Supporting information

Complementary charge-driven encapsulation of functional protein by engineered protein cages *in cellulo*

Daniel Zakaszewski[ab], Lukasz Koziej[a], Jędrzej Pankowski[ac], V. Vishal Malolan[ab], Nina Gämperli[d], Jonathan G. Heddle[a], Donald Hilvert[d] and Yusuke Azuma[ad]*

[a] *Malopolska Centre of Biotechnology (MCB), Jagiellonian University, Gronostajowa 7A, 30387 Krakow, Poland. E-mail: yusuke.azuma@uj.edu.pl*
[b] *Doctoral School of Exact and Natural Sciences, Jagiellonian University, Prof. S. Łojasiewicza 11, 30348 Krakow, Poland*
[c] *Faculty of Biochemistry, Biophysics, and Biotechnology, Jagiellonian University, Gronostajowa 7, 30387 Krakow, Poland*
[d] *Laboratory of Organic Chemistry, ETH Zurich, 8093 Zurich, Switzerland*

Materials and methods

Materials

All chemicals and biochemicals were purchased from Sigma-Aldrich (Burlington, MA, USA), New England BioLabs (Ipswich, MA, USA) or Thermo Fisher Scientific (Waltham, MA USA). Oligonucleotides were synthesized by Microsynth AG (Balgach, Switzerland) or Sigma-Aldrich. *E. coli* strains BL21-Gold(DE3), XL1-Blue, and DH5α competent cells were purchased from Agilent (Santa Clara, CA, USA), Agilent and Thermo Fisher Scientific, respectively.

Molecular cloning

The plasmids and the primers used in this study are listed in Tables S1 and S2, respectively.

pMG/*Aa*LS-Strep was prepared from pMG/*Aa*LS-wt[11] by cassette cloning *via* XhoI and SpeI sites using oligonucleotides FW-XhoI_Strep and RV-SpeI_Strep. The genes encoding *Aa*LS-neg and *Aa*LS-13 from pMG/*Aa*LS-neg and pMG/*Aa*LS-13[24] were subcloned into pMG/*Aa*LS-Strep *via* NcoI and XhoI sites, yielding pMG/*Aa*LS-neg-Strep and pMG/*Aa*LS-13-Strep, respectively. The control plasmid, pMG/empty, was prepared by digestion of the pMG/*Aa*LS-13 plasmid with XbaI and SpeI, followed by direct ligation utilizing the compatible CTAG cohesive ends generated by these enzymes.

All tetracycline-inducible protein expression vectors were prepared from the plasmid pAC-P_{tet}/T7pol-S[48] which contains a gene for a tetracycline promoter (P_{tet})-controlled T7 RNA polymerase *C*-terminally fused to the bacterial SsrA degradation tag. The RNA polymerase gene was replaced with the *N*-terminally His-tagged GFP(+36) gene from pACYC/H_6-GFP(+36)[29] via XbaI and XhoI sites, yielding pAC-P_{tet}/H_6-GFP(+36)-SsrA. A stop codon was introduced between the genes for GFP(+36) and SsrA *via* QuickChange mutagenesis using oligonucleotides FW-GFP_stop and RV-GFP_stop as mutagenic PCR primers, yielding pAC-P_{tet}/H_6-GFP(+36).

V206K mutation was introduced into sfGFP gene harboured by pACYC/H_6-sfGFP plasmid[11] *via* site-directed mutagenesis using oligonucleotides FW-sfGFP_V206K and RV-sfGFP_V206K as mutagenic PCR primers, yielding pACYC/H_6-msfGFP. The msfGFP gene was subcloned into pAC-P_{tet}/H_6-GFP(+36)-SsrA and pAC-P_{tet}/H_6-GFP(+36) via BamHI and XhoI sites, yielding pAC-P_{tet}/H_6-msfGFP-SsrA and pAC-P_{tet}/H_6-msfGFP, respectively.

E. coli strains XL1 Blue or DH5α were used as the host cells for every cloning step. Sequences of plasmids were confirmed by DNA Sanger sequencing performed by Microsynth AG or Eurofins Genomics Europe Sequencing GmbH (München, Germany).

Protein production

All cage and guest proteins were coproduced in *E. coli* strain BL21-Gold(DE3) transformed with pMG and pAC-P_{tet} vectors, which encode the guest proteins and host cages, respectively. In general, cells were cultured at 37 °C in lysogeny broth, Miller formulation (LB) medium supplemented with 100 μg/ml carbenicillin and 35 μg/ml chloramphenicol until the OD_{600} reached ~0.7–0.9, at which point protein production was induced by the addition of isopropyl β-D-1-thiogalactopyranoside (IPTG) (0.2 mM) for host cages and tetracycline (2 μM) for guest proteins. After culturing at 30 °C for 20 h, cells were subjected to flow cytometry, confocal microscopic analysis, and protein purification.

The protein production conditions, 2 μM tetracycline and 20 h culturing time, were employed based on the results of preliminary titration experiments (Fig. S6a), in which GFP(+36)-SsrA abundance was examined 4.5 or 20 h after induction with 0–8 μM tetracycline. For the assay using the salicylate operon promoter (Fig. S1b–c), protein production was induced with sodium salicylate (0.2 mM) and tetracycline (2 μM) instead.

Analytical flow cytometry

Liquid cultures prepared from separate biological replicates were diluted 333× in phosphate-buffered saline (PBS). Flow cytometry data were acquired using a Navios Flow Cytometer (Beckman Coulter, Brea, CA, USA), equipped with a solid-state blue laser operating at 488 nm. Forward and side scatter (FS and SS, respectively) were detected by a photodiode through a 488/10 nm bandpass filter, and green fluorescence by a photomultiplier tube through a 525/40 nm bandpass filter. The starting sample size was set to 33,000 events. Cytometric data were then processed employing a gating protocol in Kaluza C software (Beckman Coulter) and only the events that fell within the range of the gates were considered (Fig. S7).

Confocal laser scanning microscopy

Cells were harvested by centrifugation at 4,000 g, 4 °C for 2 min and resuspended in PBS (25% original volume). Live cell imaging was performed using a Zeiss Axio Observer inverted microscope with the LSM 880 laser scanning unit (Carl Zeiss, Thornwood, NY, USA), equipped with an α Plan-Apochromat 100x/1.46 Oil DIC M27 objective, an argon laser operating at 488 nm, an AxioCam MRm camera, and an Airyscan detector. The detection wavelength was set to 545/55 nm and the pixel time to 1.7 μs. Detector gain was adjusted for each sample since the fluorescence intensity of the cells varied greatly depending on the specific proteins that were produced. Therefore, the images were used for qualitative purposes only. Data were analysed with ZEISS ZEN software (Carl Zeiss).

Protein purification

Cells were harvested by centrifugation at 3,000 g, 4 °C for 10 min and then stored at –20 °C until purification. The cell pellet from 4 ml culture was suspended in 200 μl of 100 mM Tris-HCl buffer (pH 7.5) containing 150 mM NaCl, 1 mM ethylenediaminetetraacetate (EDTA), and 5 mM sodium phosphate buffer (pH 7.5) supplemented with 0.1 mg/ml lysozyme, 0.01 mg/ml DNase I, 0.01 mg/ml RNase A, and protease inhibitor cocktail. After incubation at 37 °C for 1 h, the sample was lysed by sonication and cleared by centrifugation at 22,000 g, 25 °C. Supernatant was loaded onto 50 μl StrepTactin Sepharose High Performance resin (Cytiva Europe GmbH, Freiburg im Breisgau, Germany) in a spin column. After washing three times with 200 μl of 100 mM Tris-HCl buffer (pH 7.5) containing 150 mM NaCl, 1 mM EDTA, and 5 mM sodium phosphate buffer (pH 7.5), the protein was eluted with 100 μl phosphate-buffered saline containing 1 mM EDTA (PBS-E) and 2.5 mM desthiobiotin. Total protein concentration was determined using absorbance at 280 nm using the molar absorption coefficient of *Aa*LS proteins, ε_{280} = 19,630 M^{-1} cm^{-1}.[49] Isolated proteins were characterized by SDS-PAGE, native AGE, spectrofluorimetry, and cryogenic transmission electron microscopy (cryo-EM).

N-Terminally His-tagged GFP(+36) was produced and purified as previously reported.[29] The molar absorption coefficient of GFP(+36) in PBS was determined to be ε_{488} = 54,900 ± 2,100 M^{-1} cm^{-1} based on the absorbance of the chromophore obtained following protein denaturation in 0.1 M NaOH ($\varepsilon_{447,den}$ = 44,100 M^{-1} cm^{-1}).[50]

SDS-PAGE

Samples were prepared in an SDS-PAGE loading buffer (62.5 mM Tris-HCl (pH 6.8), 2.5% SDS, 10% glycerol, 5% β-mercaptoethanol, 0.005 % bromophenol blue), and subjected to heat denaturation. Cells harvested from 10 μl liquid culture or 2 μg of each purified protein were loaded onto 14% polyacrylamide gels and electrophoresed. Protein bands were visualized using ReadyBlue Protein Gel Stain (Sigma-Aldrich).

Native agarose gel electrophoresis

Samples were prepared in a native AGE loading buffer (50 mM BisTris-HCl (pH 7.2), 10% glycerol, 0.004% bromophenol blue). 2 μg of protein was loaded onto the gel. Electrophoresis was performed in 2% agarose-BIS-TRIS-tricine gels (150 mM BIS-TRIS, 50 mM tricine (pH 7.0)), running at 70 V for 40 min. Fluorescent bands were visualized using ChemiDoc MP (Bio-Rad) in fluorescein detection mode (460–490 nm excitation and 532/28 nm emission filters). The same gels were also stained with ReadyBlue Protein Gel Stain.

Spectrofluorimetry

Fluorescence spectra were recorded on an RF-6000 spectrofluorometer (Shimadzu, Kyoto, Japan) using 0.5 μM of proteins in PBS-E in a 1-cm-light-pass polystyrene cuvette. Upon excitation at 450/10 nm, the spectra were acquired over the range of 480–600 nm in 1 nm steps, at a 10 nm bandwidth and a scan speed of 2,000 nm/min. Light scattering from the cuvette and buffer was subtracted from the observed data. Measurements were performed on three separate batches of purified protein. GFP(+36)-SsrA loading efficiency was determined using purified GFP(+36) as a standard.

Cryogenic transmission electron microscopy

Approximately 3.5 μl of the 0.5 mg/ml protein solution were applied to freshly glow-discharged TEM grids (Quantifoil R2/1, Cu 200 mesh) and plunge-frozen in liquid ethane by a Vitrobot Mark IV (Thermo Fisher Scientific) using the following parameters: humidity: 95%, temperature: 4 °C, wait time: 30 s; blot total: 1; blot force: 5; blot time: 5 s; drain time: 0 s. Cryo-EM micrographs were collected at the National Cryo-EM Centre SOLARIS (Krakow, Poland) using a Glacios Cryo-Transmission Electron Microscope equipped with a Falcon 4 detector (Thermo Fisher Scientific), operating at a 200 kV accelerating voltage, a 190,000× magnification, and a corresponding pixel size of 0.73 Å/px. Micrographs were processed in Fiji software.[51]

Supplementary data

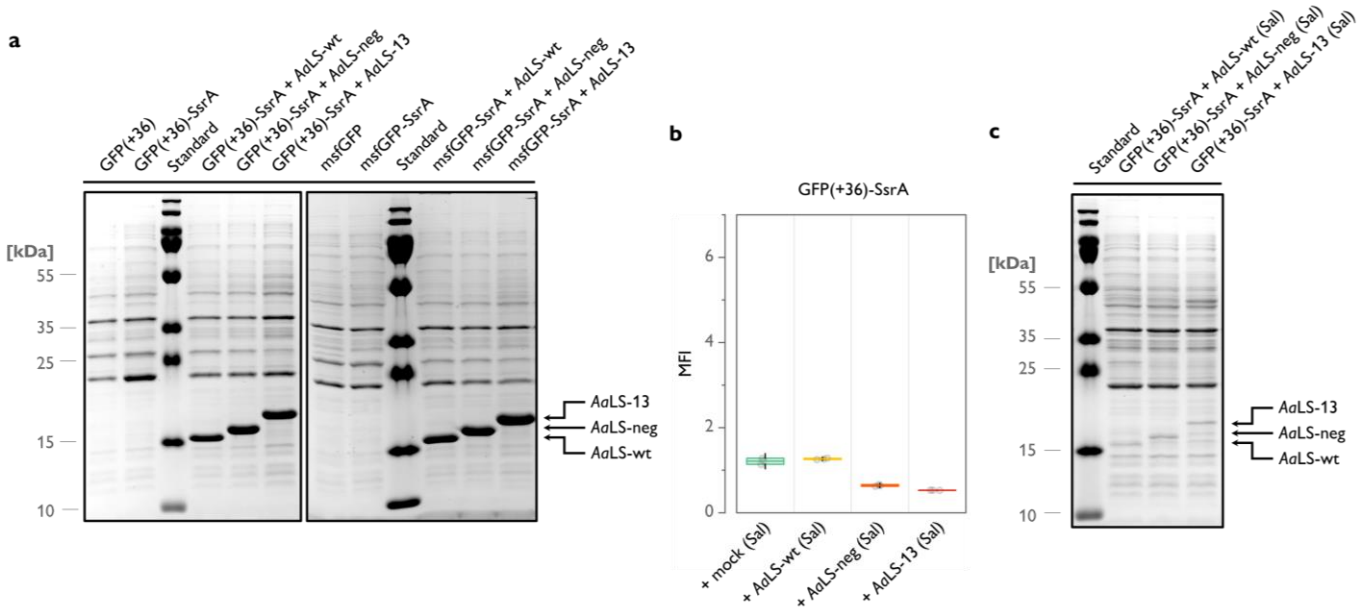

Fig. S1 Production levels of *Aa*LS proteins in cells. (a) SDS-PAGE analysis of whole-cell proteins from the cells used for the flow cytometry analysis shown in Fig. 2a–b. Calculated molecular masses of each protein are the following: *Aa*LS-wt, 18.1 kDa; *Aa*LS-neg, 18.1 kDa; *Aa*LS-13, 18.0 kDa; GFP(+36), 29.5 kDa; GFP(+36)-SsrA, 30.6 kDa. **(b)** Flow cytometry and **(c)** SDS-PAGE analysis of cells coproducing GFP(+36)-SsrA with *Aa*LS proteins upon induction with salicylate (Sal). The flow cytometry data are shown as the mean fluorescence intensity (MFI) of the cells, where boxes and whiskers indicate sample standard deviations and 95% confidence intervals, respectively (n_i = 3). For the gel images, bands corresponding to *Aa*LS proteins are indicated by arrows.

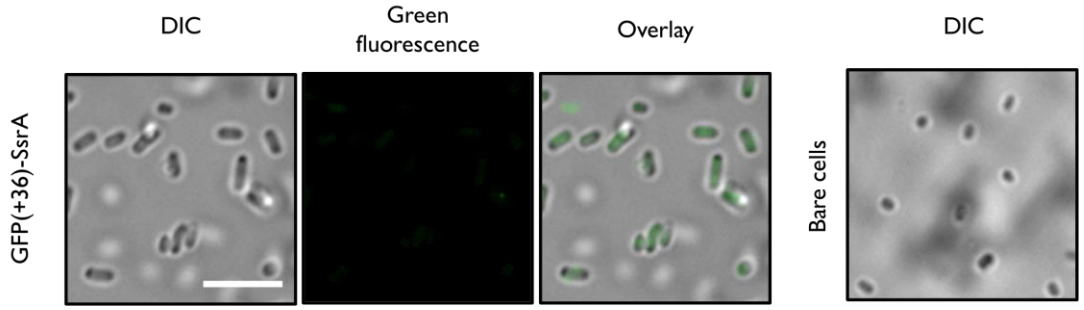

Fig. S2 Differential interference contrast (DIC) and confocal microscopy images of cells producing GFP(+36)-SsrA and non-transformed ("bare") cells. Scale bar – 5 μm. Fluorescence is only well visible in the overlay image due to the low signal attributed to the low abundance of GFP(+36)-SsrA in the cells. Importantly, SsrA-tagged GFP(+36) shows no tendency to form inclusion bodies, as opposed to GFP(+36) at the same concentration of inducer. Bare cells appear smaller than those producing recombinant proteins, which was taken into account for the flow cytometry gating procedure (Fig. S7).

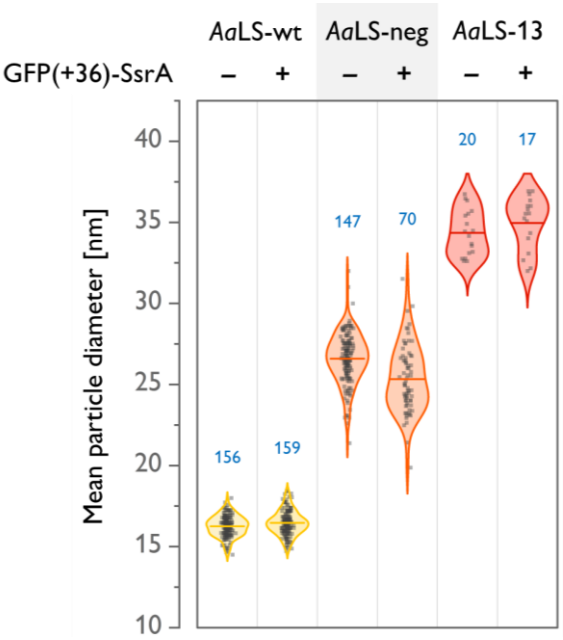

Fig. S3 Distribution of *Aa*LS-neg and *Aa*LS-13 diameters for samples purified with or without GFP(+36)-SsrA, as quantified from cryo-EM micrographs. Surfaces occupied by each particle in the micrographs were measured, and then the mean radius (and diameter) was calculated for each. Horizontal lines and outlines indicate means and probability densities, respectively, the latter scaled relative to total sample size and smoothed by a kernel density estimator. Numbers above each dataset represent the sample size.

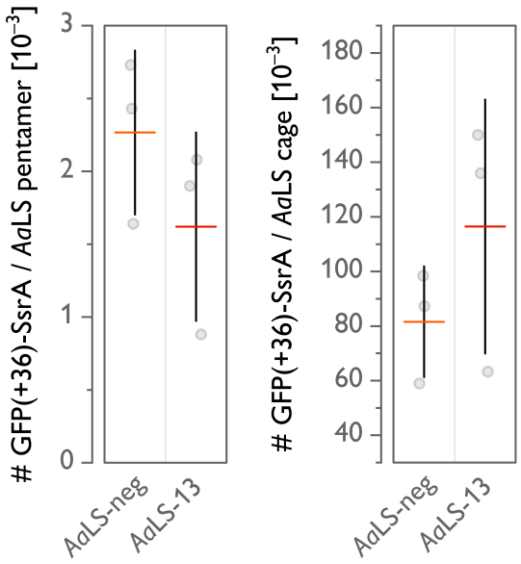

Fig. S4 Quantification of GFP(+36)-SsrA copurified with *Aa*LS-neg and *Aa*LS-13. Data are shown as the number of GFP molecules per *Aa*LS pentamer (left) and per cage (right), assuming that *Aa*LS-neg and *Aa*LS-13 assemble into 180-mer and 360-mer cages, respectively. Experiments were performed with 3 different batches of protein. Horizontal lines and whiskers indicate means and sample standard deviations, respectively.

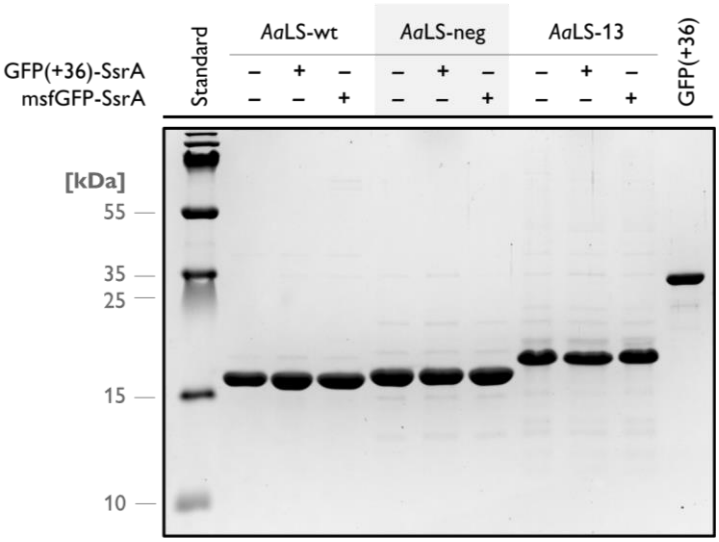

Fig. S5 SDS-PAGE analysis of purified *Aa*LS proteins produced alone or coproduced with SsrA-tagged GFPs. Calculated molecular mass of each protein is the following: *Aa*LS-wt, 18.1 kDa; *Aa*LS-neg, 18.1 kDa; *Aa*LS-13, 18.0 kDa; GFP(+36), 29.5 kDa; GFP(+36)-SsrA, 30.6 kDa.

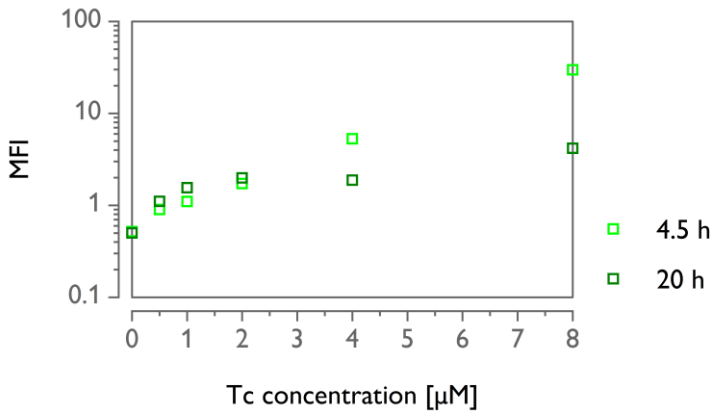

Fig. S6 Titration of the tetracycline promoter regulating GFP(+36)-SsrA gene expression. Median fluorescence intensities (MFI) from the flow cytometry analysis of cells harbouring the pMG/empty mock plasmid and pAC-P_{tet}/H_6-GFP(+36)-SsrA are plotted relative to the tetracycline (Tc) concentration. At Tc concentrations higher than 2 μM, fluorescence signals at the 4.5 h time point were significantly higher than those at 20 h, indicating that the rate of protein production can temporarily overtake cellular proteolysis capacity if gene expression is too intense. Since insufficient protein degradation could obscure the protein cage-triggered rescue effect, all the other experiments in this study were performed with 2 μM Tc and 20 h cell culturing.

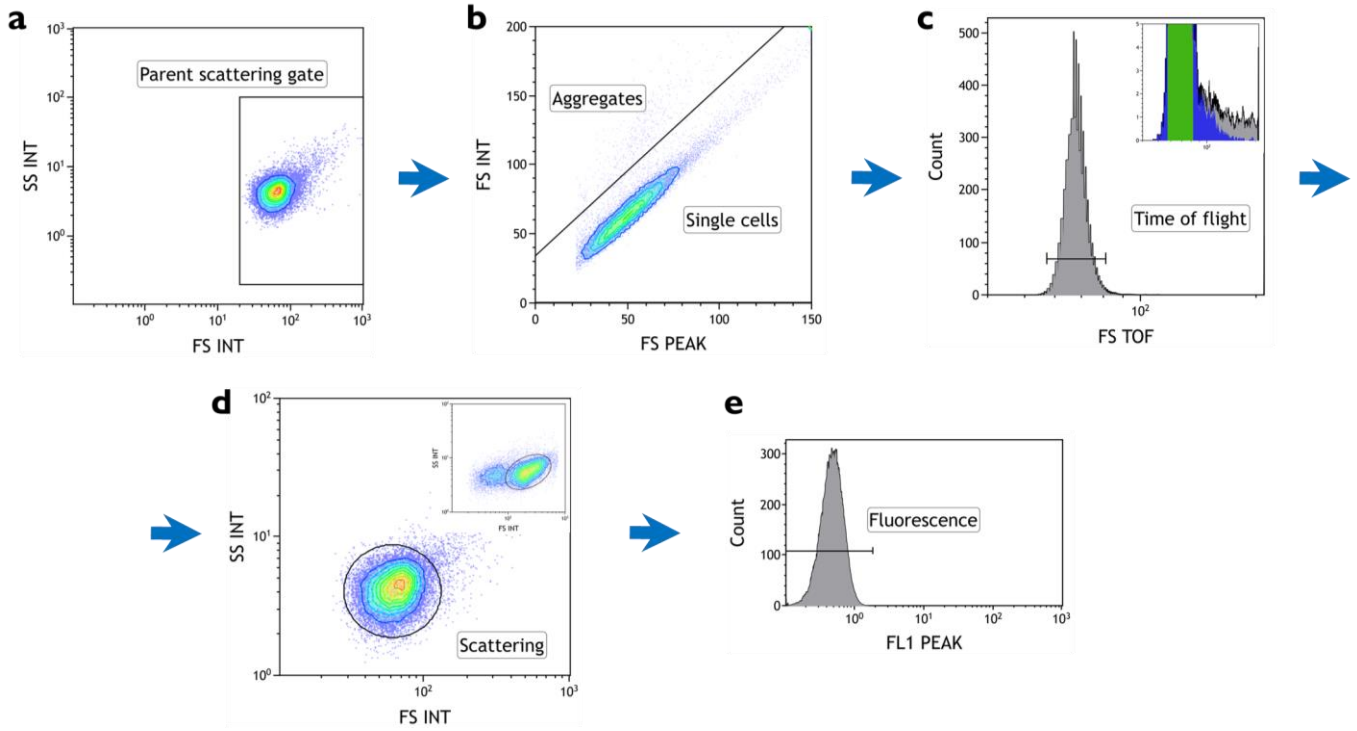

Fig. S7 The logic of gating in analytical flow cytometry of bacterial cells, shown on a sample of non-transformed cells. (a) The parent scattering gate was set at the stage of data acquisition. **(b)** Aggregates were then excluded from the analysis based on a cytogram of the front scatter integral (FS INT) *vs.* the front scatter peak value (FS PEAK). FS INT and FS PEAK should be linearly correlated for singlet objects, while aggregates yield an abnormally high FS INT signal. Thus, the subpopulation localized above the diagonal in the cytogram was excluded. **(c)** Aggregates that were overlooked by the previous gate were subsequently removed using the FS-based time of flight (FS TOF) parameter. Because of their large size, they yield abnormally high TOF values (blue in the inset graph with shortened *y*-axis). **(d)** The desired population was chosen assuming that the SS *vs.* FS pattern reflects the morphology of the objects. We varied the position of this gate for the different samples since the SS/FS pattern changed depending on which recombinant protein was overproduced (see inset cytogram for cells coproducing GFP(+36)-SsrA with *Aa*LS-neg as an example). **(e)** Lastly, the median fluorescence intensity (MFI) was determined from a log-scale graph of counts *vs.* green fluorescence intensity (FL1 PEAK).

Table S1 PCR primers used in the study.

Name	Sequence
FW-XhoI-Strep	TCGAGGGCGGCTGGAGCCATCCGCAGTTCGAAAAGTAA
RV-SpeI_Strep	CTAGTTACTTTTCGAACTGCGGATGGCTCCAGCCGCCC
FW-GFP_stop	GGCGCTCGAGTAAGCGGCGAACGATGAAAACTATGCG
RV-GFP_stop	CGTTCGCCGCTTACTCGAGCGCCATGGAGCCG
FW-sfGFP_V206K	CAACCATTACCTGTCGACACAATCTAAGCTTTCGAAAGATCCCAACGAAAAGC
RV-sfGFP_V206K	GCTTTTCGTTGGGATCTTTCGAAAGCTTA-GATTGTGTCGACAGGTAATGGTTG

Table S2 Plasmids used in the study.

Name	Cloned gene / Source organism	*N*-tag	*C*-tag	Promoter	ARG[d]	Ori
pMG/*Aa*LS-wt-Strep	lumazine synthase / *Aquifex aeolicus*	—	Strep	T7*lac*[b], *sal*[c]	*bla*[e]	pBR322
pMG/*Aa*LS-neg-Strep	lumazine synthase var. neg / synthetic	—	Strep	T7*lac*, *sal*	*bla*	pBR322
pMG/*Aa*LS-13-Strep	lumazine synthase var. 13 / synthetic	—	Strep	T7*lac*, *sal*	*bla*	pBR322
pMG/empty (mock)	n/a[a]	n/a	n/a	T7*lac*, *sal*	*bla*	pBR322
pAC-P_{tet}/H_6-GFP(+36)	GFP(+36) / synthetic	His-6	—	*tet*	*cat*[f]	p15A
pAC-P_{tet}/H_6-GFP(+36)-SsrA	GFP(+36) / synthetic	His-6	SsrA	*tet*	*cat*	p15A
pAC-P_{tet}/H_6-msfGFP	msfGFP / synthetic	His-6	—	*tet*	*cat*	p15A
pAC-P_{tet}/H_6-msfGFP-SsrA	msfGFP / synthetic	His-6	SsrA	*tet*	*cat*	p15A

[a] not applicable
[b] T7 promoter with lactose operon operator
[c] salicylate operon promoter
[d] antibiotic resistance gene
[e] beta-lactamase gene
[f] chloramphenicol acetyltransferase gene

References

52. Neuenschwander M. High-throughput assays for detecting enzymatic activities in living cells (PhD Thesis, ETH Zurich, 2007).
53. Lear S & Cobb SL. Pep-Calc.com: a set of web utilities for the calculation of peptide and peptoid properties and automatic mass spectral peak assignment. *J Comput Aided Mol Des* **30**, 271–277 (2016).
54. Ward WW. Properties of the coelenterate green-fluorescent proteins. in *Bioluminescence and Chemiluminescence: Basic Chemistry and Analytical Applications* (eds. Deluca MA & McElroy WD) 235–242 (Academic Press Inc., 1981).
55. Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, Preibisch S, Rueden C, Saalfeld S, Schmid B, Tinevez J-Y, White DJ, Hartenstein V, Eliceiri K, Tomancak P & Cardona A. Fiji: an open-source platform for biological-image analysis. *Nat Methods* **9**, 676–682 (2012).