Supplementary Material

TSGIT: an N- and C-terminal tandem tag system for purification of native and intein-mediated ligation-ready proteins

Vlad-Stefan Raducanu[#], Daniela-Violeta Raducanu[#], Yujing Ouyang[#], Muhammad Tehseen[#], Masateru Takahashi[#] and Samir M. Hamdan[#, *]

[#] King Abdullah University of Science and Technology, Division of Biological and Environmental Sciences and Engineering, Thuwal 23955, Saudi Arabia.

*Corresponding author.
E-mail address: samir.hamdan@kaust.edu.sa (S.M. Hamdan)

Supplementary Materials and Methods

1. Protein sample analysis

The purity of the eluted protein fractions was estimated by SDS-PAGE analysis as described previously [1,2] but with two modifications. First, the 5X electrophoresis sample buffer was changed to [10% SDS, 50 mM TCEP, 50% Glycerol, 250 mM Tris-HCl and 0.5% Bromophenol Blue dye, pH 6.8] to replace sulfhydryl-containing DTT with TCEP as reductant. Second, the samples were heated up at 96^{o}C only for 30 sec. These precautions were taken to ensure that the degree of Intein-mediated cleavage is not overestimated on SDS-PAGE [3,4] (for more information consult the manual of the IMPACT kit, New England BioLabs, E6901S) and that SDS-PAGE sample preparation does not lead to a significant Peptide Backbone Fragmentation [5] of the mRuby3 fluorescent protein of interest (for more information consult the manufacturer's description of commercially available purified Red Fluorescent Protein, Recombinant RFP, Cell Biolabs, STA-202). The samples were separated on 10% SDS-PAGE gels (Invitrogen NuPAGE 10% Bis-Tris gels, 10 wells and 1.0 mm thickness), run at 200 V for 1 hr in 1X MOPS SDS running buffer (Invitrogen Novex 20X NuPAGE MOPS SDS Running Buffer). For a better separation of the TSGIT-gp2.5 cleaved products, the corresponding samples were also separated on 12% SDS-PAGE gels (Invitrogen NuPAGE 12% Bis-Tris gels, 10 wells and 1.0 mm thickness) that was run at 75 V for 2.5 hr in 1X MOPS SDS running buffer. The molecular weight ladder marker used for all SDS-PAGE gels was PageRuler Prestained Protein Ladder (Thermo Scientific, 26616).

The final yields and concentrations of pure mRuby3, mRuby3-BioP and gp2.5 were determined by A_{280} measurements using NanoDrop spectrophotometer (Thermo Scientific) [1,2]. The extinction coefficient for both mRuby3 and mRuby3-BioP was considered 27390 $M^{-1}cm^{-1}$. The extinction coefficient for gp2.5 as monomer was considered 32890 $M^{-1}cm^{-1}$. The reported values of protein amount and concentration throughout the manuscript were the average and standard deviation generated by six repeated measurements of A_{280}. SDS-PAGE gels were imaged using the iBright CL1000 system (Invitrogen) and quantification was performed using the built-in option for gel analysis of the ImageJ software.

For the samples exhibiting less than 90% purity for mRuby3 or less than 98% purity for gp2.5, the protein yields were estimated by using the Pierce BCA protein assay kit (Thermo Scientific) with a Bovine Serum Albumin (BSA) standard, similarly to the protocol previously described in [1], but under denaturing conditions (5% SDS, 1 mM DTT and heating up to 95^{o}C for 10 min) to eliminate mRuby's natural absorbance that can interfere with the A_{562} measurement. For gp2.5 samples, the measurement was performed identically as described in [1], without the requirement of denaturing conditions. Six serial dilutions of the samples of interest were prepared in denaturing buffer and mixed with working reagent as per manufacturer's instructions. For A_{562} absorbance measurements, the resulting test samples were placed in a clear bottom 96-well microplate (Corning) and absorbance was measured using a xMark Microplate Spectrophotometer (Bio-Rad)

set to 562 nm. The BCA standard curve was fit to a linear dependence of A_{562} versus BSA concentration. The A_{562} measurements of the unknown samples were converted to protein concentrations by using this standard curve and by taking into account the dilution factor. All the reported protein yields and errors represent the average and the standard deviation of six measurements.

2. Steady-state fluorescence measurements

Steady-state fluorescence measurements for the mRuby3/ NeutrAvidin^{DyLight650} system were conducted at room temperature using Fluoromax-4 (HORIBA Jobin Yvon). All emission spectra were measured in storage buffer. In all cases, excitation was set to 520 nm and emission spectra were collected between 530 and 750 nm. Both excitation and emission slit widths were set to 5 nm. Measurements were recorded with an integration time of 0.2 sec. The emission spectra were corrected by subtracting the background emission of a blank solution comprised of storage buffer. The reported spectra are the average of three independent replicates. The spectra were then corrected and normalized as described below. NeutrAvidin^{DyLight650} was purchased from Thermo Fisher Scientific.

3. Correction and normalization of emission spectra

Steady state emission spectra for the mRuby3/ NeutrAvidin^{DyLigth650} were collected as described above. For all emission spectra, the excitation wavelength was fixed to $\lambda_{ex} = 520\,nm$. The concentrations of biotin-labeled mRuby3 and unlabeled mRuby3 were both fixed to 50 nM. Emission spectra of various concentrations of NeutrAvidin^{DyLigth650} were collected and corrected by blank subtraction. The resulting set of spectra is denoted as I. Emission spectra of various concentrations of NeutrAvidin^{DyLigth650} in the presence of biotin-labeled mRuby3 and unlabeled mRuby3 were collected and corrected by blank subtraction. The resulted sets of spectra are denoted as $\dot{I}_{bio}$ and $\dot{I}_{unlabled}$, respectively. Including their full dependence, the sets of emission spectra can be written as:

$$\begin{cases} I = I(\lambda_{ex} = 520\,nm; \lambda_{em}, c) \\ \dot{I}_{bio} = \dot{I}_{bio}(\lambda_{ex} = 520\,nm; \lambda_{em}, c) \\ \dot{I}_{unlabled} = \dot{I}_{unlabled}(\lambda_{ex} = 520\,nm; \lambda_{em}, c) \end{cases} , \quad (1)$$

where λ_{em} is the current emission wavelength and c is the current concentration of NeutrAvidin^{DyLigth650}. The contribution of emission of NeutrAvidin^{DyLigth650} at a given wavelength upon direct excitation at 520 nm, i.e., $I(\lambda_{ex} = 520\,nm; \lambda_{em}, c)$ was then subtracted from each spectrum. Simultaneously, the spectra were normalized to a total area of 1 A.U. by integration. The resulting corrected and normalized emission spectra of biotin-labeled mRuby3 and unlabeled mRuby3, in the presence of various concentration of NeutrAvidin^{DyLigth650}, are given by:

$$\begin{cases} \ddot{I}_{bio}(\lambda_{ex} = 520\ nm; \lambda_{em}, c) = \dfrac{\dot{I}_{bio}(\lambda_{em}, c) - I(\lambda_{em}, c)}{\int\left[\dot{I}_{bio}(\lambda_{em}, c) - I(\lambda_{em}, c)\right] d\lambda_{em}} \\ \ddot{I}_{unlabeled}(\lambda_{ex} = 520\ n; \lambda_{em}, c) = \dfrac{\dot{I}_{unlabeled}(\lambda_{em}, c) - I(\lambda_{em}, c)}{\int\left[\dot{I}_{unlabeled}(\lambda_{em}, c) - I(\lambda_{em}, c)\right] d\lambda_{em}} \end{cases} , \quad (2)$$

where integration is performed over the whole collected emission spectrum. The explicit dependence on the fixed excitation wavelength was omitted for the R.H.S. terms for simplicity. For any given NeutrAvidin^{DyLigth650} concentration, the fluorescence emission enhancement at $\lambda_{em} = 673\ nm$ (i.e., the position of the emission maximum of NeutrAvidin^{DyLigth650}) was calculated for biotin-labeled mRuby3 and unlabeled mRuby3 relative to their emission spectrum in the absence of NeutrAvidin^{DyLigth650} [i.e., $\ddot{I}_{bio}(\lambda_{ex} = 520\ nm, \lambda_{em} = 673\ nm; 0)$ and $\ddot{I}_{unlabeled}(\lambda_{ex} = 520\ nm, \lambda_{em} = 673\ nm; 0)$ respectively] as:

$$\begin{cases} \Delta\ddot{I}_{bio}(\lambda_{ex} = 520\ nm, \lambda_{em} = 673\ nm; c) = \ddot{I}_{bio}(c) - \ddot{I}_{bio}(0) \\ \Delta\ddot{I}_{unlabeled}(\lambda_{ex} = 520\ nm, \lambda_{em} = 673\ nm; c) = \ddot{I}_{unlabeled}(c) - \ddot{I}_{unlabeled}(0) \end{cases} , \quad (3)$$

where the explicit dependence on the fixed excitation and emission wavelengths was omitted for the R.H.S. terms for simplicity. In the case of biotin-labeled mRuby3, where significant enhancement was observed, the dependence of fluorescence enhancement in emission at 673 nm upon excitation at 520 nm, as a function of NeutrAvidin^{DyLigth650} concentration [denoted as $\Delta\ddot{I}_{bio}(c)$], was fitted to a Hill-type dependence, similar to the one described in [6], as:

$$\Delta\ddot{I}_{bio}(c) = \Delta\ddot{I}_{max} \times \frac{c^n}{{K_{1/2}}^n + c^n} \quad , \quad (4)$$

where $K_{1/2}$ represents the monomeric concentration of NeutrAvidin^{DyLigth650} at which half of the maximum emission enhancement ($\Delta\ddot{I}_{max}$) is produced and n represents the Hill coefficient.

4. Time-resolved fluorescence measurements

Time-resolved fluorescence lifetime measurements were carried out using QuantaMaster 800 spectrofluorometer (Photon Technology International Inc.) equipped with a Fianuim supercontinuum fiber laser source (Fianium, Southampton, U.K.) operating at 20 MHz repetition rate as described previously [7,8]. Arrival time of each photon was measured with a Becker-Hickl SPC-130 time-correlated single photon counting module (Becker-Hickl GmbH, Berlin, Germany). Measurements were collected under magic angle (54.7°) conditions and photons were counted using time to amplitude converter (TAC). To reduce the collection of scattered light, a longpass filter (550 nm) was placed at the emission side. In all measurements, 10,000 counts were acquired.

The instrument response function (IRF) was estimated using a Ludox colloidal silica suspension dissolved in water. Measurements were recorded at room temperature in gp2.5 binding buffer [50 mM HEPES-KOH pH (7.5), 50 mM KCl, 10 mM MgCl2, 1 mM DTT, 5% Glycerol and 0.1 mg/mL BSA]. The samples were excited at 532 nm and emission was collected at 565 nm with 5 nm slit width for both the excitation and emission. Cy3-labeled ssDNA was kept at a limiting concentration of 50 nM. Increasing concentrations of gp2.5 were then added to the Cy3-ssDNA-containing samples. The fluorophore lifetime decays were then obtained using FluoFit software package (PicoQuant) applying the IRF and fitted to two-exponential decays. The best fit was chosen based on reduced chi-square and randomness of the residuals. The final lifetimes at each gp2.5 concentration represent the mean of amplitude-averaged lifetimes [9] of three independent replicates. The increase in Cy3 fluorescence lifetime upon gp2.5 binding at various concentrations is reported as a difference in ns compared to the fluorescence lifetime of the Cy3-labeled oligo in the absence of protein. The resulting binding isotherms at various concentrations of gp2.5 (c) versus the increase in Cy3 fluorescence lifetime were fitted to Hill-type dependencies similar to the one presented in Eq.(4) as:

$$\Delta\tau(c) = \Delta\tau_{max} \times \frac{c^n}{{K_d}^n + c^n} \quad , \quad (5)$$

where K_d represents the monomeric concentration of gp2.5 at which half of the maximum fluorescence lifetime enhancement ($\Delta\tau_{max}$) is produced and n represents the Hill coefficient.

5. Size-exclusion chromatography analysis

For size-exclusion chromatography analysis, a 120 ml Superdex 16/600 75 pg (GE Healthcare) column and a 120 ml Superdex 16/600 200 pg (GE Healthcare) were pre-equilibrated with analysis buffer [50 mM HEPES pH (8), 250 mM NaCl, 0.1 mM TCEP and 5% Glycerol]. The Superdex 75 pg column was calibrated with Conalbumin, Ovalbumin, Carbonic anhydrase and Ribonuclease A provided by a low molecular weight (LMW) calibration kit (GE Healthcare). The Superdex 200 pg column was calibrated with Ferritin, Aldolase, Conalbumin and Ovalbumin provided by a high molecular weight (HMW) calibration kit (GE Healthcare).

Cleaved gp2.5 was analyzed using the Superdex 75 pg column, while TSGIT-gp2.5 uncleaved fusion was analyzed using the Superdex 200 pg column. Size-exclusion chromatography was performed using an FPLC system (ÄKTA, GE Healthcare) equipped with an absorbance module to continuously monitor the protein elution. Column run and elution were performed in analysis buffer at a flow-rate of 1 ml/min. To estimate the peak positions, the experimental A_{280} versus retention volume dependencies were fitted to a smoothing spline with a smoothing parameter of 0.98. The positions of the elution peaks were then determined by the points were the first derivative of the fitted splines intercept the x-axis. Prior to fitting, baselines were corrected for each of the four chromatograms using the build-in function of the Unicorn software (GE Healthcare).

Retention volumes for each peak were then converted to proportion of resin pores available to the respective molecule (K_{av}) as previously described [10].

6. Expression and purification of His-tagged Ulp1 SUMO protease

The gene encoding Trx-SUMO-Ulp1-His as a fusion protein was custom synthesized by IDT as gBlock and cloned into a pRSF-1b plasmid by Gibson assembly. This plasmid is denoted as pTS-Ulp1. This expression plasmid was transformed into *E. coli* strain BL21 (DE3) competent cells (Novagen) and colonies were selected on LB-agar plates containing 50 µg/ml Kanamycin. Ulp1 was overproduced by growing the transformed cells in 10 liters of 2xYT media (Teknova) supplemented with the same concentration of Kanamycin. Cells were grown at 37°C to an OD_{600} of 1.0 and then protein expression was induced by the addition of 0.5 mM isopropyl β-D-thiogalactopyranoside (IPTG) and incubated further for 6 hr at 37°C. Cells were collected by centrifugation at 5,500*g* for 10 min and re-suspended in lysis buffer [20 mM Tris pH (8), 300 mM NaCl, 20 mM Imidazole, 5 mM β-Mercaptoethanol, 10% Glycerol and one EDTA free protease inhibitor cocktail tablet per 50 ml (Roche, UK)]. All further steps were performed at 4°C.

Cells were lysed enzymatically by adding 2 mg/ml lysozyme and mechanically by sonication using the same cycle conditions describe in the main Materials and Methods section. Cell debris was removed by centrifugation (22,040*g*, 45 min) and the clear supernatant was directly loaded onto a custom-assembled 30 ml His-affinity column filled with Ni-NTA Superflow resin (QIAGEN) pre-equilibrated with binding buffer [20 mM Tris pH (8), 300 mM NaCl, 20 mM Imidazole, 5 mM β-Mercaptoethanol and 10% Glycerol]. We have previously noticed that the use of HisTrap HP 5 ml affinity columns (GE Healthcare) rapidly saturated with Ulp1 and a high amount of protein was lost in the flow-through fraction. Therefore, we increased the volume of the column to the 30 ml custom-assembled one described above. The 30 ml column was then washed with 10 column volumes of binding buffer followed by gradient elution with 10 column volumes of elution buffer [50 mM Tris pH (8), 300 mM NaCl, 500 mM Imidazole, 5 mM β-Mercaptoethanol and 10% Glycerol]. The peak fractions were pooled and dialyzed overnight in a dialysis buffer [25 mM Tris pH (7.5), 150 mM NaCl, 1 mM DTT and 50% Glycerol], flash frozen and stored at -80°C.

During protein expression Trx-SUMO as an N-terminal fusion tag provides increased expression and solubility levels. Ulp1 cleaves itself from the fusion, while remaining attached to its C-terminal His-tag which is then used for purification. The Trx-SUMO cleaved tag will pass as flow-through in the His affinity column. The design of self-cleaving Ulp1 in fusion with N-terminal SUMO is inspired from [11]. Ulp1 SUMO protease cleaved itself from the fusion *in vivo* prior to cell lysis. The typical yield of this method generates 50–100 mg of His-tagged Ulp1 from a 10 liters culture. The custom-purified Ulp1 SUMO protease purity is shown in Figure S1. Under the mentioned storage conditions the protein can also be stored at -20°C for several months while retaining activity.

7. Single-molecule flow-stretching bead assay

The primed single-stranded DNA (ssDNA) substrate was generated by annealing circular M13mp18 ssDNA (New England BioLabs) to 100-fold excess of the 5'-Digoxigenin-CTAGAGGATCCCCGGGTACCGAGCTCGAATTCGTAATCA-Biotin-TGGTCATAGCTGTTTCCTGTGTG-3' primer (Integrated DNA Technologies), which contains two orthogonal attachment modifications and which upon hybridization to M13mp18 ssDNA generates an EcoR1 restriction site. The annealed DNA was linearized with EcoRI (New England BioLabs), and the reaction was stopped by EcoRI inactivation at 65°C for 20 min. Excess unannealed primers and heat-inactivated EcoRI were removed by using a QIAquick PCR purification kit (QIAGEN). The final concentration of the DNA was quantified by using UV-visible absorption spectroscopy at 260 nm with an extinction coefficient of 91801.1 $mM^{-1}cm^{-1}$.

Single-molecule experiments were performed at room temperature in a custom-built microfluidic flow cell as described previously [12,13]. Briefly, bacteriophage T7 ssDNA binding protein gp2.5 was introduced into the flow cell at concentration of 2 μM in T7 reaction buffer [40 mM Tris-HCl pH (7.5), 50 mM KGlu, 10 mM $MgCl_2$, 10 mM DTT, and 0.1 mg/mL BSA]. The ssDNA stretching reaction was performed under continuous presence of protein in solution. Data acquisition and processing methods were identical to the previously described ones [14-16]. The centroid position of the DNA-attached beads during each acquisition time point (500 ms sampling rate) was determined by fitting a 2-dimensional Gaussian distribution to the bead intensities by using the DiaTrack particle-tracking software (SemaSopht). Residual instabilities in the flow were corrected by subtracting traces corresponding to tethered DNA molecules that were not enzymatically altered. Displacement of the beads due to the conversion of the template strand from free ssDNA to gp2.5-coated ssDNA was transformed into the numbers of equivalent dsDNA base pairs (bp) by using a conversion factor of 3.76 bp/nm. This conversion factor was derived from the difference in the length between ssDNA and dsDNA at the applied stretching force of ~2.6 pN [15-17].

General considerations on Intein-tag fusion cleavage

This discussion is based on the chemical cascade scheme presented in Figure 2b for thiol-cleavable contiguous mini-inteins. In the figure, Cys1 denotes the first cysteine residue of the Intein-tag. Briefly, the reaction is initiated by the peptide bond rearrangement between the last amino acid of the protein of interest and Cys1 of Intein-tag via N→S acyl shift in the precursor fusion protein. The resulting linear thioester intermediate can be attacked by the sulfhydryl group of a thiol reagent such as Dithiothreitol (DTT), β-Mercaptoethanol (2-ME), 2-Mercaptoethanesulfonic acid (2-MESNA), Thiophenol, or even free cysteine amino acid which results in cleavage of the Intein-tag and therefore of the C-terminal fusion tag from the protein of interest via thiol-thioester exchange. The resulting protein of interest has an activated thioester C-terminus which is required for a subsequent IPL reaction. If IPL is not of interest and the excess free thiol reagent is removed, at basic pH, water can act as a nucleophile resulting in the hydrolysis of the activated thioester and

formation of a native peptide carboxyl C-terminus. The reactions are driven forward (down in the scheme presented in Figure 2b) toward the cleaved product by basic pH, presence of conserved Intein's B-block residues [18-20] (for *Mxe* GyrA amino acids TANH) and large excess of free thiol reagent (Figure 2c). In this mechanism, basic pH aids the deprotonation of the sulfhydryl groups to increase their nucleophilic attack power [21], the conserved Intein's residues (especially H75) promote the initial N→S acyl shift [18-20], while excess thiol reagents provide the power for the thiol-thioester exchange and prevent the re-ligation of the cleaved Intein-tag to the target protein. For the purification of oxidation-sensitive proteins, sulfhydryl-free reducing agents (Figure 2d) can be used in buffer composition to maintain the cysteines of the protein of interest in their reduced form and prevent aggregation prior to the addition of the thiol reagent used for Intein-tag cleavage. If a thiol reducing agent would be included from the beginning of the protein purification, it can induce undesired premature cleavage of the fusion. In Figure 2e, which shows the schematic representation of the cascade of chemical reactions of IPL, 2-MESNA is illustrated in its deprotonated form (sulfonate) as the reaction occurs at pH 8.5 above the pKa of its sulfo group. Finally, an S→N acyl shift component reaction of IPL ensures that the protein-peptide fusion is linked through a native peptide bond. The peptide can contain various modifications such as specific reactive groups or attachment moieties. The IPL reaction is typically driven forward by basic pH as described above and a high excess of peptide over target protein.

General considerations on SUMO-tag fusion design

The sequence encoding the gene of interest can be assembled together with the TSGIT N- and C-terminal fusion tags by Gibson assembly cloning between the SUMO-tag and Intein-tag regions (Figure 2a) into any desired destination plasmid. As most fusion systems, TSGIT does not require the inserted gene of interest to contain the N-terminal methionine since it already contains an initiator methionine at the N-terminus of the fusion. In *E. coli*, only one methionine aminopeptidase (MAP) exists [22]. In general, removal of the initiator methionine by MAPs is based on the size of the residue immediately adjacent to the initiator methionine [22]. If the initiator methionine-adjacent amino acid has a radius of gyration of 1.29 Å or less then the initiation methionine will be cleaved by MAP. In fusion proteins, the N-terminal methionine encoded by the ATG codon of the inserted gene cannot be accessed by MAPs since it becomes a regular internal methionine rather than initiator methionine. If not manually removed, upon *in vitro* cleavage of the SUMO-tag by Ulp1, this methionine would be present at the N-terminus of the target protein and MAPs would not be present to ensure its removal. Failure to remove the initiator methionine can lead to dramatic decrease in protein activity. In general, the principle presented in [22] should be consulted to decide whether the first methionine should be included in the inserted gene. If the initiator methionine is removed, particular care should be taken in cases where the resulting cleaved protein of interest would have an N-terminal cysteine, as this residue can be attacked by the activated thioester C-terminus generated by Intein-tag cleavage, resulting in undesired circular protein forms and/ or head-to-tail protein multimers.

Supplementary Figures and Tables

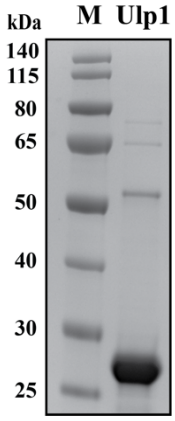

Figure S1. **Custom-purified His-tagged Ulp1 SUMO protease.** Image of a 10% SDS-PAGE gel showing the purity of the purified Ulp1. The marker (M) is PageRuler Prestained Protein Ladder.

mRuby3 insert

```
  1 Val Ser Lys Gly Glu Glu Leu Ile Lys Glu Asn Met Arg Met Lys Val Val Met Glu Gly Ser Val Asn Gly His Gln Phe Lys Cys Thr Gly Glu Gly Glu Gly Arg Pro Tyr Glu Gly  40
1036 GTA AGT AAA GGG GAG GAG CTT ATC AAA GAA AAT ATG CGC ATG AAG GTG GTC ATG GAA GGG TCA GTG AAC GGT CAC CAG TTC AAA TGT ACC GGT GAA GGC GAA GGC CGT CCG TAC GAA GGC 1155

 41 Val Gln Thr Met Arg Ile Lys Val Ile Glu Gly Gly Pro Leu Pro Phe Ala Phe Asp Ile Leu Ala Thr Ser Phe Met Tyr Gly Ser Arg Thr Phe Ile Lys Tyr Pro Ala Asp Ile Pro  80
1156 GTA CAG ACC ATG CGC ATT AAA GTT ATT GAG GGA GGA CCG TTA CCC TTC GCA TTT GAC ATC CTT GCC ACA TCC TTC ATG TAC GGT TCA CGC ACC TTC ATT AAG TAT CCG GCG GAC ATT CCA 1275

 81 Asp Phe Phe Lys Gln Ser Phe Pro Glu Gly Phe Thr Trp Glu Arg Val Thr Arg Tyr Glu Asp Gly Gly Val Val Thr Val Thr Gln Asp Thr Ser Leu Glu Asp Gly Glu Leu Val Tyr  120
1276 GAC TTT TTC AAA CAA TCT TTC CCC GAA GGC TTT ACC TGG GAG CGC GTT ACT CGC TAC GAA GAC GGC GGG GTT GTT ACT GTG ACG CAG GAC ACT TCT TTG GAA GAC GGT GAA TTA GTT TAC 1395

121 Asn Val Lys Val Arg Gly Val Asn Phe Pro Ser Asn Gly Pro Val Met Gln Lys Lys Thr Lys Gly Trp Glu Pro Asn Thr Glu Met Met Tyr Pro Ala Asp Gly Gly Leu Arg Gly Tyr  160
1396 AAC GTA AAA GTG CGT GGC GTG AAC TTC CCA AGT AAC GGT CCG GTG ATG CAA AAG AAG ACC AAG GGA TGG GAA CCA AAT ACG GAG ATG ATG TAC CCG GCA GAC GGC GGA CTG CGT GGA TAT 1515

161 Thr Asp Ile Ala Leu Lys Val Asp Gly Gly Gly His Leu His Cys Asn Phe Val Thr Thr Tyr Arg Ser Lys Lys Thr Val Gly Asn Ile Lys Met Pro Gly Val His Ala Val Asp His  200
1516 ACG GAC ATC GCA TTA AAG GTT GAC GGC GGG GGC CAC TTA CAT TGC AAT TTT GTA ACT ACG TAC CGT TCG AAG AAG ACG GTT GGA AAC ATC AAG ATG CCC GGA GTC CAC GCT GTG GAC CAT 1635

201 Arg Leu Glu Arg Ile Glu Glu Ser Asp Asn Glu Thr Tyr Val Val Gln Arg Glu Val Ala Val Ala Lys Tyr Ser Asn Leu Gly Gly Gly Met Asp Glu Leu Tyr Lys  236
1636 CGC CTT GAG CGC ATC GAG GAG TCA GAT AAC GAA ACC TAT GTT GTT CAA CGT GAG GTA GCT GTT GCA AAA TAC AGC AAT CTG GGG GGG GGT ATG GAT GAG TTG TAT AAG 1743
```

Figure S2. **Amino acid sequence of the mRuby3 insert.** The amino acids sequence and the corresponding nucleotide sequence are presented in FASTA format. The initial methionine is absent as it is not required by the TSGIT fusion expression system. This sequence was assembled together with the TSGIT N- and C-terminal fusion tags between the SUMO-tag and Intein-tag regions by Gibson assembly into a pRSF-1b plasmid.

gp2.5 insert

```
  1 Ala Lys Lys Ile Phe Thr Ser Ala Leu Gly Thr Ala Glu Pro Tyr Ala Tyr Ile Ala Lys Pro Asp Tyr Gly Asn Glu Glu Arg Gly Phe Gly Asn Pro Arg Gly Val Tyr Lys Val Asp  40
1036 GCT AAG AAG ATT TTC ACC TCT GCG CTG GGT ACC GCT GAA CCT TAC GCT TAC ATC GCC AAG CCG GAC TAC GGC AAC GAA GAG CGT GGC TTT GGG AAC CCT CGT GGT GTC TAT AAA GTT GAC 1155

 41 Leu Thr Ile Pro Asn Lys Asp Pro Arg Cys Gln Arg Met Val Asp Glu Ile Val Lys Cys His Glu Glu Ala Tyr Ala Ala Ala Val Glu Glu Tyr Glu Ala Asn Pro Pro Ala Val Ala  80
1156 CTG ACT ATT CCC AAC AAA GAC CCG CGC TGC CAG CGT ATG GTC GAT GAA ATC GTG AAG TGT CAC GAA GAG GCT TAT GCT GCT GCC GTT GAG GAA TAC GAA GCT AAT CCA CCT GCT GTA GCT 1275

 81 Arg Gly Lys Lys Pro Leu Lys Pro Tyr Glu Gly Asp Met Pro Phe Phe Asp Asn Gly Asp Gly Thr Thr Thr Phe Lys Phe Lys Cys Tyr Ala Ser Phe Gln Asp Lys Lys Thr Lys Glu  120
1276 CGT GGT AAG AAA CCG CTG AAA CCG TAT GAG GGT GAC ATG CCG TTC TTC GAT AAC GGT GAC GGT ACG ACT ACC TTT AAG TTC AAA TGC TAC GCG TCT TTC CAA GAC AAG AAG ACC AAA GAG 1395

121 Thr Lys His Ile Asn Leu Val Val Val Asp Ser Lys Gly Lys Lys Met Glu Asp Val Pro Ile Ile Gly Gly Gly Ser Lys Leu Lys Val Lys Tyr Ser Leu Val Pro Tyr Lys Trp Asn  160
1396 ACC AAG CAC ATC AAT CTG GTT GTG GTT GAC TCA AAA GGT AAG AAG ATG GAA GAC GTT CCG ATT ATC GGT GGT GGC TCT AAG CTG AAA GTT AAA TAT TCT CTG GTT CCA TAC AAG TGG AAC 1515

161 Thr Ala Val Gly Ala Ser Val Lys Leu Gln Leu Glu Ser Val Met Leu Val Glu Leu Ala Thr Phe Gly Gly Gly Glu Asp Asp Trp Ala Asp Glu Val Glu Glu Asn Gly Tyr Val Ala  200
1516 ACT GCT GTA GGT GCG AGC GTT AAG CTG CAA CTG GAA TCC GTG ATG CTG GTC GAA CTG GCT ACC TTT GGT GGC GGT GAA GAC GAT TGG GCT GAC GAA GTT GAA GAG AAC GGC TAT GTT GCC 1635

201 Ser Gly Ser Ala Lys Ala Ser Lys Pro Arg Asp Glu Glu Ser Trp Asp Glu Asp Asp Glu Glu Ser Glu Glu Ala Asp Glu Asp Gly Asp Phe  23
1636 TCT GGT TCT GCC AAA GCG AGC AAA CCA CGC GAC GAA GAA AGC TGG GAC GAA GAC GAC GAA GAG TCC GAG GAA GCA GAC GAA GAC GGA GAC TTC 1728
```

Figure S3. Amino acid sequence of the gp2.5 insert. The amino acids sequence and the corresponding nucleotide sequence are presented in FASTA format. The initial methionine is absent as it is not required by the TSGIT fusion expression system. This sequence was assembled together with the TSGIT N- and C-terminal fusion tags between the SUMO-tag and Intein-tag regions by Gibson assembly into a pRSF-1b plasmid.

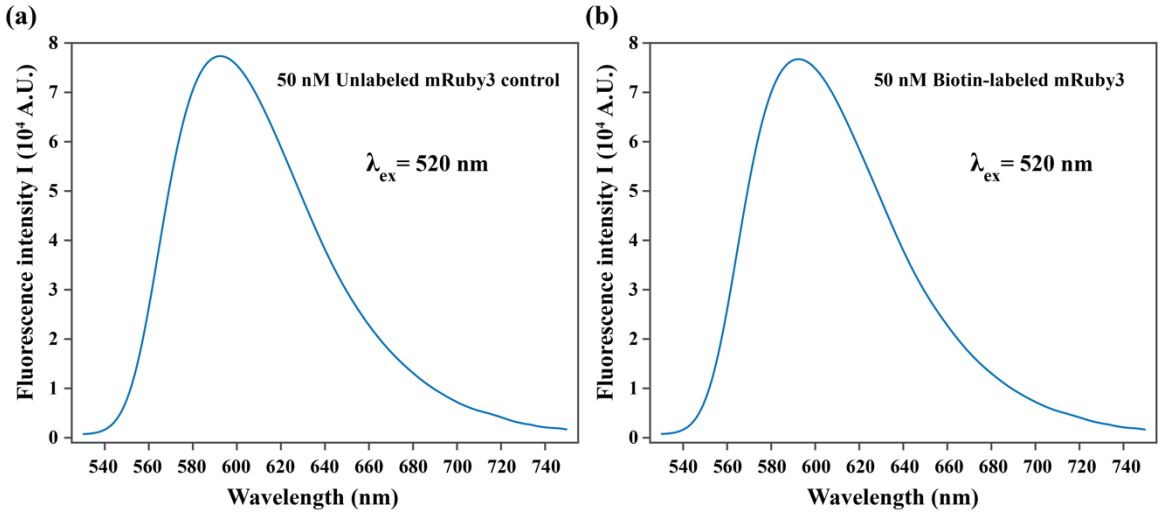

Figure S4. Fluorescence functionality of the final purified proteins. The emission spectra of (a) 50 nM unlabeled mRuby3 and (b) 50 nM IPL biotin-labeled mRuby3. All data points represent the average of three independent acquisitions. Both spectra were collected between 530 and 750 nm upon excitation at 520 nm.

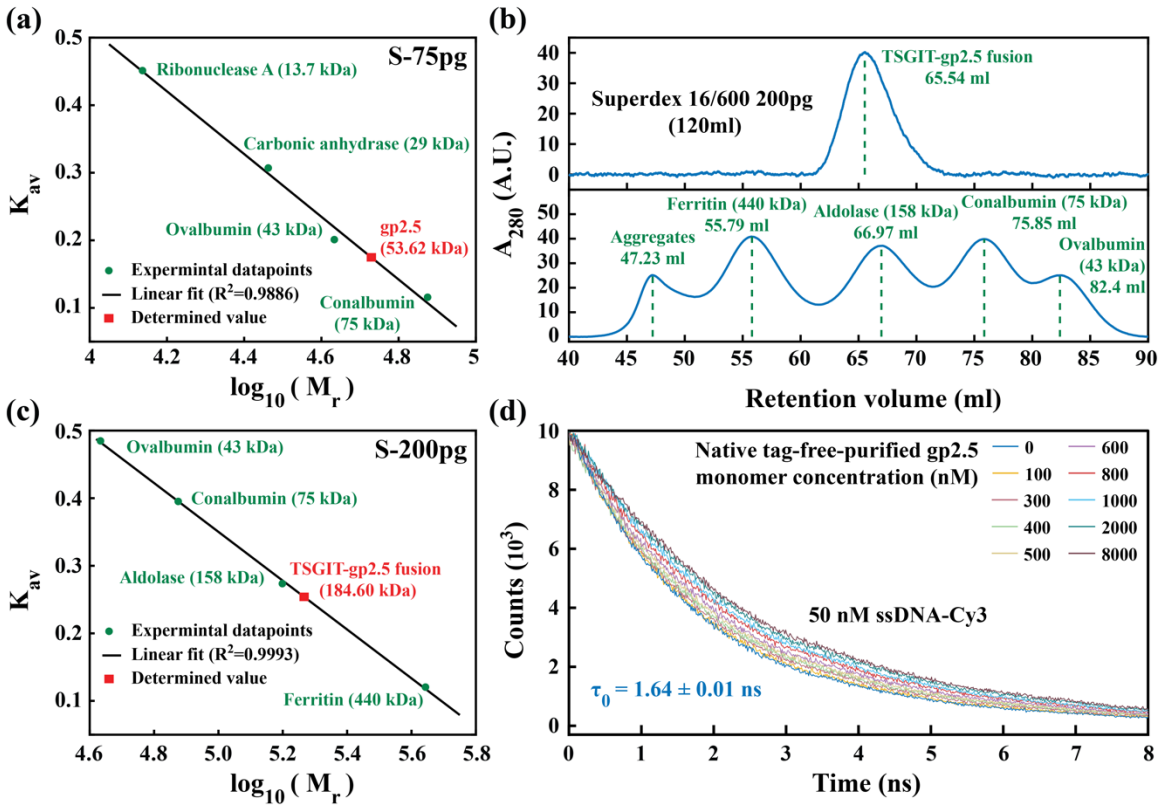

Figure S5. Characterization of the TSGIT-purified gp2.5. (a) A plot of the proportion of resin pores available to the molecule (K_{av}) versus the logarithm with base 10 of the relative molecular weight of the four molecular weight markers (green) obtained from the size-calibration of the Superdex 75 pg column shown in the bottom panel of Figure 5b. The experimental datapoints were

fitted with a linear dependence characterized by an intercept of 2.3702 and a slope of -0.4642. Using the position of the maximum of the elution peak of TSGIT-purified from the same size-exclusion column (top panel in Figure 5b), the size of the protein was estimated (red) from the linear fit by inverting the linear dependence equation. (b) Top: a chromatogram showing the elution of TSGIT-gp2.5 fusion (with both the N- and C- terminal fusion tags uncleaved, as obtained after the first Strep elution step; Lane 5 in Figure 5a) from the Superdex 200 pg size-exclusion column. Bottom: a chromatogram showing the elution of four different molecular weight markers from the Superdex 200 pg size-exclusion column. The position of the peaks (green dashed lines) were determined as described above. (c) A plot of the proportion of resin pores available to the molecule (K_{av}) versus the logarithm with base 10 of the relative molecular weight of the four molecular weight markers (green) obtained from the size-calibration of the Superdex 200 pg column shown in the bottom panel of Figure S5b. The experimental datapoints were fitted with a linear dependence characterized by an intercept of 2.1567 and a slope of -0.3613. Using the position of the maximum of the elution peak of TSGIT-gp2.5 uncleaved fusion from the same size-exclusion column (top panel in Figure S5b), the size of the protein was estimated (red) from the linear fit by inverting the linear dependence equation. (d) Examples of time-resolved fluorescence decays of the Cy3-labelled ssDNA obtained in the presence of various concentrations of native tag-free-purified gp2.5. The curves respect the color code presented in the inset legend.

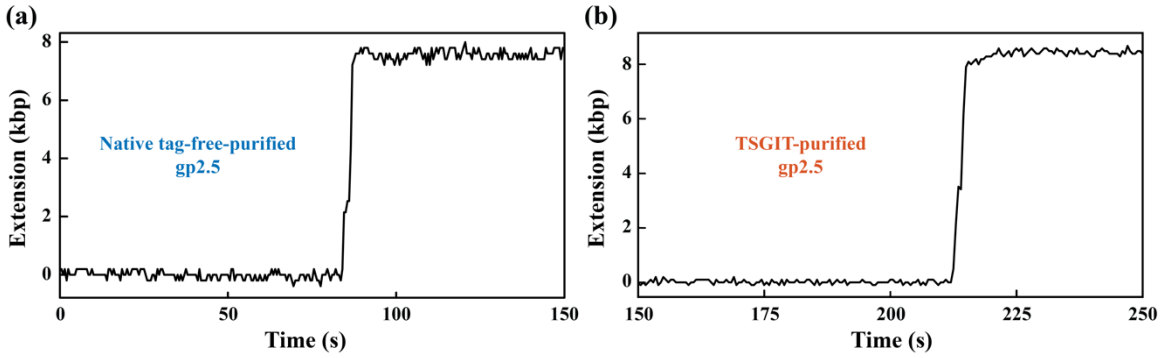

Figure S6. ssDNA stretching power of TSGIT-purified gp2.5. (a) Additional example of a single-molecule time-trace showing the stretching of the ssDNA-containing substrate upon injection of native tag-free-purified gp2.5. (b) Additional example of a single-molecule time-trace showing the stretching of the ssDNA-containing substrate upon injection of TSGIT-purified gp2.5.

Table S1. Yields and purities obtained during the purification of mRuby3 by TSGIT. For the samples exhibiting more than 90% purity, the proteins amounts were determined by A_{280} measurements using NanoDrop as described above. For the samples exhibiting less than 90% purity, the proteins amounts were determined by A_{562} measurements using Pierce BCA protein assay as described above. All the values in the table have their source indicated as: [a] purity of the band of interest, [b] purity of the sum of the bands resulted from cleavage, [c] as determined by Pierce BCA protein assay, [d] as determined by A_{280} measurements and [e] as determined for unlabeled mRuby3 fraction (half amount).

Purification step	Elution from first HisTrap column	Elution from first StrepTrap column	Cleavage via Ulp1 and 2-MESNA	Flow-through of the second HisTrap column	Flow-through of the second StrepTrap column	Elution from size-exclusion column
Purity (SDS-PAGE)	~60% [a]	~76% [a]	~86% [b] / ~23% [a]	~89% [b] / ~31% [a]	~92% [a]	>95% [a]
Total protein (mg)	25.9 ± 3.5 [c]	21.4 ± 2.1 [c]	18.6 ± 1.8 [c]	12.5 ± 2.3 [c]	2.7 ± 0.2 [d]	1.1 ± 0.2 [d, e]

Table S2. Yields and purities obtained during the purification of gp2.5 by TSGIT. For the samples exhibiting more than 98% purity, the proteins amounts were determined by A_{280} measurements using NanoDrop as described above. For the samples exhibiting less than 98% purity, the proteins amounts were determined by A_{562} measurements using Pierce BCA protein assay as described above. All the values in the table have their source indicated as: [a] purity of the band of interest, [b] purity of the sum of the bands resulted from cleavage, [c] as determined by Pierce BCA protein assay and [d] as determined by A_{280} measurements.

Purification step	Elution from first HisTrap column	Elution from first StrepTrap column	Cleavage via Ulp1 and DTT	Flow-through of the second HisTrap column	Flow-through of the second StrepTrap column	Elution from size-exclusion column
Purity (SDS-PAGE)	~67% [a]	~96% [a]	~97% [b] / ~30% [a]	~98% [b] / ~60% [a]	~98% [a]	>98% [a]
Total protein (mg)	49.1 ± 2.6 [c]	29.3 ± 1.7 [c]	28.9 ± 0.1 [c]	18.0 ± 0.6 [d]	10.1 ± 0.2 [d]	4.3 ± 0.1 [d]

References

1. Raducanu VS, Tehseen M, Shirbini A, Raducanu DV, Hamdan SM. Two chromatographic schemes for protein purification involving the biotin/avidin interaction under native conditions. J Chromatogr A 2020:461051.
2. Tehseen M, Raducanu VS, Rashid F, Shirbini A, Takahashi M, Hamdan SM. Proliferating cell nuclear antigen-agarose column: A tag-free and tag-dependent tool for protein purification affinity chromatography. J Chromatogr A 2019;1602:341-349.
3. Xu MQ, Evans TC, Jr. Purification of recombinant proteins from E. coli by engineered inteins. Methods Mol Biol 2003;205:43-68.
4. Burgess-Brown NA. Heterologous gene expression in E. coli : methods and protocols. New York, NY: Humana Press; 2017. xiv, 429 p.
5. Barondeau DP, Kassmann CJ, Tainer JA, Getzoff ED. Understanding GFP posttranslational chemistry: structures of designed variants that achieve backbone fragmentation, hydrolysis, and decarboxylation. J Am Chem Soc 2006;128(14):4685-93.
6. Leavesley SJ, Britain AL, Cichon LK, Nikolaev VO, Rich TC. Assessing FRET using spectral techniques. Cytometry A 2013;83(10):898-912.
7. Raducanu VS, Rashid F, Zaher MS, Li YY, Merzaban JS, Hamdan SM. A direct fluorescent signal transducer embedded in a DNA aptamer paves the way for versatile metal-ion detection. Sensors and Actuators B-Chemical 2020;304:127376.
8. Rashid F, Raducanu VS, Zaher MS, Tehseen M, Habuchi S, Hamdan SM. Initial state of DNA-Dye complex sets the stage for protein induced fluorescence modulation. Nat Commun 2019;10(1):2104.
9. Sillen A, Engelborghs Y. The correct use of "average" fluorescence parameters. Photochemistry and Photobiology 1998;67(5):475-486.
10. Kim YT, Tabor S, Bortner C, Griffith JD, Richardson CC. Purification and characterization of the bacteriophage T7 gene 2.5 protein. A single-stranded DNA-binding protein. J Biol Chem 1992;267(21):15022-31.
11. Kuo D, Nie M, Courey AJ. SUMO as a solubility tag and in vivo cleavage of SUMO fusion proteins with Ulp1. Methods Mol Biol 2014;1177:71-80.
12. Lee JB, Hite RK, Hamdan SM, Xie XS, Richardson CC, van Oijen AM. DNA primase acts as a molecular brake in DNA replication. Nature 2006;439(7076):621-4.
13. van Oijen AM, Blainey PC, Crampton DJ, Richardson CC, Ellenberger T, Xie XS. Single-molecule kinetics of lambda exonuclease reveal base dependence and dynamic disorder. Science 2003;301(5637):1235-8.
14. Jergic S, Horan NP, Elshenawy MM, Mason CE, Urathamakul T, Ozawa K, Robinson A, Goudsmits JM, Wang Y, Pan X and others. A direct proofreader-clamp interaction stabilizes the Pol III replicase in the polymerization mode. EMBO J 2013;32(9):1322-33.
15. Pandey M, Elshenawy MM, Jergic S, Takahashi M, Dixon NE, Hamdan SM, Patel SS. Two mechanisms coordinate replication termination by the Escherichia coli Tus-Ter complex. Nucleic Acids Res 2015;43(12):5924-35.
16. Hamdan SM, Loparo JJ, Takahashi M, Richardson CC, van Oijen AM. Dynamics of DNA replication loops reveal temporal control of lagging-strand synthesis. Nature 2009;457(7227):336-9.

17. Elshenawy MM, Jergic S, Xu ZQ, Sobhy MA, Takahashi M, Oakley AJ, Dixon NE, Hamdan SM. Replisome speed determines the efficiency of the Tus-Ter replication termination barrier. Nature 2015;525(7569):394-8.
18. Telenti A, Southworth M, Alcaide F, Daugelat S, Jacobs WR, Jr., Perler FB. The Mycobacterium xenopi GyrA protein splicing element: characterization of a minimal intein. J Bacteriol 1997;179(20):6378-82.
19. Du Z, Shemella PT, Liu Y, McCallum SA, Pereira B, Nayak SK, Belfort G, Belfort M, Wang C. Highly conserved histidine plays a dual catalytic role in protein splicing: a pKa shift mechanism. J Am Chem Soc 2009;131(32):11581-9.
20. Friedel K, Popp MA, Matern JCJ, Gazdag EM, Thiel IV, Volkmann G, Blankenfeldt W, Mootz HD. A functional interplay between intein and extein sequences in protein splicing compensates for the essential block B histidine. Chemical Science 2019;10(1):239-251.
21. Wall SB, Oh JY, Diers AR, Landar A. Oxidative modification of proteins: an emerging mechanism of cell signaling. Front Physiol 2012;3:369.
22. Wingfield PT. N-Terminal Methionine Processing. Curr Protoc Protein Sci 2017;88:6 14 1-6 14 3.