

# Heats of formation of the amino acids re-examined by means of W1-F12 and W2-F12 theories

Amir Karton · Li-Juan Yu · Manoj K. Kesharwani ·  
Jan M. L. Martin

Received: 20 January 2014 / Accepted: 15 March 2014 / Published online: 16 April 2014  
© Springer-Verlag Berlin Heidelberg 2014

**Abstract** We have obtained accurate heats of formation for the twenty natural amino acids by means of explicitly correlated high-level thermochemical procedures. Our best theoretical heats of formation, obtained by means of the ab initio W1-F12 and W2-F12 thermochemical protocols, differ significantly (RMSD = 2.3 kcal/mol, maximum deviation 4.6 kcal/mol) from recently reported values using the lower-cost G3(MP2) method. With the more recent G4(MP2) procedure, RMSD drops slightly to 1.8 kcal/mol, while full G4 theory offers a more significant improvement to 0.72 kcal/mol (max. dev. 1.4 kcal/mol for glutamine). The economical G4(MP2)-6X protocol performs equivalently at RMSD = 0.71 kcal/mol (max. dev. 1.6 kcal/mol for arginine and glutamine). Our calculations

are in excellent agreement with experiment for glycine, alanine and are in excellent agreement with the recent revised value for methionine, but suggest revisions by several kcal/mol for valine, proline, phenylalanine, and cysteine, in the latter case confirming a recent proposed revision. Our best heats of formation at 298 K ( $\Delta H_{f,298}^{\circ}$ ) are as follows: at the W2-F12 level: glycine -94.1, alanine -101.5, serine -139.2, cysteine -94.5, and methionine -102.4 kcal/mol, and at the W1-F12 level: arginine -98.8, asparagine -146.5, aspartic acid -189.6, glutamine -151.0, glutamic acid -195.5, histidine -69.8, isoleucine -118.3, leucine -118.8, lysine -110.0, phenylalanine -76.9, proline -92.8, threonine -149.0, and valine -113.6 kcal/mol. For the two largest amino acids, an average over G4, G4(MP2)-6X, and CBS-QB3 yields best estimates of -58.4 kcal/mol for tryptophan, and of -117.5 kcal/mol for tyrosine. For glycine, we were able to obtain a “quasi-W4” result corresponding to  $TAE_e = 968.1$ ,  $TAE_0 = 918.6$ ,  $\Delta H_{f,298}^{\circ} = -90.0$ , and  $\Delta H_{f,298}^{\circ} = -94.0$  kcal/mol.

Dedicated to the memory of Professor Isaiah Shavitt and published as part of the special collection of articles celebrating his many contributions.

**Electronic supplementary material** The online version of this article (doi:10.1007/s00214-014-1483-8) contains supplementary material, which is available to authorized users.

A. Karton (✉) · L.-J. Yu  
School of Chemistry and Biochemistry, The University  
of Western Australia, Perth, Australia  
e-mail: amir.karton@uwa.edu.au

M. K. Kesharwani · J. M. L. Martin (✉)  
Department of Organic Chemistry, Weizmann Institute  
of Science, 76100 Rehovot, Israel  
e-mail: gershon@weizmann.ac.il

J. M. L. Martin  
Department of Chemistry and Center for Advanced Scientific  
Computing and Modeling (CASCaM), University of North  
Texas, Denton, TX 76201, USA

**Keywords** Thermochemistry · Amino acids ·  
Explicitly correlated methods · Density functional theory ·  
Ab initio

## 1 Introduction

Due to the increasing computational power provided by supercomputers and recent advances in the development of economical ab initio methods (e.g., advances in explicitly correlated techniques [1–4]), high-level ab initio methods have now been refined to the point where they are applicable to biologically relevant systems (see Refs. [5–13] for some recent studies). Proteinogenic amino acids are the

most basic building blocks of proteins and play key roles in protein structure and function. They also serve as precursors of many biologically relevant molecules, such as polypeptides, nucleotides, hormones, neurotransmitters, and antioxidants [14]. Despite their importance, the experimental gas-phase heats of formation of most of the natural amino acids are not accurately known. Determination of these fundamental thermochemical quantities may be important in understanding why nature chose these molecules as fundamental biological building blocks—for example, by comparing the relative stabilities of  $\alpha$ - versus  $\beta$ -amino acids [15, 16]. Accurate heats of formation for the amino acids are also important from a theoretical point of view, e.g., for the validation and parameterization of computationally cost-effective procedures such as density functional theory, semiempirical molecular orbital theory, and molecular mechanics. In recent years, a large number of theoretical studies were dedicated to obtaining thermochemical properties of amino acids using high-level theoretical procedures [15–25].

In the present work, we obtain accurate theoretical heats of formation for the lowest-energy conformers for the 18 proteinogenic amino acids using the high-level, ab initio W1-F12 and W2-F12 thermochemical procedures [31]. These thermochemical procedures represent layered extrapolations to the all-electron, relativistic CCSD(T)/CBS energy (complete basis set limit coupled cluster with singles, doubles, and quasiperturbative triple excitations) and can achieve an accuracy in the sub-kcal/mol range for molecules whose wave functions are dominated by dynamical correlation [31, 32]. We use these benchmark values to evaluate the performance of a variety of  $Gn$ -type procedures [33] that were recently used for obtaining accurate thermochemical properties of amino acids [15–20].

The present paper also seeks to pay tribute to the scientific achievements of Prof. Isaiah Shavitt (OBM) and specifically to his seminal contributions to coupled cluster theory [26], to the theory and development of Gaussian basis sets [27], to accurate applied quantum chemistry [28–30], and to computational biochemistry [5].

## 2 Computational details

Most calculations were run on the CRUNTCh (Computational Research at UNT in Chemistry) Linux farm at the University of North Texas, on the high-performance computing National Computational Infrastructure (NCI) National Facility at Canberra, and on the iVEC@UWA facilities. Some additional calculations were carried out on the Faculty of Chemistry Linux farm at the Weizmann Institute of Science.

The geometries have been optimized at the B3LYP/A'VTZ level of theory [34–36] (where A'VTZ indicates the combination of the standard correlation-consistent cc-pVTZ basis set on hydrogen, [37] the aug-cc-pVTZ basis set on first-row elements, [38] and the aug-cc-pV(T+d)Z basis set on sulfur) [39]. All geometry optimizations and frequency calculations were performed using the Gaussian 09 program suite [40]. Benchmark relativistic, all-electron CCSD(T)/CBS energies were then obtained by means of our recently developed W1-F12 and W2-F12 thermochemical protocols [31] using the Molpro 2012.1 program suite [41]. The computational protocols of W1-F12 and W2-F12 theories have been specified and rationalized in reference [31].

In W1-F12 theory, the Hartree–Fock component is extrapolated from the VDZ-F12 and VTZ-F12 basis sets, using the  $E(L) = E_\infty + A/L^\alpha$  two-point extrapolation formula with  $\alpha = 5$  (where  $L$  is the highest angular momentum represented in the basis set, and  $VnZ$ -F12 denotes the cc-pVnZ-F12 basis sets of Peterson et al. [42] which were developed for explicitly correlated calculations). Optimal values for the geminal Slater exponents ( $\beta$ ) used in conjunction with the  $VnZ$ -F12 basis sets were taken from reference [43]. The valence CCSD-F12 correlation energy is extrapolated from the same basis sets, using the said two-point extrapolation formula. Extrapolation exponents ( $\alpha$ ) were taken from references [31, 43]. In all of the explicitly correlated coupled cluster calculations the diagonal, fixed-amplitude 3C(FIX) ansatz [45–47] and the CCSD-F12b approximation [48, 49] are employed. The (T) valence correlation energy is obtained in the same way as in the original Weizmann-1 (W1) theory, [50] i.e., extrapolated from the A'VDZ and A'VTZ basis sets using the above two-point extrapolation formula with  $\alpha = 3.22$ . The CCSD inner-shell contribution is calculated with the core-valence weighted correlation-consistent A'PWCVTZ basis set of Peterson and Dunning, [51] while the (T) inner-shell contribution is calculated with the PWCVTZ(no  $f$ ) basis set (where A'PWCVTZ indicates the combination of the cc-pVTZ basis set on hydrogen and the aug-cc-pwCVTZ basis set on carbon, and PWCVTZ(no  $f$ ) indicates the cc-pwCVTZ basis set without the  $f$  functions). The scalar relativistic contribution (in the second-order Douglas–Kroll–Hess approximation [52, 53]) is obtained as the difference between non-relativistic CCSD(T)/A'VDZ and relativistic CCSD(T)/A'VDZ-DK calculations [54] (where A'VDZ-DK indicates the combination of the cc-pVDZ-DK basis set on H and aug-cc-pV(D+d)Z-DK basis set on heavier elements). The atomic spin–orbit coupling terms are taken from the experimental fine structure, and the diagonal Born–Oppenheimer correction (DBOC) is calculated at the HF/A'VTZ level of theory. The zero-point vibrational energies (ZPVEs) are derived from B2PLYP/def2-TZVPP harmonic frequencies (and scaled by 0.9833, see Sect. 3.3).

In W2-F12, the Hartree–Fock component is calculated with the VQZ-F12 basis set. The valence CCSD-F12 correlation energy is extrapolated from the VTZ-F12 and VQZ-F12 basis sets, using the above two-point extrapolation formula with  $\alpha = 5.94$ . The quasiperturbative triples, (T), corrections are obtained from standard CCSD(T)/VTZ-F12 calculations (i.e., without inclusion of F12 terms) and scaled by the factor  $f = 0.987 \times E_{\text{MP2-F12}}/E_{\text{MP2}}$ . This approach has been shown to accelerate the basis set convergence [31, 49]. The CCSD inner-shell contribution is calculated with the core-valence weighted correlation-consistent A'PWCVTZ basis set, while the (T) inner-shell contribution is calculated with the PWCVTZ(no f) basis set. The scalar relativistic, spin–orbit coupling, DBOC, and ZPVE corrections are obtained in the same way as in W1-F12 theory.

The total atomization energies at 0 K ( $\text{TAE}_0$ ) are converted to heats of formation at 298 K using the Active Thermochemical Tables (ATcT) [55–59] atomic heats of formation at 0 K (H  $51.633 \pm 0.000$ , C  $170.024 \pm 0.014$ , N  $112.469 \pm 0.007$ , O  $58.997 \pm 0.000$ , and S  $65.709 \pm 0.036$  kcal/mol), and the CODATA [60] enthalpy functions,  $H_{298} - H_0$ , for the elemental reference states ( $\text{H}_2(\text{g}) = 2.024 \pm 0.000$ , C(cr,graphite) =  $0.251 \pm 0.005$ ,  $\text{N}_2(\text{g}) = 2.072 \pm 0.000$ ,  $\text{O}_2(\text{g}) = 2.075 \pm 0.000$ , and S(cr,rhombic) =  $1.054 \pm 0.001$  kcal/mol), while the enthalpy functions for the amino acids are obtained within the rigid rotor harmonic oscillator (RRHO) approximation from B3LYP/A'VTZ geometries and harmonic frequencies.

W1-F12 shows excellent performance for systems containing first-row elements (and H). Specifically, for the 97 first-row atomization energies in the W4-11 dataset, [32] W1-F12 attains a root mean square deviation (RMSD) of 0.19 kcal/mol relative to all-electron, relativistic CCSD(T) reference atomization energies at the infinite basis set limit. However, for second-row systems, it was found that the performance of W1-F12 is significantly degraded owing to shortcomings of the cc-pVDZ and cc-pVDZ-F12 basis sets for second-row elements (see Ref. [31] for details): for the 40 second-row atomization energies in the W4-11 dataset, RMSD actually exceeds 1 kcal/mol. W2-F12 does not suffer from this problem and yields similar RMSDs of 0.18 kcal/mol for first-row and 0.24 kcal/mol for second-row systems. (For further details, see reference [31]). Thus, for the sulfur-containing amino acids (cysteine and methionine) and for the small amino acids (alanine, glycine, and serine), the heats of formation are also obtained using W2-F12 theory.

The case of glycine is small enough (especially considering the  $C_s$  symmetry) that the result can be independently verified using accurate thermochemical procedures based on layered extrapolation of orbital basis sets, specifically the high-accuracy W4 method [68]. Full details of the method are given in that reference and will not be

repeated here: suffice to say that for a set of molecules where accurate experimental atomization energies are available via ATcT, the RMSD from experiment is 0.10 kcal/mol [32, 68]. The largest-scale calculation involved here, CCSD/aug'-cc-pV6Z, entails 1400 basis functions and required 3 terabyte of scratch space, yet ran to completion within a day on a machine with a large solid-state disk array. The CCSD(T)/aug'-cc-pV5Z calculation, involving 910 basis functions, ran in under a day on 32 cores and 512 GB of RAM.

The heats of formation have also been obtained using computationally more economical composite procedures, namely the Gaussian-4 (G4) protocol [33, 63] and its computationally more economical G4(MP2) and G4(MP2)-6X variants [64, 65]. These calculations were performed using the Gaussian 09 program suite [40]. The G4 and G4(MP2) protocols are widely used for the calculation of thermochemical properties and are applicable to relatively large systems (of up to 20–30 non-hydrogen atoms). They, generally, give RMSDs from experimental or high-accuracy theoretical thermochemical data of 1–2 kcal/mol [32, 63, 64]. For example, for the 454 experimental thermochemical determinations of the G3/05 test set (including heats of formation, ionization energies, and electron affinities), [66] G4 and G4(MP2) attain RMSDs of 1.2 and 1.5 kcal/mol, respectively [63, 64]. For the set of 137 very accurate theoretical atomization energies in the W4-11 set, both procedures attain an RMSD of 2.0 kcal/mol [32]. Finally, we have also considered the performance of the CBS-QB3 procedure [67] using Gaussian 09.

### 3 Results and discussion

#### 3.1 Computational cost of the W1-F12 calculations

For systems consisting of more than eight non-hydrogen atoms (with  $C_1$  symmetry), W1 theory [50] becomes prohibitively expensive with current commodity server hardware. W1-F12 theory is an explicitly correlated version of the W1 method, [50] which combines explicitly correlated F12 methods [1–4] with extrapolation techniques in order to approximate the CCSD(T)/CBS energy. Because of the drastically accelerated basis set convergence of the F12 methods [42, 43], W1-F12 is superior to the original W1 method, not only in terms of performance but also in terms of computational cost [31]. For example, the cpu times for calculating W1 and W1-F12 energies for a system containing 8 non-hydrogen atoms (with  $C_1$  symmetry) are 595 and 163 h, respectively (both calculations ran on 8 Intel Xeon Sandy Bridge cores at 2.6 GHz). In terms of disk space requirements, the W1 calculation used about five times the amount of scratch disk (660 GB) that the W1-F12 calculation required (126 GB).

In the present work, we obtain W1-F12 energies for the 18 amino acids with up to 12 non-hydrogen atoms. Of these, the largest amino acids are glutamic acid, glutamine, and lysine (10 non-hydrogen atoms); histidine (11 non-hydrogen atoms); arginine and phenylalanine (12 non-hydrogen atoms). Considering the fact that none of the amino acids (apart from glycine) have any spatial symmetry, these represent the largest W1-F12 calculations reported to date. For example, the W1-F12 calculation for arginine ran for 51 days on 6 Intel Nehalem 8837 cores at 2.67 GHz and used 253 GB of RAM and 1.1 TB of scratch disk. Due to this very steep computational cost, we obtain our best heats of formation for the two amino acids with more than 12 non-hydrogen atoms (i.e., tryptophan and tyrosine) with the *Gn* and CBS-QB3 methods [33, 67], which have a significantly reduced computational cost. In Sect. 3.4, we show that, relative to W1-F12 and W2-F12 heats of formation, G4, G4(MP2)-6X, and CBS-QB3 result in RMSDs of 0.72, 0.71, and 1.01 kcal/mol, respectively, i.e., near or below the threshold of “chemical accuracy” (traditionally arbitrarily defined as 1 kcal/mol).

### 3.2 W1-F12 and W2-F12 benchmark heats of formation

Since W1-F12 and W2-F12 theories represent a layered extrapolations to the CCSD(T) basis set limit energy, it is of interest to estimate whether the contributions from post-CCSD(T) excitations are likely to be significant for the atomization energies of the amino acids. The percentage of the total atomization energy accounted for by parenthetical connected triple excitations,  $\%TAE_e[(T)]$ , has been shown to be a reliable energy-based diagnostic for the importance of non-dynamical correlation effects [68, 74]. It has been suggested that  $\%TAE_e[(T)] < 2\%$  indicates systems that are dominated by dynamical correlation, while  $2\% < \%TAE_e[(T)] < 5\%$  indicates systems that include mild non-dynamical correlation.  $\%TAE_e[(T)]$  values for the amino acids are gathered in Table 1. The amino acids are characterized by  $\%TAE_e[(T)]$  values ranging from 1.7 (leucine) to 2.5 % (histidine). Note also that in all cases, the SCF component accounts for 69–77 % of the total atomization energy. These values suggest that our all-electron, non-relativistic, vibrationless benchmark atomization energies should, in principle, be considerably closer than 1 kcal/mol of the atomization energies at the full configuration interaction (FCI) basis set limit. For example, for systems that are associated with similar  $\%TAE_e[(T)]$  values in the W4-11 dataset [32], post-CCSD(T) contributions to the atomization energy are 0.2 kcal/mol or less, although somewhat larger values were found for benzene [75, 76].

Table 2 gives an overview of basis set convergence of the CCSD-F12 component of the total atomization energy.

**Table 1** Diagnostics indicating the importance of post-CCSD(T) contributions for the amino acids

Isomer	$\%TAE_e[SCF]^a$	$\%TAE_e[(T)]^b$
Alanine	71.4	1.99
Arginine	70.1	2.12
Asparagine	69.4	2.34
Aspartic acid	69.7	2.37
Cysteine	70.1	2.23
Glutamine	70.5	2.19
Glutamic acid	70.9	2.20
Glycine	69.6	2.17
Histidine	69.6	2.45
Isoleucine	73.7	1.73
Leucine	73.7	1.72
Lysine	72.3	1.83
Methionine	72.4	1.94
Phenylalanine	73.4	2.17
Proline	72.4	1.98
Serine	69.8	2.18
Threonine	71.0	2.05
Tryptophan <sup>c</sup>	76.6	1.85
Tyrosine <sup>c</sup>	76.5	1.76
Valine	73.1	1.79

Obtained from W1-F12 theory, unless otherwise indicated

<sup>a</sup> Percentages of the valence CCSD(T)/CBS atomization energy accounted for by the SCF component

<sup>b</sup> Percentages of the valence CCSD(T)/CBS atomization energy accounted for by the (T) component

<sup>c</sup> Obtained at the CCSD(T)/cc-pVDZ level of theory, see Ref. [32] for further details

The magnitude of the valence CCSD-F12 correlation component spans a relatively large range. For example, the CCSD-F12/V{D,T}Z-F12 results extrapolated with  $\alpha = 3.67$  (which was optimized to minimize the RMSD over 137 first- and second-row systems in the W4-11 dataset [31]) extend from 272.48 (glycine) up to 701.76 (arginine) kcal/mol. The differences between the CCSD-F12/V{D,T}Z-F12 results obtained with  $\alpha = 3.67$  (optimized over the entire W4-11 set of small molecules) and  $\alpha = 3.38$  (optimized over the subset of 97 first-row species only) can get quite significant for these medium-sized species, ranging from 0.25 kcal/mol for glycine to 0.71 kcal/mol for arginine. Note that these differences still only correspond to about 0.1 % of the valence CCSD correlation component. For comparison, for the systems in the W4-11 dataset, the absolute differences between the CCSD-F12/V{D,T}Z-F12 component extrapolated with  $\alpha = 3.67$  and 3.38 are reduced to just 0.00–0.22 kcal/mol, or 0.08 kcal/mol mean absolute—likewise, about 0.1 % of the valence CCSD correlation component of the atomization energy. Finally, using instead the extrapolation

**Table 2** Overview of the basis set convergence of the CCSD-F12 component of the total atomization energies for the amino acids (kcal/mol)

	VDZ-F12	VTZ-F12	VQZ-F12	V{D,T}Z-F12				V{T,Q}Z-F12	
				3.67 <sup>a</sup>	3.38 <sup>b</sup>	3.144 <sup>c</sup>	Best <sup>d</sup>	5.94 <sup>a</sup>	4.596 <sup>c</sup>
Alanine	327.84	334.23	335.98	336.10	336.41	336.71	336.56 ± 0.30	336.36	336.61
Arginine	683.07	697.54		701.76	702.46	703.15	702.80 ± 0.68		
Asparagine	472.43	481.98		484.76	485.23	485.68	485.46 ± 0.45		
Aspartic acid	448.56	457.12		459.61	460.03	460.44	460.23 ± 0.41		
Cysteine	357.89	363.68	365.51	365.38	365.66	365.93	365.80 ± 0.27	365.91	366.17
Glutamine	534.06	544.83		547.97	548.50	549.01	548.75 ± 0.51		
Glutamic acid	509.77	519.52		522.36	522.84	523.30	523.07 ± 0.46		
Glycine	265.77	270.96	272.46	272.48	272.73	272.97	272.85 ± 0.25	272.80 <sup>e</sup>	273.00
Histidine	570.49	582.21		585.63	586.20	586.76	586.48 ± 0.55		
Isoleucine	514.53	524.49		527.40	527.88	528.35	528.12 ± 0.47		
Leucine	513.54	523.49		526.39	526.87	527.34	527.11 ± 0.47		
Lysine	581.62	593.25		596.65	597.21	597.76	597.49 ± 0.55		
Methionine	478.88	486.87	489.21	489.20	489.59	489.97	489.78 ± 0.38	489.72	490.06
Phenylalanine	596.04	607.63		611.01	611.58	612.13	611.85 ± 0.55		
Proline	428.78	437.25		439.72	440.14	440.54	440.34 ± 0.40		
Serine	372.23	379.39	381.45	381.48	381.83	382.17	382.00 ± 0.34	381.90	382.20
Threonine	436.04	444.43		446.87	447.28	447.68	447.48 ± 0.40		
Valine	452.54	461.33		463.89	464.32	464.73	464.53 ± 0.42		
RMSD <sup>f</sup>				0.43	0.14	0.24	0.12		
MAD <sup>f</sup>				0.41	0.11	0.21	0.11		
MSD <sup>f</sup>				-0.41	-0.10	0.21	0.06		

<sup>a</sup> Extrapolation exponent ( $\alpha$ ) from Ref. [31], optimized to minimize the RMSD over the entire W4-11 dataset

<sup>b</sup> Extrapolation exponent ( $\alpha$ ) from Ref. [31], optimized to minimize the RMSD over the first-row systems in the W4-11 dataset

<sup>c</sup> Extrapolation exponent ( $\alpha$ ) from Ref. [43]

<sup>d</sup> The values are the average between the CCSD-F12 components extrapolated with  $\alpha = 3.38$  and 3.144, the difference between these two CCSD-F12 components may be regarded as a conservative error bar

<sup>e</sup> Extrapolating from the A'VQZ and A'V5Z basis sets with  $\alpha = 3.0$  results in a CCSD-F12 component of 272.94 kcal/mol

<sup>f</sup> Root mean square deviation (RMSD), mean absolute deviation (MAD), and mean signed deviation (MSD) relative to the CCSD-F12/V{T,Q}Z-F12 results for 5 systems

exponent optimized by Hill et al. [43] ( $\alpha = 3.144$ ), which was optimized over a smaller set of 14 absolute correlation energies, results in atomization energies increased by 0.24 (glycine) up to 0.69 (arginine) kcal/mol over the values with  $\alpha = 3.38$  (Table 2).

For five smaller amino acids (alanine, cystine, glycine, methionine, and serine), we were able to obtain CCSD-F12/VQZ-F12 energies. Table 2 gives the CCSD-F12/V{T,Q}Z-F12 results extrapolated with  $\alpha = 5.94$  (used in W2-F12 theory [31]) and 4.596 (from Ref. [43]). For these systems, the difference between the CCSD-F12/V{T,Q}Z-F12 contributions extrapolated with  $\alpha = 5.94$  and 4.596 ranges between 0.20 (glycine) and 0.34 (methionine) (Table 2). We note that the error statistics over the 137 systems in the W4-11 dataset are as follows: RMSD = 0.13, MAD = 0.10, and MSD = 0.01 for  $\alpha = 5.94$ , and RMSD = 0.15, MAD = 0.11, and MSD = 0.08 kcal/mol for  $\alpha = 4.596$ . Peterson and Feller [44] obtained benchmarks extrapolated

from basis sets as large as aug-cc-pV8Z for a fairly large sample of molecules that overlaps W4-11 and found that CCSD-F12b/V{T,Q}Z-F12 tends to overestimate the valence CCSD component on average: as they were using  $\alpha = 4.596$ , this is consistent with the present finding. (They also report difficulties reaching 0.1 kcal/mol convergence for CCSD-F12b energies with aug-cc-pV5Z basis sets: We were only able to apply this basis set to glycine, and in any case 0.1 kcal/mol is smaller than other potential error sources in the present work).

For the five W2-F12 amino acids, the RMSDs for CCSD-F12/V{D,T}Z-F12 with various choices of extrapolation exponent are 0.43 ( $\alpha = 3.67$ ), 0.14 ( $\alpha = 3.38$ ), and 0.24 ( $\alpha = 3.144$ ) kcal/mol. Taking the average between the CCSD-F12/V{D,T}Z-F12 components extrapolated with  $\alpha = 3.38$  and 3.144 results in an RMSD of 0.12 kcal/mol and a mean signed deviation of only +0.06 kcal/mol. We thus use this averaged CCSD-F12/V{D,T}Z-F12

component in our final W1-F12 atomization energies. The spread between the  $\alpha = 3.38$  and 3.144 values can be considered a crude gauge of the uncertainty in the basis set limit.

The component breakdowns of the W1-F12 and W2-F12 atomization energies are gathered in Table 3. The following general observations may be noted:

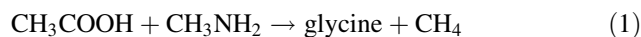
- As pointed out above, the magnitude of the valence CCSD-F12 correlation component runs a large gamut, extending from 272.85 (glycine) up to 702.80 (arginine) kcal/mol.
- The magnitude of the valence (T) correlation component can be rather large, reaching 54.28 kcal/mol for phenylalanine.
- The core–valence contribution approaches or exceeds 10 kcal/mol for the largest systems. Namely, it is 9.88 (arginine) and 11.68 (phenylalanine) kcal/mol.
- The DBOC contribution ranges from 0.28 (glycine) up to as much as 0.72 (arginine) kcal/mol.

Comparison of the W1-F12 and W2-F12 results for alanine, cystine, glycine, methionine, and serine reveals the following:

- The HF/V{D,T}Z-F12 component systematically underestimates the HF/VQZ-F12 basis set limit, namely by 0.03 (glycine), 0.04 (alanine and cysteine), 0.05 (serine), and 0.08 (methionine) kcal/mol.
- Our best CCSD-F12/V{D,T}Z-F12 component overestimates the CCSD-F12/V{T,Q}Z-F12 component by 0.05 (glycine), 0.06 (methionine), 0.10 (serine), 0.20 (alanine) kcal/mol, and underestimates it by 0.11 kcal/mol for cysteine.
- The valence (T) contribution from W1-F12 theory systematically overestimates the W2-F12 results, specifically by 0.06 (cysteine), 0.13 (methionine), 0.17 (glycine), 0.20 (alanine), and 0.25 (serine) kcal/mol.
- The core–valence contribution from W1-F12 systematically underestimates the W2-F12 result, namely by 0.09 (glycine), 0.12 (alanine), 0.14 (serine), and 0.16 (cysteine) kcal/mol (we were not able to obtain the core–valence contribution for methionine from W2-F12 theory).
- Overall, the  $TAE_e$  from W1-F12 theory overestimates the  $TAE_e$  from W2-F12 theory by 0.11 (glycine and methionine), 0.16 (serine), and 0.23 (alanine) kcal/mol, and underestimates it by 0.30 kcal/mol for cysteine.

As noted in the “Methods” section, we were able to “cross-check” the result for glycine at the W4 level: the lower-cost W2.2 level is obtained as a by-product. As seen in Table 3, the SCF, CCSD, (T), core-valence, and relativistic components of the W2-F12 calculation are all in excellent agreement with the W4 calculation, the

cumulative difference being just 0.04 kcal/mol. The higher-order correlation steps, CCSDT(Q)/cc-pVTZ, and CCSDTQ/cc-pVDZ are more problematic from a computational point of view, but their importance is typically quite small for molecules dominated by a single reference configuration (due to error compensation between “anti-bonding” higher-order  $T_3$  and “bonding”  $T_4$  contributions [68–73]). Absent a direct calculation, their importance can be estimated by assuming that their contribution to the following isodesmic reaction energy will be approximately zero:



From Table SI-II of Ref. [32], we find the post-CCSD(T) contributions to the TAEs to be  $-0.05$  kcal/mol for acetic acid,  $-0.09$  kcal/mol for methyl amine, and  $+0.01$  kcal/mol for methane, leading to an estimated post-CCSD(T) correction of  $-0.15$  kcal/mol for glycine.

### 3.3 A note on zero-point vibrational energies (ZPVEs)

In view of the magnitude of the zero-point vibrational energies (50–140 kcal/mol, see Table 4), some remarks are due concerning their calculation. Ideally, one should obtain them from accurate anharmonic force fields, and for small molecules, this is indeed a practical option [68, 85, 91]. In the present case, however, the computational cost would be prohibitive with the computational resources at hand, and multiplication of calculated harmonic frequencies with a scaling factor  $\lambda(\text{ZPVE})$  appropriate for zero-point vibrational energies [50, 83, 84, 86, 90] is the only practical option. As shown in Ref. [84], ZPVEs are typically almost exact averages of one-half the sum of the harmonics and one-half the sum of the fundamentals, the difference being just  $\text{ZPVE} - (1/4) \sum_i \omega_i + \nu_i = G_0 - \sum_i X_{ii}/4$ , where the  $X_{ii}$  are the diagonal anharmonicity constants and  $G_0$  is the polyatomic counterpart of the small  $Y_{00}$  Dunham constant [82] in diatomics. Consequently [50, 84, 90], the optimal scaling factor for ZPVEs is almost exactly midway between a  $\lambda(\omega)$  suitable for harmonic frequencies (as an approximate correction for systematic bias in the calculated frequencies) and a  $\lambda(\nu)$  suitable for fundamental frequencies (which additionally seeks to approximately corrects for anharmonicity). In fact, Alecu et al. [86] found for a large variety of basis sets and ab initio and DFT methods that  $\lambda(\omega)/\lambda(\text{ZPVE}) = 1.014 \pm 0.002$ , which is almost exactly the ratio of 1.0143 found by Perdew and coworkers [87] between harmonic frequencies and ZPVEs derived from experimental anharmonic force fields. Note that the “small” uncertainty of 0.002 on a ZPVE of 140 kcal/mol still would translate to about 0.3 kcal/mol, and even that is probably optimistic for the uncertainty in an individual

**Table 3** Component breakdown of the W1-F12 and W2-F12 atomization energies and final gas-phase heats of formation at 0 and 298 K for the lowest-energy conformers of the amino acids (kcal/mol)

	SCF	CCSD	(T)	CV	Rel.	DBOC	SO	ZPVE	TAE <sub>e</sub>	TAE <sub>0</sub>	$\Delta_f H_{0K}^{\circ}$	hc <sub>f,298</sub> <sup>RRHOH</sup>	H <sub>conf,298</sub>	$\Delta_f H_{298K}^{\circ}$
Alanine	900.58	336.56	25.06	4.51	-1.25	0.34	-0.70	66.93	1265.10	1198.17	-96.20	4.97	0.45 <sup>d</sup>	-101.73
Ditto W2-F12	900.62	336.36	24.86	4.63	-1.25	0.34	-0.70	66.93	1264.87	1197.94	-95.97	4.97	0.45 <sup>d</sup>	-101.49
Arginine	1771.89	702.80	53.45	9.88	-2.61	0.72	-0.95	138.18	2535.20	2397.02	-86.14	8.92	0.33 <sup>e</sup>	-98.78
Asparagine	1189.40	485.46	40.11	6.6	-1.91	0.52	-1.01	84.81	1719.16	1634.35	-139.26	6.39	0.61 <sup>f</sup>	-146.54
Aspartic acid	1147.26	460.23	38.93	6.12	-1.85	0.46	-1.23	76.55	1649.91	1573.37	-183.38	6.47	0.61 <sup>g</sup>	-189.57
Cysteine	926.11	365.80	29.45	4.79	-1.67	0.37	-1.26	67.37	1323.58	1256.21	-88.53	5.52	0.81 <sup>h</sup>	-94.21
Ditto W2-F12	926.15	365.91	29.39	4.95	-1.65	0.37	-1.26	67.37	1323.88	1256.50	-88.83	5.52	0.81 <sup>h</sup>	-94.50
Glutamine	1415.13	548.75	43.96	7.77	-2.11	0.58	-1.09	102.14	2012.99	1910.85	-142.47	7.40	0.64 <sup>i</sup>	-150.99
Glutamic acid	1375.14	523.07	42.74	7.29	-2.04	0.52	-1.32	94.19	1945.40	1851.22	-187.94	7.32	0.68 <sup>j</sup>	-195.50
Glycine	672.51	272.85	20.98	3.36	-1.06	0.28	-0.62	49.44	968.31	918.87	-90.19	4.16	0.49 <sup>k</sup>	-94.21
Ditto W2-F12	672.54	272.80	20.81	3.45	-1.07	0.28	-0.62	49.44	968.20	918.76	-90.08	4.16	0.49 <sup>k</sup>	-94.10
Ditto W2.2	672.50	272.72	20.92	3.43	-1.06	0.28	-0.62	49.44	968.18	918.74	-90.06	4.16	0.49 <sup>k</sup>	-94.08
Ditto W4 <sup>b</sup>	672.53	272.80	20.85	3.43	-1.06	0.28	-0.62	[49.438] <sup>a</sup>	968.07 <sup>b</sup>	918.63	-89.95	4.16	0.49 <sup>k</sup>	-93.97
Histidine	1456.87	586.48	51.29	9.33	-2.24	0.56	-0.95	100.02	2101.33	2001.31	-61.06	6.80	0.22 <sup>l</sup>	-69.84
Isoleucine	1579.21	528.12	37.11	8.00	-1.83	0.52	-0.95	119.69	2150.17	2030.48	-108.64	7.51	0.59 <sup>m</sup>	-118.31
Leucine	1580.74	527.11	36.91	8.00	-1.82	0.52	-0.95	119.68	2150.49	2030.81	-108.97	7.50	0.48 <sup>m</sup>	-118.76
Lysine	1670.47	597.49	42.32	8.54	-2.10	0.60	-0.95	132.05	2316.37	2184.32	-98.38	7.65	0.60 <sup>m</sup>	-109.95
Methionine	1380.29	489.78	36.92	7.08	-2.06	0.47	-1.43	102.81	1911.06	1808.24	-93.98	7.66	0.37 <sup>o</sup>	-102.50
Ditto W2-F12	1380.37	489.72	36.79	7.08 <sup>c</sup>	-2.06 <sup>c</sup>	0.47	-1.43	102.81	1910.95	1808.14	-93.88	7.66	0.37 <sup>o</sup>	-102.40

Table 3 continued

	SCF	CCSD	(T)	CV	Rel.	DBOC	SO	ZPVE	TAE <sub>e</sub>	TAE <sub>0</sub>	$\Delta_f H_{0K}^0$	hcf <sub>298</sub> <sup>RRHO</sup> H	H <sub>conf,298</sub>	$\Delta_f H_{298K}^0$
Phenylalanine	1839.81	611.85	54.28	11.68	-2.25	0.56	-1.21	117.86	2514.73	2396.87	-68.22	7.32	0.47 <sup>p</sup>	-76.93
	SCF	CCSD	(T)	CV	Rel.	DBOC	SO	ZPVE	TAE <sub>e</sub>	TAE <sub>0</sub>	$\Delta_f H_{0K}^0$	hcf <sub>298</sub> <sup>RRHO</sup>	H <sub>conf,298</sub>	$\Delta_f H_{298K}^0$
Proline	1241.14	440.34	33.97	6.75	-1.63	0.40	-0.87	89.92	1720.11	1630.19	-84.91	5.20	0.43 <sup>q</sup>	-92.75
Serine	950.49	382.00	29.62	4.73	-1.52	0.40	-0.92	70.49	1364.80	1294.31	-133.34	5.45	0.56 <sup>r</sup>	-139.32
Ditro W2-F12	950.54	381.90	29.37	4.87	-1.52	0.40	-0.92	70.49	1364.64	1294.15	-133.19	5.45	0.56 <sup>r</sup>	-139.16
Threonine	1178.57	447.48	34.03	5.93	-1.71	0.48	-1.01	87.98	1663.77	1575.79	-141.53	6.13	0.71 <sup>s</sup>	-148.95
Valine	1353.27	464.53	33.16	6.84	-1.63	0.46	-0.87	102.04	1855.76	1753.72	-105.17	6.64	0.42 <sup>m</sup>	-113.60

hcf stands for the heat content function (a.k.a., enthalpy function),  $H_T - H_0$

<sup>a</sup> anharmonic CCSD(T)/CBS ZPVE, Ref. [81] <sup>b</sup> values include -0.15 kcal/mol isodesmic correction for higher-order correlation effects (see text)

<sup>c</sup> Taken from W1-F12 theory

<sup>d</sup> CCSD(T)/CBS including zero-point corrections for 12 conformers from Ref. [24]

<sup>e</sup> Ref. [100]. Lowest 12 conformers, CCSD/6-31++G(d,p)

<sup>f</sup> Ref. [101]. B3LYP/6-311++G(d,p) relative energies and zero-point corrections for 62 lowest conformers taken from Table 1

<sup>g</sup> Ref. [102]. CCSD/6-311++G(d,p) total energies for 15 lowest conformers with B3LYP/6-311++G(d,p) zero-point energies. To be regarded as a lower bound. If MP2 is substituted for CCSD, drops to 0.55 kcal/mol

<sup>h</sup> MP2/cc-pVTZ structures for 71 lowest conformers taken from Ref. [23]. CCSD(T\*)-F12b/cc-pVDZ-F12 relative energies, and MP2/cc-pVTZ zero-point corrections, calculated in present work

<sup>i</sup> Ref. [104]. Lowest 21 conformers, MP2/6-311G(2df,2pd)//B3LYP/6-311++G(d,p) with zero-point energy from BHandHLYP/6-311++G(d,p)

<sup>j</sup> Ref. [103]. MP2/6-311++G(2df,p)//B3LYP/6-311++G(d,p) relative energies at the bottom of the well for the lowest 35 conformers, combined with ZPVEs at the B3LYP/6-311++G(d,p) level, both taken from Table 1

<sup>k</sup> Ref. [81]. Best CCSD(T)/CBS values including ZPVE taken for all 6 conformers

<sup>l</sup> Ref. [105]. Lowest 42 conformers, MP2/6-311++G(d,p) relative energies with B3LYP/6-311G(d) zero-point corrections

<sup>m</sup> Ref. [106]. relative energies at 0 K taken from Table 1. MP2/6-311++G(2d,2p) with scaled MP2(full)/6-31G\* ZPVEs. As only lowest 6, 8, and 12 conformers are included for VAL, LEU, and ILE, respectively, these values are to be regarded as lower limits

<sup>n</sup> Ref. [107]. Relative energies at 0 K taken from Table 1. MP2/6-31++G\*\*//B3LYP/6-31++G\*\* relative energies for 18 lowest conformers

<sup>o</sup> From approximate CCSD(T)/6-311++G(3df,2p) conformer energies in Table 7 of Ref. [21]

<sup>p</sup> Ref. [108]. Lowest 37 conformers at MP2/6-311++G(d,p) level combined with B3LYP/6-311G(d) zero-point energies. Substituting B3LYP/6-311++G(d,p) reduces value to 0.31 kcal/mol

<sup>q</sup> Ref. [109]. Focal point analysis for lowest 4 conformers. To be regarded as lower limit

<sup>r</sup> Present work. MP2/cc-pVTZ geometries from cysteine conformer set used as starting structures. For the 58 unique minima obtained, CCSD(T\*)-F12b/cc-pVDZ-F12 relative energies and MP2/cc-pVTZ zero-point corrections

<sup>s</sup> Ref. [110]. Focal point analysis on lowest 5 conformers. To be regarded as lower limit on account of the small number of conformers



**Table 4** Dependence of computed ZPVEs (kcal/mol) on the level of theory

	HF/ 6-31G(d) G3,G3MP2	B3LYP/ 6-31G(d) G3B3, G3MP2B3	B3LYP/ 6-31G(2df,p) G4,G4MP2	B3LYP/ 6-311G (2d,d,p) CBS-QB3	B3LYP/ cc- pV(T+d)Z W1	B3LYP/ aug'-cc- pV(T+d)Z	B3LYP/ def2- TZVPP	B2PLYP/ def2- TZVPP
Scaling factor <sup>a</sup>	0.8929 <sup>b</sup>	0.9600 <sup>c</sup>	0.9854 <sup>d</sup>	0.9900 <sup>e</sup>	0.9850 <sup>f</sup>	0.9896 <sup>g</sup>	0.9884 <sup>g</sup>	0.9833 <sup>g</sup>
HFREQ27 RMSD (cm <sup>-1</sup> ) w/o F <sub>2</sub>	88.1	74.2	39.1	34.0	34.6	30.9	31.9	16.2
Alanine	65.49	65.33	66.67	66.95	66.52	66.73	66.82	66.93
Arginine	135.38	134.90	137.67	138.21	137.38	137.92	137.99	138.18
Asparagine	83.29	82.91	84.67	85.04	84.43	84.61	84.77	84.81
Aspartic	75.42	74.74	76.40	76.78	76.19	76.41	76.51	76.55
Cysteine	66.21	65.78	67.07	67.43	66.97	67.16	67.25	67.37
Glutamic	92.52	91.94	93.99	94.41	93.75	94.01	94.14	94.20
Glutamine	100.08	99.83	101.85	102.35	101.63	101.99	102.07	102.13
Glycine	48.56	48.25	49.27	49.51	49.15	49.29	49.35	49.44
Histidine	98.15	97.65	99.83	100.23	99.61	99.97	100.07	100.02
Isoleucine	116.70	116.95	119.27	119.67	118.97	119.38	119.51	119.69
Leucine	116.63	116.91	119.14	119.60	118.90	119.30	119.45	119.68
Lysine	128.89	129.00	131.51	132.13	131.20	131.72	131.76	132.05
Methionine	100.66	100.42	102.35	102.79	102.12	102.54	102.59	102.81
Phenylalanine	115.16	115.18	117.60	118.10	117.47	117.94	117.97	117.86
Proline	87.94	87.91	89.70	90.09	89.54	89.86	89.95	89.92
Serine	69.18	68.74	70.29	70.65	70.07	70.22	70.37	70.49
Threonine	86.26	85.97	87.75	88.09	87.41	87.71	87.81	87.98
Valine	99.57	99.69	101.69	102.04	101.42	101.75	101.86	102.04

<sup>a</sup> Scaling factor  $\lambda(\text{ZPVE})$  appropriate for ZPVEs. All scaling factors for harmonic frequencies  $\lambda(\omega) = 1.014\lambda(\text{ZPVE})$

<sup>b</sup> As specified in Ref. [61]

<sup>c</sup> As specified in Ref. [62]

<sup>d</sup> As specified in Ref. [63, 64]

<sup>e</sup> As specified in Ref. [67]

<sup>f</sup> As specified in Ref. [50]. Sometimes in molecules with ionic bonding character, the aug'-cc-pV(T+d)Z basis set is employed with the same scaling factor: RMSD(HFREQ27) for that is 32.5 cm<sup>-1</sup> including, and 29.9 cm<sup>-1</sup> excluding F<sub>2</sub>

<sup>g</sup> Obtained in this work by minimizing the RMSD over the HFREQ27 [95] dataset. For B3LYP/cc-pV(T+d)Z, the same procedure leads  $\lambda(\text{ZPVE}) = 0.9892$ . For B2PLYP near the basis set limit, Ref. [95] has  $\lambda(\omega) = 0.997$ , which corresponds to  $\lambda(\text{ZPVE}) = 0.983$

molecule [88]. It has been argued earlier [91] (see also Ref. [92]) that for organic and bio-organic molecules that are “well-behaved” from an electronic structure point of view, the main factor limiting accuracy in computational thermochemistry may well be the treatment of the nuclear motion, rather than the electronic problem as such.

Computed zero-point vibrational energies for the amino acids at various levels of theory (including those used in the composite thermochemistry schemes compared in this work) are listed in Table 4. In search of an alternative that was more accurate than B3LYP yet still comparatively affordable, we considered the B2PLYP double hybrid functional [93] in conjunction with the def2-TZVPP basis set [94] and optimized a  $\lambda(\omega)$  scaling factor by minimizing the RMSD for the HFREQ27 dataset [95] of accurately

known harmonic frequencies. As can be seen in Table 4, the RMSD over the HFREQ27 set is only half that of B3LYP and drops to 13.2 cm<sup>-1</sup> if the anomalous F<sub>2</sub> molecule is eliminated. (For comparison, the HFREQ27 RMSD for CCSD(T)/cc-pV(Q+d)Z is still 8.4 cm<sup>-1</sup>.) The optimum scale factor  $\lambda(\omega) = 0.9971$  is very close to unity, and in conjunction with the “universal” ratio of 1.014 translated into  $\lambda(\text{ZPVE}) = 0.9833$ . As a sanity check on our procedure, we re-evaluated the  $\lambda(\text{ZPVE})$  for B3LYP/6-31G(2df,p) and B3LYP/6-311G(2d,d,p) and obtained 0.9858 and 0.9896, respectively, which agree to better than 3 decimal places with the “official” values used in G4 theory and CBS-QB3, respectively [63, 67].

It can be seen in Table 4 that the lower levels of theory used for ZPVEs in G3(MP2) [61] and G3(MP2)B3, [62] can

yield values several kcal/mol lower than the highest-level method: the RMSD from B2PLYP/def2-TZVPP are 2.12 and 2.29 kcal/mol, respectively, compared to 0.33 and 0.14 kcal/mol, respectively, for B3LYP/6-31G(2df,p) (scaled by 0.9854) as used by the G4 variants, and B3LYP/6-311G(2d,d,p) as used by CBS-QB3. (The “2d” refers to the use of an extra d function on second-row elements.) But also B3LYP/cc-pV(T+d)Z scaled by 0.985, as used in W1- and W1-F12 theory, appears to yield values that are too low, and indeed  $\lambda(\text{ZPVE})$  as obtained from the HFREQ27 set is 0.9892. For B3LYP with a basis set that is effectively at the Kohn-Sham limit,  $\lambda(\text{ZPVE}) = 1.004$  was found, which corresponds to  $\lambda(\text{ZPVE}) = 0.99$ , and the database of Radom and coworkers [90] likewise lists scaling factors near 0.99 for B3LYP with large basis sets. While a scaling factor of 0.985 vs. 0.990 may rightly be considered a distinction without a difference for small molecules (where anybody concerned about 0.1 kcal/mol in a ZPVE should seriously consider an accurate anharmonic ZPVE), the problem is much more obvious in larger systems such as presently considered.

For one system, glycine, an anharmonic value of 49.438 kcal/mol is available due to Puzzarini and coworkers [81], who combined CCSD(T)/CBS harmonic frequencies with a DFT anharmonic force field. Fortuitously, our scaled B2PLYP/def2-TZVPP value agrees to two decimal places. As an additional observation, for ethane, the accurate anharmonic ZPVE is 46.29 kcal/mol, [91] compared to 45.97 kcal/mol B3LYP/cc-pVTZ scaled by 0.985, 46.20 with a revised scaling factor of 0.99, and 46.33 kcal/mol at the B2PLYP/def2-TZVPP level scaled by 0.9833.

### 3.4 Performance of $G_n$ methods for the heats of formation of the amino acids

In this Section we use our best heats of formation from W1-F12 and W2-F12 theories (given in Table 3) to evaluate the performance of a variety of composite thermochemical Gaussian- $n$  ( $G_n$ ) procedures including G3(MP2), [61] G3(MP2)B3, [62], G4, [63] G4(MP2), [64] and G4(MP2)-6X [65]. Table 5 presents the deviations ( $G_n$ - $W_n$ -F12) from our benchmark  $W_n$ -F12 results, as well as the RMSD, mean absolute deviations (MAD), and mean signed deviations (MSD) for the  $G_n$  methods. Stover et al. [17] obtained G3(MP2) heats of formation for the amino acids: except for phenylalanine, cysteine, and methionine, the deviations between their heats of formation and our reference values exceed 1 kcal/mol. The mean signed deviation (MSD) of 1.90 kcal/mol being nearly equal to the RMSD of 2.25 kcal/mol indicates a very systematic error. Simply switching to G3(MP2)B3 cuts the MSD to 0.78 kcal/mol and the RMSD to 1.13 kcal/mol, while “upgrading” to G3B3 lowers these numbers even further to 0.45 and 0.60 kcal/mol, respectively. While both methods use MP2 rather than B3LYP reference

geometries, the entire G3 family suffers from underestimated ZPVEs for the amino acids (Table 4), so apparently some of that issue is absorbed by the empirical correction. Stover et al. [17] also obtained G3(MP2) heats of formation via isodesmic bond separation reactions. As expected this improves the performance, with RMSD = 1.48 kcal/mol and a maximum deviation of 2.40 kcal/mol for phenylalanine. We note, however, that their CCSD(T)/CBS anchor value for the heat of formation at room temperature of glycine,  $-92.6$  kcal/mol, is 1.5 kcal/mol lower than our W2-F12 value. If we substitute the latter in their isodesmic reactions, their RMSD plunges to just 0.47 kcal/mol.

We now turn our attention to the performance of the Gaussian-4 family: G4, [63] G4(MP2), [64] and G4(MP2)-6X [65]. The G4(MP2) procedure exhibits somewhat disappointing performance, its RMSD = 1.80 kcal/mol placing intermediately between G3(MP2) and G3(MP2)B3. The largest deviations are obtained for asparagine (2.48), lysine (2.32), glutamine (3.15), and arginine (3.34 kcal/mol), but all other deviations exceed 1 kcal/mol apart from phenylalanine, cysteine, and methionine. The computationally more expensive “full” G4 procedure yields much better performance with an RMSD of 0.72 kcal/mol, and just three cases exceeding 1 kcal/mol (glutamine 1.39, arginine 1.21, and lysine 1.37 kcal/mol). However, an essentially identical RMSD = 0.71 kcal/mol is afforded by the G4(MP2)-6X procedure, which involves the same computational steps and cost as G4(MP2) but entails six additional empirical scaling factors. Deviations larger than 1 kcal/mol are obtained for just four systems, namely arginine (1.63), glutamine (1.63), asparagine (1.10), and methionine ( $-1.02$  kcal/mol). Finally, we note that the CBS-QB3 method clocks in at RMSD = 1.01 kcal/mol.

Very recently, Ramabhadran et al. [21] determined the enthalpies of formation of cysteine and methionine using their connectivity-based hierarchy (CBH- $n$ ) approach [77, 78]. From their Table 3, the best enthalpies of formation obtained for the lowest-energy conformer at the CBH-2 (isoatomic) rung using experimental heats of formation for the reference species and CCSD(T)/6-311++G(3df,2p) reaction energies are  $-96.1$  (cysteine) and  $-104.3$  (methionine) kcal/mol. From their Table 7, we calculate conformer corrections of  $+0.77$  kcal/mol for cysteine and  $+0.37$  kcal/mol for methionine: The latter we actually use in the present work, while the former is slightly less than our own calculation of 0.81 kcal/mol. According to their Table 9, the heats of formation after conformer correction are  $-95.3$  and  $-104.0$  kcal/mol (the latter value presumably after roundoff), both more exothermic than our W2-F12 values (Table 3) of  $-94.5$  and  $-102.4$  kcal/mol. We do note that some of the experimental data for reference species used in Ref. [21] carry non-trivial uncertainties, which could account for at least some of the discrepancy.

**Table 5** Performance of a selection of composite procedures of the *Gn* family for the calculation of heats of formation ( $\Delta_f H_{298K}^\circ$ , exclusive of conformer correction) of the 18 amino acids in Table 3

	CBS-QB3	G3(MP2) <sup>a</sup>	G3(MP2) <sup>b</sup>	G3(MP2)B3 <sup>c</sup>	G3B3 <sup>c</sup>	G4(MP2) <sup>c</sup>	G4(MP2)-6X <sup>c</sup>	G4 <sup>c</sup>
Alanine	-0.41	1.74	1.24	0.48	-0.44	1.04	-0.02	0.25
Arginine	0.99	N/A	4.61	2.74	0.32	3.34	1.63	1.21
Asparagine	-0.64	1.35	2.75	1.44	-0.55	2.48	1.10	0.67
Aspartic acid	-1.70	0.58	1.28	-0.01	-1.55	1.64	0.09	0.20
Cysteine	-0.75	1.71	0.91	0.05	0.03	0.39	0.15	0.47
Glutamine	0.20	N/A	3.11	2.04	0.06	3.15	1.63	1.39
Glutamic acid	-0.91	1.08	1.58	0.29	-1.17	1.92	0.19	0.54
Glycine	-0.46	N/A	2.69	0.76	-0.09	1.22	0.38	0.51
Histidine	0.24	N/A	4.26	1.27	0.06	1.60	0.44	0.29
Isoleucine	0.86	1.30	1.00	0.52	-0.43	1.21	-0.46	0.37
Leucine	1.02	1.54	1.14	0.69	-0.18	1.26	-0.41	0.49
Lysine	1.50	N/A	2.65	1.74	0.53	2.32	0.61	1.37
Methionine	-0.54	1.17	0.27	-0.37	-0.37	-0.26	-1.02	-0.23
Phenylalanine	2.23	2.40	0.30	-0.47	0.29	0.68	-0.13	0.85
Proline	0.87	0.68	2.08	1.18	0.13	1.94	0.46	0.95
Serine	-1.16	1.52	1.52	0.50	-0.75	1.47	0.11	0.39
Threonine	-0.81	1.76	1.76	0.74	-0.73	1.86	0.34	0.58
Valine	0.39	1.52	1.12	0.47	-0.49	1.14	-0.32	0.31
RMSD <sup>c</sup>	1.01	1.48	2.25	1.13	0.60	1.80	0.71	0.72
MAD <sup>c</sup>	0.87	1.41	1.90	0.88	0.45	1.61	0.53	0.62
MSD <sup>c</sup>	0.05	1.41	1.90	0.78	-0.30	1.58	0.27	0.59
LD <sup>c</sup>	2.23	2.40	4.61	2.74	-1.55	3.34	1.63	1.39
ND <sup>c</sup>	Phe	Phe	Arg	Arg	Asp	Arg	Arg, glu	Glu
	6	11 <sup>e</sup>	14	6	2	15	4	3

The values listed are deviations (*Gn*-*Wn*-F12) from our benchmark W1-F12 and W2-F12 heats of formation (kcal/mol)

<sup>a</sup> From reference [17], obtained from isodesmic reactions

<sup>b</sup> From reference [17], obtained from atomization reactions

<sup>c</sup> This work

<sup>d</sup> RMSD = root mean square deviation, MAD = mean absolute deviation, MSD = mean signed deviation, LD = largest deviation, ND = number of deviations (in absolute value) exceeding 1 kcal/mol

<sup>e</sup> Out of a total of 13 determinations

### 3.5 Comparison with experiment

Comparison with experiment obviously entails thermal corrections. The RRHO approximation will cause some errors, the largest of which will be neglect of the population of the various low-energy conformers. If we neglect the difference between the rovibrational partition functions of the different conformers, then the conformer contribution to the enthalpy function  $\text{hcf}_{298} \equiv H_{T=298} - E_0$  is easily found as [96]

$$\text{hcf}_{298}^{\text{conf}} = \frac{\sum_i x_i \exp(-x_i)}{\sum_i \exp(-x_i)} \quad \text{where } x_i \equiv \frac{E_i - E_0}{RT} \quad (2)$$

where the index *i* runs over the conformers. The effect of accounting for different rovibrational partition functions in the different conformers was considered in Ref. [96] for the alkane conformers and is negligible compared to other

potential error sources in the present calculation, such as the neglect of anharmonicity and the uncertainty in the basis set extrapolation. Conformer energies were gathered from published calculations in the literature [21–24, 81, 100–112]: these range from complete basis set CCSD(T) studies for glycine [81] and alanine [24] to relatively low-level MP2 or DFT calculations for some other species. Details are given in the footnotes to Table 3.

Table 6 lists the available experimental gas-phase heats of formation at 298 K ( $\Delta_f H_{298}^\circ$ ). Our W2-F12 value for alanine (-101.5 kcal/mol) is spot on the experimental value of Dorofeeva and Ryzhova [97] ( $-101.5 \pm 0.5$  kcal/mol) and still agrees to within mutual uncertainties with that of da Silva et al. [15] ( $-101.9 \pm 0.7$ ). However, the NIST chemistry WebBook [79] value ( $-99.1 \pm 1.0$  kcal/mol) is clearly incompatible with our calculations.

**Table 6** Experimental gas-phase heats of formation at 298 K for the amino acids (kcal/mol)

$\Delta_f H_{298K}^\circ$	Expt. <sup>a</sup>	Uncert.	Best theor. <sup>b</sup>
Alanine	−99.1	1.0	−101.5
	−101.5 <sup>c</sup>	0.5	
	−101.9 <sup>d</sup>	0.7	
	−101.3 <sup>e</sup>		
Cysteine	−91.4 <sup>f</sup>	0.4	−94.2
Glycine	−93.3	1.1	−94.1, −93.97 <sup>h</sup>
	−94.1 <sup>c</sup>	0.4	
	−93.7 <sup>e</sup>		
Methionine	−98.8	1.0	−102.4
	−102.8 <sup>g</sup>	2.4	
Phenylalanine	−74.8 <sup>d,e</sup>		−76.9
Proline	−87.5	1.0	−92.8
Valine	−108.8 <sup>e</sup>		−113.6

<sup>a</sup> Values are taken from the NIST chemistry WebBook, [79] unless otherwise indicated

<sup>b</sup> From Table 3

<sup>c</sup> From Dorofeeva and Ryzhova [97]

<sup>d</sup> From da Silva et al. [15]

<sup>e</sup> From CRC Handbook [80]

<sup>f</sup> From Roux et al. 2010 [19]

<sup>g</sup> From Roux et al. 2012 [18]

<sup>h</sup> Quasi-W4 value including anharmonic ZPVE and estimated higher-order correlation correction (see Table 3 and text)

<sup>i</sup> Note that in the CRC Tables, the value for L-alanine (111.4 kcal/mol) seems to have been transposed with the value for  $\beta$ -alanine (101.3 kcal/mol)

Our W2-F12 heat of formation for cysteine (−94.2 kcal/mol) suggests that the experimental value of Roux et al. [19] should be revised downward by about 2.8 kcal/mol; the recent study of Ramabhadran et al. [21] suggests even further downward revision (*vide supra*). As for glycine, the W2-F12 heat of formation (−94.1 kcal/mol from W2-F12, −94.0 from quasi-W4) and the available experimental values agree to within overlapping uncertainties. Specifically, our calculations are spot on the experimental value of Dorofeeva and Ryzhova [97] (−94.1 ± 0.4 kcal/mol), just slightly below the experimental value from the CRC Handbook (−93.7 kcal/mol), and in the upper end of the uncertainty band of the NIST WebBook value (−93.3 ± 1.1 kcal/mol). Our W2-F12 value for methionine (−102.4 kcal/mol) agrees well with the new measurement of Roux et al. [18] (−102.8 ± 2.4 kcal/mol), and both imply a downward revision of the NIST Chemistry Webbook value (−98.8 ± 1.0 kcal/mol) by about 3–4 kcal/mol. As for phenylalanine, our W1-F12 value (−76.9 kcal/mol) suggests that the experimental value from the CRC Handbook (−74.8 kcal/mol) should be revised downward by about 2 kcal/mol. The W1-F12 values for proline

(−92.8 kcal/mol) and valine (−113.6 kcal/mol) suggest that the experimental values should be revised downward by about 5 kcal/mol (Table 6).

For the two largest amino acids, tryptophan and tyrosine, we were unable to calculate W1-F12 atomization energies. At the G4, CBS-QB3, and G4(MP2)-6X levels, respectively, we obtain heats of formation at 0 K for tryptophan of −49.60, −47.87, and −48.77 kcal/mol, and for tyrosine of −109.12, −108.58, and −108.49 kcal/mol. At room temperature, the corresponding values are −59.98, −58.27, and −58.98 kcal/mol for tryptophan and −118.56, −118.03, and −117.78 kcal/mol for tyrosine. Averaging over all three levels of theory, and adding in conformer corrections for tryptophan of 0.71 kcal/mol [111] and for tyrosine of 0.65 kcal/mol, we finally obtain estimated heats of formation at 298 K of −58.37 kcal/mol for tryptophan, and of −117.47 kcal/mol for tyrosine.

## 4 Conclusions

We have obtained benchmark heats of formation at the CCSD(T)/CBS limit for the 20 natural amino acids. Our best heats of formation at 298 K ( $\Delta_f H_{f,298}^\circ$ ) are −101.5 (alanine), −98.8 (arginine), −146.5 (asparagine), −189.6 (aspartic acid), −94.5 (cysteine), −151.0 (glutamine), −195.5 (glutamic acid), −94.0 (glycine, quasi-W4) or −94.1 (glycine, W2-F12), −69.8 (histidine), −118.3 (isoleucine), −118.8 (leucine), −110.0 (lysine), −102.4 (methionine), −76.9 (phenylalanine), −92.8 (proline), −139.2 (serine), −149.0 (threonine), and −113.6 (valine) kcal/mol. These heats of formation are obtained at the W2-F12 level for alanine, cysteine, glycine, methionine, and serine, and at the W1-F12 level for all of the rest. For the two largest amino acids, an average over G4, G4(MP2)-6X, and CBS-QB3 yields best estimates of −58.4 kcal/mol for tryptophan, and of −117.5 kcal/mol for tyrosine.

Uncertainties caused by issues with the zero-point vibrational energy and the conformer corrections rival, and probably exceed, those directly related to the electronic structure treatment. The overall uncertainty is somewhat difficult to quantify, but a semi-quantitative estimate would range from about ±0.5 kcal/mol for the smaller, to about ±1 kcal/mol for the larger, amino acids.

For glycine, by way of validation, we were able to obtain a “quasi-W4” result corresponding to  $TAE_e = 968.1$ ,  $TAE_0 = 918.6$ ,  $\Delta_f H_{f,298}^\circ = -90.0$ , and  $\Delta_f H_{f,298}^\circ = -94.0$  kcal/mol.

Our best theoretical values suggest that the experimental gas-phase heats of formation from the NIST WebBook should be revised downward by 2.4 (alanine), 0.7–0.8 (glycine), 3.2 (methionine), and 5.3 (proline) kcal/mol. Similarly, we suggest that the experimental values from the

CRC Handbook should be revised downward by 0.4 (glycine), 2.0 (phenylalanine), and 4.8 (valine) kcal/mol. Our best theoretical values are in good agreement with the recently reported experimental values of Roux and coworkers for alanine [15] and methionine, [18] but suggest that their experimental value for cysteine should be revised downward by 2.8 kcal/mol. Finally, our best theoretical values for alanine and glycine are in excellent agreement with the recent values of Dorofeeva and Ryzhova [97].

Using our W1-F12 and W2-F12 benchmark heats of formation, we benchmark the performance of the empirical composite *Gn* procedures. We obtain the following RMSDs: 2.25 (G3(MP2)), 1.13 (G3(MP2)B3), 0.60 (G3B3), 1.80 (G4(MP2)), 0.71 (G4(MP2)-6X), and 0.72 (G4) kcal/mol. Particularly G4(MP2)-6X appears to offer an excellent performance-to-computational cost ratio.

Finally, it appears that for W1- and W1-F12, the scaling factor for the B3LYP/cc-pV(T+dZ)Z or B3LYP/aug'-cc-pV(T+d)Z zero-point vibrational energy should be revised upward to 0.990.

## 5 Supporting information

B3LYP/A'VTZ optimized geometries for the species considered in the present work (Table S1). Full references for ref [40] (Gaussian 09) and ref [41] (Molpro 2010) (Table S2). B2PLYP/def2-TZVPP harmonic frequencies for all amino acids except tryptophan and tyrosine, and B3LYP/aug'-cc-pV(T+d)Z frequencies for all amino acids.

**Acknowledgments** JMLM is the Baroness Thatcher Professor of Chemistry at the Weizmann Institute of Science and acknowledges partial financial support from the Lise Meitner-Minerva Center for Computational Quantum Chemistry and the Helen and Martin Kimmel Center for Molecular Design. This research was supported in part by the Weizmann AERI (Alternative Energy Research Initiative) and by a startup grant from the University of North Texas from which the Martin group Linux cluster was purchased. The authors would like to thank Dr. David Hrovat for assistance with procurement and management of the latter. A.K. is the recipient of an Australian Research Council (ARC) Discovery Early Career Researcher Award (project number: DE140100311). We also acknowledge the generous allocation of computing time from the National Computational Infrastructure (NCI) National Facility and the support of iVEC through the use of advanced computing resources located at iVEC@UWA.

**Note added in revision** For nine of the amino acids, we were able to compute correlation corrections to the DBOC at the CCSD/cc-pVDZ level using the CFOUR package [98, 99]. (We have shown earlier [91] that the correlation contribution to DBOCs converges very rapidly with the basis set.) They uniformly reduce the DBOCs: values are Gly -0.10, Ala -0.13, Ser -0.14, Cys -0.13, Asn -0.16, Asp -0.15, Pro -0.16, Thr -0.17, and Val -0.18 kcal/mol. Linear regression through the origin reveals that DBOC[CCSD] ~ 0.642 DBOC[HF]: This suggests TAE reductions of up to 0.26 kcal/mol (arginine) for the remaining

amino acids if correlation were included in the DBOC. Once again, this highlights that the nuclear motion, rather than the clamped-nuclei electronic structure, is the accuracy-limiting factor here.

## References

- Hättig C, Klopper W, Köhn A, Tew DP (2012) Chem Rev 112:4
- Kong L, Bischoff FA, Valeev E (2012) Chem Rev 112:75
- Ten-no S, Noga J (2012) WIREs Comput Mol Sci 2:114
- Peterson KA, Feller D, Dixon DA (2012) Theor Chem Acc 131:1079
- Quinn JR, Zimmerman SC, Del Bene JE, Shavitt I (2007) J Am Chem Soc 129:934
- Distasio RA Jr, Steele RP, Rhee YM, Shao Y, Head-Gordon M (2007) J Comp Chem 28:839
- Šponer J, Riley KE, Hobza P (2008) Phys Chem Chem Phys 10:2595
- Valdes H, Pluháčková K, Pitoňák M, Řezáč J, Hobza P (2008) Phys Chem Chem Phys 10:2747
- Valdes H, Pluháčková K, Hobza P (2009) J Chem Theory Comput 5:2248
- Jiang J, Wu Y, Wang Z-X, Wu C (2010) J Chem Theory Comput 6:1199
- Tkatchenko A, Rossi M, Blum V, Ireta J, Scheffler M (2011) Phys Rev Lett 106:118102
- Bokatjian-Johnson SS, Stover ML, Dixon DA (2012) J Phys Chem B 116:14844
- Goerigk L, Karton A, Martin JML, Radom L (2013) Phys Chem Chem Phys 15:7028
- Nelson DL, Cox MM (2008) Lehninger principles of biochemistry. Palgrave-Macmillan, New York
- Ribeiro da Silva MAV, Ribeiro da Silva MDMC, Santos AFLOM, Roux MV, Foces-Foces C, Notario R, Guzman-Majia R, Juaristi E (2010) J Phys Chem B 114:16471
- Notario R, Roux MV, Foces-Foces C, da Silva MAVR, da Silva MDMCR, Santos AFLOM, Guzman-Mejia R, Juaristi E (2011) J Phys Chem B 115:9401
- Stover ML, Jackson VE, Matus MH, Adams MA, Cassidy CJ, Dixon DA (2012) J Phys Chem B 116:2905
- Roux MV, Notario R, Segura M, Chickos JS, Liebman JF (2012) J Phys Org Chem 25:916
- Roux MV, Foces-Foces C, Notario R, da Silva MAVR, da Silva MDMC, Santos AFLOM, Juaristi E (2010) J Phys Chem B 114:10530
- Brás NF, Perez MAS, Fernandes PA, Silva PJ, Ramos MJ (2011) J Chem Theory Comput 7:3898
- Ramabhadran RO, Sengupta A, Raghavachari K (2013) J Phys Chem A 117:4973
- Jaeger HM, Schaefer HF III, Demaison J, Császár AG, Allen WD (2010) J Chem Theory Comput 6:3066
- Wilke JJ, Lind MC, Schaefer HF III, Császár AG, Allen WD (2009) J Chem Theory Comput 5:1511
- Balabin RM (2011) Comp Theor Chem 965:15
- Balabin RM (2009) Chem Phys Lett 479:195
- Shavitt I, Bartlett RJ (2009) Many-body methods in chemistry and physics: MBPT and coupled-cluster theory. Cambridge Molecular Science, Cambridge
- Shavitt I (1993) The history and evolution of Gaussian basis sets. Isr J Chem 33:357
- Bartlett RJ, Cole SJ, Purvis GD, Ermler WC, Hsieh HC, Shavitt I (1987) J Chem Phys 87:6579
- Shavitt I (1985) Tetrahedron 41:1531
- Comeau DC, Shavitt I, Jensen P, Bunker PR (1989) J Chem Phys 90:6491

31. Karton A, Martin JML (2012) *J Chem Phys* 136:124114
32. Karton A, Daon S, Martin JML (2011) *Chem Phys Lett* 510:165
33. Curtiss LA, Redfern PC, Raghavachari K (2011) *WIREs Comput Mol Sci* 1:810
34. Lee C, Yang W, Parr RG (1988) *Phys Rev B* 37:785
35. Becke AD (1993) *J Chem Phys* 98:5648
36. Stephens PJ, Devlin FJ, Chabalowski CF, Frisch MJ (1994) *J Phys Chem* 98:11623
37. Dunning TH (1989) *J Chem Phys* 90:1007
38. Kendall RA, Dunning TH, Harrison RJ (1992) *J Chem Phys* 96:6796
39. Dunning TH Jr, Peterson KA, Wilson AK (2001) *J Chem Phys* 114:9244
40. Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR, Scalmani G, Barone V, Mennucci B, Petersson GA, et al (2009) Gaussian 09, Revision D01, Gaussian Inc, Wallingford CT. See also: URL: <http://www.gaussian.com>
41. Werner H-J, Knowles PJ, Manby FR, Schütz M, Celani P, Knizia G, Korona T, Lindh R, Mitrushenkov A, Rauhut G et al (2012) Molpro 2012.1, University College Cardiff Consultants Limited: Cardiff U.K. See also: <http://www.molpro.net>.
42. Peterson KA, Adler TB, Werner H-J (2008) *J Chem Phys* 128:084102
43. Hill JG, Peterson KA, Knizia G, Werner H-J (2009) *J Chem Phys* 131:194105
44. Feller D, Peterson KA (2013) *J Chem Phys* 139:084110
45. Ten-no S (2004) *Chem Phys Lett* 398:56
46. Werner H-J, Adler TB, Manby FR (2007) *J Chem Phys* 126:164102
47. Knizia G, Werner H-J (2008) *J Chem Phys* 128:154103
48. Adler TB, Knizia G, Werner H-J (2007) *J Chem Phys* 127:221106
49. Knizia G, Adler TB, Werner H-J (2009) *J Chem Phys* 130:054104
50. Martin JML, de Oliveira G (1999) *J Chem Phys* 111:1843
51. Peterson KA, Dunning TH (2002) *J Chem Phys* 117:10548
52. Douglas M, Kroll NM (1974) *Ann Phys* 82:89
53. Heß BA (1986) *Phys Rev A* 33:3742
54. de Jong WA, Harrison RJ, Dixon DA (2001) *J Chem Phys* 114:48
55. Ruscic B, Pinzon RE, Morton ML, von Laszewski G, Bittner S, Nijssure SG, Amin KA, Minkoff M, Wagner AF (2004) *J Phys Chem A* 108:9979
56. Ruscic B (2004) *Encyclopedia of Science and Technology* (2005 Yearbook of Science and Technology). McGraw-Hill, New York, p 3
57. Ruscic B, Pinzon RE, Morton ML, Srinivasan NK, Su M-C, Sutherland JW, Michael JV (2006) *J Phys Chem A* 110:6592
58. Stevens WR, Ruscic B, Baer T (2010) *J Phys Chem A* 114:13134
59. Ruscic B, Feller D, Peterson KA (2014) *Theor Chem Acc* 133:1415. doi:10.1007/s00214-013-1415-z
60. Cox JD, Wagman DD, Medvedev VA (1989) CODATA key values for thermodynamics, Hemisphere Publishing Corp.: New York. <http://www.codata.org/resources/databases/key1.html>
61. Curtiss LA, Redfern PC, Raghavachari K, Rassolov V, Pople JA (1999) *J Chem Phys* 110:4703
62. Baboul AG, Curtiss LA, Redfern PC, Raghavachari K (1999) *J Chem Phys* 110:7650
63. Curtiss LA, Redfern PC, Raghavachari K (2007) *J Chem Phys* 126:84108
64. Curtiss LA, Redfern PC, Raghavachari K (2007) *J Chem Phys* 127:124105
65. Chan B, Deng J, Radom L (2011) *J Chem Theory Comput* 7:112
66. Curtiss LA, Redfern PC, Raghavachari K (2005) *J Chem Phys* 123:124107
67. Montgomery JA, Frisch MJ, Ochterski JW, Petersson GA (1999) *J Chem Phys* 110:2822
68. Karton A, Rabinovich E, Martin JML, Ruscic B (2006) *J Chem Phys* 125:144108
69. Bak KL, Jørgensen P, Olsen J, Helgaker T, Gauss J (2000) *Chem Phys Lett* 317:116
70. Boese AD, Oren M, Atasoylu O, Martin JML, Kállay M, Gauss J (2004) *J Chem Phys* 120:4129
71. Stanton JF (1997) *Chem Phys Lett* 281:130
72. Karton A, Taylor PR, Martin JML (2007) *J Chem Phys* 127:064104
73. Harding ME, Vázquez J, Ruscic B, Wilson AK, Gauss J, Stanton JF (2008) *J Chem Phys* 128:114111 and references therein
74. Fogueri UR, Kozuch S, Karton A, Martin JML (2013) *Theor Chem Acc* 132:1291
75. Karton A, Kaminker I, Martin JML (2009) *J Phys Chem A* 113:7610
76. Harding ME, Vázquez J, Gauss J, Stanton JF, Kállay M (2011) *J Chem Phys* 135:044513
77. Ramabhadran RO, Raghavachari K (2011) *J Chem Theory Comput* 7:2094
78. Ramabhadran RO, Raghavachari K (2012) *J Phys Chem A* 116:7531
79. Afeefy HY, Liebman JF, Stein SE “Neutral Thermochemical Data” in NIST Chemistry WebBook, NIST Standard Reference Database Number 69, Eds Linstrom PJ, Mallard WG, National Institute of Standards and Technology, Gaithersburg MD, 20899. <http://webbook.nist.gov>. Retrieved October 22, 2013
80. (2012) CRC handbook of chemistry and physics, 93rd edn. CRC Press, Boca Raton 2013
81. Barone V, Biczysko M, Bloino J, Puzzarini C (2013) *Phys Chem Chem Phys* 15:10094–10111. The absolute ZPVE for the lowest-energy conformer is not given explicitly in the paper, but is 49.438 kcal/mol: Puzzarini C, personal communication to authors (January, 2014).
82. Dunham JL (1932) *Phys Rev* 41:721
83. Del Bene JE, Aue DH (1992) *Shavitt I* 114:1631
84. Grev RS, Janssen CL, Schaefer HF (1991) *J Chem Phys* 95:5128
85. Schuurman MS, Allen WD, Schaefer HF (2005) *J Comput Chem* 26:1106
86. Alecu IM, Zheng J, Zhao Y, Truhlar DG (2010) *J Chem Theor Comput* 6:2872. <http://comp.chem.umn.edu/freqscale/index.html>
87. Csonka G, Ruzsinszky A, Perdew JP (2005) *J Phys Chem A* 109:6779
88. Irikura KK, Johnson RD, Kacker RN, Kessel R (2009) *J Chem Phys* 130:114102
89. Sinha P, Boesch SE, Gu C, Wheeler RA, Wilson AK (2004) *J Phys Chem A* 108:9213
90. Merrick JP, Moran D, Radom L (2007) *J Phys Chem A* 111:11683. <http://groups.chem.usyd.edu.au/radom/More/ScaleFactor.html>
91. Karton A, Ruscic B, Martin JML (2007) *J Mol Struct Theochem* 811:345
92. Pfeiffer F, Rauhut G, Feller D, Peterson KA (2013) *J Chem Phys* 138:044311
93. Grimme S (2006) *J Chem Phys* 124:034108
94. Weigend F, Ahlrichs R (2005) *Phys Chem Chem Phys* 7:3297
95. Kozuch S, Gruzman D, Martin JML (2011) *J Phys Chem C* 114:20801, Table S-16.
96. Gruzman D, Karton A, Martin JML (2009) *J Phys Chem A* 113:11974
97. Dorofeeva OV, Ryzhova ON (2009) *J Chem Thermodyn* 41:433
98. CFOUR, Coupled-Cluster techniques for Computational Chemistry, a quantum-chemical program package by Stanton JF,

- Gauss J, Harding ME, Szalay PG with contributions from Auer AA, Bartlett RJ, Benedikt U, Berger C, Bernholdt DE, Bomble YJ, Cheng L, Christiansen O, Heckert M, Heun O, Huber C, Jagau TC, Jonsson D, Juslius J, Klein K, Lauderdale WJ, Matthews DA, Metzroth T, Mück LA, O'Neill DP, Price DR, Prochnow E, Puzzarini C, Ruud K, Schiffmann F, Schwalbach W, Simmons C, Stopkowicz S, Tajti A, Vázquez J, Wang F, Watts JD and the integral packages MOLECULE (Almlöf J, Taylor PR), PROPS (Taylor PR), ABACUS (Helgaker T, Jensen HJA, Jørgensen P, Olsen J), and ECP routines by Mitin AV, van Wüllen C. For the current version. <http://www.cfour.de>
99. Gauss J, Tajti A, Kállay M, Stanton JF, Szalay PG (2006) *J Chem Phys* 125:144111
  100. Ling S, Yu W, Huang Z, Lin Z, Haranczyk M, Gutowski M (2006) *J Phys Chem A* 110:12282–12291
  101. Chen M, Huang Z, Lin Z (2005) *J Mol Struct Theochem* 719:153–158
  102. Chen M, Lin Z (2007) *J Chem Phys* 127:154314
  103. Meng L, Lin Z (2011) *Comp Theor Chem* 976:42–50
  104. Pang R, Guo M, Ling S, Lin Z (2013) *Comp Theory Chem* 1020:14–21
  105. Huang Z, Yu W, Lin Z (2006) *J Mol Struct Theochem* 801:7–20
  106. Dokmairijan S, Lee VS, Nimmanpipug P (2010) *J Mol Struct Theochem* 953:28–38
  107. Boeckx B, Maes G (2012) *J Phys Chem B* 116:12441–12449
  108. Huang Z, Yu W, Lin Z (2006) *J Mol Struct Theochem* 758:195–202
  109. Czinki E, Császár AG (2003) *Chem Eur J* 9:1008–1019
  110. Szidarovszky T, Czakó G, Császár AG (2009) *Mol Phys* 107:761–775
  111. Huang Z, Lin Z (2005) *J. Phys. Chem. A* 109:2656. MP2/6-311++G(d, p) total energies for 45 lowest conformers, with B3LYP/6-311G\* zero-point energies.
  112. Zhang M, Huang Z, Lin Z (2005) *J Chem Phys* 122: 134313. Lowest 36 conformers, MP2/6-311G(2df, p)//B3LYP/6-311++G(d, p) with zero-point energy from B3LYP/6-311++G(d, p).