JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

Communication

pubs.acs.org/JACS

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

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 ^{13}C 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, $\bar{\nu}_{CN}$, 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, $\bar{\nu}_{CN}$ 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 technique[17] to decompose IR shifts of nitriles into H-bonding and electrostatic components based on an analysis that requires both $\bar{\nu}_{CN}$ and the ^{13}C chemical shift, δ^{13}_{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 ^{13}CN, which is necessary for measuring δ^{13}_{CN} in the protein. Knowledge of both $\bar{\nu}_{CN}$ and δ^{13}_{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 ^{13}CN into aromatic nitriles, in particular p-^{13}CN-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, $\bar{\nu}_{CN}$ 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

   dx.doi.org/10.1021/ja303895k | *J. Am. Chem. Soc.* 2012, 134, 10373−10376

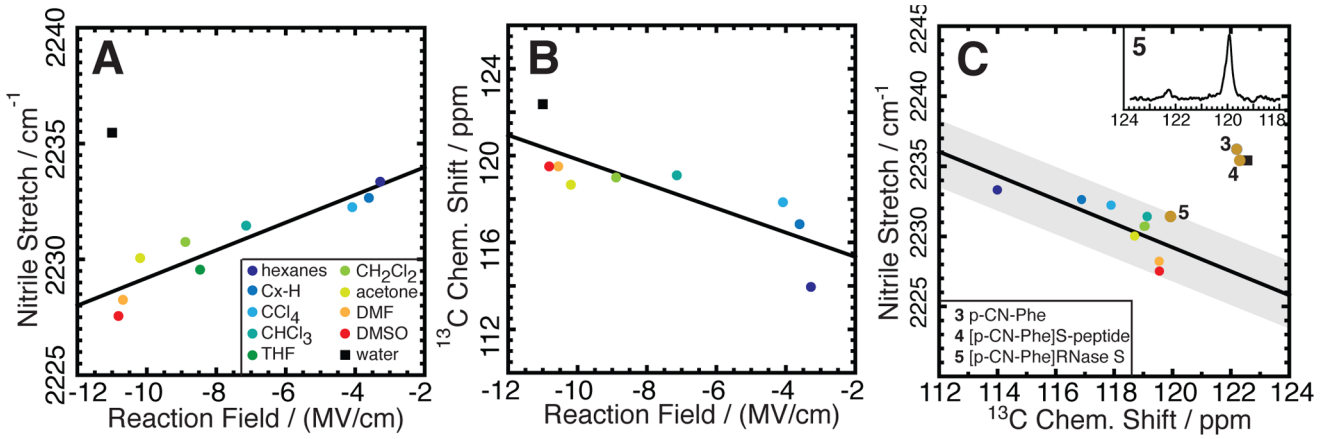


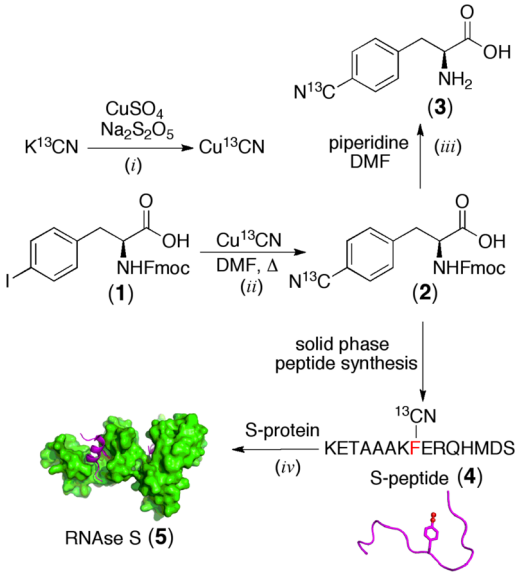
Figure 1. Solvatochromism calibrates the sensitivity of the C−N stretch and the ^{13}C chemical shift of benzonitrile to electric field and H-bonding. (A) $\bar{\nu}_{CN}$ of PhCN in various non-H-bonding solvents (colored circles) and in water (black square) compared against the electric field each solvent exercises on PhCN (see footnote 20). $\bar{\nu}_{CN}$ = 2235.2 + 0.597***F***; 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. δ^{13}_{CN} = 114.12 − 0.561***F***; R^2 = 0.66. (C) Comparison between $\bar{\nu}_{CN}$ and δ^{13}_{CN}. $\bar{\nu}_{CN}$ = 2331.7 − 0.854δ^{13}_{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 $cm^{-1}/(MV/cm)$) shows excellent agreement with the independently measured Stark tuning rate of PhCN (0.61 $cm^{-1}/(MV/cm)$),[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 $\bar{\nu}_{CN}$ 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 $\bar{\nu}_{CN}$.[17]

As described in earlier work on thiocyanate probes, a plot of $\bar{\nu}_{CN}$ 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 $\bar{\nu}_{CN}$ and δ^{13}_{CN} in environments free of H-bonds. Therefore, deviations from this line represent effects of specific interactions, which alter $\bar{\nu}_{CN}$ and δ^{13}_{CN} differently and confound the mutual linear dependence of $\bar{\nu}_{CN}$ 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}CN-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]



[a]Conditions: (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 chemistry[24] (i). N-Fmoc-p-^{13}CN-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

dx.doi.org/10.1021/ja303895k | J. Am. Chem. Soc. 2012, 134, 10373−10376

label other aromatic nitriles[26] 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 semisynthetic 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-^{13}CN-Phe]S-peptide, **4**), which in turn can be used to form [p-^{13}CN-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.

^{13}C 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 $\bar{\nu}_{CN}$'s of these three species, allow us to apply the analysis described in Figure 1. The three ($\delta_{^{13}CN}$, $\bar{\nu}_{CN}$) 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**), ($\delta_{^{13}CN}$, $\bar{\nu}_{CN}$) 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 structure[6] and from 2D IR experiments.[27] This conclusion could not have been reached if one only knew $\bar{\nu}_{CN}$ 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 $\bar{\nu}_{CN}$ 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.

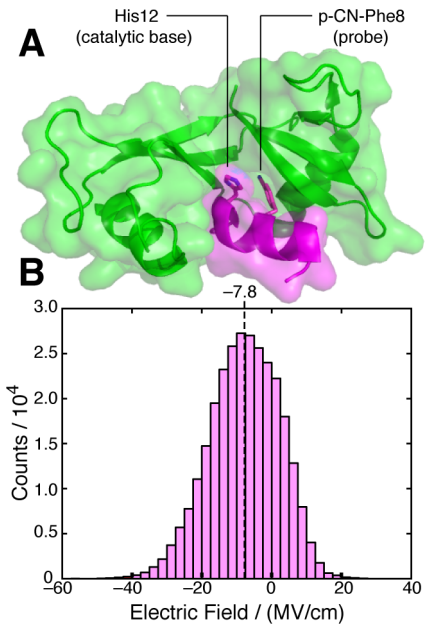


Figure 2. (A) Structure of [p-CN-Phe]RNase S (**5**). The probe, located at position 8 on the S-peptide, is shown in sticks. It is buried in a hydrophobic pocket and is close to the catalytic proton shuttle His12, shown in sticks. (B) Histogram of the electric fields experienced by the nitrile in **5**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

dx.doi.org/10.1021/ja303895k | J. Am. Chem. Soc. 2012, 134, 10373−10376

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-^{13}CN-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 $\bar{\nu}_{CN}$ 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, $\boldsymbol{F}_{Onsager}$, is given by the expression $\boldsymbol{F}_{Onsager} = (\boldsymbol{\mu}_0/a^3)[2(\varepsilon - 1)(n^2 + 2)/3(2\varepsilon + n^2)]$. It is a function of the solvent's static dielectric constant, ε, the solute's gas-phase dipole moment, $\boldsymbol{\mu}_0$, 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 Å^3, 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.; Daha, 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.