Highlights

Cascaded group-additivity ONIOM: A new method to approach CCSD(T)/CBS energies of large aliphatic hydrocarbons

Junjun Wu [a], Hongbo Ning [a,b], Liuhao Ma [a], Peng Zhang [c], and Wei Ren [a,b,*]

[a] *Department of Mechanical and Automation Engineering, The Chinese University of Hong Kong, New Territories, Hong Kong*

[b] *Shenzhen Research Institute, The Chinese University of Hong Kong, New Territories, Hong Kong*

[c] *Department of Mechanical Engineering, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong*

[*] Corresponding author. Fax: +852 2603 6002. Email: renwei@mae.cuhk.edu.hk (W. Ren)

1

Abstract: We report a cascaded group-additivity (CGA) ONIOM method for high-level energy calculations of large aliphatic hydrocarbon molecules by combining the group additivity and two-layer ONIOM methods. This hybrid method is implemented by partitioning the target molecule into individual groups, which are cascaded via the overlapping between them. The energy of the entire molecule is first calculated at a low level of theory such as M06-2x/cc-pVTZ. Then all the groups and their overlappings are treated at the levels of CCSD(T)/CBS and M06-2x/cc-pVTZ to obtain their energy difference to be used as the energy correction. We selected small-to-middle size aliphatic hydrocarbons including 79 C_4-C_8 molecules as the validation set to demonstrate the feasibility of the CGA-ONIOM method, followed by the calculations of 12 representative C_{10}, C_{12} and C_{16} aliphatic hydrocarbons (including normal-, branched-, cyclo- and unsaturated categories). Our calculations agree well with the reference values available in the literature with the modest deviation around 1.0 kcal·mol^{-1}. Compared with conventional CCSD(T)/CBS calculation of the whole molecule, the computational cost can be dramatically reduced by a factor of ~10^2 for molecules with 10 carbons and ~10^4 for molecules with 16 carbons. Considering its outstanding computational efficiency and accuracy, our proposed CGA-ONIOM method is promising for combustion chemistry studies of large fuel molecules at a high level of theory.

Keywords: ONIOM, group additivity, CCSD(T), large molecule, enthalpy of formation

1. Introduction

Accurate knowledge of energies of fuel molecules is the prerequisite for developing reliable thermodynamic and kinetic models in combustion studies. Energies calculated at the CCSD(T) complete basis set (CBS) level of theory is treated as the gold standard [1]. However, it is computationally unfeasible or practically even impossible to directly apply CCSD(T)/CBS calculations for large fuel molecules with more than ten heavy atoms. Note that real fuels consist of various large fuel molecules with up to twenty heavy atoms [2–5]. As a result, possible solutions are in urgent need to obtain accurate thermodynamic and chemical information of these large fuel molecules with a reasonable computational cost.

Many empirical and theoretical efforts have been made in the past decades to advance the computational methods. In early times, Benson et al. [6] proposed the group additivity (GA) method to estimate the energies of large molecules by summarizing the functional groups of known thermodynamics. This successful method is widely used in many modern calculations and demonstrates good accuracy and computational efficiency. However, GA may be not reliable enough to predict the thermochemical properties of complex molecules, where interactions between the nearing functional groups become significant [7,8]. Many other methods based on Kohn-Sham density functional theory (DFT) were developed to compensate the computational cost and accuracy [9,10]. Despite the success of approaching the accuracy at the CCSD(T) level, the DFT method still requires further improvements in the overall accuracy [11].

In comparison, ONIOM (our own n-layered integrated molecular orbital and molecular mechanics) has been historically developed by Morokuma and co-workers [12], which divides the whole system into different layers and computes these layers at different levels of theory. Many efforts have been made to explore the ONIOM method applied in different research areas like organic system [13,14], catalysis [15–17], and bio-molecular system [18–

20]. Until recently, the ONIOM method has been used in combustion science [21] to determine bond dissociation energies [13,22] and energy barriers of hydrogen abstraction reactions [23]. A two-layer ONIOM method is commonly adopted in these studies, which treats the whole system (low layer) at a low level of theory and the selected part of the system (high layer) at a high level of theory. Such a configuration dramatically reduces the computational cost. When implementing the two-layer ONIOM method, only the high layer is treated at a high level of theory, for example, the CCSD(T) method with a large basis set. Thus the ONIOM method features locally accurate energies [24] and can be used to predict accurate energy differences like BDE or energy barrier. However, the accurate prediction of energy for the whole molecule is also desired for combustion chemistry research, while the intrinsic local accuracy of ONIOM in energy prediction impedes its further applications in combustion.

High-level energy calculations of large molecules at the level of CCSD(T) is in urgent need but still unavailable. In this work, we present a new method to obtain energies of large fuel molecules at the CCSD(T)/CBS level by combining the group additivity strategy and the ONIOM method, namely cascaded group-additivity ONIOM (CGA-ONIOM). Compared with the conventional GA method, CGA-ONIOM accounts for the interaction effects from the neighboring functional groups enabled by the feature of ONIOM [24]. Compared with the conventional ONIOM method, CGA-ONIOM can approach the globally accurate energy (rather than the locally accurate energy) enabled by the additional GA method. As a proof of principle, the enthalpy of formation values of various important aliphatic C_4-C_8 hydrocarbon molecules were calculated using the proposed new method and compared with the data available in the literature. Note that the enthalpy of formation is preferred since it is the most important thermochemical property [7] and also the most straightforward indicator to check the accuracy of *ab initio* calculations. Herein, the CGA-ONIOM method was further

4

implemented for calculating large C_{10}, C_{12} and C_{16} hydrocarbons, which are typical components in jet fuels and diesel [2], to demonstrate the merits of the proposed method. The CGA-ONIOM method is proved to be an accurate and computationally efficient method for combustion chemistry research.

2. Methods

2.1 Principle of CGA-ONIOM

It is necessary to briefly review the conventional ONIOM method before we introduce the CGA-ONIOM. The ONIOM method treats the entire molecule at different levels of theory. For instance, the molecular system is divided into the model system and environment system for the two-layer ONIOM method. The model system (high layer), usually the chemically important portion, is treated by the accurate and expensive electronic structure method; whereas the environment system (low layer) is treated by the less-accurate but more-efficient method. Note that at the layer boundary the H atom (adopted in this work) or CH_3 group can be used as the linkage to saturate the dangling bonds. Therefore, the ONIOM calculation features locally accurate energy E^M as:

$$E^M = E_{low}^M\left(Real\right) + E_{high}\left(Model\right) - E_{low}\left(Model\right). \tag{1}$$

where E^M is the total energy of the whole ("real") molecule, which is a hybrid energy containing both low-level and high-level components; $E_{low}^M\left(Real\right)$ is the total low-level energy of the whole molecule, E_{high} (*Model*) and E_{low} (*Model*) are the high-level and low-level energies of the key chemically important portion ("model").

In this work, to approach the accuracy of the CCSD(T) complete basis set, we computed the high-level energy of the model system at both the CCSD(T)/cc-pVTZ [25] and CCSD(T)/cc-pVQZ [25] levels and extrapolated them to the complete basis using the scheme [26]:

$$E\left(l_{\max}\right) = E_{\infty} + \frac{B}{\left(l_{\max}+1\right)^{4}}, \qquad (2)$$

where $l_{\max}$ is the maximum angular momentum function within the two basis sets ($l_{\max}= 3$ for cc-pVTZ basis set and $l_{\max}= 4$ for the cc-pVQZ basis set); B is the system-specific parameter; and E_{∞} is the energy extrapolated to the complete basis set, denoted as $E_{CCSD(T)/CBS}$. The low-level energies were determined by the M06-2x method [27] with the basis set cc-pVTZ [25]. Note that in principle other methods such as MP2 or B3LYP with reasonably large basis sets are also applicable.

In Equation (1), the last two terms are treated as the energy difference of a model system n determined at two theory levels, i.e. CCSD(T)/CBS and M06-2x/cc-pVTZ. Hence, the locally accurate energy of the whole molecule is expressed as:

$$E^{M} = E_{low}^{M}\left(real\right) + \Delta E_{CCSD(T)/CBS}(Model_{n}), \qquad (3)$$

Equation (3) lays the foundation of the CGA-ONIOM method. Conceptually, the whole molecule can be deemed as the assembly of a series of small model systems, which are named "group" in this study as the model system often refers to a single chemically important portion of the molecule. With the energies of these groups accurately determined, one can expect to approach the globally accurate energy of the entire molecule. In the proposed CGA-ONIOM method, the fuel molecule is first partitioned into different groups, which are cascaded via the overlapping between them. All these groups and overlappings are taken as the individual model system, and thus a set of ONIOM calculations are implemented with the energy differences of the groups and overlappings determined by Equation (3). Eventually, the total energy of the molecule is obtained by counting the energy differences of all the groups and their overlapping parts:

$$E_{CCSD(T)/CBS}^{M} = E_{low}^{M}\left(Real\right) + \sum_{n=1}^{N}\Delta E_{CCSD(T)/CBS}\left(GP_{n}\right) - \sum_{n=1}^{N-1}\Delta E_{CCSD(T)/CBS}\left(GP_{n}\cap GP_{n+1}\right), \quad (4)$$

where N is the total number of the groups and GP denotes the group (black outlines in Figure 1); $\Delta E\left(GP_n \cap GP_{n+1}\right)$ represents the energy difference due to the overlapping of two adjacent groups (red regions in Figure 1). Such an overlapping cascades all the functional groups and is indispensable in the CGA-ONIOM method.
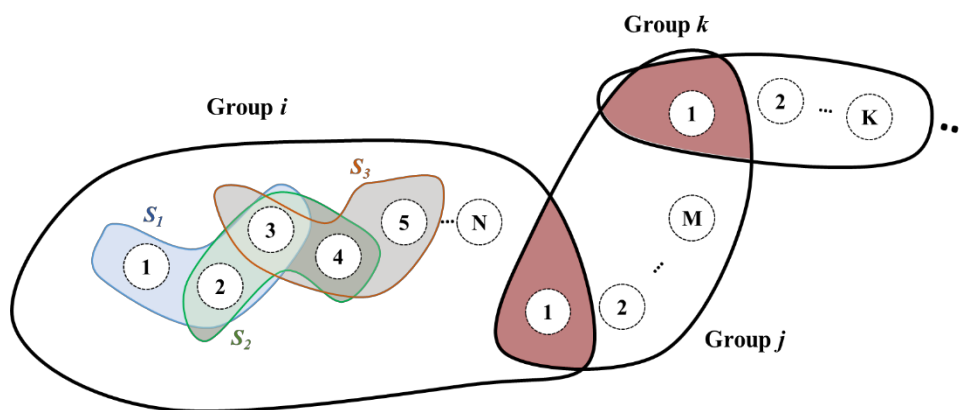

Figure 1. Schematic of the cascaded group-additivity ONIOM method applied for calculating the energies of large fuel molecules. Dotted circle, the heavy atoms in the molecule; shaded region S_i, the angle scheme of three heavy atoms inside the group; red shaded region, the overlapping of adjacent groups.

The CGA-ONIOM is seemingly similar to the many-body expansion method [28] (especially the two-body truncation method) as both methods are based on (1) the summation of energy difference of each group, and (2) the subtraction of interaction energies caused by the groups. However, CGA-ONIOM differs from the conventional many-body expansion method. Particularly, CGA-ONIOM only considers the overlapping or interaction of adjacent groups (See Figure 1), while many-body expansion method considers the interactions of any two groups or three groups according to the combination algorithm [28].

2.2 Feasibility of CGA-ONIOM

We select n-hexane, a middle-sized hydrocarbon, as the example to illustrate the feasibility of the summation strategy in CGA-ONIOM by comparing the energy by the CGA-ONIOM method to that obtained by the CCSD(T) method.
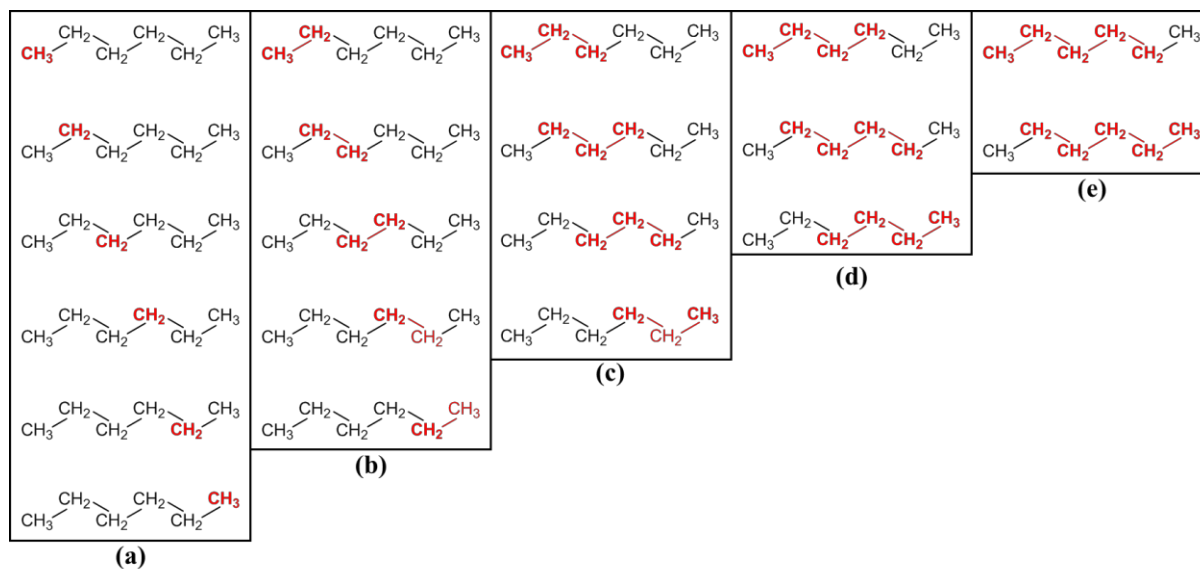
Figure 2. Representative schemes to partition n-hexane into different groups: (a) point scheme; (b) dumbbell scheme, (c) angle scheme, (d) N-shape scheme and (e) M-shape scheme.

Figure 2 illustrates the five representative schemes available to implement CGA-ONIOM calculations. Scheme (a) takes CH_3— and —CH_2— as the basic groups without any overlapping; scheme (b) takes CH_3CH_2— and —$(CH_2)_2$— as the basic groups with —CH_2— overlapping; scheme (c) takes $CH_3(CH_2)_2$— and —$(CH_2)_3$— as the basic groups with —$(CH_2)_2$— overlapping; scheme (d) takes $CH_3(CH_2)_3$— and —$(CH_2)_4$— as the basic groups with —$(CH_2)_3$— overlapping; and scheme (e) takes $CH_3(CH_2)_4$— as the basic group with —$(CH_2)_4$— overlapping. For convenience, we call these schemes as point scheme, dumbbell scheme, angle scheme, N-shape scheme, and M-shape scheme, respectively.

Table 1 compares the energies calculated by the CGA-ONIOM method using different partitioning schemes with the exact energy of the entire molecule. The point scheme fails to

predict the energy with a very large deviation of 14.3 kcal·mol^{-1}, which is mainly because this scheme ignores the neighboring bonded interaction. In comparison, the dumbbell scheme and angle scheme give relatively close results with a small deviation of 0.4 and 0.3 kcal·mol^{-1}, respectively. Additionally, the N-shape scheme consisting of four carbon atoms performs even better with a difference of 0.1 kcal·mol^{-1}, and the M-shape scheme with five carbon atoms performs the best with a negligible difference of -0.002 kcal·mol^{-1}. Hence, the summation strategy adopted by CGA-ONIOM is feasible except the point scheme. The N/M-shape scheme is more accurate but incurs more computational cost. In the practical application, the dumbbell scheme and angle scheme are desirable as they allow quite accurate and efficient predictions with a deviation of < 1.0 kcal·mol^{-1}.

Table 1 Systematical comparison of the energy of n-hexane calculated by CGA-ONIOM with the conventional quantum chemistry calculation at the CCSD(T)/cc-pVDZ level of theory.

	E (Hartree)	Truncation Error (kcal·mol^{-1})
E_{tot} (CCSD(T)/cc-pVDZ)	-236.3747416	-
E_1 (Point Scheme)	-236.3519799	14.3
E_2 (Dumbbell Scheme)	-236.3741433	0.4
E_3 (Angle Scheme)	-236.3742458	0.3
E_4 (N-shape Scheme)	-236.3746608	0.1
E_5 (M-shape Scheme)	-236.3747445	0.0 (-0.002)

2.3 Neighboring Interaction & Selection of Partitioning Scheme

Regarding the neighboring interaction, He et al. [7] suggested that the neighboring interaction can be approximated by the exponential relationship $\exp(-3d_{ij})$, where d_{ij} is the distance between the group i and j. This relationship indicates an asymptotic damping of the neighboring interactions with the increased distance (see Figure S1 in Supporting

Information). Herein, we still use n-hexane as the example to investigate the neighboring interaction among the functional groups.

Figure 3 shows the intramolecular distance between the center atom (C_3) and the rest atoms at M06-2x/cc-pVTZ. The neighboring interactions with the C_3 atom, either bonded or non-bonded, were estimated with the relationship suggested by He et al. [7]. Particularly, the neighboring interaction of 1.0 Å was taken as the reference to scale the intramolecular interactions, i.e., $\exp(-3d_{ij})/\exp(-3)$. We presume that the neighboring interaction with C_3 is summable and thus evaluate the performance of different partitioning schemes. Note that all the neighboring interactions with C_3 are summarized and defined as the total interaction.
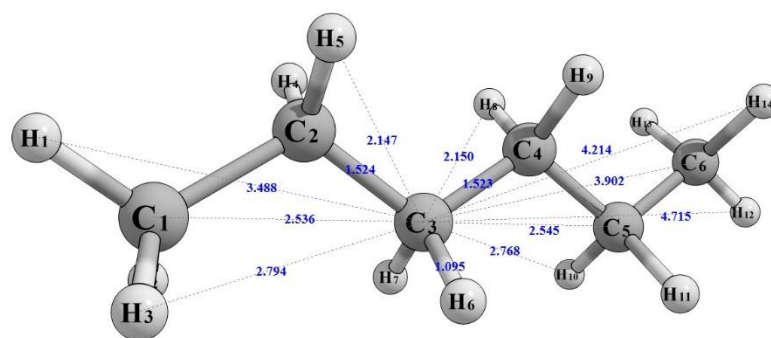
Figure 3. Intramolecular distance of n-hexane with C_3 as the center at the M06-2x/cc-pVTZ level of theory.

Figure 4 compares the neighboring interactions including bonded and non-bonded interactions obtained by different partitioning schemes with the total interactions (more details in Table S1 of Supporting Information). The neighboring interaction is almost governed by the bonded interactions with a significant quota of 92.0%. Among the five schemes, the angle, N-shape and M-shape schemes are capable of predicting the bonded interactions properly. Regarding the non-bonded interactions, the dumbbell, angle, N-shape and M-shape schemes can predict 38.1%, 76.2%, 88.0%, 99.8% of the overall non-bonded interactions. Note that the non-bonded interactions contribute only 8% to the overall neighboring interaction in this case. This fact thus indicates the possibility of avoiding the use

of M-shape scheme that requires more computational effort. Therefore, a smaller partitioning scheme that retains both the accuracy and efficiency is more attractive.

We further compare the overall neighboring interaction determined by the five schemes. The angle, N-shape and M-shape schemes are found with very good performance, predicting almost identical neighboring interactions to the total interactions by a factor of 0.98, 0.99 and 0.999, respectively. Thus we believe the angle scheme is sufficient to account for the bonded and non-bonded interactions for most fuel molecules. Indeed, Table 1 shows that the energy predicted by the angle scheme only remains a small deviation of 0.3 kcal·mol^{-1} compared to the real energy at CCSD(T)/cc-pVDZ.
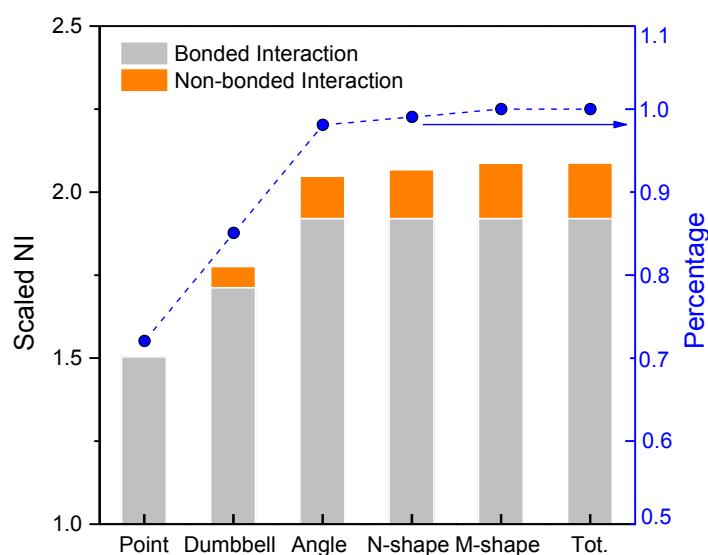
Figure 4. Scaled neighboring interaction (NI) to the C_3 atom in n-hexane with different partitioning schemes. The percentage is obtained by taking ratio of NI of different partitioning scheme to the total interactions.

It should be pointed out that for molecules with very strong non-bonded interactions, i.e., oxygenates that involves H-bonds, the M-shape scheme or large scheme is recommended to implement CGA-ONIOM calculations. This is supported by Hua et al. [29] that a critical distance of 3.0~4.0 Å to the centered functional group is suggested for considering the

neighboring interaction. Another study by Chi et al.[30] further verified this idea that a critical size of 3.5 Å was adopted to successfully predict the energies of a series of ester molecules. Therefore, the M-shape scheme should be a good candidate to treat the large fuel molecules, especially those with very strong non-bonded interactions.

In this work, we prefer the angle scheme as the generalized basis for all the CGA-ONIOM calculations of aliphatic hydrocarbons due to its acceptable accuracy and higher efficiency. In general, one can directly divide the molecule into angle groups and determine the energy differences of all the groups and their overlappings. Note that one should be very careful in partitioning the large polycyclic molecule manually with various combinations of angle schemes to avoid possible over-counting or under-counting. Hence, we suggest an alternative two-step process that the polycyclic molecule is firstly partitioned into cyclic functional groups, followed by a further step of partitioning of the cyclic functional group into the final angle schemes. The energy difference of each functional group is given by:

$$\Delta E_{CCSD(T)/CBS}(GP) = \sum_{i=1}^{I}\left\{E_{CCSD(T)/CBS}(AS_i) - E_{low}(AS_i)\right\} - \sum_{j=0}^{J}\left\{E_{CCSD(T)/CBS}(BS_j) - E_{low}(BS_j)\right\}, \quad (5)$$

where I is the total number of the angle arrays, J is the total number of overlapping bonds in each functional group, and $E(AS)$ is the energy corresponding to the angle scheme. Additionally, $E(BS)$ is the energy of the overlapping bond that is be determined by:

$$E(BS) = E(AS_i \cap AS_j). \quad (6)$$

It should be emphasized that the two-step process is an alternative method for partitioning polycyclic molecules (JP-10 in this work) and is unnecessary for simple molecules. For simple molecules ($I = 1$, $J = 0$), Equation (5) is reduced as:

$$\Delta E(GP) = \Delta E(AS), \quad (7)$$

Finally, based on Equations (1) to (7), we obtain the generalized energy of the entire fuel molecule at the CCSD(T)/CBS level as:

$$E_{CCSD(T)/CBS}^{M} = E_{low}^{M}\left(Real\right) + \sum_{n=1}^{N}\sum_{i=1}^{I}\left\{E_{CCSD(T)/CBS}\left(AS_{n,i}\right) - E_{low}\left(AS_{n,i}\right)\right\}$$

$$-\sum_{n=1}^{N}\sum_{j=0}^{J}\left\{E_{CCSD(T)/CBS}\left(BS_{n,j}\right) - E_{low}\left(BS_{n,j}\right)\right\} - \sum_{n=1}^{N-1}\Delta E_{CCSD(T)/CBS}\left(GP_n \cap GP_{n+1}\right).$$

(8)

2.4 Angle Scheme Allocation

For the long-chain aliphatic hydrocarbons, the CGA-ONIOM calculation with angle-scheme is straightforward to implement like the example n-hexane. However, for molecules with more complicated structures such as the branched or polycyclic alkanes, there exist several ways to allocate the angles inside the molecule. As a result, different allocation methods may yield different results. It is thus necessary to assess the effect of angle scheme allocation on the final energy before implementing the CGA-ONIOM calculation. In light of the conventional GA method, the typical functional groups in aliphatic hydrocarbons include C/C/H3, C/C2/H2, C/C3, C/C4, etc. [31]. Therefore, we choose 2-methyl-pentane and 2,2-dimethyl-butane as the examples to assess the effect of angle scheme allocations on CGA-ONIOM energy calculations. Note that these two molecules are preferred since they are relatively small in size and retain the mentioned nominal groups.
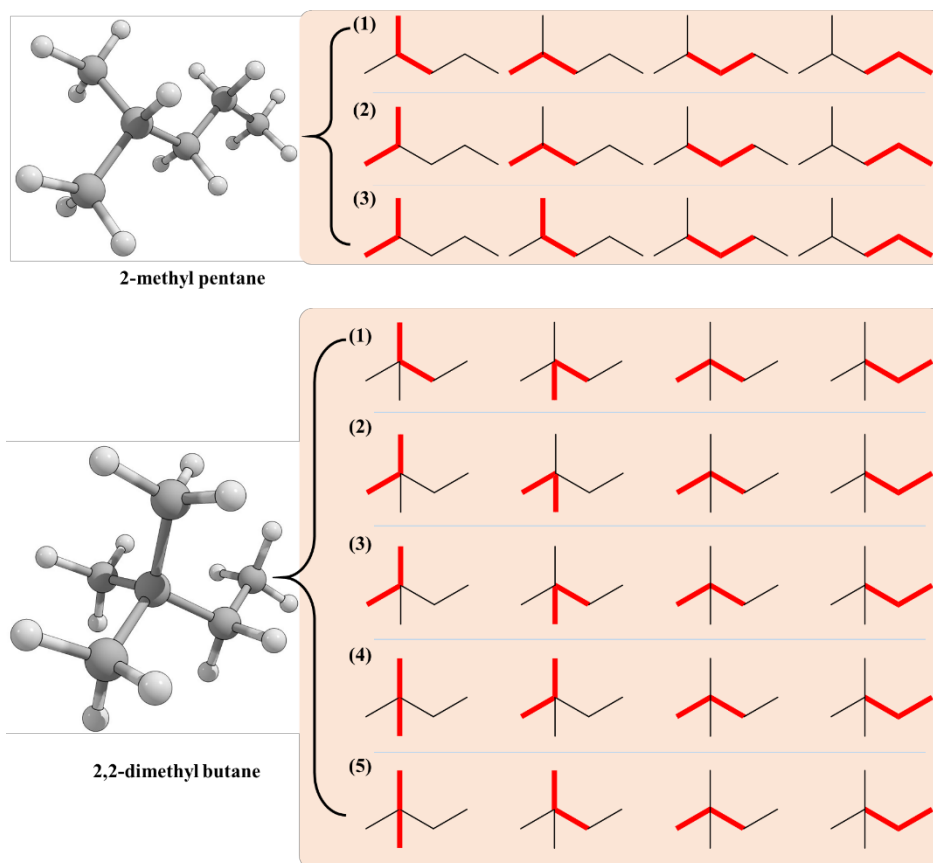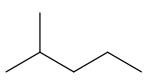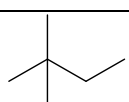
Figure 5. Representative angle scheme allocations of 2-methyl-butane and 2,2-dimethyl-butane for CGA-ONIOM calculation. The red bold bonds indicate the angle selected. Each selected angle is treated as the high layer and the remainder is treated as the low layer.

Table 2 CGA-ONIOM-determined electronic energies of 2-methyl pentane and 2,2-dimethyl butane with different angle scheme allocations. The high layer is treated by the CCSD(T) method with the cc-pVXZ (X=D, T, Q) basis set, and the low layer is treated by the M06-2x/cc-pVTZ method.

Scheme	CCSD(T) calculation for high layer (in Hartree)			
	cc-pVDZ	cc-pVTZ	cc-pVQZ	CBS(T-Q)
1	-236.37527	-236.63409	-236.70485	-236.75393
2	-236.37462	-236.63407	-236.70483	-236.75392
3	-236.37556	-236.63410	-236.70486	-236.75395
Conventional Calc.	-236.37603	-236.63460	-236.70539	-236.75451
Mean truncation error [a] (kcal·mol^{-1})	**0.552**	**0.323**	**0.345**	**0.360**
Standard deviation (kcal·mol^{-1})	**0.302**	**0.010**	**0.010**	**0.010**
1	-236.37738	-236.63670	-236.70747	-236.75657
2	-236.37811	-236.63680	-236.70750	-236.75655

14

3	-236.37741	-236.63675	-236.70749	-236.75656
4	-236.37807	-236.63673	-236.70747	-236.75655
5	-236.37842	-236.63674	-236.70748	-236.75656
Conventional Calc.	-236.37871	-236.63736	-236.7081	-236.75717
Mean Truncation Error [a] (kcal·mol^{-1})	**0.520**	**0.390**	**0.384**	**0.381**
Standard Deviation (kcal·mol^{-1})	**0.290**	**0.023**	**0.008**	**0.005**

[a] Truncation error is obtained by taking the difference of the conventional CCSD(T) calculation and the CGA-ONIOM calculation.
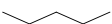
Figure 5 depicts the possible angle scheme allocations of 2-methyl-pentane and 2,2-dimethyl-butane available for CGA-ONIOM calculations. The high layer (red bold angle) was treated at the CCSD(T) method with cc-pVDZ, cc-pVTZ and cc-pVQZ basis sets, respectively; the low layer was treated at the M06-2x/cc-pVTZ level of theory. We also calculated the electronic energies of the entire molecules. The truncation errors and the standard deviations of the CGA-ONIOM energies determined by different angle scheme allocations are provided in Table 2. The CGA-ONIOM method with the high layer determined by CCSD(T)/cc-pVDZ gives a mean truncation error of ~0.5 kcal·mol^{-1}, which is reduced with the high layer determined at CCSD(T)/cc-pVTZ or higher theory levels. The standard deviation caused by the angle scheme allocation is reduced by improving the theory level from CCSD(T)/cc-pVDZ to CCSD(T)/CBS. Particularly, the standard deviation is only 0.010 kcal·mol^{-1} for 2-methyl pentane and 0.005 kcal·mol^{-1} for 2,2-dimethyl butane if the high layer treated at the CCSD(T)/CBS level of theory. Such a small deviation in energy is trivial, and thus one can freely divide the large molecule into angles as long as the dumbbell overlapping is guaranteed. Additionally, it is observed that 2-methyl pentane with one branch treated by CGA-ONIOM at CCSD(T)/CBS gives a truncation error of 0.360 kcal·mol^{-1}, which is increased to 0.381 kcal·mol^{-1} for 2,2-dimethyl butane with two branches. It indicates the truncation error increases by ~0.02 kcal·mol^{-1} per additional

15

branch. Such a rough estimation tells that the angle scheme allocation contributes minor to the truncation error, even for highly-branched or polycyclic alkanes. In fact, the group size in the CGA-ONIOM calculation is the major source of truncation error. Note that the truncation error could be reduced by adopting a larger partitioning scheme or avoided by using the method of error cancelling balanced reaction [32].

2.5 Enthalpy of Formation

Ideally, the performance of CGA-ONIOM should be assessed by directly comparing the CGA-ONIOM-determined energies with the CCSD(T) calculations. Herein, we first compared the electronic energies of six medium-size molecules (n-butane, n-pentane, n-hexane, cyclo-butane, cyclo-pentane and cyclo-hexane) for a perception of the possible error by CGA-ONIOM. Table 3 compares the CCSD(T)/CBS energies of these alkanes determined by the CGA-ONIOM and conventional calculations. The truncation error increases slightly from ~0.1 kcal·mol^{-1} for C_4 alkanes to ~0.3 kcal·mol^{-1} for C_6 alkanes, indicating a rough estimation of error increasing rate of 0.1 kcal·mol^{-1} per heavy atom. One may expect a moderate truncation error of ~1.3 kcal·mol^{-1} for C_{16} aliphatic hydrocarbons. Again, it should be emphasized that the truncation error can be reduced by adopting larger partitioning schemes (N/M scheme or larger), which is not pursed in this work.

Table 3 Comparison of the electronic energies of C_4-C_6 alkanes determined by CGA-ONIOM and conventional CCSD(T)/CBS calculations.

Molecule	Structure	CCSD(T)/CBS energy (in Hartree)		Truncation Error (in kcal·mol^{-1})
		CGA-ONIOM	Convention	
n-butane		-158.2341140	-158.2342499	0.085
n-pentane		-197.4932907	-197.4936239	0.209
n-hexane		-236.7524817	-236.7530358	0.348
cyclo-butane		-156.9916839	-156.9914486	0.148

16

| cyclo-pentane | | -196.2834027 | -196.2829681 | 0.273 |
| cyclo-hexane | | -235.5516917 | -235.552361 | 0.420 |

However, it becomes extremely computationally expensive to conduct large-scale CCSD(T)/CBS calculations of large-size molecules. As a compromise, we selected the enthalpy of formation (EOF) as a measure to assess the performance of CGA-ONIOM in energy calculations. Note that EOF is one of the most important thermochemical properties and has also been intensively adopted to assess the accuracy of electronic structure method [33].

In the course of EOF calculation, the target molecule was optimized using Gaussian 09 [34] at the M06-2x/cc-pVTZ level of theory and confirmed without imaginary frequencies, followed by the application of CGA-ONIOM for energy determination. The EOF was determined using two methods: one is to take atoms (^{3}C, H) as the reference species (ΔH_f^A) and the other is to take molecules (CH_4, H_2) as the reference species (ΔH_f^M) [35]. The EOF depends not only on the electronic energy but also the zero-point vibrational energy (ZPVE) [35], and the scaling factor for ZPVE is 0.955±0.019 [36] as recommended by CCCBDB [36]. Note that the existing uncertainty in the ZPVE may contribute to the uncertainty in EOF. In fact, Truhlar et al. [37] suggested a scaling factor of ~0.971 for the ZPVE correction, slightly different from the 0.955 but still within its uncertainty range. To reduce the uncertainty caused by ZPVE, we computed the EOF with both scaling factors and took the average value. Finally, the EOF (ΔH_f) is determined by:

$$\Delta H_f = \frac{1}{2}\left(\Delta H_f^A + \Delta H_f^M\right), \tag{9}$$

with the corresponding error ε defined as:

$$\varepsilon = \pm\left|\Delta H_f - \Delta H_f^{A/M}\right|. \tag{10}$$

Another uncertainty source of EOF is the multiple conformers of long-chain molecules [7,38]. Herein, we only consider the single-structure EOF, which may incurs an uncertainty of < 1.0 kcal·mol^{-1} in EOF, as indicated by He et al. [7].

3. Results and Discussion

3.1 Validation of CGA-ONIOM on C$_4$-C$_8$ Hydrocarbons

We first systematically calculated the EOFs of 38 alkanes and 41 alkenes ranging from C$_4$ to C$_8$ at both 0 K and 298 K; the full set of EOFs as well as a sample implementation of CGA-ONIOM are provided in Supporting Information. Figure 6 compares the full set of the calculated EOFs with the reference values provided by NIST database [39], Active Thermochemical Tables (ATcT) [35,40,41], and Thermochemical Data of Organic Compounds (TDOC) [42]. Note that NIST [39] and TDOC [42] only provides EOF values at 298 K, and ATcT [35,40,41] provides EOFs of small molecules at both 0 K and 298 K. Thus the test set includes 13 EOFs at 0 K and 79 EOFs at 298 K, totaling 92 EOFs. It is interesting to see that our calculations present a very good agreement with all the literature values, suggesting the accuracy and reliability of the CGA-ONIOM method.
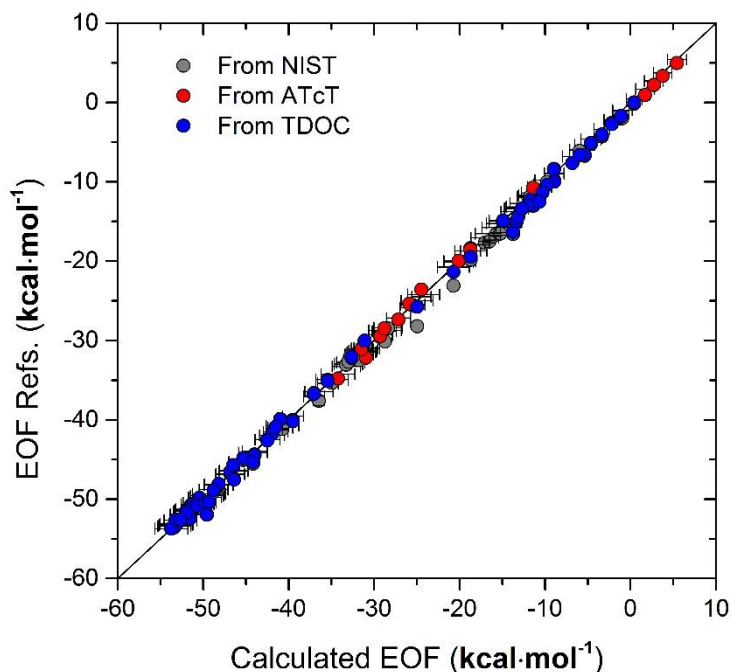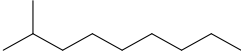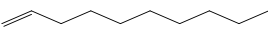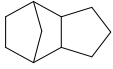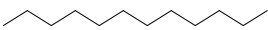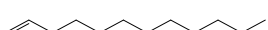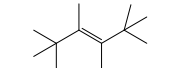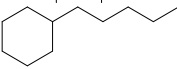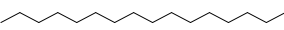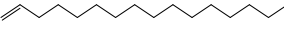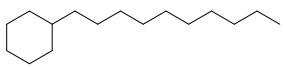
Figure 6. Comparison of the calculated EOFs by the CGA-ONIOM method with the reference values from NIST [39], AtcT [35,40,41] and TDOC [42] database. The horizontal error bar indicates the error in CGA-ONIOM calculations.

Table 4 lists EOFs of a series of representative aliphatic hydrocarbons calculated by CGA-ONIOM. Most of the calculated EOFs agree well with the reference values, only with a small deviation of < 1.0 kcal·mol^{-1}. Interestingly, we observe certain discrepancies in EOFs between the NIST and TDOC databases. For instance, the EOFs for Z-2,2-dimethyl-3-hexene (Z22D3H) and E-2,2,-Dimethyl-3-hexene (E22D3H) in NIST are -23.1 and -28.2 kcal·mol^{-1}, respectively. However, the corresponding experimental values in TDOC are -21.3 and -25.7 kcal·mol^{-1}, respectively. Our calculated EOFs of the two molecules are much closer to the TDOC values (-20.7 kcal·mol^{-1} for Z-22D3H and -25.0 kcal·mol^{-1} for E-22D3H).

Table 4. EOFs of a series of representative aliphatic hydrocarbons calculated by CGA-ONIOM. The reference values are taken from NIST [39], ATcT [35,40,41] and TDOC [42] databases.

Name	Structure	This work	NIST	ATcT	TDOC
Butane		-31.1±1.0	-30.0,-30.4	-30.1	-30.0

Compound				
Iso-butane	-32.6±1.0	-32.1,-32.4	-32.2	-32.1
1-butene	0.4±0.9	-0.2	-	0.0
Trans-2-butene	-2.2±0.9	-2.6	-2.7	-2.7
Cis-2-butene	-1.1±0.9	-1.8	-1.7	-1.7
Pentane	-35.4±1.2	-35.0,-35.1,-35.2	-35.0	-35.1
2-methyl-pentane	-37.0±1.2	-36.7,-36.8,-36.9	-36.6	-36.7
Neo-pentane	-39.5±1.2	-40.1,-40.3,-39.7	-40.0	-40.2
Cyclopentane	-18.7±1.2	-18.3,-18.4,-18.5	-18.5	-
1-pentene	-4.6±1.1	-5.3	-	-5.1
Hexane	-41.0±1.4	-39.9,-40.0	-39.9	-39.9
2-methyl-pentane	-41.9±1.4	-41.7	-	-41.8
2,2-dimethyl-pentane	-44.0±1.4	-44.4	-	-44.5
Cyclohexane	-29.3±1.4	-29.8, -29.4, -29.5	-29.4	-
Methyl-cyclopentane	-25.5±1.4	-25.3,-25.5,-25.8	-	-
1-hexene	-9.8±1.4	-10.0	-	-10.4
3-methylene-pentane	-12.7±1.4	-13.4	-	-13.4
Heptane	-45.3±1.7	-44.9,-45.2	-44.8	-44.9
2-methyl-Hexane	-46.9±1.6	-46.6,-46.9	-	-46.5
2,4-dimethyl-pentane	-48.5±1.7	-48.3,-49.0	-	-48.2
3,3-dimethyl-pentane	-48.2±1.7	-48.2,-49.2	-	-48.1
1-heptene	-14.9±1.6	-15.1,-15.2,-14.9	-	-14.9
E-2-heptene	-17.0±1.6	-17.5,-17.7,-17.8	-	-
Z-2-heptene	-15.8±1.6	-16.4,-16.6,-16.9	-	-
Cycloheptane	-28.4±1.6	-28.5	-	-
Methyl-cyclohexane	-36.5±1.6	-37.0, -38.2	-	-
Octane	-50.4±1.9	-49.8,-49.9	-	-49.9
2-methyl-heptane	-52.0±1.9	-51.5	-	-51.5
3-methyl-heptane	-51.4±1.9	-50.8	-	-50.8
2,3-dimethyl-hexane	-51.3±1.9	-51.1	-	-51.1
2,3,3-trimethyl-Pentane	-51.9±1.9	-51.7	-	-51.7
1-octene	-18.7±1.9	-19.8	-	-19.5
Z-22-dimethyl-3-Hexene	-20.7±1.9	-23.1	-	-21.3
E-22-dimethyl-3-Hexene	-25.0±1.9	-28.2	-	-25.7

20

cyclooctane		-28.7±1.9	-30.1	-	-
n-Decane ($C_{10}H_{22}$)		-60.3±0.8	-59.7±0.3	-	-59.6±0.3
2-methyl-nonane ($C_{10}H_{22}$)		-61.2±0.8	-62.2		-62.1±0.5
1-decene ($C_{10}H_{20}$)		-29.6±0.9	-29.8	-	-29.5
JP-10 ($C_{10}H_{18}$)		-17.8±0.9	-17.6 [43],	-18.5[44],	-17.4±1.3[45]
n-Dodecane ($C_{12}H_{26}$)		-70.4±0.9	-69.5±0.3	-	-69.8±0.5
1-dodecene ($C_{12}H_{24}$)		-38.7±0.8	-39.5±0.5	-	-
E-2,3-Di-tert-butyl-2-butene ($C_{12}H_{24}$)		-33.9±0.8	-35.7	-	-
Hexyl-cyclohexane ($C_{12}H_{24}$)		-59.9±0.9	-60.8±0.4	-	-
n-Hexadecane ($C_{16}H_{34}$)		-89.9±1.6	-89.6	-	-
2-methylpentadecane ($C_{16}H_{34}$)		-91.8±1.6	-	-	-
1-cetene ($C_{16}H_{32}$)		-58.2±1.1	-59.1±0.6	-	-59.4±0.6
Decyl-cyclohexane ($C_{16}H_{32}$)		-79.3±1.1	-80.5±0.6	-	-

3.2 CGA-ONIOM Uncertainty Quantification

We further evaluated the performance of CGA-ONIOM method from the statistic point of view including the mean unsigned error (MUSE), mean signed error (MSE), and standard deviation (STD). Figure 7 plots the corresponding probability distribution function (PDF) and the probability distribution of the deviations between CGA-ONIOM and ATcT [35,40,41], NIST [39], TDOC [42], for the 79 C_4 to C_8 aliphatic molecules. The PDF of CGA-ONIOM relative to ATcT centers around 0 kcal·mol^{-1} (i.e. MSE of 0.01 kcal·mol^{-1}), and there are ~88% of the molecules within the deviation of ±1.0 kcal·mol^{-1}. In comparison, the PDFs relative to NIST and TDOC shift to 0.44 kcal·mol^{-1} and 0.33 kcal·mol^{-1}, with the probability of 77% and 76% within ±1.0 kcal·mol^{-1}, respectively. Nevertheless, the slight shift (< 0.5 kcal·mol^{-1}) is reasonable and acceptable, which may be caused by the truncation error of the angle scheme, ZPVE uncertainty, multi-conformer error or the uncertainties in the database.
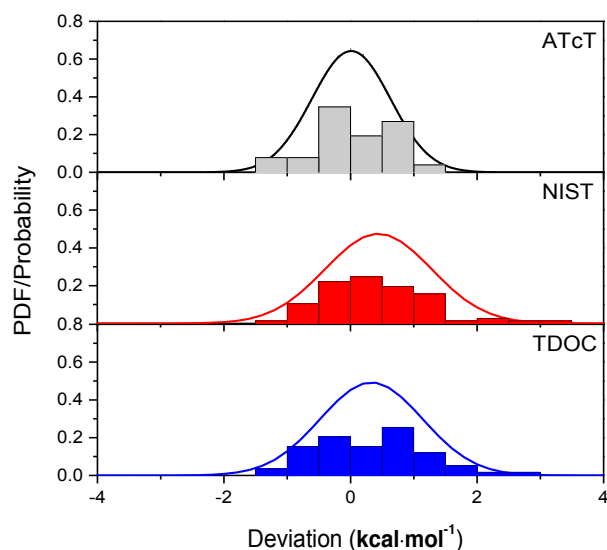
Figure 7. Probability density function (PDF, in line) and the probability distribution (in histogram) of the EOF deviation between CGA-ONIOM and ATcT (in black), NIST (in red) and TDOC (in blue) for the 79 aliphatic hydrocarbons.

Table 5. The MUSE, MSE, STD and 95% confidence interval of the CGA-ONIOM predicted EOF compared to the ATcT, NIST and TDOC databases. (units in $kcal \cdot mol^{-1}$)

	MUSE	MSE	STD	95% confidence interval
ATcT	0.54	0.01	0.61	0.01±1.03
NIST	1.42	0.44	0.84	0.44±1.43
TDOC	0.70	0.33	0.81	0.33±1.37
Tot.	0.68	0.33	0.79	0.33±1.33

Furthermore, we computed the 95% confidence interval of the EOFs by CGA-ONIOM compared to the three databases; see Table 5 for details. Note that 95% confidence interval is a very stringent indicator [7,46] to assess the performance of CGA-ONIOM. The best consistency is seen between the CGA-ONIOM and the ATcT database, presenting a 95% confidence interval of 0.01±1.03 $kcal \cdot mol^{-1}$. Regarding the overall performance, the CGA-ONIOM method is still fairly accurate and reliable, as it predicts EOFs with a MUSE of 0.68 $kcal \cdot mol^{-1}$, a MSE of 0.33 $kcal \cdot mol^{-1}$ and a 95% confidence interval of 0.33±1.33 $kcal \cdot mol^{-1}$.

3.3 Enthalpy of Formation of Large Hydrocarbons

Now that CGA-ONIOM has been validated against the small-to-middle sized hydrocarbons, we further applied this method to large hydrocarbons. In this work, 12 representative large aliphatic hydrocarbons including normal-, branched-, cyclo- and unsaturated categories, were selected to demonstrate the potential of CGA-ONIOM in accurate and efficient energy calculations. The C_{10} molecules include n-decane, 2-methyl-nonane, 1-decene, and exo-tricyclodecane (named as JP-10 for short); the C_{12} molecules contain n-dodecane, E-2,3-Di-tert-butyl-2-butene, 1-dodecene, and hexyl-cyclohexane; and the C_{16} molecules include n-hexadecane, 2-methyl-pentadecane, 1-hexadecene, and decyl-cyclohexane. All the calculated EOFs were listed in Table 4 and compared with available reference values in database or literature. We also estimated the EOFs at 298 K using the GA method [47] for comparison (details in Supporting Information). In general, the EOF calculated by CGA-ONIOM shows a very slight deviation (~1.0 kcal·mol^{-1}) from the reference values [39,42]. Besides, our calculations are also very close to the estimations using the conventional GA except for JP-10. It is noted that GA estimates a higher enthalpy of formation (-14.3 kcal·mol^{-1}) than CGA-ONIOM (-17.8 kcal·mol^{-1}) and other literature values [43–45] (-17.4 ~ -18.5 kcal·mol^{-1}). All the calculations follow the same procedure, so we only choose the simple n-alkane, i.e. n-dodecane and the complicated polycyclic alkane i.e. JP-10 as two representatives to demonstrate the process of using CGA-ONIOM, which follow a one-step and two-step partitioning processes, respectively.
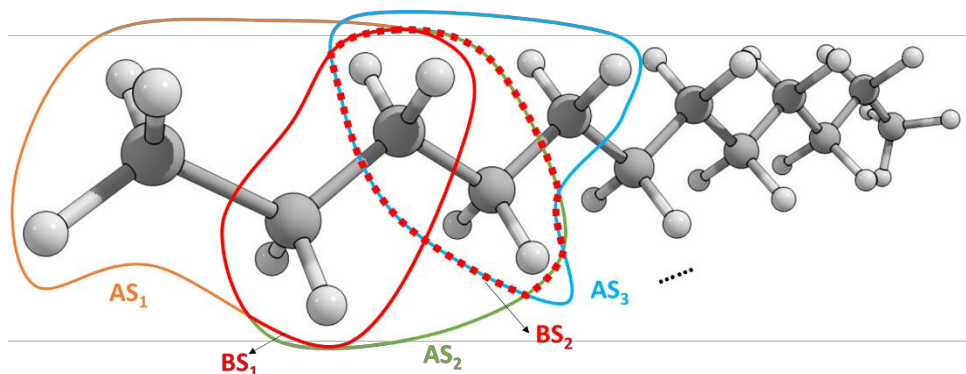
Figure 8. Schematic diagram of applying CGA-ONIOM for n-dodecane ($C_{12}H_{26}$) using the angle scheme (AS). This molecule is divided into 10 ASs with 9 bond scheme (BS) overlappings.

Figure 8 depicts the schematic diagram of applying CGA-ONIOM for n-dodecane ($C_{12}H_{26}$). It is straightforward to divide this molecule into groups, i.e. 2 propyl and 8 $-(CH_2)_3-$ angle schemes, which are cascaded via 9 CH_2-CH_2 bonds. Note that it is highly recommended to consider the molecular symmetry to reduce the computational efforts when applying CGA-ONIOM for energy calculations. Hence, only 5 angles (1 propyl and 4 $-(CH_2)_3-$) and 5 bonds ($CH_2- CH_2$) were considered for n-dodecane due to its symmetric feature. The CCSD(T)/CBS energy of n-dodecane is then approached by the CGA-ONIOM calculation:

$$E_{CCSD(T)/CBS}^{M} = E_{low}^{M}\left(Real\right) + \sum_{n=1}^{10}\Delta E_{CCSD(T)/CBS}\left(AS_n\right) - \sum_{n=1}^{9}\Delta E_{CCSD(T)/CBS}\left(BS_n\right) \quad (11)$$

On this basis, we calculated the EOF at 298.15 K using the CCSD(T)/CBS energy calculated by CGA-ONIOM method. Table 6 compares the EOF of n-dodecane obtained by the CGA-ONIOM method with the experimental data [39,42,48], group additivity estimation [47], and theoretical predictions [49] by the CBS-QB3, G2, G3 and G4 methods. Here we take the average of the three sets of experimental measurements as the EOF benchmark, i.e. -69.6 kcal·mol^{-1}. The comparison shows that the GA method estimates almost identical value to the benchmark, indicating the reliability of GA method for structure-simple, large fuel

molecules. Our CGA-ONIOM calculation predicts an EOF of -70.4±0.9 kcal·mol⁻¹, slightly lower than the benchmark by 0.8 kcal·mol⁻¹. In comparison, the EOFs determined by CBS-QB3, G2, G3 and G4 methods show a larger deviation from the benchmark by 1.1-2.3 kcal·mol⁻¹.

Table 6 Comparison of the EOF of n-dodecane ($C_{12}H_{26}$) obtained by the CGA-ONIOM-based CCSD(T)/CBS method with the experimental data [39,42,48], group additivity estimation [47], and theoretical calculations by CBS-QB3, G2, G3 and G4 methods [49].

	EOF @ 298.15 K	Deviation
Ref.	-69.4±0.2[a], 69.5±0.3[b], -69.8±0.5[c] (**-69.6**)	-
GA[47]	-69.6	0.0
CGA-ONIOM	**-70.4±0.9**	**-0.8**
CBS-QB3[49]	-67.5	+2.1
G2[49]	-71.3	-1.7
G3[49]	-71.9	-2.3
G4[49]	-70.7	-1.1

[a] Reference values collected from [48].

[b] Reference value collected from NIST Chemistry WebBook [39].

[c] Reference value collected from TDOC [42].

Similarly, we investigated the performance of using CGA-ONIOM to calculate the energy of JP-10 that is a typical polycyclic molecule. The structure of JP-10 molecule is relatively complex, incurring various angle scheme allocations shown as arrows in Figure 9(a). As previously discussed, the error caused by angle allocations has a minor effect on the final CGA-ONIOM-based energy. One can directly divide the JP-10 (or other polycyclic molecules) into a set of angle fragments as long as the dumbbell overlapping is guaranteed. This one-step process must be carefully conducted to avoid the possible miscounting of the angle and dumbbell combinations. In practice, we find it is quite challenging for a beginner to correctly use the process for complex molecules. Alternatively, one can adopt the two-step

process that involves an additional step to divide JP-10 into cyclic functional groups as such the following division into angle fragments is less prone to errors for a beginner. There remains a hierarchical structure of the polycyclic molecule, cyclic functional group and angle scheme: the JP-10 molecule comprises several cyclic functional groups; each cyclic functional group is comprised of a series of angle schemes. As both the one-step and two-step processes give the same results, here we used the two-step approach as an example for CGA-ONIOM calculations.
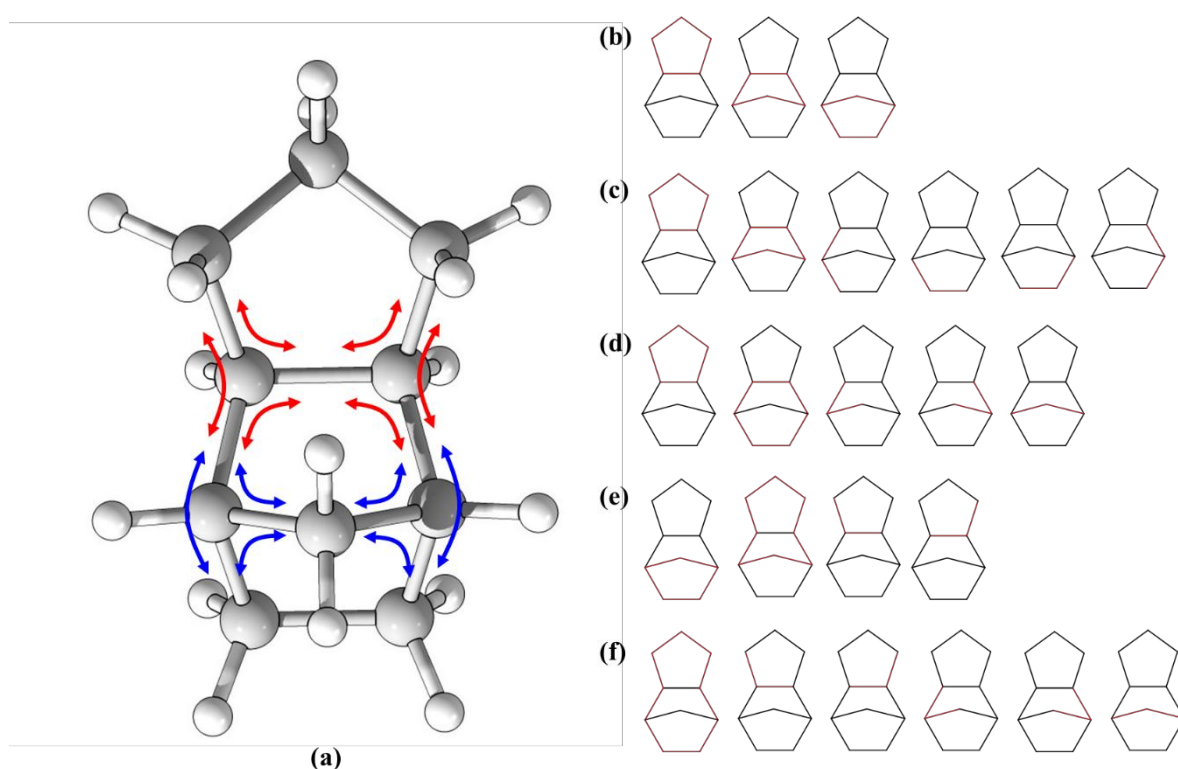
Figure 9. Application of CGA-ONIOM for JP-10 (boat): (a) 3D view of JP-10 with angle scheme allocations, and different ways of partitioning JP-10 into (b) three five-member ring, (c) two five-member ring and four angles, (d) one five-member ring, one six-member ring and three angles, (e) one five-member ring, one eight-member ring and two angles, and (f) one nine-member ring and five angles. Among (d)-(f), the red cyclic functional groups and angles are included in the high layer and treated by CCSD(T)/CBS.

We conducted an exhaustive search for the possible combinations of functional groups and found 13 combinations with the details provided in Supporting Information. Figure 9 depicts five representative combinations of different functional groups. Specifically, Figure

26

9(b) shows the combination of three five-member rings, cascaded by one bond overlapping and one angle overlapping. Figure 9(c) shows the combination of two five-member rings and four angles, cascaded by six bond overlappings. Figure 9(d) shows the combination of one five-member ring, one six-member ring, and three angles, which are also cascaded by five bond overlappings. Figure 9(e) shows the combination of one five-member ring, one eight-member ring and two angles, which are cascaded by one angle and two bond overlappings. Figure 9(f) shows the combination of one nine-member ring and five angles, which are cascaded by seven bond overlappings.

Table 7 lists the EOFs of JP-10 molecule including experimental data [43], GA estimation, 13 CGA-ONIOM calculations and other theoretical predictions [44,45]. The empirical estimation by the GA method gives an EOF of -14.3 kcal·mol^{-1}, which is 3.3 kcal·mol^{-1} larger than measurement. In comparison, CGA-ONIOM predicts an EOF ranging from -18.4 kcal·mol^{-1} to -17.3 kcal·mol^{-1}, leading to a mean EOF of -17.8 kcal·mol^{-1}. Among the previous theoretical predictions [44,45], it is notable that Bozzelli et al. [45] adopted the five isodemic work reactions to derive an EOF of -17.4±1.3 kcal·mol^{-1} by the CBS-QB3 and G3MP2B3 methods, which agrees well with our CGA-ONIOM calculations. We also calculated the EOF using the composite method, i.e. -18.1 kcal·mol^{-1} by CBS-QB3//M06-2x/cc-pVTZ and -16.9 kcal·mol^{-1} by G3//M06-2x/cc-pVTZ method, giving rise to a mean value of -17.5 kcal·mol^{-1}. The three sets of theoretical calculations, i.e., -17.8 kcal·mol^{-1} by CGA-ONIOM method, -17.4 kcal·mol^{-1} by isodemic reaction method and -17.5 kcal·mol^{-1} by CBS-QB3 and G3 method, agree well with the experimental data -17.6±0.6 kcal·mol^{-1}. Therefore, the CGA-ONIOM method features a better accuracy than the GA method for complex molecules, and a less computational cost than conventional composite methods for large molecules.

Table 7 EOFs of JP-10 molecule at 298.15 K predicted by CGA-ONIOM with different partitioning combinations. The group additivity estimation and theoretical calculations (units in kcal·mol^{-1}) are provided for comparison.
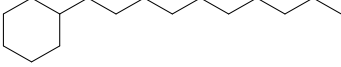
	EOF @ 298.15 K	Deviation
Expt. [43]	-17.6±0.6	-
GA	-14.3	+3.3
Scheme 1 (CGA-ONIOM)	-18.4	-0.8
Scheme 2 (CGA-ONIOM)	-17.9	-0.3
Scheme 3 (CGA-ONIOM)	-18.4	-0.8
Scheme 4 (CGA-ONIOM)	-18.4	-0.8
Scheme 5 (CGA-ONIOM)	-17.9	-0.3
Scheme 6 (CGA-ONIOM)	-17.9	-0.3
Scheme 7 (CGA-ONIOM)	-17.9	-0.3
Scheme 8 (CGA-ONIOM)	-17.8	-0.2
Scheme 9 (CGA-ONIOM)	-17.9	-0.3
Scheme 10 (CGA-ONIOM)	-17.3	+0.3
Scheme 11 (CGA-ONIOM)	-17.3	+0.3
Scheme 12 (CGA-ONIOM)	-17.4	+0.2
Scheme 13 (CGA-ONIOM)	-17.3	+0.3
Average.	**-17.8**	**-0.2**
G2[44]	-20.7	-3.1
G2(MP2) [44]	-19.8	-2.2
CBS-QB3[44]	-18.8	-1.2
CBS-QB3[45]	-17.5	-0.1
CBS-QB3//M06-2x/cc-pVTZ	-18.1	-0.5
G3(MP2) [44]	-18.2	-0.6
G3(MP2)//B3LYP/6-31G(d) [44]	-18.2	-0.6
G3MP2B3[45]	-17.4	+0.1
G3//M06-2x/cc-pVTZ	-16.9	+0.7

Finally, we evaluated the computational efficiency of CGA-ONIOM. Here we define the ratio of computational cost of conventional CCSD(T) to that of CGA-ONIOM as the computational efficiency ϕ. The CGA-ONIOM method only requires individual CCSD(T) calculations for small groups divided from the parent molecule. Thus the total computational cost primarily depends on the group number and scales polynomially with the molecular size. In this work, the corresponding cost depends on the number of angle scheme l_{AS} and number of dumbbell scheme l_{DS}. In comparison, the computational cost pf the conventional CCSD(T)

calculation scales exponentially with the molecular size [50]. Table 8 compares the computational cost required for CGA-ONIOM calculations with that of the conventional CCSD(T). The computational cost for large molecules with 16 heavy atoms can be dramatically reduced by a factor of $\sim 10^4$ (from 7759 for decyl-cyclohexane to 8348 for n-2-methylpentadecane). The required computational cost for molecules with fewer carbon atoms such as C_{10} species can be reduced by a factor of ~ 500 using CGA-ONIOM. Besides the reduced computational cost, the CGA-ONIOM method also enables parallel calculations. Each group partitioned from the target molecule can be treated individually on the separate nodes, making the CGA-ONIOM method more efficient in energy calculations.

Table 8 Comparison of the computational cost required by the CGA-ONIOM calculation (Ω_{CGA}) with the conventional CCSD(T) calculation ($\Omega_{conv.}$). Ω_{CGA} depends on the number of angle schemes (l_{AS}) and dumbbell schemes (l_{DS}). The computational efficiency (ϕ) is defined as the ratio of $\Omega_{conv.}:\Omega_{CGA}$.

Name	Structure	l_{AS}	l_{DS}	Ω_{CGA}	$\Omega_{conv.}$	ϕ
n-decane ($C_{10}H_{22}$)		8	7	1.83×10^{11}	9.96×10^{13}	544
2-methyl-nonane ($C_{10}H_{22}$)		8	6	1.82×10^{11}	9.96×10^{13}	548
1-decene ($C_{10}H_{20}$)		8	7	1.83×10^{11}	9.96×10^{13}	544
JP-10 ($C_{10}H_{18}$)		7	6	1.60×10^{11}	9.96×10^{13}	622
n-dodecane ($C_{12}H_{24}$)		10	9	2.29×10^{11}	3.57×10^{14}	1556
1-dodecene ($C_{12}H_{22}$)		10	9	2.29×10^{11}	3.57×10^{14}	1556
E-2,3-Di-tert-butyl-2-butene ($C_{12}H_{24}$)		10	3	2.22×10^{11}	3.57×10^{14}	1611
Hexyl-cyclohexane ($C_{12}H_{22}$)		11	10	2.52×10^{11}	3.57×10^{14}	1414
n-hexadecane ($C_{16}H_{34}$)		14	13	3.22×10^{11}	2.67×10^{15}	8315
2-methylpentadecane ($C_{16}H_{34}$)		14	12	3.20×10^{11}	2.67×10^{15}	8348
1-cetene ($C_{16}H_{32}$)		14	13	3.22×10^{11}	2.67×10^{15}	8315

Decyl-cyclohexane ($C_{16}H_{32}$)		15	14	3.45×10^{11}	2.67×10^{15}	7759

4. Summary

In this work, we developed the CGA-ONIOM method for accurate energy calculations of large fuel molecules at an affordable computational cost. The angle scheme involving three heavy atoms was proposed to obtain the energy of the large molecule with a high accuracy and efficiency. The feasibility of CGA-ONIOM method was scrutinized by calculating the enthalpy of formation of 79 C_4-C_8 aliphatic hydrocarbons. By comparing with ATcT, NIST and TDOC databases, the overall performance of CGA-ONIOM was quantified with the 95% confidence interval of 0.33 ±1.33 kcal·mol^{-1}. Further applications of the proposed CGA-ONIOM method for 12 large molecules including C_{10}, C_{12} and C_{16} hydrocarbons were also demonstrated, showing very good agreement with the reference values. The CGA-ONIOM method is found with a high computational efficiency as it dramatically reduces the computational cost by a factor of ~10^2 for C_{10} hydrocarbons and a factor of ~10^4 for C_{16} hydrocarbons.

Although this method was only validated against aliphatic hydrocarbons in this work, we believe that it is a promising method to be employed for combustion chemistry studies of other large fuel molecules. It is critical to select the group size properly to reach a trade-off between the calculation accuracy and efficiency when applying CGA-ONIOM for the parent molecule. It should be noted that truncation error caused by angle scheme can be reduced by adopting a large group size or treated by error-cancelling reaction methods, which however are not considered in this work. Besides, we also remind the readers that a larger partitioning scheme (i.e., N/M-shape schemes or above) is needed to cover the delocalized region for molecules with regionally delocalized electrons. The rest could still be treated using the angle

scheme in principle. However, the CGA-ONIOM method should be cautiously used for the molecules with electrons fully delocalized over the entire molecule due to the general failure of ONIOM under this circumstance [51]. Our future work involves the systematical uncertainty analysis and the application of CGA-ONIOM for aromatics, oxygenates, radicals and transition states. This study provides the first attempt of using the CGA-ONIOM method for the accurate energy determination of large fuel molecules in combustion research.

Reference

[1]   J.D. Watts, J. Gauss, R.J. Bartlett, Coupled-cluster methods with noniterative triple excitations for restricted open-shell Hartree–Fock and other general single determinant reference functions. Energies and analytical gradients, J Chem Phys. 98 (1993) 8718–8733.

[2]   W.J. Pitz, C.J. Mueller, Recent progress in the development of diesel surrogate fuels, Prog Energy Combust Sci. 37 (2011) 330–350.

[3]   J.Y.W. Lai, K.C. Lin, A. Violi, Biodiesel combustion: Advances in chemical kinetic modeling, Prog Energy Combust Sci. 37 (2011) 1–14.

[4]   C.K. Westbrook, Biofuels combustion, Annu Rev Phys Chem. 64 (2013) 201–219.

[5]   K. Kohse-Hoinghaus, P. Oßwald, T.A. Cool, T. Kasper, N. Hansen, F. Qi, C.K. Westbrook, P.R. Westmoreland, Biofuel combustion chemistry: From ethanol to biodiesel, Angew Chemie - Int Ed. 49 (2010) 3572–3597.

[6]   S.W. Benson, D.M. Golden, G.R. Haugen, R. Shaw, F.R. Cruickshank, A.S. Rodgers, H.E. O'neal, R. Walsh, Additivity rules for the estimation of thermochemical properties, Chem Rev. 69 (1969) 279–324.

[7]   T. He, S. Li, Y. Chi, H.-B. Zhang, Z. Wang, B. Yang, X. He, X. You, An adaptive distance-based group contribution method for thermodynamic property prediction, Phys Chem Chem Phys Phys Chem Chem Phys. 18 (2016) 23822–23830.

[8]   V. Van Speybroeck, R. Gani, R.J. Meier, The calculation of thermodynamic properties of molecules, Chem Soc Rev. 39 (2010) 1764–1779.

[9]   J. Sun, R.C. Remsing, Y. Zhang, Z. Sun, A. Ruzsinszky, H. Peng, Z. Yang, A. Paul, U. Waghmare, X. Wu, M.L. Klein, J.P. Perdew, Accurate first-principles structures and energies of diversely bonded systems from an efficient density functional, Nat Chem. 8 (2016) 831–836.

[10]   R. Peverati, D.G. Truhlar, Quest for a universal density functional: the accuracy of density functionals across a broad spectrum of databases in chemistry and physics, Philos Trans R Soc A Math Phys Eng Sci. 372 (2014) 20120476–20120476.

[11]   Y. Wang, X. Jin, H.S. Yu, D.G. Truhlar, X. He, Revised M06-L functional for improved accuracy on chemical reaction barrier heights, noncovalent interactions, and solid-state physics, Proc Natl Acad Sci. 114 (2017) 8487–8492.

[12]   T. Vreven, K.S. Byun, I. Komáromi, S. Dapprich, J.A. Montgomery, K. Morokuma, M.J. Frisch, Combining quantum mechanics methods with molecular mechanics methods in ONIOM, J Chem Theory Comput. 2 (2006) 815–826.

[13]   T. Vreven, K. Morokuma, The accurate calculation and prediction of the bond dissociation energies in a series of hydrocarbons using the IMOMO (integrated molecular orbital+molecular orbital) methods, J Chem Phys. 111 (1999) 8799.

[14]   R.D.J. Froese, K. Morokuma, IMOMO-G2MS Approaches to Accurate Calculations of Bond Dissociation Energies of Large Molecules, J Phys Chem A. 103 (1999) 4580–4586.

[15]   O. Eisenstein, P.B. Hitchcock, A. V. Khvostov, M.F. Lappert, L. Maron, L. Perrin, A. V. Protchenko, Mono-, di-, and trianionic β-diketiminato ligands: A computational study and the synthesis and structure of [(YbL)3(THF)], L = [{N(SiMe3)C(PH)}2CH], J Am Chem Soc. 125 (2003) 10790–10791.

[16]   R. Noyori, M. Kitamura, T. Ohkuma, Asymmetric Catalysis Special Feature Part I: Toward efficient asymmetric hydrogenation: Architectural and functional engineering of chiral molecular catalysts, Proc Natl Acad Sci. 101 (2004) 5356–5362.

[17]   Y. Kitagawa, T. Saito, M. Ito, Ms. Shoji, K. Koizumi, S. Yamanaka, T. Kawakami, M. Okumura, K. Yamaguchi, Approximately spin-projected geometry optimization method and its application to di-

chromium systems, Chem Phys Lett. 442 (2007) 445–450.

[18]   E. Lipiec, R. Sekine, J. Bielecki, W.M. Kwiatek, B.R. Wood, Molecular characterization of DNA double strand breaks with tip-enhanced Raman scattering, Angew Chemie - Int Ed. 53 (2014) 169–172.

[19]   H. Hirao, K. Morokuma, What is the real nature of ferrous soybean lipoxygenase-1? a new two-conformation model based on combined ONIOM(DFT:MM) and multireference configuration interaction characterization, J Phys Chem Lett. 1 (2010) 901–906.

[20]   H.M. Senn, W. Thiel, QM/MM methods for biomolecular systems, Angew Chemie - Int Ed. 48 (2009) 1198–1229.

[21]   S.J. Klippenstein, From theoretical reaction dynamics to chemical modeling of combustion, Proc Combust Inst. 36 (2017) 77–111.

[22]   J. Wu, H. Ning, L. Ma, W. Ren, Accurate Prediction of Bond Dissociation Energies of Large n-Alkanes Using ONIOM-CCSD(T)/CBS Methods, Chem Phys Lett. (2018).

[23]   L. Zhang, P. Zhang, Towards high-level theoretical studies of large biodiesel molecules: an ONIOM [QCISD(T)/CBS:DFT] study of hydrogen abstraction reactions of $CnH2n+1$ $COOCmH2m+1$ + H, Phys. Chem. Chem. Phys. 17 (2015) 200–208.

[24]   L.W. Chung, W.M.C. Sameera, R. Ramozzi, A.J. Page, M. Hatanaka, G.P. Petrova, T. V. Harris, X. Li, Z. Ke, F. Liu, H.B. Li, L. Ding, K. Morokuma, The ONIOM Method and Its Applications, Chem Rev. 115 (2015) 5678–5796.

[25]   T.H. Dunning, Gaussian basis sets for use in correlated molecular calculations through neon and hydrogen. I . The atoms boron through neon and hydrogen, 90 (1989).

[26]   D. Feller, D.A. Dixon, Extended benchmark studies of coupled cluster theory through triple excitations, J Chem Phys. 115 (2001) 3484–3496.

[27]   Y. Zhao, D.G. Truhlar, The M06 suite of density functionals for main group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and transition elements: Two new functionals and systematic testing of four M06-class functionals and 12 other function, Theor Chem Acc. 120 (2008) 215–241.

[28]   M.J. Elrod, R.J. Saykally, Many-Body Effects in Intermolecular Forces, Chem Rev. 94 (1994) 1975–1997.

[29]   S. Hua, W. Li, S. Li, The generalized energy-based fragmentation approach with an improved fragmentation scheme: Benchmark results and illustrative applications, ChemPhysChem. 14 (2013)

108–115.

[30] Y. Chi, X. You, L. Zhang, W. Li, Utilization of generalized energy-based fragmentation method on the study of hydrogen abstraction reactions of large methyl esters, Combust Flame. 190 (2018) 467–476.

[31] N. Cohen, S.W. Benson, Estimation of heats of formation of organic compounds by additivity methods, Chem Rev. 93 (1993) 2419–2438.

[32] P. Buerger, J. Akroyd, S. Mosbach, M. Kraft, A systematic method to estimate and validate enthalpies of formation using error-cancelling balanced reactions, Combust Flame. 187 (2018) 105–121.

[33] Y. Minenkov, H. Wang, Z. Wang, S.M. Sarathy, L. Cavallo, Heats of Formation of Medium-Sized Organic Compounds from Contemporary Electronic Structure Methods, J Chem Theory Comput. 13 (2017) 3537–3560.

[34] H. Frisch, M.J.E.A., Trucks, G.W., Schlegel, H.B., Scuseria, G.E., Robb, M.A., Cheeseman, J.R., Scalmani, G., Barone, V., Mennucci, B., Petersson, G.A. and Nakatsuji, Gaussian 09, revision D. 01., (2009).

[35] S.J. Klippenstein, L.B. Harding, B. Ruscic, Ab Initio Computations and Active Thermochemical Tables Hand in Hand: Heats of Formation of Core Combustion Species, J Phys Chem A. 121 (2017) 6580–6602.

[36] R.D. Johnson III, NIST Computational Chemistry Comparison and Benchmark Database, NIST Standard Reference Database Number 101, Release 15b. 2011, (n.d.).

[37] I.M. Alecu, J. Zheng, Y. Zhao, D.G. Truhlar, Computational Thermochemistry: Scale Factor Databases and Scale Factors for Vibrational Frequencies Obtained from Electronic Model Chemistries, J Chem Theory Comput. 6 (2010) 2872–2887.

[38] Y. Li, H.J. Curran, An Extensive Theoretical Study of the Thermochemical Properties of Unsaturated Hydrocarbons, Allylic and Super-Allylic Radicals: The Development and Optimization of Group Additivity Values, J Phys Chem A. 122 (2018) 4736–4749.

[39] NIST Chemistry WebBook, SRD 69:, (n.d.). http://webbook.nist.gov/chemistry/.

[40] B. Ruscic, R.E. Pinzon, G. Von Laszewski, D. Kodeboyina, A. Burcat, D. Leahy, D. Montoy, A.F. Wagner, Active Thermochemical Tables: Thermochemistry for the 21st century, J Phys Conf Ser. 16 (2005) 561–570.

[41] B. Ruscic, R.E. Pinzon, M.L. Morton, G. Von Laszevski, S.J. Bittner, S.G. Nijsure, K.A. Amin, M. Minkoff, A.F. Wagner, Introduction to active thermochemical tables: Several "Key" enthalpies of

formation revisited, J Phys Chem A. 108 (2004) 9979–9997.

[42]   JB Pedley, Thermochemical data of organic compounds, Springer: Netherlands, 2012.

[43]   J.S. Chickos, D. Hillesheim, G. Nichols, The enthalpies of vaporization and sublimation of exo- and endo-tetrahydrodicyclopentadienes at T = 298.15 K, J Chem Thermodyn. 34 (2002) 1647–1658.

[44]   M.J. Zehe, R.L. Jaffe, Theoretical calculation of jet fuel thermochemistry. 1. Tetrahydrodicylopentadiene (JP10) thermochemistry using the cbs-qb3 and G3(MP2)//B3LYP methods, J Org Chem. 75 (2010) 4387–4391.

[45]   J.M. Hudzik, R. Asatryan, J.W. Bozzelli, Thermochemical Properties of exo-Tricyclo[5.2.1.02,6]decane (JP-10 Jet Fuel) and Derived Tricyclodecyl Radicals, J Phys Chem A. 114 (2010) 9545–9553.

[46]   B. Ruscic, Uncertainty quantification in thermochemistry, benchmarking electronic structure computations, and Active Thermochemical Tables, Int J Quantum Chem. 114 (2014) 1097–1101.

[47]   Group Additivity Based Estimates, NIST Chem WebBook, SRD 69. (n.d.). https://webbook.nist.gov/chemistry/grp-add/.

[48]   William M. Haynes, CRC Handbook of Chemistry and Physics, 95th Editi, CRC press, 2014.

[49]   M.M. Ghahremanpour, P.J. Van Maaren, J.C. Ditz, R. Lindh, D. Van Der Spoel, Large-scale calculations of gas phase thermochemistry: Enthalpy of formation, standard entropy, and heat capacity, J Chem Phys. 145 (2016).

[50]   D.G. Truhlar, Basis-set extrapolation, Chem Phys Lett. 294 (1998) 45–48.

[51]   F.R. Clemente, T. Vreven, M.J. Frisch, Getting the Most out of ONIOM: Guidelines and Pitfalls, Quantum Biochem. (2010) 61–83. s