**BMC Biochemistry**

CrossMark

# Structural plasticity of green fluorescent protein to amino acid deletions and fluorescence rescue by folding-enhancing mutations

Shu-su Liu[1], Xuan Wei[1], Xue Dong[1], Liang Xu[1], Jia Liu[1,2]* and Biao Jiang[1]*

## Abstract

**Background:** Green fluorescent protein (GFP) and its derivative fluorescent proteins (FPs) are among the most commonly used reporter systems for studying gene expression and protein interaction in biomedical research. Most commercially available FPs have been optimized for their oligomerization state to prevent potential structural constraints that may interfere with the native function of fused proteins. Other approach to reducing structural constraints may include minimizing the structure of GFPs. Previous studies in an enhanced GFP variant (EGFP) identified a series of deletions that can retain GFP fluorescence. In this study, we interrogated the structural plasticity of a UV-optimized GFP variant (GFP$_{UV}$) to amino acid deletions, characterized the effects of deletions and explored the feasibility of rescuing the fluorescence of deletion mutants using folding-enhancing mutations.

**Methods:** Transposon mutagenesis was used to screen amino acid deletions in GFP that led to fluorescent and nonfluorescent phenotypes. The fluorescent GFP mutants were characterized for their whole-cell fluorescence and fraction soluble. Fluorescent GFP mutants with internal deletions were purified and characterized for their spectral and folding properties. Folding-ehancing mutations were introduced to deletion mutants to rescue their compromised fluorescence.

**Results:** We identified twelve amino acid deletions that can retain the fluorescence of GFP$_{UV}$. Seven of these deletions are either at the N- or C- terminus, while the other five are located at internal helices or strands. Further analysis suggested that the five internal deletions diminished the efficiency of protein folding and chromophore maturation. Protein expression under hypothermic condition or incorporation of folding-enhancing mutations could rescue the compromised fluorescence of deletion mutants. In addition, we generated dual deletion mutants that can retain GFP fluorescence.

**Conclusion:** Our results suggested that a "size-minimized" GFP may be developed by iterative incorporation of amino acid deletions, followed by fluorescence rescue with folding-enhancing mutations.

**Keywords:** Green fluorescent protein (GFP), Transposon mutagenesis, Amino acid deletions, Protein folding, Chromophore maturation

## Background

The discovery and application of green fluorescent proteins (GFP) have revolutionized biomedical research during the past several decades. GFP was identified in jellyfish *Aequorea victoria* in the 1960s [1] and then isolated and characterized by Prasher *et al.* in 1992 [2]. Soon after this, GFP was adapted as a fluorescent tag to unravel the details of cellular events, such as protein localization, trafficking and interaction [3]. Extensive engineering studies resulted in a panel of fluorescent

protein (FP) variants with a wide range of spectral properties [4]. Other properties such as brightness, cytotoxicity and photostability have been also optimized to facilitate the application of FPs.

Notably, considerable efforts have been made to optimize the oligomerization state of FPs. Many of the naturally occurring FPs are dimeric or tetrameric [4]. Oligomerization does not limit the application of GFPs as reporters for gene expression, but may interfere with the native function of fused proteins. Tsien's group first demonstrated that tetrameric *Discosoma sp.* red fluorescent protein (DsRed) could be engineered to be monomeric [5]. In this study, they first disabled

* Correspondence: liujia@shanghaitech.edu.cn; jiangbiao@shanghaitech.edu.cn
[1]Shanghai Institute for Advanced Immunochemical Studies, ShanghaiTech University, Shanghai, China
Full list of author information is available at the end of the article

Liu *et al. BMC Biochemistry* (2015) 16:17

Page 2 of 11

the formation of tetrameric structure by site-directed mutagenesis and then used random mutagenesis to rescue the fluorescence. It is now known that many oligomeric FPs can be converted into monomers by point mutations without appreciable deleterious effects [6].

Oligomerization rarely caused troubles for the native function of GFPs, however it may introduce structural constraints and unfavorably affect the role of GFPs as a genetic tag for biological applications such as protein-protein interaction and subcellular localization. Despite the extensive optimization, the artifacts associated with GFP oligomerization still require careful assessment in each experiment [7]. In addition to traditional strategy, an alternative approach to reducing the structural constraints associated with GFP tag is to develop a size-minimized construct. This is extremely challenging in the case of GFP because it is generally considered as a "compact" protein due to the well-shaped β-can structure [8–10].

The wild-type *A. victoria* GFP is a 27 kDa protein containing 238 amino acids. GFP and all its known variants adopt a β-can structure assembled by 11 antiparallel β-strands [3, 11]. Most strands are connected by small loops consisting of one to four amino acids. Two larger loops appear at positions 129 ~ 143 and 189 ~ 197. The chromophore of GFP is located in the central α helix, surrounded by β-strands. The top and bottom "lids" composed mainly of residues 74–91 and 128–145, respectively [11, 12].

Several early studies have been performed to understand the tolerance of GFP to amino acid deletions. GFP is well tolerant to deletions at N or C terminus [8, 9]. One study showed that the minimal domain required for the fluorescence of GFP contains residues from 2 to 232 [8]. An enhanced version of GFP (EGFP, mutation F64L/S65T) can still be fluorescent with 5 amino acids deletion from N terminus or 10 amino acid deletion from C terminus [9]. However, existing evidence suggested that GFP is very sensitive to deletions at internal positions. Deletions of the two large loops or the two small helices eliminated the fluorescence of EGFP [9]. Targeted deletion analysis of the longest loop (129 ~ 143) 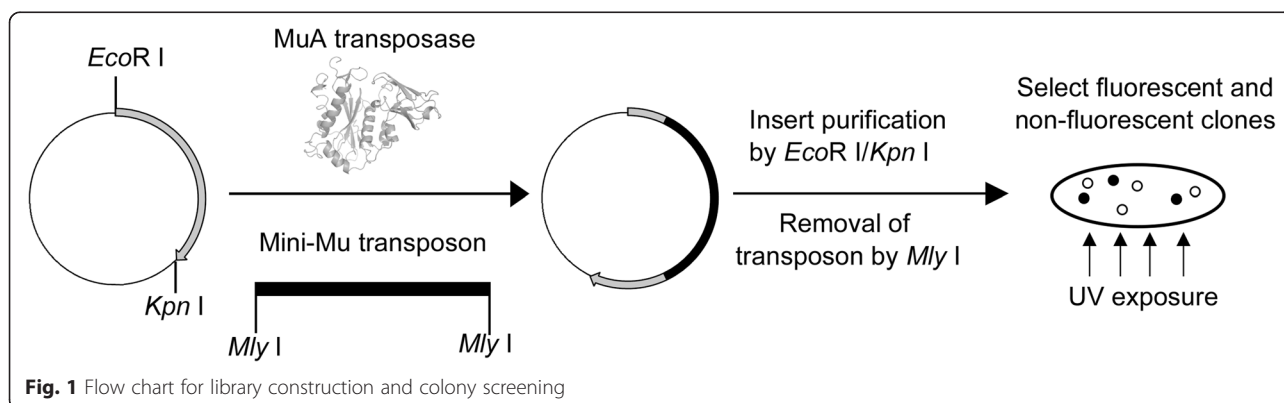in SuperGlo GFP (sgGFP, mutation F64L/S65C/I167T) showed that I128Δ and D129Δ are the only two single deletions that can retain GFP fluorescence [10].

Nevertheless, all early studies focused on analyzing deletions at the termini or large loops of GFP. In recent studies, Jones *et al.* used a transposon-based mutagenesis method [13] to investigate the global structural plasticity of EGFP to amino acid deletions [14]. A series of deletion mutants that can retain GFP fluorescence have been identified and some mutants even exhibited improved cellular fluorescence [14, 15]. These studies shed the light on the important roles of amino acid deletions on GFP fluorescence and indicated that incorporation of amino acid deletions might be a feasible approach to the development of size-minimized GFP construct. In this study, we created a set of $GFP_{UV}$ deletion mutants using a similar approach, characterized their properties in a comprehensive manner and explored the feasibility of rescuing the fluorescence of deletion mutants by introducing folding-enhancing mutations.

## Results

### Transposon-mediated deletion mutagenesis and colony screening

The transposon mutagenesis used in this study was described by Jones *et al.* [13]. It relies on an *in vitro* transposition reaction using MuA transposase and a mini-Mu transposon DNA [16] with engineered *Mly* I recognition site at each end [13]. In the presence of the donor transposon DNA, MuA transposase cleaves the acceptor DNA at a random position in a five nucleotide staggered manner. Following transposon insertion and transformation of the transposition product into bacteria, the five nucleotide overhang is repaired by bacterial DNA repair machinery, leading to duplication of the five nucleotides at the transposon insertion site. Subsequent removal of the transposon DNA using *Mly* I enzyme cleaves four nucleotide from the acceptor DNA at each end of transposon, resulting in a net 3 bp deletion from the target DNA (Fig. 1).



**Fig. 1** Flow chart for library construction and colony screening

Liu *et al. BMC Biochemistry* (2015) 16:17

Page 3 of 11

In this study, we performed transposon reaction using *Mly* I Mu transposon [13] as the donor DNA and pGFP$_{UV}$ plasmid as the acceptor DNA. We collected 38,000 individual colonies from the transformation of transposition product, which were sufficient to cover more than 95 % of all possible insertion sites even considering the minor substrate preference of the transposon system [16]. We next used *EcoR* I/*Kpn* I to isolate the transposon inserts within the GFP$_{UV}$ gene (Additional file 1: Figure S1). The final deletion library was obtained by releasing the transposon DNA from pGFP$_{UV}$ plasmid using *Mly* I restriction digestion.
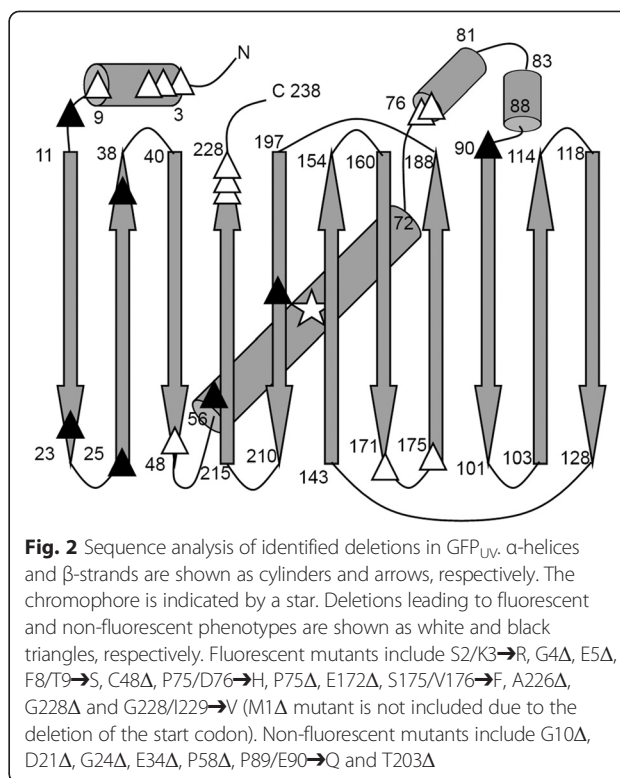
pGFP$_{UV}$ plasmid encodes a GFP variant with enhanced fluorescence under UV light (GFP$_{UV}$) [17]. pGFP$_{UV}$ plasmid constitutively expresses GFP$_{UV}$ protein and thus can be used directly for functional screen. In pilot experiments, we observed that some mutants exhibited considerable difference in fluorescence when grown at different temperature. We chose to perform the fluorescence screening at low temperature (20 °C). Forty fluorescent and 24 non-fluorescent colonies were identified and selected for sequencing for further analysis.

### Sequence analysis of deletion mutants

Sequencing results revealed that a small fraction of mutants contained four nucleotide deletion, which presumably resulted from excessive *Mly* I digestion (Additional file 2: Table S1). This result was consistent with previous study [18] and suggested that *Mly* I digestion might require further optimization. In total, we identified thirteen fluorescent and seven non-fluorescent mutants with unique deletions. Of the thirteen fluorescent mutants, eight contained deletions at N- or C-terminus whereas five carried deletions at internal positions. The five internal deletions were all located at the termini of α-helices or β-strands. For the seven deletions that eliminated GFP fluorescence, one was found in the N-terminal region, one was in the middle of β-strand and the rest were all found at the termini of α-helices or β-strands (Fig. 2). It is worth noting that the random nucleotide deletions may occur in two neighboring residues, leading to amino acid mutations in addition to deletions.
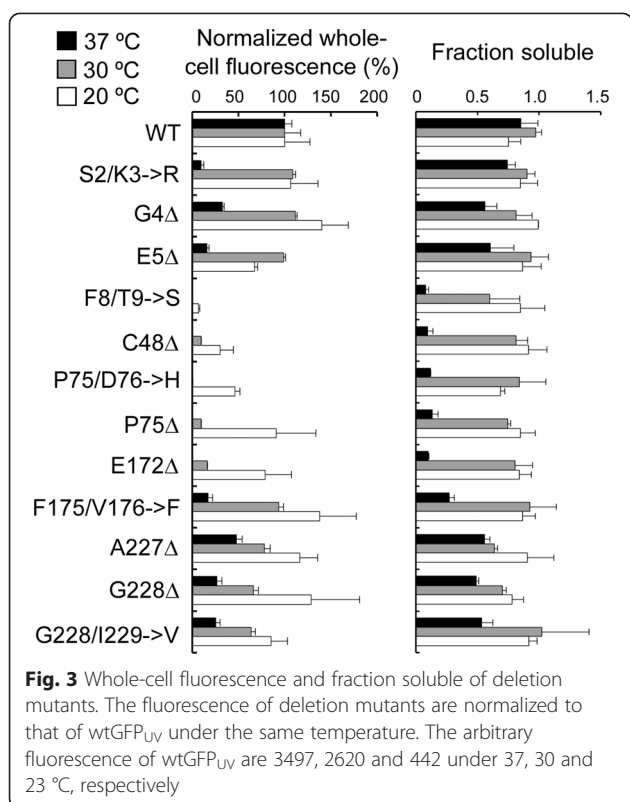
### Whole-cell fluorescence and fraction soluble of deletion mutants

In order to quantitatively analyse the identified fluorescent deletion mutants, we characterized their whole-cell fluorescence in the context of pGFP$_{UV}$ plasmid. As temperature may affect the fluorescence of some mutants, whole-cell fluorescence was determined under different temperature: 37, 30 and 20 °C. It was found that, for all temperature, terminal deletions generally retained higher degree of fluorescence compared with internal deletions. The fluorescence of all deletion mutants was



**Fig. 2** Sequence analysis of identified deletions in GFP$_{UV}$. α-helices and β-strands are shown as cylinders and arrows, respectively. The chromophore is indicated by a star. Deletions leading to fluorescent and non-fluorescent phenotypes are shown as white and black triangles, respectively. Fluorescent mutants include S2/K3➔R, G4Δ, E5Δ, F8/T9➔S, C48Δ, P75/D76➔H, P75Δ, E172Δ, S175/V176➔F, A226Δ, G228Δ and G228/I229➔V (M1Δ mutant is not included due to the deletion of the start codon). Non-fluorescent mutants include G10Δ, D21Δ, G24Δ, E34Δ, P58Δ, P89/E90➔Q and T203Δ

compromised at high temperature (37 °C). Decrease of temperature to 30 °C restored the fluorescence of most mutants except F8/T9➔S and P75/D76➔H. Expression of protein at 20 °C essentially recovered the fluorescence of all mutants (Fig. 3).

The recovery of fluorescence at low temperature suggested that these deletions might impair protein folding, which is a temperature-dependent process. In order to understand the effect of amino acid deletions on protein folding of GFP$_{UV}$, we sought to determine the fraction of soluble GFP$_{UV}$ at different temperature using a previously described method [19]. Wild-type and mutant GFP$_{UV}$ was expressed from pGFP$_{UV}$ plasmid and the bacterial cells were then lysed to assess the GFP present in the entire or soluble fraction of the cell extracts (Additional file 3: Figure S2). The soluble fraction of deletion mutants was clearly temperature dependent (Fig. 3) and plausibly correlated with the whole-cell fluorescence. For example, the mutants with no fluorescence at 37 °C (F8/T9➔S, C48Δ, P75/D76➔H, P75Δ and E172Δ) had dramatically reduced level of soluble fraction of proteins compared with other mutants. However, factors other than protein folding need be considered to explain the reduced fluorescence of deletion mutants. In one case, P75/D76➔H mutant exhibited similar level of soluble fraction of proteins at 30 and 20 °C, but it was only fluorescent at 20 °C. It is known that both

Liu *et al. BMC Biochemistry* (2015) 16:17

Page 4 of 11



**Fig. 3** Whole-cell fluorescence and fraction soluble of deletion mutants. The fluorescence of deletion mutants are normalized to that of wtGFP$_{UV}$ under the same temperature. The arbitrary fluorescence of wtGFP$_{UV}$ are 3497, 2620 and 442 under 37, 30 and 23 °C, respectively

protein folding and chromophore maturation of GFP are temperature dependent [20]. Therefore, we intended to purify the deletion mutants and characterize their biochemical and biophysical properties. We chose to analyze the fluorescent mutants with internal deletions. Characterization of these mutants and search of feasible means to rescuing their compromised fluorescence might be the first step toward generating a size-minimized GFP.

### Protein purification and the effect of amino acid deletions on spectral properties

Deletion mutants C48Δ, P75/D76➔H, P75Δ, E172Δ, S175/V176➔F were purified to more than 95 % homogeneity (Fig. 4a). The emission spectra of all variants under 397 and 475 nm excitation can be superimposed with that of wild-type GFP$_{UV}$ (wtGFP$_{UV}$) (Fig. 4b), suggesting that the chromophores of these mutants did not have structural alteration. For excitation spectra, two peaks (397 and 475 nm) were observed for all samples. These deletion mutants showed variations in the intensity ratio between the major and minor peaks. For example, P75Δ showed slightly increased minor peak whereas the minor peak in S175/V176➔F mutant was considerably reduced (Fig. 4c). The 395 nm and 475 nm excitation peaks represent neutral and anionic chromophore, respectively [20]. Mutations that alter the ratio of the two chromophore species are frequently found in

GFP variants. In one case, enhanced GFP (EGFP) only has a single excitation peak at 475 nm [21] because S65T mutation transform the chromophore to be completely ionized [20]. It is interesting to observe in this study that amino acid deletions, in addition to substitutions, can have impact on the ionization state of chromophore.

We next determined the extinction coefficients at 397 and 475 nm (ε$_{397}$ and ε$_{475}$) and quantum yield of the GFP$_{UV}$ variants. Consistent with the spectra, ε$_{397}$ of all deletion mutant presented no appreciable difference from that of wtGFP$_{UV}$, while modest variation was observed for ε$_{475}$ (Table 1). We also found that all variants have very similar quantum yield to wtGFP$_{UV}$ protein, suggesting that these internal amino acid deletions did not alter the intrinsic brightness of GFP (Table 1). Taken together, these results indicated that the diminished fluorescence of deletion mutants is mainly attributed to the disrupted folding process, rather than intrinsic brightness.

### Characterization of refolding kinetics and chromophore maturation of deletion mutants

As whole-cell lysis study indicated that amino acid deletions may affect protein folding, we next sought to determine the refolding kinetics of deletion variants using purified proteins. Previous study suggested that the refolding process of GFP$_{UV}$ could be fitted equally well with sequential or parallel model [22]. Here we performed refolding experiments using guanidine hydrochloride and recorded the recovered fluorescence as described [22]. The data were fitted into a double exponential equation with a parallel model using Prism 4.0 (Fig. 5). The $k_1$ and $k_2$ of wtGFP$_{UV}$ were generally consistent with the reported values [23]. Deletion C48Δ led to notable decrease in $k_1$ value, suggesting that this deletion was involved in the major folding pathway of GFP. It was also found that P75/D76➔H and P75Δ reduced the refolding rate in the slow phase (decreased $k_2$ value) (Fig. 5). Interestingly, the *cis-trans* isomerization of P75 was proposed as a rate-limiting step in the folding process of GFP [24]. It was also noted that deletion E172Δ impacted neither fast nor slow phase of protein folding. This observation suggested that factors other than protein folding may account for the compromised fluorescence in deletion mutants.

Instructed by these results, we then analyzed the efficiency of chromophore maturation of deletion mutants. In this experiment, we denatured 7.0 μM protein of each sample and determined their absorbance at 450 nm (Fig. 6). As the extinction coefficient of matured chromophore at 450 nm has been determined in previous study [25], the concentration of matured chromophore can be calculated using Beer's law. It is evident
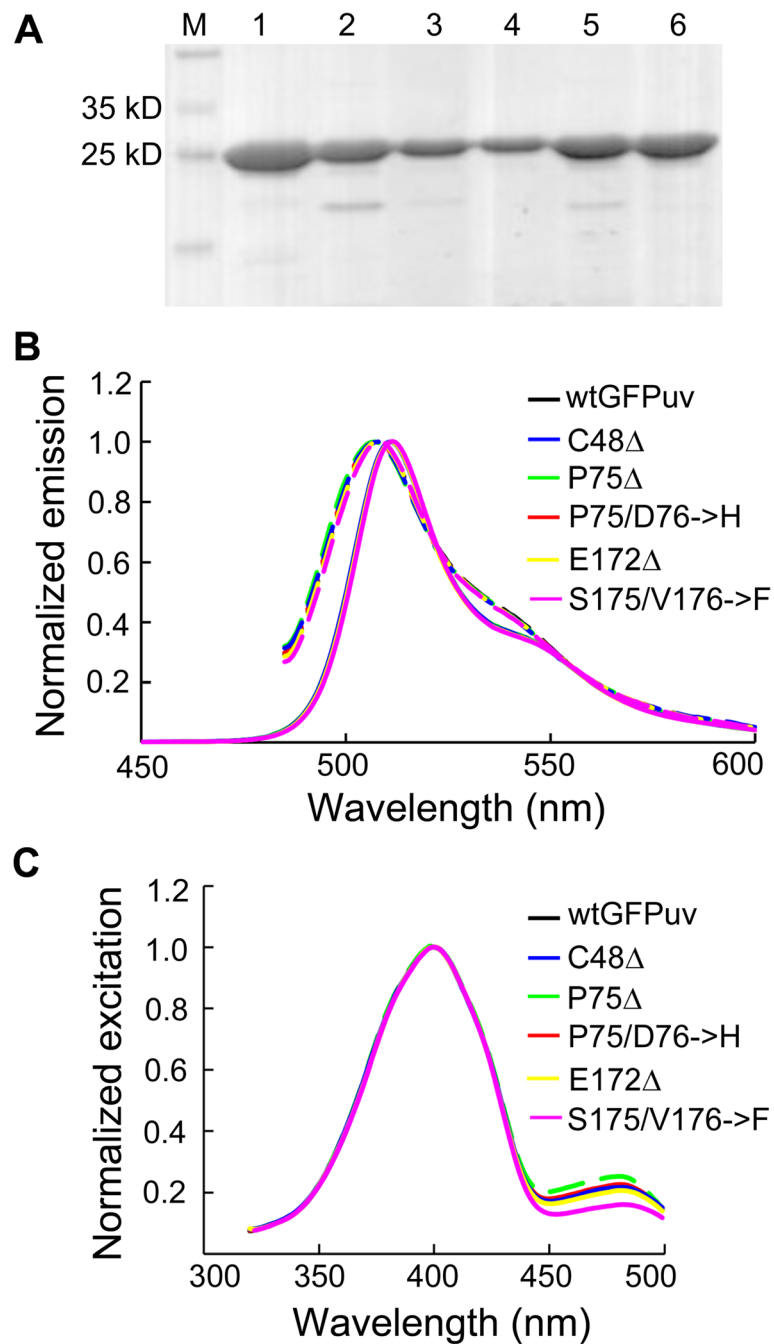
Liu *et al. BMC Biochemistry* (2015) 16:17

Page 5 of 11



**Fig. 4** Excitation and emission spectra of purified GPF$_{UV}$ variants. **a** Purified wtGFP$_{UV}$ and deletion mutants, resolved on 12 % SDS-PAGE. M, marker. Lane 1 to 6, wtGFP$_{UV}$, C48Δ, P75Δ, P75/D76➔H, E172Δ and S175/V176➔F, respectively. **b** Normalized excitation spectra. **c** Normalized emission spectra. Note that the spectra of these variants are largely overlapped

that all deletions reduced the efficiency of chromophore maturation (Fig. 6). It is worth mentioning that the protein samples used for this experiment was purified at 20 °C. Previous analysis of the whole-cell lysis showed that the five mutant with internal deletions displayed only trace fluorescence but had considerable amount of soluble proteins when expressed at 30 °C. According to

the analysis of chromophore, this discrepancy is likely attributed to the impaired chromophore maturation of GFP proteins present in the soluble fraction.

### Fluorescence rescue by folding-enhancing mutations

Based on the above results, the compromised fluorescence of GFP deletion mutants resulted from disrupted

Liu *et al. BMC Biochemistry* (2015) 16:17

Page 6 of 11

**Table 1** Extinction coefficient and quantum yield of deletion mutants

| | $\varepsilon_{397}$ (M$^{-1}$ cm$^{-1}$) | $\varepsilon_{475}$ (M$^{-1}$ cm$^{-1}$) | Quantum yield ($\lambda_{ex}$ = 397 nm) |
|---|---|---|---|
| wtGFP$_{UV}$ | 29600 | 7500 | 0.79[a] |
| C48Δ | 31600 | 8500 | 0.80 |
| P75/D76➔H | 29700 | 6800 | 0.79 |
| P75Δ | 29400 | 7100 | 0.79 |
| E172Δ | 29300 | 7600 | 0.81 |
| S175/V176➔F | 30900 | 5000 | 0.79 |

[a]Patterson *et al.,* [26]

protein folding and chromophore maturation. As a first step toward generating a size-minimized GFP, we speculated whether the defects in these deletion mutants can be restored by folding-enhancing mutations. We chose to test the effect of two previously described folding mutations F64L [26] and S30R [19] on the mutants with internal deletions. Remarkably, F64L recovered a considerable fraction of fluorescence in all variants except P75Δ, whereas S30R only improved the fluorescence of S175/V176➔F (Fig. 7a). This result illustrated that folding-enhancing mutations can rescue the compromised fluorescence of deletion mutants and that some mutations may be more effective than others. Encouraged by these results, we attempted to use F64L mutation to rescue mutants carrying multiple internal deletions. We first generated nine dual deletion mutants by combining the identified five internal deletions. As expected, GFP fluorescence was completely abolished by dual deletions, even with prolonged incubation at 4 °C. We next incorporated F64L mutation into these dual mutants. Although F64L could not recover GFP fluorescence at 37 °C, it successfully restored the fluorescence of all mutants to a considerable degree when proteins were

expressed at 20 °C (Fig. 7b). To the best of our knowledge, it has not been reported before that GFP mutants carrying multiple internal deletions can retain fluorescence.

## Discussion

Amino acid deletion, insertion and substitution are all important mechanisms for protein evolution in nature. While amino acid substitutions are responsible for the alteration of protein properties in many cases, insertions and deletions (indels) are critical for generating length variation of proteins [27, 28]. Although less commonly used, indels are efficient means for improving protein functions [29, 30]. Established methods for introducing random indels into proteins include RID mutagenesis reported by Murakami *et al.* [31] and RAISE mutagenesis described by Fujii *et al.* [32]. Notably, Jones *et al.* developed a facile transposon-based method for introducing random triplet nucleotide deletions within the full sequence of a given protein [13] and has gone to employ this method to investigate the structural tolerance of TEM-1 β-lactamase to amino acid deletions [33]. This method has been soon adapted for generating nucleotide substitutions with canonical [34] and non-canonical [18, 35–37] amino acids as well as domain insertions [38, 39]. Subsequent modification of this method allowed generation of multiple in-frame codon mutations [40, 41]. In this study, we used this transposon mutagenesis [13] to interrogate the global plasticity of GFP to amino acid deletions.

Early studies suggested that GFP is well tolerated to amino acid deletions at termini [8, 9] but not to those at internal positions [10]. However, recent deletion analysis of EGFP identified a set of internal deletions that can retain or even improve GFP fluorescence [14, 15]. Herein, we generated a deletion library of GFP$_{UV}$ [17] using a similar approach. This GFP variant (mutation F99S/M153T/V163A) has enhanced absorbance under UV light and thus functional screen can be performed directly by visual inspection
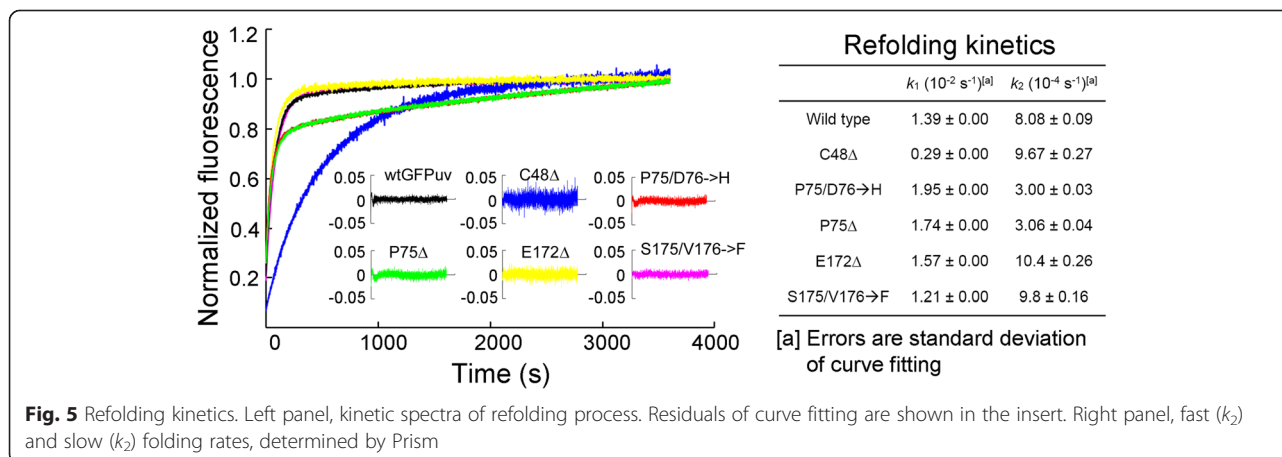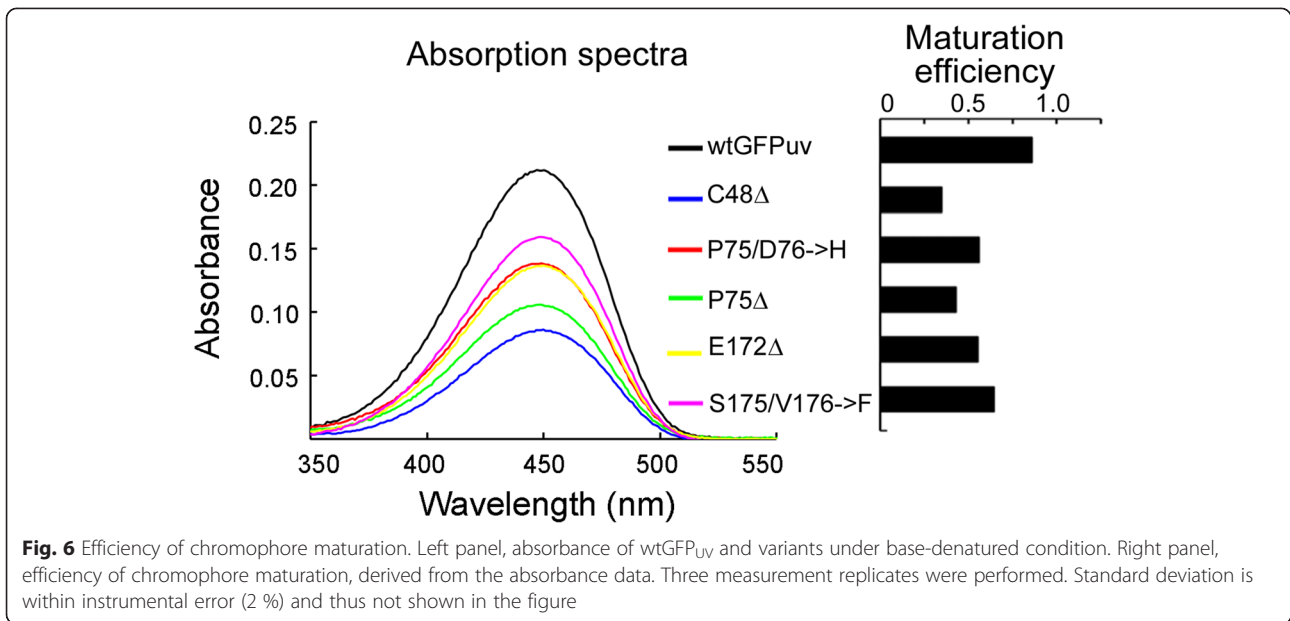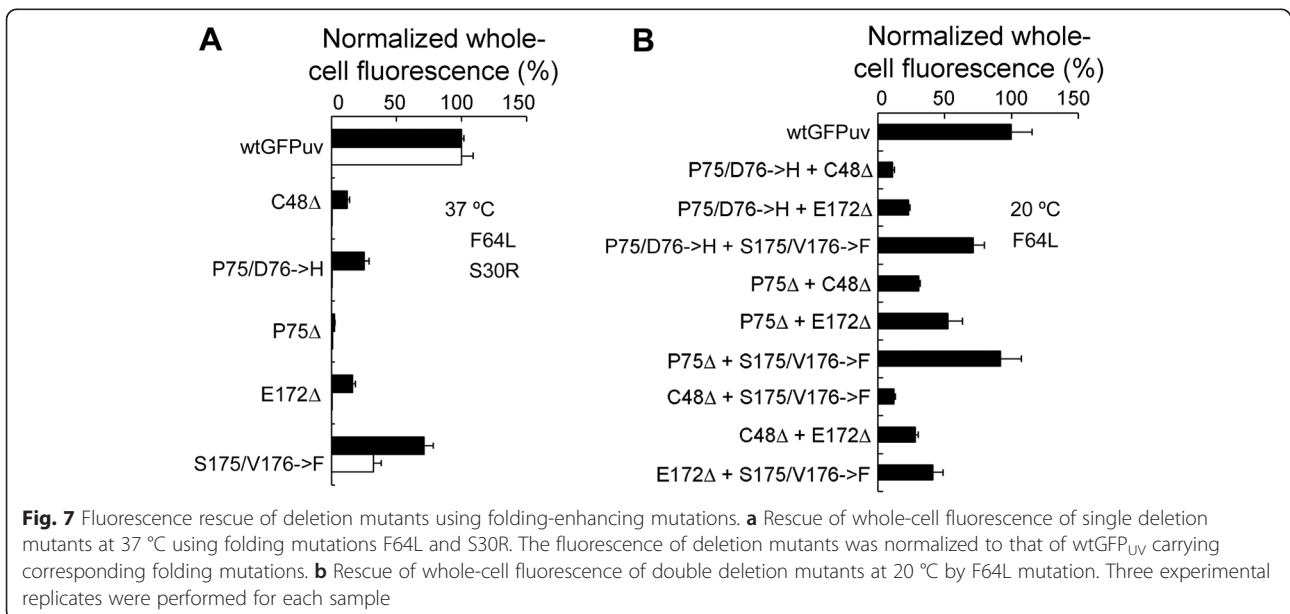


## Refolding kinetics

| | $k_1$ (10$^{-2}$ s$^{-1}$)[a] | $k_2$ (10$^{-4}$ s$^{-1}$)[a] |
|---|---|---|
| Wild type | 1.39 ± 0.00 | 8.08 ± 0.09 |
| C48Δ | 0.29 ± 0.00 | 9.67 ± 0.27 |
| P75/D76➔H | 1.95 ± 0.00 | 3.00 ± 0.03 |
| P75Δ | 1.74 ± 0.00 | 3.06 ± 0.04 |
| E172Δ | 1.57 ± 0.00 | 10.4 ± 0.26 |
| S175/V176➔F | 1.21 ± 0.00 | 9.8 ± 0.16 |

[a] Errors are standard deviation of curve fitting

**Fig. 5** Refolding kinetics. Left panel, kinetic spectra of refolding process. Residuals of curve fitting are shown in the insert. Right panel, fast ($k_2$) and slow ($k_2$) folding rates, determined by Prism

Liu *et al. BMC Biochemistry* (2015) 16:17

Page 7 of 11



**Fig. 6** Efficiency of chromophore maturation. Left panel, absorbance of wtGFP$_{UV}$ and variants under base-denatured condition. Right panel, efficiency of chromophore maturation, derived from the absorbance data. Three measurement replicates were performed. Standard deviation is within instrumental error (2 %) and thus not shown in the figure

using UV excitation. We identified 13 fluorescent and 7 non-fluorescent mutants with unique deletions. We first analysed the whole-cell fluorescence of the mutants that retained GFP fluorescence and found that their fluorescence is strongly temperature dependent. Remarkably, several mutants (G4Δ, F175/V176➔F, A227Δ and G228Δ) that showed decreased fluorescence at high temperature (37 ℃) are even more fluorescent than wtGFP$_{UV}$ when the temperature is reduced to 4 ℃. This suggested that these deletions may enhance GFP fluorescence at the cost of protein stability. Further analysis of the fraction of soluble GFP in the whole-cell lysis suggested that the impaired

fluorescence of mutants at high temperature was attributed to the disrupted protein folding and that other process need be considered to explain the loss of fluorescence. Characterization of purified GFP proteins carrying internal deletions confirmed the decreased refolding rates in some variants. Additional analysis showed that all internal deletions significantly reduced the efficiency of chromophore maturation. Interestingly, these mutants have the same intrinsic brightness (quantum yield) with wtGFP$_{UV}$. This finding indicated that these mutants used the same autocatalytic process to form GFP chromophore. The present and previous [14, 15] results suggested that it



**Fig. 7** Fluorescence rescue of deletion mutants using folding-enhancing mutations. **a** Rescue of whole-cell fluorescence of single deletion mutants at 37 ℃ using folding mutations F64L and S30R. The fluorescence of deletion mutants was normalized to that of wtGFP$_{UV}$ carrying corresponding folding mutations. **b** Rescue of whole-cell fluorescence of double deletion mutants at 20 ℃ by F64L mutation. Three experimental replicates were performed for each sample

Liu *et al. BMC Biochemistry* (2015) 16:17

Page 8 of 11

is feasible to recover protein folding and chromophore maturation by introducing beneficial mutations.

Next we attempted to recover the fluorescence of deletion mutants by employing two folding mutations previously identified in GFP. It was shown that F64L mutation successfully rescued the fluorescence of most mutants whereas S30R was less effective. The different roles of F64L and S30R can be explained by their distinct functions in GFP. F64L is a central mutation in close proximity to chromophore and many studies have demonstrated its critical role in protein folding and chromophore maturation [21, 24, 26]. S30R, in contrast, is a distal mutation that enhanced protein folding in an indirect manner [19]. In addition, we analysed the effect of F64L on the non-fluorescent deletion mutants (G10Δ, D21Δ, G24Δ, E34Δ, P58Δ, P89/E90➜Q and T203Δ) and found that only G10Δ can recover trace fluorescence at 20 °C (data not shown). It is also worth noting that GFP variants containing F64L mutation can be also used as the starting template for engineering experiments. Previous studies suggested that in the presence of F64L, more deletions may be found to retain GFP fluorescence [14, 15].

Toward the goal of generating a size-minimized GFP, we next set to generate dual deletion mutants and then determined the effects of F64L on these variants. It was found that F64L was capable of rescuing their fluorescence at 20 °C but not that at 37 °C. We also generated GFP$_{UV}$ variants containing triple internal deletions and found that their fluorescence could not be restored by F64L. This result indicated that additional folding mutations or global optimization might be required for compensating the deleterious effects of multiple amino acid deletions. Recent study highlighted the importance of epistasis in protein evolution [42], therefore it may be also interesting to combine the beneficial deletions (G4Δ, F175/V176➜F, A227Δ and G228Δ) to explore their interactions.

## Conclusion
We have explored the structural plasticity of GFP to amino acid deletion on a whole-protein scale and showed that compromised fluorescence associated with deletions can be recovered by introducing folding-enhancing mutations. Our results suggested that a "size-optimized" GFP might be developed by iterative deletions of amino acids, followed by fluorescence rescue using folding mutations.

## Methods
### Construction of deletion library
pGFP$_{UV}$ vector (Clontech, Mountain View, CA) was used as the transposon target plasmid for *in vitro* transposition reaction. GFP$_{UV}$ protein carrying a 24 amino acid N-terminal tag is constitutively expressed from this

vector. All of the four *Mly* I sites (249, 2334, 2836 and 3322) in pGFP$_{UV}$ vector were removed by site-directed mutagenesis without changing the amino acid sequence (see Additional file 4: Table S2 for primers). The engineered Mu transposon bearing chloramphenicol-resistant gene and *Mly* I sites was constructed as described [13]. Transposon DNA was released from pUC19 vector by *Bgl* II digestion, gel purified and then resolved on a 1 % agarose gel to determine DNA concentration and purity. Transposition reaction was performed in a 20 μL mixture containing 50 mM Tris-acetate, pH 7.5, 150 mM potassium acetate, 10 mM magnesium acetate, 4 mM spermidine, 570 ng of pGFPuv vector, 140 ng of transposon DNA (~1.3 molar excess) and 1 unit of HyperMu MuA transposase (Epicentre Biotechnologies, Madison, WI). The reaction was kept at 30 °C for 4 hrs and stopped by addition of 0.1 % SDS, followed by heat-inactivation at 70 °C for 10 min. The reaction product was transformed into chemically competent GeneHogs *Escherichia coli* cells and plated on LB agar containing 100 μg/mL ampicillin and 10 μg/mL chloramphenicol for selection of pGFP$_{UV}$ vector with transposon insertion. Approximately 38,000 colonies were collected and maxi-prepped to build the pGFP$_{UV}$-MuDel library.

pGFP$_{UV}$-transposon library DNA was digested with *EcoR* I/*Kpn* I, and the 2.0 kb DNA fragment corresponding to transposon-carrying GFP$_{UV}$ gene was re-ligated with the 2.6 kb vector backbone. This purified pGFP$_{UV}$-transposon library was then digested with *Mly* I to remove transposon DNA, leaving a three nucleotide scar at a random position of GFP$_{UV}$. The blunt-end intramolecular ligation was performed in a 20 μL reaction containing 50 mM Tris–HCl, pH7.5, 10 mM MgCl$_2$, 10 mM dithiothreitol (DTT) and 0.5 mM ATP, 300 ng DNA, 400 cohesive end units of T4 DNA ligase (NEB). The ligation product was transformed into chemically competent GeneHogs *E. coli* and the transformants were plated on LB agar containing 100 μg/mL ampicillin. In total, 10,000 colonies were collected, and maxi-prepped to build the triplet nucleotide deletion library.

### Screening for deletion mutants
The deletion library DNA was transformed into GeneHogs and the transformants were plated on LB agar supplemented with 100 μg/mL ampicillin at a density of 500 colonies per plate. Transformed cells were grown at 20 °C for 30 h and screened for fluorescence by visual inspection using Spectronics (Westbury, New York) model TC312E UV transilluminator under 310 nm wavelength. Forty fluorescent and 24 non-fluorescent colonies were selected and sequenced for further characterization.

Liu *et al. BMC Biochemistry* (2015) 16:17

Page 9 of 11

## Liquid culture whole-cell fluorescence

Plasmids encoding wild-type GFP$_{UV}$ (wtGFP$_{UV}$) and deletion mutants were transformed into GeneHogs and plated on LB agar containing 100 µg/mL ampicillin at a density of ~200 colonies per plate. Single colonies with a diameter of ~0.5 mm were inoculated into 2 mL of liquid LB medium supplemented with 100 µg/mL ampicillin and grown at 37, 30 and 23 °C for 14, 18 and 25 h, respectively. Cells were harvested by centrifugation, washed with 500 µL TNG buffer (100 mM Tris–HCl, pH 7.5, 150 mM NaCl and 10 % glycerol) and then resuspended in 100 µL TNG buffer [19]. One milliliter of cell resuspension with an OD$_{600}$ of $0.150 \pm 0.003$ was prepared for fluorescence assay and the remaining cells in TNG buffer were stored at –20 °C for further experiments. Whole-cell fluorescence was determined using a model F4500 fluorescence spectrophotometer (Hitachi, Tokyo, Japan) as described [19]. The excitation and emission wavelength was set as 397 nm and 509 nm, respectively. Background fluorescence of empty GeneHogs cells was subtracted from each reading. All data were normalized to those of wtGFP$_{UV}$ under the same conditions. Three experimental replicates were performed for each sample.

## Protein expression and fraction soluble

For wild-type and mutant GFP$_{UV}$, 300 µL cell suspension in TNG buffer with an OD$_{600}$ of 0.100 was sonicated by Branson model 450 sonicator (Branson Ultrasonics, Danbury, CT) equipped with a 1/2 inch horn and a 1/8 in. tip. The power output and duty time were both set 50 %. Cell suspension was forced to two sequences of 10 pulse sonication with an interval of 3 min. Following sonication, 150 µL cell lysis was centrifuged at $12,000\,g$ for 10 min and the supernatant was transferred into a new tube. SDS loading samples were prepared in a 30 µL solution containing 15 µL crude cell lysis (whole protein) or supernatant (soluble fraction) and 15 µg bovine serum albumin (BSA) protein (Sigma, St. Louis, MO) as an internal standard. Protein samples were resolved in 12 % acrylamide SDS-PAGE gels and analyzed by Image J (http://rsbweb.nih.gov/ij/). GFP expression was quantified based on the density ratio of GFP and BSA bands as described [19]. The fraction soluble of each sample was extrapolated from the density ratio of soluble GFP present in supernatant and overall GFP present in crude cell lysis. Three experimental replicates were performed for each sample.

## Protein expression and purification and determination

wtGFP$_{UV}$ and mutants with internal deletions were subcloned into pET28b(+) vector (EMD Chemicals Inc., San Diego, CA) using primers CTAgctagcATGAGTAAAGG AGAAGAACTT (*Nhe* I site in lowercase) and CCCaa gcttTTATTTGTAGAGCTCATC (*Hind* III site in lowercase). pET28b vector encoding wild-type and mutant GFP$_{UV}$ with N-terminal His$_6$ tags were transformed into *E. coli* BL21 (DE3) cells (Stratagene Inc., La Jolla, CA) and spread on LB agar plates supplemented with 50 µg/mL kanamycin. A single colony was inoculated into 500 mL LB media supplemented with 50 µg/mL kanamycin and grown at 37 °C to an OD$_{600}$ of 0.8. Protein expression was induced with 1 mM isopropyl-β-D-1-galactopyronaside (IPTG) for 12 h at 20 °C. The next day, cells were centrifuged and resuspended in 25 mL binding buffer (100 mM HEPES, pH 7.5, 5 mM imidazole, 1 mM phenylmethanesulfonyl fluoride, PMSF). Cells were lysed by sonication and then centrifuged at 12,000 g for 30 min. Supernatant was transferred into a new tube and then loaded on to a column pre-packed with 1 mL HisLink resins (Promega Corporation, Madison, WI). The protein-bound resins were washed with 20 mL wash buffer (100 mM HEPES, pH 7.5, 20 mM imidazole) and then eluted with 5 mL elution buffer (100 mM HEPES, pH 7.5, 300 mM imidazole). Purified proteins were concentrated and buffer-exchanged into TNG buffer (100 mM Tris–HCl, pH 7.5, 150 mM NaCl and 10 % glycerol). The concentration of purified proteins were determined using Pierce BCA Protein Assay kit (Thermo Fisher Scientific, Rockford, IL).

## Biophysical properties of deletion mutants

The excitation and emission scan were performed using 6 µg/mL purified proteins in TNG buffer. An emission wavelength of 509 nm was used for excitation scan, while emission scan was performed using excitation wavelengths of 397 nm and 495 nm.

The concentration of matured GFP was determined using "base-denatured" method [25]. Briefly, 7 µM purified proteins were denatured in 0.1 M NaOH for 5 min at 25 °C. Absorption spectra from 300 nm to 600 nm were recorded. The concentration of matured GFPs was calculated from experimentally determined A$_{447}$ and previously reported ε$_{447}$ (44,000 M$^{-1}$ cm$^{-1}$ for denatured GFPs) [25] using Beer's Law. The efficiency of chromophore maturation was calculated based on the concentrations of matured and overall GFP.

The extinction coefficients of each sample at 397 nm (ε$_{397}$) and 495 nm (ε$_{495}$) were determined using Beer's Law: $A = \varepsilon \times l \times c$, where $A$ is experimentally determined absorbance, $l$ is path length and $c$ is the concentration of matured GFPs (= total GFP concentration × efficiency of chromophore maturation).

For the measurement of quantum yield, all protein samples were diluted to an OD$_{397}$ of 0.100 and then further diluted 100-fold with water. The emission spectra from 450 nm to 600 nm were scanned using an excitation wavelength of 397 nm. The quantum yield of wtGPF$_{UV}$ has been defined in a previous study as 0.79

Liu *et al. BMC Biochemistry* (2015) 16:17

Page 10 of 11

[7]. The quantum yield of mutants was calculated by comparing their integrated area of emission spectra with that of wtGPF$_{UV}$. Three experimental replicates were performed for each measurement. The instrumental error was estimated to be 2 %.

## Kinetic refolding experiments

To characterize the refolding ability of the deletion mutants, protein samples were denatured in the following solutions: 20 mM Tris–HCl, pH 7.5, 100 mM NaCl, 1 mM ethylenediaminetetraacetic acid (EDTA), 1 mM DTT, 0.20 mg/mL protein and 6 M guanidine hydrochloride (GdnHCl). Protein unfolding was processed at 25 °C for 24 h. Refolding process was initiated by diluting the denaturation solution 20 fold using the same buffer without GdnHCl. The fluorescence recovery was monitored at 25 °C for 60 min. The refolding data were fitted into a double exponential equation with a parallel refolding model using Prism 4.0 (GraphPad Software Inc., La Jolla, CA).

## Fluorescence rescue by folding-enhancing mutations

Mutations F64L [21] and S30R [19] were introduced into wtGFP$_{UV}$ and deletion mutants by site-directed mutagenesis (see Additional file 4: Table S2 for primers). The whole-cell fluorescence of F64L- or S30R-rescued mutants was assayed as described above. Mutants with double internal deletions were constructed using site-directed mutagenesis and mutation F64L was then introduced into the double deletion mutants. The rescued whole-cell fluorescence was assayed at 20 °C as described above.

## Additional files

**Additional file 1: Figure S1.** Isolation of transposon inserts inside GFP$_{UV}$ gene. (PDF 179 kb)

**Additional file 2: Table S1.** Sequencing results of 64 selected clones. (PDF 120 kb)

**Additional file 3: Figure S2.** Quantification of the soluble fraction of GFP$_{UV}$ samples using Image J. (PDF 152 kb)

**Additional file 4: Table S2.** Primers used in this study. (PDF 121 kb)

## Abbreviations

GFP: Green fluorescent protein; EGFP: Enhanced green fluorescent protein; GFP$_{UV}$: UV-optimized GFP variant; wtGFP$_{UV}$: Wide-type GFP$_{UV}$; sgGFP: SuperGlo GFP; BSA: Bovine serum albumin; IPTG: Isopropyl-β-D-1-galactopyronaside; PMSF: Phenylmethanesulfonyl fluoride; HEPES: 2-[4-(2-Hydroxyethyl)-1-piperazinyl] ethanesulfonic acid; EDTA: Ethylenediaminetetraacetic acid; GdnHCl: Guanidine hydrochloride; LB: Lysogeny broth.

## Competing interests

The authors declare that they have no competing interests.

## Authors' contributions

SSL, JL and BJ designed experiments. SSL, XW, XD and LX performed the experiments. SSL, JL and BJ analysed the data and wrote the paper. All authors read and approved the final manuscript.

## Author details

[1]Shanghai Institute for Advanced Immunochemical Studies, ShanghaiTech University, Shanghai, China. [2]Department of Chemistry and Biochemistry, University of Maryland, College Park, USA.

## References

1. Shimomura O, Johnson FH, Saiga Y. Extraction, purification and properties of aequorin, a bioluminescent protein from the luminous hydromedusan, Aequorea. J Cell Comp Physiol. 1962;59:223–39.
2. Prasher DC, Eckenrode VK, Ward WW, Prendergast FG, Cormier MJ. Primary structure of the *Aequorea victoria* green-fluorescent protein. Gene. 1992;111(2):229–33.
3. Shaner NC, Patterson GH, Davidson MW. Advances in fluorescent protein technology. J Cell Sci. 2007;120(Pt 24):4247–60.
4. Shaner NC, Steinbach PA, Tsien RY. A guide to choosing fluorescent proteins. Nat Methods. 2005;2(12):905–9.
5. Campbell RE, Tour O, Palmer AE, Steinbach PA, Baird GS, Zacharias DA, et al. A monomeric red fluorescent protein. Proc Natl Acad Sci U S A. 2002;99(12):7877–82.
6. Zacharias DA, Violin JD, Newton AC, Tsien RY. Partitioning of lipid-modified monomeric GFPs into membrane microdomains of live cells. Science. 2002;296(5569):913–6.
7. Voss U, Larrieu A, Wells DM. From jellyfish to biosensors: the use of fluorescent proteins in plants. Int J Dev Biol. 2013;57(6–8):525–33.
8. Dopf J, Horiagon TM. Deletion mapping of the *Aequorea victoria* green fluorescent protein. Gene. 1996;173(1 Spec No):39–44.
9. Li X, Zhang G, Ngo N, Zhao X, Kain SR, Huang CC. Deletions of the *Aequorea victoria* green fluorescent protein define the minimal domain required for fluorescence. J Biol Chem. 1997;272(45):28545–9.
10. Flores-Ramirez G, Rivera M, Morales-Pablos A, Osuna J, Soberon X, Gaytan P. The effect of amino acid deletions and substitutions in the longest loop of GFP. BMC Chem Biol. 2007;7:1.
11. Ormo M, Cubitt AB, Kallio K, Gross LA, Tsien RY, Remington SJ. Crystal structure of the *Aequorea victoria* green fluorescent protein. Science. 1996;273(5280):1392–5.
12. Yang F, Moss LG, Phillips Jr GN. The molecular structure of green fluorescent protein. Nat Biotechnol. 1996;14(10):1246–51.
13. Jones DD. Triplet nucleotide removal at random positions in a target gene: the tolerance of TEM-1 beta-lactamase to an amino acid deletion. Nucleic Acids Res. 2005;33(9):e80.
14. Arpino JA, Reddington SC, Halliwell LM, Rizkallah PJ, Jones DD. Random single amino acid deletion sampling unveils structural tolerance and the benefits of helical registry shift on GFP folding and structure. Structure. 2014;22(6):889–98.
15. Arpino JA, Rizkallah PJ, Jones DD. Structural and dynamic changes associated with beneficial engineered single-amino-acid deletion mutations in enhanced green fluorescent protein. Acta Crystallogr D Biol Crystallogr. 2014;70(Pt 8):2152–62.
16. Haapa S, Taira S, Heikkinen E, Savilahti H. An efficient and accurate integration of mini-Mu transposons in vitro: a general methodology for functional genetic analysis and molecular biology applications. Nucleic Acids Res. 1999;27(13):2777–84.
17. Crameri A, Whitehorn EA, Tate E, Stemmer WP. Improved green fluorescent protein by molecular evolution using DNA shuffling. Nat Biotechnol. 1996;14(3):315–9.
18. Daggett KA, Layer M, Cropp TA. A general method for scanning unnatural amino acid mutagenesis. ACS Chem Biol. 2009;4(2):109–13.
19. Pedelacq JD, Cabantous S, Tran T, Terwilliger TC, Waldo GS. Engineering and characterization of a superfolder green fluorescent protein. Nat Biotechnol. 2006;24(1):79–88.
20. Tsien RY. The green fluorescent protein. Annu Rev Biochem. 1998;67:509–44.

Liu *et al. BMC Biochemistry* (2015) 16:17

Page 11 of 11

21. Cormack BP, Valdivia RH, Falkow S. FACS-optimized mutants of the green fluorescent protein (GFP). Gene. 1996;173(1 Spec No):33–8.

22. Fukuda H, Arai M, Kuwajima K. Folding of green fluorescent protein and the cycle3 mutant. Biochemistry (Mosc). 2000;39(39):12025–32.

23. Battistutta R, Negro A, Zanotti G. Crystal structure and refolding properties of the mutant F99S/M153T/V163A of the green fluorescent protein. Proteins. 2000;41(4):429–37.

24. Andrews BT, Schoenfish AR, Roy M, Waldo G, Jennings PA. The rough energy landscape of superfolder GFP is linked to the chromophore. J Mol Biol. 2007;373(2):476–90.

25. Ward WW. Bioluminescence and chemiluminescence. New York: Academic; 1981.

26. Patterson GH, Knobel SM, Sharif WD, Kain SR, Piston DW. Use of the green fluorescent protein and its mutants in quantitative fluorescence microscopy. Biophys J. 1997;73(5):2782–90.

27. Pascarella S, Argos P. Analysis of insertions/deletions in protein structures. J Mol Biol. 1992;224(2):461–71.

28. Toth-Petroczy A, Tawfik DS. Protein insertions and deletions enabled by neutral roaming in sequence space. Mol Biol Evol. 2013;30(4):761–71.

29. Afriat-Jurnou L, Jackson CJ, Tawfik DS. Reconstructing a missing link in the evolution of a recently diverged phosphotriesterase by active-site loop remodeling. Biochemistry (Mosc). 2012;51(31):6047–55.

30. Goldsmith M, Tawfik DS. Directed enzyme evolution: beyond the low-hanging fruit. Curr Opin Struct Biol. 2012;22(4):406–12.

31. Murakami H, Hohsaka T, Sisido M. Random insertion and deletion of arbitrary number of bases for codon-based random mutation of DNAs. Nat Biotechnol. 2002;20(1):76–81.

32. Fujii R, Kitaoka M, Hayashi K. RAISE: a simple and novel method of generating random insertion and deletion mutations. Nucleic Acids Res. 2006;34(4):e30.

33. Simm AM, Baldwin AJ, Busse K, Jones DD. Investigating protein structural plasticity by surveying the consequence of an amino acid deletion from TEM-1 beta-lactamase. FEBS Lett. 2007;581(21):3904–8.

34. Baldwin AJ, Busse K, Simm AM, Jones DD. Expanded molecular diversity generation during directed evolution by trinucleotide exchange (TriNEx). Nucleic Acids Res. 2008;36(13):e77.

35. Liu J, Cropp TA. Experimental methods for scanning unnatural amino acid mutagenesis. Methods Mol Biol. 2012;794:187–97.

36. Baldwin AJ, Arpino JA, Edwards WR, Tippmann EM, Jones DD. Expanded chemical diversity sampling through whole protein evolution. Mol Biosyst. 2009;5(7):764–6.

37. Arpino JA, Baldwin AJ, McGarrity AR, Tippmann EM, Jones DD. In-frame amber stop codon replacement mutagenesis for the directed evolution of proteins containing non-canonical amino acids: identification of residues open to bio-orthogonal modification. PLoS One. 2015;10(5):e0127504.

38. Edwards WR, Busse K, Allemann RK, Jones DD. Linking the functions of unrelated proteins using a novel directed evolution domain insertion method. Nucleic Acids Res. 2008;36(13):e78.

39. Arpino JA, Czapinska H, Piasecka A, Edwards WR, Barker P, Gajda MJ, et al. Structural basis for efficient chromophore communication and energy transfer in a constructed didomain protein scaffold. J Am Chem Soc. 2012;134(33):13632–40.

40. Liu J, Cropp TA. A method for multi-codon scanning mutagenesis of proteins based on asymmetric transposons. Protein Eng Des Sel. 2012;25(2):67–72.

41. Liu J, Cropp TA. Rational protein sequence diversification by multi-codon scanning mutagenesis. Methods Mol Biol. 2013;978:217–28.

42. Dellus-Gur E, Elias M, Caselli E, Prati F, Salverda ML, de Visser JA, et al. Negative Epistasis and Evolvability in TEM-1 beta-Lactamase-The Thin Line between an Enzyme's Conformational Freedom and Disorder. J Mol Biol. 2015;427(14):2396–409.