Electronic Supplementary Material (ESI) for Physical Chemistry Chemical Physics.
This journal is © the Owner Societies 2016

An adaptive distance-based group contribution method for thermodynamic property prediction

Tanjin He,[ab] Shuang Li,[db] Yawei Chi,[bc] Hongbo Zhang,[bc] Zhi Wang,*[a] Bin Yang,[bc] Xin He,[e]
Xiaoqing You*[bc]

[a] State Key Laboratory of Automotive Safety and Energy, Tsinghua University, Beijing 100084, China
[b] Center for Combustion Energy, Tsinghua University, Beijing 100084, China
[c] Key Laboratory for Thermal Science and Power Engineering of the Ministry of Education, Tsinghua University, Beijing 100084, China
[d] National Synchrotron Radiation Laboratory, University of Science and Technology of China, Hefei, Anhui 230029, China
[e] Aramco Services Company: Aramco Research Center-Detroit, Novi, MI 48377, USA

Corresponding Author:

Zhi Wang
State Key Laboratory of Automotive Safety and Energy,
Tsinghua University, Beijing, 100084 China.
Phone: +86-10-62772515
E-mail: wangzhi@tsinghua.edu.cn

Xiaoqing You
Center for Combustion Energy, and Department of Thermal Engineering,
Tsinghua University Beijing 100084, China
Phone: +86-10-62771854
E-mail: xiaoqing.you@tsinghua.edu.cn

Content

1. Thermodynamic database

The file “currentDatabase.xlsx” is the thermodynamic database used in the training and test. Currently this database has 925 alkanes ranging from C3 to C13, 634 alkenes ranging from C3 to C10, 871 alkyl radicals ranging from C3 to C10, and 1458 alkenyl radicals ranging from C3 to C9. The standard enthalpy of formation at M062X/def2TZVP//B3LYP/6-31G(d) level of theory, the optimized structure and frequencies at B3LYP/6-31G(d) level of theory, the group contribution vector, SMILES, connectivity, formula, atoms number and multiplicity are recorded in this database file. In addition, the alkane database with the conventional GA groups, which used in the comparison between the distance-based contribution method and the conventional GA method is also provided as the file “conventionalGAGroups.xlsx”. For updating conveniently in the future, the database can be also downloaded on the github website https://github.com/DBGConline/DBGC_distance-based_group_contribution.

2. Source code, fitted parameters, and example

The source code implementing the distance-based group contribution method is in the directory “DBGC_distance-based_group_contribution”. The fitted parameters based on the training set containing 305 alkanes (C3-C11), 113 alkenes (C3-C8), 364 alkyl radicals (C3-C9), and 546 alkenyl radicals (C3-C8) is in the directory “TrainingTestPrediction/savedNet”. A step-by-step instruction to experience the use of the distance-based group contribution method is in the file “Example.txt”. More detailed help with how to use the code can be find in the file “ReadMe.txt”. An online demo can be found on the website http://DBGC.online. For updating conveniently in the future, the code can be also downloaded on the github website https://github.com/DBGConline/DBGC_distance-based_group_contribution.

3. Standard enthalpy of formation calculation

A virtual reaction, x C(g) + y/2 H_2(g) + z/2 O_2(g) = $C_xH_yO_z$(g) was constructed to calculated the gas phase standard enthalpy of formation of $C_xH_yO_z$, which should be the enthalpy difference between the substance to be calculated and the constituent chemical elements in the standard state as Eq. (S1),

$$\begin{aligned}\Delta_f H^o(C_xH_yO_z;298K) = H^o_{calc}(C_xH_yO_z;298K) - xH^o_{calc}(C;298K)\\ -\frac{y}{2}H^o_{calc}(H_2;298K) - \frac{z}{2}H^o_{calc}(O_2;298K)\\ +x\Delta_f H^o(C;298K) + \frac{y}{2}\Delta_f H^o(H_2;298K)\\ +\frac{z}{2}\Delta_f H^o(O_2;298K)\end{aligned} \tag{S1}$$

The standard enthalpies of formation of H_2 and O_2 are both zero. Because the standard state of carbon is solid graphite while the substance studies with *ab initio* methods is usually gas, it brings trouble to the determination of standard enthalpy of formation. The standard enthalpy of formation of C(g) is required. The experimental enthalpies of formation of 89 organic molecules, including alkane, alkene and oxygenated hydrocarbon, were used to calculate an averaged enthalpy of gas carbon with the same level of theoretical method in the paper. The species and the experimental enthalpies of formation are listed in Table S1. The calculated enthalpy of gas carbon is presented in Fig. S1. The red line is the experimental formation enthalpy of gas carbon on NIST Chemistry WebBook [1], 171.29 kcal/mol. The blue dots are the calculated formation enthalpy of gas carbon at M06-2X/def2-TZVP//B3LYP/6-31G(d) level of theory. The green line is the averaged value of computation, 172.04 kcal/mol. In this study the averaged value is used because we think this value is more consistent with the computational methodology, thus bringing an effect of error cancellation. Similar method was used in ref. [2], namely, decreasing the systematic error through replacement of the atomic standard enthalpy of formation. The predicted enthalpies of formation at M06-2X/def2-TZVP//B3LYP/6-31G(d) level of theory are listed in Table S1.

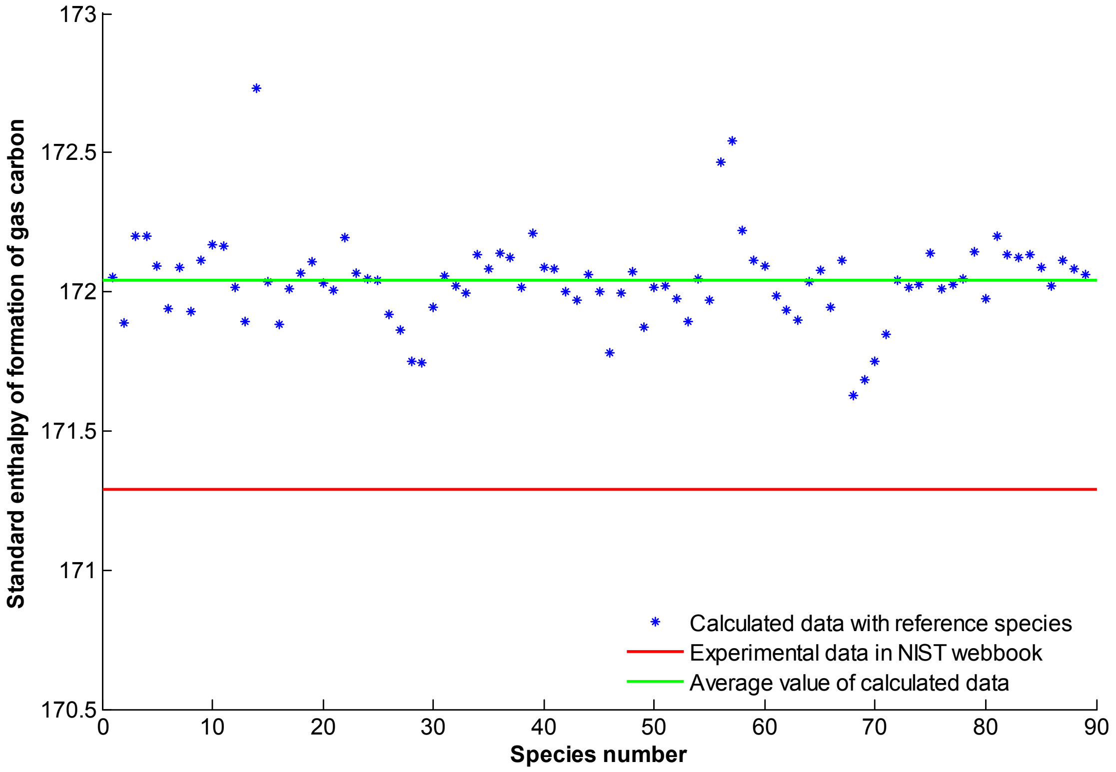
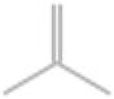
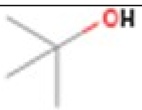
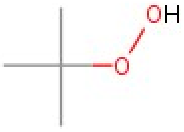
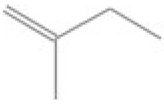
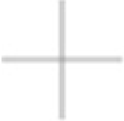
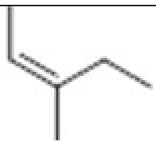
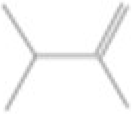


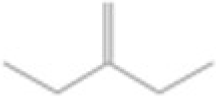
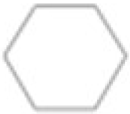
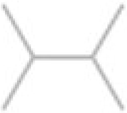
Fig. S1 Calculated standard enthalpy of formation of gas carbon at M06-2X/def2-TZVP//B3LYP/6-31G(d) level of theory and RRHO approximation

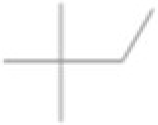
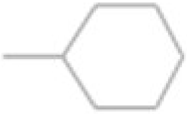
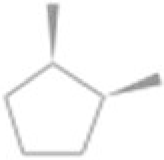
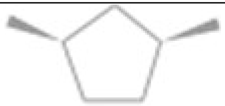
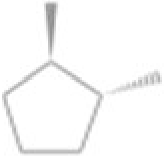
Table S1. Experimental and computational standard enthalpy of formation of reference species

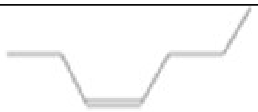
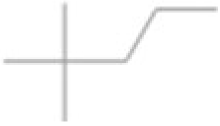
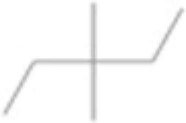
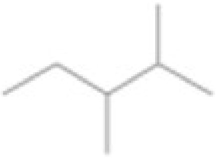
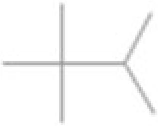
ID	Formula	Structure	Predicted standard enthalpy of formation [kcal/mol]	Experimental standard enthalpy of formation [kcal/mol]	Reference
1	$C3H7OH$		-61.24	-61.2±0.7	3-10
2	$C3H7OOH$		-46.66	-47.11	11
3	$C3H8$		-25.50	-25.02±0.12	7,12
4	$C4H10$		-30.67	-30.03±0.16	7,12

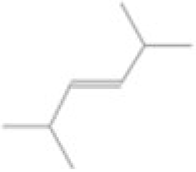
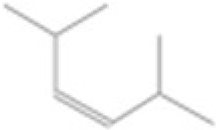
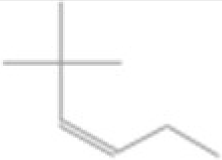
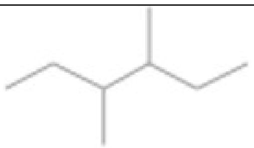
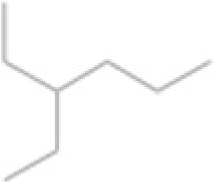
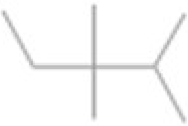
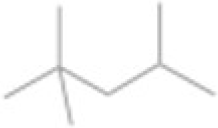
5	C4H10		-32.29	-32.07±0.15	7,12
6	C4H8		0.26	-0.15±0.19	7,13
7	C4H8		-2.77	-2.58±0.24	7,13
8	C4H8		-3.85	-4.29±0.26	7,13
9	C4H9OH		-66.30	-66±1	3,5-8,11,15-18
10	C4H9OH		-68.31	-67.8±0.2	7,19
11	C4H9OH		-70.55	-70.05	5,8,14
12	C4H9OH		-74.63	-74.72±0.21	7,20
13	C4H9OOH		-55.55	-56.14	11
14	C5H10		-5.44	-1.97±0.28	7,21
15	C5H10		-4.99	-5±2	5,22-26
16	C5H10		-6.21	-7±1	5,7,23,24,27-29
17	C5H10		-7.55	-7.7±0.4	5,23,24,27-30

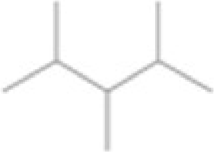
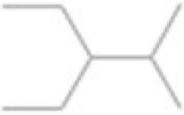
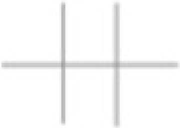
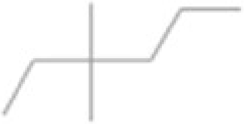
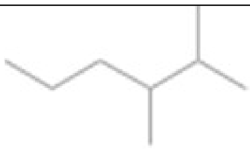
18	C5H10		-6.23	-6.09	5,24
19	C5H10		-8.73	-8.39±0.2	7,20
20	C5H10		-9.89	-9.92±0.21	7,20
21	C5H10		-18.09	-18.26±0.19	7,31
22	C5H12		-35.85	-35.08±0.14	7,32
23	C5H12		-36.87	-36.73±0.14	7,32
24	C5H12		-40.18	-40.14±0.15	7,32
25	C6H12		-11.82	-11.82±0.16	33
26	C6H12		-14.06	-14.79±0.21	20
27	C6H12		-14.11	-15.18±0.21	20
28	C6H12		-13.99	-15.74±0.36	20
29	C6H12		-15.03	-16.8±0.36	7,20

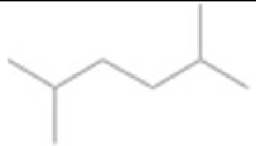
30	C6H12		-12.82	-13.4±0.21	20
31	C6H12		-29.88	-29.78	5,34,35
32	C6H12		-10.09	-10.2±0.6	5,7,25,36-43
33	C6H12		-25.07	-25.33	5,35,44
34	C6H12		-7.17	-6.6±0.2	7,45
35	C6H12		-11.46	-11.2±0.26	37,46
36	C6H12		-12.80	-12.2±0.2	37,46
37	C6H12		-12.32	-11.8±0.24	37,46
38	C6H12		-10.87	-11±0.2	37,46
39	C6H14		-40.95	-39.94	44,47
40	C6H14		-41.95	-41.66±0.25	7,47
41	C6H14		-41.29	-41.02±0.23	47
42	C6H14		-42.25	-42.49±0.24	7,47

43	C6H14		-43.93	-44.35±0.23	7,47
44	C7H14		-15.25	-15.1	23,48
45	C7H14		-16.63	-16.9	23
46	C7H14		-35.19	-36.99±0.25	7,35
47	C7H14		-30.04	-30.37±0.25	7,35
48	C7H14		-17.94	-17.7	22
49	C7H14		-31.87	-33.05±0.28	7,49
50	C7H14		-30.80	-30.96±0.32	7,49
51	C7H14		-31.78	-31.93±0.35	7,49
52	C7H14		-32.22	-32.67±0.3	7,49
53	C7H14		-31.45	-32.47±0.29	7,49
54	C7H14		-17.56	-17.5	23

55	C7H14		-16.10	-16.6	23
56	C7H14		-13.66	-10.66±0.39	50,51
57	C7H14		-15.26	-11.74±0.43	50,51
58	C7H16		-46.15	-44.89±0.19	7,47
59	C7H16		-47.10	-46.6±0.3	7,47
60	C7H16		-46.34	-45.96±0.3	7,47
61	C7H16		-48.89	-49.29±0.32	7,47
62	C7H16		-47.44	-48.17±0.22	7,47
63	C7H16		-46.62	-47.62±0.3	7,47
64	C7H16		-48.28	-48.3±0.23	7,47
65	C7H16		-45.61	-45.34±0.28	7,47
66	C7H16		-48.30	-48.96±0.27	7,47

67	C8H16		-20.41	-19.82	52
68	C8H16		-25.27	-28.56	53
69	C8H16		-23.74	-26.6	53
70	C8H16		-25.88	-28.2	53
71	C8H16		-21.56	-23.1	53
72	C8H18		-50.94	-50.91±0.37	7,47
73	C8H18		-50.19	-50.4±0.28	7,47
74	C8H18		-52.52	-52.61±0.36	47
75	C8H18		-52.54	-51.73±0.33	7,47
76	C8H18		-53.34	-53.57±0.32	7,47

77	C8H18		-51.86	-51.97±0.4	7,47
78	C8H18		-50.52	-50.48±0.31	7,47
79	C8H18		-52.21	-51.38±0.3	7,47
80	C8H18		-53.53	-54.06	5,47,54
81	C8H18		-51.17	-49.88	5,47,54
82	C8H18		-52.25	-51.5±0.31	7,47
83	C8H18		-51.49	-50.82±0.27	7,47
84	C8H18		-51.44	-50.69±0.28	7,47
85	C8H18		-54.08	-53.71±0.24	7,47
86	C8H18		-52.48	-52.61±0.26	7,47
87	C8H18		-51.71	-51.13±0.36	7,47
88	C8H18		-52.78	-52.44±0.27	7,47

89	C8H18		-53.41	-53.21±0.36	7,47

4. Minimum energy conformer approximation

For long chain molecules, multiple conformers may exist, because the arrangements of atoms in space can be interconverted by rotation around single bonds. In this study, conformational analysis was performed to obtain conformers with the minimum energy, which were used to represent the corresponding molecules. To catch a glimpse of the influence without considering other low-energy conformers, we randomly chose 7 C_{10} molecules as a test set. Assuming the population of conformers follows the Boltzmann distribution, the enthalpy difference between the approach using only the minimum energy conformer and that using multiple lowest-energy conformers can be calculated with Eq. (S2).

$$H_T(X) - H_T(1) = \sum_{i=1}^{n} \frac{\sigma_i \exp(-(G_T(i) - G_T(1))/RT)}{\sum_{i=1}^{n} \sigma_i \exp(-(G_T(i) - G_T(1))/RT)} (H_T(i) - H_T(1)) \tag{S2}$$

where X represents the species name, $H_T(X)$ is the enthalpy calculated when considering multiple conformers for species X, $H_T(i)$ is the enthalpy of the i^{th} conformer at temperature T, $G_T(i)$ is the Gibbs free energy of the i^{th} conformer at temperature T, the 1st conformer is the minimum energy one, σ accounts for degeneracies, and R is the gas constant. In this test, the initial structures of conformers were generated with MSTor [55-58], by systematically rotating all rotatable bonds through 360° using a fixed increment of 120°, and 3^m initial structures would be generated when there are m rotatable bonds. These initial structures were all optimized at the B3LYP/6-31G(d) level of theory, and the enthalpies were calculated at the same level of theory. For a further simplification, the entropies are assumed to be the same for the conformers corresponding to the same molecule. Thus the free energy difference and enthalpy different in Eq. (S2) can be substituted with enthalpy difference between conformers:

$$H_T(X) - H_T(1) = \sum_{i=1}^{n} \frac{\sigma_i \exp(-(H_T(i) - H_T(1))/RT)}{\sum_{i=1}^{n} \sigma_i \exp(-(H_T(i) - H_T(1))/RT)} (H_T(i) - H_T(1)) \tag{S3}$$

The values of $H_{298K}(X)$-$H_{298K}(1)$ for the test cases are presented in Fig. S2. With more and more conformers considered, the enthalpy difference at 298.15 K between approaches using minimum energy conformer and multiple conformers increases gradually, and moves close to an upper limit. Typically, the enthalpy difference is less than 1 kcal/mol between these two approaches. The influence of conformers is limited is because that the conformer with a large population fraction also has a similar energy to the minimum energy conformer, although their structures can be substantially different.

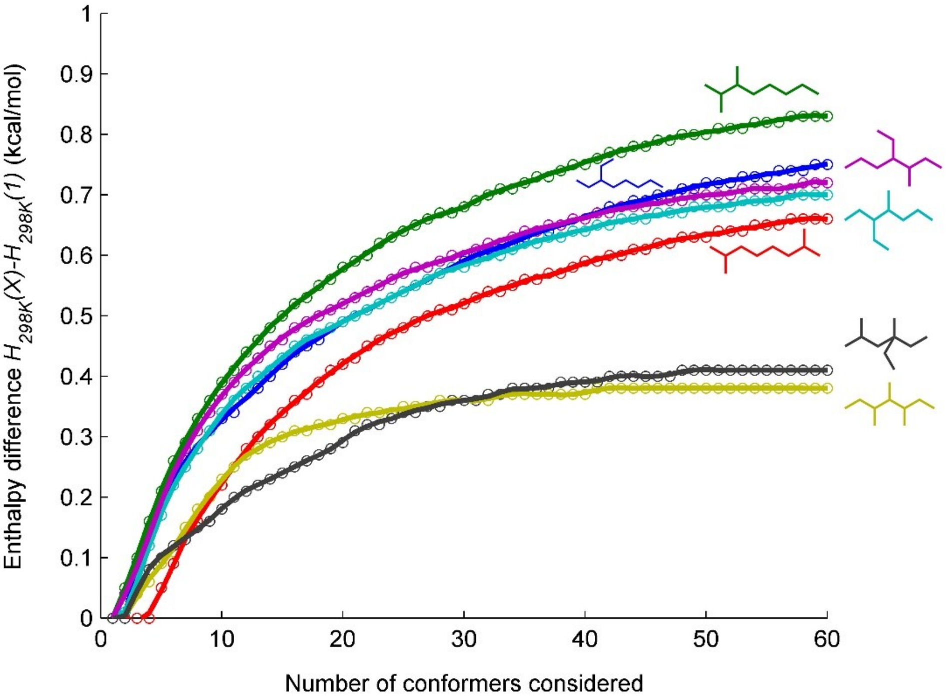


Fig. S2 The variation of enthalpy difference at 298.15 K between approaches using minimum energy conformer and multiple conformers ($H_{298K}(X)$-$H_{298K}(1)$) with the increasing number of conformers considered for 7 test C_{10} molecules.

We may also estimate the enthalpy difference between the two approaches in another way. Assuming the entropies of conformers corresponding to the same molecule are the same, Eq. (S2) can be written as

$$H_T(X) - H_T(1) = RT\sum_{i=1}^{n} \frac{\sigma_i \exp(-h_i)}{\sum_{i=1}^{n} \sigma_i \exp(-h_i)} h_i \tag{S4}$$

where

$$h_i = (H_T(i) - H_T(1)) / RT \tag{S5}$$

To simplify the calculation, let's assume h_i is evenly distributed with space ΔH ($\Delta H = h_{i+1} - h_i$), and the degeneracy number σ is constant for the conformers corresponding to the same molecule. As shown in Fig. S3, the function *exp(-x)* is a monotonic decreasing one. The function *x*exp(-x)* is monotonically increasing in the interval [0, h_m], ($h_m = 1$), monotonically decreasing in the interval [h_m, +∞], and reaches the maximum value at point (1, *exp*(-1)). Then,

$$\begin{aligned}
H_T(X) - H_T(1) &= RT\sum_{i=1}^{n} \frac{\exp(-h_i)}{\sum_{i=1}^{n} \exp(-h_i)} h_i \\
&= RT\frac{\sum_{i=1}^{n} h_i \exp(-h_i)\Delta H}{\sum_{i=1}^{n} \exp(-h_i)\Delta H} \\
&< RT\frac{\sum_{h_i \le h_m - \Delta H} h_i \exp(-h_i)\Delta H + \sum_{h_m - \Delta H < h_i < h_m + \Delta H} h_i \exp(-h_i)\Delta H + \sum_{h_i \ge h_m + \Delta H} h_i \exp(-h_i)\Delta H}{\sum_{i=1}^{n} \int_{h_i}^{h_{i+1}} \exp(-h)dh} \\
&< RT\frac{\sum_{h_i \le h_m - \Delta H} \int_{h_i}^{h_{i+1}} h_i \exp(-h_i)dh + 2\exp(-1)\Delta H + \sum_{h_i \ge h_m + \Delta H} \int_{h_{i-1}}^{h_i} h_i \exp(-h_i)dh}{\sum_{i=1}^{n} \int_{h_i}^{h_{i+1}} \exp(-h)dh} \\
&\le RT\frac{\int_0^{h_m} h_i \exp(-h_i)dh + 2\exp(-1)\Delta H + \int_{h_m}^{+\infty} h_i \exp(-h_i)dh}{\int_0^{+\infty} \exp(-h)dh} \\
&= RT(1 + 2\exp(-1)\Delta H)
\end{aligned} \tag{S6}$$

When there are many conformers lying densely in energy, ΔH is a small number and basically less than 1 kcal/mol, making $H_{298K}(X)$-$H_{298K}(1) \le 1.03$ *kcal/mol*. This theoretical analysis is consistent with the results in Fig. S2. Considering the very accurate quantum chemistry calculation is beyond the scope of this work, and the error cancellation in the calculation of standard enthalpy of formation also helps decrease the systematic error, the simplification to use the minimum energy conformer is a proper approximation.

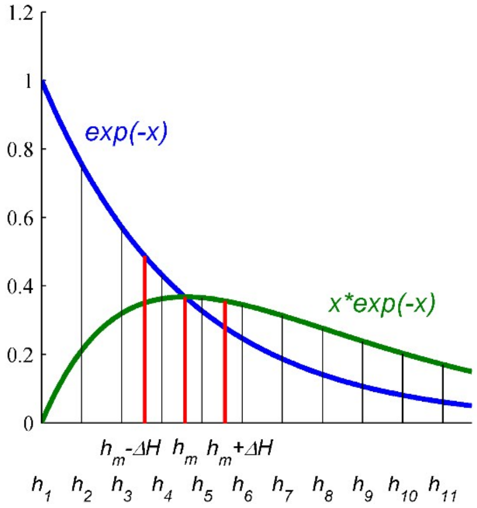
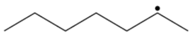
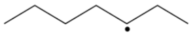


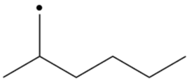
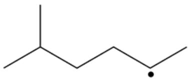
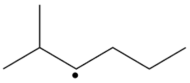
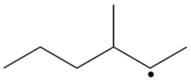
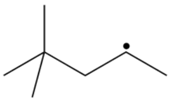
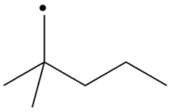
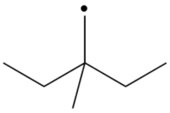
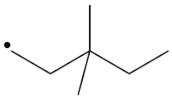
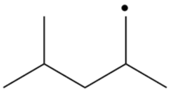
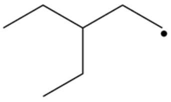
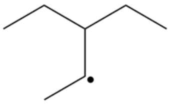
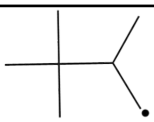
Fig. S3 Sketch of function *exp(-x)* and *x*exp(-x)*. The function *x*exp(-x)* reaches maximum value at (h_m, $h_m*exp(-h_m)$), where $h_m = 1$.

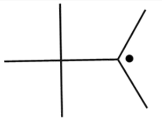
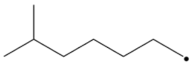
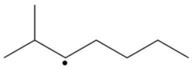
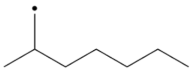
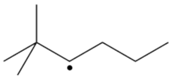
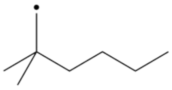
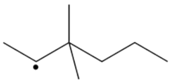
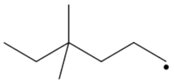
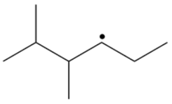
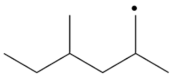
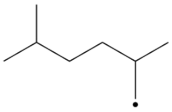
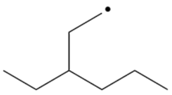
5. Simplification in obtaining radical structure

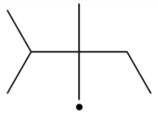
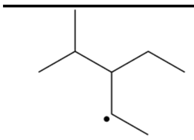
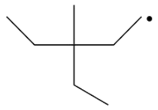
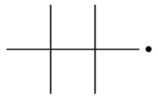
Conformational analysis for radicals is challenging because there can be several corresponding radicals with different radical sites for each molecule, which greatly increase the amount of computational work. For simplification, we obtain the initial geometry by removing a hydrogen atom from the parent molecule, and assume the corresponding optimized geometry represents the radical conformer with the minimum energy. We acknowledge that radical structure does not always follow the parent molecule. However, to some extent, the molecular structure of a radical is similar to that of the corresponding stable molecule [59], and this assumption is a balance between accuracy and efficiency. To explore whether this assumption is proper, 15 C_7 radicals and 15 C_8 radicals were randomly chosen to compare the difference between the radical conformer optimized based on the parent molecule and the minimum energy conformer. Two approaches were used: in the first one (approach A), the initial geometry was obtained by removing a hydrogen atom from the minimum energy conformer of the parent molecule, and then the geometry was optimized at the B3LYP/6-31G(d) level of theory; in the other one (approach B), a systematic conformational search was conducted to find the minimum energy conformer with MSTor [55-58], namely, each of all rotatable bonds was systematically rotated through 360° using a fixed increment of 120°, and 3n initial structures would be generated when there are n rotatable bonds. These initial structures were all optimized at the B3LYP/6-31G(d) level of theory, and the conformer with the minimum self-consistent field (SCF) energy was selected as the final structure of the radical. The enthalpies of radical structures produced by these two approaches was compared in Table S2 at the B3LYP/6-31G(d) level of theory. Two thirds of the minimum energy conformers are the same as the structures optimized based on the parent molecules. The mean unsigned enthalpy difference between these two approaches is 0.17 kcal/mol. The unbiased sample standard deviation of the enthalpy difference is 0.33 kcal/mol. The maximum unsigned enthalpy difference is ~1 kcal/mol. Considering the very accurate quantum chemistry calculation is beyond the scope of this work, the assumption about radical structure is a proper balance between accuracy and efficiency.

Table S2. Comparison of the enthalpies of radical structures obtained through approach A (optimizing initial structure obtained by removing a hydrogen atom from the parent molecule) and approach B (systematically rotating all rotatable bonds to find the minimum energy conformer) at the B3LYP/6-31G(d) level of theory.

Radical	Enthalpy of structure in approach *A* (hartree)	Enthalpy of structure in approach *B* (hartree)	Enthalpy difference (kcal/mol)
	-275.523064	-275.523064	0.00
	-275.522697	-275.522697	0.00

	-275.516670	-275.516670	0.00
	-275.522779	-275.522779	0.00
	-275.522834	-275.522834	0.00
	-275.521437	-275.522731	-0.81
	-275.523338	-275.523338	0.00
	-275.515828	-275.515828	0.00
	-275.513516	-275.513516	0.00
	-275.514233	-275.514459	-0.14
	-275.516432	-275.516432	0.00
	-275.513013	-275.513953	-0.59
	-275.520088	-275.521634	-0.97
	-275.513736	-275.513736	0.00

	-275.524174	-275.524269	-0.06
	-314.800829	-314.800829	0.00
	-314.807249	-314.807249	0.00
	-314.801093	-314.801093	0.00
	-314.806290	-314.806290	0.00
	-314.800252	-314.800252	0.00
	-314.802899	-314.804405	-0.94
	-314.796407	-314.796968	-0.35
	-314.803751	-314.805368	-1.01
	-314.799109	-314.799109	0.00
	-314.800752	-314.800761	-0.01
	-314.798290	-314.798290	0.00

	-314.794425	-314.794425	0.00
	-314.804184	-314.804345	-0.10
	-314.795473	-314.795473	0.00
	-314.792631	-314.792631	0.00

6. Observation on interaction between groups

Interacting quantum atoms (IQA) analysis [60,61] was performed for a simple observation on the interaction between two groups. We partitioned the energy of the molecule with atoms as chemically meaningful fragments. Each atom consists of its nucleus and its 3D atomic basin as defined in the Atom In Molecules (AIM) method of Bader and co-workers [62-64]. The interatomic interaction energy between atom A and atom B is defined with the same manner as in Blanco *et al.* [60]:

$$\begin{aligned} E_{\text{int}}^{AB} &= V_{nn}^{AB} + V_{en}^{AB} + V_{ne}^{AB} + V_{ee}^{AB} \\ &= \frac{Z^A Z^B}{r_{AB}} - \int_{\Omega_A} \frac{\rho_1(r_1) Z^B}{r_{1B}} dr_1 - \int_{\Omega_B} \frac{Z^A \rho_1(r_2)}{r_{A2}} dr_2 + \int_{\Omega_A}\int_{\Omega_B} \frac{\rho_2(r_1, r_2)}{r_{12}} dr_1 dr_2 \qquad \text{(S7)} \\ &(A \neq B) \end{aligned}$$

where V_{nn}^{AB} is the internuclear repulsion, V_{en}^{AB} is the interatomic nuclear attraction between electrons in basin of atom A and the nucleus of atom B, V_{ne}^{AB} is the interatomic nuclear attraction between the nucleus of atom A and electrons in basin of atom B, V_{ee}^{AB} is the interelectronic repulsion interaction between electrons in basin of atom A and electrons in basin of atom *B*, Z^X is the nuclear charge of atom *X*, r_{xy} is the distance between electron/nucleus *x* and electron/nucleus *y*, ρ_1 is the electron density function, ρ_2 is the electron pair-density function, and Ω_X is the atomic basin of atom *X*. In this IQA analysis, the group is regarded as a polyvalent atom (ligancy > 2) in a molecule together with all of its hydrogen ligands, such as CH_3- and -CH_2-. The interaction energy between group *i* and group *j* is defined as the summation of interatomic interaction energy between all atom pairs cross these two groups:

$$E_{\text{int}}^{ij} = \sum_{A\in i}\sum_{B\in j} E_{\text{int}}^{AB} \tag{S8}$$

where $A \in i$ means atom *A* belongs to group *i*, and $B \in j$ means atom *B* belongs to group *j*. To explore the interaction energy variation against different group distance d_{ij}, the molecules C_2H_6, C_3H_8, n-C_4H_{10}, n-C_5H_{12}, and n-C_6H_{14} were used for demonstration. For example, the distance d_{ij} between CH_3- and CH_3- groups in the molecule C_2H_6 is 1, while in the molecule C_3H_8 it 2. The interaction energy of three group pairs, CH3- and CH3-, CH3- and -CH2-, and -CH2- and -CH2-, are calculated at the HF/6-311G(d,p) level of theory as shown in Fig. S4. For the same type of group pair, there can be several pairs in the some molecules. For example, in the molecule n-C_4H_{10}, there are four pairs of CH_3- and –CH_2- group pair, and the group distance ranges from 1 to 2. All the group pairs with different group distance were taken into consideration. The interaction energy corresponding to d_{ij} larger than 3 is not presented, because sometimes $E_{\text{int}}^{ij}(d_{ij} = 4)$ and $E_{\text{int}}^{ij}(d_{ij} < 4)$ are opposite in sign, and usually the absolute value of $E_{\text{int}}^{ij}(d_{ij} \geq 4)$ is very small and can be neglected compared with that of $E_{\text{int}}^{ij}(d_{ij} < 4)$. When d_{ij} is less than or equal to 3, there is an approximately linear relationship between $\ln(-E_{\text{int}}^{ij})$ and d_{ij}, and the slope is about 3. This simple observation inspired us to use *exp(-3*d_{ij}*)* to approximate the interaction between two groups in the distance-based group contribution method. We acknowledge that it cannot be exactly accurate, thus artificial neural network (ANN) was used to taken some potential non-linear effect into consideration. It should be noted that the dimensionless value of *exp(-3*d_{ij}*)* is a relative scalar. After passed to ANN, weighted by weights in the network, and transformed by transfer functions in neuron nodes, the absolute contribution to the enthalpy of formation is combined into the last output of ANN, which carries the dimension of energy.

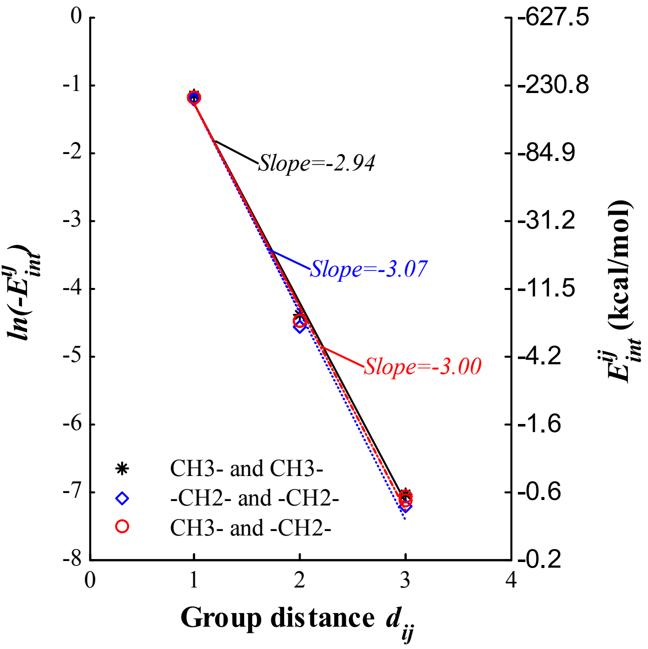
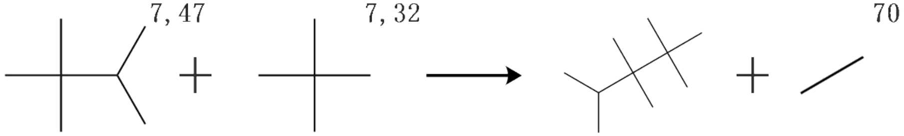
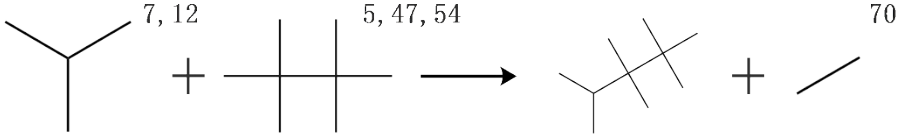
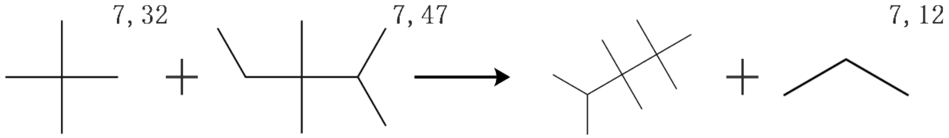
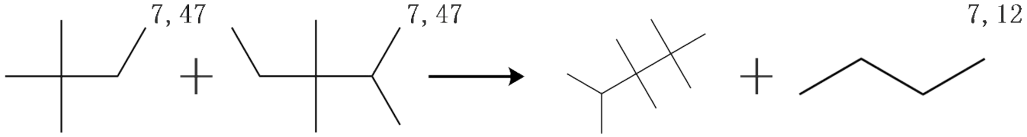
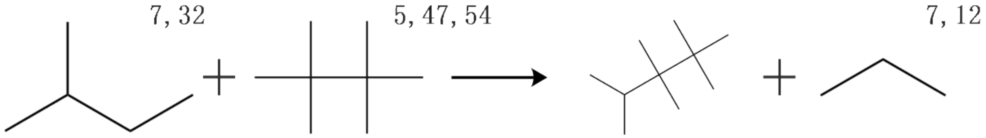


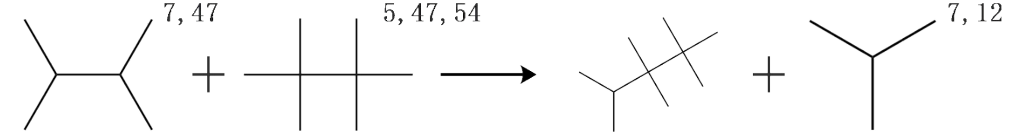
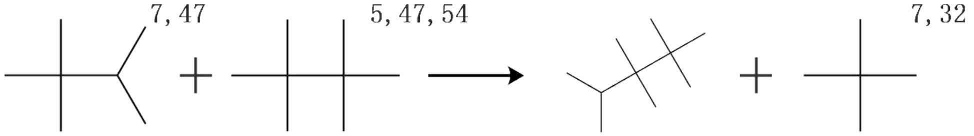
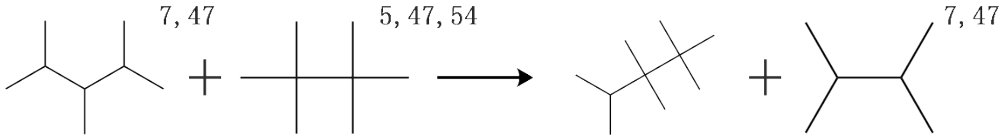
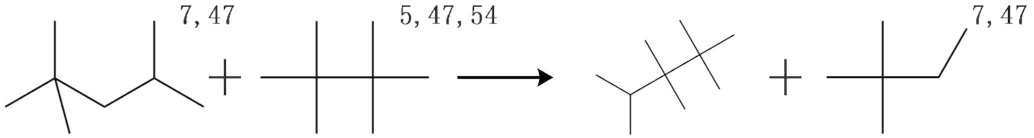
Fig. S4. Interaction energy between three kinds of group pairs with different group distance d_{ij}. ln(-E_{int}^{ij}) has a approximate linear relationship with d_{ij}, and the slope is about 3. The value of E_{int}^{ij} is also presented for reference. The colored lines are the linear fitting corresponding to the data points with the same color.

7. Homodesmotic reactions for 2,2,3,3,4-pentamethylpentane

The standard enthalpy of formation predicted by the distance-based group contribution method for 2,2,3,3,4-pentamethylpentane, which also has many branches, is 5~7 kcal/mol larger than that from the conventional group additivity (GA) methods. Ten homodesmotic reactions [65] were constructed to calculate the standard enthalpy of formation of 2,2,3,3,4-pentamethylpentane. In a homodesmotic reaction, the numbers of each type of carbon-carbon bond (C_{sp3}-C_{sp3}), and the numbers of each type of carbon atom (sp^3) with zero, one, two, and three hydrogens attached are the same in reactants and products. Due to the cancellation of errors in computational method, this type of reaction is often used as a hypothetical reaction to calculate the enthalpy of formation. The ten reactions are listed as follows in Table S3. An extended table can be found in the supplemented material “homodesmotic reactions.xlsx”. The experimental data of all species other than 2,2,3,3,4-pentamethylpentane in the homodesmotic reactions are from the same source as Table S1 and are labeled with superscript. The reaction enthalpies were calculated under the CBS-QB3 level of theory. The rigid rotor harmonic oscillator (RRHO) approximation with corrections for hindered rotors were used during the calculation. Vibrations corresponding to torsions around all single bonds were separated out and treated as hindered rotors with the method suggested by Sharma et al. [66], which has been implemented in Mesmer 4.0 [67,68]. The internal rotor potentials were calculated at the B3LYP/6-31G(d) level of theory. Relaxed scan was performed for the torsion about every single bond between 0° and 360° with an incremental step of 10°. The structures were re-optimized when a lower energy conformer, relative to the initial low-energy conformer, was found. And then the internal rotor potentials were rescanned to make sure the updated conformer is the lowest-energy one. Besides the chemical equation of the ten homodesmotic reactions, the reaction enthalpy, which equals to the total enthalpy of products minus that of reactants, and the calculated enthalpy for 2,2,3,3,4-pentamethylpentane are also presented. The average of the calculated enthalpy value is -58.42 kcal/mol, and the unbiased sample standard deviation (SD) is 0.60 kcal/mol. This result suggests that the conventional GA prediction is about 4~6 kcal/mol lower. Comparing with the computed value -58.42 kcal/mol, the deviations of the distance-based group contribution prediction, NIST Structures and Properties, RMG, and THERM are 1.23 kcal/mol, -5.78 kcal/mol, -4.38 kcal/mol and -4.38 kcal/mol, respectively. The data in ref. [69] was originally -64.20 kcal/mol, and now the authors have updated it with our computed value after our private communication with them.

Table S3. Ten homodesmotic reactions used to calculate the standard enthalpy of formation for 2,2,3,3,4-pentamethylpentane

ID	***Reaction***	***Reaction enthalpy [kcal/mol]***	***Calculated enthalpy for 2,2,3,3,4-pentamethylpentane [kcal/mol]***
1		10.10	-59.00
2		7.30	-58.83
3		8.75	-58.10
4		8.13	-57.92
5		6.80	-58.97

6		5.55	-58.93
7		4.65	-58.23
8		6.76	-58.93
9		6.28	-57.26
10		5.29	-57.99

Reference

[1] P.J. Linstrom and W.G. Mallard, Eds., NIST Chemistry WebBook, NIST Standard Reference Database Number 69, National Institute of Standards and Technology, Gaithersburg MD, 20899, http://webbook.nist.gov, (retrieved April 23, 2016).

[2] M. Saeys, M.F. Reyniers, G.B. Marin, V. V. Speybroeck, M. Waroquier, J. Phys. Chem. A. 107 (2003) 9147-9159.

[3] C. Mosselman, H. Dekker, J. Chem. Soc. Faraday Trans. 1 (1975) 417-424.

[4] J.E. Connett, J. Chem. Thermodyn. 4 (1972) 233-237.

[5] D.R. Burgess, "Thermochemical Data" in NIST Chemistry WebBook, NIST Standard Reference Database Number 69, National Institute of Standards and Technology, Gaithersburg MD.

[6] I. Wadso, Acta Chem. Scand. 20 (1966) 544-552.

[7] H.Y. Afeefy, J.F. Liebman, and S.E. Stein, "Neutral Thermochemical Data" in NIST Chemistry WebBook, NIST Standard Reference Database Number 69, National Institute of Standards and Technology, Gaithersburg MD.

[8] J. Chao, F.D. Rossini, J. Chem. Eng. Data 10 (1965) 374-379.

[9] A. Snelson, H.A. Skinner, Trans. Faraday Soc. 57 (1961) 2125-2131.

[10] J.H.S. Green, Chem. Ind. (London) (1960) 1215-1216.

[11] S.L. Khursan, V.S. Martem'yanov, Russ. J. Phys. Chem. (Engl. Transl.) 65 (1991) 321-325.

[12] D.A. Pittam, G. Pilcher, J. Chem. Soc. Faraday Trans. 1 68 (1972) 2224-2229.

[13] E.J. Prosen, F.W. Maron, F.D. Rossini, J. Res. NBS 46 (1951) 106-112.

[14] H.A. Skinner, A. Snelson, Trans. Faraday Soc. 56 (1960) 1776-1783.

[15] H.A. Gundry, D. Harrop, A.J. Head, G.B. Lewis, J. Chem. Thermodyn. 1 (1969) 321-332.

[16] C. Mosselman, H. Dekker, Recl. Trav. Chim. Pays-Bas 88 (3) (1969) 257-265.

[17] J. Tjebbes, Acta Chem. Scand. 14 (1960) 180-188.

[18] G.S. Parks, Thermal data on organic compounds I. The heat capacities and free energies of methyl, ethyl and normal-butyl alcohols, J. Am. Chem. Soc. 47 (2) (1925) 338-345.

[19] J.E. Connett, J. Chem. Thermodyn. 7 (1975) 1159-1162.

[20] K.B. Wiberg, S. Hao, J. Org. Chem. 56 (1991) 5108-5110.

[21] W.D. Good, R.T. Moore, A.G. Osborn, D.R. Douslin, J. Chem. Thermodyn. 6 (1974) 303-310.

[22] A. Molnar, R. Rachford, G.V. Smith, R. Liu, Appl. Catal. 9 (1984) 219-223.

[23] K.B. Wiberg, D.J. Wasserman, E. Martin, J. Phys. Chem. 88 (1984) 3684-3688.

[24] W.D. Good, N.K. Smith, J. Chem. Thermodyn. 11 (1979) 111-118.

[25] D.W. Rogers, S. Skanupong, J. Phys. Chem. 78 (1974) 2569-2572.

[26] D.W. Rogers, F.J. McLafferty, Tetrahedron 27 (1971) 3765-3775.

[27] W.V. Steele, R.D. Chirico, J. Phys. Chem. Ref. Data 22 (1993) 377-430.

[28] K.W. Egger, S.W. Benson, J. Am. Chem. Soc. 88 (1966) 236-240.

[29] J. Coops, D. Mulder, J.W. Dienske, J. Smittenberg, Rec. Trav. Chim. Pays-Bas 65 (1946) 128.

[30] J.D. Cox, G. Pilcher, Thermochemistry of Organic and Organometallic Compounds, Academic Press, New York, 1970, p. 1-636.

[31] J.P. McCullough, R.E. Pennington, J.C. Smith, I.A. Hossenlopp, G. Waddington, J. Am. Chem. Soc. 81 (1959) 5880-5883.

[32] W.D. Good, J. Chem. Thermodyn. 2 (1970) 237-244.

[33] W.V. Steele, R.D. Chirico, A.B. Cowell, S.E. Knipmeyer, A. Nguyen, J. Chem. Eng. Data 42 (1997) 1052-1066

[34] R. Spitzer, H.M. Huffman, J. Am. Chem. Soc. 69 (1947) 211-213.

[35] E.J. Prosen, W.H. Johnson, F.D. Rossini, J. Res. NBS 37 (1946) 51-56.

[36] A. Molnar, R. Rachford, G.V. Smith, R. Liu, Appl. Catal. 9 (1984) 219-223.

[37] W.V. Steele, R.D. Chirico, J. Phys. Chem. Ref. Data 22 (1993) 377-430.

[38] K.B. Wiberg, D.J. Wasserman, J. Am. Chem. Soc. 103 (1981) 6563-6566.

[39] D.W. Rogers, O.A. Dagdagan, N.L. Allinger, J. Am. Chem. Soc. 101 (1979) 671-676.

[40] D.W. Rogers, P.M. Papadimetriou, N.A. Siddiqui, Mikrochim. Acta 2 (1975) 389-400.

[41] E. Bretschneider, D.W. Rogers, Mikrochim. Acta. 58 (3) (1970) 482-490.

[42] H.A. Skinner, A. Snelson, Trans. Faraday Soc. 55 (1959) 405-407.

[43] T. Flitcroft, H.A. Skinner, M.C. Whiting, Trans. Faraday Soc. 53 (1957) 784-790.

[44] W.D. Good, N.K. Smith, J. Chem. Eng. Data 14 (1969) 102-106.

[45] R. Fuchs, J.H. Hallman, Can. J. Chem. 61 (1983) 503-505.

[46] D.W. Rogers, E. Crooks, K. Dejroongruang, J. Chem. Thermodyn. 19 (1987) 1209-1215.

[47] E.J. Prosen, F.D. Rossini, J. Res. NBS 34 (1945) 263-267.

[48] J.D. Rockenfeller, F.D. Rossini, J. Phys. Chem. 65 (1961) 267-272.

[49] W.H. Johnson, E.J. Prosen, F.D. Rossini, J. Res. NBS 42 (1949) 251-255.

[50] K.B. Wiberg, E.C. Jr. Lupton, D.J. Wasserman, A. de Meijere, S.R. Kass, J. Am. Chem. Soc. 106 (1984) 1740-1744.

[51] E.C. Jr. Lupton, Diss. Abs., 31 (1970) 1174-1175.

[52] E. Bretschneider, D.W. Rogers, Mikrochim. Acta 58 (3) (1970) 482-490.

[53] K. Yates, R.S. McDonald, J. Org. Chem. 38 (1973) 2465-2478.

[54] W.D. Good, J. Chem. Thermodynamics 4 (5) (1972) 709-714.

[55] J. Zheng, T. Yu, E. Papajak, I, M. Alecu, S. L. Mielke, and D. G. Truhlar, *Phys. Chem. Chem. Phys.*, 2011, 13, 10885-10907.

[56] J. Zheng and D. G. Truhlar, *J. Chem. Theory Comput.*, 2013, 9, 1356-1367.

[57] J. Zheng, S. L. Mielke, K. L. Clarkson, and D. G. Truhlar, *Comput. Phys. Commun.*, 2012, 183, 1803-1812.

[58] J. Zheng, R. Meana-Pañeda, and D. G. Truhlar, *Comput. Phys. Commun.*, 2013, 184, 2032-2033.

[59] T. H. Lay, J. W. Bozzelli, A. M. Dean, E. R. Ritter, J. Phys. Chem. A., 1995, 99, 14514-14527.

[60] M.A. Blanco, A.M. Pendás, E. Francisco, J. Chem. Theory Comput. 1 (2005) 1096-1109.

[61] AIMAll (Version 16.01.09), Todd A. Keith, TK Gristmill Software, Overland Park KS, USA, 2016 (aim.tkgristmill.com).

[62] R.F.W. Bader, Chem. Rev. 91 (1991) 893-928.

[63] P.L.A. Popelier, Atoms in Molecules: An introduction, Pearson Education, 1999.

[64] R.F.W. Bader, Atoms in Molecules, Clarendon Press, Oxford, 1990.

[65] S.E. Wheeler, K.N. Houk, P.R. Schleyer, W.D. Allen, J. Am. Chem. Soc. 131 (2009) 2547-2560.

[66] S. Sharma, S. Ramans, W.H. Green, J. Phys. Chem. A. 114 (2010) 5689–5701.

[67] D.R. Glowacki, C.H. Liang, C. Morley, M.J. Pilling, S.H. Robertson, J. Phys. Chem. A. 116 (2012) 9545–9560.

[68] S.H. Robertson, D.R. Glowacki, C.-H. Liang, C. Morley, R. Shannon, M. Blitz, A. Tomlin, P.W. Seakins, M.J. Pilling, MESMER (Master Equation Solver for Multi-Energy Well Reactions), 2008-2013; an object oriented C++ program implementing master equation methods for gas phase reactions with arbitrary multiple wells. http://sourceforge.net/projects/mesmer.

[69] E. Goos, A. Burcat, B. Ruscic, Extended Third Millennium Ideal Gas and Condensed Phase Thermochemical Database for Combustion with Updates from Active Thermochemical Tables, http://garfield.chem.elte.hu/Burcat/burcat.html, accessed April 23, 2016.

[70] J.A. Manion, J. Phys. Chem. Ref. Data 31 (2002) 123-172.