# PCCP

This article can be cited before page numbers have been issued, to do this please use: B. Basiak, A. W. Ritchie, L. J. Webb and M. Cho, *Phys. Chem. Chem. Phys.*, 2016, DOI: 10.1039/C6CP01578F.

View Article Online
DOI: 10.1039/C6CP01578F

Journal Name

ARTICLE

# Vibrational solvatochromism of nitrile infrared probes: Beyond vibrational Stark dipole approach†

Received 00th January 20xx,
Accepted 00th January 20xx

DOI: 10.1039/x0xx00000x

**www.rsc.org/**

Bartosz Błasiak,[a,b] Andrew W. Ritchie,[c] Lauren J. Webb[c] and Minhaeng Cho[a,b,*]

Systematic probing of local environments around biopolymers is important in understanding their functions. Therefore, there has been a growing interest in *in-situ* measurements of molecular granularity and heterogeneity through the systematic analysis of vibrational frequency shifts of carbonyl and nitrile infrared probes with vibrational Stark dipole theory. However, we here show that the nitrile vibrational frequency shift induced by its interaction with surrounding molecules cannot be solely described by electric field-based theory because of the exchange-repulsion and dispersion interaction contributions. Considering a variety of molecular environments ranging from bulk solutions to protein environments, we explore the distinct scenarios of solute-environment contacts and their traces in vibrational frequency shifts. We believe that the present work could provide a set of clues that could be potentially used to design a rigorous theoretical model linking vibrational solvatochromism and molecular topology in complex heterogeneous environments.

## 1 Introduction

Vibrational solvatochromic response of small infrared probe molecules has been widely used to track changes of the three dimensional structures of biomolecules.[1-3] In particular, the vibrational Stark effect theory, which is based on the assumption that the vibrational frequency shifts are mainly determined by interaction between vibrational transition dipole and local electric field, has been widely used to extract quantitative information on the local electric field around an IR probe.[1, 4, 5] One of the most versatile IR probes used so far were carbonyl (CO) stretches and nitrile (CN) stretches. Despite that a variety of nitrile-derivatized amino acids were selectively and specifically introduced into a macromolecular framework and used extensively[6-11], still there is a lack of quantitatively reliable model for describing intermolecular interaction-induced vibrational frequency shifts.

It was demonstrated before that solvatochromism of CN stretch vibration in various solvents is quite complicated and cannot be explained simply by the Coulombic interaction of vibrational dipole with local electric field produced by surrounding molecules[12-17], which is in notable contrast to the solvatochromism of CO stretch modes.[1, 18, 19] However, it was recently shown that even in the case of amide I mode of *N*-methylacetamide (NMA) the origin of the solvation-induced frequency shift is far more complicated than it was previously conjectured.[20] For instance, Pauli's exclusion principle manifests in strong blue shifts of vibrational frequencies as opposed to red shifting Coulomb, induction and dispersion effects. Such strong blue-shifting behaviour due to 'repulsion wall'[21-23] was attributed to the large slope of potential energy that is 'felt' by vibrational probe when interacting strongly with water molecule via hydrogen bond.[24] The importance of exchange-repulsion effect even on the NMA amide I mode frequency shift in water suggests that the repulsive interaction-induced solvatochromism could be the dominant contribution to the frequency blue-shifts of nitrile stretch modes in strongly H-bonding solvents. In fact, by the analysis of the forces computed from the classical molecular dynamics simulations and their effects on the vibrational frequency, it was found that the vibrational frequency blue-shifts of CN stretches of acetonitrile or $CN^-$ anion in water result from van der Waals interactions.[15, 16]

The fact that the non-specific electrostatic and specific H-bonding effects contribute differently to the vibrational shifts was shown experimentally by the Boxer group[13, 25], in which they combined IR and NMR techniques together to separate the H-bonding contributions. Recently, Zhang et al.[26] studied the specific and nonspecific nitrile frequency shifts in analogs of 5-cyanotryptophan by utilizing empirically determined Kamlet-Taft parameters[27-29] of various solvents. They found a simple three-parameter linear formula that can be used to relate the CN absorption frequency with the solvent polarity

---

[a] *Center of Molecular Spectroscopy and Dynamics, Institute of Basic Science (IBS), Seoul 02841, Republic of Korea.*
[b] *Department of Chemistry, Korea University, Seoul 02841, Republic of Korea.*
[c] *Department of Chemistry, Center for Nano- and Molecular Science and Technology, and Institute for Cell and Molecular Biology, The University of Texas at Austin, 105 E. 24th Street, STOP A5300, Austin, Texas 78712, U.S.A.*
** email: mcho@korea.ac.kr*

† Electronic Supplementary Information (ESI) available: Experimentally measured FTIR spectra, vibrational potential maps of CN stretch mode, approximate aspects of SolEFP theory, distance-dependence of the vibrational solvatochromic interactions between MeSCN and solvent molecules, structures of the model systems, directionality of the induced effects of the CN stretch solvatochromic dipole, test of the fragmentation approach to model the protein environment by EFP, short-range nitrile frequency shifts of free RalGDS with SCN label and CN frequency shifts in MeSCN-$CH_4$ dimer. See DOI: 10.1039/x0xx00000x

View Article Online
DOI: 10.1039/C6CP01578F

and H-bonding properties. However, this and other similar approaches[12] are based on the empirical data and, therefore, cannot provide detailed insights into the physical origins of the vibrational frequency shifts. Nevertheless, it is apparent that H-bonding interaction between nitrile and protic solvent exerts a blue-shift of nitrile stretch frequency, which is probably beyond Coulomb interaction effect.

To show the complicated relationship between nitrile stretch mode frequency and solvent polarity, the experimentally measured vibrational frequencies of CN stretch modes of acetonitrile (MeCN) and methyl thiocyanate (MeSCN) dissolved in solvents of varying polarity and proticity are plotted with respect to the corresponding Onsager factors determined by solvent dielectric constants (Figure 1). For aprotic and non-polar solvents, the linear dependence of nitrile frequency shift on the Onsager factor is evident, suggesting the validity of the Kirkwood-Bauer-Magat (KBM) theory.[30, 31] However, as the solvent polarity increases, the reaction field theory for vibrational solvatochromism fails. Furthermore, as the H-bond donating ability of solvent molecule increases, the nitrile stretch frequency becomes strongly blue-shifted, as compared to those in DMSO for instance.[14] Theoretically, recent quantum mechanical calculations for solute-solvent clusters showed that the interaction between oscillator's quadrupole and the solvent electric field gradient is at least equally important in quantitatively describing vibrational frequency shifts of nitrile and other related IR probe modes in water, indicating the breakdown of simple Stark-dipole theory.[18, 32, 33] Thus, from the experimental data in Figure 1 as well as from classical and quantum chemistry calculation studies, it becomes quite clear that a more refined theory beyond the KBM and Stark dipole models is needed. In particular, it is strongly desirable to construct a reliable and robust vibrational solvatochromic model for highly heterogeneous environment such as proteins and nucleic acids in aqueous solutions. Up to the present time, there is no such model beyond simple vibrational Stark dipole theory with including contributions from non-Coulombic interactions.

In the present work, we investigate the significance of exchange-repulsion interaction and dispersion effect on the vibrational frequency shifts of various nitrile stretch modes. Using our first-principles solvatochromism theory based on the effective fragment potential method[20] (hereafter it will be referred to as SolEFP), we studied vibrational solvatochromism of nitriles in a variety of molecular environments ranging from bulk solutions to proteins. A detailed description on the SolEFP theory is presented in Section 2. In Section 3.1 we first elaborate on the vibrational properties of nitrile stretch mode in vacuum and in Section 3.2 relatively small model systems involving hydrogen bonds are considered. The direct comparisons with experimental results are made for the bulk solutions in Section 3.3. The nature of the specific solvatochromic frequency shifts of nitrile probes in proteins are discussed in Section 3.4. The main results are summarized in Section 4 with a few concluding remarks.

## 2 Theory

### 2.1 Vibrational electrochromism

The vibrational electrochromism theory on the vibrational spectral change in response to a static external electric field was already fully developed.[33-37] If the IR chromophore is exposed to a general electric field $\mathbf{F}$, the vibrational frequency shift of a *j*th normal mode from its gas-phase frequency $\omega_{j,0}$ can be a function of the external electric field:

$$\Delta\omega_j(\mathbf{F}) \equiv \omega_j(\mathbf{F}) - \omega_{j,0} = -\boldsymbol{\mu}_{j,0}(\mathbf{r}_0)\cdot\mathbf{F} - \frac{1}{2}\boldsymbol{\alpha}_{j,0}(\mathbf{r}_0):\mathbf{F}\otimes\mathbf{F} - \frac{1}{3}\boldsymbol{\Theta}_{j,0}(\mathbf{r}_0):\nabla\mathbf{F} + \ldots \tag{1}$$

where $\omega_j(\mathbf{F})$ is the electric-field-dependent frequency and $\boldsymbol{\mu}_{j,0}$, $\boldsymbol{\alpha}_{j,0}$ and $\boldsymbol{\Theta}_{j,0}$ are the vibrational solvatochromic difference dipole, polarizability, and quadrupole moment, respectively, defined with respect to the origin at $\mathbf{r}_0$. If the applied electric field is uniform in space, all its spatial derivatives vanish so that only the first two terms on the right hand side of Eq. (1) are to be considered for vibrational electrochromism. The simplest choice is to take into account the first (dipole-field) term, which leads to the most frequently used model, i.e., Stark-dipole model,

$$\Delta\omega_j(\mathbf{F}) \approx -\boldsymbol{\mu}_{j,0}(\mathbf{r}_0)\cdot\mathbf{F}(\mathbf{r}_0). \tag{2}$$

The magnitude of the proportionality constant $\boldsymbol{\mu}_{j,0}$ in Eq. (2) is called the Stark tuning rate that can be measured with vibrational Stark spectroscopy with an external electric field applied to solid-state sample where IR chromophores are fixed in space.[38, 39]

### 2.2 Vibrational solvatochromism: A general theory

Over the years, to describe the vibrational solvatochromism of nitrile probes, most of the previous studies assumed that the *solvatochromism* (not electrochromism) of those CN and SCN probes is strictly *electrostatic* in nature, even including H-bonding.[18, 40, 41] Probably this is because the Coulombic interactions are relatively large and furthermore straightforward to take into consideration. However, due to the fact that an IR probe under the interaction with electrostatic environment can involve either making or breaking of hydrogen bond(s), the non-electrostatic short-range effects change drastically leading to the need of significant modifications of Eqs. (1) and (2). Within the SolEFP theory, the vibrational Coulombic frequency shift of the *j*th IR-active mode when it interacts with solvent electric potential $\phi$ can be written as

$$\Delta\omega_j^{\mathrm{Coul}}(\phi) \equiv \omega_j(\phi) - \omega_{j,0} = \Delta\omega_j^{\mathrm{SolCAMM}} + \frac{1}{2M_{j,0}\omega_{j,0}} \times \sum_{\substack{x\in\mathrm{solute}\\ \mathrm{atoms}}} \sum_{\substack{y\in\mathrm{solvent}\\ \mathrm{atoms}}} \sum_{i} \left\{ \frac{g_{ijj}}{M_{i,0}\omega_{i,0}^2} q_x - 2\delta_{ij}\left(\frac{\partial q_x}{\partial Q_i}\right)_{\mathbf{Q}_0} \right\} q_y \frac{\mathbf{v}_x^{(i)}\cdot\mathbf{r}_{xy}}{r_{xy}^3} \tag{3}$$

where

View Article Online
DOI: 10.1039/C6CP01578F

$$\Delta\omega_j^{\text{SolCAMM}} = \sum_{\substack{x\in\text{solute}\\ \text{atoms}}} \Big\{ l_{jx}\phi(\mathbf{r}_x) + \mathbf{L}_{jx}\cdot\boldsymbol{\nabla}\phi(\mathbf{r}_x) + \frac{1}{3}\boldsymbol{\lambda}_{jx}\colon\boldsymbol{\nabla\nabla}\phi(\mathbf{r}_x) + \frac{1}{15}\boldsymbol{\Lambda}_{jx}\,\vdots\,\boldsymbol{\nabla\nabla\nabla}\phi(\mathbf{r}_x) + \ldots\Big\} \tag{4}$$

Here, $l_{jx}$, $\mathbf{L}_{jx}$, $\boldsymbol{\lambda}_{jx}$, and $\boldsymbol{\Lambda}_{jx}$ are the distributed solvatochromic charges, dipoles, quadrupoles, and octupoles, respectively, that can be estimated with the first-principles quantum chemistry calculation method (see Ref.[42]). In Eq. (3), $q_x$ is the charge of the *x*th solute atom at $\mathbf{r}_x$, $\mathbf{v}_x^{(i)}$ is the *i*th normal mode vibrational eigenvector of the *x*th solute atom, $\mathbf{r}_{xy}$ is the interatomic position vector defined as $\mathbf{r}_{xy} = \mathbf{r}_x - \mathbf{r}_y$, $g_{ijj} = (\partial^3 E/\partial Q_i \partial Q_j^2)_{\mathbf{Q}_0}$is the gas-phase cubic anharmonic constant, $E$ is the solute total energy, and $M_{i,0}$ and $\omega_{i,0}$ are the reduced mass and harmonic frequency of the solute molecule in the gas phase, respectively. In Eq. (4), we used the cumulative atomic multipole moments (CAMM)[43] to describe the distribution of electric potential (hence the superscript SolCAMM in Eqs. (3) and (4) means that the vibrational solvatochromic shift is calculated with CAMM.

The exchange-repulsion contribution to the first-order (with respect to the solute-solvent interaction) electrostatic interaction-induced vibrational frequency shift can be extracted directly from the wavefunction and its derivatives with respect to the normal coordinates,

$$\begin{aligned}\Delta\omega_j^{\text{Ex-Rep}} = &-\frac{1}{2M_{j,0}\omega_{j,0}}\sum_i \frac{g_{ijj}}{M_{i,0}\omega_{i,0}^2}\\ &\times\frac{\partial}{\partial Q_i}\Bigg\{\sum_{a\in\text{A}}\sum_{b\in\text{B}}\Bigg[-4\sqrt{\frac{-2\ln|S_{ab}|}{\pi}}\frac{S_{ab}^2}{r_{ab}}\\ &-2S_{ab}\Bigg(\sum_{c\in\text{A}}G_{ac}^{\text{A}}S_{cb} + \sum_{d\in\text{B}}G_{bd}^{\text{B}}S_{da} - 2T_{ab}\Bigg)\\ &-2S_{ab}^2\Bigg(\frac{1}{r_{ab}} + \sum_{x\in\text{A}}\frac{Z_x}{r_{xb}} + \sum_{y\in\text{B}}\frac{Z_y}{r_{ya}} - \sum_{c\in\text{A}}\frac{2}{r_{cb}} - \sum_{d\in\text{B}}\frac{2}{r_{da}}\Bigg)\Bigg]\Bigg\}_{\mathbf{Q}_0}\end{aligned} \tag{5}$$

where $G_{ac}^{\text{X}}$ is the Fock matrix element of molecule X in the localized molecular orbital (LMO) space, $S_{ab}$ and $T_{ab}$ are the overlap and kinetic integrals between *a*th and *b*th LMO's and the relative distance between the LMO's is defined as $r_{kl} = |\mathbf{r}_k - \mathbf{r}_l|$. In Eq. (5), the labels A and B denote solute and solvent molecules, the indices *a*, *b*, *c*, and *d* refer to LMO's, whereas the indices *x* and *y* to atomic sites, respectively.[24]

Since all the molecules are polarizable when they are exposed to electric field, the induction effect on vibrational frequency shift cannot be ignored, even though its magnitude can often be smaller than other contributions. Nonetheless, we showed that the induction (polarization) frequency shift can be recast in the following form

$$\Delta\omega_j^{\text{Ind}} = -\frac{1}{2}\mathbf{a}_j\cdot\mathbf{F} - \frac{1}{2}\mathbf{F}^T\cdot\mathbf{A}_j\cdot\mathbf{F} \tag{6}$$

where $\mathbf{a}_j$ and $\mathbf{F}$ are the vectors consisting of distributed vibrational solvatochromic induced-dipoles and electric fields at the distributed LMO sites, respectively:

$$\begin{aligned}\mathbf{a}_j &= \begin{pmatrix}\mu_{j;1,x}^{\text{Ind}} & \mu_{j;1,y}^{\text{Ind}} & \mu_{j;1,z}^{\text{Ind}} & \mu_{j;2,x}^{\text{Ind}} & \mu_{j;2,y}^{\text{Ind}} & \cdots\end{pmatrix}\\ \mathbf{F} &= \begin{pmatrix}F_{1,x} & F_{1,y} & F_{1,z} & F_{2,x} & F_{2,y} & \cdots\end{pmatrix}\end{aligned} \tag{7}$$

In the case when the system has $N$ polarizable centers, $\mathbf{F}$ and $\mathbf{a}_j$ have $3N$ elements and the dot product '·' in Eq. (6) is defined as $\mathbf{a}_j\cdot\mathbf{F} = \sum_a^N\left\{[\mathbf{a}_j]_{ax}F_{a,x} + [\mathbf{a}_j]_{ay}F_{a,y} + [\mathbf{a}_j]_{az}F_{a,z}\right\}$ with the index *a* running over all the polarizable centers – note that the vibrational solvatochromic dipole and electric field in Eqs. (1) and (2) are vectors with three elements.

The matrix $\mathbf{A}_j$ in Eq. (6) is also determined by the distributed vibrational solvatochromic polarizabilities. In Ref.[24], we showed that

$$\mathbf{a}_j = -\frac{1}{2M_{j,0}\omega_{j,0}}\sum_i\frac{g_{ijj}}{M_{i,0}\omega_{i,0}^2}\left(\frac{\partial\mathbf{F}^T}{\partial Q_i}\right)_{\mathbf{Q}_0}\cdot\left[\mathbf{D}^{-1} + (\mathbf{D}^{-1})^T\right] \tag{8}$$

and

$$\mathbf{A}_j = \frac{1}{2M_{j,0}\omega_{j,0}}\mathbf{D}^{-1}\cdot\left[\sum_i\frac{g_{ijj}}{M_{i,0}\omega_{i,0}^2}\left(\frac{\partial\mathbf{D}}{\partial Q_i}\right)_{\mathbf{Q}_0}\right]\cdot\mathbf{D}^{-1} \tag{9}$$

Here, the $\mathbf{D}$ matrix is defined as

$$\mathbf{D}_{ab} = \begin{cases}\boldsymbol{\alpha}_a^{-1}(0)\delta_{ab} & \text{if } a, b \text{ belong to the same molecule}\\ -\mathbf{T}_{ab} & \text{if } a, b \text{ belong to different molecules}\end{cases} \tag{10}$$

and the derivatives of $\mathbf{D}$ matrix with respect to the normal coordinates are analogously given as

$$\left(\frac{\partial\mathbf{D}_{ab}}{\partial Q_i}\right)_{\mathbf{Q}_0} = \begin{cases}-\boldsymbol{\alpha}_a^{-1}(0)\left(\dfrac{\partial\boldsymbol{\alpha}_a(0)}{\partial Q_i}\right)_{\mathbf{Q}_0}\boldsymbol{\alpha}_a^{-1}(0)\delta_{ab}\\ -\left(\dfrac{\partial\mathbf{T}_{ab}}{\partial Q_i}\right)_{\mathbf{Q}_0}\end{cases} \tag{11}$$

In the above equation, $\mathbf{T}_{ab}$ is the dipole-dipole interaction tensor. The expression given here for the induction contribution to the vibrational frequency shift is quite general enough to take into account the many-body polarization-induced electric fields.[24, 44] For the sake of notational simplicity, we additionally define $\mathbf{a}'_j = \mathbf{a}_j + \mathbf{F}^T\cdot\mathbf{A}_j$, so that the induction term can be written as

$$\Delta\omega_j^{\text{Ind}} = -\frac{1}{2}\mathbf{a}'_j\cdot\mathbf{F} \tag{12}$$

which is identical to Eq. (32) in Ref.[24]

The dispersion contribution to the vibrational frequency shift was recently found to be[20]

$$\begin{aligned}\Delta\omega_j^{\text{Disp}} \approx &\frac{1}{4\pi M_{j,0}\omega_{j,0}}\sum_i\frac{g_{ijj}}{M_{i,0}\omega_{i,0}^2}\\ &\times\sum_{a\in\text{A}}\sum_{b\in\text{B}}\sum_{\alpha\beta\gamma\delta}^{x,y,z}\Bigg\{T_{\alpha\beta}^{ab}T_{\gamma\delta}^{ab}\int_0^\infty\left(\frac{\partial\alpha_{\alpha\gamma}^a(i\Omega)}{\partial Q_i}\right)_{\mathbf{Q}_0}\alpha_{\beta\delta}^b(i\Omega)d\,\Omega\\ &+\left(\frac{\partial\left[T_{\alpha\beta}^{ab}T_{\gamma\delta}^{ab}\right]}{\partial Q_i}\right)_{\mathbf{Q}_0}\int_0^\infty\alpha_{\alpha\gamma}^a(i\Omega)\alpha_{\beta\delta}^b(i\Omega)d\,\Omega\Bigg\}\end{aligned} \tag{13}$$

In the above calculation, one needs to use the imaginary frequency-dependent (dynamic) polarizabilities, whereas the induction term in Eq. (12) requires the static polarizabilities.

In summary, the electric field-dependent frequency shift is given as the sum of the two contributions

View Article Online
DOI: 10.1039/C6CP01578F

$$\Delta\omega_j^{\text{Electric}} = \Delta\omega_j^{\text{Coul}} + \Delta\omega_j^{\text{Ind}} \tag{14}$$

The total frequency shift is therefore given as

$$\Delta\omega_j = \Delta\omega_j^{\text{Electric}} + \Delta\omega_j^{\text{Ex-Rep}} + \Delta\omega_j^{\text{Disp}} \tag{15}$$

Note that, while $\Delta\omega_j^{\text{Electric}}$ can provide information on local electric field around the IR probe, the frequency shifts induced by exchange-repulsion and dispersion interactions, which are the second and third terms in Eq. (15), are not directly related to the solvent electric field produced by the distributed charges of surrounding solvent molecules. This is the major source making the vibrational solvatochromism of a variety of nitrile-derivatized IR probes deviate from the prediction with Stark-dipole theory.

In this work, we make use of yet another way to partition frequency shifts, which was referred to as the vibrational solvatochromism theory[20, 24] based on the hybrid variational-perturbational interaction energy decomposition scheme (EDS)[45-48], or shortly SolEDS. This method was found to be very accurate, but due to its huge computational cost, it could only be used to describe vibrational frequency shifts of molecular dimers. Nevertheless, the SolEDS results could be useful to quantitatively test the chemical accuracy of the SolEFP method. Depending on the level of theory [either Hartree-Fock (HF) or second-order Møller-Plesset perturbation theory (MP2)], the SolEDS frequency shift is given by a sum of different terms as

$$\Delta\omega_j^{\text{HF}} = \Delta\omega_{j,\text{el}}^{(10)} + \Delta\omega_{j,\text{ex}}^{\text{HL}} + \Delta\omega_{j,\text{del}}^{\text{HF}} \tag{16}$$

or

$$\Delta\omega_j^{\text{MP2}} = \Delta\omega_j^{\text{HF}} + \Delta\omega_{j,\text{disp}}^{(20)} + \Delta\omega_{j,\text{el,r}}^{(12)} + \Delta\omega_{j,\text{ex}}^{(2)} \tag{17}$$

where the Hartree-Fock frequency shift consists of three contributions: $\Delta\omega_{j,\text{el}}^{(10)}$ represents the contribution from the Coulomb interaction between unperturbed solute and solvent charge densities (i.e., without including charge redistribution due to the induction effects), $\Delta\omega_{j,\text{ex}}^{\text{HL}}$ is the exchange-repulsion contribution, and $\Delta\omega_{j,\text{del}}^{\text{HF}}$ is the charge delocalization term. At the level of MP2, the electron correlation effects should be included in the vibrational frequency shift calculation so that the second-order dispersion term, $\Delta\omega_{j,\text{disp}}^{(20)}$ as well as the electron correlation corrections to the Coulomb and exchange-delocalization effects, denoted as $\Delta\omega_{j,\text{el,r}}^{(12)}$ and $\Delta\omega_{j,\text{ex}}^{(2)}$, respectively, contribute to the total frequency shift too. Note, however, that the calculations performed at HF and MP2 levels are based on two different vibrational potential energy surfaces so that the $\Delta\omega_j^{\text{HF}}$ part of $\Delta\omega_j^{\text{MP2}}$ in Eq. (17) quantitatively differs from $\Delta\omega_j^{\text{HF}}$ obtained at the HF level of theory, though the same notation $\Delta\omega_j^{\text{HF}}$ is used in both equations (16) and (17).

### 2.3 Extended Stark-dipole model with polarization effect

The frequency shift in Eq. (2) is approximate not only because it neglects the contributions from the interactions of quadrupole and higher-order multipole moments with electric field and other non-electrostatic (exchange-repulsion) effect, but also because it does not take into consideration the induction contribution in Eq. (6). From Eqs. (4) and (12) one could recast the frequency shift due to the interaction with an applied external electric field as

$$\Delta\omega_j^{\text{Stark}} = -\boldsymbol{\mu}_{j,0}(\mathbf{r_0})\cdot\mathbf{F}(\mathbf{r_0}) - \frac{1}{2}\mathbf{a}_j'\cdot\mathbf{F} \tag{18}$$

where the set of induced dipoles $\mathbf{a}_j'$ were given above in Sec. 2.2. Here, it should be noted that the induced dipole itself depends on the direction and strength of the electric field too so that it is not possible to re-write Eq. (18) simply as a product of Stark tuning rate and external electric field that are separately dependent on solute molecule and solvent charge distribution, respectively. However, if only the polarizability projected onto the bond axis of nitrile, which is denoted as $a_{j,\text{proj}}'$, is taken into consideration, the electric field-induced vibrational frequency shift is approximately given as

$$\Delta\omega_j^{\text{Stark}} \approx -\mu_j'\cdot F(\mathbf{r_0}) \tag{19}$$

where the modified Stark tuning rate, which includes the induction term approximately, is given as $\mu_j' = \mu_j + (1/2)\langle a_{j,\text{proj}}'\rangle$. This is the Stark-dipole theory proposed recently by Fried et al.[49] Later in Section 3, we shall present quantitative analysis result with the above approximate expression in Eq. (19).

## 3 Results and discussion

Here, we consider MeSCN as the model of all the nitrile IR probes, because the vibrational solvatochromism of MeSCN is quantitatively similar to that of MeCN (Figure 1) – note further that MeSCN probe is theoretically more convenient to study because, unlike in MeCN case[50], its CN stretch band is not contaminated by other Fermi resonance bands.

### 3.1 Vibrational properties of isolated MeSCN

First of all, let us consider the vibrational solvatochromic parameters of MeSCN in the gas phase. The vibrational solvatochromic multipole moments up to octupoles for the nitrile stretch mode of MeSCN were calculated by using SolCAMM technique proposed earlier.[42] In Table 1, the vibrational solvatochromic (molecular not distributed) dipoles and quadrupoles of MeSCN calculated with a few different methods are directly compared. Our *ab initio* calculated values of the Stark dipole of the CN stretch mode are in the range from 0.2 to 0.5 cm$^{-1}$/(MV/cm), which is close to the experimental value of MeSCN in 2-methyltetrahydrofuran glass at 74 K (0.55-0.64 au).[38] Evidently, our MP2 estimation somewhat underestimates the value of the Stark dipole (0.2 cm$^{-1}$/(MV/cm)) and method treating the electron correlation more accurately is needed (CCSD gives the result 0.31 cm$^{-1}$/(MV/cm), closer to the experimental value). This might be also a result of too small basis set used for such calculations but we did not extend it beyond 6-311++G(d,p) because of the resource limitations. Nevertheless, except for MP2 results, SolCAMM/6-311++G(d,p) method predicts the Stark tuning rate well. In Table 1, for the sake of comparison we also present the Stark tuning rates estimated by using the

View Article Online
DOI: 10.1039/C6CP01578F

**Table 1.** Vibrational solvatochromic dipole and traceless quadrupole moments (at the mid-point of CN bond) of MeSCN that were obtained with our SolMMM method. The fitting results in this table were obtained by carrying out multivariate least square fitting analyses of HF and density functional theory (B3LYP) calculations of MeSCN-water clusters.

| | **Method** | $\mu_{\text{CN str.}}$ [cm$^{-1}$/(MV/cm)] | $\Theta_{\text{CN str.}}$ [cm$^{-1}$/(MV/cm$^2$)] $\times 10^{-8}$ [e)] |
|---|---|---|---|
| **SolMMM/ 6-311++G**** | HF | 0.47 | 1.16 |
| | MP2 | 0.20 | 0.96 |
| | CCSD | 0.31 | 1.06 |
| | B3LYP | 0.43 | 1.26 |
| **Fitting** | HF[a)] | 0.54 | |
| | B3LYP[b)] | 0.43 | 1.02 |
| **Exp.**[c)] | | 0.55~0.64[d)] | - |

[a)] Ref.[40], [b)] Ref.[18], [c)] Ref.[38], [d)] Local field correction factor is assumed to be in the range of 1.1-1.3, [e)] The average value of the traceless quadrupole was estimated from the formula: $\sqrt{Q_{zz}^2+\frac{4}{3}(Q_{xy}^2+Q_{xz}^2+Q_{yz}^2)+\frac{1}{3}(Q_{xx}-Q_{yy})^2}$ . Note that in Ref.[18] and Ref.[42] the traceless quadrupole is defined according to Jackson[51] whereas we use Buckingham's convention here[52] (to convert value of traceless quadrupole element $Q_{ij}$ from Jackson's to Buckingham's convention multiply it by ½).

*distributed* vibrational solvatochromic charges that were obtained from multivariate least square fitting analyses of the HF and DFT (density functional theory) calculation results from a number of MeSCN-water clusters. They are in excellent agreement with the experimental value. Nevertheless, it should be emphasized that the present SolCAMM calculation results for the (molecular not distributed) vibrational solvatochromic dipole and quadrupole moments are fully derived from the first-principles without using any empirical or fitting procedure.

In Figure 2(a), the spatial distribution of the electric potential produced by the *vibrational solvatochromic* multipole moments (SolCAMM/CCSD/6-311++G**) is plotted (see also Figure S2 for similar maps obtained with other methods). Clearly, the distribution of this electric potential around MeSCN differs from that of the vibrational solvatochromic dipole (Figure 2(b)). Interestingly, if the vibrational solvatochromic quadrupole contribution is additionally taken into account (Figure 2(c)), the corresponding electric potential around the H-bond accepting N-atom of the nitrile changes its sign (compare Figures 2(b) with 2(c)). Note that this important change in the distribution of electric potential produced by the vibrational solvatochromic multipoles cannot be described by the Stark dipole theory at all. Thus, even within the approximation that the vibrational solvatochromic frequency shift solely results from the interactions of the vibrational solvatochromic multipoles of the solute molecule with the electric field produced by the atomic charges of surrounding solvent molecules, the simple dipolar description of the complicated vibrational solvatochromism is not acceptable.

### 3.2 H-bonding effect on nitrile stretch frequency shift

Although the H-bonding interaction-induced frequency shifts of various nitrile stretches in protic solvents have been observed, an all-encompassing theory for such phenomena has not been developed yet. To theoretically study such H-bonding effects on the nitrile frequency shift, we considered sixty seven MeSCN-water clusters that were fully optimized at HF/6-311++G** level of theory. For the sake of comparisons, we also considered fifteen MeSCN-DMSO clusters, where DMSO is not a protic solvent molecule. The SolEFP-calculated frequency shifts are then directly compared with the full HF/6-311++G** harmonic analysis results. Our SolEFP results are found to be in good agreement with the benchmark HF/6-311++G** results, despite the approximate nature of the SolEFP theory (see Section III in ESI).

Our SolEFP calculation results (Figure 3(a)) show that the exchange-repulsion contribution (blue dots) to the nitrile frequency shift is in general very large. In fact, the $\Delta\omega^{\text{Ex-Rep}}$ contributions are almost comparable to or even larger than the Coulomb (red dot) and induction (green dot) contributions. To further examine the intermolecular distance dependences of each contribution, we calculated them separately for varying H-bond distance between nitrile's N atom and water H atom (see Figure 3(b)). What is surprising to us is that the exchange repulsion contribution, $\Delta\omega^{\text{Ex-Rep}}$, is the largest among all the other contributions in the distance range around the equilibrium (optimum) H-bond distance (~2.25 Å). The electrostatic contribution, which is the sum of the Coulomb and induction terms, is rather constant in the distance range from 2.0 to 2.5 Å. To examine both the distance- and orientation-dependence of the vibrational frequency shift induced by its interaction with a water molecule, we carried out further SolEFP calculations at the level of HF/6-311++G** (Figure 4). Starting from the optimized geometry of MeSCN-$H_2O$ dimer, the intermolecular distance $R_d$ was increased with a distance interval of 0.05 Bohr along the H-bond direction. In addition, the water molecule on the plane of SCN…HOH was in-plane rotated by angle $\theta$(with 1° interval) around the center of mass of $H_2O$. Thus, we considered 25,200 such MeSCN…HOH configurations in total and the resulting SolEFP frequency shifts are plotted in Figure 4. From the distributions of the vibrational frequency shift components with respect to $R_d$ and $\theta$, one can easily note that the frequency shifting behavior of the Coulomb contribution, $\Delta\omega^{\text{Coul}}$, is highly directional, similar to the case of the amide I frequency shift of NMA by H-bonded water molecules[20]. However, the overall frequency shift near the optimum H-bonding configuration is still mainly determined by the exchange-repulsion contribution, which was not expected and differs from the vibrational solvatochromic frequency shift of the amide I mode of NMA in water reported previously by us.[20, 24] Note that, in the case of NMA, the Coulomb contribution, $\Delta\omega^{\text{Coul}}$, was found to be the largest. This clearly indicates that the origin of the nitrile frequency shift induced by the solute-solvent interaction is very different from that of all the other carbonyl stretch modes. In particular, if the nitrile experiences a strong steric (repulsive) interaction with surrounding molecules or chemical groups, most of the previous vibrational solvatochromism models based on purely electrostatic interaction-induced effects may not be quantitatively reliable.

View Article Online
DOI: 10.1039/C6CP01578F

**Table 2.** The nature of the frequency shifts of CN stretch mode in various H-bonding complexes involving MeCN or MeSCN and $H_2O$, MeOH and $CF_3CH_2OH$ (structures are shown in Figure S7 in ESI). The frequency shift partitioning was performed by using SolEDS//MP2/6-311++G** method. Bond lengths are in Bohr and frequency shifts in $cm^{-1}$. The details about SolEDS method can be found in Sec. 2.2 and Refs.[20,24]

| | MeCN | | | MeSCN | | |
|---|---|---|---|---|---|---|
| | $H_2O$ | MeOH | $CF_3CH_2OH$ | $H_2O$ | MeOH | $CF_3CH_2OH$ |
| $\boldsymbol{R}_{\textbf{H-bond}}$ | 3.9814 | 3.9570 | 3.7894 | 3.9569 | 3.9285 | 3.7550 |
| $\Delta\boldsymbol{\omega}_{\mathbf{el}}^{(10)}$ | -7.32 | -8.71 | -19.45 | -14.08 | -15.89 | -21.12 |
| $\Delta\boldsymbol{\omega}_{\mathbf{ex}}^{\mathrm{HL}}$ | +32.31 | +36.63 | +44.36 | +40.60 | +46.06 | +58.16 |
| $\Delta\boldsymbol{\omega}_{\mathbf{del}}^{\mathrm{HF}}$ | -10.49 | -11.29 | -17.97 | -14.06 | -15.35 | -23.83 |
| $\Delta\boldsymbol{\omega}_{\mathbf{disp}}^{(20)}$ | -10.63 | -13.51 | -15.83 | -12.64 | -16.00 | -20.62 |
| $\Delta\boldsymbol{\omega}_{\mathbf{el,r}}^{(12)}$ | +4.55 | +3.48 | +17.76 | +7.80 | +7.20 | +6.49 |
| $\Delta\boldsymbol{\omega}_{\mathbf{ex}}^{(2)}$ | +6.79 | +8.40 | +3.89 | +6.23 | +7.73 | +9.84 |
| $\Delta\boldsymbol{\omega}^{\mathrm{MP2}}$ | **+15.21** | **+15.00** | **+12.76** | **+13.85** | **+13.75** | **+8.92** |
| $\Delta\boldsymbol{\omega}_{\mathbf{el}}^{(10)} + \Delta\boldsymbol{\omega}_{\mathbf{el,r}}^{(12)}$ | -2.77 | -5.23 | -1.69 | -6.28 | -8.69 | -14.63 |
| $\Delta\boldsymbol{\omega}_{\mathbf{ex}}^{\mathrm{HL}} + \Delta\boldsymbol{\omega}_{\mathbf{ex}}^{(2)}$ | +39.10 | +45.03 | +48.25 | +46.83 | +53.79 | +68.00 |
| $\Delta\boldsymbol{\omega}_{\mathbf{del}}^{\mathrm{HF}} + \Delta\boldsymbol{\omega}_{\mathbf{disp}}^{(20)}$ | -21.12 | -24.80 | -33.80 | -26.70 | -31.35 | -44.45 |
| **Full QM** | **+12.83** | **+12.58** | **+8**[a)] | **+10.54** | **+10.16** | **+2.86** |

a) Partially optimized due to energy optimization problem

### 3.3 CN stretch in bulk solutions

As shown in Figure 1, the nitrile frequency shifts are strongly solvent-dependent. Therefore, we now apply our SolEFP theory to the vibrational frequency shift calculations in combination with MD simulations of MeSCN dissolved in four representative solvents. The main results are presented in Figure 5. Because the application of SolEFP theory is the main issue here, we did not attempt to re-parameterize any force field parameters of the solvent molecules. The four solvents are aprotic non-polar $CCl_4$, aprotic and weakly polar $CHCl_3$, aprotic and strongly polar DMSO, and protic polar $H_2O$. Fortunately, fairly accurate force field parameters for these solvent molecules are available so that we could use them without any further re-parameterization or re-optimization of the classical MD force fields.

The experimentally measured average nitrile frequency shifts are cyan squares in Figure 5. The total frequency shifts calculated with SolEFP and MD simulation methods are black diamonds in the same figure (see Table S1 for numerical values of the data presented in Figure 5). Although the differences between SolEFP and experimental results are found to be non-negligibly large by about 10 $cm^{-1}$, the overall trend of the vibrational frequency shift with respect to the solvent polarity and H-bonding ability is correctly described by the SolEFP model. In contrast, the Stark-dipole theory predicts even a very large red shift when MeSCN interacts with water molecules. Here, it should also be noted that the previous *ab initio* semi-empirical maps for nitrile frequency shift[18, 41] cannot describe the relative blue shift of nitrile stretch mode in water, because they were based on the assumption that the vibrational solvatochromic frequency shifts are induced by electrostatic interactions only.

As can be seen in Figure 5, the exchange-repulsion contribution, $\Delta\omega^{\text{Ex-Rep}}$, to the nitrile frequency blue shift in water is about 8 to 10 $cm^{-1}$ larger than those in aprotic solvents considered here (compare blue bars in Figure 5). To further understand the H-bond strength-dependence of the nitrile frequency shift, we performed first-principles SolEDS calculations for Me(S)CN-X dimers with X being $H_2O$, MeOH or $CF_3CH_2OH$, which are presented in Table 2 (the corresponding molecular structures are depicted in Figure S7). The exchange-repulsion interaction contributes to the very strong frequency blue shift ($\Delta\omega_{\text{ex}}^{\text{HL}}$). Furthermore, the magnitude of $\Delta\omega_{\text{ex}}^{\text{HL}}$ increases as the H-bond length of X molecule increases. This can be understood by noting that the stronger H-bond becomes the closer solute and solvent molecules are brought to each other. This in turn causes a sharp increase in the exchange-repulsion contribution to the frequency shift (see Eq. (5) and Figure 3(b)). This approximately explains all the frequency blue shifts induced by the interaction between nitrile and protic solvent molecules, when compared to aprotic solvents. It is also interesting that, after taking into account the electron correlation effects ($\Delta\omega_{\text{el,r}}^{(12)}$, $\Delta\omega_{\text{ex}}^{(2)}$ and $\Delta\omega_{\text{disp}}^{(20)}$ terms in Table 2), the Coulombic effects on the nitrile stretch mode frequency shifts in these dimers are roughly 6-10 times smaller than the non-Coulombic effects, and the red shifting contribution is thus due to the charge delocalization and dispersion interactions. For example, in the case of the MeSCN-MeOH dimer, the Coulombic frequency shift is just −8.7 $cm^{-1}$, while the shifts by exchange-repulsion effect and the sum of charge-delocalization and dispersion effects are +53.8 and −31.4 $cm^{-1}$, respectively. The relative contributions to the frequency shifts from them are similarly found in all the other dimers studied here (see Table 2). Therefore, not only the short-range repulsive interaction but also the non-Coulombic electrostatic effects like induction and dispersion are also quite important in quantitatively describing the nitrile frequency shifts in H-bonded systems. Indeed, among the red-shifting attractive interaction-induced contributions, the dispersion term is *dominant*, not just in aprotic solvents, but even in water too (Figure 5). Note also that in the case of $CCl_4$, Coulomb contribution completely vanishes, but yet the induction contribution is not negligible. In Figure 6, we plot the distributions of different frequency shift components separately with the full width at half maximum (FWHM) values

View Article Online
DOI: 10.1039/C6CP01578F

**Table 3.** Full width at half maximum (FWHM) in cm$^{-1}$ of the distribution of MeSCN CN stretch mode frequency shift. Frequency shift distributions were obtained by combining SolEFP method with MD simulation.

| | $\boldsymbol{\Gamma}_{\mathrm{Coul}}^{\mathrm{static}}$ | $\boldsymbol{\Gamma}_{\mathrm{Ex\text{-}Rep}}^{\mathrm{static}}$ | $\boldsymbol{\Gamma}_{\mathrm{Ind}}^{\mathrm{static}}$ | $\boldsymbol{\Gamma}_{\mathrm{Disp}}^{\mathrm{static}}$ | $\boldsymbol{\Gamma}_{\mathrm{SolEFP}}^{\mathrm{static}}$ |
|---|---|---|---|---|---|
| **$CCl_4$** | 4.0 ± 0.5 | 24 ± 3 | 1.4 ± 0.5 | 5.6 ± 1.0 | 21 ± 4 |
| **$CHCl_3$** | 12.0 ± 0.5 | 21 ± 3 | 2.5 ± 1.5 | 5.1 ± 1.0 | 22 ± 4 |
| **DMSO** | 16.4 ± 0.5 | 23 ± 3 | 4.8 ± 1.5 | 6.4 ± 1.0 | 25 ± 4 |
| **$H_2O$** | 18.2 ± 0.5 | 23 ± 5 | 12.9 ± 3.0 | 6.2 ± 1.5 | 27 ± 5 |

listed in Table 3. The distributions (blue curve in Figure 6) of exchange-repulsion frequency shifts are quite broad for all four solutions. As the solvent polarity increases from $CHCl_3$ to water, the red-shifting Coulomb contribution (red curve) increases. However, the dispersion contribution remains more or less the same. From the present SolEFP results shown in Figures 5 and 6, it is quite clear now that the nitrile frequency shift cannot be fully described by the contributions, $\Delta\omega^{\mathrm{Coul}}$ and $\Delta\omega^{\mathrm{Ind}}$, that are linearly dependent on electric field.

The substantial offset of our SolEFP results with respect to experimental results could be caused primarily by the difference between the theoretical models describing EFP2-EFP2 interaction potential and MD force field. The offset values are systematic and positive, which suggests that the exchange-repulsion-induced vibrational frequency shift is likely to be affected by such difference strongly. Note that this repulsion contribution is extremely sensitive to a minute change in the solute-solvent distance when they are in close proximity (Figures 3(b) and 4). In addition, our SolEFP model accounts only for the dipole-dipole distributed polarizability effects, while the dipole-quadrupole and higher-order polarizability effects were ignored. Previously, other research groups noted that the distributed dipole-dipole EFP2 dispersion interaction needs to be rescaled by 4/3 to approximately reproduce the total dispersion interaction energy at equilibrium solute-solvent geometry.[53, 54] Therefore, any deviation induced by slightly less accurate interaction potential models could cause a large error in estimating repulsion-induced frequency shift. Secondly, noting that the dispersion interaction with dipole-dipole polarizability terms included induces a very large frequency red shift, we believe that the distributed dipole-quadrupole dispersion interactions could also contribute to the vibrational frequency shift. Other plausible sources of error could originate from either our neglect of the vibrational second derivatives of the solute-solvent interaction potential when the non-Coulombic interaction-induced frequency shifts were calculated or insufficient treatment of electron correlation effects beyond the intermolecular dispersion.

Nevertheless, our analysis is consistent with the previous empirical calculation of the CN stretch mode of acetonitrile[12], in which Reimers and Hall found, by applying the solvent descriptor model of Fawcett et al.[55], that the dispersion interaction contributes the most to the observed frequency red shifts. They also found a drastic increase in ‘specific’ frequency blue shifts as they were using more protic solvents like water, 2,2,2-trifluoroethanol and trifluoroacetic acid. For instance, the non-specific electrostatic, dispersion and specific (electrostatic and non-electrostatic) frequency shifts of CN mode of acetonitrile in water were found to be −6.4, −9.7 and +6.7 cm$^{-1}$, whereas in $CCl_4$ −1.9, −13.0 and 0.0 cm$^{-1}$, respectively. Therefore, it is evident that those short-range blue-shifting effects originate from the exchange-repulsion interactions that were previously analyzed by Rey and Hynes[15] and Morales and Thompson[16] from a perspective of a repulsive part of the van der Waals intermolecular potential.

Before we apply the present SolEFP model to the vibrational frequency shift calculations of nitriles incorporated into proteins, it is necessary to point out the limitation of the extended Stark theory discussed at the end of Section 2. Since the SolEFP approach is the first-principles theory that includes all the vibrational solvatochromic induced multipole moments, we could test the validity of the extended but still approximate expression for the Stark-dipole model in Eq. (19). The evaluation of the solvatochromic induced dipole moment is straightforward within the SolEFP theory:

$$\boldsymbol{\mu}_j^{\mathrm{Ind}} = \sum_{a}^{\substack{\text{Polarizable}\\ \text{sites}}} \boldsymbol{\mu}_{j,a}^{\mathrm{Ind}} \tag{20}$$

where the summation runs over all polarizable sites in the system (including solvent sites). Note that the solvatochromic induced dipole moment is independent on the origin of the coordinate system (which is also true for ions). Furthermore, it is possible to evaluate the molecular (not distributed) solvatochromic quadrupole moment as well:

$$\boldsymbol{\Theta}_j^{\mathrm{Ind}} = -\sum_{a}^{\substack{\text{Polarizable}\\ \text{sites}}} \left\{\boldsymbol{\mu}_{j,a}^{\mathrm{Ind}} \otimes (\mathbf{r}_{\mathrm{CN}} - \mathbf{r}_a) + (\mathbf{r}_{\mathrm{CN}} - \mathbf{r}_a) \otimes \boldsymbol{\mu}_{j,a}^{\mathrm{Ind}}\right\} \tag{21}$$

where $\mathbf{r}_{\mathrm{CN}}$ is the position vector of the center of CN bond.

**Table 4.** Frequency shifts originating from the Coulomb interaction and induction (polarization) contributions of CN stretch mode of MeSCN in four different solvents. Here, the exact SolEFP induction frequency shift is compared with the approximate schemes (see the main text for explanation). The vibrational solvatochromic dipole and quadrupole magnitudes are presented in the last two columns. $\mathbf{F}$ denotes the set of electric fields at 19 distributed sites of MeSCN whereas $\mathbf{F}_{\mathrm{CN}}$ is the electric field at the CN mid-bond. Frequency shifts are given in cm$^{-1}$ and solvatochromic dipoles and quadrupoles in cm$^{-1}$/(MV/cm) and $10^{-8}$ × cm$^{-1}$/(MV/cm$^2$), respectively.

| | **Frequency shifts** | | | | **Molecular solvatochromic multipoles** | |
|---|---|---|---|---|---|---|
| | $\Delta\omega_{\mathrm{CN\,str.}}^{\mathrm{Coul}}$ | $\Delta\omega_{\mathrm{CN\,str.}}^{\mathrm{Ind}}$ | $\langle -\frac{1}{2}\langle\boldsymbol{\mu}_{\mathrm{CN\,str.}}^{\mathrm{Ind}}\rangle \cdot \mathbf{F}_{\mathrm{CN}}\rangle$ | $\langle -\frac{1}{2}\langle\mathbf{a}_{\mathrm{CN\,str.}}^{'}\rangle \cdot \mathbf{F}\rangle$ | $\lvert\langle\boldsymbol{\mu}_{\mathrm{CN\,str.}}\rangle\rvert$ | $\lvert\langle\boldsymbol{\Theta}_{\mathrm{CN\,str.}}\rangle\rvert$ |
| **Vacuum** | 0.0 | 0.0 | 0.0 | 0.0 | 0.466 | 1.16 |
| **$CCl_4$** | -0.7 | -3.3 | 0.0 | 0.0 | 0.436 | 1.05 |
| **$CHCl_3$** | -5.6 | -3.8 | 0.0 | -0.6 | 0.477 | 1.05 |
| **DMSO** | -10.6 | -5.2 | -0.2 | -1.6 | 0.506 | 1.07 |
| **$H_2O$** | -18.5 | -10.1 | -3.2 | -4.3 | 0.639 | 1.04 |

View Article Online
DOI: 10.1039/C6CP01578F

In Table 4, we compare the exact induction frequency shifts that can be calculated with Eq. (12) with (i) those obtained by considering time-averaged vibrational solvatochromic induced dipole in Eq. (20) interacting with the instantaneous electric field evaluated at the CN mid-bond, i.e., $\Delta\omega^{\mathrm{Ind}}_{\mathrm{CN\,str.}} \approx \langle -\frac{1}{2}\langle \boldsymbol{\mu}^{\mathrm{Ind}}_{\mathrm{CN\,str.}}\rangle \cdot \mathbf{F}_{\mathrm{CN}}\rangle$ and (ii) those obtained by considering time-averaged vibrational solvatochromic induced dipoles distributed over 19 different locations in MeSCN molecule that interact with the instantaneous electric fields at those polarizable sites, i.e., $\Delta\omega^{\mathrm{Ind}}_{\mathrm{CN\,str.}} \approx \langle -\frac{1}{2}\langle \mathbf{a}'_{\mathrm{CN\,str.}}\rangle \cdot \mathbf{F}\rangle$ (see Eqs.(7), (8), and (12)). In addition, the total magnitudes of the solvatochromic dipole and quadrupole moments, as well as the Coulomb frequency shifts, are also given in Table 4.

As can be seen in Table 4, it is concluded that the exact induction frequency shift ($\Delta\omega^{\mathrm{Ind}}_{\mathrm{CN\,str.}}$ in Table 4) cannot be reproduced by the two approximate calculation methods (i) and (ii) discussed above. To examine the directionality of the vibrational solvatochromic induced dipoles, we examined the relative orientation of the total vibrational solvatochromic induced dipole with respect to the permanent vibrational solvatochromic dipole of the nitrile stretch mode (Figure S8 in ESI). It turns out that the relative angle between the two dipoles varies from 0 to 180° in all the aprotic solvents. This means that the induction frequency shifts are determined by instantaneous solute-solvent configurations so that the averaged value over all possible solute-solvent configurations cannot provide information on local electric field [note that the total magnitudes of the molecular vibrational solvatochromic dipoles and quadrupoles do not depend on solvent much (see Table 4)]. However, in the case of water, the vector component of the average induced vibrational solvatochromic dipole is large and it induces small red shift (Table 4). This indicates that the extended Stark-dipole theory including both the molecular vibrational solvatochromic dipole and solvent-induced dipole contributions[49] is still incapable of fully describing even the induction effects correctly in the cases of nitrile probes. To prove this quantitatively, for both a small cluster consisting of one MeSCN and three water molecules (Figure 7(a)) and a large cluster taken from MD trajectory, we calculated the instantaneous distributed vibrational solvatochromic dipoles ($\mathbf{a}'_j$ in Eq. (12)) and depict them with arrows. Note that quite a few dipoles at different sites contribute to $\Delta\omega^{\mathrm{Ind}}_{\mathrm{CN\,str.}}$, where they all have quite different directions and magnitudes. If they all are added up at the mid-point of the CN bond, they would largely cancel out with one another. Furthermore, they are spatially delocalized even on surrounding water molecules so that $\Delta\omega^{\mathrm{Ind}}_{\mathrm{CN\,str.}}$ cannot be simply estimated by considering the induced dipoles of MeSCN only. We emphasize that the SolEFP induction theory fully takes into account such non-additive nature in calculating $\Delta\omega^{\mathrm{Ind}}$ and that $\mathbf{a}'_j$ *spans the entire system* including both solute and solvent molecules.[24] Therefore, it becomes clear that the Stark-dipole model is a highly approximate approach to predict the frequency shifts of nitrile stretch mode not only because it does not include other important contributions from vibrational solvatochromic quadrupolar, exchange-repulsion and dispersion effects, but also because it does not take into account even the induction effect correctly.

### 3.4 SCN probe embedded within protein framework

In the above subsections, we showed that the vibrational solvatochromism of CN probes originates not only from the electrostatic interaction with solvent electric field but also from other short-range effects that *cannot* be directly correlated with solvent electric field. In particular, the exchange-repulsion contribution caused by the steric contacts of the IR probe with surrounding molecules or chemical groups causes a strong blue shift of the nitrile stretch modes and the dispersion interaction-induced term makes it red-shifted. Such delicate balance among many different contributions to the vibrational frequency shift is particularly important when those IR probes are incorporated into proteins. To study these aspects in detail, one needs to construct a rigorous vibrational solvatochromism model for protein environment, which properly takes into account Coulomb, induction, dispersion, and exchange-repulsion interaction-induced terms systematically. Here, we make the first attempt to develop such *ab initio* model.

**Constructing the model for a protein environment.** This task is extremely challenging because, unlike the simple cases of MeSCN in various bulk solutions, we need to consider vibrational probe which is covalently attached to the macromolecule (denoted as Prot-SCN) – note that there is no well-defined theoretical procedure to decouple any IR probe vibration from the protein backbone. However, we believe that the vibrational solvatochromism of any Prot-SCN probe can still be modeled, to a reasonable level of accuracy, by using the model MeSCN molecule because CN stretch modes are fairly localized and isolated from other intramolecular protein vibrations.

To test the transferability of the *ab initio*-calculated SolEFP solvatochromic parameters we studied a few SCN probe systems embedded in model amino acids of which structures can be found in Figure S9 in ESI. We assume that the vibrational frequency shift induced by neighboring peptide groups could be described by the following two factors: (i) Coulombic frequency shift due to the electric field by the distributed charges on the amide group (CONH); and (ii) through-bond effects that are associated with frequency shift originating from change of relative configuration of covalently bonded chemical group such as methylene. The first effect can be modeled by considering the interaction (Eq. (3)) of MeSCN vibrational solvatochromic multipoles with atomic partial charges of CONH group. The second effect (ii) cannot be directly quantified, but we could approximately estimate its nature by performing additional analyses of the CN stretch mode frequencies of several MeSCN analogs (Figure 8(a)), where the relative configuration of neighboring methylene (or methyl) groups are different from one another. If one more methyl (methylene) group is added to the MeSCN, the CN stretch mode frequency is red-shifted by about 3 $\mathrm{cm}^{-1}$. Thus, the overall frequency shift of Prot-SCN due to the local side-chain configuration of Pro-SCN such as proteins with

View Article Online
DOI: 10.1039/C6CP01578F

cyanylated cysteine residues and the two neighboring amide groups of the cyanylated cysteine can be approximately written as

$$\Delta\omega_{\text{Pept-SCN}} \approx \Delta\omega^{\text{Coul}}_{\text{MeSCN-CONH}} + \Delta\omega_{\text{Through-bond}} \tag{22}$$

where $\Delta\omega^{\text{Coul}}_{\text{MeSCN-CONH}}$ is evaluated from the interaction of SolCAMM with ESP charges fitted on *N*-methylacetamide (NMA) peptide group (see Sec. IX in ESI for details). Using the above scheme, we could calculate the vibrational frequency shifts and compared them with *ab initio* calculation results in Figure 8(b). The correlation is excellent. The constant offset, which is roughly equal to −3.1 cm$^{-1}$, can indeed be attributed to the through-bond (or additional methylene group) effect, which is fully consistent with the through-bond effect (shifting the frequency by −2.8 cm$^{-1}$ on average) traced in Figure 8(a) when the methyl group of MeSCN is replaced with ethyl or even longer alkyl chain. The results here show that the SolEFP parameters extracted from the *ab initio* calculations of MeSCN can indeed be used to describe the vibrational solvatochromism of any Prot-SCN's.

**Table 5.** The minimalistic model of amino acid side chains that is used in Eq. (23). In this model the protein is mimicked by a set of small EFP2 fragments that capture the fundamental physical components of the interaction between the IR probe and molecular environment. Peptide units are treated by NMA molecules except those being attached next to SCN probes (see the main text).

| **Amino acid side-chain** | **EFP fragments** | **Model fragment** |
|---|---|---|
| **Asp** | · $CH_3COO^-$ | · Carboxyl group (deprotonated) |
| **Asn** | · $CHONH_2$<br>· $CH_4$ | · Amide group<br>· $C_\beta H_2$ |
| **Glu** | · $CH_3COO^-$<br>· $CH_4$ | · Carboxyl group (deprotonated)<br>· $C_\beta H_2$ |
| **Gln** | · $CHONH_2$<br>· 2 x $CH_4$ | · Amide group<br>· $C_\beta H_2$, $C_\gamma H_2$ |
| **Ser** | · $CH_3OH$ | · C-OH group |
| **Thr** | · $CH_3OH$<br>· $CH_4$ | · C-OH group<br>· $C_\beta H_2$ |
| **Tyr** | · PhOH<br>· $CH_4$ | · Ar-OH group<br>· $C_\beta H_2$ |
| **Phe** | · Benzene<br>· $CH_4$ | · Aromatic ring<br>· $C_\beta H_2$ |
| **Lys** | · $CH_3NH_3^+$<br>· 3 x $CH_4$ | · Amine group (protonated)<br>· $C_\beta H_2$, $C_\gamma H_2$, $C_\delta H_2$ |
| **Arg** | · Methylguanidinium $^+$<br>· 2 x $CH_4$ | · Guanidinium group (protonated)<br>· $C_\beta H_2$, $C_\gamma H_2$ |
| **Gly** | - | - |
| **His** | · Imidazole<br>· $CH_4$ | · Aromatic ring (N is not protonated)<br>· $C_\beta H_2$ |
| **Trp** | · Indole<br>· $CH_4$ | · Indole group<br>· $C_\beta H_2$ |
| **Met** | · $(CH_3)_2S$<br>· $CH_4$ | · Methyl sulfide group<br>· $C_\beta H_2$ |
| **Cys** | · $CH_3SH$ | · Thiol group |
| **Pro** | · 3 x $CH_4$ | · Ring |
| **Ala** | · $CH_4$ | · $C_\beta H_2$ |
| **Leu** | · 3 x $CH_4$ | · *sec*-buthyl chain |
| **Ile** | · 3 x $CH_4$ | · 2-methyl-*n*-propyl chain |
| **Val** | · 2 x $CH_4$ | · *iso*-propyl chain |

The next and perhaps the more difficult task is to model the vibrational solvatochromic influences of surrounding amino-acid side chains in protein on the CN frequency of Prot-SCN. To use our SolEFP approach, a properly chosen set of building blocks that faithfully mimic the amino-acid side chains as well as the remaining peptide groups are need. Here we use the following procedure: i) aliphatic group of side chain is represented by an appropriately superimposed methane; ii) aromatic side chain of phenylalanine is by benzene, iii) histidine side chain is modeled by imidazole; iv) polar groups can be modeled by the methylated group (for example MeOH is the model for the serine side chain); and v) peptide bond is modeled by NMA. In Table 5, the model chemical groups representing twenty amino acids are summarized and we obtained the EFP2's of those model compounds. One of the problems originates from the overlap between an EFP2 fragment and an NMA modelling an amide group. Here, we assume that the 'spurious' atoms introduced by all of the EFP2's do not contribute much to the vibrational frequency shift. Then, we adopt the following strategy:

1) The SolEFP pairwise-additive frequency shifts originating from Coulomb, dispersion, and exchange-repulsion interactions between model amino acid side chains and nitrile are evaluated by using the standard SolEFP method described here.
2) The (polarization) induction effects are included as follows: the vibrational frequency shifts due to the interaction of nitrile with all the overlapping fragments such as $CH_4$ and NMA are considered separately. The remaining fragments including polar fragments as well as solvent water molecules are considered by taking into account the many-body induction effects.

This scheme is formulated as

$$\Delta\omega^{\text{SolEFP}} \approx \left( \sum_{X}^{\text{all EFP2's}} \Delta\omega^{\text{Coul}}_{\text{Solute-}X} + \Delta\omega^{\text{Ex-Rep}}_{\text{Solute-}X} + \Delta\omega^{\text{Disp}}_{\text{Solute-}X} \right) + \Delta\omega^{\text{Ind}} \tag{23}$$

where

$$\Delta\omega^{\text{Ind}} \approx \Delta\omega^{\text{Ind}}_{\text{Polar EFP2's}} + \sum_{X}^{\substack{\text{remaining}\\ \text{EFP2's}}} \Delta\omega^{\text{Ind}}_{\text{Solute-}X} \tag{24}$$

To test the validity of the above approach, Eqs. (23) and (24), we specifically considered five dimeric complexes consisting of one MeSCN and serine, tyrosine, or lysine model peptide (see Figure S10 in ESI). First, we carried out *ab initio* harmonic vibrational analyses of those complexes to obtain the CN

View Article Online
DOI: 10.1039/C6CP01578F

frequency shifts. Then, using the fragment approach with the SolEFP parameters of MeSCN and the EFP2's representing the serine, tyrosine, and lysine model peptides (see the details in Table S2 in ESI). Surprisingly, the fragment approach reproduces both $\Delta\omega^{\mathrm{Disp}}$ and $\Delta\omega^{\mathrm{Ex\text{-}Rep}}$ very well. Our Eq. (24) however underestimates the induction effect (compare "Full" and "Frag" entries in Table S2). Perhaps this is because the cooperative induction effects by bonding orbital electrons of two different fragments were ignored, which is the problem that cannot be easily resolved and is beyond the scope of the present work. Nonetheless, our SolEDS/HF frequency shifts are in excellent agreement with full *ab initio* results with the errors of just ~1 cm$^{-1}$ or lesser. When dispersion is not included in SolEFP, those analysis results can be compared to SolEDS/HF results. Overall, the SolEFP, as compared to the SolEDS, underestimates the frequency shifts by about 4 to 10 cm$^{-1}$. This is mainly due to the approximate nature of the SolEFP method in taking into account the non-Coulombic interactions[20, 24] (see also ESI, Sec. III). Nevertheless, Eq. (23) works reasonably well, though the present approach is highly simple and approximate in nature. What is important here is that the short-range repulsion and dispersion effects, which are often the main contributions to the nitrile frequency shifts in solutions, are at least correctly described by the present fragment SolEFP model for vibrational solvatochromism of nitrile IR probes in protein environments.

**Short-range frequency shifts of CN stretch in Ras-binding domain of RalGDS.** To get the detailed insight into the origin of the short-range interaction-induced vibrational frequency shifts of CN probes embedded in proteins, we have specifically studied the Ral guanine nucleotide dissociation stimulator (RalGDS), which is a downstream effector of Rap1A (Rap). The Ras-binding domain of RalGDS was systematically mutated at 6 different positions to introduce an SCN probe (Figure 9).[56] We have performed the umbrella sampling molecular dynamics simulations of the 6 mutants of RalGDS in water and the same mutants docked to human oncoprotein p21$^{\mathrm{Ras}}$ (Ras) mutant (here labeled as Ras'; see Computations). Subsequently, from the MD trajectories, we have considered the local environments within 10 Å from SCN probes (see Figure S11 for example), where the most probable configurations were taken into consideration based on the weighted histogram analysis method. In total, about 1000 Prot-SCN configurations were taken from the MD trajectories for our SolEFP fragment analyses. The average values of the short-range frequency shifts of SCN stretch modes in RalGDS are given in Figure S12, whereas those in RalGDS-Ras' complexes are presented in Figure 10.

In the cases of RalGDS proteins without binding to Ras' (Figure S12), all the SCN probes are highly exposed to water so that the contributions from protein environment to the SCN stretch frequency shifts are relatively small and mainly are Coulombic in nature. $\Delta\omega^{\mathrm{Ex\text{-}Rep}}$ due to solvating water molecules is very large, about 35 cm$^{-1}$. This is even larger than that of MeSCN in water, which is about 26 cm$^{-1}$ (see Figure 5) – note that this difference is likely to be caused by the differences in the force field parameters (of SCN group and water) that were used to simulate bulk solutions in Sec. 3.3 and the protein systems in water. The induction and dispersion effects are generally similar to those in bulk water and the Coulomb contribution varies from case to case, because the SCN probes have quite different orientations with respect to water phase – note that $\Delta\omega^{\mathrm{Coul}}$ is strongly dependent on the angle between SCN and H-bonded water molecule.

The situation is much more complicated when SCN probe is located at the interface of RalGDS and Ras' proteins (Figure 10). The calculated $\Delta\omega^{\mathrm{Ex\text{-}Rep}}$ values are still quite large, which range from +24 cm$^{-1}$ in RalGDS/G28+Ras' to +40 cm$^{-1}$ in RalGDS/Y31C+Ras'. Furthermore, the blue shifts have quite different origins depending on the specific position of the SCN probe. Non-bonding close-contact of SCN with aliphatic or aromatic part of amino acid side chain (Val, Leu, Ile or Phe) is the dominant interaction contributing to $\Delta\omega^{\mathrm{Ex\text{-}Rep}}$ in the case of G28 (see also a model MeSCN-$CH_4$ distance-dependence scan in Figure S13 in ESI). On the other hand, in the cases of K32 and N29, the interaction of SCN with protein backbone induces a substantial blue shift, where the H-bond between SCN and amide proton and the repulsive interaction between peptide-oxygen and SCN-nitrogen atoms are important. In the case of Y31, SCN probe is located in the vicinity of Ser residues that form strong hydrogen bonds through OH groups with SCN and induce a substantial blue shift. It is interesting to note here that when Ras' protein is not docked to RalGDS, SCN probe does not make such H-bonding interaction with Ser residue because, in that case, the probe prefers to interact with water rather than protein. This results in a lack of the exchange-repulsion blue shift due to Ser. Also, in the case of RalGDS/G28 (without docked protein), SCN interacts with Tyr residue via making an H-bond with OH group of Tyr. However, typical H-bond lengths, estimated from the MD trajectory, are longer than those of RalGDS/Y31+Ras' so that the exchange-repulsion interaction-induced frequency shift due to the tyrosine OH group is small (note that $\Delta\omega^{\mathrm{Ex\text{-}Rep}}$ is a very short-range effect and becomes vanishingly small as solute-solvent distance increases; Figure 3(b)). Finally, the SCN probe in the case of N54 can form H-bonding interactions with water molecules penetrated into the protein-protein interface, which induces a strong blue shift. This is an exceptional position at the RalGDS protein that is likely to be exposed to water in both free and bound states of RalGDS, whereas the SCN group at the other positions does not form an H-bonding interaction with water molecules when RalGDS binds to the Ras mutant studied here.

Now, it is clear that both water and protein residues can cause a large blue shift of SCN stretch frequency due to their exchange-repulsion interactions with SCN. Another distinctive feature found here is that the dispersion effects are generally larger at the protein interface (by about 5 to 10 cm$^{-1}$), when compared with those in the cases of free RalGDS proteins. This is mainly because many-electron hydrocarbon groups including aromatic rings are close to the SCN probe in the cases of RalGDS-Ras complexes. Unfortunately, our calculations of the SCN frequency shifts in RalGDS proteins cannot be directly compared with experimental results previously reported in Ref.[56] since our Prot-SCN MD simulation systems considered in

View Article Online
DOI: 10.1039/C6CP01578F

this work are still too small because of the short cut-off distance of 10 Å used in the present Coulomb-induced frequency shift calculation. Furthermore, it should be noted that the experimentally measured frequency shifts differ from one another by at most 3 $cm^{-1}$ – note that none of the existing vibrational solvatochromic maps have this chemical accuracy yet. Therefore we do not aim to reproduce those relative frequency shifts in this work. Rather, our qualitative analysis results presented here shows the importance of the exchange-repulsion and the dispersion interaction effects, which cannot be described with other electric field-based theories previously used. Our results suggest also that the magnitudes of $\Delta\omega^{\text{Ex-Rep}}$ at the protein surface are larger than those in bulk solutions. Interestingly, the repulsion effects are quantitatively similar in both aqueous and protein environments (note that $\Delta\omega^{\text{Ex-Rep}}$ of CN stretch mode of MeSCN in water is higher by about 7 $cm^{-1}$ than that in DMSO, $CHCl_3$ or $CCl_4$). The spatial confinement effects at the interface between RalGDS and Ras' proteins is of particular importance because the environment around SCN is crowded and congested, which results in a strong repulsion interaction. Only in the case of RalGDS/G28C+Ras', the repulsion frequency shift is smaller than 30 $cm^{-1}$. From this, it is shown that the frequency blue-shift due to the H-bonding interactions of SCN with water molecules cannot be distinguished from that to the repulsive interactions of SCN with protein residues.

We have to mark here that the amount of protein samples taken from MD is very little at a present time (roughly 80-100 MD snapshots per system) so that a further study along this line is necessary. Still, we believe that most of the probable configurations based on the weighted histogram analysis method were taken into consideration in the present calculation study and that most of the salient features about the distribution of the frequency shifts of CN IR probe embedded in a highly heterogeneous protein environment were captured here. Our fragmentation method would be useful in the cases when an IR probe forms H-bonds with water, peptide, and amino-acid side chains or is in close contact with protein residues (such as molecular crowding-induced steric interaction, for example). However, due to the intrinsic limitation of fragmentation approach to the calculation of vibrational frequency shift, our estimated deviations are in a range from 3 to 7 $cm^{-1}$ based on Table S2. We thus need to refine the approach to make it to the spectroscopic accuracy in the future. Nevertheless, it will be desirable to carry out further works for comparatively small proteins to test the validity of the SolEFP approach developed here.

One of the related works was reported by Zou et al.[57] Carrying out MD simulations, they obtained van der Waals and electrostatic forces exerting on the CN group of benzonitrile embedded at four different sites in a model ion channel protein to elucidate the origins of the CN frequency change caused by the binding of halothane molecule. They found that electrostatic forces are roughly 2-2.5 times larger in magnitude than van der Waals forces and they used Stark dipole model for further analyses. The agreement with average experimental value of the frequency shift (about 6 $cm^{-1}$) was found to be quite good (roughly 3 and 6 $cm^{-1}$ obtained from two separate MD simulations). However, their approach does not provide all the contributions to the vibrational frequency shifts separately. In fact, their analysis of the van der Waals forces included not only blue-shifting exchange-repulsion but also red-shifting dispersion effect that have opposite signs and very different distance-dependences [dispersion is a relatively long-range effect as compared to exchange-repulsion; see Fig.S2-S5]. Therefore, the actual importance of the non-Coulombic frequency shifts was not traced fully in the study of Zou et al.[57] due to the cancellation of attractive and repulsive contributions that result in the frequency shifts of different nature.

Recently, Zhang et al. have proposed a new IR probe based on 5-cyanoindole that was found to be a good candidate for monitoring the local dynamics in proteins, especially because of its sensitivity to the molecular environment and notably long vibrational lifetime (12.3 ps for the ground state bleach decay).[26] However, they showed that the vibrational response of their probe is governed by the local electrostatic potential around CN group, i.e., intermolecular H-bonds with CN group and intramolecular H-bond of CN with indole's NH group. This cannot be easily modeled by the Stark dipole theory. Moreover, it is expected that even the contemporary (empirical or first-principles) solvatochromic maps based on the distributed site multipoles[18, 41, 42] might have faced certain problems in quantitatively describing the vibrational frequency shifts of such IR probes due to the complicated nature of H-bonding interaction in general. We believe that the SolEFP model along with its protein-model extension presented in this work for the first time is a promising first step toward developing more refined and chemically accurate model capable of capturing all the vibrational solvatochromic effects from many-body intermolecular interactions, especially when an IR probe is surrounded by highly heterogeneous molecular environments like proteins.

## 4 Conclusions

In this study, we have analyzed the solvatochromic vibrational frequency shifts of nitrile stretch mode in a variety of different molecular environments by using the first principles SolEFP model, which takes into account the Coulomb, induction, dispersion and exchange-repulsion contributions. Our model is capable of reproducing the relative experimental vibrational frequency shifts of CN stretch of MeSCN dissolved in $CCl_4$, $CHCl_3$, DMSO and $H_2O$, which proves that this approach is better than previous ones based on Coulombic interaction only including the widely used Stark-dipole theory. Indeed, the electric-field-based theories cannot properly describe the CN vibrational frequency shift induced by H-bonding interaction of water molecules because they ignored the exchange-repulsion effect that causes a strong frequency blue shift.

View Article Online
DOI: 10.1039/C6CP01578F

In addition, we here showed that the dispersion interaction between polarizable CN group and polarizable solvent molecules is yet another important red-shifting contribution to the frequency shift. Our extended vibrational solvatochromism theory was used to elucidate the origin of CN frequency shifts of SCN probes in highly heterogeneous protein environments of Ral guanine nucleotide dissociation stimulator (RalGDS) and its complexes with human oncoprotein p21$^{Ras}$ (Ras) mutants. In such protein environments, the exchange-repulsion frequency blue-shifts result not just from H-bonding interaction with solvent water molecules but also from the close contacts with neighbouring aliphatic or aromatic side chains inside the proteins. This is the first theoretical attempt to develop a vibrational solvatochromism model beyond Coulomb approximation, which is applicable to any IR probes incorporated into proteins. Once the accuracy of the present approach is improved in the future, it will be possible to make quantitative and direct comparisons of computational results with experimentally measured frequency shifts. We anticipate that the vibrational solvatochromism theory combined with frequency- and/or time-resolved vibrational spectroscopy such as FTIR, femtosecond IR pump-probe, and two-dimensional IR technique would then be of use to measure molecular granularity, dynamics, and conformational heterogeneity of IR probe-incorporated biomolecules.

## Methodology

### Computations

Geometry optimizations and vibrational analyses of model QM clusters of MeCN and MeSCN molecules were performed by using Gaussian09 package.[58] Cubic anharmonic constants[59] were obtained by numerical differentiation of Hessian matrices as described in Ref.[20] Hybrid variational-perturbational interaction energy calculations were performed by utilizing GAMESS package[60] modified by Robert W. Góra.[61] SolEDS calculations were undertaken as in Ref.[20] SolEFP parameters for CN stretch mode of MeSCN were computed in a similar way as described previously.[20] In short, cumulative atomic multipole moments (CAMM)[43], wavefunctions, Fock matrices and distributed polarizabilities were differentiated numerically with respect to all the normal coordinates of MeSCN. The differentiation steps in Cartesian and normal mode spaces were set to be 0.006 and 0.084 Å, respectively. Pipek-Mezey method[62] were used to localize molecular orbitals including core electrons. Distributed polarizabilities centered at localized molecular orbitals (LMO) were computed by using GAMESS package.[60] Geometry optimization and harmonic vibrational analyses were performed by using Gaussian09 package. Throughout all of the computations of SolEFP parameters as well as frequency analyses, HF method[63] and 6-311++G** Cartesian basis set[64] was used as implemented in Gaussian09. MP2 calculations were performed assuming a frozen core approximation.

Hybrid QM/MM molecular dynamics simulations of MeSCN in $H_2O$, DMSO, $CHCl_3$ and $CCl_4$ were performed by using Amber11 program[65], Parm99 force field[66] was utilized to describe the solvent whereas MeSCN was modeled by PM3 semi-empirical Hamiltonian[67]. Force field parameters for $CCl_4$, $CHCl_3$ and DMSO were taken from the work of Fox and Kollman[68] and water was treated by TIP4P model[69]. MeSCN was consecutively solvated by 817 $CCl_4$, 1309 $CHCl_3$, 1042 DMSO and 4615 $H_2O$ molecules. We generated such very large simulation boxes because we have not implemented any kind of Ewald summation for the convergence of the electrostatic frequency shifts. 30000 steps of steepest descent energy minimization were performed and the cascade of 1 ns NVT and NPT simulations were run to equilibrate the system densities. Finally, 2 ns production NVT simulations were run and 20000 frames for each system (which were analyzed by using SolEFP theory) were saved every 0.10 ps. Throughout all these simulations, bonds involving hydrogen atoms in MM part of the systems were constrained using SHAKE algorithm.[70] To control the total temperatures (NVT) and pressures (NPT) Langevin thermostat[71] with coupling frequency of 0.1 ps and Berendsen barostat[72] with pressure relaxation time of 1 ps were used. 12 Å cut-off for non-bonding interactions was set and long-range electrostatics was treated by using the particle-mesh Ewald summation method.[73] The time step for integration of equations of motion was assumed to be 1 fs. The SolEFP analysis of the resulting MD trajectories was performed by using the rigid molecule algorithm[42] that utilizes the Kabsch method[74] implemented in the Biopython package.[75] MD trajectories were interfaced with our in-house SolEFP code by using MDAnalysis package.[76] The vibrational Coulomb, induction, dispersion and exchange-repulsion interactions between MeSCN and solvent molecules were cut at 40, 17, 17 and 13 Bohrs, respectively, to achieve the compromise between the speed and accuracy. Unfortunately, despite the chosen cut-off was sufficient to converge Coulomb frequency shifts, the non-Coulombic contributions were not fully converged and the errors of ± 1-3 $cm^{-1}$ for each contribution are to be expected. However, due to the significant computational demands for the evaluation of the exchange-repulsion and induction frequency shifts, we believe that the errors are at least comparable between the studied systems and hence the relative frequency shifts are still well described within the 3 $cm^{-1}$ accuracy (for more details see Figures S2-5 and discussion in ESI).

Umbrella sampling molecular dynamics simulations were performed on the six SCN probe locations (N27$C_{SCN}$, G28$C_{SCN}$, N29$C_{SCN}$, Y31$C_{SCN}$, K32$C_{SCN}$ and N54$C_{SCN}$) of both free RalGDS and RalGDS+Ras D30E (RalGDS+Ras') complexes in water as has been discussed previously.[56, 77] In brief, 2D umbrella sampling strategy was utilized to sample the SCN probe orientation relative to the protein. Two dihedral angles specifying the SCN orientation were varied independently every 30° resulting in 144 different biased simulations per each protein system. AMBER03 forcefield[78] was used with TIP3P water model[69] as implemented in GROMACS[79]. Note that SCN probe in our simulations of MeSCN in bulk solvents from Section 3.3 was modeled with PM3 Hamiltonian[67].

### Experiments

Absorption frequencies of CN stretch modes in MeCN and MeSCN dissolved in various solvents were measured by using FTIR (Bruker VERTEX 70) spectrometer at 22°C with 1 $cm^{-1}$ resolution. In the cases of very weak solubility, the saturated solutions of Me(S)CN were prepared by ultrasonification (JEOITECH UC-02) of the mixture of two immiscible phases through 1 hour and subsequent separation of the sample phase with a syringe. All compounds were obtained in pure forms from Sigma Aldrich and were not further purified.

## Acknowledgements

This work was supported by IBS-R023-D1 and the Welch Foundation (Grant No. F1722). Quantum chemistry calculations were performed at Wrocław Centre of Networking and Supercomputing (WCSS) of which support is greatly acknowledged. The authors gratefully acknowledge the Texas Advanced Computing Center (TACC) at The University of Texas at Austin for providing high-performance computing resources that have contributed to the results within this paper.

## Notes and references

1. S. D. Fried and S. G. Boxer, *Acc. Chem. Res.*, 2015, **48**, 998-1006.
2. N. M. Levinson, S. D. Fried and S. G. Boxer, *J. Phys. Chem. B*, 2012, **116**, 10470-10476.
3. N. M. Levinson, E. E. Bolte, C. S. Miller, S. A. Corcelli and S. G. Boxer, *J. Am. Chem. Soc.*, 2011, **133**, 13236-13239.
4. P. Hanoian, C. T. Liu, S. Hammes-Schiffer and S. Benkovic, *Acc. Chem. Res.*, 2015, **48**, 482-489.
5. C. M. Ragain, R. W. Newberry, A. W. Ritchie and L. J. Webb, *J. Phys. Chem. B*, 2012, **116**, 9326-9336.

View Article Online
DOI: 10.1039/C6CP01578F

6. A. T. Fafarman and S. G. Boxer, *J. Phys. Chem. B*, 2010, **114**, 13536-13544.
7. A. J. Stafford, D. L. Ensign and L. J. Webb, *J. Phys. Chem. B*, 2010, **114**, 15331-15344.
8. A. T. Fafarman, L. J. Webb, J. I. Chuang and S. G. Boxer, *J. Am. Chem. Soc.*, 2006, **128**, 13356-13357.
9. Z. Getahun, C.-Y. Huang, T. Wang, B. De León, W. F. DeGrado and F. Gai, *J. Am. Chem. Soc.*, 2003, **125**, 405-411.
10. J. Liu, J. Strzalka, A. Tronin, J. S. Johansson and J. K. Blasie, *Biophys. J.*, 2009, **96**, 4176-4187.
11. K. C. Schultz, L. Supekova, Y. Ryu, J. Xie, R. Perera and P. G. Schultz, *J. Am. Chem. Soc.*, 2006, **128**, 13984-13985.
12. J. R. Reimers and L. E. Hall, *J. Am. Chem. Soc.*, 1999, **121**, 3730-3744.
13. A. T. Fafarman, P. A. Sigala, D. Herschlag and S. G. Boxer, *J. Am. Chem. Soc.*, 2010, **132**, 12811-12813.
14. L. J. G. W. van Wilderen, D. Kern-Michler, H. M. Muller-Werkmeister and J. Bredenbeck, *Phys. Chem. Chem. Phys.*, 2014, **16**, 19643-19653.
15. R. Rey and J. T. Hynes, *J. Chem. Phys.*, 1998, **108**, 142-153.
16. C. M. Morales and W. H. Thompson, *J. Phys. Chem. B*, 2011, **115**, 7597-7605.
17. D. Ben-Amotz, M. R. Lee, S. Y. Cho and D. J. List, *J. Chem. Phys.*, 1992, **96**, 8781-8792.
18. H. Lee, J.-H. Choi and M. Cho, *J. Chem. Phys.*, 2012, **137**, 114307-114314.
19. S. D. Fried, S. Bagchi and S. G. Boxer, *J. Am. Chem. Soc.*, 2013, **135**, 11181-11192.
20. B. Błasiak and M. Cho, *J. Chem. Phys.*, 2015, **143**, 164111.
21. W. Zierkiewicz, P. Jurečka and P. Hobza, *ChemPhysChem*, 2005, **6**, 609-617.
22. P. Rodziewicz, K. S. Rutkowski, S. M. Melikova and A. Koll, *ChemPhysChem*, 2005, **6**, 1282-1292.
23. X. Li, L. Liu and H. B. Schlegel, *J. Am. Chem. Soc.*, 2002, **124**, 9639-9647.
24. B. Błasiak and M. Cho, *J. Chem. Phys.*, 2014, **140**, 164107.
25. S. Bagchi, S. D. Fried and S. G. Boxer, *J. Am. Chem. Soc.*, 2012, **134**, 10373-10376.
26. W. Zhang, B. N. Markiewicz, R. S. Doerksen, I. I. I. A. B. Smith and F. Gai, *Phys. Chem. Chem. Phys.*, 2015, DOI: 10.1039/C5CP04413H.
27. M. J. Kamlet, J. L. Abboud and R. W. Taft, *J. Am. Chem. Soc.*, 1977, **99**, 6027-6038.
28. M. J. Kamlet and R. W. Taft, *J. Am. Chem. Soc.*, 1976, **98**, 377-383.
29. R. W. Taft and M. J. Kamlet, *J. Am. Chem. Soc.*, 1976, **98**, 2886-2894.
30. J. G. Kirkwood, *J. Chem. Phys.*, 1934, **2**, 351-361.
31. L. Onsager, *J. Am. Chem. Soc.*, 1936, **58**, 1486-1493.
32. J.-H. Choi and M. Cho, *J. Chem. Phys.*, 2011, **134**, 154513.
33. M. Cho, *J. Chem. Phys.*, 2009, **130**, 094505.
34. M. Cho, *J. Chem. Phys.*, 2003, **118**, 3480-3490.
35. S. S. Andrews and S. G. Boxer, *J. Phys. Chem. A*, 2002, **106**, 469-477.
36. J. R. Reimers, J. Zeng and N. S. Hush, *J. Phys. Chem.*, 1996, **100**, 1498-1504.
37. N. S. Hush and J. R. Reimers, *J. Phys. Chem.*, 1995, **99**, 15798-15805.
38. I. T. Suydam and S. G. Boxer, *Biochemistry*, 2003, **42**, 12050-12055.
39. S. S. Andrews and S. G. Boxer, *J. Phys. Chem. A*, 2000, **104**, 11853-11863.
40. J.-H. Choi, K.-I. Oh, H. Lee, C. Lee and M. Cho, *J. Chem. Phys.*, 2008, **128**, 134506-134508.
41. M. M. Waegele and F. Gai, *J. Phys. Chem. Lett.*, 2010, **1**, 781-786.
42. B. Błasiak, H. Lee and M. Cho, *J. Chem. Phys.*, 2013, **139**, 044111.
43. W. A. Sokalski and R. A. Poirier, *Chem. Phys. Lett.*, 1983, **98**, 86-92.
44. H. Li, H. M. Netzloff and M. S. Gordon, *J. Chem. Phys.*, 2006, **125**, 194103.
45. S. M. Cybulski, G. Chałasiński and R. Moszyński, *J. Chem. Phys.*, 1990, **92**, 4357-4363.
46. W. A. Sokalski, S. Roszak and K. Pecul, *Chem. Phys. Lett.*, 1988, **153**, 153-159.
47. G. Chałasiński and M. M. Szczęśniak, *Mol. Phys.*, 1988, **63**, 205-224.
48. R. W. Góra, W. Bartkowiak, S. Roszak and J. Leszczyński, *J. Chem. Phys.*, 2004, **120**, 2802-2813.
49. S. D. Fried, L.-P. Wang, S. G. Boxer, P. Ren and V. S. Pande, *J. Phys. Chem. B*, 2013, **117**, 16236-16248.
50. I. Suzuki, J. Nakagawa and T. Fujiyama, *Spectrochim. Acta Part A*, 1977, **33**, 689-698.
51. J. D. Jackson, in *Classical Electrodynamics, 2nd ed.*, Wiley, Singapore, 1990, 1996.
52. A. D. Buckingham, *Q. Rev., Chem. Soc.*, 1959, **13**, 183-214.
53. M. S. Gordon, D. G. Fedorov, S. R. Pruitt and L. V. Slipchenko, *Chem. Rev.*, 2012, **112**, 632-672.
54. I. Adamovic and M. S. Gordon, *Mol. Phys.*, 2005, **103**, 379-387.
55. W. R. Fawcett, G. Liu and T. E. Kessler, *J. Phys. Chem.*, 1993, **97**, 9293-9298.
56. A. W. Ritchie and L. J. Webb, *J. Phys. Chem. B*, 2015, **119**, 13945-13957.
57. H. Zou, J. Liu and J. K. Blasie, *Biophys. J.*, 2009, **96**, 4188-4199.
58. M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P, Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. J. A. Montgomery, J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, Ö. Farkas, J. B.

View Article Online
DOI: 10.1039/C6CP01578F

Foresman, J. V. Ortiz, J. Cioslowski and D. J. Fox, *Gaussian 09, Revision D.1*, Gaussian, Inc, Wallingford CT, 2009.

59. V. Barone, *J. Chem. Phys.*, 2005, **122**, 014108-014110.
60. M. W. Schmidt, K. K. Baldridge, J. A. Boatz, S. T. Elbert, M. S. Gordon, J. H. Jensen, S. Koseki, N. Matsunaga, K. A. Nguyen, S. Su, T. L. Windus, M. Dupuis and J. A. Montgomery, *J. Comput. Chem.*, 1993, **14**, 1347-1363.
61. R. W. Góra, *EDS, Revision 2.8.1*, Wrocław University of Technology, Wrocław, Poland, 1998-2010.
62. J. Pipek and P. G. Mezey, *J. Chem. Phys.*, 1989, **90**, 4916-4926.
63. C. C. J. Roothaan, *Rev. Mod. Phys.*, 1951, **23**, 69-89.
64. R. Krishnan, J. S. Binkley, R. Seeger and J. A. Pople, *J. Chem. Phys.*, 1980, **72**, 650-654.
65. D. A. Case, T. A. Darden, T. E. Cheatham, III, C. L. Simmerling, J. Wang, R. E. Duke, R. Luo, R. C. Walker, W. Zhang, K. M. Merz, B. Roberts, B. Wang, S. Hayik, A. Roitberg, G. Seabra, I. Kolossváry, K. F. Wong, F. Paesani, J. Vanicek, J. Liu, X. Wu, S. R. Brozell, T. Steinbrecher, H. Gohlke, Q. Cai, X. Ye, J. Wang, M.-J. Hsieh, G. Cui, D. R. Roe, D. H. Mathews, M. G. Seetin, C. Sagui, V. Babin, T. Luchko, S. Gusarow, A. Kovalenko and P. A. Kollman, *AMBER 11*, University of California, San Fransisco, 2010.
66. J. Wang, P. Cieplak and P. A. Kollman, *J. Comput. Chem.*, 2000, **21**, 1049-1074.
67. J. J. P. Stewart, *J. Comput. Chem.*, 1989, **10**, 209-220.
68. T. Fox and P. A. Kollman, *J. Phys. Chem. B*, 1998, **102**, 8070-8079.
69. W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Klein, *J. Chem. Phys.*, 1983, **79**, 926-935.
70. J.-P. Ryckaert, G. Ciccotti and H. J. C. Berendsen, *J. Comp. Phys.*, 1977, **23**, 327-341.
71. J. A. Izaguirre, D. P. Catarello, J. M. Wozniak and R. D. Skeel, *J. Chem. Phys.*, 2001, **114**, 2090-2098.
72. H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola and J. R. Haak, *J. Chem. Phys.*, 1984, **81**, 3684-3690.
73. T. Darden, D. York and L. Pedersen, *J. Chem. Phys.*, 1993, **98**, 10089-10092.
74. W. Kabsch, *Acta Cryst. A*, 1976, **32**, 922-923.
75. P. J. A. Cock, T. Antao, J. T. Chang, B. A. Chapman, C. J. Cox, A. Dalke, I. Friedberg, T. Hamelryck, F. Kauff, B. Wilczynski and M. J. L. de Hoon, *Bioinformatics*, 2009, **25**, 1422-1423.
76. N. Michaud-Agrawal, E. J. Denning, T. B. Woolf and O. Beckstein, *J. Comput. Chem.*, 2011, **32**, 2319-2327.
77. A. W. Ritchie and L. J. Webb, *J. Phys. Chem. B*, 2013, **117**, 11473-11489.
78. Y. Duan, C. Wu, S. Chowdhury, M. C. Lee, G. Xiong, W. Zhang, R. Yang, P. Cieplak, R. Luo, T. Lee, J. Caldwell, J. Wang and P. Kollman, *J. Comput. Chem.*, 2003, **24**, 1999-2012.
79. D. Van Der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark and H. J. C. Berendsen, *J. Comput. Chem.*, 2005, **26**, 1701-1718.
80. Y. Koga, S. Kondo, S. Saeki and W. B. Person, *J. Phys. Chem.*, 1984, **88**, 3152-3157.
81. F. Cerceau, F. Raulin, R. Courtin and D. Gautier, *Icarus*, 1985, **62**, 207-220.
82. F. W. Parker, A. H. Nielsen and W. H. Fletcher, *J. Mol. Spectrosc.*, 1957, **1**, 107-123.
83. G. Eaton, A. S. Pena-Nunez and M. C. R. Symons, *J. Chem. Soc., Faraday Trans. 1*, 1988, **84**, 2181-2193.
84. K.-I. Oh, J.-H. Choi, J.-H. Lee, J.-B. Han, H. Lee and M. Cho, *J. Chem. Phys.*, 2008, **128**, 154504-154510.
85. J. F. Sullivan, H. L. Heusel and J. R. Durig, *J. Mol. Struct.*, 1984, **115**, 391-396.
86. M. G. Maienschein-Cline and C. H. Londergan, *J. Phys. Chem. A*, 2007, **111**, 10020-10025.

View Article Online
DOI: 10.1039/C6CP01578F

## Figure Captions:

**Figure 1.** Experimental frequencies of CN stretch mode of MeCN and MeSCN dissolved in various solvents at room temperature (see Figure S1 in ESI for the FTIR spectra). In this Figure, $\varepsilon$ stands for the solvent dielectric constant. Frequencies that were not directly measured for this work were taken from the following references: MeCN in gas-phase (Refs.[80-82]), MeCN in $CCl_4$, THF, MeOH, $H_2O$, $CF_3CH_2OH$ and $(CF_3)_2CHOH$ (Ref.[83]), MeCN in $CH_3(CH_2)_2OH$ (Ref.[84]), MeSCN in gas-phase (Ref.[85]), MeSCN in $CCl_4$, $CHCl_3$, $CH_2Cl_2$, MeCN, EtOH, MeOH, $H_2O$, $CF_3CH_2OH$ and $(CF_3)_2CHOH$ (Ref.[86]). In the cases of solvents with vanishingly small dipole moment (heptane, hexane, cyclohexane, isooctane, $CCl_4$), the linear relationship between CN frequency and Onsager factor indicates that the Kirkwood-Bauer-Magat law[30, 31] (blue line) works well.

**Figure 2.** Electric potential produced by the vibrational solvatochromic multipole moments of the vibrational 0-1 transition of nitrile (CN) stretch mode in MeSCN, that were computed by using SolCAMM//CCSD/6-311++G** method. (a) potential derived from SolCAMM (up to octupoles) distributed over all atoms; (b), (c) and (d) potential derived from the total molecular multipole moments (SolMMM) centered at the middle of CN bond. Spatial dimensions are in Bohrs.

**Figure 3.** SolEFP(HF/6-311++G**) frequency shift components. (a) The SolEFP components contributing to the CN frequency shifts of MeSCN in various water and DMSO clusters, MeSCN-$(H_2O)_{n=1-27}$ and MeSCN-$(DMSO)_{n=1-7}$.The total SolEFP frequency shifts are also directly compared with the results from HF/6-311++G** calculations. (b) The SolEFP components for MeSCN/water dimer with varying intermolecular distance are plotted with respect to the distance between H(water) and N(MeSCN) atoms. The optimum H-bond distance is at 2.2536 Å.

**Figure 4.** Distance- and orientation-dependence of CN stretch mode frequency shift of MeSCN interacting with water molecule. Frequency shifts were calculated by using SolEFP method. Water molecule was translated in the direction along the line between N and O atoms by a distance $R_d$. Water molecule was then rotated in a plane of SCN-O atoms around its center of mass by angle $\theta$. The structure at $R_d = 0$ and $\theta = 0$ corresponds to the equilibrium (optimal H-bonding) geometry of MeSCN-$H_2O$ dimer, which was obtained with HF/6-311++G** method.

**Figure 5.** The ensemble-averaged values of the frequency shifts of CN stretch mode of MeSCN dissolved in four different solvents at room temperature, that were obtained by applying the SolEFP model of MeSCN CN stretch mode to the snapshot configurations taken from the MD trajectories. The experimental data (black diamonds) are also shown for comparison. In this Table, “Dipole” denotes the Stark-dipole model (with the dipole centered at CN mid-bond) whereas “Multipoles” describes the site-distributed field-dependent frequency shift in Eq. (14), with the solvatochromic multipole moments and polarizabilities distributed over all atoms and LMO-centroids, respectively. Both take into account induction effects.

**Figure 6.** Static frequency shift distributions of CN stretch mode in MeSCN dissolved in four different solvent at room temperature that were obtained by applying the SolEFP model of MeSCN CN stretch mode to MD simulation trajectories.

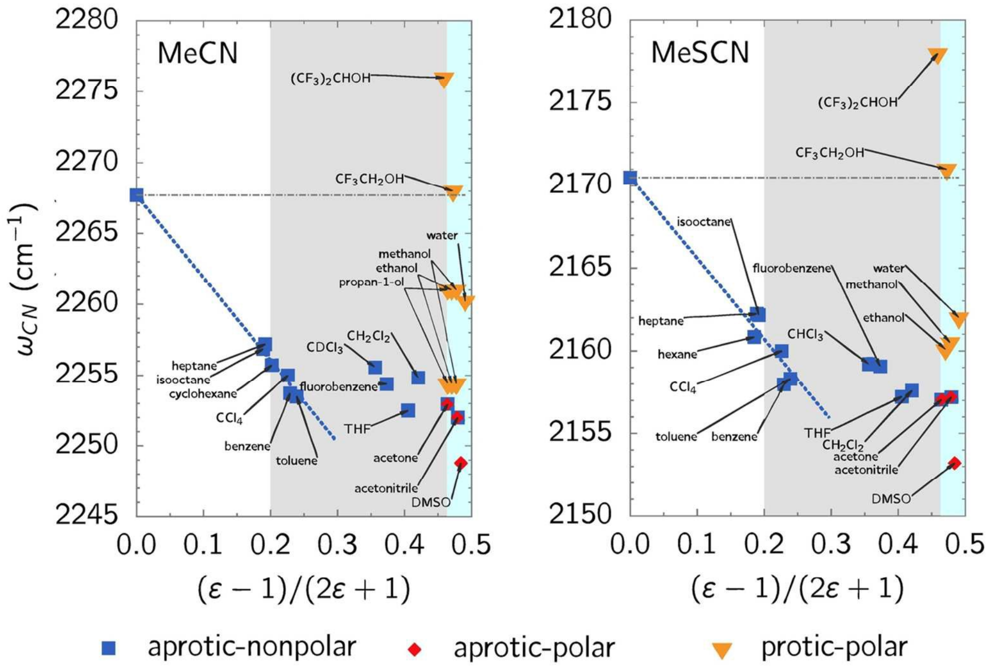
**Figure 7.** Vibrational solvatochromic (SolEFP) induced dipole moments (gray arrows, $\mathbf{a}'$ in the main text) of CN stretch mode of MeSCN interacting with water molecules. Note that the induced dipoles are delocalized even on water molecules. (a) A cluster consisting of MeSCN and three water molecules. (b) A configuration taken from MD trajectory.

**Figure 8.** Through-bond effect on CN frequency shift. (a) The CN stretch mode frequency depends on the length and configuration of the attached aliphatic chain. (b) For a few representative model peptides containing SCN IR probe (see Figure S9 in ESI), we carried out HF vibrational analyses and the resulting frequency shifts are compared with the approximate equation (22).

**Figure 9.** Locations of the SCN probes within RalGDS protein bound to Ras’.

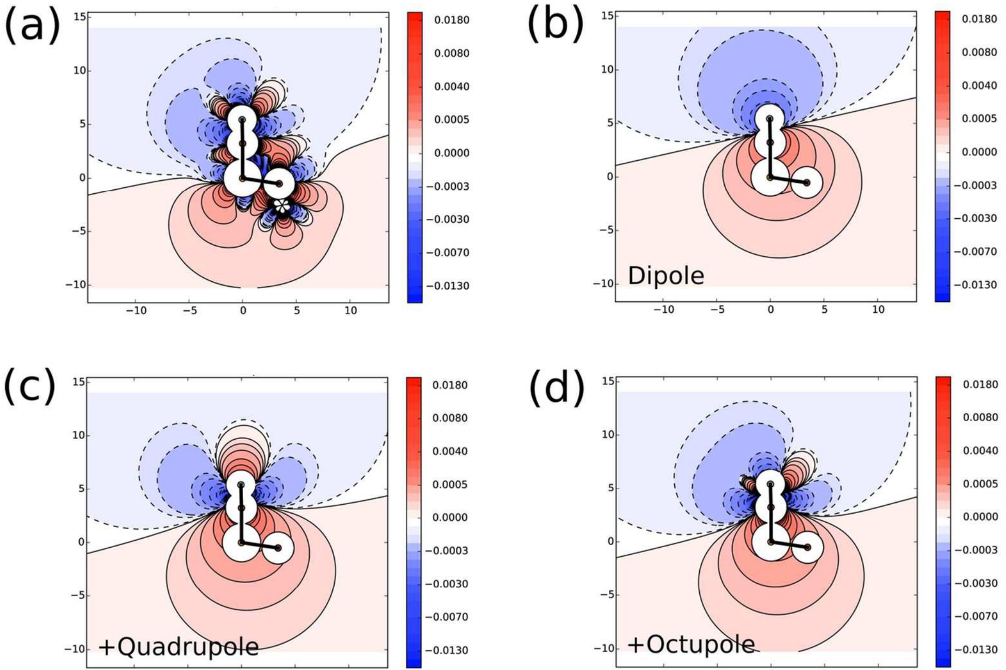
**Figure 10.** The SolEFP short-range frequency shift components of CN stretch mode of SCN probe incorporated at six different sites of RalGDS protein that is non-covalently bound to Ras’. The frequency shifts were averaged over roughly 80 configurations in each case, taken from the umbrella sampling simulations. Only the most probable configurations were selected. The standard deviations for each frequency shift contribution are displayed with black error bars. All values in $cm^{-1}$.

View Article Online
DOI: 10.1039/C6CP01578F



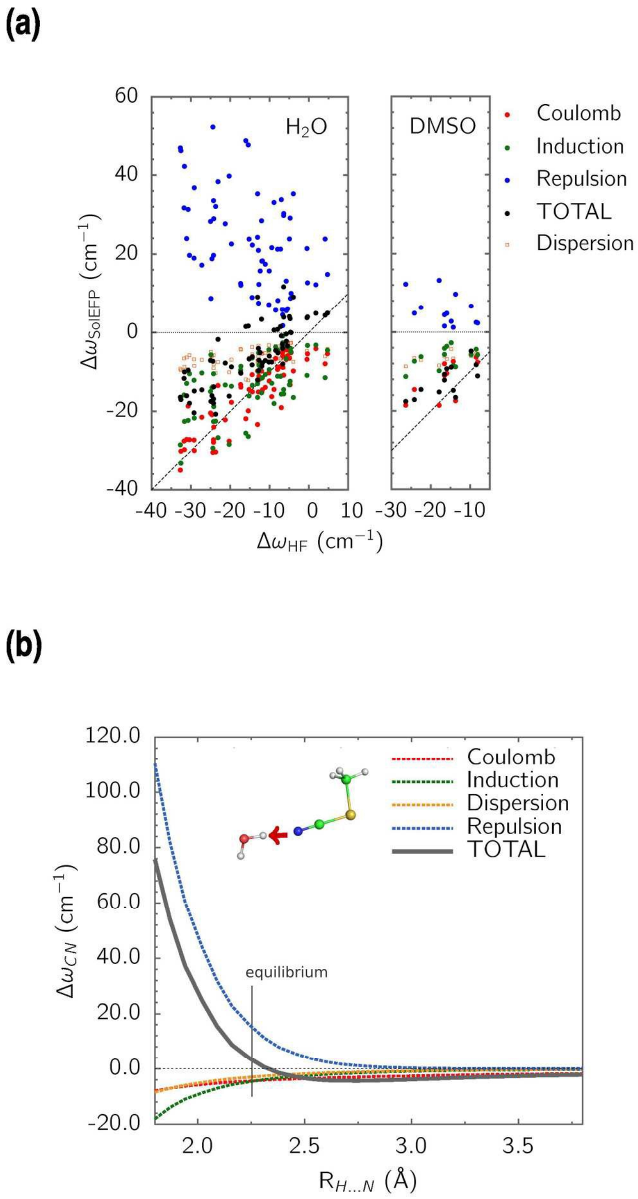
Experimental frequencies of CN stretch mode of MeCN and MeSCN dissolved in various solvents at room temperature (see Figure S1 in ESI for the FTIR spectra). In this Figure, ε stands for the solvent dielectric constant. Frequencies that were not directly measured for this work were taken from the following references: MeCN in gas-phase (Refs.[80-82]), MeCN in $CCl_4$, THF, MeOH, $H_2O$, $CF_3CH_2OH$ and $(CF_3)_2CHOH$ (Ref.[83]), MeCN in $CH_2(CH_3)_2OH$ (Ref.[84]), MeSCN in gas-phase (Ref.[85]), MeSCN in $CCl_4$, $CHCl_3$, $CH_2Cl_2$, MeCN, EtOH, MeOH, $H_2O$, $CF_3CH_2OH$ and $(CF_3)_2CHOH$ (Ref.[86]). In the cases of solvents with vanishingly small dipole moment (heptane, hexane, cyclohexane, isooctane, $CCl_4$), the linear relationship between CN frequency and Onsager factor indicates that the Kirkwood-Bauer-Magat law[30,31] (blue line) works well.
117x80mm (300 x 300 DPI)

View Article Online
DOI: 10.1039/C6CP01578F



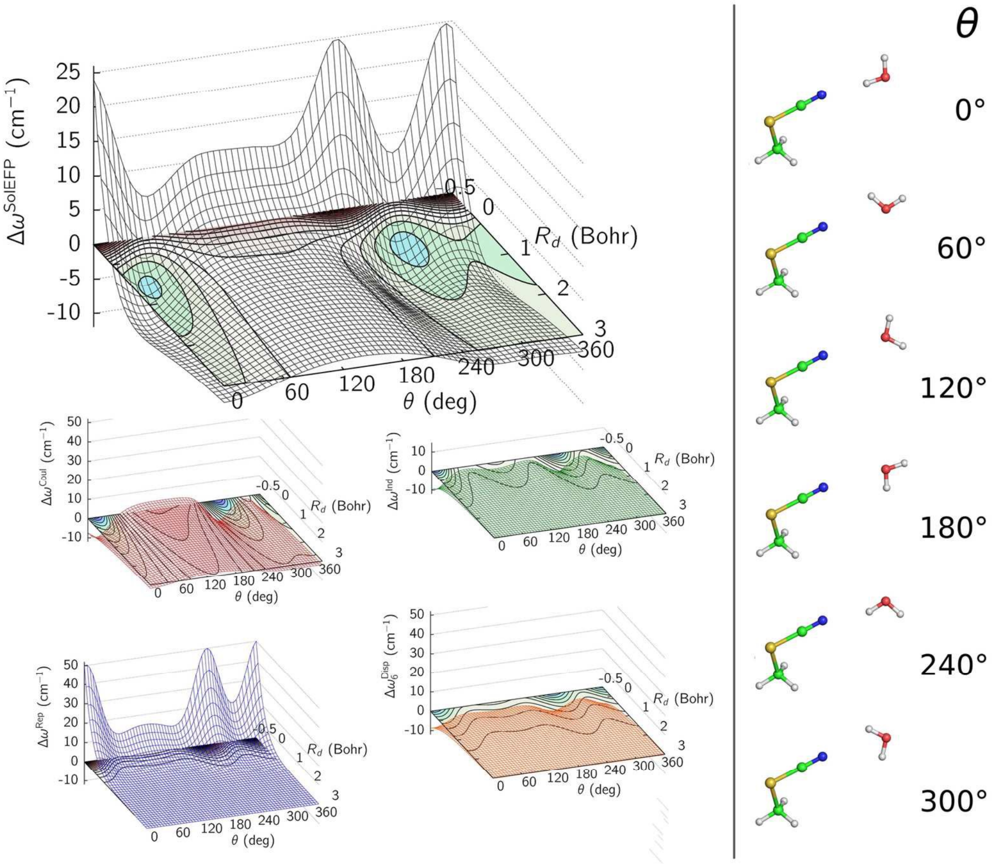
Electric potential produced by the vibrational solvatochromic multipole moments of the vibrational 0-1 transition of nitrile (CN) stretch mode in MeSCN, that were computed by using SolCAMM//CCSD/6-311++G** method. (a) potential derived from SolCAMM (up to octupoles) distributed over all atoms; (b), (c) and (d) potential derived from the total molecular multipole moments (SolMMM) centered at the middle of CN bond. Spatial dimensions are in Bohrs.
118x82mm (300 x 300 DPI)

View Article Online
DOI: 10.1039/C6CP01578F



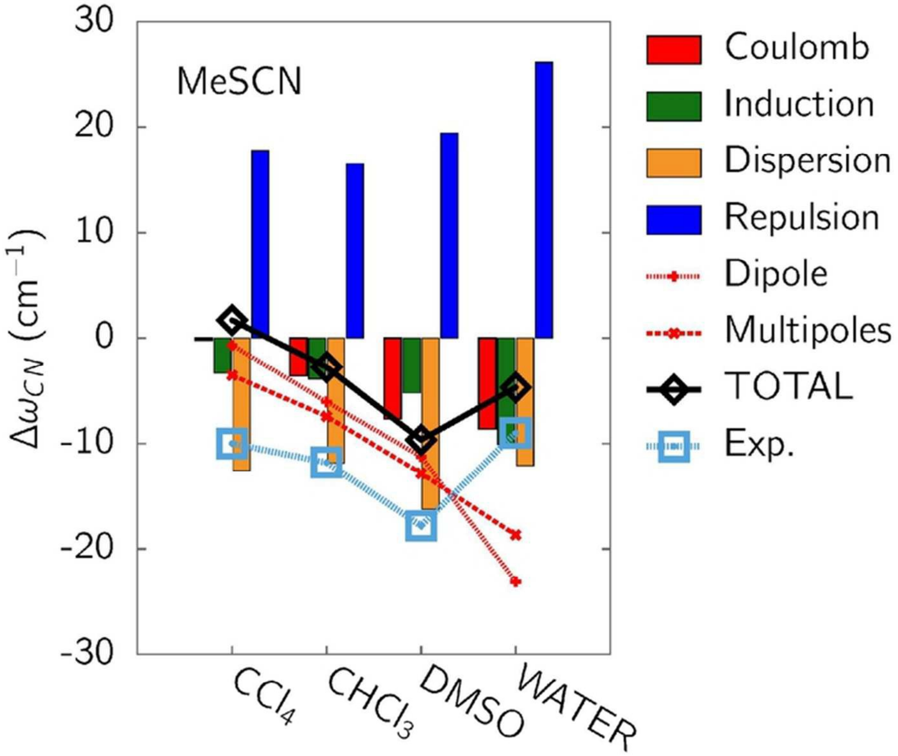
SolEFP(HF/6-311++G**) frequency shift components. (a) The SolEFP components contributing to the CN frequency shifts of MeSCN in various water and DMSO clusters, MeSCN-$(H_2O)_{n=1-27}$ and MeSCN-$(DMSO)_{n=1-7}$.The total SolEFP frequency shifts are also directly compared with the results from HF/6-311++G** calculations. (b) The SolEFP components for MeSCN/water dimer with varying intermolecular distance are plotted with respect to the distance between H(water) and N(MeSCN) atoms. The optimum H-bond distance is at 2.2536 Å.
153x282mm (300 x 300 DPI)

View Article Online
DOI: 10.1039/C6CP01578F



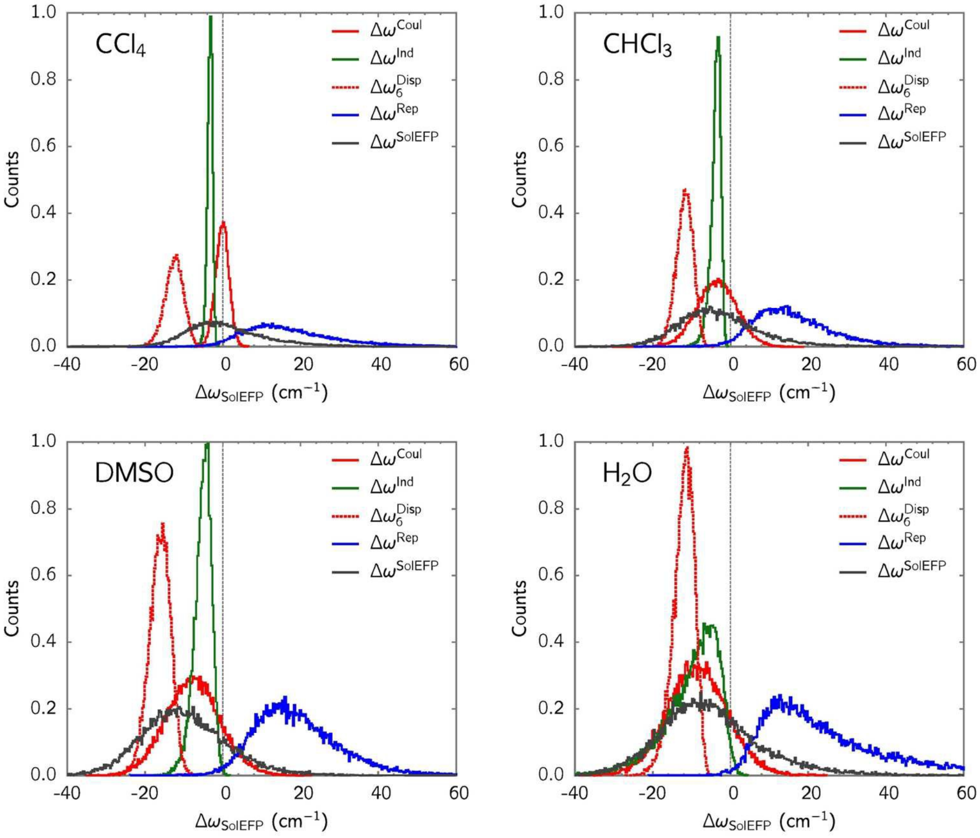
Distance- and orientation-dependence of CN stretch mode frequency shift of MeSCN interacting with water molecule. Frequency shifts were calculated by using SolEFP method. Water molecule was translated in the direction along the line between N and O atoms by a distance $R_d$. Water molecule was then rotated in a plane of SCN-O atoms around its center of mass by angle $\theta$. The structure at $R_d$=0 and $\theta$=0 corresponds to the equilibrium (optimal H-bonding) geometry of MeSCN-$H_2O$ dimer, which was obtained with HF/6-311++G** method.
149x130mm (300 x 300 DPI)

View Article Online
DOI: 10.1039/C6CP01578F



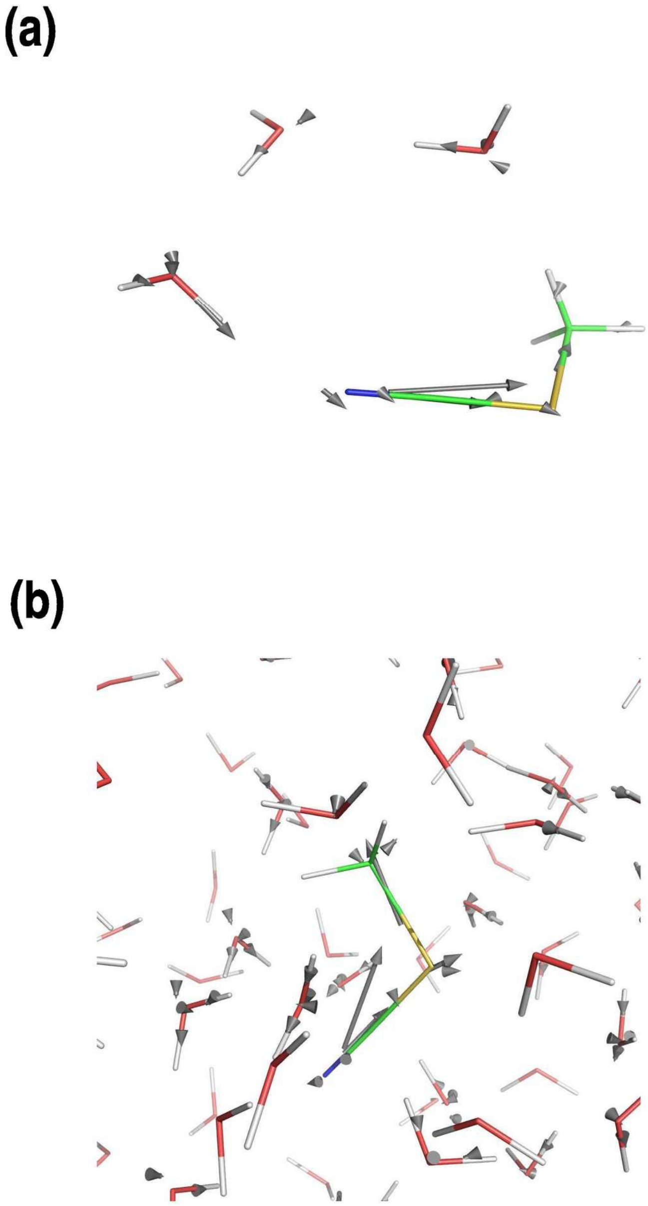
The ensemble-averaged values of the frequency shifts of CN stretch mode of MeSCN dissolved in four different solvents at room temperature, that were obtained by applying the SolEFP model of MeSCN CN stretch mode to the snapshot configurations taken from the MD trajectories. The experimental data (black diamonds) are also shown for comparison. In this Table, "Dipole" denotes the Stark-dipole model (with the dipole centered at CN mid-bond) whereas "Multipoles" describes the site-distributed field-dependent frequency shift in Eq. (14), with the solvatochromic multipole moments and polarizabilities distributed over all atoms and LMO-centroids, respectively. Both take into account induction effects.
70x59mm (300 x 300 DPI)

View Article Online
DOI: 10.1039/C6CP01578F



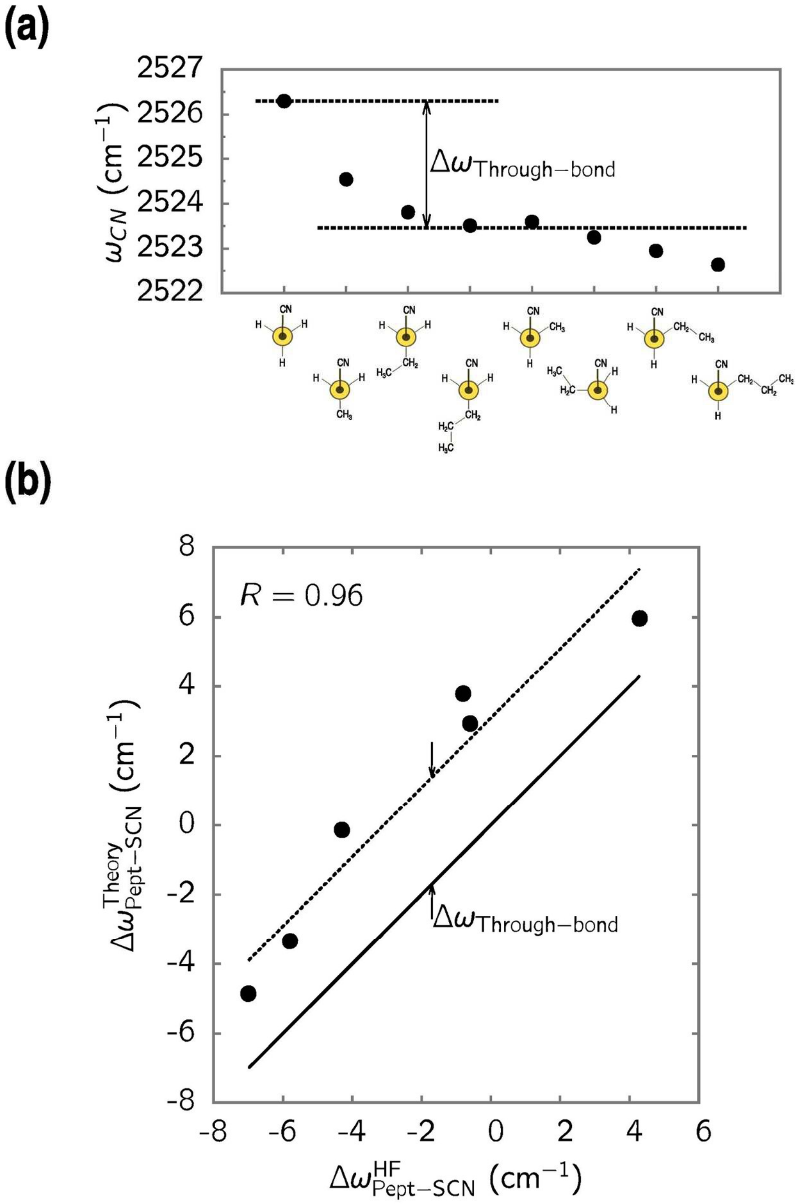
Static frequency shift distributions of CN stretch mode in MeSCN dissolved in four different solvent at room temperature that were obtained by applying the SolEFP model of MeSCN CN stretch mode to MD simulation trajectories.
144x122mm (300 x 300 DPI)

View Article Online
DOI: 10.1039/C6CP01578F



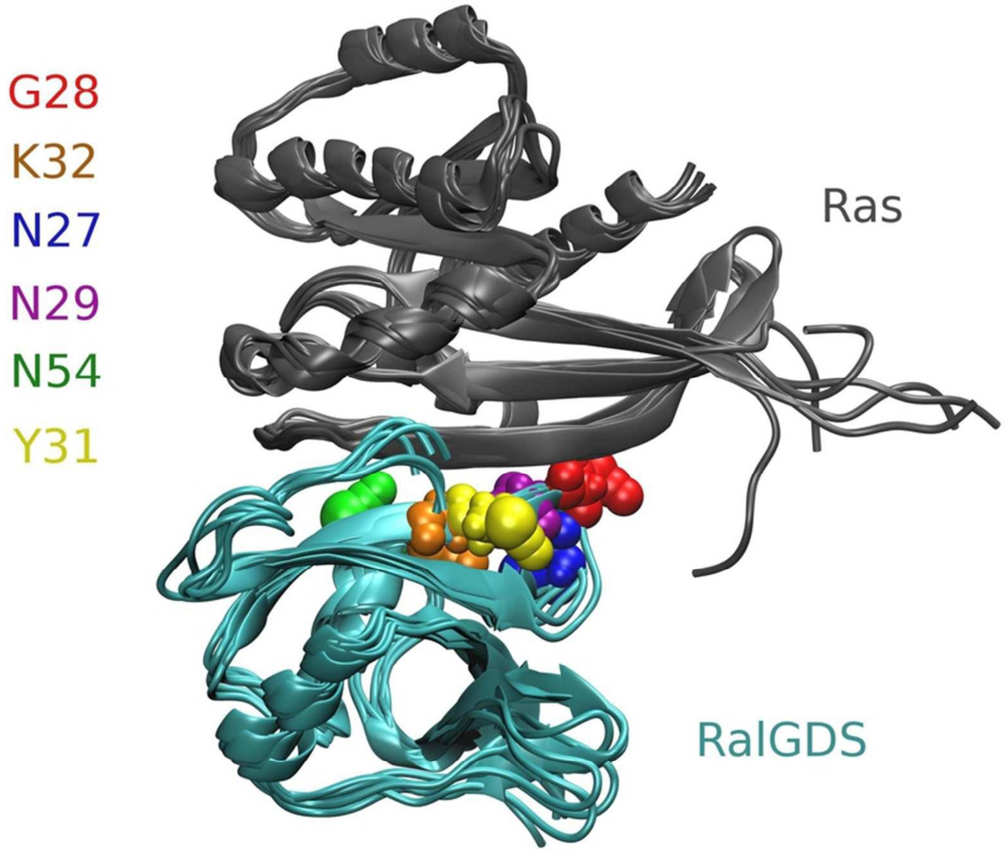
Vibrational solvatochromic (SolEFP) induced dipole moments (gray arrows, **a**′ in the main text) of CN stretch mode of MeSCN interacting with water molecules. Note that the induced dipoles are delocalized even on water molecules. (a) A cluster consisting of MeSCN and three water molecules. (b) A configuration taken from MD trajectory.
143x248mm (300 x 300 DPI)

View Article Online
DOI: 10.1039/C6CP01578F



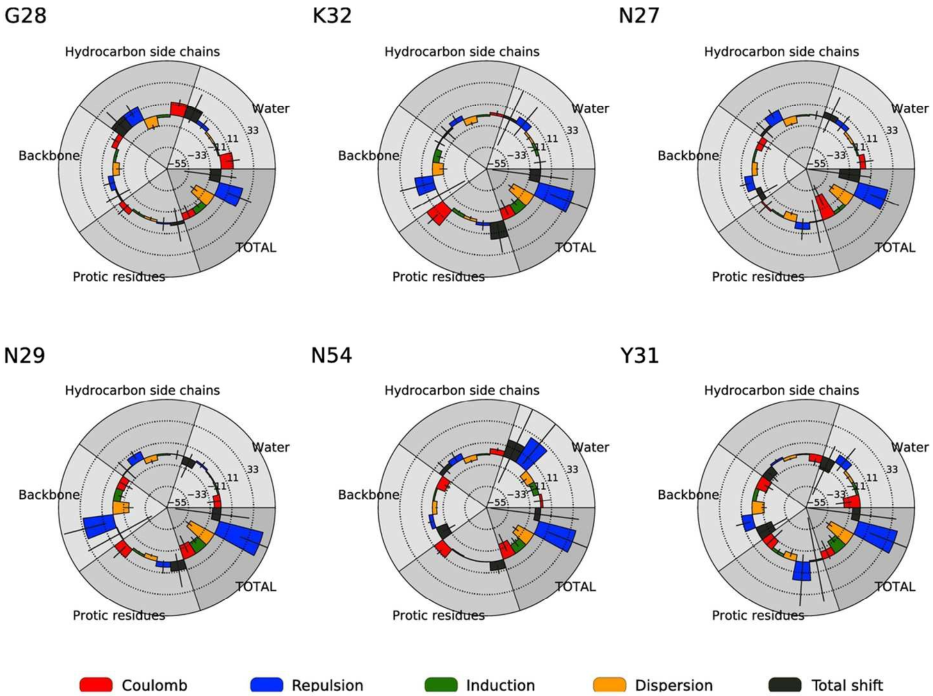
Through-bond effect on CN frequency shift. (a) The CN stretch mode frequency depends on the length and configuration of the attached aliphatic chain. (b) For a few representative model peptides containing SCN IR probe (see Figure S8 in ESI), we carried out HF vibrational analyses and the resulting frequency shifts are compared with the approximate equation (22).
126x191mm (300 x 300 DPI)

View Article Online
DOI: 10.1039/C6CP01578F



Locations of the SCN probes within RalGDS protein bound to Ras'.
71x61mm (300 x 300 DPI)

View Article Online
DOI: 10.1039/C6CP01578F



The SolEFP short-range frequency shift components of CN stretch mode of SCN probe incorporated at six different sites of RalGDS protein that is non-covalently bound to Ras'. The frequency shifts were averaged over roughly 80 configurations in each case, taken from the umbrella sampling simulations. Only the most probable configurations were selected. The standard deviations for each frequency shift contribution are displayed with black error bars. All values in $cm^{-1}$.
117x81mm (300 x 300 DPI)