QuantumSi

Control More of Your Protein Research

Introducing Platinum™ *Next-Generation* Protein Sequencer

The power of protein sequencing is in your hands! Now, you can sequence proteins right in your lab with Platinum™, the NEW benchtop solution from Quantum-Si.

Our platform offers single-molecule resolution, a simple workflow and automated, cloud-based data analysis software that delivers proteoform information without added turnaround time or extensive infrastructure.

Platinum provides deeper insights into protein biology to complement your existing "omics" approaches by enabling you to:

- Conduct proteomics experiments in your lab *at your bench*
- Interrogate protein variants and PTMs, and correlate with biological function
- Achieve deeper proteomics insights faster
- Perform analytics with no bioinformatics expertise required

Introducing Platinum™

The Protein Sequencing Company™

FULL-LENGTH PAPER

THE PROTEIN SOCIETY WILEY

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] | Gustavo Fuertes[1]

[1]Institute of Biotechnology of the Czech Academy of Sciences, Czech Republic

[2]Faculty of Science, Charles University, Prague, Czech Republic

[3]Escola Superior de Tecnologia do Barreiro, Instituto Politécnico de Setúbal, Portugal

[4]Department of Cell and Molecular Biology, Uppsala University, Uppsala, Sweden

[5]Institute of Molecular Science, Universitat de València, Paterna, Spain

Correspondence
Víctor A. Lórenz-Fonfría, Institute of Molecular Science, Universitat de València, Carrer del Catedrático José Beltrán Martínez 2, 46980 Paterna, Spain.
Email: victor.lorenz@uv.es

Gustavo Fuertes, Institute of Biotechnology of the Czech Academy of Sciences, Prumyslova 595, 25250 Vestec, Czech Republic.
Email: 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 45 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.

KEYWORDS
flavoproteins, FTIR spectroscopy, genetic code expansion, kinetics, photosensory receptors, protein structural dynamics, signal transduction, site-specific vibrational probes, time-resolved methods, UV/vis spectroscopy

1 | 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 (Kottke et al. 2018). 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 (Nash et al. 2011). 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 (Nash et al. 2011; Takakado et al. 2018; Takakado et al. 2017; Iuliano et al. 2018; Zoltowski et al. 2013). UV/vis spectroscopy has been instrumental in monitoring the photoconversion of FMN (Figures 1 and S1). Non-irradiated LOV proteins absorb maximally at ~450 nm while their irradiated counterparts peak at ~390 nm (Figure 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 via the C4A atom (see FMN atom numbering in Figure S2a) (Maia et al. 2021). FMN-Cys covalent bond formation takes place in the nanosecond to microsecond time scale (Iuliano et al. 2018; Andrikopoulos et al. 2021; Liu et al. 2023). The rupture of such an adduct state, sometimes referred to as A_{390} species (Iuliano et al. 2018), is much slower, typically returning to the resting state in seconds to hours, depending on the particular LOV domain (Lebedev et al. 2014). The dark recovery kinetics of isolated LOV domain suggests synchronous relaxation of chromophore and protein chain (Harper et al. 2004), 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 (Iuliano et al. 2018; Kottke et al. 2017). 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 (Figure 1b bottom panel) (Lorenz-Fonfria 2020). Analysis of the protein amide I band, mostly arising from coupled backbone C=O

stretching vibrations (Yang et al. 2016; Arrondo et al. 1993), 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) (Konold et al. 2016; Goett-Zink et al. 2020). In spite of its utility as a reporter for global structure changes in proteins (Yang et al. 2016; Arrondo et al. 1993), 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 (Figure 1a left panel). As an alternative, chemical groups with absorption frequencies in the range from ~1,800 cm^{-1} to ~2,500 cm^{-1}, known as the "transparent window" because most natural proteins do not absorb there (Figure 1b gray area), make the assignment straightforward while providing valuable site-specific information (Lim et al. 1998; Fayer 2001; Getahun et al. 2003; Suydam and Boxer 2003; Adhikary et al. 2017; Thielges 2021). These extrinsic non-native moieties, which allow researchers to spectrally single out individual residues, can be incorporated into proteins using a variety of methods (Fafarman et al. 2006; Oh et al. 2008; Taskent-Sezgin et al. 2010). 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 (Manandhar et al. 2021; Schultz et al. 2006). 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 Figure 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 (Adhikary et al. 2017; Thielges 2021; Löffler et al. 2022). Additionally, the system can be designed so that the included ncAA tunes protein functions and helps elucidating reaction mechanisms (Tolentino Collado et al. 2022). Up-to-date, the three most popular ncAA used as IR reporters have been 4-cyanophenylalanine (CNF) (Getahun et al. 2003; Kraskov et al. 2021), 4-azidophenylalanine (Ye et al. 2009; Kurttila et al. 2021), and cyanocysteine (Fafarman et al. 2006; Blankenburg et al. 2019). Thiocyanates, which have the smallest
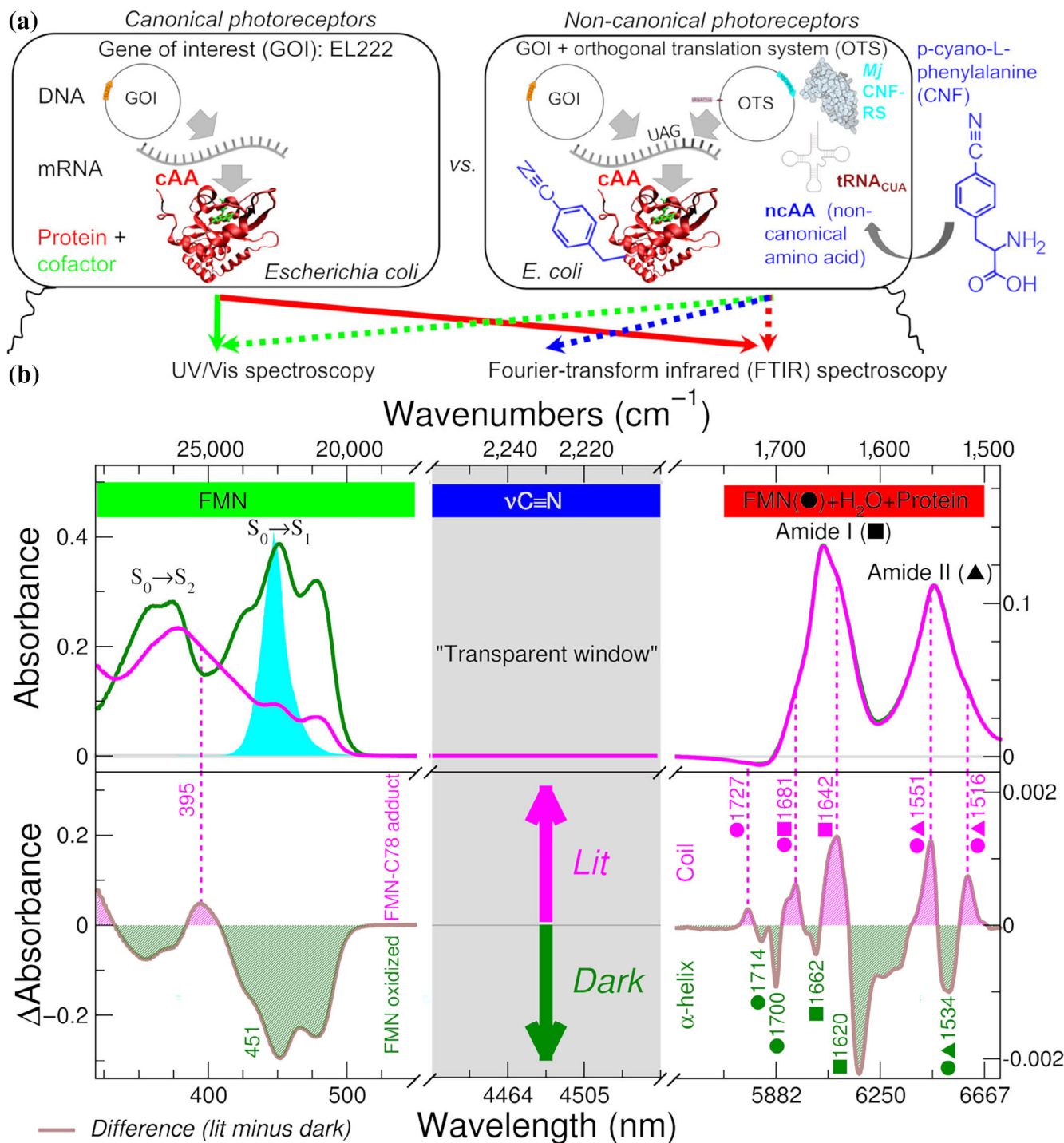
FIGURE 1 Protein engineering via genetic code expansion meets electronic/vibrational spectroscopy. (a) Comparison between photoreceptors based on canonical (left) and non-canonical (right) amino acids (ncAA) carrying vibrational probes. The ncAA p-cyano-L-phenylalanine (CNF) contains a triple bond C≡N group vibrating in the "transparent window" region (gray 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. Absorbance 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.

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 (Milles et al. 2012), and their complex

absorption profile makes interpretation challenging (Zhang et al. 2018). 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 (Figure 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 (Kurttila et al. 2021; Blankenburg et al. 2019). 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 were aimed at testing the role of specific residues in the photocycle (according to a defined working hypothesis) and have therefore focused mainly on a small number of residues and mainly proximal to the chromophore site (Kraskov et al. 2021; Ye et al. 2009; Kurttila et al. 2021; Blankenburg et al. 2019; Krause et al. 2019; Hall et al. 2019).

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 (Table S1). 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.

2 | RESULTS

2.1 | Preparation of EL222 variants and initial screening

For amber suppression in *Escherichia coli* we used an orthogonal translation system (OTS) comprised of an orthogonal aminoacyl-tRNA synthetase/tRNA pair derived from tyrosyl tRNA synthetase of the archaea *Methanocaldococcus jannaschii* (*Mj*CNFRS) and its cognate tRNA (Figure 1a and Table S2). The modified suppressor tRNA incorporates the 5′-CUA-3′ anticodon (tRNA$_{CUA}$). The enzyme is able to attach phenylalanine derivatives (CNF and others) to the suppressor tRNA$_{CUA}$, thus allowing decoding UAG codons during protein translation (Schultz et al. 2006; Miyake-Stoner et al. 2009). All proteins engineered using this strategy (see sequences in Table S2) were purified to >95%. Mass spectrometry analyses confirmed the successful incorporation of CNF in all 45 single-mutant proteins (Table S3). Although the fidelity of *Mj*CNFRS is known to be high, mis-incorporation of phenylalanine is still possible. Mass spectrometry confirmed that mis-incorporation of Phe was only significant in two mutants (W89CNF, I94CNF), where the phenylalanine-carrying EL222 was the most abundant species detected. In these cases, we minimized the presence of Phe by re-expression of these constructs in minimal media (Table S3). In another mutant (A42CNF) we only found evidence of truncated products, which cannot be attributed to the UAG codon being recognized as STOP since the purification tag is found at the C-terminus, but to post-translational protein instability enhanced by the introduced mutation, consistently with the known tendency of EL222 to undergo proteolytic degradation under certain conditions (Nash et al. 2011). Prior to the kinetic studies, several rounds of mutant screening were conducted to select the most suitable candidates (Note S1 and Figure S3). In the first selection round, two criteria were employed: the degree of FMN incorporation, and the oligomeric state in the dark (Figure S4). Wt EL222 binds stoichiometric amounts of FMN and it is a monomer in the absence of illumination (Table S3). Thus, we discarded mutants with no or little FMN incorporation, and variants with oligomeric species in the absence of illumination.

2.2 | Steady-state spectroscopy

We measured EL222 samples 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/vis) region, with electronic transitions from the FMN chromophore ($S_0 \rightarrow S_1$ and $S_0 \rightarrow S_2$, Figure 1b top left panel). The second is the 1,750 to 1,500 cm^{-1} (mid-IR) region, which includes vibrational transitions from the protein

backbone (amide I and II), FMN (C=O stretching among others) and water (Figure 1b top right panel). The third region, also in the mid-IR but around 2,230 cm^{-1}, includes the C≡N stretching from the selectively introduced CNF. Peaks in the latter region are absent in wt EL222 (compare Figures 1 and 2). The similarity between the UV/vis dark spectra of CNF variants relative to the wt was used as a selection criterion for the second round of screening (Note S1 and Figure S5).

2.2.1 | Infrared absorption region of native protein and flavin mononucleotide

The main band in this region comes from amide I vibrations (1,700–1,600 cm^{-1}), primarily coupled C=O vibrations from the peptide backbone, known for its secondary structure sensitivity (Krimm and Bandekar 1986). The nearby amide II band (1,600–1,500 cm^{-1}) is due to peptide backbone C—N stretching and N—H bending modes (Barth 2007). The two carbonyl stretching modes of flavins (C2=O2 and C4=O2, see Figure S2a) also show up in the 1,750–1,650 cm^{-1} spectral region (Andrikopoulos et al. 2021; Andrikopoulos et al. 2020; Spexard et al. 2011). Absolute FTIR spectra in the 1,750–1,500 cm^{-1} region of dark-adapted and lit-adapted wt EL222 in H$_2$O are hardly distinguishable (Figure 1b top right). In contrast, lit-minus-dark difference spectra show significant changes between both (Figure 1b bottom right), although some spectra might be distorted by the strong background absorbance from liquid water, a potential problem that can be circumvented by dissolving the protein into deuterated water (D$_2$O). Absolute and difference IR spectra of wt EL222 in D$_2$O are shown in Figure S1a. Some band interpretations in the amide I′ (I in H$_2$O) region are possible (see Note S2 for further assignments). The negative band (bleach) at 1,647 cm^{-1} (−) (1,662 cm^{-1} in H$_2$O) can be attributed to loss of α-helices, with concomitant positive bands at 1,661 cm^{-1} (+) and 1,630 cm^{-1} (+) (1,681 and 1,642 cm^{-1} in H$_2$O) reflecting gain in unordered structures and/or β-sheet elements. Thus, from the point of view of secondary structure changes in the protein, IR spectroscopy suggests a light-induced helix-to-coil transition *somewhere* in EL222. In other LOV domains, a similar IR pattern has been interpreted as unwinding/unfolding of the Jα helix (Konold et al. 2016; Herman et al. 2013), and unfolding of the A′α helix (Herman and Kottke 2015). Hence, we assume that in the EL222 protein, the Jα and A′α helices are, individually or in combination, the likely candidates to unfold upon blue-light irradiation.

2.2.2 | Infrared absorption region of genetically encoded nitrile groups

We tagged EL222 with cyano probes at 45 different positions. For the 40 single-CNF EL222 mutants able to incorporate substantial amounts of FMN we measured absolute IR spectra for the C≡N stretching, which we converted to extinction coefficient IR spectra (Figure 2a top panels). 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 nitrile stretching bands using three parameters: (i) the peak maximum (ν_{max}), (ii) the local electric field projected on the C≡N bond ($|\vec{F}|$), and (iii) the non-electrostatic H-bonding contribution to the nitrile vibration frequency ($\Delta\nu_{HB}$) (Note S3 and Table S4) (Weaver et al. 2022). Peak maxima span ∼14 cm^{-1}, from 2,223 cm^{-1} for CNF120 to 2,237 cm^{-1} for CNF68, a variability similar to that reported before for nitrile groups in other proteins (∼18 cm^{-1}) (First et al. 2018). On the other hand, the local electric field spans from −108 MV/cm for CNF137 to −16 MV/cm for CNF149. CNF120 is also the residue whose peak maximum is the least affected by H-bonding effects (0 cm^{-1}), while in the other extreme we have CNF137 (+24 cm^{-1}). Indeed, CNF120 is the lone residue whose peak position can be predicted by electric field-frequency calibration curves valid for non-protic solvents (Figure S6) suggesting that all other CNF residues of EL222 are engaged in H-bonding to one extent or another. Thus, although there is no consensus among the three metrics, we can conclude that CNF120 is one the residues that experiences a more apolar and non-H-bonding environment, while CNF137 is one of the residues that senses a more polar and H-bonding environment.

To observe spectral changes upon light stimulation we recorded difference IR spectra in the nitrile stretching region for each individual EL222 mutant (Figure 2a bottom panels). Mutants were classified in three categories based on the light-induced frequency shift ($\Delta\nu_{max}$), change in the local electric field ($\Delta|\vec{F}|$), and change in H-bonding status ($\Delta\Delta\nu_{HB}$) (Note S3 and Table S4). Sixteen mutants experienced a red-shift in ν_{max}, that is, the illuminated samples had lower mean frequency for its nitrile group after illumination. Six proteins experienced, in contrast, a blue-shift, and 17 mutants did not show significant frequency shifts (below 0.3 cm^{-1}). For both red-
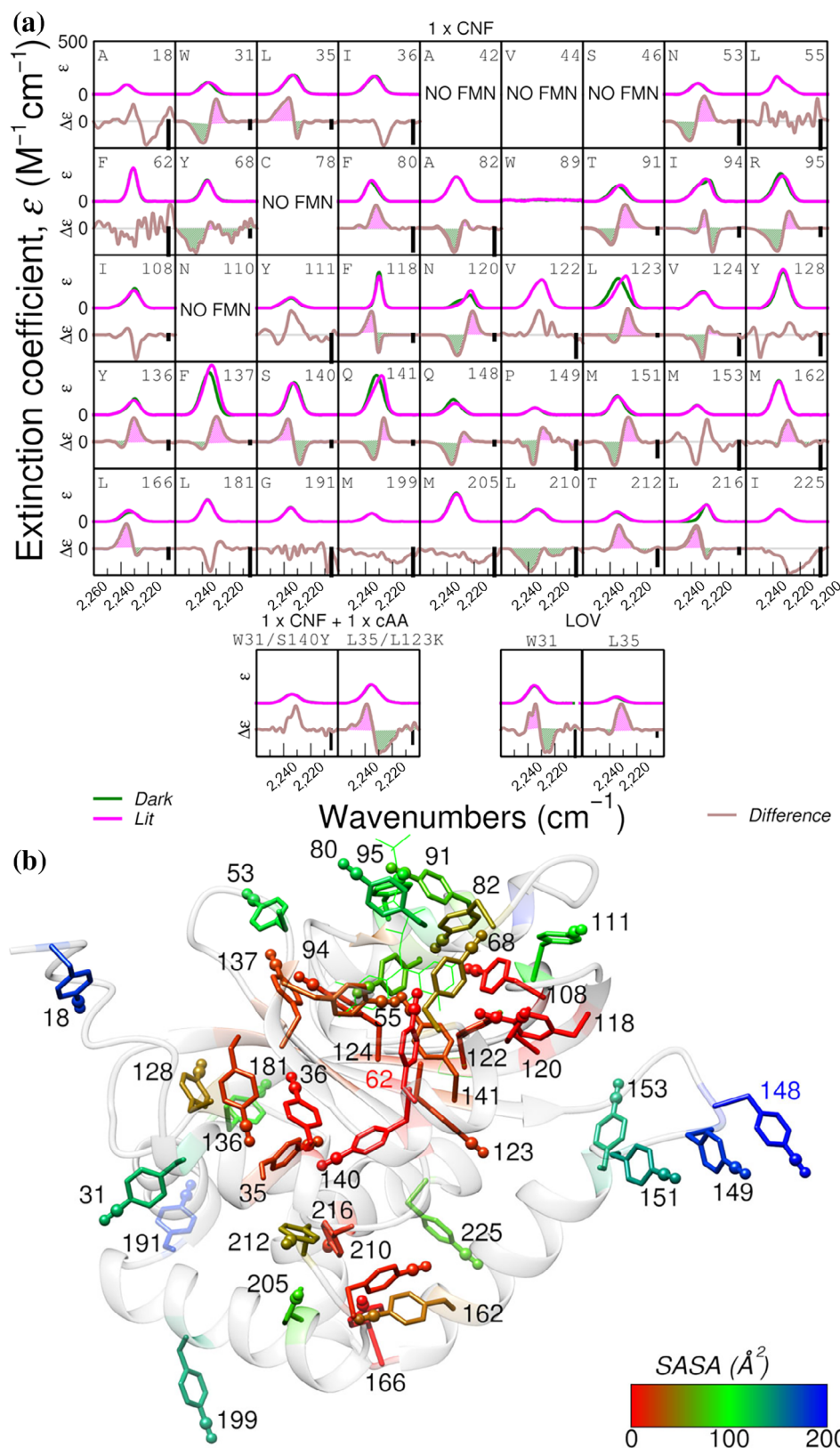
FIGURE 2 Site-specific infrared spectroscopy of EL222 mutants in the dark and lit states. (a) Absolute absorption spectra (top) and steady-state light-induced difference spectra (bottom) in the nitrile absorption region, normalized by concentration (extinction coefficient spectra). Each of the 45 single-CNF EL222 mutants carries a CNF in place of the indicated native residue. Two mixed mutants and two single-CNF mutants of an EL222 variant lacking the HTH domain (referred to as LOV) are shown in the bottom row. Negative bands in the difference spectra correspond to the dark state (green areas) while positive bands indicate the lit state (magenta areas). Difference spectra without shaded areas indicate absence of significant changes (<0.3 cm^{-1} frequency shift or <2 MV/cm electric field difference). The vertical scale bar (black) amounts to 2.5 mM^{-1} cm^{-1} (b) Molecular model of EL222 highlighting the approximate location of the CNF residues colored according to their surface-accessible solvent area (SASA).

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 (<2 MV/cm difference). For most residues $\Delta\Delta\nu_{HB}$ was ≤ 1 cm^{-1}, suggesting no significant light-induced changes in their H-bonding interactions, CNF123 being one notable exception.

2.3 | Visualization and interpretation of *p*-cyanophenylalanine spectra

2.3.1 | Mapping absolute and light-induced *p*-cyanophenylalanine-derived shifts

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 S4 and Table S5), and mapped the surface accessible solvent area of the CNF residues to quantify their exposure to water (Figure 2b and Table S5). To visualize the microenvironment for each CNF in a more global context, we mapped the values for absorption frequencies, local electric fields, and H-bonding contributions to the vibration frequency in the dark on the sequence and overall fold of EL222 (Figure S7). We did a similar mapping to visualize light-induced changes in vibration frequencies, local electric fields, and H-bonding contributions to the vibration frequency (Figure S8). We found that the magnitude of $\Delta\nu_{max}$, $\Delta|\vec{F}|$ or $\Delta\Delta\nu_{HB}$ does not correlate with proximity to the light-harvesting site (the FMN, Figure S9). See further discussion in Note S5.

2.3.2 | Toward a molecular-level interpretation of nitrile frequencies

Although introducing CNF groups at many different positions in EL222 provided us with valuable information about photoinduced changes on the local electric field and H-bonding, no clear pattern emerged. Another established strategy to identify changes in solvent accessibility of nitrile groups, even in the dynamic and complex environment found in a protein (First et al. 2018), is through frequency–temperature line slope (FTLS) plots (Adhikary et al. 2015). 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}/$^\circ$C ($\sim$1 cm^{-1} downshift every 24°C increase) (First et al. 2018). Thus, the slope in FTLS plots for CNF groups reports on their degree of exposure to water, with a zero slope indicative of a buried residue (Adhikary et al. 2015). A disadvantage of this method is that it requires spectral measurements at different temperatures and is therefore time-consuming and inconvenient to apply to many residues.

For FTLS experiments we concentrated on four residues located in different structural elements of EL222 that show a wt-like FMN spectrum and display clear light-induced IR spectral differences in the CNF region: 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 Figure S10. 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 (Figure 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 become less exposed to water upon perturbation by light. On the other hand, a large change in slope occurring between the dark and lit states of EL222 W31CNF (Figure 3b) indicates a dramatic change in polarity around the residue 31, from a slope corresponding to full solvation in the dark state to a null slope indicative of a solvent-excluded residue in the lit state. Since EL222 is known to oligomerize in the lit state (Takakado et al. 2018; Takakado et al. 2017; Zoltowski et al. 2013), we suggest that CNF31, and by extension W31, participates in a protein–protein interaction surface (Figure 3c). In fact, the dimer interface seen in the X-ray structures of three LOV photosensors crystallized in the lit state is largely constituted by interactions involving the A'α helix (Vaidya et al. 2011; Bocola et al. 2015; Heintz and Schlichting 2016). To test such hypothesis, we designed one EL222 mutant as positive (constitutive oligomerization) and one as negative (diminished oligomerization) controls. First, we prepared the S140Y mutation, known to promote EL222 oligomerization already in the dark-adapted state (Zoltowski et al. 2013), in combination with W31CNF as a reporter (Figure 3c). We found that the EL222 W31CNF/S140Y mutant had enhanced oligomerization propensity in the absence of illumination with respect to the wt version (Table S3), and its light-induced difference IR spectra in the C≡N region did not show appreciable changes (Figure 2a). Moreover, FTLS analysis of W31CNF/S140Y displayed a reduced slope compared to W31CNF, suggesting a lower degree of solvent accessibility around CNF31 in the dark state of the EL222 W31CNF/S140Y variant compared to the dark state of EL222 W31CNF, albeit not as dehydrated as the lit state
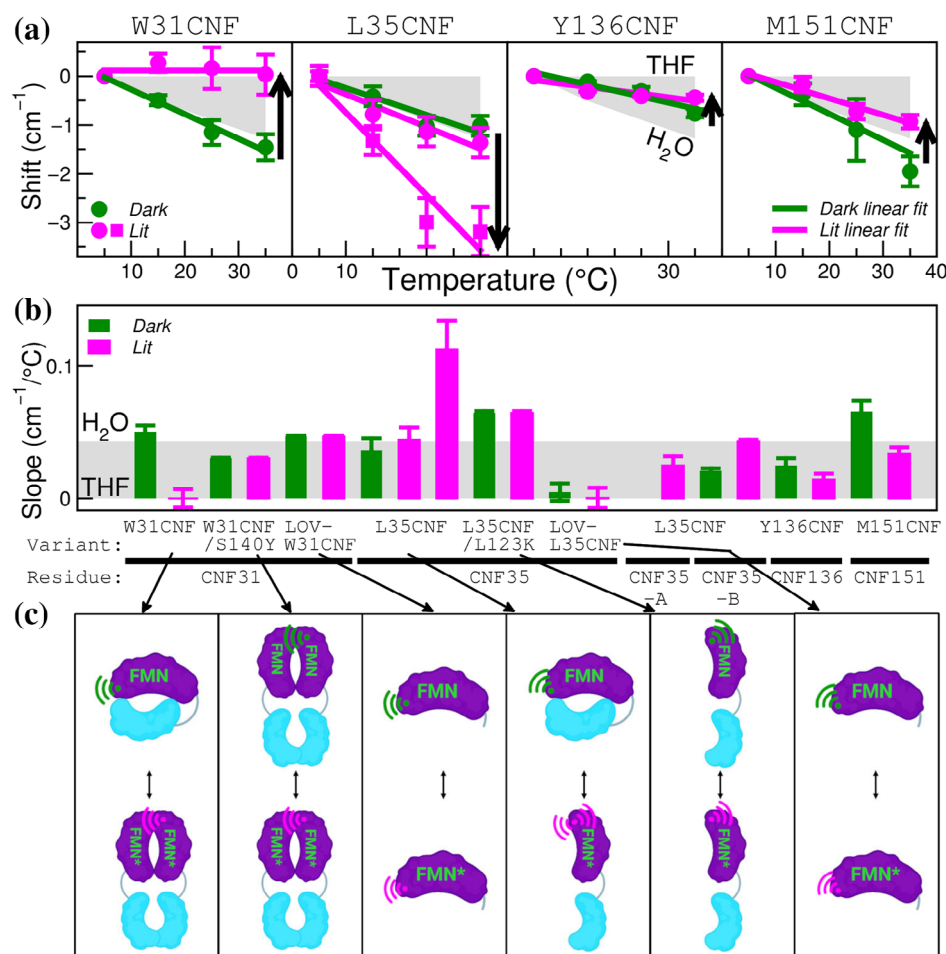
FIGURE 3 Light-induced changes in solvent exposure and H-bonding around selected CNF residues. (a) Frequency–temperature line slope (FTLS) analyses of the C≡N absorption bands of four single-CNF EL222 variants (indicated on top of the panels). Frequency shifts, calculated with respect to spectra acquired 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 (Figure 5). As a reference, the gray area represents the behavior of free CNF dissolved in tetrahydrofuran (THF, zero-slope) versus water (H_2O, negative slope of 0.043 $cm^{-1}/°C$). Error bars represent the standard deviation of 3 independent measurements. (c) Cartoons illustrating the expected behavior in dark (top) and lit (bottom) states of 6 mutants shown in panel B.

of EL222 W31CNF (Figure 3b). As a second control, we produced a truncated EL222 lacking the HTH domain, known to drive homo-dimerization (Nash et al. 2011), again including the mutation W31CNF as a reporter (LOV-W31CNF) (Figure 3c). This variant was mostly monomeric in the dark (Table S3) and showed small light-induced spectral changes (Figure 2a). Accordingly, the slope in the FTLS experiments was similar between LOV-W31CNF and the parental two-domain EL222 W31CNF mutant in the dark state (Figure 3b). Taken altogether, these results are consistent with the nitrile

moiety at position 31 reporting primarily on the oligomerization state of EL222.

In the case of the L35CNF mutant, the observed asymmetry in the C≡N band (Figure 2a), possibly indicating the presence of more than one CNF environment, complicated the analysis. Indeed, three bands (one negative and two positive) were clearly resolved at all temperatures in the photostationary light-minus-dark difference IR spectra of EL222 L35CNF (Figure S10). We fitted multiple Gaussian distributions to the difference spectra of L35CNF, and used the mean peak positions to construct

FTLS plots (Figure S10). The slope of the negative band (corresponding to the dark state) agrees with a fully water-exposed residue. The slope of one of the positive bands was similar to the slope of the negative band, corresponding to a CNF environment fully water-exposed, while the other positive band had a steeper slope, more than twice that reported before for model CNF groups dissolved in water (Figure 3b). Therefore, we conclude that in one of the two environments (or conformations) of L35CNF populated in the light, the CNF probe experiences a stronger hydrogen-bonding environment than in liquid water. Because L35 resides in A'α element (residues 30–37) and its side-chain lies at the interface between the A'α and HTH domains (Figure S1b), we asked ourselves whether CNF35 could be sensing the proximity between the A'α-LOV and HTH halves of EL222 (Figure 3c). To test such hypothesis, we designed two EL222 variants as negative controls. First, we prepared a double mutant containing L35CNF and L123K. The 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 (Nash et al. 2011) (Figure 3c). 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 Figure 2a. In contrast to L35CNF, FTLS analysis of L35CNF/L123K showed no clear change in slope between dark and lit states (Figure 3b). Secondly, we introduced the L35CNF mutation in an EL222 variant lacking the HTH domain (LOV-L35CNF, Figure 3c). There was evidence of only a single lit environment for residue CNF35 in LOV-L35CNF (Figure 2a), which did not differ significantly from the dark-state (Figure 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.

2.4 | Time-resolved spectroscopy of EL222

To characterize the dynamics of the transition from lit to dark states with single-residue resolution, we measured the photorecovery kinetics of EL222 using three different spectral regions (Note S6): 320–550 nm (FMN probe), $1,500$–$1,750$ cm^{-1} (amide probe), and $2,200$–$2,270$ cm^{-1} (CNF probe) (Figure 4a). The UV/vis spectra mainly reveal adduct rupture (A_{390}-to-D_{450} transition), bands in the $1,500$–$1,750$ 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) (Figure 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 senses local changes around 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 (Figure 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. 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.

2.4.1 | WT EL222

Time-resolved UV/vis and IR spectra for wt EL222 in H_2O and D_2O buffer are reproduced in Figures S11 and S12 (panel a). Experimental traces were analyzed by maximum entropy lifetime distribution analysis (LDA) (Lórenz-Fonfría and Kandori 2016). 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 (van Stokkum et al. 2004). 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 (Davis et al. 2018). An example of experimental and fitted time traces can be found in Figure 4c. The output of LDA is typically a lifetime density map (Figures S11 and S12 panel b) where each point represents the exponential amplitude for a given lifetime and frequency. One way to summarize the information content of a lifetime density map is to calculate the square root of the sum over the squared amplitudes for all frequencies, providing a lifetime distribution known as the "average dynamical content", abbreviated as D (Note S3)
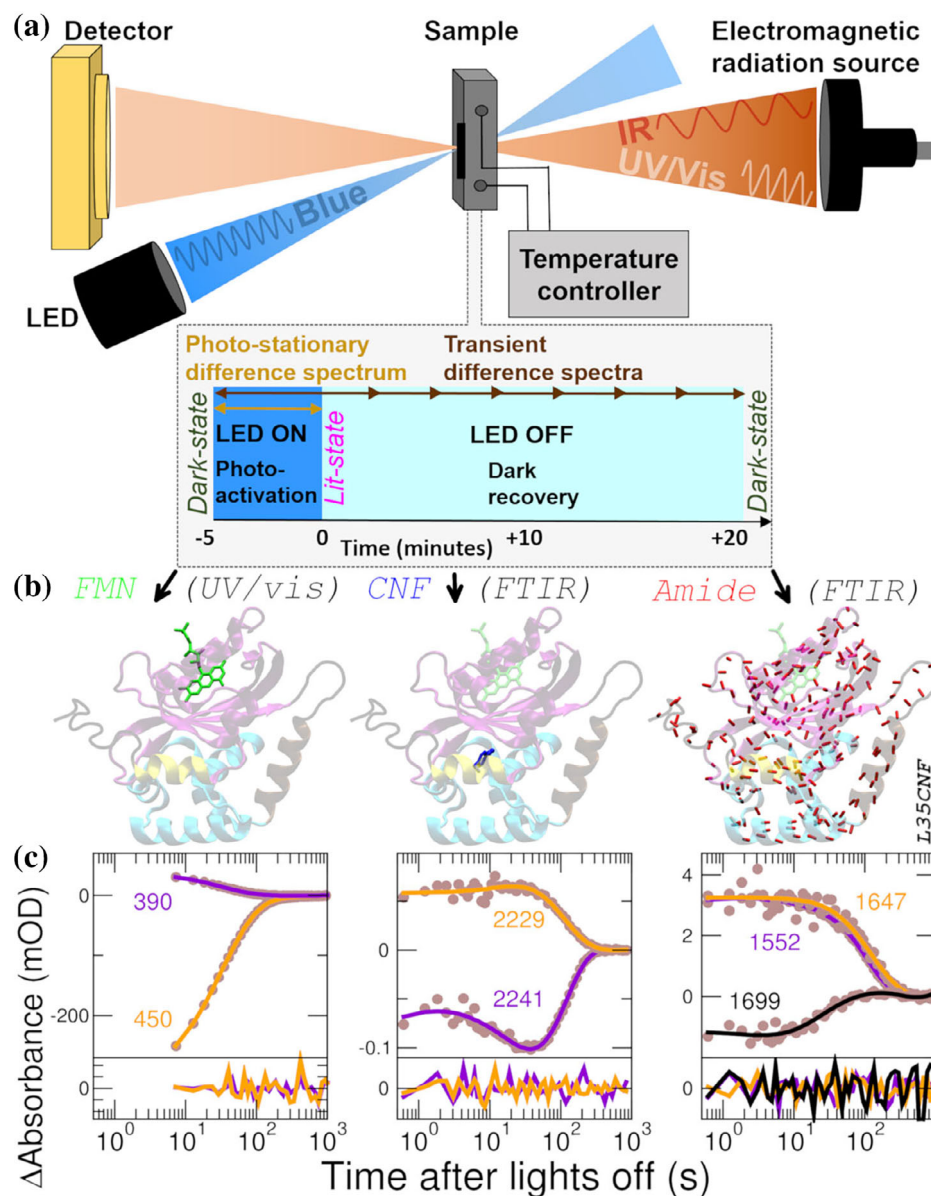
FIGURE 4 Time-resolved spectroscopy of the lit-to-dark transition of EL222 using multiple probes (FMN, CNF, amide). (a) Data were recorded in solution (typically H_2O-based buffer but a few experiments were conducted with D_2O-based buffer) at 20°C (except for the Arrhenius plots) by first accumulating the lit state population under continuous blue-light irradiation (3 min at 25 mW/cm^2), then switching lights off (time zero) and recording rapid-scan spectra for 10 min. 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 2,270 and 2,200 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 1,750 and 1,500 cm^{-1} (this region includes also contributions from FMN and other vibrations, see main text) with a cell of 6 μm path-length. (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.

(Stock and Hamm 2018). The amplitudes of *D* lifetime distributions provide a visual summary about the magnitude of the spectroscopic changes that occur at a given time constant. *D* lifetime distributions for the UV/vis and IR data are plotted in Figure 5a (top panel) and lifetimes

(τ) with the largest *D* values are listed in Table S6. For comparison, we also include in Table S6 the lifetimes extracted from global mono- or bi-exponential fits. In general, both methods yielded similar values but we focused on LDA since it provides a more direct
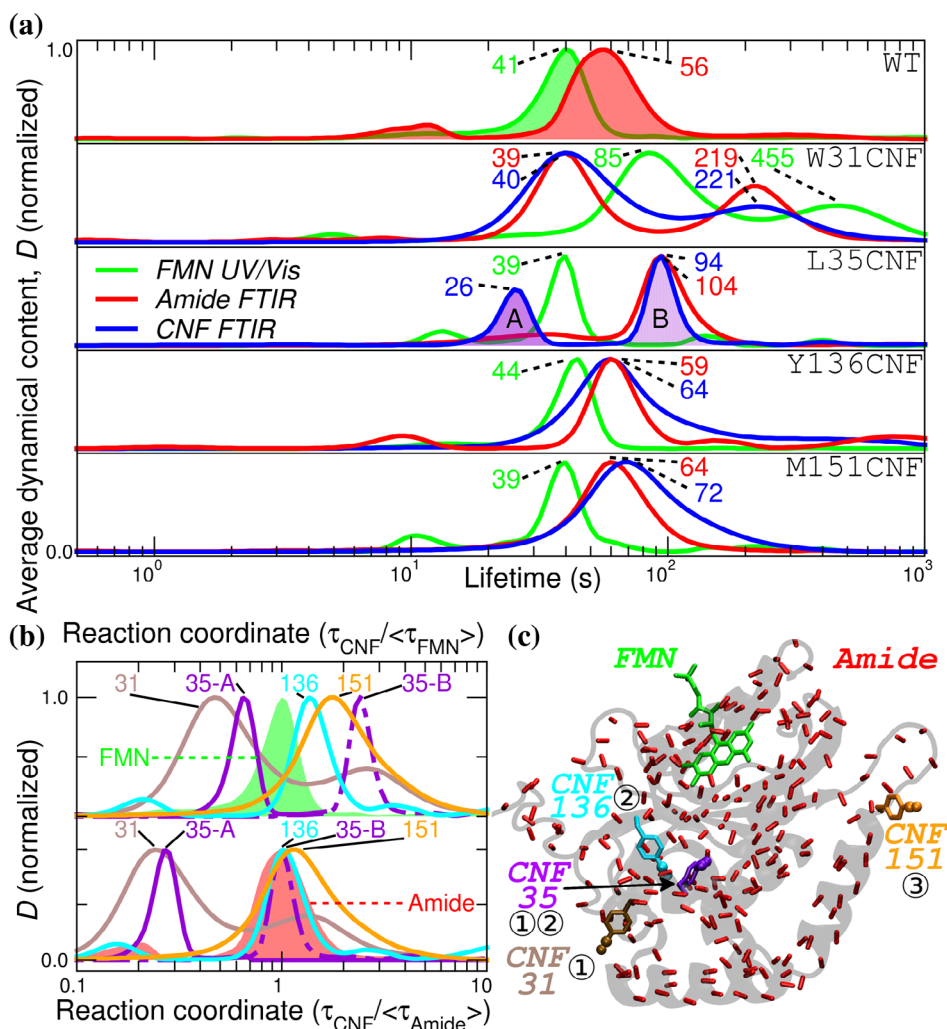
FIGURE 5 Lifetime distribution analysis (LDA) of the dark recovery kinetics of single-CNF EL222 variants using the maximum entropy method. (a) Normalized average dynamical content (D) as a function of lifetime for five variants (wt EL222 and single-CNF W31CNF, L35CNF, Y136CNF, and M151CNF mutants) using 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. Clear cases of bimodal lifetime distributions are found for the CNF relaxation kinetics of L35CNF, named A and B (violet-shaded areas), and for W31CNF (all 3 probes). (b) 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). 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. (c) Molecular model of EL222 showing the approximate location of residues CNF31 (brown), CNF35 (violet), CNF136 (cyan) and CNF151 (orange). The numbers indicate the relative order of dark-state recovery.

visualization of the time scales at which spectral changes occur without any assumption about the underlying model.

Regarding the FTIR data, we also obtained D lifetime distributions for three spectral sub-ranges: (a) 1,750–1,680 cm^{-1}, which is dominated by the two bands from FMN carbonyls, (b) 1,680–1,600 cm^{-1}, with bands mainly ascribable to the protein backbone amide I mode, and (c) 1,600–1,500 cm^{-1}, which has mixed contributions from the protein backbone amide II bands and FMN ring

modes. For wt EL222 in H$_2$O, one dominant photorecovery time constant was found for all triplicates at 41 s for UV/vis (320–550 nm), 44 s for FMN C=O (1,750–1,680 cm^{-1}), 47 s for the Amide II (1,600–1,500 cm^{-1}), and 59 s for the Amide I (1,680–1,600 cm^{-1}) (Figure S13a top).

Although the recovery lifetime in the UV/vis (τ_{FMN}) clearly preceded that in the amide I region (τ_{amide}), it might be unclear if this was a reliable feature or induced by the higher noise and baseline fluctuations in the latter

region caused by the strong background absorbance from liquid water. To resolve this point, as well as to gain more insight into the mechanism of dark reversion, we repeated the kinetics as a function of multiple parameters (Note S7): solvent (Figure S13a bottom), temperature (Figure S14 and Table S7), protein concentration (Figure S15), and sequence (Figure S13b).

Three key conclusions can be extracted from time-resolved studies using the native chromophores present in wt EL222 and its LOV domain. The first is that recovery kinetics manifest a single dominant kinetic process, although additional but minor events cannot be excluded. The second is that the main kinetic event reported by UV/vis spectroscopy (FMN-C78 adduct rupture) seems to precede the main kinetic event probed by IR spectroscopy (global protein refolding to the dark-state conformation), at least at temperatures below 35°C (Figure 5a top row). The third is that sub-dividing the 1,750–1,500 cm^{-1} IR region into smaller zones does not lead to a clear increase in the information content, that is, the number of kinetic components stays unchanged. We next asked whether the non-native CNF probes could resolve additional kinetic events in the lit-to-dark transition of EL222.

2.4.2 | Single-*p*-cyanophenylalanine EL222 mutants

As a result of the third (and last) round of screening based on the FMN recovery kinetics (Note S1, Table S8, and Figure S16), we performed time-resolved UV/vis and IR experiments with five mutants (L35CNF, N53CNF, Y136CNF, M151CNF, and L216CNF). Exceptionally, we included also the W31CNF variant because, although the kinetics of its FMN group is slowed relative to the wt, the kinetics of its nitrile group may be sensitive to oligomer dissociation, as we concluded in static FTLS experiments.

We studied the dark recovery kinetics of the six selected EL222 CNF mutants using two different IR frequency ranges: 1,750–1,500 cm^{-1}, plus 2,270–2,200 cm^{-1} (Figure 4c for the case of L35CNF mutant). Complete kinetic traces for the six single-CNF mutants can be found in Figures S17–S22 (panel a). Most single-CNF mutants displayed a single dominant component in the CNF-derived *D* lifetime distributions, equivalent to a mono-exponential behavior, which was used to calculate the recovery lifetime of the CNF moiety (τ_{CNF}) (Figure 5a and Table S6). In these cases, the transient decay-associated difference spectra were similar to the steady-state difference spectra (Figure S23a for the nitrile absorption region and Figure S23b for the amide bands). Unlike wt EL222, two mutants (N53CNF and L216CNF,

Figure S24) showed a similar recovery kinetics for all three probes (FMN, amide, and CNF), suggesting a mechanistic perturbation induced by the CNF label. For the remaining four CNF variants (W31CNF, L35CNF, Y136CNF, and M151CNF), the FMN recovery happens before the protein backbone recovery, similar to the behavior observed for wt EL222 (Figure 5a and Table S6). Two mutants showed particular behaviors: W31CNF and L35CNF.

EL222 W31CNF featured the most intricate kinetics of all single-CNF variants, with two events clearly distinguishable by all three probes, although with time constants twice slower for UV/vis (Figure 5a). The IR spectrum from the slowest kinetic event of W31CNF resembled more the spectrum of W31CNF under photostationary conditions (Figure S23b). The fastest kinetic event was sensed by both the amide and CNF probes, but its temperature-dependence for CNF31 showed non-Arrhenius kinetic behavior, in sharp contrast to all other dynamical processes (Figure S14c). This result suggests that CNF31 monitors a unique kinetic event. In combination with the previous mutagenesis and FTLS results, we propose that such an extra process reports on the monomer/oligomer equilibrium of EL222.

The EL222 L35CNF mutant showed two similarly intense dynamical events linked to the relaxation of CNF35 residue, with associated lifetimes of $\sim$26 and $\sim$94 s (labeled "A" and "B," respectively in Figure 5a middle). The slowest component correlates with the protein backbone recovery, while the FMN recovery sensed by UV/vis largely precedes it. On the other hand, the fastest CNF component takes place before the FMN moiety relaxes. To facilitate the interpretation of the two dynamical processes of CNF35, CNF35-A as the fastest and CNF35-B as the slowest, we looked at their transient spectra in the C≡N absorption region. The two events have distinct associated spectra and thus report on dissimilar changes around the CNF residue: a red shift from $\sim$2,241 to $\sim$2,230 cm^{-1} for the fastest component, and a blue-shift from $\sim$2,229 to $\sim$2,240 cm^{-1} for the slowest one (Figure S23a). The transient difference spectrum of CNF35-B resembles more the steady-state difference spectrum of the whole CNF population, where gain in α-helicity occurs, than the transient difference spectra of CNF35-A (Figure S23a). Because the CNF35 label is located at the A′α helix, we previously concluded from FTLS analysis of steady-state difference spectra that one of the two spectral components of L35CNF could be monitoring the decaged ("open")/caged ("close") conformational equilibrium of EL222, but changes in the folding state of A′α cannot be ruled out. To link stationary and kinetic evidence, we performed time-resolved experiments at different temperatures. First, we did FTLS

TABLE 1 Normalized lit-to-dark recovery lifetimes.

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

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

analyses on transient decay-associated difference spectra (DADS) obtained upon LDA (Figure S10). The main positive band in the DADS of CNF35-B had a similar FTLS slope as one of the lit states of L35CNF, while the main positive band of the DADS of CNF35-A had a flatter slope (lower solvent exposure) than either state of L35CNF (Figure 3b). Thus, the positive band in the "slow" component (CNF35-B) may correspond to one of the lit environments previously found for CNF35 under continuous illumination. On the other hand, hydrogen bonding around the CNF environment corresponding to the "fast" kinetic component (CNF35-A) seems to deviate from the environment sensed under photostationary conditions. Secondly, we prepared Arrhenius plots for the two kinetic events (Figure S14d) and found that the CNF35-B component had similar activation energy as the backbone relaxation calculated from the amide region of wt EL222 but 1.2 times higher than CNF35-A (Table S7). Therefore, we conclude that the slowest CNF35 component of L35CNF (CNF35-B) reports on the same (or similar) relaxation event as the backbone amide vibrations, that is, the global increase in α-helical secondary structure content. In contrast, the fastest CNF35 component of L35CNF (CNF35-A) monitors another dynamical process. Although the precise nature of such a process is difficult to define, considering all the mutagenesis and FTLS results, we tentatively propose that CNF35-A could be sensing the interaction between A′α-LOV and HTH halves of EL222 but not the folding state of A′α.

To compare the temporal evolution among mutants, given the altered and complex kinetics in some cases, we defined two reaction coordinates, one relative to the FMN relaxation and the other relative to amide relaxation. Subsequently, we computed normalized lifetimes (Table 1) and displayed *D* distributions against normalized lifetimes (Figure 5b). In such representations, the relaxation events monitored by CNF31 and CNF35-A, which partially overlap due to the broad nature of the CNF31 distribution, clearly take place before any other

processes (Figure 5b). CNF136, CNF35-B, and CNF151 relax after the FMN adduct breaks (Figure 5b top). The relaxation of CNF136 and CNF35-B are nearly simultaneous with the global protein relaxation (Figure 5b bottom). Finally, the relaxation of CNF151 is slower and the distribution of rates broader than the protein backbone relaxation (Figure 5b bottom). Therefore, from the point of view of the CNF probes, the results above point to at least three distinct kinetic classes of residue-specific relaxation times in EL222 (Figure 5c). First, a fast relaxation preceding both changes in the FMN and protein backbone and sensed as a red shift of CNF35-A and a red shift of CNF31. Second, a relaxation simultaneous with the protein backbone relaxation and sensed as a red shift of CNF136 and a blue shift of CNF35-B. Third, a slow relaxation posterior to the relaxation of the protein backbone and sensed as a red shift of CNF151. We conclude that CNF labels located in different parts of the protein track several aspects of the relaxation to the dark-state that are not concurrent with changes in the FMN nor in the protein backbone.

2.4.3 | Improved photocycle model of EL222

Based on the synergistic utilization of multiple probes (FMN, backbone, and CNF) we propose the following model of the dark recovery reaction of the LOV-based transcription factor EL222 (Figure 6). The dark state conformation can be safely assumed to represent the structure solved by X-ray crystallography (Nash et al. 2011). While the lit-state is presently unknown, it should be more oligomeric (dimer held together through interactions at least between the HTH domains), "open" (HTH domain decaged from LOV domain) and have less α-helical content (i.e. more random coil content) than the dark-state. Our data are in agreement with multiple parallel refolding pathways where different equilibria are present, each one with a distinct rate constant.
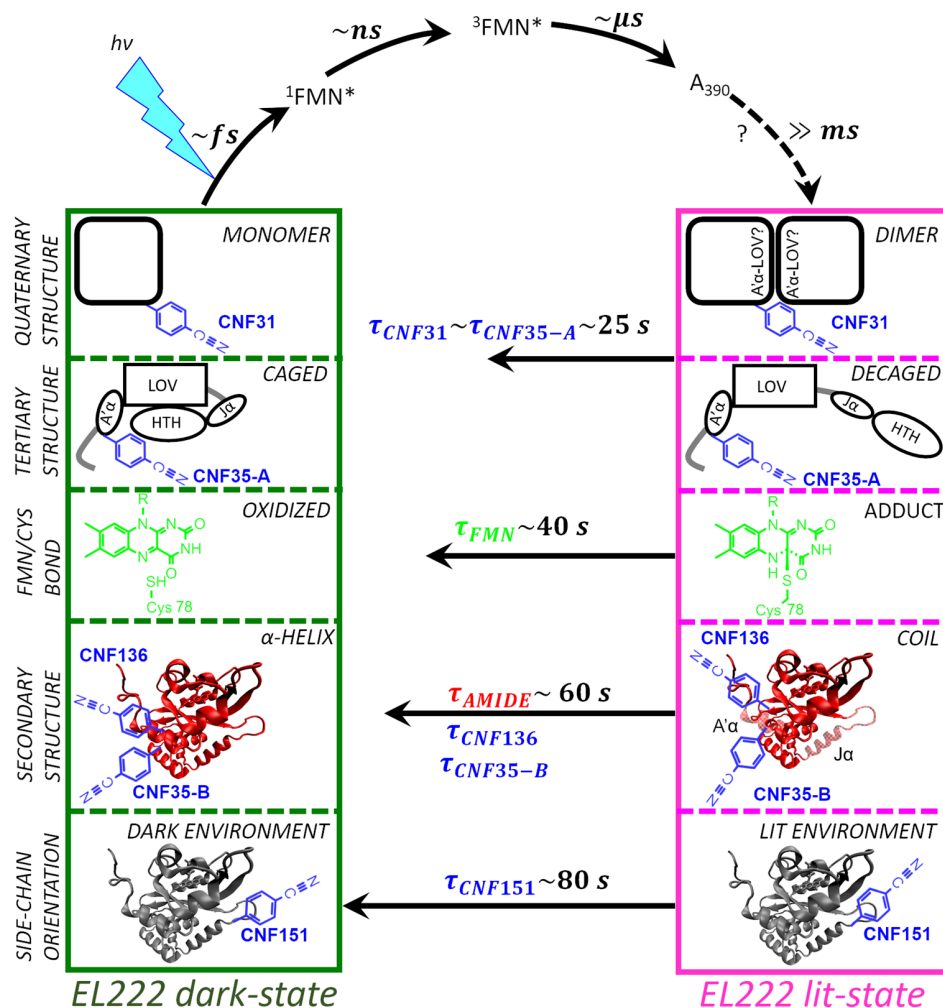
FIGURE 6 Kinetic-molecular model of EL222 photocycle. The model is defined in terms of five loosely coupled equilibria: monomer/oligomer, caged/decaged (close/open) conformation, oxidized/C78-bound FMN, helix/coil, and side-chain dark environment/lit environment. For the dark-to-lit transition, several 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.

Nevertheless, we cannot exclude the presence of sequential intermediates or more complex kinetic schemes, particularly concerning CNF35 relaxation. For instance, a mixture of EL222 monomers and dimers may be present upon illumination, adding complexity to CNF31 relaxation.

The lit-to-dark transition of EL222 can be described by up to seven equilibria, including one for the backbone relaxation, another one for the FMN and five for the residues (environments) CNF31, CNF35-A, CNF35-B, CNF136, and CNF151. These seven equilibria can be grouped into four categories based on their measured average recovery lifetimes in H_2O. Ordered from fastest to slowest kinetics but without assuming any dependence among them, we have the following molecular events. In ∼25 s, EL222 collapses and the A'α-LOV fragment engages in interactions with the HTH domain (sensed by CNF35-A probe). The oligomer falls apart in similar time scales (CNF31 probe). In ∼40 s, the FMN-C78 covalent bond is broken (sensed by the UV/vis absorption of the FMN probe). The protein backbone (most likely Jα, A'α,

or both) increases its α-helical secondary structure content in ~60 s (sensed by amide, CNF35-B and CNF136 probes). Equilibration of the environment around CNF151 takes place ~80 s upon stopping illumination.

The dark-to-lit transition of EL222 measured by time-resolved vibrational spectroscopy is typically analyzed using a sequential kinetic model. A singlet species (^{1}FMN*) is nearly instantaneously formed upon photon absorption, followed by the triplet species (^{3}FMN*) in ~3 ns, and finally the adduct state (A_{390}) in ~5 μs (or ~600 ns in H_2O-based buffer) (Iuliano et al. 2018; Andrikopoulos et al. 2021; Liu et al. 2023). Because the steady-state difference spectra are different from the transient spectra of the adduct state obtained a few microseconds after irradiation (Figure S25), the existence of an additional intermediate species in the multi-millisecond time scale is envisioned, as previously postulated (Iuliano et al. 2018). 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 (Takakado et al. 2018; Takakado et al. 2017).

3 | DISCUSSION

3.1 | 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); and (ii) the high degree of incorporation efficiency (only 2 cases out of 45 showed clear levels of mis-incorporation). 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 latter feature is a clear advantage over other commonly used probes, like azidophenylalanine. 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 labeling 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 S8).

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 <10) or did not comprehensively address their potential perturbation (Kraskov et al. 2021; Ye et al. 2009; Kurttila et al. 2021; Blankenburg et al. 2019; Krause et al. 2019; Hall et al. 2019; Ye et al. 2010). 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 2 (Y136CNF, M151CNF) out of 45 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 the infrared spectra in the "transparent window" region of CNF clearly shows long-distance signaling (allosteric communication) inside EL222. Perturbation of the FMN chromophore and binding pocket (the allosteric site) by photon absorption is transferred to the HTH module (the active site where DNA binding occurs) and other regions like A′α and Jα elements. This observation is in agreement with previous results based on NMR spectroscopy, although the residue-by-residue sensitivity of the two techniques differs (Nash et al. 2011). Importantly, the spectral observables (e.g., peak positions and areas) may be transformed

into physical variables, like local electric fields and H-bonding, thus providing new avenues for the synergistic use of experimental and physics-based computational approaches. In summary, ncAA-assisted optical spectroscopy complements the information content provided by common structural biology techniques and helps delineating signal transduction pathways in EL222.

3.2 | Dark reversion kinetics as witnessed by the chromophore, protein and residues

Previous studies on LOV photorecovery have largely concentrated on the UV/vis region (Lebedev et al. 2014) rather than on the IR region (Herman et al. 2013). Here, we measured the thermal recovery of EL222 by both UV/vis and FTIR spectroscopies. In principle, the relaxation times of different structural elements in proteins could be different, which may be inferred by independently analyzing different IR frequencies (Davis et al. 2018; Davis and Dyer 2014). This was not the case for EL222 because the Arrhenius plots were linear and essentially identical at all three probed frequency ranges $(1,750–1,680\ cm^{-1}, 1,680–1,600\ cm^{-1}, and\ 1,600–1,500\ cm^{-1})$. In turn, these results suggest that the dark recovery of EL222 conforms to a two-state folding transition (Davis et al. 2018). Previously, non-Arrhenius behavior (two different slopes) was observed for the recovery kinetics of EL222 based on FMN absorbance (Zoltowski et al. 2011). Here, we restricted ourselves to temperatures at or below 35°C to avoid partial unfolding of the protein (melting temperature is 55°C). Indeed, if we compare only the data below 35°C, the two studies agree on the same activation energy (∼15 kcal/mol) for the rupture of the flavin-cysteinyl bond. The ∼18 kcal/mol barrier seen in the IR region agrees with its slower kinetics and indicates a larger energy expenditure involved in rearranging the structure of EL222 in the transition state relative to adduct breakage. The observed kinetic isotope effect (KIE), more than 4-fold slower relaxation kinetics in D_2O than H_2O, is in agreement with previous reports of similar LOV domains based on UV/vis data (Lebedev et al. 2014; Corchnoy et al. 2003). The high KIE can be explained because the adduct breakage requires releasing a proton and an electron, and the folding of α-helices necessitates the breaking of protein-water H-bonds and formation of new intramolecular H-bonds.

For wt EL222 at 20°C, D values suggest a partial decoupling between FMN and protein relaxation where, on average, the FMN-C78 bond is broken earlier (41 ± 2 s in H_2O or 177 ± 10 s in D_2O) and then the protein recovers its secondary structure content (57 ± 3 s in H_2O or 223 ± 4 s in D_2O). We compared our results indicating asynchronous relaxation of protein and FMN structures,

with former studies done in other LOV domains by circular dichroism (CD), NMR and UV/vis spectroscopies (Harper et al. 2004; Corchnoy et al. 2003; Nakasone et al. 2007). In a previous NMR study (Harper et al. 2004), $^{15}N/^1H$ chemical shifts (reporting on local tertiary structure) yielded a mean recovery lifetime significantly longer than the recovery lifetime derived from UV/vis (reporting on FMN-Cys adduct). This observation is in agreement with our results that some CNF residues and the amide bands recover slightly after the FMN does. In addition, the NMR study showed a dispersion in the value of the recovery lifetimes, in line with the spread of recovery lifetimes we observed for our CNF side-chains. Unlike in the NMR study, we clearly identified two residues that recovered faster than the FMN. Another previous study also compared the recovery kinetics of the protein structure and of the FMN using far UV-CD to monitor the former and visible-CD to monitor the later (Corchnoy et al. 2003). Similar recovery lifetimes but slightly faster for the FMN were estimated. However, the FMN chromophore also absorbs in the far UV region and might contribute to the kinetics by attenuating differences in the estimated time constants. Such a potential problem is not present in our case since protein backbone carbonyls vibrate in a specific IR spectral region that is free from FMN bands. We should also note that different LOV proteins were used in previous works (Harper et al. 2004; Corchnoy et al. 2003; Nakasone et al. 2007), which may indeed experience a more synchronous relaxation of different moieties (as we see in our N53CNF and L216CNF variants of EL222). At present, we do not know how common the asynchronous dark reversion among LOV photoreceptors is.

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 (Hasegawa et al. 2005). 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 spectroscopy (Bannister et al. 2019).

3.3 | Implications for protein folding dynamics and signal transduction

Our results suggest significant light-induced structural changes in EL222 at the secondary (loss of helicity), tertiary (domain opening), and quaternary (oligomerization) levels of structure, in agreement with the literature (Nash et al. 2011; Takakado et al. 2018; Takakado et al. 2017;

Iuliano et al. 2018; Zoltowski et al. 2013; Andrikopoulos et al. 2021). Therefore, EL222 should be considered as a fold-switching protein, much like other photoreceptors, for example, phytochromes (Porter and Looger 2018; Kulkarni et al. 2018). The lit-to-dark transition of EL222 resembles the denatured-to-native refolding of unfolded proteins, where secondary structure acquisition ensues chain collapse (Uzawa et al. 2004). 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 (Ihalainen et al. 2008; Davis et al. 2015), even between the main-chain and side-chain of the same residue (Nagarajan et al. 2011), or between global and local probes (Ma and Gruebele 2005; Sukenik et al. 2016). Because our methodological approach is restricted to the ensemble level, the proposed relative order of events in our four-step kinetic model (Figure 6) 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 (Henry et al. 2013).

If the intrinsic time scales of thio-adduct breakage (picoseconds) (Kennis et al. 2004), chain collapse (microseconds) (Hagen and Eaton 2000), protein monomerization (milliseconds) (Miyamori et al. 2015), helix-coil transitions (nanoseconds) (Jas and Kuczera 2018), and side-chain rotation (picoseconds) (Cousin et al. 2018), 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 (Gil et al. 2017; Berntsson et al. 2019; Lee et al. 2022). 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 (Iuliano et al. 2018; Andrikopoulos et al. 2021). 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 (Ihalainen et al. 2008). In other LOV sensors, dynamic interactions at protein/protein interfaces are considered additional rate-limiting factors for the conformational recovery (Bannister et al. 2019). 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. This property adds to the multiple roles played by A'α in other LOV domains, including unfolding (Herman and Kottke 2015) and participation in protein–protein interaction interfaces (Vaidya et al. 2011; Bocola et al. 2015; Heintz and Schlichting 2016). Furthermore, we hypothesize that the dissociation of the A'α-LOV/A'α-LOV interface, that is, 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 (Takala et al. 2018). Overall, our work stresses again the importance of integrative approaches using multiple probes, residues and techniques to comprehensively characterize protein folding landscapes (Davis and Dyer 2014; Prigozhin et al. 2015).

4 | CONCLUSIONS

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: $\tau_{\mathrm{CNF31}} \sim \tau_{\mathrm{CNF35-A}} < \tau_{\mathrm{FMN}} < \tau_{\mathrm{Amide}} \sim \tau_{\mathrm{CNF35-B}} \sim \tau_{\mathrm{CNF136}} \leq \tau_{\mathrm{CNF151}}$. Thus, CNF31, and partially CNF35 (CNF35-A species), 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 species) relax simultaneously with backbone relaxation. Finally, the relaxation of CNF151 is delayed with respect to that of the protein backbone. Therefore, our evidence suggests that some CNF sites recover to their 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 on the A'α-LOV/A'α-LOV interface (oligomer/monomer EL222 equilibrium), and

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

Importantly, stringent selection criteria are critical to identify non-perturbative CNF residues. CNF kinetics tend to resemble more amide kinetics than FMN kinetics. Indeed, CNF and amide probes rendered partially redundant kinetic information for several mutants. However, in the absence of information from amide relaxation, the interpretation of CNF-derived kinetics would have been challenging if not impossible to achieve. Lifetime distributions from CNF and FMN probes only overlapped each other in the case of two mutants. In short, none of the probes considered in isolation (side-chain C≡N, backbone C=O, and FMN) can deliver a complete picture of EL222 dark recovery mechanism and kinetics, thus making a case for the complementarity of the three.

5 | MATERIALS AND METHODS

5.1 | Materials

CNF was purchased from Bide Pharmatech. The plasmid encoding EL222 was a kind gift of Dr. Kevin Gardner. The plasmid encoding the intein-chitin binding domain-12his fusion protein was a kind gift of Dr. Edward Lemke. The final plasmid for protein expression encoded an N-terminal EL222 protein (amino acids 17–225) followed by a C-terminal intein-CBD-12his purification tag. The plasmid pDule2-pCNF encoding the OTS composed of a polyspecific aminoacyl tRNA synthetase (MjCNFRS) and suppressor tRNA (MjtRNA$_{CUA}$) was a kind gift from Dr. Ryan Mehl. All other materials can be found in the Key Resource Table (Table S0).

5.2 | Protein expression

TAG sequences were introduced by site-directed mutagenesis and checked by DNA sequencing. *Escherichia coli* BL21(DE3) were co-transformed by electroporation (2.5 kV for 5 ms in a 2 mm path-length cuvette) with an ampicillin-resistant plasmid encoding EL222 (wt or with one TAG triplet at desired positions) and another spectinomcyin-resistant plasmid encoding an OTS able to incorporate CNF in response to UAG codons. The transformation mixture was spread on plates containing both antibiotics and grown overnight at 37°C. A single bacterial colony was picked, grown overnight at 37°C in LB medium and kept as glycerol stock at −80°C for later use. Bacteria were grown in terrific broth (TB) media (or exceptionally glucose-based minimal media in the case of W89CNF and I94CNF mutants) at 37°C up to an optical density at 600 nm of 0.4 at which point CNF (freshly dissolved in 3 M NaOH) was added to the medium at a final concentration of 1 mM. Cells were grown for 30 more minutes, then the temperature was lowered to 20°C and gene expression was induced with 0.5 mM IPTG in the dark.

5.3 | Protein purification and characterization

All protein purification steps were conducted under very dim light and 4°C unless otherwise stated. After 16 hr, cells were harvested by centrifugation at 6,000 g, resuspended in Tris–HCl 50 mM pH = 8 NaCl 100 mM PMSF (phenylmethylsulfonyl fluoride) 1 mM, and disrupted by sonication for 4 min on ice (50% of duty cycle, 50 mW). Cell suspensions were clarified by centrifugation at 60,000g for 30 min and the supernatant was purified by immobilized metal affinity chromatography (IMAC) with a Ni^{2+} resin. Loosely bound proteins were removed with 40 mM imidazole washes and fusion proteins were eluted with 500 mM imidazole. Addition of 100 mM DTT activated the self-splicing intein moiety that cleaved off itself from EL222 (48 hr at 23°C). Cleavage mixtures were purified by a second IMAC step to remove non-cleaved proteins. The flow-through was polished by size-exclusion chromatography (SEC) in a Superdex75 Increase column with 50 mM MES, 100 mM NaCl, pH = 6.8 as running buffer. The oligomeric states of all the purified proteins were estimated by interpolating the measured elution volumes on a calibration curve constructed using proteins of known molecular weight. In cases where the oligomeric state could not be unambiguously identified the mutants are labeled with a question mark. Fractions were analyzed by SDS-PAGE, pooled and concentrated (Table S3) using 10 kDa molecular-weight cut-off filters. Proteins were quantified by two methods: UV/vis spectroscopy (FMN concentration) assuming an extinction coefficient at 450 nm of 13,000 M^{-1} cm^{-1}, and IR spectroscopy (protein concentration). The latter is proportional to the concentration of C=O bonds (mostly arising from the protein backbone, with minor contributions from the two FMN carbonyls and side-chain carbonyls). The yields of CNF mutants were variable and in general lower by at least one order of magnitude with respect to the wt (Table S3).

All proteins were analyzed by mass spectrometry to determine the molecular weight and check for the presence of the CNF residue (Table S3). Proteins were diluted with 100 μL of 5% acetic acid in water and loaded onto Opti-trap C4 cartridge (Optimize Technologies), washed 4× with 250 μL of 5% acetic acid in water and eluted with 100 μL of 80% acetonitrile, 5% acetic acid. Proteins were analyzed by direct infusion using syringe pump at a flow

rate 2 μL/min connected with an electrospray ion source of 15 T solariX XR FT-ICR mass spectrometer (Bruker Daltonics). The mass spectrometer was externally calibrated using 1% (w/w) sodium trifluoroacetate. Proteins were measured in positive mode with 2 M data acquisition. The data were processed using SNAP algorithm, a part of DataAnalysis 4.4 software (Bruker Daltonics).

5.4 | Steady-state FTIR

The stationary infrared spectra of all the EL222 variants were measured with a Bruker Vertex 70v FTIR spectrometer equipped with a globar source, KBr beamsplitter and a liquid nitrogen cooled mercury cadmium telluride (MCT) detector. The protein samples were centrifuged at 18,000g for 15 min to remove any precipitants before measurements. D_2O buffer exchange was done by centrifugal concentration method using 10 kDa cutoff filters. The aperture value was set in accordance to get maximum signal without saturating the detector. For measurements in the "transparent window" region or in D_2O buffer, ~25 μL of sample was loaded into a temperature controllable demountable liquid cell with CaF_2 windows and a 50 μm Teflon spacer. For measuring amide region in H_2O buffer, ~10 μL of sample and 6 μm Teflon spacer was used instead. For each sample, three spectra (dark state, lit state and difference) were recorded at 20°C and 200–300 scans were averaged in the spectral range of 4,000 to 900 cm^{-1} with a resolution of 2 cm^{-1}. Protein lit state was induced by illuminating blue light of 450 nm LED with power of 25 mW/cm^2 on the sample throughout the measurement. Dark and lit samples were measured against buffer and the difference spectra were recorded as the lit spectra against the dark spectra. Protein samples had concentration of 1–2 mM in storage buffer (Table S3).

The resulting spectra were processed with Spectra-Gryph software. Baseline correction and Gaussian curve fitting was done with OriginPro (OriginLab).

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

5.5 | Time-resolved rapid-scan FTIR experiments

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

5.6 | Steady-state UV/vis spectroscopy

UV/vis absorbance spectra in the range of 320–550 nm of EL222 variants were measured with a Specord 50 Plus spectrometer (Analytik Jena) using a quartz cuvette (Hellma Analytics) having 10 mm path length at 20°C and 70 μL sample volume. The samples were the same as for the FTIR experiments but they were diluted ~70 times to a concentration between 10 and 30 μM, equivalent to an absorbance at 450 nm ~0.5.

5.7 | Time-resolved UV/vis spectroscopy

Dark state recovery kinetics of protein variants were measured by first continuously illuminating the sample for 60 s with blue light of 450 nm LED with power of approximately 25 mW/cm^2 (M450L1, Thorlabs). LED was then turned off and spectra were subsequently taken for 10 min in the case of H_2O-based experiments or 15 min for D_2O-based experiments.

AUTHOR CONTRIBUTIONS

Aditya S. Chaudhari: Formal analysis (equal); investigation (equal); methodology (equal); visualization (equal); writing – review and editing (equal). **Aditi Chatterjee:** Investigation (equal). **Catarina A.O. Domingos:** Investigation (equal). **Prokopis C. Andrikopoulos:** Resources (equal). **Yingliang Liu:** Resources (equal). **Inger Andersson:** Writing – review and editing (equal). **Bohdan Schneider:** Funding acquisition (equal); supervision (equal). **Victor A. Lorenz-Fonfria:** Formal analysis (equal); methodology (equal); software (equal); writing – review and editing (equal). **Gustavo Fuertes:** Conceptualization (equal); formal analysis (equal); investigation (equal); methodology (equal); visualization (equal); writing – original draft (equal); writing – review and editing (equal).

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

The data that supports the findings of this study are available in the supplementary material of this article. Large datasets that support the findings of this study are openly available in Zenodo at https://dx.doi.org/10.5281/zenodo.7086623

ORCID

Aditya S. Chaudhari https://orcid.org/0000-0003-3538-7471

Aditi Chatterjee https://orcid.org/0000-0002-6886-0651

Catarina A. O. Domingos https://orcid.org/0000-0003-3800-2303

Prokopis C. Andrikopoulos https://orcid.org/0000-0003-0255-1108

Yingliang Liu https://orcid.org/0000-0002-4089-688X

Inger Andersson https://orcid.org/0000-0002-2729-0787

Bohdan Schneider https://orcid.org/0000-0001-7855-3690

Victor A. Lórenz-Fonfría https://orcid.org/0000-0002-8859-8347

Gustavo Fuertes https://orcid.org/0000-0002-8564-8644

REFERENCES

Adhikary R, Zimmermann J, Dawson PE, Romesberg FE. Temperature dependence of CN and SCN IR absorptions facilitates their interpretation and use as probes of proteins. Anal Chem. 2015;87:11561–7.

Adhikary R, Zimmermann J, Romesberg FE. Transparent window vibrational probes for the characterization of proteins with high structural and temporal resolution. Chem Rev. 2017;117:1927–69.

Andrikopoulos PC, Liu Y, Picchiotti A, Lenngren N, Kloz M, Chaudhari AS, et al. Femtosecond-to-nanosecond dynamics of flavin mononucleotide monitored by stimulated Raman spectroscopy and simulations. Phys Chem Chem Phys. 2020;22:6538–52.

Andrikopoulos PC, Chaudhari AS, Liu Y, Konold PE, Kennis JTM, Schneider B, et al. QM calculations predict the energetics and infrared spectra of transient glutamine isomers in LOV photoreceptors. Phys Chem Chem Phys. 2021;23:13934–50.

Arrondo JLR, Muga A, Castresana J, Goñi FM. Quantitative studies of the structure of proteins in solution by fourier-transform infrared spectroscopy. Prog Biophys Mol Biol. 1993;59:23–56.

Bannister S, Böhm E, Zinn T, Hellweg T, Kottke T. Arguments for an additional long-lived intermediate in the photocycle of the full-length aureochrome 1c receptor: a time-resolved small-angle X-ray scattering study. Struct Dyn. 2019;6:034701.

Barth A. Infrared spectroscopy of proteins. Biochim Biophys Acta Bioenergetics. 2007;1767:1073–101.

Berntsson O, Rodriguez R, Henry L, Panman MR, Hughes AJ, Einholz C, et al. Photoactivation of *Drosophila melanogaster* cryptochrome through sequential conformational transitions. Sci Adv. 2019;5:eaaw1531.

Blankenburg L, Schroeder L, Habenstein F, Blasiak B, Kottke T, Bredenbeck J. Following local light-induced structure changes and dynamics of the photoreceptor PYP with the thiocyanate IR label. Phys Chem Chem Phys. 2019;21:6622–34.

Bocola M, Schwaneberg U, Jaeger K-E, Krauss U. Light-induced structural changes in a short light, oxygen, voltage (LOV) protein revealed by molecular dynamics simulations—implications for the understanding of LOV photoactivation. Front Mol Biosci. 2015;2.

Corchnoy SB, Swartz TE, Lewis JW, Szundi I, Briggs WR, Bogomolni RA. Intramolecular proton transfers and structural changes during the photocycle of the LOV2 domain of Phototropin 1. J Biol Chem. 2003;278:724–31.

Cousin SF, Kadeřávek P, Bolik-Coulon N, Gu Y, Charlier C, Carlier L, et al. Time-resolved protein side-chain motions unraveled by high-resolution relaxometry and molecular dynamics simulations. J Am Chem Soc. 2018;140:13456–65.

Davis CM, Dyer RB. WW domain folding complexity revealed by infrared spectroscopy. Biochemistry. 2014;53:5476–84.

Davis CM, Cooper AK, Dyer RB. Fast helix formation in the B domain of protein a revealed by site-specific infrared probes. Biochemistry. 2015;54:1758–66.

Davis CM, Zanetti-Polzi L, Gruebele M, Amadei A, Dyer RB, Daidone I. A quantitative connection of experimental and simulated folding landscapes by vibrational spectroscopy. Chem Sci. 2018;9:9002–11.

Fafarman AT, Webb LJ, Chuang JI, Boxer SG. Site-specific conversion of cysteine thiols into thiocyanate creates an IR probe for electric fields in proteins. J Am Chem Soc. 2006;128:13356–7.

Fayer MD. Fast protein dynamics probed with infrared vibrational Echo experiments. Annu Rev Phys Chem. 2001;52:315–56.

First JT, Slocum JD, Webb LJ. Quantifying the effects of hydrogen bonding on nitrile frequencies in GFP: beyond solvent exposure. J Phys Chem B. 2018;122:6733–43.

Getahun Z, Huang CY, Wang T, De Leon B, DeGrado WF, Gai F. Using nitrile-derivatized amino acids as infrared probes of local environment. J Am Chem Soc. 2003;125:405–11.

Gil AA, Laptenok SP, French JB, Iuliano JN, Lukacs A, Hall CR, et al. Femtosecond to millisecond dynamics of light induced Allostery in the Avena sativa LOV domain. J Phys Chem B. 2017;121:1010–9.

Goett-Zink L, Klocke JL, Bögeholz LAK, Kottke T. In-cell infrared difference spectroscopy of LOV photoreceptors reveals structural responses to light altered in living cells. J Biol Chem. 2020;295:11729–41.

Hagen SJ, Eaton WA. Two-state expansion and collapse of a polypeptide 1 1Edited by A. R Fersht. J Mol Biol. 2000;301:1019–27.

Hall CR, Tolentino Collado J, Iuliano JN, Gil AA, Adamczyk K, Lukacs A, et al. Site-specific protein dynamics probed by ultrafast infrared spectroscopy of a noncanonical amino acid. J Phys Chem B. 2019;123:9592–7.

Harper SM, Neil LC, Day IJ, Hore PJ, Gardner KH. Conformational changes in a photosensory LOV domain monitored by time-resolved NMR spectroscopy. J Am Chem Soc. 2004;126:3390–1.

Hasegawa K, Masuda S, Ono T-A. Spectroscopic analysis of the dark relaxation process of a photocycle in a sensor of blue light using FAD (BLUF) protein Slr1694 of the Cyanobacterium synechocystis sp. PCC6803. Plant Cell Physiol. 2005;46:136–46.

Heintz U, Schlichting I. Blue light-induced LOV domain dimerization enhances the affinity of Aureochrome 1a for its target DNA sequence. Elife. 2016;5:e11860.

Henry ER, Best RB, Eaton WA. Comparing a simple theoretical model for protein folding with all-atom molecular dynamics simulations. Proc Natl Acad Sci U S A. 2013;110:17880–5.

Herman E, Kottke T. Allosterically regulated unfolding of the A'α helix exposes the dimerization site of the blue-light-sensing Aureochrome-LOV domain. Biochemistry. 2015;54:1484–92.

Herman E, Sachse M, Kroth PG, Kottke T. Blue-light-induced unfolding of the Jα helix allows for the dimerization of Aureochrome-LOV from the diatom Phaeodactylum tricornutum. Biochemistry. 2013;52:3094–101.

Ihalainen JA, Paoli B, Muff S, Backus EHG, Bredenbeck J, Woolley GA, et al. Helix folding in the presence of structural constraints. Proc Natl Acad Sci U S A. 2008;105:9588–93.

Iuliano JN, Gil AA, Laptenok SP, Hall CR, Tolentino Collado J, Lukacs A, et al. Variation in LOV photoreceptor activation dynamics probed by time-resolved infrared spectroscopy. Biochemistry. 2018;57:620–30.

Jas GS, Kuczera K. Helix–coil transition courses through multiple pathways and intermediates: fast kinetic measurements and dimensionality reduction. J Phys Chem B. 2018;122:10806–16.

Kennis JTM, van Stokkum IHM, Crosson S, Gauden M, Moffat K, van Grondelle R. The LOV2 domain of Phototropin: a reversible photochromic switch. J Am Chem Soc. 2004;126:4512–3.

Konold PE, Mathes T, Weißenborn J, Groot ML, Hegemann P, Kennis JTM. Unfolding of the C-terminal Jα helix in the LOV2 photoreceptor domain observed by time-resolved vibrational spectroscopy. J Phys Chem Lett. 2016;7:3472–6.

Kottke T, Lorenz-Fonfria VA, Heberle J. The grateful infrared: sequential protein structural changes resolved by infrared difference spectroscopy. J Phys Chem B. 2017;121:335–50.

Kottke T, Xie A, Larsen DS, Hoff WD. Photoreceptors take charge: emerging principles for light sensing. Annu Rev Biophys. 2018; 47:291–313.

Kraskov A, von Sass J, Nguyen AD, Hoang TO, Buhrke D, Katz S, et al. Local electric field changes during the photoconversion of the bathy phytochrome Agp2. Biochemistry. 2021;60:2967–77.

Krause BS, Kaufmann JCD, Kuhne J, Vierock J, Huber T, Sakmar TP, et al. Tracking pore hydration in Channelrhodopsin by site-directed infrared-active azido probes. Biochemistry. 2019;58:1275–86.

Krimm S, Bandekar J. Vibrational Spectroscopy and Conformation of Peptides, Polypeptides, and Proteins. Advances in Protein Chemistry. Advances in Protein Chemistry. Elsevier Academic Press. Volume 38; 1986. p. 181–364.

Kulkarni P, Solomon TL, He Y, Chen Y, Bryan PN, Orban J. Structural metamorphism and polymorphism in proteins on the brink of thermodynamic stability. Protein Sci. 2018;27:1557–67.

Kurttila M, Stucki-Buchli B, Rumfeldt J, Schroeder L, Häkkänen H, Liukkonen A, et al. Site-by-site tracking of signal transduction in an azidophenylalanine-labeled bacteriophytochrome with step-scan FTIR spectroscopy. Phys Chem Chem Phys. 2021;23: 5615–28.

Lebedev N, Zayner JP, Sosnick TR. Factors that control the chemistry of the LOV domain photocycle. PLoS ONE. 2014;9.

Lee SJ, Kim TW, Kim JG, Yang C, Yun SR, Kim C, et al. Light-induced protein structural dynamics in bacteriophytochrome revealed by time-resolved x-ray solution scattering. Sci Adv. 2022;8:eabm6278.

Lim M, Hamm P, Hochstrasser RM. Protein fluctuations are sensed by stimulated infrared echoes of the vibrations of carbon monoxide and azide probes. Proc Natl Acad Sci U S A. 1998;95: 15315–20.

Liu Y, Chaudhari AS, Chatterjee A, Andrikopoulos PC, Picchiotti A, Rebarz M, et al. Sub-millisecond photoinduced dynamics of free and EL222-bound FMN by stimulated Raman and visible absorption spectroscopies. Biomolecules. 2023;13.

Löffler JG, Deniz E, Feid C, Franz VG, Bredenbeck J. Versatile vibrational energy sensors for proteins. Angew Chem Int Ed. 2022;61.

Lorenz-Fonfria VA. Infrared difference spectroscopy of proteins: from bands to bonds. Chem Rev. 2020;120:3466–576.

Lórenz-Fonfría VA, Kandori H. Bayesian maximum entropy (two-dimensional) lifetime distribution reconstruction from time-resolved spectroscopic data. Appl Spectrosc. 2016;61:428–43.

Ma H, Gruebele M. Kinetics are probe-dependent during downhill folding of an engineered 6-85 protein. Proc Natl Acad Sci U S A. 2005;102:2283–7.

Maia RNA, Ehrenberg D, Oldemeyer S, Knieps-Grünhagen E, Krauss U, Heberle J. Real-time tracking of proton transfer from the reactive cysteine to the flavin chromophore of a Photosensing light oxygen voltage protein. J Am Chem Soc. 2021;143: 12535–42.

Manandhar M, Chun E, Romesberg FE. Genetic code expansion: inception, development, commercialization. J Am Chem Soc. 2021;143:4859–78.

Milles S, Tyagi S, Banterle N, Koehler C, VanDelinder V, Plass T, et al. Click strategies for single-molecule protein fluorescence. J Am Chem Soc. 2012;134:5187–95.

Miyake-Stoner SJ, Miller AM, Hammill JT, Peeler JC, Hess KR, Mehl RA, et al. Probing protein folding using site-specifically encoded unnatural amino acids as FRET donors with tryptophan. Biochemistry. 2009;48:5953–62.

Miyamori T, Nakasone Y, Hitomi K, Christie JM, Getzoff ED, Terazima M. Reaction dynamics of the UV-B photosensor UVR8. Photochem Photobiol Sci. 2015;14:995–1004.

Nagarajan S, Taskent-Sezgin H, Parul D, Carrico I, Raleigh DP, Dyer RB. Differential ordering of the protein backbone and side chains during protein folding revealed by site-specific recombinant infrared probes. J Am Chem Soc. 2011;133:20335–40.

Nakasone Y, Eitoku T, Matsuoka D, Tokutomi S, Terazima M. Dynamics of conformational changes of Arabidopsis Phototropin 1 LOV2 with the linker domain. J Mol Biol. 2007;367: 432–42.

Nash AI, McNulty R, Shillito ME, Swartz TE, Bogomolni RA, Luecke H, et al. Structural basis of photosensitivity in a bacterial light-oxygen-voltage/helix-turn-helix (LOV-HTH) DNA-binding protein. Proc Natl Acad Sci U S A. 2011;108:9449–54.

Oh K-I, Lee J-H, Joo C, Han H, Cho M. β-Azidoalanine as an IR probe: application to amyloid Aβ(16-22) aggregation. J Phys Chem B. 2008;112:10352–7.

Porter LL, Looger LL. Extant fold-switching proteins are widespread. Proc Natl Acad Sci U S A. 2018;115:5968–73.

Prigozhin MB, Chao SH, Sukenik S, Pogorelov TV, Gruebele M. Mapping fast protein folding with multiple-site fluorescent probes. Proc Natl Acad Sci U S A. 2015;112:7966–71.

Schultz KC, Supekova L, Ryu Y, Xie J, Perera R, Schultz PG. A genetically encoded infrared probe. J Am Chem Soc. 2006;128: 13984–5.

Spexard M, Immeln D, Thöing C, Kottke T. Infrared spectrum and absorption coefficient of the cofactor flavin in water. Vibrat Spectrosc. 2011;57:282–7.

Stock G, Hamm P. A non-equilibrium approach to allosteric communication. Philos Trans R Soc B Biol Sci. 2018;373:20170187.

van Stokkum IHM, Larsen DS, van Grondelle R. Global and target analysis of time-resolved spectra. Biochim Biophys Acta Bioenergetics. 2004;1657:82–104.

Sukenik S, Pogorelov TV, Gruebele M. Can local probes go global? A joint experiment-simulation analysis of lambda(6-85) folding. J Phys Chem Lett. 2016;7:1960–5.

Suydam IT, Boxer SG. Vibrational stark effects calibrate the sensitivity of vibrational probes for electric fields in proteins. Biochemistry. 2003;42:12050–5.

Takakado A, Nakasone Y, Terazima M. Photoinduced dimerization of a photosensory DNA-binding protein EL222 and its LOV domain. Phys Chem Chem Phys. 2017;19:24855–65.

Takakado A, Nakasone Y, Terazima M. Sequential DNA binding and dimerization processes of the photosensory protein EL222. Biochemistry. 2018;57:1603–10.

Takala H, Lehtivuori HK, Berntsson O, Hughes A, Nanekar R, Niebling S, et al. On the (un)coupling of the chromophore, tongue interactions, and overall conformation in a bacterial phytochrome. J Biol Chem. 2018;293:8161–72.

Taskent-Sezgin H, Chung J, Banerjee PS, Nagarajan S, Dyer RB, Carrico I, et al. Azidohomoalanine: a conformationally sensitive IR probe of protein folding, protein structure, and electrostatics. Angew Chem Int Ed Engl. 2010;49:7473–5.

Thielges MC. Transparent window 2D IR spectroscopy of proteins. J Chem Phys. 2021;155:40903.

Tolentino Collado J, Iuliano JN, Pirisi K, Jewlikar S, Adamczyk K, Greetham GM, et al. Unraveling the photoactivation mechanism of a light-activated adenylyl cyclase using ultrafast spectroscopy coupled with unnatural amino acid mutagenesis. ACS Chem Biol. 2022;17:2643–54.

Uzawa T, Akiyama S, Kimura T, Takahashi S, Ishimori K, Morishima I, et al. Collapse and search dynamics of apomyoglobin folding revealed by submillisecond observations of -helical content and compactness. Proc Natl Acad Sci U S A. 2004;101:1171–6.

Vaidya AT, Chen CH, Dunlap JC, Loros JJ, Crane BR. Structure of a light-activated LOV protein dimer that regulates transcription. Sci Signal. 2011;4:ra50-ra50.

Weaver JB, Kozuch J, Kirsh JM, Boxer SG. Nitrile infrared intensities characterize electric fields and hydrogen bonding in protic, aprotic, and protein environments. J Am Chem Soc. 2022;144: 7562–7.

Yang W-J, Griffiths PR, Byler DM, Susi H. Protein conformation by infrared spectroscopy: resolution enhancement by Fourier self-deconvolution. Appl Spectrosc. 2016;39:282–7.

Ye S, Huber T, Vogel R, Sakmar TP. FTIR analysis of GPCR activation using azido probes. Nat Chem Biol. 2009;5:397–9.

Ye S, Zaitseva E, Caltabiano G, Schertler GF, Sakmar TP, Deupi X, et al. Tracking G-protein-coupled receptor activation using genetically encoded infrared probes. Nature. 2010;464:1386–9.

Zhang J, Wang L, Zhang J, Zhu J, Pan X, Cui Z, et al. Identifying and modulating accidental Fermi resonance: 2D IR and DFT study of 4-azido-l-phenylalanine. J Phys Chem B. 2018;122: 8122–33.

Zoltowski BD, Nash AI, Gardner KH. Variations in protein-flavin hydrogen bonding in a light, oxygen, voltage domain produce non-Arrhenius kinetics of adduct decay. Biochemistry. 2011;50:8771–9.

Zoltowski BD, Motta-Mena LB, Gardner KH. Blue light-induced dimerization of a bacterial LOV-HTH DNA-binding protein. Biochemistry. 2013;52:6653–61.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.