Supplementary Information – HyDRA challenge

Giacomo Mandelli [ORCID: 0000-0001-9861-7612][1,*], Cecilia Lanzi [ORCID: 0000-0001-7633-9658][1,*], Riccardo Conte [ORCID: 0000-0003-3026-3875, email: riccardo.conte1@unimi.it][1,**], and Michele Ceotto [ORCID: 0000-0002-8270-3409, email: michele.ceotto@unimi.it][1,**]

[1] Dipartimento di Chimica, Università degli Studi di Milano, via C. Golgi 19, 20133 Milano, Italy
[*] Performed the calculations
[**] Supervised the work

1 Computational Details

1.1 Overview of the Theory

The submitted frequency values for the symmetric stretching of the water molecule in the complexes are obtained through quasi-classical trajectory simulations. In our quasi-classical trajectory method (QCT), the power spectrum is obtained by Fourier transform of the velocity autocorrelation function along a single trajectory [1, 2]. To derive the QCT frequency value for a certain normal mode, the formulation used is the following [3]:

$$I_j(\omega) = \frac{1}{2T}\left|\int_0^T dt e^{i\omega t}\mathbf{p}_j(t)\right|^2 \tag{1}$$

where T is the trajectory simulation time, $\mathbf{p}_j(t)$ is the momentum along the j-th normal mode at time t and ω is the frequency. It is important to note that equation 1 is the result of a time average filtering procedure [1].

In our quasiclassical on-the-fly simulations a single classical trajectory is run on the ab initio Potential Energy Surface (PES) with the following harmonic initial conditions:

$$\begin{cases} \mathbf{q}_0 & = \mathbf{q}_{eq} \\ \mathbf{p}_0^j & = \sqrt{(2n_j+1)\omega_j} \end{cases} \tag{2}$$

where the initial positions $\mathbf{q}_0$ for the system are the coordinates of the system in its minimum geometry on the PES, and the initial momenta $\mathbf{p}_0^j$, for each degree of freedom j, are taken as a harmonic guess where ω_j is the harmonic frequency and n_j is the vibrational quantum number.

We chose this method for the Hydra challenge because of its simplicity and effectiveness in reproducing the experimental values in the training set. Indeed, as the QCT spectra are classical spectra on ab initio potential, they are very simple to obtain, and given the performances, we were able to compute all the spectra of all the global and local minima for the training and test set molecules.

The QCT method samples efficiently the anharmonicities of the system and may reveal possible classical beatings among modes of vibrations, but it cannot show quantum mechanical effects such as anharmonic overtones, interferences and Zero Point Energy (ZPE).

1.2 Computational Setup and Workflow

To be consistent in the calculations of the training and test sets, all the parameters, software, and methods used are the same for all the molecules.
The software used for all the ab initio calculations is NWChem in its 7.2.1 release [4].
For each step involving ab initio calculations we employed the following computational setup:

Level of theory: Density Functional Theory (DFT) with Grimme's D3 dispersion corrections [5]

Functional: B3LYP exchange-correlation functional

Basis set: Dunning correlation consistent basis set aug-cc-pVDZ

The first step towards the QCT spectrum is the **geometry optimization** to get the global or local minimum geometry of the water complex. We considered and submitted the QCT values for local minima up to $5\,\mathrm{kJmol^{-1}}$ above the global minimum energy. We used the standard geometry optimization parameters provided by the Driver module in NWChem.
After the geometry optimization, the **Hessian** at the equilibrium geometry is calculated.
Once the Hessian is obtained, the **initial velocities** are computed.
The initial velocities are then given as input to the NWChem **adiabatic ab initio molecular dynamics** module (QMD). With QMD we ran 2500 steps simulations with a time step of $10\,\mathrm{au}$ and no thermostat. The relevant QMD module parameters to be set are the following: nstep_nucl=2500, dt_nucl=10.0, thermostat=none, print_xyz=1. From the QMD output file, containing the velocities and the geometries at each step, the **power spectrum** is obtained using the procedure described in section 1.1 above. Vibrational frequencies are obtained from the peaks of the power spectra. For a straightforward peak assignment it is possible to visualise the power spectrum of each normal mode independently.
For the QCT on a single trajectory we do not have an **error estimate** definition, and talking about uncertainty of the method is not meaningful. Anyway, we decided to provide an uncertainty value for the QCT frequencies of the water symmetric stretch based upon the QCT results on the training set molecules. Indeed, we chose an **uncertainty** of $\pm 16\,\mathrm{cm^{-1}}$, slightly higher than the Mean Absolute Error (MAE), such that the 70% of the QCT training set results agree with the experimental ones.

2 Additional Computed Data

2.1 QCT Power Spectra of the Training and Test Sets

All the power spectra here shown refer to the conformations submitted as global minima for the training and test sets molecules.

2.1.1 Training Set

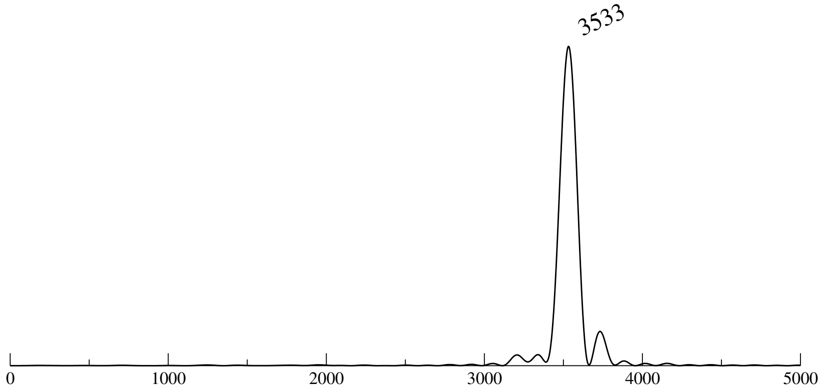

Figure 1: Acetone monohydrate QCT spectrum for the water symmetric stretch (mode 38), harmonic estimate $3606.38\,\mathrm{cm}^{-1}$

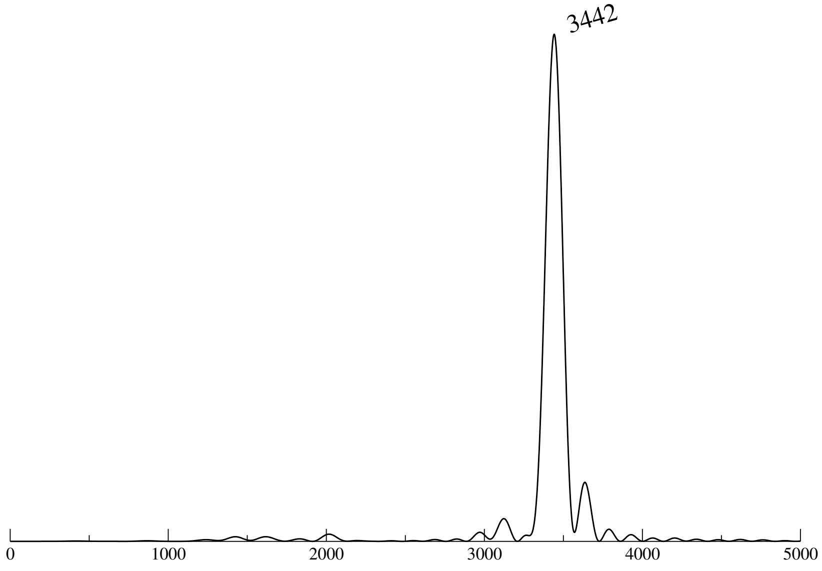

Figure 2: Imidazole monohydrate QCT spectrum for the water symmetric stretch (mode 28), harmonic estimate $3496.05\,\mathrm{cm}^{-1}$

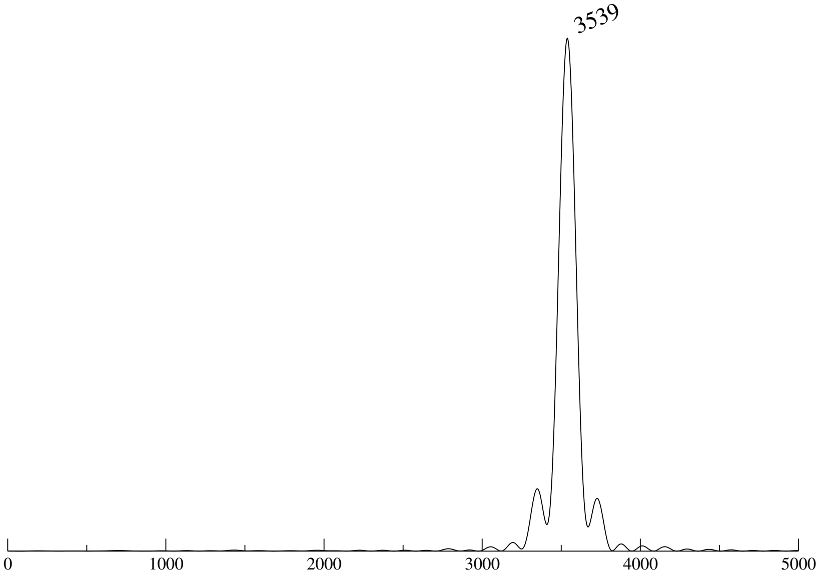

Figure 3: Cyclobutanone monohydrate QCT spectrum for the water symmetric stretch (mode 35), harmonic estimate $3630.2\,\mathrm{cm}^{-1}$

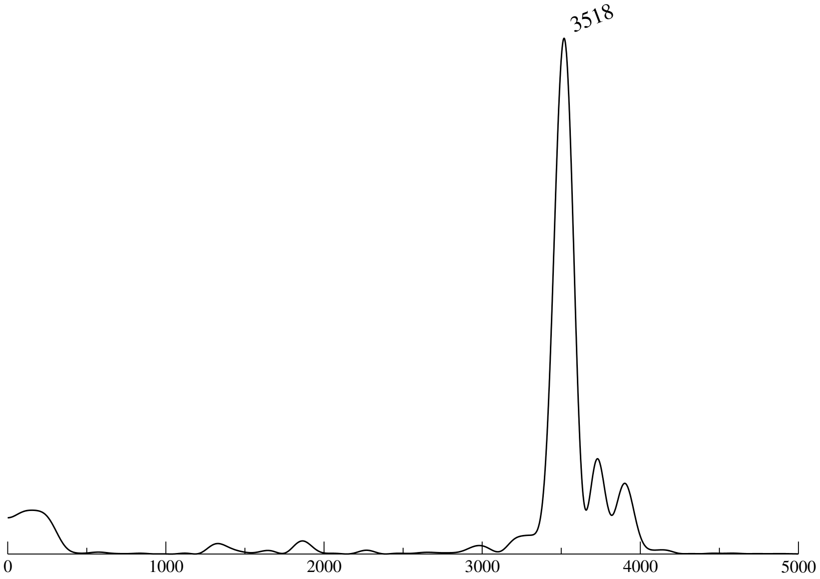

Figure 4: Aniline monohydrate QCT spectrum for the water symmetric stretch (mode 43), harmonic estimate $3583.38\,\mathrm{cm}^{-1}$

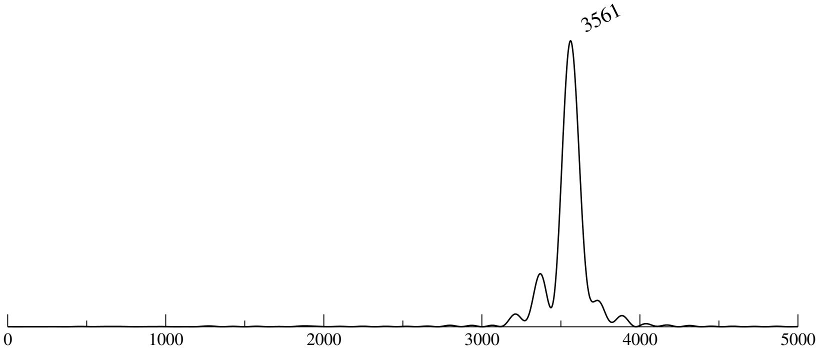

Figure 5: *o*-Cyanophenol monohydrate QCT spectrum for the water symmetric stretch (mode 44), harmonic estimate $3700.99\,\mathrm{cm}^{-1}$

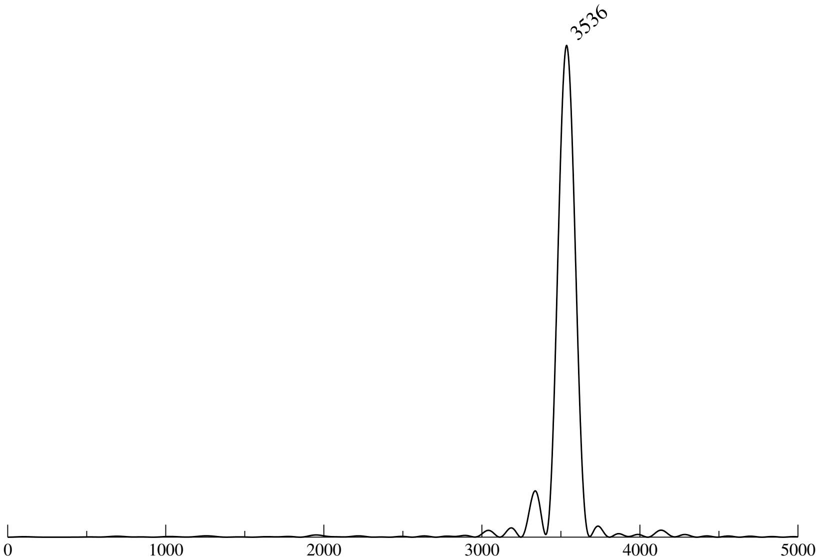

Figure 6: Acetophenone monohydrate isomer I QCT spectrum for the water symmetric stretch (mode 53), harmonic estimate $3602.22\,\mathrm{cm}^{-1}$

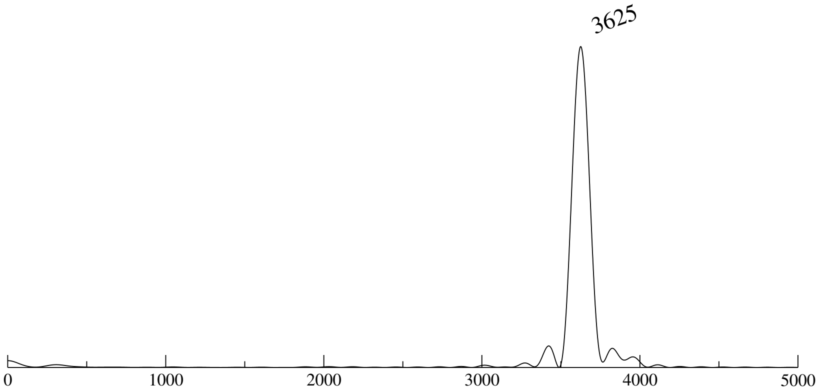

Figure 7: 1-Phenylethanol monohydrate QCT spectrum for the water symmetric stretch (mode 59), harmonic estimate $3773.18\,\mathrm{cm}^{-1}$

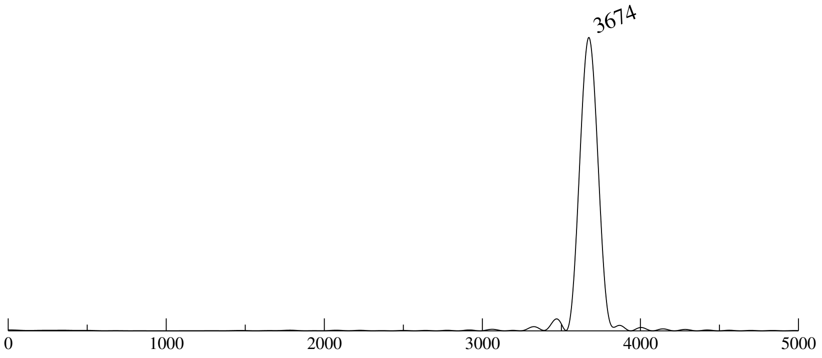

Figure 8: 1,2,4,5-Tetrafluorobenzene monohydrate QCT spectrum for the water symmetric stretch (mode 38), harmonic estimate $3788.61\,\mathrm{cm}^{-1}$

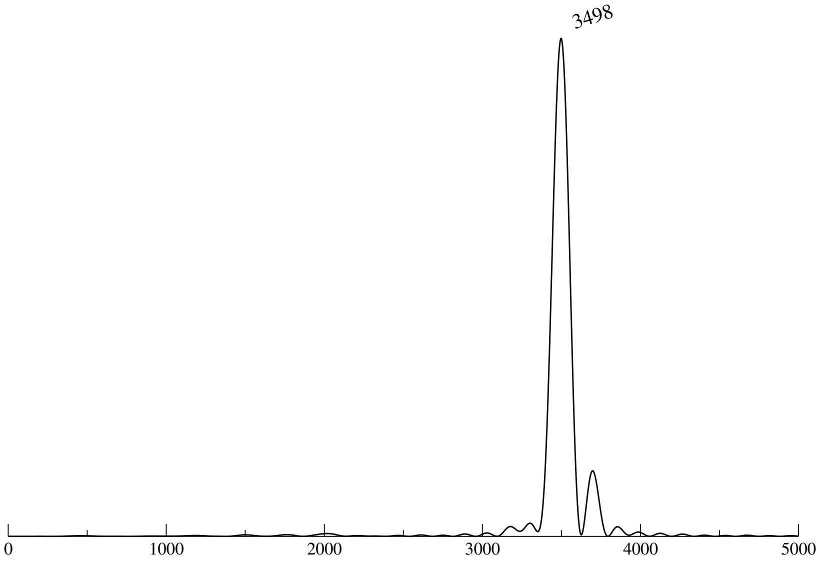

Figure 9: Di-tert-butyl nitroxide monohydrate QCT spectrum for the water symmetric stretch (mode 86), harmonic estimate $3598.74\,\mathrm{cm}^{-1}$

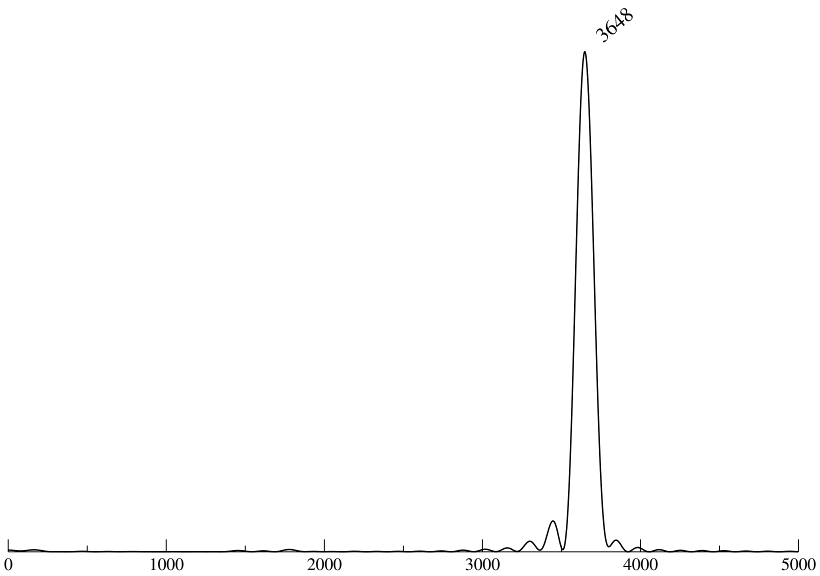

Figure 10: Dibenzofuran monohydrate QCT spectrum for the water symmetric stretch (mode 65), harmonic estimate $3741.20\,\mathrm{cm}^{-1}$

2.1.2 Test Set

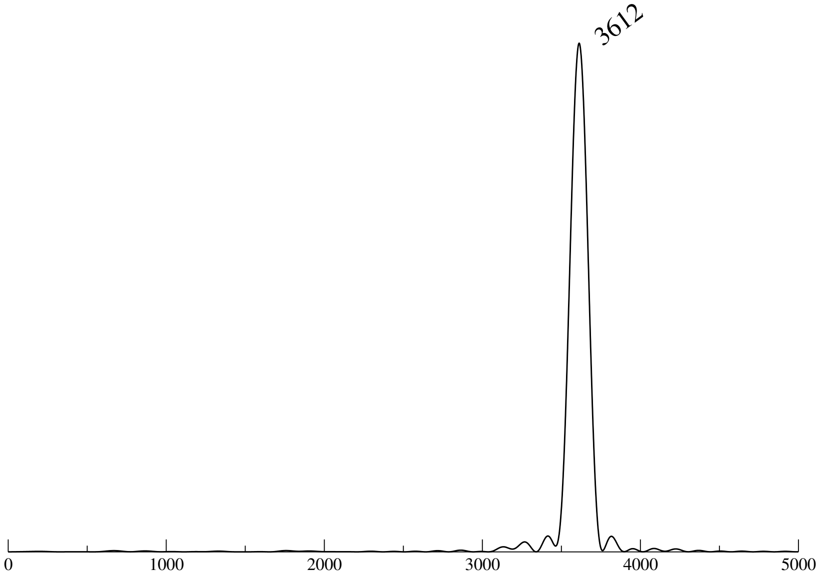

Figure 11: Formaldehyde monohydrate QCT spectrum for the water symmetric stretch (mode 14), harmonic estimate $3683.82\,\mathrm{cm}^{-1}$

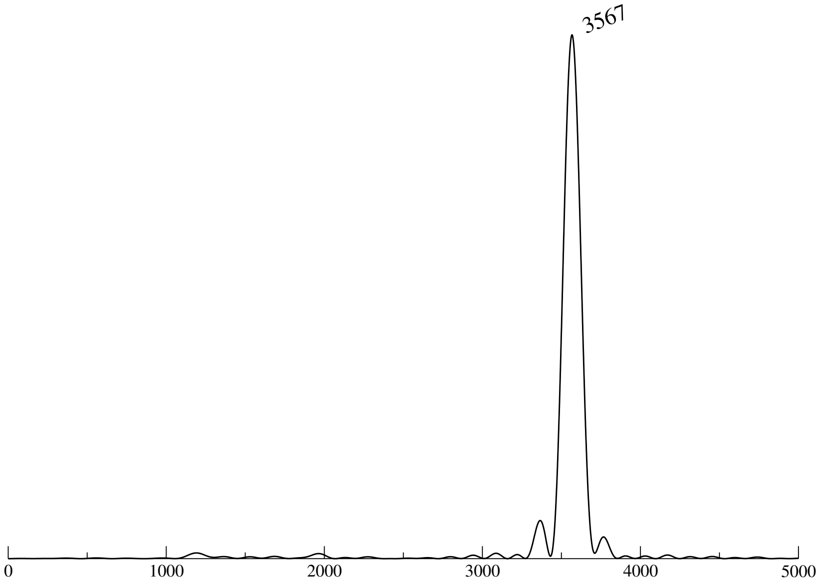

Figure 12: Tetrahydrofuran monohydrate QCT spectrum for the water symmetric stretch (mode 41), harmonic estimate $3574.76\,\mathrm{cm}^{-1}$

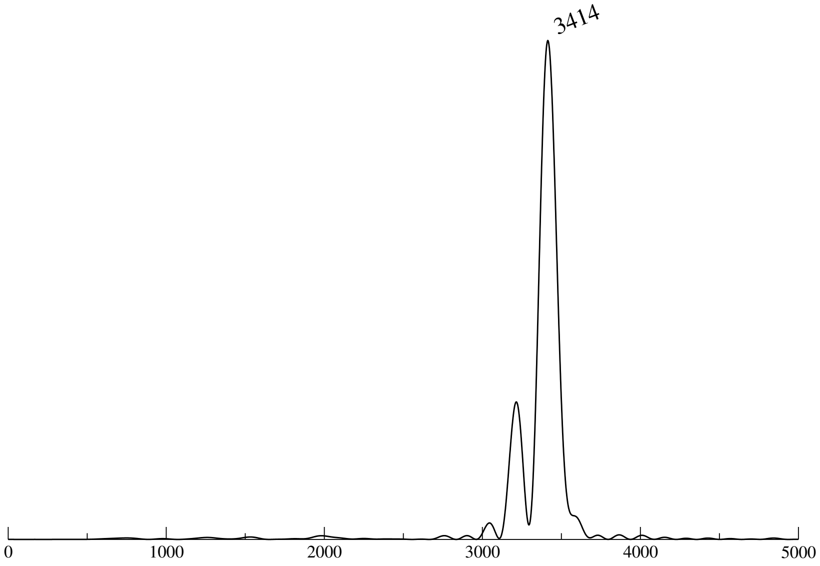

Figure 13: Pyridine monohydrate QCT spectrum for the water symmetric stretch (mode 35), harmonic estimate $3481.27\,\mathrm{cm}^{-1}$

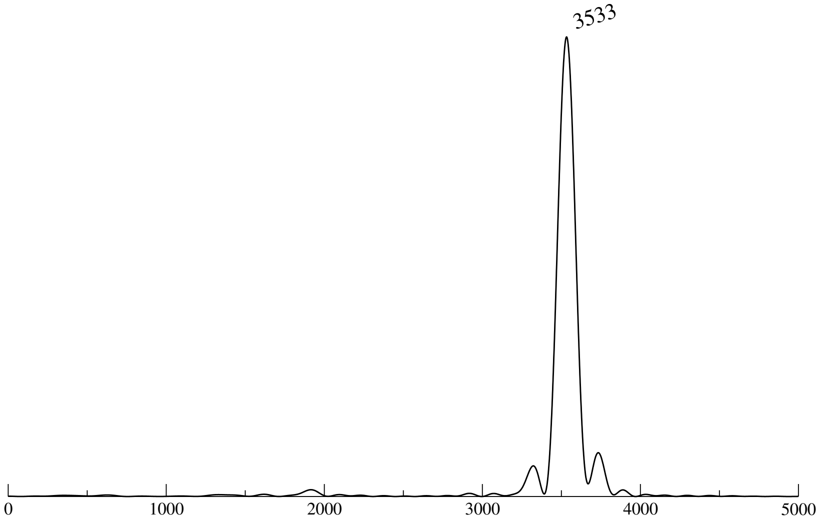

Figure 14: Tetrahydrothiophene monohydrate QCT spectrum for the water symmetric stretch (mode 41), harmonic estimate $3583.10\,\mathrm{cm}^{-1}$

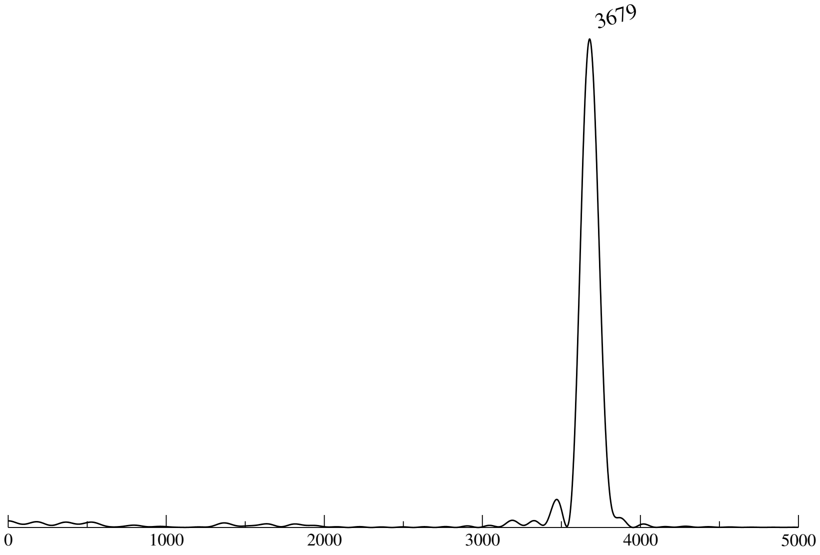

Figure 15: QCT spectrum of the IG conformer of the 2,2,2-trifluoroethan-1-ol monohydrate for the water symmetric stretch (mode 29), harmonic estimate $3786.58\,\mathrm{cm}^{-1}$

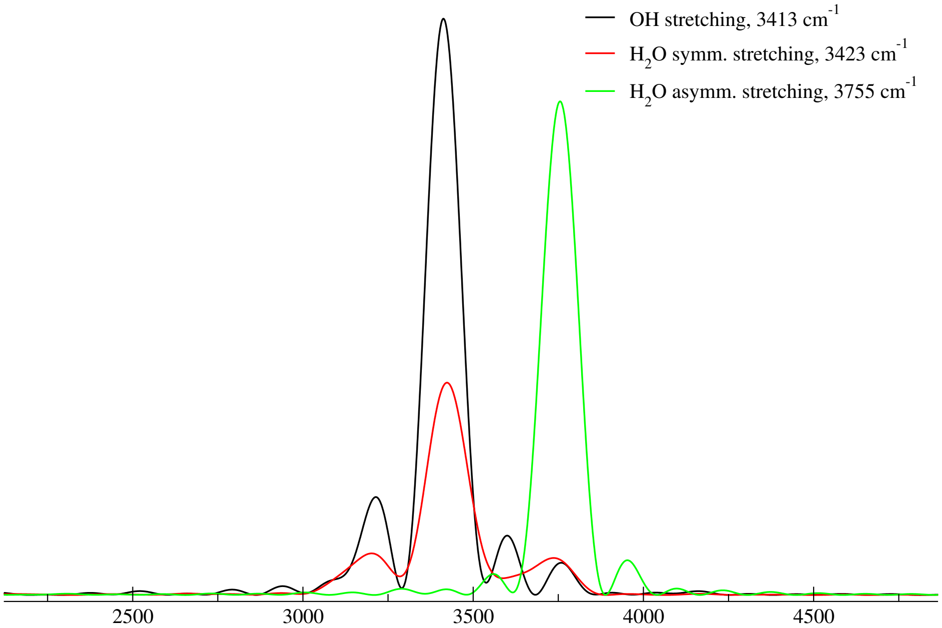

Figure 16: QCT spectrum of the i-I conformer of the methyl lactate monohydrate. OH stretch harmonic estimate $3545.70\,\mathrm{cm}^{-1}$, water symmetric stretch harmonic estimate $3601.74\,\mathrm{cm}^{-1}$, water asymmetric stretch harmonic estimate $3869.36\,\mathrm{cm}^{-1}$

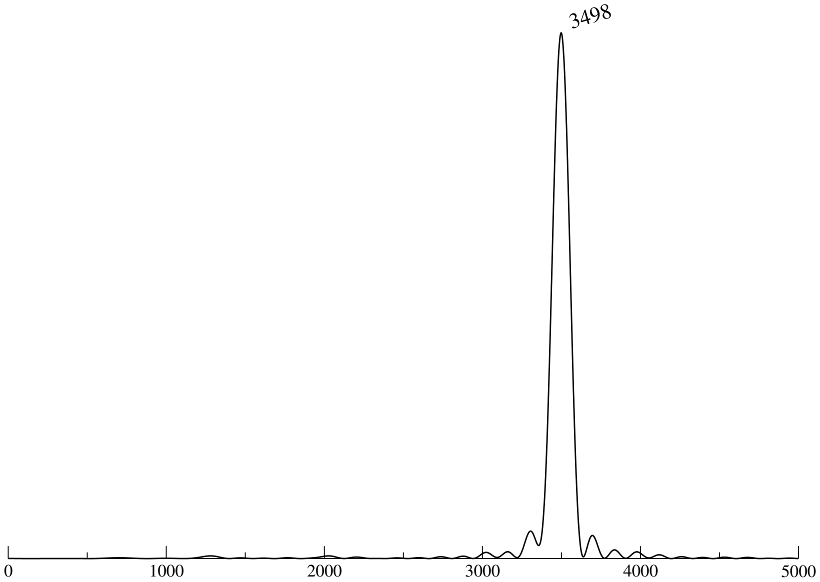

Figure 17: 1,3-dimethyl-2-imidazolidinone monohydrate QCT spectrum for the water symmetric stretch (mode 56), harmonic estimate $3569.42\,\mathrm{cm}^{-1}$

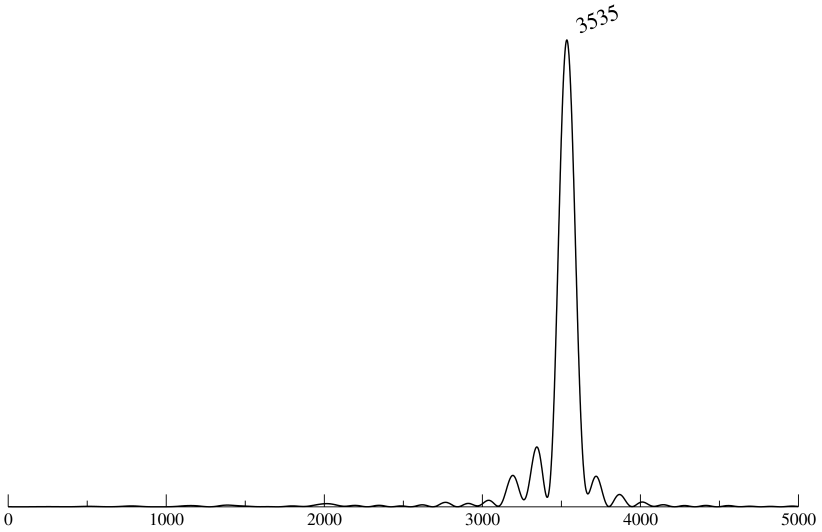

Figure 18: QCT spectrum of the BC1b conformer of the cyclooctanone monohydrate for the water symmetric stretch (mode 71), harmonic estimate $3568.93\,\mathrm{cm}^{-1}$

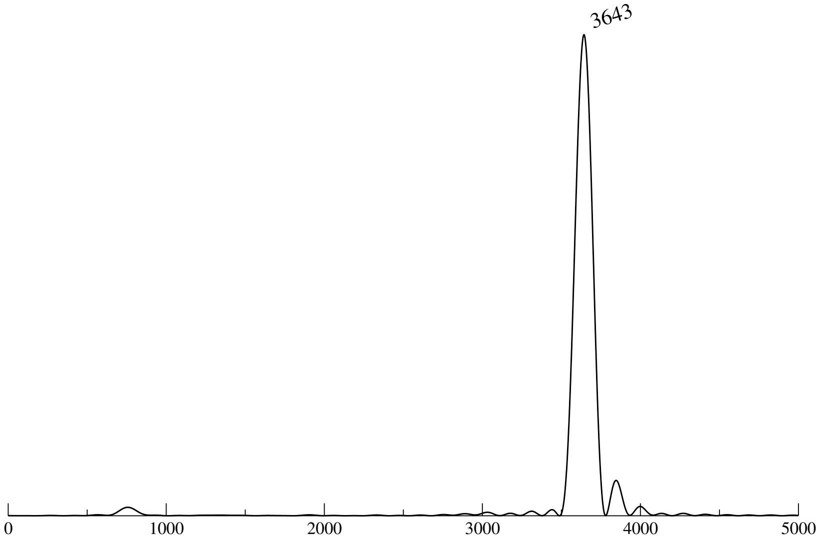

Figure 19: QCT spectrum of the TPH I conformer of the 2,2,2-trifluoroacetophenone monohydrate for the water symmetric stretch (mode 53), harmonic estimate $3707.10\,\mathrm{cm}^{-1}$

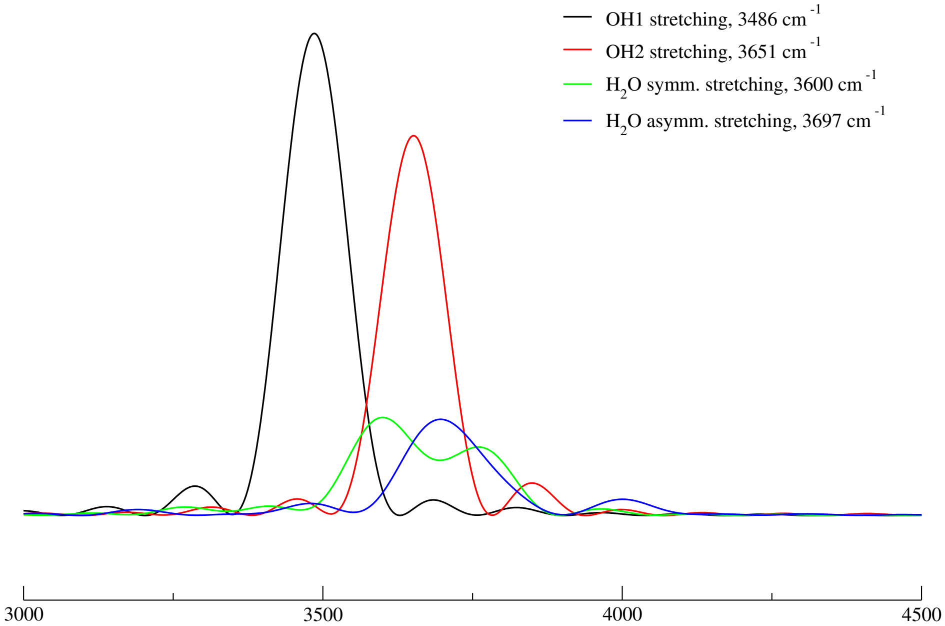

Figure 20: QCT spectrum of the PCD1 conformer of the 1-phenylcyclohexane-cis-1,2-diol monohydrate. OH_1 harmonic estimate $3590.51\,\mathrm{cm}^{-1}$, OH_2 harmonic estimate $3717.05\,\mathrm{cm}^{-1}$, water symmetric stretch harmonic estimate $3749.65\,\mathrm{cm}^{-1}$, water asymmetric OH stretch harmonic estimate $3874.93\,\mathrm{cm}^{-1}$

2.2 OH-stretching Fundamentals for the Test Set

Code	CAS	fundamentals (cm^{-1})	description
CON	502-49-8	3757	water asymm. stretch
DMI	80-73-9	3781	water asymm. stretch
FAH	50-0-0	3768	water asymm. stretch
MLA	547-64-8	3755	water asymm. stretch
		3413	OH stretch
PCD	125132-75-4	3697	water asymm. stretch
		3486	OH_1 stretch *(a)*
		3651	OH_2 stretch *(a)*
PYR	110-86-1	3781	water asymm. stretch
THF	109-99-9	3770	water asymm. stretch
THT	110-01-0	3741	water asymm. stretch
TPH	434-45-7	3784	water asymm. stretch
TFE	75-89-8	3756	water asymm. stretch
		3622	OH stretch

Table 1: QCT values for the water asymmetric stretching frequencies and other OH stretchings in the monohydrates. All the values refer to the global minimum geometry. The local minima spectra can be found in section 2.4.2. *(a)* OH_1 is the intramolecular H-bond acceptor alcohol moiety, while OH_2 is the intramolecular H-bond donor alcohol moiety.

2.3 Isotopologues information

When necessary, we carried out isotope substitutions to find the normal modes of vibration with the strongest water symmetric stretching character. The masses used for the isotope substitutions are the following:

O^{18}: $17.9991\,\mathrm{amu}$

D: $2.0141\,\mathrm{amu}$

For the PCD2 monohydrate and the TFE monohydrate we only used harmonic estimates to carry out such analysis (see tables 2 and 3), while for the MLA monohydrate we computed the QCT spectra shown in figures 21, 22 and 23.
We also computed the QCT spectra for the local minimum isotopologue of the MLA monohydrate (i-III conformer, see paragraph 2.4.2.1), see figure 24.

Code	harmonic freq. (cm^{-1})	OD$_a$ *(a)* isotopologue harm. freq. (cm^{-1})	description
PCD2	3866.44	3866.44	water asymm. stretch
	3765.34	3765.34	OH$_b$ stretch *(a)*
	3535.34	2615.42	OH$_a$ stretch *(a)*
	3604.92	3551.49	water symm. stretch

Table 2: Harmonic frequencies of the PCD2 conformer of the 1-phenylcyclohexane-cis-1,2-diol monohydrate (see paragraph 2.4.2.4). *(a)* OH$_a$ is the H bond donor alcohol moiety, while OH$_b$ is the H bond acceptor alcohol moiety

Code	harmonic freq. (cm^{-1})	OD isotopologue harm. freq. (cm^{-1})	description
TFE	3896.30	3896.30	water asymm. stretch
	3786.58	3785.88	water symm. stretch
	3604.40	2626.30	OH stretch

Table 3: Harmonic frequencies of the 2,2,2-trifluoroethan-1-ol monohydrate

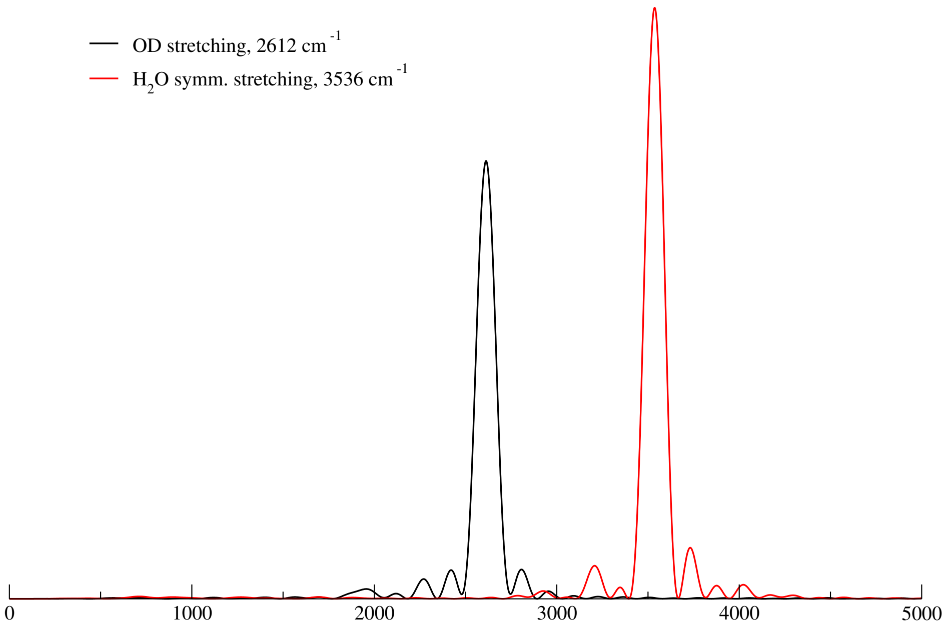

Figure 21: QCT spectrum of the i-I conformer of the MLA monohydrate with the deuterium isotope substitution on the OH moiety. OD stretch harmonic estimate $2605.49\,\mathrm{cm}^{-1}$, water symmetric stretch harmonic estimate $3571.19\,\mathrm{cm}^{-1}$, water asymmetric stretch harmonic estimate $3869.32\,\mathrm{cm}^{-1}$

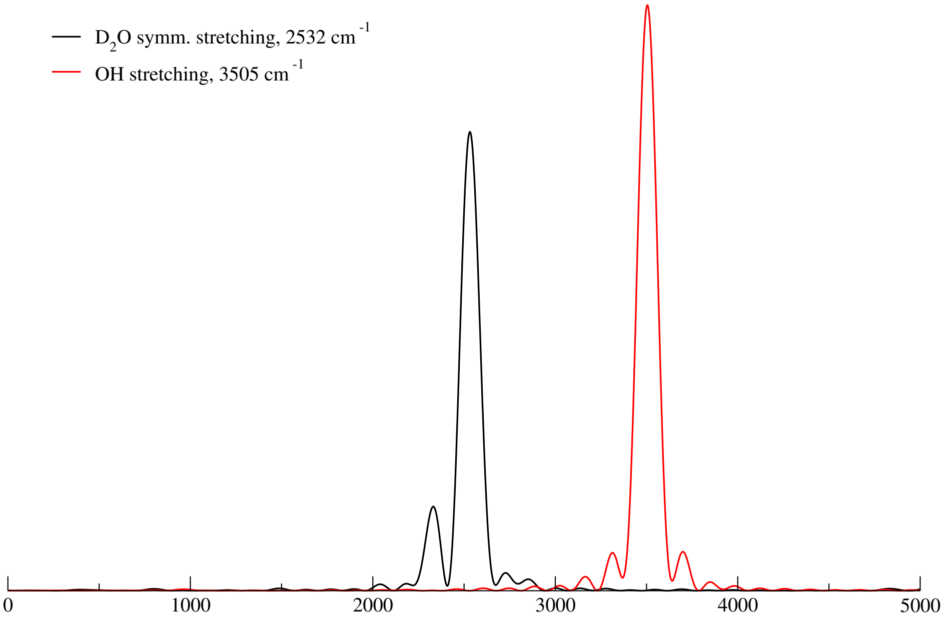

Figure 22: QCT spectrum of the i-I conformer of the MLA monohydrate with the deuterium isotope substitution on the water molecule. OH stretch harmonic estimate $3577.40\,\mathrm{cm}^{-1}$, water symmetric stretch harmonic estimate $2588.66\,\mathrm{cm}^{-1}$, water asymmetric stretch harmonic estimate $2822.83\,\mathrm{cm}^{-1}$

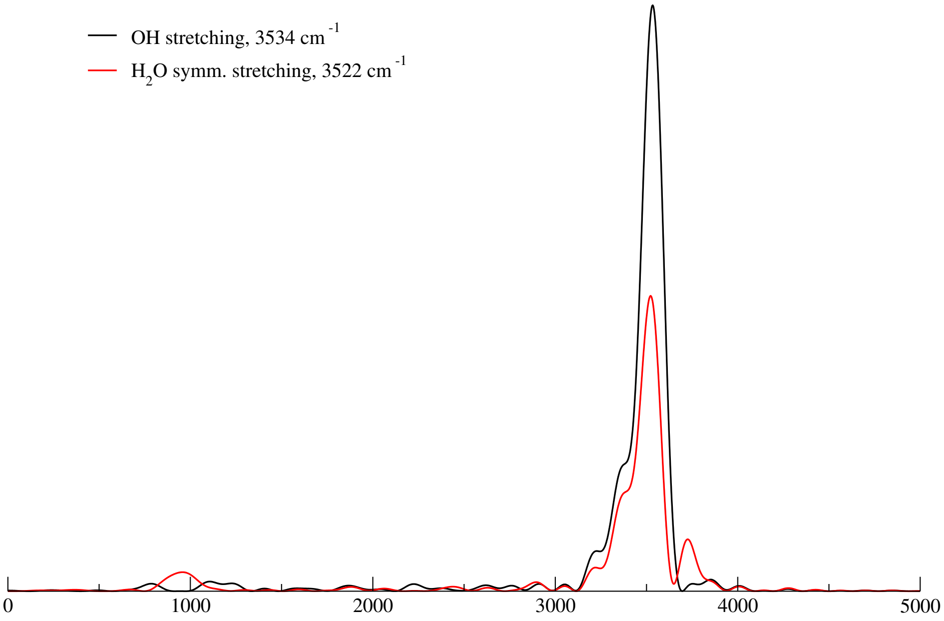

Figure 23: QCT spectrum of the i-I conformer of the MLA monohydrate with the O^{18} isotope substitution on the carbonyl moiety

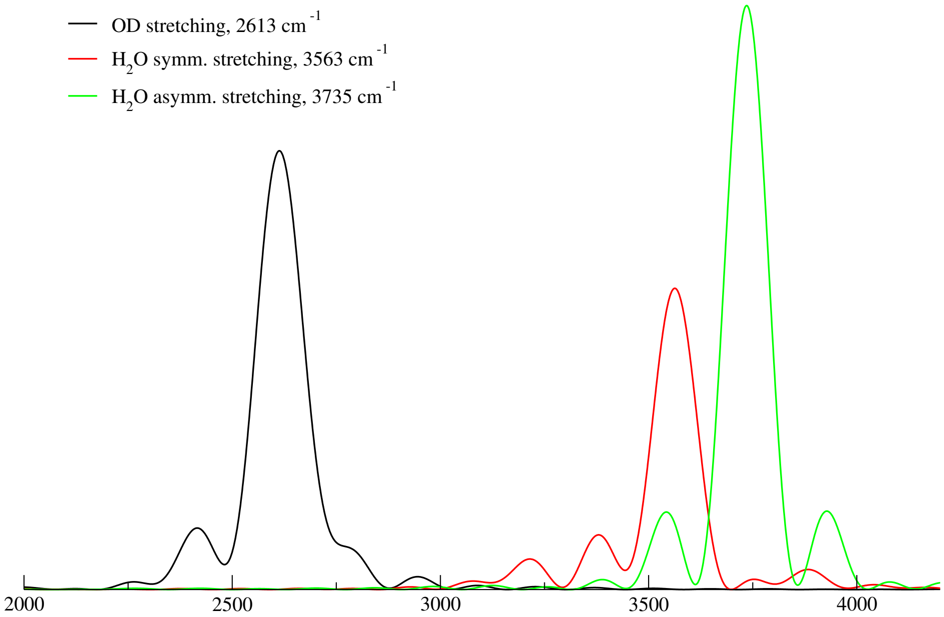

Figure 24: QCT spectrum of the i-III conformer of the MLA monohydrate with the deuterium isotope substitution on the OH moiety. OD stretch harmonic estimate $2610.76\,\mathrm{cm}^{-1}$, water symmetric stretch harmonic estimate $3615.76\,\mathrm{cm}^{-1}$, water asymmetric stretch harmonic estimate $3866.12\,\mathrm{cm}^{-1}$

2.4 Relative Energies for Local Minima and Spectral Properties

We considered all the local minima within a range of $5\,\mathrm{kJmol^{-1}}$ from the global minimum. Energy differences are provided with and without harmonic Zero Point Energy (ZPE) corrections.

2.4.1 Training Set

In the training set we found two possible conformers for the **acetophenone monohydrate** (APH). Indeed, according to the literature [6], the acetophenone monohydrate has two different stable conformers called isomer I and isomer II, with the former representing the global minimum. Our calculations show that isomer I is $1.187\,\mathrm{kJmol^{-1}}$ lower in energy than isomer II. However, if the ZPE correction is also included, the result changes and isomer I becomes the highest in energy, with a difference of $1.018\,\mathrm{kJ\,mol^{-1}}$. We decided to consider isomer I as the global minimum of this complex and to provide its value for the symmetric water stretching frequency.
The local minimum geometry (isomer II) is shown in table 4, while the power spectrum for the water symmetric stretch is shown in figure 25.

O	-0.80696031	3.59269528	0.00889066
O	1.29182941	1.71327983	0.38072124
C	2.75271697	-0.13884867	0.63234146
C	1.40286018	0.49156953	0.35059221
C	-0.72376020	-2.60124380	-0.31906458
C	-1.97612169	-2.02088212	-0.54474329
C	0.37574981	-1.79227047	-0.02547814
C	0.23734454	-0.39334497	0.04240368
C	-1.02799812	0.18048095	-0.18655767
C	-2.12475271	-0.63023378	-0.47607224
H	-0.01754610	3.03876567	0.14714796
H	-0.50586133	4.50036464	0.12116161
H	3.09899152	-0.71537896	-0.23745809
H	3.47248793	0.65370981	0.85654024
H	2.68776156	-0.83184301	1.48307793
H	1.34211763	-2.26205992	0.14828431
H	-0.60359742	-3.68349831	-0.37156290
H	-2.83596777	-2.65152332	-0.77448170
H	-3.09999180	-0.17479515	-0.64909735
H	-1.15173264	1.25974236	-0.13715802

Table 4: Acetophenone monohydrate isomer II

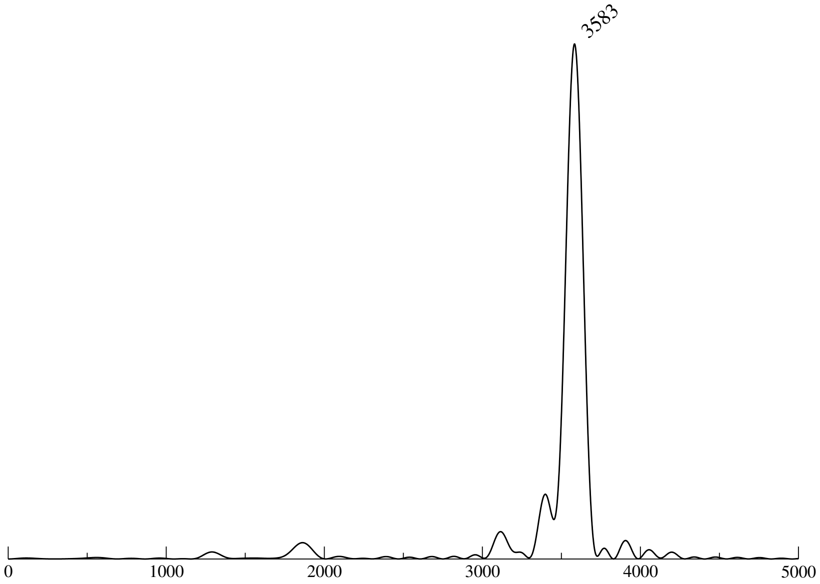

Figure 25: Acetophenone monohydrate isomer II QCT spectrum for the water symmetric stretch (mode 53), harmonic estimate $3640.90\,\mathrm{cm}^{-1}$

2.4.2 Test Set

2.4.2.1 Methyl Lactate Monohydrate

For the methyl lactate monohydrate (MLA) we have two possible coordination options for the water molecule. It can either add to the carbonyl group and the OH group or it can insert into the weak intramolecular hydrogen bond between the OH and the carbonyl group. The different conformers will be named using the nomenclature from the available literature [7]. According to the literature calculations (MP2/6-311++G(d,p) with ZPE corrections), we can have seven possible conformers with the most stable ones being the insertions I, II and III. Then, we have the addition IV, V and the insertion VI. In our studies the insertion VII conformer has been discarded because of its high energy with respect to the global minimum i-I (insertion I). We also discarded the isomers i-II and i-VI because they are respectively identical to the isomers i-I and i-III with the water slightly changed in its position upon rotation.
As found in the literature, the global minimum geometry is i-I. The positions in energy of the other local minima with respect to the global one are summarised in table 5.

Conformers	i-I	i-III	a-IV	a-V	
i-I	0	3.545	3.401	2.975	ZPE included
i-I	0	4.422	11.581	10.967	ZPE not included

Table 5: MLA energy differences with respect to the global minimum i-I in kJmol^{-1} (with and without the ZPE)

The geometries of the computed local minima are shown in tables 6, 7 and 8. Their QCT spectra are shown in figures 26, 27 and 28.

C	-2.71147425	-0.90745749	-0.21103262
O	-1.30253267	-1.12946777	-0.44290812
H	2.43050828	-2.07564721	-0.30412677
H	-2.91955160	-0.89144495	0.86559755
H	-3.02282517	0.04593634	-0.65485519
H	-3.21942399	-1.74700246	-0.69315336
C	-0.47140742	-0.19473045	0.04718586
H	1.28095042	-2.13192750	1.05326870
H	0.75360836	-2.63583213	-0.58038247
C	0.98906181	-0.48579450	-0.32140888
O	-0.86800603	0.77601552	0.66824447
O	1.88826248	0.35124729	0.36298895
C	1.38090444	-1.93300063	-0.02183375
H	1.65902941	1.29006631	0.22335890
H	1.04005447	-0.32078534	-1.41640533
O	0.79089611	2.95721102	0.21818087
H	0.06257451	2.41141043	0.57152325
H	1.05109929	3.54629641	0.93558122

Table 6: Methyl lactate monohydrate isomer i-III

C	-2.95750916	-1.25496680	-0.52142333
O	-1.54365993	-1.00916674	-0.70149691
C	-1.01037027	-0.06929050	0.08949327
H	-3.52830871	-0.34052575	-0.72306902
H	-3.20811187	-2.03857144	-1.24109873
H	-3.15505493	-1.58782216	0.50461720
H	2.34437829	-0.93440091	-0.04597597
H	1.13488255	-1.47327065	1.15221337
H	0.96747801	-1.95144247	-0.56571697
C	0.47830890	0.12376458	-0.15985269
O	0.94359333	1.19125784	0.66094160
O	-1.63458830	0.56740354	0.91761259
C	1.28169963	-1.14937157	0.11331464
H	0.60148630	0.41163634	-1.21681227
H	0.18792387	1.51150866	1.17979723
O	3.55501721	1.33248749	-0.37769893
H	4.08195587	2.13790384	-0.34697523
H	2.74185912	1.53345452	0.11674670

Table 7: Methyl lactate monohydrate isomer a-IV

C	-1.88595448	-1.15084025	0.16915797
O	-0.50692099	-1.03575558	0.61066914
H	2.71005298	0.39915048	2.37830605
H	-2.41900699	-0.20951345	0.34061422
H	-1.92273887	-1.39706464	-0.89750238
H	-2.30974226	-1.95703991	0.77246452
C	0.23943318	-0.11003329	0.01184461
H	1.13809335	-0.41145230	2.64465985
H	1.19744067	1.31603243	2.17250766
C	1.66013880	-0.06124145	0.55813935
O	-0.15121164	0.63319331	-0.87554561
O	2.44367474	0.83972850	-0.20331549
C	1.66945766	0.33345125	2.03938472
H	1.85781463	1.25986259	-0.85184404
H	2.07835082	-1.07664134	0.45143892
O	-2.76499684	0.90077595	-2.06640757
H	-2.81380044	1.58655555	-2.74080724
H	-1.84363305	0.91075871	-1.75603395

Table 8: Methyl lactate monohydrate isomer a-V

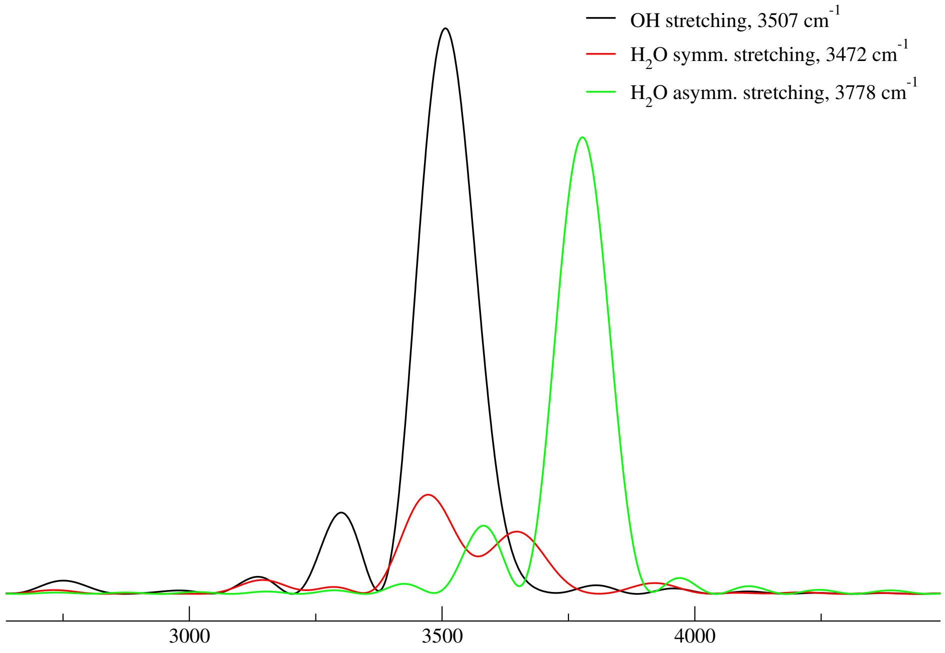

Figure 26: QCT spectrum of the i-III conformer of the methyl lactate monohydrate. OH stretch harmonic estimate $3566.57\,\mathrm{cm}^{-1}$, water symmetric stretch harmonic estimate $3631.24\,\mathrm{cm}^{-1}$, water asymmetric stretch harmonic estimate $3866.15\,\mathrm{cm}^{-1}$

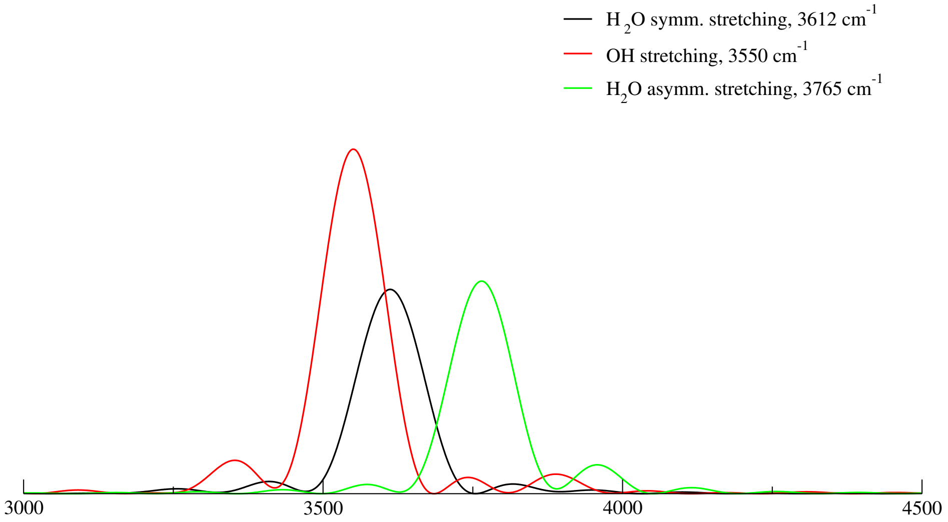

Figure 27: QCT spectrum of the a-IV conformer of the methyl lactate monohydrate. OH stretch harmonic estimate $3670.00\,\mathrm{cm}^{-1}$, water symmetric stretch harmonic estimate $3707.85\,\mathrm{cm}^{-1}$, water asymmetric stretch harmonic estimate $3880.69\,\mathrm{cm}^{-1}$

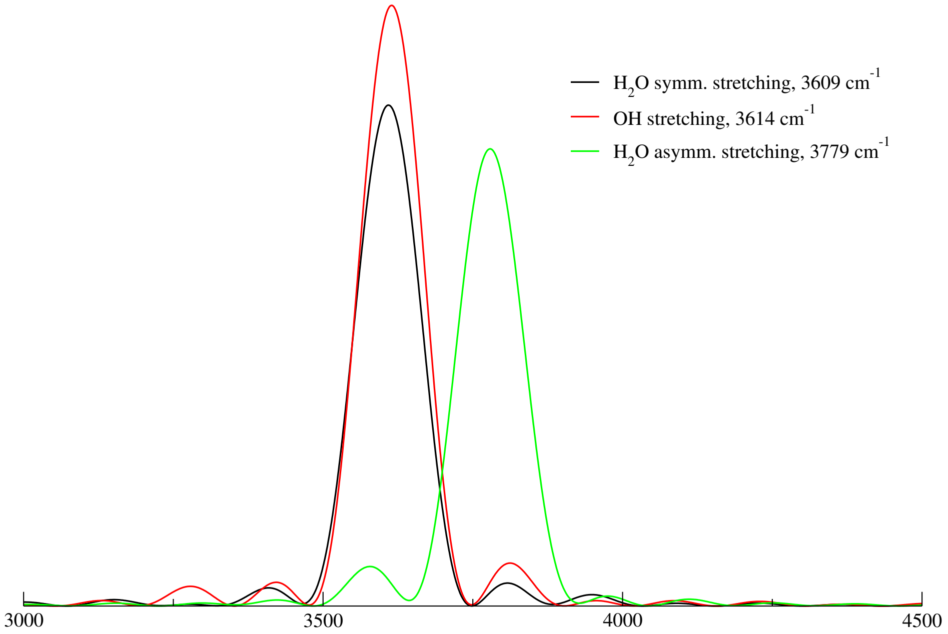

Figure 28: QCT spectrum of the a-V conformer of the methyl lactate monohydrate. OH stretch harmonic estimate $3679.26\,\mathrm{cm}^{-1}$, water symmetric stretch harmonic estimate $3731.92\,\mathrm{cm}^{-1}$, water asymmetric stretch harmonic estimate $3879.54\,\mathrm{cm}^{-1}$

2.4.2.2 Cyclooctanone Monohydrate

The cyclooctanone monomer structure has been characterised extensively [8] and the minimum (and dominant) conformation is the boat-chair with non equivalent oxygen lone pairs (BC1). The water molecule can be placed in two positions giving two possible docking isomers: BC1a and BC1b. From the MP2/6-311++G(d,p) calculations found in the literature [9], the global minimum for the cyclooctanone monohydrate (CON) is supposed to be the BC1a, with the BC1b minimum $0.622\,\mathrm{kJmol^{-1}}$ higher in energy. Anyway, with our DFT calculations we find the global minimum for the complex to be in the BC1b conformation with the BC1a conformer $3.725\,\mathrm{kJmol^{-1}}$ higher in energy (ZPE included). Without ZPE we have an energy difference of $0.996\,\mathrm{kJmol^{-1}}$. Given this discrepancy we carried out an MP2/aug-cc-pVDZ calculation and the BC1a isomer resulted to be the global minimum with an energy difference of $0.359\,\mathrm{kJmol^{-1}}$ (ZPE included) from the BC1b isomer. Without ZPE we have an energy difference of $0.07\,\mathrm{kJmol^{-1}}$. Given these subtle differences we decided to provide the DFT QCT estimates for both the isomers. The BC1a geometry is shown in table 9 and its QCT power spectrum for the symmetric water stretch is shown in figure 29.

C	-1.61235938	1.54051040	0.18989947
C	-2.21161352	0.14860831	-0.07552425
C	-0.30069910	1.87025480	-0.56095802
H	-0.01707189	2.90544165	-0.32845922
O	1.35897500	1.19431091	0.99745021
C	0.84628935	0.98947831	-0.09737203
H	2.02859773	-0.75265761	-0.44085896
H	1.91861839	0.39116638	-1.79595741
C	-1.34399018	-1.04113483	0.40422550
H	-3.18299882	0.11595489	0.43837168
H	-2.44101678	0.04503707	-1.14837041
C	-0.59266947	-1.81048269	-0.69893193
H	-0.62991221	-0.70170917	1.16705194
H	-1.98424392	-1.76839755	0.92327556
C	0.23578580	-0.96089011	-1.67392693
H	0.06965971	-2.54246016	-0.21317155
H	-1.32211980	-2.38521045	-1.29158235
C	1.33682905	-0.11815717	-1.00751750
H	0.71945886	-1.63158736	-2.39843605
H	-0.42641887	-0.31328779	-2.26605885
H	-2.35232320	2.30711570	-0.08228481
H	-1.42004325	1.65165871	1.26737631
H	-0.45302633	1.78718870	-1.64512638
O	2.03244143	-1.34628461	2.02499722
H	2.42504671	-1.37911247	2.90379261
H	1.94904721	-0.39671314	1.81433509

Table 9: Cyclooctanone monohydrate conformer BC1a

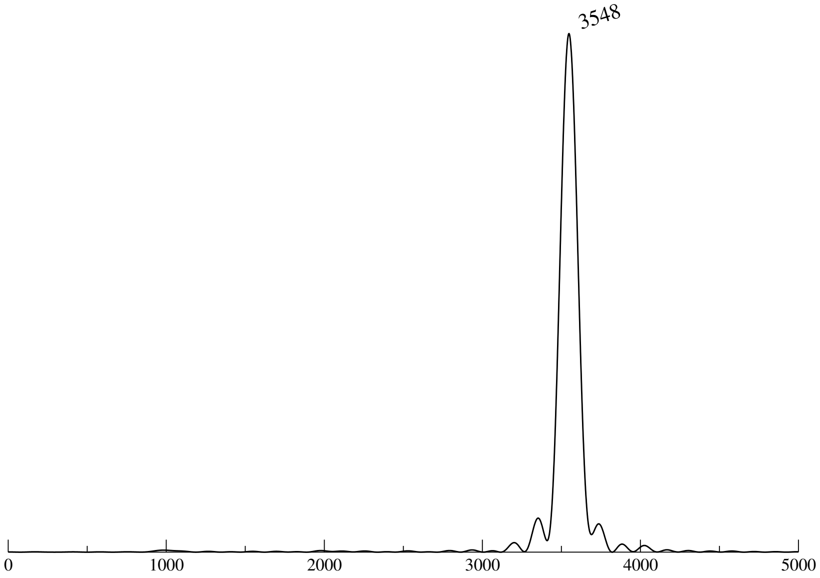

Figure 29: QCT spectrum of the BC1a conformer of the cyclooctanone monohydrate for the water symmetric stretch (mode 71), harmonic estimate $3598.27\,\mathrm{cm}^{-1}$

2.4.2.3 2,2,2-Trifluoroacetophenone Monohydrate

For the 2,2,2-trifluoroacetophenone monohydrate (TPH), from the literature [10], two possible isomers have been predicted, with the water acting as hydrogen bond donor in both of them. In the isomer I we have the water oxygen atom interacting with an hydrogen atom from the phenyl ring, while in the second isomer we have the water molecule interacting with the fluorine atoms. According to both the literature and our calculations, the TPH I conformer is the global minimum. The local minimum geometry TPH II, see table 10, is $8.897\,\mathrm{kJmol}^{-1}$ higher in energy (ZPE included) with respect to the global minimum. Without ZPE we have an energy difference of $6.587\,\mathrm{kJmol}^{-1}$. From the literature this energy difference is only $4.5\,\mathrm{kJmol}^{-1}$ (ZPE included) and given the discrepancies we decided to provide also the QCT spectrum for the TPH II conformer shown in figure 30.

Note that the APH I global minimum (from the training set) corresponds to the TPH II minimum, and given the same water position in the two complexes, we can see that the red shift is larger for the APH complex. Therefore, we can assume that the fluorine atoms change the hydrogen bond interaction and this results in smaller red shifts.

C	3.05211762	1.53971487	0.00000086
C	3.78163775	0.34323510	-0.00000040
C	3.11637372	-0.88710517	-0.00000043
C	1.72146970	-0.92941870	-0.00000017
C	0.98267270	0.26950202	-0.00000013
C	1.66107504	1.50500739	0.00000056
H	3.57226539	2.49765213	-0.00000024
H	4.87180324	0.37188287	-0.00000001
H	3.68561588	-1.81663533	0.00000011
H	1.21965657	-1.89371346	0.00000021
H	1.07749030	2.42426684	-0.00000038
C	-0.49660495	0.30983581	-0.00000031
O	-1.14678966	1.33702305	0.00000042
C	-1.28141192	-1.04009324	-0.00000048
F	-0.97375502	-1.77887111	-1.09526928
F	-0.97375628	-1.77886987	1.09526969
F	-2.60535904	-0.83634098	-0.00000034
O	-4.02047581	2.16517379	0.00000014
H	-4.54849466	1.35914079	0.00000042
H	-3.10291260	1.85317403	-0.00000023

Table 10: 2,2,2-Trifluoroacetophenone monohydrate conformer TPH II

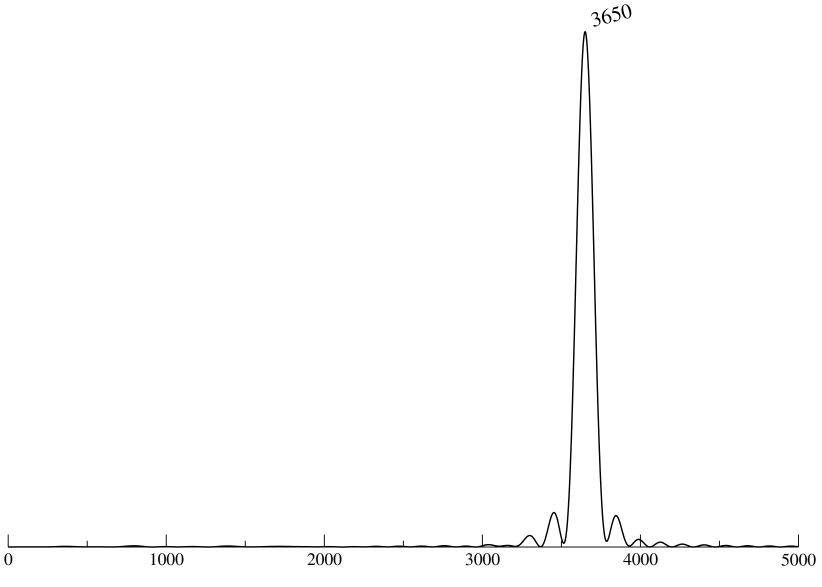

Figure 30: QCT spectrum of the TPH II isomer of the 2,2,2-trifluoroacetophenone monohydrate for water symmetric stretch (mode 53), harmonic estimate $3743.04\,\mathrm{cm}^{-1}$

2.4.2.4 1-phenylcyclohexane-cis-1,2-diol Monohydrate

In our theoretical study, in the 1-phenylcyclohexane-cis-1,2-diol monohydrate (PCD), the water can have three different binding positions: bridging in the intramolecular hydrogen bond interaction between the two alcohol moieties, bridging in the interaction between the alcohol group and the phenyl ring and lastly the water molecule can be in an external position behaving as an hydrogen bond donor.

The global minimum geometry (PCD1) is the one with the water bridging in the interaction between the alcohol moiety and the phenyl ring. The closest local minimum we found (PCD2) is shown in table 11, with an energy difference of $4.052\,\mathrm{kJmol}^{-1}$ (ZPE corrected) from the global minimum. Without ZPE we have an energy difference of $0.788\,\mathrm{kJmol}^{-1}$.

The other two relevant local minima that we found (PCD3 and PCD4) are shown in figures 31 and 32. Their energy gaps from the global minimum are respectively $12.522\,\mathrm{kJmol}^{-1}$ and $8.669\,\mathrm{kJmol}^{-1}$ (including the ZPE). Without ZPE the energy gaps are respectively $10.439\,\mathrm{kJmol}^{-1}$ and $8.628\,\mathrm{kJmol}^{-1}$, therefore the QCT spectra for these isomers will not be computed.

In the PCD2 conformer we call OH_a the H bond donor alcohol and OH_b the H bond acceptor alcohol. The QCT spectrum of the PCD2 isomer is shown in figure 33.

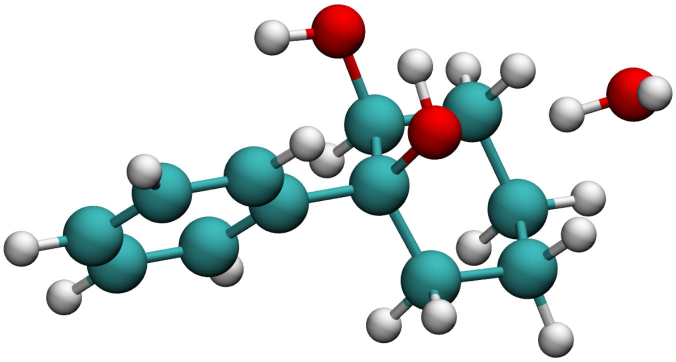

Figure 31: 1-phenylcyclohexane-cis-1,2-diol monohydrate PCD3

C	-0.53848716	-0.26747855	0.18866092
C	-2.05128194	-0.38397293	-0.10040017
H	-1.96633997	-2.65400471	-1.79045626
C	-0.02146044	1.14952259	-0.09552590
C	0.20751365	-1.28538934	-0.72276953
H	1.93235428	-0.33631951	-0.76387105
H	0.30826473	-3.39086573	-1.12056726
H	-0.02530566	-2.97844977	0.57401703
C	-2.55871044	-1.81587779	0.11920080
H	-2.57207099	0.32190681	0.55978668
H	-2.24736655	-0.06733923	-1.13494684
C	-1.76653733	-2.82925032	-0.71994746
H	-2.46073013	-2.06566797	1.18396851
H	-3.62767866	-1.86881414	-0.13288538
C	-0.25541130	-2.71193924	-0.46650608
H	-2.10206088	-3.85269361	-0.50079682
H	0.53999339	-0.76314398	1.79682943
O	-0.39485248	-0.61219400	1.56163861
O	1.63842462	-1.23019284	-0.53711877
H	-0.01069027	-1.02702711	-1.77139928
C	0.53594988	1.93257465	0.92463509
C	1.02362677	3.21706438	0.65843265
C	0.96504201	3.74225351	-0.63479727
C	0.41129827	2.97157478	-1.66255932
C	-0.07593886	1.68990158	-1.39332778
H	0.57980839	1.53582331	1.93645216
H	1.45144158	3.80828281	1.46904234
H	1.34561383	4.74300531	-0.84163580
H	0.35612112	3.36926025	-2.67672590
H	-0.50982742	1.11288564	-2.21021536
O	2.32167769	-1.28677348	2.12308918
H	2.28367511	-1.47803883	1.16378771
H	2.45036833	-2.13586388	2.56024112

Table 11: 1-phenylcyclohexane-cis-1,2-diol monohydrate conformer PCD2

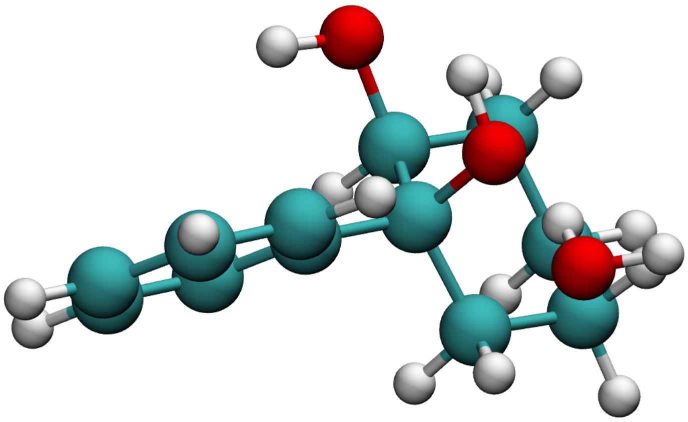

Figure 32: 1-phenylcyclohexane-cis-1,2-diol monohydrate PCD4

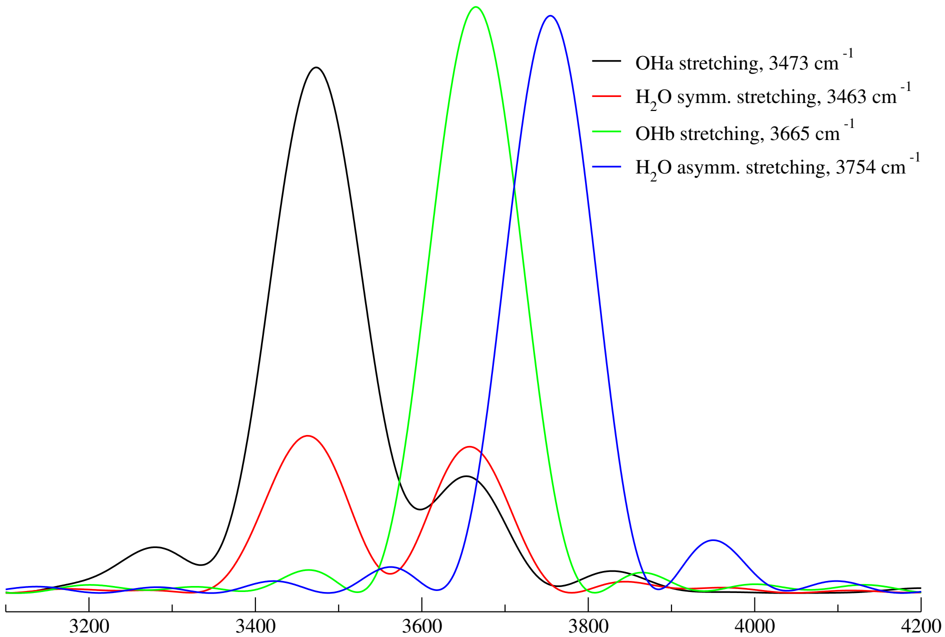

Figure 33: QCT spectrum of the PCD2 isomer of the PCD complex. OH_a harmonic estimate $3535.34\,cm^{-1}$, OH_b harmonic estimate $3765.34\,cm^{-1}$, water symmetric OH stretch harmonic estimate $3604.92\,cm^{-1}$, water asymmetric OH stretch harmonic estimate $3866.44\,cm^{-1}$

2.5 Other Computed Quantities for the Training and Test Sets

For the **2,2,2-trifluoroethan-1-ol monohydrate** (TFE) we have six possible conformations [11] but we considered only three of them as the others are rather high in energy. The nomenclature of these compounds is composed by the conformation of the 2,2,2-trifluoroethan-1-ol monomer: gauche or trans (g, t), and the position of the water molecule: insertion or addition (i, a). The global minimum sits in the IG conformation and the other minima sit relatively high in energy with respect to the global minimum, therefore we do not expect to see them in the spectrum, given the experimental conditions. Indeed, the AT$_2$ minimum (see table 12) lies at $12.246\,\mathrm{kJmol}^{-1}$ ($8.970\,\mathrm{kJmol}^{-1}$ if the ZPE is included) with respect to the global one and the AG minimum (see table 13) lies at $12.176\,\mathrm{kJmol}^{-1}$ ($9.218\,\mathrm{kJmol}^{-1}$ if ZPE corrected). Given the blue shift of the symmetric water stretch in the IG global minimum conformer, we decided to compute the QCT spectrum for the AT$_2$ conformer, shown in figure 34. In this conformer we only have a pure H bond acceptor interaction of the water molecule, and the magnitude of the blue shift is smaller if compared to the IG conformer where we also have the donor interaction with the fluorine atoms.

C	0.50788581	0.07523174	0.30695390
C	-0.93490382	-0.35572924	0.49893180
H	0.67642529	0.92820338	0.98567466
H	1.13672277	-0.76914518	0.63577917
F	-1.81437911	0.62224126	0.18303757
O	0.69637706	0.40563832	-1.04737462
H	1.62215830	0.67770731	-1.17636213
F	-1.25455275	-1.44133892	-0.24201196
F	-1.14312844	-0.68136413	1.80555638
O	3.40030804	1.18939884	-1.39914098
H	3.51301791	2.09465267	-1.71635261
H	3.88936894	0.63908394	-2.02423118

Table 12: 2,2,2-Trifluoroethan-1-ol monohydrate conformer AT$_2$

C	0.41540348	0.54396736	0.38047665
C	-0.98117865	-0.05474905	0.32324947
H	0.36398324	1.44505953	1.01010504
H	1.09184557	-0.18526763	0.83944885
F	-1.83899886	0.78255860	-0.32936816
O	0.91364250	0.81370652	-0.91847461
H	0.37273099	1.49364171	-1.33967289
F	-1.01683614	-1.23982146	-0.32167760
F	-1.46915460	-0.24547061	1.57260164
O	3.27619082	-0.70102922	-0.43281917
H	3.53593226	-1.44585423	-0.98685545
H	2.65547898	-0.18588773	-0.97060058

Table 13: 2,2,2-Trifluoroethan-1-ol monohydrate conformer AG

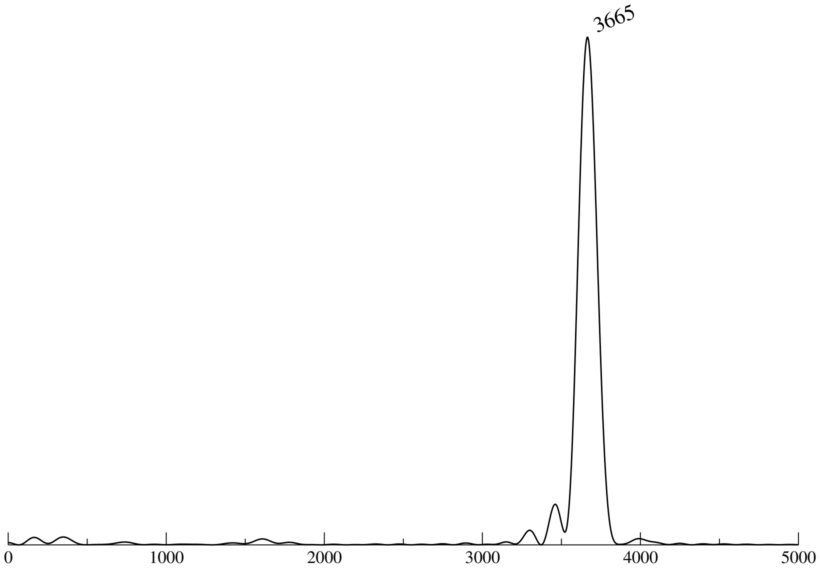

Figure 34: QCT spectrum of the AT_2 conformer of the 2,2,2-trifluoroethan-1-ol monohydrate for the symmetric water stretch (mode 29), harmonic estimate $3789.96\,\mathrm{cm}^{-1}$

References

(1) Rognoni, A.; Conte, R.; Ceotto, M. *Journal of Chemical Physics* **2021**, *154*, 094106, DOI: 10.1063/5.0040494.

(2) Rognoni, A.; Conte, R.; Ceotto, M. *Chemical Science* **2021**, *12*, 2060–2064.

(3) Botti, G.; Ceotto, M.; Conte, R. *Journal of Chemical Physics* **2021**, *155*, 234102, DOI: 10.1063/5.0075220.

(4) Aprà, E. et al. *Journal of Chemical Physics* **2020**, *152*, 184102, DOI: 10.1063/5.0004997.

(5) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. *Journal of Chemical Physics* **2010**, *132*, 154104, DOI: 10.1063/1.3382344.

(6) Lei, J.; Zhang, J.; Feng, G.; Grabow, J. U.; Gou, Q. *Physical Chemistry Chemical Physics* **2019**, *21*, 22888–22894.

(7) Thomas, J.; Sukhorukov, O.; Jäger, W.; Xu, Y. *Angewandte Chemie - International Edition* **2014**, *53*, 1156–1159.

(8) Burevschi, E.; Penã, I.; Sanz, M. E. *Physical Chemistry Chemical Physics* **2019**, *21*, 4331–4338.

(9) Burevschi, E.; Sanz, M.; Pena, I. In 2019, pp 1–1.

(10) Lei, J.; Alessandrini, S.; Chen, J.; Zheng, Y.; Spada, L.; Gou, Q.; Puzzarini, C.; Barone, V. *Molecules (Basel, Switzerland)* **2020**, *25*, 4899, DOI: 10.3390/molecules25214899.

(11) Thomas, J.; Xu, Y. *Journal of Chemical Physics* **2014**, *140*, 06B616_1, DOI: 10.1063/1.4883518.