Vibrational solvatochromism of nitrile infrared probes: beyond the vibrational Stark dipole approach†

Bartosz Błasiak, Andrew W. Ritchie, Lauren J. Webb and Minhaeng Cho*

Systematic probing of local environments around biopolymers is important for understanding their functions. Therefore, there has been 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 by vibrational Stark dipole theory. However, here we show that the nitrile vibrational frequency shift induced by its interaction with the 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. In particular, the vibrational Stark effect theory, which is based on the assumption that vibrational frequency shifts are mainly determined by interaction between the vibrational transition dipole and the local electric field, has been widely used to extract quantitative information on the local electric field around an IR probe. One of the most versatile IR probes used so far are carbonyl (CO) stretches and nitrile (CN) stretches. Despite the fact that a variety of nitrile-derivatized amino acids are selectively and specifically introduced into a macromolecular framework and are used extensively, still there is a lack of a quantitatively reliable model for describing intermolecular interaction-induced vibrational frequency shifts.

It was demonstrated before that solvatochromism of CN stretching vibration in various solvents is quite complicated and cannot be explained simply by the Coulombic interaction of the vibrational dipole with the local electric field produced by the surrounding molecules, which is in notable contrast to the solvatochromism of CO stretching modes. 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. 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 a strong blue-shifting behaviour due to a ‘repulsion wall’ was attributed to the large slope of potential energy that is ‘felt’ by the vibrational probe when interacting strongly with the water molecule via a hydrogen bond. The importance of the 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 stretching 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 anions in water result from van der Waals interactions.
The fact that the non-specific electrostatic and specific H-bonding effects contribute differently to the vibrational shifts was shown experimentally by Boxer’s group,\textsuperscript{13,25} they combined IR and NMR techniques together to separate H-bonding contributions. Recently, Zhang et al.\textsuperscript{26} studied the specific and non-specific nitrile frequency shifts in analogs of 5-cyanotryptophan by utilizing empirically determined Kamlet–Taft parameters\textsuperscript{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 and H-bonding properties. However, this and other similar approaches\textsuperscript{12,30} 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 stretching frequency, which is probably beyond the Coulomb interaction effect.

To show the complicated relationship between nitrile stretching mode frequency and solvent polarity, the experimentally measured vibrational frequencies of the CN stretching 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 (Fig. 1). For aprotic and non-polar solvents, the linear dependence of the nitrile frequency shift on the Onsager factor is evident, suggesting the validity of the Kirkwood–Bauer–Magat (KBM) theory.\textsuperscript{31,32} However, as the solvent polarity increases, the reaction field theory for vibrational solvatochromism fails. Furthermore, as the H-bond donating ability of the solvent molecules increases, the nitrile stretching frequency is strongly blue-shifted, as compared to those in DMSO for instance.\textsuperscript{14} Theoretically, recent quantum mechanical calculations for solute–solvent clusters showed that the interaction between the oscillator’s quadrupole and the solvent electric field gradient is at least equally important for quantitatively describing vibrational frequency shifts of nitrile and other related IR probe modes in water, indicating the breakdown of simple Stark-dipole theory.\textsuperscript{18,33,34} Thus, from the experimental data in Fig. 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 a highly heterogeneous environment such as proteins and nucleic acids in aqueous solutions. To date, there is no such model beyond simple vibrational Stark dipole theory 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 stretching modes. Using our first-principles solvatochromism theory based on the effective

---

**Fig. 1** Experimental frequencies of the CN stretching mode of MeCN and MeSCN dissolved in various solvents at room temperature (see Fig. 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 the gas-phase (ref. 81–83), MeCN in CCl\(_4\), THF, MeOH, H\(_2\)O, CF\(_3\)CH\(_2\)OH and (CF\(_3\))\(_2\)CHOH (ref. 84), MeCN in CH\(_2\)(CH\(_2\))\(_2\)OH (ref. 85), MeSCN in gas-phase (ref. 86), MeSCN in CCl\(_4\), CHCl\(_3\), CH\(_2\)Cl\(_2\), MeCN, EtOH, MeOH, H\(_2\)O, CF\(_3\)CH\(_2\)OH and (CF\(_3\))\(_2\)CHOH (ref. 87). In the cases of solvents with vanishingly small dipole moment (heptane, hexane, cyclohexane, isooctane, CCl\(_4\)), the linear relationship between the CN frequency and the Onsager factor indicates the Kirkwood–Bauer–Magat law\textsuperscript{31,32} (blue line) works well.
fragment potential method\textsuperscript{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 of the SolEFP theory is presented in Section 2. In Section 3.1 we first elaborate on the vibrational properties of the nitrile stretching mode under 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 has already been fully developed.\textsuperscript{34–38} 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:

$$\Delta \omega_j (F) \equiv \omega_j (F) - \omega_{j,0} = \Delta \omega_j^{\text{SolCAMM}} + \frac{1}{2 M_j} \sum_{x, y \in \text{solvent atoms}} \sum_{r \in \text{solvent atoms}} g_{ij} \left( \frac{\partial^2 \phi}{\partial Q^i \partial Q^j} \right) Q_0 \mathbf{r}_x \cdot \mathbf{r}_y \frac{V_j (\mathbf{r}_x)}{r_{xy}}$$

(3)

where

$$\Delta \omega_j^{\text{SolCAMM}} = \sum_{x \in \text{solute atoms}} \left\{ j_{xx} \phi (\mathbf{r}_x) + \mathbf{L}_{xx} \cdot \nabla \phi (\mathbf{r}_x) + \frac{1}{2} \mathbf{A}_{xx} : \nabla \nabla \phi (\mathbf{r}_x) \right\}$$

(4)

Here, $j_{xx}$, $\mathbf{L}_{xx}$, $\mathbf{A}_{xx}$ and $\mathbf{A}_{xx}$ are the distributed solvatochromic charges, dipoles, quadrupoles, and octupoles, respectively, which can be estimated using the first-principles quantum chemistry calculation method (see ref. 43). In eqn (3), $Q_0$ is the charge of the $x$th solute atom at $\mathbf{r}_x$, $V_j (\mathbf{r}_x)$ is the $j$th normal mode vibrational eigenvector of the $x$th solute atom, $\mathbf{r}_x$ is the interatomic position vector defined as $\mathbf{r}_x = \mathbf{r}_x - \mathbf{r}_0$, $\mathbf{g}_{ij} = \{(E / \partial \mathbf{Q} / \partial \mathbf{Q}^2) \mathbf{q}_{ij}\}$ is the gas-phase cubic anharmonic constant, $E$ is the solute total energy, and $M_j$ and $\omega_{j,0}$ are the reduced mass and harmonic frequency of the solute molecule in the gas phase, respectively. In eqn (4), we used cumulative atomic multipole moments (CAMM)\textsuperscript{14} to describe the distribution of electric potential (hence the superscript SolCAMM in eqn (3) and (4) means that the vibrational solvatochromic shift is calculated using 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,

$$\Delta \omega_j^{\text{Ex-Rep}} = -\frac{1}{2M_j} \sum_{x, y \in \text{solute atoms}} g_{ij} \mathbf{M}_{ij} \frac{\partial}{\partial \mathbf{Q}} \left[ -4 \mathbf{2} \mathbf{ln} | \mathbf{S}_{ab} | | \mathbf{S}_{ab} | | \mathbf{S}_{ab} | | \mathbf{S}_{ab} | \frac{1}{r_{ab}} \right]$$

(5)

$$-2 S_{ab} \left( \sum_{x \in A} G_{xc}^a S_{xb} + \sum_{d \in B} G_{bd}^b S_{da} - 2 T_{ab} \right)$$

$$-2 S_{ab} \left( \sum_{x \in A} G_{xc}^a S_{xb} + \sum_{d \in B} G_{bd}^b S_{da} - 2 T_{ab} \right)$$

where $G_{xc}^a$ is Fock’s 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 LMOs and the relative distance between the LMOs and nuclei is defined as $r_{ab} = | \mathbf{r}_a - \mathbf{r}_b |$. In eqn (5), the labels A and B denote solute and...
solvent molecules, the indices $a$, $b$, $c$, and $d$ refer to LMOs, whereas the indices $x$ and $y$ refer to atomic sites, respectively.\textsuperscript{24}

Since all the molecules are polarizable when they are exposed to the electric field, the induction effect on the 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:

$$\Delta\omega_{ij}^{\text{Ind}} = -\frac{1}{2}a_{j}^T F - \frac{1}{2}P^{T}A_{j} F$$

(6)

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

$$a_{j} = (\mu_{j,1x}, \mu_{j,1y}, \mu_{j,1z}, \mu_{j,2x}, \mu_{j,2y}, \mu_{j,2z}, \ldots)$$

$$F = (F_{1x}, F_{1y}, F_{2x}, F_{2y}, F_{2z}, \ldots)$$

(7)

In the case when the system has $N$ polarizable centers, $F$ and $a_{j}$ have $3N$ elements and the dot product $\cdot$ in eqn (6) is also determined by the distributed vibrational solvatochromic polarization.

In ref. 24, we showed that

$$a_{j} = -\frac{1}{2M_{j,0\omega_{j,0}}^{\text{P}}} \sum_{i} \frac{g_{ij}}{M_{i,0\omega_{i,0}}^{\text{P}}} \left( \frac{\partial F}{\partial \omega_{j}} \right)_{\omega_{0}} \cdot \left[ D^{-1} + (D^{-1})^{T} \right]$$

(8)

and

$$A_{j} = \frac{1}{2M_{j,0\omega_{j,0}}^{\text{P}}} \cdot \left[ \sum_{i} \frac{g_{ij}}{M_{i,0\omega_{i,0}}^{\text{P}}} \left( \frac{\partial D}{\partial \omega_{j}} \right)_{\omega_{0}} \right] \cdot D^{-1}$$

(9)

Here, the $D$ matrix is defined as

$$D_{ab} = \begin{cases} a_{a}^{-1}(0)\delta_{ab} & \text{if } a, b \text{ belong to the same molecule} \\ -T_{ab} & \text{if } a, b \text{ belong to different molecules} \end{cases}$$

(10)

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

$$\left( \frac{\partial D_{ab}}{\partial \omega_{j}} \right)_{\omega_{0}} = \begin{cases} -a_{a}^{-1}(0) \left( \frac{\partial a_{a}(0)}{\partial \omega_{j}} \right)_{\omega_{0}} & \text{if } a, b \text{ belong to the same molecule} \\ -T_{ab} \left( \frac{\partial \omega_{j}}{\partial \omega_{j}} \right)_{\omega_{0}} & \text{if } a, b \text{ belong to different molecules} \end{cases}$$

(11)

In the above equation, $T_{ab}$ 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.\textsuperscript{24,45}

For the sake of notational simplicity, we additionally define $a_{j}' = a_{j} + F^{T}A_{j}$, so that the induction term can be written as

$$\Delta\omega_{ij}^{\text{Ind}} = -\frac{1}{2}a'_{j} \cdot F$$

(12)

which is identical to eqn (32) in ref. 24.

The dispersion contribution to the vibrational frequency shift was recently found to be\textsuperscript{20}

$$\Delta\omega_{ij}^{\text{Disp}} \approx \frac{1}{4\pi M_{j,0\omega_{j,0}}} \sum_{i} \frac{g_{ij}}{M_{i,0\omega_{i,0}}} \left( T_{\omega_{j},\omega_{j}}^{ab} T_{\omega_{j},\omega_{j}}^{bc} \right)_{\omega_{0}} \int_{0}^{\infty} \left( \frac{\partial \omega_{j}}{\partial \omega_{j}} \right)_{\omega_{0}} \left( \frac{\partial \omega_{j}}{\partial \omega_{j}} \right)_{\omega_{0}} \left( \frac{\partial \omega_{j}}{\partial \omega_{j}} \right)_{\omega_{0}} d\omega$$

$$+ \left( \frac{\partial T_{\omega_{j},\omega_{j}}^{ab}}{\partial \omega_{j}} \right)_{\omega_{0}} \left( \frac{\partial T_{\omega_{j},\omega_{j}}^{bc}}{\partial \omega_{j}} \right)_{\omega_{0}} \left( \frac{\partial \omega_{j}}{\partial \omega_{j}} \right)_{\omega_{0}} d\omega$$

(13)

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

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

$$\Delta\omega_{ij}^{\text{Electric}} = \Delta\omega_{ij}^{\text{Coal}} + \Delta\omega_{ij}^{\text{Ind}}$$

(14)

The total frequency shift is therefore given as

$$\Delta\omega_{ij} = \Delta\omega_{ij}^{\text{Electric}} + \Delta\omega_{ij}^{\text{Rep-Rep}} + \Delta\omega_{ij}^{\text{Disp}}$$

(15)

Note that, while $\Delta\omega_{ij}^{\text{Electric}}$ can provide information on the 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 eqn (15), are not directly related to the solvent electric field produced by the distributed charges of the surrounding solvent molecules. This is the major source making the vibrational solvatochromism of a variety of nitrile-derivatized IR probes deviate from the prediction by 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\textsuperscript{20,24} based on the hybrid variational–perturbational interaction energy decomposition scheme (EDS),\textsuperscript{46–49} or shortly SoIEDS. 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 SoIEDS results could be useful to quantitatively test the chemical accuracy of the SoIEFPP method. Depending on the level of theory [either Hartree–Fock (HF) or second-order Møller–Plesset perturbation theory (MP2)], the SoIEDS frequency shift is given by a sum of different terms as

$$\Delta\omega_{ij}^{\text{HF}} = \Delta\omega_{ij}^{(10)} + \Delta\omega_{ij}^{(12)} + \Delta\omega_{ij}^{(14)}$$

(16)

or

$$\Delta\omega_{ij}^{\text{MP2}} = \Delta\omega_{ij}^{(10)} + \Delta\omega_{ij}^{(20)} + \Delta\omega_{ij}^{(12)} + \Delta\omega_{ij}^{(24)}$$

(17)

where the Hartree–Fock frequency shift consists of three contributions: $\Delta\omega_{ij}^{(10)}$ represents the contribution from the Coulomb interaction between unperturbed solute and solvent charge densities (i.e., without including charge redistribution due to induction effects), $\Delta\omega_{ij}^{(12)}$ is the exchange–repulsion contribution, and $\Delta\omega_{ij}^{(14)}$ 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, $\Delta \omega_{\text{disp}}^{(2)}$ as well as the electron correlation corrections to the Coulomb and exchange–delocalization effects, denoted as $\Delta \omega_{\text{el}}^{(1)}$ and $\Delta \omega_{\text{el}}^{(2)}$, respectively, contributes 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 $\Delta \omega_{\text{HF}}^{(2)}$ part of $\Delta \omega_{\text{MP2}}^{(2)}$ in eqn (17) quantitatively differs from $\Delta \omega_{\text{HF}}^{(2)}$ obtained at the HF level of theory, though the same notation $\Delta \omega_{\text{HF}}^{(2)}$ is used in both eqn (16) and (17).

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

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

$$\Delta \omega_{\text{Stark}}^{\text{Stark}} = -\mathbf{a}_j \cdot \mathbf{F}(\mathbf{r}_0) - \frac{1}{2} \mathbf{a}_j \cdot \mathbf{F}$$

(18)

where the set of induced dipoles $\mathbf{a}_j$ are given above in Section 2.2. Here, it should be noted that the induced dipole itself depends on the direction and strength of the electric field also, so that it is not possible to re-write eqn (18) simply as a product of the Stark tuning rate and the external electric field that are separately dependent on solute molecule and solvent charge distribution, respectively. However, if only the polarizability projected on the bond axis of nitrile, which is denoted as $\rho_j \rho_0$, is taken into consideration, the electric field-induced vibrational frequency shift is approximately given as

$$\Delta \omega_{\text{Stark}}^{\text{Stark}} \approx -\mu_j \cdot \mathbf{F}(\mathbf{r}_0)$$

(19)

where the modified Stark tuning rate, which includes the induction term approximately, is given as $\mu_j = \mu_j + (1/2)(\rho_j \rho_0)$. This is the Stark-dipole theory proposed recently by Fried et al.\(^{50}\) Later in Section 3, we shall present quantitative analysis results with the above approximate expression in eqn (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 (Fig. 1) – note further that the MeSCN probe is theoretically more convenient to study because, unlike in the MeCN case,\(^{51}\) its CN stretching 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 stretching mode of MeSCN were calculated by using SolCamm technique proposed earlier.\(^{43}\) In Table 1, the vibrational solvatochromic (molecular not distributed) dipoles and quadrupoles of MeSCN calculated using a few different methods are directly compared. Our ab initio calculated values of the Stark dipole of the CN stretching mode are in the range from 0.2 to 0.5 cm\(^{-1}\) (MV cm\(^{-1}\))\(^{-1}\), which is close to the experimental value of MeSCN in 2-methyltetrahydrofuran glass at 74 K (0.55–0.64 cm\(^{-1}\) (MV cm\(^{-1}\))\(^{-1}\)).\(^{39}\) Evidently, our MP2 estimation somewhat underestimates the value of the Stark dipole (0.2 cm\(^{-1}\) (MV cm\(^{-1}\))\(^{-1}\)) and the method for treating the electron correlation more accurately is needed (CCSD gives the result 0.31 cm\(^{-1}\) (MV cm\(^{-1}\))\(^{-1}\), closer to the experimental value). This might also be a result of the basis sets used for such calculations being too small, but we did not extend it beyond 6-311+G** because of resource limitations. Nevertheless, except for MP2 results, the SolCamm/6-311+G** 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 distributed vibrational solvatochromic charges that were obtained from multivariate least squares fitting analyses of the HF and DFT (density functional theory) calculation results from a number of MeSN–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 Fig. 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 Fig. S2, ESI\(^{+}\) for similar maps obtained using other methods). Clearly, the distribution of this electric potential around MeSCN differs from that of the vibrational solvatochromic dipole (Fig. 2(b)). Interestingly, if the vibrational solvatochromic quadrupole contribution is additionally taken into account (Fig. 2(c)), the corresponding electric potential around the H-bond accepting the N-atom of the

<table>
<thead>
<tr>
<th>Method</th>
<th>$\mu_{\text{CN str.}}$ [cm(^{-1}) (MV cm(^{-1}))(^{-1})]</th>
<th>$\Theta_{\text{CN str.}}$ [cm(^{-1}) (MV cm(^{-1}))(^{-1})] \times 10(^{-8})</th>
</tr>
</thead>
<tbody>
<tr>
<td>SolCamm/6-311+G**</td>
<td>0.47</td>
<td>1.16</td>
</tr>
<tr>
<td>MP2</td>
<td>0.20</td>
<td>0.96</td>
</tr>
<tr>
<td>CCSD</td>
<td>0.31</td>
<td>1.06</td>
</tr>
<tr>
<td>B3LYP</td>
<td>0.43</td>
<td>1.26</td>
</tr>
<tr>
<td>Fitting</td>
<td>0.54</td>
<td></td>
</tr>
<tr>
<td>B3LYP (\rho)</td>
<td>0.43</td>
<td>1.02</td>
</tr>
</tbody>
</table>

*Ref. 41. 5 Ref. 18. 6 Ref. 39. 7 Local field correction factor is assumed to be in the range of 1.1–1.3. 8 The average value of the traceless quadrupole was estimated from the formula: $\sqrt{Q_{xx}^2 + Q_{yy}^2 + Q_{zz}^2 + 3(Q_{xx}^2 + Q_{yy}^2 + Q_{zz}^2)}$. Note that in ref. 18 and 43 the traceless quadrupole is defined according to Jackson\(^{52}\) whereas we use Buckingham’s convention here\(^{31}\) (to convert value of traceless quadrupole element $Q_{zz}$ from Jackson’s to Buckingham’s convention multiply it by 1/2).
nitrile changes its sign (compare Fig. 2(b) with Fig. 2(c)). Note that this important change in the distribution of the 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 the nitrile stretching 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 the HF/6-311++G** level of theory. For the sake of comparisons, we also considered fifteen MeSCN–DMSO clusters, in which DMSO is not a protic solvent molecule. The SolEFP-calculated frequency shifts were 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 (Fig. 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 the H-bond distance between the nitrile’s N atom and the water H atom (see Fig. 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** (Fig. 4). Starting from the optimized geometry of the MeSCN–H$_2$O 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⋯OH⋯H was in-plane rotated by angle $\theta$ (with 1° interval) around the center of mass of H$_2$O.
Thus, we considered 25,200 such MeSCN⋯HOH configurations in total and the resulting SolEFP frequency shifts are plotted in Fig. 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, $\Delta_0^{\text{Coul}}$, 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, $\Delta_0^{\text{Coul}}$, 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 stretching modes. In particular, if the nitrile experiences a strong steric (repulsive) interaction with the surrounding molecules or chemical groups, most of the previous vibrational solvatochromism models based on purely electrostatic interaction-induced effects may not be quantitatively reliable.

### 3.3 CN stretching in bulk solutions

As shown in Fig. 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 Fig. 5. Because the application of SolEFP theory is the central issue here, we did not attempt to re-parameterize any force field parameters of the solvent molecules. The four solvents are aprotic non-polar $\text{CCl}_4$, aprotic and weakly polar $\text{CHCl}_3$, aprotic and strongly polar DMSO, and protic polar $\text{H}_2\text{O}$. 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 Fig. 5. The total frequency shifts calculated using SolEFP and MD simulation methods are black diamonds in the same figure (see Table S1, ESI† for numerical values of the data presented in Fig. 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 $\text{ab initio}$ semi-empirical maps for the nitrile frequency shift cannot describe the relative blue shift of the nitrile stretching 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 Fig. 5, the exchange–repulsion contribution, \( \Delta \alpha_{\text{ex}} \), 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 Fig. 5). To further understand the H-bond strength-dependence of the nitrile frequency shift, we performed first-principles SolEEDS calculations for Me(S)CN–X dimers with X being H\(_2\)O, MeOH or CF\(_3\)CH\(_2\)OH, which are presented in Table 2 (the corresponding molecular structures are depicted in Fig. S7, ESI†). The exchange–repulsion interaction contributes to the very strong frequency blue shift \( \Delta \alpha_{\text{ex}} \). Furthermore, the magnitude of \( \Delta \alpha_{\text{ex}} \) increases as the H-bond length of the X molecule increases. This can be understood by noting that the H-bond becomes stronger when the solute and solvent molecules are brought closer to each other. This in turn causes a sharp increase in the exchange–repulsion contribution to the frequency shift (see eqn (5) and Fig. 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 \( \Delta \alpha_{\text{corr}} \), the Coulombic effects on the nitrile stretching 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 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 the 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 for 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 (Fig. 5). Note also that in the case of CCl\(_4\), the Coulomb contribution completely vanishes, but yet the induction contribution is not negligible. In Fig. 6, we plot the distributions of different frequency shift components separately with the full width at half maximum (FWHM) values listed in Table 3. The distributions (blue curve in Fig. 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

![Table 2](Image)

**Table 2** The nature of the frequency shifts of the CN stretching mode in various H-bonding complexes involving MeCN or MeSCN and H\(_2\)O, MeOH and CF\(_3\)CH\(_2\)OH (structures are shown in Fig. S7 in ESI). The frequency shift partitioning was performed by using the SolEEDS//MP2/6-311++G** method. Bond lengths are in Bohr and frequency shifts in cm\(^{-1}\). The details about the SolEDS method can be found in Section 2.2 and ref. 20 and 24

<table>
<thead>
<tr>
<th></th>
<th>MeCN</th>
<th>MeCN</th>
<th>MeSCN</th>
<th>MeSCN</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>H(_2)O</td>
<td>MeOH</td>
<td>CF(_3)CH(_2)OH</td>
<td>H(_2)O</td>
</tr>
<tr>
<td>( R_{\text{H-bond}} )</td>
<td>3.9814</td>
<td>3.9570</td>
<td>3.7894</td>
<td>3.9569</td>
</tr>
<tr>
<td>( \Delta \alpha_{\text{ex}} )</td>
<td>-7.32</td>
<td>-8.71</td>
<td>-19.45</td>
<td>-14.08</td>
</tr>
<tr>
<td>( \Delta \alpha_{\text{ex}}^{(12)} )</td>
<td>+32.31</td>
<td>+36.63</td>
<td>+44.36</td>
<td>+46.06</td>
</tr>
<tr>
<td>( \Delta \alpha_{\text{ex}}^{(20)} )</td>
<td>-10.49</td>
<td>-11.29</td>
<td>-17.97</td>
<td>-14.06</td>
</tr>
<tr>
<td>( \Delta \alpha_{\text{disp}} )</td>
<td>-10.63</td>
<td>-13.51</td>
<td>-15.83</td>
<td>-12.64</td>
</tr>
<tr>
<td>( \Delta \alpha_{\text{corr}}^{(12)} )</td>
<td>+4.55</td>
<td>+3.48</td>
<td>+17.76</td>
<td>+7.80</td>
</tr>
<tr>
<td>( \Delta \alpha_{\text{corr}}^{(20)} )</td>
<td>+6.79</td>
<td>+8.40</td>
<td>+3.89</td>
<td>+6.23</td>
</tr>
<tr>
<td>( \Delta \alpha_{\text{Me2}} )</td>
<td>+15.21</td>
<td>+15.00</td>
<td>+12.76</td>
<td>+13.85</td>
</tr>
<tr>
<td>( \Delta \alpha_{\text{HL}}^{(12)} + \Delta \alpha_{\text{HL}}^{(20)} )</td>
<td>-2.77</td>
<td>-5.23</td>
<td>-1.69</td>
<td>-6.28</td>
</tr>
<tr>
<td>( \Delta \alpha_{\text{ex}}^{(12)} + \Delta \alpha_{\text{ex}}^{(20)} )</td>
<td>+39.10</td>
<td>+45.03</td>
<td>+48.25</td>
<td>+46.83</td>
</tr>
<tr>
<td>( \Delta \alpha_{\text{corr}}^{(12)} + \Delta \alpha_{\text{corr}}^{(20)} )</td>
<td>-21.12</td>
<td>-24.80</td>
<td>-33.80</td>
<td>-26.70</td>
</tr>
<tr>
<td><strong>Full QM</strong></td>
<td>+12.83</td>
<td>+12.58</td>
<td>+8(^{a})</td>
<td>+10.54</td>
</tr>
</tbody>
</table>

\(^{a}\) Partially optimized due to the energy optimization problem.
contribution (red curve) increases. However, the dispersion contribution remains more or less the same. From the present SolEFP results shown in Fig. 5 and 6, it is quite clear now that the nitrile frequency shift cannot be fully described by the contributions, $D_{\text{Coul}}$ and $D_{\text{Ind}}$, that are linearly dependent on the 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 the EFP2–EFP2 interaction potential and the 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 strongly by such a difference. Note that this repulsion contribution is extremely sensitive to a minute change in the solute–solvent distance when they are in close proximity (Fig. 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.\cite{54,55}

Therefore, any deviation induced by slightly less accurate interaction potential models could cause a large error in estimating the 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 intermolecular dispersion.

Nevertheless, our analysis is consistent with the previous empirical calculation of the CN stretching mode of acetonitrile,\cite{12} in which Reimers and Hall found, by applying the solvent descriptor model of Fawcett et al.,\cite{56} 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 the CN mode of

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

<table>
<thead>
<tr>
<th>MeSCN</th>
<th>$\Delta\omega_{\text{Coul}}$</th>
<th>$\Delta\omega_{\text{Ind}}$</th>
<th>$\Delta\omega_{\text{Disp}}$</th>
<th>$\Delta\omega_{\text{Ex-Rep}}$</th>
<th>$\Delta\omega_{\text{SolEFP}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>CCl$_4$</td>
<td>4.0 ± 0.5</td>
<td>1.4 ± 0.5</td>
<td>5.6 ± 1.0</td>
<td>21 ± 4</td>
<td>4.0 ± 0.5</td>
</tr>
<tr>
<td>CHCl$_3$</td>
<td>12.0 ± 0.5</td>
<td>2.5 ± 1.5</td>
<td>5.1 ± 1.0</td>
<td>22 ± 4</td>
<td>12.0 ± 0.5</td>
</tr>
<tr>
<td>DMSO</td>
<td>16.4 ± 0.5</td>
<td>4.8 ± 1.5</td>
<td>6.4 ± 1.0</td>
<td>25 ± 4</td>
<td>16.4 ± 0.5</td>
</tr>
<tr>
<td>H$_2$O</td>
<td>18.2 ± 0.5</td>
<td>12.9 ± 3.0</td>
<td>6.2 ± 1.5</td>
<td>27 ± 5</td>
<td>18.2 ± 0.5</td>
</tr>
</tbody>
</table>

**Fig. 6** Static frequency shift distributions of the CN stretching mode in MeSCN dissolved in four different solvents at room temperature, which were obtained by applying the SolEFP model of the MeSCN CN stretching mode to MD simulation trajectories.
acetonitrile in water were found to be $-6.4, -9.7$ and $+6.7$ cm$^{-1}$, whereas in CCl$_4$ they were $-1.9, -13.0$ 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$^{13}$ 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 eqn (19). The evaluation of the solvatochromic induced dipole moment is straightforward within the SolEFP theory:

$$\mathbf{\mu}_{\text{Ind}} = \sum_{a} \mathbf{\mu}_{\text{Ind},a}$$

(20)

where the summation runs over all polarizable sites in the system (including solvent sites). Note that the solvatochromic induced dipole moment is independent of 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:

$$\mathbf{\Theta}_{\text{Ind}} = \sum_{a} \left\{ \mathbf{\mu}_{\text{Ind},a} \otimes (\mathbf{r}_{\text{CN}} - \mathbf{r}_a) + (\mathbf{r}_{\text{CN}} - \mathbf{r}_a) \otimes \mathbf{\mu}_{\text{Ind},a} \right\}$$

(21)

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

In Table 4, we compare the exact induction frequency shifts that can be calculated using eqn (12) with (i) those obtained by considering time-averaged vibrational solvatochromic induced dipoles in eqn (20) interacting with the instantaneous electric field evaluated at the CN mid-bond, i.e., $\Delta\omega_{\text{Ind},\text{CN str.}} \approx \langle -\frac{1}{2} \mathbf{\mu}_{\text{Ind},\text{CN str.}} \rangle \mathbf{F}_{\text{CN}}$; and (ii) those obtained by considering time-averaged vibrational solvatochromic induced dipoles distributed over 19 different locations in MeSCN molecules that interact with the instantaneous electric fields at those polarizable sites, i.e., $\Delta\omega_{\text{Ind},\text{CN str.}} = \langle -\frac{1}{2} \mathbf{\Theta}_{\text{Ind},\text{CN str.}} \rangle \mathbf{F}$ (see eqn (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 ($\Delta\omega_{\text{Ind},\text{CN str.}}$ in Table 4) cannot be reproduced using 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 stretching mode (Fig. 8 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 the 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 a 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$^{50}$ 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 (Fig. 7(a)) and a large cluster taken from the MD trajectory, we calculated the instantaneously distributed vibrational solvatochromic dipoles ($\mathbf{a}_j$ in eqn (12)) and depict them with arrows. Note that quite a few dipoles at different sites contribute to $\Delta\omega_{\text{Ind},\text{CN str.}}$, 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 $\Delta\omega_{\text{Ind},\text{CN str.}}$ cannot be simply estimated by considering the induced dipoles of MeSCN only. We emphasize that the SolEFP induction theory fully takes into account such a non-additive nature for calculating $\Delta\omega_{\text{Ind}}$ 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 the nitrile stretching mode not only because it does not include other important contributions from

<table>
<thead>
<tr>
<th>Frequency shifts</th>
<th>Molecular solvatochromic multipoles</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\langle</td>
</tr>
<tr>
<td>Vacuum</td>
<td>0.0</td>
</tr>
<tr>
<td>CCl$_4$</td>
<td>-0.7</td>
</tr>
<tr>
<td>CHCl$_3$</td>
<td>-5.6</td>
</tr>
<tr>
<td>DMSO</td>
<td>-10.6</td>
</tr>
<tr>
<td>H$_2$O</td>
<td>-18.5</td>
</tr>
</tbody>
</table>

Paper PCCP


This journal is © the Owner Societies 2016
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 the protein framework

In the above subsections, we showed that the vibrational solvatochromism of CN probes originates not only from the electrostatic interaction with the solvent electric field but also from other short-range effects that cannot be directly correlated with the solvent electric field. In particular, the exchange–repulsion contribution caused by the steric contacts of the IR probe with the surrounding molecules or chemical groups leads to a strong blue shift of the nitrile stretching modes and the dispersion interaction-induced term makes it red-shifted. Such a 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 the 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 an *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 a 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 stretching 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 whose structures can be found in Fig. S9 in ESI † We assume that the vibrational frequency shift induced by neighboring peptide groups could be described by the following two factors: (i) Coulomb 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 the change of relative configuration of the covalently bonded chemical group such as methylene.

The first effect (i) can be modeled by considering the interaction (eqn (3)) of MeSCN vibrational solvatochromic multipoles with atomic partial charges of the CONH group. The second effect (ii) cannot be directly quantified, but we could approximately estimate its nature by performing additional analyses of the CN stretching mode frequencies of several MeSCN analogs (Fig. 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 stretching mode frequency is red-shifted by about 3 cm⁻¹. Thus, the overall frequency shift of Prot-SCN due to the local side-chain configuration of Prot-SCN such as proteins with cyanylated cysteine residues and the two neighboring amide groups of the cyanylated cysteine can be approximately written as

\[
\Delta \omega_{\text{Prot-SCN}} \approx \Delta \omega_{\text{MeSCN-CONH}} + \Delta \omega_{\text{Through-bond}} \tag{22}
\]

where \(\Delta \omega_{\text{MeSCN-CONH}}\) is evaluated from the interaction of SolCAMM with ESP charges fitted on the N-methylacetamide (NMA) peptide group (see Section IX in ESI † for details). Using the above scheme, we could calculate the vibrational frequency shifts and compare them with *ab initio* calculation results shown in Fig. 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 Fig. 8(a) when the methyl group of MeSCN is replaced with ethyl or even a 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-SCNs.

The next and perhaps the more difficult task is to model the vibrational solvatochromic influences of the surrounding
amino-acid side chains in proteins 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 needed. Here we use the following procedure: (i) the aliphatic group of the side chain is represented by an appropriately superimposed methane; (ii) the aromatic side chain of phenylalanine is by benzene; (iii) the 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) the peptide bond is modeled by NMA. In Table 5, the model chemical groups representing twenty amino acids are summarized and we obtained the EFP2 fragments 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 fragments 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– dispersion 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 CH4 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

\[ \Delta \omega_{\text{SolEFP}} \approx \left( \sum_{X} \Delta \omega_{\text{Solute-X}}^{\text{Coul}} + \Delta \omega_{\text{Solute-X}}^{\text{Ex-Rep}} + \Delta \omega_{\text{Solute-X}}^{\text{Disp}} \right) + \Delta \omega_{\text{Ind}} \]

(23)

where

\[ \Delta \omega_{\text{Ind}} \approx \Delta \omega_{\text{Ind-Polar EFP2s}}^{\text{Solute-X}} + \sum_{X} \Delta \omega_{\text{Ind-Solvent}}^{\text{Solute-X}} \]

(24)

To test the validity of the above approach, eqn (23) and (24), we specifically considered five dimeric complexes consisting of one MeSCN and serine, tyrosine, or lysine model peptide (see Fig. S10 in ESI†). First, we carried out ab initio harmonic vibrational analyses of those complexes to obtain the CN frequency shifts. Then, the fragment approach was employed with the SolEFP

Table 5 The minimalistic model of amino acid side chains, which is used in eqn (22). 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 the molecular environment. Peptide units are treated by NMA molecules except those being attached next to SCN probes (see the main text).

<table>
<thead>
<tr>
<th>Amino acid side-chain</th>
<th>EFP fragments</th>
<th>Model fragment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Asp</td>
<td>CH2COO−</td>
<td>Carboxyl group (deprotonated)</td>
</tr>
<tr>
<td>Asn</td>
<td>CHONH2−</td>
<td>Amide group</td>
</tr>
<tr>
<td>Glu</td>
<td>CH2COO−</td>
<td>Carboxyl group (deprotonated)</td>
</tr>
<tr>
<td>Gln</td>
<td>CHONH2−</td>
<td>Amide group</td>
</tr>
<tr>
<td>Ser</td>
<td>CH2OH−</td>
<td>C-OH group</td>
</tr>
<tr>
<td>Thr</td>
<td>CH3OH−</td>
<td>C-OH group</td>
</tr>
<tr>
<td>Val</td>
<td>CH2−</td>
<td>CH3H2</td>
</tr>
<tr>
<td>Trp</td>
<td>PhOH</td>
<td>Ar-OH group</td>
</tr>
<tr>
<td>Phe</td>
<td>CH4</td>
<td>CH3H2</td>
</tr>
<tr>
<td>Lys</td>
<td>CH3NH+</td>
<td>Amine group (protonated)</td>
</tr>
<tr>
<td></td>
<td>3 × CH4</td>
<td>C3H7, C3H8, C3H12</td>
</tr>
<tr>
<td>Arg</td>
<td>Methylguanidinium+</td>
<td>Guanidinium group (protonated)</td>
</tr>
<tr>
<td>Gly</td>
<td>2 × CH4</td>
<td>C3H7, C3H8, C3H12</td>
</tr>
<tr>
<td>His</td>
<td>Imidazole</td>
<td>Aromatic ring (N is not protonated)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C3H7</td>
</tr>
<tr>
<td>Met</td>
<td>(CH3)2S</td>
<td>Methyl sulfide group</td>
</tr>
<tr>
<td></td>
<td>CH2H2</td>
<td>C3H7</td>
</tr>
<tr>
<td>Cys</td>
<td>CH2SH</td>
<td>Thiol group</td>
</tr>
<tr>
<td>Pro</td>
<td>3 × CH4</td>
<td>Ring</td>
</tr>
<tr>
<td>Ala</td>
<td>CH4</td>
<td>C3H7</td>
</tr>
<tr>
<td>Leu</td>
<td>3 × CH4</td>
<td>2-Butyl chain</td>
</tr>
<tr>
<td>Ile</td>
<td>3 × CH4</td>
<td>2-Methyl-x-propyl chain</td>
</tr>
<tr>
<td>Val</td>
<td>2 × CH4</td>
<td>Iso-propyl chain</td>
</tr>
</tbody>
</table>
parameters of MeSCN and the EFP2s representing the serine, tyrosine, and lysine model peptides (see the details in Table S2 in ESI†). Surprisingly, the fragment approach reproduces both $\Delta_0^{\text{Disp}}$ and $\Delta_0^{\text{Ex-Rep}}$ very well. Our eqn (24) however, underestimates the induction effect (compare “Full” and “Frag” entries in Table S2, ESI†). 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 $\sim 1$ 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,† Section III). Nevertheless, eqn (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 stretching 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 Ras 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 (Fig. 9).$^{57}$ We performed umbrella sampling molecular dynamics simulations of the 6 mutants of RalGDS in water and the same mutants docked to the human oncoprotein p21$^{\text{Ras}}$ (Ras) mutant (here labeled as Ras$^\prime$; see Computations). Subsequently, from the MD trajectories, we have considered the local environments within 10 Å from SCN probes (see Fig. S11, ESI† 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 stretching modes in Ras$^\prime$GDS are given in Fig. S12 (ESI†), whereas those in Ras$^\prime$GDS–Ras$^\prime$ complexes are presented in Fig. 10.

In the cases of Ras$^\prime$GDS proteins without binding to Ras$^\prime$ (Fig. S12, ESI†), all the SCN probes are highly exposed to water so that the contributions from the protein environment to the SCN stretching frequency shifts are relatively small and are mainly Coulombic in nature. $\Delta_0^{\text{Ex-Rep}}$ 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 Fig. 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 Section 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 the water phase – note that $\Delta_0^{\text{Coul}}$ is strongly dependent on the angle between the SCN and the H-bonded water molecule.

The situation is much more complicated when the SCN probe is located at the interface of Ras$^\prime$GDS and Ras$^\prime$ proteins (Fig. 10). The calculated $\Delta_0^{\text{Ex-Rep}}$ values are still quite large, which range from $+24$ cm$^{-1}$ in Ras$^\prime$GDS/G28 + Ras$^\prime$ to $+40$ cm$^{-1}$ in Ras$^\prime$GDS/Y31 + Ras$^\prime$. Furthermore, the blue shifts have quite different origins depending on the specific position of the SCN probe. Non-bonding close-contact of SCN with the aliphatic or aromatic part of the amino acid side chain (Val, Leu, Ile or Phe) is the dominant interaction contributing to $\Delta_0^{\text{Ex-Rep}}$ in the case of G28 (see also a model MeSCN–CH$_2$ distance-dependence scan in Fig. S13 in the ESI†). On the other hand, in the cases of K32 and N29, the interaction of SCN with the protein backbone induces a substantial blue shift, where the H-bond between SCN and the amide proton and the repulsive interaction between peptide–oxygen and SCN–nitrogen atoms are important. In the case of Y31, the 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 the Ras$^\prime$ protein is not docked to Ras$^\prime$GDS, the 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 Ras$^\prime$GDS/G28 (without docked protein), SCN interacts with the Tyr residue via forming a H-bond with the OH group of Tyr. However, typical H-bond lengths, estimated from the MD trajectory, are longer than those of Ras$^\prime$GDS/Y31 + Ras$^\prime$ so that the exchange–repulsion interaction-induced frequency shift due to the tyrosine OH group is small (note that $\Delta_0^{\text{Ex-Rep}}$ is a very short-range effect and becomes vanishingly small as the solute–solvent distance increases; Fig. 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 Ras$^\prime$GDS.

Fig. 9  Locations of the SCN probes within the Ras$^\prime$GDS protein bound to Ras$^\prime$. 

G28  K32  N27  N29  N54  Y31

Ras  Ras$^\prime$GDS
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.

It is clear that both water and protein residues can cause a large blue shift of the SCN stretching 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. 57 since our Prot-SCN MD simulation systems considered in 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 show the importance of the exchange–repulsion and the dispersion interaction effects, which cannot be described by other electric field-based theories used previously. 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 the CN stretching 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 are 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/G28 + Ras', the repulsion frequency shift is smaller than 30 cm$^{-1}$. From this, it is shown that the frequency is blue-shifted because the H-bonding interactions of SCN with water molecules cannot be distinguished from 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 present (roughly 80–100 MD snapshots per system) so 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 the 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 the fragmentation approach to the calculation of the vibrational frequency shift, our estimated deviations are in a range from 3 to 7 cm$^1$ based on Table S2 (ESI†). We thus need to refine the approach to

![Fig. 10](image-url)
make it to the spectroscopic accuracy in the future. Nevertheless, it will be desirable to carry out further studies for comparatively small proteins to test the validity of the SolEFP approach developed here.

One of the related studies was reported by Zou et al.\cite{Zou18094} 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 the halothane molecule. They found that electrostatic forces are roughly 2–2.5 times larger in magnitude than van der Waals forces and they used the Stark dipole model for further analyses. The agreement with the 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 a 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. S3–S6, ESI\(^+\)]. Therefore, the actual importance of the non-Coulombic frequency shifts was not traced fully in the study of Zou et al.\cite{Zou18094} due to the cancellation of attractive and repulsive contributions that result in the frequency shifts of different natures.

Recently, Zhang et al. 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).\cite{Zhang18094} However, they showed that the vibrational response of their probe is governed by the local molecular environment around CN group, i.e., intermolecular H-bonds with the CN group and the intramolecular coupling of CN with the 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\cite{Kollman94,Go94,Go96} 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 a 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 the nitrile stretching 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 the CN stretching of MeSCN dissolved in CCl\(_4\), CHCl\(_3\), DMSO and H\(_2\)O, which proves that this approach is better than previous ones based on Couplin 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.

In addition, here we showed that the dispersion interaction between the polarizable CN group and the 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 the Ras guanine nucleotide dissociation stimulator (RasGDS) and its complexes with human oncoprotein p21\(^{\text{Ras}}\) (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 due to close contacts with the 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 techniques would then be of use to measure molecular granularity, dynamics, and conformational heterogeneity of IR probe-incorporated biomolecules.

## 5. Methodology

### 5.1 Computations

Geometry optimizations and vibrational analyses of model QM clusters of MeCN and MeSCN molecules were performed by using the Gaussian09 package.\cite{Gaussian09} Cubic anharmonic constants\cite{G03} were obtained by numerical differentiation of Hessian matrices as described in ref. 20. Hybrid variational–perturbational interaction energy calculations were performed by utilizing the GAMESS package\cite{GAMESS} modified by Robert W. Góra.\cite{G03} SolEFP calculations were made as in ref. 20. SolEFP parameters for the CN stretching mode of MeSCN were computed in a similar way as described previously.\cite{Zhang18094} In short, cumulative atomic multipole moments (CAMM),\cite{Go96} wavefunctions, Fock’s 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 0.006 and 0.084 Å, respectively. The Pipek–Mezey method\cite{Pipek83} was used to localize molecular orbitals including core electrons. Distributed polarizabilities centered at localized molecular
Hybrid QM/MM molecular dynamics simulations of MeSCN in H$_2$O, DMSO, CHCl$_3$ and CCl$_4$ were performed by using the Amber11 program. The Parm99 force field was utilized to describe the solvent whereas MeSCN was modeled by PM3 semi-empirical Hamiltonian. Force field parameters for CCl$_4$, CHCl$_3$ and DMSO were taken from the work of Fox and Kollman and water was treated by the TIP4P model. MeSCN was consecutively solvated by 817 CCl$_4$, 1309 CHCl$_3$, 1042 DMSO and 4615 H$_2$O 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. 30 000 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 20 000 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 the MM part of the systems were constrained using the SHAKE algorithm. To control the total temperatures (NVT) and pressures (NPT) a Langevin thermostat with coupling frequency of 0.1 ps and a Berendsen barostat with a 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. 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 that utilizes the Kabsch method implemented in the Biopython package. MD trajectories were interfaced with our in-house SolEFP code by using the MDAnalysis package. The vibrational Coulomb, induction, dispersion and exchange–repulsion interactions between MeSCN and solvent molecules were cut at 40, 17, 13 and 17 Bohr, respectively, to achieve the compromise between the speed and accuracy. Unfortunately, despite the fact that 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 Fig. S3–S6 and discussion in the ESI!).

Umbrella sampling molecular dynamics simulations were performed at the six SCN probe locations (N27C$_{SCN}$, G28C$_{SCN}$, N29C$_{SCN}$, Y31C$_{SCN}$, K32C$_{SCN}$ and N54C$_{SCN}$) of both free RafGDS and RafGDS + Ras D30E (RafGDS + Ras$^*$) complexes in water as has been discussed previously. In brief, the 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 for each protein system. The AMBER03 forcefield was used with the TIP3P water model as implemented in GROMACS. Note that the SCN probe in our simulations of MeSCN in bulk solvents from Section 3.3 was modeled using PM3 Hamiltonian.

5.2 Experiments

Absorption frequencies of CN stretching modes in MeCN and MeSCN dissolved in various solvents were measured by using a FTIR (Bruker VERTEX 70) spectrometer at 22 °C with 1 cm$^{-1}$ resolution. In the case of very weak solubility, the saturated solutions of Me(S)CN were prepared by ultrasonication of MeSCN in the mixture of two immiscible phases for 1 hour and subsequent separation of the sample phase using 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. F-1722). The authors gratefully acknowledge the Texas Advanced Computing Center (TACC) at The University of Texas at Austin and the Wrocław Centre for Networking and Supercomputing (WCSS) at Wroclaw University of Technology in Poland for providing high-performance computing resources that have contributed to the results within this paper.

Notes and references