Intein-Mediated Reconstitution of Split Lumazine Synthase for Programmable Protein Nanocage Assembly

Supplementary Materials

AaLS$_N$-Int^N (27796.46 Da)

MGSSHHHHHHSQDPMQIYEGKLTAEGLRFGIVASRFNHALVDRLVEGAIDSIVRHGGREEDITLVRVPGS
WEIPVAAGELARKEDIDAVIAIGVLIRGATPHFDYIASEVSKGLANLSLELCKPITFGVITADTLEQAIERA
GTAEYCLSYETEILTVEYGLLPIGKIVEKRIECTVYSVDNNGNIYTQPVAQWHDRGEQEVFEYCLEDGSL
IRATKDHKFMTVDGQMLPIDEIFERELDLMRVDNLPN

AaLS$_N$-ST-Int^N (29452.40 Da)

MGSSHHHHHHSQDPMQIYEGKLTAEGLRFGIVASRFNHALVDRLVEGAIDSIVRHGGREEDITLVRVPGS
WEIPVAAGELARKEDIDAVIAIGVLIRGSGGAHIVMVDAYKPTKATPHFDYIASEVSKGLANLSLELCKPI
TFGVITADTLEQAIERAGTAEYCLSYETEILTVEYGLLPIGKIVEKRIECTVYSVDNNGNIYTQPVAQWHD
RGEQEVFEYCLEDGSLIRATKDHKFMTVDGQMLPIDEIFERELDLMRVDNLPN

Int^C-AaLS$_C$-EGFRAfb (16209.30 Da)

MGRYTMGSSHHHHHHGSGGMIKIATRKYLGKQNVYDIGVERDHNFALKNGFIASNCFNKHGNKGWE
AALSAIEMANLFKSLRGGASVDNKFNKEMWAAWEEIRNLPNLNGWQMTAFIASLVDDPSQSANLLAE
AKKLNDAQAPKVD

AaLS(R108C)-Z-ST (29215.72 Da)

MGSSHHHHHHSQDPNSSSMQIYEGKLTAEGLRFGIVASRFNHALVDRLVEGAIDSIVRHGGREEDITLVR
VPGSWEIPVAAGELARKEDIDAVIAIGVLIRGATPHFDYIASEVSKGLANLSLELCKPITFGVITADTLEQA
IERAGTKHGNKGWEAALSAIEMANLFKSLRKDPNSGGGLVARGSGGGSGGGTGGGSGGGVDNKFNKE
QQNAFYEILHLPNLNEEQRNAFIQSLKDDPSQSANLLAEAKKLNDAQAPKGSGGAHIVMVDAYKPTK

SC-mEGFP (42514.38 Da)

MGSSHHHHHHSQDPMVDTLSGLSSEQGQSGDMTIEEDSATHIKFSKRDEDGKELAGATMELRDSSGKT
ISTWISDGQVKDFYLYPGKYTFVETAAPDGYEVATAITFTVNEQGQVTVNGKATKGDAHIKNSTSTGGG
SGGGVRERGMVSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFICTTGKLPVPWP
TLVTTLTYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLVNRIELK
GIDFKEDGNILGHKLEYNYNSHNVYIMADKQKNGIKVNFKIRHNIEDGSVQLADHYQQNTPIGDGPVL
LPDNHYLSTQSKLSKDPNEKRDHMVLLEFVTAAGITLGMDELYK

SC-SuperNovaRed (40660.10 Da)

MGSSHHHHHHSQDPMVDTLSGLSSEQGQSGDMTIEEDSATHIKFSKRDEDGKELAGATMELRDSSGKT
ISTWISDGQVKDFYLYPGKYTFVETAAPDGYEVATAITFTVNEQGQVTVNGKATKGDAHIKNSTGSEVG
PALFQSDMTFKIFIDGEVNGQKFTIVADGSSKFPHGDFNVHAVCETGKLPMSWKPICHLIQYGEPFFARY
PDGISHFAQECFPEGLSIDRTVRFENDGTMTSHHTYELDDTCVVSRITVNCDGFQPDGPIMRDQLVDILP
SETHMFPHGPNAVRQTATIGFTTADGGKMMGHFDSKMTFNGSRAIEIPGPHFVTIITKQTRDTSDKRDH
VCQREVAYAHSVPRITSAIGSDED

SC-SuperNovaRed-EGFRAfb (47615.89 Da)

MGSSHHHHHHSQDPMVDTLSGLSSEQGQSGDMTIEEDSATHIKFSKRDEDGKELAGATMELRDSSGKT
ISTWISDGQVKDFYLYPGKYTFVETAAPDGYEVATAITFTVNEQGQVTVNGKATKGDAHIKNSTGSEVG
PALFQSDMTFKIFIDGEVNGQKFTIVADGSSKFPHGDFNVHAVCETGKLPMSWKPICHLIQYGEPFFARY
PDGISHFAQECFPEGLSIDRTVRFENDGTMTSHHTYELDDTCVVSRITVNCDGFQPDGPIMRDQLVDILP
SETHMFPHGPNAVRQTATIGFTTADGGKMMGHFDSKMTFNGSRAIEIPGPHFVTIITKQTRDTSDKRDH
VCQREVAYAHSVPRITSAIGSDEDGASVDNKFNKEMWAAWEEIRNLPNLNGWQMTAFIASLVDDPSQS
ANLLAEAKKLNDAQAPKVD

Figure S1. Amino acid sequence of proteins. The sequence of each protein is presented in the specified color, while the linker region is shown in black.

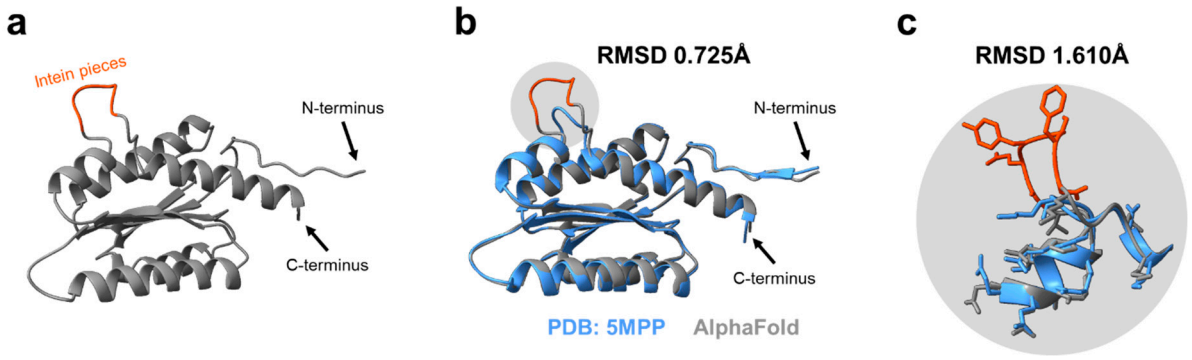


Figure S2. Structural comparison between spliced monomers and AaLS wild-type monomer. (a) Structure of the post-splicing monomer containing the intein pieces predicted by AlphaFold. (b) Comparison of structural similarity between the predicted structure and the AaLS-wt (PDB: 5MPP) monomer. (c) Structural effects of intein pieces remaining between monomer sequences after splicing. RMSD was calculated using Chimera X software. Orange = Intein pieces after splicing, Blue = AaLS-wt, Gray = spliced AaLS.

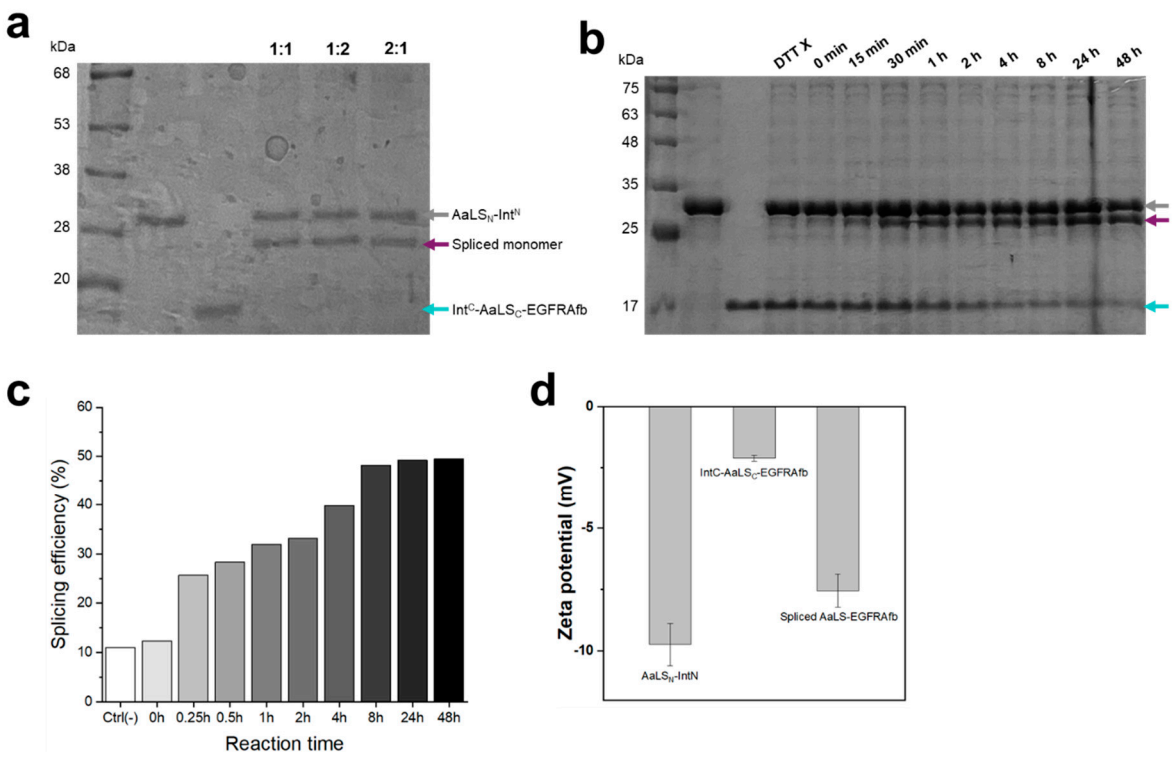


Figure S3. Optimization of intein-mediated *trans*-splicing. (a) SDS-PAGE analysis comparing splicing reactions after mixing N-fragments (AaLS$_N$-Int^N) and C-fragments (Int^C-AaLS$_C$-EGFRAfb) at three ratio conditions (1:1, 1:2, 2:1) with 4mM DTT for 1-hour. (b) Comparison of splicing efficiency based on DTT presence/absence and reaction time after mixing N-fragments and C-fragments at a 1:1 ratio. All splicing reactions were performed at 37 ℃. The protein standard marker is present on the left side of the gel. Gray arrow = AaLS$_N$-Int^N, Purple arrow = Spliced monomer, Light blue arrow = Int^C-AaLS$_C$-EGFRAfb. (c) Calculated splicing efficiency depending on SDS-PAGE by Image J. (d) Surface charge differences between fragments and spliced proteins. Data are presented as mean ± standard deviation (SD) (n=3).

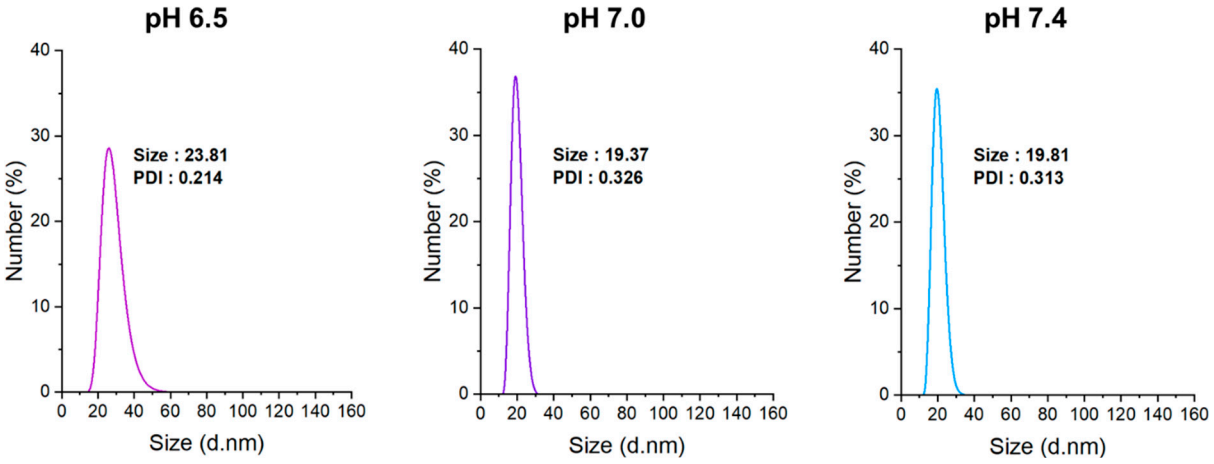


Figure S4. DLS measurements of reassembled spliced monomers under various pH conditions. Protein that had completed splicing reactions were dialyzed at three selected points within the pH range from slightly acid pH, mimicking the tumor microenvironment (TME), to physiological conditions, and their size and PDI were measured.

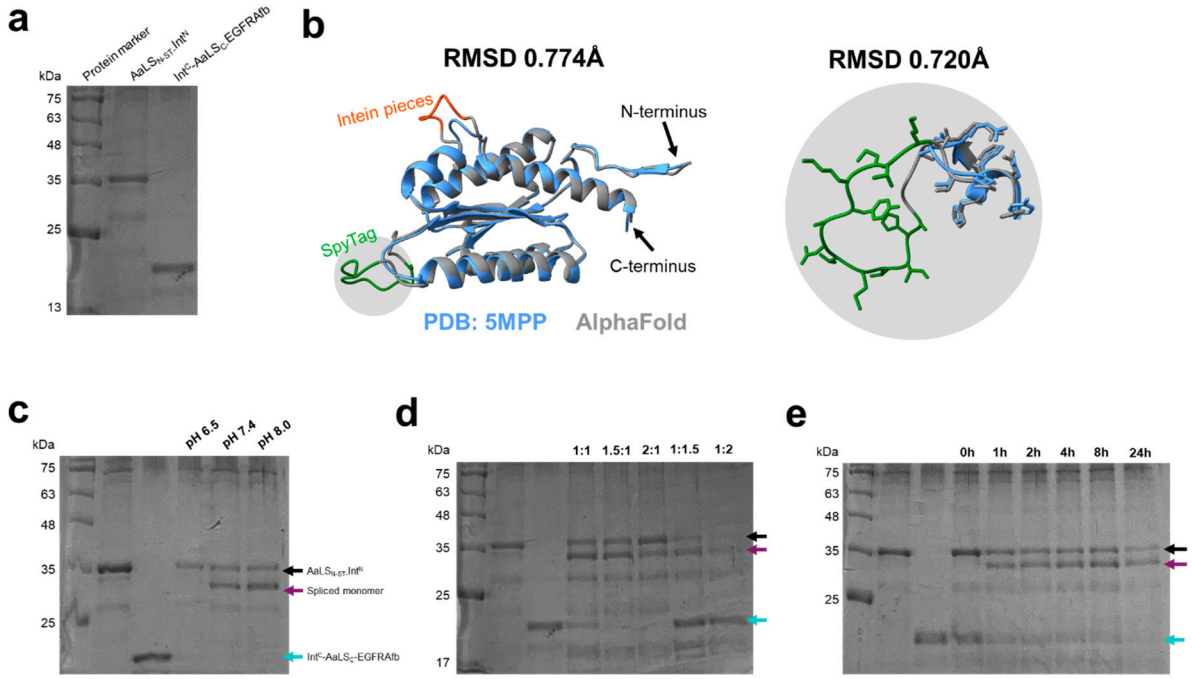


Figure S5. Structural analysis of proteins designed for cargo selective encapsulation and comparative analysis of intein-mediated *trans*-splicing. (a) SDS-PAGE analysis of purified AaLS$_{N-ST}$-Int^{N} and Int^{C}-AaLS$_{C}$-EGFRAfb proteins. (b) Comparison of structural similarity between the predicted post-splicing protein by AlphaFold and AaLS-wt (PDB: 5MPP). The predicted structure includes the intein pieces and SpyTag and exhibits a structure highly similar to the existing monomer. RMSD was calculated using Chimera X software. Orange = Intein pieces after splicing, Green = SpyTag, Blue = AaLS-wt, Gray = spliced AaLS. (c) Comparison of splicing reactions for N-fragments containing SpyTag and C-fragments by pH, (d) ratios, and (e) times. Black arrow = AaLS$_{N-ST}$-Int^{N}, Purple arrow = Spliced monomer, Light blue = Int^{C}-AaLS$_{C}$-EGFRAfb.

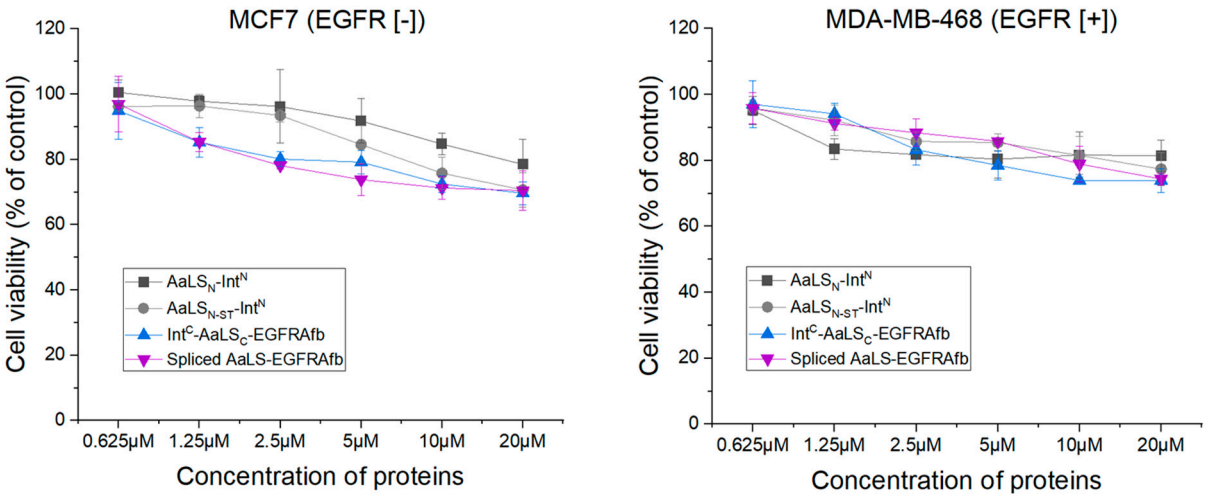


Figure S6. Cytotoxicity of proteins in EGFR positive and negative breast cancer cell lines. Each protein exhibits no EGFR expression-dependent cytotoxicity and maintains high cell viability even at high concentrations. Data are presented as mean ± standard deviation (SD) (n=3).

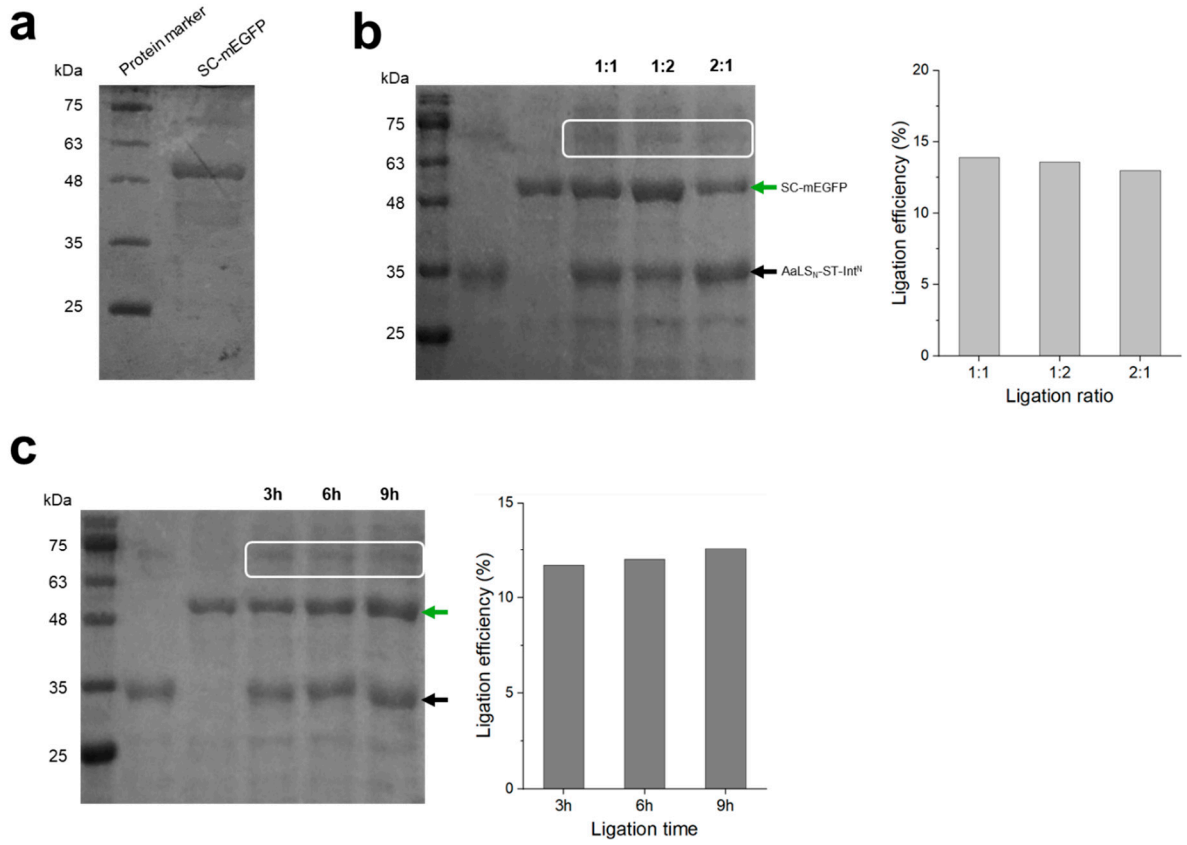


Figure S7. Optimization of protein nanocage formation as a biosensor. (a) SDS-PAGE analysis of SC-mEGFP protein. (b) Comparative analysis of fusion efficiency at various ratios of $AaLS_{N\text{-}ST}$-Int^N and SC-mEGFP. (c) Comparison of reaction time optimization and conjugation efficiency of $AaLS_{N\text{-}ST}$-Int^N and SC-mEGFP at a 1:1 ratio. Green arrow = SC-mEGFP, Black arrow = $AaLS_{N\text{-}ST}$-Int^N, White line = ST/SC ligated protein.

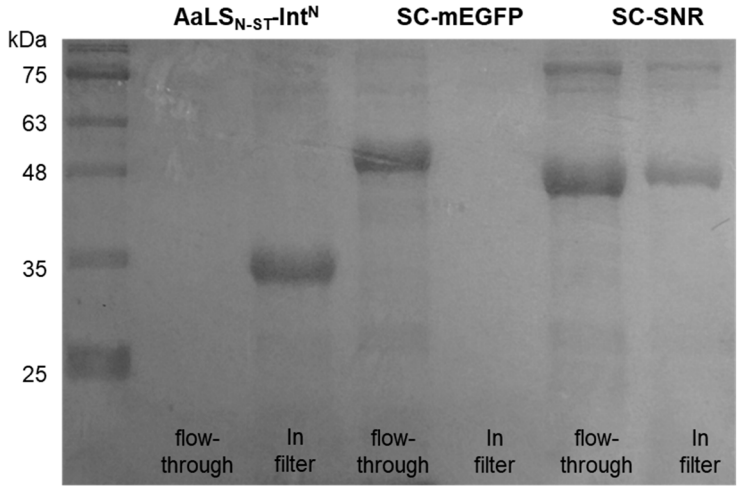


Figure S8. Removal of SC-fusion proteins. The SC-mEGFP and SC-SuperNovaRed (SNR) that remained unbound to AaLS$_{N-ST}$-Int^N were removed as flow-through (more than 90%) using a 50 kDa cut-off centrifugal filter. The AaLS$_{N-ST}$-Int^N protein, which is smaller in size than the SC-fusion proteins, was not removed by the same filter and remained in the sample, as confirmed by SDS-PAGE analysis.

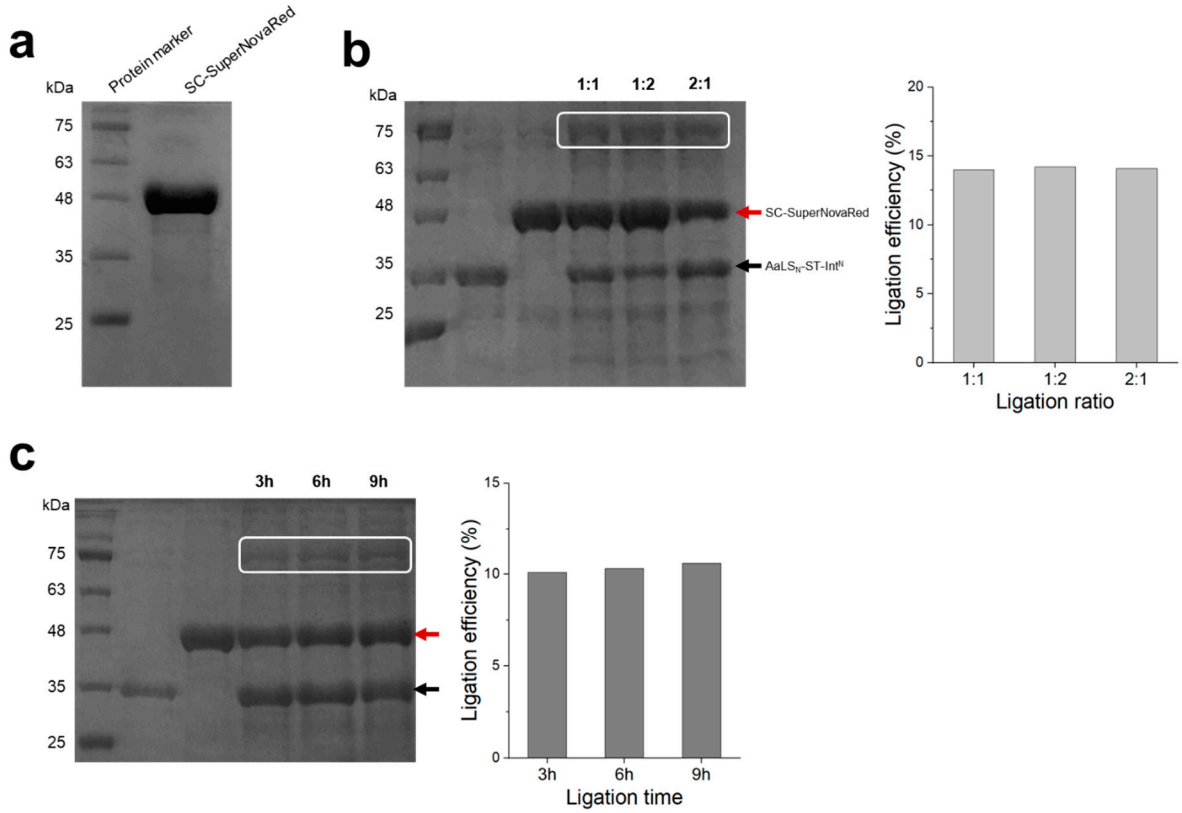


Figure S9. Optimization of protein nanocage formation as a bioreactor. (a) SDS-PAGE analysis of purified SC-SuperNovaRed protein. (b) Comparative analysis of fusion efficiency at various ratios of $AaLS_{N\text{-}ST}$-Int^{N} and SC-SuperNovaRed. (c) Comparison of reaction time optimization and conjugation efficiency of $AaLS_{N\text{-}ST}$-Int^{N} and SC-SuperNovaRed at a 1:1 ratio. Red arrow = SC-SuperNovaRed, Black arrow = $AaLS_{N\text{-}ST}$-Int^{N}, White line = ST/SC ligated protein. Ligation efficiency was analyzed using Image J.

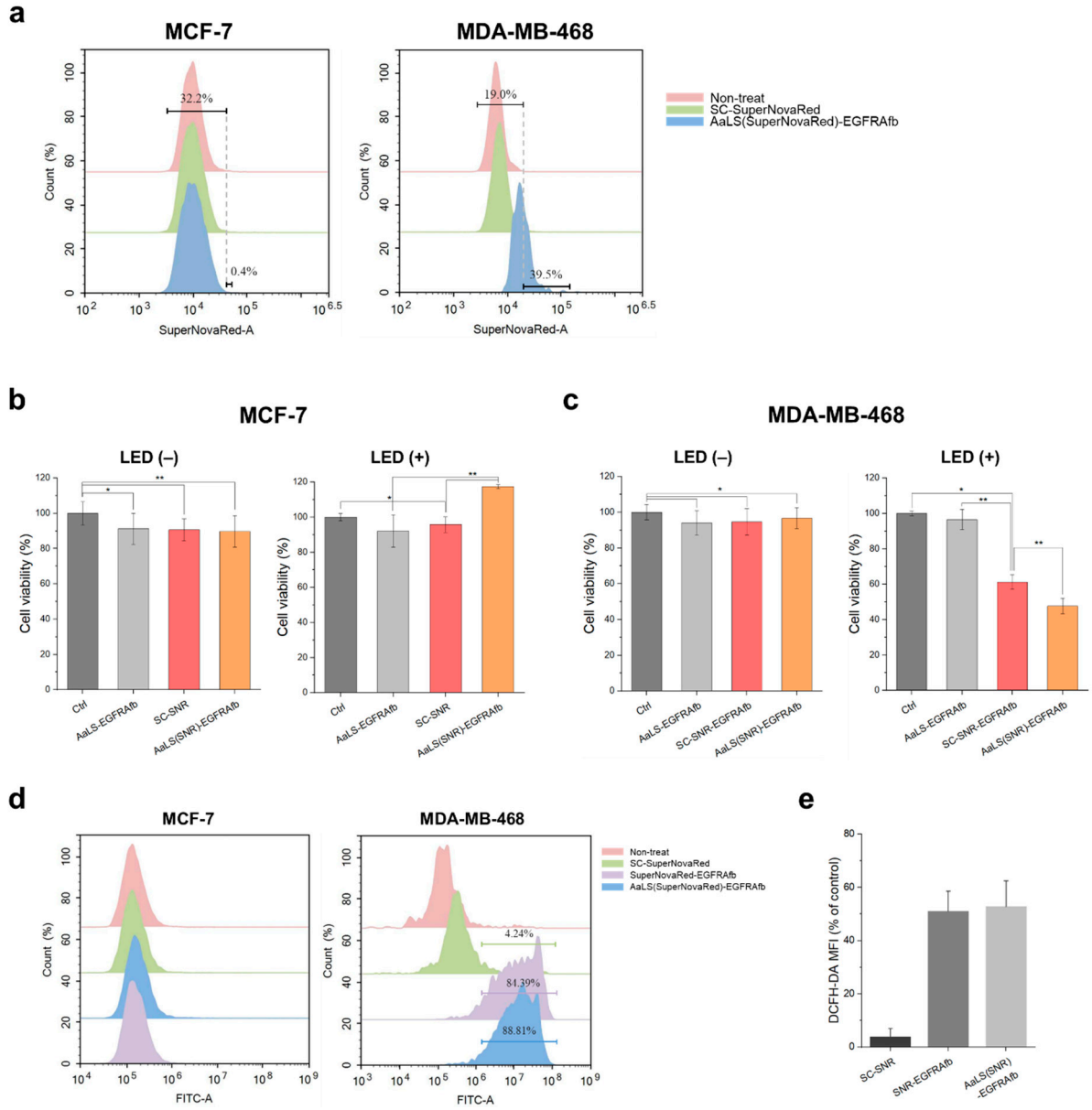


Figure S10. EGFR-specific delivery of protein nanocages encapsulating SuperNovaRed and the generation of reactive oxygen species (ROS) by LED. (a) Flow cytometry histograms of the SC-SuperNovaRed and SuperNovaRed-encapsulated protein nanocage platforms in EGFR-low (MCF-7) and EGFR-overexpressing (MDA-MB-468) cell lines. (b) LED-induced cell death in MCF-7 and (c) MDA-MB-468 cells caused by a protein nanocage reassembled without SC-SuperNovaRed (SC-SNR), SC-SNR (negative control protein) or SC-SNR-EGFRAfb (positive control protein), and a protein nanocage platform encapsulating SNR. (d) Flow cytometry histogram of ROS formation in a targeted protein nanocage platform. (e) Quantitative analysis of ROS generated by LED exposure in MDA-MB-468 cells. Data are presented as mean ± standard deviation (SD) (n=3). Statistical significance is indicated as $p < 0.05$ (*), $p < 0.01$ (**), and $p < 0.001$ (***).

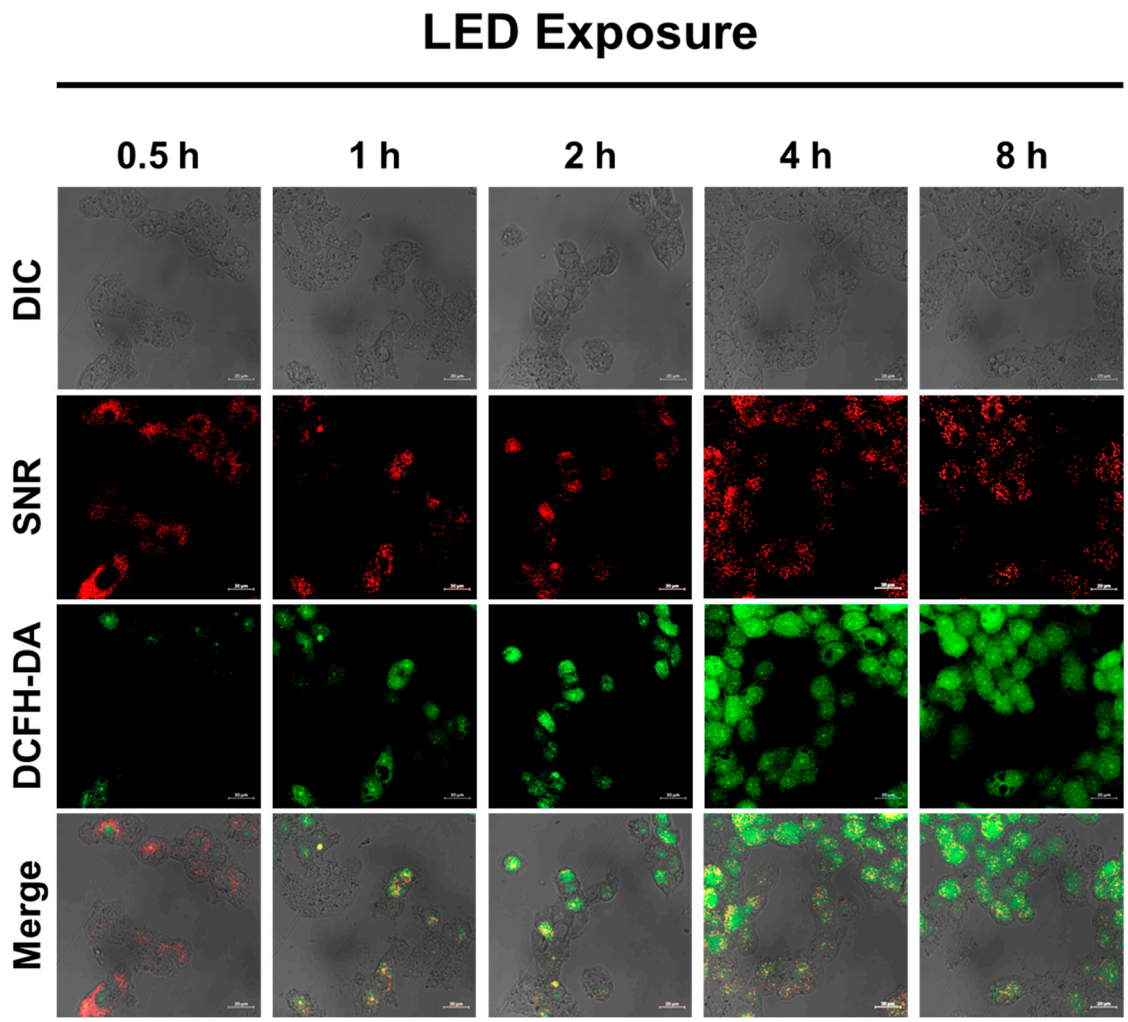


Figure S11. ROS formation in an LED exposure-time-dependent SuperNoraRed-encapsulated protein nanocage platform. The reassembled protein nanocage platform encapsulating SuperNovaRed (SNR) was delivered via receptor-mediated endocytosis by targeting EGFR overexpressed on the cell surface, and ROS formation was analyzed using fluorescence imaging at 0.5, 1, 2, 4, and 8 hours after exposure to LED light. The MDA-MB-468 cell line was used, and DCFH-DA was detected on the FITC channel. Scale bar = 20 μm.

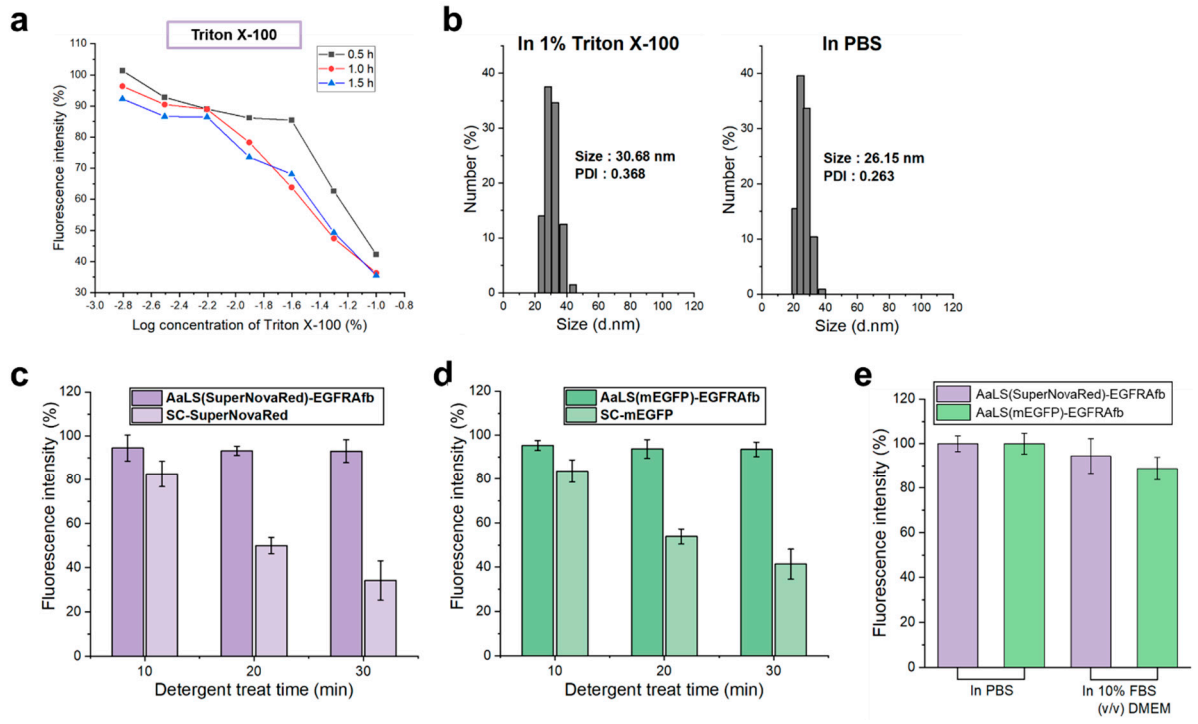


Figure S12. Protein nanocages that stably encapsulate cargo molecules protect the molecules inside from detergents. (a) Quenching curves showing the reaction of Triton X-100 and 10 μM SC-SuperNovaRed at various dilutions ranging from 1% to 0.016% over time (0.5, 1.0, 1.5h). (b) DLS measurement of the diameter of SuperNovaRed-encapsulated protein nanocage in 1% Triton X-100 or 1X PBS at 1 hour. (c-d) Difference in fluorescence intensity between protein nanocages encapsulating fluorescent proteins and free fluorescent proteins in 1% Triton X-100. (e) Observation of fluorescence quenching of fluorescent proteins (SuperNovaRed and mEGFP) inside nanocages in DMEM containing 10% FBS, which mimics an *in vivo* environment. Data are presented as mean ± standard deviation (SD) (n=3).