Genetically encoded non-canonical amino acids reveal asynchronous dark reversion of chromophore, backbone and side-chains in EL222

Aditya S. Chaudhari,[1,2] Aditi Chatterjee,[1,2] Catarina A.O. Domingos,[1,3] Prokopis C. Andrikopoulos,[1] Yingliang Liu,[1] Inger Andersson,[1,4] Bohdan Schneider,[1] Víctor A. Lórenz-Fonfría,[5,*] and Gustavo Fuertes[1,*]

[1] Institute of Biotechnology CAS v.v.i, Prumyslova 595, 25250 Vestec, Czech Republic
[2] Faculty of Science, Charles University, Albertov 2038, 128 00 Prague, Czech Republic
[3] Escola Superior de Tecnologia do Barreiro, Instituto Politécnico de Setúbal, Rua Américo da Silva Marinho, s/n, 2839-001 Lavradio, Portugal
[4] Department of Cell and Molecular Biology, Uppsala University, Box 596, S-751 24 Uppsala, Sweden
[5] Institute of Molecular Science, Universitat de València, Carrer del Catedrático José Beltrán Martínez 2, 46980 Paterna, Spain.

*Corresponding authors. Email: victor.lorenz@uv.es, gustavo.fuertes@ibt.cas.cz

ABSTRACT.

Photoreceptors containing the light-oxygen-voltage (LOV) domain elicit biological responses upon excitation of their flavin mononucleotide (FMN) chromophore by blue light. The mechanism and kinetics of dark-state recovery are not well understood. Here we incorporated the non-canonical amino acid p-cyanophenylalanine (CNF) by genetic code expansion technology at forty-five positions of the bacterial transcription factor EL222. Screening of light-induced changes in infrared (IR) absorption frequency, electric field and hydration of the nitrile groups identified residues CNF31 and CNF35 as reporters of monomer/oligomer and caged/decaged equilibria, respectively. Time-resolved multi-probe UV/Visible and IR spectroscopy experiments of the lit-to-dark transition revealed four dynamical events. Predominantly, rearrangements around the A'α helix interface (CNF31 and CNF35) precede FMN-cysteinyl adduct scission, folding of α-helices (amide bands), and relaxation of residue CNF151. This study illustrates the importance of characterizing all parts of a protein and suggests a key role for the N-terminal A'α extension of the LOV domain in controlling EL222 photocycle length.

Significance.

The kinetics of fold switching between non-illuminated and blue-light-irradiated states in the transcription factor EL222 is important for understanding the signal transduction mechanism of LOV photoreceptors. Here we combine two native probes, the FMN chromophore (absorption bands in the UV/Visible region) and the protein backbone (amide bands in the infrared region), with genetically encoded cyano (C≡N)-containing phenylalanine residues as infrared reporters of protein microenvironments. EL222 structural dynamics is more complex than expected if using a single type of probe. Local changes around residues 31 and 35 precede FMN-protein adduct rupture, which in turn precedes the global protein conformational relaxation. Our findings point the way forward to obtaining comprehensive descriptions of kinetic transitions in LOV and other photosensors.

Introduction.

The conformation and dynamics of photoreceptor proteins are orchestrated by light, which regulates changes at the secondary, tertiary and sometimes quaternary levels of structure (*1*). In the photocontrolled transcription factor EL222 from *Erythrobacter litoralis*, the well-studied dark state is characterized by a monomeric "closed" conformation featuring extensive interactions between the light-oxygen-voltage (LOV) and helix-turn-helix (HTH) domains (*2*). Although high-resolution structural information of the lit state(s) of EL222 is missing, it is known that blue-light excitation of the flavin mononucleotide (FMN) chromophore embedded in the LOV domain leads to uncaging of the associated HTH module, oligomerization and subsequent DNA binding and gene expression (*2-6*). UV/Vis spectroscopy has been instrumental in monitoring the photoconversion of FMN (**Fig. 1** and **Fig. S1**). Non-irradiated LOV proteins absorb maximally at ~450 nm while their irradiated counterparts peak at ~390 nm (**Fig. 1B** left panel). This frequency shift is caused by photon absorption by the oxidized FMN cofactor, which triggers its partial reduction to a semiquinone radical and adduct formation with a conserved cysteine residue (*7*). FMN-Cys covalent bond formation takes place in the nanosecond to microsecond time scale (*5, 8*). The rupture of such an adduct state, sometimes referred to as A_{390} species (*5*), is much slower, typically returning to the resting state in seconds to hours, depending on the particular LOV domain (*9*). The dark recovery kinetics of isolated LOV domain suggests synchronous relaxation of chromophore and protein chain (*10*), but the behavior of multi-domain LOV-containing proteins is largely unknown.

Vibrational spectroscopies are *a priori* ideally suited to monitor protein-specific changes due to their intrinsic sensitivity to chemical bonds and inherently high time resolution, down to the picoseconds time range in the case of infrared experiments (*5, 11*). Light-induced difference infrared (IR) spectra of photosensitive proteins, where the signal from the dark state is subtracted from its lit state, are particularly informative (**Fig. 1B** bottom panel) (*12*). Analysis of the protein amide I band, mostly arising from coupled backbone C=O stretching vibrations (*13, 14*), has revealed secondary structure changes in other LOV photoreceptors, often interpreted as unfolding of the α-helices flanking the core LOV domain (N-terminal A'α and C-terminal Jα) (*15, 16*). In spite of its utility as a reporter for global structure changes in proteins (*13, 14*), the amide I vibration generally lacks spatial detail. Another difficulty is that even IR difference spectra of proteins are strongly congested, with multiple vibrations falling in common frequency regions, making it challenging to ascribe spectral changes to particular residues or vibrations.

Most studies to date have focused on "canonical" photoreceptors composed of the 20 standard amino acids (**Fig. 1A** left panel). As an alternative, chemical groups with absorption frequencies in the range from ~1800 cm^{-1} to ~2500 cm^{-1}, known as the "transparent window" because most natural proteins do not absorb there (**Fig. 1B** grey area), make the assignment straightforward while providing valuable site-specific information (*17, 18*). These extrinsic non-native moieties, which allow researchers to spectrally single out individual residues, can be incorporated into proteins using a variety of methods (*19*). Genetic code expansion (GCE) technology is arguably the most versatile of such methods as it allows the introduction of non-canonical amino acids (ncAA) at virtually any desired position along the protein sequence (*20*). An engineered aminoacyl-tRNA synthetase recognizes a particular ncAA and attaches it to an orthogonal transfer RNA (tRNA), this one most often including the CUA anticodon

(complementary to the amber stop codon, UAG), as illustrated in **Fig. 1A** (right panel). NcAAs can be utilized as chemical handles to introduce reporter groups, or as "spectator" vibrational probes that inform non-intrusively on local changes (*17, 18, 21*). Additionally, the system can be designed so that the included ncAA tunes protein functions and helps elucidating reaction mechanisms (*22*). To date, the three most popular ncAA used as IR reporters have been 4-cyanophenylalanine (CNF) (*23*), 4-azidophenylalanine (*24*), and cyanocysteine (*25*). Thiocyanates, which have the smallest transition dipole strength of the three, are generated by chemical cyanylation of cysteine residues. Azides are potentially unstable due to their chemical and photochemical reactivity (*26*), and their complex absorption profile makes interpretation challenging (*27*). For those reasons we chose CNF, a genetically encodable ncAA featuring a cyano (C≡N) group attached to a phenyl ring at the *para* position (**Fig. 1A** right), as our infrared reporter of local EL222 environments. Being appended to a side-chain, the vibrations of the reporter group are expected to report on local changes in the tertiary and quaternary protein structure, which might differ from global secondary structure changes experienced by the protein and sensed by amide backbone vibrations (*25, 28*). Due to its local nature, a sufficiently large number of tagged positions is imperative for the reporter group to render a complete picture of protein structural changes. However, previous IR studies on photoreceptors decorated with unnatural amino acids have focused mainly on a small number of residues and mainly proximal to the chromophore site (*23-25, 28-30*), with its associated limitations.

Here, we exploit the potential of merging protein engineering tools with spectroscopy by tagging in EL222 up to 45 different residue positions with the unnatural amino acid CNF. We characterized the spectral properties of CNF mutants and mapped them on the 1D (sequence) and 3D (global fold) structural levels of EL222. After a thorough screening phase to identify both intrusive and spectrally silent CNF probes, we choose the most promising EL222-CNF protein variants for kinetic studies by time-resolved UV/Vis and IR spectroscopies. We focused on the less understood transition from the "lit" state (or mixture of states) present under continuous illumination, to the "dark" state of EL222 repopulated in the absence of light. In total, four kinetic events were identified and interpreted in molecular terms. Thus, the joint use of spectral signals from the chromophore, the protein backbone and genetically encoded side-chain infrared probes enabled a multi-probe and multi-site characterization of non-equilibrium protein structural dynamics in EL222, a strategy that may benefit other LOV proteins and photosensory receptors in general.

Results

Preparation of EL222 variants and quality control.

Mutant selection.

Ideally, the IR reporters should cover as much of the EL222 sequence as possible while being minimally perturbative. CNF is chemically similar to the canonical amino acids phenylalanine and tyrosine, which can be envisioned as the natural choices for replacement. However, the four native phenylalanines and four native tyrosines of EL222 are all located in the photosensitive LOV domain and are therefore insufficient for a complete understanding of EL222 structural dynamics. Thus, we looked for other replacement candidates distributed in the protein structure. We selected residues both in the FMN binding pocket (**Fig. S2A**), to track changes in the active site, and along the LOV/HTH interface (**Fig. S2B**), to identify and characterize changes between the putative "decaged" (absence of LOV/HTH

120 interactions) and "caged" (presence of LOV/HTH interactions) states of EL222. In total, we chose a set of
121 45 residue positions along the EL222 sequence. From the N-terminus to the C-terminus we have (**Table**
122 **S1**): (i) one residue (A18) in the disordered N-terminal domain, (ii) 3 residues in the A'α helix, (iii) 27
123 residues in the LOV domain, (iv) 5 residues in the linker domain (including the bridging Jα helix), and (v)
124 9 residues in the HTH domain. Such an extensive labeling corresponds, on average, to one local probe
125 inserted every six residues.

126 *Mutant production.*

127 For amber suppression in *E. coli* we used an orthogonal translation system (OTS) comprised of
128 an orthogonal aminoacyl-tRNA synthetase/tRNA pair derived from tyrosyl tRNA synthetase of the
129 archaea *Methanocaldococcus jannaschii* (*Mj*CNFRS) and its cognate tRNA (**Fig. 1A** and **Table S2**). The
130 modified suppressor tRNA incorporates the 5'-CUA-3' anticodon ($tRNA_{CUA}$) that is efficiently recognized
131 by the modified synthetase. The enzyme is also able to attach phenylalanine derivatives (CNF and
132 others) to the suppressor $tRNA_{CUA}$, thus allowing decoding UAG codons during protein translation (*31,*
133 *32*). All proteins engineered using this strategy (see sequences in **Table S2**) were purified to > 95%. Mass
134 spectrometry analyses confirmed the successful incorporation of CNF in all 45 single-mutant proteins
135 (**Table S3**). Although the fidelity of MjCNFRS is known to be high, mis-incorporation of phenylalanine is
136 still possible. Mass spectrometry confirmed that mis-incorporation of Phe was only significant in two
137 mutants (W89CNF, I94CNF), where the phenylalanine-carrying EL222 was the most abundant species
138 detected. In these cases, we were able to minimize the presence of Phe by re-expression of these
139 constructs in minimal media (**Table S3**). In another mutant (A42CNF) we only found evidence of
140 truncated products, which cannot be attributed to the UAG codon being recognized as STOP since the
141 purification tag is found at the C-terminus, but to post-translational protein instability enhanced by the
142 introduced mutation, consistently with the known tendency of EL222 to undergo proteolytic
143 degradation under certain conditions (*2*).

144 While for double-CNF mutants the amount and purity of protein was still sufficient for
145 subsequent studies, both the yield and protein translation fidelity of triple-CNF mutants were
146 unsatisfactory low. This a known issue of nonsense suppression methods and the reason why the large
147 majority of studies limit themselves to typically one or exceptionally two target positions for ncAA
148 replacement. We turned to the *E. coli* strain C321.ΔA.exp(DE3) as the expression host, a genomically
149 recoded organism that supports multiple instances of UAG codon suppression with high efficiency, at
150 least in model proteins (*33*). Gratifyingly, we could purify a few milligrams of high-quality EL222 (**Table**
151 **S3**) and subject it to further experiments.

152 *Initial screening of mutants.*

153 Much alike site-directed mutagenesis, the introduction of an unnatural side-chain may give rise
154 to non-native interactions that can alter the structure, thermodynamics and kinetics of the system
155 under study (*34*). Altered mutants need not be discarded since they can provide valuable mechanistic
156 insights, but their interpretation must be done with extreme care. To ascertain the perturbative effect of
157 the introduced CNF residues, we initially monitored two parameters: the level of incorporation of FMN,
158 used to discard gross structural perturbations, and the oligomeric state of EL222 in the dark, to discard
159 constitutive protein-protein interactions in the absence of illumination. Further screening of CNF

mutants criteria based on quantifying their degree of similarity with respect to wt EL222 (vide infra) allowed us to improve their classification as intrusive/non-intrusive.

Most of the 44 full-length mutants of EL222 were able to incorporate FMN stoichiometrically (one mole of FMN per mole of protein). However, four mutants did not bind FMN at all (**Table S3**): V44CNF, S46CNF, C78CNF, and N110CNF. The mutations C78CNF and N110CNF involve residues participating in FMN binding (**Fig. S2A**). Indeed, both mutations were originally intended as negative controls: C78 is the critical residue that makes a covalent adduct with the C4a atom of FMN upon blue-light illumination, and N110 makes two hydrogen-bonds with N3 and C2=O2 of FMN that might be critical for the binding of this cofactor (*2*). Additionally, the bulky CNF side-chain, compared to asparagine, would reduce the space available to FMN in the binding pocket in the N110CNF mutant. Although V44 and S46 are not involved directly in FMN binding, the mutation of these small residues by CNF may introduce detrimental steric clashes, probably leading to long-range structure changes that may reach the binding site of FMN and prevent its incorporation. Out of the 40 EL222 mutants able to incorporate FMN, in four cases (L55CNF, N120CNF, V122CNF and V124CNF) we found a significant drop (less than 50%) in FMN loading. Thus, for them, we expect sample heterogeneity in the form of a mixture of apo-EL222 and holo-EL222.

Concerning the oligomeric state of holo-EL222 wt, it is a monomer in the dark but readily oligomerizes upon light stimulation (*3, 4, 6*). We verified the oligomeric state of wt EL222 by size-exclusion chromatography (SEC) and found that its elution volume, related to the particle size, was in agreement with the expected molecular weight of a monomer. Similarly, nearly all mutants were found to exist as monomers in the dark, with similar elution volumes as wt EL222 (**Table S3**). Three notable exceptions were S140CNF, L123CNF and Q141CNF. The elution volume of S140CNF was significantly smaller than the elution volume of wt EL222, suggesting a multimeric organization in the dark. The latter two showed elution volumes in between those expected for monomeric and dimeric assemblies. This behavior is akin to those previously reported for L123K and S140Y mutations, which resulted in the assembly of oligomers (*2, 6*). Hence, our results suggest that the three mentioned mutants (L123CNF, S140CNF and Q141CNF) may partially or fully adopt in the dark a structure resembling the conformation competent for DNA binding adopted by wt EL222 after illumination, i.e., resembling the lit state.

Steady-state spectroscopy.

EL222 samples were measured in the dark as well as under continuous blue-light illumination, achieving a photostationary state. We focused on three spectral regions. The first is the 320-550 nm (UV/Visible) region, which is due to absorption bands of the FMN chromophore ($S_0 \rightarrow S_1$ and $S_0 \rightarrow S_2$ electronic transitions, **Fig. 1B** top left panel). The second is the 1750 to 1500 cm^{-1} mid-IR region which includes absorption bands from both the protein backbone (amide I and II), FMN (C=O and C=C stretching vibrations among others) and water (**Fig. 1B** top right panel). The third region includes the band around 2230 cm^{-1} from the selectively introduced nitrile moieties. Peaks in the latter region are absent in wt EL222 (compare **Figs. 1** and **2**).

UV/Visible absorption region of FMN.

We first compared the UV/Vis spectra of wt EL222 and CNF mutants to identify those mutations with a preserved FMN spectrum and, by extension, with protein-chromophore interactions similar to those in the wt (**Fig. S3**). We characterized the spectral similarity in the 400-550 nm region of the

201 different mutants with respect to wt by a linear correlation analysis (**Fig. S3A**), which we complemented
202 by determining shifts in maximum absorbance wavelength ($\Delta\lambda_{max}$) with respect to wt (**Fig. S3B**). With
203 few exceptions, all variants showed a high spectral similarity (correlation coefficient above 0.9) and
204 $\Delta\lambda_{max}$ less than 2 nm, sustaining a preserved environment of the FMN chromophore in the dark.
205 However, three mutants showed clear indications of an altered FMN environment (L55CNF, N120CNF,
206 and Q141CNF), with a correlation coefficient below 0.7 and $\Delta\lambda_{max}$ larger than 4 nm.

207 *IR absorption region of native protein and FMN.*

208 The main band in this region is the amide I band (1700 to 1600 cm^{-1}), primary arising from
209 peptide backbone carbonyl C=O stretching modes, known by its secondary structure sensitivity (*35*). The
210 nearby amide II band (1600 to 1500 cm^{-1}) is due to C-N stretching and N-H bending modes, which can be
211 used to report on solvent accessibility of the protein backbone via hydrogen/deuterium exchange
212 experiments (*36*). The two carbonyl stretching modes of flavins (C2=O2 and C4=O2, see **Fig. S2A**) also
213 show up in the 1750 to 1650 cm^{-1} spectral region (*8, 37, 38*). Absolute FTIR spectra in the 1750-1500 cm^{-}
214 ^{1} region of dark-adapted and lit-adapted wt EL222 in H_2O are hardly distinguishable (**Fig. 1B** top right). In
215 contrast, lit-minus-dark difference spectra show significant changes between both (**Fig. 1B** bottom
216 right). The positive/negative features at 1727 cm^{-1} (+)/1714 cm^{-1} (-) and 1681 cm^{-1} (+)/1700 cm^{-1} (-) can
217 be attributed to the C4=O4 and C2=O2 stretches of FMN, respectively (*8, 39, 40*). The positive-negative-
218 positive feature at 1551 cm^{-1}(+)/1534 cm^{-1} (-)/1516 cm^{-1} (+) is probably due to changes in the protein
219 amide II band, but potential contributions from side-chains (e.g. tyrosine ring vibration) or FMN modes
220 (particularly the C4A-C10 ring stretching vibration, see numbering in **Fig. S2A**) cannot be ruled out (*5, 16,*
221 *41*). Even with a clear assignment of the above bands to amide II vibrations, their interpretation in
222 structural terms would be difficult as a result of the complex dependence of amide II frequency on
223 secondary structure (42). The negative bands (bleaches) at 1662 cm^{-1} (-) and 1620 cm^{-1} (-) and the
224 positive bands at 1681 cm^{-1} (+) and 1642 cm^{-1} (+) cm^{-1} could be assigned to the secondary structure-
225 sensitive amide I vibrations from the protein backbone. However, the O-H scissoring mode of liquid H_2O
226 has a strong background absorbance in the amide I region, potentially distorting the intensity of bands
227 in this region when working in solution as in the present case. The background absorbance from water
228 in the amide I region can be dramatically reduced by dissolving the protein into deuterated water (D_2O),
229 which also induces the H/D exchange of labile hydrogen atoms in the protein and causes bands from
230 vibrations involving these hydrogens to shift significantly. Absolute and difference IR spectra of wt EL222
231 in D_2O are shown in **Fig. S1A**. Although a definitive assignment of all observed bands is out of the scope
232 of this study, some interpretations in the amide I' (I in H_2O) region are possible. The negative band
233 (bleach) at 1647 cm^{-1} (-) (1662 cm^{-1} in H_2O) can be attributed to loss of α-helices, with concomitant
234 positive bands at 1661 cm^{-1} (+) and 1630 cm^{-1} (+) (1681 and 1642 cm^{-1} in H_2O) reflecting gain in
235 unordered structures and/or β-sheet elements. Thus, from the point of view of secondary structure
236 changes in the protein, IR spectroscopy suggests a light-induced helix-to-coil transition *somewhere* in
237 EL222. In other LOV domains, a similar IR pattern has been interpreted as unwinding/unfolding of the Jα
238 helix (*15, 43*), and unfolding of the A'α helix (*44*). Hence, we assume that in the EL222 protein, the Jα
239 and A'α helices are, individually or in combination, the likely candidates to unfold upon blue-light
240 irradiation.

IR absorption region of genetically encoded nitrile groups.

We tagged EL222 with cyano probes at different positions to address light-induced structural changes residue-by-residue. In **Fig. 2A** (top panels) we show the absolute spectra of the C≡N stretching vibration of the 40 single-CNF EL222 mutants able to incorporate substantial amounts of FMN. Exceptionally, the EL222-W89CNF mutant showed no measurable peak in that spectral region despite all quality-control analyses confirming the presence of the nitrile moiety (**Table S3**). We characterized the vibration frequency of the C≡N stretching from the peak maximum (ν_{max}). We also looked at the band areas, proportional to the integrated extinction coefficient: larger areas indicate a larger transition dipole moment for the C≡N bond. Recently, it has been proposed that this dipole senses the local electric field by a Stark effect and, thus, the average local electric field ($|\vec{F}|$) projected on the C≡N direction can be derived from the peak area of the C≡N stretch (*45*). Hence, we finally extracted two parameters from the C≡N band of 39 single-CNF mutants: ν_{max} and $|\vec{F}|$ (**Note S1** and **Table S4**). Peak maxima span ~14 cm^{-1}, from 2223 cm^{-1} for N120CNF to 2237 cm^{-1} for Y68CNF, a variability similar to that reported before for nitrile groups in other proteins (~18 cm^{-1}) (*46*). As a reference, for model compounds the nitrile frequency in aprotic solvents decreases with the solvent polarity, from 2233 cm^{-1} in hexane to 2227 cm^{-1} in DMSO. However, for polar protic solvents (able to form H-bonds), the nitrile frequency increases, being 2236 cm^{-1} in water (*47*). Thus, as a first approximation, one could infer that residue CNF68 experiences a very polar and H-bonding environment, and CNF120 a rather apolar and non-H bonding environment. Indeed, CNF120 is the only residue whose peak maxima can be predicted by electric field-frequency calibration curves valid for non-protic solvents (**Fig. S4**) suggesting that all other CNF residues of EL222 are engaged in H-bonding to one extent or another (see further results below).

As done above for the amide I band, a simple procedure to observe changes upon light stimulation is to record a difference spectrum. Difference IR spectra in the nitrile stretching region are shown in **Fig. 2A** (bottom panels) for each individual EL222 mutant. Mutants were classified in three categories based on the observed frequency shift ($\Delta\nu_{max}$), and the estimated change in electric field ($\Delta|\vec{F}|$) of the nitrile band upon illumination. Sixteen mutants experienced a red-shift in ν_{max}, i.e., the illuminated samples had lower mean frequency for its nitrile group after illumination. Six proteins experienced, in contrast, a blue-shift, and seventeen mutants did not show significant frequency shifts (below 0.25 cm^{-1}). For both red-shifted and blue-shifted difference FTIR spectra, a simple pattern is typically observed with a single negative band (i.e. the disappearing dark-state) and a single positive one (i.e. the appearing lit-state). One outlier is L35CNF whose difference IR spectra suggest two components in the lit state. Concerning the changes in the electric field between dark (non-irradiated) and lit (irradiated) states, 12 CNF labels sensed a smaller electric field upon illumination, 11 a larger field, and another 16 sensed virtually no change (less than 1 MV/cm difference).

Visualization and interpretation of CNF spectra.

Mapping absolute and light-induced CNF-derived shifts.

We started first by mapping the characterized CNF microenvironments in the dark state (frequency and electric field) into the primary structure of EL222 (**Fig. S5A**). No evident sequence-dependent pattern was found for the nitrile frequency. However, the larger electric fields tend to concentrate in the middle of the sequence of EL222, between residues 122 and 141 (Hβ and Iβ strands

of the LOV domain). In order to better interpret the spectral changes at the atomic level, we next mapped the nitrile frequencies and electric fields into the dark state crystal structure of EL222 (**Fig. S5B**) (*2*). To this end, we created structural models of the mutants with substitution of the native residues by CNF, considering only the most energetically favored side-chain rotamer (**Note S2** and **Table S5**). Although in these structural models the positions of the cyano side-chains are approximate, it seemed that residues close in space behaved spectrally similar. Exceptions were found. For instance, while CNF148 and CNF151 display similar areas for their nitrile vibration, the peak area of CNF149, and therefore the sensed electric field, is significantly different.

Next, we performed a similar analysis to quantify the extent to which the local environment sensed by the CNF probe changes upon illumination. The light-induced differences in the C≡N stretching wavenumber ($\Delta\nu_{max}$), and local electric field ($\Delta\left|\vec{F}\right|$) were plotted as a function of the residue number for all the variants (**Fig. 2B**). Significant light-induced IR shifts were observed throughout the EL222 sequence from the N-terminal A'α helix (CNF31) to the C-terminal HTH DNA-binding domain (CNF216). Propagation of conformational changes far beyond the FMN locus has been also proposed based on the NMR chemical shift differences between dark and lit EL222 states (*2*). Our results using the cyano groups have unveiled a "hot spot" between residues 136 and 153, where on average half of the residues showed notable light-induced frequency and/or area shifts. This region encompasses the C-terminal part of the LOV domain (Iβ strand), the intrinsically disordered region between the folded domains, and the N-terminal part of the bridging Jα helix. In other LOV-containing proteins, this part of the β-sheet has been proposed as an interaction interface between LOV domains (*48-50*), and the Jα helix has been shown to unfold upon blue-light irradiation (*15, 43*).

The magnitude of the photoinduced changes in frequency or electric fields (areas) does not correlate with closeness to the light-harvesting site (the FMN, see **Fig. S6**). On the other hand, spatial proximity of residues does not always imply a similar spectral response of the nitrile group of CNF to the protein structural changes induced by light (**Fig. 2C**). For instance, CNF151 and CNF153 (two residues in the Jα helix) experience comparable red shifts in the lit-state without clear area changes, possibly connected to a partial Jα unfolding. However, the three residues sitting on A'α (CNF31, CNF35, and CNF36) respond quite differently in the lit-state: red shift and area decrease for CNF31, blue shift and area increase for CNF35, and blue shift and area decrease for CNF36, thus complicating the interpretation about the nature of light-induced structural changes in the A'α helix. This is in line with the multiple roles played by A'α in other LOV domains, including unfolding (*44*) and participation in protein-protein interaction interfaces (*48-50*).

It is informative to compare our IR results with the published NMR assignments of EL222 (*2*). We should keep in mind that the ^{1}H-^{15}N chemical shifts report primarily on the environment around peptide backbone amide groups while site-specific IR frequency shifts monitor the environment around side-chain C≡N triple bonds. Additionally, the NMR shifts correspond to the fraction of EL222 molecules that remain monomeric under illumination, while IR is sensitive to all species (monomeric or else) present in solution. Some residues were only accessible to IR (e.g. W31) or NMR (e.g. A42). For some positions both techniques agree and point to changes in the A'α (L35), LOV domain (N120), 1α-2α loop (G181), and 4α helix (L216) of the HTH domain. Other locations show differential sensitivity. For instance, F80 is the residue with the largest chemical shift difference but the nitrile frequency of CNF80 did not vary

substantially with light and only a modest 2 MV/cm change in electric field was observed. On the other side, IR clearly detects a change in the surroundings of residues CNF118 (LOV domain) and CNF166 (HTH domain), that is largely absent in NMR studies for the corresponding native residue F118 and L166. These results emphasize the complementarity between NMR and IR spectroscopies for the interpretation of structural dynamics in complex biological machines.

Towards a molecular-level interpretation of nitrile frequencies.

Introducing CNF groups at many different positions in EL222 provided us with information about light-induced changes on the local electric field from variations in the nitrile band area. We also resolved changes in the nitrile frequency, whose interpretation is more challenging by its dual dependence on local electric fields and H-bonding (*45*). In H-bonding environments, e.g. water, a blue-shift of the nitrile stretching frequency upon increasing the magnitude of the electric field is observed (*45, 47*). In contrast, in non-H bonding environments, e.g. DMSO, the nitrile stretching frequency red-shifts as the electric field increases (*47*). A combined analysis of peak frequency and band area of nitrile probes has been recently proposed to characterize their H-bonding (**Note S1**) (*45*). We applied this analysis to our data, and the estimated frequency shifts ascribable to H-bonding ($\Delta\nu_{HB}$) are listed in **Table S4** for each CNF group. Surprisingly, little or no significant changes in the H-bonding strength of nitrile groups are predicted between dark and lit states of EL222 by such analysis, CNF123 being one notable exception. Our results do not dismiss the utility of such a method but rather suggest that further benchmarking is necessary.

Another more established way to identify changes in the hydrogen bonding interactions of nitrile groups, even in the dynamic and complex environment found in a protein (*46*), is through frequency-temperature line slope (FTLS) plots (*51*). In FTLS analysis, the vibrational frequency of a selected bond is recorded as a function of temperature. For model nitrile groups dissolved in aprotic solvents, unable to form H-bonds with the nitrile group, the slope is zero. In contrast, when dissolved in water, a protic solvent, the slope is approximately -0.04 $cm^{-1}/°C$ (equivalent to ~1 cm^{-1} downshift every 24 °C increase) (*46*). Thus, the slope in FTLS plots for CNF groups is related to the degree of exposure to water, with a zero slope indicative of a buried residue (*51*). Even more, the slope is directly related to molecular-level details of the specific hydrogen bonding interactions to the nitrile probe. In particular, for a given CNF residue there is a strong correlation between the FTLS and the solvent accessible surface area, a structure-sensitive parameter (*46*). A disadvantage of this method is that it requires spectral measurements at different temperatures and is therefore time-consuming when applied to a large number of proteins.

We concentrated on four positions that cover different parts of the sequence, show wt-like FMN spectra and display clear light-induced spectral differences: W31CNF (A'α, red-shift and smaller electric field), L35CNF (A'α, blue-shift and larger electric field), Y136CNF (LOV, red-shift and larger electric field), and M151CNF (Jα, red-shift and no change in electric field). The experimentally measured spectra in the nitrile absorption region of several EL222 variants is plotted in **Fig. S7**. The chosen temperature range (5-35 °C) is well below the melting temperature of wt EL222 (45 °C). Two residues (Y136CNF and M151CNF) show relatively small but clear changes in slope between their dark and lit states (**Fig. 3A**). In the absence of illumination, CNF151 displays a slope indicative of full solvent exposure, while for CNF136 the slope lies in between that expected for a fully exposed and a fully buried residue. Both residues

364 become less exposed to water upon perturbation by light. On the other hand, a large change in slope
365 occurring between the dark and lit states of EL222 W31CNF (**Fig. 3B**) indicates a dramatic change in
366 polarity around the residue 31, from a slope corresponding to full solvation in the dark state to a null
367 slope indicative of a solvent-excluded residue in the lit state. Since EL222 is known to oligomerize in the
368 lit state (*3, 4, 6*), we suggest that CNF31, and by extension W31, participates in a protein-protein
369 interaction surface. In fact, the dimer interface seen in the X-ray structures of three LOV photosensors
370 crystallized in the lit state is largely constituted by interactions involving the A'α helix (*48-50*). To test
371 such hypothesis we designed two EL222 mutants as positive (constitutive oligomerization) and negative
372 (diminished oligomerization) controls. First, we prepared the S140Y mutation, known to promote EL222
373 oligomerization already in the dark-adapted state (*6*), in combination with W31CNF as a reporter. Wt
374 EL222 was nearly 100% monomeric without light stimulation (**Table S3**). We found that the EL222
375 W31CNF/S140Y mutant had enhanced oligomerization propensity in the absence of illumination with
376 respect to the wt version (**Table S3**), and its light-induced difference IR spectra in the C≡N region did not
377 show appreciable changes (**Fig. 2A**). Moreover, FTLS analysis of W31CNF/S140Y displayed a reduced
378 slope compared to W31CNF, suggesting a lower degree of solvent accessibility around CNF31 in the
379 dark state of the EL222 W31CNF/S140Y variant compared to the dark state of EL222 W31CNF, albeit not
380 as dehydrated as the lit state of EL222 W31CNF (**Fig. 3B**). As a second control, we produced a truncated
381 EL222 lacking the HTH domain, known to drive homo-dimerization (*2*), again including the mutation
382 W31CNF as a reporter (LOV-W31CNF). This variant was mostly monomeric in the dark (**Table S3**) and
383 showed no clear light-induced spectral changes (**Fig. S7**). Accordingly, the slope in the FTLS experiments
384 were similar between LOV-W31CNF and the parental two-domain EL222 W31CNF mutant in the dark
385 state, suggesting a similar environment around residue CNF31 (**Fig. 3B**). Taken altogether, these results
386 are consistent with the nitrile moiety at position 31 reporting primarily on the oligomerization state of
387 EL222.

388 In the case of the L35CNF mutant, the observed asymmetry in the C≡N band (**Fig. 2A**), possibly
389 indicating the presence of more than one CNF environment, complicated the analysis. Indeed, three
390 bands (one negative and two positive) were clearly resolved at all temperatures in the photostationary
391 light-minus-dark difference IR spectra of EL222 L35CNF (**Fig. S7**). In principle, two bands in the absolute
392 spectra should give rise to two pairs of positive/negative bands in the difference spectra. Hence, the
393 missing band could be possibly due to cancellation of one negative band by others. We fitted multiple
394 Gaussian distributions to the difference spectra of L35CNF, and used the mean peak positions to
395 construct FTLS plots (**Fig. S7**). The slope of the negative band (corresponding to the dark state) agrees
396 with a fully water exposed residue. The slope of one of the positive bands was similar to the slope of the
397 dark state, while the other positive band had a steeper slope, more than twice that reported before for
398 model CNF groups dissolved in water (**Fig. 3B**). Therefore, we conclude that in one of the two
399 environments (or conformations) of L35CNF populated in the light, the CNF probe experiences a
400 stronger hydrogen-bonding environment than in liquid water. Because L35 resides in A'α element
401 (residues 30-37) and its side-chain lies at the interface between the A'α and HTH domains (**Fig. S1B**), we
402 asked ourselves whether CNF35 could be sensing the folding status (helix *vs.* coil) of A'α or the proximity
403 between the A'α-LOV and HTH halves of EL222. To test the latter hypothesis, we designed two EL222
404 variants as negative controls. First, we prepared a double mutant containing L35CNF and L123K. The
405 latter mutation has been reported to increase the size (hydrodynamic radius) and proteolytic sensitivity

of EL222, probably by favoring an "open" conformation (HTH decaged from the LOV domain) even in the absence of illumination (*2*). The elution time of L35CNF/L123K was intermediate between that of monomeric and dimeric wt EL222 (**Table S3**), compatible with monomeric EL222 in a more extended conformation than that seen in the crystal structure. The absolute and difference IR spectra of EL222 L35CNF/L123K are shown in **Fig. 2A**. The amplitude of the difference spectrum was too low and noisy to allow further analysis. Nevertheless, FTLS analysis of L35CNF/L123K rendered a slope different from that seen in the dark state of L35CNF (**Fig. 3B**). Secondly, we introduced the L35CNF mutation in an EL222 variant lacking the HTH domain (LOV-L35CNF). There was evidence of only a single lit environment for residue CNF35 in LOV-L35CNF (**Fig. S7**), which did not differ significantly from the dark-state (**Fig. 3B**). We conclude that in one of the two environments experienced by CNF35 in the EL222 L35CNF mutant, the CNF probe might be sensing the light-induced open/close (decaged/caged) conformational equilibrium of the protein.

Time-resolved spectroscopy of EL222.

To characterize the transition from lit to dark states with single-residue resolution, we measured the photorecovery kinetics of EL222 using three different spectral regions: 320-550 nm (FMN probe), 1500-1750 cm^{-1} (amide probe), and 2200-2270 cm^{-1} (CNF probe) (**Fig. 4A**). The UV/Visible spectra mainly reveal adduct rupture (A_{390}-to-D_{450} transition), bands in the 1500-1750 cm^{-1} region disclose global protein secondary structure changes ($Amide_{coil}$-to-$Amide_{helix}$ transition) among others, and cyano bands in the "transparent window" region report on local environment changes around a particular CNF residue ($CNF_{lit\text{-}environment}$-to-$CNF_{dark\text{-}environment}$ transition) (**Fig. 4B**). Thus, as opposed to FMN and amide bands, which report on local changes around the chromophore and global protein backbone changes, respectively, the nitrile group witnesses local changes around the C≡N group of the CNF residue to which it is attached. Of note, a direct comparison of all these three probes can only be accomplished using H_2O as a solvent. While D_2O as a solvent facilitates the study of the structure-sensitive amide I region, as already mentioned, it literally wipes out the tiny infrared signal from the nitrile stretching vibration (**Fig. S1A**). Measuring secondary structure changes of soluble proteins in bulk H_2O is possible but technically challenging by the strong background absorption from H_2O (see **Note S1**). Prior to studies with CNF mutants, we thoroughly studied the recovery kinetics of wt EL222, natively devoid of nitrile bands, by UV/Vis/IR spectroscopies using two intrinsic probes: FMN and amide.

WT EL222.

Full kinetic traces for wt EL222 in H_2O and D_2O buffer (wavelength or frequency *vs.* time) can be found in **Fig. S8-S9** (panel A), respectively. Experimental traces were analyzed by a maximum entropy lifetime distribution analysis (LDA) (52). The commonly employed global exponential fit and related kinetic analyses use a fixed number of exponential terms with adjustable but common time constants, which implicitly assume a discrete and known number of intermediary states (*53*). In contrast, LDA uses a distribution of exponential terms, able to adapt to more complex scenarios such as when the number of intermediates is unknown or even not well-defined, when the exponential terms are not discrete but distributed, or when the same time constants are not shared at all frequencies (*54*). An example of experimental and fitted time traces can be found in **Fig. 4C**. The output of LDA is typically a lifetime density map (**Fig. S8-S9** panel B) where each point represents the exponential amplitude for a given lifetime and frequency (either wavenumber in the case of IR or wavelength in the case of UV/Vis

447 spectroscopy). One way to summarize the information content of a lifetime density map is to calculate
448 the square root of the sum over the squared amplitudes for all frequencies, providing a lifetime
449 distribution known as the "average dynamical content", hereafter abbreviated as *D* (**Note S1**) (*52, 55*).
450 The amplitudes of *D* lifetime distributions provide a visual summary about the magnitude of the
451 spectroscopic changes that occur at a given time constant. *D* lifetime distributions for the UV/Vis and IR
452 data are plotted in **Fig. 5A** (top panel) and lifetimes (τ) with the largest *D* values are listed in **Table S6**.
453 Regarding the FTIR data, we also obtained *D* lifetime distributions for three spectral sub-ranges:
454 a) 1750-1680 cm^{-1}, which is dominated by the two bands from FMN carbonyls, b) 1680-1600 cm^{-1}, with
455 bands mainly ascribable to the protein backbone amide I mode, and c) 1600-1500 cm^{-1}, which has mixed
456 contributions from the protein backbone amide II bands and FMN ring modes. For wt EL222 in H_2O, one
457 dominant photorecovery time constant was found for all triplicates at 41 s for UV/Vis (320-550 nm), 44 s
458 for FMN C=O (1750-1680 cm^{-1}), 47 s for the Amide II (1600-1500 cm^{-1}), and 59 s for the Amide I (1680-
459 1600 cm^{-1}) (**Fig. S10A** top). Although the recovery lifetime in the UV/Vis clearly preceded that in the
460 amide I region, it might be unclear if this was a reliable feature or induced by the higher noise and
461 baseline fluctuations in the latter region caused by the strong background absorbance from liquid water.
462 To resolve this point, as well as to gain more insight into the mechanism of dark reversion, we repeated
463 the kinetics in D_2O-based buffer, as a function of temperature and protein concentration.
464 In D_2O-based buffer, dark recovery kinetics of wt EL222 was clearly slower than in H_2O-based
465 buffer (**Fig. S10A** bottom). The kinetic isotope effect (KIE) for the main kinetic event was 4.6 for the
466 amide II, 4.3 for the FMN, 4.2 for the FMN C=O, and 3.8 for the amide I. Importantly, the decoupling
467 between UV/Vis-based (320-550 nm) and FTIR-based (1750-1500 cm^{-1}) kinetics is confirmed in D_2O
468 buffer, with FMN recovering to the dark state in ~178 s and the protein backbone in ~220 s (**Fig. S10A**
469 bottom): FMN recovery occurs prior to protein backbone recovery also in D_2O solvent.
470 From recovery kinetics measured in the UV/Vis and IR regions at various temperatures, we
471 prepared Arrhenius plots (natural logarithm of rate constant *vs.* reciprocal temperature). As shown in
472 **Fig. S11A**, the rate constants were basically independent of the IR probed frequency range at all tested
473 temperatures. Therefore, from here on, *D* lifetime distributions and associated lifetimes (τ_{Amide}) will refer
474 to the whole 1750-1500 cm^{-1} region. The recovery kinetics in the UV/Vis region was faster i.e. smaller
475 lifetimes (τ_{FMN}) than in the IR region at all temperatures, both in H_2O and D_2O, with the difference being
476 reduced as the temperature increased, both converging at 35 °C (**Fig. S11B**). The activation energies
477 from the IR and UV/Vis regions, derived from the slopes in Arrhenius plots, were similar but not identical
478 (~18 kcal/mol and ~15 kcal/mol, respectively), both H_2O and D_2O (**Table S7**). We conclude that
479 acquisition of secondary structure and adduct breakage are two separate events.
480 Due to the large difference in extinction coefficients and concentrations between the FMN
481 chromophore, peptide carbonyls groups, and nitrile groups, together with the necessity to minimize the
482 background absorption of liquid water in the IR, we had to use lower protein concentrations in UV/Vis
483 (50 µM) than in FTIR (1-2 mM) experiments. Because such difference in concentration could be partially
484 behind the faster recovery kinetics measured in the UV/Vis than in the IR, we measured the dark
485 recovery kinetics of wt EL222 at various protein concentrations by both UV/Vis and FTIR time-resolved
486 spectroscopy (**Fig. S12**). A concentration-independent time constant of ~40 s was found in the 320-550
487 nm region from 10 µM to 400 µM, while a concentration-independent time constant of ~55 s was
488 obtained in the 1750-1500 cm^{-1} region from 400 µM to 2 mM. Therefore, the recovery reaction of EL222

489 is approximately zeroth-order from 10 μM to 2 mM. We also confirm that the kinetic parameters derived
490 by UV/Vis and FTIR spectroscopies can be safely compared to each other even if derived from
491 experiments at dissimilar protein concentration, and we reaffirm the conclusion that the recovery of the
492 FMN chromophore monitored by the UV/Vis is faster than the recovery of the protein backbone
493 monitored by IR spectroscopy.

494 Since protein domains are independent folding units, the observed lit-to-dark refolding kinetics
495 of EL222 could be affected by the presence of two domains (LOV and HTH). We thus prepared and
496 measured a variant (EL222-LOV) containing the A'α and Jα extensions flanking the LOV domain but
497 lacking the C-terminal DNA-binding module. The steady-state difference spectra of EL222-LOV (**Fig. S1B**
498 in H_2O and **S1C** in D_2O) were similar to that of wt EL222 suggesting that most of the secondary structure
499 changes are localized in the A'α-LOV-Jα region of EL222. In addition, a similar lit-to-dark recovery
500 kinetics is still present, with FMN relaxation clearly preceding the relaxation of the protein backbone,
501 although with the difference that the rates were accelerated for both H_2O and D_2O experiments (**Fig.**
502 **S10B** and **Table S6**). Therefore, we dismiss the assumption that the asynchronous relaxation of FMN and
503 the protein backbone in EL222 could be a consequence of EL222 being a multi-domain protein.

504 All these results taken together, three key conclusions can be extracted from time-resolved
505 studies using the native chromophores present in wt EL222 and its LOV domain. The first is that recovery
506 kinetics manifest a single dominant kinetic process, although additional but minor events cannot be
507 excluded. The second is that the main kinetic event reported by UV/Vis spectroscopy (FMN-C78 adduct
508 rupture) seems to precede the main kinetic event probed by IR spectroscopy (global protein refolding to
509 the dark-state conformation), at least at temperatures below 35 °C (**Fig. 5A** top row). The third is that
510 sub-dividing the 1750-1500 cm^{-1} IR region into smaller zones does not lead to a clear increase in the
511 information content i.e. the number of kinetic components stays unchanged. We next asked whether
512 the non-native CNF probes could resolve additional kinetic events in the lit-to-dark transition of EL222.

513 *Single-CNF EL222 mutants.*

514 To facilitate the kinetic comparison among the CNF mutants, as well as with respect to wt EL222,
515 it is desirable that the introduced CNF residues minimally alter the thermodynamics and kinetics of the
516 photochemical reaction. To this end, we first screened single-CNF mutants according to their
517 photoconversion yields and FMN photorecovery kinetics.

518 The fraction of EL222 molecules in the lit-state (the adduct state or A_{390} species) upon
519 illumination at 450 nm was estimated based on UV/Vis absorbance measurements (**Note S3**). For wt
520 EL222, the photoconversion yield was close to one (**Table S8**), indicating that EL222 molecules covalently
521 bound to FMN via C78 predominate under steady-state conditions. Photoconversion yields were also
522 large for essentially all mutants (except F80CNF), suggesting that CNF residues minimally perturbed the
523 equilibrium composition of the photostationary state.

524 The FMN recovery lifetimes for single-CNF mutants, obtained from the main components in the
525 lifetime *D* distribution from time-resolved UV/Vis spectroscopy, are collected in **Table S6**. Almost all
526 variants feature a single kinetic component. One important exception is the W31CNF mutant. Out of the
527 39 CNF mutants (excluding FMN-less and CNF-less variants), 14 mutants displayed recovery times similar
528 to the wt (~41 s); 13 mutants were found to recover significantly faster than the wt; and 12 of the
529 mutants returned to the dark state at much lower rates than the wt **Fig. S13A**). The fastest mutant was

530 found to be F62CNF (~8 s) while the slowest mutant was N120CNF (~370 s), suggesting a strong
531 destabilization of the lit-state in the former case and a strong stabilization of the lit-state in the latter
532 case. To the best of our knowledge, N120CNF is the slowest single mutant of EL222 so far described and
533 may be useful in optogenetic applications (*56*). Strikingly, five of the "fast" mutations are found in the
534 DNA-binding domain, far away from the FMN (**Fig. S13B**). These results add to the number of studies
535 reporting on the extreme sensitivity of FMN recovery times to even minute perturbations in the LOV
536 folding status and raises concerns about how such perturbations are allosterically coupled to the, often
537 distant, FMN site (*9, 57*).

538 From the pool of mutants that are kinetically non-perturbative (similar FMN lifetimes as wt),
539 nine of them (L35CNF, N53CNF, I108CNF, Y136CNF, P149CNF, M151CNF, M153CNF, L181CNF, and
540 L216CNF) show a clear frequency shift in the CNF stretch region upon illumination and could therefore
541 be suitable for time-resolved IR studies. Unfortunately, the signal-to-noise ratio of the CNF difference
542 spectra of L108CNF, P149CNF, M153CNF and L181CNF was too low to allow accurate time-resolved
543 measurements. As a result, we performed kinetic analyses with five mutants (L35CNF, N53CNF,
544 Y136CNF, M151CNF, and L216CNF). Exceptionally, we included also the W31CNF variant because,
545 although the kinetics of its FMN group is perturbed, the kinetics of its nitrile group may be sensitive to
546 oligomer dissociation, as we concluded in static FTLS experiments (see above).

547 We studied the dark recovery kinetics of the six selected EL222 CNF mutants using two different
548 IR frequency ranges: 1750-1500 cm^{-1}, plus 2270-2200 cm^{-1} (see **Fig. 4C** for the case of L35CNF mutant).
549 Complete kinetic traces for the six single-CNF mutants can be found in **Fig. S14-S19** (panel A). Most
550 single-CNF mutants displayed a single dominant component in the CNF-derived *D* lifetime distributions,
551 equivalent to a mono-exponential behavior, which was used to calculate the recovery lifetime of the
552 CNF moiety (τ_{CNF}) (**Fig. 5A** and **Table S6**). In these cases, the transient decay-associated difference
553 spectra were similar to the steady-state difference spectra (**Fig. S20A** for the nitrile absorption region
554 and **Fig. S20B** for the amide bands). Unlike wt EL222, two mutants (N53CNF and L216CNF, **Fig. S21**)
555 showed a similar recovery kinetics for all three probes (FMN, amide and CNF), suggesting a mechanistic
556 perturbation induced by the CNF label. For the remaining four CNF variants (W31CNF, L35CNF, Y136CNF,
557 M151CNF), the FMN recovery happens before the protein backbone recovery, similar to the behavior
558 observed for wt EL222 (**Fig. 5A** and **Table S6**).

559 Two mutants showed particular behaviors: W31CNF and L35CNF.

560 EL222 W31CNF featured the most intricate kinetics of all single-CNF variants, with two events
561 clearly distinguishable by all three probes, although with unequal amplitudes (**Fig. 5A**). For UV/Vis and
562 CNF relaxation there was a major kinetic event, while for amide-derived kinetics the two components
563 had similar amplitudes. Indeed, the slowest relaxation event of CNF31, which has also the lowest
564 amplitude, was not resolved at all temperatures, so it will be ignored in the following. The spectrum
565 from the slowest kinetic event of W31CNF resembled more the spectrum of W31CNF under
566 photostationary conditions (**Fig. S20B**). The temperature-dependence of the main relaxation component
567 of CNF31 showed non-Arrhenius kinetic behavior, in sharp contrast to all other dynamical processes (**Fig.
568 S11C**), suggesting that CNF31 monitors a unique kinetic event. Therefore, it seems that in the W31CNF
569 mutant, apart from protein backbone and FMN relaxations, an additional kinetic event is taking place
570 and is sensed by both the amide and CNF probes. In combination with the previous mutagenesis and

571 FTLS results, we propose that such an extra process reports on the monomer/oligomer equilibrium of
572 EL222.
573 The EL222 L35CNF mutant showed two similarly intense dynamical events linked to the
574 relaxation of CNF35 residue, with associated lifetimes of ~26 and ~94 seconds (labeled "A" and "B",
575 respectively in **Fig. 5A** middle). The slowest component correlates with the protein backbone recovery,
576 while the FMN recovery sensed by UV/Vis largely precedes it. On the other hand, the fastest CNF
577 component takes places before the FMN moiety relaxes. To facilitate the interpretation of the two
578 dynamical processes of CNF35, CNF35-A as the fastest and CNF35-B as the slowest, we looked at their
579 transient spectra. The two events have distinct associated spectra and thus report on dissimilar changes
580 around the CNF residue: a red shift from ~2241 to ~2230 cm^{-1} for the fastest component, and a blue-
581 shift from ~2229 to ~2240 cm^{-1} for the slowest one (**Fig. S20A**). The transient difference spectrum of
582 CNF35-B resembles more the steady-state difference spectrum of the whole CNF population, where gain
583 in α-helicity occurs, than the transient difference spectra of CNF35-A (**Fig. S20A**). Because the CNF35
584 label is located in the A'α helix, we previously concluded from FTLS analysis of steady-state difference
585 spectra that one of the two spectral components of L35CNF could be monitoring the decaged
586 ("open")/caged ("close") conformational equilibrium of EL222 but the folding state of A'α could not be
587 ruled out. To link stationary and kinetic evidence, we performed time-resolved experiments at different
588 temperatures. First, we did FTLS analyses on transient decay-associated difference spectra (DADS)
589 obtained upon lifetime distribution analysis (**Fig. S7**). The main positive band in the DADS of CNF35-B
590 had a similar FTLS slope as one of the lit states of L35CNF, while the main positive band of the DADS of
591 CNF35-A had a flatter slope (lower solvent exposure) than either state of L35CNF (**Fig. 3B**). Thus, the
592 positive band in the "slow" component (CNF35-B) may correspond to one of the lit environments
593 previously found for CNF35 under continuous illumination. On the other hand, hydrogen bonding
594 around the CNF environment corresponding to the "fast" kinetic component (CNF35-A) seems to deviate
595 from the environment sensed under photostationary conditions. Secondly, we prepared Arrhenius plots
596 for the two kinetic events (**Fig. S11D**) and found that the CNF35-B component had similar activation
597 energy as the backbone relaxation calculated from the amide region of wt EL222 but 1.2 times higher
598 than CNF35-A (**Table S7**). Therefore, we conclude that the slowest CNF35 component of L35CNF (CNF35-
599 B) reports on the same (or similar) relaxation event as the backbone amide vibrations i.e. the global
600 increase in α-helical secondary structure content. In contrast, the fastest CNF35 component of L35CNF
601 (CNF35-A) monitors another dynamical process. Although the precise nature of such a process is difficult
602 to define, considering all the mutagenesis and FTLS results, we tentatively propose that CNF35-A could
603 be sensing the interaction between A'α-LOV and HTH halves of EL222 but not the folding state of A'α.
604 We gained additional information about how homogenous the relaxation kinetics was for the
605 different mutants through comparison of width of the distribution of *D* values, appropriately normalized
606 by dividing the time constant axis by the mean lifetime of a probed range. In such a way, all probes were
607 centered at 1, thus facilitating their comparison when the absolute lifetimes differ. As can be seen from
608 **Fig. 5B**, the lifetime distributions of the nitrile groups in the W31CNF and M151CNF mutants were wider
609 than the rest. Such an effect is unlikely to be an artefact, since all data sets were processed alike and we
610 used essentially the same regularization values in order to obtain the lifetime distributions (**Table S9**). It
611 might rather suggest a heterogeneity in the kinetic barrier for the dark relaxation of CNF31 and CNF151
612 residues. Actually, non-Arrhenius kinetic behavior was observed for both CNF31 and CNF151 (**Fig. S11D**

613 and **S11E**) in contrast with UV/Vis- and amide-derived kinetics. In the particular case of CNF151 we
614 found a nearly zero activation energy at temperatures beyond 15 °C thus suggesting a barrierless
615 transition between its lit and dark environments (**Table S7**).

616 To compare the temporal evolution among mutants, given the altered and complex kinetics in
617 some cases, we defined two reaction coordinates, one relative to the FMN relaxation and the other
618 relative to amide relaxation. Subsequently, we computed normalized lifetimes (**Table 1**) and displayed *D*
619 distributions against normalized lifetimes (**Fig. 5C**). In such representations, the relaxation events
620 monitored by CNF31 and CNF35-A, which partially overlap due to the broad nature of the CNF31
621 distribution, clearly take place before any other processes (**Fig. 5C**). CNF136, CNF35-B, and CNF151 relax
622 after the FMN adduct breaks (**Fig. 5C** top). The relaxation of CNF136 and CNF35-B are nearly
623 simultaneous with the global protein relaxation (**Fig. 5C** bottom). Finally, the relaxation of CNF151 is
624 slower and the distribution of rates broader than the protein backbone relaxation (**Fig. 5C** bottom).
625 Therefore, from the point of view of the CNF probes, the results above point to at least three distinct
626 kinetic classes of residue-specific relaxation times in EL222. First, a fast relaxation preceding both
627 changes in the FMN and protein backbone and sensed as a red shift of CNF35-A and a red shift of CNF31.
628 Second, a relaxation simultaneous with the protein backbone relaxation and sensed as a red shift of
629 CNF136 and a blue shift of CNF35-B. Third, a slow relaxation posterior to the relaxation of the protein
630 backbone and sensed as a red shift of CNF151.

631 We conclude that CNF labels located in different parts of the protein track several aspects of the
632 relaxation to the dark-state that are not concurrent with changes in the FMN nor in the protein
633 backbone. To directly compare the kinetics sensed by more than one CNF probe, without the need of
634 using normalized lifetime distributions, we measured the dark recovery of EL222 variants carrying
635 multiple CNF residues.

636 *Multiple-CNF EL222 mutants.*

637 First, we prepared four double-CNF EL222 mutants: W31CNF/L35CNF, W31CNF/M151CNF,
638 L35CNF/M151CNF, and Y136CNF/M151CNF (**Table S2** and **S3**). Absolute and difference infrared spectra
639 in the CNF region are displayed in **Fig. 2A** bottom row. The lit-minus-dark steady-state difference spectra
640 (SSDS) of three mutants were simple, with a single negative/positive signal pair. Only the pattern of the
641 difference IR spectrum of the L35CNF/M151CNF mutant was more convoluted, presenting two positive
642 bands at both edges together with a nearly zero differential absorbance in the central spectral region.
643 This is consistent with our previous observation that the corresponding single-CNF mutants, L35CNF and
644 M151CNF, showed frequency shifts in opposite directions: blue shifted and red shifted, respectively.
645 Actually, the SSDS in the nitrile region of double-CNF mutants could be well reproduced as a linear
646 combination of the two constituent single-CNF SSDS (**Fig. S22**), suggesting that the two introduced CNF
647 groups worked as two independent probes. The single exception was W31CNF/L35CNF, probably as a
648 result of the close spatial proximity of the two CNF moieties in the double mutant.

649 We next monitored the recovery kinetics after ceasing the illumination by time-resolved UV/Vis
650 and IR spectroscopy (see full datasets in **Fig. S23-S26**). The *D* lifetime distributions of FMN and amide
651 were more complicated than for single-CNF mutants, featuring shoulders, asymmetry and double peaks
652 in some cases (**Fig. 6A**). Regarding the C≡N stretching mode, two dynamical events were clearly
653 identified in three out of four cases (**Fig. 6A**). We attempted to assign the observed time-resolved

spectra by assuming that each transient spectrum of the double-CNF mutants is the result of a linear combination of the two transient spectra of the corresponding single-CNF mutants (**Fig. 6B**), as explained in **Note S1**. In all three cases, the slowest dynamical event seen upon analysis of the time traces in the CNF absorption region is enriched in CNF151 relative to the fastest event (see molecular models and fractional contributions in **Fig. 6C**). Thus, the double-CNF mutants suggest that CNF151 returns to its dark state environment at a slower pace compared to the other three CNF probes (CNF31, CNF35 and CNF136), and that the relaxation times of CNF31 and CNF35 are similar. These observations are in agreement with the results obtained with single-CNF mutants.

Is it feasible to follow the relaxation for more than two CNF groups simultaneously in a single EL222 protein? To test this, we prepared two triple-CNF mutants: W31CNF/L35CNF/M151CNF and L35CNF/Y136CNF/M151CNF (**Table S2** and **S3**). Steady-state absolute and difference infrared spectra in the CNF region are displayed in **Fig. 2A** bottom row. Regarding the lit-to-dark recovery kinetics, two events were resolved for W31CNF/L35CNF/M1151CNF while one event, with an asymmetric and broad peak in the lifetime distribution *D*, was resolved for L35CNF/Y136CNF/M151CNF (**Fig. S27-S29**). Hence, increasing the number of CNF reporters did not lead to a concomitant increase in the number of peaks in the *D* lifetime distribution. On the contrary, in the CNF relaxation in the triple-CNF mutant L35CNF/Y136CNF/M151CNF we resolved less kinetic events (one) than in the double-CNF mutant Y136CNF/M151CNF (two). To clarify to what extent this effect could be due to analytical limitations of the lifetime distribution analysis when several kinetic components overlap, we generated realistic synthetic datasets of multiple-CNF variants by averaging the individual experimental time-resolved matrices of the corresponding single-CNF variants (**Fig. S30**). Only two major dynamical events were retrieved by LDA using the maximum entropy method despite the fact that, individually, the data contain four different dynamical events. Even more revealing is the synthetic dataset corresponding to a hypothetical quadruple-CNF mutant (W31CNF/L35CNF/Y136CNF/M151CNF). In such a case, only a single dynamical event, corresponding to the fastest relaxation of CNF35, is found. Hence, it seems that there is a "dilution" effect in multiple-CNF mutants where the response of some CNF residues may be spectrally masked by other CNF sites depending on their relative intensities and their temporal overlap with other probes. Therefore, the experiments with mutants carrying three CNF probes do not rule out the presence of multiple (more than two) CNF-associated dynamical events, but rather point to limitations of the current lifetime distribution analysis when many kinetic components of different intensities spectrally and temporally overlap. Other limitations of our approach are detailed in **Note S4**.

Taken together, our integrative time-resolved UV/Vis/IR approach suggests that after removal of the light stimulus, the side-chain C≡N moieties, FMN chromophore and protein backbone do not return to the initial state simultaneously. Specifically, the recovery lifetimes follow the order: $\tau_{CNF31} \sim \tau_{CNF35\text{-}A} < \tau_{FMN} < \tau_{Amide} \sim \tau_{CNF35\text{-}B} \sim \tau_{CNF136} \leq \tau_{CNF151}$. Thus, CNF31, and partially CNF35 (CNF35-A), sense the relaxation path back to the dark state prior to all other probes, although the full recovery of CNF35 takes place later together with the protein backbone relaxation. The recovery of the FMN precedes that of the protein backbone. CNF136 and partially CNF35 (CNF35-B) relax simultaneously with backbone relaxation. Finally, the relaxation of CNF151 is delayed with respect to that of the protein backbone. Therefore, we have evidence that some CNF sites sense their recovery to the initial dark state environment at a rate significantly different from either the protein backbone structure or the FMN chromophore. With the help of FTLS and mutational analyses we tentatively assume that CNF31 reports

696 on the A'α-LOV/A'α-LOV interface (oligomer/monomer EL222 equilibrium), and the fastest relaxation
697 event of CNF35 reports on changes at the A'α-LOV/HTH interface (open/close equilibrium between the
698 lit and dark state of EL222).

699 Critically, inclusion of CNF labels disclosed three additional dynamical events including the
700 relaxation around residues CNF31, CNF35 and CNF151. CNF kinetics tend to resemble more amide
701 kinetics than FMN kinetics. Indeed, CNF and amide probes rendered partially redundant kinetic
702 information for several mutants. However, in the absence of information from amide relaxation, the
703 interpretation of CNF-derived kinetics would have been challenging if not impossible to achieve. Lifetime
704 distributions from CNF and FMN probes only overlapped each other in the case of two mutants. In short,
705 none of the probes considered in isolation (side-chain C≡N, backbone C=O, and FMN) can deliver a
706 complete picture of EL222 dark recovery mechanism and kinetics, thus making a case for the
707 complementarity of the three.

708 *Improved photocycle model of EL222.*

709 Based on the synergistic utilization of multiple probes (FMN, backbone, and CNF) and multiple
710 sites (one, two, and three CNF residues), we propose the following model of the dark recovery reaction
711 of the LOV-based transcription factor EL222 (**Fig. 7**). The dark state conformation can be safely assumed
712 to represent the structure solved by X-ray crystallography (*2*). While the lit-state is presently unknown, it
713 should be more oligomeric (dimer held together through interactions at least between the HTH
714 domains), "open" (HTH domain decaged from LOV domain) and have less α-helical content (i.e. more
715 random coil content) than the dark-state. Our data are in agreement with multiple parallel refolding
716 pathways where different equilibria are present, each one with a distinct rate constant, but do not
717 exclude the presence of sequential intermediates or more complex kinetic schemes, particularly
718 concerning CNF35 relaxation. Indeed, not all kinetic events will be experienced by all proteins present in
719 solution. For instance, the fraction of EL222 molecules that stay as monomers upon illumination may be
720 spectrally silent with respect to CNF31 relaxation. Specifically, the lit-to-dark transition of EL222 can be
721 described by up to seven equilibria, including one for the backbone relaxation, another one for the FMN
722 and five for the residues (environments) CNF31, CNF35-A, CNF35-B, CNF136 and CNF151. These seven
723 equilibria can be grouped into four categories based on their measured average recovery lifetimes in
724 H_2O. Ordered from fastest to slowest kinetics but without assuming any dependence among them, we
725 have the following molecular events. In ~25 s, EL222 collapses and the A'α-LOV fragment engages in
726 interactions with the HTH domain (sensed by CNF35-A probe). The oligomer falls apart in similar time
727 scales (CNF31 probe). In ~40 s, the FMN-C78 covalent bond is broken (sensed by the UV/Vis absorption
728 of the FMN probe). The protein backbone (most likely Jα, A'α, or both) increases its α-helical secondary
729 structure content in ~60 s (sensed by amide, CNF35-B and CNF136 probes). Equilibration of the
730 environment around CNF151 takes place ~80 s upon stopping illumination.

731 In the dark-to-lit transition of EL222 measured in D_2O, the time-resolved IR data is typically analyzed
732 using a sequential kinetic model according to which a singlet species (^{1}FMN*) is nearly instantaneously
733 formed upon photon absorption, followed by the triplet species in ~3 ns (^{3}FMN*), and finally the adduct
734 state in ~5 μs (A_{390}) (*5, 8*). Because the steady-state difference spectra are different from the transient
735 spectra of the adduct state obtained a few microseconds after irradiation (**Fig. S31**), the existence of an
736 additional intermediate species in the multi-millisecond time scale is envisioned, as previously

postulated (*5*). Moreover, since some of the largest apparent CNF spectral changes are observed for those mutants that have higher tendency to self-associate than the wt (L123CNF, S140CNF, and Q141CNF), we hypothesize that such an additional intermediate may correspond to sparsely populated oligomeric species. This is in line with the low photo-dimerization yield of EL222 in the absence of its target DNA (*3, 4*).

Discussion

Non-canonical amino acids for site-specific infrared structural biology.

We have studied the dark-state relaxation of EL222 from changes in the UV/Vis absorption spectra of the FMN chromophore and in the infrared absorption spectra resulting from protein backbone amide vibrations, complemented by spectral changes in the "transparent window" region of EL222 mutants containing the genetically encoded ncAA CNF. Major advantages of the co-translational incorporation of ncAA carrying vibrational reporters include: i) the freedom of choice regarding the labeling positions (we successfully introduced CNF at 45 different positions out of 45 attempts); ii) the high degree of incorporation efficiency (only two cases out of 45 showed clear levels of mis-incorporation); and iii) the possibility of simultaneous introduction of several reporters (so far we achieved triple CNF labeling in EL222). General benefits of using CNF are the sensitivity of the nitrile vibration to changes in the local electric field and solvent exposure, and its spectral simplicity. The CNF region typically contains one or at most two peaks, a situation that contrasts with the multitude of overlapping bands in the amide region, facilitating the isolation of dynamic changes as illustrated by cleaner lifetime distributions obtained for the former region. A drawback of the CNF region is its intrinsically local information content, which we compensated for by labelling many different residue locations and by simultaneously studying vibrational changes in the amide region (reporting on global secondary changes in structure), and changes in UV/Vis spectra (reporting on covalent bond formation between the FMN chromophore and C78). In the future, simultaneous use of CNF and other ncAA may provide access to even more dynamical events (**Note S4**).

Our results indicate that CNF residues inserted in selected positions of EL222 can, after a careful scrutiny, provide residue-specific information on the photocycle not attainable by other probes naturally present in EL222, like the FMN chromophore or the protein backbone. This observation is in line with previous studies on other photoreceptors that, in contrast to our work, focused on a relatively low number of unnatural residues (typically less than ten) or did not comprehensively address their potential perturbation (*23, 24, 28-30, 58*). Based on our results with a large collection of mutants (45 for a protein of 210 residues), a major caveat of genetically encoded non-canonical side-chains as site-specific infrared reporters is the risk of mis- or over-interpretations when only a limited number of such probes are tested. As a matter of fact, only two (Y136CNF, M151CNF) out of forty-five CNF residues faithfully recapitulated the kinetics seen in wt EL222. Others, like W31CNF, departed kinetically from wt behavior although they could still provide useful mechanistic information as reporters for changes in solvent accessibility (addressed by FTLS experiments). In total, only three CNF-derived dynamical events (CNF31, the fastest component of CNF35 and CNF151) provided genuinely unique kinetic information.

Additionally, looking at many CNF residues gives an overview of spectral and structural changes that can be mapped on the 1D (sequence level) and the 3D (conformational level) structure of EL222, the latter possible thanks to the availability of a high-resolution structure in its dark-state. Analysis of

778 the infrared spectra in the "transparent window" region of CNF clearly shows long-distance signaling
779 (allosteric communication) inside EL222. Perturbation of the FMN chromophore and binding pocket (the
780 allosteric site) by photon absorption is transferred to the HTH module (the active site where DNA
781 binding occurs) and other regions like A'α and Jα elements. This observation is in agreement with
782 previous results based on NMR spectroscopy, although the residue-by-residue sensitivity of the two
783 techniques differs (*2*).

784 Importantly, the spectral observables (e.g. peak positions and areas) may be transformed into
785 physical variables, like local electric fields and H-bonding, thus providing new avenues for the synergistic
786 use of experimental and physics-based computational approaches. In summary, ncAA-assisted optical
787 spectroscopy complements the information content provided by common structural biology techniques
788 and helps delineating signal transduction pathways in EL222.

789 ## Dark reversion kinetics as witnessed by the chromophore, protein and residues.

790 Previous studies on LOV photorecovery have largely concentrated on the UV/Vis region (*9*)
791 rather than on the IR region (*43*). Here we measured the thermal recovery of EL222 by both UV/Vis and
792 FTIR spectroscopies. In principle, the relaxation times of different structural elements in proteins may be
793 different, which may be inferred by independently analyzing different IR frequencies (*54, 59*). This was
794 not the case for EL222 because the Arrhenius plots were linear and essentially identical at all three
795 probed frequency ranges (1750-1680 cm^{-1}, 1680-1600 cm^{-1} and 1600-1500 cm^{-1}). In turn, these results
796 suggest that the dark recovery of EL222 conforms to a two-state folding transition (*54*). Previously, non-
797 Arrhenius behavior (two different slopes) was observed for the recovery kinetics of EL222 based on FMN
798 absorbance (*60*). Here we restricted ourselves to temperatures at or below 35 °C to avoid partial
799 unfolding of the protein (melting temperature is 55 °C). Indeed, if we compare only the data below 35
800 °C, the two studies agree on the same activation energy (~15 kcal/mol) for the rupture of the flavin-
801 cysteinyl bond. The ~18 kcal/mol barrier seen in the IR region agrees with its slower kinetics and
802 indicates a larger energy expenditure involved in rearranging the structure of EL222 in the transition
803 state relative to adduct breakage. The observed KIE (more than 4-fold slower relaxation kinetics in D_2O
804 than H_2O) is in agreement with previous reports of similar LOV domains based on UV/Vis data (*9, 61*).
805 The high KIE can be explained because the adduct breakage requires releasing a proton and an electron,
806 and the folding of α-helices necessitates the breaking of protein-water H-bonds and formation of new
807 intramolecular H-bonds.

808 For wt EL222 at 20 °C, *D* values suggest a decoupling between FMN and protein relaxation
809 where, on average, the FMN-C78 bond is broken earlier (~41 s in H_2O or ~178 s in D_2O) and then the
810 protein recovers its secondary structure content (~57 s in H_2O or ~220 s in D_2O). These results contrast
811 with the synchronous relaxation of protein and FMN structures seen in other LOV domains by circular
812 dichroism (CD), NMR and UV/Vis spectroscopies (*10, 61, 62*). Such discrepancies are rather puzzling
813 since CD, NMR and IR techniques are all sensitive to protein secondary structure changes. One could
814 argue that the spectral overlap between the protein backbone and the FMN absorption bands in the far-
815 UV region could contaminate the CD kinetics, biasing it toward the UV/Vis-determined rates, which is
816 solely due to FMN. Concerning the previous NMR study, we notice that the method is sensitive to
817 backbone dynamics whose relaxation kinetics could be simpler than the side-chain relaxation kinetics
818 probed by the CNF. Nevertheless, asynchronous dark relaxation of chromophore, backbone and native

side-chains was previously reported for Slr1694, a related photoreceptor belonging to the BLUF (blue-light using flavin adenine dinucleotide) family, based on a hybrid UV/Vis/IR approach (*63*). Also in agreement with our observations, a delayed recovery of the protein structure with respect to the flavin adduct decay has been recently described for a dimeric two-domain LOV photoreceptor based on small-angle X-ray scattering and UV/Vis (*64*).

Implications for protein folding dynamics and signal transduction.

The lit-to-dark transition of EL222 resembles the denatured-to-native refolding of unfolded proteins, where secondary structure acquisition ensues chain collapse (*65*). High cooperativity would imply that once the rate-limiting step has occurred, all subsequent steps would follow at essentially the same time. Such a situation is only observed in the case of N53CNF and L216CNF variants. For the rest of the CNF mutants, our kinetic results indicate a more complex picture where (i) CNF dark recovery rates are residue-specific and (ii) CNF recovery rates need not match neither the local FMN relaxation nor the global protein relaxation rates. These results add to the growing body of reports on protein folding dynamics showing different relaxation times for different residues (*66, 67*), even between the main-chain and side-chain of the same residue (*68*), or between global and local probes (*69, 70*). Because our methodological approach is restricted to the ensemble level, the proposed relative order of events in our four-step kinetic model (**Fig. 7**) is the most probable one. In other words, at the single-molecule level, another order of events would be a priori possible, as shown by molecular dynamics simulations of protein folding (*71*).

If the intrinsic time scales of thio-adduct breakage (picoseconds) (72), chain collapse (microseconds) (*73*), protein monomerization (milliseconds) (*74*), helix-coil transitions (nanoseconds) (*75*), and side-chain rotation (picoseconds) (*76*), are all much faster than the experimentally determined dark recovery times of the whole population of EL222, what is then the reason for observed lifetimes (multi-seconds) and what is the nature of the rate-limiting step? Light-induced protein structural dynamics and allostery typically follows a defined set of sequential conformational transitions (*41, 77, 78*). The light-induced dark-to-lit transition of EL222 is not an exception: FMN photochemistry drives, and therefore precedes, all subsequent conformational changes through a cascade of intermediates. Indeed, secondary structure changes (α-helix unfolding) are observed concomitantly with adduct formation (*5, 8*). In contrast, the dark-induced lit-to-dark transition of EL222 is not necessarily caused by adduct breakage since other sites (CNF31, CNF35) experience a change in their surroundings at earlier times. The escape from heterogeneous traps rather than the nucleation rate has been proposed to account for the residue-specific folding rates of α-helices (*66*). In other LOV sensors, dynamic interactions at protein/protein interfaces are considered additional rate-limiting factors for the conformational recovery (*64*). Our results highlight the A'α element, where both CNF31 and CNF35 reside, as a central hub in controlling the dark recovery rate of EL222. Furthermore, we hypothesize that the dissociation of the A'α-LOV/A'α-LOV interface i.e. dimer disassembly, and/or the formation of A'α-LOV/HTH interface might be the rate-determining step(s) in the lit-to-dark transition of EL222. Our kinetic model is also compatible with weakly coupled equilibria between the chromophore, local protein structure and global protein conformation, as recently described for another photoreceptor (*79*). This would imply that photosensory receptors have redundant signaling pathways that may be important to ensure robust responses to light or absence of it (*79*). Overall, our work stresses again the importance of integrative approaches using

860 multiple probes, residues and techniques to comprehensively characterize protein folding landscapes
861 (*59, 80*).

862 Materials and Methods.

863 Materials.

864 CNF was purchased from Bide Pharmatech. The plasmid encoding EL222 was a kind gift of Dr.
865 Kevin Gardner. The plasmid encoding the intein-chitin binding domain-12his fusion protein was a kind
866 gift of Dr. Edward Lemke. The final plasmid for genetic code expansion encoded an N-terminal EL222
867 protein (amino acids 17 to 225) followed by a C-terminal intein-CBD-12his purification tag. The plasmid
868 pDule2-pCNF encoding the orthogonal translation system (OTS) composed of a polyspecific aminoacyl
869 tRNA synthetase (*Mj*CNFRS) and suppressor tRNA (*Mj*tRNA$_{CUA}$) was a kind gift from Dr. Ryan Mehl. All
870 other materials can be found in the **Key Resource Table**.

871 Protein expression.

872 TAG sequences were introduced by site-directed mutagenesis. *Escherichia coli* BL21(DE3) (or
873 C321.ΔA.exp(DE3) in case of the triple mutants) were co-transformed with a plasmid encoding EL222
874 (with one, two or three TAG triplets at desired positions) and another plasmid encoding an OTS able to
875 incorporate CNF in response to UAG codons. Bacteria were grown in terrific broth (TB) media (or
876 exceptionally glucose-based minimal media in the case of W89CNF and I94CNF mutants) at 37 °C up to
877 an optical density at 600 nm of 0.4 at which point CNF (freshly dissolved in 3M NaOH) was added to the
878 medium at a final concentration of 1 mM. Cells were grown for thirty more minutes, then the
879 temperature was lowered to 20 °C and gene expression was induced with 0.5 mM IPTG in the dark.

880 Protein purification and characterization.

881 All protein purification steps were conducted under very dim light and 4 °C unless otherwise
882 stated. After 16 hours, cells were harvested by centrifugation at 6000 g and disrupted by sonication for 4
883 minutes on ice (50% of duty cycle, 50 mW). Cell suspensions were clarified by centrifugation at 60000 g
884 for 30 minutes and the supernatant was purified by immobilized metal affinity chromatography (IMAC)
885 with a Ni^{2+} resin. Loosely bound proteins were removed with 40 mM imidazole washes and fusion
886 proteins were eluted with 500 mM imidazole. Addition of 100 mM DTT activated the self-splicing intein
887 moiety that cleaved off itself from EL222 (48 hours at 23 °C). Cleavage mixtures were purified by a
888 second IMAC step to remove non-cleaved proteins. The flow-through was polished by size-exclusion
889 chromatography (SEC) in a Superdex75Increase column with 50 mM MES, 100 mM NaCl, pH=6.8 as
890 running buffer. The oligomeric states of all the purified proteins were estimated by interpolating the
891 measured elution volumes on a calibration curve constructed using proteins of known molecular weight.
892 In cases where the oligomeric state could not be unambiguously identified the mutants are labeled as
893 multimers. Fractions were analyzed by SDS-PAGE, pooled and concentrated (**Table S3**) using 10 kDa
894 molecular-weight cut-off filters. Proteins were quantified by two methods: UV/visible spectroscopy
895 (FMN concentration) assuming an extinction coefficient at 450 nm of 13000 $M^{-1}cm^{-1}$, and IR
896 spectroscopy (protein concentration). The latter is proportional to the concentration of C=O bonds
897 (mostly arising from the protein backbone, with minor contributions from the two FMN carbonyls and
898 side-chain carbonyls). The yields of CNF mutants were variable and in general lower by at least one
899 order of magnitude (**Table S3**).

900 All proteins were analyzed by mass spectrometry to determine the molecular weight and check
901 for the presence of the CNF residue (**Table S3**). Proteins were diluted with 100 µL of 5% acetic acid in
902 water and loaded onto Opti-trap C4 cartridge (Optimize Technologies), washed 4 x with 250 µL of 5%
903 acetic acid in water and eluted with 100 µL of 80% acetonitrile, 5% acetic acid. Proteins were analyzed
904 by direct infusion using syringe pump at a flow rate 2 µL/min connected with an electrospray ion source
905 of 15T solariX XR FT-ICR mass spectrometer (Bruker Daltonics). The mass spectrometer was externally
906 calibrated using 1% (w/w) sodium trifluoracetate. Proteins were measured in positive mode with 2M
907 data acquisition. The data were processed using SNAP algorithm, a part of DataAnalysis 4.4 software
908 (Bruker Daltonics).

909 ## Steady-state FTIR.

910 The stationary infrared spectra of all the EL222 variants were measured with a Bruker Vertex
911 70v FTIR spectrometer equipped with a globar source, KBr beamsplitter and a liquid nitrogen cooled
912 mercury cadmium telluride (MCT) detector. The protein samples were centrifuged at 18000 g for 15
913 minutes to remove any precipitants before measurements. D_2O buffer exchange was done by centrifugal
914 concentration method using 10 kDa cutoff filters. The aperture value was set in accordance to get
915 maximum signal without saturating the detector. For measurements in the "transparent window" region
916 or in D_2O buffer, ~25 µl of sample was loaded into a temperature controllable demountable liquid cell
917 with CaF2 windows and a 50 µm Teflon spacer. For measuring amide region in H_2O buffer, ~10 µl of
918 sample and 6 µm Teflon spacer was used instead. For each sample, three spectra (dark state, lit state
919 and difference) were recorded at 20 °C and 200-300 scans were averaged in the spectral range of 4000
920 to 900 cm^{-1} with a resolution of 2 cm^{-1}. Protein lit state was induced by illuminating blue light of 450 nm
921 LED with power of 25 mW/cm^2 on the sample throughout the measurement. Dark and lit samples were
922 measured against buffer and the difference spectra were recorded as the lit spectra against the dark
923 spectra. Protein samples had concentration of 1-2 mM in storage buffer (**Table S3**).

924 The resulting spectra where processed with SpectraGryph software. Baseline correction and
925 Gaussian curve fitting was done with OriginPro (OriginLab).

926 Same setup of FTIR spectrometer mentioned above was used for frequency-temperature line
927 slope (FTLS) measurements. Spectra were recorded at 5, 15, 25, and 35 °C averaging over 300 scans at a
928 resolution of 1 cm^{-1}. First, the reference spectrum for each temperature was measured followed by the
929 protein samples. The samples were equilibrated for 10 minutes before measurement.

930 ## Time-resolved rapid-scan FTIR experiments.

931 Transient infrared spectra with a time resolution of tens of milliseconds were recorded on a
932 Bruker Vertex 70v FTIR spectrometer. For the nitrile region, we used a 4.50 µm bandpass filter. Same
933 sample amount, concentration, path length and temperature was used as in steady state FTIR. For the
934 amide region, no filter was used and a 6 µm path-length was used to minimize the contribution of water
935 absorption. The spectral resolution was 4 cm^{-1}. At a single time point, 10 scans were averaged and for
936 the cyano region it was further averaged 3 times. Samples were first illuminated for three minutes under
937 continuous irradiation to accumulate the lit state. Then the LED was switched off and recovery kinetics
938 were recorded over 10 minutes in the case of H_2O-based experiments or 15 min for D_2O-based
939 experiments.

940 ### Steady-state UV/Vis spectroscopy.

941 UV-visible absorbance spectra in the range of 320 to 550 nm of EL222 variants were measured
942 with a Specord 50 Plus spectrometer (Analytik Jena) using a quartz cuvette (Hellma Analytics) having 10
943 mm path length at 20 °C and 70 µL sample volume. Protein samples were diluted to get approximately
944 0.5 optical density at 450 nm.

945 ### Time-resolved UV/Vis spectroscopy.

946 Dark state recovery kinetics of protein variants were measured by first continuously illuminating
947 the sample for 60 seconds with blue light of 450 nm LED with power of approximately 25 mW/cm^2
948 (M450L1, Thorlabs). LED was then turned off and spectra were subsequently taken for 10 min in the
949 case of H_2O-based experiments or 15 min for D_2O-based experiments. The samples were the same as for
950 the FTIR experiments but they were diluted ~70 times to a concentration between 10 and 30 µM,
951 equivalent to an absorbance at 450 nm ~0.5.
952 Additional methods can be found in Supplementary **Notes S1**, **S2**, and **S3**.

953 ## REFERENCES

954 1. T. Kottke, A. Xie, D. S. Larsen, W. D. Hoff, Photoreceptors Take Charge: Emerging Principles for
955 Light Sensing. *Annual Review of Biophysics* **47**, 291-313 (2018).
956 2. A. I. Nash, R. McNulty, M. E. Shillito, T. E. Swartz, R. A. Bogomolni, H. Luecke, K. H. Gardner,
957 Structural basis of photosensitivity in a bacterial light-oxygen-voltage/helix-turn-helix (LOV-HTH)
958 DNA-binding protein. *Proc Natl Acad Sci U S A* **108**, 9449-9454 (2011).
959 3. A. Takakado, Y. Nakasone, M. Terazima, Sequential DNA Binding and Dimerization Processes of
960 the Photosensory Protein EL222. *Biochemistry* **57**, 1603-1610 (2018).
961 4. A. Takakado, Y. Nakasone, M. Terazima, Photoinduced dimerization of a photosensory DNA-
962 binding protein EL222 and its LOV domain. *Phys Chem Chem Phys* **19**, 24855-24865 (2017).
963 5. J. N. Iuliano, A. A. Gil, S. P. Laptenok, C. R. Hall, J. Tolentino Collado, A. Lukacs, S. A. Hag Ahmed,
964 J. Abyad, T. Daryaee, G. M. Greetham, I. V. Sazanovich, B. Illarionov, A. Bacher, M. Fischer, M.
965 Towrie, J. B. French, S. R. Meech, P. J. Tonge, Variation in LOV Photoreceptor Activation
966 Dynamics Probed by Time-Resolved Infrared Spectroscopy. *Biochemistry* **57**, 620-630 (2018).
967 6. B. D. Zoltowski, L. B. Motta-Mena, K. H. Gardner, Blue light-induced dimerization of a bacterial
968 LOV-HTH DNA-binding protein. *Biochemistry* **52**, 6653-6661 (2013).
969 7. R. N. A. Maia, D. Ehrenberg, S. Oldemeyer, E. Knieps-Grünhagen, U. Krauss, J. Heberle, Real-Time
970 Tracking of Proton Transfer from the Reactive Cysteine to the Flavin Chromophore of a
971 Photosensing Light Oxygen Voltage Protein. *Journal of the American Chemical Society* **143**,
972 12535-12542 (2021).
973 8. P. C. Andrikopoulos, A. S. Chaudhari, Y. Liu, P. E. Konold, J. T. M. Kennis, B. Schneider, G. Fuertes,
974 QM calculations predict the energetics and infrared spectra of transient glutamine isomers in
975 LOV photoreceptors. *Physical Chemistry Chemical Physics* **23**, 13934-13950 (2021).
976 9. N. Lebedev, J. P. Zayner, T. R. Sosnick, Factors That Control the Chemistry of the LOV Domain
977 Photocycle. *PLoS ONE* **9**, (2014).
978 10. S. M. Harper, L. C. Neil, I. J. Day, P. J. Hore, K. H. Gardner, Conformational Changes in a
979 Photosensory LOV Domain Monitored by Time-Resolved NMR Spectroscopy. *Journal of the*
980 *American Chemical Society* **126**, 3390-3391 (2004).
981 11. T. Kottke, V. A. Lorenz-Fonfria, J. Heberle, The Grateful Infrared: Sequential Protein Structural
982 Changes Resolved by Infrared Difference Spectroscopy. *J Phys Chem B* **121**, 335-350 (2017).
983 12. V. A. Lorenz-Fonfria, Infrared Difference Spectroscopy of Proteins: From Bands to Bonds. *Chem*
984 *Rev* **120**, 3466-3576 (2020).
985 13. W.-J. Yang, P. R. Griffiths, D. M. Byler, H. Susi, Protein Conformation by Infrared Spectroscopy:
986 Resolution Enhancement by Fourier Self-Deconvolution. *Applied Spectroscopy* **39**, 282-287
987 (2016).
988 14. J. L. R. Arrondo, A. Muga, J. Castresana, F. M. Goñi, Quantitative studies of the structure of
989 proteins in solution by fourier-transform infrared spectroscopy. *Progress in Biophysics and*
990 *Molecular Biology* **59**, 23-56 (1993).
991 15. P. E. Konold, T. Mathes, J. Weiβenborn, M. L. Groot, P. Hegemann, J. T. M. Kennis, Unfolding of
992 the C-Terminal Jα Helix in the LOV2 Photoreceptor Domain Observed by Time-Resolved
993 Vibrational Spectroscopy. *The Journal of Physical Chemistry Letters* **7**, 3472-3476 (2016).

16. L. Goett-Zink, J. L. Klocke, L. A. K. Bögeholz, T. Kottke, In-cell infrared difference spectroscopy of LOV photoreceptors reveals structural responses to light altered in living cells. *Journal of Biological Chemistry* **295**, 11729-11741 (2020).
17. R. Adhikary, J. Zimmermann, F. E. Romesberg, Transparent Window Vibrational Probes for the Characterization of Proteins With High Structural and Temporal Resolution. *Chem Rev* **117**, 1927-1969 (2017).
18. M. C. Thielges, Transparent window 2D IR spectroscopy of proteins. *The Journal of Chemical Physics* **155**, (2021).
19. A. M. Saleh, K. M. Wilding, S. Calve, B. C. Bundy, T. L. Kinzer-Ursem, Non-canonical amino acid labeling in proteomics and biotechnology. *Journal of Biological Engineering* **13**, (2019).
20. M. Manandhar, E. Chun, F. E. Romesberg, Genetic Code Expansion: Inception, Development, Commercialization. *Journal of the American Chemical Society* **143**, 4859-4878 (2021).
21. J. G. Löffler, E. Deniz, C. Feid, V. G. Franz, J. Bredenbeck, Versatile Vibrational Energy Sensors for Proteins. *Angewandte Chemie International Edition*, (2022).
22. J. Tolentino Collado, J. N. Iuliano, K. Pirisi, S. Jewlikar, K. Adamczyk, G. M. Greetham, M. Towrie, J. R. H. Tame, S. R. Meech, P. J. Tonge, A. Lukacs, Unraveling the Photoactivation Mechanism of a Light-Activated Adenylyl Cyclase Using Ultrafast Spectroscopy Coupled with Unnatural Amino Acid Mutagenesis. *ACS Chemical Biology*, (2022).
23. A. Kraskov, J. von Sass, A. D. Nguyen, T. O. Hoang, D. Buhrke, S. Katz, N. Michael, J. Kozuch, I. Zebger, F. Siebert, P. Scheerer, M. A. Mroginski, N. Budisa, P. Hildebrandt, Local Electric Field Changes during the Photoconversion of the Bathy Phytochrome Agp2. *Biochemistry*, (2021).
24. S. Ye, T. Huber, R. Vogel, T. P. Sakmar, FTIR analysis of GPCR activation using azido probes. *Nat Chem Biol* **5**, 397-399 (2009).
25. L. Blankenburg, L. Schroeder, F. Habenstein, B. Blasiak, T. Kottke, J. Bredenbeck, Following local light-induced structure changes and dynamics of the photoreceptor PYP with the thiocyanate IR label. *Phys Chem Chem Phys* **21**, 6622-6634 (2019).
26. S. Milles, S. Tyagi, N. Banterle, C. Koehler, V. VanDelinder, T. Plass, A. P. Neal, E. A. Lemke, Click Strategies for Single-Molecule Protein Fluorescence. *Journal of the American Chemical Society* **134**, 5187-5195 (2012).
27. J. Zhang, L. Wang, J. Zhang, J. Zhu, X. Pan, Z. Cui, J. Wang, W. Fang, Y. Li, Identifying and Modulating Accidental Fermi Resonance: 2D IR and DFT Study of 4-Azido-l-phenylalanine. *The Journal of Physical Chemistry B* **122**, 8122-8133 (2018).
28. M. Kurttila, B. Stucki-Buchli, J. Rumfeldt, L. Schroeder, H. Häkkänen, A. Liukkonen, H. Takala, T. Kottke, J. A. Ihalainen, Site-by-site tracking of signal transduction in an azidophenylalanine-labeled bacteriophytochrome with step-scan FTIR spectroscopy. *Physical Chemistry Chemical Physics* **23**, 5615-5628 (2021).
29. B. S. Krause, J. C. D. Kaufmann, J. Kuhne, J. Vierock, T. Huber, T. P. Sakmar, K. Gerwert, F. J. Bartl, P. Hegemann, Tracking Pore Hydration in Channelrhodopsin by Site-Directed Infrared-Active Azido Probes. *Biochemistry* **58**, 1275-1286 (2019).
30. C. R. Hall, J. Tolentino Collado, J. N. Iuliano, A. A. Gil, K. Adamczyk, A. Lukacs, G. M. Greetham, I. Sazanovich, P. J. Tonge, S. R. Meech, Site-Specific Protein Dynamics Probed by Ultrafast Infrared Spectroscopy of a Noncanonical Amino Acid. *J Phys Chem B* **123**, 9592-9597 (2019).

1036 31. K. C. Schultz, L. Supekova, Y. Ryu, J. Xie, R. Perera, P. G. Schultz, A genetically encoded infrared
1037 probe. *J Am Chem Soc* **128**, 13984-13985 (2006).
1038 32. S. J. Miyake-Stoner, A. M. Miller, J. T. Hammill, J. C. Peeler, K. R. Hess, R. A. Mehl, S. H. Brewer,
1039 Probing protein folding using site-specifically encoded unnatural amino acids as FRET donors
1040 with tryptophan. *Biochemistry* **48**, 5953-5962 (2009).
1041 33. M. J. Lajoie, A. J. Rovner, D. B. Goodman, H.-R. Aerni, A. D. Haimovich, G. Kuznetsov, J. A.
1042 Mercer, H. H. Wang, P. A. Carr, J. A. Mosberg, N. Rohland, P. G. Schultz, J. M. Jacobson, J.
1043 Rinehart, G. M. Church, F. J. Isaacs, Genomically Recoded Organisms Expand Biological
1044 Functions. *Science* **342**, 357-360 (2013).
1045 34. R. Adhikary, J. Zimmermann, P. E. Dawson, F. E. Romesberg, IR probes of protein
1046 microenvironments: utility and potential for perturbation. *Chemphyschem* **15**, 849-853 (2014).
1047 35. S. Krimm, J. Bandekar, "Vibrational Spectroscopy and Conformation of Peptides, Polypeptides,
1048 and Proteins" in *Advances in Protein Chemistry Volume 38* (Advances in Protein Chemistry,
1049 1986), pp. 181-364.
1050 36. A. Barth, Infrared spectroscopy of proteins. *Biochimica et Biophysica Acta (BBA) - Bioenergetics*
1051 **1767**, 1073-1101 (2007).
1052 37. M. Spexard, D. Immeln, C. Thöing, T. Kottke, Infrared spectrum and absorption coefficient of the
1053 cofactor flavin in water. *Vibrational Spectroscopy* **57**, 282-287 (2011).
1054 38. P. C. Andrikopoulos, Y. Liu, A. Picchiotti, N. Lenngren, M. Kloz, A. S. Chaudhari, M. Precek, M.
1055 Rebarz, J. Andreasson, J. Hajdu, B. Schneider, G. Fuertes, Femtosecond-to-nanosecond dynamics
1056 of flavin mononucleotide monitored by stimulated Raman spectroscopy and simulations.
1057 *Physical Chemistry Chemical Physics* **22**, 6538-6552 (2020).
1058 39. K. Ataka, P. Hegemann, J. Heberle, Vibrational Spectroscopy of an Algal Phot-LOV1 Domain
1059 Probes the Molecular Changes Associated with Blue-Light Reception. *Biophysical Journal* **84**,
1060 466-474 (2003).
1061 40. T. Iwata, D. Nozaki, Y. Sato, K. Sato, Y. Nishina, K. Shiga, S. Tokutomi, H. Kandori, Identification of
1062 the CO Stretching Vibrations of FMN and Peptide Backbone by 13C-Labeling of the LOV2 Domain
1063 of Adiantum Phytochrome3. *Biochemistry* **45**, 15384-15391 (2006).
1064 41. A. A. Gil, S. P. Laptenok, J. B. French, J. N. Iuliano, A. Lukacs, C. R. Hall, I. V. Sazanovich, G. M.
1065 Greetham, A. Bacher, B. Illarionov, M. Fischer, P. J. Tonge, S. R. Meech, Femtosecond to
1066 Millisecond Dynamics of Light Induced Allostery in the Avena sativa LOV Domain. *J Phys Chem B*
1067 **121**, 1010-1019 (2017).
1068 42. F. Dousseau, M. Pezolet, Determination of the secondary structure content of proteins in
1069 aqueous solutions from their amide I and amide II infrared bands. Comparison between classical
1070 and partial least-squares methods. *Biochemistry* **29**, 8771-8779 (2002).
1071 43. E. Herman, M. Sachse, P. G. Kroth, T. Kottke, Blue-Light-Induced Unfolding of the Jα Helix Allows
1072 for the Dimerization of Aureochrome-LOV from the Diatom Phaeodactylum tricornutum.
1073 *Biochemistry* **52**, 3094-3101 (2013).
1074 44. E. Herman, T. Kottke, Allosterically Regulated Unfolding of the A'α Helix Exposes the
1075 Dimerization Site of the Blue-Light-Sensing Aureochrome-LOV Domain. *Biochemistry* **54**, 1484-
1076 1492 (2015).

45. J. B. Weaver, J. Kozuch, J. M. Kirsh, S. G. Boxer, Nitrile Infrared Intensities Characterize Electric Fields and Hydrogen Bonding in Protic, Aprotic, and Protein Environments. *Journal of the American Chemical Society* **144**, 7562-7567 (2022).
46. J. T. First, J. D. Slocum, L. J. Webb, Quantifying the Effects of Hydrogen Bonding on Nitrile Frequencies in GFP: Beyond Solvent Exposure. *J Phys Chem B* **122**, 6733-6743 (2018).
47. P. Deb, T. Haldar, S. M. Kashid, S. Banerjee, S. Chakrabarty, S. Bagchi, Correlating Nitrile IR Frequencies to Local Electrostatics Quantifies Noncovalent Interactions of Peptides and Proteins. *J Phys Chem B* **120**, 4034-4046 (2016).
48. A. T. Vaidya, C. H. Chen, J. C. Dunlap, J. J. Loros, B. R. Crane, Structure of a Light-Activated LOV Protein Dimer That Regulates Transcription. *Science Signaling* **4**, ra50-ra50 (2011).
49. M. Bocola, U. Schwaneberg, K.-E. Jaeger, U. Krauss, Light-induced structural changes in a short light, oxygen, voltage (LOV) protein revealed by molecular dynamics simulations—implications for the understanding of LOV photoactivation. *Frontiers in Molecular Biosciences* **2**, (2015).
50. U. Heintz, I. Schlichting, Blue light-induced LOV domain dimerization enhances the affinity of Aureochrome 1a for its target DNA sequence. *eLife* **5**, (2016).
51. R. Adhikary, J. Zimmermann, P. E. Dawson, F. E. Romesberg, Temperature Dependence of CN and SCN IR Absorptions Facilitates Their Interpretation and Use as Probes of Proteins. *Anal Chem* **87**, 11561-11567 (2015).
52. V. A. Lórenz-Fonfría, H. Kandori, Bayesian Maximum Entropy (Two-Dimensional) Lifetime Distribution Reconstruction from Time-Resolved Spectroscopic Data. *Applied Spectroscopy* **61**, 428-443 (2016).
53. I. H. M. van Stokkum, D. S. Larsen, R. van Grondelle, Global and target analysis of time-resolved spectra. *Biochimica et Biophysica Acta (BBA) - Bioenergetics* **1657**, 82-104 (2004).
54. C. M. Davis, L. Zanetti-Polzi, M. Gruebele, A. Amadei, R. B. Dyer, I. Daidone, A quantitative connection of experimental and simulated folding landscapes by vibrational spectroscopy. *Chemical Science* **9**, 9002-9011 (2018).
55. G. Stock, P. Hamm, A non-equilibrium approach to allosteric communication. *Philosophical Transactions of the Royal Society B: Biological Sciences* **373**, (2018).
56. A. Pudasaini, K. K. El-Arab, B. D. Zoltowski, LOV-based optogenetic devices: light-driven modules to impart photoregulated control of cellular signaling. *Frontiers in Molecular Biosciences* **2**, (2015).
57. S. H. Song, P. L. Freddolino, A. I. Nash, E. C. Carroll, K. Schulten, K. H. Gardner, D. S. Larsen, Modulating LOV domain photodynamics with a residue alteration outside the chromophore binding site. *Biochemistry* **50**, 2411-2423 (2011).
58. S. Ye, E. Zaitseva, G. Caltabiano, G. F. Schertler, T. P. Sakmar, X. Deupi, R. Vogel, Tracking G-protein-coupled receptor activation using genetically encoded infrared probes. *Nature* **464**, 1386-1389 (2010).
59. C. M. Davis, R. B. Dyer, WW Domain Folding Complexity Revealed by Infrared Spectroscopy. *Biochemistry* **53**, 5476-5484 (2014).
60. B. D. Zoltowski, A. I. Nash, K. H. Gardner, Variations in protein-flavin hydrogen bonding in a light, oxygen, voltage domain produce non-Arrhenius kinetics of adduct decay. *Biochemistry* **50**, 8771-8779 (2011).

1119 61. S. B. Corchnoy, T. E. Swartz, J. W. Lewis, I. Szundi, W. R. Briggs, R. A. Bogomolni, Intramolecular
1120 Proton Transfers and Structural Changes during the Photocycle of the LOV2 Domain of
1121 Phototropin 1. *Journal of Biological Chemistry* **278**, 724-731 (2003).
1122 62. Y. Nakasone, T. Eitoku, D. Matsuoka, S. Tokutomi, M. Terazima, Dynamics of Conformational
1123 Changes of Arabidopsis Phototropin 1 LOV2 with the Linker Domain. *Journal of Molecular*
1124 *Biology* **367**, 432-442 (2007).
1125 63. K. Hasegawa, S. Masuda, T.-a. Ono, Spectroscopic Analysis of the Dark Relaxation Process of a
1126 Photocycle in a Sensor of Blue Light using FAD (BLUF) Protein Slr1694 of the Cyanobacterium
1127 Synechocystis sp. PCC6803. *Plant and Cell Physiology* **46**, 136-146 (2005).
1128 64. S. Bannister, E. Böhm, T. Zinn, T. Hellweg, T. Kottke, Arguments for an additional long-lived
1129 intermediate in the photocycle of the full-length aureochrome 1c receptor: A time-resolved
1130 small-angle X-ray scattering study. *Structural Dynamics* **6**, (2019).
1131 65. T. Uzawa, S. Akiyama, T. Kimura, S. Takahashi, K. Ishimori, I. Morishima, T. Fujisawa, Collapse
1132 and search dynamics of apomyoglobin folding revealed by submillisecond observations of -
1133 helical content and compactness. *Proceedings of the National Academy of Sciences* **101**, 1171-
1134 1176 (2004).
1135 66. J. A. Ihalainen, B. Paoli, S. Muff, E. H. G. Backus, J. Bredenbeck, G. A. Woolley, A. Caflisch, P.
1136 Hamm, -Helix folding in the presence of structural constraints. *Proceedings of the National*
1137 *Academy of Sciences* **105**, 9588-9593 (2008).
1138 67. C. M. Davis, A. K. Cooper, R. B. Dyer, Fast Helix Formation in the B Domain of Protein A Revealed
1139 by Site-Specific Infrared Probes. *Biochemistry* **54**, 1758-1766 (2015).
1140 68. S. Nagarajan, H. Taskent-Sezgin, D. Parul, I. Carrico, D. P. Raleigh, R. B. Dyer, Differential ordering
1141 of the protein backbone and side chains during protein folding revealed by site-specific
1142 recombinant infrared probes. *J Am Chem Soc* **133**, 20335-20340 (2011).
1143 69. H. Ma, M. Gruebele, Kinetics are probe-dependent during downhill folding of an engineered 6-
1144 85 protein. *Proceedings of the National Academy of Sciences* **102**, 2283-2287 (2005).
1145 70. S. Sukenik, T. V. Pogorelov, M. Gruebele, Can Local Probes Go Global? A Joint Experiment-
1146 Simulation Analysis of lambda(6-85) Folding. *J Phys Chem Lett* **7**, 1960-1965 (2016).
1147 71. E. R. Henry, R. B. Best, W. A. Eaton, Comparing a simple theoretical model for protein folding
1148 with all-atom molecular dynamics simulations. *Proceedings of the National Academy of Sciences*
1149 **110**, 17880-17885 (2013).
1150 72. J. T. M. Kennis, I. H. M. van Stokkum, S. Crosson, M. Gauden, K. Moffat, R. van Grondelle, The
1151 LOV2 Domain of Phototropin: A Reversible Photochromic Switch. *Journal of the American*
1152 *Chemical Society* **126**, 4512-4513 (2004).
1153 73. S. J. Hagen, W. A. Eaton, Two-state expansion and collapse of a polypeptide 1 1Edited by A. R.
1154 Fersht. *Journal of Molecular Biology* **301**, 1019-1027 (2000).
1155 74. T. Miyamori, Y. Nakasone, K. Hitomi, J. M. Christie, E. D. Getzoff, M. Terazima, Reaction
1156 dynamics of the UV-B photosensor UVR8. *Photochemical & Photobiological Sciences* **14**, 995-
1157 1004 (2015).
1158 75. G. S. Jas, K. Kuczera, Helix–Coil Transition Courses Through Multiple Pathways and
1159 Intermediates: Fast Kinetic Measurements and Dimensionality Reduction. *The Journal of*
1160 *Physical Chemistry B* **122**, 10806-10816 (2018).

76. S. F. Cousin, P. Kadeřávek, N. Bolik-Coulon, Y. Gu, C. Charlier, L. Carlier, L. Bruschweiler-Li, T. Marquardsen, J.-M. Tyburn, R. Brüschweiler, F. Ferrage, Time-Resolved Protein Side-Chain Motions Unraveled by High-Resolution Relaxometry and Molecular Dynamics Simulations. *Journal of the American Chemical Society* **140**, 13456-13465 (2018).

77. O. Berntsson, R. Rodriguez, L. Henry, M. R. Panman, A. J. Hughes, C. Einholz, S. Weber, J. A. Ihalainen, R. Henning, I. Kosheleva, E. Schleicher, S. Westenhoff, Photoactivation of *Drosophila melanogaster* cryptochrome through sequential conformational transitions. *Science Advances* **5**, (2019).

78. S. J. Lee, T. W. Kim, J. G. Kim, C. Yang, S. R. Yun, C. Kim, Z. Ren, I. Kumarapperuma, J. Kuk, K. Moffat, X. Yang, H. Ihee, Light-induced protein structural dynamics in bacteriophytochrome revealed by time-resolved x-ray solution scattering. *Science Advances* **8**, (2022).

79. H. Takala, H. K. Lehtivuori, O. Berntsson, A. Hughes, R. Nanekar, S. Niebling, M. Panman, L. Henry, A. Menzel, S. Westenhoff, J. A. Ihalainen, On the (un)coupling of the chromophore, tongue interactions, and overall conformation in a bacterial phytochrome. *Journal of Biological Chemistry* **293**, 8161-8172 (2018).

80. M. B. Prigozhin, S. H. Chao, S. Sukenik, T. V. Pogorelov, M. Gruebele, Mapping fast protein folding with multiple-site fluorescent probes. *Proc Natl Acad Sci U S A* **112**, 7966-7971 (2015).

Author contributions.

Conceptualization: VALF, GF
Formal analysis, methodology: ASC, VALF, GF
Funding acquisition: BS
Investigation: ASC, AC, CAOD, GF
Resources: PCA, YL
Software: VALF
Visualization: ASC, GF
Writing—original draft: GF
Writing—review & editing: ASC, IA, VALF, GF

Conflicts of interest.

There are no conflicts of interest to declare.

Acknowledgments.

The work was supported by the project Structural dynamics of biomolecular systems (ELIBIO) (CZ.02.1.01/0.0/0.0/15_003/0000447) from the European Regional Development Fund (ERDF) and the Ministry of Education, Youth and Sports (MEYS) of the Czech Republic. The Institute of Biotechnology of the Czech Academy of Sciences acknowledges the institutional grant RVO86652036. We acknowledge CF Biophysic, CF SMS of CIISB, Instruct-CZ Centre BIOCEV, supported by MEYS CR (LM2018127)) and ERDF-Project „UP CIISB“ (No. CZ.02.1.01/0.0/0.0/18_046/0015974). We thank the financial support provided by the Ministerio de Ciencia e Innovación (MICINN) - Agencia Estatal de Investigación (AEI) through projects BFU2016-768050-P, BFU2017-91559-EXP, PID2019-106103GBI00 and CTQ2017-87372-P, and by the Generalitat Valenciana through the project PROMETEU/2019/066.

Data and materials availability.

Steady-state and time-resolved datasets will be deposited in Zenodo (10.5281/zenodo.7086623). Other data, software and materials are available from the corresponding authors upon reasonable request. Plasmids can be shared upon signing of a Material Transfer Agreement (MTA).

1208 TABLE 1.

1209 **Normalized lit-to-dark recovery lifetimes.** Normalized CNF-derived recovery times of single-CNF EL222
1210 variants measured in H_2O buffer and a possible explanation of the corresponding dynamical event based
1211 on mutagenesis and FTLS evidence. Absolute (non-normalized) lit-to-dark lifetimes appear in **Table S6**.

Probe	τ_{CNF}/τ_{FMN}	τ_{CNF}/τ_{Amide}	Tentative assignment
CNF31	0.5	0.2	Disruption of A'α-LOV/ A'α-LOV interactions (oligomer dissociation)
CNF35-A	0.6	0.2	A'α packing against HTH (formation of caged conformation)
CNF35-B	2.5	0.9	Fold switching from disordered to α-helical (affecting primarily A'α and Jα elements)
CNF136	1.4	1.1	
CNF151	2.0	1.3	Reconfiguration disordered loop

1212

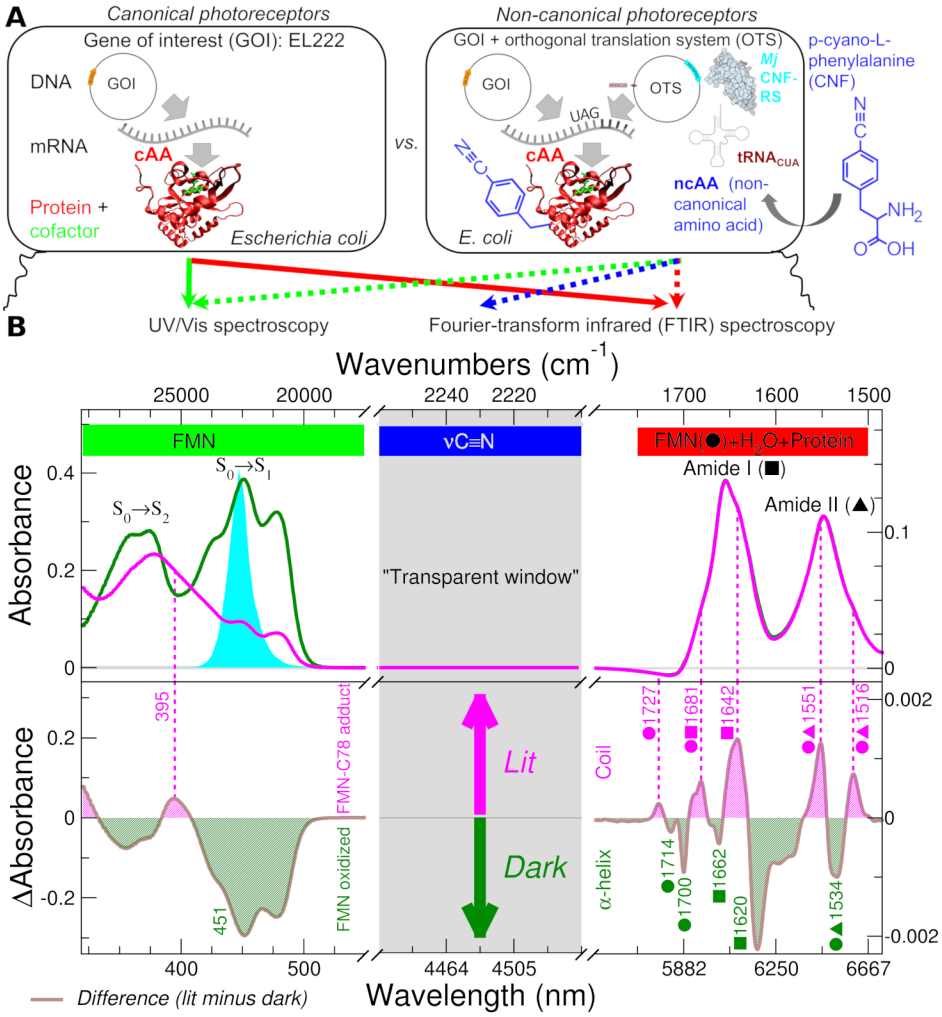
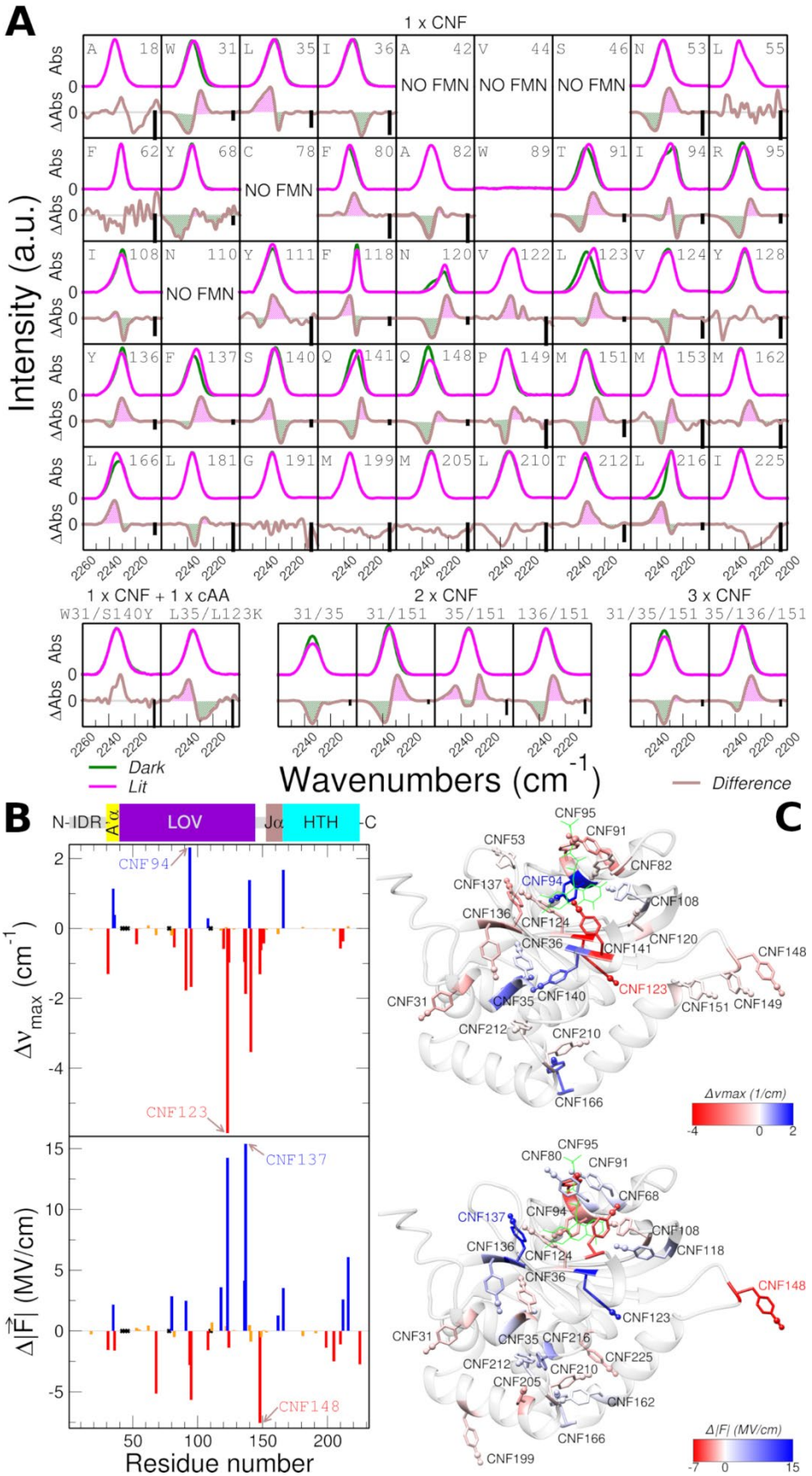


FIGURE 1.

Protein engineering via genetic code expansion meets electronic/vibrational spectroscopy. (A) Comparison between photoreceptors based on canonical (left) and non-canonical (right) amino acids (ncAA) carrying vibrational probes. The ncAA p-cyano-L-phenylalanine (CNF) contains a triple bond C≡N group vibrating in the "transparent window" region (grey area in panel B) and delivers local residue-specific information. **(B)** Absolute (top) and difference (bottom) electronic and vibrational spectra of wt EL222 in H_2O-based buffer. In the case of UV/Vis spectra (left), dark state (green) and lit state (magenta) samples have been collected before and immediately after blue-light illumination (spectral output in the top panel, cyan area), respectively. Transmission IR spectra of hydrated EL222 (right) were collected in the absence (dark state) and presence (lit state) of continuous blue-light illumination. The positions of the main bands discussed in the text are marked as dashed lines and numbers correspond to their wavelength/frequency.


A
1 x CNF
Intensity (a.u.)
Abs
ΔAbs
NO FMN
1 x CNF + 1 x cAA
W31/S140Y L35/L123K
2 x CNF
31/35 31/151 35/151 136/151
3 x CNF
31/35/151 35/136/151
Dark
Lit
Difference
Wavenumbers (cm-1)
B
N-IDR
Aα
LOV
Jα
HTH
-C
CNF94
CNF123
CNF137
CNF148
Residue number
C
Δvmax (1/cm)
Δ|F| (MV/cm)

1226

1227 FIGURE 2.

1228 **Site-specific infrared spectroscopy of EL222 mutants in the dark and lit states. (A)** Absolute (top) and
1229 steady-state light-induced difference (bottom) IR spectra in the nitrile absorption region of forty-five
1230 single-CNF EL222 mutants (each carrying a single CNF ncAA in place of the indicated native residue),
1231 together with four double-CNF, two triple-CNF and two single-CNF/single cAA mutants (bottom row).
1232 Negative bands in the lit-minus-dark difference spectra correspond to the dark state (green areas) while
1233 positive bands indicate the lit state (magenta areas). Difference spectra without shaded areas indicate
1234 absence of significant changes (less than 0.25 cm^{-1} frequency shifts and less than 1 % area difference).
1235 The vertical scale bar (black) shown in the difference spectra amounts to 0.5 mOD. **(B)** Mapping site-
1236 specific changes reported by the CNF probe along the sequence. The light-minus-dark difference
1237 between the frequency of the maximum absorbance ($\Delta\nu_{max}$) and changes in electric fields ($\Delta|\vec{F}|$) are
1238 plotted as a function of the residue number. Residues with the largest differences are indicated. A
1239 scheme of different EL222 regions is shown on top, displaying, from the N- to the C-terminus, an
1240 intrinsically disordered region (IDR, grey), the A'α extension (yellow), the LOV domain (purple), another
1241 IDR (grey), the Jα bridging helix (brown), and the HTH domain (cyan). **(C)** Mapping site-specific changes
1242 reported by the CNF probe onto the known 3D structure of dark-adapted EL222 (PDB: 3P7N). Separate
1243 models with CNF residues colored according to the two criteria in panel B, $\Delta\nu_{max}$ (top) and $\Delta|\vec{F}|$ (bottom)
1244 are shown.

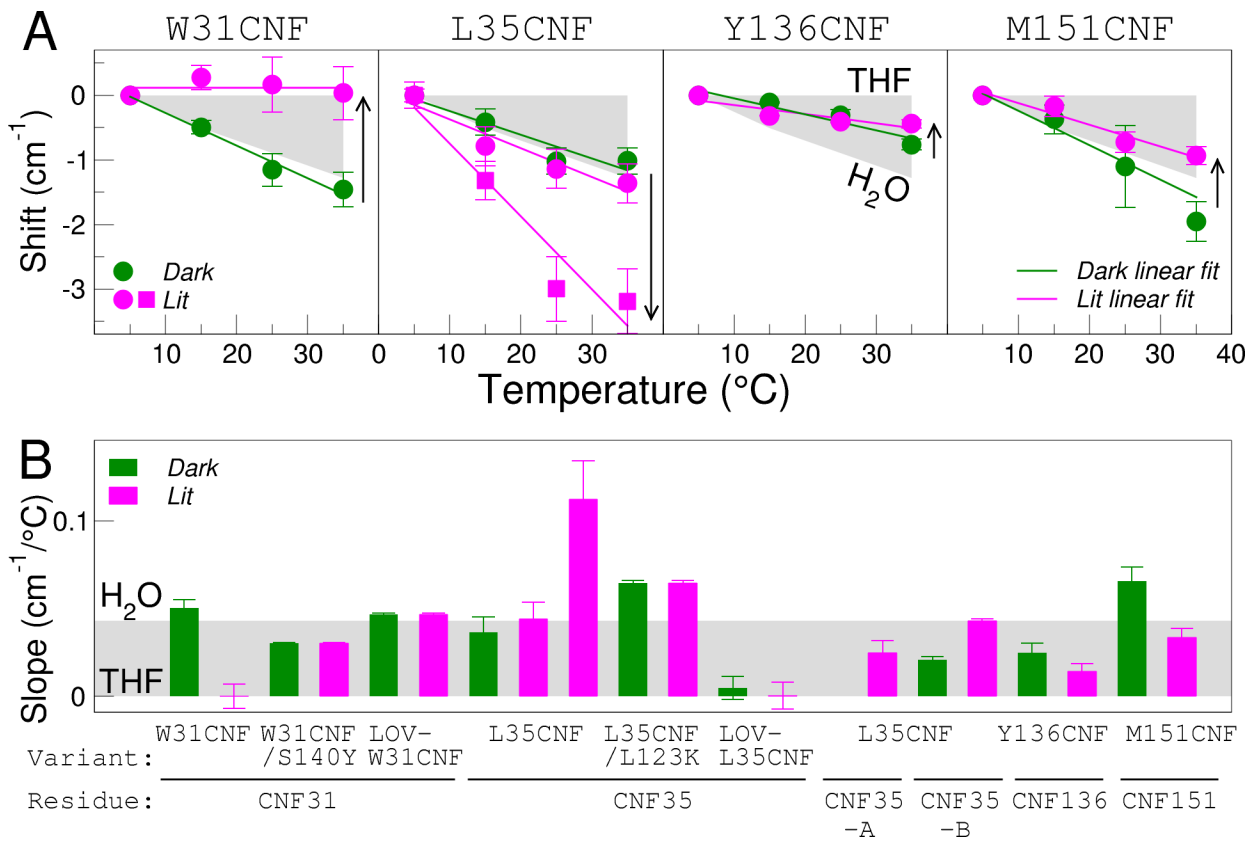


FIGURE 3.

Light-induced changes in solvent exposure and H-bonding around selected CNF residues. (A) Frequency-temperature line slope (FTLS) analyses of the C≡N absorption bands of four single-CNF EL222 variants (indicated on top of the panels). Frequency shifts, calculated with respect to spectra acquired at a temperature of 5 °C, of non-illuminated and illuminated samples are indicated as green and magenta circles, respectively. In the EL222 L35CNF panel, the shifts of the additional lit species found for this mutant are represented as magenta squares. Arrows indicate the magnitude of the change in site-specific hydration for the dark-to-lit transition. **(B)** Slopes calculated from linear fits to FTLS plots for the indicated EL222 variants and CNF residues. EL222 W31CNF/S140Y is a constitutively oligomeric variant (A'α-LOV/A'α-LOV interactions in the absence of illumination). EL222 L35CNF/L123K is constitutively "open" variant (lack of interaction between A'α-LOV and DNA-binding domains in the absence of illumination). LOV-W31CNF and LOV-L35CNF are both truncated versions of EL222 lacking the C-terminal DNA-binding domain. CNF35-A and CNF35-B are the "fast" and "slow" kinetic components of the cyano relaxation of L35CNF variant (see **Fig. 5**). As a reference, the grey area represents the behavior of free CNF dissolved in tetrahydrofuran (THF, zero-slope) versus water (H_2O, negative slope of 0.043 cm^{-1}/°C). Raw data can be found in **Fig. S7**. Error bars represent the standard deviation of 3 independent measurements.

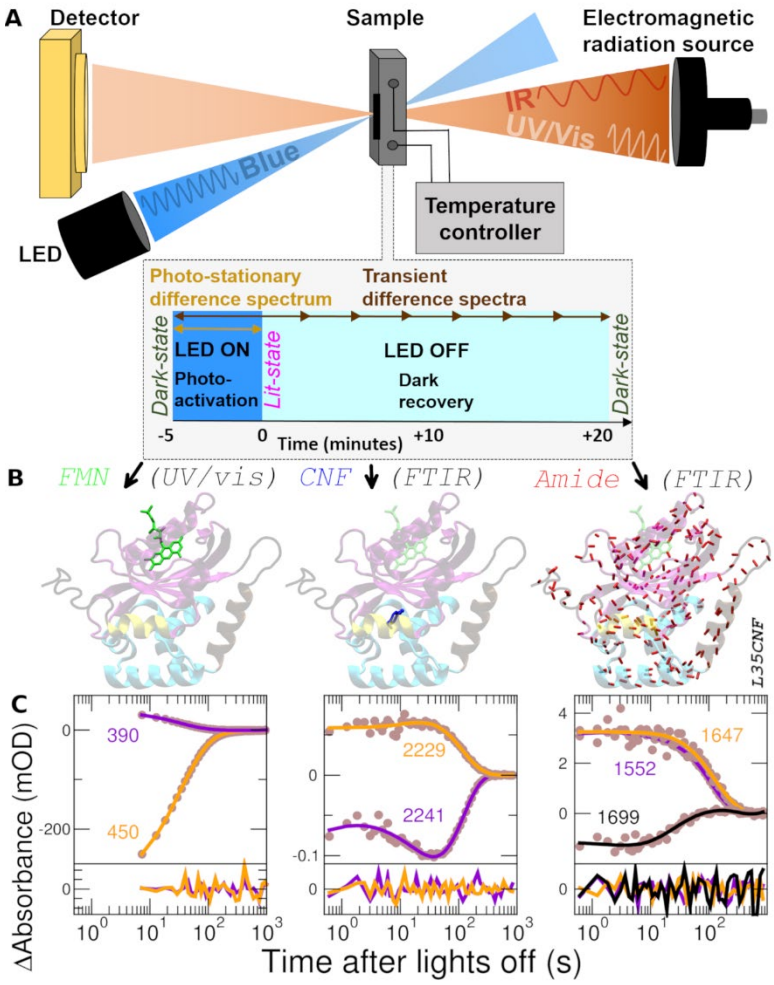


FIGURE 4.

Time-resolved spectroscopy of the lit-to-dark transition of EL222 using multiple probes (FMN, CNF, amide). (A) Data were recorded in solution (typically H_2O-based buffer but a few experiments were conducted with D_2O-based buffer) at 20 °C (except for the Arrhenius plots) by first accumulating the lit state population under continuous blue-light irradiation (5 minutes at 25 mW/cm^2), then switching lights off (time zero) and recording rapid-scan spectra for 10 minutes. The same protein was subjected to three independent experiments. **(B)** *Left*, local chromophore-specific time-resolved UV/Vis spectra (FMN probe, green) were taken between 320 and 550 nm in a cell of 10 mm path-length. *Middle*, local residue-specific time-resolved IR spectra (CNF probe, blue) were taken between 2270 and 2200 cm^{-1} with a cell of 50 µm path-length. *Right*, global protein time-resolved IR spectra (amide probe, mostly backbone carbonyls, red) were taken between 1750 and 1500 cm^{-1} (this region includes also contributions from FMN and other vibrations, see main text) with a cell of 8 µm path-length. Molecular models are color-coded as in the scheme of **Fig. 2B**. **(C)** Kinetic traces of differential absorption (circles) and fits (lines) at two or three selected frequencies as a function of time after ceasing illumination for the three probes (EL222-L35CNF mutant): FMN (left), CNF (middle), and amide (right). Fits were done by maximum entropy lifetime distribution method. Fit residuals are shown in the bottom panels.

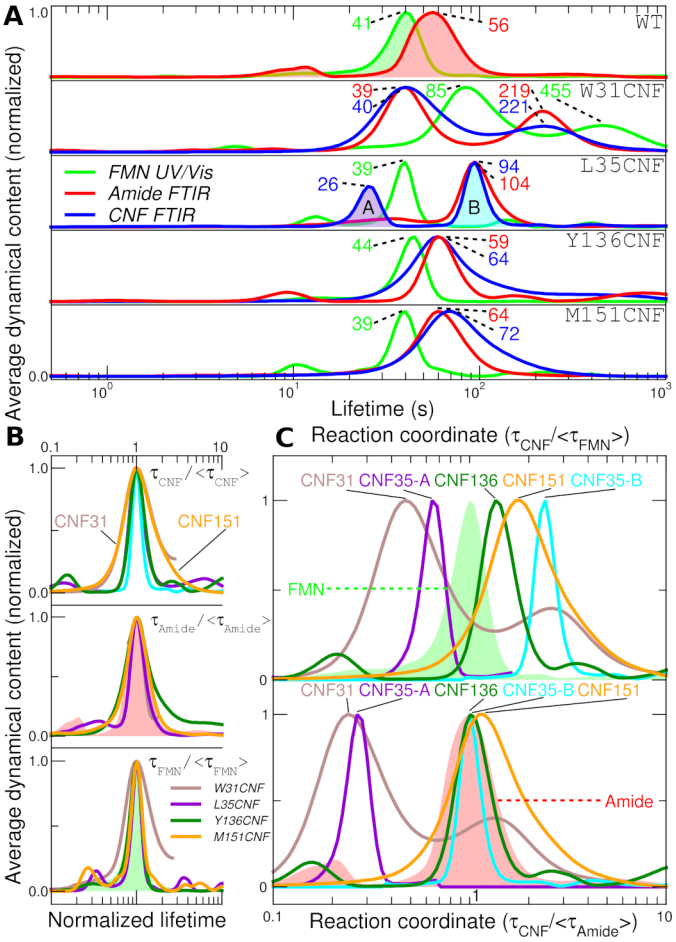


FIGURE 5.

Lifetime distribution analysis (LDA) of the dark recovery kinetics of single-CNF EL222 variants using the maximum entropy method. (A) Normalized average dynamical content (*D*) as a function of lifetime for five variants (wt EL222 and single-CNF W31CNF, L35CNF, Y136CNF, and M151CNF mutants) using three probes: FMN (green), amide (red) and the cyano moiety of the CNF residue (blue). The numbers report the lifetimes of the main dynamical events (mean peak value) again color-coded as green, red and blue for FMN, amide, and CNF probes, respectively. Two clear cases of bimodal lifetime distributions are found for the CNF relaxation kinetics of L35CNF, named CNF35-A (violet-shaded area) and CNF35-B (cyan-shaded area). **(B)** Normalized average dynamical content (*D*) as a function of normalized lifetime calculated as the ratio between a given lifetime and its mean value (peak maxima). The top, middle and bottom panels show the CNF-, amide- and FMN-derived distributions, respectively. CNF31 and CNF151 stand out as the moieties whose distributions have the largest width. **(C)** Normalized average dynamical content as a function of the reaction coordinate calculated as the ratio between a given CNF-derived lifetime and the mean peak value of FMN-derived *D* lifetime distribution (top panel), or the mean peak value of amide-derived *D* lifetime distribution (bottom panel). Mean peak values for all *D* lifetime distributions can be found in **Table S6**. For easy referencing, the *D* lifetime distributions of the amide and FMN native probes of wt EL222 are shown as filled areas in red and green colors, respectively.

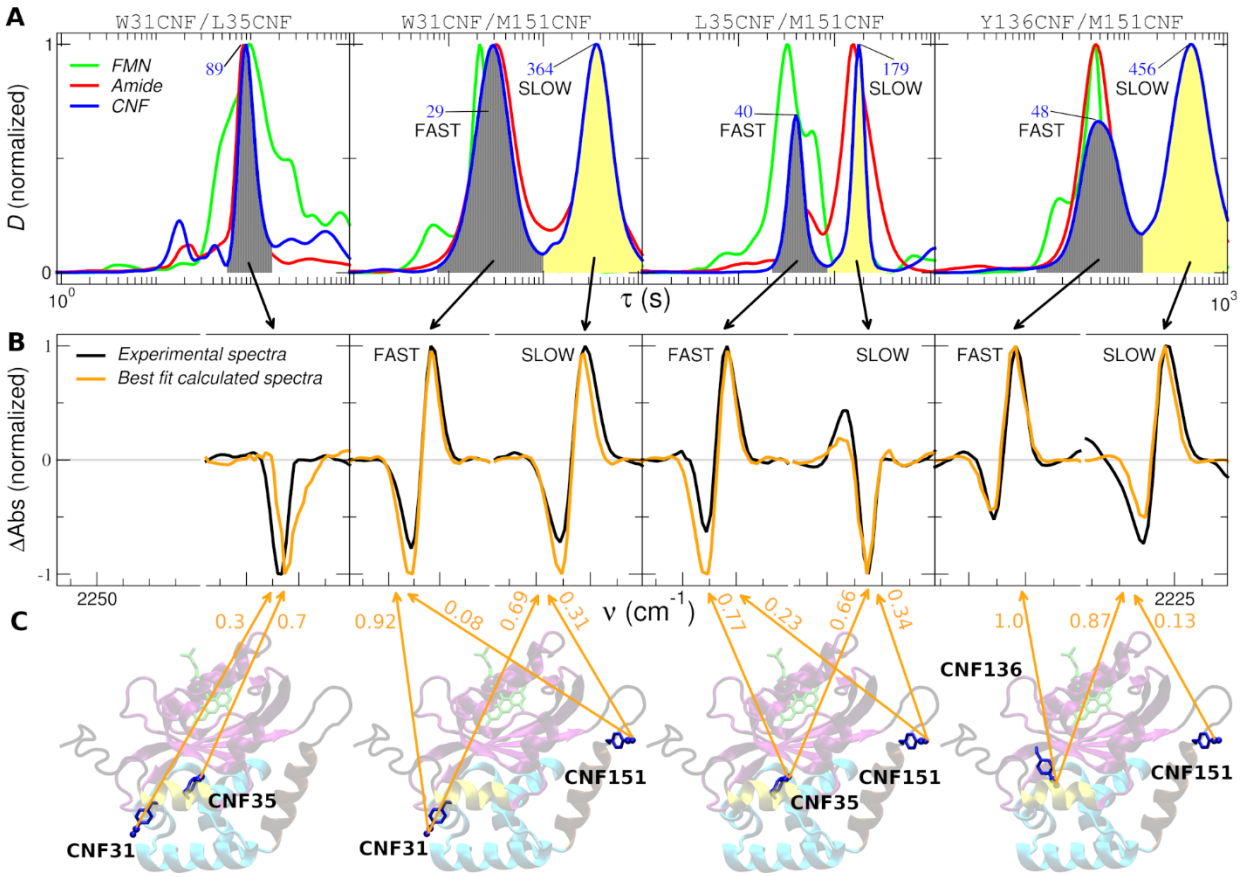


FIGURE 6.

Disentangling residue-by-residue relaxation rates with double-CNF EL222 mutants. (A) Average dynamical content (*D*) as a function of lifetime (τ), derived from lifetime distribution analysis using the maximum entropy method for four double-CNF mutants (W31CNF/L35CNF, W31CNF/M151CNF, L35CNF/M151CNF, and Y136CNF/M151CNF). Curves were calculated for the relaxation of the FMN (green lines), amide bands (red lines), and the cyano moiety of the CNF residue (blue lines). The peak maxima, corresponding to the most probable lifetime (in seconds) are indicated in the case of CNF relaxation. **(B)** The transient decay-associated difference spectra of the kinetic events seen in the lit-to-dark transition of the cyano moiety were derived by integration over the shaded areas (black for the fastest event and yellow for the slowest event). Assuming that the transient spectra of double-CNF mutants are a linear combination of the corresponding single-CNF transient spectra, the relative contribution of each residue was retrieved by minimizing the discrepancy between experimental (black lines) and calculated (orange lines) spectra. The orange numbers on top of the structures (in C) report the fractional contribution of each CNF site to the best-fit calculated spectra. **(C)** Structural models of the double-CNF mutants in the dark state showing the location of the CNF residues and nitrile moiety (Van der Waals representation).

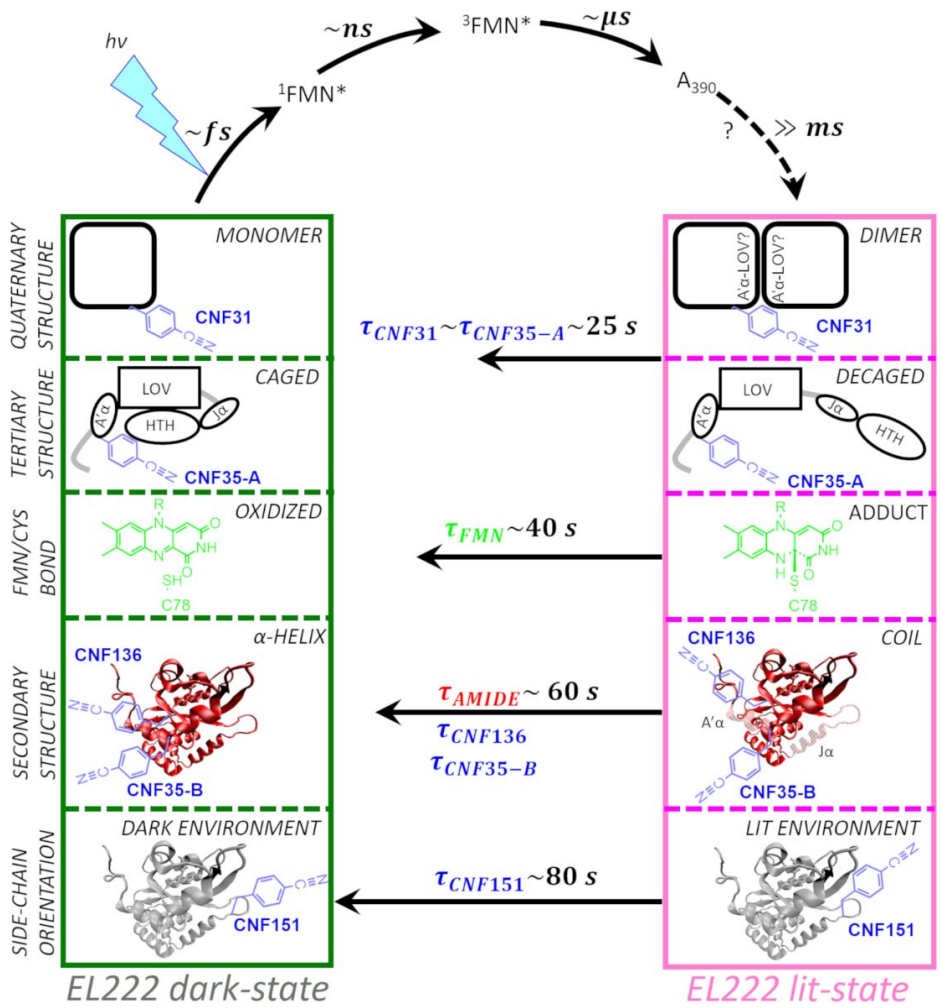


FIGURE 7.

Kinetic-molecular model of EL222 photocycle. The model is defined in terms of five loosely coupled equilibria: monomer/oligomer, caged/decaged conformation, oxidized/C78-bound FMN, caged/decaged, helix/coil, and side-chain dark environment/lit environment. For the dark-to-lit transition, several intermediate species are formed (sequential model): the singlet (^{1}FMN*), the triplet (^{3}FMN*), the adduct (A_{390}), and potentially others. For the lit-to-dark transition, each equilibrium exchanges at a specific rate (parallel model). The relative time scale of events appears to follow the order (from fastest to slowest): 1st) CNF31~CNF35-A (the fastest kinetic event monitored by CNF35), 2nd) FMN, 3rd) Amide~CNF136~CNF35-B (the slowest kinetic event of CNF35), 4th) CNF151. Assuming that CNF31 informs on the A'α-LOV/A'α-LOV interface and that the "fast" CNF35 (CNF35-A) process reports on the A'α-LOV/HTH interface, a molecular interpretation of our results would be as follows. On average (population level), the fastest event is oligomer dissociation and formation of the "caged" conformation. The second step is the rupture of the FMN-C78 adduct. The third event is the increase in α-helical content of the protein backbone, which happens at a similar rate as the equilibration of CNF136 and some CNF35 side-chains. Finally, the slowest phase is the reorganization of the CNF151 side-chain. CNF, FMN, and amide probes are colored blue, green, and red, respectively.