



A New Protein-Protein Interaction Sensor Based on Tripartite Split-GFP Association

Stéphanie Cabantous, Hau Nguyen, Jean-Denis Pedelacq, Faten Koraïchi, Anu Chaudhary, Kumkum Ganguly, Meghan Lockard, Gilles Favre, Thomas Terwilliger, Geoffrey Waldo

► To cite this version:

Stéphanie Cabantous, Hau Nguyen, Jean-Denis Pedelacq, Faten Koraïchi, Anu Chaudhary, et al.. A New Protein-Protein Interaction Sensor Based on Tripartite Split-GFP Association. Scientific Reports, 2013, 3 (1), 10.1038/srep02854 . hal-02378714

HAL Id: hal-02378714

<https://hal.science/hal-02378714>

Submitted on 24 Nov 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



OPEN

SUBJECT AREAS:

MOLECULAR
ENGINEERING
SENSORS AND PROBES
PROTEIN DESIGN
FLUORESCENT PROTEINSReceived
5 June 2013Accepted
16 September 2013Published
4 October 2013Correspondence and
requests for materials
should be addressed to
S.C. (cabantous.
stephanie@
claudiusregaud.fr) or
G.S.W. (waldo@lanl.
gov)

A New Protein-Protein Interaction Sensor Based on Tripartite Split-GFP Association

Stéphanie Cabantous¹, Hau B. Nguyen², Jean-Denis Pedelacq³, Faten Koraichi¹, Anu Chaudhary⁴, Kumkum Ganguly², Meghan A. Lockard⁵, Gilles Favre¹, Thomas C. Terwilliger² & Geoffrey S. Waldo²

¹INSERM UMR1037, Cancer Research Center of Toulouse, Université de Toulouse, Institut Claudius Regaud, F-31052 Toulouse, France, ²Bioscience Division, MS-M888, Los Alamos National Laboratory, Los Alamos, NM 87545, USA, ³CNRS, IPBS (Institut de Pharmacologie et de Biologie Structurale), F-31077 Toulouse, France; Université de Toulouse; UPS, IPBS, F-31077 Toulouse, France, ⁴Department of Microbiology, University of Washington, Seattle, WA 98195, USA, ⁵Rockefeller University, New York, NY 10065, USA.

Monitoring protein-protein interactions in living cells is key to unraveling their roles in numerous cellular processes and various diseases. Previously described split-GFP based sensors suffer from poor folding and/or self-assembly background fluorescence. Here, we have engineered a micro-tagging system to monitor protein-protein interactions *in vivo* and *in vitro*. The assay is based on tripartite association between two twenty amino-acids long GFP tags, GFP10 and GFP11, fused to interacting protein partners, and the complementary GFP1-9 detector. When proteins interact, GFP10 and GFP11 self-associate with GFP1-9 to reconstitute a functional GFP. Using coiled-coils and FRB/FKBP12 model systems we characterize the sensor *in vitro* and in *Escherichia coli*. We extend the studies to mammalian cells and examine the FK-506 inhibition of the rapamycin-induced association of FRB/FKBP12. The small size of these tags and their minimal effect on fusion protein behavior and solubility should enable new experiments for monitoring protein-protein association by fluorescence.

Signal transduction and gene expression often involves protein-protein interactions. Optical tracking of these interactions can help to illuminate regulatory mechanisms and identify aberrant processes in diseases. Fluorescent protein biosensors have been developed to measure protein complexes in living cells¹. For example, bioluminescence resonance energy transfer (FRET and BRET) enable dynamic observations of protein-protein interactions². More recent developments include protein-fragment complementation assays (PCA) that monitor protein-protein interactions by reconstitution of fragments of various enzymes split into two pieces (as fused tags on passenger proteins) including fragments of dihydrofolate reductase³, β -galactosidase⁴, β -lactamase⁵, and the firefly and *Gaussia* luciferases⁶. Though sensitive, these have the limitation that the observed products diffuse away from the protein interaction site. PCA based on split green fluorescent protein (GFP) and color variants⁷ is termed bimolecular fluorescence complementation (BiFC). BiFC relies on (i) interactions between bait and prey proteins that bring together two non-fluorescent split protein domains and (ii) subsequent co-folding into the β -barrel structure to form the chromophore⁸. Because assembly of the GFP fragments is irreversible, the stability of the BiFC enables integration, accumulation, and subsequent detection even of transient interactions and low affinity complexes^{9,10}. These remain attached to the interacting proteins, enabling tracking of the complex. However, improving existing BiFCs based on large and bulky fragments is challenging as increasing BiFC fragment solubility and folding can increase background signals from spontaneous assembly of the fluorescent protein fragments¹¹. For example, BiFC fragments¹² obtained by fragmentation of folding-reporter GFP (FR-GFP)¹³ and superfolder GFP (sfGFP)¹⁴ at permissive sites 156 and 173 were aggregation-prone and had high backgrounds from self-assembly (Supplementary Fig. S1). Using a smaller tag can reduce aggregation and folding interference. For example, one split GFP uses a very small 15 amino acid tag for quantification of soluble protein and tracking proteins *in vivo*^{15–17}. However the large size of the remaining piece and self-assembly of the fragments makes this system less suited for protein-protein interaction studies. We had previously developed a spontaneously assembling split GFP based on GFP 10–11 (residues 194–233, the tag) and GFP1-9 (residues 1–193, the detector) (Waldo lab, unpublished results). Here we describe the further engineering of this split GFP to yield an entirely new protein interaction assay based on the association of *three* fragments of the GFP: two short peptides GFP10 (residues 194–212) and GFP11 (residues 213–233) each tagged to one of the interacting partners, and a third large GFP1-9 (residues 1–193) detector fragment (Fig. 1). We present the characterization of the assay using attractive and repulsive pairs of charged coiled-coils peptides¹⁸ and the rapamycin mediated heterodimerization

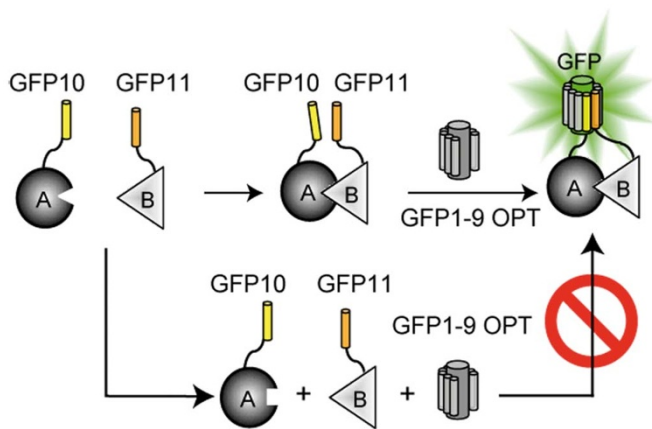


Figure 1 | Principle of the tripartite split-GFP complementation assay. β -strand 10 (GFP10) and β -strand 11 (GFP11) are fused to bait (A) and prey (B) proteins, respectively and the detector fragment GFP1-9 (β -strand 1-9) is added separately. When protein interaction occurs, GFP10 and GFP11 are tethered and then spontaneously associate with GFP1-9 fragment to form a full-length GFP. If proteins A and B do not interact, GFP10 and GFP11 are not tethered and entropy is too high to allow complementation with GFP1-9.

of FK506 binding protein (FKBP) and FKBP12-*rapamycin* binding domain (FRB)¹⁹. The assay correctly reports the localization of protein-protein complexes in mammalian cells with very low background fluorescence levels. The chief advantage over BiFC is the small size of the tags (ca. 20 amino acids), and the concomitant reduced interference with passenger protein folding. The tripartite split GFP assay may find broad utility in the visualization of protein complexes specially where bulkier BiFC fragments might impede localization or interfere with folding.

Results

Engineering a three-body split-GFP system for improved solubility and complementation. We previously identified self-assembling split-GFP fragments corresponding to the C-terminal β -hairpin GFP10-11 of sfGFP¹⁴ (residues 194-238) and the large fragment GFP1-9 (residues 1-193) (Fig. 2). In order to improve GFP 1-9 folding efficiency, we performed two rounds of directed evolution and obtained a new variant of GFP1-9 named GFP1-9 M1 that contained five additional mutations relative to sfGFP1-9 (Fig. 2, Supplementary Note). We converted this bimolecular pair into a three-body split-GFP in a stepwise manner. First, in attempt to destabilize the GFP10-11 β -hairpin self-assembly, we inserted a long flexible linker that included a cloning site between GFP10 and GFP11 and evolved the whole cassette for improved solubility and complementation efficiency with GFP1-9 M1 (See Methods). Next, we spaced out GFP10 and GFP11 further by inserting a partially soluble bait protein, hexulose phosphate synthase or HPS¹⁵ in the cloning site of the linker. HPS alone was ~60% soluble and was used in our earlier study as a bait protein to drive the evolution of more soluble split-GFP tags¹⁵ (Supplementary table 1 and Supplementary Note). In the final step, we evolved GFP1-9 M1 to recognize this “sandwich” bait topology. When complemented with the optimized GFP10 and GFP11 tags attached to the N and C termini of the HPS¹⁵, the resulting variant named GFP1-9 OPT, exhibited a 40-fold improvement in the rate of formation of fluorescence (Fig. 3a, black line) relative to GFP 1-9 M1 (Fig. 3a, gray line). Differences between the GFP1-9 M1 were less obvious when complemented with the GFP10-11 hairpin (displayed on permissive loop of a cherry fluorescent protein, Waldo lab, manuscript under review) (Fig. 3a, right panel), presumably due to the reduced entropy of the GFP 10-11 hairpin relative to the GFP10-HPS-GFP11 “sandwich” format. GFP1-9 OPT amino-acid sequence revealed the presence of four additional mutations relative to GFP1-9 M1 (Fig. 2). We then

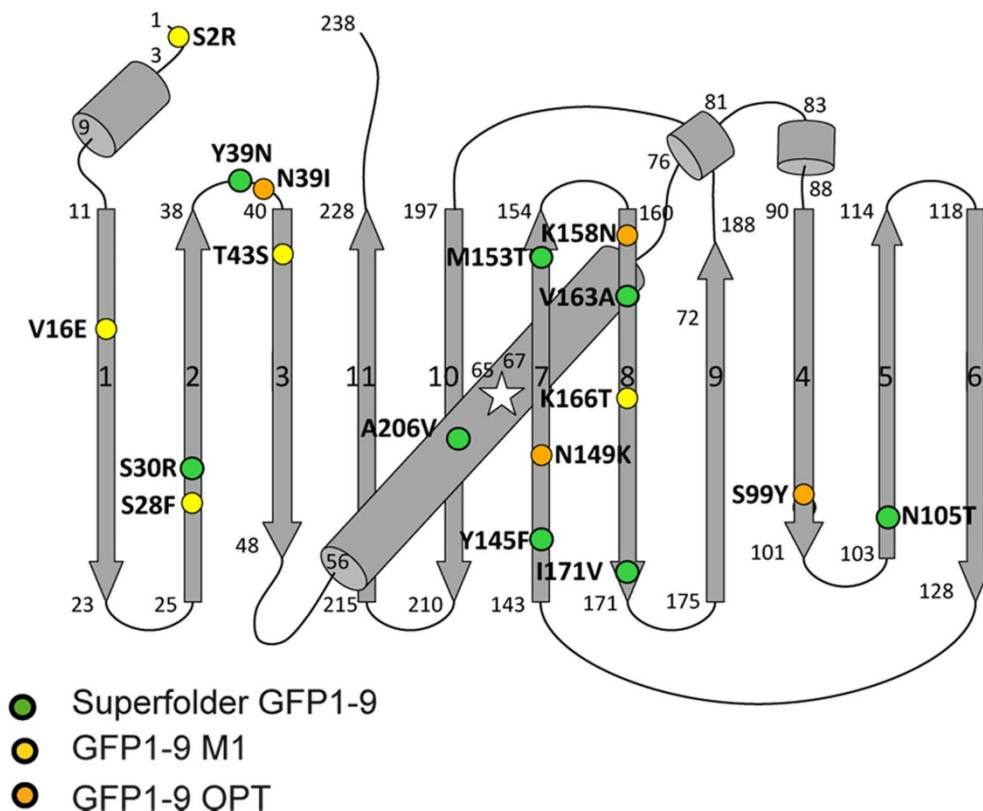


Figure 2 | Secondary structure diagram of GFP 1-9 variants with corresponding mutations. Superfolder GFP1-9 (green dots), GFP1-9 M1 (yellow), and GFP1-9 OPT (orange).

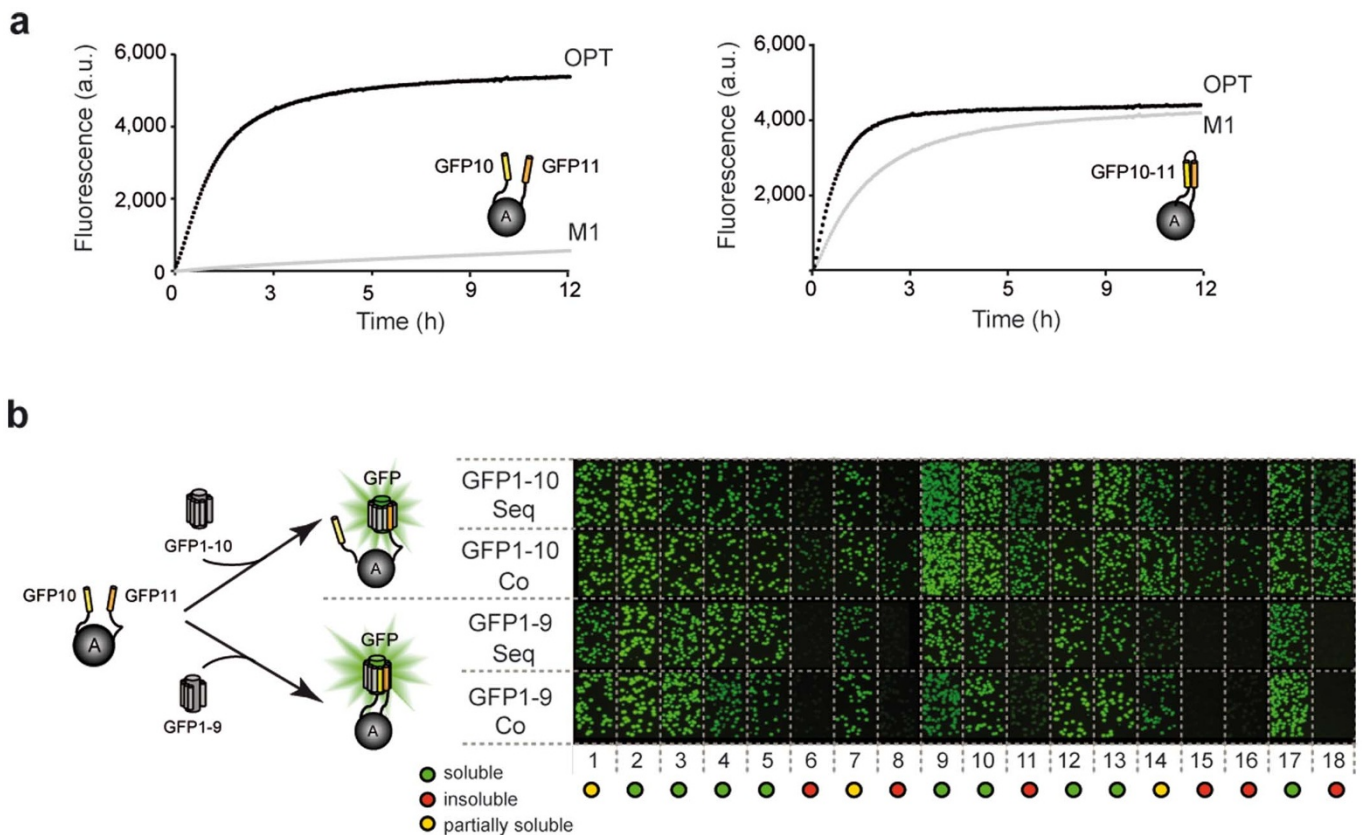


Figure 3 | Complementation of the tripartite split-GFP *in vitro* and *in vivo*. (a) Complementation curves of GFP1-9 M1 (gray line) and GFP1-9 OPT (black line) with equimolar GFP10-sulfite reductase (SR)-GFP11 fusion protein (left) or GFP10-11 hairpin domain displayed on a permissive loop of a superfolder red fluorescent protein (right) (Waldo lab, manuscript under review). (b) *In vivo* solubility screen of 18 *Pyrobaculum* test proteins¹⁵ expressed in a “sandwich” configuration with N-terminal GFP10 and C-terminal GFP11 (GFP10-A-GFP11, A = protein of interest) from pTET-GFP10/11 plasmid assayed with either GFP1-10 (top) or GFP1-9 (bottom) expressed from a pET vector in BL21(DE3) *E. coli*.¹⁵ Fluorescence pictures of *E. coli* colonies on plates after 1–1/2 h Anet induction followed by IPTG induction (sequential induction) (SEQ) or 3 h co-induction (CO). Legend indicates of the tagged proteins determined by SDS-PAGE (soluble (green), partially soluble (yellow), insoluble (red)).

examined whether optimized GFP10 and GFP11 would affect the solubility of fused proteins. We cloned set of eighteen *P. aerophilum* test proteins into a modified pTET vector as GFP10-POI-GFP11 “sandwiches” (POI = protein of interest) to assay solubility levels as previously described¹⁶. We had previously measured the solubility of these eighteen controls without tags and shown that the GFP11 M3 tag engineered in the earlier study did not perturb passenger solubility¹⁶. We first assayed the sandwich fusions using GFP1-10 to detect only the GFP11 tag (Fig. 3b). Co-expressing GFP1-10 and the sandwich resulted in bright colonies reflecting total protein expression (Fig. 3b, top). On the other hand, expressing the sandwich first, shutting off expression to allow the proteins to remain soluble or aggregate, then expressing the GFP1-10 detector (sequential induction) gave weak fluorescence for the protein expressed in insoluble form (proteins #6,8,15,16) as expected. This pattern was very similar to that previously seen for the GFP11 M3 tag study¹⁶ indicating that the GFP10-POI-GFP11 sandwich did not strongly perturb the solubility of the test proteins. Note that the version of GFP11 developed in this manuscript differs slightly from the GFP11 M3 tag described earlier¹⁶. Next we studied the complementation of the sandwich with GFP1-9 OPT (subsequently referred to as GFP1-9) (Fig. 3b, bottom). Interestingly, whether the sandwich and GFP1-9 were co-expressed or sequentially expressed, only the soluble proteins gave fluorescent colonies (Fig. 3b, bottom). When GFP1-10 is co-expressed, insoluble protein can be labeled (Fig. 3b, top), ostensibly because the available GFP1-10 can capture the GFP11 M3 tag early on, committing the fluorophore to

form regardless of the subsequent fate of the fusion. In contrast, even when co-expressed with GFP1-9, one or both tags of insoluble proteins must become inaccessible prior to productive binding with GFP1-9. This suggests the GFP 1-9 could be used to stringently assess protein solubility, without the concern of false-positives caused by capturing transiently soluble proteins as with GFP1-10.

Validating a protein-protein interaction sensor using coiled-coil heterodimerization. To test whether GFP1-9 is capable of detecting GFP10 and GFP11 fused to separate interacting protein molecules, we selected two pairs of coiled-coils developed by Hodges and coworkers¹⁸ based on their ability to direct preferential interaction between oppositely charged K1/E1 coiled-coils or repulsion for the negatively charged E1/E1 pair (Supplementary Fig. S2). This model system features rapid association of the K/E heterodimer and a lack of E coil homodimerization¹⁸. Pair-wise combinations of E1-GFP11 and GFP10-E1, or GFP10-K1 and E1-GFP11 were expressed from an anhydrotetracycline (Antet) inducible bicistronic pTET vector (Supplementary Fig. S3), and transformed into *E. coli* cells expressing the GFP1-9 from an isopropyl β-D-1-thiogalactopyranoside (IPTG) inducible pET vector (Fig. 4a). After 3 h *E. coli* GFP1-9 cells co-expressing GFP10-K1 and E1-GFP11 turned brightly fluorescent, thus demonstrating efficient K/E coiled-coil heterodimerization. In contrast, co-expression of GFP10-E1 and E1-GFP11 produced residual background fluorescence, comparable with mock induction after either Antet or IPTG induction

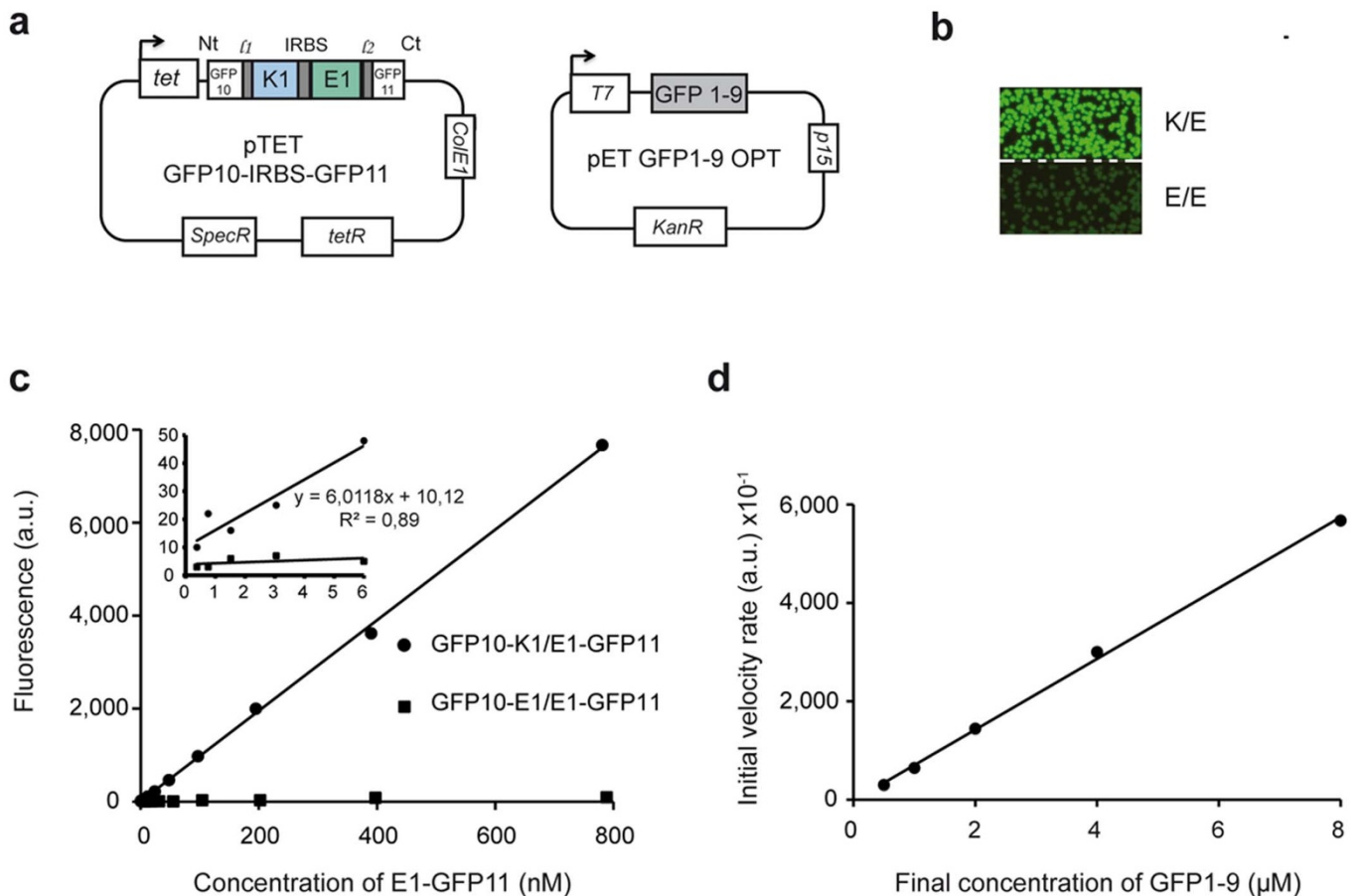


Figure 4 | Characterization of protein-protein interactions using coiled-coil heterodimerization. (a) Expression vectors for tripartite split-GFP interaction assays in *E. coli*: bicistronic pTET tetracycline inducible vector harboring GFP10 and GFP11 tags with cloning sites for test proteins; IPTG inducible pET vector (T7 promoter) for stable expression of GFP1-9 OPT. (b) Colony fluorescence on plates after Antet and IPTG co-induction driving expression of GFP10-K1/E1-GFP11 pair (K/E) (top) and GFP10-E1/E1-GFP11 (E1/E1) (bottom) with GFP1-9 detection fragment. (c) *In vitro* quantification of K1/E1 or E1/E1 interaction with increasing amount of E-GFP11 (0.4 nM up to 800 nM). Fluorescence values 1 h after initiating complementation with GFP1-9. (d) Initial velocity rates of complementation reactions with various concentrations of GFP1-9 detection reagent (0.5 μM up to 8 μM) in GFP10-K1/E1-GFP11 assay (1 μM of each tagged species).

in *E. coli* (Fig. 4b). These results suggest minimal GFP10 and GFP11 self-association in the absence of interacting fused proteins. Using the K1/E1 model, we also chose the GFP10 construct (residue 194–212) as the optimal ‘break point’ for detecting protein-protein interactions (Supplementary Fig. S4). To assess the sensitivity of the assay, GFP10-K1, GFP10-E1, and E1-GFP11 fusion proteins were individually expressed from pET vectors in *E. coli* and purified using immobilized metal affinity chromatography (Supplementary Fig. S5a). Quantified GFP10-K1 or GFP10-E1 were mixed with E1-GFP11 at varying concentrations, and supplemented with a molar excess of the GFP1-9 fragment (8 μM) to initiate rapid complementation and ensure that kinetic rate is independent of GFP1-9 concentration. Fluorescence of the GFP10-K1/E1-GFP11 interaction assay was linearly related to the concentration of E1-GFP11 substrate, as previously observed with self-assembling split-GFP assays¹⁵ (Figure 4c). The assay could detect K1/E1 coil interaction down to 1 nM which corresponds roughly to the dissociation constant of the heterodimer¹⁸. As expected, the E1/E1 pair interaction signal remained at a basal level of fluorescence even at very high concentrations (1 mM) (Fig. 4c). These results are completely consistent with a model in which tripartite split-GFP complementation occurs only when GFP10 and GFP11 are brought into close proximity by interacting protein partners, presumably by reduction in entropy. To study how different concentrations of the detecting reagent (GFP1-9) affected the

assay, two-fold successive dilutions of GFP1-9 were added to the protein mixtures containing GFP10-K1 and E1-GFP11 at the same final concentration (1 μM each). We observed a slower kinetic complementation rate as the reagent concentration decreased, completely consistent with simple bimolecular reaction kinetics (Fig. 4d). For example, the initial rate for appearance of fluorescence drops by a factor of two each time the concentration of GFP1-9 is halved. Although further studies are needed to understand the precise mechanism of assembly, these observations are consistent with a model in which the coils bring the smaller fragments into proximity prior to detection by GFP1-9, i.e., approximating a two-body system with the tethered GFP10 and GFP11 acting as a module interacting with GFP1-9.

Inducible FRB/FKBP interactions. To further test the specificity and dynamic range of the method, we used the well-studied FKBP12-FRB rapamycin inducible protein interaction¹⁹. We fused GFP10 to the rapamycin-binding domain of the mammalian target of rapamycin (mTOR) kinase (FRB) and GFP11 to the FK506-binding protein 12 (FKBP) and tested their association *in vitro* upon addition of recombinant GFP 1-9 protein (Fig. 5a). GFP10-FRB and FKBP-GFP11 fusion proteins were co-expressed from bicistronic pTET GFP10-IRBS-GFP11 vector (Supplementary Fig. S3). In earlier studies of interacting proteins in cell extracts using BiFC, misfolding and aggregation issues interfered with complementation unless the

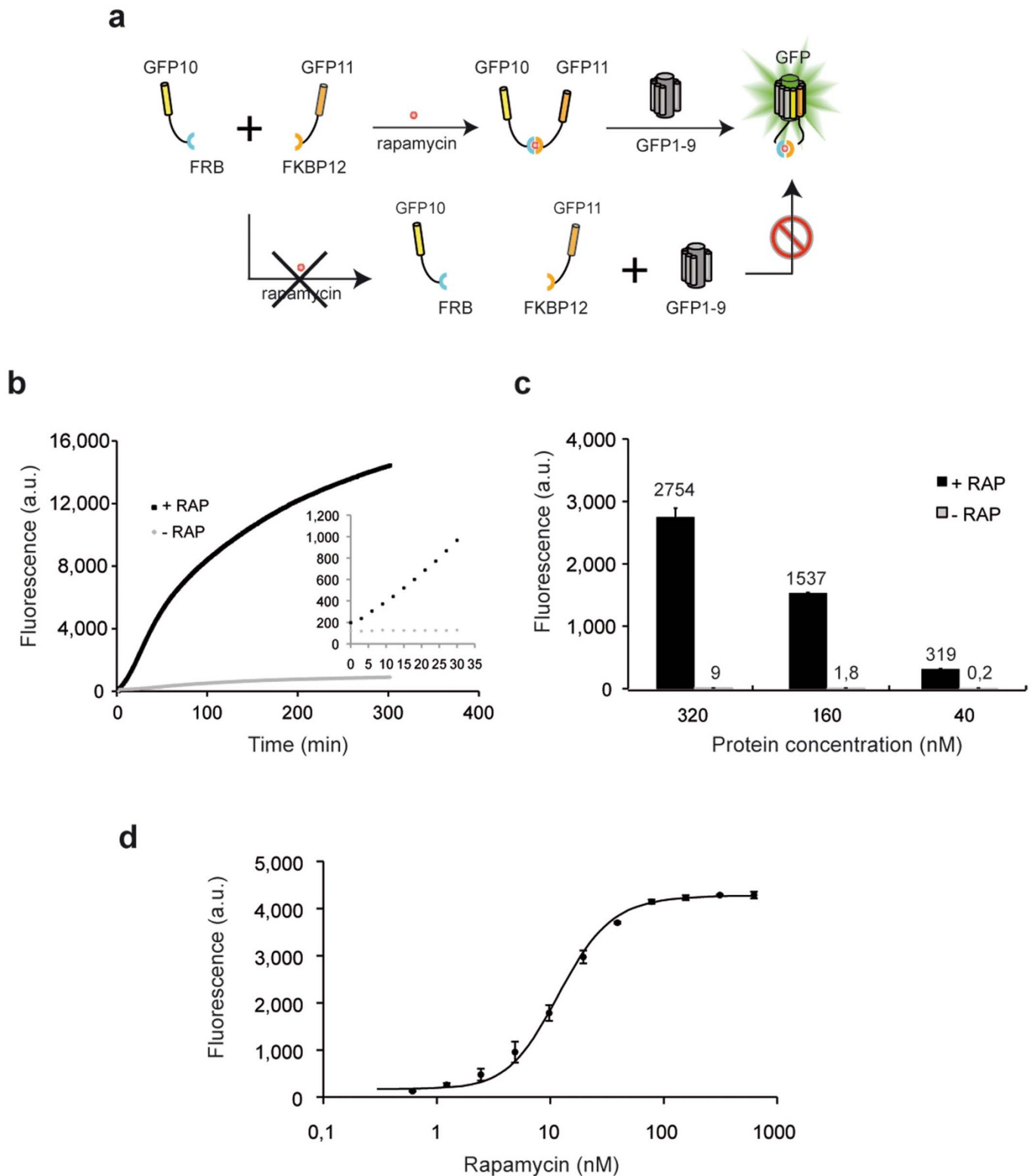


Figure 5 | Application of the tripartite split-GFP to study the rapamycin inducible FRB/FKBP interaction. (a) GFP10 and GFP11 tags were fused to FRB and FKBP proteins. Rapamycin ligand binding brings both protein fusions into proximity, permitting GFP fluorescence reconstitution upon addition of GFP1-9. (b) Raw fluorescence progress curves for GFP1-9 complementation with soluble extracts of GFP10-FRB and FKBP-GFP11 fusions in presence (+RAP) or absence (−RAP) of rapamycin (starting time marked by addition of rapamycin to initiate complementation) (c) Fluorescence levels of GFP10-FRB and FKBP-GFP11 assayed at various concentrations (320, 160, and 40 nM in each) initiated by addition of rapamycin to 150 nM final concentration (black bar) and no rapamycin (gray bar). (d) Rapamycin dose curve (0.6 to 300 nM) for FRB/FKBP binding *in vitro* measured as final fluorescence after 1 h.



proteins had been co-expressed or were refolded after denaturing⁸. In contrast, our soluble crude *E. coli* cell extracts could be mixed with a 4 μ M solution of purified GFP1–9 and incubated for several minutes prior to addition of 150 nM rapamycin, after which the rapid increase in fluorescence indicated soluble interacting GFP10-FRB and FKBP-GFP11 protein. A two-fold increase in fluorescence was easily detectable after only 10 minutes incubation (Fig. 5b, inset), with a half-life of 60 minutes. Without added rapamycin fluorescence remained near blank levels (Fig. 5b). To further examine the behavior of the protein interaction reporter, we studied purified 6-His-GFP10-FRB and 6-His-FKBP-GFP11 fusion proteins (Supplementary Fig. S5b). In the presence of 150 nM rapamycin (much greater than the reported K_d of ca. 12 nM¹⁹), endpoint fluorescence was directly proportional to the amount of added protein complex as expected (Fig. 5c). To further evaluate reaction kinetics and equilibria of rapamycin-FKBP/FRB protein association, we tested the effect of increasing concentrations of rapamycin (0.05 to 300 nM). The rapamycin dose response curve presents a half maximal value of 13.5 nM, in accordance with previously published data¹⁹ (Fig. 5d).

Monitoring the association of protein complexes in *E. coli*. Protein signaling modules often involve more than two interacting partners. To examine the ability of our tagging system to detect multimeric complexes, we selected two polycistrons from the *E. coli* genome encoding the heterotrimer *Tus* BCD complex (YheNML)²⁰ and a

putative allophanate hydrolase constituted by the dimeric assembly YBGJ/YBGK²¹. We also studied an alternate version of the *Tus* complex where YheN subunit was omitted, thus preventing stable association of the complex as previously reported²⁰. These polycistrons were PCR amplified from genomic DNA including their internal ribosome binding sites. Translation of individual complex subunits was dependent on natural IRBS present in the polycistron, so the first subunit had an N-terminal GFP10 and the last subunit had a C-terminal GFP11 furnished by the pTET SpecR vector (Fig. 6a). To ensure sufficient flexibility between complex subunits, especially for the YheNML whose subunits have nearly buried N-termini in the assembled complex²⁰, 30 and 25-mer linkers were inserted between the polycistron and the split-GFP tags (Fig. 6b and Supplementary Fig. S3). These constructs were transformed into *E. coli* cells expressing either the GFP1–9 fragment for complex detection (requiring both the GFP10 and GFP11 tag), or the GFP1–10 fragment¹⁶ to assess the expression level and solubility by monitoring the GFP11 tag (Fig. 6b and 6c). All the complexes were well expressed and highly soluble as shown in Fig. 6c, column A, B. Cells expressing the trimeric YheNML complex became fluorescent after 4h complementation with GFP1–9, as expected (Fig. 6c, 1D). In contrast, weak fluorescence was seen for colonies expressing only YheM and YheL (missing the YheN subunit) with GFP1–9 (Fig. 6c, 2D). Colonies expressing the YBGJ/YBGK control complex were brightly fluorescent after

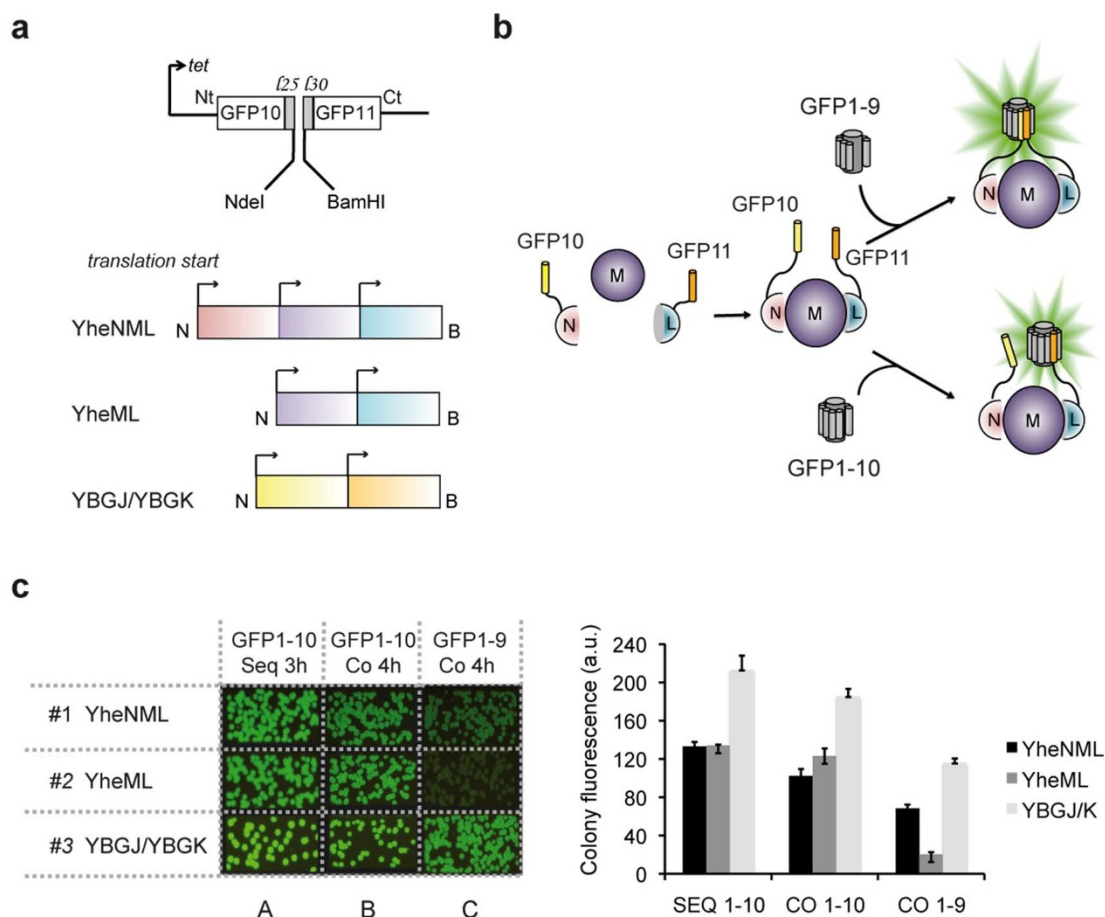


Figure 6 | Monitoring complex formation and stability in *E. coli*. (a) Analysis of YheNML heterotrimeric complex formation using split-GFP. GFP1–9 complementation provides information on interaction of GFP10 and GFP11 tagged domains, while GFP1–10 binds to GFP11 tagged protein, thus providing expression and solubility levels of the GFP11 tagged protein or complex. (b) *E. coli* polycistrons YheNML, YheML and YBGJ/K were subcloned between GFP10 and GFP11 strands with extended linkers (l25 and l30) into pTET GFP10/11 *tet* inducible plasmids. (c) Split-GFP assay in *E. coli* cells expressing either GFP1–9 or GFP1–10. Sequential or co-induction was performed as previously described. Corresponding semi-quantitative measure of colony fluorescence using NIH Image (right).

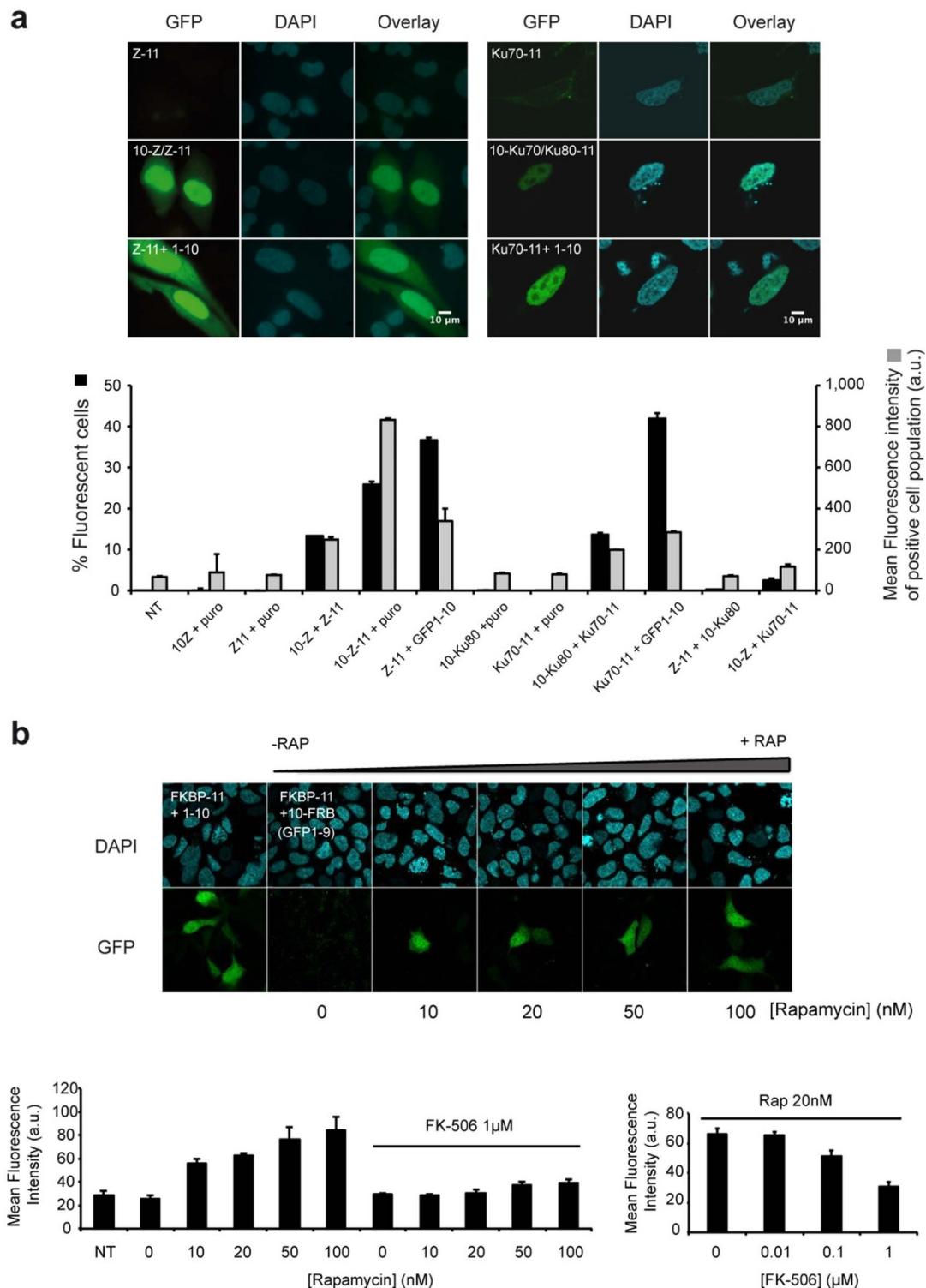


Figure 7 | Visualization of complex formation in mammalian cells. (a) Leucine zipper and Ku70/80 heterodimerization. Cells transiently expressing GFP1–9 with GCN4 zipper (Z-11) (left panel) or Ku70-GFP11 (Ku70-11) (right panel) display no background fluorescence. Heterodimerization is visualized as fluorescence with interacting leucine zippers GFP10-Z and Z-GFP11 in CHO cells expressing GFP1–9. Complementation of GFP11-Z with GFP1–10 confirms localization of the zipper alone (bottom). Ku70/80 complex formation is visualized in cell nuclei (right panel). Expression of one Ku component tagged with GFP11 is monitored with GFP1–10. FACS analysis of HEK 293_GFP 1–9 cell lines transfected with corresponding constructs (24 h after transfection). Percentage of fluorescent cells (black bars); mean fluorescence intensity of the positive cell population (gray bars). Self-associating GFP10-Z-GFP11 domain is used as positive control of transfection and complementation with GFP1–9 (mean $\pm$ SD; N = 3). (b) Rapamycin induced FRB/FKBP interaction in mammalian cells. Stable HEK 293 cells expressing GFP1–9 were co-transfected with GFP10-FRB and FKBP-GFP11 constructs and stimulated with increasing concentrations of rapamycin (RAP) (0, 10, 20, 50 and 100 nM). Bipartite complementation of FKBP-11 and GFP1–10 is shown in the most left image; Green fluorescence at 488 nm excitation (GFP), DAPI nuclear staining (cyan). Scale bars = 10 μ m. Bottom panel: FACS quantification of rapamycin induced association with or without addition of competitive inhibitor FK-506 (1 μ M). Right graph: addition of increasing concentrations of FK-506 (20 nM rapamycin; 0.01, 0.1 and 1 μ M FK-506) (mean $\pm$ SD; N = 3).



complementation with GFP1–9 as expected (Fig. 6c, 3D). These results are in good agreement with those obtained in our previous bead-based assays using GFP1–10 complementation²¹.

Visualization of protein-protein interactions in mammalian cells.

We adapted the split-GFP fragments for optimized expression in mammalian cells and tested the formation of several eukaryotic complexes by fluorescence microscopy. As proof of principle, we first fused the yeast GCN4 leucine zipper²² to each of the GFP10 and GFP11 tags and co-transfected both constructs along with the GFP1–9 detector fragment into CHO cells to study the formation of GCN4 heterodimer (Fig. 7a, left panel). Nuclear fluorescence corresponding to leucine zipper heterodimerization was observed in cells expressing GFP1–9 after 24 h of transfection (Fig. 7a, lane 2). As expected, co-expression of GFP1–9 and one of the leucine zipper domain C-terminally fused to GFP11 did not produce fluorescence (Fig. 7a, lane 1). In a separate experiment we observed the fluorescence complementation of the GFP11 tagged proteins with GFP1–10, which measures soluble protein levels in cells¹⁵. As expected, the leucine zipper domain alone was strongly localized in the nucleus and only weakly in the cytoplasm (Fig. 7a, lane 3). We were also able to detect much larger protein complexes, such as the Ku70–Ku80 complex²³ that is required in the non-homologous end-joining (NHEJ) pathway in DNA repair. We successfully detected heterodimerization between GFP10–Ku80 and Ku70–GFP11 in HEK 293 cells expressing stably the GFP1–9 fragment (Fig. 7a, right panel). To shed additional light on possible background from the tripartite assay in living cells, we co-expressed leucine-zipper and Ku proteins that localize in the same subcellular compartment. Quantification of the fluorescence levels by flow activated cell sorting (FACS) indicated basal levels of fluorescence from individual split-GFP controls and between non-interacting proteins tested. Pairwise leucine zipper and Ku subunit heterodimerization led to fluorescence levels comparable to the bipartite split-GFP assay, which titrates soluble protein levels in cells¹⁵ (Fig. 7a). In parallel, we co-expressed GFP10-FRB and FKBP–GFP11 protein fusions in HEK 293_GFP1–9 cells, stimulated with increasing concentrations of rapamycin. Fluorescence levels for the unstimulated cells remain basal, whereas bright fluorescent cells could be visualized with as little as 10 nM rapamycin, in the range of the EC₅₀ value measured for the ternary complex *in vitro*²⁴ (Fig. 7b). Moreover, pre-addition of an excess of competitive inhibitor FK-506 totally prevented FRB/FKBP association by rapamycin in an FK-506 dose dependent manner (Fig. 7b, bottom panel), thus demonstrating that tripartite split-GFP complementation can be used to monitor inhibitors of protein-protein interactions by small molecule compounds prior to association of interacting partners.

Discussion

Here we describe the first protein-protein interaction reporter based on tripartite split-GFP association. Instead of the bulky and poorly folded BiFC fragments of GFP^{12,25}, our tags based on small engineered β -strands of GFP (~20 amino acids long) minimize protein interference and aggregation. Interaction assays using the E-coil/K-coil model¹⁸ have a sensitivity limit in the picomole range. Chemically induced interactions of the FRB/FKBP complex are detectable using the split-GFP within a few minutes after addition of rapamycin *in vitro*. Bulky BiFC fragments from various fluorescent proteins are expressed largely in *E. coli* in inclusion bodies⁸, and the effect of the fragments from other enzyme-based PCA assays on the folding of fused proteins is poorly understood²⁶. Here we show that our tagging system is highly soluble, permitting production of high yield of fusion proteins at 37°C in *E. coli*. The assay is therefore not temperature dependent, unlike other BiFC for which co-expression and decreasing growth temperature, or refolding denatured

extracts are the only ways to improve assembly of split-GFP fragments^{8,27}. We have shown that the method is sensitive enough to detect association of proteins expressed from *E. coli* polycistronic mRNAs, and fluorescence is correlated with independently measured protein complex formation and stability. As shown for the TusABC (YheNML) studies above, we could monitor the formation of larger trimeric complexes by extending the length of the linkers between the GFP tag and the protein of interest. This suggests that precise geometry of the GFP10 and GFP11 is not critical for interaction with GFP1–9, but rather that reduction of entropy is important for triggering assembly of the GFP10, GFP 11, and GFP1–9.

The newly described tripartite split-GFP assay is a promising tool to study protein-protein interactions *in vitro* and in living cells. Its chief advantage over the bulky aggregation-prone fragments of existing BiFC is the small sizes of the GFP10 and GFP11 tagging peptides. It is therefore well suited to study interactions of unstable protein complexes that are difficult to detect with larger GFP tags. Our system greatly expands the prospect for protein-interaction screening and the design of new biosensors for protein complex assembly and association²⁸. Although association of the 3-body GFP fragments is currently irreversible, this system can be exploited to turn on the detection of protein complex formation by simple addition of GFP1–9 reagent. The method is particularly well adapted to high-throughput interaction screens of libraries of protein-protein interfaces and domains²⁹, the study of complex formation stability, and the robust detection of soluble proteins and protein complexes by flow cytometry. The technology should prove useful for screening small molecule compounds, for example inhibitors for the interfaces of macromolecular protein complexes. Further work might yield split tripartite GFPs whose assembly can be regulated with light³⁰.

Methods

Cloning. K1-coil and E1-coil DNA, FRB (NM_004958) and FKBP (NM_000801) sequences were amplified by PCR using synthetic oligonucleotides (SI note) and cloned into a pTET ColE1 SpecR GFP10/11 via *NdeI*:*KpnI* (GFP10 fusion) or *SpeI*:*BamHI* (GFP11 fusion). Linker extension of pTET GFP10/11 was performed using inverse PCR with synthetic oligonucleotides (Supplementary Fig. S3). GFP1–9 was cloned into a pET28a p15 Kan vector via *NdeI*:*BamHI* sites. Proteins expressed with GFP10 (E1 and K1 coils, FRB) and proteins expressed with GFP11 (E1 coil, FKBP) were subcloned into a pET vector bearing a C-terminal or N-terminal 6-His tag respectively, prior to purification and characterization. For mammalian expression, all the constructs used in the study were derived from pcDNA 3.1 Zeo vector backbone encoding GCN4-hGLuc1 and hLuc2-GCN4 kindly provided by Stephen Michnick. Split-luciferase domains were replaced with GFP10 and GFP11 fragments. Ku80, FRB were cloned into *BspEI*:*XbaI* sites of pcDNA_GFP10; Ku70 and FKBP were amplified from cDNA and inserted into *NotI*:*Clal* cloning sites of pcDNA_GFP11 vector.

Directed evolution of split-GFP fragments for tripartite complementation. The DNA construct encoding (GFP10)-I1(DVSGSGGS)-*NdeI*:GGGSGSGG:*BamHI*-I2 (GGGSGSGGS)-(GFP11), where GFP10 and GFP11 sequences were derived from superfolder GFP (Supplementary Note), was evolved by DNA shuffling. Libraries of GFP10–11 variants expressed from pTET SpecR vector were screened for improved solubility using a sequential induction protocol¹⁶ in *E. coli* cells containing GFP1–9 M1 on a pET p15 vector. At each round, protein solubility of selected optima was verified by complementation with GFP1–10 *in vitro*. From the six brightest clones sequenced after three rounds of evolution, one best mutant (#5) was identified, termed GFP11 M4 (Supplementary Note). Upstream GFP10 fragment (GFP10 M1) was further evolved by DNA shuffling and primer doping mutagenesis with a pool of fourteen synthetic oligonucleotide primers. Each primer was centered at one of the fourteen amino acids of the GFP10 M1 domain, containing an NNN coding degeneracy at the central target amino acid and flanking homology to the GFP10 M1 in the context of the cloning vector. A partially soluble protein HPS was then inserted in the *NdeI*:*BamHI* cloning site to obtain (GFP10 mutant M1)-I1-HPS-I2-(GFP11 mutant M1) in pTET SpecR vector. After three rounds of selection using the sequential induction format from the pTET and pET plasmids¹⁶, one best-performing clone, termed (GFP10 M2)-L1-*NdeI*:HPS:*BamHI*-L2-(GFP11 M4) was isolated. GFP10 M2-HPS-GFP11 M4 fusion and GFP1–9 M1 inserts were swapped between pTET and pET plasmids to perform directed evolution of GFP1–9 M1. Briefly, cDNAs libraries of GFP1–9 were expressed from pTET SpecR plasmids and tested for *in vivo* complementation assays in an *E. coli* strain expressing GFP10 M2-HPS-GFP11 M4 fusion in pET p15. At the second round, GFP1–9 OPT was isolated and sequenced.



GFP1-9 in vitro assays. Production of recombinant GFP1-9 OPT protein fragment was performed according to previous protocol described for split-GFP1-10¹⁶. GFP1-9 was expressed from a pET p15 vector without a 6His tag and produced after induction 1 mM IPTG for 4 hours at 37°C. The protein was refolded from washed inclusions bodies (see GFP1-10 protocol¹⁶) and solubilized in TNG buffer. To perform kinetic characterization of the newly engineered GFP1-9 OPT and GFP1-9 M1, 180 µl of equal amounts of GFP1-9 refolded pellet fractions (ca. 0.5 mg/ml) were mixed with 20 µl of a soluble protein control sulfite reductase (SR) in fusion with sandwich GFP10 and GFP11 (GFP10-SR-GFP11) or GFP10-11 peptide (3.5 µM each). For K1/E1 and E1/E1 kinetic saturation studies, 25 µl of E1-GFP11 (6.25 µM) were mixed with 50 µl of GFP10-K1 or GFP10-E1 at various concentrations (800 nM down to 0.4 nM). For ligand induced interactions, rapamycin (LC laboratories) was diluted in DMSO and added in 10 µl aliquot to the final FRB/FKBP protein assay mix 20 µl of GFP10-FRB + 20 µl FKBP-GFP11 (0.5 mg/ml). Assay was repeated by diluting FRB/FKBP samples in TNG buffer. Complementation was induced by addition of a large excess of GFP1-9 OPT (150 µl, 0.25 mg/ml). Fluorescence kinetic ($\lambda_{exc} = 488 \text{ nm}$ / $\lambda_{em} = 530 \text{ nm}$) was monitored with a FL600 Microplate Fluorescence Reader (Bio-Tek), at 3 min intervals, for 15 h. The background fluorescence of a blank sample (150 µl of GFP1-9 OPT, 0.25 mg/ml and 50 µl 0.5% (w/v) BSA in TNG buffer) was subtracted from the final fluorescence values.

Mammalian cell interactions assays. CHO cells were grown in Ham's F-12 medium (Gibco, Invitrogen Co) and 10% (v/v) fetal bovine serum (FBS); HEK 293 cells were grown in Dulbecco's Modified Eagle Medium (DMEM) and 10% (v/v) FBS. CHO cells were co-transfected with Lipofectamine 2000 (Gibco, Invitrogen Co.) with plasmids encoding for GFP1-9 OPT, GCN4-GFP11, GFP10-GCN4 or GCN4-GFP11 M4+ GFP1-10. Stable GFP1-9 cell lines were produced by lentiviral transduction of HEK 293. HEK 293 GFP1-9 cells were co-transfected with jetPRIME reagent (Polyplus transfection) with 0.5 µg of each plasmid (GFP10 and GFP11 fusions); Twenty-four hours after transfection, cells were washed with PBS and mounted in DAPI-containing ProLong Gold antifading reagent (Molecular Probes). Imaging was performed using a LEICA DM-RB fluorescence microscope, with a 40× and 100× oil immersion objective to visualize the stained cells. Images from CHO cells were acquired with a Photometric CoolSNAP HQ camera and analysed with Metamorph or ImageJ softwares. Imaging of HEK cells was performed using a Zeiss (Carl Zeiss) confocal laser scanning inverted microscope (LSM 710 NLO with Quazar spectral detector array). Flow cytometry measurements were performed using a BD Biosciences FACSCaliburTM cytometer. Data analysis was performed using CellQwest[®] software (BD Biosciences).

- Day, R. N. & Davidson, M. W. The fluorescent protein palette: tools for cellular imaging. *Chem Soc Rev* **38**, 2887–2921 (2009).
- Pfleger, K. D. & Eidne, K. A. Illuminating insights into protein-protein interactions using bioluminescence resonance energy transfer (BRET). *Nat Methods* **3**, 165–174 (2006).
- Pelletier, J. N., Arndt, K. M., Pluckthun, A. & Michnick, S. W. An in vivo library-versus-library selection of optimized protein-protein interactions. *Nat Biotechnol* **17**, 683–690 (1999).
- Rossi, F., Charlton, C. A. & Blau, H. M. Monitoring protein-protein interactions in intact eukaryotic cells by beta-galactosidase complementation. *Proc Natl Acad Sci U S A* **94**, 8405–8410 (1997).
- Galarneau, A., Primeau, M., Trudeau, L. E. & Michnick, S. W. Beta-lactamase protein fragment complementation assays as in vivo and in vitro sensors of protein protein interactions. *Nat Biotechnol* **20**, 619–622 (2002).
- Shekhawat, S. S. & Ghosh, I. Split-protein systems: beyond binary protein-protein interactions. *Curr Opin Chem Biol* **15**, 789–797 (2011).
- Hu, C. D. & Kerppola, T. K. Simultaneous visualization of multiple protein interactions in living cells using multicolor fluorescence complementation analysis. *Nat Biotechnol* **21**, 539–545 (2003).
- Magliery, T. J. *et al.* Detecting protein-protein interactions with a green fluorescent protein fragment reassembly trap: scope and mechanism. *J Am Chem Soc* **127**, 146–157 (2005).
- Morell, M., Espargaro, A., Aviles, F. X. & Ventura, S. Detection of transient protein-protein interactions by bimolecular fluorescence complementation: the Abl-SH3 case. *Proteomics* **7**, 1023–1036 (2007).
- MacDonald, M. L. *et al.* Identifying off-target effects and hidden phenotypes of drugs in human cells. *Nat Chem Biol* **2**, 329–337 (2006).
- Kodama, Y. & Hu, C. D. An improved bimolecular fluorescence complementation assay with a high signal-to-noise ratio. *Biotechniques* **49**, 793–805 (2010).
- Hu, C. D., Chinenov, Y. & Kerppola, T. K. Visualization of interactions among bZIP and Rel family proteins in living cells using bimolecular fluorescence complementation. *Mol Cell* **9**, 789–798 (2002).
- Waldo, G. S., Standish, B. M., Berendzen, J. & Terwilliger, T. C. Rapid protein-folding assay using green fluorescent protein. *Nat Biotechnol* **17**, 691–695 (1999).

- Pedelacq, J. D., Cabantous, S., Tran, T., Terwilliger, T. C. & Waldo, G. S. Engineering and characterization of a superfolder green fluorescent protein. *Nat Biotechnol* **24**, 79–88 (2006).
- Cabantous, S., Terwilliger, T. C. & Waldo, G. S. Protein tagging and detection with engineered self-assembling fragments of green fluorescent protein. *Nat Biotechnol* **23**, 102–107 (2005).
- Cabantous, S. & Waldo, G. S. In vivo and in vitro protein solubility assays using split GFP. *Nat Methods* **3**, 845–854 (2006).
- Kaddoum, L., Magdeleine, E., Waldo, G. S., Joly, E. & Cabantous, S. One-step split GFP staining for sensitive protein detection and localization in mammalian cells. *Biotechniques* **49**, 727–728, 730, 732 passim (2010).
- Tripet, B. *et al.* Engineering a de novo designed coiled-coil heterodimerization domain for the rapid detection, purification and characterization of recombinantly expressed peptides and proteins. *Protein Eng* **10**, 299 (1997).
- Banaszynski, L. A., Liu, C. W. & Wandless, T. J. Characterization of the FKBP. rapamycin. FRB ternary complex. *J Am Chem Soc* **127**, 4715–4721 (2005).
- Numata, T., Fukai, S., Ikeuchi, Y., Suzuki, T. & Nureki, O. Structural basis for sulfur relay to RNA mediated by heterohexameric TusBCD complex. *Structure* **14**, 357–366 (2006).
- Lockard, M. A. *et al.* A high-throughput immobilized bead screen for stable proteins and multi-protein complexes. *Protein Eng Des Sel* **24**, 565–578 (2011).
- Remy, I. & Michnick, S. W. A highly sensitive protein-protein interaction assay based on Gaussia luciferase. *Nat Methods* **3**, 977–979 (2006).
- Cary, R. B., Chen, F., Shen, Z. & Chen, D. J. A central region of Ku80 mediates interaction with Ku70 in vivo. *Nucleic Acids Res* **26**, 974–979 (1998).
- Marz, A. M., Fabian, A. K., Kozany, C., Bracher, A. & Hausch, F. Large FK506-binding proteins shape the pharmacology of rapamycin. *Mol Cell Biol* **33**, 1357–1367 (2013).
- Ghosh, I. Antiparallel Leucine Zipper-Directed Protein Reassembly: Application to the Green Fluorescent Protein. *J. Am. Chem. Soc.* **122**, 5658–5659 (2000).
- Ozawa, T. Designing split reporter proteins for analytical tools. *Anal Chim Acta* **556**, 58–68 (2006).
- Robida, A. M. & Kerppola, T. K. Bimolecular fluorescence complementation analysis of inducible protein interactions: effects of factors affecting protein folding on fluorescent protein fragment association. *J Mol Biol* **394**, 391–409 (2009).
- Kellermann, S. J., Rath, A. K. & Rentmeister, A. Tetramolecular fluorescence complementation for detection of specific RNAs in vitro. *Chembiochem* **14**, 200–204 (2013).
- Pedelacq, J. D. *et al.* Experimental mapping of soluble protein domains using a hierarchical approach. *Nucleic Acids Res* (2011).
- Do, K. & Boxer, S. G. Thermodynamics, kinetics, and photochemistry of beta-strand association and dissociation in a split-GFP system. *J Am Chem Soc* **133**, 18078–18081 (2011).

Acknowledgements

We gratefully acknowledge Dr. P. Calsou and Dr. P. Frit for providing Ku80 and Ku70 cDNAs, S. Michnick for the GCN4-hGluC1 and hLuc2-GCN4 vectors. We thank Sophie Allart and Astrid Canivet for technical assistance at the Cellular Imaging Facility of CPTP, Toulouse. This work was supported by the National Institutes of Health NIH 1 U54 GM074946-01 to SC, GSW and TCT (LANL), and LDRD/DOE ER projects 20110443ER and 20100520ER (LANL), the French 'Région Midi-Pyrénées', INSERM and University of Toulouse to FK, SC, JDP, GF, and a DTRA-TMTI grant to AC.

Author contributions

G.W. and S.C. designed the study. G.W., S.C., H.N., F.K., K.G. and M.L. performed experiments. S.C., H.N., J.D.P. and G.W. wrote the main manuscript text. All authors revised the manuscript. A.C., G.F. provided reagents. G.W. and T.T. supervised the project.

Additional information

Supplementary information accompanies this paper at <http://www.nature.com/scientificreports>

Competing financial interests: The authors declare competing financial interests. The split-GFP technologies are the subject of domestic and foreign patent applications by Los Alamos National Laboratories on behalf of the Department of Energy and LANS, L.L.C.

How to cite this article: Cabantous, S. *et al.* A New Protein-Protein Interaction Sensor Based on Tripartite Split-GFP Association. *Sci. Rep.* **3**, 2854; DOI:10.1038/srep02854 (2013).



This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivs 3.0 Unported license. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-nd/3.0>

A New Protein-Protein Interaction Sensor Based on Tripartite Split-GFP Association

Supplementary information

Stéphanie Cabantous^{1,*}, Hau B. Nguyen², Jean-Denis Pedelacq³, Faten Koraïchi¹, Anu Chaudhary⁴, Kumkum Ganguly², Meghan A. Lockard⁵, Gilles Favre¹, Thomas C. Terwilliger², & Geoffrey S. Waldo^{2,*}

Supplementary Fig.S1: Solubility and *in vitro* complementation of folding reporter (FR) and superfolder (SF) GFP BIFC fragments.

Supplementary Fig. S2: Helical coiled-coils model sequences.

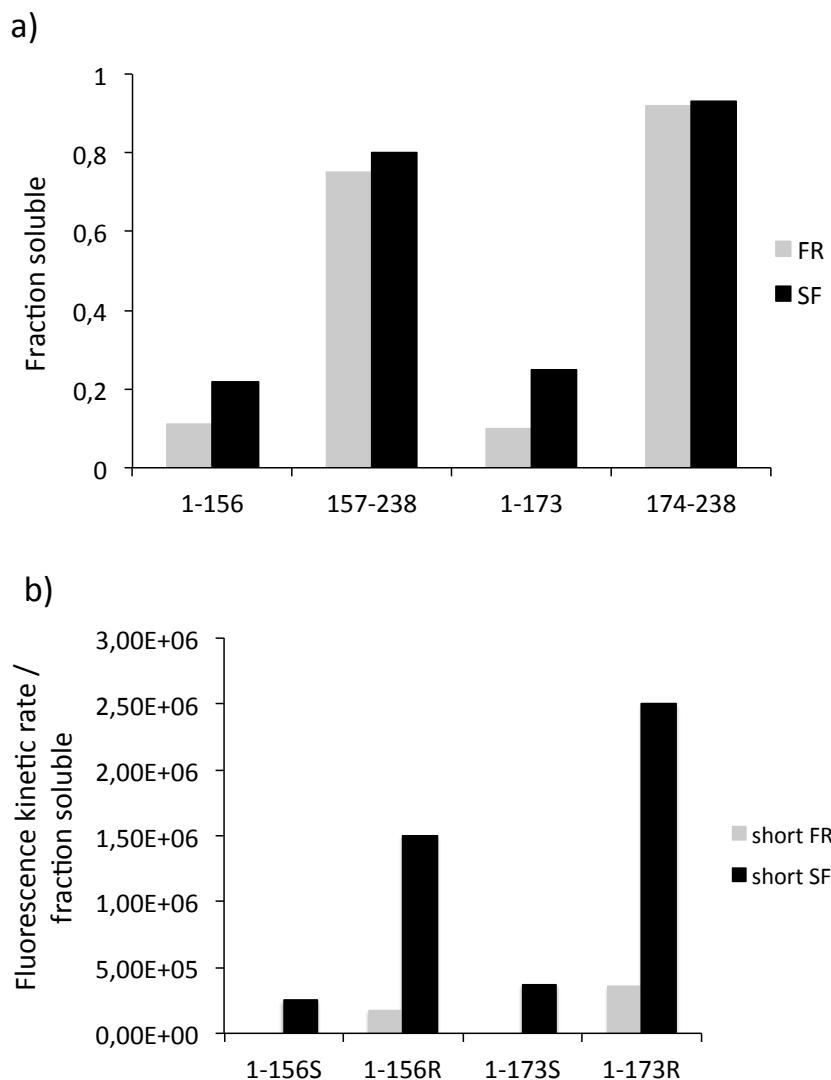
Supplementary Fig. S3: Vector map of pTET Bicistronic GFP10-IRBS-GFP11 vector.

Supplementary Fig. S4: Screen of the junction split-point between GFP9 and GFP10 strands.

Supplementary Fig. S5: SDS-PAGE of soluble and Talon® IMAC purified fractions from 6-His-tagged GFP10 and GFP11 fusion proteins.

Supplementary Table S1: Amino-acid sequence of GFP10 and GFP11 tags

Supplementary Note: Methods and sequences.



Supplementary Figure S1: Solubility and *in vitro* complementation of folding reporter (FR) and superfolder (SF) GFP BiFC fragments expressed individually in *E. coli*. We previously developed a bipartite split-GFP system to quantify protein solubility and follow protein localization *in vivo*^{1,2}. Considering that self-association between GFP11 and GFP1-10 makes this system unsuitable for protein-protein interaction studies, we then studied the behavior of BiFC fragments³ obtained by fragmentation of folding-reporter GFP⁴ (FR-GFP) and Superfolder GFP⁵ (SF-GFP) at permissive sites 156 and 173⁶. **a)** FR and SF Split-GFP fragments were produced by splitting GFP at permissive 156 and 173 positions, as described in BiFC assays³. Fragments (1-157) and (158-238), (1-173) and (174-238) were expressed from pET vectors in *E. coli* at 37°C for 3h. Soluble and insoluble fractions were quantified by SDS-PAGE. Small GFP fragments are mostly soluble whereas large fragments are largely insoluble. **b)** *In vitro* complementation assay combining soluble fractions from small split-GFP fragments with the soluble (S) or refolded (R) fractions of the corresponding large fragments (1-156) and (1-173). As observed with other fluorescent protein fragments used in refolding-based BiFC³, split-GFP fragments from refolded FR-GFP self-associate detectably (gray bars). Improved folding in SF-GFP fragments resulted in improved solubility, but also increased spontaneous fluorescence for all conditions (whether refolded or soluble extracts) (black bar).

Attractive coiled coil pair

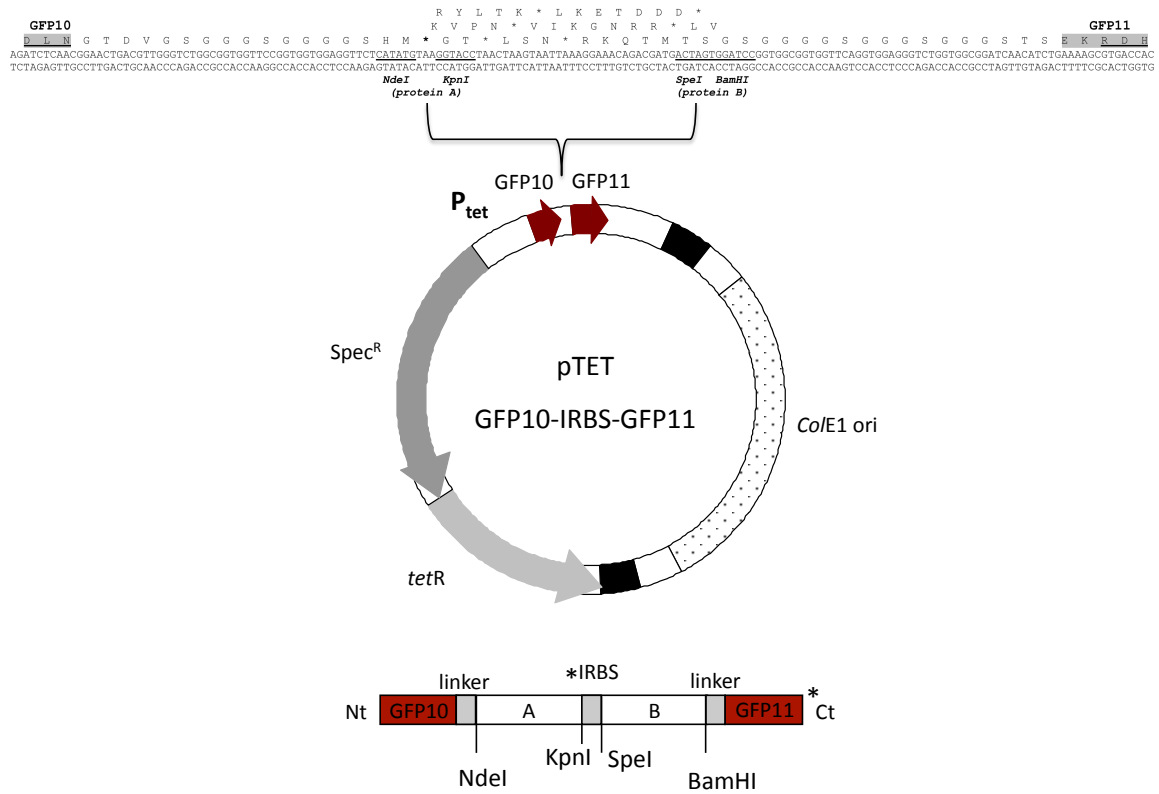
K1 coil KVSALKENVSALKEKVSALTEKVSALKEKVSALKE
 ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓
E1 coil KVSALENEVSALEKEKVSVLEKEKVSALKEKVRALEK

Repulsive coiled coil pair

E1 coil KVSALENEVSALEKEKVSVLEKEKVSALKEKVRALEK
 ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓
E1 coil KVSALENEVSALEKEKVSVLEKEKVSALKEKVRALEK

Supplementary Figure S2: Helical coiled-coils as a model to study protein-protein interactions. Developed by Hodges and coworkers⁷, E and K coils are constituted by five repeated heptads of the EVSALEK or KVSALKE motifs (gray boxes). Electrostatic interactions introduced in the first and sixth positions of the heptads direct preferential interaction, allowing interchain ionic attraction (K/E) or repulsion (E/E) (black arrows). Valine (V) and Leucine (L) residues maintain hydrophobic interactions between coils (bold). Directed evolution of wild-type sequences (G. S. Waldo, unpublished data) resulted in additional mutations that greatly increased solubility of K coil mutant 1 (K1) and E coil mutant 1 (E1) (underlined) while maintaining the specificity and tight binding of the wild type. E1 and K1 coils were used in our interaction studies.

a)



b)

long linker (30-mer) GFP10 tag

```

D V G G G G S E G G G S G G P G S G G E G S A G G G S A G G G S
GACGTTGGTGGTGGCGGATCAGAAGGAGGCGGTAGCGGGGGCCCTGGTTTCGGGAGGGGAAGGTTCTGCTGGGGGAGGGAGCGCTGGCGGGGGGTCT
CTGCAACCACCACCGCCTAGTCTTCTCCGCCATCGCCCCGGGACCAAGCCCTCCCTTCCAAGACGACCCCTCCCTCGCGACCGCCCCCAGA

```

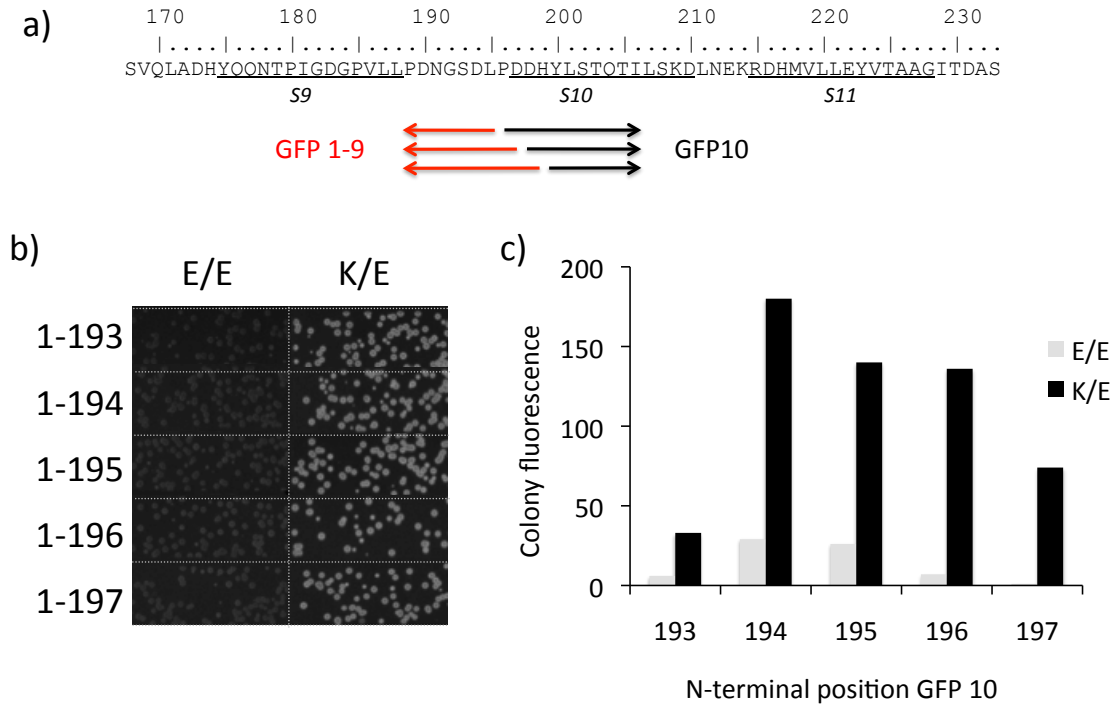
linker (25-mer) GFP11 tag

```

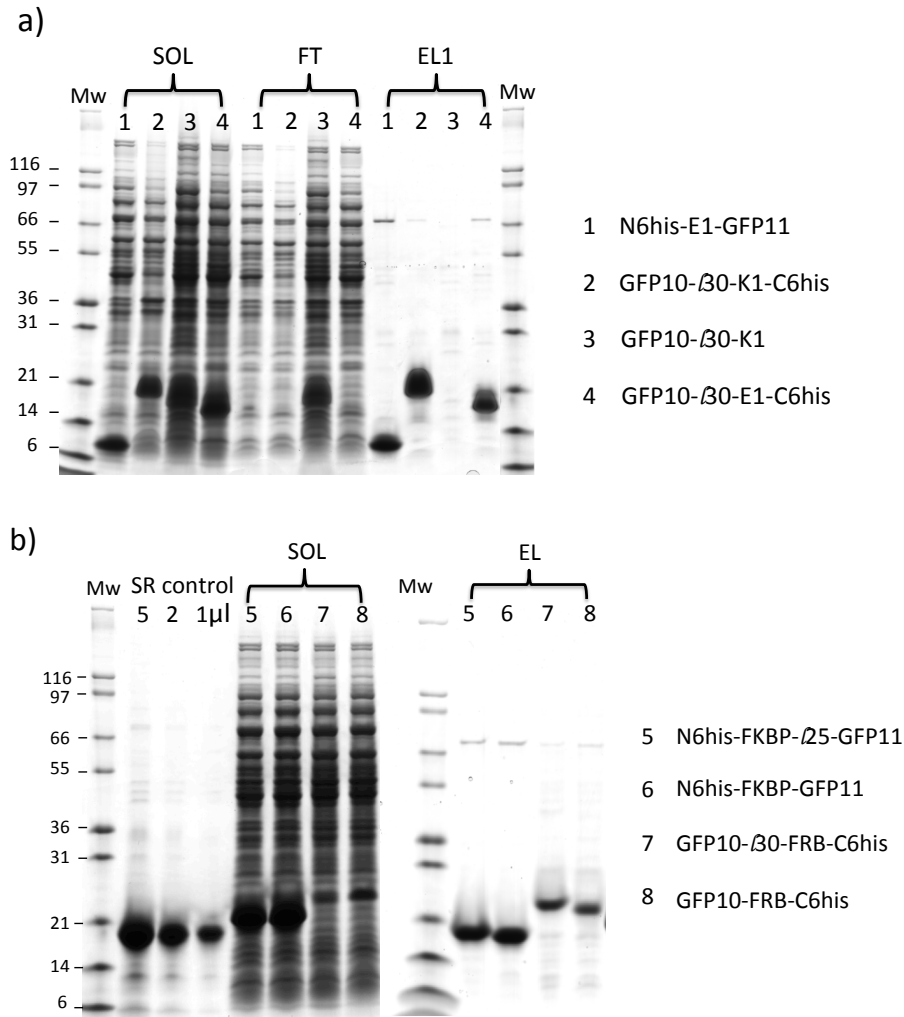
G S G A G G S P G G G S G G S G S S A S G G S T S
GGATCCGGCGCTGGCGGAAGCCCTGGGGCGGGAGCGGTGGCTCTGGTTCTTCTGCTAGTGGCGGCTCAACATCT
CCTAGGCCGCGACCGCCTTCGGGACCCCGCCCTCGCCACCGAGACCAAGAAGACGATACCGCCGAGTTGTAGA

```

Supplementary Figure S3: Vector map of pTET Bicistronic GFP10-IRBS-GFP11 vector. a) The pTET GFP10-IRBS-GFP11 vector is based on our previously designed expression vector⁸. It includes a bicistronic expression cassette under the control of the *tet* promoter. In one tagging module, N-terminal GFP10 is separated from protein A by a 12-mer linker (cloning sites *NdeI/KpnI*). An internal ribosome-binding site separates it from a second module comprising *SpeI/BamHI* cloning sites for protein B), a 12-mer linker, and the C-terminal GFP11 tag. b) Longer linkers of 30-mer (GFP10 tag) and 25-mer (GFP11 tag) were used for interaction assays of larger proteins (>50kD). The pTET *ColE ori* GFP10-IRBS-GFP11 vector was used alone for *in vitro* assays or in combination with compatible pET *p15 ori* GFP1-9 OPT plasmid for studying interactions in *E. coli*.



Supplementary Figure S4: Effect of split-point at the linker junction between GFP9 and GFP10 β -strands. **a)** To investigate the optimal split position between β -strands GFP9 and GFP10, we designed GFP1-9 with extended C-termini (red arrows) and corresponding GFP10 with truncated N-terminus (black arrows). **b)** Several truncated versions of GFP1-9 were expressed from pET vectors and tested for complementation *in vivo* with the non-interacting E/E or interacting K1/E1 coiled-coil pair expressed from pTET bicistron (GFP10-E1/E1-GFP11) or (GFP10-K1/E1-GFP11). **c)** Comparison of the output signal of the same bicistronic constructs as a function of N-terminal GFP10 truncations, GFP1-9 being kept fixed at amino acid position 193. Colony fluorescence after 2h induction at 37°C was compared for K1/E1 and E1/E1 coils pairs to evaluate the optimal split-point. Last permissive complementation was given for position 194, designating the final version of GFP10 used throughout the study. However, extension of C-termini of the GFP1-9 fragment (incrementally from amino-acid 193 to 197) had no effect on the fluorescence levels observed for interacting K1/E1 versus non-interacting E1/E1 coils.



Supplementary Figure S5: SDS-PAGE of soluble and Talon® purified fractions from 6HIS-tagged GFP10 and GFP11 fusion proteins with various linker sizes: 12mer, 25mer (indicated as *Δ25*) or 30mer (indicated as *Δ30*). To further characterize protein-protein interaction *in vitro*, K1 coil, E1 coil, or FRB were purified to 95% homogeneity. A 200 ml culture of BL21(DE3) cells expressing each construct was grown to $OD_{600nm} \sim 0.5$ in LB medium supplemented with 1 mM kanamycin, induced with 1 mM IPTG for 3 h at 37°C. Soluble extracts were purified on a 50% v/v slurry of metal affinity resin beads (Talon® resin, Clontech) in TNG buffer (100mM Tris pH 8.0, 150mM NaCl, 10% glycerol) according to standard procedures. Samples corresponding to the soluble and eluted fractions were resolved on a 4-20% gradient Criterion SDS-PAGE gel (Bio-Rad Hercules). Protein samples were stained using Gel Code Blue stain reagent (Pierce) and imaged using a GS-800 Calibrated Densitometer and quantified by Bio-Rad Protein Assay (Bio-Rad). **a)** Coiled-coils. Both C-6His (#2) and untagged (#3) versions of GFP10-K1 fusion expressed mostly in the soluble fraction. Flow-through fractions (FT) contained cellular proteins and untagged proteins. 90% pure proteins were recovered after one-step purification in the eluted fraction (EL) **b)** FKBP-GFP11 and GFP10-FRB fusions with different linker lengths. A sulfite reductase control protein (SR) sandwich with GFP10 and GFP11 tags was used as positive control of protein expression and complementation with GFP1-9.

Supplementary Table S1: Amino-acid sequence of GFP10 and GFP11 β -strands

<i>Fragment</i>	^{a,b,c} <i>Amino acid sequence</i>
GFP10 SF	195 200 205 210 MGLPD<u>NHYLSTQ</u>SVLSKDPN
GFP10 M1 (evolved 10-linker-11) (G194D, S205T, A206I, S208L, P211L)	MDLPDNHYLSTQ TILL KD LN
GFP10 M2 (evolved 10-HPS-11) (G194D, N198D, S205T, A206I, S208, P211L)	MDLPD D HYLSTQ TILS KD LN
GFP11 SF	215 220 225 230 EKRDH<u>MVLL</u>EFVTAAGITGAS^d
GFP11 M4 (F223Y)	EKRDH <u>MVLL</u> E Y VTAAGITD AS

^aUnderlined residues correspond to structural β -strand sequence from Superfolder GFP.

^bBold residues correspond to the additional point mutations introduced by mutagenesis.

^cNumbering corresponds to position in full-length SF GFP.

^dResidues following amino-acid 230 correspond to linker sequences.

Supplementary Note

1. Amino-acid sequences of evolved split-GFP fragments

Amino-acid sequence of GFP 10-11 hairpin

GFP10 and GFP11 are separated by a long flexible linker (wt, gray shade) that includes two cloning sites *NdeI* and *BamHI*. The entire cassette was shuffled and selected for optimized complementation with GFP1-9. Six optima were isolated and sequenced. SM5 corresponds to the best variant, with mutations identified both in GFP10 (GFP10 M1) and only one mutation in GFP11 (GFP11 M4), thus differing from the previous GFP11M3 split-GFP domain identified with GFP1-10 complementation¹. Amino-acids corresponding to structural β -strands GFP10 and GFP11 are underlined.

	GFP10	<i>NdeI</i>	<i>BamHI</i>	GFP11
WT	YTM <u>GL</u> PDNHYLSTQ <u>SVLSKDP</u> NGTGGGSGGGSHMGGGS <u>GS</u> GGGSGGGSTSEKRDHMLVLEFVTAAGITGAS*			
SM1	YTM <u>DLP</u> DNHYLSTQ <u>TILLKDL</u> NGTGVGSGGGSHMGGGSGSGGGSGGGSTSEKRDHMLVLEYVTAAGITDAS*			
SM2	YTM <u>DLP</u> DNHYLSTQ <u>TILLKDL</u> NGTGVGSGGGSHMGGGSGSGGE <u>SGGG</u> STSEKRDHMLVLEYVTAAGITGAS*			
SM3	YTM <u>DLP</u> DNHYLSTQ <u>TILLKDL</u> NGTGVGSGGGSHMGGGSGSGGGSGGGSTSEKRDHMLVLEYVTAAGITDAS*			
SM4	YTM <u>DLP</u> DNHYLSTQ <u>TILLKDL</u> NGTDVGSGGGSHMGGGSGSDGGSGGGSTSEKRDHMLVLEYVTAAGITGAS*			
SM5	YTM <u>DLP</u> DNHYLSTQ <u>TILLKDL</u> NGTGVGSGGGSHMGGGSGSGGGSGGGSTSEKRDHMLVLEYVTAAGITDAS*			
SM6	YTM <u>DLP</u> DNHYLSTQ <u>TILLKDL</u> NGTGGGSGDGCHMDGGSGSGGGSGGGSTSEKRDHMLVLEYVTAAGITGAS*			

Amino-acid sequence of GFP10 optima in HPS fusion

A partially soluble protein hexulose phosphate synthase (HPS) was cloned into the *NdeI* and *BamHI* cloning sites and the cassette was evolved, excluding HPS. Additional rounds of doped mutagenesis with oligonucleotides targeting GFP10 M1 were realized, and resulted in best mutant GFP10 M2. GFP11 sequence was unchanged, as solubility of this fragment was optimized.

GFP10 M1 YTMDLPDNHYLSTQTILLKDLNGTGVGSGGGSHM-(HPS)-GSGGGSGGGSTSEKRDHMLVLEYVTAAGITDAS*
GFP10 M2 YTMDLPDDHYLSTQTILSKDLNGTDVGSGGGSHM-(HPS)-GSGGGSGGGSTSEKRDHMLVLEYVTAAGITDAS*

Alignment of GFP 1-9 variants obtained by directed evolution (aa 1-196)

	10	20	30	40	50	60
GFP1-9 SF	MSKGEELFTGVVPILVELDGDVNGHKFSVRGEGG	DATNGKLT	TKLFI	CTTGKLPVPWPTL		
GFP1-9 M1	.R.....I.....F.....S.....					
GFP1-9 OPT	.R.....I.....F.....I.....S.....					
	70	80	90	100	110	120
GFP1-9 SF	VTTLTYGVQCFSRYPDHMKRHDFFKSAMPEGYVQERTISFKDDGTYKTRA	EVKFE	GGDTLV			
GFP1-9 M1						
GFP1-9 OPT	Y.....					
	130	140	150	160	170	180
GFP1-9 SF	NRIELKGIDFKEDGNILGHKLEYNFSHN	VYITADKQKNGIKANFKIRHNV	EDGSVQLAD			
GFP1-9 M1						
GFP1-9 OPT	K.....N.....T.....					
	190					
GFP1-9 SF	HYQQNTPIGDGPVLLP*					
GFP1-9 M1	*					
GFP1-9 OPT	*					

Several consensus mutations such as S2R, T43S, K166T appeared during directed evolution of GFP1-9 M1 against the GFP10-11 hairpin. Additional key mutations confer GFP1-9 OPT specific binding to sandwich 10-HPS-11: N39I, S99Y, K149N and K158N.

2. Oligonucleotides used to amplify *E. coli* polycistrons

Forward NdeI primer to YBGJ of YBGJ/YBGK *E.coli*

AGATATACATATGCAACGAGCGCGTTGTTATCTGATAGG

Bottom BamHI non-stop to YBGK of YBGJ/YBGK *E.Coli*

AATTCGGATCCATTTTCATTGTGCAGCCGCCACGCTA

Forward NdeI primer to YheN *E.coli*

AGATATACATATGCGTTTTGCCATCGTGGTGACCGGGCC

Forward NdeI primer to YheM *E.coli*

AGATATACATATGAAACGAATTGCGTTGTTTTTCTAC

Reverse BamHI primer to YheL *E coli*

AATTCGGATCCCCAGGCCATCTGGCTGGAGTGCTTAA

Supplementary information references

1. Cabantous, S., Terwilliger, T.C. & Waldo, G.S. Protein tagging and detection with engineered self-assembling fragments of green fluorescent protein. *Nat Biotechnol* **23**, 102-107 (2005).
2. Kaddoum, L., Magdeleine, E., Waldo, G.S., Joly, E. & Cabantous, S. One-step split GFP staining for sensitive protein detection and localization in mammalian cells. *Biotechniques* **49**, 727-728, 730, 732 passim (2010).
3. Hu, C.D., Chinenov, Y. & Kerppola, T.K. Visualization of interactions among bZIP and Rel family proteins in living cells using bimolecular fluorescence complementation. *Mol Cell* **9**, 789-798 (2002).
4. Waldo, G.S., Standish, B.M., Berendzen, J. & Terwilliger, T.C. Rapid protein-folding assay using green fluorescent protein. *Nat Biotechnol* **17**, 691-695 (1999).
5. Pedelacq, J.D., Cabantous, S., Tran, T., Terwilliger, T.C. & Waldo, G.S. Engineering and characterization of a superfolder green fluorescent protein. *Nat Biotechnol* **24**, 79-88 (2006).
6. Hu, C.D. & Kerppola, T.K. Simultaneous visualization of multiple protein interactions in living cells using multicolor fluorescence complementation analysis. *Nat Biotechnol* **21**, 539-545 (2003).
7. Tripet, B., *et al.* Engineering a de novo designed coiled-coil heterodimerization domain for the rapid detection, purification and characterization of recombinantly expressed peptides and proteins. *Protein Eng* **10**, 299 (1997).
8. Cabantous, S. & Waldo, G.S. In vivo and in vitro protein solubility assays using split GFP. *Nat Methods* **3**, 845-854 (2006).
9. Cary, R.B., Chen, F., Shen, Z. & Chen, D.J. A central region of Ku80 mediates interaction with Ku70 in vivo. *Nucleic Acids Res* **26**, 974-979 (1998).
10. Osipovich, O., Durum, S.K. & Muegge, K. Defining the minimal domain of Ku80 for interaction with Ku70. *J Biol Chem* **272**, 27259-27265 (1997).

