Electronic Supplementary Material (ESI) for Chemical Science.
This journal is © The Royal Society of Chemistry 2018

Genetically Encoded Tags for Real Time Dissection of Protein Assembly in Living Cells

Guolin Ma[‡, *, a] Qian Zhang[‡, a, b] Lian He[a], Nhung T. Nguyen[a], Shuzhong Liu[a], Zuojiong Gong[b], Yun Huang*[,c, d], Yubin Zhou*[,a, e]

Supplementary Information

Includes

- Supplementary Method
- Supplementary Notes
- 14 Supplementary Figures
- 1 Supplementary Table
- 5 Supplementary Videos
- Supplementary References

METHODS.

Reagents. The KOD Hot Start DNA Polymerase was purchased from EMD Millipore Corporation. Restriction endonucleases, NEBuilder HiFi DNA Assembly Master Mix and T4 DNA ligase were obtained from New England BioLabs, QuikChange Multi Site-directed Mutagenesis Kit was purchased from Agilent Technologies. Rapamycin (CAS, 53123-88-9) and thapsigargin (CAS, 67526-95-8) were purchased from Sigma-Aldrich. These compounds were dissolved in DMSO and prepared as stock solutions (1 mg/mL).

Plasmid Construction. The microtubule associated proteins were mainly purchased from Addgene (EB1-GFP, #17234; GFP-α-tubulin, #12298; CLIP170, #54044; CAMSAP1 and CAMSAP2, # 59036 and 59037). To generate mCherry tagged fusion constructs, we firstly amplified the corresponding fragments of microtubule association/binding domains (MADs; see Figs S2-4) by standard PCR and then inserted them individually into a modified pmCherry-C1 vector (Clontech). A 24-aa flexible linker made of $(SGGGGGGG)_3$ was inserted between mCh and MADs. To generate oligomeric mCh-MAD fusions, we used the cDNA encoding glutathione S-transferase (GST) from the GST expression vector pGEX-6p-1 (GE Healthcare Life Sciences). AtHAL3 was purchased from TAIR (U67024), while DsRed2, human HSF1 and human p53 were all purchased from Addgene (#54611, #71724 and #10838). cDNA sequences encoding GCN4 (a dimeric coiled coil)[1] and OsHAL3 (a trimeric plant protein homologous to AtHAL3)[2, 3] were synthesized by Integrated DNA Technologies. STIM1 variants were generated as we described previously[4]. These various oligomeric proteins were amplified by standard PCR and inserted into the upstream of the linker in the indicated mCh-MAD plasmids.

Rapamycin induced dimerization (FKBP-FRB) and tetramerization (FRB-FKBP) constructs were obtained as described previously[5]. For light induced dimerization of MoTag2, iLID and sspB components were amplified form Venus-iLID-CAAX and MBP-sspB (Addgene # 60411 and # 60409)[6]. The P2A version, mCh-iLID-MoTag2-P2A-mCh-sspB-MoTag2, was generated using the HiFi DNA assembly method by ligating amplified fragments composed of mCh-iLID-MoTag2 and P2A-mCh-sspB-MoTag2. For CRY2 fused MoTag2, we amplified the photosensitive N-terminal PHR domain (aa 1-498) of *Arabidopsis thaliana* CRY2 (Addgene # 70159)[7] and inserted it into mCh-MoTag1. To generate p53-MoTag2, HSF1-MoTag2 or STIM1-MoTag2 fusion variants, a series of truncated or deleted fragments derived from p53 and HSF1 were amplified, respectively, using the KOD hot start DNA polymerase and then inserted into the mCh-MoTag2 backbone between the HindIII and XhoI restriction sites. Mutations were introduced by using the QuikChange Lightning Multi Site-Directed Mutagenesis Kit.

The constructs for expressing the recombinant proteins as His_6-tagged proteins in *E. coli* were prepared using the pPro-EX HTb vector (Life Technologies). The cDNA sequences of oligomeric proteins with or without MoTag2 were amplified via PCR and cloned into the vector between the BamHI and XhoI sites for

expression of monomeric H_6-mCherry-MoTag2 (or H_6-mCherry), dimeric H_6-GST-MoTag2 (or H_6-GST) and tetrameric H_6-DsRed-MoTag2 (or H_6-DsRed).

Cell culture and transfection. The HeLa cell line was obtained from ATCC and cultured at 37 °C with 5% CO_2 in complete cell-culture medium. For fluorescence imaging, HeLa cells were seeded and cultured in 35-mm glass bottom dishes (MatTek). For transfection, 200-300 ng plasmids were mixed with Lipofectamine 3000 (Thermo Fisher Scientific) in Opti-MEM medium (Thermo Fisher Scientific) by following the manufacturer's instructions.

Live-cell imaging and analysis. Live-cell imaging was mainly carried out on a Nikon Eclipse Ti-E microscope (Nikon Instruments), installed with an A1R-A1 confocal module with LU-N4 laser sources (argon-ion: 405 and 488 nm; diode: 561 nm), CFI (chrome-free infinity) plan Apochromat VC series objective lenses (60 × oil or 40 × oil). To screen MADs, various oligoermic mCh-MAD fusion were transfected and imaged after transfection 18 h. To exam how the protein expression levels would affect the quantification of MT- or comet-to-cytosol ratios, varying amounts (200 ng, 500 ng and 1000 ng) of plasmids encoding the MoTag-fusion proteins were transfected in HeLa cells and imaged after transfection for 18 h. Rapamycin (at a final concentration of 5 µM) was added to induce mCherry-FKBP-FRB-MoTag dimerization or mCherry-FRB-FKBP-MoTag tetramerization. Photostimulation was applied with an external light source (470 nm, 4 µW/mm^2, ThorLabs Inc., Newton, NJ, USA) to elicit blue light induced dimerization (iLID/sspB) and tetramerization (CRY2-PHR).

All the acquired confocal images were analyzed by using the NIS-Elements AR microscope imaging software (Nikon, NIS-element AR version 4.0). The defined regions of interest such as MT or MT plus ends (comets) and the nearby cytosolic areas were measured with the "Intensity Line Profile" tool by drawing a line to extract the intensity distribution profiles. Next, the fluorescence intensity values of MT region versus the neighboring cytosolic mean intensity were determined to obtain the MT-to-cytosol ratio ($F_{MT}/F_{cytosol}$ or $F_{comet}/F_{cytosol}$; see examples in Figures S5-6). 6 ~ 8 different areas within each cell were typically chosen to obtain the averaged ratios. 15 ~ 20 cells were always selected to quantify the target protein's final MT- or Comet-to-cytosol ratio. The collected data were further analyzed and plotted by GraphPad Prism 5. The relationship between the comet-to-cytosol fluorescent signal ratio of mCh-MoTag2-tagged proteins and their oligomeric states was further discussed in Supplementary Notes.

Epifluorescence Microscope (DeltaVision Elite Microscope, GE Healthcare) with a 60 ×/1.45 oil lens and a CoolSNAP EMCCD camera was used to verify MoTag's applications. Images acquired by DeltaVision Microscope were saved in tiff format and further processed by Image J (NIH). Similar with NIS-element software, the MT-to-cytosol or comet-to-cytosol analysis was calculated with the command of [Analyze] – [Plot profile] in Image J.

Protein Expression and Purification. The constructs encoding various oligomeric proteins were transformed into *Escherichia coli* strain BL21 (DE3) cells (EMD Millipore), and grown at 37 °C in LB medium with 100 mg/L of ampicillin. Protein expression was induced by the addition of 500 µM IPTG when OD_{600} of the culture reached 0.6~0.8. For mCherry and DsRed fusions, the bacteria were incubated for another 12–16 hour at 16 °C; while for GST fusions, the bacteria were incubated for another 3 – 4 hour at 37 °C. Harvested cells were resuspended in a resuspension buffer containing 20 mM Tris-HCl pH 8.0, 10 mM imidazole, 150 mM NaCl, 2 mM DTT. The cellular debris was removed by centrifugation. The lysate was loaded to Ni^{2+}-nitrilotriacetic acid (Ni-NTA)-agarose resin (Qiagen). Bound recombinant proteins were eluted in 20 mM Tris pH 8.0, 250 mM imidazole, 150 mM NaCl, 2 mM DTT. The eluents were further purified by gel filtration on a Superdex 200 10/300 GL column (GE Healthcare). The peak fractions were further confirmed by SDS-PAGE.

Cell cycle and viability analysis. HeLa cells were transfected with DsRed only, tetrameric DsRed-MoTag1 or DsRed-MoTag2. After transfection for 24 h, the harvested HeLa cells were washed with PBS, and then fixed in ice-cold 70% ethanol at 4°C for 30 min. After wash with PBS, the cells were treated with 1 µg/ml of DAPI (Sigma D9542) then analyzed by flow cytometry. The histograms of cell distribution in various cell cycle stages were acquired by using a BD LSRII flow cytometer (BD Biosciences). The FlowJo software was used to analyze cell cycle distribution and determine the fraction of cells in the G_0/G_1, S, and G_2/M stages. Each sample was assayed in triplicate. Cell viability was analyzed by using the standard trypan blue staining assay.

Statistical analysis. Quantitative data are presented as the mean and s.e.m. unless otherwise noted. Sample sizes (n) were listed for each experiment. For MoTag related experiments, two-tailed Student's *t*-test was used to analyze significant differences between two group means. For rapamycin or blue light inducible dimerization and tetramerization, two-tailed Student's *t*-test was used to analyze significant difference. For all statistics, ***P < 0.001.

SUPPLEMENTARY NOTES

The relationship between F_{comet}/ $F_{cytosol}$ ratios and the oligomeric state of assayed proteins

Note: F_{comet}/ $F_{cytosol}$ = comet-to-cytosol ratio of mCh-[Protein]-MoTag2, MoTag2 = SxIP motif from $DST_{5474-5485}$
To clearly explain SxIP-EB1-MT interactions, herein, mCh-[Protein]-MoTag2 was replace by mCh-[Protein]-SxIP.

EB1 plays a central role in regulating the dynamics of the +TIP interaction networks[8, 9]. It contains an N-terminal calponin homology (CH) domain (binding to microtubule) and a unique C-terminal EB homology domain (EBH). EBH recognize SxIP motifs from binding partners such as APC, DST and STIM1[8, 9]. Thus, proteins bearing SxIP motifs can be indirectly recruited to microtubule plus ends through physically interacting with the EBH domain of EB1[8, 9].

The following equations (1-3) describe the binding of **mCh-[Protein]-SxIP** (abbr. **P**) to MT ends (**MT**):

$$[P] + [EB1] \underset{Kd_1}{\rightleftharpoons} [P\text{-}EB1] \quad (1) \qquad Kd_1 = \frac{[P]*[EB1]}{[P\text{-}EB1]}$$

$$[P\text{-}EB1] + [MT] \underset{Kd_2}{\rightleftharpoons} [P\text{-}EB1\text{-}MT] \quad (2) \qquad Kd_2 = \frac{[P\text{-}EB1]*[MT]}{[P\text{-}EB1\text{-}MT]}$$

$$[P] + [EB1] + [MT] \underset{Kd_1*Kd_2}{\rightleftharpoons} [P\text{-}EB1\text{-}MT] \quad (3) = (1) + (2) \qquad Kd_1*Kd_2 = \frac{[P]*[MT]*[EB1]}{[P\text{-}EB1\text{-}MT]}$$

[P] = [mCh-Protein-SxIP], [MT] = [Tubulin], K_d = Dissociation constant

If **P** is a monomeric protein (n = 1, n represents the oligomeric state):
The fraction (θ) of MT-bound **mCh-[Protein]-SxIP** could be expressed as (equation 4),

$$\theta_{[P\text{-}EB1\text{-}MT]} = \frac{[P\text{-}EB1\text{-}MT]}{[P\text{-}EB1\text{-}MT] + [P\text{-}EB1] + [P]} = \frac{[MT]*[EB1]}{[MT]*[EB1] + Kd_2*[EB1] + Kd_1*Kd_2} \quad (4)$$

Because SxIP-EB1 and EB1-MT share similar binding mechanisms[8, 10], their dissociation constants (K_{d1} and K_{d2}) have been reported to fall in a similar range with comparable values[8, 10]. SxIP-containing peptides have been shown to interact with EB1 with a 1:1 stoichiometry and the dissociation constant (K_{d1}) ranges from 0.5 μM to 10 μM (based on SxIP motifs from APC and MACF)[11-14]. For EB1 binding to MT, the dissociation constant (K_{d2}) was determined to be ~10-20 μM[15]. In living cells, the intracellular concentrations of tubulin ([MT]) and EB1 ([EB1]) have been determined to be 20~40 μM[16-18], and 0.25~2 μM[19-21], respectively. Given that [MT] generally is smaller than K_{d2}, the MT-bound fraction ($\theta_{[P\text{-}EB1\text{-}MT]}$) of monomeric **mCh-[Protein]-**

MoTag2 should be less than 0.5 (equation 5). Therefore, the ratio Fraction $_{(bound)}$/ Fraction $_{(unbound)}$, also expressed as $\theta_{[P\text{-}EB1\text{-}MT]}/(1- \theta_{[P\text{-}EB1\text{-}MT]})$, is expected to be no more than 1 (equation 6).

$$\theta_{[P\text{-}EB1\text{-}MT]} = \frac{[MT]*[EB1]}{[MT]*[EB1] + Kd_2*[EB1] + Kd_1*Kd_2} \leq 0.5 \qquad (n = 1) \qquad (5)$$

$$\frac{\theta_{[P\text{-}EB1\text{-}MT]}}{1 - \theta_{[P\text{-}EB1\text{-}MT]}} \leq 1 \qquad (n = 1) \qquad (6)$$

Experimentally, the comet-to-cytosol fluorescent ratio ($F_{comet}/F_{cytosol}$; proportional to Fraction $_{(bound)}$/ Fraction $_{(unbound)}$) of monomeric **mCh-[Protein]-SxIP**, which did not show overt MT plus end tracking, was measured as:

$$\frac{F_{comet}}{F_{cytosol}} \approx 1 \qquad \text{Monomeric mCh-[Protein]-SxIP} \qquad (7)$$

$$\text{Log}\left[\frac{F_{comet}}{F_{cytosol}}\right] = 0 \qquad \text{Monomeric mCh-[Protein]-SxIP} \qquad (8)$$

Under confocal microscopy, the monomeric mCh-[Protein]-SxIP showed an even cytosolic distribution of mCherry signals without MT tips tracking (or comet formation). The fluorescent signal ratio ($F_{comet}/ F_{cytosol}$) can be considered as Fraction $_{(bound)}$/ Fraction $_{(unbound)}$. Our measured values of $F_{comet}/ F_{cytosol}$ is ~ 1.0, corresponding to log [$F_{comet}/ F_{cytosol}$] as 0 (equations 7-8, Figure 1e, Figure 2g and Figure S9). Our observation for monomeric mCh-[Protein]-SxIP is also consistent with previous reports, in which monomeric EB1 or SxIP motif fails to show MT end plus tracking whereas dimerization could facilitate their MT end plus tracking[22].

If **P** assumes an oligomeric state and interact with EB1/MT in a non-cooperative manner (n > 1):
The oligomeric **mCh-[Protein]-SxIP** binds to MT as follows:

$$[P] + n*[EB1] \underset{Kd_1}{\rightleftharpoons} [P\text{-}EB1_n] \qquad (9) \qquad Kd_1 = \frac{[P]*[EB1]^n}{[P\text{-}EB1_n]}$$

$$[P\text{-}EB1_n] + n[MT] \underset{Kd_2}{\rightleftharpoons} [P\text{-}EB1_n\text{-}MT_n] \qquad (10) \qquad Kd_2 = \frac{[P\text{-}EB1]*[MT]^n}{[P\text{-}EB1_n\text{-}MT_n]}$$

$$[P] + [EB1_n] + [MT_n] \underset{Kd_1*Kd_2}{\rightleftharpoons} [P\text{-}EB1_n\text{-}MT_n] \qquad (11) = (9) + (10) \qquad Kd_1*Kd_2 = \frac{[P]*[EB1]^n*[MT]^n}{[P\text{-}EB1_n\text{-}MT_n]}$$

The fraction (θ) of MT-bound **mCh-[Protein]-SxIP**:

$$\theta_{[P\text{-}EB1n\text{-}MTn]} = \frac{[P\text{-}EB1_n\text{-}MT_n]}{[P\text{-}EB1_n\text{-}MT_n] + [P\text{-}EB1_n] + [P]} = \frac{[MT]^n * [EB1]^n}{[MT]^n * [EB1]^n + Kd_2 * [EB1]^n + Kd_1 * Kd_2} \quad (12)$$

The logarithmic relationship between the ratio (Fraction $_{(bound)}$/ Fraction $_{(unbound)}$) and oligomeric state (n) is expressed as,

$$\text{Log}\left[\frac{\theta_{[P\text{-}EB1n\text{-}MTn]}}{1-\theta_{[P\text{-}EB1n\text{-}MTn]}}\right] = \text{Log}\left[\frac{1}{\frac{1}{\theta_{[P\text{-}EB1n\text{-}MTn]}} - 1}\right] \quad \text{Log}_{10} \text{ was presented as Log} \quad (13)$$

$$= \text{Log}\left[\frac{1}{\frac{Kd_2}{[MT]^n} + \frac{Kd_1 * Kd_2}{[MT]^n * [EB1]^n}}\right] \quad (13)$$

$$= \text{Log}\left[\frac{[MT]^n * [EB1]^n}{Kd_2 * [EB1]^n + Kd_1 * Kd_2}\right] \quad (13)$$

It has been reported that the binding affinity of dimeric or oligomeric SxIP with EB1 is 100 ~ 1000-fold stronger than monomeric SxIP[12]. Therefore, $Kd_1 * Kd_2$ is considered to be $<< Kd_2 * [EB1]^n$ and thus negligible (when $n > 1$). Equation (13) can thus be transformed and simplified as,

$$Kd_1 * Kd_2 << Kd_2 * [EB1]^n \qquad n > 1 \quad (14)$$

$$\text{Log}\left[\frac{\theta_{[P\text{-}EB1n\text{-}MTn]}}{1-\theta_{[P\text{-}EB1n\text{-}MTn]}}\right] = \text{Log}\left[\frac{[MT]^n}{Kd_2}\right] = n * \text{Log}[MT] - \text{Log}[Kd_2] \quad (15)$$

$$a = \text{Log}[MT] \qquad b = \text{Log}[Kd_2]$$

$$\text{Log}\left[\frac{\theta_{[P\text{-}EB1n\text{-}MTn]}}{1-\theta_{[P\text{-}EB1n\text{-}MTn]}}\right] = n * a - b \quad (16)$$

Based on equation (16), the ratio of Fraction $_{(bound)}$/ Fraction $_{(unbound)}$ (in Log_{10} scale) and the oligomeric state (n) displays a linear relationship (a as the slope).

Experimentally, we have extensively examined the comet-to-cytosol fluorescent signals (F_{comet}/ $F_{cytosol}$; equivalent to Fraction $_{(bound)}$/ Fraction $_{(unbound)}$ in theory) of a dozen proteins with known oligomeric states (Figures 1-2, S5-6 and S8). The comet-to-cytosol ratio of fluorescent signals (in Log_{10} scale) plotted against the oligomeric states of indicated proteins indeed exhibited a linear relationship (equation 17, Figure 2g). The slope is determined to be 0.21. These results clearly justified the use of the aforementioned simple mathematical model to explain the acquired experimental data. In some low endogenous level of EB1, overexpression of EB1 can improve the visualization of comet formation.

$$\text{Log}\left[\frac{F_{comet}}{F_{cytosol}}\right] \approx K*(n-1) \quad (17)$$

K = 0.21; n represents the number of oligomeric state, n = 1, 2, 3 or 4

Based on our extensive testing on a Nikon confocal microscope or a DeltaVision Elite Microscope that only can acquire epifluorescence imgages (see Figure S13), the fusion of SxIP to proteins of interest, followed by convenient recording of $F_{comet}/ F_{cytosol}$, can rapidly discriminate proteins assembled as monomer, dimer, trimer and tetramer.

SUPPLEMENTARY FIGURES

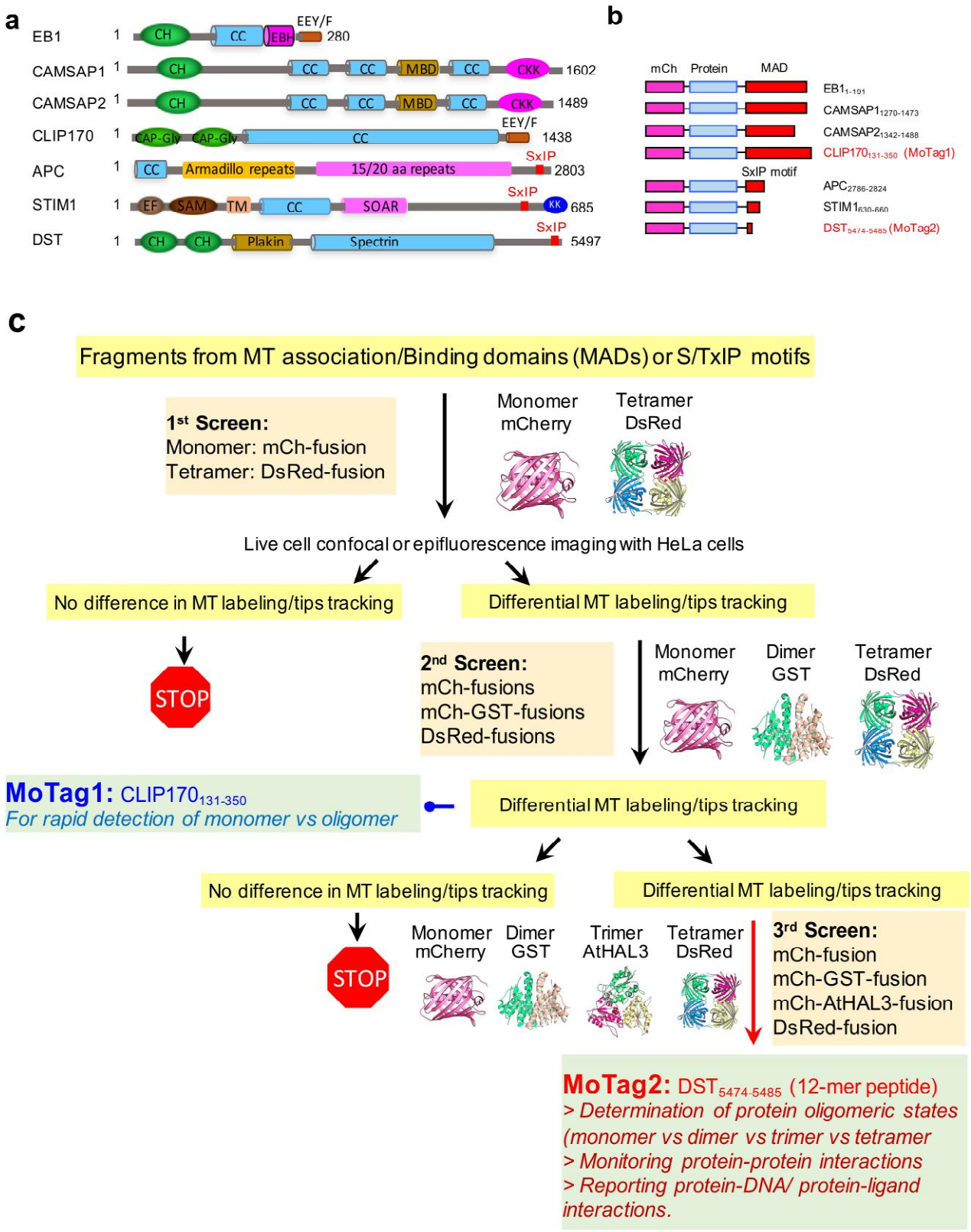


Figure S1 | Design and scheme for screening tags to enable quantitative discrimination of proteins assembled as monomer *versus* oligomer (dimer, trimer and tetramer).

(**a**) MT associated proteins to be engineered in the current study. See Figures S2-4 and S8 for full annotations.
(**b**) Domain architecture of proteins fused with MT association/binding domains (MADs).
(**c**) Schematic diagram showing the flowchart of screening procedures. In the initial screen, MT association/binding domains (MADs) or S/TxIP motifs isolated from proteins shown in panel (**a**) were fused with monomeric mCherry or tetrameric DsRed. Constructs showing differential MT labeling or comet formation were chosen for a second round of screening to discriminate monomer vs dimer (as GST fusion) vs tetramer. MoTag1 made of a fragment of the CAP-Gly domain of CLIP170 met the criteria to efficiently detect monomer from dimer or high-order oligomer. For the last round screening, candidates were fused to monomeric mCherry, dimeric GST, trimeric HAL3 or tetrameric DsRed. The 12-mer peptide $DST_{5474-5485}$ (MoTag2) emerged as the only construct that could determine the oligomeric state of proteins assembled as monomer, dimer, trimer and tetramer. We further illustrated the use of MoTag to probe protein self-association, as well as its heteromerization with substrates or ligands (*e.g.,* DNA for p53 and HSF1 [Figures S11-12] or Ca^{2+} for STIM1 [Figure 3g-i]).

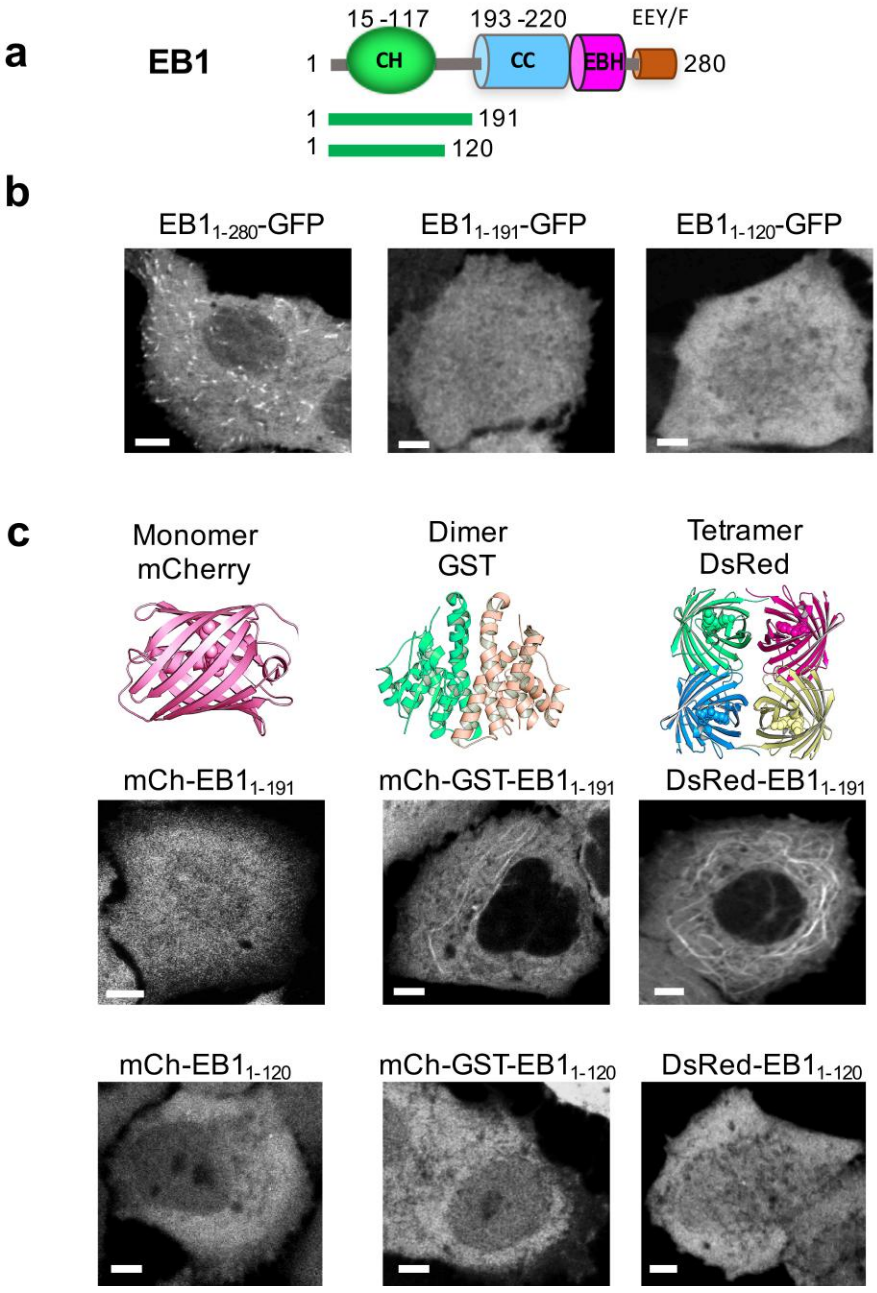


Figure S2 | Optimization and characterization of MT binding domain (CH) in EB1.

(**a**) Schematic of the domain organization of EB1. CH: calponin homology domain that binds to tubulin; CC: coiled-coil region; EBH: EB-homology domain that recognizes SxIP motif; EEY/F: C-terminal acidic tail. The fragments tested in the study was listed below the cartoon.

(**b-c**) Representative confocal images of HeLa cells expressing the indicated constructs. 3D structures of indicated proteins: monomeric mCherry (mCh), dimeric glutathione S-transferase (GST) and tetrameric DsRed were shown on the top. EB1$_{1-120}$ showed no difference in MT labeling, whereas EB1$_{1-280}$ displayed background MT labeling. EB1$_{1-191}$ showed oligomeric state-dependent staining of MT, but failed to detect the difference between dimer (mCh-GST-EB1$_{1-191}$) and tetramer (DsRed-EB1$_{1-191}$). Scale bar, 5 μm.

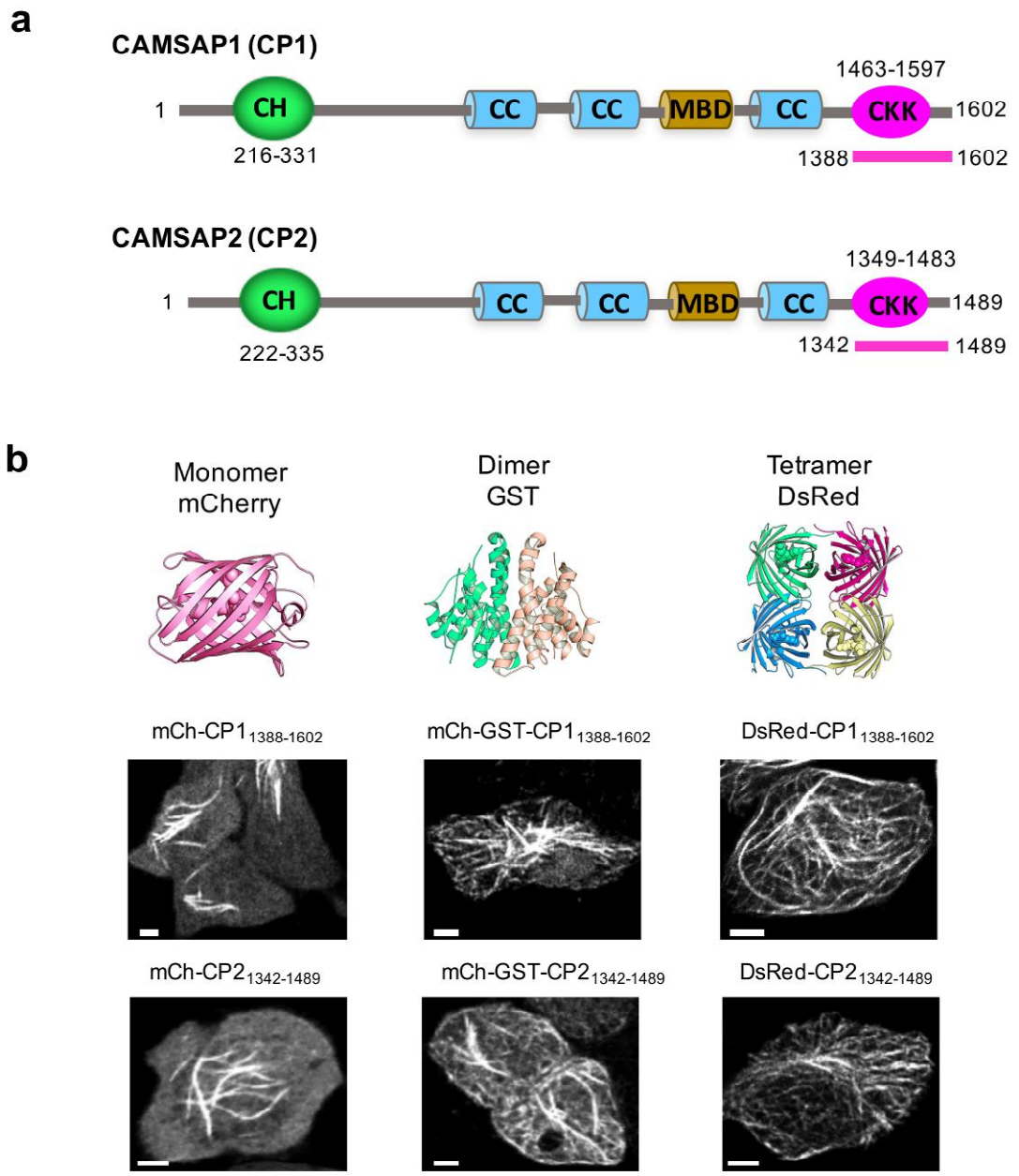


Figure S3 | Characterization of MT binding domain (CKK) isolated from CAMSAP1 and CAMSAP2.

(**a**) Cartoon representation of the domain architecture of CAMSAP1 (CP1) and CAMSAP2 (CP2). CH: calponin homology domain; CC: coiled-coil domain; MBD: the microtubule-binding domain. CKK: the carboxy-terminal (CAMSAP, KIAA1078 and KIAA1543) reserved domain that binds to MT.

(**b**) Representative confocal images of HeLa cells transfected with oligomeric proteins fused to mCh-CP1$_{1388-1602}$ or mCh-CP1$_{1342-1489}$. 3D structures of monomeric mCherry, dimeric GST and tetrameric DsRed were shown on the top. Although the images of monomeric mCh-CP$_{CKK}$ and dimeric GST-CP$_{CKK}$ (or tetrameric DsRed-CP$_{CKK}$) showed difference in the degrees of MT labeling, monomeric mCh-CP$_{CKK}$ displayed strong background MT binding, thus making it not suitable for further optimization. Scale bar, 5 µm.

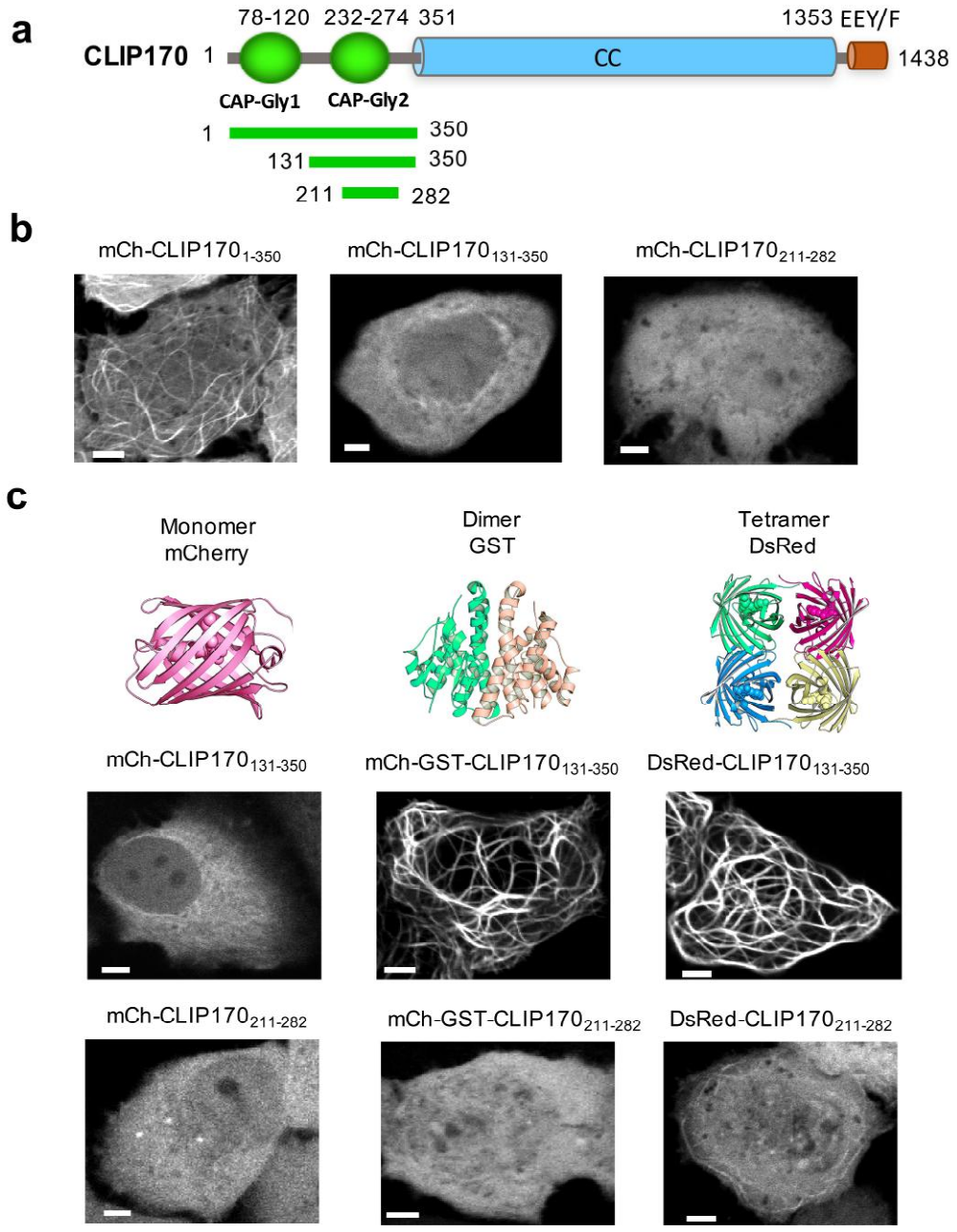


Figure S4 | Optimization and characterization of MT binding domain (CAP-Gly) derived from CLIP170.

(**a**) Domain organization of CLIP170. CAP-Gly: Cytoskeleton-Associated Proteins conserved glycine-rich domain, binding to tubulin; CC: coiled-coil domain; EEY/F: C-terminal acidic tail. The fragments used in the study were listed.

(**b**) Confocal images of HeLa cells expressing the indicated constructs.

(**c**) Representative confocal images of HeLa cells transfected with oligomeric proteins fused to mCh-CLIP170$_{131-350}$ or mCh-CLIP170$_{211-282}$. Monomeric mCh-CLIP170$_{131-350}$ and dimeric GST-CLIP170$_{131-350}$ (or tetrameric DsRed- CLIP170$_{131-350}$) showed significant difference in the degrees of MT labeling. CLIP170$_{131-350}$ was selected to discriminate monomer and oligomer, and was named as MoTag1. Scale bar, 5 μm.

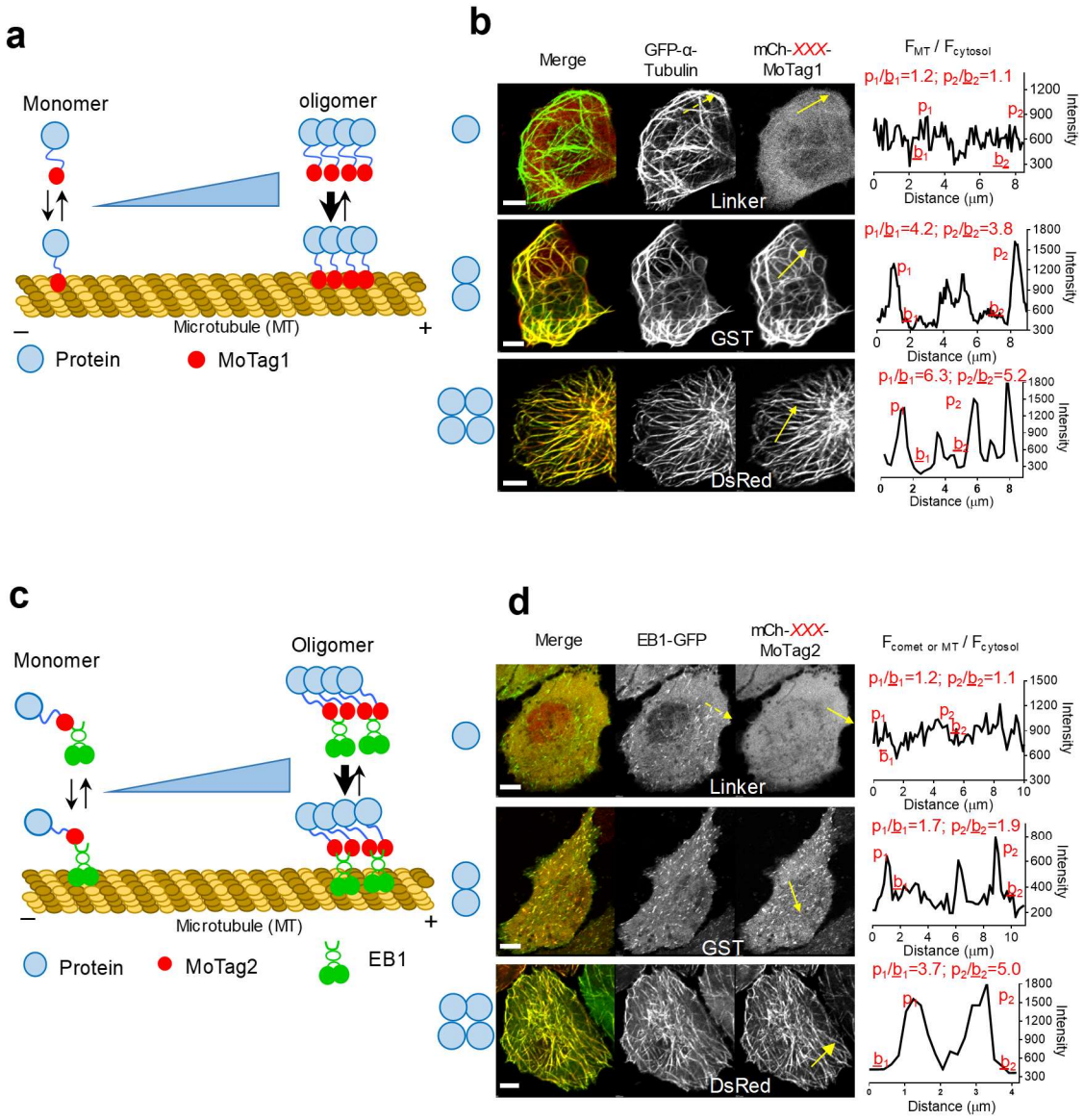


Figure S5 | Examples illustrating quantitative analysis on MT-to-cytosol or comet-to-cytosol fluorescent signals in HeLa cells.

(**a**) A simplified view of MoTag1 fused with different oligomeric proteins showing different degrees of MT labeling.

(**b**) (*Left*) Confocal images of HeLa cells coexpressing GFP-α-Tubulin (green; MT marker) and mCherry tagged (red) monomeric MoTag1 ($CLIP170_{131-350}$), dimeric GST-MoTag1 and tetrameric DsRed-MoTag1. mCherry-MoTag1 showed smooth distribution in the cytosol, while dimeric mCh-GST-MoTag2 or tetrameric DsRed-MoTag2 displayed excellent MT labeling. (*Right*) The MT-over-cytosol fluorescence intensity (F_{MT} / $F_{cytosol}$) ratio was acquired by using the "Plot profile" tool in the Nikon-element software package. The measured regions were indicated by the yellow arrow. The intensity versus to distance (μm) was plotted. Based on the graph, the maximum value of the peak was selected as F_{MT} (marked as p_n in the plot), while the neighboring cytosolic region (3~4 points) was averaged and selected as $F_{cytosol}$ (marked as b_n for

baseline). Herein, F_{MT} / $F_{cytosol}$ (p_n/b_n) values from two selected regions were shown. In a typical analysis, at least 6~8 regions per cell were randomly selected to yield an averaged value for F_{MT} / $F_{cytosol}$. Scale bar, 5 μm.

(**c**) Schematic showing MoTag2 binding to MT via EB1.

(**d**) (*Left*) Confocal images of HeLa cells coexpressing EB1-GFP (green) and mCherry tagged (red) monomeric MoTag2 (DST12, $DST_{5474-5485}$), dimeric GST- MoTag2 and tetrameric DsRed- MoTag2. For monomeric MoTag2 coexpressed with EB1-GFP, it showed an even cytosolic distribution. By contrast, EB1-GFP constantly tracked MT plus ends (Fig S2). Dimeric mCh-GST-MoTag2 showed good colocalization with EB1-GFP comets. For high-order oligomer fused with MoTag2 (DsRed), it seemed to synergize with EB1 to clearly mark the MT cytoskeleton as MoTag1 did. (Right) The comet-over-cytosol fluorescence intensity (F_{comet} / $F_{cytosol}$) ratio was calculated as described in panel b. Scale bar, 5 μm.

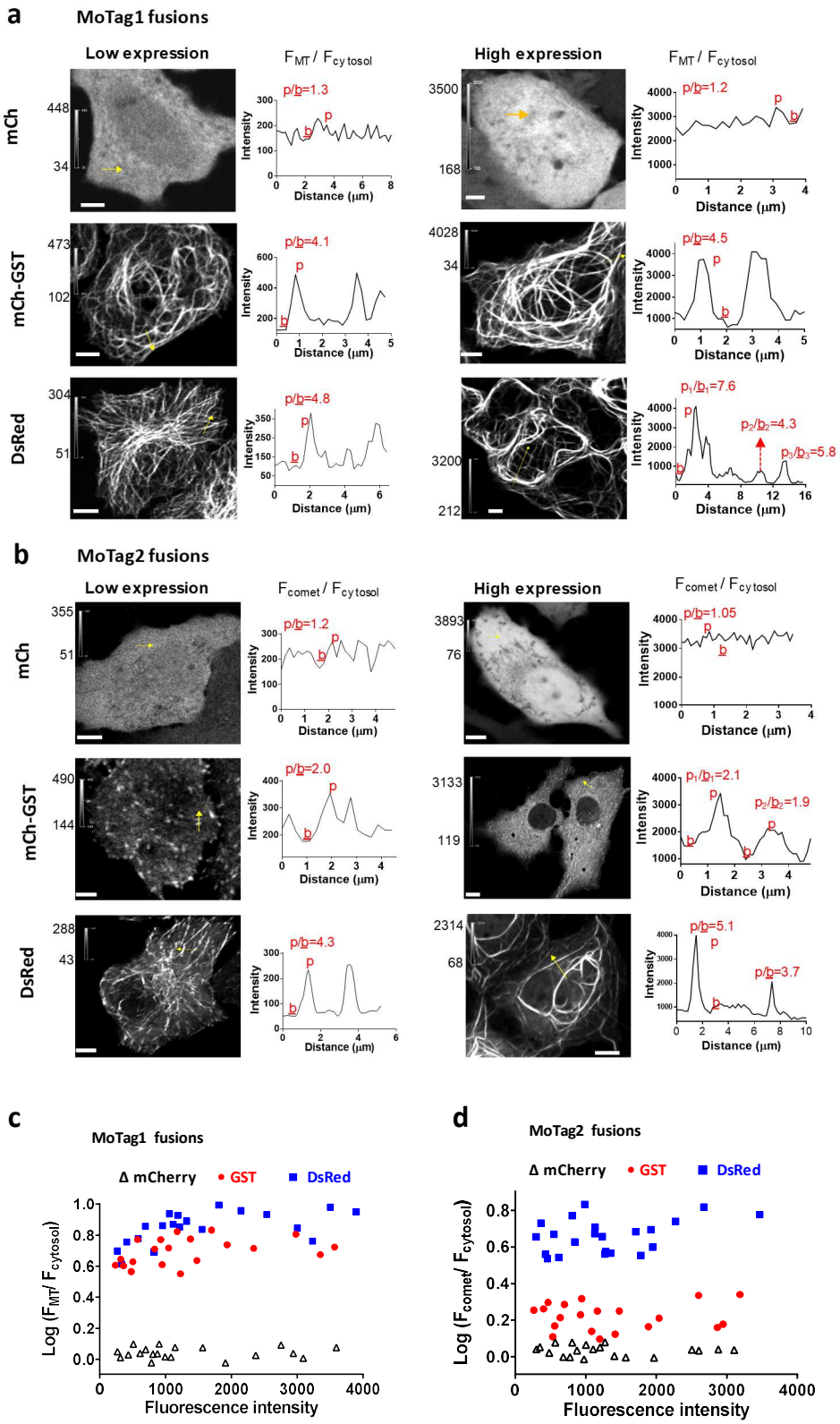


Figure S6 | Statistics of MT-over-cytosol or comet-over-cytosol fluorescence ratio ($F_{MT}/F_{cytosol}$, or $F_{comet}/F_{cytosol}$) of MoTags at varying expression levels. Scale bar, 5 μm.

(**a**) Representative confocal images of HeLa cells expressing MoTag1 fusions (monomeric mCh, dimeric GST and Tetrameric) at low (left) or high levels (right). The fluorescence intensity range was labeled in the top left of the image. Graph on the right showed the fluorescence intensity across the indicated arrows (randomly selected to obtain $F_{comet}/F_{cytosol}$ values). See Figure **S5** for more details.

(**b**) Representative confocal images of HeLa cells expressing MoTag2 fusions at low (left) or high expression levels (right). The current intensity range was labeled in the top left of the image. The random selected comets (arrows) were analyzed for obtaining $F_{MT}/F_{cytosol}$ values (graphs next to the confocal images). See Figure S5 for more details.

(**c-d**) The MT-to-cytosol ratio (**c**, MoTag1 fusions) or comet-to-cytosol (**d**, MoTag2 fusions) ratio of fluorescent signals (in Log10 scale) plotted against protein expression levels (indicated by the arbitrary fluorescent intensities). mCherry: black trianle; GST: red dot; DsRed: blue square. n = 20 cells.

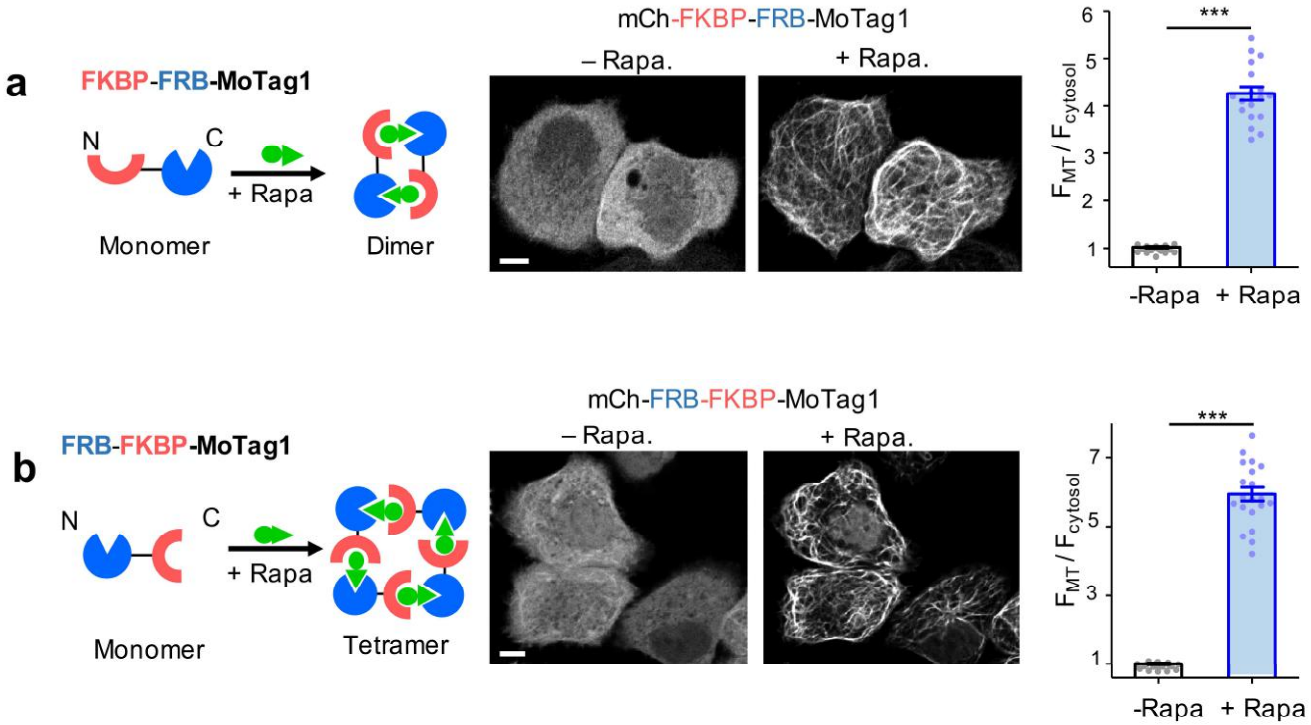


Figure S7 | Chemical inducible protein oligomerization monitored by MoTags.

(**a**) (Left) Schematic depicting rapamycin induced FKBP-FRB dimerization[5]. (Middle) Representative views showing confocal images of mCh-FKBP-FRB-MoTag1 in HeLa cells before and after addition of 5 μM rapamycin. (Right) Measurements of $F_{MT} / F_{cytosol}$ before and after rapamycin treatment in cells shown in the middle panel.

(**b**) (Left) Schematic depicting rapamycin induced FRB-FKBP tetramerization[5]. (Middle) Representative views showing confocal images of mCh-FRB-FKBP-MoTag1 in HeLa cells before and after addition of 5 μM rapamycin. (Right) Measurements of $F_{MT} / F_{cytosol}$ before and after rapamycin treatment in cells shown in the middle panel.

Data were acquired from 6-8 regions of interest from at least 21 cells from three independent experiments. Error bars denote s.e.m.. ***$P < 0.001$ (paired Student's *t*-test). Scale bar, 5 μm.

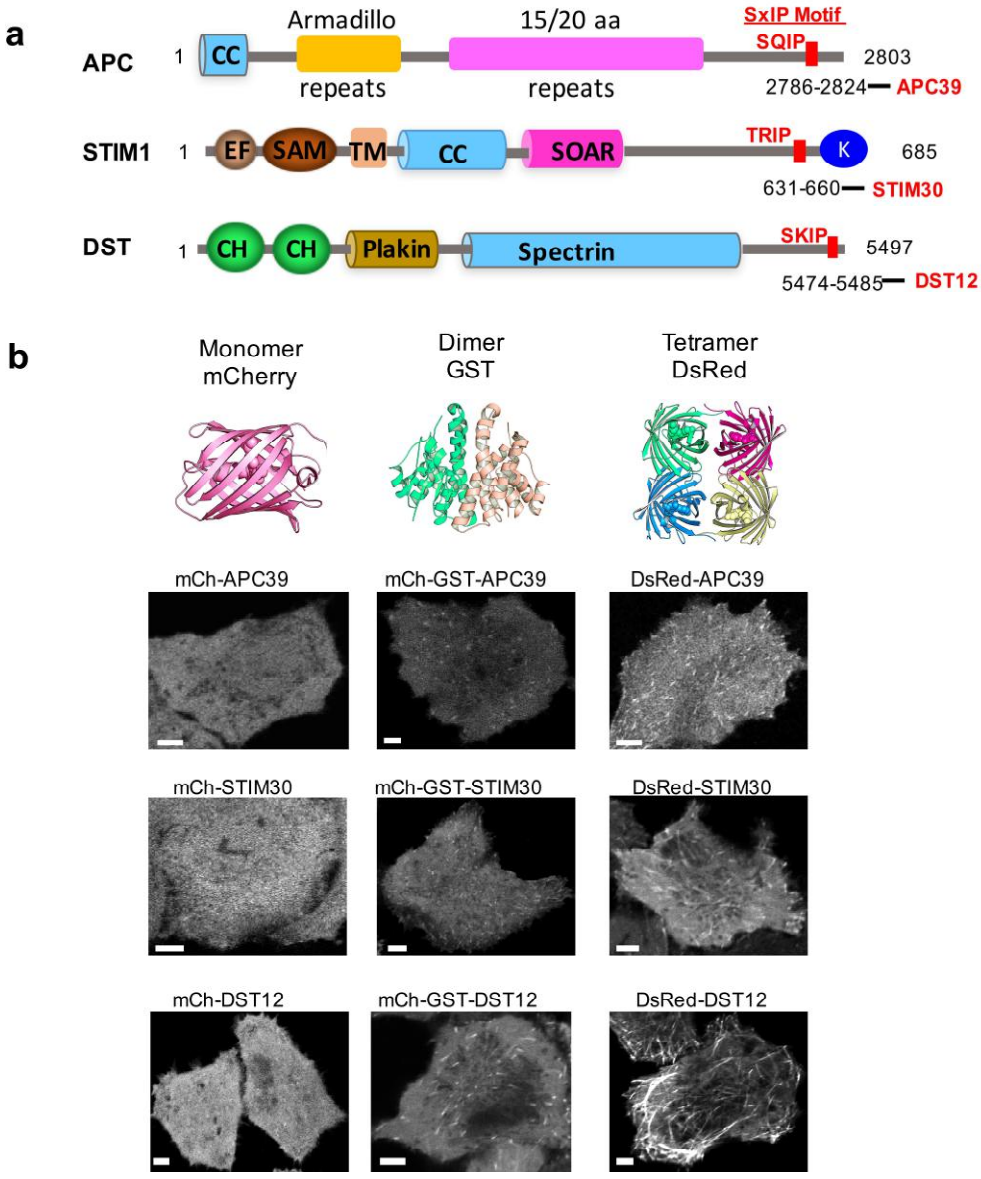


Figure S8| Characterization of EB1-binding SxIP motifs isolated from APC, STIM1 and DST.

(**a**) Domain architectures of the indicated EB1 binding proteins. APC, adenomatous polyposis coli (APC); DST, dystonin; STIM1, stromal interaction molecule 1; SxIP motif, an amino acid sequence motif of Ser-any amino acid-Ile-Pro, which is specifically recognized by the EB homology domain (EBH) of EBs. CC, coiled-coil; EF-SAM, EF-hand and sterile alpha motif; TM, transmembrane domain; SOAR, STIM1 Orai-activating region; CH: calponin homology domain. The selected SxIP sequence from APC, STIM1 and DST were indicated and named as APC39, STIM30 and DST12, respectively.

(**b**) Representative confocal images of HeLa cells transfected with oligomeric proteins fused to mCh-APC39 (top), mCh-STIM30 (middle), or mCh-DST12 (bottom). Compared with APC39 and STIM30, DST12 showed much stronger signals to enable clear discrimination among monomeric, dimeric and tetrameric versions. Hence, we named DST12 as MoTag2 and used it for further applications. Scale bar, 5 μm.

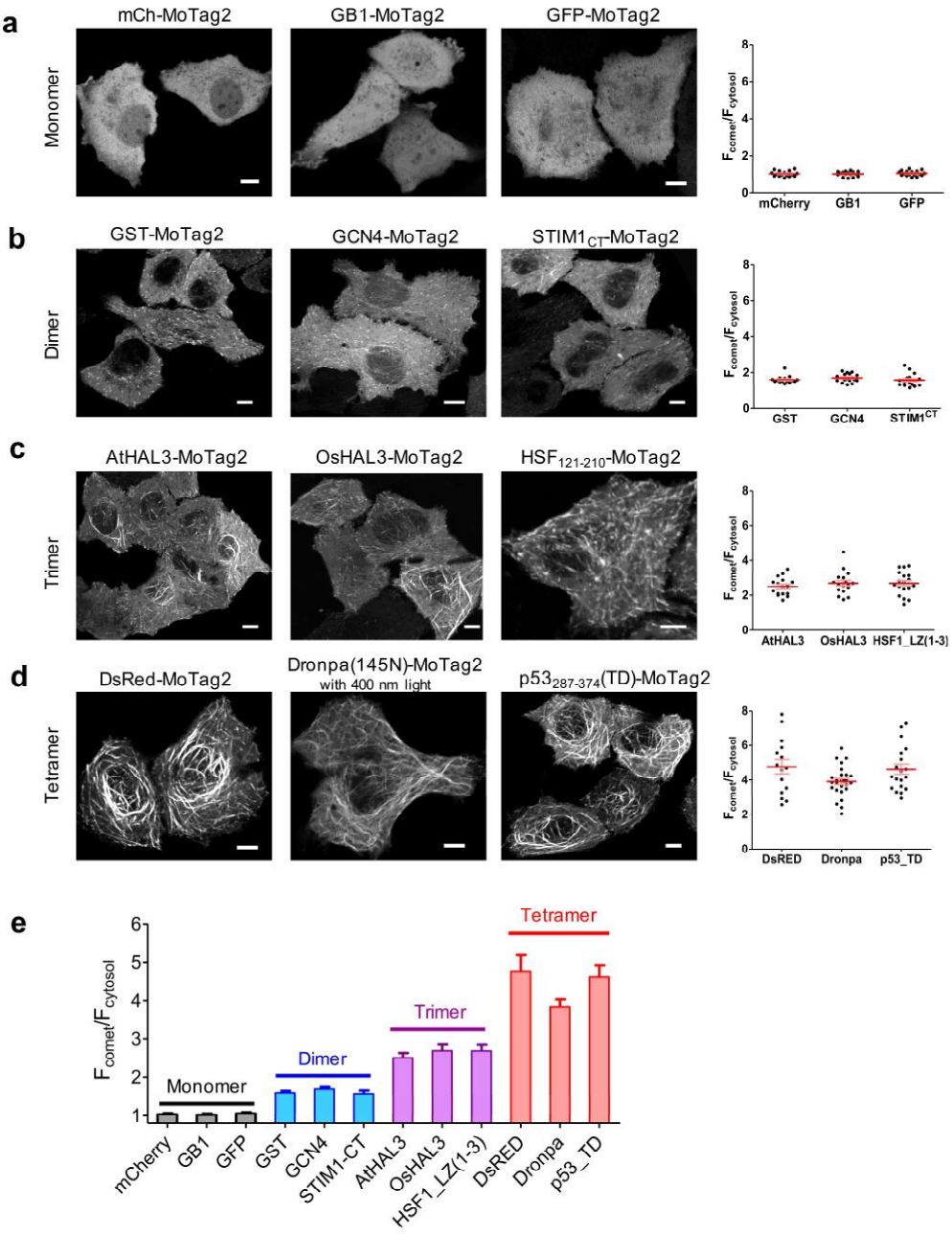
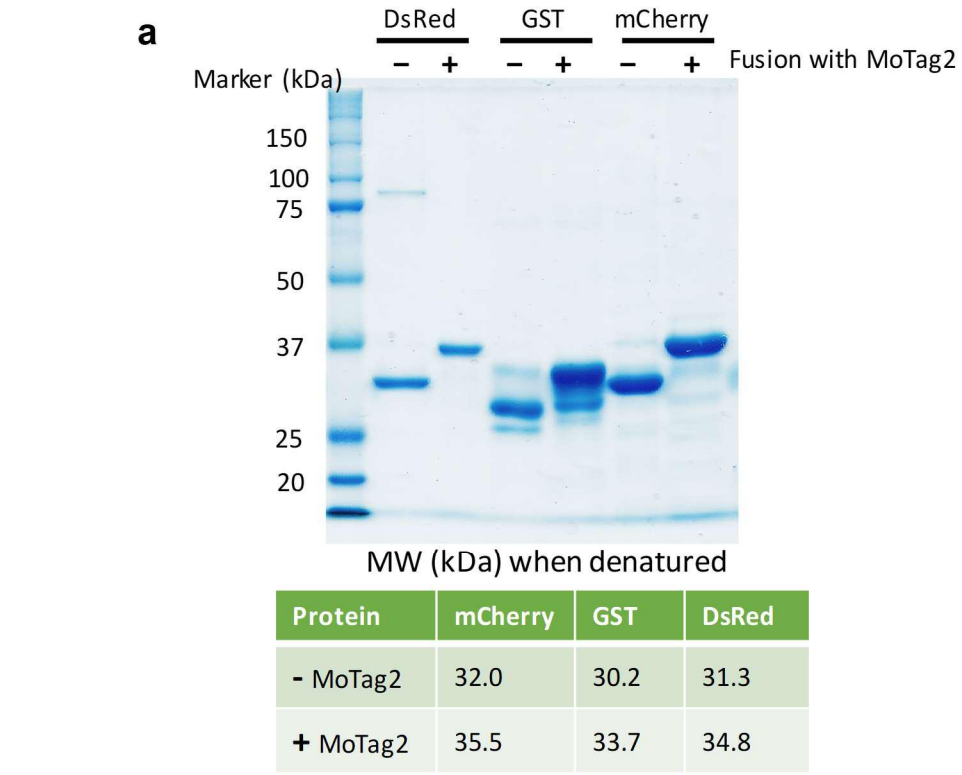


Figure S9 | MoTag2 as a fusion tag to induce comet formation and monitor the oligomeric states of proteins.

(**a**-**d**) *Left*, Representative confocal images of HeLa cells expressing proteins fused to mCh-MoTag2 ($DST_{5474-5485}$, DST12). The oligomeric proteins tracked MT plus ends through interacting with endogenous EB1. Proteins with higher order of oligomers (trimer or tetramer) exhibited stronger comet intensities or longer lengths of MT comets. *Right*, Quantification of the MT comet intensity over cytosolic fluorescence intensity.

(**e**) Quantification of the comet-over-cytosol fluorescence ratio ($F_{comet}/F_{cytosol}$) presented as bar graph for proteins assayed in the study. All data were acquired from 6-8 regions of interest from at least 16 cells, and presented as mean ± s.e.m.. Scale bar, 5 µm.



MW (kDa) when denatured

Protein	mCherry	GST	DsRed
- MoTag2	32.0	30.2	31.3
+ MoTag2	35.5	33.7	34.8

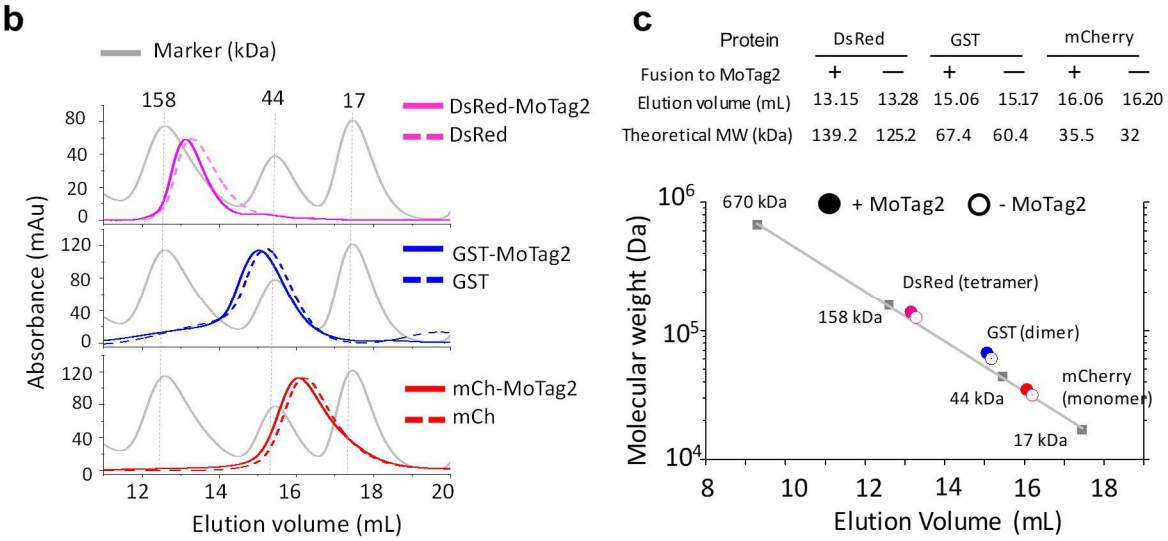


Figure S10 | Characterization of the oligomeric states of MoTag2 fusion proteins in solution.

(**a**) *Top*, Purified recombinant proteins, with or without fusion with MoTag2, were resolved on SDS-PAGE. Coomassie Brilliant Blue staining was applied to visualize the peak fraction for each protein after elution from size exclusion chromatography (SEC). *Bottom*, the calculated molecular weight (MW) of indicated proteins, with or without fusion to MoTag2.

(**b**) SEC elution profiles of indicated proteins (with or without fusion with MoTag2) after injection into a Superdex 200 10/300 GL column. Top: DsRed; Middle: GST; Bottom: mCherry. Solid line, MoTag2-fused proteins; dashed line, protein without MoTag2.

(**c**) Determination of the molecular sizes of the indicated proteins. Gel filtration standard with a lyophilized mixture of molecular weight markers (17/44/158/670 kDa; shown in gray line in panel **b**) was used to calibrate the SEC system when determining the size of proteins. The elution volume and MWs were listed on the top. Dot (●): MoTag2 fusions, Circle (○): free proteins. The measured MWs were very close to their theoretical MWs, implying that MoTag2 did not affect protein assembly.

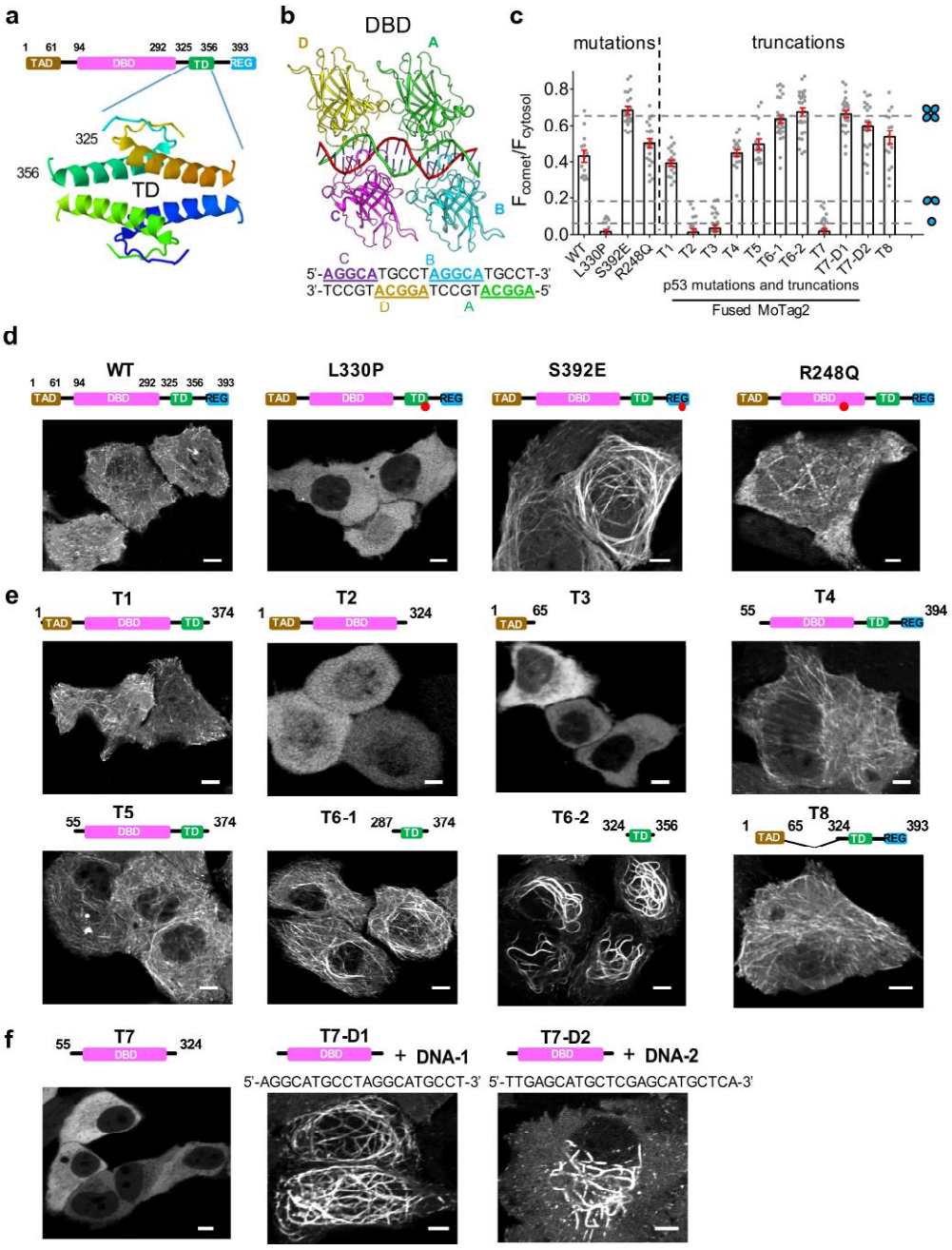


Figure S11 | Assessment of the effects of mutations or truncations on p53 oligomerization by MoTag2.

(**a**) Domain architecture of human p53 (top) and the 3D structure of the p53 tetramerization domain (TD; PDB entry: 1C26[23]). TAD, activation domain; DBD, DNA binding domain; REG, regulatory domain.

(**b**) The 3D structure of the p53 DNA binding domain (DBD) bound to DNA as a homo-tetramer (units A-D; PDB entry: 3KMD[24]). The nucleotide sequence of the bound DNA duplex was shown below the cartoon.

(**c**) Quantification of the $F_{comet}/F_{cytosol}$ ratio for p53 variants assayed in the study. The dashed lines represent the standard values corresponding to proteins assembled as monomer, dimer or tetramer (see Figure 3g and Figure S9e). Selected data were also presented in Figure 3b-f. At least 15 cells were selected from three independent experiments. The data were shown as mean ± s.e.m..

(**d-f**) Representative confocal images of HeLa cells expressing NES-mCh-p53 variants fused to MoTag2 ($DST_{5474-5485}$). An NES derived from the cAMP-dependent protein kinase inhibitor was added to minimize the nuclear accumulation of p53 variants. (**d**) WT and mutant p53. L330P is known to abolish the TD structure and maintains p53 as a monomer[25]; whereas S392E is reported to enhance p53 tetramerization[26]. The R248Q mutation tends to cause p53 aggregation[27, 28]. (**e**) p53 truncation or deletion variants. (**f**) p53 DBD in the absence or presence of two cognate DNA duplexes. 6 hours prior to confocal imaging, mCh-DBD-MoTag2 transfected HeLa cells were co-incubated with 10 μg DNA oligos in the presence of Lipofectamine 3000. All scale bars, 5 μm.

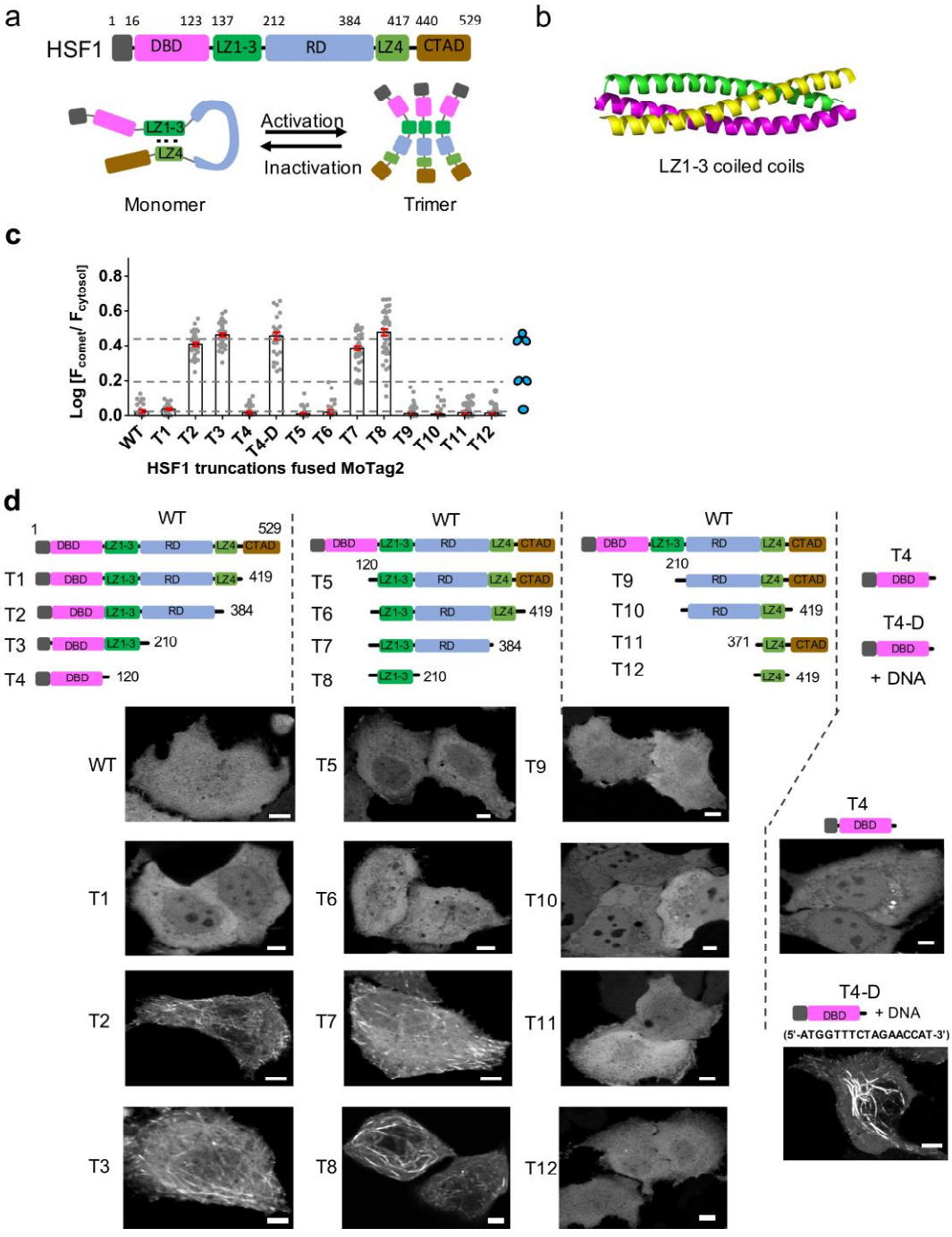


Figure S12 | Real-time assessment of the oligomeric states of human HSF1 truncations by MoTag2.

(**a**) The domain organization of HSF1. The 529-residue HSF1 protein comprises an N-terminal DNA-binding domain (DBD), leucine zipper regions 1-3 (LZ1-3, also named as HR-A/B: heptad repeat regions A and B), a regulatory domain (RD), leucine zipper domain 4 (LZ4, also named as HR-C), and a C-terminal regulatory domain (CTAD). The weakening of intramolecular coiled coil interactions between LZ1-3 and LZ4 could trigger the monomer-to-tetramer transition of HSF1.

(**b**) Crystal structure of the HSF1 homolog CtSkn7 as a homotrimer (PDB entry: 5D5Z[29]).

(**c**) Quantification of $F_{comet}/F_{cytosol}$ ratio for HSF1 variants assayed in the study. The dashed lines represent the standard values corresponding to proteins assembled as monomer, dimer or trimer (see Figure 2g and Figure S9e).

(**d**) Representative confocal images of HeLa cells expressing MoTag2 ($DST_{5474-5485}$) fused to mCherry-HSF1 truncation variants. MoTag2 fused HSF1 showed even distribution in cytosol, implying that the full length HSF1 existed as an inactivated monomeric state. By comparison, truncation of the LZ4 domain resulted in liberation of trimerization domain LZ1-3 to enable a trimeric assembly of truncated HSF1based on the $F_{comet}/F_{cytosol}$ ratio (e.g., constructs T2, T3, T7 and T8). Upon binding to its cognate DNA sequence, HSF1 undergoes monomer-to-oligomer transition (constructs T4 vs T4-D).

At least 27 cells were selected from three independent experiments. The data were shown as mean ± s.e.m.. Scale bars, 5 μm.

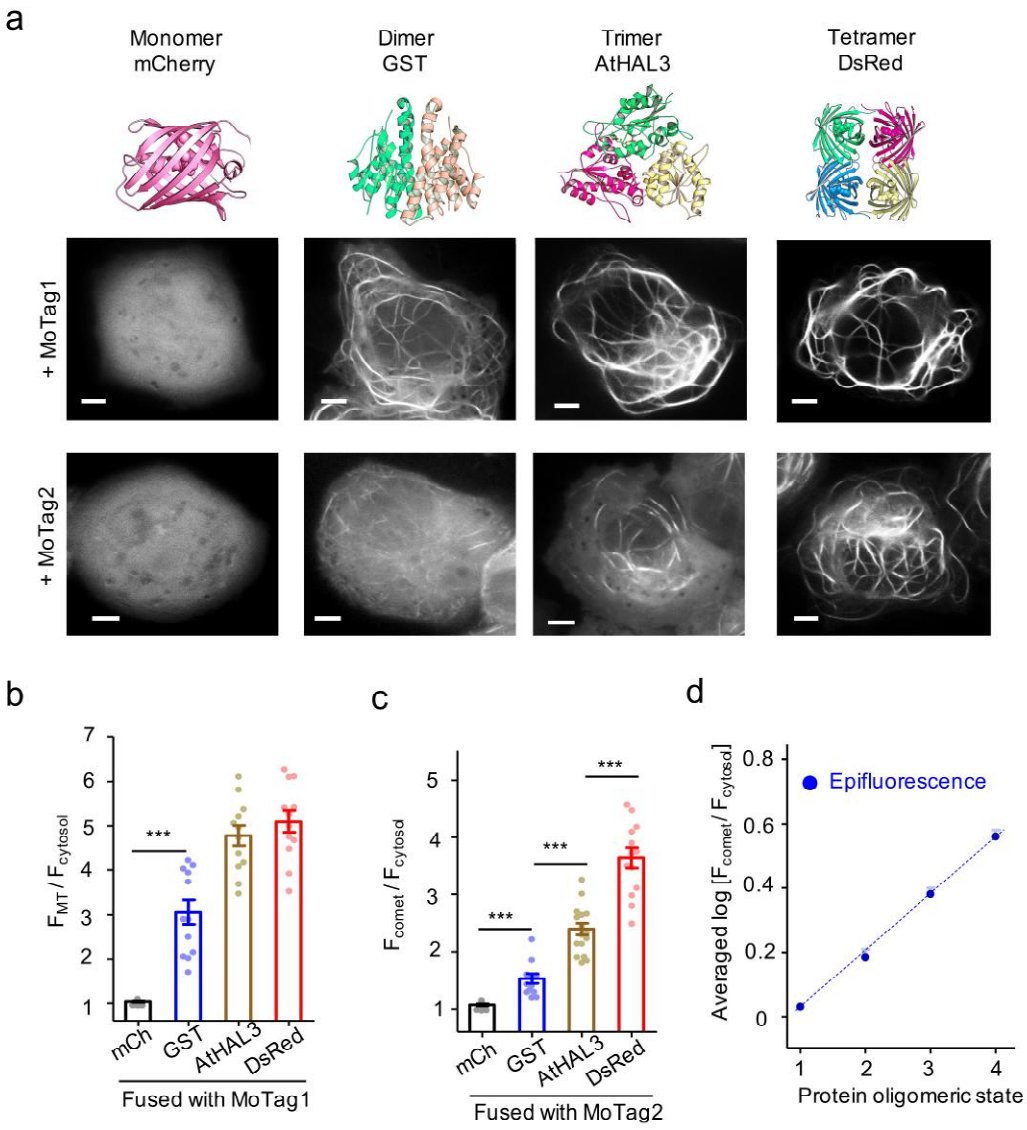


Figure S13 | Epifluorescence microscopy used to probe protein oligomeric states with MoTags.

(**a**) Epifluorescence images of HeLa cells expressing oligomeric proteins fused with mCh-MoTag1 ($CLIP170_{131-350}$, top panel) and mCh-MoTag2 ($DST_{5474-5485}$, bottom panel). Shown on the top were 3D structures of indicated proteins: monomeric mCherry (mCh), dimeric glutathione S-transferase (GST), trimeric halotolerance protein 3 (AtHAL3), and tetrameric DsRed. Scale bar, 5 µm.

(**b-d**) Quantification of the ratio of MT or comet intensity over cytosolic intensity of mCherry signals of MoTag1 (**b**) and MoTag2 (**c**) fusion proteins expressed in HeLa cells. (**d**) The comet-to-cytosol ratio (in Log10 scale) was plotted against the oligomeric states of indicated proteins and also showed a positive correlation. Each group represents the average of 6-8 regions of interest measured from transfected cells. n= 12 cells from three independent experiments. All data were shown as mean ± s.e.m. ***P<0.001 (two-tailed Student's t-test).

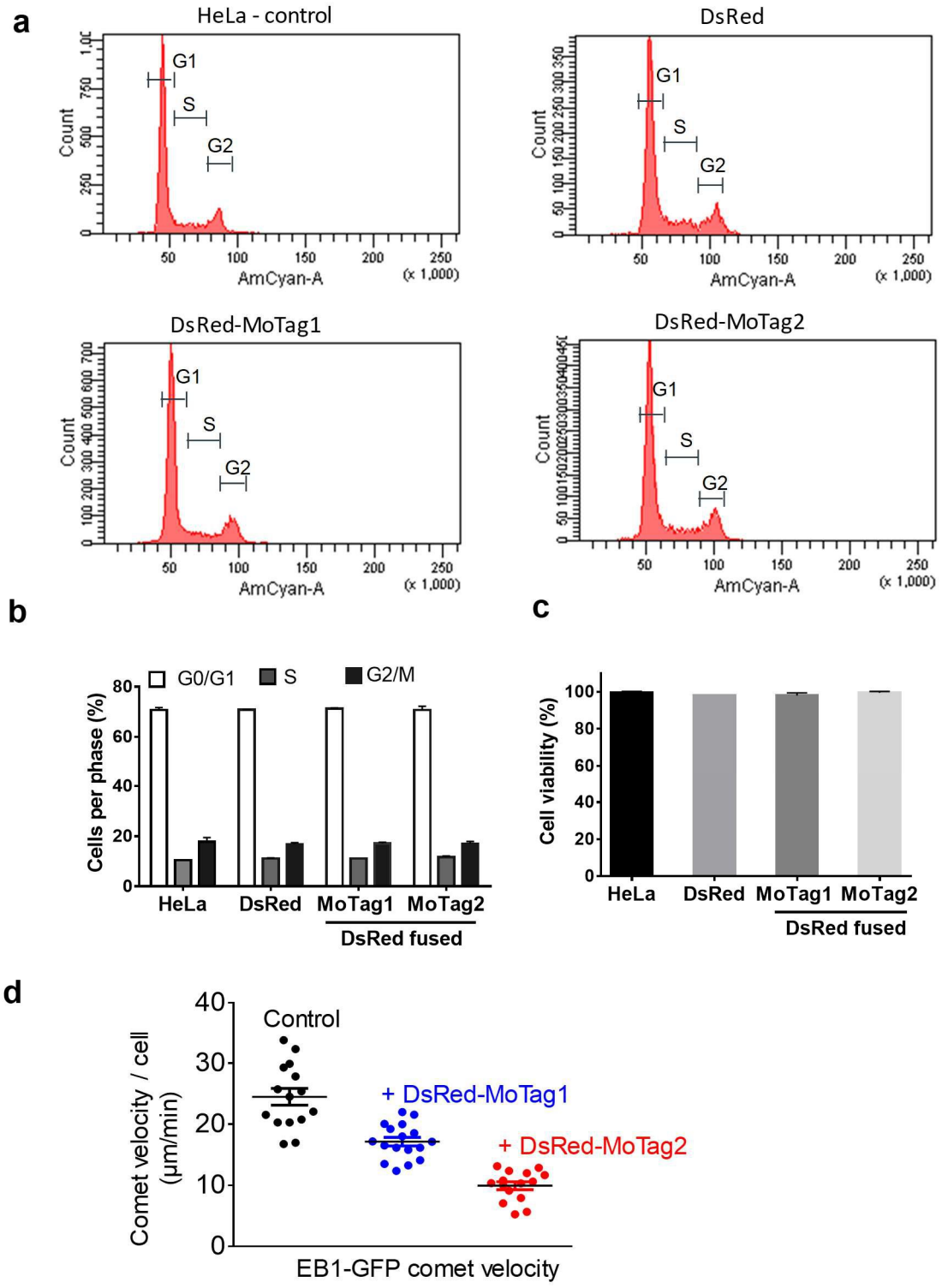


Figure S14 | Overexpression of MoTag-fused proteins in HeLa cells did not cause overt changes in cell cycle and cell viability, but slowed down the comet velocity of EB1-GFP.

(**a**) Cell cycle analysis of HeLa cells transfected with DsRed (control), and DsRed fused MoTag1 or MoTag2.

(**b**) Quantification of cell populations at each representative cell cycle stage.

(**c**) Quantification of cell viability by trypan blue staining. Data were presented as mean ± s.e.m. (n=3)

(**d**) Scatter plot summarizing the comet velocity of EB1-GFP in HeLa cells with coexpression of DsRed (control), DsRed-MoTag1 (blue dot) and DsRed-MoTag2 (red dot).

Table S1. Summary of reported dissociation constants (K_d) of homo-oligomers and chemical (FRB:FKBP) or optical dimerizers (iLID:sspB) used in this study.

Oligomeric states	Name	Dissociation constants (K_d)	Reference
Dimer	GST	5.1 µM	PMID: 18796433
Dimer	GST	0.34 µM	PMID: 15035604
Dimer	GST	0.16 µM	PMID:16681369
Dimer	GST	<< 1 nM	PMID: 19795889
Dimer	GCN4	~8 nM	PMID: 8993335 PMID: 7548036
Dimer	GCN4	2.5 µM	PMID: 7696274
Dimer	iLID/sspB, nano	132 nM	PMID: 25535392
Dimer	iLID/sspB, micro	0.8 µM	PMID: 25535392
Dimer	FRB/FKBP	12 nM	PMID: 15796538
Tetramer	DsRED	< 1 nM	PMID: 11724946 PMID: 11050231
Tetramer	Dronpa (145N)	9.9 µM	PMID: 28232577
Tetramer	p53	~ 1 nM	PMID: 20030809 PMID: 28760960

Based on our experimental results and published data, we conclude that MoTags are compatible with proteins with the oligomerization (self-association) affinities *ranging from 1 nM to at least 10 µM*.

SUPPLEMENTARY VIDEOS

Movie S1 | Rapamycin-induced dimerization of FKBP-FRB-MoTag1 led to obvious MT labeling. 5 μM rapamycin (at time point 50 sec) was added to HeLa cells expressing mCherry tagged FKBP-FRB-MoTag1.

Movie S2 | Rapamycin-induced tetramerization of FRB- FKBP-MoTag1 led to strong MT labeling. 5 μM rapamycin (at time point 35 sec) was added to HeLa cells expressing mCherry tagged FKBP-FRB-MoTag1.

Movie S3 | Rapamycin-induced dimerization of FKBP-FRB-MoTag2 led to noticeable tracking of MT plus ends. 5 μM rapamycin (at time point 30 sec) was added to HeLa cells expressing mCherry tagged FKBP-FRB-MoTag2.

Movie S4 | Rapamycin-induced tetramerization of FRB-FKBP-MoTag2 led to noticeable MT labeling. 5 μM rapamycin (at time point 15 sec) was added to HeLa cells expressing mCherry tagged FKBP-FRB-MoTag2. Scale bar, 10 μm.

Movie S5 | Blue light induced comet formation in HeLa cells co-expressing iLID-MoTag2 and sspB-MoTag2. The video showed the view of HeLa cells expressing mCh-iLID-MoTag2-P2A-mCh-sspB-MoTag2 responding to two repeated light-dark cycles (470 nm, 40 $\mu W/mm^2$).

Supplementary References

1. O'Shea, E.K., Klemm, J.D., Kim, P.S. & Alber, T. X-ray structure of the GCN4 leucine zipper, a two-stranded, parallel coiled coil. *Science* **254**, 539-544 (1991).
2. Sun, S.Y. et al. OsHAL3 mediates a new pathway in the light-regulated growth of rice. *Nature cell biology* **11**, 845-851 (2009).
3. Albert, A. et al. The X-ray structure of the FMN-binding protein AtHal3 provides the structural basis for the activity of a regulatory subunit involved in signal transduction. *Structure* **8**, 961-969 (2000).
4. Ma, G. et al. Inside-out Ca(2+) signalling prompted by STIM1 conformational switch. *Nat Commun* **6**, 7826 (2015).
5. Inobe, T. & Nukina, N. Rapamycin-induced oligomer formation system of FRB-FKBP fusion proteins. *Journal of bioscience and bioengineering* **122**, 40-46 (2016).
6. Guntas, G. et al. Engineering an improved light-induced dimer (iLID) for controlling the localization and activity of signaling proteins. *Proc Natl Acad Sci U S A* **112**, 112-117 (2015).
7. Kyung, T. et al. Optogenetic control of endogenous Ca(2+) channels in vivo. *Nature biotechnology* **33**, 1092-1096 (2015).
8. Akhmanova, A. & Steinmetz, M.O. Control of microtubule organization and dynamics: two ends in the limelight. *Nat Rev Mol Cell Biol* **16**, 711-726 (2015).
9. Akhmanova, A. & Steinmetz, M.O. Microtubule +TIPs at a glance. *Journal of cell science* **123**, 3415-3419 (2010).
10. Akhmanova, A. & Steinmetz, M.O. Tracking the ends: a dynamic protein network controls the fate of microtubule tips. *Nat Rev Mol Cell Biol* **9**, 309-322 (2008).
11. Lesniewska, K., Warbrick, E. & Ohkura, H. Peptide aptamers define distinct EB1- and EB3-binding motifs and interfere with microtubule dynamics. *Mol Biol Cell* **25**, 1025-1036 (2014).
12. Buey, R.M. et al. Sequence determinants of a microtubule tip localization signal (MtLS). *The Journal of biological chemistry* **287**, 28227-28242 (2012).
13. Honnappa, S. et al. An EB1-binding motif acts as a microtubule tip localization signal. *Cell* **138**, 366-376 (2009).
14. Honnappa, S., John, C.M., Kostrewa, D., Winkler, F.K. & Steinmetz, M.O. Structural insights into the EB1-APC interaction. *The EMBO journal* **24**, 261-269 (2005).
15. Zhu, Z.C. et al. Interactions between EB1 and microtubules: dramatic effect of affinity tags and evidence for cooperative behavior. *The Journal of biological chemistry* **284**, 32651-32661 (2009).
16. Lodish, H. Molecular Cell Biology. *Book*, 763 (2008).
17. Yvon, A.M., Wadsworth, P. & Jordan, M.A. Taxol suppresses dynamics of individual microtubules in living human tumor cells. *Mol Biol Cell* **10**, 947-959 (1999).
18. Gard, D.L. & Kirschner, M.W. Microtubule assembly in cytoplasmic extracts of Xenopus oocytes and eggs. *J Cell Biol* **105**, 2191-2201 (1987).
19. Alberico, E.O. et al. Biochemical evidence that human EB1 does not bind preferentially to the microtubule seam. *Cytoskeleton* **70**, 317-327 (2013).
20. Kronja, I., Kruljac-Letunic, A., Caudron-Herger, M., Bieling, P. & Karsenti, E. XMAP215-EB1 interaction is required for proper spindle assembly and chromosome segregation in Xenopus egg extract. *Mol Biol Cell* **20**, 2684-2696 (2009).
21. Tirnauer, J.S., Grego, S., Salmon, E.D. & Mitchison, T.J. EB1-microtubule interactions in Xenopus egg extracts: role of EB1 in microtubule stabilization and mechanisms of targeting to microtubules. *Mol Biol Cell* **13**, 3614-3626 (2002).
22. Slep, K.C. & Vale, R.D. Structural basis of microtubule plus end tracking by XMAP215, CLIP-170, and EB1. *Mol Cell* **27**, 976-991 (2007).

23. Jeffrey, P.D., Gorina, S. & Pavletich, N.P. Crystal structure of the tetramerization domain of the p53 tumor suppressor at 1.7 angstroms. *Science* **267**, 1498-1502 (1995).
24. Chen, Y., Dey, R. & Chen, L. Crystal structure of the p53 core domain bound to a full consensus site as a self-assembled tetramer. *Structure* **18**, 246-256 (2010).
25. Kawaguchi, T. et al. The relationship among p53 oligomer formation, structure and transcriptional activity using a comprehensive missense mutation library. *Oncogene* **24**, 6976-6981 (2005).
26. Nichols, N.M. & Matthews, K.S. Human p53 phosphorylation mimic, S392E, increases nonspecific DNA affinity and thermal stability. *Biochemistry* **41**, 170-178 (2002).
27. Ano Bom, A.P. et al. Mutant p53 aggregates into prion-like amyloid oligomers and fibrils: implications for cancer. *The Journal of biological chemistry* **287**, 28152-28162 (2012).
28. Xu, J. et al. Gain of function of mutant p53 by coaggregation with multiple tumor suppressors. *Nat Chem Biol* **7**, 285-295 (2011).
29. Neudegger, T., Verghese, J., Hayer-Hartl, M., Hartl, F.U. & Bracher, A. Structure of human heat-shock transcription factor 1 in complex with DNA. *Nature structural & molecular biology* **23**, 140-146 (2016).