Rainbow tags: a visual tag system for recombinant protein expression and purification

Robert D. Finn, Iouri Kapelioukh, and Mark J.I. Paine
Ninewells Hospital and Medical School, Dundee, UK

BioTechniques 38:387-392 (March 2005)

Visualization systems for tracking proteins are standard experimental tools in most areas of biological research apart from protein purification. Here, we have sought to plug this gap by producing red and yellow visual tags using the heme-binding domain of mosquito cytochrome b5 and the flavin mononucleotide (FMN)-binding domain of human P450 reductase. Tests with colorless glutathione-S-transferase (GST) show them to be simple and effective tools for visually identifying correctly folded protein and tracking protein molecules through protein expression and purification. Furthermore, the characteristic absorbance signatures of the colored tags can be used to quantify protein concentrations directly, which allows purification to be linked to colorimetric detection. This technology, which we call Rainbow Tagging, facilitates expression and downstream processing of recombinant proteins, paving the way for the development of automated high-throughput protein expression systems.

INTRODUCTION

The rationale behind the development of a visible tag system derives from the fact that protein isolation procedures are much easier to perform with visible polypeptides such as flavoenzymes or hemoproteins that contain colored chromophores. For instance, the flavodoxin-like flavin mononucleotide (FMN)-binding domain of human cytochrome P450 reductase (CPR) lends itself particularly well to protein purification, as it is small (21 kDa), visible (blue-green or yellow, depending on the oxidation state of the FMN cofactor), and expresses in soluble form at extremely high levels (30–40 mg/L) in *Escherichia coli* (1,2). The same applies to the small (approximately 17 kDa) red heme-binding molecule cytochrome b5 (b5)(3). Thus, the combination of visibility, solubility, and small size make these molecules ideal fusion tags for generating visible recombinant proteins that can be tracked through the purification process. To exploit this, we have adapted a red mosquito cytochrome b5 molecule and a yellow human FMN-binding domain to use as visual tags. These have extremely useful practical

applications in facilitating normal protein purification procedures.

MATERIALS AND METHODS

Chemicals and Reagents

All chemicals were purchased from Sigma (Poole, Dorset, UK), and all enzymes from Life Technologies (Paisley, UK), except where stated. Nickel agarose affinity resin was from Qiagen (Valencia, CA, USA).

Plasmid Construction

Plasmids used in this study are shown in Figure 1 and oligonucleotide primer sequences (RF1–7) are listed in Table 1. A two-step PCR using *Pfu* DNA polymerase (Stratagene, La Jolla, CA, USA) was used to generate pFMN (Figure 1), which contains the FMN-binding domain of human CPR fused to an N-terminal hexahistidine (His₆) tag and a C-terminal thrombin cleavage sequence (TCS). The first PCR used pET15b (EMD Biosciences, San Diego, CA, USA) as template with RF1 and RF2 as forward and reverse primers, respectively, to produce a

DNA fragment containing the *Xba*I site of pET15b and a His₆ sequence fused to the first 21 nucleotides of CPR exon 2 (2). This fragment was gel-purified and used as the forward primer along with RF3 as the reverse primer in a PCR with pMP plasmid as template, which contains the CPR FMN domain sequence (2). The resultant 600-bp DNA fragment, encoding a His₆-CPR FMN-TCS sequence, was subcloned into the unique *Xba*I/*Xho*I sites of pET15b to generate the yellow pFMN vector.

A 700-bp fragment encoding schistosomal glutathione-S-transferase (GST) was amplified by PCR using RF4 and RF5 as forward and reverse primers, respectively, and pGEX 6P-1 (Amersham Biosciences Ltd., Little Chalfont, Buckinghamshire, UK) as a DNA template. The resulting fragment was subcloned into the unique *Nde*I/*Xho*I sites of pET15b and pFMN to create pGST and pFMN-GST (Figure 1), which express colorless and yellow GST protein, respectively.

To generate a red fusion protein (b5-GST), nucleotides 2–300 of the *Anopheles gambiae* cDNA sequence (4) were amplified using RF6 and RF7 as forward and reverse primers, which respectively incorporate an N-terminal His₆ tag to facilitate affinity purification and a C-terminal thrombin cleavage sequence immediately downstream of the b5 coding sequence. The resultant PCR product was subcloned as an *Nco*I/*Nde*I fragment into pFMN-GST, so as to replace the yellow FMN tag with a red b5 tag to generate the plasmid pb5-GST.

Recombinant Protein Expression and Purification

For *E. coli* expression of the various constructs, BLR (De3/pLysS) strains transformed with the appropriate plasmid were grown at 37°C in Terrific Broth (TB) containing 50 µg/mL ampicillin and 30 µg/mL chloramphenicol to an A₆₀₀ of 0.4–0.8. For expression using the b5 tag (b5-GST), the TB growth media was supplemented with 1 mM δ-aminolevulinic acid, a heme precursor that affects efficient heme incorporation. Recombinant protein expression was

Table 1. Oligonucleotide Primers Used for PCR

Name	Nucleotide Sequence
RF1	5'-CCTCTAGAAATAATTTTGTTTAACTTTAAGAAGGAGATATACCATG-3'
RF2	5'-GCTCTCTCTGACAGAGGAGGTGCCGCTGCTGTGATGATG-3'
RF3	5'-CCTCGAGCATATGGCTGCCGCGCGGCACCAGGCTGGACTCCTCGCCAGT-3'
RF4	5'-GAGGAATTCATATGTCCCCTATACTAGGTTAT-3'
RF5	5'-CGGGGAATTCGGGGATCCCAGGGGCCCTAGAACAG-3'
RF6	5'-GATATACCATGGGCAGCAGCCATCATCATCATTCGGAAGTGAAC ACGTACTCGC-3'
RF7	5'-CCTCGAGCATATGGCTGCCGCGCGGCACCAGCTGCTGGTCCATTTCCAGTC-3'

induced by adding isopropyl-β-D-thiogalactopyranoside (IPTG) to 1 mM and growing cultures at 30°C for 15 h. Cells from 50-mL cultures were harvested by centrifugation at 5000× g for 20 min and resuspended in 5 mL

binding buffer (20 mM Tris, pH 8.0, 500 mM NaCl). Cell suspensions were lysed by freeze-thawing and sonication (Soniprep 150 probe; MSE Ltd, Kent, UK) using several short bursts at high power, and then centrifuged at 16,000×

g for 10 min at 4°C in Eppendorf® tubes. The supernatants were loaded onto approximately 2-mL nickel-nitrilotriacetic acid (NTA) agarose columns previously equilibrated in binding buffer. Proteins were washed consecutively with 10 mL binding buffer and 10 mL binding

buffer/40 mM imidazole and eluted with binding buffer/300 mM imidazole. Purified fractions were collected into Eppendorf tubes by hand.

For fast protein liquid chromatography (FPLC) purification, the supernatant containing FMN-GST was loaded onto a 1 mL Hi-Trap™ nickel column (Amersham Biosciences) using an AKTA™ FPLC machine (Amersham Biosciences). The column was washed with 5 column volumes of binding buffer, and bound protein was eluted using a 10-column volume linear gradient of 0–350 mM imidazole. Fractions (0.2 mL) were collected in a 96-well microplate using a Frac-950 fraction collector (Amersham Biosciences). Fractions were spectrally monitored at 450 nm.

Analysis of Fractions

Extinction coefficients were calculated from the formula $E = A/c$, where A is the absorbance of the solution in a cuvette with a 1-cm light path, and c is the concentration in μM. Protein fractions of approximately 95% purity, as estimated by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE), were analyzed. b5-GST, with an estimated molecular weight of 38 kDa, was measured at 412 nm, and FMN-GST, with an estimated molecular weight of 50 KDa, was measured at 450 nm. Analysis of GST activity was performed using 1-chloro-2,4-dinitrobenzene (CDNB) and glutathione (GSH) as substrates. Measurement of the formation of CDNB-GSH conjugate was carried out at 37°C using a Cobas Fara centrifugal analyzer (Roche Diagnostics, Welwyn Garden City, Herts, UK) as previously described (5). For thrombin cleavage,

Figure 1. Plasmid maps. Construction details are described in Materials and Methods. The nucleotide sequences across the cloning junctions of the FMN-GST and b5-GST fusions are shown at the bottom. TCS, Thrombin cleavage sequence; FMN, flavin mononucleotide; GST, glutathione-S-transferase; b5, cytochrome b5.

proteins were buffer-exchanged into thrombin cleavage buffer (20 mM Tris, pH 8.4, 150 mM NaCl, 2.5 mM CaCl_2) and cleaved at 4°C overnight with 0.5 U/mg of thrombin. Protein concentrations were determined using Bradford assays (Sigma) and bovine serum albumin standard. Electrophoresis was carried out using precast NuPage® 4%–12% Bis-Tris gels (Invitrogen, Carlsbad, CA, USA).

RESULTS AND DISCUSSION

To test the visual tagging concept, the coding sequence for a colorless enzyme, *Schistosoma mansoni* GST, was cloned downstream of the b5 and FMN genes, expressed in *E. coli*, and purified. A schematic overview of the origin of the red and yellow tags is shown in Figure 2A. To facilitate protein purification, we engineered a His₆ tag sequence at the amino terminus of the colored molecules, thus enabling metal chelating affinity chromatography to be performed and a TCS to the carboxy-terminal end of each domain to allow the fusion tag to be removed by thrombin digestion following purification (Figure 2B).

To simulate the type of small-scale experiment commonly used for pilot expression studies, we purified the red and yellow-tagged GSTs from 50-mL cultures using standard benchtop proce-

dures and equipment. Recombinant protein expression was detectable from the color of the harvested cell pellets, with b5- and FMN-tagged GSTs producing red and blue-green colored pellets, respectively. The b5 tag retained its pink/red color throughout the purification process, whereas the FMN tag changed color from an initial dark blue-green color in the pellet and cell lysates to a bright yellow color (after which the tag is named) during column purification, reflecting redox changes to the FMN isoalloxazine ring (6).

Visualization of tagged protein is largely dependent on the presence of correctly bound flavin or heme cofactor, which in turn is dependent on the correct folding of the protein. Thus, the presence of the green or pink color in the harvested cell pellet and supernatant is a very effective early indicator of the quantity and solubility of the recombinant fusion protein being expressed. As lack of solubility is a major problem experienced when expressing recombinant protein in *E. coli*, a major advantage of this system is that it allows the identification of clones that express soluble recombinant protein, thus permitting extensive rapid screening and optimization to encourage elevated levels of recombinant protein prior to chromatography. It is worth noting that the recombinant protein expressed in inclusion bodies

also appears to retain cofactor and is therefore visible. This provides another useful visual aid to the quantity and whereabouts of the protein being expressed.

During nickel-agarose affinity purification, a localized accumulation of tagged protein is visible at the top of the column. With the addition of the competitive ligand (imidazole) one observes a displacement of the His₆-tagged fusion proteins (Figure 3A) down the column, which allows the dynamics of protein binding with the column matrix to be visualized in real-time. This ability to track protein movement allows the precise capture of elution fractions containing visible tagged protein. In addition, it also reduces the need for time-consuming methods, such as immunoblotting or enzyme activity measurements, which are commonly used to identify fractions containing proteins of interest.

The absorbance spectra of the FMN and b5 domains produce wavelength maxima in the regions of approximately 450 and 412 nm, respectively, which can be used to directly quantify the protein concentrations of the tagged molecules. Here, we have calculated the extinction coefficients of the b5- and FMN-tagged GSTs to be 0.061 and 0.008 $\mu\text{M}^{-1}\cdot\text{cm}^{-1}$, respectively. Thus, a single absorbance unit (1 AU) is equivalent to protein concentrations of approximately 16 and 118 μM for the yellow and red tags, respectively. These measurements show that the red tag is significantly more sensitive (approximately 10-fold) than the yellow tag. The limit of detection by eye is about 0.1 AU, which means the system is sensitive down to approximately 1.6 μM (0.1 mg b5-GST/mL) for the red tag and approximately 12 μM (0.7 mg FMN-GST/mL) for the yellow tag. Specific activities of the elution fractions towards a typical GST substrate, CDNB, were determined. Both the b5-GST and FMN-GST fusions catalyzed the formation of CDNB-GSH conjugated products with specific activities of 4.4 and 2.8 μmol CDNB-GSH produced/min/mg, respectively. Following thrombin cleavage (Figure 3B), the GST released maintained similar specific activity, indicating the structure and function of

Figure 2. Origin of red and yellow tags and construction of colored GST fusions. (A) The yellow tag corresponds to the 21-kDa flavin mononucleotide (FMN) domain (amino acid residues 61–241) of human cytochrome P450 reductase (CPR). The red tag corresponds to a soluble approximately 13-kDa heme (Fe)-binding region (amino acid residues 1–100) of the mosquito (*Anopheles gambiae*) cytochrome b5 (b5) gene. An, membrane anchor; FAD, flavin-adenine dinucleotide (FAD)-binding domain. (B) The N termini of the FMN and b5 domains were fused to hexahistidine (His₆) and linked to schistosomal glutathione-S-transferase (GST) via a thrombin cleavage sequence (TCS) using PCR and standard cloning techniques. The fusion products are expressed off the T7 promoter. Ap, β -lactamase gene.

the GST was not compromised.

The characteristic absorbance signatures of the tagged proteins means they can be tracked spectrophotometrically, thus increasing the sensitivity of detection and allowing selective monitoring of color-tagged proteins using automated protein purification procedures, such as FPLC. An example of this is shown in Figure 3C, where the yellow FMN-GST has been purified using FPLC, and the fractions were tracked by absorbance at 450 nm.

These experiments show the FMN domain and b5 molecules to be extremely effective visual tags for facilitating recombinant protein expression and purification. Apart from

color, the intrinsic advantages these tag molecules have are their small size (FMN is approximately 21 kDa, while b5 is only approximately 13 kDa) and solubility when expressed at high levels in *E. coli*, which is helpful in conferring solubility to other fused polypeptides (7). As well as GST, we have been successful in using the Rainbow system to express two human kinases, an insect sulfite reductase, and a viral peptide sequence.

While we have employed *E. coli* for expression, the visible tag system could also be used in conjunction with other commonly used protein expression systems such as baculovirus, yeast, or cell-free translation

Figure 3. Visualization of GST fusions during purification. (A) Red [cytochrome b5 (b5)] and yellow [flavin mononucleotide (FMN)] and glutathione-S-transferase (GST) fusions being eluted off nickel agarose with 300 mM imidazole. The GST column is loaded with an equivalent quantity of hexahistidine (His_6)-tagged GST. (B) Removal of red (b5) and yellow (FMN) tags by thrombin cleavage. C, cut; U, uncut. Positions of GST, b5, and FMN are shown on the right. (C) Spectrophotometric detection of the yellow FMN-tagged GST fractions. The FMN-GST fusion was purified by fast protein liquid chromatography (FPLC) as described in Materials and Methods. Absorbance (A) was monitored at 450 nm. The circles represent FMN-GST, and the squares represent His_6 -GST. Shown immediately below the graph are the elution fractions (top 2 rows) of FMN-GST and GST and the corresponding sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE; bottom two rows) of equivalent volumes of eluate (2 μL).

systems. Furthermore, the concept of visible-tagged polypeptides could be extended to any nucleotide sequence that produces a colored peptide. For instance, the small prokaryotic flavodoxins that are structurally related to the FMN domain of CPR are alternative yellow fusion tags that we are currently investigating. Finally, fusing a polypeptide tag to the N terminus may be detrimental to the activity of some proteins. Thus, we also intend to examine if Rainbow tags are effective when used as C-terminal fusions.

Fusion tags are most commonly employed as affinity ligands to facilitate protein purification. Common examples include His₆ (8,9), GST (10–12), and maltose-binding protein (13,14). Visual tagging systems such as green fluorescent protein (GFP) or β -galactosidase are well known and used extensively in most areas of biological research (15–19). However, to be visually effective, these systems tend to require an external energy source or catalytic reaction. In contrast to these systems, flavin- and heme-binding proteins are easily visible with the naked eye.

Surprisingly, simple visible tags or tracker systems are not yet readily available to facilitate scale-up expression and purification of colorless molecules. The visual tagging system described here is designed to plug this gap. One important area where this would be applicable is in the translation of genome data into functional products and high-throughput protein structure determination by X-ray crystallography or nuclear magnetic resonance (NMR) (20), where systems to monitor the production of soluble protein would be enormously beneficial.

ACKNOWLEDGMENTS

This work was supported by NESTech. We are grateful to Dr. Hilary Ranson from the Liverpool School of Tropical Medicine for the Anopheles gambiae b5 cDNA. The visual tag concept is the subject of patent application no. PCT/GB2004/003543.

COMPETING INTERESTS STATEMENT

The authors declare no competing interests.

REFERENCES

- Zhao, Q., S. Modi, G. Smith, M.J.I. Paine, C.R. Wolf, D. Tew, L.-Y. Lian, G.C.K. Roberts, and H.P.C. Driessen. 1999. Crystal structure of the FMN-binding domain of human cytochrome P450 reductase at 1.93 Å resolution. *Prot. Sci.* 8:298–306.
- Barsukov, I., S. Modi, L.-Y. Lian, K.H. Sze, M.J.I. Paine, C.R. Wolf, and G.C.K. Roberts. 1997. ¹H, ¹⁵N and ¹³C NMR resonance assignment, secondary structure and global fold of the FMN-binding domain of human cytochrome P450 reductase. *J. Biomed. NMR* 10:63–75.
- Voice, M.W., Y. Zhang, C.R. Wolf, B. Burchell, and T. Friedberg. 1999. Effects of human cytochrome b5 on CYP3A4 activity and stability in vivo. *Arch. Biochem. Biophys.* 366:116–124.
- Nikou, D., H. Ranson, and J. Hemingway. 2003. An adult-specific CYP6 P450 gene is overexpressed in a pyrethroid-resistant strain of the malaria vector, *Anopheles gambiae*. *Gene* 318:91–102.
- McLellan, L.I., L.A. Kerr, A.D. Cronshaw, and J.D. Hayes. 1991. Regulation of mouse glutathione S-transferases by chemoprotectors. Molecular evidence for the existence of three distinct alpha-class glutathione S-transferase subunits, Ya1, Ya2, and Ya3, in mouse liver. *Biochem. J.* 276:461–469.
- Munro, A., M. Noble, L. Robledo, S. Daff, and S. Chapman. 2001. Determination of the redox properties of human NADPH-cytochrome P450 reductase. *Biochemistry* 40:1956–1963.
- Davis, G., C. Elisee, D. Newham, and R. Harrison. 1998. New fusion protein systems designed to give soluble expression in *Escherichia coli*. *Biotechnol. Bioeng.* 65:382–388.
- Franklin, A., T. Jetton, K. Shelton, and M. Magnuson. 1994. BZP, a novel serum-responsive zinc finger protein that inhibits gene transcription. *Mol. Cell. Biol.* 14:6773–6788.
- Winter, M. and M. Welsh. 1997. Stimulation of CFTR activity by its phosphorylated R domain. *Nature* 389:294–296.
- Parker, M., M. Lo Bello, and G. Federici. 1990. Crystallization of glutathione S-transferase from human placenta. *J. Mol. Biol.* 213:221–222.
- Ji, X., P. Zhang, R. Armstrong, and G. Gililand. 1992. The three-dimensional structure of a glutathione S-transferase from the Mu gene class. Structural analysis of the binary complex of isoenzyme 3-3 and glutathione at 2.2 Å resolution. *Biochemistry* 31:10169–10184.
- Maru, Y., D. Afar, O. Witte, and M. Shibuya. 1996. The dimerization property of glutathione S-transferase partially reactivates Bcr-Abl lacking the oligomerization domain. *J. Biol. Chem.* 271:15353–15357.
- Di Guan, C., P. Li, P. Riggs, and H. Inouye. 1988. Vectors that facilitate the expression and purification of foreign peptides in *Escherichia coli* by fusion to maltose-binding protein. *Gene* 67:21–30.
- Maina, C., P. Riggs, A. Grande, B. Slatko, L. Moran, J. Tagliamonte, L. McReynolds, and C. Guan. 1988. An *Escherichia coli* vector to express and purify foreign proteins by fusions to and separation from maltose-binding protein. *Gene* 74:365–373.
- Chalfie, M., Y. Tu, G. Euskirchen, W. Ward, and D. Prasher. 1994. Green fluorescent protein as a marker for gene expression. *Science* 263:802–805.
- Cramer, A., E. Whitehorn, E. Tate, and W. Stemmer. 1996. Improved green fluorescent protein by molecular evolution using DNA shuffling. *Nat. Biotechnol.* 14:315–319.
- Poppenberg, L., K. Friehs, and E. Flaschel. 1997. The green fluorescent protein is a versatile reporter for bioprocess monitoring. *J. Biotechnol.* 58:79–88.
- Paramban, R.I., R.C. Bugos, and W.W. Su. 2004. Engineering green fluorescent protein as a dual functional tag. *Biotechnol. Bioeng.* 86:687–697.
- Muller-Hill, B. and J. Kania. 1974. Lac repressor can be fused to β -galactosidase. *Nature* 249:561–563.
- Heinemann, U., G. Illing, and H. Oschkinat. 2001. High-throughput three-dimensional protein structure determination. *Curr. Opin. Biotechnol.* 12:348–354.

Received 8 September 2004; accepted 5 October 2004.

Address correspondence to Mark J.I. Paine, Biomedical Research Centre, University of Dundee, Ninewells Hospital and Medical School, Dundee, DD1 9SY, UK. e-mail: m.j.paine@dundee.ac.uk