SUPPORTING INFORMATION

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

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

[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*

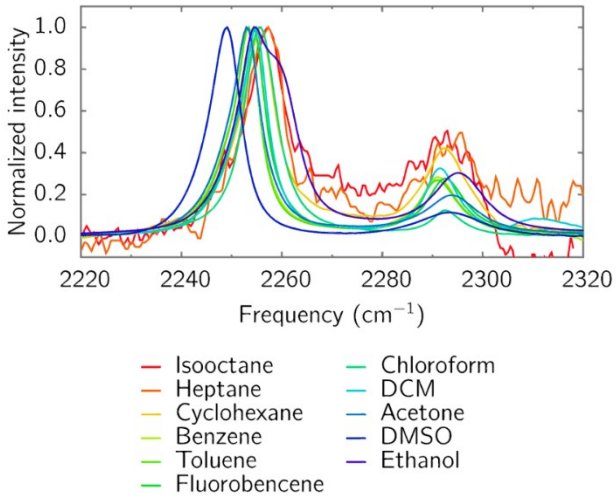
[c] *Department of Chemistry, Center for Nano- and Molecular Science and Technology, and Institute for Cell and Molecular Biology, The University of Texas at Austin, 105 E. 24th Street, STOP A5300, Austin, Texas 78712, U.S.A.*

Contents

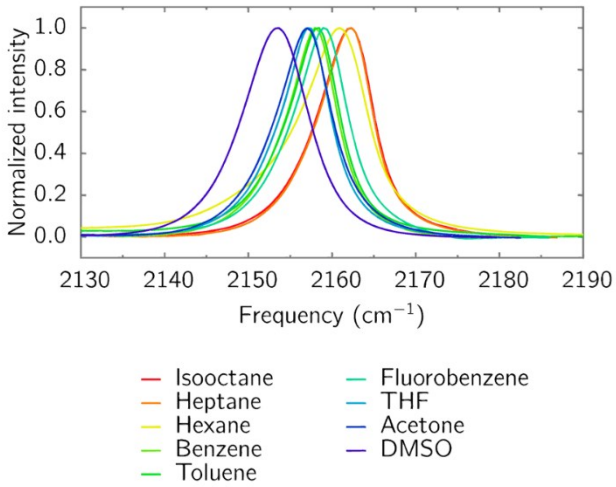
I. The CN stretch FTIR spectra of MeCN and MeSCN in various solvents

Figure S1. Experimentally measured FTIR spectra of the CN stretch mode. (a) Those of MeCN (see the low frequency bands in the spectra). (b) Those of MeSCN. Note that the high frequency bands in the spectra of MeCN are combination bands. Due to the low solubility of MeCN in isooctane and heptane, the S/N ratio is not good as compared to the other solutions. Nonetheless, it is still possible to estimate the average frequency of the CN stretch mode.

(a)



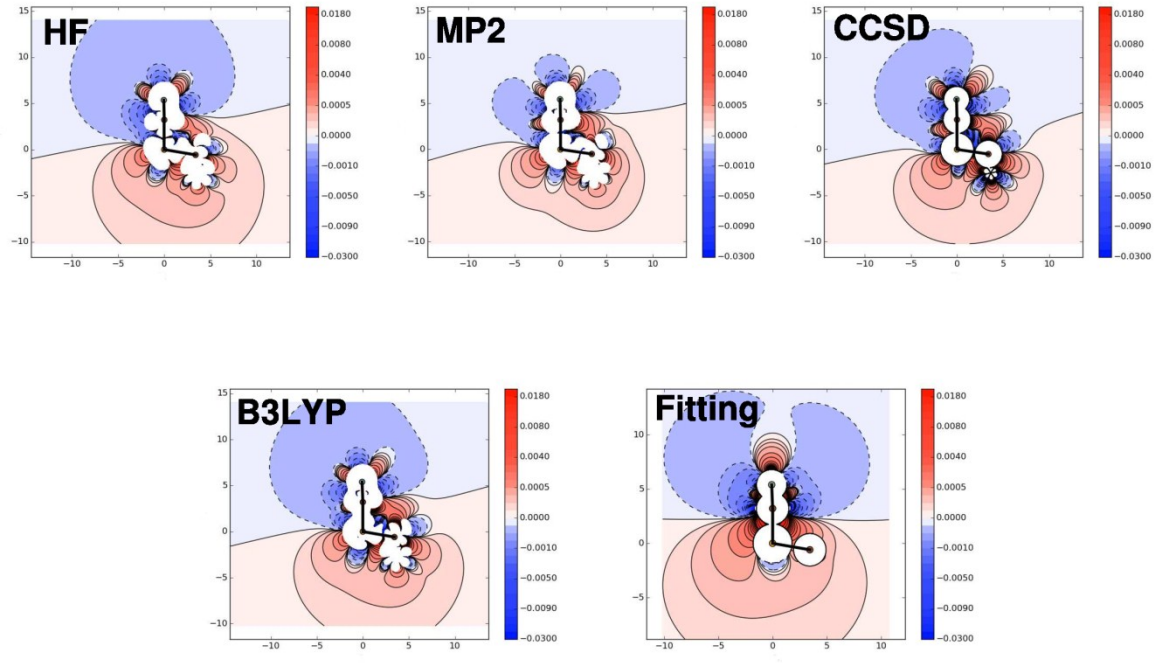
(b)



II. Comparison of electrostatic potentials computed from distributed solvatochromic multipole moments of MeSCN CN stretch mode by using various methods

In Figure S2, we plot the electrostatic potential produced by the set of vibrational distributed solvatochromic multipole moments by using *ab initio* solvatochromic (SolCAMM) methods with HF, MP2, CCSD and B3LYP (basis set is 6-311++G**) and the fitting, distributed-charge-based method described in Ref. 18 in the main text. The resulting distributions of the electrostatic potential are generally similar in each case. However, one can notice a somewhat weaker electrostatic potential at MP2 level. Note also that the potential produced by distributed charges ("Fitting") generates the vibrational Stark dipole moment which is almost collinear with the S-C-N bond axis (the angle of deviation is only 1°). In contrast, all the *ab initio* vibrational methods predict that Stark dipole forms an angle of about 11° with respect to the S-C-N bond axis. This is because the distributed charges that were obtained by multivariate least squares analysis method were constrained to the subset of points lying within the SCN group only, consequently neglecting the contribution due to the methyl group. Therefore, SolCAMM models take into account the bonding of SCN group with the carbon atom which may be a part of a protein.

Figure S2. 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 various methods. HF, MP2, CCSD and B3LYP data were computed by using SolCAMM method and 6-311++G** basis set whereas "Fitting" refers to the semi-empirically derived solvatochromic charges of the antenna model from Ref. 18 in the main text. Spatial dimensions are in Bohrs.



III. Approximate aspects of SolEFP theory

The SolEFP model developed by us is based on a few approximations:

1) The charge-penetration effects have been ignored, since they are likely to be weak. Nevertheless, since they were found to be of importance in describing Coulomb interaction energy (not frequency shift), such charge-penetration effects need to be studied in the future.
2) Due to the complexity of the SolEFP equations and to the difficulty in calculating all those terms, we here didn't take into account the electronic anharmonicity contributions (that are associated with the second derivatives of solute-solvent interaction potential terms with respect to normal coordinates), except for that associated with the Coulomb interaction-induced frequency shift. When the exchange-repulsion, induction, and dispersion terms in the frequency shift calculation were estimated, only the first derivatives of potential energy with respect to normal coordinates were considered (see the second term in Eq. (3) in Ref.[1]). We already showed that this approximation is quite acceptable for describing the carbonyl and CN stretch modes.[1, 2] However, for completeness it will be necessary to further test the validity of this approximation for other IR probes.
3) The exchange-repulsion interaction-induced frequency shift was based on the approximation that only a single exchange of electron pair between solute and solvent is considered. In fact, we believe that this is an excellent approximation because the direct comparisons of our SolEFP results with (completely *ab initio*) SolEDS calculation results indicate that the current SolEFP exchange-repulsion frequency shifts are quite accurate.
4) The induction and dispersion interaction-induced frequency shifts were treated by taking into consideration only the dipole-dipole interactions between LMO polarizable centers. Perhaps, the dispersion interaction-induced contribution originating from distributed quadrupole-dipole and quadrupole-quadrupole interactions could be of importance.
5) The vibrational solvatochromism theory in Eq. (3) in Ref.[1], which is the starting equation for all the subsequent SolEFP and SolEDS approaches, is based on the approximation that only the diagonal force constants (calculated in the gas-phase normal coordinate space) are affected by solvation. Any changes in the off-diagonal Hessian matrix elements were not fully taken into account. However, based on our SolEDS calculation studies over the years it is likely that this approximation is valid for those spatially localized vibrations such as CO or CN stretch modes.
6) Since the SolEFP theory is to calculate all the vibrational solvatochromism parameters that can be obtained from *ab initio* calculations of the solute molecule *in the gas phase*, the same structure should be used when the theory is applied to the solution systems with a solute molecule surrounded by solvent molecules. Therefore, there exists a certain ambiguity (difficulty) in perfectly superimposing it onto the solute molecule in solutions – note that the detailed molecular structures (bond lengths, bond

angles, and so on) of solute molecules in solutions are slightly different from that in the gas phase due to electronic structure change induced by solute-solvent intermolecular interactions. Nevertheless, it should be emphasized that the present vibrational solvatochromism theory correctly includes the effects from solvation-induced structural distortions since they are related to mechanical anharmonicity in the corresponding potential energy surface. Therefore, the above difficulty may not be the most crucial one, but still it needs to be investigated more in detail.

IV. Distance-dependence of the vibrational solvatochromic interaction between MeSCN and solvent molecules

Any vibrational frequency shift induced by solute-solvent interactions decays monotonically to zero as intermolecular distance, R, increases. Therefore, for successful and reliable applications of the SolEFP theory to condensed phase systems, a proper convergence of the particular frequency shift contributions when R is increased needs to be confirmed. However, in practice, the molecular dynamics (MD) simulation boxes are finite in size, which essentially leads to a problem of convergence, similar to the one with the conventional interaction energy calculations in MD. The use of periodic boundary conditions (PBC) enables to apply Ewald summation method for long-range electrostatic interactions. Here, we have not developed similar approach to the calculation of the electrostatic frequency shifts. To avoid this difficulty, we deliberately used very large simulation boxes with volume of around 90^3 Å and performed all the computations in the real space only.

Coulomb, induction, exchange-repulsion and dispersion interactions (and the corresponding frequency shifts) decay differently with respect to R. Therefore, it is important to estimate the optimal R_{Max}, i.e., the maximum (or cutoff) solute-solvent distance that is considered for a particular contribution to the frequency shift. In SolEFP theory, exchange-repulsion and induction contributions are the most time-consuming ones in terms of computational cost. Evaluation of $\Delta\omega^{\mathrm{Ind}}$ and **a'** requires inversions of large matrices many times per one MD step. The size of the corresponding matrices grows quadratically with the increase of the number of polarizable sites in the system. Computation of $\Delta\omega^{\mathrm{Rep}}$ requires calculations of millions of one-electron integrals and their derivatives with respect to the solute geometry. Therefore, a reasonable compromise between accuracy and speed needs to be made.

In the Figures S3-6 we present the detailed analysis of the convergence of the frequency shift contributions when R_{Max} increases. From those analyses we have chosen the R_{Max} to be 40, 17, 17 and 13 Bohrs, respectively, for Coulomb, exchange-repulsion, induction, and dispersion terms. While Coulomb terms are mostly well converged, the exchange-repulsion, dispersion and induction frequency shifts are not fully converged within the given cutoff distances in many cases. This will lead to certain small positive ($\Delta\omega^{\mathrm{Rep}}$) and negative ($\Delta\omega^{\mathrm{Ind}}$ and $\Delta\omega_6^{\mathrm{Disp}}$) deviations (errors) that are about ±3 cm^{-1}. We believe that those errors

cancel with each other to some extent so that the cutoff distance values 17 and 13 Bohrs are acceptable. Note that increasing R_{Max} by 1 Bohr results in a dramatic increase in computational time required to evaluate $\Delta\omega^{\mathrm{Ind}}$ and $\Delta\omega^{\mathrm{Rep}}$. With the above given cutoff distances, we achieved a reasonable time of 4-8 days that is needed to analyze 20,000 MD frames at 3.50 GHz Intel® Xeon® E3-1270 V2 computer (with parallelization on 4 CPU, each analyzing one MD frame at a time).

Figure S3. Cutoff distance-dependence of the CN stretch mode frequency shift in MeSCN/CCl_4 solution studied in this work. a) Coulomb, b) exchange-repulsion, c) induction and d) dispersion frequency shifts. We have analyzed 12 configurations taken every 1000 frames from the MD trajectory (simulation details in the main text). The sampled frequency shifts are coloured in blue-to-green to facilitate tracking the convergence of a particular sample. The average number of solvent molecules that correspond to a particular R_{Max} value is plotted by brown lines and circles.

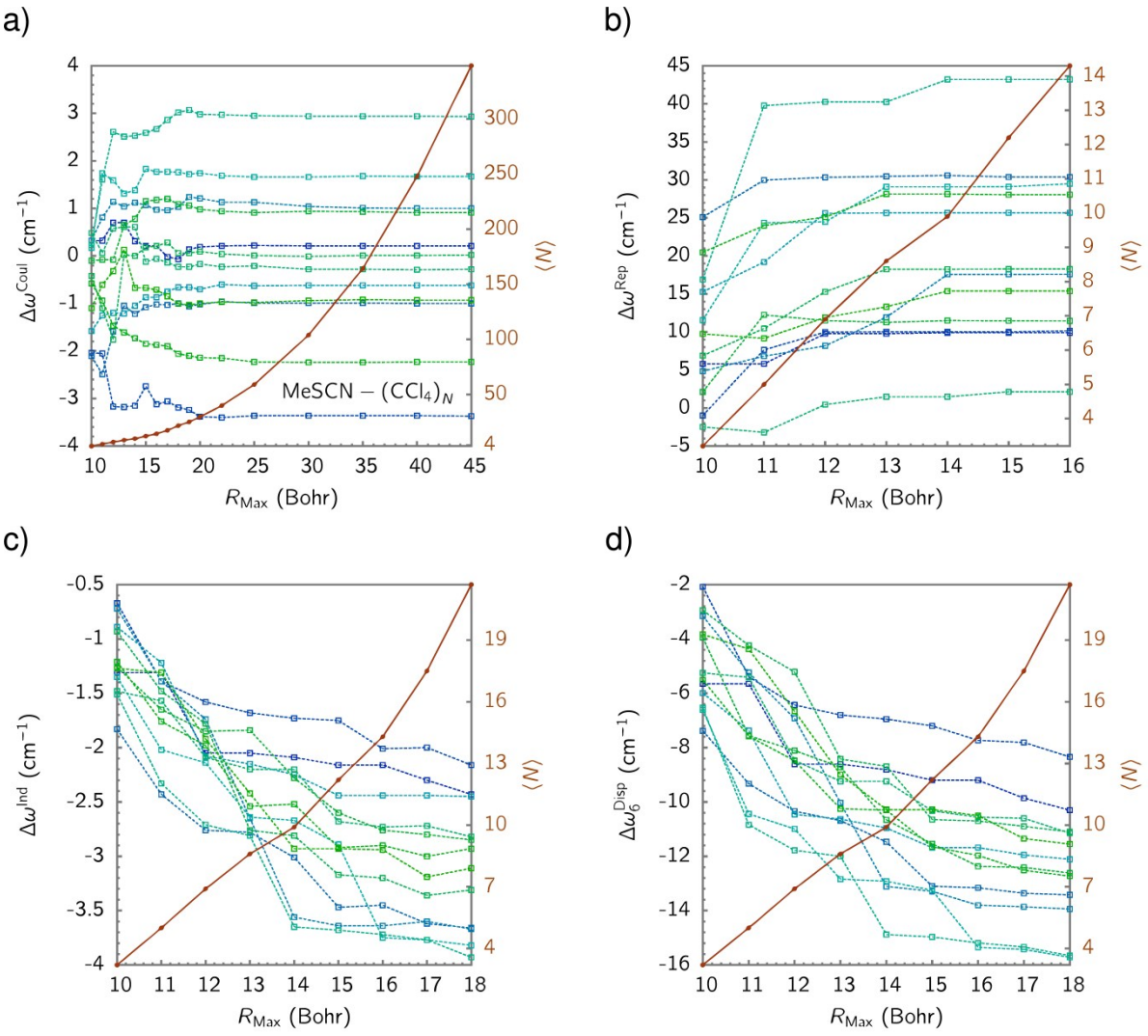


Figure S4. Cutoff distance-dependence of the CN stretch mode frequency shift in MeSCN/$CHCl_3$ solution studied in this work. a) Coulomb, b) exchange-repulsion, c) induction and d) dispersion frequency shifts. We analyzed 12 configurations taken every 1000 frames from the MD trajectory (simulation details are given in the main text). The sampled frequency shifts are coloured in blue-to-green to facilitate tracking the convergence of a particular sample. The average number of solvent molecules that correspond to a particular R_{Max} value is plotted by brown lines and circles.

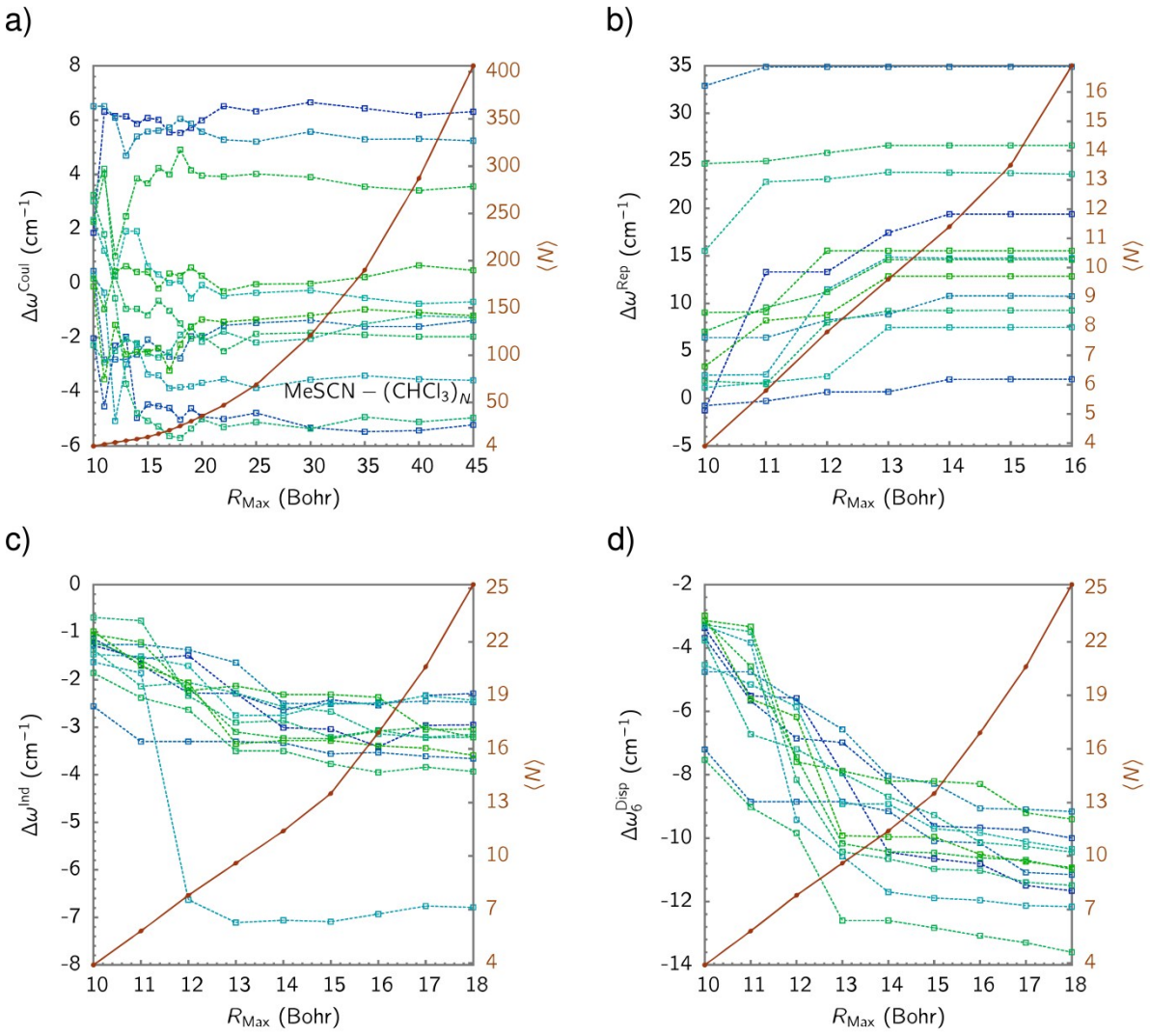


Figure S5. Cutoff distance-dependence of the CN stretch mode frequency shift in MeSCN/DMSO solution studied in this work. a) Coulomb, b) exchange-repulsion, c) induction and d) dispersion frequency shifts. We have analyzed 12 configurations taken every 1000 frames from the MD trajectory (simulation details in the main text). The sampled frequency shifts are coloured in blue-to-green to facilitate tracking the convergence of a particular sample. The average number of solvent molecules that correspond to a particular R_{Max} value is plotted by brown lines and circles.

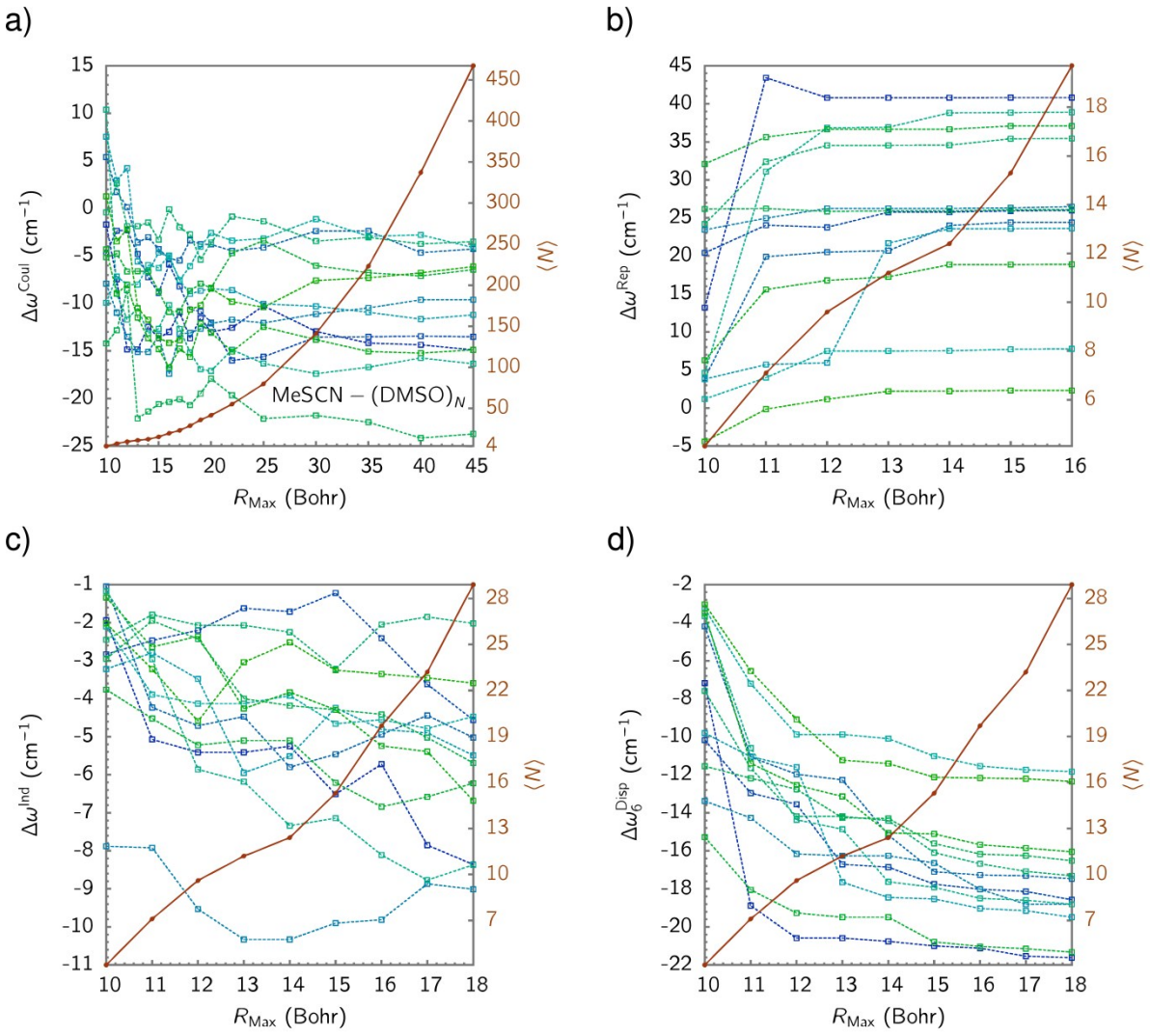
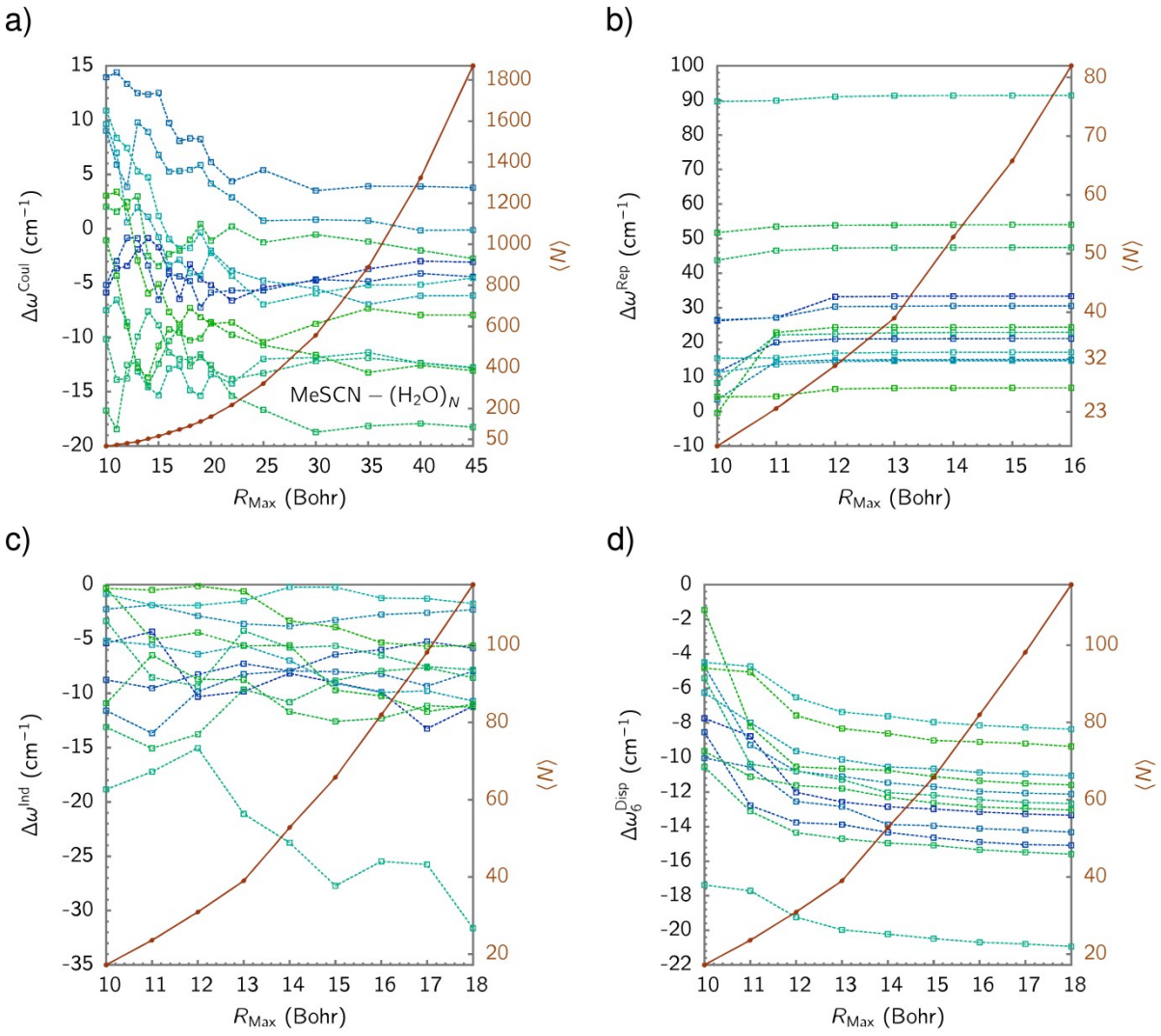
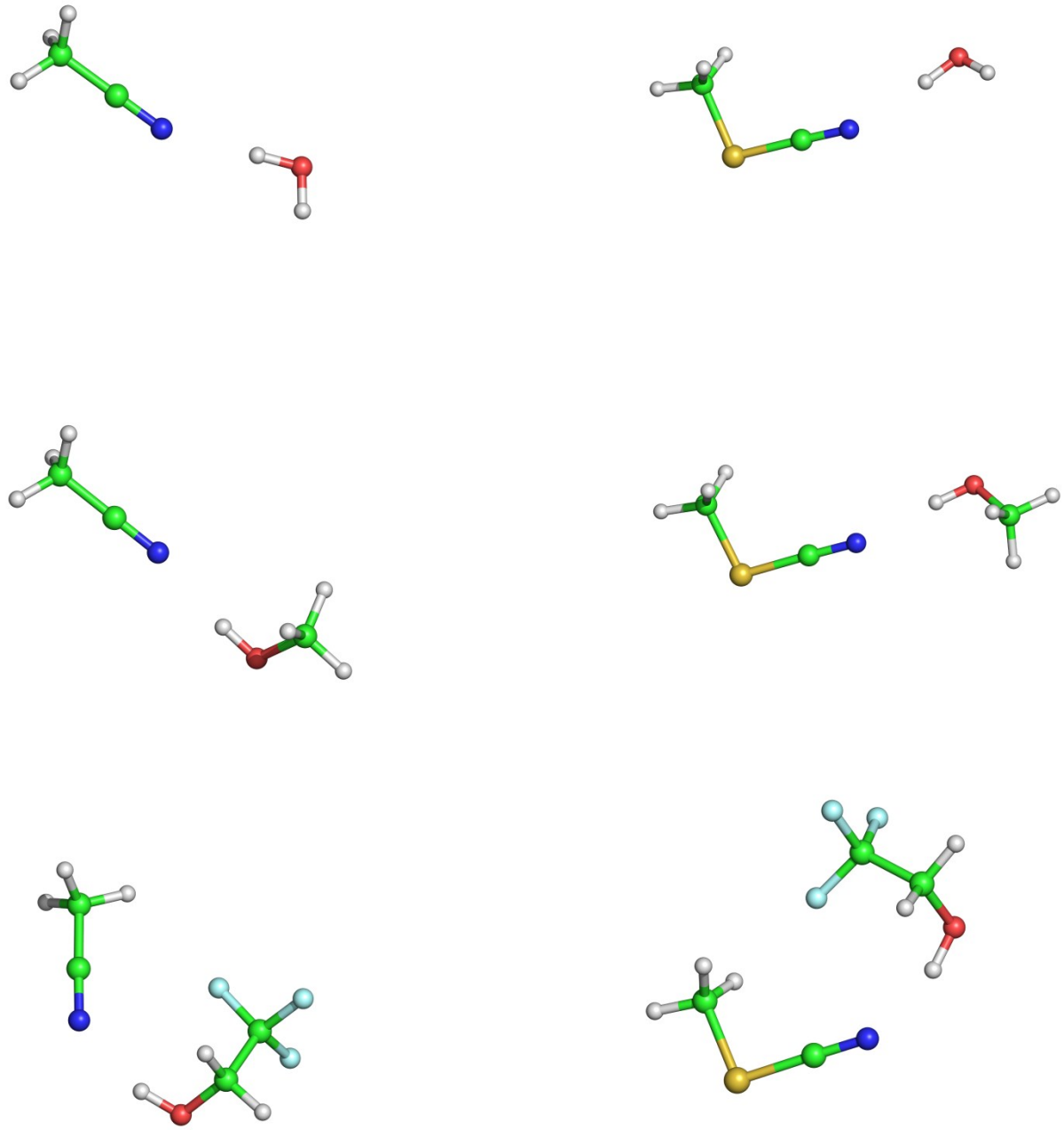


Figure S6. Cutoff distance-dependence of the CN stretch mode frequency shift in MeSCN/H_2O solution studied in this work. a) Coulomb, b) exchange-repulsion, c) induction and d) dispersion frequency shifts. We have analyzed 12 configurations taken every 1000 frames from the MD trajectory (simulation details in the main text). The sampled frequency shifts are coloured in blue-to-green to facilitate tracking the convergence of a particular sample. The average number of solvent molecules that correspond to a particular R_{Max} value is plotted by brown lines and circles.



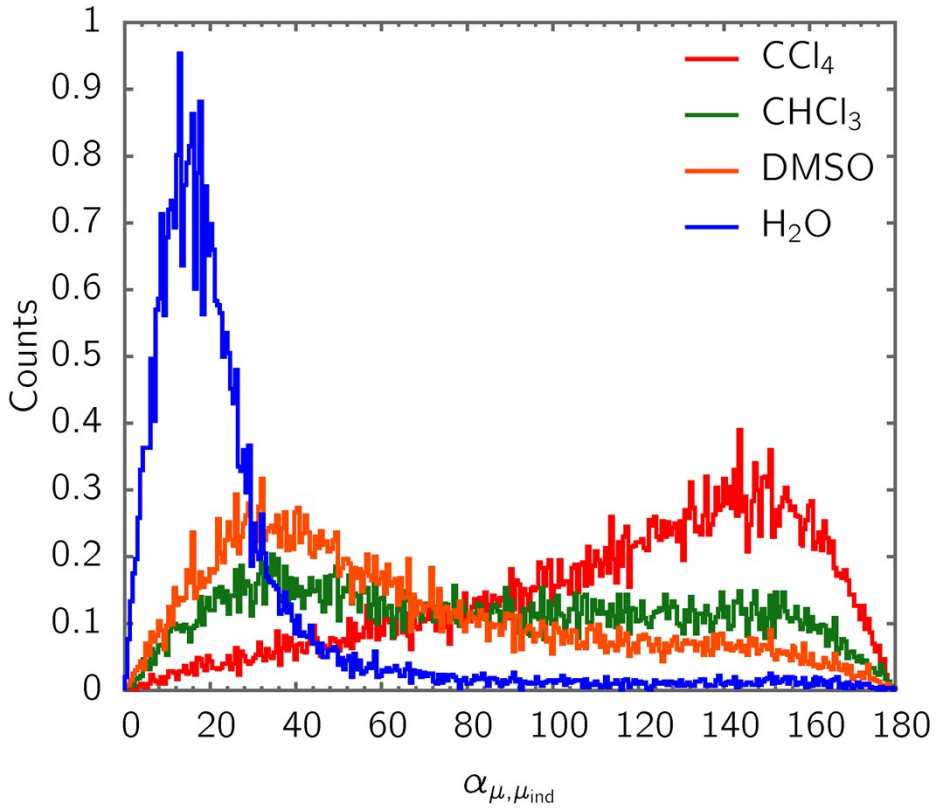
V. Structures of MeCN (MeSCN)-X (X=water, methanol, 2,2,2-trifluoroethanol) dimers

Figure S7.



VI. Directionality of the induced effect on the vibrational solvatochromic dipole of MeSCN nitrile stretch mode in the bulk solutions

Figure S8. The distributions of the angles between the induced molecular solvatochromic dipole moments of CN stretch mode in MeSCN with the static solvatochromic dipole, centered at CN mid-bond. The analysis was performed by applying the SolEFP model of MeSCN CN stretch mode to MD simulation trajectories. Angles are given in degrees.



VII. Numerical values of the data from Figure 5 in the main text

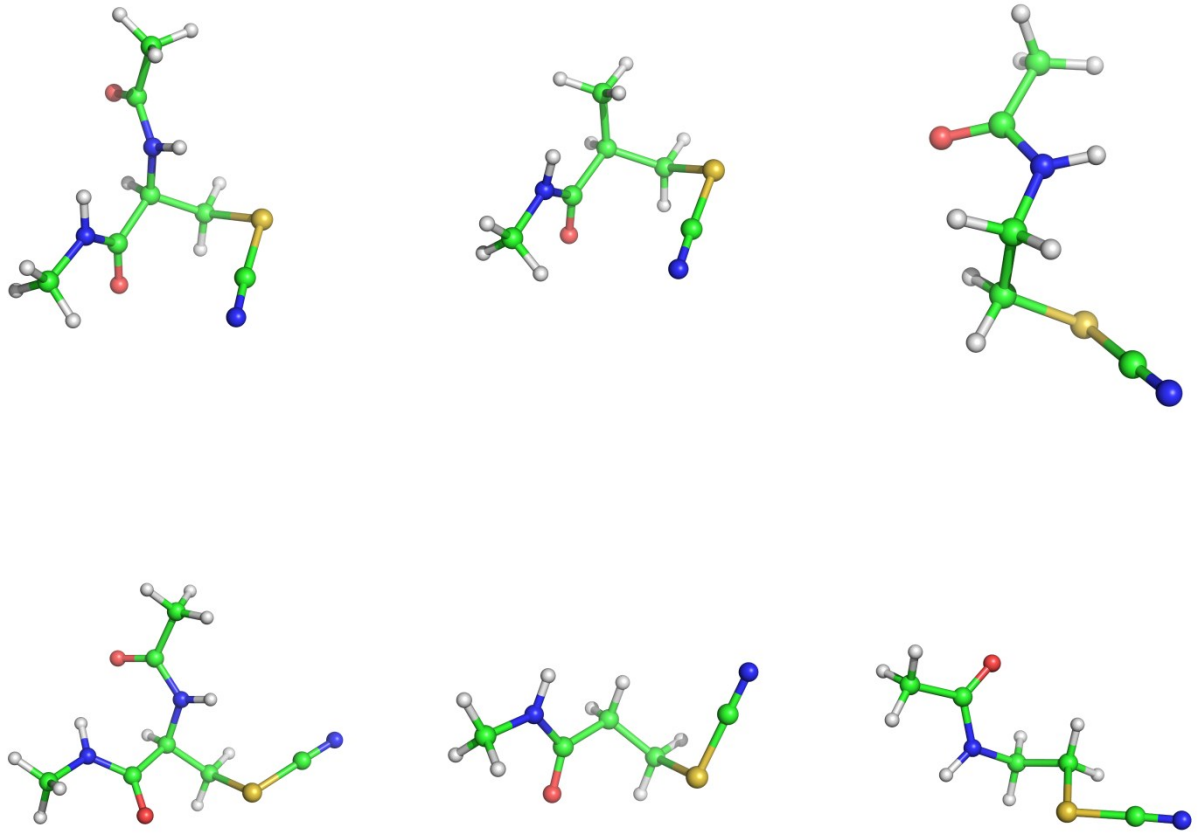
Table S1. Frequency shifts are labeled as in Figure 5. The numerical uncertainties of SolEFP frequency shifts are given as rough estimates of errors associated with SolEFP/MD method. This does not include the variation in qualities of MD forcefields used for describing solvents. All values in cm^{-1}.

	CCl_4	$CHCl_3$	DMSO	H_2O
Coulomb	−0.2 ± 1.0	−3.6 ± 1.0	−7.6 ± 1.0	−8.6 ± 1.0
Induction	−3.3 ± 1.0	−3.8 ± 1.0	−5.2 ± 1.0	−10.1 ± 1.0
Dispersion	−12.6 ± 1.0	−11.8 ± 1.0	−16.2 ± 1.0	−12.1 ± 1.0
Repulsion	+17.8 ± 5.0	+16.5 ± 5.0	+19.4 ± 5.0	+26.1 ± 5.0
Dipole	−0.7 ± 1.0	−6.0 ± 1.0	−11.2 ± 1.0	−23.1 ± 1.0
Multipoles	−3.5 ± 2.0	−7.4 ± 2.0	−12.8 ± 2.0	−18.7 ± 2.0
TOTAL[a)]	**+1.7 ± 8.0**	**−2.7 ± 8.0**	**−9.6 ± 8.0**	**−4.7 ± 8.0**
Exp.	**−10.0 ± 1.0**	**−11.8 ± 1.0**	**−17.8 ± 1.0**	**−9.0 ± 1.0**

a) TOTAL denotes the sum of “Coulomb”, “Induction”, “Dispersion” and “Repulsion”.

VIII. Structures of a few Prot-SCN model compounds

Figure S9.



IX. Fitting of atomic charges located at the peptide group of NMA

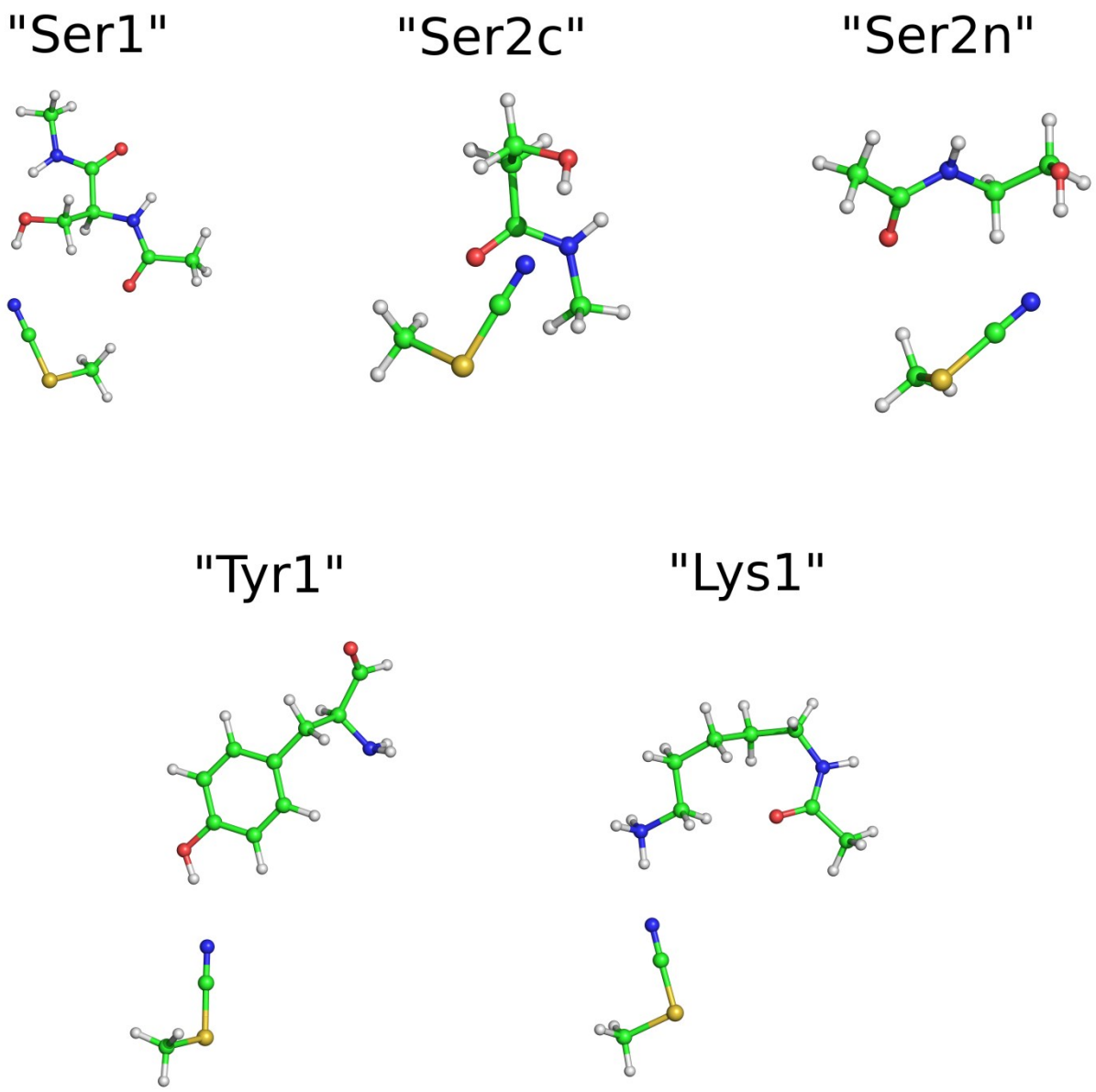
NMA molecule was optimized by using HF/6-311++G(d,p) method. Subsequently, the four electrostatic charges that were placed at C, N, O and amide-H atoms of NMA were fit from electrostatic potential derived from the cumulative atomic multipole moments (CAMM)[3]. In the variational fitting of the charges, the CAMM electrostatic potential was calculated in 30000 randomly selected points around an NMA molecule in a rectangular box of size 26 x 29 x 23 Bohrs with the origin at the center of geometry of NMA molecule (orientation of the structure is given below). The above fitting procedure resulted in the following charges (in a.u.): q_C=0.486770, q_O=−0.568832, q_N=−0.038462 and q_H=0.120524. The distribution of charges obtained in this way creates the dipole moment of 1.633 a.u. in magnitude which is reasonably close to the exact dipole moment magnitude of NMA molecule at HF/6-311++G(d,p) level of theory (1.623 a.u.; the basis set is Cartesian as implemented in the Gaussian 09 package[4]).

Structure of NMA optimized by using HF method and 6-311++G(d,p) Cartesian basis set (coordinates are given in Angstroms):

C	-0.643759	-1.848146	0.000000
N	-0.785028	-0.406404	0.000000
C	0.289224	0.411909	0.000000
C	0.000000	1.897303	0.000000
O	1.411421	-0.006666	0.000000
H	-0.105964	-2.184903	0.878095
H	0.463161	2.336182	-0.875792
H	-1.693273	-0.011260	0.000000
H	-1.058368	2.131423	0.000000
H	0.463161	2.336182	0.875792
H	-1.631713	-2.290957	0.000000
H	-0.105964	-2.184903	-0.878095

X. Structures of MeSCN-peptide dimers

Figure S10.



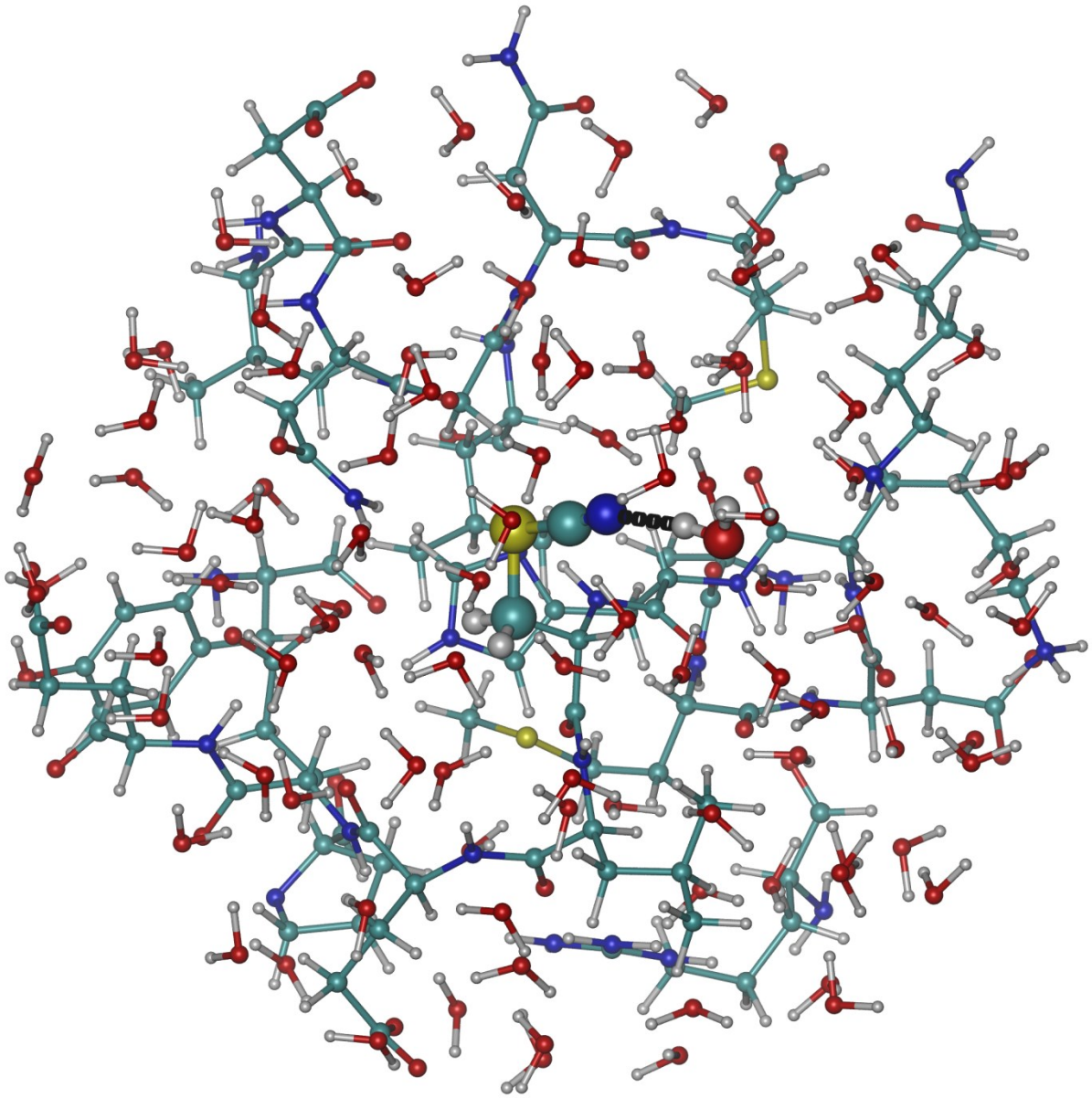
XI. Test of the fragmentation approach in the modeling of the protein environment by using small EFP fragments

Table S2. Test of the EFP fragmentation model used in the minimalistic model of the protein (Table 5). In this Table, the test protein environment mimicking molecules ["Ser1", "Ser2c", "Ser2n", "Tyr1", "Lys1"; see the structures above in Figure S5] are treated by SolEFP method. First, the interaction with an entire EFP parameter ("Full") set is considered. Then, the same molecules are approximated by small EFP models and Eq. (23) is used to evaluate SolEFP shifts ("Frag"). To test our solvatochromic theory we also show more accurate SolEDS calculations in this table and compare it with "Full QM" frequency shifts, those which were obtained from harmonic normal mode analysis. To compare SolEFP results with SolEDS at HF level, the SolEFP frequency shifts without dispersion ($\Delta\omega^{\text{SolEFP}}_{\text{no disp}}$) are presented here as well. Note that, while the SolEDS method takes into account both mechanical and electronic anharmonicities for all the interaction energy contributions, in the SolEFP method the electronic anharmonicity is included only for $\Delta\omega^{\text{Coul}}$ (see Sec. II. of this Document and also the discussion in Ref.[2]). All calculations were performed at HF/6-311++G** level of theory.

		"Ser1"	"Ser2c"	"Ser2n"	"Tyr1"	"Lys1"
SolEFP						
$\Delta\omega^{\text{Coul}}$	**Frag**	-8.3	-5.9	-7.7	-5.7	-39.9
	Full	-9.0	-7.2	-10.8	-6.7	-38.3
$\Delta\omega^{\text{Ex-Rep}}$	**Frag**	24.9	14.4	10.3	24.4	70.0
	Full	24.0	14.9	11.1	22.5	72.7
$\Delta\omega^{\text{Ind}}$	**Frag**	-6.1	-4.4	-4.1	-10.8	-46.9
	Full	-7.4	-8.0	-6.4	-14.6	-51.0
$\Delta\omega^{\text{Disp}}$	**Frag**	-6.9	-6.3	-6.2	-4.6	-13.9
	Full	-6.1	-6.4	-6.1	-5.0	-14.9
$\Delta\omega^{\text{SolEFP}}$	**Frag**	3.6	-2.2	-7.7	3.3	-30.7
	Full	1.5	-6.7	-12.2	-3.8	-31.5
$\Delta\omega^{\text{SolEFP}}_{\text{no disp}}$	**Frag**	**10.5**	**4.1**	**-1.5**	**7.9**	**-16.8**
	Full	**7.6**	**-0.3**	**-6.1**	**1.2**	**-16.6**
SolEDS						
$\Delta\omega^{(10)}_{\text{el}}$		-11.9	-10.1	-11.8	-9.7	-34.5
$\Delta\omega^{\text{HL}}_{\text{ex}}$		25.0	16.4	14.7	26.4	69.2
$\Delta\omega^{\text{HF}}_{\text{del}}$		-12.5	-8.5	-7.6	-10.1	-44.6
$\Delta\omega^{\text{HF}}$		0.6	-2.2	-4.7	6.6	-10.0
Full QM						
$\Delta\omega^{\text{HF}}$		-0.5	-3.1	-5.4	5.3	-10.7

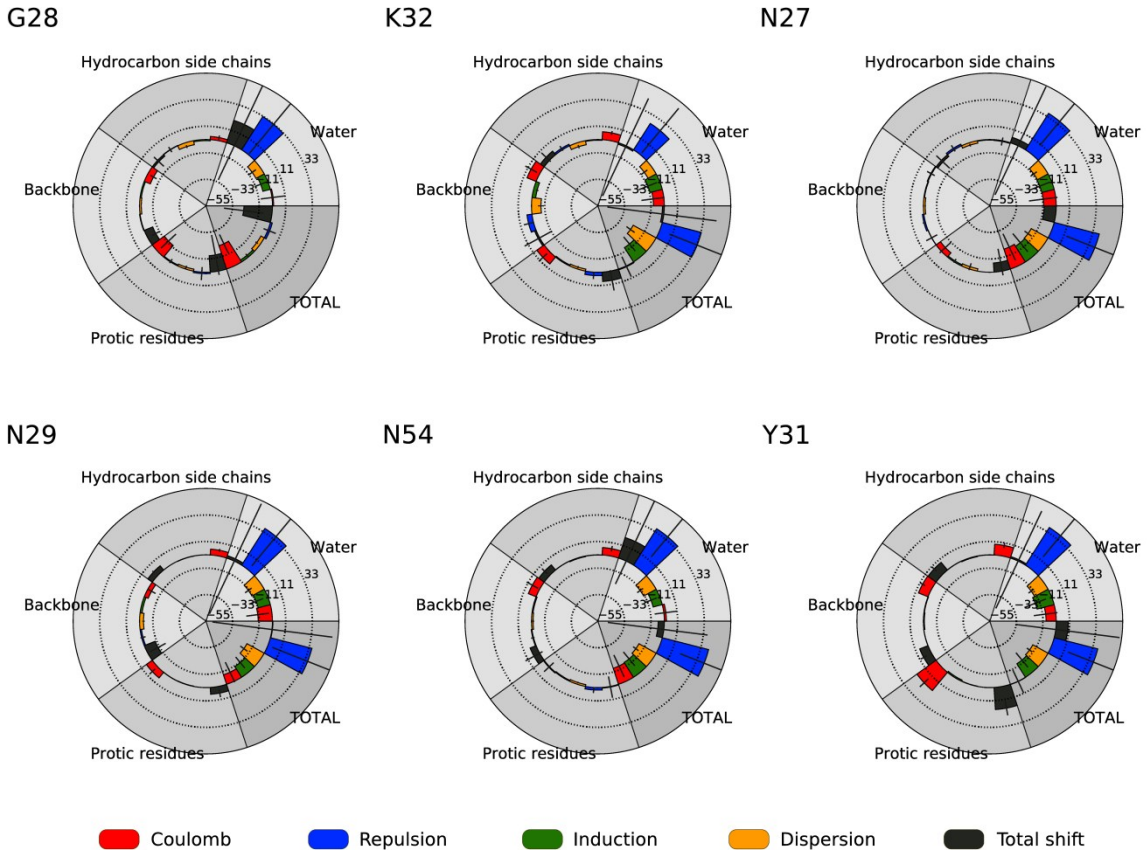
XII. A representative snapshot structure of RalGDS protein closest vicinity of the SCN probe

Figure S11.



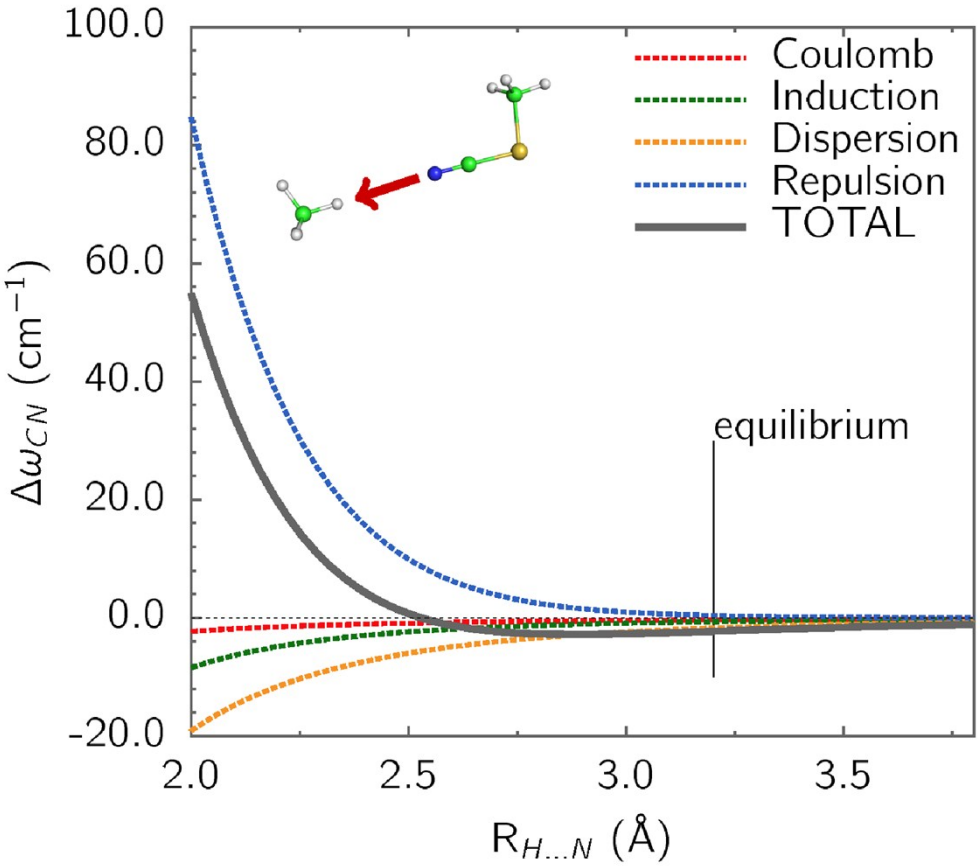
XIII. Short-range nitrile frequency shifts of free RalGDS with SCN label

Figure S12. The SolEFP short-range frequency shift components of CN stretch mode of SCN probe incorporated at six different sites of Ras-binding domain of RalGDS. 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}.



XIV. Distance-dependence of CN stretch mode frequency in MeSCN-CH_4 complex

Figure S13. The SolEFP CN stretch frequency shift components for MeSCN-CH_4 dimer with varying intermolecular distance are plotted with respect to the distance between H(CH_4) and N(MeSCN) atoms. The optimum H-bond distance is at 3.2021 Å (HF/6-311++G** level of theory). Note that the van der Waals interactions of the SCN probe at G28 position of RalGDS that is docked to Ras' protein (Figure 10 in the main text) are of similar nature as in this simple model example. In particular, in the highly crowded molecular environments where the steric clashes occur frequently, the blue-shift of the CN stretch mode can be substantial.



Supporting References

1. B. Błasiak and M. Cho, J. Chem. Phys. **143** (16), 164111 (2015).
2. B. Błasiak and M. Cho, J. Chem. Phys. **140** (16), 164107 (2014).
3. W. A. Sokalski and R. A. Poirier, Chem. Phys. Lett. **98** (1), 86-92 (1983).
4. 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. Foresman, J. V. Ortiz, J. Cioslowski and D. J. Fox, *Gaussian 09, Revision D.1*. (Gaussian, Inc, Wallingford CT, 2009).