Zero-Cost Estimation of Zero-Point Energies

2 AUTHORS:

Attila Csaszar
Eötvös Loránd University
197 PUBLICATIONS **5,725** CITATIONS

SEE PROFILE

Tibor Furtenbacher
Eötvös Loránd University
31 PUBLICATIONS **703** CITATIONS

SEE PROFILE

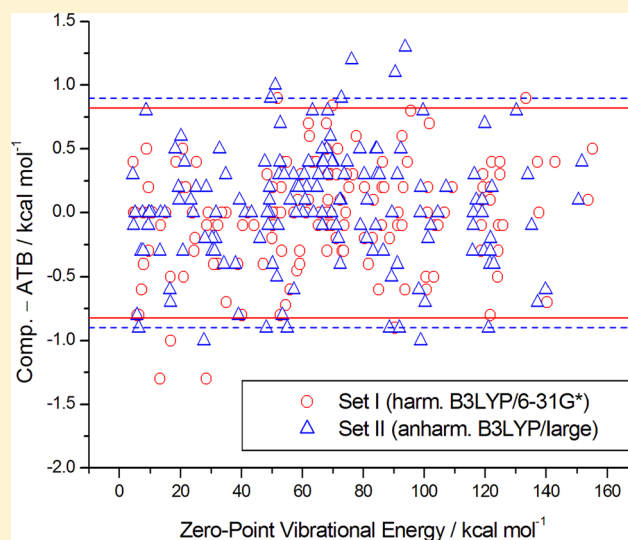# Zero-Cost Estimation of Zero-Point Energies

Published as part of The Journal of Physical Chemistry A virtual special issue "Spectroscopy and Dynamics of Medium-Sized Molecules and Clusters: Theory, Experiment, and Applications".

Attila G. Császár* and Tibor Furtenbacher

MTA-ELTE Complex Chemical Systems Research Group, P.O. Box 32, H-1518 Budapest 112, Hungary

Laboratory of Molecular Structure and Dynamics, Institute of Chemistry, Eötvös University, Pázmány Péter sétány 1/A, H-1117 Budapest, Hungary

ABSTRACT: An additive, linear, atom-type-based (ATB) scheme is developed allowing no-cost estimation of zero-point vibrational energies (ZPVE) of neutral, closed-shell molecules in their ground electronic states. The atom types employed correspond to those defined within the MM2 molecular mechanics force field approach. The reference training set of 156 molecules cover chained and branched alkanes, alkenes, cycloalkanes and cycloalkenes, alkynes, alcohols, aldehydes, carboxylic acids, amines, amides, ethers, esters, ketones, benzene derivatives, heterocycles, nucleobases, all the natural amino acids, some dipeptides and sugars, as well as further simple molecules and ones containing several structural units, including several vitamins. A weighted linear least-squares fit of atom-type-based ZPVE increments results in recommended values for the following atoms, with the number of atom types defined in parentheses: H(8), D(1), B(1), C(6), N(7), O(3), F(1), Si(1), P(2), S(3), and Cl(1). The average accuracy of the ATB ZPVEs is considerably better than 1 kcal mol^{-1}, that is, better than chemical accuracy. The proposed ATB scheme could be extended to many more atoms and atom types, following a careful validation procedure; deviation from the MM2 atom types seems to be necessary, especially for third-row elements.

1. INTRODUCTION

Although zero-point vibrational energies (ZPVE) of molecules can be considered as unmeasurable quantities, they do play an important role in several areas of molecular sciences, eminently in thermochemistry. Thus, during the last 80 years, several empirical schemes have been developed allowing the estimation of ZPVEs of gas-phase molecules without the need to perform extensive and costly quantum-chemical computations.[1−17] The empirical schemes important for the present study are based on the idea of the additivity of atomic, bond, or group increments. The basic idea behind most of the empirical expressions is that, following the Born−Oppenheimer approximation and the harmonic oscillator model of molecular vibrations, the ZPVE of a molecule is related to a simple sum of wavenumbers, and thus errors introduced during the estimation of the individual wavenumbers might get avareged out in a ZPVE.

Toward the end of the 1930s, Pitzer[1] investigated how to compute thermodynamic functions for gaseous hydrocarbons based on empirical values for the vibrational frequencies of stretching and bending normal modes. In a follow-up study, Cottrel[2] realized that for *n*-paraffin chains there is an incremental increase in the ZPVEs with successive additions of the −CH$_2$− unit. Another decade later, Pitzer and Catalano[3] used a constant value of 17.7 kcal mol^{-1} per CH$_2$ group to estimate ZPVE values of paraffins via a linear expression. The empirical equation of Pitzer and Catalano resulted in accuracies from 0.1% to 1% for C$_n$H$_{2n+2}$ molecules with $n \leq 5$. Accepting that chemical accuracy is defined as 1 kcal mol^{-1}, a 0.1% accuracy would be sufficient to estimate a ZPVE of chemical accuracy for a ZPVE of 1000 kcal mol^{-1}, corresponding to a hydrocarbon with $n \approx 50$ or a dekapeptide, rather large molecular systems.

In the 1950s and 1960s, Bernstein[4−7] performed multivariate linear least-squares analyses and obtained formulas for the rapid estimation of ZPVEs of halomethanes, haloethylenes, and benzene homologues. Bernstein considered not only chemical bonds as ZPVE building elements but also interactions between atoms that are two or three bonds away from each other. The accuracy of the empirical methods Bernstein proposed is generally between 0.1% and a few percent.

Somewhat later, Fujimoto and Shingu[8] employed three- and five-parameter linear expressions, based on C−H and C−C bonds and an end-of-chain correction, to estimate ZPVEs of normal and branched hydrocarbons. The simple expressions could be parametrized very well for this relatively simple class of molecules.

Even later, Flanigan et al.[9] and Schulman and Disch[11] showed how a simple empirical scheme, based on the number of carbon and hydrogen atoms a molecule contains, can be used to obtain ZPVE estimates of C_nH_m hydrocarbons. The simple approximate relationship proposed by Flanigan et al.[9] is $ZPVE(n,m)/kcal\ mol^{-1} = 2n + 7m$. On the basis of 55 experimental and theoretical reference ZPVE values of semirigid molecules ranging from methane, CH_4, to naphthalene, $C_{10}H_8$, Schulman and Disch[11] improved the predicting power of this linear relation. Their improved equation reads as $ZPVE(n,m)/kcal\ mol^{-1} = 3.88n + 7.12m − 6.19$, with a root-mean-square (rms) deviation of 1.22 kcal mol^{-1} for the complete set they used. The formula of Schulman and Disch was extended to molecules containing N, O, F, and Cl atoms by Ibrahim and Fataftah,[12] with atomic prefactors of 4.05 for N, 3.40 for O, 3.27 for F, and 2.22 for Cl. In a similar study, Grice and Politzer[13] presented a simple linear relation between the molecular stoichiometry and the ZPVE for a number of molecules.

In the 1980s, in a series of papers Oi and co-workers[15−17] attempted to provide a theoretical foundation explaining the additivity of ZPVE estimates based on a system of harmonic oscillators. Oi et al. investigated the contributions based on valence internal coordinate types characterizing the (harmonic) vibrations of molecules. They argued that the simple incremental scheme is not valid near the ends of a chain or near branching points. They attributed this to the fact that the related force constant and inverse mass matrix elements are rather different from those of the midchain region. In a somewhat related study, Fliszar et al.[18] pointed out that the values of the standard thermochemical quantity $H(T) − H(0) + ZPVE$, where H is the enthalpy and T is the thermodynamic temperature, follow additivity rules based on the number of atoms and branching points characterizing the given molecule.

Rahal et al.[14] linked the ZPVEs of molecules to bond contributions (BC) and determined a large number of single- and multiple-bond BC_i values. Their empirical relation can be written as $ZPVE(BC)/kcal\ mol^{-1} = \sum_i^P N_i BC_i − 2.09$, where P is the overall number of bond types and N_i is the number of a certain bond type in the molecule. They tested their empirical formula for various classes of molecules and atoms, including H, C, N, O, S, Cl, and F. Except for a few isolated cases, the empirical relationship worked with an accuracy better than 2 kcal mol^{-1} for the 48 molecules studied.

Ruzsinszky et al.[19] investigated how the partial charges computed within density functional theory (DFT) would correlate with ZPVE values. They investigated 117 closed-shell, neutral molecules containing H, C, N, O, and F atoms. Their scaled Hartree−Fock and DFT computations show that atomic partial charges can be employed to estimate ZPVEs with chemical accuracy. Although this method does not require computation of derivative information, it still requires quantum chemical computations to estimate the ZPVEs of molecules.

In a classic development of thermochemistry, Benson[20] proposed a group additivity method for estimating gas-phase enthalpies of formation, entropies, and heat capacities. Benson's table of group increments has been widely used, lately especially for larger molecular systems. One major problem with Benson's method is the excessive use of corrections, especially for flexible, cyclic molecules. Nevertheless, the considerable utility of Benson's simple tables cannot be questioned and has been supported by several studies, including those that use atom equivalents for relating ab initio energies to enthalpies of formation.[21,22]

While stoichiometry-based empirical relations can be quite useful, it is suggested here that even higher accuracy can be expected if the atoms are substituted by atom types commonly occurring in molecular mechanics (force field) approaches.[23] Furthermore, as of today, we can take advantage of a number of highly accurate ZPVE estimates of small molecules and of the simple determination of a large number of sufficiently accurate ZPVE values of medium-sized molecules estimated via first-principles approaches. This idea of atom-type-based (ATB) ZPVE increments is pursued in the present investigation.

In the fourth age of quantum chemistry[24] interest in the first-principles prediction of thermochemical properties, including temperature-dependent standard enthalpies of formation, $\Delta_f H_T^0$, of small and medium-sized gas-phase molecules, as well as those of radicals, increased dramatically.[25−28] For example, $\Delta_f H_T^0$ values are often employed in the calculation of temperature-dependent enthalpies of reactions helping detailed reaction kinetics studies. The best first-principles techniques, often following recommendations of the composite focal-point analysis (FPA) scheme,[29,30] are capable of producing (temperature-dependent) enthalpies of formation with uncertainties smaller than those characterizing traditional experimental determinations.[26,27,31] It must be noted, furthermore, that small experimental uncertainties of $\Delta_f H_0^0$ or $\Delta_f H_{298}^0$ values are sometimes unfounded. A case in mind is the $^2\Pi$ OH radical, where the experimental $\Delta_f H_0^0 = 39.12 \pm 0.21$ kJ mol^{-1} value, based on a spectroscopic approach, was widely accepted for a long time, but then it had to be lowered substantially, based on a positive ion cycle approach, to 36.94 ± 0.33 kJ mol^{-1}.[32] The revised value has been fully supported by high-level quantum-chemical computations.[27,32] A comparison of experimental and ab initio thermochemical quantities must be concluded with the statement that the approach producing the most accurate thermochemical quantities seems to be a mixed one, the joint utilization of all theoretical and experimental information, like in the Active Thermochemical Tables (ATcT) technique developed by Ruscic et al.[33] First-principles thermochemical quantities can also be improved by the active database technology, as proven by results from the Network of computed reaction Enthalpies to Atom-based Thermochemistry (NEAT)[34] approach of the present authors.

On the basis of electronic structure computations one can readily afford these days for small and medium-sized molecules, it seems that one of the largest uncertainties in the converged ab initio determination of enthalpies of formation is the ZPVE contribution.[35] More precisely, the ZPVE contribution is not only considerably larger than the uncertainty of the computed pure electronic enthalpy of formation but the error stemming from the use of simple harmonic ZPVE estimates is also of comparable magnitude (unless some kind of scaling is attempted and applied successfully). For small and medium-sized semirigid molecules accurate ZPVE estimates can be obtained if the potential energy surface (PES) is known accurately around the equilibrium position, perhaps in the form of an anharmonic force field[36] or some other expansion.[37] Once an accurate, even local, representation of the PES is available,

Table 1. Highly Accurate "Experimental" (expt) Zero-Point Vibrational Energies of Small Molecules, in cm^{-1}, and the Corresponding Harmonic and Anharmonic Estimates at Different Levels of Electronic Structure Theory[a]

molecule	expt value	source	harmonic values				anharmonic contribution		
			B3LYP/ 6-31G*	B3LYP/ large	MP2(FC)/ cc-pVTZ	CCSD(T)_FC/ cc-pVQZ	B3LYP/ 6-31G*	B3LYP/ large	MP2(FC)/ cc-pVTZ
N_2O	2371	ref 61	2462	2457	2362	2398			
CO_2	2526	ref 61	2544	2573	2535	2545			
H_2S	3288	ref 62	3330	3303	3395	3333	−45	−43	−39
D_2O	3390	ref 63	3380	3413	3450	3440	−39	−41	−38
HCN	3481	ref 61	3614	3575	3470	3502			
D_2O_2	4326	ref 38	4338	4391	4408	4389	−55	−54	−57
H_2O	4638	ref 63	4645	4688	4741	4728	−73	−76	−72
C_2H_2	5716	ref 64	5846	5916	5823	5801			
H_2O_2	5725	ref 38	5762	5835	5861	5834	−96	−97	−102
H_2CO	5786	ref 65	5888	5822	5914	5860	−81	−79	−77
H_2CCO	6832	ref 66	6967	6919	6947	6907	−79	−75	−76
NH_3	7437	ref 67	7579	7508	7641	7571	−123	−120	−134
C_2D_4	8374	ref 68	8533	8481	8565	8462	−84	−82	−81
CH_3F	8548	ref 69	8672	8581	8779	8674	−122	−125	−125
CH_4	9692	ref 70	9924	9782	9930	9842	−138	−137	−137
C_2H_4	11 004	ref 68	11 242	11 174	11 292	11 156	−149	−144	−142

[a]Large = 6-311+G(3df,2pd); FC = frozen core. No anharmonic contributions were computed for the linear molecules of the data set.

sophisticated variational or diffusion Monte Carlo (DMC) techniques can yield the ZPVE of the molecule with outstanding accuracy limited principally by the accuracy of the PES.[38] Nevertheless, for semirigid molecules larger than ~15−20 atoms the high-level computation of ZPVE estimates is still time-consuming, and it is highly desirable to have a procedure at hand that would provide, at zero cost, ZPVE estimates of chemical accuracy. Development of such a scheme is attempted in this study by utilizing the usual atom types of molecular mechanics (force field) approaches[23] and a linear additivity approximation of the ATB ZPVE increments.

2. THEORETICAL FOUNDATIONS

Weighted linear least-squares fits have been employed successfully in thermochemistry within at least the ATcT[33] and NEAT[34] approaches to relate a large number of measurable or computed quantities, like reaction enthalpies, to a (much) smaller set of basic thermochemical quantities, enthalpies of formation. A similar tool has been used in high-resolution molecular spectroscopy, within the Measured Active Rotational−Vibrational Energy Levels[39,40] approach, to determine a relatively small number of energy levels from a relatively large number of measured rovibrational transitions. The basis of these refinements is provided by the underlying thermochemical[33] and spectroscopic[41,42] networks. The same scheme is used here to determine atom-type ZPVE increments.

Our basic hypothesis is that the ZPVE value of a molecule can be written as the sum of atom-type ZPVE increments:

$$ZPVE = \sum_{i=1}^{n} c_i \xi_i \tag{1}$$

where n is the number of different atom types in the molecule, c_i is the number of the atoms with atom type i, and ξ_i is the value of the ith atom-type increment. Since the relation between ZPVE and ξ_i can be written in a similar way as in eq 1 for all the molecules, a matrix equation can be formulated, with N_m molecules and N_a different atom-types:

$$\mathbf{aX} = \mathbf{Y} \tag{2}$$

where the vector $\mathbf{Y}$, of dimension N_m, contains the N_m computed ZPVEs, and the vector $\mathbf{X}$, of dimension N_a, contains the ξ_i values as results. The ijth element of the design matrix $\mathbf{a}$, of dimension $N_m \times N_a$, is c_j if atom-type j exists in the ith molecule; otherwise, it is zero. The weighted least-squares solution of eq 2 is obtained by solving

$$\mathbf{AX} = \mathbf{B} \tag{3}$$

where $\mathbf{A} = \mathbf{a}^T\mathbf{ga}$, $\mathbf{B} = \mathbf{a}^T\mathbf{gY}$, and $g_i = 1/\delta_i^2$, where δ_i is the uncertainty attached to the ith ZPVE. The dimension of the extremely sparse matrix $\mathbf{A}$ is $N_a \times N_a$, and eq 3 is a simple system of linear equations that can be solved much more efficiently than the overdetermined system of linear equations given by eq 2.

3. COMPUTATIONAL DETAILS

Motivated by the success of the earliest systematic investigations[43,44] proving the utility of second-order vibrational perturbation theory (VPT2)[45,46] to computational molecular spectroscopy, several electronic structure packages[47,48] have by now the capability to compute, in an automatic fashion and even analytically,[49] not only harmonic (quadratic) force fields but also anharmonic (quartic) force fields in Cartesian and normal coordinate spaces. The computed force fields can then be utilized to compute harmonic and anharmonic ZPVE estimates. To avoid the nonzero-force dilemma,[50] all harmonic and anharmonic force fields of this study were computed at the respective optimized structures.

The D.01 version of the Gaussian09 code[48] was used for all DFT, second-order Møller−Plesset (MP2), and coupled cluster theory with single and double excitations and a perturbative estimate of triples (CCSD(T)) electronic structure computations of this study, including geometry optimizations as well as harmonic and anharmonic force-field computations. Only one widely employed DFT functional was utilized, abbreviated as B3LYP,[51,52] whose accuracy for the present purposes was established in ref 53 and elsewhere. Two atom-centered, fixed-

Table 2. Computed Zero-Point Vibrational Energy Estimates, in kcal mol^{-1}, Obtained at the Harmonic (harm) and Anharmonic (anharm) Levels of Approximation at Two Levels of Electronic Structure Theory, Compared to Two Different Empirical (emp) Atom-Type-Based Estimates Obtained as Part of This Study[a]

molecular type	molecule	harm B3LYP/6-31G*	emp	Δ	anharm B3LYP/large	expt	emp	Δ
alkane	methane, CH_4	28.4	29.7	−1.3	27.7	27.7	28.7	−1.0
	ethane, CH_3CH_3	47.2	47.6	−0.4	46.0		46.2	−0.2
	propane, $CH_3CH_2CH_3$	65.3	65.4	−0.1	63.7		63.7	0.0
	isobutane, $(CH_3)_3CH$	83.1	83.3	−0.2	81.1		81.2	−0.2
	n-butane, $CH_3(CH_2)_2CH_3$	83.3	83.3	0.1	81.3		81.2	0.1
	neopentane, $C(CH_3)_4$	100.6	101.1	−0.5	98.2		98.7	−0.6
	isopentane, $C_2H_5CH(CH_3)_2$	101.1	101.1	0.0	98.7		98.7	0.0
	n-pentane, $CH_3(CH_2)_3CH_3$	101.3	101.1	0.2	98.9		98.7	0.2
	neohexane	118.7	119.0	−0.3	115.9		116.2	−0.3
	2,3-dimethyl-butane	118.9	119.0	−0.1	116.1		116.2	−0.1
	3-methyl-pentane	119.1	119.0	0.1	116.2		116.2	0.0
	isohexane	119.1	119.0	0.1	116.2		116.2	0.0
	n-hexane, $CH_3(CH_2)_4CH_3$	119.3	119.0	0.3	116.5		116.2	0.3
	n-heptane, $CH_3(CH_2)_5CH_3$	137.2	136.8	0.4	134.0		133.7	0.3
	n-octane, $CH_3(CH_2)_6CH_3$	155.2	154.7	0.5	151.6		151.2	0.4
cycloalkane	cyclopropane, C_3H_6	51.3	51.3	0.0	50.2		50.6	−0.4
	cyclobutane, C_4H_8	69.9	69.1	0.8	68.3		67.4	0.8
alkene	ethylene, C_2H_4	32.2	32.5	−0.4	*31.5*	31.5	31.7	−0.2
	propylene, C_3H_6	50.3	50.4	−0.1	49.2		49.2	0.0
	1,3-butadiene, C_4H_6	53.0	53.1	0.0	52.6		52.2	0.4
	cis-butene, C_4H_8	68.2	68.3	−0.1	66.6		66.7	−0.1
	isoprene, C_5H_8	71.6	71.1	0.5	70.1		69.7	0.4
cycloalkene	cyclobutene, C_4H_6	54.6	55.3	−0.7	53.4		54.2	−0.8
	cyclopentadiene, C_5H_6	58.3	58.7	−0.4	57.2		57.8	−0.6
alkyne	acetylene, C_2H_2	16.7	17.2	−0.5	*16.7*	16.2	16.9	−0.7
	propyne, C_3H_4	35.0	35.0	0.0	34.8		34.5	0.3
	2-butyne, CH_3CCCH_3	53.1	52.9	0.2	52.7		52.0	0.7
	1-butyne, $HCCCH_2CH_3$	53.3	52.9	0.4	52.3		52.0	0.3
alcohol	methanol, CH_3OH	32.3	32.7	−0.4	31.6		31.6	0.0
	ethanol, C_2H_5OH	50.4	50.5	−0.1	49.2		49.1	0.1
	2-propanol	68.1	68.4	−0.3	66.5		66.7	−0.1
	propanol, $CH_3(CH_2)_2OH$	68.4	68.4	0.1	66.5		66.7	−0.1
	glycerole	74.3	74.3	0.0	72.6		72.5	0.1
	isobutyl alcohol	86.1	86.2	−0.1	84.1		84.2	−0.1
	butanol, $CH_3(CH_2)_3OH$	86.4	86.2	0.2	84.5		84.2	0.3
aldehyde	formaldehyde, H_2CO	16.8	17.9	−1.0	*16.5*	16.5	17.1	−0.6
	acetaldehyde, CH_3CHO	35.0	35.7	−0.7	34.2		34.6	−0.4
	propionaldehyde, $CH_3CH_3CH_2CHO$	53.3	53.6	−0.3	52.0		52.1	−0.1
	glyceraldehyde	59.2	59.5	−0.3	58.1		57.9	0.2
	butyraldehyde, $CH_3(CH_2)_2CHO$	71.2	71.4	−0.2	69.6		69.6	0.0
carboxylic acid	formic acid, $HCOOH$	21.0	21.5	−0.5	20.7		20.9	−0.3
	oxalic acid, $(COOH)_2$	30.8	31.2	−0.4	30.3		30.6	−0.3
	acetic acid, CH_3COOH	38.9	39.4	−0.4	38.0		38.4	−0.4
	propionic acid, C_2H_5COOH	57.2	57.2	0.0	56.0		55.9	0.1
	butyric acid, C_3H_7COOH	75.3	75.1	0.2	73.7		73.4	0.3
	pentanoic acid, C_4H_9COOH	93.2	93.0	0.2	91.2		90.9	0.3
amine	methylamine, CH_3NH_2	40.4	40.6	−0.1	39.4		39.3	0.1
	ethylamine, $C_2H_5NH_2$	58.5	58.4	0.1	57.1		56.8	0.3
	propylamine, $C_3H_7NH_2$	76.6	76.3	0.3	74.7		74.3	0.4
	butylamine, $C_4H_9NH_2$	94.5	94.1	0.4	92.3		91.8	0.5
nitro	nitroethane	49.6	49.6	0.0	48.5		48.5	0.0
amide	acrylamide, $CH_2CHCONH_2$	50.0	50.1	−0.2	48.8		48.9	−0.1
ketone	acetone, CH_3COCH_3	52.8	53.6	−0.8	51.6		52.1	−0.5
ether	diethyl ether	86.3	86.0	0.2	84.1		83.6	0.5
ester	ethyl acetate	74.5	74.2	0.3	72.8		71.9	0.9
	ethylpropionate	92.7	92.1	0.6	90.5		89.4	1.1
benzene	1-F-2,4-di-NO_2-benzene	61.2	61.2	0.0	60.2		60.2	0.0

Table 2. continued

molecular		harm			anharm			
type	molecule	B3LYP/6-31G*	emp	Δ	B3LYP/large	expt	emp	Δ
derivative	benzonitrile, C_6H_5CN	62.4	61.8	0.6	61.4		61.2	0.2
	benzene, C_6H_6	63.2	62.9	0.4	62.1		61.7	0.4
	phenol, C_6H_5OH	65.8	65.8	0.0	64.8		64.6	0.2
	benzaldehyde, C_6H_5CHO	69.2	68.9	0.3	68.0		67.5	0.5
	aniline, $C_6H_5NH_2$	73.7	73.7	−0.1	72.3		72.2	0.1
	toluene, $C_6H_5CH_3$	80.5	80.7	−0.2	79.1		79.2	−0.1
	acetophenone, $C_6H_5C(O)CH_3$	86.9	86.7	0.2	85.3		85.0	0.3
heterocycle	thiophene, C_4H_4S	42.0	42.0	0.0	41.3		41.3	0.0
	furan, C_4H_4O	44.1	44.1	−0.1	43.3		43.3	0.0
	pyrimidine	48.5	48.2	0.3	47.7		47.3	0.4
	pyrrole, C_4H_4NH	51.9	50.9	0.9	51.1		50.1	1.0
	pyridine, C_5H_5N	55.9	56.4	−0.6	55.0		55.9	−0.9
	purine	59.7	59.4	0.3	58.8		58.6	0.2
	pyrane, C_5H_6O	62.0	62.0	0.1	60.9		60.8	0.1
	2-methylpyrrole	69.3	69.3	0.0	68.1		67.8	0.3
	2-methylpyridine	73.1	73.4	−0.3	71.8		72.0	−0.2
	2,6-dimethylpyridine	90.4	91.2	−0.9	88.6		89.5	−0.9
	2-propylpyridine	109.3	109.1	0.2	107.2		107.0	0.2
nucleobase	uracil	54.8	54.5	0.4	54.0		53.7	0.3
	cytosine, $C_4H_5N_3O$	62.0	61.3	0.7	60.6		60.3	0.3
	adenine	69.7	69.8	−0.1	69.2		68.6	0.6
	thymine	72.4	72.3	0.1	71.1		71.2	−0.1
	guanine	73.6	73.9	−0.3	72.4		72.8	−0.4
amino acid	glycine, Gly	50.3	50.2	0.0	48.1		49.0	−0.9
	cysteine, Cys	68.0	67.5	0.5	66.5		66.0	0.5
	alanine, Ala	68.1	68.1	0.0	66.5		66.5	0.0
	serine, Ser	71.3	71.0	0.3	69.8		69.4	0.4
	aspartate, Asp	77.7	77.1	0.6	76.2		75.0	1.2
	asparagine, Asn	85.1	85.7	−0.6	83.2		83.5	−0.3
	threonine, Thr	88.7	88.9	−0.2	86.6		86.9	−0.3
	proline, Pro	91.3	91.3	−0.1	89.3		89.8	−0.5
	glutamic acid, Glu	95.7	95.0	0.8	93.8		92.5	1.3
	histidine, His	100.8	101.4	−0.6	98.9		99.9	−1.0
	glutamine, Gln	103.1	103.6	−0.5	100.3		101.0	−0.7
	valine, Val	103.7	103.8	−0.1	101.3		101.5	−0.2
	methionine, Met	104.6	104.7	−0.1	102.2		102.3	−0.1
	phenylalanine, Phe	119.3	119.1	0.2	117.1		117.0	0.1
	leucine, Leu	121.8	121.7	0.1	119.0		119.0	0.0
	isoleucine, Ile	121.8	121.7	0.2	119.1		119.0	0.1
	tyrozine, Tyr	121.8	122.6	−0.8	119.6		119.9	−0.3
	lysine, Lys	133.4	132.5	0.9	130.3		129.5	0.8
	tryptophane, Trp	137.8	137.8	0.0	135.3		135.4	−0.1
	arginine, Arg	140.3	141.0	−0.7	137.2		137.9	−0.7
sugar	erythrose, $C_4H_8O_4$	80.5	80.3	0.2	78.9		78.4	0.5
	ribose, $C_5H_{10}O_5$	101.7	101.1	0.7	99.6		98.8	0.8
	fructose, $C_6H_{12}O_6$	122.2	121.9	0.4	119.9		119.2	0.7
	altrose, $C_6H_{12}O_6$	124.1	124.6	−0.5	121.6		122.0	−0.4
	galactose, $C_6H_{12}O_6$	124.1	124.6	−0.5	121.8		122.0	−0.2
	glucose, $C_6H_{12}O_6$	124.4	124.6	−0.3	121.7		122.0	−0.3
	mannopyranose, $C_6H_{12}O_6$	124.5	124.6	−0.1	121.8		122.0	−0.2
	mannofuranose, $C_6H_{12}O_6$	125.0	124.6	0.4	122.2		122.0	0.2
	allose, $C_6H_{12}O_6$	125.0	124.6	0.3	121.1		122.0	−0.9
dipeptide	For-Gly-NH_2	64.6	64.2	0.4	63.3		62.5	0.8
	For-Ala-NH_2	81.7	82.1	−0.4	80.3		80.0	0.3
B-containing	BCl_3	4.8	4.8	0.0	4.7		4.8	−0.1
molecule	BF_3	7.9	7.9	0.0	7.7		7.7	0.0
F-containing	perfluoromethane, CF_4	10.8	10.8	0.0	10.5		10.5	0.0
molecule	perfluoroethylene, C_2F_4	13.5	13.6	−0.1	13.4		13.5	0.0
	difluoromethane, CH_2F_2	20.7	20.2	0.5	20.2		19.6	0.6

Table 2. continued

molecular type	molecule	harm B3LYP/6-31G*	emp	Δ	anharm B3LYP/large	expt	emp	Δ
	methyl fluoride, CH$_3$F	24.8	25.0	−0.2	24.2		24.2	0.0
	perfluorocyclobutene, C$_4$F$_6$	25.3	24.8	0.4	25.0		24.8	0.2
Si-containing molecule	silane, SiH$_4$	30.9	31.0	−0.1	19.3		19.1	0.2
	disilane, Si$_2$H$_6$	15.2	15.2	0.0	30.3		30.5	−0.2
P-containing molecule	phosphine, PH$_3$	24.8	25.0	−0.2	14.8		14.8	0.0
S-containing molecule	hydrogen sulfide, H$_2$S	9.5	9.9	−0.3	9.4		9.5	−0.1
	methanethiol, CH$_3$SH	29.1	29.1	0.0	28.4		28.2	0.2
Cl-containing molecule	perchloromethane, CCl$_4$	5.8	6.6	−0.8	5.8		6.6	−0.8
	perchloroethylene, C$_2$Cl$_4$	9.6	9.5	0.2	9.6		9.6	0.0
	dichloromethane, CH$_2$Cl$_2$	18.6	18.2	0.4	18.2		17.7	0.5
	methyl chloride, CH$_3$Cl	23.9	23.9	0.0	23.3		23.2	0.1
other molecule	SiCl$_4$	4.6	4.2	0.4	4.5		4.2	0.3
	PF$_3$	5.3	5.3	0.0	5.2		5.2	0.0
	NF$_3$	6.5	7.2	−0.8	6.4		7.2	−0.9
	SiF$_4$	8.0	8.3	−0.4	7.8		8.0	−0.3
	carbonyl fluoride, F$_2$CO	8.9	8.4	0.5	8.7		8.0	0.8
	heavy water, D$_2$O	9.8	9.8	0.0	9.7	9.7	9.7	0.0
	H$_2$O	13.3	14.6	−1.3	*13.2*	13.3	13.6	−0.3
	ketene, CH$_2$CO	19.9	19.8	0.2	*19.5*	19.5	19.5	0.1
	NH$_3$	21.7	21.4	0.2	*21.3*	21.3	20.9	0.4
	deuterated ethylene, C$_2$D$_4$	24.4	24.7	−0.3	*24.2*	23.9	23.9	0.0
	methylcyanide, CH$_3$CN	28.6	28.7	−0.1	28.0		28.2	−0.2
	NH$_2$COOH	32.3	32.4	−0.1	31.2		31.5	−0.3
	hydrazine, N$_2$H$_4$	33.6	33.6	0.0	32.8		32.3	0.5
	methylformate, HCO(O)CH$_3$	39.2	38.5	0.7	37.3		36.9	0.4
	urea, NH$_2$CONH$_2$	40.2	41.1	−0.8	39.0		39.8	−0.8
	NH$_2$CH$_2$COOH	50.2	50.2	0.0	49.2		49.0	0.2
	mercaptoethanol	50.2	49.9	0.2	49.6		48.7	0.9
	dihydroxyacetone	59.1	59.5	−0.4	57.9		57.9	0.0
	methylglycinate	67.9	67.2	0.7	65.3		65.0	0.3
	sarcosine	68.0	67.5	0.5	66.5		66.1	0.4
	NH$_2$(CH$_2$)$_2$COOH	68.4	68.1	0.3	66.9		66.5	0.4
	NH$_2$(CH$_2$)$_3$COOH	86.4	86.0	0.4	84.5		84.0	0.5
	sarine	91.6	91.6	0.0	89.4		89.4	0.0
	naphthalene, C$_{10}$H$_8$	92.8	92.9	−0.1	91.2		91.6	−0.4
	hydroxyproline	93.7	94.3	−0.6	91.9		92.8	−0.9
	methionine sulfoxide	106.7	106.7	0.0	104.5		104.5	0.0
	cystine	125.9	126.1	−0.1	122.9		123.3	−0.4
	nicotine	142.9	142.5	0.4	139.9		140.5	−0.6
	N-acetylglucoseamine	153.8	153.7	0.1	150.5		150.4	0.1

^{a}Large = 6-311+G(3df,2pd). When "experimental" (expt) values were employed in the linear least-squares refinement concerning Set II (see Section 5 for the details), the estimated anharmonic B3LYP/6-311+G(3df,2pd) values, not used in the analysis, are presented in italics. The anharmonic B3LYP/6-311+G(3df,2pd) ZPVEs were obtained via a simple sum of the B3LYP/6-311+G(3df,2pd) harmonic value and the anharmonic correction obtained at the B3LYP/6-31G* level.

exponent Gaussian basis sets were used in the DFT computations, namely, a small 6-31G*[54] and a medium-sized 6-311+G(3df,2pd)[55] basis. For nonlinear molecules having non-Abelian point groups, the ZPVE energies were obtained after slightly distorting the molecule to break degeneracy. The VPT2 facility of Gaussian09 is based upon refs 56 and 57. Unlike previous versions, version D.01 of Gaussian09 utilizes the formula proposed by Schuurman et al.[58] to compute resonance-free ZPVE estimates for asymmetric tops.

The atom types utilized in this study, see eqs 1−3, strictly correspond to those defined in the MM2 variant of molecular mechanics (MM).[23] The Chem3D program[59] was employed to determine atom types within the molecules as well as the starting geometries of the molecules for the quantum chemical geometry optimizations. It must be noted that since ATB ZPVEs are insensitive to conformational changes, it was not checked which conformer a given MM2 starting geometry corresponds to and whether the optimized structure is the "global" minimum in the conformational space of the given molecule. This is an especially relevant remark for amino acids,[60] but the consequences are well tolerable within the present study.

4. HIGHLY ACCURATE ZPVES OF SMALL MOLECULES

For a number of small molecules highly accurate ZPVE values are available from the literature; a subset of these[38,61−70] is collected into Table 1. Except for the data for H$_2$O$_2$ and D$_2$O$_2$,

Table 3. List of Atom-Type-Based (ATB) Increments, in kcal mol^{-1}, Applicable to Empirical ZPVE Estimates

atom	atom-type	no. of occurences		atom-type-based increments	
		molecule	atom	harmonic B3LYP/6-31G*	recommended
H	H, nonpolar	134	842	5.9374	5.6057
	>NH, amine	41	94	6.5552	6.0377
	−H amide, enol	10	20	6.6531	6.1827
	−OH, alcohol	24	64	6.0558	6.1726
	−COOH, carboxyl	34	38	6.7789	7.0989
	−SiH	2	10	4.0034	3.7998
	−PH	1	3	4.4976	4.2769
	−SH, thiol	4	5	4.4892	4.3763
	D, nonpolar	2	6	3.1815	3.6685
B	>B−, trigonal	2	2	4.2837	4.5445
C	C(sp^3), alkane	108	320	5.9741	6.2918
	C(sp^2), alkene	20	57	4.3972	4.6305
	C(sp), alkyne,C=C=O	7	12	2.6492	2.8766
	C(sp^2), carbonyl	60	69	3.4202	3.5050
	benzene	22	123	4.5400	4.6272
	cyclopropane	4	10	5.2226	5.6437
N	N(sp^3)	41	47	3.6749	4.1331
	N(sp^2), amide	11	16	4.2329	4.5871
	N(sp^2), pyrrole	7	9	3.0221	3.0917
	N(sp)	2	2	2.2555	2.2401
	−N= azo, pyridine	8	11	3.1483	3.0902
	=N− imine, oxime	3	4	2.5429	2.4601
	NO$_2$	2	3	2.2884	2.4612
O	−O−, alcohol, ether	65	123	2.8259	2.3599
	=O, carbonyl	63	76	2.5795	2.3320
	=O, carboxylate	1	2	2.8447	2.7063
F	F, fluoride	12	34	1.2016	1.0486
Si	Si, silane	4	5	3.5025	3.8596
P	>P− phosphine	2	2	1.7181	2.0132
	phosphorus(V)	1	1	1.6738	2.4731
S	−S− sulfide	6	7	0.8744	0.7424
	sp^2, thiophene	1	1	0.0663	0.1874
	sulfoxide	1	1	0.3432	0.6817
Cl	Cl, chloride	6	18	0.1632	0.0601

which come from DMC computations,[38] the accurate ZPVE values reported in Table 1 for the other 14 molecules come from variational nuclear motion computations, based on an exact kinetic energy operator and accurate representations of the ground electronic state PESs of the molecules. Uncertainties of ZPVE values determined this way could be obtained using the standard harmonic sum definition of the ZPVE as follows: a 2σ uncertainty of the computed ZPVE can be taken simply as the signed sum of the deviation of the fundamentals from their measured counterparts, readily available for the small molecules in question. These estimates are not given in Table 1 as they are so small that they, being on the order of a couple of cm^{-1}, are irrelevant for the present study.

In Table 1, the highly accurate ZPVE values are compared to computed estimates obtained from harmonic force fields, determined at four levels of electronic structure theory. For all but the linear molecules anharmonic estimates, obtained within the VPT2 approach, are also given. There was no attempt here to be comprehensive; a more detailed assessment of the performance of different DFT and wave function techniques is out of the scope of the present study.

To understand the accuracy of quantum chemically estimated ZPVEs,[71] especially those of larger molecules, it is important to notice in Table 1 that the anharmonic contributions to the ZPVEs appear to be model-independent even if subchemical (calibration)[72] accuracy is desired. In fact, even the anharmonic B3LYP/6-31G* values are accurate for almost all practical purposes, including those of the present investigation. A conclusion that can be drawn from Table 1 is that for most larger molecules accurate anharmonic ZPVE values can be obtained as a sum of B3LYP/6-311+G(3df,2pd) harmonic and B3LYP/6-31G* anharmonic contributions. Of course, the large basis DFT results can be replaced by even more accurate large basis CCSD(T) harmonic values if even higher accuracy is required (note that B3LYP performs better than frozen-core MP2). Nevertheless, B3LYP/6-31G* DFT results appear to provide excellent reference values for the present investigation. It is also clear from this table that a harmonic B3LYP/6-31G* ZPVE scale factor of 0.984 would greatly improve the agreement between these approximate computed harmonic values and the highly accurate "fully anharmonic" experimental ZPVEs. A similar scale factor can be obtained for the B3LYP/6-311+G(3df,2pd) harmonic ZPVEs; it is 0.985. Note that, while the individual differences among the ZPVE estimates are somewhat different for the B3LYP and MP2 values, there is considerable similarity for the scale factors, and they are rather close to one.

5. ATOM TYPES

Table 2 contains the ZPVE values computed for the two reference sets of the present investigation, including 27 molecular classes and 156 molecules. The first set of reference ZPVE values, Set I, was simply obtained as harmonic B3LYP/6-31G* values. Set I is only used to test the concept of this paper (see Section 6). The second, composite set (Set II) contains variational ("experimental") ZPVEs, when available (see Table 1 for the list of such molecules), and estimated anharmonic ZPVE values obtained from a harmonic B3LYP/6-311+G(3df,2pd) value amended by anharmonic corrections obtained at the B3LYP/6-31G* level.

The test molecules include chained and branched alkanes up to C_8, alkenes, cycloalkanes and cycloalkenes, alkynes, alcohols, aldehydes, carboxylic acids, amines, molecules containing the nitro group, amides, ketones, ethers, esters, benzene derivatives, heterocycles, nucleobases, all the natural amino acids, dipeptides, sugars, F-, Si-, P-, S-, and Cl-containing molecules, as well as a large set of "mixed" molecules containing several structural elements. To avoid "strange" atom types, unusual from the point of view of the present reference list, diatomic molecules were left out of consideration.

There are 75 atom types in the MM2 force field approach of Allinger.[23] Not all the atoms are of interest for the present study; here, 8, 6, 7, 3, 2, and 3 atom types occur for H, C, N, O, P, and S, respectively, while only one atom type was utilized for D, B, F, Si, and Cl. Note that although within MM2 Allinger[23] did not define a separate H atom type when hydrogen is connected to P or Si, it turned out that this distinction is important for the present study, and thus strictly in the cases of H(P) and H(Si), we needed to deviate from the MM2 recommendations. Note that MM force fields developed later than MM2 started to employ considerably more atom types to increase the accuracy of their predictions.

Furthermore, since in the MM2 formalism cyclopropane has a separate C atom type, cyclopropane was left in the reference set, but the reproduction of the computed ZPVE and the increment itself should not be taken literally as they are based on only a single reference.

The reference B3LYP/6-31G* ZPVE values (Set I) received equal weight during the linear least-squares refinement of the ZPVE increments. The uncertainties chosen for the refinement of the ZPVE increments of Set II are 0.1 kcal mol^{-1} for the variational ZPVE data and 0.5 kcal mol^{-1} for the anharmonic data.

Table 3 contains the information how many times a given atom type occurs in the reference set, both as part of molecules and individually. More importantly, Table 3 lists one of the end results of this study, the ATB ZPVE increments determined during the least-squares refinements based on the two reference data sets.

6. DISCUSSION

The proof of the concept of the present study utilized the harmonic B3LYP/6-31G* ZPVE values, given in Table 2, as use of a single level of electronic structure theory ensures the most consistent set of reference ZPVE values against which the least-squares estimation scheme can be checked.

The data presented in Table 2, and the associated Figure 1, clearly prove the viability and utility of the proposed scheme. As can be expected at the end of a linear least-squares fit, the signed average deviation is very close to 0.00 both for Set I and
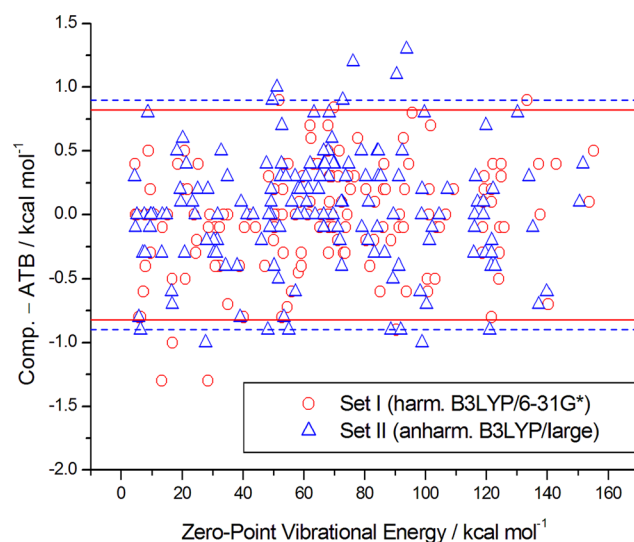

Figure 1. Differences between the DFT-computed Set I, the mixed Set II, and the empirical ATB ZPVE estimates of this study. For further details see the Δ values of Table 2, upon which this Figure is based. The horizontal full and dotted lines correspond to ±2 rms deviations for Set I and Set II, respectively.

Set II. More informative is the rms deviation between the reference and the predicted ATB ZPVE values. In the case of Set I, the rms deviation is a mere 0.4 kcal mol^{-1} (twice the rms value is indicated as a horizontal straight line in Figure 1). Thus, the incremental ATB ZPVE estimation scheme, even with the limited set of atom types prescribed by the MM2 approach, can be expected to yield ZPVE estimates of a large number of organic molecules with much better than chemical accuracy, approaching calibration accuracy.

Figure 1 vividly demonstrates that the accuracy of the empirical ATB ZPVE estimations does not worsen for larger ZPVE values. The predictive power of the present ATB ZPVE estimation scheme can be further tested when applied to molecules considerably larger than those in the reference set. Vitamins provide an excellent test set for this purpose. In Table 4, the estimated ZPVE values, based on Set I ATB increments,

Table 4. Harmonic B3LYP/6-31G* and Empirical Atom-Type-Based (ATB) ZPVE Estimates, in kcal mol^{-1}, for Several Vitamins

name	computed ZPVE	ATB ZPVE	difference
vitamin A (retinol)	285.6	285.2	0.3
vitamin B$_2$ (riboflavin)	234.9	235.3	−0.4
vitamin C (ascorbic acid)	94.2	94.8	−0.6
vitamin K$_1$ (phylloquinone)	444.7	444.2	0.5
vitamin K$_2$	398.8	399.1	−0.2

are compared with the computed harmonic ones for vitamins A (retinol), B$_2$ (riboflavin), C (ascorbic acid), K$_1$ (phylloquinone), and K$_2$. As can be seen, reproduction of the computed values is outstanding, well within the stated goal of this study, 1 kcal mol^{-1}.

The extensive set of data obtained for hydrocarbons, presented in Table 2, supports a number of interesting observations: (1) On the basis of the present results, there appears to be no real need to distinguish between midchain and end-of-chain carbon atoms in the case of paraffins; chemical accuracy can be maintained with the "standard" MM2 atom

types. (2) n-Paraffins always have the highest computed ZPVE among the isomeric C_n forms. For the molecular formula C_4H_{10}, n-butane has a ZPVE higher than that of isobutane by 0.3 kcal mol^{-1}. For C_5H_{12}, the order is neopentane, isopentane, and n-pentane; the spread of the values is small, just 0.6 kcal mol^{-1}. For C_6H_{14}, the order of the computed ZPVEs is again neohexane, isohexane, and n-hexane; the spread of the ZPVE values of the five possible isomers is only 0.6 kcal mol^{-1}, whereby the branched paraffins 2,3-dimethylbutane and 3-methyl-pentane have ZPVE values in between those of neohexane and isohexane. Since the ATB ZPVE estimates of all the n-paraffins are always slightly smaller than the reference values, and in the case of n-octane the difference is still only 0.7 kcal mol^{-1}, and in the n-hexane, n-heptane, n-octane series the increase of the error is a mere 0.1 kcal mol^{-1} per CH_2 unit, the ZPVE estimates of all branched paraffins should be highly accurate and expected to be within chemical accuracy at least up to C_{15}. (3) As already noted by Schulman and Disch,[11] the difference between the ZPVEs of alkanes and their alkene counterparts seem to be very regular, 15.0 and 15.1 kcal mol^{-1} for C_3 and C_4, respectively. The estimated deviations are 15.4 and 15.0 for C_3 and C_4, respectively, perfectly adequate for accurate predictions.

The ZPVE contribution of the methylene (CH_2) group discussed by Pitzer and Catalano[3] is now defined as $\zeta(C,sp^3) + 2\zeta(H,\text{nonpolar})$, which turns out to be 17.9 for the B3LYP/6-31G* harmonic test set, while it is 17.5 based on the recommended ATB increments of Table 3, based on the anharmonic Set II reference set. The difference between the ZPVE values of Gly and Ala, as well as of benzene and toluene, are fully consistent with the $\zeta(C,sp^3) + 2\zeta(H,\text{nonpolar})$ estimate and the additivity assumption.

While the present ATB estimation scheme cannot distinguish among the ZPVEs of conformers, a particularly stringent test of the present scheme is when ZPVE values of isomers are determined. The present reference set of molecules contains several relevant and interesting cases: (1) The case of normal and branched alkanes has been discussed above, and our ATB scheme performs extremely well, because the isomeric computed ZPVE values are always found to be highly similar. (2) The three molecules obtained by replacing a H atom of glycine with a methyl group are the amino acid alanine (replacing one of the α-hydrogens with CH_3), sarcosine (replacing a N−H hydrogen with CH_3), and methylglycinate (replacing the O−H hydrogen with CH_3). The three structural isomers have harmonic B3LYP/6-31G* ZPVE values in the narrow range of 67.9−68.1 kcal mol^{-1}, while the ATB ZPVEs are between 67.2 and 68.1 kcal mol^{-1}. Therefore, the reproduction of the computed ZPVEs is significantly better than the stated goal of chemical accuracy of 1 kcal mol^{-1}. (3) Sugars provide another interesting and important test case. While erythrose and ribose are unique sugars of their kind, there are six C_6 sugars in the reference set: altrose, galactose, glucose, mannopyranose, mannofuranose, and allose. While it is true that our ZPVE prediction scheme cannot distinguish between these sugars, thus all have a predicted value of 124.6 kcal mol^{-1}, the computed B3LYP/6-31G* ZPVE values are within the narrow range of 124.1−125.0 kcal mol^{-1}; thus, reproduction of these reference values is outstanding, once again.

The ZPVE increments associated with third-row elements require detailed consideration; the preliminary data presented suggest the following: (1) The problem with diatomic molecules can be exemplified on HCl, not part of the training set, whose ZPVE is a mere 4.2 kcal mol^{-1}. The H atom data of Table 3 show that there are no atom types with ZPVE increments below 5.9 kcal mol^{-1} when H is bonded to a second-row atom. Thus, third-row atom types require considerably lower increments, and there is no real reason to develop a H(Cl) ZPVE increment. (2) The ZPVE of CCl_4 is 5.8 kcal mol^{-1}, whereas $\zeta(C,sp^3)$ is 6.3. The case of C_2Cl_4 is slightly better though still bad, as $\zeta(C,sp^2)$ is 4.66, while the ZPVE of C_2Cl_4 is 9.6, suggesting an overly low Cl increment. Nevertheless, the CH_2Cl_2 and CH_3Cl data also suggest that the Cl atom-type ZPVE increment must be low, (well) below 0.5 kcal mol^{-1}. (3) An alternative to the present approach is the utilization of several atom types for third-row elements and atoms attached to them. A step in this direction is taken in the present study, and we introduce H atom types attached to Si and P atoms to make the ZPVE values of SiH_4 and PH_3 compatible with other molecular ZPVE estimates. This approach is not developed fully in the present study but can be done later, and perhaps it should be based on atom types defined by an MM approach other than MM2. As is clear from Table 3, and also in agreement with expectation, the increments are substantially lower when the H atom is attached to a third-row element as compared to a second-row element: while $\zeta(H,\text{third row})$ is ∼4.3 kcal mol^{-1}, $\zeta(H,\text{second row})$ is ∼6.3 kcal mol^{-1}.

The present study also supports, as well as shows the limitations of, the atom-based (stoichiometry-based) incremental schemes of Flanigan et al.,[9] Schulman and Disch,[11] and Ibrahim and Fataftah.[12] In some of these studies there is a large zero-order term (−6.19 kcal mol^{-1}) making direct comparison of the results difficult, but it can be stated that the values of ∼4 for C and ∼7 for H are supported, as averaging the H and C ATB increments we obtain ∼5.3 for C and ∼6.2 for H. The average N and O atomic increments are 3.4 and 2.5, in reasonable agreement with the values of 4.05 and 3.4 advocated by Ibrahim and Fataftah,[12] who employed a zero-order term and thus needed to come up with higher atomic increments than those obtained in this study.

The present study also offers limited evidence how substituting H to D affects the ZPVE value of the molecule. While the recommended H(nonpolar) ATB value is 5.6057, the D(nonpolar) value is considerably smaller, 3.6685. As expected, the D(nonpolar) increment is reasonably close to the $\sqrt{2H}$ (nonpolar) increment.

The ATB increments can be combined to provide contributions of usually defined chemical groups. A selected few group ZPVE increments based on the recommended Set I and Set II ATB increments are listed in Table 5; a lot more can be derived based on the ATB increments of Table 3. It is perhaps preferable to maintain ATB increments as it is easier to make the recognition of atom types more automatic than that of chemical groups. The "recommended" ZPVE group

Table 5. Group Contributions to Empirical ZPVE Estimates

group	B3LYP/6-31G* harmonic	ATB recommended
−CH_2−	17.85	17.50
−CH_3	23.79	23.11
−OH	8.88	8.53
−NH_2	16.78	16.21

increments are characteristically smaller than the harmonic contributions, just as expected.

7. CONCLUSIONS

The conclusions of general importance of this study are as follows:

(1) For most medium-sized and large neutral, closed-shell molecules of interest to organic chemistry ZPVE values can be estimated with better than chemical accuracy using the ATB increments determined in this study.

(2) The ATB ZPVE increments developed in this study allow to obtain ZPVE values not only at zero cost but also with an accuracy higher than the quantum chemical technique B3LYP/6-31G* offers via harmonic ZPVE estimates. A simple scaling of the B3LYP/6-31G* values by a factor of 0.984 improves the accuracy of the directly computed values and results in an accuracy comparable to that of the ATB scheme.

(3) Overall, determining ZPVE values of even large molecules can be done within the present scheme with zero cost. Introduction of further atom types could further improve the quality of the predictions, and the scheme proposed can be easily extended to further atoms, as well. Nevertheless, a detailed validation is necessary, as a careful selection of atom types is necessary to maintain the near-calibration accuracy observed in this study.

(4) The ZPVE values of isomers are very close to each other; for slightly larger systems, the deviations do not exceed chemical accuracy, that is, 1 kcal mol^{-1}.

■ AUTHOR INFORMATION

Notes
The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

■ REFERENCES
(1) Pitzer, K. S. Thermodynamics of Gaseous Hydrocarbons: Ethane, Propane, Propylene, *n*-Butane, Isobutane, 1-Butene, *Cis* and *Trans* 2-Butenes, Isobutene, and Neopentane (Tetramethylmethane). *J. Chem. Phys.* **1937**, *5*, 473−479.

(2) Cottrell, T. L. Binding Energies in Hydrocarbons. *J. Chem. Soc.* **1948**, 1448−1453.

(3) Pitzer, K. S.; Catalano, E. Electronic Correlation in Molecules. III. The Paraffin Hydrocarbons. *J. Am. Chem. Soc.* **1956**, *78*, 4844−4846.

(4) Bernstein, H. J. The Physical Properties of Molecules in Relation to Their Structure. I. Relations between Additive Molecular Properties in Several Homologous Series. *J. Chem. Phys.* **1952**, *20*, 263−269, 1328(E).

(5) Bernstein, H. J.; Pullin, A. D. E. A Sum Rule for the Vibrational Frequencies of Homologs. *J. Chem. Phys.* **1953**, *21*, 2188−2195.

(6) Bernstein, H. J. Constitutive Nature of the Heat Capacity of Gases. *J. Chem. Phys.* **1956**, *24*, 911.

(7) Bernstein, H. J. Bond and Interaction Contributions for Calculating the Heat of Formation, Diamagnetic Susceptibility, Molar Refraction and Volume, and Thermodynamic Properties of Some Substituted Methanes. *J. Phys. Chem.* **1965**, *69*, 1550−1564.

(8) Shingu, H.; Fujimoto, T. Empirical Correlation between Molecular Structure and Zero-Point Energies in Hydrocarbons. *J. Chem. Phys.* **1959**, *31*, 556−557.

(9) Flanigan, M. C.; Komornicki, A.; McIver, J. W., Jr. In *Electronic Structure Calculation*; Segal, G. A., Ed.; Plenum Press: New York, 1977.

(10) Arimoto, S. Preliminary Considerations Towards a Theoretical Foundation of the Zero Energy Additivity Rules. *Phys. Lett. A* **1985**, *113A*, 126−132.

(11) Schulman, J. M.; Disch, R. L. A Simple Formula for the Zero-Point Energies of Hydrocarbons. *Chem. Phys. Lett.* **1985**, *113*, 291−293.

(12) Ibrahim, M. R.; Fataftah, Z. A. Estimation of Zero-Point Energies of Compounds Containing N, O, Cl and F Atoms. *Chem. Phys. Lett.* **1986**, *125*, 149−154.

(13) Grice, M. E.; Politzer, P. Use of Molecular Stoichiometry to Estimate Vibrational Energy. *Chem. Phys. Lett.* **1995**, *244*, 295−298.

(14) Rahal, M.; Hilali, M.; El Hammadi, A.; El Mouhtadi, M.; El Hajbi, A. Calculation of Vibrational Zero-Point Energy. *J. Mol. Struct.: THEOCHEM* **2001**, *572*, 73−80.

(15) Oi, T.; Popowicz, A.; Ishida, T. Additivity of Zero-Point Energy. *J. Phys. Chem.* **1986**, *90*, 3080−3088.

(16) Oi, T.; Ishida, T. Correlation of Zero-Point Energy with Molecular Structure and Molecular Forces. 1. Development of the Approximation. *J. Phys. Chem.* **1983**, *87*, 1067−1073.

(17) Oi, T.; Ishida, T. Correlation of Zero-Point Energy with Molecular Structure and Molecular Forces. 3. Approximation for H/D Isotope Shifts and Linear Frequency Sum Rule. *J. Phys. Chem.* **1984**, *88*, 1507, 2057−2061.

(18) Fliszar, S.; Poliquin, F.; Badilescu, I.; Vauthier, E. Structure Dependent Regularities of Zero-Point Plus Heat Content Energies in Organic Molecules. *Can. J. Chem.* **1988**, *66*, 300−303.

(19) Ruzsinszky, A.; Kristyán, S.; Margitfalvi, J. L.; Csonka, G. I. Rapid Estimation of Zero-Point Energies of Molecules Using Hartree-Fock Atomic Partial Charges. *J. Phys. Chem. A* **2003**, *107*, 1833−1839.

(20) Benson, S. W. *Thermochemical Kinetics,* 2nd ed.; Wiley: New York, 1976.

(21) Ibrahim, M. R.; von Rague Schleyer, P. Atom Equivalents for Relating *ab initio* Energies to Enthalpies of Formation. *J. Comput. Chem.* **1985**, *6*, 157−167.

(22) Wiberg, K. B. Group Equivalents for Converting *ab initio* Energies to Enthalpies of Formation. *J. Comput. Chem.* **1984**, *5*, 197−199.

(23) Allinger, N. L. Conformational Analysis. 130. MM2. A Hydrocarbon Force Field Utilizing V_1 and V_2 Torsional Terms. *J. Am. Chem. Soc.* **1977**, *99*, 8127−8134.

(24) Császár, A. G.; Fábri, C.; Szidarovszky, T.; Mátyus, E.; Furtenbacher, T.; Czakó, G. Fourth Age of Quantum Chemistry: Molecules in Motion. *Phys. Chem. Chem. Phys.* **2012**, *14*, 1085−1106.

(25) East, A. L. L.; Allen, W. D. The Heat of Formation of NCO. *J. Chem. Phys.* **1993**, *99*, 4638−4650.

(26) Tajti, A.; Szalay, P. G.; Császár, A. G.; Kállay, M.; Gauss, J.; Valeev, E. F.; Flowers, B. A.; Vazquéz, J.; Stanton, J. F. HEAT: High Accuracy Extrapolated ab initio Thermochemistry. *J. Chem. Phys.* **2004**, *121*, 11599−11613.

(27) Ruscic, B.; Boggs, J. E.; Burcat, A.; Császár, A. G.; Demaison, J.; Janoschek, R.; Martin, J. M. L.; Rossi, M. J.; Morton, M. L.; Stanton, J. F.; et al. IUPAC Critical Evaluation of Thermochemical Properties of Selected Radicals. Part 1. *J. Phys. Chem. Ref. Data* **2005**, *34*, 573−656.

(28) Ruscic, B.; Litorja, M.; Asher, R. L. Ionization Energy of Methylene Revisited: Improved Values for the Enthalpy of Formation of CH_2 and the Bond Dissociation Energy of CH_3 via Simultaneous Solution of the Local Thermochemical Network. *J. Phys. Chem. A* **1999**, *103*, 8625−8633.

(29) Császár, A. G.; Allen, W. D.; Schaefer, H. F., III In Pursuit of the ab initio Limit for Conformational Energy Prototypes. *J. Chem. Phys.* **1998**, *108*, 9751−9764.

(30) Allen, W. D.; East, A. L. L.; Császár, A. G. In *Structures and Conformations of Non-Rigid Molecules*; Laane, J., Dakkouri, M., van der

Veken, B., Oberhammer, H., Eds.; Kluwer: Dordrecht, 1993; pp 343–373.

(31) Karton, A.; Rabinovich, E.; Martin, J. M. L.; Ruscic, B. W4 Theory for Computational Thermochemistry: In Pursuit of Confident sub-kJ/mol Predictions. *J. Chem. Phys.* **2006**, *125*, 144108.

(32) Ruscic, B.; Wagner, A. F.; Harding, L. B.; Asher, R. L.; Feller, D.; Dixon, D. A.; Peterson, K. A.; Song, Y.; Qian, X.; Ng, C. Y.; et al. On the Enthalpy of Formation of Hydroxyl Radical and Gas-Phase Bond Dissociation Energies of Water and Hydroxyl. *J. Phys. Chem. A* **2002**, *106*, 2727–2747.

(33) Ruscic, B.; Pinzon, R. E.; Morton, M. L.; von Laszevski, G.; Bittner, S. J.; Nijsure, S. G.; Amin, K. A.; Minkoff, M.; Wagner, A. F. Introduction to Active Thermochemical Tables: Several "Key" Enthalpies of Formation Revisited. *J. Phys. Chem. A* **2004**, *108*, 9979–9997.

(34) Császár, A. G.; Furtenbacher, T. From a Network of Computed Reaction Enthalpies to Atom-Based Thermochemistry (NEAT). *Chem. - Eur. J.* **2010**, *16*, 4826–4835.

(35) Curtiss, L. A.; Redfern, P. C.; Raghavachari, K.; Pople, J. A. Gaussian-3X (G3X) theory: Use of Improved Geometries, Zero-Point Energies, and Hartree–Fock Basis Sets. *J. Chem. Phys.* **2001**, *114*, 108–117.

(36) Császár, A. G. Anharmonic Molecular Force Fields. *WIREs Comput. Mol. Sci.* **2012**, *2*, 273–289.

(37) Rauhut, G. Efficient Calculation of Potential Energy Surfaces for the Generation of Vibrational Wave Functions. *J. Chem. Phys.* **2004**, *121*, 9313–9322.

(38) McCoy, A. B. Diffusion Monte Carlo Approaches for Investigating the Structure and Vibrational Spectra of Fluxional Systems. *Int. Rev. Phys. Chem.* **2006**, *25*, 77–107.

(39) Furtenbacher, T.; Császár, A. G.; Tennyson, J. MARVEL: Measured Active Rotational-Vibrational Energy Levels. *J. Mol. Spectrosc.* **2007**, *245*, 115–125.

(40) Furtenbacher, T.; Császár, A. G. MARVEL: Measured Active Rotational-Vibrational Energy Levels. II. Algorithmic Improvements. *J. Quant. Spectrosc. Radiat. Transfer* **2012**, *113*, 929–935.

(41) Császár, A. G.; Furtenbacher, T. Spectroscopic Networks. *J. Mol. Spectrosc.* **2011**, *266*, 99–103.

(42) Furtenbacher, T.; Árendás, P.; Mellau, G.; Császár, A. G. Simple Molecules as Complex Systems. *Sci. Rep.* **2014**, *4*, 4654.

(43) Clabo, D. A.; Allen, W. D.; Remington, R. B.; Yamaguchi, Y.; Schaefer, H. F., III A Systematic Study of Molecular Vibrational Anharmonicity and Vibration-Rotation Interaction by Self-Consistent-Field Higher-Derivative Methods. Asymmetric Top Molecules. *Chem. Phys.* **1988**, *123*, 187–239.

(44) Allen, W. D.; Yamaguchi, Y.; Császár, A. G.; Clabo, D. A.; Remington, R. B.; Schaefer, H. F., III A Systematic Study of Molecular Vibrational Anharmonicity and Vibration-Rotation Interaction by Self-Consistent-Field Higher Derivative Methods. Linear Polyatomic Molecules. *Chem. Phys.* **1990**, *145*, 427–466.

(45) Nielsen, H. H. The Vibration-Rotation Energies of Molecules. *Rev. Mod. Phys.* **1951**, *23*, 90–136.

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

(47) CFOUR, Coupled-Cluster techniques for Computational Chemistry, a quantum-chemical program package by Stanton, J. F.; Gauss, J.; Harding, M. E.; Szalay, P. G. with contributions from Auer, A. A.; Bartlett, R. J.; Benedikt, U.; Berger, C.; Bernholdt, D. E.; Bomble, Y. J.; Cheng, L.; Christiansen, O.; Heckert, M.; Heun, O.; et al. and the integral packages MOLECULE (Almlöf, J.; Taylor, P.R.), PROPS (Taylor, P.R.), ABACUS (Helgaker, T.; Jensen, H. J. Aa.; Jørgensen, P.; Olsen, J.), and ECP routines by A. V. Mitin and C. van Wüllen. For the current version, see http://www.cfour.de. Most recently accessed: September 9, 2015.

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

(49) Ringholm, M.; Jonsson, D.; Bast, R.; Gao, B.; Thorvaldsen, A. J.; Ekström, U.; Helgaker, T.; Ruud, K. Analytic Cubic and Quartic Force Fields Using Density-Functional Theory. *J. Chem. Phys.* **2014**, *140*, 034103.

(50) Allen, W. D.; Császár, A. G. On the ab initio Determination of Higher-Order Force Constants at Nonstationary Reference Geometries. *J. Chem. Phys.* **1993**, *98*, 2983–3015.

(51) Becke, A. D. Density-Functional Thermochemistry. III. The Role of Exact Exchange. *J. Chem. Phys.* **1993**, *98*, 5648–5652.

(52) Lee, C. T.; Yang, W. T.; 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*, 785–789.

(53) Scott, A. P.; Radom, L. Harmonic Vibrational Frequencies: An Evaluation of Hartree–Fock, Møller–Plesset, Quadratic Configuration Interaction, Density Functional Theory, and Semiempirical Scale Factors. *J. Phys. Chem.* **1996**, *100*, 16502–16513.

(54) Hariharan, P. C.; Pople, J. A. The Influence of Polarization Functions on Molecular Orbital Hydrogenation Energies. *Theoret. Chim. Acta* **1973**, *28*, 213–222.

(55) Krishnan, R.; Binkley, J. S.; Seeger, R.; Pople, J. A. Self-Consistent Molecular Orbital Methods. XX. A Basis Set for Correlated Wave Functions. *J. Chem. Phys.* **1980**, *72*, 650–654.

(56) Barone, V. Anharmonic Vibrational Properties by a Fully Automated Second-Order Perturbative Approach. *J. Chem. Phys.* **2005**, *122*, 014108.

(57) Bloino, J.; Biczysko, M.; Barone, V. General Perturbative Approach for Spectroscopy, Thermodynamics, and Kinetics: Methodological Background and Benchmark Studies. *J. Chem. Theory Comput.* **2012**, *8*, 1015–1036.

(58) Schuurman, M. S.; Allen, W. D.; von Rague Schleyer, P.; Schaefer, H. F., III The Highly Anharmonic BH_5 Potential Energy Surface Characterized in the ab initio Limit. *J. Chem. Phys.* **2005**, *122*, 104302.

(59) See https://www.cambridgesoft.com/Ensemble_for_Chemistry/ChemOffice/ChemOfficeProfessional/.Chem3D. Most recently accessed: September 9, 2015.

(60) Császár, A. G.; Perczel, A. Ab initio Characterization of Peptide Building Units. *Prog. Biophys. Mol. Biol.* **1999**, *71*, 243–309.

(61) Czakó, G.; Furtenbacher, T.; Császár, A. G.; Szalay, V. Variational Vibrational Calculations Using High-Order Anharmonic Force Fields. *Mol. Phys.* **2004**, *102*, 2411–2423.

(62) Tarczay, G.; Császár, A. G.; Polyansky, O. L.; Tennyson, J. Ab initio Rovibrational Spectroscopy of H_2S. *J. Chem. Phys.* **2001**, *115*, 1229–1242.

(63) Barletta, P.; Shirin, S. V.; Zobov, N. F.; Polyansky, O. L.; Tennyson, J.; Valeev, E. F.; Császár, A. G. The CVRQD *ab initio* Ground-State Adiabatic Potential Energy Surfaces for the Water Molecule. *J. Chem. Phys.* **2006**, *125*, 204307.

(64) Kozin, I. N.; Law, M. M.; Tennyson, J.; Hutson, J. M. Calculating Energy Levels of Isomerizing Tetra-Atomic Molecules. II. The Vibrational States of Acetylene and Vinylidene. *J. Chem. Phys.* **2005**, *122*, 064309.

(65) Czakó, G.; Nagy, B.; Tasi, G.; Somogyi, A.; Simunek, J.; Noga, J.; Braams, B. J.; Bowman, J. M.; Császár, A. G. Proton Affinity and Enthalpy of Formation of Formaldehyde. *Int. J. Quantum Chem.* **2009**, *109*, 2393–2409.

(66) Fábri, C.; Mátyus, E.; Furtenbacher, T.; Mihály, B.; Zoltáni, T.; Nemes, L.; Császár, A. G. Variational Quantum Mechanical and Active Database Approaches to the Rotational-Vibrational Spectroscopy of Ketene. *J. Chem. Phys.* **2011**, *135*, 094307.

(67) Mátyus, E.; Czakó, G.; Császár, A. G. Toward Black-Box-Type Full- and Reduced-Dimensional Variational (Ro)Vibrational Computations. *J. Chem. Phys.* **2009**, *130*, 134112.

(68) Fábri, C.; Furtenbacher, T.; Császár, A. G. A Hybrid Variational-Perturbational Nuclear Motion Algorithm. *Mol. Phys.* **2014**, *112*, 2462–2467.

(69) Bettens, R. P. A. Bound State Potential Energy Surface Construction: ab initio Zero-Point Energies and Vibrationally Averaged Rotational Constants. *J. Am. Chem. Soc.* **2003**, *125*, 584−587.

(70) Mátyus, E.; Czakó, G.; Sutcliffe, B. T.; Császár, A. G. Vibrational Energy Levels with Arbitrary Potentials using the Eckart-Watson Hamiltonians and the Discrete Variable Representation. *J. Chem. Phys.* **2007**, *127*, 084102.

(71) Grev, R. S.; Janssen, C. L.; Schaefer, H. F., III Concerning Zero-Point Vibrational Energy Corrections to Electronic Energies. *J. Chem. Phys.* **1991**, *95*, 5128−5132.

(72) Császár, A. G.; Allen, W. D.; Yamaguchi, Y.; Schaefer III, H. F. *Ab Initio* Determination of Accurate Potential Energy Hypersurfaces for the Ground Electronic States of Molecules. In *Computational Molecular Spectroscopy*; Jensen, P., Bunker, P. R., Eds.; Wiley: New York, 2000.