The Journal of Chemical Physics

RESEARCH ARTICLE | APRIL 16 2026

A comprehensive test of the AMOEBA force field using spectroscopy, structures, and simulations of nitrile protein environments

Special Collection: Festschrift in honor of Minhaeng Cho: Ultrafast Spectroscopy

Jacob M. Kirsh ; Jared Bryce Weaver ; Jacek Kozuch ; Steven G. Boxer

J. Chem. Phys. 164, 155102 (2026)
https://doi.org/10.1063/5.0318326

The Journal of Chemical Physics ARTICLE pubs.aip.org/aip/jcp

A comprehensive test of the AMOEBA force field using spectroscopy, structures, and simulations of nitrile protein environments

Cite as: J. Chem. Phys. 164, 155102 (2026); doi: 10.1063/5.0318326
Submitted: 19 December 2025 • Accepted: 24 March 2026 •
Published Online: 16 April 2026

Jacob M. Kirsh,[1,a] Jared Bryce Weaver,[1,b] Jacek Kozuch,[2,3,c] and Steven G. Boxer[1,c]

AFFILIATIONS

[1] Department of Chemistry, Stanford University, Stanford, California 94305-5012, USA
[2] Institute of Physical Chemistry and Theoretical Chemistry, BRICS–Braunschweig Integrated Centre of Systems Biology, Rebenring 56, 38106 Braunschweig, Germany
[3] Department of Physics, Freie Universität Berlin, D-14195 Berlin, Germany

Note: This paper is part of the Special Topic, Ultrafast Spectroscopy and Dynamics.
[a] **Present address:** Division of Chemistry and Chemical Engineering, California Institute of Technology, Pasadena, CA 91125, USA.
[b] **Present address:** Drug Discovery, Insitro, Inc. 279 E. Grand Ave., South San Francisco, California 94080, USA.
[c] **Authors to whom correspondence should be addressed:** jacek.kozuch@tu-braunschweig.de and sboxer@stanford.edu

ABSTRACT

Local noncovalent interactions, including hydrogen bonds (H-bonds), generate electric fields that are essential for biological assembly and function. We recently demonstrated that a nitrile's (–C≡N) infrared (IR) transition dipole moment and anomalous H-bond frequency blueshift can report on its environmental electric field and H-bond geometry and dynamics, respectively. Here, we expand on prior work with nitriles site-selectively incorporated into photoactive yellow protein by introducing mutations designed to alter nitrile H-bonding and local electrostatics. A comprehensive analysis combining IR data, high-resolution X-ray crystal structures, and extensive molecular dynamics simulations demonstrates that the multipolar, polarizable AMOEBA force field accurately models electrostatics and H-bond geometries in both fast and slow H-bond exchange regimes. This finding is reached by correlating experimentally and computationally derived –C≡N electric fields and H-bond blueshifts exhibiting different H-bond fluctuation timescales. This result implies that AMOEBA correctly models thermodynamic and kinetic aspects of noncovalent interactions. The diverse, thoroughly characterized collection of –C≡N protein environments reported herein provides a benchmark for next-generation molecular dynamics force fields that incorporate higher level descriptions of molecular electrostatics.

INTRODUCTION

Noncovalent interactions dictate the native conformations and dynamics adopted in condensed phase systems, and alterations to these interactions tune the efficiency of myriad chemical and biological processes. Accordingly, molecular dynamics (MD) simulations have become a prominent tool to explore noncovalent interactions.[1] At the same time, the accuracy of MD simulations is limited by the benchmarking methods used to obtain force field parameters, so direct comparison of MD noncovalent interactions against experimental assessments is crucial.[2–4] Vibrational spectroscopy is a powerful technique for this purpose because the frequency and intensity of molecular vibrations carry information on their local environment.[5–11] In particular, interpreting vibrational spectra within the framework of the vibrational Stark effect (VSE) enables assessments of local electric fields projected on a vibrational probe, and these electric fields provide a quantitative metric for the strengths of local noncovalent interactions.[5] As such, comparing experimental VSE-derived and computational electric fields has become an important strategy for assessing MD simulations,[12–15] as well as for exploring the role of electrostatics at interfaces[16,17] and in solvents and enzymes.[18–22]

Nitriles (–C≡N) have become one of the most popular VSE probes owing to their absorption in an uncluttered region of the infrared (IR) and relatively strong IR intensities.[9,10,23–26] However, previous MD benchmarking efforts utilizing them[27–29] were hampered by instances of specific nitrile–hydrogen-bond ("H-bond" or "HB") interactions, where bluer –C≡N frequencies are observed than would be predicted by the VSE.[30–33] Many empirically and physically based models have been proposed to describe this anomalous "H-bond blueshift" of –C≡N frequencies (see Sec. S17 in Ref. 30).[34–39] In particular, Cho's group has described –C≡N frequency tuning as arising from separate contributions of the local electric field, higher-order multipolar electrostatics, and Pauli repulsion, among other factors.[6,40–42] We have opted for a minimal, experimentally based expression with a term dependent on the electric field to first-order and an H-bond shift term that accounts for all non-field-based (and higher-order field) effects,

$$\begin{aligned}\tilde{\nu}_{\mathrm{obs}} &= \tilde{\nu}_0 - \Delta\vec{\mu}\cdot\vec{F} + \Delta\tilde{\nu}_{\mathrm{HB}} \\ &= \tilde{\nu}_0 + \Delta\tilde{\nu}_{\mathrm{non-HB}} + \Delta\tilde{\nu}_{\mathrm{HB}},\end{aligned} \tag{1}$$

where $\tilde{\nu}_{\mathrm{obs}}$ is the observed frequency; $\tilde{\nu}_0$ is the zero-field (gas phase) frequency; $\Delta\vec{\mu}$ is the vibrational Stark tuning rate, which determines the sensitivity of a vibrational transition to the local electric field, $\vec{F}$; $\Delta\tilde{\nu}_{\mathrm{non-HB}}$ is the frequency shift due to the VSE and is equal to $-\Delta\vec{\mu}\cdot\vec{F}$; and $\Delta\tilde{\nu}_{\mathrm{HB}}$ is the H-bond contribution to the frequency not captured by the linear, dipolar VSE.[30] By definition, $\Delta\tilde{\nu}_{\mathrm{HB}} = 0$ in non-H-bonding environments. Because of the separate contributions of $\Delta\tilde{\nu}_{\mathrm{non-HB}}$ and $\Delta\tilde{\nu}_{\mathrm{HB}}$, nitrile frequency shifts by themselves can be difficult to interpret, as confusing situations arise where the blueshift due to H-bonds counteracts redshifts due to environmental electric fields. To avoid this complication, early work on the connection between nitrile frequency shifts and electric fields was based only on non-H-bonding solvents.[43]

In a recent report, we found an alternative approach to determine nitrile electric fields.[30] We observed that the –C≡N transition dipole moment (TDM), $\vec{m}$, shows a linear dependence on the average MD solvent electric fields in H-bonding and non-H-bonding environments and is expressed as

$$\vec{m}\left(\vec{F}\right) = \vec{m}_0 - \underline{A}\cdot\vec{F}, \tag{2}$$

where $\underline{A}$ is the transition polarizability and $\vec{m}_0$ is the zero-field TDM.[30] Importantly, Eq. (2) describes that local electric fields are the origin for significant non-Condon effects (i.e., dependence of an oscillator's vibrational transition dipole on the coordinates of the surrounding environment[44,45]) for nitriles in both protic and aprotic environments;[42,44,46–48] such effects have often been assumed to be negligible.[49,50] Practically, it implies that changes to the –C≡N TDM (or, "TDM tuning") can be directly used to determine environmental electric fields. Furthermore, by inputting electric fields obtained by TDM tuning [Eq. (2)] into Eq. (1), $\Delta\tilde{\nu}_{\mathrm{HB}}$ can be isolated so that nitrile intensities and frequency shifts can report on both electric fields and H-bond shifts. In addition, $\Delta\tilde{\nu}_{\mathrm{HB}}$ has recently become a useful observable in and of itself:[38] we found that H-bond shifts are a function of nitrile–H-bond geometry and dynamics and derived an analytical expression that uses the H-bond distance and angle to predict the blueshift [Eq. (S1)]. Thus, the dual, independent tuning of –C≡N frequencies and TDMs enables two unique assessments of nitrile noncovalent interactions.

To investigate nitrile frequency and TDM tuning in a protein setting, we used amber suppression to place nitrile-bearing *o*-cyanophenylalanine (oCNF) at four diverse sites in photoactive yellow protein (PYP),[30] a 14 kDa protein from the PAS domain superfamily.[51] PYP is an attractive model system because atomic resolution X-ray crystal structures, which provide precise information on local environments, are readily obtained. In addition, it contains the chromophore *p*-coumaric acid, whose absorption provides an internal standard for absolute concentration determination, a necessary component to determine vibrational TDMs.[30] The incorporated nitriles experience very different electric fields and H-bonding geometries and dynamics as interpreted by their TDMs and H-bond blueshifts.[30] This information enabled a variety of experimental and computational tests. First, it allowed us to test the efficacy of –C≡N TDMs in proteins.[30] We found that TDMs reliably report on environmental electric fields in cases where frequencies cannot serve as field reporters due to offsetting red- and blueshifts. Furthermore, the TDM-derived fields were well-rationalized by crystal structures. Second, it enabled us to benchmark MD force fields for protein protic and aprotic environments, with a focus on testing the type of electrostatic description needed to recapitulate the PYP nitrile electric fields.[52] To investigate this, we compared MD simulations with the fixed-charge AMBER force field[53] (which describes electrostatics via monopoles) and the multipolar, polarizable AMOEBA force field[54] (which also includes dipoles, quadrupoles, and polarizability). We found that using AMOEBA is necessary to accurately model nitrile H-bonding environments and therefore nitrile electric fields, which is similar to the case of carbonyl probes in enzyme active sites.[12] The PYP blueshifts were then used to test the geometry-dependent model for $\Delta\tilde{\nu}_{\mathrm{HB}}$ in Eq. (S1), and predicted values successfully recapitulated the experimental results.[38] While making H-bond shift comparisons, we recognized that they can be a useful observable for benchmarking MD force fields beyond electric fields: blueshifts can be derived from a nitrile's average MD H-bond distance and angle, and the resulting value can be compared with the TDM-derived $\Delta\tilde{\nu}_{\mathrm{HB}}$.

These studies utilized four oCNF-containing PYP variants, which constitute an informative but small test set, so more thorough benchmarking of AMOEBA against nitrile-derived quantities is warranted. In the current study, we expand this set by creating a collection of 12 diverse nitrile H-bonding and electrostatic environments via mutagenesis. We systematically characterize the new –C≡N noncovalent interactions with a comprehensive analysis combining IR data, high-resolution X-ray crystal structures, and extensive MD simulations. Our characterization shows that the main effect of these mutations is to slow exchange between protic and aprotic populations of the nitrile probe, i.e., the nitriles experience heterogeneous settings with fast and slow dynamics. A new, extended TDM tuning method is developed to quantify subpopulation electric fields and H-bond shifts. This characterization allows us to more stringently test AMOEBA and thereby implicitly test its kinetic and thermodynamic modeling of noncovalent interactions. Using our extended collection, strong correlations are found for comparisons of experimental and MD-derived nitrile electric fields and H-bond shifts, highlighting the accuracy of the AMOEBA force field in simulating subtly to significantly different noncovalent interactions. At the same time, the detailed view of local noncovalent interactions that emerges from this study uniquely

allows us to investigate the limits of the TDM tuning method and the VSE approach broadly.

RESULTS

Variant and mutant selection

In our previous report,[30] four phenylalanine sites in PYP (positions 28, 62, 92, and 96) were replaced with oCNF to produce four variants, and the nitriles were found to experience a diverse set of electric field strengths. The nitrile in the F96oCNF PYP variant (F96oCNF henceforth) experienced the smallest electric field, so we chose it as our template for mutagenesis with the goal of designing new H-bonds or environments of varying polarity. For convenience, we use the term "variants" to refer to proteins with an F → oCNF replacement and reserve the term "mutants" for proteins with canonical mutations.

Three sites, M100, T103, and I49, were chosen for mutagenesis because they are the closest amino acids to oCNF96 [Fig. 1; distances shown in Fig. 1(b)]. Eight total mutations were made: T103V, I49T, M100D, M100E, M100H, M100K, M100N, and M100Q. All mutations except T103V were introduced with the goal to create a new H-bond to the –CN. T103V was selected to test whether the nitrile's electric field could be further decreased by making its environment even more hydrophobic. If a new –C≡N H-bond was not formed, we reasoned that I49T, M100N, and M100Q could still increase the polarity of the nitrile environment (and therefore the electric field projected on the –C≡N bond); M100K could introduce a positive charge to the environment; and M100D, M100E, and M100H could create polar or charged environments depending on their pK_a values. In the following sections, we provide a detailed description of the nitriles' local environments as interpreted from IR spectra, X-ray crystal structures, and AMOEBA simulations.

IR spectroscopy for F96oCNF and mutants indicates four nitrile environmental phenotypes in different chemical exchange regimes

Nitrile frequency and TDM Stark tuning rates $\Delta\vec{\mu}$ and $\underline{A}$, which collectively enable the extraction of electric fields and H-bond shifts using Eqs. (1) and (2), were previously obtained for the oCNF model compound *o*-tolunitrile (oTN).[30] To ensure these calibrations are suitable for incorporated oCNF, we obtained calibrations for a Boc-protected analog, N-Boc-2-cyano-L-phenylalanine (Fig. S1). We found that N-Boc-2-cyano-L-phenylalanine's calibrations (Fig. S4) are highly similar to oTN's (Fig. S5), indicating that oTN is an appropriate model compound. Here, we utilize parameters from the oTN calibration curves ($\tilde{\nu}_0 = 2231.4\ \mathrm{cm}^{-1}$; $|\Delta\vec{\mu}| = 0.19\ \frac{\mathrm{cm}^{-1}}{\frac{\mathrm{MV}}{\mathrm{cm}}}$; $|\vec{m}_0| = 0.0388\ \mathrm{D}$; $A = -0.000\ 92\ \frac{\mathrm{D}}{\frac{\mathrm{MV}}{\mathrm{cm}}}$) for consistency with our prior results.[30,38,52]

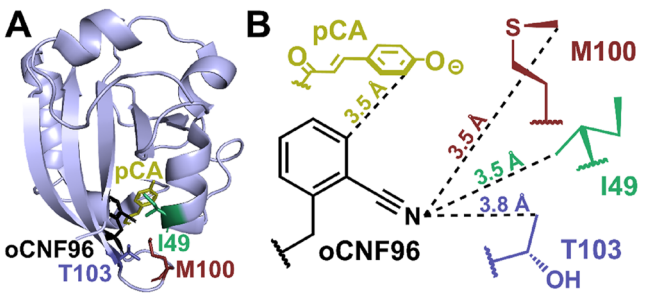

FIG. 1. Two perspectives of F96oCNF PYP. (a) Global view of the crystal structure (7SJJ[30]) in ribbon form. oCNF96 is shown in black, and the three residues mutated in this study, T103, I49, and M100, are shown in blue, green, and red, respectively. The chromophore *p*-coumaric acid (pCA) is shown in gold. The residues and chromophore are shown as sticks. (b) Local view of the nitrile environment. The closest contacts to the –C≡N N are carbons, consistent with the observation that the nitrile electric field is relatively small.[30]

Room temperature FTIR spectra were obtained for F96oCNF mutants at pH 6.0, matching conditions from the previous study.[30] The spectra for the mutants and F96oCNF are shown in Fig. 2(a), and the other variants' spectra are also shown for comparison. In analogy to the naming convention for the variants, we refer to the mutated F96oCNF PYPs by their mutation from this point forward. Spectra were also obtained at pH 8.5 for M100E, M100D, and M100H to interrogate the effect of potential protonation state changes on the nitrile environment. No changes in the –C≡N spectral shape were observed between the two pH values (Fig. S9). Given the proximity of M100 to oCNF96 in F96oCNF [Fig. 1(b)] and their exposure to solvent (*vide infra*), this invariance suggests that the protonation states of D100, E100, and H100 are the same at both pH values, i.e., D100 and E100 are deprotonated (anionic), and H100 is neutral.

The nine spectra at pH 6.0 for the nitriles in F96oCNF and mutants display four distinct lineshapes [Fig. 2(a); see also Fig. S7]. We relate these lineshapes to four environmental phenotypes experienced by the –C≡N, and these are denoted phenotype 1–4 (P1–P4) for convenience. P1 contains *F96oCNF* and T103V, P2 contains *I49T*, P3 contains *M100Q*, and P4 contains M100E, *M100D*, M100N, M100K, and M100H (IR spectra of nitriles grouped by phenotype are shown in Fig. S8). Throughout the text, the italicized PYPs are employed as phenotype representatives, and their spectra are overlaid in Fig. 2(b) to highlight their differences. P1- and P2-type nitriles possess spectra with a single band, albeit with different intensities. Although F96oCNF (and F28oCNF and F92oCNF) show a single symmetric IR band at room temperature, we previously found that this is not because the band represents a single type of environment: low temperature IR spectra showed two bands, indicating that the room temperature band reflects fast exchange between H-bonded and non-H-bonded nitrile populations.[52] The high spectral similarity for T103V and F96oCNF suggests that T103V also undergoes fast exchange. In the following, nitrile exchange regimes are treated in an approximately binary way for simplicity, i.e., they are treated as being in fast or slow exchange depending on their lineshape. We note that a more rigorous approach would be to assess exchange regimes by modeling lineshapes using Kubo theory,[55,56] as in Ref. 49. This approach was successfully used in our prior work,[38] where IR spectra from AMOEBA simulations recapitulated spectra for nitriles in solvents and proteins. Peak frequencies and full widths at half maxima (FWHMs) from band fitting of P1- and P2-type spectra are provided in Tables I and S6.

In contrast to P1- and P2-, P3- and P4-type nitriles manifest asymmetric lineshapes suggestive of multiple populations in slower exchange. Importantly, the lineshape asymmetry suggests that these nitriles represent new types of environments for benchmarking compared to the single-band PYP variants (Fig. 2). P3- and P4-type

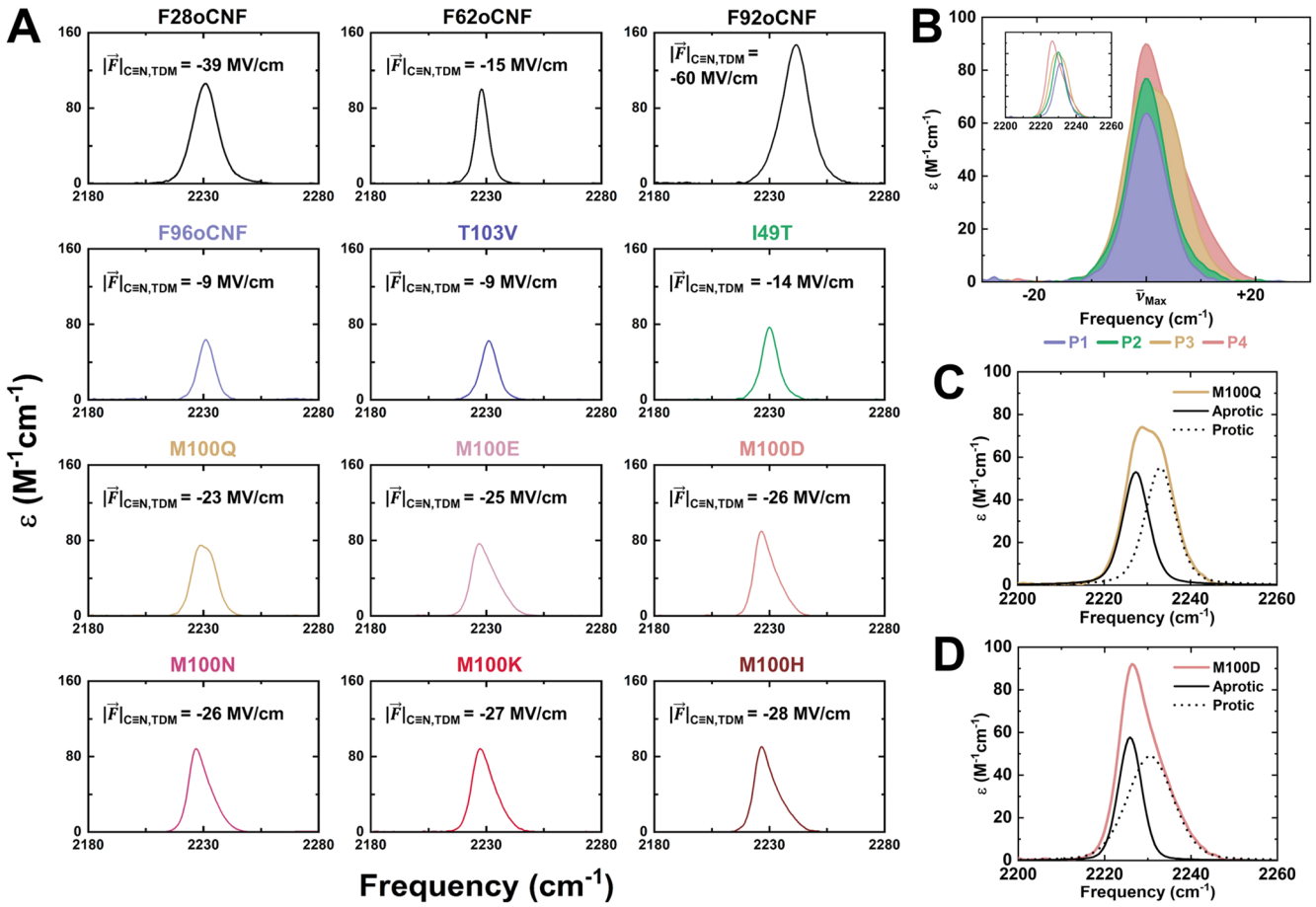

FIG. 2. Nitriles incorporated into PYPs demonstrate heterogeneity in intensity and lineshape. (a) Room temperature IR spectra of the –C≡N stretch for all variant[30] and mutant PYPs at pH 6.0. Spectra for F96oCNF and mutants yield four consistent IR lineshapes, termed environmental phenotypes. Phenotypes and members, with representatives italicized, are P1 (*F96oCNF* and T103V), P2 (*I49T*), P3 (*M100Q*), and P4 (M100E, *M100D*, M100N, M100K, and M100H). Species are colored by their phenotype (P1: blue; P2: green; P3: gold; and P4: red/pink). TDM-derived average environmental electric fields inferred using the oTN calibration curve are displayed,[30] and the complete set of extracted IR peak parameters for all nitrile-containing PYPs is provided in Table S6. (b) Shaded IR overlays (centered to $\bar{\nu}_{Max}$) of representatives of the four environmental phenotypes P1–P4. The inset shows non-centered spectra. In (c) and (d), two band fits are shown for P3 and P4 representative spectra, respectively. Raw spectra are shown in gold or pink, bands corresponding with non-H-bonding nitrile populations are shown as solid black lines, and bands corresponding with H-bonding nitrile populations are shown as black dotted lines. Additional P4 spectra are shown in Fig. S10. Spectra for F28oCNF, F62oCNF, F92oCNF, and F96oCNF in (a) and (b) are reproduced with permission from Weaver *et al.*, J. Am. Chem. Soc. **144**(17), 7562–7567 (2022). Copyright 2022 American Chemical Society.

TABLE I. Extracted IR peak parameters for the –C≡N stretch in environmental phenotype representatives.

Environment	$\tilde{\nu}_{obs}$ (cm^{-1})[a]	FWHM (cm^{-1})[a]	Total $\lvert\vec{m}\rvert$ (D)	Avg. $\lvert\vec{F}\rvert_{C\equiv N,TDM}$ (MV/cm)	$\Delta\tilde{\nu}_{non\text{-}HB,TDM}$ (cm^{-1})[b]	$\Delta\tilde{\nu}_{HB,TDM}$ (cm^{-1})
F96oCNF (P1)	2231.2[c]	8.4[c]	0.047 ± 0.002[c]	−9 ± 2[c]	−2[c]	2[c]
I49T (P2)	2230.2	8.3	0.052 ± 0.002	−14 ± 2	−2	2
M100Q (P3)	[d]	[d]	0.060 ± 0.002	−23 ± 2	[d]	[d]
M100D (P4)	[d]	[d]	0.063 ± 0.002	−26 ± 2	[d]	[d]

[a]Peak positions ($\tilde{\nu}_{obs}$) and FWHMs have an error of <0.1 cm^{-1}.
[b]Where $\Delta\tilde{\nu}_{non-HB} = -\Delta\vec{\mu}\cdot\vec{F}$, Eq. (1).
[c]Reproduced with permission from Weaver *et al.*, J. Am. Chem. Soc. **144**(17), 7562–7567 (2022). Copyright 2022 American Chemical Society.
[d]N/A since multiple bands are observed.

spectra were fit with two bands, shown in Figs. 2(c), 2(d), and S10, which adequately fit all observed spectra; the associated frequencies of the lower and higher energy bands and their FWHM values are provided in Table S8. The frequencies of the higher energy bands are comparable to or bluer than oTN's frequency in water (2231.7 cm^{-1}),[30] implying that the bands arise from nitrile H-bonding populations. Previous observations of multiple IR bands for a single nitrile in a protein were explained by the presence of both non-H-bonding and H-bonding populations in MD simulations.[29,57] Moreover, a nitrile on a peptide interacting with either the solvent or a hydrophobic protein pocket exhibited a highly similar spectral shape to the P4-type spectra.[58] These points together suggest that the heterogeneous P3- and P4-type spectra arise from mixtures of protic and aprotic nitrile populations with a slower exchange rate than that for F96oCNF.

Two bands were required to fit many of the mutants' IR spectra, but the process for extracting electric fields and H-bond shifts was previously developed only for single bands.[30] Therefore, we first apply our single-band analysis method using TDMs based on the integral of the observed spectra to obtain *averaged* environmental electric fields.[59] Later, we develop and apply a methodology for extracting electric fields and H-bond shifts when there are two populations. For all protein spectra in Fig. 2(a), peak areas were translated to TDMs,[30,60] and TDMs were converted to electric fields using Eq. (2). TDM-derived electric fields, referred to as $|\vec{F}|_{C\equiv N,TDM}$, are provided in Table I for the phenotype representatives and in Fig. 2(a) and Table S6 for all nitrile-containing PYPs. For P1- and P2-type spectra, electric fields were multiplied by the frequency Stark tuning rate to determine the $\Delta\tilde{\nu}_{non-HB}$ terms, i.e., the shifts due purely to the VSE [Eq. (1); Table I]. We then isolated the H-bond shifts, $\Delta\tilde{\nu}_{HB}$, by subtracting the zero-field frequency $\tilde{\nu}_0$ and the $\Delta\tilde{\nu}_{non-HB}$ values from the observed frequencies ($\tilde{\nu}_{obs}$) [Eq. (1); Tables I and S6; see Fig. 3 in Ref. 30 for a visualization of this procedure].

Across the entire collection of variants and mutants, we obtained TDMs between 0.047 and 0.094 D, where each TDM has an error of ±0.002 D. These correspond to a nitrile electric field range of −9 to −60 MV/cm, with errors of ±2 MV/cm. The F96oCNF mutants' fields span from −9 ± 2 to −28 ± 2 MV/cm. The smaller nitrile electric field range exhibited by the mutants compared with the variants [20 MV/cm compared with 50 MV/cm; Fig. 2(a)] implies that the mutants' nitrile environments are more hydrophilic than the parent F96oCNF, but the changes are more subtle than initially intended. Among P4, which possesses residues with sidechain charges ranging from −1 to +1, –C≡N electric fields values are the same within error (Table S6).

Overall, the extracted fields and H-bond shifts corroborate our assessment of the mutants based on their IR lineshapes. T103V (P1) has the same electric field and H-bond shift (2 cm^{-1}) as its progenitor F96oCNF, consistent with it being the only mutant not expected to have a more polar nitrile environment. In general, $\Delta\tilde{\nu}_{HB} \neq 0$ indicates the presence of –C≡N H-bonding, and values can be as large as 50 cm^{-1} depending on the nitrile H-bond geometry and dynamics.[38] In this case, T103V's small H-bond shift and single IR band indicate that it experiences fast H-bonding exchange dynamics.[38] In contrast to T103V, all other mutants possess more negative (i.e., larger magnitude) fields due to changes to the –C≡N environment (Table S6). For I49T (P2), $\Delta\tilde{\nu}_{HB}$ remains the same as for F96oCNF at 2 cm^{-1} (Table I), but the average electric field magnitude increases, indicating that the –C≡N experiences a more polar environment. As for the P1-type nitriles, I49T's small, nonzero H-bond shift implies that it undergoes fast exchange between protic and aprotic populations. P3- and P4-type nitriles have even more negative average electric fields than I49T, indicating they possess the most polar environments of the mutants; at the same time, their spectra possess two peaks, indicating more slowly exchanging H-bonding populations. In that case, all F96oCNF mutant nitriles have an underlying H-bonding population. To further investigate this finding and identify potential H-bond donors (which need to be known to obtain MD H-bond blueshifts), we turned to X-ray crystallography.

X-ray crystal structures suggest potential –C≡N H-bond donors

We obtained low temperature (100 K), high resolution (<1.25 Å) X-ray crystal structures for all eight F96oCNF mutants at pH 6.0 (Table S10) to enable interpretation of the environmental phenotypes from IR spectroscopy. In addition, these structures served as necessary starting points for MD simulations. Global structural alignment of F96oCNF and all mutants indicates that the tertiary structure is highly conserved (Fig. S11), which is reflected in the root mean square deviations all being <0.3 Å (Table S11). The structures enable us to compare the local nitrile environment in F96oCNF[30] with the environments in the mutants. Here, we present a comparison of the phenotype representatives' local environments (Fig. 3), while a comparison for F96oCNF and all mutants is provided in the supplementary material (Fig. S13). In both figures, interactions with the nearest potential nitrile H-bond donor are shown with cyan dashed lines.

F96oCNF and its mutants possess a single nitrile orientation with the oCNF plane essentially identical to that of F96 in wild type PYP.[61] In F96oCNF (P1), oCNF96 is in a predominately hydrophobic environment, where the closest contacts from the surrounding amino acids are all carbons [Fig. 1(b)]. The IR analysis shows that it has a nonzero $\Delta\tilde{\nu}_{HB}$ (Table I), and our previous work demonstrated that a value of $\Delta\tilde{\nu}_{HB} \neq 0$ indicates the presence of H-bonding.[38] The closest potential H-bond donor, a water termed Water 1, is ~5 Å from the nitrile [Fig. 3(a)], and this is also the case in T103V (Fig. S13). Like F96oCNF, I49T (P2) has a nonzero $\Delta\tilde{\nu}_{HB}$ value, and Water 1 remains at a similar distance from its nitrile [Fig. 3(b)]. In I49T, T49's sidechain hydroxyl points into the bulk solvent, suggesting the nitrile does not H-bond with T49. As such, the most likely P1- and P2-type nitrile H-bond donor in solution is water.

In M100Q (P3), Q100 demonstrates a broadened electron density at the location of the amide head group; as such, the density was fit with two conformers [Fig. 3(c)], both of which are 3.4 Å from the nitrile (Table S13). Electron density is observed for two waters at an O–O-distance of 2.1 Å, which is too short to correspond with a noncovalent interaction between them. We therefore modeled the waters as alternative positions of Water 1, and they can be matched to the Q100 conformers at Q100 $N_{\varepsilon 1}$–Water 1 H-bonding distances of 3.1 Å (Fig. S14). As Water 1 is 4.0 Å from the nitrile in one of the conformations, the structure suggests that M100Q's –C≡N H-bonds with Q100 and/or water in solution.

In the crystal structure for M100D (P4), D100 was fit with two conformers, and the major conformer (conformer A) faces away from the nitrile and points toward the surrounding bulk water

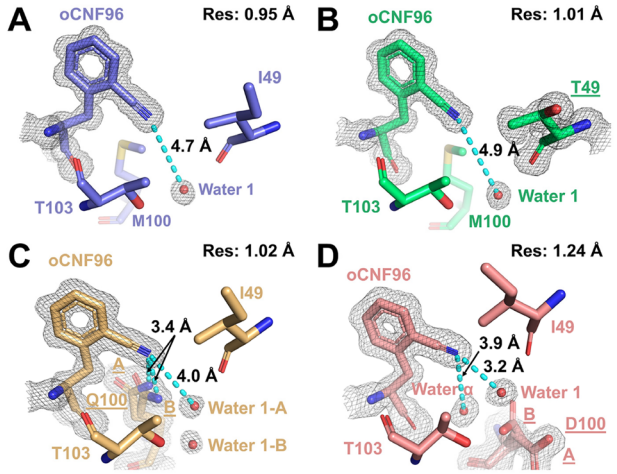

FIG. 3. Low temperature (100 K) crystal structures for environmental phenotypes. Panels display closeups of oCNF96 and local environments highlighting the potential H-bond donor(s) for the –C≡N in phenotype representatives, i.e., (a) F96oCNF (P1),[30] (b) I49T (P2), (c) M100Q (P3), and (d) M100D (P4). $2mF_o$-DF_c electron density maps contoured at 1σ are displayed for the nitriles, mutated residues, and relevant water(s). M100Q and M100D show multiple conformers for the mutated residue; occupancies can be found in Table S12. Labels for mutated amino acids are shown underlined, and structure resolutions are shown in the upper right corner. Contacts for closest potential nitrile–donor H-bond(s) are shown with cyan dashed lines; explicit N–donor distances and angles are provided in Table S13. An analogous comparison for F96oCNF and all mutants is shown in Fig. S13.

[Fig. 3(d)]. This opens up space near the nitrile such that both Water 1 and an additional water, Water α, are observed within 4.0 Å of the nitrile. Water α is also observed in two of the other four P4-type nitriles (Fig. S13), implying the mutations in P4 create a more solvent-accessible pocket in solution. Thus, M100D's structure suggests that water is the major –C≡N H-bond donor for P4-type nitriles rather than the newly introduced residues.

Overall, the crystal structures are consistent with the environmental phenotypes from the IR spectra. The P1 and P2 structures are highly analogous, while the P3 and P4 structures show larger deviations from the P1 nitrile local environment; these observations match the relative (dis)similarity of the phenotype IR spectra [Fig. 2(b)]. Moreover, water is observed closer to the nitrile in the P3 and P4 structures, consistent with those phenotypes experiencing larger average nitrile electric fields (Table I). At the same time, the structures imply all nitriles in F96oCNF or the mutants experience a mixture of hydrophobic and hydrophilic interactions, which is consistent with their electric field range being only 20 MV/cm (Table I). Next, we turned to MD simulations to characterize the nitriles' electric fields and H-bonding interactions for benchmarking; the simulations were also used to clarify H-bonding for the M100Q nitrile.

Molecular dynamics with the AMOEBA force field enable quantification of nitrile electric fields, H-bonding fractions, and H-bond distances and angles

MD simulations starting from crystal structures were performed with the AMOEBABIO18 force field[54] to produce 200 ns (eight repeats of 25 ns) aggregate trajectories for the F96oCNF mutants. Based on our findings from IR spectroscopy, M100E and M100D were run with anionic residues at position 100, while H100 in M100H was treated as net neutral (protonated at $N_{\delta 2}$—see discussion in the supplementary material, Sec. S3). In addition, 100 ns of MD were produced for the original PYP variants to supplement the previously procured 100 ns,[52] which brought each variant to a total of 200 ns. The electric fields from the protein MD simulations are referred to as $|\vec{F}|_{C\equiv N,MD}$, and field distributions for all nitrile-containing PYPs are shown as black traces in Fig. 4. These were calculated instead of IR spectra—which have been used for benchmarking elsewhere[38,49]—because we aim to directly compare computationally and experimentally derived electric fields.

The MD distributions of electric fields for F96oCNF and mutants are asymmetric or bimodal (Fig. 4), indicating that multiple populations are present. To better understand these underlying populations, each MD frame was classified as H-bonding or non-H-bonding with respect to the nitriles using a heavy-atom H-bond distance cutoff (N···X) of 4.0 Å and an H-bond cone (N···X–H) of 30°, as we did previously (Fig. S2),[52] and the resulting –C≡N H-bonding population fractions (f_{protic}) and non-H-bonding population fractions ($f_{aprotic}$) are provided in Table II. The two underlying populations are visualized in Fig. 4 as magenta and gray histograms for H-bonding and non-H-bonding environments, respectively. They separate into distinct electric field distributions with the H-bonding distributions centered at larger electric fields. We quantified the median electric fields for the non-H-bonding ($|\vec{F}|_{C\equiv N,aprotic,MD}$) and H-bonding ($|\vec{F}|_{C\equiv N,protic,MD}$) populations (Table II; shown as dashed lines in Fig. 4) and found that the P1-, P2-, and P4-type nitriles experience similar median values for both subpopulations. Furthermore, for P1-, P2-, and P4-type nitriles, the H-bonding fractions demonstrate more variability than the median field values (Table II). Indeed, the H-bonding fractions vary from 20% to 55%, a 2.8-fold change, while the analogous changes for the non-H-bonding and H-bonding median electric fields are 1.3-fold and 1.1-fold, respectively. In general, we find that P1-type nitriles have the smallest H-bonding fraction, followed by P2-type and then P4-type. The average TDM-derived fields increase along the same order (Table I), implying that the mutations primarily affect the average $|\vec{F}|_{C\equiv N,TDM}$ values by modulating the nitriles' H-bonding fractions. Using the median fields and H-bonding/non-H-bonding fractions, we obtained average MD electric fields for comparisons with average TDM-derived fields, and these fraction-weighted median $|\vec{F}|_{C\equiv N,MD}$ values are provided in Table II and illustrated as solid lines in Fig. 4.

Interestingly, M100K does not follow the MD H-bonding fraction and electric field trends just described. Instead, its H-bonding fraction is significantly smaller than that of any other P4-type nitrile (Table II). To understand this effect, we examined nitrile H-bond residence times in the simulations (supplementary material, Sec. 8). Our analysis found M100K's –C≡N engages in many more >1 ns H-bonds than the other P4-type nitriles (Table S19). We inspected M100K's MD trajectory and found several instances of long-lived K100–T103 H-bonds (Fig. S26). These H-bonds artificially reduce M100K's H-bonding fraction on the timescale of our MD simulations, so we exclude M100K from forthcoming correlations for benchmarking.

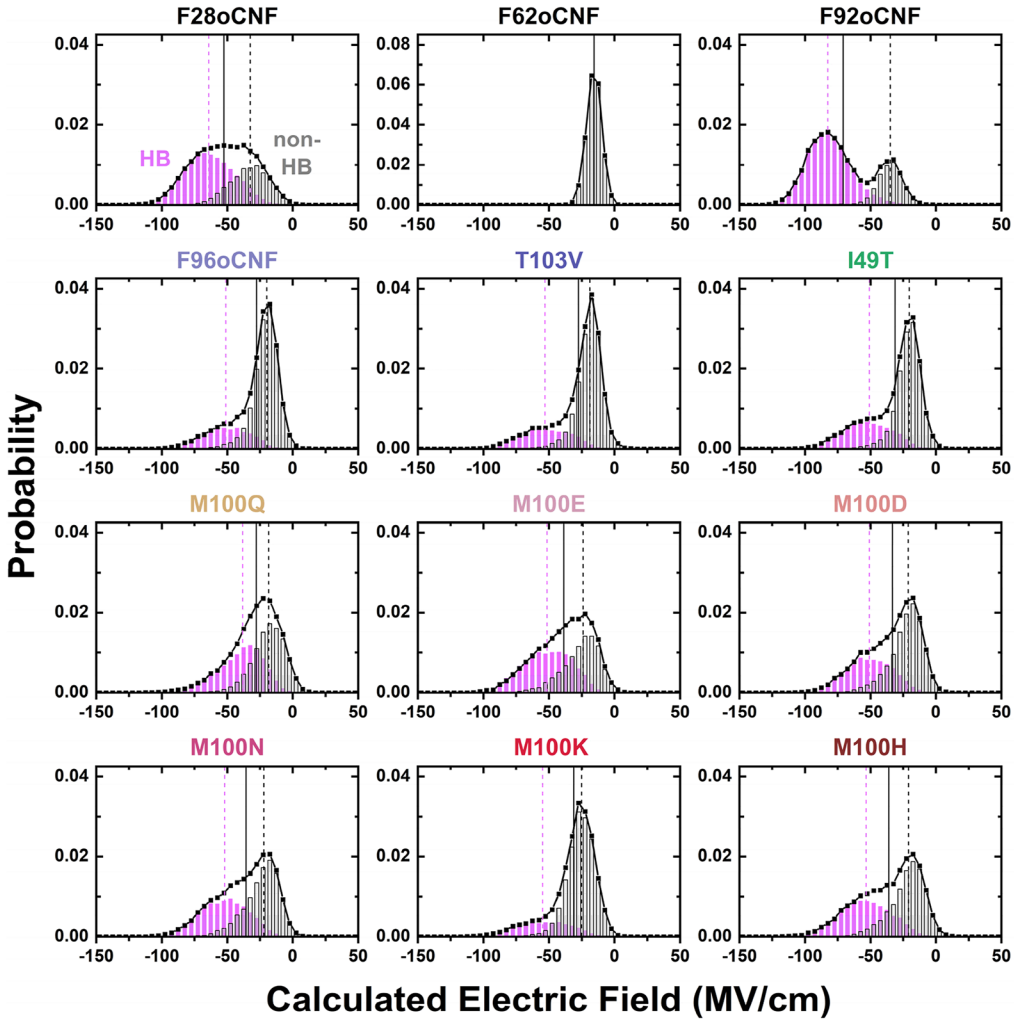

FIG. 4. MD electric field distributions for nitrile-incorporated PYPs. Simulations were run eight times at 25 ns each for 200 ns of data per protein. Electric field distributions (black traces) are decomposed into their H-bonding (magenta) and non-H-bonding populations (gray). Magenta and black dashed lines indicate the median H-bonding and non-H-bonding electric fields, respectively, and the solid lines indicate the fraction-weighted electric fields. Field distributions for M100E, M100D, and M100H come from simulations with E100 and D100 deprotonated and H100 protonated at $N_{\delta 2}$ (i.e., net neutral). M100K's distribution deviates significantly from the other P4-type distributions, causing this mutant not to be used for benchmarking.

Notably, the above analysis focuses on the P1-, P2-, and P4-type nitriles due to the similarity of their H-bonding and non-H-bonding median electric fields (Table II). In contrast, M100Q (P3) is the only mutant where the median H-bonding field is significantly smaller than that for the other phenotypes (Table II). To understand why this is the case, we identified the nitrile H-bond donors for all H-bonding frames, with particular interest in the distribution of protein and solvent H-bonds (Table S20). We found that all non-M100Q mutant nitriles predominately H-bond with the solvent, while M100Q's nitrile is split nearly 50:50 between solvent and protein H-bonds, with nearly all protein H-bonds coming from Q100. These findings are consistent with our qualitative assessment of potential –C≡N H-bond donors from the X-ray crystal structures (Fig. 3). By breaking down M100Q's H-bonding population into nitrile–water H-bonds and nitrile–Q100 H-bonds, we found that the median water $|\vec{F}|_{\mathrm{C{\equiv}N,protic,MD}}$ is similar to F96oCNF's median $|\vec{F}|_{\mathrm{C{\equiv}N,protic,MD}}$, while the median Q100 $|\vec{F}|_{\mathrm{C{\equiv}N,protic,MD}}$ is significantly smaller (compare Tables II and S21); this explains M100Q's overall smaller $|\vec{F}|_{\mathrm{C{\equiv}N,protic,MD}}$.

With the MD electric fields and H-bonding fraction/donor information in hand, we needed to obtain the average nitrile H-bond distances and angles to obtain computationally derived H-bond blueshifts [using Eq. (S1)] for benchmarking. These were procured by plotting the –C≡N H-bond distance/angle distributions for F96oCNF and mutants as contour plots (Figs. S31–S42) and fitting the plots with one or two Gaussians depending on the extent of

TABLE II. Calculated nitrile H-bonding/non-H-bonding fractions, median electric fields, fraction-weighted electric fields, and associated errors from averaging eight independent 25 ns MD trajectories.

PYP environment (phenotype)	H-bonding fraction f_{protic} (%)	Non-H-bonding fraction $f_{aprotic}$ (%)[a]	H-bonding median $\lvert\vec{F}\rvert_{C\equiv N,protic,MD}$ (MV/cm)	Non-H-bonding median $\lvert\vec{F}\rvert_{C\equiv N,aprotic,MD}$ (MV/cm)	Fraction-weighted $\lvert\vec{F}\rvert_{C\equiv N,MD}$ (MV/cm)[b]
F28oCNF	63.6 ± 5.7	36.4	−64.0 ± 1.5	−32.5 ± 2.4	−52.6 ± 0.5
F62oCNF	0	100	N/A	−15.7 ± 1.0	−15.7 ± 1.0
F92oCNF	75.1 ± 8.5	24.9	−82.4 ± 4.1	−34.9 ± 1.9	−70.7 ± 5.5
F96oCNF (P1)	25.0 ± 8.2	75.0	−51.0 ± 2.4	−19.8 ± 0.9	−27.5 ± 2.5
F96oCNF T103V (P1)	25.7 ± 12.7	74.3	−52.9 ± 6.3	−18.8 ± 1.7	−27.6 ± 5.8
F96oCNF I49T (P2)	33.5 ± 16.5	66.5	−50.9 ± 5.0	−20.4 ± 1.5	−31.2 ± 6.8
F96oCNF M100Q (P3)	48.4 ± 12.5	51.6	−38.1 ± 7.0	−18.2 ± 4.7	−27.8 ± 7.5
F96oCNF M100E (P4)	54.3 ± 7.8	45.7	−51.4 ± 5.7	−24.1 ± 2.4	−38.8 ± 5.0
F96oCNF M100D (P4)	40.5 ± 6.5	59.5	−50.8 ± 2.1	−21.1 ± 2.1	−33.1 ± 3.8
F96oCNF M100N (P4)	45.3 ± 3.6	54.7	−52.0 ± 3.0	−22.0 ± 1.6	−35.6 ± 3.2
F96oCNF M100K (P4)	20.5 ± 10.9	79.5	−54.8 ± 6.3	−25.1 ± 2.3	−31.1 ± 3.3
F96oCNF M100H (P4)	45.8 ± 11.3	54.2	−53.2 ± 4.9	−20.8 ± 3.2	−35.9 ± 7.4

[a]Error same as for the H-bonding fraction.
[b]Sum of the products of the H-bonding/non-H-bonding fractions and the H-bonding/non-H-bonding median $\lvert\vec{F}\rvert_{C\equiv N,MD}$ values.

protein H-bonding (Table S23). All contour plots were well-fit with this scheme, and the average distances and angles are provided in Tables S23–S25. Below, we use the cumulative MD information from this section for our benchmarking efforts. Prior to this, we develop and apply a method to extract nitrile electric fields and H-bond shifts when two IR bands are present.

Development and application of an extended TDM tuning method for two IR bands

Analysis of a single nitrile IR band, as demonstrated above for P1- and P2-type nitriles, is straightforward because two observables, the nitrile's TDM $|\vec{m}|$ and frequency $\tilde{\nu}_{obs}$, can be mapped onto two unknowns, electric field $|\vec{F}|$ and H-bond shift $\Delta\tilde{\nu}_{HB}$, using Eqs. (1) and (2). However, if a single nitrile splits into two or more bands, as for the P3-/P4-type nitriles, then multiple sets of two unknowns must be determined. This situation is more complicated because we need to know additional information—such as the species' population fractions—to determine their TDMs. Relative fractions are often estimated from IR peak intensity ratios,[62] but –C≡N TDM tuning [Eq. (2)] compromises this approach since protic and aprotic environments have very different electric fields and associated TDMs.[5,30,52] Furthermore, because the IR spectra cannot provide the fractions, the system is underdetermined by one observable for two IR bands. To circumvent this problem, an alternative approach was recently developed that relied upon comparing nitrile intensities across different photoproduct states in a phytochrome at different temperatures.[59] However, as we sought a generally applicable two-band TDM tuning method, we did not pursue this approach here.

We opted to develop a method that depends on explicit knowledge of the population fractions, i.e., f_{protic} and $f_{aprotic}$ (Fig. 5). This is because once the population fractions are fixed, the parameters for each IR band become fully determined and mutually independent. Given that H-bonds often exchange on the (sub)-picosecond timescale, an appropriate method to directly, experimentally determine these fractions is 2D IR spectroscopy.[63–65] In the absence of these data, we used f_{protic} and $f_{aprotic}$ from the AMOEBA MD simulations. With these values in hand, the non-H-bonding and H-bonding IR bands are scaled by the reciprocal of f_{protic} and $f_{aprotic}$, respectively, yielding the bands that would result if either population were the only one present [Fig. 5(a)]. Then, the TDMs of these scaled bands are obtained by integrating them.[30,60] These TDMs are referred to as $|\vec{m}|_{100\%,protic}$ or $|\vec{m}|_{100\%,aprotic}$ [Fig. 5(a)]. Once the $|\vec{m}|_{100\%}$ values are determined, the process is the same as detailed above: electric fields ($|\vec{F}|_{protic,TDM}$ or $|\vec{F}|_{aprotic,TDM}$) are inferred from the TDMs using the oTN field-TDM calibration[30] [Fig. 5(b)], which are used to infer $\tilde{\nu}_0 + \Delta\tilde{\nu}_{non\text{-}HB,protic/aprotic}$ [Fig. 5(c)]. Then, by taking the difference of $\tilde{\nu}_{obs,protic/aprotic}$ and these values, $\Delta\tilde{\nu}_{HB,protic/aprotic}$ is obtained [Fig. 5(c)]. Significantly, this approach does not depend on the origin of f_{protic} and $f_{aprotic}$. Therefore, the method in Fig. 5 is completely general, and it may be used for cases where more than two nitrile IR bands are present[57] so long as the fractions can be determined.

To test this method, we compared TDM- and MD-derived electric fields for oTN in methanol, since we previously found that the experimental IR spectrum exhibits two bands [see the inset of Fig. 5(a)], and AMOEBA MD simulations showed that oTN's –C≡N's low and high energy IR bands represent non-H-bonding and H-bonding populations, respectively.[38] In addition, the average TDM-derived field using the oTN calibration,[30] −41 ± 1 MV/cm, and the average MD electric field, −38 MV/cm (Table S26), are in close agreement. Using f_{protic} = 22.0% and $f_{aprotic}$ = 78.0% from MD simulations (Table S26), we obtained $|\vec{m}|_{100\%,aprotic}$ = 0.074 ± 0.001 D and $|\vec{m}|_{100\%,protic}$ = 0.093 ± 0.001 D, resulting in $|\vec{F}|_{aprotic,TDM}$ = −39 ± 1 MV/cm and $|\vec{F}|_{aprotic,TDM}$ = −59 ± 1 MV/cm

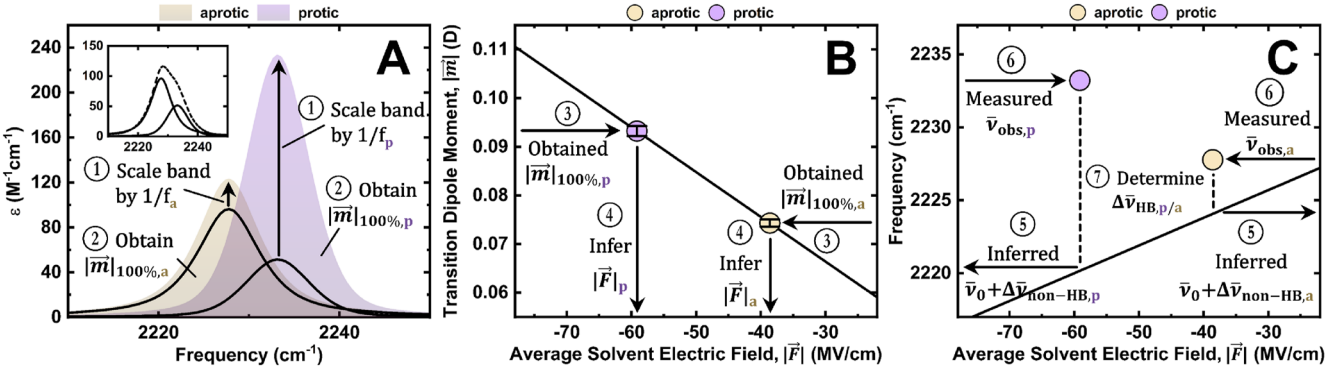

FIG. 5. Procedure to extract electric fields and H-bond shifts from nitrile IR spectra with two bands. (a) The IR spectrum for oTN's –C≡N in methanol[38] (inset dashed line) is fit with two bands (inset solid lines), which are assigned to protic (**p**) and aprotic (**a**) nitrile populations. These bands are scaled by the reciprocal of the respective protic (f_{protic}) and aprotic ($f_{aprotic}$) fractions from AMOEBA MD simulations (1). The integrals of the scaled bands are used to determine the transition dipole moments[60] that would result if the protic or aprotic populations were the only one present in solution (2); these values are referred to as $|\vec{m}|_{100\%,protic}$ and $|\vec{m}|_{100\%,aprotic}$, respectively. (b) Using $|\vec{m}|_{100\%,protic}$ and $|\vec{m}|_{100\%,aprotic}$ (3), the oTN TDM vs field calibration curve[30] enables inference of $|\vec{F}|_{protic,TDM}$ and $|\vec{F}|_{aprotic,TDM}$, respectively (4). (c) Plugging $|\vec{F}|_{protic,TDM}$ and $|\vec{F}|_{aprotic,TDM}$ into the oTN aprotic field–frequency curve,[30] $\bar{\nu}_0 + \Delta\bar{\nu}_{non\text{-}HB,protic,TDM}$ and $\bar{\nu}_0 + \Delta\bar{\nu}_{non\text{-}HB,aprotic,TDM}$ can be inferred [Eq. (1)] (5), and by using the measured $\bar{\nu}_{obs,protic}$ and $\bar{\nu}_{obs,aprotic}$ values (6), the H-bond blueshifts $\Delta\bar{\nu}_{HB,protic,TDM}$ and $\Delta\bar{\nu}_{HB,aprotic,TDM}$ are determined [Eq. (2)] (7) (Table S27). The spectrum of oTN in methanol in (a) is reproduced with permission from J. M. Kirsh and J. Kozuch, JACS Au **4**(12), 4844–4855 (2024). Copyright 2024 Author(s), licensed under a Creative Commons Attribution 4.0 License.

using the procedure in Fig. 5 (Table S27). From these fields, we can deduce that $\Delta\bar{\nu}_{HB,protic,TDM}$ = 13 ± 1 cm^{-1} and $\Delta\bar{\nu}_{HB,aprotic,TDM}$ = 4 ± 1 cm^{-1} according to Fig. 5(c) (Table S27).

To assess the efficacy of the two-band TDM tuning method for oTN in methanol, we compared the electric fields and H-bond shifts derived using the procedure in Fig. 5 with the corresponding AMOEBA values. We previously found that blueshifts for nitriles H-bonding with solvent molecules are approximately half the Eq. (S1) value due to motional narrowing from fast fluctuations between H-bonding and non-H-bonding states,[38] and this approach is used here and in benchmarking below. To summarize the comparisons, we found that the TDM-derived electric fields are in reasonable agreement ($|\vec{F}|_{aprotic,MD}$ = −30 MV/cm and $|\vec{F}|_{protic,MD}$ = −68 MV/cm), $\Delta\bar{\nu}_{HB,aprotic,TDM}$ is in decent agreement ($\Delta\bar{\nu}_{HB,aprotic,MD}$ = 0 cm^{-1}), and $\Delta\bar{\nu}_{HB,protic,TDM}$ is in excellent agreement ($\Delta\bar{\nu}_{HB,protic,MD}$ = 12 cm^{-1}) (see Sec. S9 for further discussion). We investigated potential origins of uncertainty for values derived from the extended TDM tuning method and found that the main source for oTN in methanol comes from the nitrile bands' unscaled areas, which is likely a consequence of the lack of clear separation between the –C≡N's protic and aprotic bands (Sec. S9). As the P3- and P4-type nitriles have similar frequency gaps between their bands (Table S8), we expect band fitting to be the limiting factor in those cases as well. Overall, the two-band TDM tuning method produces reasonable to excellent values when the AMOEBA H-bonding and non-H-bonding fractions are used, implying that the approach outlined in Fig. 5 can be used for benchmarking the P3- and P4-type nitriles.

Using the unscaled band areas from P3- and P4-type nitrile IR spectral fits (Fig. S10) and the MD-based $f_{protic}/f_{aprotic}$ from Table II, we derived IR peak parameters for each underlying population; the results for P3 and P4 representatives are given in Table III and for the remaining P4-type nitriles in Tables S29 and S30. The electric field difference for M100Q's protic and aprotic populations (−6 ± 4 MV/cm) is smaller than that for M100D (−29 ± 4 MV/cm), mirroring the trend observed for the MD values (Table II). Moreover, $\Delta\bar{\nu}_{HB,aprotic,TDM}$ is as close to 0 cm^{-1} for both phenotypes (0 ± 1 and −3 ± 1 cm^{-1} for P3 and P4, respectively) as it was for oTN in methanol (4 ± 1 cm^{-1}), so the values derived here are useful for benchmarking.

TABLE III. Extracted IR peak parameters for P3- and P4-type nitrile representatives.

Environment	$\|\vec{m}\|_{100\%}$ (D)	$\|\vec{F}\|_{C\equiv N,TDM}$ (MV/cm)	$\Delta\bar{\nu}_{non\text{-}HB,TDM}$ (cm^{-1})	$\Delta\bar{\nu}_{HB,TDM}$ (cm^{-1})
$M100Q_a$ ($P3_{aprotic}$)	0.059 ± 0.002	−22 ± 3	−4 ± 1	0 ± 1
$M100Q_p$ ($P3_{protic}$)	0.064 ± 0.002	−28 ± 2	−5 ± 1	7 ± 1
$M100D_a$ ($P4_{aprotic}$)	0.053 ± 0.002	−15 ± 2	−3 ± 1	−3 ± 1
$M100D_p$ ($P4_{protic}$)	0.079 ± 0.003	−44 ± 3	−8 ± 1	8 ± 1

DISCUSSION

To assess AMOEBA's description of local noncovalent interactions, we compare TDM- and MD-derived $|\vec{F}|$ values and H-bond blueshifts for the 12 nitrile-containing PYPs. These comparisons are broken down into four categories: comparisons of average electric fields, which can be performed independent of the number of IR bands; comparisons of H-bond blueshifts for single IR bands; comparisons of individual fields for nitrile subpopulations using the procedure in Fig. 5; and comparisons of blueshifts for protic nitrile subpopulations using the same procedure. We first compare the electric fields and H-bond blueshifts from the single-band TDM tuning analysis. The electric field comparisons directly extend upon our prior benchmarking study.[52] Then, we correlate fields and blueshifts derived from the extended TDM tuning method. These test the modeling of individual nitrile protic and aprotic environments for the first time, enabling the most comprehensive VSE-based assessment of AMOEBA with this approach to date. Together, these comparisons are a much more stringent test of AMOEBA than comparing the average fields alone, as we did previously[52] (see Table S31).

Correlation of average TDM- and MD-derived electric fields and H-bond blueshifts

As described above, total –C≡N IR band areas can be translated into average $|\vec{F}|_{C\equiv N,TDM}$ values irrespective of band (a)symmetry (Fig. 2 and Table S6). For benchmarking, these values correspond to the averaged MD electric fields $|\vec{F}|_{C\equiv N,MD}$, which represent fraction-weighted convolutions of all nitrile environments (Table II and solid lines in Fig. 4). Figure 6(a) shows the correlation between these experimental and computational electric field values for all nitrile-containing PYPs. Upon inspection, we find that the data points cluster in four regions of the plot where $|\vec{F}|_{C\equiv N,TDM}$ is roughly −15, −25, −40, and −60 MV/cm. From low to high field, the clusters are comprised of hydrophobic F62oCNF and the P1-/P2-type nitriles, the P3-/P4-type nitriles, F28oCNF, which H-bonds with solvent, and F92oCNF, which H-bonds with the protein.[30,52] Importantly, only the F62oCNF point stands for a homogeneous environment, while the rest represent environments undergoing H-bonding exchange at room temperature.

A linear fit weighting errors in both dimensions was obtained to quantify the relationship between the electric fields.[52] In the case of perfect agreement, a linear fit to the values would have a slope of unity. In line with our previous finding using the original four PYP variants,[52] we observe a strong linear correlation between experimentally and computationally derived electric fields for the extended collection of PYP variants/mutants, as indicated by the R^2 value of 0.84. The regression provides a best fit line of $|\vec{F}|_{C\equiv N,MD} = (1.16 \pm 0.09)\ |\vec{F}|_{C\equiv N,TDM} - (5.8 \pm 2.5)$, which is nearly identical to the regression line we previously obtained for the limited set of original variants $\{|\vec{F}|_{C\equiv N,MD} = (1.17 \pm 0.09)\ |\vec{F}|_{C\equiv N,TDM} - (5.3 \pm 2.7)\}$. Both lines possess nonzero offsets as a consequence of subtle differences in the MD parameters for oTN, which is used for TDM-field calibrations, and the protein-bound oCNF amino acid (see Ref. 52 for further discussion).

The similarity of the lines, particularly with respect to the slope, has important implications: AMOEBA's ability to recapitulate electric fields for the mutants, which introduced new types of environments with more complex lineshapes and H-bond exchange dynamics, strengthens our earlier argument that its sophisticated electrostatic description makes it uniquely versatile.[52] A similar conclusion was also reached by Bradshaw *et al.* when they tested electric fields for a small protein test set using AMOEBA, AMBER, and CHARMM force fields.[69,70] Now, using our more comprehensive collection of mutants and variants, this conclusion regarding AMOEBA's modeling accuracy can be made more confidently.

For additional benchmarking, we turned to –C≡N H-bond blueshifts, which can be quantified from TDM and frequency tuning.[30] The H-bond shift is a metric for H-bond distance, angle, and dynamics,[38] so comparing experimental and MD-derived shifts implicitly tests how well AMOEBA models the kinetic and thermodynamic aspects of nitrile–environment interactions. Accordingly, we compared the experimentally derived blueshifts for single IR bands, $\Delta\tilde{\nu}_{HB,TDM}$ (Table S6), against the average MD blueshifts. In general, experimental nitrile blueshifts are directly dependent on the average H-bond distance and angle according to Eq. (S1).[38] As such, Eq. (S1) was used to determine computationally derived $\Delta\tilde{\nu}_{HB,MD}$ values based on the average nitrile MD H-bond geometry of the protic fraction (Tables S23 and S25). If this H-bonding population ($\Delta\tilde{\nu}_{HB} \neq 0$) shows extensive fluctuations and/or exchanges rapidly with a non-H-bonding population ($\Delta\tilde{\nu}_{HB} = 0$), the experimental IR spectrum will be averaged due to motional narrowing. As both scenarios can occur for the PYP nitriles, we sought approaches to account for them. First, if the H-bonding population fluctuates rapidly relative to IR timescales and the on- and off-rates are equal, experimentally derived H-bond shifts take on half of their original value, i.e., the average of $\Delta\tilde{\nu}_{HB}$ and 0 cm^{-1}.[38] We previously found that treating the rates as equal effectively describes the limiting case of rapid chemical exchange for model molecule oTN in methanol or water.[38] As such, we treat MD solvent–nitrile H-bonds as being in rapid exchange, i.e., values from Eq. (S1) are halved. In addition, we treat protein–nitrile H-bonds as slowly exchanging, i.e., no averaging is applied: this is because AMOEBA simulations showed that these H-bonds have relatively longer residence times, and this approach was previously used to successfully model MD-based H-bond shifts.[38,52] Finally, we account for rapid exchange between H-bonding and non-H-bonding populations in MD simulations by fraction-weighting all H-bond shift values, i.e., by multiplying full or halved values from Eq. (S1) by f_{protic}. We note that accounting for both effects is important as we used an H-bond cutoff of <4.0 Å, which captures the entire first (fluctuating) H-bonding coordination sphere observed in nitrile radial distribution functions (see Fig. S3 in Ref. 52).

This overall approach was applied to P1- and P2-type nitriles to obtain $\Delta\tilde{\nu}_{HB,MD}$ values. These values are compiled in Table S32 along with the experimentally derived blueshifts. The correlation of experimental vs MD-derived H-bond blueshifts for nitriles with single IR bands is provided in Fig. 6(b). The H-bond shifts increase in the same order as the average fields in Fig. 6(a), demonstrating the close connection between nitrile H-bonding characteristics and electrostatic environment. Fitting the data gives an excellent linear relationship: $\Delta\tilde{\nu}_{HB,MD} = (0.96 \pm 0.03)\ \Delta\tilde{\nu}_{HB,TDM} + (1.0 \pm 0.3)$, with $R^2 = 0.98$. This demonstrates that AMOEBA

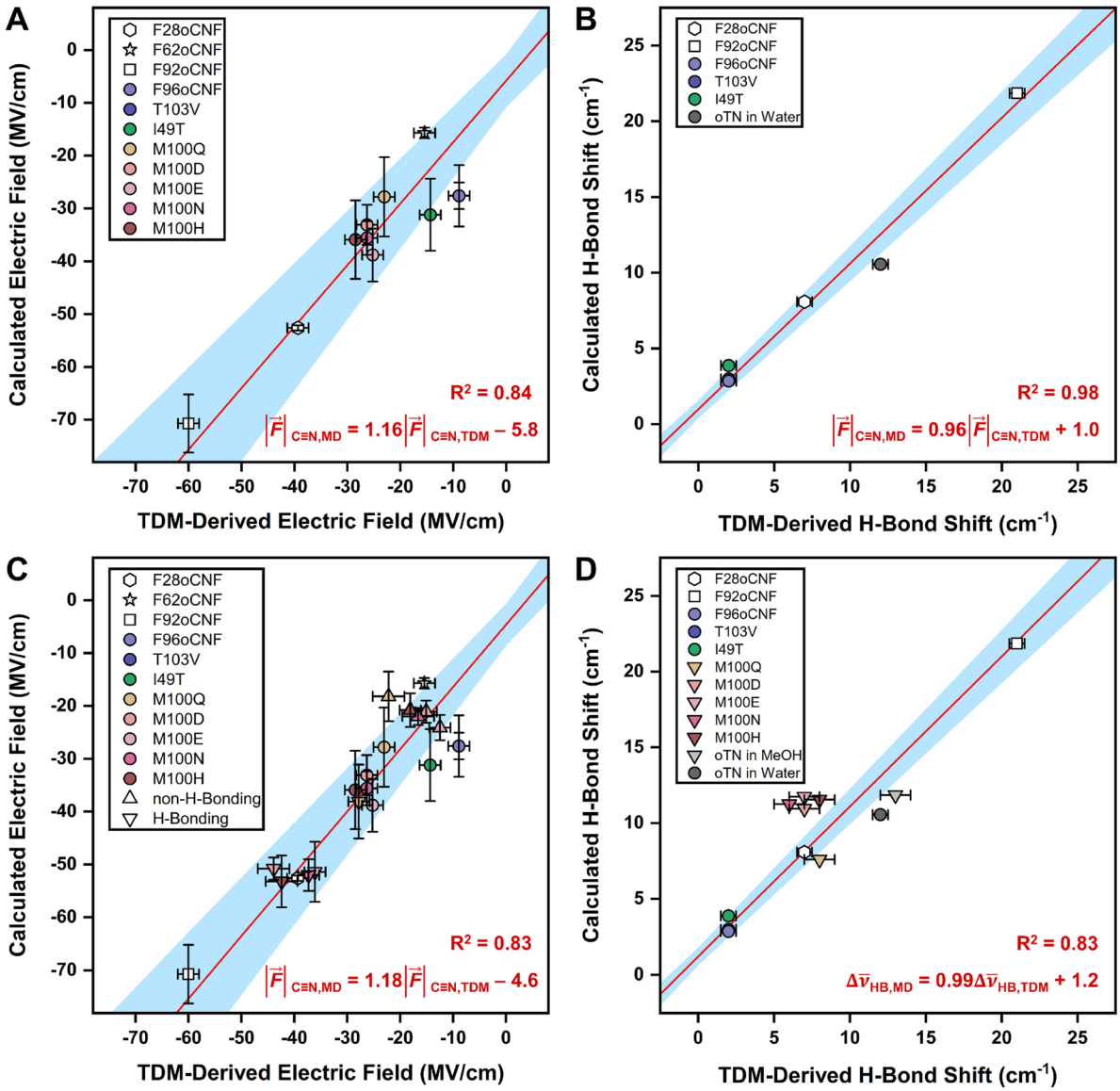

FIG. 6. Strong correlations between (a) average $|\vec{F}|_{C\equiv N,TDM}$ and fraction-weighted $|\vec{F}|_{C\equiv N,MD}$ values, (b) average $\Delta\bar{\nu}_{HB,TDM}$ and $\Delta\bar{\nu}_{HB,MD}$ values, (c) TDM- and MD-derived $|\vec{F}|_{C\equiv N,aprotic}$ and $|\vec{F}|_{C\equiv N,protic}$ values, and (d) protic $\Delta\bar{\nu}_{HB,TDM}$ and $\Delta\bar{\nu}_{HB,MD}$ values for nitrile-containing PYPs indicate AMOEBA accurately models diverse protein environments. In (a) and (c), $|\vec{F}|_{C\equiv N,TDM}$ values are derived from the oTN AMOEBA electric field-TDM calibration,[30] with the method for extracting values for two IR bands provided in Fig. 5; $|\vec{F}|_{C\equiv N,MD}$ values come from MD simulations of the proteins (Table II). In (b) and (d), $\Delta\bar{\nu}_{HB,TDM}$ values are obtained from the difference in the observed –C≡N peak frequency and $\Delta\bar{\nu}_{non\text{-}HB,TDM}$ (see Fig. 5 for an illustration); $\Delta\bar{\nu}_{HB,MD}$ values come from inputting the average MD H-bond distance and angle into Eq. (S1) and accounting for H-bond donor type and band averaging. All experimentally and computationally derived blueshifts can be found in Table S32. Triangular points are for individual nitrile subpopulations, while non-triangular points represent average nitrile populations. Coefficients of determination and error-weighted linear regressions[66,67] are shown in red in with $R^2 = 0.98$ and $|\vec{F}|_{C\equiv N,MD} = (1.16 \pm 0.09)\,|\vec{F}|_{C\equiv N,TDM} - (5.8 \pm 2.5)$ in (a), $R^2 = 0.83$ and $|\vec{F}|_{C\equiv N,MD} = (1.18 \pm 0.07)\,|\vec{F}|_{C\equiv N,TDM} - (4.6 \pm 2.0)$ in (c), $R^2 = 0.98$ and $\Delta\bar{\nu}_{HB,MD} = (0.96 \pm 0.03)\,\Delta\bar{\nu}_{HB,TDM} + (1.0 \pm 0.3)$ in (b), and $R^2 = 0.83$ and $\Delta\bar{\nu}_{HB,MD} = (0.99 \pm 0.03)\,\Delta\bar{\nu}_{HB,TDM} + (1.2 \pm 0.3)$ in (d); the light blue regions indicate 2σ confidence intervals. The constant offsets arise from the challenges of evaluating the self-field[68] for protein-incorporated oCNF and/or our oCNF parameterization method (see Sec. S1 in Ref. 52 for details). For all fits, perfect MD recapitulation of the TDM-derived values would result in a best fit line with unit slope. Note that the point for T103V is covered by F96oCNF in all panels. Errors of ±0.5 and ±1 cm^{-1} were used for TDM-derived H-bond shifts from one- and two-band TDM analysis, respectively (Sec. S11).

can effectively recapitulate nitrile H-bonding fractions and average angles and distances for the limiting case of fast H-bond exchange, which was observed spectroscopically as P1- and P2-type nitrile environments.

Correlation of electric fields and H-bond shifts for nitriles with two IR bands

The P3- and P4-type nitriles show asymmetric band shapes due to individual H-bonding and non-H-bonding populations in slower chemical exchange. These cases provide the opportunity to test the application of the dual TDM and frequency tuning method for complex cases with overlapping –C≡N bands. This provides an even more stringent test of AMOEBA by expanding our benchmarking to explicitly include populations with relatively slow H-bonding dynamics. Moreover, it enables a test of subpopulation fields and H-bond shifts rather than averaged values for the first time.

As described in the section titled "Results", we determined the individual TDM- and MD-derived average electric fields for the H-bonding and non-H-bonding subpopulations of nitriles exhibiting two IR bands (Tables II, S29, and S30). A comparison of all experiment- and simulation-based fields, i.e., the ones from both the single- and two-band analyses, is shown in Fig. 6(c). We find that the P3- and P4-type nitrile data points, which originally clustered around −25 MV/cm [Fig. 6(a)], now separate into the hydrophobic cluster at −15 MV/cm and the cluster for nitriles H-bonded to solvent at −40 MV/cm. This latter partitioning agrees with crystallographic evidence that these nitriles are exposed to bulk water (Figs. 3 and S13). When we apply a linear fit, we once again obtain a very good correlation with $R^2 = 0.83$ and a best fit line of $\left|\vec{F}\right|_{C\equiv N,MD} = (1.18 \pm 0.07)\left|\vec{F}\right|_{C\equiv N,TDM} - (4.6 \pm 2.0)$. Significantly, the correlation and fit remain effectively unchanged from the case using average fields alone in Fig. 6(a). Since the procedure in Fig. 5 required the experimental band shapes to be scaled by the MD H-bonding/non-H-bonding fractions, the invariance of the fitting parameters implies that AMOEBA properly models subpopulations which are in the slow H-bond exchange regime. It relatedly implies that hundreds-of-nanoseconds simulations can reasonably estimate H-bonding fractions in proteins without large conformational changes, like PYP.[71]

In analogy to the approach for single –C≡N IR bands, we determined the H-bond blueshifts for the H-bonding populations from experiments (Table S30) and MD. We again derived the MD H-bond shift via the H-bond geometry for the H-bonding population and halved those values to obtain $\Delta\bar{\nu}_{HB,MD}$ when the H-bond rapidly fluctuates relative to the single ps timescale of oCNF's vibrational dephasing (Table S32).[72,73] In contrast to $\Delta\bar{\nu}_{HB,MD}$ values for single nitrile bands, multiplication by f_{protic} is unnecessary here since comparisons with the experimental protic population are explicitly made. Comparing all TDM- and MD-derived H-bond blueshifts, we obtain a best fit of $\Delta\bar{\nu}_{HB,MD} = (0.99 \pm 0.03)\,\Delta\bar{\nu}_{HB,TDM} + (1.2 \pm 0.3)$ with $R^2 = 0.83$ [Fig. 6(d)]. Again, we find that the best fit line is largely unchanged when we include values derived using the two-band TDM tuning method. However, R^2 in Fig. 6(d) is lower than that in Fig. 6(b), and inspection of Fig. 6(d) suggests this stems from deviations in the MD H-bond blueshifts of the P4-type nitrile protic populations. We investigated these blueshifts and found that they are likely overestimated due to our assumption that solvent H-bond shifts are halved compared to their value from Eq. (S1) due to fast H-bond fluctuations. Indeed, when the MD-derived shifts for P4-type nitriles are instead indirectly obtained from the MD electric fields, the $\Delta\bar{\nu}_{HB,MD}$ values decrease, and the same linear fit (within error) as in Fig. 6(d) is obtained but with an R^2 of 0.99 (see Table S33, Fig. S44, and discussion in Sec. S11). The similarity of the best fit lines in Figs. 6(b) and 6(d) and the good to excellent R^2 values directly show that nitrile H-bond geometries and dynamics are properly simulated independent of the H-bonding exchange regime. The slopes and correlations are highly similar to those in our prior work comparing TDM- and MD-derived H-bond shifts,[38] with the latter obtained from AMOEBA MD-based IR spectra. This finding implies a robustness in the assessment of H-bond shifts in solvent and diverse protein environments; this is notable given our pragmatic treatment of motional narrowing effects via averaging. It also establishes the H-bond blueshift as an independent observable for benchmarking applications.

Implications for the AMOEBA force field and using nitriles to probe local noncovalent interactions

The comparisons in Fig. 6 using the extended collection of nitrile-containing PYPs show that AMOEBA correctly models local electric fields and H-bond shifts, a metric for nitrile H-bond geometry. Importantly, these nitriles undergo fast and slow H-bond exchange dynamics while experiencing a wide range of average electric field strengths and H-bond donors. These points together indicate that AMOEBA properly simulates nitrile noncovalent interactions and thermodynamic and kinetic parameters in highly diverse protein environments. Specifically, accurate modeling of electric fields, H-bond geometries, and H-bonding fractions supports thermodynamic fidelity, while the ability to reproduce these parameters across fast and slow H-bond exchange regimes supports kinetic accuracy. Each correlation in Fig. 6 probes a distinct convolution of these factors (see explicit relations in Table S31), and their individual and collective strength justifies our conclusion. In contrast, we previously found that the fixed-charge AMBER force field is much less accurate at modeling nitrile H-bonding species,[52] and reanalysis of the PYP variants' AMBER results implies that this is because it does not properly model the kinetics or thermodynamics of nitriles interacting with their environments (Sec. S12).

The strong correlations observed here demonstrate that nitrile TDM–frequency dual tuning is a versatile tool for probing noncovalent interactions across environments with both fast and slow H-bond dynamics. This is an important finding, as many studies employing nitriles have been hampered by the presence of H-bonding and complicated IR lineshapes that impede VSE-based electric field assessments via frequency shifts.[25,74–76] In this context, we recall that our calibration for measuring –C≡N electric fields in proteins is based on correlating nitrile TDMs for the oTN model molecule with AMOEBA electric fields in solvents.[30,52] This makes the strong correlations in Fig. 6 even more impressive and strengthens the use of –C≡N TDMs as a robust observable for assessment of electric fields in the condensed phase, with successful application in a diverse range of 11 solvents and 12 proteins so far.[30,52]

This study was greatly aided by the versatility of PYP as a model system: it is amenable to nitrile incorporation via amber

suppression, its crystals diffract to high-resolution, its visible absorption spectrum can be used to assess the nitrile concentration for TDM calculation, and it is a small protein with conformational dynamics that are adequately captured in ~100 ns MD simulations.[30,52] Of these features, the most important for TDM tuning analysis was the ability to precisely determine the nitrile concentration. This may be difficult to do in some contexts, such as on electrode surfaces[16,32,77] or in metal organic frameworks.[10] Furthermore, nitrile probes in these systems are often more readily observed using Raman spectroscopy or vibrational sum-frequency generation,[10,77] where intensities do not simply depend on TDMs. At the same time, nitrile IR intensities can be particularly useful for systems that undergo time-dependent structural changes at constant concentration, e.g., the PYP photocycle[78] or other light-driven proteins.[59,76,79,80]

CONCLUSIONS AND OUTLOOK

Benchmarking MD force fields with local observables is critical to ensure that they are physically accurate and provide realistic chemical modeling.[81,82] In this study, we report a collection of nitrile-containing PYP mutants, which we characterized with IR spectroscopy, X-ray crystallography, and MD simulations. This characterization enabled assessment of nitrile local environments via one- and (newly established) two-band TDM tuning methods. Benchmarking AMOEBA using electric fields and H-bond shifts revealed strong correlations between experiments and simulations, showing that it accurately captures subtle and significant differences in nitrile electrostatic and H-bonding noncovalent interactions. Notably, strong correlations were found for settings with slow and fast exchange dynamics, suggesting that AMOEBA models thermodynamic and kinetic parameters of nitrile–environment interactions well. Altogether, our previous[52] and present tests of force fields at both chemical and physical levels show that while fixed-charge models often achieve qualitative consensus with experiments, force fields with more extensive electrostatic descriptions, such as AMOEBA, are needed to consistently attain quantitative agreement. Consequently, AMOEBA holds extraordinary potential as a force field for proteins and biomolecules in general.[83–91]

To conclude, we highlight future experimental and computational opportunities based on this study and our recent work on nitriles.[30,38,52] Just as measuring –C≡N frequencies and TDMs enables two unique assessments of local noncovalent interactions, similar dual tuning applications may enhance the utility of alkynes (–C≡C–)[92] and isonitriles (–N≡C)[64] to inform on molecular environments. Relatedly, using AMOEBA to model small molecules with these probes can likely help with recapitulating IR spectra from vibrational spectroscopic maps,[93,94] as was the case when we modeled the –C≡N blueshift.[38] In addition, employing 2D IR spectroscopy to study nitriles[73,95,96] can expand their use as electric field reporters, as absolute and relative TDMs can be determined more accurately than with linear IR spectroscopy.[64,65,97,98] Coupling this method with TDM tuning analysis would likely be useful for the common situation where –C≡N spectra in proteins report on heterogeneous environments with multiple bands.[29,57,59,76,99–101] Furthermore, exchange rates can be determined directly using 2D IR spectroscopy,[63,64] and these can be used as another observable for benchmarking; isonitriles may prove useful for this, as they have larger extinction coefficients than nitriles[102] and slower exchange dynamics, which often leads to larger band separation.[64,93,103,104] Finally, our benchmarking was enabled by generating and extensively characterizing 12 diverse nitrile-containing PYPs. We hope this collection can aid future benchmarking for extensions of AMOEBA,[105–109] other polarizable force fields, such as the Drude-type implementations,[110,111] and machine learning force fields.[112–117]

ACCESSION CODES

All X-ray density maps and atomic models for the proteins have been deposited in the Protein Data Bank. PYP variants: F28oCNF, 7SPX; F62oCNF, 7SPW; F92oCNF, 7SPV; F96oCNF, 7SJJ. F96oCNF mutants: T103V, 8E02; I49T, 8DZY; M100Q, 8E1K; M100E, 8E09; M100D, 8E03; M100N, 8 E1L; M100K, 8DZX; M100H, 8DZU.

SUPPLEMENTARY MATERIAL

See the supplementary material for experimental details on sample preparation and IR spectroscopy, computational details on MD simulations, IR spectra and TDM calibration curve for Boc-protected version of oCNF, UV–Vis spectra and visible extinction coefficient determination of F96oCNF PYP mutants, IR spectra of F96oCNF mutants at pH 8.5, X-ray crystallography methods, discussion on crystallography of F96oCNF mutants, MD analysis of nitrile-incorporated PYPs, discussion on extended TDM tuning method, MD H-bond shifts, alternative H-bond blueshift correlations, broader implications from this study, and attempts to characterize MD nitrile H-bonding and non-H-bonding residence times.

ACKNOWLEDGMENTS

J.M.K. was supported by a Burt and Deedee McMurtry Stanford Graduate Fellowship. J.B.W. was supported by a Stanford Center for Molecular Analysis and Design Fellowship. This work was supported by grants from the NSF (Grant No. MCB-1915727) and NIH (Grant No. R35GM118044) to S.G.B. J.K. acknowledges the German Research Foundation (DFG) for financial support via the DFG Individual Research Grant No. KO 5464-4 (Grant No. 493270578). We thank the HPC Service of ZEDAT (Freie Universität Berlin) for computing time (10.17169/refubium-26754) and Dr. Silvia Russi at the Stanford Synchrotron Radiation Lightsource (SSRL) for technical assistance during X-ray data collection. Use of the SSRL, SLAC National Accelerator Laboratory, is supported by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences, under Contract No. DE-AC02-76SF00515. The SSRL Structural Molecular Biology Program is supported by the DOE Office of Biological and Environmental Research and by the National Institutes of Health, National Institute of General Medical Sciences (Grant No. P30GM133894). The contents of this publication are solely the responsibility of the authors and do not necessarily represent the official views of NIGMS or NIH.

AUTHOR DECLARATIONS

Conflict of Interest

The authors have no conflicts to disclose.

Author Contributions

Jacob M. Kirsh: Conceptualization (equal); Formal analysis (equal); Investigation (equal); Methodology (equal); Writing – original draft (equal); Writing – review & editing (equal). **Jared Bryce Weaver**: Conceptualization (equal); Formal analysis (supporting); Investigation (equal); Methodology (equal). **Jacek Kozuch**: Formal analysis (supporting); Funding acquisition (supporting); Investigation (equal); Methodology (equal); Writing – review & editing (equal). **Steven G. Boxer**: Funding acquisition (lead); Supervision (lead); Writing – review & editing (equal).

DATA AVAILABILITY

All MD parameters and Python scripts used in this work can be found at https://github.com/KozuchLab/Publications/tree/main/oCNPhe_GROMACS_TINKER. Other data that support the findings of this study are available from the corresponding authors upon reasonable request.

REFERENCES

[1] H. A. L. Filipe and L. M. S. Loura, "Molecular dynamics simulations: Advances and applications," Molecules **27**(7), 2105 (2022).
[2] S. A. Showalter and R. Brüschweiler, "Validation of molecular dynamics simulations of biomolecules using NMR spin relaxation as benchmarks: Application to the AMBER99SB force field," J. Chem. Theory Comput. **3**(3), 961–975 (2007).
[3] K. Sarthak, D. Winogradoff, Y. Ge, S. Myong, and A. Aksimentiev, "Benchmarking molecular dynamics force fields for all-atom simulations of biological condensates," J. Chem. Theory Comput. **19**(12), 3721–3740 (2023).
[4] K. Kříž, L. Schmidt, A. T. Andersson, M.-M. Walz, and D. van der Spoel, "An imbalance in the force: The need for standardized benchmarks for molecular simulation," J. Chem. Inf. Model. **63**(2), 412–431 (2023).
[5] S. D. Fried and S. G. Boxer, "Measuring electric fields and noncovalent interactions using the vibrational Stark effect," Acc. Chem. Res. **48**(4), 998–1006 (2015).
[6] B. Błasiak, C. H. Londergan, L. J. Webb, and M. Cho, "Vibrational probes: From small molecule solvatochromism theory and experiments to applications in complex systems," Acc. Chem. Res. **50**(4), 968–976 (2017).
[7] C. R. Baiz, B. Błasiak, J. Bredenbeck, M. Cho, J.-H. Choi, S. A. Corcelli, A. G. Dijkstra, C.-J. Feng, S. Garrett-Roe, N.-H. Ge, M. W. D. Hanson-Heine, J. D. Hirst, T. L. C. Jansen, K. Kwac, K. J. Kubarych, C. H. Londergan, H. Maekawa, M. Reppert, S. Saito, S. Roy, J. L. Skinner, G. Stock, J. E. Straub, M. C. Thielges, K. Tominaga, A. Tokmakoff, H. Torii, L. Wang, L. J. Webb, and M. T. Zanni, "Vibrational spectroscopic map, vibrational spectroscopy, and intermolecular interaction," Chem. Rev. **120**(15), 7152–7218 (2020).
[8] J. Kozuch, K. Ataka, and J. Heberle, "Surface-enhanced infrared absorption spectroscopy," Nat. Rev. Methods Primers **3**(1), 70 (2023).
[9] P. A. Kocheril, H. Wang, D. Lee, N. Naji, and L. Wei, "Nitrile vibrational lifetimes as probes of local electric fields," J. Phys. Chem. Lett. **15**(19), 5306–5314 (2024).
[10] Z. Ji, S. Mukherjee, J. Andreo, A. Sinelshchikova, F. Peccati, G. Jiménez-Osés, S. Wuttke, and S. G. Boxer, "Electrostatic atlas of non-covalent interactions built into metal–organic frameworks," Nat. Chem. **17**(12), 1920–1927 (2025).
[11] S. Mukherjee, S. D. E. Fried, N. Y. Hong, N. Bagheri, J. Kozuch, I. I. Mathews, J. M. Kirsh, and S. G. Boxer, "Covalent drug binding in live cells monitored by mid-infrared quantum cascade laser spectroscopy: Photoactive yellow protein as a model system," J. Am. Chem. Soc. **147**(51), 47213–47222 (2025).
[12] V. V. Welborn and T. Head-Gordon, "Fluctuations of electric fields in the active site of the enzyme ketosteroid isomerase," J. Am. Chem. Soc. **141**(32), 12487–12492 (2019).
[13] X. Wang, X. He, and J. Z. H. Zhang, "Predicting mutation-induced Stark shifts in the active site of a protein with a polarized force field," J. Phys. Chem. A **117**(29), 6015–6023 (2013).
[14] D. J. Sandberg, A. N. Rudnitskaya, and J. A. Gascón, "QM/MM prediction of the Stark shift in the active site of a protein," J. Chem. Theory Comput. **8**(8), 2817–2823 (2012).
[15] X. Wang, J. Z. H. Zhang, and X. He, "Quantum mechanical calculation of electric fields and vibrational Stark shifts at active site of human aldose reductase," J. Chem. Phys. **143**(18), 184111 (2015).
[16] S. Sarkar, C. Tseng, A. Maitra, M. J. Voegtle, and J. M. Dawlaty, *Advances in Vibrational Stark Shift Spectroscopy for Measuring Interfacial Electric Fields*, *ACS Symposium Series Vol. 1398* (American Chemical Society, 2021), pp. 199–224.
[17] L. Feuerstein, E. E. Bas, D. Golze, T. Heine, M. Oschatz, and I. M. Weidinger, "Nitrile groups as build-in molecular sensors for interfacial effects at electrocatalytically active carbon–nitrogen materials," ACS Appl. Mater. Interfaces **17**(16), 23996–24004 (2025).
[18] S. Bagchi, S. D. Fried, and S. G. Boxer, "A solvatochromic model calibrates nitriles' vibrational frequencies to electrostatic fields," J. Am. Chem. Soc. **134**(25), 10373–10376 (2012).
[19] A. T. Fafarman and S. G. Boxer, "Nitrile bonds as infrared probes of electrostatics in ribonuclease S," J. Phys. Chem. B **114**(42), 13536–13544 (2010).
[20] S. D. Fried and S. G. Boxe, "Electric fields and enzyme catalysis," Annu. Rev. Biochem. **86**, 387–415 (2017).
[21] C. Zheng, Y. Mao, J. Kozuch, A. O. Atsango, Z. Ji, T. E. Markland, and S. G. Boxer, "A two-directional vibrational probe reveals different electric field orientations in solution and an enzyme active site," Nat. Chem. **14**(8), 891–897 (2022).
[22] C. Zheng, Z. Ji, I. I. Mathews, and S. G. Boxer, "Enhanced active-site electric field accelerates enzyme catalysis," Nat. Chem. **15**(12), 1715–1721 (2023).
[23] J. D. Slocum and L. J. Webb, "Measuring electric fields in biological matter using the vibrational Stark effect of nitrile probes," Annu. Rev. Phys. Chem. **69**(1), 253–271 (2018).
[24] A. Maitra, W. R. Lake, A. Mohamed, S. C. Edington, P. Das, B. C. Thompson, S. Hammes-Schiffer, M. Johnson, and J. M. Dawlaty, "Measuring the electric fields of ions captured in crown ethers," J. Phys. Chem. Lett. **15**(29), 7458–7465 (2024).
[25] J. C. Fink and L. J. Webb, "The effect of phosphoserine-containing membranes on electrostatic fields at the protein–protein interface measured through vibrational Stark effect spectroscopy," Biochemistry **64**(10), 2280–2290 (2025).
[26] Q. Li, M. Wang, B. Zhuang, R.-R. Feng, W. Zhang, and F. Gai, "Nitrile infrared intensity is more sensitive to local electric field change than its frequency: Application to probe protein hydration dynamics," J. Chem. Phys. **163**(22), 225102 (2025).
[27] I. T. Suydam, C. D. Snow, V. S. Pande, and S. G. Boxer, "Electric fields at the active site of an enzyme: Direct comparison of experiment with theory," Science **313**(5784), 200–204 (2006).
[28] L. J. Webb and S. G. Boxer, "Electrostatic fields near the active site of human aldose reductase: 1. New inhibitors and vibrational Stark effect measurements," Biochemistry **47**(6), 1588–1598 (2008).
[29] L. Xu, A. E. Cohen, and S. G. Boxer, "Electrostatic fields near the active site of human aldose reductase: 2. New inhibitors and complications caused by hydrogen bonds," Biochemistry **50**(39), 8311–8322 (2011).
[30] J. B. Weaver, J. Kozuch, J. M. Kirsh, and S. G. Boxer, "Nitrile infrared intensities characterize electric fields and hydrogen bonding in protic, aprotic, and protein environments," J. Am. Chem. Soc. **144**(17), 7562–7567 (2022).
[31] H. A. McMahon, K. N. Alfieri, K. A. A. Clark, and C. H. Londergan, "Cyanylated cysteine: A covalently attached vibrational probe of protein–lipid contacts," J. Phys. Chem. Lett. **1**(5), 850–855 (2010).
[32] H. Biava, T. Schreiber, S. Katz, J.-S. Völler, M. Stolarski, C. Schulz, N. Michael, N. Budisa, J. Kozuch, T. Utesch, and P. Hildebrandt, "Long-range modulations of electric fields in proteins," J. Phys. Chem. B **122**(35), 8330–8342 (2018).
[33] J. T. First, J. D. Slocum, and L. J. Webb, "Quantifying the effects of hydrogen bonding on nitrile frequencies in GFP: Beyond solvent exposure," J. Phys. Chem. B **122**(26), 6733–6743 (2018).
[34] D. Ben-Amotz, M. Lee, S. Y. Cho, and D. J. List, "Solvent and pressure-induced perturbations of the vibrational potential surface of acetonitrile," J. Chem. Phys. **96**(12), 8781–8792 (1992).
[35] W. R. Fawcett, G. Liu, and T. E. Kessler, "Solvent-induced frequency shifts in the infrared spectrum of acetonitrile in organic solvents," J. Phys. Chem. **97**(37), 9293–9298 (1993).

[36]H. Torii, "Unified electrostatic understanding on the solvation-induced changes in the CN stretching frequency and the NMR chemical shifts of a nitrile," J. Phys. Chem. A **120**(36), 7137–7144 (2016).
[37]M. Hirose and H. Torii, "Role of the electrostatic interactions in the changes in the CN stretching frequency of benzonitrile interacting with hydrogen-bond donating molecules," J. Mol. Liq. **362**, 119714 (2022).
[38]J. M. Kirsh and J. Kozuch, "Hydrogen bond blueshifts in nitrile vibrational spectra are dictated by hydrogen bond geometry and dynamics," JACS Au **4**(12), 4844–4855 (2024).
[39]R. A. LaCour, R. Zhao, and T. Head-Gordon, "The interplay of Pauli repulsion, electrostatics, and field inhomogeneity for blueshifting and redshifting vibrational probe molecules," J. Chem. Theory Comput. (published online 2026).
[40]J.-H. Choi, K.-I. Oh, H. Lee, C. Lee, and M. Cho, "Nitrile and thiocyanate IR probes: Quantum chemistry calculation studies and multivariate least-square fitting analysis," J. Chem. Phys. **128**(13), 134506 (2008).
[41]H. Lee, J.-H. Choi, and M. Cho, "Vibrational solvatochromism and electrochromism. II. Multipole analysis," J. Chem. Phys. **137**(11), 114307 (2012).
[42]B. Błasiak, A. W. Ritchie, L. J. Webb, and M. Cho, "Vibrational solvatochromism of nitrile infrared probes: Beyond the vibrational Stark dipole approach," Phys. Chem. Chem. Phys. **18**(27), 18094–18111 (2016).
[43]N. M. Levinson, S. D. Fried, and S. G. Boxer, "Solvent-induced infrared frequency shifts in aromatic nitriles are quantitatively described by the vibrational Stark effect," J. Phys. Chem. B **116**(35), 10470–10476 (2012).
[44]J. R. Schmidt, S. A. Corcelli, and J. L. Skinner, "Pronounced non-Condon effects in the ultrafast infrared spectroscopy of water," J. Chem. Phys. **123**(4), 044513 (2005).
[45]R. Malik, A. Chandra, B. Das, and A. Chandra, "Temperature dependence of non-Condon effects in two-dimensional vibrational spectroscopy of water," J. Phys. Chem. B **127**(11), 2488–2498 (2023).
[46]P. W. Schultz, G. E. Leroi, and A. I. Popov, "Solvation of SCN- and SeCN- anions in hydrogen-bonding solvents," J. Am. Chem. Soc. **118**(43), 10617–10625 (1996).
[47]B. A. Lindquist and S. A. Corcelli, "Nitrile groups as vibrational probes: Calculations of the C≡N infrared absorption line shape of acetonitrile in water and tetrahydrofuran," J. Phys. Chem. B **112**(20), 6301–6303 (2008).
[48]S. A. Yamada, W. H. Thompson, and M. D. Fayer, "Water-anion hydrogen bonding dynamics: Ultrafast IR experiments and simulations," J. Chem. Phys. **146**(23), 234501 (2017).
[49]R. J. Xu, B. Blasiak, M. Cho, J. P. Layfield, and C. H. Londergan, "A direct, quantitative connection between molecular dynamics simulations and vibrational probe line shapes," J. Phys. Chem. Lett. **9**(10), 2560–2567 (2018).
[50]M. G. Maienschein-Cline and C. H. Londergan, "The CN stretching band of aliphatic thiocyanate is sensitive to solvent dynamics and specific solvation," J. Phys. Chem. A **111**(40), 10020–10025 (2007).
[51]J. T. Henry and S. Crosson, "Ligand-binding PAS domains in a genomic, cellular, and structural context," Annu. Rev. Microbiol. **65**, 261–286 (2011).
[52]J. M. Kirsh, J. B. Weaver, S. G. Boxer, and J. Kozuch, "Critical evaluation of polarizable and nonpolarizable force fields for proteins using experimentally derived nitrile electric fields," J. Am. Chem. Soc. **146**(10), 6983–6991 (2024).
[53]K. Lindorff-Larsen, S. Piana, K. Palmo, P. Maragakis, J. L. Klepeis, R. O. Dror, and D. E. Shaw, "Improved side-chain torsion potentials for the amber ff99SB protein force field," Proteins: Struct., Funct., Bioinf. **78**(8), 1950–1958 (2010).
[54]Y. Shi, Z. Xia, J. Zhang, R. Best, C. Wu, J. W. Ponder, and P. Ren, "Polarizable atomic multipole-based AMOEBA force field for proteins," J. Chem. Theory Comput. **9**(9), 4046–4063 (2013).
[55]R. Kubo, "Statistical-mechanical theory of irreversible processes. I. General theory and simple applications to magnetic and conduction problems," J. Phys. Soc. Jpn. **12**(6), 570–586 (1957).
[56]R. Kubo, M. Yokota, and S. Nakajima, "Statistical-mechanical theory of irreversible processes. II. Response to thermal disturbance," J. Phys. Soc. Jpn. **12**(11), 1203–1211 (1957).
[57]B. Lee, B. M. Papoutsis, N. Y. Wong, J. Piacentini, C. Kearney, N. A. Huggins, N. Cruz, T. T. Ng, K. H. Hao, J. S. Kramer, E. E. Fenlon, P. S. Nerenberg, C. M. Phillips-Piro, and S. H. Brewer, "Unraveling complex local protein environments with 4-cyano-l-phenylalanine," J. Phys. Chem. B **126**(44), 8957–8969 (2022).
[58]B. A. Lindquist, K. E. Furse, and S. A. Corcelli, "Nitrile groups as vibrational probes of biomolecular structure and dynamics: An overview," Phys. Chem. Chem. Phys. **11**(37), 8119–8132 (2009).
[59]A. D. Nguyen, N. Michael, L. Sauthof, J. von Sass, O. T. Hoang, A. Schmidt, M. La Greca, R. Schlesinger, N. Budisa, P. Scheerer, M. A. Mroginski, A. Kraskov, and P. Hildebrandt, "Hydrogen bonding and noncovalent electric field effects in the photoconversion of a phytochrome," J. Phys. Chem. B **128**(47), 11644–11657 (2024).
[60]E. S. Park and S. G. Boxer, "Origins of the sensitivity of molecular vibrations to electric fields: Carbonyl and nitrosyl stretches in model compounds and proteins," J. Phys. Chem. B **106**(22), 5800–5806 (2002).
[61]E. D. Getzoff, K. N. Gutwin, and U. K. Genick, "Anticipatory active-site motions and chromophore distortion prime photoreceptor PYP for light activation," Nat. Struct. Mol. Biol. **10**(8), 663–668 (2003).
[62]J. Zheng, K. Kwak, J. Asbury, X. Chen, I. R. Piletic, and M. D. Fayer, "Ultrafast dynamics of solute-solvent complexation observed at thermal equilibrium in real time," Science **309**(5739), 1338–1343 (2005).
[63]J. Zheng, K. Kwak, and M. D. Fayer, "Ultrafast 2D IR vibrational echo spectroscopy," Acc. Chem. Res. **40**(1), 75–83 (2007).
[64]J. Kübel, G. Lee, S. A. Ooi, S. Westenhoff, H. Han, M. Cho, and M. Maj, "Ultrafast chemical exchange dynamics of hydrogen bonds observed via isonitrile infrared sensors: Implications for biomolecular studies," J. Phys. Chem. Lett. **10**(24), 7878–7883 (2019).
[65]P. M. Donaldson, "Spectrophotometric concentration analysis without molar absorption coefficients by two-dimensional-infrared and Fourier transform infrared spectroscopy," Anal. Chem. **94**(51), 17988–17999 (2022).
[66]D. York, N. M. Evensen, M. L. Martınez, and J. De Basabe Delgado, "Unified equations for the slope, intercept, and standard errors of the best straight line," Am. J. Phys. **72**(3), 367–375 (2004).
[67]Travis Wiens, Linear Regression with Errors in X and Y, https://www.Mathworks.Com/Matlabcentral/Fileexchange/26586-Linear-Regression-with-Errors-in-x-and-y, 2023, *MATLAB Central File Exchange*, retrieved 30 September 2023.
[68]S. D. Fried, L.-P. Wang, S. G. Boxer, P. Ren, and V. S. Pande, "Calculations of the electric fields in liquid solutions," J. Phys. Chem. B **117**(50), 16236–16248 (2013).
[69]R. T. Bradshaw, J. Dziedzic, C.-K. Skylaris, and J. W. Essex, "The role of electrostatics in enzymes: Do biomolecular force fields reflect protein electric fields?," J. Chem. Inf. Model. **60**(6), 3131–3144 (2020).
[70]J. Huang, S. Rauscher, G. Nawrocki, T. Ran, M. Feig, B. L. de Groot, H. Grubmüller, and A. D. MacKerell, "CHARMM36m: An improved force field for folded and intrinsically disordered proteins," Nat. Methods **14**(1), 71–73 (2017).
[71]P. Düx, G. Rubinstenn, G. W. Vuister, R. Boelens, F. A. A. Mulder, K. Hård, W. D. Hoff, A. R. Kroon, W. Crielaard, K. J. Hellingwerf, and R. Kaptein, "Solution structure and backbone dynamics of the photoactive yellow protein," Biochemistry **37**(37), 12689–12699 (1998).
[72]J. M. Rodgers, W. Zhang, C. G. Bazewicz, J. Chen, S. H. Brewer, and F. Gai, "Kinetic isotope effect provides insight into the vibrational relaxation mechanism of aromatic molecules: Application to cyano-phenylalanine," J. Phys. Chem. Lett. **7**(7), 1281–1287 (2016).
[73]S. Bagchi, S. G. Boxer, and M. D. Fayer, "Ribonuclease S dynamics measured using a nitrile label with 2D IR vibrational echo spectroscopy," J. Phys. Chem. B **116**(13), 4034–4042 (2012).
[74]A. S. Chaudhari, A. Chatterjee, C. A. O. Domingos, P. C. Andrikopoulos, Y. Liu, I. Andersson, B. Schneider, V. A. Lórenz-Fonfría, and G. Fuertes, "Genetically encoded non-canonical amino acids reveal asynchronous dark reversion of chromophore, backbone, and side-chains in EL222," Protein Sci. **32**(4), e4590 (2023).
[75]H. Mohrmann, I. Kube, V. A. Lórenz-Fonfría, M. Engelhard, and J. Heberle, "Transient conformational changes of sensory rhodopsin II investigated by vibrational Stark effect probes," J. Phys. Chem. B **120**(19), 4383–4387 (2016).
[76]A. Kraskov, J. von Sass, A. D. Nguyen, T. O. Hoang, D. Buhrke, S. Katz, N. Michael, J. Kozuch, I. Zebger, F. Siebert, P. Scheerer, M. A. Mroginski, N. Budisa, and P. Hildebrandt, "Local electric field changes during the photoconversion of the bathy phytochrome Agp2," Biochemistry **60**(40), 2967–2977 (2021).

[77]S. Sarkar, A. Maitra, S. Banerjee, V. S. Thoi, and J. M. Dawlaty, "Electric fields at metal–surfactant interfaces: A combined vibrational spectroscopy and capacitance study," J. Phys. Chem. B **124**(7), 1311–1321 (2020).
[78]F. Schotte, H. S. Cho, F. Dyda, and P. Anfinrud, "Watching a signaling protein function: What has been learned over four decades of time-resolved studies of photoactive yellow protein," Struct. Dyn. **11**(2), 021303 (2024).
[79]J. D. Slocum and L. J. Webb, "Nitrile probes of electric field agree with independently measured fields in green fluorescent protein even in the presence of hydrogen bonding," J. Am. Chem. Soc. **138**(20), 6561–6570 (2016).
[80]J. B. Weaver, S. G. Boxer, C. E. D. Chidsey, and A. R. Dunn, "Using genetic code expansion to investigate electron transfer and protein electrostatics in the photosynthetic reaction center," Ph.D. thesis, Stanford University, Stanford, CA, 2022.
[81]K. Lindorff-Larsen, P. Maragakis, S. Piana, M. P. Eastwood, R. O. Dror, and D. E. Shaw, "Systematic validation of protein force fields against experimental data," PLoS One **7**(2), e32131 (2012).
[82]S. Bottaro and K. Lindorff-Larsen, "Biophysical experiments and biomolecular simulations: A perfect match?," Science **361**(6400), 355–360 (2018).
[83]C. Zhang, C. Lu, Z. Jing, C. Wu, J.-P. Piquemal, J. W. Ponder, and P. Ren, "AMOEBA polarizable atomic multipole force field for nucleic acids," J. Chem. Theory Comput. **14**(4), 2084–2108 (2018).
[84]J. M. Delgado, P. R. Nagy, and S. Varma, "Polarizable AMOEBA model for simulating Mg^{2+}·protein·nucleotide complexes," J. Chem. Inf. Model. **64**(2), 378–392 (2024).
[85]H. S. Antila, S. Dixit, B. Kav, J. J. Madsen, M. S. Miettinen, and O. H. S. Ollila, "Evaluating polarizable biomembrane simulations against experiments," J. Chem. Theory Comput. **20**(10), 4325–4337 (2024).
[86]C. Song and L.-P. Wang, "A polarizable QM/MM model that combines the state-averaged CASSCF and AMOEBA force field for photoreactions in proteins," J. Chem. Theory Comput. **20**(15), 6632–6651 (2024).
[87]M. Deegbey, E. W. Sumner, and V. Vaissier Welborn, "Is AMOEBA a good force field for molecular dynamics simulations of carbohydrates?," J. Chem. Inf. Model. **65**(11), 5289–5300 (2025).
[88]D. El Ahdab, L. Lagardère, Z. Hobaika, T. Jaffrelot Inizan, F. Célerse, N. Gresh, R. G. Maroun, and J.-P. Piquemal, "AMOEBA polarizable molecular dynamics simulations of guanine quadruplexes: From the c-kit proto-oncogene to HIV-1," J. Chem. Inf. Model. **65**(9), 4488–4500 (2025).
[89]M. Roca, Y. Maghsoud, G. A. Cisneros, K. Świderek, and V. Moliner, "Comparing force field treatments in QM/MM studies of the SARS-CoV-2 RNA-dependent RNA polymerase (RdRp) mechanism," J. Chem. Theory Comput. **21**(23), 12342–12355 (2025).
[90]J. M. Delgado, G. Schillaci, and S. Varma, "Phosphates and phosphorylated amino acids for the AMOEBA-HFC polarizable force field," J. Chem. Theory Comput. **21**(22), 11645–11656 (2025).
[91]M. A. Schulz and J. Kozuch, "Catalytic activity of TEM β-lactamases is modulated by ligand sidechain conformation during evolution of antibiotic resistance," chemRxiv:2025-9tlxj-v2 (2025).
[92]M. G. Romei, E. V. Von Krusenstiern, S. T. Ridings, R. N. King, J. C. Fortier, C. A. McKeon, K. M. Nichols, L. K. Charkoudian, and C. H. Londergan, "Frequency changes in terminal alkynes provide strong, sensitive, and solvatochromic Raman probes of biochemical environments," J. Phys. Chem. B **127**(1), 85–94 (2023).
[93]M. Maj, "Solvatochromic charge model of isonitrile probes for investigating hydrogen-bond dynamics with 2DIR spectroscopy," J. Chem. Phys. **162**(1), 014113 (2025).
[94]A. Aneesh, K. Streu, and C. A. Daly, Jr., "Development of a spectroscopic map to explain the broad Raman peak for alkynes solvated in triethylamine," J. Phys. Chem. B **129**(33), 8509–8520 (2025).
[95]C. Fang, J. D. Bauman, K. Das, A. Remorino, E. Arnold, and R. M. Hochstrasser, "Two-dimensional infrared spectra reveal relaxation of the nonnucleoside inhibitor TMC278 complexed with HIV-1 reverse transcriptase," Proc. Natl. Acad. Sci. U. S. A. **105**(5), 1472–1477 (2008).
[96]A. Ghosh, A. Remorino, M. J. Tucker, and R. M. Hochstrasser, "2D IR photon echo spectroscopy reveals hydrogen bond dynamics of aromatic nitriles," Chem. Phys. Lett. **469**(4–6), 325–330 (2009).
[97]S. M. Fica-Contreras, R. Daniels, O. Yassin, D. J. Hoffman, J. Pan, G. Sotzing, and M. D. Fayer, "Long vibrational lifetime R-selenocyanate probes for ultrafast infrared spectroscopy: Properties and synthesis," J. Phys. Chem. B **125**(31), 8907–8918 (2021).
[98]M. Grechko and M. T. Zanni, "Quantification of transition dipole strengths using 1D and 2D spectroscopy for the identification of molecular structures via exciton delocalization: Application to α-helices," J. Chem. Phys. **137**(18), 184202 (2012).
[99]S. K. Jha, M. Ji, K. J. Gaffney, and S. G. Boxer, "Direct measurement of the protein response to an electrostatic perturbation that mimics the catalytic cycle in ketosteroid isomerase," Proc. Natl. Acad. Sci. U. S. A. **108**(40), 16612–16617 (2011).
[100]A. T. Fafarman, P. A. Sigala, J. P. Schwans, T. D. Fenn, D. Herschlag, and S. G. Boxer, "Quantitative, directional measurement of electric field heterogeneity in the active site of ketosteroid isomerase," Proc. Natl. Acad. Sci. U. S. A. **109**(6), E299–E308 (2012).
[101]P. A. Sigala, A. T. Fafarman, J. P. Schwans, S. D. Fried, T. D. Fenn, J. M. M. Caaveiro, B. Pybus, D. Ringe, G. A. Petsko, S. G. Boxer, and D. Herschlag, "Quantitative dissection of hydrogen bond-mediated proton transfer in the ketosteroid isomerase active site," Proc. Natl. Acad. Sci. U. S. A. **110**(28), E2552–E2561 (2013).
[102]A. Kanzow, M. A. Suhm, and M. Bödecker, "Isocyanides versus nitriles: Divergent hydrogen bonding behavior driven by the balance between dispersive and electrostatic forces," ChemPhysChem **27**(5), e202500834 (2026).
[103]M. Maj, C. Ahn, D. Kossowska, K. Park, K. Kwak, H. Han, and M. Cho, "β-isocyanoalanine as an IR probe: Comparison of vibrational dynamics between isonitrile and nitrile-derivatized IR probes," Phys. Chem. Chem. Phys. **17**(17), 11770–11778 (2015).
[104]M. Maj, C. Ahn, B. Błasiak, K. Kwak, H. Han, and M. Cho, "Isonitrile as an ultrasensitive infrared reporter of hydrogen-bonding structure and dynamics," J. Phys. Chem. B **120**(39), 10167–10180 (2016).
[105]Y. Yang, Q. Jin, and S. Yin, "Development of an anisotropic polarizable model for the all-atom AMOEBA force field," Phys. Chem. Chem. Phys. **26**(35), 22900–22911 (2024).
[106]N. Mauger, T. Plé, L. Lagardère, S. Huppert, and J.-P. Piquemal, "The Q-AMOEBA (CF) polarizable potential," J. Phys. Chem. Lett. **16**(23), 5723–5731 (2025).
[107]A. C. Thiel, M. J. Speranza, S. Jadhav, L. L. Stevens, D. K. Unruh, P. Ren, J. W. Ponder, J. Shen, and M. J. Schnieders, "Constant-pH simulations with the polarizable atomic multipole AMOEBA force field," J. Chem. Theory Comput. **20**(7), 2921–2933 (2024).
[108]X. Yang, C. Liu, and P. Ren, "Exploring biomolecular conformational dynamics with polarizable force field AMOEBA and enhanced sampling method milestoning," J. Chem. Theory Comput. **20**(10), 4065–4075 (2024).
[109]Y. Wang, T. J. Inizan, C. Liu, J.-P. Piquemal, and P. Ren, "Incorporating neural networks into the AMOEBA polarizable force field," J. Phys. Chem. B **128**(10), 2381–2388 (2024).
[110]J. A. Lemkul, J. Huang, B. Roux, and A. D. MacKerell, Jr., "An empirical polarizable force field based on the classical Drude oscillator model: Development history and recent applications," Chem. Rev. **116**(9), 4983–5013 (2016).
[111]A. Khabibrakhmanov, D. V. Fedorov, A. Ambrosetti, J. Crain, K. L. C. Hunt, E. R. Johnson, K. D. Jordan, S. Góger, M. Gori, M. R. Karimpour, R. J. Maurer, M. Sadhukhan, M. Stöhr, and A. Tkatchenko, "Accurate noncovalent interactions in atomistic systems via quantum Drude oscillators," J. Chem. Phys. **163**(15), 151001 (2025).
[112]O. T. Unke, S. Chmiela, H. E. Sauceda, M. Gastegger, I. Poltavsky, K. T. Schütt, A. Tkatchenko, and K.-R. Müller, "Machine learning force fields," Chem. Rev. **121**(16), 10142–10186 (2021).
[113]O. T. Unke, M. Stöhr, S. Ganscha, T. Unterthiner, H. Maennel, S. Kashubin, D. Ahlin, M. Gastegger, L. Medrano Sandonas, J. T. Berryman, A. Tkatchenko, and K.-R. Müller, "Biomolecular dynamics with machine-learned quantum-mechanical force fields trained on diverse chemical fragments," Sci. Adv. **10**(14), eadn4397 (2024).

[114]D. P. Kovács, J. H. Moore, N. J. Browning, I. Batatia, J. T. Horton, Y. Pu, V. Kapil, W. C. Witt, I.-B. Magdău, D. J. Cole, and G. Csányi, "MACE-OFF: Short-range transferable machine learning force fields for organic molecules," J. Am. Chem. Soc. **147**(21), 17598–17611 (2025).
[115]A. Kabylda, J. T. Frank, S. Suárez-Dou, A. Khabibrakhmanov, L. Medrano Sandonas, O. T. Unke, S. Chmiela, K.-R. Müller, and A. Tkatchenko, "Molecular simulations with a pretrained neural network and universal pairwise force fields," J. Am. Chem. Soc. **147**(37), 33723–33734 (2025).
[116]S. Suárez-Dou, M. Gallegos, K. Han, F. N. Brünig, J. T. Berryman, and A. Tkatchenko, "Stability and vibrations of proteins in vacuum and water: Bridging quantum accuracy and force-field efficiency," arXiv:2601.09845 (2026).
[117]V. Kostal, B. L. Shanks, P. Jungwirth, and H. Martinez-Seara, "Bayesian learning for accurate and robust biomolecular force fields," J. Chem. Theory Comput. **22**(5), 2652–2663 (2026).

Supporting Information for

A Comprehensive Test of the AMOEBA Force Field Using Spectroscopy, Structures, and Simulations of Nitrile Protein Environments

Jacob M. Kirsh,[1,§] Jared Bryce Weaver,[1,#] Jacek Kozuch,[2,3,*] and Steven G. Boxer[1,*]

[1] Department of Chemistry, Stanford University, Stanford, California 94305, United States.

[2] Institute of Physical Chemistry and Theoretical Chemistry, Braunschweig Integrated Centre of Systems Biology, Rebenring 56, 38106 Braunschweig, Germany.

[3] Department of Physics, Freie Universität Berlin, D-14195 Berlin, Germany.

[§] Present Address: Division of Chemistry and Chemical Engineering, California Institute of Technology, 1200 E California Blvd, Pasadena, CA, 91125.

[#]Present Address: Drug Discovery, Insitro, Inc. 279 E. Grand Ave., South San Francisco, California 94080, United States.

* Correspondence to J.K. (jacek.kozuch@tu-braunschweig.de) and S.G.B. (sboxer@stanford.edu)

Table of Contents

S1 Sample Preparation

Plasmid Construction

Constructs for the Photoactive Yellow Protein (PYP) F96oCNF mutants were generated using the previously reported pBAD-PYP-F96TAG plasmid[1] using an Agilent QuikChange Lightning Site-Directed Mutagenesis Kit. Resultant PYP plasmids were co-transformed with pEVOLpylT-N346AC348A (Addgene) into MAX Efficiency™ DH10B Competent Cells (ThermoFisher).

DNA Sequence

pBAD-PYP-F96TAG (codon for chromophore attachment site underlined; amber suppression TAG site in **bold**; I49, M100, T103 codons in green, blue, and red, respectively)

ATGGGGGGTTCTCATCATCATCATCATCATGGTATGGCTAGCATGACTGGTGGACAGCAAATGGGTCGGGATCTGTACGAC
GATGACGATAAGGATCCGAGACTCGAGGATGACGATGACAAAATGGAACACGTAGCCTTCGGTAGCGAGGACATCGAGA
ACACCCTCGCCAAGATGGACGACGGCCAGCTCGACGGCCTGGCCTTCGGCGCCATCCAGCTCGACGGCGACGGCAACATCC
TTCAGTACAACGCCGCGGAGGGCGACATCACCGGCCGCGACCCGAAGCAGGTCATCGGCAAGAACTTCTTCAAGGACGTG
GCCCCGTGCACTGACAGCCCGGAGTTCTACGGCAAGTTCAAGGAAGGGGTGGCCTCGGGCAACCTGAACACGATGTTCGA
GTACACC**TAG**GATTACCAAATGACGCCCACGAAGGTGAAGGTGCACATGAAGAAGGCCCTCTCCGGCGACAGCTACTGGG
TCTTCGTCAAGCGCGTCTAA

Note, the plasmid map and entire DNA sequence for the progenitor plasmid pBAD-PYP were previously provided in ref. [1].

Protein Expression and Purification

Protein expression, apo-PYP isolation, chromophore activation and incorporation, and purification of reconstituted protein were performed as previously reported.[1] As before, heterologous overexpression was performed in defined auto-induction media[2] with phenylalanine withheld and addition of 1 mM 2-cyano-L-phenylalanine (oCNF; Chem-Impex International). The buffer used from cell pellet resuspension through His-tag nickel affinity chromatography was Buffer A (20 mM TrisHCl (Fischer Scientific), 10 mM NaCl (Fischer Scientific), pH 8.0). PYPs reconstituted with para-coumaric acid (pCA) and with an intact N-terminal His-tag were purified via anion exchange chromatography on a linear program beginning in Buffer A and ending at 40% Buffer B (20 mM Tris HCl, 1 M NaCl, pH 8.0); proteins did not elute until ≥25% Buffer B. His-tags were cleaved by enterokinase (New England Biolabs), and PYP re-purified with anion exchange chromatography eluted between 13 and 21% Buffer B. The identity of all proteins was confirmed with liquid chromatography-electrospray ionization mass spectrometry (LC-ESI/MS) measured on a Waters 2795 Acquity HPLC with ZQ single quadrupole detector and C8 column attachment at Stanford University Mass Spectrometry (SUMS) facility. Expected and observed masses following His-tag removal are summarized in **Table S1**.

Table S1. Expected and observed mass for each F96oCNF PYP construct with chromophore incorporated.

PYP Construct	Expected Mass[a] (Da)	Observed Mass[b] (Da)
F96oCNF	14,045	14,050
F96oCNF I49T	14,033	14,038
F96oCNF M100D	14,029	14,034
F96oCNF M100E	14,043	14,048
F96oCNF M100H	14,051	14,056
F96oCNF M100K	14,042	14,047
F96oCNF M100N	14,028	14,033
F96oCNF M100Q	14,042	14,047
F96oCNF T103V	14,043	14,048

[a] Predicted from the primary sequence with N-terminal His-tag removed.
[b] Proteins with ~15kDa mass have ±5 Da deviations, depending on protonation states.

Protein Preparation for Fourier Transform Infrared (FTIR) Spectroscopy

PYPs were buffer exchanged on a 10 kDa Amicon centrifugal filter unit from the pH 8.0 buffers into a pH 6.0 buffer containing 20 mM NaH_2PO_4 (Fischer Scientific) and 10 mM NaCl. IR was also performed for some PYP mutants at pH 8.5 in 20 mM Tris-HCl and 10 mM NaCl (see **Figure S9** and discussion). 3 mg of PYP was exchanged by centrifugation at 14,000 G at 4°C until the sample was concentrated to ~0.02 mL, at which point the sample was spun an additional hour at the same conditions. Concentrated samples were pipetted up and down to ensure homogeneity and transferred to a microcentrifuge tube for storage. Immediately prior to use for FTIR spectroscopy, samples were spun down for one hour at 21,000 G at 4°C. Several PYPs demonstrated changes in their UV-Vis profile upon exposure to ambient light (see **Section S2**), so light exposure was minimized during sampling loading into the IR cell and handling to ensure that the protein remained in its dark-adapted ground state.

Small Molecule Preparation for FTIR Spectroscopy

Inspired by a report using a tert-butyloxycarbonyl (Boc)-protected amino acid to measure nitrile frequency tuning in an organic solvent,[3] we measured the nitrile frequency and transition dipole moment (TDM) of N-Boc-2-cyano-L-phenylalanine (BoCNF; Chem-Impex International; **Figure S1**) in a range of solvents to calibrate the observable responses to the nitrile's environmental electric field (**Figures S3–S4**). BoCNF was prepared to various concentrations in different solvents to ensure full solubility: 5 mM in dibutyl ether (DBE), 20 mM in cyclohexanone (CHX), N,N-dimethylformamide (DMF), dimethyl sulfoxide (DMSO), pyridine (PYR), tetrahydrofuran (THF), and toluene (TOL), and 10 mM in 100 mM HEPES, pH 7.0 buffer (Water; buffer was used to ensure a uniform protonation state of the (deprotonated) C-terminal carboxylate to ease comparisons with MD simulations and to increase solubility). All solvents were anhydrous and of 99.0% purity or higher (Sigma-Aldrich). Sample concentrations were determined by successively weighing the solute and then solute plus solvent to allow for

assessment of TDMs; solvent volume had a relative error of ± 0.1%. Solute mass error was estimated to be <1% of total mass, where mass error was estimated by reweighing all solute masses. To ensure the sample was fully dissolved, BoCNF was vortexed in solvent until no visible solid solute remained and was then vortexed for an additional five minutes. Heating and/or sonication to increase solubility were avoided due to potential heating effects on the amino acid. Unlike BoCNF in the other solvents, the amino acid's solubility in hexanes (HEX) was too low to reliably measure the nitrile TDM. So, 5-10 mgs of sample were added to ∼20 mLs of HEX solvent and vortexed for 10 minutes (additional vortexing was found to minimally increase the amino acid concentration). Samples in HEX were centrifuged and the supernatant was removed for spectroscopic use. BoCNF samples used to measure TDMs were independently prepared three times in each solvent; the sample in HEX, used only for determination of the nitrile IR absorption energy, was prepared once.

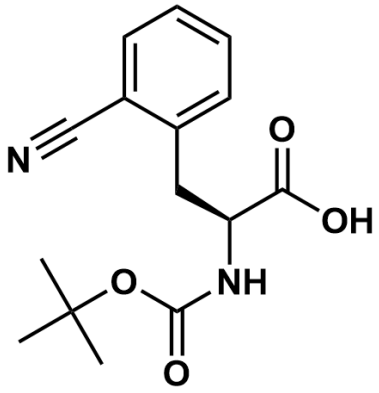

Figure S1. Structure of N-Boc-2-cyano-L-phenylalanine (BoCNF).

S2 Spectroscopic Methods

UV-Vis Absorption Measurements

UV-Vis absorption measurements were performed on a PerkinElmer Lambda 365 UV-Vis spectrometer. To determine the protein concentration for IR spectra, a UV-Vis absorption measurement was performed on the IR cell filled with sample. The protein concentration was determined using the visible absorption of PYP chromophore, para-coumaric acid (pCA). The nitrile concentration – which is necessary to determine the TDM – is equivalent to the protein concentration in the event of 100% fidelity of nitrile incorporation, which was previously confirmed by mass spectrometry (see Figure S12 in ref. [1]). All other spectra were acquired with a 1 mL quartz cuvette. Data were acquired every 1.0 nm at a maximum scan rate of 480 nm/min.

We note that while handling several of the F96oCNF PYP mutants (called F96oCNF mutants for short) under ambient light, namely M100Q, M100E, M100D, M100N, and M100H, we noticed a gradual change in pCA's absorption profile, where a new, higher energy peak corresponding with the protonated form of pCA grew in over time (see Figure 2 and discussion in ref. [4]). Consequently, all UV-Vis and IR spectra in this work are taken for the dark-adapted species to ensure measurements were uniformly taken in PYP's ground state.

Extinction Coefficient Determination of PYPs

A protocol for extinction coefficient determination of PYPs was established in our previous work.[1] It was determined that pCA has an $\varepsilon_{343nm} = 21.3\ mM^{-1}cm^{-1}$ when PYP is fully denatured in 4.5M guanidinium chloride. A UV-Vis absorption was taken before and after denaturation, so the extinction coefficient of pCA in intact PYP can be determined, and concentration normalized electronic (**Figure S6**) and vibrational spectra (**Figure S3**) were determined for all mutants. Note that measurements were carried out in the pH 6.0 buffer used for IR spectroscopy unless otherwise noted.

FTIR Spectroscopy

The procedure for acquisition and analysis of FTIR spectra was previously described in detail.[1] Briefly, spectra were obtained on a Bruker Vertex 70 spectrometer with a liquid-nitrogen-cooled mercury cadmium telluride (MCT) detector. The sample cell was composed of two CaF_2 windows separated by two nonoverlapping spacers, and exact cell pathlengths were determined by measuring the interference fringes from the empty IR cell (pathlengths ranging from 54-56 μm). The sample compartment was purged for 5-10 minutes with dry air prior to acquisition of spectra. For BoCNF samples, an aperture of 2 to 4 mm was used for acquisition and 256 or 512 scans were averaged to produce each spectrum at 1 cm^{-1} resolution. Three concentration normalized IR spectra were acquired for each sample preparation (*i.e.*, 9 spectra acquired per solvent except BoCNF in hexanes (3)). For protein samples, the sample compartment was purged for 10 minutes prior to acquisition and 9 spectra were acquired (512 scans per spectrum). The conditions for *o*-tolunitrile (oTN) in methanol are provided in ref. [5]; as a note, no purging was performed in that case to mitigate sample evaporation. All IR spectra were acquired at room temperature.

IR spectra were processed by subtracting the relevant blank from the sample spectrum and then baselining as minimally as possible, taking care to baseline far from the wings of the spectra to ensure the integrity of the peak areas (relatedly, automated baselining procedures have recently been developed for fidelitous determination of intensities[6]). All spectra were independently processed. For BoCNF and oTN spectra, the solvent was used as the blank, while for protein samples wild type (WT) PYP was used. BoCNF peaks were fit using a pseudo-Voigt lineshape in OPUS 6.5. A second band was included for BoCNF spectra in DMF and toluene to account for a small shoulder to the blue of the primary peak – this did not affect the analysis of nitrile frequencies. Peak areas were determined in Excel using the baselined spectra (i.e., not the band fits). As for BoCNF, fitting of PYP single band –C≡N spectra for environmental phenotypes P1 and P2 was performed with pseudo-Voigt lineshapes in OPUS 6.5. P3-type and P4-type nitrile spectra are asymmetric, but the underlying bands are not very well resolved at the 1 cm^{-1} resolution employed for FTIR, making it difficult to reproduce the entire bandshape solely by minimizing the residual root mean square error (RMSE) of pseudo-Voigts against the absorption spectrum. Consequently, P3-type and P4-type spectra were fit with two pseudo-Voigts where the nitrile absorption and the second derivative of the absorption were simultaneously fit by minimizing the RMSE's of both (the RMSE of the second derivative of the absorption was scaled by 100-fold to make the RMSE of the absorption and the second derivative on the same order);[5,7] the fits to the P3-type and P4-type absorption spectra are shown in **Figure S10**. Peak areas for oTN and PYPs were determined as above. Errors for the TDMs of individual bands in oTN, P3-, and P4-type spectra are the convolved errors from averaging the TDM of the entire bandshape and from averaging the unscaled band areas from different spectral fits.

S3 Molecular Dynamics (MD) Simulations and Analysis

BoCNF (Small Molecule) Solvent Electric Field Calculations

Molecular dynamics (MD) simulations were performed with Tinker 8.7[8] using the AMEOBA09 force field[9] as previously described.[1,10] Solvent parameters were taken as implemented in Tinker and from our previous work.[1,11] The protocol for the simulation of oTN in methanol can be found in ref. [5]. Neutral BoCNF (the relevant species in organic solvents) and anionic BoCNF (deprotonated carboxylate group; the relevant species in water) were parametrized using Poltype[12] with a multipole fitting to the electrostatic potentials from MP2/6-311g++(2d,2p) calculations in Gaussian 16.[13] (Note, the proper protonation state of BoCNF's C-terminal carboxylic acid/carboxylate for MD was identified using the BoCNF IR spectra from the appearance of an ~1750 cm^{-1} band for the carboxylic acid carbonyl stretch or an ~1600 cm^{-1} band for the carboxylate antisymmetric stretch). Details on MD setup, minimization, production runs, and electric field analysis are provided in ref [1]. Note that *environmental* electric fields were extracted, *i.e.*, MD electric fields had self-fields removed.[14] Environmental electric fields were obtained by first getting the induced dipoles from simulations with solvent present (referred to as the *absolute fields*), then a given trajectory was rerun with solvent removed to obtain the induced dipoles coming from BoCNF itself (which enables assessment of the self-field): the environmental fields were obtained by subtracting the self-field from the absolute fields.

Protein MD Calculations and Electric Field Calculations

AMOEBA MD simulations for the original variants F28oCNF, F62oCNF, F92oCNF, and F96oCNF PYP were started from previously deposited X-ray crystal structures with PDB IDs 7SPX, 7SPW, 7SPV, and 7SJJ, respectively.[1] Simulations for newly reported F96oCNF mutants T103V, I49T, M100Q, M100E, M100D, M100N, M100K, and M100H were started from X-ray crystal structures with PDB IDs 8E02, 8DZY, 8E1K, 8E09, 8E03, 8E1L, 8DZX, and 8DZU, respectively (**Table S10**).

MD simulations were performed with Tinker9[15] using the AMOEBABIO18 force field.[16] A 7.5 nm x 7.5 nm x 7.5 nm cubic box was filled with water and NaCl at 0.05 M and minimized (steepest descent until all forces were below 1 kcal mol^{-1} $Å^{-1}$) with van der Waals and electrostatics cutoffs of 9 and 7 Å, respectively. The dipole convergence criterion was set to 0.01 D with a mutual polarization scheme. NVT and NPT equilibrations were performed, where the van der Waals cutoff was increased to 12 Å, and the RESPA integrator, the Bussi thermostat, and Monte-Carlo barostat (300 K and 1 bar) were employed. MD production runs were performed with 1 fs time steps for a total time of 25 ns. Runs were repeated eight times to produce 200 ns aggregate MD trajectories for each F96oCNF mutant. Additionally, four repeats of 25 ns were performed for the PYP variants and added to the 100 ns aggregate trajectories from our prior work[17] to bring all nitrile-containing PYPs to a total of 200 ns simulation time. Glu 46 was simulated as a protonated carboxylic acid, as determined previously.[18] Note that these simulations required the inclusion of two "noncanonical" amino acids: oCNF and pCA covalently linked with Cys 69 (*i.e.,* the chromophore for PYP).[19,20] Details on the parameterization of these amino acids are described in ref. [17], Section S1.

In one of the crystal structures (F28oCNF), two nitrile orientations were observed, and in four of the mutant crystal structures (T103V, M110Q, M100D, and M100N), two to three conformations of the newly mutated amino acid were observed. **Table S2** details how many of the eight total simulations per PYP were begun from a given conformation.

Table S2. The number of runs began from a given nitrile or mutant conformation of the 8 total MD runs per PYP.

PYP Construct	Conformation A	Conformation B	Conformation C
F28oCNF[a]	4	4	–
F62oCNF	8	–	–
F92oCNF	8	–	–
F96oCNF	8	–	–
F96oCNF I49T	8	–	–
F96oCNF M100D	4	4	–
F96oCNF M100E	8	–	–
F96oCNF M100H	8	–	–
F96oCNF M100K	8	–	–
F96oCNF M100N	4	2	2
F96oCNF M100Q	4	4	–
F96oCNF T103V	4	4	–

[a] In this case, different conformations means distinct *nitrile* conformers. Conformation A is defined as the F28oCNF nitrile that is closer to H_2O-A in the crystal structure.[1]

MD-based electric fields ($\left|\vec{F}\right|_{C\equiv N,MD}$) were determined as the mean electric field between the C and N atoms of the nitrile group along the –C≡N bond axis. In TINKER, the induced dipoles on the C and N atoms were saved every 10 ps and were divided by their polarizability parameters to provide the atomic electric fields. These values were dotted into the unit –C≡N bond vector to produce scalar electric fields and averaged to produce the reported $\left|\vec{F}\right|_{C\equiv N,MD}$ values. Note that absolute electric fields, or the fields directly outputted by the simulations, were used to avoid complications with assessing the self-field of the oCNF adduct in PYP[17] (see further details above in the description for obtaining MD fields of the BoCNF small molecule).

Choice of Protonation States in MD Simulations for PYPs with Newly Introduced Titratable Residues

IR spectroscopy performed at pH 6.0 and 8.5 for M100E, M100D, and M100H indicated minimal spectral changes, so we determined the E100 and D100 are carboxylates in solution, and H100 is net neutral (**Figure S9** and discussion). As such, E100 and D100 were simulated in their deprotonated form, and the results of these simulations are used in the main text. H100 remained ambiguous, however, as the proton can be at $N_{\delta 2}$ or $N_{\varepsilon 1}$. Crystallography for M100H (**Figure S19**) suggested H100's $N_{\delta 2}$ is protonated in solution, so the "HID" form of histidine was simulated and used in the main text. As a further test, 100 ns of MD simulations were also performed with the neutral, carboxylic acid forms of E100 and D100 and the "HIP" ($N_{\delta 2}$ and $N_{\varepsilon 1}$ protonated) form of H100 to test the impact on the nitrile environment. The nitrile electric field distributions or H-bonding fractions for these

simulations show clear deviations from M100N's distribution in main text **Figure 4** (**Figure S24** & **Table S17**), giving us confidence that the proper protonation states were used for benchmarking. This analysis highlights how comparisons of MD outputs against experimental observables helps to assign protonation states in solution.[21–24]

H-Bond Definitions and MD H-Bonding Characterizations

Throughout this work, H-bond distances and angles are defined by the heavy atoms, as indicated in **Figure S2**:

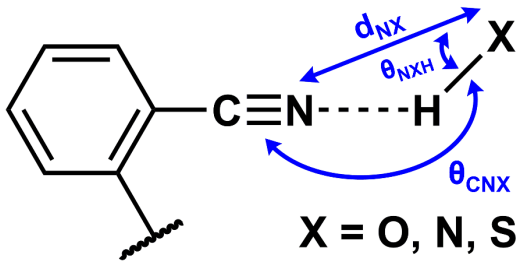

Figure S2. Geometric definitions for oCNF's nitrile's H-bonds. X refers to a heavy-atom H-bond donor. The dashed line indicates a potential H-bonding interaction. θ_{CNX} is the CNX angle, θ_{NXH} is the NXH angle, and d_{NX} is the distance from the nitrile nitrogen (N) to the heavy-atom H-bond donor (X).

In our prior work,[17] radial distribution functions of F28oCNF, F92oCNF, and F96oCNF suggested 4 Å was an appropriate H-bond distance (d_{NX}) cutoff, while previous studies of small molecule crystal structures suggested 30° as an appropriate cutoff for θ_{NXH} (referred to as an H-bond cone in the main text).[25] Further, we found empirically that it was best not to set a constraint on the H-bond angle (θ_{CNX}) so long as the other two constraints were satisfied. These same H-bond conditions were adopted here.

Using this H-bond definition, several types of MD characterization were performed. Histograms of nitrile H-bond distances (binned every 0.05 Å) and angles (binned every 30°) were generated in OriginPro 2021b and are shown in **Figures S29–S42**, along with contour plots of the distances vs the angles with contours describing the relative frequency of occurrence; parameters associated with 2D Gaussian fits of the contour plots are provided in **Tables S23–S25**. In addition, analysis of nitrile H-bond donor identities is provided in **Table S20**. Finally, nitrile H-bond residence times were extracted in an effort to obtain H-bonding and non-H-bonding exchange lifetimes, using the same method detailed in ref. [17], Section S1; the resulting analysis is provided in **Figures S46–S48** and **Table S34**.

Calculation of Nitrile H-Bond Shifts from MD Simulations

In our previous work, we found an explicit form to calculate the nitrile H-bond blueshift when the H-bond distance (d) and angle (θ) are known:[5]

$$\Delta\bar{\nu}_{HB}(d,\theta) = \Delta\bar{\nu}_{HB,0} \cdot \left\{ \left(\frac{d}{d_0}\right)^{-4} \cdot \cos[m \cdot (\theta - 180°)] + \left[e^{-b(d-d_0)} - \left(\frac{d}{d_0}\right)^{-8} \right] \cdot [1 - \cos[m \cdot (\theta - 180°)]] \right\} \quad \text{(S1)}$$

where $\Delta\bar{\nu}_{HB,0} = 16.6$ cm^{-1} is the H-bond blueshift at a reference distance (d_0 = 3.36 Å), m = 0.91 is the cosine period modulation, and b = 2.85 Å^{-1} is the exponential decay constant. Further, comparisons of experimentally and computationally derived H-bond shifts indicated protein H-bonds follow eq S1 exactly, while solvent H-bond are approximately one half the value from eq S1 due to motional narrowing.[5] As such, in this work, protein H-bonds take on the full predicted value, while solvent H-bonds are half the predicted value, unless otherwise indicated.

S4 BoCNF IR Spectra, MD Electric Fields, and Stark Tuning Rates

BoCNF Nitrile Spectra

We previously performed field-frequency and field-TDM calibrations for oTN to extract electric fields and H-bond blueshifts for oCNF incorporated into PYP.[1] To ensure these calibrations are appropriate, we obtained calibrations for a Boc-protected version of oCNF, N-Boc-2-cyano-L-phenylalanine (BoCNF; **Figure S1**). FTIR spectra of BoCNF were measured in a variety of aprotic solvents and water (**Figure S3**). AMOEBA MD simulations with AMOEBA09[9] were performed in the same solvents to obtain the average solvent electric field, $|\vec{F}|$, projected onto the –C≡N probe. Correlation of the aprotic peak frequencies (**Table S3**) versus the average $|\vec{F}|$ values (**Table S4**) exhibits a strong linear trend (**Figure S4A**), and the slope of the best fit line is the frequency Stark tuning rate (0.18 $\frac{cm^{-1}}{\frac{MV}{cm}}$). In contrast, the nitrile frequency in water, 2230.9 cm^{-1}, is bluer than that in hexanes, 2228.7 cm^{-1}, despite a significantly larger average MD solvent electric field (the field is 50 MV/cm larger in magnitude in water; **Table S4**), highlighting the H-bond blueshift (**Figure S4A**). A linear correlation is observed between $|\vec{m}|$ (**Table S3**) and the average $|\vec{F}|$ values (**Figure S4B)**, as previously observed.[1] We found oTN and BoCNF have very similar frequency and TDM calibration curves (see **Figure S5** and discussion); this suggests that model compounds may reasonably be used to calibrate VSE probes in place of ligands or amino acids which present solubility and/or simulation challenges.[10]

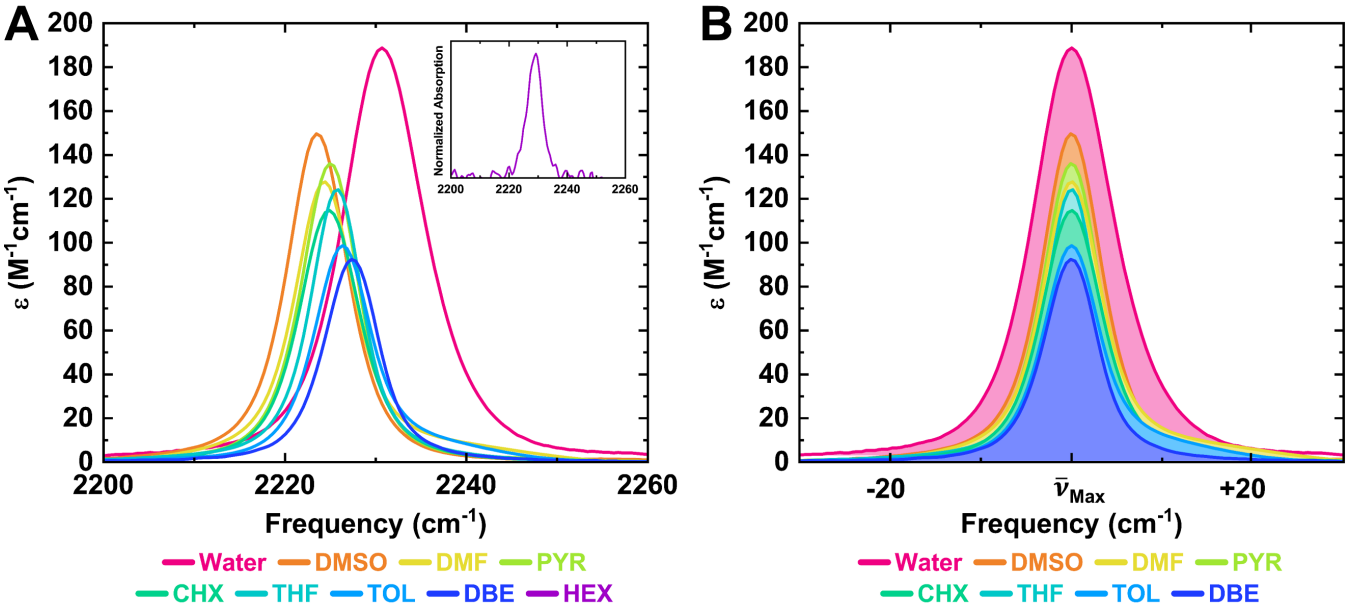

Figure S3. Room-temperature IR solvatochromism of BoCNF nitrile normalized to extinction coefficients. (A) Significant peak area tuning is observed in protic and aprotic solvents as previously observed for oCNF model compound oTN.[1] Inset shows normalized absorption in HEX, where extinction coefficients could not be determined (see **Section S1**). (B) Overlay of IR spectra of (A) where bands have been centered at their absorption maxima ($\bar{\nu}_{Max}$) and shaded to highlight peak area changes.

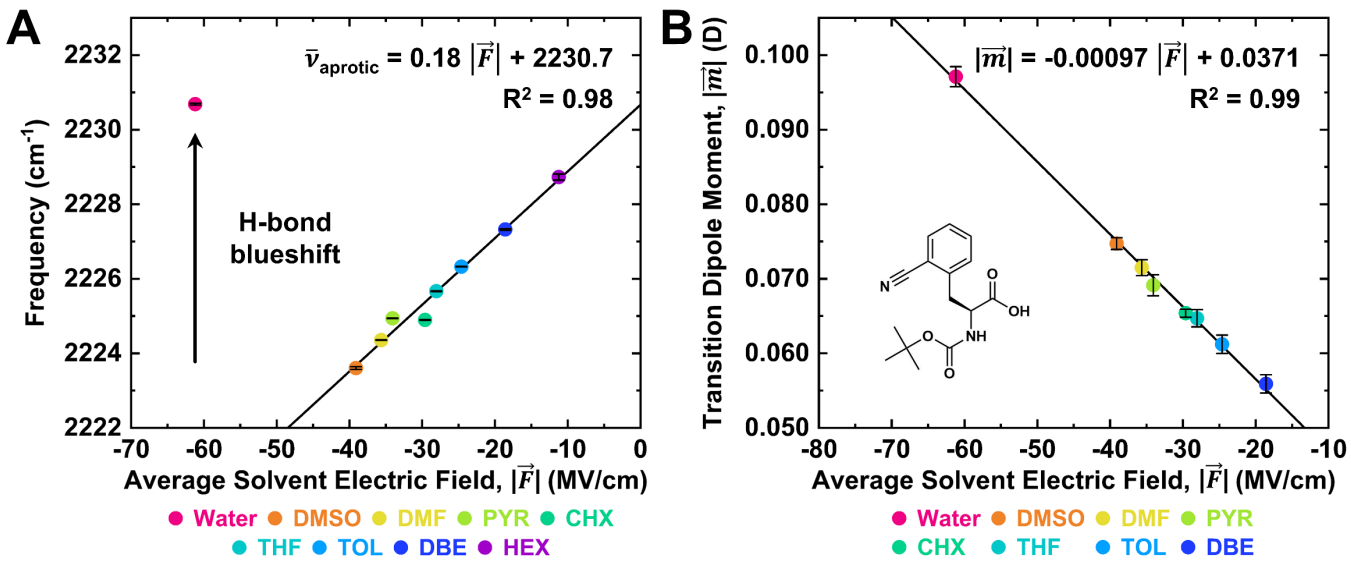

Figure S4. Responses of BoCNF frequency and TDM to average MD solvent electric field. (A) Peak frequency versus average solvent electric field obtained in aprotic solvents and water using AMOEBA MD simulations. The linear fit includes only the aprotic points, as the nitrile frequency deviates from the VSE in water. (B) TDM, $|\vec{m}|$, versus average solvent electric field. The slopes of the lines in (A) and (B) are the frequency and TDM Stark tuning rates, respectively, while the intercepts are the zero-field (gas phase) frequency and TDM. Error bars represent 1σ standard deviations. Raw spectra used to extract frequencies and TDMs are shown in **Figure S3**.

BoCNF IR Observables and Electric Fields Calculated from MD

Table S3. Extracted IR peak parameters for BoCNF as a function of solvent.

Environment	Peak (cm^{-1})	FWHM (cm^{-1})	$\lvert\vec{m}\rvert$ (D)	ε_{max} ($M^{-1}cm^{-1}$)[a]
Water	2230.88 ± 0.02	10.72 ± 0.04	0.0971 ± 0.0013	188.7
DMSO	2223.61 ± 0.04	8.02 ± 0.05	0.0747 ± 0.0008	149.7
DMF	2224.35 ± 0.01	7.83 ± 0.04	0.0715 ± 0.0011	127.8
PYR	2224.94 ± 0.01	7.53 ± 0.03	0.0691 ± 0.0014	136.1
CHX	2224.90 ± 0.01	7.97 ± 0.01	0.0654 ± 0.0005	114.7
THF	2225.67 ± 0.01	6.79 ± 0.02	0.0647 ± 0.0012	124.3
TOL	2226.33 ± 0.01	7.46 ± 0.01	0.0612 ± 0.0012	98.6
DBE	2227.32 ± 0.02	7.13 ± 0.04	0.0559 ± 0.0012	92.5
HEX	2228.73 ± 0.08	6.6 ± 0.3	ND	ND

[a] Provided for spectra shown in **Figure S3**.

Table S4. Average calculated solvent electric fields for BoCNF using AMOEBA MD simulations. BoCNF was simulated with a C-terminal carboxylate in water and with a C-terminal carboxylic acid in the organic solvents.

Environment	Environmental $\lvert\vec{F}\rvert$ (MV/cm)
Water	-61.2
DMSO	-39.1
DMF	-35.6
PYR	-34.1
CHX	-29.6
THF	-28.1
TOL	-24.6
DBE	-18.6
HEX	-11.2

Comparison Between oTN and BoCNF Electrooptic Parameters from Experiment and DFT

To the best of our knowledge, our BoCNF calibrations (**Figure S4**) are the first report for the vibrational Stark tuning rate(s) of a nitrile on an amino acid calibrated via solvatochromism; prior calibrations have been performed with vibrational Stark spectroscopy for amino acids in solution[26] and incorporated into peptides and proteins.[26–28] Previously determined calibrations for oTN[1] are $\bar{\nu}_{\mathrm{aprotic}}$ = 0.19$\left|\vec{F}\right|$ + 2231.4 and $|\vec{m}|$ = -0.00092$\left|\vec{F}\right|$ + 0.0388, while for BoCNF we found $\bar{\nu}_{\mathrm{aprotic}}$ = 0.18$\left|\vec{F}\right|$ + 2230.7 and $|\vec{m}|$ = -0.00097$\left|\vec{F}\right|$ + 0.0371 (**Figure S5**; black solid lines for BoCNF and red dotted lines for oTN; frequencies in cm^{-1} and TDMs in D). The absolute differences in the calibrations are small, and the calibrations would therefore give rise to similar inferred electric fields and H-bond shifts when employed. At the same time, we were curious if the changes may be significantly different. The error in the y-intercept of the BoCNF frequency-field calibration is ± 0.1 cm^{-1} (**Figure S5**), and assuming the same error for oTN, the 0.7 cm^{-1} difference in the zero-field frequencies from solvatochrimc calibrations is significant out to 4σ. In contrast, the error in the zero-field TDM for BoCNF is ± 0.007 D (**Figure S5**), so the difference in the values between oTN and BoCNF is only significant at the 1σ level using the same error for oTN. Analogously, the BoCNF TDM and frequency Stark tuning rates have errors of 0.0002 $\frac{D}{\frac{MV}{cm}}$ and 0.005 $\frac{cm^{-1}}{\frac{MV}{cm}}$, respectively. Once again assuming oTN has the same errors, the Stark tuning rates are significantly different at the 1σ level and not significantly different, respectively. Overall, the comparison between the two compounds' parameters with errors considered further highlights their high level of similarity.

We were curious if DFT calculations could capture the difference in the zero-field frequencies and the similarity in the zero-field TDMs, so we performed gas phase geometry optimization and vibrational frequency analysis at the b3lyp/6-311++g(d,p) level with Gaussian16.[13] The calculations yielded (unscaled) values for $\bar{\nu}_0$ and $|\vec{m}_0|$ of 2326.3 cm^{-1}/0.0789 D and 2324.5 cm^{-1}/0.0786 D for the oTN and BoCNF nitriles, respectively. The gas phase frequency and TDM for BoCNF (and oTN) are larger than the experimental values, but to very different degrees: for BoCNF, the frequency is overestimated by 4.2%, while the TDM is overestimated by 102% (consistent with previous TDM calculations[5]), and the DFT TDM value falls between the experimental TDMs in DMSO and water (**Table S3**). It may be that more absolutely accurate TDM calculations require an alternative approach to DFT, such as explicitly representing the vibrational wavefunctions.[29]

Using the scaling factor for harmonic frequencies at this level of theory (0.9679),[30] the predicted difference in zero-field frequencies from DFT for oTN minus BoCNF is 1.8 cm^{-1}; this represents impressive accuracy, as the difference in the extrapolated values from experimental solvatochromism is 0.7 cm^{-1}. For the DFT zero-field TDMs, the (unscaled) difference in the predicted values for oTN minus BoCNF is 0.0003 D, close to the magnitude of the experimental error. Together, these results suggest DFT can accurately predict the relative difference in zero-field frequencies and can discern when the TDMs are only marginally different for highly similar aromatic nitrile-bearing compounds.

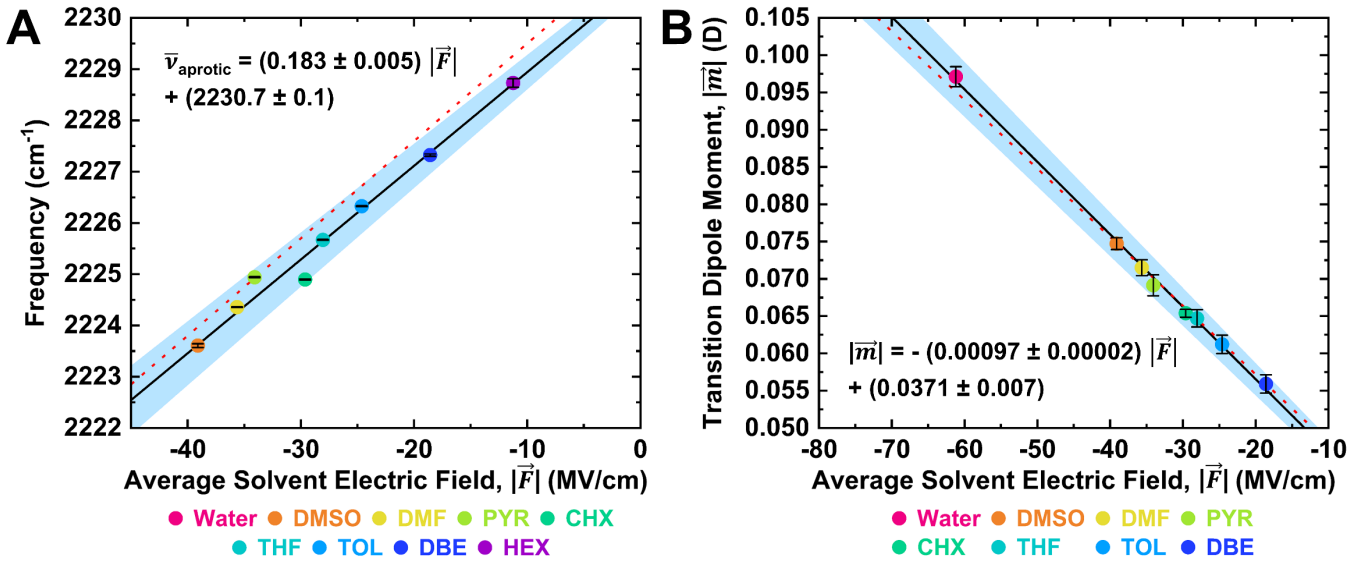

Figure S5. BoCNF (A) frequency- and (B) TDM-field calibrations with 1σ errors shown in the equations of best fit. The light blue region indicates the 2σ confidence interval. oTN calibrations are shown as red dotted lines.

S5 Protein UV-Vis Spectra and Characterization

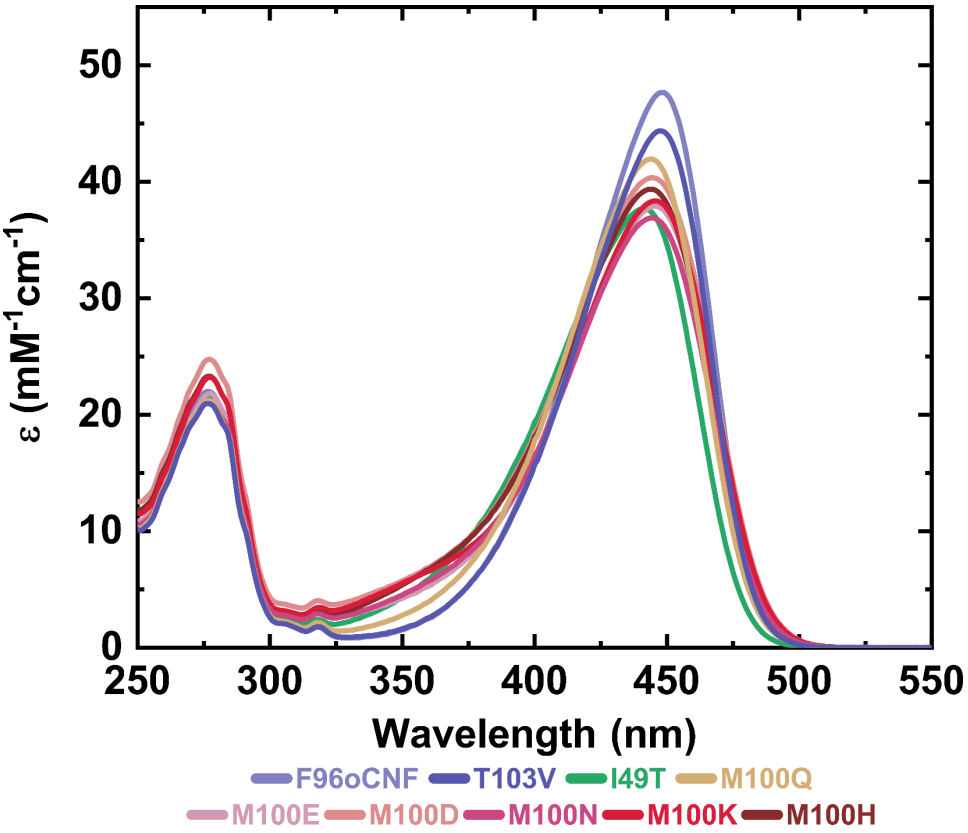

Figure S6. Concentration normalized UV-Vis spectra of dark-adapted F96oCNF and mutants at pH 6.0. Extinction coefficients were determined by taking UV-Vis spectra of folded proteins, denaturing the proteins in guanidinium chloride, and using the previously determined chromophore extinction coefficient in guanidinium chloride to determine concentrations of folded proteins (see **Section S2**; method analogous to base denaturation used for GFP extinction coefficient determination[31,32]). Mutations do not affect the absorption of the chromophore when the protein is unfolded, *so changes to the visible absorption spectrum of a folded protein do not affect the ability to determine protein, and therefore nitrile, concentration*.

All PYP mutants near position 96 (main text **Figure 1**) display bluer absorption maxima and decreased extinction coefficient at absorption maxima compared to F96oCNF (**Table S5**). Part of the explanation for the decreased ε_{max} values in the mutants is the appearance of the protonated pCA chromophore to the blue of the typically deprotonated pCA absorption in the wild type protein. This has been previously observed in (non-nitrile-containing) PYP M100 mutants, M100A,[33] M100L,[34,35] M100K,[35] and M100E,[4,35,36] the latter two of which were made in this study (with oCNF96 background). oCNF96 and all mutations near position 96 are near the pCA chromophore (main text **Figure 1**); given the changes to the nitrile IR absorption in the mutants, corresponding changes in pCA's visible absorption are unsurprising. Additionally, exposure of the F96oCNF M100 mutants to room light induced a bleach of the deprotonated chromophore band and a gain in intensity of the protonated chromophore band (see examples in ref. [4]), and this observation led us to dark-adapt all proteins prior to the measurement of IR spectra (**Section S1**). This bleaching phenomenon has been observed in the M100 mutants listed above and attributed to accumulation of the final photocycle intermediate pB (also occasionally referred to

as I_2 or PYP_M) at steady-state, explained by a substantial (one to three orders of magnitude) decrease in rate for pB to pG (*i.e.*, dark-adapted ground state) conversion from the wild type PYP kinetics.[33,35] This behavior is in contrast to wild type PYP, which requires a significant laser flash intensity to induce substantial bleaching.[33]

Table S5. Absorption maxima and extinction coefficients for pCA in F96oCNF and mutants. Measurements were performed at pH 6.0 unless specified otherwise.

PYP Construct	λ_{max} (nm)	ε_{max} ($mM^{-1}cm^{-1}$)
F96oCNF[a]	449	47.7
T103V	448	44.4
I49T	441	37.7
M100Q	444	42.0
M100E	445	37.9
M100E, pH 8.5	446	40.2
M100D	444	40.3
M100D, pH 8.5	445	44.1
M100N	444	36.9
M100K	446	38.4
M100H	444	39.4
M100H, pH 8.5	445	43.5

[a] Taken from ref. [1].

S6 Protein IR Spectra and Extracted Parameters

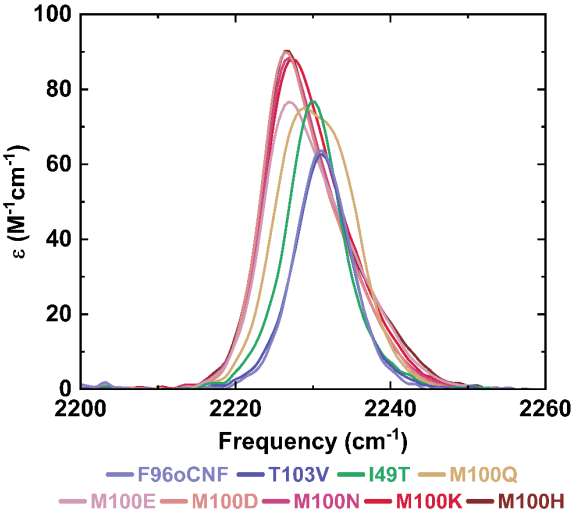

Figure S7. Overlay of IR spectra for the nitriles in F96oCNF and all mutants.

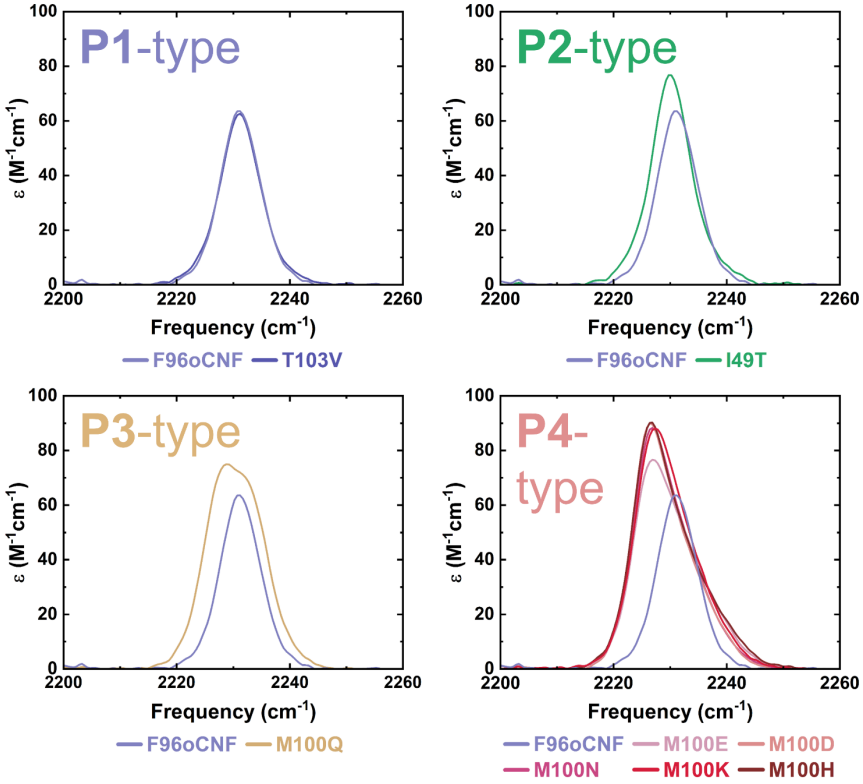

Figure S8. IR spectra for F96oCNF mutants categorized by phenotype, with F96oCNF shown in each panel for comparison. In all phenotypes, the nitrile integrated peak area with respect to F96oCNF has remained constant (P1-type) or increased (P2-, P3-, and P4-type), indicating that all mutants experience a nitrile electric field greater than or equal to that in F96oCNF.

Table S6. Extracted IR peak parameters for the oCNF nitrile incorporated in PYPs. Phenotype categorizations are indicated by coloring: P1, blue; P2, green; P3, gold; P4, red/pink.

Environment	$\bar{\nu}_{obs}$ (cm^{-1})[a]	FWHM (cm^{-1})[a]	Total $\lvert\vec{m}\rvert$ (D)	$\lvert\vec{F}\rvert_{C\equiv N,TDM}$ (MV/cm)	$\Delta\bar{\nu}_{non\text{-}HB}$ (cm^{-1})	$\Delta\bar{\nu}_{HB}$ (cm^{-1})
F28oCNF[b]	2230.8	12.7	0.075 ± 0.002	-39 ± 2	-8	7
F62oCNF[b]	2228.1	6.5	0.053 ± 0.002	-15 ± 2	-3	0
F92oCNF[b]	2241.3	14.4	0.094 ± 0.002	-60 ± 2	-11	21
F96oCNF[b,c]	2231.2	8.4	0.047 ± 0.002	-9 ± 2	-2	2
T103V	2231.2	8.5	0.047 ± 0.002	-9 ± 2	-2	2
I49T[c]	2230.2	8.3	0.052 ± 0.002	-14 ± 2	-3	2
M100Q[c]	—	—	0.060 ± 0.002	-23 ± 2	—	—
M100E	—	—	0.062 ± 0.002	-25 ± 2	—	—
M100D[c]	—	—	0.063 ± 0.002	-26 ± 2	—	—
M100N	—	—	0.063 ± 0.002	-26 ± 2	—	—
M100K	—	—	0.064 ± 0.002	-27 ± 2	—	—
M100H	—	—	0.065 ± 0.002	-28 ± 2	—	—

[a] Peak positions ($\bar{\nu}_{obs}$) and FWHMs have an error of <0.1 cm^{-1}.
[b] Reproduced/updated from ref. [1].
[c] Reproduced from main text **Table 1**.

IR Spectra at pH 6.0 and 8.5

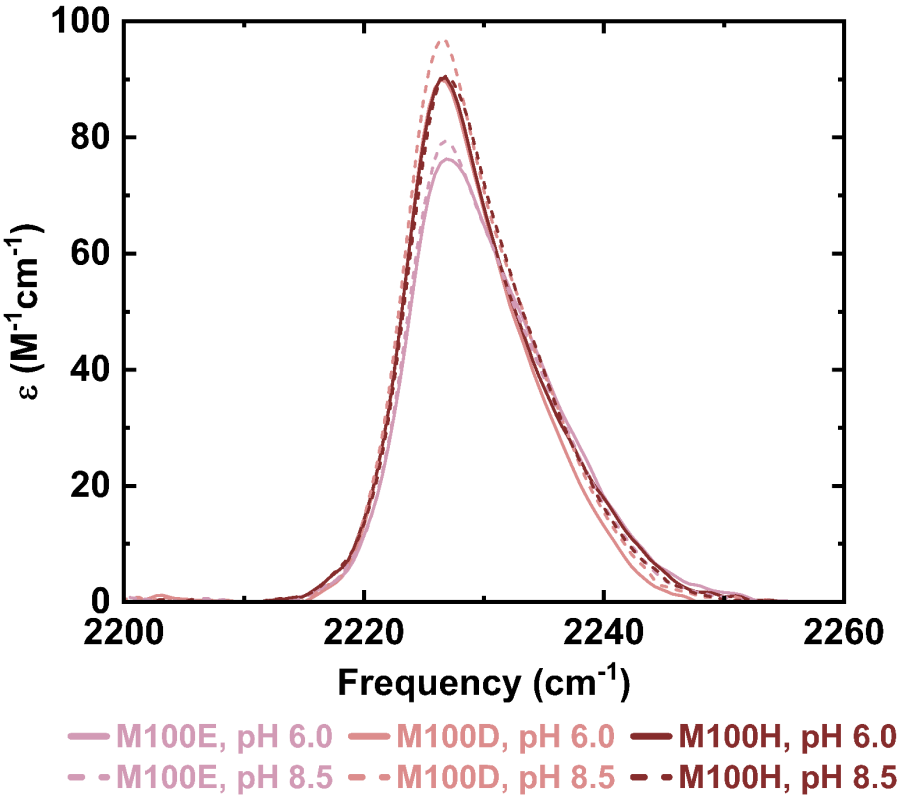

Figure S9. Comparison of nitrile IR spectra for PYP mutants M100E, M100D, and M100H at pH's 6.0 and 8.5. Electric field analysis for pH 8.5 spectra can be found in **Table S7**.

FTIR was performed at two pH values to assess whether possible changes in protonation of the titratable residues in M100E, M100D, and M100H would affect the nitrile's absorption spectra. No pH dependent change is observed in the spectral shape of any mutants (**Figure S9**). The small changes observed for the TDMs and electric fields at pH 8.5 (**Table S7**) as compared with pH 6.0 are likely a consequence of small errors in the concentration or small baselining differences. The spectra indicate that E100 and D100 are deprotonated and H100 is neutral in solution**.** We attempted to further corroborate these assignments by obtaining extinction coefficient normalized IR spectra below pH 6.0, but these attempts were unsuccessful due to protein precipitation. This experiment highlights vibrational spectroscopy's ability to provide information on local protonation states in comparison with X-ray crystallography, where very high resolutions are needed to model protons, and MD simulations, where proton placement must be assigned.

Table S7. Extracted IR peak parameters for the oCNF96 nitrile incorporated in F96oCNF mutants at pH 8.5.

Environment	Total $\lvert\vec{m}\rvert$ (D)	Avg. $\lvert\vec{F}\rvert_{C\equiv N,TDM}$ (MV/cm)
M100E	0.062 ± 0.002	-25 ± 2
M100D	0.066 ± 0.002	-30 ± 2
M100H	0.065 ± 0.002	-28 ± 2

IR Band Fitting for P3- and P4-Type Nitrile Spectra

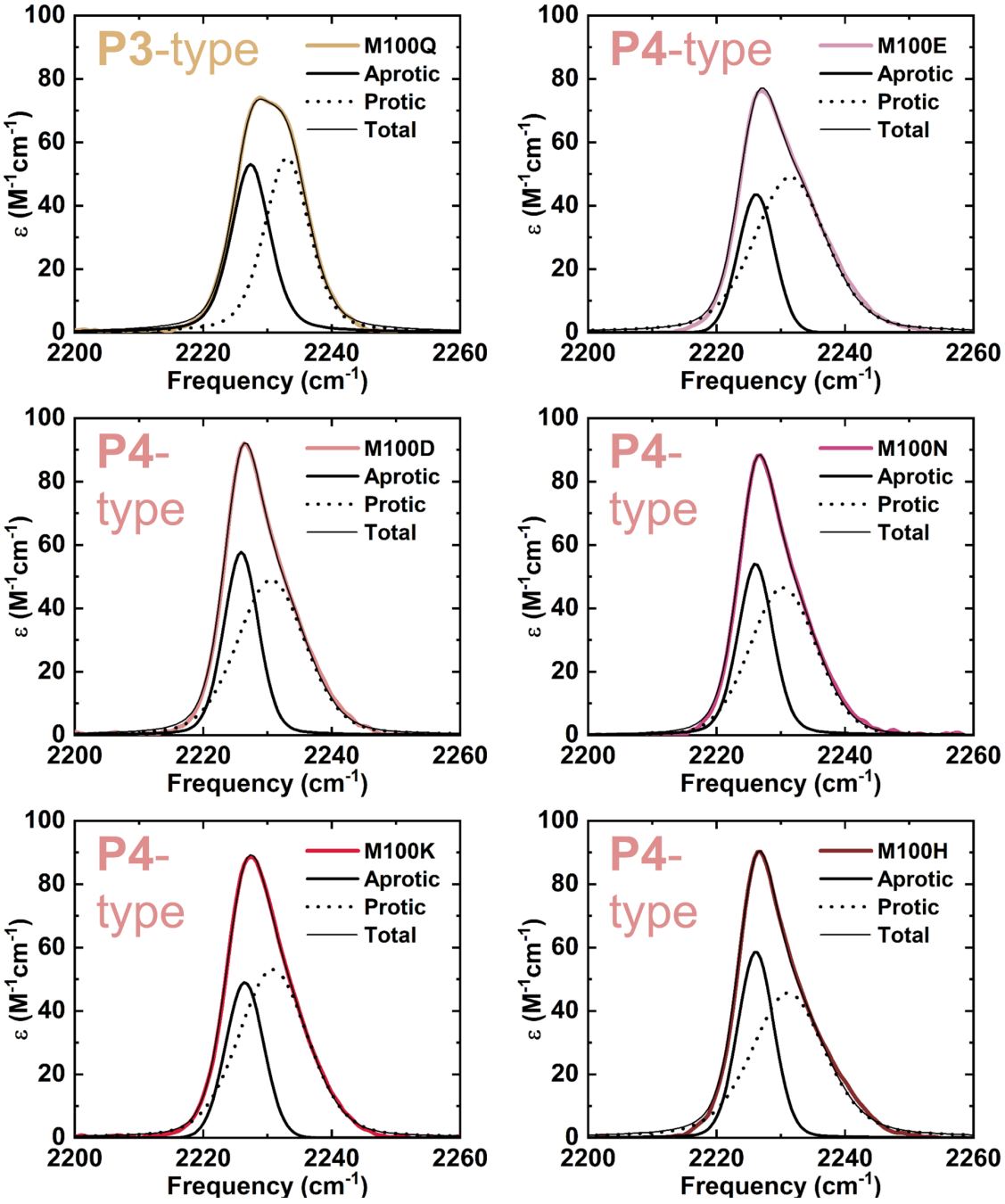

Figure S10. Two band fits for P3- and P4-type nitriles.

Table S8. Extracted aprotic and protic population IR peak parameters for the oCNF96 nitrile incorporated in P3 and P4 PYPs.

Environment	$\bar{\nu}_{obs}$ (cm^{-1})	FWHM (cm^{-1})
$M100Q_{aprotic}$[a]	2227.6 ± 0.2	7.4 ± 0.2
$M100Q_{protic}$[b]	2233.4 ± 0.3	8.0 ± 0.2
$M100E_{aprotic}$[c]	2226.16 ± 0.03	6.65 ± 0.02
$M100E_{protic}$[d]	2231.3 ± 0.1	13.4 ± 0.1
$M100D_{aprotic}$[c]	2225.94 ± 0.04	6.6 ± 0.1
$M100D_{protic}$[d]	2230.7 ± 0.1	12.7 ± 0.2
$M100N_{aprotic}$[c]	2226.08 ± 0.02	6.5 ± 0.1
$M100N_{protic}$[d]	2230.6 ± 0.2	12.1 ± 0.1
$M100K_{aprotic}$[c]	2226.49 ± 0.03	7.1 ± 0.1
$M100K_{protic}$[d]	2230.5 ± 0.1	12.8 ± 0.1
$M100H_{aprotic}$[c]	2226.11 ± 0.01	6.63 ± 0.03
$M100H_{protic}$[d]	2231.14 ± 0.02	13.5 ± 0.1

[a] Fits had 0-45% Lorentzian contribution.

[b] Fits had 30-55% Lorentzian contribution.

[c] Fits had 0-40% Lorentzian contribution.

[d] Fits had 10-50% Lorentzian contribution.

S7 X-ray Crystallography

Protein Crystallization

PYP crystallization was performed as before.[1] PYPs were buffer exchanged into an aqueous buffer at pH 6.0, 20 mM KH_2PO_4,[37] concentrated to 20 mg/mL, and set as hanging drops on a coverslip (1 µL of protein mixed with 1 µL of mother liquor; wells contained 500 µL mother liquor). Well solutions contained 1 M NaCl and a variable amount of $(NH_4)_2SO_4$ (1.8–2.6 M in 0.1 M increments). Hanging drops were typically seeded with crushed crystals from WT PYP or F96oCNF PYP which resulted in crystal formation within two days. Crystal trays were kept at room temperature and covered in aluminum foil to minimize exposure to light.

Crystals of the F96oCNF M100X mutants were particularly sensitive to light, displaying cracks if excessively exposed. This may be related to the altered photocycle kinetics of the M100 mutants in ambient light described under **Figure S6**. As a precaution, all F96oCNF mutants were set and seeded at low light exposure (i.e., no overhead lighting). The M100H mutant was particularly sensitive, so handling was done under red light.

Crystals of sufficient quality were looped, dipped in polyfluoroether cryo-oil (Hampton Research), and flash frozen into a Stanford Synchrotron Radiation Lightsource (SSRL) cassette immersed in liquid nitrogen. Looping at low light exposure was found sufficient to maintain good crystal diffraction. The $(NH_4)_2SO_4$ concentration leading to crystal diffraction at the highest resolution for each mutant is shown in **Table S9**.

Table S9. Optimal ammonium sulfate concentration for crystallization of each F96oCNF PYP mutant.

F96oCNF PYP Mutant	$(NH_4)_2SO_4$ Concentration (M)
T103V	2.0
I49T	2.0
M100Q	2.1
M100E	2.2
M100D	2.4
M100N	2.6
M100K	2.5
M100H	2.5

X-Ray Data Collection and Structure Refinement

X-ray diffraction was performed at SSRL (Menlo Park, CA) Beam Line 12-1 at 100K. Data collection was performed using Blu-Ice.[38] Data processing was performed using the X-ray Detector Software (XDS)[39,40] with the autoxds script.[41] Molecular replacement was performed in Phenix[42] with phenix.phaser using WT PYP (PDB: 1NWZ) as the search model for all mutants. Coot[43] was used to perform numerous rounds of model building and refinement followed by automated refinement via phenix.refine. PDB-REDO[44] was occasionally used to further refine structures. Refinement continued until the R-free score could no longer be lowered. Resulting PDB codes and data collection/refinement statistics can be found in **Table S10**.

Table S10. X-ray data collection and refinement statistics for F96oCNF mutants.

	T103V	I49T	M100Q	M100E
PDB Entry	8E02	8DZY	8E1K	8E09
Data collection statistics				
Beamline	BL 12-1			
Wavelength (Å)	0.7126			
Detector distance (mm)	150			
Resolution range (Å)	33.20 – 1.09 (1.12 – 1.09)	40.90 – 1.01 (1.04 – 1.01)	40.57 – 1.02 (1.04 – 1.02)	40.61 – 1.01 (1.04 – 1.01)
Space group	P 6_3 (No. 173)			
Unit cell dimensions a, b, c (Å) α, β, γ (°)	66.30, 66.30, 40.69 90.0, 90.0, 120.0	65.91, 65.91, 40.96 90.0, 90.0, 120.0	66.36, 66.36, 40.57 90.0, 90.0, 120.0	66.21, 66.21, 40.61 90.0, 90.0, 120.0
Total observations	1720362 (114308)	2119169 (119891)	2068596 (114016)	2179198 (130346)
Unique reflections	41659 (3025)	51960 (3724)	50640 (3668)	53209 (3930)
Redundancy	41.3 (37.8)	40.8 (32.2)	40.8 (31.1)	41.0 (33.2)
Completeness (%)	99.9 (98.4)	99.8 (97.9)	99.9 (98.7)	100 (99.5)
Mean I/σI	23.8 (3.5)	16.4 (1.6)	19.4 (1.5)	16.3 (1.5)
Wilson B-factor (Å^2)	9.94	8.78	10.7	8.89
R_{meas}	0.123 (1.26)	0.133 (1.58)	0.102 (2.04)	0.132 (2.20)
$CC_{1/2}$	1 (0.904)	0.999 (0.754)	1 (0.667)	0.999 (0.700)

	T103V	I49T	M100Q	M100E
Refinement statistics				
Reflections used in refinement	41659	51951	50637	53177
Reflections used for R_{free}	2083	2584	2527	2671
R_{work}	0.1401	0.1478	0.1477	0.1460
R_{free}	0.1701	0.1604	0.1596	0.1614
Number of non-H Atoms: Protein Ligands Solvent	1235 1089 24 122	1308 1160 24 124	1291 1167 24 100	1277 1151 24 102
Protein residues	124	124	123	123
RMSD bond lengths (Å)	0.013	0.009	0.007	0.013
RMSD bond angles (°)	1.57	1.46	1.26	1.58
Ramachandran favored (%)	96.67	97.50	96.64	95.80
Ramachandran allowed (%)	3.33	2.50	3.36	4.20
Ramachandran outliers (%)	0.00	0.00	0.00	0.00
Rotamer outliers (%)	2.59	0.80	0.00	0.82
Clashscore	5.96	2.59	3.88	2.19
Average B-factor (Å^2): Protein Ligand Solvent	13.89 13.27 7.18 20.71	13.88 13.33 7.77 20.22	15.52 15.10 8.88 21.95	14.03 13.60 8.94 20.07

	M100D	M100N	M100K	M100H
PDB Entry	8E03	8E1L	8DZX	8DZU
Data collection statistics				
Beamline	BL 12-1			
Wavelength (Å)	0.7126			
Detector distance (mm)	150			
Resolution range (Å)	33.26 – 1.24 (1.28 – 1.24)	33.22 – 1.20 (1.24 – 1.20)	41.40 – 1.14 (1.17 – 1.14)	41.10 – 1.04 (1.06 – 1.04)
Space group	P 6_3 (No. 173)			
Unit cell dimensions a, b, c (Å) α, β, γ (°)	65.71, 65.71, 40.99 90.0, 90.0, 120.0	66.27, 66.27, 40.74 90.0, 90.0, 120.0	65.76, 65.76, 41.19 90.0, 90.0, 120.0	66.25, 66.25, 40.80 90.0, 90.0, 120.0
Total observations	1158678 (72683)	1294756 (84237)	1498489 (96486)	1966115 (113071)
Unique reflections	28191 (2020)	31300 (2284)	36301 (2625)	47986 (3460)
Redundancy	41.1 (36.0)	41.4 (36.9)	41.3 (36.8)	41.0 (32.7)
Completeness (%)	99.9 (98.8)	100 (99.8)	99.7 (97.6)	99.9 (98.9)
Mean I/σI	21.3 (1.5)	14.2 (1.5)	20.7 (1.5)	17.1 (1.6)
Wilson B-factor (Å^2)	13.7	11.7	11.8	9.52
R_{meas}	0.104 (2.08)	0.175 (1.61)	0.113 (1.59)	0.125 (2.04)
$CC_{1/2}$	1 (0.675)	0.999 (0.769)	1 (0.778)	1 (0.713)

	M100D	M100N	M100K	M100H
Refinement statistics				
Reflections used in refinement	28189	31297	36289	47981
Reflections used for R_{free}	1392	1547	1802	2405
R_{work}	0.1500	0.1601	0.1637	0.1453
R_{free}	0.1797	0.1877	0.1878	0.1620
Number of non-H Atoms: Protein Ligands Solvent	1266 1164 24 78	1271 1145 24 102	1235 1124 37 87	1234 1103 24 107
Protein residues	123	124	123	123
RMSD bond lengths (Å)	0.012	0.011	0.011	0.011
RMSD bond angles (°)	1.50	1.50	1.50	1.51
Ramachandran favored (%)	95.80	97.50	95.80	96.64
Ramachandran allowed (%)	4.20	2.50	4.20	3.36
Ramachandran outliers (%)	0.00	0.00	0.00	0.00
Rotamer outliers (%)	1.60	1.63	2.48	1.69
Clashscore	2.17	6.58	1.34	3.64
Average B-factor (Å^2): Protein Ligand Solvent	18.25 17.74 12.29 27.59	15.99 15.38 11.32 23.95	19.04 18.78 11.06 24.61	13.50 12.92 8.12 20.64

PYP Structures Focusing on Tertiary and Secondary Structure Perspectives

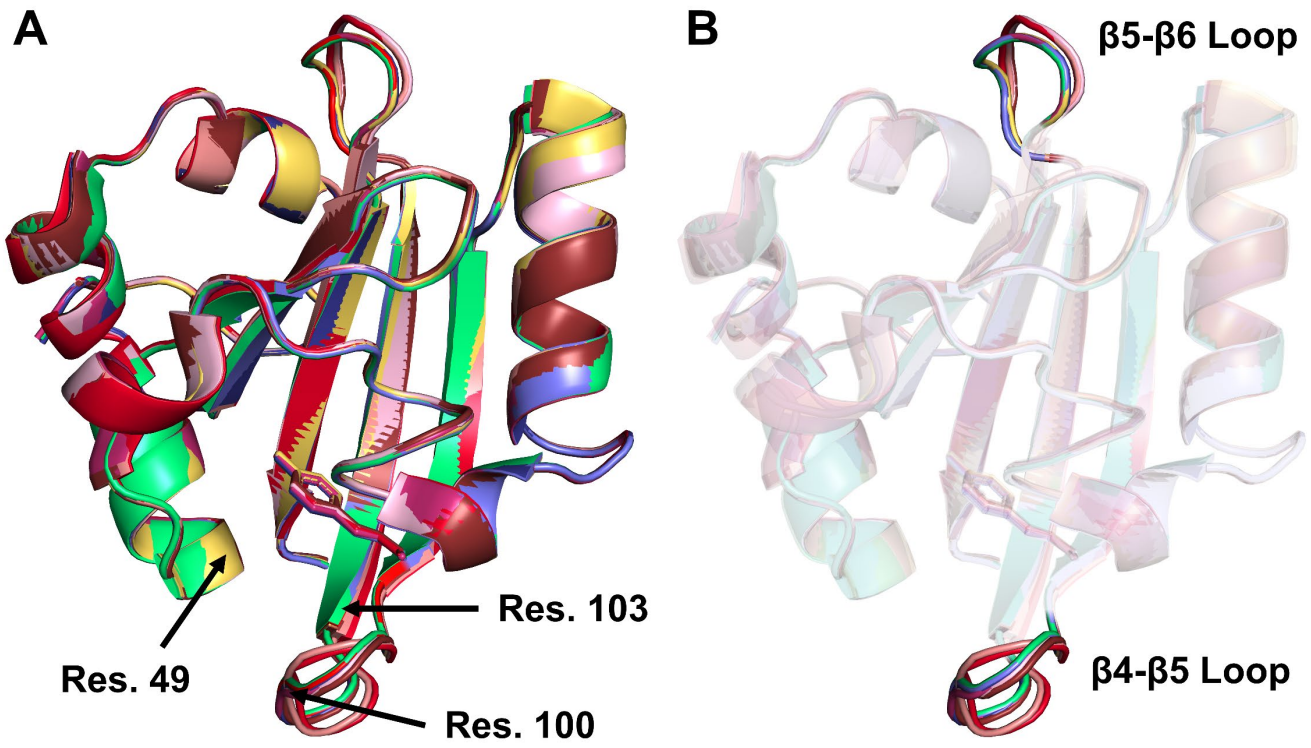

Figure S11. Overlay of crystal structures of F96oCNF and mutants. (A) Global structural alignment indicates the tertiary structure is highly conserved. The locations of residues 49, 100, and 103 are denoted. The chromophore pCA is shown in the stick representation. (B) Transparent view of the PYPs highlighting the most variable regions, the β4-β5 loop (residues 97-101 in F96oCNF; for nomenclature see Figure 2 in ref. [45]) and the β5-β6 loop (residues 112-117). The β4-β5 loop contains the M100X residue, and mutations at that site result in the largest structural deviations from F96oCNF (**Table S11**). Closer views of the β4-β5 loops can be found in **Figure S12**.

Table S11. Root-mean-square deviations (RMSD) for global alignment between F96oCNF and mutants.

F96oCNF PYP Mutant	RMSD (Å)
T103V	0.10
I49T	0.07
M100Q	0.17
M100E	0.19
M100D	0.20
M100N	0.15
M100K	0.23
M100H	0.21

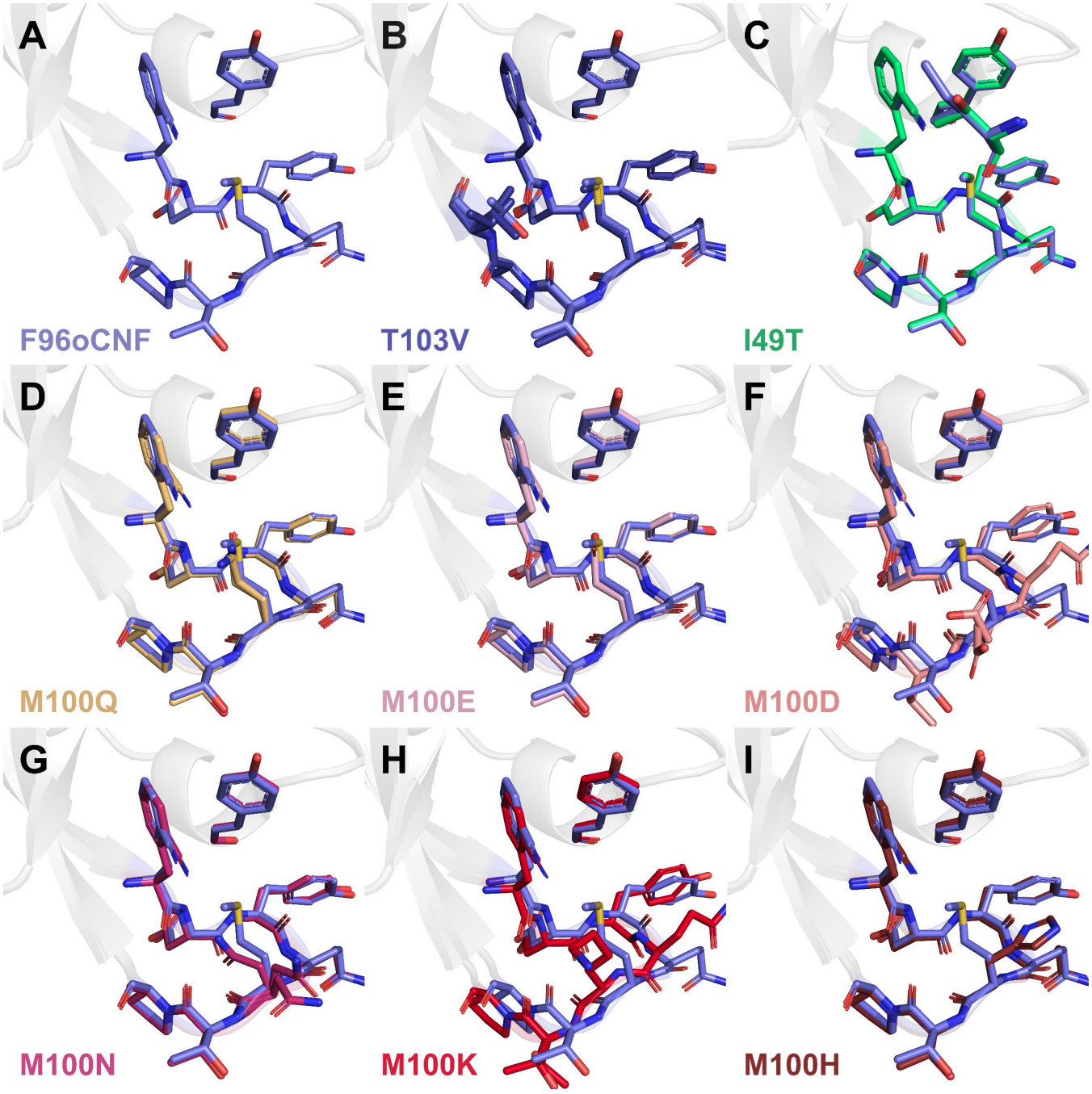

Figure S12. View of the β4-β5 loop and pCA chromophore for (A) F96oCNF and F96oCNF globally aligned with (B) T103V, (C) I49T, (D) M100Q, (E) M100E, (F) M100D, (G) M100N, (H) M100K, and (I) M100H. T/V103 and I/T49 are also displayed in (B) and (C), respectively. In (B) and (C), oCNF96 of F96oCNF and the mutants overlap, while in (D)–(I) the mutant's nitrile is tilted in toward the mutated residue. In M100D (F) and M100K (H), alternative β4-β5 loop conformations are observed. M100D and M100K have an RMSD of 0.11 Å, lower than the RMSD of either protein with F96oCNF (**Table S11**), and inspection of panels F and H indicates their β4-β5 loop conformations are similar. Changes to the loop from low temperature crystal structures have been observed when comparing wild type (WT) PYP, which possess the same β4-β5 loop conformation as F96oCNF, with the Y98Q mutant[46] or with PYP from Rc. centenum;[45,47] neither of those conformations are the same as that adopted by M100D and M100K.

Main Text Figure 3 for all F96oCNF Mutants

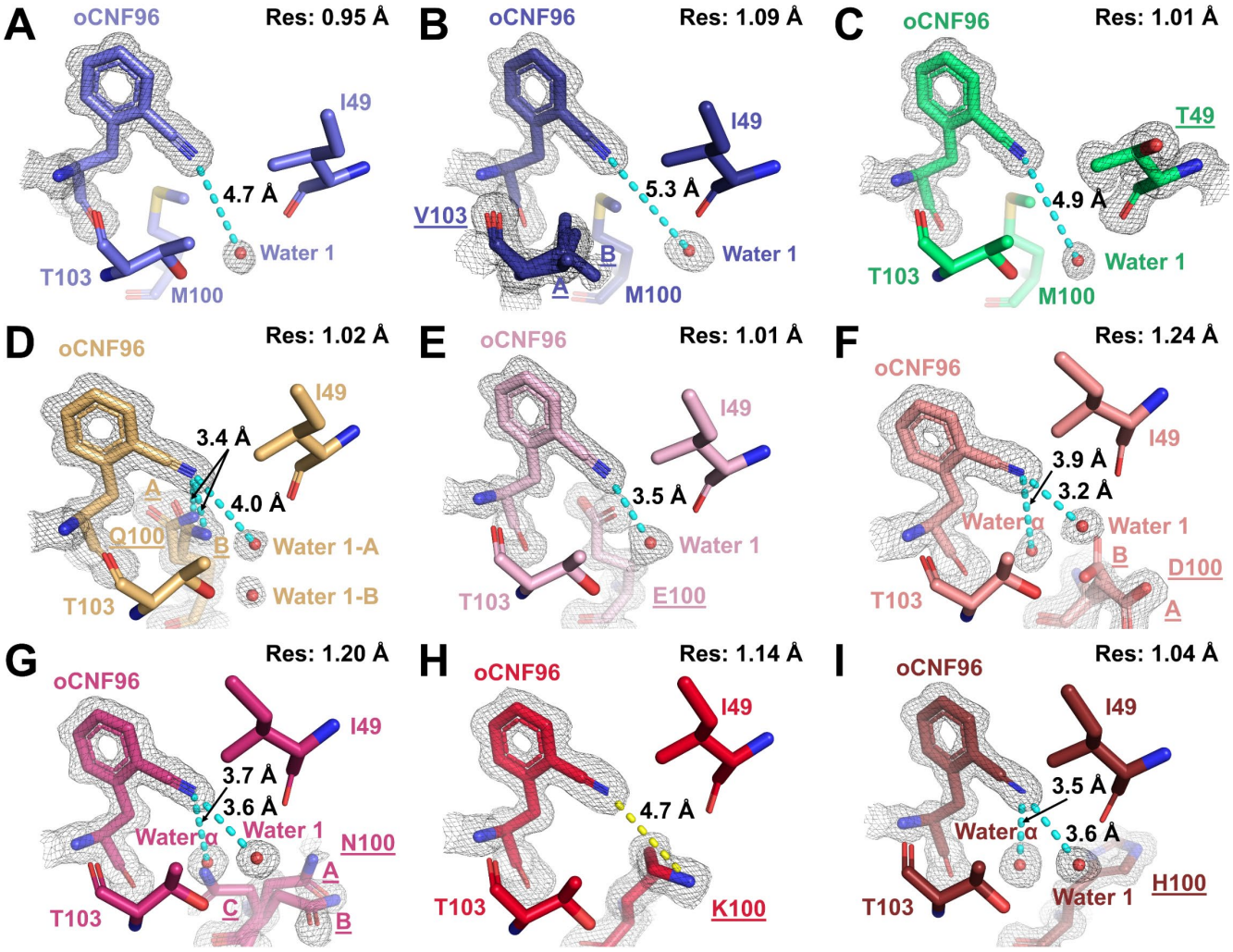

Figure S13. Low temperature (100K) crystal structures for F96oCNF and mutants. Panels display closeups of oCNF96 and local environments highlighting the potential H-bond donor(s) in (A) F96oCNF,[1] (B) T103V, (C) I49T, (D) M100Q, (E) M100E, (F) M100D, (G) M100N, (H) M100K, and (I) M100H. $2mF_o$-DF_c electron density maps contoured at 1σ are displayed for the nitriles, mutated residues, and relevant water(s). M100Q, M100D, and M100N show multiple conformers for the mutated residue; occupancies can be found in **Table S12**. Labels for mutated amino acids are shown underlined, and structure resolutions are shown in the upper right corner. Contacts for closest potential nitrile–donor H-bond(s) are shown with cyan dashed lines; explicit N–donor distances and angles are provided in **Table S13**. For all nitriles except M100K, a water (Water 1) is found within approximately 5 Å of the –C≡N. In M100K, no water is observed near the nitrile as K100 occupies Water 1's position, and the K100–nitrile interaction is shown in yellow to demarcate this difference. Structures are colored according to the similarity of their IR spectra.

In the crystal structures for the P4-type nitriles, Water 1 is observed within 4 Å of each nitrile except M100K's, where it is not observed at all (**Figure S13E-I**). Instead, N_ζ of K100 occupies the space where Water 1 resides in the other M100X mutants (**Figure S15**), where it is pointed away from the nitrile and out towards the solvent (**Figure S16**). To demarcate this contrast from the other P4 nitriles – as well as the other nitrile

environments considered here – the nitrile–K100 interaction is shown in yellow in **Figure S13I**. In M100E, E100 adopts a single conformation that is highly similar to M100Q's Q100-A conformation (**Figure S18A**). In M100H, M100D, and M100N, the mutated residue's only/major conformer (conformer A) faces away from the nitrile and points towards the surrounding water molecules. An additional water, Water α, is observed within 4 Å of these nitriles, implying the mutations in P4 create a more solvent-accessible pocket in solution. In solution at pH 6.0, E100 and D100 are deprotonated, and measurements of the difference in sidechain carbon-oxygen bond lengths for those residues indicate they are carboxylates in the crystal structures as well (see **Figure S21** and discussion).[48] In M100H, H100 appears to engage in two H-bonds, suggesting H100 is doubly protonated in the crystal (**Figure S19**). This would seem to contradict the assignment that H100 is neutral from IR spectroscopy (**Figure S9**), but one of the H-bonds in **Figure S19** is with a symmetry mate that would not be present in solution.

Crystallographic Occupancies and Nearest Potential H-Bonds

Table S12. Crystallographic occupancies for conformers at mutated residue in F96oCNF mutants.

F96oCNF Mutant	A (%)	B (%)	C (%)
T103V	73	27	–
M100Q	54	46	–
M100D	84	16	–
M100N	38	32	30

Table S13. Observed distance and heavy-atom angle between nitrile and potential H-bond donors.

PYP Construct	Distance (Å)	Angle (°)	Identity
F96oCNF	4.70	152	Water 1
T103V	5.29	151	Water 1
I49T	4.92	147	Water 1
M100Q	3.36	124	Q100-B
	3.39	110	Q100-A
	3.99	156	Water 1-A
	5.69	143	Water 1-B
M100E	3.51	145	Water 1
M100D	3.18	151	Water 1
	3.89	106	Water α
M100N	3.57	155	Water 1
	3.73	109	Water α
	4.55	105	N100-C
M100K	4.71	157	K100
M100H	3.59	154	Water 1
	3.51	106	Water α

Structures of oCNF96 Local Environments

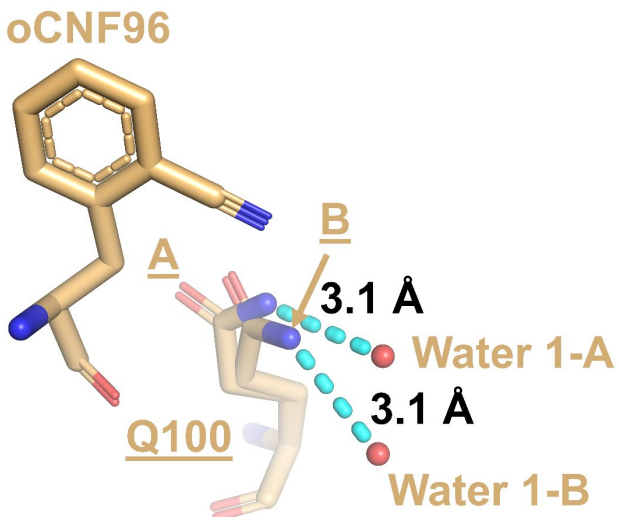

Figure S14. Repeat of main text **Figure 3C** with Water 1-A and Water 1-B shown as H-bond partners for Q100-A and Q100-B, respectively.

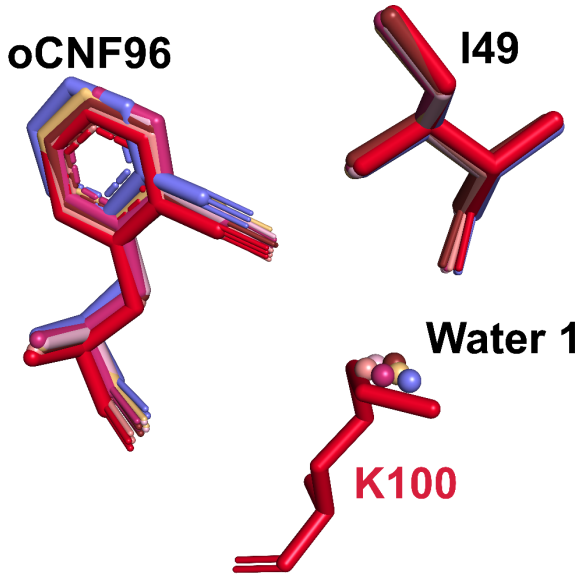

Figure S15. Overlay of oCNF96, I49, and Water 1 for F96oCNF and P3 and P4 PYPs. In F96oCNF, the Water 1–I49 backbone carbonyl O distance is 2.6 Å, indicating an H-bonding interaction; the similarity in position of Water 1 and I49 in the mutants suggests this interaction is maintained. In M100K, K100's N_ζ approximately replaces Water 1 and remains within H-bonding distance of I49 (**Figure S16**). Compared to the nitrile's position in F96oCNF, the P3-/P4-type nitriles move towards Water 1/K100. Atoms of a given structure are uniformly colored for clarity.

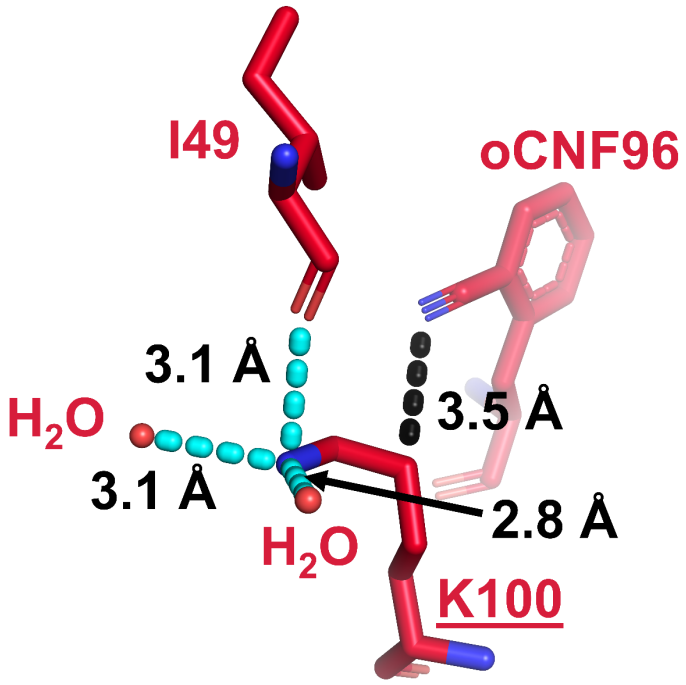

Figure S16. M100K K100 local environment. Potential K100 H-bonds are arranged in an approximately tetrahedral pattern and are indicated in cyan. Closest oCNF96– K100 contact shown in black.

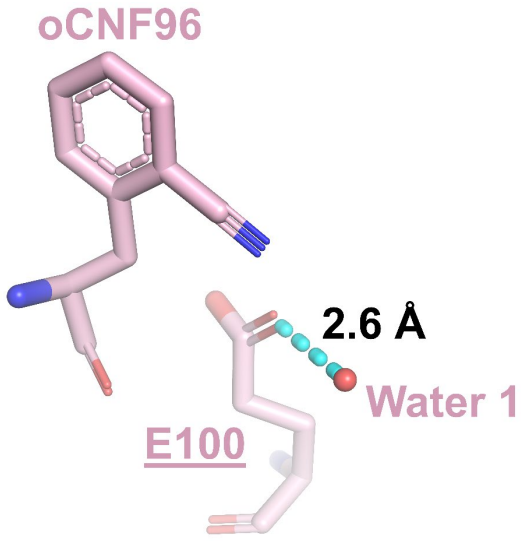

Figure S17. (A) Repeat of **Figure S13E** with E100–Water 1 contact indicated.

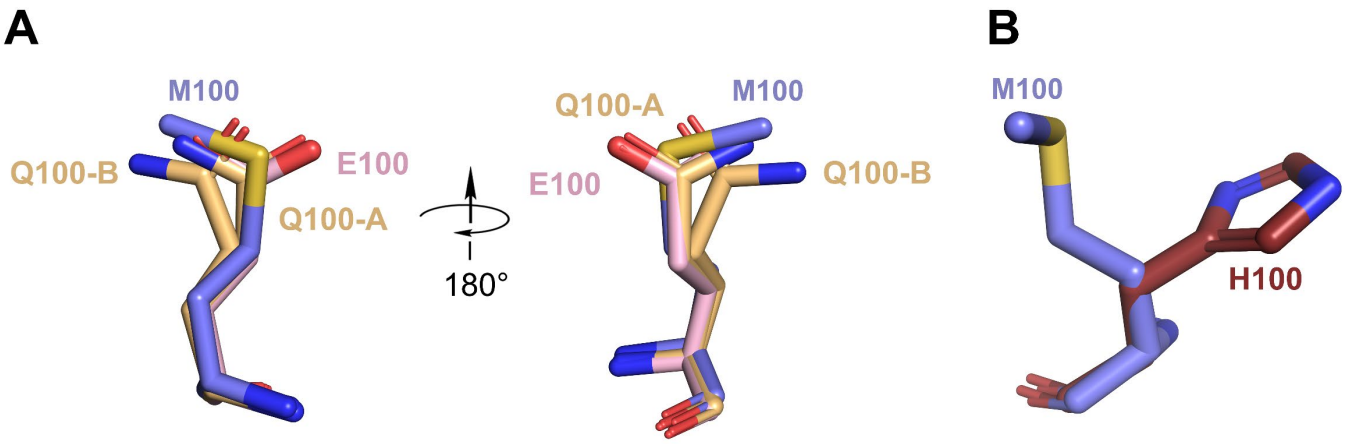

Figure S18. (A) Overlay of M100, Q100, and E100 residues from F96oCNF, M100Q, and M100E, respectively. The two views are related by a 180° rotation. M100, Q100-A, and E100 overlap very closely, particularly Q100-A and E100 due to M100's elongated C – S bonds. (B) Overlay of M100 and H100 from M100H. The C_α and C_β carbons overlap but, unlike Q100 or E100, H100 shows a marked deviation from M100 beginning at C_γ.

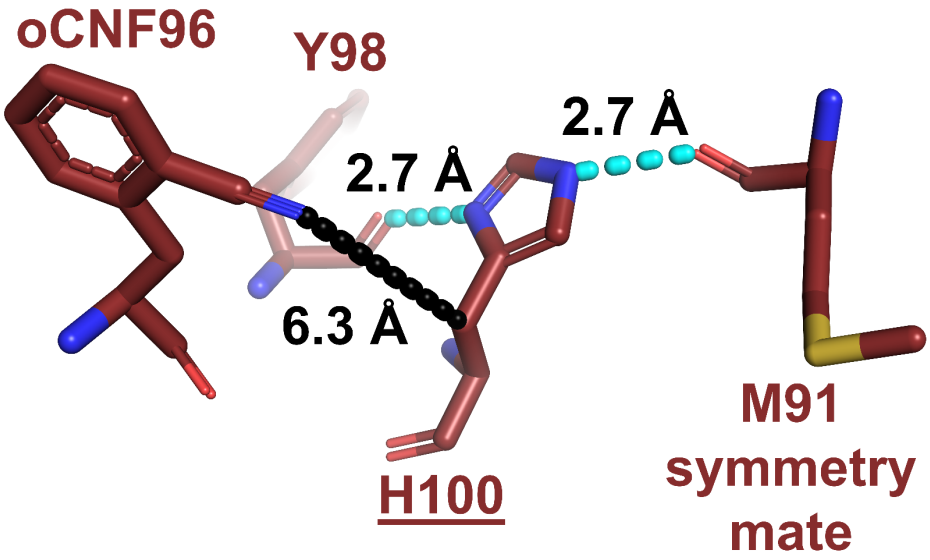

Figure S19. M100H H100 local environment. Potential H-bonds are indicated in cyan, and the presence of two carbonyls within 3 Å of H100's sidechain nitrogens suggests H100 is doubly protonated in the crystal. Closest oCNF96–H100 contact shown in black.

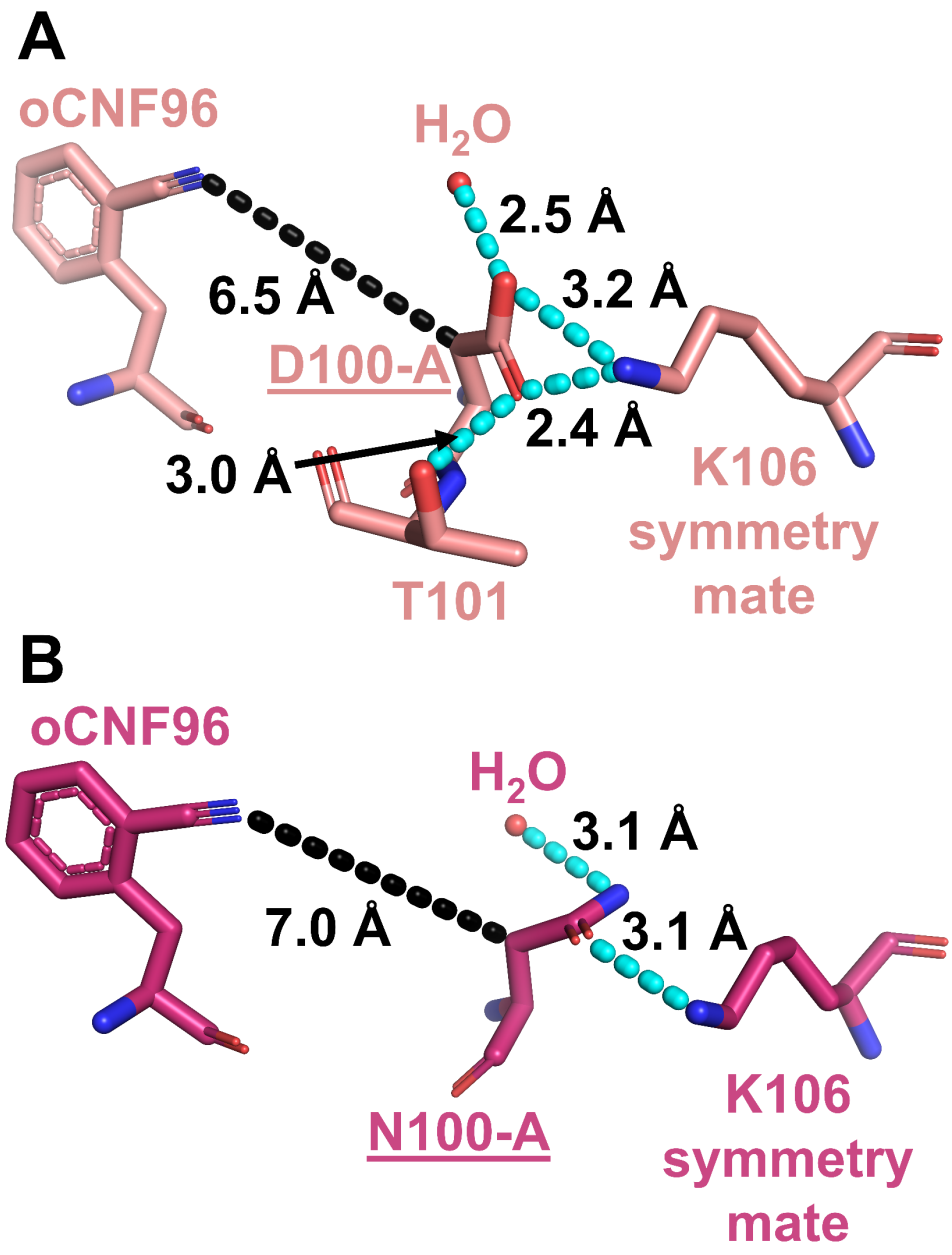

Figure S20. Local environments for (A) D100 in M100D and (B) N100 for M100N. Environments are shown for major conformers. Potential H-bonds are indicated in cyan. Both D100-A and N100-A have a potential H-bond with a K106 symmetry mate; this interaction is also a salt bridge for D100 since it is anionic (see **Figure S21**). The closest oCNF96–D100/N100 contact is shown in black.

Protonation States of E100/D100/H100 in Crystallo

Prior characterization of X-ray structures of small molecules indicates carboxylic acids have a bond length of 1.21 Å for C=O and 1.30 Å for C–OH while carboxylates have equivalent bond lengths of 1.26 Å.[48,49] The estimated standard error on bond lengths for X-ray structures at atomic resolution (≤ 1.2 Å) is 5-10 times smaller than the ~0.1 Å difference between carboxylic acid bond lengths,[48] so direct measurement of glutamic or aspartic acid sidechain bond lengths can elucidate the protonation state.[48,50–52] As a calibration, we measured the sidechain bond lengths for the 19 endogenous Glu or Asp residues in M100E PYP (1.01 Å resolution) and display the differences between the $C_{\varepsilon/\delta}$–$O_{\varepsilon1/\delta1}$ and $C_{\varepsilon/\delta}$–$O_{\varepsilon2/\delta2}$ bond lengths as a histogram (**Figure S21**). We then measured the sidechain bond length difference for E100 and for D100 in M100D PYP. Based on our calibration and others',[48,50] we identified residues with bond length differences of 0.00 – 0.02 Å as carboxylates, 0.03 – 0.04 Å as indeterminate, and 0.06 – 0.08 Å as carboxylic acids. E100 was found to have a sidechain bond length difference of 0.02 Å in M100E, and in M100D, D100-A had a difference of 0.02 Å and D100-B had a difference of 0.00 Å; as such, all E100 or D100 residues are assigned as carboxylates in the crystal structures. These assignments are consistent with PROPKA[53–55] (PDB2PQR v3.6.1), which predicts pK_a values (E100: 4.40; D100-A: 2.81; D100-B: 3.84) significantly below the pH in the crystals (6.0).

In M100H PYP, the crystal structure suggests H100 has H-bonds at $N_{\delta1}$ with Y98's backbone carbonyl and at $N_{\varepsilon2}$ with a symmetry partner (**Figure S19**). The protonation state is difficult to corroborate by direct measurements of H100's angles/bond lengths: the $N_{\delta1}$–$C_{\varepsilon1}$–$N_{\varepsilon2}$ and $N_{\varepsilon2}$–$C_{\delta2}$–C_{γ} angles are nearly equivalent (difference of -0.1°),[56] supporting the assignment of a doubly protonated histidine, while the difference in bond lengths between the $N_{\delta1}$–$C_{\varepsilon1}$ and $C_{\varepsilon1}$–$N_{\varepsilon2}$ bonds is 0.01 Å, supporting the assignment of a singly protonated histidine at $N_{\delta1}$.[57]

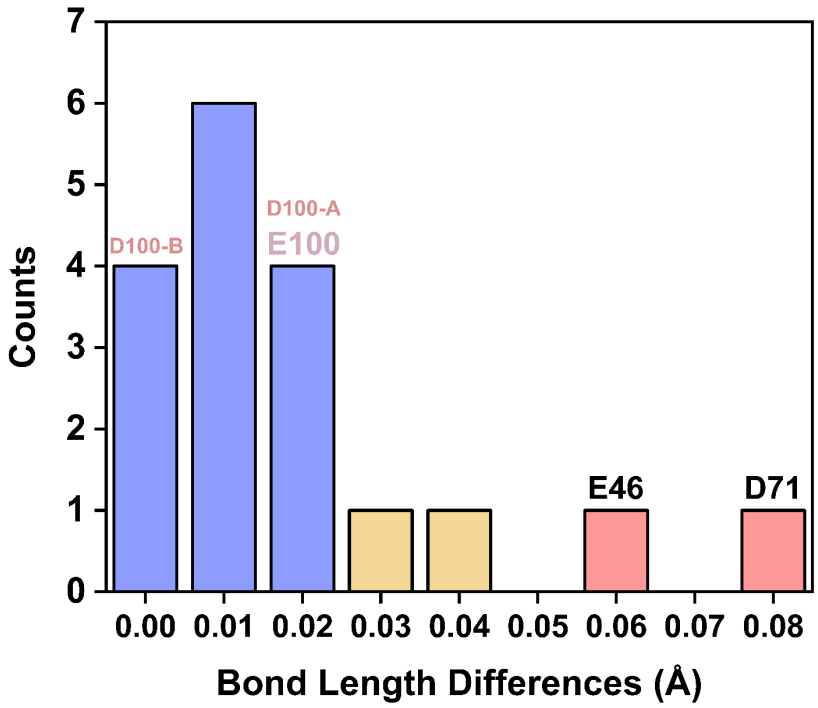

Figure S21. Histogram of sidechain bond length differences between $C_{\varepsilon/\delta}$–$O_{\varepsilon1/\delta1}$ and $C_{\varepsilon/\delta}$–$O_{\varepsilon2/\delta2}$ for 19 endogenous Glu/Asp residues in M100E PYP. In cases where two conformers were modeled, the larger of the two differences was taken. Blue indicates sidechains we identified as carboxylates, yellow indicates that the protonation could not be determined, and red indicates carboxylic acids. E46, a known carboxylic acid which H-bonds with the pCA chromophore,[18,58,59] was found to have a bond length difference of 0.06 Å. D71 was the only other GLU/ASP residue assigned as a carboxylic acid. Since its pK_a in solution has previously been measured as 4.6[60] (PROPKA predicts a similar value of 4.14), the pK_a shift is likely a consequence of crystal packing effects.

B-Factor Analysis

The anomalous local environment of K100 in M100K compared to the other P4 proteins (**Figure S13**) and the alterations to the β4-β5 loop conformations in M100D and M100K (**Figure S12F,H**) – on which residue 100 resides – made us curious to consider the crystallographic B-factors, a measure of the mean displacement of a scattering center in a crystal, to assess the flexibility/rigidity of the loop (and the proteins overall; **Table S14**). Atomic B-factors are affected both by dynamic disorder of the protein, resulting from temperature-dependent vibrations, and by static crystalline disorder,[61] so comparing B-factors between crystals requires normalization with respect to the average B-factor for a given sample (referred to as normalized B-factors, or B_{norm}-values; **Table S15**).[62,63] F96oCNF and mutants are colored by normalized B-factors where chain thicknesses also indicate B_{norm}-values in **Figure S22**. In **Figure S23**, we considered the B_{norm}-values for residues around the nitriles in F96oCNF, M100K, and M100D. We observed significant increases in the β4-β5 loop flexibility for M100K and M100D compared with F96oCNF, indicated by the loop's change from blue in F96oCNF to white in M100K and red in M100D, as well as the concomitant increases in thickness. Correspondingly, residue K100 has a normalized B-factor approximately 1σ larger than M100, while D100 is 2.5σ larger than M100 (**Table S15**). The other M100 mutations also increased the B_{norm}-values at the mutation site, but not to the extent of D100 and K100 since no other mutation increased the value by more than 0.5σ (**Table S15**). Additionally, a comparison of residue 100's B_{norm}-values for non-nitrile-containing WT PYP and Y98Q, another mutant with an altered β4-β5 loop,[46] indicated an increased value of roughly 1σ for Q98 vs Y98. This finding suggests a correlation between alternative β4-β5 loop conformations and elevated B_{norm}-values that extends beyond our PYPs (**Table S15**).

Table S14. Average B-factors for F96oCNF and mutants of the entire structure, protein alone, and waters alone. Note that absolute B-factors include contributions from protein heterogeneity and crystalline disorder. Intervals are for 1σ s.d. All values in Å^2.

PYP Construct	Average overall B-factor (Å^2)	Average protein B-factor (Å^2)	Average solvent B-factor (Å^2)
F96oCNF	12.9 ± 6.0	12.2 ± 5.3	19.3 ± 7.7
T103V	13.9 ± 6.9	13.1 ± 6.4	20.7 ± 7.2
I49T	13.9 ± 5.9	13.2 ± 5.4	20.1 ± 7.2
M100Q	15.5 ± 6.8	15.0 ± 6.4	21.8 ± 7.4
M100E	14.0 ± 6.1	13.5 ± 5.8	20.1 ± 6.5
M100D	18.3 ± 7.7	17.6 ± 7.3	27.6 ± 7.6
M100N	16.0 ± 6.4	15.3 ± 5.9	24.0 ± 6.8
M100K	19.0 ± 13.4	18.6 ± 13.6	24.6 ± 7.3
M100H	13.5 ± 6.0	12.8 ± 5.5	20.6 ± 6.9

Table S15. Normalized B-factors at selected protein and water sites. Mean and standard deviation used for normalization are from the "Average overall B-factor" column in **Table S14**. B-factors for protein residues are averaged over all atoms of the residue, potentially including different conformers. All values in Å^2.

PYP Construct	Residue 49	oCNF96	Residue 100	Residue 103	Water 1	Water A
F96oCNF	-0.33	-1.07	-0.83	-0.90	0.55	–
T103V	–	-0.98	–	-0.57	-0.12	–
I49T	-0.36	-1.07	–	–	1.48	–
M100Q	–	-0.95	-0.34	–	-0.12	–
M100E	–	-0.81	-0.57	–	1.23	–
M100D	–	-0.66	1.36	–	1.43	1.67
M100N	–	-0.69	-0.70	–	0.57	0.73
M100K	–	-0.53	0.09	–	–	–
M100H	–	-0.83	-0.77	–	0.80	-0.29
WT (1NWZ)[64,a]	–	–	-0.62	–	–	–
Y98Q (2I9V)[46,a]	–	–	0.50	–	–	–

[a] Does not contain incorporated nitrile.

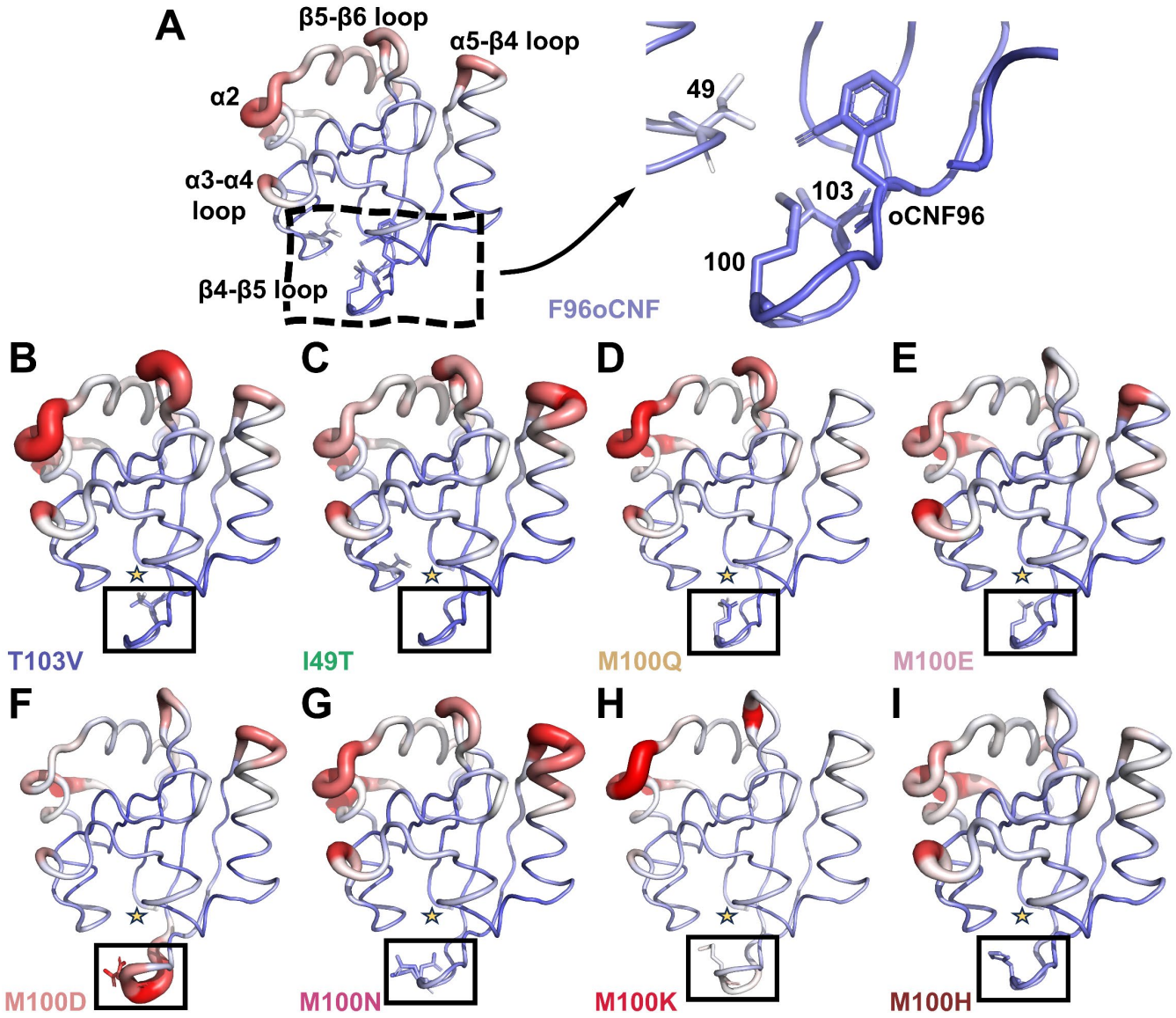

Figure S22. Sausage plots of normalized B-factors (B_{norm}-values) for F96oCNF and mutants. Color and thickness represent values for residue alpha carbons. Values range from -2σ (blue; thinnest) to 0σ (white) to +2σ (red; thickest). For each of the (potentially) mutated residues (shown in stick representation), heavy atoms are colored by B_{norm}-value. Panels display B_{norm}-values for (A) F96oCNF, (B) T103V, (C) I49T, (D) M100Q, (E) M100E, (F) M100D, (G) M100N, (H) M100K, and (I) M100H. In (A), secondary structural features α2, the α3-α4 loop, the α5-β4 loop, the β4-β5 loop, and the β5-β6 loop are specified along with the locations of residues 49, 100, 103, and the nitrile at position 96. Panels (B) through (I) display the mutated residue and a gold star at the approximate nitrile location. Mutations D100 (F) and K100 (H) lead to significantly increased B_{norm}-values, and therefore increased flexibility, at the β4-β5 loop proximal to the nitrile (see **Figure S23** for closeup). Mutations can also lead to increased (or decreased) flexibility compared to F96oCNF throughout the protein: significantly increased normalized B-factors are observed at α2 in (B), (D), and (H), at the α3-α4 loop in (E), at the α5-β4 loop in (C) and (G), and at the β5-β6 loop in (B), while significantly decreased flexibility (i.e., increased rigidity) is observed at the α5-β4 loop in (D) and (H). Normalized B-factors for individual residues of interest are provided in **Table S15**. In the closeup in (A), a part of one loop was removed for clarity.

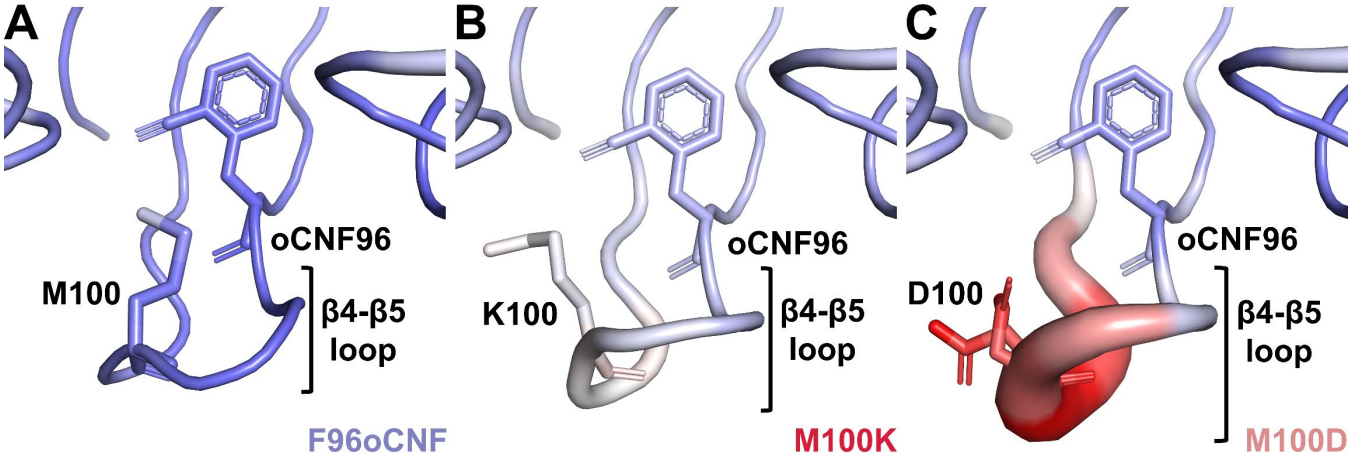

Figure S23. Normalized B-factors (B_{norm}-values) for residues nearby the nitrile in (A) F96oCNF, (B) M100K, and (C) M100D. The protein backbone is shown in putty form where each residue's color and thickness are proportional to B_{norm}-values, ranging from -2σ (blue; thinnest) to 0σ (white) to +2σ (red; thickest). Negative and positive B_{norm}-values indicate decreased and increased flexibility compared to the average heavy atom in the crystal structure. Backbone coloring and thickness are from C_α normalized B-factors, while each heavy atom's B_{norm}-value is indicated by color for oCNF96 and residue 100. Changing residue 100's identity from M100 to K100 to D100 leads to progressively increased flexibility at the β4-β5 loop. Mutation from M100 to D100 or K100 also caused an alternative β4-β5 loop conformation compared to that observed in F96oCNF (see **Figure S12** for overlays). Normalized B-factors for individual residues of interest are provided in **Table S15**.

Mutant Nitriles Predominately H-Bond with Water: Implications for the Design of Nitrile Noncovalent Interactions

In this study, we were able to alter the PYP nitrile electric fields and H-bonding, but this predominately occurred for reasons we did not intend, i.e., many of the nitriles H-bond with water rather than the mutated residue. The logic and systematic design of various nitrile noncovalent interactions here is highly analogous to our recent paper where chemical modifications were made around a –C≡N in metal organic frameworks (MOFs) to electrostatically characterize a range of noncovalent interactions.[65] In that study, nitrile frequencies/inferred electric fields strongly suggested the design goals were achieved. Comparing the design efforts here with those in ref. [65] highlights the difference of exploring nitrile noncovalent interactions in proteins vs MOFs. MOFs are rigid and can be evacuated, which helps ensure chemical modifications conform to the design goal when H-bonding interactions are introduced. In contrast, soluble proteins are soft matter which are always surrounded by water, and the PYP mutants demonstrate the challenges of designing noncovalent interactions in these settings. Our inability to design nitrile environments suggests that nitriles are reasonably innocent spectators, at least when they are placed close to the protein surface. The success of the MOF designs suggest nitriles may be successfully used to benchmark force fields in that setting.[66]

S8 Molecular Dynamics Figures & Tables

MD Electric Field and H-Bonding Characterization

Table S16. Standard deviations (s.d.'s) of the electric field distributions of MD nitrile H-bonding and non-H-bonding populations. Field distributions for nitrile-containing PYPs are shown in main text **Figure 4**.

PYP Construct	s.d. H-bonding $\lvert\vec{F}\rvert_{C\equiv N,MD}$ (MV/cm)	s.d. non-H-bonding $\lvert\vec{F}\rvert_{C\equiv N,MD}$ (MV/cm)
F28oCNF	19.1	15.0
F62oCNF	N/A	5.6
F92oCNF	17.0	10.7
F96oCNF	17.9	9.9
T103V	18.9	9.6
I49T	18.7	9.6
M100Q	18.0	12.4
M100E	19.0	14.5
M100D	17.8	12.7
M100N	18.4	13.3
M100K	20.5	10.7
M100H	19.6	13.4

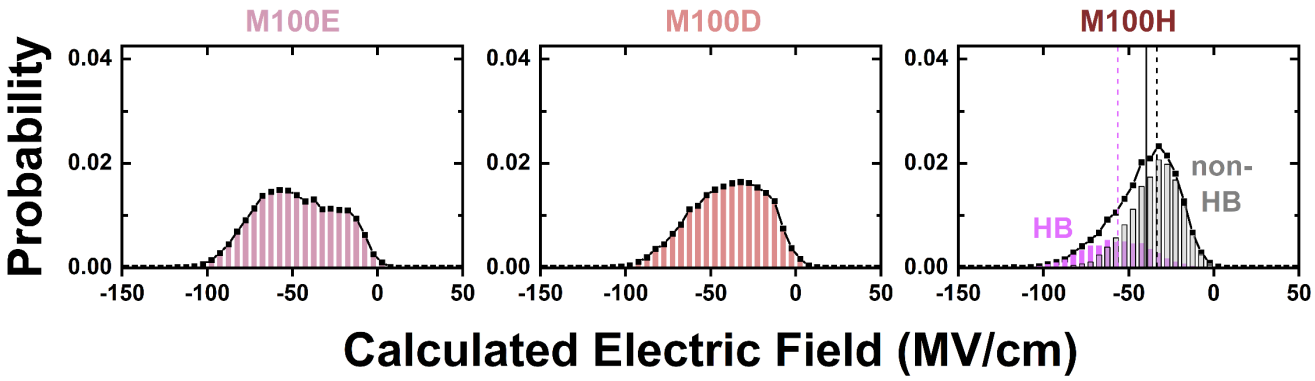

Figure S24. Electric field distributions (black traces) for M100E, M100D, and M100H when titratable residues E100 and D100 are protonated and H100 is doubly protonated with net positive charge. For M100H, distributions are decomposed into their H-bonding (magenta) and non-H-bonding populations (gray). Magenta and black dashed lines indicate the median H-bonding and non-H-bonding electric fields, respectively, and the solid lines indicate the fraction-weighted electric fields. The shapes of M100E and M100D's MD distributions look appreciably different from M100N's (another P4 nitrile which does not have a titrate mutation; main text **Figure 4**), implying E100 and D100 are deprotonated as suggested by IR spectroscopy and X-ray crystallography. M100H looks similar to M100N in terms of its field distribution, but its H-bonding and non-H-bonding fractions (**Table S17**) deviate significantly from those of M100N (main text **Table 2**); in contrast, neutral M100H has more similar H-bonding fractions to M100N and a more similar electric field distribution, suggesting H100 is neutral in solution (as suggested by IR spectroscopy at pH 6.0 and 8.5; **Figure S9**).

Table S17. Calculated nitrile H-bonding/non-H-bonding fractions, median electric fields, fraction-weighted electric field, and associated errors for M100H with positively charged H100. Errors come from averaging four independent 25 ns MD trajectories. The distribution is shown in **Figure S24**.

Environment	H-bonding fraction (%)	non-H-bonding Fraction (%)	H-bonding median $\left\|\vec{F}\right\|_{C\equiv N,MD}$ (MV/cm)	non-H-bonding median $\left\|\vec{F}\right\|_{C\equiv N,MD}$ (MV/cm)	Fraction-weighted $\left\|\vec{F}\right\|_{C\equiv N,MD}$ (MV/cm)
M100H	26.6 ± 9.8	73.4 ± 9.8	-56.5 ± 2.7	-33.3 ± 1.7	-39.7 ± 2.3

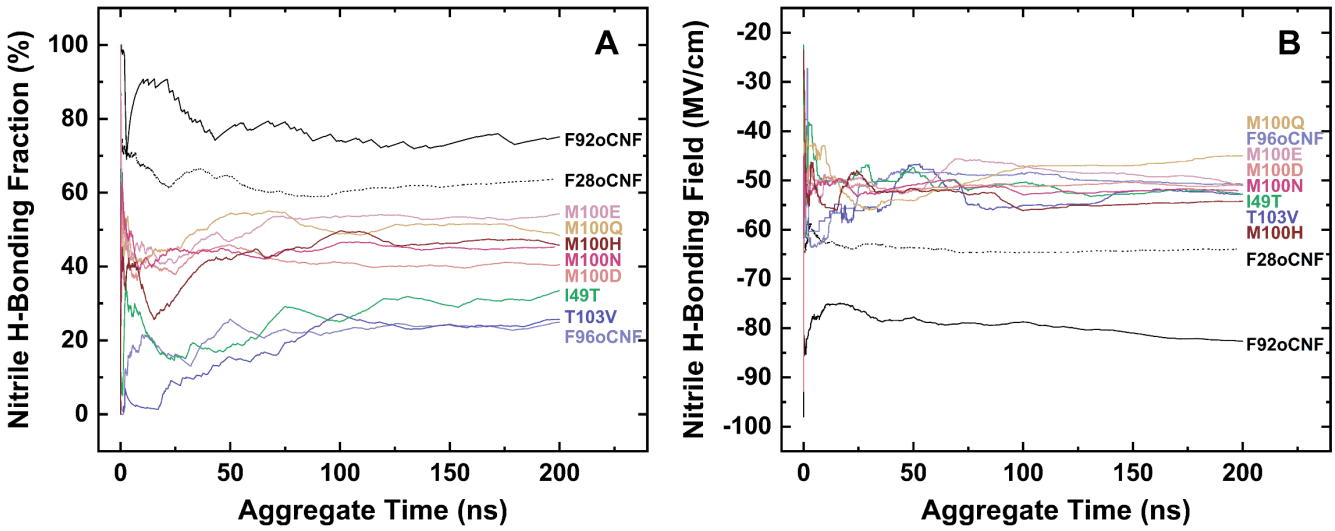

Figure S25. Moving average of the nitrile H-bonding fraction and moving median of the nitrile H-bonding electric field for all nitrile-containing PYPs except F62oCNF, which has no H-bonding population. The plots indicate our aggregate MD simulations are reasonably converged.

On the Errors for the Fraction-Weighted Median MD Electric Fields

While we might have assumed that nitriles in fast chemical exchange would have smaller errors for their fraction-weighted median $|\vec{F}|_{C\equiv N,MD}$ values since their local dynamics are faster than for those nitriles which demonstrate two resolved IR populations, several P4-type nitriles have lower errors than those for T103V or I49T, belying this assumption. Specifically, we found that M100N and M100D, whose crystal structures have multiple conformations observed for the mutated residues (**Figure S13F,G**), have the smallest errors in $|\vec{F}|_{C\equiv N,MD}$ values for any F96oCNF mutant considered for benchmarking. This observation suggests that running simulations from multiple starting conformations aided in exploring the nitrile electric field sampling spaces. M100Q has the largest error for any nitrile-containing PYP (main text **Table 2**), which seems to contradict this hypothesis since two poses were observed in the crystal structure (main text **Figure 3C**). To investigate this, we calculated the H-bonding fractions and fraction weighted-median $|\vec{F}|_{C\equiv N,MD}$ values for runs starting from the "A" conformer (i.e., with a nitrile–water H-bond) and the "B" conformer (i.e., with a nitrile–Q100 H-bond; see **Table S18** for fields of the subpopulations). We found $|\vec{F}|_{C\equiv N,MD,M100Q\ A}$ is -34.2 ± 3.4 and $|\vec{F}|_{C\equiv N,MD,M100Q\ B}$ is -21.4 ± 4.3, and runs starting from the A conformer have 15% less nitrile H-bonding populations than runs starting from the B-conformer (**Table S18**). The large disparity in $|\vec{F}|_{C\equiv N,MD}$ depending on the starting point explains M100Q's larger error, and altogether these results highlight how the crystal structure pose(s) can affect the nitrile environments sampled during the MD trajectories and the resulting MD electric fields.

Table S18. Calculated nitrile H-bonding fractions, median electric fields, fraction-weighted electric fields, and associated errors from averaging four independent 25 ns MD trajectories (**Table S2**).

PYP Construct	Runs starting from "A" conformation	Runs starting from "B" conformation
H-bonding fraction (%)	56 ± 7	41 ± 12
Protein H-bonding fraction (%)	50 ± 25	55 ± 8
non-H-bonding median field (MV/cm)	-22.4 ± 1.3	-14.0 ± 2.5
H-bonding median field (MV/cm)	-44.0 ± 5.3	-32.3 ± 1.4
Fraction-weighted median field (MV/cm)	-34.2 ± 3.4	-21.3 ± 4.3

Frequency of "Long" Nitrile H-Bonding and Non-H-Bonding Residence Times

Table S19. Number of nitrile H-bonding and non-H-bonding residence events of >250 ps and >1000 ps observed for F96oCNF and mutants.

PYP Construct	State	>250 ps	>1000 ps
F96oCNF	Non-H-Bonding	97	14
F96oCNF	H-Bonding	0	0
T103V	Non-H-Bonding	111	22
T103V	H-Bonding	6	0
I49T	Non-H-Bonding	110	11
I49T	H-Bonding	6	0
M100Q	Non-H-Bonding	18	1
M100Q	H-Bonding	1	0
M100E	Non-H-Bonding	2	0
M100E	H-Bonding	14	0
M100D	Non-H-Bonding	7	0
M100D	H-Bonding	0	0
M100N	Non-H-Bonding	1	0
M100N	H-Bonding	0	0
M100K	Non-H-Bonding	117	17
M100K	H-Bonding	0	0
M100H	Non-H-Bonding	40	3
M100H	H-Bonding	5	0

Structural Origins of M100K's and M100E's Altered H-Bonding Fractions Relative to the Other P4-Type Nitriles

M100K has an anomalously low MD H-bonding fraction compared with the other P4-type nitriles (main text **Table 2**). To understand why this is the case, we tallied the number of times each nitrile for F96oCNF and mutants remained in an H-bonding or non-H-bonding state for >1 ns (**Table S19**). We found M100K's nitrile had five times the number of these residences in the non-H-bonding state compared with any other P4-type nitrile. We investigated the protein structures associated with >1 ns residences and found K100–T103 H-bonds in many of these snapshots (see **Figure S26** for a representative), an interaction that is not present in the crystal structure (**Figure S13I**). K100 H-bonds with T103 in 33% of frames – about the same difference in H-bonding fraction seen between M100K and other P4-type nitriles (main text **Table 2**). As such, the K100–T103 H-bonds are the origin for M100K's deviation from the other P4-type nitrile MD field distributions (main text **Figure 4**); because of this artificial dissimilarity, M100K was not used for benchmarking.

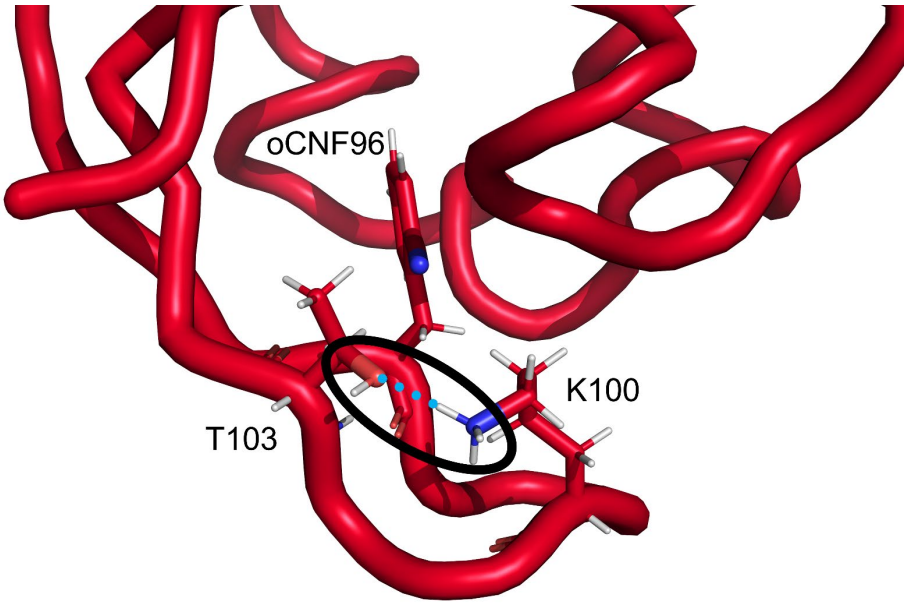

Figure S26. Representative snapshot of a K100 – T103 H-bond (cyan dotted line) in M100K with oCNF96, K100, and T103 shown as sticks. The N_{K100}–O_{T103} H-bond distance is 2.73 Å and the C_{T103}–O_{T103}–N_{T103} H-bond angle is 141°. Water molecules have been removed for clarity.

In addition, M100E has a larger H-bonding fractions than the other P4-type nitriles in the AMOEBA simulations (main text **Table 2**). To understand why, we carried out a similar analysis to what was done for M100K. In considering "long" residence times in the H-bonding state, M100E exhibited twice as many >250 ps residences as any other P4-type nitrile used for benchmarking (**Table S19**). We previously found protein H-bonds tend to exhibit longer residence times than solvent H-bonds,[5,17] so we investigated M100E's nitrile's H-bond partners and found it H-bonds with a partner on the protein 13% of the time, an order of magnitude more than in M100D, M100N, or M100H, and a 4-fold increase from F96oCNF (**Table S20**). Of M100E's nitrile–protein H-bonds, over 95% were with T50 (**Table S20**), a residue in the same alpha helix as I49 (main text **Figure 1A**). These T50 H-bonds can be observed in M100E's H-bond distance vs angle contour plot: M100E could not be well-fit unless two Gaussians were used, with one Gaussian having similar average H-bond distance and angle to the other P4-type

nitriles and the other Gaussian centered at 3.25 Å and 110° (**Table S24**). Fitting contour plots for M100E nitrile–water or nitrile–T50 H-bonds indicates T50 is the origin of the population with longer distances and smaller angles, while the water H-bonds sample the same geometries as in the other P4-type nitriles(**Table S25**). As such, the T50 H-bonds are the origin of M100E's smaller non-H-bonding fraction compared with the other P4-type nitriles (main text **Table 2**), which in turn can explain why M100E's IR band has a smaller extinction coefficient at its absorption maximum – which corresponds with the aprotic peak position – than any other P4-type nitrile (main text **Figure 2A** and **Figure S10**).

MD H-Bond Donor Identities

Table S20. Fraction of nitrile H-bonds to either a protein residue or solvent from MD simulations. Cases with multiple H-bonds (which constitute <2% of all MD frames) were not considered to simplify donor assignment.

Environment/ Force Field	Protein (%)	Solvent (%)
F96oCNF	3.0 (84.0 with T50)	97.0
T103V	6.6	93.4
I49T	8.4	91.6
M100Q	53.0 (97.4 with Q100)	47.0[a]
M100E	13.3 (96.4 with T50)	86.7
M100D	0.5	99.5
M100N	1.7	98.3
M100K	11.7 (39.8 with K100)	88.3
M100H	0.6	99.4

[a] Crystallographic occupancies in **Table S12** are remarkably similar (conformer A corresponds with a solvent H-bonding pose).

Investigating the Origins of M100Q's MD Electric Field Distribution

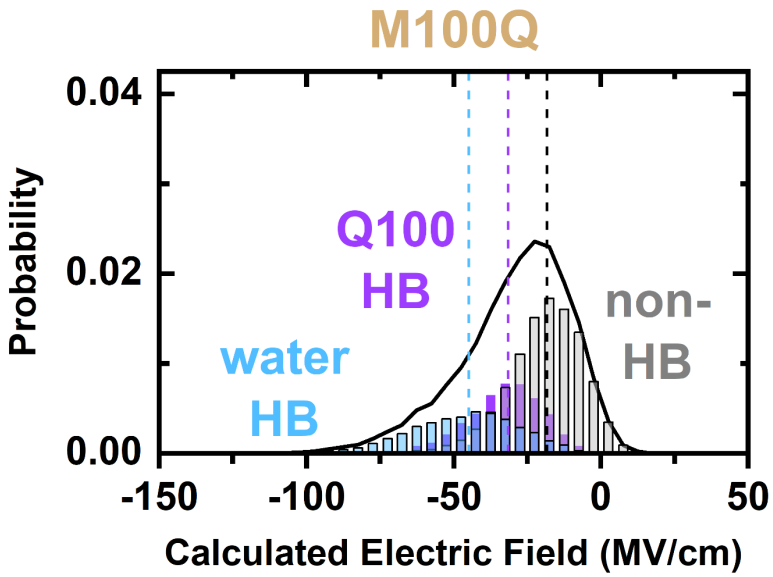

Figure S27. Decomposition of the M100Q nitrile MD electric field distribution from main text **Figure 4** into contributions from the non-H-bonding (grey), Q100 H-bonding (purple), and water H-bonding populations (blue). Fields from frames with both a Q100 and a water H-bond – about 8% of H-bonding frames – were excluded.

Table S21. Median and standard deviation of the electric field distributions for M100Q nitrile–water H-bonds and nitrile–Q100 H-bonds (which make up the vast majority of the protein H-bonding population; see **Table S20**) from aggregate MD simulations. Median electric fields and associated errors from averaging over eight independent 25 ns trajectories are shown in parentheses.

PYP Construct	Median Electric Field (MV/cm)	s.d. Electric Field (MV/cm)
$M100Q_{Water}$	-44.8 (-43.6 ± 7.3)	18.3
$M100Q_{Q100}$	-31.5 (-31.3 ± 3.7)	12.6

Investigating the Seeming Discrepancy Between the Number of IR and MD Populations for M100Q

While the identification of two nitrile H-bonding species in M100Q in **Table S20** and **Figure S27** explains why its median H-bonding field is smaller than those for F96oCNF or the other mutants (main text **Figure 4** and **Table 2**), it also raises a separate question: why are only two nitrile IR bands present (main text **Figure 2**) when there are three nitrile species in the MD simulations (non-H-bonding, H-bonding with water, and H-bonding with Q100)? To probe this question, we generated a predicted M100Q IR spectrum that does not account for TDM tuning by calculating a nitrile frequency for every output MD frame (**Figure S28**). This was done using eq 1, where the gas phase frequency and Stark tuning rate are that for oTN,[1] the fields are from the AMOEBA simulations, and the H-bond blueshifts were derived using eq S1 (see **Section S3**). The frequencies were binned every 0.25 cm^{-1} to create a spectrum which indicates that the MD M100Q H-bonding population is centered at slightly higher energy than the non-H-bonding population, as we observe in the experimental IR spectrum (main text **Figure 2**). When we quantified the median values for both MD populations, we obtained values of 2228.0 cm^{-1} for the non-H-bonding population and 2232.3 cm^{-1} for the H-bonding population (**Table S22**): these values are within 1 cm^{-1} of the experimentally observed frequencies (**Table S8**), in excellent agreement. Further dissecting the H-bonding population, we find the Q100 MD H-bonding population has a median frequency of 2233.2 cm^{-1}, which is within the error of the experimental value (**Table S8**), while the water H-bonding population has a median value of 2231.8 cm^{-1}. Since the MD nitrile–Q100 and nitrile–water H-bonding populations' center frequencies are separated by less than 1.5 cm^{-1}, this exercise implies the two populations overlap in the experimental IR spectrum, as is observed in **Figure S28**. Therefore, we can clarify the apparent population discrepancy between IR and MD using eqs 1 and S1, an illustration of the VSE formalism's ability to successfully model nitrile frequencies with the new H-bond blueshift equation.[5]

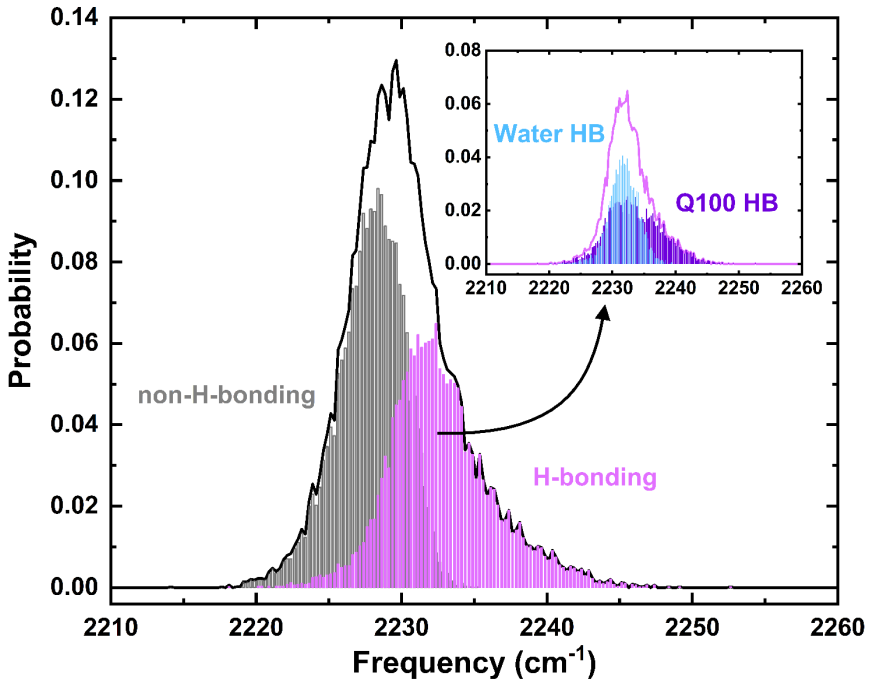

Figure S28. M100Q MD IR spectrum. The spectrum was generated by using the fields and H-bond distance/angle from every output frame and inputting the quantities into main text eq 1, with the H-bond blueshift described by eq S1. Frequencies were binned every 0.25 cm^{-1}. Frames with H-bonds to both water and Q100 were excluded for simplicity. Note that the H-bonding "band" appears smaller than the experimental H-bonding band (main text **Figure 2C**) since the MD generated IR spectrum does not incorporate the TDM tuning effect on the band intensity.

Table S22. M100Q MD nitrile frequencies derived using main text eq 1.

PYP Construct	Median MD Nitrile Frequency (cm^{-1})
$M100Q_{non\text{-}H\text{-}bonding}$	2228.0
$M100Q_{H\text{-}bonding}$	2232.3
$M100Q_{Water}$	2233.2
$M100Q_{Q100}$	2231.8

Nitrile H-Bond Contour Plots

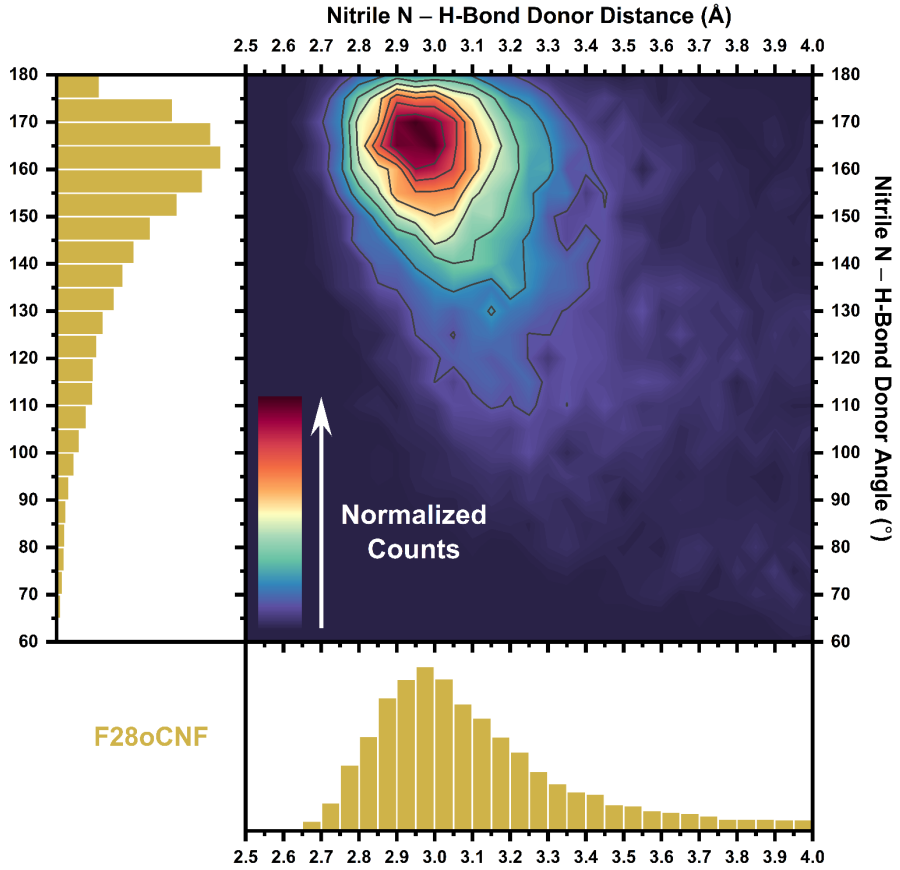

Figure S29. Nitrile H-bond angle vs distance for H-bonds in F28oCNF PYP depicted as a contour plot describing H-bond angle/distance sampling space. 1D histograms show the sampling space projected along the individual variables.

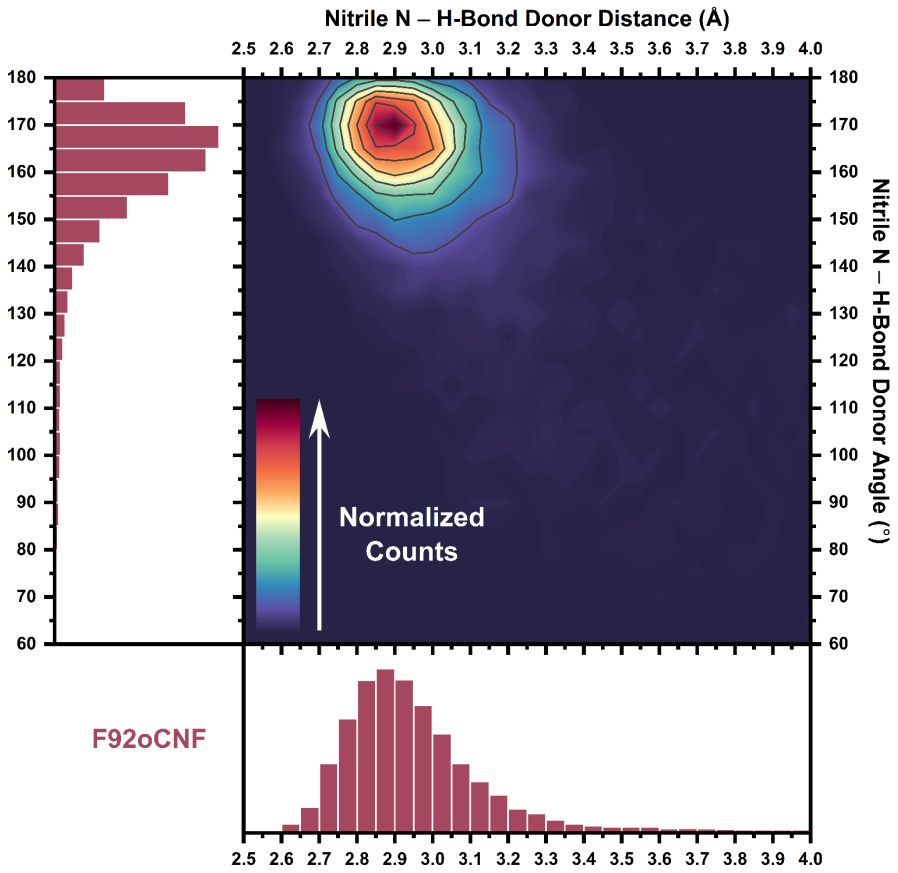

Figure S30. Nitrile H-bond angle vs distance for H-bonds in F92oCNF PYP depicted as a contour plot describing H-bond angle/distance sampling space. 1D histograms show the sampling space projected along the individual variables.

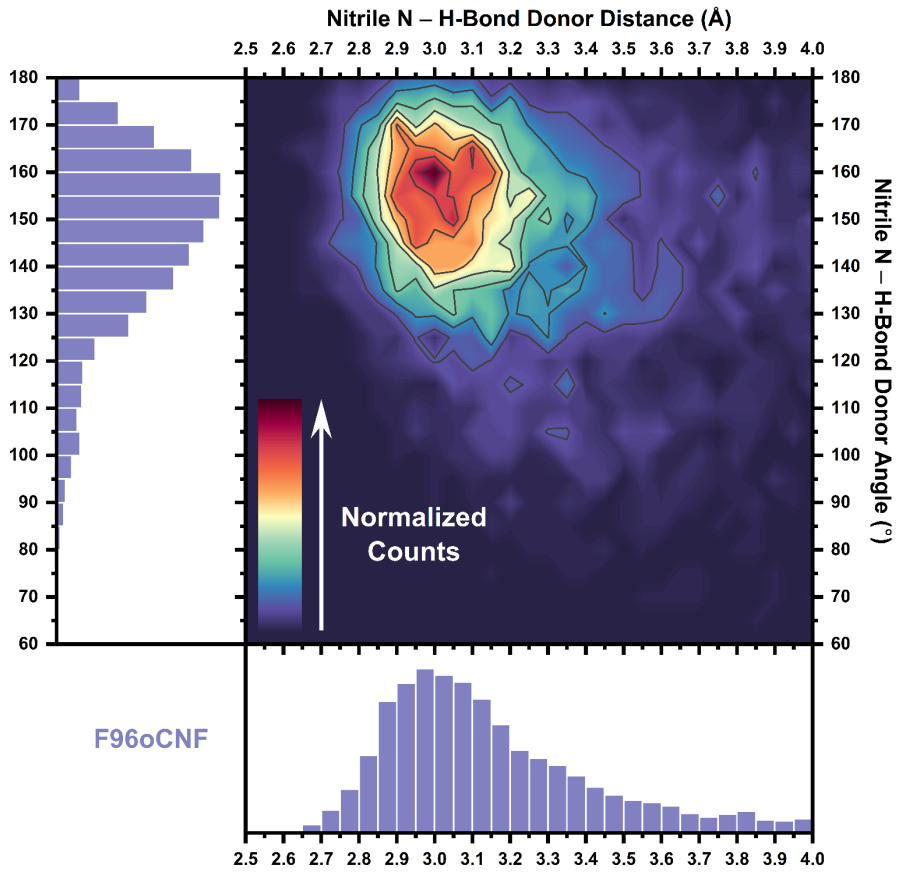

Figure S31. Nitrile H-bond angle vs distance for H-bonds in F96oCNF PYP depicted as a contour plot describing H-bond angle/distance sampling space. 1D histograms show the sampling space projected along the individual variables.

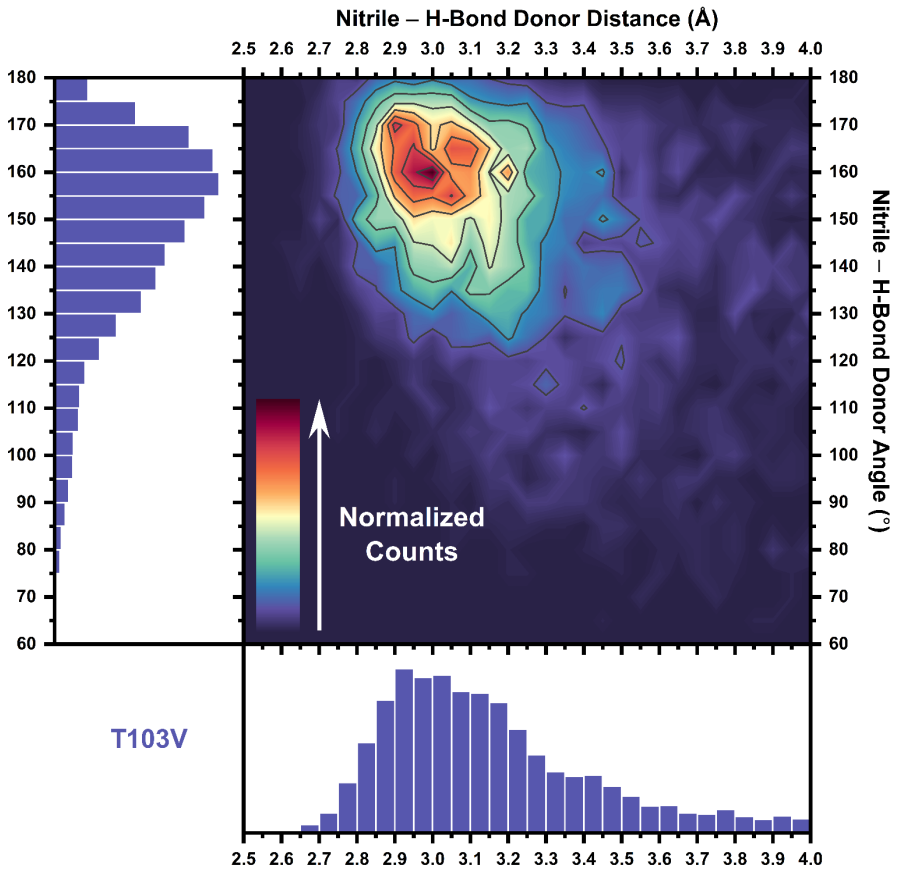

Figure S32. Nitrile H-bond angle vs distance for H-bonds in F96oCNF T103V PYP depicted as a contour plot describing H-bond angle/distance sampling space. 1D histograms show the sampling space projected along the individual variables.

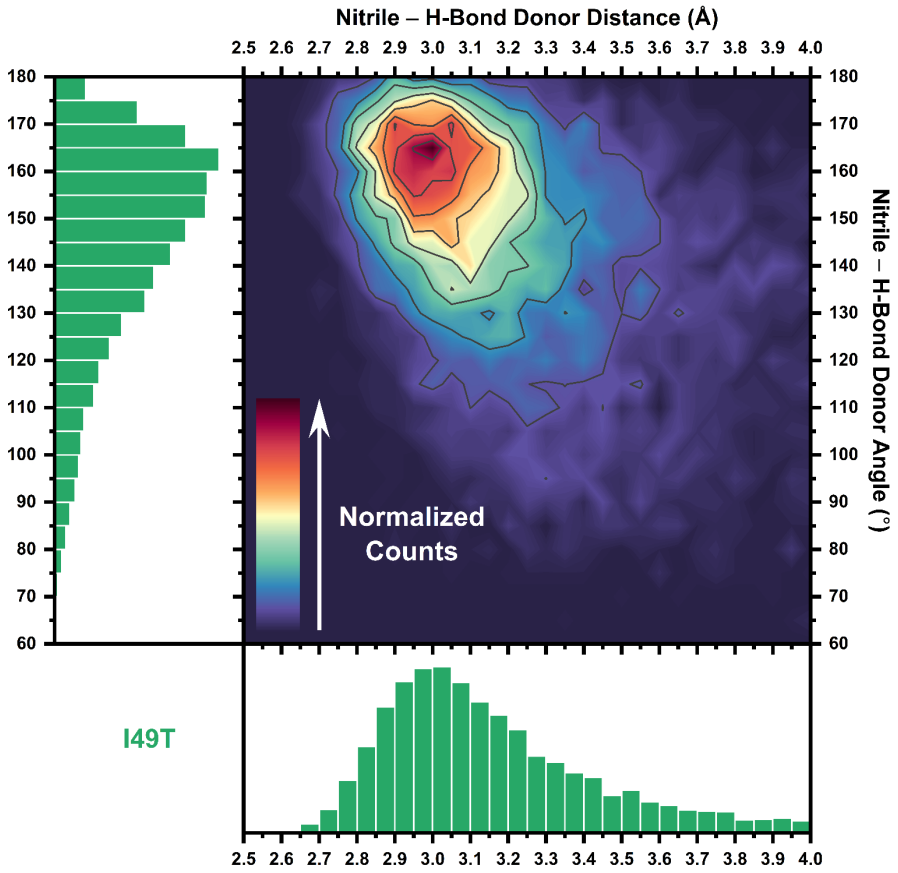

Figure S33. Nitrile H-bond angle vs distance for H-bonds in F96oCNF I49T PYP depicted as a contour plot describing H-bond angle/distance sampling space. 1D histograms show the sampling space projected along the individual variables.

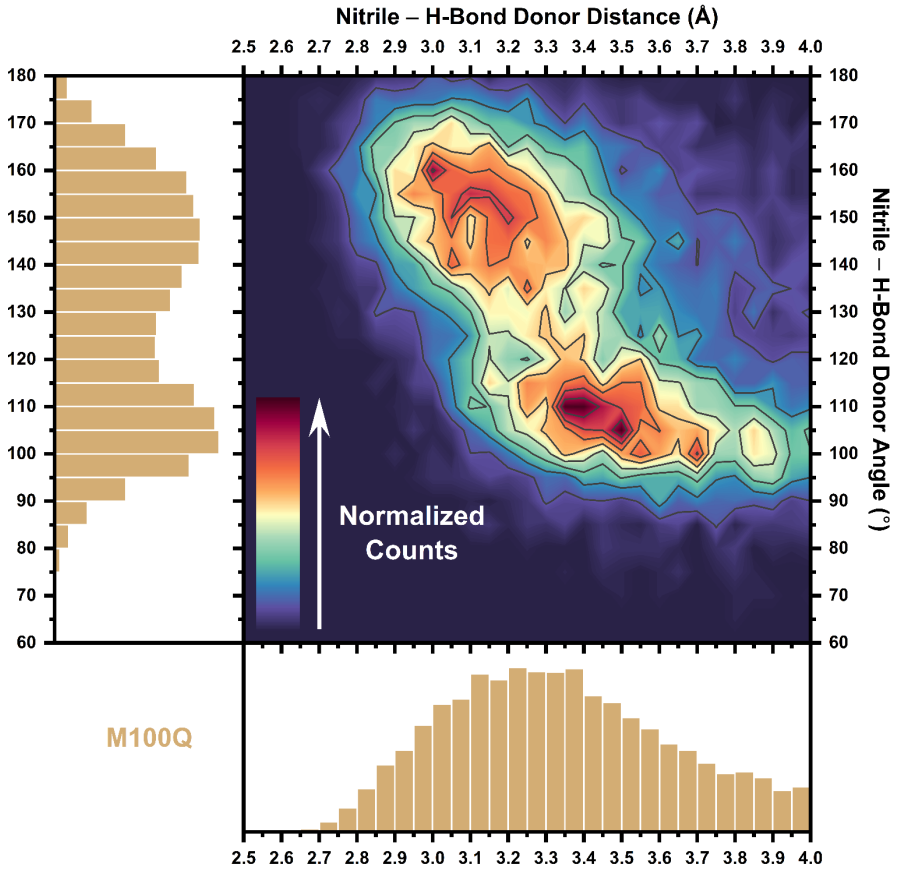

Figure S34. Nitrile H-bond angle vs distance for H-bonds in F96oCNF M100Q PYP depicted as a contour plot describing H-bond angle/distance sampling space. 1D histograms show the sampling space projected along the individual variables. More sampling is observed at smaller angles and longer distances than in the other F96oCNF mutants. Due to this, we explored what the contour plot looks like out to longer distances (**Figure S35**) and further broke the contour plot down into contributions from water and Q100 H-bonds with the nitrile (**Figure S36**; see **Table S20** for relative H-bonding fractions).

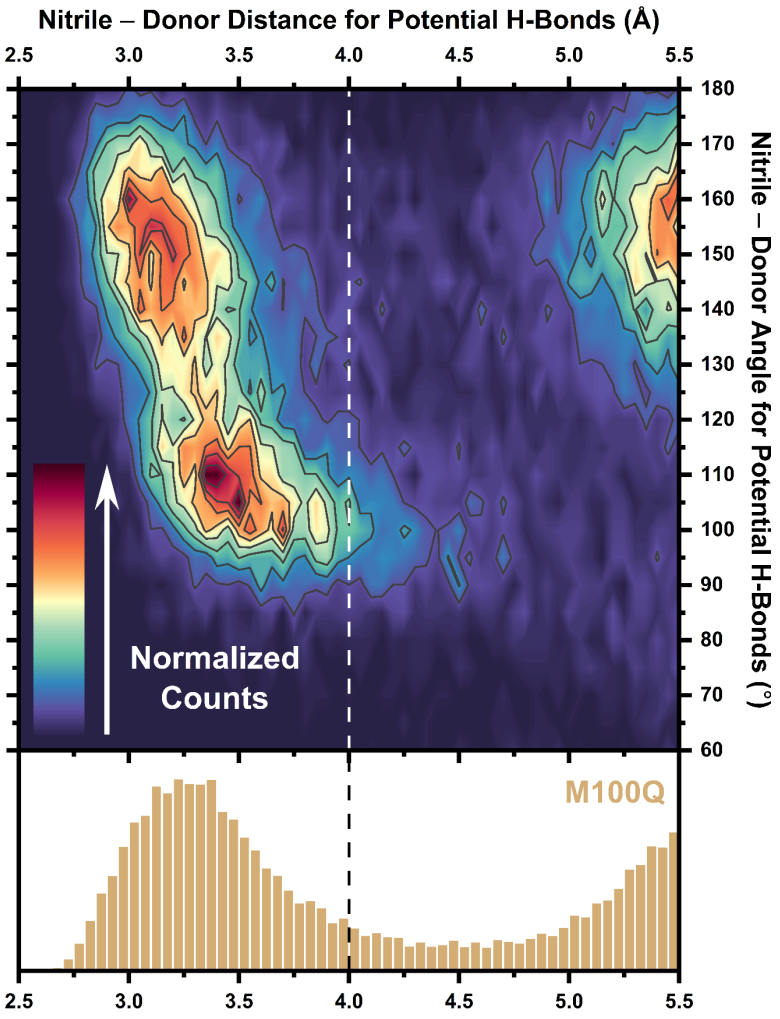

Figure S35. Repeat of the F96oCNF M100Q PYP nitrile H-bond angle vs distance contour plot and H-bond distance histogram in **Figure S34** extending the distance axis out to 5.5 Å. The H-bond cutoff distance of 4.0 Å is shown as a dashed line. If the H-bond cutoff distance were moved out to 4.5 Å (closer to the end of the first solvation shell here), the H-bonding fraction would increase by 1.4 %, comparable to the increase in the fraction for F92oCNF using a distance cutoff of 3.5 Å vs 4.0 Å.[17] This does not represent a large increase and, hence, a cutoff distance of 4.0 Å is still reasonable to adopt here.

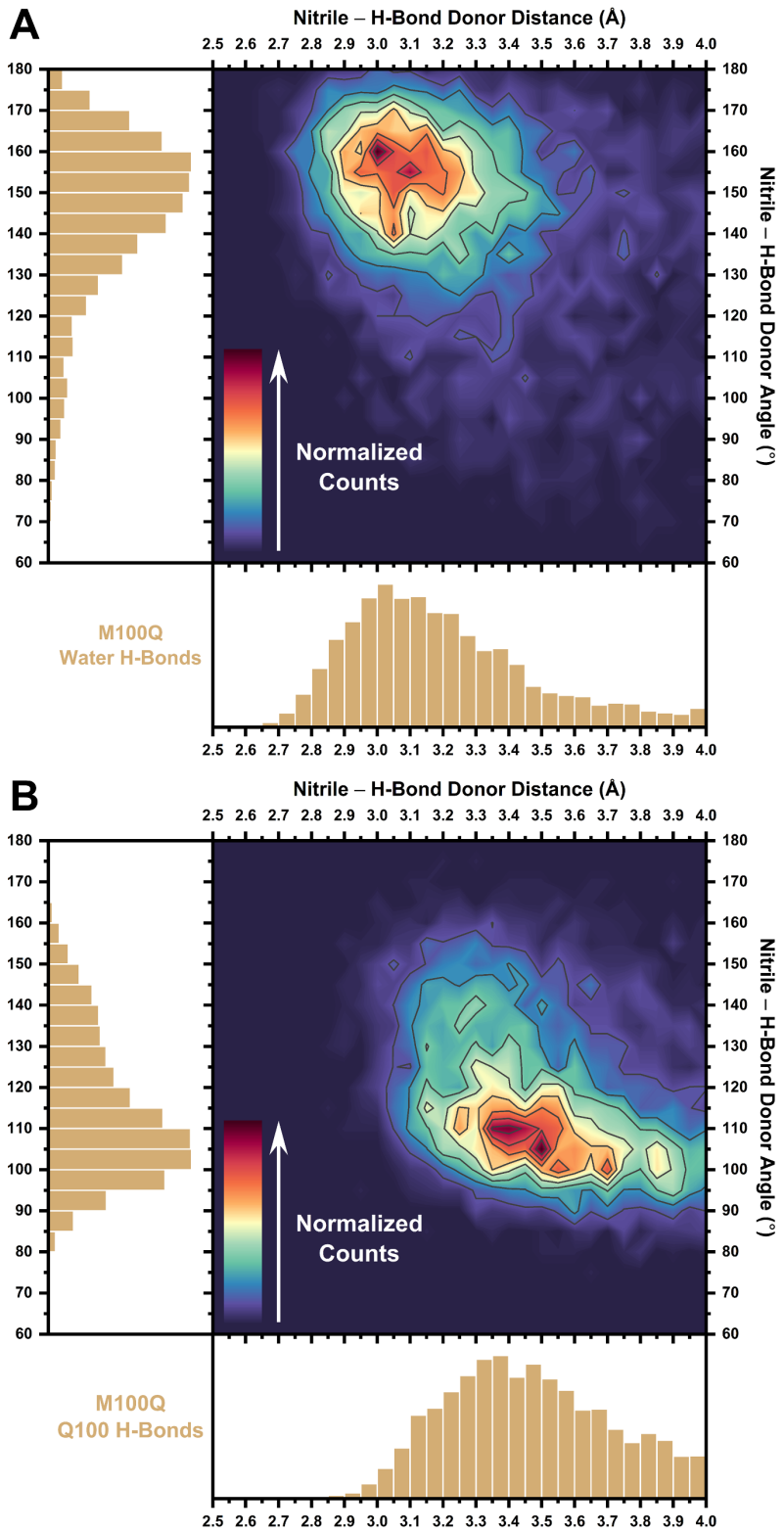

Figure S36. Nitrile H-bond angle vs distance for H-bonds with (A) water and (B) Q100 in F96oCNF M100Q PYP depicted as a contour plot describing H-bond angle/distance sampling space. 1D histograms show the sampling space projected along the individual variables.

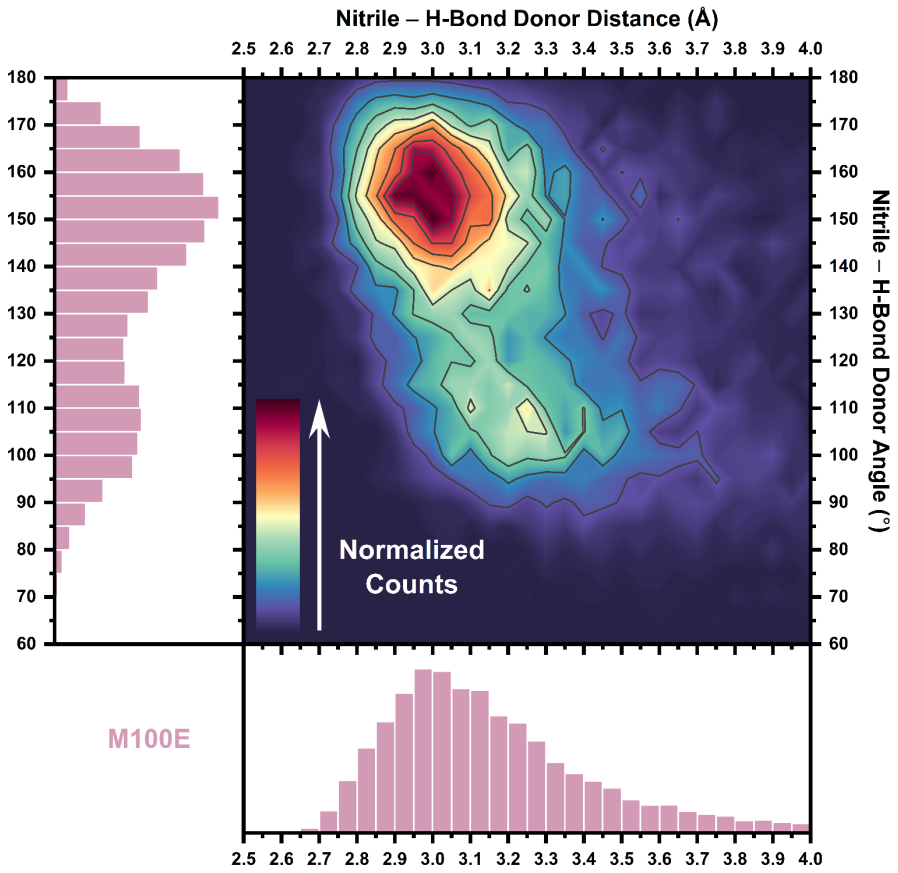

Figure S37. Nitrile H-bond angle vs distance for H-bonds in F96oCNF M100E PYP depicted as a contour plot describing H-bond angle/distance sampling space. 1D histograms show the sampling space projected along the individual variables. More sampling is observed at smaller angles than in the other F96oCNF mutants. Due to this, we further broke the contour plot down into contributions from water and T50 H-bonds with the nitrile (**Figure S38**; see **Table S20** for relative H-bonding fractions).

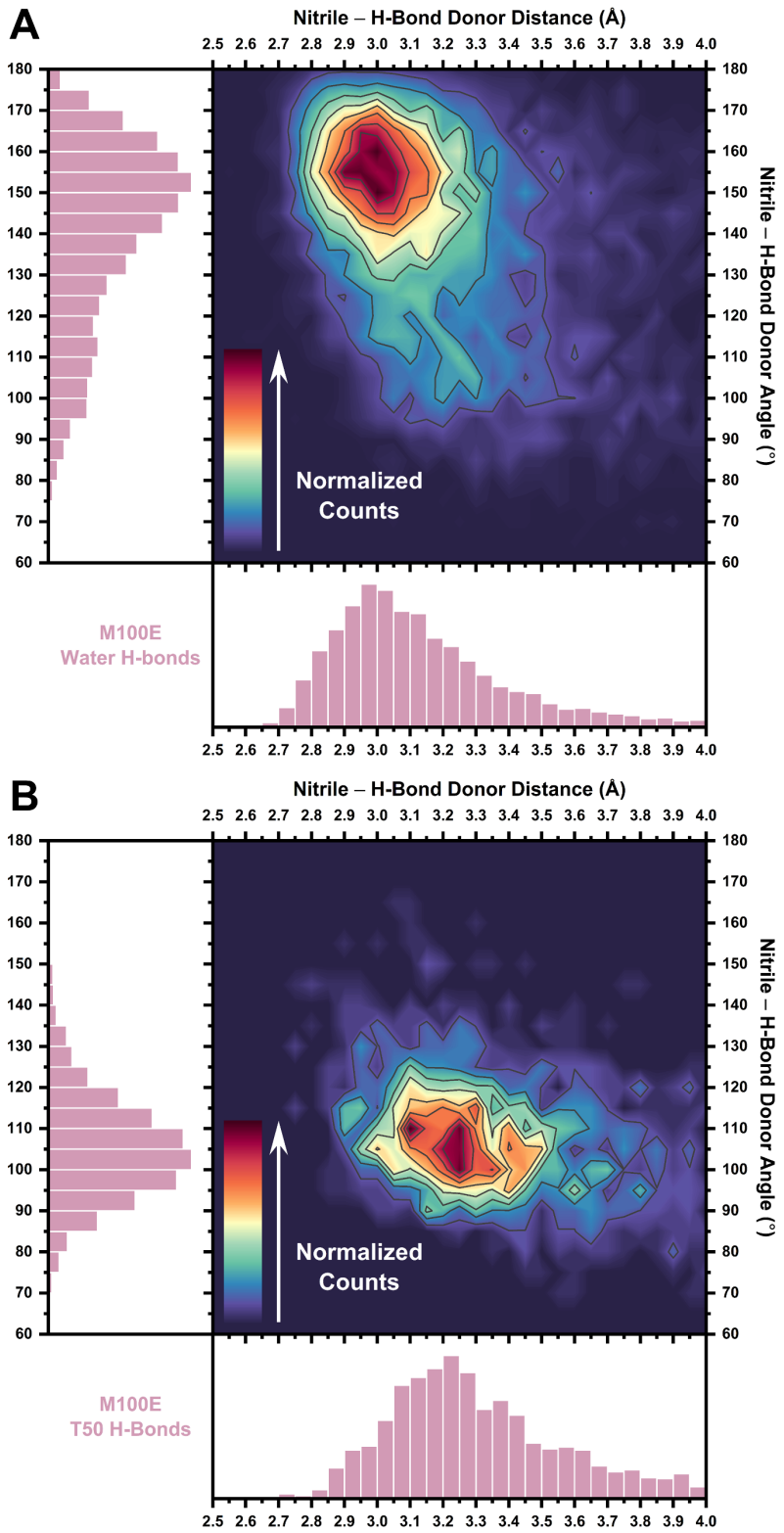

Figure S38. Nitrile H-bond angle vs distance for H-bonds with (A) water and (B) T50 in F96oCNF M100E PYP depicted as a contour plot describing H-bond angle/distance sampling space. 1D histograms show the sampling space projected along the individual variables.

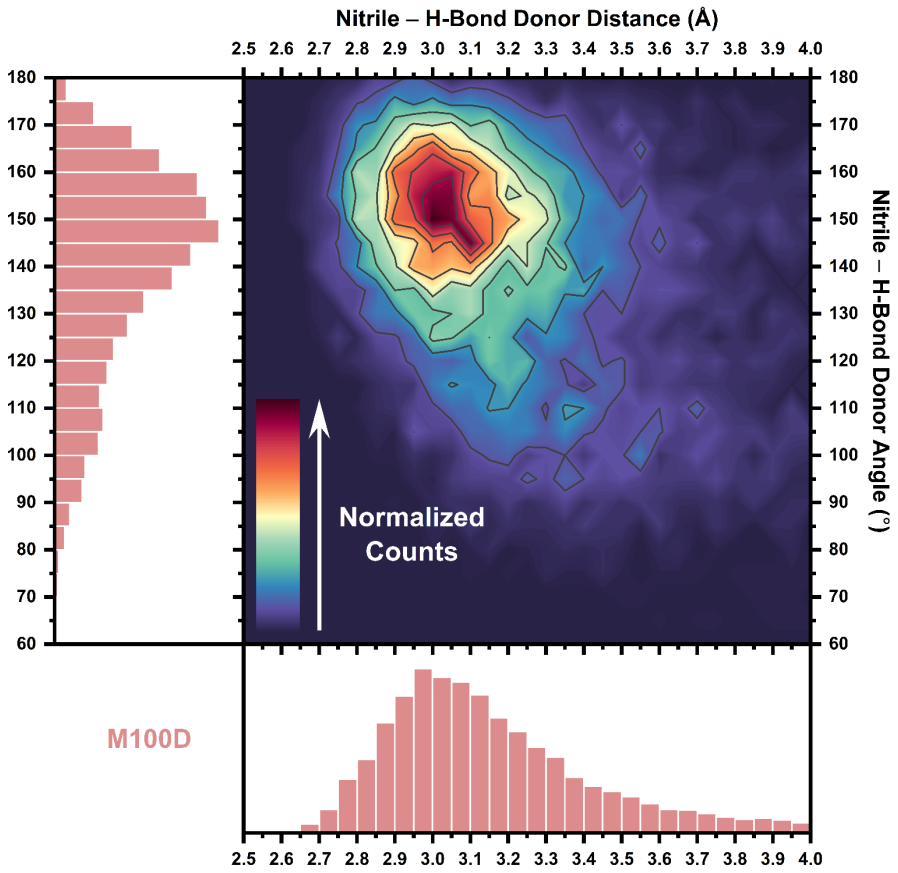

Figure S39. Nitrile H-bond angle vs distance for H-bonds in F96oCNF M100D PYP depicted as a contour plot describing H-bond angle/distance sampling space. 1D histograms show the sampling space projected along the individual variables.

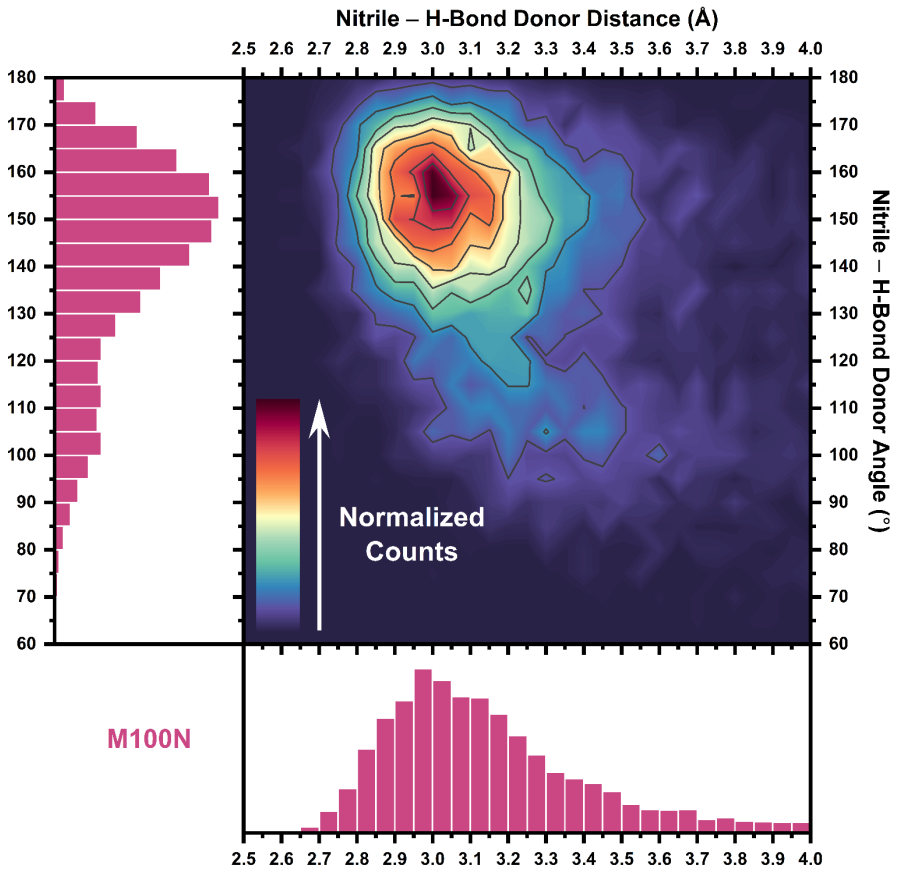

Figure S40. Nitrile H-bond angle vs distance for H-bonds in F96oCNF M100N PYP depicted as a contour plot describing H-bond angle/distance sampling space. 1D histograms show the sampling space projected along the individual variables.

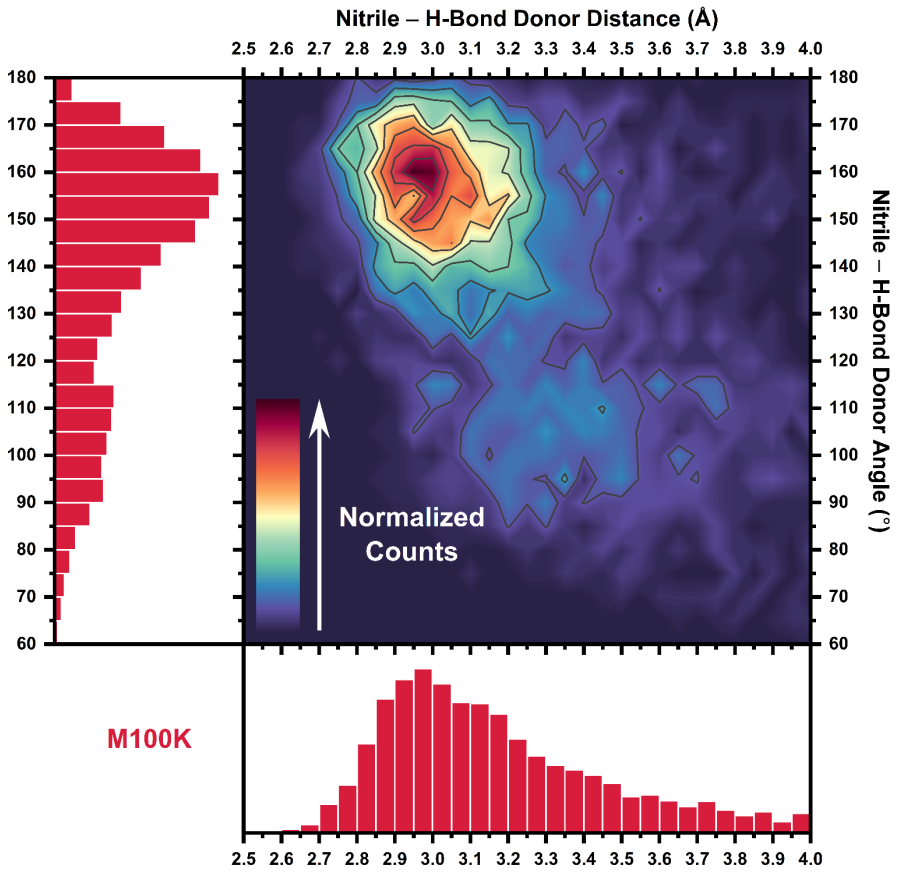

Figure S41. Nitrile H-bond angle vs distance for H-bonds in F96oCNF M100K PYP depicted as a contour plot describing H-bond angle/distance sampling space. 1D histograms show the sampling space projected along the individual variables.

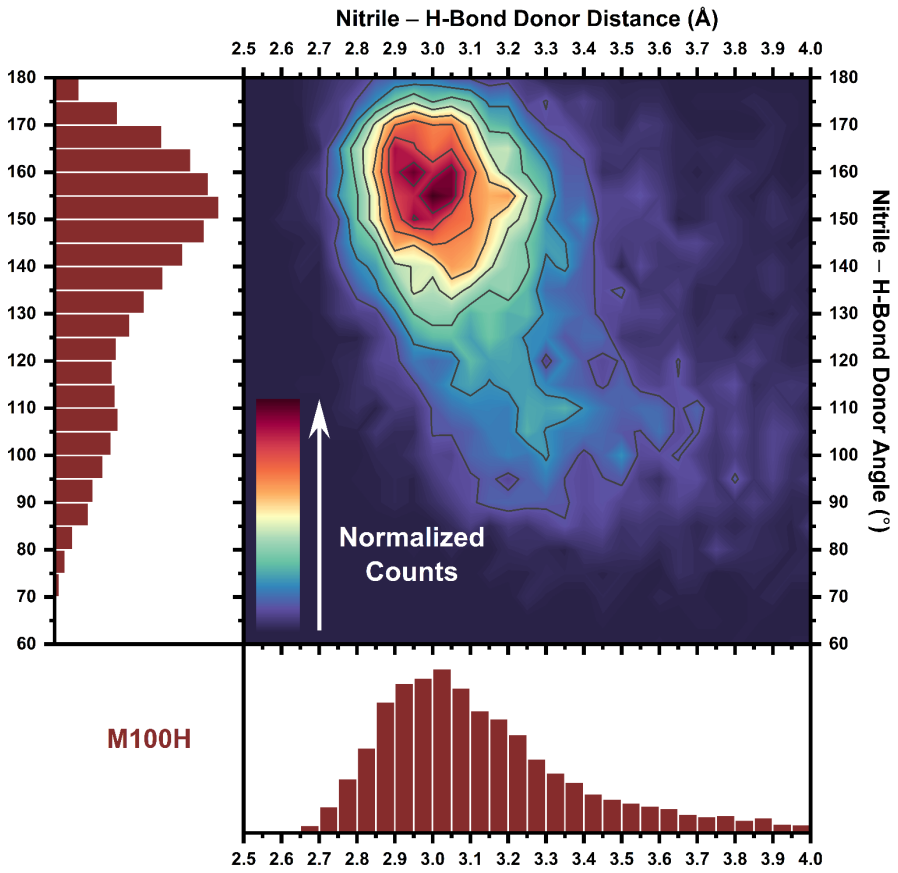

Figure S42. Nitrile H-bond angle vs distance for H-bonds in F96oCNF M100H PYP depicted as a contour plot describing H-bond angle/distance sampling space. 1D histograms show the sampling space projected along the individual variables.

Nitrile H-bond Distances and Angles Extracted from Contour Plots

To extract average H-bond donor distances and angles for the contour plots in **Figures S29–S42**, the plots were fit with a rotated Gaussian surface (*i.e.*, 2D Gaussian):

$$C = A\exp\left\{-\frac{1}{2}\left(\frac{d_{NX}\cos(\varphi)+\theta_{CNX}\sin(\varphi)-\mu_{d_{NX}}\cos(\varphi)+\mu_{\theta_{CNX}}\sin(\varphi)}{\sigma_{d_{NX}}}\right)^2-\frac{1}{2}\left(\frac{-d_{NX}\sin(\varphi)+\theta_{CNX}\cos(\varphi)-\mu_{d_{NX}}\sin(\varphi)+\mu_{\theta_{CNX}}\cos(\varphi)}{\sigma_{\theta_{CNX}}}\right)^2\right\}$$

where C is the relative number of counts, A is a scaling factor, d_{NX} is the H-bond donor distance, θ_{CNX} is the H-bond donor angle, φ is the rotation angle in the plane of the contour plot, $\mu_{d_{NX}}$ is the H-bond distance mean, $\mu_{\theta_{CNX}}$ is the H-bond angle mean, $\sigma_{d_{NX}}$ is the H-bond distance standard deviation, and $\sigma_{\theta_{CNX}}$ is the H-bond angle standard deviation (**Table S23**).

Most contour plots are well-fit by single Gaussians with average distances and angles of approximately 3.0 Å and 150°, respectively (**Table S23**). In contrast, M100Q's and M100E's contour plots show the presence of two populations (**Figure S34** & **Figure S37**). For M100Q, when fit with two Gaussians, one population is observed with a similar average H-bond geometry as for F96oCNF and the other mutants, while the other population has an average distance and angle of 3.5 Å and 110° (**Table S24**). We generated separate contour plots for M100Q nitrile–water H-bonds and nitrile–Q100 H-bonds (**Table S25**), and single Gaussian fits to both plots indicate the water H-bonds are the population present for the other nitriles, while the Q100 H-bonds are the population unique to M100Q (compare **Tables S23–S25**). An analysis of M100E in the same way reveals that the two populations arise from nitrile–water H-bonds, which also have average distances and angles of about 3.0 Å and 150°, and nitrile–T50 H-bonds, which exhibit an average distance and angle of 3.3 Å and 110° (**Table S25**; see also discussion around **Figure S26**).

Table S23. Average nitrile H-bond distances and angles from fits with rotated Gaussian surfaces in **Figures S29–S42**. Errors provided are 1σ standard errors of the fit. M100Q and M100E demonstrate a marked decrease in R^2 for one band fits compared with all other PYPs (except M100K); the presence of a second population is evident in the contour plots (**Figure S34** & **Figure S37**), and two band fits are provided in **Table S24**.

PYP Construct	Avg. H-Bond Distance, $\mu_{d_{NX}}$ (Å)	Avg. H-Bond Angle, $\mu_{\theta_{CNX}}$ (°)	H-Bond Distance s.d., $\sigma_{d_{NX}}$ (Å)	H-Bond Angle s.d., $\sigma_{\theta_{CNX}}$ (°)	R^2
F28oCNF	3.00 ± 0.003	162.8 ± 0.3	0.16 ± 0.002	16.1 ± 0.3	0.89
F92oCNF	2.91 ± 0.001	167.7 ± 0.1	0.13 ± 0.001	9.6 ± 0.1	0.96
F96oCNF	3.05 ± 0.003	154.1 ± 0.2	0.18 ± 0.003	16.2 ± 0.3	0.89
T103V	3.05 ± 0.003	157.4 ± 0.3	0.18 ± 0.003	17.5 ± 0.3	0.87
I49T	3.05 ± 0.003	157.2 ± 0.4	0.18 ± 0.003	19.5 ± 0.4	0.88
M100Q	3.31 ± 0.007	132.9 ± 0.7	0.25 ± 0.005	31.3 ± 0.7	0.79
M100E	3.06 ± 0.004	147.3 ± 0.7	0.18 ± 0.004	29.1 ± 0.8	0.81
M100D	3.06 ± 0.003	150.4 ± 0.3	0.19 ± 0.003	17.6 ± 0.3	0.88
M100N	3.05 ± 0.003	152.2 ± 0.3	0.18 ± 0.003	16.1 ± 0.3	0.89
M100K	3.02 ± 0.004	156.2 ± 0.3	0.17 ± 0.003	15.1 ± 0.3	0.80
M100H	3.03 ± 0.003	152.9 ± 0.4	0.17 ± 0.003	20.7 ± 0.4	0.87

Table S24. Average H-bond distance and angle between nitrile and H-bond donor from simultaneous fits with two 2D Gaussians. Errors provided are 1σ standard error of the fits.

PYP Construct	Avg. H-Bond Distance (Å)	Avg. H-Bond Angle (°)	H-Bond Distance s.d. (Å)	H-Bond Angle s.d. (°)	R^2
M100Q	3.15 ± 0.005	152.8 ± 0.4	0.23 ± 0.004	14.7 ± 0.4	0.88
M100Q	3.50 ± 0.009	109.1 ± 0.4	0.29 ± 0.008	12.8 ± 0.5	
M100E	3.01 ± 0.002	155.2 ± 0.2	0.16 ± 0.002	13.1 ± 0.3	0.94
M100E	3.25 ± 0.007	111.7 ± 0.6	0.24 ± 0.006	16.5 ± 0.6	
M100K	3.01 ± 0.003	158.3 ± 0.2	0.15 ± 0.003	12.5 ± 0.3	0.92
M100K	3.38 ± 0.015	117.8 ± 1.6	0.29 ± 0.01	29.8 ± 0.6	

Table S25. Average H-bond distance and angle between nitrile and H-bond donor from single Gaussian fits. Errors provided are 1σ standard error of the fits.

PYP Construct	Avg. H-Bond Distance (Å)	Avg. H-Bond Angle (°)	H-Bond Distance s.d. (Å)	H-Bond Angle s.d. (°)	R^2
$M100Q_{Water}$	3.12 ± 0.003	153.6 ± 0.2	0.21 ± 0.003	15.3 ± 0.2	0.90
$M100Q_{Q100}$	3.49 ± 0.005	111.5 ± 0.3	0.25 ± 0.005	15.1 ± 0.3	0.84
$M100E_{Water}$	3.04 ± 0.003	151.8 ± 0.3	0.17 ± 0.003	18.6 ± 0.3	0.87
$M100E_{T50}$	3.26 ± 0.004	106.6 ± 0.2	0.21 ± 0.003	11.1 ± 0.2	0.88

S9 oTN in Methanol MD and IR Analysis

AMOEBA MD Nitrile H-Bonding/Non-H-Bonding Fraction and Electric Field Characterization

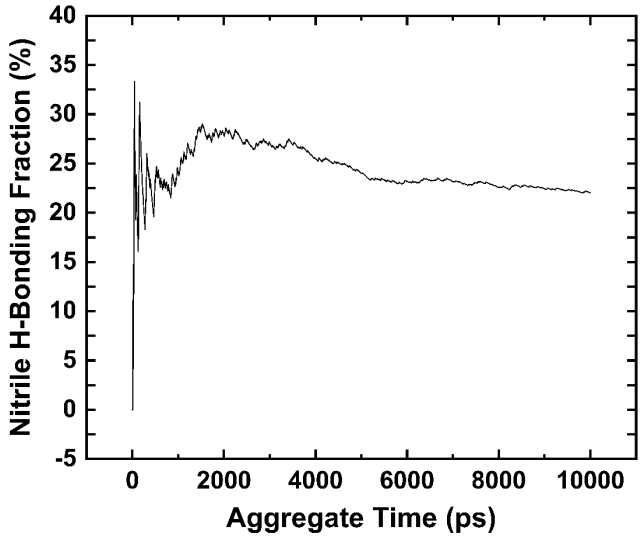

Figure S43. Moving average of the nitrile H-bonding fraction for oTN in methanol (MeOH) indicates the simulation is reasonably well converged.

Table S26. Average MD solvent electric fields and population fractions for oTN in MeOH.[5]

Environment	Median Environmental $\|\vec{F}\|$ (MV/cm)	Population Fraction (%)
MeOH Aprotic	-29.8	78.0
MeOH Protic	-67.9	22.0
MeOH	-37.9[a]	–

[a] Fraction-weighted field that is the sum of the products of the aprotic fraction and field and the protic fraction and field.

Table S27. Extracted IR peak parameters for oTN in MeOH using population fractions shown in **Table S26**.[5]

Environment	$\bar{\nu}_{\mathrm{obs}}$ (cm^{-1})[a]	FWHM (cm^{-1})[a]	$\|\vec{m}\|_{100\%}$ (D)	$\|\vec{F}\|_{\mathrm{C\equiv N,TDM}}$ (MV/cm)	$\Delta\bar{\nu}_{\mathrm{non\text{-}HB}}$ (cm^{-1})	$\Delta\bar{\nu}_{\mathrm{HB}}$ (cm^{-1})
MeOH	–	–	0.0766 ± 0.0005	-41 ± 1	–	–
MeOH Aprotic	2227.77 ± 0.02	8.42 ± 0.06	0.0743 ± 0.0009	-39 ± 1	-7 ± 1	4 ± 1
MeOH Protic	2233.21 ± 0.03	8.26 ± 0.06	0.0932 ± 0.0011	-59 ± 1	-11 ± 1	13 ± 1

[a] Peak positions ($\bar{\nu}_{\mathrm{obs}}$) and FWHM have an error of <0.1 cm^{-1}.

Assessing the Efficacy and Origins of Error in the Two-Band TDM Tuning Analysis

To assess the efficacy of the two-band TDM tuning method for oTN in MeOH, we considered the values we obtained from the method against the analogous MD values. For these comparisons, we treated the values for oTN in MeOH derived from AMOEBA as the ground truth. This is justified by our ability to recapitulate the IR spectrum of oTN in MeOH by applying the fast frequency approximation[67,68] to the instantaneous –C≡N frequencies and TDMs from AMOEBA simulations,[5] as well as the strong correlations observed between experimental and MD electric fields and H-bond shifts.[5,17] The TDM-derived electric fields (-39 ± 1 MV/cm and -59 ± 1 MV/cm for non-H-bonding and H-bonding species; **Table S27**) can be compared against the median MD H-bonding and non-H-bonding environmental electric fields, which are -30 MV/cm and -68 MV/cm, respectively; these are within 10 MV/cm of the experimentally derived values, in reasonable agreement. The H-bond shifts can be compared against expected or MD-derived values: $\Delta\bar{\nu}_{\text{HB,aprotic,MD}}$ should be 0 cm^{-1} by definition, while $\Delta\bar{\nu}_{\text{HB,protic,MD}}$ can be obtained by inputting the average H-bond distance and angle from the trajectory (3.04 Å and 159.7°; see Figure S8 and Table S8 in ref. [5]) into eq S1 and halving it, a consequence of motional narrowing from quickly fluctuating solvent H-bonding interactions:[5] this process results in a value of 12 cm^{-1}. Then, we find the TDM tuning $\Delta\bar{\nu}_{\text{HB,aprotic,TDM}}$ value (4 ± 1 cm^{-1}) is in decent agreement, while the $\Delta\bar{\nu}_{\text{HB,protic,TDM}}$ value (13 ± 1 cm^{-1}) is in excellent agreement. Overall, the two-band TDM tuning method produces reasonable to excellent values.

Aside from the error inherent in the field-TDM and field-frequency small molecule calibrations (see **Figure S5**), the sources of error in this method come from band fitting and the H-bonding/non-H-bonding fractions. To diagnose where the difference between the TDM-derived and MD-derived fields for oTN in MeOH arise from, we considered what happens when the unscaled band areas and the fractions are independently altered. **Table S28** shows the results for two scenarios: one, when the unscaled aprotic band area is made 10% smaller and the protic band area is made 10% larger, and the second, when the H-bond fraction is decreased by 3.5%. These scenarios were chosen because they are conceivable given limitations to band fitting or AMOEBA's ability to reproduce the true H-bonding fraction. When the H-bonding fraction is decreased, the protic field is closer to the AMOEBA value, but the aprotic value is largely unaltered (**Table S28**), so this alteration does not improve the aprotic field compared with AMOEBA. In contrast, when the aprotic and protic bands areas are simultaneously decreased and increased, respectively, in a manner which conserves the total band area, both the aprotic and protic fields are in good agreement with values from AMOEBA. As such, we deem it more likely the discrepancy between TDM-derived and MD-derived protic and aprotic fields in the main text is a result of band fitting rather than from incorrect H-bonding and non-H-bonding fractions. The error in the band fitting itself likely arises from the lack of clear separation between the –C≡N's protic and aprotic bands (main text **Figure 2**); this is a common problem in spectral fitting and therefore does not reflect on the efficacy of the method developed here.

It is worth noting that even in the scenario where individual electric fields are in excellent agreement, the aprotic H-bond shift is still 2 cm^{-1} rather than 0 cm^{-1} (**Table S28**). This indicates that choosing a value for $\Delta\bar{\nu}_{\text{HB,aprotic}}$ (e.g., 0 cm^{-1}) to enable two-band TDM tuning analysis, since doing so would reconcile the number of observables and unknowns when two IR bands are present, is not advisable: it is too stringent given the propagation of error of our TDM-field and frequency-field calibrations and the limitations in accuracy expected for the experiments and simulations used here. At the same time, assessments of the difference between $\Delta\bar{\nu}_{\text{HB,aprotic,TDM}}$ and 0 cm^{-1} are particularly useful since nitriles in aprotic environments have no H-bond shift. As long as $\Delta\bar{\nu}_{\text{HB,aprotic,TDM}}$ is close to

zero, the other parameters derived from this method should be physically meaningful, independent of where f_{protic} is derived from: the case of oTN in MeOH exemplifies this point.

Table S28. Hypothetical extracted IR peak parameters for oTN in MeOH if unscaled band areas or population fractions were altered from the values in the main text.

Environment	$\mathbf{f_{protic}}$ (%)	$\mathbf{f_{aprotic}}$ (%)	$\mathbf{\lvert\vec{m}\rvert_{100\%}}$ (D)	$\mathbf{\lvert\vec{F}\rvert_{C\equiv N,TDM}}$ (MV/cm)	$\mathbf{\Delta\bar{\nu}_{non\text{-}HB}}$ (cm^{-1})	$\mathbf{\Delta\bar{\nu}_{HB}}$ (cm^{-1})
MeOH Aprotic[a]	22.0%	78.0%	0.0665 ± 0.0009	-30 ± 1	-6 ± 1	2 ± 1
MeOH Protic[a]	22.0%	78.0%	0.1021 ± 0.0013	-69 ± 1	-13 ± 1	15 ± 1
MeOH Aprotic	18.5%	81.5%	0.0727 ± 0.0007	-37 ± 1	-7 ± 1	3 ± 1
MeOH Protic	18.5%	81.5%	0.1017 ± 0.0013	-68 ± 1	-13 ± 1	15 ± 1

[a] Parameters if nitrile bands in main text **Figure 5A** were altered such that the unscaled aprotic band area was 10% smaller and the protic band area was 10% larger.

S10 Nitrile Electric Fields and H-Bond Blueshifts Obtained from TDM Tuning for Multiple IR Bands

Table S29. Extracted aprotic population IR peak parameters for the P3- and P4-type nitriles.

Environment	$\bar{\nu}_{obs}$ (cm^{-1})[a]	FWHM (cm^{-1})[a]	$\lvert\vec{m}\rvert_{100\%}$ (D)	$\lvert\vec{F}\rvert_{C\equiv N,TDM}$ (MV/cm)	$\Delta\bar{\nu}_{non\text{-}HB}$ (cm^{-1})	$\Delta\bar{\nu}_{HB}$ (cm^{-1})
$M100Q_{aprotic}$[b]	2227.6 ± 0.2	7.4 ± 0.2	0.059 ± 0.002	-22 ± 3	-4 ± 1	0 ± 1
$M100E_{aprotic}$[c]	2226.16 ± 0.03	6.65 ± 0.02	0.050 ± 0.002	-13 ± 2	-2 ± 1	-3 ± 1
$M100D_{aprotic}$[c]	2225.94 ± 0.04	6.6 ± 0.1	0.053 ± 0.002	-15 ± 2	-3 ± 1	-3 ± 1
$M100N_{aprotic}$[c]	2226.08 ± 0.02	6.5 ± 0.1	0.054 ± 0.002	-17 ± 2	-3 ± 1	-2 ± 1
$M100H_{aprotic}$[c]	2226.11 ± 0.01	6.63 ± 0.03	0.055 ± 0.002	-18 ± 2	-3 ± 1	-2 ± 1

[a] Repeated from **Table S8**.

[b] Fits had 0-45% Lorentzian contribution.

[c] Fits had 0-40% Lorentzian contribution.

Table S30. Extracted protic population IR peak parameters for the P3- and P4-type nitriles.

Environment	$\bar{\nu}_{obs}$ (cm^{-1})[a]	FWHM (cm^{-1})[a]	$\lvert\vec{m}\rvert_{100\%}$ (D)	$\lvert\vec{F}\rvert_{C\equiv N,TDM}$ (MV/cm)	$\Delta\bar{\nu}_{non\text{-}HB}$ (cm^{-1})	$\Delta\bar{\nu}_{HB}$ (cm^{-1})
$M100Q_{aprotic}$[b]	2233.4 ± 0.3	8.0 ± 0.2	0.064 ± 0.002	-28 ± 2	-5 ± 1	7 ± 1
$M100E_{aprotic}$[c]	2231.3 ± 0.1	13.4 ± 0.1	0.072 ± 0.002	-36 ± 2	-7 ± 1	7 ± 1
$M100D_{aprotic}$[c]	2230.7 ± 0.1	12.7 ± 0.2	0.079 ± 0.003	-44 ± 3	-8 ± 1	8 ± 1
$M100N_{aprotic}$[c]	2230.6 ± 0.2	12.1 ± 0.1	0.073 ± 0.002	-37 ± 2	-7 ± 1	6 ± 1
$M100H_{aprotic}$[c]	2231.14 ± 0.02	13.5 ± 0.1	0.078 ± 0.003	-42 ± 3	-8 ± 1	8 ± 1

[a] Repeated from **Table S8**.

[b] Fits had 30-55% Lorentzian contribution.

[c] Fits had 10-50% Lorentzian contribution.

S11 Correlations of Experimentally and Computationally Derived Electric Fields and H-Bond Shifts

Table S31. Correlations of average and individual fields and blueshifts compare different convolutions of fields/blueshifts with H-bonding and non-H-bonding fractions.

	Average Field	**Individual Field**	**Average Blueshift**	**Individual Blueshift (Protic Population)**
Experimentally Derived	$\left\|\vec{F}\right\|_{\mathrm{C{\equiv}N,TDM}}$ (Always attainable)	$\left\|\vec{F}\right\|_{\mathrm{C{\equiv}N,aprotic,TDM}}$ *or* $\left\|\vec{F}\right\|_{\mathrm{C{\equiv}N,protic,MD}}$ (Requires input of $\mathrm{f_{aprotic}}$ *or* $\mathrm{f_{protic}}$;[a] Attainable when multiple bands present)	$\Delta\bar{\nu}_{\mathrm{HB,TDM}}$ (Attainable for single IR bands)	$\Delta\bar{\nu}_{\mathrm{HB,protic,TDM}}$ (Requires input of $\mathrm{f_{protic}}$[b])
MD-Derived	$\mathrm{f_{protic}} * \left\|\vec{F}\right\|_{\mathrm{C{\equiv}N,protic,MD}} + \mathrm{f_{aprotic}} * \left\|\vec{F}\right\|_{\mathrm{C{\equiv}N,aprotic,MD}}$	$\left\|\vec{F}\right\|_{\mathrm{C{\equiv}N,protic,MD}}$ *or* $\left\|\vec{F}\right\|_{\mathrm{C{\equiv}N,aprotic,MD}}$	$\mathrm{f_{protic}} * \Delta\bar{\nu}_{\mathrm{HB,protic,MD}}$[b,c,d]	$\Delta\bar{\nu}_{\mathrm{HB,protic,MD}}$[d]

[a] $(|\vec{m}|_{100\%,\mathrm{protic}} - |\vec{m}|_0)/|\underline{A}| = \left|\vec{F}\right|_{\mathrm{C{\equiv}N,protic,TDM}}$, where "unscaled band area"$/\mathrm{f_{protic}} = |\vec{m}|_{100\%,\mathrm{protic}}$.

[b] $\mathrm{f_{protic}} * \Delta\bar{\nu}_{\mathrm{HB,protic,MD}} + \mathrm{f_{aprotic}} * \Delta\bar{\nu}_{\mathrm{HB,aprotic,MD}} = \mathrm{f_{protic}} * \Delta\bar{\nu}_{\mathrm{HB,protic,MD}} + \mathrm{f_{aprotic}} * 0 = \mathrm{f_{protic}} * \Delta\bar{\nu}_{\mathrm{HB,protic,MD}}$.

[c] $\bar{\nu}_{\mathrm{obs}} - \bar{\nu}_0 - |\Delta\vec{\mu}|\left|\vec{F}\right|_{\mathrm{C{\equiv}N,protic,TDM}} = \Delta\bar{\nu}_{\mathrm{HB,protic,TDM}}$, where $\left|\vec{F}\right|_{\mathrm{C{\equiv}N,protic,TDM}}$ is derived as in [a].

[d] Derived by inputting the average H-bonding distance and angle (**Table S23** & **Table S25**) into eq S1 and dividing the result by 2 for solvent H-bonds.

Details on Error Propagation for $\Delta\bar{\nu}_{HB,TDM}$

When the TDM tuning method is employed for single IR bands, the error in $\Delta\bar{\nu}_{\mathrm{HB,TDM}}$ arises from the error in the fields inferred via the TDM. Then, the absolute error in $\Delta\bar{\nu}_{\mathrm{HB,TDM}}$ is estimated by multiplying the inferred electric field error from TDM tuning by the Stark tuning rate ($0.18\ \frac{cm^{-1}}{\frac{MV}{cm}}$)[1]. In this study, nitriles in the proteins have a field error of 2 MV/cm (**Table S6**). Then, for the benchmarking in main text **Figure 6**, we conservatively use an error of ± 0.5 cm^{-1} for $\Delta\bar{\nu}_{\mathrm{HB,TDM}}$ for single IR bands.

For the extended TDM tuning method with two bands, we propagate the error for the TDM of the entire bandshape with the fitting error for either the protic or aprotic band to obtain the error for the unscaled band areas. Our fitting typically does not lead to electric field errors larger than 2 MV/cm (**Table S27**; **Table S29**; **Table S30**), so the error estimates for H-bond shifts would be similar to those for single band nitriles. However, in the two band TDM tuning method implemented in the main text, we use the AMOEBA MD H-bonding and non-H-bonding fractions to obtain the scaled band areas from the unscaled band areas. Therefore, to reflect the increased uncertainty in values from the two-band method, we use an error of ± 1 cm^{-1} for $\Delta\bar{\nu}_{\mathrm{HB,protic,TDM}}$ for nitriles with two IR bands.

Calculation of $\Delta\bar{\nu}_{HB,MD}$

Values for $\Delta\bar{\nu}_{HB,MD}$ and all values necessary to obtain it are provided in **Table S32**. The values were calculated as follows: for all H-bonding nitrile species, the average H-bonding distance and angle (**Table S23** & **Table S25**) were input into eq S1.[5] In cases where multiple H-bond donors were observed in the MD simulations, $\Delta\bar{\nu}_{HB,MD}$ is the average of the H-bond shifts from each environment. If the H-bond donor was the solvent, the value from eq S1 was divided by 2. This is because the H-bond shift is motionally narrowed to an averaged value when the H-bond donor is quickly making and breaking the hydrogen bond over the nitrile's vibrational excited state dephasing time.[5] While prior work indicated the most accurate computational method to obtain $\Delta\bar{\nu}_{HB,MD}$ for nitriles H-bonded with solvents is to directly simulate their IR spectra, we found halving the prediction from eq S1 was a reasonable approximation.[5] Additionally, for single band nitriles, the value from eq S1 or half the value was multiplied by the H-bonding fraction, since the experimental values are averaged with the H-bond shift for the aprotic fraction (0 cm^{-1}), in analogy to the fraction-weighted MD electric fields. For nitriles exhibiting two IR bands, no weighting is necessary, as the method in main text **Figure 5** obtains all TDM tuning parameters as if the band were 100% of the sample concentration. We distinguish between bands that need to be multiplied by the H-bonding fraction or not by the presence or absence of a value for f_{protic} in **Table S32.**

We previously found eq S1, which was derived using oTN's nitrile, also worked well to describe H-bond shifts for various aliphatic nitriles in metal organic frameworks (MOFs).[5,65] We include H-bond shifts for these MOFs in some of the following comparisons. Experimentally and computationally derived MOF nitrile H-bond shifts were taken from ref. [5]. Experimentally derived values were originally obtained from the difference in the observed Raman shift and the expected frequency using AMOEBA electric fields if the VSE were the only frequency tuning mechanism.[65] Since these points are derived from frequencies, which have much less error than TDMs, but rely on the accuracy of the MD fields, we used an error of 0.5 cm^{-1} to reflect our relative confidence in the values. Despite not being derived from TDMs, we still refer to these values as $\Delta\bar{\nu}_{HB,TDM}$ for simplicity. MOF $\Delta\bar{\nu}_{HB,MD}$ values were originally obtained by inputting the average H-bond distance and angle information from ref. [65] into eq S1. For the nitriles in MOFs, we assume the nitrile H-bonding fraction is 100%.

Table S32. Compiled TDM- and MD-derived H-bond blueshifts using average H-bond distance and angle in **Table S23** or **Table S25**. MOF nitrile H-bond distance and angle are from ref. [65].

Environment	f_{protic} (%)	eq S1 Output	H-Bond Donor	$\Delta\bar{\nu}_{HB,MD}$ (cm^{-1})	$\Delta\bar{\nu}_{HB,TDM}$ (cm^{-1})
F28oCNF	63.6%	25.4	Solvent	8.1	7
F92oCNF	75.1%	29.1	Protein	21.9	21
F96oCNF	25.0%	25.4	Solvent	2.9	2
T103V	25.7%	23.2	Solvent	3.0	2
I49T	33.5%	23.1	Solvent	3.9	2
M100Q	–	19.7 (47.0%) 5.6 (53.0%)	Solvent Protein	7.6	7
M100E	–	24.2 (86.7%) 9.4 (13.3%)	Solvent Protein	11.7	7
M100D	–	21.9	Solvent	11.0	8
M100N	–	22.5	Solvent	11.3	6
M100H	–	23.1	Solvent	11.6	8
$oTN_{methanol}$	–	23.7[a]	Solvent	11.9	13
oTN_{water}	81.8%[a]	25.8[a]	Solvent	10.6	12
Allylic Acid (AA)	100%[b]	31.8	MOF	31.8	33.5
Benzoic Acid (CPh)	100%[b]	28.2	MOF	28.2	25.6
Isophthalic Acid (DCPh)	100%[b]	32.2	MOF	32.2	28.6

[a] Values derived from ref. [5].
[b] By assumption.

Alternative Approach to Obtain $\Delta\bar{\nu}_{HB,MD}$

MD- and TDM-derived H-bond blueshifts for P4-type nitriles had values between 11–12 cm^{-1} and 6–8 cm^{-1}, respectively (**Table S32**). Because these values have a larger difference than other MD- and TDM-derived shifts (**Table S32**), we were curious to understand the origin of the difference. To investigate this, we calculated $\Delta\bar{\nu}_{HB,MD}$ for all nitrile-containing PYPs using a method which derives the blueshift by comparing the experimental frequency and the AMEOBA H-bonding median field (we used this method for previous comparisons of experimentally and computationally derived H-bond shifts[5]). Specifically, to facilitate comparisons with $\Delta\bar{\nu}_{HB,TDM}$, we used the best fit line in main text **Figure 6C** to convert the AMOEBA H-bonding and non-H-bonding fields to "TDM-derived" fields. These were then converted to frequencies using the oTN Stark tuning rate.[1] For nitriles with one IR band, the observed frequency is equal to the fraction-weighted frequency. To derive $\Delta\bar{\nu}_{HB,MD}$, the predicted, MD protic band frequency was first obtained by manipulating this relationship for the experimental frequency (**Table S6**), the MD non-H-bonding frequency, and the MD H-bonding and non-H-bonding fractions. Then, $\Delta\bar{\nu}_{HB,MD}$ was obtained by taking the difference of the predicted protic band frequency and the MD "TDM-derived" H-bonding frequency and multiplying by the H-bonding fraction. For two-band nitriles, obtaining $\Delta\bar{\nu}_{HB,MD}$ is simpler, as it is the difference between the experimental frequency (**Table S8**) and the "TDM-derived" H-bonding frequency. Values derived from this method are provided in **Table S33**. $\Delta\bar{\nu}_{HB,MD}$ ranges from 6–8 cm^{-1} for the P4-type nitriles, the same range as was observed for $\Delta\bar{\nu}_{HB,protic,TDM}$ (**Table S32**). Moreover, the rest of the nitriles have very similar H-bond shifts to the values in **Table S32**. An alternative comparison of main text **Figure 6D** using these values is shown in **Figure S44**.

Our analysis indicates the difference between $\Delta\bar{\nu}_{HB,protic,TDM}$ and $\Delta\bar{\nu}_{HB,protic,MD}$ values for P4-type nitriles comes from deviations in $\Delta\bar{\nu}_{HB,protic,MD}$, which are likely due to the assumption that MD-derived H-bond shifts can be obtained by halving values from eq S1 due to motional averaging.[5] This assumption is violated in the case where the nitrile H-bonding and non-H-bonding exchange lifetimes are not equal, which may be the case for the P4-type nitriles. Unfortunately, we cannot investigate this further with our current MD outputs (see **Section S13**).

Table S33. H-bond blueshifts obtained by comparing AMOEBA-derived nitrile frequencies with experimental values (using **Table S6** for single IR band nitriles and **Table S8** for two band nitriles).

Environment	$\Delta\bar{\nu}_{HB,MD}$ (cm^{-1})
F28oCNF	7.1
F92oCNF	20.5
F96oCNF	3.5
T103V	3.5
I49T	3.0
M100Q	7.4
M100E	7.4
M100D	6.7
M100N	6.8
M100H	7.8
M100E	7.4
oTN$_{methanol}$	13.3
oTN$_{water}$	11.5

Alternative H-Bond Blueshift Correlations

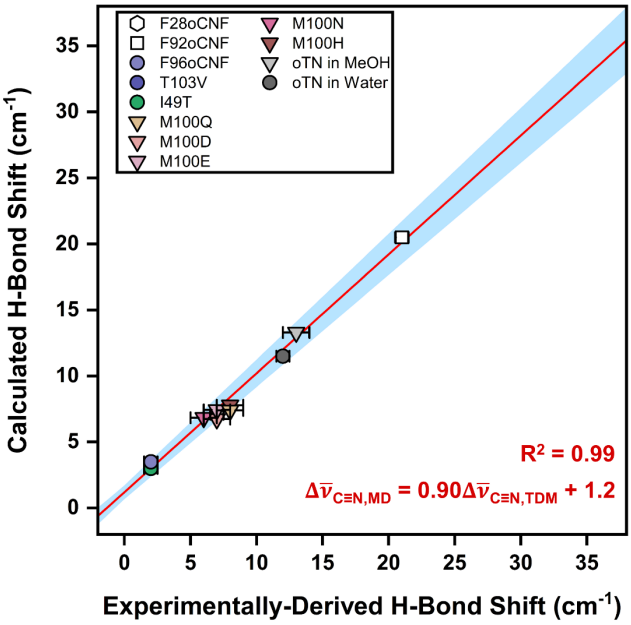

Figure S44. Correlation between $\Delta\bar{\nu}_{HB,TDM}$ and $\Delta\bar{\nu}_{HB,MD}$ values for the alternative method of obtaining $\Delta\bar{\nu}_{HB,MD}$ described above. $\Delta\bar{\nu}_{HB,TDM}$ values are from **Table S32** and $\Delta\bar{\nu}_{HB,MD}$ values are from **Table S33.** The correlation coefficient squared and error-weighted line regression are shown in red with $R^2 = 0.99$ and $\Delta\bar{\nu}_{HB,MD} = (0.90 \pm 0.03)\ \Delta\bar{\nu}_{HB,TDM} + (1.2 \pm 0.3)$, respectively. The light blue region indicates the 2σ confidence interval. Downward-facing triangular symbols indicate blueshifts for individual protic populations, while non-downward-facing triangular symbols indicate blueshifts for averaged populations.

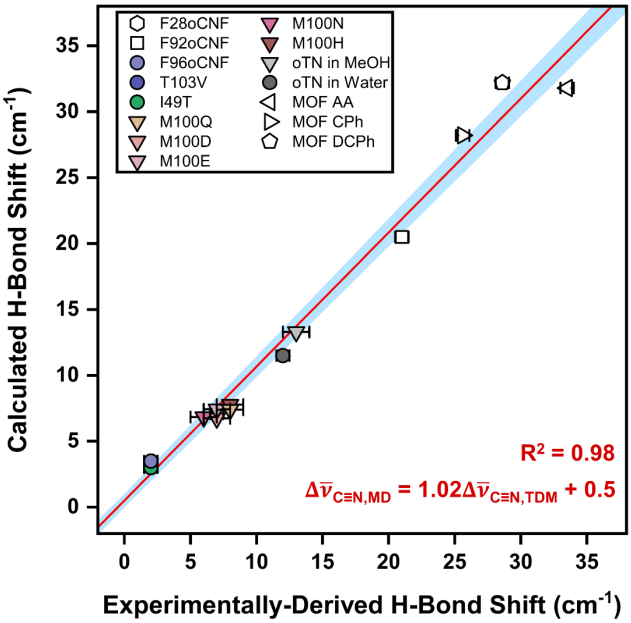

Figure S45. Correlation between $\Delta\bar{\nu}_{HB,TDM}$ and $\Delta\bar{\nu}_{HB,MD}$ values shown in **Figure S44** with MOF data points[65]. The correlation coefficient squared and error-weighted line regression are shown in red with R^2 = 0.98 and $\Delta\bar{\nu}_{HB,MD}$ = (1.02 ± 0.01) $\Delta\bar{\nu}_{HB,TDM}$ + (0.5 ± 0.3), respectively. The light blue region indicates the 2σ confidence interval. Downward-facing triangular symbols indicate blueshifts for individual protic populations, while non-downward-facing triangular symbols indicate blueshifts for averaged populations.

S12 Implications from this Study

Reinterpreting Issues of the AMBER Force Field to Recapitulate TDM-Derived Electric Fields

With our new understanding that AMOEBA correctly simulates nitrile H-bonding environments because it properly models thermodynamic and kinetic parameters in both fast and slow H-bonding exchange settings, we briefly return to our benchmarking study of the original PYP variants to better understand why AMBER could not recapitulate nitrile electric fields in H-bonding envrironments.[17] Specifically, we investigated AMBER kinetics and thermodynamics of noncovalent interactions in those settings by comparing the H-bonding fractions and electric fields for the nitriles with some H-bonding, i.e., F28oCNF, F92oCNF, and F96oCNF. Using the fractions and fields for AMOEBA from main text **Table 2** and for AMBER from Table 2 in ref. [17], we found that the AMBER fractions are 0.1 – 0.8 fold smaller than the AMOEBA fractions, and the AMBER fields are 0.4 – 0.5 fold smaller than the AMOEBA fields. If AMBER's kinetics for nitrile H-bonding environments were properly modeled, we would expect the fractions to be comparable to AMOEBA's, and if the thermodynamics were well-modeled, we would expect the fields to be approximately ¾'s of the AMOEBA fields (see Figure S19 in ref. [17]). The comparisons imply that AMBER can misrepresent kinetics *or* thermodynamics for H-bonding nitriles, where the extent of incorrectness depends on the specific context. This finding suggests it may be difficult to make simple modifications to AMBER to better model interactions for nitrile probes.

On the Accessibility and Potential of the AMOEBA Force Field

Polarizable force fields like AMOEBA have generally been less utilized due to their computational cost and the complexity of their implementation.[69] However, recent advances have sought to ameliorate these issues.[70–72] For AMOEBA specifically, the Tinker and OpenMM platforms which run the force field have integrated GPUs,[15,73,74] and new software like TinkerModeller facilitate the conversion of MD input files from classical to polarizable force fields.[75] With these tools, AMOEBA simulations are more accessible, and they routinely reach the hundreds of nanoseconds to microseconds timescale for small to medium sized proteins.[76–80] As a case in point, we have cumulatively simulated over 2.5 microseconds of MD for the nitrile-containing PYPs on Tinker9,[15] and recent studies have collected over 5 and 25 microseconds of AMOEBA data on Tinker-HP.[74,81–83] In addition to the nitrile fields and H-bond blueshifts studied here and previously,[5,17] the utilization of AMOEBA has been found necessary for recapitulating experimental observables in proteins like electric fields at enzyme active sites,[84,85] binding free energies,[86–88] NMR chemical shifts[89] and relaxation parameters,[90] pK_a values,[91–93] and ion channel conductances,[76,94] among others, highlighting its capabilities.

S13 Nitrile Non-H-Bonding/H-Bonding State Residence Times: Attempting to Predict Chemical Exchange Regimes

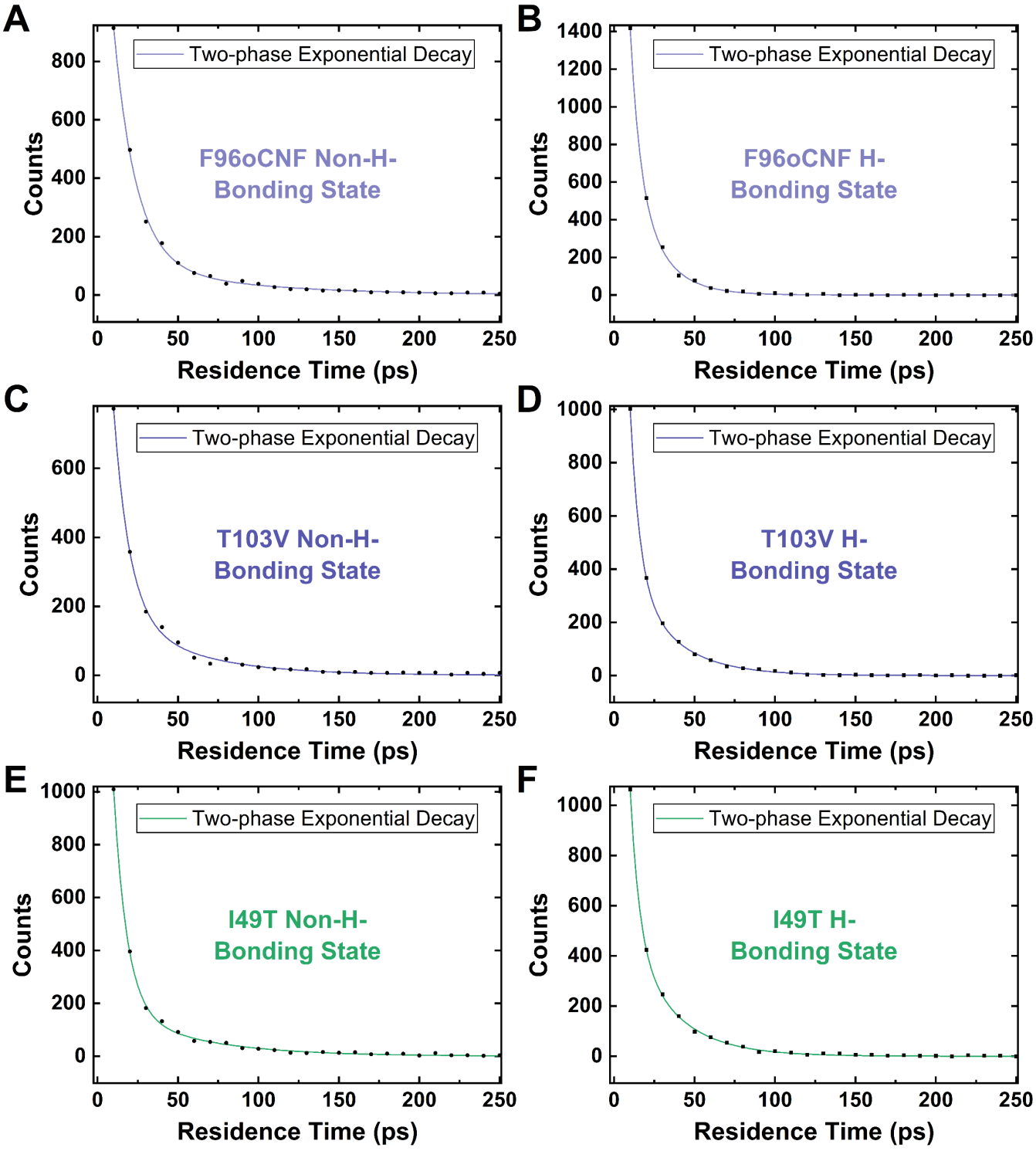

Figure S46. Frequency of residence times in MD of the non-H-bonding (circles) and H-bonding states (squares) for (A/B) F96oCNF, (C/D) T103V, and (E/F) I49T. Fit parameters from two-phase exponentials are shown in **Table S34.**

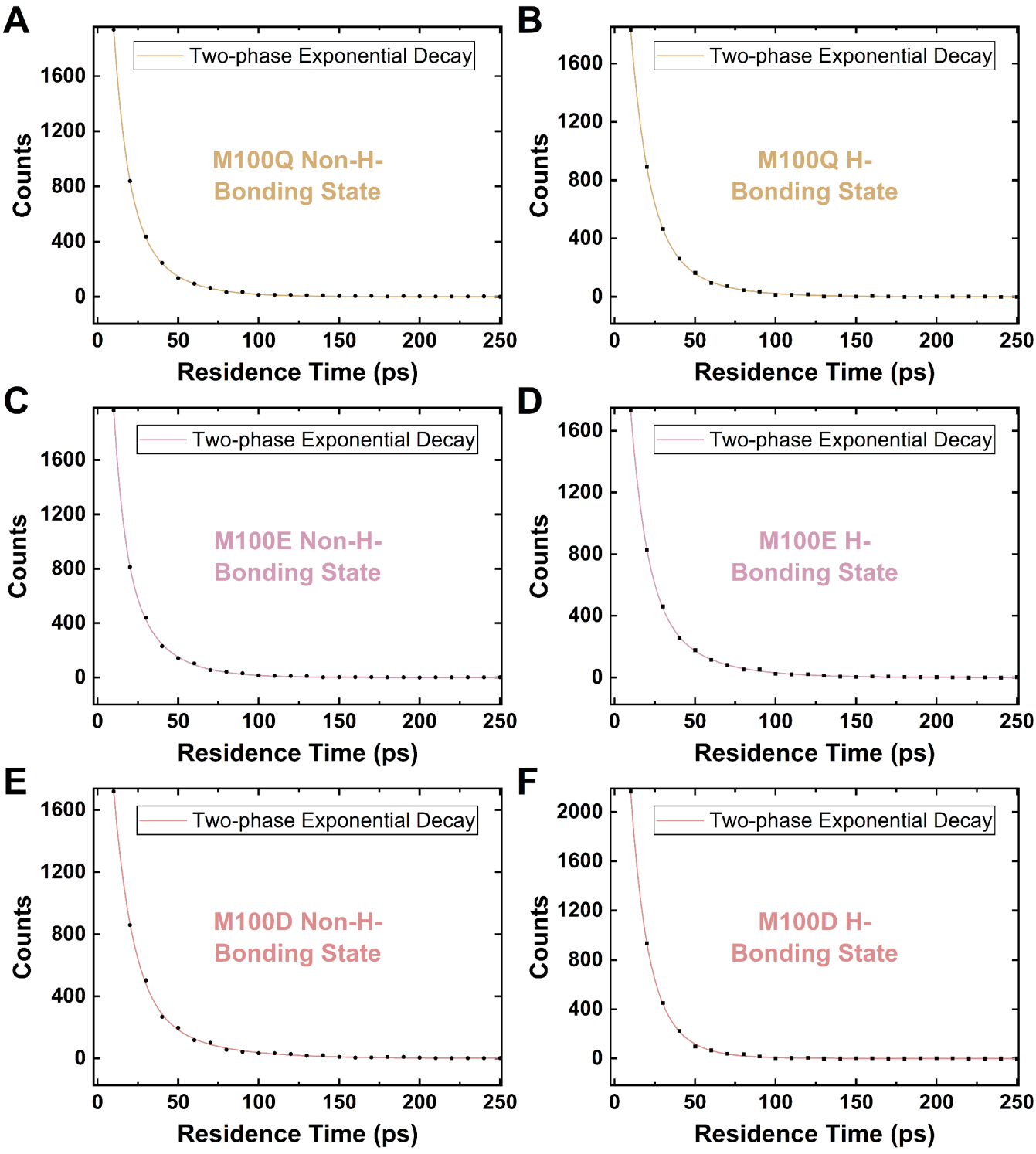

Figure S47. Frequency of residence times in MD of the non-H-bonding (circles) and H-bonding states (squares) for (A/B) M100E, (C/D) M100Q, and (E/F) M100D. Fit parameters from two-phase exponentials are shown in **Table S34.**

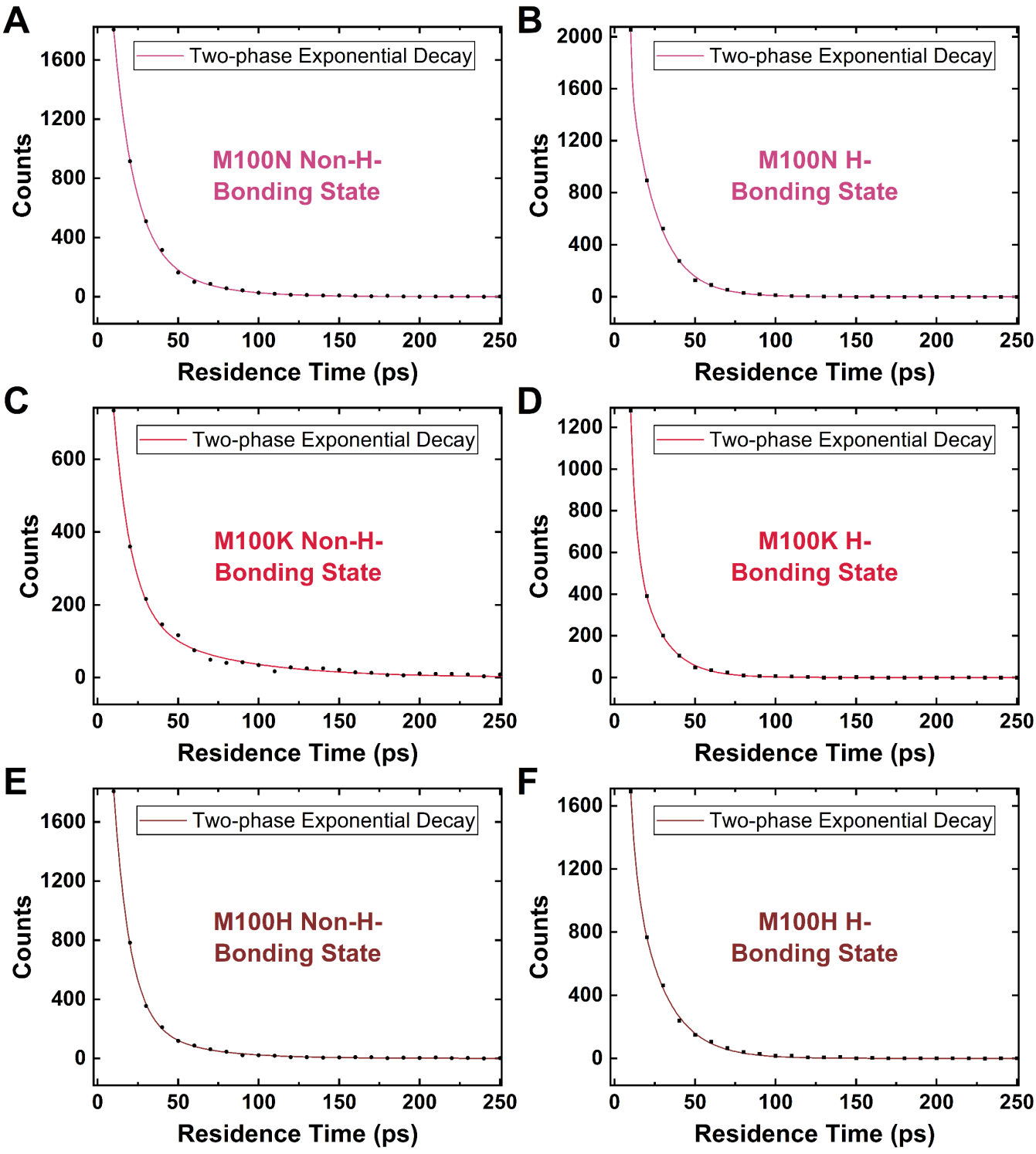

Figure S48. Frequency of residence times in MD of the non-H-bonding (circles) and H-bonding states (squares) for (A/B) M100N, (C/D) M100K, and (E/F) M100H. Fit parameters from two-phase exponentials are shown in **Table S34.**

Fitting Nitrile Non-H-Bonding/H-Bonding State Lifetimes

Single or double decaying exponentials fits to the plots in **Figures S46–S48** have the form:

$$c = A_1 exp\left(\frac{-t_r}{\tau_1}\right) + A_2 exp\left(\frac{-t_r}{\tau_2}\right)$$

where c is the number of counts, t_r is the residence time, and A_1/A_2 and τ_1/τ_2 are the amplitudes and lifetimes, respectively, of the first and second exponentials. The fitting coefficients for F96oCNF and mutants are provided in **Table S34**. Note that MD frames were output every 10 ps, a consequence of the longer simulation times needed to sample the protein degrees of freedom.

While we expected to find that P1-type and P2-type nitriles would have different state lifetimes on average compared with P3-type and P4-type nitriles (particularly with respect to the fast lifetime component), a survey of **Table S34** does not indicate this. We do not attribute this to the simulations' inability to distinguish between nitriles in various exchange regimes, but rather to the 10 ps spacing between saved coordinates, which functionally blinds us from discerning lifetimes much shorter than that interval. This attribution is supported by our recent work where 2 ns aggregate AMOEBA simulations for F28oCNF and oTN in water and methanol were carried out with data saved every 20 fs:[5] this strategy enabled the recapitulation of IR lineshapes from the MD simulations using the fluctuating frequency approximation.[68,95] As the nitriles in F28oCNF and oTN in water are in fast exchange while oTN's nitrile in methanol is in a relatively slower exchange regime (evidenced by the presence of two bands in main text **Figure 5**), it appears that AMOEBA can properly model nitrile H-bonding/non-H-bonding states with variable lifetimes if the outputs are saved frequently enough (i.e., approximately every 10 to 100 fs[96–98]). Consequently, a more rigorous analysis of the lifetimes for the nitriles in F96oCNF and the mutants would require rerunning MD simulations and adopting an analogous strategy, which is outside the scope of this work.

Table S34. Parameters from exponential fits for non-H-bonding and H-bonding nitrile populations.

PYP Construct	State	A_1	τ_1 (ps)	A_2	τ_2 (ps)
F96oCNF	Non-H-Bonding	1664 ± 15	13.7 ± 0.3	129 ± 13	72.6 ± 0.5
F96oCNF	H-Bonding	3983 ± 153	6.0 ± 0.4	1168 ± 144	17.6 ± 0.8
T103V	Non-H-Bonding	1642 ± 32	9.9 ± 0.3	219 ± 23	46.9 ± 3.0
T103V	H-Bonding	3509 ± 63	5.9 ± 0.1	526 ± 14	27.2 ± 0.4
I49T	Non-H-Bonding	2801 ± 46	8.3 ± 0.2	215 ± 14	50.7 ± 2.2
I49T	H-Bonding	3504 ± 134	5.6 ± 0.2	675 ± 24	27.3 ± 0.5
M100Q	Non-H-Bonding	4019 ± 45	8.4 ± 0.2	1070 ± 81	24.4 ± 0.5
M100Q	H-Bonding	3315 ± 36	11.0 ± 0.2	697 ± 50	28.9 ± 0.8
M100E	Non-H-Bonding	4565 ± 81	6.9 ± 0.2	1417 ± 81	22.1 ± 0.5
M100E	H-Bonding	3206 ± 29	10.2 ± 0.2	726 ± 46	31.6 ± 0.9
M100D	Non-H-Bonding	3064 ± 44	11.7 ± 0.3	543 ± 65	37.2 ± 2.1
M100D	H-Bonding	3720 ± 257	8.8 ± 0.5	1711 ± 329	17.9 ± 0.9
M100N	Non-H-Bonding	3002 ± 86	11.9 ± 0.4	707 ± 113	30.4 ± 1.8
M100N[a]	H-Bonding	$3.2 \times 10^7 \pm 7.2 \times 10^{10}$	0.9 ± 177	2908 ± 85	17.1 ± 0.2
M100K	Non-H-Bonding	1393 ± 25	11.0 ± 0.4	204 ± 17	57.5 ± 3.2
M100K	H-Bonding	15158 ± 3359	3.1 ± 0.3	1259 ± 44	16.2 ± 0.3
M100H	Non-H-Bonding	4141 ± 10	9.9 ± 0.1	392 ± 35	35.4 ± 1.6
M100H	H-Bonding	8563 ± 4223	3.4 ± 0.6	2088 ± 74	19.4 ± 0.3

[a] The lifetime of the first component is small enough that the amplitude and lifetime cannot be meaningfully determined. Values provided are for the fit shown in **Figure S48**.

S14 References

(1) 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.

(2) Hammill, J. T.; Miyake-Stoner, S.; Hazen, J. L.; Jackson, J. C.; Mehl, R. A. Preparation of Site-Specifically Labeled Fluorinated Proteins for 19F-NMR Structural Characterization. *Nat Protoc* **2007**, *2* (10), 2601–2607.

(3) Lee, B.; Papoutsis, B. M.; Wong, N. Y.; Piacentini, J.; Kearney, C.; Huggins, N. A.; Cruz, N.; Ng, T. T.; Hao, K. H.; Kramer, J. S.; Fenlon, E. E.; Nerenberg, P. S.; Phillips-Piro, C. M.; Brewer, S. H. Unraveling Complex Local Protein Environments with 4-Cyano-l-Phenylalanine. *J. Phys. Chem. B* **2022**, *126* (44), 8957–8969.

(4) Kumar, A.; Woolley, G. A. Origins of the Intermediate Spectral Form in M100 Mutants of Photoactive Yellow Protein. *Photochemistry and Photobiology* **2015**, *91* (4), 985–991.

(5) 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.

(6) Boutwell, D. R.; Cao, A. L.; Walker, A. S.; Buchanan, L. E. Refining Structural Analysis of Proteins: Automated Methods to Measure Transition Dipole Strength of Single Residues. *J. Phys. Chem. B* **2025**, *129* (33), 8360–8367.

(7) Heermant, S.; Paschke, R. R.; Kozuch, J. Antiparallel to Parallel β-Sheet Transition of Amyloid-β Is Promoted by the Ability to Incor-Porate into Membranes. ChemRxiv February 24, 2025.

(8) Rackers, J. A.; Wang, Z.; Lu, C.; Laury, M. L.; Lagardère, L.; Schnieders, M. J.; Piquemal, J.-P.; Ren, P.; Ponder, J. W. Tinker 8: Software Tools for Molecular Design. *J. Chem. Theory Comput.* **2018**, *14* (10), 5273–5289.

(9) 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. Jr.; 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.

(10) Kozuch, J.; Schneider, S. H.; Zheng, C.; Ji, Z.; Bradshaw, R. T.; Boxer, S. G. Testing the Limitations of MD-Based Local Electric Fields Using the Vibrational Stark Effect in Solution: Penicillin G as a Test Case. *J. Phys. Chem. B* **2021**, *125* (17), 4415–4427.

(11) Zheng, C.; Mao, Y.; Kozuch, J.; Atsango, A. O.; Ji, Z.; Markland, T. E.; Boxer, S. G. A Two-Directional Vibrational Probe Reveals Different Electric Field Orientations in Solution and an Enzyme Active Site. *Nat. Chem.* **2022**, *14* (8), 891–897.

(12) Walker, B.; Liu, C.; Wait, E.; Ren, P. Automation of AMOEBA Polarizable Force Field for Small Molecules: Poltype 2. *Journal of Computational Chemistry* **2022**, *43* (23), 1530–1542.

(13) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Rob, M. A.; Cheeseman, J. R.; Scalmani, G.; et al. Gaussian 16; Gaussian, Inc.: Wallingford, CT, 2016.

(14) Fried, S. D.; Wang, L.-P.; Boxer, S. G.; Ren, P.; Pande, V. S. Calculations of the Electric Fields in Liquid Solutions. *The Journal of Physical Chemistry B* **2013**, *117* (50), 16236–16248.

(15) Tinker9: Next Generation of Tinker with GPU Support. Zhi Wang, Washington University in St. Louis, 2021, Https://Github.Com/TinkerTools/Tinker9.

(16) Shi, Y.; Xia, Z.; Zhang, J.; Best, R.; Wu, C.; Ponder, J. W.; Ren, P. Polarizable Atomic Multipole-Based AMOEBA Force Field for Proteins. *J. Chem. Theory Comput.* **2013**, *9* (9), 4046–4063.

(17) 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.

(18) Xie, A.; Hoff, W. D.; Kroon, A. R.; Hellingwerf, K. J. Glu46 Donates a Proton to the 4-Hydroxycinnamate Anion Chromophore During the Photocycle of Photoactive Yellow Protein. *Biochemistry* **1996**, *35* (47), 14671–14678.

(19) Van Beeumen, J. J.; Devreese, B. V.; Van Bun, S. M.; Hoff, W. D.; Hellingwerf, K. J.; Meyer, T. E.; Cusanovich, M. A.; Mcree, D. E. Primary structure of a photoactive yellow protein from the phototrophic bacterium Ectothiorhodospira halophila, with evidence for the mass and the binding site of the chromophore. *Protein Science* **1993**, *2* (7), 1114–1125.

(20) Hoff, W. D.; Dux, P.; Hard, K.; Devreese, B.; Nugteren-Roodzant, I. M.; Crielaard, W.; Boelens, R.; Kaptein, R.; Van Beeumen, J.; Hellingwerf, K. J. Thiol Ester-Linked p-Coumaric Acid as a New Photoactive Prosthetic Group in a Protein with Rhodopsin-like Photochemistry. *Biochemistry* **1994**, *33* (47), 13959–13962.
(21) Sandberg, D. J.; Rudnitskaya, A. N.; Gascón, J. A. QM/MM Prediction of the Stark Shift in the Active Site of a Protein. *J. Chem. Theory Comput.* **2012**, *8* (8), 2817–2823.
(22) Vaissier Welborn, V. Understanding Cysteine Reactivity in Protein Environments with Electric Fields. *J. Phys. Chem. B* **2023**, *127* (46), 9936–9942.
(23) Lasham, J.; Djurabekova, A.; Zickermann, V.; Vonck, J.; Sharma, V. Role of Protonation States in the Stability of Molecular Dynamics Simulations of High-Resolution Membrane Protein Structures. *J. Phys. Chem. B* **2024**, *128* (10), 2304–2316.
(24) Huang, Z.; Wu, Z.; Lai, S.; Chen, X.; Liu, J. Conserved Protonation Pattern in the Extended Active Site of Human Carboxylesterase 1 and Its Impact on Enzyme Catalysis. *J. Chem. Inf. Model.* **2025**, *65* (16), 8794–8805.
(25) Le Questel, J.-Y.; Berthelot, M.; Laurence, C. Hydrogen-Bond Acceptor Properties of Nitriles: A Combined Crystallographic and Ab Initio Theoretical Investigation. *Journal of Physical Organic Chemistry* **2000**, *13* (6), 347–358.
(26) Fafarman, A. T.; Boxer, S. G. Nitrile Bonds as Infrared Probes of Electrostatics in Ribonuclease S. *J. Phys. Chem. B* **2010**, *114* (42), 13536–13544.
(27) Fafarman, A. T.; Webb, L. J.; Chuang, J. I.; Boxer, S. G. Site-Specific Conversion of Cysteine Thiols into Thiocyanate Creates an IR Probe for Electric Fields in Proteins. *J. Am. Chem. Soc.* **2006**, *128* (41), 13356–13357.
(28) Fafarman, A. T.; Sigala, P. A.; Schwans, J. P.; Fenn, T. D.; Herschlag, D.; Boxer, S. G. Quantitative, Directional Measurement of Electric Field Heterogeneity in the Active Site of Ketosteroid Isomerase. *Proceedings of the National Academy of Sciences* **2012**, *109* (6), E299–E308.
(29) Streu, K.; Hunsberger, S.; Patel, J.; Wan, X.; Daly, C. A., Jr. Development of a Universal Method for Vibrational Analysis of the Terminal Alkyne C≡C Stretch. *The Journal of Chemical Physics* **2024**, *160* (7), 074106.
(30) Andersson, M. P.; Uvdal, P. New Scale Factors for Harmonic Vibrational Frequencies Using the B3LYP Density Functional Method with the Triple-ζ Basis Set 6-311+G(d,p). *J. Phys. Chem. A* **2005**, *109* (12), 2937–2941.
(31) *Ward, W. Properties of the Coelenterate Green-Fluorescent Proteins. In Bioluminescence and Chemiluminescence; 1981; Pp 235–242.*
(32) Lin, C.-Y.; Both, J.; Do, K.; Boxer, S. G. Mechanism and Bottlenecks in Strand Photodissociation of Split Green Fluorescent Proteins (GFPs). *Proceedings of the National Academy of Sciences* **2017**, *114* (11), E2146–E2155.
(33) Devanathan, S.; Genick, U. K.; Canestrelli, I. L.; Meyer, T. E.; Cusanovich, M. A.; Getzoff, E. D.; Tollin, G. New Insights into the Photocycle of Ectothiorhodospira Halophila Photoactive Yellow Protein: Photorecovery of the Long-Lived Photobleached Intermediate in the Met100Ala Mutant. *Biochemistry* **1998**, *37* (33), 11563–11568.
(34) Sasaki, J.; Kumauchi, M.; Hamada, N.; Oka, T.; Tokunaga, F. Light-Induced Unfolding of Photoactive Yellow Protein Mutant M100L. *Biochemistry* **2002**, *41* (6), 1915–1922.
(35) Kumauchi, M.; Hamada, N.; Sasaki, J.; Tokunaga, F. A Role of Methionine100 in Facilitating PYPM-Decay Process in the Photocycle of Photoactive Yellow Protein1. *The Journal of Biochemistry* **2002**, *132* (2), 205–210.
(36) Kumar, A.; Burns, D. C.; Al-Abdul-Wahid, M. S.; Woolley, G. A. A Circularly Permuted Photoactive Yellow Protein as a Scaffold for Photoswitch Design. *Biochemistry* **2013**, *52* (19), 3320–3331.
(37) Romei, M. G.; Lin, C.-Y.; Boxer, S. G. Structural and Spectroscopic Characterization of Photoactive Yellow Protein and Photoswitchable Fluorescent Protein Constructs Containing Heavy Atoms. *Journal of Photochemistry and Photobiology A: Chemistry* **2020**, *401*, 112738.
(38) McPhillips, T. M.; McPhillips, S. E.; Chiu, H.-J.; Cohen, A. E.; Deacon, A. M.; Ellis, P. J.; Garman, E.; Gonzalez, A.; Sauter, N. K.; Phizackerley, R. P.; Soltis, S. M.; Kuhn, P. Blu-Ice and the Distributed Control System: Software for Data Acquisition and Instrument Control at Macromolecular Crystallography Beamlines. *J Synchrotron Rad* **2002**, *9* (6), 401–406.
(39) Kabsch, W. XDS. *Acta Cryst D* **2010**, *66* (2), 125–132.

(40) Kabsch, W. Integration, Scaling, Space-Group Assignment and Post-Refinement. *Acta Cryst D* **2010**, *66* (2), 133–144.
(41) *A Quick XDS Tutorial for SSRL*. https://smb.slac.stanford.edu/facilities/software/xds/ (accessed 2023-03-03).
(42) Adams, P. D.; Afonine, P. V.; Bunkóczi, G.; Chen, V. B.; Davis, I. W.; Echols, N.; Headd, J. J.; Hung, L.-W.; Kapral, G. J.; Grosse-Kunstleve, R. W.; McCoy, A. J.; Moriarty, N. W.; Oeffner, R.; Read, R. J.; Richardson, D. C.; Richardson, J. S.; Terwilliger, T. C.; Zwart, P. H. PHENIX: A Comprehensive Python-Based System for Macromolecular Structure Solution. *Acta Cryst D* **2010**, *66* (2), 213–221.
(43) Emsley, P.; Cowtan, K. Coot: Model-Building Tools for Molecular Graphics. *Acta Cryst D* **2004**, *60* (12), 2126–2132.
(44) Joosten, R. P.; Joosten, K.; Cohen, S. X.; Vriend, G.; Perrakis, A. Automatic Rebuilding and Optimization of Crystallographic Structures in the Protein Data Bank. *Bioinformatics* **2011**, *27* (24), 3392–3398.
(45) Rajagopal, S.; Moffat, K. Crystal Structure of a Photoactive Yellow Protein from a Sensor Histidine Kinase: Conformational Variability and Signal Transduction. *Proceedings of the National Academy of Sciences* **2003**, *100* (4), 1649–1654.
(46) Kyndt, J. A.; Savvides, S. N.; Memmi, S.; Koh, M.; Fitch, J. C.; Meyer, T. E.; Heyn, M. P.; Van Beeumen, J. J.; Cusanovich, M. A. Structural Role of Tyrosine 98 in Photoactive Yellow Protein: Effects on Fluorescence, Gateway, and Photocycle Recovery,. *Biochemistry* **2007**, *46* (1), 95–105.
(47) Kyndt, J. A.; Meyer, T. E.; Cusanovich, M. A. Photoactive Yellow Protein, Bacteriophytochrome, and Sensory Rhodopsin in Purple Phototrophic Bacteria. *Photochem Photobiol Sci* **2004**, *3* (6), 519–530.
(48) Lin, J.; Pozharski, E.; Wilson, M. A. Short Carboxylic Acid–Carboxylate Hydrogen Bonds Can Have Fully Localized Protons. *Biochemistry* **2017**, *56* (2), 391–402.
(49) Engh, R. A.; Huber, R. Accurate Bond and Angle Parameters for X-Ray Protein Structure Refinement. *Acta Cryst A* **1991**, *47* (4), 392–400.
(50) Coates, L.; Erskine, P. T.; Crump, M. P.; Wood, S. P.; Cooper, J. B. Five Atomic Resolution Structures of Endothiapepsin Inhibitor Complexes: Implications for the Aspartic Proteinase Mechanism. *Journal of Molecular Biology* **2002**, *318* (5), 1405–1415.
(51) Coates, L.; Tuan, H.-F.; Tomanicek, S.; Kovalevsky, A.; Mustyakimov, M.; Erskine, P.; Cooper, J. The Catalytic Mechanism of an Aspartic Proteinase Explored with Neutron and X-Ray Diffraction. *J. Am. Chem. Soc.* **2008**, *130* (23), 7235–7237.
(52) Witt, A. C.; Lakshminarasimhan, M.; Remington, B. C.; Hasim, S.; Pozharski, E.; Wilson, M. A. Cysteine pKa Depression by a Protonated Glutamic Acid in Human DJ-1. *Biochemistry* **2008**, *47* (28), 7430–7440.
(53) Søndergaard, C. R.; Olsson, M. H. M.; Rostkowski, M.; Jensen, J. H. Improved Treatment of Ligands and Coupling Effects in Empirical Calculation and Rationalization of pKa Values. *J. Chem. Theory Comput.* **2011**, *7* (7), 2284–2295.
(54) Olsson, M. H. M.; Søndergaard, C. R.; Rostkowski, M.; Jensen, J. H. PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical pKa Predictions. *J. Chem. Theory Comput.* **2011**, *7* (2), 525–537.
(55) Jurrus, E.; Engel, D.; Star, K.; Monson, K.; Brandi, J.; Felberg, L. E.; Brookes, D. H.; Wilson, L.; Chen, J.; Liles, K.; Chun, M.; Li, P.; Gohara, D. W.; Dolinsky, T.; Konecny, R.; Koes, D. R.; Nielsen, J. E.; Head-Gordon, T.; Geng, W.; Krasny, R.; Wei, G.-W.; Holst, M. J.; McCammon, J. A.; Baker, N. A. Improvements to the APBS Biomolecular Solvation Software Suite. *Protein Science* **2018**, *27* (1), 112–128.
(56) Fisher, S. J.; Blakeley, M. P.; Cianci, M.; McSweeney, S.; Helliwell, J. R. Protonation-State Determination in Proteins Using High-Resolution X-Ray Crystallography: Effects of Resolution and Completeness. *Acta Crystallographica Section D* **2012**, *68* (7), 800–809.
(57) Ahmed, H. U.; Blakeley, M. P.; Cianci, M.; Cruickshank, D. W. J.; Hubbard, J. A.; Helliwell, J. R. The Determination of Protonation States in Proteins. *Acta Cryst D* **2007**, *63* (8), 906–922.
(58) Borgstahl, G. E. O.; Williams, D. R.; Getzoff, E. D. 1.4 .ANG. Structure of Photoactive Yellow Protein, a Cytosolic Photoreceptor: Unusual Fold, Active Site, and Chromophore. *Biochemistry* **1995**, *34* (19), 6278–6287.

(59) Wang, J. Visualization of H Atoms in the X-Ray Crystal Structure of Photoactive Yellow Protein: Does It Contain Low-Barrier Hydrogen Bonds? *Protein Science* **2019**, *28* (11), 1966–1972.
(60) Oktaviani, N. A.; Pool, T. J.; Yoshimura, Y.; Kamikubo, H.; Scheek, R. M.; Kataoka, M.; Mulder, F. A. A. Active-Site pKa Determination for Photoactive Yellow Protein Rationalizes Slow Ground-State Recovery. *Biophysical Journal* **2017**, *112* (10), 2109–2116.
(61) Creighton, T. E. *Proteins: Structures and Molecular Properties*; Macmillan, 1993.
(62) Sun, Z.; Liu, Q.; Qu, G.; Feng, Y.; Reetz, M. T. Utility of B-Factors in Protein Science: Interpreting Rigidity, Flexibility, and Internal Motion and Engineering Thermostability. *Chem. Rev.* **2019**, *119* (3), 1626–1665.
(63) Broom, A.; Rakotoharisoa, R. V.; Thompson, M. C.; Zarifi, N.; Nguyen, E.; Mukhametzhanov, N.; Liu, L.; Fraser, J. S.; Chica, R. A. Ensemble-Based Enzyme Design Can Recapitulate the Effects of Laboratory Directed Evolution in Silico. *Nat Commun* **2020**, *11* (1), 4808.
(64) Getzoff, E. D.; Gutwin, K. N.; Genick, U. K. Anticipatory Active-Site Motions and Chromophore Distortion Prime Photoreceptor PYP for Light Activation. *Nat Struct Mol Biol* **2003**, *10* (8), 663–668.
(65) 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* (12), 1920–1927.
(66) Heinen, J.; Dubbeldam, D. On Flexible Force Fields for Metal–Organic Frameworks: Recent Developments and Future Prospects. *WIREs Computational Molecular Science* **2018**, *8* (4), e1363.
(67) Schmidt, J. R.; Corcelli, S. A. Infrared Absorption Line Shapes in the Classical Limit: A Comparison of the Classical Dipole and Fluctuating Frequency Approximations. *The Journal of Chemical Physics* **2008**, *128* (18), 184504.
(68) Layfield, J. P.; Hammes-Schiffer, S. Calculation of Vibrational Shifts of Nitrile Probes in the Active Site of Ketosteroid Isomerase upon Ligand Binding. *J. Am. Chem. Soc.* **2013**, *135* (2), 717–725.
(69) Baker, C. M. Polarizable Force Fields for Molecular Dynamics Simulations of Biomolecules. *WIREs Computational Molecular Science* **2015**, *5* (2), 241–254.
(70) Rackers, J. A.; Silva, R. R.; Wang, Z.; Ponder, J. W. Polarizable Water Potential Derived from a Model Electron Density. *J. Chem. Theory Comput.* **2021**, *17* (11), 7056–7084.
(71) Guan, X.; Leven, I.; Heidar-Zadeh, F.; Head-Gordon, T. Protein C-GeM: A Coarse-Grained Electron Model for Fast and Accurate Protein Electrostatics Prediction. *J. Chem. Inf. Model.* **2021**, *61* (9), 4357–4369.
(72) Witek, J.; Heindel, J. P.; Guan, X.; Leven, I.; Hao, H.; Naullage, P.; LaCour, A.; Sami, S.; Menger, M. F. S. J.; Cofer-Shabica, D. V.; Berquist, E.; Faraji, S.; Epifanovsky, E.; Head-Gordon, T. M-Chem: A Modular Software Package for Molecular Simulation That Spans Scientific Domains. *Molecular Physics* **2023**, *121* (9–10), e2129500.
(73) Harger, M.; Li, D.; Wang, Z.; Dalby, K.; Lagardère, L.; Piquemal, J.-P.; Ponder, J.; Ren, P. Tinker-OpenMM: Absolute and Relative Alchemical Free Energies Using AMOEBA on GPUs. *Journal of Computational Chemistry* **2017**, *38* (23), 2047–2055.
(74) Adjoua, O.; Lagardère, L.; Jolly, L.-H.; Durocher, A.; Very, T.; Dupays, I.; Wang, Z.; Inizan, T. J.; Célerse, F.; Ren, P.; Ponder, J. W.; Piquemal, J.-P. Tinker-HP: Accelerating Molecular Dynamics Simulations of Large Complex Systems with Advanced Point Dipole Polarizable Force Fields Using GPUs and Multi-GPU Systems. *J. Chem. Theory Comput.* **2021**, *17* (4), 2034–2053.
(75) Wang, X.; Liu, H.; Li, Y.; Li, J.; Li, W.-L. TinkerModeller: An Efficient Tool for Building Biological Systems in Tinker Simulations. *J. Chem. Theory Comput.* **2025**, *21* (5), 2712–2722.
(76) Peng, X.; Zhang, Y.; Chu, H.; Li, Y.; Zhang, D.; Cao, L.; Li, G. Accurate Evaluation of Ion Conductivity of the Gramicidin A Channel Using a Polarizable Force Field without Any Corrections. *J. Chem. Theory Comput.* **2016**, *12* (6), 2973–2982.
(77) El Khoury, L.; Célerse, F.; Lagardère, L.; Jolly, L.-H.; Derat, E.; Hobaika, Z.; Maroun, R. G.; Ren, P.; Bouaziz, S.; Gresh, N.; Piquemal, J.-P. Reconciling NMR Structures of the HIV-1 Nucleocapsid Protein NCp7 Using Extensive Polarizable Force Field Free-Energy Simulations. *J. Chem. Theory Comput.* **2020**, *16* (4), 2013–2020.

(78) Jing, Z.; Ren, P. Molecular Dynamics Simulations of Protein RNA Complexes by Using an Advanced Electrostatic Model. *J. Phys. Chem. B* **2022**, *126* (38), 7343–7353.
(79) Baltrukevich, H.; Bartos, P. RNA-Protein Complexes and Force Field Polarizability. *Front Chem* **2023**, *11*, 1217506.
(80) Xu, Z.; Schahl, A.; Jolivet, M.-D.; Legrand, A.; Grélard, A.; Berbon, M.; Morvan, E.; Lagardere, L.; Piquemal, J.-P.; Loquet, A.; Germain, V.; Chavent, M.; Mongrand, S.; Habenstein, B. Dynamic Pre-Structuration of Lipid Nanodomain-Segregating Remorin Proteins. *Commun Biol* **2024**, *7* (1), 1–17.
(81) Lagardère, L.; Jolly, L.-H.; Lipparini, F.; Aviat, F.; Stamm, B.; Jing, Z. F.; Harger, M.; Torabifard, H.; Cisneros, G. A.; Schnieders, M. J.; Gresh, N.; Maday, Y.; Ren, P. Y.; Ponder, J. W.; Piquemal, J.-P. Tinker-HP: A Massively Parallel Molecular Dynamics Package for Multiscale Simulations of Large Complex Systems with Advanced Point Dipole Polarizable Force Fields. *Chem. Sci.* **2018**, *9* (4), 956–972.
(82) Delobelle, Q.; Inizan, T. J.; Adjoua, O.; Lagardère, L.; Célerse, F.; Maréchal, V.; Piquemal, J.-P. High-Resolution Molecular-Dynamics Simulations of the Pyruvate Kinase Muscle Isoform 1 and 2 (PKM1/2). *Chemistry – A European Journal* **2025**, *31* (20), e202402534.
(83) Schahl, A.; Lagardère, L.; Walker, B.; Ren, P.; Wioland, H.; Ballet, M.; Jégou, A.; Chavent, M.; Piquemal, J.-P. Histidine 73 Methylation Coordinates β-Actin Plasticity in Response to Key Environmental Factors. *Nat Commun* **2025**, *16* (1), 2304.
(84) 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.
(85) Ji, Z.; Kozuch, J.; Mathews, I. I.; Diercks, C. S.; Shamsudin, Y.; Schulz, M. A.; Boxer, S. G. Protein Electric Fields Enable Faster and Longer-Lasting Covalent Inhibition of β-Lactamases. *J. Am. Chem. Soc.* **2022**, *144* (45), 20947–20954.
(86) Shi, Y.; Laury, M. L.; Wang, Z.; Ponder, J. W. AMOEBA Binding Free Energies for the SAMPL7 TrimerTrip Host–Guest Challenge. *J Comput Aided Mol Des* **2021**, *35* (1), 79–93.
(87) Qu, X.; Dong, L.; Si, Y.; Zhao, Y.; Wang, Q.; Su, P.; Wang, B. Reliable Prediction of the Protein–Ligand Binding Affinity Using a Charge Penetration Corrected AMOEBA Force Field: A Case Study of Drug Resistance Mutations in Abl Kinase. *J. Chem. Theory Comput.* **2022**, *18* (3), 1692–1700.
(88) Chung, M. K. J.; Miller, R. J.; Novak, B.; Wang, Z.; Ponder, J. W. Accurate Host–Guest Binding Free Energies Using the AMOEBA Polarizable Force Field. *J. Chem. Inf. Model.* **2023**, *63* (9), 2769–2782.
(89) Wang, A.; Peng, X.; Li, Y.; Zhang, D.; Zhang, Z.; Li, G. Quality of Force Fields and Sampling Methods in Simulating pepX Peptides: A Case Study for Intrinsically Disordered Proteins. *Phys. Chem. Chem. Phys.* **2021**, *23* (3), 2430–2437.
(90) Marcellini, M.; Nguyen, M.-H.; Martin, M.; Hologne, M.; Walker, O. Accurate Prediction of Protein NMR Spin Relaxation by Means of Polarizable Force Fields. Application to Strongly Anisotropic Rotational Diffusion. *J. Phys. Chem. B* **2020**, *124* (25), 5103–5112.
(91) Lin, Y.-C.; Ren, P.; Webb, L. J. AMOEBA Force Field Trajectories Improve Predictions of Accurate pKa Values of the GFP Fluorophore: The Importance of Polarizability and Water Interactions. *J. Phys. Chem. B* **2022**, *126* (40), 7806–7817.
(92) Vaissier Welborn, V. Understanding Cysteine Reactivity in Protein Environments with Electric Fields. *J. Phys. Chem. B* **2023**, *127* (46), 9936–9942.
(93) Song, Z.; Wang, R.; Jiao, X.; Huang, Z. Graph-Based Deep Learning Models for Predicting pKa Values of Protein-Ionizable Residues via Physically Inspired Feature Engineering. *J. Chem. Inf. Model.* **2026**, *66* (3), 1742–1756.
(94) Weng, J.; Liao, C.; Li, Y.; Zhang, D.; Li, G.; Wang, A. Effect of Protein Dimerization on Ion Conductivity of Gramicidin A Channel Studied Using Polarizable Force Field. *cjcp* **2021**, *34* (4), 471–479.
(95) Schmidt, J. R.; Corcelli, S. A. Infrared Absorption Line Shapes in the Classical Limit: A Comparison of the Classical Dipole and Fluctuating Frequency Approximations. *The Journal of Chemical Physics* **2008**, *128* (18), 184504.

(96) Kwac, K.; Yang, N.; Ryan, M. J.; Zanni, M. T.; Cho, M. Molecular Dynamics Simulation Study of Water Structure and Dynamics on the Gold Electrode Surface with Adsorbed 4-Mercaptobenzonitrile. *J. Chem. Phys.* **2024**, *160* (6), 064701.
(97) Aneesh, A.; Streu, K.; Daly, C. A. Development of a Spectroscopic Map to Explain the Broad Raman Peak for Alkynes Solvated in Triethylamine. *J. Phys. Chem. B* **2025**, *129* (33), 8509–8520.
(98) Xu, C.; Fitzgerald, J. E.; Lyman, E.; Baiz, C. R. Sphingomyelin Slows Interfacial Hydrogen-Bonding Dynamics in Lipid Membranes. *Biophysical Journal* **2025**, *124* (7), 1158–1165.