Supplementary Information – HyDRA challenge

Michael Hippler [ORCID: 0000-0002-3956-3922, email: M.Hippler@sheffield.ac.uk]

Department of Chemistry, University of Sheffield, Sheffield S3 7HF, United Kingdom

1 Computational Details

High-level quantum chemical calculations have been performed using the General Atomic and Molecular Electronic Structure System (GAMESS) US program package, version 30 SEP 2018 (R3), on various 64 bit Linux personal computers (office computers, typically Intel i5 and i7 quad core processors, 8 GB RAM) [MH1]. The standard 6-311++G(d,p) basis set (Pople triple split valence set with one added polarization function and one diffuse function on all atoms, including hydrogen) has been employed. Dynamic electron correlation has been included with 2nd order Møller-Plesset perturbation theory (MP2, frozen core). Calculations on molecular clusters are affected by the basis set superposition error (BSSE). For energy, geometry optimization (gradients) and vibrational calculations (Hessian) on molecular clusters, the BSSE was corrected using a suite of self-written programmes and scripts [MH2-5], implementing the counterpoise correction procedure (CP) as proposed by Boys and Bernardi [MH6] within GAMESS. In the following, we call this method CP MP2/6311++G(d,p). In addition, after trialing various alternative basis sets on the training set, we performed calculations with a reduced basis set, 6-31++G(2d,p), a Pople double split valence set with 2 added d polarisation functions and one p function on hydrogen, and one diffuse function on all atoms, including hydrogens. We refer to this method as CP MP2/631++G(2d,p). Uncertainties in frequency shift calculations are obviously very difficult to estimate, but the RMS of errors in the training set is about 8%. In the manuscript, this is submission PH5.

Geometry optimizations were performed with very tight convergence criteria with remaining gradients below 10⁻⁶ Hartree/Bohr in order to obtain well-defined equilibrium geometries (for comparison, the default is 10⁻⁴ Hartree/Bohr). The nature of stationary states (minimum structures or transition states) was confirmed by CP corrected frequency calculations with numerical derivatives of the gradients, where positive and negative Cartesian displacements were enforced to increase the accuracy. A vibrational scaling factor of 0.95 was used to partially compensate for anharmonicities and other effects for zero-point energies (ΔE_0) and thermodynamic properties.

In calculations for the training set it has been found that CP MP2/631++G(2d,p) performs (surprisingly) well, better than the more involved CP MP2/6311++G(d,p). This surely must be due to some extent of error compensations. For convenience, we submit values using this method as the first of our 2 submissions. We believe that more accurate calculations which do not rely as much to fortuitous error compensations must include harmonic couplings and anharmonic effects in the water vibrational modes. In isolated water, the O-H stretching local vibrations combine in the harmonic approximation to the asymmetric O-H stretching normal mode (higher in frequency than the O-H local mode), and the symmetric O-H stretching normal mode (lower in frequency than the O-H local mode). In the complex, the harmonic O-H

stretching modes resemble more local modes, however. If frequency shifts are referenced to the combined normal modes in isolated water and the (approximately) local mode in the complex, then the shift depends on the accuracy of the description of the hydrogen bonding in the complex, and in addition on how well the splitting of harmonic modes in isolated water is modeled. Frequency shifts are also affected by anharmonic interactions, in particular the Fermi resonance between the bending and (symmetric) stretching modes in water. This is an issue in accurate frequency shift calculations, in particular since the Fermi resonance will become more pronounced in the complex where the red shift of the symmetric stretching vibration will bring the two interacting vibrations closer together and thus tune more into resonance. We therefore decided to include in the CP MP2/6311++G(d,p) calculations anharmonic effects for the water vibrations, all the while continuing with CP corrections. Since full dimensional potential energy surface calculations are not feasible, we have to use some severe approximations to capture the essential effects of anharmonic interactions of water vibrations. We have self-written programmes and scripts to implement the following within GAMESS:

Using the harmonic normal mode water vibrations (bending, symmetric and asymmetric stretching) following from the CP MP2/6311++G(d,p) calculation of the complex, find the displacement of water atoms during these modes. Energies are then sampled along the directions of the harmonic normal modes and along pairs of harmonic normal modes (2-D grids along each pair of normal modes), using the Quartic Force Field (QFF) approximation of Yagi et al. [MH7], which computes a fit to the derivatives up to fourth order by computing a specialized set of geometry points. These calculations are only done for water normal modes; modes and displacements of the complex partner are not included. In the CP corrected energy calculations, the complex partner is only included in its frozen geometry of the CP corrected equilibrium complex. The justification for not including modes of the complex partner is first it would be too computationally involved, and second, in the weakly hydrogen bound complex it is thought that water and complex partner vibrations do not interact much (but see below). Vibrational wavefunctions are obtained then obtained by GAMESS at an SCF-like level, termed VSCF, using product nuclear wavefunctions, along with an MP2-like correction to the vibrational energy, which is termed correlation corrected (cc-VSCF) [MH8-10]. For IR intensity calculations, the harmonic dipole derivative tensor is used, rather than computing dipoles along the anharmonic grid points.

In some occasions, we have found that water normal modes (e.g., the water bending mode) are mixed with complex partner normal modes in the harmonic approximation, which causes problems in our simplified approach assuming isolated water modes. An effective way to avoid this unwanted mixing is to do frequency calculations where all H of the complex partner are deuterated to D which in almost all cases shift vibrations away to leave the water vibrations isolated. Where there is no mixing to start with, this procedure gives almost the same frequency shift than without deuteration. We finally call this method including CP corrections and selected anharmonic effects in MP2 calculations of water complexes 'CP MP2/6311++G(d,p) 2-D QFF cc-VSCF'. In the manuscript, this is submission AC4. More details will be presented in a forthcoming publication [MH11]. We consider this method as our best effort concerning methodology and submit results as the second set of our 2 submissions. Again, uncertainties are very difficult to estimate for this method, but the RMS of errors in the training set is about 21%. This rather large RMS is dominated by 2 outliers; reasons for these outliers are unclear at present, but might be related to anharmonic interactions not captured by our simplified treatment.

2

- [MH1] Schmidt, M. W.; Baldridge, K. K.; Boatz, J. A.; Elbert, S. T.; Gordon, M. S.; Jensen, J. H.; Koseki, S.; Matsunaga, N.; Nguyen, K. A.; Su, S.; Windus, T. L.; Dupuis, M.; Montgomery Jr, J. A., General atomic and molecular electronic structure system. J. Comp. Chem. 1993, 14 (11), 1347-1363. DOI: 10.1002/jcc.540141112
- [MH2] Hippler, M., Quantum-chemical study of CHCl₃–SO₂ association. J. Chem. Phys. 2005, 123 (20), 204311. DOI: 10.1063/1.2121609
- [MH3] Chung, S.; Hippler, M., Infrared spectroscopy of hydrogen-bonded CHCl₃–SO₂ in the gas phase. The J. Chem. Phys. 2006, 124 (21), 214316. DOI: 10.1063/1.2207617
- [MH4] Hippler, M., Quantum chemical study and infrared spectroscopy of hydrogen-bonded CHCl₃–NH₃ in the gas phase. J. Chem. Phys. 2007, 127 (8), 084306. DOI: 10.1063/1.2757176
- [MH5] Hippler, M.; Hesse, S.; Suhm, M. A., Quantum-chemical study and FTIR jet spectroscopy of CHCl3–NH3 association in the gas phase. Phys. Chem. Chem. Phys. 2010, 12 (41), 13555-13565. DOI: 10.1039/c0cp00530d
- [MH6] Boys, S. F.; Bernardi, F., The calculation of small molecular interactions by the differences of separate total energies. Some procedures with reduced errors. Mol. Phys. 1970, 19 (4), 553-566. DOI: 10.1080/00268977000101561
- [MH7] Yagi, K.; Hirao, K.; Taketsugu, T.; Schmidt, M. W.; Gordon, M. S., Ab initio vibrational state calculations with a quartic force field: applications to H₂CO, C₂H₄, CH₃OH, CH₃CCH, and C₆H₆. J. Chem. Phys. 2004, 121, 1383-1389. DOI: 10.1063/1.1764501
- [MH8] Gerber, R. B.; Jung, J. O., in "Computational Molecular Spectroscopy", Jensen, P.; Bunker, P.R., eds., Wiley and Sons, Chichester, 2000, pp 365-390.
- [MH9] Chaban, G. M.; Jung, J. O.; Gerber, R.B., Ab initio calculation of anharmonic vibrational states of polyatomic systems: Electronic structure combined with vibrational self-consistent field. J. Chem. Phys. 1999, 111, 1823-1829. DOI: 10.1063/1.479452
- [MH10] Matsunaga, N.; Chaban, G. M.; Gerber, R. B., Degenerate perturbation theory corrections for the vibrational self-consistent field approximation: Method and applications. J. Chem. Phys. 2002, 117, 3541-3547. DOI: 10.1063/1.1494978
- [MH11] Hippler, M., in preparation, to be published.

2 Additional computed data

Summary of results:

Set 1 (PH5), CP MP2/631++G(2d,p):

Calculated reference for water symmetric stretching fundamental: 3850.03 cm⁻¹, IR intensity 11.6 km mol⁻¹.

HvDRA Training Set

Molecule	CAS- number	exp. OHb shift / cm ⁻¹	calc. OHb shift / cm ⁻¹	IR int. / km mol ⁻¹	∆E _{el} / kJ mol ⁻¹	∆ <i>E</i> ₀ / kJ mol⁻¹	r _e / Å
ACE Acetone	67-64-1	119 / 126 ^{a)}	120.2	322.7	-24.2	-17.0	1.98 OHO=C
APH Acetophenone	98-86-2	121 / 127 ^{a)}	119.1	434.8	-23.8	-17.1	1.98 OH O=C
TFB 1,2,4,5-Tetrafluorobenzene	327-54-8	10	9.5	9.8	-14.1	-10.0	2.35 HO HC, 2.36 OH F
POH 1-Phenylethanol	98-85-1	37	38.4	47.4	-26.1	-19.7	2.08 HO HOC
IMZ Imidazole	288-32-4	199	188.4	535.8	-28.2	-21.1	2.00 OHN
ANL Aniline	62-53-3	133	152.8	437.1	-21.8	-15.1	2.08 OHN
DBF Dibenzofuran	132-64-9	34	38.9	116.0	-15.2	-10.0	2.12 OH […] O
DBN Di-tert-butyl nitroxide	2406-25-9	173	151.7	350.6	-28.3	-21.7	1.90 OH O
OCP o-Cyanophenol	611-20-1	62	60.0	60.8	-35.8	-27.8	1.86 HO HOC
CBU Cyclobutanone	1191-95-3	109	108.9	251.9	-23.4	-16.6	2.01 OHO

^{a)} Deperturbed.

HyDRA Challenge Set

abbreviation - Molecule	CAS- number	calc. OHb shift / cm ⁻¹	IR int. / km mol ⁻¹	∆E _{el} / kJ mol⁻¹	∆ <i>E</i> ₀ / kJ mol⁻¹	r _e / Å
FAH Formaldehyde	50-00-0	77.3	145.4	-19.9	-12.4	2.06 OH O
THF Tetrahydrofuran	109-99-9	176.4	378.1	-26.5	-18.8	1.90 OHO
THT Tetrahydrothiophene	110-01-0	129.8	240.2	-21.0	-14.9	2.45 OH S
PYR Pyridine	110-86-1	195.8	603.8	-25.7	-18.8	2.01 OH N
CON Cyclooctanone ^{b)}	502-49-8	141.8 ^{b)}	363.2	-26.4	-19.2	1.96 OH O
		129.7 ^{b)}	277.5	-26.0	-19.0	1.97 OH O
DMI 1,3-Dimethyl-2-imidazolidinone	80-73-9	143.0	454.3	-28.2	-21.3	1.92 OH O
TPH 2,2,2-Trifluoroacetophenone	434-45-7	43.5	112.9	-18.9	-13.3	2.08 OH O
TFE 2,2,2-Trifluoroethan-1-ol	75-89-8	11.9	13.0	-27.7	-21.0	1.94 HO HOC
MLA Methyl lactate	547-64-8	109.4	361.3	-29.0	-19.4	1.98 OH O, 1.98 HO HOC
PCD	125132-75-4	50.6 ^{c)}	66.6	-30.0	-23.2	1.96 HO HOC
1-Phenylcyclohexane-cis-1,2-diol °)		124.6 ^{c)}	347.6	-28.4	-20.2	1.96 OH OC, 1.99 HO HOC

^{b)} 2 complexes found with nearly the same energy, complex b (slightly more stable) and complex a. ^{c)} 2 complexes found with close energy, complex acc (more stable) and complex int.

Set 2 (AC4), CP MP2/6311++G(d,p) 2-D QFF cc-VSCF:

Calculated reference for water symmetric stretching fundamental: 3732.40 cm⁻¹, IR intensity 12.7 km mol⁻¹.

HYDRA Haining Set							
Abbreviation - Molecule	CAS- number	exp. OH♭ shift / cm⁻¹	calc. OHb shift / cm ⁻¹	IR int. / km mol ⁻¹	∆E _{el} / kJ mol⁻¹	∆E₀ / kJ mol⁻¹	r _e / Å
ACE Acetone	67-64-1	119 / 126 ^{a)}	127.7	264.1	-22.4	-15.5	2.02 OHO=C
APH Acetophenone	98-86-2	121 / 127 ^{a)}	128.0	358.2	-22.1	-15.8	2.02 OHO=C
TFB 1,2,4,5-Tetrafluorobenzene	327-54-8	10	5.8	11.0	-13.8	-9.5	2.35 HO HC, 2.43 OH F
POH 1-Phenylethanol	98-85-1	37	34.1	33.2	-24.9	-18.8	2.08 HO HOC
IMZ Imidazole	288-32-4	199	234.9	532.3	-28.0	-20.7	2.02 OH N
ANL Aniline	62-53-3	133	182.4	383.2	-21.6	-14.8	2.10 OH N
DBF Dibenzofuran	132-64-9	34	42.0	102.7	-14.6	-9.5	2.14 OH O
DBN Di-tert-butyl nitroxide	2406-25-9	173	174.5	318.1	-27.6	-21.0	1.92 OH O
OCP o-Cyanophenol	611-20-1	62	47.4	36.6	-33.3	-26.3	1.88 HO HOC
CBU Cyclobutanone	1191-95-3	109	111.4	197.4	-21.1	-14.5	2.06 OH O

HyDRA Training Set

^{a)} Deperturbed.

HyDRA Challenge Set

abbreviation - Molecule	CAS- number	calc. OHb shift / cm ⁻¹	IR int. / km mol ⁻¹	∆E _{el} / kJ mol ⁻¹	∆ <i>E</i> ₀ / kJ mol⁻¹	<i>r</i> _e / Å
FAH Formaldehyde	50-00-0	78.2	111.9	-17.7	-10.6	2.10 OHO
THF Tetrahydrofuran	109-99-9	191.9	349.4	-24.8	-17.8	1.93 OHO
THT Tetrahydrothiophene	110-01-0	158.4	217.0	-19.0	-12.9	2.47 OH S
PYR Pyridine	110-86-1	238.8	579.2	-25.3	-18.3	2.02 OHN
CON Cyclooctanone ^{b)}	502-49-8	150.3 ^{b)}	302.4	-24.2	-17.3	2.00 OHO
		141.0 ^{b)}	251.8	-23.8	-17.0	2.01 OHO
DMI 1,3-Dimethyl-2-imidazolidinone	80-73-9	155.7	406.5	-26.9	-19.8	1.95 OHO
TPH 2,2,2-Trifluoroacetophenone	434-45-7	40.8	87.5	-17.6	-12.0	2.12 OHO
TFE 2,2,2-Trifluoroethan-1-ol	75-89-8	1.8	13.3	-26.9	-20.9	1.96 HOHOC
MLA Methyl lactate	547-64-8	109.6	237.7	-28.0	-18.6	2.03 OH O, 2.02 HO HOC
PCD	125132-75-4	50.5 ^{c)}	52.4	-27.7	-21.8	1.96 HOHOC
1-Phenylcyclohexane-cis-1,2-diol °)		128.9 ^{c)}	210.4	-26.1	-18.2	2.01 OH OC, 2.05 HO HOC

^{b)} 2 complexes found with nearly the same energy, complex b (slightly more stable) and complex a. ^{c)} 2 complexes found with close energy, complex acc (more stable) and complex int.



Figures showing CP-corrected equilibrium geometries of the challenge set, CP MP2/6311++G(d,p)

Figures showing CP-corrected equilibrium geometries of the training set, CP MP2/6311++G(d,p)



DBF

7

3 Additional computed data: the water dimer

After completing the challenge, and as a reference, we have also calculated the water dimer using both our methods. For comparison, experimental values are 3657 cm⁻¹ for the symmetric OH- stretching vibration of the water monomer (Herzberg). The experimental value for the water dimer H-donor bound OH-stretching vibration is still subject to some debate, but the accepted value seems to be 3601 cm⁻¹ (Huisken, F.; Kaloudis, M.; Kulcke, A., Infrared spectroscopy of small size-selected water clusters. J. Chem. Phys. 1996, 104, 17. DOI: 10.1063/1.470871). This would imply an experimental OH_p shift of **-56** cm⁻¹.

Summary of calculations:

Set 1 (PH5), CP MP2/631++G(2d,p):

calc. OH _b shift / cm ⁻¹ :	-79.1
IR int. / km mol ⁻¹	233.2
$\Delta E_{ m el}$ / kJ mol ⁻¹	-19.3
$\Delta E_{ m 0}$ / kJ mol $^{ m -1}$	-11.1
r _e / Å	2.03 (OHO)

Set 2 (AC4), CP MP2/6311++G(d,p) 2-D QFF cc-VSCF:

calc. OH _b shift / cm ⁻¹ :	-76.8
IR int. / km mol ⁻¹	191.9
$\Delta E_{ m el}$ / kJ mol ⁻¹	-19.0
$\Delta E_{ m 0}$ / kJ mol $^{ m -1}$	-11.4
r _e / Å	2.06 (OH O)



CP-corrected equilibrium geometry of water dimer, CP MP2/6311++G(d,p).

As a curious aside, CP MP2/6311++G(d,p) (without anharmonic corrections) gives a shift of - 55.5 cm⁻¹, in perfect agreement with the expected -56 cm⁻¹.

In conclusion, the water dimer calculations seem to be far less satisfactory, predicting shifts which seem significantly too large. The reasons are unclear, but probably related to the wide amplitude motions of these light molecules and harmonic and anharmonic couplings of close vibrational modes of the two water units. Clearly, more sophisticated anharmonic calculations, possibly fully-dimensional, are required, and also a definite resolution, assignment and conclusion of experimental data.