Positive functional synergy of structurally integrated, designed artificial protein dimers assembled by fully genetically encoded Click chemistry

Harley L Worthy[1]‡, Husam Sabah Auhim[1,2]‡, W. David Jamieson[3]‡, Jacob R. Pope[1], Aaron Wall[1,4], Daniel Watkins[1], Rachel Johnson[1], Pierre Rizkallah[4], Oliver Castell[3], D. Dafydd Jones[1]*.

‡ Joint first authors

1. Molecular Biosciences, School of Biosciences, Cardiff University, Cardiff, UK.
2. Department of Biology, College of Science, Baghdad University, Baghdad, Iraq.
3. School of Pharmacy and Pharmaceutical Sciences, Cardiff University, Cardiff, UK.
4. School of Medicine, Cardiff University, Cardiff, UK.

* Corresponding author. D. Dafydd Jones, Molecular Biosciences Division, School of Biosciences, Cardiff University, Cardiff, UK. CF10 3AT. Email: jonesdd@cardiff.ac.uk. Tel: +44 29 20874290.

Abstract

Constructing artificial higher order protein complexes is currently of great interest as they sample structural architectures and functional features not accessible by classical monomeric proteins. A desirable yet hard to achieve feature is synergy between subunits, whereby function of the assembly is greater than the sum of its parts. We combined *in silico* modelling with fully genetically encoded strain promoted azide-alkyne cycloaddition, to construct bespoke protein dimers. Using fluorescent proteins GFP and Venus as models, homo and heterodimers were constructed that switched ON once assembled and displayed enhanced spectral properties. The determined molecular structure reveals long range polar bond networks involving amino acids and structured water molecules play a key role in activation and functional enhancement by directly linking the two functional centres. Single molecule analysis revealed the dimer is more resistant to photobleaching spending longer times in the ON state with only one CRO likely to be active at any one time. Thus, genetically encoded bioorthogonal chemistry can be used beyond simple passive linkage approaches to generate new and truly integrated protein complexes that form long range bonds networks, which have a profound effect on function and our understanding of fluorescent protein function.

Introduction

The importance of protein oligomerisation to biology is illustrated clearly in nature [1, 2], with dimers being the most commonly observed final structural state of proteins [3, 4]. Oligomerisation is thus now emerging as an alternative route to engineer proteins and peptides to construct higher order complexes with new and useful properties from a limited repertoire of monomeric building blocks [5-7]. There are challenges associated with constructing protein oligomers as the subunit interfaces normally comprise numerous weak non-covalent interactions. [1, 8] Great strides have been made in generating assembled peptide and protein oligomeric systems using approaches such as helix-helix interactions [9, 10], metal coordination [11-13], fusion domains [6], disulphide bridging [14] and remodelled naturally inspired protein-protein interfaces [15-17]. However, one key aspect is generally missing that is common in natural protein oligomer systems: there is little or no functional synergy between the individual components, so complexes manifest the properties of the starting components. This is because long range bond networks beyond the local direct interactions at the interface region (which drive the initial assembly event) need to be considered for connecting functional centres, which is a fundamentally more challenging proposition. For example, an impressive range of GFP oligomers have been constructed through disulphide or metal-mediated approaches but functional communication was not apparent [14].

Here we describe the construction of fluorescent protein dimers assembled by designed covalent linking using genetically encoded strain promoted azide-alkyne cycloaddition chemistry [18](Fig 1a). The benefit of such an approach is that the design can be a simpler, defined and predictable in terms of linkage position and be easily combined with approaches noted above if required. The different but mutually compatible reaction handles can be placed at different sites in each monomer protein to generate a single covalent crosslink position that acts as a stable molecular “bolt”. NcAAs such as tyrosine or lysine derivatives (Figure 1a for example) are ideal for such an approach. Their relatively long side chains will reduce steric clashes while maintaining structural intimacy to promote and stabilise favourable non-covalent interactions required for dimerisation and inter-monomer communication. They are also generally form less labile bonds compared to, for example, disulphide bridges. While previous work has generated proteins linked by bioorthogonal crosslinks, normally via extended linker molecules [19-22], the proteins

are structurally and functionally distinct so provide little improvement on classical genetic fusions. Strain promoted azide-alkyne cycloaddition (SPAAC) [23] is a biocompatible 1-to-1 reaction that does not require any toxic catalysts and allows both regioisomers (Figure 1a) around the linking triazole to be sampled. It can also be fully genetically encoded using two separate ncAAs (azF and SCO in Figure 1a) so bypassing the requirement of linking molecules thus enabling an intimate interaction between the monomers.

Green fluorescent protein (GFP) [24] together with its yellow relative Venus [25] were chosen as the target proteins, as GFP in particular is proving to be an excellent model for investigating and understanding the molecular influence of ncAA incorporation on protein function (see [26-29] for examples), including BioClick reactions [30] and biohybrid assemblies [31-33]. Fluorescent proteins in general are also proving an excellent system to study important chemical and biological process, such as charge transfer networks, long range proton wires and coupled photochemistry [34, 35]. We show that dimer interface regions can be identified and designed through the use of *in silico* docking approaches, and that dimerisation switched ON and enhances function. Our experimentally determined structure highlights the formation of a new long-range interaction network between the protein's functional centres. Heterodimers were also constructed, which showed apparent integrated function combining facets of each individual monomer so the dimer acts as one single functional unit.

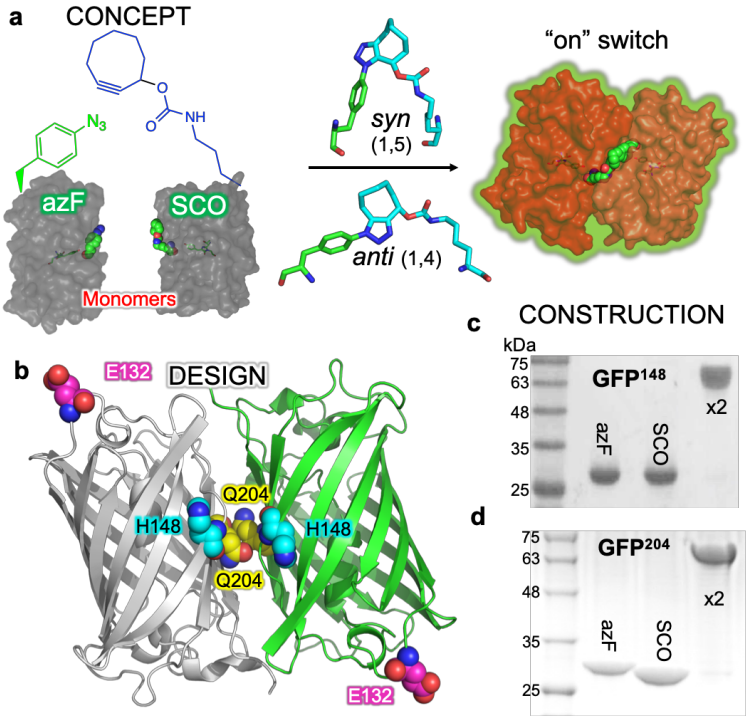


Figure 1. Bioorthogonal-driven protein dimersation and implementation phases. (a) Concept. SPAAC reaction between the two genetically encoded ncAAs (azF, green and SCO, blue) intimately links two monomeric proteins via either a *syn* or *anti* triazole link to promote the formation additional non-covalent interactions. Shown are the 148 variants. (b) Design. *In silico* modelling to predict potential dimer interfaces and residues contributing to the interface. Shown is the highest ranked GFP-GFP dimer model (see Table S1 for statistics), with residues E132 (magenta), H148 (cyan) and Q204 (yellow) selected for replacement with either azF or SCO shown. (c and d) Construction. Dimerisation as analysed by gel mobility shift; (c) formation of dimeric GFP^{148x2} from monomers GFP^{148azF} and GFP^{148SCO}; (d) formation of dimeric GFP^{204x2} from monomers GFP^{204azF} and GFP^{204SCO}.

Results and discussion

Design of Click chemistry interface sites. We surmised that areas of a protein's surface that are compatible in terms of association are more likely to generate an integrated structure through the formation of mutually compatible non-covalent interactions. Also, as the proteins are naturally monomeric these interactions are at best, weak and transient so do not persist, we reasoned that we needed a molecular "bolt" as part of the interface site to promote and stabilise any new interactions. The first step is to identify regions on the target proteins that can interact. Initially ClustPro 2.0 [36] (cluspro.org) was used to generate potential dimer configurations. The output GFP homodimer models were refined, analysed and ranked using

RosettaDock [37, 38] (see Table S1). The highest ranked configuration is shown in Figure 1a (which is the closest model to the determined structure below; vide infra), with the next 4 ranked configurations shown in Figure S1. While different orientations of one GFP to the other were observed, docking revealed residues 145-148, 202-207 and 221-224 were routinely found to contribute to the dimer interface. To bolt the two proteins together, genetically encoded bioorthgonal Click chemistry was used (Figure 1a). The benefit of Click chemistry compared to for example disulphide linkages, is longer side chains to overcome potential steric clashes, improved stability of the linkage and the ability to generate heterodimers in a designed 1-to-1 manner (vide infra). The additional benefit of the *in silico* approach will allow identification of areas compatible for general Click chemistry-based direct protein-protein conjugation.

Based on the dimer models, 3 residues were selected for replacement with the two Click compatible ncAAs, SCO [39] (strained alkyne) and azF (azide) [40, 41] (Figure 1a). H148 and Q204 were chosen based on their location at the putative dimer interface (Figure 1b and Figure S1). Both 148 and 204 residues are also known to be readily modified with small molecule cyclooctyne adducts on azF incorporation [30, 42]. As shown in Figure 1c-d, gel mobility shift analysis revealed that dimerisation was successful; this was confirmed by mass spectrometry analysis (Figure S2). Residue 132 was not predicted to be at the dimer interface (Figure 1b and S1) but is known to be compatible with a range of strained alkyne adducts ranging from dyes [42] to carbon nanotubes [31] to single stranded DNA [32]. Thus, it acts as a good test of our ability to predict protein-protein interfaces and the Click reaction compatibility. No dimer product was observed using GFP^{132azF} with SCO containing protein (Figure S3) indicating the importance of surface interface compatibility and how the design process can be used to predict compatible sites.

Positive functional switching on forming GFP^{148x2} dimer. Regulated activity is a feature common to natural protein oligomers [2], including control of protein function through changes in interaction networks. Here we show that we can switch ON GFP by modulating the charged state of the chromophore (CRO) and enhance function through dimerisation. H148 forms a H-bond with CRO and plays an important role in proton shuttling that regulates the population of the preferred unfavoured neutral A (λ_{max} ~400 nm) and anion B state (λ_{max} ~490 nm) [43] that is fluorescently active from

of CRO; incorporating azF in place of H148 (GFP^{148azF}) removes the H-bond resulting in a functionally compromised A state predominating [29, 30] (Figure 2a and Table S2); there is also a red shift in the minor B form (~15 nm from wt GFP; λ_{max} = 500 nm; Table S2). Incorporating SCO at residue 148 (GFP^{148SCO}) elicits a similar effect, with the A form of CRO predominating but with a smaller red shift (λ_{max} 492 nm) in the minor B form (Figure 2a and Table S2). SCO incorporation at residue 148 also leads to a significant decrease in brightness at both excitation wavelengths (Table S2).

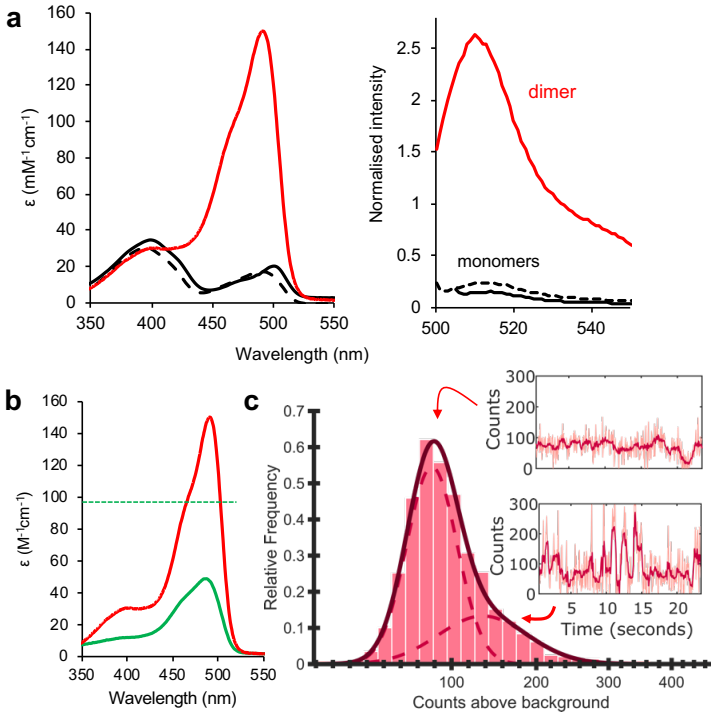


Figure 2. Spectral properties of GFP 148 variants before and after dimerisation. (a) Absorbance and fluorescence emission (right; on excitation at 492 nm) of GFP^{148x2} (red), GFP^{148SCO} (black dashed) and GFP^{148azF} (black). Fluorescence emission was normalised to wt GFP. (b) Comparison of GFP^{WT} absorbance spectra (green) with GFP^{148x2} (red). The green dashed line represents the expected value if ε at λ_{max} is simply doubled for GFP^{WT}. (c) Single molecule fluorescence intensity histogram for GFP^{148x2} dimers (115 trajectories comprised of 1742 spots), with two representative fluorescence time course traces of individual dimers inset (both with raw and Cheung-Kennedy filtered data). The histogram of observed GFP^{148x2} fluorescence intensities is described by a two-component mixed log normal distribution. Representative fluorescent time course traces illustrate typically observed fluorescent behavior of the dimer. With prolonged fluorescence observed at ~80-100 counts corresponding to the first component in the histogram. Some dimers exhibit rapid and brief forays to higher intensity states, giving rise to the second higher intensity peak in the histogram. Additional traces can be found in Supporting Figure S6.

Dimerisation of GFP^{148azF} and GFP^{148SCO} produces two significant positive effects: (i) switches ON fluorescence at ~490 nm; (ii) greatly enhanced brightness through increased molar absorbance coefficient at 490 nm (Figure 2a and Table S2). The major excitation peak is red shifted on dimerisation (λ_{max} 492 nm) compared to wt GFP (λ_{max} 485 nm) (Table S2). The 490:400 nm absorbance ratio shifts by an order of magnitude from ~0.5 for the monomers to ~5 for the dimer, with the CRO B form now dominating in the dimer absorbance spectrum (Figure 2a). Previous examples of modifying GFP^{azF148} with small molecule adducts or light at best result in partial conversion to CRO B form [29, 30]. The 10 fold switch in absorbance is also mirrored in fluorescence emission; excitation at 490 nm results ~20 fold higher emission than either monomer (Figure 2a). Additionally, the dimer shows overall enhanced function even when compared to the original superfolder GFP^{WT} (Figure 2b and Table S2). Molar absorbance and brightness increased ~320% for GFP^{148x2} (~160% on a per CRO basis) (Figure S5), higher than expected for a simple additive increase if monomer units are acting independently of each other.

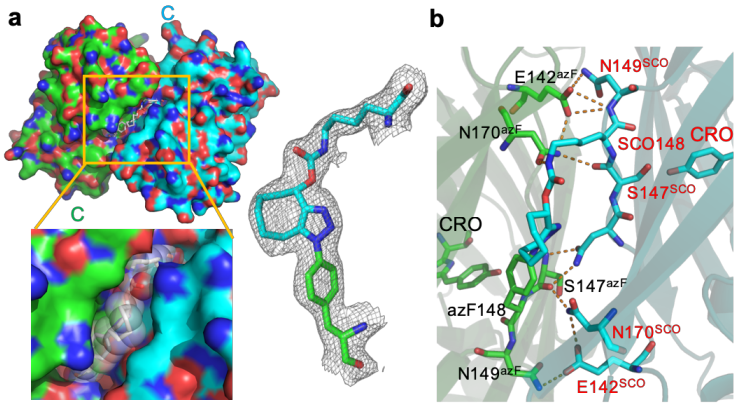


Figure 3. Structure of GFP^{148x2} (PDB 5nhn). The azF bearing protein is coloured green and SCO bearing protein is coloured cyan. (a) Overall monomer arrangement and interface, with the contribution of the azF-SCO crosslink shown in the magnified panel. The electron density for the crosslink is shown to the right. (b) H-bond network contributing to the dimer interface.

Molecular basis for functional switching in GFP^{148x2}. The crystal structure of GFP^{148x2} (see Table S3 for statistics) reveals that the monomers forms an extensive dimer interface with long range interactions linking the two CRO centres. The

monomer units of GFP^{148x2} arrange in a quasi-symmetrical head-to-tail arrangement with each monomer offset by ~45° to each other (Figure 3a and Figure S4a-b). The anti-parallel side-by-side arrangement of the two monomers is closest to that of the highest ranked model (Figure S1 and Table S1). The crosslink forms the elongated *anti* 1,4 triazole link that is partially buried and intimately associated with both monomer units so forming an integral part of the dimer interface (Figure 3a and S4). The CROs are 15 Å apart pointing towards each other (Figure 3b and S4b).

The interface is relatively extensive and has similar characteristics to natural dimers [44]. Interface buried area is ~1300 Å^2, with generally the same residues from each monomer contributing (Figures 3a-b and S4c-d). H-bonding plays an important role with residues E142, N146, S147, N149 and N170 from both monomers contributing to 8 inter-subunit H-bonds (Figure 3b). The structure shows that natural dimer interfaces can be mimicked and stabilised through the use of Click-linked monomers, which the original designs outlined in Figure 1c and S1 suggested were feasible but where probably too weak or transient to persist without the embedded link. Thus, it may be feasible that our approach could be used to stabilise more broadly transient weak protein-protein interactions so forming defined interfaces.

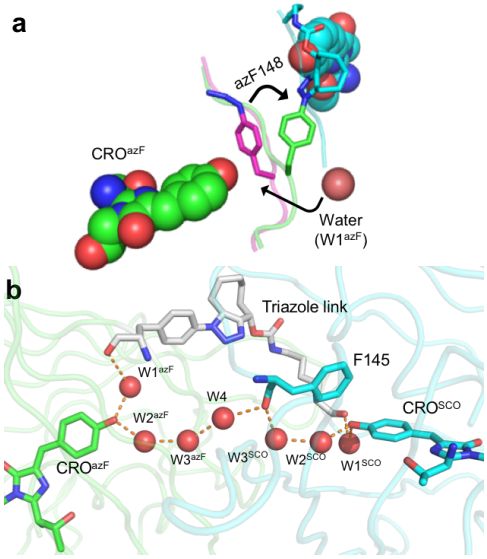


Figure 4. Activation via conformational changes and inter-subunit communication networks on formation of GFP^{148x2}. The azF bearing protein is coloured green and SCO bearing protein is coloured cyan. (a) Conformational change to azF148 on dimerisation. The GFP^{148azF} (PDB 5bt0 [30]) is coloured in magenta. (b) Water dominated H-bond and potential proton transfer wire linking the CRO from the azF (CRO^{azF}) and SCO (CRO^{SCO}) monomers. The solvent tunnel determined by CAVER analysis [45] is shown in Figure S5.

Dimerisation induces a series of conformational changes to form a long-range interaction networks that underlies the mechanism by which GFP is switched ON. The position of residue 148azF changes compared to the monomer resulting in a hole that is occupied by a water in the dimer ($W1^{azF}$ in Figure 4a); the new water can H-bond to CRO^{azF} and the backbone of 148azF (Figure 4); an equivalent water is present in the GFP^{148SCO} monomer unit, ($W1^{SCO}$) which forms similar interactions. These structured water molecules replace the lost H-bond on removal of H148, and form part of an extended predominantly water wire network that spans the dimer interface linking the two CROs (Figures 4b and S5). The three water molecules in each unit ($W1^{azF/SCO}$, $W2^{azF/SCO}$ and $W3^{azF/SCO}$) are symmetrical; W4 combined with the backbone of $F145^{SCO}$ provide the bridge across the dimer interface to link the two water networks together. Thus, dimerisation generates an extended, inter-monomer water-rich wire that can contribute to H-bond and proton shuttling network so promoting a switch from the A CRO to the B form.

Single molecule analysis reveals a potential "ping-pong" fluorescence mechanism. Total internal reflection fluorescence (TIRF) microscopy was used to investigate the fluorescent behaviour of GFP^{148x2} dimers at the single molecule level. The fluorescence intensity time-course of single GFP^{148x2} dimers demonstrated a range of intensity states, with fluorescence at ~80-100 counts predominating and displaying greater longevity than the sub-population of higher intensity states, with these characterised by brief forays to a range of intensities from ~100-~300 counts (Figure 2c). The fluorescence traces also demonstrate prolonged photostability with long periods to photobleaching (Figure 2c and S6). In comparison, GFP^{WT} photobleaches more rapidly, with fluorescence traces showing a single intensity state in which the on states generally last for shorter periods (Figure S7). Furthermore, monomeric GFP^{WT} was sometimes found to exist in an initial dark, non-fluorescent state, prior to initiation of fluorescence and subsequent photobleaching (Figure S7). Extraction of the average consecutive fluorescence 'on time' prior to photobleaching or occupancy of transient non-fluorescent states (blinking) finds that GFP^{148x2} displays longer periods of continuous fluorescence (mean 0.9 s), compared to GFP^{WT} (0.65 s). Given the similarity in measured single molecule fluorescence

intensity, the increased ON times and photobleaching lifetime likely account for the increased fluorescence observed in steady state ensemble measurements of GFP^{148x2} (Figure 2a).

In an attempt to rationalise the range of fluorescence states observed in the dimer traces, a histogram of all measured intensities was generated (Figure 2C). Unlike GFP^{WT} (Figure S7) the resulting distribution was found to be poorly described by a single log-normal distribution [46]. Using mixed log-normal distribution models [47], the calculated AIC favoured a two-component fit (Figure 2C). The measured intensity distribution shows a predominant lower intensity peak (~90 counts) and a partially overlapping higher intensity peak, as a consequence of the brief forays to higher intensity states observed in the single molecule fluorescence traces. Whilst a bimodal intensity distribution might ordinarily be expected in a dimer comprised of two co-located independently active fluorophores, with each fluorophore sequentially photobleaching, the single molecule intensity time-course traces are not consistent with this model and show a lack two well defined states. The simple on/off state behaviour of GFP^{WT} (Figure S7) is infrequently observed in the dimer traces which themselves do not present as the anticipated adduct of two monomeric traces, instead showing more complex behaviour.

A potential explanation for the behaviour observed in the GFP^{148x2} single molecule intensity traces may be found by considering the effect of protonation states on the GFP^{148} monomers and the inter-CRO bond network of the dimer as revealed by the crystal structure. Both GFP^{148} monomers predominantly occupy its protonated A form in the ground state with an absorption maxima at ~400nm. This would be expected to give rise to no or extremely limited fluorescence under the 473nm TIRF illumination. However, the buried water network connecting the CROs in the dimer interior provides a putative proton wire to shuttle a proton from the CRO in the ground state [48]. We anticipate that rapid proton shuttling between the two CROs in a synchronised "ping-pong" mechanism of action manifests [49] in a situation where the majority of the time one dimer CRO exists in its minimally fluorescent protonated (A) state while the other exists in the fluorescent deprotonated (B) state. This is supported by the steady state ensemble absorbance data in which a shoulder around 400nm is indicative of the coexistence of both the predominant B and minor A forms of the CRO in the dimer population, a feature absent from GFP^{WT} absorbance data (Figure 2b). Such a scenario, with a single CRO of the dimer active

at any one time, would give rise to the dominant lower intensity peak as measured in the GFP^{148x2} single molecule intensity histogram arising from the lower persistent intensity state seen in the single molecule traces (Figures 2C & S6a, g, i, m & o), and be consistent with the absence of classical two-step sequential photobleaching. It is notable that comparable single molecule intensities (~100 counts) are observed in the GFP^{WT} where a single CRO is present and predominantly exists in its fluorescent B form. In the dimer, the less frequented higher intensity state which is transiently visited during the fluorescent time-course may be indicative of a more rarely encountered simultaneous activity of both CROs.

Long-range polar networks and role of water. The structure of GFP^{148x2} highlights the structural role of internal water molecules and provides a rationale for the observed synergy: the formation of long-range interaction networks that linked the two functional centres (Figure 3). Moreover, the network was quasi-symmetrical with water molecules contributing to a proton shuttling system. Formation of the triazole link is critical as it results in a local conformation change allowing entry of a water molecule to replace the functional role of the original histidine (Figure 4a). Long-range polar networks and charge transfer processes play an important role in chemistry and biology, including autofluorescent proteins [50, 51]. Yet, it represents one of the most challenging to construct through protein engineering due to difficultly including water as part of the design process; the GPF^{148x2} system represents an excellent model in which to study the interplay of water and functionally interlinked proteins in the process and how it can be used to link functional centres.

We [30, 52] and others [35, 50, 53, 54] have proposed that water networks and dynamics play an important role in GFP fluorescence. In GFP and other monomeric *A. victoria* FPs, water molecules beyond the directly bonded CRO water molecule (termed W1; see Figure 4) that contribute to an extended network are largely exposed to the solvent (Figure S8) in the momoners and thus subject to dynamic exchange with the bulk solvent. In the dimer, these water molecules now lie at the monomer interface forming a buried putative H-bond network (Figure 4). The simple effect of surface burial and reduced dynamics may be an important contributor to the improved functional affects we see here and can potentially be applied in other systems. Additionally, solvent burial and reduced dynamics is likely to be important to permitting the formation of a more permanent, organised and concerted proton wire

network, as observed for GFP^{148x2} resulting in changes the inherent fluorescence properties.

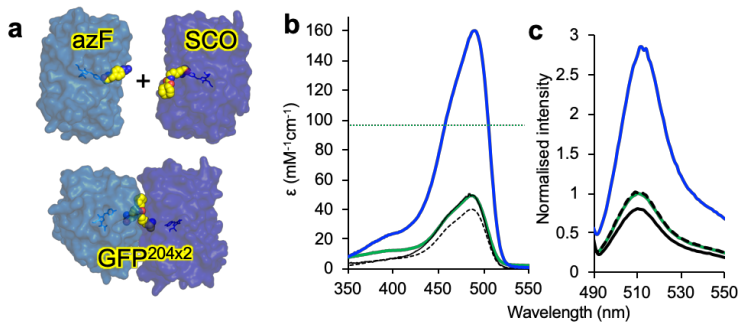


Figure 5. Spectral properties of GFP 204 variants before and after dimerisation. (a) schematic of dimerisation of GFP^{204azF} and GFP^{204SCO} to form GFP^{204x2} (b) Absorbance and (c) fluorescence (on excitation at 487 nm) of GFP^{204x2} (blue), GFP^{204SCO} (black dashed) and GFP^{204azF} (black) and wt GFP (green). Fluorescence emission was normalised to wt GFP. The red dashed line represent the molar absorbance value for a simple addition of two individual GFP^{WT} at λ_{max}.

Enhanced function on forming GFP^{204x2} dimer. To explore how different linkage sites can elicit functional affects, we investigated the alternative dimer GFP^{204x2} (Figure 5a) constructed above (Figure 1d). We found that dimerisation enhanced the spectral properties above that of simple addition of the monomeric or GFP^{WT} proteins highlighting again the synergistic benefits of dimerisation. Incorporation of either azF or SCO at residue 204 had little effect on the spectral properties compared to GFP^{WT} [42] (Figure 5b and Table S2). The B CRO form predominated in the monomeric forms; both molar absorbance and emission intensities where similar to each other and GFP^{WT}. The fluorescence emission of GFP^{204SCO} was slightly reduced (80% of GFP^{WT}). On forming the GFP^{204x2} dimer (Figure 1d and Figure S2 for mass analysis), spectral analysis showed functional enhancement in terms of the core spectral parameters: molar absorbance coefficient (ε) and fluorescence emission (Figure 5b and Table S2). On dimerization, ε increased up to 400% compared to the starting monomers to 160,000 $M^{-1}cm^{-1}$. This equates to an average per chromophore ε of 80,000 $M^{-1}cm^{-1}$, almost doubling the brightness compared to the starting monomers, and 31,000 $M^{-1}cm^{-1}$ higher compared to GFP^{WT}. In line with the increased capacity to absorb light, fluorescence emission was also enhanced; the normalised per chromophore emission was 1.8 fold higher than the GFP^{204azF} monomer. Thus, as with GFP^{148x2} the dimeric structure

of GFP^{204x2} has an increased probability of electronic excitation and fluorescence output compared to monomeric forms (see Figure S9 for spectral comparison of dimers). This is all the more impressive for both dimeric forms as our starting superfolder GFP^{WT} is a benchmark for green fluorescent protein performance.

Heterodimers and functional integration. Heterodimers, in which a dimer is composed of two different proteins, is a commonly observed alternative dimerisation state [1, 2]. It also allows us to design new complexes in which functionally distinct proteins can be linked. An advantage of bioorthogonal coupling is the ability to generate defined, single species (hetero)dimers comprised of two different protein units (i.e. A + B = A-B not A-A, B-B, A-B mixture which can be difficult to separate). The yellow fluorescent protein Venus [25] was chosen as the partner protein, given the spectral overlap between the two (Figure S10). Sequence differences are shown in Figure S11.

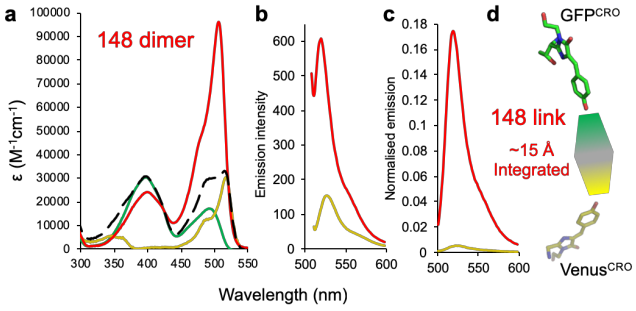


Figure 6. Communication between heterodimers. (a) Absorbance spectra of GFVen^{148}. Red, gold, green, black dashed lines represent GFVen^{148}, Venus^{148azF}, GFP^{148SCO} and monomer addition spectrum, respectively. (b) Emission intensity of 0.5 μM GFVen^{148} (red) and Venus^{148azF} (gold) on excitation at 505 nm. (c) Normalised emission of GFVen^{148} (red) and Venus^{148azF} (gold) on excitation at 400 nm. (d) Spatial arrangement of GFP and Venus CRO based on the GFP^{148x2} structure.

GFP^{148SCO} was combined with the equivalent azF containing Venus (Venus^{148azF}) to generate GFVen^{148} (see Figure S12 for evidence). When Venus combines with GFP through the 148 SPAAC linkage, new spectral characteristics emerge suggesting that an integrated system has been generated. Both GFP^{148SCO} and Venus^{148azF} are functionally inferior in terms of their spectral properties compared to the dimer, GFVen^{148x2} (Table S2). Interestingly, the dimer has spectral properties intermediate of individual monomers but without any significant peak broadening

(Figure 6a-b and Figure S13a) suggesting that the two CRO centres have become functionally integrated in terms of fluorescence emission (Figure 6b). The major λ_{max} is 505 nm, intermediate between GFP (492 nm) and Venus (517 nm). The ε equivalent to the B CRO form (490-510 nm region) increases significantly (~4-5 fold), higher than the simple additive spectra of the monomers, while the A CRO population decreases but is still observed (Figure 6a). This is matched by a ~4 fold increase in emission intensity on excitation at 505 nm (Figure 4b). A single emission peak is observed that is also intermediate between the two monomers, irrespective of the excitation wavelength (λ_{EM} at 517 nm; Figure 4b-c); a single peak rather than a double or broadened peak was observed on excitation at 490 nm (capable of exciting both CROs) suggesting that a single species is emitting. An additive spectrum of individual monomer spectra that simulates two independently acting proteins supports the idea of a new integrated function as it is broader and red-shifted compared to the measured GFVen^{148x2} emission profile (Figure S13b). Thus, the GFVen^{148} appears to be acting as a single entity in terms of fluorescence output despite the differences in the starting properties. Emission on excitation at 400 nm was also measured as Venus^{148azF} has little absorbance at this wavelength compared to GFVen^{148}. Emission intensity was 30 fold higher for GFVen^{148} compared to monomeric Venus^{148azF} with emission still peaking at 517 nm (Figure 6c and S13c).

By linking together Venus and GFP via residue 148, new properties emerged that integrated the spectral features (in terms of λ_{max} and λ_{EM}) of the original monomeric proteins. Rather than displaying classical FRET, as would be expected for an assembly of two fluorescent proteins, a single emission intermediate between the two original proteins was observed (Figure 4a-c). This could suggest that two CROs are now acting predominantly as one species in GFVen^{148} with the components observed for GFP^{148x2} (such as the water network) playing a role. The presence of significant neutral A state of the CRO (Figure 4a) does suggest two monomeric units are not fully synchronised, mostly likely in the GFP monomer unit so the new intermediate spectral properties may be the result structural changes in Venus^{148azF} on dimerisation. However, it is clear that dimerisation had a significant positive impact with regards to GFVen^{148}, with structural mechanisms similar to the enhancements observed for GFP^{148x2} likely playing a key role.

Conclusion

We have successfully demonstrated that dimers can be constructed from normally monomeric units using genetically encoded bioorthogonal chemistry through predicting protein interface regions. Importantly, we go beyond simple passive linking of individual proteins by constructing truly structurally integrated complexes that enhance inherent function. The generation of new protein oligomerisation systems is currently a hot topic in protein engineering [7, 16] as it allows us to understand a commonly observed molecular feature in biology, explores new functional and structural space, and expand uses in the nanosciences. Here we have addressed one of the key challenges in engineered protein oligomerisation systems: interunit communication and networking beyond the direct interface region. In this work, homodimers displayed enhanced function, including switching ON through assembly. The latter aspect could have applications for the generation of proximity-based biosensors that does not rely on the complexities concerning FRET. Our structure reveals that an extensive dimer interface is formed through mutually compatible interactions, mimicking natural dimer interfaces without having to extensively engineer such weak interactions; the new Click link enables otherwise weak and/or transient interactions to persist, likely through an entropic mechanism. Indeed, the ability to predict potential interaction interfaces and stabilise them through covalent linkage may provide a general strategy of constructing structurally interacting protein dimers and higher order oligomers, so moving beyond simple passive linkage. The approach is not restricted to the ncAA used here, with different strained alkyne regioisomers and even linking chemistries available [18] so allowing a broader sampling of dimer conformations. Nor are the sites or targets of coupling restricted, with alternative non-symmetrical architectures and linkage of disparate proteins feasible. With developments in codon reprogramming for incorporation of multiple bioorthogonal chemistries into a single protein at defined positions [55] and codon replacement cell lines [56, 57], coupled with integration with more classical protein engineering approaches, our approach can aid the design of higher order functional oligomeric species for use in both synthetic biology and nanoscience.

Experimental.

Detailed methods concerning protein production, protein structure determination and mass spectrometry are outlined in the Supporting Information.

In silico modelling of GFP dimer interfaces. ClustPro is a global docking rigid-body approach that requires no prior information on interface regions, and has been shown to be a good predictor of dimer interfaces [36]. ClusPro was used in multimer mode (set to dimers) using the structure of wild type sfGFP (2B3P) as a starting model. ClusPro generates ~100,000 structures and scores them using balanced energy coefficients as described by Kozakov et al [36](Eq 1). E is the energy score of the complex; E_{rep} is the energy of the repulsive contribution of van der Walls interactions and E_{att} is the attractive interaction equivalent. E_{elec} is a term generated by electrostatic energy and E_{DARS} is a term that mainly accounts for free energy change due to exclusion of water from the interface.

Eq 1: $E = 0.40E_{rep} + -0.40E_{att} + 600E_{elec} + 1.00E_{DARS}$

The server then takes the 1000 models with the lowest scores and clusters them using pairwise to generate I-RMSD (interface root mean squared deviation). Doing so creates clusters centred on the structure with the most neighbours within a 9 Å radius. Of the remaining models that do not fall within the first cluster the one with the most neighbours becomes the centre of the next cluster and so on until all models are part of a cluster. The centre models of each cluster are energy minimised using the CHARMM force-field for 300 steps with fixed backbone to minimise steric clashes.
A model for each cluster was downloaded and run through ROSETTA's high resolution docking protocol. This added extra rotamers and subsequent minimisation of side chains. The docking protocol also rescores the models and adds an interface score [37, 38]. The interface score is the total complex score minus the sum of the separate monomer energies and is used as a metric of how good a model is. Total score and interface score were plotted against I-RMSD to highlight any outliers and remove them. The top 5 models were then used as a basis for determining which residues would be most suitable for crosslinking to form dimers.
Protein dimerisation by SPAAC. Concentrations of monomer variants were determined using the Bio-RAD DC Protein Assay using wild type wt GFP as a standard and correlated to the 280 nm absorbance. Dimers were generated by mixing azF and SCO monomers (100 μM, 50 mM Tris-HCl) overnight at room temperature. Dimers were purified by size exclusion chromatography and concentrations determined, as described above. The spectral properties of the dimers were characterised as described below. Fluorescence spectra were taken at a concentration of 0.25 μM (equivalent chromophore number to 0.5 μM of monomers). Protein dimerisation was also monitored by SDS PAGE gel mobility assays. Dimer yields were between 35-80%.
Protein spectral analysis. Proteins were diluted to 5 μM and 0.5 μM in 50 mM Tris-HCl pH 8.0 for absorbance and fluorescence spectra, respectively. Absorbance spectra were taken on a Cary Win UV, using a 300 nm/min scan rate at 1 nm intervals. Absorbance at λmax for each variant, was used to determine the molar extinction coefficients (ε) for each variant, using the Beer-Lambert equation. Emission spectra were collected on a Cary Varian fluorimeter at a scan rate of 150 nm/min and either 0.5 or 1 nm intervals. Emission and excitation slit widths were set to 5 nm and a detector voltage of 600 mV. Samples were excited at various wavelengths as stated in the main text and emission was scanned from the excitation wavelength to 650 nm. Quantum yields were calculated as previously described using a fluorescein standard dissolved in 0.1M NaOH [29, 30] In brief

samples were diluted to an optical density of 0.05 at the chosen excitation wavelength, in 50 mM Tris-HCl pH 8.0. An emission spectrum was then taken as above with a reduced slit width of 2.5 nm and increased PMT voltage of 800 mV. The integral of the emission spectrum from 5 nm after the excitation wavelength to 650 nm.

Single molecule imaging and data processing. Single molecule imaging was performed using a custom built total internal reflection fluorescence (TIRF) microscope based on a Nikon Ti-U inverted microscope and Andor iXon ultra 897 EMCCD camera as outlined in detail in the Supporting Infromation. Single molecule imaging data was processed and analysed using ImageJ[58] and Matlab (R2017a) (MathWorks U.S.A.) as outlined in detail in the Supporting Information.

Acknowledgments.

We would like to thank the Lemke group at EMBL Heidelberg for donating the pEVOL-SCO plasmid. We would like to thank the staff at the Diamond Light Source (Harwell, UK) for the supply of facilities and beam time, especially Beamline I03 and I04 staff. We thank BBSRC (BB/H003746/1 and BB/M000249/1), EPSRC (EP/J015318/1) and Cardiff SynBio Initiative/SynBioCite for supporting this work. HLW was supported by a BBSRC-facing Cardiff University PhD studentship, HSA by the Higher Committee for Education Development in Iraq, and RLJ by KESS studentship. We would like to thank Mr Thomas Williams of the Mass Spectrometry Suite in the School of Chemistry, Cardiff University, and the Protein Technology Hub, School of Biosciences, Cardiff University for use of facilities.

References

1. M. H. Ali and B. Imperiali, *Bioorg Med Chem*, 2005, **13**, 5013-5020.
2. D. S. Goodsell and A. J. Olson, *Annu Rev Biophys Biomol Struct*, 2000, **29**, 105-153.
3. N. J. Marianayagam, M. Sunde and J. M. Matthews, *Trends Biochem Sci*, 2004, **29**, 618-625.
4. G. Mei, A. Di Venere, N. Rosato and A. Finazzi-Agro, *FEBS J*, 2005, **272**, 16-27.
5. C. H. Norn and I. Andre, *Curr Opin Struct Biol*, 2016, **39**, 39-45.
6. N. Kobayashi and R. Arai, *Curr Opin Biotechnol*, 2017, **46**, 57-65.
7. T. O. Yeates, Y. Liu and J. Laniado, *Curr Opin Struct Biol*, 2016, **39**, 134-143.
8. S. Miller, A. M. Lesk, J. Janin and C. Chothia, *Nature*, 1987, **328**, 834-836.
9. A. R. Thomson, C. W. Wood, A. J. Burton, G. J. Bartlett, R. B. Sessions, R. L. Brady and D. N. Woolfson, *Science*, 2014, **346**, 485-488.
10. Y. Mou, P. S. Huang, F. C. Hsu, S. J. Huang and S. L. Mayo, *Proc Natl Acad Sci U S A*, 2015, **112**, 10714-10719.
11. B. S. Der, M. Machius, M. J. Miley, J. L. Mills, T. Szyperski and B. Kuhlman, *J Am Chem Soc*, 2012, **134**, 375-385.
12. J. D. Brodin, X. I. Ambroggio, C. Tang, K. N. Parent, T. S. Baker and F. A. Tezcan, *Nat Chem*, 2012, **4**, 375-382.
13. W. J. Song and F. A. Tezcan, *Science*, 2014, **346**, 1525-1528.
14. D. J. Leibly, M. A. Arbing, I. Pashkov, N. DeVore, G. S. Waldo, T. C. Terwilliger and T. O. Yeates, *Structure*, 2015, **23**, 1754-1768.
15. A. Sahasrabuddhe, Y. Hsia, F. Busch, W. Sheffler, N. P. King, D. Baker and V. H. Wysocki, *Proc Natl Acad Sci U S A*, 2018, DOI: 10.1073/pnas.1713646115.
16. J. A. Fallas, G. Ueda, W. Sheffler, V. Nguyen, D. E. McNamara, B. Sankaran, J. H. Pereira, F. Parmeggiani, T. J. Brunette, D. Cascio, T. R. Yeates, P. Zwart and D. Baker, *Nat Chem*, 2017, **9**, 353-360.

17. A. Chevalier, D. A. Silva, G. J. Rocklin, D. R. Hicks, R. Vergara, P. Murapa, S. M. Bernard, L. Zhang, K. H. Lam, G. Yao, C. D. Bahl, S. I. Miyashita, I. Goreshnik, J. T. Fuller, M. T. Koday, C. M. Jenkins, T. Colvin, L. Carter, A. Bohn, C. M. Bryan, D. A. Fernandez-Velasco, L. Stewart, M. Dong, X. Huang, R. Jin, I. A. Wilson, D. H. Fuller and D. Baker, *Nature*, 2017, **550**, 74-79.
18. D. M. Patterson and J. A. Prescher, *Curr Opin Chem Biol*, 2015, **28**, 141-149.
19. C. H. Kim, J. Y. Axup, A. Dubrovska, S. A. Kazane, B. A. Hutchins, E. D. Wold, V. V. Smider and P. G. Schultz, *J Am Chem Soc*, 2012, **134**, 9918-9921.
20. C. J. White and J. W. Bode, *ACS Cent Sci*, 2018, **4**, 197-206.
21. S. Schoffelen, J. Beekwilder, M. F. Debets, D. Bosch and J. C. van Hest, *Bioconjug Chem*, 2013, **24**, 987-996.
22. J. Torres-Kolbus, C. Chou, J. Liu and A. Deiters, *PLoS One*, 2014, **9**, e105467.
23. E. M. Sletten and C. R. Bertozzi, *Acc Chem Res*, 2011, **44**, 666-676.
24. J. D. Pedelacq, S. Cabantous, T. Tran, T. C. Terwilliger and G. S. Waldo, *Nat Biotechnol*, 2006, **24**, 79-88.
25. A. Rekas, J. R. Alattia, T. Nagai, A. Miyawaki and M. Ikura, *J Biol Chem*, 2002, **277**, 50573-50578.
26. W. Niu and J. Guo, *Mol Biosyst*, 2013, **9**, 2961-2970.
27. J. H. Bae, M. Rubini, G. Jung, G. Wiegand, M. H. Seifert, M. K. Azim, J. S. Kim, A. Zumbusch, T. A. Holak, L. Moroder, R. Huber and N. Budisa, *J Mol Biol*, 2003, **328**, 1071-1081.
28. F. Wang, W. Niu, J. Guo and P. G. Schultz, *Angew Chem Int Ed Engl*, 2012, **51**, 10132-10135.
29. S. C. Reddington, P. J. Rizkallah, P. D. Watson, R. Pearson, E. M. Tippmann and D. D. Jones, *Angew Chem Int Ed Engl*, 2013, **52**, 5974-5977.
30. A. M. Hartley, H. L. Worthy, S. C. Reddington, P. J. Rizkallah and D. D. Jones, *Chemical Science*, 2016, DOI: 10.1039/C6SC00944A.
31. M. Freeley, H. L. Worthy, R. Ahmed, B. Bowen, D. Watkins, J. E. Macdonald, M. Zheng, D. D. Jones and M. Palma, *J Am Chem Soc*, 2017, **139**, 17834-17840.
32. G. Marth, A. M. Hartley, S. C. Reddington, L. L. Sargisson, M. Parcollet, K. E. Dunn, D. D. Jones and E. Stulz, *ACS Nano*, 2017, **11**, 5003-5010.
33. A. J. Zaki, A. Hartley, S. C. Reddington, S. K. Thomas, P. Watson, A. Hayes, A. V. Moskalenko, M. F. Craciun, J. E. Macdonald, D. D. Jones and M. Elliott, *RSC Advances*, 2018, **8**, 5768-5775.
34. S. J. Remington, *Curr Opin Struct Biol*, 2006, **16**, 714-721.
35. B. Salna, A. Benabbas, J. T. Sage, J. van Thor and P. M. Champion, *Nat Chem*, 2016, **8**, 874-880.
36. D. Kozakov, D. R. Hall, B. Xia, K. A. Porter, D. Padhorny, C. Yueh, D. Beglov and S. Vajda, *Nat Protoc*, 2017, **12**, 255-278.
37. A. Leaver-Fay, M. Tyka, S. M. Lewis, O. F. Lange, J. Thompson, R. Jacak, K. Kaufman, P. D. Renfrew, C. A. Smith, W. Sheffler, I. W. Davis, S. Cooper, A. Treuille, D. J. Mandell, F. Richter, Y. E. Ban, S. J. Fleishman, J. E. Corn, D. E. Kim, S. Lyskov, M. Berrondo, S. Mentzer, Z. Popovic, J. J. Havranek, J. Karanicolas, R. Das, J. Meiler, T. Kortemme, J. J. Gray, B. Kuhlman, D. Baker and P. Bradley, *Methods Enzymol*, 2011, **487**, 545-574.
38. R. F. Alford, A. Leaver-Fay, J. R. Jeliazkov, M. J. O'Meara, F. P. DiMaio, H. Park, M. V. Shapovalov, P. D. Renfrew, V. K. Mulligan, K. Kappel, J. W.

Labonte, M. S. Pacella, R. Bonneau, P. Bradley, R. L. Dunbrack, Jr., R. Das, D. Baker, B. Kuhlman, T. Kortemme and J. J. Gray, *J Chem Theory Comput*, 2017, **13**, 3031-3048.
39. T. Plass, S. Milles, C. Koehler, C. Schultz and E. A. Lemke, *Angew Chem Int Ed Engl*, 2011, **50**, 3878-3881.
40. J. Chin, S. Santoro, A. Martin, D. King, L. Wang and P. Schultz, *J Am Chem Soc*, 2002, **124**, 9026-9027.
41. S. Reddington, P. Watson, P. Rizkallah, E. Tippmann and D. D. Jones, *Biochemical Society transactions*, 2013, **41**, 1177-1182.
42. S. C. Reddington, E. M. Tippmann and D. D. Jones, *Chemical communications*, 2012, **48**, 8419-8421.
43. M. H. Seifert, J. Georgescu, D. Ksiazek, P. Smialowski, T. Rehm, B. Steipe and T. A. Holak, *Biochemistry*, 2003, **42**, 2500-2512.
44. T. A. Larsen, A. J. Olson and D. S. Goodsell, *Structure*, 1998, **6**, 421-427.
45. E. Chovancova, A. Pavelka, P. Benes, O. Strnad, J. Brezovsky, B. Kozlikova, A. Gora, V. Sustr, M. Klvana, P. Medek, L. Biedermannova, J. Sochor and J. Damborsky, *PLoS Comput Biol*, 2012, **8**, e1002708.
46. S. A. Mutch, B. S. Fujimoto, C. L. Kuyper, J. S. Kuo, S. M. Bajjalieh and D. T. Chiu, *Biophys J*, 2007, **92**, 2926-2943.
47. P. M. Dijkman, O. K. Castell, A. D. Goddard, J. C. Munoz-Garcia, C. de Graaf, M. I. Wallace and A. Watts, *Nat Commun*, 2018, **9**, 1710.
48. A. Acharya, A. M. Bogdanov, B. L. Grigorenko, K. B. Bravaya, A. V. Nemukhin, K. A. Lukyanov and A. I. Krylov, *Chem Rev*, 2017, **117**, 758-795.
49. R. A. Frank, C. M. Titman, J. V. Pratap, B. F. Luisi and R. N. Perham, *Science*, 2004, **306**, 872-876.
50. A. Shinobu, G. J. Palm, A. J. Schierbeek and N. Agmon, *J Am Chem Soc*, 2010, **132**, 11093-11102.
51. J. J. van Thor, *Chem Soc Rev*, 2009, **38**, 2935-2950.
52. S. C. Reddington, S. Driezis, A. M. Hartley, P. D. Watson, P. J. Rizkallah and D. D. Jones, *RSC Advances*, 2015, **5**, 77734-77738.
53. N. Agmon, *Biophys J*, 2005, **88**, 2452-2461.
54. A. Shinobu and N. Agmon, *J Chem Theory Comput*, 2017, **13**, 353-369.
55. W. Wan, Y. Huang, Z. Wang, W. K. Russell, P. J. Pai, D. H. Russell and W. R. Liu, *Angew Chem Int Ed Engl*, 2010, **49**, 3211-3214.
56. A. J. Rovner, A. D. Haimovich, S. R. Katz, Z. Li, M. W. Grome, B. M. Gassaway, M. Amiram, J. R. Patel, R. R. Gallagher, J. Rinehart and F. J. Isaacs, *Nature*, 2015, **518**, 89-93.
57. M. J. Lajoie, A. J. Rovner, D. B. Goodman, H. R. Aerni, A. D. Haimovich, G. Kuznetsov, J. A. Mercer, H. H. Wang, P. A. Carr, J. A. Mosberg, N. Rohland, P. G. Schultz, J. M. Jacobson, J. Rinehart, G. M. Church and F. J. Isaacs, *Science*, 2013, **342**, 357-360.
58. C. A. Schneider, W. S. Rasband and K. W. Eliceiri, *Nat Methods*, 2012, **9**, 671-675.

Supporting Information.

Methods.

Gene Sequences

> sfGFP E132TAG

ATG GTT AGC AAA GGT GAA GAA CTG TTT ACC GGC GTT GTG CCG ATT CTG GTG GAA CTG GAT
GGT GAT GTG AAT GGC CAT AAA TTT AGC GTT CGT GGC GAA GGC GAA GGT GAT GCG ACC AAC
GGT AAA CTG ACC CTG AAA TTT ATT TGC ACC ACC GGT AAA CTG CCG GTT CCG TGG CCG ACC
CTG GTG ACC ACC CTG ACC TAT GGC GTT CAG TGC TTT AGC CGC TAT CCG GAT CAT ATG AAA
CGC CAT GAT TTC TTT AAA AGC GCG ATG CCG GAA GGC TAT GTG CAG GAA CGT ACC ATT AGC
TTC AAA GAT GAT GGC ACC TAT AAA ACC CGT GCG GAA GTT AAA TTT GAA GGC GAT ACC CTG
GTG AAC CGC ATT GAA CTG AAA GGT ATT GAT TTT AAA **TAG** GAT GGC AAC ATT CTG GGT CAT
AAA CTG GAA TAT AAT TTC AAC AGC CAT AAT GTG TAT ATT ACC GCC GAT AAA CAG AAA AAT
GGC ATC AAA GCG AAC TTT AAA ATC CGT CAC AAC GTG GAA GAT GGT AGC GTG CAG CTG GCG
GAT CAT TAT CAG CAG AAT ACC CCG ATT GGT GAT GGC CCG GTG CTG CTG CCG GAT AAT CAT
TAT CTG AGC ACC CAG AGC GTT CTG AGC AAA GAT CCG AAT GAA AAA CGT GAT CAT ATG GTG
CTG CTG GAA TTT GTT ACC GCC GCG GGC ATT ACC CAC GGT ATG GAT GAA CTG TAT AAA GGC
AGC CAC CAT CAT CAT CAC CAT TAA

> sfGFP H148TAG

ATG GTT AGC AAA GGT GAA GAA CTG TTT ACC GGC GTT GTG CCG ATT CTG GTG GAA CTG GAT
GGT GAT GTG AAT GGC CAT AAA TTT AGC GTT CGT GGC GAA GGC GAA GGT GAT GCG ACC AAC
GGT AAA CTG ACC CTG AAA TTT ATT TGC ACC ACC GGT AAA CTG CCG GTT CCG TGG CCG ACC
CTG GTG ACC ACC CTG ACC TAT GGC GTT CAG TGC TTT AGC CGC TAT CCG GAT CAT ATG AAA
CGC CAT GAT TTC TTT AAA AGC GCG ATG CCG GAA GGC TAT GTG CAG GAA CGT ACC ATT AGC
TTC AAA GAT GAT GGC ACC TAT AAA ACC CGT GCG GAA GTT AAA TTT GAA GGC GAT ACC CTG
GTG AAC CGC ATT GAA CTG AAA GGT ATT GAT TTT AAA GAA GAT GGC AAC ATT CTG GGT CAT
AAA CTG GAA TAT AAT TTC AAC AGC **TAG** AAT GTG TAT ATT ACC GCC GAT AAA CAG AAA AAT
GGC ATC AAA GCG AAC TTT AAA ATC CGT CAC AAC GTG GAA GAT GGT AGC GTG CAG CTG GCG
GAT CAT TAT CAG CAG AAT ACC CCG ATT GGT GAT GGC CCG GTG CTG CTG CCG GAT AAT CAT
TAT CTG AGC ACC CAG AGC GTT CTG AGC AAA GAT CCG AAT GAA AAA CGT GAT CAT ATG GTG
CTG CTG GAA TTT GTT ACC GCC GCG GGC ATT ACC CAC GGT ATG GAT GAA CTG TAT AAA GGC
AGC CAC CAT CAT CAT CAC CAT TAA

> sfGFP Q204TAG

ATG GTT AGC AAA GGT GAA GAA CTG TTT ACC GGC GTT GTG CCG ATT CTG GTG GAA CTG GAT
GGT GAT GTG AAT GGC CAT AAA TTT AGC GTT CGT GGC GAA GGC GAA GGT GAT GCG ACC AAC
GGT AAA CTG ACC CTG AAA TTT ATT TGC ACC ACC GGT AAA CTG CCG GTT CCG TGG CCG ACC
CTG GTG ACC ACC CTG ACC TAT GGC GTT CAG TGC TTT AGC CGC TAT CCG GAT CAT ATG AAA
CGC CAT GAT TTC TTT AAA AGC GCG ATG CCG GAA GGC TAT GTG CAG GAA CGT ACC ATT AGC
TTC AAA GAT GAT GGC ACC TAT AAA ACC CGT GCG GAA GTT AAA TTT GAA GGC GAT ACC CTG
GTG AAC CGC ATT GAA CTG AAA GGT ATT GAT TTT AAA GAA GAT GGC AAC ATT CTG GGT CAT
AAA CTG GAA TAT AAT TTC AAC AGC CAT AAT GTG TAT ATT ACC GCC GAT AAA CAG AAA AAT
GGC ATC AAA GCG AAC TTT AAA ATC CGT CAC AAC GTG GAA GAT GGT AGC GTG CAG CTG GCG
GAT CAT TAT CAG CAG AAT ACC CCG ATT GGT GAT GGC CCG GTG CTG CTG CCG GAT AAT CAT
TAT CTG AGC ACC **TAG** AGC GTT CTG AGC AAA GAT CCG AAT GAA AAA CGT GAT CAT ATG GTG
CTG CTG GAA TTT GTT ACC GCC GCG GGC ATT ACC CAC GGT ATG GAT GAA CTG TAT AAA GGC
AGC CAC CAT CAT CAT CAC CAT TAA

> Venus H148TAG

ATG CGG GGT TCT CAT CAT CAT CAT CAT CAT GGT ATG GCT AGC ATG ACT GGT GGA CAG CAA

ATG GGT CGG GAT CTG TAC GAG AAC CTG TAC TTC CAG GGC TCG AGC ATG GTG AGC AAG GGC
GAG GAG CTG TTC ACC GGG GTG GTG CCC ATC CTG GTC GAG CTG GAC GGC GAC GTA AAC GGC
CAC AAG TTC AGC GTG TCC GGC GAG GGC GAG GGC GAT GCC ACC TAC GGC AAG CTG ACC CTG
AAG CTG ATC TGC ACC ACC GGC AAG CTG CCC GTG CCC TGG CCC ACC CTC GTG ACC ACC CTG
GGC TAC GGC CTG CAG TGC TTC GCC CGC TAC CCC GAC CAC ATG AAG CAG CAC GAC TTC TTC
AAG TCC GCC ATG CCC GAA GGC TAC GTC CAG GAG CGC ACC ATC TTC TTC AAG GAC GAC GGC
AAC TAC AAG ACC CGC GCC GAG GTG AAG TTC GAG GGC GAC ACC CTG GTG AAC CGC ATC GAG
CTG AAG GGC ATC GAC TTC AAG GAG GAC GGC AAC ATC CTG GGG CAC AAG CTG GAG TAC AAC
TAC AAC AGC **TAG** AAC GTC TAT ATC ACC GCC GAC AAG CAG AAG AAC GGC ATC AAG GCC AAC
TTC AAG ATC CGC CAC AAC ATC GAG GAC GGC GGC GTG CAG CTC GCC GAC CAC TAC CAG CAG
AAC ACC CCC ATC GGC GAC GGC CCC GTG CTG CTG CCC GAC AAC CAC TAC CTG AGC TAC CAG
TCC GCC CTG AGC AAA GAC CCC AAC GAG AAG CGC GAT CAC ATG GTC CTG CTG GAG TTC GTG
ACC GCC GCC GGG ATC ACT CTC GGC ATG GAC GAG CTG TAC AAG TAA

In silico modelling of GFP dimer interfaces. ClustPro is a global docking rigid-body approach that requires no prior information on interface regions, and has been shown to be a good predictor of dimer interfaces [1]. ClusPro generates ~100,000 structures and scores them using balanced energy coefficients as described by Kozakov et al [1](Eq 1). E is the energy score of the complex; E_{rep} is the energy of the repulsive contribution of van der Walls interactions and E_{att} is the attractive interaction equivalent. E_{elec} is a term generated by electrostatic energy and EDARS is a term that mainly accounts for free energy change due to exclusion of water from the interface.

Eq 1: $E = 0.40E_{rep} + -0.40E_{att} + 600Ee_{lec} + 1.00E_{DARS}$

The server then takes the 1000 models with the lowest scores and clusters them using pairwise to generate I-RMSD (interface root mean squared deviation). Doing so creates clusters centred on the structure with the most neighbours within a 9 Å radius. Of the remaining models that do not fall within the first cluster the one with the most neighbours becomes the centre of the next cluster and so on until all models are part of a cluster. The centre models of each cluster are energy minimised using the CHARMM force-field for 300 steps with fixed backbone to minimise steric clashes.

Protein production: GFP variants.

The GFP H148TAG, E132TAG and Q204TAG [2] mutants were constructed previously[3, 4]. Super-folder GFP mutant plasmids (based on the pBAD vector) $GFP^{Q204TAG}$ and $GFP^{H148TAG}$ (gene sequence above) were co-transformed by electroporation into *E. coli* Top10 cells (Invitrogen) with either pDULE-cyanoRS (p-azido-L-phenylalanine [azF] incorporation) [5] or pEVOL-SCO (s-cyclooctyne-L-lysine [SCO] incorporation) [6]. The transformed cells were used to inoculate 1L flasks of autoinduction media according to the recipe defined in Studier *et al.* [7] and supplemented with 50 µg/mL carbenicillin and either, 25 µg/mL tetracycline or 35 µg/mL chloramphenicol dependant on whether expressing protein incorporating azF or SCO, respectively. Cultures were grown overnight at 37 °C in a shaking incubator. After 1 hour of growth cultures were inoculated with appropriate non-canonical amino acid to a final concentration of 0.5 mM. Cultures containing azF were kept in the dark until after dimerisation with the SCO-containing protein.

Cells were harvested via centrifugation at 5000 xg for 20 mins. The supernatant was discarded and cells resuspended in 20 mL of 50 mM Tris-HCl pH8.0, 300 mM NaCl, 20 mM imidazole. The cells were lysed using a French press and the resulting lysate was clarified by centrifugation at 25,000 xg for at least 30 minutes. Cell lysates were then loaded onto a 5 mL HisTrapHP™ (GE Healthcare) equilibrated in lysis buffer. Bound GFP was eluted by washing the column in 250 mM Imidazole. Samples were then loaded onto a Superdex 75 column equilibrated in 50 mM Tris-HCl pH8.0 and purity was checked via SDS-PAGE analysis. Concentrations of monomer variants were determined using the Bio-RAD DC Protein Assay using wild type wt GFP as a standard and correlated to the 280 nm absorbance.

Protein production: Venus variants.

The plasmid housing Venus (based on the pBAD vector and procured from Addgene) was used to prepare the Venus variant H148TAG (gene sequence above) via site-directed mutagenesis using Phusion HF polymerase (Finnzymes, Loughborough, Leicestershire). The primer pair, Venus148 F(AACAGCTAGAACGTCTATATCACC) and Venus148 R(GTAGTTGTACTCCAGCTTGTGC) were used. Venus was co-transformed by electroporation into *E. coli* Top10 cells with pDULE-cyanoRS (p-azido-L-phenylalanine [azF] incorporation). The transformed cells were used to inoculate 1L flasks of LB media supplemented with 100 µg/mL ampicillin, 25 µg/mL tetracycline and 0.1 mM of azF. Cultures were grown for 1 hour at 37 °C in a shaking incubator before expression was induced by addition of 0.1% of arabinose and incubated for 24 hours at 25°C. Cultures were kept in the dark until after dimerisation with SCO.

Cells were harvested via centrifugation at 5000 xg for 20 mins. The supernatant was discarded and cells resuspended in 20 mL of 50 mM Tris-HCl pH8.0, 1 mM EDTA. The cells were lysed using a French press and the resulting lysate was clarified by centrifugation at 25,000 xg for at least 30 minutes. Cell lysates were then loaded onto a Protino^{R} Ni-TED 2000 Packed Columns (Machery-Nagel, Germany) equilibrated in equilibration-wash buffer (50 mM Na H_2PO_4, 300 mM NaCl, pH 8) then allowed to drain by gravity. Bound Venus was eluted with 3 bed volumes of elution buffer (50 mM Na H_2PO_4, 300 mM NaCl, 250 mM imidazole, pH 8). Samples were then loaded onto a Superdex 75 column equilibrated in 50 mM Tris-HCl pH8.0 and purity was checked via SDS-PAGE analysis. Concentrations of monomer variants were determined using the Bio-RAD DC Protein Assay using wild type wt GFP as a standard and correlated to the 280 nm absorbance.

Single molecule imaging and data processing

Single molecule imaging was performed using a custom built total internal reflection fluorescence (TIRF) microscope based on a Nikon Ti-U inverted microscope and Andor iXon ultra 897 EMCCD camera. Illumination was provided by a Ventus 473nm DPSS laser with a power output of 100mW. Laser coupling into the microscope was achieved via a custom built optical circuit (components were sourced from Thorlabs, Chroma and Semrock) followed by a single mode fibre-optic launch. Laser power at the microscope stage averaged at 5.8µW/µm^2. The total internal reflection illumination angle was generated using a combination of fibre-optic micro-positioning and a high numerical aperture TIRF objective (Nikon, CFI Apochromat TIRF 60X oil, NA1.49). The Excitation and fluorescence emission wavelengths were separated using a dichroic mirror with a 488nm edge (Chroma zt488rdc-xr). Emitted wavelengths were further filtered using a 500nm edge long pass filter (Chroma hhq500lp) and a 525nm band pass filter (Chroma et525/50m). Acquisitions were controlled using the Andor Solis software package. Frame exposure times were set to 60ms and an EM gain of 250 was used. Coverslips used for TIRF imaging underwent oxygen plasma treated to remove fluorescent contaminants prior to use. Protein solutions were diluted to concentrations suitable for single molecule measurements before droplets were placed onto coverslips for imaging.

Single molecule imaging data was processed and analysed using ImageJ[8] and Matlab (R2017a) (MathWorks U.S.A.). 32 bit floating point TIFF image stacks were used throughout. The first acquisition frame was removed from all image sequences to account for latency of shutter opening by the camera TTL trigger. All images were processed to normalise for spatial variation in intensity profile of the laser illumination using a reference image look-up of relative spatial illumination intensity, mapping the laser illumination created from a Gaussian blurred (20 pixel radius) median z-projection of a fluorescent image stack. The resulting image stack was then corrected for temporal laser intensity fluctuations to minimise the noise in extracted traces. This was achieved by quantifying fluctuations in the global image background and scaling the corresponding frame accordingly, relative to the mean. Practically, this was achieved by removing bright fluorescent spots, defined as any pixel with an intensity greater than 0.05 standard deviations above the median pixel intensity of that frame. Identified pixels were assigned a value equal to the median pixel intensity, effectively erasing them to give a background only image stack. Each frame was scaled relative to the mean intensity of all frames (all pixels) and used to create a temporal lookup table of relative frame to frame laser power fluctuations. This enabled correction of the main image stack. Background counts were subtracted by the pixel-wise subtraction of time averaged median pixel intensity of a background region of interest. Spots were detected using the ImageJ plugin trackmate [9], integrated in the FIJI[10] distribution of ImageJ. Detection was used as a means of automatically identifying spots

and removing distinct "off" states which due to their abundance can mask peaks within intensity distributions. These dark states occur either as a result of photo-bleaching or as part of a natural fluorophore blinking phenomenon. Trackmate detects spots occurring above a background threshold thus spots which are either photobleached or existing in a dark state for the total duration of any given frame are not included in the detection process. An estimated spot diameter of 4 pixels was applied with a difference of Gaussian (DoG) detection routine. This applies differently sized Gaussian blurs (greater or lesser than the estimated spot diameter) to two copies of each frame which are then subtracted from one another. This process acts as a spatial bandpass filter enhancing features in the range of the estimated spot diameter enabling detection. As spots were static, linking was performed using spot linking and gap closing distances of 1 pixel. A frame gap closing distance of 3 was also used to link spots displaying long "off" states. Data was exported to matlab where a Gaussian mixture model was fitted to the logarithm of resultant frequency of intensity values for all spots of a given dimer. To assess the suitability of fitting, four gaussian mixture models with components ranging from 1-4 were fit to both GFP^{148x2} and GFP^{wt} (1000 replicates/model). The mean Akaike information criteria was calculated for each model with the minimal value indicative of the most probable fit.In addition spatial coordinates of tracked spots were used to generate representative traces from corrected stacks. For each dimer, data sets consisted of two separate acquisitions 24 seconds (400 frames) in length amassing information from ~200 dimer pairs each.

Protein structure determination

Samples of GFP^{148x2} were concentrated to 10 mg/mL using spin concentrators (10,000 Da mW cut-off). Crystallisation trays were set up using either JCB or PACT pre-made crystallography screens. Trays were monitored regularly to check for crystal formation. Data were collected at the Diamond Light Source (Harwell, UK). Data were reduced using the XIA2 package [11] assigned a space group using POINTLESS [12], scaled using SCALA [12] and merged using TRUNCATE [13]. Structures were solved by molecular replacement with PHASER, using a previously determined sfGFP structure (PDB code 2B3P). Structures were then adjusted manually using COOT [14] and refined by TLS restrained refinement using RefMac [15]. All the above programs were accessed via the CCP4 package (http://www.ccp4.ac.uk/) [13].

Mass Spectrometry

Protein samples were buffer exchanged into fresh 50 mM Tris-HCl pH 8.0 and diluted to 10 μM, for mass spectrometry analysis. Samples were recorded by liquid chromatography time of flight mass spectrometry (LC/TOF-MS) using a Waters Synapt G2-Si QT in positive Electrospray ionisation mode. Mass peaks between 200-2,000 Da were recorded in positive Electrospray ionisation mode using Leucine Enkephalin as a calibrant. The data was processed using MassLynx 4.1 programme using the Maximum entropy 1 add on. Proteins were passed through a Waters Acquity UPLC CSH 130 C18 (80°C) and eluted using a gradient of acetonitrile (5-95%) in 0.1% formic acid over 5 minutes.

Table S1. Statistics for *in silico* modelling of GFP dimer interfaces.

Model [a]	Total energy (kJ/mole)	Interface Energy (kJ/mole)	I-RSMD ($Å^2$)	RMSD GFP^{148x2} ($Å^2$) [b]
Model 5	-503.94	-13.434	1.755	**4.72**
Model 1	-501.966	-4.192	0.278	9.53
Model 4	-497.071	-6.993	0.397	18.96
Model 2	-497.112	-6.375	0.56	11.95
Model 3	-494.492	-3.079	0.195	8.31

[a] models ordered according to their rank. [b] compared to the determined structure

Table S2. Spectral properties of GFP and Venus variants.

Variant	λ_{max} (nm)	λ_{EM} (nm)	ε ($M^{-1}cm^{-1}$)	QY	Brightness
wt GFP	485	511	49000[a]	0.75[a]	36750
GFP^{148AzF} *b*	400	511	34200	0.69	23598
	500	511	19800	0.32	6336
GFP^{148SCO}	395	511	31000	0.52	16120
	492	511	17300	0.84	14532
GFP^{148x2}	492	511	150200	0.80	120160
GFP^{204AzF}	485	511	51000	0.68	34680
GFP^{204SCO}	485	511	39800	0.66	26268
GFP^{204x2}	490	513	160000	0.71	113600
wt Venus	515	528	92200[c]	0.65	59930
$Venus^{148azF}$	517	525	30100	0.45	13545
$GFVen^{148}$	400	517	24000	0.42	10080
	505	517	96000	0.46	44160

[a] We have reported previously a significant shortfall in the molar absorbance coefficient we routinely calculate (here and [2-4]) and that published by Pedelacq et al [16]. [b] Values reported previously [3]. [c] Published previously by Nagai et al [17] and measured in the current study as 95000 $M^{-1}cm^{-1}$. Given that our value is close to the reported value, we have used the reported value.

Table S3: Crystallographic statistics for GFP^{148x2}

	GFP^{148x2}
PDB ID	5NHN

Wavelength (Å)	0.979
Beamline	Diamond IO4
Space group	$P6_5$
a (Å)	99.80
b (Å)	99.8
c (Å)	108.92
Resolution range (Å)	67.71-1.96
Total reflections measured	964841
Unique reflections	41,916
Completeness (%) (last shell)	100 (99.9)
Multplicity (last shell)	21.8 (14.1)
I/σ (last shell)	22.9 (4.0)
CC1/2	1.000 (0.680)
R(merge)[a] (%) (last shell)	7.9 (68.8)
B(iso) from Wilson ($Å^2$)	41.03
B(iso) from refinement	50.8
Log Likelihood Coordinate rms	0.126
Non-H atoms	3877
Solvent molecules	226
R-factor[b] (%)	18.2
R-free[c] (%)	21.1
Rmsd bond lengths (Å)	0.015
Rmsd bond angles (°)	1.868
Core region (%)	98.42
Allowed region (%)	1.13
Additionally allowed region (%)	0
Disallowed Region (%)	0.46

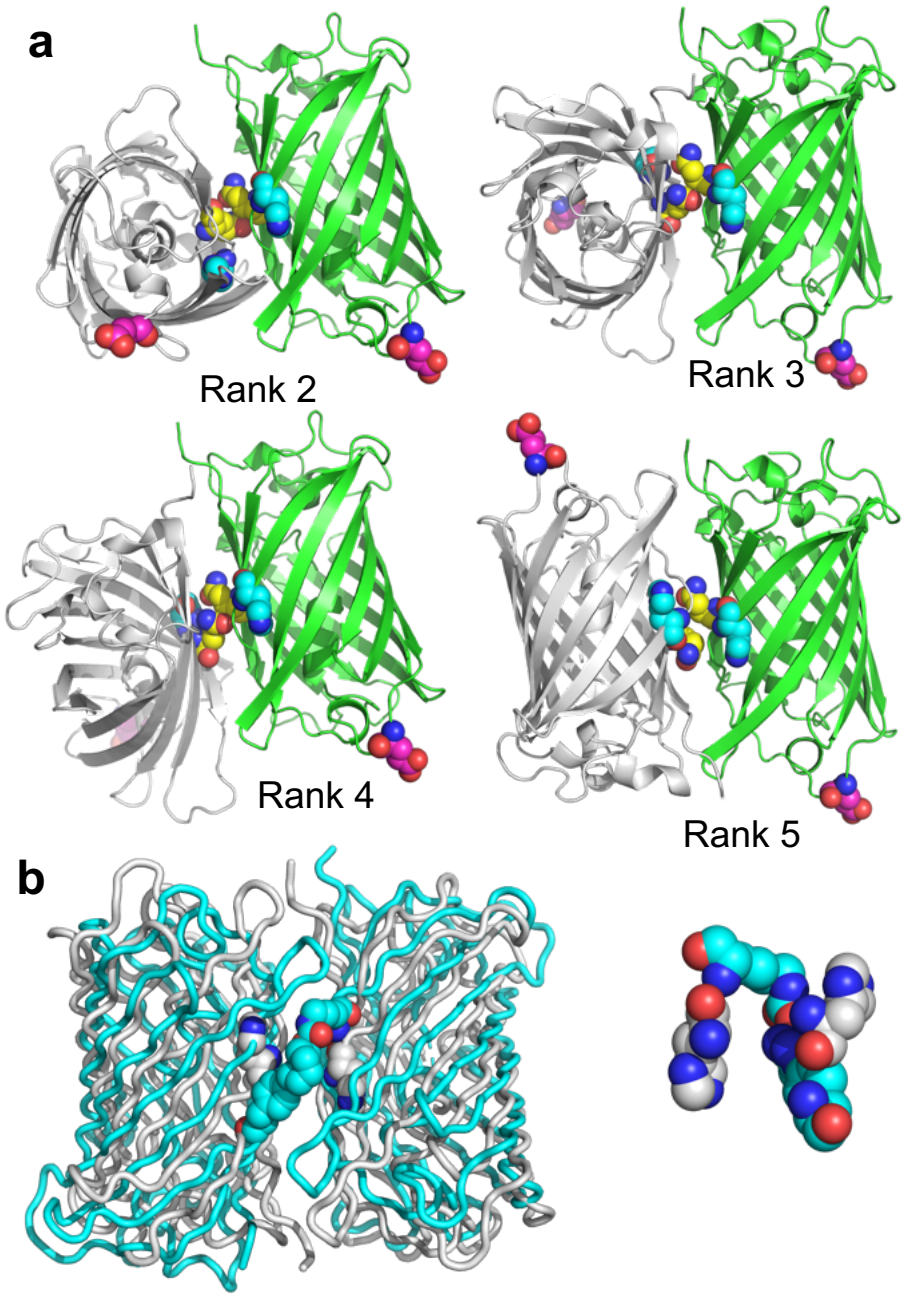
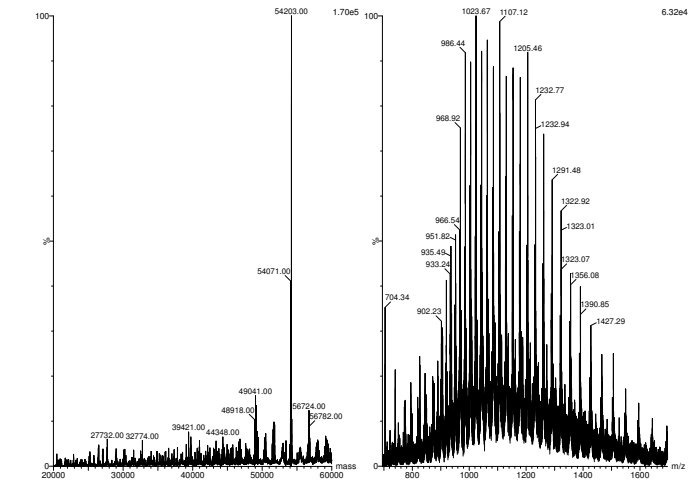


Figure S1. Models of GFP dimerisation. (a) The 2nd to 5th ranked models of GFP dimerisation. The top ranked model is shown in the main text. Ranking was performed as described in the Supporting Methods. Statistics are shown in Table S1. The Glu132, His148 and Gln204 are coloured magenta, cyan and yellow, respectively. The reference GFP structure is coloured green. (b) Overlay of GFP^{148x2} structure (cyan) with the closest model (grey, model rank 1st), with a calculated RMSD of 4.7 Å. The 148 residues are also shown separated as spheres.

a



b

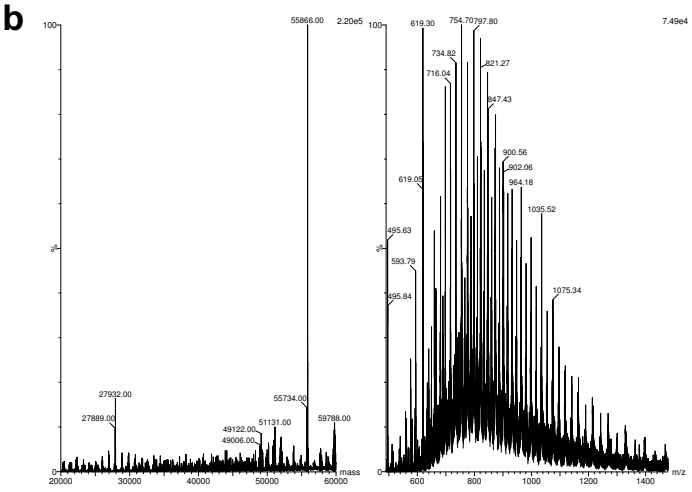


Figure S2. GFP dimer formation. (a) Mass spectrum of the GFP^{148x2} dimer. The theoretical molecular weight for full length dimerised protein is 55846 Da. The observed mass (54203 Da) matches the loss of the His tag from each monomer (823 Da x 2 = 1646 Da; 54200 Da). (b) Mass spectrum of the GFP^{204x2} dimer. The theoretical molecular weight for full length dimerised protein is 55864 Da, with a mass of 55866 Da observed.

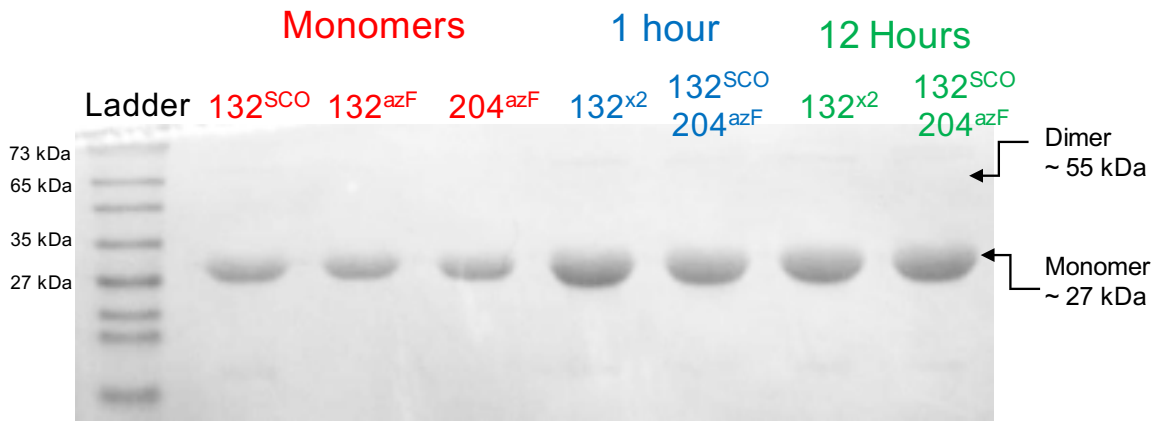


Figure S3. Dimerisation potential of a non-dimer interface residue, as predicted by *in silico* modelling. The residue predicted not to form part of the interface is Glu132 of GFP. The SCO ncAA was incorporated at residue 132 (132^{SCO}) of GFP and dimerisation with azF incorporated at either residue 132 (132^{azF}) or 204 (204^{azF}) of GFP. No clear dimerisation product was observed for the 132^{azF}-132^{SCO} (132^{x2}) or 132^{SCO}-204^{azF} by gel mobility shift assay.

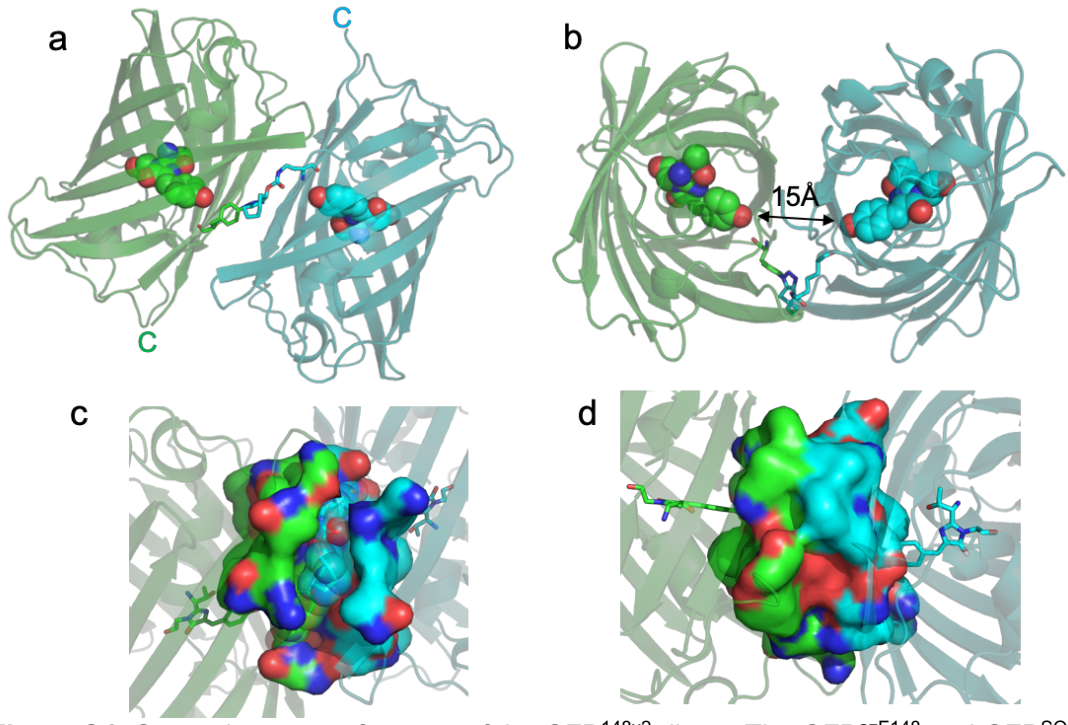


Figure S4. General structure features of the GFP^{148x2} dimer. The GFP^{azF148} and GFP^{SOC148} monomers are coloured green and cyan, respectively. Side on (a) and top-down view of the dimer arrangement with the SPAAC crosslinking shown in sticks, the CRO as spacefill and the C-terminal labled with a 'C'. The hydrophobic packing around the dimer interface as seen from two separate views (c and d). The SPAAC crosslink is shown as transparent spheres in d.

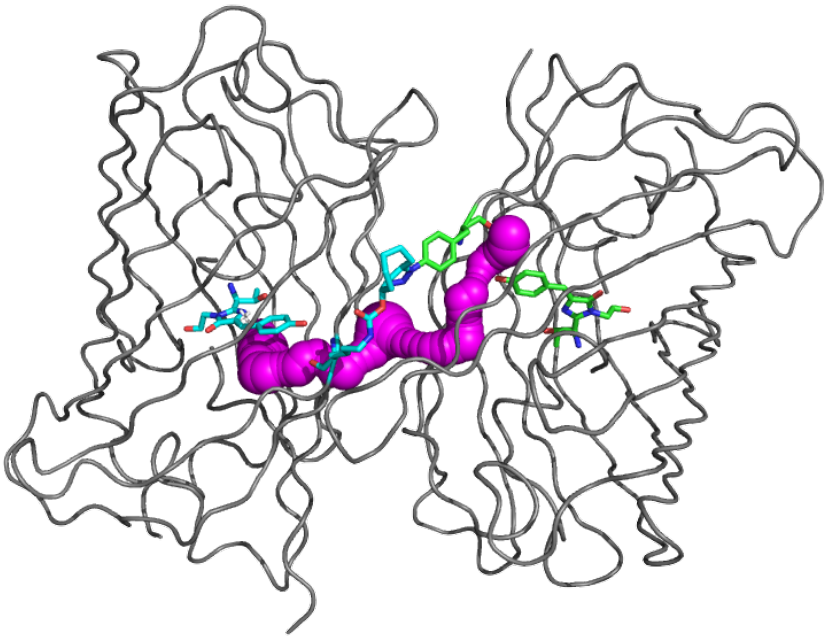


Figure S5. CAVER [18] analysis of the channel linking the two CRO of GFP^{148x2}.

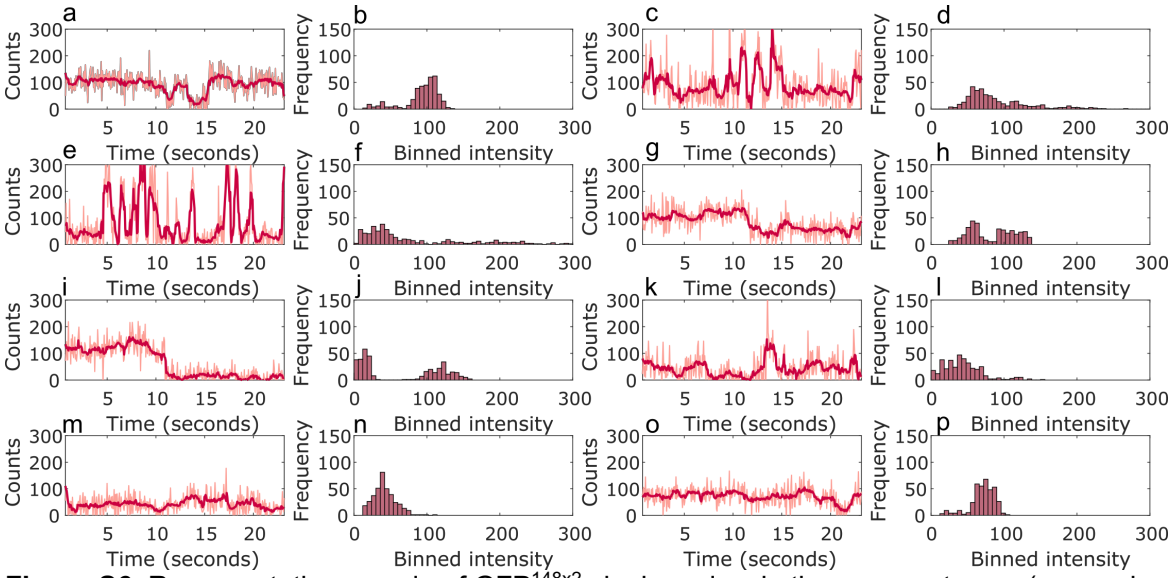


Figure S6. Representative sample of GFP^{148x2} single molecule time course traces (raw and Cheung-Kennedy filtered) coupled with paired intensity frequency histograms (generated from Cheung Kennedy filtered data) to the right of each trace. Traces highlight the complexity of behaviour demonstrated by dimers at the single molecule level, showing a range of fluorescence states, transitions and on times. Histograms show no well defined or recurring intensity peaks emphasising the inherent intensity variability of GFP^{148x2} in contrast to GFP^{WT}.

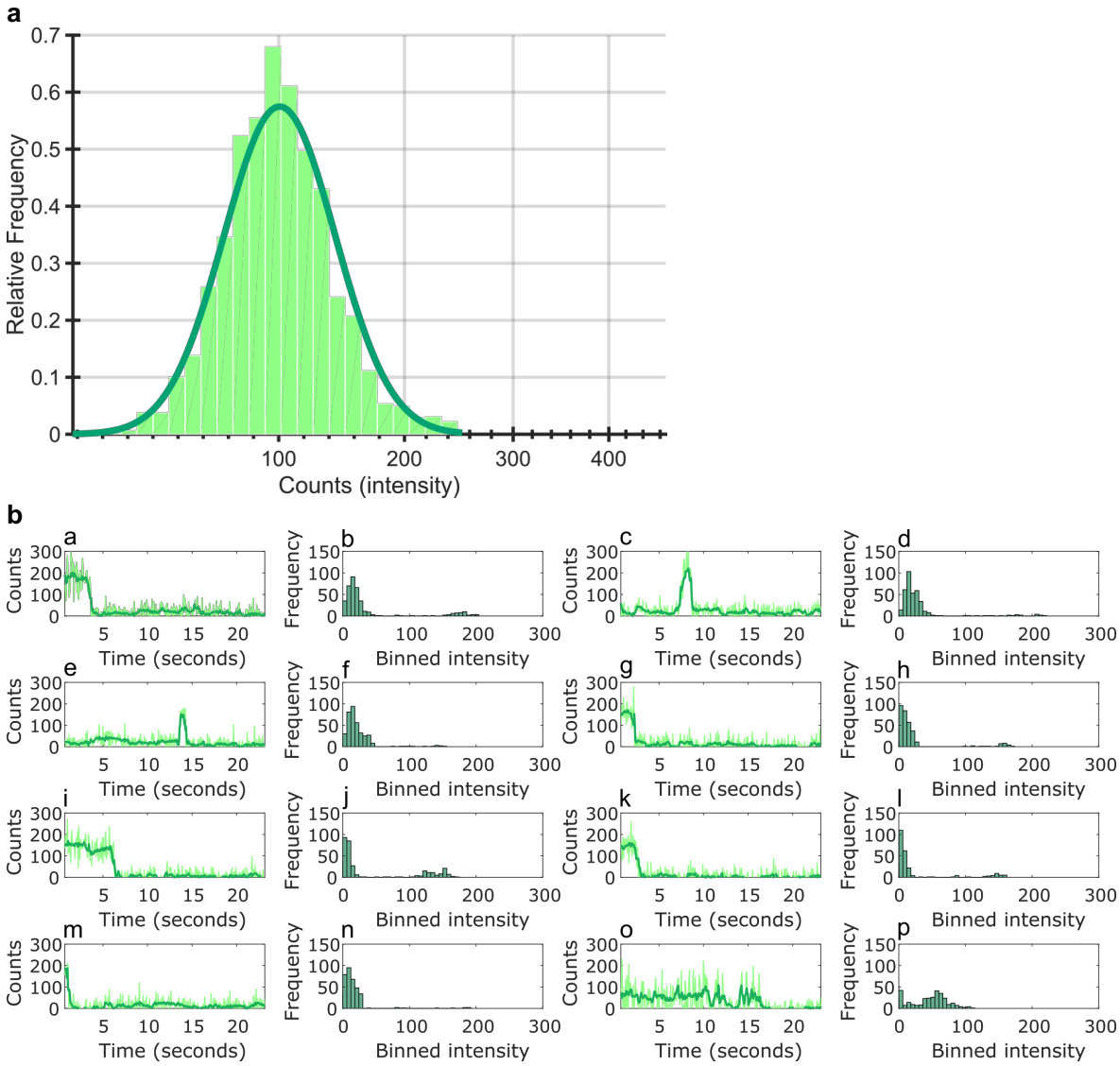


Figure S7. (a) A single molecule fluorescence intensity histogram for wt GFP consisting of 204 trajectories (2244 spots). The histogram data fits to a single log normal distribution centred around 100 counts. (b) Representative sample of monomeric wt GFP single molecule time course traces (raw and Cheung-Kennedy filtered) coupled with paired intensity frequency histograms (generated from Cheung Kennedy filtered data) to the right of each trace. Common intensity states are predominant in traces which often contain single clear-cut transitions alongside long lived dark states. In the majority of traces, transition to a darkstate occurs after a relatively short period of time. Histograms generally show clear separation between baseline and intensity peaks which commonly arise between counts of 100-200.

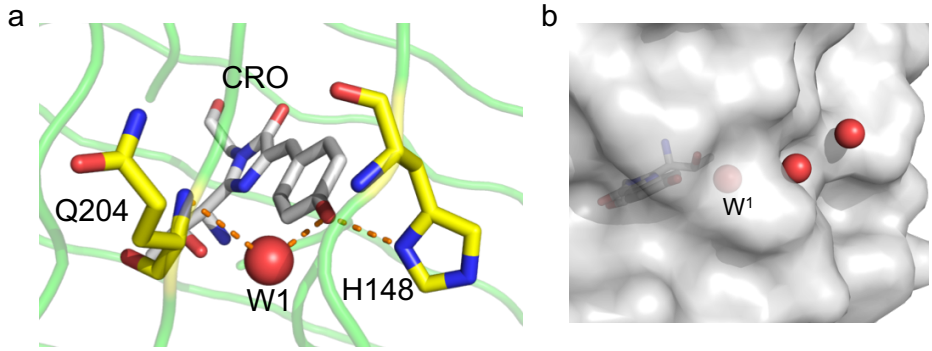


Figure S8. (a) Molecular interactions involving H148 and Q204 (yellow), the chromophore (CRO; grey) and water molecule W1 (red ball). (b) Extended water molecule (red balls) network from CRO (sticks) to surface in wt GFP. Water molecule W^1 is indicated on the figure. GFP is shown in surface representation.

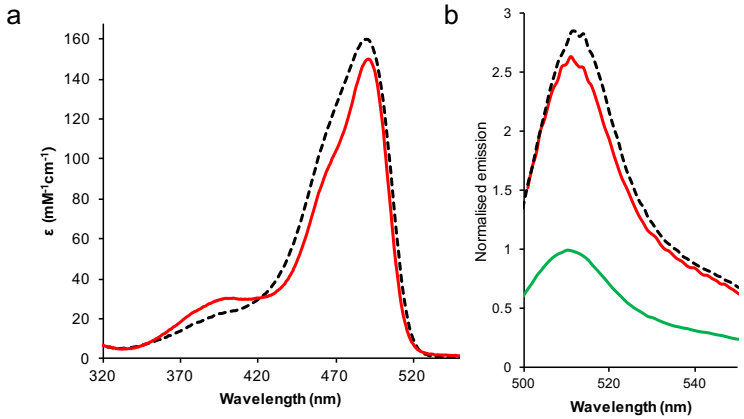


Figure S9. Comparison of the GFP^{148x2} (red line), GFP^{204x2} (dashed line) and wt GFP (green line). (a) Absorbance spectra of the two dimer species. (b) Emission spectra of two dimers and monomeric wt GFP (0.5 μM, excitation at λ_{max} (Table S1)). Fluorescence is normalised to the wt GFP intensity.

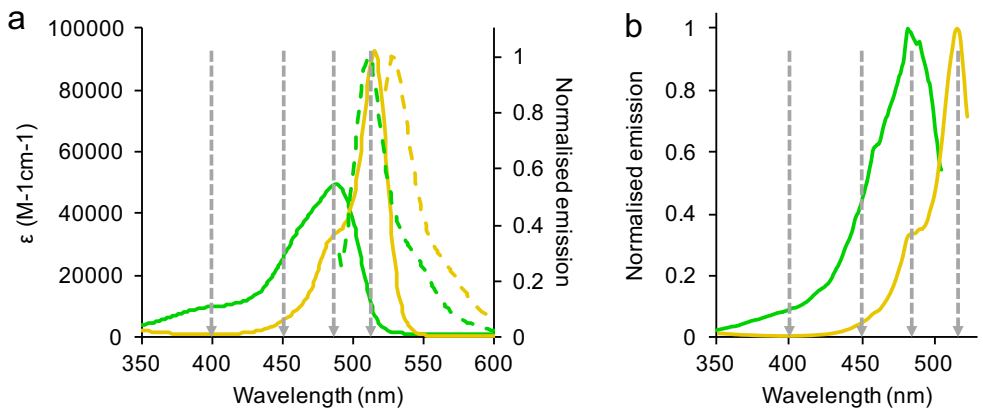


Figure S10. Comparison of the GFP (green) and Venus (gold) absorbance and fluorescence spectra. (a) absorbance (solid line) and emission spectra (dashed line). (b) Excitation spectra (on monitoring emission at λ_{EM}; Table S1). The grey dashed arrows indicate wavelengths used to monitor communication between two monomers in GFVen dimers.

```
Venus          -35 MRGSHHHHHHGMASMTGGQQMGRDLYENLYFQGSSMVSKGEELFTGVVPI     14
                                                   X|||||||||||||||
sfGFP                                            1-MSKGEELFTGVVPI     14

Venus           15 LVELDGDVNGHKFSVSGEGEGDATYGKLTLKLICTTGKLPVPWPTLVTTL     64
                   ||||||||||||||||X|||||||X||||||X||||||||||||||||||
sfGFP           15 LVELDGDVNGHKFSVRGEGEGDATNGKLTLKFICTTGKLPVPWPTLVTTL     64

Venus           65 GYGLQCFARYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKF     114
                   X||X|||X|||||||X|||||||||||||||||||X||||||X|||||||||
sfGFP           65 TYGVQCFSRYPDHMKRHDFFKSAMPEGYVQERTISFKDDGTYKTRAEVKF     114

Venus          115 EGDTLVNRIELKGIDFKEDGNILGHKLEYNYNSHNVYITADKQKNGIKAN     164
                   ||||||||||||||||||||||||||||||X|||||||||||||||||||
sfGFP          115 EGDTLVNRIELKGIDFKEDGNILGHKLEYNFNSHNVYITADKQKNGIKAN     164
                                                      ^
Venus          165 FKIRHNIEDGGVQLADHYQQNTPIGDGPVLLPDNHYLSYQSALSKDPNEK     214
                   ||||||X|||X|||||||||||||||||||||||||||X||X||||||||
sfGFP          165 FKIRHNVEDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSVLSKDPNEK     214

Venus          215 RDHMVLLEFVTAAGITLGMDELYK--------     274
                   ||||||||||||||||X|||||||
sfGFP          215 RDHMVLLEFVTAAGITHGMDELYKLEHHHHHH     246
```

Figure S11. Sequence alignment between the versions of Venus and GFP used in the current study. The mutated H148 is shown in bold. Blue, red and green highlighted residues correspond to His tags, TEV protease cleavage motif and CRO forming residue, respectively. Residues that differ between the two are highlighted by a X.

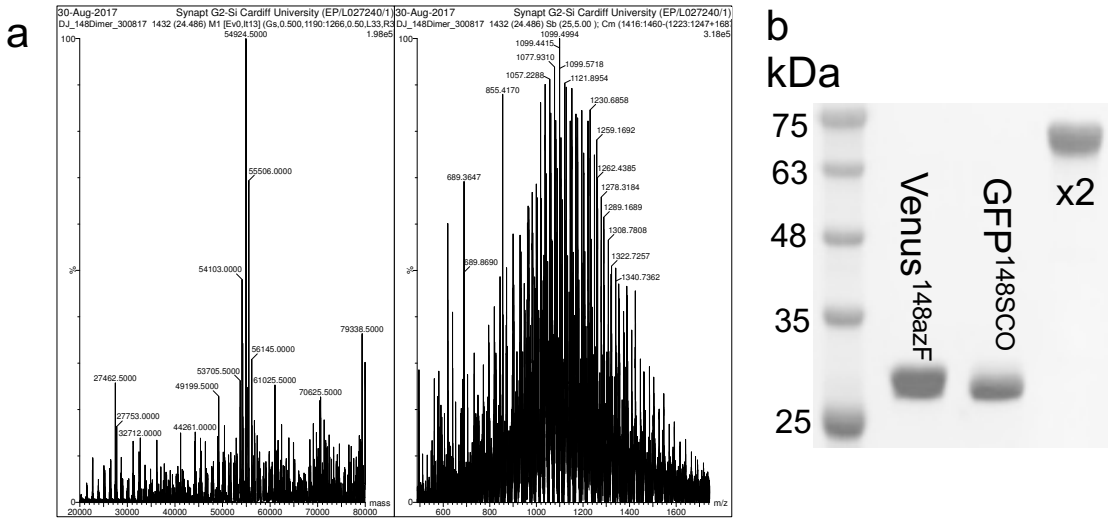


Figure S12. Dimerisation of GFVen^{148}. (a) Mass spectra of GFVen^{148}. The major peak at 54924 Da corresponds to GFP^{204SCO} (27698 Da) and Venus^{148azF} with a truncated N-terminal extension (up to -4-GSSM in Figure S15; 26956 Da), with has a calculated molecular mass of 54924 Da. The second smaller peak at 55506 Da, corresponds to GFP^{204SCO} and Venus^{148azF} minus the N-terminal extension up (to -7-YFQG; 27539Da), with a calculated mass of 55507 Da. The third minor peak at 54103 Da corresponds to GFP^{148SCO} with the loss of its C-terminal His-tag (27145 Da) and Venus^{148azF} with the N-terminal extension truncated (up to -4-GSSM in Figure S13; 26956 Da), which has a calculated molecular mass of 54101 Da. The GFVen^{148} dimer was confirmed independently by SDS-PAGE, as shown in the Figure 1 of the main text. Terminal processing of the H148 variants seems to be a common theme (see Figure S4). SDS PAGE analysis of GFVen^{148} is shown in Figure 1 of the main manuscript. (b) SDS PAGE analysis of GFVen^{148} dimerisation.

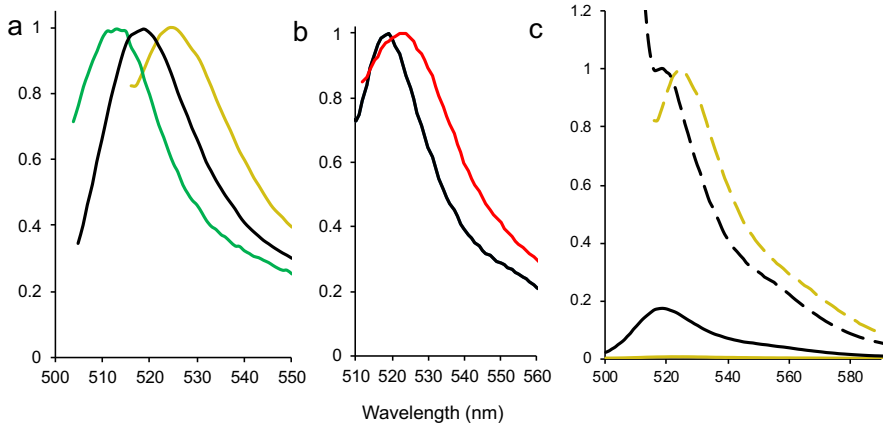


Figure S13. Comparison of emission spectra. (a) GFP^{148azF} (green; excitation 490 nm), Venus^{148azF} (gold; excitation 510 nm) and GFVen^{148} (black; excitation 490 nm). (b) Comparison of measured emission spectra (on excitation at 505 nm) of GFVen^{148} (black line) and additive emission spectrum of GFP^{148SCO} (excitation at 490 nm) and Venus^{148azF} (excitation at 505 nm). The molar absorbance coefficient for each monomer at their excitation wavelengths was similar (~17,200 and ~17,800 $M^{-1}cm^{-1}$, respectively). (c) Emission of Venus^{148azF} (gold) and GFVen^{148} (black) on excitation at 400 nm (solid line) or 510 nm (dashed line).

Supporting References

1. D. Kozakov, D. R. Hall, B. Xia, K. A. Porter, D. Padhorny, C. Yueh, D. Beglov and S. Vajda, *Nat Protoc*, 2017, **12**, 255-278.
2. S. C. Reddington, E. M. Tippmann and D. D. Jones, *Chemical communications*, 2012, **48**, 8419-8421.
3. A. M. Hartley, H. L. Worthy, S. C. Reddington, P. J. Rizkallah and D. D. Jones, *Chemical Science*, 2016, DOI: 10.1039/C6SC00944A.
4. S. C. Reddington, P. J. Rizkallah, P. D. Watson, R. Pearson, E. M. Tippmann and D. D. Jones, *Angew Chem Int Ed Engl*, 2013, **52**, 5974-5977.
5. S. J. Miyake-Stoner, C. A. Refakis, J. T. Hammill, H. Lusic, J. L. Hazen, A. Deiters and R. A. Mehl, *Biochemistry*, 2010, **49**, 1667-1677.
6. T. Plass, S. Milles, C. Koehler, C. Schultz and E. A. Lemke, *Angew Chem Int Ed Engl*, 2011, **50**, 3878-3881.
7. F. W. Studier, *Protein Expr Purif*, 2005, **41**, 207-234.
8. C. A. Schneider, W. S. Rasband and K. W. Eliceiri, *Nat Methods*, 2012, **9**, 671-675.
9. J. Y. Tinevez, N. Perry, J. Schindelin, G. M. Hoopes, G. D. Reynolds, E. Laplantine, S. Y. Bednarek, S. L. Shorte and K. W. Eliceiri, *Methods*, 2017, **115**, 80-90.
10. J. Schindelin, I. Arganda-Carreras, E. Frise, V. Kaynig, M. Longair, T. Pietzsch, S. Preibisch, C. Rueden, S. Saalfeld, B. Schmid, J. Y. Tinevez, D. J. White, V. Hartenstein, K. Eliceiri, P. Tomancak and A. Cardona, *Nat Methods*, 2012, **9**, 676-682.
11. G. Winter, *J Appl Crystallogr*, 2009, **43**, 196-190.
12. P. Evans, *Acta Crystallogr D Biol Crystallogr*, 2006, **62**, 72-82.
13. S. Bailey, *Acta Crystallographica Section D-Biological Crystallography*, 1994, **50**, 760-763.
14. P. Emsley and K. Cowtan, *Acta Crystallogr D Biol Crystallogr*, 2004, **60**, 2126-2132.
15. G. N. Murshudov, A. A. Vagin and E. J. Dodson, *Acta Crystallogr D Biol Crystallogr*, 1997, **53**, 240-255.
16. J. D. Pedelacq, S. Cabantous, T. Tran, T. C. Terwilliger and G. S. Waldo, *Nat Biotechnol*, 2006, **24**, 79-88.
17. T. Nagai, K. Ibata, E. S. Park, M. Kubota, K. Mikoshiba and A. Miyawaki, *Nat Biotechnol*, 2002, **20**, 87-90.
18. E. Chovancova, A. Pavelka, P. Benes, O. Strnad, J. Brezovsky, B. Kozlikova, A. Gora, V. Sustr, M. Klvana, P. Medek, L. Biedermannova, J. Sochor and J. Damborsky, *PLoS Comput Biol*, 2012, **8**, e1002708.