Research Article

Journal of PeptideScience

Received: 26 April 2010 Revised: 19 May 2010 Accepted: 20 May 2010 Published online in Wiley Online Library: 13 July 2010

(wileyonlinelibrary.com) DOI 10.1002/psc.1263

Azatryptophans as tools to study polarity requirements for folding of green fluorescent protein‡

Michael Georg Hoesl, Maud Larregola, Haissi Cui and Nediljko Budisa*

Aequorea victoria **green fluorescent protein and its widely used mutants enhanced green fluorescent protein and enhanced cyan fluorescent protein (ECFP) are ideal target proteins to study protein folding. The spectral signals of their chromophores are directly correlated with the folding status of the surrounding protein matrix. Previous studies revealed that tryptophan at position 57 (Trp57) plays a crucial role for the green fluorescent protein's structural and functional integrity. To precisely dissect its role in ECFP folding, we performed its substitution with the isosteric analogs 4-azatryptophan [(4-Aza)Trp] and 7-azatryptophan [(7-Aza)Trp]. Although Trp is moderately hydrophobic, these isosteric analogs are hydrophilic, which makes them an almost ideal tool to study the role of Trp57 in ECFP folding. We achieved high-level expression of both (4-Aza)Trp-ECFP and (7-Aza)Trp-ECFP. However, great portions (70–90%) of protein samples were insoluble and did not contain a maturated chromophore. All attempts to refold the insoluble protein fractions failed. Nevertheless, low amounts of fully labeled, soluble, chromophore containing fractions with altered spectral features were also isolated and identified. The most probable reason for the high yield of misfolding is the introduction of strong hydrophilicity at position 57 which strongly interferes with productive and efficient folding of ECFP. In addition, the results support a strong correlation between translational kinetics of non-canonical amino acids in the ribosome and *in vivo* folding of the related modified protein sequence. Copyright © 2010 European Peptide Society and John Wiley & Sons, Ltd.**

Supporting information may be found in the online version of this article

Keywords: azatryptophan; GFP; ECFP; Trp57; folding

Introduction

Autofluorescent proteins from various marine organisms are unique within the group of fluorescent proteins because their chromophores do not originate from a separately added prosthetic group [1,2]. In *Aequorea victoria* green fluorescent protein (*av*GFP), the chromophore is generated via a self-catalyzed intramolecular posttranslational modification of three amino acids (Ser65-Tyr66-Gly67) within the polypeptide chain [3]. This tripeptide sequence can also be found in a number of other proteins; but the cyclization takes place only under the special structural features of the *av*GFP's β-barrel fold. Thus, the tendency to form such a chromophore is not an intrinsic property of this tripeptide but rather of the folding of the surrounding protein matrix. A great number of different *av*GFP variants and mutants were produced over the past years. Among them, also an enhanced cyan fluorescent mutant (ECFP) was generated which contains a Trp residue at position 66 in the chromophore (Figure 1a).

To understand the process of chromophore maturation and to identify crucial determinants of the *av*GFP folding process, systematic and extensive variations of the chromophore itself as well as of the *av*GFP protein matrix have been performed (e.g. by site directed mutagenesis or random mutagenesis coupled with screening). Particular attention has been paid to the role of the conserved amino acids Y66, G67, R96 and E222 [4], whereas other important residues like Trp57 were not analyzed in detail. As Trp57 in *av*GFP was reported to be sensitive when mutated to other canonical amino acid [5], we envisaged that the method of genetic code engineering with non-canonical amino acids could provide an efficient tool to gain more insight into the general role of this residue in *av*GFP folding. For this purpose, the supplementation-based incorporation (SPI) method was applied by which canonical amino acid residues can selectively be replaced by isosteric analogs [6].

By this procedure, in preceding studies different mono-fluorinated Trp analogs were incorporated into ECFP [7] and by exchanging Trp66 with (4-Amino)Trp a 'gold' class of fluorescent proteins (the most red-shifted *av*GFP variant known to date) was obtained [8]. Conversely, (5-Amino)- and (6-Amino)Trp or hydroxylated Trp analogs could not be incorporated into ECFP or EGFP [7]. Obviously, the reason for this failure has to be strongly correlated with the structure of both EGFP and ECFP and especially with the role of Trp57 in the folding process.

The intent of this work was to address these issues with azatryptophans, the most isosteric Trp analogs as they differ only by a CH- to N-substitution that results in a 1 Da mass difference [9,10]. The Trp57 to (Aza)Trp57 substitution should not only

* *Correspondence to: Nediljko Budisa, Max Planck Institute of Biochemistry, Independent Molecular Biotechnology Group, Martinsried, Germany. E-mail: budisa@biochem.mpg.de*

Max Planck Institute of Biochemistry, Independent Molecular Biotechnology Group, Martinsried, Germany

‡ *Special issue devoted to contributions presented at the Symposium 'Probing Protein Function through Chemistry', 20–23 September 2009, Ringberg Castle, Germany.*

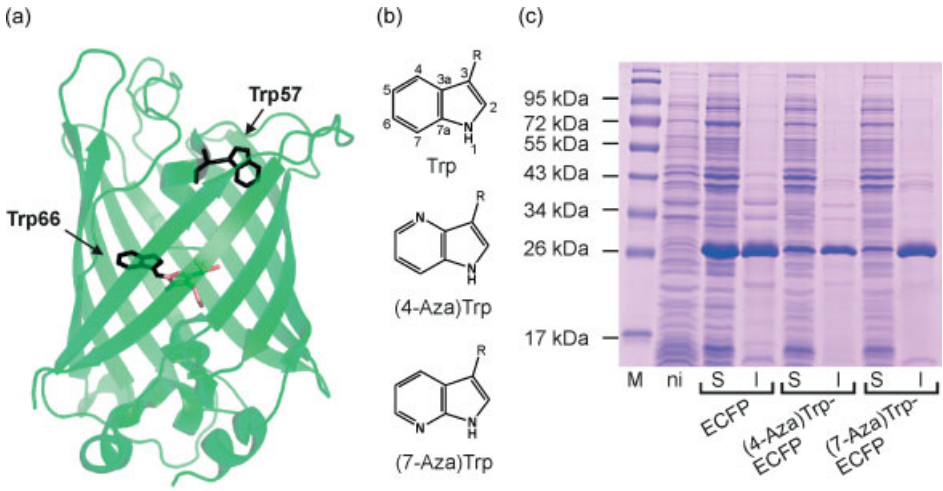

Figure 1. Protein structure of ECFP, the chemical structures of the azatryptophans and the expression gel. (a) Structure of ECFP with the two Trp residues highlighted in black; Trp66 is an integral part of the chromophore. (b) Chemical structures of Trp, (4-Aza)Trp and (7-Aza)Trp; R stands for CH_2CHNH_2COOH. (c) Separation of proteins: the soluble and insoluble fractions of the cell lysates (12% SDS–PAGE) clearly show that a high portion of ECFP is soluble. On the contrary, both (Aza)Trp-ECFP variants are mainly insoluble. M, marker (Fermentas PageRuler Prestained Protein Ladder); ni, non-induced sample; S, soluble protein fraction; I, insoluble protein fraction.

provide subtle steric local changes but also deliver dramatically different physicochemical properties into this microenvironment of the *av*GFP structure. In contrast to Trp, the azatryptophans are rather hydrophilic amino acids. Thus, their presence at position 57 should provide a tool to precisely determine which local physicochemical determinants are essential for the *av*GFP structural and functional integrity.

Results

Protein Expression and Solubility

The SPI method was applied for exchanging the tryptophan residues of ECFP by (4-Aza)Trp and (7-Aza)Trp (Figure 1a and b). Fermentation and expression procedures generally resulted in high overall protein yields. In details, ECFP showed the highest expression, whereas lowest yield of soluble protein was obtained from the (7-Aza)Trp-ECFP expression (Figure 1c). After cell harvest, pelleted cells expressing ECFP exhibited a strongly green color, whereas cells expressing ECFP variants were only faintly green.

An inspection of soluble and insoluble protein fractions by SDS–PAGE revealed the (Aza)Trp-containing variants as mainly insoluble (>70%). However, even for ECFP an insoluble fraction of about 40% was obtained which is known to mainly result from the presence of surface hydrophobic residues involved in dimerization [11,12]. Among the two azatryptophan containing ECFPs, the (7-Aza)Trp-ECFP was more insoluble (~90%) than the (4-Aza)Trp-ECFP (~70%). Both the soluble and insoluble fractions of ECFP and its variants were purified by Ni-NTA affinity chromatography (see section on Materials and Methods for details).

Discovery of Fully Labeled, Insoluble and Chromophore-Free ECFP Variants

Different refolding protocols were applied to obtain native, fluorescent protein from the insoluble fraction of (7-Aza)Trp-ECFP. However, neither the use of different buffer compositions nor refolding by dilution or on-column produced any functionally refolded protein (see section on Materials and Methods for details).

The insoluble fraction was solubilized by dialysis in 8 M urea against Tris–HCl buffer (100 mM, pH 8) supplemented with 400 mM arginine and 1 mM dithiothreitol (DTT) at room temperature over night. NMR analyses of this sample revealed that such solubilized (7-Aza)Trp-ECFP variant does not contain any secondary structure elements (Figure S8A and S8B, Supporting Information).

ESI-LC-MS of the insoluble protein fractions gave exclusive mass peaks at 28 305.53 ± 0.3 Da for (4-Aza)Trp-ECFP and 28 306.43 ± 0.3 Da for (7-Aza)Trp-ECFP (Figure S1B and S1C). Compared to the theoretical mass of matured ECFP (28 283.65 Da), this leads to mass differences of +21.88 and +22.78 Da, respectively. An additional mass difference of −20 Da occurs in the posttranslational chromophore formation which is characterized by an oxidation (−2 Da) and dehydration (−18 Da) process. Above experimental masses clearly indicate the absence of this process in the analyzed insoluble (Aza)Trp-ECFP fractions. Besides impaired chromophore maturation, the masses confirm the successful incorporation of two (Aza)Trp residues leading to an additional mass difference of 2 Da.

To obtain precise sequence mapping of the modifications in (7-Aza)Trp-ECFP, trypsination was performed in-gel with subsequent NanoLC-MS/MS analyses of the tryptic fragments. Exclusively, the peptide with both (7-Aza)Trps incorporated (LPVP(Aza)WPTLVTTLT(Aza)WGVQCFSR) was identified in the sample (see Section 2.1 in Supporting Information). This confirms that the insoluble fraction of (7-Aza)Trp-ECFP contained fully labeled, but non-folded chromophore-free protein. Further evidence for (Aza)Trp incorporation and absence of chromophore was also provided by absorbance and fluorescence measurements under native and denaturing conditions (see Sections 6.1 and 6.2 in Supporting Information).

Discovery of Fully Labeled, Soluble Fluorescent ECFP Variants

In contrast to the insoluble fractions, the soluble portions of the ECFP variants were heterogeneous materials according to the LC chromatograms of the ESI-LC-MS measurements. Correspondingly, the different species of the heterogeneous soluble fractions were separated by analytical HPLC (Figure 2a).

Journal of PeptideScience

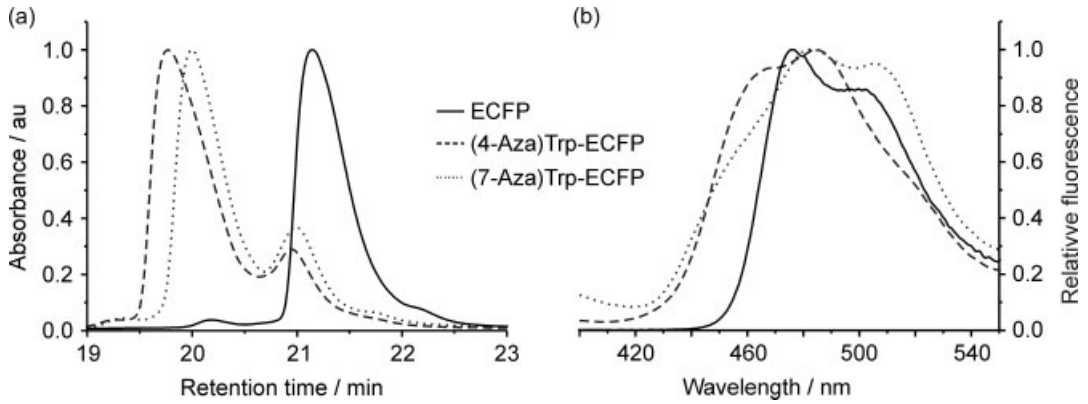

Figure 2. HPLC analysis and spectroscopy of labeled soluble ECFP fractions. Analytical HPLC runs were performed with ECFP and its (Aza)Trp variants (a). Fractions of the different peaks were collected and ESI-MS was performed. Peaks with fully labeled (Aza)Trp-ECFP [at t_R = 19.77 min for (4-Aza)Trp-ECFP and t_R = 20.00 min for (7-Aza)Trp-ECFP)] were analyzed by fluorescence spectroscopy (b). Spectra were recorded in 10 mM Tris–HCl (pH 7.5) with traces of residual TFA and acetonitrile at an excitation wavelength of 278 nm with an excitation/emission slit of 15/15.

In the HPLC profiles, two peaks are detectable: a major [at t_R of 19.77 min for (4-Aza)Trp-ECFP and t_R = 20.00 min for (7-Aza)Trp-ECFP)] and a minor peak (at t_R = 21.00 min in both samples). ESI-MS confirmed that the major peak contains fully labeled protein variants, whereas protein eluates from the minor peak were only partially substituted (see Figures S1D–S1E for detailed mass analyses). By peak integration of the HPLC chromatograms, the fully labeled protein corresponds to 86% of the sample for (4-Aza)Trp-ECFP and 82% for (7-Aza)Trp-ECFP.

In addition, the HPLC chromatograms of the parent ECFP and its variants revealed remarkable differences in hydrophobicity between those protein species. These differences doubtlessly result from the presence of hydrophilic azatryptophans in the ECFP structure, indicating that such isosteric and subtle substitutions of only two atoms in the whole protein molecule produces dramatic differences in overall protein hydrophobicity.

Additionally, analyses of the tryptic digests of the heterogeneous, non-HPLC-separated samples with NanoLC-MS/MS revealed the presence of peptides containing (Aza)Trp at both possible positions (57 and 66). Although at position 66 (chromophore) either Trp or (4-Aza)Trp/(7-Aza)Trp were detected, at 57 position only peptides containing (4-Aza)Trp/(7-Aza)Trp were identified. However, from this analysis it cannot be concluded that no Trp incorporation took place at Trp57 because in an analysis with parent ECFP (which has a Trp at position 57) the corresponding peptide could also not be identified (see Section 2.2 in Supporting Information). Possibly, the high hydrophobicity of the peptide prevents either efficient chromatographic separation or most probably peptides which do not contain an hydrophilic (Aza)Trp at position 57 do not ionize efficiently in the NanoLC-MS/MS setup.

As the main interest was on fully labeled homogeneous samples, the corresponding HPLC peaks of both the (4-Aza)Trp-ECFP and (7-Aza)Trp-ECFP were collected and their fluorescence properties were analyzed (Figure 2b). The spectra exhibit clear differences when compared to ECFP with its characteristic fluorescence maxima at 476 and 500 nm. The (4-Aza)Trp-containing ECFP shows red-shifted emission maxima at 483 and 505 nm, whereas (7-Aza)Trp exhibits a blue-shifted fluorescence (maxima at 470 and 485 nm). It should also be noted, that the spectra were broader than those of the parent protein, a feature that was also observed in ECFP with (4-Amino)Trp) [8].

Discussion

There are four totally conserved residues in all known natural fluorescent protein families (Y66, G67, R96 and E222; numbering according to *av*GFP). However, fluorescence can also be observed with Trp, His or even non-canonical Trp and Tyr analogs at position 66 in engineered mutants [7,8,13–15]. Thus, for the correct folding of the *av*GFP into its tertiary structure, formation of the chromophore is much more important than its composition [16]. This is not surprising because it is well known that chromophore cyclization is not strictly dependent on its tripeptide sequence. For example, *av*GFP mutants with GGG instead of the natural SYG tripeptide were also found to undergo cyclization [17]. To totally exclude the influence of chromophore modification in ECFP as possible reason for the observed large insoluble fraction, control experiments were performed with EGFP (which contains Tyr at position 66). Identical results were obtained with this protein (see Supporting Information).

Non-fluorescent but soluble *av*GFP mutants are well known from literature. For example, the double mutation Y74M/M88Y was found to prevent chromophore formation due to the fact that these residues affect the *cis*/*trans* isomerization of neighboring Pro residues [18]. Interestingly, Trp57 is an integral part of the proline-rich motif PVPWP that shields the *av*GFP chromophore cavity from solvent water [19]. This motif is crucial for *av*GFP structural and functional integrity as mutagenesis of Val55 (to Ala) and Trp57 (to Leu) abolished protein folding and thus fluorescence [5]. However, it is difficult to judge whether the proline *cis/trans* isomerization is influenced by the presence of the azatryptophans, although this option cannot be excluded.

The Trp57 residue is not uniformly conserved among autofluorescent proteins. Nevertheless, the results of this study support a key role of Trp57 in the co-translational folding of the *av*GFP scaffold. Thereby, the type of indole ring substitutions at the position 57 seems to decisively affect ECFP folding and stability. An inspection of the high-resolution crystal structure of ECFP (pdb: 1OXD) indicates that Trp57 is accommodated in the minicore cavity of the protein structure (Figure 3a). The positions C4–C7 of the indole moiety of Trp57 are buried in the hydrophobic core, whereas only N1 and C2 are solvent exposed. The presence of (7-Aza)Trp in this cavity leads to higher amounts of insoluble protein than (4-Aza)Trp.

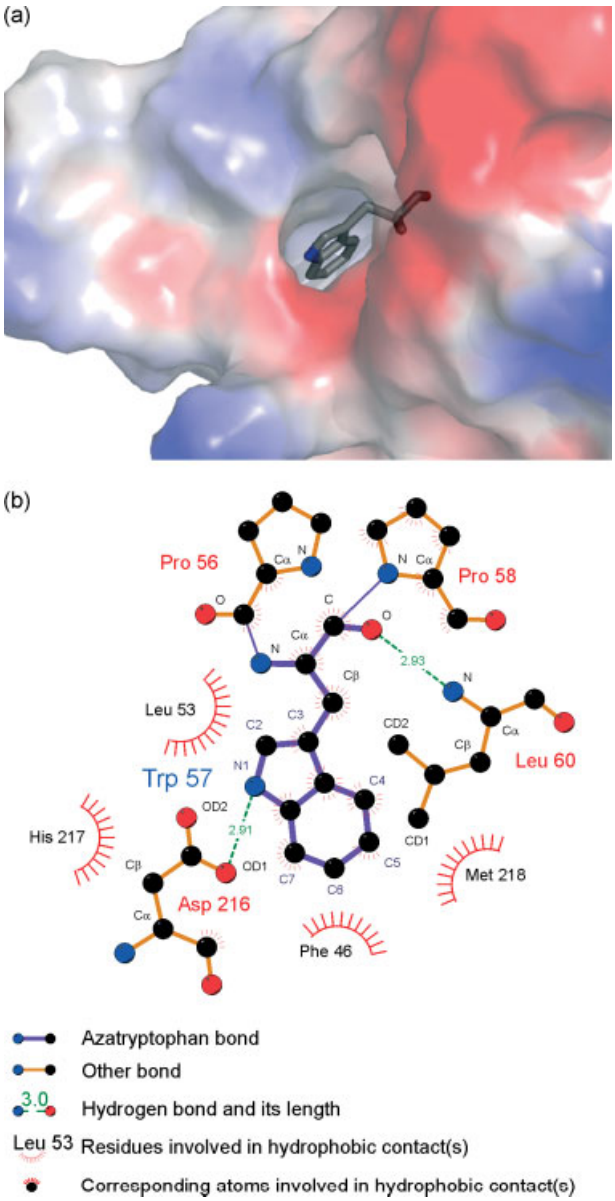

Figure 3. Local structural environment of Trp57 in ECFP. (a) Surface representation of Trp57 positioning in the context of the whole ECFP molecule. (b) Interaction network-map of Trp57 with the residues in its vicinity. Interestingly, these analyses indicate that there should be more space for the substituent accommodation at position 4 than at position 7 of the indole side chain.

According to the ECFP structure (Figure 3a), position 4 of the indole ring should be tolerant to various substitutions as there is sufficient space for their accommodation (no crucial interaction partners in the local radius of 4.5 Å). However, position C7 of the Trp57 ring makes a number of crystallographic distance interactions with various neighboring nitrogen and oxygen atoms, e.g. the backbone nitrogens of C48 (3.89 Å), M218 (4.17 Å) and H217 (3.74 Å). Additional interactions of C7 include the backbone oxygens of D216 (3.61 Å), H217 (3.5 Å) and F46 (4.2 Å) as well as the oxygen of the D216 side chain (4.37 Å; Figure 3b).

By incorporating different Trp-analogs substituted at indole position 4, we could confirm above described structural considerations. For example, (4-Amino)Trp and (4-Methyl)Trp are well accommodated in ECFP structures, whereas (5-Amino)Trp and (6-Amino)Trp are not translatable into the protein at all [8]. The 4-fluorination and 4-methylation of Trp57 are also beneficial for protein stability because such substitutions enhance the indole ring hydrophobicity. The same holds true for the aromatic and hydrophobic sulfur- and seleno-containing Trp surrogates seleno-pyrrolyl- and thieno-pyrrolyl-alanines (a detailed survey of the structures with these substitutions is available in the literature [7]). However, from the ECFP structure it is not clear why (4-Aza)Trp is not well tolerated at position 57.

Introduction of (7-Aza)Trp with its free electron pair at position 57, however, could lead to numerous repulsive interactions between the free electron pair of the (7-Aza)Trp nitrogen and other oxygen and nitrogen atoms in proximity. In agreement with this, (7-F)Trp was the only successful incorporation of a position 7 substituted Trp into ECFP and EGFP [7]. Obviously, the small fluorine atom is the only well-tolerated substituent at this position.

In conclusion, the structural surrounding of the C7 indole ring position could well explain the impaired folding of (7-Aza)Trp-ECFP. On the contrary, according to the ECFP structure, (4-Aza)Trp should be tolerated. However, the introduction of a higher overall hydrophilicity by (4-Aza)- and (7-Aza)Trp incorporation into the hydrophobic minicore around position 57 is the most plausible explanation for the significantly reduced folding capacity of (4-Aza)- and (7-Aza)Trp-ECFP.

In the context of a living cell, the protein synthesis machinery is part of a complex and crowded environment in which the polypeptide chain must fold into a tertiary structure [20]. The folding of a newly synthesized polypeptide chain can take place via co- or posttranslational pathways [21]. In ECFP or EGFP, these processes are additionally complicated by the fact that initial protein folding is followed by chromophore formation. Correspondingly, the nature of the non-canonical amino acid represents a crucial factor in the *in vivo* folding mechanism and thus determines the outcome of an expression experiment. Our study strongly supports a correlation between polypeptide synthesis with synthetic amino acids and *in vivo* protein folding.

Summary

The results of this study demonstrate that Trp57 with its large surface, aromaticity and hydrophobicity is crucial for an efficient *av*GFP folding and subsequent chromophore formation. Although the isosteric analogs 4- and 7-azatryptophan mimic Trp in size and shape, their hydrophilic properties are proposed to be the main cause for impaired folding and excessive loss in solubility in both fully labeled EGFP and ECFP structures, thus strongly affecting the *in vivo* folding mechanism.

Materials and Methods

All standard chemicals were purchased from Sigma (Sternheim, Germany) or Merck KGaA (Darmstadt, Germany) unless otherwise indicated. 4-Azaindole and 7-Azaindole were obtained from Biosynth (Staad, Switzerland). (4-Aza)Trp and (7-Aza)Trp were prepared as described elsewhere [8]. The crude products were added to the cells without further purification.

Journal of PeptideScience

Fermentation, Analog Incorporation and Protein Purification

Protein expression was performed using the Trp auxotrophic *Escherichia coli* strain JE7345 [NBRP (NIG, Japan) : *E. coli*]. Cells were transformed with a pQE80L vector (Qiagen, Hilden, Germany) carrying the ECFP or EGFP gene with an *N*-terminal His-tag. Transformants were grown in new minimal medium [22] containing 7.5 μM Trp as natural substrate until depletion of Trp and an optical density at 600 nm between 0.6 and 0.8. Subsequently, the crude (Aza)Trp preparations were added to the medium, and target protein expression was induced by addition of 1 mM isopropyl β-D-1-thiogalactopyranoside. Protein expression was performed for 4 h at 30 °C with vigorous shaking. After cell harvest, cells were lysed with a French Press and the lysate cleared by high-speed centrifugation (15 000 *g*, 4 °C, 30 min). Soluble target protein was purified from the supernatant using a Ni–NTA column (GE Healthcare, Munich, Germany) equilibrated with sodium dihydrogen phosphate buffer (50 mM, pH 8.0) containing 500 mM NaCl. Proteins were eluted with an imidazole gradient (0–250 mM) and buffer was exchanged to Tris–HCl buffer (50 mM, pH 8.0). Protein purity was analyzed by SDS–PAGE and Coomassie staining. For purification of insoluble target protein fractions, the pellet from high-speed centrifugation was solubilized in sodium dihydrogen phosphate buffer (50 mM, pH 8.0) containing 8 M urea and 5 mM DTT. Purification procedure was carried out like for the soluble protein fractions but under denaturing conditions with 8 M urea and 5 mM DTT. Purified insoluble protein fractions were brought to solubility by subsequent dialysis against Tris–HCl buffer (50 mM, pH 8.0) supplemented with 400 mM arginine and 1 mM DTT at room temperature overnight and then against Tris–HCl buffer (50 mM, pH 8.0) for 3.5 h.

Refolding of Insoluble Protein Fractions

Refolding trials were performed with the FoldIt kit from Hampton Research (Aliso Viejo, CA, USA) according to the manufacturer instructions. On-column refolding was performed as follows. The insoluble protein fractions were solubilized in sodium dihydrogen phosphate buffer (50 mM, pH 8.0) containing 8 M urea and 5 mM DTT and bound to a Ni–NTA column. Protein was refolded on the column by buffer exchange to dihydrogen phosphate (50 mM, pH 8.0) with 500 mM NaCl. The buffer exchange rate was 0.75 column volumes (CVs) per hour and the total buffer volume 15 CV. After refolding, proteins were eluted with an imidazole gradient (0–250 mM).

Refolding by dilution was carried out by slow dilution of insoluble protein fractions (in sodium dihydrogen phosphate buffer (50 mM, pH 8.0) containing 8 M urea and 5 mM DTT) in 50-fold of their volume dihydrogen phosphate (50 mM, pH 8.0) with 500 mM NaCl. None of the refolding attempts led to functional fluorescent ECFP variants.

Estimation of the Protein Concentration by Bradford Protein Assay

The protein concentrations were determined by the method of Bradford [23] using ECFP for calibration. The concentration of ECFP for the calibration curve was determined by UV_{280} absorbance and using the ε_M from the software ProtParam provided by the Expasy Proteomics Server (www.expasy.ch/tools/#proteome).

High Performance Liquid Chromatography

Proteins (100 μg) were purified by RP-HPLC (Waters, Milford, MA, USA) on a EC 125/4 nucleosil 100-5 C8 column (Macherey-Nagel, Düren, Germany) by eluting with an acetonitrile gradient in aqueous 0.1% (v/v) TFA: from 20 to 30% B within 5 min then to 60% B in 30 min (eluent B: 0.08% (v/v) TFA in acetonitrile). A flow rate of 1.5 ml min^{-1} was used and the absorbance was monitored between 200 and 450 nm.

Mass Analysis

For LC-ESI-MS, 20 μl aliquots of the purified proteins were pre-separated on a Discovery Bio Wide Pore C5 column (3.5 μm particle size, 100 × 2.1 mm, Supelco, Bellefonte, PA, USA) by eluting with a gradient from 20 to 90% B within 15 min and 90 to 95% B in 2 min [eluent A: 0.05% (v/v) TFA in water, eluent B: 0.05% (v/v) TFA in acetonitrile]. A flow rate of 250 μl min^{-1} was used. The masses of the eluted fractions were analyzed on a MicroTOF ESI-MS (Bruker Daltonics, Bremen, Germany). The absorbance was detected at 210 nm. With our ESI-LC-MS experimental setup, we can unambiguously detect a 1 Da mass difference with an error range of ±0.3 Da.

In-gel digestions of the proteins and subsequent mass analysis of the resulting peptide fragments were performed as follows.

In-Gel Digestion

Gel bands were cut into 1 mm^3 cubes and washed two times with 50 mM ammonium bicarbonate, 50% ethanol. For protein reduction, gel pieces were incubated with 10 mM DTT in 50 mM ammonium bicarbonate for 1 h at 56 °C. Alkylation of cysteines was performed by incubating the samples with 10 mM iodoacetamide in 50 mM ammonium bicarbonate for 45 min at 25 °C in the dark. Gel pieces were washed two times with 50 mM ammonium bicarbonate, 50% ethanol, dehydrated with 100% ethanol and dried in a vacuum concentrator. The gel pieces were rehydrated with 12.5 ng μl^{-1} trypsin (sequencing grade, Promega, Mannheim, Germany) in 50 mM ammonium bicarbonate and incubated overnight at 37 °C for protein digestion [24]. Supernatants were transferred to fresh tubes, and the remaining peptides were extracted by incubating the gel pieces two times with 30% acetonitrile in 3% TFA followed by dehydration with 100% acetonitrile. The extracts were combined and desalted using RP-C18 StageTip columns, and the eluted peptides were subjected to mass spectrometric analysis.

Nanoliquid Chromatography Mass Spectrometry/Mass Spectrometry

All digested peptide mixtures were separated by on-line NanoLC and analyzed by ESI-MS. The experiments were performed on an Agilent 1200 nanoflow system connected to an LTQ Orbitrap mass spectrometer (Thermo Electron, Bremen, Germany) equipped with a nanoelectrospray ion source (Proxeon Biosystems, Odense, Denmark). Binding and chromatographic separation of the peptides took place in a 15-cm fused silica emitter (75 μm inner diameter from Proxeon Biosystems) in-house packed with reversed-phase ReproSil-Pur C18-AQ 3 μm resin (Dr Maisch GmbH, Ammerbuch-Entringen, Germany). Peptide mixtures were injected onto the column with a flow of 500 nl min^{-1} and subsequently eluted with a flow of 250 nl min^{-1} from 2 to 40% acetonitrile in 0.5% acetic acid, in a 100-min gradient. The precursor ion spectra

were acquired in the Orbitrap analyzer (m/z 300–1800, $R = 60\,000$ and ion accumulation to a target value of 1 000 000), and the five most intense ions were fragmented and recorded in the ion trap. The lock mass option enabled accurate mass measurement in both MS and Orbitrap MS/MS mode as described previously [25]. Target ions already selected for MS/MS were dynamically excluded for 90 s.

Peptide Identification Via MASCOT Database Search

The data analysis was performed with the MaxQuant software supported by Mascot as the database search engine for peptide identifications as described [26]. Peaks in MS scans were determined as three-dimensional hills in the mass-retention time plane. MS/MS peak lists were filtered to contain at most six peaks per 100 Da interval and searched by Mascot (Matrix Science, London, UK) against a concatenated forward and reversed version of *E. coli* K12 data (extracted from NCBInr and containing the sequence of ECFP with frequently observed contaminants like proteases and human keratins). The initial mass tolerance in MS mode was set to 7 ppm and MS/MS mass tolerance was 0.5 Da. Cysteine carbamidomethylation was searched as a fixed modification, whereas *N*-acetyl protein, oxidized methionine and azatryptophan incorporation were searched as variable modifications.

UV/Vis Spectroscopy

UV/Vis spectra were recorded on a Lambda 19 photometer from Perkin Elmer (Überlingen, Germany). Spectra of native proteins were measured at 20 °C in Tris-HCl buffer (100 mM, pH 7.8). For measurement of absorbance under denaturing conditions, solutions of ECFP and (Aza)Trp variants were supplemented with 8 M urea and boiled for 10 min before measurement.

Fluorescence Spectroscopy

Unless otherwise stated, fluorescence spectra of the native proteins were measured at 20 °C in Tris–HCl buffer (100 mM, pH 7.4) using a luminescence spectrometer LS50B (PerkinElmer and Life Sciences, Boston, MA, USA). The protein concentration was 0.5 μM. Spectra were recorded from 320 to 600 nm with an excitation wavelength of 295 nm (slit 5/5). Spectra under denaturing conditions were measured after boiling proteins for 10 min in Tris–HCl buffer (100 mM, pH 7.8) supplemented with 8 M urea and 1 mM tris(2-carboxyethyl)phosphine. In this case, the protein concentration was 5 μM.

NMR Spectroscopy

NMR spectra were acquired at 300 K on a Bruker DRX 600 MHz spectrometer equipped with a cryoprobe. Samples contained 120 μM protein in sodium phosphate buffer (50 mM, pH 7.4). Water suppression was carried out using the WATERGATE 3-9-19 [27] sequence. NMR data were processed using the Bruker program Xwin-NMR version 3.5.

Analysis of Intramolecular Interactions

The analysis of intramolecular interactions of Trp57 was performed by Pymol (DeLano Scientific, Palo Alto, CA, USA) and LIGPLOT v4.5.4 [28].

Acknowledgements

Thomas Steiner is greatly acknowledged for his help with ECFP structure analysis and Dr Lars Merkel for critical reading of the manuscript. We thank Elisabeth Weyher-Stringl and Dr Cyril Boulegue for mass spectrometric analysis of our protein preparations. We also thank Oliver Baker for performing initial experiments.

Supporting information

Supporting information may be found in the online version of this article.

References

1 Phillips GN. Structure and dynamics of green fluorescent protein. *Curr. Opin. Struct. Biol.* 1997; **7**: 821–827.
2 Labas YA, Gurskaya NG, Yanushevich YG, Fradkov AF, Lukyanov KA, Lukyanov SA, Matz MV. Diversity and evolution of the green fluorescent protein family. *Proc. Natl. Acad. Sci. U. S. A.* 2002; **99**: 4256–4261.
3 Shimomura O. Structure of the chromophore of the *Aequorea* green fluorescent protein. *FEBS Lett.* 1979; **104**: 220–222.
4 Craggs TD. Green fluorescent protein: structure, folding and chromophore maturation. *Chem. Soc. Rev.* 2009; **38**: 2865–2875.
5 Palm GJ, Wlodawer A. *Green Fluorescent Protein*. Academic Press Inc.: San Diego, 1999; 378–394.
6 Budisa N, Pal PP. Designing novel spectral classes of proteins with a tryptophan-expanded genetic code. *Biol. Chem.* 2004; **385**: 893–904.
7 Budisa N, Pal PP, Alefelder S, Birle P, Krywcun T, Rubini M, Wenger W, Bae JH, Steiner T. Probing the role of tryptophans in *Aequorea victoria* green fluorescent proteins with an expanded genetic code. *Biol. Chem.* 2004; **385**: 191–202.
8 Bae JH, Rubini M, Jung G, Wiegand G, Seifert MHJ, Azim MK, Kim JS, Zumbusch A, Holak TA, Moroder L, Huber R, Budisa N. Expansion of the genetic code enables design of a novel "gold" class of green fluorescent proteins. *J. Mol. Biol.* 2003; **328**: 1071–1081.
9 Cash MT, Schreiner PR, Phillips RS. Excited state tautomerization of azaindole. *Org. Biomol. Chem.* 2005; **3**: 3701–3706.
10 Lepthien S, Hoesl MG, Merkel L, Budisa N. Azatryptophans endow proteins with intrinsic blue fluorescence. *Proc. Natl. Acad. Sci. U. S. A.* 2008; **105**: 16095–16100.
11 Fukuda H, Arai M, Kuwajima K. Folding of green fluorescent protein and the cycle3 mutant. *Biochemistry* 2000; **39**: 12025–12032.
12 Jackson SE, Craggs TD, Huang JR. Understanding the folding of GFP using biophysical techniques. *Expert Rev. Proteomics* 2006; **3**: 545–559.
13 Pal PP, Bae JH, Azim MK, Hess P, Friedrich R, Huber R, Moroder L, Budisa N. Structural and spectral response of *Aequorea victoria* green fluorescent proteins to chromophore fluorination. *Biochemistry* 2005; **44**: 3663–3672.
14 Kwon I, Tirrell DA. Site-specific incorporation of tryptophan analogues into recombinant proteins in bacterial cells. *J. Am. Chem. Soc.* 2007; **129**: 10431–10437.
15 Wang L, Xie JM, Deniz AA, Schultz PG. Unnatural amino acid mutagenesis of green fluorescent protein. *J. Org. Chem.* 2003; **68**: 174–176.
16 Reid BG, Flynn GC. Chromophore formation in green fluorescent protein. *Biochemistry* 1997; **36**: 6786–6791.
17 Barondeau DP, Putnam CD, Kassmann CJ, Tainer JA, Getzoff ED. Mechanism and energetics of green fluorescent protein chromophore synthesis revealed by trapped intermediate structures. *Proc. Natl. Acad. Sci. U. S. A.* 2003; **100**: 12111–12116.
18 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**: 476–490.
19 Steiner T, Hess P, Bae JH, Wiltschi B, Moroder L, Budisa N. Synthetic biology of proteins: tuning GFPs folding and stability with fluoroproline. *PLoS One* 2008; **3**: Article No. e1680.
20 Zhang G, Hubalewska M, Ignatova Z. Transient ribosomal attenuation coordinates protein synthesis and co-translational folding. *Nat. Struct. Mol. Biol.* 2009; **16**: 274–280.

21 Zhang G, Ignatova Z. Generic algorithm to predict the speed of translational elongation: implications for protein biogenesis. *Plos One* 2009; **4**: Article No. e5036.
22 Budisa N, Steipe B, Demange P, Eckerskorn C, Kellermann J, Huber R. High-level biosynthetic substitution of methionine in proteins by its analogs 2-aminohexanoic acid, selenomethionine, telluromethionine and ethionine in *Escherichia coli*. *Eur. J. Biochem.* 1995; **230**: 788–796.
23 Bradford MM. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.* 1976; **72**: 248–254.
24 Shevchenko A, Wilm M, Vorm O, Mann M. Mass spectrometric sequencing of proteins from silver stained polyacrylamide gels. *Anal. Chem.* 1996; **68**: 850–858.
25 Olsen JV, de Godoy LMF, Li GQ, Macek B, Mortensen P, Pesch R, Makarov A, Lange O, Horning S, Mann M. Parts per million mass accuracy on an orbitrap mass spectrometer via lock mass injection into a C-trap. *Mol. Cell. Proteomics* 2005; **4**: 2010–2021.
26 Gruhler A, Olsen JV, Mohammed S, Mortensen P, Faergeman NJ, Mann M, Jensen ON. Quantitative phosphoproteomics applied to the yeast pheromone signaling pathway. *Mol. Cell. Proteomics* 2005; **4**: 310–327.
27 Piotto M, Saudek V, Sklenar V. Gradient-tailored excitation for single-quantum NMR spectroscopy of aqueous solutions *J. Biomol. NMR* 1992; **2**: 661–665.
28 Wallace AC, Laskowski RA, Thornton JM. LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions. *Protein Eng.* 1995; **8**: 127–134.

