UC San Diego
UC San Diego Electronic Theses and Dissertations

Title

Optimization of protein and RNA detection methodologies and a new approach for manipulating protein activity in living cells

Permalink

https://escholarship.org/uc/item/2b03q5zx

Author

Martin, Brent R.

Publication Date

2006

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA, SAN DIEGO

Optimization of protein and RNA detection methodologies and a new

approach for manipulating protein activity in living cells

A dissertation submitted in partial satisfaction of the requirements

for the degree Doctor of Philosophy

in

Biomedical Sciences

by

Brent R. Martin

Committee in charge:

        Professor Roger Tsien, Chair
        Professor Mark Ellisman
        Professor Xiang-Dong Fu
        Professor Gerald Joyce
        Professor Susan Taylor
        Professor Inder Verma

2006

This dissertation of Brent R. Martin is approved,

and is acceptable in quality and form for publication on microfilm:

_____

_____

_____

_____

_____

_____
Chair

University of California, San Diego

2006

I dedicate this work to my father, Albert Martin, my first mentor and most

successful collaborator.

Table of Contents

Chapter 1: Mammalian Cell-Based Optimization of the Biarsenical-binding Tetracysteine Motif for Improved Fluorescence and Affinity

Chapter 2: Inducible aggregation of tetracysteine-GFP fusion proteins for reversible protein inactivation

Chapter 3: New strategies and progress towards enhancing the specificity of trans-splicing RNAs in mammalian cells

List of Figures and Tables

Acknowledgements

I would like to thank several former members of the Tsien lab, including Grant Walkup, David Zacharias, Alice Ting, Robert Campbell, Jin Zhang, Amy Palmer, and Coyt Jackson for passing down their knowledge and helping me achieve my research goals. Also, many thanks to Oded Tour, Christina Hauser, Paul Steinbach, Qing Xiong, Tom Deerinck, and other members of the FlAsHers group for assistance and advice. Most importantly, I would like to acknowledge my closest collaborators, Stephen Adams and Ben Giepmans whose guidance and encouragement has been invaluable. I would also like to thank James Lim for assistance with several experiments discussed in Chapter 3, specifically those involving RIα knockout fibroblasts. Finally I would like to thank Roger Tsien for supporting me and placing me in such an excellent research environment and the members of my thesis committee for making time and giving invaluable advice. On a personal note, I would like to thank my parents and wife Monica for their constant support.

The text of Chapter 1, in part, is a reprint of the material as it appears in Nature Biotechnology (Citation: Martin, B.R., Giepmans, B.N.G., Adams, S.R., Tsien, R.Y. Nature Biotechnology 23, 1308-1314 (2005), http://www.nature.com/nbt). I was the primary researcher and author and the co-authors listed in this publication contributed or supervised the research which forms the basis for this chapter.

Vita

Education

9/95 – 6/99   B.S. in Molecular Biology, Cum Laude, Provost Honors, University

of California, San Diego

9/99 – present   Graduate Student, Biomedical Sciences Graduate Program,

University of California, San Diego

5/06   Ph.D., Biomedical Sciences, University of California, San Diego

Publications

Adams SR, Campbell RE, Gross LA, Martin BR, Walkup GK, Yao Y, Llopis J, and Tsien

RY. 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.

Martin BR, Giepmans BN, Adams SR, Tsien RY. 2005 Mammalian cell-based

optimization of the biarsenical-binding tetracysteine motif for improved fluorescence and

affinity. *Nature Biotechnol.* 10:1308-1314.

Martin BR and Tsien RY. Inducible aggregation of tetracysteine-GFP fusion proteins for

reversible protein inactivation. *In Preparation.*

Poster Presentations

Martin BR, Giepmans BN, Adams SR, Tsien RY. Optimization of the Biarsenical Binding

Tetracysteine Motif for Fluorescence and Affinity and Discovery of a Reversible Tag for

Protein Aggregation. Imaging Technology, The American Society for Cell Biology Annual

Meeting (2005), San Francisco. *(Highlighted in Nature Chemical Biology 2, 119-122 (2006))*.

Martin BR, Jackson WC, Tsien RY. Mammalian cell-based directed evolution of the biarsenical binding tetracysteine peptide for improved fluorescence and dithiol resistance. Program No. 124.11, Society for Neuroscience, Annual Meeting (2004), San Diego.

Giepmans BN, Martin BR, Gaietta GM, Deerinck TJ, Adams SR, Tsien RY, Ellisman MH Visualizing Cytoskeletal Dynamics and Ultrastructure at Cell-cell Junctions Using Genetically Encoded Tags. Cytoskeleton-Membrane Interactions I, The American Society for Cell Biology Annual Meeting (2004), Washington D.C. (Presented by BN Giepmans).

Martin BR, Jackson WC, Tsien RY. Mammalian cell-based directed evolution of the biarsenical-binding tetracysteine motif for improved fluorescence and affinity. Imaging Technology, The American Society for Cell Biology Annual Meeting (2003), San Francisco.

ABSTRACT OF THE DISSERATION

Optimization of protein and RNA detection methodologies and a new approach

for manipulating protein activity in living cells

by

Brent R. Martin

Doctor of Philosophy in Biomedical Sciences

University of California, San Diego, 2006

Professor Roger Tsien, Chair

The orchestrations that underlie the existence of even the simplest organisms are quite complex and extremely dynamic. In order to gain a greater understanding of the biochemistry underlying many unsolved biological problems, new tools are required to first visualize a phenomenon, and then perturb it to study its significance. Visualizing the dynamics of intracellular biochemistry has been enhanced greatly with widespread adoption of genetically encoded fluorescent proteins.

Due to the large size of fluorescent proteins and their lack of chemical flexibility, the tetracysteine-biarsenical system was developed. This technology uses the combination of a short genetically encoded tag and a specific class of exogenous, membrane-permeant dyes. Since its introduction, the biarsenical-tetracysteine system has suffered from spontaneous background staining, preventing the detection of dilute proteins. To remedy this problem, a library of tetracysteine sequences was screened for improved dithiol resistance and brightness. Several new sequences were discovered

and background staining was reduced significantly, resulting in a 20-fold increase in labeling contrast.

One unique sequence isolated from the library found an unexpected mechanism to ensure its selection. Upon ReAsH labeling, YRECCPGCCMWR-GFP rapidly aggregates into tiny, highly fluorescent speckles. Upon bleaching ReAsH, the aggregates dissociate, dispersing the tagged protein throughout the cell. Fusions of this tag on several cellular proteins led to ReAsH dependent aggregation of the tagged protein as wells as endogenous binding partners. By sequestering protein complexes in the aggregates, activity is inhibited.

Finally, the detection of specific RNAs in living cells remains a major challenge in biology with numerous potential applications. Trans-splicing repair of clinically relevant transcripts has been reported as an efficient and specific method for delivering exogenous message for translation. Therefore, a crippled reporter gene lacking translation initiation sites gene was targeted using existing trans-splicing techniques to an expressed RNAs. Trans-splicing then leads to the conversion of the targeted mRNA into a chimeric mRNA capable of translating an active protein. After significant effort and several novel approaches to enhance specificity, it became clear that new methods of RNA detection will be required to prevent non-specific splicing of cargo RNAs in cells.

Chapter 1

Mammalian Cell-Based Optimization of the Biarsenical-binding Tetracysteine

Motif for Improved Fluorescence and Affinity

<u>Abstract</u>

Membrane-permeant biarsenical dyes such as FlAsH and ReAsH fluoresce upon binding to genetically encoded tetracysteine motifs expressed in living cells[1,2], yet spontaneous non-specific background staining conceals weakly expressed or dilute proteins from detection[1,3]. If the affinity of the tetracysteine peptide could be increased, then stringent dithiol washes should increase the contrast between specific and nonspecific staining. Residues surrounding the tetracysteine motif were randomized and fused to GFP, then retrovirally transduced into mammalian cells and iteratively sorted by fluorescence-activated cell sorting for high FRET from GFP to ReAsH despite increasing concentrations of dithiol competitors. Selected sequences demonstrate higher fluorescence quantum yields and drastically improved dithiol resistance, culminating in a >20-fold increase in contrast. The best tetracysteine sequences, FLNCCPGCCMEP and HRWCCPGCCKTF, maintain their enhanced properties as fusions to either termini of GFP or directly to β-actin. The new biarsenical-tetracysteine motif and should enable detection of a much broader spectrum of cellular proteins.

<u>Introduction</u>

Biarsenical-tetracysteine labels are analogous to fluorescent protein fusions[2], yet offer several unique capabilities such as correlative fluorescence and electron microscopy (EM)[4], determination of protein age by multi-color fluorescence pulse chase[4,5], chromophore-assisted light inactivation (CALI) for spatio-temporal inactivation

of proteins[6-8], and numerous other *in vitro* applications[1]. Additionally, the tetracysteine

sequence consists of only a few amino acids, far smaller and potentially less

perturbative than incorporation of a fluorescent protein[9-13]. Furthermore, biarsenical-

tetracysteines are detectable immediately following tetracysteine translation[14,15], allowing

visualization of early events in protein synthesis, in contrast to the intrinsic delays

required for fluorescent protein maturation[16]. Several other protein fusion partners can

also trap distinctive tags in or on living cells[17], however, these proteins are either an

order of magnitude larger than tetracysteines, require secondary processing enzymes,

lack a general ability to label intracellular targets, or have no demonstrated expanded

functionality.

The earliest designs of tetracysteine sequences were intended to encourage α-

helicity under the assumption that the biarsenical would ideally fit into the i, i+1, i+4, and

i+5 positions of an α-helix[18]. With these sequences, non-specific biarsenical background

staining was estimated to equal the fluorescence of several micromolar of labeled

protein[1,3], and was partially reduced by increasing the concentration of the dithiols 1,2-

ethanedithiol (EDT) or 2,3-dimercaptopropanol (BAL) in washes to remove thiol

dependent background or by including non-fluorescent dyes to block hydrophobic

binding sites[2]. When the helix-breaking amino acids PG were inserted between the CC

pairs, the resulting tetracysteine significantly enhanced the affinity and contrast of FlAsH

labeled tetracysteine fusion proteins in cells, increasing the tolerable concentration of

dithiol competitors without detrimental loss of specific fluorescence[1]. However, only a

few pairs of amino acids were tested between the cysteines, and the surrounding

residues were left unaltered, maintaining the α-helical bias.

To optimize the tetracysteine sequence for improved ReAsH affinity and

fluorescence, we developed a retrovirally transduced mammalian cell-based library

approach for fluorescent selection of optimal residues surrounding the tetracysteine motif by fluorescence-activated cell sorting (FACS). Other complementary approaches, such as surface display on phage or bacteria[19], pan libraries for high affinity binders *in vitro*, disregarding maintenance of desirable fluorescence properties. By performing these selections in the reducing environment of mammalian cytosol, we intended to avoid disulfide formation and evolve peptides with improved specificity, activity, toxicity, and expression in the environment most important to us.

Results and Discussion

We created our first library, ReAsH Retroviral Library 1 (RRL1), by ligating a semi-randomized oligonucleotide cassette to the C terminus of green fluorescent protein (GFP) in a retroviral cloning vector (**Fig. 1.1a**). NIH3T3 cells were infected with the recombinant viral library at a low multiplicity of infection (MOI), stained with ReAsH and analyzed by flow cytometry. Measurement of the GFP quench and GFP-sensitized FRET (fluorescence resonance energy transfer) to ReAsH emission allows for determination of the kinetics and extent of ReAsH labeling in a single cell[1]. ReAsH binding was detectable in cells expressing GFP fused to AEAAARECCRECCARA[18] (αRE), our first generation tetracysteine sequence, and RRL1 cells, as compared to cells expressing GFP alone (**Fig. 1.1b**). Interestingly, the RRL1 cells showed varying levels of FRET after dithiol washing, indicating different amino acid combinations near the tetracysteine are capable of modulating dithiol resistance and/or fluorescent properties of the complex. FRET positive RRL1 cells were collected and expanded (**Fig. 1.1c**, left). To discriminate higher affinity peptides, three additional rounds of sorting were performed, each time increasing the selection pressure by escalating the dithiol

concentration during washing. Finally, single cells were sorted (**Fig. 1.1c**, right) on a 96-well plate.
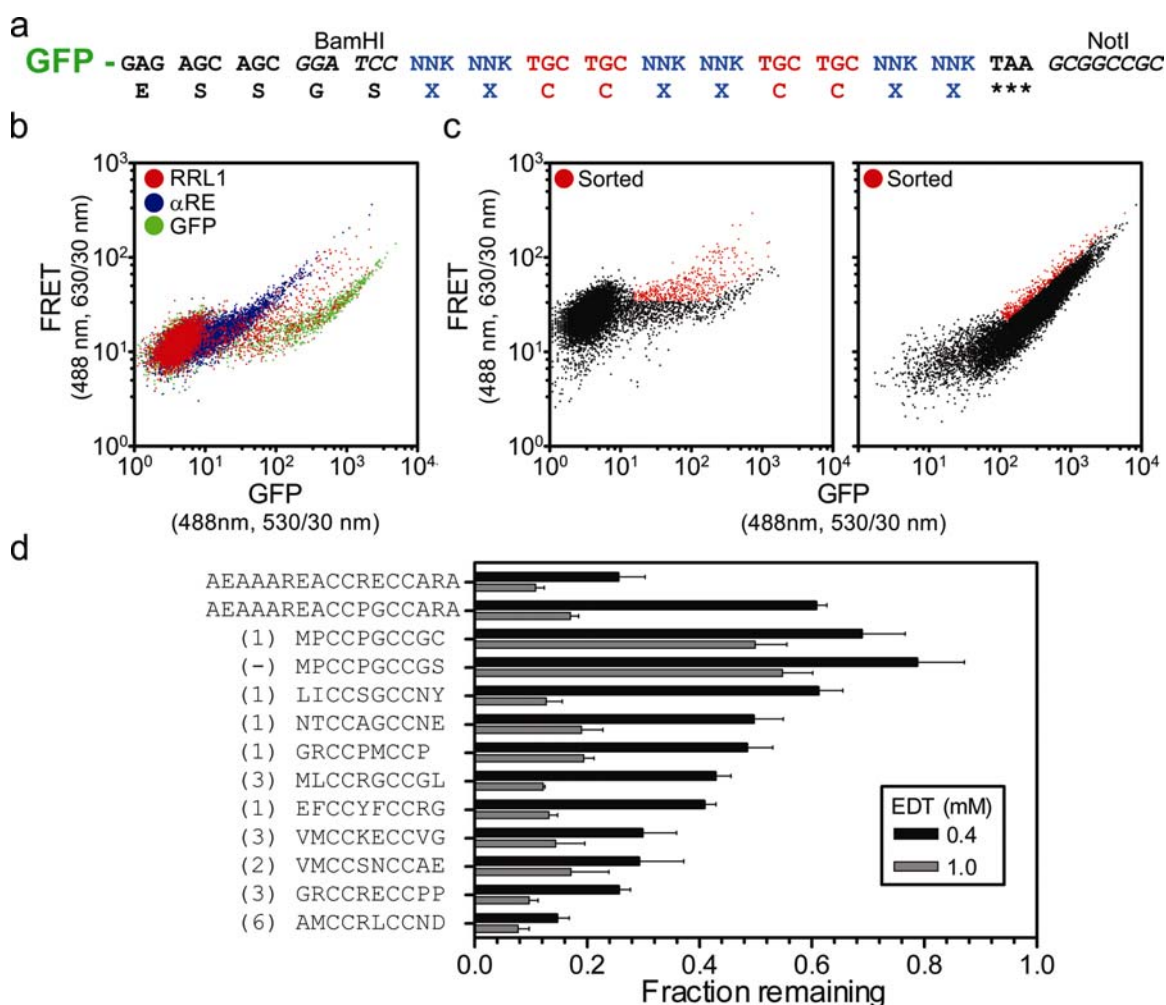
Figure 1.1. RRL1 selection for improved tetracysteine sequences. (a) Schematic of RRL1 cloning strategy. Unique restriction sites (italic), randomized codons (blue), and cysteine codons (red) are indicated. (b) FACS analysis of ReAsH binding by FRET. NIH3T3 cells virally transduced with either GFP-RRL1 (red), GFP-αRE (blue), or GFP alone (green) following ReAsH staining and a 30 min 0.1 mM BAL wash. ReAsH binding is characterized by an increase in FRET (630/22 nm) and a decrease in GFP (530/30 nm) fluorescence, which appear on a log scale as a shifts upwards and leftwards. (c) RRL1 FACS selections. Cells collected (red) in sort 1 (left) and sort 4 (right). (d) Sequence results and analysis for dithiol resistance. Unique sequences isolated in the RRL1 selection are listed, with the number of identical clones isolated in parenthesis. (-) indicates an additional peptide deliberately generated rather than isolated from the selection. The dithiol resistance of ReAsH fluorescence is shown for each sequence determined from live cell imaging experiments. Measurements represent the average of more than five cells following acute treatment with 0.4 mM and 1.0 mM EDT to ReAsH labeled cells. Background subtracted, total ReAsH fluorescence before washing is normalized to 1, representing saturated labeling.

Sequence analysis of the isolated clones established ten novel tetracysteine sequences (**Fig. 1.1d**). MPCCPGCCGC was highly resistant to EDT, maintaining 50% of its total ReAsH fluorescence in the face of 1.0 mM EDT, while αRE and AEAAARECCPGCCARA[1] (αPG), our second generation tetracysteine, both retained less than 20% (**Fig. 1.1d**). The next best four peptides all contained either internal prolines or glycines, corroborating the superiority of hairpin turns over helical conformations. Replacement of the cysteine in the final randomized position of MPCCPGCCGC to a serine showed no effect on dithiol resistance, proving the last cysteine was a fortuitous non-participant in arsenical binding.

Instead of optimizing a single tetracysteine sequence for improved contrast, an overlooked approach for increasing the biarsenical-tetracysteine contrast is to attach multiple tetracysteines to a single protein. By fusing several tetracysteines locally to a single protein, it should be possible to increase the local concentration of the fluorescent complex, enhancing the relative brightness as compared to non-specific background fluorescence. To test this idea, a series of tandem tetracysteines was constructed as C-terminal fusions to ECFP, diagramed as ECFP-ESSGS(MPCCPGCCGS)$_n$. Expression levels in both bacteria and mammalian cells were inhibited by increasing multiples of tetracysteines. N-terminally 6-his-tagged recombinant protein was expressed in bacteria, labeled with FlAsH, and then purified by a Ni-NTA column. The resulting protein was nearly completely composed of non-oxidized, monomeric, FlAsH-labeled protein (**Fig. 1.2a**). After measuring the quantum yield of each multiple FlAsH-tetracysteine complex, it was observed that by increasing the number of fluorophores, the quantum yield was quenched (**Fig. 1.2b**). Two tetracysteines on ECFP doubled the overall brightness of FlAsH, yet further addition of tetracysteines gave diminishing results. When expressed in cells, no additional brightness was observed with two tetracysteines versus one

tetracysteine following FlAsH labeling (**Fig. 1.2c**). To explore this effect further, GFP with

5 tandem tetracysteines was attached to the C-terminus of the gap junction protein Cx43

(**Fig. 1.2d-e**). Following addition of ReAsH, the fluorescence increased quickly, then

slowly decayed as labeling became saturated. Following incremental dithiol washing, the

fluorescence increased again. These results imply that the level of tetracysteine

occupancy correlates with the output fluorescence. Early in staining as ReAsH-

tetracysteine complexes first form, no fluorescence quenching occurs. Later in staining,

as each vacant tetracysteine site is filled, dye-dye interactions lead to strong

fluorescence quenching and diminished fluorescence signal. This quenching can be

relieved by the incremental disruption of ReAsH-tetracysteine complexes using dithiol

washes. Overall, as the number of tetracysteines linked in tandem increases, the

fluorescence diminishes. Furthermore, no increases in contrast were observed following

ReAsH photo-oxidation for EM (data not shown). Because of the lower expression and

decreased fluorescence, this avenue towards increased contrast was set aside, and

attention was refocused at increasing the brightness and affinity of a single tetracysteine

tag.
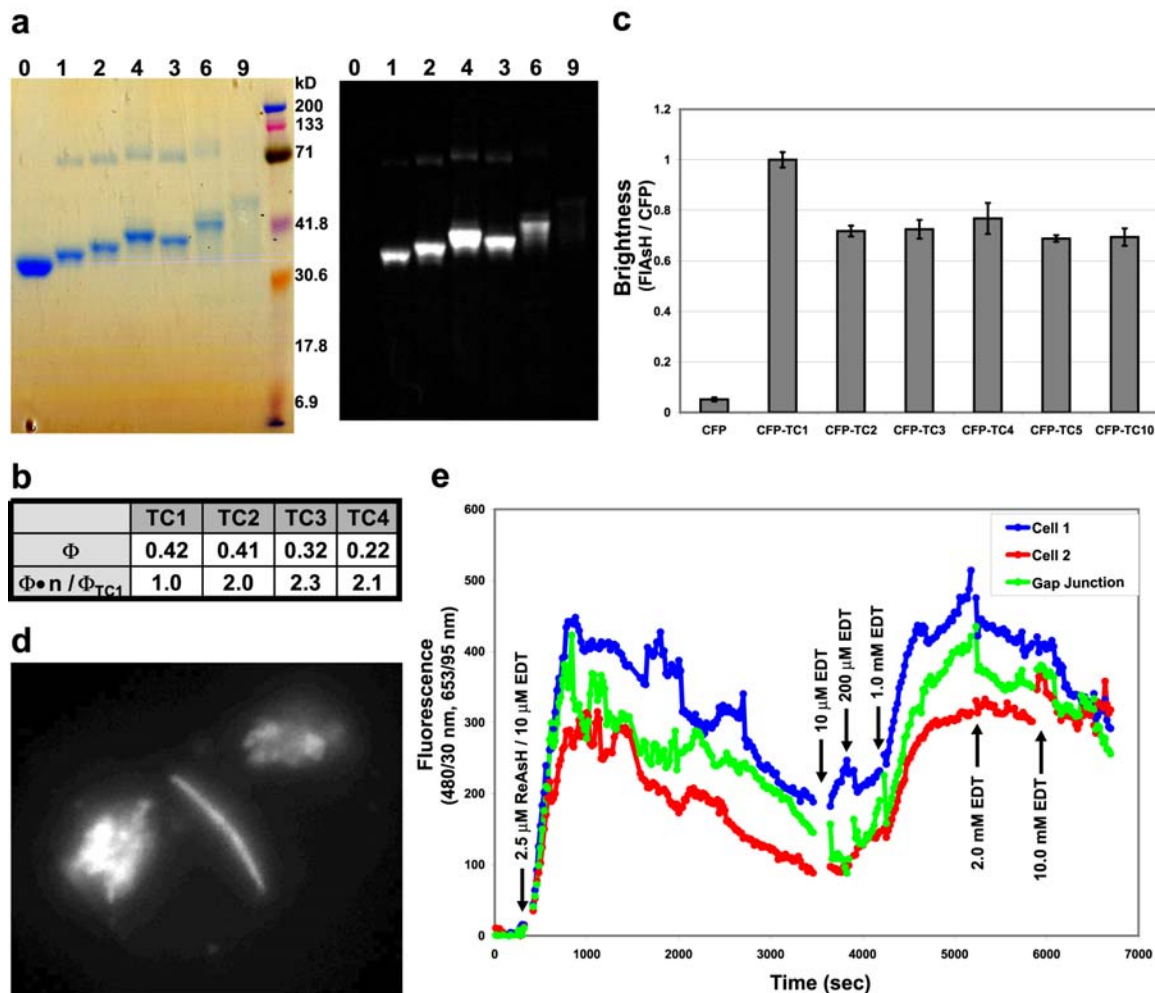
Figure 1.2. Multiple tetracysteines do not enhance contrast in cells. (a) Gel analysis of bacterially expressed, FlAsH-labeled and purified 6-His-ECFP-$(TC)_n$ protein. Coomassie stained SDS-PAGE gel of purified protein (left), with contrast enhanced. The protein runs predominantly as a monomer, yet some oxidized dimer is visible in samples containing tetracysteine sequences, implying incomplete or partial FlAsH labeling. This incomplete binding is due to the Ni-NTA purification scheme, which does not exclude oxidized or modified tetracysteines, as does the FlAsH bead purification protocol. Also, FlAsH fluorescence quantum yields resulting from Ni-NTA purifications are generally lower, due to the mixture of oxidized tetracysteines in the protein sample. FlAsH fluorescence of the labeled protein is also observed (right) and shows a trend of increased fluorescence, relative to the amount of protein seen in the Coomassie stained gel, as the number of tetracysteines increases. (b) Quantum yields of multiple tandem tetracysteines. Φ = fluorescence quantum yield, n = number of tetracysteines. The improvement in the overall fluorescence output of a single protein is written as $\Phi \bullet n / \Phi_{TC1}$. (c) FlAsH brightness is decreased in cells expressing multiply tetracysteine tagged CFP. Background subtracted FlAsH fluorescence per measured amount of pre-stained CFP fluorescence is lower with multiple tetracysteine tags. (d) HeLa cells expressing Cx43-GFP-TC5 are linked by a large gap junction, as seen by GFP sensitized ReAsH FRET emission approximately 1000 sec after labeling. (e) Timecourse of ReAsH staining of Cx43-GFP-TC5 in HeLa cells shows quenching of fluorophores after initial increases in florescence. Regions were drawn around each cell and the shared gap junction.

The results from the RRL1 selection confirmed the consensus sequence CCPGCC and verified the usefulness of the mammalian cell-based library approach for optimization of the tetracysteine motif. In an effort to further optimize the ReAsH binding tetracysteine peptide, a new library, RRL2, was devised, fixing PG as the internal residues, while randomizing the three external residues on either side of the tetracysteine, XXXCCPGCCXXX, and abbreviated XXX#XXX (# = CCPGCC) (**Fig. 1.3a**).

Three hundred million HEK293 cells were infected with RRL2 virus and pre-sorted for GFP expression, isolating thirty million cells. A 568 nm laser was added to directly excite ReAsH, allowing cells to be sorted based on two criteria: FRET ratio (GFP-sensitized ReAsH emission divided by GFP emission) and directly-excited ReAsH emission. The GFP+ cells were stained with ReAsH and sorted for multiple rounds, each round selecting the best 10-15% of the total population (**Fig. 1.3b**) with the goal of eliminating unfavorable cells over the course of several selections, each time amplifying the pool of selected cells in culture for better sampling of each genotype. After ten rounds, the population fell into two categories: one exhibiting high ReAsH fluorescence and a low FRET ratio, and the other displaying a high FRET ratio but lower ReAsH fluorescence. Each phenotype was simultaneously separated by sorting with two streams, one pool for each phenotype. After four more rounds of sorting, cells were washed with a low concentration of dithiol and sorted into 96-well plates to determine the composition of each population (**Fig. 1.3b**, middle).
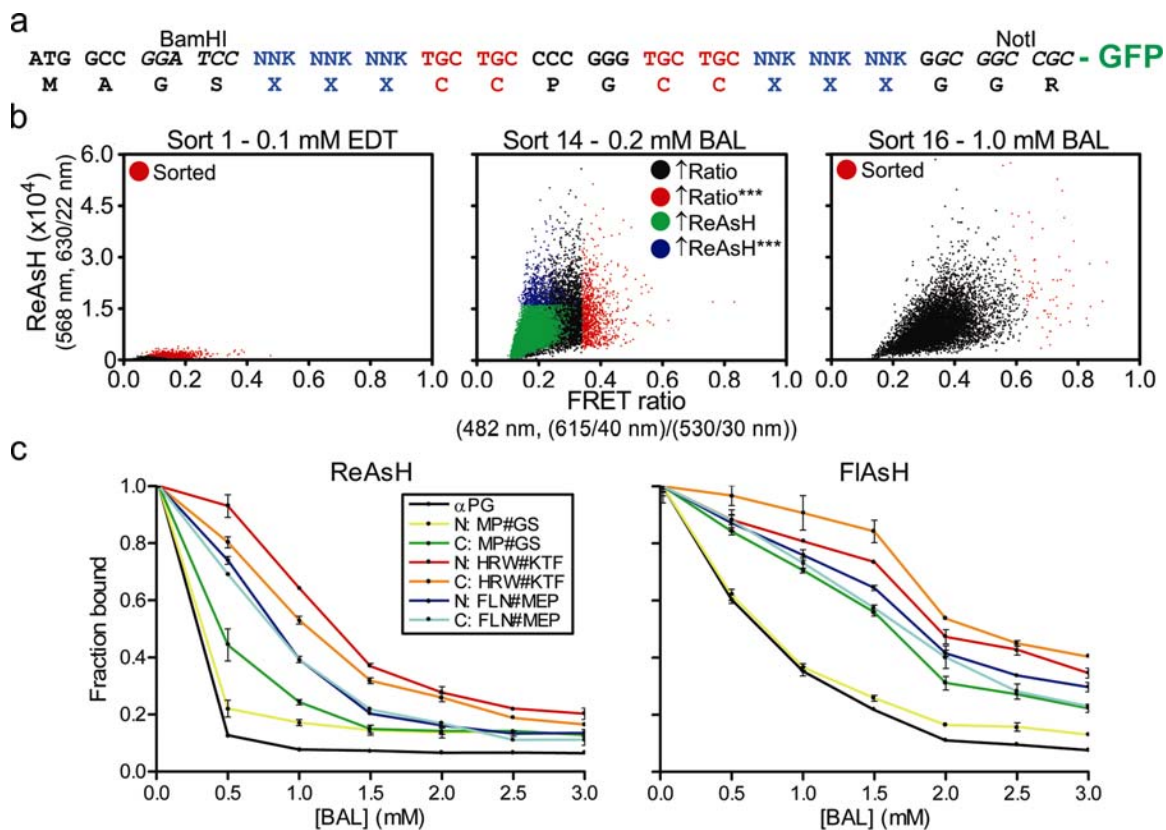
Figure 1.3. RRL2 selection and analysis of optimized flanking residues. (a) Schematic of RRL2 cloning strategy. Notation as in **Fig. 1.1a**. (b) RRL2 sort history. The FRET ratio is plotted versus ReAsH intensity in individual cells following a given BAL wash. In sort 1 (left), cells with high FRET ratios and ReAsH intensities were collected (red). Sort 14 (middle) shows the separation of high FRET ratio cells (↑Ratio, red/black) from high ReAsH intensity cells (↑ReAsH, blue/green) and the corresponding sorted fraction (***). In sort 16 (right), the final clones were selected from the top few percent in the high FRET ratio population. (c) Dithiol resistance of final optimized tetracysteines. BAL titration of ReAsH (left) and FlAsH (right) labeled N and C terminal optimized tetracysteine fusions to GFP and Cerulean respectively. Tetracysteine color notation is the same in both ReAsH and FlAsH titrations. Dithiol resistance is shown as the average fraction of the FRET ratio remaining following incremental washes with high concentrations of BAL, shown with corresponding standard deviations derived from three or more duplicate wells on a 96-well plate.

Following sequencing and analysis, the high ReAsH intensity, low FRET ratio clones displayed massive over-expression of the tetracysteine-GFP fusion, but weak dithiol resistance and poor labeling efficiency (data not shown). By excluding the GFP excitation data during the selections, protein expression levels were left uncorrected, leading to overexpression rather than high affinity. On the contrary, the high FRET ratio, lower ReAsH fluorescence population was dominated by sequences with better or equal

dithiol resistance as MP#GS (**Fig. 1.4**), while still expressing high levels of the fusion protein.
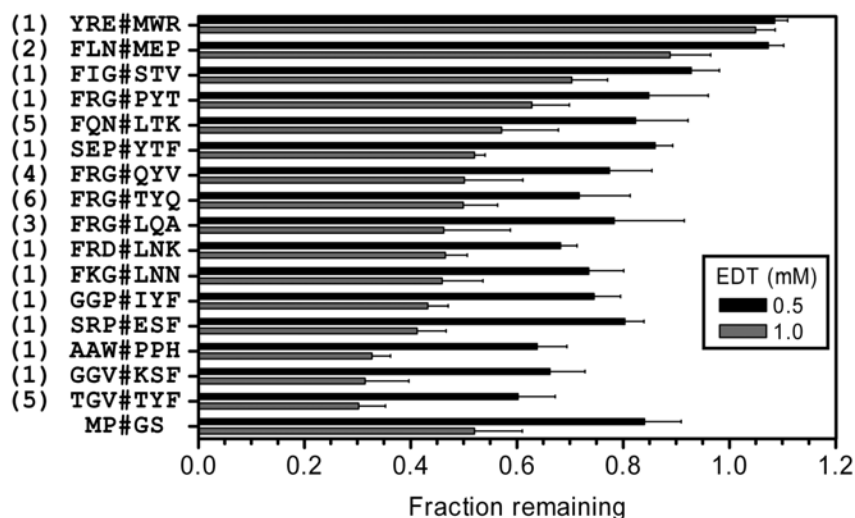
Figure 1.4. Analysis of unique sequences isolated in Sort 14. Unique sequences are listed next to their frequency of occurrence, in parenthesis. Dithiol resistance is shown as the fraction of the FRET ratio remaining following washes with high concentrations of EDT, and corresponding standard deviations calculated from three or more duplicate wells on a 96-well plate.

After two more rounds of selection with higher stringency, single cells were sorted onto 96-well plates for clonal expansion (**Fig. 1.3b**, right). All twenty-two isolated clones converged on three sequences, HRW#KTF, FLN#MEP, and YRE#MWR. Each peptide was tested as both N and C terminal fusions to GFP (Emerald) and CFP (Cerulean[20]), and exhibited significantly improved dithiol resistance for both ReAsH and FlAsH (**Fig. 1.3c**), while demonstrating little preference for either biarsenical and confirming the weaker dithiol resistance of ReAsH as compared to FlAsH[1]. Upon ReAsH labeling, cells expressing YRE#MWR, but neither of the other two sequences, quickly formed tiny subcellular, highly red fluorescent aggregates capable of evading even the highest dithiol washes. This sequence was therefore excluded from further analysis, though the ability to precipitate a protein in living cells merely by addition of a permeant fluorogenic small molecule may be useful in other contexts.

In certain cell types, a small fraction of the tetracysteine-GFP protein becomes palmitoylated, blocking biarsenical binding and targeting the fusion protein to the plasma membrane. Fusion of a single FLAG, HA, or MYC epitope upstream of the tetracysteine or addition of the palmitoylation inhibitor 2-bromo-palmitate[21] blocked artifactual membrane localization without affecting dithiol resistance (**Fig. 1.5**). In several examples, including β-actin and α-tubulin, genetic fusion to a cellular protein is sufficient to block palmitoylation.

Increasing the FRET ratio of acceptor to donor emissions can be accomplished by altering the orientation of the GFP and ReAsH chromophores, as well as by increasing the fluorescence quantum yield of ReAsH directly. To address this, mammalian expressed tetracysteine-GFP protein was affinity purified to homogeneity from clonal lysates using FlAsH-agarose beads[1], then labeled with ReAsH *in vitro*. The selected peptides increased the quantum yield of ReAsH from 0.28 to 0.47 (**Table 1.1**). Both optimized sequences retain much of the improved fluorescence properties when transferred to the C-terminus of GFP. Replacement of the GFP with CFP permitted similar analysis of FlAsH, which likewise showed marked improvement in fluorescence quantum yield associated with the optimized tetracysteines.
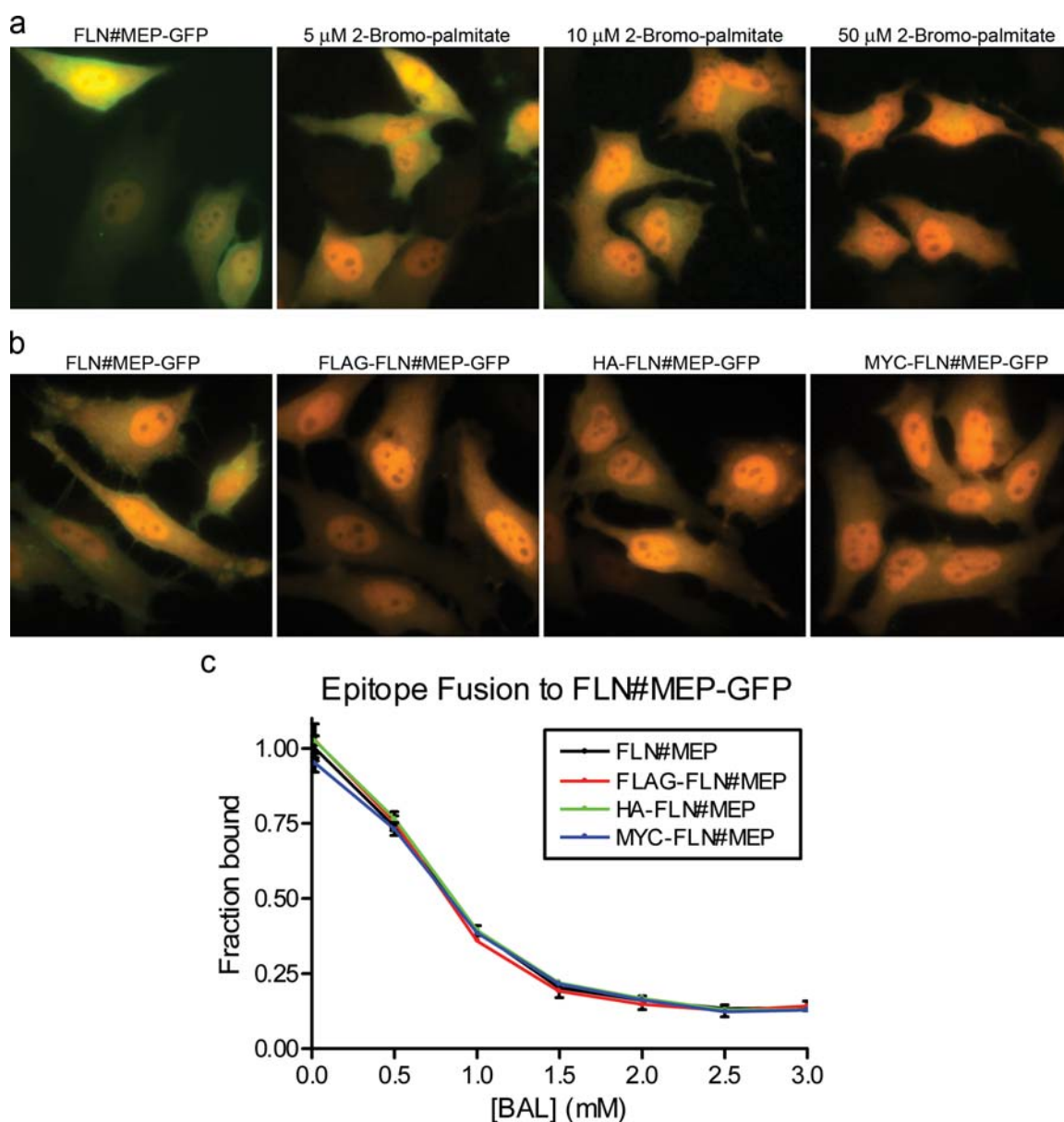
Figure 1.5. Inhibition of tetracysteine-specific membrane localization. (a) Inhibition of membrane localization by addition of 2-bromopalmitate. HeLa cells were transduced with FLN#MEP-GFP virus in media containing increasing concentrations of 2-bromopalmitate (Fluka). Two days later, cells were ReAsH stained and imaged on a Zeiss Axiovert 200M microscope with a cooled charge-coupled device camera (Roper Scientific), controlled by METAFLUOR 6.1 software (Universal Imaging). Imaging of the tetracysteine-GFP fusion was achieved by using a 480/30 nm excitation filter, 505 nm dichroic mirror, and two emission filters (535/45 nm for GFP and 653/95 for ReAsH). The two emission images were scaled and combined as individual channels for GFP (green) and GFP-mediated FRET to ReAsH (red). (b) Addition of various epitope tags N-terminal to the tetracysteine block palmitoylation and membrane localization in HeLa cells. HeLa cells were transduced with recombinant virus then two days later ReAsH stained and imaged. (c) Palmitoylation inhibition by epitope tag fusion does not affect the dithiol resistance of FLN#MEP performed as previously described. Similar results were observed for HRW#KTF.

Table 1.1. Quantum yields of FlAsH and ReAsH bound to optimized tetracysteine sequences fused to fluorescent proteins. Fluorescent protein quantum yields were unaltered by fusion to the tetracysteine. Furthermore, synthetic FLN#MEP-amide gave extinction coefficients and quantum yields for both bound FlAsH (70,000 $M^{-1}cm^{-1}$ and 0.85) and ReAsH (69,000 $M^{-1}cm^{-1}$ and 0.48) equivalent to or slightly higher than the genetically encoded fusions to GFP. The pK_a's of FlAsH (5.5) and ReAsH (4.7) bound to were little changed from earlier peptides[6]. Oxidation of the Met to the sulfoxide barely changed the extinction coefficients or quantum yields

	FlAsH Quantum Yield on CFP		ReAsH Quantum Yield on GFP	
Tetracysteine	N-terminus	C-terminus	N-terminus	C-terminus
αPG	nd	0.59	nd	0.28
MP#GS	0.72	0.50	0.30	0.18
HRW#KTF	0.65	0.61	0.42	0.34
FLN#MEP	0.78	0.70	0.47	0.42

Next, we tested whether the improved dithiol resistance and higher quantum yields could improve ReAsH contrast in cells. Populations of ReAsH labeled HEK293T cells expressing distinct tetracysteine-GFP fusions were analyzed by flow cytometry using dual laser excitation to monitor both GFP and ReAsH fluorescence. FLN#MEP cells showed improved dithiol resistance and enhanced ReAsH fluorescence when compared αPG cells (**Fig. 1.6a**, left). As expected, the contrast is linearly proportional to the amount of tetracysteine-GFP in the cell. The slope of the linear regression of contrast versus the tetracysteine-GFP concentration normalizes for expression, allowing direct comparison of the effectiveness of different staining conditions, dithiol washes, or tetracysteine sequences (**Fig. 1.6a**, right).
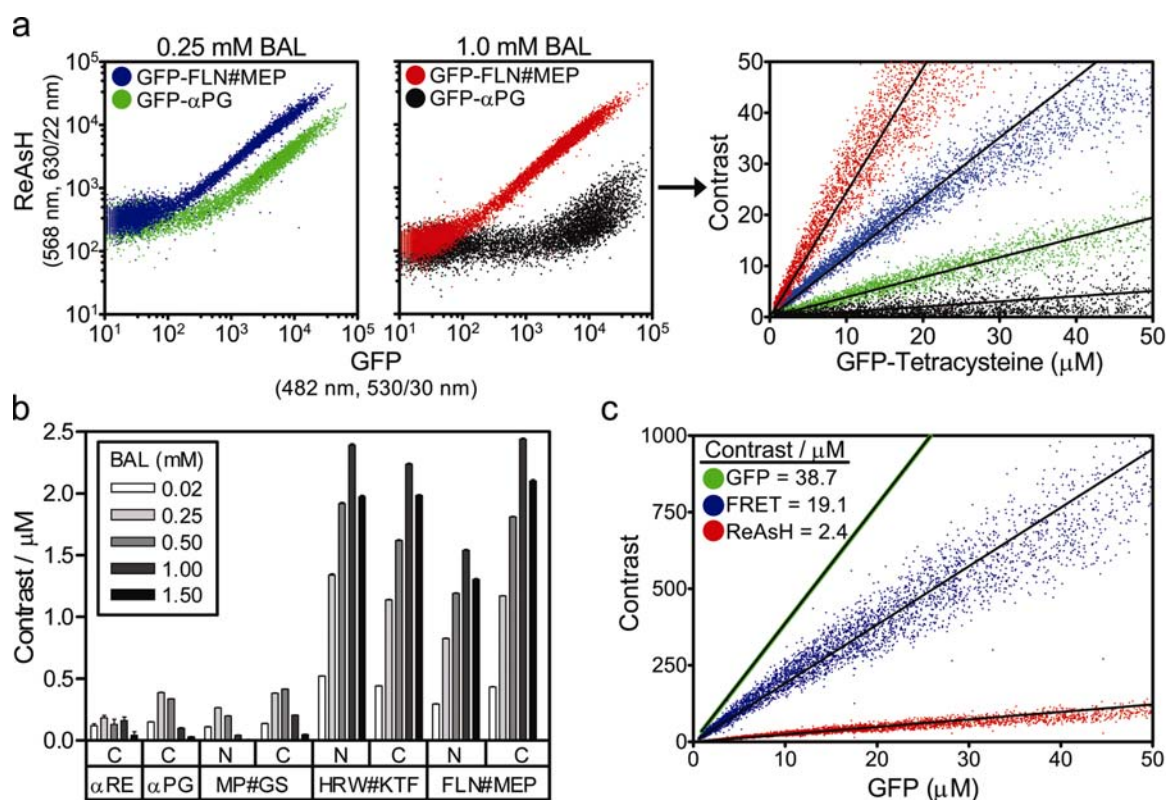
Figure 1.6. Contrast improvement quantitated by flow cytometry. (a) Comparison of ReAsH and GFP fluorescence in αPG (green/black) and FLN#MEP - expressing (blue/red) HEK293 cells following dithiol washes. Overlay of populations comparing ReAsH and GFP values following 0.25 mM BAL (left) or 1.0 mM BAL (middle) washes. ReAsH contrast in a single tetracysteine-GFP expressing cell is defined as specific ReAsH fluorescence divided by the mean non-specific ReAsH fluorescence determined from stained, non-transduced cells. Background ReAsH fluorescence drops significantly following the high BAL wash. The calculated contrast, comparing the GFP concentration to ReAsH contrast in single cells, is shown with corresponding color notation (right). Linear regression lines are shown for each population, demonstrating the steeper slope of FLN#MEP (red/blue) versus αPG (green/black). (b) Concentration independent analysis of tetracysteine contrast. Bars represent the slope of the linear regression and the corresponding standard error. N and C terminal fusions of the optimized tetracysteine sequences were compared to the α-helical sequences at five different concentrations of BAL. Moving the tetracysteine to the C-terminus of GFP had little effect on the overall contrast of HRW#KTF, and modestly improved the contrast of FLN#MEP. (c) Contrast of GFP (green), FRET from GFP to ReAsH (blue), and ReAsH (red). The slope of each linear regression is noted in the legend, corresponding to the contrast achieved per micromolar of recombinant protein. GFP contrast is calculated from GFP transduced HEK293T cells relative to non-transduced cells. FRET and ReAsH contrast is calculated from GFP-FLN#MEP transduced cells labeled with ReAsH and washed with 1.0 mM BAL.

The improved dithiol resistance and quantum yields of HRW#KTF and FLN#MEP

considerably increase the contrast of the tetracysteine-biarsenical complex in living cells

(**Fig. 1.6b**). These effects are most obvious after washing with high concentrations of

dithiol, which minimizes the non-specific background staining without affecting the desired specific fluorescence. The optimized tetracysteine peptides increase ReAsH contrast ~20-fold and >6-fold over the αRE and αPG sequences respectively (**Fig. 1.6**).

Despite the improvements described above, the absolute contrast of ReAsH under optimal conditions is still approximately 16-fold lower than that of GFP (**Fig. 1.6c**). Fusion of a tetracysteine tagged CFP (FlAsH or ReAsH) or GFP (ReAsH) to a cellular protein allows biarsenical fluorescence to be monitored by FRET from the donor fluorescent protein to the nearby, genetically fused biarsenical-tetracysteine acceptor. By excluding non-specific biarsenicals from being excited, FRET mediated detection increases ReAsH contrast almost 8-fold, making it only 2-fold less than GFP (**Fig. 1.6c**). Fusion of a tetracysteine barely increases the size of the fluorescent protein, but potentially provides the extended functionalities of the biarsenical-tetracysteine system, such as fluorescent pulse-chases[4,5], CALI[7,8], and EM photoconversion[4].

As a demonstration of fluorescence based pulse chases using a tetracysteine-GFP fusions, Cx43-GFP-FLN#MEP was transfected into HeLa cells and stained with ReAsH. Five hours later, new protein was visualized by the presence of GFP labeling and the absence of ReAsH staining (**Fig. 1.7**). New protein can be visualized by either looking for GFP fluorescence and the absence of ReAsH fluorescence (**Fig. 1.7a**), or more precisely by taking the ratio of GFP-sensitized ReAsH FRET emission to GFP fluorescence (**Fig. 1.7b**). Observing the ratio is more accurate, since not all GFP fluorescence is removed upon ReAsH binding and quenching. In this example, the cells in upper right form new gap junctions, while the cells on the bottom left have a pre-existing gap junction that is turning over older protein from the middle and accepting new protein from the peripheral membranes.
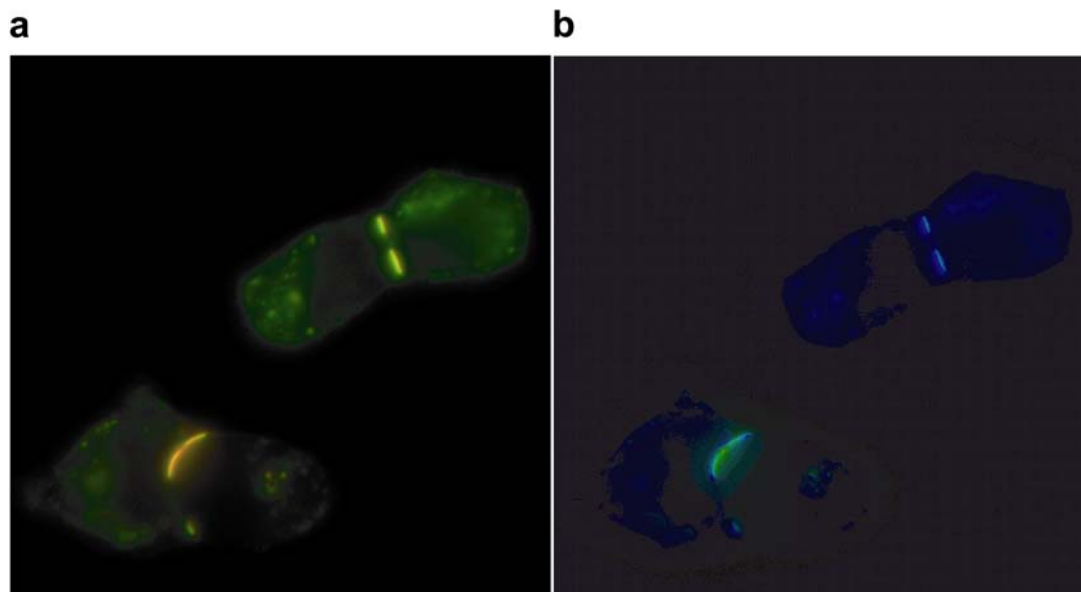
Figure 1.7. Tetracysteine-GFP based fluorescence pulse chase. (a) Overlay of GFP-ReAsH FRET (480/30 nm, 653/95 nm) and GFP (480/30 nm, 535/25 nm) 5 hours post-ReAsH staining and washing with 0.25 mM BAL. (b) Ratiometric analysis of new protein. The displayed image shows a ratio range of 0.5 – 2.5. Further timepoints show removal of ReAsH fluorescence, but the protein half-life is greater than 4 hours, as reported for Cx43-αRE[4].

To demonstrate EM photoconversion by FRET, Cx43-GFP-MP#GS was transfected into HeLa cells and stained with ReAsH and photo-oxidized for EM. Prior to fixation, both GFP and ReAsH fluorescence show large gap junctions shared between two cells. The resulting correlated EM images have high contrast attained by avoiding excitation of background ReAsH during the photo-oxidation procedure. Using the FLN#MEP tetracysteine tag, FRET-mediated photoconversion has also been used for EM detection of recombinant protein in several other systems, including recent work tracking golgi dynamics during cell division.
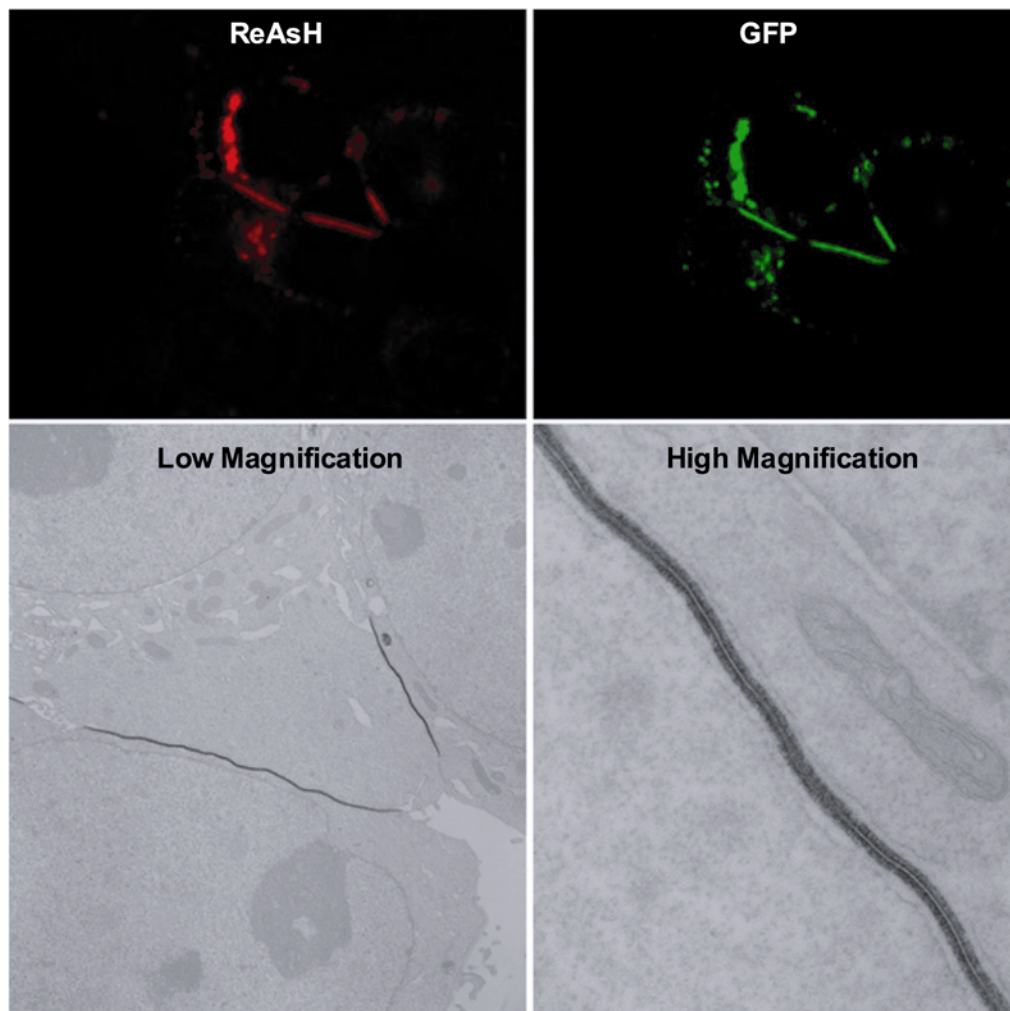
Figure 1.8. FRET-mediated Photoconversion of Cx43-GFP-tetracysteine. Correlated fluorescence and EM of FRET mediated photoconversion of Cx43-GFP-MP#GS at both low and high magnification.

To test whether the optimized tetracysteines retain greater contrast when fused to biologically relevant proteins, not just GFP, we transduced primary human foreskin fibroblasts for stable expression of MP#GS, FLN#MEP, HRW#KTF as N-terminal tetracysteine-GFP fusions to β-actin. After ReAsH labeling and washing with a high concentration of dithiol, actin stress fibers were easily identified by GFP fluorescence, but ReAsH fluorescence was only visible with FLN#MEP and HRW#KTF, not with MP#GS (**Fig. 1.9a**). Next, the optimized sequences were directly fused to β-actin without

an intervening GFP. Tetracysteine expressing cells were first labeled with ReAsH, then

washed with a high concentration of dithiol. Subsequently, cells were labeled with FlAsH

to fill all tetracysteine vacancies caused by the high dithiol wash and analyzed for both

FlAsH and ReAsH fluorescence. FLN#MEP and HRW#KTF were capable of resisting

high dithiol washes and retaining ReAsH fluorescence, while ReAsH labeling of MP#GS

was completely replaced by FlAsH (**Fig. 1.9b**).
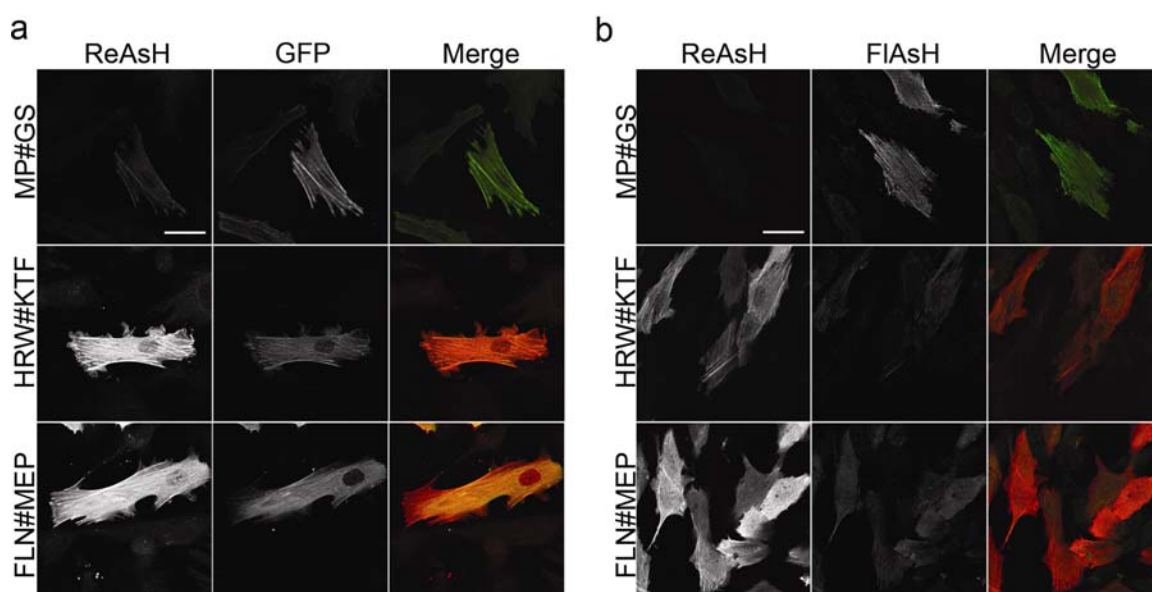
Figure 1.9. Fusion of optimized tetracysteines to β-actin. (a) Tetracysteine-GFP-β-actin fusions. Confocal images comparing ReAsH labeled tetracysteine-GFP-actin fusions in human primary fibroblasts following a 0.75 mM BAL wash. Significant GFP fluorescence is expected even in the absence of dithiol washing, since the FRET efficiency is less than 100%. (b) Tetracysteine-β-actin fusions. Cells were labeled with ReAsH, then washed with 0.75 mM BAL, and relabeled with FlAsH to bind free sites, then weakly washed with 0.1 mM BAL to reduce FlAsH background. The length of the experiment was concluded to be too fast to detect significant new protein expression, as determined by ReAsH/FlAsH pulse-chase time course experiments or addition of cycloheximide (data not shown). Scale bar: 40 μm.

In addition to fluorescence, fusion of an optimized tetracysteine to β-actin

provides another successful example of correlated fluorescence (**Fig. 1.10a**) and

electron microscopy by ReAsH-mediated EM photoconversion (**Fig. 1.10b**). Clearly,

addition of FLN#MEP or HWR#KTF as the flanking residues of CCPGCC radically

increases the dithiol resistance of the tetracysteine-biarsenical complex, providing better

contrast in combination with high concentration dithiol washes, independent of fusion to GFP.
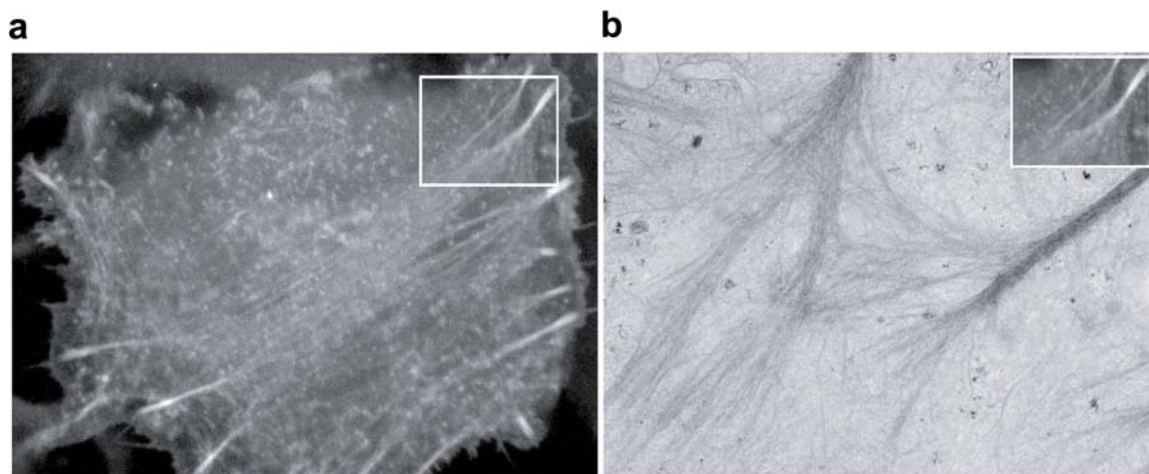
Figure 1.10. Correlated fluorescence and EM of tetracysteine tagged β-actin.

We currently do not have a complete molecular explanation of the higher affinity of the final improved tetracysteines, HRW#KTF and FLN#MEP. Alanine scanning suggests that large aromatic residues are valuable (**Fig. 1.11)**, possibly shielding the biarsenical-tetracysteine complex from competing dithiols and other quenchers. Charge provides no simple explanation, because HRW#KTF has a charge of +2 to +3 while FLN#MEP has a -1 charge, and the key residues contributing to the high dithiol resistance are neutral (**Fig. 1.11**). In both sequences, hydrophobic aromatic residues such as Trp and Phe were found to be essential for improvements in both affinity and brightness. These residues alone are insufficient to explain the high dithiol resistance phenotype, since each of the sequences isolated in the high FRET ratio population show a similar Phe at the first or last variable position or a Trp at the third variable position (**Fig. 1.4**), yet all are still less than optimal. HRW#KTF contains both a Trp and Phe,

which in combination may lead to the observed superior dithiol resistance. Equally

important in FLN#MEP is the Asn, which only contributes to affinity, not brightness.

Overall, the Ala scan demonstrates that each residue except the Leu contributes at least

partially to the enhanced dithiol resistance of FLN#MEP. In FLN#MEP, E11A increases

the fluorescence quantum yield another 16% to 0.54, with little loss of affinity. As for

HRW#KTF, all residues contribute except the His and Arg, whose replacement by Ala

slightly increases the dithiol resistance without affecting brightness. Simple point

mutations demonstrate that further improvements are possible, and the selected

sequences are not the final optimization. These sequences may be amenable to further

optimization, although the benefits of such optimizations are likely marginal. Stacking of

a tryptophan indole ring is known to quench fluorescein bound to anti-fluorescein

antibodies[22], so it is surprising that tryptophan immediately adjacent to a cysteine pair in

HRW#KTF is compatible with, let alone contributory towards high quantum yield for

ReAsH. These residues may have further roles in modulating the expression or

fluorescent properties of the biarsenical-tetracysteine complex. Further understanding of

the affinity and photophysical properties of the complexes will likely require high-

resolution structures. Meanwhile we prefer FLN#MEP over HRW#KTF, mainly because

the former gives somewhat higher quantum yields.
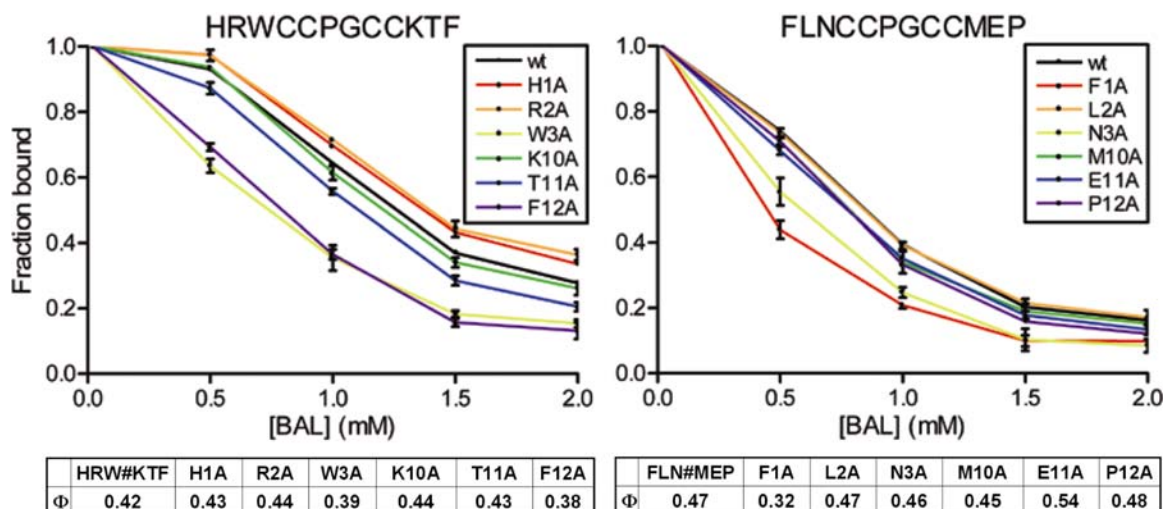
Figure 1.11. Dithiol resistance of alanine mutants point to key residues. Alanine mutants of HRW#KTF-GFP (left) and FLN#MEP-GFP (right) were transduced into HEK293T cells and analyzed for dithiol resistance on a plate reader. All data points differ from wild-type to statistical significance (P < 0.05 by a two-tailed t-test) except FLN#MEP L2A at all concentrations of BAL tested, FLN#MEP M4A at 1.5 mM and 2.0 mM BAL, and HRW#KTF K4A at 1.0 mM and 1.5mM BAL. Data shown normalized to **Fig. 1.2** titrations. Fluorescence quantum yields (Φ) of alanine point mutations are listed.

Most attempts to use FACS to screen libraries have been performed in bacteria[23-25], yet screening in mammalian cells is the best guarantee that the resulting optimizations will function properly in mammalian cells. Additionally, many genetically encoded reporters detect mammalian specific biochemistry, making bacterial optimization unfeasible. The potential to screen large mammalian cell libraries will be a crucial tool in the further improvement of genetically encoded fluorescent reporters.

Materials and Methods

RRL1 production

Oligonucleotide primers sequences are listed in **Table 1.2**. Fifty micromolar of 5'

phosphorylated degenerate oligonucleotide primers (Primer1, Primer2) were annealed in

1x T4 DNA Ligase Buffer (NEB). The retrovirus vector pCLNCX (Imgenex) was modified

to destroy the BamHI site upstream of the second CMV promoter. EGFP was amplified

by PCR (Primer3; Primer4) to introduce cloning sites and a short linker, then ligated into

the modified pCLNCX. Vector DNA was digested with BamHI and NotI twice, then

ligated with an optimized concentration of the annealed library overnight. The ligation

reaction was purified and concentrated using a QIA-quick column (Qiagen), then

electroporated into TG1 cells (Stratagene). Based on serial dilutions of the

electroporated cells, the library contained 2.6×10^7 members, or 2.4% of the total

nucleotide diversity. After overnight growth in 250 milliliter (ml) LB-ampicillin, the plasmid

library was purified and co-transfected with pCL-Eco (Imgenex) using calcium-phosphate

onto a 10 cm dish of 80% confluent HEK293 cells. The medium was replaced 24 hours

later, and 48 hours later 8 ml of virus was harvested, filtered, and frozen in liquid

nitrogen. 6.5×10^6 NIH3T3 cells were infected with 1 ml of 1.2×10^6 Green Fluorescent

Units (GFU) / ml virus in the presence of 8 μg / ml polybrene (Sigma).

RRL2 production

Fifty micromolar of both degenerate oligonucleotide primers (Primer5, Primer6), each

containing 18 bp complementarity over the region encoding for the CCPGCC motif, were

annealed. Overhangs were filled in with Klenow Fragment ($3' \rightarrow 5'$ exo$^-$) (NEB). After

heat-inactivation, DNA was extracted with phenol/chloroform and precipitated with

ethanol. The product was digested with BamHI and NotI (NEB), extracted with

phenol/chloroform and concentrated in a Microcon YM-10 column (Millipore), then separated by non-denaturing PAGE. The digested fragment was excised and electro-eluted (Bio-Rad Model 422 Electro-Eluter, 12-15kD cutoff), extracted with phenol/chloroform, and concentrated. In RRL2, NIH3T3 cells were replaced with the smaller, more robust and higher expressing HEK293 cells and the WPRE[26] was added to the 3' untranslated region of the enhanced folding mutant, Emerald GFP[27]. These improvements increased the expression of the tetracysteine-GFP library, amplifying the mean signal from a single cell by >10-fold. The WPRE element was amplified by PCR from BluescriptII SK+ WPRE (Primer7, Primer8) and cloned between the XhoI sites and ClaI sites of pCLNCX. Emerald GFP was amplified by PCR (Primer9, Primer10) to introduce restriction sites into the N-terminus of GFP, digested with HindIII and XhoI, and ligated into the pCLNCX-WPRE vector. Next, the GFP vector was digested with BamHI and NotI, extracted with phenol/chloroform, precipitated with ethanol. An optimized concentration of cassette was ligated with 15 μg of vector, extracted with phenol/chloroform, and precipitated with ethanol in the presence of 20 μg yeast tRNA (Ambion). The ligation was electroporated into Electro-Ten Blue cells (Stratagene), and grown overnight in 1 liter (l) LB-Ampicillin. The RRL2 plasmid library was calculated to contain 8.6×10^8 members, or 80% of the total nucleotide diversity. Purified plasmid was co-transfected with an equal concentration of pCL-Ampho (Imgenex) into four 10 cm plates of 80% confluent HEK293 cells with Fugene6 (Roche Diagnostics). Following virus production and storage at $-80°$ C, thawed virus titered by flow cytometry contained 2.5×10^6 GFU / ml in HEK293 cells. Three hundred million HEK293 cells were infected at an MOI of 0.14.

ReAsH staining and flow cytometry

Cells were cultured in DMEM supplemented with 10% FBS, 100 units / ml penicillin G,

100 μg / ml streptomycin, and 2.5 μg / ml amphotericin B. Confluent cells were stained

with 0.25 μM or 0.5 μM ReAsH / 10 μM EDT in HBSS (Hanks Balanced Saline Solution

supplemented with 2 g / l glucose and 20 mM Hepes) for one hour, rinsed once with

HBSS, then incubated with specific concentrations of dithiol (EDT or BAL) in HBSS for

30 minutes at room temperature. Following trypsinization, cells were pelleted and

resuspended in HBSS and sorted into 15 ml tubes containing 5 ml media containing

30% FBS. Sorted cells were pelleted and cultured in media containing 0.5 mg / ml G418.

Single cells were collected in 96-well plates with 0.3 ml media per well supplemented

with 10% Optimem and an additional 10% FBS. RRL1 was sorted on a MoFlow flow

cytometer using a single 488 nm laser and 530/30 nm and 630/22 nm emissions. RRL2

was performed on a BD FACSVantage DiVa with two lasers, 482 nm (530/30 nm and

615/40 nm emission) and 568 nm (630/22 nm emission). Laser alignment changes on

successive rounds of selection were normalized using readings of alignment beads.

Sequence retrieval and mutagenesis

Total RNA was isolated using Trizol Reagent (Invitrogen) and reverse transcribed with

ImPromII reverse transcriptase (Promega) using a gene specific primer in the 3'UTR of

pCLNCX (Primer11), then amplified by PCR (Primer11, Primer12). The PCR product

was gel purified and sequenced with either a 5' or 3' primer pCLNCX primer. All unique

PCR products from RRL1 were digested with HindIII and NotI and sub-cloned into

pCDNA3 (Invitrogen). MP#GC was mutated to MP#GS by PCR (Primer7, Primer13).

The RT-PCR products of several unique clones from RRL2 were digested with HindIII

and XhoI for ligation into the pCLNCX-WPRE vector. N-terminal YRE#MWR-GFP was

subcloned onto the C-terminus of the GFP from RRL1 using BamHI and XhoI sites. Next, the tetracysteine-GFP fusion was amplified by PCR to introduce a stop codon before the NotI site (Primer12, Primer14), digested, and ligated into pCLNCX WPRE. Other tetracysteine clones were then substituted by sub-cloning. C-terminal MP#GS was amplified by PCR to remove the stop codon (Primer7, Primer15), then ligated onto the N-terminus of RRL2 GFP with HindIII/NotI sites, and finally subcloned to RRL2 vector using BamHI/XhoI sites. Alanine mutations of the first three residues of HRW#KTF and FLN#MEP were induced by Quikchange [(Primer29; Primer30), (Primer31; Primer32), (Primer33; Primer34), (Primer35; Primer36), (Primer37; Primer38), (Primer39; Primer40)] as N-terminal fusions to emerald GFP in pCLHCX. Alanine mutations of the final three residues were introduced by PCR (Primer41; Primer42, Primer43, Primer44, Primer45, Primer46, Primer47). Epitope fusions were generated by PCR and cloned into pCLNCX-WPRE to generate the following fusion proteins: (FLAG) MDYKDDDDKGS-FLN#MEP-GFP (Primer20, Primer10), (HA) MVYPYDVPDYAGS-FLN#MEP-GFP (Primer21, Primer10), (MYC) MVQKLISEEDLGS-FLN#MEP-GFP (Primer22, Primer10). Virus was generated and cell lines were selected and grown to confluency for plate reader dithiol titrations and for protein purification using FlAsH affinity beads.

Dithiol titrations

RRL1 unique clones were analyzed for dithiol resistance in transiently transfected HeLa cells by fluorescence microscopy, using the following filters: GFP (excitation 480/30 nm, emission 535/25 nm), FRET (excitation 480/30 nm, emission 635/55 nm), ReAsH (excitation 540/25 nm, emission 635/55 nm), dichroic 505LP. Maximal ReAsH fluorescence was reached in approximately 30 to 45 minutes, as determined by a lack of further GFP fluorescence quench. EDT or BAL were dissolved in DMSO and premixed

with HBSS, then added in increasing concentrations to cells at intervals of 15-20 minutes. The average ReAsH fluorescence in non-transfected cells was subtracted from tetracysteine-GFP expressing cells. RRL2 clonal cell lines were plated in multiple wells on 96-well plates with black walls and clear bottoms and grown to confluency. Fluorescence measurements were taken on a Tecan Sapphire fluorescence plate reader at each fluorophore's specific wavelengths. Cells were measured prior to staining to obtain initial baseline fluorescence, and then stained with ReAsH for 1 hour. The staining solution was replaced with HBSS with 10 μM EDT to measure the maximal fluorescence. Next, various concentrations of dithiol in HBSS were added to predetermined wells and incubated for 30 minutes before the final measurement. Background fluorescence was determined using non-transduced HEK293 cells and subtracted from each individual wavelength.

Multiple Tetracysteines

Multiple tandem tetracysteines were added serially to the C-terminus of GFP using BamHI-BglII fusions. First, the CFP was amplified with a short linker (Primer23; Primer24) and cloned into pRSETb. Next, the MP#GS sequence was cloned to the C-terminus with a 5'BamHI site and a 3'BglII site (Primer25; Primer26). Vector was digested with either BamHI/ScaI or BglII/ScaI, and then fragments from each digest containing the tetracysteine were ligated, leading to the formation of two tandem tetracysteines. These tetracysteines were then excised and transferred to target proteins using BamHI/ScaI. This method ensures directional serial addition of tandem tetracysteines and is rapidly additive, since once two tandems tetracysteines are produced, the same vector can be digested separately with BamHI/ScaI and BglII/ScaI and re-ligated to make 4 tandem tetracysteines. These fusions were also transferred to

GFP in pRSETb, and later pCDNA3 by KpnI/EcoRI digestion and ligation. GFP tetracysteines were transferred to Cx43 by first amplifying Cx43 to remove the internal HindIII site and add a KpnI site for cloning (Primer27; Primer28) in pCDNA3 and pCLNCX-WPRE.

Protein purification and fluorescence measurements

For multiple tetracysteines, recombinant protein was expressed using IPTG induced JM109 bacterial cultures using the pRSETb vector. Bacteria were lysed using BPERII (Pierce) bacterial lysis reagent supplemented with 10 mM MES (Sodium 2-mercaptoethanesulfonate, Aldrich), 1 mM TCEP (tris-(2-carboxyethyl)phosphine HCl, Molecular Probes), and Complete protease inhibitor (Roche Diagnostics). All solutions used for all protein purification and *in vitro* characterization experiments contained 10 mM MES and 1 mM TCEP. Three equivalents of FlAsH-EDT were added simultaneously during cell lysis immediately bind free tetracysteines to prevent oxidation. If FlAsH was not bound during lysis, significant oxidation occurred that could not be reversed by TCEP and MES treatment. Free FlAsH was removed by dialysis in PBS and analyzed for purity by SDS-PAGE in the absence of DTT. For optimized tetracysteines, a confluent monolayer of clonal tetracysteine tagged GFP or CFP HEK293 cells was lysed in RIPA buffer (1% NP-40, 1% sodium deoxycholate, 0.1% SDS, 0.15 M NaCl, 0.1 M sodium phosphate pH 7.2, 2 mM EDTA). Cell debris was pelleted and the soluble cell lysate was filtered through 0.45 μm syringe filter, then diluted 1:2 in Phosphate Buffered Saline (PBS). Pre-washed FlAsH- beads were added to the lysate and incubated at 4° C overnight, then pelleted and washed 3 times with PBS, then once in PBS supplemented with 0.1 mM EDT or BAL. Protein was eluted with 0.25 M DTT (dithiothreitol, Invitrogen), then buffer exchanged four times in PBS by centrifugation in a 30 kD cutoff Microcon

filter. Next, protein was labeled with 3-5 fold excess biarsenical overnight at 4° C, then buffer exchanged 4 times with 0.1 mM BAL in PBS to remove free biarsenical. No other bands were detectable by SDS-PAGE determined by either direct ReAsH fluorescence or coomassie blue staining. Mass spectrometry analysis of purified, trypsin digested protein revealed a loss of expected mass observed in N-terminal most tetracysteine containing peptide, corresponding to N-terminal methionine cleavage and N-terminal acetylation. FlAsH and ReAsH quantum yields were determined using dichlorofluorescein in 0.1 N NaOH (Φ = 0.92) or Rhodamine 101 in ethanol (Φ = 1.0, then corrected for refractive index) respectively as standards.

Determining ReAsH contrast using flow cytometry

HEK293T cells (ATCC) were transduced with retrovirus encoding a tetracysteine-GFP fusion and stained two days later with 0.5 μM ReAsH / 10 μM EDT for one hour, then washed 1x with Dulbecco's PBS / 1 mM EDTA and trypsinized. Suspended cells were aliquoted into various concentrations of BAL, incubated for 30 minutes, then pelleted and resuspended in HBSS supplemented with 20 μM BAL to prevent further ReAsH binding, and analyzed on a BD FACSVantage DiVa flow cytometer. One aliquot was washed with 10 mM BAL to remove all ReAsH to measure the maximal GFP fluorescence of the population, and hence the average GFP quench at the tested BAL concentrations. Lower staining concentrations of ReAsH slightly enhanced contrast, yet resulted in a lower fraction of labeled tetracysteine, and subsequently less signal. Inclusion of Disperse Blue 3[2] or long overnight labeling procedures did not enhance ReAsH contrast. GFP contrast was measured in the absence of ReAsH labeling to avoid background contributions from non-specific background ReAsH labeling. Quantum FITC MESF High Level beads (Bangs Labs) were analyzed and peak averages were plotted against the

fluorophore number to produce a standard curve. HEK293T cell volume was determined

by measuring the average cell diameter of suspended cells by DIC microscopy. Data

was calculated using the following formulae:

Φ = Fluorescence Quantum Yield, ε = Extinction coefficient, A = Absorbance

Laser wavelength = 482 nm; Peak $A_{\text{Emerald GFP}}$ = 484 nm, Peak $A_{\text{Fluorescein}}$ = 490 nm

V = Cell Volume = 0.9 ± 0.2 picoliters

B = Brightness = $(\Phi_{\text{GFP}} * \varepsilon_{\text{GFP}} * A_{\text{GFP:482nm}}/A_{\text{GFP:484nm}})/(\Phi_{\text{Fluorescein}} * \varepsilon_{\text{Fluorescein}} *$

$A_{\text{Fluorescein:482nm}}/A_{\text{Fluorescein:490nm}})$

Q = Quench = (Mean GFP fluorescence at given [BAL]) / (Mean GFP fluorescence at 10

mM BAL)

From standard curve, K = moles Fluorescein / relative fluorescent units

GFP_{cell} = GFP fluorescence signal from a single event, gated against both forward and

side scatter

[GFP] = GFP_{cell} * K * B / (Q * V)

T = Transduced = GFP positive cell, NT = Non-transduced HEK293T cell

ReAsH Contrast = $(\text{ReAsH}_T - \text{Mean ReAsH}_{NT})$ / Mean ReAsH_{NT}

GFP Contrast = $(\text{GFP}_T - \text{Mean GFP}_{NT})$ / Mean GFP_{NT}

Contrast was plotted against GFP concentration and slopes were calculated from linear

regression analysis using $1/X^2$ weighting to minimize percentage errors.

Production and analysis of β-actin fusions

HEK293 total RNA was isolated and reverse transcribed using an oligo d(T) primer to

generate cDNA. Human β-actin was amplified by PCR (Primer16, Primer17) using Taq

polymerase (Roche), digested with HindIII/XhoI, and ligated in the N-terminal FLN#MEP

pCLNCX-WPRE vector. The resulting plasmid was used as a PCR template to generate

an in-frame fusion of actin (Primer18, Primer17) and FLN#MEP GFP pCLNCX-WPRE (Primer12, Primer19). The vector, and the two PCR products were digested (pCLNCX; HindIII/XhoI, 4cys-GFP; HindIII/BglII, Actin; BglII/XhoI) and ligated. The final actin or GFP-actin fusions were subcloned into N-terminal HRW#KTF and MP#GS pCLNCX-WPRE vectors. Human primary fibroblasts were transduced with tetracysteine β-actin and tetracysteine-GFP-β-actin fusions and maintained without drug selection to preserve heterogeneous expression levels. Cells were grown and imaged on glass-bottom poly-D-lysine coated dishes (MatTek Corp.). All labeling and imaging was performed at 37° C. Cells were labeled with 0.5 μM ReAsH / 10 μM EDT in DMEM for one hour, then rinsed and incubated with 0.75 mM BAL for 30 minutes. Tetracysteine-actin fusions were further labeled with 0.5 μM FlAsH / 10 μM EDT for one hour, then washed with 0.1 mM BAL. Next, the washing media was rinsed away several times and replaced with media lacking phenol red. Imaging was performed on a BioRad1024 MRC confocal system and maintained at 37° C using a Bioptech lens heater and a 3 cm water-circulation-based dish heater. Cells were imaged in a single large focal plane positioned at the cell base through a 63x oil objective at 1024x1024 resolution with the following settings: ReAsH: excitation 568 nm, emission 585 nm LP; FlAsH: excitation 488 nm, emission 540/30 nm; GFP: excitation 488 nm, emission 522/35 nm). Following the high concentration dithiol washes required for optimal biarsenical-tetracysteine contrast, primary human foreskin fibroblast proliferation slows down for a while, dividing only once in 72 hours compared to two divisions for untreated cells. Overall, the cells retain high viability and actin dynamics are unaffected and remain observable for several days after labeling. Photoconversion was performed according to published protocols[4].

Table 1.2. Primer sequences.

Primer	Sequence
1	GATCCNNKNNKTGCTGCNNKNNKTGCTGCNNKNNKTAAGC (5'Phosph)
2	GGCCGCTTAMNNMNNGCAGCAMNNMNNGCAGCAMNNMNNG (5'Phosph)
3	ATTTTATAGCGGCCGCTTAGGATCCACTGCTTTCCTTGGTACAGCTCGTCCATGCC GAGAG
4	CCCAAGCTTGCCACCATGGTGAGCAAGGGCGAGGAG
5	TTTGGATCCNNKNNKNNKTGCTGCCCCGGCTGCTGC
6	TTTTGCGGCCGCCMNNMNNMNNGCAGCAGCCGGGGCAGCA
7	AATACGACTCACTATAGGGA
8	GTAATCCAGAGGTTGATTCTCGAGAAAA
9	TTTTAAGCTTGCCACCATGGCCGGATCCTAAGCGGCCGCAGCAAGGGCGAGGAG CTG
10	CCCCATCGATCTCGAGTTACTTGTACAGCTCGTCCAT
11	ACCTACAGGTGGGGTCTTTCATTCCC
12	AGCTCGTTTAGTGAACCGTCAGATC
13	GACAAGCGGCCGCTTAAGAACCGC
14	AAACTCGAGTTAGCGGCCGCCCCTCCACATGCAG
15	AAAGCGGCCGCCAGAACCGCAGCACCCGGGGCA
16	TTTGCGGCCGCATGGATGATGATATCGCCGCG
17	TTTCTCGAGCTAGAAGCATTTGCGGTGGAC
18	CTCAGATCTCGGGCTATGGATGATGATATCGCCGC
19	TCGAGATCTGAGTCCGGACTTGTACAGCTCGTCCATG
20	TTTAAGCTTGCCACCATGGATTACAAGGATGACGACGATAAGGGATCCGCCGGAT CCTTTTTGAATTG

Table 1.2. Continued.

21	TTTAAGCTTGCCACCATGGTGTACCCCTACGACGTGCCCGACTACGCCGGATCCG CCGGATCCTTTTTGAATTG
22	TTTAAGCTTGCCACCATGGTGCAGAAGCTGATCTCAGAGGAGGACCTGGGATCCG CCGGATCCTTTTTGAATTG
23	GGGGAGATCTGGGTACCGCCACCATGGTGAGCAAGGGC
24	CTTAGGATCCGGAGCTCTCCTTGTACAGCTCGTC
25	5'PHOS-GATCAATGCCTTGCTGCCCCGGGTGCTGCGGCTCCG
26	5'PHOS-GAACCGGAGCCGCAGCACCCGGGGCAGCAAGGCATT
27	GGGAAGCTTGCCACCATGGGTGACTGGAGTGCCTTGGGGAAATTACTGG ACAAGG
28	AAAAGGTACCGACCGGTTGAACCGCAATCTCCAGGTCATCAG
29	ACCATGGCCGGATCCGCTCGGTGGTGCTGCCC
30	GGGCAGCACCACCGAGCGGATCCGGCCATGGT
31	ATGGCCGGATCCCATGCGTGGTGCTGCCCCGG
32	CCGGGGCAGCACCACGCATGGGATCCGGCCAT
33	GCCGGATCCCATCGGGCGTGCTGCCCCGGCTG
34	CAGCCGGGGCAGCACGCCCGATGGGATCCGGC
35	ACCATGGCCGGATCCGCTTTGAATTGCTGCCC
36	GGGCAGCAATTCAAAGCGGATCCGGCCATGGT
37	ATGGCCGGATCCTTTGCGAATTGCTGCCCCGG
38	CCGGGGCAGCAATTCGCAAAGGATCCGCCAT
39	GCCGGATCCTTTTTGGCTTGCTGCCCCGGCTG
40	CAGCCGGGGCAGCAAGCCAAAAAGGATCCGGC
41	GGGGCATATGCCAAGTACGCCCCCTATT
42	CGGCCGCCAAAAGCTGCGCAGCAGCCGGGGC
43	CTGCGGCCGCCAAAAGCCTTGCAGCAGCCGG
44	TTGCTGCGGCCGCCAGCAGTCTTGCAGCAGCC
45	CGGCCGCCCGGCTCCGCGCAGCAGCCGGGGCA
46	CTGCGGCCGCCCGGCGCCATGCAGCAGCCGG
47	TGCTGCGGCCGCCCGCCTCCATGCAGCAGCC

<u>References</u>

1. Adams,S.R., Campbell,R.E., Gross,L.A., Martin,B.R., Walkup,G.K., Yao,Y., Llopis,J., & Tsien,R.Y. 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 (2002).

2. Griffin,B.A., Adams,S.R., Jones,J., & Tsien,R.Y. Fluorescent labeling of recombinant proteins in living cells with FlAsH. *Methods Enzymol.* **327**, 565-578 (2000).

3. Stroffekova,K., Proenza,C., & Beam,K.G. The protein-labeling reagent FLASH-EDT2 binds not only to CCXXCC motifs but also non-specifically to endogenous cysteine-rich proteins. *Pflugers Archiv-European Journal of Physiology* **442**, 859-866 (2001).

4. Gaietta,G., Deerinck,T.J., Adams,S.R., Bouwer,J., Tour,O., Laird,D.W., Sosinsky,G.E., Tsien,R.Y., & Ellisman,M.H. Multicolor and electron microscopic imaging of connexin trafficking. *Science* **296**, 503-507 (2002).

5. Ju,W., Morishita,W., Tsui,J., Gaietta,G., Deerinck,T.J., Adams,S.R., Garner,C.C., Tsien,R.Y., Ellisman,M.H., & Malenka,R.C. Activity-dependent regulation of dendritic synthesis and trafficking of AMPA receptors. *Nature Neuroscience* **7**, 244-253 (2004).

6. Poskanzer,K.E., Marek,K.W., Sweeney,S.T., & Davis,G.W. Synaptotagmin I is necessary for compensatory synaptic vesicle endocytosis in vivo. *Nature* **426**, 559-563 (2003).

7. Tour,O., Meijer,R.M., Zacharias,D.A., Adams,S.R., & Tsien,R.Y. Genetically targeted chromophore-assisted light inactivation. *Nature Biotechnology* **21**, 1505-1508 (2003).

8. Marek,K.W. & Davis,G.W. Transgenically encoded protein photoinactivation (FlAsH-FALI): Acute inactivation of synaptotagmin I. *Neuron* **36**, 805-813 (2002).

9. Panchal,R.G., Ruthel,G., Kenny,T.A., Kallstrom,G.H., Lane,D., Badie,S.S., Li,L.M., Bavari,S., & Aman,M.J. In vivo oligomerization and raft localization of Ebola virus protein VP40 during vesicular budding. *Proceedings of the National Academy of Sciences of the United States of America* **100**, 15936-15941 (2003).

10. Andresen,M., Schmitz-Salue,R., & Jakobs,S. Short tetracysteine tags to beta-tubulin demonstrate the significance of small labels for live cell imaging. *Molecular Biology of the Cell* **15**, 5616-5622 (2004).

11. Enninga,J., Mounier,J., Sansonetti,P., & Van Nhieu,G.T. Secretion of type III effectors into host cells in real time. *Nature Methods* **2**, 959-965 (2005).

12. Robia,S.L., Flohr,N.C., & Thomas,D.D. Phospholamban pentamer quaternary conformation determined by in-gel fluorescence anisotropy. *Biochemistry* **44**, 4302-4311 (2005).

13. Hoffmann,C., Gaietta,G., Bunemann,M., Adams,S.R., Oberdorff-Maass,S., Behr,B., Vilardaga,J.P., Tsien,R.Y., Eisman,M.H., & Lohse,M.J. A FlAsH-based FRET approach to determine G protein - coupled receptor activation in living cells. *Nature Methods* **2**, 171-176 (2005).

14. Rice,M.C., Bruner,M., Czymmek,K., & Kmiec,E.B. In vitro and in vivo nucleotide exchange directed by chimeric RNA/DNA oligonucleotides in Saccharomyces cerevisae. *Molecular Microbiology* **40**, 857-868 (2001).

15. Rice,M.C., Czymmek,K., & Kmiec,E.B. The potential of nucleic acid repair in functional genomics. *Nature Biotechnology* **19**, 321-326 (2001).

16. Tsien,R.Y. The green fluorescent protein. *Annu. Rev. Biochem.* **67**, 509-544 (1998).

17. Chen,I. & Ting,A.Y. Site-specific labeling of proteins with small molecules in live cells. *Current Opinion in Biotechnology* **16**, 35-40 (2005).

18. Griffin,B.A., Adams,S.R., & Tsien,R.Y. Specific covalent labeling of recombinant protein molecules inside live cells. *Science* **281**, 269-272 (1998).

19. Benhar,I. Biotechnological applications of phage and cell display. *Biotechnology Advances* **19**, 1-33 (2001).

20. Rizzo,M.A., Springer,G.H., Granada,B., & Piston,D.W. An improved cyan fluorescent protein variant useful for FRET. *Nature Biotechnology* **22**, 445-449 (2004).

21. Webb,Y., Hermida-Matsumoto,L., & Resh,M.D. Inhibition of protein palmitoylation, raft localization, and T cell signaling by 2-bromopalmitate and polyunsaturated fatty acids. *Journal of Biological Chemistry* **275**, 261-270 (2000).

22. Watt,R.M. & Voss,E.W. Mechanism of Quenching of Fluorescein by Anti-Fluorescein Igg Antibodies. *Immunochemistry* **14**, 533-541 (1977).

23. Daugherty,P.S., Iverson,B.L., & Georgiou,G. Flow cytometric screening of cell-based libraries. *J. Immunol. Methods* **243**, 211-227 (2000).

24. Cormack,B.P., Valdivia,R.H., & Falkow,S. FACS-optimized mutants of the green fluorescent protein (GFP). *Gene* **173**, 33-38 (1996).

25. Nguyen,A.W. & Daugherty,P.S. Evolutionary optimization of fluorescent proteins for intracellular FRET. *Nature Biotechnology* **23**, 355-360 (2005).

26. Zufferey,R., Donello,J.E., Trono,D., & Hope,T.J. Woodchuck hepatitis virus posttranscriptional regulatory element enhances expression of transgenes delivered by retroviral vectors. *Journal of Virology* **73**, 2886-2892 (1999).

27. Cubitt,A.B., Heim,R., Adams,S.R., Boyd,A.E., Gross,L.A., & Tsien,R.Y. Understanding, improving and using green fluorescent proteins. *Trends Biochem. Sci.* **20**, 448-455 (1995).

Chapter 2

Inducible aggregation of tetracysteine-GFP fusion proteins for reversible protein inactivation

Abstract

The ability to rapidly and reversibly control the activity and mobility of cellular proteins would be a powerful method to study complicated protein dynamics in cells. Because of the sparse availability and huge effort required for the development of specific pharmacological agents, new general methods for protein modulation would be highly desirable. During selection for optimized tetracysteine tags with improved affinity and fluorescence, one novel sequence, YRECCPGCCMWR, was discovered capable of inducing protein aggregation upon biarsenical labeling in living cells or *in vitro*[1]. Aggregation is initiated by labeling with ReAsH or CHoXAsH, but not with FlAsH. Additionally, aggregation is highly dependent on fusion to the N-terminus of GFP, whereas fusions to CFP, YFP, or the C-terminus of GFP fail to aggregate. Various fiber morphologies are attainable by N-terminal fusion to epitope tags, or by fusion to various proteins. The tetracysteine-GFP fusion is entirely resistant to high concentrations of dithiol competitors and persists in cells for days with no observable toxicity. Upon intense spatial illumination, biarsenical fluorescence is rapidly eliminated and GFP fluorescence is released throughout the cell. Fusion of the aggregating tetracysteine-GFP tag to several cellular proteins confers biarsenical-dependent aggregation. CHoXAsH labeling of the regulatory subunit of PKA (RIα) or the catalytic subunit (Cα) led to rapid aggregation of PKA holoenzyme, including endogenous or recombinantly expressed mCherry-Cα or mCherry-RIα respectively. Following forskolin stimulation, Cα fails to release from the aggregated PKA holoenzyme, preventing full PKA activity. In

summary, a new light reversible, aggregation-based method for reversible protein aggregation has been developed capable of modulating protein activity in living cells.

Introduction

Nature has many methods for disrupting protein activity; including chemical modification, spatial sequestration or compartmentalization, associations with inhibitory partners, or degradation. By far, the most direct means to inactivate a protein is to eliminate it. Genetic knockouts are the most tried means for studying the individual contribution of a single protein in a biological process. Standard methods for knocking out a gene are especially laborious for mammalian systems, and RNA interference techniques are generally limited to cultured cell lines. In each case, the knockout target is the genetic material, not the protein itself or its nearby binding partners. Long-term genetic inactivation often leads to the initiation of compensatory pathways, masking the true biological role of the targeted gene. Additionally, inactivation by siRNA pathways may require extended periods for full phenotype penetration, since the turnover rate of each protein target varies. The efficiency of nucleotide-based methods for protein inactivation can only be measured by protein blotting or immunofluorescence, which requires the availability of specific antibodies. Furthermore, elimination of a gene irreversibly prevents activity throughout the entire cell, tissue, or organism with no spatial or temporal control.

Recently, a method has been developed for rapid, genetically targeted protein inactivation with spatial and temporal resolution, named chromophore assisted light inactivation, or CALI[2]. This technique is based on the propensity of certain fluorophores to generate short-lived oxygen radicals, such as singlet oxygen[3], following intense illumination. Singlet oxygen exerts its damaging effects over the range of several

nanometers, specifically damaging nearby proteins by intense oxidation[4]. Targeted oxidation by CALI has been successfully shown as a potent method for protein inactivation[5]. For example, extracellular proteins targeted with fluorescein-labeled monoclonal antibodies or scFv phage has been used to selectively inactivate proteins on the cell surface. Screening of CALI targeted candidates has lead to the discovery of Hsp90 as a functional mediator of tumor invasion[6]. Unfortunately, fluorescently labeled antibody and phage technologies cannot access intracellular targets without microinjection or other perturbative strategies.

Genetic targeting of CALI was first demonstrated using the biarsenical-tetracysteine system on the primary sensor for calcium-mediated exocytosis, synaptotagmin I (SytI)[7,8]. In SytI deficient cells, exocytosis and vesicular trafficking are blocked. By temporally inactivating SytI after vesicle exocytosis by FlAsH-mediated CALI, a previously uncharacterized functional role for SytI in endocytosis was discovered. This example precisely demonstrates the power of spatial temporal activation over genetic ablation for uncovering rapid dynamics in live cells. Other general examples of biarsenical-tetracysteine targeted CALI include inactivation of tetracysteine-tagged gap junctions (Cx43) and calcium channels $(\alpha_1 C)$[9]. Furthermore, inactivation was greatly enhanced by labeling with ReAsH, and evidence was provided supporting the primary involvement of singlet oxygen.

Despite the precise spatial and temporal control attained using biarsenical-tetracysteine CALI, the singlet-oxygen mediated damage is likely irreversible, requiring new protein synthesis and turnover for activity restoration. Pharmacological competitive inhibitors have ideal dynamics for reversible inactivation, yet they lack a generic targeting mechanism. Fusions to the FK506 binding protein (FKBP) has found widespread use as a general method for chemically dimerizing fused proteins in cells

upon addition of a divalent ligand[10]. Mutation of F36M in FKBP causes spontaneous

homodimerization that reverses upon ligand addition[11]. Four tandem copies of the self-

dimerizing F36M mutant of FKBP cause spontaneous cross-linking and formation of

large, multi-protein aggregates that are dispersed upon addition of ligand, AP22543.

This method has been used to sequester proteins in the secretory pathway, such as

insulin[12], for drug-dependent release. Also, the ability to target large protein aggregates

in the secretory pathway has been used as a tool to study the transport of mega-vesicles

carrying cargo too large for normal vesicular trafficking[11]. In each case, to monitor

aggregate formation, GFP was additionally fused to the tandem aggregation domains,

resulting in a very sizeable tag (> 75kD). Furthermore, the basal state of F36M FKBP is

aggregation, so expression of fusions to biologically relevant proteins in cells would

require continuous incubation with ligand to prevent sequestration or inactivation.

Neither aggregation nor dispersion is particularly rapid, since they depend on the

addition and removal of ligand though incubations and washing steps.

The FKBP dimerization system can also be used directly to induce protein

inactivation. Clathrin light chain (CLC) naturally forms interlinking trimers that coat

vesicles and mediate endocytosis[13]. Due to the natural tendency of CLC to form higher

order multimers, expression of FKBP-CLC in cells and addition of the dimerizing ligand

FK1012-A leads to aggregate formation[14]. Again, to monitor the formation of CLC

aggregates, GFP was additionally fused to the FKBP-CLC gene. Rapid inhibition of up to

70% of endocytosis was observed following ligand addition, as assayed by the uptake of

labeled transferrin or LDL. This approach is not very transferable to other proteins, since

most proteins do not form high order multimers. Yet, this problem could be avoided by

using a similar tandem FKBP approach, as used with F36M FKBP, which could induce

aggregation upon FK1012-A addition by cross-linking in tagged proteins in trans.

During efforts to increase the fluorescence and affinity of the biarsenical-tetracysteine system, a novel tetracysteine sequence was discovered that in combination with GFP is capable of ligand-induced aggregation[1]. The biochemical and cellular properties of these aggregates will be discussed, as well as efforts to transform this artifact from a serendipitous selection for a universal tag for light-reversible, ligand-depended, aggregation-based protein inactivation.

Results

To optimize the fluorescence and affinity of the biarsenical tetracysteine system, a mammalian cell-based library was generated on the N-terminus of GFP for fluorescence selections by FACS[1]. Following ReAsH labeling, binding was monitored by FRET from GFP to ReAsH. By increasing the concentration of dithiol competitors over 16 rounds of sorting, non-toxic, highly expressed, bright, and highly dithiol resistant sequences were isolated. Despite the large number of sequence variants present in the library ($>3x10^7$), it was surprising that one of the three sequences selected after 16 rounds of enrichment achieved each of the selection criteria by an unexpected mechanism.

While investigating the properties of the three resulting sequences, one sequence, YRE#MWR-GFP (# = CCPGCC), was observed to form tiny fluorescent aggregates (**Fig. 2.1a-b**) upon ReAsH labeling[1]. These aggregates had no precise subcellular localization, forming randomly throughout the cell. Essentially all of the tetracysteine-GFP protein aggregated, with the exception of a small fraction of membrane-localized palmitoylated protein, as observed with other tetracysteine sequences. Furthermore, aggregates remained relatively autonomous and had limited diffusibility or movement from their site of induction, with no observed coalescence into

larger aggregates. Inclusion of the anti-dimerization A206K mutation in GFP (mGFP)[15]

did not affect the ability to form aggregates, implying the mechanism does not require

cooperative aggregation mediated by GFP dimerization (**Fig. 2.1c**). When the FLAG

epitope tags were fused to the N-terminus of the tetracysteine to block palmitoylation,

the morphology of the aggregates changed dramatically from punctae to more fiber-like

ribbons (**Fig. 2.1d**). This was the first observation of the variety of polymerized states

observed upon biarsenical-mediated aggregation.
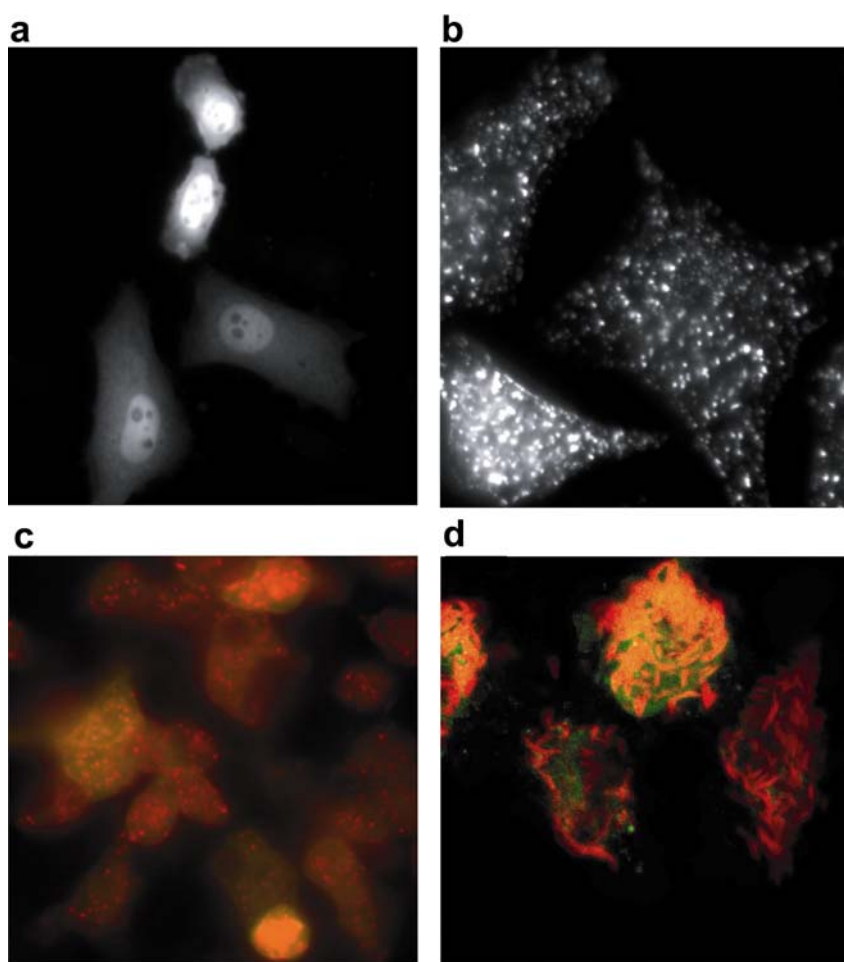
Figure 2.1. YRE#MWR-GFP aggregates following ReAsH labeling. (a) FLN#MEP-GFP cells labeled with ReAsH show diffuse localization. (b) YRE#MWR-GFP cells labeled with ReAsH show distinct punctae. (c) Monomerization of GFP with the A206K mutation has no effect on aggregation. Shown as an overlay of both GFP and FRET emission. (d) FLAG-YRE#MWR-mGFP labeled with ReAsH. Confocal acquisition and fluorescence overlay of GFP and ReAsH fluorescence. Exact fluorescence and region scaling vary between each image.

FACS only measures total cellular fluorescence regardless of aggregation, since spatial information is lost during the bulk fluorescence collection by photomultipliers. YRE#MWR-GFP expresses approximately 30% higher in 293T cells than FLN#MEP-GFP, less than 10% less than untagged GFP, and 5.9 fold better than GFP-AEAAARECCPGCCARA (αPG). Following labeling, the efficiency of YRE#MWR-GFP distribution into discrete aggregates is striking and apparently led to its selective advantage. Further analysis of ReAsH labeled YRE#MWR-GFP unmistakably demonstrates the optimal selectable phenotype: high resistance to dithiols and a consistently high FRET ratio (**Fig. 2.2**). Cells expressing YRE#MWR-GFP aggregates are essentially immune to dithiol competitors, including the 10 mM 2,3-dimercaptopropanol (BAL), which we regard as the maximal soluble concentration of dithiol which reverses even the highest affinity tags[16]. During the high dithiol washes, FACS analysis revealed the critical concentration required for aggregation of YRE#MWR-GFP protein is equal to or less than 2 μM of diffuse protein in a cell. Furthermore, there is an increase in the slope of ReAsH to GFP fluorescence in YRE#MWR-GFP labeled cells. This increase is due to intermolecular FRET caused by the extremely close proximity of ReAsH-GFP protein in the aggregates. Clearly, aggregated protein is extremely protected from the cellular environment, possibly even changing to a solid precipitate, since small molecules such as BAL are excluded.
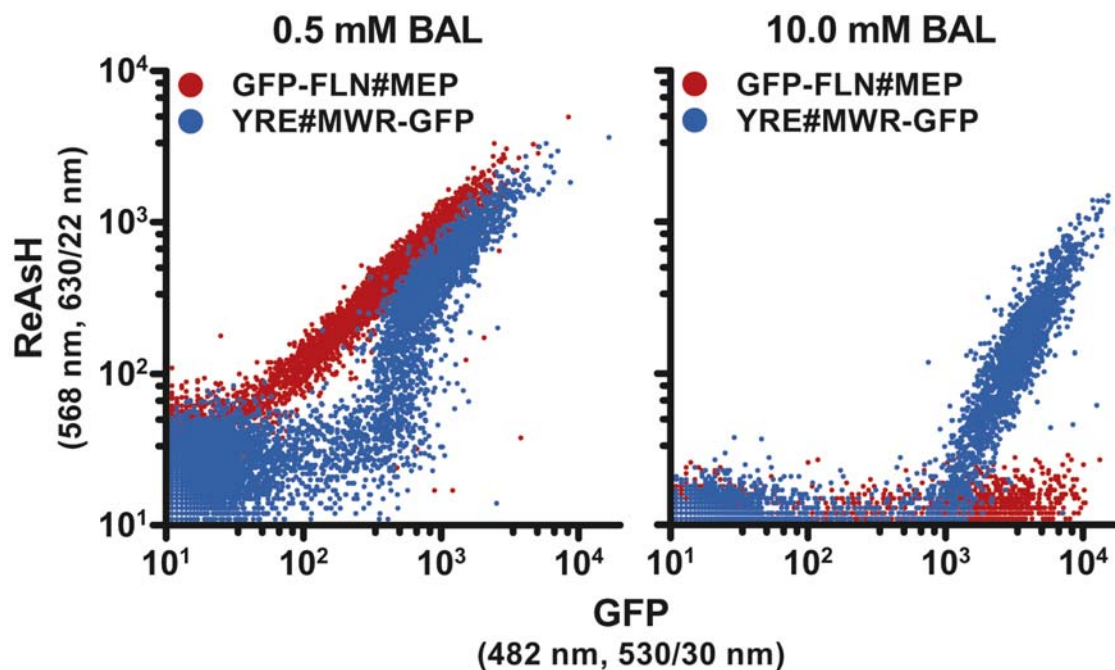
Figure 2.2. FACS analysis of YRE#MWR-GFP expressing cells. Clearly, YRE#MWR-GFP is highly resistant to dithiol competitors, since addition of 10mM BAL causes only a drop in the background fluorescence, but little affect on the ReAsH emission. The slope of GFP versus ReAsH fluorescence is slightly increased at the higher BAL concentration. Also, aggregates only form in cells exhibiting at least $2x10^6$ counts of GFP fluorescence, which is equivalent to approximately 2 μM GFP protein, as determined by standard beads.

YRE#MWR-GFP protein purified from either bacteria or mammalian cell lysates aggregates upon ReAsH binding *in vitro*. These aggregates are essentially precipitates, since even moderate speed centrifugation leads to sedimentation of the labeled protein. Interestingly, aggregation also occurs when protein is labeled with CHoXAsH, but not with FlAsH (**Fig. 2.3**). CHoXAsH and ReAsH are structurally similar[16], so it is most likely the additional benzoic ring of fluorescein prevents FlAsH induced aggregation either sterically or by affecting the solubility.
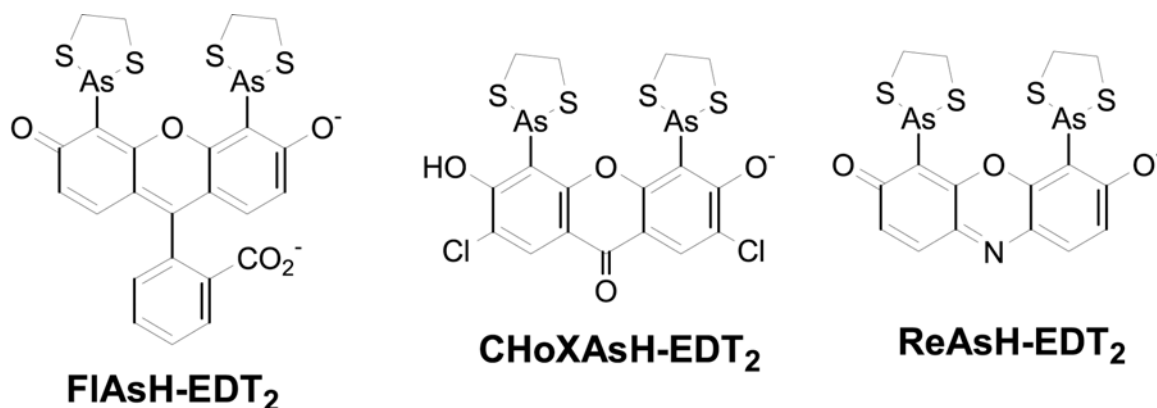
Figure 2.3. Chemical structures of three biarsenical dyes. FlAsH is green fluorescent, CHoXAsH is blue fluorescent, and ReAsH is red fluorescent.

The fluorescent quantum yield of ReAsH and FlAsH labeled YRE#MWR-GFP is 0.25 and 0.45 respectively. In comparison, the non-aggregating optimal tetracysteine FLN#MEP-GFP has ReAsH and FlAsH quantum yields of 0.49 and 0.78[1]. The mechanism of fluorescence enhancement for FLN#MEP and HRW#KTF is highly dependent on the presence of a phenylalanine at the first or last randomized position, or a tryptophan at the third position[1]. Neither of these proven enhancements is present in YRE#MWR, although the tyrosine at the first position is structurally similar to phenylalanine, and a tryptophan is present at the fifth position rather than the third position.

To collect further biochemical insight to the contribution of charge in the formation of biarsenical-tetracysteine-GFP aggregates, salt was removed from ReAsH labeled protein. GFP is extremely resilient, remaining folded and fluorescent even in the presence of denaturants, detergents, or changes in ionic strength. Following salt removal, ReAsH aggregation was disrupted. Next, SDS, an ionic detergent, or NP-40, a non-ionic detergent, were tested for their ability to solubilize the aggregates. One percent SDS was sufficient for aggregate solubilization, whereas addition of one percent

NP-40 had no effect. In addition, the fluorescence quantum yield of ReAsH bound to YRE#MWR-GFP was only slightly affected in the absence of salt or in the presence of SDS, but was quenched 5-fold (0.05) in the presence of NP-40 (**Fig. 2.4**). These results imply a degree of charge dependence in the ReAsH-induced tetracysteine-GFP aggregates. The YRE#MWR sequence contains several large amino acids, two positively charged arginines, and a negatively charged glutamate. The imbalance in charge could provide a mechanism for salt bridge formation between separate ReAsH-tetracysteine proteins in the aggregates. Also, as demonstrated by the inability of NP-40 to solubilize aggregates, hydrophobic interactions are not alone sufficient for aggregation.
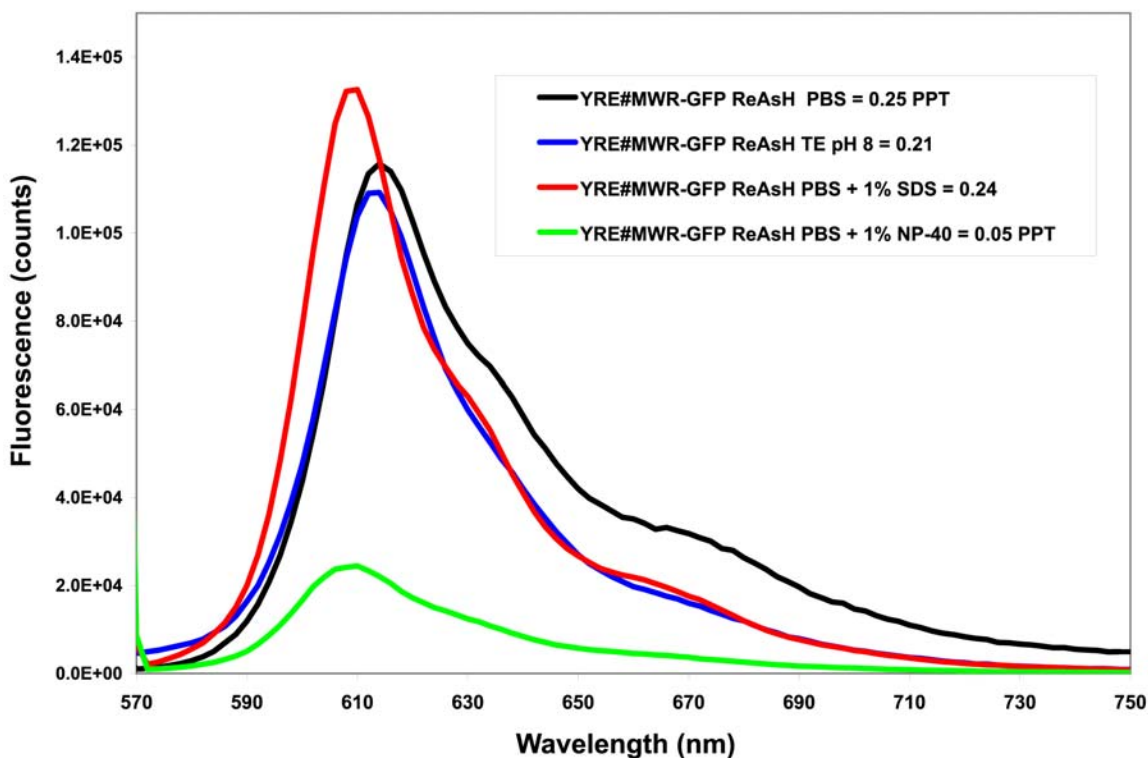
Figure 2.4. Detergents and salts alter properties of YRE#MWR-GFP. Lower salt or adding detergents affects the shape of the fluorescence emission curve. SDS shifts the emission peak lower, while NP-40 quenches fluorescence 5-fold.

Transfer of the tetracysteine to the C-terminus of GFP prevents aggregate formation, showing that the short tag is not sufficient for aggregation. N-terminal fusion of YRE#MWR alone, without GFP, to β-actin also showed no ReAsH dependent aggregation. Clearly, the N-terminal fusion to GFP is obligatory, and may either stabilize an aggregation prone conformation or mediate unique interactions between the biarsenical tetracysteine complex and the GFP protein. Replacement of Emerald GFP (F64L, S65T, S72A, N149K, M153T, V163A, I167T)[17] with EGFP (F64L, S65T, V163A)[18] did not affect aggregation. On the contrary, transfer of the YRE#MWR using the exact same linkers to the N-terminus of ECFP (F64L, S65T, Y66W, N146I, M153T, V163A)[17] or Citrine YFP (S65G, V68L, Q69M, S72A, T203Y)[19] also prevents aggregation, despite CFP only differing from Emerald and EGFP by two single point mutations present (Y66W and N146I) (**Fig. 2.5**).
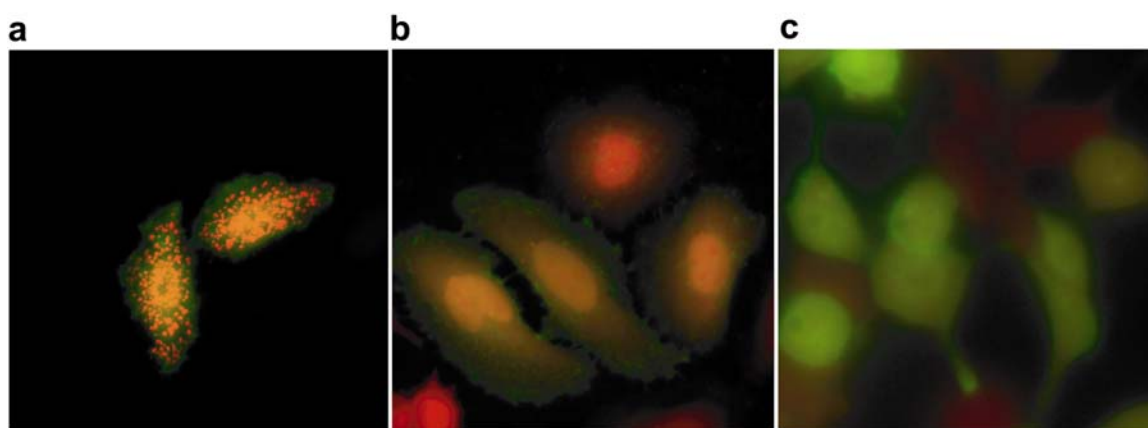
Figure 2.5. Aggregate formation is blocked in some fluorescent protein mutants. (a) YRE#MWR-EGFP forms aggregates, but YRE#MEP-Citrine (b) and YRE#MWR-CFP (c) fail to form aggregates. Palmitoylation is revealed upon labeling in HeLa cells (a-b) but not as severe in 293T cells (c).

Y66W is present on the internal α-helix that forms the chromophore and N146I is on the opposite end of the β-barrel where it interacts with the chromophore (**Fig. 2.6**). Neither of these mutations can easily explain either the mechanism or the structural elements

required for aggregation, although any interaction would likely occur with the loops on the top end of GFP. The precise mechanism may require higher resolution structural analysis, but significant knowledge could be gained by performing an alanine scan across the YRE#MWR sequence. Once the relevant residues are assigned, the interaction with nearby GFP residues can be correlated with the slight structural changes between CFP and GFP.
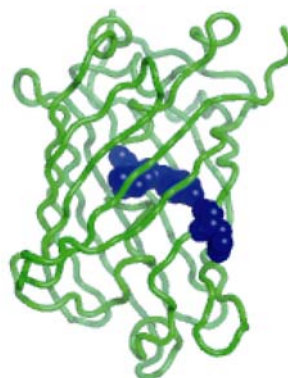
Figure 2.6. Location of Y66W and N146I mutation on GFP. The N-terminus is visible in the upper right, and is far from the blue colored mutations at the chromophore. The large loop facing the N-terminus is a potential site of interaction of the ReAsH-YRE#MWR complex, especially since this loop initiates from the internal α-helix containing the chromophore.

While observing the dynamics of ReAsH YRE#MWR-GFP aggregates in cells, it immediately was apparent the rate of ReAsH photobleaching was greatly enhanced in aggregates. Following the capture of several images in time-lapse, a majority of ReAsH fluorescence had vanished. This is presumably due to the extremely tight association biarsenical-tetracysteine-GFP complexes in the aggregates, localizing the effects of any oxygen radicals generated during the photobleaching reaction to the entire aggregate. Astonishingly, as ReAsH photobleached, GFP fluorescence rapidly redistributed out of the aggregates (**Fig. 2.7**). Similar results were later demonstrated for CHoXAsH, which has previously been reported as the most photosensitive biarsenical[16]. By spatially restricting the ReAsH illumination one side of a cell, GFP can be locally released and

diffuses throughout the entire cell. ReAsH inducible aggregation is therefore reversible, and holds significant potential as a novel approach for rapid spatially dispersion of the aggregated protein.

| Before Bleach | Mask ½ cell with Iris, photobleach | After photobleach |
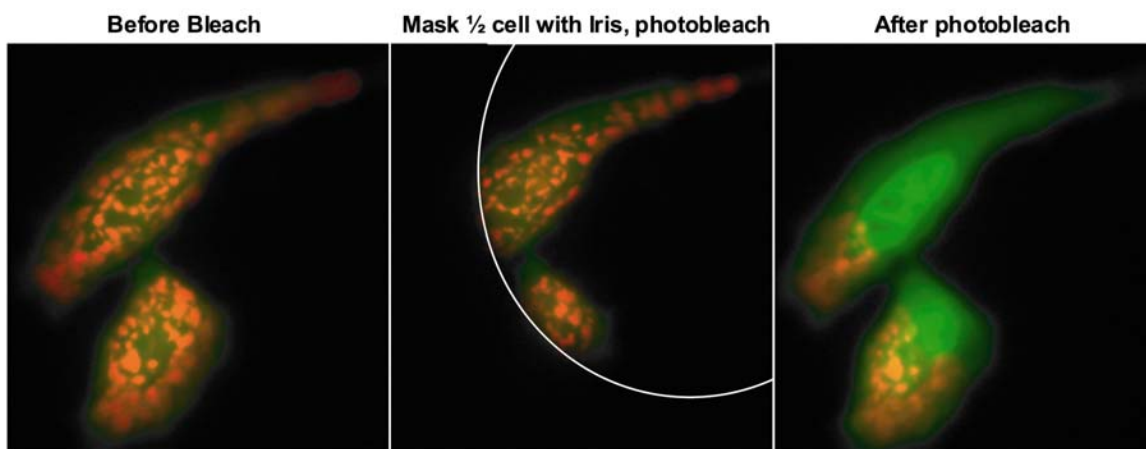
Figure 2.7. YRE#MWR-GFP aggregates are released by photobleaching. Closing the microscope diaphragm alls for selective bleaching of one region of a cell. Following photobleaching, GFP rapidly diffuse throughout the entire cell.

The mechanism of aggregate dispersion following ReAsH photobleaching is not understood, and likely involves multiple complicated photochemical reactions. Since nothing is known about the mechanism and products of ReAsH photobleaching[9], it will be difficult to determine the mechanism of aggregate dispersion. ReAsH is likely oxidized, but it is not clear if it falls off or remains attached to the tetracysteine. To determine the state of the tetracysteine motif following aggregate dispersion by photobleaching, ReAsH was applied again to the bleached cells in an attempt to re-aggregate the tetracysteine-GFP protein (**Fig. 2.8**). YRE#MWR-GFP cells were labeled with ReAsH and photobleached, then washed with high levels of dithiol to remove any remaining arsenicals attached to the tetracysteine. After additional washes to remove excess dithiol, ReAsH was re-applied, but failed to induce further aggregate formation. Similar experiments with CHoXAsH gave equivalent results.
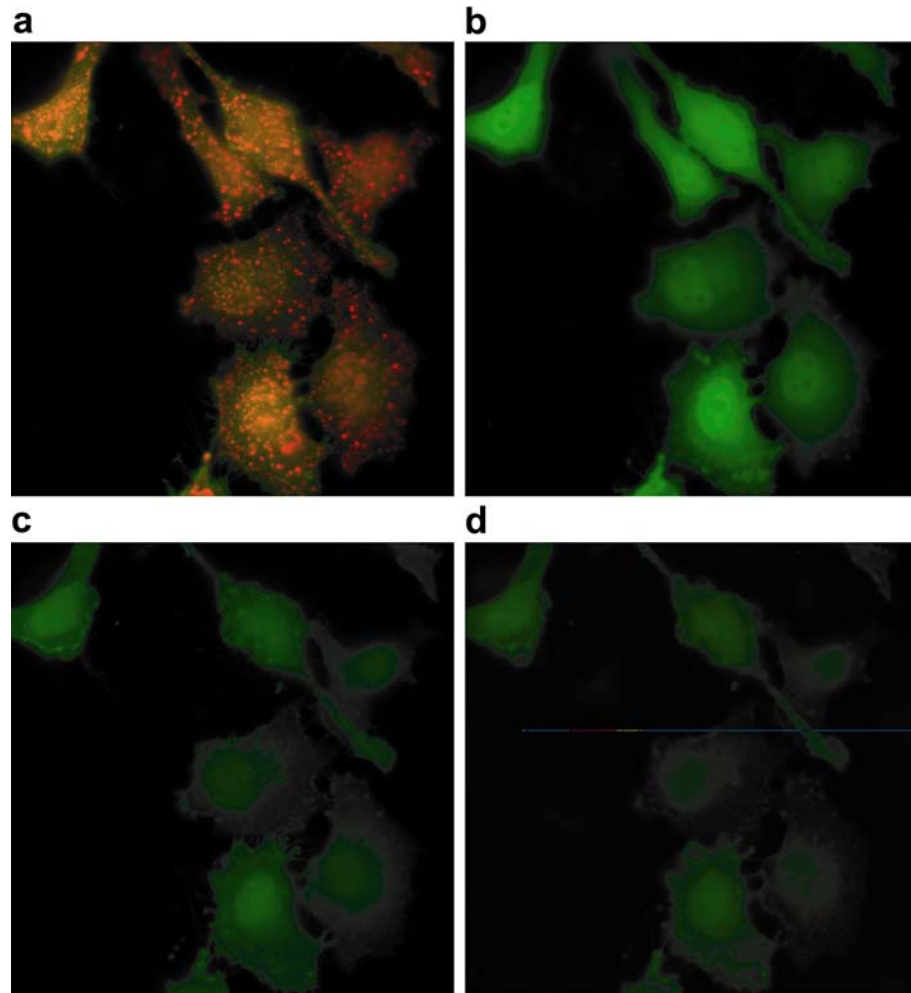
Figure 2.8. YRE#MWR-GFP re-aggregation blocked after photobleaching. (a) ReAsH labeling was performed for 45 minutes to induce aggregates, and then washed with 0.1mM BAL to remove excess dye. (b) Cells immediately following a one-minute ReAsH photobleach. (c) Cells washed with 0.6 mM BAL for 20 minutes, then rinsed four times. (d) No aggregates are visible after a similar 45 minute ReAsH stain.

The tetracysteine-GFP protein becomes permanently modified during photobleaching, preventing future biarsenical binding. Oxidation of the cysteines to disulfides would not permanently prevent biarsenical binding, since intracellular glutathione or the addition of high concentrations of dithiols would each reduce the disulfides and restore the tetracysteine motif. To irreversibly damage the tetracysteine site, one or more cysteine thiols could become oxidized to a sulfinic ($-SO_2H$) or sulfonic acid ($-SO_3H$). It is also

possible other regions of the tetracysteine-GFP have accrued oxidative damage that

block further aggregate formation. During ReAsH photobleaching, GFP also accrues

damage, since the fluorescence quench of GFP by ReAsH FRET is only partially

restored. As demonstrated by these results, photo-release does not innocuously release

proteins from the aggregates, but may be damaging to tethered or nearby proteins (**Fig.

2.9**).
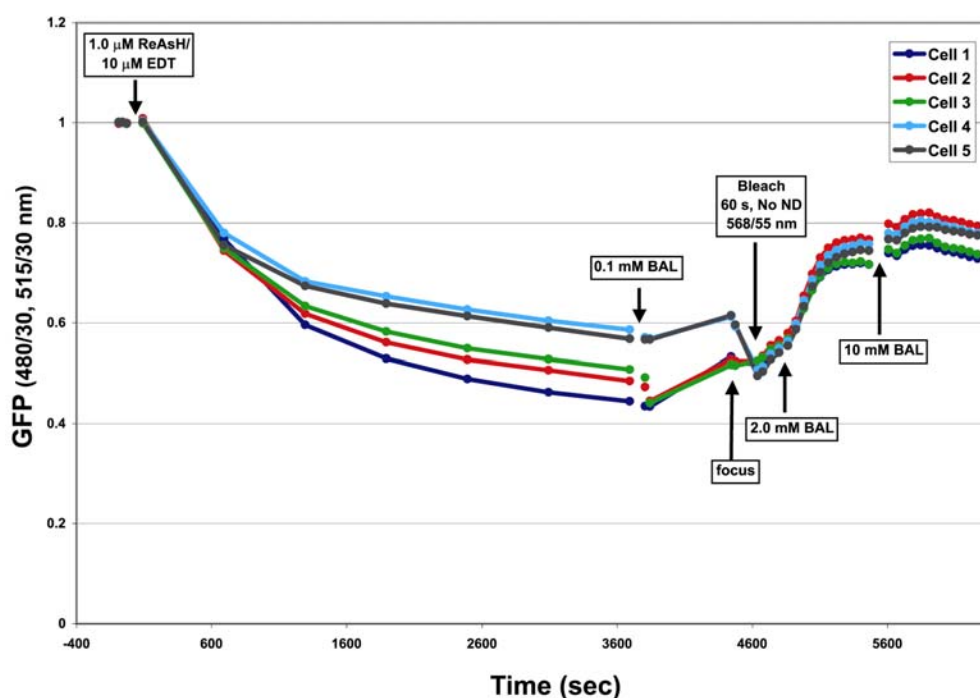
Figure 2.9. Timecourse of labeling and bleaching of ReAsH. Bleaching does not restore GFP
fluorescence. Washing with 10 mM BAL only partially restores GFP fluorescence.

The most exciting application of ReAsH induced aggregation is the potential to

selectively inactivate proteins in living cells, followed by precise spatial and temporal

control of their release. To achieve this goal, YRE#MWR-GFP was fused onto the N-

terminus of β-actin and α-tubulin. These fusions resemble standard GFP fusions, except

they contain an additional eighteen amino acid tetracysteine tag on their N-terminus.

ReAsH labeled protein did indeed aggregate in transfected cells, but the aggregation

was slower and required extended incubations at room temperature to depolymerize the

cytoskeleton (**Fig. 2.10**). Only depolymerized β-actin and α-tubulin become aggregated,

leaving the cytoskeleton intact. High levels of endogenous untagged protein are most

likely at fault, diluting the tagged protein in polymers and preventing the initiation of

aggregation. Importantly, this example demonstrates that without gene replacement,

active endogenous protein can still support activity. It is unclear if gene replacement with

the aggregating tag would lead to more efficient aggregation, or if the cytoskeleton is

sufficiently rigid that labeled protein cannot freely associate to form larger aggregates.
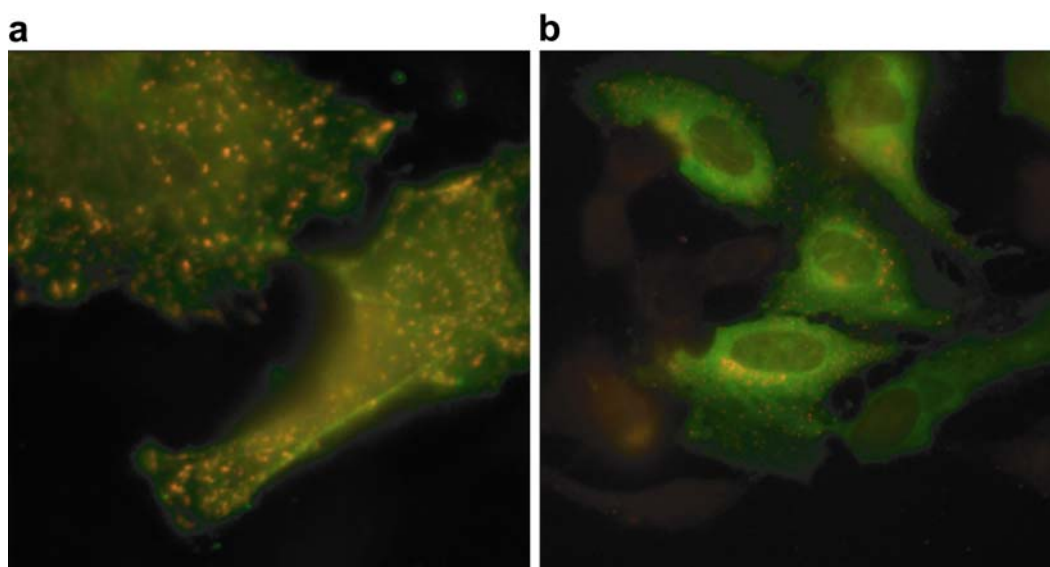
Figure 2.10. ReAsH labeling of YRE#MWR-GFP tagged β-actin and α-tubulin. (a) ReAsH aggregates of YRE#MWR-GFP- β-actin are visible, but actin stress fibers do not disassemble and are not aggregated. Image taken with a 100X objective. (b) YRE#MWR-GFP-a-tubulin also forms aggregates only from non-polymerized α-tubulin.

In an effort to avoid high level expression of endogenous rigidly structured

proteins, YRE#MWR-mGFP was fused to the N-terminus of the cytoplasmically

expressed bacterial enzyme β-lactamase gene[20]. B-lactamase activity can be monitored

with high sensitivity in live cells using the membrane permeant, fluorescent substrate

CCF2[20]. Using this system, enzyme levels down to only a few hundred molecules can be

easily detected, far lower than other gene expression reporters[21]. Furthermore, since β-lactamase is not of mammalian origin, no endogenous activity is observed. HEK293 cells stably expressing tetracysteine-GFP fusions to β-lactamase were labeled with CHoXAsH or ReAsH and monitored for the formation of aggregates. Only YRE#MWR-mGFP-β-lactamase expressing cells formed aggregates, whereas FLN#MEP-mGFP-β-lactamase distribution was unaltered (**Fig. 2.11**). As was observed with N-terminal epitope tag fusions, ribbon-like aggregate morphologies can form. Unfortunately, efforts to measure any potential disruption in activity were largely uninterpretable and did not show a linear relationship between ReAsH intensity and CCF2 activity. Due to these problems, we refocused our efforts towards other enzymes.
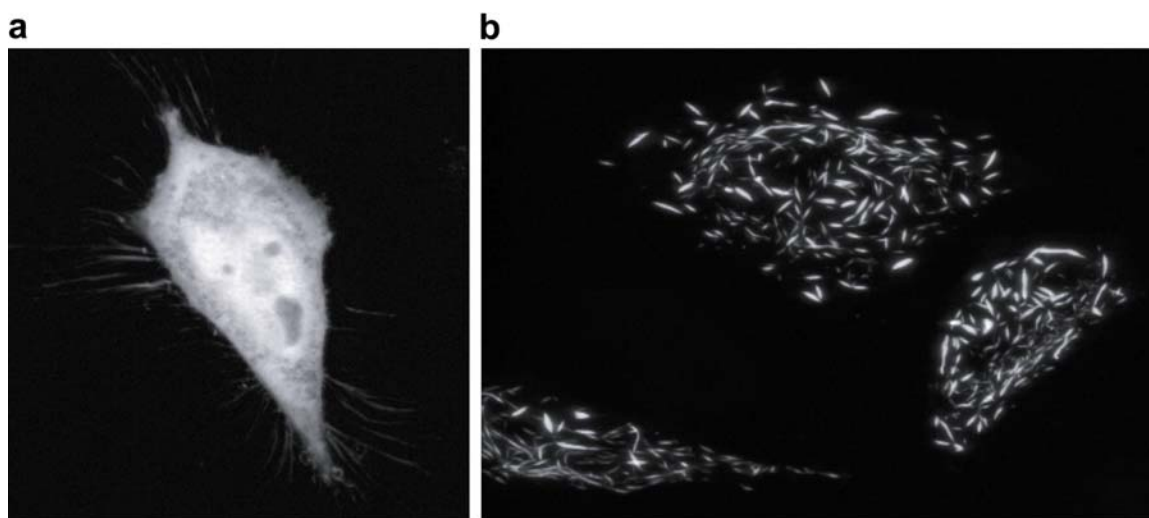
Figure 2.11. CHoXAsH labeled tetracysteine-mGFP-β-lactamase. (a) FLN#MEP-mGFP-β-lactamase shows diffuse labeling and no disruption of GFP localization. (b) YRE#MWR-mGFP-β-lactamase is has a fiber-like morphology, reminiscent of ReAsH labeled FLAG-YRE#MWR-mGFP.

Genetically encoded fluorescent reporters of protein kinase activity have recently been developed that are capable of dynamically monitoring phosphorylation dynamics in living cells[22-24]. AKAR, or A-kinase activity reporter, was the first serine/threonine kinase reporter developed specifically to detect the activity of protein kinase A (PKA)[23].

Phosphorylation of the kinase substrate sequence by PKA leads to an intramolecular association with a fused FHA phospho-binding motif, resulting in a dramatic conformational change[25]. This structural rearrangement results an alteration of the distance and orientation between the fused spectrally overlapping CFP and YFP chromophores, changing the FRET efficiency and the ratio of the CFP and YFP emission. By genetic fusion to specific targeting motifs, the reporter can me directed to precise sub-cellular localizations for defined spatial detection of activity. Because of the robust and dynamic activity observed with this reporter, PKA was chosen as a model system for assaying inactivation by ReAsH-YRE#MWR-GFP aggregation.

Protein kinase A (PKA) is essential for the regulation of diverse cellular processes, including ion channel conductance, metabolism, cell migration, and gene expression[26,27]. The targets of PKA phosphorylation are very diverse and localized throughout the cell, including the nucleus, the cytoplasm, and the plasma membrane. PKA is a heterotetrameric protein consisting of two regulatory (R) subunits and two catalytic (C) subunits. There are four genes for R (RIα/β and RIIα/β) and two genes for C (Cα/β), that form either type I or type II PKA holoenzyme. Upon binding cAMP, R dissociates and releases active C, initiating the phosphorylation of target sites and nuclear translocation.

RIα was chosen as the initial aggregation target, in the hope it could inactivate endogenous C subunit in *trans*. Unfortunately, fusion of YRE#MWR-GFP to RIα led to aggregate formation in the absence of ReAsH addition (**Fig. 2.12a**). Replacement of GFP with mGFP corrected this problem, and ReAsH addition quickly triggered the soluble protein into aggregates (**Fig. 2.12b**). Conversely, fusion of YRE#MWR-mGFP to the C-terminus of RIα did not lead to aggregation following ReAsH labeling, demonstrating the sensitivity of the N-terminus of YRE#MWR-mGFP to modification

(**Fig. 2.12c**). Also, type II PKA holoenzyme could not be aggregated by YRE#MWR-mGFP fusion to RIIα, although only a minimal linker was tested (**Fig. 2.12d**). The N-terminus of RIIα mediates both homodimerization and binding to PKA anchoring proteins (AKAPs)[28], so it is possible these rigid associations prevent aggregation.
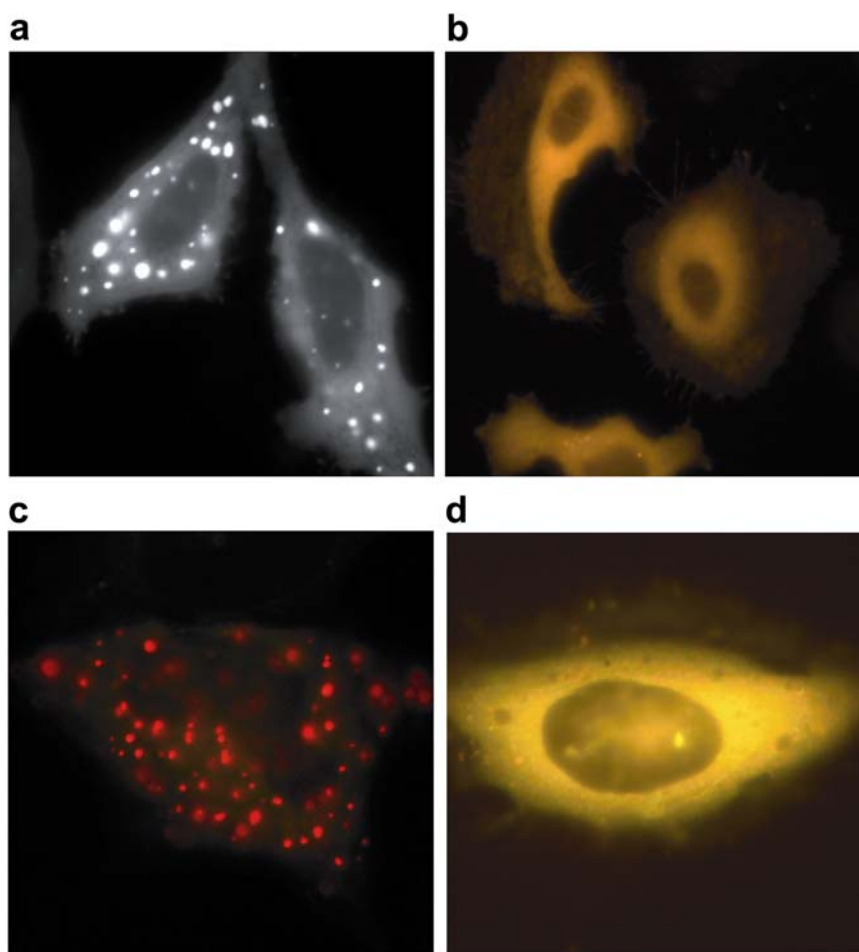
Figure 2.12. YRE#MWR-GFP fusions to PKA regulatory subunits. (a) GFP fluorescence of YRE#MWR-GFP-RIα shows aggregation independent of ReAsH labeling. Similar results were observed in the absence of tetracysteine fusion or with FLN#MEP. Cells were transfected two days prior to imaging. (b) RIα-YRE#MWR-mGFP does not aggregate following ReAsH staining, demonstrating the dependence of a free N-terminal tetracysteine tag. Green fluorescence is GFP and red fluorescence is ReAsH. (c) ReAsH dependent aggregation of YRE#MWR-mGFP-RIα is achieved by introducing the GFP monomerizing mutation A206K. (d) CHoXAsH labeling of cells co-transfected with YRE#MWR-mGFP-RIIα and mCherry-Cα does not initiate aggregation.

The timecourse of RIα aggregation is similar to YRE#MWR-GFP, and is
complete with 15 minutes of CHoXAsH labeling (**Fig. 2.13**). Sparse aggregates are
visible prior to labeling, but the majority of the expressed protein is diffusely cytoplasmic.
Interestingly, a lower expressing cell aggregates quicker than a highly expressing cell.
Therefore, the initiation of aggregate formation may require additional factors other than
concentration alone, but likely also requires triggering or seeding of a specific
conformation.
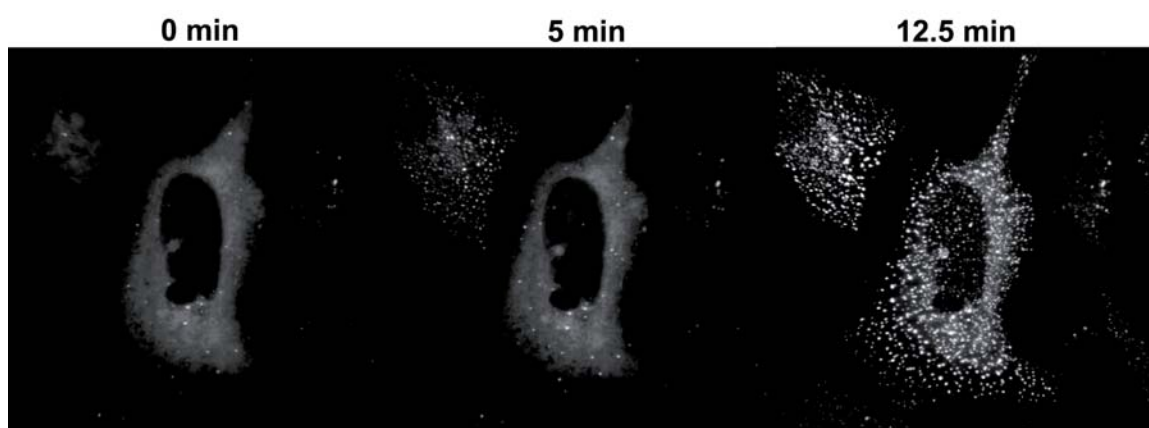
Figure 2.13. Timecourse of YRE#MWR-mGFP-RIα aggregation. Stably expressing HeLa cells
labeled with CHoXAsH displayed as a Z-projection of a confocal series. Aggregation in lower
expressing cells can take up to 30 minutes.

To test the trans-acting effects of aggregation, mCherry-Cα was co-transfected
into cells with YRE#MWR-mGFP-RIα and labeled with CHoXAsH (**Fig. 2.14a**).
Aggregation drew both mCherry and GFP into aggregates, visualized by co-localization
of red and green fluorescence. Therefore, mCherry-Cα can form holoenzyme with
YRE#MWR-mGFP-RIα, signifying endogenous C subunits also likely associate with
tagged RIα. Forskolin addition did not cause release of mCherry-Cα from the
aggregates (**Fig. 2.14b**), demonstrating inhibition of the natural biochemistry following

holoenzyme trapping. Because forskolin stimulation has little effect, a drop in forskolin
induced PKA activity should be detectable in cells containing aggregates.
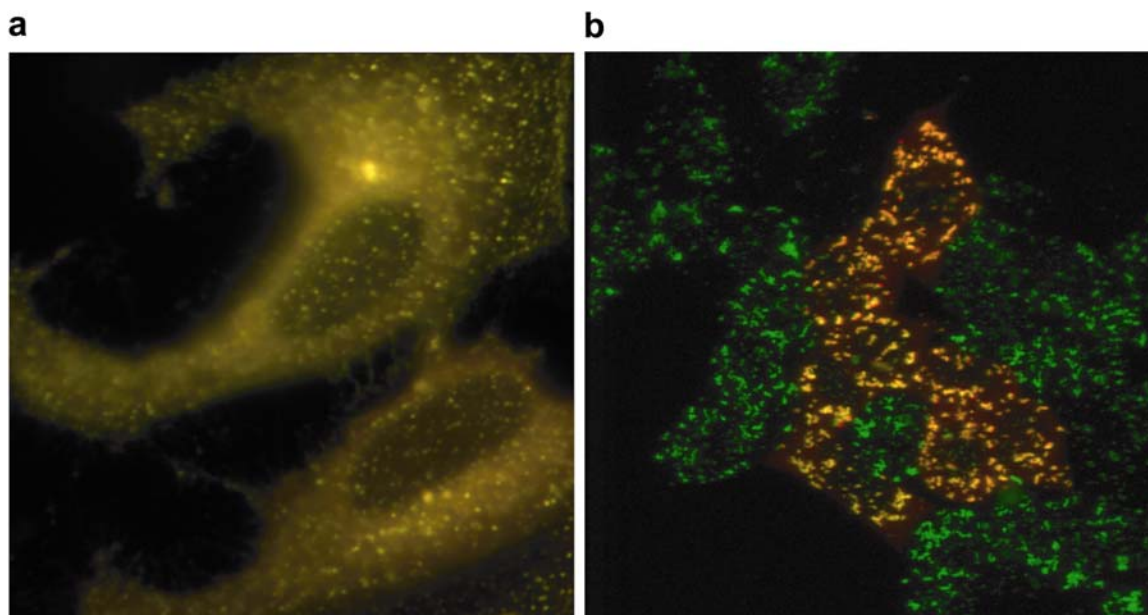
Figure 2.14. Co-localization of RIα and Cα in aggregates. (a) YRE#MWR-mGFP-RIα co-localizes with mCherry-Cα in HeLa cells labeled with CHoXAsH. (b) Confocal Z-projection of the same fusions expressed 293T cells aggregated with CHoXAsH then stimulated with 50 μM forskolin for 15 minutes. Not all cells are expressing mCherry-Cα, because the cells were generated by first making a stable cell line of YRE#MWR-mGFP-RIα, then infecting the cell line with mCherry-Cα virus.

Next, YRE#MWR-mGFP-RIα cell were transfected with AKAR2 and aggregated with either ReAsH or CHoXAsH, then strongly stimulated with forskolin and IBMX (**Fig. 2.15**). GFP fluorescence should not be affected by kinase-induced AKAR conformational changes, although GFP leakage into the FRET channel can decrease the observed percent ratio change. The overall differences in ratios should be unaffected by co-expression of GFP and AKAR2. PKA activity in the cytosol was marginally inhibited by CHoXAsH, reducing the rate of reporter phosphorylation. ReAsH induced aggregation was much more effective, reducing the kinase activity by 50%. Incomplete inactivation is

expected, since the transfected cells contain endogenous RI and RII subunits that bind

away Cα from the aggregates.
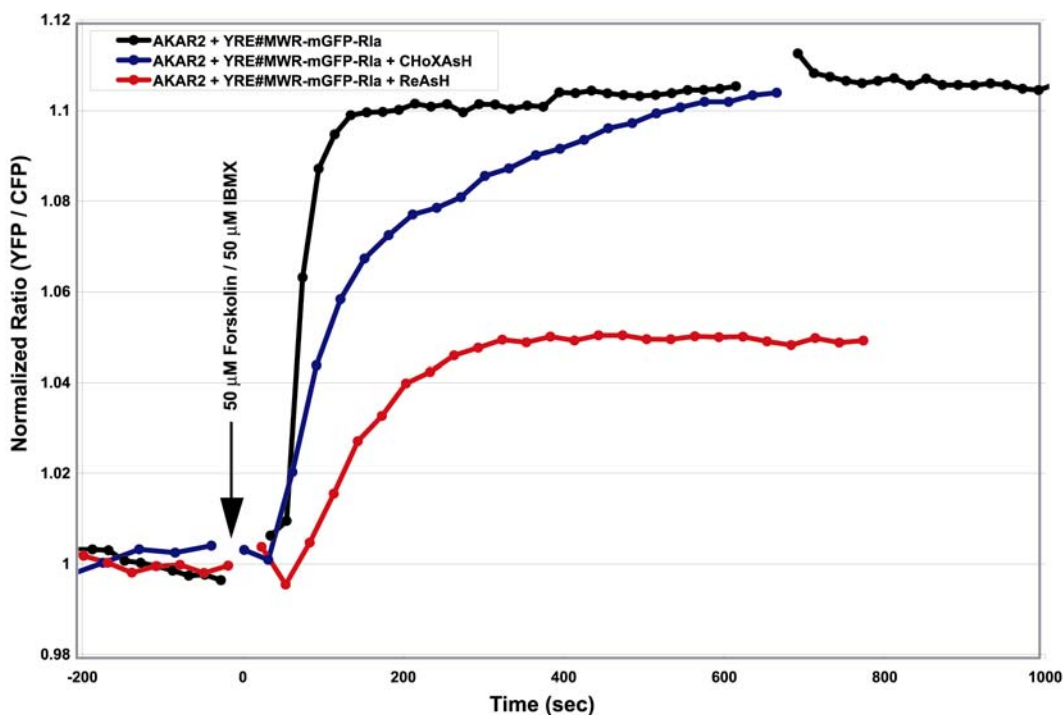
Figure 2.15. Cytosolic PKA is partially inhibited by RIα aggregation. AKAR2 was co-expressed with tagged RIα in HeLa cells and strongly stimulated with forskolin and IBMX.

In order to impose a more stringent test of PKA activity, AKAR2 was targeted to

the nucleus by fusion to an NLS sequence, spatially separating the reporter from the

aggregates[23]. Activated Cα must first be released from the aggregates, then translocate

to the nucleus to phosphorylate the reporter. Following stimulation, CHoXAsH and

ReAsH labeled cells demonstrated an approximate 4-fold reduction in nuclear PKA

activity (**Fig. 2.16**). When CHoXAsH is bleached with UV light, CFP is also illuminated by

as much as 20% of its maximal excitation. This leakage bleaches the AKAR reporter and

shifts the ratio. The same is true during ReAsH bleaching, leading to a similar offset.

Therefore, monitoring the light induced reversal of PKA aggregation is difficult using the

AKAR reporter.
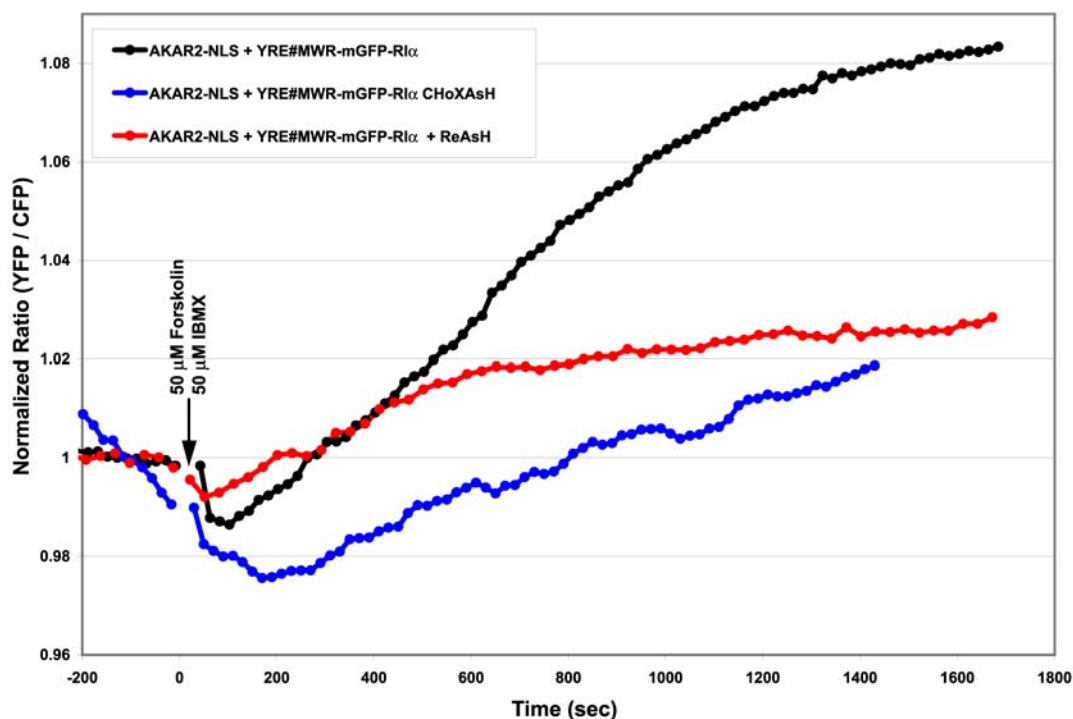
Figure 2.16. Nuclear PKA activity further inactivated by RIα aggregation. Nuclear localized AKAR2 detects a strong inhibition of PKA activity dependent on aggregation of YRE#MWR-mGFP-RIα in HeLa cells.

Aggregation-mediated PKA inactivation was next applied to the Cα subunit itself.

Fusion of the YRE#MWR-mGFP tag to Cα and CHoXAsH addition led to the formation

of punctuate aggregates, which co-localized with mGFP-RIα (**Fig. 2.17**, inset). Forskolin

treatment failed to release mGFP-RIα, reinforcing the observation that binding partners

are co-aggregated upon CHoXAsH addition. PKA activity was measured in cells by co-

transfecting AKAR2-NLS, mCherry-RIα, and YRE#MWR-mGFP-Cα, then stimulating

with forskolin and IBMX (**Fig. 2.17**). Co-expression of RIα is essential to prevent

excessive free Cα, which results in constitutively active kinase activity and eventual

toxicity. CHoXAsH labeling led to complete inactivation of PKA in YRE#MWR-mGFP-Cα

expressing cells, but no inactivation in FLN#MEP-mGFP-Cα expressing cells. These

results are surprising, since GFP-tagged Cα reportedly does not efficiently translocate to

the nucleus due to its increased size. Therefore, the measured nuclear PKA activity is

likely entirely from untagged, endogenous C subunit. Therefore, tagged and untagged

Cα are mixed in assembled PKA holoenzyme, and aggregation induced by the presence

of a single tagged Cα can act in a dominant negative fashion to sequester all Cα

subunits.
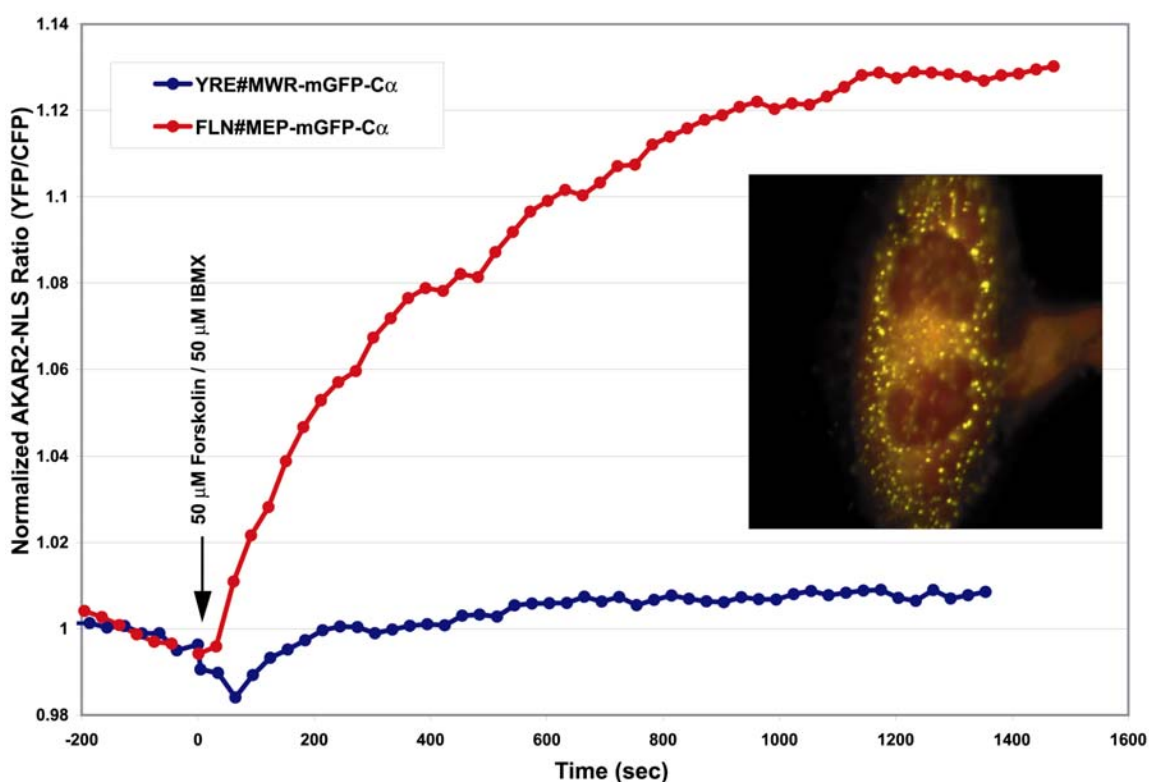
Figure 2.17. Inactivation of PKA by Cα aggregation Summary of multiple experiments, excluding one data set pair where FLN#MEP-mGFP-Cα was non-functional, and aggregated YRE#MWR-mGFP-Cα was highly active.

The interfering effects of endogenously expressed PKA subunits complicate and

diminish the resulting effects of aggregation-mediated inactivation. Also, observation of a

morphological phenotype would bypass the need for fluorescent activity reporters, which

suffer from spectral overlap problems upon biarsenical bleaching. To clearly determine

the potential ability of aggregation-mediated inactivation, mutant cells deficient in the

tagged components would be ideal.

Mouse embryonic RIα knockout fibroblasts have many striking phenotypes,

including increased basal PKA activity, enhanced wound healing, altered cytoskeletal

organization[29]. These cells were cultured and then rescued by viral transduction of

tetracysteine-mGFP-RIα. To test the ability of the rescued cells form functional PKA

holoenzyme, cells were stimulated with forskolin, and lysates were probed for enhanced

PKA activity using an anti-phospho-PKA substrate antibody (**Fig. 2.18**). Indeed,

expression of the tagged RIα did re-establish cAMP regulated kinase activity.
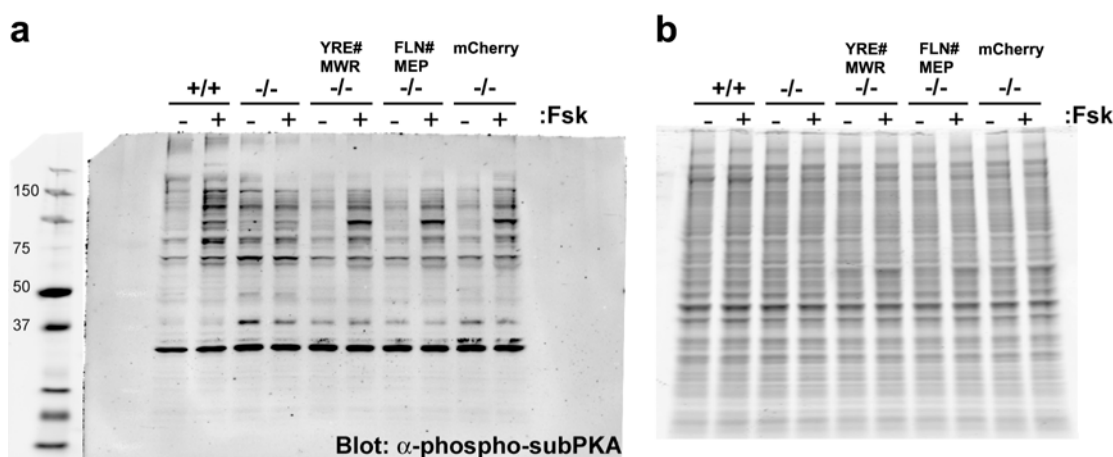
Figure 2.18. RIα fusions restore cAMP regulation in RIα null cells (a) Western blot analysis of phospho-PKA substrates shows forskolin inducible activity in wildtype cells, but not in the RIα$^{-/-}$ cells. This regulation is restored in the RIα$^{-/-}$ cells transduced with tagged RIα expressing virus. (b) Coomassie staining shows equivalent amounts of loaded protein.

Next, the cytoskeletal morphology was examined (**Fig. 2.19**). Null cells exhibit a

round shape with increased cortical actin and fewer filopodia-like extensions. Significant

reversion of the mutant cytoskeletal morphology was observed in the rescued cells.
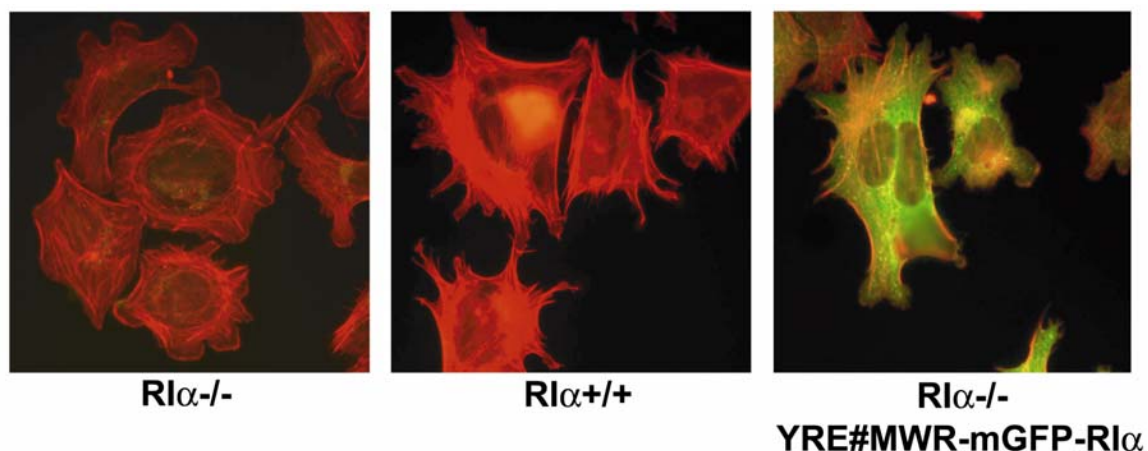
Figure 2.19. Cytoskeletal morphology is rescued by tagged RIα expression. Green fluorescence is from GFP expression and red is from rhodamine–phalloidin staining.

Next, cells were CHoXAsH labeled to induce aggregation, which occurred quickly as observed in other cell types. In preliminary observations, RIα aggregation leads to changes in the cytoskeleton and attachment to the matrix, causing the cell to retract and round. Further experiments will be required to follow up on this observation, including reversal of retraction by bleaching and release of PKA holoenzyme from the aggregates. Also, with the establishment of cell lines solely expressing tagged RIα, the aggregation-mediated inactivation can again be monitored by the AKAR reporter without interfering endogenous RIα subunit.

Discussion

The serendipitous isolation of YRE#MWR-GFP from a library of randomized tetracysteine sequences demonstrates an additional solution to the selection strategy for enhanced affinity and brightness. Because our selection criteria tolerated different subcellular localizations, a ReAsH dependent tag for protein aggregation was discovered

that functions in cells and on purified protein. Unfortunately, the linkage between the tetracysteine and the GFP is required and is not yet entirely understood, although only a few point mutations in GFP can disrupt this phenomenon. Although the aggregating tetracysteine cannot act alone as a short tag, the presence of the GFP is potentially advantageous, since the extent of aggregation can easily be monitored by microscopy.

Interestingly, fusion of the YRE#MWR-GFP tag to proteins such as β-actin, α-tubulin, β-lactamase, and multiple PKA subunits all result in aggregation. Inhibition of activity has been measured in live cells on two separate proteins: RIα, and Cα. Also, protein-protein interactions can also be determined in cells by co-transfecting one protein tagged with YRE#MWR-GFP and the other tagged with mCherry. Co-localization of the two colors within the aggregates demonstrates association. Surprisingly, photobleaching of the biarsenical can release proteins from the aggregate, potentially reversing the sequestration and inactivation. Unfortunately, each strategy for detecting protein activity uses fluorescent activity reporters that partially overlap with the spectra of the biarsenical-tetracysteine system, inhibiting light dependent re-activation by spatially controlled photobleaching.

The major limiting aspect of this technique is the dependence on a free N-terminus. Fusion of the YRE#MWR-GFP tag to the C-terminus of a protein has inhibited aggregation in all cases except when fused to short epitope tags. Clearly, a certain amount of conformational flexibility and space is required for aggregates to form, which could be ameliorated by the addition of linkers for added separation.

Possibly the strangest phenomenon is the ability to form various aggregate morphologies depending on the fusion protein. Addition of as few as a dozen residues upstream of the tetracysteine can alter the aggregates from punctae to fibers. Similar multiple aggregate morphologies have also been observed studying prions[30], but the

conformational changes are mediated by much larger protein domains, and generally insoluble to detergents. The fiber-like morphology is not restricted to short N-terminal tags, but has also been observed upon aggregation of β-lactamase, a 30 kD protein.

Other techniques for controlled inactivation have been found to biologically useful for studying otherwise intractable problems. CALI permanently destroys fluorophore labeled proteins, can be used with the biarsenical tetracysteine alone without GFP fusion, and does not intentionally disrupt protein localization[7,9]. FKBPs linked in tandem can be used in a similar manner to YRE#MWR-GFP aggregation[11], but is non-fluorescent and requires GFP fusion to visualize the extent of aggregation. Also, reversing aggregation is achieved by washing away excess ligand, which is quite slow.

Theoretically, the YRE#MWR-GFP aggregation system offers the most flexibility. Proteins are tagged with a slightly longer GFP containing the YRE#MWR tetracysteine, then labeled and aggregated with in 15-20 minutes. Endogenous binding partners co-localize in aggregates, and are resistant to pharmacological manipulation. Finally, the aggregated protein can be released on demand with precise spatial-temporal resolution by photobleaching. It is still unclear if photobleached, disaggregated protein retains fully active. Because of the known damaging affects caused during bleaching ReAsH, we prefer to use CHoXAsH as the aggregation agent, since it bleaches rapidly, preventing prolonged generation of oxygen radicals.

Another exciting application could be to capture in vivo interaction partners by aggregation in live cells. Following aggregation, the insoluble aggregates could be collected from lysates by centrifugation, then photobleached or solubilized in SDS to release them from other insoluble contaminants. Because the aggregates are highly fluorescent, it is quite easy to follow the fate of the tagged protein during any additional purification steps.

Further characterization and application of the YRE#MWR-GFP tag is still required before it can be advertised as a general protein inactivation technique. First, it is important to determine the structural components involved in aggregation. This knowledge is important for the designing new fusions to other proteins, as well as help gain insight into the mechanism of light induced solubilization. Additionally, the photo-reversibility will need to further characterized and whether full activity can be restored upon bleaching and aggregate breakdown. Finally, significant progress has been made with the use of genetically deficient cell lines rescued with the aggregating tag, which serves as the ideal model system for complete protein inactivation. There are currently few general methods for manipulating protein activity and localization in cells, and none that are rapidly reversible. Therefore, application of YRE#MWR-GFP as a general tag for aggregation-mediated inactivation would be of significant interest to many researchers.

Materials and Methods

Biarsenical Labeling

Biarsenical labeling was performed with 1.0 μM ReAsH or 2.5 μM CHoXAsH with 10 μM

EDT in HBSS (Hanks Balanced Saline Solution supplemented with 2 g / l glucose and

20 mM Hepes) for 30 to 60 minutes at room temperature. Cells were then washed with

0.2-0.5 mM BAL to remove excess non-specific biarsenical staining for 15 minutes.

Cell Culture, Transfection, and Transduction

HEK293T cells, HeLa cells, and MEFs were cultured in DMEM supplemented with 10%

FBS, 100 units / ml penicillin G, and 100 μg / ml streptomycin. Transfections and virus

preparations virus transduced stable cell lines were generated as previously described[1].

Flow cytometry Analysis and CCF2 staining

ReAsH labeled cells were analyzed simultaneously with non-aggregating tetracysteines,

as described previously[1]. Cells were stained with CCF2 according to manufacturer

protocols (GeneBlazer, Invitrogen) in which 1 μl of a 1 mM stock of CCF2 is diluted 1:10

in a solution containing 100 mg/ml Pluronic-F127 and 0.1% acetic acid in DMSO. This

mix is then diluted in 1 ml of HBSS supplemented with 2 g/L glucose and 20 mM Hepes.

To this, 10 μl of 250 mM probenecid was added per milliliter. After removing the growth

medium, the staining solution was added for approximately 45 minutes to 1 hour at room

temperature. Cells were then trypsinized and pelleted, then resuspended in the

supplemented HBSS for flow cytometry analysis. Cells were kept in the presence of

probenecid to prevent dye leakage and analyzed within 15 minutes. CCF2 cleavage and

ReAsH fluorescence were sequentially monitored on a FACS Vantage DiVa by dual

laser excitation at 413 nm and 568 nm, respectively. CCF2 cleavage by β-lactamase

was monitored by taking the ratio of blue (450/20 nm) to green (510/21 nm) emission. ReAsH intensity was collected with a 630/30 nm emission filter.

Analysis of Purified Protein

Protein was purified from stably expressing HEK293T cells using FlAsH-agarose affinity beads, as previously described[1,16]. Upon overnight labeling of purified YRE#MWR-GFP protein at 4°C, ReAsH or CHoXAsH aggregates were formed. Fluorescent quantum yields were measured as previously described[1] in the presence of 1% NP-40 (Sigma) or 1% SDS (Sigma).

Microscopy

Epifluorescent microscopy of GFP and ReAsH was performed as previously described[1]. AKAR2 monitored by sequential excitation for CFP and FRET fluorescence respectively Confocal microscopy was performed on a Zeiss LSM Live5 confocal microscope equipped with a 488 nm laser for GFP excitation and a 532 nm laser for ReAsH or mCherry excitation. All imaging experiments were performed at room temperature.

Table 2.1: Filter sets used for epifluorescence.

Fluorophore	Excitation (nm)	Dichroic (nm)	Emission (nm)
CFP	436/20	450	475/40
GFP	480/30	505	515/30
GFP ReAsH FRET	480/30	505	653/95
CFP-YFP FRET	436/20	450	535/25
mCherry	540/25	560	595/50
ReAsH	568/55	600	653/95

Generation of YRE#MWR fusions

YRE#MWR-GFP was isolated by RT-PCR and subcloned into pCLNCX-WPRE as
previously described[1]. GFP-YRE#MWR was generated by transferring the tetracysteine
to the C-terminus through several subcloning step, as previously described for GFP-
FLN#MEP[1]. FLAG-FLN#MEP-GFP was digested with BamHI and XhoI to remove the
tetracysteine-GFP fragment, and replaced with YRE#MWR-GFP to make FLAG-
YRE#MWR-GFP. YRE#MWR-GFP was digested with NotI and XhoI to remove the GFP
fragment, then replaced with CFP digested from FLN#MEP-CFP[1]. Citrine was amplified
by PCR (Primer1; Primer2) and ligated in the NotI/XhoI digested YRE#MWR-GFP vector
to make YRE#MWR-Citrine. GFP was monomerized by introduction of the A206K
mutation by QuikChange mutagenesis (Stratagene) from the published primers[15]. N-
terminal actin and tubulin fusions to FLN#MEP-GFP[1] were digested with NotI and XhoI,
then ligated into YRE#MWR-GFP plasmid. B-lactamase (Primer3; Primer4) was
amplified to introduce restriction sites for in-frame fusion with various tags. YRE#MWR-
mGFP (Primer5; Primer6), FLN#MEP-mGFP (Primer5; Primer6), mGFP (Primer7;
Primer6), and mCherry (Primer7; Primer6) were amplified and then ligated together with

β-lactamase. RIα (Primer8; Primer9), RIIα (Primer10; Primer11), and Cα (Primer12; Primer13) were amplified and digested with BglII or BamHI and XhoI to replacement of β-lactamase in the previously described vectors. To make C-terminal fusion to RIα, RIα (Primer14; Primer15) was amplified and digested with HindIII and BglII. YRE#MWR-GFP was digested with BamHI and XhoI and ligated with the RIα fragment into pCLNCX-WPRE.

Table 2.2. Primer sequences.

Number	Name	Sequence
Primer1	NRRL	TTTTAAGCTTGCCACCATGGCCGGATCCTAAGCGGCCGCAGCAAGGGCGAGGAGCTG
Primer2	3'FP XhoI	CCCCATCGATCTCGAGTTACTTGTACAGCTCGTCCAT
Primer3	5'BglII Bla	CTCAGATCTCGGGCTCACCCAGAAACGCTGGTGAAAG
Primer4	3'bla XhoI	GGGGCTCGAGTTACCAATGCTTAA
Primer5	5'pCL	AGCTCGTTTAGTGAACCGTCAGATC
Primer6	3' FP BglII	TCGAGATCTGAGTCCGGACTTGTACAGCTCGTCCATG
Primer7	5'H3KzFP	CCCAAGCTTGCCACCATGGTGAGCAAGGGCGAGGAG
Primer8	5'RI BglII	TTTAAGCTTAGATCTGCTTCCGGCACCACCGCCAGC
Primer9	3' NotI RI	TTTGCGGCCGCTTAGACAGACAGGGACACGAAGC
Primer10	5'RII	TTTGGATCCGGTTCAATGGGCCACATCCAGATCCCG
Primer11	3'RII	TTTCTCGAGTTACTGCCCGGGGTCCATCAG
Primer12	5'Cat	TTTGGATCCGGTTCAATGGGCAACGCCGCCGCCGCC
Primer13	3'Cat	TTTCTCGAGTTAAAACTCAGTAAACTCCTTGCC
Primer14	5'RI H3	TTTAAGCTTGCCACCATGGCTTCCGGCACCACCGCC
Primer15	3'RI Bgl2	TTTAGATCTGGCGACAGACAGGGACACGAAGC

References

1. Martin,B.R., Giepmans,B.N.G., Adams,S.R., & Tsien,R.Y. Mammalian cell-based optimization of the biarsenical-binding tetracysteine motif for improved fluorescence and affinity. *Nature Biotechnology* **23**, 1308-1314 (2005).

2. Jay,D.G. Selective Destruction of Protein Function by Chromophore-Assisted Laser Inactivation. *Proceedings of the National Academy of Sciences of the United States of America* **85**, 5454-5458 (1988).

3. Horstkotte,E., Schroder,T., Niewohner,J., Thiel,E., Jay,D.G., & Henning,S.W. Toward understanding the mechanism of chromophore-assisted laser inactivation - Evidence for the primary photochemical steps. *Photochemistry and Photobiology* **81**, 358-366 (2005).

4. Liao,J.C., Roider,J., & Jay,D.G. Chromophore-Assisted Laser Inactivation of Proteins Is Mediated by the Photogeneration of Free-Radicals. *Proceedings of the National Academy of Sciences of the United States of America* **91**, 2659-2663 (1994).

5. Eustace,B.K. & Jay,D.G. Fluorophore-assisted light inactivation for multiplex analysis of protein function in cellular processes. (2003).

6. Eustace,B.K., Sakurai,T., Stewart,J.K., Yimlamai,D., Unger,C., Zehetmeier,C., Lain,B., Torella,C., Henning,S.W., Beste,G., Scroggins,B.T., Neckers,L., Ilag,L.L., & Jay,D.G. Functional proteomic screens reveal an essential extracellular role for Hsp90α in cancer cell invasiveness. *Nat Cell Biol* **6**, 507-514 (2004).

7. Marek,K.W. & Davis,G.W. Transgenically encoded protein photoinactivation (FlAsH-FALI): Acute inactivation of synaptotagmin I. *Neuron* **36**, 805-813 (2002).

8. Poskanzer,K.E., Marek,K.W., Sweeney,S.T., & Davis,G.W. Synaptotagmin I is necessary for compensatory synaptic vesicle endocytosis in vivo. *Nature* **426**, 559-563 (2003).

9. Tour,O., Meijer,R.M., Zacharias,D.A., Adams,S.R., & Tsien,R.Y. Genetically targeted chromophore-assisted light inactivation. *Nature Biotechnology* **21**, 1505-1508 (2003).

10. Ho,S.N., Biggar,S.R., Spencer,D.M., Schreiber,S.L., & Crabtree,G.R. Dimeric ligands define a role for transcriptional activation domains in reinitiation. *Nature* **382**, 822-826 (1996).

11. Rollins,C.T., Rivera,V.M., Woolfson,D.N., Keenan,T., Hatada,M., Adams,S.E., Andrade,L.J., Yaeger,D., van Schravendijk,M.R., Holt,D.A., Gilman,M., & Clackson,T. A ligand-reversible dimerization system for controlling protein-protein interactions. *Proceedings of the National Academy of Sciences of the United States of America* **97**, 7096-7101 (2000).

12. Rivera,V.M., Wang,X., Wardwell,S., Courage,N.L., Volchuk,A., Keenan,T., Holt,D.A., Gilman,M., Orci,L., Cerasoli,F., Jr., Rothman,J.E., & Clackson,T. Regulation of Protein Secretion Through Controlled Aggregation in the Endoplasmic Reticulum. *Science* **287**, 826-830 (2000).

13. Edeling,M.A., Smith,C., & Owen,D. Life of a clathrin coat: insights from clathrin and AP structures. *Nature Reviews Molecular Cell Biology* **7**, 32-44 (2006).

14. Moskowitz,H.S., Heuser,J., McGraw,T.E., & Ryan,T.A. Targeted Chemical Disruption of Clathrin Function in Living Cells. *Mol. Biol. Cell* **14**, 4437-4447 (2003).

15. Zacharias,D.A., Violin,J.D., Newton,A.C., & Tsien,R.Y. Partitioning of lipid-modified monomeric GFPs into membrane microdomains of live cells. *Science* **296**, 913-916 (2002).

16. Adams,S.R., Campbell,R.E., Gross,L.A., Martin,B.R., Walkup,G.K., Yao,Y., Llopis,J., & Tsien,R.Y. 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 (2002).

17. Cubitt,A.B., Woollenweber,L.A., & Heim,R. Understanding structure-function relationships in the Aequorea victoria green fluorescent protein. *Methods in Cell Biology, Vol 58* **58**, 19-+ (1999).

18. Tsien,R.Y. The green fluorescent protein. *Annu. Rev. Biochem.* **67**, 509-544 (1998).

19. Griesbeck,O., Baird,G.S., Campbell,R.E., Zacharias,D.A., & Tsien,R.Y. Reducing the environmental sensitivity of yellow fluorescent protein - Mechanism and applications. *J. Biol. Chem.* **276**, 29188-29194 (2001).

20. Zlokarnik,G., Negulescu,P.A., Knapp,T.E., Mere,L., Burres,N., Feng,L.X., Whitney,M., Roemer,K., & Tsien,R.Y. Quantitation of transcription and clonal selection of single living cells with beta-lactamase as reporter. *Science* **279**, 84-88 (1998).

21. Zlokarnik,G. Fusions to beta-lactamase as a reporter for gene expression in live mammalian cells. *Applications of Chimeric Genes and Hybrid Proteins, Pt A* **326**, 221-241 (2000).

22. Ting,A.Y., Kain,K.H., Klemke,R.L., & Tsien,R.Y. Genetically encoded fluorescent reporters of protein tyrosine kinase activities in living cells. *Proc. Natl. Acad. Sci. U. S. A* **98**, 15003-15008 (2001).

23. Zhang,J., Ma,Y., Taylor,S.S., & Tsien,R.Y. Genetically encoded reporters of protein kinase A activity reveal impact of substrate tethering. *Proc. Natl. Acad. Sci. U. S. A* **98**, 14997-15002 (2001).

24. Zhang,J., Campbell,R.E., Ting,A.Y., & Tsien,R.Y. Creating new fluorescent probes for cell biology (vol 3, pg 906, 2002). *Nature Reviews Molecular Cell Biology* **4**, (2003).

25. Zhang,J., Hupfeld,C.J., Taylor,S.S., Olefsky,J.M., & Tsien,R.Y. Insulin disrupts beta-adrenergic signalling to protein kinase A in adipocytes. *Nature* **437**, 569-573 (2005).

26. Kopperud,R., Krakstad,C., Selheim,F., & Doskeland,S.O. cAMP effector mechanisms. Novel twists for an "old" signaling system. *FEBS Letters* **546**, 121-126 (2003).

27. Howe,A.K. Regulation of actin-based cell migration by cAMP/PKA. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* **1692**, 159-174 (2004).

28. Michel,J.J.C. & Scott,J.D. AKAP Mediated Signal Transduction. *Annual Review of Pharmacology and Toxicology* **42**, 235-257 (2002).

29. Amieux,P.S., Howe,D.G., Knickerbocker,H., Lee,D.C., Su,T., Laszlo,G.S., Idzerda,R.L., & McKnight,G.S. Increased Basal cAMP-dependent Protein Kinase Activity Inhibits the Formation of Mesoderm-derived Structures in the Developing Mouse Embryo. *J. Biol. Chem.* **277**, 27294-27304 (2002).

30. Shorter,J. & Lindquist,S. Prions as Adaptive Conduits of Memory and Inheritance. *Nat Rev Genet* **6**, 435-450 (2005).

Chapter 3

New strategies and progress towards enhancing the specificity of trans-splicing RNAs in mammalian cells

Abstract

Visualization of endogenous gene expression from native genes in live cells is an important goal for the detection of oncogenic or developmentally regulated mRNAs in live animals. Current methods for visualizing gene expression rely on the recapitulation of endogenous regulation upon introduction of promoter-reporter gene hybrids. These transcriptional reporters often lack the complete regulatory information, or lose precise regulation when introduced at high copy number. Our goal is to directly reprogram endogenous RNAs, expressed solely under the influence of the endogenous promoter and upstream regulatory elements, for the translation of a sensitive *in vivo* reporter gene. Two methods using RNA trans-splicing have been demonstrated for the direct conversion of endogenous mRNAs to detectable signals. Both methods physically cleave the endogenous mRNA at a predetermined site, followed by ligation of a new recombinant exon. The first involves an engineered trans-splicing group I ribozyme from *Tetrahymena thermophila*, which requires no additional cofactors or endogenous proteins for complete activity[1]. The second system, SMaRT (Spliceosome Mediated RNA Trans-splicing) binds an intron-exon splice junction in a pre-mRNA, masking the endogenous consensus splicing sites and re-directing splicing in trans to a tethered recombinant exon[2]. A simple system was developed to monitor the efficiency of trans-splicing in live cells using a red fluorescent protein gene as the target RNA. The highly sensitive reporter gene, β-lactamase[3], was inserted as the cargo RNA for in-frame splicing. Detection of specific trans-splicing was visualized and quantified by conversion of the fluorescent β-lactamase substrate CCF2[4] from green to blue in red fluorescent

cells. Each method of trans-splicing lacked strong activity and specificity in mammalian cells, leading to the design and implementation of several novel schemes for increasing specificity. After significant effort, all strategies of trans-splicing failed to generate reproducible, high efficiency, specific introduction of the reporter gene.

Introduction

Current methods for detection of endogenous mRNA in cells are quite brutal, requiring fixation or injection of fluorescently labeled complementary oligonucleotides. It is unlikely these methods defined for *in situ* hybridization will be readily transferable to a genetically encoded system. Also, without signal amplification and the ability to wash away unbound label, all methods to inject fluorescently oligonucleotides will likely fail to provide sufficient contrast for *in vivo* imaging. Other methods for measuring RNAs indirectly involve transfection of "knock-in" reporter genes, such as GFP[5], luciferase[6], or thymidine kinase[6,7], downstream of a specific promoter. For *in vivo* imaging, these methods require time consuming construction and breeding of transgenic mice. These transgenic methods are essentially useless for clinical use, due to potentially deleterious genetic manipulations. A new method is required to permit mRNA detection and selective marking of cells expressing a specific endogenous mRNA *in vivo*, without direct genetic manipulation of genomic DNA.

The ability to specifically mark cells expressing a pathogenically relevant transcript by a genetically encoded reporter is a stepping stone in developing methods for visualizing gene expression in live animals or even humans. By delivery of a reporter gene to specifically designated cells expressing a specific transcript, it would be possible to generate contrast using various whole animal imaging modalities. After imaging cells specifically expressing a known pathogenically relevant transcript, cytotoxic substrates

could then be delivered locally by the activated reporter. Theoretically, if a ribozyme could be introduced into cells, it would face far fewer problems than other gene therapy methods. The expression needs to only be transient, since the delivery of the substrate and the imaging of reporter activity would only require a short amount of time, and not lifelong expression of an integrated gene. Introduction of a trans-splicing gene into cells should lead to transcription of many active trans-splicing molecules, each capable of converting a single endogenous RNA. Chimeric mRNAs can be translated several times, especially if mRNA stabilization sequences are included on the reporter exon. A good reporter gene can quickly amplify the signal by processing large quantities of substrate. All of this amplification should enable sufficient signal for any number of imaging modalities.

Trans-splicing is a process evolved in many species throughout nature. Nematodes trans-splice a short leader sequence to the start of all coding RNAs[8]. Trans-splicing has been observed for two genes in Drosophila *mod[mgd4]*[9], required for chromatin regulation, and *lola*[10], required for axon guidance. In mammals examples include SV40 transcripts[11] and the rat carnitine octanoyl transferase mRNA[12]. Homotypic trans-splicing has also been observed as a more common phenomenon in genes with long introns[13,14]. The long delays between 5' and 3' splice site transcription in long introns can lead to co-transcriptional trans-splicing from two actively transcribing RNAs from two RNA polymerases on the same gene. This leads to homotypic duplications of exons in the resulting transcript. This demonstrates proximity of a splice donor to a splice acceptor can alone lead to promiscuous trans-splicing. Nature has demonstrated it is possible to trans-splice RNA in cells, but it is unexplored how this phenomenon could be used to detect endogenous mRNAs.

Pioneered by Sullenger and Cech[1], targeted trans-splicing in cells was first achieved using the self-splicing intron from Tetrahymena. The group I cis-splicing ribozyme was split in two pieces, removing the 5' exon. The target RNA is recognized by the six-nucleotide internal guide sequence (IGS) of the ribozyme, catalyzing cleavage and ligation to the cargo 3' exon. Because the ribozyme requires no additional cellular factors for activity, trans-splicing activity easily observed in bacteria to reconstitute full-length β-galactosidase RNA. The chimeric mRNA was successfully translated, as determined by β-galactosidase activity. This milestone encouraged further investigation in mammalian cells to determine if the more complicated RNA regulatory mechanisms inhibit trans-splicing activity.

Following this work, trans-splicing ribozymes was further optimized and used in many situations to repair mutant mRNAs in mammalian cells. By targeting the ribozyme upstream of undesired mutations, the mutant RNA fragment was replaced with the correct sequence and enzymatic function was restored. Several ribozymes were developed to repair therapeutically relevant transcripts, including β-globin[1,15], myotonic dystrophy[16], p16[17], and p53[18]. Successful payloads of greater than 1100 bases were delivered with variable efficiencies between 2-50%. In many cases, the low efficiency of trans-splicing led to some downstream benefit due to the dominant action of the repaired protein[18].

Rational design of trans-splicing ribozymes is not straightforward. RNA secondary structure surrounding the IGS is critical for splicing activity. An unbiased, randomized IGS library (NNNNNG) was developed as clever way to map accessible splice sites, scanning the target RNA for accessible splicing sites *in vitro*[15] or *in vivo*[19]. While optimizing the IGS can lead to increased activity, it does not address the problems of specificity. The six base pair IGS sequence does not allow for enough sequence

discrimination to direct specific, targeted trans-splicing in the milieu of endogenous mRNA transcripts. Using the split LacZ reporter system, Kohler and colleagues systematically extended the P1 and P10 helices to mimic the wild type Tetrahymena ribozyme intron in bacteria[20]. Exploiting the sensitivity of the LacZ assay, additional antisense regions were attached upstream of the P1 helix to enhance interaction with the IGS. Optimal activity was obtained by extending the antisense region greater than thirty bases. With the inclusion of the extended antisense region and optimization of the P1 and P10 helices, activity and specificity were significantly improved. Further studies have demonstrated that antisense-binding regions may also need selection for accessibility. An elaborate selection for optimal antisense binding regions was carried out in yeast using a Gal4-VP16 fusion as the ribozyme cargo and UAS-prototrophy markers for nutrient selection [21]. It was determined that simply increasing the size of the binding region had less affect than finding a smaller, but highly accessible site for antisense binding to the target RNA.

Direct transfection of ribozyme RNA into mammalian cells has been shown to enhance trans-splicing efficiency and is preferred for maximal conversion[15]. When ribozyme RNA is transcribed in cells from plasmid DNA, the efficiency is significantly less, and even further less when using single copy integrant stable cell lines. These observations demonstrate the importance of the ribozyme:substrate ratio. It is also possible that expression from a stably integrated ribozyme gene may diminish splicing activity by interactions with RNA regulatory proteins. To test the activity of cis-splicing in mammalian cells, a Tetrahymena self-slicing ribozyme was inserted at several positions inside the luciferase gene[22]. Splicing efficiency varied depending on the exon context, with less activity closer to the 3' end. Surprisingly, translated protein did not correspond to the levels of functionally spliced mRNAs. Insertion of the cis-splicing intron caused

>100-fold drop in luciferase activity, despite >50% conversion to spliced mRNA.
Messenger RNA transport or mRNA surveillance mechanisms likely trigger the improper
translation of self-splicing ribozymes in mammalian cells, diminishing reporter
expression and activity.

Despite the clear warnings of improper translation of ribozyme-mediated trans-
spliced RNAs, several groups have very recently made significant progress on trans-
splicing ribozyme based delivery of reporter genes for detection of endogenous, or
transfected substrate RNAs. Hasegawa and Rao have demonstrated robust cis and
trans-splicing of β-lactamase targeted to a red fluorescent protein in COS cells[23,24].
Additionally they found a position in the ribozyme capable of tolerating insertions, and
screened several variants to find a new ribozyme with 4-fold higher cis-splicing activity.
A 211 antisense binding region was included to increase targeting, and splicing
efficiency was affected by splice site selection. In work in press, the same group has
designed a breakpoint in the L2 loop of the ribozyme, capable of destroying splicing
activity[25]. When a binding domain of complementary nucleotides is attached to each
fragment, the ribozyme refolds and activity is restored, leading to trans-splicing of the
two cargos. They also show that the two halves of the ribozyme can be brought together
by a bridging sequence on a third RNA to restore activity. Lee and colleagues have also
achieved several of our goals, including cancer specific targeting of endogenously
expressed hTERT RNA, specific delivery of the reporter genes luciferase and HSV-tk,
and inhibition xenograft tumor growth[26]. The authors also claim no there is no evidence
of activation of the PKR-interferon pathway upon transfection of the ribozyme into
multiple cell types.

A second method for RNA trans-splicing, SMaRT, may offer advantages over
ribozyme-based approaches[27,28]. Importantly, SMaRT does not affect RNA localization

or transport, and relies entirely on endogenous splicing machinery. To direct specific trans-splicing, a pre-trans-splicing molecule (PTM) is introduced into mammalian cells expressing a specific pre-mRNA. During the process of pre-mRNA processing, the PTM specifically binds to the 3' end of the target intron, masking the branchpoint and polypyrimidine tract. By tethering an engineered optimal branchpoint sequence and polypyrimidine tract in close proximity to the endogenous splicing machinery, it is possible redirect splicing in trans to an exogenous exon. The resulting chimeric mRNA is efficiently translated without frame loss or inappropriate regulation.

Another advantage of spliceosome-mediated trans-splicing is the ability to replace the 5' end of an endogenous RNA[29], whereas ribozyme methods can only replace the 3' end. The PTM masks the 5' splice donor by antisense binding, and presents a new consensus splice donor. This approach may be less effective, since the 5' end of the first exon is subject to translation and early termination, which could lead to activation of the non-sense-mediated decay pathway. Interestingly, by placing 5' and 3' PTM sequences on either side of an exon, targeted exon replacement can occur, but the efficiency has not been reported[28].

Because the targeting mechanism of SMaRT relies on bases pairing to consensus regions in introns, specificity is quite low. A creative approach to block promiscuous trans-splicing was developed, termed a "safety PTM"[2]. By forming a weak intra-PTM base-paired helix, it was hypothesized the consensus splicing elements upstream of the recombinant exon could be masked to prevent untargeted splicing. Upon recognition of the optimal binding intron, the safety PTM would unravel, revealing the consensus sites for specific targeted trans-splicing. Adoption of this design reduces activity, yet shows impressive gains in specificity. Further rational optimizations have demonstrated the safety PTM strategy is less effective than simply extending the binding region, which has

now been adopted as the standard approach for PTM design[30]. As observed with

ribozymes, the longer the antisense binding region, the higher the specificity. Typical

binding domains reported in recent papers are greater than 300 bp.

Beyond simple tissue culture experiments, the SMaRT method has been used to

repair therapeutically relevant mutant pre-mRNAs in cell culture models, transgenic or

xenograft mice[17,30-34], and following lentiviral delivery[35]. Targeted repair of a conductance

mutant CFTR chloride channel in xenografts exhibited therapeutic levels of mRNA

repair, increasing conductance up to 12-15% of wild type channels[32]. SMaRT activity is

not only detectable by assays directed at phenotype correction, but the protein product

can also be detected by western blot or even coomassie staining of immunoprecipitated

proteins[36]. The rapid progress and successful application of SMaRT in simple and

complex model systems, as well as the ease of design and relatively high RNA

conversion efficiency signifies SMaRT as a more robust approach and a leap forward in

the rational design of trans-splicing RNAs.

Even though the PTMs are capable of redirecting a large portion of pre-mRNA

splicing to trans-splicing, specificity is largely unaddressed on a genome-wide scale.

Specificity has only been assayed by monitoring the conversion of targeted PTMs onto

non-targeted transcripts by RT-PCR, ignoring the complexity and number of pre-mRNAs

present in an individual cell. The only evidence implying the low fidelity of SMaRT in

mammalian cells comes from attempts to repair the RET oncogene[37]. Failed attempts

using safety and regular PTMs prompted the investigation to the true intracellular targets

of the RET PTM. 5' RACE experiments pinpointed dozens of non-specific spliced

targets, despite predicted differences in designed complementary binding regions.

Undoubtedly, without exploration of the binding properties and specificity of SMaRT in

mammalian cells, it is unclear the amount of contrast between targeted and non-targeted splicing.

Current efforts at SMaRT-based RNA reprogramming have focused on RNA repair and design optimization. Clearly the efficiency and specificity are highly dependent on contributions from tethering trans-RNAs in local proximity to an active splicing reaction and from masking cis-splicing elements with antisense regions. The competition between the cis- and the trans-splicing is also important in determining the rate of RNA reprogramming. When the cis-linkage of the 5' and 3' ends of the splicing reaction is separated, it is unclear how much trans-splicing will occur between the fragments. When matching sense and antisense regions are placed at ends of these fragments in the intronic region, splicing is restored, as measured by reconstitution of a split reporter[38]. Therefore, splicing can be performed in trans with endogenous splicing machinery by simply tethering the 5' and 3' splice at sites in close proximity. As a potential application of this observation, a toxin gene was split and fused to two halves of the VEGF intron 6, and an overlapping 50 bp sense/antisense region was attached to the 5' and 3' ends of each component, leaving the natural 5' splice donor and 3' splice acceptor, along with their cis-components intact[38,39]. Toxicity was observed in cells and mice transfected with both components. Despite these promising results, no controls were included to determine the specificity of the reaction, such as removing or mutating the binding regions. This technique has also been used to expand the packaging capacity of gene transfer vectors. Infection of two halves of a gene can be trans-spliced together to reconstitute the full-length transgene[40]. Neither of these examples depends on highly specificity since they only assay for a dominant effect, such as toxicity or rescue of a mutant phenotype. Additionally, neither approach directly targets endogenous RNAs.

Despite these issues, it is clear that trans-splicing can occur by tethering two RNAs in close proximity with intact splice elements.

The first report of live animal imaging of trans-splicing activity was recently reported[41], yet it is far from reaching the goal of imaging endogenous gene expression. Using a split luciferase reporter, xenografts of genetically modified cells expressing a 5' fragment of luciferase linked to an HPV intron and 3' exon were grown in nude mice. The HPV exon was replaced by trans-splicing following transfection of a PTM carrying the 3' fragment of luciferase. Reconstitution of the gene restored 22% of the luciferase signal observed from a cis-splicing construct. Specific luciferase signals of 5-6 fold above background were achieved. Unfortunately, by using a split reporter gene trans-splicing assay, non-specific splicing activity masked. This issue will need to be addressed to attempt imaging of endogenous RNAs by trans-splicing delivery of a reporter gene.

A third trans-splicing approach uses a secondary enzyme to catalyze trans-splicing[42], potentially increasing the specificity and control of the reaction. Expression of the archaeal tRNA processing enzyme from *Methanococcus jannaschii* catalyzes cis- and trans-splicing in mouse cells. The endoribonuclease catalyzes specific cleavage at a BHB helix, the specific motif containing unique bulges in the recognition site. This motif can be rationally designed to accommodate any endogenous transcript. Following cleavage, an unidentified cellular ligase completes the splicing reaction. By requiring an orthogonal enzyme and a unique target motif, one would imagine an increase in specificity. Cis-splicing is highly efficient, capable of converting greater than 65% of RNAs with a low concentration of cargo RNA. However, trans-splicing was far less efficient, converting 3% of an artificial target RNA and less than 1% of an endogenous target RNA. Only 11 base pairs of antisense binding regions were included on either side of the BHB motif, implying further activity may be added by extending the antisense

binding regions. More work is required to explore the specificity and activity of this system, but it is the concept of an orthogonal trans-splicing enzyme may be the best approach for enhancing specificity for high fidelity trans-splicing for molecular imaging.

As thoroughly discussed above, two major properties need to be optimized to improve trans-splicing in mammalian cells: activity and specificity. To accomplish this task, we set out to develop a dynamic fluorescence based screening assay for quantitative, high throughput flow cytometric of large pools of diversified trans-splicing molecules in living cells. Positive and negative selections are required to enrich highly specific events and to eliminate non-specific events. As discussed earlier, it is unlikely major improvements in activity or specificity with single point mutations, so more complicated mutagenesis protocols could be required, such as combining mutagenic PCR, gene shuffling, possibly directly insertion of randomized oligonucleotides.

To meet these requirements, a novel selection strategy has been devised capable of selecting either improved ribozymes or improved PTMs (**Fig. 3.1**). First, a cell line conditionally expressing a target RNA with large dynamic range is required. We have chosen to use a HEK 293 tetracycline inducible dsRed cell line capable modulating dsRed mRNA by the presence or absence of the transcriptional inducer doxycycline. The level of induction is easily monitored by the presence or absence of red fluorescence, allowing for an initial selection for highly inducible clones by flow cytometry. Our initial plan was to create genetically diverse libraries of trans-splicing molecules carrying a recombinant β-lactamase as a reporter. These libraries would be transferred to viral vectors for stable transduction to the target mammalian cells carrying a tetracycline inducible dsRed gene. The extremely high sensitivity of the β-lactamase reporter allows for detection of even the weakest signals, otherwise undetectable with other nondestructive reporters, such as GFP. If trans-splicing and translation of the

dsRed-beta lactamase fusion protein occurs, activity could be measured by conversion

of CCF2 substrate to blue in individual cells. Positive (blue) cells will be sorted and

expanded in the absence of doxycycline to eliminate dsRed mRNA. After several days in

culture, the cells should no longer be red fluorescent, and β-lactamase negative cells

(green) will be sorted then expanded in the presence of doxycycline. By ping-ponging

back and forth between positive and negative selections, it could be possible to select

for optimized trans-splicing RNAs with higher fidelity and activity in mammalian cells.
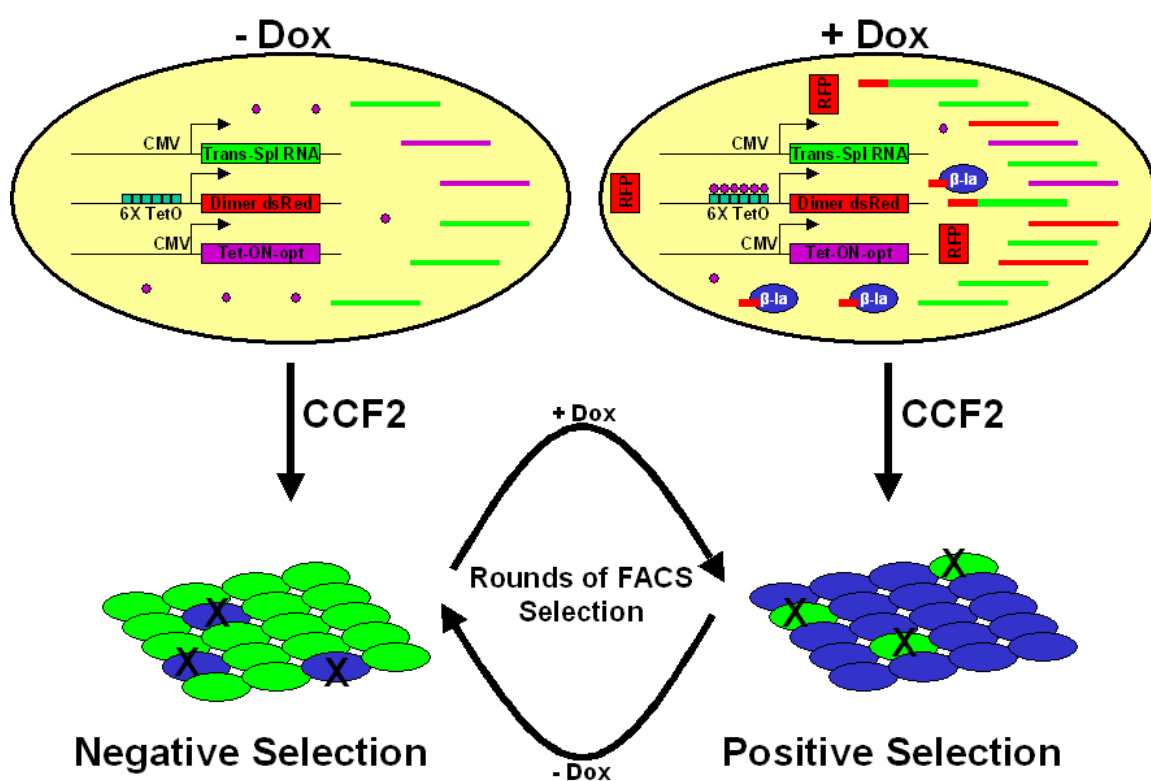
Figure 3.1. Mammalian cell-based libraries for optimizing trans-splicing. Lines represent transcribed RNAs with different colors to note unique genes. Purple represents the tet-ON-opt gene[43], an enhanced version of the reverse tetracycline transactivator. Red represents the Dimer2 fluorescent protein gene. Green represents the trans-splicing RNA gene. Hybrids of these lines represent successful trans-spliced RNAs. Blue represents β-lactamase protein. Doxycycline is abbreviated by "Dox". Blue cells are β-lactamase positive and detected by cleavage of the fluorescent ratiometric substrate CCF2. Green cells do not express β-lactamase, and do not cleave CCF2. Cells with an "X" are eliminated during FACS selection.

In theory, this approach could establish a powerful directed evolution platform for mammalian cell-based optimization of trans-splicing. Unfortunately, as will be discussed in this chapter, our goal to optimize trans-splicing was never reached. Difficulties included low activity and high non-specific splicing. Despite many efforts, novel approaches, and countless generations of improvements, detection of specific β-lactamase activity over background activity was never accomplished.

Results

The activity and specificity of RNA trans-splicing by group I ribozymes in mammalian cells was assessed by generating two rationally designed ribozymes, each targeting a specific uridine in the dsRed mRNA (**Fig. 3.2**). To assay successful splicing, the β-lactamase gene was attached as the 3' exon payload in frame with the predicted dsRed fusion. Β-lactamase is the ideal reporter to assess trans-splicing in living cells, since the ratiometric fluorescent substrate CCF2 is membrane permeable, nondestructive, and highly sensitive[44]. When excited with violet light, positive cells fluoresce blue, and negative cells fluoresce green. Depending on the incubation procedure, the detection limit of β-lactamase is estimated to approach 100 copies of enzyme in a single cell, ~100X more sensitive than GFP[45]. Each ribozyme contains recent optimizations, including extended P1 and P10 helices and 35 bases of extended antisense sequence. The expected trans-splicing product containing the N-terminal fragment of dsRed fused to β-lactamase displayed functional β-lactamase activity, indistinguishable from wild-type β-lactamase, under the conditions tested (data not shown). Each ribozyme was sub-cloned into a modified CMV-based expression vector with deleted sequences between the start of the transcriptional start site and the beginning of the ribozyme. These plasmids were transfected with or without a dsRed expression vector into HEK293 cells

and left for two days. The transfected cells were then incubated for several hours with

CCF2 under conditions for optimal sensitivity and examined on a fluorescent

microscope. Both rationally designed ribozymes failed to express detectable levels of β-

lactamase in the presence or absence of dsRed mRNA. The low activity was

hypothesized to be linked to the accessibility of the IGS, which is required for efficient

ribozyme recognition and splicing.
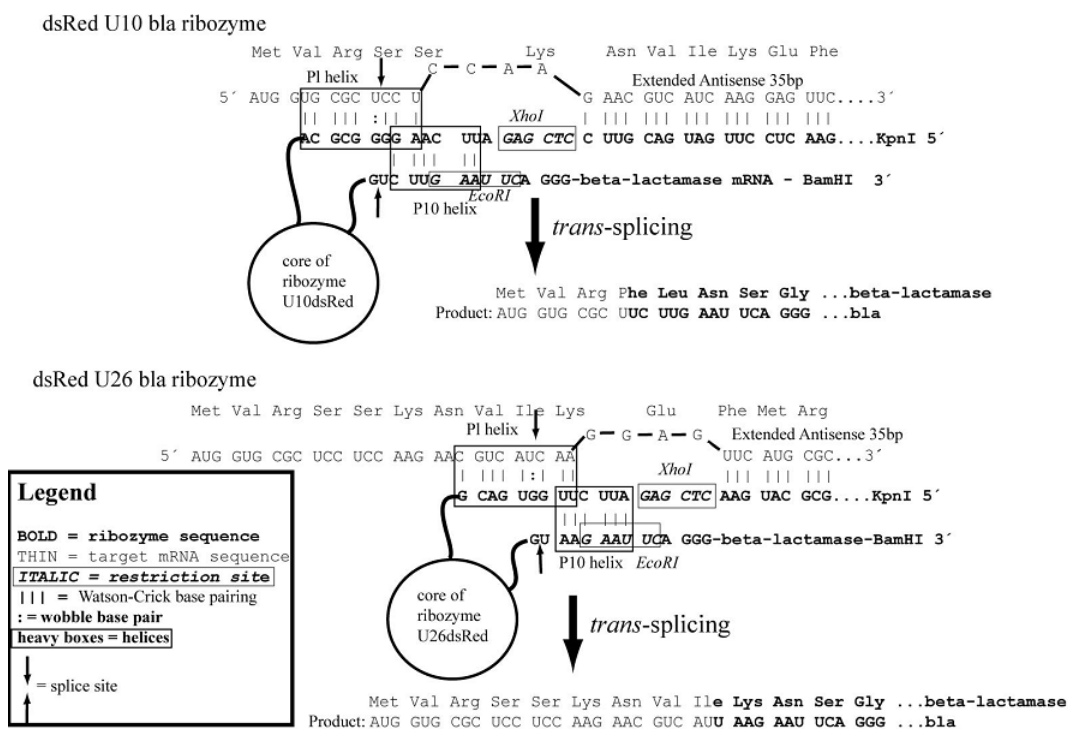
Figure 3.2. Designed dsRed targeted trans-splicing ribozymes. Schematic diagram of both tested dsRed targeted trans-splicing ribozymes and the resulting fusion proteins. Key identifiers are described in the legend provided.

To explore all possible IGS sites, a mapping library strategy was developed for

discovery of optimal IGS sequences in the dsRed mRNA (**Fig. 3.3a-b**). The components

were transcribed *in vitro* to high yields and purity (**Fig. 3.3c**). The two RNAs were then

reacted at 37° C and aliquots were removed at specific time-points for RT-PCR and

analysis by denaturing PAGE (**Fig. 3.4a**). Despite no obvious conversion to products,

the resulting reactions were reverse transcribed with a LacZ specific primer and amplified by PCR (**Fig. 3.4b**). Analysis and sub-cloning of the resulting products showed no specific trans-splicing of 14 of 15 clones, resulting from PCR artifacts generated by non-specific annealing and reverse transcription from weak consensus sites in the target dsRed RNA (**Fig. 3.4c**). Although one sequence was identified: U44 of dsRed Phe15, once reconstituted with an extended P10 helix and an extended antisense region, it failed to transcribe as a full-length product *in vitro*, and was not pursued further.
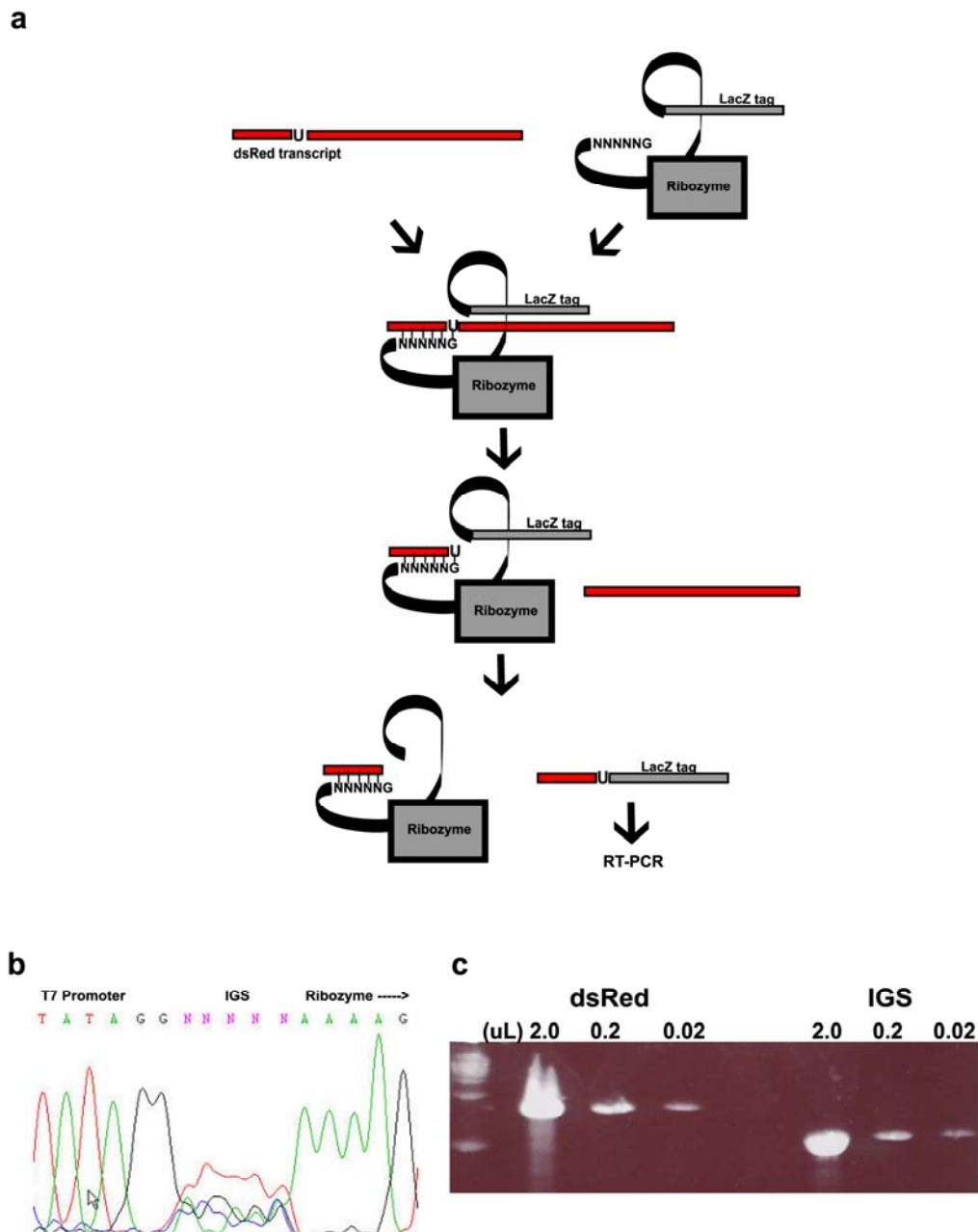
Figure 3.3. DsRed targeted IGS library for *in vitro* IGS mapping. (a) Strategy for *in vitro* mapping of internal guide sequences. A randomized library of ribozymes lacking any extended antisense sequence is reacted with dsRed RNA. Accessible uridines match with members of the IGS library, leading to cleavage, strand transfer, and ligation of the LacZ tag. Accessible uridines are then mapped by sequencing RT-PCR products using 5' dsRed and 3' LacZ tag primers. (b) Sequence results of the cloned IGS library. The sequence chromatograph shows random nucleotides over the five randomized nucleotides. (c) *In vitro* transcription products of dsRed and IGS RNA. Sequences ran as expected, as resolved by denaturing PAGE and ethidium bromide staining. The molecular weight marker resolves two bands, 1 kb and 0.5 kb.

Additionally, *in vitro* reactions using total RNA from several cell lines expressing dsRed RNA showed no products when reacted with *in vitro* transcribed ribozyme (**Fig. 3.5a**). Cells transfected with both the IGS library and dsRed also showed expression of each component (**Fig. 3.5b**), but no specific products following RT-PCR analysis of total cellular RNA (**Fig. 3.5c**). In summary, there are several possible explanations for the lack of detectable specific trans-splicing. First, RT-PCR assay was insensitive or inhibited by the appearance of non-specific products that dominated the amplification during early PCR cycles. Second, the ribozyme may not have been expressed functionally. And third, the reaction conditions could have been poor leading to inefficient ribozyme activity.
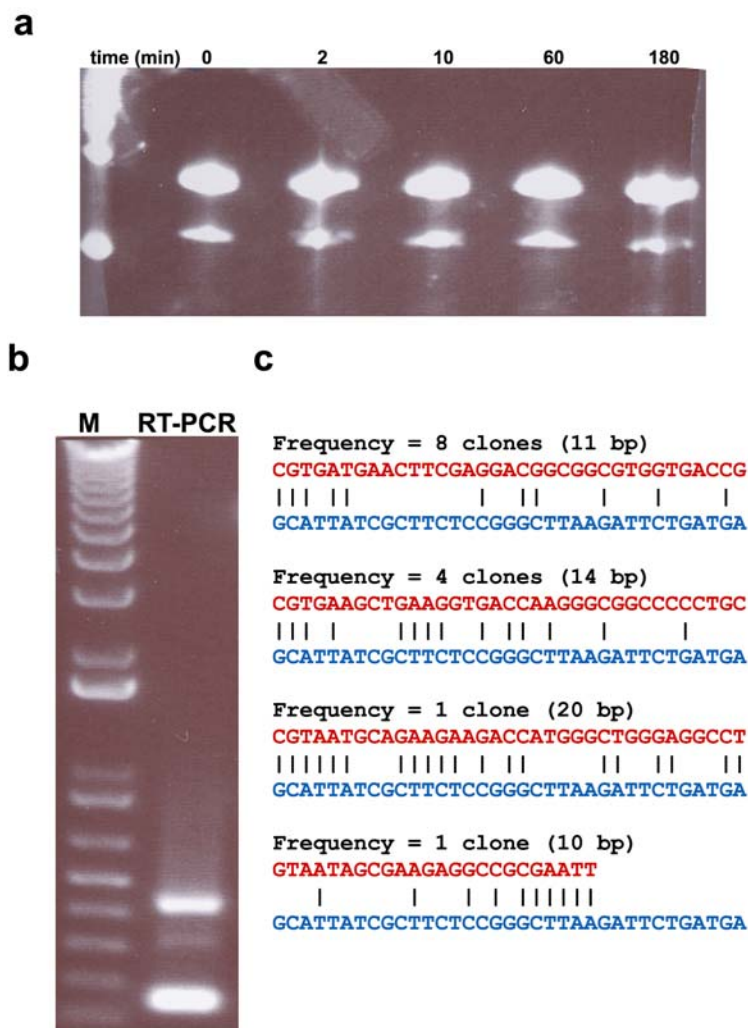
Figure 3.4. *In vitro* trans-splicing targeting dsRed using the IGS library. (a) Timecourse of trans-splicing reaction. Aliquots were removed at specific time-points and analyzed by denaturing PAGE. (b) Products from RT-PCR analysis. Three bands appeared following RT-PCR analysis of the final time-point collected at 180 minutes. (c) RT-PCR products resulted from non-specific primer annealing. The frequency of each product is shown, with the most frequent representing the lowest band, the next most frequent representing the top band, and the other two falling somewhere in the middle.
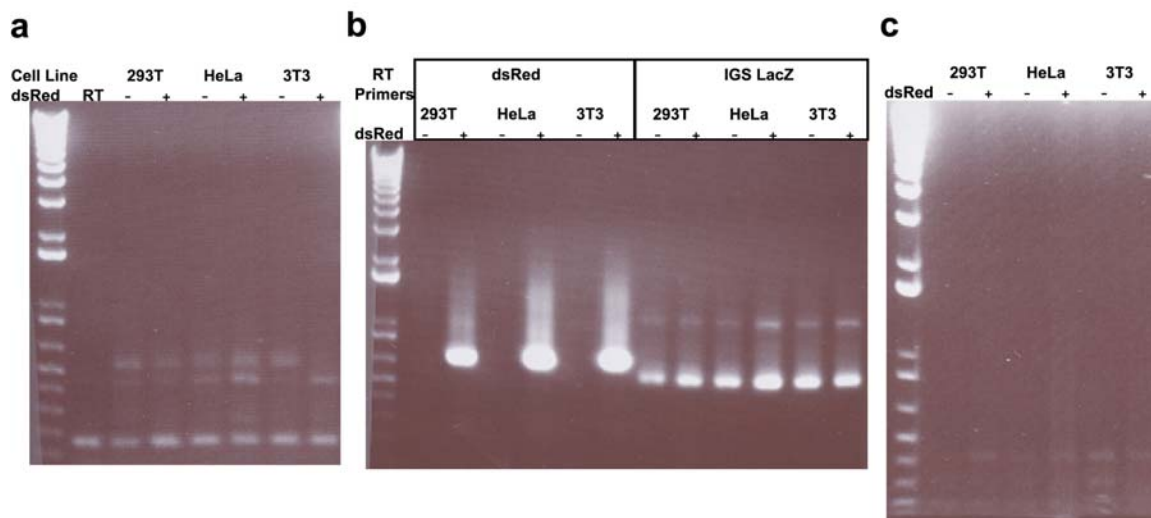
Figure 3.5. Trans-splicing in the context of total cellular RNA. (a) RT-PCR products from the trans-splicing reaction of *in vitro* transcribed IGS mapping library RNA with total cellular RNA *in vitro*. (b) Positive RT-PCR detection of expressed IGS library ribozyme and dsRed transcripts from isolated total RNA from transfected cells. (c) No specific products detected by RT-PCR of total RNA from transfected cells using primers specific for dsRed and the LacZ tag.

In an effort to troubleshoot the lack of detectable trans-splicing, new IGS library ribozyme was synthesized. Unlike before, upon *in vitro* transcription, the ribozyme library underwent self-cleavage, rendering the majority of the reaction mixture catalytically dead (**Fig. 3.6a**). The cleavage is likely due to the high concentration of Mg^{2+} present in high yield commercial *in vitro* transcription kits. Reducing the Mg^{2+} concentration prevented self-cleavage, but the yields are at least two orders of magnitude lower (**Fig. 3.6a**). Interestingly, supplementing the transcription reaction with the ribozyme inhibitor L-argininamide did not inhibit self-splicing. Despite this problem, a reasonable amount of full-length IGS library ribozyme was gel purified and used with dsRed RNA for an *in vitro* IGS mapping trans-splicing reaction. Following the reaction, the products were reverse transcribed with new primer designed to be less prone to aberrant annealing on non-specific sites of the dsRed RNA. Also, the reverse transcription reaction was performed using the more thermostable AMV reverse transcriptase enzyme at 65°C. Unfortunately,

non-specific PCR products dominated the reaction, despite increasing the reaction temperature of both the reverse transcription and the PCR reaction (**Fig. 3.6b**).
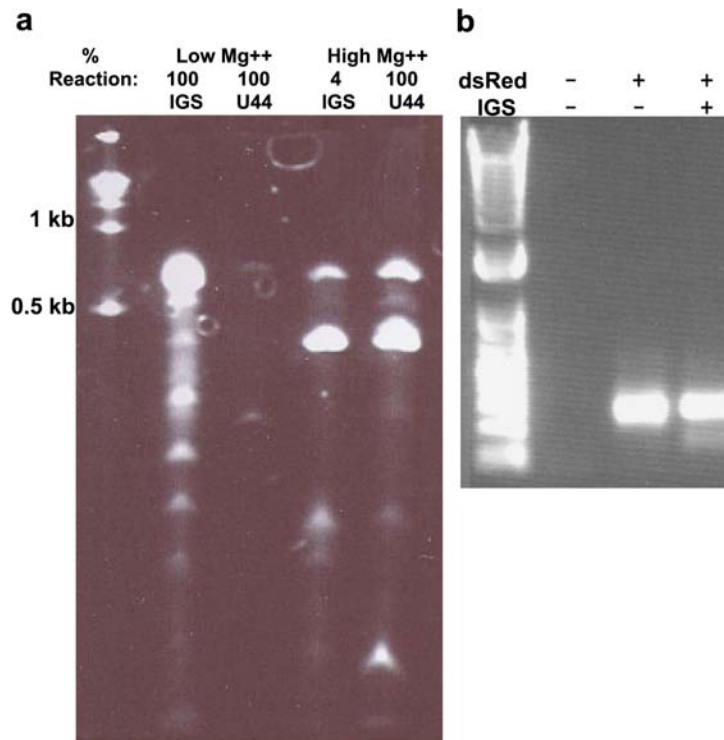
Figure 3.6. In vitro transcription and reaction using newer protocols. (a) Comparison of two buffer systems. The trans-splicing ribozyme protocol uses low Mg^{2+}, whereas the MegaScript kit uses high Mg^{2+} concentrations in its reaction buffer. MegaScript product of the IGS mapping library RNA was gel purified in sufficient quantities for further analysis in a separate preparative gel. The U44 ribozyme could not be gel purified in significant quantities for further analysis. (b) No significant specific products are detectable following reverse transcription at 65^0C and PCR annealing at 55^0C for 36 cycles.

Despite a significant effort, no verified IGS sites were identified due to high non-specific RT-PCR background and low efficiency *in vitro* transcription. Higher temperature reverse transcriptase reactions and several modified primer pairs failed to distinguish non-specific background products from true trans-splicing events. There is no direct evidence the *in vitro* splicing reactions failed, because RT-PCR assays often require optimization to find conditions of selective amplification of specific products. Overall,

implementation of ribozyme-based RNA trans-splicing proved to be less than straight

forward, and a significant effort would be required to establish the technology. Due to

the lack of immediate results, and less than optimistic personal communications with

prominent researchers in the trans-splicing field, we decide to re-direct our efforts to

spliceosome mediated RNA trans-splicing (SMaRT).

SMaRT reprograms cis-splicing to an exogenous pre-trans-spicing molecule (PTM).

The target RNA was generated by inserting a 125 bp modified adenoviral intron an

optimized red fluorescent dimer protein (Dimer2) (**Fig. 3.7a**). Inserting the intron

significantly enhanced the expression of Dimer2 and the spliced product was visible by
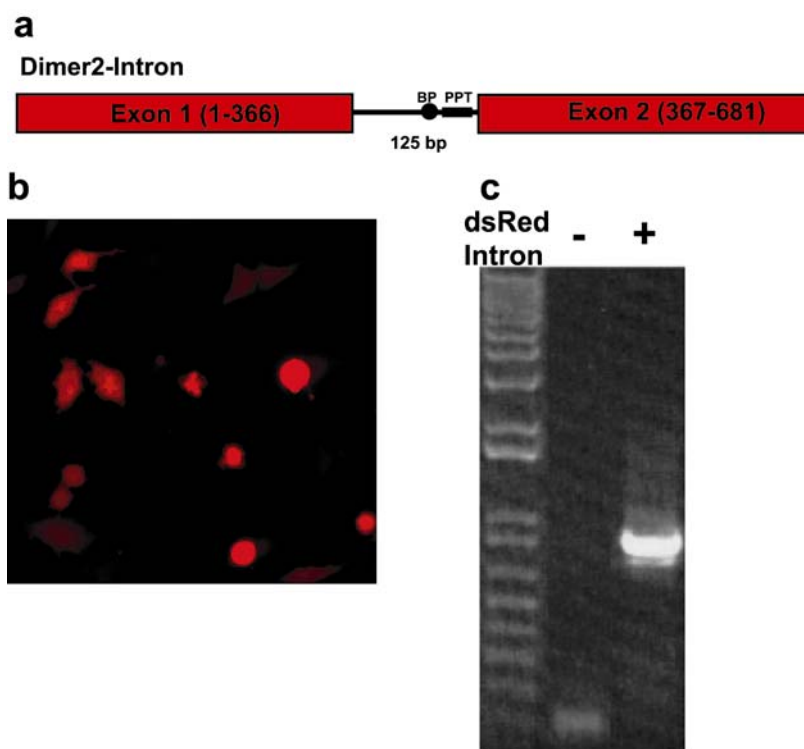
RT-PCR analysis (**Fig. 3.7b-c**).

Figure 3.7. Construction and testing of Dimer2-Intron. (a) Schematic of Dimer2 design, including the branchpoint sequence (BP) and polypyrimidine tract (PPT). (b) Fluorescence of Dimer2-Intron in HeLa cells two days post transfection. (c) RT-PCR analysis of Dimer2-Intron from total RNA of transfected HeLa cells.

To regulate the transcription of Dimer2 RNA, the intron-containing gene was sub-cloned to a tetracycline-inducible recombinant retroviral vector. Introduction of a PEST sequence was essential due to the extremely high stability and long half-life of the Dimer2 protein in cells (**Fig. 3.8a**). Virus was generated and HEK293 cells were infected and sorted for brightness in the presence or dimness in the absence of the transcriptional inducer doxycycline. Following selection, RNA was isolated and analyzed by RT-PCR (**Fig. 3.8b-c**). Due to the viral RNA intermediate generated by using retrovirus mediated gene delivery, the intron was spliced out in the final inducible cells. This example demonstrates one pitfall of using retroviral libraries for the optimization of reactive RNAs. If splicing occurs in the viral packaging cells, the resulting virions will not contain the same gene as the vector DNA. During viral production, the RNA undergoes a preliminary selection against spontaneous activity in the cellular environment of the packaging cells. One method to avoid this is to reverse the orientation of the promoter and trans-splicing gene in the viral vector. This would allow full-length viral RNA transcription from the 5' UTR promoter without cis-interference from an active trans-splicing RNA. Because of these problems, transient transfection of Dimer2 plasmid was used in all experiments, although FACS sorting of transiently transfected cells over an extended time period could have also generated an inducible stable cell line.
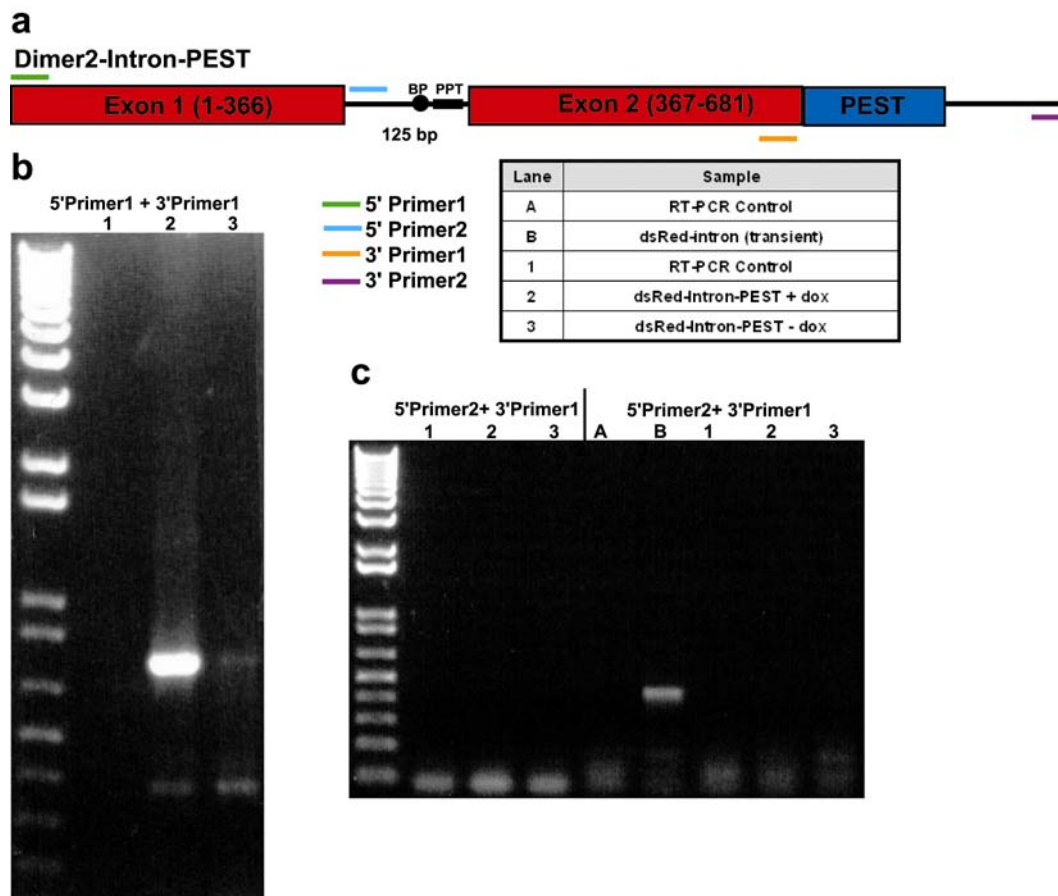
Figure 3.8. Virus-transduced Dimer2-intron-PEST cells have no intron. (a) Schematic of Dimer2-Intron-PEST and the primers used for amplification. (b) RT-PCR of Dimer2 from total cellular RNA shows high doxycycline induced expression, but no unspliced RNA. (c) Absence of any intron in virally transduced cells. Transiently transfected cells contain the intron, but virally derived cell lines do not.

The Dimer2-intron was targeted for trans-splicing by designing a specific PTM. This RNA contained 135 bp of antisense to mask the cis-elements essential for the 3' splice site selection as well as a section of the intron-exon junction (**Fig. 3.9a**). An epitope-tagged cytosolic β-lactamase gene with the methionine initiation codon removed was added to the 3' end of the PTM in frame for proper 3' exon replacement. The PTM was designed to carrying restriction sites for modular replacement of the spacer, binding domain, or cargo with only minor adjustments from functional published designs (**Fig. 3.9b**).
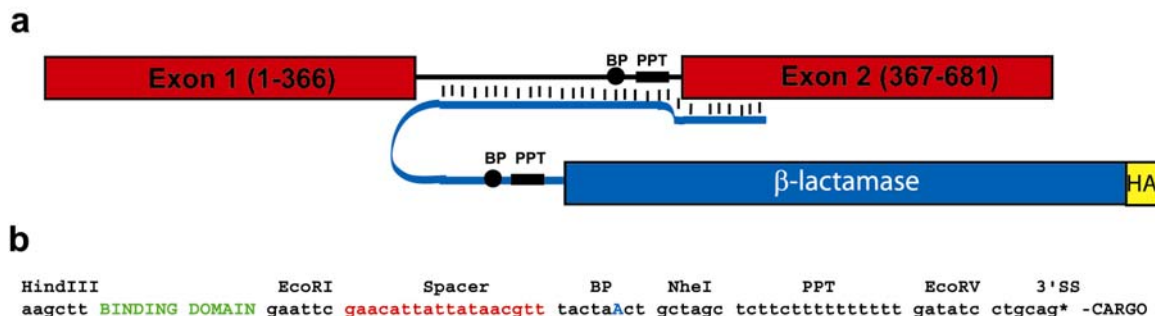
Figure 3.9. Diagram of spliceosome mediated trans-splicing targeting Dimer2-Intron. (a) Schematic diagram showing tethering of full-length β-lactamase PTM to the intron and 3' exon. (b) Nucleotide map of PTM components and modular cloning strategy.

HeLa cells were co-transfected with the Dimer2-intron and the β-lactamase-HA PTM for two days, then assayed for β-lactamase activity. Strong β-lactamase activity was present independent of Dimer2 expression, signifying low specificity (**Fig. 3.10a**). Total cellular RNA was isolated and analyzed for specific products by RT-PCR using primers from the 5' end of Dimer2 and a reverse primer for the HA epitope. A specific trans-splicing product was present in the Dimer2-PTM co-transfection, indicating the occurrence of specific trans-splicing in cells (**Fig. 3.10b**). To determine the nature of the non-specific products, and the relative ratio of specific to non-specific trans-splicing, cell lysates from transfected cells were probed by western blot with an anti-HA antibody. No products, either specific or non-specific, were detectable by western blot despite the high β-lactamase activity visualized by microscopy (**Fig. 3.10c**). Clearly, the sensitivity of the β-lactamase reporter is significantly greater than the sensitivity of a western blot, since no bands were present from any cells except those directly transfected with β-lactamase-HA protein. Conversely, the direct transfection of the β-lactamase-HA gene is extremely well expressed, which could saturate the western blot signal and mask weak signals from trans-spliced protein. Nevertheless, the efficiency of SMaRT is significantly

lower than reported elsewhere. This low activity and non-specificity may explain why

SMaRT is advertised as a powerful method for RNA repair of genes whose restored

activity is dominant.
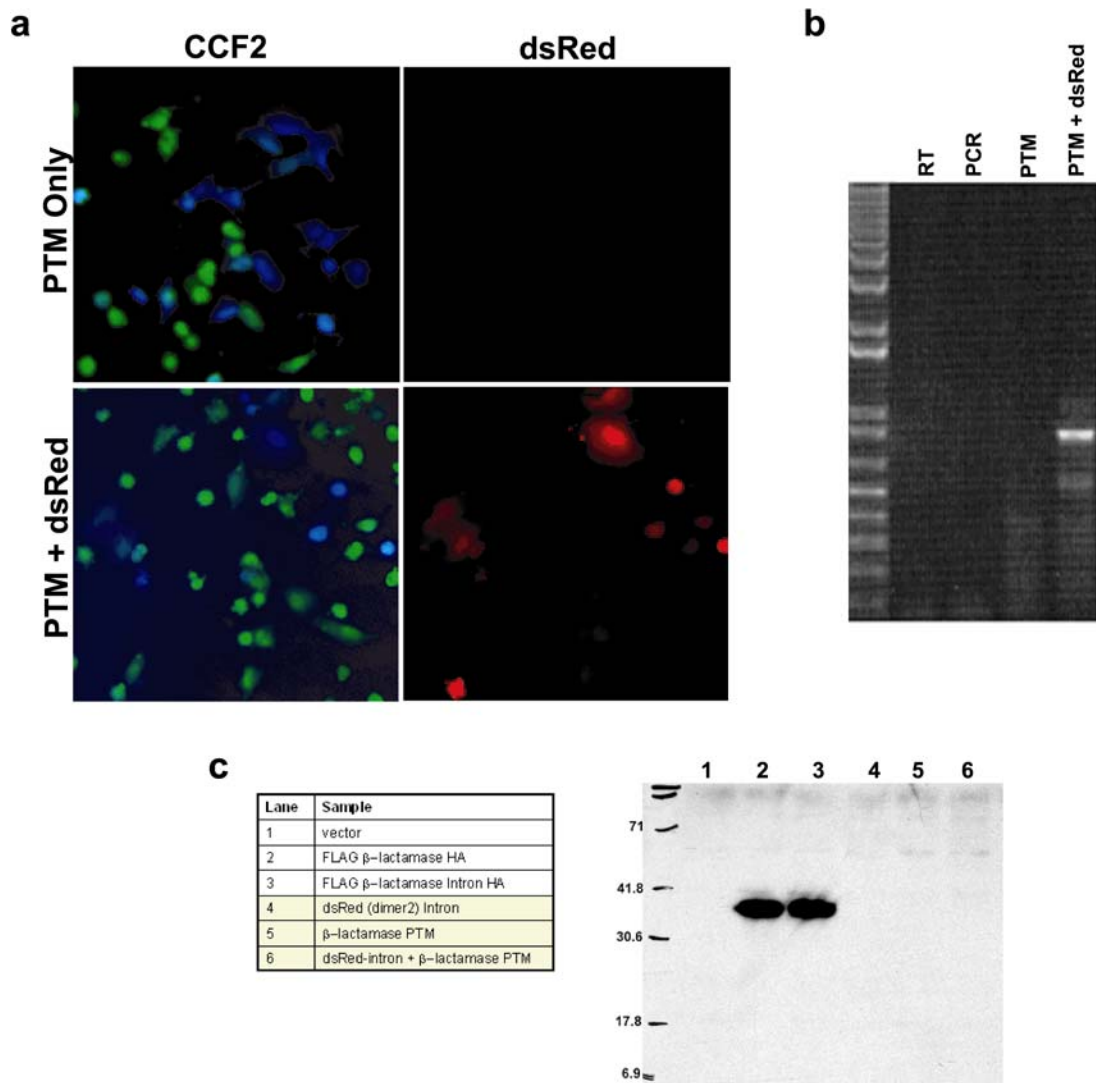
Figure 3.10. Analysis of a Dimer2-intron targeted PTM in HeLa cells. (a) Fluorescence analysis of β-lactamase activity using the CCF2 substrate and correlated red fluorescence from Dimer2 expression. Note that blue cells are present in the absence of Dimer2-Intron transfection. (b) RT-PCR analysis shows the presence of the desired trans-spliced product. (c) Western blot of total protein from transfected cells blotted with αHA antibody. No non-specific or specific trans-spliced products were visible. Only directly transfected β-lactamase or β-lactamase-Intron were detectable.

The driving force behind our efforts to develop specific trans-splicing RNAs has been to develop a method to specifically deliver reporter genes for imaging of gene expression. Obviously, the promiscuity of spliceosome-mediated trans-splicing in its present form will not allow us to achieve this goal. Therefore, a new methodology was envisioned as an approach to dampen non-specific trans-splicing. As shown above, specific 3 ' exon replacement (3'ER) is achievable in cells and reports have also demonstrated 5' exon replacement (5'ER)[29]. If both 3'ER and 5'ER could be targeted to the same intron, then simply by proximity, trans-splicing could occur between the two PTMs. Therefore, we designed 5'ER and 3'ER PTMs, each carrying half of the β-lactamase gene. Upon binding the splice junctions of the Dimer2-intron RNA, the two PTMs could be trans-spliced together to generate an mRNA encoding the full length β-lactamase gene (**Fig. 3.11a**). Most importantly, by requiring two recognition events, the specificity of trans-splicing should be enhanced. Along these lines, it could also be possible to direct trans-splicing in an intron-independent manner by tethering each PTM to nearby antisense regions (**Fig. 3.11b**). By avoiding splice junction targeting, this method would also avoid competition between cis- and trans-splicing and could increase both activity and specificity.
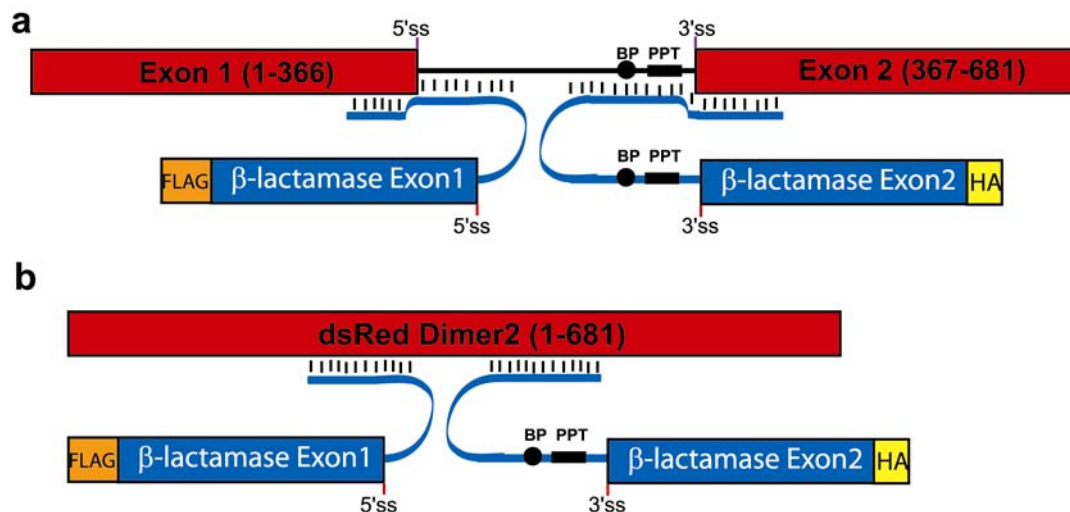
Figure 3.11. Trans-splicing by tethering β-lactamase gene fragments. (a) Intron-dependent tethering strategy. 5'ER and 3'ER PTMs are targeted to each end of an intron to induce trans-splicing between PTMs. (b) Intron-independent trans-splicing strategy. Trans-splicing is relies on nearby tethering of two PTMs presenting active 5' and 3' exons.

To test 5'ER, the adenoviral intron was transferred from the Dimer2 plasmid to three locations (BLA314, BLA363, and BLA514) in a doubly epitope tagged FLAG-β-lactamase-HA gene (**Fig. 3.12a-c**). The first two positions (314 and 363) were chosen due to their proximity to unique restriction sites for cloning purposes. BLA514 should have minimal spontaneous β-lactamase, since the intron was inserted exactly at the amino acid position determined by others to prevent protein fragment complementation[46,47]. This is quite relevant for 5'ER, since the PTM should get translated until it reaches stop codons in the intron. If expression of the N-terminal fragment of β-lactamase was sufficient for enzymatic activity, then the reporter assay would be corrupted and specific splicing would be difficult to detect. All three insertions sites demonstrated robust β-lactamase activity (**Fig. 3.12d**).
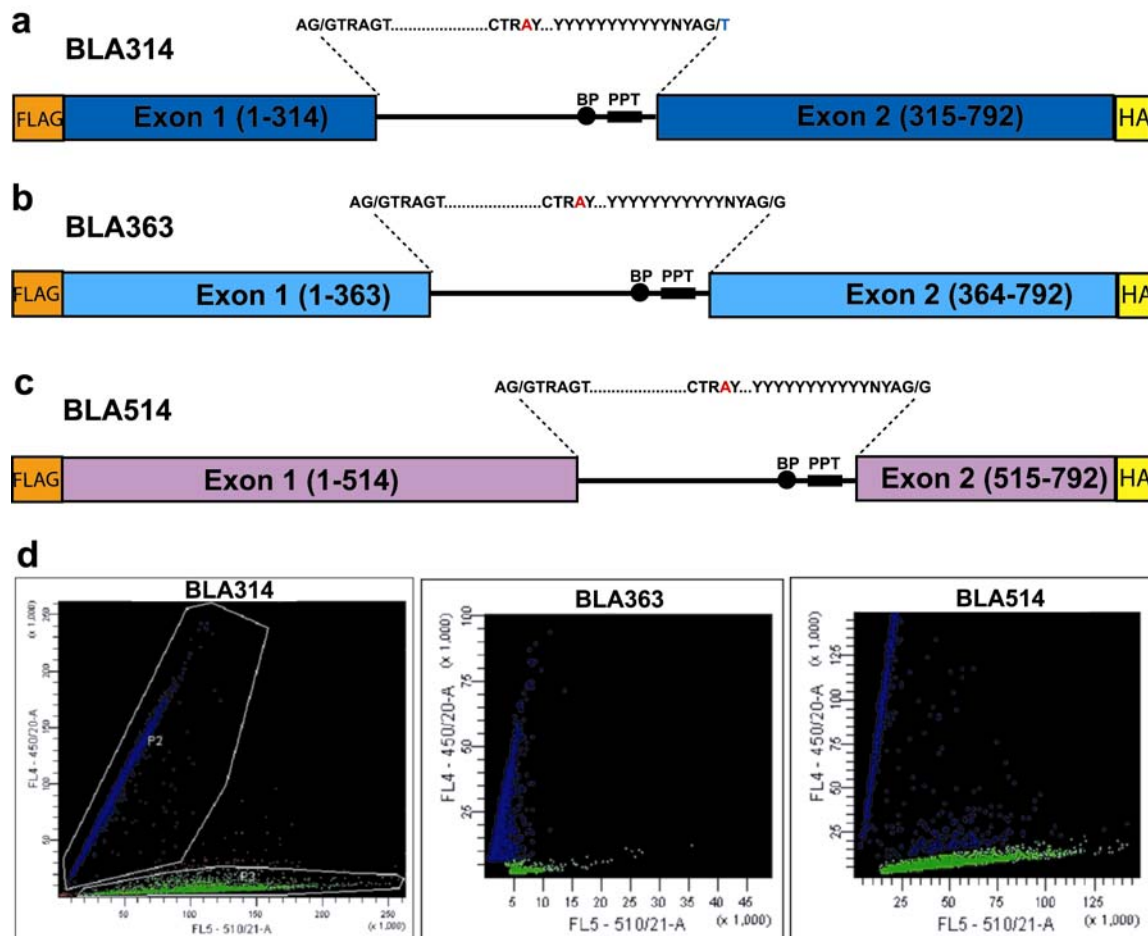
Figure 3.12. Testing β-lactamase intron insertions at three positions. (a) Schematic diagram of cloning strategy highlighting important conserved splicing elements for BLA314, BLA363, and BLA514. (b) Flow cytometry analysis of CCF2 loaded cells for β-lactamase activity. Cells co-transfected with β-lactamase and Dimer2. Dimer2 positive cells shown. Each insertion sites leads to efficient reporter activity in HeLa cells. Experiments performed on separate days with varying loading efficiencies and instrument settings.

To test whether protein complementation could be a contributing factor to any non-specific background, each pair of exons from β-lactamase was expressed in HeLa cells, without any intron fragments or splice sites. BLA314 and BLA363 showed no protein complementation, whereas the exons designed according to the reported anti-protein complementation breakpoint[46,47] in BLA514 showed high β-lactamase activity (**Fig. 3.13a**). Further analysis revealed that only protein from BLA514 exons was expressed in cells, whereas those from BLA314 and BLA363 were undetectable by western blot (**Fig.**

3.13b). Therefore, spurious activity of trans-splicing using 314 and 363 splice sites is not

due to protein complementation, and must be from either non-specific RNA or DNA

events.
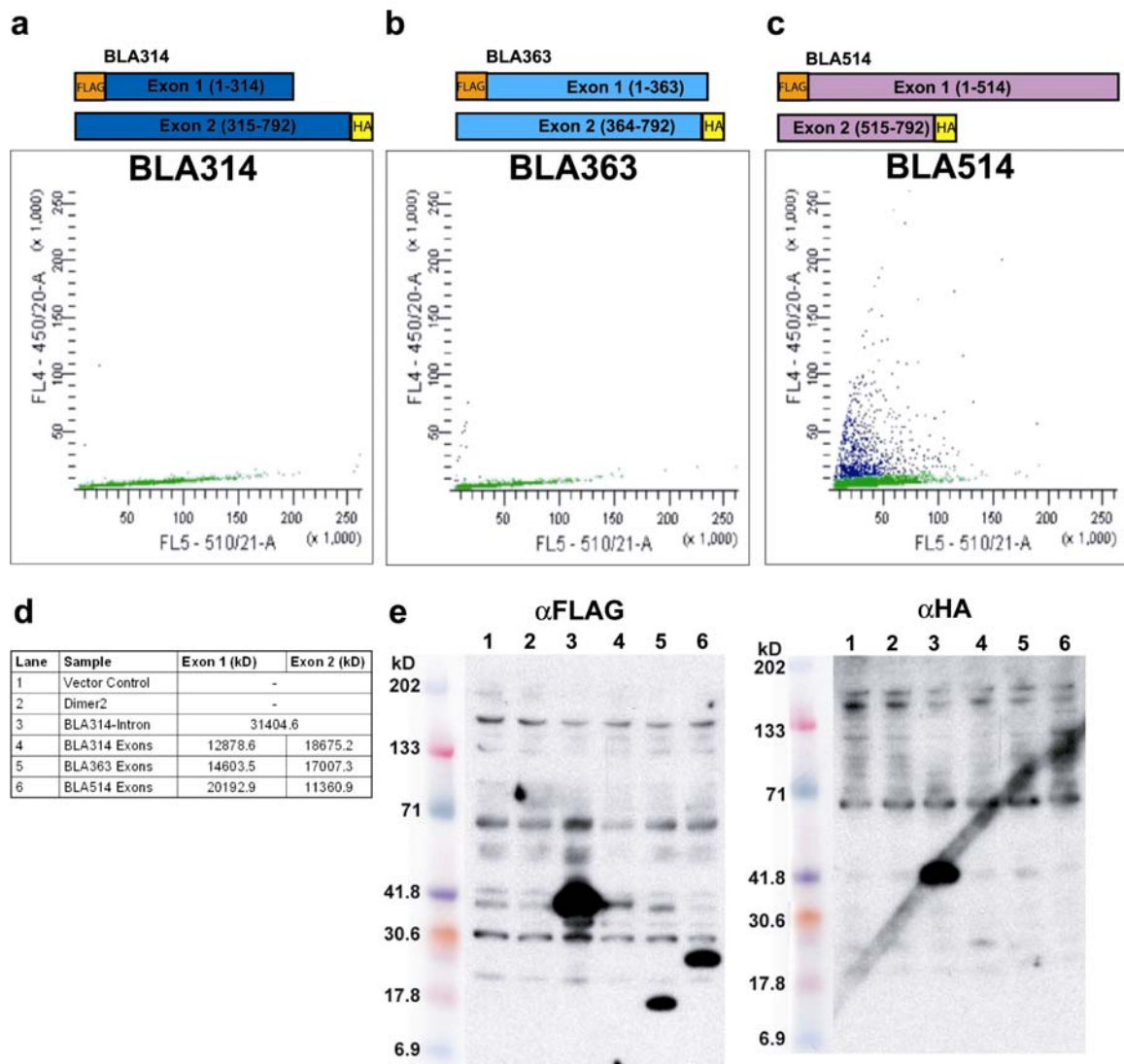
Figure 3.13. Detection of β-lactamase fragment expression in cells. (a-c) Diagram of β-lactamase exon breakpoints and their activity following individual expression of each in 293T cells. Only BLA514 shows any activity. (d) Predicted fragment sizes and lane designations for western blot. (e) Western blot analysis shows only BLA514 fragments are present and could act to restore β-lactamase activity. BLA314 and BLA363 do not lead to any protein expression or β-lactamase activity.

Next, 3'ER was tested by replacing the 5' exon in the Dimer2-intron gene with the

5' exon of β-lactamase from BLA314, BLA363, and BLA514 (**Fig. 3.14**). This created a

fusion protein containing the first exon of β-lactamase, followed by the intron and the

second exon of Dimer2 and named BID. PTMs were generated to match each BID, and

then transfected into cells then analyzed for activity after 2 days by flow cytometry.
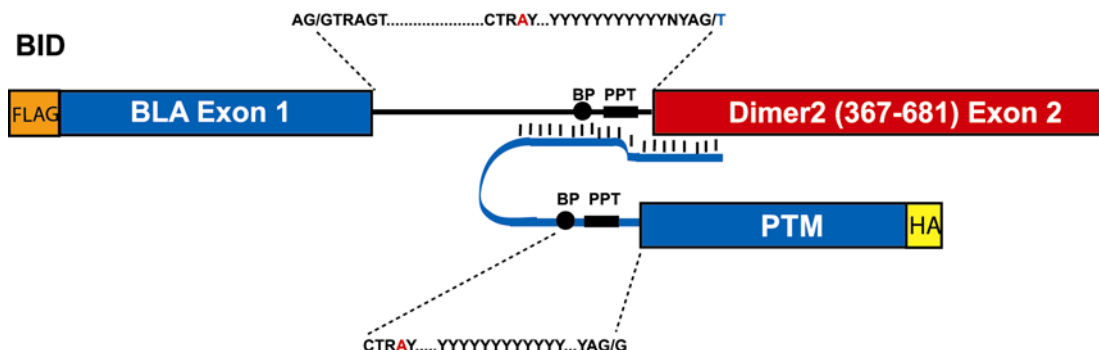
Figure 3.14. Schematic of the 3'ER split β-lactamase reporter. Consensus splice elements are shown, including the BP and PPT sequences. Three versions of 3'ER were constructed for each intron insertion site. The PTM contains the second half of the β-lactamase gene.

During preliminary experiments using the mammalian expression vector

pCDNA3, PTMs carrying partial β-lactamase fragments incapable could alone lead to

intense β-lactamase activity (**Fig. 3.15a**). The likely mechanism accounting for this is

recombination in bacteria or during the transfection between the β-lactamase in the

vector backbone and the PTM cargo, since there was no observed β-lactamase activity

in cells transfected with the vector alone. To correct this problem, the ampicillin gene

used for bacterial drug selection was replaced with a kanamycin resistance gene (APH-

IIIa). This was sufficient to prevent spurious β-lactamase expression in transfected cells

(**Fig. 3.16**), but did not correct the non-specific background seen in the full-length β-

lactamase PTM (**Fig. 3.10**, data shown from APH-IIIa vector containing plasmids).

Again, the high sensitivity of the β-lactamase reporter system is sensitive enough to

detect low frequency events that would otherwise be undetectable.

	Blue
Vector Control	0.3%
β-lactamase-Intron	11.4%
BID314	10.9%
PTM314	5.6%
BID314 + PTM314	13.5%
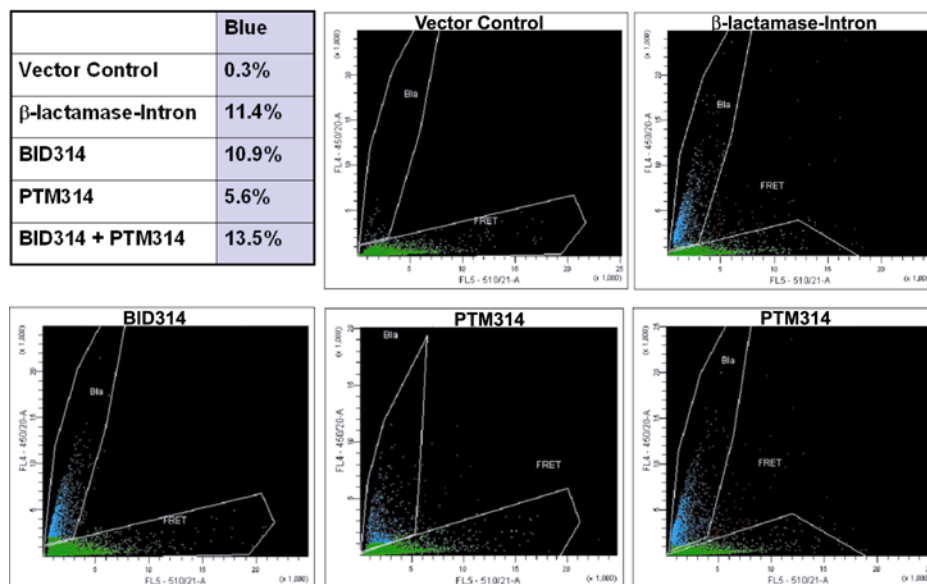
Figure 3.15. Spontaneous β-lactamase activity in 3'ER HeLa cells. Following 413 nm excitation, blue cells display a high ratio of 450/20 nm to 510/21 nm fluorescence and green cells display a low 450/20 nm to 510/21 nm ratio.

Once the vector was corrected, BID-dependent activity was seen for PTM314 (**Fig. 3.16**), PTM363, and PTM514. Next, total cellular RNA was isolated from the BID-PTM314 for RT-PCR analysis (**Fig. 3.17a-d**).

	Blue
Vector Control	0.0%
β-lactamase-Intron	19.3%
BID314	0.0%
PTM314	0.0%
BID314 + PTM314	9.5%
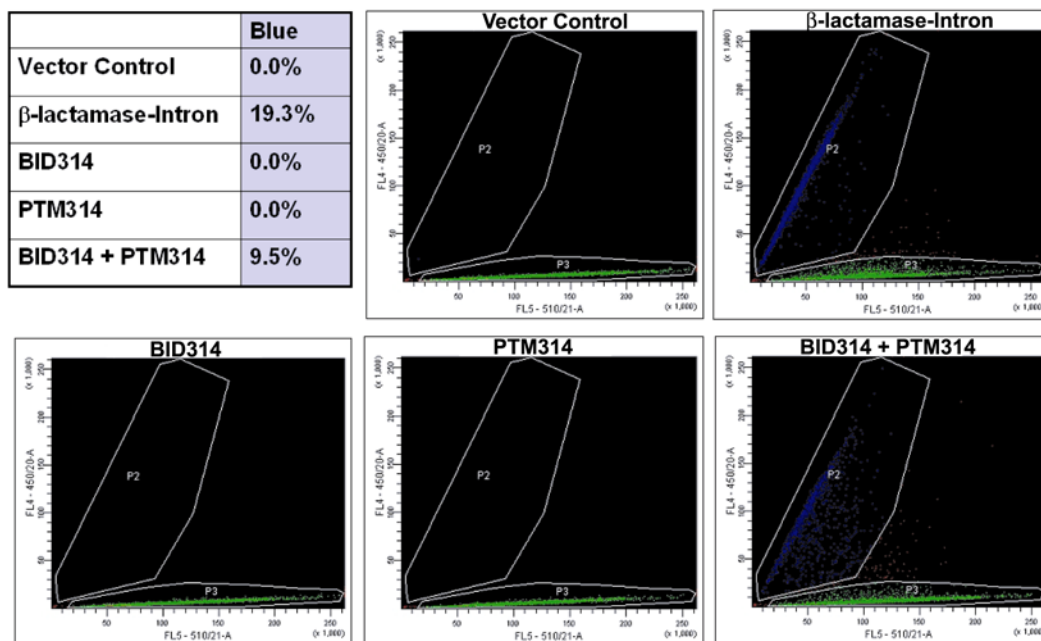
Figure 3.16. Activity of 3'ER in cells. Controls for expression of each component individually show no spurious activity. Only expression of BID314 and PTM314 reconstitute β-lactamase activity.
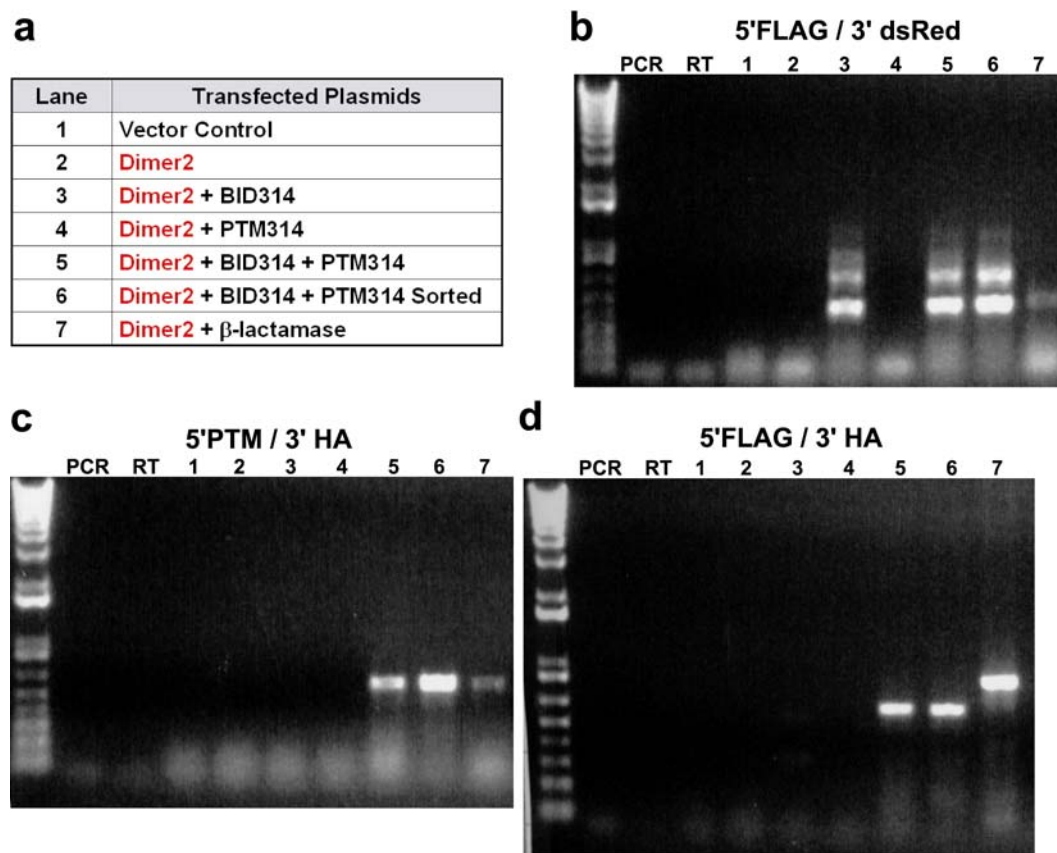
Figure 3.17. RT-PCR analysis of 3'ER from transfected 293T cells. (a) Lane legend. "Sorted" sample was enriched for β-lactamase positive cells by FACS. (b) Expression of BID314 verified by RT-PCR with a 5' primer for the FLAG epitope tag and a 3' primer specific to Dimer2 dsRed. (c) Expression of PTM314 verified in cells by RT-PCR using primers for the 5' end of the PTM and the 3' end of the HA epitope tag of the PTM cargo. (d) Detection of trans-spliced products by RT-PCR using primers specific for 5' FLAG and 3' HA epitope tag sequences.

Interestingly, despite strong β-lactamase activity, no specific full-length β-lactamase RNA was isolated in the co-transfected cells. Instead, a 650 bp fragment was amplified in both BLA314 and the PTM314-BID314 co-transfection, implying a cryptic spice site or non-specific annealing during the RT-PCR. Rather than immediately troubleshoot the RT-PCR results, cell lysates were collected and a western blot was performed to find trans-spliced products. Expression of BID was confirmed by probing the blots with αFLAG, but no specific expression of trans-spliced products were apparent by αHA blotting (**Fig. 3.18**). In summary, trans-splicing was detected by β-lactamase

activity, and possibly by the appearance of a smaller ~650 bp fragment by RT-PCR, but no protein was detected in lysates by western blot.
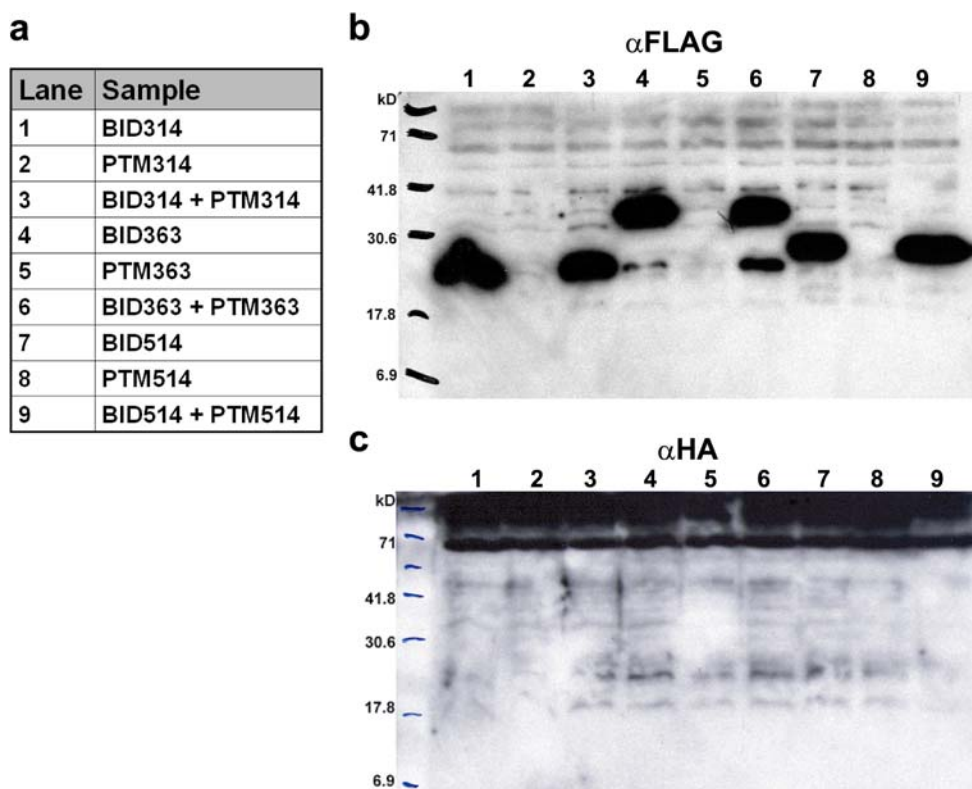
Figure 3.18. No specific trans-splicing is detectable by western blotting. (a) Legend describing lane markers. (b) Anti-FLAG western detects expression of BID. (c) Anti-HA western does not show any PTM specific products, but does show high MW non-specific reactive proteins.

Next, PTMs were developed to target a fusion of the first exon of Dimer2 fused to the intron and the second exon of β-lactamase, called DIB (**Fig. 3.19a**). These PTMs were designed to mask the 5' splice donor site of the DIB target RNA to re-direct splicing in trans to generate a full-length β-lactamase mRNA. Cells were transfected and later analyzed by flow cytometry for β-lactamase activity. All three PTM versions (PTM314, PTM363, and PTM514) showed β-lactamase activity dependent on DIB co-transfection, although the number of cells with saturating β-lactamase activity was lower as seen by the increased spread in the population (**Fig. 3.19b**). Next, the cells were analyzed by

RT-PCR for generation of full-length doubly epitope tagged β-lactamase mRNA (**Fig. 3.19c**). Again, despite β-lactamase activity, no full-length β-lactamase RNA was detectable, and BLA314 and BLA363 showed smaller fragments of ~650 bp and ~500 bp respectively, confirming the previously noted deficiencies. Western blot analysis probing with αHA showed no expression of DIB314 or DIB514, but strong expression of DIB363 (**Fig. 3.19d**). There was no appearance of new bands reactive to αHA dependent on the co-transfection of the PTM. Blotting with αFLAG showed some PTM dependent expression for PTM514, but again this activity was not dependent on DIB514. As was seen with 3'ER, 5'ER did generate β-lactamase activity dependent on the PTM, but spliced products were not detectable by RT-PCR or by western blot.
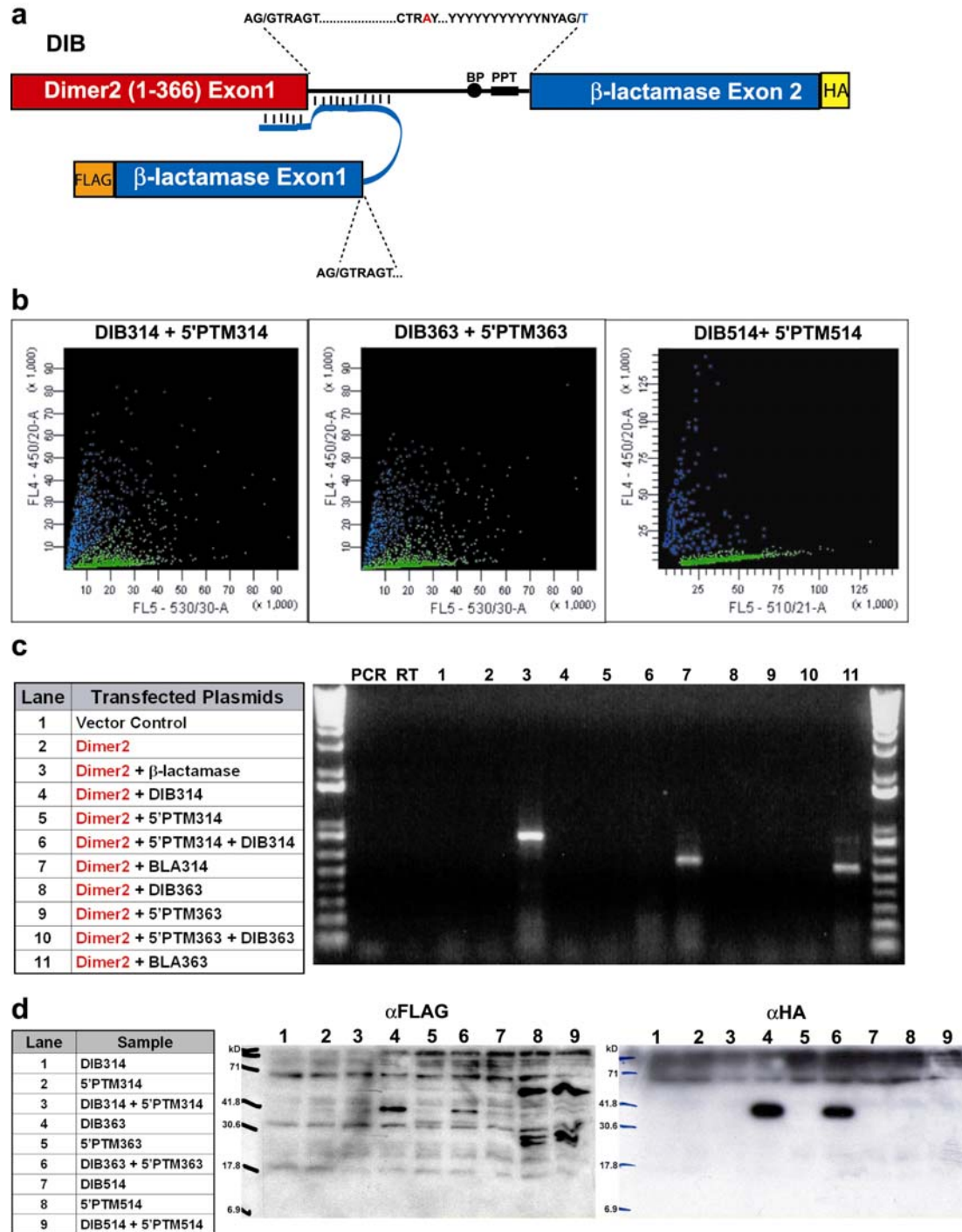
Figure 3.19. Design and testing of split β-lactamase reporters for 5'ER. (a) Schematic diagram of split reporter strategy and PTM design to mask 5' splicing elements. (b) Activity is detectable for 5'ER of the split b-lactamase reporter. Each individual component showed no activity in the absence of the other β-lactamase component. (c) RT-PCR analysis using 5'FLAG and 3'HA primers to amplify trans-splicing products. No trans-splicing was detectable, despite apparent activity. BLA314 and BLA363 show shortened products, implying a cryptic splice site.

Finally, because β-lactamase activity was achieved in 5'ER and 3'ER, and because the vectors were already generated, the double trans-splicing approach was attempted (**Fig. 3.11a**). Deletions were made to the PTM binding domains to prevent competition and provide unique binding sites for each PTM. Each PTM was transfected with its cognate PTM in the presence or absence of the Dimer2-intron co-transfection. In all experiments, β-lactamase activity was present in cells lacking Dimer2-intron expression (**Fig. 3.20a-b**). In one set of preliminary experiments, co-transfection of Dimer2 increased β-lactamase activity 2 to 3-fold, yet these results were not reproducible (**Fig. 3.20a-b**). RT-PCR analysis of the cells transfected with both PTMs along with Dimer2-intron showed expression of each transfected component, but no specific trans-splicing, even after another round of nested PCR. Additionally, since the double trans-splicing reaction was competing with cis-splicing, trans-splicing from the 5'ER PTM to the 3' exon as well as trans-splicing from the 5' exon to the 3'ER PTM was tested, but showed no activity.

a

Sample	Experiment 1			Experiment 2		
% Positive for X	β-lactamase	Dimer2	Dimer2 AND β-lactamase	β-lactamase	Dimer2	Dimer2 AND β-lactamase
5'PTM363 + PTM363	5.9	-	-	11.4	-	-
5'PTM363 + PTM363 + Dimer2-Intron	18.5	33.8	37.1	9.4	53.5	15.7

b

Lane	Sample
1	Vector Control
2	Dimer2-Intron
3	Dimer2-Intron + β-lactamase
4	5'PTM363 + PTM363
5	Dimer2-Intron + 5'PTM363 + PTM363
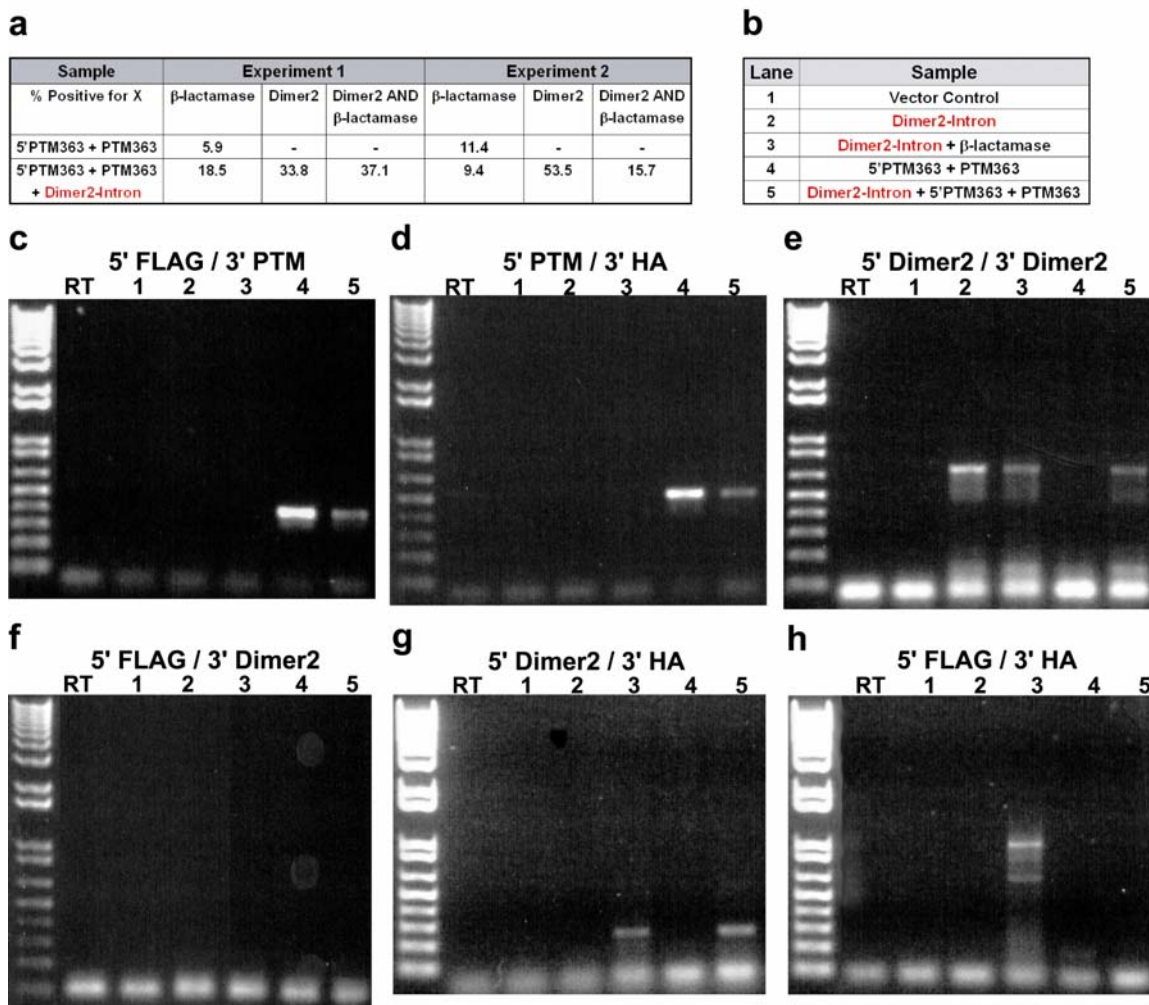
Figure 3.20. Double trans-splicing generates background activity. (a) Summary of flow cytometry analysis of β-lactamase activity from two experiments. Experiment 1 shows enhancement of activity upon transfection of the Dimer2-intron target. Experiment 2 shows now enhancement, and was repeated with similar results several times. (b) Lane assignments for RT-PCR analysis. (c) Detection of 5'PTM363. (d) Detection of PTM363. (e) Detection of Dimer2-Intron. (f) Detection of 5'ER from 5'PTM363 to the Dimer2-intron target RNA. (g) Detection of 3'ER from PTM363 to the Dimer2-Intron target RNA. (h) Detection of full-length β-lactamase, the expected product of PTM to PTM trans-splicing.

To test whether tethering alone is sufficient to mediate trans-splicing, the BLA314 or BLA363 genes were split in the middle of the intron and cloned into separate expression vectors, one containing the first exon and half of the intron (5'STS) and the second containing a 50 bp binding domain of either sense or antisense sequence followed by the second half of the intron and the second β-lactamase exon (3'STS) (**Fig. 3.21a**). This

method has since been published and called segmental trans-splicing, or STS. This experiment should help understand the nature of the non-specific background. If transfection of two halves of the β-lactamase-intron gene with no targeting generates significant activity, then trans-splicing background is likely inherent, either caused by DNA recombination of the vectors upon co-transfection, or by extremely active trans-splicing of transcripts containing free splice sites. Both halves of the β-lactamase-intron gene were transfected into cells and analyzed by flow cytometry for trans-splicing activity (**Fig. 3.21b**). Non-targeted β-lactamase activity was high, capable of saturating CCF2 cleavage in 24-31% of transfected cells. Upon targeting with 50 bp of antisense sequence, the fraction of saturated cells rose to 45-56%, or approximately double. A two-fold increase is unlikely to provide enough dynamic range for any useful applications, suggesting antisense targeting is not strong enough alone to impart specificity.
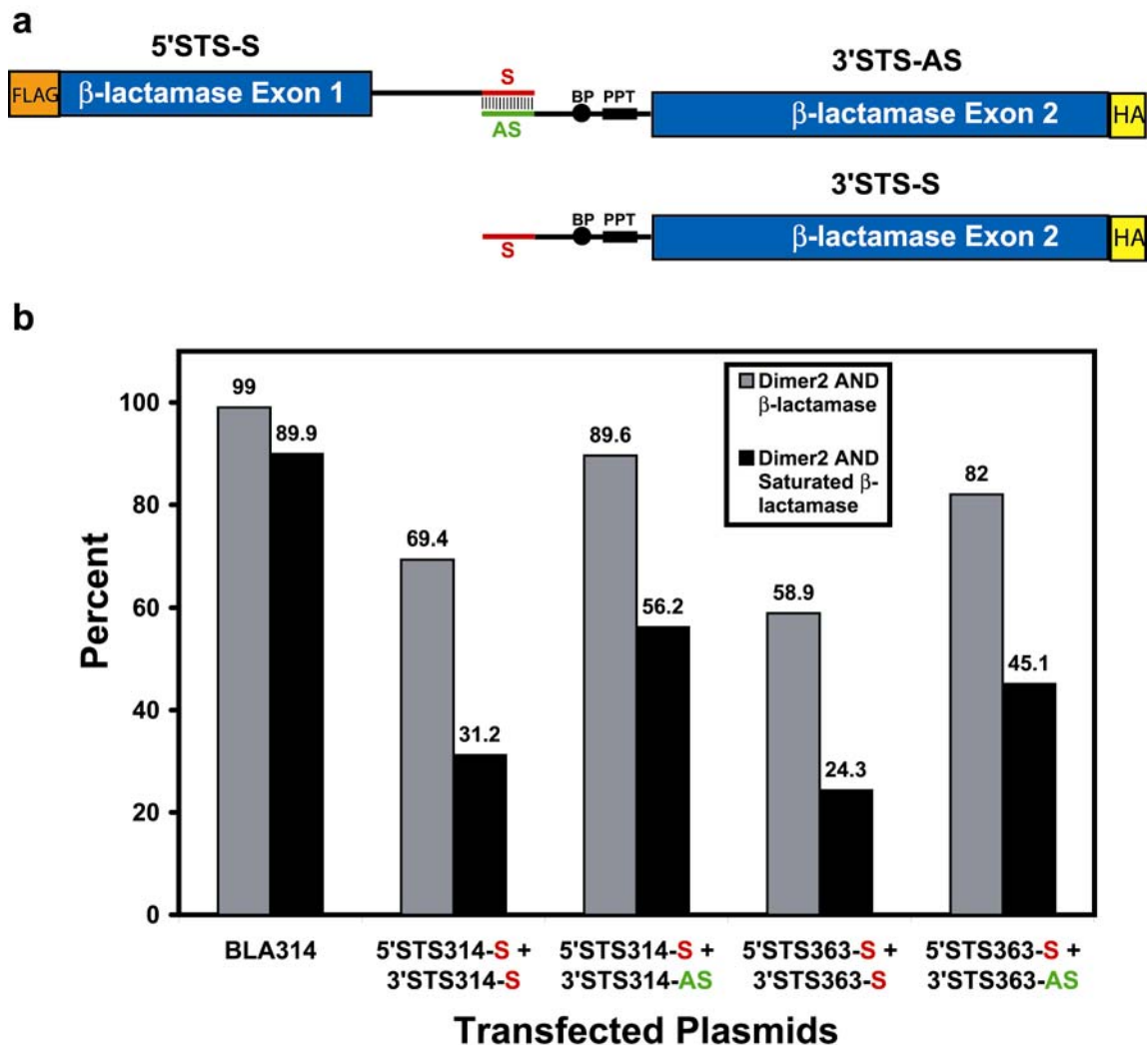
Figure 3.21. Segmental trans-splicing is only partially sequence dependent. (a) Design of split β-lactamase reporters for STS. Green represents antisense sequences and red represents sense sequences used for the binding domains of each fragment. As a control, binding domains on both the 3' and 5' components were tested with the same sense, which should not contribute to any recognition. (b) Summary of β-lactamase reporter results measuring the percent β-lactamase positive cells. Transfected cells were detected by the presence of Dimer2 fluorescence from co-transfection. Saturated β-lactamase activity is also scored, and represents cells with a maximal CCF2 FRET ratio. These cells have processed all the CCF2 substrate, and therefore express enough β-lactamase enzyme to fall out of the linear range
.

Discussion

Our efforts to use trans-splicing for the specific delivery of genetically encoded

reporters for visualizing gene expression was not achieved. Despite published reports to

the contrary, we were unable to copy the design features used by others to generate

trans-splicing ribozymes with any cellular activity. Attempts to map the internal guide sequence using *in vitro* approaches were also fruitless. Clearly our attempts at ribozyme-mediated trans-splicing were not exhaustive. The U10 and U26 targeting ribozymes were never tested *in vitro*. Also, because of complicated material transfer agreements, we were unable to receive any published positive controls, so our negative results are unsupported. Others have since successfully designed ribozymes targeting dsRed-β-lactamase hybrids, and restored β-lactamase activity[23].

Due to these complications, we refocused our efforts based on the early promise of spliceosome-mediated RNA trans-splicing. Delivery of full-length β-lactamase was successful, but exposed the high level of non-specific splicing. New approaches for decreasing non-specificity were developed that required two individual trans-splicing events. Despite successful target dependent 3' and 5' exon replacement, non-specific splicing was still not reduced. On no occasion was β-lactamase protein resulting from trans-spliced RNA detected, despite robust visualization of β-lactamase activity.

In hindsight, our designed PTMs and introns had several problems. One problem (**Fig. 3.19c**) was the shortened β-lactamase product visualized by RT-PCR of β-lactamase-intron transfected cells. Analysis of expressed protein by western blot showed BLA314 translated full-length protein at levels approaching β-lactamase expressed without an intron (**Fig. 3.10**). This implies either the smaller RT-PCR product is the result of a RT-PCR artifact, or that cis-splicing can also occur at a cryptic site, generating a spliced mRNA that is not efficiently translated or has poor protein stability. Clearly, β-lactamase activity was present, so specific splicing occurred, but for unknown reasons, the full-length product was undetectable using our RT-PCR procedure.

Finally, much of the non-specific activity seemed to be driven by an unknown mechanism. The source of the spontaneous activity could be to be either non-specific splicing or DNA recombination of the plasmid DNA during co-transfection. In the case of the full-length β-lactamase PTM, it is less probable plasmid recombination resulted in the expression of the reporter. The high non-specific background is more likely due to non-specific splicing. Any untargeted β-lactamase splicing is subject to the triplet codon rules of the genetic code, restricting successful splicing events to a single translation frame. Therefore, only a fraction of non-specific splicing events are reported, implying widespread illegitimate splicing. 5' RACE could find the exact source of the spurious β-lactamase expression. This would have directly explained how expression of the full length β-lactamase PTM, lacking an initiator methionine codon, could be expressed.

Next, to avoid the possible recombination of plasmid DNA in transient transfections, it may have been useful to use a viral method to establish single site integrants of each component. Plasmid-based over-expression could have lead to increased non-specific splicing by localizing the transcription from large numbers of nearby plasmids, decreasing the separation of target and PTM RNAs and minimizing the impact of antisense targeting. This issue could be avoided by generating each component as virions separately in packaging cells, where separate expression would prevent any DNA recombination or premature trans-splicing. To accomplish this, a new viral vector would need to be developed to reverse the sense of the any intron or PTM to prevent any activity occurring during viral RNA processing and packaging. Also, the β-lactamase gene within the vector backbone would need to be replaced with different drug selection marker, which is not a non-trivial task based on the sparse availability of unique restriction sites. By serial transduction, the presence plasmid recombination would be avoided, and allow for FACS selection of stable cells with β-lactamase activity.

Furthermore, it would have been useful to obtain a larger intron for targeting. Several reports have noted that SMaRT is most efficient with long binding domains targeted to even longer introns. It is possible the 125 bp adenoviral intron was too short to efficiently separate the limited accessibility and strong cis-splicing due to the close proximity of the 5' and 3' splice sites. To address this issue, intron 3 from the hTERT gene was amplified from a human BAC clone containing the hTERT gene, the protein subunit of the telomerase and a well-known cancer marker. The 3 kb intron was inserted in two reporter genes, mCherry and β-lactamase. Successful splicing was observed in cells as red fluorescence for mCherry and CCF2 cleavage for β-lactamase. Alas, due to the negative results achieved from our work with the shorter intron, little hope remained for gaining specific trans-splicing.

The reduction of non-specific trans-splicing is a significant challenge, and may require new ideas and new designs. It may be possible to limit non-specific splicing by optimizing linker lengths and binding domain sites, but this approach will require a high-throughput library screening approach for each new target, similar to our initial proposal. One idea for increasing the specific expression of a reporter gene is to filter out non-specific splicing using the cells own mechanisms. If the PTM could be targeted to an export sequence for extracellular display, then membrane impermeant contrast agents, such as labeled antibodies, could be used for detection. Ideally, successful splicing would lead to the export and display of an epitope tag for detection by flow cytometry. Again, this approach will not limit the non-specific splicing, but could prevent a significant portion from detection.

Trans-splicing RNAs have tremendous potential for the repair of clinically relevant mutant RNAs. The nature of non-specific splicing has been difficult to study, since no live cell assays are available. Our system of inducible target gene expression could be

used to determine the extent of non-specific splicing, and serve as a tool for the

discovery of novel mechanisms for its inhibition. Trans-splicing of reporters of

endogenous gene expression has yet to be optimized and will require significant new

ideas and approaches.

<u>Materials and Methods</u>

<u>Trans-splicing ribozymes.</u>

Ribozymes were constructed as diagramed in **Figure 2**, using a modified pCDNA3 mammalian expression vector (Invitrogen) to eliminate 60 bp of 5' untranslated sequence, but leaving the T7 phage promoter. The pCDNA3-UTR was made by PCR cloning a fragment of the CMV promoter (Primer1; Primer2). The β-lactamase gene was amplified from pCDNA3 and the bacterial periplasmic export sequences were removed (Primer3; Primer4) for fusion to the 3' end of the ribozyme. The *Tetrahymena thermophilia* self-splicing group I ribozyme was obtained as a gift from Tom Cech (U. Colorado). Fusions of the desired spicing reaction were generated to test the functionality of the final fusion (Primer5, Primer6; Primer7). The ribozymes IGS, P1 and P10, were cloned to the 5' end of the ribozyme [(Primer8; Primer10) and (Primer9; Primer11)]. Next, extended antisense regions were added (Primer12, Primer13; Primer7). The pieces were then all pasted together using standard restriction enzyme cloning to make the diagramed U10 and U26 dsRed targeted ribozymes. The target dsRed gene was cloned by PCR into pCDNA3 and pCLNCX from pRSETb (Primer14; Primer15). The IGS mapping library was generated by PCR (Primer16; Primer17) and cloned into the pCDNA3-UTR. RNA was transcribed using the T7 MegaScript kit (Ambion) from gel purified (QIAquick, Qiagen) restriction fragments. RNA was purified adding ammonium acetate and phenol/chloroform extracted, chloroform extracted twice, isopropanol precipitated, washed with ethanol, and resuspended in TE and quantitated by absorbance. The purified RNA was analyzed by running a 4% polyacrylamide urea TBE gel, and then stained in ethidium bromide. Trans-splicing was performed according to published methods[48] and products were reverse transcribed (Primer17 or Primer18) using SuperScriptII or AMV (Invitrogen) at 42°C to 65°C for 1hr, heat inactivated at

70°C for 15 minutes, and then treated with RNaseH for 30 minutes at 37°C and amplified by PCR (Primer14; Primer17). RNA was isolated from NIH3T3, 293T, and HeLa cells using Trizol (Invitrogen) for *in vitro* trans-splicing with cellular RNA. *In vivo* trans-splicing was tested by transfecting cells with dsRed and the IGS library, extracting the RNA, treating the samples with DNase to remove any co-purified plasmid DNA, and processed by RT-PCR for analysis of resulting products by agarose gel electrophoresis. The U44 dsRed targeted ribozyme was generated by PCR (Primer19; Primer17). Samples were excised following separation by urea-polyacrylamide gel electrophoresis, crushed, and eluted in 0.5 M ammonium acetate, 1.0% SDS, 1.0 mM EDTA and heated to 55°C for several hours. The polyacrylamide fragments were pelleted and the supernatant was phenol/chloroform extracted. The full length purified ribozyme RNA was used for subsequent in vitro splicing reactions.

Spliceosome Mediated Trans-splicing. The Dimer2-Intron was received as a gift from Mariano Garcia-Blanco (Duke University). The vector pCDNA3 was modified to replace the β-lactamase gene with the kanamycin resistance gene. This required a Quikchange (Stratagene) reaction (Primer20; Primer21) to remove a restriction site, followed by PCR of the kanamycin gene (Primer22, Primer23) from pET30a (Novagen). Following digestion and ligation, the modified vector was named pCDNK3. The Dimer2-Intron gene was amplified from pCDNA6 (Primer24, Primer25) and cloned into pCDNK3. FLAG-BLA-HA was amplified from pCDNA3 vector (Primer26, Primer27) and cloned into pCDNK3.

BLA314 was generated in several steps. First, exon 1 was amplified by PCR (Primer28, Primer29), and then cloned into pCDNK3 Dimer2-Intron, replacing the first Dimer2 exon to generate BID314. Second, the adenoviral intron was re-engineered to destroy several restriction enzymes. The XbaI site was removed by PCR (Primer30,

Primer31) of a the intron and exon 2 with a primer carrying an NheI site, an isoschizomer

of XbaI. Next, the restriction sites surrounding the 3' splice junction were modified by

PCR [(Primer32; Primer31) and (Primer28; Primer33)] and inserted into the vector.

Finally, the second exon of BLA314 was amplified (Primer34; Primer35) and replaced in

the Dimer2 intron and in BID to generate BLA314 and DIB314. BLA363 was generated

by PCR of each component, [Exon1 (Primer28; Primer36), Intron (Primer37; Primer38),

and Exon2 (Primer39; Primer35)], followed by ligation. The first exon was swapped with

the first exon of Dimer2-Intron to generate BID363. The first exon of Dimer2-Intron was

cloned to the 5' end of the intron and the second exon to generate DIB363. Fusion of the

all three components generated BLA363. BLA514 was made by PCR of each

component, [Exon1 (Primer28; Primer40), Intron (Primer37; Primer41), Exon2

(Primer42; Primer35)]. Each component was recombined, as above to generate BID514,

DIB514, and BLA514. The full-length PTM-BLA was generated by PCR first of a scaffold

(Primer43; Primer35), then by adding on the antisense binding domain (Primer44;

Primer45). PTM314 was made by PCR (Primer46; Primer35), then fusion of the binding

domain used in the PTM-BLA. PTM363 (Primer47; Primer35) and PTM514 (Primer48;

Primer35) were made similarly with the same binding domain. A shortened binding

domain version of each was also made (Primer49; Primer50). The 5'ER constructs were

made by PCR [5'PTM314 (Primer28;Primer51), 5'PTM363 (Primer28; Primer52),

5'PTM514 (Primer28; Primer53)] of each exon with the appropriate linker, then fusion to

a binding domain (Primer54; Primer55). A shortened binding domain was also generated

(Primer56; Primer57). To express each exon individually, each was amplified by PCR

[Exon1-314 (Primer28; Primer58), Exon1-363 (Primer28; Primer59), Exon1-514

(Primer28;,Primer60), Exon2-314 (Primer61; Primer35), Exon2-363 (Primer62;

Primer35), Exon2-514 (Primer63; Primer35)] and sub-cloned into pCDNK3 for

expression. To test segmental trans-splicing, each component was amplified by PCR [STS363-S (Primer28; Primer64), STS314-S (Primer28; Primer65), 3STS314-AS (Primer66; Primer35), 3STS314-S (Primer67; Primer35), 3STS363-AS (Primer68, Primer35), 3STS363-S (Primer69; Primer35)] and sub-cloned into pCDNK3.

Cell Culture and transfection. Cells were grown in DMEM supplemented with 10% FBS and 1% penicillin/streptomycin. Fugene6 (Roche) was used for all transfections, using ~ 1 μg of total plasmid DNA and 3 μl of transfection reagent per 2 ml of media (35 mm dish). Empty vector DNA was used to raise the DNA concentrations to a steady level across all samples. Cells were analyzed two days post-transfection.

RNA isolation and analysis. RNA was purified from transfected cells using Trizol (Invitrogen) and eluted in TE. Approximately 1 μg of RNA was DNase treated and reverse transcribed with SuperScriptII (Invitrogen), SuperScriptIII (Invitrogen, high temperature), ImpromII (Promega, low temperature). The reaction temperature was varied according to the enzyme specifications. RNA was degraded following reverse transcription with RNaseH, and PCR amplified with Taq polymerase (Roche).

CCF2 staining and flow cytometry. Cells were stained according to manufacturer protocols (GeneBlazer, Invitrogen) in which 1 μl of a 1 mM stock of CCF2 is diluted 1:10 in a solution containing 100 mg/ml Pluronic-F127 and 0.1% acetic acid in DMSO. This mix is then diluted in 1 ml of HBSS supplemented with 2 g/l glucose and 20 mM Hepes. To this, 10 μl of 250 mM probenecid was added per milliliter. After removing the growth medium, the staining solution was added for approximately 45 minutes to 1 hour at room temperature. Cells were then trypsinized and pelleted, then resuspended in the

supplemented HBSS for flow cytometry analysis. Cells were kept in the presence of
probenecid to prevent dye leakage and analyzed within 15 minutes. Flow cytometry was
performed on a BD FACSVantage DiVa flow cytometer equipped with two lasers, a 568
nm primary laser and a 413 nm secondary laser. The 568 nm laser was used to excite
red fluorescent proteins, including dsRed and Dimer2, with a 630/30 nm emission filter.
The 413 nm laser was used to excite CCF2, and the emission was split using a 505 SP
dichroic to collect 450/20 nm (blue) and 510/21 nm (green) emission. Cells were gated
by forward and side scatter, and then gated for transfection by red fluorescence.

Western blotting. Cells were lysed in 25 mM Tris pH 7.5, 150 mM NaCl, 1% Triton X100,
2.5 mM DTT, and Complete protease inhibitor (Roche Diagnostics), then mixed 1:1 with
2X sample loading buffer and boiled 15 minutes. Samples were separated on either a
10-20% gradient tricine gel (Invitrogen) (protein fragment complementation) or on 15%
with a standard tris buffered polyacrylamide gel with a 5% stacking gel (trans-splicing),
then transferred to Immobilon (Millipore) and blocked in TTBS (50 mM Tris pH 8, 138
mM NaCl, 2.7 mM KCl, 0.05% tween 20) with 10% whole milk. Membranes were
incubated with anti-HA (Sigma) (1:5,000) or anti-FLAG (Sigma) (1:10,000) in 2% BSA
overnight, then washed 4 times with 10% milk in TTBS. Added goat anti-mouse HRP
antibody (EMD Biosciences) 1:10,000 for 1 hour in 7.5% milk, then washed 4 times in
10% milk TTBS and developed using a luminescent ECL substrate (Amersham
Biosciences).

Table 3.1. Primer Sequences

Primer	Primer Name	Sequence
Primer1	5'NdeCMV	GGGGCATATGCCAAGTACGCCCCCTATT
Primer2	5'KpnI start site	AAAAAAGGTACCTAGTTAGCCAGAGAGCTCTG CTTA
Primer3	5' EcoRI BLA	AAAAATGAATTCAGGGCACCCAGAAACGCTGG TGAAA
Primer4	3'BLA BamHI	AAAAATGGATCCTTACCAATGCTTAATCAGTGA
Primer5	5'N-dsRed U26 spliced	TTTTTAGGTACCATGGTGCGCTCCTCCAAGAA CGTCATTAAGAATTCAGGGCACCCAGAAACG
Primer6	5'N-dsRed U10 spliced	GGGGGGTACCGCCACCATGGTGCGCTTCTTG AATTCAGGGCACCCAGAAACG
Primer7	3' XhoI BLA	GGGGCTCGAGTTACCAATGCTTAA
Primer8	5'rz-dsRed(U10)P1	GGGGCTCGAGATTCAAGGGGCGCAAAAGTTA TCAG
Primer9	5'rz-dsRed(U26)P1	GGGGCTCGAGATTCTTGGTGACGAAAAGTTAT CAGGCATGCACC
Primer10	3'rz-dsRed(U10)P10	GGGGGAATTCAAGACGAGTACTCCAAAACTAA TCA
Primer11	3'rz-dsRed(U26)P10	GGGGGAATTCTTACGAGTACTCCAAAACTAAT CA
Primer12	U10 Extended antisense	GGGGGGTACCGCACCTTGAAGCGCATGAACT CCTTGATGACGTTCCTCGAGTTCAAGGGGCG CAA
Primer13	U26 extended antisense	GGGGGGTACCACGGTGCCCTCCATGCGCACC TTGAAGCGCATGAACTCGAGATTCTTGGTGAC GA
Primer14	5'KpnKzDsRed	TTTTGGTACCGCCACCATGGTGCGCTCCTCCA AG
Primer15	3'Cla dsRed	GGGGTAGCTACTACAGGAACAGGTGGTGGCG
Primer16	5'mapping library	GGGGGGTACCTAATACGACTCACTATAGGNN NNNAAAAGTTATCAGGCATGCACC

Table 3.1. Continued.

Primer17	3' mapping library	AGTAGTCTTAGAATTCGGGCCTCTTCGCTATT ACG
Primer18	3' mapping library V2	GGGGGAATTCGGATGTGCTGCAAGGCGATTA AGTTGGG
Primer19	5'IVU44P1	GGGGGGTACCTAATACGACTCACTATAGGAC GCCAAAAGT
Primer20	5'QC pcDNA3 Bgl2	GGTCATGAGATTATCAAAAAAGATCTTCACCTA GATCC
Primer21	3' QC pCDNA3 Bgl2	GGATCTAGGTGAAGATCTTTTTTGATAATCTCA TGACC
Primer22	5' Bgl2Kan	GAAAGATCTCAAGAAGATCCTTTG
Primer23	3' Bgl2Kan	CCCAGATCTGTGCGCGGAACCCCTATTTG
Primer24	HindIII dsRed Intron	GCGAAGCTTGGATCCCTCTAAAAGCGGGCAT GAC
Primer25	EcoRI dsRd intron	TATGAATTCGGTGCCGCGGAACTTCACCTGAG G
Primer26	5'H3kzFlagBLA	TTTAAGCTTGCCACCATGGATTACAAGGATGA CGACGATAAGGGGCACCCAGAAACGCTG
Primer27	3' c-term HA BLA	AAACTCGAGTTAGGCGTAGTCGGGCACGTCG TAGGGGTAGCTGGTCCAATGCTTAATCAGTGA
Primer28	T7	AATACGACTCACTATAGGGA
Primer29	3'bla exon1 BamHI	AGAGGGATCCATACTCACCTCATGGTTATGGC AGCACT
Primer30	5'Red/Intron NheI	GACTGCTAGCGTAGTCCAGGGTTTCCGAG
Primer31	BGH Reverse	TAGAAGGCACAGTCGAGG
Primer32	5'Red/Intron SalI	CGTGTCGACGGCGTACTAACTCGGCCCTTTC C
Primer33	3'sacII bla-intron	TGGCCGCGGTGTTATCACTGAGGAAAAAAAG GGAAAGGG
Primer34	5' spit bla exon2 sacII	AACACCGCGGCCAACTTACTTCTGAC

Table 3.1. Continued.

Primer35	3' Reverse HA XhoI	AAACTCGAGTTAGGCGTAGTCGGGC
Primer36	3'Bla Intron2exon1	GAGGGATCCATACTCACCTTCGGTCCTCCGAT CGTTG
Primer37	Intron BamHI	TATGGATCCCTCTAAAAGCGGGCATGAC
Primer38	Intron SacII	AAACCGCGGAGTTAGTACGCCGTCGAG
Primer39	Bla intron2exon2 sacII	ACTCCGCGGTTTCCCTTTTTTTTCCTCAGGAG CTAACCGCTTTTTTG
Primer40	3'bla exon1v3	GAGGGATCCATACTCACCTGTTAATAGTTTGC GCAACG
Primer41	3'IntronV3	GAAGCTAGCGTAAGTAGTTCGCCTGAGGAAAA AAAAGGGAAAGGGC
Primer42	5'blaexon2V3	TACGCTAGCTTCCCGGCAACAATTA
Primer43	PTM	GGGGGAATTCGAACATTATTATAACGTTTACTA ACTGCTAGCTCTTCTTTTTTTTTTGATATCCTG CAGGGGCACCCAGAAACGCTGGTG
Primer44	5'Hind3BD	GCGAAGCTTGGATCCCTCTAAAAGCG GGCATGAC
Primer45	3'EcoRIBD	TATGAATTCGGTGCCGCGGAACTTCACCTGAG G
Primer46	split bla PTM EcoRV	GGGGGATATCCTGCAGTGATAACACCGCGGC CAAC
Primer47	5'splblax2v2	GATATCCTGCAGGAGCTAACCGCTTTTTTGCA C
Primer48	5'x2v3E5	TTTGATATCCTGCAGGCGAACTACTTACGCTA GCTTC
Primer49	PTM 3' anchor H3	GGGAAGCTTGGGTCTCCCTATAGTGAGTCG
Primer50	3'-16BDxatg	AAAGGATCCGACTTCTAGCGTAGTCCAG
Primer51	3'x1v1ss5	ACTGAATTCTACGTTATAATAATGTTCACTCAC CTCATGGTTATGGCAGCAC
Primer52	3'x1v2ss5	ACTGAATTCTACGTTATAATAATGTTCACTCAC CTTCGGTCCTCCGATCGTTG

Table 3.1. Continued.

Primer53	3'x1v3	ACTGAATTCTACGTTATAATAATGTTCACTCAC CTGTTAATAGTTTGCGCAAC
Primer54	Ex1PTMIntV1 EcoRI	GTTGAATTCAGTCGACGGAAACCCTCGG
Primer55	3'Ex1PTMIntV1 Xho	AAACTCGAGGGCACGCTGATCTACAAGG
Primer56	PTM 5' anchor EcoRI	GGGGAATTCGAACATTATTATAACGTTTAC
Primer57	PTM 3'-42bp EcoRI	GGGGAATTCGCAGTTAGTACGCCGTCGAGAC GAGC
Primer58	3'x1v2PC	TTTCTCGAGTTACATGGTTATGGCAGCACTGC
Primer59	3'x1v2PC	TTTCTCGAGTTACTTCGGTCCTCCGATCGTTG
Primer60	3'x1v3PC	TTTCTCGAGTTATGTTAATAGTTTGCGCAACG
Primer61	5'X2V1PC	TTTAAGCTTGCCACCATGAGTGATACCACCGC GGCCACC
Primer62	5'X2V2PC	TTTAAGCTTGCCACCATGGAGCTAACCGCTTT TTTGCAC
Primer63	5'X2V3PC	TTTAAGCTTGCCACCATGGGCGAACTACTTAC GCTAGC
Primer64	Ex2XhoI	AAACTCGAGTCGACGGAAACCCTCGGAAAC
Primer65	Ex1XhoI	AAACTCGAGTCGACGGAAACCCTCGGAAAC
Primer66	ASEx2Overlap	TTTAAGCTTCGACGGAAACCCTCGGAAACCCT GGACTACGCTAGAAGTCTCAGCTCGTCTCGA
Primer67	5'H3Ex2	TTTAAGCTTACTTCTAGCGTAGTCCAGGG
Primer68	ASEx2Overlap	TTTAAGCTTCGACGGAAACCCTCGGAAACCCT GGACTACGCTAGAAGTCTCAGCTCGTCTCGA
Primer69	5'H3Ex2	TTTAAGCTTACTTCTAGCGTAGTCCAGGG

References

1. Sullenger,B.A. & Cech,T.R. Ribozyme-Mediated Repair of Defective Messenger-Rna by Targeted Transsplicing. *Nature* **371**, 619-622 (1994).

2. Puttaraju,M., Jamison,S.F., Mansfield,S.G., Garcia-Blanco,M.A., & Mitchell,L.G. Spliceosome-mediated RNA trans-splicing as a tool for gene therapy. *Nature Biotechnology* **17**, 246-252 (1999).

3. Zlokarnik,G. Quantitation of transcription and clonal selection of single living cells with beta-lactamase as reporter. (1998).

4. Zlokarnik,G. Quantitation of transcription and clonal selection of single living cells with beta-lactamase as reporter. (1998).

5. Chalfie,M., Tu,Y., Euskirchen,G., Ward,W.W., & Prasher,D.C. Green Fluorescent Protein As A Marker for Gene-Expression. *Science* **263**, 802-805 (1994).

6. Contag,P.R., Olomu,I.N., Stevenson,D.K., & Contag,C.H. Bioluminescent indicators in living mammals. *Nature Medicine* **4**, 245-247 (1998).

7. Yu,Y.J., Annala,A.J., Barrio,J.R., Toyokuni,T., Satyamurthy,N., Namavari,M., Cherry,S.R., Phelps,M.E., Herschman,H.R., & Gambhir,S.S. Quantification of target gene expression by imaging reporter gene expression in living animals. *Nature Medicine* **6**, 933-937 (2000).

8. Murphy,W.J., Watkins,K.P., & Agabian,N. Identification of A Novel Y-Branch Structure As An Intermediate in Trypanosome Messenger-Rna Processing - Evidence for Trans Splicing. *Cell* **47**, 517-525 (1986).

9. Dorn,R., Reuter,G., & Loewendorf,A. Transgene analysis proves mRNA trans-splicing at the complex mod(mdg4) locus in Drosophila. *PNAS* **98**, 9724-9729 (2001).

10. Horiuchi,T., Giniger,E., & Aigaki,T. Alternative trans-splicing of constant and variable exons of a Drosophila axon guidance gene, lola. *Genes Dev.* **17**, 2496-2501 (2003).

11. Eul,J., Graessmann,M., & Graessmann,A. Trans-splicing and alternative-tandem-cis-splicing: Two ways by which mammalian cells generate a truncated SV40 T-antigen. *Nucl. Acids Res.* **24**, 1653-1661 (1996).

12. Li,B.L., Li,X.L., Duan,Z.J., Lee,O., Lin,S., Ma,Z.M., Chang,C.C., Yang,X.Y., Park,J.P., Mohandas,T.K., Noll,W., Chan,L., & Chang,T.Y. Human Acyl-CoA:Cholesterol Acyltransferase-1 (ACAT-1) Gene Organization and Evidence That the 4.3-Kilobase ACAT-1 mRNA Is Produced from Two Different Chromosomes. *J. Biol. Chem.* **274**, 11060-11071 (1999).

13. Takahara,T., Kanazu,S.i., Yanagisawa,S., & Akanuma,H. Heterogeneous Sp1 mRNAs in Human HepG2 Cells Include a Product of Homotypic trans-Splicing. *J. Biol. Chem.* **275**, 38067-38072 (2000).

14. Takahara,T., Tasic,B., Maniatis,T., Akanuma,H., & Yanagisawa,S. Delay in Synthesis of the 3' Splice Site Promotes trans-Splicing of the Preceding 5' Splice Site. *Molecular Cell* **18**, 245-251 (2005).

15. Lan,N., Howrey,R.P., Lee,S.W., Smith,C.A., & Sullenger,B.A. Ribozyme-mediated repair of sickle beta-globin mRNAs in erythrocyte precursors. *Science* **280**, 1593-1596 (1998).

16. Phylactou,L.A., Darrah,C., & Wood,M.J.A. Ribozyme-mediated trans-splicing of a trinucleotide repeat. *Nat Genet* **18**, 378-381 (1998).

17. Kastanos,E., Hjiantoniou,E., & Phylactou,L.A. Restoration of protein synthesis in pancreatic cancer cells by trans-splicing ribozymes. *Biochemical and Biophysical Research Communications* **322**, 930-934 (2004).

18. Watanabe,T. & Sullenger,B.A. Induction of wild-type p53 activity in human cancer cells by ribozymes that repair mutant p53 transcripts. *Proceedings of the National Academy of Sciences of the United States of America* **97**, 8490-8494 (2000).

19. Lan,N., Rooney,B.L., Lee,S.W., Howrey,R.P., Smith,C.A., & Sullenger,B.A. Enhancing RNA repair efficiency by combining trans-splicing ribozymes that recognize different accessible sites on a target RNA. *Molecular Therapy* **2**, 245-255 (2000).

20. Kohler,U., Ayre,B.G., Goodman,H.M., & Haseloff,J. Trans-splicing ribozymes for targeted gene delivery. *Journal of Molecular Biology* **285**, 1935-1950 (1999).

21. Ayre,B.G., Kohler,U., Turgeon,R., & Haseloff,J. Optimization of trans-splicing ribozyme efficiency and specificity by in vivo genetic selection. *Nucl. Acids Res.* **30**, (2002).

22. Hagen,M. & Cech,T.R. Self-splicing of the Tetrahymena intron from mRNA in mammalian cells. *Embo Journal* **18**, 6491-6500 (1999).

23. Hasegawa,S., Choi,J.W., & Rao,J.H. Single-cell detection of trans-splicing ribozyme in vivo activity. *Journal of the American Chemical Society* **126**, 7158-7159 (2004).

24. Hasegawa,S., Jackson,W.C., Tsien,R.Y., & Rao,J. Imaging Tetrahymena ribozyme splicing activity in single live mammalian cells. *PNAS* **100**, 14892-14896 (2003).

25. Hasegawa,S. & Rao,J. Modulating the splicing activity of Tetrahymena ribozyme via RNA self-assembly. *FEBS Letters* **In Press** (2006)

26. Kwon,B.S., Jung,H.S., Song,M.S., Cho,K.S., Kim,S.C., Kimm,K., Jeong,J.S., Kim,I.H., & Lee,S.W. Specific Regression of Human Cancer Cells by Ribozyme-Mediated Targeted Replacement of Tumor-Specific Transcript. *Molecular Therapy* **12**, 824-834 (2005).

27. Garcia-Blanco,M.A. Messenger RNA reprogramming by spliceosome-mediated RNA trans-splicing. *Journal of Clinical Investigation* **112**, 474-480 (2003).

28. Mitchell,L.G. & McGarrity,G.J. Gene therapy progress and prospects: Reprograming gene expression by trans-splicing. *Gene Therapy* **12**, 1477-1485 (2005).

29. Mansfield,S.G., Clark,R.H., Puttaraju,M., Kole,J., Cohn,J.A., Mitchell,L.G., & Garcia-Blanco,M.A. 5 ' Exon replacement and repair by spliceosome-mediated RNA trans-splicing. *Rna-A Publication of the Rna Society* **9**, 1290-1297 (2003).

30. Mansfield,S.G., Kole,J., Puttaraju,M., Yang,C.C., Garcia-Blanco,M.A., Cohn,J.A., & Mitchell,L.G. Repair of CFTR mRNA by spliceosome-mediated RNA trans-splicing. *Gene Therapy* **7**, 1885-1895 (2000).

31. Chao,H.J., Mansfield,S.G., Bartel,R.C., Hiriyanna,S., Mitchell,L.G., Garcia-Blanco,M., & Walsh,C.E. Phenotype correction of hemophilia A mice by spliceosome-mediated RNA trans-splicing. *Nature Medicine* **9**, 1015-1019 (2003).

32. Liu,X.M., Jiang,Q.S., Mansfield,S.G., Puttaraju,M., Zhang,Y.L., Zhou,W.H., Cohn,J.A., Garcia-Blanco,M.A., Mitchell,L.G., & Engelhardt,J.F. Partial correction of endogenous Delta F508 CFTR in human cystic fibrosis airway epithelia by spliceosome-mediated RNA trans-splicing. *Nature Biotechnology* **20**, 47-52 (2002).

33. Rodriguez-Martin,T., Garcia-Blanco,M.A., Mansfield,S.G., Grover,A.C., Hutton,M., Yu,Q.M., Zhou,J.H., Anderton,B.H., & Gallo,J.M. Reprogramming of tau alternative splicing by spliceosome-mediated RNA trans-splicing: Implications for tauopathies. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 15659-15664 (2005).

34. Rogers,C.S., Vanoye,C.G., Sullenger,B.A., & George,A.L. Functional repair of a mutant chloride channel using a trans-splicing ribozyme. *Journal of Clinical Investigation* **110**, 1783-1789 (2002).

35. Tahara,M., Pergolizzi,R.G., Kobayashi,H., Krause,A., Luettich,K., Lesser,M.L., & Crystal,R.G. Trans-splicing repair of CD40 ligand deficiency results in naturally regulated correction of a mouse model of hyper-IgM X-linked immunodeficiency. *Nature Medicine* **10**, 835-841 (2004).

36. Puttaraju,M., DiPasquale,J., Baker,C.C., Mitchell,L.G., & Garcia-Blanco,M.A. Messenger RNA repair and restoration of protein function by spliceosome-mediated RNA trans-splicing. *Molecular Therapy* **4**, 105-114 (2001).

37. Kikumori,T., Cote,G.J., & Gagel,R.F. Promiscuity of pre-mRNA spliceosome-mediated trans splicing: A problem for gene therapy? *Human Gene Therapy* **12**, 1429-1441 (2001).

38. Pergolizzi,R.G., Ropper,A.E., Dragos,R., Reid,A.C., Nakayama,K., Tan,Y., Ehteshami,J.R., Coleman,S.H., Silver,R.B., Hackett,N.R., Menez,A., & Crystal,R.G. In vivo trans-splicing of 5 ' and 3 ' segments of pre-mRNA directed by corresponding DNA sequences delivered by gene transfer. *Molecular Therapy* **8**, 999-1008 (2003).

39. Nakayama,K., Pergolizzi,R.G., & Crystal,R.G. Gene transfer-mediated pre-mRNA segmental trans-splicing as a strategy to deliver intracellular toxins for cancer therapy. *Cancer Res* **65**, 254-263 (2005).

40. Pergolizzi,R., Ropper,A., Dragos,R., Reid,A., Nakayama,K., Silver,R., Hackett,N., Menez,A., & Crystal,R. In vivo trans-splicing of 5 ' and 3 ' segments of pre-mRNA directed by corresponding DNA sequences delivered by gene transfer. *Molecular Therapy* **7**, S21 (2003).

41. Bhaumik,S., Walls,Z., Puttaraju,M., Mitchell,L.G., & Gambhir,S.S. Molecular imaging of gene expression in living subjects by spliceosome-mediated RNA trans-splicing. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 8693-8698 (2004).

42. Deidda,G., Rossi,N., & Tocchini-Valentini,G.P. An archaeal endoribonuclease catalyzes cis- and trans- nonspliceosomal splicing in mouse cells. *Nat Biotech* **21**, 1499-1504 (2003).

43. Urlinger,S., Baron,U., Thellmann,M., Hasan,M.T., Bujard,H., & Hillen,W. Exploring the sequence space for tetracycline-dependent transcriptional activators: Novel mutations yield expanded range and sensitivity. *PNAS* **97**, 7963-7968 (2000).

44. Zlokarnik,G., Negulescu,P.A., Knapp,T.E., Mere,L., Burres,N., Feng,L.X., Whitney,M., Roemer,K., & Tsien,R.Y. Quantitation of transcription and clonal selection of single living cells with beta-lactamase as reporter. *Science* **279**, 84-88 (1998).

45. Zlokarnik,G. Fusions to beta-lactamase as a reporter for gene expression in live mammalian cells. *Applications of Chimeric Genes and Hybrid Proteins, Pt A* **326**, 221-241 (2000).

46. Galarneau,A., Primeau,M., Trudeau,L.E., & Michnick,S.W. Β-lactamase protein fragment complementation assays as in vivo and in vitro sensors of protein-protein interactions. *Nat Biotech* **20**, 619-622 (2002).

47. Wehrman,T., Kleaveland,B., Her,J.H., Balint,R.F., & Blau,H.M. Protein-protein interactions monitored in mammalian cells via complementation of beta -lactamase enzyme fragments. *PNAS* **99**, 3469-3474 (2002).

48. Jones J.T., Lee S.W., and Sullenger, B. A. Trans-splicing reactions by ribozymes. *Methods Mol Biol.* 74, 341-348. 1997.