JCTC Journal of Chemical Theory and Computation

Article

pubs.acs.org/JCTC

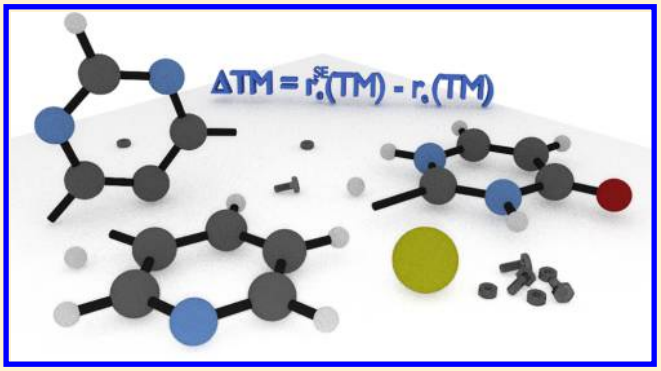
Semiexperimental Equilibrium Structures for Building Blocks of Organic and Biological Molecules: The B2PLYP Route

Emanuele Penocchio, Matteo Piccardo, and Vincenzo Barone*

Scuola Normale Superiore, 56126 Pisa, Italy

Supporting Information

ABSTRACT: The B2PLYP double hybrid functional, coupled with the correlation-consistent triple-ζ cc-pVTZ (VTZ) basis set, has been validated in the framework of the semiexperimental (SE) approach for deriving accurate equilibrium structures of molecules containing up to 15 atoms. A systematic comparison between new B2PLYP/VTZ results and several equilibrium SE structures previously determined at other levels, in particular B3LYP/SNSD and CCSD(T) with various basis sets, has put in evidence the accuracy and the remarkable stability of such model chemistry for both equilibrium structures and vibrational corrections. New SE equilibrium structures for phenylacetylene, pyruvic acid, peroxyformic acid, and phenyl radical are discussed and compared with literature data. Particular attention has been devoted to the discussion of systems for which lack of sufficient experimental data prevents a complete SE determination. In order to obtain an accurate equilibrium SE structure for these situations, the so-called templating molecule approach is discussed and generalized with respect to our previous work. Important applications are those involving biological building blocks, like uracil and thiouracil. In addition, for more general situations the linear regression approach has been proposed and validated.



1. INTRODUCTION

The availability of accurate equilibrium structures of molecules of medium- to large-size in the gas phase is a mandatory prerequisite for a deeper understanding of their intrinsic physical−chemical properties and of the role of dynamical and environmental effects in tuning their behavior.[1−16] Moreover, accurate reference geometries are of paramount interest for testing the accuracy of different quantum mechanical (QM) methods[2,17−21] and for developing accurate force fields to be used in molecular mechanics (MM) and/or molecular dynamics (MD) simulations.[22−29] This has stimulated a large experimental and computational effort, which has, however, faced a number of difficulties. From the experimental point of view, the information needed for reaching accurate equilibrium geometries becomes prohibitive except for very small systems,[2,3,30,31] and the same happens for the most sophisticated QM methods due to their very unfavorable scaling with the number of active electrons.[1,2,32−35]

A breakthrough in this field is represented by the introduction of the so-called semiexperimental (SE) equilibrium geometry (r_e^{SE}), which is obtained by a least-squares fit (LSF) of ground vibrational state experimental rotational constants (B_0^β) of different isotopologues corrected by computed vibrational contributions (ΔB_{vib}^β).[1,2,36] After the pioneering work of Pulay and co-workers,[36] this approach has led to accurate geometries of several molecular systems thanks to the terrific development of hardware and software in the last years.[37]

From a computational point of view, the bottleneck of the SE protocol is the calculation of the semidiagonal cubic force field at a sufficiently accurate level. Several studies have shown that, thanks to a fortunate, but systematic error cancellation, the level of theory needed to obtain accurate vibrational corrections is significantly lower than that required for obtaining accurate equilibrium geometries.[37] This consideration prompted us to start a systematic study of different systems and computational levels with the aim of providing a large set of reference structures for the main building blocks of organic and biological systems by means of a reliable yet feasible computational model.

The results collected in the present paper are both a continuation and an improvement of those presented in a recent work,[38] where we showed how the use of DFT methods, in particular the B3LYP functional in conjunction with a polarized double-ζ basis set, in the cubic force field calculations strongly reduces the computational cost required for the determination of the vibrational corrections, without affecting the accuracy of the resulting SE equilibrium geometries. This allowed us to obtain accurate SE equilibrium structures for

Received: June 30, 2015
Published: August 20, 2015

DOI: 10.1021/acs.jctc.5b00622
J. Chem. Theory Comput. 2015, 11, 4689−4707

molecules containing up to 15 atoms, mostly including molecular building blocks of chemical and biological interest.

In the present work we report the complete SE equilibrium structure determination of four new systems (i.e., phenylacetylene, pyruvic acid, peroxyformic acid, and phenyl radical) obtained by use of ΔB_{vib}^{β} corrections calculated at the B3LYP level (referred to as B3LYP r_e^{SE}), which have been added to the collection of B3LYP r_e^{SE} geometries previously determined (B3se set).

Studying pyruvic acid, we found quite large uncertainties on B3LYP r_e^{SE} parameters and significant discrepancies with r_e values optimized at the CCSD(T)/CBS+CV level.[39] Thus, we decided to validate the use of the more accurate double hybrid B2PLYP functional,[40] coupled with a valence triple-ζ basis set,[41,42] for SE geometrical determinations and redetermine all the r_e^{SE} structures of the molecules collected in the B3se set. The new B2PLYP SE equilibrium structures have been collected in the B2se set, which has been made available in a graphical interactive form on our web site dreams.sns.it, together with the B3se, B2th, and B3th sets (the latter two containing the purely theoretical equilibrium structures).[43]

Comparison between the B3LYP and B2PLYP SE results allows to study the stability of the SE determinations when DFT ΔB_{vib}^{β} corrections are used. It has been found that B3LYP and B2PLYP r_e^{SE} geometries do not show significant differences for the majority of the systems, thus further confirming the accuracy of the equilibrium geometries determined in our previous work,[38] B3LYP ΔB_{vib}^{β} corrections representing thus the best compromise between accuracy and computational cost.

Significant discrepancies on some geometrical parameters have been found just for few systems. In those cases, B2PLYP SE equilibrium geometries are affected by lower uncertainties and are in better agreement with high level post-Hartree−Fock r_e optimizations. In the light of these results, the B2se set is proposed as the reference set for equilibrium structures.

The r_e^{SE} equilibrium structures represent a high quality benchmark for structural studies and validation of computational models, as well as a set of possible "structural synthons" that can be used, in the framework of the previously presented template approach,[38] in order to build reliable equilibrium structures of bigger molecules for which sufficient experimental data are not available and for which high level calculations, such as CCSDT/CBS+CV ones, are not feasible.

Unfortunately in the absence of sufficient experimental data, we are unable to obtain complete r_e^{SE} equilibrium structures for some important classes of molecules, like, for instance, nucleosides. This consideration prompted us to include in our two sets of SE equilibrium geometries also the partial SE structure of uracil, obtained keeping fixed in the least-square fit the non determinable parameters at the highest possible QM level. This last choice allowed us to obtain, without using additional expensive calculations, a very accurate equilibrium structure for 2-thiouracil. This result shows, as a sort of proof of concept, how the template approach represents a powerful technique that can be used to draw accurate information about equilibrium structure for molecules of biological interest.

Finally, the introduction of some technical improvements, explained in the following sections, allows to suggest a general strategy to deal with systems for which the lack of experimental information for a sufficient number of isotopologues prevents the derivation of a complete SE equilibrium structure, and that, in a sense, generalizes the template approach previously presented.

The strategy presented and validated in this work to face the problem of the absence of a sufficient number of experimental data is reminiscent of similar approaches adopted in other fields, e.g. in the structure determination via gas electron diffraction,[44−46] in which the interplay of theory and experiment allows to go beyond the intrinsic limitations of both approaches.

2. METHODOLOGY AND COMPUTATIONAL DETAILS

The semiexperimental structure r_e^{SE} is obtained by a least-squares fit of the molecular parameters to the experimental ground-state rotational constants $(B_0^{\beta})^{EXP}$ of a set of isotopologues, or their corresponding moments of inertia $(I_0^{\beta})^{EXP}$, where $\beta = x, y$, or z is one of the principal inertial axes in the molecule-fixed reference frame, corrected subtracting the vibrational and electronic contributions (see section 4 for cases when there are more unknown parameters than data). The equilibrium SE rotational constants $(B_e^{\beta})^{SE}$ are then given by

$$(B_e^{\beta})^{SE} = (B_0^{\beta})^{EXP} - (\Delta B_0^{\beta})^{QM} \quad (1)$$

where $(\Delta B_0^{\beta})^{QM}$ is explicitly given by

$$(\Delta B_0^{\beta})^{QM} = \frac{m_e}{M_p} g^{\beta\beta} B_e^{\beta} - \sum_i \frac{\alpha_i^{\beta} d_i}{2} = \Delta B_{el}^{\beta} + \Delta B_{vib}^{\beta} \quad (2)$$

ΔB_{el}^{β} is an electronic contribution, which derives from the rotational $\mathbf{g}$ tensor and the ratio between electron (m_e) and proton (M_p) masses.[47−49] It will be systematically included in our computations, even if often negligible. Within the BO approximation and enforcing Eckart−Sayvetz conditions,[50−53] the vibrational contribution ΔB_{vib}^{β} is obtained by applying second-order vibrational perturbation theory (VPT2) to the molecular ro-vibrational Hamiltonian expressed in normal coordinates.[54−56] In the summation, α_i^{β} are the vibration−rotation interaction constants and d_i is the degeneracy of the ith vibrational normal mode.[47,56,57]

In the present investigation, the quadratic and cubic force fields required for the computation of the ΔB_{vib}^{β} term have been evaluated by two computational models rooted into the density functional theory (DFT). The hybrid B3LYP functional[58−60] was used in conjunction with the polarized double-ζ SNSD basis set,[61−63] which is obtained adding to the 6-31G** basis set one contracted and one diffuse s function on each atom, together with diffuse p and d functions on non-hydrogen atoms, all with optimized orbital exponents. Several papers have shown that the SNSD basis set represents an excellent compromise between accuracy and computational cost for several spectroscopic properties, and in particular within the anharmonic computational schemes widely adopted in our group.[21,61,62,64−67] In addition, the remarkable results obtained in the previous work[38] with molecules containing up to 15 atoms encourage us to believe that the B3LYP/SNSD computational model can pave the route to the extension of the SE approach to the structural analysis of larger systems, provided its limits are clearly defined. The double-hybrid B2PLYP functional[40,68,69] was used in conjunction with the correlation-consistent polarized triple-ζ cc-pVTZ basis set[41,70−72] (shortly denoted as VTZ), which has shown to provide very accurate harmonic frequencies and anharmonic corrections.[35,69] The harmonic parts have been evaluated from analytical second derivatives, whereas the cubic force fields by

DOI: 10.1021/acs.jctc.5b00622
J. Chem. Theory Comput. 2015, 11, 4689−4707

Table 1. r_e^{SE} and r_e Geometries for Phenyl Radical[a]

	r_e^{SE}[b]				r_e			
	B3LYP/SNSD[d]	B2PLYP/VTZ[d]	CCSD(T)/ANO0[d,e]	CCSD(T)/ANO0[e,f]	B3LYP/SNSD	B2PLYP/VTZ	CCSD(T)/CVQZ[g]	CC/CBS+CV+aug[h]
$r(C_i-C_o)$	1.3726(7)[c]	1.3728(7)[c]	1.3722(3)[c]	1.3731(47)[c]	1.3764	1.3716	1.3741	1.3693
$r(C_o-C_m)$	1.3989(3)	1.3990(3)	1.3984(1)	1.3980(17)	1.4043	1.3994	1.3993	1.3947
$r(C_m-C_p)$	1.3936(2)	1.3937(2)	1.3931(1)	1.3927(16)	1.3968	1.3918	1.3926	1.3882
$r(C_o-H)$	1.0799(2)	1.0799(2)	1.0808(1)	1.0805(12)	1.0864	1.0804	1.0806	1.0811
$r(C_m-H)$	1.0811(2)	1.0813(2)	1.0819(1)	1.0814(12)	1.0872	1.0814	1.0817	1.0822
$r(C_p-H)$	1.0802(2)	1.0804(2)	1.0809(1)	1.0818(14)	1.0861	1.0804	1.0808	1.0812
$a(C_o-C_i-C_o)$	125.87(3)	125.82(3)	125.85(1)	125.80(20)	125.98	125.93	125.74	125.78
$a(C_i-C_o-C_m)$	116.59(3)	116.62(3)	116.60(1)	116.60(23)	116.49	116.53	116.61	116.56
$a(C_o-C_m-C_p)$	120.17(2)	120.15(2)	120.16(1)	120.18(13)	120.19	120.18	120.17	120.29
$a(C_m-C_p-C_m)$	120.62(2)	120.63(2)	120.63(1)	120.65(12)	120.66	120.66	120.69	120.52
$a(C_i-C_o-H)$	122.97(3)	122.96(3)	122.90(1)	122.35(16)	122.47	122.43	122.35	122.44
$a(C_o-C_m-H)$	120.41(3)	120.44(3)	120.39(1)	119.83(18)	119.63	119.63	119.66	119.56
$a(C_m-C_p-H)$	119.69(1)	119.68(1)	119.68(1)	119.68(8)	119.67	119.67	119.65	119.74
Rms resid. [MHz]	0.0005	0.0005	0.0002	0.0035				
mean Δ_e [uÅ^2]	−0.00162	−0.00272	−0.00249	−0.00251				

[a]See Figure 3 for atom numbering. Distances in angstroms, angles in degrees. *i*, *o*, *m*, *p* stands for *ipso*, *orto*, *meta* and *para*, respectively. [b]The digits within parentheses are the uncertainties on the geometrical parameters. [c]Indicates the inclusion of ΔB_{el}^{β} and $\Delta_e = I^C - I^B - I^A$ is the inertial defect. [d]The fits have been performed on SE I_e^C moments of inertia of all nine isotopomers and I_e^A moments of inertia of C_6D_5, C_6H_5, *o*-C_6H_4D, and *p*-C_6H_4D isotopomers (scheme 2). [e]CCSD(T)/ANO0 ΔB_{vib}^{β} from ref 79. [f]The fits have been performed on SE I_e^C moments of inertia of all nine isotopomers and I_e^A moments of inertia of C_6D_5, C_6H_5, *o*-C_6H_4D, *m*-C_6H_4D, and *p*-C_6H_4D isotopomers (scheme 1). [g]CCSD(T)/CVQZ r_e from ref 79. [h]CCSD(T)/CBS+CV+aug r_e from ref 89.

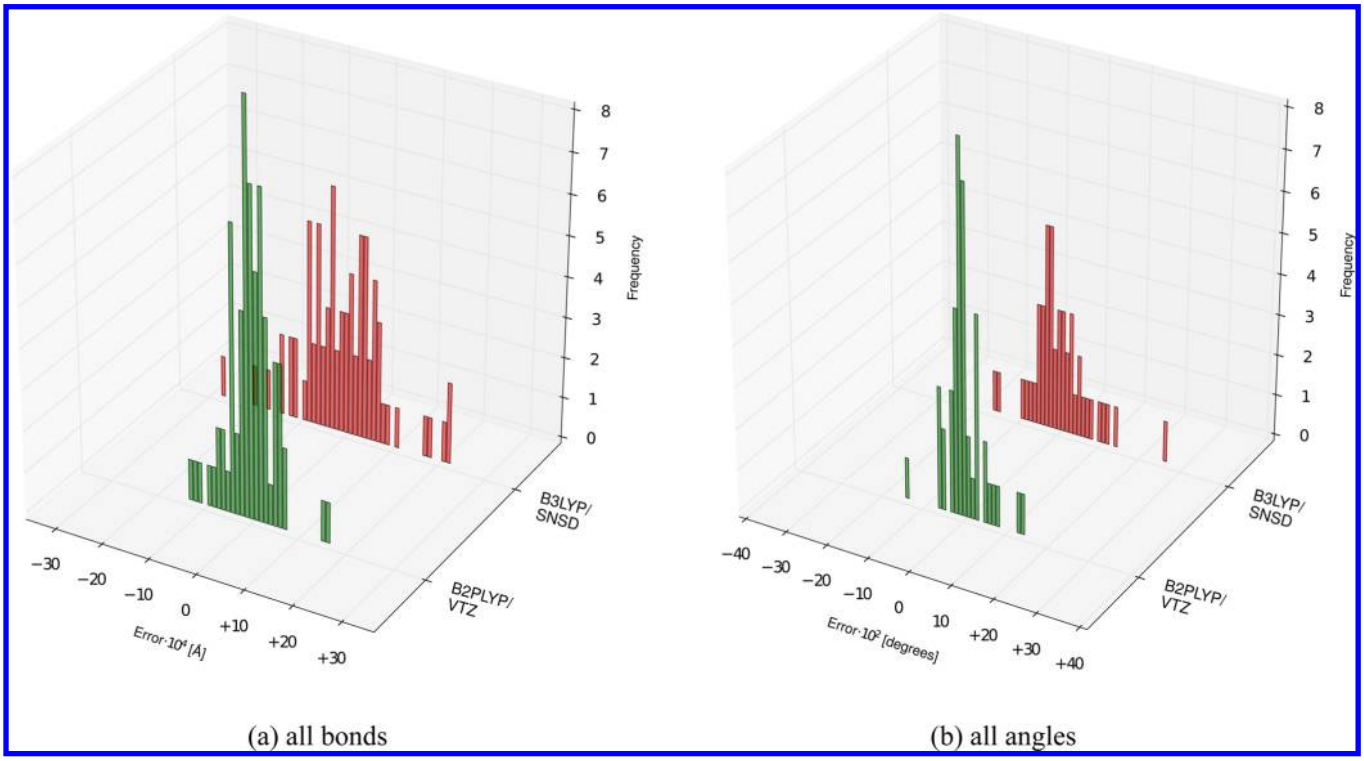


Figure 1. Statistical distributions of the B2PLYP/VTZ and B3LYP/SNSD deviations from CCSD(T) SE equilibrium parameters for the molecules belonging to the CCse set.

numerical differentiation in the normal-coordinate representation.[73−78] Very-tight criteria have been used for the SCF and geometry optimization convergence, together with an ultrafine grid for the numerical integration connected to the exchange-correlation functional and its derivatives. All the calculations have been performed with the Gaussian package, with the default step of 0.01 Å for the numerical differentiations.

The ΔB_{el}^{β} contributions have been evaluated by calculating the $g^{\beta\beta}$ constants by the B3LYP functional in conjunction with the aug-cc-pVTZ (hereafter AVTZ) basis set.[42,70]

3. SEMIEXPERIMENTAL EQUILIBRIUM STRUCTURE

3.1. Performance of B2PLYP/VTZ Force Fields. The performance of the B2PLYP/VTZ force fields in the

DOI: 10.1021/acs.jctc.5b00622
J. Chem. Theory Comput. 2015, 11, 4689−4707

computation of the vibrational contributions to ground-state rotational constants, used in the derivation of SE equilibrium geometries, has been systematically tested on the 21 molecules previously included in the CCse set[38] (HCN, HNC, HCO^+, $HNCCN^+$, HCCH, HCCCCH, H_2CCCH_2, SH_3^+, NH_3, H_2O, H_2CO, CH_2ClF, CH_2CHF, *cis*-CHFCHCl, oxirane, dioxirane, cyclobutene, *trans*-glyoxal, *cis* and *trans*-acrolein, pyridazine) and on phenyl radical, leading to a total number of 22 molecules.

For the 21 molecules already discussed in our previous paper,[38] the new B2PLYP/VTZ SE equilibrium structures are explicitly reported in Table S1 (see the Supporting Information, hereafter SI), together with the B2PLYP/VTZ r_e geometries. Regarding phenyl radical, the recent results in the detection of its microwave spectra[79] and the determination of its cubic force field at the CCSD(T)/ANO0 level[79] have allowed a CC SE determination, that is reported in Table 1 together with DFT SE results, and will be discussed in detail in the next section. In all cases the B3LYP/AVTZ electronic corrections have been also included.

In the tables we report the root-mean-square (RMS) of the residuals in terms of equilibrium rotational constants (hereafter simply referred to as residuals), and for planar molecules, the mean inertial defects $\Delta_e = I^C - I^B - I^A$, as indicators of the quality of the fits. For all molecules studied in this work, all the B2PLYP/VTZ vibrational corrections have been collected in the SI.

The differences in the geometrical parameters of the B2PLYP/VTZ and B3LYP/SNSD SE equilibrium geometries with respect to the CCSD(T) SE equilibrium structures are graphically reported in Figure 1. On inspection of the figure, it is apparent that B2PLYP results show deviations strongly clustered around zero, which never exceeds 0.0020 Å. From the statistical analysis reported in Table 2, the deviations for all B2PLYP/VTZ bond lengths and valence angles show mean values of zero, very low standard deviations (0.0005 Å and 0.05°) and mean absolute errors (MAEs) (0.0004 Å and 0.03°). A general improvement is observed with respect to the B3LYP/SNSD distributions, although the differences are small. The R^2 of the linear regressions obtained plotting the CCSD(T) r_e^{SE} values as functions of the corresponding B3LYP and B2PLYP r_e^{SE} values (see Figure 2 and Table 3) does not point out any significant deviation from linearity, confirming that the use of both B2PLYP and B3LYP corrections in the SE approach leads to results that reproduce very well the best SE equilibrium structures.

3.2. B3se and B2se Sets. In the light of the excellent results obtained, B2PLYP/VTZ SE equilibrium structures have been determined for all molecules belonging the B3se set[38] not yet taken into account in the previous section (CH_2F_2, CCl_2F_2, CH_2Cl_2, $CHClF_2$, ethene, ethenol, propene, butadiene, cyclopropane, aziridine, benzene, pyrrole, pyrazole, imidazole, furan, thiophene, maleic anhydride, pyridine, dimethyl ether, *cis* and *trans*-formic acid, *cis*-methyl formate, glycolaldehyde, and propanal; see Figure 3, Table S2 (in the SI) and Tables 4 and 5 in the text), adopting the same set-ups used in B3LYP/SNSD fits, if not explicitly indicated in the following, and collecting all the B2PLYP/VTZ SE equilibrium geometries in the B2se set.

Moreover, new B3LYP/SNSD and B2PLYP/VTZ r_e^{SE} structures have been estimated for phenyl radical (Table 1), phenylacetylene (Table 6), peroxyformic (Table 4), and pyruvic (Table 7) acids, increasing to 51 the number of molecules included in B3se and B2se sets. All these structures have been reported in graphical interactive form on our web site dreams.sns.it,[43] while all the B2PLYP/VTZ ΔB_{vib}^{β} corrections are summarized in Tables S3 and S4 of the SI, together with the B3LYP/SNSD ΔB_{vib}^{β} corrections for the molecules not yet studied in ref 38.

The agreement between the parameters of the B3se and B2se sets is generally very satisfactory. The few problematic cases are collected in Table 4 and discussed in the following.

A quite large discrepancy is found for the C3−C4 bond length of furan, where the B3LYP r_e^{SE} value (1.4344 Å) is larger than its B2PLYP counterpart (1.4317 Å), which is, in turn, in remarkable agreement with the MP2/VTZ result (1.432 Å), taken from ref 80.

Conversely, the B2PLYP r_e^{SE} parameters for thiophene show a very good agreement with B3LYP ones, which present relevant differences with respect to those previously derived from a combined use of electron diffraction (ED), microwave spectroscopy (MW) and computed vibrational contributions at the B3LYP/6-311+G* level,[81] as discussed in ref 38. The largest discrepancies are found on S−C2 (1.704 and 1.7126 Å for literature and present determinations, respectively), C2−C3 (1.372 versus 1.3625 Å), C2−H (1.085 versus 1.0772 Å), and C3−H (1.088 versus 1.0792 Å) bond lengths, and C2−S−C5 (94.2° versus 91.88°).

Significant differences are found for the H−C═O bond angles of *cis*- and *trans*-formic acids. The B3LYP value for the *cis* and *trans* rotamers are respectively 0.34°and 0.24° larger than the B2PLYP ones, (124.21°and 125.38°), which are in better agreement with the MP2/VTZ SE equilibrium values (123.26°and 125.04°).[82] All other parameters calculated by employing B2PLYP vibrational corrections are in satisfactory agreement with their B3LYP counterparts.

Table 2. Mean, Standard Deviation, and MAE for the B2PLYP/VTZ and B3LYP/SNSD Deviations from CCSD(T) SE Equilibrium Parameters for the Molecules Belonging to the CCse set[a]

	B2PLYP/VTZ	B3LYP/SNSD
	All Bonds (74 Items)	
mean	+0.0000	−0.0001
st dev	0.0005	0.0009
MAE	0.0004	0.0007
	CH Bonds (30 Items)	
mean	−0.0002	−0.0005
st dev	0.0004	0.0005
MAE	0.0003	0.0006
	CC Bonds (21 Items)	
mean	+0.0002	+0.0005
st dev	0.0005	0.0009
MAE	0.0004	0.0009
	CO Bonds (7 Items)	
mean	+0.0002	+0.0003
st dev	0.0002	0.0004
MAE	0.0002	0.0005
	All Angles (46 Items)	
mean	+0.00	+0.00
st dev	0.05	0.07
MAE	0.03	0.05

[a]For the different types of bonds, only the sets having at least seven items have been considered.

DOI: 10.1021/acs.jctc.5b00622
J. Chem. Theory Comput. 2015, 11, 4689−4707

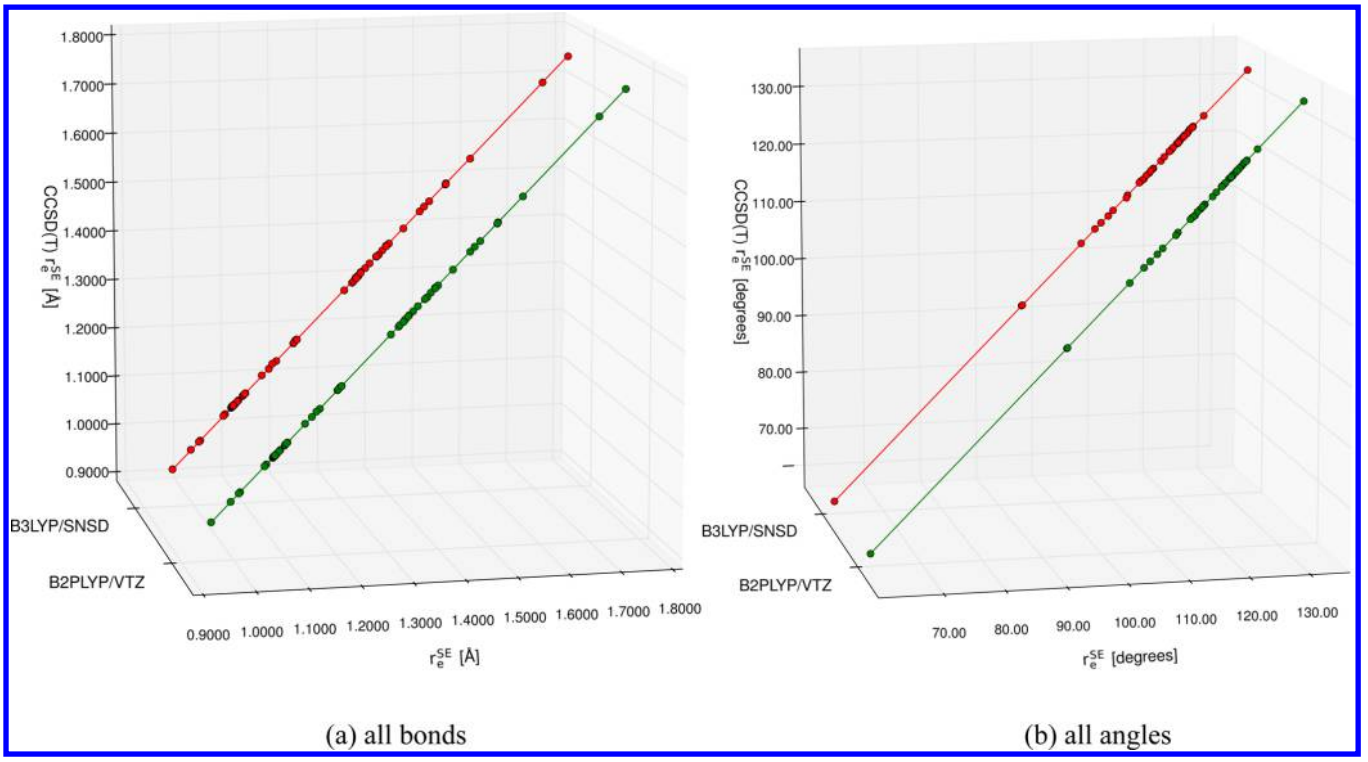


Figure 2. CCSD(T) r_e^{SE} equilibrium parameters plotted versus the B2PLYP/VTZ and B3LYP/SNSD r_e^{SE} values for the molecules belonging to the CCse set.

Table 3. Parameters for Linear Regressions of the CCSD(T) r_e^{SE} Parameters versus the B2PLYP/VTZ and B3LYP/SNSD r_e^{SE} Ones for the Molecules Belonging to the CCse Set[a]

	B2PLYP/VTZ	B3LYP/SNSD
	All Bonds	
A	0.999068	0.998004
B	0.001152	0.002514
R^2	0.999993	0.999978
	All Angles	
A	1.000425	1.000032
B	−0.050991	−0.005710
R^2	0.999982	0.999957

[a] $r_e^{SE}(CCSD(T)) = A r_e^{SE}(DFT) + B$.

Peroxyformic acid (see Figure 3) is the simplest organic peroxyacid exhibiting internal hydrogen bonding. It shows a planar structure completely characterized by nine geometrical parameters. The B2PLYP/VTZ and B3LYP/SNSD r_e^{SE} structures, obtained by fitting the SE I_A^e and I_C^e moments of inertia, show a very good agreement for all parameters except the OH bond length, for which the B3LYP value (0.9770 Å) is larger than the B2PLYP one (0.9720 Å). To the best of our knowledge, the most accurate r_e geometry presented in the literature, optimized at MP2/AVQZ level,[83] is not accurate enough to allow a reliable quantitative comparison.

For *cis*-methyl formate, in ref 38 we noted discrepancies between the MP2/VTZ and B3LYP/SNSD SE values for the parameters related to the H_{plane} atom. In particular, $r(C_m-H_{plane})$ is 1.0845 and 1.0793 Å and $a(O-C_m-H_{plane})$ is 105.35° and 106.05° at the B3LYP and MP2 level, respectively. B2PLYP results are very close to B3LYP ones, showing a C_m-H_{plane} bond length of 1.0848 Å and $O-C_m-H_{plane}$ angles of 105.40°.

The B2PLYP SE structure of propanal is close to the B3LYP one, with small differences on C1−C2 (1.5022 versus 1.5037 Å) and C1−H (1.1057 versus 1.1040 Å) bond lengths, where the B2PLYP SE equilibrium parameters reproduce well the results obtained employing MP2/VTZ vibrational corrections[84] (1.5023 and 1.1056 Å).

Propene deserves some considerations. We remember here that the fit based on B3LYP/SNSD vibrational corrections has been performed on the SE equilibrium moments of inertia corresponding to the $(B_B^0)^{EXP}$ and $(B_C^0)^{EXP}$ rotational constants, due to large uncertainties affecting the $(B_A^0)^{EXP}$ of some isotopologues[85] that lead to ill-conditioned results and excluding $CHD_{cis}{=}CDCH_3$ and $CH_2{=}{}^{13}CHCH_3$ isotopologues, because of the corresponding large residuals affecting the equilibrium rotational constants. Differently, the best results for the fit with B2PLYP ΔB_{vib}^{β} corrections have been obtained excluding $CHD_{cis}{=}CDCH_3$, $CH_2{=}{}^{13}CHCH_3$, $CH_2{=}CHCH_2D_{plane}$, $CHD_{cis}{=}CHCH_2D_{plane}$, and $CHD_{cis}{=}CHCH_2D_{out}$ isotopologues. The latter results are reported in Table 5, together with the SE equilibrium structure recently evaluated employing ΔB_{vib}^{β} contributions evaluated at the MP2/VTZ(fc) level[86] and the high-level r_e structure, optimized at CCSD(T)/VQZ + MP2[wCVQZ(ae) − wCVQZ] + MP2[V-(Q,5)Z − VQZ] level for CC bond lengths, and CCC bond angle, and at CCSD(T)/VQZ + MP2[wCVQZ(ae) − wCVQZ] level for the other parameters.[86] B2PLYP SE equilibrium values show a general better agreement with the MP2 ones with respect to the B3LYP results. Some fitted geometrical parameters defining the methyl hydrogen atoms lying outside the molecular C−C−C plane, in particular the C3−H_{out} bond length (1.0895 and 1.0817 Å for B3LYP and B2PLYP, respectively) and the C1=C2−C3−H_{out} dihedral angle (120.47° and 120.70°), are significantly smaller than the corresponding values obtained using MP2 vibrational con-

DOI: 10.1021/acs.jctc.5b00622
J. Chem. Theory Comput. 2015, 11, 4689−4707

Table 4. B3LYP/SNSD and B2PLYP/VTZ r_e^{SE} and B2PLYP/VTZ r_e Geometries of Furan, Thiophene, *cis* and *trans*-Formic Acid, Peroxyformic Acid, and *cis*-Methyl Formate[a]

	r_e^{SE}[b]		r_e
	B3LYP/SNSD	B2PLYP/VTZ	B2PLYP/VTZ
	furan		
r(C2−O)	1.3598(4)[c]	1.3598(1)[c]	1.3615
r(C2=C3)	1.3542(4)	1.3551(2)	1.3569
r(C3−C4)	1.4344(19)	1.4317(0)	1.4311
r(C2−H)	1.0739(3)	1.0731(2)	1.0734
r(C3−H)	1.0743(3)	1.0751(1)	1.0748
a(C2−O−C5)	106.50(3)	106.61(2)	106.69
a(O−C2−C3)	110.79(4)	110.66(2)	110.53
a(C2−C3−C4)	105.96(6)	106.04(3)	106.12
a(H−C2−O)	115.82(3)	115.88(1)	115.90
a(H−C3−C4)	127.61(3)	127.68(1)	127.51
rms resid [MHz]	0.0010	0.0003	
mean Δ_e [uÅ^2]	0.00116	−0.00097	
	thiophene		
r(S−C2)	1.7127(5)[c]	1.7126(5)[c]	1.7215
r(C2=C3)	1.3625(9)	1.3622(8)	1.3671
r(C3−C4)	1.4233(21)	1.4233(21)	1.4210
r(C2−H)	1.0772(5)	1.0771(5)	1.0757
r(C3−H)	1.0792(3)	1.0794(3)	1.0786
a(C2−S−C5)	91.88(4)	91.88(4)	91.83
a(S−C2−C3)	111.66(3)	111.66(3)	111.45
a(C2=C3−C4)	112.40(7)	112.40(7)	112.63
a(H−C2−S)	120.06(10)	120.11(10)	120.19
a(H−C3−C4)	124.14(3)	124.14(2)	124.08
rms resid [MHz]	0.0010	0.0009	
mean Δ_e [uÅ^2]	0.00340	0.00147	
	cis-formic acid		
r(C−H)	1.0985(3)[c]	1.0981(1)[c]	1.1001
r(C=O)	1.1910(3)	1.1911(1)	1.1933
r(C−O)	1.3485(3)	1.3482(1)	1.3515
r(O−H)	0.9619(2)	0.9612(1)	0.9636
a(H−C=O)	124.21(18)	123.87(3)	124.06
a(O−C=O)	122.30(1)	122.28(1)	122.60
a(C−O−H)	109.00(2)	109.18(1)	109.29
rms resid [MHz]	0.0003	0.0001	
mean Δ_e [uÅ^2]	−0.00067	0.00008	
	trans-formic acid		
r(C−H)	1.0918(3)[c]	1.0918(2)[c]	1.0934
r(C=O)	1.1973(5)	1.1976(3)	1.1999
r(C−O)	1.3417(5)	1.3411(3)	1.3451
r(O−H)	0.9656(4)	0.9660(2)	0.9691
a(H−C=O)	125.38(22)	125.14(15)	125.31
a(O−C=O)	124.78(1)	124.81(1)	125.14
a(C−O−H)	106.78(2)	106.81(2)	106.91
rms resid [MHz]	0.0005	0.0004	
mean Δ_e [uÅ^2]	0.00057	0.00005	
	peroxyformic acid		
r(C−H)	1.0901(1)[c]	1.0905(1)[c]	1.0919
r(C=O)	1.2018(1)	1.2015(1)	1.2042
r(C−O)	1.3374(1)	1.3368(1)	1.3406
r(O−O)	1.4389(1)	1.4390(1)	1.4414
r(O−H)	0.9762(3)	0.9720(3)	0.9808
a(H−C=O)	126.98(2)	126.96(1)	127.05
a(H−C−O)	108.73(1)	108.72(1)	108.43
a(O=C−O)	124.29(2)	124.32(2)	124.52
a(C−O−O)	110.29(1)	110.31(1)	110.70
a(O−O−H)	100.56(1)	100.78(1)	100.42
rms resid [MHz]	0.0002	0.0002	
mean Δ_e [uÅ^2]	−0.01526	−0.00787	
	cis-methyl formate		
r(C_m−O)	1.4358(16)[c]	1.4347(13)[c]	1.4395
r(C−O)	1.3343(15)	1.3346(13)	1.3393
r(C_m−H_{plane})	1.0845(30)	1.0848(25)	1.0833
r(C_m−H_{out})	1.0875(8)	1.0870(7)	1.0865
r(C−H)	1.0925(18)	1.0923(15)	1.0941
r(C=O)	1.2001(17)	1.2003(14)	1.2022
a(C_m−O−C)	114.26(13)	114.30(11)	114.92
a(O−C_m−H_{plane})	105.35(42)	105.40(35)	105.55
a(O−C_m−H_{out})	110.07(5)	110.15(5)	110.32
a(O−C−H)	109.54(17)	109.58(14)	109.08
a(O−C=O)	125.50(16)	125.47(14)	125.82
d(H_{out}−C−O−C)	−60.36(8)	−60.37(7)	−60.32
rms resid [MHz]	0.0025	0.0021	

[a]Distances in angstroms, angles in degrees. Atom numbering is the same as ref 38, and it is also reported in Figure 3. [b]All fits have been performed on moments of inertia. The uncertainties on the geometrical parameters are reported within parentheses, rounded to 1×10^{-4} Å for lengths and 1×10^{-2} deg for angles if smaller than these values. $\Delta_e = I^C - I^B - I^A$ is the inertial defect. [c]Denotes the inclusion of ΔB_{el}^{β}. B3LYP/SNSD r_e^{SE} is taken from ref 38 for all molecules, with the exception of peroxyformic acid, which is from this work.

tributions (1.0949 Å and 121.08°), with B2PLYP results being in better agreement with the MP2 ones. It is noteworthy that, in contrast to the B3LYP trend, the MP2 and B2PLYP SE C3−H_{out} bond length differs significantly from the other C−H bonds, which range between 1.0805 and 1.0862 Å. On the contrary, the B2PLYP value for the C1=C2−H angle (119.07°) is larger than both B3LYP (118.84°) and MP2 (118.75°) results. The B2PLYP SE results show a fully satisfactory agreement with the CC r_e ones.

3.3. Phenylacetylene, Pyruvic Acid, and Glyoxylic Acid. The new B3LYP/SNSD and B2PLYP/VTZ SE equilibrium geometries of phenylacetylene (see Figure 3), obtained by fitting the SE I_A^e and I_B^e moments of inertia of all the isotopologues experimentally determined (see SI), are shown in Table 6. This molecule presents a planar structure (C_{2v} symmetry), completely characterized by 12 geometrical parameters. Different SE equilibrium structures of phenylacetylene have been recently proposed in ref 87, where the B3LYP/6-311+G(3df,2pd) and B3LYP/6-31G* cubic force fields have been used in the vibrational correction calculations and are compared with the high-level CCSD(T)/wCVQZ(AE) + [MP2(AE)/AwCV5Z − MP2(AE)/wCVQZ] r_e geometry, reported in Table 6. The B3LYP/SNSD and B2PLYP/VTZ r_e^{SE} geometries determined in this work show a very good agreement to one another and with the CC results.

A further system of wide interest, for which it has been possible to determine the complete SE equilibrium geometry, is pyruvic acid (see Figure 3). It is the simplest alpha-keto acid,

DOI: 10.1021/acs.jctc.5b00622
J. Chem. Theory Comput. 2015, 11, 4689−4707

Table 5. r_e^{SE} and r_e Geometries of Propene[a]

	r_e^{SE}[b]			r_e	
	MP2/VTZ[d]	B3LYP/SNSD[e]	B2PLYP/VTZ[f]	B2PLYP/VTZ	CCSD(T)[g]
		propene			
r(C1=C2)	1.3310(7)[c]	1.3326(2)[c]	1.3315(4)[c]	1.3293	1.3317
r(C2−C3)	1.4956(7)	1.4956(2)	1.4958(4)	1.4963	1.4954
r(C1−H_{cis})	1.0834(6)	1.0818(2)	1.0824(5)	1.0821	1.0825
r(C1−H_{trans})	1.0805(12)	1.0804(2)	1.0800(5)	1.0801	1.0804
r(C2−H)	1.0857(4)	1.0841(2)	1.0844(3)	1.0845	1.0845
r(C3−H_{plane})	1.0862(8)	1.0880(4)	1.0873(9)	1.0886	1.0888
r(C3−H_{out})	1.0949(9)	1.0895(7)	1.0917(16)	1.0912	1.0907
a(C1=C2−C3)	124.47(2)	124.43(1)	124.44(2)	124.96	124.481
a(C2−C1−H_{cis})	121.08(5)	121.13(2)	121.17(3)	121.36	121.149
a(C2−C1−H_{trans})	121.55(14)	121.31(3)	121.44(6)	121.52	121.4548
a(C1=C2−H)	118.75(13)	118.84(4)	119.07(7)	118.75	118.856
a(C2=C3−H_{plane})	111.10(4)	111.07(2)	111.08(3)	111.30	111.027
a(C2−C3−H_{out})	110.53(11)	111.02(7)	110.85(16)	111.00	110.972
d(C1=C2−C3−H_{out})	121.08(14)	120.47(8)	120.70(18)	120.72	120.581
rms resid [MHz]		0.0002	0.0004		

[a]Distances in angstroms, angles in degrees. See Figure 3 for atom numbering. [b]All fits have been performed on SE I_e^B and I_e^C moments of inertia. The uncertainties on the geometrical parameters are reported within parentheses. [c]Denotes the inclusion of ΔB_{el}^{β}. [d]MP2/VTZ r_e^{SE} from ref 86. [e]The CHD_{cis}=$CDCH_3$ and CH_2=$^{13}CHCH_3$ isotopomers have been excluded from the fit. [f]CHD_{cis}=$CDCH_3$, CH_2=$^{13}CHCH_3$, CH_2=$CHCH_2D_{plane}$, CHD_{cis}=$CHCH_2D_{plane}$, and CHD_{cis}=$CHCH_2D_{out}$ isotopomers have been excluded from the fit. [g]The CCSD(T) r_e has been calculated at CCSD(T)/VQZ + MP2[wCVQZ(ae) − wCVQZ] + MP2[V(Q,5)Z − VQZ] level for CC bond lengths, and CCC bond angle, and at CCSD(T)/VQZ + MP2[wCVQZ(ae) − wCVQZ] level for all other parameters, from ref 86.

Table 6. r_e^{SE} and r_e Geometries for Phenylacetylene[a]

	r_e^{SE}[b]		r_e		
	B3LYP/SNSD	B2PLYP/VTZ	B3LYP/SNSD	B2PLYP/VTZ	CCSD(T)[d]
r(C_i−C_o)	1.3988(3)[c]	1.3981(3)[c]	1.4059	1.4012	1.3985
r(C_o−C_m)	1.3887(3)	1.3892(3)	1.3921	1.3883	1.3886
r(C_m−C_p)	1.3916(10)	1.3913(11)	1.3961	1.3918	1.3915
r(C_i−C_α)	1.4311(3)	1.4316(4)	1.4304	1.4273	1.4322
r(C_α−C_β)	1.2075(1)	1.2075(1)	1.2102	1.2077	1.2075
r(C_o−H)	1.0788(4)	1.0799(4)	1.0853	1.0800	1.0806
r(C_m−H)	1.0803(2)	1.0806(2)	1.0862	1.0806	1.0808
r(C_p−H)	1.0801(1)	1.0806(1)	1.0861	1.0805	1.0808
r(C_β−H)	1.0608(1)	1.0607(1)	1.0666	1.0602	1.0618
a(C_o−C_i−C_o)	119.44(3)	119.53(3)	119.02	119.14	119.45
a(C_i−C_o−C_m)	120.12(2)	120.08(2)	120.32	120.26	120.13
a(C_o−C_m−C_p)	120.24(4)	120.22(5)	120.29	120.27	120.21
a(C_m−C_p−C_m)	119.84(4)	119.88(4)	119.75	119.80	119.87
a(C_i−C_o−H)	119.43(4)	119.38(4)	119.24	119.22	119.26
a(C_p−C_m−H)	120.14(4)	120.13(5)	120.08	120.08	120.09
a(C_m−C_o−H)	120.45(5)	120.54(5)	120.43	120.51	120.62
a(C_o−C_m−H)	119.62(1)	119.65(2)	119.63	119.65	119.70
rms resid [MHz]	0.0005	0.0004			
mean Δ_e [uÅ^2]	−0.00033	0.00258			

[a]Distances in angstroms, angles in degrees. See Figure 3 for atom numbering. [b]All fits have been performed on SE I_e^A and I_e^B moments of inertia. The uncertainties on the geometrical parameters are reported within parentheses. [c]Denotes the inclusion of ΔB_{el}^{β}. $\Delta_e = I^C - I^B - I^A$ is the inertial defect. [d]CCSD(T)/wCVQZ(AE) + [MP2(AE)/AwCV5Z − MP2(AE)/wCVQZ] r_e, from ref 87.

with a carboxylic and a ketonic functional group, and is defined by 17 geometrical parameters (C_s symmetry). Its first r_e^{SE} determinations, where the ΔB_{vib}^{β} corrections have been calculated at the B3LYP/SNSD and B2PLYP/VTZ level, are collected in Table 7, together with the B3LYP/SNSD, B2PLYP/VTZ, and CCSD(T)/CBS+CV r_e structures. The fits have been performed using the SE I_A^e, I_B^e, and I_C^e moments of inertia of $CH_3COC^{18}OOH$, $CH_3COCO^{18}OH$, and $CH_3COCOOD$ isotopologues, with 3, 2, and 1 as respective weights, and the SE I_A^e and I_C moments of inertia of all other isotopologues experimetnally investigated, with weights of 3 and 1, respectively (see SI).

Characterized by a similar molecular structure, glyoxylic acid, which is the simplest α-oxoacid (defined by 11 geometrical parameters, C_s symmetry), is also reported in Table 7. The new B2PLYP/VTZ SE equilibrium structure is compared with the B3LYP/SNSD SE determination previously presented in ref 38, and the B3LYP/SNSD, B2PLYP/VTZ, and CCSD(T)/VQZ r_e

DOI: 10.1021/acs.jctc.5b00622
J. Chem. Theory Comput. 2015, 11, 4689−4707

Table 7. r_e^{SE} and r_e Geometries for Glyoxylic and Pyruvic Acids[a]

	r_e^{SE}[b]		r_e		
	B3LYP/SNSD	B2PLYP/VTZ	B3LYP/SNSD	B2PLYP/VTZ	CCSD(T)
		glyoxylic acid[d]			
r(C1−C2)	1.5211(3)[c]	1.5244(3)[c]	1.5345	1.5276	1.5256
r(C1−H)	1.0964(3)	1.0966(3)	1.1045	1.0977	1.0963
r(C1═O)	1.2067(3)	1.2054(3)	1.2080	1.2081	1.2087
r(C2═O)	1.1994(3)	1.1967(3)	1.2034	1.2011	1.1977
r(C2−O)	1.3325(3)	1.3316(4)	1.3373	1.3356	1.3317
r(O−H)	0.9692(4)	0.9697(4)	0.9764	0.9727	0.9697
a(C2−C1−H)	115.59(2)	115.48(2)	115.13	115.13	115.41
a(C2−C1═O)	120.60(3)	120.69(3)	121.03	121.01	120.66
a(C1−C2═O)	121.95(3)	121.86(3)	121.74	121.97	121.90
a(C1−C2−O)	113.70(3)	113.48(3)	113.78	113.25	113.35
a(C−O−H)	106.84(2)	107.10(2)	107.67	106.95	106.74
rms resid [MHz]	0.0003	0.0003			
mean Δ_e [uÅ^2]	−0.01110	−0.01327			
		pyruvic acid[e]			
r(C1−C2)	1.5356(13)[c]	1.5382(8)[c]	1.5507	1.5434	1.5387
r(C1−C_m)	1.4877(12)	1.4898(7)	1.5042	1.4916	1.4893
r(C_m−H_p)	1.0812(9)	1.0819(5)	1.0904	1.0845	1.0845
r(C_m−H_o)	1.0909(3)	1.0902(2)	1.0955	1.0895	1.0893
r(C1═O)	1.2157(9)	1.2115(5)	1.2075	1.2153	1.2114
r(C2═O)	1.2019(9)	1.1980(5)	1.2091	1.2021	1.1979
r(C2−O)	1.3289(10)	1.3311(6)	1.3401	1.3346	1.3297
r(O−H)	0.9725(6)	0.9678(4)	0.9717	0.9746	0.9706
a(C1−C_m−H_p)	110.22(7)	110.19(4)	109.39	110.06	109.88
a(C1−C_m−H_o)	109.16(4)	109.26(2)	110.07	109.51	109.35
a(C2−C1−C_m)	117.27(12)	116.93(7)	114.95	116.98	116.91
a(C2−C1═O)	117.53(9)	117.75(5)	120.29	117.80	117.70
a(C1−C2═O)	122.71(9)	122.87(6)	122.84	123.10	122.80
a(C1−C2−O)	113.14(12)	112.81(7)	112.64	112.45	112.81
a(C2−O−H)	106.01(4)	106.39(3)	107.01	106.03	106.40
d(C2−C1−C_m−H_o)	57.77(6)	57.80(4)	58.36	58.01	58.07
rms resid [MHz]	0.0013	0.0008			

[a]Distances in angstroms, angles in degrees. See Figure 3 for atom numbering. [b]The uncertainties on the geometrical parameters are reported within parentheses. $\Delta_e = I^C - I^B - I^A$ is the inertial defect. [c]Denotes the inclusion of ΔB_{el}^{β}. [d]The fits have been performed using SE I_e^B and I_e^C moments of inertia. CCSD(T)/VQZ r_e from ref 88. [e]The fits have been performed using SE I_e^A, I_e^B, and I_e^C moments of inertia, with 3, 2, and 1 weight respectively, on $CH_3COC^{18}OOH$, $CH_3COCO^{18}OH$, and $CH_3COCOOD$ and SE I_e^A and I_e^C, with 3 and 1 weight respectively, on all other isotopomers. CCSD(T)/CBS+CV r_e from this work.

results, the latter taken from ref 88. All the SE equilibrium structures have been obtained by fitting the SE I_B^e and I_C^e moments of inertia of 8 out of 9 experimentally observed isotopologues, where the $H^{13}COCOOH$ isotopomer has been excluded from the fits because of the large residuals shown by its equilibrium rotational constants of the fitted geometry.

The structures of both these molecules deserve some considerations. Quite large discrepancies are found in the C1−C2 bond lengths, which are underestimated at the B3LYP/SNSD level (1.5211 and 1.5356 Å for glyoxylic and pyruvic acids, respectively) with respect to the B2PLYP/VTZ r_e^{SE} values (1.5244 and 1.5382 Å) and the CCSD(T) r_e ones (1.5256 and 1.5387 Å), and for the C═O bond lengths, where r(C2═O) in pyruvic acid shows the largest differences (1.2019, 1.1980, and 1.1979 Å for B3LYP r_e^{SE}, B2PLYP r_e^{SE}, and CCSD(T) r_e, respectively). B2PLYP/VTZ SE results are generally in better agreement with the fully theoretical values, while B3LYP/SNSD SE equilibrium geometries show significant differences.

3.4. Phenyl Radical. The SE equilibrium geometry of phenyl radical (see Figure 3) is of great interest both for testing the performances of ab initio computational methods employed in cubic force field calculations of radical species, and for providing an important biochemical building block. Phenyl radical is a planar molecule (C_{2v} symmetry), completely defined by 10 internal parameters. Its SE equilibrium geometries have been obtained performing the fits on the SE moments of inertia, derived from the ground-state rotational constants experimentally investigated corrected by CCSD(T)/ANO0, B3LYP/SNSD, and B2PLYP/VTZ ΔB_{vib}^{β} corrections, together with B3LYP/AVTZ electronic corrections.

Due to large uncertainties on experimental data, not all ground-state rotational constants of all isotopic species have been taken into account in the fitting procedures. At first, the fits have been performed using the scheme proposed in ref 79 (hereafter scheme 1), which includes the SE I_e^C moments of inertia of all nine isotopomers experimentally observed, I_e^A moments of inertia of the three C_6H_4D isotopomers, and C_6D_5, and C_6H_5 species (see Table S5 in SI). This setup is justified by the fact that the C_0 rotational constants can be determined with extreme precision for a nearly oblate top, whereas A_0 constants are most accurately known for the deuterated and normal species. The CCSD(T)/ANO0 SE

DOI: 10.1021/acs.jctc.5b00622
J. Chem. Theory Comput. 2015, 11, 4689−4707

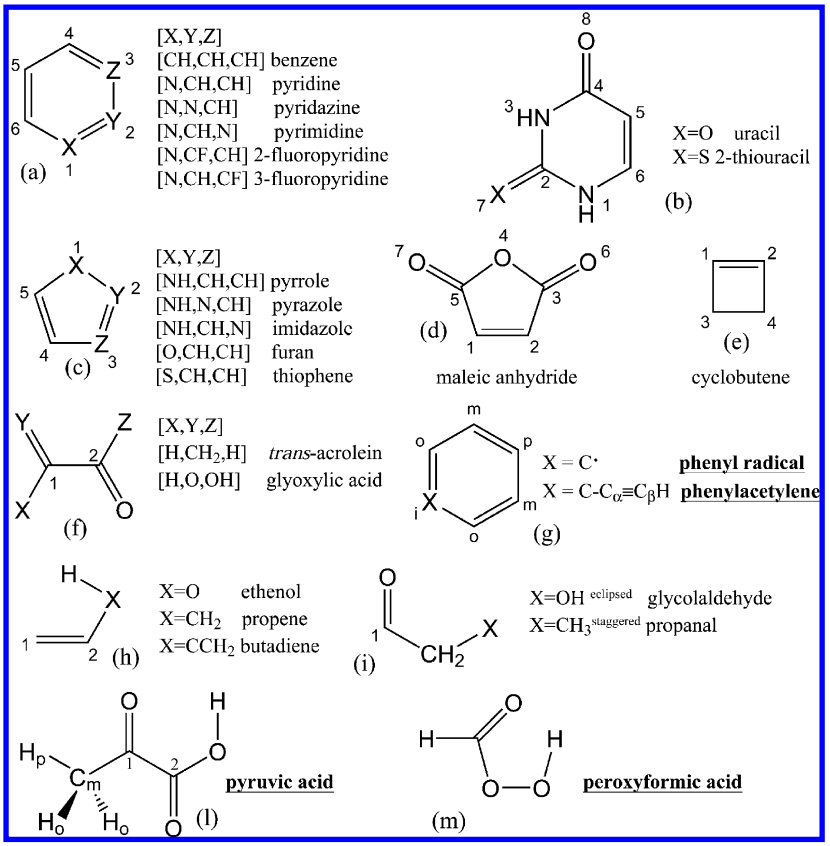


Figure 3. Sketch of the molecules for which we have used in the tables a specific atom labeling. The four new molecules included in B3se and B2se sets are underlined.

results obtained with scheme 1 are reported in Table 1. They show quite large uncertainties on fitted parameters (about ±0.0020 Å on bond lengths and ±0.20° on angles), and an increasing order in magnitude for the CH bond lengths of $r(C_o–H) < r(C_m–H) < r(C_p–H)$ (1.0805, 1.0814, and 1.0818 Å), which is in disagreement with that of all r_e geometries (optimized at the B3LYP/SNSD, B2PLYP/VTZ, and CCSD(T) levels), where $r(C_o–H)$ is approximately equal to $r(C_p–H)$ and significantly shorter than $r(C_m–H)$. It has been found that this result is due to a poor compatibility of the A_0 constants of the three monodeuterated species. Then, different fits have been performed starting from scheme 1 and further removing from fit procedures one, two or all three A_0 constants of the *o*-, *p*-, or *m*-C_6H_4D isotopologues. The most satisfactory and stable results have been obtained removing the A_0 constant of *m*-C_6H_4D isotopomer (hereafter scheme 2). The B3LYP/SNSD, B2PLYP/VTZ, and CCSD(T)/ANO0 SE equilibrium geometries obtained with scheme 2 are reported in Table 1. They show a significant reduction of the RMS residuals as well as of the uncertainties affecting geometrical parameters. Moreover, the trends of CH bond lengths in r_e geometries is well reproduced, and the three different SE equilibrium geometries are in very good agreement to one another. The largest differences between scheme 2 and scheme 1 results concern $a(C_i–C_o–H)$ (122.35 versus 122.90°, for CCSD(T)/ANO0 r_e^{SE}) and $a(C_o–C_m–H)$ (119.83° versus 120.39°), where for both parameters the values issued from scheme 1 are in better agreement with r_e results.

It is noteworthy that there are non-negligible differences between the CCSD(T)/CVQZ[79] and CCSD(T)/CBS+CV +aug[89] equilibrium parameters. Specifically, the CC bond lengths calculated at the CBS+CV+aug level are shorter than the CCSD(T)/CVQZ ones (1.3693 Å versus 1.3741 Å for $r(C_i–C_o)$, 1.3947 versus 1.3993 Å for $r(C_o–C_m)$, and 1.3882 versus 1.3926 Å for $r(C_m–C_p)$), where the latter are in a better agreement with the r_e^{SE} results. The r_e geometries optimized at the B3LYP/SNSD and B2PLYP/VTZ level are also reported in Table 1 in order to be easily accessible for the template approach presented in ref 38, and that will be further discussed in this paper.

Fully theoretical equilibrium geometries for phenyl radical have been also optimized at the unrestricted MP2 (UMP2) and restricted-open shell MP2 (ROMP2) levels, in conjunction with the VTZ basis set, to test the performances of both methods in vibrational correction estimations. The UMP2 method leads to an unreliable equilibrium geometry, showing too short CC bond lengths (by about 0.0260–0.0300 Å with respect to the CCSD(T) r_e^{SE} values), and does not allow to achieve a good cubic force field for SE determinations. On the other hand, the ROMP2 geometrical parameters are in better agreement with the CCSD(T) SE ones, showing quite large values for the CC bond lengths (1.3754, 1.4011, and 1.3943 Å for $r(C_i–C_o)$, $r(C_o–C_m)$, and $r(C_m–C_p)$, respectively), and a satisfactory agreement for the CH bond lengths (1.0807, 1.0819, and 1.0811 Å for $r(C_o–H)$, $r(C_m–H)$, and $r(C_p–H)$, respectively) and all bond angles. Unfortunately, the lack of implementation of the analytical second-order energy derivatives for ROMP2, in most computational softwares, makes this method not eligible to achieve the necessary precision for cubic force fields, and then not usable for SE equilibrium determinations.

DOI: 10.1021/acs.jctc.5b00622
J. Chem. Theory Comput. 2015, 11, 4689–4707

4. TOWARD LARGER SYSTEMS

A step further regarding the setup of a complete collection of biomolecule building blocks is represented by uracil, the simplest DNA base. Its equilibrium structure has been investigated using both theoretical and experimental data.[90] Characterized by a planar structure (C_s symmetry), uracil is completely defined by 21 geometrical parameters. Its 10 isotopomers experimentally investigated (where the isotopic substitutions are on the carbon, nitrogen and oxygen atoms, see SI) provide 30 ground-state rotational constants, but only two of the three moments of inertia for each isotopic species are independent due to the planarity, and only 20 rotational constants can be included in the fit. The insufficient number of rotational constants imposes to fix some parameters during the fitting procedures, in particular those defining the position of the H atoms, due to the lack of experimental data on deuterated isotopologues.

It has been already pointed out in our previous work that in order to achieve good accuracy it is important to fix molecular parameters to the r_e values computed at very high levels of theory.[38] Thus, we decided to fix the position of H atoms at the best ab initio value available in the literature, optimized at the CCSD(T)/CBS+CV level,[90] using the distances from the corresponding ring atoms (Y) and the δ as parameters. A δ coordinate is defined as the angle between the line passing through H and Y atoms, and the bisector the XYZ in-ring angle, where Y is in the vertex of the angle (for a pictorial representation of δ coordinates see Figure 4). Note that,

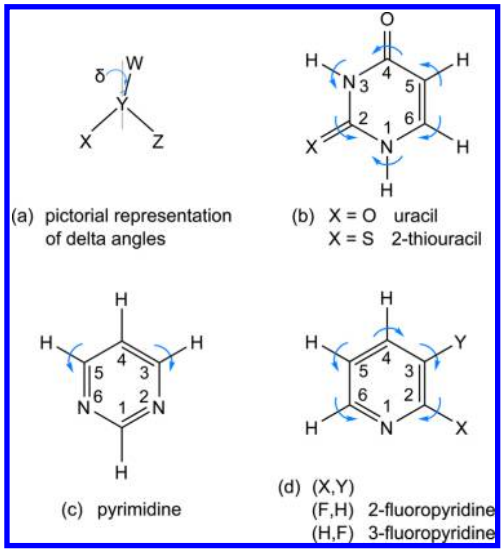


Figure 4. Sketch of delta angles used in this work, with arrows showing the sign convention adopted in various cases.

when more than one δ is present in the molecule, the label of the corresponding Y atom is added as subscript. The use of the δ coordinates to describe terminal out-of-ring atoms, when dealing with planar cycles, introduces several advantages. First of all, it removes the arbitrariness in the choice of the angles defining the out-of-ring atom (i.e., the choice of using XYH or ZYH angles); then it preserves the local symmetry during fitting optimization when present (i.e., when the H atom is constrained on the bisector due to the symmetry). Finally, differently from the use of the XYH angle, it accounts for the variation of the fitted XYZ angle. Indeed, a great stability in the fitting results has been found when using δ parameters, for which the sign convention adopted is indicated in Figure 4, together with atom numbering.

In this framework, the fits of uracil have been performed on its I_0^A and I_0^C moments of inertia, corrected by B3LYP/SNSD and B2PLYP/VTZ vibrational contribution as well as by B3LYP/AVTZ electronic contributions. In Table 8, the new B3LYP/SNSD and B2PLYP/VTZ SE equilibrium geometries are compared with the results of previous determinations, where the H atoms have been fixed using the XYH or ZYH bond angles.[90] All the SE equilibrium structures are in very good agreement.

4.1. Templating Molecule Approach. In our previous work[38] we proposed the template approach (hereafter referred to as templating molecule approach, TMA) for situations where lack of experimental data for some isotopic species does not allow the determination of all geometrical parameters defining a molecular system. Within this method, the parameters kept fixed in the fits are determined by

$$r_e(\text{fixed}) = r_e + \Delta\text{TM} \tag{3}$$

with Δ TM defined as

$$\Delta\text{TM} = r_e^{SE}(\text{TM}) - r_e(\text{TM}) \tag{4}$$

r_e being the geometrical parameter of interest calculated at the same level for both the molecule under consideration and that chosen as a reference, denoted as templating molecule (TM), which contains the functional group under investigation. Let us take 2-thiouracil as a case study to sketch the main features of the TMA.

2-Thiouracil is a widely studied molecule because of its important applications in medicine. It has a planar structure (C_s symmetry) completely defined by 21 internal parameters and skeched in Figure 4. Unfortunately, the ground-state rotational constants experimentally observed for 2-thiouracil are available only for two isotopic species (parent and ^{34}S substituted species), which can provide accurate informations within the SE approach just on the thioketone moiety of the molecule. In other words, as previously pointed out,[67] only two parameters can be determined at SE level, so that this molecule represents a very challenging example for the TMA. In order to obtain an accurate equilibrium structure without the use of expensive computations, the TMA has been applied to fix all parameters except r(C2−S7) and δ_2 of 2-thiouracil, using uracil as TM. Results are shown in Table 9. It is noteworthy that the use of δs angles rather than one of the two possible a(X−Y−Z) angles provides us with an unambiguous choice for the fixed parameters. This is a remarkable improvement compared with the procedure followed in our previous work,[38] because we have found that in some cases, in particular when the fitting problem is ill-conditioned, the resulting geometry can vary depending on the choice of fixed angles. As displayed in Table 9, we have applied the TMA to 2-thiouracil using both B3LYP/SNSD and B2PLYP/VTZ as theoretical method, taking the r_e^{SE}(B2PLYP/VTZ) structure of uracil reported in Table 8 as TM for deriving the parameters to be fixed according to eq 4. We also report, for comparison, the geometry obtained fixing the nondeterminable parameters at the CCSD(T)/CBS+CV +diff+T level. All these three partial SE structure are in a satisfactory agreement with each other and also with the totally ab initio CCSD(T)/CBS+CV+diff+T structure.[67] This is a remarkable result because it shows not only that with the TMA one can obtain very accurate equilibrium structures of

DOI: 10.1021/acs.jctc.5b00622
J. Chem. Theory Comput. 2015, 11, 4689−4707

Table 8. r_e and r_e^{SE} Equilibrium Geometries for Uracil[a]

	r_e^{SE}[b]				r_e	
	B3LYP/N07D[e,f]	B3LYP/SNSD[f]	B2PLYP/VTZ[f]	B2PLYP/VTZ[g]	B2PLYP/VTZ	CCSD(T)[e]
r(N1−C2)	1.3818(5)	1.3814(4)[d]	1.3817(2)[d]	1.3817(2)[d]	1.3880	1.3785
r(C2−N3)	1.3763	1.3768(3)	1.3763(6)	1.3762(6)	1.3796	1.3756
r(N3−C4)	1.3979(4)	1.3980(5)	1.3989(2)	1.3990(2)	1.4067	1.3974
r(C4−C5)	1.4550(6)	1.4564(5)	1.4555(2)	1.4554(2)	1.4541	1.4539
r(C5−C6)	1.3450(6)	1.3419(6)	1.3440(3)	1.3443(3)	1.3452	1.3433
r(C6−N1)	1.3720(6)	1.3734(6)	1.3712(3)	1.3711(3)	1.3707	1.3723
r(C2−O)	1.2103(2)	1.2101(2)	1.2097(1)	1.2096(1)	1.2127	1.2112
r(C4−O)	1.2128(2)	1.2121(3)	1.2124(1)	1.2123(1)	1.2157	1.2138
r(N1−H)	1.0046[c]	1.0046[c]	1.0046[c]	1.0043[c]	1.0053	1.0046
r(N3−H)	1.0090[c]	1.0090[c]	1.0090[c]	1.0085[c]	1.0095	1.0090
r(C5−H)	1.0766[c]	1.0766[c]	1.0766[c]	1.0759[c]	1.0756	1.0766
r(C6−H)	1.0793[c]	1.0793[c]	1.0793[c]	1.0793[c]	1.0792	1.0793
a(C2−N1−C6)	123.37(2)	123.38(2)	123.47(1)	123.47(1)	123.57	123.38
a(N1−C6−C5)	121.92(1)	121.90(1)	121.92(1)	121.92(1)	121.99	121.91
a(C6−C5−C4)	119.52(2)	119.62(2)	119.55(1)	119.54(1)	119.81	119.49
a(C5−C4−N3)	113.86(2)	113.77(2)	113.79(1)	113.79(1)	113.47	113.97
a(C4−N3−C2)	127.94	127.96(6)	127.98(3)	127.99(3)	128.25	127.75
a(N3−C2−N1)	113.38	113.36(5)	113.29(2)	113.29(2)	112.90	113.51
a(N1−C2−O)	123.88(4)	123.86(4)	123.86(2)	123.86(2)	122.83	123.62
a(C5−C4−O)	125.77(5)	125.79(5)	125.81(2)	125.80(2)	126.18	125.83
a(C2−N1−H)	115.22[c]	115.22(1)	115.17(1)	115.21(1)	115.17	115.22
a(C2−N3−H)	115.70[c]	115.60(3)	115.58(1)	115.71(1)	115.58	115.70
a(C6−C5−H)	122.11[c]	122.04(1)	122.08(1)	121.93(1)	121.80	122.11
a(N1−C6−H)	115.34[c]	115.34(1)	115.33(1)	115.44(1)	115.40	115.34
δ_1	3.09	3.09[c]	3.09[c]	3.05[c]	3.05	3.09
δ_3	0.33	0.43[c]	0.43[c]	0.30[c]	0.30	0.43
δ_5	1.87	1.85[c]	1.85[c]	1.70[c]	1.70	1.85
δ_6	3.70	3.71[c]	3.71[c]	3.60[c]	3.60	3.71
rms resid [MHz]		0.0005	0.0002	0.0002		
mean Δ_e [uÅ^2]		−0.02547	−0.01082	−0.01082		

[a]Distances in angstroms, angles in degrees. Atom numbering and sign convention on δ angles are reported in Figure 4. [b]All fits have been performed on SE I_e^A and I_e^C moments of inertia, weighted respectively as 3 and 1 in the fits performed in this work. The digits within parentheses are the uncertainties on the geometrical parameters. [c]Denotes that the parameters are kept fixed. [d]Indicates the inclusion of ΔB_{el}^{β} and $\Delta_e = I^C - I^B - I^A$ is the inertial defect. [e]B3LYP/N07D r_e^{SE} and CCSD(T)/CBS+CV r_e from ref 90. [f]Fixed parameters at CCSD(T)/CBS+CV r_e values. [g]CH bond lengths fixed at LSR r_e values (see Table 11), and δs at B2PLYP/VTZ r_e values.

molecules when there is a scarcity of experimental data, but also that it becomes possible using lower levels of theory. In particular, the B2PLYP/VTZ computational model gives a final RMS of the residuals about 5 times lower than in the other cases, denoting a better agreement of these parameters with the experimental data. As an additional remark it is noteworthy that, although in Table 8 we have only reported fits obtained using B2PLYP/VTZ ΔB_{vib}^{β} corrections, the use of the B3LYP/SNSD ones gives almost the same results, with a much reduced computational effort. In conclusion, in order to get reliable equilibrium structures of medium-large molecules lacking experimental data to reach a complete SE determination, a good strategy could be a kind of hybrid template approach that uses B2PLYP/VTZ equilibrium geometries and B3LYP/SNSD ΔB_{vib}^{β} corrections.

4.2. Linear Regression Approach. After generalization of eq 3 to

$$r_e(\text{fixed}) = r_e + \Delta r \tag{5}$$

some considerations can be made on the Δr and regarding the choice of the r_e geometries. Using the B3th and B2th sets, we systematically studied the errors affecting the B3LYP/SNSD and B2PLYP/VTZ r_e geometries, taking the B2PLYP/VTZ r_e^{SE} parameters as references. In particular, we estimated the mean errors (ME) and the standard deviations (SD) of the error distributions for the geometrical parameters of interest, as well as, the root-mean-square errors (RMSE), which can be used as indicators of the mean errors affecting the r_e determinations. Moreover, we studied the linear dependence of the errors on the r_e values from the linear regressions of the fitting of B2PLYP/VTZ r_e^{SE} parameters versus the corresponding r_e ones. In this way, Δr can be estimated as

$$\Delta r = A r_e + B \tag{6}$$

A and B parameters being (1 − slope) and the intercept for the linear regressions. Introduction of eq 6 in eq 5 defines the linear regression approach (LRA). A measure of the errors affecting parameters derived by the LRA is given by the standard error of the estimate (SEE), which also indicates if the errors are linearly correlated to r_e values (low SEE) or not (hight SEE).

The ME, SD, RMSE, and SEE for CH, CC, CO, CN, OH, NH, CF, and CCl bond lengths are reported in Table 10, together with the A and B parameters for linear regressions. As expected, B3LYP r_e results show significantly larger RMSE with respect to the B2PLYP r_e ones for all kinds of bonds. The ME for B2PLYP/VTZ errors are about half the ME of B3LYP/

DOI: 10.1021/acs.jctc.5b00622
J. Chem. Theory Comput. 2015, 11, 4689−4707

Table 9. r_e and r_e^{SE} Equilibrium Geometries for 2-Thiouracil[a]

	r_e^{SE}[b]			r_e		
	B2PLYP/VTZ[d]	B2PLYP/VTZ[e]	B2PLYP/VTZ[f]	B3LYP/SNSD	B2PLYP/VTZ	CCSD(T)[g]
r(N1−C2)	1.3654[c]	1.3659[c]	1.3665[c]	1.3770	1.3728	1.3654
r(C2−N3)	1.3631	1.3621	1.3623	1.3688	1.3656	1.3631
r(N3−C4)	1.4017	1.4034	1.4037	1.4157	1.4115	1.4017
r(C4−C5)	1.4522	1.4538	1.4536	1.4560	1.4522	1.4522
r(C5−C6)	1.3448	1.3447	1.345	1.3505	1.3462	1.3448
r(C6−N1)	1.3714	1.3706	1.3706	1.3744	1.3701	1.3714
r(C2−S)	1.6501(1)	1.6489(1)	1.6479(1)	1.6693	1.6567	1.6496
r(C4−O)	1.213	1.2110	1.2112	1.2176	1.2145	1.2130
r(N1−H)	1.0058	1.0053	1.0056	1.0115	1.0063	1.0058
a(N3−H)	1.0098	1.0093	1.0096	1.0149	1.0101	1.0098
a(C5−H)	1.0768	1.0767	1.0768	1.0813	1.0758	1.0768
a(C6−H)	1.0796	1.0792	1.0792	1.0847	1.0791	1.0796
a(C2−N1−C6)	123.69	123.72	123.71	123.78	123.81	123.69
a(N1−C6−C5)	121.67	121.72	121.7	121.71	121.77	121.67
a(C6−C5−C4)	119.24	119.26	119.32	119.51	119.58	119.24
a(C5−C4−N3)	113.84	113.66	113.65	113.58	113.33	113.84
a(C4−N3−C2)	127.78	127.93	127.91	127.88	128.18	127.78
a(N3−C2−N1)	113.78	113.70	113.73	113.54	113.34	113.78
a(N1−C2−S)	123.74(1)	123.70(1)	123.74(1)	122.29	122.41	123.69
a(C5−C4−O)	126.17	126.17	126.21	126.51	126.58	126.17
a(C2−N1−H)	115.18	115.24	115.15	115.25	115.15	115.18
a(C2−N3−H)	116.06	116.09	116.01	116.20	116.01	116.06
a(C6−C5−H)	122.19	122.21	122.17	122.07	121.89	122.19
a(N1−C6−H)	115.29	115.23	115.24	115.22	115.31	115.29
δ_1	2.975	2.89	2.982	2.86	2.94	2.98
δ_2	0.63(1)	0.55(1)	0.60(1)	−0.14	0.92	0.58
δ_3	0.05	−0.05	0.026	1.82	−0.10	0.05
δ_4	3.09	3.00	3.035	3.92	3.24	3.09
δ_5	1.81	1.83	1.833	3.30	1.68	1.81
δ_6	3.875	3.91	3.911	0.93	3.80	3.88
rms resid [MHz]	0.0077	0.0076	0.0017			
mean Δ_e [uÅ^2]	−0.03384	−0.03384	−0.03384			

[a]Distances in angstroms, angles in degrees. Atom numbering and sign convention on δ angles are reported in Figure 4. [b]All fits have been performed on SE I_e^A and I_e^B moments of inertia, weighted respectively as 3.7 and 1.3. The digits within parentheses are the uncertainties on the fitted geometrical parameters. [c]Denotes the inclusion of ΔB_{el}^{β} and $\Delta_e = I^C - I^B - I^A$ is the inertial defect. [d]Fixed parameters at CCSD(T)/CBS+CV+diff+T r_e values. [e]r_e(fixed) = r_e(B3LYP/SNSD) + ΔTM; ΔTM = r_e^{SE}(B2PLYP/VTZ) − r_e(B3LYP/SNSD), with uracil as TM. [f]r_e(fixed) = r_e(B2PLYP/VTZ) + ΔTM; ΔTM = r_e^{SE}(B2PLYP/VTZ) − r_e(B2PLYP/VTZ), with uracil as TM. [g]CCSD(T)/CBS+CV+diff+T r_e from ref 67.

SNSD for all typologies of bonds. At variance, the differences between the B3LYP and B2PLYP SEEs, affecting the LRC r_e parameters, are quite small, with values that never exceed ±0.0020 Å. The decreasing of the errors moving from r_e to r_e evaluations based on the LRA is particularly relevant for B3LYP/SNSD CH and NH bond lengths, where the RMSE (0.0065 and 0.0068 Å, respectively) are about 10 times larger than the corresponding SEE (0.0009 and 0.0006 Å). The improvements of the r_e values obtained using the LRA can be observed also from Figure 5, where the frequencies of the Δr for r_e and LRA r_e are graphically reported for CH bond lengths.

In the light of the small SEE shown, the CH bond lengths for uracil, 2-thiouracil, and 2- and 3-fluoropiridine have been determined as LRA r_e and compared with the CCSD(T) ones and with those estimated within the template approach in Table 11. The agreement between the estimated parameters, both within the template and linear regression approaches, and the fully ab initio ones is remarkable, showing the feasibility of using LRA r_e parameters to fix the CH bond lengths.

The results for bond angles are less satisfactory. In the majority of cases, the SEEs are not very different from the corresponding RMSEs, which indicates that the corrections provided by the linear regressions do not significantly improve the parameters, and then the errors in r_e determinations for bond angles are not as systematic as for bond lengths. The *A* and *B* parameters of linear regression estimations for the bond angles with B2PLYP/VTZ SEEs smaller than 0.20 degrees are shown in Table 12. The angles including one N atom, i.e. NCH, NCC, and CNC, present the lowest SEEs, both at the B3LYP/SNSD and B2PLYP/VTZ levels, which are not very different from the corresponding RMSEs. The largest improvement of the error is for COH angles, where the RMSE for r_e (0.54°) decreases to 0.07° for LRA r_e estimations. Moreover, the comparison of RMSEs and SEEs for HCH and CCC angles shows that r_e parameters are not significantly different from LRA r_e ones, calculated by eq 5.

Finally, the study of the errors affecting δ parameters for planar cyclic molecules leads to the result that also for these parameters the Δr corrections given by eq 6 are negligible. However, differently from bond angles, the RMSEs affecting the δ parameters calculated at the B2PLYP/VTZ level are about 0.10°. This results confirms that δ parameters are generally

DOI: 10.1021/acs.jctc.5b00622
J. Chem. Theory Comput. 2015, 11, 4689−4707

Table 10. Statistical Parameters for CH, CC, CO, CN, OH, NH, CF, and CCl Bond Lengths[a]

	r_e			r_e	
	B3LYP	B2PLYP		B3LYP	B2PLYP
	CH bonds (82 items)			OH bonds (8 items)	
ME	0.0063	0.0002	ME	0.0078	0.0045
SD	0.0013	0.0011	SD	0.0027	0.0024
RMSE	0.0065	0.0011	RMSE	0.0082	0.0051
A	−0.093027	−0.074795	*A*	−0.281475	−0.289543
B	0.095014	0.080798	*B*	0.265886	0.276062
R^2	0.988016	0.990739	R^2	0.807087	0.911140
SEE	0.0009	0.0008	SEE	0.0021	0.0014
	CC bonds (45 items)			NH bonds (7 items)	
ME	0.0048	0.0001	ME	0.0067	0.0010
SD	0.0035	0.0028	SD	0.0009	0.0007
RMSE	0.0059	0.0028	RMSE	0.0068	0.0012
A	−0.023958	−0.015977	*A*	0.101599	0.014185
B	0.028995	0.022342	*B*	−0.109546	−0.015295
R^2	0.998736	0.999063	R^2	0.990162	0.988951
SEE	0.0029	0.0025	SEE	0.0006	0.0007
	CO bonds (30 items)			CF bonds (6 items)	
ME	0.0049	0.0027	ME	0.0103	0.0025
SD	0.0029	0.0015	SD	0.0016	0.0011
RMSE	0.0057	0.0031	RMSE	0.0105	0.0027
A	−0.015281	−0.004834	*A*	−0.065292	−0.000953
B	0.014730	0.003455	*B*	0.077948	−0.001231
R^2	0.999250	0.999746	R^2	0.983616	0.990184
SEE	0.0025	0.0015	SEE	0.0014	0.0011
	CN bonds (14 items)			CCl bonds (5 items)	
ME	0.0043	0.0015	ME	0.0328	0.0168
SD	0.0022	0.0018	SD	0.0029	0.0030
RMSE	0.0048	0.0023	RMSE	0.0329	0.0171
A	−0.003991	0.006962	*A*	−0.115068	−0.108290
B	0.000901	−0.010544	*B*	0.172324	0.174512
R^2	0.999539	0.999714	R^2	0.994461	0.989961
SEE	0.0021	0.0017	SEE	0.0014	0.0019

[a]B3LYP and B2PLYP in conjunction respectively with SNSD and VTZ basis sets. ME and SD stand for mean and standard deviation of the (r_e − r_e^{SE}) differences, while RMSE is the root-mean-square error. *A* and *B* are (1 − slope) and the intercept for the linear regression of the fit of B2PLYP/VTZ r_e^{SE} parameters, taken as references, versus the corresponding B3LYP/SNSD and B2PLYP/VTZ r_e values, and SEE is the corresponding standard error of the estimations.

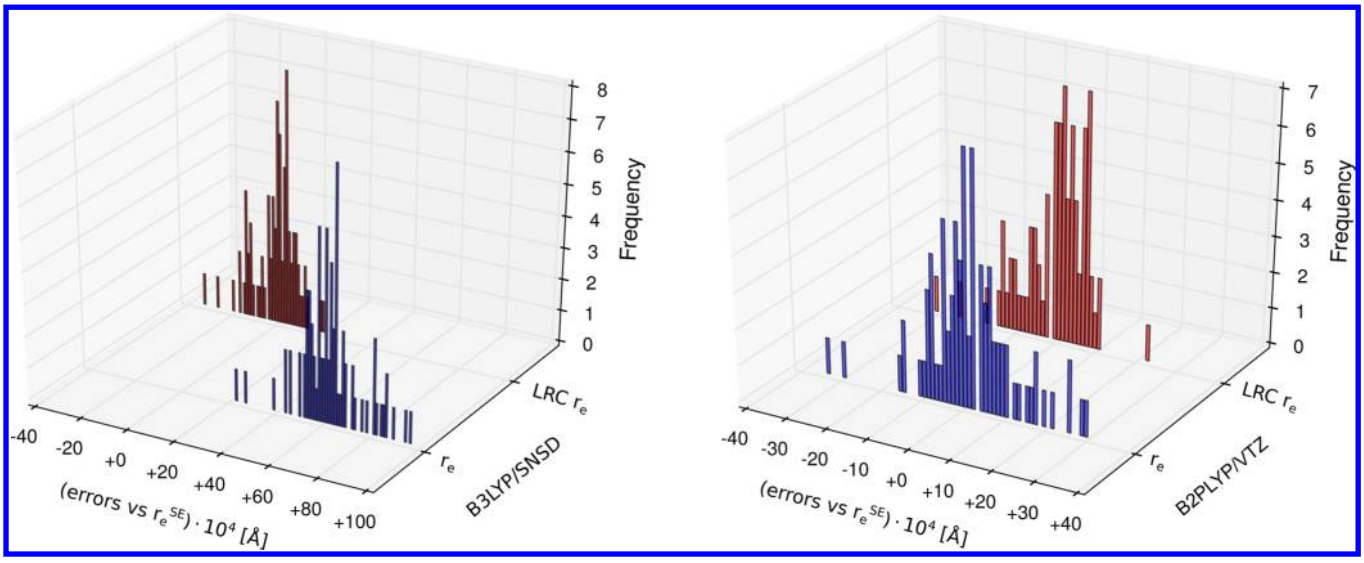


Figure 5. Statistical distributions of B2PLYP/VTZ and B3LYP/SNSD r_e and LRE(r_e) deviations from the SE equilibrium parameters (B2PLYP/VTZ r_e^{SE}) for the molecules belonging to the B2se set.

DOI: 10.1021/acs.jctc.5b00622
J. Chem. Theory Comput. 2015, 11, 4689−4707

Table 11. r_e Parameters Calculated with Template and Line Regressions Approaches

	r_e	r_e(fixed) = r_e + Δ(TM)		r_e(fixed) = r_e + Δr	
	CCSD(T)	B3LYP	B2PLYP	B3LYP	B2PLYP
		uracil[a]			
r(C5−H)	1.0766	1.0757	1.0757	1.0756	1.0760
r(C6−H)	1.0793	1.0794	1.0793	1.0790	1.0793
r(N1−H)	1.0046	1.0039	1.0041	1.0040	1.0043
r(N3−H)	1.0090	1.0077	1.0083	1.0082	1.0085
		2-fluoropiridine[b]			
r(C3−H)	1.0787	1.0781	1.0785	1.0781	1.0784
r(C4−H)	1.0807	1.0801	1.0803	1.0800	1.0803
r(C5−H)	1.0794	1.0788	1.0792	1.0789	1.0791
r(C6−H)	1.0815	1.0809	1.0815	1.0814	1.0816
		3-fluoropiridine[b]			
r(C2−H)	1.0817	1.0811	1.0813	1.0815	1.0818
r(C4−H)	1.0797	1.0791	1.0795	1.0793	1.0796
r(C5−H)	1.0800	1.0794	1.0797	1.0796	1.0799
r(C6−H)	1.0814	1.0808	1.0811	1.0815	1.0816
		2-thiouracil[c]			
r(C5−H)	1.0768	1.0767	1.0768	1.0757	1.0761
r(C6−H)	1.0796	1.0792	1.0792	1.0788	1.0792
r(N1−H)	1.0058	1.0053	1.0056	1.0047	1.0053
r(N3−H)	1.0098	1.0093	1.0096	1.0085	1.0091

[a]CCSD(T)/CBS+CV r_e from ref 90, pyrrole as TM. [b]CCSD(T)/CBS +CV r_e from ref 38, pyridine as TM. [c]CCSD(T)/CBS+CV+diff+T r_e from ref 67, uracil as TM.

affected by smaller errors with respect to the corresponding angles, describing the out-of-cycle atoms and can be thus confidently used in the template approach.

Some examples of application of the linear regression approach are given in Tables 13, 14, and 15. New SE equilibrium structures for pyrimidine (C_{2v} symmetry, see Figure 3) are presented in the first table. For this molecule, the lack of experimental data on deuterated species does not allow the determination of all CH bond lengths and the corresponding angles within the complete SE approach. Differently from the fits proposed in ref 38, where r(C2−H), r(C4−H), r(C5−H), and a(N3−C4−H) have been fixed by use of eq 3 taking pyridine as TM, in the fits proposed in this work the three CH bond lengths have been estimated by eq 5 for B2PLYP/VTZ (see Table 11), and δ_4 has been used to describe the position of H atom bonded to the C4 atom. The fits for equilibrium structure have been obtained by correcting the experimental rotational constants with Δ^{β}_{vib} contributions both at the B3LYP/SNSD and B2PLYP/VTZ levels, and fitting the I^A_e and I^C_e moments of inertia, with the inclusion of Δ^{β}_{el} corrections. The r^{SE}_e structure determined using B3LYP/6-311+G(3df,2pd) Δ^{β}_{vib} and the predicate approach, taken from ref 91, is also reported in the table for comparison purposes. The agreement among all SE structures is remarkable.

The accuracy of SE equilibrium geometries obtained fixing δ parameters at the B2PLYP/VTZ level, thus avoiding expensive CC computations, has been tested also on 2- and 3-fluoropiridine (C_s symmetries). The results are shown in Tables 14 and 15, while in Figure 3 atom numbering is shown together with sign convention on δ angles. Because of the lack of rotational data for deuterated species (see SI), for the first system only one of the eight parameters defining the out-of-ring H atoms can be fitted, while for 3-fluoropiridine also the CF bond length must be fixed in order to converge the fitting procedure. As done before, the fits have been performed on the SE I^A_e and I^C_e moments of inertia for 2-fluoropiridine, and on the SE I^A_e and I^B_e moments of inertia for 3-fluoropiridine. First fits have been performed using both B3LYP/SNSD and B2PLYP/VTZ Δ^{β}_{vib} corrections, together with Δ^{β}_{el} corrections, and fixing the CH bond lengths and δ angles describing the H atoms at the CCSD(T)/CBS+CV r_e values.[38] The agreement with the previously determined SE equilibrium structures, where the valence angles have been used instead of δ parameters, is fully

Table 12. A and B Parameters of Linear Regression Estimations for All Angles with B2PLYP/VTZ SEEs Smaller than 0.20°

	r_e[a]			r_e[a]	
	B3LYP	B2PLYP		B3LYP	B2PLYP
	HCH angles (29 items)			NCC angles (8 items)	
RMSE	0.53	0.35	RMSE	0.15	0.13
A	0.020715	0.018715	A	0.000544	0.000206
B	−1.993258	−1.819612	B	0.059088	0.051879
R^2	0.992041	0.998167	R^2	0.999790	0.999969
SEE	0.40	0.19	SEE	0.09	0.11
	CCC angles (22 items)			CNC angles (5 items)	
RMSE	0.47	0.24	RMSE	0.26	0.17
A	−0.003057	−0.001425	A	−0.000524	−0.002065
B	0.089102	0.018197	B	−0.167985	0.062961
R^2	0.999644	0.999912	R^2	0.999959	0.999981
SEE	0.39	0.19	SEE	0.13	0.09
	NCH angles (11 items)			COH angles (5 items)	
RMSE	0.20	0.20	RMSE	0.54	0.20
A	0.006236	0.003990	A	−0.055857	−0.116754
B	−0.882748	−0.628329	B	5.504899	12.607223
R^2	0.998033	0.998477	R^2	0.995909	0.989637
SEE	0.15	0.13	SEE	0.07	0.12

[a]B3LYP and B2PLYP in conjunction respectively with SNSD and VTZ basis sets. SEE and RMSE stand for standard error of the estimate and root-mean-square error, respectively. Pyruvic acid has been excluded from all analysis, as well as, the all linear molecules. r_e(fixed) = r_e + Δr, where Δr = Ar_e + B

DOI: 10.1021/acs.jctc.5b00622
J. Chem. Theory Comput. 2015, 11, 4689−4707

Table 13. r_e and r_e^{SE} Equilibrium Geometries for Pyrimidine[a]

	r_e^{SE}[b]				r_e
	B3LYP/6-311+G[e]	B3LYP/SNSD[f]	B3LYP/SNSD[g]	B2PLYP/VTZ[g]	B2PLYP/VTZ
r(C2−N3)	1.3331(3)[c]	1.3334(1)[c]	1.3334(3)[c]	1.3333(3)[c]	1.3343
r(N3−C4)	1.3349(6)	1.3355(1)	1.3353(2)	1.3350(2)	1.3355
r(C4−C5)	1.3874(4)	1.3867(3)	1.3869(3)	1.3868(2)	1.3878
r(C2−H)	1.0820(23)	1.0822[d]	1.0820[d]	1.0820[d]	1.0821
r(C4−H)	1.0843(17)	1.0826[d]	1.0825[d]	1.0825[d]	1.0826
r(C5−H)	1.0799(23)	1.0795[d]	1.0794[d]	1.0794[d]	1.0793
a(C2−N3−C4)	115.71(2)	115.69(1)	115.69(1)	115.70(1)	115.76
a(N3−C4−C5)	122.28(2)	122.27(2)	122.27(1)	122.27(1)	122.28
a(C4−C5−C6)	116.66(4)	116.72(1)	116.71(1)	116.70(1)	116.68
a(N1−C2−H)	116.32(2)	116.31(1)	116.31(1)	116.32(1)	116.37
a(N3−C4−H)	116.48(2)	116.37[d]	116.53(1)	116.52(1)	116.52
a(C4−C5−H)	121.67(2)	121.64(1)	121.65(1)	121.65(1)	121.67
δ_4	2.38	2.49	2.34[d]	2.34[d]	2.34
rms resid [MHz]		0.0003	0.0003	0.0002	
mean Δ_e [uÅ^2]		−0.00133	−0.00123	−0.00266	

[a]Distances in angstroms, angles in degrees. Atom numbering and sign convention on δ angles are reported in Figure 4. [b]All fits have been performed on SE I_e^A and I_e^C moments of inertia. [c]Indicates the inclusion of ΔB_{el}^{β}. The digits within parentheses are the uncertainties on the geometrical parameters. [d]Denotes that the parameters are kept fixed. $\Delta_e = I^C - I^B - I^A$ is the inertial defect. [e]r_e^{SE} structure determined using B3LYP/6-311+G(3df,2pd) ΔB_{vib}^{β} and the predicate approach, from ref 91. [f]r_e^{SE} structure indicates as fit2 in Table 8 of ref 38 (see also Table 11). [g]CH bond lengths fixed at LSR r_e values (see Table 11), and δ_4 at B2PLYP/VTZ r_e value.

Table 14. r_e and r_e^{SE} Equilibrium Geometries for 2-Fluoropyridine[a]

	r_e^{SE}[b]				r_e	
	B3LYP/SNSD[e]	B3LYP/SNSD[f]	B2PLYP/VTZ[f]	B2PLYP/VTZ[g]	B2PLYP/VTZ	CCSD(T)[h]
r(N−C2)	1.3138(10)[c]	1.3140(14)[c]	1.3085(8)[c]	1.3043(7)[c]	1.3090	1.3063
r(N−C6)	1.3402(5)	1.3401(3)	1.3422(2)	1.3421(1)	1.3409	1.3410
r(C2−C3)	1.3838(14)	1.3834(18)	1.3888(11)	1.3929(9)	1.9301	1.3898
r(C3−C4)	1.3837(3)	1.3838(4)	1.3829(2)	1.3827(1)	1.3854	1.3836
r(C4−C5)	1.3949(4)	1.3949(5)	1.3945(3)	1.3933(3)	1.3936	1.3933
r(C5−C6)	1.3836(4)	1.3836(5)	1.3842(3)	1.3851(2)	1.3861	1.3844
r(C2−F)	1.3357(2)	1.3356(2)	1.3343(1)	1.3346(1)	1.3394	1.3344
r(C3−H)	1.0781[d]	1.0787[d]	1.0787[d]	1.0785[d]	1.0784	1.0787
r(C4−H)	1.0801[d]	1.0807[d]	1.0807[d]	1.0803[d]	1.0803	1.0807
r(C5−H)	1.0788[d]	1.0794[d]	1.0794[d]	1.0792[d]	1.0791	1.0794
r(C6−H)	1.0809[d]	1.0815[d]	1.0815[d]	1.0815[d]	1.0816	1.0815
a(C6−N−C2)	116.24(6)	116.24(7)	116.30(4)	116.39(4)	116.60	116.42
a(N−C2−C3)	126.17(7)	126.17(8)	126.16(5)	126.16(4)	125.90	126.17
a(C2−C3−C4)	116.73(3)	116.73(3)	116.69(2)	116.60(1)	116.70	116.59
a(C3−C4−C5)	119.06(1)	119.05(1)	119.07(1)	119.10(1)	119.15	119.09
a(C4−C5−C6)	118.37(0)	118.37(1)	118.36(1)	118.36(1)	118.30	118.39
a(C5−C6−N)	123.44(2)	123.43(1)	123.42(1)	123.39(1)	123.34	123.35
a(C3−C2−F)	118.43(12)	118.45(16)	118.06(10)	117.70(8)	118.03	117.87
a(C2−C3−H)	120.38[d]	120.35(2)	120.37(1)	120.57(1)	120.52	120.42
a(C3−C4−H)	120.33(1)	120.37(1)	120.36(1)	120.17(1)	120.14	120.34
a(C6−C5−H)	120.26[d]	120.31(1)	120.31(1)	120.31(1)	120.38	120.30
a(C5−C6−H)	120.86[d]	120.87(1)	120.88(1)	120.82(1)	120.85	120.91
δ_2	1.52(10)	1.54(16)	1.14(9)	0.78(8)	0.98	0.95
δ_3	1.26(2)	1.28[d]	1.28[d]	1.13[d]	1.13	1.28
δ_4	0.14(1)	0.11[d]	0.11[d]	0.29[d]	0.29	0.11
δ_5	0.56(1)	0.51[d]	0.51[d]	0.51[d]	0.51	0.51
δ_6	2.58(1)	2.59[d]	2.59[d]	2.52[d]	2.52	2.59
rms resid [MHz]	0.0004	0.0002	0.0001	0.0001		
mean Δ_e [uÅ^2]	−0.00133	−0.00133	−0.00085	−0.00085		

[a]Distances in angstroms, angles in degrees. Atom numbering and sign convention on δ angles are reported in Figure 4. [b]The fits have been performed on the SE I_e^A and I_e^C moments of inertia. [c]Indicates the inclusion of ΔB_{el}^{β}. The digits within parentheses are the uncertainties on the geometrical parameters. [d]Denotes that the parameters are kept fixed. $\Delta_e = I^C - I^B - I^A$ is the inertial defect. [e]r_e^{SE} structure indicates as fit1 in Table 9 of ref 38 (see also Table 11). [f]CH bond lengths and δs fixed at CCSD(T)/CBS+CV r_e values. [g]CH bond lengths fixed at LSR r_e values (see Table 11), and δs at B2PLYP/VTZ r_e values. [h]CCSD(T)/CSB+CV from ref 38.

DOI: 10.1021/acs.jctc.5b00622
J. Chem. Theory Comput. 2015, 11, 4689−4707

Table 15. r_e and r_e^{SE} Equilibrium Geometries for 3-Fluoropyridine[a]

	r_e^{SE}[b]				r_e	
	B3LYP/SNSD[e]	B3LYP/SNSD[f]	B2PLYP/VTZ[f]	B2PLYP/VTZ[g]	B2PLYP/VTZ	CCSD(T)[h]
r(N−C2)	1.3346(9)[c]	1.3339(9)[c]	1.3356(4)[c]	1.3350(7)[c]	1.3338	1.3323
r(C2−C3)	1.3931(5)	1.3868(4)	1.3857(1)	1.3839(3)	1.3868	1.3859
r(C3−C4)	1.3714(6)	1.3793(3)	1.3801(1)	1.3819(3)	1.3817	1.3786
r(C4−C5)	1.3895(8)	1.3897(5)	1.3883(2)	1.3875(4)	1.3895	1.3894
r(C5−C6)	1.3940(5)	1.3906(3)	1.3913(1)	1.3910(2)	1.3908	1.3888
r(C6−N)	1.3314(9)	1.3353(3)	1.3345(1)	1.3349(3)	1.3360	1.3337
r(C2−H)	1.0811[d]	1.0817[d]	1.0817[d]	1.0818[d]	1.0819	1.0817
r(C3−F)	1.3406[d]	1.3393[d]	1.3393[d]	1.3402[d]	1.3440	1.3393
r(C4−H)	1.0791[d]	1.0797[d]	1.0797[d]	1.0796[d]	1.0796	1.0797
r(C5−H)	1.0794[d]	1.0800[d]	1.0800[d]	1.0799[d]	1.0799	1.0800
r(C6−H)	1.0808[d]	1.0814[d]	1.0814[d]	1.0816[d]	1.0817	1.0814
a(C6−N−C2)	117.68(6)	117.59(3)	117.63(1)	117.63(2)	117.78	117.73
a(N−C2−C3)	121.75(6)	121.97(3)	121.88(1)	121.92(3)	121.97	122.09
a(C2−C3−C4)	121.07(3)	121.01(1)	121.04(1)	121.05(1)	120.82	120.83
a(C3−C4−C5)	117.04(2)	116.88(1)	116.91(1)	116.87(1)	117.03	116.84
a(C4−C5−C6)	118.88(2)	118.95(1)	118.95(1)	118.96(1)	118.95	119.18
a(C5−C6−N)	123.58(5)	123.60(1)	123.58(1)	123.57(1)	123.44	123.33
a(C3−C2−H)	119.90[d]	120.02(2)	120.06(1)	120.05(1)	120.02	119.95
a(C4−C3−F)	120.59[d]	119.94(1)	119.92(1)	119.75(1)	119.87	119.14
a(C3−C4−H)	121.88(1)	120.55(16)	120.32(7)	120.48(13)	120.42	121.96
a(C6−C5−H)	120.24[d]	120.39(1)	120.39(1)	120.36(1)	120.36	120.28
a(C5−C6−H)	120.48[d]	120.39(1)	120.40(1)	120.36(1)	120.42	120.53
δ_2	0.78(3)	1.00[d]	1.00[d]	1.01[d]	1.01	1.00
δ_3	1.12(2)	0.44[d]	0.44[d]	0.28[d]	0.28	0.44
δ_4	−0.40(1)	1.01(16)	1.22(7)	1.08(13)	1.06	−0.38
δ_5	0.32(1)	0.13[d]	0.13[d]	0.16[d]	0.16	0.13
δ_6	2.27(2)	2.19[d]	2.19[d]	2.14[d]	2.14	2.19
rms resid [MHz]	0.0004	0.0003	0.0001	0.0002		
mean Δ_e [uÅ^2]	−0.00180	−0.00180	−0.00149	−0.00149		

[a]Distances in angstoms, angles in degrees. Atom numbering and sign convention on δ angles are reported in Figure 4. [b]The fits have been performed on the SE I_e^A and I_e^B moments of inertia. [c]Indicates the inclusion of ΔB_{el}^{β}. The digits within parentheses are the uncertainties on the geometrical parameters. [d]Denotes that the parameters are kept fixed. $\Delta_e = I^C - I^B - I^A$ is the inertial defect. [e]r_e^{SE} structure indicates as fit1 in Table 10 of ref 38 (see also Table 11). [f]CH and CF bond lengths and δs fixed at CCSD(T)/CBS+CV r_e values. [g]CH and CF bond lengths fixed at LSR r_e values (see Table 11), and δs at B2PLYP/VTZ r_e values. [h]CCSD(T)/CSB+CV from ref 38.

satisfactory, showing the stability of the fit with respect to the choice of these parameters. Then, fits have been performed fixing the CH bond lengths, plus the CF one for 3-fluoropiridine, at the corresponding LRA r_e values, while δ parameters at the corresponding B2PLYP/VTZ r_e values. The agreement between all the SE equilibrium structures is remarkable.

Finally, the SE equilibrium structure for uracil, obtained within the LRA approach, is shown in Table 8. Also in this case, the LRA SE equilibrium structure is in very good agreement with those obtained fixing the indeterminable parameters at the CCSD(T) r_e values.

In order to face the need of a larger number of accurate equilibrium structures, useful both as templating molecules in the TMA and as items in statistical analysis to improve LRA parameters, we have decided at this point to include uracil, 2-thiouracil, pyrimidine, 2-fluoropiridine and 3-fluoropiridine r_e^{SE} equilibrium structures in the B2se and B3se sets even if only their partial SE structures are available. This choice is fully justified by the excellent accuracy reached with TMA and LRA, and it allows us to add to our database, consisting now of 56 molecules, the structures of some important biological building blocks.

5. CONCLUSIONS

The large set of accurate equilibrium geometries proposed in our previous work[38] has been extended by new SE equilibrium geometries for phenyl radical, which is of interest for its radical nature, pyruvic and peroxyformic acids, and phenylacetylene, correcting the experimentally observed rotational constants by B3LYP/SNSD and B2PLYP/VTZ Δ_{vib}^{β} values, as well as, Δ_{el}^{β} ones. In the light of the discrepancies found between B3LYP/SNSD, B2PLYP/VTZ r_e^{SE} and high-level CCSD(T) r_e results for some parameters of glyoxylic and pyruvic acids, the use of B2PLYP/VTZ cubic force fields in the calculation of the Δ_{vib}^{β} employed in SE determinations has been systematically studied for all molecules belonging to the CCse, as well as B3se, sets of molecules. This approach delivers results with an accuracy comparable to the CCSD(T) ones, and leads to equilibrium geometries not significantly different from the B3LYP/SNSD SE determinations for the majority of systems, confirming the feasibility of using the low computational cost B3LYP/SNSD level to achieve accurate results. For the few problematic cases found, the B2PLYP/VTZ SE equilibrium structures have shown a better agreement with high-level theoretical results. The new B2PLYP/VTZ SE equilibrium structures have been collected in the B2se set.

DOI: 10.1021/acs.jctc.5b00622
J. Chem. Theory Comput. 2015, 11, 4689−4707

Concerning molecular systems for which the lack of experimental data does not allow the determination of a full semiexperimental equilibrium structure, different approaches to fix some geometrical parameters in the fitting procedure have been discussed and validated. One possible choice is to fix the non determinable parameters at high computational level r_e values. The CCSD(T) r_e parameters, in conjunction with at least quadruple-ζ quality basis sets, have been found to have a sufficient accuracy to be employed. However, to avoid the highly expensive CC computations, the templating molecule approach proposed in our previous work has been further tested and generalized. We have proposed here also the new linear regression approach, where the corrections to the r_e parameters are calculated by linear functions of the same parameters. We could also consider the use of predicates[92] as a third approach to treat more problematic systems, but the need of defining sufficiently accurate initial guess structures, and error bars for the fitting procedure, would have required further analysis that we have preferred to postpone to a successive work. As a general consideration, in all the considered approaches, we have tried to maximize the independence of the chosen geometrical parameters. In this way, the use of δ parameters to fix the angles for the out-of-ring atoms in planar cyclic molecules is preferred. In the light of high accuracy reached by partial r_e^{SE} structures analyzed in this work, we decided to include them in our database of equilibrium SE structures, that can be consulted on our web site dreams.sns.it.[43] As a final remark, we point out how in this work we were able to determine accurate equilibrium structures of semirigid molecules, also in the presence of limited experimental data. This is a satisfactory result that paves the route for extension of this approach to larger system possibly containing heavier atoms. Additional refinements are instead needed for flexible molecules, that represent a wide part of interesting chemical species. Work is in progress in our laboratory in this and other related directions.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jctc.5b00622.

r_e^{SE} and r_e geometries for the systems not explicitly discussed in the text, together with $(B_0^\beta)^{EXP}$ and B2PLYP/VTZ ΔB_{vib}^β corrections for all the molecules studied in the paper. The B3LYP/SNSD ΔB_{vib}^β and B3LYP/AVTZ ΔB_{el}^β are also reported for the new molecules not previously presented in ref 38 (PDF)

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: vincenzo.barone@sns.it.

Notes

The authors declare no competing financial interest.

■ REFERENCES

(1) Domenicano, A., Hargittai, I., Eds. *Accurate molecular structures. Their determination and importance*; Oxford University Press: New York, 1992.

(2) Demaison, J., Boggs, J., Császár, A., Eds. *Equilibrium molecular structures: from spectroscopy to quantum chemistry*; CRC Press: Boca Raton, FL, 2011.

(3) Barone, V., Ed. *Computational strategies for spectroscopy: from small molecules to nano systems*; Wiley & Sons, Inc.: Hoboken, NJ, 2011.

(4) Grunenberg, J., Ed. *Computational spectroscopy: methods, experiments and applications*; Wiley-VCH Verlag GmbH & Co. KGaA: Weinheim, Germany, 2010.

(5) Quack, M., Merkt, F., Eds. *Handbook of high-resolution spectroscopy*; John Wiley & Sons, Inc.: Weinheim, Germany, 2011; p 2182.

(6) Puzzarini, C.; Biczysko, M. In *Structure elucidation in organic chemistry*; Cid, M.-M., Ed.; Wiley-VCH Verlag GmbH & Co. KGaA: Weinheim, Germany, 2015; pp 27−64.

(7) Bak, K. L.; Gauss, J.; Jørgensen, P.; Olsen, J.; Helgaker, T.; Stanton, J. F. The accurate determination of molecular equilibrium structures. *J. Chem. Phys.* **2001**, *114*, 6548.

(8) Demaison, J. Experimental, semi-experimental and *ab initio* equilibrium structures. *Mol. Phys.* **2007**, *105*, 3109.

(9) Puzzarini, C.; Stanton, J. F.; Gauss, J. Quantum-chemical calculation of spectroscopic parameters for rotational spectroscopy. *Int. Rev. Phys. Chem.* **2010**, *29*, 273.

(10) Pérez, C.; Muckle, M. T.; Zaleski, D. P.; Seifert, N. A.; Temelso, B.; Shields, G. C.; Kisiel, Z.; Pate, B. H. Structures of cage, prism, and book isomers of water hexamer from broadband rotational spectroscopy. *Science* **2012**, *336*, 897−901.

(11) Caminati, W. Nucleic acid bases in the gas phase. *Angew. Chem., Int. Ed.* **2009**, *48*, 9030−9033.

(12) Pietraperzia, G.; Pasquini, M.; Schiccheri, N.; Piani, G.; Becucci, M.; Castellucci, E.; Biczysko, M.; Bloino, J.; Barone, V. The gas phase anisole dimer: a combined high-resolution spectroscopy and computational study of a stacked molecular system. *J. Phys. Chem. A* **2009**, *113*, 14343−14351.

(13) Blanco, S.; Lesarri, A.; López, J. C.; Alonso, J. L. The gas-phase structure of alanine. *J. Am. Chem. Soc.* **2004**, *126*, 11675−11683.

(14) Lovas, F. J.; McMahon, R. J.; Grabow, J.-U.; Schnell, M.; Mack, J.; Scott, L. T.; Kuczkowski, R. L. Interstellar chemistry: a strategy for detecting polycyclic aromatic hydrocarbons in space. *J. Am. Chem. Soc.* **2005**, *127*, 4345−4349.

(15) Peña, I.; Daly, A. M.; Cabezas, C.; Mata, S.; Bermúdez, C.; Niño, A.; López, J. C.; Grabow, J.-U.; Alonso, J. L. Disentangling the puzzle of hydrogen bonding in vitamin C. *J. Phys. Chem. Lett.* **2013**, *4*, 65−69.

(16) Puzzarini, C.; Biczysko, M.; Barone, V.; Largo, L.; Peña, I.; Cabezas, C.; Alonso, J. L. Accurate characterization of the peptide linkage in the gas phase: a joint quantum-chemical and rotational spectroscopy study of the glycine dipeptide analogue. *J. Phys. Chem. Lett.* **2014**, *5*, 534−540.

(17) Grimme, S.; Steinmetz, M. Effects of london dispersion correction in density functional theory on the structures of organic molecules in the gas phase. *Phys. Chem. Chem. Phys.* **2013**, *15*, 16031−16042.

(18) Jurecka, P.; Sponer, J.; Cerny, J.; Hobza, P. Benchmark database of accurate (MP2 and CCSD(T) complete basis set limit) interaction energies of small model complexes, DNA base pairs, and amino acid pairs. *Phys. Chem. Chem. Phys.* **2006**, *8*, 1985−1993.

(19) Zhao, Y.; Truhlar, D. G. Density functionals for noncovalent interaction energies of biological importance. *J. Chem. Theory Comput.* **2007**, *3*, 289−300.

(20) Barone, V.; Biczysko, M.; Pavone, M. The role of dispersion correction to DFT for modelling weakly bound molecular complexes in the ground and excited electronic states. *Chem. Phys.* **2008**, *346*, 247−256.

(21) Barone, V.; Biczysko, M.; Bloino, J.; Puzzarini, C. Accurate molecular structures and infrared spectra of trans-2,3-dideuterooxirane, methyloxirane, and trans-2,3-dimethyloxirane. *J. Chem. Phys.* **2014**, *141*, 034107.

(22) Senn, H. M.; Thiel, W. QM/MM methods for biomolecular systems. *Angew. Chem., Int. Ed.* **2009**, *48*, 1198−1229.

(23) Brooks, B. R.; Brooks, C. L.; Mackerell, A. D.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; Caflisch, A.; Caves, L.; Cui, Q.; Dinner, A. R.; Feig, M.; Fischer, S.;

DOI: 10.1021/acs.jctc.5b00622
J. Chem. Theory Comput. 2015, 11, 4689−4707

Gao, J.; Hodoscek, M.; Im, W.; Kuczera, K.; Lazaridis, T.; Ma, J.; Ovchinnikov, V.; Paci, E.; Pastor, R. W.; Post, C. B.; Pu, J. Z.; Schaefer, M.; Tidor, B.; Venable, R. M.; Woodcock, H. L.; Wu, X.; Yang, W.; York, D. M.; Karplus, M. CHARMM: The biomolecular simulation program. *J. Comput. Chem.* **2009**, *30*, 1545−1614.
(24) Pronk, S.; Páll, S.; Schulz, R.; Larsson, P.; Bjelkmar, P.; Apostolov, R.; Shirts, M. R.; Smith, J. C.; Kasson, P. M.; van der Spoel, D.; Hess, B.; Lindahl, E. GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics* **2013**, *29*, 845−854.
(25) Grubisic, S.; Brancato, G.; Pedone, A.; Barone, V. Extension of the AMBER force field to cyclic α,α dialkylated peptides. *Phys. Chem. Chem. Phys.* **2012**, *14*, 15308−15320.
(26) Maple, J. R.; Dinur, U.; Hagler, A. T. Derivation of force fields for molecular mechanics and dynamics from ab initio energy surfaces. *Proc. Natl. Acad. Sci. U. S. A.* **1988**, *85*, 5350−5354.
(27) Dasgupta, S.; Yamasaki, T.; Goddard, W. A. The Hessian biased singular value decomposition method for optimization and analysis of force fields. *J. Chem. Phys.* **1996**, *104*, 2898−2920.
(28) Biczysko, M.; Bloino, J.; Brancato, G.; Cacelli, I.; Cappelli, C.; Ferretti, A.; Lami, A.; Monti, S.; Pedone, A.; Prampolini, G.; Puzzarini, C.; Santoro, F.; Trani, F.; Villani, G. Integrated computational approaches for spectroscopic studies of molecular systems in the gas phase and in solution: pyrimidine as a test case. *Theor. Chem. Acc.* **2012**, *131*, 1201.
(29) Barone, V.; Cacelli, I.; de Mitri, N.; Licari, D.; Monti, S.; Prampolini, G. Joyce and Ulysses: integrated and user-friendly tools for the parameterization of intramolecular force fields from quantum mechanical data. *Phys. Chem. Chem. Phys.* **2013**, *15*, 3736−3751.
(30) Császár, A. In *Equilibrium molecular structures: from spectroscopy to quantum chemistry*; Demaison, J., Boggs, J., Császár, A., Eds.; CRC Press: Boca Raton, FL, 2011; pp 233−261.
(31) Kuchitsu, K. In *Accurate molecular structures. Their determination and importance*; Domenicano, A., Hargittai, I., Eds.; Oxford University Press: New York, 1992; p 14.
(32) Raghavachari, K.; Trucks, G. W.; Pople, J. A.; Head-Gordon, M. A fifth-order perturbation comparison of electron correlation theories. *Chem. Phys. Lett.* **1989**, *157*, 479.
(33) Heckert, M.; Kállay, M.; Gauss, J. Molecular equilibrium geometries based on coupled-cluster calculations including quadruple excitations. *Mol. Phys.* **2005**, *103*, 2109.
(34) Heckert, M.; Kállay, M.; Tew, D. P.; Klopper, W.; Gauss, J. Basis-set extrapolation techniques for the accurate calculation of molecular equilibrium geometries using coupled-cluster theory. *J. Chem. Phys.* **2006**, *125*, 044108.
(35) Barone, V.; Biczysko, M.; Bloino, J.; Puzzarini, C. The performance of composite schemes and hybrid CC/DFT model in predicting structure, thermodynamic and spectroscopic parameters: the challenge of the conformational equilibrium in glycine. *Phys. Chem. Chem. Phys.* **2013**, *15*, 10094−10111.
(36) Pulay, P.; Meyer, W.; Boggs, J. E. Cubic force constants and equilibrium geometry of methane from Hartree-Fock and correlated wavefunctions. *J. Chem. Phys.* **1978**, *68*, 5077.
(37) Pawłowski, F.; Jørgensen, P.; Olsen, J.; Hegelund, F.; Helgaker, T.; Gauss, J.; Bak, K.; Stanton, J. Molecular equilibrium structures from experimental rotational constants and calculated vibration-rotation interaction constants. *J. Chem. Phys.* **2002**, *116*, 6482.
(38) Piccardo, M.; Penocchio, E.; Puzzarini, C.; Biczysko, M.; Barone, V. Semi-experimental equilibrium structure determinations by employing B3LYP/SNSD anharmonic force fields: validation and application to semirigid organic molecules. *J. Phys. Chem. A* **2015**, *119*, 2058−2082.
(39) Barone, V.; Biczysko, M.; Bloino, J.; Penocchio, E.; Puzzarini, C.; Cimino, P. The CC/DFT Route towards Accurate Structures and Spectroscopic Features for Observed and Elusive Conformers of Flexible Molecules: Pyruvic Acid as Case Study. *J. Chem. Theory Comput.* **2015**, DOI: 10.1021/acs.jctc.5b00580.
(40) Grimme, S. Semiempirical hybrid density functional with perturbative second-order correlation. *J. Chem. Phys.* **2006**, *124*, 034108.
(41) Dunning, T. H., Jr. Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen. *J. Chem. Phys.* **1989**, *90*, 1007−1023.
(42) Kendall, R. A.; Dunning, T. H., Jr.; Harrison, R. J. Electron affinities of the first row atoms revisited. Systematic basis sets and wave functions. *J. Chem. Phys.* **1992**, *96*, 6796.
(43) CCse, B3se, and B2se sets are available in the download section. http://dreamslab.sns.it (accessed Feb 2015).
(44) Schäfer, L.; Alsenoy, C. V.; Scarsdale, J. Estimates for systematic empirical corrections of consistent 4−21G ab initio geometries and their correlations to total energy group increments. *J. Mol. Struct.: THEOCHEM* **1982**, *86*, 349−364.
(45) Chiu, N.; Ewbank, J.; Schäfer, L. Ab initio studies of structural features not easily amenable to experiment: Part 21. Structural trends in the C-H bond distances and C-C-H bond angles of cyclohexane. *J. Mol. Struct.: THEOCHEM* **1982**, *86*, 397−399.
(46) Van Alsenoy, C.; Scarsdale, J. N.; Schäfer, L. Ab initio studies of structural features not easily amenable to experiment. 12. The molecular structure of bicyclo(2.1.0)pentane and the usefulness of abinitio geometries in interpreting microwave rs-structure model uncertainties. *J. Chem. Phys.* **1981**, *74*, 6278−6284.
(47) Papoušek, D.; Aliev, M. *Molecular vibrational-rotational spectra*; Elsevier: New York, 1982.
(48) Gauss, J.; Ruud, K.; Helgaker, T. Perturbation-dependent atomic orbitals for the calculation of spin-rotation constants and rotational g tensors. *J. Chem. Phys.* **1996**, *105*, 2804−2812.
(49) Flygare, W. Magnetic interactions in molecules and an analysis of molecular electronic charge distribution from magnetic parameters. *Chem. Rev.* **1974**, *74*, 653−687.
(50) Born, M.; Oppenheimer, R. Zur quantentheorie der molekeln. *Ann. Phys. (Berlin, Ger.)* **1927**, *389*, 457−484.
(51) Sayvetz, A. The kinetic energy of polyatomic molecules. *J. Chem. Phys.* **1939**, *7*, 383−389.
(52) Eckart, C. Some studies concerning rotating axes and polyatomic molecules. *Phys. Rev.* **1935**, *47*, 552−558.
(53) Jensen, P.; Bunker, P. In *Computational molecular spectroscopy*; Jensen, P., Bunker, P., Eds.; Wiley: New York, 2000; pp 3−12.
(54) Nielsen, H. H. The vibration-rotation energies of molecules. *Rev. Mod. Phys.* **1951**, *23*, 90−136.
(55) Watson, J. K. G. Simplification of the molecular vibration-rotation hamiltonian. *Mol. Phys.* **1968**, *15*, 479−490.
(56) Aliev, M.; Watson, J. K. G. In *Molecular spectroscopy: modern research*; Rao, K. N., Ed.; Academic Press: Ohio, 1985; pp 1−67.
(57) Mills, I. M. *Molecular spectroscopy: modern research*; Rao, K. N., Mathews, C. W., Eds.; Elsevier, 1972; pp 115−140.
(58) Lee, C.; Yang, W.; Parr, R. G. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1988**, *37*, 785.
(59) Becke, A. D. Density-functional thermochemistry. III. The role of exact exchange. *J. Chem. Phys.* **1993**, *98*, 5648−5652.
(60) Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J. Ab initio calculation of vibrational absorption and circular dichroism spectra using density functional force fields. *J. Phys. Chem.* **1994**, *98*, 11623.
(61) Carnimeo, I.; Puzzarini, C.; Tasinato, N.; Stoppa, P.; Pietropolli-Charmet, A.; Biczysko, M.; Cappelli, C.; Barone, V. Anharmonic theoretical simulations of infrared spectra of halogenated organic compounds. *J. Chem. Phys.* **2013**, *139*, 074310.
(62) Barone, V.; Biczysko, M.; Bloino, J. Fully anharmonic IR and Raman spectra of medium-size molecular systems: accuracy and interpretation. *Phys. Chem. Chem. Phys.* **2014**, *16*, 1759−1787.
(63) The SNSD basis set is available in the download section. http://dreamslab.sns.it (accessed Oct 2014).
(64) Puzzarini, C.; Biczysko, M.; Barone, V. Accurate harmonic/anharmonic vibrational frequencies for open-shell systems: performances of the B3LYP/N07D model for semirigid free radicals

benchmarked by CCSD(T) computations. *J. Chem. Theory Comput.* **2010**, *6*, 828.
(65) Barone, V.; Bloino, J.; Biczysko, M. Validation of the DFT/N07D computational model on the magnetic, vibrational and electronic properties of vinyl radical. *Phys. Chem. Chem. Phys.* **2010**, *12*, 1092.
(66) Barone, V.; Biczysko, M.; Bloino, J.; Puzzarini, C. Characterization of the elusive conformers of glycine from state-of-the-art structural, thermodynamic, and spectroscopic computations: theory complements experiment. *J. Chem. Theory Comput.* **2013**, *9*, 1533.
(67) Puzzarini, C.; Biczysko, M.; Barone, V.; Peña, I.; Cabezas, C.; Alonso, J. Accurate molecular structure and spectroscopic properties of nucleobases: a combined computational-microwave investigation of 2-thiouracil as a case study. *Phys. Chem. Chem. Phys.* **2013**, *15*, 16965.
(68) Neese, F.; Schwabe, T.; Grimme, S. Analytic derivatives for perturbatively corrected "double hybrid" density functionals: Theory, implementation, and applications. *J. Chem. Phys.* **2007**, *126*, 124115.
(69) Biczysko, M.; Panek, P.; Scalmani, G.; Bloino, J.; Barone, V. Harmonic and Anharmonic Vibrational Frequency Calculations with the Double-Hybrid B2PLYP Method: Analytic Second Derivatives and Benchmark Studies. *J. Chem. Theory Comput.* **2010**, *6*, 2115−2125.
(70) Woon, D. E.; Dunning, T. H., Jr. Gaussian basis sets for use in correlated molecular calculations. III. The atoms aluminum through argon. *J. Chem. Phys.* **1993**, *98*, 1358−1371.
(71) Woon, D. E.; Dunning, T. H., Jr. Gaussian basis sets for use in correlated molecular calculations. V. Core-valence basis sets for boron through neon. *J. Chem. Phys.* **1995**, *103*, 4572−4585.
(72) Peterson, K. A.; Dunning, T. H., Jr. Accurate correlation consistent basis sets for molecular core-valence correlation effects: The second row atoms Al-Ar, and the first row atoms B-Ne revisited. *J. Chem. Phys.* **2002**, *117*, 10548−10560.
(73) Schneider, W.; Thiel, W. Anharmonic force fields from analytic second derivatives: Method and application to methyl bromide. *Chem. Phys. Lett.* **1989**, *157*, 367.
(74) Stanton, J. F.; Gauss, J. Analytic second derivatives in high-order many-body perturbation and coupled-cluster theories: Computational considerations and applications. *Int. Rev. Phys. Chem.* **2000**, *19*, 61−95.
(75) Thiel, W.; Scuseria, G.; Schaefer, H. F. S.; Allen, W. D. The anharmonic force fields of HOF and F_2O. *J. Chem. Phys.* **1988**, *89*, 4965−4975.
(76) Barone, V. Anharmonic vibrational properties by a fully automated second-order perturbative approach. *J. Chem. Phys.* **2005**, *122*, 014108.
(77) Barone, V. Characterization of the potential energy surface of the HO_2 molecular system by a density functional approach. *J. Chem. Phys.* **1994**, *101*, 10666−10676.
(78) Stanton, J. F.; Lopreore, C. L.; Gauss, J. The equilibrium structure and fundamental vibrational frequencies of dioxirane. *J. Chem. Phys.* **1998**, *108*, 7190−7196.
(79) Martinez, O.; Crabtree, K. N.; Gottlieb, C. A.; Stanton, J. F.; McCarthy, M. C. An accurate molecular structure of phenyl, the simplest aryl radical. *Angew. Chem., Int. Ed.* **2015**, *54*, 1808−1811.
(80) Demaison, J.; Császár, A. G.; Margulès, L. D.; Rudolph, H. D. Equilibrium structures of heterocyclic molecules with large principal axis rotations upon isotopic substitution. *J. Phys. Chem. A* **2011**, *115*, 14078−14091.
(81) Kochikov, I. V.; Tarasov, Y. I.; Spiridonov, V. P.; Kuramshina, G. M.; Rankin, D. W. H.; Saakjan, A. S.; Yagola, A. G. The equilibrium structure of thiophene by the combined use of electron diffraction, vibrational spectroscopy and microwave spectroscopy guided by theoretical calculations. *J. Mol. Struct.* **2001**, *567*, 29−40.
(82) Demaison, J.; Herman, M.; Liévin, J. Anharmonic force field of *cis*- and *trans*-formic acid from high-level ab initio calculations, and analysis of resonance polyads. *J. Chem. Phys.* **2007**, *126*, 164305.
(83) Hazra, M. K.; Sinha, A. Spectra and integrated band intensities of the low order OH stretching overtones in peroxyformic acid: an atmospheric molecule with prototypical intramolecular hydrogen bonding. *J. Phys. Chem. A* **2011**, *115*, 5294−5306.
(84) Vogt, N.; Demaison, J.; Vogt, J.; Rudolph, H. D. Why it is sometimes difficult to determine the accurate position of a hydrogen atom by the semiexperimental method: Structure of molecules containing the OH or the CH_3 group. *J. Comput. Chem.* **2014**, *35*, 2333−2342.
(85) Lide, D. R.; Christensen, D. Molecular Structure of Propylene. *J. Chem. Phys.* **1961**, *35*, 1374−1378.
(86) Demaison, J.; Rudolph, H. Ab initio anharmonic force field and equilibrium structure of propene. *J. Mol. Spectrosc.* **2008**, *248*, 66−76.
(87) Rudolph, H. D.; Demaison, J.; Császár, A. G. Accurate determination of the deformation of the benzene ring upon substitution: equilibrium structures of benzonitrile and phenylacetylene. *J. Phys. Chem. A* **2013**, *117*, 12969−12982.
(88) Bakri, B.; Demaison, J.; Margulès, L.; Møllendal, H. The submillimeter-wave spectrum and quantum chemical calculations of glyoxylic acid. *J. Mol. Spectrosc.* **2001**, *208*, 92−100.
(89) Barone, V.; Biczysko, M.; Bloino, J.; Egidi, F.; Puzzarini, C. Accurate structure, thermodynamics, and spectroscopy of medium-sized radicals by hybrid coupled cluster/density functional theory approaches: The case of phenyl radical. *J. Chem. Phys.* **2013**, *138*, 234303.
(90) Puzzarini, C.; Barone, V. Extending the molecular size in accurate quantum-chemical calculations: the equilibrium structure and spectroscopic properties of uracil. *Phys. Chem. Chem. Phys.* **2011**, *13*, 7189.
(91) Császár, A. G.; Demaison, J.; Rudolph, H. D. Equilibrium structures of 3-, 4-, 5-, 6-, and 7-membered unsaturated N-containing heterocycles. *J. Phys. Chem. A* **2015**, *119*, 1731−1746.
(92) Demaison, J. In *Equilibrium molecular structures: from spectroscopy to quantum chemistry*; Demaison, J., Boggs, J., Császár, A., Eds.; CRC Press: Boca Raton, FL, 2011; pp 29−52.

DOI: 10.1021/acs.jctc.5b00622
J. Chem. Theory Comput. 2015, 11, 4689−4707