

A Solvatochromic Model Calibrates Nitriles' Vibrational Frequencies to Electrostatic Fields

Sayan Bagchi,[‡] Stephen D. Fried,[‡] and Steven G. Boxer*

Department of Chemistry, Stanford University, Stanford, California 94305-5080, United States

S Supporting Information

ABSTRACT: Electrostatic interactions provide a primary connection between a protein's three-dimensional structure and its function. Infrared probes are useful because vibrational frequencies of certain chemical groups, such as nitriles, are linearly sensitive to local electrostatic field and can serve as a molecular electric field meter. IR spectroscopy has been used to study electrostatic changes or fluctuations in proteins, but measured peak frequencies have not been previously mapped to total electric fields, because of the absence of a field-frequency calibration and the complication of local chemical effects such as H-bonds. We report a solvatochromic model that provides a means to assess the H-bonding status of aromatic nitrile vibrational probes and calibrates their vibrational frequencies to electrostatic field. The analysis involves correlations between the nitrile's IR frequency and its ¹³C chemical shift, whose observation is facilitated by a robust method for introducing isotopes into aromatic nitriles. The method is tested on the model protein ribonuclease S (RNase S) containing a labeled p-CN-Phe near the active site. Comparison of the measurements in RNase S against solvatochromic data gives an estimate of the average total electrostatic field at this location. The value determined agrees quantitatively with molecular dynamics simulations, suggesting broader potential for the use of IR probes in the study of protein electrostatics.

Recent advances toward characterizing the complex molecular environment present inside proteins and biological macromolecules have focused on the use of infrared probes.^{1,2} One of the important strengths of IR-based methodologies is the linear dependence of vibrational frequencies on an electrostatic field, a phenomenon known as the vibrational Stark effect.^{3,4} A number of studies have employed extrinsic IR probes as a molecular tool to measure electrostatic field changes inside proteins.^{4–7} In particular, the nitrile stretching mode has been proposed as an ideal IR probe for investigations of protein structure^{4–10} and dynamics,¹¹ because it is a local mode in an uncluttered region of the IR spectrum that is particularly sensitive to its local electrostatic field.^{3,5,7}

While the vibrational Stark effect allows one to quantify the relationship between changes in electrostatic field and an accompanying IR frequency shift, e.g., in response to mutation or pH change, it does not provide a calibration to associate a particular total electrostatic field with any observed frequency.

Moreover, a nitrile's frequency, $\bar{\nu}_{\text{CN}}$, is impacted both by electrostatic fields and by hydrogen bonding.^{12,13} When a nitrile probe is H-bonded, it is particularly difficult to determine from IR measurements what electric field it is experiencing. Upon accepting an H-bond, $\bar{\nu}_{\text{CN}}$ shifts to the blue and broadens.^{12–16} The shift has been shown theoretically to depend on the distance and the angle between the nitrile and the proton on the H-bond donor,¹⁴ in a manner that is not described by the vibrational Stark effect.

Recently, we reported a technique¹⁷ to decompose IR shifts of nitriles into H-bonding and electrostatic components based on an analysis that requires both $\bar{\nu}_{\text{CN}}$ and the ¹³C chemical shift, $\delta^{13}\text{CN}$, of the nitrile carbon, effectively using these two observations to specify two unknowns. We initially applied this approach to the cysteine thiocyanate probe, Cys-SCN, which can be prepared by cyanylating cysteine with KCN.¹⁸ The synthetic procedure readily allowed for incorporation of isotopically labeled ¹³CN, which is necessary for measuring $\delta^{13}\text{CN}$ in the protein. Knowledge of both $\bar{\nu}_{\text{CN}}$ and $\delta^{13}\text{CN}$ for the nitrile probe *in situ* was used to determine whether a series of Cys-SCN probes installed at the active site of the enzyme ketosteroid isomerase were H-bonded.^{7,17} Aromatic nitriles can also be introduced into proteins, either as p-CN-Phe^{6,8,10} or as a common functional group on inhibitors, including many drugs, that bind to the active site of enzymes and signaling proteins.^{4,5,11} For these reasons, we sought to generalize the methods and analysis previously described for cysteine thiocyanates^{17,18} to aromatic nitriles, beginning with a simple synthetic route for incorporating ¹³CN into aromatic nitriles, in particular p-¹³CN-Phe, and then testing whether similar behavior is observed that would facilitate distinguishing H-bonded from non-H-bonded aromatic nitriles.

Our analysis is based on solvatochromic trends of benzonitrile (PhCN), a model compound used to represent p-CN-Phe. As seen in Figure 1A, when PhCN is dissolved in various non-H-bonding solvents, $\bar{\nu}_{\text{CN}}$ varies from 2233.4 cm^{–1} in the most nonpolar solvent (hexane) to 2227.6 cm^{–1} in the most polar solvent (DMSO). This trend can largely be attributed to the effective electric field created by the polarization of the solvent. Using a simple analytic model such as the Onsager reaction field approach,^{19,20} the magnitude of the electric fields due to each solvent can be approximated (the x-axis of Figure 1A).²¹ The Onsager equation estimates the total electric field due to dielectric polarization, so it is unable to describe chemical interactions such as H-bonds; however, it

Received: April 23, 2012

Published: June 13, 2012

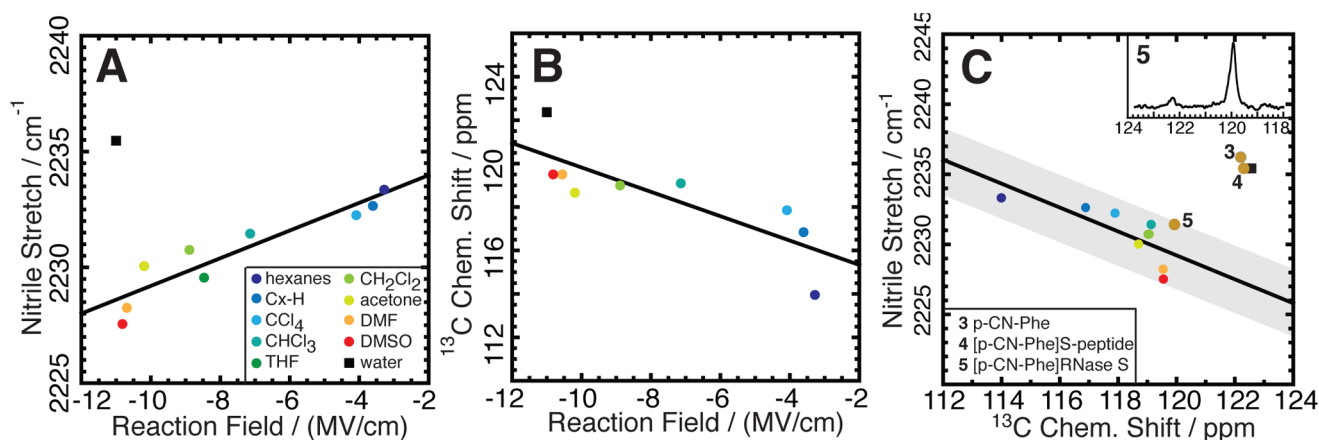


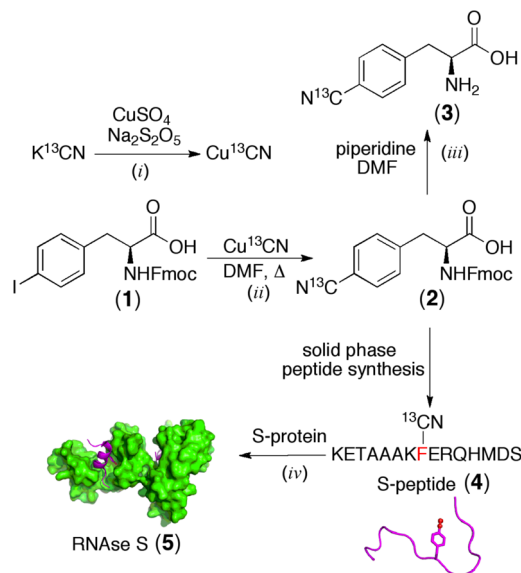
Figure 1. Solvatochromism calibrates the sensitivity of the C–N stretch and the ^{13}C chemical shift of benzonitrile to electric field and H-bonding. (A) $\bar{\nu}_{\text{CN}}$ of PhCN in various non-H-bonding solvents (colored circles) and in water (black square) compared against the electric field each solvent exercises on PhCN (see footnote 20). $\bar{\nu}_{\text{CN}} = 2235.2 + 0.597F$; $R^2 = 0.86$ (excluding water). (B) δ^{13}_{CN} of PhCN in various non-H-bonding solvents (colored circles) and in water (black square) compared similarly. $\delta^{13}_{\text{CN}} = 114.12 - 0.561F$; $R^2 = 0.66$. (C) Comparison between $\bar{\nu}_{\text{CN}}$ and δ^{13}_{CN} . $\bar{\nu}_{\text{CN}} = 2331.7 - 0.854\delta^{13}_{\text{CN}}$; $R^2 = 0.60$. All non-H-bonding solvents follow an approximate linear trend, but water is significantly removed from that region. H-bonding is responsible for blue-shifting PhCN's frequency in water away from what would be expected based on electrostatics. Three additional points (brown circles) are compared against the solvatochromic model: amino acid p-CN-Phe (3), [p-CN-Phe]S-peptide (4), and [p-CN-Phe]RNase S (5), all in 20 mM HEPES, pH 8.0. 3 and 4 fall far away from the electrostatic line, consistent with the nitrile being H-bonded, while 5 falls near the electrostatic line, suggesting an absence of H-bonding. Inset shows ^{13}C NMR spectrum of 5.

qualitatively describes the electric field caused by non-H-bond-donating solvents. Electric fields so calculated display a good correlation to observed frequencies ($R^2 = 0.86$); moreover, the slope of the best-fitting line ($0.60 \text{ cm}^{-1}/(\text{MV}/\text{cm})$) shows excellent agreement with the independently measured Stark tuning rate of PhCN ($0.61 \text{ cm}^{-1}/(\text{MV}/\text{cm})$),^{9,22} which is the frequency shift per unit field found when an *external* electric field is applied to the compound. This agreement reinforces the view that the solvatochromic trend is reporting largely on electric fields. On the other hand, the frequency of PhCN in water (the black square in Figure 1A) does not follow the trend, consistent with an additional non-electrostatic effect associated with the nitrile being H-bonded.^{12,14,17} A parallel study on δ^{13}_{CN} of PhCN in the same solvents shows a similar correlation (albeit weaker) with the calculated electric field (Figure 1B). In this case, unlike $\bar{\nu}_{\text{CN}}$ in water, δ^{13}_{CN} in water is relatively similar to its value in another very polar solvent (DMSO). In the context of the Onsager model, which ascribes water and DMSO very similar reaction fields, this observation suggests that δ^{13}_{CN} is sensitive to electrostatic field but not as sensitive to H-bonding as $\bar{\nu}_{\text{CN}}$.¹⁷

As described in earlier work on thiocyanate probes, a plot of $\bar{\nu}_{\text{CN}}$ versus δ^{13}_{CN} (Figure 1C) recapitulates the linear sensitivity of both observables to electrostatic field but is independent of any specific model to calculate it.¹⁷ The best-fit line in Figure 1C, drawn for the non-H-bonding solvents' points, describes the covariance of $\bar{\nu}_{\text{CN}}$ and δ^{13}_{CN} in environments free of H-bonds. Therefore, deviations from this line represent effects of specific interactions, which alter $\bar{\nu}_{\text{CN}}$ and δ^{13}_{CN} differently and confound the mutual linear dependence of $\bar{\nu}_{\text{CN}}$ and δ^{13}_{CN} on field. Points that occur in the region near the line given by electrostatics (shown in gray in Figure 1C) correspond to environments free of H-bonds, while points that fall significantly off this line appear to experience a non-electrostatic contribution, assigned to H-bond effects. Noting this observation, IR and NMR data can be used to assess whether a nitrile is H-bonded in complicated environments where it is not known *a priori*.

To selectively label the carbon in the nitrile of p- ^{13}C -Phe, we adopted the Rosenmund–von Braun reaction,²³ which exchanges the iodine in aromatic iodides for a nitrile using a cuprous reagent. As shown in Scheme 1 (synthetic methods are

Scheme 1. Incorporation of Isotopically Labeled Nitriles into Amino Acids, Peptides, and Proteins^a



^aConditions: (i) 1 h, 46 °C; (ii) 1 h, 150 °C, microwave; (iii) 2 h, RT; (iv) RT, 20 mM HEPES, pH 8.0.

given in the Supporting Information), Cu^{13}CN was prepared from commercial K^{13}CN with aqueous chemistry²⁴ (i). N-Fmoc-p- ^{13}C -Phe (2) was prepared in moderate (40%) yield using a microwave-mediated reaction (ii) between the corresponding N-Fmoc-p-I-Phe (1) and Cu^{13}CN . With 2 in hand, Fmoc can be easily removed to give the isotopically labeled amino acid (3).²⁵ This same approach has been used to

label other aromatic nitriles²⁶ and in more complex nitrile-bearing enzyme inhibitors (unpublished results).

When ribonuclease A is digested by subtilisin, a 20-residue peptide called the S-peptide is liberated, forming a truncated ribonuclease called the S-protein. Subsequent combination of S-peptide with S-protein reconstitutes a functional semi-synthetic protein called RNase S.⁶ We chose to use this system as a test case because the S-peptide can be exploited to deploy an IR/NMR probe into RNase S by replacing Phe8 with p-CN-Phe. This nitrile-bearing RNase S has been characterized, and its crystal structure has been solved (PDB: 3OQY).⁶ We used **2** as a reagent to generate an S-peptide bearing an isotopically labeled nitrile ([p-¹³CN-Phe]S-peptide, **4**), which in turn can be used to form [p-¹³CN-Phe]RNase S (**5**) as shown in Scheme 1. **5** has been shown to have nearly the same catalytic properties as the native ribonuclease.⁶ The nitrile in **5** is buried in a hydrophobic region near the active site, with no access to the solvent and no H-bonding partners within a reasonable distance.

¹³C NMR experiments were carried out on the labeled amino acid (**3**), the labeled peptide fragment (**4**), and the labeled split protein (**5**). With quantitative isotopic enrichment of a single atom, high-quality NMR spectra (see inset of Figure 1C) could be obtained on protein samples with fewer than 100 scans. These measurements, in combination with $\bar{\nu}_{\text{CN}}$'s of these three species, allow us to apply the analysis described in Figure 1. The three (δ^{13}_{CN} , $\bar{\nu}_{\text{CN}}$) ordered pairs for the free amino acid (**3**), S-peptide (**4**), and RNase S (**5**) can be compared against the solvatochromic model (brown circles) and are shown alongside it in Figure 1C. It is apparent that the points for **3** and **4** lie well off the line described by a purely electrostatic model, and in fact are very close to the point representing PhCN in water (the black square). These data suggest that the nitriles of p-CN-Phe in the free amino acid and in the S-peptide are H-bonded (most likely to water), and that their local electrostatic environments are largely determined by the surrounding water molecules.

When the nitrile is embedded into a protein environment, by incorporating the S-peptide into RNase S (**5**), (δ^{13}_{CN} , $\bar{\nu}_{\text{CN}}$) moves into a region very close to the line described by the electrostatic model, within the range found for other non-H-bonding solvents. This result indicates that the nitrile in RNase S is *not* H-bonded, consistent with expectations about the H-bonding status of this nitrile from the crystal structure⁶ and from 2D IR experiments.²⁷ This conclusion could not have been reached if one only knew $\bar{\nu}_{\text{CN}}$ in **4** and **5**, because the red-shift upon complexation of the S-peptide to the truncated protein is only 4.0 cm⁻¹, which is too small to assign to the breaking of an H-bond. Figure 1C reveals that the modest red-shift occurs because there are two partially counteracting contributions: when the nitrile is embedded into a protein, it loses an H-bond but also gets placed in a comparatively weaker electric field. This weaker electric field can be explained qualitatively by the fact that the local hydrophobic environment inside RNase S is less polar than the environment in aqueous solution. The superposition of these two effects results in a shift that is difficult to interpret from the IR measurements alone. However, in combination with NMR measurements, we can deconvolute the two effects. This discussion highlights how the dual IR/NMR technique can be used to dissect an IR frequency shift and resolve a puzzle such as understanding how a large alteration in local environment (associated with reconstituting

an unstructured peptide into a complete protein) can result in a fairly small shift.

We recently proposed a semiempirical strategy to determine total electrostatic fields in proteins:²⁶ the vibrational frequencies of a model compound in a series of solvents paired with their calculated electric fields (such as from Onsager's theory) might be used as a reference to assign a vibrational frequency recorded in a protein to an electrostatic field. This concept rests on the assumptions that (1) the model being used to calculate solvent fields is accurate and (2) the solvatochromic frequency shifts are principally due to electrostatics.

Evidence supporting these two assumptions for the non-H-bonding case was presented,²⁶ and the solvatochromic study on PhCN presented here (Figure 1) further supports the previous work. The NMR/IR data for RNase S (point **5** in Figure 1C) confirm that the nitrile probe is not H-bonded, indicating that the solvatochromic data might be a useful reference to assign a total electrostatic field to $\bar{\nu}_{\text{CN}}$ observed in the protein complex. Applying the field–frequency correlation from Figure 1A, the electric field corresponding to the nitrile stretching frequency in **5** is −7 MV/cm. Molecular dynamics (MD) simulations using the Amber-99 force field were carried out on the [p-CN-Phe]RNase S construct as previously described (details found in Supporting Information).^{6,27} The electric field at the midpoint of the nitrile bond was calculated every 2 fs steps during the 20 ns trajectory, and the average total electric field experienced by the nitrile was found to be −7.8 MV/cm (Figure 2), which agrees well with the value determined by the proposed semiempirical method. This agreement supports the use of a solvatochromic scheme to calibrate a mapping between vibrational frequency and electrostatic field; it also provides an uncommonly convincing testament to the accuracy of the electrostatic parameters in a standard MD force field.

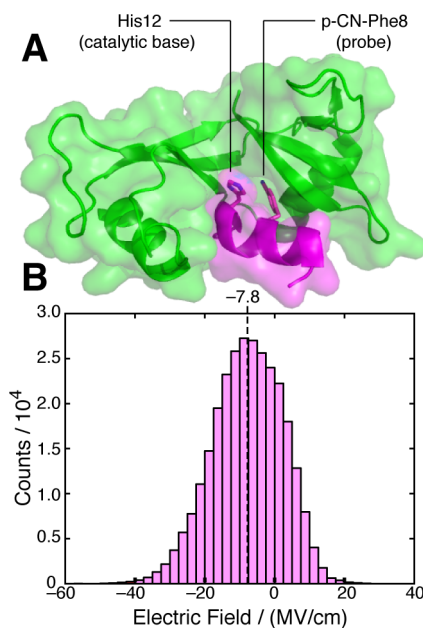


Figure 2. (A) Structure of [p-CN-Phe]RNase S (**5**). The probe, located at position 8 on the S-peptide, is shown in sticks. It is buried in a hydrophobic pocket and is close to the catalytic proton shuttle His12, shown in sticks. (B) Histogram of the electric fields experienced by the nitrile in **5** during a 20 ns MD trajectory. The mean electric field is −7.8 MV/cm; the standard deviation is 10.0 MV/cm.

To the best of our knowledge, this work constitutes the first instance of a meaningful comparison between computation and experiment of a single-state total electric field. Previous studies have focused on electric field *differences* between two states^{2,4–7,28,29} or electric field fluctuations,^{27,30} which do not require an external reference and depend only on the Stark tuning rate. The total electric field that a protein exercises on a target biomacromolecule, ligand, or transition state defines the energetics that underlie molecular recognition, binding, and catalysis, respectively. We therefore expect that translating IR spectroscopic data into semiempirical absolute electric field maps in proteins will lead to a deeper physical understanding of protein function. Two aspects of our system likely made this agreement between experiment and theory possible: (1) the nitrile probe is not H-bonded at any point during the MD trajectory, which would introduce non-electrostatic contributions to the vibrational frequency, and (2) the absence of slow dynamics, as evidenced by 2D IR studies on this construct,²⁷ allowed the relatively short simulation to adequately sample configurations of the protein present in the equilibrium ensemble.

In summary, we have reported the synthesis and application of a dual NMR/IR probe (p-¹³CN-Phe), which was able to determine that a nitrile probe installed in RNase S is not H-bonded. Using experimental solvatochromic data as a calibration tool, we then translated the nitrile stretching frequency to a total electrostatic field and found excellent agreement with simulation. Although this single example's agreement is encouraging, current work is underway to further benchmark this method and examine its range of validity.

■ ASSOCIATED CONTENT

■ Supporting Information

Synthetic, spectroscopic, and simulation methods. This material is available free of charge via the Internet at <http://pubs.acs.org>.

■ AUTHOR INFORMATION

Corresponding Author

sboxer@stanford.edu

Author Contributions

[‡]These authors contributed equally.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We thank the Trost laboratory for use of their microwave reactor. S.D.F. thanks the NSF for a predoctoral fellowship. This work is supported in part by a grant from the NIH (GM27738).

■ REFERENCES

- (1) Taskent-Sezgin, H.; Chung, J.; Banerjee, P. S.; Nagarajan, S.; Dyer, R. B.; Carrico, I.; Raleigh, D. P. *Angew. Chem., Int. Ed.* **2010**, *49*, 7473–7475.
- (2) Hu, W.; Webb, L. J. *J. Phys. Chem. Lett.* **2011**, *2*, 1925–1930.
- (3) Suydam, I. T.; Boxer, S. G. *Biochemistry* **2003**, *42*, 12050–12055.
- (4) Suydam, I. T.; Snow, C. D.; Pande, V. S.; Boxer, S. G. *Science* **2006**, *313*, 200–204.
- (5) Webb, L. J.; Boxer, S. G. *Biochemistry* **2008**, *47*, 1588–1598.
- (6) Fafarman, A. T.; Boxer, S. G. *J. Phys. Chem. B* **2010**, *114*, 13536–13544.

- (7) Fafarman, A. T.; Sigala, P. A.; Schwans, J. P.; Fenn, T. D.; Herschlag, D.; Boxer, S. G. *Proc. Natl. Acad. Sci. U.S.A.* **2012**, *109*, E299–E308.
- (8) Lindquist, B. A.; Furse, K. E.; Corcelli, S. A. *Phys. Chem. Chem. Phys.* **2009**, *11*, 8119–8132.
- (9) Dalosto, S. D.; Vanderkooi, J. M.; Sharp, K. A. *J. Phys. Chem. B* **2004**, *108*, 6450–6457. Dalosto et al. calculated from DFT that the linear dependence of $\bar{\nu}_{\text{CN}}$ on electrostatic field breaks down at very large fields, ca. -50 MV/cm. This effect would complicate the simple linear solvatochromic model presented herein; however, our measurements suggest that average total electrostatic fields in solvents and proteins are low enough to remain in a linear regime.
- (10) Getahun, Z.; Huang, C.-Y.; Wang, T.; De León, B.; DeGrado, W. F.; Gai, F. *J. Am. Chem. Soc.* **2002**, *125*, 405–411.
- (11) Fang, C.; Bauman, J. D.; Das, K.; Remorino, A.; Arnold, E.; Hochstrasser, R. M. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 1472–1477.
- (12) Aschaffenburg, D. J.; Moog, R. S. *J. Phys. Chem. B* **2009**, *113*, 12736–12743.
- (13) Maienschein-Cline, M. G.; Londergan, C. H. *J. Phys. Chem. A* **2007**, *111*, 10020–10025.
- (14) Choi, J.-H.; Oh, K.-I.; Lee, H.; Lee, C.; Cho, M. *J. Chem. Phys.* **2008**, *128*, 134506.
- (15) Choi, J.-H.; Cho, M. *J. Chem. Phys.* **2011**, *134*, 154513.
- (16) Lindquist, B. A.; Corcelli, S. A. *J. Phys. Chem. B* **2008**, *112*, 6301–6303.
- (17) Fafarman, A. T.; Sigala, P. A.; Herschlag, D.; Boxer, S. G. *J. Am. Chem. Soc.* **2010**, *132*, 12811–12813.
- (18) Fafarman, A. T.; Webb, L. J.; Chuang, J. I.; Boxer, S. G. *J. Am. Chem. Soc.* **2006**, *128*, 13356–13357.
- (19) Onsager, L. *J. Am. Chem. Soc.* **1936**, *58*, 1486–1493.
- (20) The Onsager field, F_{Onsager} is given by the expression $F_{\text{Onsager}} = (\mu_0/a^3)[2(\epsilon - 1)(n^2 + 2)/3(2\epsilon + n^2)]$. It is a function of the solvent's static dielectric constant, ϵ , the solute's gas-phase dipole moment, μ_0 , and the solute's refractive index, n . The term a is the Onsager cavity radius and is related to the molecular volume of the solute.
- (21) The dipole of PhCN was taken to be 4.48 D (Borst, D. R.; Korter, T. M.; Pratt, D. W. *Chem. Phys. Lett.* **2001**, *350*, 485–490). Its index of refraction is 1.528 (CRC Handbook), and its volume factor was taken to be 171 Å³, as determined from its formula weight 103 g/mol and density 1.0 g/mL. Static dielectric constants for all the solvents were taken from the CRC Handbook.
- (22) Andrews, S. S.; Boxer, S. G. *J. Phys. Chem. A* **2000**, *104*, 11853–11863.
- (23) Ellis, G. P.; Romney-Alexander, T. M. *Chem. Rev.* **1987**, *87*, 779–794.
- (24) Matloubi, H.; Shafiee, A.; Saemian, N.; Shirvani, G.; Dahan, F. J. *J. Labelled Compd. Radiopharm.* **2004**, *47*, 31–36.
- (25) Wang, L.; Zhang, Z. W.; Brock, A.; Schultz, P. G. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 56–61.
- (26) Levinson, N. M.; Fried, S. D.; Boxer, S. G. *J. Phys. Chem. B* **2012**, DOI: 10.1021/jp301054e.
- (27) Bagchi, S.; Boxer, S. G.; Fayer, M. D. *J. Phys. Chem. B* **2012**, *116*, 4034–4042.
- (28) Park, E. S.; Andrews, S. S.; Hu, R. B.; Boxer, S. G. *J. Phys. Chem. B* **1999**, *103*, 9813–9817.
- (29) Laberge, M.; Vanderkooi, J. M.; Sharp, K. A. *J. Phys. Chem.* **1996**, *100*, 10793–10801.
- (30) Merchant, K. A.; Noid, W. G.; Akiyama, R.; Finkelstein, I. J.; Goun, A.; McClain, B. L.; Loring, R. F.; Fayer, M. D. *J. Am. Chem. Soc.* **2003**, *125*, 13804–13818.

Supporting Information for

A Solvatochromic Model Calibrates Nitriles' Vibrational Frequencies to Electrostatic Fields

Sayan Bagchi¹, Stephen D. Fried¹, Steven G. Boxer*

Department of Chemistry, Stanford University, Stanford, CA 94305-5080

Methods

(1) Synthetic Methods.....	2-6
(2) Spectroscopic Methods.....	7-8
(3) Simulation Methods.....	9
(4) References and Notes.....	10

Synthetic Methods.

Copper(I) cyanide. We adapted the method from Matloubi *et al.*¹ illustrated in Scheme 1. To a 20 mL scintillation vial was added copper(II) sulfate (Baker & Adamson, 401 mg, 2.51 mmol, 1.25 equiv). Dissolution into 3 mL ddH₂O (Millipore) upon shaking produced a blue solution. To a separate 20 mL scintillation vial was added sodium metabisulfite (Alfa Aesar, 129 mg, 0.68 mmol, .338 equiv) and 1.5 mL ddH₂O. To a third 20 mL scintillation vial was dissolved potassium hydroxide (EMD, 57.7 mg, 1.03 mmol, .512 equiv) and potassium cyanide (Cambridge Isotopes, 2 mmol, 1 equiv) in 1.25 mL ddH₂O. The potassium cyanide used was either KCN or K¹³CN defining the isotopic composition of the cuprous product. The vial containing copper(II) sulfate was heated in an oil bath at 46°C, and stirred with a magnetic stir bar. The metabisulfite solution was added, and then the KOH/KCN (or K¹³CN) solution was slowly transferred dropwise to the heated vial over the course of ca. 5 min. The reaction appeared to occur instantly as evinced by formation of white precipitate. The suspension was allowed to stir for 1 h at 46°C. After 1 h, the vial was transferred to a different stir-plate with the heat off, and stirred for an additional 30 m at room temperature. The suspension was filtered over a fine frit, washed with water (2 × 5 mL), ethanol (2 × 5 mL), and diethyl ether (1 × 5 mL), and the white precipitate was collected (95% yield).

N-Fmoc-L-p-[¹³C]cyanophenylalanine. A conical heavy-walled microwave tube was charged with N-Fmoc-L-4-iodophenylalanine (Peptech, 70 mg, .136 mmol, 1

equiv), Cu¹³CN (25 mg, .273 mmol, 2 equiv), and 0.5 mL anhydrous DMF (Acros Organics). No stir bar was added; however, vigorous shaking of the microwave tube was employed to completely dissolve the reactants. The solution turned green during dissolution. The microwave tube was sealed with a crimped septum, and inserted into a microwave apparatus (Biotage Initiator). A program to apply heating at a particular temperature (140°C or 150°C) for a particular time (1 h or 1.5 h) was selected (see Table below) at high absorption level. Following irradiation, the amber-colored liquid was rinsed with ethyl acetate (15 mL) and water (15 mL) into a separatory funnel, and insoluble material adhered to the microwave tube was left behind. The aqueous phase was extracted with ethyl acetate (2 × 15 mL), and then the combined organic phase was washed with water (2 × 25 mL). The washed organic phase was dried over magnesium sulfate and reduced by rotary evaporation to furnish an orange oil. The oil was redissolved in 1.5 mL EtOH and loaded onto an HPLC (Shimadzu LC-20AT prominence) C18 column (A = 0.1% (vol/vol) TFA in ddH₂O; B = acetonitrile) running at 5 mL/min under a gradient where the fraction B linearly steps up from 30% to 100% at a rate of 1 %/min over 70 min. The chromatogram of the crude product contained three bands (32 – 36 min, 38 – 42 min, 47 – 51 min), and was compared to that of the starting material and of an authentic sample of product (Peptech, no isotopic enrichment) to assign the bands. The major fraction by UV absorption @ 254 nm (38 – 42 min) corresponding to product was collected and lyophilized overnight (VirTis Freezemobile 12EL) to afford a light yellow powder of the title compound (16.5 mg, 40% yield). ¹H NMR(DMSO-*d*₆ @ 300 MHz): δ 7.89 (d, *J* = 7 Hz, 2 H), 7.75 (dd^(a), *J* = {8 Hz, 5 Hz}, 2

H), 7.61 (d, $J = 7$ Hz, 2 H), 7.48-7.39 (m (= dd + d), 4 H), 7.29 (dd, $J = \{8$ Hz, 6 Hz}, 2 H), 4.20 (m, 1 H), 3.16 (dd, $J = \{6$ Hz, 14 Hz}, 1 H), 2.94 (dd, $J = \{6$ Hz, 14 Hz}, 1 H), 2.73 (t, $J = 7$ Hz, 1 H), 2.26 (d, $J = 6$ Hz, 2 H). ^{13}C NMR(DMSO- d_6 @ 125 MHz): δ 118.97.^(b) LCMS: mass calc'd for [^{13}M] $\text{C}_{24}^{13}\text{C}_1\text{H}_{20}\text{O}_4\text{N}_2$ is 413.4 and for [^{13}M – Fmoc] $\text{C}_9^{13}\text{C}_1\text{H}_{10}\text{O}_2\text{N}_2$ is 191.2. Found: (+) 414.1715 [$^{13}\text{M}+1$] and 192.1128 [^{13}M – Fmoc + 1]; (–) 412.3611 [$^{13}\text{M} - 1$] and 190.2275 [^{13}M – Fmoc – 1]. *N.B.*, the following masses are obtained on the non-isotopically enriched analog of this compound: (+) 413.1566 [$\text{M}+1$] and 191.1201 [M – Fmoc + 1]; (–) 411.3666 [$\text{M} - 1$] and 189.2350 [M – Fmoc – 1].

Test Conditions		
Temperature / °C	Time / h	Yield
140	1	5.0 mg
140	1.5	12.6 mg
150	1	11.6 mg

L-p-[^{13}C]cyanophenylalanine. The Fmoc protecting group was removed under the standard conditions of solid-phase peptide synthesis, although adapted to the solution phase.² An aliquot of N-Fmoc-L-p-[^{13}C]cyanophenylalanine (12 mg, .029 mmol) was dissolved in 0.5 mL anhydrous DMF in a scintillation vial, and 0.125 mL piperidine (20% (vol/vol)) was added. The solution was allowed to stand at room temperature for 2 h, at which point it was diluted with 10 mL ddH₂O. The aqueous phase was washed with ether (3 × 10 mL) in order to remove DMF and organic byproducts, and then lyophilized overnight. The resulting colorless film was

redissolved in buffer A (0.1% (vol/vol) TFA in ddH₂O) and loaded onto an HPLC C18 column, (A = 0.1% (vol/vol) TFA in ddH₂O; B = acetonitrile) running at 5 mL/min under a gradient where the fraction B linearly steps up from 5% to 30% at a rate of 1.2 %/min over 30 min. The chromatogram of the product possessed a single major band that lined up perfectly with that of an authentic sample of product (Peptech, no isotopic enrichment). The fraction (23 – 26 min) where product eluted was collected and lyophilized overnight to afford a colorless solid (3.9 mg, 71% yield). ¹H NMR(DMSO-*d*₆ @ 300 MHz): δ 7.83 (dd^(a), *J* = {5 Hz, 8 Hz}, 2 H), 7.48 (d, *J* = 8 Hz, 2 H), 4.06 (m, 1 H), 3.21 (dd, *J* = {6 Hz, 14 Hz}, 1 H), 3.10 (dd, *J* = {6 Hz, 14 Hz}, 1 H). ¹³C NMR(DMSO-*d*₆ @ 125 MHz): δ 118.91.^(b) LCMS: mass calc'd for [¹³**M**] C₉¹³C₁H₁₀O₂N₂ is 191.2 and for [¹³**M** – CO₂H] C₈¹³C₁H₉N₂ is 146.2. Found: (+) 192.1021 [¹³**M**+1] and 146.0880 [¹³**M** – CO₂H]; (–) 190.2225 [¹³**M** – 1]. *N.B.*, the following masses are obtained on the non-isotopically enriched analog of this compound: (+) 145.0851 [**M** – CO₂H]; (–) 189.0244 [**M** – 1].

[p-¹³CN-Phe]S-peptide (KETAAAKF[¹³CN]ERQHMDs). N-Fmoc-L-p-[¹³C]cyanophenylalanine, as prepared above, was transferred to Elim BioPharm (Hayward, CA) which employed standard Fmoc-based solid-phase peptide synthesis to generate the S-peptide. LCMS: mass calc'd for KET AAA KF¹³CNE RQH MDS is 1775. Found: (+) 1774.7 [¹³**M**], 1775.7 [¹³**M** + 1].

[p-¹³CN-Phe]RNase S. S-protein preparation and S-peptide reinsertion were done in accordance with previously described methods.^{3,4} In brief, 100 mg of bovine

pancreatic RNase A (Sigma) was digested by subtilisin (Sigma) at 0°C overnight. The digest products were denatured by lowering the pH to 2 with 1.0 M HCl, and were subjected to HPLC. The major fraction corresponding to S-protein was gathered and lyophilized overnight to afford a colorless solid (64 mg, ca. 73% yield). 37.5 mg (ca. 2.5 μ mol) of S-protein was dissolved in 350 μ L of 20 mM HEPES, pH 8.0. 4.0 mg (2.25 μ mol, .9 equiv) of [p-CN-Phe]S-peptide was dissolved in 150 μ L of 20 mM HEPES, pH 8.0. Both mixtures were inhomogeneous, but when added together, result in a homogeneous solution, consisting of ca. 5 mM [p-¹³CN-Phe] RNase S. This preparation of RNase S was used directly for NMR.

Spectroscopic Methods.

FTIR Spectroscopy. Solutions of benzonitrile (10 mM) in various solvents were prepared. In all cases, solvents of chromatography grade or higher were used. For some solvents, benzonitrile was not completely soluble up to 10 mM; the actual concentration in these cases was not explicitly determined. 20 μ L of the benzonitrile solution was loaded into a demountable liquid cell (Bruker) with two sapphire windows (.750" thick, Meller Optics). The windows were separated by using two off-set semicircular mylar spacers (of thickness 75 μ m and 100 μ m). All spectra were obtained on a Bruker Vertex 70 FTIR with a glowbar blackbody source, a KBr beamsplitter, and a sample compartment connected to a nitrogen tank to purge atmospheric gasses. A liquid N₂ cooled InSb detector was employed to detect the signal. For each run, 64 scans were collected between 2000-2500 cm^{-1} at 1 cm^{-1} resolution. Absorption spectra were calculated from the log-difference of the sample and background transmissions. The absorption spectra were baselined using a polynomial fit. The processed absorption spectra were fit to a Gaussian to obtain the peak frequencies. Nitrile frequencies for p-CN-Phe, [p-CN-Phe]S-peptide, and [p-CN-Phe]RNase S were obtained using the same FTIR methods.

¹³C NMR Spectroscopy of Benzonitrile. Solutions of benzonitrile (1% vol/vol) in various solvents were prepared. For some solvents, benzonitrile was not completely soluble up to 1%; the actual concentration in these cases was not explicitly determined. When possible, deuterated solvents were used for locking the

NMR onto deuterium's frequency: D₂O, CDCl₃, CD₂Cl₂, DMSO-*d*₆, acetone-*d*₆, cyclohexane-*d*₁₂, DMF-*d*₇. For CCl₄ and hexanes, non-deuterated solvents were used. The spectra were referenced to an external sample of 200 mM TSP (40 mM KPi, 5% vol/vol D₂O) set to 0 ppm. 600 µL of the solution was transferred to a standard NMR tube and inserted in the bore of a 500 MHz Varian NMR spectrometer.

¹³C NMR Spectroscopy of p-CN-Phe, [p-¹³CN-Phe]S-Peptide, and [p-¹³CN-Phe]RNase S. Solutions of 10 mM Phe¹³CN, 5 mM [p-¹³CN-Phe]S-peptide, and 5 mM [p-¹³CN-Phe]RNase S were prepared in 20 mM HEPES, pH 8.0, with 5% (vol/vol) D₂O. For the amino acid and the S-peptide, solutions were made by adding the material as a powder to the proper buffer to a volume of 500 µL. For RNase S, the system was synthesized in the form of an NMR sample (see the synthesis of RNase S). The solutions were transferred to a standard NMR tube. Spectra were acquired at the Stanford Magnetic Resonance Laboratory on a Bruker Avance 500 MHz spectrometer running TopSpin v1.3, and equipped with a cryogenic broadband probe (Bruker). One-dimensional carbon-13 spectra were obtained at 25°C using a single pulse sequence (90° hard-pulse) with no proton decoupling during detect, 2.5 s pre-scan delay, 33 kHz spectral window, and 99k total data points. Between 50-150 scans were averaged to furnish the FIDs, which were processed in Topspin using a 10 Hz exponential line broadening function and auto-phased. The observed resonance chemical shifts were determined using the Topspin “peak-picking” option. The spectra were referenced to an external sample of 200 mM TSP (40 mM KPi, 5% vol/vol D₂O) set to 0 ppm.

Simulation Methods.

Molecular dynamics simulation of [p-CN-Phe]RNase S was conducted in GROMACS 3.3.1 using the Amber-99 force field.⁵ The simulation strategy used here is the same as the approach previously developed for other systems with nitrile probes.^{4,6,7} In brief, the protein was solvated in explicit SPC water. A cutoff of 10 Å was used for the nonbonded interactions. Simulations were conducted with periodic boundary conditions, and the long-range electrostatics interactions were calculated with the particle mesh Ewald model. The system was energy minimized, equilibrated for 100 ps, and the production run was performed at a time step of 2 fs for 20 ns using the Nose-Hoover thermostat and the Parinello-Raman barostat. Electric fields $\vec{F}(t)$ were calculated at every time step (2fs) by summing the pairwise electrostatic forces between a charged, mass-less, virtual particle located at the midpoint of the nitrile bond and all atoms of the protein and explicit water. The H-bonding status of the nitrile in each snapshot was evaluated using the consensus cutoffs built into GROMACS' g_hbond tool: viz. a donor-acceptor distance of less than 3.5 Å, and an acceptor-donor-hydrogen angle of less than 30°.

We found that longer equilibration periods (>50 ps) were important in order to achieve the agreement with experimental electrostatic fields, as described in the main text. Averaging over a longer duration of the trajectory led generally to better agreement with experimental electrostatic fields, but shorter time-windows led to similar results, so long as they were longer than the slowest time-component of the field fluctuation.⁷

References and Notes.

- (1) Matloubi, H.; Shafiee, A.; Saemian, N.; Shirvani, G.; Daha, F. J. *J. Labelled Compd. Radiopharm.* **2004**, *47*, 31–36.
 - (2) Wang, L.; Zhang, Z. W.; Brock, A.; Schultz, P. G. *Proc. Natl. Acad. Sci.* **2003**, *100*, 56–61.
 - (3) Doscher, M. S.; Hirs, C. H. W. *Biochemistry* **1967**, *6*, 304–312.
 - (4) Fafarman, A. T.; Boxer, S. G. *J. Phys. Chem. B* **2010**, *114*, 13536–13544.
 - (5) Sorin, E. J.; Pande, V. S. *Biophys. J.* **2005**, *88*, 2472–2493.
 - (6) Xu, L.; Cohen, A. E.; Boxer, S. G. *Biochemistry* **2011**, *50*, 8311–8322.
 - (7) Bagchi, S.; Boxer, S. G.; Fayer, M. D. *J. Phys. Chem. B* **2012**, *116*, 4034–4042.
- (a) This doublet of doublets corresponds to the protons that are ortho to the nitrile, and are split by the spin- $\frac{1}{2}$ on ^{13}C .
- (b) Short ^{13}C scans showed only a single signal, and were used to identify the isotopic enrichment at the nitrile.