Supporting Information

Sortase A-mediated N-terminal modification of Cowpea Chlorotic Mottle Virus for highly efficient cargo loading

Lise Schoonen, Jan Pille, Annika Borrmann, Roeland J.M. Nolte, Jan C. M. van Hest

Institute for Molecules and Materials, Radboud University Nijmegen, Heyendaalseweg 135, 6525 AJ Nijmegen, The Netherlands.

List of contents

1 Materials and methods

1.1 Materials

Hot start II HF DNA polymerase, restriction enzymes, T4 DNA ligase and Antarctic phosphatase were obtained from New England Biolabs. The DNA oligos were synthesized by Biolegio. Ampicillin and kanamycin were purchased from MP Biomedicals. Chloramphenicol was obtained from Sigma-Aldrich. Isopropyl β-D-1-thiogalactopyranoside (IPTG) was purchased from Acros. Ni-NTA agarose beads were obtained from Qiagen. DyLight 650 NHS ester was purchased from ThermoFisher Scientific.

1.2 Buffers

pH 5.0 buffer: 50 mM NaOAc, 500 mM NaCl, 10 mM $MgCl_2$, 1 mM EDTA, pH 5.0.
pH 7.5 buffer: 50 mM TrisHCl, 500 mM NaCl, 10 mM $MgCl_2$, 1 mM EDTA, pH 7.5.
Sortase buffer: 50 mM HEPES, 150 mM NaCl, 5 mM $CaCl_2$, pH 7.5.
TBST buffer: 50 mM TrisHCl, 150 mM NaCl, 0.05% Tween-20, pH 8.0.

1.3 UV-vis measurements

Protein concentrations were measured on a Varian Cary 50 Conc UV-vis spectrometer using a quartz cuvette with a path length of 3 mm. Protein concentrations were calculated using the theoretical extinction coefficients.[1] Samples were centrifuged prior to the measurements.

1.4 Western blot analysis

Protein samples were loaded on a 12% (w/v) polyacrylamide gel. Proteins were separated (150 V) and transferred on a nitrocellulose membrane using the TransBlot Turbo semi dry transfer system (BioRad) according to manufacturer's protocol. The membrane was blocked in 5% milk in TBST buffer. The membrane was incubated with mouse anti-GFP antibody (1:3000 dilution in 5% milk in TBST buffer, 11814460001 Roche) overnight at 4 °C. The membrane was washed 3 x 5 minutes in TBST buffer and incubated with rabbit anti-mouse antibody (1:3000 dilution in 5% milk in TBST buffer, P0161 Dako) for 1h at rt. The membrane was washed 3 x 5 minutes in TBST buffer. GFP expression was subsequently visualized using the SuperSignal Chemiluminescent substrates kit (Thermo Scientific) (ImageQuant LAS4000, GE Healthcare).

1.5 Mass spectrometry

Protein mass characterization was performed by electrospray ionization time-of-flight (ESI-TOF) on a JEOL AccuTOF CS. Deconvoluted mass spectra were obtained using MagTran 1.03 b2. Isotopically averaged molecular weights were calculated using the 'Protein Calculator v3.4' at http://protcalc.sourceforge.net. Protein samples were desalted and concentrated to 10-100 μM by spin filtration with MQ.

1.6 Size exclusion chromatography (SEC)

SEC measurements were performed on a Superose 6 10/300 column or a Superdex 75 PC 10/300 column (GE Healthcare). Analytical and preparative SEC measurements were executed on a Shimadzu LC-2010AHT HPLC and Agilent 1260 bio-inert HPLC, respectively. Samples (50-200 μg) were separated on the column with a flow rate of 0.5 mL/min.

1.7 Transmission electron microscopy (TEM)

TEM grids (FCF-200-Cu, EMS) were glow-discharged using a Cressington carbon coater and power unit. Protein samples (0.2 mg/mL, 5 μL) were applied on the glow-discharged grids and incubated for 1 min. The samples were carefully removed using a filter paper and the grid was allowed to dry for at least 15 minutes. Then the grid was negatively stained by applying 2% uranyl acetate in water (5 μL). The staining solution was removed after 15 seconds and the grid was allowed to dry for at least 15 minutes. The samples were analyzed on a JEOL JEM-1010 TEM.

1.8 Reversed phase HPLC

Reversed phase HPLC purifications were performed on a Shimadzu LC-20A system equipped with a Reprosil C18 column, 250 x 10 mm (particle size: 5 μm) (Dr. Maisch GmbH, Screening Devices, Amersfoort, The Netherlands). Elution was achieved using an acetonitrile/water gradient containing 0.1% trifluoroacetic acid (5-100% MeCN, 5-55 min, flow 4 mL/min).

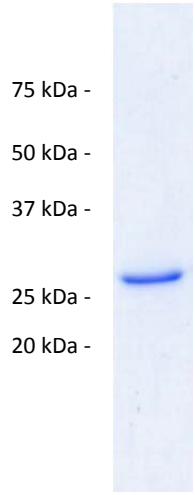
1.9 LC-MS measurements

LC-MS was performed on a Thermo Finnigan LCQ-Fleet ESI-ion trap (Thermo Fischer) equipped with an Alltima C18 column (50 mm x 2 mm, particle size 3 μm) (Alltech Applied Sciences BV). An CH_3CN/H_2O gradient containing 0.1% HCO_2H was used for elution (5-100%, over 15 min). Samples were diluted in MeOH to a concentration of 10-15 μM.

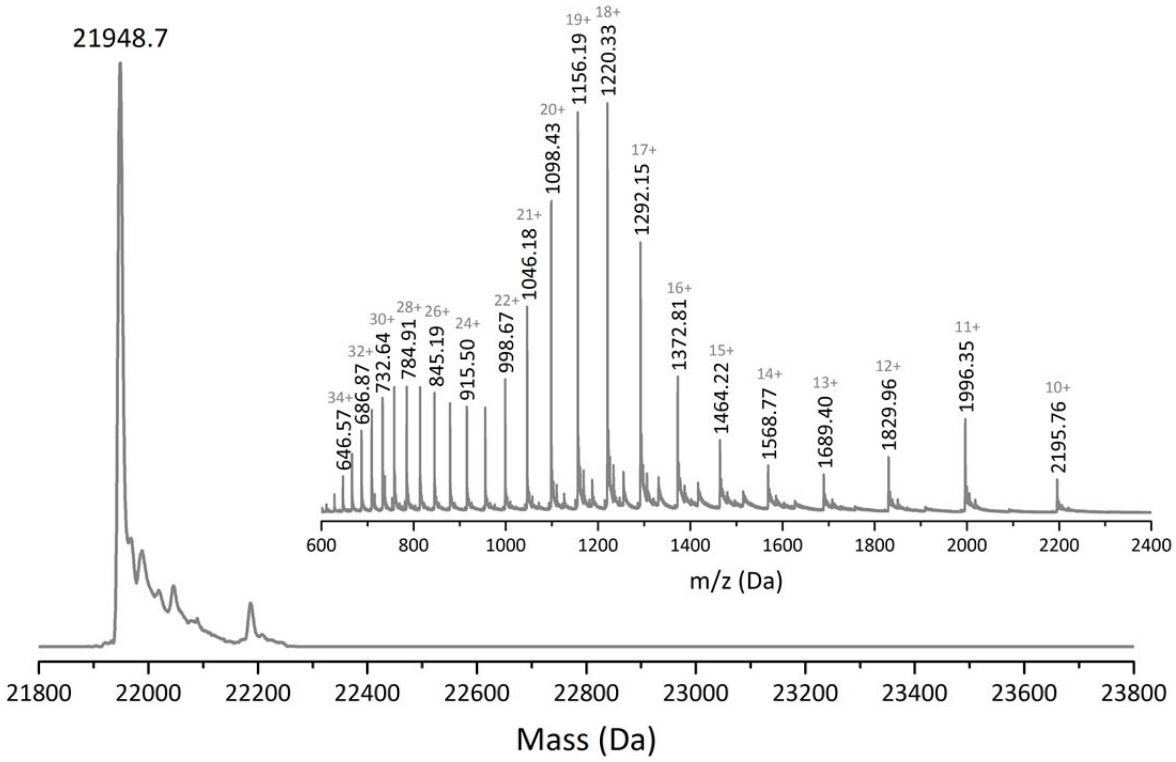
2 Experimental section

2.1 Expression of Sortase A

E.coli BL21 AI cells were transformed with a pQE30 plasmid carrying the Sortase gene, followed by incubation in LB medium (1 mL) for 1h at 37 °C.[2] After this short incubation phase, the cells were transferred into fresh LB medium (4 mL) with ampicillin (100 mg/L) and were incubated at 37 °C for 4h. This preculture was then transferred into TB medium (500 mL) with ampicillin (100 mg/L) and cells were incubated for 24h at 37 °C. Cells were pelleted and resuspended in lysis buffer (50 mM NaH_2PO_4, 300 mM NaCl, 5 mM imidazole and supplemented with 1 mM phenylmethanesulfonyl fluoride, pH 8.0) and lysed by sonication. The lysate was centrifuged (14.000 g, 30 min, 4 °C) and the supernatant was incubated with Ni-NTA beads for 2 h at 4 °C. Ni-NTA beads were washed with wash buffer (50 mM NaH_2PO_4, 300 mM NaCl, 10 mM imidazole, pH 8.0) and the purified protein was eluted from the beads with elution buffer (50 mM NaH_2PO_4, 300 mM NaCl, 500 mM imidazole, pH 8.0). For storage the protein was dialyzed against Sortase buffer. The purity was verified by SDS-PAGE. ESI-TOF: calculated 21947.5 Da, found 21948.7 Da.

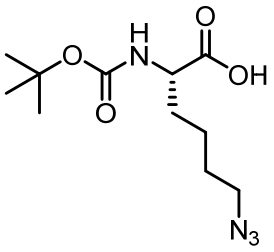


SDS-PAGE of purified Sortase A.

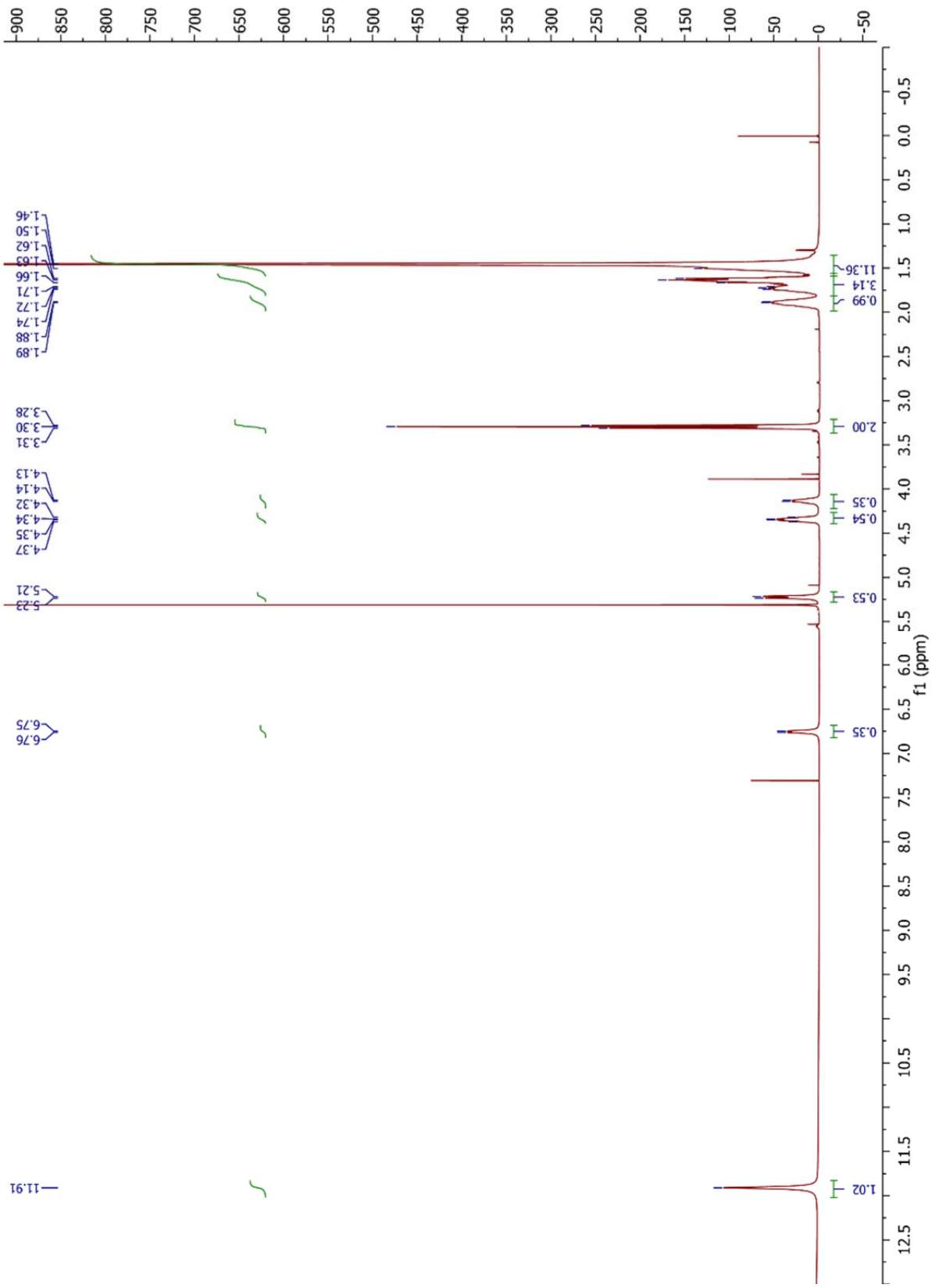


ESI-TOF mass spectrometry of purified Sortase A. Deconvoluted total mass spectrum and multiply charged ion series (inset). The expected molecular weight is 21947.5 Da.

2.2 Synthesis of Boc-azidonorleucine

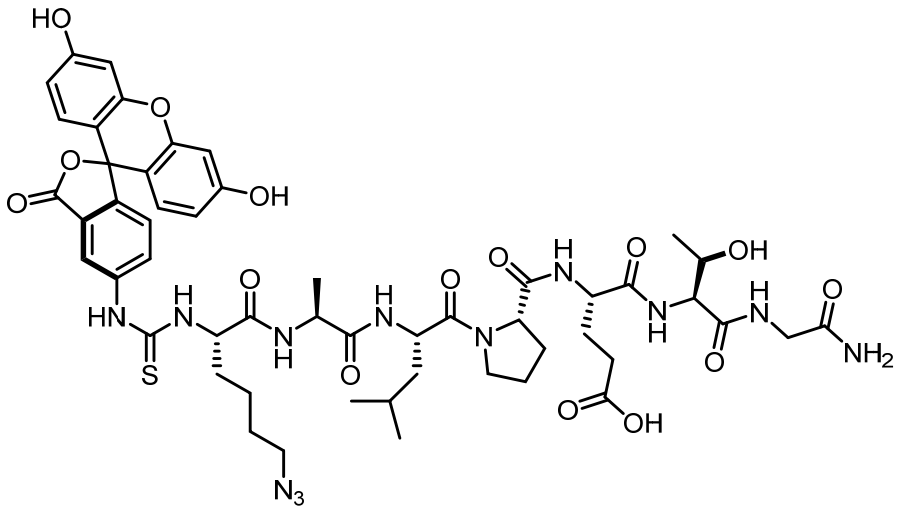


To a solution of Boc-protected L-lysine (2.5 g, 10.15 mmol) in MeOH (50 mL) was added K_2CO_3 (2.23 g, 18.09 mmol, 1.8 eq.), copper(II) sulfate pentahydrate (253 mg, 1.01 mmol, 0.1 eq.) and imidazole-1-sulfonyl azide HCl salt (3.18 g, 15.17 mmol, 1.5 eq.). After stirring for 16 h at rt, half of the solvent was evaporated *in vacuo* before the solution was acidified with 2M HCl. The mixture was extracted with DCM (3×100 mL) and the organic layer was dried with Na_2SO_4 and concentrated *in vacuo*. The crude product was then redissolved in DCM (100 mL) and extracted with 5% aq. $NaHCO_3$ (3×100 mL). The combined aqueous layers were washed twice with DCM (100 mL), acidified with 1M HCl and extracted with DCM (3×100 mL). The organic layer was dried with Na_2SO_4 and concentrated *in vacuo* to afford the product as a colorless oil (2.45 mg, 88%). R_f = 0.33 (1:9, MeOH:DCM). ESI-MS: m/z for $C_{11}H_{20}N_4O_4Na^+$ (M + Na^+): calculated 295.13822, found 295.13910. ^{1}H NMR ($CDCl_3$, 400 MHz): δ 11.91 (broad s, 1H), 6.82-5.19 (m, 1H), 4.41-4.08 (m, 1H), 3.3 (t, *J* = 6.8 Hz, 3H), 1.97-1.81 (m, 1H), 1.81-1.58 (m, 3H), 1.56-1.37 (m, 11H).[3]



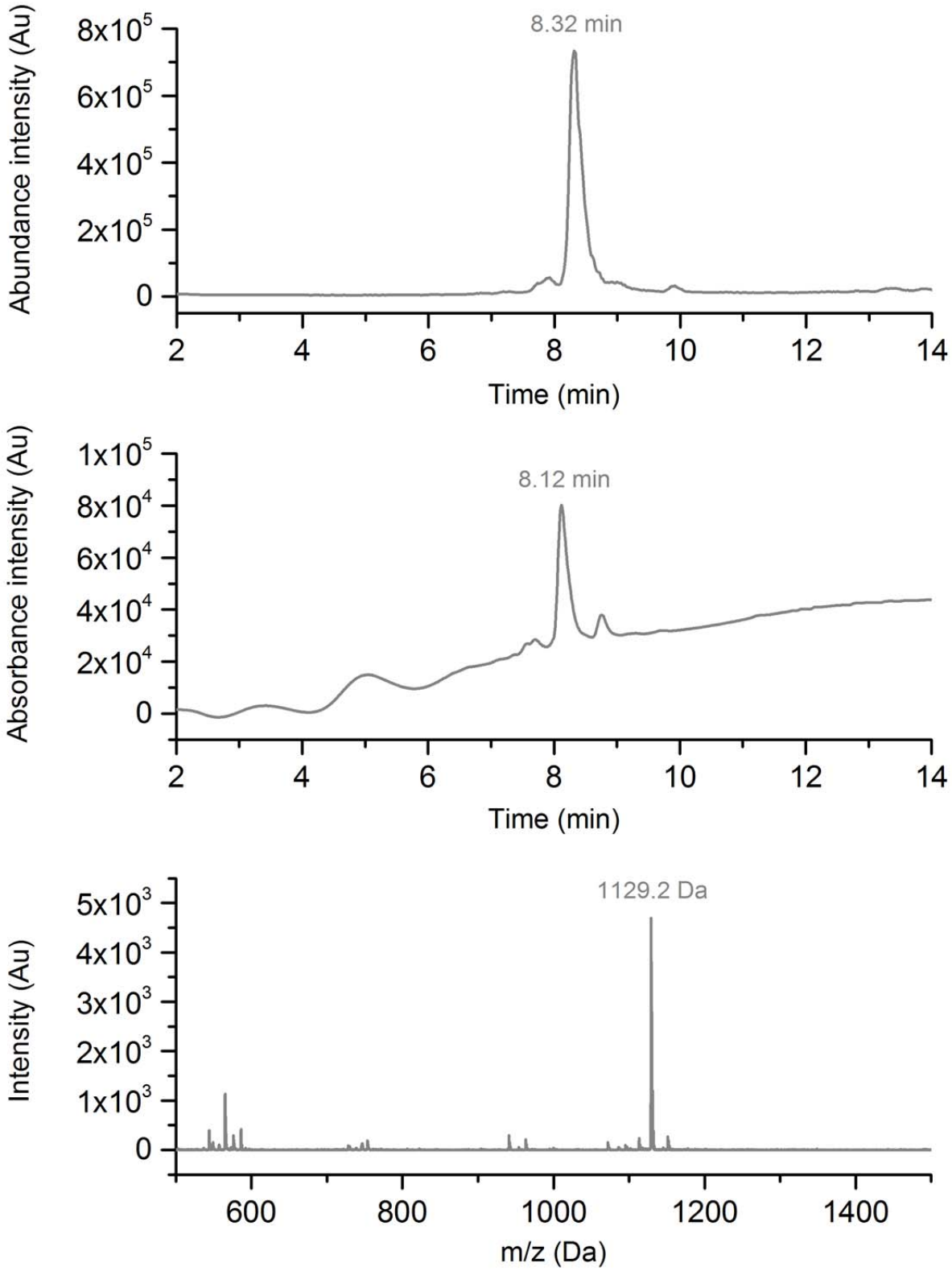
[1]H-NMR spectrum of Boc-azidonorleucine.

2.3 Synthesis of FITC-Anl-Ala-Leu-Pro-Glu-Thr-Gly-NH_2



Anl-Ala-Leu-Pro-Glu-Thr-Gly-NH_2 was synthesized from 2.0 g Fmoc-4-methoxy-4'-(γ-carboxypropyloxy)-benzhydrylamine linked to alanyl-aminomethyl resin (Bachem) using Fmoc solid-phase peptide synthesis (SPPS). The resin was swollen in DMF for 20 min prior to use. The Fmoc-groups from the resin were removed with piperidine in DMF (20%, v/v) for 30 min. Functionalization of the resin with the first Fmoc-protected amino acid was performed for 45 min with 3 eq. of Fmoc-Gly-OH and 3.3 eq. of *N*,*N*'-diisopropylcarbodiimide (DIPCDI) and 3.6 eq. of *N*-hydroxybenzotriazole (HOBt) in DMF. Deprotection of the Fmoc-groups was done with piperidine in DMF (20%, v/v) for 30 min. For all subsequent coupling reactions, again 3 eq. of the required Fmoc-protected amino acid, 3.3 eq. of DIPCDI and 3.6 eq. HOBt in DMF was added to the resin and allowed to react for 45 min. After each coupling reaction and Fmoc deprotection, a Kaiser test was performed to ensure completion of the reaction. Boc-protected azidonorleucine (3 eq.) was added as the final residue. The peptide was cleaved from the resin by treatment with a mixture of 1,2-ethanedithiol:H_2O:TFA (2.5:2.5:95) for 3 h. The peptide was precipitated in Et_2O and separated from solvent by centrifugation yielding a white solid.

To a solution of crude peptide (20 mg, 0.027 mmol) in DMF (2 mL) was added fluorescein isothiocyanate isomer I (11.6 mg, 0.027 mmol, 1.0 eq.) and DIPEA (9.4 µL, 0.054 mmol, 2.0 eq.). After 2.5 hours stirring at rt, the mixture was concentrated *in vacuo*. The crude peptide was purified by reversed phase HPLC. The fractions containing the peptide were combined and lyophilized yielding a yellow powder. LC-MS: m/z for $C_{52}H_{65}N_{12}O_{15}S^+$: ($M + H^+$) calculated 1130.2, found 1129.2 (retention time: 8.32 min).

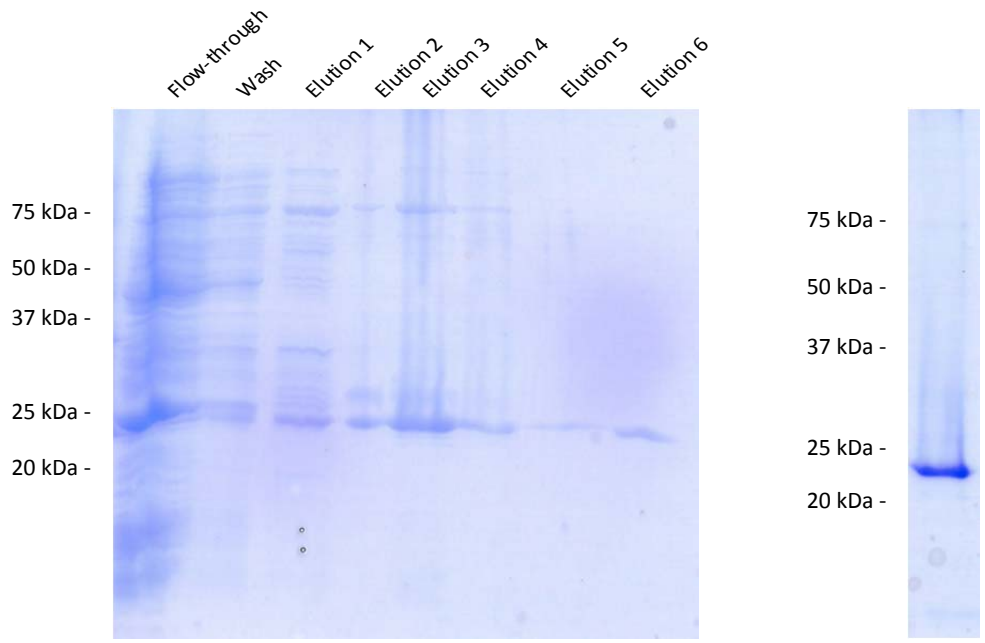


LS-MS analysis of FITC-Anl-Ala-Leu-Pro-Glu-Thr-Gly-NH_2. Top: Abundance intensity, middle: absorbance intensity, bottom: mass spectrum of the peak at 8.32 min. The expected molecular weight is 1130.2 Da.

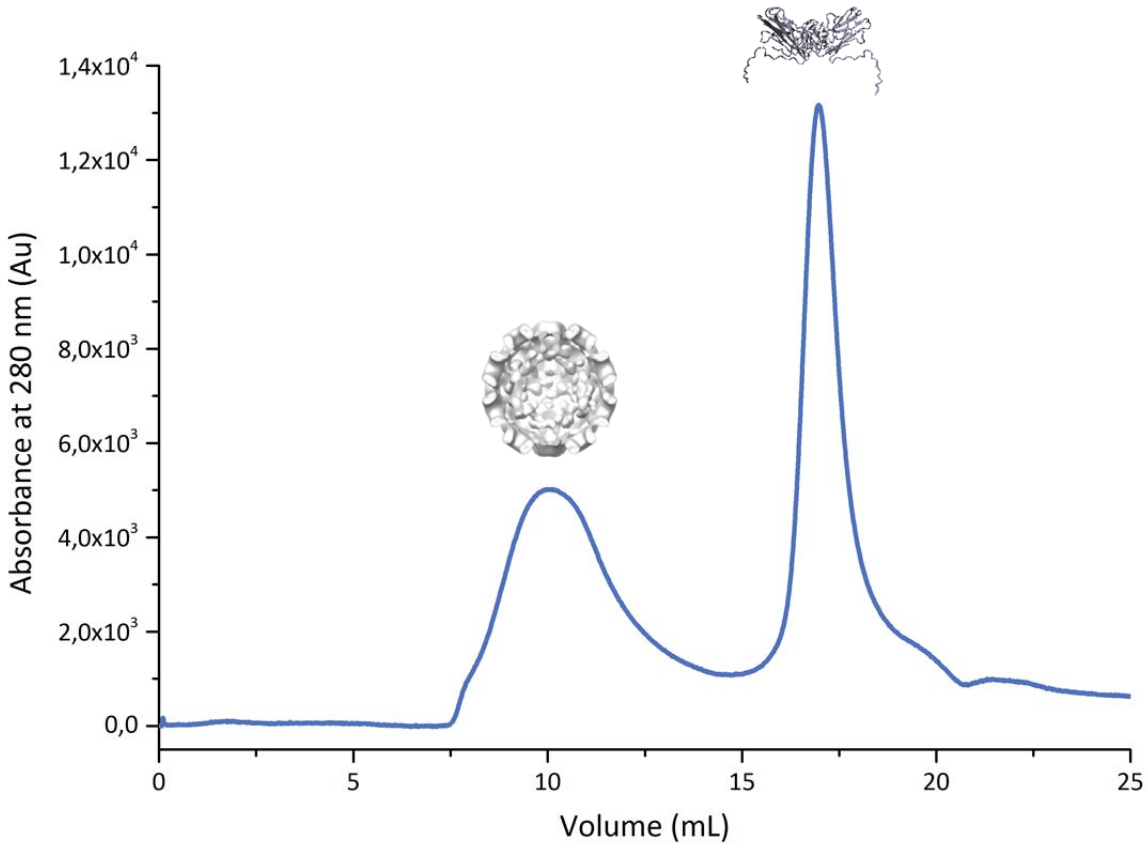
2.4 Expression of G-CCMV

The pET-15b-H_6-CCMV vector encoding for bacterial expression of the hexahistidine-tagged CCMV protein was previously constructed as described by Minten *et al.*[4] For a typical expression, LB medium (50 mL), containing ampicillin (100 mg/L) and chloramphenicol (50 mg/L), was inoculated with a single colony of *E. coli* BLR(DE3)pLysS containing pET-15b-G-H_6-CCMV, and was incubated overnight at 37 °C. This overnight culture was used to inoculate 2xTY medium (1 L), supplemented with ampicillin (100 mg/L) and chloramphenicol (50 mg/L). The culture was grown at 37 °C and protein expression was induced during logarithmic growth (OD_{600} = 0.4-0.6) by addition of IPTG (1 mM). After 6 h of expression at 30 °C, the cells were harvested by centrifugation (2700 g, 15 min, 4 °C) and the pellets were stored overnight at -20 °C.

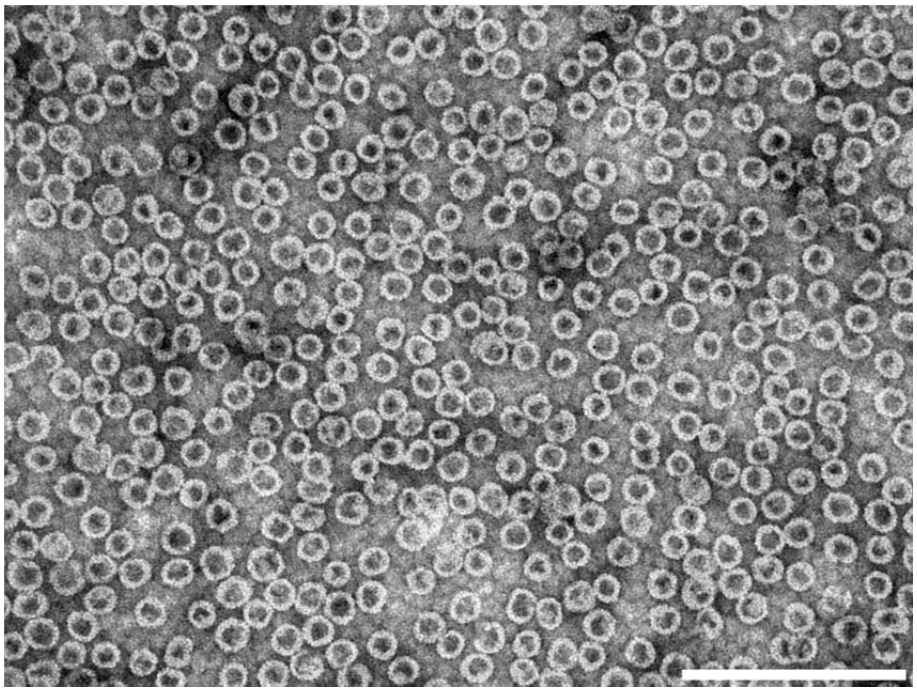
After thawing, the cell pellet was resuspended in lysis buffer (50 mM NaH_2PO_4, 1.5 M NaCl, 10 mM imidazole, pH 8.0; 25 mL). The cells were lysed by ultrasonic disruption (2 times 30 s, 100% duty cycle, output control 3, Branson Sonifier 250, Marius Instruments). The lysate was incubated with DNase (10 mg/L) and RNase A (5 mg/L) for 10 min at 4 °C. Then the lysate was centrifuged (16.400 g, 15 min, 4 °C) to remove the cellular debris. The supernatant was incubated with Ni-NTA agarose beads (3 mL) for 1 h at 4 °C. The suspension was loaded onto a column, the flow-through was collected and the beads were washed twice with wash buffer (50 mM NaH_2PO_4, 1.5 M NaCl, 20 mM imidazole, pH 8.0; 20 mL). Then, the protein of interest was eluted from the column with elution buffer (50 mM NaH_2PO_4, 1.5 M NaCl, 250 mM imidazole, pH 8.0; 1 time 0.5 mL, 5 times 1.5 mL). The purification was analyzed by SDS-PAGE. The fractions containing G-CCMV were combined and dialyzed against pH 7.5 buffer to obtain the protein dimers. For storage the proteins were assembled by dialysis against pH 5.0 buffer. The pure protein was obtained with a yield of 12-20 mg/L of culture. The purity was verified by SDS-PAGE. The geometry and assembly properties were analyzed by SEC using a Superose 6 GL 10/300 column with pH 5.0 buffer as the eluent and TEM. ESI-TOF: calculated 22506.6 Da, found 22507.0 Da.



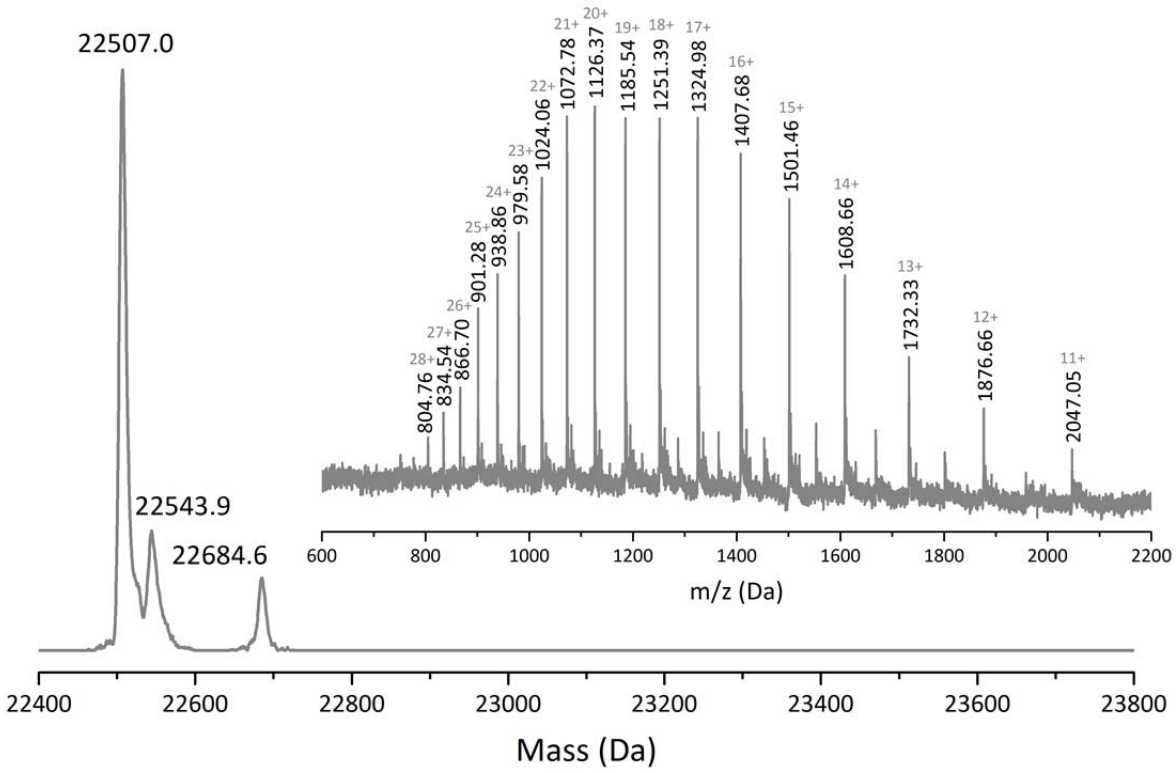
SDS-PAGE of affinity purification of G-CCMV (left) and purified G-CCMV (right).



Size exclusion chromatogram of purified G-CCMV in pH 5 buffer.



Uranyl acetate stained TEM micrograph of G-CCMV. Average particle size = 27.5 ± 1.9 nm. Scale bar corresponds to 200 nm.



ESI-TOF mass spectrometry of purified G-CCMV. Deconvoluted total mass spectrum and multiply charged ion series (inset). The expected molecular weight is 22506.6 Da.

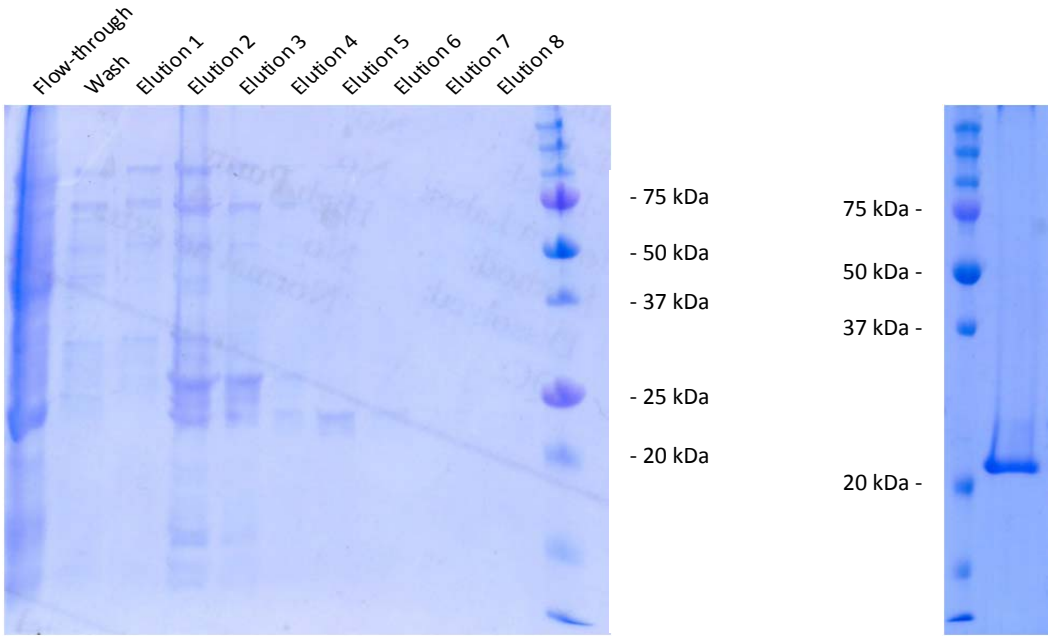
2.5 Cloning of G_3-CCMV

The pET-15b-H_6-CCMV vector encoding for the hexahistidine-tagged CCMV protein was previously constructed as described by Minten *et al.*[4] For the introduction of the G_3-tag, a set of DNA oligos was designed (Table S1). The oligos were annealed and the resulting insert encoded for a G_3-tag, a hexahistidine tag and the CCMV capsid protein with a 5' NcoI and a 3' BamHI restriction site. The product after PCR was purified by agarose gel electrophoresis. Both the purified insert and the pET-15b-H_6-CCMV vector were digested with NcoI-HF® and BamHI-HF® and the products were again purified by agarose gel electrophoresis. Subsequently, the inserts were ligated into the digested vector to yield pET-15b-G_3-H_6-CCMV. This plasmid was transformed into *E. coli* XL1-BLUE cells, the DNA was extracted and the sequence was confirmed by DNA sequencing (Table S2). For expression of G_3-CCMV, the plasmid was transformed into *E. coli* BLR(DE3)pLysS cells (Novagen, MERCK).

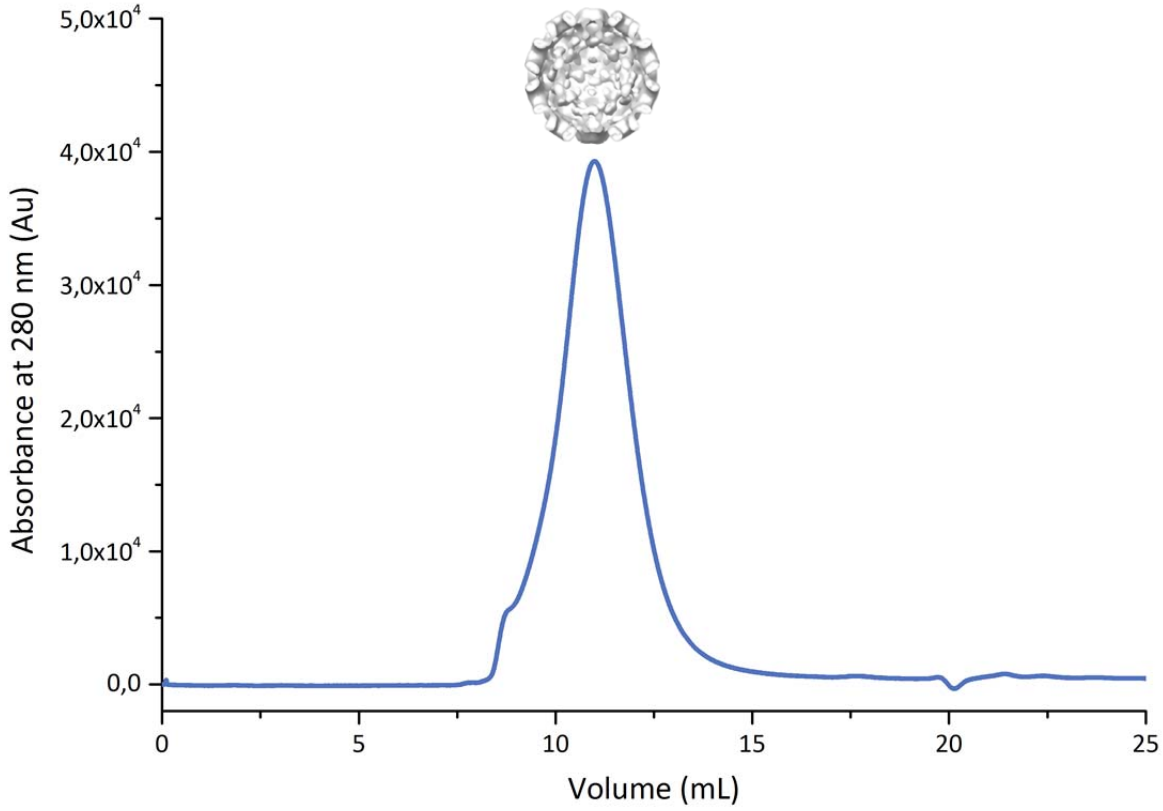
2.6 Expression of G_3-CCMV

For a typical expression, LB medium (50 mL), containing ampicillin (100 mg/L) and chloramphenicol (50 mg/L), was inoculated with a single colony of *E. coli* BLR(DE3)pLysS containing pET-15b-G_3-H_6-CCMV, and was incubated overnight at 37 °C. This overnight culture was used to inoculate 2xTY medium (1 L), supplemented with ampicillin (100 mg/L) and chloramphenicol (50 mg/L). The culture was grown at 37 °C and protein expression was induced during logarithmic growth (OD_{600} = 0.4-0.6) by addition of IPTG (1 mM). After 6 h of expression at 30 °C, the cells were harvested by centrifugation (2700 g, 15 min, 4 °C) and the pellets were stored overnight at -20 °C.

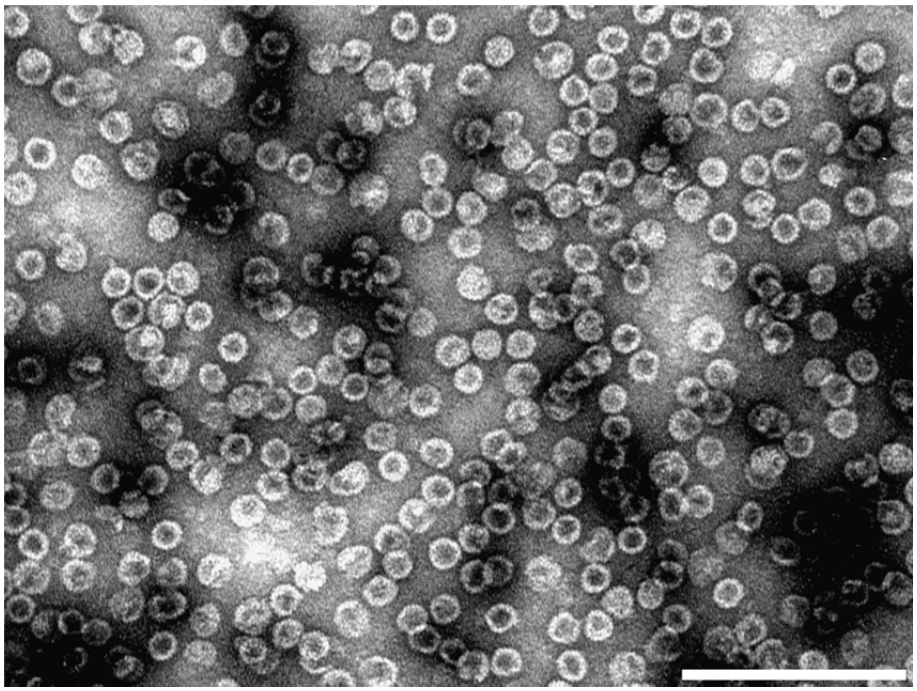
After thawing, the cell pellet was resuspended in lysis buffer (50 mM NaH_2PO_4, 1.5 M NaCl, 10 mM imidazole, pH 8.0; 25 mL). The cells were lysed by ultrasonic disruption (5 times 30 s, 100% duty cycle, output control 3, Branson Sonifier 250, Marius Instruments). The lysate was incubated with DNase (10 mg/L) and RNase A (5 mg/L) for 10 min at 4 °C. Then the lysate was centrifuged (16.400 g, 15 min, 4 °C) to remove the cellular debris. The supernatant was incubated with Ni-NTA agarose beads (3 mL) for 1 h at 4 °C. The suspension was loaded onto a column, the flow-through was collected and the beads were washed twice with wash buffer (50 mM NaH_2PO_4, 1.5 M NaCl, 20 mM imidazole, pH 8.0; 20 mL). Then, the protein of interest was eluted from the column with elution buffer (50 mM NaH_2PO_4, 1.5 M NaCl, 250 mM imidazole, pH 8.0; 1 time 0.5 mL, 7 times 1.5 mL). The purification was analyzed by SDS-PAGE. The fractions containing G_3-CCMV were combined and dialyzed against pH 7.5 buffer to obtain the protein dimers. For storage the proteins were assembled by dialysis against pH 5.0 buffer. The sample was concentrated (Amicon® Ultra-4 Centrifugal Filter Device 10.000 NMWL) and further purification was performed by preparative SEC using a Superose 6 10/300 GL column with pH 5.0 buffer as the eluent. The pure protein was obtained with a yield of 2-14 mg/L of culture. The purity was verified by SDS-PAGE. The geometry and assembly properties were analyzed by SEC using a Superose 6 increase 10/300 column with pH 5.0 buffer as the eluent and TEM. ESI-TOF: calculated 21451.4 Da, found 21451.7 Da.



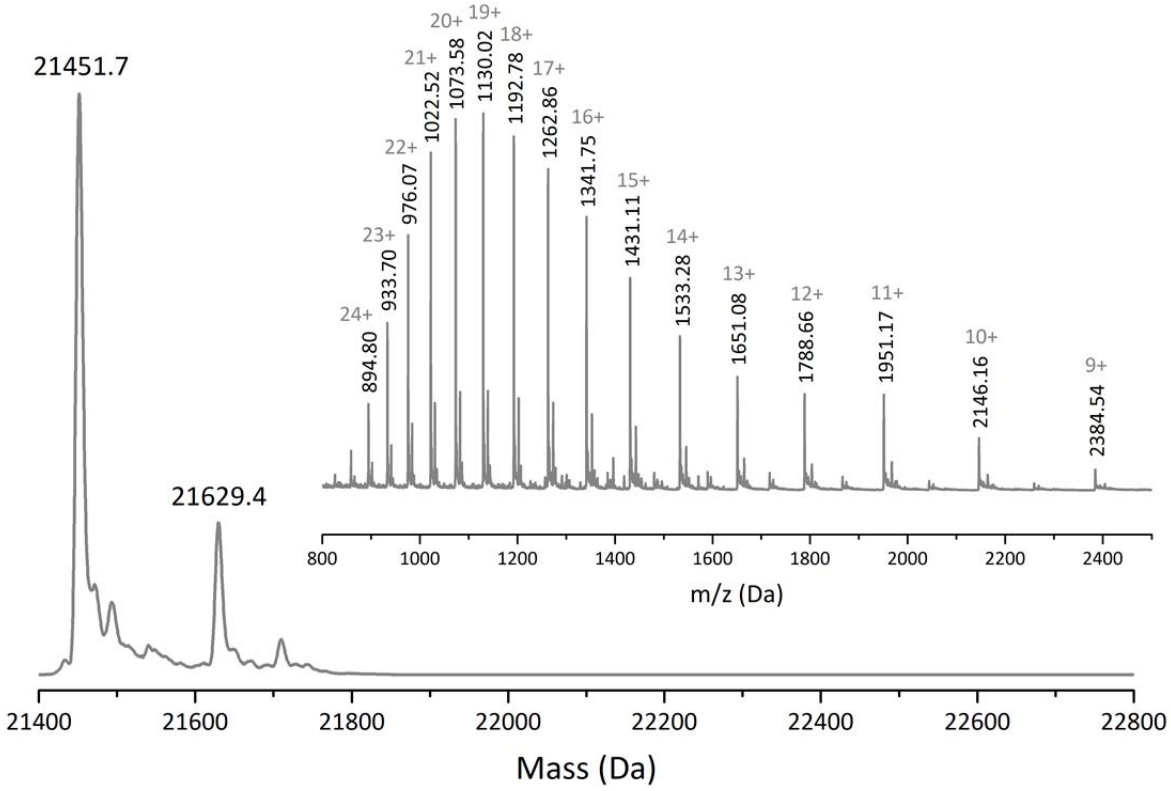
SDS-PAGE of affinity purification of G_3-CCMV (left) and purified G_3-CCMV (right).



Size exclusion chromatogram of purified G_3-CCMV in pH 5 buffer.



Uranyl acetate stained TEM micrograph of G_3-CCMV. Average particle size = 28.6 ± 2.1 nm. Scale bar corresponds to 200 nm.

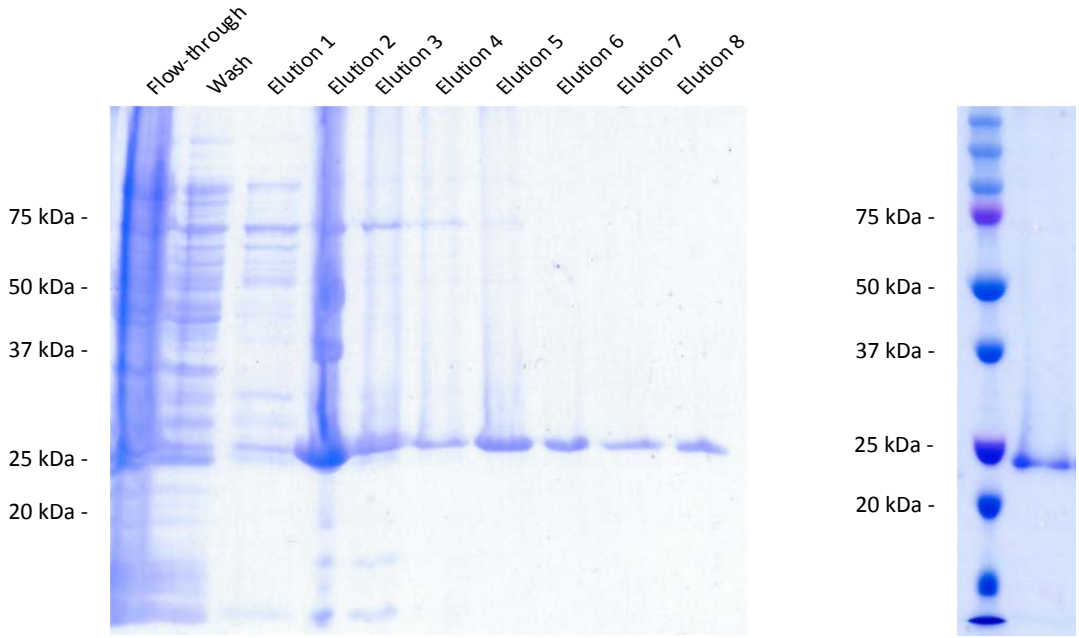


ESI-TOF mass spectrometry of purified G_3-CCMV. Deconvoluted total mass spectrum and multiply charged ion series (inset). The expected molecular weight is 21451.4 Da.

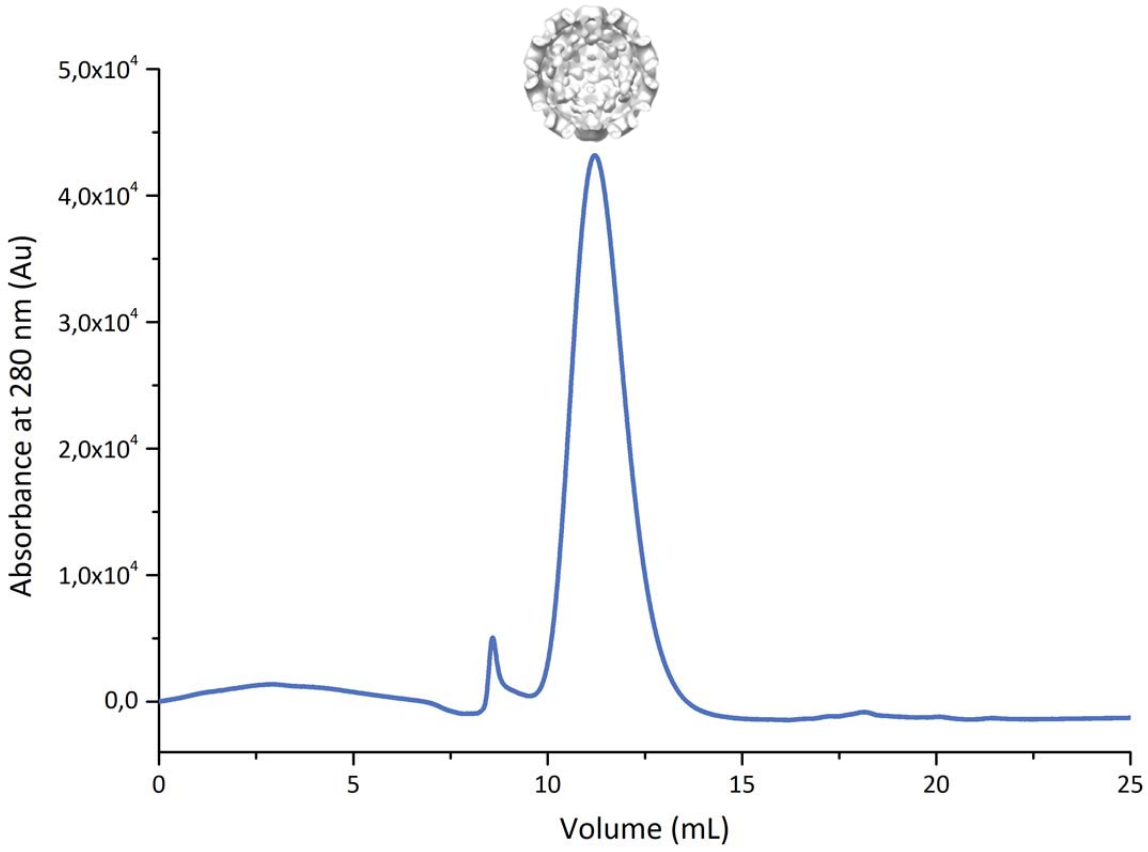
2.7 Expression of G-ELP-CCMV

The pET-15b-G-H_6-[$V_4L_4G_1$-9]-CCMV(ΔN26) vector encoding for the hexahistidine-tagged ELP-CCMV protein was previously constructed as described by van Eldijk *et al.*[5] For a typical expression, LB medium (50 mL), containing ampicillin (100 mg/L) and chloramphenicol (50 mg/L), was inoculated with a single colony of *E. coli* BLR(DE3)pLysS containing pET-15b-G-H_6-[$V_4L_4G_1$-9]-CCMV(ΔN26), and was incubated overnight at 37 °C. This overnight culture was used to inoculate 2xTY medium (1 L), supplemented with ampicillin (100 mg/L) and chloramphenicol (50 mg/L). The culture was grown at 37 °C and protein expression was induced during logarithmic growth (OD_{600} = 0.4-0.6) by addition of IPTG (1 mM). After 6 h of expression at 30 °C, the cells were harvested by centrifugation (2700 g, 15 min, 4 °C) and the pellets were stored overnight at -20 °C.

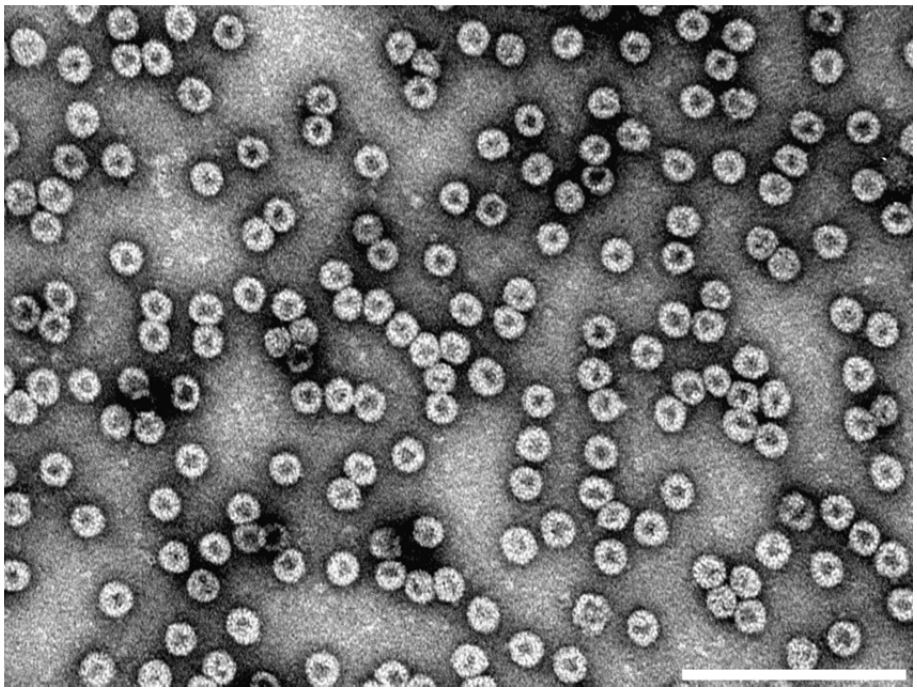
After thawing, the cell pellet was resuspended in lysis buffer (50 mM NaH_2PO_4, 1.3 M NaCl, 10 mM imidazole, pH 8.0; 25 mL). The cells were lysed by ultrasonic disruption (5 times 30 s, 100% duty cycle, output control 3, Branson Sonifier 250, Marius Instruments). The lysate was incubated with DNase (10 mg/L) and RNase A (5 mg/L) for 10 min at 4 °C. Then the lysate was centrifuged (16.400 g, 15 min, 4 °C) to remove the cellular debris. The supernatant was incubated with Ni-NTA agarose beads (3 mL) for 1 h at 4 °C. The suspension was loaded onto a column, the flow-through was collected and the beads were washed twice with wash buffer (50 mM NaH_2PO_4, 1.3 M NaCl, 20 mM imidazole, pH 8.0; 20 mL). Then, the protein of interest was eluted from the column with elution buffer (50 mM NaH_2PO_4, 1.3 M NaCl, 250 mM imidazole, pH 8.0; 1 time 0.5 mL, 7 times 1.5 mL). The purification was analyzed by SDS-PAGE. The fractions containing G-ELP-CCMV were combined and dialyzed against pH 7.5 buffer to obtain the protein dimers. For storage the proteins were assembled by dialysis against pH 5.0 buffer. The pure protein was obtained with a yield of 100 mg/L of culture. The purity was verified by SDS-PAGE. The geometry and assembly properties were analyzed by SEC using a Superose 6 increase 10/300 column with pH 5.0 buffer as the eluent and TEM. ESI-TOF: calculated 22253.4 Da, found 22253.5 Da.



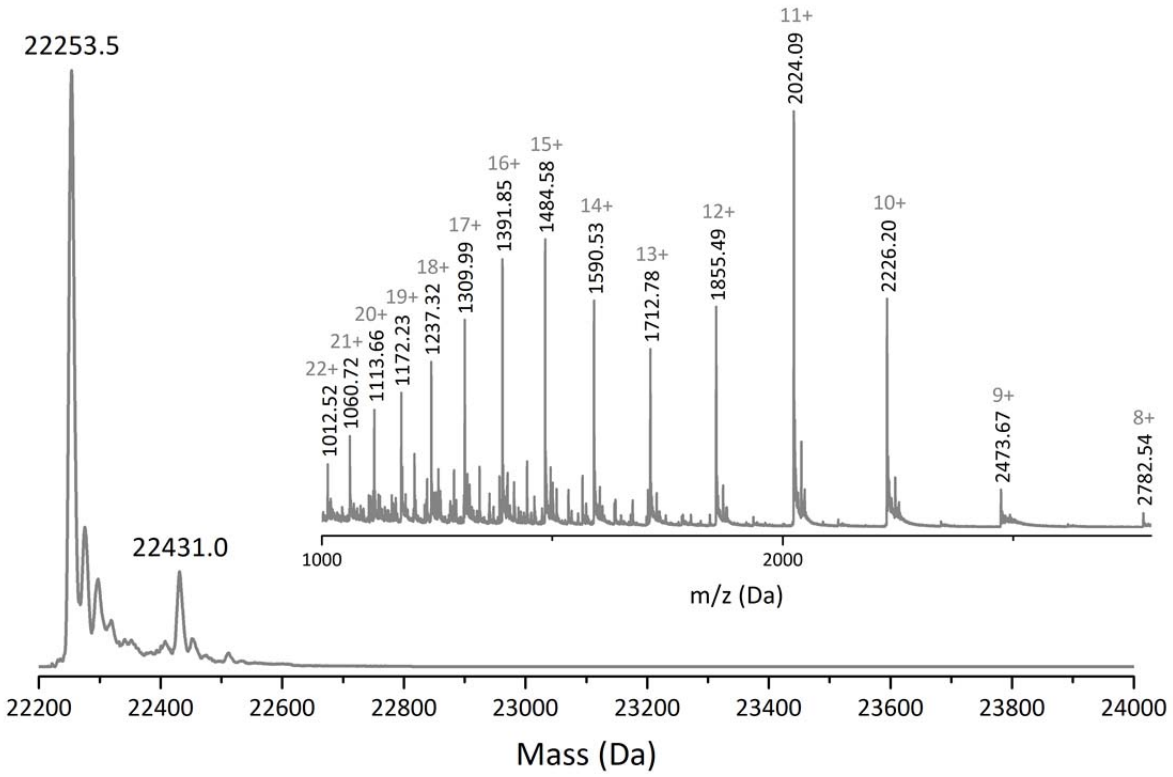
SDS-PAGE of affinity purification of G-ELP-CCMV (left) and purified G-ELP-CCMV (right).



Size exclusion chromatogram of purified G-ELP-CCMV in pH 5 buffer.



Uranyl acetate stained TEM micrograph of G-ELP-CCMV. Average particle size = 29.2 ± 1.5 nm. Scale bar corresponds to 200 nm.



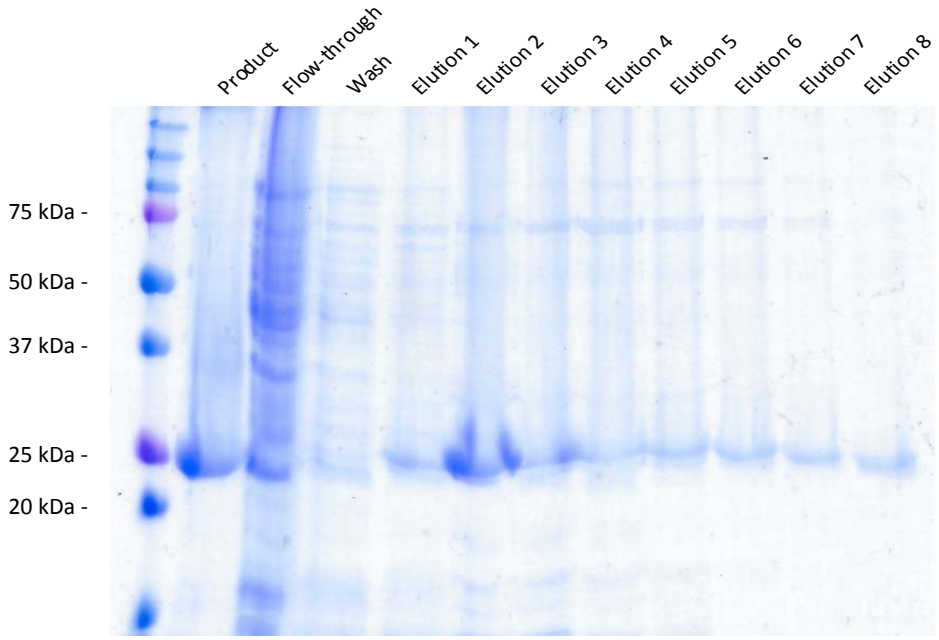
ESI-TOF mass spectrometry of purified G-ELP-CCMV. Deconvoluted total mass spectrum and multiply charged ion series (inset). The expected molecular weight is 22253.4 Da.

2.8 Cloning of G_3-ELP-CCMV

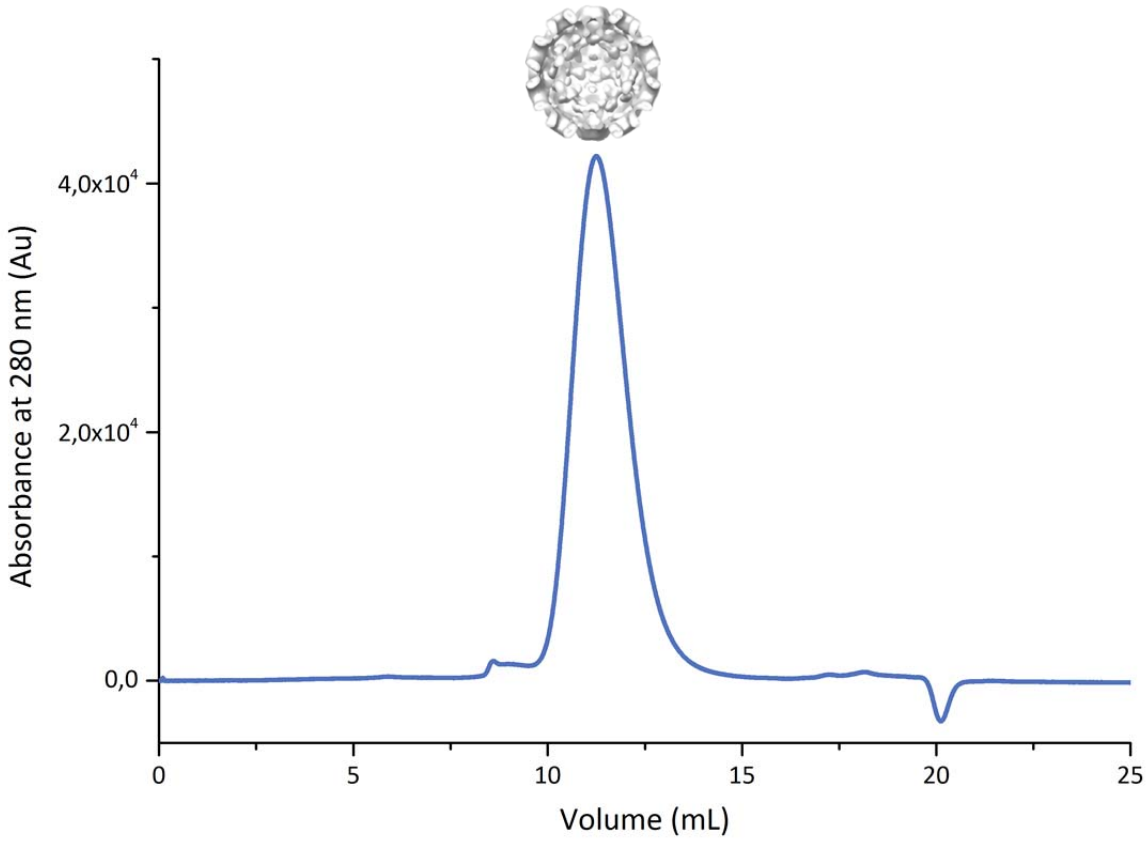
The pET-15b-G-H_6-[$V_4L_4G_1$-9]-CCMV(ΔN26) vector encoding for the hexahistidine-tagged ELP-CCMV protein was previously constructed as described by van Eldijk *et al.*[5] For the introduction of the G_3-tag, a set of DNA oligos was designed (Table S1). The oligos were annealed and the resulting insert encoded for a G_3-tag, a hexahistidine tag, 9 ELP repeats and the CCMV capsid protein with a 5' NcoI and a 3' BamHI restriction site. The product after PCR was purified by agarose gel electrophoresis. Both the purified insert and the pET-15b-G-H_6-[$V_4L_4G_1$-9]-CCMV(ΔN26) vector were digested with NcoI-HF® and BamHI-HF® and the products were again purified by agarose gel electrophoresis. Subsequently, the inserts were ligated into the digested vector to yield pET-15b-G_3-H_6-[$V_4L_4G_1$-9]-CCMV(ΔN26). This plasmid was transformed into *E. coli* XL1-BLUE cells, the DNA was extracted and the sequence was confirmed by DNA sequencing (Table S2). For expression of G_3-CCMV, the plasmid was transformed into *E. coli* BLR(DE3)pLysS cells (Novagen, MERCK).

2.9 Expression of G_3-ELP-CCMV

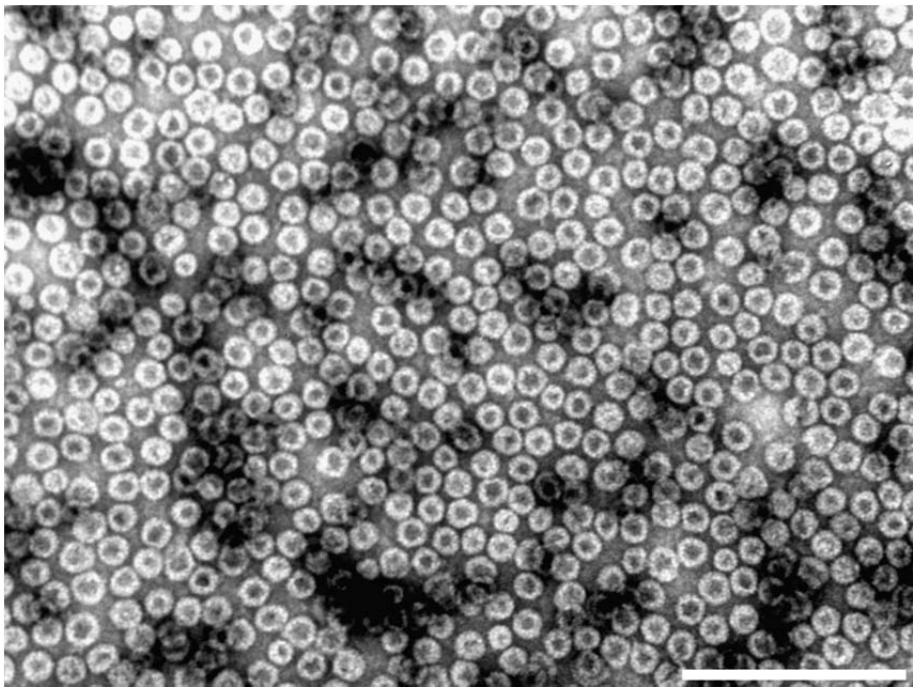
For a typical expression, LB medium (50 mL), containing ampicillin (100 mg/L) and chloramphenicol (50 mg/L), was inoculated with a single colony of *E. coli* BLR(DE3)pLysS containing pET-15b-G_3-H_6-[$V_4L_4G_1$-9]-CCMV(ΔN26), and was incubated overnight at 37 °C. This overnight culture was used to inoculate 2xTY medium (1 L), supplemented with ampicillin (100 mg/L) and chloramphenicol (50 mg/L). The culture was grown at 37 °C and protein expression was induced during logarithmic growth (OD_{600} = 0.4-0.6) by addition of IPTG (1 mM). After 5-6 h of expression at 30 °C, the cells were harvested by centrifugation (2700 g, 15 min, 4 °C) and the pellets were stored overnight at -20 °C.
After thawing, the cell pellet was resuspended in lysis buffer (50 mM NaH_2PO_4, 1.3 M NaCl, 10 mM imidazole, pH 8.0; 25 mL). The cells were lysed by ultrasonic disruption (5 times 30 s, 100% duty cycle, output control 3, Branson Sonifier 250, Marius Instruments). The lysate was incubated with DNase (10 mg/L) and RNase A (5 mg/L) for 10 min at 4 °C. Then the lysate was centrifuged (16.400 g, 15 min, 4 °C) to remove the cellular debris. The supernatant was incubated with Ni-NTA agarose beads (3 mL) for 1 h at 4 °C. The suspension was loaded onto a column, the flow-through was collected and the beads were washed twice with wash buffer (50 mM NaH_2PO_4, 1.3 M NaCl, 20 mM imidazole, pH 8.0; 20 mL). Then, the protein of interest was eluted from the column with elution buffer (50 mM NaH_2PO_4, 1.3 M NaCl, 250 mM imidazole, pH 8.0; 1 time 0.5 mL, 7 times 1.5 mL). The purification was analyzed by SDS-PAGE. The fractions containing G_3-ELP-CCMV were combined and dialyzed against pH 7.5 buffer to obtain the protein dimers. For storage the proteins were assembled by dialysis against pH 5.0 buffer. The pure protein was obtained with a yield of 70 mg/L of culture. The purity was verified by SDS-PAGE. The geometry and assembly properties were analyzed by SEC using a Superose 6 increase 10/300 column with pH 5.0 buffer as the eluent and TEM. ESI-TOF: calculated 22367.6 Da, found 22366.8 Da.



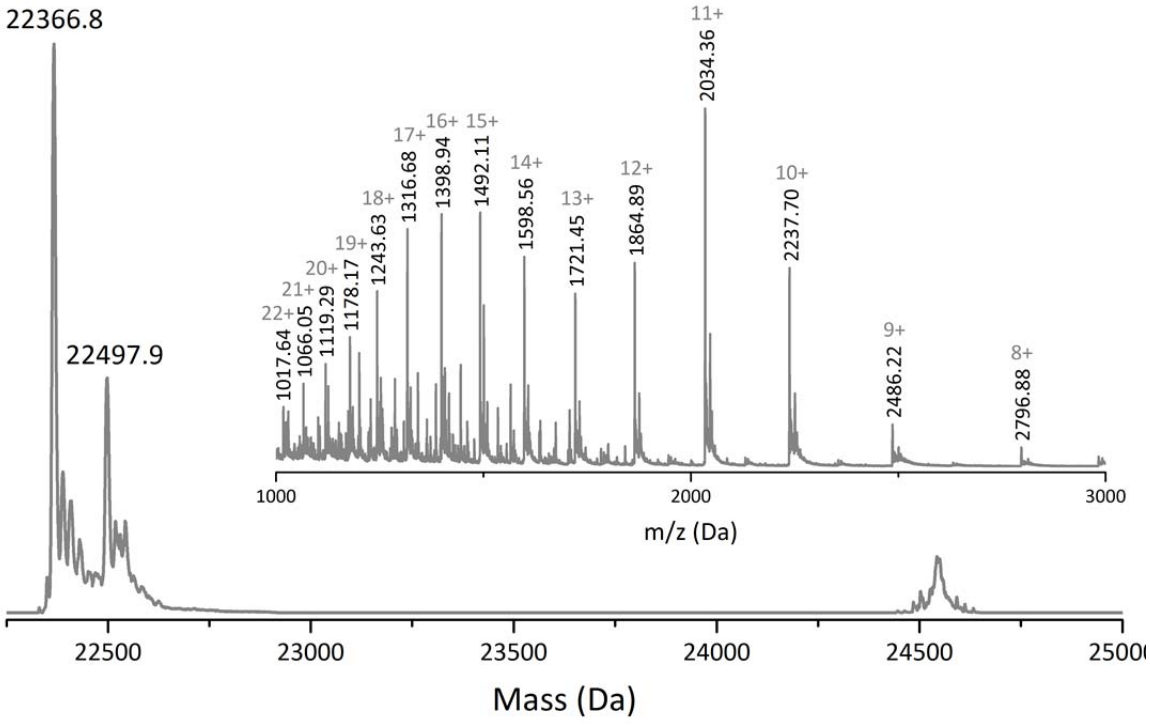
SDS-PAGE of affinity purification of G_3-ELP-CCMV and purified G_3-ELP-CCMV.



Size exclusion chromatogram of purified G_3-ELP-CCMV in pH 5 buffer.



Uranyl acetate stained TEM micrograph of G_3-ELP-CCMV. Average particle size = 26.6 ± 1.6 nm. Scale bar corresponds to 200 nm.

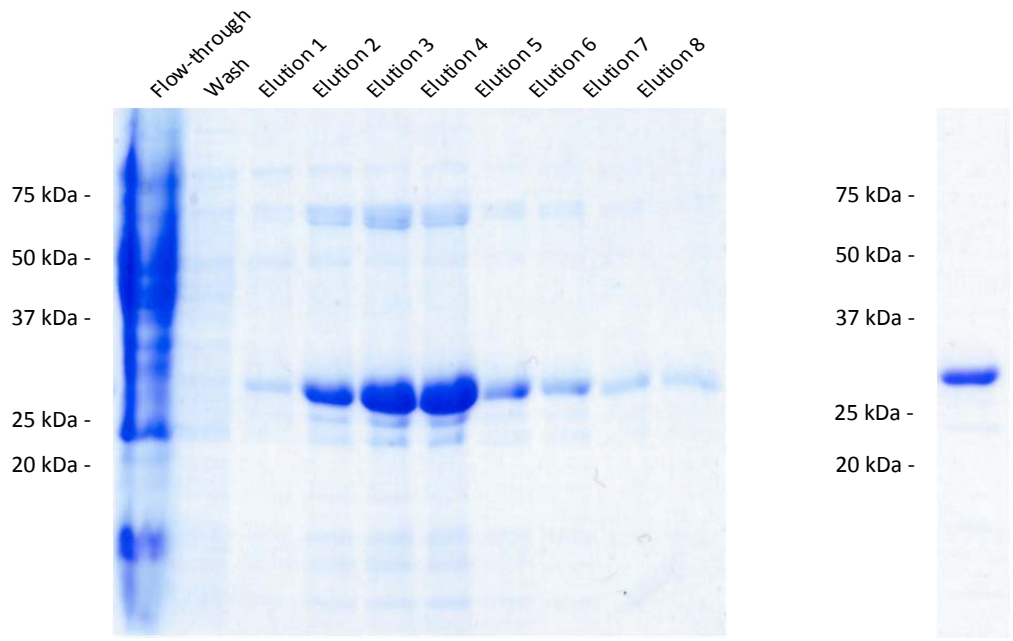


ESI-TOF mass spectrometry of purified G_3-ELP-CCMV. Deconvoluted total mass spectrum and multiply charged ion series (inset). The expected molecular weight is 22367.6 Da.

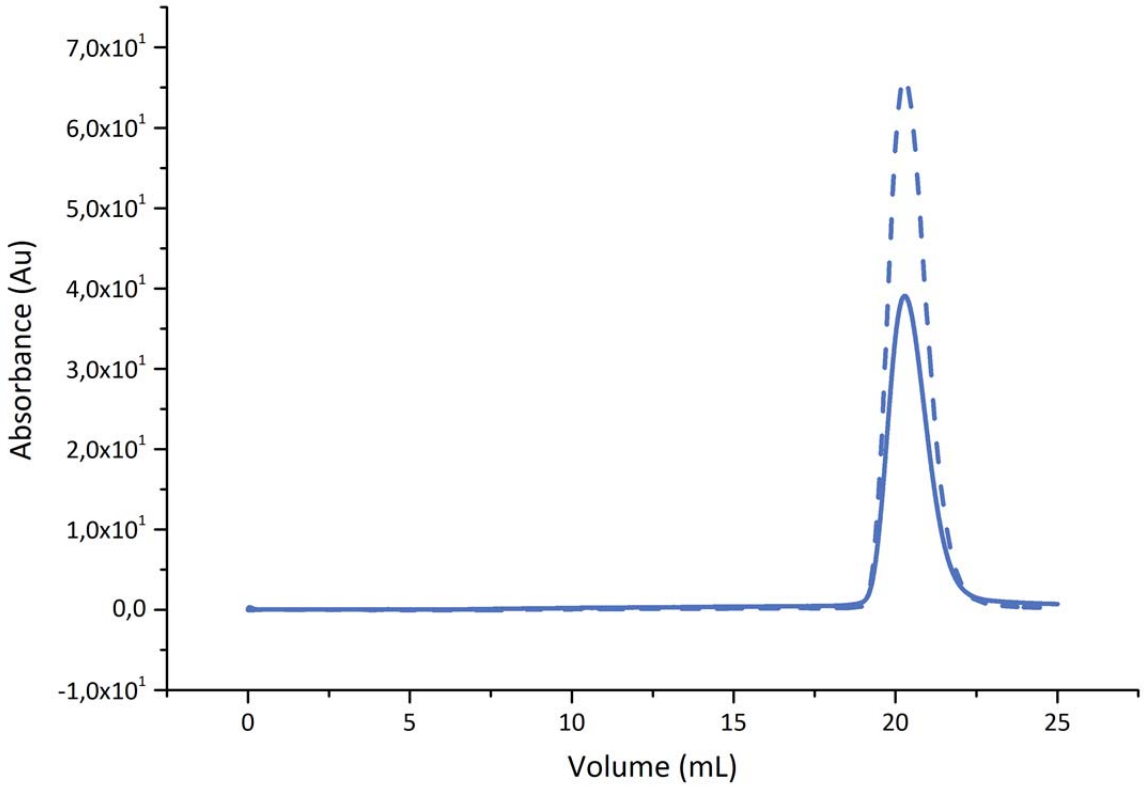
2.10 Expression of GFP-LPETG-H_6

The pLEICS-05-GFP-LPETG-H_6 vector encoding for the hexahistidine-tagged GFP-LPETG protein was previously constructed by the group of Geerten W. Vuister (Department of Biochemistry, University of Leicester) and kindly donated to our group. For a typical expression, LB medium (50 mL), containing ampicillin (100 mg/L) and chloramphenicol (50 mg/L), was inoculated with a single colony of *E. coli* BLR(DE3)pLysS containing pLEICS-05-GFP-LPETG-H_6 and was incubated overnight at 37 °C. This overnight culture was used to inoculate 2xTY medium (1 L), supplemented with ampicillin (100 mg/L) and chloramphenicol (50 mg/L). The culture was grown at 37 °C and protein expression was induced during logarithmic growth (OD_{600} = 0.4-0.6) by addition of IPTG (1 mM). After 5-6 h of expression at 30 °C, the cells were harvested by centrifugation (2700 g, 15 min, 4 °C) and the pellets were stored overnight at -20 °C.

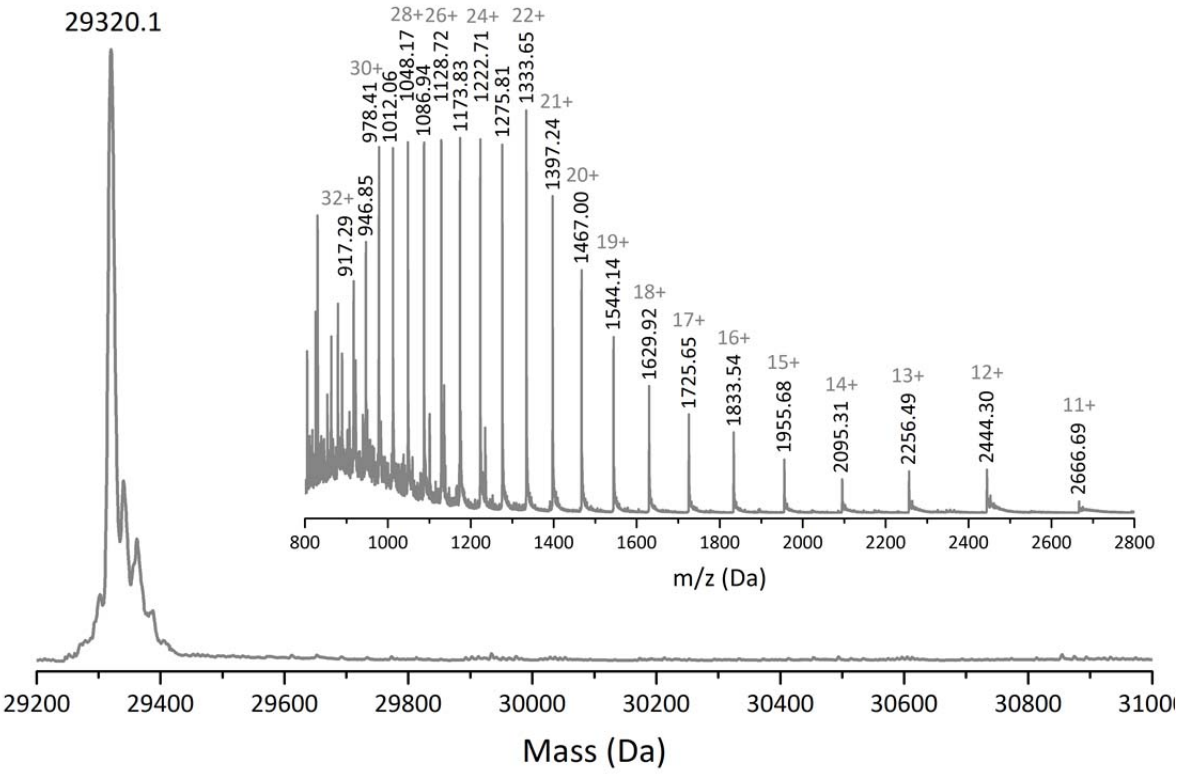
After thawing, the cell pellet was resuspended in lysis buffer (50 mM NaH_2PO_4, 0.5 M NaCl, 10 mM imidazole, pH 8.0; 25 mL). The cells were lysed by ultrasonic disruption (5 times 30 s, 100% duty cycle, output control 3, Branson Sonifier 250, Marius Instruments). Then the lysate was centrifuged (16.400 g, 15 min, 4 °C) to remove the cellular debris. The supernatant was incubated with Ni-NTA agarose beads (3 mL) for 1 h at 4 °C. The suspension was loaded onto a column, the flow-through was collected and the beads were washed twice with wash buffer (50 mM NaH_2PO_4, 0.5 M NaCl, 20 mM imidazole, pH 8.0; 20 mL). Then, the protein of interest was eluted from the column with elution buffer (50 mM NaH_2PO_4, 0.5 M NaCl, 250 mM imidazole, pH 8.0; 1 time 0.5 mL, 7 times 1.5 mL). The purification was analyzed by SDS-PAGE. Elution fractions containing GFP-LPETG-H_6 were combined and concentrated (Amicon® Ultra-15 Centrifugal Filter Device 10.000 NMWL). Further purification was performed by preparative SEC using a Superdex 200 10/300 column and PBS buffer as the eluent. The pure protein was obtained with a yield of 7.5 mg/L of culture. The purity was verified by SDS-PAGE and SEC using a Superose 6 10/300 GL column with PBS buffer as the eluent. ESI-TOF: calculated 29321.9 Da, found 29320.1 Da.



SDS-PAGE of affinity purification of GFP-LPETG-H_6 (left) and purified GFP-LPETG-H_6 (right).



Size exclusion chromatogram of purified GFP-LPETG-H_6 in pH 5 buffer. Solid line = 280 nm, dashed line = 495 nm.



ESI-TOF mass spectrometry of purified GFP-LPETG-H_6. Deconvoluted total mass spectrum and multiply charged ion series (inset). The expected molecular weight is 29321.9 Da.

2.11 SrtA-mediated coupling experiments

For a typical SrtA-mediated coupling experiment, stock solutions of SrtA, CCMV and the LPETG-containing reactant were prepared in Sortase buffer. If a component had been dissolved in another buffer, it was spin filtrated to Sortase buffer (10 kDa MWCO, 3 x 10 min). The components were added together to final concentrations of 0-150 μM SrtA and 50 μM CCMV and LPETG-substrate. The solutions were shaken at 21 ^{o}C for 24 hours. The reaction progress was followed by SDS-PAGE analysis.

2.12 Fluorescent labeling of CCMV

G-ELP-CCMV **1c** (511.3 μM, 100 μL) was diluted with phosphate buffer (0.1 M NaH_2PO_4, 0.15 M NaCl, pH 7.5) (400 μL) and dialyzed against the phosphate buffer. The solution was added to a vial containing DyLight 650 NHS-ester (50 μg) and shaken at 21 ^{o}C for 1 hour in the dark. The excess of dye was removed by dialysis to Sortase buffer. The labeling was confirmed by SDS-PAGE analysis and UV-vis spectroscopy. The dialyzed protein was used for SrtA-mediated coupling experiments.

2.13 Calculation of GFP encapsulation based on absorption ratios

The number of GFP proteins inside the CCMV capsid can be calculated based on the absorption of the capsid peak at 280 nm and 395 nm.[3] CCMV absorbs at 280 nm only, whereas GFP absorbs at both 280 and 395 nm. Therefore, the ratio between these absorbances can be used to calculate the GFP to CCMV ratio. The law of Lambert-Beer states that the absorption (A) is equal to the extinction coefficient (ε) times the concentration (c) times the path length of the cuvette. Since the path length is the same for all measurements, we can omit this from the equation, thus $A=\varepsilon \cdot c$.

$$\frac{A_{280}}{A_{395}} = \frac{A_{280,C}+A_{280,CG}}{A_{395,CG}}$$

We can substitute the absorption by $\varepsilon \cdot c$:

$$\frac{A_{280}}{A_{395}} = \frac{\varepsilon_{280,C} \cdot c_C + \varepsilon_{280,CG} \cdot c_{CG}}{\varepsilon_{395,CG} \cdot c_{CG}}$$

$$\frac{A_{280}}{A_{395}} \cdot \varepsilon_{395,CG} \cdot c_{CG} = \varepsilon_{280,C} \cdot c_C + \varepsilon_{280,CG} \cdot c_{CG}$$

$$\frac{A_{280}}{A_{395}} \cdot \varepsilon_{395,CG} = \frac{\varepsilon_{280,C} \cdot c_C}{c_{CG}} + \varepsilon_{280,CG}$$

$$\frac{\frac{A_{280}}{A_{395}} \cdot \varepsilon_{395,CG} - \varepsilon_{280,CG}}{\varepsilon_{280,C}} = \frac{c_C}{c_{CG}}$$

$$\frac{c_{CG}}{c_C} = \frac{\varepsilon_{280,C}}{\frac{A_{280}}{A_{395}} \cdot \varepsilon_{395,CG} - \varepsilon_{280,CG}}$$

For the calculations, we assumed the following extinction coefficients at pH 5, which have previously been determined by Minten *et al.*[4]:

$\varepsilon_{280,C}=24075\ M^{-1}cm^{-1}$, $\varepsilon_{280,CG}=26348\ M^{-1}cm^{-1}$, $\varepsilon_{395,CG}=28506\ M^{-1}cm^{-1}$

$$\frac{c_{CG}}{c_C} = \frac{24075}{\frac{A_{280}}{A_{395}} \cdot 28506 - 26348}$$

In Figure 4, values of $A_{280}=21.5936$ and $A_{395}=2.0709$ were found. This results in a GFP to CCMV ratio of approximately 1:11. Thus in a capsid, consisiting of 180 capsid proteins, approximately 16 GFP proteins are present.

$$\frac{c_{CG}}{c_C} = \frac{24075}{\frac{21.5936}{2.0709} \cdot 28506 - 26348} = 8.89 \cdot 10^{-2}$$

Table S1 - DNA sequence of the oligos used for the construction of expression vectors in this study.

Name	Sequence
G_3-CCMV Forward	5′-ATATATCCATGGGCGGTGGCCATCATCATCATCACGGTGGCATGTCTACAGTCGGA-3′
G_3-CCMV Reverse	5′-ATATATGGATCCCTAATACACCGGAGT-3′
G_3-ELP-CCMV Forward	5′-ATATATCCATGGGCGGTGGCCATCATCATCATCACGTTCCGGGCGTCGGTGTT-3′
G_3-ELP-CCMV Reverse	5′-ATATATGGATCCCTAATACACCGGAGT-3′

Table S2 - Protein sequences of the proteins used in this study.

Name	Sequence
Sortase A	TGSHHHHHHGSKPHIDNYLHDKDKDEKIEQYDKNVKEQASKDKKQQAKPQIPKDKSKVAGYIEIPDADIKEPVYP GPATPEQLNRGVSFAEENESLDDQNISIAGHTFIDRPNYQFTNLKAAKKGSMVYFKVGNETRKYKMTSIRDVKPT DVGVLDEQKGKDKQLTLITCDDYNEKTGVWEKRKIFVATEVK
wt CCMV	MSTVGTGKLTRAQRRAAARKNKRNTRVVQPVIVEPIASGQGKAIKAWTGYSVSKWTASCAAAEAKVTSAITISLP NELSSERNKQLKVGRVLLWLGLLPSVSGTVKSCVTETQTTAAASFQVALAVADNSKDVVAAMYPEAFKGITLEQL TADLTIYLYSSAALTEGDVIVHLEVEHVRPTFDDSFTPVY
G-CCMV	GSSHHHHHHSSGLVPRGSHMMSTVGTGKLTRAQRRAAARKNKRNTRVVQPVIVEPIASGQGKAIKAWTGYSVSKW TASCAAAEAKVTSAITISLPNELSSERNKQLKVGRVLLWLGLLPSVSGTVKSCVTETQTTAAASFQVALAVADNS KDVVAAMYPEAFKGITLEQLTADLTIYLYSSAALTEGDVIVHLEVEHVRPTFDDSFTPVY
G_3-CCMV	GGGHHHHHHGGMSTVGTGKLTRAQRRAAARKNKRNTRVVQPVIVEPIASGQGKAIKAWTGYSVSKWTASCAAAEA KVTSAITISLPNELSSERNKQLKVGRVLLWLGLLPSVSGTVKSCVTETQTTAAASFQVALAVADNSKDVVAAMYP EAFKGITLEQLTADLTIYLYSSAALTEGDVIVHLEVEHVRPTFDDSFTPVY
G-ELP-CCMV	GHHHHHHVPGVGVPGLGVPGVGVPGLGVPGVGVPGLGVPGGGVPGVGVPGLGLEVVQPVIVEPIASGQGKAIKAW TGYSVSKWTASCAAAEAKVTSAITISLPNELSSERNKQLKVGRVLLWLGLLPSVSGTVKSCVTETQTTAAASFQV ALAVADNSKDVVAAMYPEAFKGITLEQLTADLTIYLYSSAALTEGDVIVHLEVEHVRPTFDDSFTPVY
G_3-ELP-CCMV	GGGHHHHHHVPGVGVPGLGVPGVGVPGLGVPGVGVPGLGVPGGGVPGVGVPGLGLEVVQPVIVEPIASGQGKAIK AWTGYSVSKWTASCAAAEAKVTSAITISLPNELSSERNKQLKVGRVLLWLGLLPSVSGTVKSCVTETQTTAAASF QVALAVADNSKDVVAAMYPEAFKGITLEQLTADLTIYLYSSAALTEGDVIVHLEVEHVRPTFDDSFTPVY
GFP-LPETG-H_6	SVSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFICTTGKLPVPWPTLVTTLTYGVQCFSRY PDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNYNSHN VYIMADKQKNGIKVNFKIRHNIEDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMVLLEFV TAAGITLGMDELYKLPETGENLYFQSLEHHHHHH

3 Supplemental figures

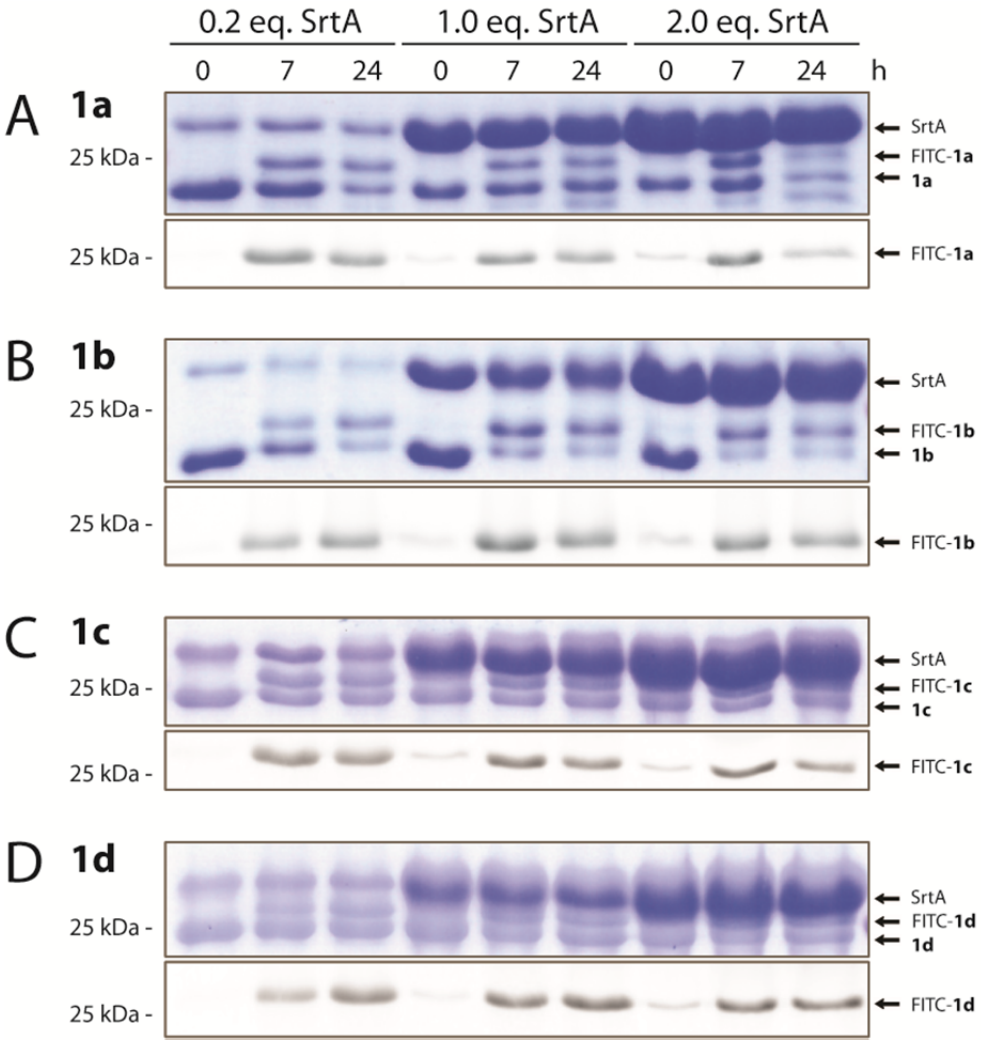


Figure S1 - SDS-PAGE of sortase-mediated N-terminal labeling of CCMV with FITC-Anl-Ala-Leu-Pro-Glu-Thr-Gly-NH_2. Gels were visualized both by Coomassie blue staining (top) and by in-gel fluorescence (bottom). A: Modification of G-CCMV **1a**. B: Modification of G_3-CCMV **1b**. C: Modification of G-ELP-CCMV **1c**. D: Modification of G_3-ELP-CCMV **1d**.

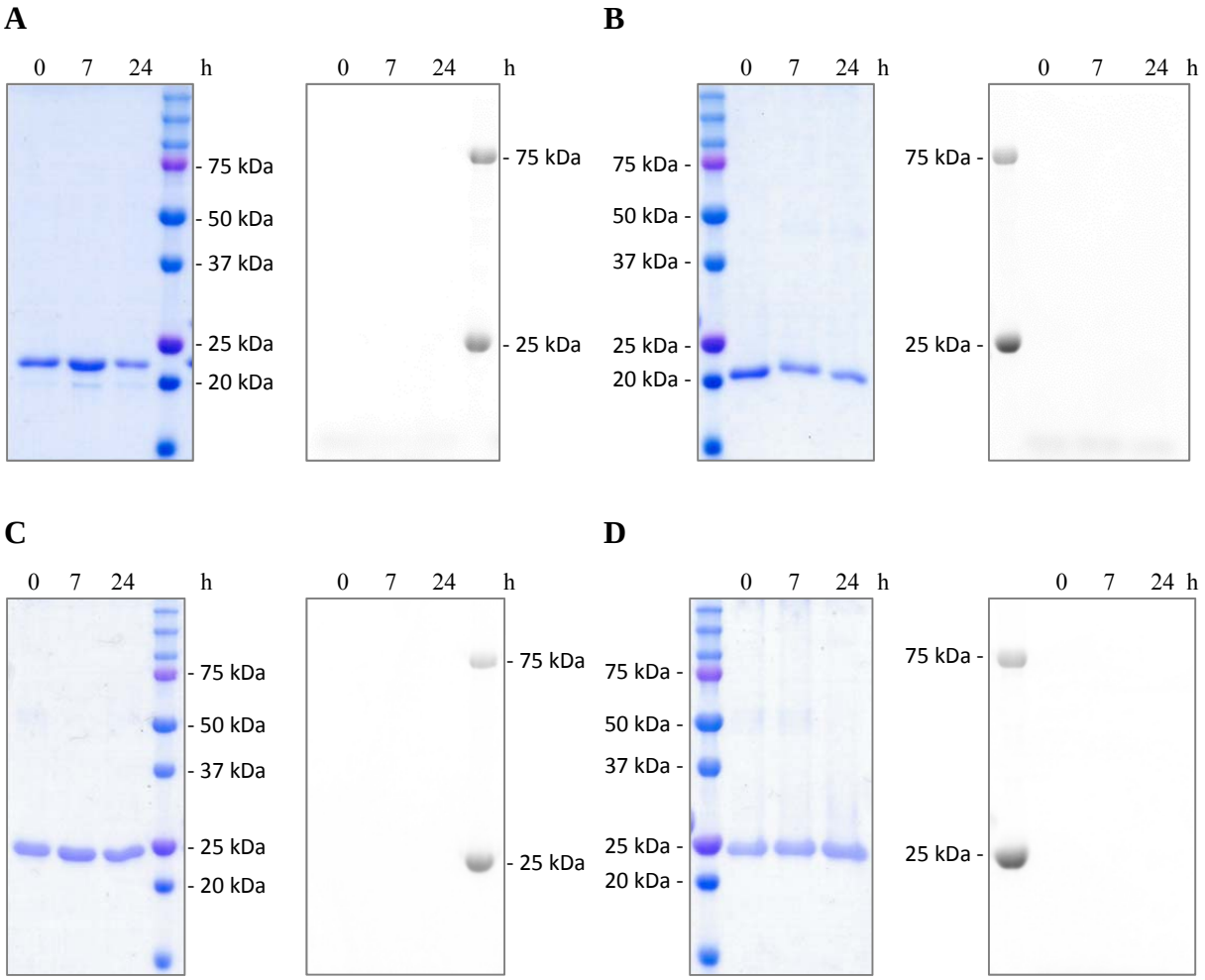
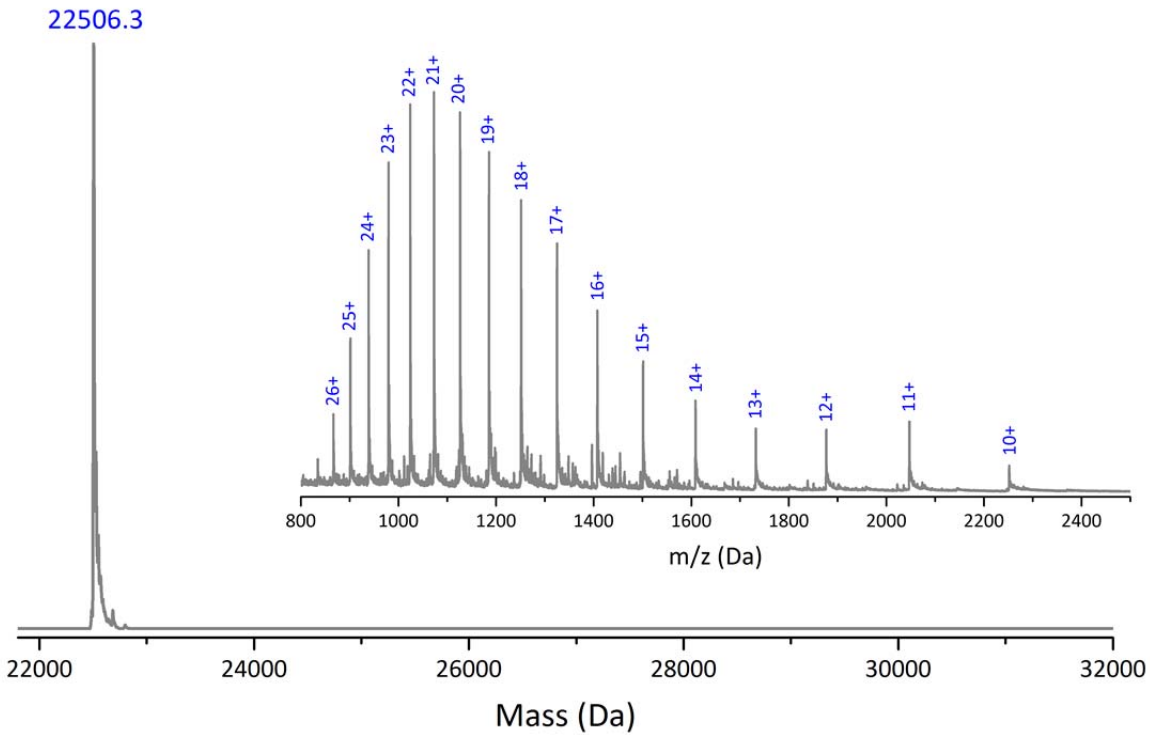
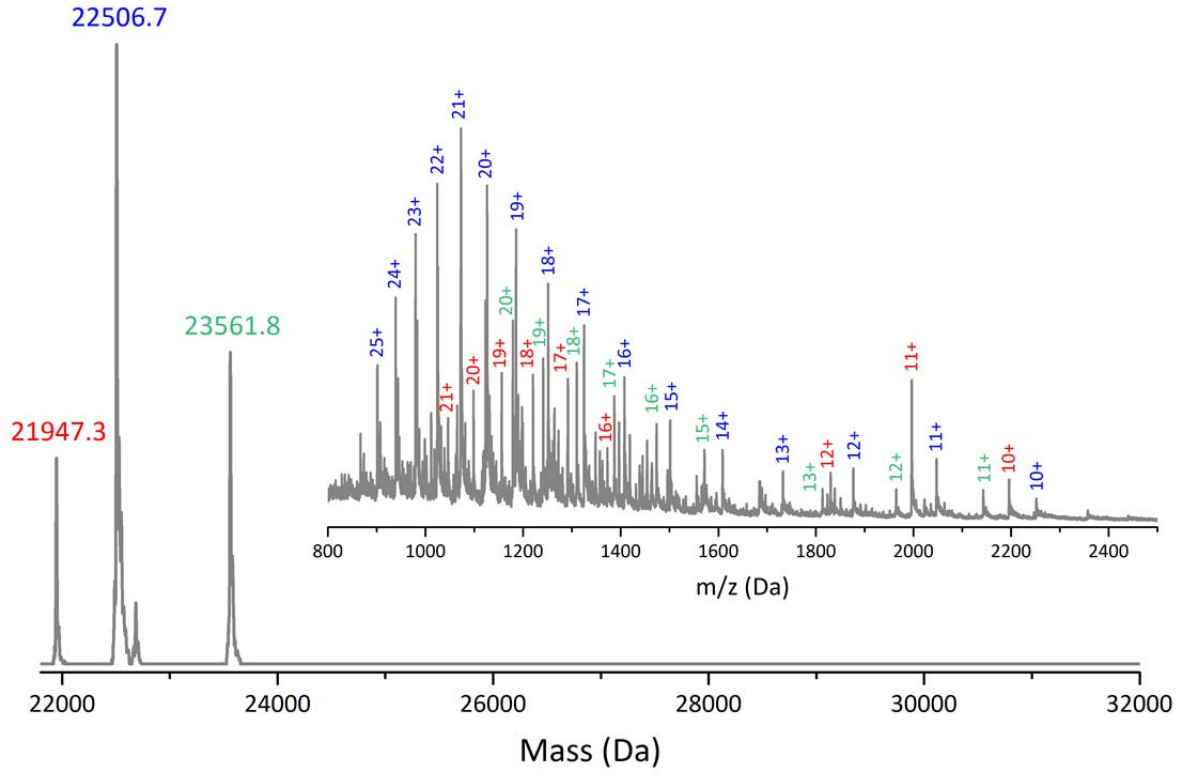


Figure S2 - SDS-PAGE of CCMV with FITC-Anl-Ala-Leu-Pro-Glu-Thr-Gly-NH_2 and 0 eq. SrtA. Gels were visualized both by Coomassie blue staining (left) and by in-gel fluorescence (right). A: Modification of G-CCMV. B: Modification of G_3-CCMV. C: Modification of G-ELP-CCMV. D: Modification of G_3-ELP-CCMV.

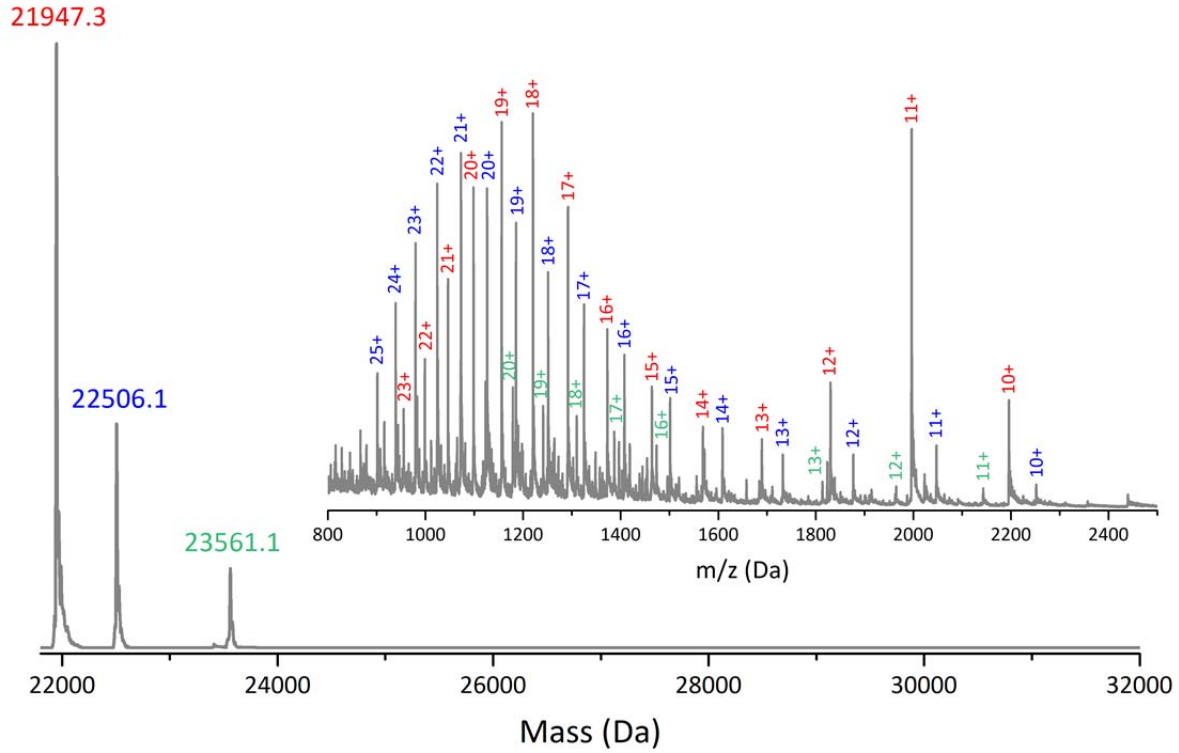
A


22506.3
26+
25+
24+
23+
22+
21+
20+
19+
18+
17+
16+
15+
14+
13+
12+
11+
10+
800
1000
1200
1400
1600
1800
2000
2200
2400
m/z (Da)
22000
24000
26000
28000
30000
32000
Mass (Da)

B


22506.7
23561.8
21947.3
800
1000
1200
1400
1600
1800
2000
2200
2400
m/z (Da)
22000
24000
26000
28000
30000
32000
Mass (Da)

C



D

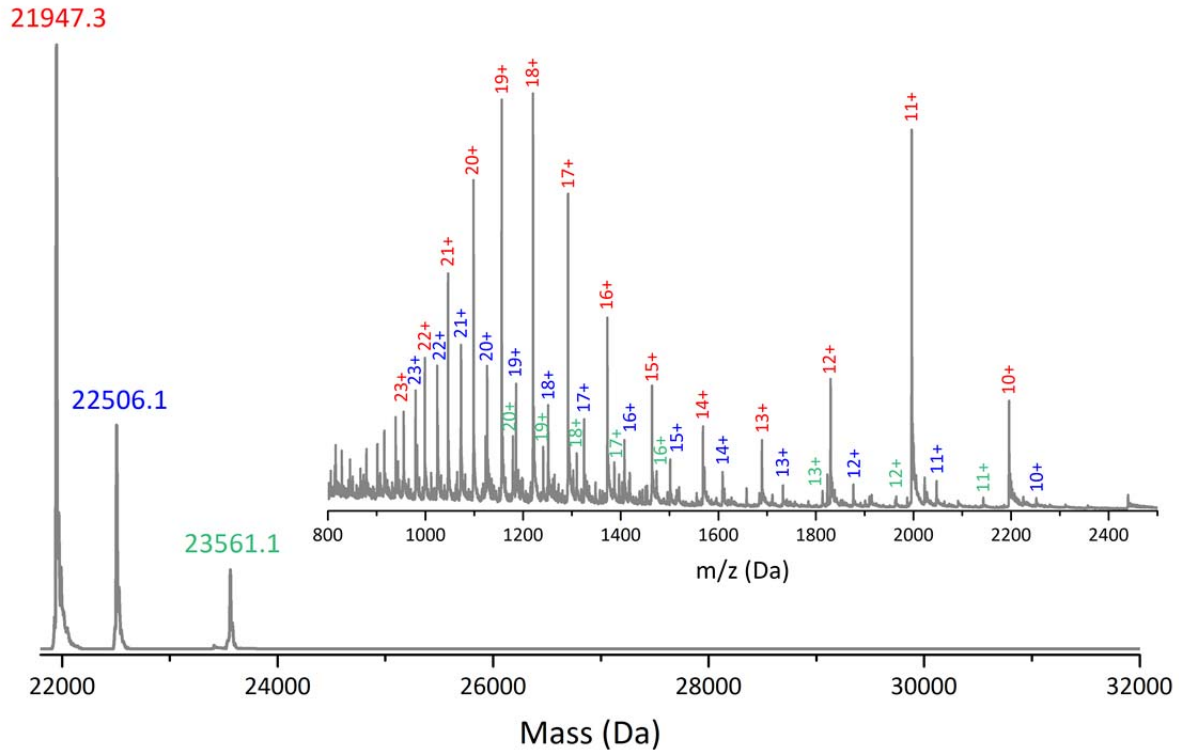
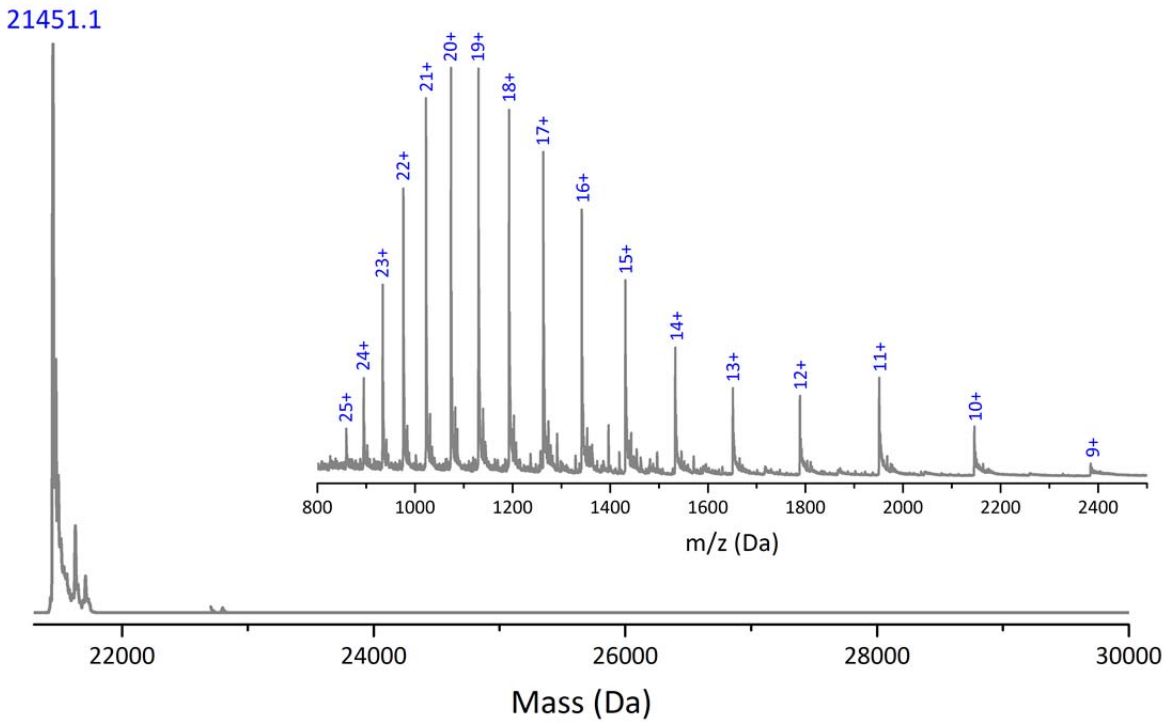
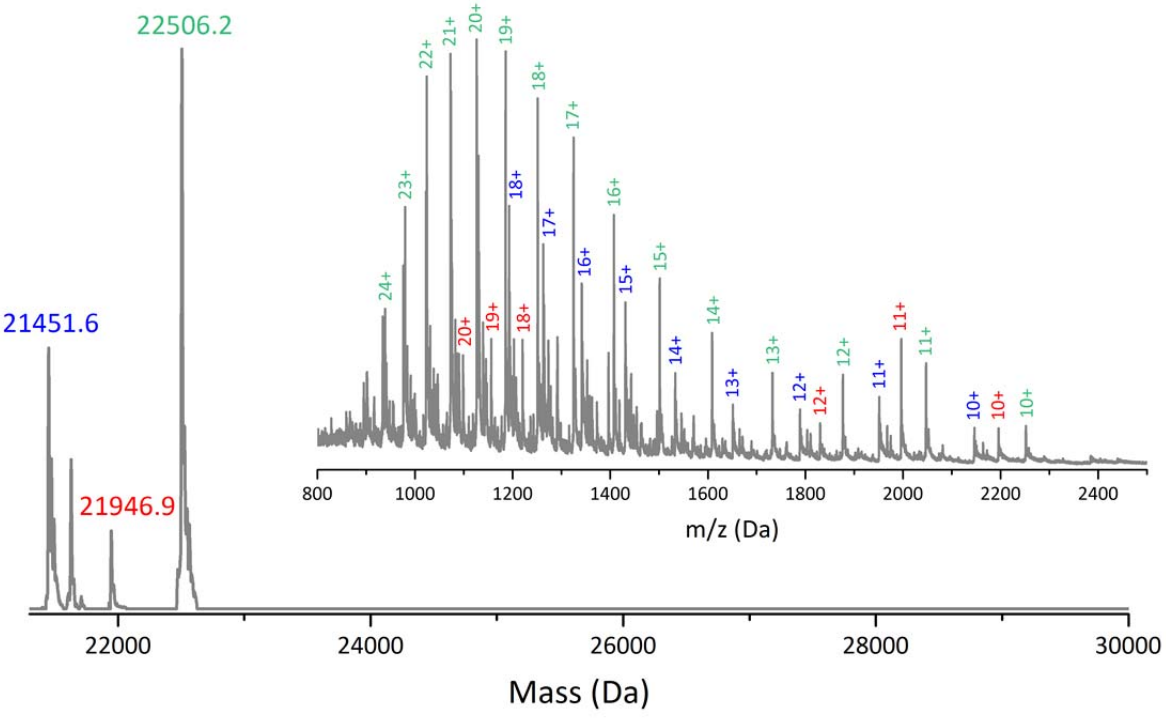


Figure S3 - ESI-TOF mass spectrometry of SrtA-catalyzed modifications of G-CCMV with FITC-Anl-Ala-Leu-Pro-Glu-Thr-Gly-NH_2. Deconvoluted total mass spectra and multiply charged ion series (insets). The expected molecular weights are 21947.5 Da (SrtA, red), 22506.6 Da (G-CCMV, blue), 23561.7 (FITC-AnlALPETG-CCMV, green). A: 0 eq. SrtA. B: 0.2 eq. SrtA. C: 1.0 eq. SrtA. D: 2.0 eq. SrtA.

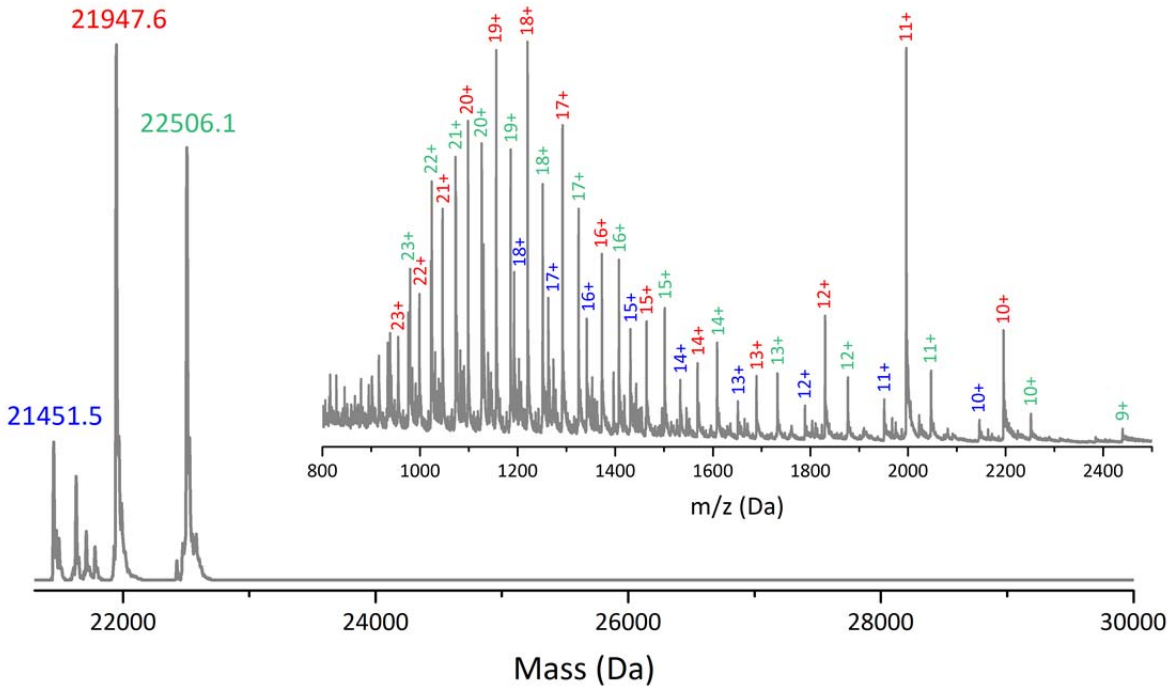
A


21451.1
25+
24+
23+
22+
21+
20+
19+
18+
17+
16+
15+
14+
13+
12+
11+
10+
9+
800
1000
1200
1400
1600
1800
2000
2200
2400
m/z (Da)
22000
24000
26000
28000
30000
Mass (Da)

B


22506.2
21451.6
21946.9
800
1000
1200
1400
1600
1800
2000
2200
2400
m/z (Da)
22000
24000
26000
28000
30000
Mass (Da)

C



D

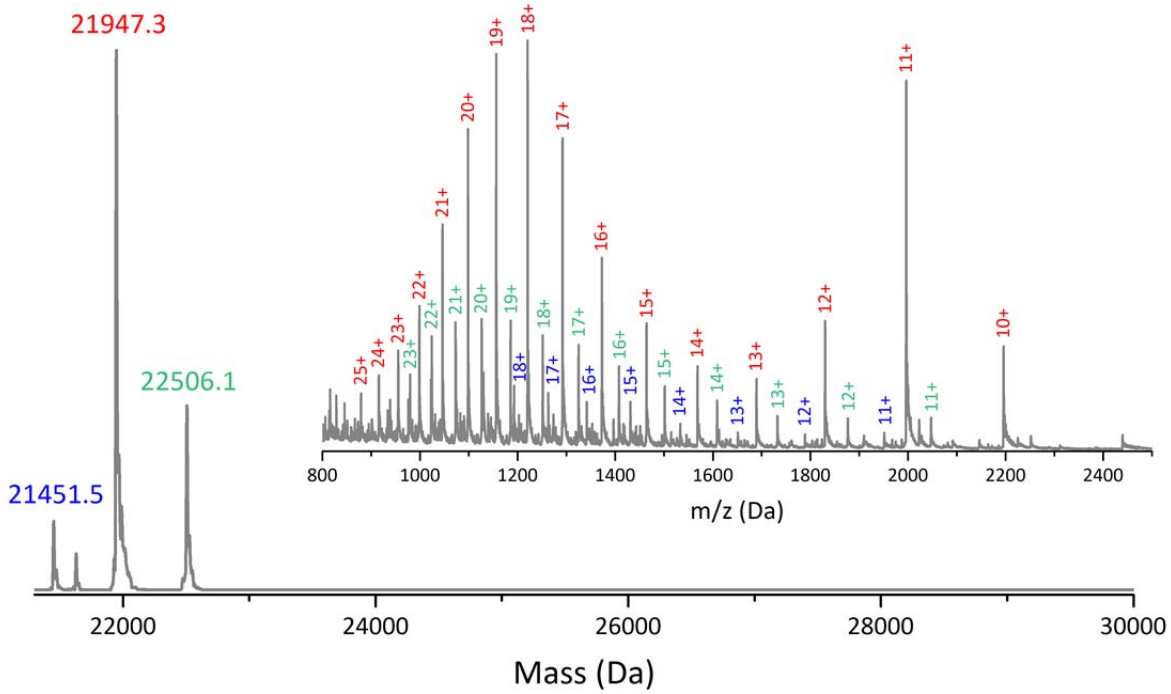
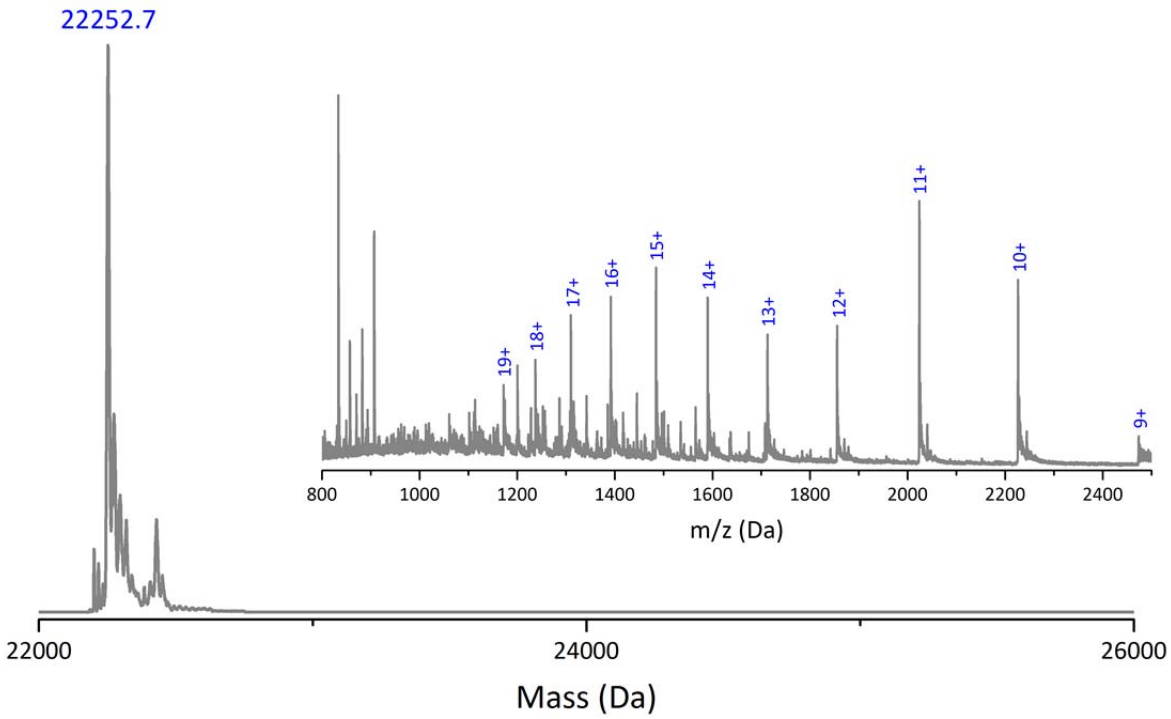
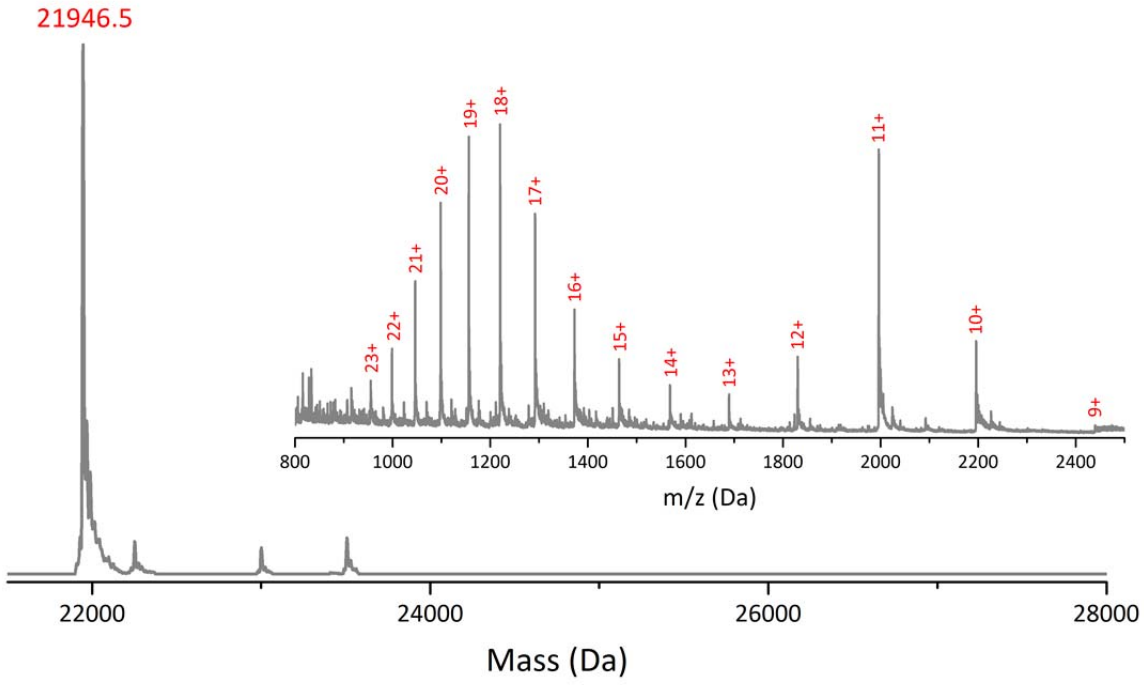


Figure S4 - ESI-TOF mass spectrometry of SrtA-catalyzed modifications of G_3-CCMV with FITC-Anl-Ala-Leu-Pro-Glu-Thr-Gly-NH_2. Deconvoluted total mass spectra and multiply charged ion series (insets). The expected molecular weights are 21947.5 Da (SrtA, red), 21451.4 Da (G_3-CCMV, blue), 22506.6 (FITC-AnlALPETG$_3$-CCMV, green). A: 0 eq. SrtA. B: 0.2 eq. SrtA. C: 1.0 eq. SrtA. D: 2.0 eq. SrtA.

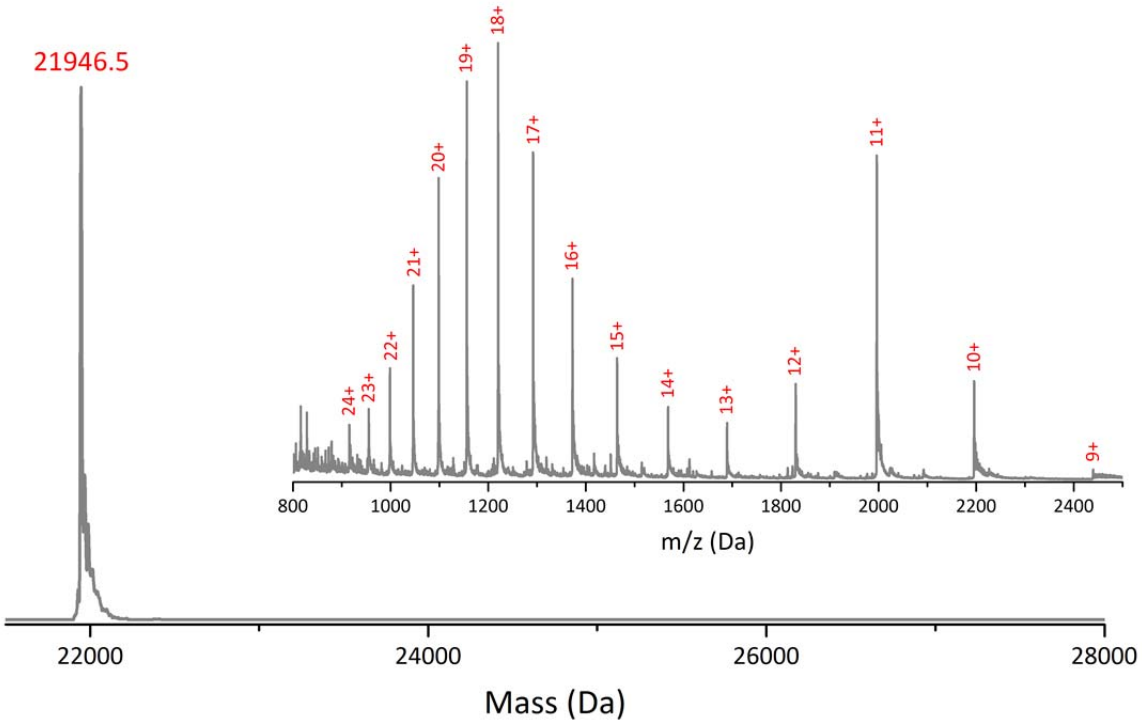
A


22252.7
19+
18+
17+
16+
15+
14+
13+
12+
11+
10+
9+
800
1000
1200
1400
1600
1800
2000
2200
2400
m/z (Da)
22000
24000
26000
Mass (Da)

B


21946.5
23+
22+
21+
20+
19+
18+
17+
16+
15+
14+
13+
12+
11+
10+
9+
800
1000
1200
1400
1600
1800
2000
2200
2400
m/z (Da)
22000
24000
26000
28000
Mass (Da)

C



D

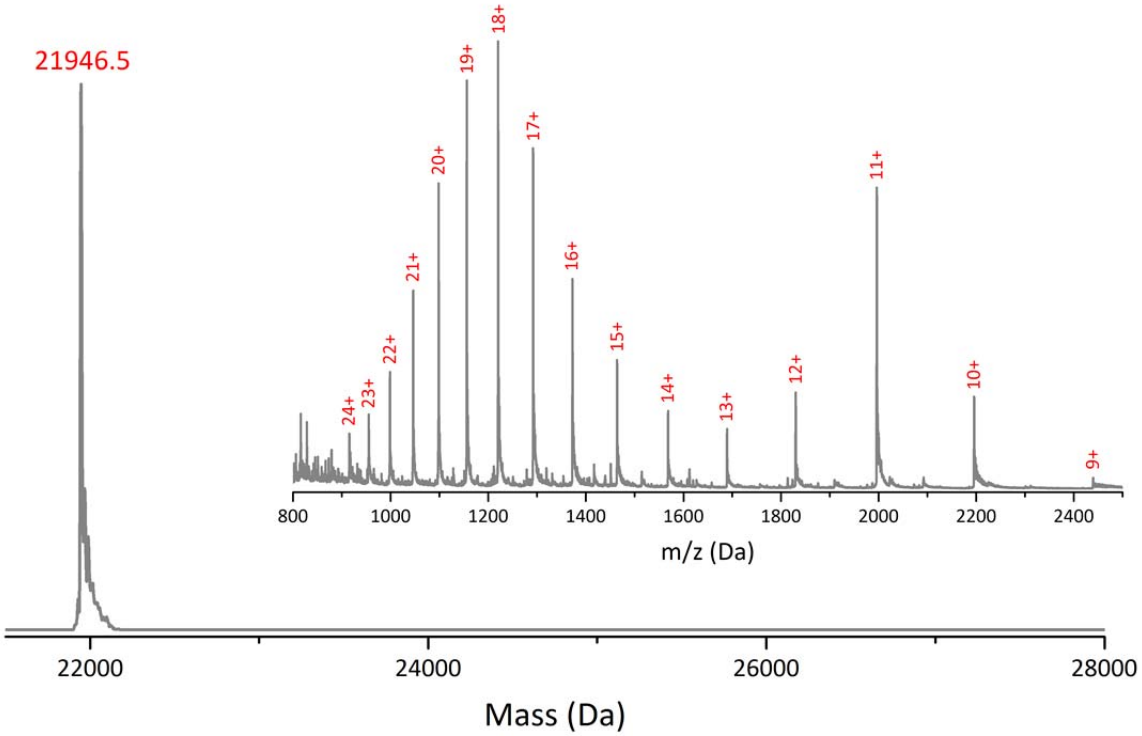
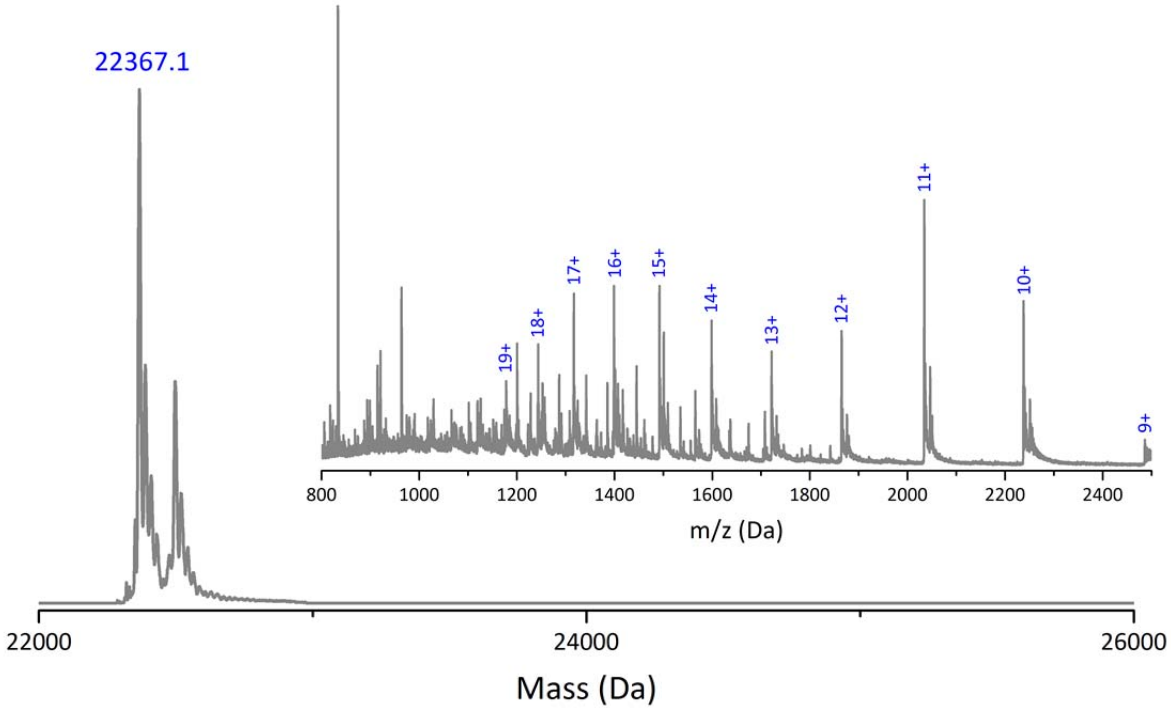
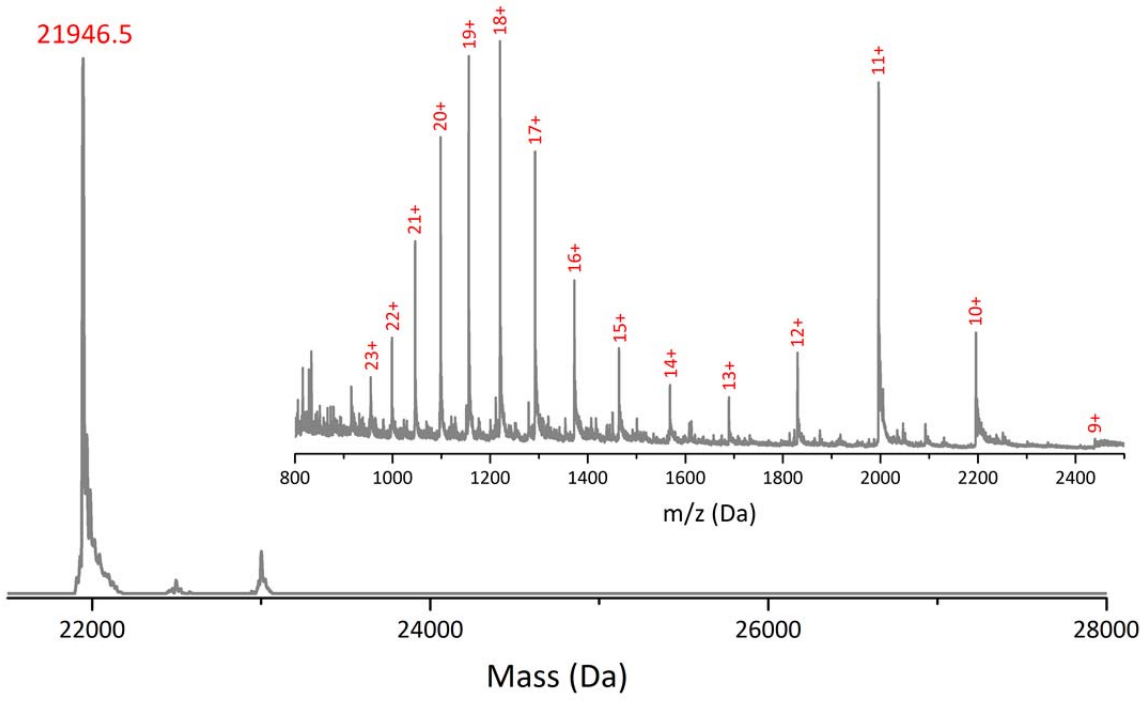


Figure S5 - ESI-TOF mass spectrometry of SrtA-catalyzed modifications of G-ELP-CCMV with FITC-Anl-Ala-Leu-Pro-Glu-Thr-Gly-NH_2. Deconvoluted total mass spectra and multiply charged ion series (insets). The expected molecular weights are 21947.5 Da (SrtA, red), 22253.4 Da (G-ELP-CCMV, blue), 23308.6 (FITC-AnlALPETG-ELP-CCMV, green). A: 0 eq. SrtA. B: 0.2 eq. SrtA. C: 1.0 eq. SrtA. D: 2.0 eq. SrtA.

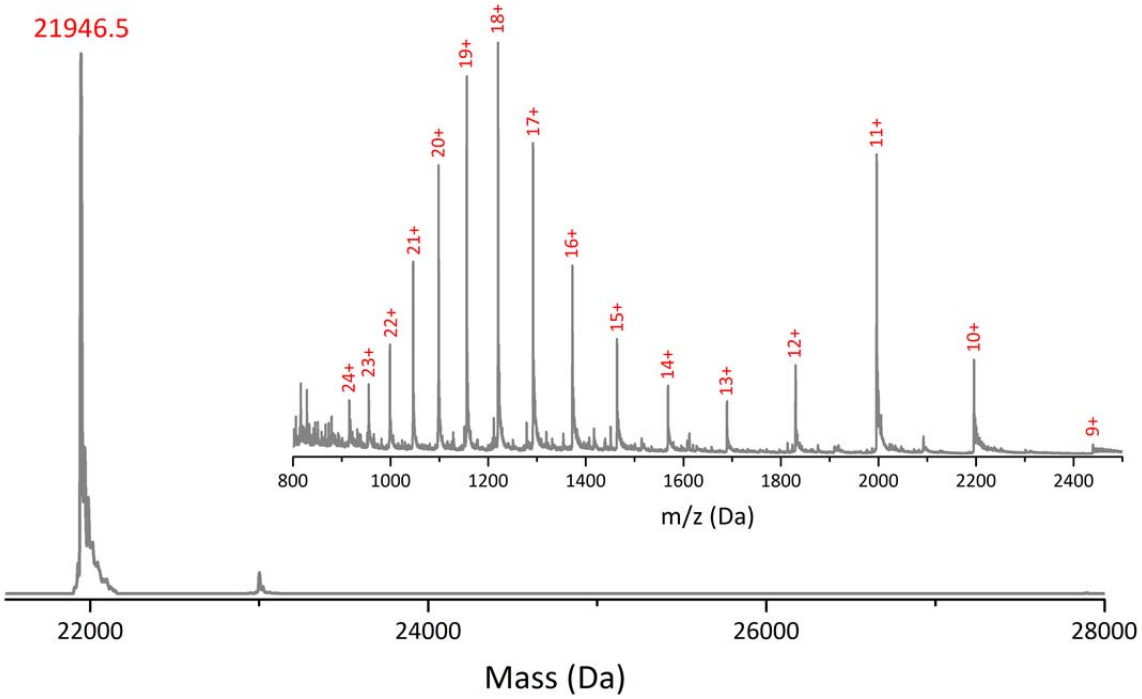
A


22367.1
19+
18+
17+
16+
15+
14+
13+
12+
11+
10+
9+
800
1000
1200
1400
1600
1800
2000
2200
2400
m/z (Da)
22000
24000
26000
Mass (Da)

B


21946.5
23+
22+
21+
20+
19+
18+
17+
16+
15+
14+
13+
12+
11+
10+
9+
800
1000
1200
1400
1600
1800
2000
2200
2400
m/z (Da)
22000
24000
26000
28000
Mass (Da)

C



D

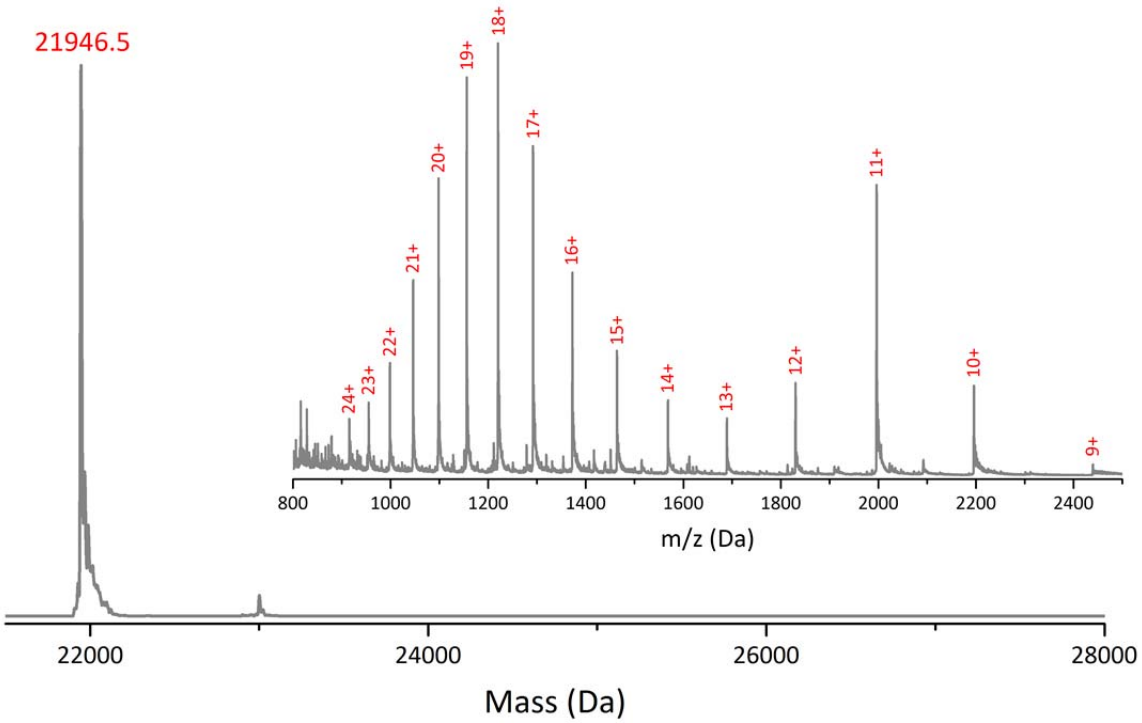
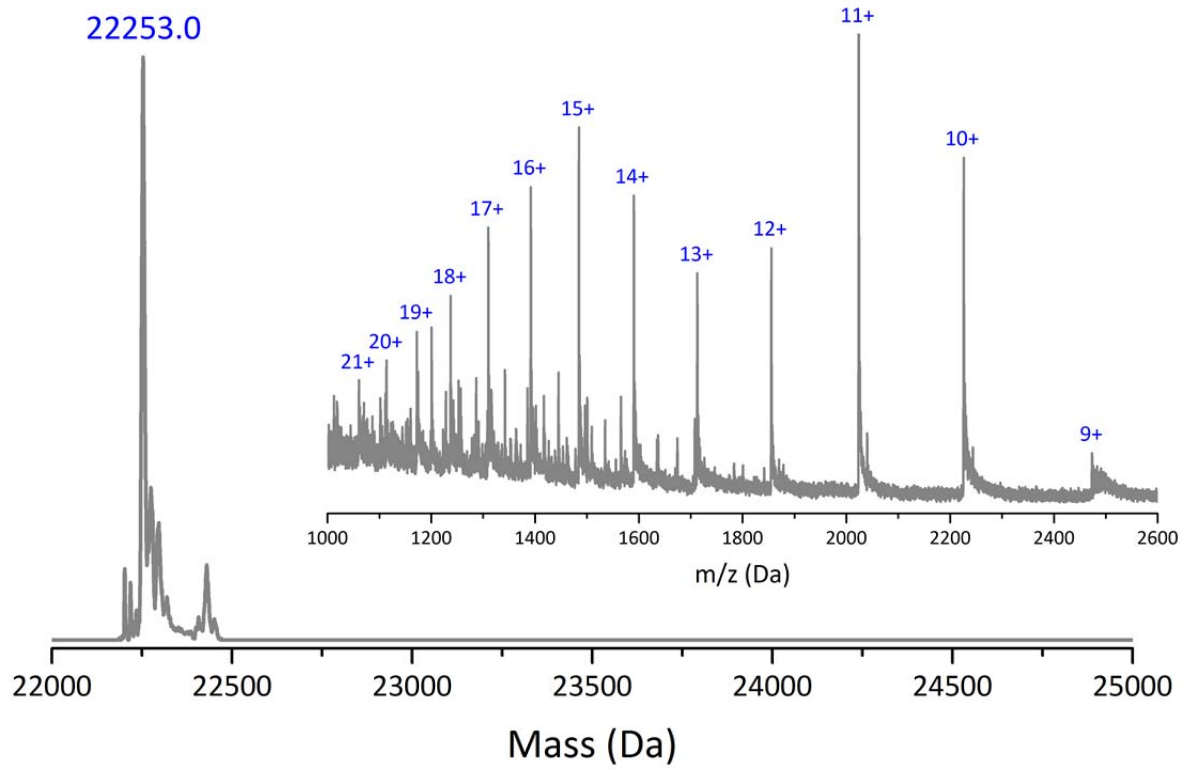
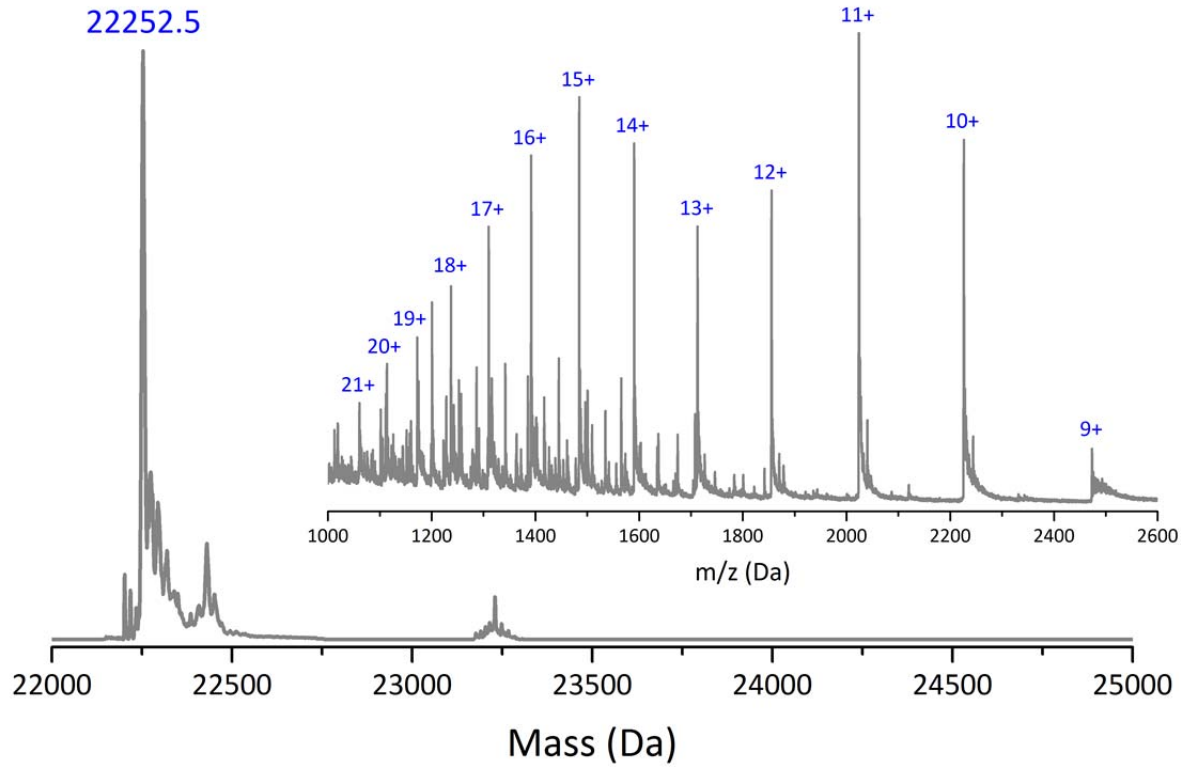


Figure S6 - ESI-TOF mass spectrometry of SrtA-catalyzed modifications of G_3-ELP-CCMV with FITC-Anl-Ala-Leu-Pro-Glu-Thr-Gly-NH_2. Deconvoluted total mass spectra and multiply charged ion series (insets). The expected molecular weights are 21947.5 Da (SrtA, red), 22367.6 Da (G_3-ELP-CCMV, blue), 23422.7 (FITC-AnlALPETG$_3$-ELP-CCMV, green). A: 0 eq. SrtA. B: 0.2 eq. SrtA. C: 1.0 eq. SrtA. D: 2.0 eq. SrtA.

A


22253.0
11+
15+
10+
16+
14+
17+
12+
13+
18+
19+
20+
21+
9+
1000
1200
1400
1600
1800
2000
2200
2400
2600
m/z (Da)
22000
22500
23000
23500
24000
24500
25000
Mass (Da)

B


22252.5
11+
15+
10+
16+
14+
12+
17+
13+
18+
19+
20+
21+
9+
1000
1200
1400
1600
1800
2000
2200
2400
2600
m/z (Da)
22000
22500
23000
23500
24000
24500
25000
Mass (Da)

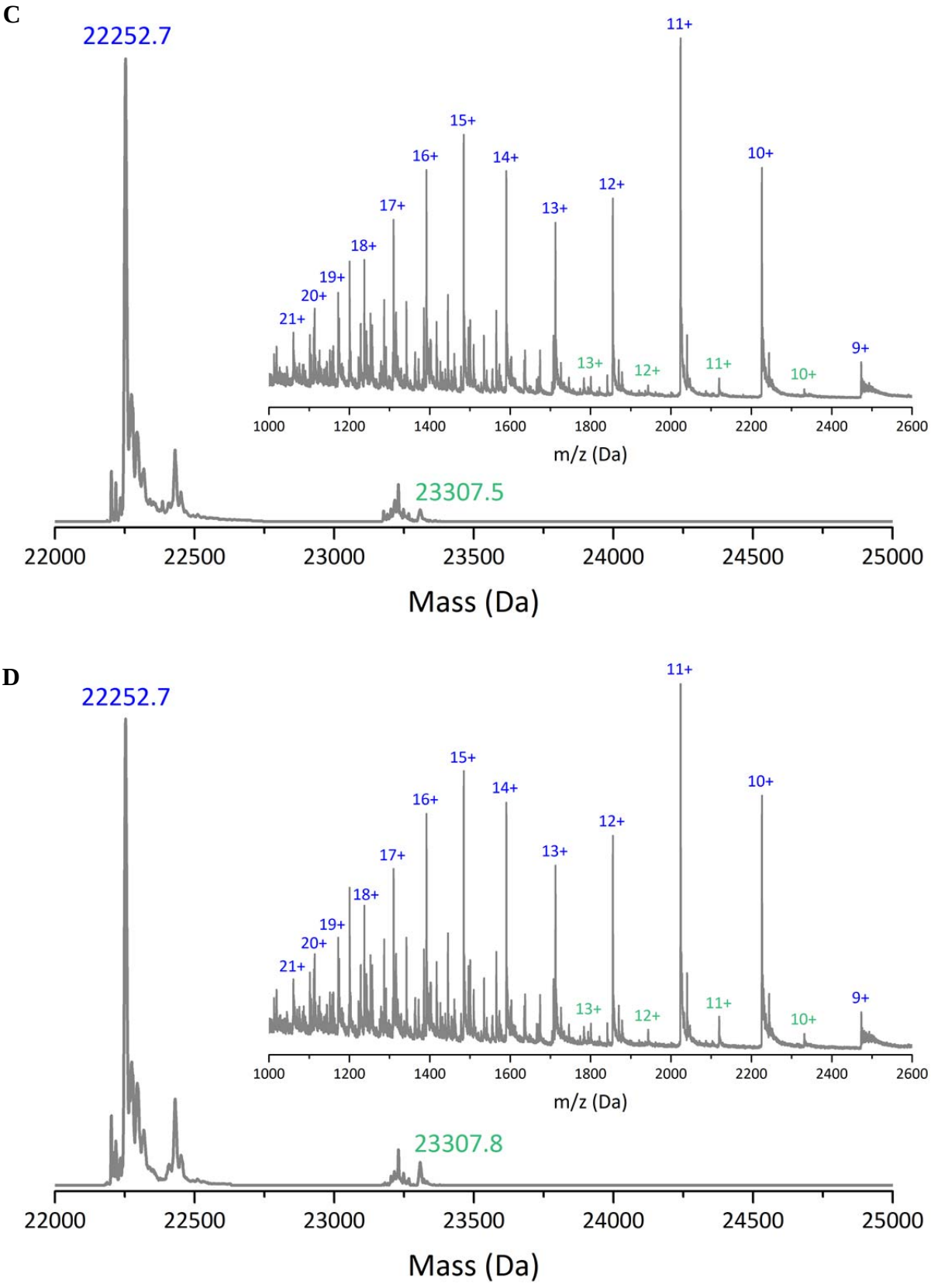


Figure S7 - ESI-TOF mass spectrometry of SrtA-catalyzed modifications of G-ELP-CCMV with FITC-Anl-Ala-Leu-Pro-Glu-Thr-Gly-NH_2 after capsid purification. Deconvoluted total mass spectra and multiply charged ion series (insets). The expected molecular weights are 22253.4 Da (G-ELP-CCMV, blue) and 23308.6 (FITC-AnlALPETG-ELP-CCMV, green). A: 0 eq. SrtA. B: 0.1 eq. SrtA. C: 0.2 eq. SrtA. D: 0.5 eq. SrtA.

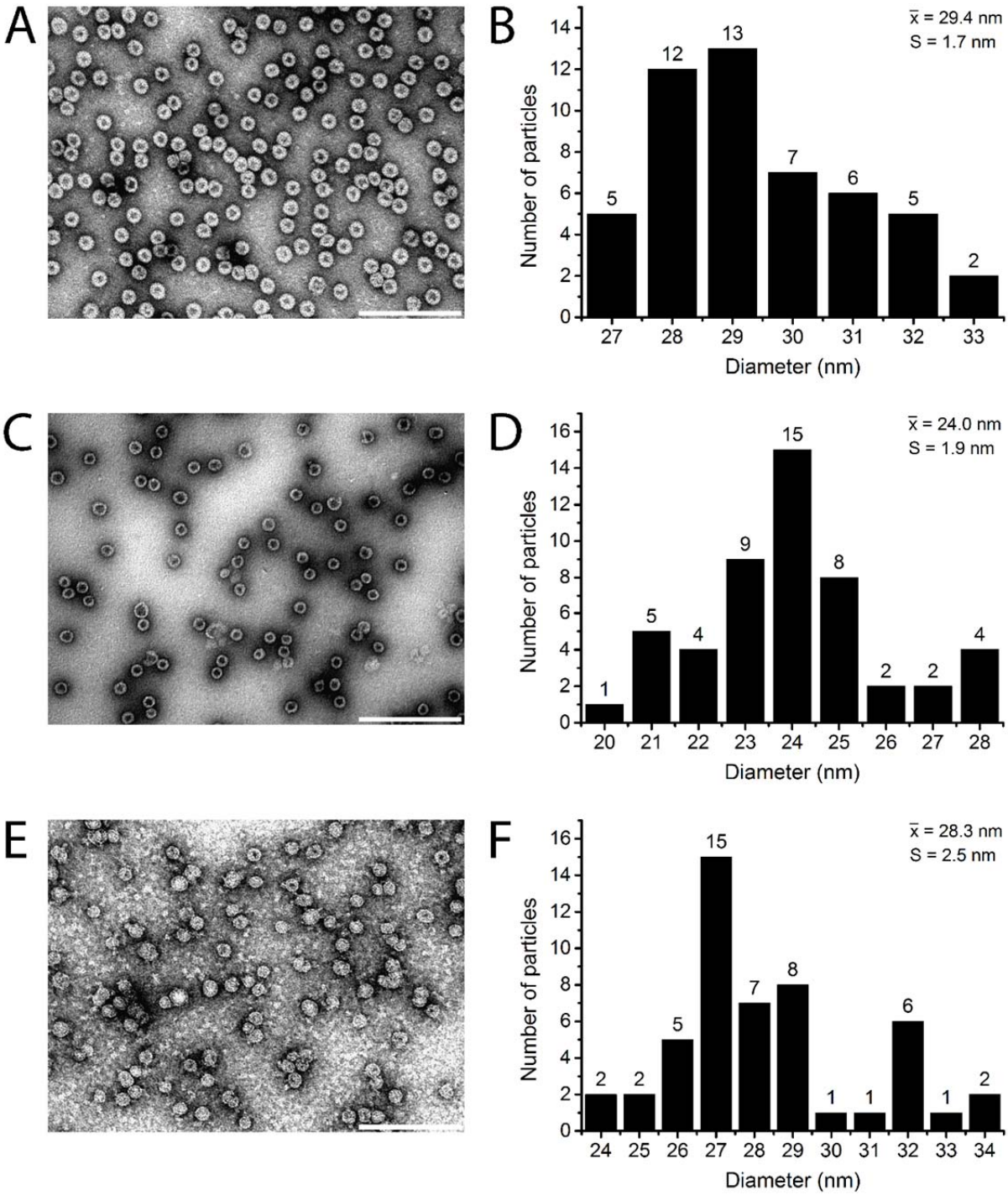


Figure S8 - A) Uranyl acetate-stained TEM micrograph of G-ELP-CCMV **1c**. Scale bar corresponds to 200 nm. B) Size distribution of the G-ELP-CCMV particles shown in A. C) Uranyl acetate-stained TEM micrograph of G-ELP-CCMV **1c** after reaction with FITC-Anl-Ala-Leu-Pro-Glu-Thr-Gly-NH_2 (0 eq. SrtA). Scale bar corresponds to 200 nm. D) Size distribution of the FITC-modified G-ELP-CCMV particles shown in C. E) Uranyl acetate-stained TEM micrograph of G_3-CCMV **1b** after reaction with FITC-Anl-Ala-Leu-Pro-Glu-Thr-Gly-NH_2 (2.0 eq. SrtA). Scale bar corresponds to 200 nm. F) Size distribution of the FITC-modified G_3-CCMV particles shown in E.

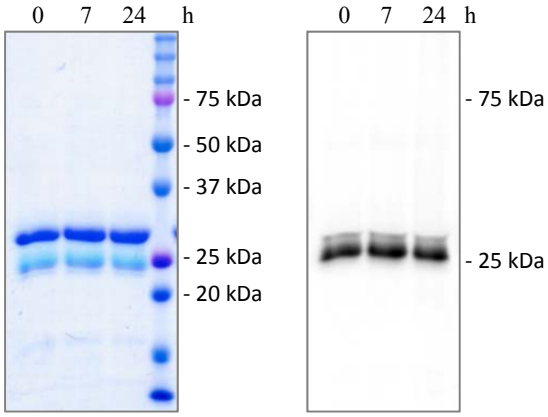
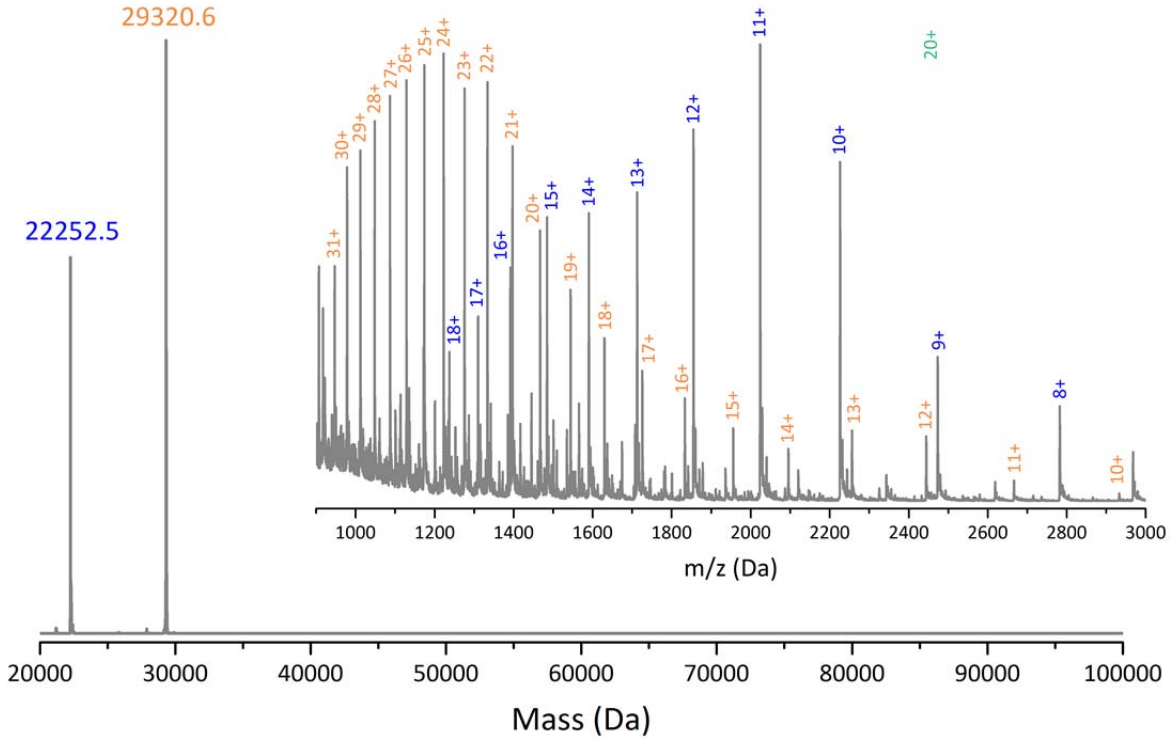
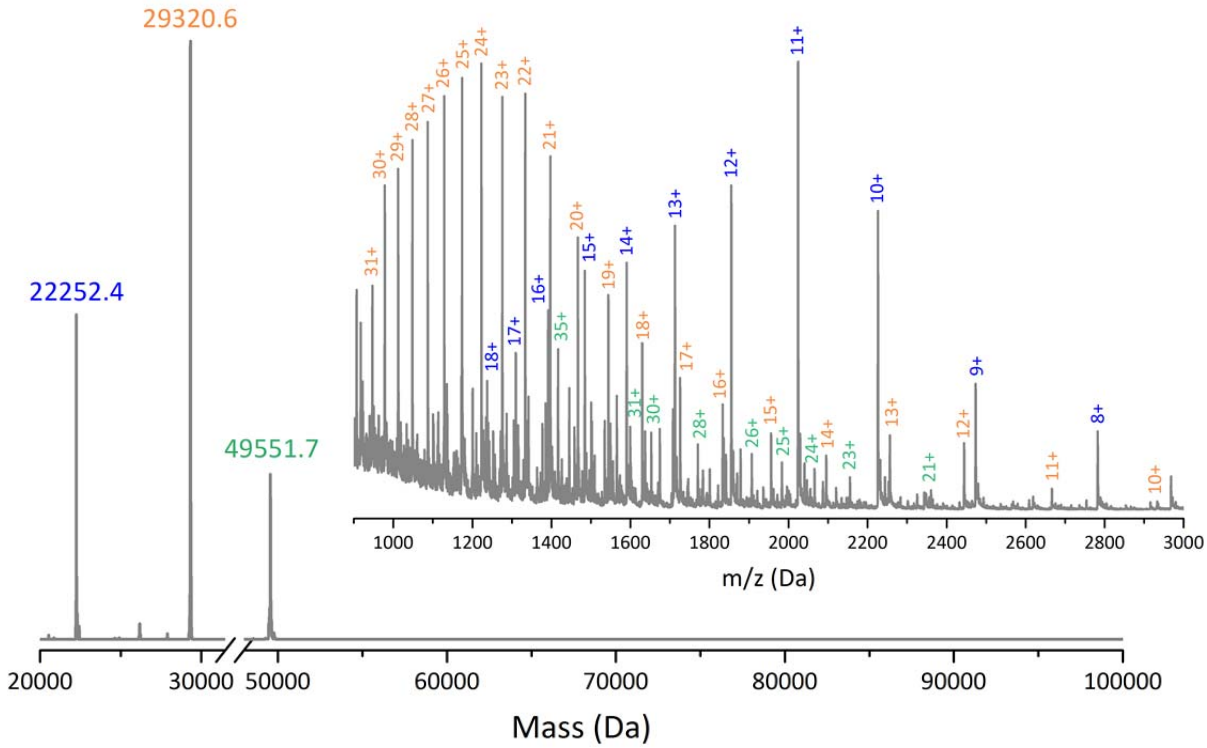


Figure S9 - SDS-PAGE of CCMV with GFP-LPETG and 0 eq. SrtA. The gel was visualized both by Coomassie blue staining (left) and by in-gel fluorescence (right).

A


29320.6
22252.5
20000
30000
40000
50000
60000
70000
80000
90000
100000
Mass (Da)
1000
1200
1400
1600
1800
2000
2200
2400
2600
2800
3000
m/z (Da)

B


29320.6
22252.4
49551.7
20000
30000
50000
60000
70000
80000
90000
100000
Mass (Da)
1000
1200
1400
1600
1800
2000
2200
2400
2600
2800
3000
m/z (Da)

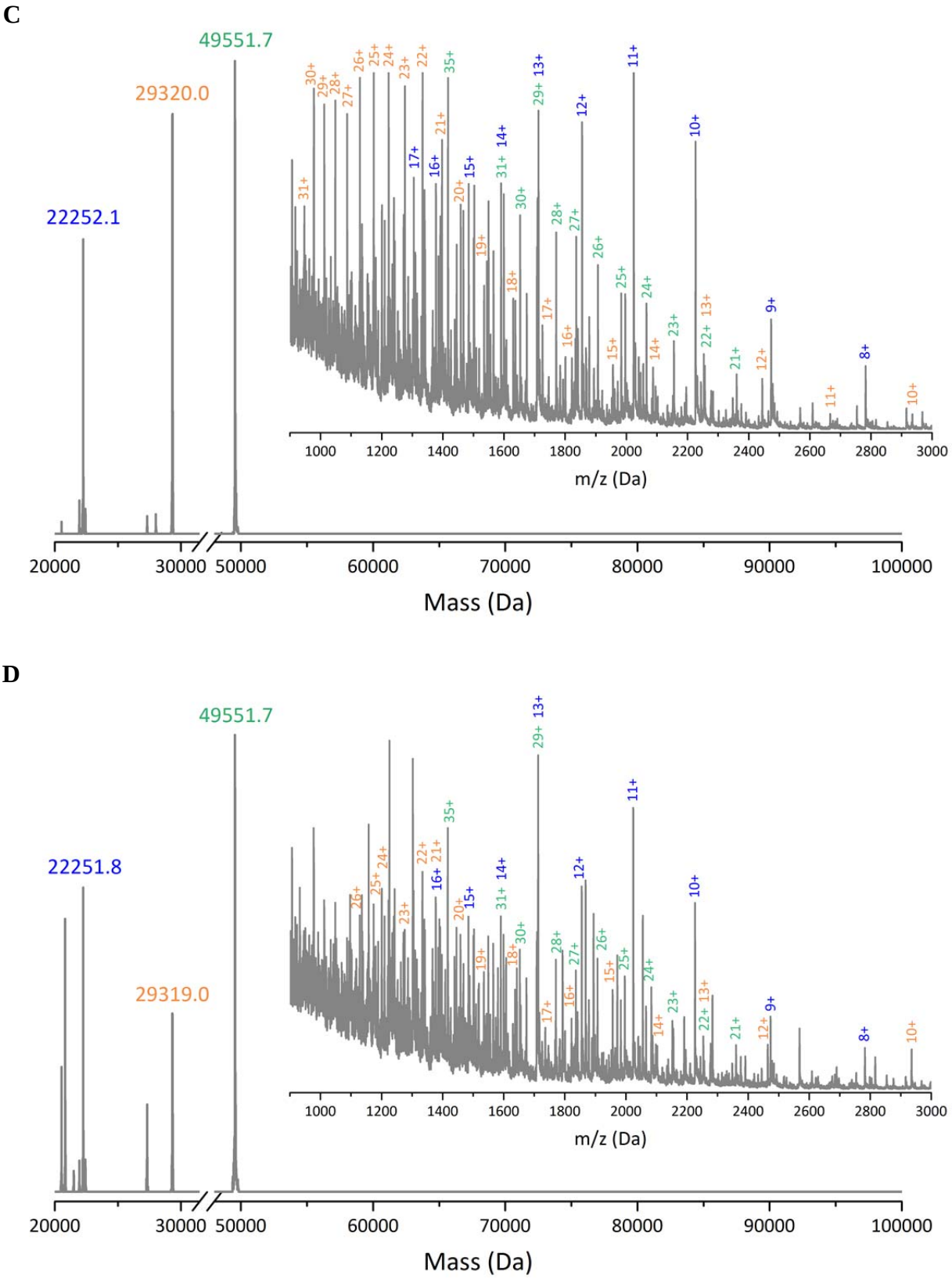


Figure S10 - ESI-TOF mass spectrometry of SrtA-catalyzed modifications of G-ELP-CCMV with GFP-LPETG. Deconvoluted total mass spectra and multiply charged ion series (insets). The expected molecular weights are 21947.5 Da (SrtA, red), 22253.4 Da (G-ELP-CCMV, blue), 29321.9 (GFP-LPETG, orange) 49553.2 (GFP-LPETG-ELP-CCMV, green). A: 0 eq. SrtA. B: 0.2 eq. SrtA. C: 1.0 eq. SrtA. D: 2.0 eq. SrtA.

4 References

[1] Gill, S. C., and von Hippel, P. H. (1989) Calculation of protein extinction coefficients from amino acid sequence data. *Anal. Biochem. 182*, 319-26.

[2] Ton-That, H., Liu, G., Mazmanian, S. K., Faull, K. F., and Schneewind, O. (1999) Purification and characterization of sortase, the transpeptidase that cleaves surface proteins of Staphylococcus aureus at the LPXTG motif. *Proc. Natl. Acad. Sci. U. S. A. 96*, 12424-9.

[3] Yang, R., Bi, X., Li, F., Cao, Y., Liu, C.-F. (2014) Native chemical ubiquitination using a genetically incorporated azidonorleucine. *Chem. Commun. 50*, 7971-4.

[4] Minten, I. J., Hendriks, L. J., Nolte, R. J., and Cornelissen, J. J. (2009) Controlled encapsulation of multiple proteins in virus capsids, *J. Am. Chem. Soc. 131*, 17771-3.

[5] van Eldijk, M. B., Wang, J. C.-Y., Minten, I. J., Li., C., Zlotnick, A., Nolte, R. J. M., Cornelissen, J. J. L. M., and van Hest, J. C. M. (2012) Designing two self-assembly mechanisms into one viral capsid protein, *J. Am. Chem. Soc. 134*, 18506-9.