Journal: Nature Chemical Biology

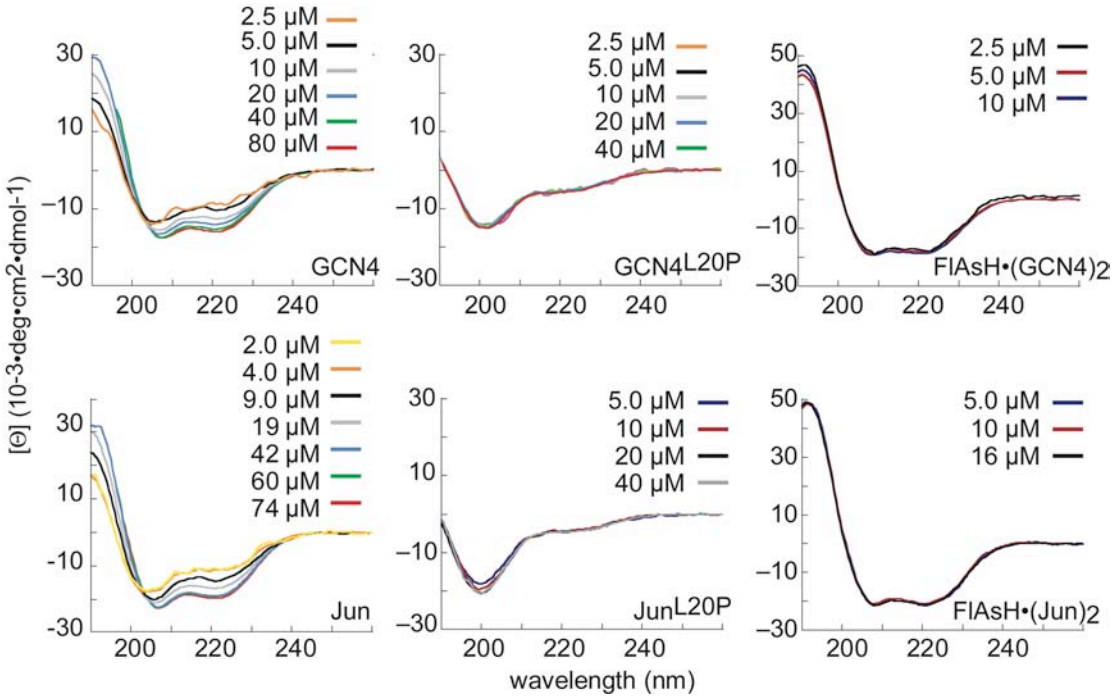
Article Title: Surveying polypeptide and protein domain conformation and association with FlAsH and ReAsH

Corresponding author: Alanna Schepartz (alanna.schepartz@yale.edu)

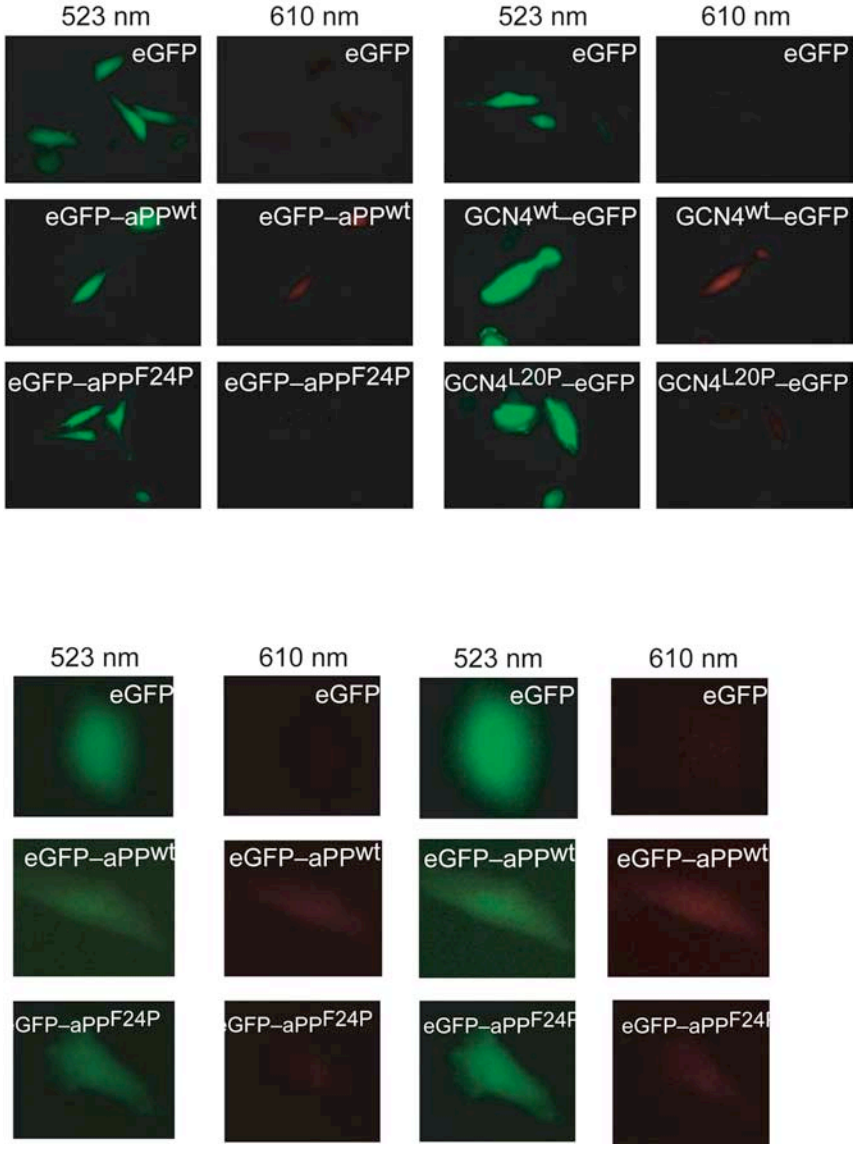
Authors: Nathan W. Luedtke, Rachel J. Dexter, Daniel B. Fried, Alanna Schepartz

Supplementary Information Guide

Item #	Name	Description
1	Supplementary Fig. 1	Concentration-dependent circular dichroism (CD) spectra of GCN4, Jun, $GCN4^{L20P}$, Jun^{L20P}, FlAsH•$(GCN4)_2$ and FlAsH•$(Jun)_2$
2	Supplementary Fig. 2	Differentiation of folded and misfolded proteins and assemblies in living cells with ReAsH-EDT_2.
3	Supplementary Fig. 3	FACS analysis of cells after treatment with ReAsH-EDT_2.
4	Supplementary Fig. 4	HPLC traces of purified polypeptides and protein domains used in this work.
5	Supplementary Table 1	Mass spectral data
6	Supplementary Table 2	Secondary structures of free and FlAsH-bound polypeptides and protein domains estimated from CD spectra in **Fig. 3** according to the CDSSTR method employing the 37-protein basis set SP37A.
7	Supplementary Table 3	Oligonucleotide sequences used in vector construction
8	Supplementary Methods	Descriptions of materials, polypeptide synthesis and characterization, quantum yield calculations, and vector constructions

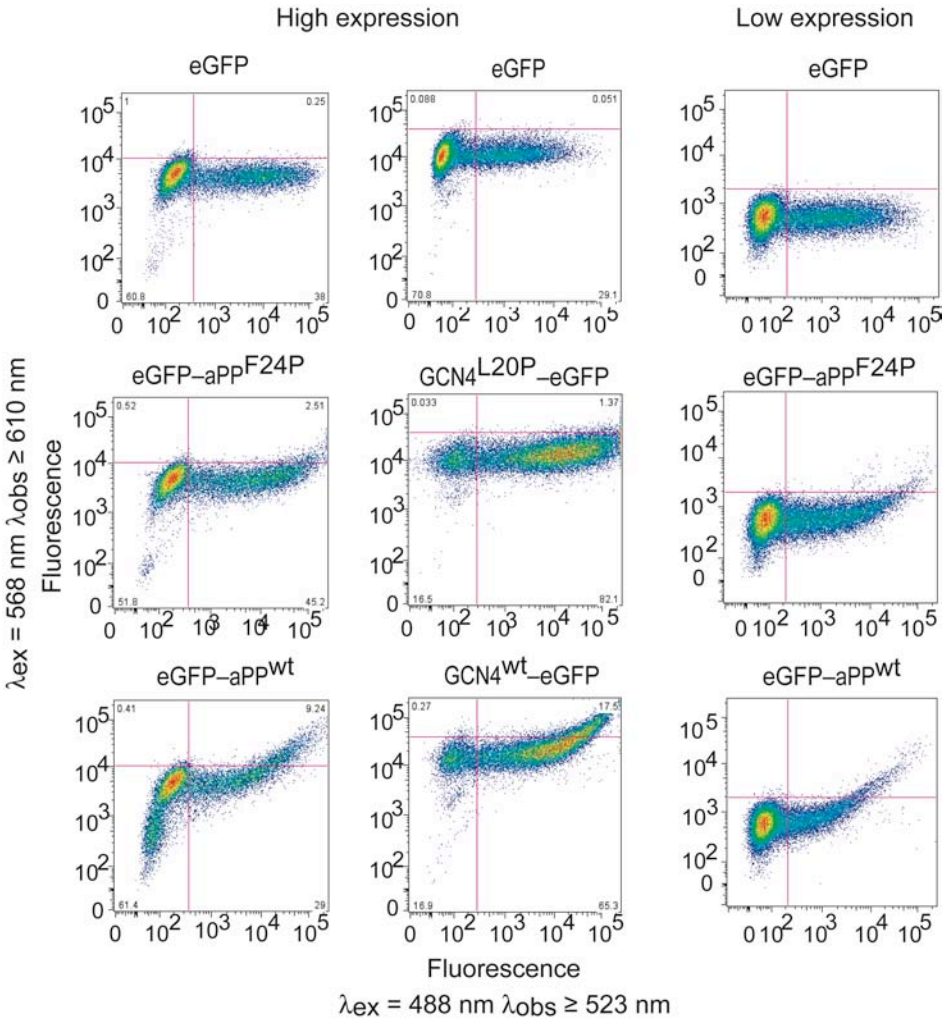


Supplementary Fig. 1. Concentration-dependent circular dichroism (CD) spectra of GCN4, Jun, GCN4^{L20P}, Jun^{L20P}, FlAsH•(GCN4)$_2$ and FlAsH•(Jun)$_2$ at the concentrations shown. CD data were acquired at 25 ^{o}C in 10 mM phosphate buffer (pH 7.0) in the absence of competing dithiols.

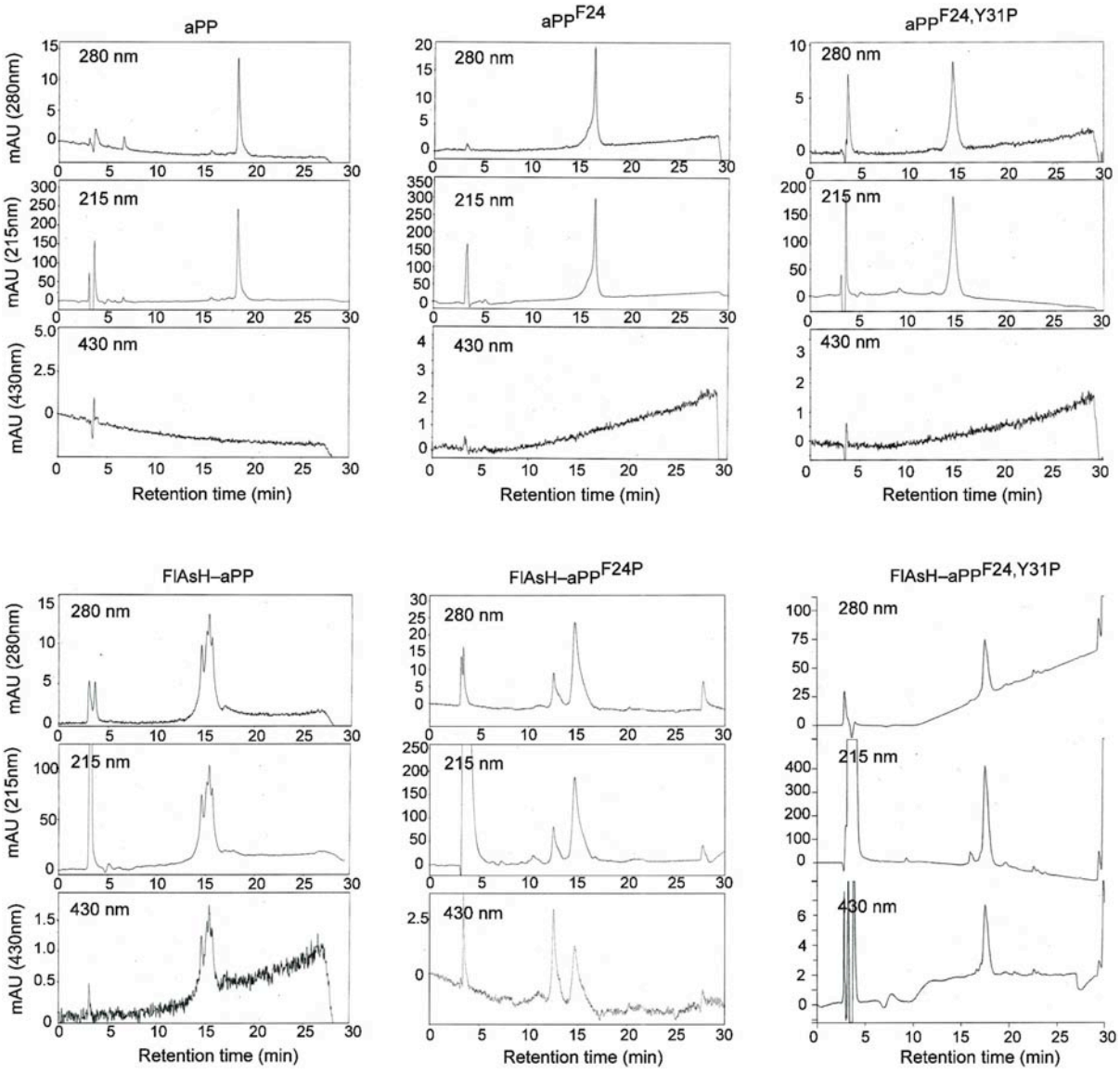


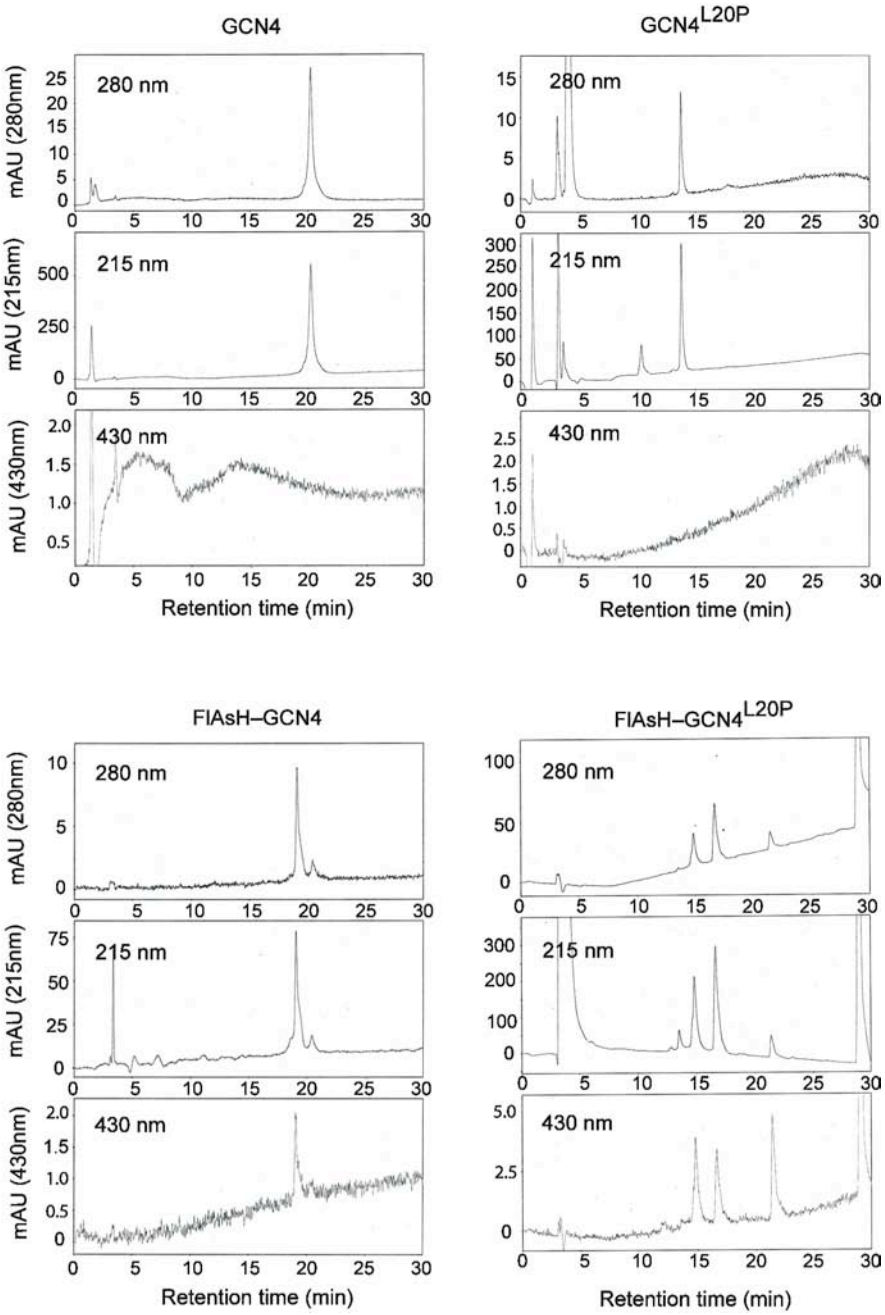
Supplementary Fig. 2. Differentiation of folded and misfolded proteins and assemblies in living cells with ReAsH-EDT_2. (**a**) Close-up images of cells shown in **Fig. 5b** and **d** in the main text. See **Supplementary Fig. 3** online for FACS analysis of each cell population. (**b**) Close-up images of cells expressing folded and misfolded proteins and assemblies at low levels after treatment with ReAsH-EDT_2. Live HeLa cells expressing eGFP, eGFP-aPP^{wt}, eGFP-aPP^{F24P}, $GCN4^{wt}$-eGFP, or $GCN4^{L20P}$-eGFP for 10 h were treated with 500 nM ReAsH-EDT_2 and 100 μM EDT for

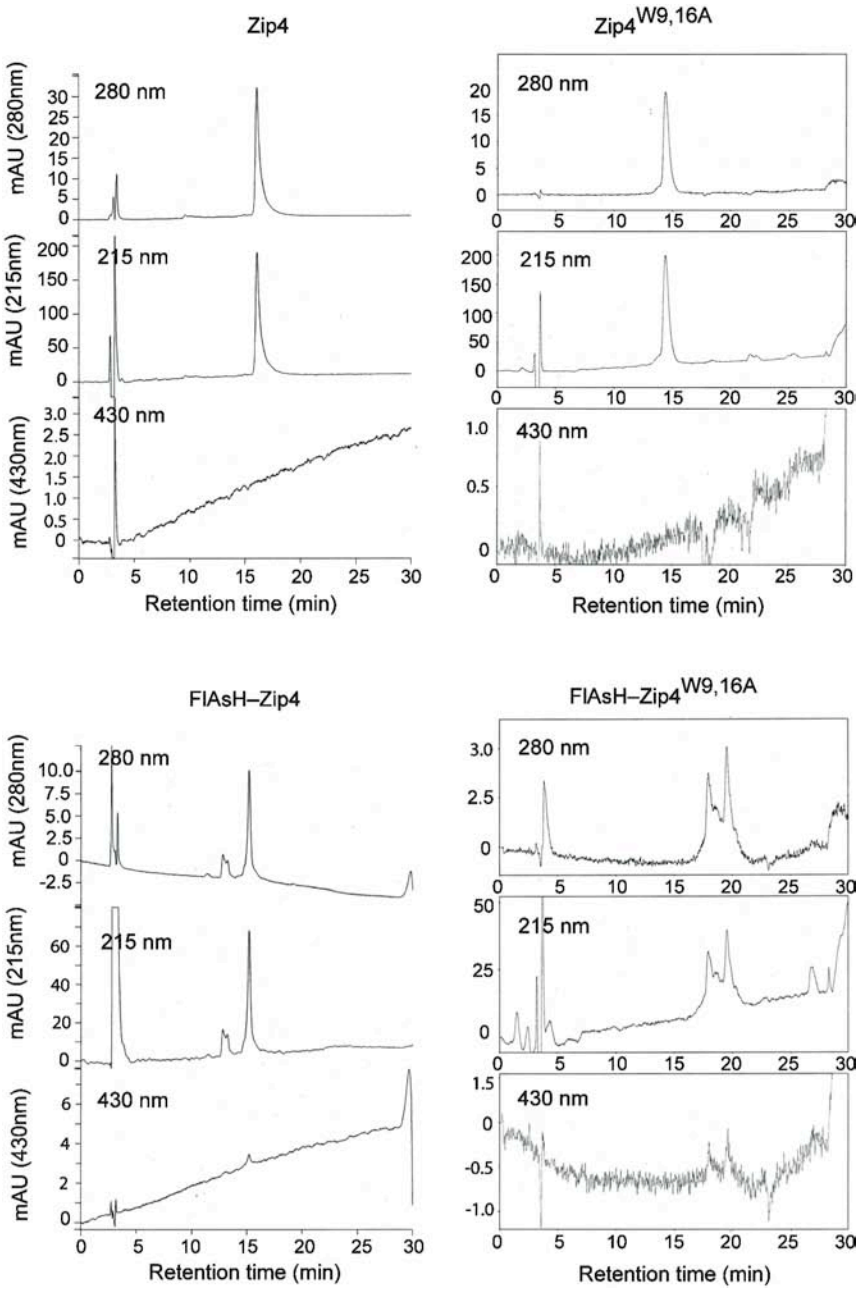
30 min and washed with 0.75 mM BAL prior to imaging. See **Supplementary Fig. 3** online for FACS analysis of each cell population.

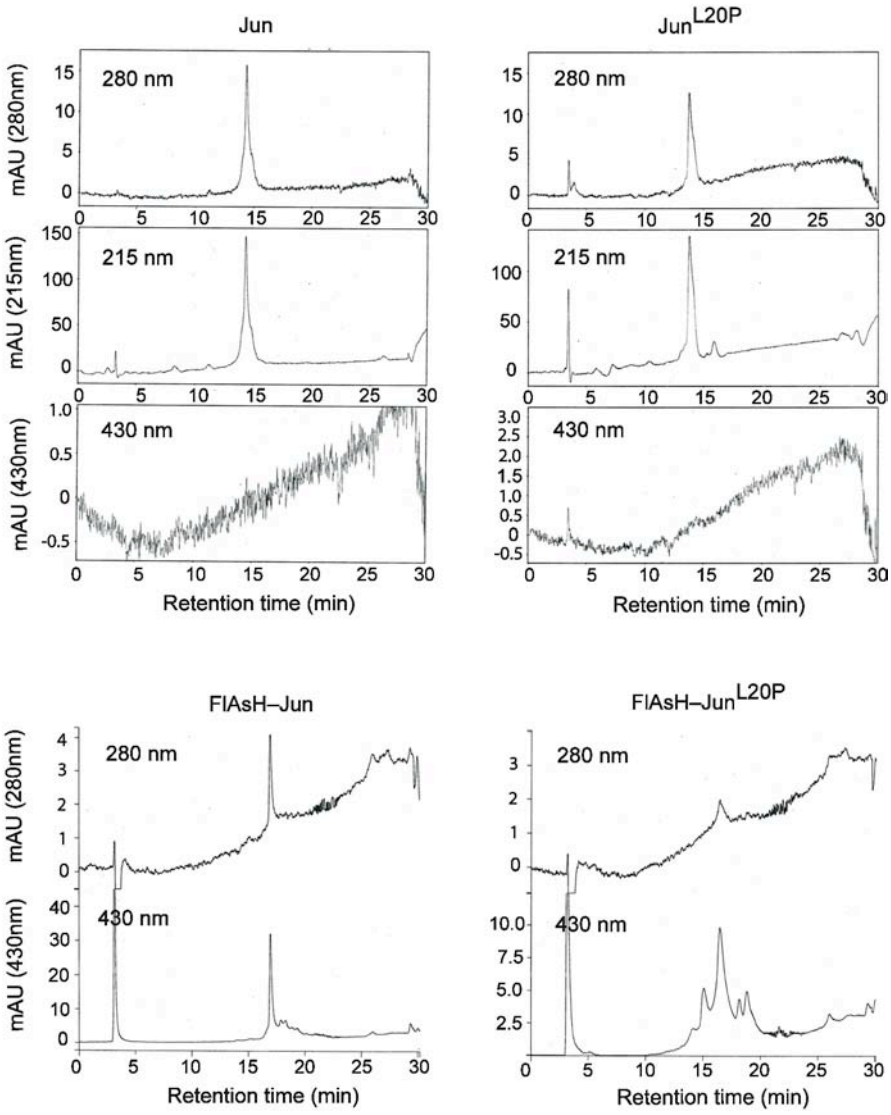


Supplementary Fig. 3. FACS analysis of cells after treatment with ReAsH-EDT_2. FACS analysis of cells expressing the indicated protein at folded and misfolded proteins and assemblies expressed at (**a**) high or (**b**) low levels in living cells with ReAsH-EDT_2. Primary dot plots in arbitrary units on a log scale illustrating the relative eGFP and ReAsH fluorescence of the indicated cell populations.


aPP
aPP^F24
aPP^F24,Y31P
FlAsH–aPP
FlAsH–aPP^F24P
FlAsH–aPP^F24,Y31P
280 nm
215 nm
430 nm
mAU (280nm)
mAU (215nm)
mAU (430nm)
Retention time (min)


GCN4
GCN4L20P
FlAsH–GCN4
FlAsH–GCN4L20P
280 nm
215 nm
430 nm
mAU (280nm)
mAU (215nm)
mAU (430nm)
Retention time (min)


Zip4
Zip4W9,16A
280 nm
215 nm
430 nm
mAU (280nm)
mAU (215nm)
mAU (430nm)
Retention time (min)
FlAsH–Zip4
FlAsH–Zip4W9,16A



Supplementary Fig. 4. Analytical HPLC traces of purified polypeptides and protein domains used in this work. Samples (typically 100 μL) were injected onto a Rainin Dynamax HPLC equipped with a Vydac analytical column (C18, 300 Å, 10 μm, 1 mm x 250 mm) with a flow rate of 1 mL min^{-1}, and a gradient of CH_3CN in H_2O (containing 0.1% TFA) from 1 – 30 min. Peaks were detected at 215, 280 and 430 nm (280 and 430 only for FlAsH complexes). Although all unmodified polypeptides and protein domains migrated as single peaks upon reinjection, their FlAsH

complexes typically migrated as between 1-4 peaks of identical mass that represent alternative thiol-arsenic pairings.

Supplementary Table 1. Mass spectral data obtained for polypeptides and protein domains studied in this work.

Unmodified protein domains	**Formula**	**Calc**	**Meas**
$_{Ac}$-FLNCCPSQPTYPGDDAPVEDLIRFYDNLQQYLNVCCMEP-$_{NH2}$	$C_{200}H_{295}N_{49}O_{63}S_5$	4554	4554
$_{Ac}$-FLNCCPSQPTYPGDDAPVEDLIR**P**YDNLQQYLNVCCMEP-$_{NH2}$	$C_{196}H_{293}N_{49}O_{63}S_5$	4504	4503
$_{Ac}$-FLNCCPSQPTYPGDDAPVEDLIR**P**YDNLQQ**P**LNVCCMEP-$_{NH2}$	$C_{192}H_{291}N_{49}O_{62}S_5$	4438	4437
$_{Ac}$-FLNCCEWTWDDATKTWTWTCCMEP-$_{NH2}$	$C_{134}H_{183}N_{31}O_{39}S_5$	3012	3014
$_{Ac}$-FLNCCEWT**A**DDATKT**A**TWTCCMEP-$_{NH2}$	$C_{118}H_{173}N_{29}O_{39}S_5$	2782	2785
$_{Ac}$-WRIARLEEKVKTLKAQNSELASTANMLREQVAQLKQKVCC-$_{NH2}$	$C_{200}H_{344}N_{62}O_{59}S_3$	4657	4658
$_{Ac}$-WRIARLEEKVKTLKAQNSE**P**ASTANMLREQVAQLKQKVCC-$_{NH2}$	$C_{199}H_{340}N_{62}O_{59}S_3$	4641	4641
$_{Ac}$-CCGGQLEDKVEELLSKNYHLENEVARLKKLVG-$_{NH2}$	$C_{158}H_{263}N_{45}O_{50}S_2$	3657	3657
$_{Ac}$-CCGGQLEDKVEELLSKNYH**P**ENEVARLKKLVG-$_{NH2}$	$C_{157}H_{259}N_{45}O_{50}S_2$	3641	3638
$_{Ac}$-FLNCCPGCCMEP-$_{NH2}$	$C_{55}H_{84}N_{14}O_{16}S_5$	1358	1359
$_{Ac}$-FLNCCPPPPPPPPCCMEP-$_{NH}$	$C_{93}H_{137}N_{21}O_{23}S_5$	2078	2079
FlAsH•protein complexes	**Formula**	**Calc**	**Meas**
FlAsH•$_{Ac}$-FLNCCPSQPTYPGDDAPVEDLIRFYDNLQQYLNVCCMEP-$_{NH2}$	$C_{220}H_{303}N_{49}As_2O_{68}S_5$	5032	5031
FlAsH•$_{Ac}$-FLNCCPSQPTYPGDDAPVEDLIR**P**YDNLQQYLNVCCMEP-$_{NH2}$	$C_{216}H_{301}N_{49}As_2O_{68}S_5$	4982	4976
FlAsH•$_{Ac}$-FLNCCPSQPTYPGDDAPVEDLIR**P**YDNLQQ**P**LNVCCMEP-$_{NH2}$	$C_{212}H_{299}N_{49}As_2O_{67}S_5$	4916	4912
FlAsH•$_{Ac}$-FLNCCEWTWDDATKTWTWTCCMEP-$_{NH2}$	$C_{154}H_{191}N_{31}As_2O_{44}S_5$	3491	3493
FlAsH•$_{Ac}$-FLNCCEWT**A**DDATKT**A**TWTCCMEP-$_{NH2}$	$C_{138}H_{181}N_{29}As_2O_{44}S_5$	3260	3266
FlAsH•[$_{Ac}$-WRIARLEEKVKTLKAQNSELASTANMLREQVAQLKQKVCC-$_{NH2}$**]**$_2$	$C_{420}H_{696}N_{124}As_2O_{123}S_6$	9793	9796
FlAsH•[$_{Ac}$-WRIARLEEKVKTLKAQNSE**P**ASTANMLREQVAQLKQKVCC-$_{NH2}$**]**$_2$	$C_{418}H_{688}N_{124}As_2O_{123}S_6$	9761	9756
FlAsH•[$_{Ac}$CCGGQLEDKVEELLSKNYHLENEVARLKKLVG-$_{NH2}$**]**$_2$	$C_{336}H_{534}N_{90}As_2O_{105}S_4$	7792	7792
FlAsH•[$_{Ac}$-CCGGQLEDKVEELLSKNYH**P**ENEVARLKKLVG-$_{NH2}$**]**$_2$	$C_{334}H_{526}N_{90}As_2O_{105}S_4$	7760	7764
FlAsH•$_{Ac}$-FLNCCPGCCMEP-$_{NH2}$	$C_{75}H_{92}N_{14}As_2O_{21}S_5$	1836	1839
FlAsH•$_{Ac}$-FLNCCPPPPPPPPCCMEP-$_{NH}$	$C_{113}H_{145}N_{21}As_2O_{28}S_5$	2556	2556

Supplementary Table 2. Secondary structures of free and FlAsH-bound polypeptides and protein domains estimated from CD spectra in Figure 3 of the main text according to the CDSSTR / SELCON3 / CONTILL methods, respectively, employing the 37-protein basis set SP37A.

Construct	α-helix	β-sheet	turn	PPII	unstructured
aPP	.38 / .35 / .36	.12 / .08 / .10	.12 / .14 / .13	.07 / .06 / .07	.31 / .36 / .35
FlAsH•aPP	.31 / .27 / .31	.16 / .16 / .14	.12 / .13 / .13	.07 / .06 / .06	.34 / .37 / .37
aPP^{F24P}	.07 / .13 / .10	.19 / .15 / .15	.17 / .13 / .10	.10 / .11 / .06	.47 / .49 / .61
FlAsH•aPP^{F24P}	.07 / .11 / .14	.26 / .36 / .23	.14 / .17 / .14	.10 / .12 / .09	.43 / .33 / .40
$aPP^{F24P,Y31P}$	.06 / .09 / .09	.18 / .15 / .17	.16 / .13 / .11	.11 / .11 / .07	.49 / .51 / .56
FlAsH• $aPP^{F24P,Y31P}$	.04 / .08 / .07	.22 / .15 / .20	.15 / .10 / .12	.11 / .08 / .11	.48 / .47 / .50
GCN4	.50 / .43 / .48	.08 / .08 / .03	.11 / .11 / .11	.07 / .04 / .07	.25 / .34 / .32
FlAsH• $(GCN4)_2$	.60 / .56 / .55	.06 / .05 / .05	.09 / .08 / .09	.05 / .04 / .05	.20 / .26 / .26
$GCN4^{L20P}$	.16 / .18 / .18	.17 / .22 / .15	.14 / .14 / .14	.10 / .11 / .10	.43 / .33 / .43

FlAsH•$(GCN4^{L20P})_2$	.07 / n.c.[b] / .11	.20 / n.c.[b] / .22	.17 / n.c.[b] / .13	.08 / n.c.[b] / .10	.49 / n.c.[b] / .44
Jun	.62 / .48 / .54	.05 / .13 / .00	.08 / .11 / .17	.05 / .04 / .04	.20 / .30 / .30
FlAsH•$(Jun)_2$	.65 / .61 / .60	.04 / .03 / .00	.08 / .08 / .08	.05 / .04 / .05	.19 / .26 / .27
Jun^{L20P}	.08 / .16 / .14	.14 / .17 / .13	.17 / .14 / .15	.13 / .13 / .11	.50 / .40 / .47
FlAsH•$(Jun^{L20P})_2$	.06 / .12 / .10	.18 / .17 / .19	.15 / .11 / .12	.11 / .08 / .11	.50 / .38 / .48
Zip	n.c.[b] / .11 / .00	n.c.[b] / .23 / .59	n.c.[b] / .15 / .25	n.c.[b] / .12 / .19	n.c.[b] / .44 / .05
FlAsH•Zip	n.c.[b] / .03 / .00	n.c.[b] / .27 / .59	n.c.[b] / .17 / .19	n.c.[b] / .10 / .14	n.c.[b] / .49 / .08
$Zip^{W9,16A}$	.02 / .12 / .05	.26 / .20 / .20	.17 / .14 / .16	.10 / .09 / .08	.46 / .45 / .52
FlAsH•$Zip^{W9,16A}$	.05 / .12 / .09	.18 / .19 / .19	.15 / .14 / .15	.12 / .11 / .10	.49 / .43 / .47

[a] Analysis conducted using CDPro software http://lamar.colostate.edu/~sreeram/CDPro/main.html

[b] n.c. = calculation not converged.

Supplementary Table 3. Oligonucletide sequences used in vector construction.

Name	Sequence
Oligo 1	GGCGGCGGAATTCCAAGTAGTGGAAGTTTTTTAAATTGTTGTCCAAGTCAAC CAACATATCCAGGAGATGATGCACCAGTAGAAGATTTAATAAGATTTTATGAT AATTTACAACAATATTTAAATGTATGTTGTATGGAACCATAACTGGATCCGGG CGCG
Oligo 2	GGCGGCGGAATTCCAAGTAGTGGAAGTTTTTTAAATTGTTGTCCAAGTCAACCAACA TATCCAGGAGATGATGCACCAGTAGAAGATTTAATAAGACCATATGATAATTTACAA CAATATTTAAATCTATGTTGTATGGAACCATAACTGGATCCGGGCGCG
Oligo 3	GGC GGC CTC GAG ACC ATG GCG GGA AGT TGC TGC GGA GGA CAA TTA GAA GAT AAG GTA GAA GAA CTC TTA AGT AAG AAC TAT CAC TTA GAA AAC GAA GTA GCA AGA TTA AAG AAG TTA GTA GGA GGA AGT AGT GGA AGT AGT GGA ATT CGG CGC GG
Oligo 4	GGC GGC CTC GAG ACC ATG GCG GGA AGT TGC TGC GGA GGA CAA TTA GAA GAT AAG GTA GAA GAA CTC TTA AGT AAG AAC TAT CAC CCA GAA AAC GAA GTA GCA AGA TTA AAG AAG TTA GTA GGA GGA AGT AGT GGA AGT AGT GGA ATT CGG CGC GG
Primer 1	CGCGCCCGGATCCAGTTATGG
Primer 2	CCGCGCCGAATTCC

Supplementary Methods

Materials.

Fmoc-protected amino acids and Rink amide resin were purchased from Novabiochem. All other reagents required for solid phase peptide synthesis were obtained from American Bioanalytical. 1M Tris-HCl (pH 8) was obtained from American Bioanalytical, ethanedithiol (EDT) (purum grade) was from Fluka, tris-(2-carboxyyethyl)-phosphine HCl (TCEP) was from Pierce, and ethylenediaminetetraacetic acid (EDTA) was obtained as a 0.5 M solution (pH 8.0) from American Bioanalytical. FlAsH-EDT_2 (**1**) and ReAsH-EDT_2 (**2**) were synthesized according to previously reported methods [1, 2] and characterized using ^{1}H NMR and ESI-MS. In both cases the ^{1}H NMR and MS data matched previously reported values. FlAsH-EDT_2 [**1**]: ^{1}H-NMR (400 MHz, d_6-DMSO/D_2O (19:1)) δ 7.97 (d, J = 8 Hz, 1H), 7.81 (t, J = 8 Hz, 1H), 7.72 (t, J = 8 Hz, 1H), 7.26 (d, J = 8 Hz, 1H), 6.57 (d, J = 9 Hz, 2H), 6.53 (d, J = 9 Hz, 2H), 3.37 – 3.32 (m, 8H). ESI MS (m/z): $[M-H]^-$ calcd for $C_{24}H_{17}As_2O_5S_4$: 663, found 663. ReAsH-EDT_2 [**2**]: ^{1}H-NMR (400 MHz, d_6-DMSO/d_3-MeOD (19:1)) δ 6.30 (d, J = 9 Hz, 2H), 6.20 (d, J = 9 Hz, 2H), 3.48 – 3.47 (m, 8H). ESI MS (m/z): $[M-H]^-$ calcd for $C_{16}H_{12}As_2NO_3S_4$: 544, found 544.

Polypeptide Synthesis and Characterization.

All peptides and proteins were synthesized using a Symphony® multi-channel solid-phase synthesis apparatus (Protein Technologies, Inc.) on a 25 µmole scale using standard Fmoc chemistry and Rink amide resin. Following automated synthesis and N-terminal acylation, peptides were cleaved from the resin and deprotected upon 2 hr treatment at RT with 5 mL of 95% TFA containing water (1%), phenol (1%),

ethanedithiol (1%), thioanisol (0.5%), and triisopropylsilane (2%). The peptides were then precipitated by the addition of 40 mL of diethyl ether. The precipitate was washed 1x with 40 mL of diethyl ether, dried under vacuum, and purified using a Rainin Dynamax HPLC and a Vydac semipreparative column (C18, 300 Å, 10 μm, 10 mm x 250 mm) using a variable gradient of CH_3CN in H_2O (containing 0.1% TFA) from 1 – 25 min and a flow rate of 5 mL•min^{-1}. The major product from each synthesis was collected, lyophilized, and characterized for purity by analytical HPLC (>98% in all cases) and analyzed by a Voyager (Applied Biosystems) MALDI-TOF (matrix-assisted laser desorption-ionization time-of-flight) spectrometer (337 nm laser, α-cyano-4-hydroxycinnaminic acid matrix). Following purification, peptides were lyophilized, kept at –20 °C, and reconstituted into the appropriate buffer immediately before use. Molecular formulas, calculated masses, and measured (m/z) ratios of polypeptides and protein domains as determined by MALDI TOF MS (positive mode) are found in **Supplementary Table 1** 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 (**Figs. 2** and **4**). Titrations were performed in freshly prepared, argon-sparged TTEE buffer (100 mM Tris-HCl (pH 7.8) containing 3.5 mM TCEP, 1 mM EDT, and 1 mM EDTA). Fresh solutions of FlAsH-EDT_2 or ReAsH-EDT_2 were prepared by dissolving a small amount (~0.1 mg) in DMSO and quantifying each stock solution (FlAsH ε_{508nm} = 41,000 $cm^{-1}M^{-1}$; ReAsH ε_{578nm} = 63,000 $cm^{-1}M^{-1}$) upon dilution into 100 mM Tris-HCl (pH 8.0) or 100 mM NaOH,

respectively. FlAsH-EDT_2 and ReAsH-EDT_2 stock solutions were then diluted into fresh TTEE buffer to a concentration of 50 nM and used immediately. 1:1 mixtures of buffered FlAsH/ReAsH solutions and each peptide dilution were then added to 384-well, black polystyrene, non-binding Costar® plates (Corning Incorp), sealed with Parafilm-M for 90 min, and the fluorescence emission at λ_{530} (FlAsH) or λ_{630} (ReAsH) detected using an Analyst™ AD plate reader (LJL Biosystems). The changes in fluorescence emission as a function of peptide or protein concentration were plotted using Kaleidagraph v 3.6 (Synergy Software) and apparent K_d values fit to the equation below, which is derived from first principals with no assumptions[3].

$$F_{obs} = F_{min} + ((F_{max} - F_{min})/(2[dye]))([dye] + [protein] + K_d - (([dye] + [protein] + K_d)^2 - 4[dye][protein])^{0.5})$$

In this equation, F_{obs} represents the observed fluorescence at any total peptide or protein concentration [protein], [dye] is the total concentration of FlAsH-EDT_2 or ReAsH-EDT_2, and F_{max} and F_{min} are the maximum and minimum fluorescence values, respectively. K_d is the apparent equilibrium dissociation constant. The reported K_d values represent the average ± standard deviation of the calculated K_d's of three or more independent trials.

Determination of quantum yield (Φ).

Quantum yields were determined by dissolving HPLC-purified FlAsH•protein complexes of known stoichiometry and concentration in 100 mM Tris-Cl buffer (pH 7.8) alongside a fluorescein standard. Samples with optical densities ranging from

0.001 – 0.02 at 508 nm were irradiated using a Photon Technology International Quanta Master 4L-format scanning spectrofluorimeter (Birmingham, NJ) equipped with an LPS-220B 75 W Xenon lamp power supply. The quantum yield of each FlAsH•protein complex ($\Phi_{complex}$) was determined using the relationship $\Phi_{complex} = \Phi_{Flu} \bullet (m_{complex}/m_{Flu})$, where m represents the slope of a plot of integrated emission between 450 and 650 nm as a function of the optical density of the solution, and taking the quantum yield of fluorescein (Φ_{Flu}) = 0.95.

Construction of *peGFP*, *peGFP-aPP^{wt}*, *peGFP-aPP^{F24P}*.

The oligonucleotide sequences used in the construction of these vectors are listed in **Supplementary Table 2** online. *Primer 1* (2 μM) was annealed to *oligo 1* or *oligo 2* (encoding residues 1-39 of aPP^{wt} and aPP^{F24P}, respectively) (2 μM) and extended using Sequenase Version 2.0 DNA polymerase (USB) according to protocols recommended by the supplier. After heat inactivation, DNA was precipitated with ethanol and resuspended in sterile water. The product was digested simultaneously with *EcoR*I and *Bam*HI (NEB) for at least 3 h, precipitated in ethanol, and the digested duplex oligonucleotides resuspended in sterile water. *peGFP-C3* (Clontech) was also digested simultaneously with *EcoR*I and *BamHI;* in this case the products were separated on a 2% agarose gel (2% low melt agarose in 1x TAE), and the desired vector DNA excised and extracted using a Gel Extraction Kit (Qiagen). Digested duplex oligonucleotides were ligated with digested vector DNA using T4 DNA Ligase (Roche Rapid Ligation kit) at a 5:1 molar ratio of duplex oligo:vector (200 ng DNA total per reaction) on ice overnight. In each case the products of ten identical ligation reactions were combined, precipitated with ethanol, electroporated

(Electro Cell Manipulator 600) into XL1 Blue cells (Stratagene), and grown on LB kanamycin plates. Sequencing was performed to identify correct clones, which were then amplified in XL1 Blue cells.

Construction of *peGFP*, *pGCN4^{wt}-eGFP*, *pGCN4^{L20P}-eGFP*.

The *peGFP*, *pGCN4^{wt}-eGFP*, *pGCN4^{L20P}-eGFP* vectors were constructed as described for the analogous aPP vectors, using *oligos 3* and *4* (encoding residues 1-32 of GCN4^{wt} and GCN4^{L20P}), *primer 2*, vector *peGFP-N1* (Clontech) and digestion with *EcoRI* and *XhoI*.