

# Chemical Science

Accepted Manuscript



This is an *Accepted Manuscript*, which has been through the Royal Society of Chemistry peer review process and has been accepted for publication.

*Accepted Manuscripts* are published online shortly after acceptance, before technical editing, formatting and proof reading. Using this free service, authors can make their results available to the community, in citable form, before we publish the edited article. We will replace this *Accepted Manuscript* with the edited and formatted *Advance Article* as soon as it is available.

You can find more information about *Accepted Manuscripts* in the [Information for Authors](#).

Please note that technical editing may introduce minor changes to the text and/or graphics, which may alter content. The journal's standard [Terms & Conditions](#) and the [Ethical guidelines](#) still apply. In no event shall the Royal Society of Chemistry be held responsible for any errors or omissions in this *Accepted Manuscript* or any consequences arising from the use of any information it contains.



[www.rsc.org/chemicalscience](http://www.rsc.org/chemicalscience)

## Directed evolution of GFP with non-natural amino acids identifies residues for augmenting and photoswitching fluorescence

Cite this: DOI: 10.1039/x0xx00000x

Received 00th January 2012,  
Accepted 00th January 2012

DOI: 10.1039/x0xx00000x

www.rsc.org/

Samuel C. Reddington<sup>a,b,\$</sup>, Amy J. Baldwin<sup>a,b,\$</sup>, Rebecca Thompson<sup>a,†</sup>,  
Andrea Brancale<sup>c</sup>, Eric M. Tippmann<sup>b,†</sup> & D. Dafydd Jones<sup>a,\*</sup>.

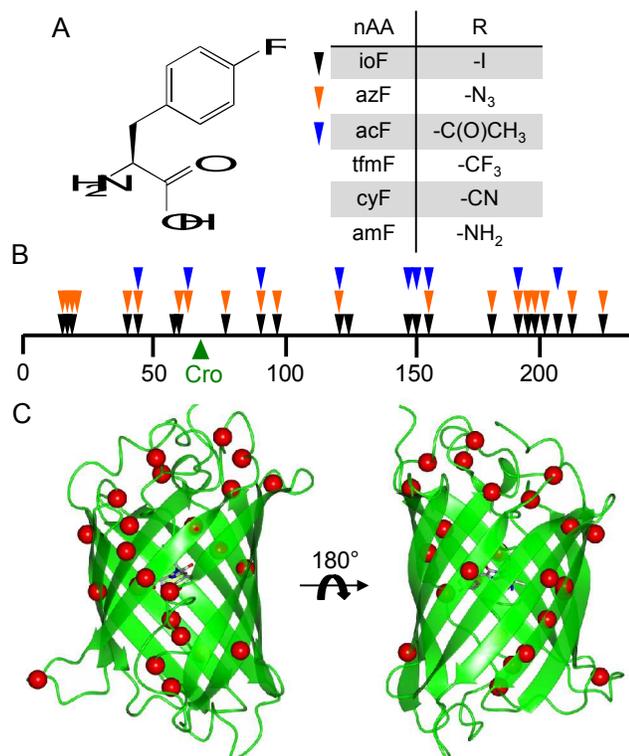
Genetic code reprogramming allows proteins to sample new chemistry through the defined and targeted introduction of non-natural amino acids (nAAs). Many useful nAAs are derivatives of the natural aromatic amino acid tyrosine, with the *para* OH group replaced with useful but often bulkier substituents. Extending residue sampling by directed evolution identified positions in Green Fluorescent Protein tolerant to aromatic nAAs, including identification of novel sites that modulate fluorescence. Replacement of the buried L44 residue by photosensitive *p*-azidophenylalanine (azF) conferred environmentally sensitive photoswitching. *In silico* modelling of the L44azF dark state provided an insight into the mechanism of action through modulation of the hydrogen bonding network surrounding the chromophore. Targeted mutagenesis of T203 with aromatic nAAs to introduce  $\pi$ -stacking with the chromophore successfully generated red shifted versions of GFP. Incorporation of azF at residue 203 conferred high photosensitivity on sfGFP with even ambient light mediating a functional switch. Thus, engineering proteins with non-natural aromatic amino acids by sampling a wide residue set can introduce new and beneficial properties into a protein through the sampling of non-intuitive mutations. Coupled with retrospective *in silico* modelling, this will facilitate both our understanding of the impact of nAAs on protein structure and function, and future design endeavours.

### Introduction

Reprogramming the genetic code to allow incorporation of non-natural amino acids (nAAs) is a powerful way of engineering proteins by expanding the chemistry sampled (see [1-4] for recent reviews). A wide variety of both aliphatic and aromatic nAAs can now be used to introduce physicochemical properties not normally sampled by the 20 natural amino acids. One of the most developed recombinant methods to incorporate nAAs into proteins *in situ* is amber stop codon (TAG) reprogramming through the use of orthogonal tRNA/aminoacyl-tRNA-synthetase pairs engineered for a specific nAA [5]. Recently, *E. coli* has been engineered to remove all amber stop codons from the genome and the associated termination factor (RF-1) to truly reprogram TAG for nAA incorporation [6]. Aromatic nAAs (see Fig. 1 for some examples) based around the natural amino acid tyrosine have proved a particularly fruitful and well-utilised source of new chemistry [2, 7, 8], driven by the use of engineered versions of the versatile tyrosyl-tRNA synthetase.

While rational protein engineering using nAAs is becoming more widespread, it still suffers from the drawbacks of traditional rational site-directed mutagenesis commonly used to implement the approach: predicting the impact of nAA incorporation on protein structure, function and folding. For protein engineering with nAAs to reach its true potential, more information is required on the tolerance of a protein to nAA incorporation and the subsequent impact on the structure-function relationship. This is beginning to be addressed, including with GFP [9-12], through more detailed structure-function investigations of site directed mutations. The concept of directed evolution [13-15] was developed to overcome such predictive problems encountered during traditional protein engineering, and genetic library generation methods have recently been developed for broader nAA sampling [16, 17]. By sampling the whole protein backbone, it is possible to uncover influential residues that would have otherwise been overlooked because, for example, they are distal from the active site or their actions are exerted indirectly on active site residues. Retrospective analysis of such unexpected beneficial mutations can then feedback to the protein design process.

Here we report the use of directed evolution to sample different residues positions and aromatic nAAs to investigate tolerance by and impact on Green Fluorescent Protein. GFP has been the subject of extensive protein engineering endeavours (including by directed evolution) to expand its usefulness as a genetically encoded imaging and reporting agent. Current GFP engineering endeavours focus on generating responsive variants, especially photoswitching for super-resolution imaging [18, 19]. The “superfolder” GFP variant (sfGFP) that is the focus of this work is a fast folding, maturing and highly stable engineered version [21] of the protein but has no inherent photoswitching behaviour. A library generated by random amber stop codon replacement mutagenesis and one site-directed variant (T203) was screened with different aromatic nAAs to select variants that retained fluorescence so assessing sfGFP’s tolerance to nAA incorporation. T203 was targeted, as replacement with aromatic residues is known to influence the fluorescent properties of GFP through  $\pi$  stacking with the chromophore [24] but was not sampled during selection. A number of variants were identified that had altered spectral properties and displayed photosensitive behaviour, with some nAA incorporation sites hitherto unexplored by traditional mutagenesis. *In silico* modelling was used to investigate the impact of a nAA at residue 44 on GFP structure and thus function, with the aim of facilitating the future design of useful proteins containing new chemistry.



**Fig. 1. Tolerance of sfGFP to aromatic nAA incorporation.** (a) Structure of the aromatic non-natural amino acids used in this study including the 3 aromatic nAAs used in library screening (ioF, azF and acF). (b) Linear map of sfGFP sequence showing sites tolerant to aromatic nAA incorporation as arrows. Each colour represents variants found from screening with three different aromatic nAAs (ioF, black; azF, orange; acF, blue). The position of the sfGFP chromophore (“Cro”) is shown as a green arrow. A more detailed description is provided in

Supporting Table 1. (c) 3D structure of sfGFP showing the positions of tolerated aromatic nAA incorporation sites as red spheres.

## Results and discussion

### RANDOM TAG CODON REPLACEMENT MUTAGENESIS

The nAAs selected as part of this study are derivatives of phenylalanine (Fig. 1a) with the *para* position substituted with various chemical groups not present in the natural amino acid repertoire. To sample aromatic nAA incorporation across the breadth of sfGFP, a recently developed directed evolution method based on trinucleotide exchange (TriNEX; [20]) (Supp. Fig. 1) was used to introduce TAG codons at random positions throughout the gene. A detailed description of the library construction process is outlined in the Supplementary Information.

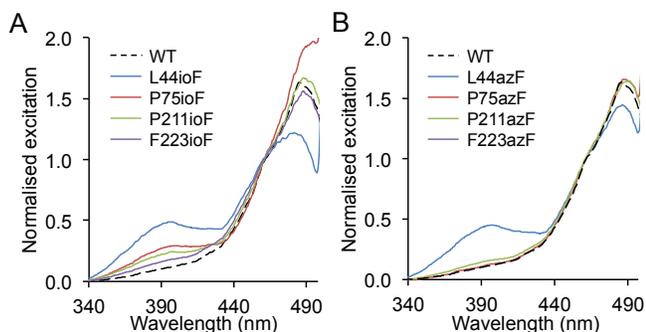
The protein appeared tolerant to a variety of aromatic nAAs incorporated at different positions, demonstrating the structural plasticity of sfGFP, even to novel chemical substituents (Fig. 1). Sequence analysis of randomly selected library members that conferred cellular fluorescence when grown in the presence of the aromatic nAAs ioF, azF and acF revealed that all of the variants had in-frame TAG codons distributed throughout the *sfGFP* gene (Fig. 1b; Supp. Table 1). Identical mutations were observed independently on selection in the presence of different aromatic nAAs, with 5 of the mutation sites observed when screened with all three aromatic nAAs (Fig. 1b).

The majority of tolerated aromatic nAA substitutions involved residues resident in  $\beta$ -strands (60%), in keeping with the predominance of this secondary structure element (comprising ~50% of residues) in sfGFP (Fig. 1c) [21]. Residues normally buried within the core of sfGFP, which generally comprise aliphatic side chains, were remarkably tolerant to aromatic nAA incorporation despite their bulkier nature and polar *para* substituents; 10 buried residues and 5 partially exposed residues were tolerant to aromatic nAA incorporation (Supp. Table 1). These include hydrophobic core residues V16, L44, L119, V150, L201 that are close to the chromophore. Three residues in the core helix, P56, W57 and L60, housing the buried chromophore (comprised of G65-Y66-T67) were also found to be tolerant to at least one of the chosen aromatic nAAs (Fig. 1c; Supp Table 1).

### FLUORESCENCE PROPERTIES OF AROMATIC NAA SUBSTITUTED GFP VARIANTS.

The chromophore at the heart of all GFP-like fluorescent proteins is very sensitive to its immediate chemical environment, a feature that has been utilised by both nature and protein engineers to generate a broad colour palette and introduce new responsive properties [18, 22]. To assess the impact of aromatic nAA incorporation on sfGFP function, the fluorescence spectra of each unique mutation site observed during the selection phase were investigated using ioF or azF. In most cases, replacement of the native amino acid with ioF or azF had little or no effect on the excitation or emission spectra (Supp. Table 2) but several mutations did elicit an effect, predominantly on the excitation spectra (Fig. 2). The most

significant change observed was the promotion or suppression of the minor excitation peak at  $\sim 400$  nm, which reports on the charge state of the phenolic group of the chromophore [23, 24]. In the ground state, the neutral, protonated phenol group excites at  $\sim 400$  nm while the anionic, deprotonated form excites at  $\sim 485$  nm.



**Fig. 2. Spectral properties of sfGFP<sup>NAA</sup> variants.** Fluorescence excitation spectra of (a) sfGFP<sup>ioF</sup> and (b) sfGFP<sup>azF</sup> variants. Spectra were recorded on cell lysates (soluble fraction) by monitoring emission at 511 nm. Spectra were normalised to a value of 1 at 460 nm.

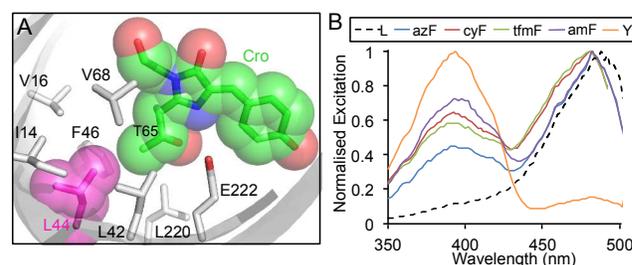
Replacing L44, P75, P211 and F223 with ioF resulted in promotion of the excitation peak at  $\sim 400$  nm (Fig. 2a; Supp. Table 2). The L44ioF variant generated the largest change in the 485:400 nm ratio (1:0.4) of all the variants. L44 is buried within the core of the protein (Supp. Table 1) and lies close to the chromophore but does not directly interact with the phenol group. In the case of P75, P211 and F223, it is not clear how ioF is eliciting an effect given their distance, surface exposure and relative positioning with relation to the chromophore but is most likely to be via small and propagated changes in the protein structure. However, the effects appear to be specific to the steric bulk of ioF. Replacement of the iodo group with an azido results in only the L44azF retaining a significant excitation peak at  $\sim 400$  nm (Fig. 2b). Therefore, there is merit to sampling non-obvious residues with different aromatic nAAs to accelerate the generation of proteins with novel properties. Replacement of several residues with ioF or azF suppressed the minor 400 nm excitation peak and thus the population of the neutral chromophore in the ground state of sfGFP (Supp. Fig. 2). However, these changes were relatively minor in comparison to the variants that increase the proportion of the neutral form.

#### THE EFFECTS OF AROMATIC NAA REPLACEMENT AT RESIDUE 44.

It was clear that replacement of L44 with either ioF or azF produced the most significant changes in sfGFP fluorescence excitation of all the variants sampled. L44 has not to our knowledge been the target of traditional or nAA mutagenesis approaches to engineer the fluorescent properties of GFP and therefore the directed evolution approach has uncovered a novel mutation site. The L44 side chain is tightly packed in a largely hydrophobic core environment in the same plane as the chromophore (Fig. 3a). It packs close to the T65 moiety of the

chromophore, at the opposite end to the phenol (contributed by Y66) group. The T65 moiety together with other nearby residues are known modulators of GFP fluorescence [25, 26]; small structural perturbations can have significant effects through disruption of charge-transfer networks that define the protonation state of the chromophore phenol group. Especially pertinent is the position of L44 parallel to E222 (Fig 3a), the later being a pivotal residue in defining this polar bond network [25, 27, 28].

Further investigation with additional aromatic nAAs demonstrated that changing the nature of the *para* substituent group modulated the excitation profile of the protein (Fig. 3). Three additional aromatic nAAs, *p*-cyano-L-phenylalanine (cyF), *p*-trifluoromethyl-L-phenylalanine (tfmF) and *p*-amino-L-phenylalanine (amF) (Fig. 1a), together with the natural analogue tyrosine were incorporated at residue 44. The different aromatic amino acids chosen have *para* substituents that vary in their polarity, electronegativity and hydrogen-bonding potential. All are factors known to influence the fluorescence properties of the chromophore [24]. Furthermore, amF represents one of the potential photochemical endpoints for azF photolysis (*vide infra*) [9]. The effect of each aromatic nAA was to alter the ratio of neutral ( $\lambda_{EX} \sim 400$  nm) to charged ( $\lambda_{EX} \sim 484$  nm) forms of the chromophore in the ground state (Fig. 3b). The proportion of the neutral peak increased with the pattern Tyr > amF > cyF  $\approx$  tfmF > azF > Leu. Y44 essentially forced the chromophore to adopt entirely the neutral form over the anionic. The amino (amF), cyano (cyF) and trifluoromethyl (tfmF) groups also promoted the neutral form but with the phenolate form dominating to different degrees depending on the aromatic nAA (neutral:anionic ratios of 0.7:1, 0.64:1 and 0.58:1, respectively). Additionally, all of the sampled aromatic nAAs resulted in a slight blue-shift in excitation ( $\sim 5$  nm), which may be indicative of changes in residue packing and contacts around the chromophore.

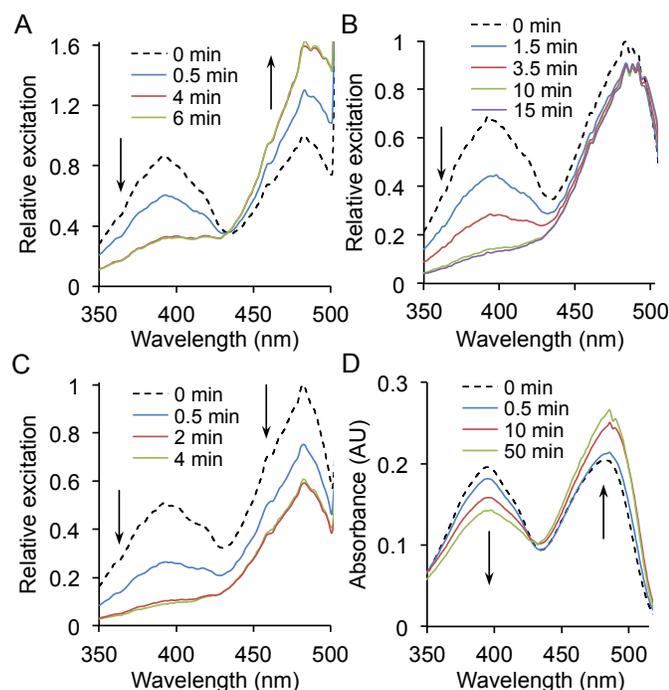


**Fig. 3. The influence of L44 mutations on sfGFP fluorescence.** (a) Position of L44 (magenta) in sfGFP in relation to the chromophore (Cro; green) with neighbouring residues shown as sticks. PDB accession 2B3P. (b) Fluorescence excitation spectra of sfGFP with the indicated amino acid at position 44. Spectra were measured by monitoring emission at 511 nm on the soluble fraction of cell lysates from protein production cultures that were diluted to equivalent OD<sub>600</sub> of 0.5.

#### PHOTOCONTROL OF sfGFP<sup>L44AZF</sup>

Photoswitching autofluorescent proteins are of current interest as probes for super resolution imaging [18]. Phenyl azide chemistry could provide a general route to genetically encode photocontrol in autofluorescent proteins (and proteins in

general) with fast and large magnitude switching characteristics [4, 9]. Phenyl azides are known to be photochemically sensitive to light in the near UV and blue region and have classically been used as photocrosslinking reagents in biology through the generation of a reactive singlet nitrene radical on irradiation at physiologically relevant temperatures [4, 29-31]. Singlet nitrene is a highly reactive electrophile that can undergo a number of reactions including insertion into single bonds, addition to double bonds and reduction to a phenyl amine, which are environment dependant. Recent work from our lab using targeted mutagenesis has shown how different phenyl azide photochemical pathways can be used to modulate GFP fluorescence both *in vitro* and *in situ*, including by reduction to the phenyl amine (azF to amF) and addition reactions [9].



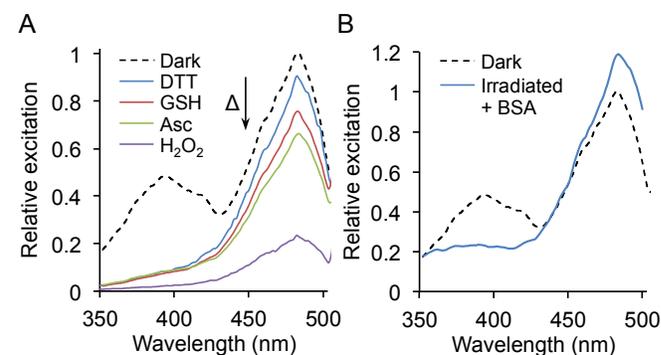
**Fig. 4. Photoswitching properties of sfGFP<sup>L44azF</sup>.** Fluorescence photoswitching of sfGFP<sup>L44azF</sup> was monitored on (a) whole cell, (b) cell lysate and (c) 1  $\mu$ M pure protein samples. Samples in (a) and (b) were standardised to an OD<sub>600</sub> of 0.5. Relative excitations were calculated by normalisation of fluorescence excitation intensity to 1 for the major 484 nm peak at time point 0 (no irradiation). Fluorescence excitation spectra were recorded following UV irradiation (302 nm, 6 W) for the indicated amounts of time by monitoring emission at 511 nm. Arrows indicate the change in fluorescence intensity over irradiation time. (d) Photoswitching behaviour of sfGFP<sup>L44azF</sup> monitored by UV-vis absorbance. Spectra were recorded on 10  $\mu$ M pure protein in phosphate buffer (100 mM, pH 8, 300 mM NaCl). The corresponding emission spectra are shown in Supp. Fig. 5

Each of the sfGFP variants described above were produced containing azF and their light-dependent fluorescence was investigated. The orthogonality and fidelity of the tRNA-synthetase system used incorporate azF is high, with the TAG codon reverting to a stop codon in the absence of azF from cell cultures (Supp. Fig. 3; [9]). The parent sfGFP is largely insensitive to UV exposure of this kind [9], with the majority of aromatic nAA variants also displaying similar insensitivity (see P75azF as an example in Supp. Fig. 4). However, some variants

displayed significant changes in their excitation and emission spectra on irradiation. Amongst these were the A37azF and L60azF variants, which exhibited a decrease in fluorescence emission intensity of between 35-40% on increasing exposure to UV (Supp. Fig. 4).

The most notable effect was observed for L44azF (sfGFP<sup>L44azF</sup>). Production of the variant in the dark resulted in whole cell fluorescence excitation peak ratio (484:394 nm) of  $\sim$ 1:1 (Fig. 4a and Supporting Figure 5a and 5b). On irradiation the 394 nm excitation peak decreased and the 484 nm peak increased ratiometrically characterised by an isobestic point at 433 nm (Fig. 4a); the 394:484 nm excitation ratio stabilised at  $\sim$ 1:5. As discussed above, the red shift in  $\lambda_{EX}$  by 90 nm is indicative of a change in the ground state chromophore population from the neutral phenol to anionic phenolate.

The fluorescence changes of sfGFP<sup>L44azF</sup> on photolysis appeared to be environmentally sensitive. In comparison to whole cell samples, a ratiometric response was not observed for lysed cells; the 394 nm excitation peak almost disappeared while the 484 nm peak remained constant (Fig. 4b and Supp. Fig. 5c and 5d). The change was different again for purified sfGFP<sup>L44azF</sup>, with irradiation resulting in a decrease in both excitation peaks and a near complete loss of the 394 nm peak (Fig. 4c and Supp. Fig. 5e and 5f). The bacterial cytosol is a complex mixture with proteins densely packed at a high effective concentration. Various properties such as ionic conditions, pH and redox potential are also regulated to maintain a constant intracellular environment. On preparation of cell lysate samples, cellular components are diluted and on protein purification are completely replaced by phosphate buffer. GFP is known to be sensitive to changes in pH and salt but both these had little effect on the photolysis characteristics apart from changing the initial fluorescence intensity (Supp. Fig. 6 and Supp. Fig. 7).



**Fig. 5. Redox and molecular crowding sensitive photoswitching of sfGFP<sup>L44azF</sup>.** (a) Photoswitching of pure sfGFP<sup>L44azF</sup> in the presence of different redox agents. The initial ('dark' reading) and final time points of irradiation for each redox agent are shown. (b) Photoswitching of sfGFP<sup>L44azF</sup> in the presence of 200 mg/mL BSA showing the initial ('dark') and final time point of irradiation (60 min). Fluorescence excitation spectra are shown in (a) and (b) and were recorded on 1  $\mu$ M sfGFP<sup>L44azF</sup> by monitoring emission at 511 nm.

Phenyl azides can be sensitive to redox reagents, with one of the photochemical pathways being reduction to the phenyl amine, which can in turn influence GFP fluorescence [9]. To

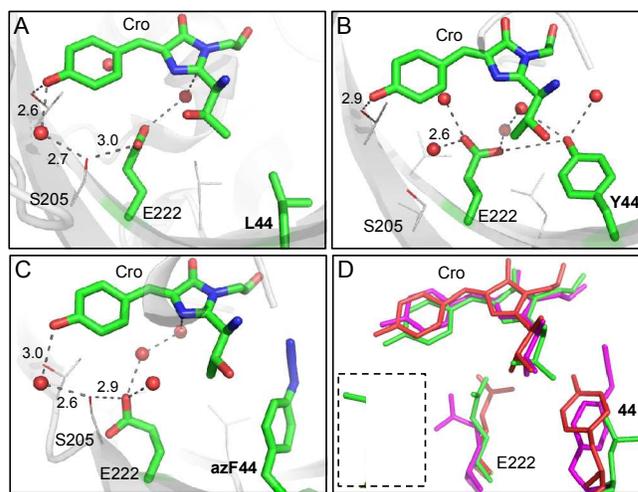
test the effect of different redox agents on the photolysis characteristics of sfGFP<sup>L44azF</sup>, irradiation was performed with pure protein in buffers containing 1 mM dithiothreitol (DTT), reduced glutathione (GSH), ascorbate (Asc) or hydrogen peroxide (H<sub>2</sub>O<sub>2</sub>). While the general photolysis characteristics were similar to that in the absence of redox agent, the extent of the 485 nm peak decrease depended on the reducing agent used (Fig. 5a); stronger reducing agents giving rise to smaller proportional decreases at 485 nm compared to 400 nm with the pattern DTT > GSH > Asc (Fig. 5a). DTT gave a very similar photolysis pattern to that observed in cell lysates. The presence of the sole oxidizing agent, H<sub>2</sub>O<sub>2</sub>, was largely destructive with only ~20% fluorescence intensity retained after irradiation (Fig. 5a). Incubation of the sfGFP<sup>L44azF</sup> with each redox agent in the dark (no irradiation) had no effect on the excitation spectra profile, confirming requirement for irradiation to elicit the redox agent mediated effect.

Interestingly, under all conditions photolysis promoted conversion to the phenolate chromophore as suggested by increasing absorbance at 484 nm, with an isobestic point at ~426 nm (Fig. 4d and Supp. Fig. 8). The different redox agents did affect the extent that the 484 nm peak increased on prolonged irradiation as represented by the differences in molar absorbance coefficients pre- and post- irradiation (Supp. Table 3); for example the presence of 1 mM DTT resulted in a 45 % increase whereas H<sub>2</sub>O<sub>2</sub> only increased by 23 %. The blue-to-red absorbance shift on irradiation for sfGFP<sup>L44azF</sup> suggests that the photochemical conversion is similar under the different redox conditions but the endpoint is redox sensitive, in terms of fluorescence emission. This is borne out in the changes in quantum yield (QY) on excitation at 484 nm. QY dropped from 0.69 to 0.51 on irradiation, tallying with the apparent drop in fluorescence emission intensity despite the increase in absorbance at 484 nm. In the presence of DTT the decrease in QY is less pronounced on irradiation (0.69 to 0.63). Combined, the data suggests that photolysis of the pure protein still promotes the anionic form of the chromophore over the neutral form but that the fluorescent capacity of the anionic form is reduced with a more significant non-radiative energy loss that is redox agent-dependent.

The initial loss of N<sub>2</sub> from sfGFP<sup>L44azF</sup> on photolysis could possibly lead to a rearrangement of the hydrogen bond network linked to the chromophore. The photochemical endpoint of the reactive nitrene thus may determine local fluorescence or, more accurately, quenching environment. We attempted to determine the photochemical endpoint by mass spectrometry but this was not feasible, as explained in the Supporting Information. One redox sensitive photochemical pathway is reduction of phenyl azide to a phenyl amine. This can be probed directly by incorporating *p*-amino-phenylalanine (amF; Fig. 1a) into sfGFP at residue 44 as in Fig. 3b. A peak at 394 nm corresponding to the neutral chromophore dominates the absorbance spectrum (Supp. Fig. 9a); the fluorescence excitation peak ratio (394:484 nm) is 0.7:1 (Fig. 3b and Supp. Fig. 9b-c). Both of these are dissimilar to the endpoint of sfGFP<sup>L44azF</sup> irradiation suggesting that the phenyl amine is unlikely to be the photochemical

endpoint. Despite our current limited understanding of the redox-dependent endpoint, these observed responses opens up the possibility of using sfGFP<sup>L44azF</sup> as a genetically encoded cellular redox sensor for monitoring oxidative stress.

One condition that we were not able to fully replicate was high protein concentration. Intact *E. coli* cells experience high protein concentration, typically in the 1-10 mM range, and the cytosol is close to gel-like conditions. In an attempt to mimic intracellular conditions more closely, a high concentration (200 mg/mL) of the commonly used inert protein BSA (bovine serum albumen) was added to 1 μM sfGFP<sup>L44azF</sup> (in PBS) to act as a crowding agent. Subsequent photolysis resulted in the blue-to-red fluorescence switching characteristic of that observed for whole cell samples (Fig. 5b). However, no single isobestic point was observed as with whole cell samples. These results confirm the complexity of sfGFP<sup>L44azF</sup> photoswitching but suggest a key role of molecular crowding.



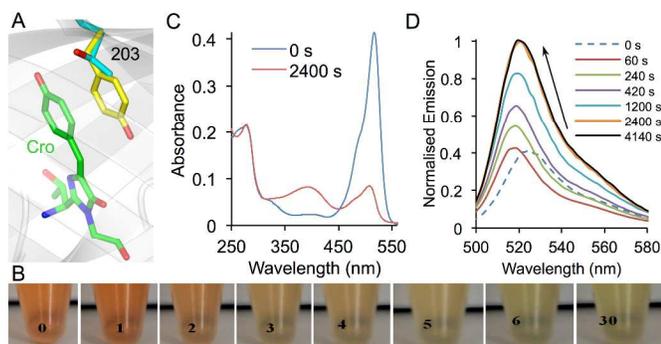
**Fig. 6. Molecular models of the sfGFP<sup>L44</sup> aromatic nAA mutants.** (a) X-ray crystal structure of the parent protein sfGFP (PDB accession 2B3P). Representative structural models of (b) sfGFP<sup>L44Y</sup> and (c) sfGFP<sup>L44azF</sup> obtained by molecular dynamics (50 ns). Residue 44, E222 and the chromophore (Cro) are shown as sticks and coloured by element with green carbon atoms in each case. Nearby residues are shown as lines and coloured by element with grey carbon atoms. Suggested hydrogen bonds are shown as black dashed lines and structural water molecules as red spheres with distance shown in Å. (d) Overlay of sfGFP (green), sfGFP<sup>L44Y</sup> (red) and sfGFP<sup>L44azF</sup> (pink) showing residues 44, E222 and the chromophore. The inset shows E222 from a different orientation (~90° rotation).

#### UNDERSTANDING THE STRUCTURAL IMPACT OF L44 MUTATIONS TO AROMATIC NAA BY MOLECULAR MODELLING.

The ability to model *in silico* the impact of mutations on proteins is a powerful way of matching phenotypic observations with underlying structural mechanisms, and form the basis for future protein design endeavours. There is currently limited *in silico* approaches [32] to modelling and thus designing proteins containing nAAs. Here, we use a molecular dynamics approach to model the effect of mutating L44 to tyrosine or azF. The sfGFP<sup>L44Y</sup> model suggests that a shift in the position of E222 side chain that directly impacts on the local hydrogen bonding network, including removing the indirect interaction through S205 to the phenol group of the chromophore (Fig. 6b). The model suggests that the side chain hydroxyl group of the T65

moiety forming the chromophore rotates away from E222 and the hydroxyl group of Y44 to form a new interaction with E222 both directly and via a structural water molecule. This new putative polar bond network is likely to stabilise the charged state of E222, which in turn prevents deprotonation of the chromophore phenol moiety by charge repulsion [25], as suggested by the experimental data (Fig. 3b).

The replacement of the *para* OH group with azido at residue 44 seemed to preclude either the wild type (L44) or Y44 arrangements and resulted in azF44 flipping away from E222 into a hydrophobic pocket (Fig. 6c). The consequence is that E222 shifts away from residue 44 whose side chain now occupies a position intermediate between sfGFP and sfGFP<sup>L44Y</sup> (Fig 6d). The shift in E222 does not appear to remove the key polar interaction with S205 but the position of S205 itself is shifted by ~1.6 Å (Fig. 6). The consequence of this shift in S205 appears to be that the H-bond with the bridging water and the chromophore now lengthens by ~0.4 Å and the chromophore loses the hydrogen bond with the OH group of Thr203. The intermediate position of E222 and its knock on effect on hydrogen bond network may correlate with excitation spectrum observed for sfGFP<sup>L44azF</sup> suggesting the presence of both the charged and neutral form of the chromophore (Fig. 3b). Without knowing the endpoint product of sfGFP<sup>L44azF</sup> it is difficult to build models concerning what might be occurring on photolysis. It is interesting to speculate that the loss of N<sub>2</sub> will potentially provide more rotational freedom thus changing the relative position of the aromatic moiety, which in turn may influence the local polar network and thus chromophore charge.



**Fig. 7. Photoswitching properties of sfGFP<sup>T203azF</sup>.** (a) Structure of sfGFP showing the position of residue 203 in relation to the chromophore (Cro; green). Residue 203 is shown as the native Thr (grey) and Tyr as (yellow) as found in YFP (PDB accession 1YFP) demonstrating the  $\pi$ - $\pi$  stacking interaction with the chromophore. (b) Ambient light photoswitching of sfGFP<sup>T203azF</sup>. Images of a cell lysate sample left in ambient room light for the indicated amount of time (in min). Photoswitching of sfGFP<sup>T203azF</sup> monitored by (c) absorbance and (d) fluorescence emission. Photolysis was performed with a handheld UV lamp (302 nm, 6 W) for the indicated amount of time. Emission spectra were recorded on 1  $\mu$ M protein after excitation at 485 nm and absorbance spectra with 10  $\mu$ M protein.

#### AROMATIC NAA INCORPORATION AT T203 AND PHOTOSWITCHING

Residue T203 lies directly above the plane of the chromophore (Fig. 7a) and has proved important in engineering the

fluorescent properties of GFP. Replacement of T203 with natural aromatic residues red shifts  $\lambda_{EM}$  and  $\lambda_{EX}$  through  $\pi$ - $\pi$  stacking interactions with the chromophore and is the major mutation generating Yellow Fluorescent Protein (YFP; T203Y) [33, 34]. This residue was not observed in the original library screening process but this may be due to its colour shift in fluorescence and its acute sensitivity to light (*vide infra*). A more likely explanation is that it may be a rarely occurring codon exchange mutation in the library [20] so was not sampled during selection. Thus, residue 203 was the subject of targeted mutagenesis to assess the influence of aromatic nAAs. As expected, all of the aromatic nAAs sampled at residue 203 (azF, cyF, tmF and amF) as well as tyrosine resulted in red-shifted fluorescence (Supp. Fig. 10) due to the predicted  $\pi$  stacking with the chromophore (Fig. 7a).

Given the proximity and intimate interaction between residue 203 and the chromophore, the introduction of azF at residue 203 should confer photosensitivity on sfGFP. sfGFP<sup>T203azF</sup> was indeed very sensitive to irradiation with even ambient light eliciting a major change in the absorption properties and thus the transmitted light (Fig. 7b). In contrast to YFP, which confers a yellowish colour on *E. coli*, cells expressing sfGFP<sup>T203azF</sup> were red in colour. The red transmittance properties suggest that the sfGFP<sup>T203azF</sup> is absorbing in the blue, green and yellow region. On removal from the dark, the cell lysates changed from red to light green over the course of 30 min. More detailed studies on the purified protein carefully prepared to minimise light exposure revealed that sfGFP<sup>T203azF</sup> was produced as a relatively poor fluorescent protein (compared to sfGFP) most likely due to the presence of the electron rich azido group, which can act as a quencher [4, 35]. Before irradiation, the sfGFP<sup>T203azF</sup> had  $\lambda_{EX}$  and  $\lambda_{EM}$  of 511 nm and 525 nm, respectively (*vide supra*). A minor excitation peak at ~485 nm was also observed as a shoulder to the major 511 nm peak (Supp. Fig. 11). The absorbance spectrum showed an equivalent peak at 517 nm with an extinction coefficient of ~41,400 M<sup>-1</sup>cm<sup>-1</sup> (Fig. 7c). On irradiation, the emission peak initially blue shifted by 8 nm and on further irradiation increased in intensity ~3 fold and red shifted by 4 nm (to 521 nm; Fig. 7c). The rate of conversion to the endpoint was slower for the chromophore  $\pi$ -stacked phenyl azide at residue 203 than observed for sfGFP<sup>L44azF</sup> (compare Fig. 4C and 7D). This is probably a function of the local microenvironment within the core of the protein. Interestingly, photoconversion of azF embedded within the highly delocalised chromophore (Y66azF mutation reported previously) with an extended conjugated system beyond the phenyl azide was equally as slow compared to positions that abut the chromophore [9]. This suggests that photoreactivity of the phenyl azide could be influenced by the local electronic environment with electron donating and extended/interacting  $\pi$  systems slowing the rate.

Absorbance changes matched the shift in wavelength (517 to 509 nm), however the peak intensity dropped to ~20 % of the original with the emergence of a secondary peak at 393 nm suggesting the formation of a second non-radiative species (Fig. 7c). The increased fluorescence despite the drop in absorbance

is evidence that a major change chemical environment of the chromophore occurs on photolysis. This could include a crosslink to the chromophore, which has been observed previously and is known to influence the electronic excitation properties (including loss of fluorescence) of the chromophore [9]. The general red shift of the major excitation peak suggests that a significant population of photolysed sfGFP<sup>T203azF</sup> retains the aromatic stacking configuration. The split population is unlikely to be due to the presence of neutral and phenolate chromophore states, as the blue absorbance peak is not fluorescent. The spectral properties of sfGFP<sup>T203amF</sup> suggest that the final endpoint of sfGFP<sup>T203azF</sup> photolysis is not the phenyl amine (Supp. Fig. 9d and e). Thus, it is clear that incorporation of azF at residue 203 instils highly sensitive photoswitching properties on sfGFP as well as red shifting its fluorescence. The photochemistry again appears to quite complex and would benefit for further structural and biophysical investigations. Further protein engineering or incorporation of additional photosensitive aromatic nAA may attenuate the sensitivity thus generating a potentially useful photoswitching autofluorescent protein.

## Conclusions

Here we have shown that aromatic nAAs with useful chemistry are tolerated within the  $\beta$ -barrel scaffold common to autofluorescent proteins, epitomised by sfGFP. This included within the core and at positions close to functional centres, despite the bulky nature and different chemical properties of the aromatic nAAs. The directed evolution approach unearthed residue positions not normally considered when engineering the fluorescence properties of GFP-based autofluorescent proteins. Phenyl azide chemistry was particularly useful as it instilled photoswitching properties not normally present in the parent protein when incorporated at particular residues. A key challenge lies in understanding how the protein environment defines the chemical route taken from the nitrene radical on photolysis [36]. The use of *in silico* modelling with nAAs allows us to understand the impact of mutations outside the natural 20 amino acid set, which will lead the way towards *in silico* design of nAA containing proteins with useful and predictable properties.

## Experimental

Detailed experimental methods are provided in the Supporting Information, including TAG library construction, protein production, spectroscopy and molecular dynamic simulations.

## Acknowledgements

The authors would like to thank Johanna Jones, Katharina Streichert and Matthew Edmundson for their technical input with regards to variant construction and protein production. The authors would like to thank the BBSRC (BB/H003746/1 and BB/E007384), EPSRC (EP/J015318/1) and Cardiff SynBio Initiative/SynBioCite for supporting this work. SCR was supported by a MRC studentship supplemented by the School of Chemistry.

## Notes and references

<sup>a</sup> School of Biosciences, Cardiff University, Cardiff, UK.

<sup>b</sup> School of Chemistry, Cardiff University, Cardiff, UK.

<sup>c</sup> School of Pharmacy and Pharmaceutical Sciences, Cardiff University, Cardiff, UK.

† SCR and AJB contributed equally to this work as indicated by \$. SCR current address Dept of Biochemistry, Oxford. RT current address Astbury Centre for Structural Biology, University of Leeds, UK. EMT current address Department of Chemistry, Indiana–Purdue University Fort Wayne, Fort Wayne, IN 46815, USA.

Electronic Supplementary Information (ESI) available: [Detailed experimental methods, Supplementary Figures 1 to 11 and Supplementary Tables 1 to 3]. See DOI: 10.1039/b000000x/

- Kim, C. H., Axup, J. Y. & Schultz, P. G. (2013) Protein conjugation with genetically encoded unnatural amino acids, *Current opinion in chemical biology*, **17**, 412-419.
- Zhang, W. H., Otting, G. & Jackson, C. J. (2013) Protein engineering with unnatural amino acids, *Current opinion in structural biology*, **23**, 581-7.
- Antonczak, A. K., Morris, J. & Tippmann, E. M. (2011) Advances in the mechanism and understanding of site-selective noncanonical amino acid incorporation, *Current opinion in structural biology*, **21**, 481-487.
- Reddington, S., Watson, P., Rizkallah, P., Tippmann, E. & Jones, D. D. (2013) Genetically encoding phenyl azide chemistry: new uses and ideas for classical biochemistry, *Biochem Soc T*, **41**, 1177-1182.
- Wang, L., Magliery, T., Liu, D. & Schultz, P. (2000) A New Functional Suppressor tRNA/Aminoacyl- tRNA Synthetase Pair for the *in Vivo* Incorporation of Unnatural Amino Acids into Proteins, *J Am Chem Soc*, **122**, 5010-5011.
- Lajoie, M. J., Rovner, A. J., Goodman, D. B., Aerni, H. R., Haimovich, A. D., Kuznetsov, G., Mercer, J. A., Wang, H. H., Carr, P. A., Mosberg, J. A., Rohland, N., Schultz, P. G., Jacobson, J. M., Rinehart, J., Church, G. M. & Isaacs, F. J. (2013) Genomically Recoded Organisms Expand Biological Functions, *Science*, **342**, 357-360.
- Liu, C. C. & Schultz, P. G. (2010) Adding new chemistries to the genetic code, *Annu Rev Biochem*, **79**, 413-44.
- Lang, K. & Chin, J. W. (2014) Cellular Incorporation of Unnatural Amino Acids and Bioorthogonal Labeling of Proteins, *Chem Rev*.
- Reddington, S. C., Rizkallah, P. J., Watson, P. D., Pearson, R., Tippmann, E. M. & Jones, D. D. (2013) Different Photochemical Events of a Genetically Encoded Phenyl Azide Define and Modulate GFP Fluorescence, *Angewandte Chemie International Edition*, **52**.
- Reddington, S. C., Tippmann, E. M. & Jones, D. D. (2012) Residue choice defines efficiency and influence of bioorthogonal protein modification via genetically encoded strain promoted Click chemistry, *Chemical communications*, **48**, 8419-8421.
- Bae, J. H., Rubini, M., Jung, G., Wiegand, G., Seifert, M. H. J., Azim, M. K., Kim, J. S., Zumbusch, A., Holak, T. A., Moroder, L., Huber, R. & Budisa, N. (2003) Expansion of the genetic code enables design of a novel "gold" class of green fluorescent proteins, *J Mol Biol*, **328**, 1071-1081.

12. Wang, F., Niu, W., Guo, J. T. & Schultz, P. G. (2012) Unnatural Amino Acid Mutagenesis of Fluorescent Proteins, *Angew Chem Int Edit.* **51**, 10132-10135.
13. Arnold, F. H. (2001) Combinatorial and computational challenges for biocatalyst design, *Nature.* **409**, 253-257.
14. Dalby, P. A. (2003) Optimising enzyme function by directed evolution, *Current opinion in structural biology.* **13**, 500-505.
15. Yuen, C. M. & Liu, D. R. (2007) Dissecting protein structure and function using directed evolution, *Nature Methods.* **4**, 995-997.
16. Baldwin, A., Arpino, J., Edwards, W., Tippmann, E. & Jones, D. (2009) Expanded chemical diversity sampling through whole protein evolution, *Molecular BioSystems.* **5**, 764-766.
17. Daggett, K. A., Layer, M. & Cropp, T. A. (2009) A General Method for Scanning Unnatural Amino Acid Mutagenesis, *ACS Chemical Biology.* **4**, 109-113.
18. Lippincott-Schwartz, J. & Patterson, G. H. (2009) Photoactivatable fluorescent proteins for diffraction-limited and super-resolution imaging, *Trends Cell Biol.* **19**, 555-565.
19. Patterson, G. H. & Lippincott-Schwartz, J. (2002) A photoactivatable GFP for selective photolabeling of proteins and cells, *Science.* **297**, 1873-1877.
20. Baldwin, A., Busse, K., Simm, A. & Jones, D. (2008) Expanded molecular diversity generation during directed evolution by trinucleotide exchange (TriNEx), *Nucleic acids research.* **36**, 77-86.
21. Pedelacq, J. D., Cabantous, S., Tran, T., Terwilliger, T. C. & Waldo, G. S. (2006) Engineering and characterization of a superfolder green fluorescent protein, *Nature biotechnology.* **24**, 79-88.
22. Shaner, N. C., Campbell, R. E., Steinbach, P. A., Giepmans, B. N., Palmer, A. E. & Tsien, R. Y. (2004) Improved monomeric red, orange and yellow fluorescent proteins derived from *Discosoma* sp. red fluorescent protein, *Nat Biotechnol.* **22**, 1567-72.
23. Chattoraj, M., King, B. A., Bublitz, G. U. & Boxer, S. G. (1996) Ultra-fast excited state dynamics in green fluorescent protein: Multiple states and proton transfer, *Proceedings of the National Academy of Sciences of the United States of America.* **93**, 8362-8367.
24. Tsien, R. (1998) The green fluorescent protein, *Annual review of biochemistry.* **67**, 509-544.
25. Brejc, K., Sixma, T. K., Kitts, P. A., Kain, S. R., Tsien, R. Y., Ormo, M. & Remington, S. J. (1997) Structural basis for dual excitation and photoisomerization of the *Aequorea victoria* green fluorescent protein, *Proceedings of the National Academy of Sciences of the United States of America.* **94**, 2306-2311.
26. Ormo, M., Cubitt, A. B., Kallio, K., Gross, L. A., Tsien, R. Y. & Remington, S. J. (1996) Crystal structure of the *Aequorea victoria* green fluorescent protein, *Science.* **273**, 1392-1395.
27. Arpino, J. A. J., Rizkallah, P. J. & Jones, D. D. (2012) Crystal Structure of Enhanced Green Fluorescent Protein to 1.35 angstrom Resolution Reveals Alternative Conformations for Glu222, *PLoS one.* **7**.
28. van Thor, J. J. (2009) Photoreactions and dynamics of the green fluorescent protein, *Chemical Society reviews.* **38**, 2935-50.
29. Chen, Y., Ebright, Y. & Ebright, R. (1994) Identification of the target of a transcription activator protein by protein-protein photocrosslinking, *Science.* **265**, 90-92.
30. Fleet, G. W. J., Porter, R. R. & Knowles, J. R. (1969) Affinity Labelling of Antibodies with Aryl Nitrene as Reactive Group, *Nature.* **224**, 511-512.
31. Schuster, G. & Platz, M. (1992) Photochemistry of phenyl azide, *Advances in photochemistry.* **17**, 69-143.
32. Mills, J. H., Khare, S. D., Bolduc, J. M., Forouhar, F., Mulligan, V. K., Lew, S., Seetharaman, J., Tong, L., Stoddard, B. L. & Baker, D. (2013) Computational Design of an Unnatural Amino Acid Dependent Metalloprotein with Atomic Level Accuracy, *Journal of the American Chemical Society.* **135**, 13393-13399.
33. Cubitt, A. B., Woollenweber, L. A. & Heim, R. (1999) Understanding structure-function relationships in the *Aequorea victoria* green fluorescent protein, *Methods in Cell Biology.* **58**, 19-30.
34. Wachter, R., Elsliger, M., Kallio, K., Hanson, G. & Remington, S. (1998) Structural basis of spectral shifts in the yellow-emission variants of green fluorescent protein, *Structure.* **6**, 1267-1277.
35. Sivakumar, K., Xie, F., Cash, B. M., Long, S., Barnhill, H. N. & Wang, Q. (2004) A fluorogenic 1,3-dipolar cycloaddition reaction of 3-azidocoumarins and acetylenes, *Organic Letters.* **6**, 4603-4606.
36. Morris, J. L., Reddington, S. C., Murphy, D. M., Jones, D. D., Platts, J. A. & Tippmann, E. M. (2013) Aryl Azide Photochemistry in Defined Protein Environments, *Organic Letters.* **15**, 728-731.