Supplementary Information – HyDRA challenge

Roman M. Balabin [ORCID: 0000-0001-9374-1809, email: balabinrm@gmail.com]¹

¹ Bond Street Holdings, Long Point Road, KN-1002 Henville Building 9, Charlestown, KN10 Nevis, St. Kitts and Nevis

1 Computational Details

Model LS4. The structures of monomers and 1:1 hydrates of all compounds were obtained from manual starting structures and available literature data. The geometries of all hydrates were optimized at dispersion-corrected density functional theory level (DFT-D) with closed- and open-shell Perdew, Burke and Ernzerhof hybrid density functional (PBE0, commonly referred to as PBE1PBE) that uses one-fourth of exact exchange energy. Three-body empirical dispersion correction and Becke-Johnson damping (D3BJ) was applied with Grimme's parameters: $s_8 = 1.2177$, $a_1 = 0.4145$, and $a_2 = 4.8593$. Complex Dunning's correlation consistent basis set of triple/quadruple quality was used to build molecular orbitals: Truhlar's may-cc-pVTZ set was applied to non-hydrated monomeric fragment, while fully augmented aug-cc-pVQZ basis set was added to water atoms. Force constants and the resulting vibrational frequencies were obtained at "very tightly" converged geometries with RMS force threshold of 10⁻⁶ Hartree/Bohr. Pruned "ultra fine" grids with 99 radial shells and 590 angular points per shell were used for all calculations.

Gaussian 09 (revision D.01) software package was used without correction for inaccurate second derivatives of Grimme's dispersion term since this correction was found to be below 1 cm⁻¹ for OH-stretching range under study. A wavefunction analyzer Multiwfn (version 3.8 dev, special update delivered on 7 February 2022) was used for visualization on H-bonding domains and integration of several real space functions, including the Lagrangian kinetic energy density $G(\mathbf{r})$ and PBE/Becke exchange energies. Optimization Toolbox of MATLAB R2020a was used for non-linear regression and function analysis. Self-written MATLAB software was used to extract and analyze electron densities on Lebedev grids with Becke partitioning weight function: to ensure the integration convergence the number of points for numerical quadratures was set to 250 (radial) and 1454 (spherical).

Harmonic wavenumbers of hydrogen-bonded OH bond stretching vibrations (ω_H) and lengths of the OH bonds (r_b) were used to predict anharmonic stretching fundamentals (v_F). Empirical model with five parameters – to compensate, inter alia, for GGA DFT self-interaction error – was applied. (i) Elongation of OH bond was calculated as $\Delta r_b = r_b - r_{b,W}$, where $r_{b,W}$ is the equilibrium bond length of free water at PBE0 level ($r_{b,W}$ is 95.77 and 96.28 pm for closed- and open-shell systems, respectively); (ii) nondimensional normalized elongation was calculated as $\delta = k_r^*(\Delta r_b - \Delta r_{b,0})$ where $\Delta r_{b,0} = 0.694$ pm and $k_r = 1.653$ pm⁻¹; (iii) sigmoid function of δ was built as $S_2(r_b) = \delta^*(1+\delta^2)^{-1/2}$; (iv) the final stretching fundamental was calculated as a linear combination of the harmonic wavenumber and a correction for anharmonicity and PBE0 deficiencies, $v_F = k_H^*\omega_H + A_s^*S_2(r_b) + B_0$, where parameters k_H and B_0 were set to reproduce $H_2O/HOD/D_2O$ experimental frequencies and parameter A_s was fitted to available experimental and anharmonic coupled cluster data for hydrates of organic molecules.

The model was based on forty-two neutral systems in ground electronic state and resulted in MAD of 3.5 cm⁻¹. Exact parameters are $r_{b,W} = 95.772069$ pm, $\Delta r_{b,0} = 0.69393343$ pm, $k_r = 1.6529957$ pm⁻¹, $k_H = 0.917051675$, $A_s = 50.602948$ cm⁻¹ and $B_0 = 152.970453$ cm⁻¹.

Model limitations: (a) OH-stretching fundamentals from 3364 to 3567 cm⁻¹ were used for training; (b) vibrations that could involve a significant fraction of bonded NH (e.g., 2-aminopyridine) or free OH stretching (e.g., water dimer) were excluded since they would probably require a special correction; (c) vibrations with low red-shifts from 3657 cm⁻¹ might be dependent on D3BJ parameters, especially for systems that are not bound on GGA DFT level itself – so higher error (up to 8 cm⁻¹) is expected; (d) sulphur and, especially, phosphorus containing systems are underrepresented in the model set – the same could be said about radicals; (e) complex vibrational interactions (e.g., resonances or intramolecular-intermolecular couplings) are completely out of model scope.

Quantum chemical calculations of formaldehyde-water and methanol-water clusters were performed in years 2009-2011 on the Obelix cluster at the Competence Center for Computational Chemistry in the Department of Chemistry and Applied Biosciences of the Swiss Federal Institute of Technology (ETH Zurich, Switzerland). Three systems used for the model calibration – monohydrates of acridine, 3-aminopyridine, and benzimidazole – were calculated as a part of R&D studies at Brainware Analytics GmbH (Schlieren, Zurich) in years 2013-2015.

2 Additional computed data (optional)

Code	CAS	fundamentals (cm ⁻¹)	description
CON	502-49-8		
DMI	80-73-9		
FAH	50-0-0		
MLA	547-64-8		
PCD	125132-75-4		
PYR	110-86-1		
THF	109-99-9		
THT	110-01-0		
TPH	434-45-7		
TFE	75-89-8		

2.1 OH-stretching fundamentals for test set

2.2 IR intensities and Raman scattering activities

No further information has been made available.

2.3 Isotopolog information

No further information has been made available.

2.4 Relative energies for local minima and spectral properties

No further information has been made available.

2.5 Other computed quantities for the training and test sets

No further information has been made available.