Submission to: Photochemical & Photobiological Sciences

Manuscript type: Full paper

Thermostable flavin-based fluorescent protein from *Chloroflexus aggregans*: a framework for ultra-high resolution structural studies

Vera V. Nazarenko[1], Alina Remeeva[1], Anna Yudenko[1], Kirill Kovalev[1,2,3,4], Anton Dubenko[1], Ivan M. Goncharov[1], Pavel Kuzmichev[1], Andrey V. Rogachev[1,5], Pavel Buslaev[1], Valentin Borshchevskiy[1], Alexey Mishin[1], Gaurao V. Dhoke[6], Ulrich Schwaneberg[6,7], Mehdi D. Davari[6], Karl-Erich Jaeger[8,9], Ulrich Krauss[8], Valentin Gordeliy[1,2,3], Ivan Gushchin[1,*]

[1] Moscow Institute of Physics and Technology, Dolgoprudny, Russia

[2] ICS-6: Structural Biochemistry, Forschungszentrum Jülich GmbH, Jülich, Germany

[3] Institut de Biologie Structurale Jean-Pierre Ebel, Université Grenoble Alpes–Commissariat à l'Energie Atomique et aux Energies Alternatives–CNRS, Grenoble, France

[4] Institute of Crystallography, RWTH Aachen University, Aachen, Germany

[5] Joint Institute for Nuclear Research, Dubna, Russia

[6] Institute of Biotechnology, RWTH Aachen University, Aachen, Germany

[7] DWI-Leibniz Institut für Interaktive Materialien, Aachen, Germany

[8] Institute of Molecular Enzyme Technology, Heinrich Heine University Düsseldorf, Forschungszentrum Jülich, Jülich, Germany

[9] IBG-1: Biotechnology, Forschungszentrum Jülich GmbH, Jülich, Germany

[*] E-mail for correspondence: ivan.gushchin@phystech.edu

Electronic supplementary information is available.

Abstract

Light-Oxygen-Voltage (LOV) domains are conserved parts of photoreceptors in plants, bacteria and fungi that bind flavins as chromophore and detect blue light. In the past, LOV domain variants have been developed as fluorescent reporter proteins (called flavin-based fluorescent proteins; FbFPs), which due to their ability to fluoresce under anaerobic conditions, fast folding kinetics as well as small size of ~12–16 kDa are a promising reporter system for quantitative real-time analysis of biological processes. Here, we present a small thermostable flavin-based fluorescent protein CagFbFP derived from a soluble LOV domain-containing histidine kinase from the thermophilic bacterium *Chloroflexus aggregans*. CagFbFP is composed of 107 amino acids with a molecular weight of 11.6 kDa and consists only of the conserved LOV core domain. The protein is thermostable with a melting point of about 68 °C. It crystallizes easily and its crystals diffract to 1.07 Å. Both the crystal structure and small angle scattering data show that the protein is a dimer. Unexpectedly, glutamine 148, which in LOV photoreceptor proteins is the key residue responsible for signal transduction, occupies two conformations. Molecular dynamics simulations show that the two conformations interconvert rapidly. The crystal structure of wild-type *Chloroflexus aggregans* LOV domain determined at 1.22 Å resolution confirmed the presence of two alternative conformations of the glutamine 148 side chain. Overall, this protein, due to its stability and ease of crystallization, appears to be a promising model for ultra-high resolution structural studies of LOV domains and for application as a fluorescent reporter.

Keywords: Light-Oxygen-Voltage (LOV) domains, flavin mononucleotide (FMN), photoreceptor, fluorescent tag, X-ray crystallography

Running title: LOV-based fluorescent protein from *Chloroflexus aggregans*

Introduction

Light-Oxygen-Voltage (LOV) domains are conserved parts of photoreceptors in plants, bacteria and fungi that detect blue light[1–4]. Structurally, as members of the PAS domain family[5], they have a mixed α/β fold. LOV domains bind flavins such as flavin mononucleotide (FMN) or flavin adenine dinucleotide (FAD) as chromophores and have characteristic absorption spectra with maxima at around 450 nm. Photoexcitation of the flavin chromophore results in formation of a covalent flavin-cysteinyl thiol adduct between the flavin C4a atom and a nearby conserved cysteine residue, which in turn induces conformational changes in the LOV domains that eventually lead to activation of various effector domains, such as kinases, esterases and DNA binding motifs[6]. One of the most conserved features of LOV domain activation, apart from adduct formation, is altered hydrogen bonding of the flavin with an adjacent glutamine side chain (Q123 in YtvA, Q182 in Vivid (VVD), Q154 in ZEITLUPE (ZTL) and Q513 in *Avena sativa* phototropin 1 LOV2), which is suggested to rotate in the active conformation in response to protonation of the flavin and formation of the cysteine-flavin adduct[2,3,7]. In the dark, the covalent cysteine-flavin bond is broken, thus concluding the LOV photocycle[3,8,9].

In the past, several LOV domain variants have been developed as fluorescent reporter proteins (called flavin-based fluorescent proteins; FbFPs), which due to their ability to fluoresce under anaerobic conditions, fast folding kinetics as well as small size of ~12–16 kDa are a promising reporter system for the quantitative real-time analysis of biological processes[10,11]. The first FbFPs, BsFbFP, EcFbFP and PpFbFP, were engineered from the LOV domains of the bacterial blue-light photoreceptor YtvA from *B. subtilis* and the PpSB2-LOV protein from *P. putida*, by replacing the adduct-forming cysteine residue of the photoreceptor with alanine, which abolishes LOV photocycling and results in sustained fluorescence emission[12]. EcFbFP represents a shorter version of BsFbFP for which the codon usage of the corresponding gene was adjusted to the *E. coli* codon usage bias[12]. Soon thereafter, shuffling of four *Arabidopsis thaliana* phototropin LOV domain-coding sequences resulted in development of the fluorescent protein iLOV, which outperformed GFP as a reporter of plant virus infection[13]. iLOV was later improved for enhanced photostability, resulting in a series of phiLOV constructs[14].

While increased thermostability is a desirable property of fluorescent proteins, for them to be used as fluorescence reporters in thermophilic microbes, the first FbFPs unfolded at moderate temperatures[15]. It was possible to increase the melting temperature of BsFbFP by 31 °C, but it was not clear how the brightness of the protein is affected by the introduced

stabilizing mutations[16]. Therefore, genome mining approaches have been employed to discover naturally thermostable proteins. The first protein identified using such an approach, CreiLOV, was derived from the phototropin LOV1 domain of the alga *Chlamydomonas reinhardtii*[17]. Later, a plethora of bacterial thermostable FbFPs have been described[18], some of which showed remarkable thermo- and photostability.

Here we describe the photophysical and structural characterization of a new thermostable flavin-based fluorescent protein, CagFbFP, derived from the LOV domain of a soluble histidine kinase from *Chloroflexus aggregans*. Our study highlights CagFbFP as a promising model for ultra-high resolution structural studies of LOV domains and for application as a fluorescent reporter.

Results and Discussion

A comprehensive bioinformatics study performed by Glantz *et al.*[6] revealed a LOV domain-containing protein Cagg_3753 in the genome of *Chloroflexus aggregans*, a thermophilic filamentous anoxygenic phototrophic bacterium, which forms dense cell aggregates and was found in a hot spring in Japan[19]. Cagg_3753 is annotated as a putative soluble histidine kinase protein[20]. It is composed of three consecutive PAS domains, with the first one being a LOV domain, rendering Cagg_3753 a putative LOV photoreceptor. The identity and function of the two other PAS domains is not clear. The PAS domains are followed by dimerization and histidine phosphotransfer domain DHp, a catalytic domain CA, and additional response regulator RR and histidine phosphotransfer Hpt domains (predicted by InterPro[20]), as is typical for sensor histidine kinases[21–23].

To test whether Cagg_3753 can be engineered into a functional flavin-binding fluorescent protein, we synthesized genes corresponding to residues 47-153 (shorter construct, CagFbFP) and 8-175 (extended construct, CagFbFPe) with the conserved adduct-forming cysteine residue Cys85 substituted by alanine. Similarly to other LOV proteins, in Cagg_3753 the LOV domain is predicted to be flanked by α-helices (traditionally called A'α and Jα). Such helices are frequently involved in dimerization (in LOV domains overall, and in histidine kinases in particular[21–23]), which would be an undesirable property for a fluorescent tag. Consequently, we constructed CagFbFP as completely devoid of A'α and Jα residues, whereas CagFbFPe includes potential helical fragments, similarly to some other described FbFP proteins (Figure 1).

Both constructs expressed well in *E. coli* in flavin-bound form and were purified using metal-affinity and size exclusion chromatography. Absorption and fluorescence excitation and

fluorescence emission spectra of CagFbFP are shown in Figure S1. Judging from the SEC profiles, it was not clear whether CagFbFP is a monomer or a dimer, whereas CagFbFPe clearly was multimeric (Figure 2A). Initial experiments conducted using Rotor-Gene Q real-time PCR cycler indicated that CagFbFP is thermostable as it unfolded at around 68 ºC (Figure 2B). CagFbFPe was even more thermostable, which is consistent with the A'α and Jα helices stabilizing the protein assembly (Figure 2B). To gain structural insight about both CagFbFP variants, we employed solution small-angle scattering, SAXS, which has been informative in previous studies of LOV domain behavior in solution[24–28]. The data indicates that CagFbFP is a dimer with the molecular weight of ~25 kDa, whereas CagFbFPe is a higher-order oligomer with the molecular weight of ~90 kDa (Figure 2C,D). Since extensive oligomerization is not desirable for a potential fluorescent tag, we continued our work with the smaller construct CagFbFP.

Interestingly, while CagFbFP is the shortest among the currently described FbFP proteins[16,18] (Figure 1), it still has remarkable thermal stability. High-throughput thermal denaturation measurements show that CagFbFP is stable in a wide range of NaCl concentrations, although refolding is hampered at salt concentrations greater than 1 M (Figure S2). The protein unfolding temperature is also essentially unaffected by urea (Figure S2). Another commonly used denaturant, GuHCl, visibly affects CagFbFP stability, but the protein is still folded in a range of temperatures (Figure S2).

Next, we conducted unfolding, thermostability and photostability measurements using the same equipment, measurement settings, and sample concentrations and volumes as used previously by Wingen *et al.*[18,29,30] so that the data are directly comparable. These data is summarized in Table 1. Exemplary data is depicted in Figures 3, 4 and S3. In the flavin fluorescence-based melting experiments we observe two transitions with melting temperatures of T_{m1} = 67.7 ± 0.5 °C and T_{m2} = 84.7 ± 0.5 °C (Figure 3A). It is not clear whether these two transitions correspond to two different species of CagFbFP in solution or it is the same CagFbFP proteins melting in two steps. In the nanoDSF-based experiments, where we monitor only Trp/Tyr fluorescence, we observe the same melting transitions (T_{m2} = 69.4 ± 0.9 °C; T_{m3} = 83.2 ± 0.8 °C) as well as an additional melting transition at lower temperature (T_{m1} = 53.4 ± 0.7 °C, Figure 3B). The additional transition could correspond either to the melting of the chromophore-free protein fraction, or dissociation of CagFbFP dimer, both of which should not be detected in the flavin-fluorescence based melting experiments. During nanoDSF-based refolding experiments, we observe the predominant transition at around 69 °C. Overall, similarly to what was observed in high-throughput measurements (Figure S2),

around 50% of the protein is refolded as judged from fluorescence emission spectra before and after melting and refolding (Figure S3).

Next, we checked the optical properties and thermostability of CagFbFP at room (20 °C) and elevated (60 °C) temperatures (Figure 4). During prolonged incubation at 60 °C the fluorescence of CagFbFP decreased slowly with an apparent rate of 2.6% per hour (Figure 4B, Table 1). CagFbFP maintains roughly 50% of its fluorescence after 24 h incubation at 60 °C with the corresponding fluorescence emission spectrum indicative of a certain fraction of released flavin as deduced from the altered ratios of the 495 nm and 531 nm emission bands (Figure 4D).

In conclusion, the thermostability of CagFbFP is somewhat lower compared to that of the currently most thermostable FbFPs MrFbFP and YNP3FbFP[18]. Still, it should be noted that CagFbFP is the shortest currently described FbFP (Figure 1) and its longer version is much more thermostable (Figure 2B).

The photobleaching experiments show that CagFbFP has a sufficient photostability compared to other bacterial thermostable FbFPs[18] (Figure 5), with the exception of extremely photostable MrFbFP. To further evaluate the photophysical properties of CagFbFP, we determined the fluorescence quantum yield (Φ_F) and extinction coefficient (ε) of the protein. In particular, the Φ_F of CagFbFP is 0.36 (Table 1), which is higher than the one of MrFbFP and YNP3FbFP[18]. With a comparable extinction coefficient, the fluorescence brightness of CagFbFP is higher than that of the currently most thermostable FbFPs, including MrFbFP. Thus, it might be possible to engineer an FbFP protein that would have even better properties by combining the best qualities of both CagFbFP and MrFbFP.

As a next step, we aimed to elucidate the structural basis of CagFbFP thermostability. The protein crystallized extremely well, as the crystals were observed in 12 out of 96 standard crystallization screen conditions[31] (Crystal Screen 1 by Hampton Research Corp., USA) and sometimes reached the size of the crystallization drop, 600 μm, along the largest dimension. The best crystals diffracted to a high resolution, and a 1.07 Å dataset was collected from a single crystal (Table 2). The data quality is comparable to that of the best-diffracting iLOV and miniSOG crystals[14,32]. CagFbFP has a traditional LOV domain fold, and the asymmetric unit contains two LOV molecules forming an antiparallel dimer, with the hydrophobic surfaces of β-sheets at the dimerization interface (Figure 6A). Similar dimerization of LOVs has been observed previously – for example, in the structure of the plant LOV domain-containing protein ZTL[33]. Interestingly, the backbones of the residues Ala58 and Asp59 adopt two alternative conformations, and thus even further stabilization of CagFbFP might be

possible, because in other LOV/FbFP proteins there is often a proline residue in the place of Ala58.

Overall, the CagFbFP structure is similar to that of other homologous proteins (backbone RMSD to iLOV[14], for example, is less than 0.5 Å). While the direct comparison of the CagFbFP structure (with an alanine instead of the adduct-forming cysteine) and photocycling, signaling-competent LOV domain structures (possessing the adduct-forming cysteine) is not possible, it is still noteworthy that the signaling-relevant glutamine 148 clearly occupies two conformations: "buried" and "exposed" (Figure 6B). The "exposed" conformation has previously been observed only in two LOV domain structures, in the flavin-free structure of W619_1-LOV from *Pseudomonas putida*[34] and in the V48I G80R mutant of the plant blue-light photoreceptor ZEITLUPE (ZTL)[33] (Figure S4). Previously, different studies revealed that in many LOV proteins this glutamine is the key residue responsible for signal transduction; its side chain flips upon formation of the illuminated state[2,35]. The residue is also important for development of novel fluorescent proteins, as it affects the electronic structure of the flavin, and its mutation was reported to result in a shift of the absorption and fluorescence emission maxima[29,30,36].

The reason for the presence of the unusual "exposed" conformation of Gln148 in CagFbFP might be its steric interaction with Ile52, an important hydrophobic amino acid located in close proximity. The nature of the side chain at this position greatly affects the kinetics of LOV signaling[33,37–40]. Overall, in the dataset of over 6,700 LOV sequences assembled by Glantz *et al.*[6], the majority possesses a valine at this position. At the same time, ~23% of bacterial, ~10% of non-tandem protist, and smaller percentages of LOV domain proteins from other organisms, including *N. crassa* VVD, have isoleucines at this position (Table S1). In most of the available crystallographic structures, the residue at the position of Ile52 is valine. Still, there are some exceptions. Threonine is observed in *Rhodobacter sphaeroides* LOV protein[28], *Erythrobacter litoralis* monomeric histidine kinase EL346[41] and *Trichoderma reesei* ENV1[42]. In the *Brucella abortus* LOV domain, the corresponding position is occupied by a leucine, but the glutamine (Gln132) is in the buried conformation[43] (Figure S4H). In VVD, the residue is isoleucine, but it is in the g^+,*t* rotameric conformation, similarly to valine in other LOVs, with the C_δ atom removed from the glutamine[35] (Figure S4I). Finally, the only structure, where there is an isoleucine side chain in the same g^-,*t* rotameric conformation as in CagFbFP, is the structure of the V48I G80R double mutant of the plant blue-light photoreceptor ZTL[33]. There, in the dark, the glutamine occupies the "exposed" conformation, but under illumination a partially occupied "buried" conformation is

observed in one of the four protein chains in the crystallographic asymmetric unit (Figure S4C,F).

Thus, CagFbFP is an unusual FbFP protein with the signaling-relevant glutamine (Gln148) clearly in two conformations. To test whether these rotamers are able to interconvert, we conducted several molecular dynamics simulations. Three 100 ns long trajectories were started with Gln148 in the "buried" conformation, and three other trajectories were started with Gln148 in the "exposed" conformation. Analysis of the distances FMN-O4···Q148-NE2 (Figure S5) and FMN-N5···Q148-NE2 (Figure 7) shows that Gln148 preserves the hydrogen bond with FMN-O4 during the simulations but rapidly switches between the "buried" and "exposed" conformations. Such behavior has not been observed in similar simulations of iLOV, where the homologous Gln489 resided solely in the "buried" conformation[30]. Thus, on the basis of crystallographic data and molecular dynamics simulations we conclude that in CagFbFP, Gln148 interconverts rapidly between the "buried" and "exposed" conformations, and signaling might proceed via a different mechanism compared to other LOV proteins.

To check whether the same Gln148 conformations are present in the wild-type (signaling-competent) Cagg_3753 LOV domain, we introduced the mutation A85C into CagFbFP and crystallized the resulting construct. The crystals diffracted to 1.22 Å (Table 2). The data quality is comparable to that of other best-diffracting wild type LOV crystals[44,45]. Overall, the structure is identical to that of CagFbFP, with RMSD of atomic positions of 0.1 Å. Electron densities reveal the adduct-forming cysteine 85 and confirm alternative conformations of glutamine 148 (Figure 8).

To conclude, due to remarkable crystallizability and high quality of the resulting crystals, CagFbFP appears to be a promising model for ultra-high resolution structural studies of LOV and FbFP proteins and their variants, whereas its optical and thermal stability make it a good fluorescent reporter.

Methods

Cloning, protein expression and purification

The nucleotide sequences encoding residues 47-153 (CagFbFP) or 8-175 (CagFbFPe) of Cagg_3753 (Uniprot accession code B8GAY9) with C85A mutation and C-terminal 6×His tag were synthesized *de novo* (Evrogen, Russia) and introduced into the pET11 expression vector (Novagen, Merck, Germany) via *NdeI* and *BamHI* restriction sites. Wild-type Cagg_3753 LOV domain construct was generated using site-directed mutagenesis. CagFbFP, CagFbFPe and wild-type Cagg_3753 LOV domain were expressed in *Escherichia coli* strain C41 (DE3). Cells were cultured in shaking baffled flasks in ZYP-5052 auto-inducing medium[46] containing 100 mg/l ampicillin. Incubation continued for 7 hours at 37 °C. Harvested cells were disrupted in M-110P Lab Homogenizer (Microfluidics, USA) at 25000 psi in a lysis buffer containing 50 mM Tris-HCl, 300 mM NaCl, pH 8.0 with addition of 0.5 mM PMSF. The lysate was clarified by removal of cell membrane fraction by ultracentrifugation at 100000 *g* for 1 h at 4º C. The supernatant was incubated with Ni-nitrilotriacetic acid (Ni-NTA) resin (Qiagen, Germany) on a rocker for 1 hour at 4º C. The Ni-NTA resin and supernatant were loaded on a gravity flow column and washed with buffer containing 20 mM imidazole, 300 mM NaCl and 50 mM Tris-HCl, pH 8.0. The protein was eluted in a buffer containing 200 mM imidazole, 300 mM NaCl and 50 mM Tris-HCl, pH 8.0. The eluate was subjected to size-exclusion chromatography on Superdex® 200 Increase 10/300 GL column (GE Healthcare Life Sciences, USA) in a buffer containing 10 mM NaCl and 10 mM sodium phosphate, pH 8.0. Protein-containing fractions were pooled and concentrated to 30 mg/ml for crystallization.

Spectroscopic methods

High-throughput thermal stability measurements were conducted using Rotor-Gene Q real-time PCR cycler (Qiagen, Germany) with excitation at 470±10 nm and fluorescence detection at 510±5 nm. 20 µl samples with a protein concentration of approximately 1 mg/ml have been used. All other spectroscopic measurements were performed as described previously[18,29,30]. In brief, to minimize inner filter and reabsorption effects, all fluorescence measurements were carried out using purified CagFbFP samples adjusted to absorption of $A_{450nm} \leq 0.1$ (corresponding to a protein concentration of approx. 0.18 mg mL^{-1}). Absorption spectra were measured using a temperature-controlled (20 ± 1 °C) Cary-60 UV/Vis spectrophotometer (Agilent Technologies, Santa Clara, CA, USA). The flavin load of CagFbFP samples was calculated from absorption spectra of the chromophore-loaded protein. The total flavin concentration in the sample was determined from the absorption at 450 nm

using the molar extinction coefficient of free FMN ($\varepsilon_{FMN,\ 450\ nm}$ = 12200 M^{-1} cm^{-1})[47,48]. The apo protein concentration was calculated by subtracting the absorbance contribution of FMN at 280 nm ($\varepsilon_{FMN,\ 280\ nm}$ = 18000 M^{-1} cm^{-1})[47,48] from the measured absorption taking into account the theoretical extinction coefficient of the protein at 280 nm ($\varepsilon_{CagFbFP,\ 280nm}$ = 8480 M^{-1} cm^{-1}), derived from the amino acid sequence using the ProtParam web service[49] (https://web.expasy.org/protparam/). Chromophore load is then expressed as the ratio of the total flavin concentration and the apo protein concentration. Fluorescence quantum yields (Φ_F) were determined employing a QuantaMaster 40 fluorescence spectrofluorimeter (Horiba Jobin Yvon GmbH, Bensheim Germany) equipped with an integrating sphere. All measurements were performed at 20 ± 2 °C in 10 x 10 mm quartz-glass cuvettes filled with 4 mL sample. The same volume of buffer was used as reference. The extinction coefficient of CagFbFP was determined by measuring the absorption of the protein at 20 °C (chromophore bound to the protein) and after heating to 95 °C (chromophore dissociated from protein). The extinction coefficient of CagFbFP at 450 nm $\varepsilon_{CagFbFP,450nm}$ (protein bound FMN) was then determined by multiplying the ratio of the absorption measured at 20 °C and 95 °C with the extinction coefficient of free FMN in aqueous solution (12200 M^{-1} cm^{-1} at 450 nm)[47,48]. Thermal-unfolding was monitored spectrofluorimetrically by detecting FbFP-specific fluorescence using a QuantaMaster 40 spectrofluorimeter (Horiba Jobin Yvon GmbH, Bensheim Germany) at a heating rate of 1 °C per minute starting at 20 °C and ending at 100 °C. All measurements were carried out in sealed 10 x 10 mm quartz glass cuvettes filled with 3.5 mL of sample. Samples were excited at 450 nm and fluorescence emission was recorded at 495 nm. Fluorescence emission spectra were recorded at 20°C before and after thermal denaturation. Melting temperatures were determined from the first derivative of the melting curve, by fitting the smoothened data (FFT filter method, 15 points, implemented in Origin 9.0G (OriginLab Corporation, Northampton, MA, USA), to multiple Gaussian functions (Origin 9.0G, Quick Peaks Gadget). In order to determine the thermal stability of CagFbFP at 60 °C, we performed a long-term thermostability experiment. All measurements were performed using a temperature-controlled QuantaMaster 40 spectrofluorimeter. Samples (1.4 mL in sealed 10 x 4 mm quartz-glass cuvettes) were heated inside of the spectrofluorimeter to 60 °C (1°C/min ramp rate). Subsequently, the temperature was kept constant for 24 hours, during which the fluorescence emission was continuously monitored. Samples were excited at 450 nm and fluorescence emission was recorded at 495 nm. Fluorescence emission spectra were recorded before heating at 20 °C, after reaching 60 °C as well as after completion of the experiment. The unfolding/denaturation rate ($k_{denat,60\ °C}$) was

determined by considering the relative (per hour) loss of the fluorescence signal over the complete incubation time. Photobleaching kinetics were measured using a temperature-controlled Cary Eclipse spectrofluorimeter (Agilent Technologies, Santa Clara, CA, USA) thermostatted to 20 ± 2 °C. For all measurements, a LUXEON High-Power LED Rebel LXML PR01 royalblue (Philips Lumileds, San Jose, CA, USA; 1.3 mW/cm^2, Kaschner *et al.*[50]) was mounted on top of a 10×2 mm quartz-glass cuvette filled with 700 µL of sample. Samples were continuously illuminated, while the bleaching of the protein bound FMN was monitored ($\lambda_{excitation}$ = 450 nm, $\lambda_{emission}$ = 495 nm, 5 nm excitation and emission band widths, data interval 1 second) over 60 minutes. The first data point below 50% of the starting intensity was taken as bleaching half time.

NanoDSF-based thermal unfolding/refolding experiments.

Thermal unfolding and refolding experiments were performed using a Prometheus NT.Plex nanoDSF system (NanoTemper Technologies GmbH, München, Germany). Purified CagFbFP protein samples, adjusted to an absorption of $A_{450nm} \approx 1.0$ (corresponding to a concentration of approx. 1.8 mg mL^{-1}), were subjected to a linear unfolding ramp (1 °C/min, from 20 to 95 °C) followed by a refolding ramp (1 °C/min, from 95 to 20°C). The intrinsic protein fluorescence was monitored continuously (18 data points per minute) at 350 and 330 nm. Unfolding transition midpoints were determined from the first derivative of the fluorescence ratio (F350/F330) by using Quick Peaks gadget implemented in Origin 9.0G (OriginLab Corporation, Northampton, MA, USA).

Small angle X-ray scattering

The SAXS experiments were performed at the EMBL BioSAXS beamline P12 on storage ring PETRA III at the Deutsches Elektronen-Synchrotron (DESY), Hamburg, Germany[51]. Scattering curves were recorded using a photon-counting PILATUS 2M detector with a sample-detector distance 3.1 m. For all experiments, two-dimensional (2D) SAXS data were recorded as a sequential set of images (frames). A single frame exposure time was 50 ms (45 ms collection time and 5 ms detector readout time). The energy was 10 keV with a wavelength of 0.124 nm. Scattering vector ($s = 4\pi\sin\theta/\lambda$) range was 0.026–5.075 nm^{-1}. Measurements were performed using the robotic sample changer with continuous in capillary sample flow through the beamline (horizontally oriented under vacuum). Before and after the protein solutions the corresponding buffer solutions were measured. The measurements was provided in the flow of protein solution at temperature T=20 ºC. The concentrations of the protein samples were 1.0, 1.5, 2.0, 2.5, 3.0, 3.5 and 10.0 mg/ml. Data was analyzed using

ATSAS software package[52]. To estimate molecular weight of particles in solution the SAXSMoW (SAXS Molecular Weight) package was used[53]. The CagFbFP dimer model was fitted using a normal mode analysis-based approach[54].Theoretical scattering curve of CagFbFP was calculated using Pepsi-SAXS[55].

Crystallization

CagFbFP and wild type Cagg_3753 LOV domain crystals were obtained using the hanging drop vapor diffusion approach. Crystallization trials were set up using NT8 robotic system (Formulatrix, USA). The crystals were grown at 22 °C and reached the final size of 100-600 μm within few days. The drops contained 150 nL concentrated protein solution and 100 nL reservoir solution. The best crystals of CagFbFP were obtained using the precipitant 200 mM Ammonium citrate dibasic, 20% (w/v) PEG 3350. The best crystals of wild type Cagg_3753 LOV domain were obtained using the precipitant solution 25 from the Morpheus crystallization kit (Molecular Dimensions, UK) containing 0.09 M sodium nitrate, 0.09 M sodium phosphate dibasic, 0.09 M ammonium sulfate, 0.1 M imidazole and MES monohydrate (acid) buffer with pH 6.5, 20% PEG 500 MME and 10% PEG 20000. To prevent drying and for cryoprotection, the same precipitant solution with 20% glycerol was added to the crystallization drops. The crystals were harvested using micromounts and then flash-cooled and stored in liquid nitrogen.

Acquisition and treatment of diffraction data

The diffraction data were collected at 100 K at the ESRF beamline ID23-1 equipped with a PILATUS 6M-F detector. The data collection statistics are reported in Table 2. Diffraction images were processed using XDS[56]. POINTLESS[57] and AIMLESS[57] were used to merge, scale and assess the quality of the data, as well as to convert intensities to structure factor amplitudes and generate Free-R labels.

Structure determination and refinement

The CagFbFP structure was solved using molecular replacement with MOLREP[58] and a model generated using the RaptorX web server[59] as a search model. The model was refined manually using Coot[60] and REFMAC5[61] resulting in a structure with 99.5/0.5/0% of the protein residues with favored/allowed/outlying backbone conformations, respectively, as determined by Ramachandran analysis. The wild type Cagg_3753 LOV domain structure was solved using molecular replacement with MOLREP[58] and CagFbFP structure with Gln148 side chain removed as a search model, and refined manually using Coot[60] and REFMAC5[61].

Molecular dynamics simulations

The initial coordinates for the monomeric subunit were taken from the solved crystal structure of CagFbFP. Two sets of independent simulations for each conformer of Gln148 ("buried" and "exposed") were carried out. The PROPKA 3.1 program[62] was used to assign the protonation states of titratable residues on the basis of pK_a values and visual inspection. Side chains of Asn and Gln residues were inspected for possible flipping. The Amber ff99SB force-field parameters[63,64] for the protein and general Amber force field (GAFF)[65] for flavin mononucleotide (FMN) were used. The atomic charges and force field parameters for FMN were adapted from previous work[30]. The phosphate group of flavin was deprotonated with a charge of −2, which leads to a total charge of −2e for CagFbFP. The whole system was neutralized by adding 2 Na+ ions. Crystal water molecules were kept and hydrogen atoms were added by employing tleap module of AmberTools14[66]. The protein was solvated in an octahedral TIP3P[67] water box centered at the center of mass to ensure a water layer of 12 Å around the protein. The systems contained ∼25500 atoms in total, including ∼7812 TIP3P[67] water molecules. Initially, the solvent and the ions were minimized followed by the whole system minimization using 10000 steps of steepest descent followed by 3000 steps of conjugate-gradient minimization. Afterwards, the system was heated slowly from 0 to 300 K for 50 ps. In MD simulations, constant pressure periodic boundary conditions using the particle mesh Ewald (PME)[68] method were employed. The electrostatic interactions were calculated using a cutoff of 10 Å. After heating step, the systems were equilibrated for 1000 ps at 300 K. Three independent production runs each for 100 ns were carried out for both the conformers. All classical molecular dynamics (MD) simulations were performed using the Amber14 program[66]. The obtained MD simulations trajectories were visualized and analyzed with Pymol[69], VMD[70], and AmberTools 14[66]. For the hydrogen bond calculations, a geometric cutoff of 3.2 Å and 150° for distance and angle was used, respectively.

Acknowledgments

Atomic coordinates and structure factors for the reported crystal structures have been deposited in the Protein Data Bank under the accession codes 6RHF (CagFbFP) and 6RHG (Cagg_3753 LOV domain). The study was supported by the Russian Foundation for Basic Research (project № 18-34-00742). We acknowledge the Structural Biology Group of European Synchrotron Radiation Facility for granting access to the synchrotron beamlines and for assistance with data collection, and especially A.N. Popov. Simulations were performed with computing resources granted by JARA-HPC from RWTH Aachen University under projects JARA0065. We thank Dr. Thomas Drepper and Dr. Marcus Wingen for valuable discussions and for providing protocols for the fluorescence measurements.

Author contributions

V.V.N. performed the gene cloning with the help of A.Y.; V.V.N. and A.R. expressed and purified the proteins with the help of A.Y., A.D., I.M.G. and P.K.; V.V.N. and A.R. performed the Rotor-Gene Q thermostability measurements; K.-E.J. and U.K. performed the spectroscopic characterization and thermostability experiments and analysis thereof; A.V.R. collected the SAXS data; A.Y. analyzed the SAXS data; A.R. crystallized the proteins; K.K. collected the diffraction data; I.G. solved and analyzed the structures with the help of P.B., V.B. and A.M.; G.V.D., U.S. and M.D.D. performed and analyzed the molecular dynamics simulations; V.G. helped with the project coordination; I.G. designed and supervised the project, analyzed the results and wrote the paper with contributions from all other coauthors.

Supplementary information

Supplementary Table S1 and Supplementary Figures S1-S5 are available.

References

1 J. Herrou and S. Crosson, Function, structure and mechanism of bacterial photosensory LOV proteins, *Nature Reviews Microbiology*, 2011, **9**, 713–723.

2 B. D. Zoltowski and K. H. Gardner, Tripping the light fantastic: blue-light photoreceptors as examples of environmentally modulated protein-protein interactions, *Biochemistry*, 2011, **50**, 4–16.

3 K. S. Conrad, C. C. Manahan and B. R. Crane, Photochemistry of flavoprotein light sensors, *Nature Chemical Biology*, 2014, **10**, 801–809.

4 A. Losi, C. Mandalari and W. Gärtner, The Evolution and Functional Role of Flavin-based Prokaryotic Photoreceptors, *Photochem. Photobiol.*, 2015, **91**, 1021–1031.

5 J. T. Henry and S. Crosson, Ligand-Binding PAS Domains in a Genomic, Cellular, and Structural Context, *Annual Review of Microbiology*, 2011, **65**, 261–286.

6 S. T. Glantz, E. J. Carpenter, M. Melkonian, K. H. Gardner, E. S. Boyden, G. K.-S. Wong and B. Y. Chow, Functional and topological diversity of LOV domain photoreceptors, *Proc. Natl. Acad. Sci. U.S.A.*, 2016, **113**, E1442-1451.

7 A. Losi and W. Gärtner, Solving Blue Light Riddles: New Lessons from Flavin-binding LOV Photoreceptors, *Photochem. Photobiol.*, 2017, **93**, 141–158.

8 T. Fettweiss, K. Röllen, J. Granzin, O. Reiners, S. Endres, T. Drepper, D. Willbold, K.-E. Jaeger, R. Batra-Safferling and U. Krauss, Mechanistic basis of the fast dark recovery of the short LOV protein DsLOV from Dinoroseobacter shibae, *Biochemistry*, , DOI:10.1021/acs.biochem.8b00645.

9 M. Salomon, J. M. Christie, E. Knieb, U. Lempert and W. R. Briggs, Photochemical and mutational analysis of the FMN-binding domains of the plant blue light receptor, phototropin, *Biochemistry*, 2000, **39**, 9401–9410.

10 A. M. Buckley, J. Petersen, A. J. Roe, G. R. Douce and J. M. Christie, LOV-based reporters for fluorescence imaging, *Curr Opin Chem Biol*, 2015, **27**, 39–45.

11 A. Mukherjee and C. M. Schroeder, Flavin-based fluorescent proteins: emerging paradigms in biological imaging, *Curr. Opin. Biotechnol.*, 2015, **31**, 16–23.

12 T. Drepper, T. Eggert, F. Circolone, A. Heck, U. Krauss, J.-K. Guterl, M. Wendorff, A. Losi, W. Gärtner and K.-E. Jaeger, Reporter proteins for in vivo fluorescence without oxygen, *Nat. Biotechnol.*, 2007, **25**, 443–445.

13 S. Chapman, C. Faulkner, E. Kaiserli, C. Garcia-Mata, E. I. Savenkov, A. G. Roberts, K. J. Oparka and J. M. Christie, The photoreversible fluorescent protein iLOV outperforms GFP as a reporter of plant virus infection, *Proc. Natl. Acad. Sci. U.S.A.*, 2008, **105**, 20038–20043.

14 J. M. Christie, K. Hitomi, A. S. Arvai, K. A. Hartfield, M. Mettlen, A. J. Pratt, J. A. Tainer and E. D. Getzoff, Structural tuning of the fluorescent protein iLOV for improved photostability, *J. Biol. Chem.*, 2012, **287**, 22295–22304.

15 A. Mukherjee, J. Walker, K. B. Weyant and C. M. Schroeder, Characterization of flavin-based fluorescent proteins: an emerging class of fluorescent reporters, *PLoS ONE*, 2013, **8**, e64753.

16 X. Song, Y. Wang, Z. Shu, J. Hong, T. Li and L. Yao, Engineering a more thermostable blue light photo receptor Bacillus subtilis YtvA LOV domain by a computer aided rational design method, *PLoS Comput. Biol.*, 2013, **9**, e1003129.

17 A. Mukherjee, K. B. Weyant, U. Agrawal, J. Walker, I. K. O. Cann and C. M. Schroeder, Engineering and characterization of new LOV-based fluorescent proteins from Chlamydomonas reinhardtii and Vaucheria frigida, *ACS Synth Biol*, 2015, **4**, 371–377.

18 M. Wingen, K.-E. Jaeger, T. Gensch and T. Drepper, Novel Thermostable Flavin-binding Fluorescent Proteins from Thermophilic Organisms, *Photochem. Photobiol.*, 2017, **93**, 849–856.

19 S. Hanada, A. Hiraishi, K. Shimada and K. Matsuura, Chloroflexus aggregans sp. nov., a filamentous phototrophic bacterium which forms dense cell aggregates by active gliding movement, *Int. J. Syst. Bacteriol.*, 1995, **45**, 676–681.
20 R. D. Finn, T. K. Attwood, P. C. Babbitt, A. Bateman, P. Bork, A. J. Bridge, H.-Y. Chang, Z. Dosztányi, S. El-Gebali, M. Fraser, J. Gough, D. Haft, G. L. Holliday, H. Huang, X. Huang, I. Letunic, R. Lopez, S. Lu, A. Marchler-Bauer, H. Mi, J. Mistry, D. A. Natale, M. Necci, G. Nuka, C. A. Orengo, Y. Park, S. Pesseat, D. Piovesan, S. C. Potter, N. D. Rawlings, N. Redaschi, L. Richardson, C. Rivoire, A. Sangrador-Vegas, C. Sigrist, I. Sillitoe, B. Smithers, S. Squizzato, G. Sutton, N. Thanki, P. D. Thomas, S. C. E. Tosatto, C. H. Wu, I. Xenarios, L.-S. Yeh, S.-Y. Young and A. L. Mitchell, InterPro in 2017—beyond protein family and domain annotations, *Nucleic Acids Res*, 2017, **45**, D190–D199.
21 C. P. Zschiedrich, V. Keidel and H. Szurmant, Molecular Mechanisms of Two-Component Signal Transduction, *Journal of Molecular Biology*, 2016, **428**, 3752–3775.
22 M. P. Bhate, K. S. Molnar, M. Goulian and W. F. DeGrado, Signal Transduction in Histidine Kinases: Insights from New Structures, *Structure*, 2015, **23**, 981–994.
23 I. Gushchin and V. Gordeliy, Transmembrane Signal Transduction in Two-Component Systems: Piston, Scissoring, or Helical Rotation?, *BioEssays*, 2018, **40**, 1700197.
24 K. Röllen, J. Granzin, R. Batra-Safferling and A. M. Stadler, Small-angle X-ray scattering study of the kinetics of light-dark transition in a LOV protein, *PLoS One*, , DOI:10.1371/journal.pone.0200746.
25 J. S. Lamb, B. D. Zoltowski, S. A. Pabit, B. R. Crane and L. Pollack, Time-resolved dimerization of a PAS-LOV protein measured with photocoupled small angle x-ray scattering, *J Am Chem Soc*, 2008, **130**, 12226–12227.
26 U. Heintz and I. Schlichting, Blue light-induced LOV domain dimerization enhances the affinity of Aureochrome 1a for its target DNA sequence, *Elife*, 2016, **5**, e11860.
27 A. Banerjee, E. Herman, T. Kottke and L.-O. Essen, Structure of a Native-like Aureochrome 1a LOV Domain Dimer from Phaeodactylum tricornutum, *Structure*, 2016, **24**, 171–178.
28 K. S. Conrad, A. M. Bilwes and B. R. Crane, Light-induced subunit dissociation by a light-oxygen-voltage domain photoreceptor from Rhodobacter sphaeroides, *Biochemistry*, 2013, **52**, 378–391.
29 M. Wingen, J. Potzkei, S. Endres, G. Casini, C. Rupprecht, C. Fahlke, U. Krauss, K.-E. Jaeger, T. Drepper and T. Gensch, The photophysics of LOV-based fluorescent proteins--new tools for cell biology, *Photochem. Photobiol. Sci.*, 2014, **13**, 875–883.
30 M. D. Davari, B. Kopka, M. Wingen, M. Bocola, T. Drepper, K.-E. Jaeger, U. Schwaneberg and U. Krauss, Photophysics of the LOV-Based Fluorescent Protein Variant iLOV-Q489K Determined by Simulation and Experiment, *J Phys Chem B*, 2016, **120**, 3344–3352.
31 J. Jancarik and S.-H. Kim, Sparse matrix sampling: a screening method for crystallization of proteins, *J Appl Cryst, J Appl Crystallogr*, 1991, **24**, 409–411.
32 J. Torra, C. Lafaye, L. Signor, S. Aumonier, C. Flors, X. Shu, S. Nonell, G. Gotthard and A. Royant, Tailing miniSOG: structural bases of the complex photophysics of a flavin-binding singlet oxygen photosensitizing protein, *Sci Rep*, 2019, **9**, 2428.
33 A. Pudasaini, J. S. Shim, Y. H. Song, H. Shi, T. Kiba, D. E. Somers, T. Imaizumi and B. D. Zoltowski, Kinetics of the LOV domain of ZEITLUPE determine its circadian function in Arabidopsis, *Elife*, , DOI:10.7554/eLife.21646.
34 V. Arinkin, J. Granzin, K. Röllen, U. Krauss, K.-E. Jaeger, D. Willbold and R. Batra-Safferling, Structure of a LOV protein in apo-state and implications for construction of LOV-based optical tools, *Sci Rep*, 2017, **7**, 42971.

35 B. D. Zoltowski, C. Schwerdtfeger, J. Widom, J. J. Loros, A. M. Bilwes, J. C. Dunlap and B. R. Crane, Conformational switching in the fungal light sensor Vivid, *Science*, 2007, **316**, 1054–1057.
36 M. G. Khrenova, A. V. Nemukhin and T. Domratcheva, Theoretical Characterization of the Flavin-Based Fluorescent Protein iLOV and its Q489K Mutant, *J Phys Chem B*, 2015, **119**, 5176–5183.
37 B. D. Zoltowski, B. Vaccaro and B. R. Crane, Mechanism-based tuning of a LOV domain photoreceptor, *Nat. Chem. Biol.*, 2009, **5**, 827–834.
38 J. Lokhandwala, R. I. Silverman Y de la Vega, H. C. Hopkins, C. W. Britton, A. Rodriguez-Iglesias, R. Bogomolni, M. Schmoll and B. D. Zoltowski, A Native Threonine Coordinates Ordered Water to Tune Light-Oxygen-Voltage (LOV) Domain Photocycle Kinetics and Osmotic Stress Signaling in Trichoderma reesei ENVOY, *J. Biol. Chem.*, 2016, **291**, 14839–14850.
39 F. Kawano, Y. Aono, H. Suzuki and M. Sato, Fluorescence imaging-based high-throughput screening of fast- and slow-cycling LOV proteins, *PLoS ONE*, 2013, **8**, e82693.
40 J. M. Christie, S. B. Corchnoy, T. E. Swartz, M. Hokenson, I.-S. Han, W. R. Briggs and R. A. Bogomolni, Steric interactions stabilize the signaling state of the LOV2 domain of phototropin 1, *Biochemistry*, 2007, **46**, 9310–9319.
41 G. Rivera-Cancel, W. Ko, D. R. Tomchick, F. Correa and K. H. Gardner, Full-length structure of a monomeric histidine kinase reveals basis for sensory regulation, *PNAS*, 2014, **111**, 17839–17844.
42 J. Lokhandwala, H. C. Hopkins, A. Rodriguez-Iglesias, C. Dattenböck, M. Schmoll and B. D. Zoltowski, Structural biochemistry of a fungal LOV domain photoreceptor reveals an evolutionarily conserved pathway integrating light and oxidative stress, *Structure*, 2015, **23**, 116–125.
43 J. Rinaldi, M. Gallo, S. Klinke, G. Paris, H. R. Bonomi, R. A. Bogomolni, D. O. Cicero and F. A. Goldbaum, The β-scaffold of the LOV domain of the Brucella light-activated histidine kinase is a key element for signal transduction, *J. Mol. Biol.*, 2012, **420**, 112–127.
44 A. Möglich and K. Moffat, Structural basis for light-dependent signaling in the dimeric LOV domain of the photosensor YtvA, *J. Mol. Biol.*, 2007, **373**, 112–126.
45 A. S. Halavaty and K. Moffat, N- and C-terminal flanking regions modulate light-induced signal transduction in the LOV2 domain of the blue light sensor phototropin 1 from Avena sativa, *Biochemistry*, 2007, **46**, 14001–14009.
46 F. W. Studier, Protein production by auto-induction in high-density shaking cultures, *Protein Expression and Purification*, 2005, **41**, 207–234.
47 L. G. Whitby, A new method for preparing flavin-adenine dinucleotide, *Biochem. J.*, 1953, **54**, 437–442.
48 M. Kataoka, S. Shimizu and H. Yamada, Purification and characterization of a novel FMN-dependent enzyme. Membrane-bound L-(+)-pantoyl lactone dehydrogenase from Nocardia asteroides, *Eur. J. Biochem.*, 1992, **204**, 799–806.
49 M. R. Wilkins, E. Gasteiger, A. Bairoch, J. C. Sanchez, K. L. Williams, R. D. Appel and D. F. Hochstrasser, Protein identification and analysis tools in the ExPASy server, *Methods Mol. Biol.*, 1999, **112**, 531–552.
50 M. Kaschner, A. Loeschcke, J. Krause, B. Q. Minh, A. Heck, S. Endres, V. Svensson, A. Wirtz, A. von Haeseler, K.-E. Jaeger, T. Drepper and U. Krauss, Discovery of the first light-dependent protochlorophyllide oxidoreductase in anoxygenic phototrophic bacteria, *Mol. Microbiol.*, 2014, **93**, 1066–1078.
51 C. E. Blanchet, A. Spilotros, F. Schwemmer, M. A. Graewert, A. Kikhney, C. M. Jeffries, D. Franke, D. Mark, R. Zengerle, F. Cipriani, S. Fiedler, M. Roessle and D. I. Svergun, Versatile sample environments and automation for biological solution X-ray scattering

experiments at the P12 beamline (PETRA III, DESY), *J Appl Crystallogr*, 2015, **48**, 431–443.
52 D. Franke, M. V. Petoukhov, P. V. Konarev, A. Panjkovich, A. Tuukkanen, H. D. T. Mertens, A. G. Kikhney, N. R. Hajizadeh, J. M. Franklin, C. M. Jeffries and D. I. Svergun, ATSAS 2.8: a comprehensive data analysis suite for small-angle scattering from macromolecular solutions, *J Appl Crystallogr*, 2017, **50**, 1212–1225.
53 H. Fischer, O. Neto, M. De, H. B. Napolitano, I. Polikarpov and A. F. Craievich, Determination of the molecular weight of proteins in solution from a single small-angle X-ray scattering measurement on a relative scale, *J Appl Cryst, J Appl Crystallogr*, 2010, **43**, 101–109.
54 A. Hoffmann and S. Grudinin, NOLB: Nonlinear Rigid Block Normal-Mode Analysis Method, *J Chem Theory Comput*, 2017, **13**, 2123–2134.
55 S. Grudinin, M. Garkavenko and A. Kazennov, Pepsi-SAXS: an adaptive method for rapid and accurate computation of small-angle X-ray scattering profiles, *Acta Crystallogr D Struct Biol*, 2017, **73**, 449–464.
56 W. Kabsch, XDS, *Acta Crystallographica Section D Biological Crystallography*, 2010, **66**, 125–132.
57 P. Evans, Scaling and assessment of data quality, *Acta Crystallographica Section D Biological Crystallography*, 2005, **62**, 72–82.
58 A. Vagin and A. Teplyakov, Molecular replacement with MOLREP, *Acta Crystallographica Section D Biological Crystallography*, 2009, **66**, 22–25.
59 M. Källberg, H. Wang, S. Wang, J. Peng, Z. Wang, H. Lu and J. Xu, Template-based protein structure modeling using the RaptorX web server, *Nat Protoc*, 2012, **7**, 1511–1522.
60 P. Emsley, B. Lohkamp, W. G. Scott and K. Cowtan, Features and development of Coot, *Acta Crystallogr. D Biol. Crystallogr.*, 2010, **66**, 486–501.
61 G. N. Murshudov, P. Skubák, A. A. Lebedev, N. S. Pannu, R. A. Steiner, R. A. Nicholls, M. D. Winn, F. Long and A. A. Vagin, REFMAC5 for the refinement of macromolecular crystal structures, *Acta Crystallographica Section D Biological Crystallography*, 2011, **67**, 355–367.
62 M. H. M. Olsson, C. R. Søndergaard, M. Rostkowski and J. H. Jensen, PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical pKa Predictions, *J Chem Theory Comput*, 2011, **7**, 525–537.
63 V. Hornak, R. Abel, A. Okur, B. Strockbine, A. Roitberg and C. Simmerling, Comparison of multiple Amber force fields and development of improved protein backbone parameters, *Proteins: Structure, Function, and Bioinformatics*, 2006, **65**, 712–725.
64 W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell and P. A. Kollman, A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules, *J. Am. Chem. Soc.*, 1995, **117**, 5179–5197.
65 J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman and D. A. Case, Development and testing of a general amber force field, *J. Comput. Chem.*, 2004, **25**, 1157–1174.
66 D. A. Case, V. Babin, J. Berryman, R. M. Betz, Q. Cai, D. S. Cerutti, C. Iii, T. E, T. A. Darden, R. E. Duke, H. Gohlke, A. W. Goetz, S. Gusarov, N. Homeyer, P. Janowski, J. Kaus, I. Kolossváry, A. Kovalenko, T. S. Lee, S. LeGrand, T. Luchko, R. Luo, B. Madej, K. M. Merz, F. Paesani, D. R. Roe, A. Roitberg, C. Sagui, R. Salomon-Ferrer, G. Seabra, C. L. Simmerling, W. Smith, J. Swails, R. C. Walker, J. Wang, R. M. Wolf, X. Wu and P. A. Kollman, *Amber 14*, University of California, San Francisco, 2014.
67 W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Klein, Comparison of simple potential functions for simulating liquid water, *J. Chem. Phys.*, 1983, **79**, 926–935.

68 U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee and L. G. Pedersen, A smooth particle mesh Ewald method, *The Journal of Chemical Physics*, 1995, **103**, 8577–8593.
69 W. L. DeLano, The PyMOL molecular graphics system.
70 W. Humphrey, A. Dalke and K. Schulten, VMD Visual molecular dynamics, *Journal of Molecular Graphics*, 1996, **14**, 33–38.
71 P. A. Karplus and K. Diederichs, Linking Crystallographic Model and Data Quality, *Science*, 2012, **336**, 1030–1033.
72 M. A. Larkin, G. Blackshields, N. P. Brown, R. Chenna, P. A. McGettigan, H. McWilliam, F. Valentin, I. M. Wallace, A. Wilm, R. Lopez, J. D. Thompson, T. J. Gibson and D. G. Higgins, Clustal W and Clustal X version 2.0, *Bioinformatics*, 2007, **23**, 2947–2948.
73 A. M. Waterhouse, J. B. Procter, D. M. A. Martin, M. Clamp and G. J. Barton, Jalview Version 2--a multiple sequence alignment editor and analysis workbench, *Bioinformatics*, 2009, **25**, 1189–1191.
74 D. Liebschner, P. V. Afonine, N. W. Moriarty, B. K. Poon, O. V. Sobolev, T. C. Terwilliger and P. D. Adams, Polder maps: improving OMIT maps by excluding bulk solvent, *Acta Crystallogr D Struct Biol*, 2017, **73**, 148–157.

Table 1. Spectral properties and thermal stability of CagFbFP. Errors correspond to the standard deviation of the mean derived from three independent measurements on three independently expressed and purified samples.

absorption λ_{max} (nm)	447
emission λ_{max} (nm)	497
ε (M^{-1} cm^{-1})	15317 ± 387
chromophore load (%)	56 ± 7
Φ_F	0.36 ± 0.01
brightness (M^{-1} cm^{-1})	5534 ± 141
$t_{bl,\ 50\%}$ (min)	4.23 ± 0.8
Tm_{Fluor} [Tm_1/Tm_2] (°C)[a]	67.7 ± 0.5 / 84.7 ± 0.5
$Tm_{nanoDSF}$ [Tm_1/Tm_2/Tm_3] (°C)[b]	53.4 ± 0.7 / 69.4 ± 0.9 / 83.2 ± 0.8
$k_{denat,\ 60\ °C}$ (% h^{-1})	2.57 ± 0.4

[a, b:] melting data displaying two or three melting transitions, respectively

Table 2. Crystallographic data collection and refinement statistics. High-resolution cutoff is applied according to Karplus and Diederichs[71].

Data collection	CagFbFP	WT Cagg_3753 LOV
Space group	$P2_12_12$	$P2_12_12$
Cell dimensions		
a, *b*, *c* (Å)	53.93, 110.00, 38.92	53.79, 110.27, 39.14
α, β, γ (°)	90, 90, 90	90, 90, 90
Wavelength (Å)	0.972	0.976
Resolution (Å)	48.43-1.07 (1.09-1.07)*	48.34-1.22 (1.24-1.22)*
R_{merge} (%)	6.0 (183.4)	8.6 (129.0)
$I/\sigma I$	9.6 (0.8)	9.4 (1.3)
$CC_{1/2}$ (%)	99.8 (29.3)	99.7 (54.1)
Completeness (%)	99.6 (97.8)	97.4 (95.6)
Unique reflections	102522 (4922)	68182 (3250)
Multiplicity	5.3 (5.0)	6.4 (6.6)
Refinement		
Resolution (Å)	48.43-1.07	48.34-1.22
No. reflections	102449	64766
R_{work}/ R_{free} (%)	12.4/14.9	13.9/17.6
No. atoms		
Protein	1800	1744
FMN	62	62
Water and glycerol	331	290
B-factors		
Protein	17.3	18.6
FMN	14.5	14.4
Water and glycerol	36.5	35.9
R.m.s deviations		
Bond lengths (Å)	0.021	0.021
Bond angles (°)	2.1	2.1

* Data for the highest resolution shell is shown in parenthesis

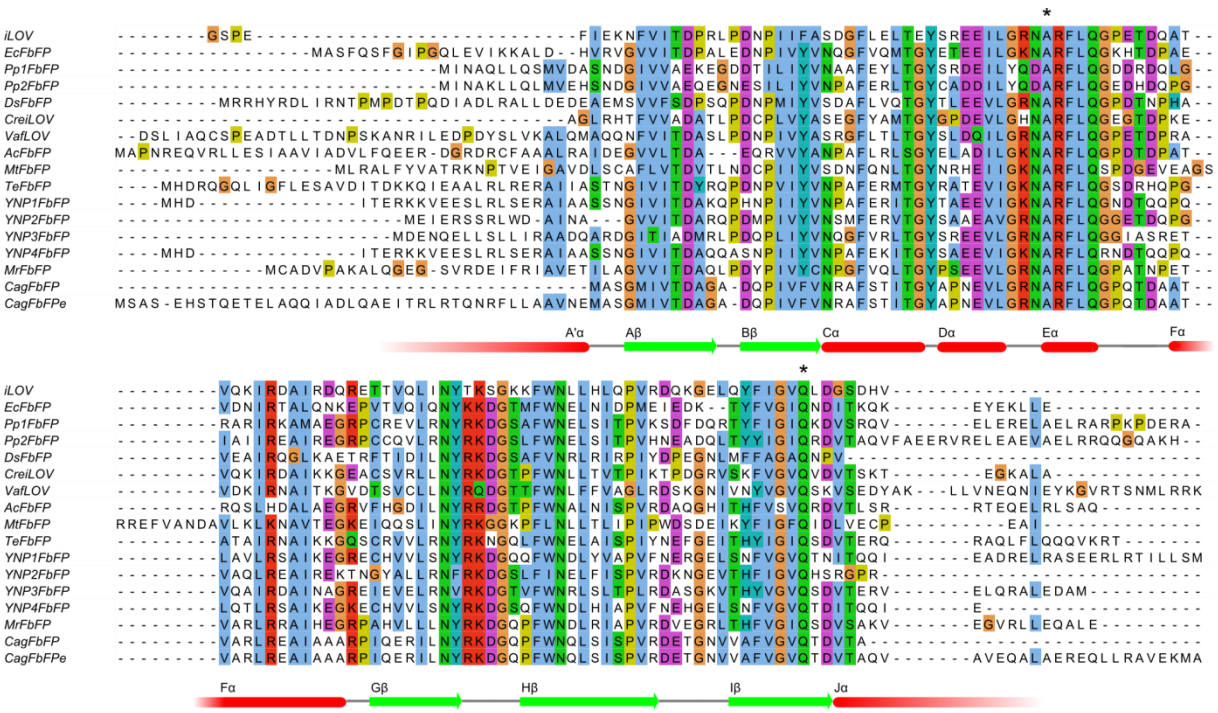


Figure 1. Sequence alignment of LOV-based flavin-binding fluorescent proteins (FbFPs): iLOV[13], EcFbFP[12], Pp1FbFP[29], Pp2FbFP (formerly PpFbFP[12]), DsFbFP[29], CreiLOV and VafLOV[17], eight different FbFPs from thermophilic microorganisms and the new *Chloroflexus aggregans* LOV-based FbFPs (CagFbFP and CagFbFPe). Conserved amino acids alanine (Ala85 in CagFbFP), which replaces the functional cysteine, and glutamine (Gln148 in CagFbFP) are marked using asterisks. The sequences were aligned using Clustal[72] and visualized using Jalview[73].

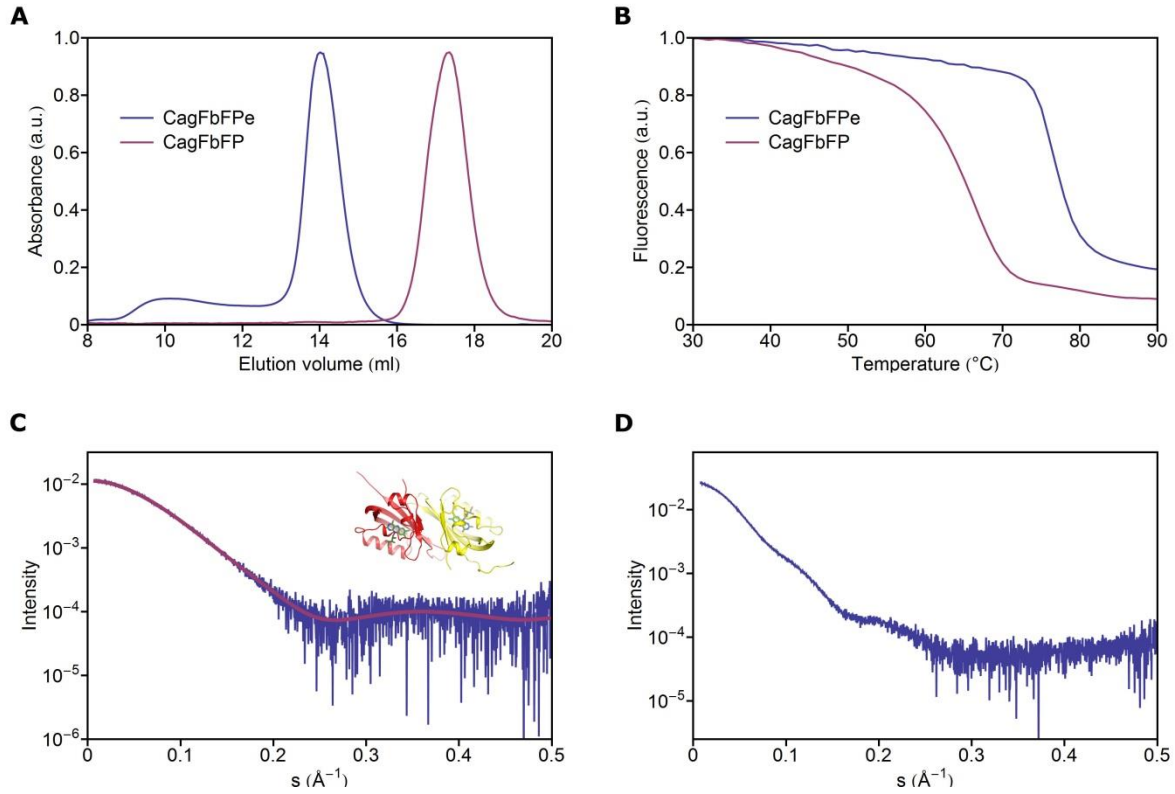


Figure 2. Comparison of the short (CagFbFP) and extended (CagFbFPe) variants of the *C. aggregans* LOV-based fluorescent protein. (A) Size-exclusion chromatography profiles of CagFbFP and CagFbFPe. (B) Melting curves of CagFbFP and CagFbFPe determined by measuring the dependence of flavin fluorescence on temperature using Rotor-Gene Q real-time PCR cycler. (C) SAXS data (blue) and model fit (magenta) of CagFbFP in solution. The inset shows the dimer model fitted using a normal mode analysis-based approach[51,52], with χ^2 of 1.064. (D) SAXS data of CagFbFPe in solution. The estimated molecular weight of CagFbFPe is ~90 kDa.

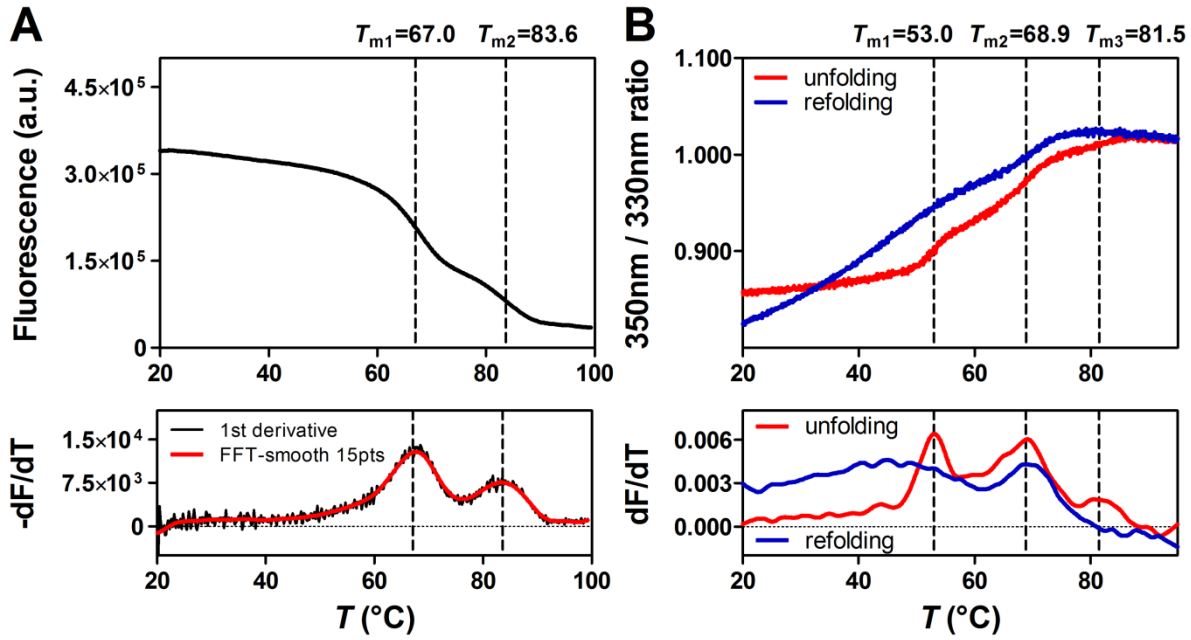


Figure 3. Exemplary melting curves of CagFbFP determined based on flavin fluorescence (A) and the fluorescence of aromatic amino acids (B). (A) The upper panel shows the decrease in the FbFP-specific fluorescence during heating from 20 °C to 100 °C (ramp rate 1 °C/min). The lower panel depicts the first derivative of the data. Melting temperatures are marked by dashed lines in all panels. The measurements were performed using a temperature controlled QuantaMaster 40 fluorescence spectrofluorimeter. (B) The change in the fluorescence of aromatic amino acids (350 nm/330 nm ratio) was monitored using a Prometheus NT.Plex nanoDSF System. The upper panel depicts the melting (red line) and refolding (blue line) behavior of CagFbFP during the respective heating (20 to 95 °C, ramp rate 1 °C/min) and cooling (95 to 20 °C, ramp rate 1 °C/min) steps. The lower panel shows the first derivative of the corresponding data. For clarity only the melting temperatures are marked by dashed lines. Only one data set out of three independent measurements on three independently prepared samples is shown.

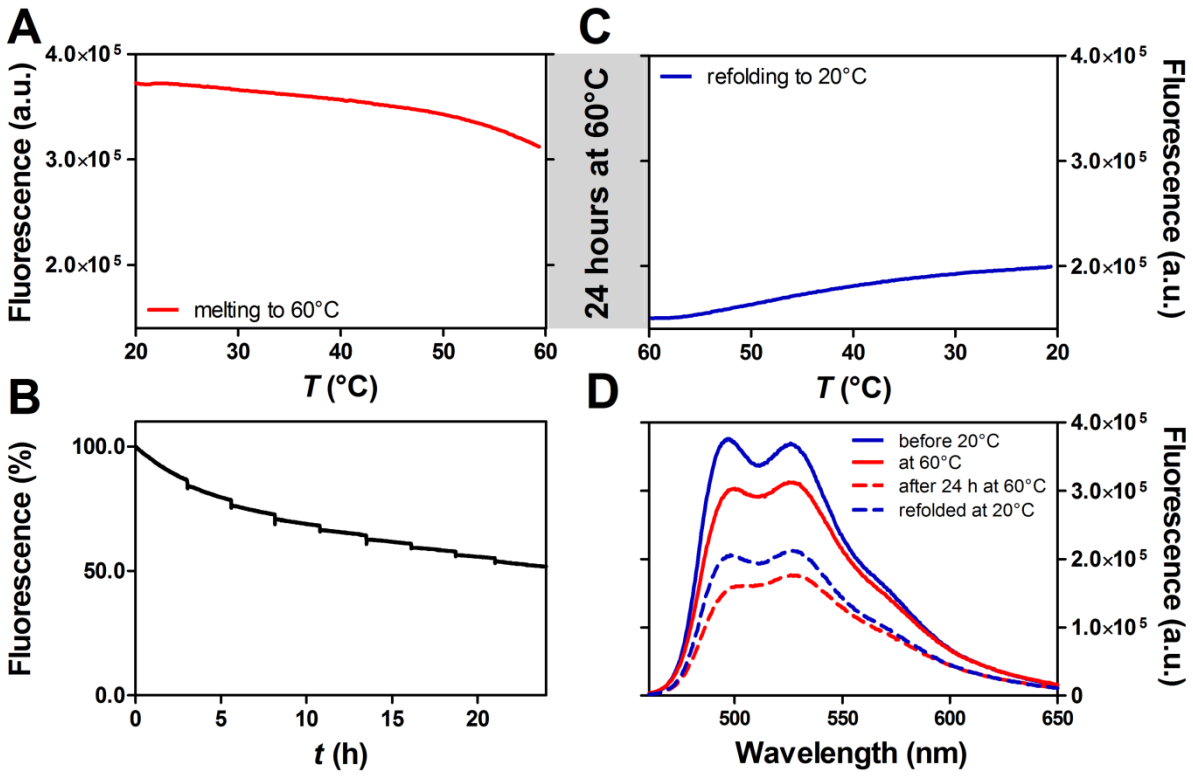


Figure 4. Fluorescence stability of CagFbFP during long-term incubation at 60 °C. Only one data set out of three independent measurements on three independently prepared samples is shown. (A) CagFbFP fluorescence during heating from 20 °C to 60 °C (ramp: 1 °C/min). (B) CagFbFP fluorescence during continuous incubation at 60°C. A decrease of fluorescence is observed proceeding with a rate of $k_{denat,60°C}$ = 2.57 ± 0.4 % per hour. (C) Recovery of CagFbFP fluorescence as the sample was cooled to 20 °C (ramp: 1 °C/min). Limited refolding of the protein is observed. (D) Emission spectra of CagFbFP before heating (solid blue line), upon heating to 60 °C (solid red line), upon 24 h incubation at 60 °C (dashed red line) and upon cooling down to 20 °C (dashed blue line).

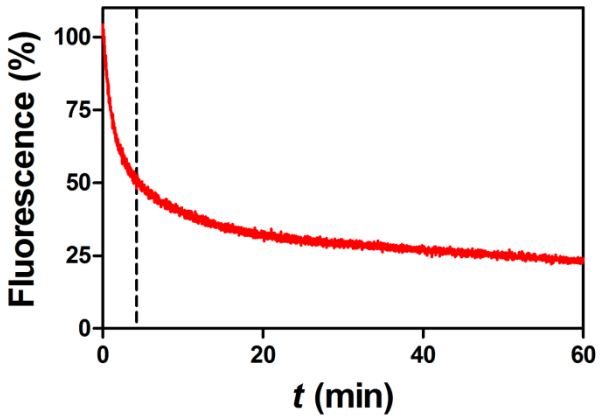


Figure 5. Exemplary photobleaching kinetics of CagFbFP. Photobleaching was monitored continuously by exciting the sample at 450 nm and recording fluorescence emission at 495 nm (1 data point per second), while the sample was continuously illuminated using a high-power LED mounted on top of the cuvette. During illumination with blue-light, the fluorescence intensity decreases over time. The first data point below 50% of the initial fluorescence was taken as photobleaching half time ($t_{bl,\ 50\%}$, marked by vertical dashed line).

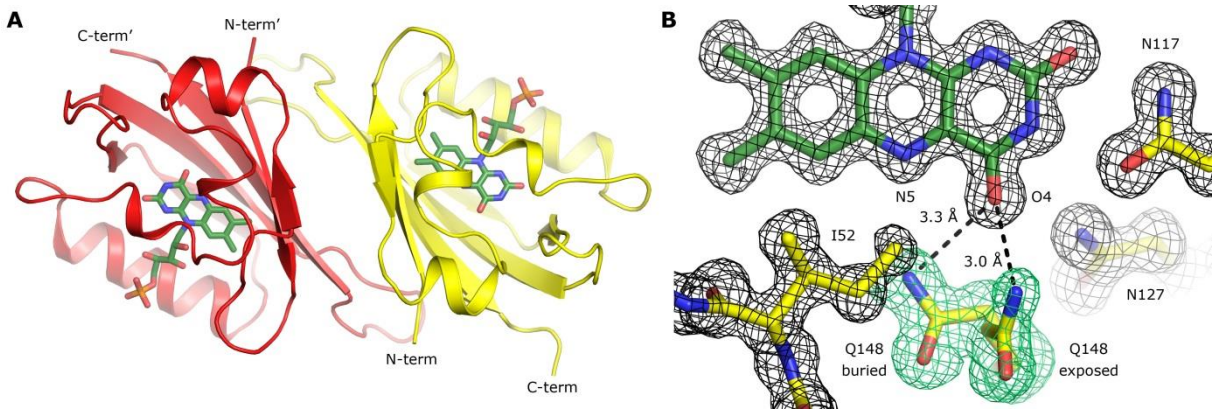


Figure 6. Structure of CagFbFP. (A) Overall structure of CagFbFP dimer. Cofactor FMN is shown in green. (B) Electron densities around FMN and Gln148. Gln148 occupies two conformations: "buried" and "exposed". Weighted $2F_o$-F_c map at the level of 1σ is shown in black, polder[74] (improved omit) map around Gln148 at the level of 3σ is shown in green. Hydrogen bonds between Gln148 and FMN's O4 are shown using the dashed lines.

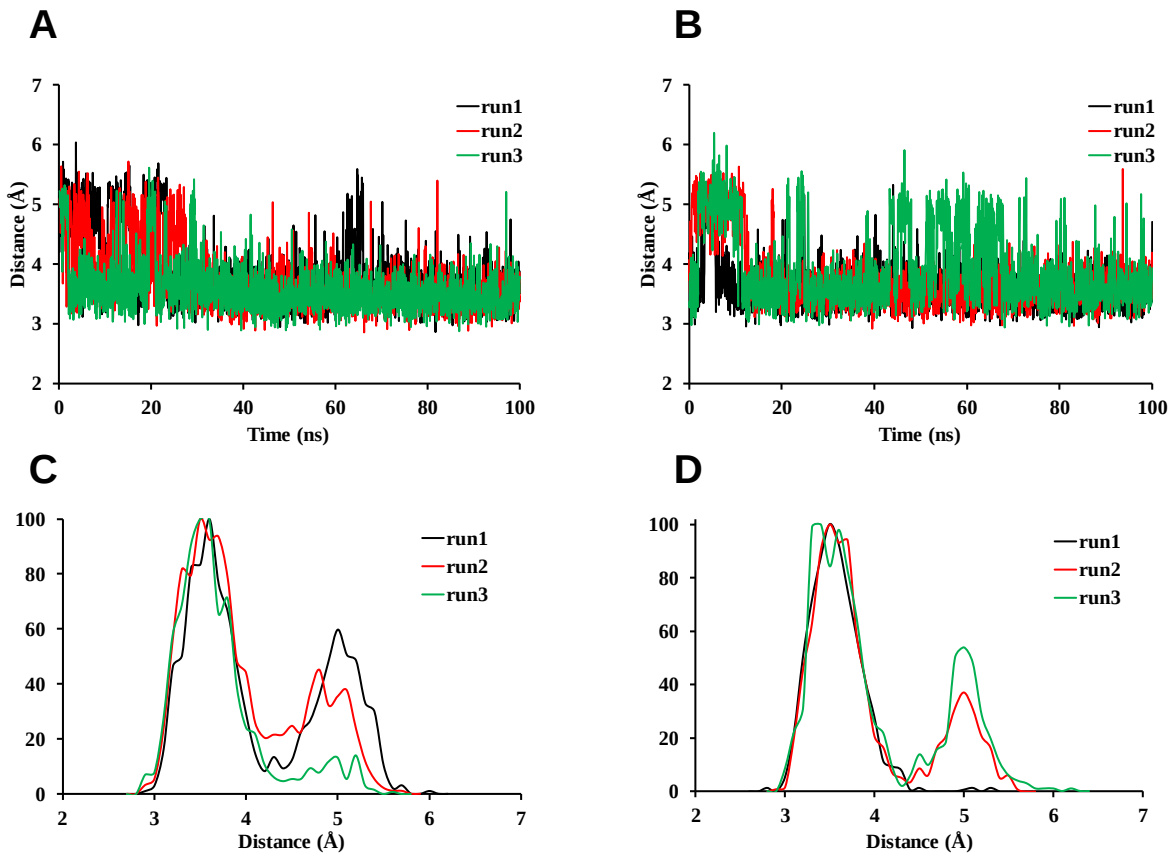


Figure 7. Switching of the Gln148 side chain of CagFbFP in molecular dynamics simulations. (A) FMN-N5 – Q148-NE2 distance as a function of time starting from the "exposed" Gln148 conformer. (B) FMN-N5 and Q148-NE2 distance as a function of time starting from the "buried" conformer. (C) Distance distribution curve of the interatomic distances between FMN-N5 and Q148-NE2 calculated over the MD trajectories starting from the "exposed" conformer. (D) Distance distribution curve of the interatomic distances between FMN-N5 and Q148-NE2 calculated over the MD trajectories starting from the "buried" conformer.

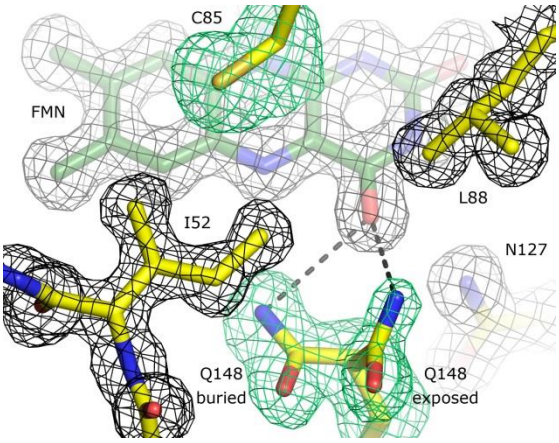
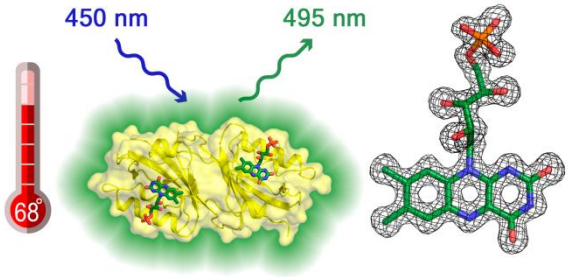


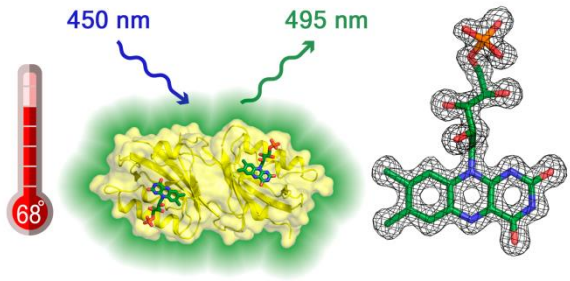
Figure 8. Structure of the FMN-binding pocket in the wild-type Cagg_3753 LOV



domain. FMN is shown in green. Gln148 occupies two conformations: "buried" and "exposed". Weighted $2F_o$-F_c map at the level of 1σ is shown in black, polder[74] (improved omit) map around Cys85 and Gln148 side chains at the level of 3σ is shown in green. Hydrogen bonds between Gln148 and FMN's O4 are shown using the dashed lines.

Thermostable flavin-based fluorescent protein from *Chloroflexus aggregans*: a framework for ultra-high resolution structural studies

Graphical abstract



Statement of novelty

A new thermostable fluorescent protein is shown to be a promising model for ultra-high resolution structural studies of LOV domains and for application as a fluorescent reporter.

Thermostable flavin-based fluorescent protein from *Chloroflexus aggregans*: a framework for ultra-high resolution structural studies

Vera V. Nazarenko, Alina Remeeva, Anna Yudenko, Kirill Kovalev, Anton Dubenko, Ivan M. Goncharov, Pavel Kuzmichev, Andrey V. Rogachev, Pavel Buslaev, Valentin Borshchevskiy, Alexey Mishin, Gaurao V. Dhoke, Ulrich Schwaneberg, Mehdi D. Davari, Karl-Erich Jaeger, Ulrich Krauss, Valentin Gordeliy, Ivan Gushchin*

* E-mail for correspondence: ivan.gushchin@phystech.edu

Supplementary information

Supplementary Table S1

Supplementary Figures S1-S5

Table S1. Frequencies of aminoacids at the position of CagFbFP's Ile52. The most frequent occurrences are highlighted. The data are based on the dataset by Glantz *et al.*, 2016.

LOV group	**Total**	**Val**	**Ile**	**Leu**	**Thr**	**Ser**	**Cys**	**Ala**	**Others**
Archaea	156	16%	3.2%	1.9%	71.2%	4.5%	-	3.2%	-
Bacteria	1146	45.1%	22.9%	12%	12.5%	2.3%	1.5%	2.4%	1.4%
Fungi	645	50.2%	3.1%	6%	2.6%	0.3%	33.2%	3.7%	0.9%
Protists single	373	45.3%	9.7%	3.8%	16.6%	1.6%	19.6%	1.6%	1.9%
Protists tandem LOV1	169	81.7%	3.6%	1.2%	5.9%	-	2.4%	5.3%	-
Protists tandem LOV2	169	71.6%	4.1%	1.2%	1.2%	1.2%	20.7%	-	-
Land plants single	2689	92.9%	3.2%	1.2%	0.9%	0.04%	0.2%	0.9%	0.7%
Land plants tandem LOV1	1587	98.6%	0.2%	0.7%	-	-	-	0.5%	-
Land plants tandem LOV2	1587	99.9%	0.1%	-	-	-	0.1%	-	-

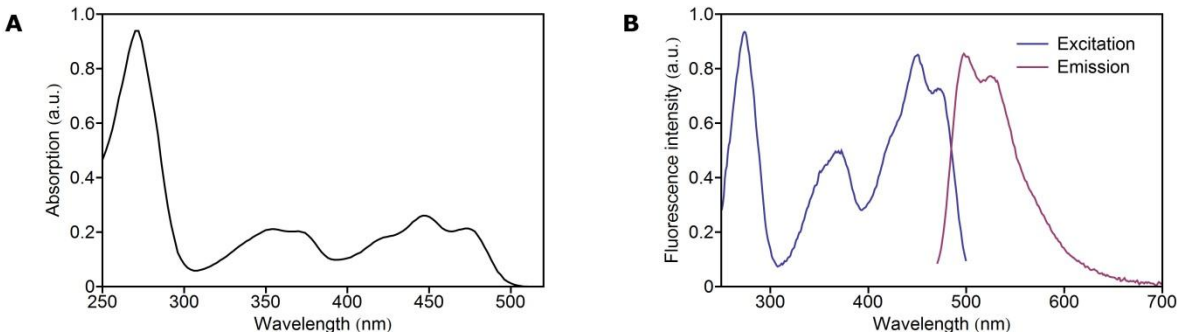


Figure S1. CagFbFP absorption (A) and fluorescence excitation and fluorescence emission (B) spectra.

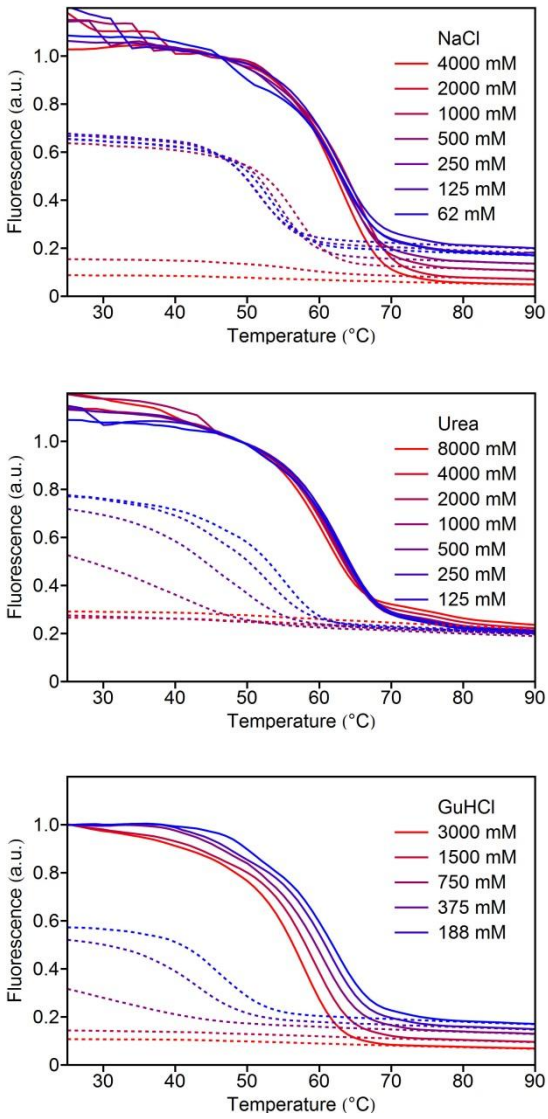


Figure S2. Temperature-induced unfolding (solid lines) and refolding (dotted lines) of CagFbFP in different solutions. Initial parts of the unfolding curves are noisy due to automatic fluorescence signal gain adjustments. To obtain the samples, purified CagFbFP was mixed with concentrated stock solutions. The measurements were performed using a Rotor-Gene Q real-time PCR cycler.

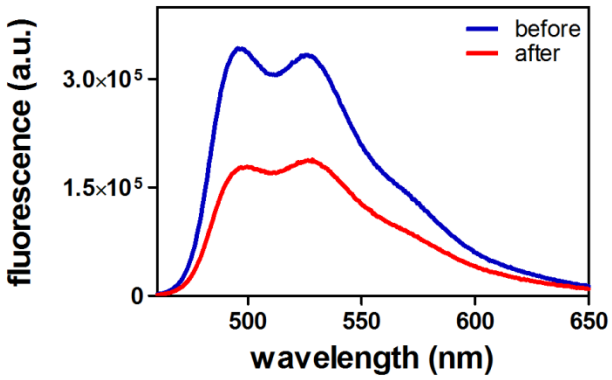


Figure S3. Exemplary fluorescence emission spectra of CagFbFP before and after melting and refolding. Samples were equilibrated for 5 minutes at 20 °C, an emission spectrum was recorded (before, blue line) and the sample was subsequently heated to 100 °C (1 °C/min ramp rate). After reaching the final temperature, the sample was cooled to 20 °C and a second emission spectrum was recorded (after, red line). In terms of fluorescence intensity at 495 nm approx. 50% of the initial value is recovered after melting. Only one data set out of three independent measurements on three independently prepared samples is shown.

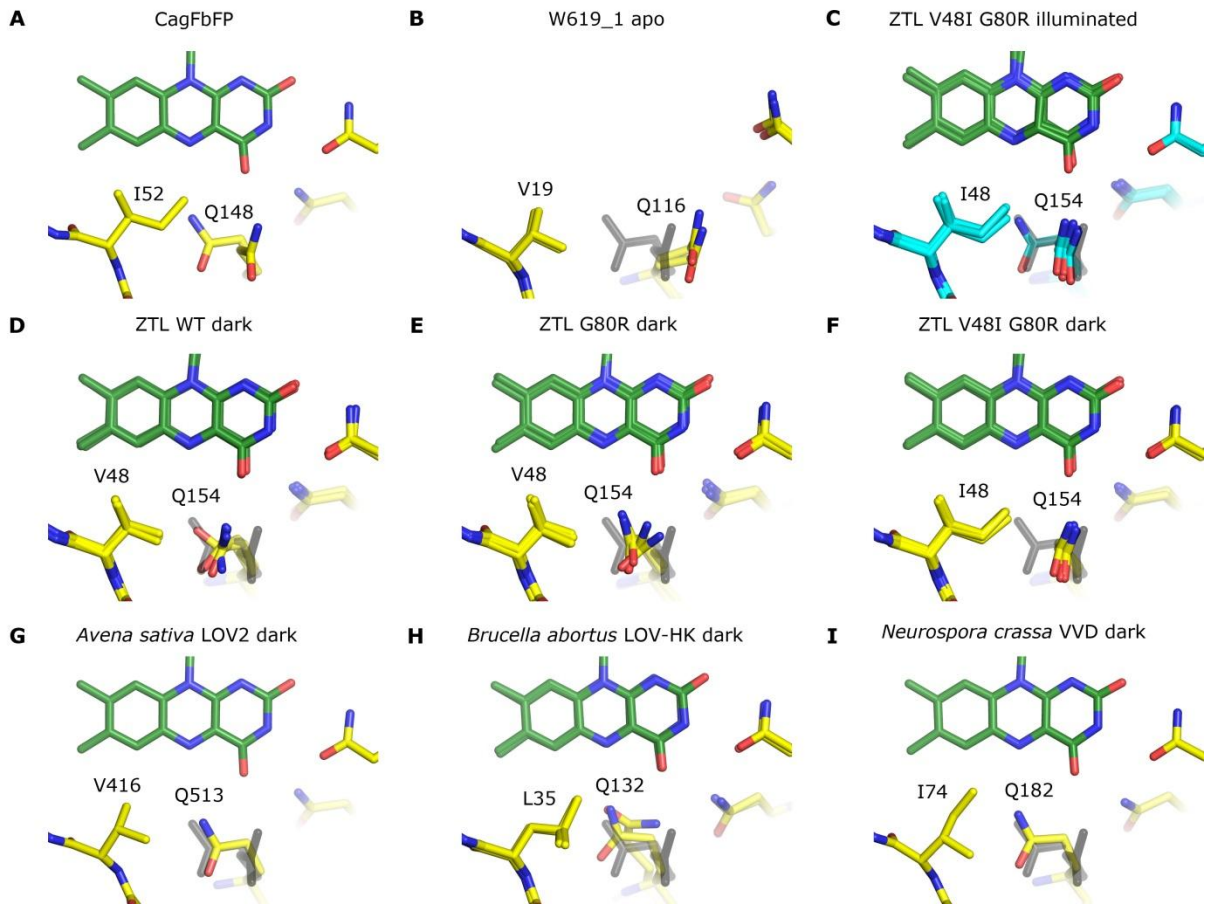


Figure S4. "Buried" and "exposed" conformations of the signaling glutamine amino acid in different crystal structures. CagFbFP Q148 positions are shown in gray for reference. (A) Structure of CagFbFP. (B) Structure of cofactor-less LOV protein W619_1 (PDB ID 5LUV). (C) Structure of V48I G80R mutant of ZTL under illumination (PDB ID 5SVW). The "buried" conformation is partially occupied in the chain C and is not observed in the chains A, B and D. (D) Structure of WT ZTL in the darkness (PDB ID 5SVG). (E) Structure of G80R mutant of ZTL in the darkness (PDB ID 5SVU). (F) Structure of V48I G80R mutant of ZTL in the darkness (PDB ID 5SVV). (G) Structure of the *Avena sativa* LOV2 in the darkness (PDB ID 2V0U). (H) Structure of the *Brucella abortus* LOV-HK in the darkness (PDB ID 3T50). (I) Structure of the *Neurospora crassa* VVD in the darkness (PDB ID 2PD7).

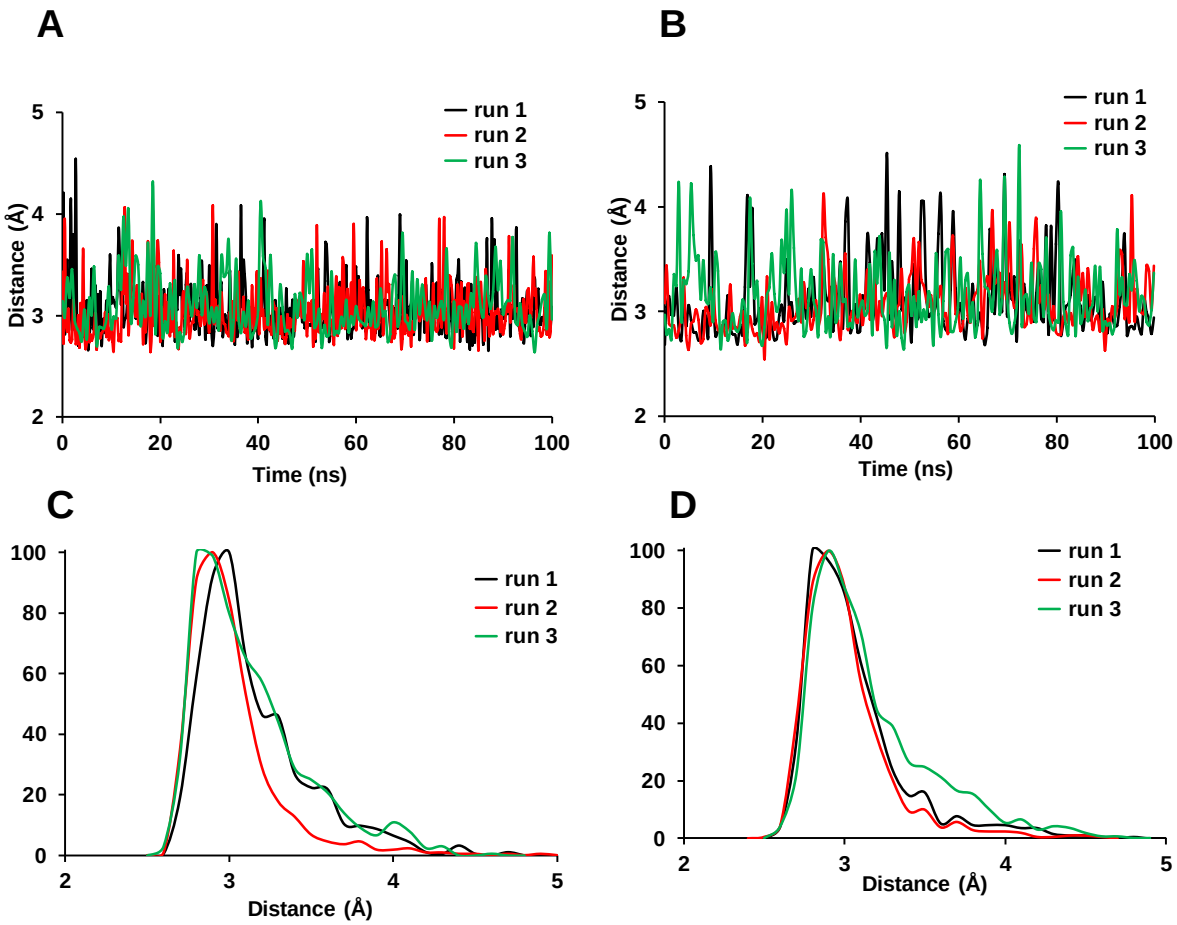


Figure S5. Distances between the Gln148 side chain and the O4 atom of FMN in molecular dynamics simulations of CagFbFP. (A) FMN-O4 and Q148-NE2 distance as a function of time starting from the "exposed" conformer. (B) FMN-O4 and Q148-NE2 distance as a function of time starting from the "buried" conformer. (C) Distance distribution curve of the interatomic distances between FMN-O4 and Q148-NE2 (in Å) calculated over the MD trajectories starting from the "exposed" conformer. (D) Distance distribution curve of the interatomic distances between FMN-O4 and Q148-NE2 (in Å) calculated over the MD trajectories starting from the "buried" conformer.