Methods in Molecular Biology 1266

Springer Protocols

Arnaud Gautier
Marlon J. Hinner *Editors*

Site-Specific Protein Labeling

Methods and Protocols

Humana Press

Contents

Chapter 1

How FlAsH Got Its Sparkle: Historical Recollections of the Biarsenical-Tetracysteine Tag

B. Albert Griffin, Stephen R. Adams, and Roger Y. Tsien

Abstract

The biarsenical-tetracysteine tagging system was the first of (and inspiration for) the now numerous methods for site-specifically labeling proteins in living cells with small molecules such as fluorophores. This historical recollection describes its conception and the trial-and-error chemical development required to become a versatile technique.

Key words FlAsH, ReAsH, Biarsenical, Tetracysteine, Protein labeling, Genetically targeted

Prior to the introduction of GFP, the ability to label and image specific proteins in living cells was severely limited and time-consuming. The protein(s) of interest were isolated from tissue or, as a recombinant protein in bacteria, labeled with a fluorescent dye after often considerable trial and error [1, 2], and then microinjected into the cytoplasm of suitably large and robust tissue culture cells. Even after the introduction of GFP [3, 4], labeling with different colors or spectroscopic phenotypes not limited to fluorescence was an unsolved problem. A simple method for modifying an expressed protein in any cell was desirable, but how could one protein be made more chemically reactive than the many others in a cell? The biarsenical-tetracysteine tag was inspired by a talk given by Jack Dixon at a FASEB meeting in 1994 on protein tyrosine phosphatases, in which he explained that their surprisingly specific inhibition by phenyl arsenoxide resulted from the requirement for vicinal cysteines and more importantly, the rarity of such sequences in proteins. Only up to a few tens of cysteine pairs were labeled by phenyl arsenoxide. This remark suggested (to R.Y.T.) that if two vicinal cysteines were better than one, then two pairs of cysteine (four total) might be still better. In other words, the desired chemical specificity might be achievable by incorporating tandem vicinal cysteines in expressed recombinant proteins in cells and labeling

Arnaud Gautier and Marlon J. Hinner (eds.), *Site-Specific Protein Labeling: Methods and Protocols*, Methods in Molecular Biology, vol. 1266, DOI 10.1007/978-1-4939-2272-7_1,

with a membrane-permeable fluorophore bearing two appropriately spaced arsenoxides. To prevent biarsenicals from binding to endogenous vicinal cysteines and thiols such as lipoic acid cofactors, one of the known 1,2-dithiol arsenic antidotes would be added at a concentration that could not compete with the expected higher tetracysteine-biarsenical affinity.

The peptide portion of the pair should be as small as possible to minimize potential disruption of the labeled protein's function. At the same time, it would be advantageous for the peptide to have some secondary structure, to mitigate the entropy cost of two arsenic atoms constraining four cysteines. The obvious candidate for a small peptide with secondary structure was an alpha-helix. Based on recently published studies on the role of natural amino acids in helix formation [5–7] we designed the following alanine-rich peptide: Acetyl-WEAAAREACCRECCARA-amide. Three arginine-glutamate pairs at i and i+4 positions form salt bridges, each across one turn of the helix. Glutamates toward the N-terminus and arginines toward the C-terminus interact favorably with the helix dipole. Cysteines are not strong alpha-helix formers by themselves, but induce alpha-helix formation in the i and i+4 positions when complexed with divalent cations [8]. Binding of the first arsenic to a pair of cysteines should nucleate an alpha-helix, positioning the other two cysteines favorably to bind the other arsenic. An even shorter peptide, without the N-terminal turn, was also prepared but proved to be difficult to maintain in the reduced state for purification and binding studies.

To validate the design, two dicysteine peptides were prepared, modeling two ways in which a single arsenic could bind to the tetracysteine binding site: i and i+1 (Acetyl-WEAAARECCARA-amide) and i and i+4 (Acetyl-WEACARECAARA-amide). Phenylarsine oxide (PAO) bound to both peptides as demonstrated by HPLC and MS. As expected, a single equivalent of the dithiol antidote 2,3-dimercaptopropanol (British anti-Lewisite or BAL) completely reversed binding. CD spectra confirmed dose-dependent alpha-helix formation when the i and i+4 peptide was titrated with 4-arsenophenylaminosulfonic acid, a more water-soluble analogue of PAO. Interestingly, in equilibrium experiments, PAO showed a preference for binding to the i and i+1 peptide. We reasoned that the arsenic atom might be too small to comfortably bridge the helix turn and perhaps a larger atom would be advantageous.

Phenylstibonoxide [9] was prepared to test this hypothesis. Antimony is larger (ionic radius of 0.76 Å vs. 0.58 Å) and shares much chemistry with arsenic. The antimony compound bound to both model peptides. In equilibrium experiments, phenylstibonoxide showed a preference for binding across alpha-helix turns (i and i+4) instead of to adjacent cysteines (i and i+1), opposite of the behavior of PAO, supporting the hypothesis that arsenic might not be sterically ideal. An organobismuth compound also bound to the i and i+4 peptide but was not further characterized. Since toxicity and biological stability of organoantimony and organobismuth

compounds were not as well studied as those of arsenic, the heavier group 15 elements would be pursued only if arsenic compounds failed to deliver the expected high tetracysteine affinity. We then set about the task of synthesis and testing of biarsenicals.

At the time of this research, reliable software was not available for modeling arsenic and peptides together, so we relied on Corey–Pauling–Koltun models to estimate that the arsenics should be spaced between 5 and 6 Å apart to fit well to the peptide. The first biarsenical compound tested for binding to the tetracysteine peptide was m-phenylenebisarsenoxide [10–12], in which the arsenics are 5.8 Å apart. There are no degrees of rotational freedom and a protruding C–H bond in the spacer between them. MS of the reaction mixture showed some of the expected mass, but also a peak that could only arise from a 2:2 or larger complex. After a few hours, no peaks were seen above noise. Our interpretation was that the early peaks were temporary kinetic products that went on to form larger polymers which precipitated. In designing the biarsenical probe, we had to be either incredibly lucky to get the spacing and orbital orientations correct, or include some degree of flexibility between the arsenics so that the molecule might adjust to the tetracysteine binding site. The first attempt was a shot in the dark that missed. Subsequent compounds included at least one degree of rotational freedom.

In a second molecule, 1,1′-bisarsenosoferrocene [13], an arsenic was attached to each cyclopentadienyl ring (Fig. 1a) via a mercurial intermediate. Rotation of the rings around the sandwiched iron ion would allow the probe to adjust to the tetracysteine binding site. The ferrocene compound induced a clean shift in HPLC retention time of the tetracysteine peptide when mixed stoichiometrically in neutral aqueous buffer. A mass spectrum of the collected fraction gave the expected molecular weight, clean of any peaks that would indicate larger complexes. However, two equivalents of BAL, one for each arsenic, rapidly returned the peptide to its original HLPC retention time. A quantitative, stable 1:1 bisarsenical/tetracysteine peptide complex had been achieved, but the hypothesized enhanced affinity that would make the system

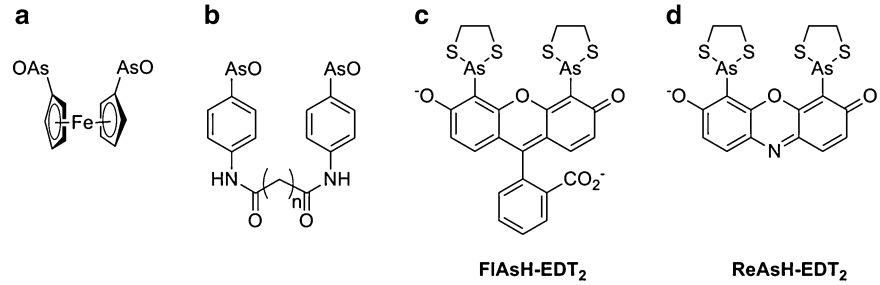

Fig. 1 The evolution of biarsenicals designed to bind tetracysteine (CCXXCC)-containing peptides. (**a**) 1,1'-bisarsenosoferrocene, an unsuccessful biarsenical. (**b**) Amides formed from 4-arsenosoaniline (also known as 4-aminophenylarsenoxide) and dicarboxylic acids, also unsuccessful. (**c**) FlAsH-EDT_2, the first successful biarsenical. (**d**) ReAsH-EDT_2, another successful biarsenical

useful in live cells was lacking. It would only be a matter of some tweaking of the spacer to achieve our goal.

Eleven compounds later the goal was no closer. Biarsenicals, some aliphatic, some aromatic, with a variety of spacers in a range of flexibilities were synthesized. Four of these bound the tetracysteine peptide, but all were easily dissociated by stoichiometric concentrations of small dithiols. These compounds were prepared by linking 4-arsenosoaniline via amide bonds (Fig. 1b) or by arsenic trioxide action on aliphatic bromides. A fresh approach was needed. We looked for commercially available compounds containing mercuric acetate or chloride groups, intending to employ the synthetic methodology used to prepare the ferrocene compound. One such compound from Aldrich was fluorescein mercuric acetate (FMA), which was sold with the compound name 2′,7′-bis(acetoxymercuri) fluorescein. In fluorescein nomenclature, substituents on the 2′ and 7′ positions protrude from opposite sides of the molecule, as displayed in the Aldrich catalog. This was of no use since the arsenics needed to be closer together. The Aldrich Library of ^{1}H FT-NMR Spectra was sitting on the benchtop and for unknown reasons we flipped it open and turned to the FMA page. It was immediately apparent from proton coupling that the hydrogens in the fluorophore region of the molecule were ortho to each other and not para as in the structure as drawn, which meant that the mercury atoms were in the 4′ and 5′ positions, close and aligned parallel to each other on the same side of the molecule. Although unlikely that a compound with no flexibility between the arsenics would succeed where our first rigid compound had failed, the chemistry was straightforward and the compound was attractive since a useful chromophore was built in the biarsenical (as a footnote, Aldrich was notified of the error in their catalog and it was subsequently corrected). Isolated yield from early preparations was quite low (<1 %). Separation of the desired compound from the complex mixture was accomplished by complexing the reaction products with aliphatic dithiols, 1,3-propanedithiol (PDT) or 1,2-ethanedithiol (EDT). Mono-substituted arsenoxides migrate cleanly on a thin-layer chromatogram when complexed with small dithiols that form a tight S-As-S containing ring, shielding the arsenic from silica oxygens. BAL was not used because it was thought that its chirality might give rise to multiple isomers, complicating product characterization. To our great delight, the compound not only bound the tetracysteine peptide, but persisted in the presence of a moderate excess of BAL. We promptly named it *Fl*uorescein *Ars*enic *H*elix-binder (FlAsH, Fig. 1c). To our even greater surprise FlAsH was totally nonfluorescent as the bis-1,2-ethanedithiol (EDT) adduct. But when EDTs were replaced by the tetracysteine peptide, it becomes brightly fluorescent. Our explanation, buttressed by molecular models, was that the thiol groups could become coplanar with the fluorescein in the ethanedithiol complex, allowing quenching in the excited state,

but were constrained to twist out of the fluorescein plane in the peptide complex.

Once the remarkable properties of the compound were confirmed, we set about improving the synthetic yield. Two alternate methods were explored, arsenic trichloride action on 4′,5′-bis-diazoniumfluorescein (Bart conditions) and direct arsonation of fluorescein using arsenic trioxide [14–16]. While both techniques had worked in our hands on other substrates, neither produced detectable quantities of the desired product. Revisiting the bis-mercuric compound, Prof. Jay Siegel suggested that a catalytic amount of palladium(II) might facilitate substitution of arsenic for mercury. This added reagent did indeed greatly improve the yield and is the method still used today to synthesize FlAsH-EDT_2. Overall, our first paper on the remarkable partnership between FlAsH and tetracysteine peptides was (surprisingly) accepted in *Science* [17] and launched the field of site-specific small-molecule labeling of proteins in living cells.

For final confirmation that the labeling system had worked in vitro as originally envisioned, CD spectra of the pair were collected to confirm the presence of an alpha-helix. When there was no evidence of helix in the spectra, we were at first perplexed. But then we concluded that the helix must be present but its CD signature was obscured by fluorescein absorbance. The PAO-dicysteine peptide adduct was helical, so we believed FlAsH-tetracysteine peptide should be as well! Later, after peptides containing a helix-breaking proline-glycine (PG) between the cysteine pairs showed a higher affinity for FlAsH [18], and sufficient material with the resorufin-based biarsenical ReAsH (Fig. 1d) was available for NMR spectroscopy [19], those peptides were found to have adopted a beta turn conformation rather than an alpha-helix. Fortunately "helix-binder" and "hairpin-binder" both start with "h", so the acronym did not have to be revised.

Studies from our group [17–29] and others have described many applications of this biarsenical-tetracysteine tagging system, summarized in a recent review [30], which is why this article is a historical reminiscence.

References

1. Taylor DL, Amato P, Luby-Phelps K et al (1984) Fluorescent analog cytochemistry. Trends Biochem Sci 9:88–91
2. Adams SR, Bacskai BJ, Taylor SS, Tsien RY (1993) Optical probes for cyclic AMP. In: Mason WT, Relf G (eds) Fluorescent probes for biological activity of living cells. Academic, London, pp 133–149
3. Chalfie M, Tu Y, Euskirchen G et al (1994) Green fluorescent protein as a marker for gene expression. Science 263:802–805
4. Heim R, Prasher DC, Tsien RY (1994) Wavelength mutations and posttranslational autoxidation of green fluorescent protein. Proc Natl Acad Sci U S A 91:12501–12504
5. O'Neil KT, DeGrado WF (1990) A thermodynamic scale for the helix-forming tendencies of the commonly occurring amino acids. Science 250:646–651
6. Merutka G, Shalongo W, Stellwagen E (1991) A model peptide with enhanced helicity. Biochemistry 30:4245–4248

7. Merutka G, Stellwagen E (1991) Effect of amino acid ion pairs on peptide helicity. Biochemistry 3:1591–1594
8. Ghadiri MR, Cho C (1990) Secondary structure nucleation in peptides. Transition metal ion stabilized alpha-helices. J Am Chem Soc 112:1630–1632
9. Doak GO, Steinman HG (1946) The preparation of stibonic acids by the scheller reaction. J Am Chem Soc 68:1987–1989
10. Steinkop W, Schmidt S, Penz H (1934) Zur kenntnis organischer arsenverbindungen, XVII. Über ein polymeres phenylarsinoxyd und über reaktionen der m-phenylendiarsinsäure. J Prakt Chem 141:301–305
11. Lieb H (1921) Aromatische diarsinsäuren und deren reduktionsprodukte (I. Mitteilung). Berichte d D Chem Gesellschaft 54:1511–1519
12. Doak GO (1940) A modified Bart reaction. J Am Chem Soc 62:167–168
13. Spang C, Edelmann FT, Noltemeyer M, Roesky HW (1989) Anorganische ringsysteme mit ferrocenyl-substituenten. Chem Ber 122:1247–1254
14. Michaelis A (1908) Über die p-dimethylaminophenyl-arsinsäure (dimethylatoxyl). Berichte d D Chem Gesellschaft 41:1514–1516
15. Jacobs WA, Heidelberger M (1919) The isomeric hydroxyphenylarsonic acids and the direct arsenation of phenol. J Am Chem Soc 41:1440–1450
16. Benda L (1909) Über o-amino-arylarsinsäuren. Berichte d D Chem Gesellschaft 42:3619–3622
17. Griffin BA, Adams SR, Tsien RY (1998) Specific covalent labeling of recombinant protein molecules inside live cells. Science 281:269–272
18. Adams SR, Campbell RE, Gross LA et al (2002) New biarsenical ligands and tetracysteine motifs for protein labeling in vitro and in vivo: synthesis and biological applications. J Am Chem Soc 124:6063–6076
19. Madani F, Lind J, Damberg P et al (2009) Hairpin structure of a biarsenical-tetracysteine motif determined by NMR spectroscopy. J Am Chem Soc 131:4613–4615
20. Griffin BA, Adams SR, Jones J, Tsien RY (2000) Fluorescent labeling of recombinant proteins in living cells with FlAsH. Methods Enzymol 327:565–578
21. Gaietta G, Deerinck TJ, Adams SR et al (2002) Multicolor and electron microscopic imaging of connexin trafficking. Science 296:503–507
22. Tour O, Meijer RM, Zacharias DA et al (2003) Genetically targeted chromophore-assisted light inactivation. Nat Biotechnol 21:1505–1508
23. Ju W, Morishita W, Tsui J et al (2004) Activity-dependent regulation of dendritic synthesis and trafficking of AMPA receptors. Nat Neurosci 7:244–253
24. Hoffmann C, Gaietta G, Bünemann M et al (2005) A FlAsH-based FRET approach to determine G protein-coupled receptor activation in living cells. Nat Methods 2:171–176
25. Martin BR, Giepmans BN, 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
26. Gaietta GM, Giepmans BN, Deerinck TJ et al (2006) Golgi twins in late mitosis revealed by genetically encoded tags for live cell imaging and correlated electron microscopy. Proc Natl Acad Sci U S A 103:17777–17782
27. Tour O, Adams SR, Kerr RA et al (2007) Calcium green FlAsH as a genetically targeted small-molecule calcium indicator. Nat Chem Biol 3:423–431
28. Martin BR, Deerinck TJ, Ellisman MH et al (2007) Isoform-specific PKA dynamics revealed by dye-triggered aggregation and DAKAP1alpha-mediated localization in living cells. Chem Biol 14:1031–1042
29. Adams SR, Tsien RY (2008) Preparation of the membrane-permeant biarsenicals FlAsH-EDT_2 and ReAsH-EDT_2 for fluorescent labeling of tetracysteine-tagged proteins. Nat Protoc 3:1527–1534
30. Hoffmann C, Gaietta G, Zürn A et al (2010) Fluorescent labeling of tetracysteine-tagged proteins in intact cells. Nat Protoc 5:1666–1677