Published in final edited form as:
Nat Chem Biol. 2007 December ; 3(12): 779–784. doi:10.1038/nchembio.2007.49.

Surveying polypeptide and protein domain conformation and association with FlAsH and ReAsH

Nathan W Luedtke[1,3], **Rachel J Dexter**[1], **Daniel B Fried**[1], and **Alanna Schepartz**[1,2]

1*Department of Chemistry, Yale University, 225 Prospect Street, New Haven, Connecticut 06520-8107, USA.*

2*Department of Chemistry and Molecular, Cellular, and Developmental Biology, Yale University, 219 Prospect Street, New Haven, Connecticut 06520-8107, USA.*

Abstract

Recombinant polypeptides and protein domains containing two cysteine pairs located distal in primary sequence but proximal in the native folded or assembled state are labeled selectively *in vitro* and in mammalian cells using the profluorescent biarsenical reagents FlAsH-EDT_2 and ReAsH-EDT_2. This strategy, termed bipartite tetracysteine display, enables the detection of protein-protein interactions and alternative protein conformations in live cells. As proof of principle, we show that the equilibrium stability and fluorescence intensity of polypeptide–biarsenical complexes correlates with the thermodynamic stability of the protein fold or assembly. Destabilized protein variants form less stable and less bright biarsenical complexes, which allows discrimination of live cells expressing folded polypeptide and protein domains from those containing disruptive point mutations. Bipartite tetracysteine display may provide a means to detect early protein misfolding events associated with Alzheimer's disease, Parkinson's disease and cystic fibrosis; it may also enable high-throughput screening of compounds that stabilize discrete protein folds.

The physical phenomenon known as fluorescence has revolutionized cell biology, and its positive impact on molecular medicine will continue to develop with time. Driving this revolution is the green fluorescent protein (GFP)[1–3] and a rainbow of natural and engineered fluorescent protein (FP) variants that can be fused at the genetic level to proteins expressed in cell cultures and even whole animals[3]. Methodologies based on Förster resonance energy transfer (FRET) that use two or more FPs with overlapping absorption and emission spectra are used frequently as sensors to probe dynamic and complex processes including protein association, metal ion binding, conformational changes and post-translational modifications[4]. However, virtually all sensors based on FRET between FP color variants show dynamic changes in fluorescence intensity (typically 20–50%) that can be smaller than normal variations in cell-to-cell intensity due to differential FP expression[5] and that may be influenced significantly by Mg^{2+}-ATP fluctuation[6]. Moreover, the steric bulk and slow folding of FPs limits their spatial and kinetic resolution[3,7].

Correspondence should be addressed to A.S. (E-mail: alanna.schepartz@yale.edu).
[3]Present address: Universität Zürich, Organisch-chemisches Institut, Winterthurerstrasse 190, CH-8057 Zürich, Switzerland.

AUTHOR CONTRIBUTIONS

N.W.L. contributed to experimental design, fluorophore synthesis, *in vitro* binding analyses, CD and manuscript drafting; R.J.D. contributed to experimental design and live cell experiments; D.B.F. contributed to *in vitro* binding analyses, CD and characterization of polypeptide-FlAsH complexes; A.S. contributed to experimental design and manuscript drafting.

Accession codes. Protein Data Bank: PDB coordinates used to generate images of aPP, Zip4, GCN4 and Jun can be found under 1PPT, 1LE3, 2ZTA and 1JUN, respectively. All structures were deposited as part of previous studies.

Published online at http://www.nature.com/naturechemicalbiology

An alternative strategy to fluorescently label recombinant proteins relies on sequence-selective protein binding by the "profluorescent" biarsenical dyes 4,5-bis(1,3,2-dithiarsolan-2-yl) fluorescein (FlAsH-EDT_2, 1) and 4,5-bis(1,3,2-dithiarsolan-2-yl)resorufin (ReAsH-EDT_2, 2) (Fig. 1)[8,9]. These cell-permeable small molecules selectively label recombinant proteins containing a CCPGCC tag in the cytosol[8], endoplasmic reticulum[10] and plasma membrane[11,12] by virtue of a thiol-arsenic ligand exchange reaction that converts the nonfluorescent 1,2-ethanedithiol (EDT)-bound forms of FlAsH and ReAsH into highly fluorescent protein-bound complexes[13]. Tetracysteine incorporation and subsequent biarsenical labeling was shown to be less disruptive to protein function than the expression of analogous FP fusion proteins in studies of the β_2-adrenergic receptor[11], yeast β-tubulin[14], G protein–coupled receptor (GPCR) activation[12], protein kinase A (PKA) subunit C-α (ref. 15) and the type III pathogenic effectors IpaB and IpaC (ref. 16). Biarsenical tetracysteine labeling is also compatible with functional assays that detect protein misfolding and aggregation in bacteria[17–19]. It is well established that addition or deletion of even a single amino acid to or from the optimized Pro-Gly sequence separating the two cysteine pairs substantially destabilizes the resulting biarsenical complexes (by up to 23,000-fold) and decreases their quantum yield (up to 500%)[13]. One interpretation of this immutability supports exclusive use of the tetracysteine motif CCPGCC as an encodable FlAsH or ReAsH tag. To us, this sensitivity suggested that the tetracysteine motif could function as a reporter of protein conformation or protein-protein binding if displayed in a bipartite mode, with two Cys-Cys pairs located distal in primary sequence (or on separate protein strands) but proximal in the folded or assembled state. If so, then bulky FPs used in cellular fluorescence assays and FRET-based sensors could be replaced by just two Cys-Cys pairs that, when assembled into close proximity (about 7 Å), bind to FlAsH or ReAsH and form a fluorescent complex.

We selected four structurally characterized polypeptides and protein domains to evaluate whether the Pro-Gly sequence within the linear tetracysteine motif could be replaced with one or more folded proteins while maintaining biarsenical affinity and fluorescence intensity. The feasibility of intramolecular bipartite tetracysteine display was evaluated using avian pancreatic polypeptide (aPP)[20] and Zip4 (ref. 21)—two well-folded polypeptides that were modified to contain one half of the linear tetracysteine motif at each terminus. Intermolecular bipartite tetracysteine display was evaluated using the protein-protein dimerization domains from the basic region leucine zipper (bZIP) proteins GCN4 (ref. 22) and Jun (ref. 23), which were modified to contain a single dicysteine motif (Fig. 1b,c).

Fluorescence titrations were performed to quantify the apparent affinity of each construct for FlAsH and ReAsH and to determine the relative brightness of the resulting complexes (Fig. 2). Each biarsenical dye (25 nM) was incubated with increasing concentrations of synthetic peptide in a buffered solution of 1 mM EDT (3), and the fluorescence emission was monitored (Fig. 2). Dithiols such as EDT and British anti-Lewisite (BAL, 4) compete with peptides containing di- or tetracysteine motifs for the biarsenical to lower the apparent affinity of these binding interactions and reduce nonspecific binding *in vitro* and in cell-based experiments[9]. Under these conditions Zip4 and Jun bound FlAsH to form complexes with apparent dissociation constants (K_{app}) comparable to those of analogous complexes containing an optimized linear tetracysteine sequence. aPP and GCN4 formed slightly less stable complexes with FlAsH characterized by values of K_{app} that were 12- and 50-fold higher, respectively (Fig. 1 and Fig. 2a). Although finite, these differences in K_{app} are small when compared with the 80,000-fold increase in K_{app} observed upon replacing the central Pro-Gly in the sequence WDCCPGCCK with Asp-Glu-Ala (ref. 13). Notably, the relative brightnesses (relative brightness = quantum yield × extinction coefficient) of the FlAsH complexes of Jun, aPP and GCN4 at saturation (Fig. 2) are comparable to that of the optimized linear tetracysteine complex.

The apparent affinity of each polypeptide for ReAsH was similar to its affinity for FlAsH (Fig. 1), although the relative fluorescence intensities of the resulting ReAsH complexes were lower, which is consistent with the three-fold lower quantum yields reported for ReAsH complexes relative to analogous FlAsH-containing complexes[8,13] (Fig. 2b). We note that the K_{app} values measured for the complexes containing Zip4 were approximately the same as for complexes containing the optimized linear tetracysteine control. However, the Zip4 complexes were less bright (Fig. 2), presumably because of tryptophan-mediated self-quenching[24]. Taken together, these *in vitro* experiments verify that well-folded polypeptides and protein-protein binding domains, when appropriately modified for bipartite tetracysteine display, bind biarsenicals with high affinity and can form complexes with fluorescence intensities that rival the optimized tetracysteine motif.

To determine whether the FlAsH complexes of aPP, Zip4, GCN4 and Jun have native-like folds and the expected stoichiometries, we performed preparative-scale binding reactions at concentrations at or above the equilibrium dissociation constant of each complex (20 μM protein and 100 μM FlAsH). FlAsH-containing complexes were purified by HPLC and analyzed by MALDI-TOF mass spectrometry (Supplementary Table 1 online) and circular dichroism (Supplementary Table 2 online). Under these conditions, aPP and Zip4 formed only 1:1 protein–FlAsH complexes with the expected molecular masses of 5,031 and 3,493 Da, respectively. GCN4 and Jun formed 2:1 protein–FlAsH complexes with the expected molecular masses of 7,792 and 9,796 Da, respectively. In no case was the intermediate 1:1:1 polypeptide–FlAsH–EDT complex observed. The circular dichroism (CD) spectrum of each wild-type polypeptide–FlAsH complex was similar to that of the analogous unmodified polypeptide (Fig. 3). Consistent with the reported dimerization of these domains[22,23], the CD spectra of unmodified Jun and GCN4 were concentration-dependent (apparent K_d = 10 μM), whereas the corresponding isolated 2:1 FlAsH complexes were not (Supplementary Fig. 1 online). Importantly, intensities characteristic of the predominant secondary structures (the α-helix signal at 222 nm for aPP, GCN4 and Jun, and the exciton-coupled signal at 230 nm for Zip4) changed minimally upon FlAsH modification. Full-spectrum analysis of the CD spectra using SELCON3, CONTINLL and CDSSTR also indicates little (if any) change in secondary structure upon FlAsH binding (Supplementary Table 2)[25]. Under these conditions, therefore, FlAsH binds these polypeptides to form complexes whose secondary structures closely resemble those of the wild-type polypeptide or protein domain. Moreover, the quantum yields (φ) of the isolated FlAsH complexes of aPP, GCN4 and Jun (0.65 ± 0.01, 0.68 ± 0.01 and 0.54 ± 0.04, respectively) compare well with the quantum yield of GFP (0.58–0.79)[3] and with that of conventional tetracysteine–FlAsH complexes (0.44–0.71)[13].

To evaluate whether bipartite tetracysteine display can differentiate folded and misfolded polypeptides *in vitro*, we constructed variants of aPP and Zip4 containing one or more destabilizing point mutations (Fig. 1c). CD analysis confirmed that these variants are largely unstructured, with little or no α-helix signal at 222 nm for aPP variants aPP^{F24P} and $aPP^{F24P,Y31P}$ and no exciton-coupled signal at 230 nm for the β-hairpin variant $Zip4^{W9,16A}$ (Fig. 3a,b). Like the analogous wild-type constructs, aPP^{F24P}, $aPP^{F24P,Y31P}$ and $Zip4^{W9,16A}$ each formed only 1:1 FlAsH complexes whose CD spectra indicate no significant change in α-helical (in the case of aPP^{F24P} and $aPP^{F24P,Y31P}$) or β-sheet (in the case of $Zip^{4W9,16A}$) secondary structure at 222 and 230 nm due to FlAsH coordination (Fig. 3a,b and Supplementary Table 2). Equilibrium FlAsH and ReAsH binding experiments in the presence of 1 mM EDT revealed that the relative stabilities of these misfolded polypeptides are from 0.9 to 2.5 kcal mol^{-1} lower than those of the analogous wild-type complexes (Fig. 4), and at low protein concentrations (< 10 μM), the emission intensities of wild-type–biarsenical complexes were from two- to ten-fold higher than those of the misfolded variants.

To assess the ability of bipartite tetracysteine display to detect intermolecular protein assemblies *in vitro*, we performed analogous experiments with GCN4 and Jun variants containing destabilizing point mutations (GCN^{4L20P} and Jun^{L20P}) at the dimerization interface of each coiled coil. CD analysis confirmed that the variants are largely unstructured (Fig. 3c,d), with little or no dimer present at concentrations as high as 40 μM (Supplementary Fig. 1). Equilibrium FlAsH and ReAsH binding experiments revealed that the biarsenical complexes of these GCN4 and Jun variants are at least 1.4 kcal mol^{-1} less stable and from 7- to 20-fold less bright than the analogous wild-type complexes (Fig. 4). These data indicate that misfolded and/or unassembled polypeptides bind to FlAsH and ReAsH with lower affinity and fluoresce less brightly than well-folded variants. Thus, in both inter- and intramolecular contexts, biarsenical affinity and brightness are sensitive to the precise tetracysteine arrangement.

We next sought to determine whether bipartite FlAsH and ReAsH display can detect protein folding and assembly in live mammalian cells (Fig. 5). We genetically fused the sequences encoding aPP and aPP^{F24P} to the C terminus, or GCN4 and $GCN4^{L20P}$ to the N terminus, of enhanced green fluorescent protein (eGFP) and transiently expressed the fusion proteins in HeLa cells alongside cells expressing eGFP alone. eGFP was chosen as a fusion partner so that fusion protein expression could be quantified independently of ReAsH binding. All fusion proteins were expressed in HeLa cells at comparable levels, and their ability to bind ReAsH was verified by gel electrophoresis (Fig. 5a). Live cells were treated with ReAsH (500 nM), washed with BAL (0.75 mM) and imaged by microscopy (Fig. 5b,d and Supplementary Fig. 2a online). eGFP emission was observed in each sample, but cells expressing eGFP-aPP and GCN4-eGFP were also brightly fluorescent at the ReAsH emission maximum. Cells expressing only eGFP, misfolded aPP (aPP^{F24P}) or destabilized GCN4 variants ($GCN4^{L20P}$) showed only weak fluorescence at 608 nm. In contrast to previous FRET studies that used pairs of FPs, the differences in brightness between cells expressing wild-type and misfolded polypeptides are clearly visible by microscopy in real time without the need for multicolor imaging or false coloration methods, even at relatively low expression levels (Supplementary Figs. 2b and 3b online). Unlike ratiometric analysis of FRET, the differences in bipartite tetracysteine protein expression predictably affect the fluorescence intensity of individual cells stained with ReAsH.

To correlate ReAsH fluorescence to cell-to-cell variations in eGFP expression, trypsinized cells were analyzed by flow cytometry using a dual laser system that selectively excites the eGFP and ReAsH chromophores (Fig. 5 and Supplementary Fig. 3). Quantification of eGFP and ReAsH emission from cells expressing unmodified eGFP showed variable green emission and no covariation in red emission (Fig. 5c,e). Cells expressing the fusion proteins eGFP-aPP and eGFP-aPP^{F24P} showed similar levels of red emission at low levels of eGFP expression, but the population of eGFP-aPP–expressing cells emitting red light increased considerably as a function of eGFP emission to 400% over background (Fig. 5c). Even greater differences in fluorescence are observed when cells expressing GCN4-eGFP and $GCN4^{F20P}$-eGFP are compared (Fig. 5e). Thus, bipartite tetracysteine display can be used to detect discrete conformational states and stable protein-protein interactions in live cells—in this case distinguishing between 40-kDa fusion proteins that differ by a single amino acid in the hydrophobic core of aPP or within the dimerization interface of GCN4.

Protein misfolding and aggregation is associated with many human diseases, including sickle cell anemia, cystic fibrosis, Alzheimer's disease and Parkinson's disease. Here we describe a fluorescent labeling strategy for the sensitive detection of discrete protein conformations and protein-protein interactions in live mammalian cells. Notably, the spatial requirement for effective bipartite tetracysteine display (~ 10 Å between the two bicysteine pairs) is substantially more demanding than for FRET between CFP and YFP (~ 50 Å) or even between small molecule pairs (which typically show 100% FRET at all distances less than ~ 20 Å), which suggests that this approach may provide high spatial resolution. Previous studies indicate

that ReAsH staining and imaging are sufficiently sensitive to accurately detect recombinant proteins expressed at 1 μM and above[12]. Bipartite tetracysteine display should be compatible with more advanced fluorescent applications including fluorescence recovery after photobleaching (FRAP), fluorescent-assisted light inactivation (FALI), and electron microscopy, thus enabling high-resolution imaging and/or inactivation of discrete, heteromeric protein-protein complexes[4]. It may also provide a strategy for high-throughput screening of compounds that stabilize discrete protein folds[26] and/or for detecting early protein misfolding events associated with Alzheimer's disease, Parkinson's disease and cystic fibrosis.

METHODS

General methods

Descriptions of materials, polypeptide synthesis and characterization, quantum yield calculations, and vector constructions may be found in Supplementary Methods online.

Determination of apparent equilibrium dissociation constants (K_d)

Equilibrium dissociation constants of biarsenical–protein complexes were determined by monitoring the increase in biarsenical fluorescence intensity as a function of peptide or protein concentration (Fig. 2 and Fig. 4). Additional details can be found in Supplementary Methods.

Preparative-scale synthesis and purification of FlAsH–peptide complexes

A fresh stock solution of FlAsH-EDT_2 was first diluted in fresh TTEE buffer containing 100 mM Tris-HCl (pH 7.8), 3.5 mM TCEP, 0.1 mM EDT and 1 mM EDTA and then added to the desired polypeptide or protein domain (approximately 20 nmol), which was also dissolved in fresh TTEE buffer. Final concentrations were 40 μM (peptide) and 80 μM (FlAsH-EDT_2). Reactions were incubated overnight at room temperature (~25 °C), purified by HPLC and characterized using MALDI-TOF mass spectrometry. Molecular formulas, calculated masses and measured (*m/z*) ratios of FlAsH–protein complexes as determined by MALDI-TOF MS (positive mode) are found in Supplementary Table 1. HPLC traces of purified polypeptides and protein domains and their FlAsH complexes are found in Supplementary Figure 4 online.

Characterization of proteins and FlAsH–protein complexes using CD

CD spectra were acquired in 10 mM sodium phosphate (pH 7.0) buffer using a Jasco J-810-150S spectropolarimeter. Peptide concentrations were 1–80 μM, and three wavelength scans (50 nm min^{-1}) were averaged for each background-corrected (buffer only) plot.

Gel electrophoretic analysis of fusion protein expression and ReAsH labeling

Cultures of approximately 1×10^6 HeLa cells expressing eGFP, eGFP-aPP or eGFP-aPP^{F24P} were stained with 0.5 μM ReAsH-EDT_2 for 30 min in phenol red–free DMEM containing 100 μM EDT. Cells were then washed with 0.75 mM BAL in the same medium for 30 min. Cells were then treated with trypsin, pelleted, redissolved in 1× phosphate-buffered saline (PBS) and lysed by 10 freeze-thaw cycles (freeze to −80 °C and then thaw to 37 °C) in the absence of protease inhibitors. Cell debris was removed upon centrifugation for 5 min at 20,000*g*. The cell extract was concentrated using a Microcon YM-10 centrifugal filter device (Millipore) and normalized to a constant level of GFP emission (508 nm) using an Analyst AD plate reader (LJL Biosystems). After normalization, samples of each extract were subjected to SDS-PAGE (12% precast polyacrylamide gels, BioRad), with samples loaded in 1% SDS sample buffer lacking DTT to minimize GFP denaturation. The gel in Figure 5a was imaged using a Typhoon Imager (GE Healthcare) by independently exciting at 488 nm and monitoring at 526 nm (eGFP), and then exciting at 532 nm and monitoring at 610 nm (ReAsH).

Fluorescence microscopy

Cells were imaged with an Olympus CKX41 microscope equipped with a mercury lamp using a 475 ± 15 nm excitation filter and an LP 520 nm barrier filter to monitor eGFP fluorescence and using a 515 ± 35 nm excitation filter and an LP 590 nm barrier filter to monitor ReAsH fluorescence.

Flow cytometry

Live HeLa cells expressing eGFP, eGFP-aPP, eGFP-aPP^{F24P}, GCN4-eGFP or GCN4^{L20P}-eGFP were treated with 500 nM ReAsH-EDT$_2$ and 100 μM EDT in phenol red–free and serum-free DMEM supplemented with 2 mM L-glutamine for 30 min, washed with PBS and treated with 0.75 mM BAL in the same medium for 30 min. Trypsinized cells were pelleted and analyzed by flow cytometry using a Becton Dickinson FACS Vantage instrument with a 488 nm laser and a 530 ± 15 nm filter to monitor eGFP fluorescence and with a 568 nm laser and a 610 ± 10 nm filter to monitor ReAsH fluorescence. The mean fluorescence intensity (arbitrary units) for ReAsH fluorescence was measured and plotted (Fig. 5c,e) as a function of each log unit of eGFP fluorescence indicated. Error bars shown represent the standard error of measurement.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

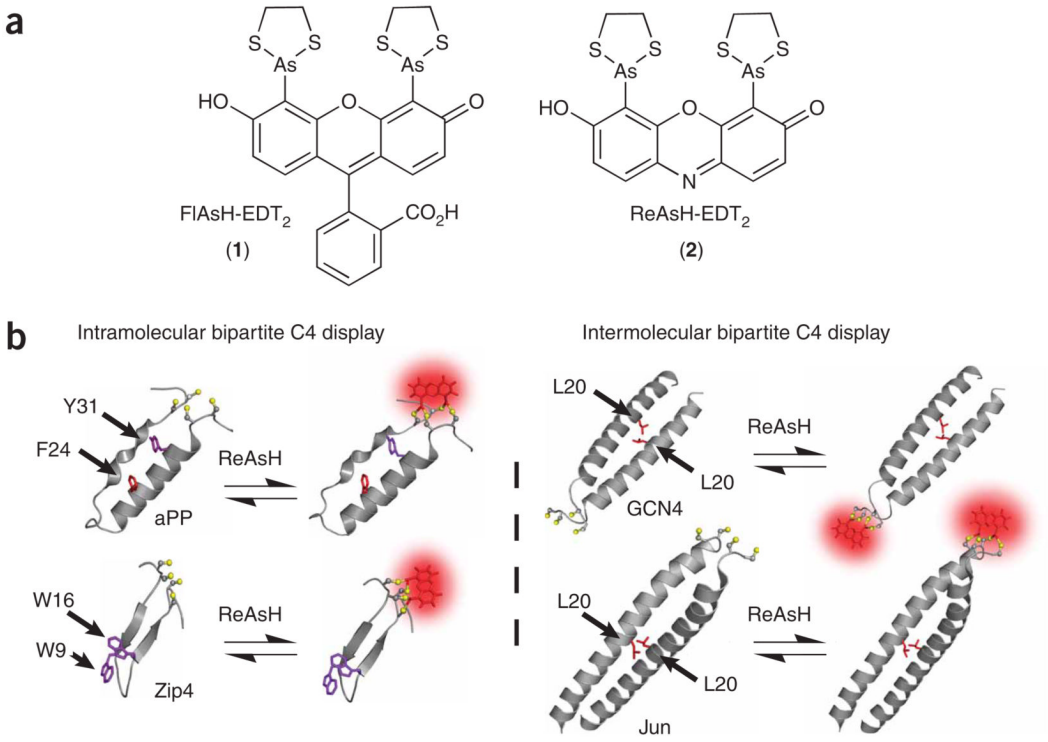
ACKNOWLEDGMENTS

This work was supported in part by the US National Institutes of Health and in part by a grant to Yale University, in support of A.S., from the Howard Hughes Medical Institute.

References

1. Shimomura O, Johnson FH, Saiga Y. Extraction, purification and properties of aequorin, a bioluminescent protein from the luminous hydromedusan, *Aequorea*. J. Cell. Comp. Physiol 1962;59:223–239. [PubMed: 13911999]
2. Prasher DC, Eckenrode VK, Ward WW, Prendergast FG, Cormier MJ. Primary structure of the *Aequorea victoria* green fluorescent protein. Gene 1992;111:229–233. [PubMed: 1347277]
3. Tsien RY. The green fluorescent protein. Annu. Rev. Biochem 1998;67:509–544. [PubMed: 9759496]
4. Giepmans BN, Adams SR, Ellisman MH, Tsien RY. The fluorescent toolbox for assessing protein location and function. Science 2006;312:217–224. [PubMed: 16614209]
5. Chudakov DM, Lukyanov S, Lukyanov KA. Fluorescent proteins as a toolkit for in vivo imaging. Trends Biotechnol 2005;23:605–613. [PubMed: 16269193]
6. Willemse M, Janssen E, de Lange F, Wieringa B, Fransen J. ATP and FRET – a cautionary note. Nat. Biotechnol 2007;25:170–172. [PubMed: 17287746]
7. Lippincott-Schwartz J, Patterson GH. Development and use of fluorescent protein markers in living cells. Science 2003;300:87–91. [PubMed: 12677058]
8. Griffin BA, Adams SR, Tsien RY. Specific covalent labeling of recombinant protein molecules inside live cells. Science 1998;281:269–272. [PubMed: 9657724]
9. Martin BR, Giepmans BNG, Adams SR, Tsien RY. Mammalian cell-based optimization of the biarsenical-binding tetracysteine motif for improved fluorescence and affinity. Nat. Biotechnol 2005;23:1308–1314. [PubMed: 16155565]
10. Gaietta GM, et al. Golgi twins in late mitosis revealed by genetically encoded tags for live cell imaging and correlated electron microscopy. Proc. Natl. Acad. Sci. USA 2006;103:17777–17782. [PubMed: 17101980]

11. Nakanishi J, Takarada T, Yunoki S, Kikuchi Y, Maeda M. FRET-based monitoring of conformational change of the β^2 adrenergic receptor in living cells. Biochem. Biophys. Res. Commun 2006;343:1191–1196. [PubMed: 16580633]
12. Hoffmann C, et al. A FlAsH-based FRET approach to determine G protein - coupled receptor activation in living cells. Nat. Methods 2005;2:171–176. [PubMed: 15782185]
13. Adams SR, et al. New biarsenical ligands and tetracysteine motifs for protein labeling in vitro and in vivo: synthesis and biological applications. J. Am. Chem. Soc 2002;124:6063–6076. [PubMed: 12022841]
14. Andresen M, Schmitz-Salue R, Jakobs S. Short tetracysteine tags to β-tubulin demonstrate the significance of small labels for live cell imaging. Mol. Biol. Cell 2004;15:5616–5622. [PubMed: 15469986]
15. Dyachok O, Isakov Y, Sagetorp J, Tengholm A. Oscillations of cyclic AMP in hormone-stimulated insulin-secreting beta-cells. Nature 2006;439:349–352. [PubMed: 16421574]
16. Enninga J, Mounier J, Sansonetti P, Tran Van Nhieu G. Secretion of type III effectors into host cells in real time. Nat. Methods 2005;2:959–965. [PubMed: 16299482]
17. Ignatova Z, Gierasch LM. Monitoring protein stability and aggregation in vivo by real-time fluorescent labeling. Proc. Natl. Acad. Sci. USA 2004;101:523–528. [PubMed: 14701904]
18. Ignatova Z, Gierasch LM. Aggregation of a slow-folding mutant of a beta-clam protein proceeds through a monomeric nucleus. Biochemistry 2005;44:7266–7274. [PubMed: 15882065]
19. Ignatova Z, Gierasch LM. Inhibition of protein aggregation in vitro and in vivo by a natural osmoprotectant. Proc. Natl. Acad. Sci. USA 2006;103:13357–13361. [PubMed: 16899544]
20. Blundell TL, Pitts JE, Tickle IJ, Wood SP, Wu CW. X-ray analysis (1.4 Å resolution) of avian pancreatic-polypeptide - small globular protein hormone. Proc. Natl. Acad. Sci. USA 1981;78:4175–4179. [PubMed: 16593056]
21. Cochran AG, Skelton NJ, Starovasnik MA. Tryptophan zippers: stable, monomeric β-hairpins. Proc. Natl. Acad. Sci. USA 2001;98:5578–5583. [PubMed: 11331745]
22. O'Shea EK, Klemm JD, Kim PS, Alber T. X-ray structure of the GCN4 leucine zipper, a 2-stranded, parallel coiled coil. Science 1991;254:539–544. [PubMed: 1948029]
23. Junius FK, O'Donoghue SI, Nilges M, Weiss AS, King GF. High resolution NMR solution structure of the leucine zipper domain of the c-Jun homodimer. J. Biol. Chem 1996;271:13663–13667. [PubMed: 8662824]
24. Watt RM, Voss EW Jr. Mechanism of quenching of fluorescein by anti-fluorescein IgG antibodies. Immunochemistry 1977;14:533–551. [PubMed: 303233]
25. Sreerama N, Venyaminov SY, Woody RW. Estimation of protein secondary structure from circular dichroism spectra: inclusion of denatured proteins with native proteins in the analysis. Anal. Biochem 2000;287:243–251. [PubMed: 11112270]
26. Hammarstrom P, Wiseman RL, Powers ET, Kelly JW. Prevention of transthyretin amyloid disease by changing protein misfolding energetics. Science 2003;299:713–716. [PubMed: 12560553]
27. Schuler B, Lipman EA, Steinbach PJ, Kumke M, Eaton WA. Polyproline and the "spectroscopic ruler" revisited with single-molecule fluorescence. Proc. Natl. Acad. Sci. USA 2005;102:2754–2759. [PubMed: 15699337]



c

Name	Sequence	K_{app} (μM) FlAsH	ReAsH
aPP	Ac-FLNCCPSQPTYPGDDAPVEDLIRFYDNLQQYLNVCCMEP-NH_2	5.8 ± 0.7	3.4 ± 1.2
aPP^{F24P}	Ac-FLNCCPSQPTYPGDDAPVEDLIRPYDNLQQYLNVCCMEP-NH_2	27 ± 11	22 ± 10
aPP^{F24,Y31P}	Ac-FLNCCPSQPTYPGDDAPVEDLIRPYDNLQQPLNVCCMEP-NH_2	72 ± 17	65 ± 12
Zip4	Ac-FLNCCEWTWDDATKTWTWTCCMEP-NH_2	0.62 ± 0.12	0.38 ± 0.17
Zip4^{W9,16A}	Ac-FLNCCEWTADDATKTATWTCCMEP-NH_2	38 ± 16	13 ± 3.6
GCN4	Ac-CCGGQLEDKVEELLSKNYHLENEVARLKKLVG-NH_2	25 ± 2.8	28 ± 6.8
GCN4^{L20P}	Ac-CCGGQLEDKVEELLSKNYHPENEVARLKKLVG-NH_2	>111	>111
Jun	Ac-WRIARLEEKVKTLKAQNSELASTANMLREQVAQLKQKVCC-NH_2	1.8 ± 0.7	10 ± 7
Jun^{L20P}	Ac-WRIARLEEKVKTLKAQNSEPASTANMLREQVAQLKQKVCC-NH_2	>111	>111
+ control	Ac-FLNCCPGCCMEP-NH_2	0.48 ± 0.16	0.37 ± 0.20
– control	Ac-FLNCCPPPPPPPPPCCMEP-NH_2	>111	>111

Figure 1.

Intra- and intermolecular bipartite tetracysteine display using FlAsH-EDT_2 and ReAsH-EDT_2. (**a**) Structures of FlAsH-EDT_2 (1) and ReAsH-EDT_2 (2). (**b**) Scheme illustrating intramolecular and intermolecular bipartite tetracysteine display. (**c**) Polypeptide and protein domains used in this work. For intramolecular bipartite tetracysteine display, the polypeptides aPP (ref. 20) and Zip4 (ref. 21) and variants thereof each carried the first five and last five residues (underlined with cysteine residues in orange) of the optimized dodecameric tetracysteine sequence FLNCCPGCCMEP (ref. 9) split between the N and C termini. For intermolecular bipartite tetracysteine display the bZip domains of GCN4 (ref. 22) and Jun (ref. 23) each carried a single CCGG (GCN4 and variants) or Cys-Cys (Jun and variants) at the N or C terminus, respectively. An optimized dodecameric tetracysteine sequence[9] and a variant obtained by replacing Pro-Gly with (Pro)[9], which adopts a rodlike type II helix in solution[27], were used as positive and negative controls, respectively. Point mutations that disrupt folding, dimerization or both of these properties are shown in red. Apparent equilibrium dissociation constants (K_{app}) of biarsenical complexes were measured in TTEE buffer

containing 100 mM Tris-HCl (pH 7.8), 3.5 mM tricarboxyethylphosphine (TCEP), 1 mM EDT and 1 mM EDTA. PDB coordinates used to generate images of aPP, Zip4, GCN4 and Jun can be found using the following PDB ID numbers, respectively: 1PPT, 1LE3, 2ZTA and 1 JUN.

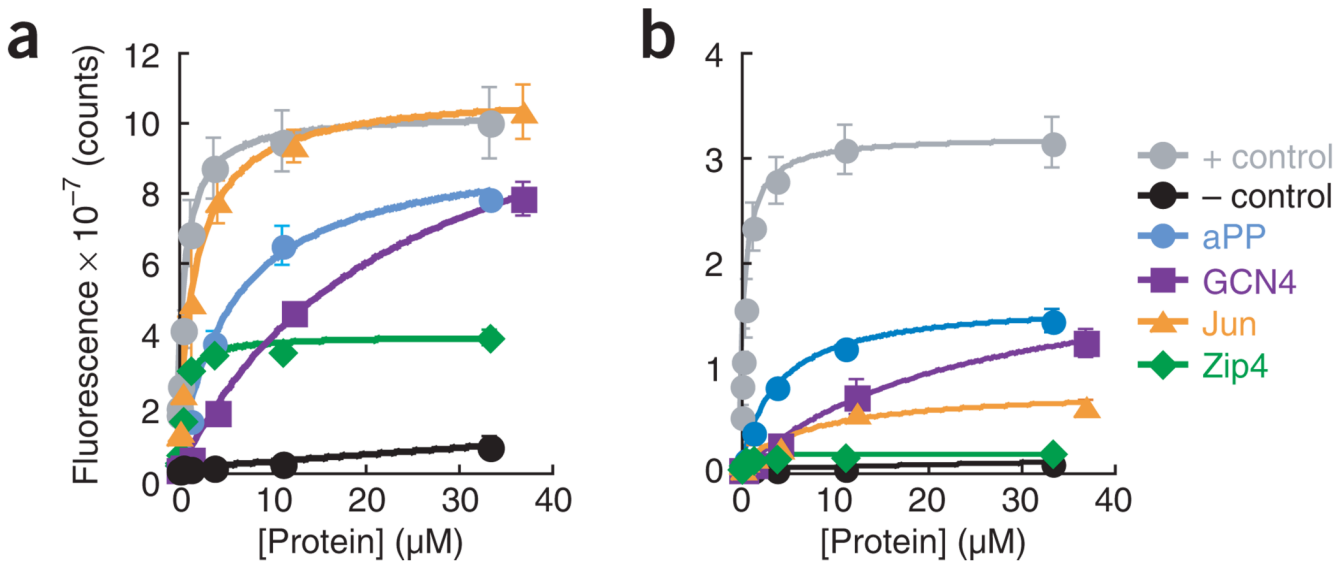


Figure 2.

Equilibrium binding of FlAsH and ReAsH to polypeptides or protein domains containing a bipartite tetracysteine motif. (**a, b**) Each plot shows the emission intensity of FlAsH (**a**) or ReAsH (**b**) solutions on addition of the indicated polypeptide or protein domain. Binding reactions were performed using 25 nM FlAsH or ReAsH in 100 mM Tris-HCl (pH 7.8) containing 3.5 mM TCEP, 1 mM EDT and 1 mM EDTA. Emission intensities were monitored at 530 ± 12.5 nm (FlAsH) or 630 ± 17.5 nm (ReAsH). Each point represents the average of three or more independent titrations ± the s.d. All binding reactions were monitored as a function of time to determine when equilibrium had been reached. These experiments indicated that all binding reactions, except that between FlAsH and aPP^{F24P}, reached equilibrium within 30–90 min.

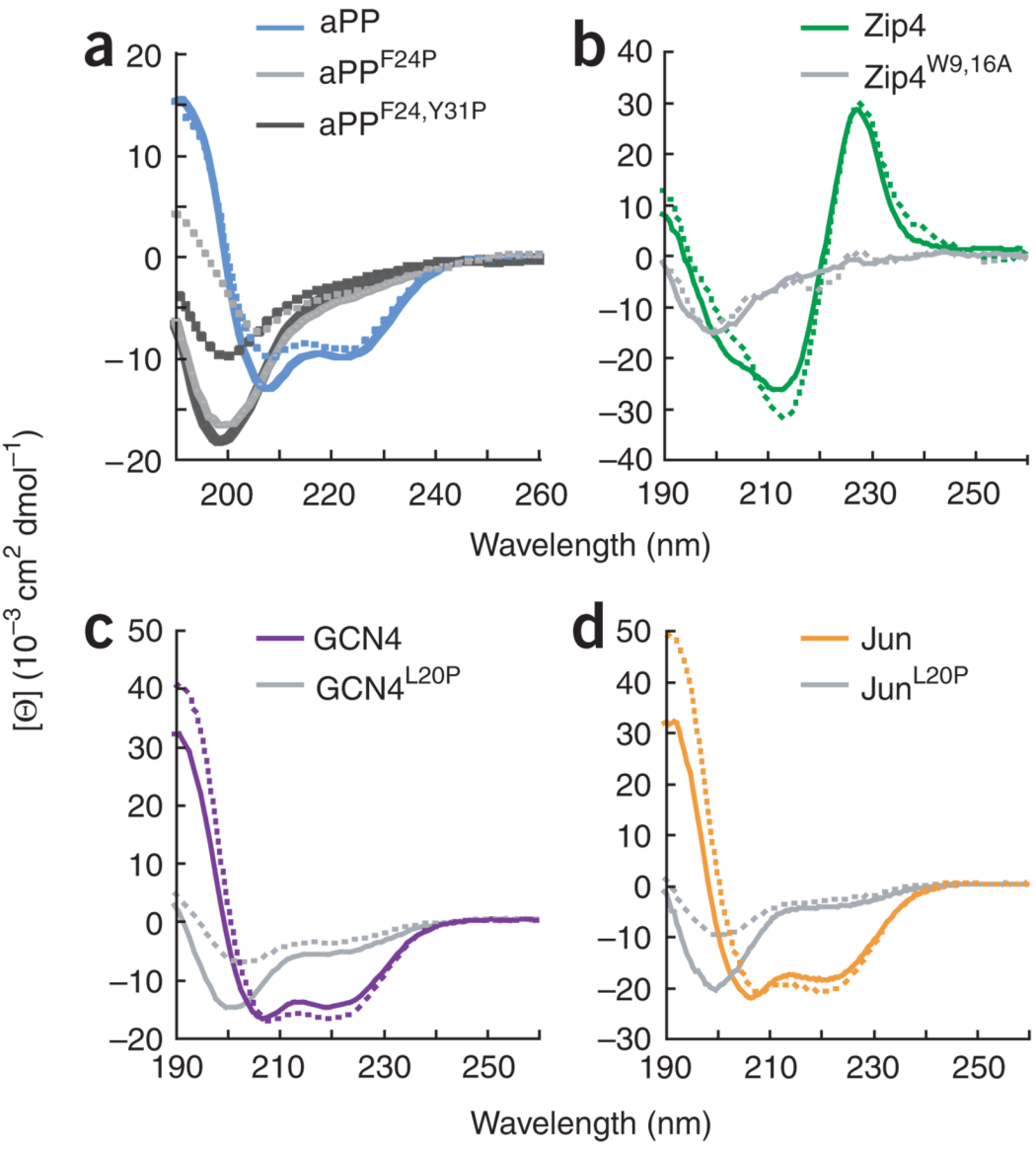


Figure 3.

Effect of FlAsH binding on the secondary structures of wild-type and variant polypeptides and protein domains. (**a**) aPP, aPP^{F24P} and aPP^{F24,Y31P}. (**b**) Zip4 and Zip4^{W9,16A}. (**c**) GCN4 and GCN4^{L20P}. (**d**) Jun and Jun^{L20P}. CD spectra of unmodified polypeptides are shown as solid lines; spectra of FlAsH complexes are shown as dotted lines. CD spectra were acquired at 25 °C in 10 mM phosphate buffer (pH 7.0) at [protein] = 10 μM (aPP, Zip4) or 20 μM (GCN4, Jun) in the absence of competing dithiols. [Θ] represents mean residue molar ellipticity. See Supplementary Figure 1 for concentration-dependent CD spectra of GCN4, Jun, GCN4^{L20P}, Jun^{L20P}, FlAsH–(GCN4)$_2$ and FlAsH–(Jun)$_2$.

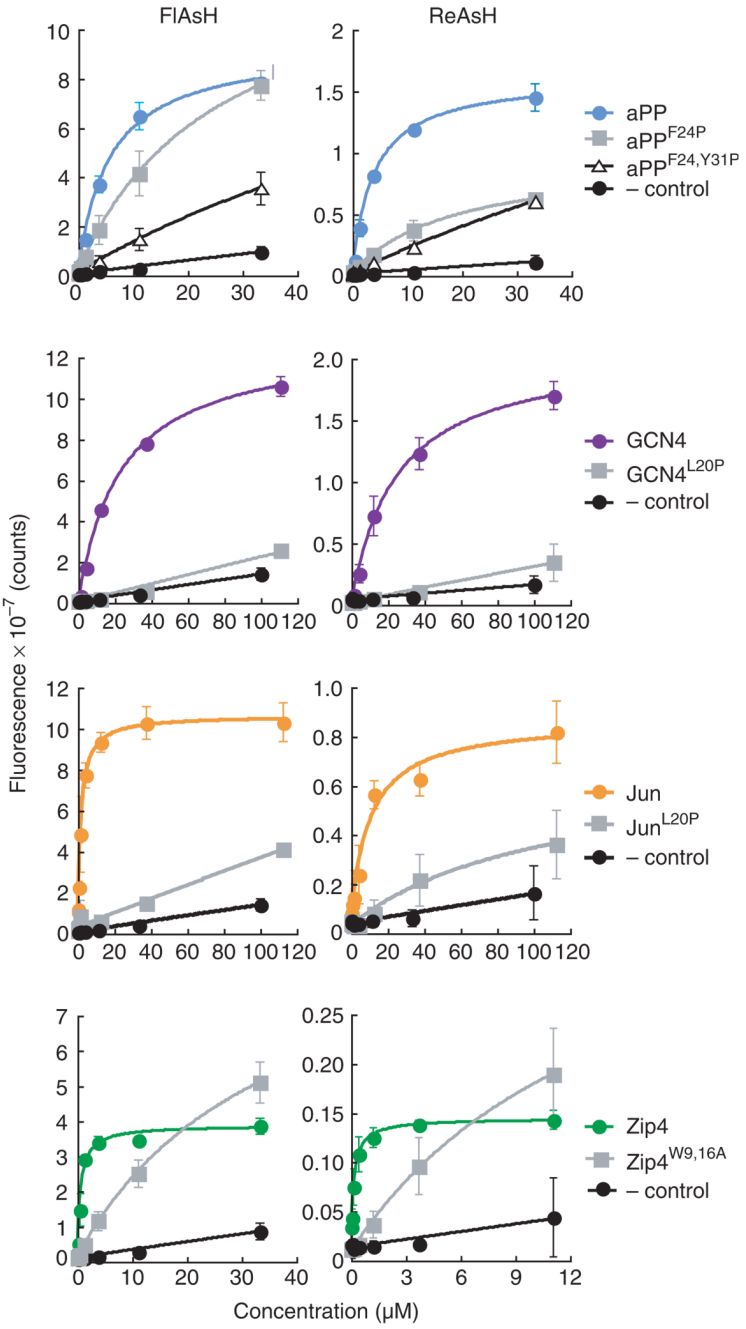


Figure 4.

Misfolded polypeptides and their assemblies bind FlAsH and ReAsH with diminished affinities. Each plot shows the emission intensity of 25 nM FlAsH or ReAsH solutions on addition of the indicated wild-type or variant polypeptide or protein domain. Black symbols illustrate data for the negative control sequence in Figure 1. Binding reactions were performed as described in Figure 2. In the case of the β-hairpin Zip4, the fluorescence intensity of the misfolded variant $Zip4^{W9,16A}$ bound to FlAsH or ReAsH exceeded that of the wild type, presumably because of the two-fold lower tryptophan content of the variant[24]. Error bars show s.d.

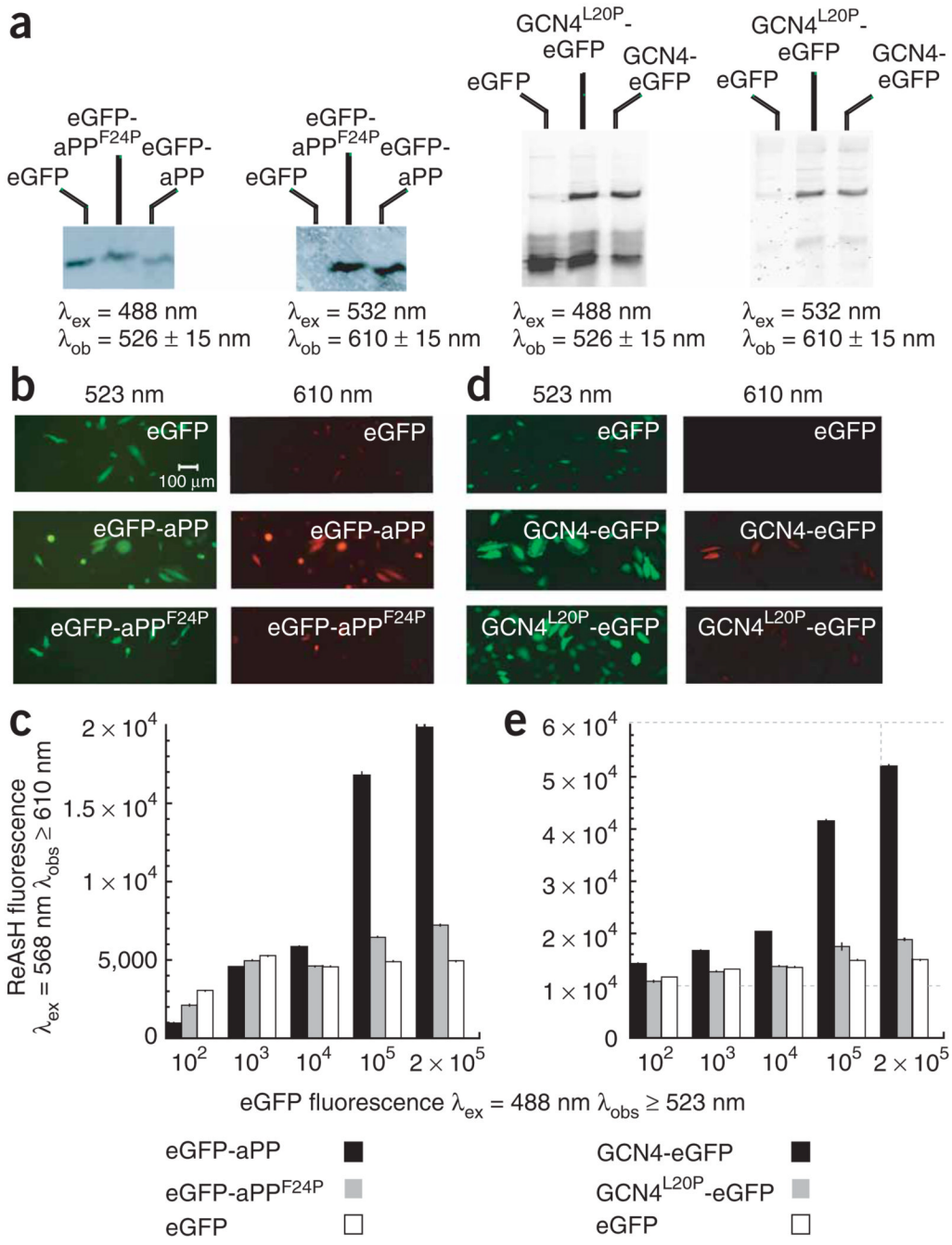


Figure 5.

Differentiation of folded and misfolded proteins and assemblies in living cells with ReAsH-EDT2. (**a**) 12% PAGE of HeLa cell extracts containing eGFP, eGFP-aPP^{F24P} and eGFP-aPP (left) or eGFP, GCN4^{L20P}-eGFP and GCN4-eGFP (right) following treatment with ReAsH-EDT_2 confirms that all fusion proteins are expressed and bind ReAsH. No other fluorescent bands were observed. Samples were loaded in 1% SDS sample buffer lacking DTT to prevent eGFP denaturation, which precludes quantitative analysis of eGFP or ReAsH emission. (**b**) Live HeLa cells expressing eGFP, eGFP-aPP and eGFP-aPP^{F24P}. See Supplementary Figure 2 for close-up images of these cells. (**c**) Trypsinized cells expressing eGFP, eGFP-aPP and eGFP-aPP^{F24P} were pelleted and analyzed by flow cytometry to monitor ReAsH fluorescence.

The mean fluorescence intensities (arbitrary units) for ReAsH fluorescence were measured and plotted as a function of each log unit of eGFP fluorescence indicated. See Supplementary Figure 3 for the flow cytometry data used to generate these averages. Error bars shown indicate the standard error. See Supplementary Table 3 online for oligonucleotide sequences used for vector construction. We estimate that each eGFP fusion protein imaged in these experiments represents between 0.04 and 0.07% of total cellular protein. (**d**) Live HeLa cells expressing eGFP, GCN4-eGFP and GCN4L20P-eGFP. See Supplementary Figure 2 for close-up images of these cells. (**e**) Analysis as in **c** of cells expressing eGFP, GCN4-eGFP and GCN4^{L20P}-eGFP.