Photochemical & Photobiological Sciences

Cite this: *Photochem. Photobiol. Sci.*, 2012, **11**, 637

www.rsc.org/pps

PAPER

Multi-domain GFP-like proteins from two species of marine hydrozoans†

Marguerite E. Hunt,[a] Chintan K. Modi,[a] Galina V. Aglyamova,[b] D. V. S. Ravikant,[c,d] Eli Meyer[b] and Mikhail V. Matz*[b]

Received 2nd August 2011, Accepted 4th November 2011
DOI: 10.1039/c1pp05238a

Proteins homologous to Green Fluorescent Protein (GFP) are widely used as genetically encoded fluorescent labels. Many developments of this technology were spurred by discoveries of novel types of GFP-like proteins (FPs) in nature. Here we report two proteins displaying primary structures never before encountered in natural FPs: they consist of multiple GFP-like domains repeated within the same polypeptide chain. A two-domain green FP (abeGFP) and a four-domain orange-fluorescent FP (Ember) were isolated from the siphonophore *Abylopsis eschscholtzii* and an unidentified juvenile jellyfish (order Anthoathecata), respectively. Only the most evolutionary ancient domain of Ember is able to synthesize an orange-emitting chromophore (emission at 571 nm), while the other three are purely green (emission at 520 nm) and putatively serve to maintain the stability and solubility of the multidomain protein. When expressed individually, two of the green Ember domains form dimers and the third one exists as a monomer. The low propensity for oligomerization of these domains would simplify their adoption as *in vivo* labels. Our results reveal a previously unrecognized direction in which natural FPs have diversified, suggesting new avenues to look for FPs with novel and potentially useful features.

Introduction

Green fluorescent protein (GFP) was first noticed almost 50 years ago.[1] Since then, almost two hundred natural GFP-like fluorescent proteins (FPs) have been described.[2] Unique to the GFP protein family is the ability to synthesize light-emitting chromophores autocatalytically from their own three amino acid residues situated near the center of the FP globule.[3,4] Because of this, FPs became indispensable in biomedical and basic research as genetically encoded fluorescent labels.[2]

The advancement of FP-based *in vivo* labeling and detection technology was tightly coupled to the development of our understanding of where FPs can be found in nature and how diverse they can be, spectrally and structurally. The emission ranges of the currently available palette of FP-derived labels span almost the whole visible spectrum from blue to red;[2] however, several years had passed between the cloning of the first green FP from *Aequorea victoria*[5] and the discovery of other FP colors.[6] This initial lag was mainly due to the fact that researchers did not expect FPs to exhibit diverse chromophore structures and potentially different biological functions. The first two FPs to be characterized were both green and were found in bioluminescent cnidarians (jellyfish *Aequorea victoria* and sea pansy *Renilla reniformis*). In these animals, the green fluorescent protein acts as the chromatic shift agent for an internal bioluminescent light source.[7,8] It seemed reasonable to assume at the time that other FPs would also be found in species endowed with bioluminescence, and would most likely be green. Eventually, a step away from this pre-conception resulted in the isolation of six new FPs, among them, a yellow and a red one (the latter became best known under its commercial name, DsRed) from non-bioluminescent representatives of class Anthozoa.[6] The discovery of non-green FPs, and of DsRed in particular, resulted in major breakthroughs in *in vivo* labeling technologies, such as robust multicolor imaging,[2] deep tissue imaging, and applications capitalizing on unique properties of DsRed, such as a fluorescent "timer".[9] Extensive searches for FPs in non-bioluminescent cnidarians yielded markers that were substantially brighter and more stable than the earlier ones[2] as well as unique tools such as photoswitchable labels[10–12] and photosensitizers.[13] In 2004, FPs were discovered outside the phylum Cnidaria[14] in the phylum Arthropoda, resulting in some of the brightest and fastest maturing green fluorescent reporters to date.[15,16] The latest news in the area of natural FPs came in 2007, when a number of green FPs were found in a representative of the phylum Chordata, the lancelet *Branchiostoma floridae*,[17] which are now being marketed for laboratory use (Allele Biotechnology; San Diego, CA).

One of the structural features shared by the majority of natural FPs is the tendency to form oligomers (stable complexes of multiple identical protein molecules), and in particular, a

[a]*Institute for Cellular and Molecular Biology, the University of Texas at Austin, Austin, TX 78712, USA*
[b]*Section of Integrative Biology, the University of Texas at Austin, Austin, TX 78712, USA. E-mail: matz@mail.utexas.edu*
[c]*Department of Computer Science, Cornell University, 4130 Upson Hill, Ithaca, NY 14853, USA*
[d]*Institute for Computational Engineering and Sciences, the University of Texas at Austin, Austin, TX 78712, USA*

†Electronic supplementary information (ESI) available. See DOI: 10.1039/c1pp05238a

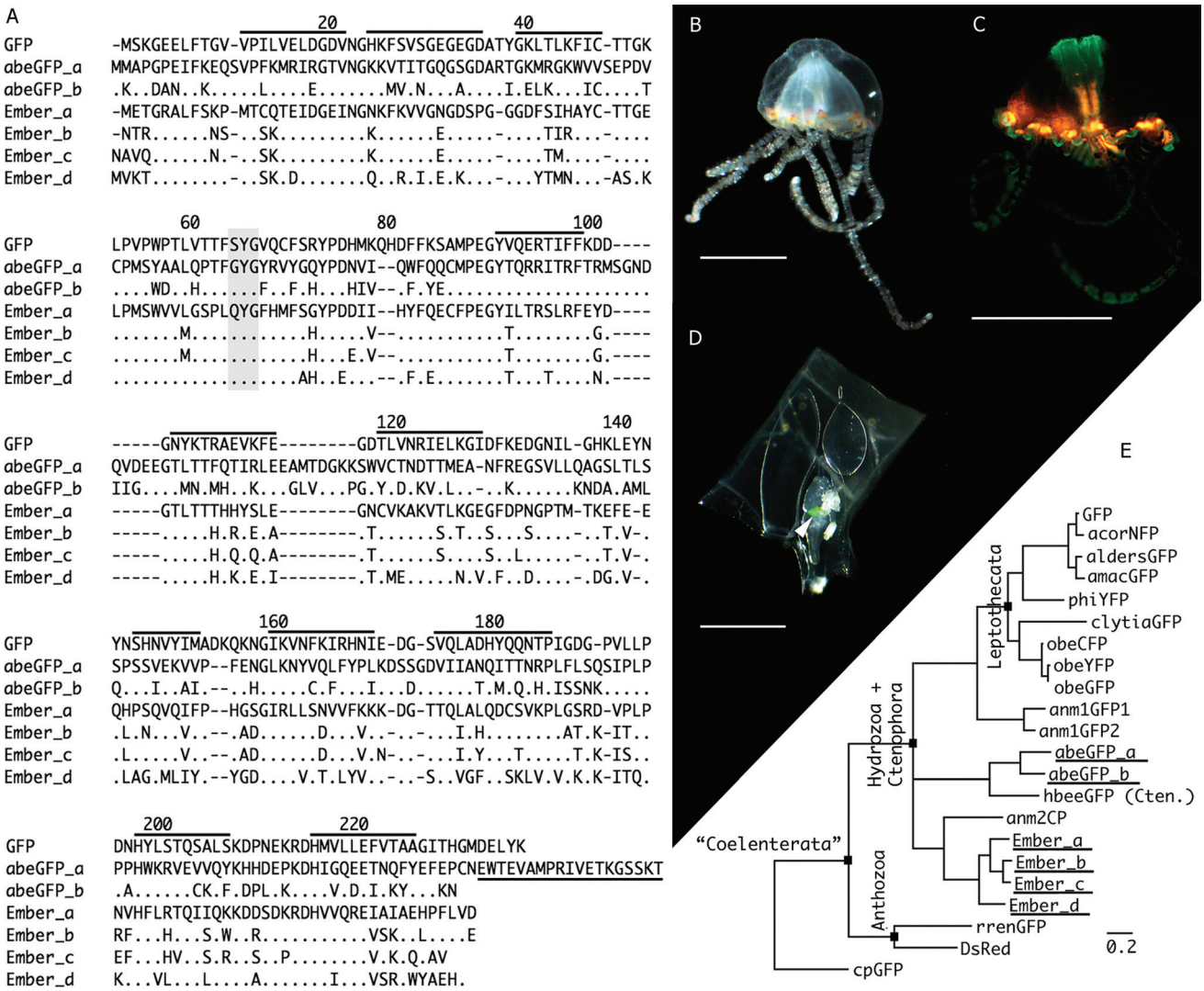

Fig. 1 Diversity of organisms and proteins studied. (A) Alignment of amino acid sequences of the GFP-like domains of the new proteins against GFP from *Aequorea victoria*. The residues in subsequent domains that are identical to the first (N-terminal) domain are shown as dots. The numbering is according to GFP sequence. The regions forming beta-sheets are indicated by lines above the alignment, the chromophore-forming triad is shaded. The linker sequence between the two domains of abeGFP is underlined (residues 234–252). (B) Unidentified anthoathecate jellyfish, white light. (C) The same specimen as in B, fluorescent optics. (D) Siphonophore *Abylopsis eschscholtzii* (without posterior nectophore), under combined white and blue light illumination. The arrowhead indicates the green-fluorescent base of the gastrozooid. On panels B–D, the scale bar is 5 mm. (E) Phylogenetic tree of the known hydrozoan GFP-like proteins. The names of GFP-like domains from the new proteins are underlined. Only the nodes with posterior probability 0.95 and higher are shown. Scale bar: 0.2 replacements per amino acid site.

characteristic tetrameric structure ("four-can pack") first observed in DsRed.[18,19] Despite the pervasive tendency to form oligomeric arrangements at the quaternary structure level, and a notable propensity for frequent gene duplications within the genome,[20,21] not a single wild-type GFP-related protein was thus far found to contain multiple GFP-like domains within a single polypeptide chain. Here we report two such proteins, isolated from representatives of two orders of class Hydrozoa (phylum Cnidaria), Siphonophorae and Anthoathecatae, and explore possible relationships between individual domains within the four-domain orange-red protein from the anthoathecate jellyfish. Our observations reveal a novel structural solution by which diverse fluorescent characteristics may be maintained in natural FPs, which calls for broadening of the scope of studies seeking to discover unusual FP types.

Results

Cloning of FP-coding open reading frames

Screening of the bacterial expression library derived from the jellyfish revealed an orange-fluorescent cDNA clone (accession number HQ699262) encoding a 907 amino acids long chain of four concatenated GFP-like domains (Fig. 1A), with a predicted molar mass of 101 kDa. Since the species of the juvenile anthoathecate jellyfish remained unidentified, we named this

four-domain protein Ember due to the fiery fluorescent appearance it confers to the organism (Fig. 1B, C). The four individual domains of Ember share 57% identical amino acid residues and are concatenated without any apparent linker sequences (Fig. 1A).

The green-fluorescent GFP-like protein from the siphonophore (accession number HQ699261) was cloned by 3′-RACE based on the N-terminal fragment identified *via* 454-sequencing of cDNA, and contained two concatenated GFP-like domains. This protein was 519 amino acids long with a 59 kDa predicted molar mass. This green fluorescent protein was named abeGFP, with "abe" to reflect its host species, *Abylopsis eschscholtzii* (Fig. 1D). The two domains of abeGFP are 59% identical, and are joined *via* a linker 19 amino acids long (Fig. 1A). Both Ember and abeGFP are expressed in bacteria as polypeptides corresponding in length to their respective primary structure, according to denaturing gel-electrophoresis (data not shown).

Phylogenetic analysis

In a phylogenetic tree of all known hydrozoan FPs rooted with the bilaterian protein cpGFP (phylum Arthropoda, class Copepoda) and two anthozoan proteins, DsRed and rrenGFP (Fig. 1E), individual domains within each protein group together, suggesting that the multi-domain structures arose *via* tandem duplications. Domains of abeGFP, representing a previously unsampled order of Hydrozoa (Siphonophora), branched off in a basal position, ostensibly reflecting the phylogenetic relationships of the host species. Grouping of abeGFP with the ctenophore protein hbeeGFP is somewhat surprising, as is the position of hbeeGFP within the clade of hydrozoan proteins, and is likely to the possibility that hbeeGFP is a hydrozoan protein derived from the ctenophore's gut content, as we have previously suggested.[2] Domains of Ember grouped with the only previously identified hydrozoan protein with a non-GFP-like chromophore, the purple chromoprotein anm2CP,[14] to the exclusion of green fluorescent proteins from the same order (Anthoathecata, anm1GFP1 and anm1GFP2). This result indicates that Ember belongs to a paralogous FP lineage in Anthoathecata, which is characterized by non-green chromophores (fluorescent red or non-fluorescent purple).

Spectroscopic characteristics

In vivo fluorescence emission maxima were 569 nm and 515 nm for the intact jellyfish and siphonophore, respectively (Fig. 2A). The heterologously expressed siphonophore protein abeGFP had a single excitation peak at 502 nm and a single emission maximum at 517 nm (Fig. 2B). The jellyfish protein Ember has two excitation peaks at 510 and 555 nm, with two corresponding peaks of fluorescence: a green one at 520 nm, and an orange-red one at 571 nm (Fig. 2C).

The individually expressed domains a–c of Ember demonstrated nearly identical purely green fluorescence (excitation max at 508 nm, emission max at 520 nm, Fig. 2D). In contrast, the fourth domain (Ember-d) was predominantly orange-red fluorescent with the excitation max at 552 nm and emission max at 571 nm (Fig. 2D). The absorption spectra of alkali- and acid-denatured Ember-d match the analogous spectra for DsRed:[14] alkali-denatured protein absorbs at 450 nm, and acid-denatured protein absorbs at 425 nm initially and gradually shifts to 390 nm absorption (data not shown). The individually expressed N-terminal FP domain of abeGFP conferred only barely detectable fluorescence to the bacteria. No fluorescence was detectable when expressing the C-terminal domain of abeGFP separately.

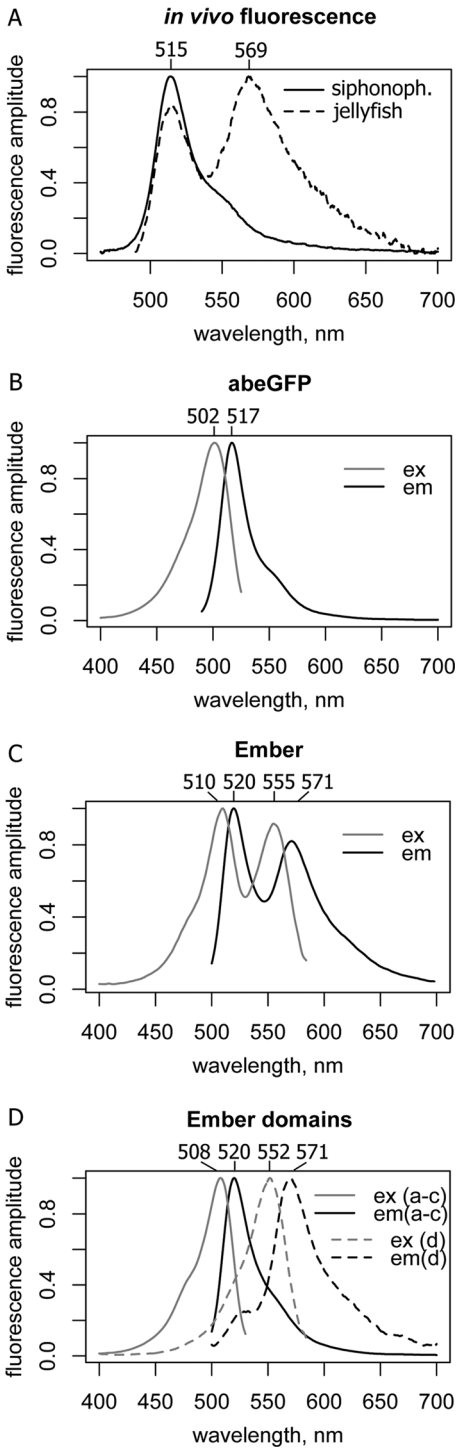

Fig. 2 Fluorescence spectra of the biological specimens (A) and excitation-emission (ex, em) spectra of the heterologously expressed new proteins and individual domains (B–D).

Table 1 Brightness characteristics of the new proteins

Protein	Source	ME, M^{-1} cm^{-1}	QY	Relative brightness
abeGFP	*Abylopsis eschscholtzii*, order Siphonophora	168 000	0.66	1.68[a]
Ember	jellyfish, order Anthoathecata	370 000	0.34	0.95[a]
Ember-a	first (N-terminal) domain of Ember	76 000	0.64	1.47
Ember-b	second domain of Ember	72 000	0.7	1.53
Ember-c	third domain of Ember	71 000	0.54	1.16
EGFP (standard)	mutant variant of GFP from *Aequorea victoria*, order Leptothecata	55 000	0.6	1

[a] Per single chromophore.

The brightness characteristics (molar extinction coefficients ME, M^{-1} cm^{-1}, and quantum yield of fluorescence QY) of abeGFP (siphonophore), full-length Ember (anthoathecate), and three green-fluorescent Ember domains (a–c) are listed in Table 1. The measurements were not successful for Ember-d domain due to poor expression yield and high aggregation tendency.

Oligomeric status

The size-exclusion chromatography with combined scattering and refractometry detection indicated that abeGFP protein is predominantly a monomer (50 kDa, corresponding to two FP domains), with a small fraction of oligomers (>400 kDa, Fig. 3A). Heterologously expressed Ember tends to form aggregates with molecular mass about 400 kDa, corresponding to four Ember chains; although a smaller peak corresponding to a single Ember chain (~100 kDa) was also present (Fig. 3B). There was also a peak corresponding to a 50–60 kDa protein, which is twice less than a single full-length Ember, ostensibly representing the product of Ember hydrolysis. The aggregation profile of Ember-d, the red fluorescent domain, largely recapitulates the full Ember pattern, with a higher proportion of the dimeric 50 kDa form (Fig. 3B). Individually expressed Ember domains a and b each eluted as a single peak containing a clearly defined dimeric form with molecular mass 53–54 kDa (Fig. 3C); however, Ember-a was eluted earlier than Ember-b, indicating that the spatial configuration of the dimer is different between them. Individually expressed Ember-c was predominantly monomeric at ~30 kDa.

Hypothetical structure of Ember

We asked whether the way the Ember domains are connected would allow for a canonical tetrameric structure to be formed from a single Ember polypeptide chain. The top-scoring structural model of Ember based on DsRed tetramer coordinates (PDB structure 1G7K[19]) indicated that straightforwardly folding Ember into a canonical tetramer requires breakage of the polypeptide chain. We then considered a hypothetical structure in which the Ember polypeptide was threaded through the interface between interacting canonical tetramers (Fig. 4A), and found that it was compatible with the way the individual domains were connected (Fig. 4B, the coordinates available in the ESI†). This hypothetical structure received the highest stability score among all the alternative models suggested by two oligomerization prediction algorithms: CombDock[22] and ASSEMBLE.[23]

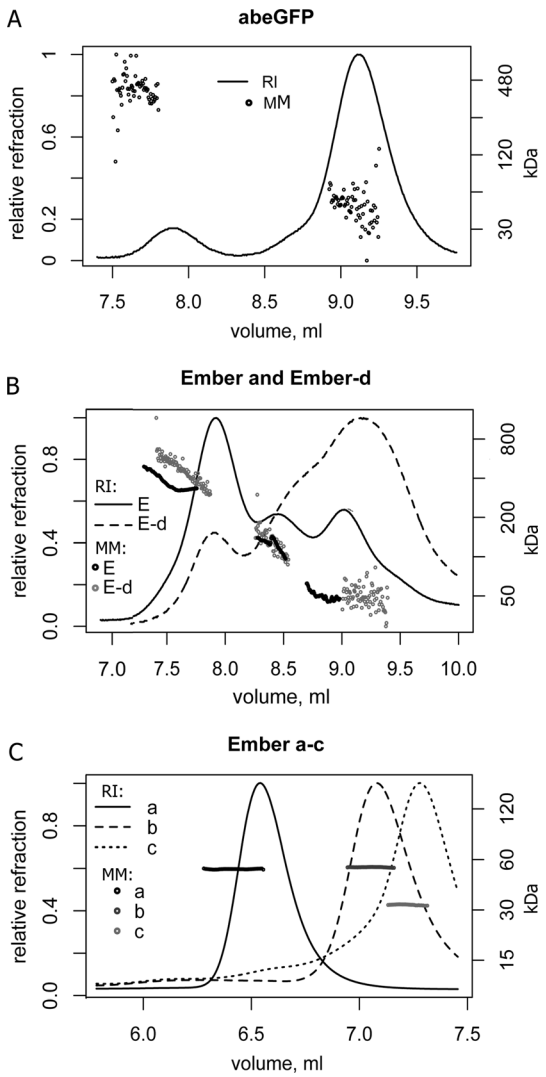

Fig. 3 Oligomerization analysis of the heterologously expressed proteins and individual domains. Horizontal axis denotes elution volume from a size-exclusion column. Curves show the elution profile based on refraction index (left vertical axis), dots indicate the molar mass of the eluted protein (right vertical axis) inferred from light scattering.

Discussion

We report two new proteins cloned from two individual specimens of Hydrozoa. Although there is technically a chance that these sequences come from some other sources – such as gut content, bacterial fauna, or some other uncontrolled

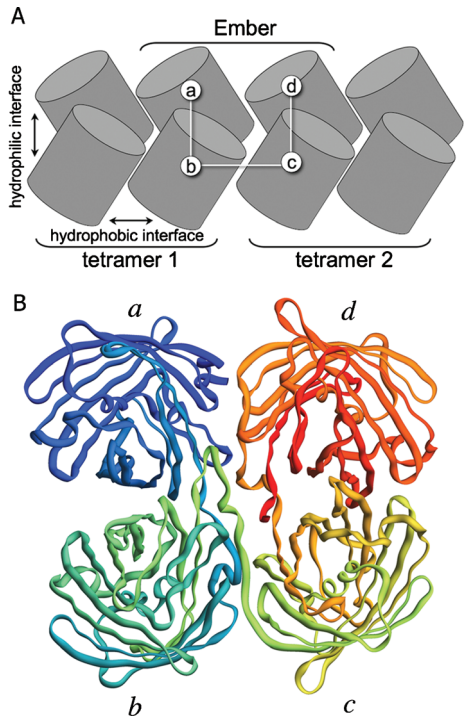

Fig. 4 Hypothetical structure of Ember, allowing for typical tetrameric arrangement of FP domains. (A) Conceptual drawing of a hypothesized complex, where Ember is threaded through the interface between canonical tetramers. (B) Top-scoring I-TASSER model of Ember following the hypothesized arrangement. Individual domains are labeled in italic letters. The molecule is colored according to the residue number, from blue (N-terminus) to red (C-terminus).

contaminations – we believe that the proteins belong to the specimens shown on Fig. 1B–D, for two reasons. First, the sequences of the proteins are decidedly hydrozoan (Fig. 1E). Second, the spectral properties of the cloned proteins match very well the native fluorescence of the animals (Fig. 2).

Ember is the first GFP-like protein endowed with orange-red fluorescence cloned from a hydrozoan representative, a jellyfish from the order Anthoathecata. Three observations indicate that, of the two types of red-fluorescent chromophores known in GFP-like proteins, DsRed-like[24] and Kaede-like,[25] Ember appears to synthesize a DsRed-like chromophore. First, the red fluorescent peak in the jellyfish and heterologously expressed Ember (Fig. 2 C) is broad and without a shoulder above 600 nm, which is much more similar to DsRed-like than to Kaede-like fluorescence. The absorption of Ember-d (the orange-fluorescent domain of Ember) is also similar to DsRed under denaturing conditions. Second, blue light illumination is not required for the red fluorescence to appear in Ember-expressing bacteria, unlike in Kaede-like proteins.[10] Third, none of the Ember domains has a histidine residue in the first position of the chromophore-forming triad (amino acid 65 according to GFP numeration), the side chain of which contributes an essential part of the Kaede-like chromophore.[25]

Of the four Ember domains, the first one to arise in evolution is the fourth one (Ember-d), the only domain exhibiting orange-red fluorescence (Fig. 2D). Since the closest Ember relative is the purple anm2CP with a DsRed-related chromophore (more specifically, an isomerised non-fluorescent variant of the DsRed-like chromophore[14]), the ability to synthesize a DsRed-chromophore is likely to be the ancestral feature of all Ember domains. The three purely green-fluorescent domains (Ember a–c), therefore, represent reversals to the basal green fluorescent chromophore type.

The loss of red fluorescence in the phylogenetically younger domains might be explained in two ways. First, due to efficient Förster resonance energy transfer (FRET) between FP domains in the polypeptide, the resulting fluorescence would stay mostly red as long as at least one domain containing the red chromophore is in close contact with the green-fluorescent domains. In such a situation, the red fluorescence might have been lost simply due to mutation pressure, without affecting the phenotype much. Alternatively, natural selection might have facilitated the evolution of green fluorescence in the duplicate domains to achieve an extended Stokes shift enabled by FRET (from 510 nm excitation to 571 nm emission, Fig. 2C). It is difficult to evaluate the likelihood of such a "photonics" scenario, since at the moment the biological function of hydrozoan fluorescent proteins remains obscure and may not even be directly related to the ability of these proteins to emit light.[26]

Phylogenetic analysis suggests that Ember domains originated *via* three tandem duplications (Fig. 1E). The first duplication generated the ancestor of the red Ember-d fused to the C-terminus of another FP domain. This N-terminal domain was duplicated twice more, inserting two more FP domains between Ember-d and the N-terminus of the four-domain protein. Notably, these duplications happened consecutively, not simultaneously, because an appreciable amount of sequence divergence had time to accumulate in between, allowing for robust phylogenetic reconstruction (Fig. 1E). This suggests that the intermediate forms – the ones consisting of two and three FP domains – might have also been fluorescent.

The red-fluorescent Ember-d domain shows higher propensity to oligomerize than other domains (Fig. 3B, C). The dependence of red fluorescence on oligomerization was first noted while engineering the naturally tetrameric DsRed into a monomeric state: any mutation that disrupted the oligomerization interface tended to be very detrimental for red fluorescence.[27] Our recent study of the evolution of Kaede-like red fluorescent proteins[28] provides even more direct evidence: of twelve historical mutations that were required to evolve red fluorescence within the green ancestral protein, seven affected the interface between FP domains in the tetrameric structure. It can be speculated that historically, the synthesis of the red chromophore in the Ember ancestor was largely enabled by interactions between individual FP domains in an oligomeric complex, and that the multi-domain structure arose to stabilize this interaction while maintaining the solubility of the protein.

All natural FPs, with the only exception of GFP from *A. victoria*, interact strongly to form dimers, tetramers or higher-order oligomers, which presents a considerable challenge during adaptation of wild-type FPs for use as molecular tags.[2] Remarkably, one of the green-fluorescent domains of Ember, Ember-c, behaves as a monomer when expressed individually (Fig. 3C). This observation suggests a novel direction in which the monomeric molecular tags can be sought in nature, that is, by cloning individual FP domains out of multi-domain FPs.

The Ember domains are connected to each other without any additional linker sequence in between (Fig. 1A), which does not allow for the formation of a canonical FP tetramer, observed in the wide diversity of natural FPs,[15,18,29–31] from a single Ember chain. This implies that Ember either assumes an alternative quaternary structure, or forms canonical tetramers *via* interaction of several polypeptides. One possible arrangement that would allow for the latter is shown on Fig. 4. This arrangement assumes interaction between canonical tetramers, in which the Ember polypeptide is threaded through domains that are bound *via* hydrophilic interactions within each tetramer, and across tetramers. Such a structure would strongly stabilize oligomeric complexes, both because the hydrophilic (weaker) interface within each tetramer is augmented by the covalent bonding, and because it binds the tetramers together. This hypothetical structure would enable formation of chains of Ember molecules (Fig. 4A), resembling the oligomeric macrohelix formed by homodimers of KillerRed protein.[32] It is perhaps notable that KillerRed is a mutated version of anm2CP, the closest relative of Ember in the phylogeny (Fig. 1E). The *in silico* experiment looking for potential alternative arrangements of Ember domains indicated that the hypothetical structure shown on Fig. 4B is the most stable one. It should be noted, however, that this result is conditional on the placement of the intervening loop regions forming the between-domain interface in the hypothetical structure (Fig. 4B). These regions could not be modeled based on existing X-ray data for other GFP-like proteins since they correspond to the disordered parts of their structure. In our *in silico* experiment, their coordinates were held as suggested by threading of the Ember sequence through the hypothetical model (Fig. 4B). We cannot exclude the possibility that a more optimal configuration of these regions may be achievable in some alternative models, but the existing computational approaches do not yet allow for thorough exploration of this structure space. Fortunately, most GFP-like proteins can be crystallized relatively easily, so the exact Ember structure will most likely be resolved experimentally in the near future.

In contrast to Ember, abeGFP is a purely green, predominantly monomeric protein, consisting of two FP domains. This fact invalidates the most extreme formulation of the hypothesis outlined above, that multi-domain FP arrangements evolve exclusively to stabilize red fluorescence. Still, the more general version of our hypothesis linking the multi-domain arrangement to the overall stability of the protein is supported by the fact that abeGFP domains developed none or barely detectable fluorescence when expressed individually in bacteria. Although it is tempting to speculate that the multi-domain structure evolved because it enabled some (yet unknown) novel aspects of the protein function, such a dependence might also have originated due to the process known as "constructive neutral evolution" or "neutral evolutionary ratchet".[33] Under such a scenario, domain duplication did not result in any changes in protein function, but was followed by accumulation of neutral mutations in the individual domains that eventually made their function dependent on their interaction within the multidomain structure.

The discovery of natural multidomain FPs opens up an avenue for future research into possible emergent features of the multidomain organization and prompts further enquiry into the mechanisms of functional diversification and complexity evolution in GFP-like proteins.[28,34,35] Such studies could advance our understanding of the biological function of FPs, their structure-function relationships, and may suggest novel biotechnology applications exploiting these features. Our study also highlights the need for broader characterization of natural FP diversity, since the multidomain organization may not be the last surprise in this fascinating protein family.

Experimental

Specimens and library preparation

A single specimen of siphonophore showing green fluorescence in the base of the gastrozooid (Fig. 1D, Casey Dunn, personal communication) and a single specimen of juvenile orange fluorescent jellyfish (Fig. 1B, C) were identified by illumination with blue light (Blue Star, Night Sea; Andover, MA) following a night-time plankton tow off Little San Salvador Island in the Bahamas on August 19, 2007 (25°1.3′N, 77°36.2′W). The specimens were identified[36] to the species level for the siphonophore (*Abylopsis eschscholtzii* [Huxley, 1859]) and to the order Anthoathecatae in case of the jellyfish. The jellyfish was inferred to be a juvenile due to the lack of well-developed gonads. Total RNAs were extracted from intact animals using RNAqueous Micro kit (Ambion; Austin, TX) and stored at −80 °C in 6.65 M LiCl until we returned to Austin. For both organisms, the cDNA was synthesized and bacterial expression libraries were constructed as we described previously.[16]

Sequencing of siphonophore cDNA with 454 GS FLX Titanium

A siphonophore cDNA library for 454 sequencing was prepared as described previously,[37] with the adaptor sequences appropriately modified to be compatible with the Titanium version of the 454 technology, and omitting the step of cDNA normalization. We amplified and sequenced only the 5′-cDNA ends to maximize the chance of encountering protein-coding regions. The latest version of the protocol, detailing these modifications, is available in the ESI and on the Matz lab website (http://www.bio.utexas.edu/research/matz_lab/matzlab/Methods.html). A total of 6459 reads were obtained. BLASTX[38] analysis of the sequence results identified a single read encoding an N-terminal fragment of a GFP-like protein. The missing 3′-cDNA fragment was obtained using step-out PCR procedure, as described previously.[39,40]

Phylogenetic analysis

The sequences of the individual domains of the new proteins were aligned with the family-wide FP alignment[2,41] using Clustal,[42] followed by manual editing in Genedoc software.[43] A subset of sequences corresponding to all hydrozoan proteins plus a ctenophore protein[44] was extracted, along with two anthozoan and one bilaterian (copepod) sequence to serve as outgroups. The accession numbers of the included sequences are: GFP, P42212; acorNOFP, AY151052; aldersGFP, ACC54354; amacGFP, AF435432; clytiaGFP, 2HPW_A; obeCFP, JN385282; obeGFP, JN385283; obeYFP, JN385284; anm1GFP1, AY485334;

anm1GFP2, AY485335; hbeeGFP, ACX47247; anm2CP, AY485336; rrenGFP, AF372525; DsRed, AF168419; cpGFP, AB185173. The resulting alignment can be obtained from the Supplementary Online Material. For phylogenetic analysis, the poorly aligned C-terminal portion (5–15 amino acids following the last beta-sheet of the FP fold[4]) was removed from the alignment. The alignment was analyzed in MrBayes 3.1,[45] with the following settings: aamodelpr = mixed, ngen = 1000000, printfreq = 500, samplefreq = 200, nchains = 4, burnin = 2500.

Heterologous expression

The fluorescent clones derived from the jellyfish, as well as 3′-cDNA fragments for the siphonophore protein, were sequenced by primer walking using the Sanger method. After identification of the complete open reading frames (ORFs), they were amplified from the original cDNA using primers bearing translation initiation signals and 5′-heels to facilitate bacterial expression, as described earlier.[20] The amplification products were ligated into pGEM-T vector (Promega; Madison, WI) and transformed into Z-competent XJb autolysis *E. coli* cells (Zymo Research; Orange, CA). The product of expression was purified from the bacteria using metal-affinity chromatography on Ni-NTA agarose (Qiagen; Valencia, CA) according to the manufacturer's protocol. Individual domains of the four-domain jellyfish FP were subcloned and expressed using the same procedure, using the construct bearing the brighter of the two detected isoforms as a template. The identity of the resulting clones was confirmed by sequencing.

Spectroscopy

The excitation and emission spectra of all the bacterial expression products were measured using LS-50B spectrofluorometer (Perkin Elmer; Waltham, MA), and corrected for the photomultiplier sensitivity. Quantum yield (QY) and molar extinction coefficient (ME), and relative brightness were evaluated as described earlier,[16] using EGFP (BioVision; Mountain View, CA) as a standard (ME = 55 000 M^{-1} cm^{-1}, QY = 0.6[46]). Briefly, absorption and fluorescence spectra were collected for a range of dilutions of native and denatured proteins. Data were plotted as regressions and the difference in the slopes between the new proteins and EGFP were used to calculate ME (maximum absorption *versus* concentration) and QY (total fluorescence *versus* absorption). Relative brightness is the product of ME and QY for each FP relative to EGFP (Table 1). For the full-length Ember protein, the ME was measured at the major absorption peak at 510 nm, and QY was determined at 470 nm excitation.

Oligomerization analysis

The oligomeric state of the products of bacterial expression was evaluated by HPLC size-exclusion chromatography with combined detection *via* multi-angle light scattering and refractometry, allowing for calculation of the molar mass of eluted proteins.[47] For chromatography, we used G4000PW_{XL} or G3000PW_{XL} columns (TosoHass, 300 × 7.8mm; Tosoh Bioscience LLC; King of Prussia, PA); the detectors were the EOS photometer and Optilab rEX differential refractometer (Wyatt Technology; Santa Barbara, CA). The data were analyzed using ASTRA software v. 4.9 and 5.3 (Wyatt Technology; Santa Barbara, CA).

Evaluating possible quaternary structure of the four-domain Ember protein

To determine whether the length of between-domain linkers in the jellyfish protein is compatible with the canonical tetrameric FP structure,[18,19] as well as to model possible alternative domain arrangements, I-TASSER software[48] was used. The alternative hypothetical domain arrangement in Ember was generated manually based on the structure of DsRed (1G7K): chains C and D were swapped and translated such that the joined C- and N-termini (the "intervening loop region") had space to thread along the interfaces while still maintaining a globular structure; chains A through D were connected consecutively. This model was then used as a template for threading using I-TASSER. To evaluate whether the model was plausible, coordinates of the individual Ember domains from the I-TASSER-optimized model were fed to the CombDock and ASSEMBLE algorithms, to see if any of the algorithm-generated models would be more stable than the proposed Ember model. CombDock generated a potential oligomer structure based on a heuristic search for shape complementarily[22] whereas ASSEMBLE[23] exhaustively searched through all possible oligomeric configurations. For the evaluation, the Ember molecule was split into domains such that the intervening loop region remained attached to the C-terminus of the preceding domain. The conformations of the intervening loop regions were held according to the I-TASSER optimized structural model. For relative scoring of the alternative oligomer configurations, PIE_{SVM} was used.[23] For more details on oligomerization modeling, see the ESI.†

Acknowledgements

We thank Dr Austen Riggs and Mrs Claire Riggs (University of Texas at Austin) for assistance with oligomerization analysis and insightful discussions, and Mr Ambrish Roy (University of Michigan) for advice on I-TASSER modeling. The Wyatt MALS instruments were purchased with funding from National Science Foundation Grant MCB-0237651 to Austen Riggs. This work was supported by the NOAA Office of Ocean Exploration Grant # NA07OAR46000289 ("Operation Deep Scope 2007"), and National Institutes of Health grants R01 GM078247 and R01 GM066243 to M.V.M.

References

1 O. Shimomura, F. H. Johnson and Y. Saiga, Extraction, purification and properties of aequorin, a bioluminescent protein from the luminous hydromedusan, Aequorea, *J. Cell. Comp. Physiol.*, 1962, **59**, 223–239.
2 D. M. Chudakov, M. V. Matz, S. Lukyanov and K. A. Lukyanov, Fluorescent Proteins and Their Applications in Imaging Living Cells and Tissues, *Physiol. Rev.*, 2010, **90**, 1103–1163.

3 R. Heim, D. C. Prasher and R. Y. Tsien, Wavelength mutations and post-translational autoxidation of green fluorescent protein, *Proc. Natl. Acad. Sci. U. S. A.*, 1994, **91**, 12501–12504.
4 M. Ormo, A. B. Cubitt, K. Kallio, L. A. Gross, R. Y. Tsien and S. J. Remington, Crystal structure of the Aequorea victoria green fluorescent protein, *Science*, 1996, **273**, 1392–1395.
5 D. C. Prasher, V. K. Eckenrode, W. W. Ward and F. G. Prendergast, Primary structure of the Aequorea victoria green-fluorescent, *Gene*, 1992, **111**, 229–233.
6 M. V. Matz, A. F. Fradkov, Y. A. Labas, A. P. Savitsky, A. G. Zaraisky, M. L. Markelov and S. A. Lukyanov, Fluorescent proteins from nonbioluminescent Anthozoa species, *Nat. Biotechnol.*, 1999, **17**, 969–973.
7 J. G. Morin and J. W. Hastings, Energy Transfer in a Bioluminescent System, *J. Cell. Physiol.*, 1971, **77**, 313–318.
8 W. W. Ward and M. J. Cormier, Energy-Transfer Via Protein-Protein Interaction in Renilla Bioluminescence, *Photochem. Photobiol.*, 1978, **27**, 389–396.
9 A. Terskikh, A. Fradkov, G. Ermakova, A. Zaraisky, P. Tan, A. V. Kajava, X. Zhao, S. Lukyanov, M. Matz, S. Kim, I. Weissman and P. Siebert, "Fluorescent timer": protein that changes color with time, *Science*, 2000, **290**, 1585–1588.
10 R. Ando, H. Hama, M. Yamamoto-Hino, H. Mizuno and A. Miyawaki, An optical marker based on the UV-induced green-to-red photoconversion of a fluorescent protein, *Proc. Natl. Acad. Sci. U. S. A.*, 2002, **99**, 12651–12656.
11 D. M. Chudakov, V. V. Belousov, A. G. Zaraisky, V. V. Novoselov, D. B. Staroverov, D. B. Zorov, S. Lukyanov and K. A. Lukyanov, Kindling fluorescent proteins for precise in vivo photolabeling, *Nat. Biotechnol.*, 2003, **21**, 452–452.
12 S. Habuchi, R. Ando, P. Dedecker, W. Verheijen, H. Mizuno, A. Miyawaki and J. Hofkens, Reversible single-molecule photoswitching in the GFP-like fluorescent protein Dronpa, *Proc. Natl. Acad. Sci. U. S. A.*, 2005, **102**, 9511–9516.
13 M. E. Bulina, D. M. Chudakov, O. V. Britanova, Y. G. Yanushevich, D. B. Staroverov, T. V. Chepurnykh, E. M. Merzlyak, M. A. Shkrob, S. Lukyanov and K. A. Lukyanov, A genetically encoded photosensitizer, *Nat. Biotechnol.*, 2005, **24**, 95–99.
14 D. A. Shagin, E. V. Barsova, Y. G. Yanushevich, A. F. Fradkov, K. A. Lukyanov, Y. A. Labas, T. N. Semenova, J. A. Ugalde, A. Meyers, J. M. Nunez, E. A. Widder, S. A. Lukyanov and M. V. Matz, GFP-like proteins as ubiquitous metazoan superfamily: Evolution of functional features and structural complexity, *Mol. Biol. Evol.*, 2004, **21**, 841–850.
15 A. G. Evdokimov, M. E. Pokross, N. S. Egorov, A. G. Zaraisky, I. V. Yampolsky, E. M. Merzlyak, A. N. Shkoporov, I. Sander, K. A. Lukyanov and D. M. Chudakov, Structural basis for the fast maturation of Arthropoda green fluorescent protein, *EMBO Rep.*, 2006, **7**, 1006–1012.
16 M. E. Hunt, M. P. Scherrer, F. D. Ferrari and M. V. Matz, Very bright green fluorescent proteins from the Pontellid copepod Pontella mimocerami, *PLoS One*, 2010, **5**, e11517.
17 D. D. Deheyn, K. Kubokawa, J. K. McCarthy, A. Murakami, M. Porrachia, G. W. Rouse and N. D. Holland, Endogenous green fluorescent protein (GFP) in amphioxus, *Biol. Bull.*, 2007, **213**, 95–100.
18 M. A. Wall, M. Socolich and R. Ranganathan, The structural basis for red fluorescence in the tetrameric GFP homolog DsRed, *Nat. Struct. Biol.*, 2000, **7**, 1133–1138.
19 D. Yarbrough, R. M. Wachter, K. Kallio, M. V. Matz and S. J. Remington, Refined crystal structure of DsRed, a red fluorescent protein from coral, at 2.0-A resolution, *Proc. Natl. Acad. Sci. U. S. A.*, 2001, **98**, 462–467.
20 I. Kelmanson and M. Matz, Molecular basis and evolutionary origins of color diversity in great star coral Montastraea cavernosa (Scleractinia: Faviida), *Mol. Biol. Evol.*, 2003, **20**, 1125–1133.
21 E. K. Bomati, G. Manning and D. D. Deheyn, Amphioxus encodes the largest known family of green fluorescent proteins, which have diversified into distinct functional classes, *BMC Evol. Biol.*, 2009, **9**, 77.
22 Y. Inbar, H. Benyamini, R. Nussinov and H. J. Wolfson, Combinatorial docking approach for structure prediction of large proteins and multi-molecular assemblies, *Phys. Biol.*, 2005, **2**, S156–165.
23 D. V. S. Ravikant and R. Elber, Energy designs for protein-protein interactions, *J. Chem. Phys.*, 2011, **135**, 065102.
24 L. A. Gross, G. S. Baird, R. C. Hoffman, K. K. Baldridge and R. Y. Tsien, The structure of the chromophore within DsRed, a red fluorescent protein from coral, *Proc. Natl. Acad. Sci. U. S. A.*, 2000, **97**, 11990–11995.
25 H. Mizuno, T. K. Mal, K. I. Tong, R. Ando, T. Furuta, M. Ikura and A. Miyawaki, Photo-induced peptide cleavage in the green-to-red conversion of a fluorescent protein, *Mol. Cell*, 2003, **12**, 1051–1058.
26 M. V. Matz, Y. A. Labas and J. Ugalde, Evolution of function and color in GFP-like proteins, *Methods Biochem. Anal.*, 2006, **47**, 139–161.
27 R. E. Campbell, O. Tour, A. E. Palmer, P. A. Steinbach, G. S. Baird, D. A. Zacharias and R. Y. Tsien, A monomeric red fluorescent protein, *Proc. Natl. Acad. Sci. U. S. A.*, 2002, **99**, 7877–7882.
28 S. F. Field and M. V. Matz, Retracing evolution of red fluorescence in GFP-like proteins from Faviina corals, *Mol. Biol. Evol.*, 2009, **27**, 225–233.
29 M. Prescott, M. Ling, T. Beddoe, A. J. Oakley, S. Dove, O. Hoegh-Guldberg, R. J. Devenish and J. Rossjohn, The 2.2 A crystal structure of a pocilloporin pigment reveals a nonplanar chromophore conformation, *Structure*, 2003, **11**, 275–284.
30 K. Nienhaus, G. U. Nienhaus, J. Wiedenmann and H. Nar, Structural basis for photo-induced protein cleavage and green-to-red conversion of fluorescent protein EosFP, *Proc. Natl. Acad. Sci. U. S. A.*, 2005, **102**, 9156–9159.
31 N. Pletneva, V. Pletnev, T. Tikhonova, A. A. Pakhomov, V. Popov, V. I. Martynov, A. Wlodawer, Z. Dauter and S. Pletnev, Refined crystal structures of red and green fluorescent proteins from the button polyp Zoanthus, *Acta Crystallogr., Sect. D: Biol. Crystallogr.*, 2007, **63**, 1082–1093.
32 S. Pletnev, N. G. Gurskaya, N. V. Pletneva, K. A. Lukyanov, D. M. Chudakov, V. I. Martynov, V. O. Popov, M. V. Kovalchuk, A. Wlodawer, Z. Dauter and V. Pletnev, Structural Basis for Phototoxicity of the Genetically Encoded Photosensitizer KillerRed, *J. Biol. Chem.*, 2009, **284**, 32028–32039.
33 J. Lukeš, J. M. Archibald, P. J. Keeling, W. F. Doolittle and M. W. Gray, How a neutral evolutionary ratchet can build cellular complexity, *IUBMB Life*, 2011, **63**, 528–537.
34 J. A. Ugalde, B. S. Chang and M. V. Matz, Evolution of coral pigments recreated, *Science*, 2004, **305**, 1433.
35 S. F. Field, M. Y. Bulina, I. V. Kelmanson, J. P. Bielawski and M. V. Matz, Adaptive evolution of multicolored fluorescent proteins in reef-building corals, *J. Mol. Evol.*, 2006, **62**, 332–339.
36 D. Boltovskoy, ed., *South Atlantic zooplankton*, Backhuys Publishers, Leiden, 1999.
37 E. Meyer, G. V. Aglyamova, S. Wang, J. Buchanan-Carter, D. Abrego, J. K. Colbourne, B. L. Willis and M. V. Matz, Sequencing and de novo analysis of a coral larval transcriptome using 454 GSFlx, *BMC Genomics*, 2009, **10**, 219.
38 S. F. Altschul, T. L. Madden, A. A. Schaffer, J. H. Zhang, Z. Zhang, W. Miller and D. J. Lipman, Gapped BLAST and PSI-BLAST: a new generation of protein database search programs, *Nucleic Acids Res.*, 1997, **25**, 3389–3402.
39 M. Matz, D. Shagin, E. Bogdanova, O. Britanova, S. Lukyanov, L. Diatchenko and A. Chenchik, Amplification of cDNA ends based on template-switching effect and step- out PCR, *Nucleic Acids Res.*, 1999, **27**, 1558–1560.
40 M. V. Matz, N. O. Alieva, A. Chenchik and S. A. Lukyanov, in *Generation of cDNA libraries: Methods and protocols*, ed. S.-H. Ying, Humana Press Inc., Totowa, NJ, 2003.
41 N. O. Alieva, K. A. Konzen, S. F. Field, E. A. Meleshkevitch, M. E. Hunt, V. Beltran-Ramirez, D. J. Miller, J. Wiedenmann, A. Salih and M. V. Matz, Diversity and Evolution of Coral Fluorescent Proteins, *PLoS One*, 2008, **3**, e2680.
42 D. G. Higgins, J. D. Thompson and T. J. Gibson, Using CLUSTAL for multiple sequence alignments, *Methods Enzymol.*, 1996, **266**, 383–402.
43 K. B. Nicholas, H. B. J. Nicholas and D. W. I. Deerfield, GeneDoc: Analysis and Visualization of Genetic Variation, *EMBNEW.NEWS*, 1997, **4**, 14.
44 S. H. Haddock, N. Mastroianni and L. M. Christianson, A photoactivatable green-fluorescent protein from the phylum Ctenophora, *Proc. R. Soc. London, Ser. B*, 2009, **277**, 1155–1160.
45 J. P. Huelsenbeck and F. Ronquist, MRBAYES: Bayesian inference of phylogenetic trees, *Bioinformatics*, 2001, **17**, 754–755.
46 R. Heim, A. B. Cubitt and R. Y. Tsien, Improved green fluorescence, *Nature*, 1995, **373**, 663–664.
47 P. J. Wyatt, Light scattering and the absolute characterization of macromolecules, *Anal. Chim. Acta*, 1993, **272**, 1–40.
48 Y. Zhang, I-TASSER server for protein 3D structure prediction, *BMC Bioinformatics*, 2008, **9**, 40.

