PCCP

ROYAL SOCIETY OF CHEMISTRY

PAPER

View Article Online
View Journal | View Issue



Cite this: *Phys. Chem. Chem. Phys.*, 2022, **24**, 26046

# Hydration dynamics and IR spectroscopy of 4-fluorophenol†

Seyedeh Maryam Salehi,[a] Silvan Käser,[a] Kai Töpfer,[a] Polydefkis Diamantis,[b] Rolf Pfister,[c] Peter Hamm,[c] Ursula Rothlisberger[b] and Markus Meuwly*[a]

Received 23rd June 2022,
Accepted 5th October 2022

DOI: 10.1039/d2cp02857c

rsc.li/pccp

Halogenated groups are relevant in pharmaceutical applications and potentially useful spectroscopic probes for infrared spectroscopy. In this work, the structural dynamics and infrared spectroscopy of *para*-fluorophenol (F-PhOH) and phenol (PhOH) is investigated in the gas phase and in water using a combination of experiment and molecular dynamics (MD) simulations. The gas phase and solvent dynamics around F-PhOH and PhOH is characterized from atomistic simulations using empirical energy functions with point charges or multipoles for the electrostatics, Machine Learning (ML) based parametrizations and with full *ab initio* (QM) and mixed Quantum Mechanical/Molecular Mechanics (QM/MM) simulations with a particular focus on the CF- and OH-stretch region. The CF-stretch band is heavily mixed with other modes whereas the OH-stretch in solution displays a characteristic high-frequency peak around 3600 $cm^{-1}$ most likely associated with the –OH group of PhOH and F-PhOH together with a characteristic progression below 3000 $cm^{-1}$ due to coupling with water modes which is also reproduced by several of the simulations. Solvent and radial distribution functions indicate that the CF-site is largely hydrophobic except for simulations using point charges which renders them unsuited for correctly describing hydration and dynamics around fluorinated sites. The hydrophobic character of the CF-group is particularly relevant for applications in pharmaceutical chemistry with a focus on local hydration and interaction with the surrounding protein.



## Introduction

Fluorination – and halogenation in general – are common chemical modifications for pharmaceuticals. Approximately 20% of all small molecule drugs used in medicinal chemistry contain X = F, Cl, Br, or I or a combination thereof. Among these compounds halogenated phenyl rings constitute an important class.[1] Because of the directionality of the interaction along the C–X bond due to the sigma hole, halogenation has emerged as one of the essential chemical modifications in medicinal materials,[2–4] and supramolecular chemistry.[5,6] By changing the halogen atom, the interactions with the environment can be tuned and the hydrophobicity around the modification site can be modulated.[3,7–12] The importance of halogenation as a fundamental concept in medicinal chemistry is highlighted by the improved binding affinities of several ligands towards their receptors.[13,14] Recently, halogenation has also been employed in the context of protein modifications, such as for insulin, to fine-tune thermodynamic stability and affinity to the insulin receptor.[15]

A halogen bond "[...]occurs when there is evidence of a net attractive interaction between an electrophilic region associated with a halogen atom in a molecular entity and a nucleophilic region in another, or the same, molecular entity."[16] Hence, halogen atoms act as electrophiles and can form an attractive interaction with a nucleophilic counterpart. Based on the analysis of the molecular surface electrostatic potential (ESP),[17] the "halogen bond" was also associated with a "$\sigma$-hole bond"[18] which is a noncovalent interaction between a covalently-bonded halogen atom X and a negative site, *e.g.* a lone pair of a Lewis base or an anion.[17] Such a "bond" involves a region of positive electrostatic potential, *i.e.* the $\sigma$-hole, and as an extension of one of the covalent bonds to the atom. The $\sigma$-hole arises as a consequence of the anisotropy of the ESP around the halogen atom. The strengths of the interactions generally correlate well with the magnitudes of the positive and negative electrostatic potentials of the $\sigma$-hole and the negative

[a] *Department of Chemistry, University of Basel, Klingelbergstrasse 80, CH-4056 Basel, Switzerland. E-mail: m.meuwly@unibas.ch*
[b] *Laboratory of Computational Chemistry and Biochemistry, Institute of Chemical Sciences and Engineering, École Polytechnique Fédérale de Lausanne (EPFL), CH-1015 Lausanne, Switzerland*
[c] *Department of Chemistry, University of Zurich, Switzerland*

† Electronic supplementary information (ESI) available: The supporting information provides Tables for the force field parametrizations (Tables S1–S3), frequency maxima for frequency distributions (Table S4), and parameters for the FFCFs (Table S5) together with Fig. S1–S12 the report quality of the PhysNet model, additional solvent distribution functions, dihedral time series and the FFCFs for the CF- and OH-stretch frequencies. See DOI: https://doi.org/10.1039/d2cp02857c

site. As fluorine has the largest electronegativity and the lowest polarizability, for some time it was in fact assumed that there is no $\sigma$-hole and that therefore fluorine is not involved in halogen bonding at all.[17,19,20] However it is now well established that it can have a positive $\sigma$-hole and form halogen bonds when it is linked to strongly electron-withdrawing groups including Cl, Br, and I.[21,22] Moreover, there is also experimental evidence for fluorine engaging in halogen bonding.[23]

Introducing a fluorine atom into organic molecules can cause major changes in the physico-chemical properties such as solubility, chemical reactivity and biological activity compared to non-fluorinated analogues.[24] In particular, fluorine often replaces hydrogen in organic molecules but the size and stereoelectronic influences of the two atoms (hydrogen *vs.* fluorine) are quite different albeit it is often regarded as isosteric substitution.[4] In bio-inorganic and medicinal chemistry, the formation of intermolecular O–H/F–C and N–H/F–C hydrogen bridges was assumed to be important in binding fluorinated compounds to enzyme active sites.[25] Such interactions affect enzyme ligand binding affinity, selectively coupled with the changes in pharmaco-kinetic properties by fluorine substitution.[10,24] The effects of fluorine substitution on the related pharmaco-kinetic properties like lipophilicity, volatility, solubility, hydrogen bonding and steric effects affect the resulting compound binding, absorption, transport and hence the related biological activity.[26] Finally, about 20% of the commercial pharmaceuticals contain fluorine[27] which underline the practical relevance of fluorinated compounds and further motivate to consider a fluorinated species in the following.

A variety of functional groups, including C–H, C–OH, C=O, and C≡N, have utilized the C–F bond as a bioisostere.[28] However, it is difficult to generalize the relative ability of fluorine to act similar to a hydrogen or hydroxy group, and different factors must be considered in each case. The van der Waals radius of fluorine (1.47 Å) lies between that of oxygen (1.57 Å) and hydrogen (1.2 Å) and as it is the element with the highest electronegativity, the C–F bond is almost identical to C–OH in terms of bond length and polarity. Despite its three electron pairs, the C–F bond interacts more weakly with the environment compared to an oxygen atom and is better described as ''weakly polar'' rather than ''hydrogen bonding''.[28,29] In pharmacological applications the replacement H → F is often considered to avoid metabolic transformation due to the high stability of the CF bond. Examples are drugs interacting with P450 for which fluorination has been widely used to block metabolic transformations.[28]

Given the different qualitative characterizations outlined so far, a more molecularly refined picture of the energetics and dynamics of fluorinated model compounds is warranted. For this, hydrated fluoro-phenol (F-PhOH) as a typical representative is considered. Using linear infrared (IR) spectroscopy together with computational characterizations at different levels of theory the structural dynamics and spectroscopy of F-PhOH is characterized. The computations use advanced empirical force fields including multipolar interactions, a machine-learned, neural network-based representation of the full-dimensional potential energy surface (NN-PES), mixed quantum mechanics/molecular mechanics and *ab initio* molecular dynamics (MD) simulation techniques. MD-based infrared spectra are well-established[30,31] and provide a powerful approach to characterized the structural dynamics in solution, in particular when combined with experiments[32–34] because the measured IR spectra can be directly compared with the computational results and can provide a structural interpretation of the spectroscopic features.

In the present work first the different energy functions employed are described and validated for F-PhOH in the gas phase. Next, results for F-PhOH in solution are presented. This is followed by a comparison of spectroscopy for F-PhOH and PhOH to quantitatively probe the consequences of the H → F replacement opposite the –OH group. Finally, the solvent structure around the two solutes is analyzed from radial distribution functions and from 2-dimensional solvent distributions.

# Methods

### Classical molecular dynamics simulations

All classical MD simulations were performed with CHARMM.[35] The bonded parameters are based on CGenFF[36] except for the CF and OH bond for which a Morse potential was used to describe their anharmonicity. To that end, a scan along the CF bond was performed at the MP2/aug-cc-pVTZ level starting from an optimized structure of F-PhOH at this level of theory. The energy of 49 points was computed on a grid ranging from $r$ = 0.75 Å to $r$ = 5.55 Å in increments of 0.1 Å. Then, the energies were fitted to a Morse potential $V(r) = D_0[1 - \exp(-\beta(r - r_0))]^2$ which yields parameters $D_0$ = 136.316 kcal mol$^{-1}$, $r_0$ = 1.349 Å, and $\beta$ = 1.603 Å$^{-1}$. For the OH bond the calculated Morse parameters are $D_0$ = 120.234 kcal mol$^{-1}$, $r_0$ = 0.971 Å, and $\beta$ = 2.088 Å$^{-1}$. To realistically describe the electrostatic interactions, a multipolar (MTP)[37–40] model was also used with MTPs on all heavy atoms up to quadrupoles and point charges for all hydrogen atoms. These parameters were fitted to the electrostatic potential using a fitting environment,[41] see Tables S1–S3 (ESI†).

Simulations for F-PhOH and PhOH were carried out in a cubic box of $30^3$ Å$^3$ ($28^3$ Å$^3$ for simulations with the NN-PES, see below) using TIP3P[42] water molecules. Minimization, heating, and equilibration procedures for 40 ps were employed to prepare the system at 300 K. This was followed by 5 ns production simulations in the *NVE* ensemble using the Velocity Verlet propagator.[43] The time step was $\Delta t$ = 1 fs and every fifth snapshot was recorded. Lennard-Jones interactions were computed with a 12 Å cutoff switched at 10 Å.[44] The electrostatic interactions for the monopoles (point charges) are treated using Particle-Mesh Ewald[45] (PME) with grid size spacing of 1 Å, characteristic reciprocal length $\kappa$ = 0.32 Å$^{-1}$, and interpolation order 6. All bonds involving hydrogen atoms are constrained *via* the SHAKE algorithm.[46,47] Additional MD simulations were also carried out for PhOH in water with the same setup that was used

for F-PhOH in order to directly compare their spectroscopy and solvent structure.

For the simulations with the NN-based PES (see below) the atomic simulation environment (ASE) was used.[48] The van der Waals interactions were those from the CGenFF[36] parametrization and the fluctuating charges are from the PhysNet representation, see below. For both terms interactions the cutoff distance is at 14 Å and switched between 13 to 14 Å. To avoid artifacts of the electrostatic Coulomb force in the cutoff range, the Coulomb force at the distance of 14 Å was shifted to zero in accordance to the shifted forces method.[49] In the gas phase 1000 trajectories, each 200 ps in length, are run to obtain an ensemble average. The *NVE* simulations are run at 300 K initialized from random momenta corresponding to a Maxwell–Boltzmann distribution, with a time step of 0.5 fs, equilibrated for 50 ps and propagated for 200 ps. Simulations in solution *NVT* using Langevin[50] thermostat at 300 K are performed for 20 trajectories of 100 ps each with a time step of 0.2 fs to obtain a total of 2 ns for PhOH and F-PhOH in solution, respectively. The IR spectra are then calculated from the dipole–dipole moment autocorrelation function[51–53] and averaged over all 1000 trajectories.

### Instantaneous normal mode analysis

From the production simulation, $10^6$ snapshots were taken as a time-ordered series for computing the frequency fluctuation correlation function (FFCF). The FFCF was determined from instantaneous harmonic vibrational frequencies based on a normal mode analysis. Such instantaneous normal modes (INM) are obtained by minimizing F-PhOH while keeping the surrounding solvent frozen. Next, normal modes were calculated using CHARMM for 5 modes ($\nu_1$ to $\nu_5$ in ascending order) between 1100 to 1400 $cm^{-1}$ in terms of participation ratio of CF stretch in that particular mode. In a separate analysis step, the participation ratios of the CF, CO, and CH stretch and the COH bending coordinates to these 5 normal modes were determined.

### Frequency fluctuation correlation function and lineshape

From the INMs the frequency trajectory $\omega_i(t)$ and the FFCF, $\langle\delta\omega(0)\delta\omega(t)\rangle$ was computed. Here, $\delta\omega(t) = \omega(t) - \langle\omega(t)\rangle$ and $\langle\omega(t)\rangle$ is the ensemble average of the transition frequency. From the FFCF the line shape function

$$g(t) = \int_0^t\int_0^{\tau'}\langle\delta\omega(\tau'')\delta\omega(0)\rangle d\tau'' d\tau'. \quad (1)$$

is determined within the cumulant approximation. To compute $g(t)$, the FFCF is numerically integrated using the trapezoidal rule and the 1D-IR spectrum is calculated from[54]

$$I(\omega) = 2\Re\int_0^\infty e^{i(\omega-\langle\omega\rangle)t}e^{-g(t)}e^{-\frac{t\alpha}{2T_1}}dt \quad (2)$$

where $\langle\omega\rangle$ is the average transition frequency obtained from the distribution, $T_1$ = 1.2 ps[55] is the vibrational relaxation time and $\alpha$ = 0.5 is a phenomenological factor to account for lifetime broadening.[54]

From the FFCF, the decay time is determined by fitting the FFCF to a general expression[56]

$$\langle\delta\omega(t)\delta\omega(0)\rangle = \sum_{i=2}^{n} a_i e^{-t/\tau_i} + \Delta_0 \quad (3)$$

where $a_i$, $\tau_i$ and $\Delta_0$ are fitting parameters. The decay times $\tau_i$ from the fits characterize the time scale of the solvent fluctuations. The absence of a minimum at short times ($\tau \sim 0.02$ ps) indicates that the interaction between F and environment is weak compare with situation in F-ACN or $N_3^-$.[55,57] The decay times $\tau_i$ of the FFCF reflect the characteristic time-scale of the solvent fluctuations to which the solute degrees of freedom are coupled. In all cases the FFCFs were fitted to an expression containing two decay times using an automated curve fitting tool from the SciPy library.[58]

### Full *ab initio* (QM) and mixed quantum mechanical/molecular mechanics (QM/MM) simulations

Full QM Simulations: The QM system was comprised of F-PhOH and 117 water molecules in a ($15.41 \times 15.44 \times 15.46$ Å$^3$) periodic box initially equilibrated classically at 300 K and 1 atm using CHARMM. The full QM equilibration and production phases lasted for 12.5 ps and 20.4 ps respectively. For the gas phase simulation, the total equilibration and production times were 23.0 ps and 28.1 ps, respectively. For the latter, the initial translations and rotations of the center of mass were removed.

For both the gas phase and the condensed phase systems, the full QM simulation protocol consisted of (i) an equilibration of the system at 300 K first with Born–Oppenheimer (BO) MD and then with Car–Parrinello (CP) MD,[59] and (ii) a production phase in the microcanonical (*NVE*) ensemble. The respective time steps for BO and CP MD were 10 and 2 atomic units (a.u.), respectively. In CP MD, the fictitious electron mass was equal to 400 a.u. In the production phase, frames were saved every 10 a.u., corresponding to a time interval of approximately 0.48 fs.

Density Functional Theory (DFT)-based *ab initio* MD simulations of F-PhOH in gas phase and in aqueous solution were carried out using the CPMD code[60] using the BLYP functional for the exchange and correlation energies[61,62] with the addition of Dispersion-Corrected Atom-Centered Potentials (DCACPs)[63–65] for the description of dispersion forces. Norm-conserving Martins–Trouiller pseudopotentials[66] were used in combination with a plane wave basis with a 175 Rydberg kinetic energy cutoff of for the expansion of the single-particle wavefunctions. The latter value was selected because it reproduces a converged equilibrium C–F bond distance at the BLYP-DCACP level of 1.35 Å for F-PhOH in gas phase, which is in good agreement with values obtained at the MP2/6-311++G(df,pd) (1.34 Å) and B3LYP/6-311++G(df,pd) (1.35 Å) levels, respectively.[67]

**Mixed QM/MM simulations.** Two QM/MM MD simulations were carried out for F-PhOH and PhOH in water, respectively, using the QM/MM interface of CPMD with the Gromos code[68] and the coupling scheme developed by Rothlisberger and coworkers.[69–71] The two systems were comprised of the solute

 

(F-PhOH or PhOH), and 331 and 311 water molecules, respectively. The system size was selected so that a direct comparison with the full QM simulation of F-PhOH in water can be made, and assess the impact of quantum description of the solvent including polarization effects on the geometric and spectral properties of F-PhOH.

The systems were first equilibrated classically, using AMBER18.[72] F-PhOH and PhOH were modelled with the GAFF2 force field,[73,74] while the TIP3P model was used for water. Following an initial minimization, the two systems were equilibrated in the isothermal–isobaric (*NPT*) (300 K, 1 atm) ensemble with the Berendsen barostat[75] and Langevin dynamics[50] for pressure and temperature control respectively, followed by further equilibration simulations in the *NVT* ensemble with Langevin dynamics, for a total of 100 ns. A time step of 2 fs was employed. In view of the small periodic box size, a reduced real space cutoff of 7 Å was used for the nonbonded interactions.

Following this preparation, production simulations with CPMD in the *NVE* ensemble were carried for both systems, whereby the solute was treated at the QM level and the solvent at the classical (MM) level. The QM setup, and the simulation time step were the same as described above for the full QM simulations. The QM/MM MD simulation protocol was also similar to the one described for the full QM simulation, apart from the use of two separate Nosé–Hoover thermostats for the QM and MM parts respectively, during the equilibration with BO and CP MD. For the F-PhOH system, the equilibration and production runs lasted 10.1 ps and 35.9 ps respectively, while for the PhOH system they lasted 12.6 ps and 25.0 ps, respectively. During the production phase, frames were saved with the same frequency as in the full QM simulations (0.48 fs).

### Machine-learned potential energy surface (PES)

To validate in particular the PC- and MTP-based simulations using an empirical force field a complementary model based on a machine-learned PES was also pursued. For this PhysNet,[76] a deep neural network (NN) of the message passing type,[77] was used to obtain a representation of the potential energy for both PhOH and F-PhOH. PhysNet uses Cartesian coordinates and nuclear charges to learn an atomic descriptor for the prediction of energies, forces, partial charges and molecular dipole moments to describe chemical systems and their properties, such as infrared spectra.

PhysNet was trained on *ab initio* energies, forces and dipole moments calculated at the MP2/6-31G(d,p) level of theory using Molpro[78] according to the protocol reported in ref. 76. The reference data, containing different geometries for both molecules, is generated from MD simulations at 50, 300 and 1000 K using CHARMM force field (5000 geometries each yielding a total of 30 000 geometries) and extended with geometries obtained from normal mode sampling[79] at temperatures between 10 and 2000 K (6600 geometries for each molecule). The complete data set thus contains 43 200 PhOH and F-PhOH structures. The performance of the PhysNet PES is reported in Fig. S1 (ESI†) which shows the correlation between the reference MP2 and the PhysNet energies for a test set of 3700 randomly selected points with $R^2$ = 0.9999 and an RMSE of 0.0037 eV. Simulations with PhysNet for the solute and an empirical water model are subsequently referred to as ML/MM MD.

### Experimental infrared spectroscopy

Experimental spectra of PhOH and F-PhOH in water have been measured in attenuated total reflection (ATR) geometry using a home-constructed ZnSe prism in a Bruker Tensor 27 FT-IR spectrometer. Identical concentrations of 0.76 M have been used for both samples, which is the saturation limit for F-PhOH. Given the low absorption cross sections of the IR modes of F-PhOH and PhOH, in particular of the Franck–Condon-like progression of the hydrogen-bond vibration, a high concentration was needed for trustable results after subtraction of the much stronger water background. As a background spectrum, pure water has been measured as well and subtracted from the PhOH and F-PhOH spectra. The subtraction procedure is deemed reliable in the reported spectra, with the exception of the range between $\approx$3100 cm$^{-1}$ to 3500 cm$^{-1}$, where the very strong OH stretch vibration dominates. In this range the difference spectrum displays a dispersion-shaped response which is attributed to a shift of the OH-stretch vibration to somewhat higher frequencies due to overall weaker water–water hydrogen bonds.

## Results

### Gas phase spectra

First, the performance of the PC- and MTP-based empirical force fields, of PhysNet, and of the DFT-BLYP/DCACP based *ab initio* MD simulations was assessed for F-PhOH in the gas phase. For this, MD simulations of F-PhOH in the gas phase were carried out and the power and infrared spectra were determined and compared with experiments.

Fig. 1 reports the infrared and CF-/OH-power spectra of F-PhOH from simulations with the experimental FT-IR spectrum[67] from 1100 to 1400 cm$^{-1}$ and between 2600 and 4000 cm$^{-1}$, respectively. For clarity, the left hand column shows the low-frequency vibrations whereas the right hand column is for the –OH-stretch region. The experiments were carried out in $CCl_4$ solvent and the spectral lines have a full width at half maximum of $\sim$10 cm$^{-1}$. As an indication for the solvent-induced shift incurred, for PhOH in $CCl_4$ the CO stretch is found at 1257 cm$^{-1}$ which amounts to a red shift of $\sim -5$ cm$^{-1}$ compared with the gas-phase frequency of 1261.7 cm$^{-1}$.[80,81] Hence, $CCl_4$-induced shifts for F-PhOH are expected to be a few wavenumbers as well. To the best of our knowledge, no gas-phase spectra are available for F-PhOH. Hence, the frequencies for F-PhOH measured in $CCl_4$ are used *in lieu* of gas phase data.

The measured CF and CO stretching modes occur mainly at 1226 and 1262 cm$^{-1}$, respectively, while they couple to one another and potentially to other modes. According to the analysis[67,81] (see Table 1), the CF stretch is coupled to the in plane bending of the ring and also the C–H bend while the CO

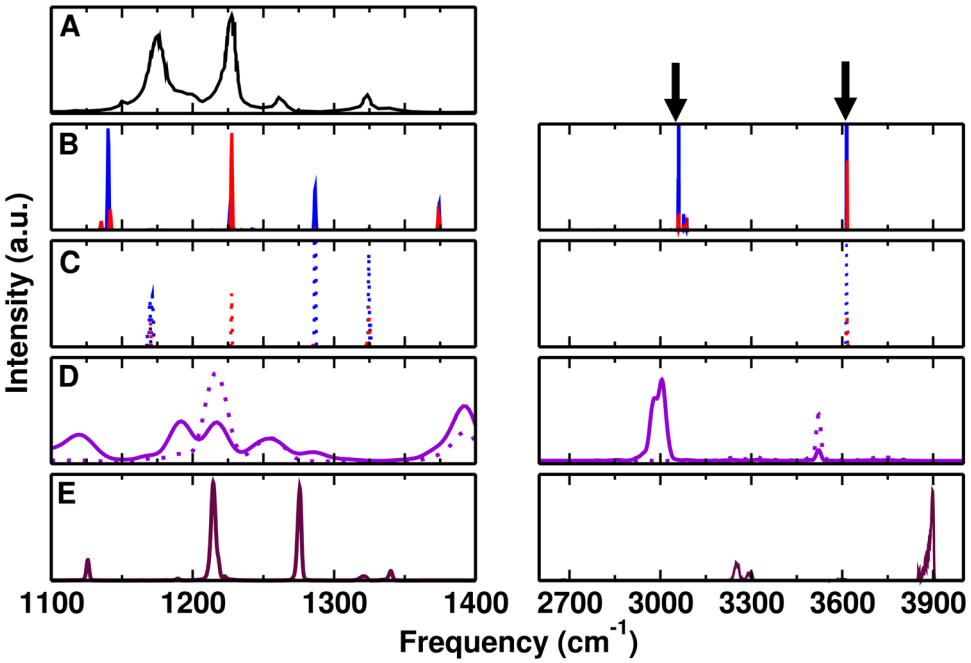


Fig. 1 Comparison of experimental (in $CCl_4$) and computed (gas phase) spectra for F-PhOH in the 1100–1400 $cm^{-1}$ (left) and 2600–4000 $cm^{-1}$ (right) frequency range. (panel A) Experimental (black) spectrum from ref. 67 extracted using g3data.[82] The arrows in the right hand column refer to the experimental frequencies for the CH and OH stretches.[67] (panels B and C) IR spectrum from the Fourier transform of the total dipole moment correlation function and CF/OH power spectra using PC (blue) and MTP (red) for F-PhOH. (panel D) Global (solid) and CF/OH (dotted) power spectrum in violet from QM simulations. (panel E) IR spectrum (maroon) from ML/MM MD simulations.

Table 1 Vibrational frequencies in $cm^{-1}$ for PhOH (gas phase) and F-PhOH (in $CCl_4$) in the range of 1100–1400 $cm^{-1}$.[67,80,81] The contributions (in terms of local deformations) to each vibrational mode indicate strong mixing and are those from the literature.[67,80] The assignment of the bands has been made on the basis of the calculated potential energy distribution.[67,83] Symbols $\nu$ and $\delta$ refer to stretching and bending modes, respectively

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| PhOH | 1150.7 | 1168.9 | 1176.5 | | 1261.7 | 1343 | | 3656 |
| | $\delta$ (CH) | $\delta$ (CH) | $\delta$ (OH) | | $\nu$ (CO) | $\delta$ (CH) | | $\nu$ (OH) |
| | $\nu$ (CC) | $\nu$ (CC) | $\delta$ (CH) | | $\delta$ (CH) | $\delta$ (OH) | | |
| | $\delta$ (OH) | | $\nu$ (CC) | | | | | |
| F-PhOH | 1149 | 1174 | | 1226 | 1262 | 1310 | 1323 | 3613 |
| | $\delta$ (CH) | $\delta$ (OH) | | $\nu$ (CF) | $\nu$ (CO) | $\delta$ (CH) | $\nu$ (CC) | $\nu$ (OH) |
| | | $\nu$ (CC) | | $\delta$ ring | $\nu$ (CC) | $\nu$ (CC) | $\delta$ (OH) | |
| | | $\delta$ (CH) | | $\delta$ (CH) | $\nu$ (CF) | | $\delta$ (CH) | |

stretch couples to the CC and CF stretching vibrations. Therefore, the CF stretch is intimately coupled with other modes and for that reason it is not possible to assign a local CF stretching mode to one particular frequency.

For the force field simulations with PC and MTP the $\beta$-parameter of the CF-Morse potential (see Methods) was slightly adjusted to $\beta$ = 1.665 $Å^{-1}$ to correctly describe the experimental spectrum in $CCl_4$ (Fig. 1A), hence the favourable comparison with the IR spectra in Fig. 1B. The CF power spectrum (Fig. 1C left) clarifies that this mode couples strongly to other vibrations (*e.g.* the CO-stretch) close in frequency for both, PC (blue) and MTP (red) force fields. The peak structure from the CF-power spectrum in Fig. 1C left matches that from the experiment whereas for the IR spectrum in panel B this is only qualitatively the case. This difference arises because power spectra are determined from temporal variations of selected internal coordinates whereas the infrared spectrum is based on the total molecular dipole moment autocorrelation function.[57] For the QM simulations in the gas phase (Fig. 1D) the two main peaks of QM are at 1217 and 1256 $cm^{-1}$ and appear to be shifted by 5 to 6 $cm^{-1}$ with respect to experiments at 1222 and 1262 $cm^{-1}$ which were recorded at $T$ = 300 K. Finally, MD simulations using the PhysNet representation of the MP2/6-31G(d,p) reference data the IR-spectrum in Fig. 1E shows two strong peaks at 1215 $cm^{-1}$ and 1276 $cm^{-1}$.

In the region of the OH-stretch vibration, the experimental spectrum for F-PhOH in $CCl_4$ reports a band at 3613 $cm^{-1}$ (black arrow)[67] compared with 3614 $cm^{-1}$ and 3616 $cm^{-1}$ from PC and MTP simulations, respectively. Using PhysNet, the main peak in the gas phase is at 3889 $cm^{-1}$ (harmonic frequency at 3882 $cm^{-1}$ at the MP2/6-31G(d,p) level; corrected frequency at 3655 $cm^{-1}$ by multiplying with 0.94 for this level of theory[84]) whereas the DFT-BLYP/DCACP QM simulations report the OH stretch at 3522 $cm^{-1}$, somewhat shifted to the blue and red, respectively, compared with experiment. Moreover, there are signatures in the infrared spectra due to the CH-stretch vibrations (black arrow) around 3035–3077 $cm^{-1}$ which are also observed in the MD simulations. For PhysNet the corresponding peak is at 3256 $cm^{-1}$ (3060 $cm^{-1}$ after correction with a scaling factor of 0.94) whereas for QM at the DFT-BLYP/DCACP level the absorption is at 3006 $cm^{-1}$. The power spectra from the PC and MTP simulations (Fig. 1C) confirm that the OH-stretch is a local mode.

 

In summary, the gas phase spectrum from finite-$T$ MD simulations find comparable patterns for the frequencies in the 1100–1400 $cm^{-1}$ region when compared with experiment. It is also found that the CF stretch is coupled to other modes in this spectral range and no local mode for this motion can be assigned.

### Spectroscopy and dynamics of F-PhOH in water

After assessing the energy functions considered in the present work, the spectroscopy of F-PhOH in solution is analyzed, see Fig. 2. The experimentally measured spectrum from the present work is the black trace in Fig. 2A with two prominent bands at 1201 and 1222 $cm^{-1}$ in the CF-stretch region together with additional unresolved shoulders to higher energy, superimposed on a broad background extending from 1170 to 1270 $cm^{-1}$.

The MTP/MD simulations for the CF power spectrum (panel C) show two prominent peaks at 1187, and 1264 $cm^{-1}$ which approximately line up with the features in the infrared spectrum (panel B) but are displaced from those observed experimentally. The peak at 1171 $cm^{-1}$ is red shifted compared to the double peak at 1201 and 1222 $cm^{-1}$ of experimental spectrum while the peak at 1264 $cm^{-1}$ is blue shifted or captured at the same position compared to two additional experimental peaks at 1244 and 1264 $cm^{-1}$. Furthermore, the smaller peak at 1357 $cm^{-1}$ is also red shifted compared to 1368 $cm^{-1}$ from experiment. The corresponding infrared spectra in panel B align with the features at 1260 $cm^{-1}$ but the signal at 1187 $cm^{-1}$ from the power spectra has no oscillator strength in the infrared spectrum. Instead, a peak in the infrared appears at 1171 $cm^{-1}$.

The *ab initio* MD simulations in solution for the CF-power spectrum (Fig. 2D solid violet) find two prominent bands at 1207 and 1262 $cm^{-1}$ compared with band maxima at 1217 $cm^{-1}$ with faint shoulders below 1200 $cm^{-1}$ from gas-phase simulations, see Fig. 1D. The two prominent bands are also found from QM/MM simulations (solid orange) with band maxima at 1225 and 1266 $cm^{-1}$. The CF-power spectra assign these features to CF-stretch involving motions and accounting for the scaling for BLYP calculations ($\sim$0.99 largely independent of basis set)[84] they shift to the red which is consistent with the experiments. Thus, the splitting between the two peaks decreases from 55 $cm^{-1}$ for the full QM simulations to 41 $cm^{-1}$ indicating the sensitivity of the solvent interactions and the resulting frequency shifts and splittings in this spectral range.

For ML/MM MD simulations a broad band is observed between 1220 to 1290 $cm^{-1}$ in the IR spectrum which is blue shifted compared to the experiment. The CF-power spectrum with peak maximum at 1273 $cm^{-1}$ (dotted maroon trace in Fig. 1E) indicates that part of this broad IR-lineshape is due to CF-stretching motion. Moreover, the CO power spectrum has a first peak at 1250 $cm^{-1}$ which contributes to the broad IR peak below 1300 $cm^{-1}$ and a second, prominent peak at 1323 $cm^{-1}$,

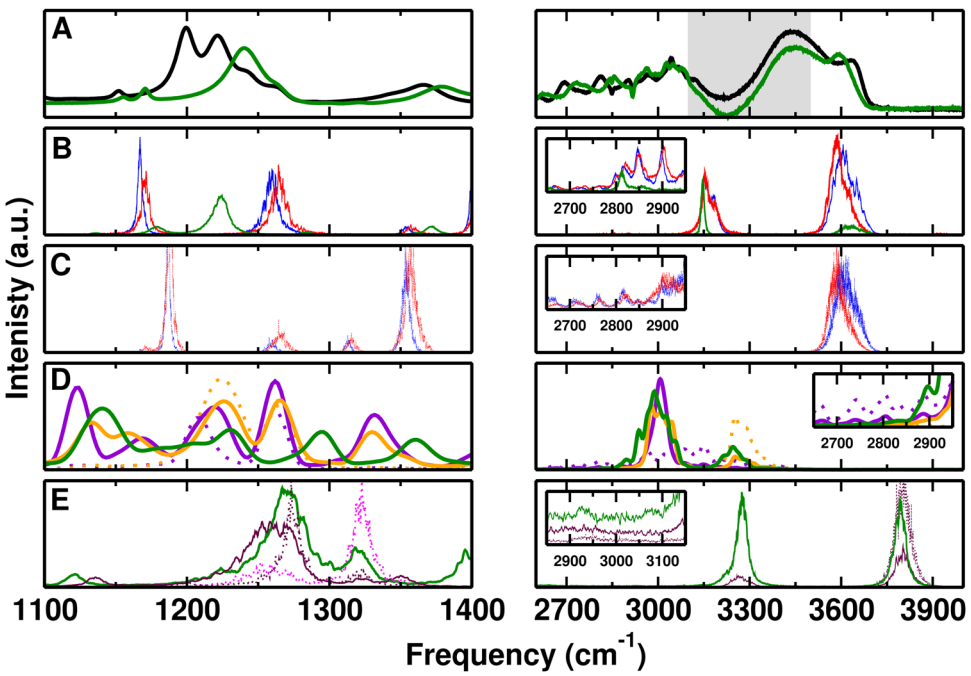


Fig. 2 Comparison of experimental and computed spectra for F-PhOH and PhOH in solution for the ranges (1100–1400 $cm^{-1}$) and (2600–4000 $cm^{-1}$). (panel A) Measured spectrum for F-PhOH (black) and PhOH (green). The gray area from 3100 to 3500 $cm^{-1}$ is unreliable due to incomplete subtraction of the water background. (panel B) IR spectra for F-PhOH (PC (blue), MTP (red)) and PhOH (MTP (green)). (panel C) CF/OH power spectra for F-PhOH from PC (blue) and MTP (red) simulations. (panel D) Global (solid) and CF/OH (dotted) power spectrum of the solute only (dotted) for F-PhOH (QM (violet) and QM/MM (orange)) and PhOH (QM/MM (green)). (panel E) IR (solid) and CF/OH (dotted) power spectrum for F-PhOH (maroon) and PhOH (green) from ML/MM simulations. The magenta dotted line is the CO power spectrum of F-PhOH. Insets in (panels B–D) show that the structured spectrum below 3000 $cm^{-1}$ is also found from the simulations whereas with PhysNet this is less realistically captured compared with experiment.

see Fig. 1E. The overlapping peaks of the CF- and CO-power spectra indicate coupling between the two types of motion.

The high-frequency region of the experimental spectrum for solvated F-PhOH (Fig. 2A) above and below 3000 $cm^{-1}$ involves a broad absorption extending from $\sim$2700 to 3100 $cm^{-1}$, a region between 3100 and 3500 $cm^{-1}$ that is not reliable due to the dominating background from the OH stretch vibration of bulk water which can not be subtracted off completely (grey area in Fig. 2A), and a high-frequency feature at 3643 $cm^{-1}$ assigned to the free OH vibration presumably originating from F-PhOH but possibly also from $H_2O$. This latter assignment is less likely, though, as the same sharp peak appears for PhOH in solution but shifted to the red by $\sim$40 $cm^{-1}$. If the signal was due to water it is expected to occur at closer frequencies given the similarity of the solutes. Furthermore, experiments on hydrated PhOH with up to 49 water molecules find the water–OH-stretch vibration at $\sim$3700 $cm^{-1}$.[85] From simulations with PCs and MTPs the high frequency peak at 3607 $cm^{-1}$ is consistent with experiment. Comparison with the spectrum for F-PhOH in $CCl_4$ and PhOH in the gas phase shows that this signal corresponds to the "free OH stretch" vibration, see Fig. 1. Features at 3150 $cm^{-1}$ in the infrared spectrum (Fig. 1B) are due to the CH-stretch vibrations which could be brought into better agreement with experiment by slight reparametrization of the force constants. These features are not present in the OH-power spectra, Fig. 1C, as expected which confirms the assignment to the CH-stretch vibration. The structured spectrum below 3000 $cm^{-1}$ is present in both, the PC and MTP simulations, albeit with lower intensity.

For the full QM and QM/MM simulations the global power spectra (Fig. 2D solid violet and orange) find a signal centered at 3000 $cm^{-1}$ which is typical for the CH-stretch modes. At higher frequency ($\sim$3250 $cm^{-1}$) the OH-stretch vibration is located which is confirmed by the OH-power spectra (dotted violet and orange traces). However, no signal in the 3600 $cm^{-1}$ region is present which suggests that the "free OH" signature in these simulations, expected around 3500 $cm^{-1}$ from the QM gas phase simulations (Fig. 1D), is absent.

Simulations with the PhysNet energy function primarily find the high frequency –OH stretch at 3803 $cm^{-1}$ with broad, largely unstructured undulations below 3000 $cm^{-1}$. It is likely that the MP2/6-31G(d,p) level is not sufficient for quantitatively describing the spectroscopy of F-PhOH. Accounting for a frequency scaling of 0.94 for harmonic frequencies[84] shifts all frequencies to the red which is more consistent with the experimentally determined spectra. Specifically, the 3800 $cm^{-1}$ and 3280 $cm^{-1}$ band maxima shift to 3572 $cm^{-1}$ and 3083 $cm^{-1}$, both of which are consistent with OH- and CH-stretching motions.

The broad feature below 3000 $cm^{-1}$ from the experiments deserves additional attention. Regular signatures in this frequency range were previously reported for thin film liquid PhOH and solid PhOH[86] and for PhOH at the air/water interface.[87] Such regular structures have been observed also in other hydrogen-bonded systems, such as the acetic acid dimer, and are typically used to characterize a medium-strong hydrogen bond.[88] They are attributed to a Franck–Condon-like progression of the hydrogen-bond vibration (with a frequency of *ca.* 50 $cm^{-1}$ in the present case) that is anharmonically coupled to the high-frequency OH stretch vibration.

Simulations for F-PhOH in solution with PC/TIP3P, MTP/TIP3P, and full QM show an extended spectroscopic response in this frequency range with pronounced peaks superimposed which are washed out in the ML/TIP3P simulation and entirely absent in the QM/MM simulations. This suggests that the spectroscopic signature below 3000 $cm^{-1}$ is due to coupling between the H-bonding motion of water around the –COH part of F-PhOH which is primarily sensitive to the nonbonded interactions. To assess whether or not flexibility of the water solvent also affects the spectral signatures, simulations with the reparametrized,[89,90] flexible KKY (Kumagai, Kawamura, Yokokawa) model were carried out.[91] One 5 ns simulation with a time step of $\Delta t$ = 0.25 fs for F-PhOH was run and analyzed. The power spectrum of the F-PhOH OH-stretch vibration confirms the pronounced, regular pattern with peaks separated by some $\sim$50 $cm^{-1}$ below 3000 $cm^{-1}$, see Fig. S2 (ESI†). In addition, the main peak between 3300 $cm^{-1}$ and 3600 $cm^{-1}$ shifts to the red by 78 $cm^{-1}$ compared with simulations using the rigid TIP3P water model. This confirms that the pattern below 3000 $cm^{-1}$ is due to anharmonic coupling through nonbonded interactions between solute and solvent and not caused by the water internal modes.

For assessing solvent-induced frequency shifts, the frequency distributions from 5 ns simulation of hydrated F-PhOH, analyzed with instantaneous normal modes (INM) for the PC and MTP model are compared with the normal modes from gas phase simulations, see Fig. 3 and/or Table S4 (ESI†) for the frequency maxima. For obtaining the instantaneous normal modes the water environment was frozen and the structure of the solute was optimized, followed by a normal mode calculation. The band positions compare well with the frequencies from Table 1. However, the bimodal distribution around 1250 $cm^{-1}$ for both, simulations with PC and MTP, can not be convincingly correlated with the experimental spectra.

The participation ratios (see Methods) of the local modes to the frequencies of the five modes in the 1100–1400 $cm^{-1}$ range

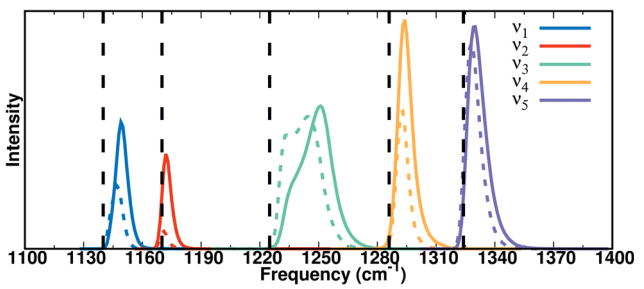


Fig. 3 Instantaneous vibrational frequency distributions from 5 ns MTP (solid colored line) and PC (dashed colored line) simulations of F-PhOH in water for five modes between 1100 and 1400 $cm^{-1}$. The black dashed lines are the harmonic frequencies for the optimized structure in the gas phase using the PC model.

 

($\nu_1$–$\nu_5$) from the MTP simulations are shown in Fig. S3 (ESI†). These were determined from the normal modes of F-PhOH over $10^5$ snapshots in solution. The contributions of the CF, CO, and CH stretch and the COH bending modes to each of the vibrations between 1100 and 1400 $cm^{-1}$ were determined by projection and the results confirm mixing of these modes.

In summary, the simulations confirm that the modes in the 1100–1400 $cm^{-1}$ frequency range in F-PhOH are strongly coupled. Assignment of individual spectral features from comparing experiment with simulations is not obvious. Consistent with experiment the force field-based simulations for the high-frequency modes find a high-frequency ($>3600$ $cm^{-1}$) phenolic –OH stretch together with broad features below 3000 $cm^{-1}$. These extended absorptions are also found from QM MD simulations without, however, the high-frequency –OH stretch.

### Comparison of the spectroscopy for hydrated F-PhOH and PhOH

To put the spectroscopy of F-PhOH in solution into context, direct comparison with PhOH provides additional insights. The experimental spectra for F-PhOH (black) and PhOH (green) in water are reported in Fig. 2A. For the experimental spectra between 1100 $cm^{-1}$ and 1400 $cm^{-1}$ pronounced differences between the two compounds are found. Most prominently, the single band with maximum at 1242 $cm^{-1}$ for PhOH, which is one of the ''X-sensitive modes'' involving considerable motion of the phenol ring and the CO group,[86] is shifted to the red for F-PhOH and split into at least two (at 1201 and 1222 $cm^{-1}$), but possibly several more peaks, some of which overlap with the peak from PhOH. Other features, such as the broader band with peak maximum at 1381 $cm^{-1}$ for PhOH are also shifted to the red (band maximum at 1368 $cm^{-1}$) for F-PhOH, see Fig. 2A.

Consistent with experiment, the number of spectral features for F-PhOH is larger than for PhOH in the 1200 to 1300 $cm^{-1}$ range. However, none of the computed spectra display the pronounced double-peak structure above 1200 $cm^{-1}$ for F-PhOH with the broad peak for PhOH to the blue of it. The simulations using MTPs find a single absorption between the low- and high-frequency absorption in F-PhOH, see Fig. 2B (green). Considering the CO-power spectrum the feature at 1224 $cm^{-1}$ involves the CO-stretch vibration together with a band at 1283 $cm^{-1}$. The QM/MM simulations report a larger number of spectroscopic features for PhOH (Fig. 2D, green) than experiment does. In particular, the single absorption at 1242 $cm^{-1}$ is not present but rather a broad absorption extending from 1150 up to $\sim$1250 $cm^{-1}$ is found. Finally, simulations using the PhysNet energy function quite well capture the absorption for PhOH at 1270 $cm^{-1}$ (Fig. 2E, green) with an additional peak above 1300 $cm^{-1}$ not present in the experiment. Correcting the position 1270 $cm^{-1}$ by 0.94 as was done for the CH- and OH-stretch vibrations shifts this band too far to the red compared with experiment, probably because for coupled vibration the standard correction factor is inappropriate. The absorption for F-PhOH is shifted to the red, in agreement with experiment but does not exhibit the double peak structure. Finally, it should be noted that the experiments are carried out for solute concentrations at which aggregation of F-PhOH and PhOH molecules can not be entirely excluded, which might affect both, the position of the absorption frequency and the line shapes of the various modes.

For the high-frequency part (Fig. 2A right panel) the spectra of solvated PhOH and F-PhOH follow each other closely except for a pronounced absorption at 3596 $cm^{-1}$ in PhOH which blue-shifts to 3643 $cm^{-1}$ upon fluorination. This suggests that either electronic coupling between the CF- and OH-sites leads to a slightly stronger OH-bond strength in F-PhOH compared with PhOH, or that the hydration structure around –OH is affected by fluorination, or a combination of the two. From normal mode calculations (MP2/6-31G(d,p)) the OH-stretch vibrations are at 3833 $cm^{-1}$ and 3829 $cm^{-1}$ for PhOH and F-PhOH, which is an insignificant difference and suggests that an electronic origin for the shift is unlikely. For the MTP (3636 and 3588 $cm^{-1}$) and ML/MM (3794 and 3793 $cm^{-1}$) simulations the –OH stretch for PhOH and F-PhOH reproduce the proximity of the two absorptions in solution, but in reverse order compared with experiment. The QM/MM simulations for PhOH (green trace Fig. 2D) report the OH-stretch at 3247 $cm^{-1}$ which is to the red of that for F-PhOH (3257 $cm^{-1}$, orange trace), in agreement with experiment. However, the absorptions are shifted by about 300 $cm^{-1}$ to the red relative to the experimental line positions.

### Radial distribution functions and solvent distribution

Radial distribution functions provide information about solvent-solute interactions. For the F–$O_{WAT}$ and F–$H_{WAT}$ distances they are reported in Fig. 4 for the PC, MTP, and ML/MM MD simulations (panels A and B) and for the QM/MM and QM simulations in panels C and D. For $g_{F\text{-}OW}$ the position of the first maximum agrees quite favourably for all methods except for simulations with PCs for which the maximum is shifted to shorter separations and the first peak is unusually sharp. Qualitatively, the four other methods find comparable shapes although the two QM simulations have a more pronounced first minimum than MTP and ML/MM simulations which may be related to a somewhat stronger interaction between the fluorine and the water-oxygens or to the shorter sampling time. For the corresponding $N(r)$, see Fig. S5 (ESI†).

The F–$H_{WAT}$ pair correlation function $g_{F-H_W}$ shows even more pronounced differences between simulations with PCs compared with all other models. The peak at 1.75 Å points towards a strong, favourable interaction between solvent hydrogen atoms and the fluorine atom which is not found in any of the other four methods. This can be explained by the negative partial charge $q_F = -0.29e$ on the Fluorine atom in the PC model. The MTP and ML/MM simulations find comparable distribution functions whereas the two QM-based simulations differ from this in that the broad first maximum is peaked at around 3 Å with the full QM simulations and shows a more pronounced local minimum. All distribution functions except that with PCs report the first maximum at an F–$O_W$ separation of $\sim$3 Å which points towards a largely hydrophobic behaviour of the CF site. This is also consistent with notions from

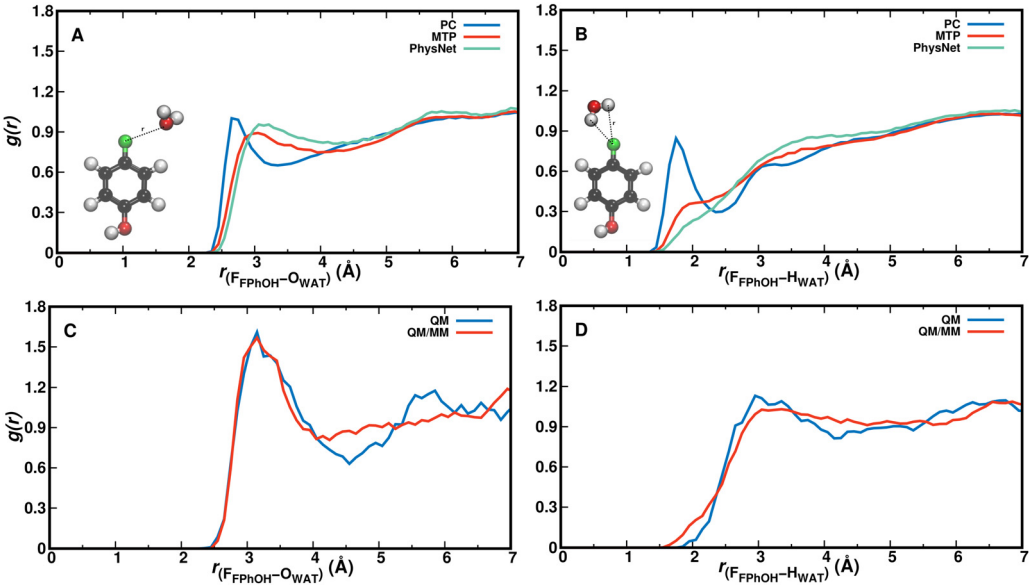


Fig. 4 The $g(r)$ for F–$O_{WAT}$ (panel A) and F–$H_{WAT}$ (panel B) separations from PC (blue), MTP (red) and PhysNet (green) simulations of F-PhOH in $H_2O$. Using PCs both $g(r)$ are more structured whereas the radial distribution functions from MTP and PhysNet – with conformationally fluctuating charges – are rather similar to one another. Both, MTP and PhysNet, point to weak interaction between the fluorinated end and the environment. Panels (C and D) show that results from QM and QM/MM simulations are similar with a somewhat more pronounced structuring for the $g_{F-O}(r)$ distribution function. The strong peak for $g_{F-H}(r)$ found in the simulation with PCs but absent for MTP and PhysNet (panel B) are also not found from QM and QM/MM simulations (panel D).

pharmaceutical chemistry in that a CF group reduces ligand solvation and increases its hydrophobicity.[28] The number of water molecules within distance $r$ is reported in Fig. S4 (ESI†) and shows that for small F–water separations ($r \leq 4$ Å) the occupation from simulations with MTP and PhysNet is similar but clearly below that of QM/MM and QM simulations (which are identical) whereas for larger separations ($r \sim 5$ Å) that from QM simulations approaches the MTP and PhysNet simulations. For the corresponding $N(r)$, see Fig. S6 (ESI†).

For analyzing hydration around the hydroxyl group the $H_{OH}$–$O_{WAT}$ and $H_{OH}$–$H_{WAT}$ pair correlation functions were considered, see Fig. 5. From simulations using the PC, MTP, and PhysNet models $g_{H_{OH}-O_W}$ characterizing the hydrogen bond between OH and water-oxygen atoms is similar up to and

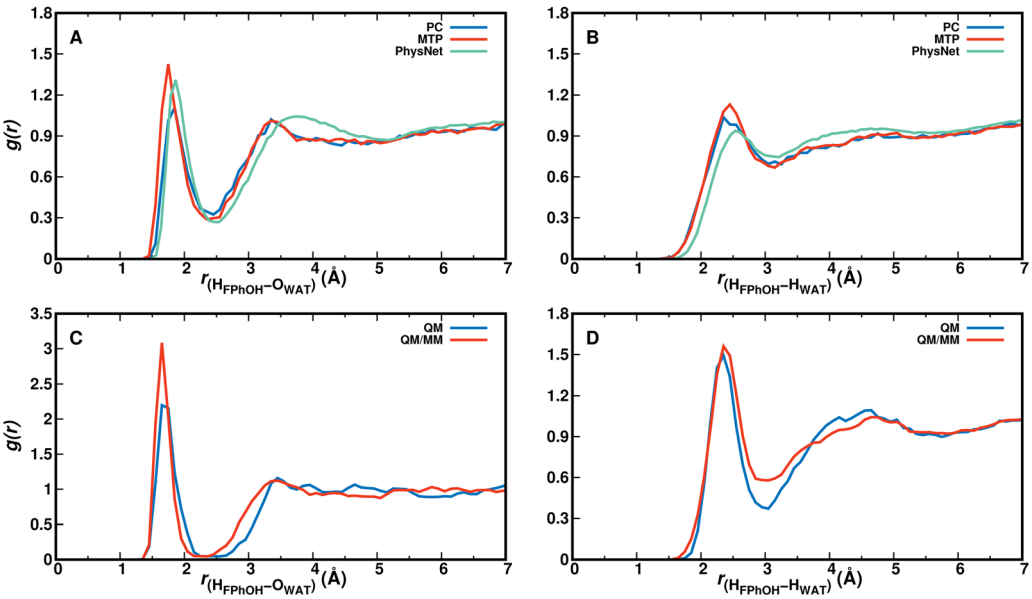


Fig. 5 The $g(r)$ for $H_{FPhOH/PhOH}$–$O_{WAT}$ (panel A) and $H_{FPhOH/PhOH}$–$H_{WAT}$ (panel B) distances between the H-atom of the solute OH-group as obtained from PC (blue), MTP (red) and PhysNet (green) simulations for F-PhOH in $H_2O$. Panels (C and D) from QM and QM/MM simulations. The scaling along all $x$-axes is identical whereas that along the $y$-axis is not.

including the first minimum, see Fig. 5A. The second maximum shifts to somewhat larger separations for the PhysNet simulations. Similar observations are made for $g_{H_{OH}-H_W}$ in panel B. The hydrogen-bond pair distribution function with a first maximum at ~1.6 Å suggests that solvent water molecules are quite strongly bound to the –OH group and that the solvent water at the –OH group exchanges.

For $g_{H_{OH}-H_{OW}}$ from the QM and QM/MM simulations (blue and red traces in Fig. 5C) the position of the first maximum is almost identical whereas the height of the first peak differs. Both pair correlation functions have a first minimum around 2.5 Å with an amplitude close to 0 which suggests that during the 25 ps simulation one solvent water molecule is strongly bound to the OH-group of F-PhOH and does not exchange with the surrounding solvent. As a consequence the "free phenolic –OH stretch" (spectroscopic feature around and above 3600 $cm^{-1}$, Fig. 2) is absent in the QM MD simulations, in contrast with what was found from the experiments and from the PC, MTP, and ML/MM MD simulations. On the other hand, the radial distribution function between the water–oxygen atoms and the two carbon atoms flanking the COH group in F-PhOH from a 5 ns MTP simulation (Fig. S7, ESI†) demonstrates that the solvent distribution dynamics is exhaustively sampled on the 5 ns time scale.

For the phenolic oxygen as an H-bond acceptor the $O–H_{WAT}$ and $O–O_{WAT}$ pair correlation functions for F-PhOH are reported in Fig S8 (ESI†). Panels A and B provide a direct comparison of the three force field-based simulations whereas panels C and D are those from the QM and QM/MM simulations, respectively. The MTP simulations (red) find strongest localization of the water followed by PC (blue) and PhysNet (green) simulations. For the QM simulations the first minimum for $g_{O-O_W}$ is deeper than for the QM–MM simulations whereas the position and height of the first maximum are comparable. The radial distribution functions from simulations with MTP are closest to those from the QM and QM/MM simulations, respectively.

Two-dimensional solvent distribution functions were generated from the positions of the water–oxygen atoms around F-PhOH. For that, the structures of the 5000 snapshots were oriented with C1 in the origin, the C1–C4 bond along the *x*-axis the [C1, C4, H] atoms in the *xy*-plane. A 2-dimensional histogram of the water positions was generated and then refined from kernel density estimation using Rstudio.[92] The distribution of the solvent water around the CF-part of F-PhOH is comparatively flat for all simulations with PC, MTP and PhysNet when contrasted with the COH-moiety of the solute, see Fig. 6. It is also found that for PhOH (top line) the solvent distributions resemble those for F-PhOH (bottom line). The asymmetry in the solvent distribution around the –COH group is due to the reference atoms chosen for the superposition of all structures. If the phenol-oxygen atom is excluded in reorienting the structures the 2d distribution becomes manifestly symmetric

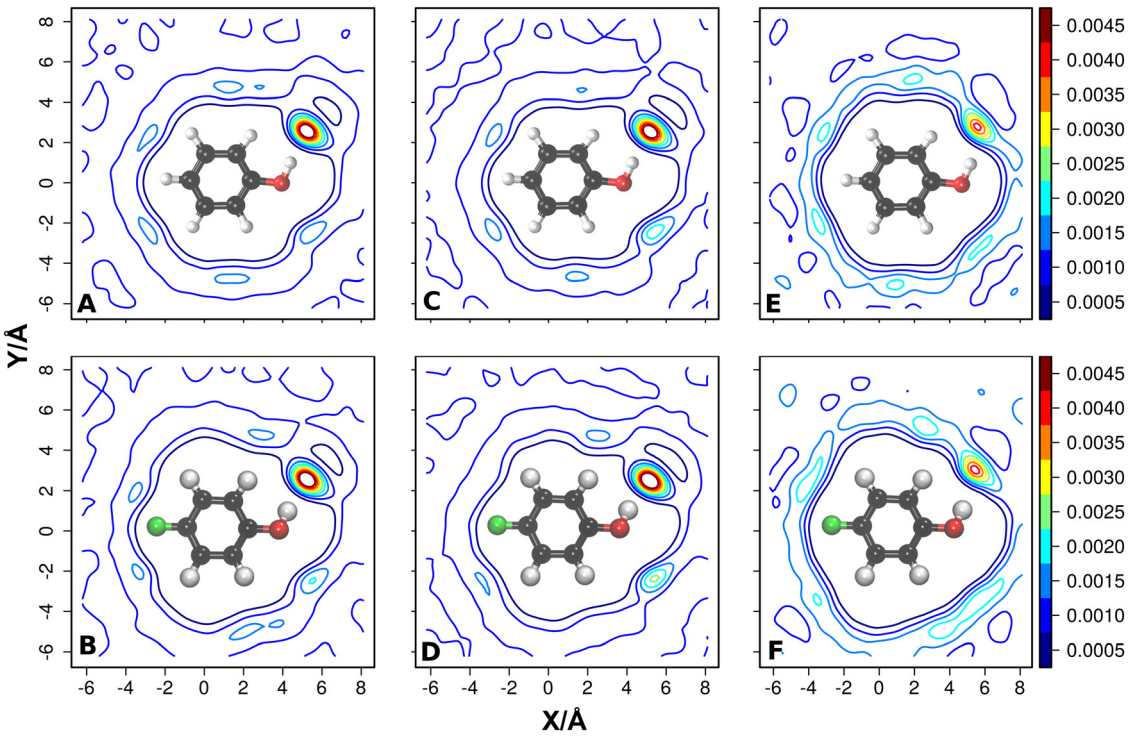


Fig. 6 The 2-dimensional solvent distributions from reorienting snapshots with respect to all heavy atoms of the solute, including the phenolic oxygen atom. Solvent distributions around PhOH and F-PhOH for PC (panels A and B), MTP (panels C and D) and from PhysNet (panels E and F) model. The isocontour values are shown in each panel. For the solvent distribution upon reorienting with respect to all heavy atoms excluding the phenolic oxygen atom, see Fig. S9 (ESI†) which manifestly shows symmetric water distribution according to the underlying dynamically averaged spatial symmetry of the solute due to rotation of the OH group around the –COH axis, see Fig. S10 (ESI†).

(Fig. S9, ESI†) and also clarifies that the –OH group of the solute rotates on the time scale of the simulations, see Fig. S10 (ESI†). Specifically, rotation of the –COH group is an activated process and occurs on the ~100 ps time scale from MTP and ML/MM simulations.

The weak interaction between solvent and solute around the CF-site of F-PhOH can also be gleaned from the behaviour of the frequency fluctuation correlation functions (FFCFs). For this, the FFCFs were determined from the frequency trajectories $\omega_i(t)$ for the five modes in the CF-stretch region, *i.e.* $\nu_1$ to $\nu_5$, between 1100 $cm^{-1}$ and 1400 $cm^{-1}$ and for the OH-stretch vibration, see Fig. S11 and S12 (ESI†). The FFCFs contain information about the coupling between a particular mode and the environmental dynamics. The fitting parameters for $\nu_1$ to $\nu_5$ in Table S5 (ESI†) show that the fast correlation is generally $\tau_1 \sim 0.1$ ps whereas the longer time scale ranges from $\tau_2 = 0.28$ ps to $\tau_2 = 0.83$ ps. Such short correlation times point towards weak solvent/solute interactions close to the CF-site. The static components $\Delta_0$ are very small, too, which also imply rapid, unspecific dynamics around the CF-group. For the OH-stretch vibration, the fast correlation time is $\tau_1 \sim 0.1$ ps while the longer one is $\tau_1 \sim 0.5$ ps which are comparatively short. Compared with the FFCFs for vibrations involving the CF-stretch the $t = 0$ amplitude and the static offset $\Delta_0$ is larger by one to two orders of magnitude which points towards somewhat slower dynamics around the OH bond of F-PhOH.

## Discussion and conclusion

The present work discusses the infrared spectroscopy of and solvent distribution around F-PhOH in water from experiments and a range of computational approaches including full QM, QM/MM, PC-based, MTP-based and ML/MM simulations. The simulations show that the modes in the range between 1200 and 1300 $cm^{-1}$ are heavily mixed which complicates the assignment of the spectroscopic features. In this frequency range and in the gas phase the ML/MM simulations capture the splitting of the bands best, see Fig. 1, although the computed frequencies are shifted to the blue compared with experiment. This is known for MP2/6-31G(d,p) calculations (scaling of 0.94 for harmonic frequencies). For F-PhOH in solution the splitting from the MTP-based simulations is too large, the QM/MM and QM simulations yield a larger number of bands than experimentally observed and the ML/MM simulations have again all frequencies shifted to the blue and the clear splitting found from experiments is not captured although the CF and CO stretch frequencies are displaced by approximately the correct amount in the power spectra. The solute–solvent pair distribution functions together with the FFCFs indicate that interaction between the fluorinated position and the environment is weak as has also been found earlier for fluoro-acetonitrile.[55]

For the phenol-OH-stretch vibration both, ''gas phase'' (around 3600 $cm^{-1}$) and ''water–phenol hydrogen bonded'' signatures (below 3000 $cm^{-1}$) are found from experiments and the simulations. A regular pattern with a spacing of ~50 $cm^{-1}$ is reminiscent of recent SFG spectra of PhOH at the air/water interface and earlier experiments on liquid and solid pure PhOH[86] and is assigned to water–phenol hydrogen bonded motions. For the high frequency (OH-stretch) modes early experiments for PhOH vapor, in apolar solvent ($CCl_4$) and as a liquid reported band positions at ~3650 $cm^{-1}$, ~3600 $cm^{-1}$, and at 3500 $cm^{-1}$, respectively.[86] In $CCl_4$ solution and the pure liquid an additional band is at 3350 $cm^{-1}$.[86] The present work finds a sharp peak at 3596 $cm^{-1}$ which is assigned to the ''free'' phenol–OH stretch in a non-hydrogen bonded environment. This is supported by the observation that the peak is only marginally shifted from the PhOH OH-stretch in vapor and the fact that the spectroscopic feature is sharp and therefore can not be due to water.

More recently, infrared spectra were recorded for PhOH complexed with variable numbers of water molecules in the gas phase,[93,94] in matrices,[95] and for PhOH at the air/water interface using vibrational sum frequency generation (SFG).[87] The cluster studies all report the phenolic-OH stretch vibration at frequencies above 3000 $cm^{-1}$ whereas the experiment at the air/water interface assigns a very broad signature in the SFG signal extending from 2550 $cm^{-1}$ to 3500 $cm^{-1}$ to the OH-stretch mode.[87] This finding is consistent with the present experiments which report a broad absorption for both, F-PhOH and PhOH extending down to ~2700 $cm^{-1}$ which is also assigned to the phenol-OH stretch for water-coordinated -COH, see Fig. 2 and Fig. S2 (ESI†).

Hydration around the –CF group as characterized by 1d- and 2d-solvent distribution functions does not feature any pronounced hydrogen bonding unless a conventional PC model is used which, however, exaggerates the directed interaction along the CF bond. On the other hand, solvent water molecules form intermittent H-bonds with the –OH group of F-PhOH which leads to the broad spectroscopic response below 3000 $cm^{-1}$ observed experimentally and reproduced by several of the computational models. Although the different computational approaches examined are able to capture part of the spectroscopy qualitatively and satisfactorily, none of them performs uniformly well. This is in part due to the coupling between the modes in the 1200 to 1300 $cm^{-1}$ region and part due to deficiencies in the methods themselves (BLYP and MP2, respectively). To further improve the intramolecular energies, transfer learning from the MP2 to the CCSD(T) level of theory can be envisaged, as has been recently done for formic acid dimer and a number of other small molecules.[96–98] If additional refinements to capture charge anisotropy are required with such an improved energy function for the bonded terms, replacing the charges by multipoles or fluctuating minimally distributed charges[99,100] can be considered.

From the perspective of halogenic modifications of organic frameworks used for drug design the present work suggests that the local hydration of –CH and –CF groups is comparable (see Fig. 4 and Fig. S7, ESI†), supporting the notion that ''F behaves like a large H-atom''.[28] It should be pointed out that this conclusion requires advanced force fields (multipoles, fluctuating charges (as in ML/MD or QM(/MM)/MD) and is

not supported by a conventional point charge force field. Although the C–F bond has three isolated electron pairs, it has weaker electrostatic interactions compared to O due to the small size and high electronegativity, which compromises its hydrogen bonding ability and can thus better be described as a weakly polar interaction rather than a hydrogen bond. Considering chlorinated and iodinated species is expected to yield altered hydration structures due to their more pronounced sigma hole. Such insights are important guidelines for rational drug discovery as they provide a basis for directed modification and evolution of ligands with specific interactions in protein binding sites.

## Conflicts of interest

There are no conflicts to declare.

## Acknowledgements

This work was supported by the Swiss National Science Foundation, NCCR MUST (to MM, PH and UR).

## References

1 L. Herrera-Rodriguez, F. Khan, K. Robins and H.-P. Meyer, Perspectives on biotechnological halogenation. Part I: Halogenated products and enzymatic halogenation, *Chim. Oggi*, 2011, **29**, 31–33.
2 M. Z. Hernandes, S. M. T. Cavalcanti, D. R. M. Moreira, J. de Azevedo, W. Filgueira and A. C. L. Leite, Halogen atoms in the modern medicinal chemistry: hints for the drug design, *Curr. Drug Targets*, 2010, **11**, 303–314.
3 H. Matter, M. Nazaré, S. Güssregen, D. Will, H. Schreuder, A. Bauer, M. Urmann, K. Ritter, M. Wagner and V. Wehner, Evidence for C-Cl/C-Br·π Interactions as an Important Contribution to Protein-Ligand Binding Affinity, *Angew. Chem., Int. Ed.*, 2009, **48**, 2911–2916.
4 K. Müller, C. Faeh and F. Diederich, Fluorine in Pharmaceuticals: Looking Beyond Intuition, *Science*, 2007, **317**, 1881–1886.
5 P. Metrangolo, H. Neukirch, T. Pilati and G. Resnati, Halogen Bonding Based Recognition Processes: A World Parallel to Hydrogen Bonding, *Acc. Chem. Res.*, 2005, **38**, 386–395.
6 P. Metrangolo, F. Meyer, T. Pilati, G. Resnati and G. Terraneo, Halogen Bonding in Supramolecular Chemistry, *Angew. Chem., Int. Ed.*, 2008, **47**, 6114–6127.
7 J. P. M. Lommerse, A. J. Stone, R. Taylor and F. H. Allen, The Nature and Geometry of Intermolecular Interactions between Halogens and Oxygen or Nitrogen, *J. Am. Chem. Soc.*, 1996, **118**, 3108–3116.
8 P. Auffinger, F. A. Hays, E. Westhof and P. S. Ho, Halogen bonds in biological molecules, *Proc. Natl. Acad. Sci. U. S. A.*, 2004, **101**, 16789–16794.
9 K. E. Riley and P. Hobza, Strength and Character of Halogen Bonds in Protein-Ligand Complexes, *Cryst. Growth Des.*, 2011, **11**, 4272–4278.
10 L. A. Hardegger, B. Kuhn, B. Spinnler, L. Anselm, R. Ecabert, M. Stihle, B. Gsell, R. Thoma, J. Diez and J. Benz, *et al.*, Systematic Investigation of Halogen Bonding in Protein-Ligand Interactions, *Angew. Chem., Int. Ed.*, 2011, **50**, 314–318.
11 K. Riley, J. Murray, J. Fanfrlík, J. Řezáč, R. Solá, M. Concha, F. Ramos and P. Politzer, Halogen bond tunability I: the effects of aromatic fluorine substitution on the strengths of halogen-bonding interactions involving chlorine, bromine, and iodine, *J. Mol. Model.*, 2011, **17**, 3309–3318.
12 K. El Hage, J.-P. Piquemal, Z. Hobaika, R. G. Maroun and N. Gresh, Could the ''Janus-like'' properties of the halobenzene CX bond (X = Cl, Br) be leveraged to enhance molecular recognition?, *J. Comp. Chem.*, 2015, **36**, 210–221.
13 Y. Lu, T. Shi, Y. Wang, H. Yang, X. Yan, X. Luo, H. Jiang and W. Zhu, Halogen Bonding A Novel Interaction for Rational Drug Design?, *J. Med. Chem.*, 2009, **52**, 2854–2862.
14 R. Wilcken, M. O. Zimmermann, A. Lange, A. C. Joerger and F. M. Boeckler, Principles and Applications of Halogen Bonding in Medicinal Chemistry and Chemical Biology, *J. Med. Chem.*, 2013, **56**, 1363–1388.
15 K. El Hage, V. Pandyarajan, N. B. Phillips, B. J. Smith, J. G. Menting, J. Whittaker, M. C. Lawrence, M. Meuwly and M. A. Weiss, Extending halogen-based medicinal chemistry to proteins, *J. Biol. Chem.*, 2016, **291**, 27023–27041.
16 G. R. Desiraju, P. S. Ho, L. Kloo, A. C. Legon, R. Marquardt, P. Metrangolo, P. Politzer, G. Resnati and K. Rissanen, Definition of the halogen bond (IUPAC Recommendations 2013), *Pure Appl. Chem.*, 2013, **85**, 1711–1713.
17 T. Clark, M. Hennemann, J. S. Murray and P. Politzer, Halogen bonding: the sigma-hole, *J. Mol. Model.*, 2007, **13**, 291–296.
18 H. Wang, W. Wang and W. J. Jin, Sigma-Hole Bond vs pi-Hole Bond: A Comparison Based on Halogen Bond, *Chem. Rev.*, 2016, **116**, 5072–5104.
19 J. B. Neaton, Chemistry A direct look at halogen bonds, *Science*, 2017, **358**, 167–168.
20 S. Kawai, A. Sadeghi, F. Xu, L. Peng, A. Orita, J. Otera, S. Goedecker and E. Meyer, Extended Halogen Bonding between Fully Fluorinated Aromatic Molecules, *ACS Nano*, 2015, **9**, 2574–2583.
21 P. Metrangolo, J. S. Murray, T. Pilati, P. Politzer, G. Resnati and G. Terraneo, The fluorine atom as a halogen bond donor, viz. a positive site, *CrystEngComm*, 2011, **13**, 6593–6596.
22 P. Politzer, J. S. Murray and T. Clark, Halogen bonding and other sigma-hole interactions: a perspective, *Phys. Chem. Chem. Phys.*, 2013, **15**, 11178–11189.
23 A. Legon, Prereactive complexes of dihalogens XY with Lewis bases B in the gas phase: A systematic case for the halogen analogue B-XY of the hydrogen bond B-HX, *Angew. Chem., Int. Ed.*, 1999, **38**, 2687–2714.
24 P. Shah and A. D. Westwell, The role of fluorine in medicinal chemistry, *J. Enzyme Inhib. Med. Chem.*, 2007, **22**, 527–540.

25 T. Barbarich, C. Rithner, S. Miller, O. Anderson and S. Strauss, Significant inter- and intramolecular O-H center dot center dot center dot FC hydrogen bonding, *J. Am. Chem. Soc.*, 1999, **121**, 4280–4281.
26 D. Chopra and T. N. G. Row, Role of organic fluorine in crystal engineering, *CrystEngComm*, 2011, **13**, 2175–2186.
27 M. Inoue, Y. Sumii and N. Shibata, Contribution of organofluorine compounds to pharmaceuticals, *ACS Omega*, 2020, **5**, 10633–10640.
28 R. Hevey, The Role of Fluorine in Glycomimetic Drug Design, *Chem. – Eur. J.*, 2021, **27**, 2240–2253.
29 D. OHagan and H. S. Rzepa, Some influences of fluorine in bioorganic chemistry, *Chem. Commun.*, 1997, 645–652.
30 P. H. Berens and K. R. Wilson, Molecular dynamics and spectra. I. Diatomic rotation and vibration, *J. Chem. Phys.*, 1981, **74**, 4872–4882.
31 S. Bakels, M.-P. Gaigeot and A. M. Rijs, Gas-phase infrared spectroscopy of neutral peptides: Insights from the far-IR and THz domain, *Chem. Rev.*, 2020, **120**, 3233–3260.
32 M. W. Lee, J. K. Carr, M. Goellner, P. Hamm and M. Meuwly, 2D IR spectra of cyanide in water investigated by molecular dynamics simulations, *J. Chem. Phys.*, 2013, **139**, 054506.
33 D. Koner, S. M. Salehi, P. Mondal and M. Meuwly, Non-conventional force fields for applications in spectroscopy and chemical reaction dynamics, *J. Chem. Phys.*, 2020, **153**, 010901.
34 P. Mondal, P.-A. Cazade, A. K. Das, T. Bereau and M. Meuwly, Multipolar force fields for amide-I spectroscopy from conformational dynamics of the alanine trimer, *J. Phys. Chem. B*, 2021, **125**, 10928–10938.
35 B. R. Brooks, C. L. Brooks III, A. D. MacKerell Jr., L. Nilsson, R. J. Petrella, B. Roux, Y. Won, G. Archontis, C. Bartels and S. Boresch, *et al.*, CHARMM: The Biomolecular Simulation Program, *J. Comp. Chem.*, 2009, **30**, 1545–1614.
36 K. Vanommeslaeghe, E. Hatcher, C. Acharya, S. Kundu, S. Zhong, J. Shim, E. Darian, O. Guvench, P. Lopes and I. Vorobyov, *et al.*, CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields, *J. Comp. Chem.*, 2010, **31**, 671–690.
37 C. Kramer, P. Gedeck and M. Meuwly, Atomic Multipoles: Electrostatic Potential Fit, Local Reference Axis Systems and Conformational Dependence, *J. Comp. Chem.*, 2012, **33**, 1673–1688.
38 T. Bereau, C. Kramer and M. Meuwly, Leveraging Symmetries of Static Atomic Multipole Electrostatics in Molecular Dynamics Simulations, *J. Chem. Theory Comput.*, 2013, **9**, 5450–5459.
39 K. El Hage, T. Bereau, S. Jakobsen and M. Meuwly, Impact of Quadrupolar Electrostatics on Atoms Adjacent to the Sigma-Hole in Condensed-Phase Simulations, *J. Chem. Theory Comput.*, 2016, **12**, 3008–3019.
40 K. El Hage, P. K. Gupta, R. Bemish and M. Meuwly, Molecular Mechanisms Underlying Solute Retention at Heterogeneous Interfaces, *J. Phys. Chem. Lett.*, 2017, **8**, 4600–4607.
41 F. Hedin, K. El Hage and M. Meuwly, A Toolkit to Fit Nonbonded Parameters from and for Condensed Phase Simulations, *J. Chem. Theory Comput.*, 2016, **56**, 1479–1489.
42 W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Klein, Comparison of Simple Potential Functions for Simulating Liquid Water, *J. Chem. Phys.*, 1983, **79**, 926–935.
43 W. C. Swope, H. C. Andersen, P. H. Berens and K. R. Wilson, A Computer Simulation Method for the Calculation of Equilibrium Constants for the Formation of Physical Clusters of Molecules: Application to Small Water Clusters, *J. Chem. Phys.*, 1982, **76**, 637–649.
44 P. J. Steinbach and B. R. Brooks, New Spherical-Cutoff Methods for Long-Range Forces in Macromolecular Simulation, *J. Comp. Chem.*, 1994, **15**, 667–683.
45 T. Darden, D. York and L. Pedersen, Particle Mesh Ewald: An Nlog(N) Method for Ewald Sums in Large Systems, *J. Chem. Phys.*, 1993, **98**, 10089–10092.
46 J.-P. Ryckaert, G. Ciccotti and H. J. C. Berendsen, Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes, *J. Comp. Phys.*, 1977, **23**, 327–341.
47 W. V. Gunsteren and H. Berendsen, Algorithms for Macromolecular Dynamics and Constraint Dynamics, *Mol. Phys.*, 1997, **34**, 1311–1327.
48 A. H. Larsen, J. J. Mortensen, J. Blomqvist, I. E. Castelli, R. Christensen, M. Dułpollak, J. Friis, M. N. Groves, B. Hammer and C. Hargus, *et al.*, The atomic simulation environment - a Python library for working with atoms, *J. Phys.: Condens. Matter*, 2017, **29**, 273002.
49 E. Spohr, Effect of electrostatic boundary conditions and system size on the interfacial properties of water and aqueous solutions, *J. Chem. Phys.*, 1997, **107**, 6342–6348.
50 R. J. Loncharich, B. R. Brooks and R. W. Pastor, Langevin dynamics of peptides: The frictional dependence of isomerization rates of N-acetylalanyl-N-methylamide, *Biopolymers*, 1992, **32**, 523–535.
51 M. Thomas, M. Brehm, R. Fligg, P. Vöhringer and B. Kirchner, Computing vibrational spectra from ab initio molecular dynamics, *Phys. Chem. Chem. Phys.*, 2013, **15**, 6608–6622.
52 M. Schmitz and P. Tavan, Vibrational spectra from atomic fluctuations in dynamics simulations. I. Theory, limitations, and a sample application, *J. Chem. Phys.*, 2004, **121**, 12233–12246.
53 M. Schmitz and P. Tavan, Vibrational spectra from atomic fluctuations in dynamics simulations. II. Solvent-induced frequency fluctuations at femtosecond time resolution, *J. Chem. Phys.*, 2004, **121**, 12247–12258.
54 P. Hamm and M. Zanni, *Concepts and Methods of 2D Infrared Spectroscopy*, Cambridge University Press, New York, 2011.
55 P.-A. Cazade, H. Tran, T. Bereau, A. K. Das, F. Klaesi, P. Hamm and M. Meuwly, Solvation of fluoro-acetonitrile

in water by 2D-IR spectroscopy: A combined experimental-computational study, *J. Chem. Phys.*, 2015, **142**, 212415.
56 K. Moller, R. Rey and J. Hynes, Hydrogen Bond Dynamics in Water and Ultrafast Infrared Spectroscopy: A Theoretical Study, *J. Phys. Chem. A*, 2004, **108**, 1275–1289.
57 M. Salehi, D. Koner and M. Meuwly, Vibrational Spectroscopy of $N_3^-$ in the Gas- and Condensed-Phase, *J. Phys. Chem. B*, 2019, **123**, 3282–3290.
58 P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser and J. Bright, *et al.*, SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python, *Nat. Methods*, 2020, **17**, 261–272.
59 R. Car and M. Parrinello, Unified Approach for Molecular Dynamics and Density-Functional Theory, *Phys. Rev. Lett.*, 1985, **55**, 2471–2474.
60 CPMD, Copyright IBM Corp 1990-2019, Copyright MPI für Festkörperforschung Stuttgart 1997-2001. **https://www.cpmd.org/**, **https://www.cpmd.org/**.
61 A. D. Becke, Density-functional exchange-energy approximation with correct asymptotic behavior, *Phys. Rev. A*, 1988, **38**, 3098.
62 C. Lee, W. Yang and R. G. Parr, Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 1988, **37**, 785–789.
63 O. A. von Lilienfeld, I. Tavernelli, U. Rothlisberger and D. Sebastiani, Optimization of Effective Atom Centered Potentials for London Dispersion Forces in Density Functional Theory, *Phys. Rev. Lett.*, 2004, **93**, 153004.
64 I.-C. Lin, M. D. Coutinho-Neto, C. Felsenheimer, O. A. von Lilienfeld, I. Tavernelli and U. Rothlisberger, Library of dispersion-corrected atom-centered potentials for generalized gradient approximation functionals: Elements H, C, N, O, He, Ne, Ar, and Kr, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 2007, **75**, 205131.
65 M. Doemer, I. Tavernelli and U. Rothlisberger, Intricacies of Describing Weak Interactions Involving Halogen Atoms within Density Functional Theory, *J. Chem. Theory Comput.*, 2013, **9**, 955–964, PMID: 26588739.
66 N. Troullier and J. L. Martins, Efficient pseudopotentials for plane-wave calculations, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 1991, **43**, 1993.
67 W. Zierkiewicz and D. Michalska, Molecular Structure and Infrared Spectra of 4-Fluorophenol: A Combined Theoretical and Spectroscopic Study, *J. Phys. Chem.*, 2003, **107**, 4547–4554.
68 W. F. van Gunsteren; coworkers, Biomolecular Simulation: The GROMOS96 Manual and User Guide; vdf Hochschulverlag AG an der ETH Zürich and BIOMOS b.v. Groningen: Zürich, Switzerland and The Netherlands, 2006. **https://www.gromos.net/**, **http://www.gromos.net/**.
69 A. Laio, J. VandeVondele and U. Rothlisberger, A Hamiltonian electrostatic coupling scheme for hybrid Car-Parrinello molecular dynamics simulations, *J. Chem. Phys.*, 2002, **116**, 6941–6947.
70 M. C. Colombo, L. Guidoni, A. Laio, A. Magistrato, P. Maurer, S. Piana, U. Röhrig, K. Spiegel, M. Sulpizi and J. VandeVondele, *et al.*, Hybrid QM/MM Car-Parrinello Simulations of Catalytic and Enzymatic Reactions, *CHIMIA International Journal for Chemistry*, 2002, **56**, 13–19.
71 E. Brunk and U. Rothlisberger, Mixed Quantum Mechanical/Molecular Mechanical Molecular Dynamics Simulations of Biological Systems in Ground and Electronically Excited States, *Chem. Rev.*, 2015, **115**, 6217–6263.
72 D. A. Case, I. Y. Ben-Shalom, S. R. Brozell, D. S. Cerutti, T. E. Cheatham III, V. W. D. Cruzeiro, T. A. Darden, R. E. Duke, D. Ghoreishi, M. K. Gilson *et al.*, AMBER 2018, University of California, San Francisco. **https://ambermd.org/**, **https://ambermd.org/**.
73 J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman and D. A. Case, Development and testing of a general amber force field, *J. Comput. Chem.*, 2004, **25**, 1157–1174.
74 GAFF2 is a public domain forcefield, an upgrade of the previously released general AMBER forcefield (GAFF). It is available with the distribution of AmberTools17 and can be downloaded from **https://ambermd.org.** A publication for it is currently under preparation. **https://ambermd.org/**, **https://ambermd.org/**.
75 H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola and J. R. Haak, Molecular dynamics with coupling to an external bath, *J. Chem. Phys.*, 1984, **81**, 3684–3690.
76 O. T. Unke and M. Meuwly, PhysNet: A neural network for predicting energies, forces, dipole moments, and partial charges, *J. Chem. Theory Comput.*, 2019, **15**, 3678–3693.
77 J. Gilmer; S. S. Schoenholz; P. F. Riley; O. Vinyals and G. E. Dahl, Neural message passing for quantum chemistry. Proc. of the 34th Int. Conf. on Machine Learning, vol. 70, 2017, pp. 1263–1272.
78 H.-J. Werner, P. J. Knowles, G. Knizia, F. R. Manby, M. Schütz, P. Celani, W. Györffy, D. Kats, T. Korona, R. Lindh *et al.*, MOLPRO, version 2019, a package of ab initio programs. 2019.
79 J. S. Smith, O. Isayev and A. E. Roitberg, ANI-1, A data set of 20 million calculated off-equilibrium conformations for organic molecules, *Sci. Data*, 2017, **4**, 170193.
80 D. Michalska, W. Zierkiewicz, D. C. Bienko, W. Wojciechowski and T. Zeegers-Huyskens, ''Troublesome'' Vibrations of Aromatic Molecules in Second-Order Moller-Plesset and Density Functional Theory Calculations: Infrared Spectra of Phenol and Phenol-OD, *J. Phys. Chem.*, 2001, **105**, 8734–8739.
81 H. Bist, J. C. Brand and D. Williams, The vibrational spectrum and torsion of phenol, *J. Mol. Spectrosc.*, 1967, **24**, 402–412.
82 B. Bauer and M. Reynolds, Recovering data from scanned graphs: Performance of Frantz's g3data software, *Behav. Res. Meth.*, 2008, **40**, 858–868.
83 Y. Morino and K. Kuchitsu, A Note on the Classification of Normal Vibrations of Molecules, *J. Chem. Phys.*, 1952, **20**, 1809–1810.

84 A. P. Scott and L. Radom, Harmonic Vibrational Frequencies: An Evaluation of HartreeFock, MllerPlesset, Quadratic Configuration Interaction, Density Functional Theory, and Semiempirical Scale Factors, *J. Phys. Chem.*, 1996, **100**, 10502–16513.

85 K. Mizuse, T. Hamashima and A. Fujii, Infrared spectroscopy of phenol-$(H_2O)$ $n > 10$: structural strains in hydrogen bond networks of neutral water clusters, *J. Phys. Chem. A*, 2009, **113**, 12134–12141.

86 J. Evans, The vibrational spectra of phenol and phenol-OD, *Spectrochim. Act.*, 1960, **16**, 1382–1392.

87 R. Kusaka, T. Ishiyama, S. Nihonyanagi, A. Morita and T. Tahara, Structure at the air/water interface in the presence of phenol: a study using heterodyne-detected vibrational sum frequency generation and molecular dynamics simulation, *Phys. Chem. Chem. Phys.*, 2018, **20**, 3002–3009.

88 E. T. Nibbering, J. Dreyer, O. Kühn, J. Bredenbeck, P. Hamm and T. Elsaesser, *Analysis and control of ultrafast photoinduced reactions*, Springer, 2007, pp. 619–687.

89 C. J. Burnham, J. C. Li and M. Leslie, Molecular Dynamics Calculations for Ice Ih, *J. Phys. Chem. B*, 1997, **101**, 6192–6195.

90 N. Plattner and M. Meuwly, Atomistic Simulations of CO Vibrations in Ices Relevant to Astrochemistry, *Chem. Phys. Chem.*, 2008, **9**, 1271–1277.

91 N. Kumagai, K. Kawamura and T. Yokokawa, An Interatomic Potential Model for $H_2O$: Applications to Water and Ice Polymorphs, *Mol. Sim.*, 1994, **12**, 177–186.

92 RStudio Team, RStudio: Integrated Development Environment for R. RStudio, PBC.: Boston, MA, 2020.

93 T. Hamashima, K. Mizuse and A. Fujii, Spectral Signatures of Four-Coordinated Sites in Water Clusters: Infrared Spectroscopy of Phenol-$(H_2O)_n$ $(20 \leq n \leq 50)$, *J. Phys. Chem. A*, 2011, **115**, 620–625.

94 T. Shimamori and A. Fujii, Infrared spectroscopy of warm and neutral phenol-water clusters, *J. Phys. Chem. A*, 2015, **119**, 1315–1322.

95 P. Banerjee, I. Bhattacharya and T. Chakraborty, Cooperative effect on phenolic $\nu$ O-H frequencies in 1: 1 hydrogen bonded complexes of o-fluorophenols with water: A matrix isolation infrared spectroscopic study, *Spectrochim. Act.*, 2017, **181**, 116–121.

96 S. Kaäser, E. D. Boittier, M. Upadhyay and M. Meuwly, Transfer Learning to CCSD (T): Accurate Anharmonic Frequencies from Machine Learning Models, *J. Chem. Theory Comput.*, 2021, **17**, 3687–3699.

97 S. Käser and M. Meuwly, Transfer learned potential energy surfaces: accurate anharmonic vibrational dynamics and dissociation energies for the formic acid monomer and dimer, *Phys. Chem. Chem. Phys.*, 2022, **24**, 5269–5281.

98 K. Töpfer, S. Käser and M. Meuwly, Double Proton Transfer in Hydrated Formic Acid Dimer: Interplay of Spatial Symmetry and Solvent-Generated Force on Reactivity, *Phys. Chem. Chem. Phys.*, 2022, **24**, 13869–13882.

99 O. T. Unke, M. Devereux and M. Meuwly, Minimal distributed charges: Multipolar quality at the cost of point charge electrostatics, *J. Chem. Phys.*, 2017, **147**, 161712.

100 E. D. Boittier, M. Devereux and M. Meuwly, Molecular Dynamics with Conformationally Dependent, Distributed Charges, arXiv, 2022, preprint, arXiv:2206.15366, DOI: **10.48550/arXiv.2206.15366**.