Supporting Information for:

Protein Encapsulation *in vivo* via Transient Disulfide Bond Formation

Xueji Zhang,[a] Jiameng Zong,[a, b] and Kenneth J. Woycechowsky [*, a, c]

[a]School of Pharmaceutical Science and Technology, Tianjin University, 92 Weijin Road, Nankai District, Tianjin, 300072, China
[b]Present Address: 210 Washtenaw Avenue, Life Sciences Institute, University of Michigan, Ann Arbor, MI 48109-2216, USA
[c]Present address: School of Molecular Sciences, Arizona State University, Tempe, AZ 85281, USA
[*]e-mail: kenneth.woycechowsky@asu.edu

SUPPORTING MATERIALS AND METHODS

Materials

E. coli strain BL21(DE3) was purchased from TransGen Biotech Co., Ltd (Beijing, China). *E. coli* strain DH5α, 5X HiFidelity PCR buffer, Fast HiFidelity Polymerase and Plasmid Minipreparation Kit were purchased from TIANGEN Biotech Co., Ltd (Beijing, China). LB broth was from Sangon Biotech Co., Ltd (Shanghai, China). LB agarose powder, ampicillin sodium salt (≥85% purity), lysozyme (ultrapure grade), DNase I (1500 U/mg), RNase A (>50 U/mg), Tris-HCl (≥99.5% purity), iodoacetamide, Tween-20, and Coomassie Brilliant Blue R-250 (ultrapure grade), and chloramphenicol (≥99% purity) were from Solarbio Life Sciences (Beijing, China). Isopropyl-β-D-thiogalactoside (IPTG) (≥99% purity) and sodium dodecyl sulfate (>93% purity) were from Genview Scientific Inc (USA). $MgCl_2$ (≥99.0% purity) and $(NH_4)S_2O_8$ (99.9% purity) were from Aladdin Bio-Chem Technology Co., Ltd. (Shanghai, China). $(NH4)_2SO_4$ (99.0% purity), Na_2HPO_4 (≥99.0% purity), and NaCl were from Kermel Chemical Reagents Co., Ltd (Tianjin China). NaH_2PO_4 (≥99.0% purity) and $EDTA(Na)_2$ (≥99.0% purity) were from Yuanli Chemical Co., Ltd (Tianjin, China). Methanol was from Jiangtian Chemical Co., Ltd (Tianjin, China). Glycine (98% purity) and Tris base (99% purity) were from HEOWNS Biochemical Technology Co., Ltd (Tianjin, China). Glycerol was from Dingguo (Tianjin, China). N,N,N',N'-Tetramethylethylenediamine (TEMED) and imidazole (99% purity) were from Meryer Biochemical Technology Co., Ltd (Shanghai, China). Protein markers, 30% acrylamide mix, and 96-well black plates were from BioSharp (Tianjin, China). 96-well transparent plates were from Greiner Bio-One (Germany). PVDF membrane (Pore size: 0.45 μm) was from Merck Millipore (USA). Mouse polyclonal anti-GFP antibodies and goat anti-mouse IgG antibodies conjugated to horseradish peroxidase were from Immunoway Biotechnology Company (Tianjin, China). Restriction enzymes DpnI, SpeI, and XhoI were from New England BioLabs (USA).

Plasmid construction

A two-stage process was carried out to create a plasmid encoding GFP fused with the last 12 amino acids of AaRS. In the first stage, oligonucleotides HX3 and HX4 (Table S2), which encode the last 11 amino acids of AaRS, were annealed to each other and then inserted into the plasmid pACYC-GFP-BsRS11 in between the XhoI and SpeI restriction sites to form pACYC-GFP-AaRS11, similar to the previously described cut-and-paste ligation strategy.[1] In the second

stage, pACYC-GFP-AaRS11 was used as a template for site-directed insertion mutagenesis to generate pACYC-GFP-AaRS12 by annealing primers RS12F and RS12R (Table S2). A codon for Lys was inserted at the beginning of the AaRS-derived sequence to add the first amino acid of the reported AaRS12 encapsulation tag,[2] giving plasmid pACYC-GFP-AaRS12. Briefly, a PCR reaction in a total volume of 50 µL was set up by mixing pACYC-GFP-AaRS11 (1 ng), RS12F and RS12R (0.3 µM each), dNTPs (0.2 mM), Fast HiFideltity DNA polymerase (2 U), and HiFidelity PCR buffer (1X final concentration) together. The mixture was put into a thermocycler and subjected to the following PCR program: initial polymerase activation at 95 ℃ for 2 min; 18 cycles of denaturation at 94 ℃ for 20 s, annealing at 60 ℃ for 10 s, and primer extension at 68 ℃ for 2.5 min; and a final incubation at 68 ℃ for 5 min. DpnI (4 U) was then added to the product mixture to digest the template plasmid at 37 ℃ for 1 h. The transformation of the product DNA into *E. coli* DH5α cells and plasmid minipreparation were done to obtain a working stock of pACYC-GFP-AaRS12, whose sequence was confirmed by DNA sequencing (Tsingke Biotechnology Co., Ltd., Beijing, China).

To create GFP-M244C, primers HCN157 and HCN158 (Table S2) were annealed to the template plasmid pACYC-GFP-R10[3] to generate the Met244Cys mutation and mutate the Ala245 codon to a stop codon. The mutagenesis reaction was carried out similarly to that described above for second stage of the construction of pACYC-GFP-AaRS12. The resulting plasmid, pACYC-GFP-M244C, was confirmed by DNA sequencing.

The plasmid pACYC-GFP0-C49S was generated *via* site-directed mutagenesis using pACYC-GFP0 as the template. PCR amplification was performed using the mutagenic primers C49S-F and C49S-R (Table S2). Following amplification, the template DNA was digested with DpnI, and the resulting PCR products were transformed into competent *E. coli* DH5*α* cells. The resulting plasmid was confirmed by DNA sequencing.

Protein production and purification

The production of individual proteins (either AaLS-IC, wt-AaLS, GFP0, GFP-AaRS12, or GFP-M244C) was performed using the corresponding plasmid, as previously described.[1] The co-production of an AaLS variant (either wt-AaLS or AaLS-IC) with a GFP variant (either GFP0, GFP-AaRS12, or GFP-M244C) from the appropriate plasmids was also carried out as described.[1]

A single colony of *E. coli* BL21 (DE3) cells, which had been transformed with pMG-AaLS-IC encoding AaLS-IC protein, was picked from a fresh transformation plate into 5 mL LB broth containing 100 μg/mL ampicillin using inoculation ring. The culture was incubated at 37 °C in the constant shaker with the speed of 220 rpm overnight.

The next day, the overnight culture was observed to be turbid. 500 mL fresh sterilized LB broth containing 100 μg/mL was inoculated with 1 mL overnight culture and followed by incubation at 37 °C with the speed of 220 rpm. When O.D.600 of the culture reached a range of 0.6-1.0, 1 mL IPTG stock solution (50 mM) was added into each culture to induce the expression of AaLS-IC. At this time, lower the temperature of the shaker to 30 °C but still with the speed of 220 rpm, overnight. After induction, the cells were harvested by centrifuge with the speed of 8000 g for 10 min at 4 °C. Then, the pellets were transferred to a 50 mL centrifuge tube and stored at -80 °C for several weeks.

E. coli strains transformed with the pACYC-GFP plasmid alone were maintained in LB medium supplemented with 30 μg/mL chloramphenicol. For the co-production of an AaLS variant (either wt-AaLS or AaLS-IC) with a GFP variant (either GFP0, GFP-AaRS12, or GFP-M244C), cells were transformed with both the pMG-based plasmid (encoding AaLS variants) and the pACYC-based plasmid (encoding GFP variants). These co-expression strains were maintained in LB medium supplemented with 100 μg/mL ampicillin and 30 μg/mL chloramphenicol to ensure the retention of both plasmids. The protein expression procedure followed the identical protocol used for AaLS-IC, with the exception of the antibiotic conditions described above.

The purification of GFP variants was carried out as previously described except for the following differences.[1] Here, the anion exchange chromatography step was performed using a HiTrap™ Q HP column. After injection onto the HiTrap™ Q HP column, GFP was eluted using a linear gradient from 0-100% high-salt buffer (50 mM sodium phosphate, 1 M NaCl, pH 8.0) over 200 mL at a flow rate of 2.0 mL/min.

The purification of the AaLS-IC nanocage, either produced individually or co-produced with a GFP variant, was basically the same as previously reported,[4] except that the running buffer for size-exclusion chromatography contained 100 mM sodium phosphate, 1 mM EDTA at pH 8.0.

Since wt-AaLS is fused to a His_6 tag, the purification of either individual wt-AaLS nanocage or protein complex co-produced with a GFP variant was performed based on previous reported

procedures[5] with following differences. The cleared cell lysate was not incubated with additional RNase A and DNase I. The Ni^{2+}-affinity chromatography step was performed using a HiPrep IMAC FF 16/10 column using the conditions and procedure previously described for an AaLS variant.[6] The running buffer for size-exclusion chromatography was the same as for AaLS-IC.

Measurement of protein concentration

Protein concentrations were determined using the Bradford assay with bovine serum albumin as a reference standard, essentially as described.[1] For the individual calibration curve samples and the samples of purified nanocages (either produced individually or with a GFP variant) or purified GFP variants, three parallel measurements were carried out.

Fluorescence Calibration Curves of GFP variants

The relationship between fluorescence emission and protein concentration was determined for the GFP variants used as model guest proteins. Three identical aliquots (200 µL each) of purified GFP solutions (0-1.5 µM) in buffer (100 mM sodium phosphate, 1 mM EDTA, pH 8.0) were added into a black 96-well plate. The average fluorescence intensities were recorded using a Tecan Infinite M200 Pro plate reader (TECAN, Männedorf, Switzerland) with an excitation wavelength of 375 nm and an emission wavelength of 508 nm. Each fluorescence measurement was corrected by subtracting the signal from a buffer-only blank sample. The average corrected fluorescence intensity of each triplicate sample was plotted against the protein concentration and the data were fit to a linear equation with the y-intercept fixed at zero. Generally, the resulting fit gave an r^2 value greater than 0.99 over the entire GFP concentration range. Calibration curves were generated on the same day that fluorescence measurements of purified encapsulation complexes were carried out.

Encapsulation yield measurements

The guest protein content of purified nanocage samples (0.8 mg protein/mL) that had been co-produced with a GFP variant was determined by measuring their GFP-specific fluorescence. Fluorescence intensity values were recorded as described above. The observed fluorescence was corrected by subtracting the signal from a sample of the corresponding nanocage (0.8 mg protein/mL) that was produced in the absence of any GFP variant. Corrected fluorescence intensities were converted to GFP concentration values by comparison with the corresponding calibration curve (see above). The nanocage protein concentration value was assumed to be the

same as the total protein concentration since the mass of nanocage protein is much greater than the mass of GFP in all of the encapsulation complexes. The encapsulation yield (mol GFP/mol nanocage) for each complex was calculated from the GFP concentration and the protein nanocage concentration in each sample. The reported encapsulation yields represent the average of a triplicate-of-triplicates for each complex; three independent batches of each complex were measured three times each.

Fluorescence emission spectra

The fluorescence emission spectra of various protein samples (200 μL in 96-well black plate) were recorded using a Tecan Infinite M200 Pro plate reader with an excitation wavelength of 375 nm. The concentration of each GFP variant was 1 μM. Encapsulation complexes and empty nanocages were measured at a protein concentration of 0.8 mg/mL. All proteins were dissolved in buffer containing 100 mM sodium phosphate and 1 mM EDTA at pH 8.0. Fluorescence emission was recorded from 450 nm to 700 nm in 2 nm increments.

Analytical size-exclusion chromatography

The assembly of encapsulation complexes was analyzed by injecting each sample (2.7-3.7 mg protein in 2 mL of running buffer) onto a HiPrep 16/60 Sephacryl S-400 HR size-exclusion column. The running buffer contained 100 mM sodium phosphate and 1 mM EDTA at pH 8.0. The flow rate was 1.0 mL/min. Fractions (4 mL each) were collected from 0 mL to 150 mL. A_{280} was measured by the in-line UV detector of ÄKTA FPLC system in 1 min time increments and averaged for each fraction. The fluorescence intensity of each fraction was recorded using a Tecan Infinite M200 Pro plate reader by transferring an aliquot (200 μL) into a 96-well black plate. The fluorescence emission wavelength was 508 nm and the excitation wavelength was 375 nm.

SDS- polyacrylamide gel electrophoresis

Denaturing gel electrophoresis was used to analyze protein production in whole cells. For the preparation of whole-cell samples, an aliquot of uninduced cells (800 μL) was removed from the culture. After induction, another aliquot of cells was removed from the culture. The volume of the aliquot from the culture of induced cells was adjusted to give an equal number of cells as the aliquot of uninduced cells, based on the $O.D._{600\ nm}$ values of the uninduced and induced cultures. The aliquots of cells were spun in a centrifuge at 12,000 rpm for 2 min at room temperature. The supernatant was discarded, and the pellet was resuspended in 1x phosphate buffered saline (PBS)

with a final volume of 20 µL. 5 µL of 5X SDS-loading buffer (50 mM Tris-HCl, 100 mM DTT, 2% SDS, 0.1 % bromophenol blue, 10 % glycerol, pH 6.8) was mixed with the resuspended cells. Subsequently, the samples were denatured at 99 ℃ for 20 min and then run on a 15% acrylamide gel for 70 min at 200 V. Proteins were visualized by staining the gel with Coomassie Blue R-250.

For the analysis of purified proteins, samples were denatured under reducing (+DTT) and non-reducing (-DTT) conditions. AaLS-IC (30 µg), AaLS-IC:GFP0 (60 µg), AaLS-IC:GFP-AaRS12 (60 µg), and AaLS-IC:GFP-M244C (60 µg) were each dissolved in a final volume of 20 µL of buffer containing 100 mM sodium phosphate and 1 mM EDTA at pH 8.0. Reduced protein samples were prepared by mixing with 5 uL of 5X SDS-loading buffer containing 100 mM DTT (as above), followed by incubation at 99 ℃ for 10 min. Non-reduced protein samples were prepared by adding 5 uL of 5X SDS-loading buffer lacking DTT, followed by incubation at 99 ℃ for 3 min. Protein separation was performed on a 15% polyacrylamide resolving gel at 200 V for 70 min. Proteins were visualized by staining the gel with Coomassie Blue R-250.

Western Blot

Western blotting was performed to analyze the GFP present in the AaLS-IC:GFP0 and AaLS-IC:GFP-AaRS12 encapsulation complexes. Prior to sample denaturation, iodoacetamide (IAA) was added at a final concentration of 250 mM to samples of AaLS-IC, AaLS-IC:GFP0, and AaLS-IC:GFP-AaRS12 to alkylate the free thiol groups in the proteins. The thiol alkylation reaction mixtures were incubated at room temperature for 3 h. IAA-treated AaLS-IC (20 µg), AaLS-IC:GFP0 (40 µg), and AaLS-IC:GFP-AaS12 (40 µg) were denatured under non-reducing and reducing conditions (as described above) and separated by SDS-PAGE on a 12% polyacrylamide resolving gel at 200V for 60 min. The proteins were then transferred to a PVDF membrane. Following the transfer step, the PVDF membrane was incubated in blocking solution (containing 5% non-fat milk powder), washed with TBST buffer, incubated with primary antibody (mouse polyclonal anti-GFP antibodies), washed again, incubated with secondary antibody (goat anti-mouse antibody conjugated to horseradish peroxidase), and washed again. The Western blot was developed using chemiluminescence, as previously described.[1]

SUPPORTING FIGURES

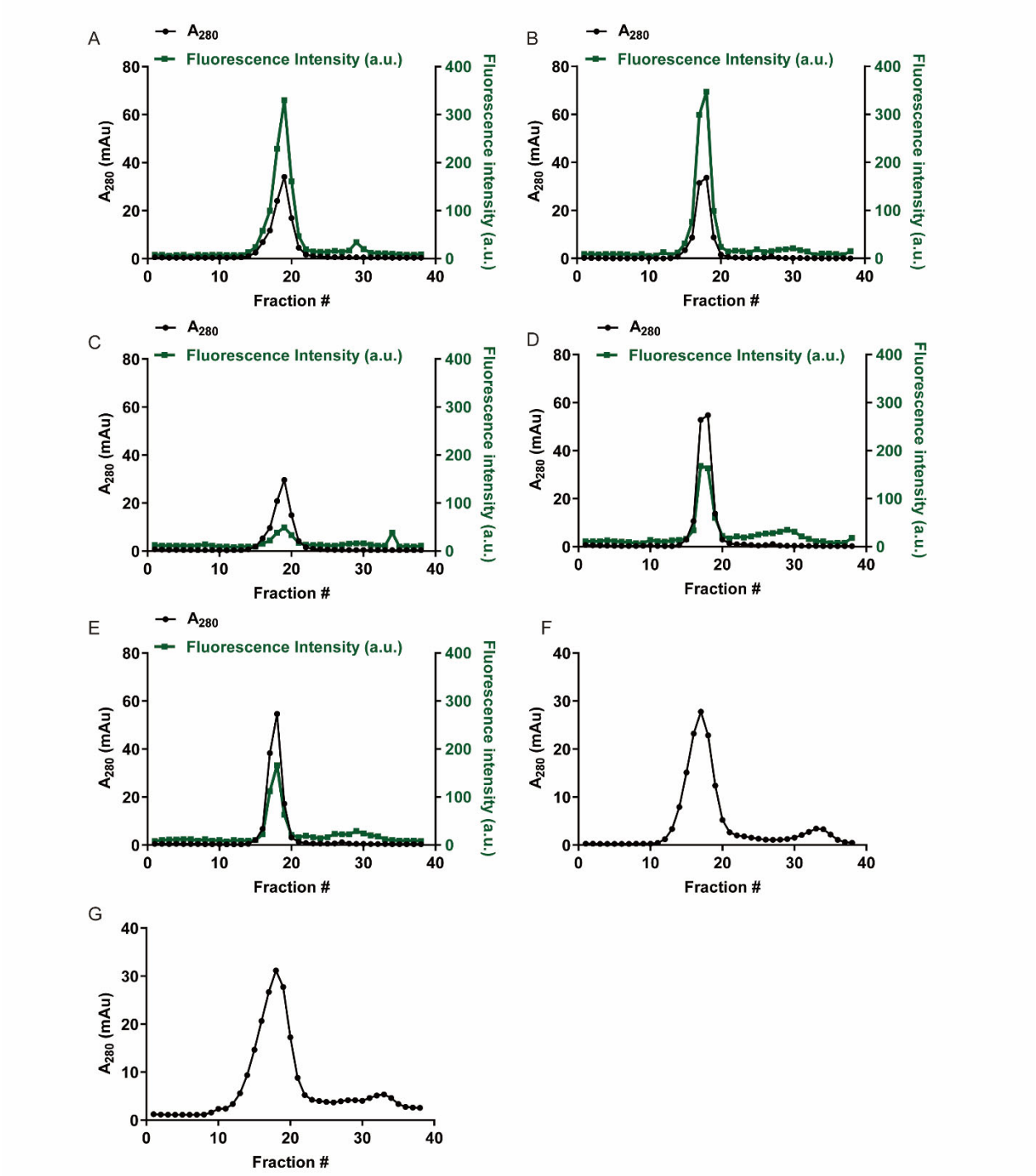

Figure S1. Representative size-exclusion chromatograms of encapsulation complexes between AaLS nanocages and GFP variants. A) wt-AaLS:GFP-AaRS12. B) AaLS-IC:GFP-AaRS12. C) wt-AaLS:GFP0. D) AaLS-IC:GFP0. E) AaLS-IC:GFP-M244C. F) Empty AaLS-IC. G) Empty wt-AaLS. A_{280} (left y-axes, black) and fluorescence emission at 508 nm (right y-axes, green) of the eluent from the Sephacryl S-400 column was measured and plotted against the fraction

number (4 mL per fraction). In Fig. S1C, the prominent fluorescence peak around fraction #35 (elution volume ~140 mL, exceeding the total column volume of 120 mL) is attributed to unencapsulated free GFP0 monomers, whose delayed elution is likely due to non-ideal secondary interactions with the Sephacryl S-400 column matrix. The faint peaks around fractions #28-31 likely correspond to lower-order complexes. This distinct separation of the fluorescence signal from the assembled cage fractions further corroborates the low encapsulation yield of the wt-AaLS:GFP0 combination.

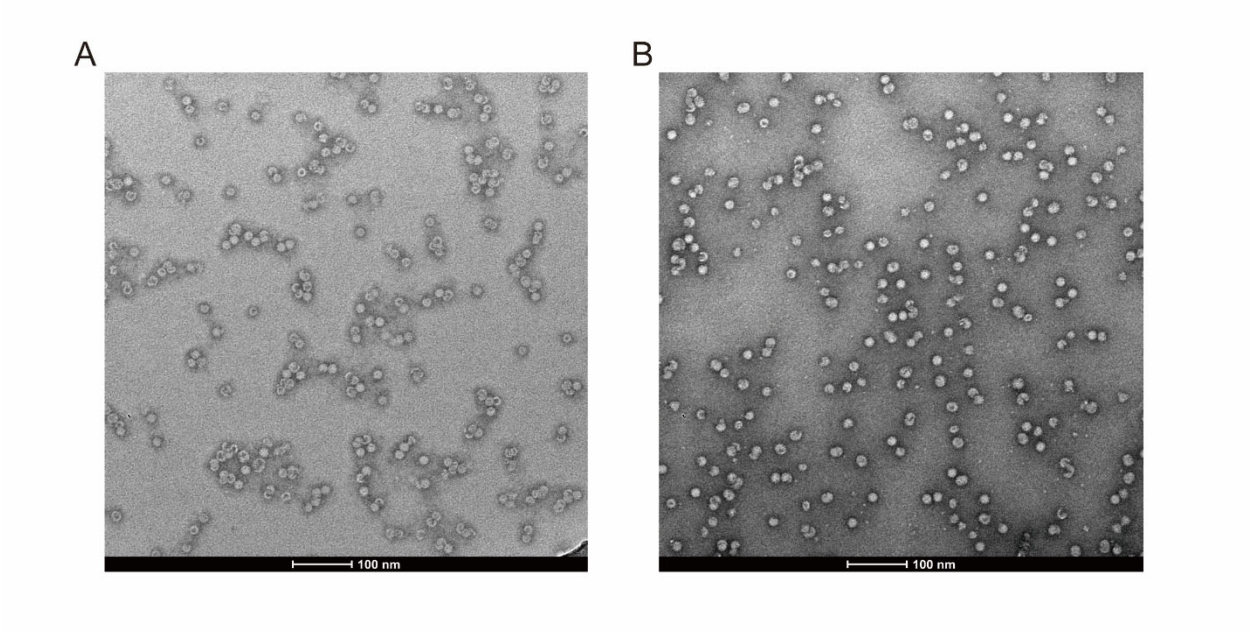

Figure S2. Negative-stain TEM images of purified AaLS-IC (A) and AaLS-IC:GFP-AaRS12 (B). Scale bars represent 100 nm.

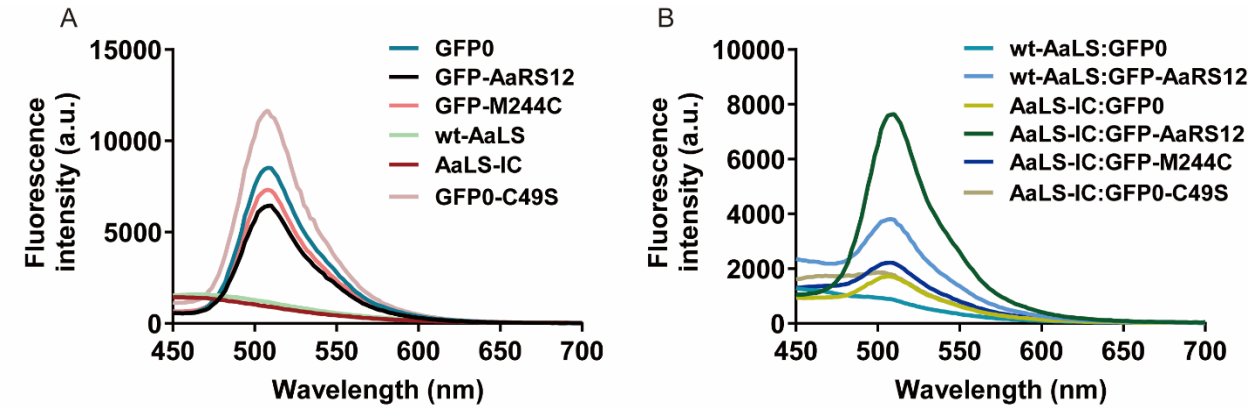

Figure S3. Fluorescence emission spectra. A) Individual GFP and AaLS variants. B) Encapsulation complexes between AaLS nanocages and GFP variants. Purified protein samples were excited at 375 nm and fluorescence emission was measured at different wavelengths in 2 nm increments.

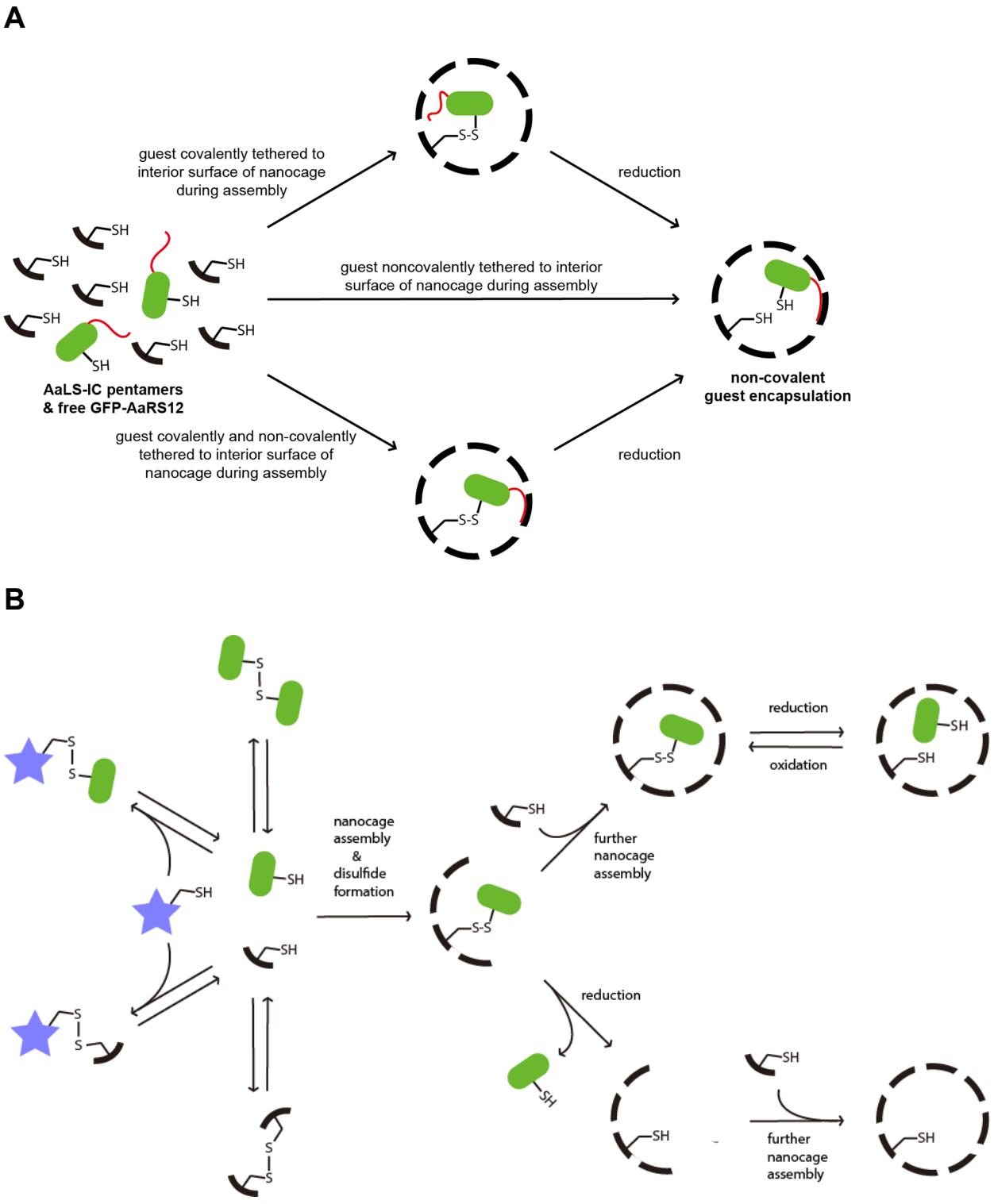

Figure S4. Schemes for loading GFP variants into AaLS-IC. A) Pathways for loading GFP-AaRS12 into AaLS-IC. During initial nanocage assembly in the cell, disulfide bond formation between GFP-AaRS12 (green) and AaLS-IC (black curved lines) increases the localization of

GFP-AaRS12 at the inner surface of the nanocage (top and bottom pathways), boosting the cargo loading yield compared to non-covalent loading alone (middle pathway). Small reducing agents, such as glutathione, can reduce the disulfide bond between the encapsulated guest and the intact nanocage (top and bottom pathways), leaving only non-covalent association between them. B) Non-specific disulfide bond formation during co-production of AaLS-IC and GFP in the cytosol. Co-production of AaLS-IC (black curved lines) and GFP (green) at high levels leads to their accumulation in the cytosol of *Escherichia coli* cells. A fraction of these proteins form disulfide bonds. Due to their high concentrations, when these proteins form disulfide bonds, they can form homodimers of AaLS-IC, homodimers of GFP, or heterodimers. Formation of these disulfide bonds should be rapidly reversible in the cytosol of *E. coli* BL21 (DE3) cells. The disulfide-linked conjugate of AaLS-IC and GFP can go on to co-assemble irreversibly with reduced AaLS-IC building blocks to form a nanocage with GFP attached to the interior surface of the nanocage. Reduction of the disulfide crosslink between AaLS-IC and GFP after nanocage assembly leaves reduced GFP physically trapped inside the nanocage. Because disulfide bond formation is likely to be non-specific, AaLS-IC building blocks or GFP can also possibly crosslink with many of the endogenous cytosolic proteins that contain a cysteine (purple stars). However, in contrast to AaLS-IC and GFP, few of the endogenous proteins will be present at high concentrations, so none of the disulfide crosslinked species between AaLS-IC and any individual endogenous protein are likely to accumulate. Rather, such species are likely to become rapidly re-reduced or undergo thiol-disulfide exchange to form the disulfide-linked AaLS-IC:GFP heterodimer. In an engineered *E. coli* strain with a more oxidizing cytosol (such as the SHuffle strain), all of the disulfide cross-linked forms could potentially accumulate. In principle, a higher level of the disulfide conjugate between AaLS-IC and GFP could lead to a higher encapsulation yield. However, disulfide crosslinked homodimers of AaLS-IC nanocage building blocks are not likely to be competent for nanocage assembly. Too much disulfide crosslinking between AaLS-IC subunits could interfere with proper assembly of this protein, possibly leading to degradation or uncontrolled oligomerization/aggregation/precipitation in the cell.

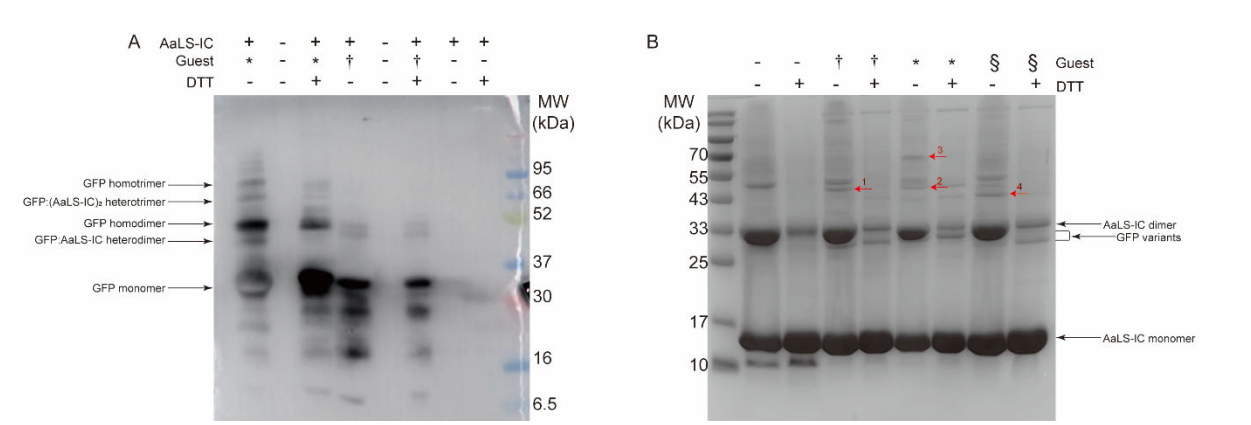

Figure S5. Most of the encapsulated GFP is not crosslinked to AaLS-IC in the isolated encapsulation complexes. A) GFP was visualized by Western blot using polyclonal mouse anti-GFP primary antibodies, goat anti-mouse secondary antibodies conjugated to horseradish peroxidase, and chemiluminescent detection. Purified AaLS-IC nanocages that were either co-produced with GFP-AaRS12 (*), co-produced with GFP0 (†), or produced in the absence of any GFP variant (-) were denatured in either the presence or absence of DTT and analyzed on the blot. The second and fifth lanes from the left were loaded with buffer only. The rightmost lane was loaded with molecular weight markers, and the corresponding molecular weights are indicated to the right of the blot. The arrows along the left side of the blot correspond to the predicted positions of bands corresponding to (from bottom to top): GFP monomer, disulfide-crosslinked heterodimer of GFP and AaLS-IC, disulfide-crosslinked homodimer of GFP, disulfide-crosslinked GFP:$(AaLS\text{-}IC)_2$ heterotrimer in which two different Cys residues from GFP form disulfide bonds with two different molecules of AaLS-IC (which only has one Cys residue, C122, per polypeptide chain), and disulfide-crosslinked GFP homotrimer. The molecular weights of the marker bands are given along the right side of the blot. B) AaLS-IC:GFP encapsulation complexes were also analyzed by SDS-PAGE. Purified AaLS-IC samples, either produced in the absence of any GFP variant (-) or co-produced with GFP0 (†), GFP-AaRS12 (*), or GFP-M244C (§), were denatured in either the presence (+) or absence (-) of DTT (indicated along the top edge of the gel). The expected positions of bands corresponding to the monomer form of AaLS-IC, the disulfide-linked homodimer of AaLS-IC, and the monomer forms of the GFP variants are indicated at the right edge of the gel. The expected positions of bands corresponding to disulfide-linked heterodimers between AaLS-IC and GFP are indicated on the gel by red arrows 1, 2, and 4. Red arrow 3 points to a band that corresponds to a disulfide-crosslinked $(AaLS\text{-}IC)_2$:GFP-AaRS12 heterotrimer. Assignments of the bands are based on comparison to the MW markers (leftmost lane), and the MW values of the proteins in the marker sample are given along the left edge of the gel.

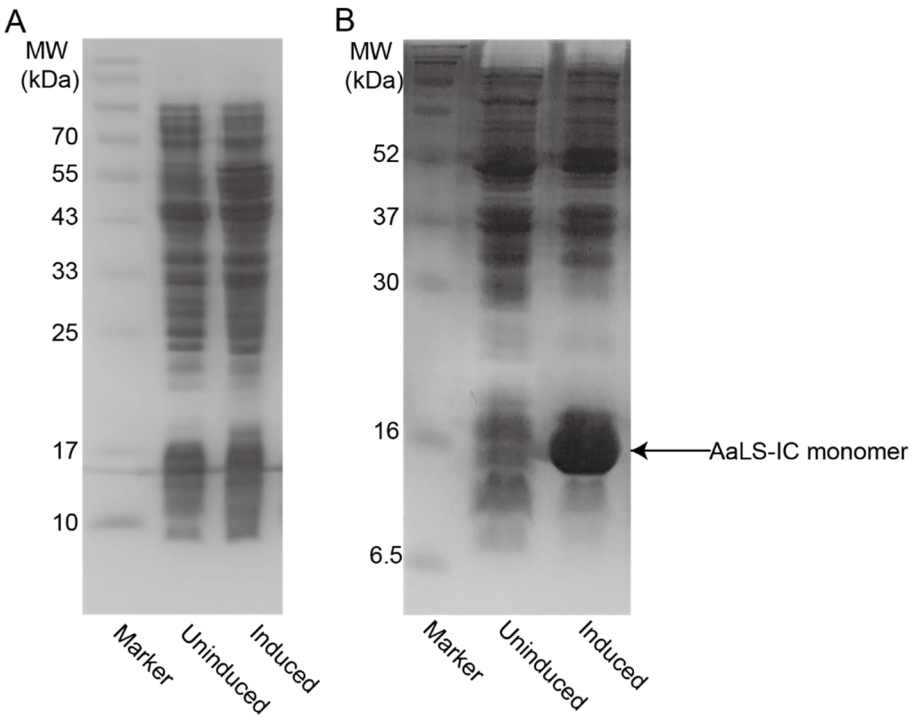

Figure S6. AaLS-IC is produced poorly in a bacterial strain with an oxidizing cytosol. SDS-PAGE analysis of AaLS-IC production in different *E. coli* strains. AaLS-IC was expressed from the plasmid pMG-AaLS-IC in *E. coli* SHuffle strain (A) and *E. coli* BL21 (DE3) (B). Whole cell samples were collected either immediately prior to induction with IPTG (Uninduced) or 14 h post-induction (Induced). The band corresponding to AaLS-IC was assigned based on comparison to protein molecular weight markers (Protein Marker). The MW values of the bands in the Protein Marker lane are listed at the left side of each gel. Arrows to the right of each gel indicate the expected mobility of AaLS-IC.

SUPPORTING TABLES

Table S1. Amino acid sequences of recombinant proteins produced in this work.

Protein	Encoding Plasmid	Amino Acid Sequence[a,b]
wt-AaLS	pMG-AaLS	MEIYEGKLTAEGLRFGIVASRFNHALVDRLVEGAI D**C**IVRHGGREEDITLVRVPGSWEIPVAAGELARKE DIDAVIAIGVLIRGATPHFDYIASEVSKGLANLSL ELRKPITFGVITADTLEQAIERAGTKHGNKGWEAA LSAIEMANLFKSLRLEHHHHHH
AaLS-IC	pMG-AaLS-IC	MEIYEGKLTAEGLRFGIVASRFNHALVDRLVEGAI DAIVRHGGREEDITLVRVPGSWEIPVAAGELARKE DIDAVIAIGVLIRGATPHFDYIASEVSKGLANLSL ELRKPITFGVITADTL**C**QAIERAGTKHGNKGWEAA LSAIEMANLFKSLRLE
GFP-AaRS12	pACYC-GFP-AaRS12	MASKGEELFTGVVPILVELDGDVNGHKFSVSGEGE GDATYGKLTLKFI**C**TTGKLPVPWPTLVTTL**C**YGVQ **C**FSRYPDHMKRHDFFKSAMPEGYVQERTIFFKDDG NYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGH KLEYNYNSHNVYIMADKQKNGIKVNFKTRHNIEDG SVQLADHYQQNTPIGDGPVLLPDNHYLSTQSALSK DPNEKRDHMVLLEFVTAAGITHGMDELYKSGGSMA LESSGGS**KKEDIFKEFLKW**
GFP0	pACYC-GFP	MASKGEELFTGVVPILVELDGDVNGHKFSVSGEGE GDATYGKLTLKFI**C**TTGKLPVPWPTLVTTL**C**YGVQ **C**FSRYPDHMKRHDFFKSAMPEGYVQERTIFFKDDG NYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGH KLEYNYNSHNVYIMADKQKNGIKVNFKTRHNIEDG SVQLADHYQQNTPIGDGPVLLPDNHYLSTQSALSK DPNEKRDHMVLLEFVTAAGITHGMDELYKSGGSMA LER
GFP-M244C	pACYC-GFP-M244C	MASKGEELFTGVVPILVELDGDVNGHKFSVSGEGE GDATYGKLTLKFI**C**TTGKLPVPWPTLVTTL**C**YGVQ **C**FSRYPDHMKRHDFFKSAMPEGYVQERTIFFKDDG NYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGH KLEYNYNSHNVYIMADKQKNGIKVNFKTRHNIEDG SVQLADHYQQNTPIGDGPVLLPDNHYLSTQSALSK DPNEKRDHMVLLEFVTAAGITHGMDELYKSGGS**C**
GFP0-C49S	pACYC-GFP0-C49S	MASKGEELFTGVVPILVELDGDVNGHKFSVSGEGE GDATYGKLTLKFI<u>S</u>TTGKLPVPWPTLVTTL**C**YGVQ **C**FSRYPDHMKRHDFFKSAMPEGYVQERTIFFKDDG NYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGH KLEYNYNSHNVYIMADKQKNGIKVNFKTRHNIEDG

SVQLADHYQQNTPIGDGPVLLPDNHYLSTQSALSK
DPNEKRDHMVLLEFVTAAGITHGMDELYKSGGSMA
LER

[a]Cysteines are colored red.
[b]The AaRS12 sequence is colored blue.

Table S2. Primers used in the construction of plasmids pACYC-GFP-AaRS12, pACYC-GFP-M244C and pACYC-GFP0-C49S.

Primer name	Sequence (5'→3')
HX3	TCGAGTCCTCTGGTGGCTCCAAAGAGGACATCTTCAAAGAATTTTTGAAGTGGTAA
HX4	CTAGTTACCACTTCAAAAATTCTTTGAAGATGTCCTCTTTGGAGCCACCAGAGGAC
RS12F	AGTCCTCTGGTGGCTCCAAAAAGGAGGACATCTTCAAAGA
RS12R	TCTTTGAAGATGTCCTCCTTTTTGGAGCCACCAGAGGACT
HCN157	GAACTATACAAATCCGGCGGCTCCTGCTAGCTCGAGCGTTAACGACGC
HCN158	GCGTCGTTAACGCTCGAGCTAGCAGGAGCCGCCGGATTTGTATAGTTC
C49S-F	TGAAGTTCATCTCCACTACTGGCAAACTGCCTGTT
C49S-R	AACAGGCAGTTTGCCAGTAGTGGAGATGAACTTCA

REFERENCES

1. X. Han and K. J. Woycechowsky, *Biochemistry,* 2017, **56,** 6211-6220.
2. Y. Azuma, R. Zschoche and D. Hilvert, *J. Biol. Chem.,* 2017, **292,** 10321-10327.3.
3. F. P. Seebeck, K. J. Woycechowsky, W. Zhuang, J. P. Rabe and D. Hilvert, *J. Am. Chem. Soc.,* 2006, **128,** 4516-4517.
4. S. Wang, A. T. Al-Soodani, G. C. Thomas, B. A. Buck-Koehntop and K. J. Woycechowsky. *Bioconjug. Chem.,* 2018, **29,** 2332-2342.
5. S. Lilavivat, D. Sardar, S. Jana, G. C. Thomas, and K. J. Woycechowsky, *J. Am. Chem. Soc.,* 2012, **134,** 13152-13155.
6. J. Fu and K. J. Woycechowsky, *Biochemistry,* 2020, **59,** 1517-1526.