Supporting Information

Correlating Nitrile IR Frequencies to Local Electrostatics Quantifies Non-covalent Interactions of Peptides and Proteins

Pranab Deb, Tapas Haldar, Somnath M Kashid, Subhrashis Banerjee, Suman Chakrabarty*, and Sayan Bagchi*

Physical and Material Chemistry Division, CSIR, National Chemical Laboratory, Dr. Homi Bhabha Road, Pashan, Pune 411008, India.
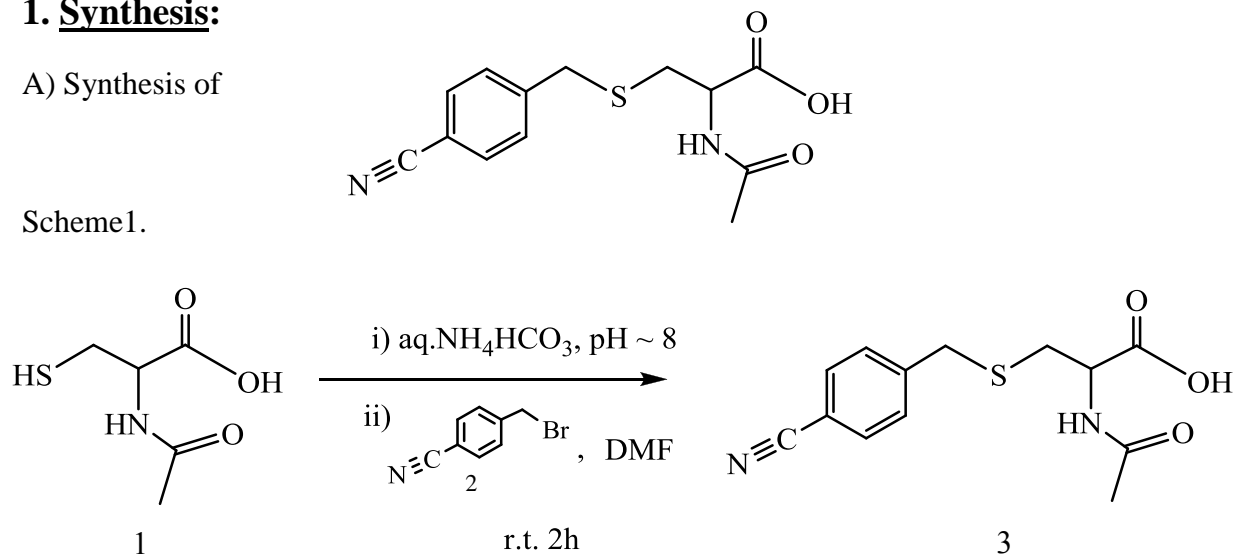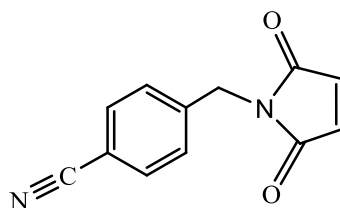
Contents:

1. **Synthesis:**

A) Synthesis of

Scheme1.

.

Preparation of 3: N-acetyl cysteine (1, 250mg, 99 mmol) is dissolved in minimum volume of degassed aq. NH_4HCO_3 solution (pH ≈ 8) in a small round bottom flask (rb). In a separate rb, 4-cyanobenzyl bromide (2, 274 mg, 118.8 mmol) is dissolved in DMF (5.5 mL) and degassed with nitrogen. The solution of 2 in DMF was added dropwise into the aq. NH_4HCO_3 solution of 1. The reaction mixture was stirred for 2 h at room temperature under nitrogen atmosphere. Product was extracted in ethyl acetate, washed with brine to remove DMF, dried over Na_2SO_4, filtered and concentrated in vacuum. The residual solid was purified by column chromatography using ethyl acetate/hexane (60:20) mixture to yield 3 (white solid).

Compound 3: ^{1}H NMR (400 MHZ, DMSO - d_6) δ 8.17 (br, 1H), 7.49-7.77 (m, 4H), 4.37 (m, 1H), 3.81 (s, 2H), 2.60-2.77 (m, 2H), 1.84 (s, 3H); ^{13}C NMR (500 MHZ, DMSO-d_6) δ 169.67, 144.75, 132.50, 130.09, 119.04, 109.75, 52.03, 35.06, 32.84, 22.57. HRMS (ESI) calculated for $C_{13}H_{15}N_2O_3S$ [M+H$^+$] is 279.3348, Found 279.0793.

B) Synthesis of

Scheme2

Step1 :

Step2 :

Preparation of 5 (Step1): To a stirring solution of 2 (500mg, 2.55 mmol) in DMF (10 mL), 5

equivalent of NaN$_3$ (explosive) added. A condenser was fitted with the rb with continuous supply

of cold water. After stirring at 60° C for 24 h, the reaction mixture was cooled to room temperature and the product was extracted with diethyl ether and washed with brine. Formation of the azide 4 was confirmed by TLC using PPh_3 (as charring agent) in hexane. Since benzyl azides are explosive, 4 was subjected to immediate reduction without further analysis.
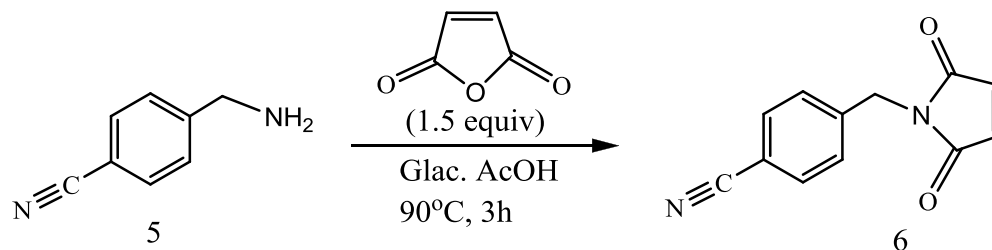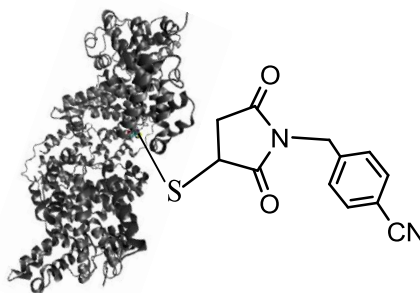
4 was dissolved in 6 mL solvent mixture of THF/water (9:1). PPh_3 (2 equiv.) added into it and the reaction mixture was stirred for 4 h. Reaction mixture was dried over Na_2SO_4. Crude product was purified by silica-gel column chromatography using EtOAc/hexane (95:5) to yield pure 5 (yellowish liquid)

Compound 5: ^{1}H NMR (200 MHZ, DMSO-d_6) δ 6.46-6.72 (m, 4H), 2.74 (s, 1H); ^{13}C NMR (200 MHZ, DMSO-d_6) δ 149.47, 132.25, 128.29, 119.31, 109.19, 45.12.; HRMS (ESI) HRMS (ESI) calculated for $C_8H_9N_2$ [M+H$^+$] is 133.1705, Found 133.0760.

Preparation of 6 (Step2): Maleic anhydride (300mg, 3.06 mmol) was added to the rb containing amine 5 (269 mg, 2.04 mmol). A water condenser was fitted with the rb. Glacial acetic acid (3.5 ml) added into the rb and the reaction mixture was refluxed at 90 °C for 3 h at stirring condition. Product was extracted with ethyl acetate followed by washing with saturated aqueous solution of $NaHCO_3$ to remove excess acetic acid. The ethyl acetate layer was dried over Na_2SO_4 and concentrated using a rotary evaporator. Finally column chromatography was performed using EtOAc/hexane (40:60) to yield 6 (yellow solid) which was characterized using NMR and mass spectrometry.

Compound 6: ^{1}H NMR (200 MHZ, CDCl$_3$) δ 7.42-7.64 (m, 4H), 6.76 (S, 2H), 4.72 (s, 2H); ^{13}C NMR (200 MHZ, CDCl3) δ 170.04, 141.10, 134.34, 132.54, 128.98, 118.43, 111.90, 40.93; HRMS (ESI) calculated for $C_{12}H_9N_2O_2$ [M+H$^+$] 213.2121, Found 213.0654.

C) Synthesis of

Scheme3:

i) aq. buffer, pH=7.6

ii) (20 equiv) , DMSO

6

4°C, overnight

BSA

7

Preparation of 7 : A 50 mM stock solution of 6 was prepared by dissolving 6 in DMSO followed by degassing with nitrogen. 6 µL of the DMSO solution was added drowise to a degassed aqueous buffer (40 mM sodium phosphate and 2mM EDTA, pH=7.6) containing 150 µM BSA (Bovine Serum Albumin). The final reaction volume was 100 µL and the concentration of reagent 6 was 20 times higher than the protein concentration. The mixture was incubated for 12 h at 4°C. Completion of the reaction was confirmed from Elmann's assay.[1] The conjugate 7 was purified using gell filtration chromatography. HRMS (ESI) analysis of the tripsin digested peptides of the conjugate revealed that only Cys-34 of BSA was modified. We have followed the tripsin digestion procedure as reported by Sen Gupta and co-workers.[2] The masses of one of the peptide fragments containing Cys-34 derived from the unmodified and modified BSA were 974.4420 and 1186.5020 respectively. The difference in the mass, 212.06, corresponds to the exact mass of the modified unit (6).

2. Empirical model to describe field-frequency correlation in H-bonded nitriles in a qualitative manner:

For an ensemble of solute molecules in any aqueous binary mixture, we can consider the experimental peak position ($\bar{\nu}^{exp}$) to be a weighted average of the average frequencies corresponding to the overlapping H-bonded ($\bar{\nu}^{HB}$) and non-H-bonded ($\bar{\nu}^{NHB}$) populations (Figure S7)

$$\bar{\nu}^{exp} = X\bar{\nu}^{HB} + (1-X)\bar{\nu}^{NHB} \quad \text{(Equation 1)}$$

where X is the population fraction of the H-bonded configurations.

For nitriles, the peak separation in $\bar{\nu}^{HB}$ and $\bar{\nu}^{NHB}$ ($\Delta\bar{\nu}$) is small (maximum between DMSO and water, ~7.5 cm^{-1}), which makes the experimental spectrum resemble a single peak. However, multiple populations have been distinguished experimentally for C=O in DMSO/water mixtures as the peak separation increases.[3] Equation 1 can be rearranged as

$$\bar{\nu}^{exp} = \bar{\nu}^{NHB} + X(\bar{\nu}^{HB} - \bar{\nu}^{NHB}) = \bar{\nu}^{NHB} + X(\Delta\bar{\nu}) \quad \text{(Equation 2)}$$

Noting that $\Delta\bar{\nu}$ is positive for nitriles and X increases with increasing water content of the binary mixtures, the above expression illustrates that $\bar{\nu}^{exp}$ will show a monotonic increase for a nitrile probe. Figure S7 also illustrates that the peak position will show an increasing blue shift as the fraction of the H-bonded population increases. However, a reverse trend, with $\bar{\nu}^{exp}$ showing a monotonic decrease, is expected for the carbonyls as $\Delta\bar{\nu}$ is negative. As a matter of fact, the experimental results on nitriles and carbonyls in DMSO/water mixture confirm the predicted

trends from our model. Moreover, the linear dependence of $\overline{v}_{C\equiv N}{}^{exp}$ to the electric field, as seen from our results, indicates that field varies linearly with X. As the overlapping peaks of the two populations are not discernible in case of nitriles, X cannot be directly obtained for each binary mixture from the experimental IR spectra.

If we suppose that X is proportional to the electric field, we can write $X = \left(\frac{\Delta X}{\Delta F}\right) F$. Replacing the expression for X in Equation 1, we get

$$\overline{v}^{exp} = \overline{v}^{NHB} + \left(\frac{\Delta X}{\Delta F}\right) F(\Delta \overline{v}) = \overline{v}^{NHB} + \left(\frac{\Delta \overline{v}}{\Delta F}\right)(\Delta X)F \quad \text{(Equation 3)}$$

As the solutes are H-bonded in water and non-H-bonded in the neat non-aqueous solvent, $\Delta X = 1$ when the change is considered between an organic solvent and water as solvents. For carbonyls, the H-bonding interactions are electrostatic and thus $\frac{\Delta \overline{v}}{\Delta F}$ equals to the Stark tuning rate. Therefore, our model predicts that if $\overline{v}^{exp}$ varies linearly with X, and the slope of the line will be equal to the Stark tuning rate. This has been confirmed by earlier published results, which show the field-frequency linearity is maintained in H-bonding and non-H bonding solvents for carbonyls without the change of the slope. As H-bonding interactions are non-electrostatic for nitriles, though non-H bonding and H-bonding solvents independently show linear field-frequency correlation, the slopes of these correlations will differ in both magnitude and sign. Our results agree with the model and qualitatively explain the non-Stark linear dependence of $\overline{v}_{C\equiv N}$ and $F_{C\equiv N}$ in H-bonding environments.
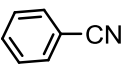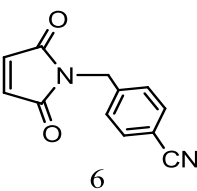
3. <u>Supplementary Tables:</u>

Table S1. Analysis of the partial charges on different atoms of the diatomic IR probes.

Species/Environment	q_C	q_N	q_O	Δq	$\Delta\Delta q$
Benzonitrile (gas)	0.326	-0.458		−0.784	0
Benzonitrile (PCM: continuum water)	0.383	-0.553		−0.936	-0.152
Benzonitrile + 1 H_2O (gas)	0.187	-0.360		−0.547	+0.237
Acetophenone (gas)	0.536		-0.502	−1.038	0
Acetophenone (pcm: continuum water)	0.582		-0.586	-1.168	-0.13
Acetophenone + 1 H_2O (gas)	0.640		-0.528	−1.168	-0.13

q_C: partial charge on carbon atom, q_N: partial charge on the nitrogen atom, q_O: partial charge on the oxygen atom

Table S2. Nitrile stretch of PhCN and **6**

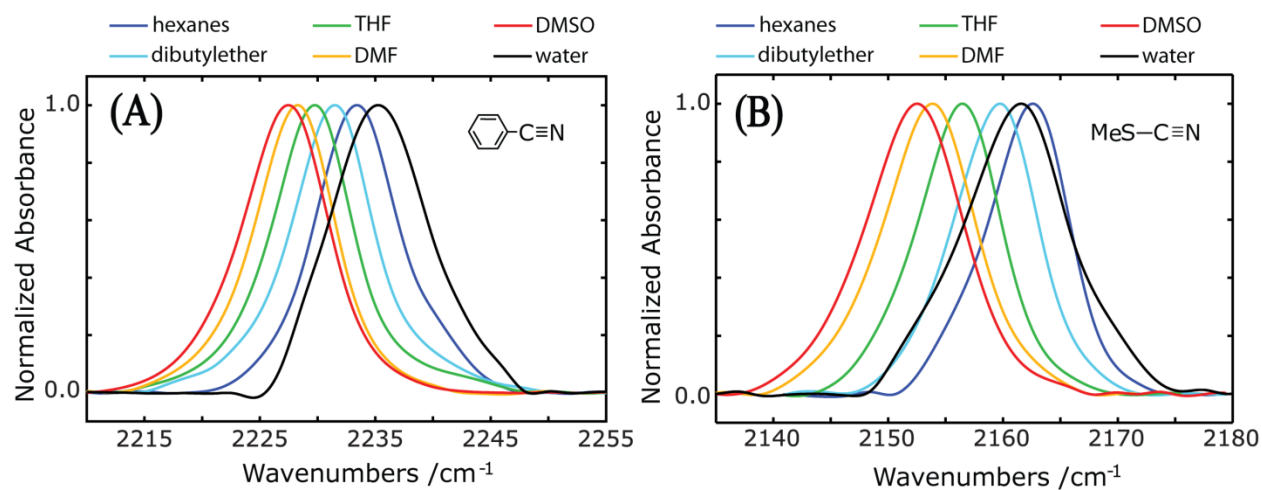| Solvent	Nitrile IR stretching frequency / cm^{-1}	
		6
THF	2229.8	2229.9
DMSO	2227.5	2227.6

Nitrile stretching frequency of PhCN and the linker (**6**) have been measured in THF and DMSO at 10 mM concentration. The nitrile frequencies of the linker, **6,** are within 0.1 cm^{-1} to the nitrile frequencies of the PhCN in both the solvents.

Table S3. Nitrile IR stretching frequencies in different THF-water mixtures.

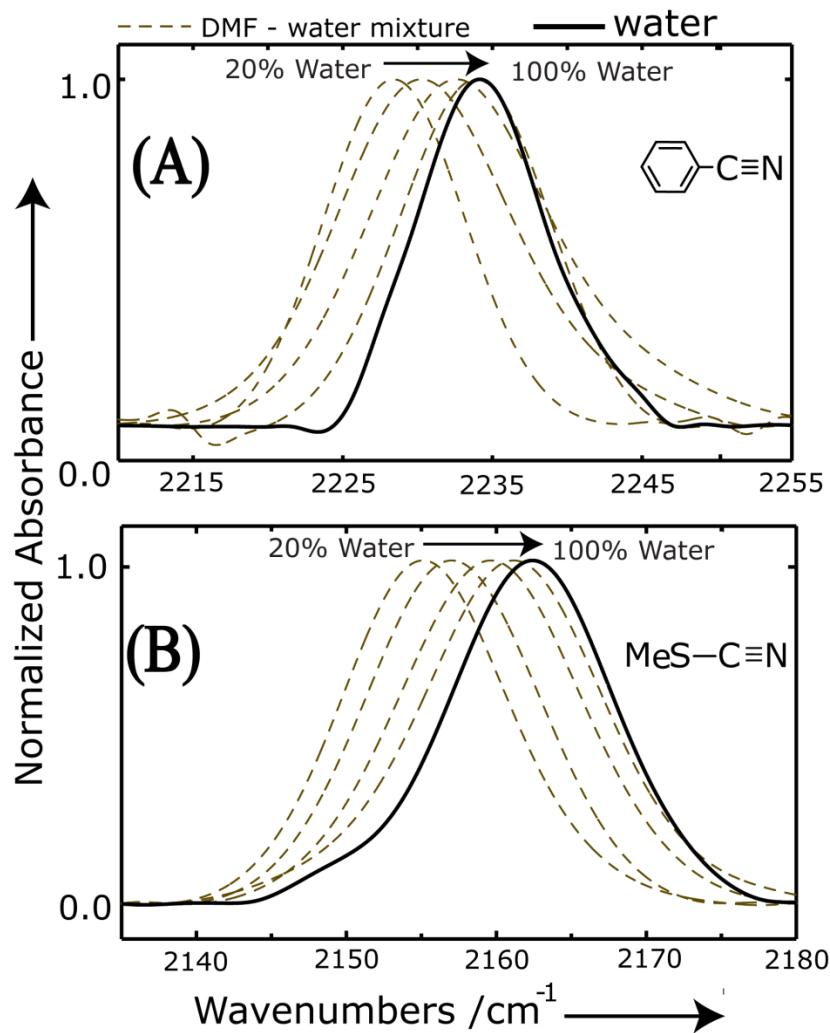| % of water in THF-water mixtures.	Nitrile IR stretch/cm^{-1}	
	PhCN	MeSCN
40	2229.8	2156.8
60	2231.6	2160.9
80	2234.8	2161.3

4. <u>Supplementary Figures:</u>

Figure S1. Vibrational Solvatochromism of nitrile vibrational probes in neat Solvents
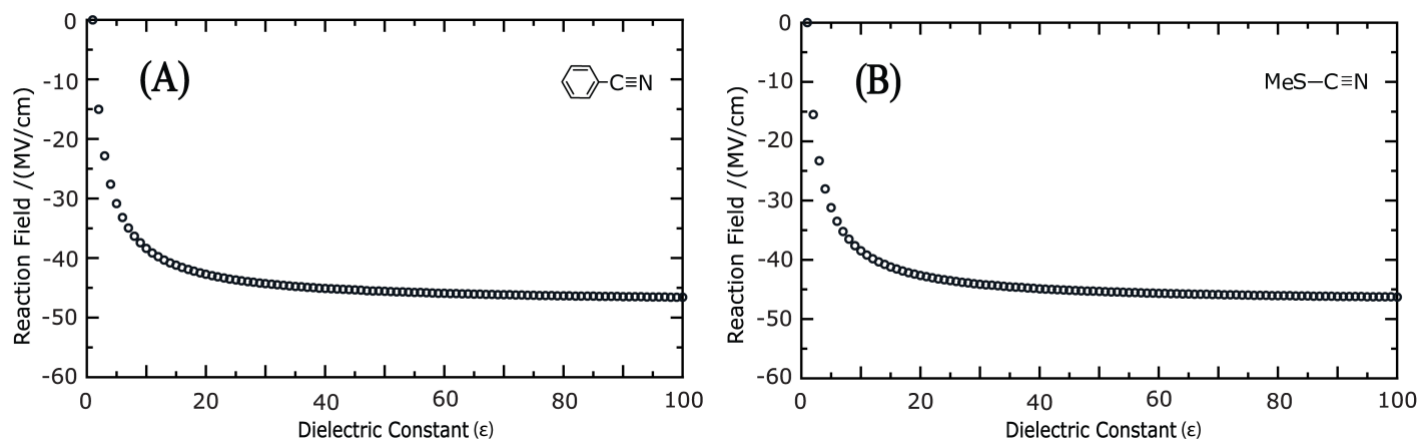

Figure S1. FTIR absorption spectra of the nitrile for (A) PhCN and (B) MeSCN in different solvents at 10 mM concentration. With increasing solvent polarity from hexane to DMSO, the nitrile stretching frequency gradually decreases (red shift). In the protic solvent, water, the nitrile stretching frequency shows a blue shift as compared to DMSO.

Figure S2. Vibrational Solvatochromism of nitrile vibrational probes in DMF/water binary mixtures.
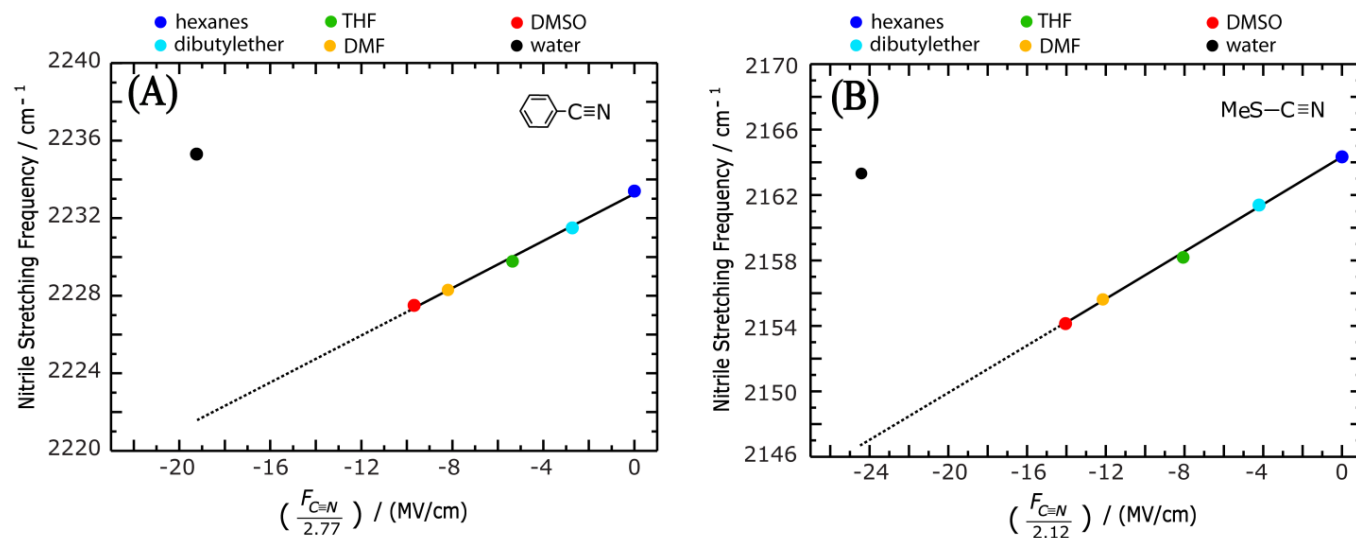

Figure S2. FTIR spectra of C≡N stretching band of (A) PhCN and (B) MeSCN in water and DMF-water binary solvent mixtures at 10 mM concentrations. Nitrile stretching frequency gradually increases upon increasing the water content in DMF/water mixtures by 20% (v/v).

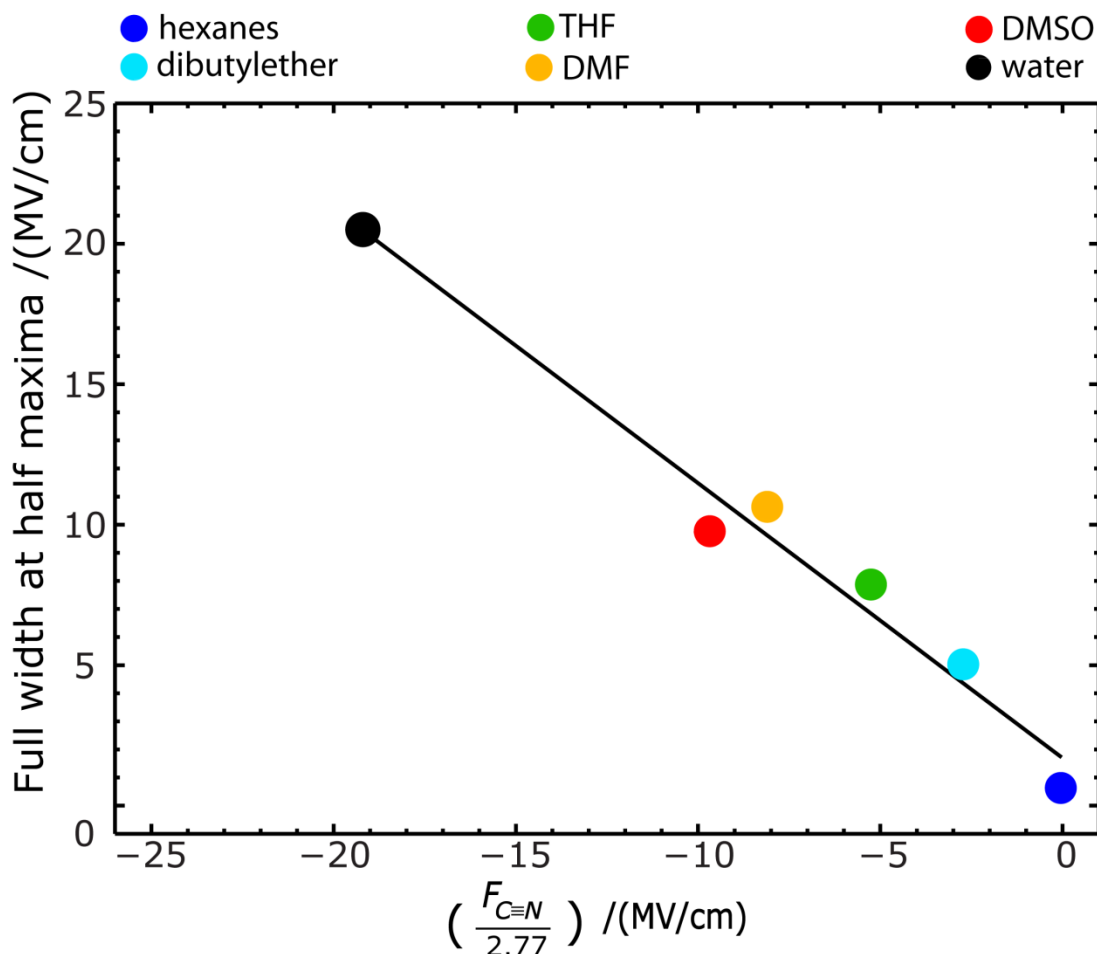**Figure S3. Dielectric constant Vs Onsager field**

Figure S3. Reaction field experienced by the nitrile probe reaches almost a constant value beyond dielectric constant ~40. Here reaction fields is calculated using the equation, $F_{Onsager} = \left(\frac{\mu_0}{a^3}\right)\left(\frac{2(\epsilon-1)(n^2+2)}{3(2\epsilon+n^2)}\right)$, where a is the radius of the spherical cavity, n is the refractive index of the solute and ϵ is the dielectric constant of the solvent. Sign of the reaction field is considered as negative since reaction field stabilizes solutes energy levels in solutions.
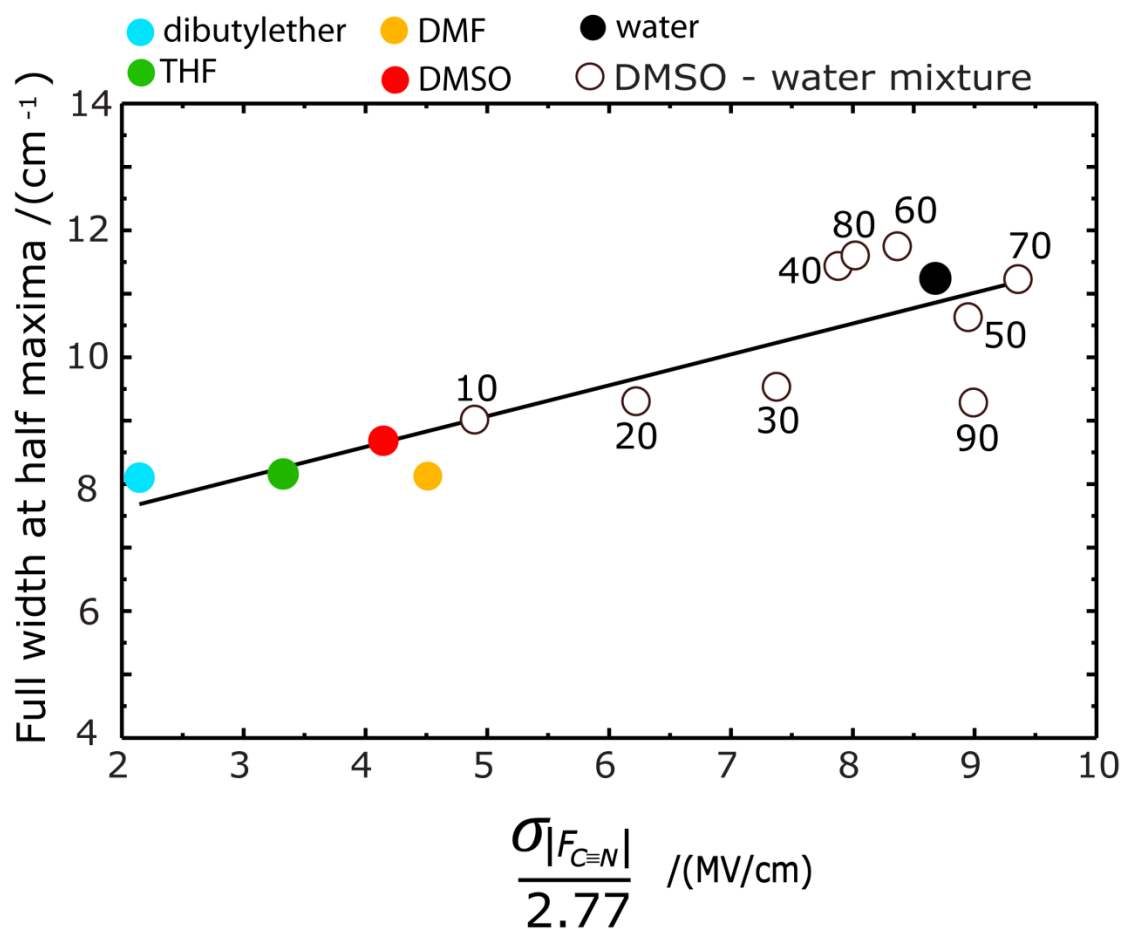
Figure S4. Field-frequency correlation

Figure S4. Field-frequency correlation of (A)PhCN and (B)MeSCN. The solid lines illustrates that the nitriles follow the Stark relation along this line. The doted lines are the extension of the Stark line at higher field values. Any point which arises as a result of non-electrostatic contribution will deviate fron this line since Stark relation is applicable as long as the interaction is electrostatic. In both plots the $(F_{C\equiv N}, \bar{v}_{C\equiv N})$ points of water are off the dotted lines indicating the presence of non-electrostatic contributions when nitrile forms H-bond. The $(F_{C\equiv N}, \bar{v}_{C\equiv N})$ points for the all DMSO/water mixtures also deviate from the dotted lines.

Figure S5. Electric field heterogeneity in neat solvents.

Figure S5. Full width half maxima (FWHM) of the electric fields (as obtained from MD) distribution and the average electric fields exerted on the nitrile probe are compared in case of benzonitrile (PhCN) in different solvents. MD calculated electric fields on the nitrile probe at each point of time have been fitted with a Gaussian distribution. The standard deviation (σ) obtained from this fitting has been used to calculate the FWHM of the field distribution using the relation, $FWHM = (2\sqrt{2ln2})\sigma$. The correlation reveals that as solvent polarity increases nitrile experiences larger electric field and heterogeniety in electric field increases at larger values of electric field. The equation of the best fit line is, $FWHM=-0.98(F_{C\equiv N}/2.77) + 1.69$, ($R^2 = 0.98$).
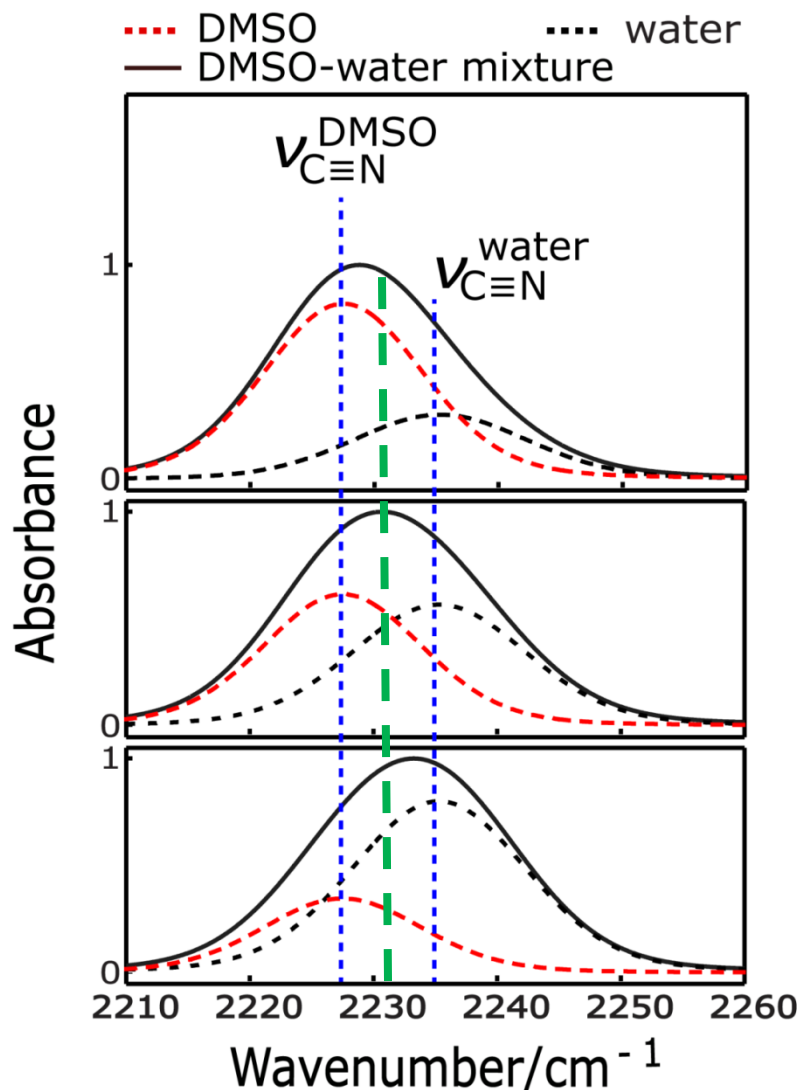
Figure S6. Linewidth Vs electric field distribution.

Figure S6: Linewidth of the experimental C≡N IR spectra of PhCN taken in different solvents and solvent mixtures is plotted against the standard deviation $(\sigma_{|F_{C≡N}|}/2.77)$ of the field distribution for different neat solvents/ solvent mixtures.The detail procedure for calculating the standard deviations is mentioned in the caption of Figure S5. The moderate strength of the linear correlation ($R^2 = 0.70$) indicates that the standard deviation is not sufficient to explain all the variations in experimental full width at half maxima (FWHM). The equation of the best fit line is, *FWHM (frequency) = 0.49($\sigma_{|F_{C≡N}|}$/2.77) + 6.64*
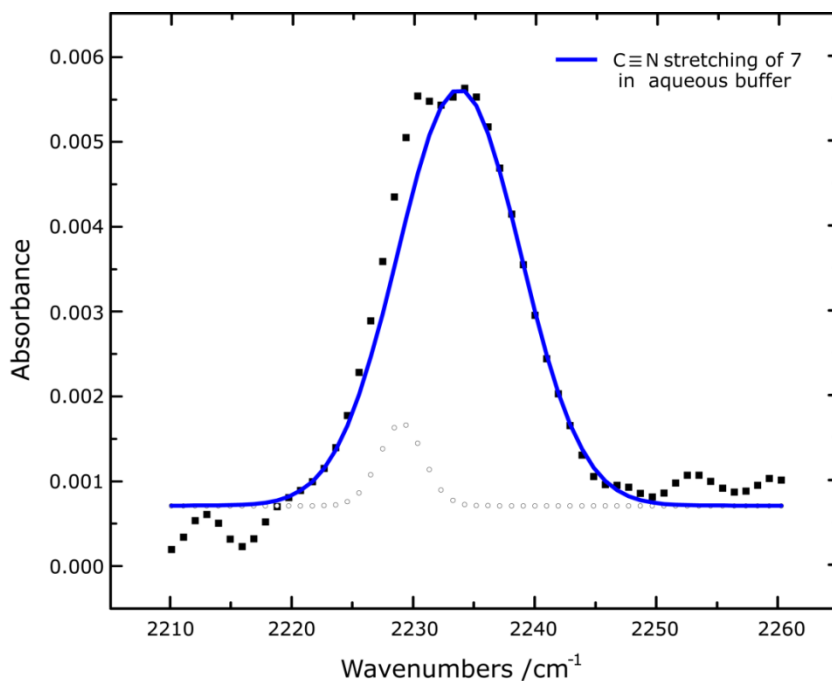
Figure S7. Model explaining blue shift of the overall frequency as the population fraction of the H-bonded nitrile increases

Figure S7. In any binary aqueous mixture the peak maxima of experimentally obtained nitrile absorption spectra (solid line) can be modeled as a weighted average of the average frequencies corresponding to the overlapping combination of non-H-bonded population (red dots) and. H-bonded populations (black dots). Small difference in the frequencies of the nitriles (PhCN & MeSCN) in DMSO (or DMF) and water makes the absorption spectra resemble a single peak. As the population of the H-bonded nitrile increases the absorption maxima shows a blue shift (the green dashed vertical line represents the peak maxima with equal populations of non-H-bonded nitriles (organic solvent) and H-bonded nitriles (water).

Figure S8. FTIR absorption of nitrile stretching mode of 7 in aqueous-buffer.

Figure S8: FTIR absorption spectra of the C≡N stretching band (at a resolution of 1 cm^{-1}) of **7** (~2.5mM) in aqueous buffer (40 mM NaH_2PO_4, pH=7.6) is represented by the black squares. The spectum has been fitted by using two Gaussian function. The high frequency peak (blue curve) at 2234.2 cm^{-1} is the nitrile stretching frequency of **7**. The low frequency peak (gray circles) at 2229.4 cm^{-1} arises from a very weak water band. The existence of this water band has been confirmed by analysing the FTIR spectrum of neat water. This weak peak contributes to the lineshape of the nitrile stretch when the nitrile peak intensity is low. The maximum possible concentration attained for the nitrile modified protein (**7**) is ~2.5 mM as compared to the 10 mM concentration used for the model nitriles. The peak arsing from water has also been verified multiple times with nitriles of known frequencies.

5. References:

(1) Ellman, G. L.Tissue Sulfhydryl Groups. *Arch. Biochem. Biophys.* **1959**, *82,* 70-77.
(2) Kumari, S.; Panda, C.; Mazumdar, S.; Sen Gupta, S.A Molecular Fe-Complex as a Catalyst Probe for in-Gel Visual Detection of Proteins Via Signal Amplification. *Chem. Commun.* **2015**.
(3) Kashid, S. M.; Bagchi, S.Experimental Determination of the Electrostatic Nature of Carbonyl Hydrogen-Bonding Interactions Using Ir-Nmr Correlations. *J. Phys. Chem. Lett.* **2014**, *5,* 3211-3215.