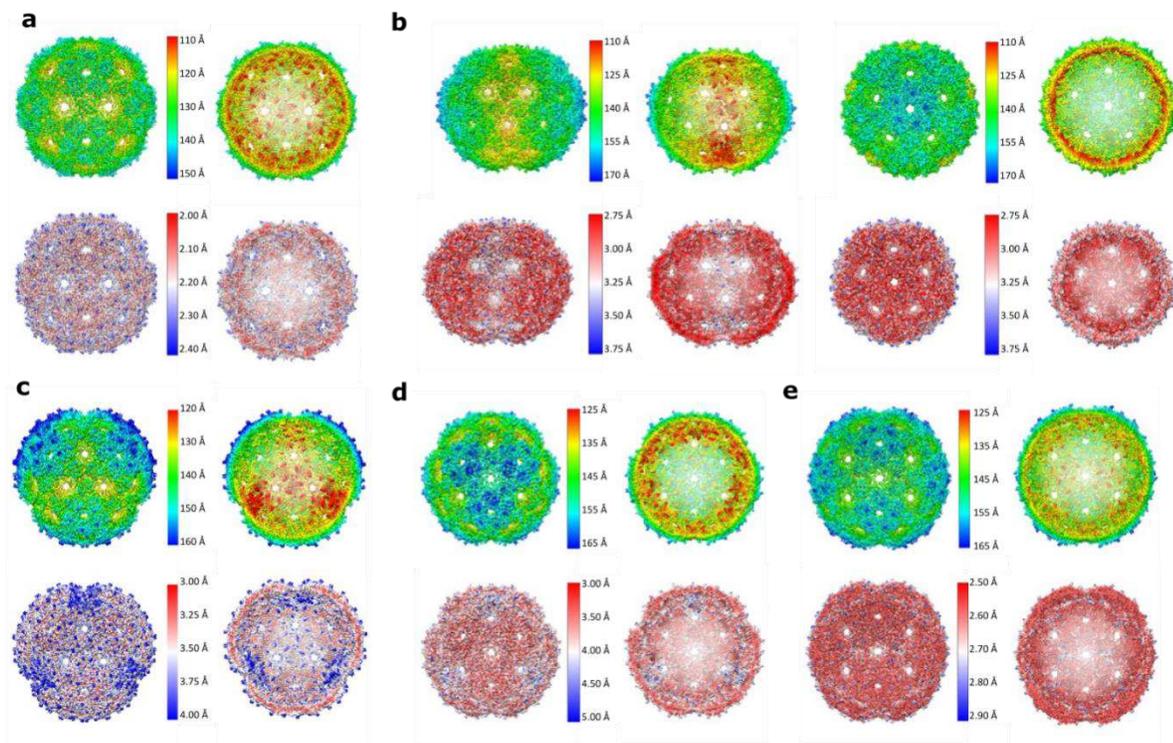


Figure S9: MS2-Random4 VLPs showed in radius colouring scheme (top) and local resolution (bottom): a. *T=3* particle, b. *D5* particle; two orthogonal views are shown (left and right), c. *D3-A* particle, d. *D3-B* particle, e. *T=4* particle. Every structure is shown in external surface view (left) as well as a cross section revealing inner surface of the VLPs (right).

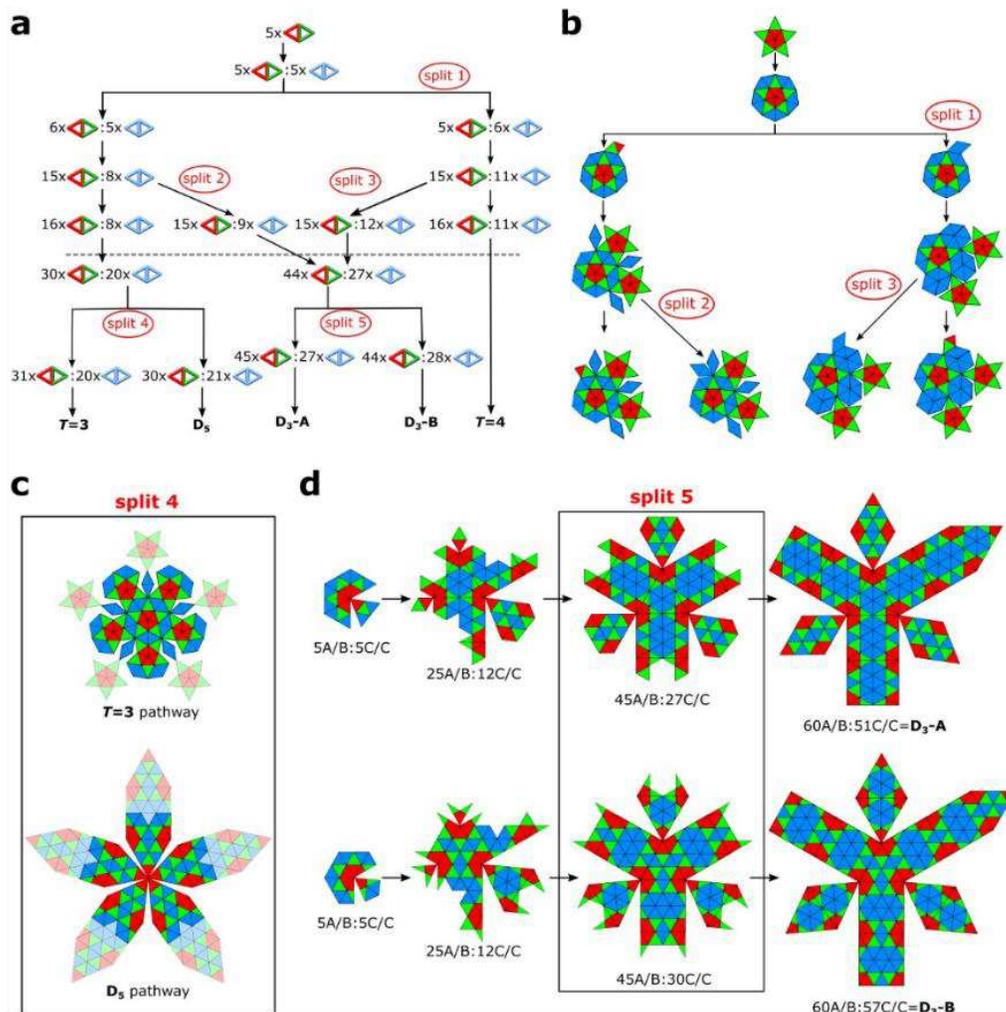


Figure S10. The bifurcation of assembly pathways and intermediate geometries shared by different particle types. **a.** Diagrammatic illustration of bifurcation points (“splits”) in the assembly pathways leading to different particle types. **b.** Tiling representation of the intermediate particle geometries at bifurcation points in the assembly pathways above the grey dashed line in **a**, with green/red and blue/blue rhomboids indicating A/B and C/C dimers, respectively. **c.** Tiling model of the particle geometry at split point 4, showing the next dimers to be recruited on the pathway towards a $T=3$ (top) and D_5 particle (bottom) in faint colours. In particular, the $T=3$ pathway must next acquire an A/B dimer (green/red), while the D_5 pathway must recruit a C/C dimer (blue/blue). **d.** Shows assembly intermediates on an assembly pathway of D_3 -A and D_3 -B particles, illustrating the geometry of the shared intermediate at split point 5.

Table S1. Primers used to generate SpyTag 4aa linkers; Spytag 7aa linkers and Random Peptide 4aa linkers insertions into the CP gene.

SpyTag 4 aa Fwd	5' CTGAGGTACCGGCGGCGGCAGCGCTCATATTGTTATGG
SpyTag 4 aa Rev	5' CTGAGGGTACCGCTGCCGCCCTTAGTTGGCTTG
SpyTag 7aa Fwd	5' cttGGTACCggtgtagtggcggcggcagcgctcat
SpyTag 7aa Rev	5' actaagggcggcggcagcggtgtagtggtaccaag
Random 4 aa Fwd	5' ctgaggtaccggcggcggcagctacgctactatgccaattgctaagcatg
Random 4 aa Rev	5' ctgaggtaccgctgcccgcaccaacatccttaacatgcttagcaattggc

Table S2 Structures of identified MS2-SpyTag VLPs – summary

Sample	MS2-SpyTag			MS2-SpyTag4				MS2-SpyTag7				Random4				
Symmetry	<i>T=3</i>	<i>D₅</i>	<i>T=4</i>	<i>T=3</i>	<i>D₅</i>	<i>D₃</i> (D3-A)	<i>T=4</i>	<i>T=3</i>	<i>D₅</i>	<i>D₃</i> (D3-A)	<i>T=4</i>	<i>T=3</i>	<i>D₅</i>	<i>D₃</i> (D3-A)	<i>D₃</i> (D3-B)	<i>T=4</i>
Resolution [Å] FSC _{0.143}	2.79	4.22	3.36	2.51	3.25	3.72	2.76	2.35	3.09	3.92	2.71	2.37	2.95	3.34	3.49	2.71
EMD-id	EMD- 12778	EMD- 12779	EMD- 12780	EMD- 12781	EMD- 12782	EMD- 12783	EMD- 12784	EMD- 12785	EMD- 12786	EMD- 12787	EMD- 12788	EMD- 12789	EMD- 12790	EMD- 12791	EMD- 12792	EMD- 12793

Table S3: Categorization of particle types into constituent numbers of A/B and C/C dimers.

Particle type	<i>T</i> =3	D5	<i>T</i> =4	D3-A	D3-B
A/B	60	60	60	60	60
C/C	30	45	60	51	57

Table S4: Number of A/B dimers in all discovered geometries of MS2 VLP variants; ^a – experimental data

sample	MS2-SpyTag			MS2-SpyTag4				MS2-SpyTag7				Random4				
	T=3	D5	T=4	T=3	D5	D3-A	T=4	T=3	D5	D3-A	T=4	T=3	D5	D3-A	D3-B	T=4
A/B dimers	60	60	60	60	60	60	60	60	60	60	60	60	60	60	60	60
C/C dimers	30	45	60	30	45	51	60	30	45	51	60	30	45	51	57	60
no. of identified particles^a	87,359	10,456	16,905	87,778	23,078	9,929	85,304	123,314	14,135	7,228	50,124	149,699	24,077	12,714	8,925	53,070
total no. of identified particles^a	114,720			206,089				194,801				248,485				
total no of A/B dimers^a	5 241,540	627,360	1 014,300	5 266,680	1 384,680	595,740	5 118,240	7 398,840	848,100	433,680	3 007,440	8 981,940	1 444,620	762,840	535,500	3 184,200
total no of C/C dimers^a	2 620,770	470,520	1 014,300	2 633,340	1 038,510	506,379	5 118,240	3 699,420	636,075	368,628	3 007,440	4 490,970	1 083,465	648,414	508,725	3 184,200
A/B dimers:all dimers ratio^a	0.667	0.571	0.500	0.667	0.571	0.541	0.500	0.667	0.571	0.541	0.500	0.667	0.571	0.541	0.513	0.500
total A/B dimers:all dimers ratio^a	0.626			0.571				0.602				0.601				

Table S5. Estimated values of parameters. f is the conversion from C/C to A/B, with backward rate b which is assumed to be fixed. Split k , $k=1$ to 5, refer to nodes in the assembly graph in Fig. 1, where the pathways to different geometries bifurcate. $f_{\text{elong}}^{\text{cc}}$ and $f_{\text{elong}}^{\text{ab}}$ are the rates at which C/C and A/B dimers bind after the nucleation step ($5A/B+5C/C$), respectively.

parameters	SpyTag	SpyTag4	SpyTag7	Random4
f	5.2 S^{-1}	0.29 S^{-1}	1.5 S^{-1}	1.64 S^{-1}
b	0.02 S^{-1}	0.02 S^{-1}	0.02 S^{-1}	0.02 S^{-1}
Split 1	0.14	0.28	0.14	0.14
Split 2	0.022	0.022	0.022	0.022
Split 3	0.031	0.031	0.031	0.031
Split 4	0.07	0.11	0.07	0.07
Split 5	0.11	0.11	0.11	0.11
$f_{\text{elong}}^{\text{cc}}$	$10^6 \text{ M}^{-1}\text{S}^{-1}$	$0.7 \times 10^6 \text{ M}^{-1}\text{S}^{-1}$	$10^6 \text{ M}^{-1}\text{S}^{-1}$	$10^6 \text{ M}^{-1}\text{S}^{-1}$
$f_{\text{elong}}^{\text{ab}}$	$10^6 \text{ M}^{-1}\text{S}^{-1}$	$10^6 \text{ M}^{-1}\text{S}^{-1}$	$10^6 \text{ M}^{-1}\text{S}^{-1}$	$10^6 \text{ M}^{-1}\text{S}^{-1}$

Table S6. The ratios for different particle types with respect to the $T=3$ particles as a function of f . This indicates that by increasing the conversion rate f the number of other particles with respect to the number of $T=3$ (wild type) particles reduce. This shows that inclusion of a tag to C/C dimers will reduce the conversion rate and leads to the formation of D_5 , D_3 , and $T=4$ particles.

Particle f	D₅	D₃-A	D₃-B	T=4
10 S^{-1}	0.06	0.02	0.003	0.1
15 S^{-1}	0.04	0.02	0.001	0.09
20 S^{-1}	0.03	0.01	0.0009	0.07
30 S^{-1}	0.02	0.009	0.0005	0.05
40 S^{-1}	0.02	0.007	0.0003	0.04

Table S7. Ratios computed based on experimental data are shown on the left which have been used for model fitting and the model results are shown on the right.

	SpyTag		SpyTag4		SpyTag7		Random4	
	Experiment	Model	Experiment	Model	Experiment	Model	Experiment	Model
T=3	1	1	1	1	1	1	1	1
D₅	0.12	0.1	0.26	0.27	0.11	0.16	0.16	0.15
D₃-A	0	0.04	0.11	0.12	0.06	0.07	0.08	0.07
D₃-B	0	0.007	0	0.05	0	0.02	0.06	0.02
T=4	0.19	0.2	0.97	0.93	0.41	0.38	0.35	0.36

Table S8: Difference between the experimentally observed percentage of $T=3$ particles and the maximal number that could be obtained from the same building blocks in the overall ensemble when formation of the icosahedral $T=3$ and $T=4$ particles was favored.

	SpyTag	SpyTag4	SpyTag7	Random4
experiment	76.15%	43.8%	64.23%	60.24%
ideal scenario	80.85%	48.19%	66.93%	63.3%
% increase	4.70%	4.39%	2.7%	3.06%

Supplementary Data S1. MS2 CP genetic constructs. Nucleotide sequences of generated CP variants were obtained from a synthetic gene provider (MS2-SpyTag) or sequencing service provider (MS2-SpyTag4, MS2-SpyTag7, MS2-Random). SpyTag sequence is represented in red, KpnI restriction site in green, linker sequences in blue.

- MS2-SpyTag (synthetic sequence):

ATGGCTTCTAACTTTACTCAGTTCGTTCTCGTCGACAATGGCGGAACTGGCGACGTGACTGTCGCCCCAAGCAACT
TCGCTAACGGGGTCGCTGAATGGATCAGCTCTAACTCGCGTTCACAGGCTTACAAAGTAACCTGTAGCGTTCGTCA
GAGCTCTGCGCAGAATCGCAAATACACCATCAAAGTCGAGGTGCCTAAAGTGGCAACCCAGACTGTTGGTGGTGT
AGAGCTTCTGTAGCCGCATGGCGTTCGACTTAAATATGGAACCTAACATTCCAATTTTCGCTACGAATCCGACT
GCGAGCTTATTGTTAAGGCAATGCAAGGTCTCCTAAAAGATGGAAACCCGATTCCCTCAGCAATCGCAGCAAACCT
CGGCATCTACGCTAACTTTACTCAGTTCGTTCTCGTCGACAATGGCGGTACCGCTCATATTGTTATGGTTGATGCTT
ACAAGCCAACTAAGGGTACCGGCGACGTGACTGTCGCCCCAAGCAACTTCGCTAACGGGGTCGCTGAATGGATCA
GCTCTAACTCGCGTTCACAGGCTTACAAAGTAACCTGTAGCGTTCGTAGAGCTCTGCGCAGAATCGCAAATACAC
CATCAAAGTCGAGGTGCCTAAAGTGGCAACCCAGACTGTTGGTGGTGTAGAGCTTCTGTAGCCGCATGGCGTTC
GTACTTAAATATGGAACCTAACATTCCAATTTTCGCTACGAATCCGACTGCGAGCTTATTGTTAAGGCAATGCAAG
GTCTCCTAAAAGATGGAAACCCGATTCCCTCAGCAATCGCAGCAAACCTCCGGCATCTACTAA

- MS2-SpyTag4:

ATGGCTTCTAACTTTACTCAGTTCGTTCTCGTCGACAATGGCGGAACTGGCGACGTGACTGTCGCCCCAAGCAACT
TCGCTAACGGGGTCGCTGAATGGATCAGCTCTAACTCGCGTTCACAGGCTTACAAAGTAACCTGTAGCGTTCGTCA
GAGCTCTGCGCAGAATCGCAAATACACCATCAAAGTCGAGGTGCCTAAAGTGGCAACCCAGACTGTTGGTGGTGT
AGAGCTTCTGTAGCCGCATGGCGTTCGACTTAAATATGGAACCTAACATTCCAATTTTCGCTACGAATCCGACT
GCGAGCTTATTGTTAAGGCAATGCAAGGTCTCCTAAAAGATGGAAACCCGATTCCCTCAGCAATCGCAGCAAACCT
CGGCATCTACGCTAACTTTACTCAGTTCGTTCTCGTCGACAATGGCGGTACCGGCGGCGGCGAGCGCTCATATTGTT
ATGGTTGATGCTTACAAGCCAACTAAGGGCGGCGGCGAGCGGTACCGGCGACGTGACTGTCGCCCCAAGCAACTTC
GCTAACGGGGTCGCTGAATGGATCAGCTCTAACTCGCGTTCACAGGCTTACAAAGTAACCTGTAGCGTTCGTAGCA
GCTCTGCGCAGAATCGCAAATACACCATCAAAGTCGAGGTGCCTAAAGTGGCAACCCAGACTGTTGGTGGTGTAG
AGCTTCTGTAGCCGCATGGCGTTCGACTTAAATATGGAACCTAACATTCCAATTTTCGCTACGAATCCGACTGC

GAGCTTATTGTTAAGGCAATGCAAGGTCTCCTAAAAGATGGAAACCCGATTCCCTCAGCAATCGCAGCAAACCTCCG
GCATCTACTAA

- MS2-SpyTag7:

ATGGCTTCTAACTTTACTCAGTTCGTTCTCGTCGACAATGGCGGAACTGGCGACGTGACTGTGCGCCCAAGCAACT
TCGCTAGCGGGGTCGCTGAATGGATCAGCTCTAACTCGCGTTCACAGGCTTGCAAAGTAACCTGTAGCGTTCGTCA
GAGCTCTGCGCAGAATCGCAAATACGCCATCAAAGTCGAGGTGCCTAAAGTGGCAACCCAGACTGTTGGTGGTGT
AGAGCTTCTGTAGCCGCATGGCGTTCGACTTAAATATGGAACCTAACATTCCAATTTTCGCTACGAATCCGACT
GCGAGCTTATTGTTAAGGCAATGCAAGGTCTCCTAAAAGATGGAAACCCGATTCCCTCAGCAATCGCAGCAAACCTC
CGGCATCTACGCTAACTTTACTCAGTTCGTTCTCGTCGACAATGGCGGTACCGGTGGTAGTGGCGGCGGCAGCGCT
CATATTGTTATGGTTGATGCTTACAAGCCAATAAGGGCGGCGGCAGCGGTGGTAGTGGTACCGGCGACGTGACT
GTCGCCCCAAGCAACTTCGCTAACGGGGTCGCTGAATGGATCAGCTCTAACTCGCGTTCACAGGCTTACAAAGTAA
CCTGTAGCGTTCGTGAGAGCTCTGCGCAGAATCGCAAATACACCATCAAAGTCGAGGTGCCTAAAGTGGCAACCC
AGACTGTTGGTGGTGTAGAGCTTCTGTAGCCGCATGGCGTTCGACTTAAATATGGAACCTAACATTCCAATTTTC
GCTACGAATCCGACTGCGAGCTTATTGTTAAGGCAATGCAAGGTCTCCTAAAAGATGGAAACCCGATTCCCTCAG
CAATCGCAGCAAACCTCCGGCATCTACTAA

- MS2-Random4:

ATGGCTTCTAACTTTACTCCGTTTCGTTCTCGTCGACAATGGCGGAACTGGCGACGTGACTGTGCGCCCAAGCAACTT
CGCTAACGGGGTCGCTGAATGGATCAGCTCTAACTCGCGTTCACAGGCTTACAAAGTAACCTGTAGCGTTCGTCA
AGCTCTGCGCAGAATCGCAAATACACCATCAAAGTCGAGGTGCCTAAAGTGGCAACCCAGACTGTTGGTGGTGT
GAGCTTCTGTAGCCGCATGGCGTTCGACTTAAATATGGAACCTAACATTCCAATTTTCGCTACGAATCCGACTG
CGAGCTTATTGTTAAGGCAATGCAAGGTCTCCTAAAAGATGGAAACCCGATTCCCTCAGCAATCGCAGCAAACCTCC
GGCATCTACGCTAACTTTACTCAGTTCGTTCTCGTCGACAATGGCGGTACCGGTGGTAGTGGTACCGGCGACGTGACT
CAATTGCTAAGCATGTTAAGGATGTTGGCGGCGGCAGCGGTACCGGCGACGTGACTGTGCGCCCAAGCAACTTCG
CTAACGGGGTCGCTGAATGGATCAGCTCTAACTCGCGTTCACAGGCTTACAAAGTAACCTGTAGCGTTCGTGAGAG
CTCTGCGCAGAATCGCAAATACACCATCAAAGTCGAGGTGCCTAAAGTGGCAACCCAGACTGTTGGTGGTGTAGA
GCTTCTGTAGCCGCATGGCGTTCGACTTAAATATGGAACCTAACATTCCAATTTTCGCTACGAATCCGACTGCG
AGCTTATTGTTAAGGCAATGCAAGGTCTCCTAAAAGATGGAAACCCGATTCCCTCAGCAATCGCAGCAAACCTCCG
CATCTACTAA

References for Supplementary Figures

- 1 Tománek, D. Guide Through the Nanocarbon Jungle: Buckyballs. *Nanotubes, Graphene, and Beyond* (2014).
- 2 Hanwell, M. D. *et al.* Avogadro: an advanced semantic chemical editor, visualization, and analysis platform. *J. Cheminform.* **4**, 1-17 (2012).