Chapter 8

Detection of Protein Cage Assembly with Bisarsenic Fluorescent Probes

Thomas A. Cornell and Brendan P. Orner

Abstract

We describe a method for the detection of specific protein-protein interactions in protein cages through the exploitation of designed binding sites for bisarsenic fluorescent probes. These sites are engineered to be protein-protein interface specific. We have adapted this method to ferritins; however, it could conceivably be applied to other protein cages. It is thought that this technique could be utilized in the thermodynamic and kinetic characterization of cage assembly mechanisms and in the high-throughput screening of protein cage libraries for the discovery of proteins with new assembly properties or of optimized conditions for assembly.

Key words FlAsH-EDT_2, Protein-protein interactions, Fluorescence detection, Oligomerization state, Ferritin

1 Introduction

Nanocage proteins play key roles across many biological systems. They perform wide-ranging functions, such as metabolite synthesis [1], protein-folding assistance [2], viral genome protection and delivery [3], and ion and metabolite storage and sequestration [4], which are dependent on, and a consequence of, their large, complex, and often self-assembly controlled, structures [5]. Many of the techniques employed to study the formation of these unique architectures are often indirect and provide only low structural and temporal resolution.

Because of their unique hollow nanostructure, protein cages are attractive for a wide range of nonnatural applications. For example, they have been employed as vehicles capable of controlled drug delivery and as size-constrained synthetic reaction vessels for mineralized materials [6, 7]. To develop such systems with greater utility, proteins with properties tailored to each specific application are required. However, the rational design of protein cages with bespoke properties faces the dual challenges of engineering both protein-folding and

Brendan P. Orner (ed.), *Protein Cages: Methods and Protocols*, Methods in Molecular Biology, vol. 1252,
DOI 10.1007/978-1-4939-2131-7_8,

protein-protein interactions. While advancing [6, 8, 9], these both are far from solved problems. Thus, it is envisioned that high-throughput techniques to discover protein cages with novel properties would not only be helpful but are necessary.

The complex and often highly symmetric structures of many protein cages are stitched together by protein-protein interactions between their protein building blocks. Understanding the fundamentals controlling the assembly could help scientists to utilize better and to engineer protein cages in the future. Research to dissect the role of the individual protein-protein interactions and their components can be slow and tedious, utilizing often limited and indirect techniques, such as alanine scanning via site-directed mutagenesis, followed by size exclusion chromatography (SEC) and transmission electron microscopy (TEM) for characterization of each of the mutants individually [10–15]. While these strategies have provided much insight, a more rapid method could enhance the field by acquiring this information more swiftly. Furthermore, a technique that more directly assesses specific protein-protein interactions could provide greater utility in this field.

A growing number of methods, such as Förster resonance energy transfer (FRET), split green fluorescent proteins (GFP), and split luciferase enzymes [16–19], have been employed to study the assembly of two protein-binding partners. Such techniques often require very large protein fusions [20], antibodies [21], or posttranslational modifications [22, 23] and are often limited to two binding partners.

Recently bisarsenic fluorescent reagents, such as FlAsH and ReAsH, have been developed as protein probes. These reagents become fluorescent when bound to a protein with four cysteine residues displayed in a specific geometric relationship [24]. The reagents provide a smaller labeling tag than GFP fusions and thus allow the probing of proteins in a more native state within cells [25]. Moreover, by splitting the tetracysteine peptide tag into two cysteine pairs that straddle a protein-protein interface, it is possible to design binding sites for these reagents that are dependent upon the formation of specific protein-protein interactions [26, 27]. While this method was originally shown to work for short peptide chains, more recently it has been used to establish structurally detailed insight into the conformationally driven transduction of binding information in the trans-membrane helix of EGFR [28].

We have recently expanded this technique to interrogate the self-assembly of a large multimeric ferritin protein cage (*see* Fig. 1) [29]. In our study, the ferritin protein, DNA-binding protein from starved cells (Dps), was engineered with a C-terminal bipartite cysteine pair. The C-termini of Dps monomers converge when this protein assembles into a tetrahedrally symmetric cage but diverge in the presumed twofold symmetric dimer intermediate of the assembly process. The convergent cysteines create a binding site for the bisarsenic fluorescent reagent, resulting in a cage-dependent signal.

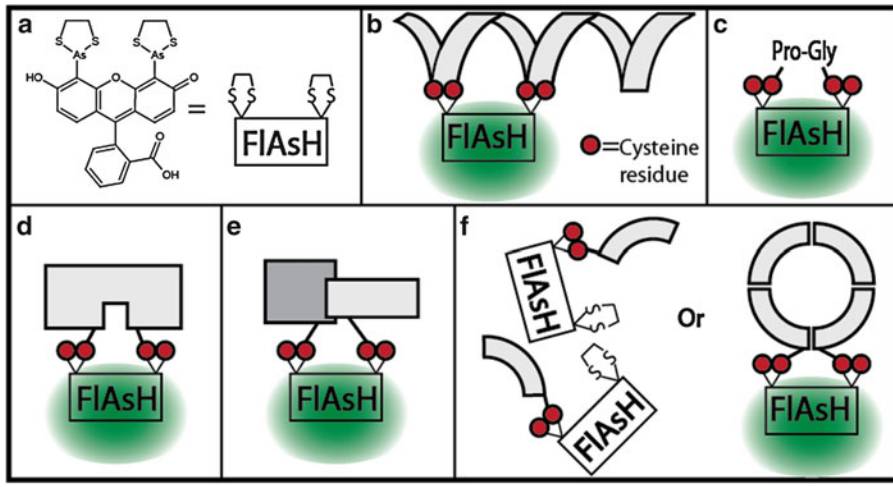

Fig. 1 Conceptual evolution of FlAsH binding site design. (**a**) Structure and schematic of FlAsH-EDT_2. (**b**) Initial presentation of four cysteines on one face of an α-helix. (**c**) Optimized hairpin peptide. (**d**) Bipartite cysteine display with the two cysteine pairs placed proximately on each of the termini of the same protein. (**e**) Bipartite cysteine display with cysteine pairs straddling a protein-protein interface between interacting proteins. (**f**) Strategy to detect self-assembly of nanocage structures

We have demonstrated that this technique works in bacterial lysates thus removing the need to purify and analyze each protein mutant individually, greatly increasing throughput. This flexible strategy could be used to establish the assembly dynamics of protein cages in vitro or in cells. In a high-throughput format, it also has potential for the discovery of mutant protein cages with enhanced properties or for screening conditions for cage stability and assembly. The following chapter describes a method for the application of this technique.

2 Materials

2.1 Protein Design and Preparation

1. PDB coordinates for the protein cage.
2. Inducible expression plasmid containing gene encoding the protein cage and carrying antibiotic resistance.
3. Primers for site-directed mutagenesis or extension PCR.
4. Expression host cell line such as Rosetta (BL21) *E. coli* (Novagen).
5. Luria broth (LB): 10 g tryptone plus (Sigma), 5 g yeast extract (Sigma), 10 g NaCl (Sigma), 1 L water. Autoclave.
6. Antibiotics for inoculation (such as carbenicillin, 50 mg/mL) and reagent to activate expression (such as IPTG, 0.5 mM final concentration).
7. FlAsH buffer: 100 mM Tris–HCl (Sigma), 100 mM NaCl (Sigma), 1 mM EDTA (Sigma), pH 7.8. Degas.

2.2 FlAsH Analysis

1. FlAsH-EDT_2 (Invitrogen): 1 μM in FlAsH buffer.
2. Additives for signal and specificity optimization: 1,2-ethanedithiol (EDT), 1 mM (Sigma); 2-mercaptoethanol (2-ME), 1 mM (Sigma); Tris(2-carboxyethyl)phosphine (TCEP), 3.5 mM (Sigma). Make all stocks in FlAsH buffer immediately before use.
3. Guanidine HCl (Sigma): 8 M in FlAsH buffer.
4. Steady state fluorimeter or fluorescent plate reader.

3 Methods

The following procedure should be used as a guide to help utilize the fluorescent reagent FlAsH-EDT_2 to detect specific protein-protein interactions in protein cages. Because these interactions arise only when the cages assemble into higher-order oligomerization states, this strategy is based on designing FlAsH binding sites at the protein-protein interfaces that appear when the cage forms. The binding sites are generated when bipartite cysteine pairs engineered into the monomers become proximal upon cage assembly. Once the binding sites emerge, in vitro or in vivo detection of oligomerization is possible by observing fluorescence.

3.1 Designing Binding Sites Dependent Upon Cage Assembly

1. Design FlAsH binding sites guided by the inspection of a protein data bank (PDB) structure. Position two pairs of cysteines across an interface that ideally only forms in the highest-order cage structure and does not appear in any assembly intermediate (unless the identification of such an intermediate is the goal of the study) (*see* **Notes 1–5**, and *see* Fig. 2).

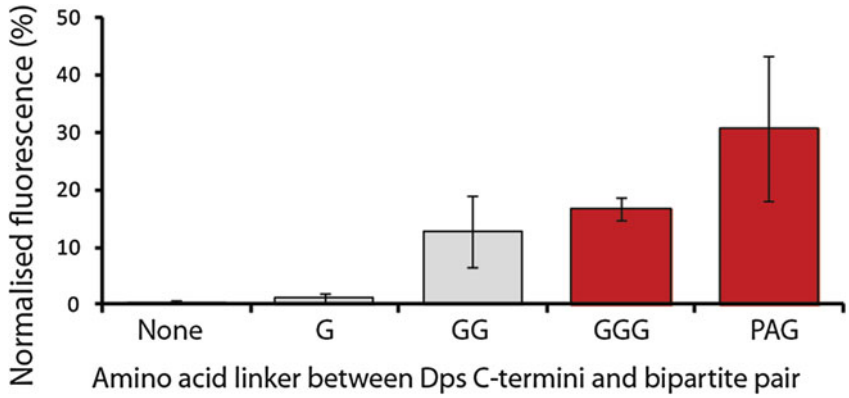

Fig. 2 Effect of the length and conformation of the sequence linking the nanocage protein Dps C-terminus to a bipartite pair (fluorescence is normalized to a positive control, G = glycine, A = alanine, and P = proline) (1 μM FlAsH-EDT_2, 1 mM EDT, 1 mM 2-ME, 3.5 mM TCEP, 0.1 mg/ml final protein concentration). Results shown are from 6 replicates of lysate samples with overexpressed proteins of interest. Error bars are SD

2. Design a positive control protein by fusing a peptide sequence containing CCPGCC (there are several optimized versions) [30] onto a monomer. This is most easily done as an extension to one of the termini (*see* **Note 6**).

3.2 Protein Preparation

1. Use standard extension or site-directed mutagenesis cloning techniques to introduce mutations required to generate the designed proteins (*see* **Notes 7** and **8**).
2. Express the designed proteins by first transferring constructed plasmids into an expression host (e.g., Rosetta, *E. coli* expression cells) and induce protein expression (such as by growing in inoculated LB until an $O.D_{600}$ of 0.6 followed by the addition of IPTG) (*see* **Note 9**).
3. Isolate cell pellet by centrifugation, disrupt cells (e.g., by sonication), and clarify by centrifugation to obtain the soluble proteins (*see* **Note 10**).
4. If choosing to assay the proteins in purified form, as opposed to in lysates, isolate and characterize the proteins with size exclusion chromatography (SEC) (*see* **Notes 11–14**) to confirm assembly.

3.3 Sample Preparation and Fluorescence Analysis

Initial screening in lysate samples can be used to rapidly guide the design process. Samples can either be screened in a high-throughput plate format, which speeds up data collection and mitigates the running of replicate experiments, or in cuvettes with a fluorimeter (*see* **Note 15**). Experiments should be designed to include a positive control protein, which not only can help establish that the dye is working as expected, but also can provide a detection upper limit to benchmark optimized dye and protein concentrations and the sensitivity of the fluorimeter. A negative control, typically a wild-type protein with no additional cysteines, should also be included. This control determines the level of the background fluorescence due to nonselective binding to the protein. A no-protein control should also be used to establish further sources of background from FlAsH itself.

1. Dilute the lysate sample to the required total protein concentration (0.1 mg/ml, BCA, Merck) with FlAsH buffer and add EDT, 2-ME, and TCEP solutions and incubate for 2 h at room temperature (*see* **Notes 16–18**, and Fig. 3).
2. Add FlAsH-EDT_2 solution and incubate for an additional 2 h at room temperature away from light (*see* **Notes 19** and **20**).
3. Read each sample in a fluorimeter or plate reader. Multiple replicates are suggested (*see* **Note 21**).

3.4 Denatured Protein Fluorescence Analysis

A useful control for multimeric proteins is to assay the designs but in the presence of a denaturant such as 6 M guanidine HCl. At high enough concentrations of the denaturant, the proteins should

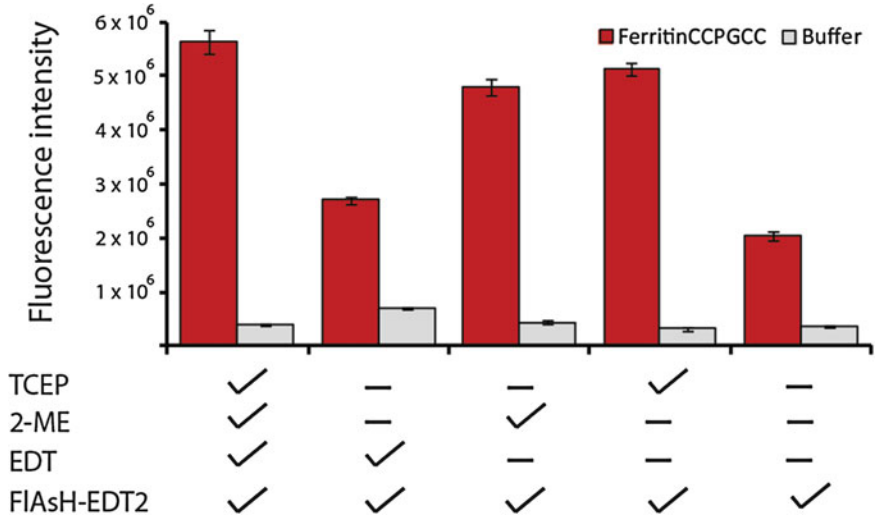

Fig. 3 Relationship between additives and the observed fluorescence intensity with an *E. coli* bacterioferritin positive control fused to CCPGCC (1 μM FlAsH-EDT_2, 1 mM EDT, 1 mM 2-ME, 3.5 mM TCEP, 0.1 mg/ml final protein concentration). Results shown are an average of 6 repeats using purified proteins. Error bars are SD

unfold and all quaternary structure will be lost. This will destroy the designed FlAsH binding sites, resulting in the loss of any fluorescent signal observed in assembly promoting conditions. This should confirm that the signal is indeed oligomerization independent especially if the positive control designs containing CCPGCC are not affected by the denaturant.

1. For protein denaturation experiments, dilute the sample to the required protein concentration with 8 M guanidine HCl in FlAsH buffer, until a final concentration of 6 M guanidine HCl is achieved. Incubate for 2 h at room temperature.
2. Add EDT, 2-ME, and TCEP solutions and incubate for a further 2 h at room temperature (*see* above).
3. Add FlAsH-EDT_2 and incubate for 2 more hours, away from light at room temperature. Run each sample in a fluorimeter and compare to non-denatured samples (*see* **Note 22**).

4 Notes

1. For large multimeric nanostructures, look for interfaces that form only when the highest oligomerization state is achieved. By directing the design toward highly symmetric interfaces on protein nanostructures, it is possible to increase the number of binding sites per structure allowing for detection at lower protein concentrations.

2. The twofold symmetry dimer is presumed to be an assembly intermediate in most mini- and maxi-ferritin cages. The protein-protein interface of this dimer is avoided for binding site engineering as it would exist regardless of the formation of higher-order cage structures.
3. Termini that converge at interfaces make for "quick and easy" locations for the addition of cysteine pairs; however, by using this strategy, the potential for an optimal geometry may be unachievable and may require flexible termini. Although some termini may be close enough in space to allow the cysteine pairs to form an appropriate FlAsH binding site, others may require screening a range of linkers in order to optimize FlAsH binding (*see* Fig. 2).
4. Binding site designs that employ a more structurally complex strategy than simply attaching cysteines to termini may initially be thought to be preferred. However, care must be taken when swapping one amino acid for a cysteine as this could lead to (a) protein misfolding due to undesired disulfide bond formation or (b) the removal of key hot-spot residues essential for stabilizing protein-protein interactions. It is reasonable to pursue both strategies in parallel.
5. Take note of all native cysteines in the structure. Any that are too close to the FlAsH binding site might need to be mutated out; on the other hand it could be imagined that some native cysteines could be included in the engineered binding site.
6. This can be a more useful control than simply using the synthesized small peptide, because this protein can also be employed as a control for oligomerization/denaturation studies to determine how assembly dependent the designed interfacial binding site is.
7. If the protein cage monomers can be expressed in a soluble form and assemble without extraordinary measures, then it is possible that the fluorescence assay can be implemented directly in the clarified lysate avoiding arduous purification and thereby increasing the throughput of designs or conditions that can be screened.
8. Expressing the protein cage with purification tags, while easing purification, could result in monomers that are assembly inhibited on account of steric blocking by the affinity label. Therefore, it is recommended to use purification strategies that involve either a very small tag (like His_6) or proteolytic removal of the tag. The small tag approach, although requiring controlled characterization to ensure native assembly, may prove most flexible in that the same protein construct could be used for screening in homogeneously purified proteins, in lysates, and in living cells. This should be taken into account while designing the cloning strategy.

9. The ideal conditions for the expression of proteins vary depending on the nature of the protein. Care must be taken with protein cages that have been observed to exist as multiple, metastable oligomerization states. Extensive characterization of the protein after expression should be performed to ensure that the desired oligomer is achieved especially when establishing benchmarks in new systems.
10. Resuspending the cell pellet in FlAsH buffer can save some steps and loss of protein especially if the screening will be undertaken with lysates. The addition of EDTA to the buffer can be helpful not only for maintaining ion concentration but can also stop some protease activity in the lysate samples (*see* **Note 12**).
11. SEC provides information about what oligomerization states exist in each sample and their ratios with respect to each other. Thus, this technique can be used to ensure that the expression method has produced the expected oligomerization states of the protein cage. Extended purification methodologies, especially ones including digestion of purification tags, can alter previously observed oligomerization state ratios, and re-characterization is essential after purification. If the initial screens were performed in lysates, further characterization should be performed with purified proteins for the top hits. Moreover, SEC should prove to be a powerful characterization tool in systems that form multiple oligomerization states of which the FlAsH binding sites were designed to probe one specifically. The FlAsH technique can be used in concert with SEC to determine which states are FlAsH active [29].
12. Purifying the protein into FlAsH buffer speeds up assaying as no buffer exchange is needed. Buffer exchange can cause protein aggregation or increase protease activity in a highly concentrated sample. Changing the concentration of the sample can also have an effect on the oligomerization state observed. Creating the simplest route from expression to screening is key to having the most reproducible results.
13. Storing lysate samples at 4 °C is recommended; however, temperature can possibly affect the oligomerization state ratio of the nanocage. Consistency between how all the samples are treated is important to maintain reproducibility.
14. Protein aggregation resulting in false positives can be the biggest problem at this step. The likelihood of this happening is increased if the design involves mutating residues in the more highly structured regions of the protein such as on a β-sheet.
15. Keep note of the sensitivity of the machine as different fluorimeters will provide different absolute readings; comparisons can be made by normalization to controls. If filters are being

used, ensure that they are the closest to the required wavelength and that their bandwidths do not overlap.

16. Protein concentration analysis for lysate samples is a measure of all the proteins in the sample. While it can be difficult to estimate how much of the sample is the protein of interest, steps such as PAGE gels and ensuring that the samples are treated identically from the protein production stage should lead to similar concentrations of the protein of interest. If multiple designs are to be compared, then additional control experiments should be conducted to ensure that each protein is expressed at a similar level.
17. In less stable nanocage systems, the oligomerization state may be altered if the sample is diluted from a stock concentration for the experiment. This also requires a control experiment.
18. The additives used in FlAsH fluorescence experiments can vary. TCEP is a commonly employed additive at concentrations ranging from 1 to 10 mM. It helps to maintain a reduced redox state. EDT is used to ensure that any FlAsH unbound to protein produces a low background. Commonly used concentrations range from 1 to 5 mM; however, high concentrations have been shown to increase stringency and even inhibit FlAsH binding completely [24]. 2-ME is used less often, but it is suggested to not only help maintain the redox state but to increase the binding kinetics so that the FlAsH binding process can more readily reach equilibrium. Concentrations of 1 mM have typically been used. Optimization for different systems is advised [27, 29] (*see* Fig. 3).
19. The amount of FlAsH dye to use can depend on the experiment and the sensitively of the fluorimeter. To quickly screen binding site designs with a yes/no output, a small amount of dye can be used (nM to μM range). For more quantitative experiments, the concentration of protein and therefore FlAsH binding sites will have an effect on how much FlAsH-EDT_2 is added so as to maintain saturated stoichiometric ratios.
20. A standard incubation time between additions is 2 h; however, this can vary if the binding kinetics of the specific system are atypical.
21. Fluorescence experiments for purified proteins are the same as the above procedure, except that the amount of FlAsH-EDT_2 dye used might be altered to maintain saturated stoichiometry. Purified proteins provide much higher confidence that any observed fluorescence is coming from the designed binding site than in lysate experiments, but this can be countered by the decreased throughput due to the time required for purification.

22. The time required to fully denature the protein may also require optimization. A positive control is especially useful here as often a small decrease in signal is observed in the presence of large amounts of guanidine HCl and this should be taken into account.

References

1. Kis K, Volk R, Bacher A (1995) Biosynthesis of riboflavin - studies on the reaction-mechanism of 6,7-dimethyl-8-ribityllumazine synthase. Biochemistry 34:2883–2892
2. Walter S, Buchner J (2002) Molecular chaperones - cellular machines for protein folding. Angew Chem Int Ed Engl 41:1098–1113
3. Homa FL, Brown JC (1997) Capsid assembly and DNA packaging in herpes simplex virus. Rev Med Virol 7:107–122
4. Aisen P, Listowsky I (1980) Iron transport and storage proteins. Annu Rev Biochem 49:357–393
5. Theil EC (1987) Ferritin - structure, gene-regulation, and cellular function in animals, plants, and microorganisms. Annu Rev Biochem 56:289–315
6. Douglas T, Strable E, Willits D, Aitouchen A, Libera M, Young M (2002) Protein engineering of a viral cage for constrained nanomaterials synthesis. Adv Mater 14:415–418
7. Fan R, Chew SW, Cheong VV, Orner BP (2010) Fabrication of gold nanoparticles inside unmodified horse spleen apoferritin. Small 6:1483–1487
8. Fletcher JM, Harniman RL, Barnes FRH, Boyle AL, Collins A, Mantell J, Sharp TH, Antognozzi M, Booth PJ, Linden N, Miles MJ, Sessions RB, Verkade P, Woolfson DN (2013) Self-assembling cages from coiled-coil peptide modules. Science 340:595–599
9. Tinberg CE, Khare SD, Dou J, Doyle L, Nelson JW, Schena A, Jankowski W, Kalodimos CG, Johnsson K, Stoddard BL, Baker D (2013) Computational design of ligand-binding proteins with high affinity and selectivity. Nature 501:212–216
10. Fan R, Boyle AL, Cheong VV, Ng SL, Orner BP (2009) A helix swapping study of two protein cages. Biochemistry 48:5623–5630
11. Ardejani MS, Li NX, Orner BP (2011) Stabilization of a protein nanocage through the plugging of a protein-protein interfacial water pocket. Biochemistry 50:4029–4037
12. Zhang Y, Fu J, Chee SY, Ang EXW, Orner BP (2011) Rational disruption of the oligomerization of the mini-ferritin *E. coli* DPS through protein-protein interface mutation. Protein Sci 20:1907–1917
13. Zhang Y, Orner BP (2011) Self-assembly in the ferritin nano-cage protein superfamily. Int J Mol Sci 12:5406–5421
14. Ardejani MS, Chok XL, Foo CJ, Orner BP (2013) Complete shift of ferritin oligomerization toward nanocage assembly via engineered protein-protein interactions. Chem Commun 49:3528–3530
15. Grueninger D, Treiber N, Ziegler MOP, Koetter JWA, Schulze M-S, Schulz GE (2008) Designed protein-protein association. Science 319:206–209
16. Ozawa T (2006) Designing split reporter proteins for analytical tools. Anal Chim Acta 556:58–68
17. Jares-Erijman EA, Jovin TM (2003) FRET imaging. Nat Biotechnol 21:1387–1395
18. Paulmurugan R, Umezawa Y, Gambhir SS (2002) Noninvasive imaging of protein–protein interactions in living subjects by using reporter protein complementation and reconstitution strategies. Proc Natl Acad Sci USA 99:15608–15613
19. Stains CI, Furman JL, Porter JR, Rajagopal S, Li Y, Wyatt RT, Ghosh I (2010) A general approach for receptor and antibody-targeted detection of native proteins utilizing split-luciferase reassembly. ACS Chem Biol 5:943–952
20. Miyawaki A, Llopis J, Heim R, McCaffery JM, Adams JA, Ikura M, Tsien RY (1997) Fluorescent indicators for Ca2+ based on green fluorescent proteins and calmodulin. Nature 388:882–887
21. Gratzner HG (1982) Monoclonal-antibody to 5-bromodeoxyuridine and 5-iododeoxyuridine - a new reagent for detection of DNA-replication. Science 218:474–475
22. Giepmans BNG, Adams SR, Ellisman MH, Tsien RY (2006) Review - the fluorescent toolbox for assessing protein location and function. Science 312:217–224
23. Neef AB, Schultz C (2009) Selective fluorescence labeling of lipids in living cells. Angew Chem Int Ed Engl 48:1498–1500

24. Griffin BA, Adams SR, Tsien RY (1998) Specific covalent labeling of recombinant protein molecules inside live cells. Science 281:269–272
25. Stroffekova K, Proenza C, Beam KG (2001) The protein-labeling reagent FLASH-EDT2 binds not only to CCXXCC motifs but also non-specifically to endogenous cysteine-rich proteins. Pflugers Arch 442:859–866
26. Luedtke NW, Dexter RJ, Fried DB, Schepartz A (2007) Surveying polypeptide and protein domain conformation and association with FlAsH and ReAsH. Nat Chem Biol 3:779–784
27. Goodman JL, Fried DB, Schepartz A (2009) Bipartite tetracysteine display requires site flexibility for ReAsH coordination. Chembiochem 10:1644–1647
28. Scheck RA, Lowder MA, Appelbaum JS, Schepartz A (2012) Bipartite tetracysteine display reveals allosteric control of ligand-specific EGFR activation. ACS Chem Biol 7:1367–1376
29. Cornell TA, Fu J, Newland SH, Orner BP (2013) Detection of specific protein–protein interactions in nanocages by engineering bipartite FlAsH binding sites. J Am Chem Soc 135:16618–16624
30. Martin BR, Giepmans BNG, Adams SR, Tsien RY (2005) Mammalian cell-based optimization of the biarsenical-binding tetracysteine motif for improved fluorescence and affinity. Nat Biotechnol 23:1308–1314