Infrared Spectroscopy of Liquid Solutions as a Benchmarking Tool of Semi-Empirical QM Methods: The case of the GFN2-xTB

Laura X. Sepulveda-Montaño,[†] *Johan F. Galindo,*[*,‡] *and Daniel G. Kuroda*[*, †]

[†] Department of Chemistry, Louisiana State University, Baton Rouge, Louisiana 70803, United States.

[‡]Department of Chemistry, Universidad Nacional de Colombia sede Bogotá, 111321 Bogotá, Colombia.

[†,*]Email: dkuroda@lsu.edu

[‡,*]Email: jfgalindoc@unal.edu.co

ABSTRACT

The accurate description of large molecular systems has triggered the development of new computational methods. Due to the computational cost of modeling large systems, the methods usually require a trade-off between accuracy and speed. Therefore, benchmarking to test the accuracy and precision of the method is an important step in their development. The typical gold standard for evaluating these methods is isolated molecules because of the low computational cost. However, the advent of high-performance computing has made it possible to benchmark computational methods using observables from more complex systems, such as liquid solutions. To this end, infrared spectroscopy provides a suitable set of observables (i.e., vibrational transitions) from liquid systems. Here, IR spectroscopy observables are used to benchmark the predictions of the newly developed GFN2-xTB semi-empirical method. Three different IR probes (i.e., N-methylacetamide, benzonitrile, and semi-heavy water) in solution are selected for this purpose. The work presented here shows that GFN2-xTB predicts central frequencies with errors lower than 10% in all probes. In addition, the method captures detailed properties of the molecular environment, such as weak interactions. Finally, the GFN2-xTB correctly assesses the vibrational solvatochromism for N-methylacetamide and semi-heavy water, but does not have the accuracy needed to properly describe benzonitrile. Overall, the results indicate not only that GFN2-xTB can be used to predict the central frequencies and their dependence on the molecular environment with reasonable accuracy, but also that IR spectroscopy data of liquid solutions provide a suitable set of observables for the benchmarking of computational methods.

INTRODUCTION

Quantum chemical calculations have become one of the most important tools used in chemistry research. In the last decades, many fields such as materials science,[1] drug discovery,[2] organic chemistry,[3] and materials synthesis design,[4] have experienced an exponential growth in the use of quantum chemical tools to accurately predict molecular properties. While there has been a large progress in the development of supercomputers with significantly more computing power, the application of quantum chemical calculations is still restricted to small systems and light atoms.[5-6] This limitation is particularly evident in *ab initio* calculations, where in some cases the Schrodinger equation is solved for all the electrons of the quantum system. To overcome these size limitations, different approximations to the Hamiltonian of the Schrodinger equation have been implemented.[7-9] One particular approximation comes from the so-called semi-empirical methods, in which experimental values are used to construct the simplified and/or parametrized Hamiltonian. In many cases, these approximations are capable of producing accurate results.[10-12] From the computational perspective, semi-empirical methods usually provide substantial savings in computational time allowing one to compute properties of much larger systems. However, these methodologies typically suffer from low accuracy when applied to systems with atomic moieties not included in the construction of the approximate Hamiltonian.

In the last five years, a new semi-empirical tight binding quantum mechanical method, GFN2-xTB, has been developed to calculate geometries, frequencies, and non-covalent interactions of systems up to a few thousand atoms.[13] To achieve better accuracy, GFN2-xTB includes multipole electrostatics and density-dependent dispersion contributions through density fluctuations and atomic multipole moments.[13] While some studies have shown that GFN2-xTB is accurate for quantum chemical calculations, the method has not been extensively benchmarked.[13-15]

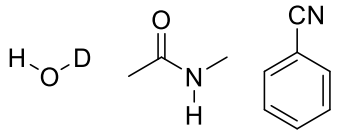
Benchmarking of chemical computational methods is typically performed against experimental values. To this end, reaction and formation enthalpies,[16] adsorption energies,[17] and electrochemical potentials are among the most widely used experimental observables.[14] In addition, some molecular properties derived from higher theoretical levels of *ab initio* calculations have been utilized to benchmark methods.[18-19] To benchmark a quantum chemical method, a molecular property should ideally provide a direct descriptor of the potential energy surface (PES), such as its minimum and curvature. However, these observables are usually not sufficient to evaluate systems with complex solute-solvent interactions, such as the hydrogen bonds or other short range interactions, because the potential energy surface of the system is strongly affected by them.[20] Therefore, the benchmarking of specific interactions, such as hydrogen bonds, requires experimental observables that are sensitive to them.

Infrared spectroscopy is a technique that measures observables directly related to the PES of the systems. The bands of the IR spectrum are associated with atomic structures and motions or equivalent, vibrational modes of the molecular system. Consequently, these IR modes are susceptible to changes in temperature and in the chemical nature of the solvent through variations in the inter- and intra-molecular potential of the molecule.[21-23] As a result, changes in the state of the system are observed via changes in the intensity, shape, and center frequency of the different IR absorption bands.[24] In particular, the variations associated with the solvation environment (i.e., vibrational solvatochromism) are a very important features of the IR band, because they are directly related to the inter-molecular potential of the molecular system.[25]

Vibrational modes and their associated frequencies can be calculated using *ab initio* methods, but an accurate prediction relies on a good description of the PES.[5,26] For systems in solution, the vibrational modes frequencies depend not only on the description of the position and curvature of

a minimum in the PES, but also on the correct quantification of the energetic contributions from the molecular environment. Thus, IR spectroscopy is able to provide molecular observables needed for benchmarking new chemical computational methods. In addition, linear IR spectroscopy is experimentally simple to perform and has a low cost when compared to other experimental techniques capable of describing the solvation effect of molecules in solution, such as NMR.[27-28]

The use of vibrational spectroscopy as a tool for benchmarking is not new, but most, if not all, of the studies have been performed using gas phase systems.[29] For example, studies have used peptide structures, [30] weakly bound complexes, [31] and ground state structures of peptides,[32] to name a few. While all the previous work is vital for benchmarking, the use of gas phase systems does not take into account for the variations of the PES caused by the solvent. The latter is an important parameter for modeling liquid phase systems. Here, the solvent contributions to the vibrational mode, or equivalent to the PES, are used as a benchmark for the GFN2-xTB methodology. For this purpose, three vibrational probes in different solvents were selected (Scheme 1, and Table 1 headers), namely water, N-methylacetamide, and benzonitrile. This IR probes along the selected solvents provide a large variety of solute-solvent and solvent-solvent interactions to represent the many different chemical environments needed to properly benchmark the GFN2-xTB using vibrational solvatochromism. In addition, the probe molecules have localized vibrational modes located in different parts of the IR spectrum, allowing us to span our investigation from ~1600 cm^{-1} to ~3500 cm^{-1}.



Scheme 1. Structure of vibrational probes. From left to right: water, N-methylacetamide and benzonitrile.

The vibrational probes selected in this work contain vibrational modes that can be used to benchmark the GFN2-xTB method. In the case of semi-heavy water, the molecule has the HOD bending and the OH stretch vibrational modes. Both the HOD bend and the OH stretch has been shown to be good reporters of the molecular environment around the water molecule.[33-39]. In particular, the OH stretch has been used to investigate the hydrogen bond network structure and dynamics in different systems.[40-43] In the case of the N-methylacetamide (NMA), the peptide bond (O=C-N-H) of the molecule is sensitive to the chemical environment,[44-46] which is readily seen through both its amide I and NH stretch vibrational modes.[47] Thus, information about the molecular environment has been derived using either the amide I mode,[48-50] or the NH stretch[51-52] of this group. Finally, benzonitrile (PhCN) contains the CN stretch mode, which is highly sensitive to its molecular environment[53] and the formation of hydrogen bonds.[54] However, the hydrogen bond interaction with the nitrile group is rather complex resulting in an intricate vibrational solvatochromism of the CN stretch,[53-56] which makes the CN stretch an ideal candidate for benchmarking quantum computational methods, such as GFN2-xTB. Overall, the investigated probes have vibrational modes that are good candidates to evaluate the suitability of the semi-empirical GFN2-xTB method for modeling the structure and interactions of these probes in different solvents using some of their vibrational mode frequencies.

METHODS

Experimental measurements and procedures can be found in the Supporting Information. The computational details are below.

Molecular Dynamics. The sampling for the frequency calculations was obtained by exploring the possible states of the molecular systems using molecular dynamics (MD) simulations. Solvent

boxes of 40 Å were built using PACKMOL.[57] Molecular dynamics simulations was performed using Amber 18 computational package.[58] The selected force fields were TIP3P [59] and OPC [60] for water, Amber force field for methanol (MeOH), and the generalized Amber force field (GAFF) [61] for the remaining solvents (Table 1). In addition, the force field for the probe molecules were ff14SB [62] for NMA and GAFF for PhCN and acetonitrile (ACN). Evaluation of the solvent boxes was conducted using radial distribution functions and densities calculations. MD of the molecular probes embedded in the solvent boxes were carried out. First an energy minimization at constant volume was performed, followed by a heating from 0 to 300.0 K for 20 ps, and finally a production run of 100 ns in the NPT ensemble was performed. A time step of 0.5 fs was used for the integration of the Newton's equations.

Table 1. Vibrational mode probes and solvents used as benchmarking sets for GFN2-xTB, where X indicates the studied vibrational mode/solvent combinations.

	Water		NMA		PhCN
Solvent	HOD bend	OD stretch	Amide I	NH stretch	CN stretch
Dimethyl sulfoxide (DMSO)	X	X	X	X	X
Acetonitrile (ACN)	X	X	X	X	X
Toluene (TOL)	N/A	N/A	X	X	X
Tetrahydrofuran (THF)*	X	X	X	X	X
Hexane (HEX)	N/A	N/A	N/A	N/A	X
Chloroform (CLF)	X	X	X	X	X
Water (WAT)	X	X	X	N/A	X
Methanol (MeOH)	N/A	N/A	X	N/A	X

N/A indicates vibrational modes that are not possible to study due to experimental limitations such as low solubility or overlapping modes with the solvent.

"*Note: The computation of THF-GBSA solvent was not possible due to geometry optimization issues that did not permit the completion of the energy minimization step.

Frequency calculations. Frequency calculations were performed on solvated probes (i.e., clusters) derived from snapshots of the MD simulation selected every 20 ps (5000 frames). These clusters contained, at least, the first solvation shell around the probe, as defined by the radial distribution functions (see the Supporting Information). In the frequency calculations, the energy minimizations were first performed at the *normal* level implemented in the GFN2-xTB package.[13,63] Two different sets of frequency computations were performed using explicit solvent and both explicit and implicit solvents. Clusters consisting only of explicit solvent molecules are denominated microsolvation, while clusters including implicit solvation models were named according to the implicit solvent model used. Specifically, the Analytical Linearized Poisson-Boltzmann (ALPB), and the Generalized Born with Surface Area (GBSA), [64] were used because they are both robust and efficient models to predict the behavior of molecules in a dielectric continuum or chemical environment.[64-66] A more detailed description of these models can be found in references [64-65,67-69]. Finally, the central frequency values were computed from the intensity weighted average [70] of the 5000 frequencies.

RESULTS AND DISCUSSION

Semi-heavy water probe. Computations of the central frequencies for the HOD bending using GFN2-xTB predict frequencies that are always smaller than the experimental values. The errors of the central frequencies are less than 6% for all solvents except for water (as solvent), where two different water models (OPC and TIP3P) produced errors of more than 8% (Figure 1 and Supporting Information). Furthermore, the HOD bending shows that microsolvation has the lowest average error compared to the other models containing implicit solvation (ALPB and GBSA), but the difference between the three models is not very large.

The suitability of the GFN2-xTB to describe more directional interactions of the probe in the different chemical environments is also evaluated from the frequency distributions. In most cases, the frequency distributions appear to be Gaussian profiles (see the Supporting Information). Moreover, the HOD bending in DMSO, chloroform, and THF shows distributions consistent with the expected chemical environment around the probe, which in the case of HDO/DMSO corresponds to a water molecule bridging two DMSO molecules with hydrogen bonds.[42,71] The histogram of frequencies for HDO in water also shows only one distribution, consistent with previous experimental work.[33] In contrast, the HOD bending in acetonitrile has two frequency distributions indicating that GFN2-xTB predicts two states of the water molecules solvated by acetonitrile.

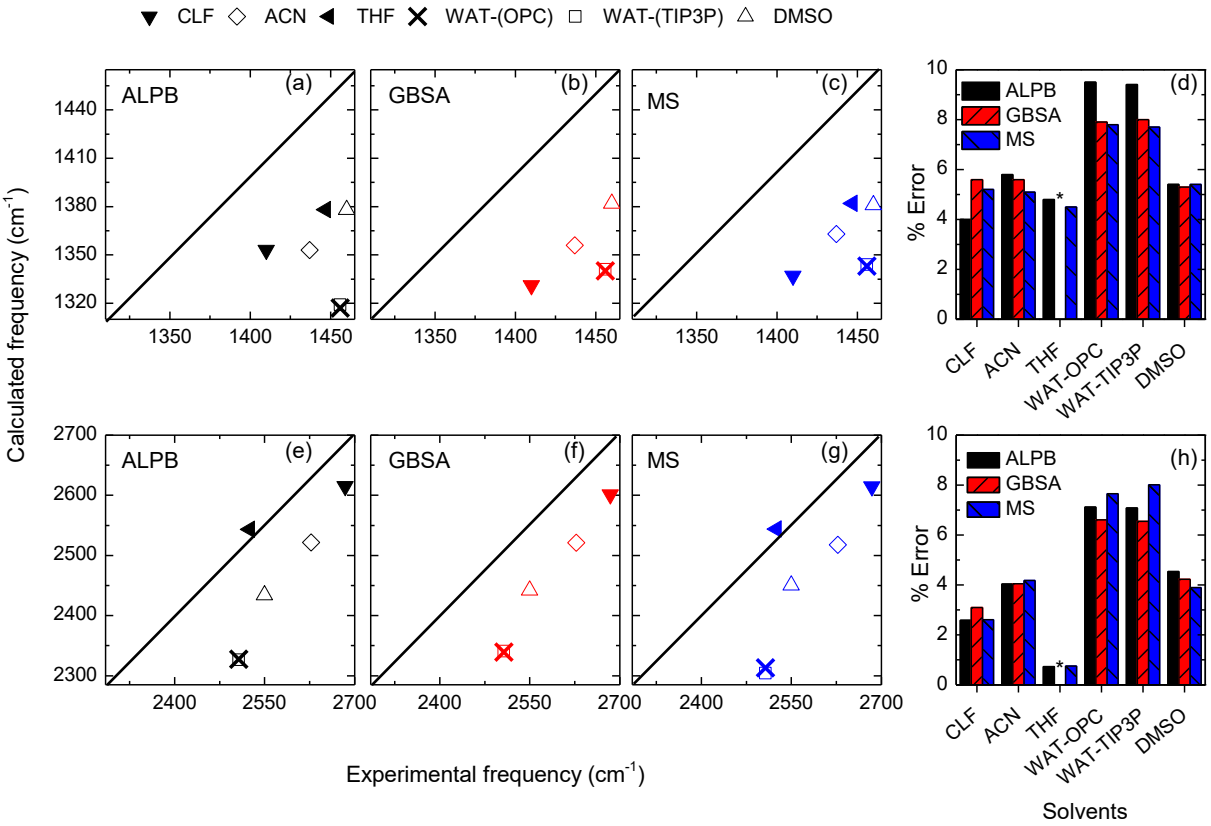


Figure 1. Solvatochromatic shift for the water vibrational modes predicted by GFN2-xTB using different solvation models. Panels (a), (b) and (c) show the results for HOD bending using ALPB, GBSA, and Microsolvation (MS), respectively, while panel (d) corresponds to %error with respect to the experimental values. Panels (e), (f) and (g) correspond to the solvatochromism of the OD

stretch with ALPB, GBSA, and Microsolvation (MS) models, respectively, while panel (h) corresponds to %error with respect to the experimental values. Star (*) means not available.

The validity of the GFN2-xTB for correctly predicting the water chemical environment is quantitatively evaluated from the vibrational solvatochromism of this molecule in the different solvents. Here, the solvatochromism of the HDO bend is determined by calculating the Pearson correlation coefficient between the experimental and calculated central frequencies in each solvent. Overall, the solvatochromism of the HOD bending mode (see the Supporting Information) appears to be reasonably well described by the GBSA (r=0.70) and microsolvation (r=0.61) solvation models, but not by ALPB (r=0.21). However, the lack of a good description of the HOD bending solvatochromism is likely due to the water solvent models since both present large deviations from their experimental values. Therefore, the HOD bending without water (TIP3P and OPC) as solvents leads to a better description of the solvatochromism of its bending mode for all models (GBSA r=0.99, Microsolvation r=0.96, and ALPB r=0.85; see Supporting Information). Note that the correlation coefficient might be overestimated in the case of the GBSA solvation model because the THF-GBSA system cannot be computed. These results indicate that the different water-solvent interactions are correctly captured by the method when the water molecule is in organic solvents, but not in water. Overall, the GFN2-xTB method correctly reproduces the solvatochromism of the HOD bending mode in different solvents, or equivalent, adequately describes the solvent contributions to the HDO potential energy surface.

A similar analysis is performed for the OD stretch of HDO. The results reveal that the average frequencies of the OD stretch in different solvents have errors in the range of 0.7 to 8.0% (Figure 1 and See the Supporting Information). Notably, HDO in THF has an error of less than 1% for both ALPB and microsolvation models. In contrast, the other systems containing organic solvents

produce less accurate, but still very good, results with errors between 2% and 4%. The best overall solvation model to describe the OD stretch is ALPB (error of 4.1%), followed by microsolvation (error of 4.3%), and GBSA (error of 4.9%). As in the case of the HOD bending, the errors of the models become significantly smaller when the systems containing water as solvent are excluded, further supporting to the idea that the classical models used to describe water (TIP3P and OPC) might not produce accurate solvation structures.

The distributions of the OD stretch frequencies for different solvents (see the Supporting Information) also confirm the validity of the method in correctly predicting the different chemical environments. For example, the results for water in DMSO show a single Gaussian distribution, in agreement with the water molecule forming predominantly one chemical environment, by forming two stable hydrogen bonds with DMSO.[42] The frequency distribution of HDO in ACN shows an two distributions, again indicating that GFN2-xTB predicts different chemical environments surrounding the HDO molecule in this solvent. This bimodal frequency distribution (see the Supporting Information) is reproduced using both implicit solvation models, but not through microsolvation. Similarly, the OD stretch of water in chloroform is expected to have a single frequency distribution because these environments lacks of directional interactions (i.e., hydrogen bonds). However, the expected single distribution is only observed for the ALPB and microsolvation models, but not for the GBSA case, where an asymmetric distribution is observed. An improper representation of the molecular environment is also derived from the OD stretch of HDO in THF when modeling with ALPB solvation and microsolvation models.

The solvatochromism of the OD stretch in the different chemical environments is correctly predicted by GFN2-xTB in all solvent systems and solvation models used, except for HDO in THF with either ALPB or microsolvation models (Figure 2). Specifically, the OD stretch using the

GBSA model has the best correlation coefficient (r=0.99), followed by ALPB (r=0.44), and microsolvation (r=0.42) (see the Supporting Information). It is important to note that these correlation values consider the water models (TIP3P and OPC). However, the solvatochromism of the OD stretch is better described when the predictions of the water solvent models are not taken into account (see the Supporting Information), as is also the case for the previously shown HOD bending.

Overall, the GFN2-xTB predicts central frequencies of OD stretch and HOD bending of water with small errors of ~4.5% and ~6.0% for the OD stretch and HOD bending, respectively. It is also observed that both modes have approximately the same error for the central frequencies, indicating that GFN2-xTB correctly captures the potential energy surface for the HDO molecule in different solvents. Moreover, the GFN2-xTB prediction of the solvatochromism for both modes in the different organic solvents shows the validity of the approach for describing different molecular environments or, equivalently, different molecular interactions. Finally, large errors are observed in the central frequency predictions for the two HDO vibrational modes studied in water. While it is possible that the GFN2-xTB method does not correctly model the water molecule, it is more likely that the force fields used to describe the water (and its complex hydrogen bond network) as a solvent (TIP3P and OPC) are responsible for the larger average errors in the central frequency predictions. This latter interpretation is supported by the systematic deviation seen in the central frequency predictions of the two investigated vibrational modes of HDO in water when compared to the organic solvents.

NMA probe. In the case of the amide I mode, the central frequencies obtained from the computations (Figure 2, and See the Supporting Information) show small deviations from experimental results with errors below 3%, but most them are less than 1%. It is also observed that

microsolvation overestimates the central frequency values for all solvents (Figure 2), while ALPB and GBSA predictions do not exhibit a clear trend in the errors. In general, the calculations that include both explicit and implicit solvation models outperform those that include only one term. The average error observed for the central amide I frequencies using the ALPB and GBSB solvation methods is less than 1.0%, while microsolvation produces a slightly larger average error of 1.4% (see the Supporting Information). Furthermore, computations using the ALPB solvation show the best performance for aprotic solvents (THF, toluene, ACN, and hexane), while those using GBSA solvation achieve the best results for protic solvents (D_2O and MeOH). However, the frequency computations containing the microsolvation model predict the central frequencies of NMA in chloroform with the lowest error.

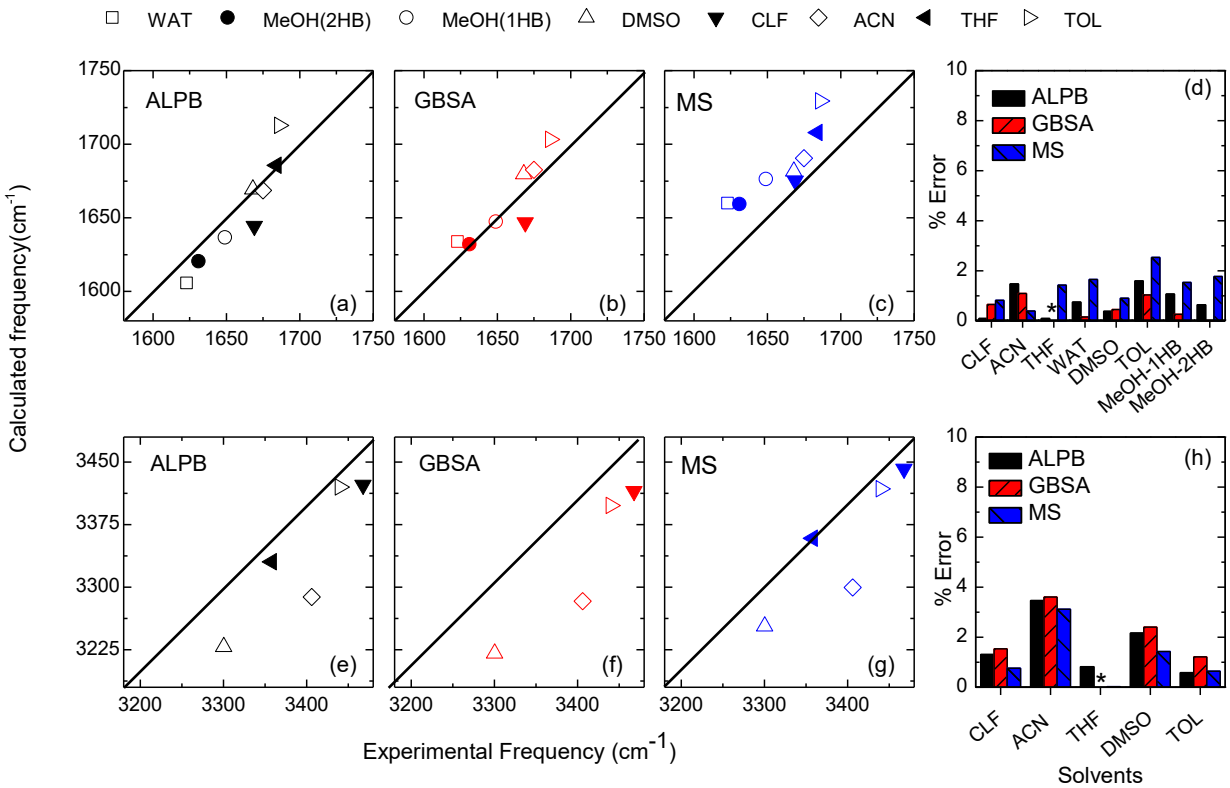


Figure 2. Solvatochromatic shift for the NMA vibrational modes predicted by GFN2-xTB using different solvation models. Panels (a), (b) and (c) show the results for amide I mode using ALPB, GBSA, and Microsolvation (MS), respectively, while panel (d) corresponds to %error with respect to the experimental values. Panels (e), (f) and (g) correspond to the solvatochromism of the NH

stretch with ALPB, GBSA, and Microsolvation (MS) models, respectively, while panel (h) corresponds to %error with respect to the experimental values. Star (*) means not available.

The solute-solvent interactions derived from the frequency distributions (see the Supporting Information) show that NMA in MeOH presents two separable Gaussian distributions, which have been experimentally observed and attributed to the NMA molecule interacting with MeOH molecules forming one or two hydrogen bonds,[72] with a chemical exchange at timescales long enough to reflect two distributions in the FTIR spectra,[73] This been the only case where the hydrogen bond dynamics is long enough to allow this type of separation and characterization on the frequency distributions. Nonetheless, a similar behavior is also observed for the distributions of amide in chloroform in accordance with previously reported hydrogen bond interactions.[74] These two cases contrast with the NMA in water where only one distribution is observed theoretically and experimentally.[23,75] In addition, NMA in all the other solvents presents single frequency distributions consistent with the lack of directional interactions (i.e., hydrogen bond) in the chemical environment surrounding the amide.

The predicted central frequencies as a function of the solvent, vibrational solvatochromism, show a linear trend with the experimental values. Specifically, the GBSA and ALPB solvation models have linear correlation coefficients of 0.97 and 0.98, respectively, while microsolvation has a lower correlation coefficient of 0.70 (see the Supporting Information). Overall, the small error as well as the strong correlation coefficient seen for all solvation models reflects the good quality of central frequency predictions for the amide in the different solvents when using the GFN2-xTB method.

In the case of the NH stretch of NMA, the predicted central frequencies exhibit lower values than those measured experimentally, regardless of the solvent model. The deviations of the

computed frequencies from the experimental values are less than 3%, except for the amide in ACN (Figure 2 and the Supporting Information). Furthermore, in contrast to the amide I mode of NMA, the smallest errors for the central frequencies are observed for the microsolvation, which is closely followed by the other two solvation models (see the Supporting Information). Since the central frequency values for amide I and NH stretch present similar deviations from the experimental frequencies, it is possible to deduce that the frequency computations using GFN2-xTB correctly predict not only the amide interactions with the different solvents, but also the molecular structure of the probe in each solvent.

The NH group is susceptible to the formation of hydrogen bonds. It is therefore expected that the number of central frequency distributions of the NH stretch will depend on the chemical environment and its ability to form hydrogen bonds. For example, solvents that cannot form hydrogen bonds are expected to have a single frequency distribution (see the Supporting Information). In contrast to the expectations, the frequency calculation of NMA in toluene using any of the solvation models, as well as NMA in chloroform using the GBSA solvation model show two distinct distributions. The presence of two distributions shows that GFN2-xTB predicts chemical environments that are unlikely to be present in the solvent. The GFN2-xTB method correctly predicts the central frequency values, as seen in the calculated solvatochromism for the NH stretch (Figure 2). In particular, the GBSA model makes the best predictions for NH stretch solvent dependence (r=0.97), followed by microsolvation (r=0.94) and ALPB (r=0.88) as shown in (see the Supporting Information).

Benzonitrile probe. The nitrile stretch of this molecule shows that the GFN2-xTB method overestimates its central frequencies irrespective of solvent and solvation model (Figure 3 insets). The large deviation is readily observed from the errors, which are above 5.0% in all cases (Figure

3). Specifically, the ALPB solvation model gives slightly better predictions for the central frequencies of the benzonitrile nitrile stretch in water, hexane, THF and methanol (see the Supporting Information), while the GBSA solvation model makes better predictions for toluene and chloroform as solvents (see the Supporting Information). Additionally, the probe embedded in ACN and DMSO yields similar errors in both ALPB and GBSA models, but significantly larger errors in the microsolvation prediction. Thus, it is evident that the inclusion of an implicit solvation model results in better predictions of the central frequencies for the nitrile stretch.

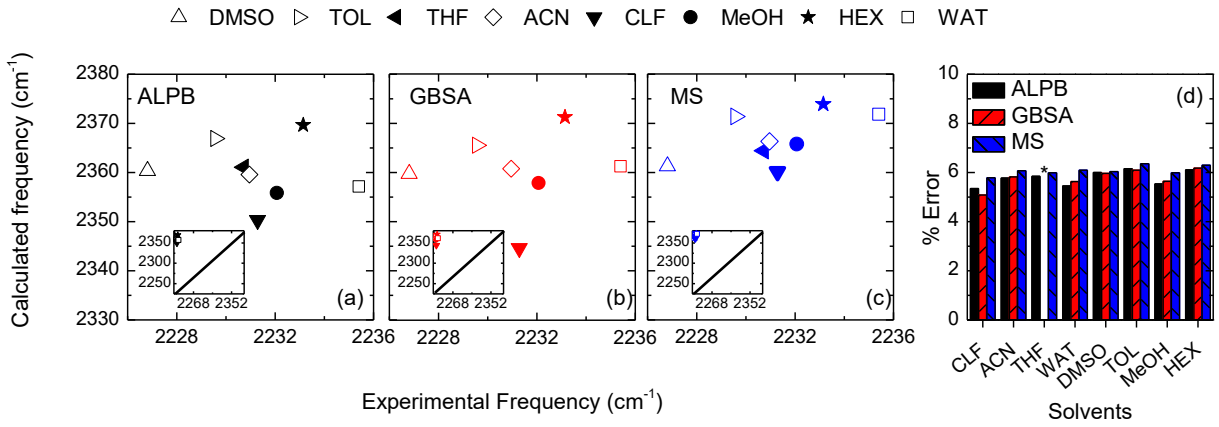


Figure 3. Solvatochromatic shift for the benzonitirle vibrational mode predicted by GFN2-xTB using different solvation models. Panels (a), (b) and (c) show the results for nitrile stretch using ALPB, GBSA, and Microsolvation (MS), respectively, while panel (d) corresponds to %error with respect to the experimental values. Insets show the deviation in an expanded region of the IR. Star (*) means not available.

The frequency distributions of the benzonitrile nitrile stretch were also analyzed (see the Supporting Information) to determine the appropriateness of the semiempirical methodology for capturing the molecular level interactions of the molecule in the different solvents. A single Gaussian frequency distribution is observed for DMSO, toluene, THF, ACN, and hexane consistent with the expected solute-solvent interactions of the probe. In contrast, benzonitrile in water shows two distributions for all the solvation models, but the presence of such states has not been observed experimentally.[54,76] Moreover, the computations predict only a single distribution

for the nitrile stretch in MeOH irrespective of the solvation model (see the Supporting Information), while experiments show two frequency distributions.[54,77-78] In addition, the probe in chloroform was expected to present a single distribution, but only the microsolvation model asserted this observation.

The vibrational solvatochromism of the benzonitrile CN stretch shows very poor predictions, as shown by the negative or very low correlations for the ALPB (r=-0.12) and GBSA (r=0.11) solvation models and slightly better modeling, but still fairly poor, for microsolvation (r=0.58) (see the Supporting Information). In general, the solvatochromism reveals an inadequate representation of the interaction potential between the solvent and the nitrile moiety of the probe.

The nitrile group of benzonitrile has a complex electronic structure, due to the electronic conjugation with the aromatic ring, which could play a role in defining the nitrile stretch frequency.[79-80] To discard the possible effects of the ring, the influence of the solvents on the nitrile group is also evaluated by using acetonitrile since this molecule only contains the nitrile group. The ACN molecule in three different solvents (water, toluene, and chloroform) shows similar levels of error (>5.0%) and overestimation of the central frequency (see the Supporting Information) as benzonitrile. In addition, the vibrational solvatochromism of the ACN nitrile stretch is also poorly captured by all solvation models (r=0.36 for ALPB, and r=0.48 for both GBSA and microsolvation, See the Supporting Information). Therefore, it is concluded that the large deviations for the nitrile stretch frequencies observed when using the GFN2-xTB method are likely a consequence of the low accuracy of the semiempirical method for this particular system. However, the above results are not surprising considering that the solvatochromism of the nitrile stretch expands only in 10 cm^{-1} region for both benzonitrile and acetonitrile,[55] indicating that very accurate calculations are required to correctly predict the solvent effect on this mode.

Cumulative frequency analysis for all probed modes. The accuracy of the GFN2-xTB method in predicting the vibrational modes of all the different molecules are also evaluated. It should be noted that all frequency calculations, when combined, appear in a region of the IR spectrum that spans ~2500cm^{-1}. Overall, the calculated central frequencies have values with small deviations from the experimental results for all solvation models (Figure 4). The average error of the evaluated central frequencies is 3.8% for both the ALPB and microsolvation models, and 3.9% for the GBSA model. Considering that the normal mode calculation requires a good equilibrium geometry of the investigated molecules,[81] it is also reasonable to state that the molecular structures determined by the GFN2-xTB method are correct for the three molecular systems studied here. The good quality of the general predictions is evident from the derived central frequencies, which show deviations from experimental results lower than 6% in all the investigated systems and vibrational modes (see the Supporting Information). Moreover, the solvatochromism of all the investigated modes has correlation coefficients greater than 0.99 irrespective of the solvation model. This last result highlights the general adequacy of the GFN2-xTB method for predicting frequency shifts of vibrational modes, even when considering the nitrile stretch and the HOD bend, which have observable deviations in the predicted central frequencies (Figure 4).

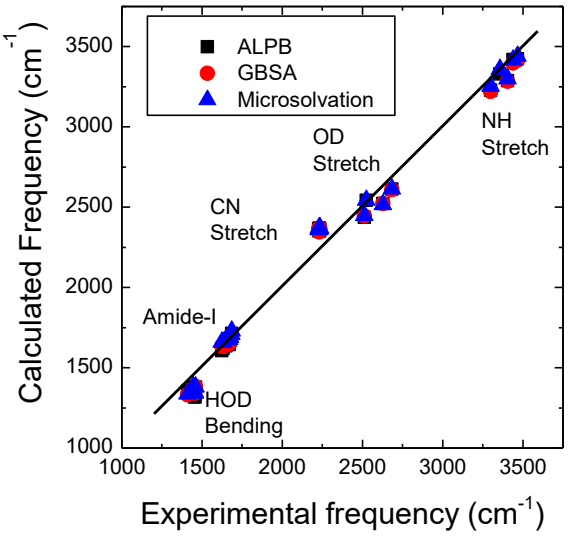


Figure 4. Vibrational frequencies of the studied vibrational modes for the different probes in various solvents. The data points represent frequency average calculated using GFN2-xTB and the ALPB (black squares), GBSA (red circles) and microsolvation (blue triangles) models against the experimental values.

CONCLUSIONS

The GFN2-xTB semi-empirical calculations provide a suitable method for computing vibrational frequencies of large systems with errors less than 10%. In most cases, the GFN2-xTB approach also appears to correctly reproduce the vibrational solvatochromism of molecules, such as water and N-methylacetamide. However, it does not have sufficient accuracy to predict the frequency shift of molecules containing nitrile groups (e.g., benzonitrile and acetonitrile) in different solvents. This is probably due to the small contribution of the bath interaction potential to the PES of the probe. In addition, the GFN2-xTB method also appears to capture detailed molecular interactions, such as hydrogen bonds, observed experimentally through the frequency distributions even for solvents that form weak hydrogen bonds, such as chloroform. Overall, GFN2-xTB frequency calculations produce accurate results that are consistent with the nature of the interactions present in the different solvents. In addition, this work demonstrates that IR spectroscopy of liquid solutions provides suitable experimental metrics for the benchmarking of chemical computational methods. In particular, the strength of the IR metrics relies on the central frequency of the vibrational mode, which is strongly dependent on the molecular structure of the probe, as well as its interaction with the solvent (i.e., intermolecular potentials). Thus, the vibrational mode serves as a metric for evaluating the predictive power of quantum chemical calculations as well as solvation models and force fields used to model the molecules under investigation.

ASSOCIATED CONTENT

The Supporting Information is available free of charge at:

Experimental specifications for FTIR measurements. Radial distributions from the molecular dynamics calculations. Calculated central frequencies and frequencies distributions of all systems mentioned in the text. Correlation coefficients indicating the solvatochromism of the systems. Results for acetonitrile molecular probe.

Corresponding Authors

[*, †]Daniel G. Kuroda. Department of Chemistry, Louisiana State University, Baton Rouge, Louisiana 70803, United States; Phone (+1) 225-578-1780; Email: dkuroda@lsu.edu; orcid.org/0000-0002-4752-7024

[*, ‡]Johan F. Galindo. Department of Chemistry, Universidad Nacional de Colombia sede Bogotá, 111321 Bogotá, Colombia.; Phone: (+57) 3165000; Email: jfgalindoc@unal.edu.co; orcid.org/0000-0001-9968-362X

Authors

†Laura X Sepulveda-Montaño; orcid.org/0000-0001-9538-4032

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENT

The authors would like to acknowledge financial support from the National Science Foundation (CHE-175135). This work has been partially supported by DIEB-UNAL (QUIPU: 201010027423). The authors would also would like to acknowledge the High Performance Computing Center at Louisiana State University and the Louisiana Optical Network Initiative

(LONI) and Center of Excellence in Scientific Computing at Universidad Nacional de Colombia for computer time.

REFERENCES

1. Chalamet, L.; Rodney, D.; Shibuta, Y., Coarse-grained molecular dynamic model for metallic materials. *Comput. Mater. Sci.* **2023,** *228*, 112306.
2. Zheng, M.; Liu, X.; Xu, Y.; Li, H.; Luo, C.; Jiang, H., Computational methods for drug design and discovery: Focus on china. *Trends Pharmacol. Sci.* **2013,** *34* (10), 549-559.
3. Castellá-Ventura, M.; Kassab, E., Comparative semiempirical and ab initio study of the harmonic vibrational frequencies of aniline—i. The ground state. *Spectrochim. Acta A Mol. Biomol. Spectrosc.* **1994,** *50* (1), 69-86.
4. Chianella, I.; Lotierzo, M.; Piletsky, S. A.; Tothill, I. E.; Chen, B.; Karim, K.; Turner, A. P., Rational design of a polymer specific for microcystin-lr using a computational approach. *Anal. Chem.* **2002,** *74* (6), 1288-1293.
5. Levine, I. N.; Busch, D. H.; Shull, H., *Quantum chemistry*. 7 ed.; Pearson Prentice Hall Upper Saddle River, NJ: 2013; Vol. 6, p 720.
6. Ishikawa, K. L.; Sato, T., A review on ab initio approaches for multielectron dynamics. *IEEE J. Sel. Top. Quantum Electron.* **2015,** *21* (5), 1-16.
7. Møller, C.; Plesset, M. S., Note on an approximation treatment for many-electron systems. *Phys. Rev.* **1934,** *46* (7), 618.
8. Coester, F., Bound states of a many-particle system. *Nucl. Phys. A* **1958,** *7*, 421-424.
9. Kohn, W.; Sham, L. J., Self-consistent equations including exchange and correlation effects. *Phys. Rev.* **1965,** *140* (4A), A1133.
10. Stewart, J. J., Optimization of parameters for semiempirical methods v: Modification of nddo approximations and application to 70 elements. *J. Mol. Model.* **2007,** *13*, 1173-1213.
11. Gaus, M.; Cui, Q.; Elstner, M., Dftb3: Extension of the self-consistent-charge density-functional tight-binding method (scc-dftb). *J. Chem. Theory Comput.* **2011,** *7* (4), 931-948.
12. Dewar, M. J.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J., Development and use of quantum mechanical molecular models. 76. Am1: A new general purpose quantum mechanical molecular model. *J. Am. Chem. Soc.* **1985,** *107* (13), 3902-3909.
13. Bannwarth, C.; Ehlert, S.; Grimme, S., Gfn2-xtb—an accurate and broadly parametrized self-consistent tight-binding quantum chemical method with multipole electrostatics and density-dependent dispersion contributions. *J. Chem. Theory Comput.* **2019,** *15* (3), 1652-1671.
14. Neugebauer, H.; Bohle, F.; Bursch, M.; Hansen, A.; Grimme, S., Benchmark study of electrochemical redox potentials calculated with semiempirical and dft methods. *J. Phys. Chem. A* **2020,** *124* (35), 7166-7176.
15. Otlyotov, A. A.; Minenkov, Y., Conformational energies of microsolvated na+ clusters with protic and aprotic solvents from gfn n-xtb methods. *J. Comput. Chem.* **2022,** *43* (27), 1856-1863.
16. Somers, K. P.; Simmie, J. M., Benchmarking compound methods (cbs-qb3, cbs-apno, g3, g4, w1bd) against the active thermochemical tables: Formation enthalpies of radicals. *J. Phys. Chem. A* **2015,** *119* (33), 8922-8933.

17. Campbell, C. T., Energies of adsorbed catalytic intermediates on transition metal surfaces: Calorimetric measurements and benchmarks for theory. *Acc. Chem. Res.* **2019,** *52* (4), 984-993.
18. Ahmadi, A.; Beheshtian, J.; Kamfiroozi, M., Benchmarking of oniom method for the study of nh3 dissociation at open ends of bnnts. *J. Mol. Model.* **2012,** *18* (5), 1729-1734.
19. Ziogos, O. G.; Kubas, A.; Futera, Z.; Xie, W.; Elstner, M.; Blumberger, J., Hab79: A new molecular dataset for benchmarking dft and dftb electronic couplings against high-level ab initio calculations. *J. Chem. Phys.* **2021,** *155* (23), 234115.
20. Hostaš, J.; Řezáč, J.; Hobza, P., On the performance of the semiempirical quantum mechanical pm6 and pm7 methods for noncovalent interactions. *Chem. Phys. Lett.* **2013,** *568*, 161-166.
21. Hamm, P.; Zanni, M., *Concepts and methods of 2d infrared spectroscopy*. Cambridge University Press: 2011; p 285.
22. Nicodemus, R. A.; Corcelli, S.; Skinner, J.; Tokmakoff, A., Collective hydrogen bond reorganization in water studied with temperature-dependent ultrafast infrared spectroscopy. *J. Phys. Chem. B* **2011,** *115* (18), 5604-5616.
23. DeCamp, M.; DeFlores, L.; McCracken, J.; Tokmakoff, A.; Kwac, K.; Cho, M., Amide i vibrational dynamics of n-methylacetamide in polar solvents: The role of electrostatic interactions. *J. Phys. Chem. B* **2005,** *109* (21), 11016-11026.
24. Bagchi, S.; Fried, S. D.; Boxer, S. G., A solvatochromic model calibrates nitriles' vibrational frequencies to electrostatic fields. *J. Am. Chem. Soc.* **2012,** *134* (25), 10373-10376.
25. Whittall, I. R.; McDonagh, A. M.; HAMPHREY, M.; Samoc, M., Organometallic complexes in nonlinear optics i: Second-order nonlinearities. *Chem. Inform.* **1998,** *29* (30), no-no.
26. Cho, M.; Fleming, G. R.; Saito, S.; Ohmine, I.; Stratt, R. M., Instantaneous normal mode analysis of liquid water. *J. Chem. Phys.* **1994,** *100* (9), 6672-6683.
27. Charisiadis, P.; Kontogianni, V. G.; Tsiafoulis, C. G.; Tzakos, A. G.; Siskos, M.; Gerothanassis, I. P., 1h-nmr as a structural and analytical tool of intra-and intermolecular hydrogen bonds of phenol-containing natural products and model compounds. *Molecules* **2014,** *19* (9), 13643-13682.
28. Kulinich, A. V.; Ishchenko, A. A.; Groth, U. M., Electronic structure and solvatochromism of merocyanines: Nmr spectroscopic point of view. *Spectrochim. Acta A Mol. Biomol. Spectrosc.* **2007,** *68* (1), 6-14.
29. Morton, T. H., Gas phase infrared spectra of small polyatomic ions as a benchmark for theory. *Inorg. Chim. Acta* **2011,** *369* (1), 140-145.
30. Nagornova, N. S.; Rizzo, T. R.; Boyarkin, O. V., Highly resolved spectra of gas-phase gramicidin s: A benchmark for peptide structure calculations. *J. Am. Chem. Soc.* **2010,** *132* (12), 4040-4041.
31. Dryza, V.; Poad, B. L.; Bieske, E. J., Spectroscopic study of the benchmark mn+─ h2 complex. *J. Phys. Chem. A* **2009,** *113* (21), 6044-6048.
32. Doemer, M.; Guglielmi, M.; Athri, P.; Nagornova, N. S.; Rizzo, T. R.; Boyarkin, O. V.; Tavernelli, I.; Rothlisberger, U., Assessing the performance of computational methods for the prediction of the ground state structure of a cyclic decapeptide. *Int. J. Quantum Chem* **2013,** *113* (6), 808-814.
33. Chuntonov, L.; Kumar, R.; Kuroda, D. G., Non-linear infrared spectroscopy of the water bending mode: Direct experimental evidence of hydration shell reorganization? *Phys. Chem. Chem. Phys.* **2014,** *16* (26), 13172-13181.

34. Seki, T.; Chiang, K.-Y.; Yu, C.-C.; Yu, X.; Okuno, M.; Hunger, J.; Nagata, Y.; Bonn, M., The bending mode of water: A powerful probe for hydrogen bond structure of aqueous systems. *J. Phys. Chem. Lett.* **2020,** *11* (19), 8459-8469.
35. Seki, T.; Sun, S.; Zhong, K.; Yu, C.-C.; Machel, K.; Dreier, L. B.; Backus, E. H.; Bonn, M.; Nagata, Y., Unveiling heterogeneity of interfacial water through the water bending mode. *J. Phys. Chem. Lett.* **2019,** *10* (21), 6936-6941.
36. Lawrence, C.; Skinner, J., Vibrational spectroscopy of hod in liquid d 2 o. Iii. Spectral diffusion, and hydrogen-bonding and rotational dynamics. *J. Chem. Phys.* **2003,** *118* (1), 264-272.
37. Schäfer, T.; Lindner, J.; Vöhringer, P.; Schwarzer, D., Od stretch vibrational relaxation of hod in liquid to supercritical h 2 o. *J. Chem. Phys.* **2009,** *130* (22), 224502.
38. Rey, R.; Møller, K. B.; Hynes, J. T., Hydrogen bond dynamics in water and ultrafast infrared spectroscopy. *J. Phys. Chem. A* **2002,** *106* (50), 11993-11996.
39. Loparo, J. J.; Roberts, S. T.; Tokmakoff, A., Multidimensional infrared spectroscopy of water. I. Vibrational dynamics in two-dimensional ir line shapes. *J. Chem. Phys.* **2006,** *125* (19), 194521.
40. Zheng, J.; Kwak, K.; Fayer, M., Ultrafast 2d ir vibrational echo spectroscopy. *Acc. Chem. Res.* **2007,** *40* (1), 75-83.
41. Fayer, M. D.; Moilanen, D. E.; Wong, D.; Rosenfeld, D. E.; Fenn, E. E.; Park, S., Water dynamics in salt solutions studied with ultrafast two-dimensional infrared (2d ir) vibrational echo spectroscopy. *Acc. Chem. Res.* **2009,** *42* (9), 1210-1219.
42. Wong, D. B.; Sokolowsky, K. P.; El-Barghouthi, M. I.; Fenn, E. E.; Giammanco, C. H.; Sturlaugson, A. L.; Fayer, M. D., Water dynamics in water/dmso binary mixtures. *J. Phys. Chem. B* **2012,** *116* (18), 5479-5490.
43. Mahanta, D. D.; Islam, S. I.; Choudhury, S.; Das, D. K.; Mitra, R. K.; Barman, A., Contrasting hydration dynamics in dme and dmso aqueous solutions: A combined optical pump-probe and ghz-thz dielectric relaxation investigation. *J. Mol. Liq.* **2019,** *290*, 111194.
44. Krummel, A. T.; Zanni, M. T., Interpreting DNA vibrational circular dichroism spectra using a coupling model from two-dimensional infrared spectroscopy. *J. Phys. Chem. B* **2006,** *110* (48), 24720-24727.
45. Mukherjee, P.; Kass, I.; Arkin, I. T.; Zanni, M. T., Structural disorder of the cd3ζ transmembrane domain studied with 2d ir spectroscopy and molecular dynamics simulations. *J. Phys. Chem. B* **2006,** *110* (48), 24740-24749.
46. Fang, C.; Senes, A.; Cristian, L.; DeGrado, W. F.; Hochstrasser, R. M., Amide vibrations are delocalized across the hydrophobic interface of a transmembrane helix dimer. *Proc. Natl. Acad. Sci. U.S.A.* **2006,** *103* (45), 16740-16745.
47. la Cour Jansen, T.; Knoester, J., A transferable electrostatic map for solvation effects on amide i vibrations and its application to linear and two-dimensional spectroscopy. *J. Chem. Phys.* **2006,** *124* (4), 044502.
48. Feng, C.-J.; Tokmakoff, A., The dynamics of peptide-water interactions in dialanine: An ultrafast amide i 2d ir and computational spectroscopy study. *J. Chem. Phys.* **2017,** *147* (8), 085101.
49. Ganim, Z.; Chung, H. S.; Smith, A. W.; DeFlores, L. P.; Jones, K. C.; Tokmakoff, A., Amide i two-dimensional infrared spectroscopy of proteins. *Acc. Chem. Res.* **2008,** *41* (3), 432-441.

50. Ye, S.; Li, H.; Yang, W.; Luo, Y., Accurate determination of interfacial protein secondary structure by combining interfacial-sensitive amide i and amide iii spectral signals. *J. Am. Chem. Soc.* **2014,** *136* (4), 1206-1209.
51. Kuznetsova, L.; Furer, V.; Maklakov, L., Infrared intensities of n-methylacetamide associates. *J. Mol. Struct.* **1996,** *380* (1-2), 23-29.
52. Köddermann, T.; Ludwig, R., N-methylacetamide/water clusters in a hydrophobic solvent. *Phys. Chem. Chem. Phys.* **2004,** *6* (8), 1867-1873.
53. Chung, J. K.; Thielges, M. C.; Fayer, M. D., Dynamics of the folded and unfolded villin headpiece (hp35) measured with ultrafast 2d ir vibrational echo spectroscopy. *Proc. Natl. Acad. Sci. U.S.A.* **2011,** *108* (9), 3578-3583.
54. Ghosh, A.; Remorino, A.; Tucker, M. J.; Hochstrasser, R. M., 2d ir photon echo spectroscopy reveals hydrogen bond dynamics of aromatic nitriles. *Chem. Phys. Lett.* **2009,** *469* (4-6), 325-330.
55. Nyquist, R., Solvent-induced nitrile frequency shifts: Acetonitrile and benzonitrile. *Appl. Spectrosc.* **1990,** *44* (8), 1405-1407.
56. Levinson, N. M.; Fried, S. D.; Boxer, S. G., Solvent-induced infrared frequency shifts in aromatic nitriles are quantitatively described by the vibrational stark effect. *J. Phys. Chem. B* **2012,** *116* (35), 10470-10476.
57. Martínez, L.; Andrade, R.; Birgin, E. G.; Martínez, J. M., Packmol: A package for building initial configurations for molecular dynamics simulations. *J. Comput. Chem.* **2009,** *30* (13), 2157-2164.
58. Case, D. A.; Darden, T.; Cheatham, T.; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Crowley, M.; Walker, R. C.; Zhang, W. *Amber 10*; University of California: 2008.
59. Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L., Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983,** *79* (2), 926-935.
60. Izadi, S.; Anandakrishnan, R.; Onufriev, A. V., Building water models: A different approach. *J. Phys. Chem. Lett.* **2014,** *5* (21), 3863-3871.
61. Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A., Development and testing of a general amber force field. *J. Comput. Chem.* **2004,** *25* (9), 1157-1174.
62. Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C., Ff14sb: Improving the accuracy of protein side chain and backbone parameters from ff99sb. *J. Chem. Theory Comput.* **2015,** *11* (8), 3696-3713.
63. Bannwarth, C.; Caldeweyher, E.; Ehlert, S.; Hansen, A.; Pracht, P.; Seibert, J.; Spicher, S.; Grimme, S., Extended tight-binding quantum chemistry methods. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2021,** *11* (2), e1493.
64. Ehlert, S.; Stahn, M.; Spicher, S.; Grimme, S., Robust and efficient implicit solvation model for fast semiempirical methods. *J. Chem. Theory Comput.* **2021,** *17* (7), 4250-4261.
65. Kleinjung, J.; Fraternali, F., Design and application of implicit solvent models in biomolecular simulations. *Curr. Opin. Struct. Biol.* **2014,** *25*, 126-134.
66. Gorges, J.; Grimme, S.; Hansen, A.; Pracht, P., Towards understanding solvation effects on the conformational entropy of non-rigid molecules. *Phys. Chem. Chem. Phys.* **2022,** *24* (20), 12249-12259.
67. Lange, A. W.; Herbert, J. M., Improving generalized born models by exploiting connections to polarizable continuum models. I. An improved effective coulomb operator. *J. Chem. Theory Comput.* **2012,** *8* (6), 1999-2011.

68. Sigalov, G.; Fenley, A.; Onufriev, A., Analytical electrostatics for biomolecules: Beyond the generalized born approximation. *J. Chem. Phys.* **2006,** *124* (12), 094511.
69. Onufriev, A.; Bashford, D.; Case, D. A., Exploring protein native states and large-scale conformational changes with a modified generalized born model. *Proteins: Struct. Funct.* **2004,** *55* (2), 383-394.
70. Navidi, W. C., *Statistics for engineers and scientists*. 3rd ed.; McGraw-Hill New York: Inc., 1221 Avenue of the Americas,
New York, NY 10020., 2006; Vol. 2, p 933.
71. Borin, I. A.; Skaf, M. S., Molecular association between water and dimethyl sulfoxide in solution: A molecular dynamics simulation study. *J. Chem. Phys.* **1999,** *110* (13), 6412-6420.
72. Kwac, K.; Lee, H.; Cho, M., Non-gaussian statistics of amide i mode frequency fluctuation of n-methylacetamide in methanol solution: Linear and nonlinear vibrational spectra. *J. Chem. Phys.* **2004,** *120* (3), 1477-1490.
73. Woutersen, S.; Mu, Y.; Stock, G.; Hamm, P., Hydrogen-bond lifetime measured by time-resolved 2d-ir spectroscopy: N-methylacetamide in methanol. *Chem. Phys.* **2001,** *266* (2-3), 137-147.
74. Galle Kankanamge, S. R.; Ma, J.; Mackin, R. T.; Leonik, F. M.; Taylor, C. M.; Rubtsov, I. V.; Kuroda, D. G., Proving and probing the presence of the elusive c− h··· o hydrogen bond in liquid solutions at room temperature. *Angew. Chem.* **2020,** *132* (39), 17160-17165.
75. Krestyaninov, M. A.; Ivlev, D. V.; Dyshin, A. A.; Makarov, D. M.; Kiselev, M. G.; Kolker, A. M., Complex investigation of h-bond in water-n-methylacetamide system: Volumetric properties, dft, ir, md analysis. *J. Mol. Liq.* **2022**, 119533.
76. Lindquist, B. A.; Corcelli, S. A., Nitrile groups as vibrational probes: Calculations of the c≡n infrared absorption line shape of acetonitrile in water and tetrahydrofuran. *J. Phys. Chem. B* **2008,** *112* (20), 6301-6303.
77. Kim, Y. S.; Hochstrasser, R. M., Comparison of linear and 2d ir spectra in the presence of fast exchange. *J. Phys. Chem. B* **2006,** *110* (17), 8531-8534.
78. Kim, Y. S.; Hochstrasser, R. M., Chemical exchange 2d ir of hydrogen-bond making and breaking. *Proc. Natl. Acad. Sci. U.S.A.* **2005,** *102* (32), 11185-11190.
79. Aschaffenburg, D. J.; Moog, R. S., Probing hydrogen bonding environments: Solvatochromic effects on the cn vibration of benzonitrile. *J. Phys. Chem. B* **2009,** *113* (38), 12736-12743.
80. Waegele, M. M.; Gai, F., Computational modeling of the nitrile stretching vibration of 5-cyanoindole in water. *J. Phys. Chem. Lett.* **2010,** *1* (4), 781-786.
81. Wilson, E. B.; Decius, J. C.; Cross, P. C., *Molecular vibrations: The theory of infrared and raman vibrational spectra*. 1 ed.; Courier Corporation: 1980; p 417.