

Covalent Drug Binding in Live Cells Monitored by Mid-Infrared Quantum Cascade Laser Spectroscopy: Photoactive Yellow Protein as a Model System

Srijit Mukherjee,[▽] Steven D. E. Fried,[▽] Nathalie Y. Hong, Nahal Bagheri, Jacek Kozuch, Irimpan I. Mathews, Jacob M. Kirsh, and Steven G. Boxer*



Cite This: *J. Am. Chem. Soc.* 2025, 147, 47213–47222



Read Online

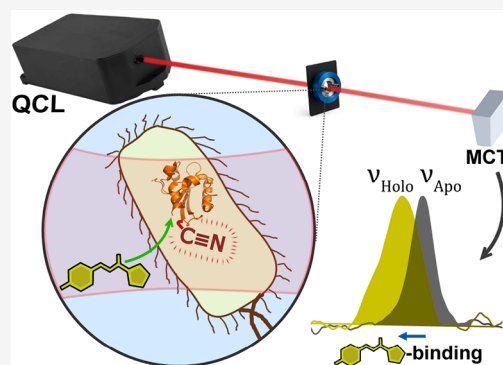
ACCESS |

Metrics & More

Article Recommendations

Supporting Information

ABSTRACT: The detection of drug–target interactions in live cells enables analysis of therapeutic compounds in a native cellular environment. Recent advances in spectroscopy and molecular biology have facilitated the development of genetically encoded vibrational probes like nitriles that can sensitively report on molecular interactions. Nitriles are powerful tools for measuring electrostatic environments within condensed media like proteins, but such measurements in live cells have been hindered by low signal-to-noise ratios. In this study, we design a spectrometer based on a double-beam quantum cascade laser (QCL)-based transmission infrared (IR) source with balanced detection that can significantly enhance sensitivity to nitrile vibrational probes embedded in proteins within cells compared to a conventional FTIR spectrometer. Using this approach, we detect small-molecule binding in *Escherichia coli*, with particular focus on the interaction between para-Coumaric acid (pCA) and nitrile-incorporated photoactive yellow protein (PYP). This system effectively serves as a model for investigating covalent drug binding in a cellular environment. Notably, we observe large spectral shifts of up to 15 cm^{-1} for nitriles embedded in PYP between the unbound and drug-bound states directly within bacteria, in agreement with observations for purified proteins. Such large spectral shifts are ascribed to the changes in the hydrogen-bonding environment around the local environment of nitriles, accurately modeled through high-level molecular dynamics simulations using the AMOEBA force field. Our findings underscore the QCL spectrometer's ability to enhance sensitivity for monitoring drug–protein interactions, offering new opportunities for advanced methodologies in drug development and biochemical research.



INTRODUCTION

Detection and quantification of drug binding is one of the most sought-after applications of analytical chemistry to human health and comprises a multibillion-dollar industry. While many drug binding techniques rely upon separative methods or highly perturbative techniques (e.g., calorimetry, surface plasmon resonance, mass spectrometry) that can only be carried out *in vitro*^{1–3} there is a decisive advantage to being able to directly probe drug–protein interactions in living cells. Measurements of drug binding in live cells can capture the native environment of the target protein of interest and account for the effects of competing variables that might limit drug binding to a target protein in a cell, including metabolic degradation, trafficking to the correct compartment, or competition of off-target sites for binding to the drug of interest.⁴

While cell-based assays remain essential for assessing efficacy and toxicity—the ultimate goals of drug discovery—they often rely on downstream, *indirect* readouts.⁵ In contrast, direct measurements of drug binding to the protein of interest can provide mechanistic insights into target engagement and

binding mode, offering valuable complementary information.^{6,7} The majority of such live-cell measurements make use of fluorescence-based assays based on fluorescence polarization,⁸ Förster Resonance Energy Transfer (FRET),⁹ fluorescence lifetime imaging (FLIM),¹⁰ fluorescence thermal shift assays (FTSA)¹¹ and single molecule fluorescence.¹² Yet, such assays are disadvantaged by limited applicability for nonfluorescent drug systems and generally require labeling by bulky fluorescent groups such as dyes or fluorescent proteins that have the potential to be highly perturbative to the system of interest.^{13,14}

Vibrational probes by contrast can serve as minimally perturbative reporters of their immediate environment,

Received: August 20, 2025

Revised: December 1, 2025

Accepted: December 3, 2025

Published: December 14, 2025



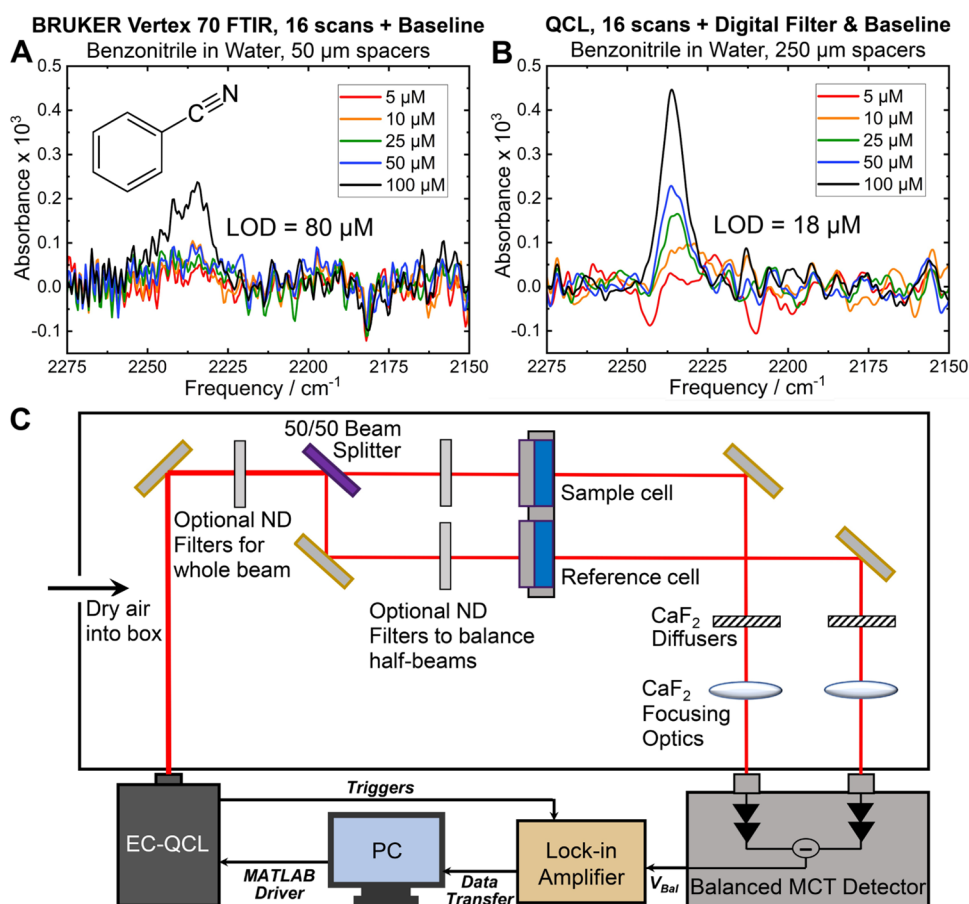


Figure 1. Comparison of detection limits for benzonitrile using (A) a conventional Bruker Vertex-70 FTIR spectrometer and (B) a QCL-based spectrometer. The FTIR, limited by aqueous sample transmission with 50 μm spacers, achieves a limit of detection (LOD) of 80 μM . The QCL-based approach improves the LOD to 18 μM , offering a ~ 5 -fold enhancement in sensitivity for IR measurements in the nitrile stretch region compared to high-end FTIR systems due to the use of a 250 μm sample path length. (C) Block diagram of double-beam QCL spectrometer with lock-in detection for balanced absorption measurement (ND = Neutral Density Filter, See SI Section S1 for details).

demonstrating frequency shifts that are physically interpretable in terms of environmental electrostatics^{15–17} molecular conformation,¹⁸ or hydrogen bonding.^{19–21} Many common vibrational probes like nitriles ($\text{C}\equiv\text{N}$) are intrinsic to drugs themselves²² or can be engineered selectively onto the target protein scaffold *via* amber suppression^{23,24} and can thus serve as effective reporters of drug–protein interactions. Nitriles are of particular interest for this purpose due to their relatively strong absorption in a clean region of the infrared spectrum. They have been used as molecular probes in various contexts, including proteins,^{21–25} lipid membranes,²⁶ surfaces,²⁷ microdroplets²⁸ and metal–organic frameworks.¹⁷

In cellulo spectroscopic analysis of site-specifically labeled vibrational probes represents a relatively unexplored frontier—one that remains limited by both sensitivity to low concentrations and interference from cellular background. Most efforts in live-cell transmission IR spectroscopy focus on line shape differences in the amide I and II regions, rather than the detection of specific vibrational frequency shifts of labeled sites.^{29–32} Despite these difficulties, a few recent and notable studies have demonstrated progress, typically using attenuated total reflectance (ATR)-based techniques to focus on strong absorbing oscillators such as $\mu\text{-CO}$ in hydrogenases in whole *Escherichia coli*³³ or flavin carbonyls as they photoreact in human cell lines.³⁴ Another recent study stemming from Heberle and co-workers has made use of high-power quantum

cascade lasers (QCLs) for time-resolved detection of cysteine protonation dynamics in halorhodopsin in live *E. coli*.³⁵ QCLs pose a significant advantage as a mid-IR light source in that their power at a given frequency is orders of magnitude higher than conventional global sources, enabling substantially longer sample pathlengths.³⁶ For example, QCLs emit tunable, narrow-bandwidth mid-infrared light that has advanced discrete frequency mid-infrared imaging for label-free, high-resolution histopathological analysis.³⁷ Lendl and co-workers have demonstrated the remarkable capabilities of external cavity, mid-IR QCLs for biomolecular spectroscopy,^{38–40} especially how the use of balanced detection with a double-beam approach can compensate for much of the noise inherent in QCL radiation and enhance the sensitivity of transmission IR spectroscopy measurements for the amide region of proteins in aqueous buffer.^{41,42}

Here we extend the capabilities of QCL-based vibrational spectroscopy of proteins into a new mid-IR frequency regime, with balanced detection for monitoring genetically encoded nitrile vibrational probes in live cells as molecular reporters of protein–small molecule binding. We first demonstrate a significant sensitivity improvement, up to 5-fold, over conventional FTIR spectroscopy for detection of aromatic nitriles due to longer pathlengths for transmission measurements accessible to the QCL setup. We then use the QCL spectrometer to detect small molecule binding to proteins—a key objective in

drug discovery—using nitrile frequency shifts as direct indicators of fine protein structural perturbations caused by binding events in live cells. High-level polarizable molecular dynamics simulations provide a physical rationale for the observed experimental results.

To model covalent drug-binding events, we present an example based on photoactive yellow protein (PYP) from *Halorhodospira halophila*. Apo-PYP, as expressed in cells, forms a covalent complex with the chromophore, p-Coumaric acid (pCA), a process that is analogous to covalent drug binding. PYP is a highly soluble globular protein from the PAS domain superfamily.⁴³ While not a drug target in the conventional sense, it shares chemical characteristics with common covalent drug targets, especially those with reactive cysteines.^{44–47} Moreover, PYP's potential as an optogenetic system has led to considerable interest in measuring its interactions with its chromophore and other binding partners in living cells.^{43,48} While a FRET-based assay using a fusion of a blue fluorescent protein and apo-PYP has previously enabled an *in cellulo* binding study,⁴⁹ we note that a vibrational spectroscopy approach without the use of a fusion protein construct is much more generalizable including nonphotoactive compounds involved in other systems. Our spectroscopic analyses of the nitrile frequency shifts using this assay, supported by detailed structural studies including X-ray crystallography and high-level AMOEBA molecular dynamics (MD) simulations, indicate subtle changes in hydrogen bonding interactions involving the embedded nitriles within the protein upon drug incorporation. These findings underscore the utility of this assay for measuring drug binding in live cells at the molecular level using QCL-based IR spectroscopy frequency shifts.

RESULTS

Design and Use of the Dual-Beam QCL-Spectrometer for Transmission IR Measurements. We designed a custom double-beam infrared spectrometer, employing an external-cavity quantum cascade laser (EC-QCL) as the mid-IR radiation source (MIRcat-QT-2000 with M2048-P, Daylight Solutions, San Diego, CA, USA), enabling spectral acquisition within the 2000–2300 cm^{-1} range (Figures S1 and S2) using an optical layout inspired from the work of Akhgar et al.⁴¹ The instrument utilizes a balanced detection scheme, where the incident laser beam is split into reference and sample paths, and subsequently focused onto a thermoelectrically cooled HgCdTe (MCT) detector (VIGO Photonics, Poland). The optical diagram is presented in Figure 1 with calibration information in Section S1, Figures S3 and S4. A high-speed SR865A lock-in amplifier (Stanford Research Systems, Sunnyvale, CA, USA) and custom MATLAB routines were used for data acquisition and signal processing, including phase correction, Savitsky-Golay filtering, and Fourier filtering.³⁸ Absorbance measurements were derived from the ratio of beam intensities taken from reference and balanced channel voltage outputs, effectively mitigating fluctuations, primarily driven by the QCL source, using the balanced detection scheme. Two identical liquid sample cells equipped with CaF_2 windows and Teflon spacers (50–250 μm ; Pike Technologies, Fitchburg, WI, USA) were mounted on machined sample holders, serving as the sample and blank sample cells.

Utilizing benzonitrile as a model nitrile probe in water in the μM range, we compared the limit of detection for the spectrometer against a conventional Bruker Vertex 70 FTIR spectrometer (Figures S5–S9 and Tables S1–S3). We also

document in the SI the increased sensitivity of the QCL for other useful vibrational probes including alkyne ($\text{C}\equiv\text{C}$), azides ($\text{N}=\text{N}^+=\text{N}^-$), and carbon-deuterium ($\text{C}-\text{D}$) stretches (Figures S10–S12 and Table S4). A notable challenge for IR spectra of aqueous samples in this spectral window is restricted transmission due to water's combination mode centered at approximately 2100 cm^{-1} .⁵⁰ Consequently, this absorption imposes a limit on the maximum spacer length that can be used in the Bruker Vertex 70, which is restricted to $\sim 50\ \mu\text{m}$ and leads to a limit of detection (LOD) for benzonitrile of $\sim 80\ \mu\text{M}$ for the nitrile stretch centered at $\sim 2235\ \text{cm}^{-1}$. In contrast, the balanced QCL-based detection (limited by the same phenomena) using 250 μm spacers significantly enhances the experimental LOD, achieving a value of 18 μM . This QCL-based method therefore provides an approximately 5-fold increase in sensitivity for IR measurements within the nitrile-stretch region, outperforming the conventional FTIR spectrometer and enabling detection of low concentrations of nitriles in aqueous environments, such as proteins or nitrile-labeled molecules directly within cells (Table S5).

Incorporation and Detection of Nitrile-Labeled Proteins in Live Bacterial Cells. We evaluated the spectrometer's ability to detect nitrile stretches within *E. coli* cells using amber suppression machinery to introduce ortho-cyanophenylalanine (oCNF) residues into target proteins of interest. Our primary focus was on apo-PYP as a model system for the equivalent of covalent drug binding. We used site-specific labeling to produce four apo-PYP variants: F28oCNF, F62oCNF, F92oCNF, and F96oCNF. As an additional test of visualizing nitriles in live cells we also incorporated a nitrile into superfolder green fluorescent protein (sfGFP), where the chromophore-forming tyrosine Y66 was replaced with oCNF. See Supporting Information Section S2 for further details on sample preparation.

The variable proximities of the four nitrile probes on PYP to the pCA binding site allowed us to observe locally specific information on subtle structure perturbations when activated pCA binds both in solution and in living cells. Each nitrile-containing PYP variant was previously characterized in the *holo* form (covalently attached to pCA) via X-ray crystallography to high-resolution ($<1.2\ \text{\AA}$), where the nitriles adopt a single, well-defined orientation at each site (Figure S13).²³ F28oCNF (PDB: 7SPX) interacts with one hydrogen bond donor, a water molecule (3.2 \AA heavy atom distance) and is located $\sim 15\ \text{\AA}$ (ring-to-ring distance) from pCA, while F92oCNF (PDB: 7SPV) engages with two donors (2.9 and 3.2 \AA for T90 and a water, respectively) and is positioned $\sim 12\ \text{\AA}$ from the pCA chromophore. In contrast, F62oCNF (PDB: 7SPW) and F96oCNF (PDB: 7SJJ) are situated in hydrophobic environments, with only carbon atoms within 3.5 \AA of the nitrile nitrogen, and they are located $\sim 10\ \text{\AA}$ and $\sim 5\ \text{\AA}$ from the pCA chromophore, respectively. As part of this study, we also determined for the first time a high-resolution X-ray crystal structure of the apo form of PYP (without the nitrile labels; PDB: 9O8V) and observed minimal deviation from the *holo* form, both with and without a nitrile incorporated (Figure 2 and SI Section S3 and Table S6). The absence of the chromophore and nitrile labels did not result in any major conformational changes in the protein structure, except for a lack of density in the disordered loop region (residues 16–20), which is located more than 20 \AA from the chromophore pocket (Figure S14).

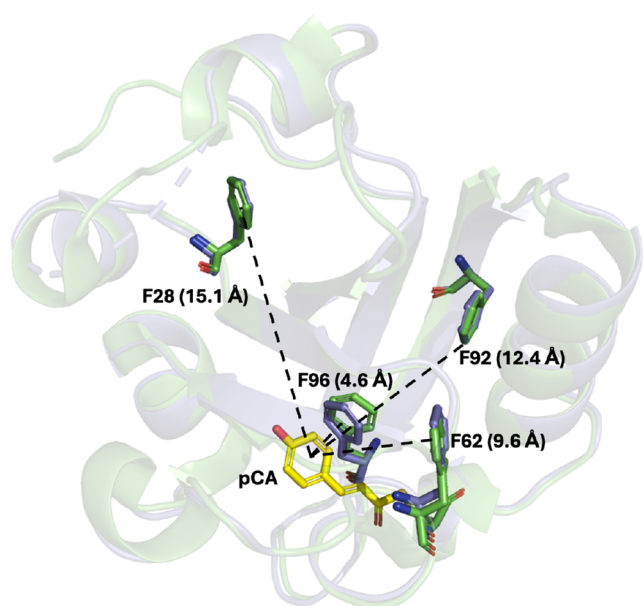


Figure 2. Strategy to study covalent drug binding using photoactive yellow protein (PYP) and nitrile vibrational spectroscopy. oCNF incorporation *via* amber suppression allows site-specific placement of nitrile reporters at F28, F62, F92, and F96. *Holo* (Green, PDB: 1NWZ) and *apo* (Gray, PDB: 9O8V) PYP crystal structures superimposed with the modified Phe residues highlighted.

Utilizing these variants and the sensitive QCL spectrometer, we measured the nitrile absorption spectra in a concentrated suspension of bacterial cells ($OD_{600} = 0.3$ when diluted 1000 \times) expressing individually nitrile-modified Phe variants of *apo*-PYP. Cells were washed three times with minimal media to remove excess unincorporated oCNF (See SI Section S2). A culture expressing unmodified *apo*-PYP (no nitrile) was used as a blank for balanced detection. We could clearly note the nitrile frequencies: for F28oCNF at 2230.8 cm^{-1} , for F62oCNF at 2227.5 cm^{-1} , for F92oCNF at 2226.4 cm^{-1} , and for F96oCNF at 2227.6 cm^{-1} (Figure 3). While similar measurements were attempted using the Bruker 70 FTIR spectrometer with 50 μm spacers (SI Section S4), the low signal-to-noise ratio hindered reliable detection of these peaks (Figure S15 and Table S7). For Y66oCNF sfGFP, the nitrile stretch was also clearly visible at 2225.8 cm^{-1} in cells using the QCL spectrometer (Figure S16).

In all cases, the *in cellulo* nitrile frequencies were distinct from that of free oCNF in aqueous media (2232.3 cm^{-1} in water; Figure S16). Furthermore, when cells expressing nitrile-containing proteins were lysed and purified by affinity column chromatography, the nitrile stretching frequencies of the isolated proteins corresponded closely to those observed in cells (Figures S16–S18). However, when *apo*-PYP nitrile frequencies were compared to previous measurements conducted on the purified *holo*-PYP (with pCA covalently attached), significant differences in peak frequency were noted for variants F92oCNF and F96oCNF (Tables S8–S9). In contrast, the frequencies for F28oCNF and F62oCNF were similar regardless of chromophore incorporation. These findings indicated that some of the nitrile environments in the protein depend on pCA binding, prompting further investigations described below.

Observation of Nitrile Frequency Shifts on pCA-Binding from the QCL-Based IR Measurement. We next

incubated the PYP-expressing bacteria with a ~ 100 -fold molar excess of pCA chemically activated with an acyl imidazolidine electrophilic warhead (SI Section S5, Figures S19–S25).^{23,24,51,52} The nitrile stretch measured by the QCL in bacteria expressing F28oCNF or F62oCNF PYP exhibited negligible shifts after activated pCA incubation, while those expressing F92oCNF and F96oCNF displayed large blue shifts of +14.9 and +6.3 cm^{-1} , respectively, when compared to the *apo* samples. These nitrile frequencies align closely with those of the previously reported, isolated *holo* F92oCNF and F96oCNF proteins.²³ Subsequent purification of these activated pCA-incubated cells yielded a characteristic yellow protein with a UV–vis absorbance maximum at 445 nm, which along with mass spectrometry (SI Section S6, Figure S26) confirmed chromophore incorporation through the cell wall and membrane and into the apoprotein. Additionally, the nitrile frequencies for these purified proteins closely matched the frequencies of both the previously reported purified *holo*-PYP²³ and the QCL-measured *in cellulo* nitrile frequencies of *holo*-PYP (Tables S8–S11).

As a control, we introduced a C69A substitution in the F96oCNF variant (F96oCNF was selected due its ~ 6 cm^{-1} blue shift between the *holo* and *apo* forms). The C69A mutation thereby disrupted the covalent linkage site for pCA in the F96oCNF-PYP variant. Even after activated pCA incubation, the nitrile frequency in this variant remained similar to that of the *apo*-F96oCNF. This confirms that the C69 thiol group is indeed responsible for pCA binding including in the cellular environment, and the covalent linkage of pCA in the binding site accounts for the observed frequency shifts (Figure S18 and Table S12).

Consequences of Drug Binding to *apo*-PYP—Structural Analysis and MD Simulations. We next sought to better understand the origin of the large frequency shifts on pCA incorporation for F92oCNF and F96oCNF PYP. As demonstrated by our group²³ and others,¹⁹ nitrile vibrational frequency shifts are affected by intermolecular interactions—electrostatic stabilization typically causes red shifts, while hydrogen bonding can induce blue shifts with respect to an unperturbed nitrile stretch.²⁰ We have shown previously that a combination of frequency shift and absolute intensity measurements can be used to obtain electric fields for nitriles;²³ however, absolute intensity quantification is unreliable and difficult to perform in living cells. Hence, when the concentration of the vibrational probe cannot be obtained with high accuracy, interpreting a spectroscopically observed frequency shift requires structural knowledge of the local electric fields and hydrogen-bonding populations. We investigated these influences on nitrile frequency shifts in PYP using molecular dynamics simulations.

We first ran 100-ns MD simulations of the *apo* form using the AMBER99SB-ILDN force field.⁵³ This revealed no global conformational changes, whether starting from the *apo* crystallographic pose first reported here (Figure 2) or the *holo* crystallographic pose 1NWZ but with pCA removed (Figure S27). As we have shown in separate work, fixed-charge MD simulations based on force-fields such as AMBER often cannot accurately recapitulate the electrostatic environment around nitrile probes in proteins, thus for these analyses higher-level simulations were carried out using the AMOEBA09 force field (Figure 4).^{54,55} Four replicate AMOEBA simulations of 25 ns each were performed using the F92oCNF (7SPV) and F96oCNF (7SJJ) PYP structures with chromo-

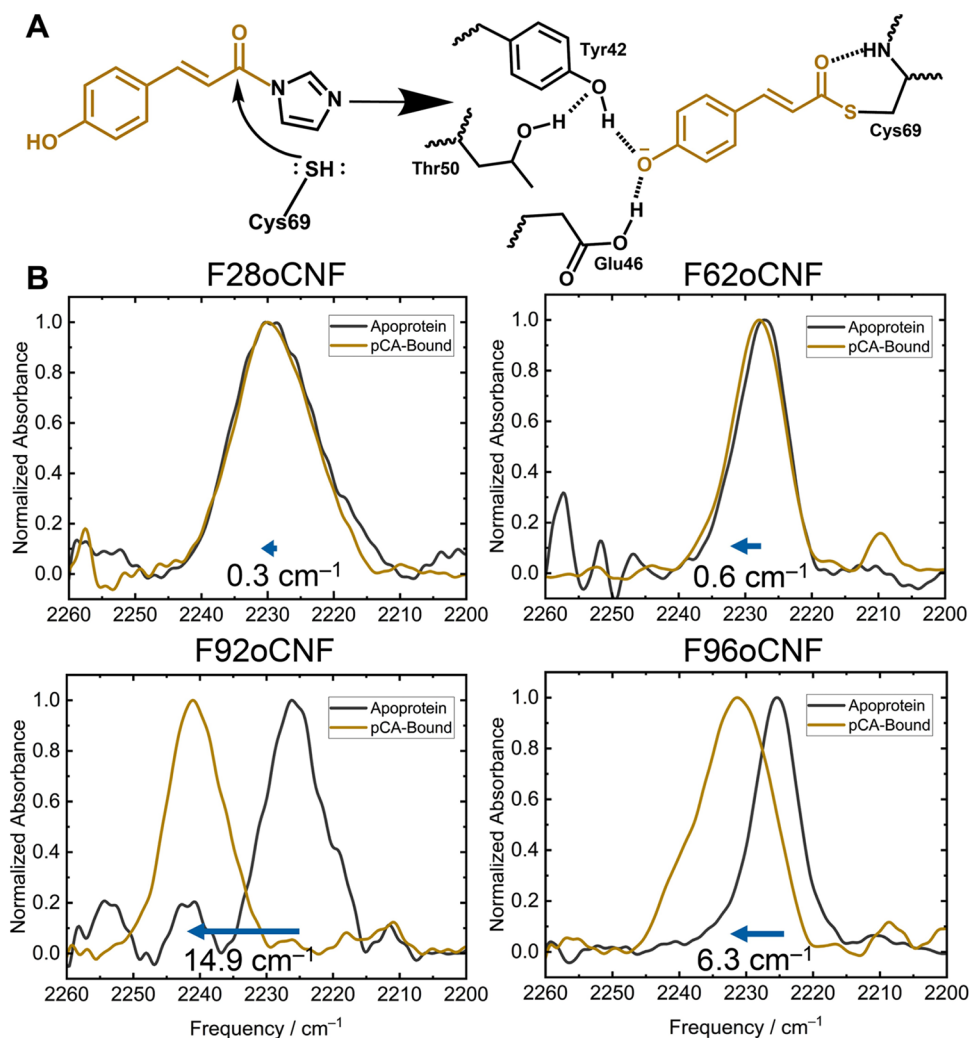


Figure 3. (A) Incorporation of the chromophore p-Coumaric acid (pCA) into PYP proceeds by nucleophilic attack at C69. The form of pCA added to the cells is an activated form with a reactive, electrophilic “warhead” that derives from a water-free reaction of pCA with carbonyldiimidazole. (B) Nitrile vibrational spectra collected in bacteria using the QCL for the F28oCNF, F62oCNF, F92oCNF, and F96oCNF variants show variable nitrile vibrational frequency shifts in response to pCA incorporation, with shifts of +0.3, +0.6, +14.9, and +6.3 cm^{-1} , respectively (analyzed by peak fitting).

phores removed as starting points,²³ again showing no global conformational changes (Figure S28, details in SI Section S7).⁵⁶ These simulations revealed that for *apo*-F92oCNF, the average electric field projected on the nitrile bond was significantly reduced to -44 MV/cm, compared to the -67 MV/cm reported by Kirsh et al. for the *holo* form.²⁴ *Apo*-F96oCNF showed only a slightly smaller average field, -23 MV/cm compared to -26 MV/cm reported for the *holo* form.²⁴ A more stabilizing field on the nitriles in the *holo*-form would typically coincide with a red-shift upon pCA binding. However, the presence of hydrogen bonds as is possible for both nitriles²⁴ obfuscates a purely electrostatic interpretation of the observed frequency differences, in contrast to the transition dipole moment approach used by Weaver et al.²³ and further discussed in ref 24.

For F92oCNF (Figure 4), we clearly observed two distinct electric field states projected on the nitrile bond: a high-field state typically corresponding to a hydrogen-bonded state (primarily with T90 side chain), and a significantly lower-field state corresponding to a non-hydrogen-bonded configuration. Distinct hydrogen-bonding populations agree with previous

low-temperature FTIR experiments of the *holo* form that slow the rate of exchange between the two populations (SI Section S8, Figure S29).²⁴ They also align with previous MD observations of a hydrogen-bonding and a non-hydrogen-bonding state in F92oCNF when the chromophore was present, where much of the simulation trajectory indicated a hydrogen-bonded state.²⁴ The current simulations in the absence of the chromophore significantly favor decreased hydrogen bonding of the nitrile to T90 in *apo*-F92oCNF, consistent with the IR measurements showing a blue shift upon chromophore incorporation. Additionally, AMOEBA simulations reveal a small increase in hydrogen-bonded population (nitrile-to-solvent) for the *apo*-F96oCNF variant upon pCA binding. We note it may be more difficult to reliably interpret simulation results in the vicinity of the chromophore pocket given the more dynamic solvation environment, particularly when pCA is absent and the pocket is solvent-exposed (Figure 2). However, the small but definite change in solvent hydrogen bonding to the nitrile in F96oCNF presents a reasonable explanation for the smaller blue shift undergone by the nitrile upon pCA binding.

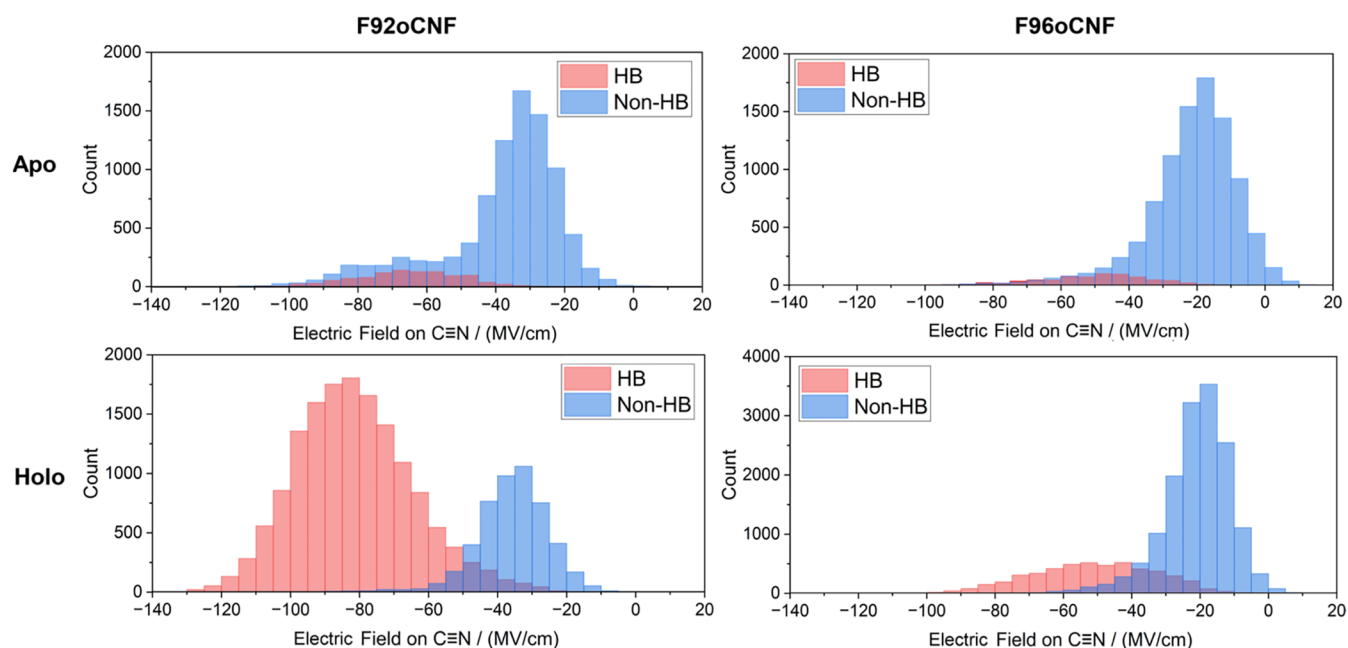


Figure 4. Local electrostatic environment around the nitrile group revealed by AMOEBA molecular dynamics simulations. Simulations were run for four replicates of 25 ns each. The distribution of electric fields experienced by the nitrile is shown, separated by whether the nitrile was engaged in hydrogen bonding. Hydrogen bonding was defined using a heavy-atom nitrile–H-bond donor–acceptor distance cutoff of 4.0 Å and a donor–acceptor–hydrogen angle cutoff of 30°. F92oCNF exhibits two distinct electrostatic states: a non-hydrogen-bonded (low-field) state and a hydrogen-bonded (high-field) state, which significantly shift in population upon pCA binding—correlating with a large $\sim 15\text{ cm}^{-1}$ blue shift in the nitrile stretch. F96oCNF shows a similar but smaller increase in the hydrogen-bonded population, consistent with the more modest 6.3 cm^{-1} blue shift observed upon pCA binding. Notably, the simulated electric field distributions are significantly broader than the experimentally observed line widths, suggesting additional structural constraints or dynamic averaging in the actual protein environment.

DISCUSSION

The IR frequency of a unique vibrational mode such as a nitrile is highly sensitive to changes in its local environment. This makes frequency shifts upon drug binding an indispensable tool for studying drug–protein interactions. However, the low sensitivity of this method *in cellulo* means it has rarely been used to date, instead opting for more traditional fluorescence-based assays. Here, we address this limitation by developing a QCL-based infrared spectrometer capable of sensitively detecting site-specific vibrational shifts in proteins bearing embedded nitrile probes, highlighting its utility for monitoring protein–drug interactions even within live-cell environments. Compared to advanced Raman techniques such as stimulated Raman spectroscopy (SRS), which can also be applied *in cellulo*,⁵⁷ linear QCL spectroscopy leveraging long path lengths can still provide great sensitivity for bulk measurements.⁵⁸ While methods like SRS offer high-resolution imaging and enhanced sensitivity when using vibrational probes conjugated to electronic chromophores (e.g., EPR-SRS, BonFIRE),⁵⁹ the chemical interpretability of spectral intensities is not straightforward, and sensitivity is typically lower when using ordinary vibrational probes due to their inherently lower molecular cross sections. (e.g., 10^5 molecules/ $0.1\text{ fL} \approx 1\text{ mM}$ for $\text{C}=\text{C}$ bonds)^{60,61}

We present the first implementation to our knowledge of quantum cascade laser (QCL)-based transmission IR spectroscopy with balanced detection specifically optimized for the nitrile stretching region ($2000\text{--}2300\text{ cm}^{-1}$), a spectrally transparent window with minimal interference from endogenous biomolecular absorption, enabling high-sensitivity detection in complex biological environments. Based on calculated protein expression levels, we estimate our working

nitrile concentration within each cell to be approximately $140\text{ }\mu\text{M}$ (SI Section S2)—well above our determined $18\text{ }\mu\text{M}$ LOD for aromatic nitriles using the QCL and at a considerably higher SNR (signal-to-noise ratio) when compared to a traditional FTIR instrument (Figures 3 vs S15). This highlights the improved or at least comparable sensitivity of our approach relative to previous methodologies.⁴¹ This high sensitivity also renders the spectrometer well-suited for detecting other nonperturbative vibrational probes in this range such as $\text{N}=\text{N}^+=\text{N}^-$, or even the weaker $\text{C}-\text{D}$ or $\text{C}\equiv\text{C}$ stretches, which typically exhibit oscillator strengths up to 10-fold lower than those of nitriles (Figures S10–S12).^{18,21} For example, nitrile or azide-labeled GPCRs⁶² are an excellent candidate for *in cellulo* IR spectroscopy using the strategy described here and a QCL spectrometer for preservation of a native membrane environment.

Utilizing this spectrometer, we employed photoactive yellow protein (PYP) and its covalently attached chromophore p-Coumaric acid (pCA) as a representative framework for many cysteine-based covalent inhibitors. For example, notable cysteine-reactive covalent drugs such as ibrutinib and acalabrutinib have transformed cancer therapy by selectively inhibiting Bruton's tyrosine kinase (BTK) through a covalent mechanism, with minimal off-target effects on other kinases.^{63,64} The PYP-pCA system provides a structurally and mechanistically analogous framework for investigating the molecular underpinnings of such selective covalent binding. While this work demonstrates covalent binding, the nitrile group's sensitivity to its local environment enables detection of a broad range of molecular interactions, including noncovalent effects and drug molecules reliant on noncovalent binding. This versatility underscores a key advantage of this vibrational

probe system over traditional fluorescence-based assays. Unlike fluorescence approaches that often require the attachment of bulky dye molecules—comparable in size to the drug itself—vibrational probes provide efficient, minimally invasive access to intrinsic molecular information, ranging from electrostatics to protein motion and conformational dynamics.

Such detailed structural insight is largely unique to spectroscopic techniques. In contrast, proteome-wide mass spectrometry assays, although powerful, face limitations. Highly specific biophysical studies of noncovalent interactions such as electrostatics or hydrogen bonding are not currently possible with mass spectrometry. Moreover, in this particular case, proteomic analysis aimed at identifying off-target binding interactions of pCA proved of limited utility. This was likely due to the standard reducing conditions for peptide linearization (involving iodoacetamide and dithiothreitol (DTT)), which can disrupt sensitive covalent linkages such as the thioester bond between PYP and pCA (SI Section S6).⁶⁵ By comparison, the IR assay allows us to include those interactions typically challenging to mass spectrometry⁶⁶—both noncovalent interactions plus covalent bonds sometimes susceptible to cleavage by DTT or tris (2-carboxyethyl)-phosphine hydrochloride (TCEP) reducing agents employed in a standard proteomics digestion. This highlights the unique advantages of IR spectroscopy for probing the highly localized interactions that could potentially govern the chemical reactivity and efficacy of covalent drugs.

Finally, we performed an in-depth analysis of the unique spectral features of the nitriles observed using our QCL-based spectrometer and the PYP-pCA model system. Specifically, we detected distinct hypsochromic shifts in the nitrile stretching frequencies of variants F92oCNF and F96oCNF upon incorporation of the chromophore. To elucidate the origin of these shifts, we conducted molecular dynamics simulations using both fixed-charge and polarizable force fields, which revealed that the frequency changes stem from localized structural and electrostatic rearrangements within the protein environment upon pCA incorporation. The ability of AMOEBA simulations to capture these subtle electrostatic rearrangements in the nitrile microenvironment further validates the use of this polarizable force field for high-level molecular dynamics studies. These findings not only benchmark the simulations but also provide insights of significant interest to force-field developers. These subtle perturbations highlight how the incorporation of a drug-like molecule such as pCA can induce site-specific changes in the protein's electrostatic landscape—even far from the binding site—underscoring the sensitivity of vibrational probes to chemically relevant microenvironments. This becomes especially relevant when the local electrostatic environment of a binding site has been shown to be critical to the reactivity of covalent binding—a concept often overlooked by traditional electrostatic complementarity approaches used in drug design.⁶⁷

Structurally, the crystallographic data seem to suggest a slight reorientation of F96 toward the empty chromophore pocket (Figure 2) and slight increase in hydrogen bonding population (Figure 4). In the case of F92oCNF, chromophore incorporation appears to increase its potential to form a hydrogen bond with T90, whereas in the *apo* state, it remains largely isolated from such interactions (Figure 4). These results are consistent with an “induced fit” model of protein rigidification upon ligand binding, corresponding to an increase in internal hydrogen bonding networks near the

ligand binding site—a trend potentially generalizable to a wide variety of drug targets.^{68,69}

CONCLUSIONS

The findings presented herein validate the capacity of the new QCL-based assay to detect subtle structural perturbations resulting from covalent bond formation at reactive cysteine residues—a common motif in many modern covalent drugs. While our current proof-of-concept studies were conducted in bacterial cells, it is worth noting that bacterial proteins account for a significant fraction of therapeutic targets of interest in modern drug discovery.⁷⁰ Furthermore, with emerging techniques in amber suppression, there is a growing potential to expand this assay to mammalian systems as well, where noncanonical amino acids can be incorporated efficiently.^{71–73} We further note the advantage in covalent drug uptake by mammalian cells, which do not have cell walls, as opposed to the *E. coli* we work with here—along with the relative stability of most true covalent drugs to hydrolysis compared to the challenges of pCA which must be chemically activated immediately prior to use.

Additionally, while in this study the vibrational probes were installed on the protein target, a similar assay could make use of vibrational reporters located on the drug itself—where a large portion of commercially available drugs already contain nitriles, alkynes, azides, or similar vibrational probes.⁷⁴ This will come at the expense of measuring selectivity of a drug for a particular protein, as labels are not specific to a target protein, but will open avenues to looking at the overall fate of the bulk of the drugs added to the cells. Another foreseeable direction with this assay involves directed evolution to optimize molecular interactions in cells. For instance, we found that subculturing pCA-incubated cells in fresh media after the assay allowed for the regeneration of a new batch of cells available for further mutagenesis or future purification. Sensitive QCL spectroscopy enables the observation of vibrational frequency shifts in live cells, offering new opportunities to study molecular interactions in their native environments. These abilities can prove especially useful for understanding functions and properties of membrane and intrinsically disordered proteins, and more applications in which the native cellular environment is key.

ASSOCIATED CONTENT

Data Availability Statement

Spectrometer control MATLAB scripts are available for download at https://github.com/sdefried/QCL_spectrometer_scripts. Data analysis scripts for molecular dynamics simulations are available for download at https://github.com/KozuchLab/Publications/tree/main/oCNPhe_GROMACS_TINKER.

Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/jacs.5c14498>.

Detailed experimental methods and additional data on QCL spectrometer assembly, cell growth, protein purification, X-ray crystallography, FTIR spectroscopy, pCA activation, mass spectrometry and proteomic analysis and computational data (Sections S1–S8) (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Steven G. Boxer – Department of Chemistry, Stanford University, Stanford, California 94305, United States; orcid.org/0000-0001-9167-4286; Email: sboxer@stanford.edu

Authors

Srijit Mukherjee – Department of Chemistry, Stanford University, Stanford, California 94305, United States; orcid.org/0000-0002-6439-9802

Steven D. E. Fried – Department of Chemistry, Stanford University, Stanford, California 94305, United States; orcid.org/0000-0002-4826-2451

Nathalie Y. Hong – Department of Chemistry, Stanford University, Stanford, California 94305, United States

Nahal Bagheri – Department of Chemistry, Stanford University, Stanford, California 94305, United States; Department of Electrical Engineering, Stanford University, Stanford, California 94305, United States

Jacek Kozuch – Fachbereich Physik, Experimentelle Molekulare Biophysik, FU Berlin, D-14195 Berlin, Germany; Forschungsbau SupraFAB, FU Berlin, D-14195 Berlin, Germany

Irimpan I. Mathews – SLAC National Accelerator Laboratory, Menlo Park, California 94025, United States; orcid.org/0000-0001-6254-3519

Jacob M. Kirsh – Department of Chemistry, Stanford University, Stanford, California 94305, United States; Department of Chemistry and Chemical Engineering, California Institute of Technology, Pasadena, California 91125, United States; orcid.org/0000-0002-1444-2913

Complete contact information is available at: <https://pubs.acs.org/10.1021/jacs.5c14498>

Author Contributions

^VS.M. and S.D.E.F. contributed equally to this work.

Notes

The authors declare the following competing financial interest(s): Stanford University, which S.D.E.F., S.M., N.Y.H., N.B., and S.G.B. are affiliated with, has filed a patent application based on this work with S.D.E.F., S.M. and S.G.B. as inventors.

■ ACKNOWLEDGMENTS

The authors thank Dr. Christopher K. Akhgar, Dr. Alicja Dabrowska, Dr. Andreas Schwaighofer, and Prof. Bernhard Lendl (TU Wien) for discussions on QCL spectrometer assembly. This work was supported in part by NIH Grant R35GM118044 (to S.G.B.). S.D.E.F. is supported by the NSF Graduate Research Fellowship (DGE-1656518) and the Stanford CMAD Fellowship; N.B. by the Stanford Bio-X Bowes Fellowship; and J.K. by the DFG Individual Research Grant KO 5464-4 (project ID 493270578). Use of the Stanford Synchrotron Radiation Lightsource (SSRL/SLAC) was supported by the DOE Office of Science under Contract DE-AC02-76SF00515 and NIH Grant P30GM133894. Computational work was supported by the Sherlock cluster (Stanford) and ZEDAT HPC (Freie Universität Berlin, 10.17169/refubium-26754). We also thank the Stanford Chemistry NMR and SUMS facilities for pCA characterization. Special thanks to Mehmet Solyali (Stanford Machine Shop),

Prof. Wei Min, Dr. Bing Xu, and Dr. Garvey McKenzie for their assistance with instrumentation, fruitful discussions and proteomics experiments, respectively.

■ REFERENCES

- (1) Hulme, E. C.; Trevethick, M. A. Ligand binding assays at equilibrium: validation and interpretation. *Br. J. Pharmacol.* **2010**, *161* (6), 1219–1237.
- (2) Vuignier, K.; Schappler, J.; Veuthey, J.-L.; Carrupt, P.-A.; Martel, S. Drug–protein binding: a critical review of analytical tools. *Anal. Bioanal. Chem.* **2010**, *398* (1), 53–66.
- (3) Gavrilidou, A. F. M.; Sokratous, K.; Yen, H.-Y.; De Colibus, L. High-Throughput Native Mass Spectrometry Screening in Drug Discovery. *Front. Mol. Biosci.* **2022**, *9*, No. 837901.
- (4) Stefaniak, J.; Huber, K. V. M. Importance of Quantifying Drug-Target Engagement in Cells. *ACS Med. Chem. Lett.* **2020**, *11* (4), 403–406.
- (5) Michelini, E.; Cevenini, L.; Mezzanotte, L.; Coppa, A.; Roda, A. Cell-based assays: fuelling drug discovery. *Anal. Bioanal. Chem.* **2010**, *398* (1), 227–238.
- (6) Renaud, J.-P.; Chung, C.; Danielson, U. H.; Egner, U.; Hennig, M.; Hubbard, R. E.; Nar, H. Biophysics in drug discovery: impact, challenges and opportunities. *Nat. Rev. Drug Discovery* **2016**, *15* (10), 679–698.
- (7) Deng, L.; Huang, X.; Ren, J. In Situ Study of the Drug–Target Protein Interaction in Single Living Cells by Combining Fluorescence Correlation Spectroscopy with Affinity Probes. *Anal. Chem.* **2020**, *92* (10), 7020–7027.
- (8) Zhu, M.-r.; Du, D.; Hu, J.; Li, L.; Liu, J.; Ding, H.; Kong, X.; Jiang, H.; Chen, K.; Luo, C. Development of a high-throughput fluorescence polarization assay for the discovery of EZH2-EED interaction inhibitors. *Acta Pharmacol. Sin.* **2018**, *39* (2), 302–310.
- (9) Gaborit, N.; Larbouret, C.; Vallaghe, J.; Peyrusson, F.; Bascoul-Mollevi, C.; Crapez, E.; Azria, D.; Chardès, T.; Poul, M.-A.; Mathis, G.; Bazin, H.; Pèlegri, A. Time-resolved fluorescence resonance energy transfer (TR-FRET) to analyze the disruption of EGFR/HER2 Dimers: A new method to evaluate the efficiency of targeted therapy using monoclonal antibodies. *J. Biol. Chem.* **2011**, *286* (13), 11337–11345.
- (10) Kanno, H.; Hiramatsu, K.; Mikami, H.; Nakayashiki, A.; Yamashita, S.; Nagai, A.; Okabe, K.; Li, F.; Yin, F.; Tominaga, K.; Bicer, O. F.; Noma, R.; Kiani, B.; Efa, O.; Büscher, M.; Wazawa, T.; Sonoshita, M.; Shintaku, H.; Nagai, T.; Goda, K.; et al. High-throughput fluorescence lifetime imaging flow cytometry. *Nat. Commun.* **2024**, *15* (1), No. 7376.
- (11) Scott, A. D. Fluorescent Thermal Shift Assays for Identifying Small Molecule Ligands. In *Biophysical Techniques in Drug Discovery*; Canales, A., Ed.; The Royal Society of Chemistry, 2017 DOI: 10.1039/9781788010016-00208.
- (12) Watanabe, D.; Hiroshima, M.; Yasui, M.; Ueda, M. Single molecule tracking based drug screening. *Nat. Commun.* **2024**, *15* (1), No. 8975.
- (13) Hughes, L. D.; Rawle, R. J.; Boxer, S. G. Choose Your Label Wisely: Water-Soluble Fluorophores Often Interact with Lipid Bilayers. *PLoS One* **2014**, *9* (2), No. e87649.
- (14) Cranfill, P. J.; Sell, B. R.; Baird, M. A.; Allen, J. R.; Lavagnino, Z.; de Gruiter, H. M.; Kremers, G.-J.; Davidson, M. W.; Ustione, A.; Piston, D. W. Quantitative assessment of fluorescent proteins. *Nat. Methods* **2016**, *13* (7), 557–562.
- (15) Fried, S. D.; Boxer, S. G. Measuring Electric Fields and Noncovalent Interactions Using the Vibrational Stark Effect. *Acc. Chem. Res.* **2015**, *48* (4), 998–1006.
- (16) Fried, S. D. E.; Zheng, C.; Mao, Y.; Markland, T. E.; Boxer, S. G. Solvent Organization and Electrostatics Tuned by Solute Electronic Structure: Amide versus Non-Amide Carbonyls. *J. Phys. Chem. B* **2022**, *126* (31), 5876–5886.
- (17) Ji, Z.; Mukherjee, S.; Andreo, J.; Sinelshchikova, A.; Peccati, F.; Jiménez-Osés, G.; Wuttke, S.; Boxer, S. G. Electrostatic atlas of non-

covalent interactions built into metal–organic frameworks. *Nat. Chem.* **2025**, *17*, 1920–1927.

(18) Fried, S. D. E.; Mukherjee, S.; Mao, Y.; Boxer, S. G. Environment- and Conformation-Induced Frequency Shifts of C–D Vibrational Stark Probes in NAD(P)H Cofactors. *J. Phys. Chem. Lett.* **2024**, *15* (43), 10826–10834.

(19) Choi, J.-H.; Cho, M. Vibrational solvatochromism and electrochromism of infrared probe molecules containing C≡O, C≡N, C=O, or C–F vibrational chromophore. *J. Chem. Phys.* **2011**, *134* (15), No. 154513.

(20) Kirsh, J. M.; Kozuch, J. Hydrogen Bond Blueshifts in Nitrile Vibrational Spectra Are Dictated by Hydrogen Bond Geometry and Dynamics. *JACS Au* **2024**, *4* (12), 4844–4855.

(21) Zheng, C.; Mao, Y.; Markland, T. E.; Boxer, S. G. Beyond the Vibrational Stark Effect: Unraveling the Large Redshifts of Alkyne C–H Bond in Solvation Environments. *J. Am. Chem. Soc.* **2025**, *147* (7), 6227–6235.

(22) Levinson, N. M.; Boxer, S. G. A conserved water-mediated hydrogen bond network defines bosutinib's kinase selectivity. *Nat. Chem. Biol.* **2014**, *10* (2), 127–132.

(23) Weaver, J. B.; Kozuch, J.; Kirsh, J. M.; Boxer, S. G. Nitrile Infrared Intensities Characterize Electric Fields and Hydrogen Bonding in Protic, Aprotic, and Protein Environments. *J. Am. Chem. Soc.* **2022**, *144* (17), 7562–7567.

(24) Kirsh, J. M.; Weaver, J. B.; Boxer, S. G.; Kozuch, J. Critical Evaluation of Polarizable and Nonpolarizable Force Fields for Proteins Using Experimentally Derived Nitrile Electric Fields. *J. Am. Chem. Soc.* **2024**, *146* (10), 6983–6991.

(25) Lin, Y.-C.; Ren, P.; Webb, L. J. AMOEBA Force Field Predicts Accurate Hydrogen Bond Counts of Nitriles in SNase by Revealing Water–Protein Interaction in Vibrational Absorption Frequencies. *J. Phys. Chem. B* **2023**, *127* (25), 5609–5619.

(26) Shrestha, R.; Cardenas, A. E.; Elber, R.; Webb, L. J. Measurement of the Membrane Dipole Electric Field in DMPC Vesicles Using Vibrational Shifts of p-Cyanophenylalanine and Molecular Dynamics Simulations. *J. Phys. Chem. B* **2015**, *119* (7), 2869–2876.

(27) Sarkar, S.; Patrow, J. G.; Voegtli, M. J.; Pennathur, A. K.; Dawlaty, J. M. Electrodes as Polarizing Functional Groups: Correlation between Hammett Parameters and Electrochemical Polarization. *J. Phys. Chem. C* **2019**, *123* (8), 4926–4937.

(28) Xiong, H.; Lee, J. K.; Zare, R. N.; Min, W. Strong Electric Field Observed at the Interface of Aqueous Microdroplets. *J. Phys. Chem. Lett.* **2020**, *11* (17), 7423–7428.

(29) Lasch, P.; Boese, M.; Pacifico, A.; Diem, M. FT-IR spectroscopic investigations of single cells on the subcellular level. *Vib. Spectrosc.* **2002**, *28* (1), 147–157.

(30) Naumann, D.; Fabian, H.; Lasch, P. FTIR Spectroscopy of Cells, Tissues and Body Fluids. In *Biological and Biomedical Infrared Spectroscopy*, Advances in Biomedical Spectroscopy; IOS Press: Amsterdam, 2009; Vol. 2, pp 312–354 DOI: 10.3233/978-1-60750-045-2-312.

(31) Mahalanabish, A.; Huang, S. H.; Tulegenov, D.; Shvets, G. Infrared Spectroscopy of Live Cells Using High-Aspect-Ratio Metal-on-Dielectric Metasurfaces. *Nano Lett.* **2024**, *24* (37), 11607–11614.

(32) Wat, J. H.; Pizzala, N. J.; Reppert, M. Isotope Reverse-Labeled Infrared Spectroscopy as a Probe of In-Cell Protein Structure. *J. Phys. Chem. B* **2024**, *128* (41), 9923–9934.

(33) Mészáros, L. S.; Ceccaldi, P.; Lorenzi, M.; Redman, H. J.; Pfützner, E.; Heberle, J.; Senger, M.; Stripp, S. T.; Berggren, G. Spectroscopic investigations under whole-cell conditions provide new insight into the metal hydride chemistry of [FeFe]-hydrogenase. *Chem. Sci.* **2020**, *11* (18), 4608–4617.

(34) Goett-Zink, L.; Karsten, L.; Mann, C.; Horstmeier, H.; Spang, J.; Müller, K. M.; Kottke, T. Photochemistry of Receptor-Bound Flavin Resolved in Living Human Cells by Infrared Spectroscopy. *J. Am. Chem. Soc.* **2025**, *147* (11), 9676–9685.

(35) Oldemeyer, S.; la Greca, M.; Langner, P.; Lê Công, K.-L.; Schlesinger, R.; Heberle, J. Nanosecond Transient IR Spectroscopy of

Halorhodopsin in Living Cells. *J. Am. Chem. Soc.* **2024**, *146* (28), 19118–19127.

(36) Yao, Y.; Hoffman, A. J.; Gmachl, C. F. Mid-infrared quantum cascade lasers. *Nat. Photonics* **2012**, *6* (7), 432–439.

(37) Bhargava, R. Digital Histopathology by Infrared Spectroscopic Imaging. *Annu. Rev. Anal. Chem.* **2023**, *16* (1), 205–230.

(38) Alcaráz, M. R.; Schwaighofer, A.; Kristament, C.; Ramer, G.; Brandstetter, M.; Goicoechea, H.; Lendl, B. External-Cavity Quantum Cascade Laser Spectroscopy for Mid-IR Transmission Measurements of Proteins in Aqueous Solution. *Anal. Chem.* **2015**, *87* (13), 6980–6987.

(39) Schwaighofer, A.; Montemurro, M.; Freitag, S.; Kristament, C.; Culzoni, M. J.; Lendl, B. Beyond Fourier Transform Infrared Spectroscopy: External Cavity Quantum Cascade Laser-Based Mid-Infrared Transmission Spectroscopy of Proteins in the Amide I and Amide II Region. *Anal. Chem.* **2018**, *90* (11), 7072–7079.

(40) Dabrowska, A.; David, M.; Freitag, S.; Andrews, A. M.; Strasser, G.; Hinkov, B.; Schwaighofer, A.; Lendl, B. Broadband laser-based mid-infrared spectroscopy employing a quantum cascade detector for milk protein analysis. *Sens. Actuators, B* **2022**, *350*, No. 130873.

(41) Akhgar, C. K.; Ramer, G.; Žbik, M.; Trajnerowicz, A.; Pawluczky, J.; Schwaighofer, A.; Lendl, B. The Next Generation of IR Spectroscopy: EC-QCL-Based Mid-IR Transmission Spectroscopy of Proteins with Balanced Detection. *Anal. Chem.* **2020**, *92* (14), 9901–9907.

(42) Hermann, D. R.; Ramer, G.; Kitzler-Zeiler, M.; Lendl, B. Quantum Cascade Laser-Based Vibrational Circular Dichroism Augmented by a Balanced Detection Scheme. *Anal. Chem.* **2022**, *94* (29), 10384–10390.

(43) Meyer, T. E. Isolation and characterization of soluble cytochromes, ferredoxins and other chromophoric proteins from the halophilic phototrophic bacterium *Ectothiorhodospira halophila*. *Biochim. Biophys. Acta, Bioenerg.* **1985**, *806* (1), 175–183.

(44) Baillie, T. A. Targeted Covalent Inhibitors for Drug Design. *Angew. Chem., Int. Ed.* **2016**, *55* (43), 13408–13421.

(45) Maurais, A. J.; Weerapana, E. Reactive-cysteine profiling for drug discovery. *Curr. Opin. Chem. Biol.* **2019**, *50*, 29–36.

(46) Visscher, M.; Arkin, M. R.; Dansen, T. B. Covalent targeting of acquired cysteines in cancer. *Curr. Opin. Chem. Biol.* **2016**, *30*, 61–67.

(47) Huang, F.; Han, X.; Xiao, X.; Zhou, J. Covalent Warheads Targeting Cysteine Residue: The Promising Approach in Drug Development. *Molecules* **2022**, *27* (22), No. 7728.

(48) Reis, J. M.; Xu, X.; McDonald, S.; Woloschuk, R. M.; Jaikaran, A. S. I.; Vizeacoumar, F. S.; Woolley, G. A.; Uppalapati, M. Discovering Selective Binders for Photoswitchable Proteins Using Phage Display. *ACS Synth. Biol.* **2018**, *7* (10), 2355–2364.

(49) Brechun, K. E.; Zhen, D.; Jaikaran, A.; Borisneko, V.; Kumauchi, M.; Hoff, W. D.; Arndt, K. M.; Woolley, G. A. Detection of Incorporation of p-Coumaric Acid into Photoactive Yellow Protein Variants in Vivo. *Biochemistry* **2019**, *58* (23), 2682–2694.

(50) Verma, P. K.; Kundu, A.; Poretz, M. S.; Dhoonmoon, C.; Chegwidden, O. S.; Londergan, C. H.; Cho, M. The Bend+Libration Combination Band Is an Intrinsic, Collective, and Strongly Solute-Dependent Reporter on the Hydrogen Bonding Network of Liquid Water. *J. Phys. Chem. B* **2018**, *122* (9), 2587–2599.

(51) Genick, U. K.; Devanathan, S.; Meyer, T. E.; Canestrelli, I. L.; Williams, E.; Cusanovich, M. A.; Tollin, G.; Getzoff, E. D. Active Site Mutants Implicate Key Residues for Control of Color and Light Cycle Kinetics of Photoactive Yellow Protein. *Biochemistry* **1997**, *36* (1), 8–14.

(52) Engstrom, K. M. Practical Considerations for the Formation of Acyl Imidazolides from Carboxylic Acids and N,N'-Carbonyldiimidazole: The Role of Acid Catalysis. *Org. Process Res. Dev.* **2018**, *22* (9), 1294–1297.

(53) Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Kepeis, J. L.; Dror, R. O.; Shaw, D. E. Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins* **2010**, *78* (8), 1950–1958.

(54) Ponder, J. W.; Wu, C.; Ren, P.; Pande, V. S.; Chodera, J. D.; Schnieders, M. J.; Haque, I.; Mobley, D. L.; Lambrecht, D. S.; DiStasio, R. A.; Head-Gordon, M.; Clark, G. N. I.; Johnson, M. E.; Head-Gordon, T. Current status of the AMOEBA polarizable force field. *J. Phys. Chem. B* **2010**, *114* (8), 2549–2564.

(55) Welborn, V. V.; Head-Gordon, T. Fluctuations of Electric Fields in the Active Site of the Enzyme Ketosteroid Isomerase. *J. Am. Chem. Soc.* **2019**, *141* (32), 12487–12492.

(56) Thomson, B.; Both, J.; Wu, Y.; Parrish, R. M.; Martínez, T. J.; Boxer, S. G. Perturbation of Short Hydrogen Bonds in Photoactive Yellow Protein via Noncanonical Amino Acid Incorporation. *J. Phys. Chem. B* **2019**, *123* (23), 4844–4849.

(57) Wei, L.; Hu, F.; Shen, Y.; Chen, Z.; Yu, Y.; Lin, C.-C.; Wang, M. C.; Min, W. Live-cell imaging of alkyne-tagged small biomolecules by stimulated Raman scattering. *Nat. Methods* **2014**, *11*, 410–412.

(58) Wang, H.; Lee, D.; Wei, L. Multimodal chemical imaging for spatial proteomics and metabolomics. *Chem. Biomed. Imaging* **2023**, *1* (1), 3–17.

(59) Wang, H.; Lee, D.; Cao, Y.; Bi, X.; Du, J.; Miao, K.; Wei, L. Bond-selective fluorescence imaging with single-molecule sensitivity. *Nat. Photonics* **2023**, *17*, 846–855.

(60) Hu, F.; Shi, L.; Min, W. Biological imaging of chemical bonds by stimulated Raman scattering microscopy. *Nat. Met.* **2019**, *16*, 830–842.

(61) Gao, X.; Min, W. Quantum theory of stimulated Raman scattering microscopy. *Chem. Phys. Rev.* **2025**, *6*, No. 021306, DOI: [10.1063/5.0248085](https://doi.org/10.1063/5.0248085).

(62) Ye, S.; Huber, T.; Vogel, R.; Sakmar, T. P. FTIR analysis of GPCR activation using azido probes. *Nat. Chem. Biol.* **2009**, *5*, 397–399.

(63) Honigberg, L. A.; Smith, A. M.; Sirisawad, M.; Buggy, J. J.; et al. The Bruton tyrosine kinase inhibitor PCI-32765 blocks B-cell activation and is efficacious in models of autoimmune disease and B-cell malignancy. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107* (29), 13075–13080.

(64) Asadi, M.; Xie, W. J.; Warshel, A. Exploring the role of chemical reactions in the selectivity of tyrosine kinase inhibitors. *J. Am. Chem. Soc.* **2022**, *144* (36), 16638–16646.

(65) Hoff, W. D.; Devreese, B.; Fokkens, R.; Nugteren-Roodzant, I. M.; Van Beeumen, J.; Nibbering, N.; Hellingwerf, K. J. Chemical Reactivity and Spectroscopy of the Thiol Ester-Linked p-Coumaric Acid Chromophore in the Photoactive Yellow Protein from *Ectothiorhodospira halophila*. *Biochemistry* **1996**, *35* (4), 1274–1281.

(66) Meissner, F.; Geddes-McAlister, J.; Mann, M.; Bantscheff, M. The emerging role of mass spectrometry-based proteomics in drug discovery. *Nat. Rev. Drug Discovery* **2022**, *21*, 637–654.

(67) Ji, Z.; Boxer, S. G. β -Lactamases evolve against antibiotics by acquiring large active-site electric fields. *J. Am. Chem. Soc.* **2022**, *144* (48), 22289–22294.

(68) Stone, M. J. NMR Relaxation Studies of the Role of Conformational Entropy in Protein Stability and Ligand Binding. *Acc. Chem. Res.* **2001**, *34* (5), 379–388.

(69) Di Cera, E. Mechanisms of Ligand Binding. *Biophys. Rev.* **2020**, *1*, No. 011303.

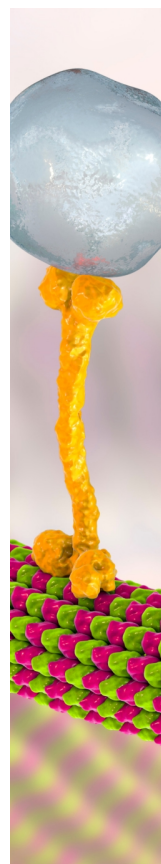
(70) Santos, R.; Ursu, O.; Gaulton, A.; Bento, A. P.; Donadi, R. S.; Bologa, C. G.; Karlsson, A.; Al-Lazikani, B.; Hersey, A.; Oprea, T. I.; Overington, J. P. A comprehensive map of molecular drug targets. *Nat. Rev. Drug Discovery* **2017**, *16* (1), 19–34.

(71) Hino, N.; Hayashi, A.; Sakamoto, K.; Yokoyama, S. Site-specific incorporation of non-natural amino acids into proteins in mammalian cells with an expanded genetic code. *Nat. Protoc.* **2006**, *1* (6), 2957–2962.

(72) Nygaard, A.; Zachariassen, L. G.; Larsen, K. S.; Kristensen, A. S.; Loland, C. J. Fluorescent non-canonical amino acid provides insight into the human serotonin transporter. *Nat. Commun.* **2024**, *15* (1), No. 9267.

(73) De Faveri, C.; Mattheisen, J. M.; Sakmar, T. P.; Coin, I. Noncanonical amino acid tools and their application to membrane protein studies. *Chem. Rev.* **2024**, *124* (22), 12498–12550.

(74) Ertl, P.; Altmann, E.; McKenna, J. M. The Most Common Functional Groups in Bioactive Molecules and How Their Popularity Has Evolved over time. *J. Med. Chem.* **2020**, *63* (15), 8408–8418.



CAS BIOFINDER DISCOVERY PLATFORM™

BRIDGE BIOLOGY AND CHEMISTRY FOR FASTER ANSWERS

Analyze target relationships,
compound effects, and disease
pathways

Explore the platform

CAS
A Division of the
American Chemical Society

Supplementary Information

Covalent Drug Binding in Live Cells Monitored by Mid-Infrared Quantum Cascade Laser Spectroscopy: Photoactive Yellow Protein as a Model System

Srijit Mukherjee^{1,†}, Steven D. E. Fried^{1,†}, Nathalie Y. Hong¹, Nahal Bagheri^{1,2}, Jacek Kozuch^{3,4}, Irimpan I. Mathews,⁵ Jacob M. Kirsh,^{1,6} and Steven G. Boxer^{1*}

[†] Authors contributed equally

* Corresponding author

¹Department of Chemistry, Stanford University, Stanford CA 94305, USA

²Department of Electrical Engineering, Stanford University, Stanford CA 94305, USA

³Fachbereich Physik, Experimentelle Molekulare Biophysik, FU Berlin, Arnimallee 14, D-14195 Berlin, Germany

⁴Forschungsbau SupraFAB, FU Berlin, Altensteinstr. 23a, D-14195 Berlin, Germany

⁵SLAC National Accelerator Laboratory, Menlo Park CA -94025, USA

⁶Department of Chemistry and Chemical Engineering, California Institute of Technology, Pasadena, CA 91125, USA

Contents

S1. QCL Spectrometer Assembly (Figs. S1-S12, Tables S1-S4)	2
S2. Cell Growth and Protein Purification (Table S5)	16
S3. X-ray Crystallography (Fig. S13-S14, Table S6)	22
S4. FTIR Spectroscopy (Figs. S15-S18, Tables S7-S12)	26
S5. Activation of p-Coumaric Acid (Figs. S19-S25)	32
S6. Mass Spectrometry (Fig. S26)	38
S7. Computational Details (Fig. S27-S28)	40
S8. Physical interpretability of nitrile shifts (Fig. S29)	43
Supplemental References	44

Section S1: QCL Spectrometer Assembly

a. Laser performance and optical assembly: The assembled double-beam transmission infrared spectrometer (Main Text Fig. 1C) uses an external-cavity quantum cascade laser (EC-QCL) as the light source (Daylight Solutions MIRcat-QT-2000 with M2048-P tunable module, cooled to 19°C using a closed-loop water chiller). The instrument design is based on the design of Akhgar et al.,¹ with several additional modifications that allow us to perform the desired analyses as described in the following sections. Firstly, we employ a different laser module to account for a new mid-IR spectral window (2000–2300 cm^{-1} , Fig. S1) in which the instrument is equipped to make high-sensitivity measurements with uniquely long pathlengths up to 250 μm when making measurements in water, where a weak but discernible IR absorption band is centered at roughly 2100 cm^{-1} . This is attributed to the combination of the HOH bending mode ($\sim 1630 \text{ cm}^{-1}$) with a librational mode ($\sim 400\text{--}700 \text{ cm}^{-1}$). The resulting combination band appears near 2100–2130 cm^{-1} . This unique capability of using longer pathlengths due to the high flux of this QCL module makes it ideal for analytes bearing oscillators such as $\text{C}\equiv\text{N}$, $\text{C}\text{--}\text{D}$, $\text{N}=\text{N}^+=\text{N}^-$, and $\text{C}\equiv\text{C}$ stretches. This module provides pulsed output up to 30% duty cycle and high output power for high signal-to-noise ratios [SNR: $>1 \text{ W}$ (peak) and $>0.5 \text{ W}$ (average)]. The laser control was implemented using custom scripts and MATLAB libraries provided by Daylight Solutions, which were modified for our specific use (Available at https://github.com/sdefried/QCL_spectrometer_scripts). For optical alignment, a collinear HeNe laser (632.8 nm) was used, which is integrated with the module and shares the same exit aperture as the infrared beam.

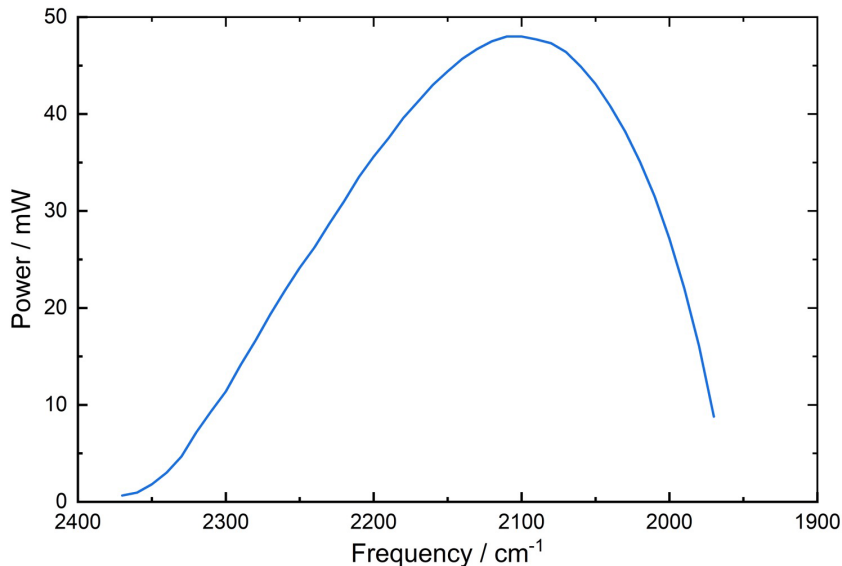


Figure S1. Spectral power of the MIRcat-QT-2000 with M2048-P at 5% duty cycle: Spectral power distribution of the MIRcat-QT-2000 with M2048-P tunable module, measured with a ThorLabs power meter measured at the exit aperture at 5% duty cycle and full current (780 mA). In the experiments to follow, a 30% duty cycle was used.

The laser beam enters a box containing all optics that is purged with dry air (Altec Air CO₂-PG28 Purge Gas Generator operated at 60 psi). The beam is split by a 50:50 ThorLabs CaF₂ Plate Beamsplitter (Coating: 2–8 μm , Cat. No. BSW510). Mid-IR-enhanced gold mirrors (ThorLabs

Cat. No. PF10-03-M02) are used to define the optical path. Infrared-compatible neutral density (ND) filters (Andover Optics Cat. No. 150FNIR-25; OD. = 1.5, 100FNIR-25; OD. = 1.0, and 030FNIR-12.5; OD. = 0.3) are used when applicable to equalize the intensity of the two beams or bring the overall power intensity into the dynamic range of the instrument (only when aqueous samples are not present that attenuate the light due to the libration and bending combination mode of water centered at $\sim 2100\text{cm}^{-1}$). CaF_2 diffusers were optionally included (Edmund Optics Cat. No. 19-734) to scramble polarization prior to detection. The two beams pass through sample and reference demountable liquid IR cells (Pike Technologies, Madison, WI, USA) designed for standard FTIR spectrometers (2 x 3" plate), which allowed a quick comparison of the same sample across different IR spectrometers used in the study. This cell features a needle plate with Luer-Lok fittings for convenient syringe filling with $\sim 75\ \mu\text{L}$ sample. The window dimensions are 32 x 3 mm, providing a 13 mm clear aperture. An O-ring seal model was utilized, employing two small O-rings around the window filling holes instead of a flat gasket. A pathlength of 250 μm for each cell was used by combining two circular Teflon spacers of 200 μm and 50 μm each. Cell holders are machined aluminum blocks that hold the demountable liquid cells in place using screws.

Inclusion of ND-filters and diffusers for balancing, and new CaF_2 focusing optics to increase incident power at the detector elements increase signal-to-noise specific for the detection of molecular vibrations in the spectral region ($2000\text{--}2300\ \text{cm}^{-1}$). Note that the full power of the beam was used in conjunction with the 250- μm pathlength aqueous samples, with only attenuation of one beam relative to the other to equalize incident power on the detector. While the laser beam waist is $< 2.5\ \text{mm}$, the beam is not focused prior to transmission through the sample cells, meaning the beam width is similar to but slightly less than that of the FTIR (6-mm aperture). The only focusing optics, consisting of CaF_2 plano-convex lenses (25.4 mm diameter, 100 mm focal length, Newport Optics Cat. No. CAPX13), are used to focus both the sample and reference beams after sample transmission into a custom-built thermoelectrically cooled balanced HgCdTe (MCT) detector, supplied by VIGO Photonics (PVI-4TE series), to detect the transmitted IR signal from both beams.¹ We note that further modification of the setup by using additional focusing optics may be possible for additional optimization.

Additional performance improvements may also arise from the specific detector modules used, as there is currently no single, unified head-to-head performance comparison published—either by manufacturers or third parties—between the VIGO PVI-4TE series and other commonly used liquid nitrogen-cooled MCT detectors, such as the Teledyne Judson models. However, detailed technical datasheets and catalogs are available for both to provide comparative performance metrics for HgCdTe detectors which we have summarized below. These resources allow for an indirect but robust comparison of key parameters such as detectivity, responsivity, noise, and spectral range.

Table S1. Parameters comparing thermoelectrically cooled VIGO HgCdTe detector used in QCL spectrometer and liquid nitrogen-cooled Teledyne Judson HgCdTe detector used in Bruker FTIR.

Parameter	PVI-4TE (VIGO)	Teledyne Judson MCT (J15D)
Cooling	4-stage TE	TE, LN ₂ , Stirling
Spectral Range (μm)	2–12 (model-dependent)	2–26 (model-dependent)
Detectivity D* (cm·Hz ^{1/2} /W)	≥5×10 ⁹ (At λ _{peak} , TE; For 4TE-8 – used here)	Up to ~10 ¹¹ (LN ₂ , at λ _{peak})
Time Constant (ns)	≤3 (at λ _{peak})	~10–100 (model-dependent)
Window	Wedged ZnSe AR	Multiple options
Area	1×1 mm ² (typical)	0.1–4 mm ² (typical)
Dark Current	Very low (deep TE)	Low (TE), very low (LN ₂)

b. Data acquisition, dynamic range and power calibrations: The detector has three output channels—namely, signal (S), reference (R), and balanced (B = S – R) channels. Channel readout voltages from S, R and B are digitized by a high-speed lock-in amplifier (Stanford Research Systems SR865A 4 MHz dual-phase lock-in amplifier) whose reference signal is dictated by the laser pulse rate (1 MHz at 30% duty cycle). The lock-in amplifier sends capture buffers of S, R, and B covering laser scans from 2360-1970 cm⁻¹ to a host computer for analysis with an in-house MATLAB script that controls both data capture and laser operation. Digitized signal packets (Fig. S2) are processed with phase correction from the lock-in detector, Savitsky-Golay filtering, removal of anomalous spectra using similarity indices, scan averaging, and Fourier filtering, like that reported by Lendl and coworkers.²

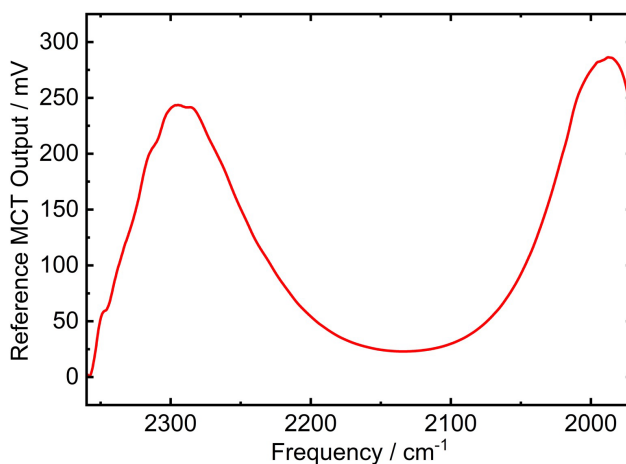


Figure S2. Processed MCT output from QCL Module M2048-P scan through water-filled cell at 30% duty cycle: Digitally processed reference MCT voltage output from 1970-2360 cm⁻¹ scan of the M2048-P tunable module at 30% duty cycle through a 250-μm sample cell filled with water. Although the laser's peak power is near 2100 cm⁻¹ (Fig. S1), water's combination mode at the same frequency significantly attenuates the beam in this frequency region. Output voltage was consistently < 300 mV for all spectral measurements reported herein.

Additionally, to increase sensitivity and make use of the full flux of photons from the laser (and extend the dynamic range of the detector), we externally calibrated the nonlinear detector voltage response in terms of laser power using a ThorLabs power meter (PM100D with S302C Thermal Power Sensor for 0.19 - 25 μm). During calibration, samples are absent in the beam path; therefore, up to three Andover Optics neutral density filters (optical densities 0.3, 1.0, and 1.5) were employed to attenuate the beam and maintain the MCT detector response within the linear dynamic range. At three different wavelengths, laser power was controlled by varying the current supplied to the QCL chip between 400 and 780 mA, while the beam intensity at both the signal and reference MCT were measured using both the MCT detector itself plus the thermal power meter. The detector voltage response to power demonstrates saturating behavior around 470 mV for both MCTs, which was well-captured using a rational instrument response function of the form $y = ax/(1 - bx)$ for y = incident power at MCT and x = MCT output voltage. The best-fit function to the data was used to convert measured detector voltage to beam incident power (Fig. S3). This calibration step was implemented to ensure maximum photon flux and optimal pathlength usage within our spectral range.¹

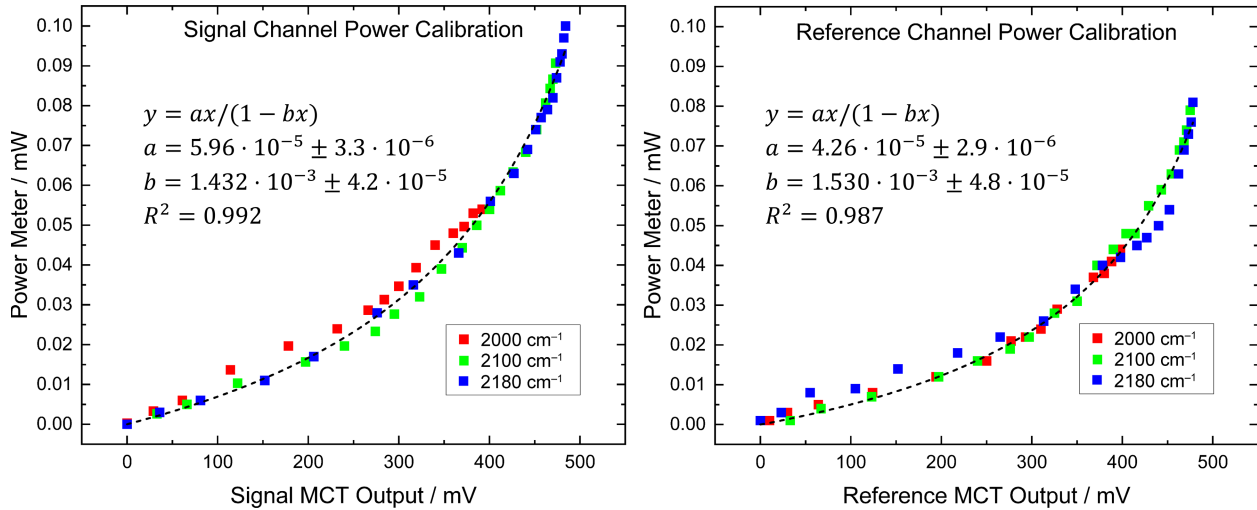


Figure S3. Calibration of MCT detector output to incident power using nonlinear curve fitting: Calibration against a thermal power meter is used to convert MCT detector output voltages to incident power on the detector, used to extend the dynamic range of the instrument given the nonlinear response of the detector at high photon incidence. Detector output voltages were consistently kept below 300 mV when transmitting through a 250- μm liquid sample cell filled with water, thus avoiding the saturation regime of the detector response, and increasing the accuracy of the transmission measurements. Fitting to the nonlinear response function was accomplished using the MATLAB curve fitting toolbox with a nonlinear least squares regression, fitting to all the data points (three separate wavelengths) at once. Errors reported are 95% confidence intervals of the parameters.

c. Frequency calibration and data capture: The frequency axes of the spectra were initially determined using the programmed laser start frequency (2360 cm^{-1}), at which a trigger was sent to the SR865A lock-in amplifier to begin a 32-kb data capture buffer, along with the SR865A capture rate (9765.6 Hz) and laser scan rate (2000 cm^{-1}/s). However, due to slight offsets between

these frequencies calculated using time and the true laser frequencies, we recalibrated the frequency axis using a linear function. To do this, we used laser wavelength TTL triggers every 10 cm^{-1} to subtract off from the MCT output as a separate channel in the lock-in amplifier, generating defined features in the voltage output capture buffer every 10 cm^{-1} . Correlating these with the original frequencies, we generated a linear calibration function that converted the time-based frequencies to corrected frequencies used as the axis for all absorption spectra reported herein (Fig. S4). By matching frequency signatures such as benzonitrile to known literature values,³ we verify that the calibration is accurate.

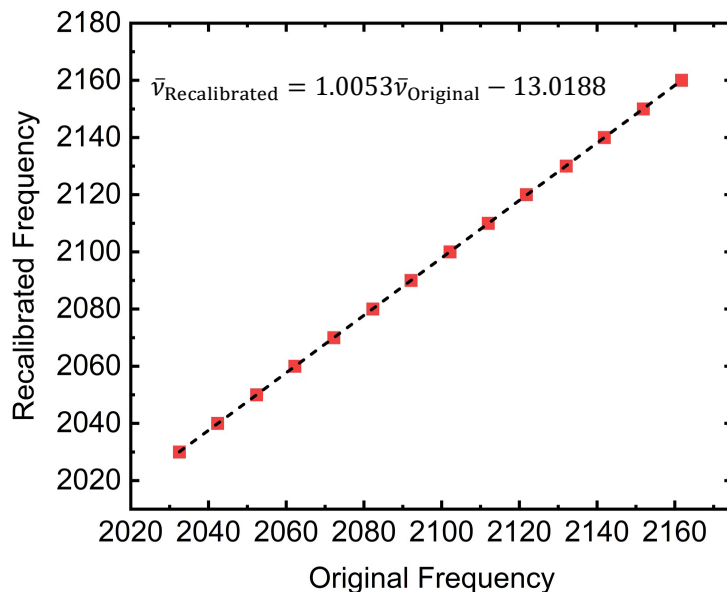


Figure S4. Frequency calibration: Recalibration of original laser scan frequencies determined by time from initial scan trigger, based on periodic wavelength triggers from the QCL.

d. Absorbance measurements using the balanced detection scheme: Absorbance of a sample relative to a reference blank was determined using a balanced detection scheme, which involves multiple measurements across different channels. Both a reference channel measurement of power versus laser frequency, and a balanced channel measurement of power difference versus laser frequency were obtained when reference blanks were loaded into both liquid cells. These are followed by loading the analyte of interest into the sample liquid cell and making a second balanced channel measurement of power difference versus frequency. Absorbance (A) is then calculated using the equation

$$A = \log [P_{\text{Signal}}(V_{\text{Ref}} + V_{\text{Bal-Blank}})/P_{\text{Signal}}(V_{\text{Ref}} + V_{\text{Bal-Sample}})],$$

where V_{Ref} is the reference channel MCT output voltage and V_{Bal} represents the balanced channel (subtracted) MCT output voltage, whether for both liquid cells filled with blanks ($V_{\text{Bal-Blank}}$) or the sample cell filled with sample against the reference filled with blank ($V_{\text{Bal-Sample}}$). The function $P_{\text{Signal}}(V)$ takes an MCT detector voltage as input and gives the beam power incident on the signal detector, according to the above calibration, $P_{\text{Signal}} = aV/(1 - bV)$. This method ensures that

background noise is eliminated using balanced detection and that the measured absorbance is consistent with the Beer-Lambert law.

e. Determination of experimental limit of detection: Following the above process, we demonstrated the capabilities of the QCL spectrometer for sensitively measuring vibrational spectra of benzonitrile in water with a 250 μm pathlength, as shown in Fig. 1 of the main text. Spectra of benzonitrile at higher concentrations, measured on both the FTIR and QCL spectrometer, are shown below, along with the concentration/absorbance correlation (Fig. S5).

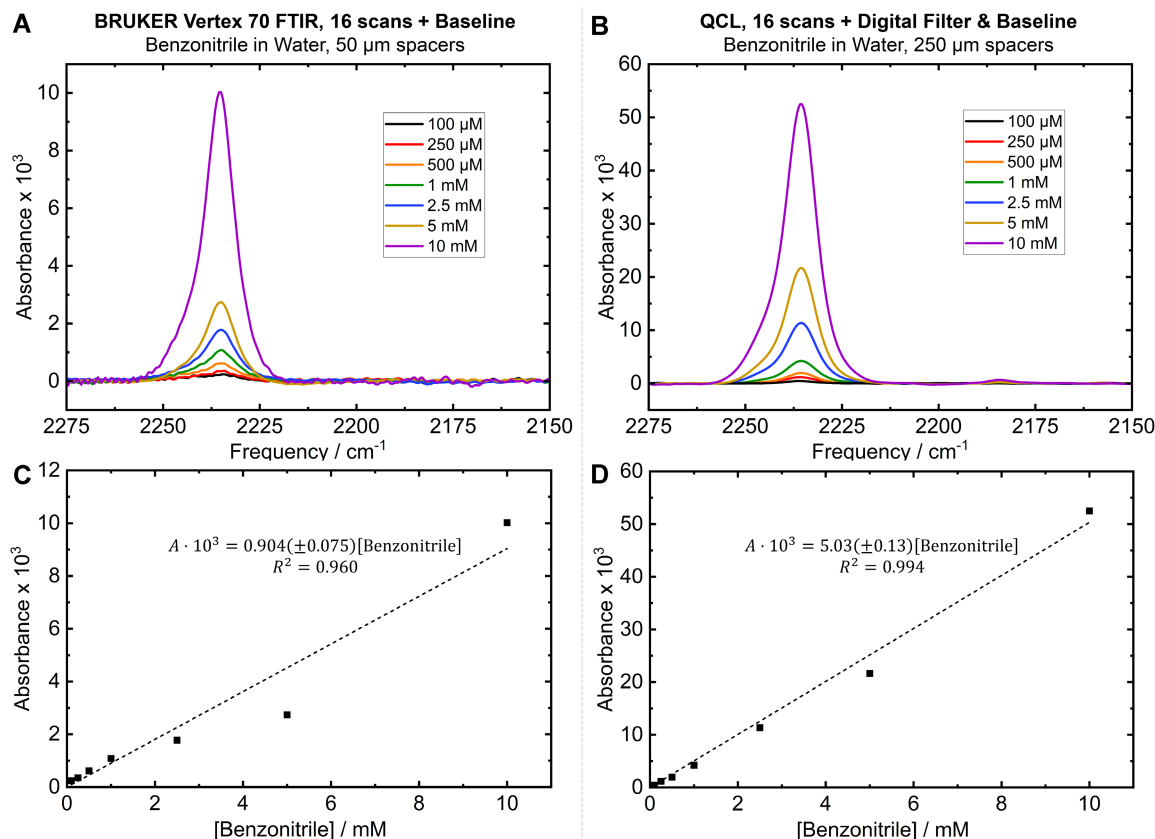


Figure S5. Benzonitrile spectra and concentration-absorbance correlations from FTIR and QCL measurements: Spectra of benzonitrile at concentrations of 100 μM – 10 mM are shown using 16 scans each on either (A) the Bruker Vertex 70 FTIR or (B) the QCL spectrometer. Correlation plots between concentration and absorbance are also shown for both (C) the FTIR and (D) the QCL spectrometer, with linear behavior according to the Beer-Lambert law, particularly in the case of the QCL spectrometer. Errors reported are standard errors. Note that the data for the FTIR carries a higher linear regression error compared to the QCL although the same samples were used for both. We attribute this to the relative difficulty in distinguishing peaks from solvent background when the absorption intensities are lower in the case of the FTIR.

The extinction coefficient of benzonitrile could be determined from both instruments. Firstly, the exact pathlength of both the FTIR and QCL spectrometer sample cells were determined using periodic spacings of etalon fringes as reported by Weaver et al.³ The true pathlength differed

slightly from the spacer widths (relating to how much the spacers were compressed), with the FTIR sample cell having pathlength 55.9 μm and the QCL sample cell having pathlength 259.4 μm . Using these values and the slopes reported below, extinction coefficients the benzonitrile nitrile stretch were calculated as $\epsilon = 162 (\pm 26) \text{ M}^{-1} \text{ cm}^{-1}$ for the FTIR and $\epsilon = 194 (\pm 10) \text{ M}^{-1} \text{ cm}^{-1}$ for the QCL (errors as 95% confidence intervals). We note that both values are within error of the $\epsilon = 186 \text{ M}^{-1} \text{ cm}^{-1}$ value reported previously,³ while meanwhile the peak maximum of 2235.5 cm^{-1} (the more important metric for the current study) similarly agrees well, within 1 cm^{-1} , of our previous value for the frequency of benzonitrile in water.³

f. Determination of theoretical limit of detection [LOD] and RMS noise analysis: Using the calibrated instrument response to benzonitrile and the RMS noise from a blank spectrum (water versus water), we determined the theoretical limit of detection for benzonitrile for both the FTIR and QCL (Fig. S6, Table S2). The RMS noise (σ_{RMS}) was measured over the range of 2260-2160 cm^{-1} , a relatively stable 100 cm^{-1} window with high photon flux even through 250 μm aqueous samples, which also includes benzonitrile's peak of interest (Fig. S2). Theoretical limit of detection was defined by the concentration of benzonitrile required to achieve a signal at least three times the RMS noise over this region, that is, $[\text{LOD}] = 3\sigma_{\text{RMS}}/(\epsilon \cdot l)$, for extinction coefficient ϵ and pathlength l , where $\epsilon \cdot l$ can be obtained from the slope of the instrument response function from Fig S5. As an additional point of comparison, we also include the scenario of detection based on the calibrated signal channel alone, without the use of balanced detection between two MCTs.

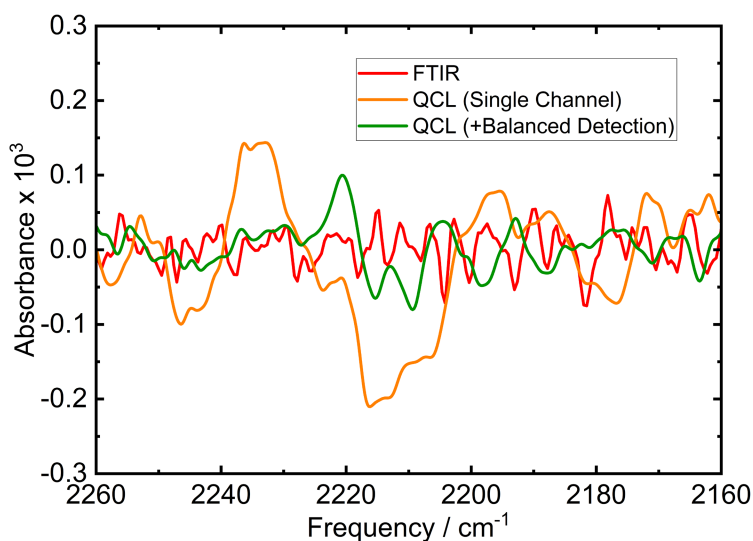


Figure S6. Baseline spectra for estimating theoretical LOD. Baseline IR spectra recorded by the QCL spectrometer using 16 scans, compared to the FTIR. Note that in all cases, 16 scans were used for ease of comparison, and the same procedure for digital processing of the QCL data was applied as described above. A manual baseline correction was employed to all spectra by fitting to broad, low-frequency polynomials to center the higher-frequency noise at zero across the spectral range.

Table S2. Limits of detection across different detection schemes (FTIR vs. QCL in balanced and single channel absorbance).

	FTIR	QCL, Single Channel Only	QCL, Balanced Detection
σ_{RMS} / Abs. Units	$2.4 \cdot 10^{-5}$	$8.4 \cdot 10^{-5}$	$3.0 \cdot 10^{-5}$
Benzonitrile LOD / μM	80	50	18

Our theoretical LODs determined from the baseline spectra (water minus water) and sensitivities match well with experimental spectra of benzonitrile at low concentrations (Fig S5 and Main text Fig. 1), where the nitrile stretch first becomes clearly visible to the FTIR between 50 and 100 μM , and to the QCL between 10 and 25 μM . The longer pathlength afforded by the higher photon flux is the primary advantage. As demonstrated above from the single channel results, the QCL light source inherently has more fluctuations than the FTIR global source, an aspect observed by Lendl and coworkers in previous research,^{1,2} although our use of a MIRcat laser equipped with ZeroPoint beam pointing control may provide some noise reduction compared to Lendl and coworkers' use of the Hedgehog laser. The ZeroPoint technology in the MIRcat provides active stabilization compared to the passive stabilization in the Hedgehog module and has lower thermal drift compared to the Hedgehog module.⁴ In the absence of balanced detection, the Lendl Lab reported nearly a 3-fold higher RMS noise for their Hedgehog (2nd Gen) QCL source compared to FTIR systems under comparable experimental conditions, with measured values of 6.2×10^{-5} (QCL single channel measurement, 53-s integration, 31- μm path length) versus 2.3×10^{-5} (FTIR, 45-s integration, 8- μm path length) in the 1700–1500 cm^{-1} spectral range. These results align closely with the noise levels reported in Table 1 of this work, despite differences in experimental parameters such as the higher number of scans (>100), shorter path lengths (8–31 μm) and a different spectral window employed in their study. The similarity in noise performance underscores the challenges inherent to QCL-based systems, where intensity fluctuations and thermal drift can dominate measurement uncertainty.

Despite the higher noise inherent to the QCL source, balanced detection decreases RMS noise into the realm of the FTIR, where the 5x longer pathlength of the QCL affords an approximately 5-fold advantage in terms of LOD. We note that in all cases described herein, 16 scans were used on the QCL spectrometer for comparison to the FTIR and measurement of the nitrile-embedded PYP within bacterial cells. This number was chosen as the optimal balance between noise-reduction and efficiency when making measurements (Fig. S7). Although Akhgar et al. demonstrated that 1000 scans could decrease noise by about 5x further beyond what we report here,¹ we aimed instead to minimize heating to our live cell samples by taking fewer scans. We found that 16 scans were sufficient to observe the nitrile peak within the cellular environments in each case.

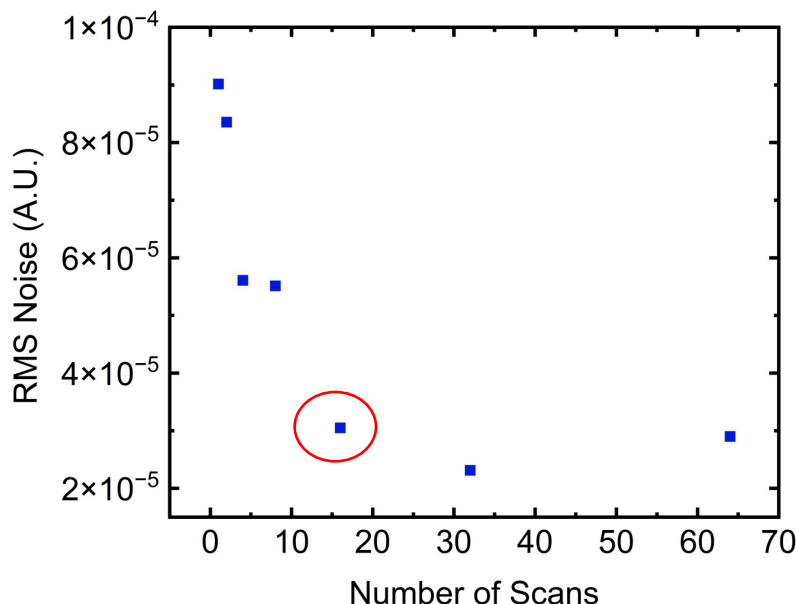


Figure S7. RMS noise for a blank (balanced) absorption spectrum through 250- μm pathlength water generally decreased with increasing number of scans based on square root law of noise averaging. We chose 16 scans as the optimal number of scans for the sake of achieving sufficient sensitivity while maintaining efficiency and minimal perturbation of the later cell-based samples.

In addition to the fewer scans, we also originally sought to reduce heating effects within the long-pathlength liquid sample cell by adding an extra 2-second dark time between scans, in addition to the typical 1.7-second time for scanning and transfer of the data capture buffer to the computer. We found no significant benefit to the additional equilibration time (Fig. S8, Table S3), indicating that on the timescale of these measurements, heating by the laser is not a major consideration. We note that further optimization of the setup should be possible to achieve a faster scan-time, such as via data streaming rather than use of a capture buffer or scanning over shorter spectral ranges (or fewer data points), in which case additional temperature controlling may be advantageous for the faster laser repetition rates. Currently, with a 1.7-second combined time for scanning over 300 cm^{-1} and data transfer, this method proves roughly one-half as fast as a scan on a Bruker FTIR (0.8 seconds)—though still with markedly better signal-to-noise, meaning the QCL spectrometer still provides an overall benefit.

We also evaluated potential local heating and sample damage caused by QCL illumination at maximum flux settings (Fig. S8). A long-wavelength infrared (LWIR) thermal imaging camera (Pembroke Instruments TM-60D) was used to monitor the sample-cell holder temperature in real time; the laser beam spot was clearly visible with the laser on and not with laser off. Temporal temperature profiles were collected at the beam position with either water or bacterial suspension present. In both cases, measurements showed no temperature change until the laser was activated. Under continuous laser illumination (1 MHz and 30% duty cycle) for 100 s, with temperature sampling every 1 s, the temperature increased from $\sim 20^\circ\text{C}$ to $\sim 21^\circ\text{C}$ over the first 20 s, plateaued, and then gradually returned toward baseline (ambient temperature of $\sim 20^\circ\text{C}$) once the laser was

switched off. In contrast, under the pulsed-illumination conditions used throughout this study with the 2s dark interval—no measurable heating was observed; the temperature remained constant for the full duration of the measurement.

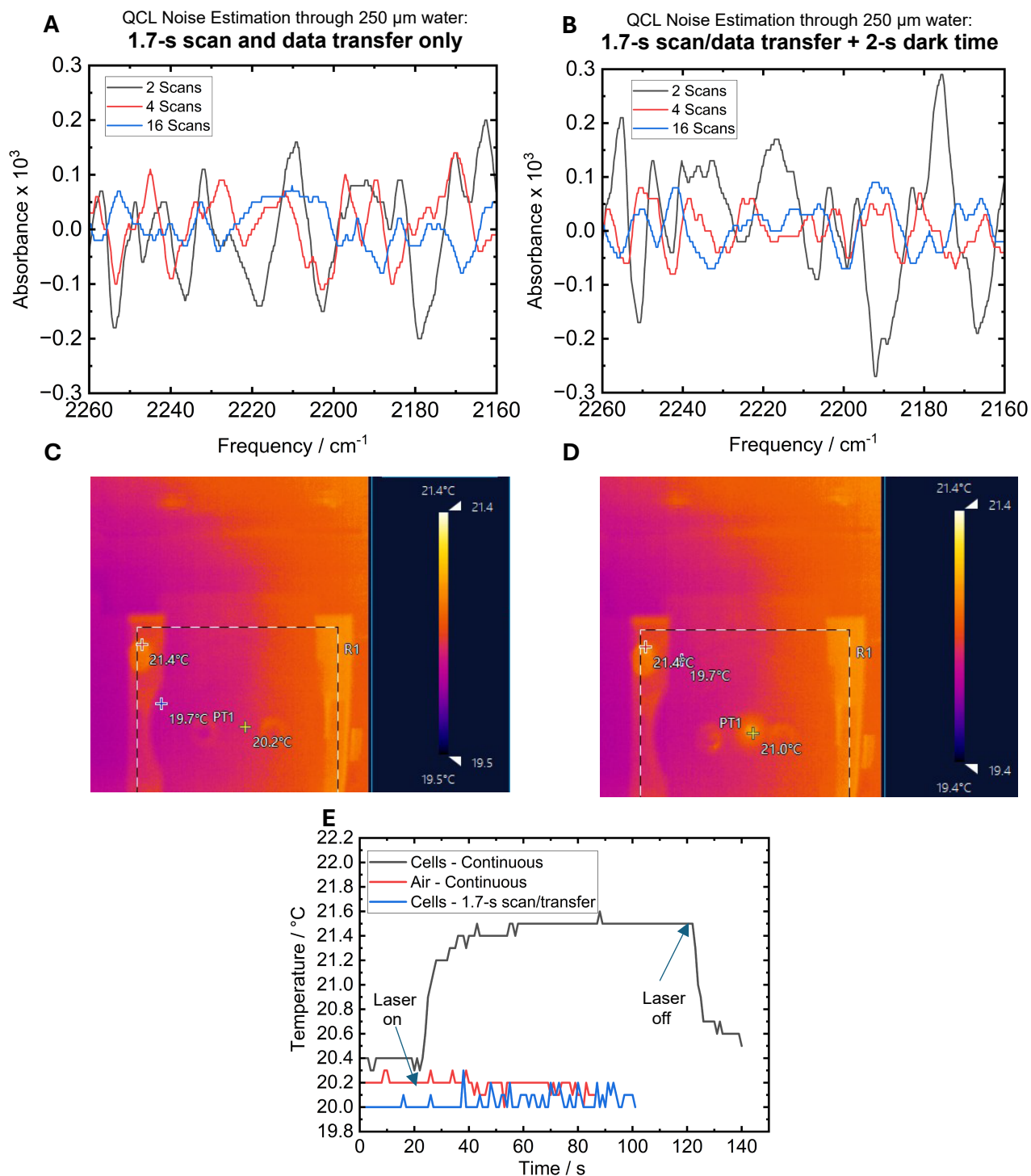


Figure S8. Verification that laser heating effects are negligible in terms of noise and cell temperature. (A-B) Baseline IR spectra recorded by the QCL spectrometer using 2, 4, or 16 scans, using (A) either the maximum repetition rate for the data buffer transfer (1.7 seconds), or (B) an additional 2 seconds (~3.7 seconds total) per scan for additional temperature equilibration. The same procedure for data processing and RMS noise calculation was followed as in Fig. S6, showing no major advantage from the additional equilibration time (and thus suggesting no heating artifacts using the 1.7-s laser repetition rate). (C-D) To further determine any possible impact of heating on the bacterial cells themselves, a LWIR thermal camera was used to measure temperature of a single pixel inside the laser spot on the sample. Thermal images are shown with (C) laser off, and (D) laser on, using a single pixel within the heated spot as a measure of temperature. (E) With continuous pulsing of the laser at 2100 cm^{-1} and a 1 MHz repetition rate, the spot increases in temperature by about 1°C in a cell suspension similar to experimental conditions, which should have minimal implication on cell survivability (far below the 37°C optimal growth threshold). Furthermore, when the 1.7-s scan and transfer time used in the experiment is incorporated, sample heating is negligible, similar to the negative control where the FTIR cell only contained air.

Table S3. RMS noise of baseline between 2260 and 2160 cm^{-1} , with varying numbers of scans and a variable dark time between scans for re-equilibration. Note that the variation between the two methods of detection does not appear significant, including compared to the previous reported baseline samples performed with 16 scans and the extra 2-s dark time (Table S1).

Number of Scans	1.7-s Scan/Data Transfer, σ_{RMS} / Abs. Units	1.7-s Scan/Data Transfer + 2-s Dark-Time σ_{RMS} / Abs. Units
2	$4.1 \cdot 10^{-5}$	$5.3 \cdot 10^{-5}$
4	$2.6 \cdot 10^{-5}$	$1.7 \cdot 10^{-5}$
16	$1.7 \cdot 10^{-5}$	$1.9 \cdot 10^{-5}$

As an additional point of comparison for our Bruker Vertex 70 FTIR, we also compared the detection of benzonitrile on an independent FTIR instrument, a ThermoFisher Nicolet iS50 with a liquid nitrogen cooled MCT detector. The sensitivity of this particular instrument, particularly in the benzonitrile region, was poorer than the Bruker Vertex 70 used for other comparative measurements reported in the study. Experimentally, the benzonitrile peak did not begin to be clearly visible in the spectrum until a concentration of $\sim 500\text{ }\mu\text{M}$, giving an experimental LOD several times higher than the $80\text{ }\mu\text{M}$ LOD on our Bruker FTIR (Fig. S9). We also calculated the RMS error for this instrument in the $2220\text{--}2245\text{ cm}^{-1}$ region of the benzonitrile peak to be $4.4 \cdot 10^{-4}$ based on the baseline noise, an order of magnitude poorer than the Bruker FTIR. Additionally, the Nicolet iS50 requires ~ 1.6 seconds for each scan, twice as slow as the Bruker instrument. Hence, we conclude that the Bruker instrument's performance is a better comparison point for the QCL spectrometer performance. Note that this is an instrument-specific measurement for research benchmark purposes and not an endorsement of Bruker instruments.

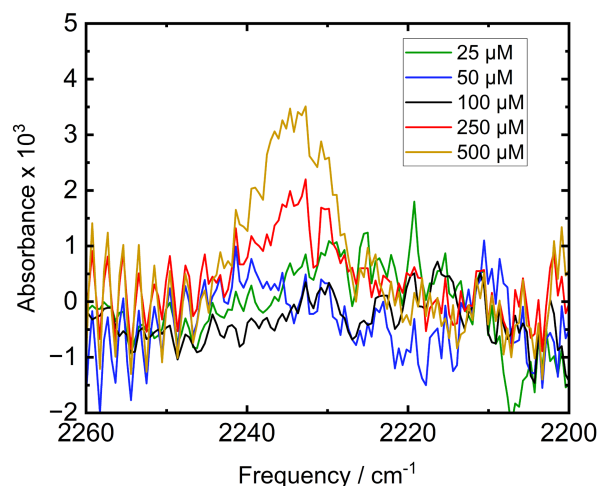


Figure S9. Benzonitrile detection on a ThermoFisher Nicolet iS50 FTIR instrument. Compared to the Bruker Vertex 70 FTIR, the ThermoFisher instrument requires a higher concentration of benzonitrile for detection and has a higher baseline noise, indicating that the Bruker Vertex 70 FTIR is a better comparison point for the QCL spectrometer performance.

g. Extending to alkyne $\text{C}\equiv\text{C}$, $\text{C}-\text{D}$, and $\text{N}=\text{N}^+=\text{N}^-$ stretches: In addition to nitriles, several other vibrational stretches that can function as readouts of molecular interactions and environments and that have been introduced into proteins by amber suppression are also reasonable candidates for detection using this QCL spectrometer at low concentrations. The $2000\text{--}2300\text{ cm}^{-1}$ region is home to other weak vibrational oscillators including carbon-deuterium ($\text{C}-\text{D}$) alkyne ($\text{C}\equiv\text{C}$) and stronger azide ($\text{N}=\text{N}^+=\text{N}^-$) bonds that have proven useful probes in a variety of biological contexts.^{5,6} We demonstrate how our instrument can sensitively detect sample vibrational probes at other neighboring frequencies including an alkyne (1-hexyne) at 2111 cm^{-1} , a deuterated (aldehyde-H) form of N-cyclohexylformamide (CXF-D) at 2182 cm^{-1} used as a probe of electric field directionality in enzymes by Zheng et al.,⁷ and an azide-labeled phenylalanine. As can be seen in Fig. S10, the QCL measurement of 10 mM 1-hexyne in water using a $250\text{ }\mu\text{m}$ pathlength gives a significantly better spectrum compared to the same number of scans (16) on the FTIR using a $50\text{ }\mu\text{m}$ pathlength. We moreover show how the QCL can give good sensitivity for detection of strongly absorbing azides such as 4-azidophenylalanine (Fig. S11). Finally, we show how the QCL again gives better sensitivity for detection of the CXF-D, with a clearly visible peak at 1 mM where the same number of scans on the FTIR at the same concentration and $50\text{ }\mu\text{m}$ pathlength does not yield a similarly clear peak (Fig. S12). Comparisons of the peak fitting parameters using either method are provided in Table S4.

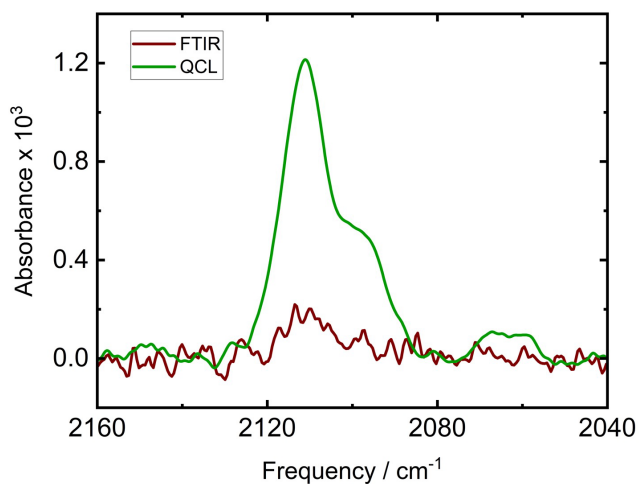


Figure S10. Spectra of 10 mM 1-Hexyne in water after 16 scans on either the Bruker Vertex 70 FTIR using a 50 μm pathlength, or on the QCL using a 250 μm pathlength, show better sensitivity of the QCL in this frequency range.

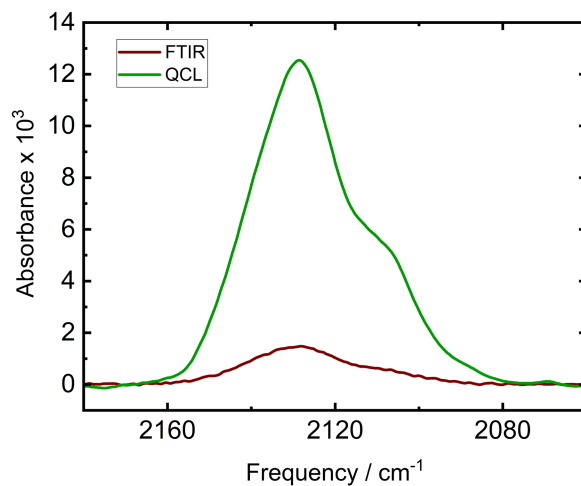


Figure S11. Spectra of 1 mM 4-azidophenylalanine in water after 16 scans on either the Bruker Vertex 70 FTIR (here using 50 and 25 μm spacers to reduce etalon fringes), or on the QCL using a 250 μm pathlength, show better sensitivity of the QCL in this frequency range.

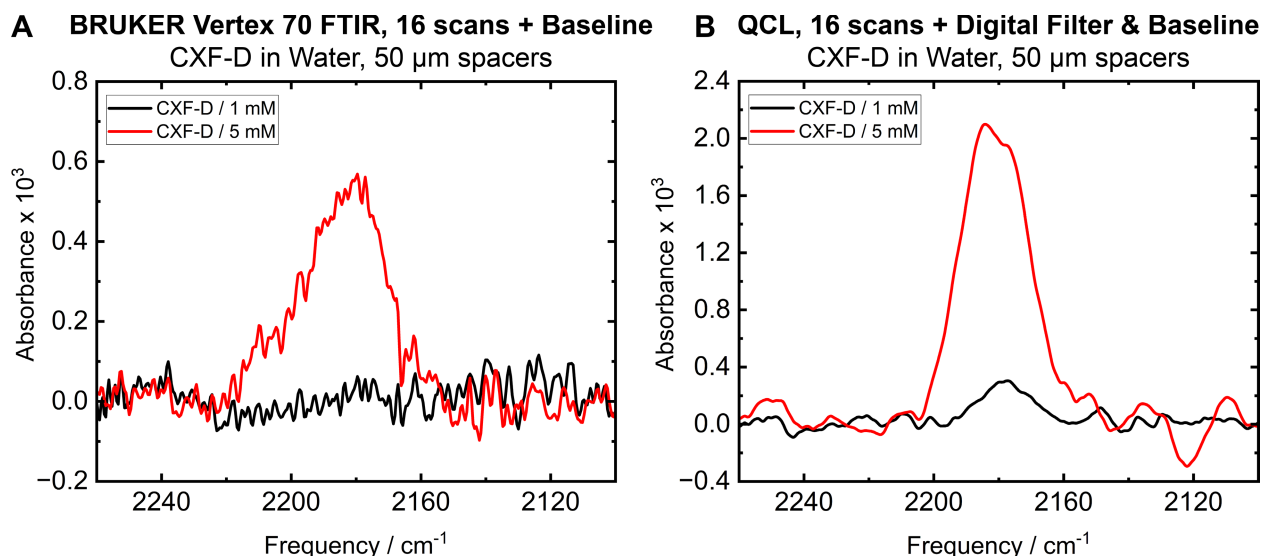


Figure S12. Spectra of CXF-D at concentrations of 1 mM and 5 mM in water to compare sensitivity of (A) the Bruker Vertex 70 FTIR at a 50 μm pathlength, and (B) the QCL spectrometer at a 250 μm , with 16 scans used on both instruments.

Table S4. Peak fitting parameters from Bruker Vertex 70 FTIR and QCL (balanced detection) measurements. Data in this table are reported for analyte concentrations where peaks were clearly identifiable in both techniques, with 16 scans performed in each case to ensure comparability. Peaks are fit in OPUS using Gaussian + Lorentzian functions of variable weights (Voigt profiles).

Analyte	Bond	Concentration	FTIR			QCL (Balanced)		
			Peak / cm^{-1}	Width / cm^{-1}	Intensity / OD	Peak / cm^{-1}	Width / cm^{-1}	Intensity / OD
Benzonitrile	$\text{C}\equiv\text{N}$	100 μM	2237.7	16.8	$2.1 \cdot 10^{-4}$	2235.8	8.6	$4.4 \cdot 10^{-4}$
1-Hexyne	$\text{C}\equiv\text{C}$	10 mM	2111.1	11.2	$1.8 \cdot 10^{-4}$	2111.4	12.1	$1.2 \cdot 10^{-3}$
			2097.6	8.1	$7.6 \cdot 10^{-5}$	2097.2	12.2	$4.4 \cdot 10^{-4}$
CXF-D	$\text{C}-\text{D}$	5 mM	2183.6	29.9	$5.3 \cdot 10^{-4}$	2181.3	24.1	$2.1 \cdot 10^{-3}$
4-Azido-phenylalnine	$\text{N}=\text{N}=\text{N}$	1 mM*	2129.6	26.3	$1.5 \cdot 10^{-3}$	2129.7	25.9	$1.2 \cdot 10^{-2}$
			2105.3	17.6	$3.3 \cdot 10^{-4}$	2105.8	17.1	$3.8 \cdot 10^{-3}$

*FTIR measurements for 4-Azidophenylalnine used 50 μm and 25 μm spacers to reduce fringes, all others used 50 μm spacers for FTIR and 250 μm for QCL. Baseline subtraction at lower SNR values can lead to inaccuracies in calculating the peak width.

Section S2: Cell growth and protein purification

a. Sequence of PYP: N-terminal 6X-His-tagged PYP with an enterokinase cleavage site, from a previously reported pBAD construct,³ was recombinantly expressed in *E. coli* both as WT and in various nitrile-incorporated variants using Amber suppression. The highlighted positions were individually replaced with the TAG stop codon for Amber suppression machinery. The positions code for Phe residues at F28, F62, F92 and F96.

```
ATGGGGGGTTCTCATCATCATCATCATGGTATGGCTAGCATGACTGGTGGACAGC
AAATGGGTTCGGGATCTGTACGACGATGACGATAAGGATCCGAGACTCGAGGATGACG
ATGACAAAATGGAACACGTAGCCTTCGGTAGCGAGGACATCGAGAACACCCTCGCCA
AGATGGACGACGGCCAGCTCGACGGCCTGGCCTTCGGCGCCATCCAGCTCGACGGC
GACGGCAACATCCTTCAGTACAACGCCGCGGAGGGCGACATCACCGGCCGCGACCC
GAAGCAGGTCATCGGCAAGAACTTCTTCAAGGACGTGGCCCCGTGCACTGACAGCC
CGGAGTTCTACGGCAAGTTCAAGGAAGGGGTGGCCTCGGGCAACCTGAACACGATG
TTCGAGTACACCTTCGATTACCAAATGACGCCACGAAGGTGAAGGTGCACATGAAG
AAGGCCCTCTCCGGCGACAGCTACTGGGTCTTCGTCAAGCGCGTCTAA
```

b. Mutagenesis and cloning: MAX Efficiency™ DH10B competent cells (ThermoFisher) underwent a double transformation. This involved introducing two plasmids: pEVOLpylT-N346AC348A and one of several pBAD-PYP variants with/without the engineered stop codons for nitrile incorporation (Addgene). The PYP plasmids, including those with the TAG mutations, were created using an Agilent QuikChange Lightning Site-Directed Mutagenesis Kit, starting from the original pBAD-PYP plasmid. A separate mutation, C69A (for chromophore knockout), was generated using a Q5 Site-Directed Mutagenesis Kit from New England Biolabs. The primers used for making the C69A mutations were the following: C69A Forwards: GCGACTGACAGCCCGGAGTTC and C69A Reverse: CGGGGCCACGTCCTTG. Annealing temperature for this reaction was set to 68°C.

c. Cell growth without Amber suppression: Plasmids for both wild-type PYP and the PYP-C69A mutant (without Amber suppression machinery) were transformed into competent BL21 (DE3) *E. coli*. Transformed cells were selected by overnight growth on agar plates containing 100 µg/mL ampicillin at 37°C. Single colonies were used to initiate 5 mL liquid cultures in LB with 100 µg/mL ampicillin, which were grown overnight at 37°C with shaking at 250 rpm. These overnight cultures were then used to inoculate 1L cultures in either Luria Bertani Broth or Terrific Broth (ThermoFisher), supplemented with 100 µg/mL ampicillin, in pre-warmed baffles flasks (Nalgene). These larger cultures were grown at 37°C with shaking at 200 rpm until they reached an OD600 of 0.6 (typically 2-4 hours). Protein expression was then induced by adding 143 mg of IPTG (0.6 mmol, ThermoFisher), and the cultures were incubated overnight at 20°C with shaking at 200 rpm.

d. Cell growth using Amber Suppression: The general protocol for sample preparation involved the incorporation of noncanonical amino acids, ortho-cyanophenylalanine (oCNF, ChemImpex, Ann Arbor, MI, USA), into overexpressed target proteins within *E. coli* utilizing an Amber suppression system. Doubly transformed cells, harboring plasmids for both the target protein and the orthogonal translation machinery, were initially cultured following established literature procedures.³ For each liter of subsequent protein expression, a starter culture was prepared by inoculating a 14 mL round-bottom culture tube containing 5 mL of Luria-Bertani (LB) growth medium (Millipore) with the transformed cells. These starter cultures were grown for 12 to 16 hours, or until they reached saturation, with the LB medium supplemented with specific antibiotics for plasmid selection: 100 µg/mL ampicillin for the pBAD plasmid and 34 µg/mL chloramphenicol for the pEVOL plasmid. Following sufficient growth, the entire starter culture was transferred into 1 L of auto-induction medium from which phenylalanine source had been omitted. This larger volume of auto-induction medium was also supplemented with the same antibiotic concentrations: 100 µg/mL ampicillin (Sigma-Aldrich) and 34 µg/mL chloramphenicol (Sigma-Aldrich) and pre-warmed to 37°C for one hour prior to inoculation. It is important to note that the auto-induction medium contained arabinose (Sigma-Aldrich), which was added after the medium had been sterilized via autoclaving to prevent degradation of the arabinose necessary for the pBad plasmid expression. The inoculated cultures were then shaken at 250 RPM at 37°C for 30 minutes, after which 10 mL of a 100 mM solution of oCNF was added. 100 mM oCNF solution was prepared approximately 1 to 2 hours before its addition to the cell culture medium using sonication and heating as oCNF is only partially soluble in water at these concentrations. Protein expression was then auto induced for approximately 19 hours at 250 RPM and 37°C.

e. Preparation of cells for QCL-based spectroscopy: After overnight expression of the cultures, a portion of ~100 mL liquid culture was introduced to freshly prepared activated pCA (acyl imidazolid form, see section S5), scraping the yellow oil product from the flask walls and mixing with the liquid cell culture, before stirring overnight at 4°C.³ The control (non-exposed) cells were similarly incubated overnight but without pCA. The following morning, both groups of cells were washed 3x (gentle centrifugation at 2000 RPM for 10-15 minutes, followed by resuspension) with minimal media (Sigma Aldrich M9 Minimal Salts; Cat. No. M6030) to remove excess oCNF and/or pCA. In the pCA-incubated batch, a noticeable yellow coloration of the loose cell pellets collected in each washing step provided a visual indication of the successful incorporation of the chromophore into the *apo*-PYP protein within the cells. Notably, excess activated pCA, which did not react with *apo*-PYP, underwent hydrolysis in water present in the medium, regenerating insoluble pCA. This re-formed pCA presented itself as a distinct white precipitate layer, often observed above or as a separate layer from the denser yellow cell pellet after the initial centrifugation (confirmed using IR, NMR and UV-vis). To isolate the cells containing the bound chromophore, the layer of the pellet containing the yellow cells was carefully scraped off using a spatula, thereby separating it from the white precipitate, for subsequent washing with minimal media in parallel with the non-pCA-incorporated cells.

Following the washing steps, the loose cell pellets were resuspended in a minimal amount of M9 minimal media. Aliquots were diluted 1000x for measurement of OD₆₀₀ by a PerkinElmer Lambda-365 UV-visible spectrophotometer (Waltham, MA), a common metric to assess the concentration of a bacterial suspension. The final OD₆₀₀ of the washed bacterial suspensions when diluted 1000x were adjusted to the range of OD₆₀₀=0.3-0.4 for measurements. Approximately 75 μ L of the OD₆₀₀-adjusted suspension was loaded into the sample demountable cell of the QCL spectrometer using a syringe. Cellular suspensions expressing the wild-type protein were prepared and washed in a similar manner for reference measurements. Using the previously described protocol for the QCL measurements, IR spectra were collected for the F28oCNF, F62oCNF, F92oCNF, and F96oCNF variant-expressing cells that had been incubated with or without the activated pCA, taken with reference to wild-type *apo*-PYP-expressing cells (lacking the nitrile reporter) serving as a crucial reference blank to identify the nitrile vibrational bands reporting on pCA binding.

f. Acquisition of nitrile IR spectra in live cells: Averaged spectra from 16 scans for each of the different nitrile-containing mutants were collected to assess the microenvironment around a ligand of interest. In all cases, the laser pulse rate was kept at 1 MHz with 300 ns duration (30% duty cycle), with scanning from 2360 cm^{-1} to 1970 cm^{-1} at a scan rate of 2000 cm^{-1}/s . A 2 second dark time was employed between consecutive scans to allow for data transfer and temperature re-equilibration (minimal over the time duration of each scan, See Figure S8). Data was collected in 16-kb data packets on the SRS865A lock-in amplifier at a rate of 9765.6 Hz (equal to the maximum collection rate divided by 16). The time constant of the SRS865A lock-in amplifier was kept as 300 μ s, the input range 1 V, and the sensitivity 500 mV. Data packets were sent via ethernet streaming to the host computer for processing. Data processing was based on the methods reported by Lendl and coworkers,² and included optional phase correction in cases where the signal was significantly distributed between *x*- and *y*- components (typically not necessary), plus removal of anomalous spectra using similarity indices (rarely employed in practice), a 3rd-order Savitsky-Golay filter on 41 data points at a time, averaging of successive scans, and a sharp (order-400) long-pass Fourier Filter with a Blackman-Harris window and 1000 Hz cutoff over the spectral window.

g. Protein purification:

- ***Apo*-PYP for vibrational spectroscopy:** To purify *apo*-PYP for measurements, the following protocol was employed: Cell pellets obtained from centrifugation were resuspended in a buffered 20 mM Tris-HCl, 10 mM NaCl (Buffer A; pH 8.0) using 2 mL of buffer per 1 mL of cell pellet. Cells were lysed by homogenization using a C3 Emulsiflex, and insoluble cell debris was removed by high-speed centrifugation at 15,000 g for 120 minutes. The resulting lysate was filtered through a 0.22 μ m Sterile PVDF filter and then loaded onto a Ni-NTA column for 6x-His tag affinity purification. The column was equilibrated with 5 column volumes (CVs) of Buffer A (1 CV = 5 mL resin) before sample loading. The protein was allowed to bind to the column for 1 hour under gentle rocking. Subsequently, the column was washed with 5 CVs each

of Buffer A containing 10 mM and 20 mM imidazole to remove non-specifically bound entities. The 6X-His-tagged PYP was then eluted with Buffer A containing 250 mM imidazole. To remove excess imidazole, the eluted protein was subjected to buffer exchange using spin filtration with Amicon Ultra 15 centrifugal filter units (3 kDa MWCO) against Buffer A. The protein concentration was determined by absorbance at 280 nm using the Expasy ProtParam-calculated extinction coefficient for the linker+his-tagged *apo*-PYP (MW=18501.5 g/mol; $\epsilon_{280}=14400 \text{ M}^{-1} \text{ cm}^{-1}$), and the purity assessed using LC-MS with a C8 reverse-phase column and single quadrupole mass spectrometer at the SUMS open access facility at Stanford University.

- ***Apo*-PYP for crystallography:** For x-ray crystallography studies, further purification steps were performed to remove the His-tag. The Ni-NTA purified *apo*-PYP was subjected to anion exchange chromatography (Akta FPLC, GE Healthcare) using an FPLC system with three 5 mL Q HP columns (Cytiva) connected in series, pre-equilibrated with Buffer A. Protein elution was achieved by applying a linear gradient of 0–40% Buffer B (1 M NaCl, 20 mM Tris-HCl, pH 8.0) over 180 mL at a flow rate of 5 mL/min. Fractions eluting between 25% and 32% Buffer B were pooled. These pooled fractions were concentrated and buffer-exchanged against Buffer A using 15 mL 3 kDa MWCO Centrifugal Filter Units (Millipore) to remove excess NaCl. The protein was then concentrated to ~10 mg/mL, and the N-terminal 6X-Histag was cleaved by adding 1 Unit of enterokinase (New England Biolabs, Catalog #P8070S) per 0.025 mg of protein and 2 μL of 1 M CaCl_2 per 1 mL of protein solution. The cleavage reaction was performed under gentle shaking at 25 RPM 4 °C for 16 hours. Following tag cleavage, the mixture was re-purified using anion exchange chromatography as described after the Ni-NTA step. Untagged PYP fractions eluted between 15% and 19% Buffer B. Fractions exhibiting maximal A_{280} were pooled. As before, NaCl was removed by buffer exchange against Buffer A using a 3 kDa Amicon centrifugal filter unit. For samples intended for protein IR spectroscopy, *apo*-PYP was further concentrated using a 0.5 mL 10 kDa Amicon centrifugal filter unit by reducing the volume to approximately 18 μL and then centrifuging for an additional hour at 14,000 g at 4 °C. The concentrated *apo*-PYP was carefully collected, gently pipetted to ensure homogeneity, and transferred to a 0.6 mL microcentrifuge tube prior to crystallography discussed in Section S3.
- ***Holo*-PYP for vibrational spectroscopy:** To purify *holo*-PYP, two distinct approaches were employed for different confirmatory measurements within the study. In the first method, adapted from the original protocol by Weaver et al.,³ the crude *apo*-PYP cell lysate obtained from one liter of expressed cells was directly reacted with the activated pCA chromophore immediately after the evaporation of THF. The activated chromophore, which adhered to the sides of the round-bottom flask in the form of a dense yellow oil, was scraped using a metal spatula into the cell lysate. This mixture was gently stirred (< 200 RPM) overnight to avoid introducing foam. The following morning, insoluble reaction precipitate, primarily a white solid form of excess pCA was removed by centrifugation at 15,000 g for 120 minutes. The resulting lysate was then

filtered and purified using the Ni-NTA column as described previously for *apo*-PYP. In the second protocol for *holo*-PYP purification, cells where the activated pCA was intended for incorporation were first lysed, and then the purification procedure was repeated as described for *apo*-PYP using the Ni-NTA column.

h. Estimation of sensitivity of in-cell IR measurements: In their work on detection of proteins via the Amide-I and Amide-II bands, Akhgar et al.¹ report an effective limit of detection (LOD) of 0.0025 mg/mL for three proteins: hemoglobin, gamma-globulin, and concanavalin A. We convert their protein LODs to molar concentrations of amides based on their numbers and this is summarized in Table S5, which we use to compare to our benzonitrile LODs.

Table S5. Estimation of molar concentrations of amides from protein concentration LODs of Akhgar et al. We assume the extinction coefficient of amides $\sim 1000 \text{ M}^{-1}\text{cm}^{-1}$.

Protein	~MW (kDa)	Residues	Protein Concentration based on 0.0025 mg ml ⁻¹ (μM)	Approx. Amide Concentration [Protein Conc x Residues] (μM)
Hemoglobin	65	578	0.038	22
IgG	150	1020	0.017	17
Concanavalin A	104	948	0.024	23

As shown above, at 0.0025 mg/mL protein, Akhgar et al. could detect amides at concentrations around $\sim 20 \text{ μM}$. Our setup utilizes a 250 μm Teflon spacer, roughly ten times longer in path length than theirs (26 μm). While this longer path length is beneficial to us, nitriles have an oscillator strength approximately five times lower than that of amides (~ 200 vs. $1000 \text{ M}^{-1} \text{ cm}^{-1}$).³ Consequently, if the numbers reported in the setup of Akhgar et al. were completely translatable to our case of nitriles, we would anticipate a benzonitrile LOD half that of carbonyls at the short pathlength, or $\sim 10 \text{ μM}$. This value is similar in magnitude to the 18 μM LOD for benzonitrile we report here. There are significant differences between the two apparatuses, most significantly in the type of laser module used over the different wavelength ranges (our MIRcat being more powerful and stable compared to the Hedgehog), and the significantly lower number of scans we used in the current work (16 scans here versus 1000 scans reported by Akhgar et al., which would give them an approximately 8-fold advantage for reducing random noise due to scan replicates). While a more exact comparison is not possible, a theoretical LOD at a similar magnitude to that anticipated by the apparatus of Akhgar et al. helps confirm the theoretical limitations of balanced QCL-based spectrometers across a variety of chemical species and wavelengths.

i. Estimation of protein concentration in cells: From the purification of 1 L cultures, we obtained 10 mg of *apo*-F92oCNF and 16 mg of *apo*-F96oCNF (both 6X-His-tagged) PYP. Assuming similar growth conditions and expression systems, OD₆₀₀ measurements of the liquid cultures taken

after 19 hours across three biological replicates (diluted 50x, with $OD_{600} = 1$ corresponding to 8×10^8 *E. coli* cells/mL) yielded average cell densities of $3.8 \pm 1.5 \times 10^9$ cells/mL for cells expressing *apo*-F92oCNF and $6.6 \pm 3.0 \times 10^9$ cells/mL for *apo*-F96oCNF. The protein per cell for both *apo*-F92oCNF and *apo*-F96oCNF was calculated based on the protein yield and average cell count, with uncertainties in the cell counts considered. For growths expressing *apo*-F92oCNF, the protein per cell was found to be approximately $2.6 \times 10^{-15} \pm 1.0 \times 10^{-15}$ g/cell, which corresponds to $1.4 \times 10^{-19} \pm 5.7 \times 10^{-20}$ mol/cell of the tagged *apo*-PYP. For *apo*-F96oCNF, the protein per cell was $2.4 \times 10^{-15} \pm 1.1 \times 10^{-15}$ g/cell, corresponding to $1.4 \times 10^{-19} \pm 6.2 \times 10^{-20}$ mol/cell. The relatively small difference in protein per cell, with overlapping error margins, suggests that the expression levels for *apo*-F92oCNF and *apo*-F96oCNF are similar in protein production per cell, corresponding to about 84000 protein molecules per cell. If we then assume each cell is $\sim 1 \mu\text{m}^3$, this results in a concentration of nitrile per cell at 140 μM , which is within the limit of detection of our QCL-spectrometer for individual cells. Recognizing that the cell suspension will be more dilute in terms of nitrile-labeled *apo*-PYP than an individual cell, we consider the concentration of cells within the final cell suspension based on a 1000x dilution giving $OD_{600}=0.35$. Again using $OD_{600} = 1$ corresponding to 8×10^8 *E. coli* cells/mL, this corresponds to 2.8×10^{11} cells/mL in the cell suspension used for the QCL measurements. Using 1.4×10^{-19} mol protein/cell, the final nitrile concentration visible to the QCL detection assay was thus calculated as 39 μM in each case—above the 18 μM detection limit for the QCL spectrometer but still smaller than the FTIR LOD for benzonitrile using the same number of scans. While we note that the extinction coefficient of the oCNF in the organized protein environment will not be identical to benzonitrile in solution, these results indicate an ability and advantage of the QCL in detecting vibrational frequencies of nitrile-tagged *apo*- or *holo*-PYP in live cell suspensions.

j. Comparison to another protein system, GFP: To show that our method is generally applicable even with proteins with different expression levels and in a different chemical environment, we considered another protein system with embedded nitriles, to show similarly that IR spectra could be recorded for the nitrile-labeled protein in a different cellular environment. Green fluorescent protein (GFP) has been a topic of significant interest—particularly in terms of properties governing its photophysics, with roles of electrostatics and sterics in tuning properties such as absorption and emission wavelength, lifetime, and fluorescence quantum yield.^{8,9} Most FPs autocatalyze their chromophore maturation within a chromophore forming triad XYG. In previous work from our group, we successfully incorporated various non-canonical Y analogs into the GFP chromophore, including halogenated tyrosines, tyrosines with strongly electron-withdrawing NO_2 groups, electron-donating ortho-methoxy groups, and in this work with ortho-cyanophenylalanine (oCNF) giving a unique nitrile stretching frequency. The exact structural and photophysical implications of this observation are beyond the scope of the current work and will be pursued elsewhere. Oleginski et al. have recently reported the successful incorporation of para-cyanophenylalanine (pCNF) into the GFP chromophore.¹⁰

To create this construct, we introduced a stop codon (TAG) at position Y66 (pBAD-sfGFP-66TAG construct). MAX Efficiency™ DH10B Competent Cells (ThermoFisher) were co-transformed with the pEVOL-pylT-N346A/C348A plasmid (Addgene) and the pBAD plasmid for protein expression via Amber suppression. Defined media supplemented with ampicillin, chloramphenicol, arabinose, and oCNF (as a phenylalanine analog) were used for cell growth. These conditions mirrored those used to generate oCNF analogs of PYP. The cells were auto induced and grown for 18 hours at 37 °C, after which the temperature was reduced to 18 °C for 4-6 hours to facilitate chromophore maturation. Following expression, sample preparation and the measurement protocol followed similarly to the PYP experiments.

Section S3: X-ray crystallography

a. Comparison to previous *holo*-PYP crystal structures: X-ray crystallography was used to compare the global structure of *apo*-PYP to the chromophore-bound state. The chromophore-bound state has been previously crystallized as wild-type at ultra-high resolution (PDB: 1NWZ, resolution = 0.82 Å)¹¹ and as the four various nitrile-containing variants we use here (resolution < 1.2 Å): F28oCNF (PDB: 7SPX), F62oCNF (PDB: 7SPW), F92oCNF (PDB: 7SPV), and F96oCNF (PDB: 7SJJ).³ These four variants station the nitrile probe at various distances from the chromophore binding site and in both hydrogen-bonding and non-hydrogen-bonding environments, as shown in Fig. S13 below.

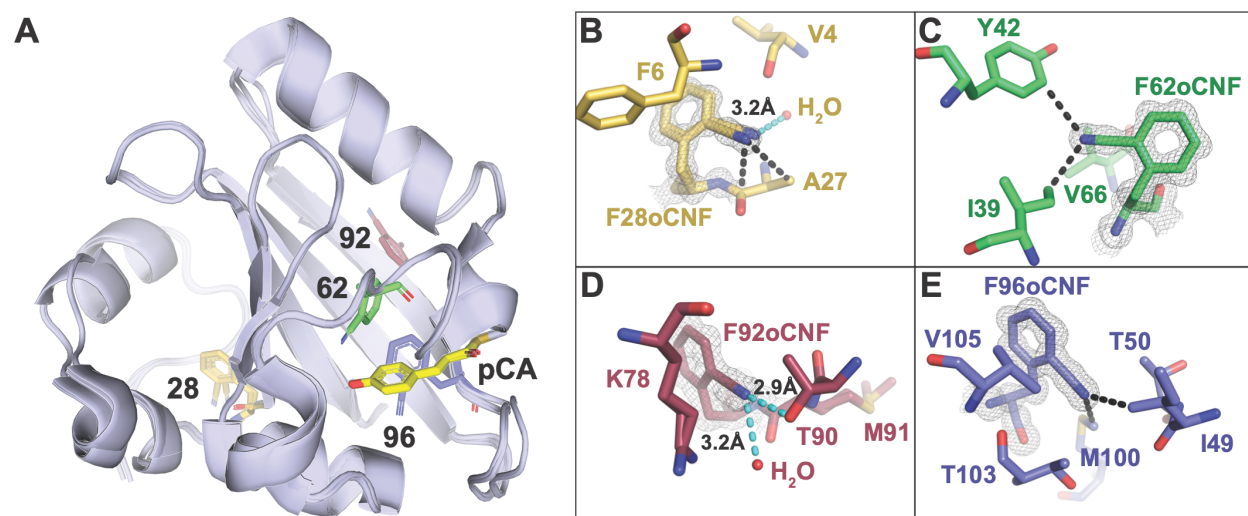


Figure S13. (A) Superimposed structures of the four nitrile-containing variants of *holo*-PYP, with the individual nitrile environments shown in detail: (B) F28oCNF interacts with one water molecule in its crystallographic pose, (C) F62oCNF lies in a hydrophobic environment, (D) F92oCNF demonstrates two hydrogen bonds to the nitrile in the crystallographic pose, from T90 and water, creating a strong electric field, and (E) F96oCNF demonstrates no hydrogen bonding partners in its crystallographic pose. Figure partially reproduced from Kirsh et al. (ACS Publications).¹²

b. Methods of *apo*-PYP Crystallography, Diffraction, and Analysis: In the present work, *apo*-PYP was crystallized using the hanging-drop method as previously described for *holo*-PYP.¹³ Freshly purified *apo*-PYP was buffer-exchanged into a buffer of pH 6.0, 20 mM potassium phosphate (Sigma-Aldrich) and concentrated to 20 mg/mL. Hanging drops were set manually using 24-well VDX plates with sealant (Hampton Research) and 22 mm diameter siliconized glass coverslips (Hampton Research). 1 μ L of protein was mixed with 1 μ L of mother liquor on each coverslip and closed over wells containing 500 μ L mother liquor. Trays were kept at room temperature and covered in aluminum foil to maintain consistency with *holo*-PYP crystallization conditions. Mother liquor solutions contained 1M NaCl and ammonium sulfate (Sigma-Aldrich) ranging from 1.8 – 2.6 M in 0.1 M increments. Crystal formation was observed within one to four weeks of initial setting.

All crystals of sufficient quality were placed on 0.1 – 0.4 mm Mounted CryoLoops (Hampton Research), briefly (1-2 seconds) incubated in polyfluoroether cryo-oil (Hampton Research), rapidly frozen in liquid nitrogen, and placed into a Stanford Synchrotron Radiation Lightsource (SSRL) cassette for X-ray diffraction data collection. X-ray diffraction was performed at SSRL (Menlo Park, CA) Beam Line 12-1 at 100K. Further detailed information on data collection strategy can be found in Romei et al.¹³ Data processing and scaling was performed using the HKL3000 software, providing the correct C2 space group.¹⁴ Molecular replacement was performed in Phenix-1.21.2 with phenix.phaser using the WT PYP (PDB Entry: 1NWZ) structure as the search model.¹⁵ Numerous rounds of model building and refinement were carried out in Coot³²¹⁶ and then with phenix.refine. Refinement continued until the R-free score could no longer be lowered. The X-ray collection parameters and refinement statistics are given below (Table S6).

Table S6. Crystallographic data summary

Entry name	<i>Apo</i> -PYP
PDB Entry	9O8V
Beamline	SSRL 12-2
Wavelength (Å)	0.97946
Detector Distance (mm)	150 mm
Space Group	C 1 2 1
Unit Cell Dimensions	104.63 Å, 50.72 Å, 71.25 Å 90.00°, 132.93°, 90.00°
Total Observations	310844 (12896)
Unique Reflections	66798 (3224)
Redundancy	4.7 (4.0)
Completeness (%)	98.5 (95.2)
Mean I/σ	42.2 (1.55)
Wilson B-factor (Å²)	17.7
R_{meas}	0.073 (0.811)

CC_{1/2}	1.00 (0.729)
Reflections used in Refinement	66760
Reflections used for R_{free}	3445
R_{work}	0.1580
R_{free}	0.1853
Number of non-H Atoms	Protein – 1770, Ligands – 0, Solvent – 177
Protein Residues	117
RMSD Bond Lengths (Å)	0.013
RMSD Bond Angles (°)	1.352
Ramachandran Favored (%)	97.35
Ramachandran Outliers (%)	2.65
Rotamer Outliers (%)	0.00
Clashscore	3.51
Average B-factor, all atoms (Å²)	25.0, Protein –24.0, Chloride – 35.9, Waters– 37.3

c. Structural comparisons between *apo*- and *holo*-PYP: The global structure of *apo*-PYP, as revealed by its crystal structure (PDB ID: 9O8V), shows minimal deviation from the *holo* form (PDB ID: 1NWZ), even in the absence of the pCA chromophore (or any incorporated nitriles). A comparative B-factor Z-score analysis between the *apo* and *holo* structures using the C_α carbons was performed to quantify local structural fluctuations. This analysis revealed generally similar patterns of disorder between the two forms. The B-factors here serve as a tool to assess differences in atomic fluctuations between two protein structures. B-factors (also called temperature factors) extracted from crystallographic data reflect atomic displacement or static disorder (a consequence of imperfect unit cell ordering in the crystals). Normalized B-factors were employed because different crystals often have different amounts of static disorder. To perform a Z-score analysis, the B-factors are normalized using the formula:

$$Z = \frac{B_i - \mu}{\sigma},$$

where B_i is the B-factor of residue i , μ the mean B-factor across the structure, and σ is the standard deviation. The resulting Z-scores help identify regions with statistically significant differences in disorder. This analysis is presented in Figure S14, where we note that, excluding the N-terminal and C-terminal regions of the *apo* form (9O8V) that are either unresolved or appear disordered, most of the other residues in both the *holo* or *apo* form have comparable Z-scores.

To further characterize structural similarity, we performed fixed-charge molecular dynamics simulations for 100 ns using two different starting configurations: (1) the *holo*-PYP structure with the chromophore coordinates removed (PDB ID: 1NWZ), and (2) the native *apo* structure (PDB ID: 9O8V). These results are presented later in Figures S24-S25; comparative analysis of these trajectories revealed no substantial conformational shifts or secondary structure rearrangements, regardless of the initial structure used. The minimal structural deviations observed throughout both

MD trajectories indicate that global conformational changes upon pCA binding are negligible. These techniques are more structurally informative than previous circular dichroism (CD) studies on *apo* and *holo* PYP,¹⁷ where the CD spectrum in the 250-300 nm region is dominated by contributions from the chromophore in the *holo* form, making structural comparison from the protein secondary structure difficult. These findings validate the computational strategy of modeling *apo*-PYP variants by using *holo*-PYP structures and simply removing the pCA chromophore coordinates. This approach was subsequently adopted in our AMOEBA polarizable force field simulations of *apo*-PYP variants, with detailed simulation protocols provided in Section S7.

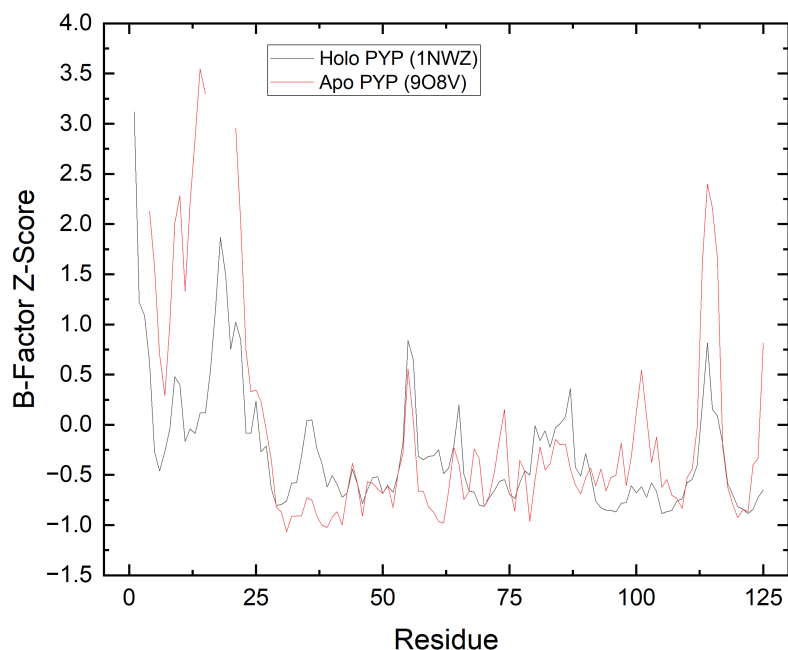


Figure S14. B-factor Z-score analysis comparing the crystal structures of *holo*-PYP (PDB ID: 1NWZ, chromophore incorporated, red) and *apo*-PYP (PDB ID: 9O8V, grey). The deviations are generally similar between the two forms. Notably, the N-terminal and C-terminal regions in the *apo* form are either missing or disordered. Overall, the crystal structures suggest minimal differences in local disorder, a conclusion further supported by the MD simulations presented in Section S7.

Section S4: FTIR spectroscopy

a. Methods and materials: For FTIR measurements involving small molecules (e.g., benzonitrile), the molecule was dissolved in solvent, where a stock solution of 10-100 mM was prepared and consequently diluted for LOD measurements. Both 1-hexyne and benzonitrile were sonicated for 10 minutes by an ultrasonication water bath to ensure complete dissolution.

For proteins (e.g. PYP), the protein was exchanged into a buffer containing 20 mM Tris HCl and 10 mM NaCl at pH 8.0. PYP was concentrated using 0.5 mL 10 kDa Amicon centrifugal filter units to approximately 18 μ L, at which point it was further concentrated for an additional hour at 14,000 g at 4°C. Samples were transferred to Eppendorf tubes, and the volume was recorded. The final concentration of the protein was adjusted to ~10mM using this protocol. In either case (small molecule or protein), ~15-20 μ L of the prepared sample was then injected into a demountable IR cell, which was assembled using two CaF₂ optical windows (19.05 mm diameter, 3 mm thickness, Lambda Research Optics, Inc.) separated by a pair of Teflon spacers (25 μ m and 50 μ m thick). FTIR spectra were recorded using a Bruker Vertex 70 spectrometer with a liquid nitrogen-cooled mercury cadmium telluride (MCT) detector (Teledyne Judson, Inc.), and the sample chamber was constantly purged with dry air to remove atmospheric water. The spectra were recorded in a wavenumber window spanning 4000–1000 cm^{-1} with 512 scans at 1 cm^{-1} spectral resolution. The obtained FTIR transmission spectra were converted to the corresponding absorption spectra using the neat solvent/background spectra taken under similar conditions as the reference. Data collection and processing were performed by the spectroscopy software OPUS 5.0.

b. FTIR experiments with cells for *apo*- and *holo*-F96oCNF PYP: We considered the ability of the FTIR to measure nitrile probe frequencies within the expressing cellular environment as we did with the QCL spectrometer. We estimated that the working concentration of nitrile-tagged proteins within the cell suspension was higher than the theoretical LOD (using that of benzonitrile, See Table S4), while it was less clear whether peaks could be observed at all using the FTIR with a shorter pathlength. As an aside, we observed that transmission in the FTIR through 250- μ m pathlength sample cells filled with water was too low and below the noise floor in this region. Using the 50- μ m pathlength, we were able to pick out the nitrile peak of both *apo*- and *holo*-F96oCNF PYP within the cell suspension from the background using 16 scans as the QCL; however, the spectra were significantly noisier and thus difficult to extract quantitative data in terms of frequency shifts between the *apo*-protein in cells and pCA-bound protein in cells (Fig. S15, Table S7). Qualitatively, a blue-shift does appear to occur upon pCA-binding, agreeing with the QCL-based results, though the lack of clarity in these spectra is a strong point in favor of using the QCL-afforded sensitivity for these measurements.

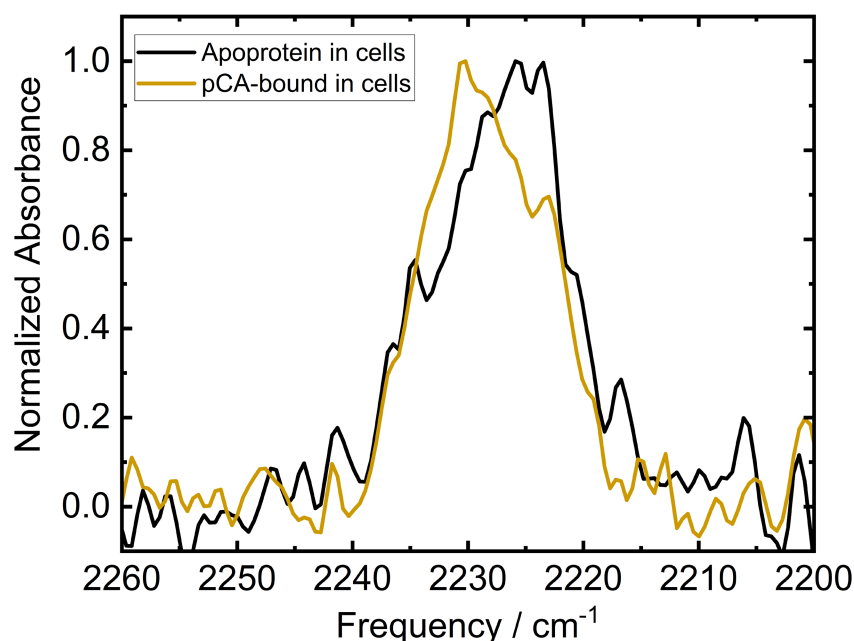


Figure S15. FTIR spectra of apo-F96oCNF PYP (black) and pCA-incubated F96oCNF PYP (gold) in expressing *E. coli* cells, showing the same qualitative blueshift upon pCA binding, albeit with significantly more noise.

Table S7. Spectral data fits, F96oCNF PYP in cells using FTIR. Peaks are fit in OPUS using Gaussian + Lorentzian functions of variable weights (Voigt profiles).

Protein	Peak / cm^{-1}	Width / cm^{-1}	Intensity/OD
F96oCNF Apo Cells	2226.9	14.2	$3.1 \cdot 10^{-4}$
F96oCNF pCA Cells	2228.5	12.7	$3.6 \cdot 10^{-4}$

c. Free oCNF spectrum (FTIR) and comparison to *in cellulo* Y66oCNF sfGFP (QCL Spectra):

Using a similar measurement protocol as for the PYP samples, we recorded 16 QCL-based IR scans of oCNF-GFP-expressing cells versus unlabeled superfolder-GFP-expressing cells as a reference. The scans revealed a clear peak at 2225.8 cm^{-1} , corresponding to the nitrile incorporated into the chromophore—distinct from free oCNF (Fig. S16)—thereby establishing another cellular system amenable to direct probing with this spectrometer.

To confirm that the observed peak arose from the nitrile-tagged GFP rather than free oCNF, we compared the vibrational spectra of the cell suspensions and purified protein with that of free oCNF (50 mM aqueous solution), collected using the Bruker Vertex 70 FTIR (black trace). The 2232.1 cm^{-1} peak of free oCNF is sufficiently distinct from the GFP spectra to confirm that the observed vibrational signature originates from the nitrile-labeled protein and not residual oCNF (which had been thoroughly washed from the cells). Further confirmation was obtained using purified proteins. The proteins were isolated from cells via Ni-affinity chromatography after 6X-His tag cleavage. Cells were pelleted at $15,000 \times g$ for 1 hour following homogenization. The supernatant was filtered using a $0.22 \mu\text{m}$ filter and loaded onto 5 mL of Ni-NTA resin. After

incubation, the proteins were eluted using 250 mM imidazole, as described for PYP in the preceding sections. The eluted proteins were then buffer exchanged using Amicon 10 kDa spin filters and further concentrated to 10 mM for FTIR analysis.

An FTIR spectrum of the purified protein (10 mM), acquired using a Bruker Vertex 70 spectrometer, displayed a single peak at 2224.9 cm^{-1} . The $\sim 1 \text{ cm}^{-1}$ red shift relative to the in-cell measurements may indicate slight differences in the cellular context (e.g., pH or salt concentration) compared to the purified protein or differences in maturation of the chromophore due to kinetics or the conditions of purification. The enhanced sensitivity of the QCL spectrometer enabled these types of measurements.

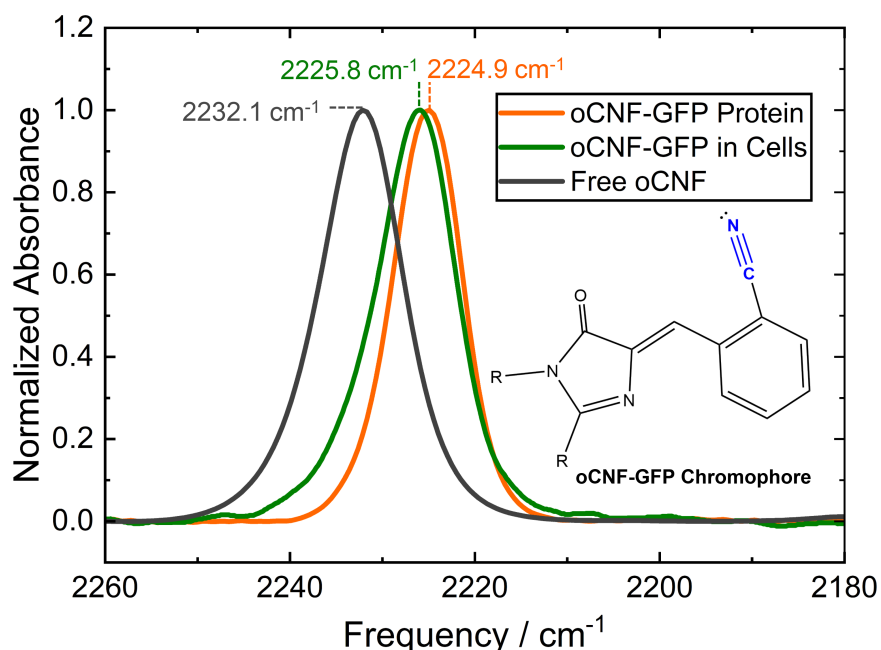


Figure S16. Comparison of QCL-based IR spectrum of Y66oCNF-superfolder GFP within expressing cells (green) to FTIR spectrum of 10 mM purified protein (orange) and of 50 mM free oCNF in water (black). Frequencies are analyzed according to peak maxima.

d. FTIR spectroscopy of PYP variants: In addition to the QCL-based measurements of the nitrile-tagged proteins in expressing cells, we also compare these results to FTIR measurements of the purified proteins, proteins in cells, and a relevant mutant. In general, the QCL-based spectra of pCA-bound *holo*-PYP proteins within the expressing cellular environments (Tables S8-S9) match well (consistently within 0.6 cm^{-1}) of the peak centers for the isolated *holo*-PYP proteins reported previously by Weaver et al.³ We also verify the similarities between the QCL-based spectra of *in cellulo* proteins and the FTIR-based spectra of purified PYP proteins—whether *apo*-PYP or the pCA-incorporated PYP after incubation of the cells with PYP for the purposes of the QCL experiments (Fig. S17, Tables S10-S11). Notably, the *in cellulo* PYP spectra are also distinct from the 2232.1 cm^{-1} for free oCNF (Fig. S16), verifying that these spectral features derive from nitrile-tagged PYP rather than free oCNF in aqueous solution. In general, the nitrile frequencies for PYP in the cell and the isolated protein are within 2 cm^{-1} of one another, with the exception of *apo*-

F92oCNF where the *in cellulo* nitrile frequency is red-shifted by $\sim 4\text{ cm}^{-1}$. Though in both cases the *apo*-form is significantly red-shifted compared to the pCA-bound form, this difference appears to be larger (by $\sim 4\text{ cm}^{-1}$) in the cell than in the isolated protein. Because the F92oCNF results are highly dependent on a dynamic hydrogen-bonding equilibrium involving the nitrile and neighboring T90 as discussed in the main text and evidenced by AMOEBA simulations (Section S7), it is possible that this equilibrium could be highly dependent on not just the chromophore presence but also the environmental conditions such as pH, ionic strength, or crowding/osmotic stress environment. Though the results agree qualitatively whether inside cells or *in vitro*, the slight differences give rise to a view as to how *in cellulo* measurements may be better-poised to capturing the true conformational ensemble of the protein, before or after ligand binding, in live cells.

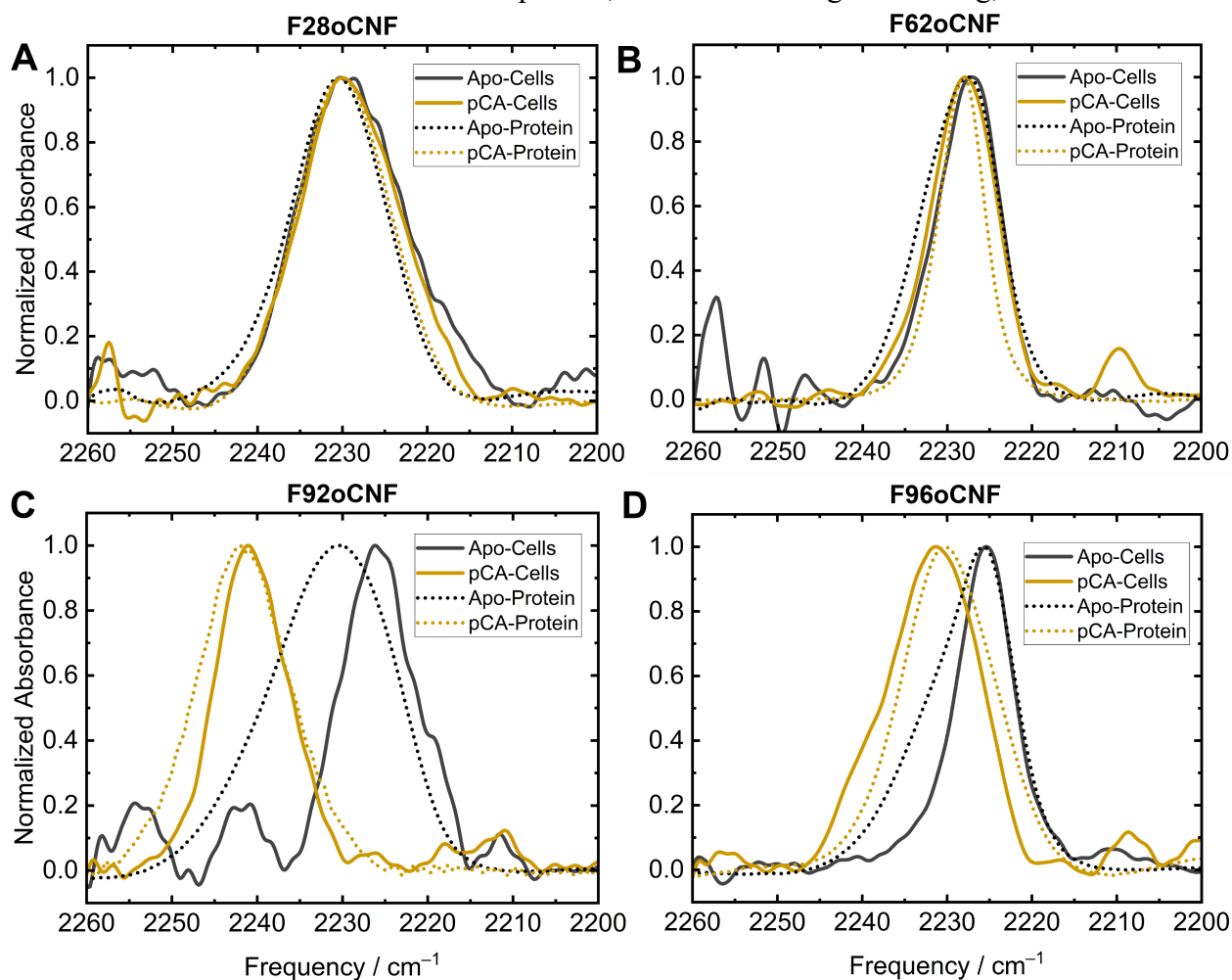


Figure S17. PYP spectra with or without incubation of cells with pCA—the *in cellulo* measurements from the QCL (solid line) are compared to FTIR measurements on the purified proteins (dotted lines). In each case, the purified proteins whose FTIR spectra are shown are the proteins extracted from the same cell batches used for the *apo* or pCA-incubated whole-cell spectroscopy, with the exception of F62oCNF, for which we show the FTIR spectrum from a previous purification by Weaver et al.³

Table S8. Observed frequency of the nitriles in PYP variants with and after pCA addition in cells using the QCL spectrometer. Frequencies are analyzed here according to **curve fitting**.

PYP Variant	Apo-PYP $\bar{\nu}_{\text{Nitrile}} / \text{cm}^{-1}$	Apo-PYP Width / cm^{-1}	pCA-PYP $\bar{\nu}_{\text{Nitrile}} / \text{cm}^{-1}$	Apo-PYP Width / cm^{-1}	$\Delta\bar{\nu}$ After pCA
F28oCNF	2228.8	14.3	2229.1	13.9	+0.3 cm^{-1}
F62oCNF	2227.5	8.7	2228.1	8.9	+0.6 cm^{-1}
F92oCNF	2225.8	9.7	2240.7	9.7	+14.9 cm^{-1}
F96oCNF	2225.4	6.8	2231.7	12.5	+6.3 cm^{-1}

Table S9. Observed frequency of the nitriles in PYP variants with and after pCA addition in cells using the QCL spectrometer. Frequencies are analyzed according to the **peak maxima**.

PYP Variant	Apo-PYP $\bar{\nu}_{\text{Nitrile}} / \text{cm}^{-1}$	pCA-PYP $\bar{\nu}_{\text{Nitrile}} / \text{cm}^{-1}$	$\Delta\bar{\nu}$ After pCA
F28oCNF	2230.5	2230.5	0 cm^{-1}
F62oCNF	2227.3	2227.9	+0.6 cm^{-1}
F92oCNF	2226.2	2241.1	+14.9 cm^{-1}
F96oCNF	2225.3	2231.3	+6.0 cm^{-1}

Table S10. Observed frequency of the nitriles in PYP variants with and after pCA addition after purification using 10mM protein on the FTIR spectrophotometer. Frequencies are analyzed according to **curve fitting**.

PYP Variant	Apo-PYP $\bar{\nu}_{\text{Nitrile}} / \text{cm}^{-1}$	Apo-PYP Width / cm^{-1}	pCA-PYP $\bar{\nu}_{\text{Nitrile}} / \text{cm}^{-1}$	Apo-PYP Width / cm^{-1}	$\Delta\bar{\nu}$ After pCA
F28oCNF	2230.5	13.1	2229.9	12.7	- 0.6 cm^{-1}
F62oCNF	2228.4	10.7	2228.3	6.5	- 0.1 cm^{-1}
F92oCNF	2231.1	16.8	2241.6	13.0	+10.5 cm^{-1}
F96oCNF	2227.1	12.1	2229.9	12.8	+2.8 cm^{-1}

Table S11. Observed frequency of the nitriles in PYP variants with and after pCA addition after purification using 10mM protein on the FTIR spectrometer. Frequencies are analyzed according to the **peak maxima**.

PYP Variant	Apo-PYP $\bar{\nu}_{\text{Nitrile}} / \text{cm}^{-1}$	pCA-PYP $\bar{\nu}_{\text{Nitrile}} / \text{cm}^{-1}$	$\Delta\bar{\nu}$ After pCA
F28oCNF	2230.4	2230.3	- 0.1 cm^{-1}
F62oCNF	2227.6	2228.2	+0.6 cm^{-1}
F92oCNF	2230.5	2242.3	+11.8 cm^{-1}
F96oCNF	2225.9	2230.2	+4.3 cm^{-1}

e. FTIR experiments using a PYP C69A control variant: In addition to FTIR controls involving the isolated nitrile-embedded proteins, we also verified that the *apo*-protein spectra were truly those of a state without any binding to C69, by examining the nitrile frequencies of the double mutant [C69A, F96oCNF]. Using the same procedure as before, we incubated the cells expressing the [C69A, F96oCNF] PYP with pCA before extracting the PYP using a Ni-NTA affinity column and compared the nitrile frequency observed by FTIR to the *apo*-protein F96oCNF and pCA-bound F96oCNF (no mutation to C69). We chose F96oCNF as the probe for this test because there is a shift upon pCA-binding that appears robust both inside and outside of the cell. The FTIR results show a clear similarity of the *apo*-protein F96oCNF to the [C69A, F96oCNF] mutant incubated with pCA (Fig. S18, Table S12), verifying that the pCA is modifying C69 inside the cells and that removal of the cysteine prevents a binding interaction from taking place.

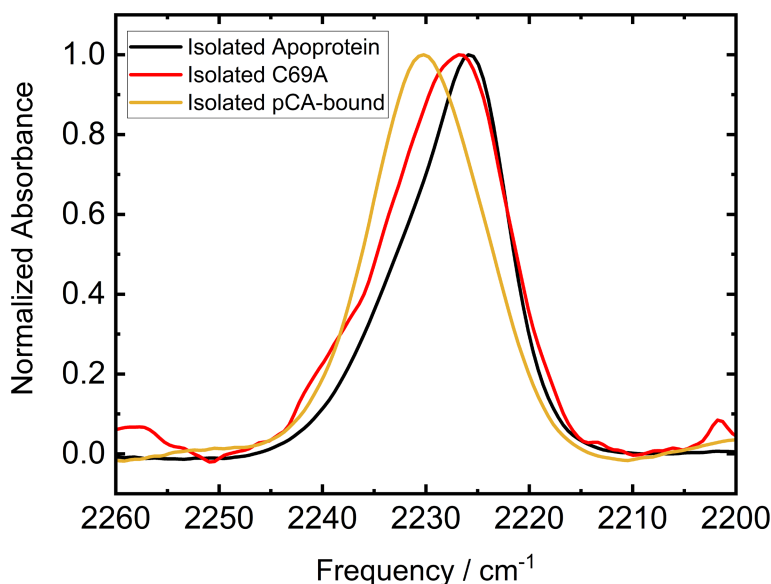


Figure S18. Comparison of FTIR spectra of isolated F96oCNF *apo*-protein (black) and isolated, pCA-bound F96oCNF (gold) to a [C69A, F96oCNF] double mutant that had been incubated with pCA while in cells (red), showing closer similarity between C69A and the *apo*-protein.

Table S12. Frequency Analysis, [C69A F96oCNF] PYP variant using FTIR.

Frequency Analysis	<i>Apo</i> -PYP $\bar{\nu}_{\text{Nitrile}} / \text{cm}^{-1}$	<i>Apo</i> -PYP Width / cm^{-1}
Spectral Fitting	2228.0	13.7
Peak Maximum	2226.7	----

Section S5: Activation of p-Coumaric Acid

a. Chemical reaction of pCA with carbonyldiimidazole: Having established the capability to observe nitrile vibrational probes at low concentrations, we next sought to demonstrate a proof-of-concept study for using these nitrile reporters to detect the binding of drug-like molecules within live cells. PYP from *Halorhodospira halophila* is a well-characterized system known to form a covalent adduct with its chromophore pCA through a nucleophilic attack by the thiol sulfur of the cysteine residue at position 69 (Cys-69) within its binding pocket. In the native organism, this reaction is carried out enzymatically;^{18,19} when expressing PYP recombinantly, the pCA is typically incorporated synthetically. In this process, the carboxylic acid motif of pCA is first activated to form a strong electrophile,^{3,20} effectively mimicking a covalent drug bearing an electrophilic “warhead” that is highly susceptible to nucleophilic attack. This electrophilic center at the carbonyl is specifically designed to undergo a rapid nucleophilic addition reaction with the highly reactive thiol side chain of the Cys-69 residue within the *apo*-PYP binding pocket, leading to the formation of a stable covalent thioester linkage.

Our activation procedure involves reacting pCA (Sigma-Aldrich) with carbonyl diimidazole (CDI, Sigma-Aldrich) to yield an acyl imidazolide “warhead.” This reaction follows the Staab reaction scheme and has been reported in several prior studies, more recently by Foreo-Doria et al.²¹ in their reaction Scheme 2. Consistent with their findings (Scheme 2 in their main text, Product 3b), our procedure uses an approximately 1:1.1 molar ratio of pCA to CDI. Specifically, 1.54 mmol (253.3 mg) of solid pCA, which possesses a carboxylic acid functional group, was combined with 1.72 mmol (279.5 mg) of solid CDI in a rigorously dried 250 mL round-bottom flask. A dry stir bar was included to ensure efficient mixing throughout the reaction.

Reaction of carboxylic acids with CDI typically occurs in two steps: nucleophilic attack of the CDI carbonyl by the deprotonated pCA carboxylate to yield an acid anhydride, followed by nucleophilic attack of the anhydride by the liberated imidazole to generate an acyl imidazolide, plus free imidazole and carbon dioxide (Fig. S19).²²

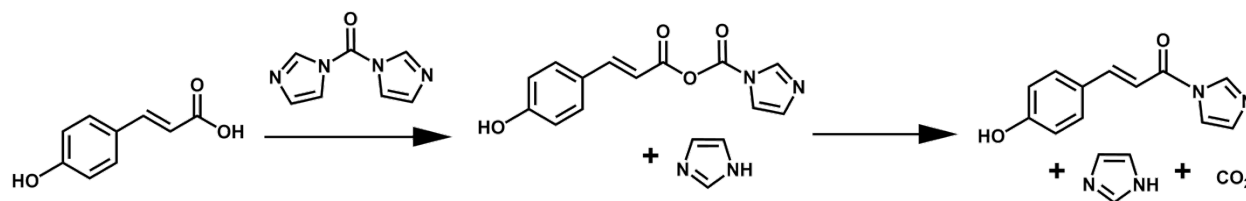


Figure S19. Anticipated reaction scheme for activation of p-Coumaric acid by carbonyl diimidazole (CDI).

To create an inert atmosphere necessary for moisture-sensitive reagents and intermediates, the flask was sealed with a rubber septum and thoroughly purged with argon gas for 30-60 minutes. This displacement of atmospheric air, particularly water vapor and oxygen, prevents unwanted side reactions that could diminish the yield and purity of the desired product. Following the argon purge, approximately 24 mL of anhydrous tetrahydrofuran (THF, Sigma-Aldrich) was carefully added to the flask while the mixture was being stirred rapidly. Anhydrous conditions and absolutely

dry THF are crucial as the activation reaction involving CDI is sensitive to water, which can hydrolyze the CDI (and or the activated intermediate).

The reaction between *p*-Coumaric acid and carbonyldiimidazole was allowed to proceed for 1 hour at room temperature, after which the solvent was removed by rotary evaporation under reduced pressure with gentle heating (maintained at 40 °C). This mild condition ensured efficient removal of tetrahydrofuran (THF) without inducing decomposition of the typically moisture-, light-, and heat-sensitive *N*-acylimidazole product. Evaporation continued until all THF had been removed, yielding the activated *p*-Coumaric acid derivative as a bright yellow oil. The identity of the product was confirmed by a combination of analytical techniques, including FTIR supported by DFT calculations (Fig. S20), ¹³C NMR spectroscopy (Figs. S21-S23), and direct-injection liquid chromatography – mass spectrometry (LC-MS) in acetonitrile (Fig. S24). We note that this protocol is specific to the activation of *p*CA; most covalent drugs have intrinsic reactive warheads and do not require prior activation. Another difference we note is that while our covalent drug analogue must cross a bacterial cell wall, most covalent drugs target mammalian cells and thus can cross the cell membrane with greater ease.

b. Vibrational Spectroscopy Characterization: Density Functional Theory (DFT) calculations were carried out using the 6-311++G(2d,2p) basis set on Gaussian 16²³ and a polarizable continuum model (PCM) for dimethyl sulfoxide (DMSO) to simulate the infrared vibrational spectra of proposed reaction products (Fig. S20). The simulated IR spectrum revealed characteristic coupled vibrational modes centered around 1345 cm⁻¹, which are consistent with the formation of an anhydride or acylimidazole intermediate. These vibrational features are notably absent in the calculated spectra of the starting materials, CDI and *p*CA, suggesting a distinct spectral signature for the product species. Experimental IR spectra of the crude reaction mixture recorded in tetrahydrofuran display an absorption band at 1388 cm⁻¹. This band does not appear in the spectra of the individual reactants (CDI and *p*CA) or in pure imidazole dissolved in THF (expected in crude product mix), further supporting the formation of a new product. The correlation between the experimental and theoretical data reinforces the assignment of the observed 1388 cm⁻¹ feature to product-specific coupled vibrational modes between the *p*CA motif coupled through bonds to the imidazole motif, characteristic of either anhydrides or acylimidazolidine products prompting us to further characterize this product using NMR.

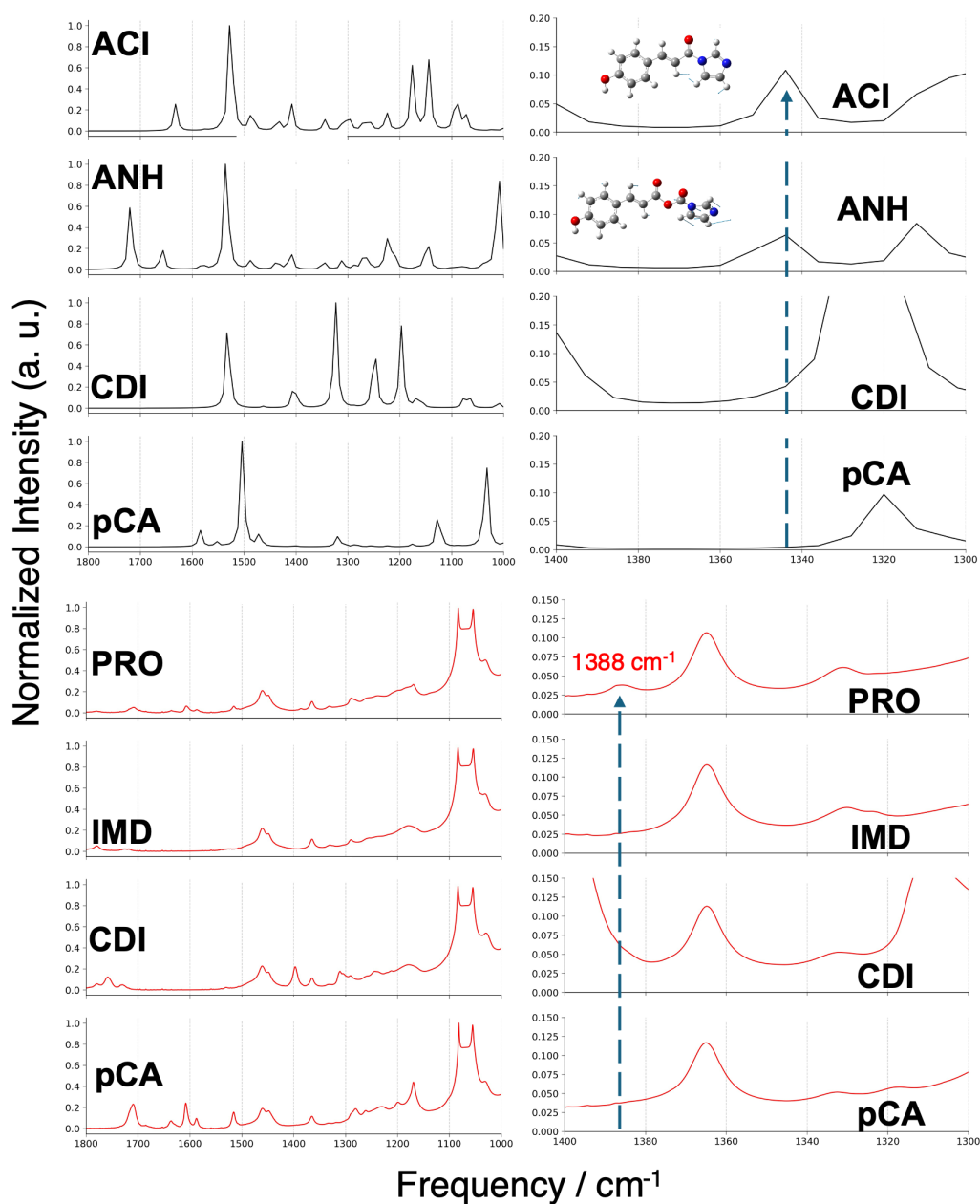


Figure S20. Simulated infrared (IR) spectrum (black) obtained from DFT calculations at the 6-311++G(2d,2p) level using a PCM solvent model for DMSO. The full spectrum (left) and an expanded view of the 1400–1300 cm⁻¹ fingerprint region (right) are shown. The calculations predict coupled vibrational modes near 1345 cm⁻¹, indicative of anhydride or acylimidazole formation. These modes are not observed in the computed spectra of the starting materials, CDI and pCA. Below: Experimental IR spectrum (red) of the crude product in THF, highlighting a distinct band at 1388 cm⁻¹. This feature is absent in the IR spectra of CDI, pCA, and pure imidazole, consistent with the formation of a new chemical species. (ACI=Acylimidazole, ANH=Anhydride, CDI=1,1'-Carbonyldiimidazole, pCA=p-Coumaric Acid, PRO= Crude Product, IMD=Imidazole)

c. ^{13}C NMR Characterization: We collected ^{13}C NMR spectra of the pCA reactant and activated pCA product upon evaporation of the THF and immediate dissolving into deuterated DMSO (DMSO- d_6 , Cambridge Isotope Laboratories, Tewksbury, MA; 256 scans 400 MHz Varian NMR). In all cases, spectra were referenced with respect to the DMSO solvent peak at 39.52 ppm. We first considered the ^{13}C NMR spectrum of the (non-activated) pCA reactant in DMSO- d_6 (Fig. S21) and assigned with reference to a previously reported and assigned spectrum²⁴ as well as an online NMR prediction algorithm.²⁵ The activated pCA product spectrum (Fig. S22) broadly looks like the unreacted pCA spectrum, with some key changes. First, new peaks at 135.6 ppm and 121.8 ppm reflect the presence of imidazole and imidazolid adducts. Two of the previously observed peaks have noticeable shifts, including the carbonyl (g) which shifts 1.5 ppm downfield, and the α -carbon (f), which shifts about 1.8 ppm downfield, indicating that these nuclei are in the closest proximity to a chemical change involving the carboxylic acid. Indeed, the peaks that we observe in our spectrum are practically identical to those observed for a previously reported ^{13}C NMR spectrum of the same activated product by Forero-Doria, et al.²¹ Their purified product, though in a different solvent (methanol- d_4), shows the same number of ^{13}C peaks as ours, but with smaller peaks corresponding to two imidazolid carbon types given that free imidazole had been removed using column chromatography. These imidazolid carbon peaks appear to overlap significantly with the free imidazole peaks so that only one set of peaks is observed in our spectrum. Another consistency is the extremely small peak corresponding to one of the imidazolid carbons at 132 ppm, also noted as miniscule in the Forero-Doria study. Although NMR chemical shifts of the carbonyl region clearly indicate a reaction has occurred, the ambiguity associated with the imidazolid region in NMR was resolved with subsequent methods. We also note X-ray crystallographic data of an O-methoxy derivative of activated pCA synthesized using the same procedure²⁶ which clearly shows the acyl imidazolid form.

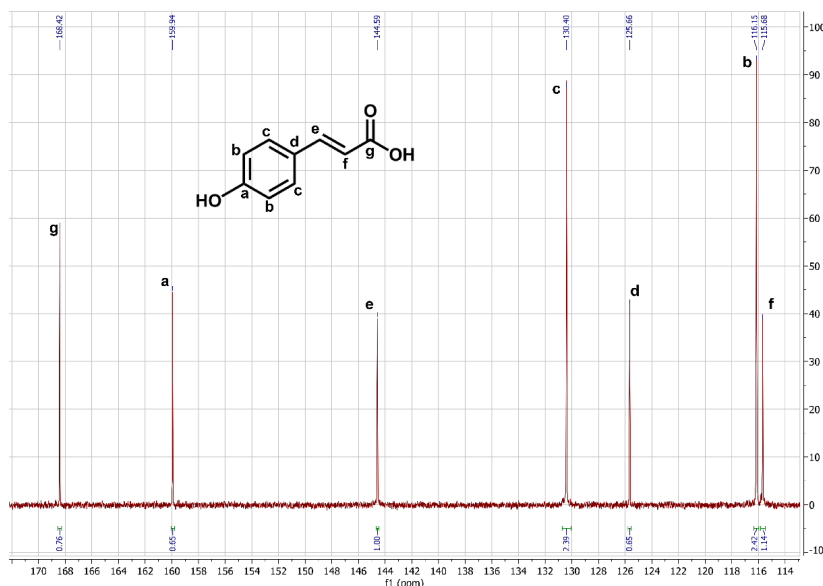


Figure S21. ^{13}C NMR spectrum of unreacted p-Coumaric acid in DMSO- d_6 , with peaks assigned to carbon nuclei in the structure.

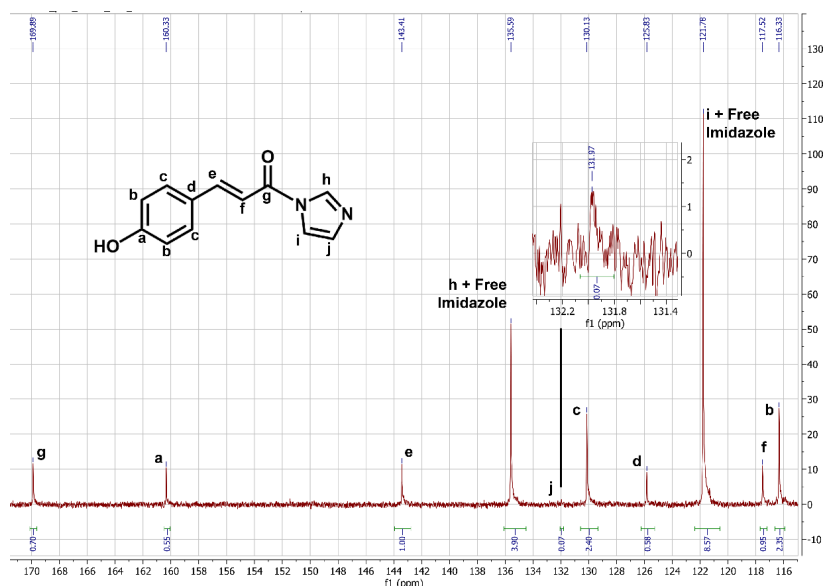


Figure S22. ^{13}C NMR spectrum of CDI-activated p-Coumaric acid in $\text{DMSO}-d_6$, with peaks assigned to carbon nuclei in the structure.

Notably, ^{13}C NMR analysis also indicated that the acyl imidazolide is relatively stable to hydrolysis (Fig. S23). A ^{13}C NMR spectrum of ~600 μl yellow oil dissolved in $\text{DMSO}-d_6$ with the addition of 100 μl H_2O and incubation for 1 hour at room temperature showed only minor chemical shift changes (~1 ppm or less for all carbon resonances), further confirming the stability of the acyl imidazolide under mildly aqueous conditions. These results support that the activated p-Coumaric acid (pCA) derivative remains intact and is thus suitable for cellular uptake and specific binding to its intended biological target.

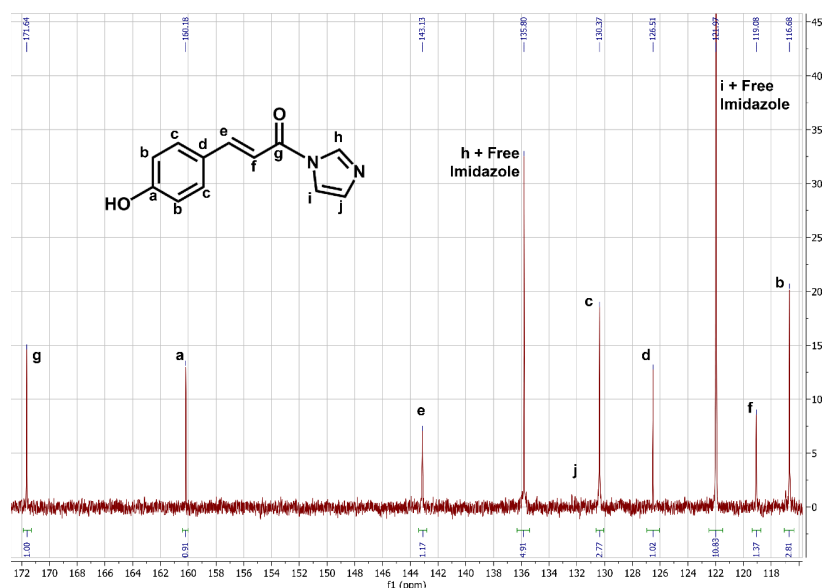


Figure S23. ^{13}C NMR spectrum of CDI-activated p-Coumaric acid in $\text{DMSO}-d_6$ after spiking in 100 μl H_2O and incubation for 1 hour at room temperature, with peaks assigned to carbon nuclei in the structure.

d. Characterization by Mass Spectrometry: The best indication of the fully formed product is a direct-injection mass spectrum of the product in acetonitrile (Fig. S24), which clearly shows its largest peak at $m/z = 215$ as expected for the full acyl imidazolid product (mass=214), and a smaller peak at $m/z=69$ for free imidazole.

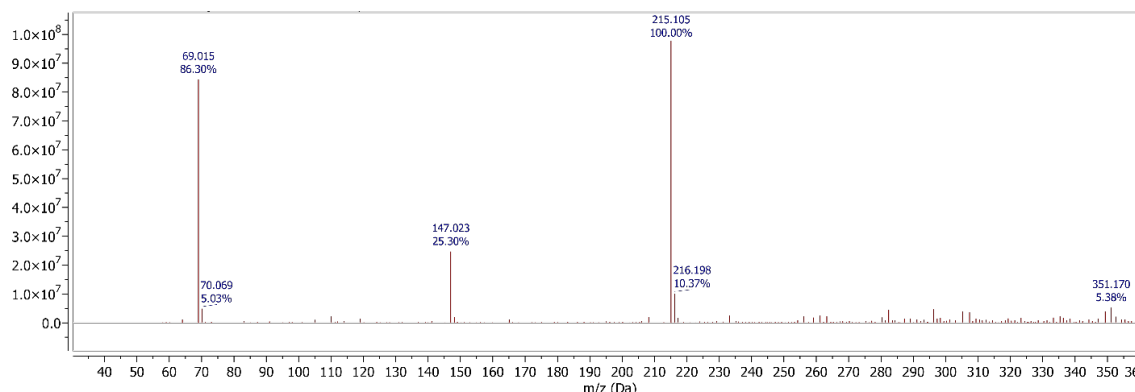


Figure S24. Direct-injection mass spectrum of CDI-activated p-Coumaric acid in acetonitrile, using a quadrupole LC-MS with 20-mV ESI cone voltage (positive ion mode).

e. Addition of activated pCA to PYP-expressing bacteria: Cells were always incubated with freshly synthesized activated pCA. The molar ratio of activated pCA to *apo*-PYP variants was estimated using the expression levels as described in section S2, and adjusted to be an approximately 100-fold excess of activated pCA to each *apo*-PYP copy. Cells were added directly to the freshly evaporated yellow oil and allowed to incubate overnight at 4 °C with gentle stirring. As shown by ^{13}C NMR results, the activated pCA is sufficiently stable in the presence of water to enable crossing of the membrane to react with PYP well before its hydrolysis. The process of cell washing and concentrating is the same with or without the incorporated pCA (Fig. S25).

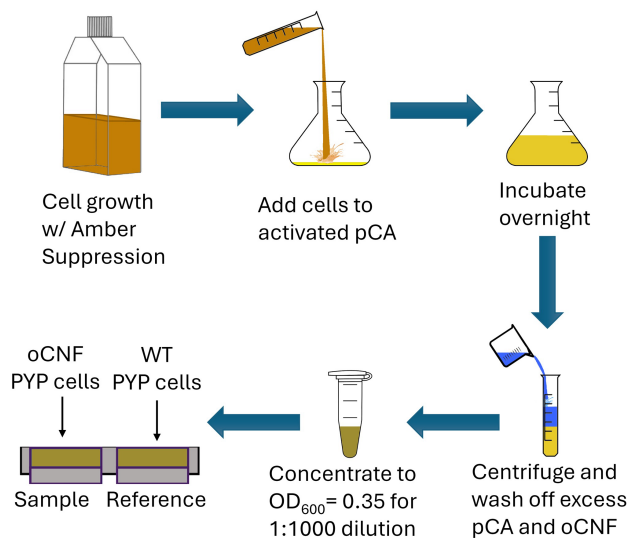


Figure S25. Sample preparation protocol showing the process of cell growth and optional reaction with activated pCA, prior to washing of excess oCNF and/or pCA, concentrating, and loading into a double liquid sample cell.

Section S6: Mass Spectrometry

a. Confirming chromophore incorporation post-lysis: An additional test verifying the incorporation of pCA chromophore into *apo*-PYP relies on mass spectrometry, where covalent binding of pCA causes a significant and measurable mass increase. Wild-type PYP was purified from its expressing *E. coli* cells, a portion of which had been incubated overnight with activated pCA using the protocol described above. A minimal purification procedure employed here focused on Ni-column affinity chromatography to purify the His-tagged PYP, with no further tag cleaving or purification. The resulting proteins were run on LC-MS using a C8 reverse-phase chromatography column using a gradient from 0.1% formic acid in water, to methanol. Mass spectrometry was carried out using electrospray ionization with a single quadrupole detector.

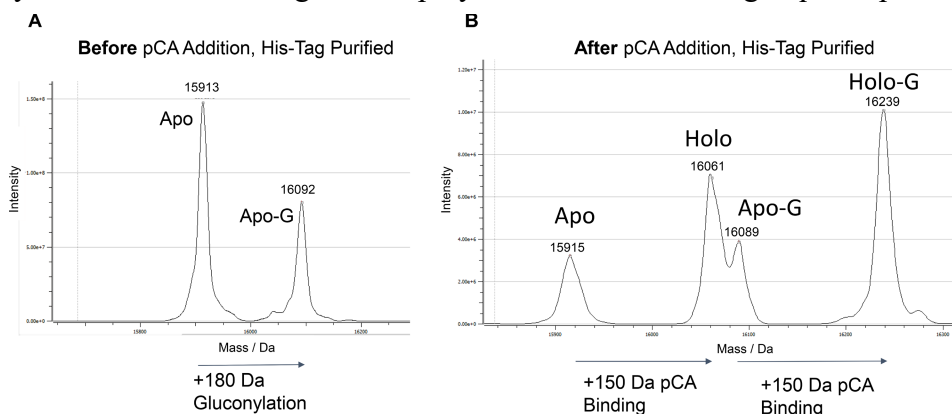


Figure S26. Mass spectrometry of Ni-column-purified, His-tagged PYP confirms pCA incorporation by the presence of ~150 Da mass-shifted peaks between (A) *apo*-PYP and (B) PYP from cells incubated with pCA. A fraction of the His-tagged PYP is gluconylated (*apo*-G and *holo*-G).

In both cases (with or without activated pCA incubation), the protein eluted as a single band whose mass spectrum was deconvolved to yield the major peaks shown in Fig. S26. The PYP from cells not incubated with activated pCA showed two primary mass peaks, corresponding to the His-tagged pBAD PYP (see sequence in S2) with and without a gluconylation, a common (+180 Da) post-translational modification for recombinantly expressed proteins in *E. coli* with N-terminal hexahistidine tags which we have observed previously for the tagged pBAD PYP.^{27,28} The protein from the cells incubated with activated pCA showed the same peaks, plus the inclusion of larger peaks shifted by approximately 150 Da and approximately reflecting the mass increase expected due to cysteine forming a thioester with activated pCA in the chromophore binding pocket (+145 Da). Hence, mass spectrometry confirms the covalent incorporation of pCA into PYP. One aspect of note is the continued presence of *apo*-protein peaks at a lower magnitude compared to the pCA-incorporated PYP, whereas in the *in cellulo* IR measurements of the nitrile-incorporated variants, we did not see significant populations of *apo*-protein. Possible sources of this difference are (1) expression of the WT-PYP is significantly higher than the oCNF-incorporated *apo*-PYP, giving more protein in the WT cases which may not completely react with activated pCA; (2) the thioester linkage between *apo*-PYP and pCA is not completely stable in all circumstances, as detailed below, and it is possible that the acidic conditions involved in the liquid chromatography may be

facilitating pCA hydrolysis on the instrument. Regardless, the ability to detect significant peaks for pCA-bound PYP confirms the covalent incorporation of chromophore detected using *in cellulo* QCL-based IR spectroscopy.

b. Limitations of mass spectrometry methods including proteomics: We note that while mass spectrometry is helpful here for the detection of small molecules that bind via stable covalent linkages to the target protein of interest, it may not necessarily always be useful for detection of ligands that bind weakly or noncovalently to the target protein inside cells. Even with covalent linkages such as the thioester bond between *apo*-PYP and pCA, the ligand may still be susceptible to hydrolytic cleavage under select conditions such as low pH as discussed above. We explored the use of mass spectrometry proteomics to study binding of pCA to proteins within the expressing *E. coli* cells. The most common methods involve a sample preparation that treats a cell lysate with sodium dodecyl sulfate (SDS) detergent to denature the proteins, followed by incubation with dithiothreitol (DTT) or tris(2-carboxyethyl)phosphine (TCEP) to reduce disulfide bonds and linearize the peptide sequences, capping of the cysteines using acrylamide or iodoacetamide to stabilize the linearized form, and digestion using a protease such as trypsin. We found that while the protein lost its 445-nm yellow color immediately upon denaturation by SDS at pH 7.4 (attributed to loss of the ordered chromophore pocket that stabilizes the deprotonated pCA chromophore), the chromophore was still attached by a thioester linkage at this point, giving rise to a 345-nm peak. Pure pCA at pH 8, by contrast, does not have a 345-nm peak but absorbs at 285 and 305 nm. Adding DTT or TCEP further altered the spectrum to a small 334-nm peak in addition to the larger protein 280-nm peak, and mass spectrometry showed no remaining pCA bound to PYP at this stage. Indeed, when the proteomics sample digestion procedure was applied in full to cell lysates using either DTT or TCEP, and iodoacetamide to cap the cysteines, no statistically significant hits in the *E. coli* genome were noted as being bound to pCA, not even the recombinantly expressed *apo*-PYP itself—indicating that the sample preparation scheme was harsh enough to break the thioester linkage by which activated pCA might bind itself to cysteines.

Unfortunately, the reducing conditions required for peptide linearization and mass spectrometry appear incompatible with the thioester linkage of interest in our system. While mass spectrometry-based proteomics has proven itself versatile in detecting covalent interactions including diverse electrophilic warheads attacked by cysteines, and new techniques within the field have even emerged for observing non-covalent complex formation,²⁹ the limitations shown by the present study indicate that challenges still remain for mass spectrometry assays involving weak covalent bonds susceptible to hydrolysis. Our method stands as an independent and entirely orthogonal, non-destructive approach to detecting interactions between target proteins and small molecules *in vivo* using the frequencies of minimally perturbative vibrational probes positioned near the binding site. Not only is our method complementary in terms of certain pitfalls characteristic to mass spectrometry methods, but it also provides new categories of information typically not available to mass spectrometry—highly specific structural details on electrostatics or hydrogen bonding patterns in the native environment, for instance.

Section S7: Computational details

Molecular dynamics simulations for the nitrile incorporated F28oCNF, F62oCNF, F92oCNF, and F96oCNF variants of PYP were initiated from their respective X-ray crystal structures of the *holo*-form, corresponding to PDB IDs 7SPX, 7SPW, 7SPV, and 7SJJ, respectively. Two approaches were taken to simulate the apo-PYP. First the coordinates of *holo*-form crystal structure PDB ID 1NWZ were used. The chromophore under the residue HC4 in 1NWZ was removed and the residue C69 was then protonated. In the second approach the apoprotein was modeled using the coordinates from PDB ID 9O8V, the new structure of the apo-PYP submitted as part of this study. Incomplete residues and sidechains were modeled using MODELLER 10.6³⁰ and using the Protein Analysis and Repair Server.³¹

a. Fixed Charge Molecular Dynamics Simulations: Fixed charge MD simulations were conducted using the AMBER99SB-ILDN force field,³² as implemented in GROMACS 2021.3³³, on the Sherlock HPC Cluster at Stanford University. Each protein was centered in a cubic simulation box (40 Å per side), solvated with TIP3P water,³⁴ and neutralized with 100 mM NaCl to ensure charge neutrality and to mimic experimental ionic strength. Energy minimization was performed using the steepest descent algorithm until the maximum force dropped below 1000 kJ mol⁻¹ nm⁻¹. Equilibration was carried out under periodic boundary conditions via sequential NVT and NPT ensembles (500 ps each), with progressively relaxed restraints: initially on all heavy atoms, then only on C α atoms, and finally with no restraints. Temperature was maintained at 300 K using a velocity-rescaling thermostat,³⁵ while pressure was kept at 1 bar using the Berendsen barostat³⁶ during equilibration and the Parrinello–Rahman barostat³⁷ during production. Production simulations were run under NPT conditions with a 1 fs integration time step. Non-bonded interactions were treated with a 1.2 nm cutoff, and long-range electrostatics were calculated using the Particle Mesh Ewald (PME) method.³⁸ All X–H bonds were constrained using the LINCS algorithm.³⁹

To evaluate the effect of chromophore removal from *holo*-form crystal structures for the simulations, we compared a 100 ns simulation of the *apo*-protein (9O8V) with that of the *holo*-protein (1NWZ) following computational chromophore removal. No significant conformational differences were observed between the runs (Fig. S27). This result validated the use of *in silico* chromophore removal as a strategy for oCNF variants, for which only *holo*-form crystal structures are available. Accordingly, we applied this approach in AMOEBA simulations to evaluate local electrostatics and investigate changes linked to IR spectral shifts. We therefore proceeded by removing the chromophore from the *holo*-form crystal structures prior to running the AMOEBA simulations (Fig. S28).

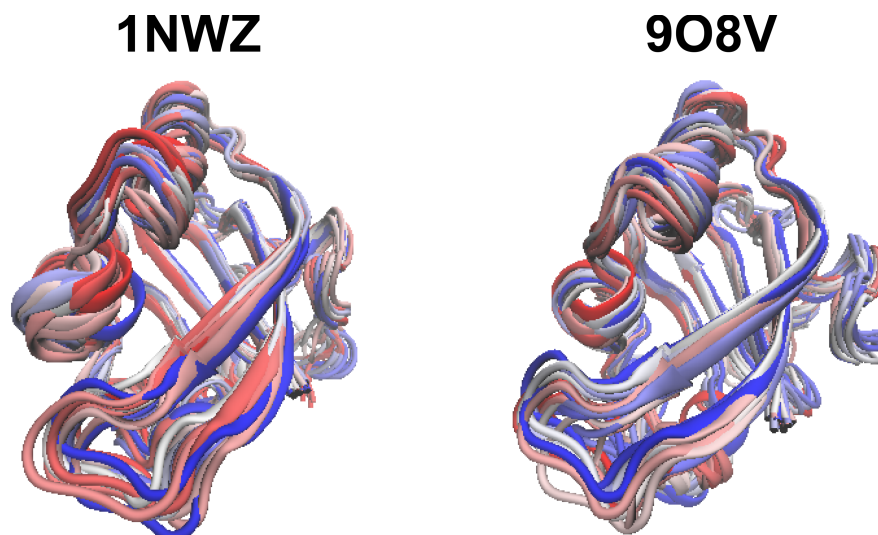


Figure S27. Fixed-charge molecular dynamics simulations (100 ns) were performed for *apo*-PYP starting from both the *holo* structure (PDB ID: 1NWZ, with the chromophore removed) and the *apo* structure (PDB ID: 9O8V). Snapshots were taken every 5 ns, with colors ranging from blue (early time points) to red (later time points). No major conformational changes or alterations in secondary structure were observed between the two simulations.

b. Polarizable Force Field Molecular Dynamics Simulations: To more accurately capture the electrostatic environment surrounding nitrile probes, we employed polarizable molecular dynamics (POL MD) simulations using the AMOEBA-BIO18 force field,⁴⁰ as implemented in Tinker 9.^{41,42} This approach, established through extensive work by Kirsh et al.,¹² effectively incorporates polarization effects critical for modeling nitrile vibrational behavior. System setup mirrored that of the fixed-charge (FC) protocol, using a 7.5 nm cubic box containing 50 mM NaCl in water. Energy minimization was performed using the steepest descent algorithm until residual forces were below $1 \text{ kcal} \cdot \text{mol}^{-1} \cdot \text{\AA}^{-1}$, with electrostatics and van der Waals cutoffs of 7 Å and 9 Å, respectively. A mutual polarization scheme with a dipole convergence threshold of 0.01 D was applied. Parameters for oCNF non-canonical amino acid incorporation have been provided earlier in Kirsh et al.¹²

Equilibration under NVT and NPT conditions followed the same restraint scheme used in the FC simulations, with the van der Waals cutoff extended to 12 Å. Simulations utilized the RESPA integrator,⁴³ the Bussi thermostat,³⁵ and a Monte Carlo barostat (300 K, 1 bar). Production simulations were run under NPT conditions with a 1 fs time step for 25 ns per replicate. Four independent runs were performed for each protein variant, yielding a total of 100 ns of aggregate trajectory per system. This was also repeated using a protonated E46 residue, but no significant changes were noted. Using Tinker, induced dipoles on the C and N atoms were recorded every 10 ps. These dipoles were divided by the corresponding atomic polarizabilities to compute the electric fields. The resulting field vectors were projected onto the $\text{--C}\equiv\text{N}$ bond axis and averaged to determine the final $F_{\text{C}\equiv\text{N}}$. Data analysis scripts can be found in:

https://github.com/KozuchLab/Publications/tree/main/oCNPhe_GROMACS_TINKER

c. Residue-Specific Fluctuation Profiles: As previously mentioned, the resolved global structure of *apo*-PYP exhibited minimal deviation in local disorder from the *holo*-form based on Z-score analyses, maintaining structural integrity even without incorporated nitrile probes. Fixed-charge molecular dynamics simulations (100-ns duration) conducted using dual starting configurations: the chromophore-removed *holo* structure (PDB ID: 1NWZ) and native *apo* structure (PDB ID: 9O8V) also supported this observation. We also now provide residue-specific AMOEBA polarizable force field RMSF analyses for both *holo* and *apo* forms of the nitrile-incorporated F92oCNF and F96oCNF (Fig. S28). The *apo* system exhibited marginally higher flexibility (<1 Å RMSF increase) compared to the *holo* form, particularly near the chromophore binding pocket residues Y42, E46, and C69. This localized flexibility enhancement aligns with chemical intuition, as pCA removal eliminates stabilizing interactions in the binding cleft while preserving overall tertiary structure. The dashed traces in Figure S28 quantitatively demonstrate this differential mobility. These dynamical profiles enable us to make quantitative predictions about local electrostatic environments derived from AMOEBA-simulated hydrogen bonding patterns. The enhanced *apo*-form flexibility correlates with modified electric field magnitudes at key chromophore-interaction sites, and these computational insights directly explain the IR shifts such as the experimentally observed hypsochromic shifts upon pCA binding of F92oCNF and F96oCNF embedded nitriles, measured via QCL vibrational spectroscopy (main text Fig. 3). The combined MD-spectroscopy approach establishes a quantitative framework linking residue-specific flexibility to spectral perturbations through electric field modulation.

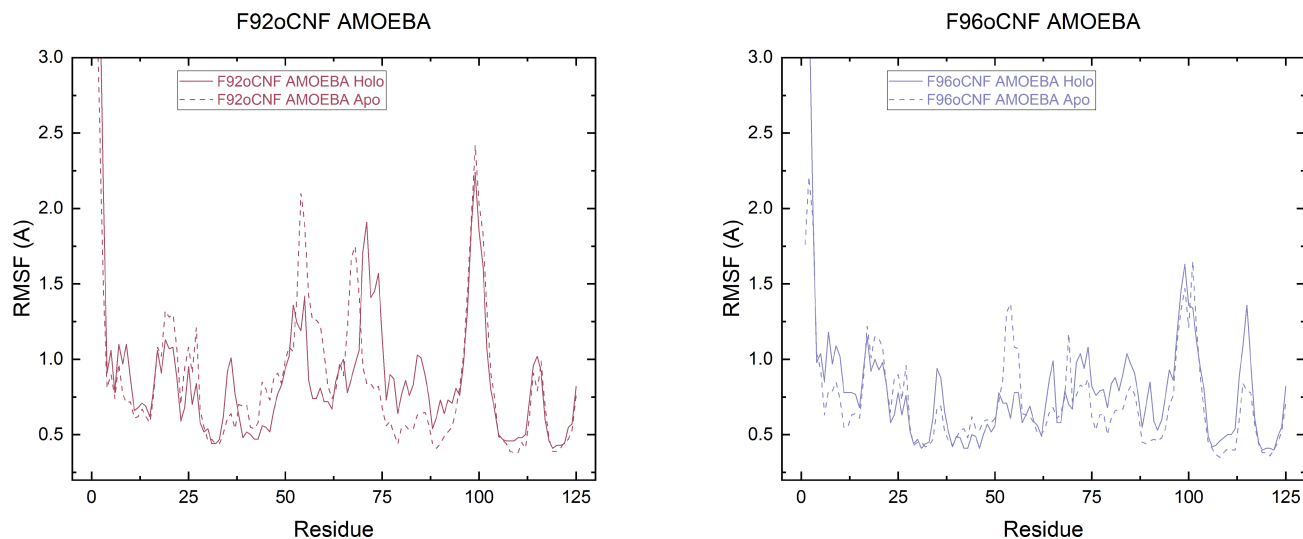


Figure S28. Residue-specific RMSF differences between AMOEBA simulations of F92oCNF (magenta) and F96oCNF (light blue), comparing *holo* (chromophore present, solid lines) and *apo* (computationally removed chromophore, dashed lines) forms. Only minor deviations in RMSF are observed between the *holo* and *apo* form generated by computational chromophore removal.

Section S8: Physical interpretability of nitrile shifts

Using the simulation results, we can provide a physical interpretation for the frequency shifts of genetically encodable nitrile probes in terms of significant physical and chemical properties, i.e., hydrogen bonding and electric field. In the case of F92oCNF and F96oCNF, the variants whose nitrile frequencies shift the most upon pCA binding, we showed in the main text how these frequency shifts are dependent on hydrogen-bonding patterns involving the nitrile. These results are showcased most clearly by the AMOEBA MD simulations but are also substantiated by temperature-dependent FTIR spectroscopy involving the purified F92oCNF and F96oCNF proteins. Below, we compare FTIR spectra of the *apo*-protein F92oCNF and F96oCNF to the pCA-bound variants at different temperatures (100K and 303K), taken from the data of Kirsh et al.¹² (Fig. S29). As can be observed, hydrogen-bonded (blue-shifted) and non-hydrogen-bonded populations become more distinguishable at lower temperatures. In pCA-bound F92oCNF, the small non-hydrogen-bonding population at 100K aligns well with the *apo*-F92oCNF at room temperature, indicating that hydrogen bonding or lack thereof can help explain the frequency differences between the *apo* and *holo* (pCA-bound) forms. At warmer temperatures such as 303K, the rate of chemical exchange between the hydrogen-bonded and non-hydrogen-bonded populations becomes faster such that the two populations appear to merge in the IR spectrum. In both F92oCNF and F96oCNF, the warmer temperature leads to a peak frequency in between the hydrogen-bonded *holo* population (the primary band at low temperature) and the non-hydrogen-bonded *apo*-protein spectrum. Hence, the hydrogen-bonding explanation for the peak shifts in *apo*-PYP is in good agreement with the explanation of Kirsh et al. in the case of *holo*-PYP.¹²

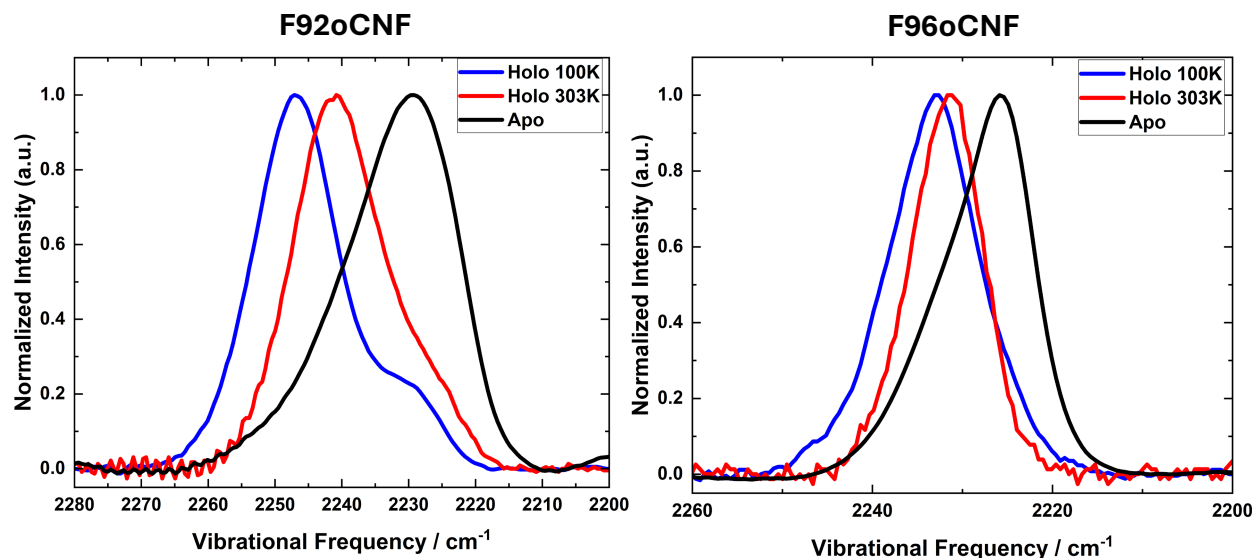


Figure S29. Isolated *apo*-protein F92oCNF and F96oCNF FTIR spectra are compared to FTIR spectra of the isolated *holo*-F92oCNF and *holo*-F96oCNF at various temperatures (taken from ref. 12). The low-temperature *holo*-proteins are more blue-shifted indicating a favoring of the hydrogen-bonding population at low temperatures, which is more distinguishable from the non-hydrogen-bonding population (red-shifted) that aligns more closely with the *apo*-protein.

Supplemental References

1. Akhgar, C. K., Ramer, G., Żbik, M., Trajnerowicz, A., Pawluczyk, J., Schwaighofer, A., & Lendl, B. (2020) The Next Generation of IR Spectroscopy: EC-QCL-Based Mid-IR Transmission Spectroscopy of Proteins with Balanced Detection. *Anal. Chem.*, 92(14), 9901–9907. <https://doi.org/10.1021/acs.analchem.0c01406>
2. Alcaráz, M. R., Schwaighofer, A., Kristament, C., Ramer, G., Brandstetter, M., Goicoechea, H., & Lendl, B. (2015) External-Cavity Quantum Cascade Laser Spectroscopy for Mid-IR Transmission Measurements of Proteins in Aqueous Solution. *Anal. Chem.*, 87(13), 6980–6987. <https://doi.org/10.1021/acs.analchem.5b01738>
3. Weaver, J. B., Kozuch, J., Kirsh, J. M., & Boxer, S. G. (2022) Nitrile Infrared Intensities Characterize Electric Fields and Hydrogen Bonding in Protic, Aprotic, and Protein Environments. *J. Am. Chem. Soc.*, 144(17), 7562–7567. <https://doi.org/10.1021/jacs.2c00675>
4. MIRcat Mid-IR Laser. *DRS Daylight Solutions*. <https://www.daylightsolutions.com/products/mircat/>
5. Fried, S. D. E., Mukherjee, S., Mao, Y., & Boxer, S. G. (2024) Environment- and Conformation-Induced Frequency Shifts of C–D Vibrational Stark Probes in NAD(P)H Cofactors. *J. Phys. Chem. Lett.*, 15(43), 10826–10834. <https://doi.org/10.1021/acs.jpcclett.4c02497>
6. Romei, M. G., von Krusenstiern, E. V., Ridings, S. T., King, R. N., Fortier, J. C., McKeon, C. A., Nichols, K. M., Charkoudian, L.K., & Londergan, C. H. (2023) Frequency Changes in Terminal Alkynes Provide Strong, Sensitive, and Solvatochromic Raman Probes of Biochemical Environments. *J. Phys. Chem. B*, 127, 1, 85-94. <https://doi.org/10.1021/acs.jpcb.2c06176>
7. Zheng, C., Mao, Y., Kozuch, Y., Atsango, A., Ji, Z., Markland, T. E., & Boxer, S. G. (2022) A two-directional vibrational probe reveals different electric field orientations in solution and an enzyme active site. *Nat. Chem.*, 14, 891-897. <https://doi.org/10.1038/s41557-022-00937-w>
8. Lin, C.-Y., Romei, M. G., Oltrogge, L. M., Mathews, I. I., & Boxer, S. G. (2019) Unified Model for Photophysical and Electro-Optical Properties of Green Fluorescent Proteins. *J. Am. Chem. Soc.*, 141, 15250-15265. <https://doi.org/10.1021/jacs.9b07152>
9. Romei, M. G., Lin, C.-Y., Mathews, I. I., & Boxer, S. G. (2020) Electrostatic control of photoisomerization pathways in proteins. *Science*, 367, 76–79. <https://doi.org/10.1126/science.aax1898>
10. Olenginski, G.M., Piacentini, J., Harris, D. R., Runko, N. A., Papoutsis, B. M., Alter, J. R., Hess, K. R., Brewer, S. H., & Phillips-Piro, C. M. (2021) Structural and spectrophotometric investigation of two unnatural amino-acid altered chromophores in the superfolder green fluorescent protein. *Acta Crystallogr.*, D77, 1010-1018. <https://doi.org/10.1107/S2059798321006525>
11. Getzoff, E. D., Gutwin, K. N., & Genick, U. K. (2003) Anticipatory active-site motions and chromophore distortions prime photoreceptor PYP for light activation. *Nat. Struct. Biol.*, 10, 663-668. <https://doi.org/10.1038/nsb958>
12. Kirsh, J. M., Weaver, J. B., Boxer, S. G., & Kozuch, J. (2024) Critical Evaluation of Polarizable and Nonpolarizable Force Fields for Proteins Using Experimentally Derived Nitrile Electric Fields. *J. Am. Chem. Soc.*, 146(10), 6983–6991. <https://doi.org/10.1021/jacs.3c14775>
13. Romei, M. G., Lin, C.-Y., & Boxer, S. G. (2020). Structural and Spectroscopic Characterization of Photoactive Yellow Protein and Photoswitchable Fluorescent Protein Constructs Containing Heavy Atoms. *J. Photochem. Photobiol. A: Chem.* 401, 1112738. <https://doi.org/10.1016/j.jphotochem.2020.112738>

14. Minor, W., Cymborowski, M., Otwinowski, Z., & Chruszcz, M. (2006) HKL-3000: The integration of data reduction and structure solution—from diffraction images to an initial model in minutes. *Acta Crystallogr. D* 62(8), 859-866. <https://doi.org/10.1107/S0907444906019949>
15. Liebschner, D., Afonine, P. V., Baker, M. L., Bunkóczi, G., Chen, V. B., Croll, T. I., Hintze, B., Hung, L. W., Jain, S., McCoy, A. J., Moriarty, N. W., Oeffner, R. D., Poon, B. K., Prisant, M. G., Read, R. J., Richardson, J. S., Richardson, D. C., Sammito, M. D., Sobolev, O. V., Stockwell, D. H., Terwilliger, T. C., Urzhumtsev, A. G., Videau, L. L., Williams, C. J., & Adams, P. D. (2019) Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. *Acta Crystallogr. D* 75(10), 861-877. <https://doi.org/10.1107/S2059798319011471>
16. Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010) Features and development of *Coot*. *Acta Crystallogr. D* 66, 486-501. doi:10.1107/S0907444910007493
17. Borucki, B., Otto, H., Meyer, T. E., Cusanovich, M. A., & Heyn, M. A. (2005). Sensitive Circular Dichroism Marker for the Chromophore Environment of Photoactive Yellow Protein: Assignment of the 307 and 318 nm Bands to the N \rightarrow π^* Transition of the Carbonyl. *J. Phys. Chem. B*, 109, 629-633. <https://doi.org/10.1021/jp046515k>
18. Meyer, T. E. (1985). Isolation and characterization of soluble cytochromes, ferredoxins and other chromophoric proteins from the halophilic phototrophic bacterium *Ectothiorhodospira halophila*. *Biochim. Biophys. Acta*, 806(1), 175–183. [https://doi.org/https://doi.org/10.1016/0005-2728\(85\)90094-5](https://doi.org/https://doi.org/10.1016/0005-2728(85)90094-5)
19. Brechun, K. E., Zhen, D., Jaikaran, A., Borisneko, V., Kumauchi, M., Hoff, W. D., Arndt, K. M., & Woolley, G. A. (2019) Detection of Incorporation of *p*-Coumaric Acid into Photoactive Yellow Protein Variants *in Vivo*. *Biochemistry*, 58(23), 2682-2694. <https://doi.org/10.1021/acs.biochem.9b00279>
20. Genick, U. K., Devanathan, S., Meyer, T. E., Canestrelli, I. L., Williams, E., Cusanovich, M. A., Tollin, G., & Getzoff, E. D. (1997) Active Site Mutants Implicate Key Residues for Control of Color and Light Cycle Kinetics of Photoactive Yellow Protein. *Biochemistry*, 36(1), 8-14. <https://doi.org/10.1021/bi9622884>
21. Forero-Doria, O., Araya-Maturana, R., Barrientos-Retamal, A., Morales-Quintana, L., Guzmán, L. N-alkylimidazolium Salts Functionalized with *p*-Coumaric and Cinnamic Acid: A Study of Their Antimicrobial and Antibiofilm Effects. *Molecules*. 2019 Sep 26;24(19):3484. doi: 10.3390/molecules24193484.
22. Engstrom, K. M. (2018) Practical Considerations for the Formation of Acyl Imidazolides from Carboxylic Acids and *N,N'*-Carbonyldiimidazole: The Role of Acid Catalysis. *Org. Process Res. Dev.*, 22(9) 1294-1297. <https://doi.org/10.1021/acs.oprd.8b00121>
23. Frisch, M. J., Trucks, G. W., Schlegel, H. B., Scuseria, G. E., Robb, M. A., Cheeseman, J. R., Scalmani, G., Barone, V., Petersson, G. A., Nakatsuji, H., Li, X., Caricato, M., Marenich, A. V., Bloino, J., Janesko, B. G., Gomperts, R., Mennucci, B., Hratchian, H. P., Ortiz, J. V., Izmaylov, A. F., Sonnenberg, J. L., Williams-Young, D., Ding, F., Lipparini, F., Egidi, F., Goings, J., Peng, B., Petrone, A., Henderson, T., Ranasinghe, D., Zakrzewski, V. G., Gao, J., Rega, N., Zheng, G., Liang, W., Hada, M., Ehara, M., Toyota, K., Fukuda, R., Hasegawa, J., Ishida, M., Nakajima, T., Honda, Y., Kitao, O., Nakai, H.; Vreven, T., Throssell, K., J. A. Montgomery, J., Peralta, J. E., Ogliaro, F., Bearpark, M., Heyd, J. J., Brothers, E. N.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A. P., Burant, J. C., Iyengar, S. S., Tomasi, J., Cossi, M., Rega, N., Millam, J. M., Klene, M., Adamo, C., Cammi, R., Ochterski, J. W., Martin, R. L., Morokuma, K., Farkas, O., Foresman, J. B., & Fox, D. J. (2016) Gaussian, Inc.: Wallingford CT, 2016.
24. Madison Metabolics Consortium—Jofre, F., Anderson, M. E., Markley, J. L. *P-Coumaric Acid*. (bmse000591). BRMB. doi:10.13018/BMSE000591
25. Banfi, D. & Patiny, L. (2008) www.nmrdb.org: Resurrecting and Processing NMR Spectra On-line. *Chimia*, 62, 280, DOI: 10.2533/chimia.2008.280

-
26. Cappelli, A., Grisci, G., Paolino, M., Guilani, g., Donati, A., Mendichi, R., Artusi, R., Demiranda, M., Zanardi, A., Giorgi, G., & Vomero, S. (2014) Hyaluronan derivatives bearing variable densities of ferulic acid residues. *J. Mater. Chem. B*, 2, 4489-4499. <https://doi.org/10.1039/C3TB21824D>
27. Aon, J. C., Caimi, R. J., Taylor, A. H., Lu, Q., Oluboyede, F., Dally, J., Kessler, M. D., Kerrigan, J. J., Lewis, T. S., Wysocki, L. A., & Patel, P. S. (2008) Suppressing Posttranslational Gluconoylation of Heterologous Proteins by Metabolic Engineering of *Escherichia coli*. *Appl. Environ. Microbiol.*, 74(4), 950-958. <https://doi.org/10.1128/AEM.01790-07>
28. Schweida, D., Barraud, P., Regl, C., Loughlin, F. E., Huber, C. G., Cabrele, C., & Schubert, M. (2010) The NMR signature of gluconoylation: a frequent N-terminal modification of isotope-labeled proteins. *J. Biomol. NMR* 73, 71–79. <https://doi.org/10.1007/s10858-019-00228-6>
29. Meissner, F., Geddes-McAlister, J., Mann, M., & Bantscheff, M. (2022) The emerging role of mass spectrometry-based proteomics in drug discovery. *Nat. Rev. Drug Discov.*, 21, 637-654. <https://doi.org/10.1038/s41573-022-00409-3>
30. Webb, B., & Sali, A. (2016) Comparative Protein Structure Modelling Using MODELLER. *Current Protocols in Bioinformatics*, 54, John Wiley & Sons, Inc., 5.6.1-5.6.37. <https://doi.org/10.1002/cpbi.3>
31. Nnyigide, O.S., Nnyigide, T.O., Lee, S.-G., & Hyun, K. (2022) Protein Repair and Analysis Server: A Web Server to Repair PDB Structures, Add Missing Heavy Atoms and Hydrogen Atoms, and Assign Secondary Structures by Amide Interactions. *J. Chem. Inf. Model.*, 62(17), 4232-4246.
32. Lindorff-Larsen, K., Piana, S., Palmo, K., Maragakis, P., Kepeis, J. L., Dror, R. O., and Shaw, D.E. (2010) Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins*, 78(8), 1950-1958. <https://doi.org/10.1002/prot.22711>
33. Abraham, M. J., Murtola, T., Schulz, R., Páll, S., Smith, J. C., Hess, B., & Lindahl, E. (2015) GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*, 1-2, 19-25. <https://doi.org/10.1016/j.softx.2015.06.001>
34. Price, D. J., & Brooks, C. L. III (2004) A modified TIP3P water potential for simulation with Ewald summation. *J. Chem. Phys.*, 121, 10096-10103. <https://doi.org/10.1063/1.1808117>
35. Bussi, G., Donadio, D., & Parrinello, M. (2007) Canonical sampling through velocity rescaling. *J. Chem. Phys.*, 126, 014101. <https://doi.org/10.1063/1.2408420>
36. Berendsen, H. J. C., Postma, J.P.M., van Gunsteren, W.F., DiNoia, A., & Haak, J. R. (1984) Molecular dynamics with coupling to an external bath. *J. Chem. Phys.*, 81, 3684-3690. <https://doi.org/10.1063/1.448118>
37. Parrinello, M. & Rahman, A. (1980) Crystal Structure and Pair Potentials: A Molecular-Dynamics Study. *Phys. Rev. Lett.*, 45, 1196. <https://doi.org/10.1103/PhysRevLett.45.1196>
38. Darden, T., York, D., & Pedersen, L. (1993) Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. *J. Chem. Phys.*, 98, 10089–10092. <https://doi.org/10.1063/1.464397>
39. Hess, B., Bekker, H., Berendsen, H. J. C., & Fraaije, J. G. E. M. (1997) LINCS: A linear constraint solver for molecular simulations. *J. Comp. Chem.*, 18(12), 1463-1472. [https://doi.org/10.1002/\(SICI\)1096-987X\(199709\)18:12%3C1463::AID-JCC4%3E3.0.CO;2-H](https://doi.org/10.1002/(SICI)1096-987X(199709)18:12%3C1463::AID-JCC4%3E3.0.CO;2-H)
40. Shi, Y., Xia, Z., Zhang, J., Best, R., Ponder, J. W., & Ren, P. (2013) Polarizable Atomic Multipole-Based AMOEBA Force Field for Proteins. *J. Chem. Theory Comput.*, 9, 4046-4063. <https://doi.org/10.1021/ct4003702>

-
41. Rackers, J. A., Wang, Z., Lu, C., Laury, M. L., Lagardère, L., Schneiders, M. J. Piquemal, J.-P., Ren, P., & Ponder, J.W. (2018) Tinker 8: Software Tools for Molecular Design. *J. Chem. Theory Comput.*, *14*, 5273-5289. <https://doi.org/10.1021/acs.jctc.8b00529>
42. Wang, Z. (2021) Tinker9: Next Generation of Tinker with GPU Support. Washington University in St. Louis, <https://github.com/TinkerTools/tinker9>
43. Tuckerman, M., Berne, B. J., & Martyna, G. J. (1992) Reversible multiple time scale molecular dynamics. *J. Chem. Phys.*, *97*(3), 1990-2001. <https://doi.org/10.1063/1.463137>