PCCP

Accepted Manuscript

ROYAL SOCIETY
OF CHEMISTRY

www.rsc.org/pccp

Reliable Vibrational Wavenumbers for C=O and N-H Stretchings of Isolated and Hydrogen-bonded Nucleic Acid Bases

Teresa Fornaro[a], **Malgorzata Biczysko***[b], **Julien Bloino**[c], **Vincenzo Barone***[a]

Received Xth XXXXXXXXXX 20XX, Accepted Xth XXXXXXXXX 20XX
First published on the web Xth XXXXXXXXXX 200X
DOI: XXX

The accurate prediction of vibrational wavenumbers for functional groups involved in hydrogen-bonded bridges remains an important challenge for computational spectroscopy. For the specific case of the C=O and N-H stretching modes of nucleobases and their oligomers, the paucity of experimental reference values, needs to be compensated by reliable computational data, which require the use of approaches going beyond the standard harmonic oscillator model. Test computations performed for model systems (formamide, acetamide and their cyclic homodimers) in the framework of the second order vibrational perturbation theory (VPT2) confirmed that anharmonic corrections can be safely computed by global hybrid (GHF) or double hybrid (DHF) functionals, whereas the harmonic part is particularly challenging. As a matter of fact, GHFs perform quite poorly and even DHFs, while fully satisfactory for C=O stretchings, face unexpected difficulties when dealing with N-H stretchings. On these grounds, a linear regression for N-H stretchings has been obtained and validated for the heterodimers formed by 4-aminopyrimidine with 6-methyl-4-pyrimidinone (4APM-M4PMN) and by uracil with water. In view of the good performance of this computational model, we have then built a training set of B2PLYP-D3/maug-cc-pVTZ harmonic wavenumbers (including linear regression scaling for N-H) for six-different uracil dimers, and a validation set including, together with 4APM-M4PMN, one of the most stable hydrogen-bonded adenine homodimers, as well as the adenine-uracil, adenine-thymine, guanine-cytosine and adenine-4-thiouracil heterodimers. Because of the unfavourable scaling of DHF harmonic wavenumbers with the dimensions of the investigated systems, we have optimized a linear regression of B3LYP-D3/N07D harmonic wavenumbers for the training set, which has been next checked against the validation set. This relatively cheap model, which shown very good agreement with experimental data (average errors of about 10 cm^{-1}), paves the route toward the reliable analysis of spectroscopic signatures for larger polynucleotides.

1 Introduction

Spectroscopic approaches have become the methods of choice for analyzing structural, dynamical, and environmental effects for molecular systems of increasing complexity in a very effective yet non-invasive way[1–7], and this is especially true for biomolecules and their building blocks, whose physical-chemical characteristics are tuned by several weak interactions of comparable strengths[8–17]. In particular, the spectroscopic fingerprints in the ground and excited electronic states, studied by variety of experiments (i.e IR, UV, IR-IR, IR-UV) are often related to the stretching vibrations ν(C=O) and ν(N-H)[16,18–27], which in turn can be involved in N-H$\cdots$O hydrogen bonds (HB) and other inter- or intra-molecular interactions. Indeed, shifts as large as hundreds of wavenumbers are sometimes observed for bands related to functional groups directly involved in hydrogen-bonded bridges,[28–35] so that interpretation of the experimental outcomes might become questionable.[32,34,35]. This is particularly true for nucleobase dimers and oligomers, in view of their importance in prebiotic chemistry.[33,36–39] Let us recall, in this connection, that supramolecular noncovalent assemblies of nucleobases, or plausible building blocks of pre-RNAs, have been proposed as key intermediates along the evolutionary pathway toward contemporary nucleic acids.[40,41] Unfortunately, unambiguous experimental frequencies are scarce for nucleobase dimers involving inter-molecular hydrogen bonds. Under such circumstances, reliable theoretical descriptions of spectroscopic fea-

---

† Electronic Supplementary Information (ESI) available: (*i*) Cartesian coordinates and structural parameters of the optimized structures of the hydrogen-bonded nucleobases dimers, computed at B3LYP-D3/N07D and B2PLYP-D3/maug-cc-pVTZ levels of theory; (*ii*) Harmonic vibrational wavenumbers for nucleobases complexes, computed at B2PLYP-D3/maug-cc-pVTZ and B3LYP-D3/N07D levels of theory; (*iii*) Comparison between B2PLYP-D3/maug-cc-pVTZ harmonic wavenumbers and B3LYP-D3/N07D ones for the uracil and 2-thiouracil monomers corrected using the linear regression parameters. See DOI: 10.1039/b000000x/

[a] Scuola Normale Superiore, piazza dei Cavalieri 7, I-56126 Pisa, Italy; e-mail: vincenzo.barone@sns.it
[b] International Centre for Quantum and Molecular Structures, College of Sciences, Shanghai University, 99 Shangda Road, Shanghai, 200444 China; e-mail: biczysko@shu.edu.cn
[c] Consiglio Nazionale delle Ricerche, Istituto di Chimica dei Composti OrganoMetallici (ICCOM-CNR), UOS di Pisa, Area della Ricerca CNR, Via G. Moruzzi 1, I-56124 Pisa, Italy

tures and, especially, of their trends, which cover the most important vibrational fingerprints would be highly desirable.

In the last few years, computational spectroscopy has extended its range of application from small semi-rigid molecules in the gas phase to larger molecular systems of current biological and/or technological interest in their natural environment.[3,42–47] Nevertheless, the accurate prediction of vibrational wavenumbers of molecular systems governed by weak intermolecular interactions remains a challenge, due to the difficulty of treating dispersion interactions[27,32] and to the enhanced accuracy requirements of the potential energy surface (PES) for hydrogen-bonded bridges[48]. Several studies have shown that second-order vibrational perturbation theory (VPT2)[49–52] in conjunction with PESs obtained by global hybrid functionals (GHFs) or, even better, double hybrid functionals (DHFs), perform a remarkable job for several molecular systems,[46,53–61] and also for their supra-molecular analogues[32,34,35,62] when complemented by empirical description of dispersion forces (e.g. the D3[63,64] model proposed by Grimme). Another remarkable conclusion of those studies, further confirmed by computations of small model systems in the present work, is that reliable, and essentially converged, anharmonic corrections are obtained by some GHFs (here B3LYP[65]) in conjunction with polarized double-ζ basis sets augmented by diffuse functions (here N07D[66–68]). The situation is more involved for the harmonic part, which requires more refined computational models. In this connection, remarkable results are usually delivered by some DHFs (here B2PLYP[69,70]) in conjunction with polarized triple-ζ basis set augmented by at least s,p diffuse functions.

Recently, we have started a comprehensive research project aimed to the accurate analysis and prediction of fully anharmonic spectra of nucleobases and their oligomers of increasing size and complexity[32,35], through a multi-step strategy. Starting from the spectroscopic properties of the isolated molecule, the investigation proceeded with the analysis of systems of increasing complexity, namely dimers, oligomers, micro-solvated species and, finally, condensed phases like the molecular solid[34]. It has been shown that fully anharmonic simulations are required to obtain realistic spectral line-shapes, which are strongly influenced by different intra- and inter-molecular interactions.[32,35] Moreover, the B3LYP-D3[63,64] computations led to remarkable improvements on the accuracy of structural parameters and binding energies for systems involving both stacking interactions and hydrogen bonds, providing at the same time accurate anharmonic frequencies.[32,33,46,62,71–73] However, some larger discrepancies with respect to experiment are still observed for the X-H and C=X (X=O, N) stretching frequencies[32,61], which can be systematically amended by hybrid computations[46,74,75] with harmonic frequencies corrected at higher-level of theory (Coupled Cluster (CC) or B2PLYP) with basis sets of at least triple-ζ quality[17,46,59]. The B2PLYP/B3LYP-D3 model has been applied for uracil-water complexes and hydrogen-bonded uracil dimers[35]. For uracil-water complexes, this method reproduces well also the vibrations involved in hydrogen bonds ($\nu(O-H)$ and $\nu(N-H)$), with average and maximum uncertainties in the spectral region of hydrogen-bonding interactions of about 10 cm^{-1} and 20 cm^{-1}, respectively, i.e. nearly halved with respect to B3LYP-D3. With the aim of developing effective protocols for the study of larger and more complex systems, within hybrid schemes, another approach has also been considered to improve the description of harmonic frequencies for modes involved in hydrogen-bonding interactions through less demanding ONIOM B2PLYP:B3LYP computations, where only the part of the molecular system forming the hydrogen bonds is treated at the B2PLYP level of theory, whereas the remaining part is described by the B3LYP functional.[35] Full and ”ONIOM” hybrid B2PLYP/B3LYP-D3 computational models yield very similar results (average uncertainties of about 10 cm^{-1}) for uracil-water complexes.[35] However, the focused schemes would become less effective for larger molecular systems, with several inter- and intra-molecular interactions, so more general approaches are required.

The difficulties experienced in predicting reliable wavenumbers for C=O and N-H functional groups involved in hydrogen bonds, which show significant red-shifts for the stretching vibrational modes and analogous blue-shifts for the out-of-plane vibrational modes[22,23], motivated us to explore in this work the ability of DHFs to deal with the vibrational wavenumbers of those moieties, either isolated or involved in hydrogen bonds within a target accuracy of about 10 cm^{-1} and 30 cm^{-1} in terms of average errors and maximum discrepancies, respectively. On the basis of previous works and in order to obtain broader applicability to the molecular systems governed by various weak inter- and intra-molecular interactions[73,76,77], we have chosen to use B2PLYP-D3 in conjunction with maug-cc-pVTZ[78] basis set. The results obtained using this method for some model systems are fully satisfactory for any kind of C=O group and for isolated N-H moieties. However, reduced accuracy of computed frequencies is observed for the N-H groups involved in hydrogen bonds, which can be improved only upon scaling by mean of linear equations. The two parameters defined in this way have been successfully validated for the heterodimers formed by 4-aminopyrimidine with 6-methyl-4-pyrimidinone (4APM-M4PMN) and by uracil with water. On these grounds, we have built a training set of B2PLYP-D3/maug-cc-pVTZ harmonic wavenumbers (including the linear regression scaling for N-H) for six-different uracil dimers (members of the T6 set) and a validation set including, together with 4APM-M4PMN, one of the most stable hydrogen-bonded adenine homodimers as well as with the adenine-uracil, adenine-thymine, guanine-cytosine and adenine-4-thiouracil heterodimers (V6 set). These molecular systems show different bonding patterns, but are all characterised by negligible proton-transfer effects, so that the VPT2 approach is expected to provide a correct description of the localized anharmonic PES. Due to the unfavourable scaling of DHF harmonic wavenumbers (in conjunction with rather large basis sets) for the bigger systems of current scientific and technological interest, we have optimized also linear regressions for C=O and N-H harmonic wavenumbers computed by GHFs and polarized double-ζ basis sets for the T6 set, which were next validated on the V6 set. The good performance of this relatively cheap model (average errors of about 10 cm^{-1}) paves the route toward the reliable analysis of spectroscopic signatures for larger polynucleotides.

2 Computational Details

All equilibrium geometries and harmonic force fields have been obtained by means of analytical gradients and Hessians of the B3LYP[65] global hybrid functional and the B2PLYP[55,69,70] double-hybrid functional, including Grimme's semi-empirical dispersion corrections (D3[63,64]). The B3LYP-D3 computations have been performed in conjunction with the double-ζ N07D basis set[66–68,79], which has been extensively validated for several spectroscopic observables[34,35,80,81]. The reference B2PLYP-D3 computations have been performed with the maug-cc-pVTZ basis set[78] (in which *d* functions on hydrogens have been removed), in view of previous experience about increased basis-set requirements[55–57,60,82,83]. The B3LYP-D3/N07D and B2PLYP-D3/maug-cc-pVTZ computational models will be simply labelled B3D3 and B2D3 in the following. Fully anharmonic vibrational computations have been performed within the VPT2 framework[49–51,84–89], using cubic and semi-diagonal quartic force fields obtained by numerical differentiation of the analytical second derivatives along each active normal coordinate (with the standard 0.01 Å step) at geometries optimized with tight convergence criteria. Fermi and Darling-Dennison resonances have been treated within the generalized VPT2 scheme (GVPT2), where nearly-resonant contributions are removed from the perturbative treatment (leading to the deperturbed model, DVPT2) and variationally treated in a second step.[51,85,90,91] This model[51,84,85], as implemented in the GAUSSIAN package,[92] has provided accurate vibrational wavenumbers for several semi-rigid systems (see for instance Ref.[46] and references therein). Anharmonic computations have been performed also with hybrid models[46,55,57,74,75,93,94], in which the harmonic part is computed at the B2D3 level and the anharmonic corrections are evaluated by the less expensive B3D3 approach. In this work we have adopted the simplest *a posteriori* hybrid approach in which anharmonic corrections are added to the best estimated harmonic results (ω_{B2D3}+(ν_{B3D3}-ω_{B3D3})). The resulting hybrid approach will be shortly labeled B2D3/B3D3 in the following.

The B2D3 model has been validated with reference to the experimental vibrational wavenumbers of 2-thiouracil, acetamide, formamide and the hydrogen-bonded cyclic homodimers of the last two molecules. In order to improve the harmonic wavenumbers of N-H stretchings involved in hydrogen bonds (without worsening those of isolated N-H moieties), the following linear scaling formula was used,

$$\omega_{\mathrm{B2D3NH}} = k_{\mathrm{B2}} * \omega_{\mathrm{B2D3}} + \Delta_{\mathrm{B2}}$$

The parameters, k_{B2} and Δ_{B2}, were fitted by solving the equation,

$$\omega_{\mathrm{exp}} = \nu_{\mathrm{exp}} - (\nu_{\mathrm{B2D3}} - \omega_{\mathrm{B2D3}}) = k_{\mathrm{B2}} * \omega_{\mathrm{B2D3}} + \Delta_{\mathrm{B2}}$$

where ν_{exp} and ν_{B2D3} are reference experimental and anharmonic wavenumbers, respectively. The model, labeled B2D3NH in the following, has been validated by comparison with the available experimental N-H stretching wavenumbers for the 4-aminopyrimidine-6-methyl-4-pyrimidinone (4APM-M4PMN) hydrogen-bonded complex, as well as for four 1:1 uracil-water complexes[25,95]. The B2D3NH model has been further applied in order to derive best theoretical estimates/reference data for ω(C=O) and ω(N-H) of larger complexes.

Next, two linear scaling equations (one for C=O and one for N-H stretchings) were derived to improve the accuracy of B3D3,

$$\omega_{\mathrm{B3D3C}} = k_{\mathrm{B3}} * \omega_{\mathrm{B3D3}} + \Delta_{\mathrm{B3}}$$

where the two pairs of k_{B3} and Δ_{B3} parameters (one for C=O and one for N-H) have been fitted by comparison with the harmonic wavenumbers of a training set (hereafter T6) of six hydrogen-bonded uracil dimers computed at the B2D3NH level. The resulting B3D3C model has been then validated with reference to B2D3NH harmonic wavenumbers computed for a validation set (hereafter V6) including the adenine homodimer (A-A), the adenine-uracil (A-U), adenine-thymine (A-T), guanine-cytosine (G-C), adenine-4-thiouracil (A-ThioU) and 4APM-M4PMN (mimicking the adenine-thymine dimer) heterodimers, which are characterized by diverse hydrogen bond patterns and quite different wavenumbers of the corresponding vibrational modes.

The structures and numbering schemes for the nucleobase complexes used in the T6 and V6 sets are shown in Figure 2, while the corresponding Cartesian coordinates are reported in the supplementary material.

All calculations have been carried out employing the GAUSSIAN suite of programs[92]. Assignments of vibrational modes were performed by means of visual inspection of the atomic displacements along normal modes and by comparison with the assignments reported in the literature. A graphical user interface (VMS-Draw)[96] has been used to visualize normal modes and analyze in detail the outcome of vibrational computations.

3 Results and discussion

3.1 Validation of the B2D3 and B2D3/B3D3 Models: 2-Thiouracil and the Dimers of Acetamide and Formamide

Harmonic and anharmonic vibrational wavenumbers and IR intensities of 2-thiouracil computed with the B2D3, B3D3, and hybrid B2D3/B3D3 models in the framework of the GVPT2 approach are compared to their experimental counterparts[53,97] in Table 1, where data relative to ν(N-H) and ν(C=O) are marked in bold. B2D3 and B3D3 harmonic wavenumbers are quite close (mean absolute error (MAE) of about 5 cm^{-1}) for all modes except ν(N3H), γ(CH) and ν(C4=O), whose discrepancies reach 10-20 cm^{-1}. As expected, B2D3 and B3D3 anharmonic corrections are essentially identical also for C-H, N-H, and C=O stretchings (MAE=2 cm^{-1}), the only discrepancies larger than 5 cm^{-1} being due to the different definition of Fermi resonances for the ν(C5H) and ν(C6H) modes. Comparison with experiment shows that the B2D3 computational model is able to improve the harmonic part of the vibrational wavenumbers for ν(C=O) and (to a lower extent) for ν(N-H). Since the anharmonic correction is the most demanding part of the calculations, it is more advantageous to adopt an hybrid scheme, where this correction is computed at a lower level of theory like B3D3. In this way, it is possible to obtain very good agreement with experiment both for ν(C=O) and ν(N-H), with mean absolute errors of about 8 cm^{-1} and 2 cm^{-1}, respectively, instead of about 14 cm^{-1} and 13 cm^{-1} in the case of full harmonic/anharmonic B3D3 computations. This further confirms the accuracy of the anharmonic shifts computed with the B3D3 method, and points out the importance of using more accurate harmonic wavenumbers (e.g B2D3) for quantitative comparison with experiments. On the whole, the MAEs obtained for the anharmonic wavenumbers computed at B2D3 and hybrid B2D3/B3D3 levels of theory are very small, i.e. 6.4 cm^{-1} and 5.5 cm^{-1}, respectively, slightly better than the full B3D3 model (7.5 cm^{-1}). Moreover, the intensities computed at both the B2D3 and B3D3 levels reproduce well the trends observed in the experimental IR spectrum. Those results confirm that B2D3 can be generally considered a reliable method for the prediction of vibrational properties and, additionally, shows better performance with respect to B3D3 for C=O and N-H stretching, which are also the ones mainly influenced by hydrogen-bonding interactions.

Following this initial analysis, the B2D3 and hybrid B2D3/B3D3 models have been tested with respect to the vibrational properties of other model systems, acetamide and formamide and their cyclic dimers featuring hydrogen-bonding interactions, focusing in particular on the ν(C=O) and ν(N-H) vibrations (see Tables 2 and 3). It is apparent that both B2D3 and hybrid B2D3/B3D3 computations provide fully satisfactory results for ν(C=O), with MAE of 5.4 cm^{-1} and 7.5 cm^{-1}, and maximum deviations of 10 cm^{-1} and 13 cm^{-1}, respectively. The situation is different for ν(N-H), which shows an overall MAE of 10.9 cm^{-1} and 12.2 cm^{-1}, respectively, but with larger maximum deviations over 40 cm^{-1} for the modes affected by hydrogen bonding (ν(N-H)s-as of both dimers). The unusually large discrepancies observed for N-H stretches involved in hydrogen bonds could be related to the parameterisation of dispersion correction (D3) as well as to intrinsic inaccuracies of the VPT2 treatment. With respect to the latter, we note that all systems studied in this work show only slight elongation of the N-H distances involved in HB interactions, so that the VPT2 approach is expected to provide a correct description of the underlying PES. Both test on the accuracy of the B2D3 PES along the hydrogen bonded bridge coordinate, and comparisons with variational vibrational computations are deferred to further studies. However, it has been shown for instance for $Tc-$pyruvic acid, featuring an intramolecular O-H$\cdots$O hydrogen bond, that for ν(O-H)$_{\mathrm{HB}}$, the discrepancies of B2D3 anharmonic wavenumbers with respect to experiment can be fully related to the harmonic part, at least for the fundamental transition[98]. Moreover, by scaling the B2D3 harmonic wavenumbers, it was possible to obtain very good agreement with experiment, on par with the best theoretical estimates employing Coupled Cluster computations and extrapolations to the complete basis set limit[98].

Although the error of B2D3 harmonic wavenumbers is not dramatic and the general trends are always reproduced, in order to improve the overall accuracy, we tried to introduce some scaling of the harmonic wavenumbers. However, since isolated N-H stretchings are well determined, a constant scaling is not the best solution, and linear scaling equations should be preferred. To this end, in this work, harmonic B2D3 wavenumbers have been fitted to “harmonic experimental” values (ω_{exp}) obtained by subtracting B2D3 anharmonic corrections from the experimental fundamental transitions obtained in gas phase measurements for all the systems reported in Table 3. Figure 3 shows a remarkable linear correlation between these two sets of data, which can be expressed as $\omega_{\mathrm{exp}} = k_{\mathrm{B2}} * \omega_{\mathrm{B2D3}} + \Delta_{\mathrm{B2}}$, with k_{B2}=0.93±0.03, Δ_{B2}=(2.9±0.9)*10^2 cm^{-1} and correlation coefficient R^2=0.9952. Such corrected harmonic wavenumbers (referred to as B2D3NH) have been applied into hybrid models with anharmonic corrections computed at the B2D3 and B3D3 levels. The results are reported in the last two columns of Table 3, showing significant improvement with MAE of 9.6 cm^{-1} and 7.6 cm^{-1} and maximum discrepancies below 20 cm^{-1}, which are well within our target accuracy.

3.2 Validation of B2D3NH model: 4APM-M4PMN and uracil-water complexes

The B2D3NH model has been further validated by comparison with available experimental data for two sets of systems, a model complex mimicking the adenine-thymine Watson-Crick pair, i.e. the 4-aminopyrimidine with 6-methyl-4-pyrimidinone[25], and four uracil-water 1:1 complexes[95]. It is worth noting that, although there are several studies devoted to vibrational spectra of isolated hydrogen-bonded complexes of biomolecule building blocks, in our opinion most of them are not adequate for validation purposes due to possible bias in the spectral analyses, which have been performed with the support of insufficiently accurate computational models. For example, in a recent study the IR photodissociation action spectra of adenine dimer hydrates[99] have been used to determine their structures by comparing the experimental results with scaled harmonic computations at M06-2X/6-311++G(2d) level. Unfortunately, the conclusions are not without ambiguities since it has been shown that this computational level is inadequate for the prediction of vibrational features of isolated adenine[72]. The same remark applies to a previous analysis of IR-UV double-resonance spectra of adenine dimers[19], in the N-H stretching frequency range, which has been performed with the aid of harmonic computations at the HF/6-31G(d,p) level, excluding the most stable symmetric structure due to the disagreement between computed and experimental spectra. Our experience suggests that improvement in theoretical methodology might eliminate such discrepancies, leading sometimes to the re-analysis of experimental data[81,100–103]. The situation is especially critical for weakly-bonded systems where it is necessary to combine structure identification of the complexes (for instance by means of rotational spectra[10,11,104]) with analysis of the vibrational transitions. In this respect both systems chosen for validation purposes provide unequivocal reference data. For 4APM-M4PMN this is related to the specific setup of Adenine-Thymine mimicking molecular system, for which the Watson-Crick type bonding is strongly preferred and no other (differently bonded) complexes are expected in the experimental mixture[25]. For the uracil-water complexes[95] the analysis of the infrared spectra has been supported by measurements of vibrational transition moment angles for various vibrational modes. These measurements allowed to identify the formation of four structural isomers of the uracil-water binary complex in helium nanodroplets, and to assign all ν(N-H) and ν(O-H) vibrations.

The accuracy of theoretical anharmonic wavenumbers has been assessed employing a database of ν(N-H) experimental wavenumbers obtained as discussed above and reported in Table 4. The database involves both free and hydrogen-bonded cases, and spans the rather large interval ranging between 2750 and 3550 cm^{-1}. Anharmonic corrections have been always obtained from GVPT2 computations at the B3D3 level, so that the results differ by the harmonic part only. As observed for smaller systems, the B2D3/B3D3 model shows excessive discrepancies with respect to experiment for ν(N-H)$_{\mathrm{HB}}$, with MAE and |MAX| of 31 cm^{-1} and 70 cm^{-1}, respectively, leading to overall MAE of 21 cm^{-1}. Significant improvement, for all anharmonic wavenumbers, is obtained with the hybrid B2D3NH/B3D3 scheme showing MAE and |MAX| of 11 cm^{-1} and 36 cm^{-1}, respectively. It is noteworthy that B2D3NH not only improves significantly ν(N-H)$_{\mathrm{HB}}$, lowering maximum discrepancies by about 35 cm^{-1}, but also refines the values of ν(N-H)$_{\mathrm{free}}$, leading to very good agreement within 1-6 cm^{-1}. On these grounds the B2D3NH model will be applied to obtain our best theoretical estimates for the harmonic wavenumbers of systems from the T6 and V6 benchmark sets.

3.3 B2D3NH harmonic wavenumbers for ν(C=O) and ν(N-H) and B3D3C model: the Nucleobase Dimers

In order to extend accurate anharmonic computations to larger systems we recall that the B2D3 harmonic wavenumbers are often more computationally demanding (mainly due to the increased memory requirements) than the evaluation of anharmonic corrections at the B3D3 level. Moreover, anharmonic corrections within reduced dimensionality schemes[46,105,106] can be quite routinely computed also for rather large systems like, for instance, uracil heptamer[34], chlorophyll[105,107], glycine adsorbed on silicon cluster[106] or metal-carbonyl-diimine complexes[108]. All these systems feature spectral fingerprints due to the carbonyl and/or amino groups, clearly showing that it would be highly desirable to correct ν(C=O) and ν(N-H) harmonic wavenumbers, also when B2D3 computations become unfeasible.

In order to build reference databases of best theoretical estimates for harmonic wavenumbers of nucleobase complexes, we have performed B2D3NH computations for a number of representative systems. As a first step, a training set (T6) of six different hydrogen-bonded uracil dimers, referred to as U-Un (n=1,2,3,4,5,6) in Figure 2, has been set up. The T6 set has been then used to derive linear scaling equations for the B3D3 wavenumbers of ν(C=O) and ν(N-H), based on the reference data from B2D3NH. The harmonic vibrational wavenumbers for these dimers, computed at the B2D3NH and B3D3 levels, are reported in the Supplementary Material, along with the assignment of the vibrational modes in the high-frequency region. These data are also compared in Figures 4 and 5, for ν(C=O) and ν(N-H), respectively. In both cases, the B2D3NH and B3D3 results show rather good linear dependence and, considering all data within the linear regression of the form $\omega_{\mathrm{B2D3NH}} = k_{\mathrm{B3}} * \omega_{\mathrm{B3D3}} + \Delta_{\mathrm{B3}}$, satisfactory fits have been obtained with coefficients of correlation,

R^2=0.996 and 0.999 for ν(C=O) and ν(N-H), respectively. In this way, it has been possible to define the correction parameters, k_{B3} and Δ_{B3}, for the B3D3 harmonic wavenumbers of ν(C=O) and ν(N-H), based on the more accurate B2D3NH ones as: $k_{\mathrm{B3}}^{\mathrm{C=O}}$=0.92±0.01 and $\Delta_{\mathrm{B3}}^{\mathrm{C=O}}$=(1.3±0.2)*$10^2$ cm^{-1}, $k_{\mathrm{B3}}^{\mathrm{N\text{-}H}}$=0.865±0.006 and $\Delta_{\mathrm{B3}}^{\mathrm{N\text{-}H}}$=(5.0±0.2)*$10^2$ cm^{-1}.

These parameters have been then applied to correct the B3D3 harmonic wavenumbers for ν(C=O) and ν(N-H) of a validation set (V6) of other homo- and hetero-dimers of nucleobases. The V6 harmonic vibrational wavenumbers are reported in the Supplementary Material, along with the assignment of the vibrational modes in the high-frequency region. All B2D3NH and B3D3 results for the validation set are also depicted in Figures 4 and 5, while the comparison between the uncorrected and corrected (C) B3D3 harmonic wavenumbers and the reference B2D3NH ones for ν(C=O) and ν(N-H) are reported in Tables 5 and 6, respectively. First of all, it can be observed that all the data from the validation set fall very close to the straight line corresponding to the regressions derived for the training set. This finding indicates that the correction parameters for B3D3 harmonic wavenumbers defined for ν(C=O) and ν(N-H) using as training set the hydrogen-bonded uracil dimers, can be transferred to improve both ν(C=O) and ν(N-H) of the other nucleobase dimers of the V6 set, even though they are characterized by different hydrogen bond patterns and quite different vibrational frequencies. Indeed, the correction parameters have been derived from hydrogen-bonded systems where only the N-H$\cdots$O=C hydrogen bond type is present. The validation set, instead, features very different hydrogen bond types, specifically $NH_2\cdots$N, N-H$\cdots$N and $NH_2\cdots$O=C. It is also noteworthy that the ν(N-H) harmonic wavenumbers from the T6 set cover the range 3200-3600 cm^{-1}, while the values from the V6 set span the much larger 2800-3800 cm^{-1} interval. Moreover, the fits derived from the training set alone and from the whole T6+V6 data agree with each other within the reported error bars. This finding strongly suggests a broad range of applicability for the proposed linear scaling equation parameters, which should hopefully cover any hydrogen-bonded system involving C=O and N-H functional groups with localized anharmonic character, independently of the partner in the hydrogen-bonded bridge.

In more detail, in the case of the A-A dimer, which presents two $NH_2\cdots$N hydrogen bonds, after correction the differences between B3D3C and B2D3NH harmonic wavenumbers are all lower than 20 cm^{-1}. For the $\nu_{\mathrm{asym}}(NH_2)$ and the ν(N9H) both corrected and uncorrected values agree within about 5 cm^{-1}, while a significant improvement (19 cm^{-1} with respect to 46 cm^{-1}) is observed for $\nu_{\mathrm{sym}}(NH_2)$.

In the case of the A-U heterodimer, featuring one $NH_2\cdots$O=C and one N-H$\cdots$N hydrogen bonds, all B3D3C wavenumbers are within 13 cm^{-1} from the reference B2D3NH values, while uncorrected B3D3 show differences reaching 60 cm^{-1}. The major improvements are observed for the harmonic wavenumber of $\nu_{\mathrm{sym}}(NH_2)$ (by about 14 cm^{-1}) and, mostly, uracil ν(N3H) is improved by about 48 cm^{-1}. Regarding the carbonyl stretching vibrational modes of uracil (ν(C2O) and ν(C4O)), involved in the hydrogen bonds within the dimer, the correction allows to improve both of them, with discrepancies from B2D3NH reference values, of -3 cm^{-1} and 1 cm^{-1}, respectively, instead of 15 cm^{-1} and 13 cm^{-1} for non-corrected B3D3.

For A-ThioU, A-T and 4APM-M4PMN the situation is analogous to that of A-U, with all discrepancies lower than 13 cm^{-1} for B3D3C values, and major improvements for N-H wavenumbers. Adenine and 4APM $\nu_{\mathrm{sym}}(NH_2)$, and mostly, thiouracil, thymine and M4PMN ν(N3H) harmonic wavenumbers undergo a remarkable improvement of about 13-21 cm^{-1} and 38-46 cm^{-1}, respectively, reaching values very close to the corresponding B2D3NH ones, with a difference of only -3 cm^{-1} in the case of thiouracil. The ν(C=O) vibration of thiouracil, thymine and M4PMN are improved as well, with discrepancies below 3 cm^{-1} instead of 10-18 cm^{-1} for non-corrected B3D3.

In the case of G-C, characterized by two $NH_2\cdots$O=C and one N-H$\cdots$N hydrogen bonds, B3D3C values agree with the B2D3NH reference within 10 cm^{-1}. The harmonic wavenumber of $\nu_{\mathrm{sym}}(NH_2)$ for guanine is improved by about 14 cm^{-1}, while even larger corrections have been observed for intermolecular modes within the hydrogen-bonded bridges, i.e. the simultaneous stretching of N1H, NH_2 of guanine and $\nu(NH_2)$ of cytosine, whose discrepancy decreases by 28 cm^{-1}, the simultaneous stretching of NH_2 of cytosine and N1H of guanine, whose discrepancy decreases by 34 cm^{-1}, and the simultaneous stretching of C6O of guanine and C2O of cytosine, with a difference of -3 cm^{-1} compared to B2D3NH, instead of 11 cm^{-1} for the non-corrected B3D3.

Given the good performance of such linear scaling equations for improving, beyond B3D3, the harmonic part of vibrational wavenumbers for ν(C=O) and ν(N-H) involved in hydrogen bonds, we have also verified their behavior for isolated uracil and 2-thiouracil monomers (see Supplementary Material). In both cases, the discrepancies of the ν(C=O) with respect to B2D3 decrease significantly upon scaling, while minor improvements are obtained for ν(N-H). Specifically, for uracil the B3D3C values for ν(C2O) and ν(C4O) agree with their B2D3 counterparts within 1-3 cm^{-1}, instead of 17-18 cm^{-1} for non-corrected B3D3 results, while for 2-thiouracil the discrepancy of 22 cm^{-1} for ν(C4O) is lowered to 6 cm^{-1}.

Finally the B3D3 and B3D3C/B3D3 ν(N-H) anharmonic wavenumbers have been compared with experimental data for the hydrogen bonded complexes considered previously, namely the formamide and acetamide dimers, 4APM-M4PMN and U-Wn. From the data reported in Table 7, it is clear that the B3D3 results are significantly improved with the B3D3C correction, reducing MAE and maximum errors by a factor of 3, from 34 cm^{-1} to 12 cm^{-1} and from 100 cm^{-1} to 37 cm^{-1}. As expected, B3D3 shows significantly larger errors for the vibrations involved in the hydrogen bonds, with MAE of 54 cm^{-1}, while the usual rather good accuracy of B3D3 is retained for ν(N-H)$_{\mathrm{free}}$ with MAE and $|$MAX$|$ of 11 cm^{-1} and 18 cm^{-1}, respectively, in line with benchmark studies for isolated molecules[32,46]. At variance, the B3D3C/B3D3 results show MAE of 11 cm^{-1}, 17 cm^{-1} and 7 cm^{-1}, on overall, for ν(N-H)$_{\mathrm{HB}}$ and ν(N-H)$_{\mathrm{free}}$, respectively. Moreover, the B3D3C/B3D3 results show an accuracy essentially on par with the B2D3NH/B3D3 one, while the B3D3C/B3D3 model requires only B3LYP-D3/N07D anharmonic computations, so it is routinely applicable for systems containing up to 50 atoms within the full-dimensionality approach and even larger by means of reduced-dimensionality GVPT2 computations.

In conclusion, we strongly believe that the proposed linear scaling equations perform a remarkable job in improving B3D3 harmonic frequencies, leading to values very close to the B2D3NH data for both ν(C=O) and ν(N-H). It can be forecast that the same scaling approach is able to improve B3D3 stretching frequencies for C=O and N-H groups involved in any kind of hydrogen bonds, as well as for isolated molecules including such moieties as far as vibrations show rather localized character.

4 Conclusions

The description of vibrational properties for functional groups like C=O and N-H, in particular when involved in hydrogen-bond networks, remains a challenge for computational spectroscopy, due to the strong anharmonicity of the corresponding stretching vibrational modes which requires methods going beyond the standard harmonic approximation and particularly accurate potential energy surfaces. Furthermore, for the case of nucleic acid bases, very limited experimental information is available about the perturbing effect of inter-molecular hydrogen bonds on vibrational frequencies. Test computations on small model systems (thiouracil, formamide, acetamide and their cyclic homodimers) have confirmed that, within the framework of VPT2, reliable anharmonic contributions can be safely computed at the B3LYP-D3/N07D level, which is cheap enough to allow the study of large systems, especially in connection with reduced dimensionality approaches. The situation is more involved for harmonic contributions, where the B3D3 computational model is clearly insufficient and even the more computationally intensive and reliable B2PLYP-D3/maug-cc-pVTZ approach is not fully satisfactory for N-H stretchings. We have thus derived a linear regression correction for this last vibrational mode and shown that the resulting B2D3NH anharmonic vibrational wavenumbers are sufficiently accurate. An even more effective approach, referred to as B2D3NH/B3D3, retaining essentially the same accuracy, is obtained correcting B2D3NH harmonic wavenumbers by B3D3 anharmonic contributions. This approach has been next validated by comparison with the available experimental data for a model complex mimicking the adenine-thymine Watson-Crick pair, i.e. 4-aminopyrimidine-6-methyl-4-pyrimidinone and for some uracil-water 1:1 complexes, showing an overall agreement with experiment within 10 cm^{-1}. Unfortunately, the B2D3NH/B3D3 model can be unpractical for larger systems in view of the quite unfavourable scaling of B2D3 harmonic wavenumbers. On these grounds, we have built a B2D3NH training set of 6 uracil homodimers (T6) and an analogous validation set (V6) of six additional dimers (including, together with 4APM-M4PMN, one of the most stable hydrogen-bonded adenine homodimer, the nucleic acids base pairs adenine-uracil, adenine-thymine, guanine-cytosine and the adenine-4-thiouracil heterodimer), which should be useful for testing, possibly correcting (by linear scaling equations), and validating cheaper quantum mechanical approaches. As a first application of these sets, we have derived and validated linear regression corrections for both C=O and N-H harmonic wavenumbers computed at the B3D3 level. Our results point out a wide applicability of such parameters for both isolated and hydrogen-bonded nucleobases, thus paving the route toward the study of large nucleosides and nucleotides of current scientific and technological interest.

Acknowledgements

The research leading to these results has received funding from the European Union's Seventh Framework Programme (FP7/2007-2013) under grant agreement No. ERC-2012-AdG-320951-DREAMS and from Italian MIUR (PRIN 2012 "STAR: Spectroscopic and computational Techniques for Astrophysical and atmospheric Research"). M.B. and J.B. acknowledge support of COST CMTS-Action CM1405 (MOLIM: MOLecules In Motion).

References

1 *Frontiers of Molecular Spectroscopy*, ed. J. Laane, Elsevier B.V., 2008, p. 740.
2 *Vibrational Spectroscopy in Life Science*, ed. F. Siebert and P. Hildebrandt, Wiley-VCH Verlag GmbH and Co. KGaA, Weinheim, Germany, 2008.
3 *Comprehensive Chiroptical Spectroscopy: Instrumentation, Methodologies, and Theoretical Simulations*, ed. N. Berova, P. L. Polavarapu, K. Nakanishi and R. W.

Woody, John Wiley & Sons, Inc. Hoboken, New Jersey, 2012.

4 G. Magyarfalvi, G. Tarczay and E. Vass, *WIREs Comput. Mol. Sci.*, 2011, **1**, 403–425.

5 R. Fausto, L. Khriachtchev and P. Hamm, in *Conformational Changes in Cryogenic Matrices*, ed. L. Khriachtchev, Pan Stanford Publishing Pte. Ltd, Singapore, 2011.

6 C. Puzzarini and M. Biczysko, in *Structure Elucidation in Organic Chemistry*, ed. M. M. Cid and J. Bravo, Wiley-VCH Verlag GmbH & Co. KGaA, 2015, ch. Computational Spectroscopy Tools for Molecular Structure Analysis, pp. 27–64.

7 *Gas-Phase IR Spectroscopy and Structure of Biological Molecules*, ed. A. M. Rijs and J. Oomens, Springer International Publishing, 2015, vol. 364.

8 T. S. Zwier, *J. Phys. Chem. A*, 2006, **110**, 4133–4150.

9 M. S. de Vries and P. Hobza, *Annu. Rev. Phys. Chem.*, 2007, **58**, 585–612.

10 M. Biczysko, G. Piani, M. Pasquini, N. Schiccheri, G. Pietraperzia, M. Becucci, M. Pavone and V. Barone, *J. Chem. Phys.*, 2007, **127**, 144303.

11 G. Pietraperzia, M. Pasquini, N. Schiccheri, G. Piani, M. Becucci, E. Castellucci, M. Biczysko, J. Bloino and V. Barone, *J. Phys. Chem. A*, 2009, **113**, 14343–14351.

12 R. Iwaura, T. Iizawa, H. Minamikawa, M. Ohnishi-Kameyama and T. Shimizu, *Small*, 2010, **6**, 1131–1139.

13 E. Garand, M. Z. Kamrath, P. A. Jordan, A. B. Wolk, C. M. Leavitt, A. B. McCoy, S. J. Miller and M. A. Johnson, *Science*, 2012, **335**, 694–698.

14 C. Puzzarini, M. Biczysko, V. Barone, L. Largo, I. Peña, C. Cabezas and J. L. Alonso, *J. Phys. Chem. Lett.*, 2014, **5**, 534–540.

15 M. E. Sanz, C. Cabezas, S. Mata and J. L. Alonso, *J. Chem. Phys.*, 2014, **140**, 204308.

16 B. Yan, S. Jaeqx, W. J. van der Zande and A. M. Rijs, *Phys. Chem. Chem. Phys.*, 2014, **16**, 10770–10778.

17 E. E. Najbauer, G. Bazsó, R. Apóstolo, R. Fausto, M. Biczysko, V. Barone and G. Tarczay, *J. Phys. Chem. B*, 2015, **119**, 10496–10510.

18 E. Nir, C. Plützer, K. Kleinermanns and M. de Vries, *Eur. Phys. J. D*, 2002, **20**, 317–329.

19 C. Plutzer, I. Hunig and K. Kleinermanns, *Phys. Chem. Chem. Phys.*, 2003, **5**, 1158–1163.

20 R. Wu, and B. Brutschy, *J. Phys. Chem. A*, 2004, **108**, 9715–9720.

21 P. Ottiger, J. A. Frey, H.-M. Frey and S. Leutwyler, *J. Phys. Chem. A*, 2009, **113**, 5280–5288.

22 S. Jarmelo, I. Reva, P. Carey and R. Fausto, *Vibrat. Spectrosc.*, 2007, **43**, 395–404.

23 M. Rozenberg, G. Shoham, I. Reva and R. Fausto, *Phys. Chem. Chem. Phys.*, 2005, **7**, 2376–2383.

24 J. C. Dean, E. G. Buchanan and T. S. Zwier, *J. Am. Chem. Soc.*, 2012, **134**, 17186–17201.

25 Y. Nosenko, M. Kunitski, T. Stark, M. Gobel, P. Tarakeshwar and B. Brutschy, *Phys. Chem. Chem. Phys.*, 2013, **15**, 11520–11530.

26 C. M. Leavitt, K. B. Moore III, P. L. Raston, J. Agarwal, G. H. Moody, C. C. Shirley, H. F. Schaefer III and G. E. Douberly, *J. Phys. Chem. A*, 2014, **118**, 9692–9700.

27 A. Bouchet, M. Schutz, B. Chiavarino, M. Elisa Crestoni, S. Fornarini and O. Dopfer, *Phys. Chem. Chem. Phys.*, 2015, **17**, 25742–25754.

28 *Theoretical Treatments of Hydrogen Bonding.*, ed. H. Dušan, John Wiley & Sons, Chichester, UK, 1997.

29 J. E. Del Bene and M. J. T. Jordan, *Int. Rev. Phys. Chem.*, 1999, **18**, 119–162.

30 A. D. Buckingham, J. Del Bene and S. A. C. McDowell, *Chem. Phys. Lett.*, 2008, **463**, 1 – 10.

31 E. Arunan, G. R. Desiraju, R. A. Klein, J. Sadlej, S. Scheiner, I. Alkorta, D. C. Clary, R. H. Crabtree, J. J. Dannenberg, P. Hobza, H. G. Kjaergaard, A. C. Legon, B. Mennucci and D. J. Nesbitt, *Pure and Applied Chemistry*, 2011, **83**, 1619–1636.

32 T. Fornaro, M. Biczysko, S. Monti and V. Barone, *Phys. Chem. Chem. Phys.*, 2014, **16**, 10112–10128.

33 T. Fornaro and I. Carnimeo, *Reference Module in Chemistry, Molecular Sciences and Chemical Engineering*, 2014, DOI:10.1016/B978–0–12–409547–2.11025–X.

34 T. Fornaro, I. Carnimeo and M. Biczysko, *J. Phys. Chem. A*, 2015, **119**, 5313–5326.

35 T. Fornaro, D. Burini, M. Biczysko and V. Barone, *J. Phys. Chem. A*, 2015, **119**, 4224–4236.

36 S. Sivakova and S. J. Rowan, *Chem. Soc. Rev.*, 2005, **34**, 9–21.

37 T. Fornaro, J. R. Brucato, S. Branciamore and A. Pucci, *Intern. Journ. Astrob.*, 2013, **12**, 78–86.

38 T. Fornaro, J. R. Brucato, E. Pace, M. C. Guidi, S. Branciamore and A. Pucci, *Icarus*, 2013, **226**, 1068 – 1085.

39 V. Barone, M. Biczysko and C. Puzzarini, *Acc. Chem. Research*, 2015, **48**, 1413–1422.

40 N. V. Hud, B. J. Cafferty, R. Krishnamurthy and L. D. Williams, *Chemistry & biology*, 2013, **20**, 466–474.

41 B. J. Cafferty and N. V. Hud, *Current opinion in chemical biology*, 2014, **22**, 146–157.

42 P. Jensen and P. R. Bunker, *Computational Molecular Spectroscopy*, John Wiley and Sons Ltd, Chichester, UK, 2000.

43 *Computational Strategies for Spectroscopy, from Small Molecules to Nano Systems.*, ed. V. Barone, John Wiley

& Sons, Inc. Hoboken, New Jersey, 2011.

44 *Computational Spectroscopy: Methods, Experiments and Applications*, ed. J. Grunenberg, Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, Germany, 2010.

45 V. Barone, A. Baiardi, M. Biczysko, J. Bloino, C. Cappelli and F. Lipparini, *Phys. Chem. Chem. Phys.*, 2012, **14**, 12404–12422.

46 V. Barone, M. Biczysko and J. Bloino, *Phys. Chem. Chem. Phys.*, 2014, **16**, 1759–1787.

47 V. Barone, *WIREs Comput. Mol. Sci.*, 2016, DOI:10.1002/wcms.1238.

48 M. Biczysko and Z. Latajka, *J. Phys. Chem. A*, 2002, **106**, 3197–3201.

49 H. H. Nielsen, *Reviews of Modern Physics*, 1951, **23**, 90–136.

50 I. M. Mills, in *Molecular Spectroscopy: Modern Research*, ed. K. N. Rao and C. W. Mathews, Academic Press, New York, 1972, ch. Vibration-Rotation Structure in Asymmetric- and Symmetric-Top Molecules, pp. 115–140.

51 V. Barone, *J. Chem. Phys.*, 2005, **122**, 014108.

52 M. Piccardo, J. Bloino and V. Barone, *Int. J. Quantum Chem.*, 2015, **115**, 948–982.

53 V. Barone, G. Festa, A. Grandi, N. Rega and N. Sanna, *Chem. Phys. Lett.*, 2004, **388**, 279–283.

54 V. Barone, *Chem. Phys. Lett.*, 2004, **383**, 528–532.

55 M. Biczysko, P. Panek, G. Scalmani, J. Bloino and V. Barone, *J. Chem. Theory Comput.*, 2010, **6**, 2115–2125.

56 M. Biczysko, J. Bloino, G. Brancato, I. Cacelli, C. Cappelli, A. Ferretti, A. Lami, S. Monti, A. Pedone, G. Prampolini, C. Puzzarini, F. Santoro, F. Trani and G. Villani, *Theor. Chem. Acc.*, 2012, **131**, 1201/1–19.

57 I. Carnimeo, C. Puzzarini, N. Tasinato, P. Stoppa, A. P. Charmet, M. Biczysko, C. Cappelli and V. Barone, *J. Chem. Phys.*, 2013, **139**, 074310.

58 V. Barone, M. Biczysko, J. Bloino and C. Puzzarini, *J. Chem. Theory Comput.*, 2013, **9**, 1533–1547.

59 V. Barone, M. Biczysko, J. Bloino and C. Puzzarini, *Phys. Chem. Chem. Phys.*, 2013, **15**, 10094–10111.

60 V. Barone, M. Biczysko, J. Bloino and C. Puzzarini, *J. Chem. Phys.*, 2014, **141**, 034107/1–17.

61 F. Vazart, C. Latouche, P. Cimino and V. Barone, *J. Chem. Theory Comput.*, 2015, **11**, 4364–4369.

62 S. Thicoipe, P. Carbonniere and C. Pouchan, *Phys. Chem. Chem. Phys.*, 2013, **15**, 11646–11652.

63 S. Grimme, J. Antony, S. Ehrlich and H. Krieg, *J. Chem. Phys.*, 2010, **132**, 154104.

64 S. Grimme, *WIREs Comput. Mol. Sci.*, 2011, **1**, 211–228.

65 A. D. Becke, *J. Chem. Phys.*, 1993, **98**, 5648–5652.

66 *Double and triple-ζ basis sets of N07 family, are available in the Download section*, http://dreams.sns.it, (last visited: 1 February, 2015).

67 V. Barone and P. Cimino, *J. Chem. Theory Comput.*, 2009, **5**, 192–199.

68 V. Barone and P. Cimino, *Chem. Phys. Lett.*, 2008, **454**, 139–143.

69 S. Grimme, *J. Chem. Phys.*, 2006, **124**, 034108.

70 T. Schwabe and S. Grimme, *Phys. Chem. Chem. Phys.*, 2007, **9**, 3397–3406.

71 V. Barone, M. Biczysko and M. Pavone, *Chem. Phys.*, 2008, **346**, 247 – 256.

72 M. Biczysko, P. Panek and V. Barone, *Chem. Phys. Lett.*, 2009, **475**, 105–110.

73 S. Grimme and M. Steinmetz, *Phys. Chem. Chem. Phys.*, 2013, **15**, 16031–16042.

74 P. Carbonniere, T. Lucca, C. Pouchan, N. Rega and V. Barone, *J. Comput. Chem.*, 2005, **26**, 384–388.

75 D. Begue, P. Carbonniere and C. Pouchan, *J. Phys. Chem. A*, 2005, **109**, 4611–4616.

76 T. Risthaus, M. Steinmetz and S. Grimme, *J. Comput. Chem.*, 2014, **35**, 1509–1516.

77 C. Latouche and V. Barone, *J. Chem. Theory Comput.*, 2014, **10**, 5586–5592.

78 E. Papajak, H. R. Leverentz, J. Zheng and D. G. Truhlar, *J. Chem. Theory Comput.*, 2009, **5**, 1197–1202.

79 V. Barone, P. Cimino and E. Stendardo, *J. Chem. Theory Comput.*, 2008, **4**, 751–764.

80 C. Puzzarini, M. Biczysko and V. Barone, *J. Chem. Theory Comput.*, 2010, **6**, 828–838.

81 V. Barone, J. Bloino and M. Biczysko, *Phys. Chem. Chem. Phys.*, 2010, **12**, 1092–1101.

82 S. Kozuch, D. Gruzman and J. M. L. Martin, *J. Phys. Chem. C*, 2010, **114**, 20801–20808.

83 C. Puzzarini and M. Biczysko, *J. Phys. Chem. A*, 2015, **119**, 5386–5395.

84 V. Barone, *J. Chem. Phys.*, 2004, **120**, 3059–3065.

85 J. Bloino, M. Biczysko and V. Barone, *J. Chem. Theory Comput.*, 2012, **8**, 1015–1036.

86 V. Barone, J. Bloino, C. A. Guido and F. Lipparini, *Chem. Phys. Lett.*, 2010, **496**, 157–161.

87 J. Bloino and V. Barone, *J. Chem. Phys.*, 2012, **136**, 124108.

88 J. Bloino, *J. Phys. Chem. A*, 2015, **119**, 5269–5287.

89 J. Bloino, M. Biczysko and V. Barone, *J. Phys. Chem. A*, 2015, **119**, 11862–11874.

90 R. D. Amos, N. C. Handy, W. H. Green, D. Jayatilaka, A. Willetts and P. Palmieri, *J. Chem. Phys.*, 1991, **95**, 8323–8336.

91 J. M. L. Martin, T. J. Lee, P. M. Taylor and J.-P. François,

J. Chem. Phys., 1995, **103**, 2589–2602.

92 M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson and *et al.*, *Gaussian 09 Revision D.01*, 2009, Gaussian Inc. Wallingford CT 2009.

93 A. D. Boese and J. Martin, *J. Phys. Chem. A*, 2004, **108**, 3085–3096.

94 C. Puzzarini, M. Biczysko and V. Barone, *J. Chem. Theory Comput.*, 2011, **7**, 3702–3710.

95 M. Y. Choi and R. E. Miller, *Phys. Chem. Chem. Phys.*, 2005, **7**, 3565–3573.

96 D. Licari, A. Baiardi, M. Biczysko, F. Egidi, C. Latouche and V. Barone, *J. Comput. Chem.*, 2015, **36**, 321–334.

97 L. Lapinski, H. Rostkowska, M. J. Nowak, J. S. Kwiatkowski and J. Leszczyński, *Vibrat. Spectrosc.*, 1996, **13**, 23 – 40.

98 V. Barone, M. Biczysko, J. Bloino, P. Cimino, E. Penocchio and C. Puzzarini, *J. Chem. Theory Comput.*, 2015, **11**, 4342–4363.

99 M. Alauddin, M. Roy, J. K. Song and S. M. Park, *Int. J. Mass Spectrom.*, 2015, **381-382**, 41 – 47.

100 M. Biczysko, J. Bloino and V. Barone, *Chem. Phys. Lett.*, 2009, **471**, 143 – 147.

101 V. Barone, M. Biczysko, J. Bloino, F. Egidi and C. Puzzarini, *J. Chem. Phys.*, 2013, **138**, 234303.

102 M. Becucci, G. Pietraperzia, M. Pasquini, G. Piani, A. Zoppi, R. Chelli, E. Castellucci and W. Demtroeder, *J. Chem. Phys.*, 2004, **120**, 5601–5607.

103 G. Piani, M. Pasquini, G. Pietraperzia, M. Becucci, N. Schiccheri, M. Biczysko, M. Pavone and V. Barone, *J. Phys. Chem. A*, 2007, **111**, 12363–12371.

104 G. Pietraperzia, M. Pasquini, F. Mazzoni, G. Piani, M. Becucci, M. Biczysko, D. Michalski, J. Bloino and V. Barone, *J. Phys. Chem. A*, 2011, **115**, 9603–9611.

105 V. Barone, M. Biczysko, J. Bloino, M. Borkowska-Panek, I. Carnimeo and P. Panek, *Int. J. Quantum Chem.*, 2012, **112**, 2185–2200.

106 I. Carnimeo, M. Biczysko, J. Bloino and V. Barone, *Phys. Chem. Chem. Phys.*, 2011, **13**, 16713–16727.

107 F. Egidi, J. Bloino, C. Cappelli and V. Barone, *J. Chem. Theory Comput.*, 2014, **10**, 346–363.

108 H. Kvapilova, A. Vlcek, V. Barone, M. Biczysko and S. Zalis, *J. Phys. Chem. A*, 2015, **119**, 10137–10146.

109 S. King, *Spectrochim. Acta A*, 1972, **28**, 165–175.

110 A. Mardyukov, E. Sánchez-García, P. Rodziewicz, N. L. Doltsinis and W. Sander, *J. Phys. Chem. A*, 2007, **111**, 10552–10561.

111 M. Albrecht, C. A. Rice and M. A. Suhm, *J. Phys. Chem. A*, 2008, **112**, 7530–7542.

112 H. Rostkowska, A. Barski, K. Szczepaniak, M. Szczesniak and W. Person, *J. Molec. Structure*, 1988, **176**, 137 – 147.

Mode	Exp[b]		B2D3		B3D3		B2D3		B3D3		B2D3/B3D3
Assignment[a]	freq	I relative	harm	I(harm)	harm	I(harm)	anh	I(anh)	anh	I(anh)	anh
ν(N1-H)	**3457**	**151**	3631	94	3634	97	**3461**	**80**	**3461**	**80**	**3458**
ν(N3-H)	**3415**	**79**	3588	64	3599	70	**3419**	**56**	**3428**	**59**	**3417**
ν(C5-H)			3266	2	3261	3	3136	1	3125	1	3129
ν(C6-H)			3226	3	3222	3	3104	2	3089	3	3093
ν(C4=O)	**1738**	**543**	1761	632	1783	668	**1732**	**335**	**1752**	**325**	**1730**
ν(C5=C6)	1634	29	1665	38	1668	63	1627	23	1633	38	1631
δ(N1-H), δ(N3-H), ν(C6-N1)	1534	623	1571	699	1573	678	1527	453	1527	522	1526
δ(N3-H) + δ(C-H)	1430	4	1457	23	1456	28	1419	8	1419	9	1421
δ(C6-H)	1389	14	1402	49	1397	72	1366	50	1362	77	1366
δ(N-H)	1376	38	1396	33	1395	10	1360	24	1356	3	1357
δ(C5-H), ν(C6-N1), δ(C6-H), ν(N3-C4)	1223	33	1243	18	1237	6	1217	4	1212	10	1219
ν(C2-N3), δ(C6-H), δ(N1-H), ν(N3-C4)	1191	150	1216	230	1212	255	1185	199	1179	189	1184
ν(C=S), δ(ring), δ(N3-H)	1148	192	1169	154	1160	133	1140	86	1132	77	1142
ν(C6-N1), δ(C5-H), ν(C5=C6)	1060	4	1085	4	1082	8	1062	4	1059	7	1062
δ(ring)	986	15	1003	13	1001	11	979	16	980	12	982
γ(C-H)			981	0	967	0	966	0	947	0	962
ν(N3-C4), ν(C4-C5), ν(C=S)	907	4	916	6	911	5	899	6	894	5	900
γ(C5-H), γ(C4=O), γ(C6-H)	806	58	825	63	816	69	816	59	804	67	813
γ(N3-H), γ(C4=O), γ(C5-H)	727	4	745	8	740	17	735	4	727	10	731
δ(ring), ν(C4-C5), ν(C=S), ν(N1-C2)	710	5	720	4	717	4	708	3	706	3	709
τ(ring), γ(N3-H)	694	67	716	64	709	64	699	65	693	53	700
γ(C=S), γ(N1-H)	643	16	649	9	642	28	645	11	637	23	643
γ(N1-H)	604	31	618	32	620	26	610	27	608	29	606
δ(ring)	530	13	536	9	538	9	529	6	531	7	529
δ(C4=O)	491	12	497	11	499	12	492	10	494	11	492
δ(ring), ν(C=S)	451	15	453	13	449	14	448	13	444	14	448
τ(ring)	395	11	404	14	402	13	402	7	398	10	400
δ(C=S)	269	10	272	9	271	9	268	7	269	5	271
τ(ring)			162	1	165	1	160	1	162	1	159
τ(ring)			128	1	131	1	129	1	130	1	127
MAE[c]			29.2		28.2		6.4		7.5		5.5
|MAX|[c]			174		184		23		27		23

[a] Abbreviations: ν = stretching; δ = in-plane bending; γ = out-of-plane bending; τ = torsional.

[b] Experimental data from low-temperature Ar Matrix measurement from Ref.[97]

[c] Mean absolute errors (MAE) and maximum absolute deviations |MAX| with respect to experimental values.

Table 1 Harmonic and anharmonic vibrational wavenumbers and IR intensities of 2-thiouracil, computed with B2PLYP-D3/maug-cc-pVTZ (B2D3) and B3LYP-D3/N07D (B3D3) methods, and the hybrid B2PLYP-D3/B3LYP-D3 (B2D3/B3D3) anharmonic wavenumbers, in comparison with experiment.[97]

System	Exp(Ar)[a]	Exp(gas)[b]	B2D3	B2D3/B3D3
2-thiouracil	1738		1732	1730
formamide	1739	1754	1744	1744
formamide dimer	1728	1740	1729	1731
acetamide	1728	1746	1727	1724
acetamide dimer	1694	1706	1697	1689
MAE[c]			5.4	7.5
|MAX|[c]			10	13

[a] Experimental data from low-temperature Ar Matrix measurements from Refs.[53,97,109,110]

[b] Experimental data from the gas phase measurements from Refs.[111,112]

[c] Mean absolute errors (MAE) and maximum absolute deviations |MAX| with respect to average experimental values.

Table 2 Anharmonic stretching vibrational wavenumbers of the C=O functional groups for the molecular systems 2-thiouracil, acetamide, formamide, and the dimers of acetamide and formamide, computed with the B2D3 and hybrid B2D3/B3D3 methods, in comparison with experiments.[53,97,109–111]

System	Assignment [a]	Exp(Ar)[b]	Exp(gas)[c]	B2D3	B2D3/B3D3	B2D3NH/B2D3	B2D3NH/B3D3
thiouracil	ν(N1-H)	3457	3460	3461	3458	3476	3473
	ν(N3-H)	3415	3424	3419	3417	3437	3435
formamide	ν(N-H)as	3547	3564	3544	3545	3550	3552
	ν(N-H)s	3427	3440	3425	3425	3442	3442
formamide dimer	ν(N-H)as-as	3515	3527	3518	3515	3528	3525
	ν(N-H)s-as	3131	3167	3122	3117	3159	3154
acetamide	ν(N-H)as	3557	3568	3550	3552	3556	3558
	ν(N-H)s	3436	3448	3435	3435	3451	3451
acetamide dimer	ν(N-H)as-as	3515	3531	3522	3519	3532	3529
	ν(N-H)s-as	3140	3149	3109	3102	3147	3140
MAE[d]				10.9	12.2	9.2	7.6
|MAX|[d]				36	42	18	16

[a] Abbreviations: ν = stretching; s = symmetric; as = asymmetric.

[b] Experimental data from low-temperature Ar Matrix measurements from Refs.[53,97,109,110]

[c] Experimental data from the gas phase measurements from Refs.[111,112]

[d] Mean absolute errors (MAE) and maximum absolute deviations |MAX| with respect to average experimental values.

Table 3 Anharmonic N-H stretching wavenumbers for the molecular systems 2-thiouracil, acetamide, formamide, and the dimers of acetamide and formamide, computed with the B2D3 and hybrid B2D3/B3D3 methods, and their corrected counterparts (see text) in comparison with experiments.[53,97,109–111]

Assignment[a]	Exp[b,c]	B2D3[d]/B3D3	B2D3NH[e]/B3D3
4APM-M4PMN			
$\nu_{asym}(NH_2)_f$(4APM)	3530	3516	3525
$\nu_{sym}(NH_2)_{HB}$ (4APM)	3130	3060	3099
ν(N3-H)$_{HB}$	2758	2714	2772
Uracil-water			
ν(N3-H)$_f$ U-W1	3443	3432	3449
ν(N1-H)$_{HB}$ U-W1	3317	3305	3333
ν(N1-H)$_f$ U-W2	3492	3480	3493
ν(N3-H)$_{HB}$ U-W2	3256	3262	3292
ν(N1-H)$_f$ U-W3	3492	3479	3492
ν(N3-H)$_{HB}$ U-W3	3271	3246	3276
ν(N1-H)$_f$ U-W4	3492	3481	3494
ν(N3-H)$_f$ U-W4	3443	3431	3448
MAE[f]		21	11
|MAX|[f]		70	36

[a] Abbreviations: ν = stretching; f= free, HB = hydrogen bonding

[b] For 4APM-M4PMN experimental data from the IR vibrational spectra were measured via double resonance scheme utilizing femtosecond multiphoton ionization in the gas-phase [25].

[c] For uracil-water complexes experimental data from IR spectra recorded in Helium nanodroplets [95].

[d] For uracil-water complexes harmonic values from B2PLYP/maug-cc-pVTZ computations.

[e] Harmonic B2PLYP-D3/maug-cc-pVTZ wavenumbers corrected based on experimental gas phase, by linear regression: $\omega_{\mathrm{B2D3NH}} = k * \omega_{\mathrm{B2D3}} + \Delta$, with k=0.9252, Δ=286 cm^{-1} , see text and Figure3).

[f] Mean absolute errors (MAE) and maximum absolute deviations |MAX| with respect to experimental values.

Table 4 Anharmonic N-H stretching wavenumbers for the molecular systems: 4-aminopyrimidine-6-methyl-4-pyrimidinone and uracil-water complexes, computed with hybrid B2D3/B3D3 and B2D3NH/B3D3 method, in comparison with experiments. [25,95]

Assignment[a]	B2D3	B3D3	B3D3C
Adenine-Uracil (A-U)			
ν(C2=O) (U)	1788	1803	1786
ν(C4=O), ν(ring), δ(N3-H) (U)	1731	1744	1732
Adenine-ThioUracil (A-ThioU)			
ν(C2=O) (ThioU)	1787	1805	1788
Adenine-Thymine (A-T)			
ν(C2=O) (T)	1786	1800	1783
ν(C4=O), δ(N3-H) (T)	1719	1731	1720
Guanine-Cytosine (G-C)			
δ(N1-H) ν(C6=O) (G), $\delta_{sciss}(NH_2)$, ν(C2=O) (C)	1751	1761	1748
4APM-M4PMN			
ν(C4=O)	1735	1745	1733

[a] Abbreviations: ν = stretching; δ = in-plane bending; sciss = scissoring; asym = asymmetric; sym = symmetric; M1 = Monomer 1; M2 = Monomer 2.

Table 5 Harmonic C=O stretching wavenumbers for the nucleobase-dimers validation set. B3D3 and B3D3C values with respect to the B2D3 reference.

Assignment[a]	B2D3NH	B3D3	B3D3C
Adenine dimer			
$\nu_{asym}(NH_2)$ (M1,M2)	3691	3679	3685
$\nu_{asym}(NH_2)$ (M1,M2)	3691	3679	3685
$\nu_{sym}(NH_2)$ (M1,M2)	3335	3254	3317
$\nu_{sym}(NH_2)$ (M1,M2)	3298	3210	3279
ν(N9-H) (M1,M2)	3670	3657	3666
ν(N9-H) (M1,M2)	3670	3657	3666
Adenine-Uracil (A-U)			
$\nu_{asym}(NH_2)$ (A)	3704	3693	3697
$\nu_{sym}(NH_2)$ (A)	3439	3387	3432
ν(N9-H) (A)	3670	3656	3665
ν(N3-H) (U)	3006	2880	2994
ν(N1-H) (U)	3659	3645	3656
Adenine-ThioUracil (A-ThioU)			
$\nu_{asym}(NH_2)$ (A)	3706	3693	3697
$\nu_{sym}(NH_2)$ (A)	3465	3409	3451
ν(N9-H) (A)	3670	3656	3665
ν(N3-H) (ThioU)	3045	2937	3043
ν(N1-H) (ThioU)	3654	3641	3652
Adenine-Thymine (A-T)			
$\nu_{asym}(NH_2)$ (A)	3704	3693	3697
$\nu_{sym}(NH_2)$ (A)	3438	3385	3431
ν(N9-H) (A)	3670	3656	3666
ν(N3-H) (T)	3020	2895	3007
ν(N1-H) (T)	3660	3647	3657
Guanine-Cytosine (G-C)			
$\nu_{asym}(NH_2)$ (G)	3713	3706	3709
$\nu_{asym}(NH_2)$ (C)	3700	3686	3691
$\nu_{sym}(NH_2)$ (G)	3430	3378	3425
ν(N1-H), $\nu(NH_2)$ (G), $\nu(NH_2)$ (C)	3252	3172	3246
$\nu(NH_2)$ (C), ν(N1-H) (G)	3185	3091	3176
ν(N9-H) (G)	3671	3657	3666
ν(N1-H) (C)	3646	3631	3644
4APM-M4PMN			
$\nu_{asym}(NH_2)$(4APM)	3709	3696	3700
$\nu_{sym}(NH_2)$ (4APM)	3342	3277	3337
ν(N3-H) (M4PMN)	3117	3011	3107

[a] Abbreviations: ν = stretching; δ = in-plane bending; sciss = scissoring; asym = asymmetric; sym = symmetric; M1 = Monomer 1; M2 = Monomer 2.

Table 6 Harmonic N-H stretching wavenumbers for the nucleobase-dimers validation set. B3D3 and B3D3C values with respect to the B2D3NH reference.

Assignment[a]	Exp[b,c,d]	B3D3	B3D3C[d]/B3D3
formamide dimer			
ν(N-H)as-as	3527	3503	3509
ν(N-H)s-as	3167	3097	3152
acetamide dimer			
ν(N-H)as-as	3531	3509	3514
ν(N-H)s-as	3149	3081	3138
4APM-M4PMN			
$\nu_{asym}(NH_2)_f$(4APM)	3530	3512	3516
$\nu_{sym}(NH_2)_{HB}$ (4APM)	3130	3034	3095
$\nu(N3\text{-}H)_{HB}$	2758	2666	2762
Uracil-water			
$\nu(N3\text{-}H)_f$ U-W1	3443	3440	3455
$\nu(N1\text{-}H)_{HB}$ U-W1	3317	3277	3318
$\nu(N1\text{-}H)_f$ U-W2	3492	3480	3490
$\nu(N3\text{-}H)_{HB}$ U-W2	3256	3252	3293
$\nu(N1\text{-}H)_f$ U-W3	3492	3478	3489
$\nu(N3\text{-}H)_{HB}$ U-W3	3271	3233	3276
$\nu(N1\text{-}H)_f$ U-W4	3492	3481	3491
$\nu(N3\text{-}H)_f$ U-W4	3443	3438	3454
MAE[g]		34	12
|MAX|[g]		96	37

[a] Abbreviations: ν = stretching; asym/as = asymmetric; sym/s = symmetric; f= free, HB = hydrogen bonding

[b] For formamide and acetamide dimers experimental data from the FTIR gas phase measurements [111]

[c] For 4APM-M4PMN experimental data from the IR vibrational spectra were measured via double resonance scheme utilizing femtosecond multiphoton ionization in the gas-phase [25].

[d] For uracil-water complexes experimental data from IR spectra recorded in Helium nanodroplets [95].

[e] Harmonic B3D3 harmonic wavenumbers corrected based on the B2D3NH, with linear regression: $\omega_{\mathrm{B3D3C}} = k * \omega_{\mathrm{B3D3}} + \Delta$, with k=0.8651, Δ=502 cm^{-1} , see text and Figure5).

[f] Mean absolute errors (MAE) and maximum absolute deviations |MAX| with respect to experimental values.

Table 7 Anharmonic N-H stretching wavenumbers for the model molecular systems: formamide and acetamide dimers, 4-aminopyrimidine-6-methyl-4-pyrimidinone and uracil-water complexes, computed with the B3D3 and B3D3C/B3D3 models, in comparison with experiment. [25,95,111]

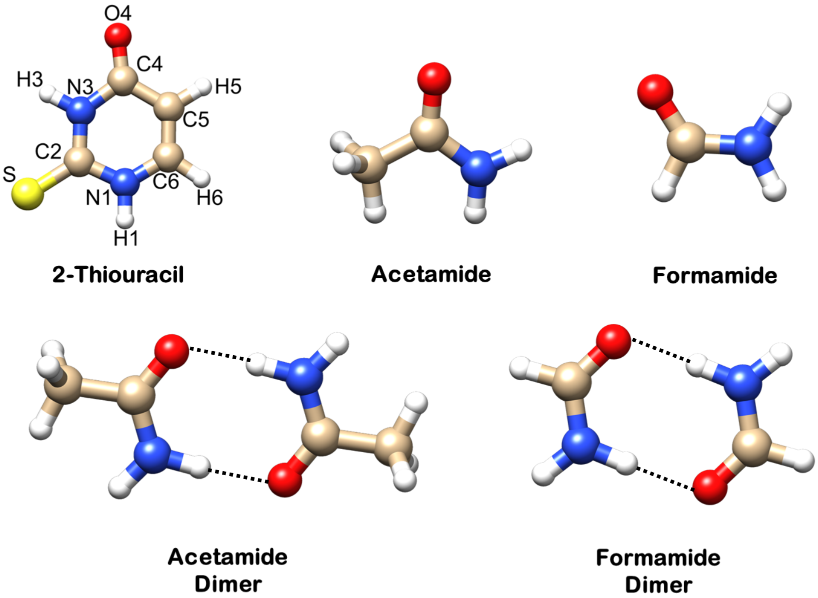

Fig. 1 Structures of 2-thiouracil, acetamide, formamide, and the dimers of acetamide and formamide.

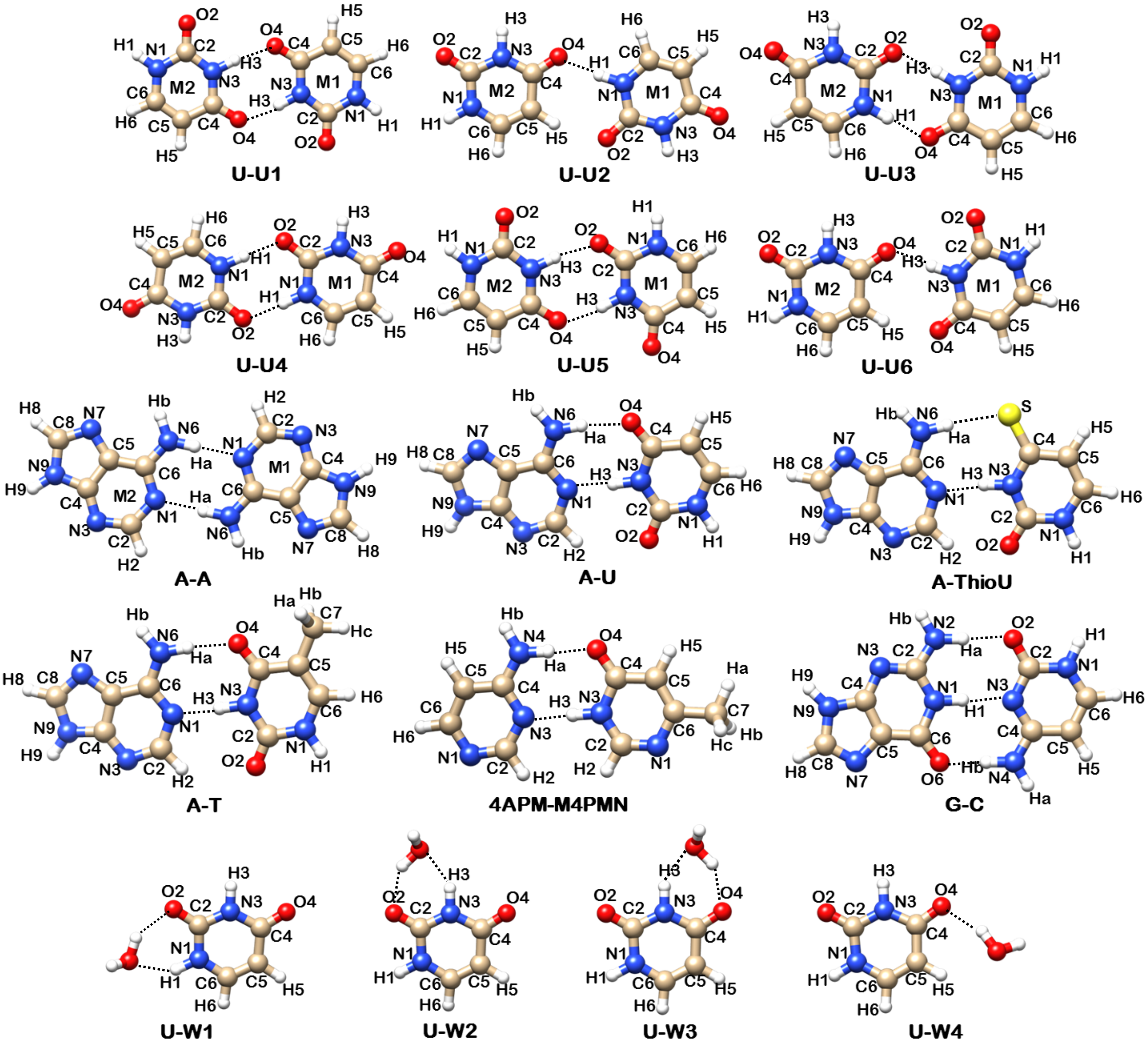

Fig. 2 Structures and numbering schemes of the nucleobase complexes used in this work as training and validation sets.

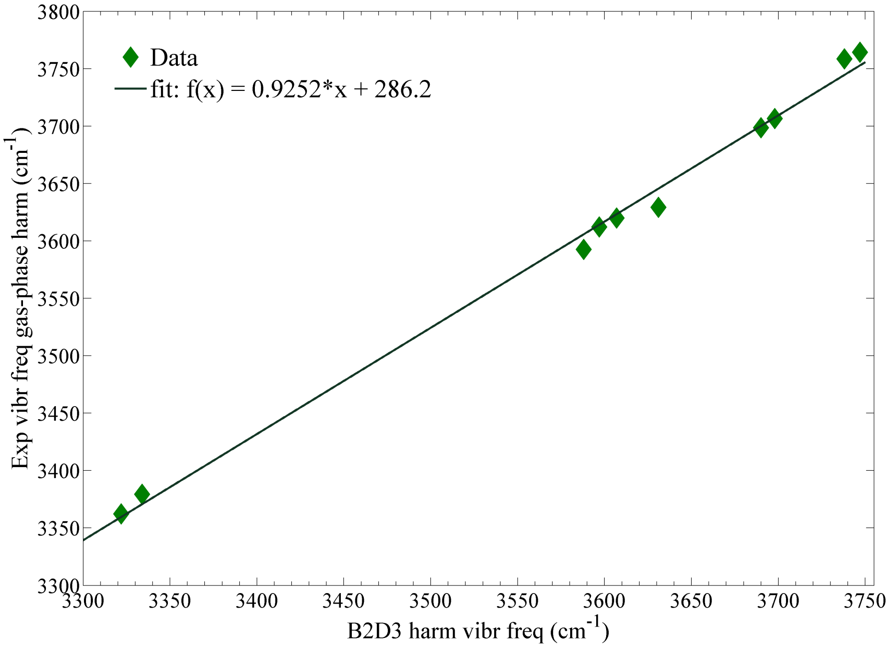

Fig. 3 Linear fits of B2PLYP-D3/maug-cc-pVTZ harmonic wavenumbers *vs.* experimental harmonic wavenumbers (estimated as ω_{exp}+Δ_{anh}^{theo}, see text), for ν(N-H) vibrations of 2-thiouracil, formamide, acetamide and cyclic dimers of the two last molecules.

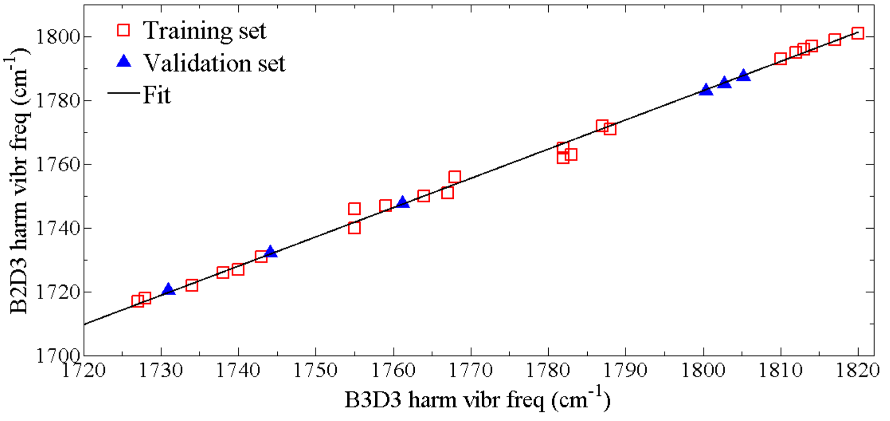

Fig. 4 Linear fits of B2D3 harmonic wavenumbers *vs.* B3D3 ones, for ν(C=O), along with data for the training and validation sets.

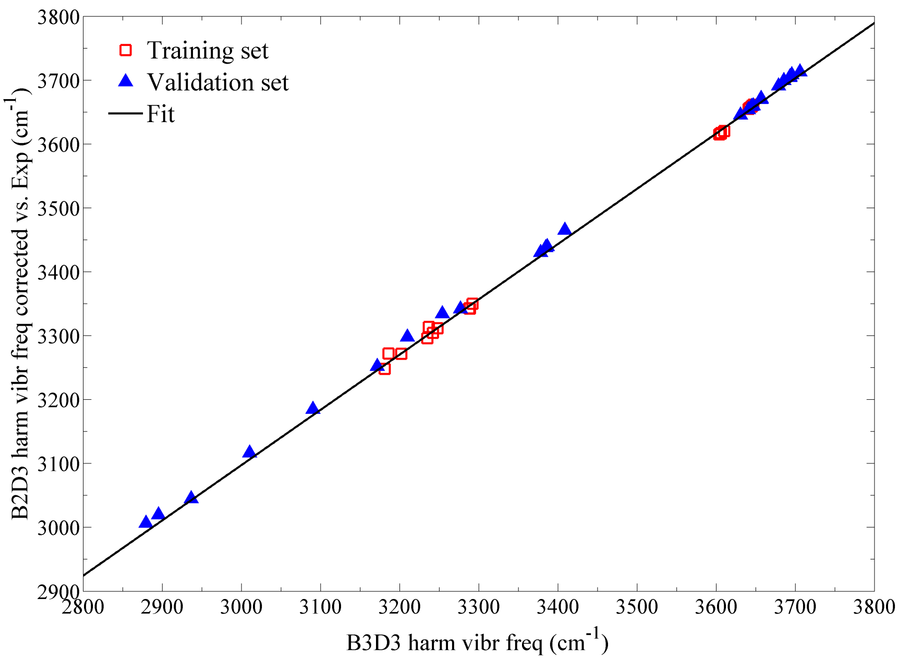

Fig. 5 Linear fits of B2D3NH harmonic wavenumbers *vs.* B3D3 ones, for ν(N-H), along with data for the training and validation sets.

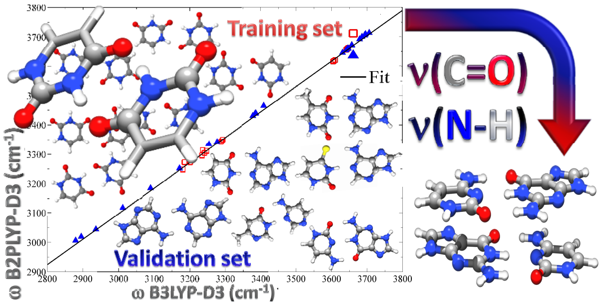

Fig. 6 Graphical abstract. The GVPT2 approach in conjunction with B3LYP-D3/N07D computations allows to obtain reliable anharmonic frequencies for C=O and N-H moieties involved in H-bridges for large systems, provided that the harmonic part is scaled by linear equations based on new B2PLYP-D3/maug-cc-pVTZ benchmarks.