



Cite this: *Sustainable Energy Fuels*,
2019, 3, 457

Received 29th July 2018
Accepted 19th November 2018

DOI: 10.1039/c8se00390d

rsc.li/sustainable-energy

Computational design of biofuels from terpenes and terpenoids†

Efrat Pahima,^{ID} Shmaryahu Hoz, Moshe Ben-Tzion and Dan Thomas Major*

Finding renewable energy sources as alternatives to petroleum-based fuels is a current global challenge. One approach to address the energy shortage problem is through biofuels. A promising family of biofuels that has many of the needed fuel characteristics is terpenes. Herein we present a combined theoretical and statistical model for calculating inherent thermodynamic properties of several promising terpenes, which show high compatibility with many criteria of petroleum-based fuels. We use density functional theory and *ab initio* quantum chemistry methods to compute the enthalpy of combustion, enthalpy of vaporization, enthalpy of formation, cetane number, boiling point and vapor pressure for a range of terpenes with good accuracy. The current *in silico* study presents a promising strategy for finding suitable petroleum substitutes, while avoiding costly experimental trial and error approaches.

Introduction

A major problem currently facing the international community is finding substitutes for petroleum-based fuels. In this field of alternative energy sources, the use of biofuels might be a viable alternative, considering that these organic molecules can be used directly in existing fuel tanks and vehicles.

Biofuels are sustainable and renewable energy materials, and their formation avoids environmentally damaging processes, in contrast to petroleum distillation, which has significant ecological side effects.^{1,2} This gives biofuels an advantage over the currently used diesel and gasoline. For example, when plants, such as soybeans, grow they employ CO₂ from the air to grow, and after the oil is extracted from the soybeans, the remains are converted into biodiesel. When the biodiesel is burnt, CO₂ and other emissions are released and return to the atmosphere. This cycle does not add to the net CO₂ concentration in the air because the next soybean crop will reuse CO₂ as it grows. In contrast, when fossil fuels, such as coal or diesel, are burnt, all the CO₂ released adds to the net CO₂ level in the air.^{1,3} Additional benefits of using biofuels are that even low concentrations of biodiesel reduce particulate matter, *i.e.* exhaust from diesel engines in the aerosol form, reduce emissions and provide significant health and compliance benefits wherever humans are exposed to higher levels of diesel exhaust.¹

On the other hand, there are also problems associated with the use of biofuels. A main drawback is facilitating mass production. For example, the global food shortage combined

with the continued growth in population makes large-scale diversion of farmland to fuel production an unsustainable technological avenue.^{4,5} Biofuel alternatives, such as sugars, require expensive thermal, chemical and biochemical treatment prior to use in fermentation processes. Biomass from algae can be generated in water and not on expensive land, but extracting it from water on a large scale raises many challenges.^{6–9}

To transform biofuels from an idea into a real industrial alternative to petroleum-based fuels, it is essential to understand how one can best design efficient biomaterials. To this end, there is a need to understand what makes biofuel molecules suitable as fuel substitutes. To be considered as a target fuel that is compatible with existing engines, bio-fuel molecules must have properties similar to fossil-based fuels, and in particular, the following factors are important:^{1,6,10,11} energy contents or the amount of energy produced during combustion (*i.e.* enthalpy of combustion), combustion quality (*e.g.* octane number, for spark ignition engines, or cetane number, for compression ignition engines), volatility, viscosity, cloud point, freezing point, density, flash point, odor, and toxicity. Many biofuel alternatives that have properties compatible with fuel requirements exist, such as fatty acid methyl esters,^{1,7,12,13} sugars, alcohols and more (Table 1).^{3,14–18}

A particular class of compounds that shows great potential as biofuels is terpenes or terpenoids.^{3,6,7,10,19,20} Terpenes are compounds derived biosynthetically from units of isoprene, C₅H₈, while isoprenoids, or terpenoids, are terpenes with additional functional groups, such as hydroxyl, epoxy, and carbonyl moieties. The basic molecular formula of terpenes is (C₅H₈)_{*n*}, where *n* is the number of isoprene units. The isoprene units may be linked together in the form of linear chains or rings, to form mono-terpenes (*n* = 2), sesquiterpenes (*n* = 3), diterpenes (*n* = 4), *etc.* These abundant organic compounds are composed of up to 60%

Department of Chemistry, Bar-Ilan University, Ramat-Gan 52900, Israel. E-mail: majort@biu.ac.il

† Electronic supplementary information (ESI) available: Gas phase calculation data, calculation explanation, thermodynamic explanation, and Cartesian coordinates of the optimized gas-phase structure. See DOI: 10.1039/c8se00390d

Table 1 Types of liquid fuels⁶

Fuel type	Major components	Important properties	Biosynthetic alternatives
Gasoline	C4–C12 hydrocarbons: linear, branched, cyclic, and aromatic; anti-knock additives	Octane number, energy content, and transportability	Ethanol, <i>n</i> -butanol and iso-butanol; short chain alcohols; short chain alkanes; ^{7,22–25}
Diesel	C9–C23 hydrocarbons: linear, branched, cyclic, and aromatic; anti-freeze additives	Cetane number, low freezing temperature, and low vapor pressure	Biodiesel; fatty acid methyl ^{1,7} and ethyl ^{17,26} esters, fatty alcohols, ^{12,13,27–29} alkanes; ^{30,31} linear or cyclic terpenes ^{7,10,19,32,33}
Jet fuel	C8–C16 hydrocarbons: linear, branched, cyclic, and aromatic; anti-freeze additives	Very low freezing temperature, enthalpy of combustion, and density	Alkanes; biodiesel; linear or cyclic terpenes ^{7,34–36}

of all natural compounds and are used in industry due to their fragrance, flavor and pharmaceutical properties.²¹

Recent advances in microbial and non-enzymatic synthesis of terpenes could facilitate future large scale generation of these molecules.^{37–39}

Linear or cyclic monoterpenes (C10) or sesquiterpenes (C15) are potential targets for biodiesel fuel. Their methyl branches, double bonds and ring structures are known to improve the fluidity at lower temperatures.⁶ Such branching substantially decreases melting and freezing points and thus can render

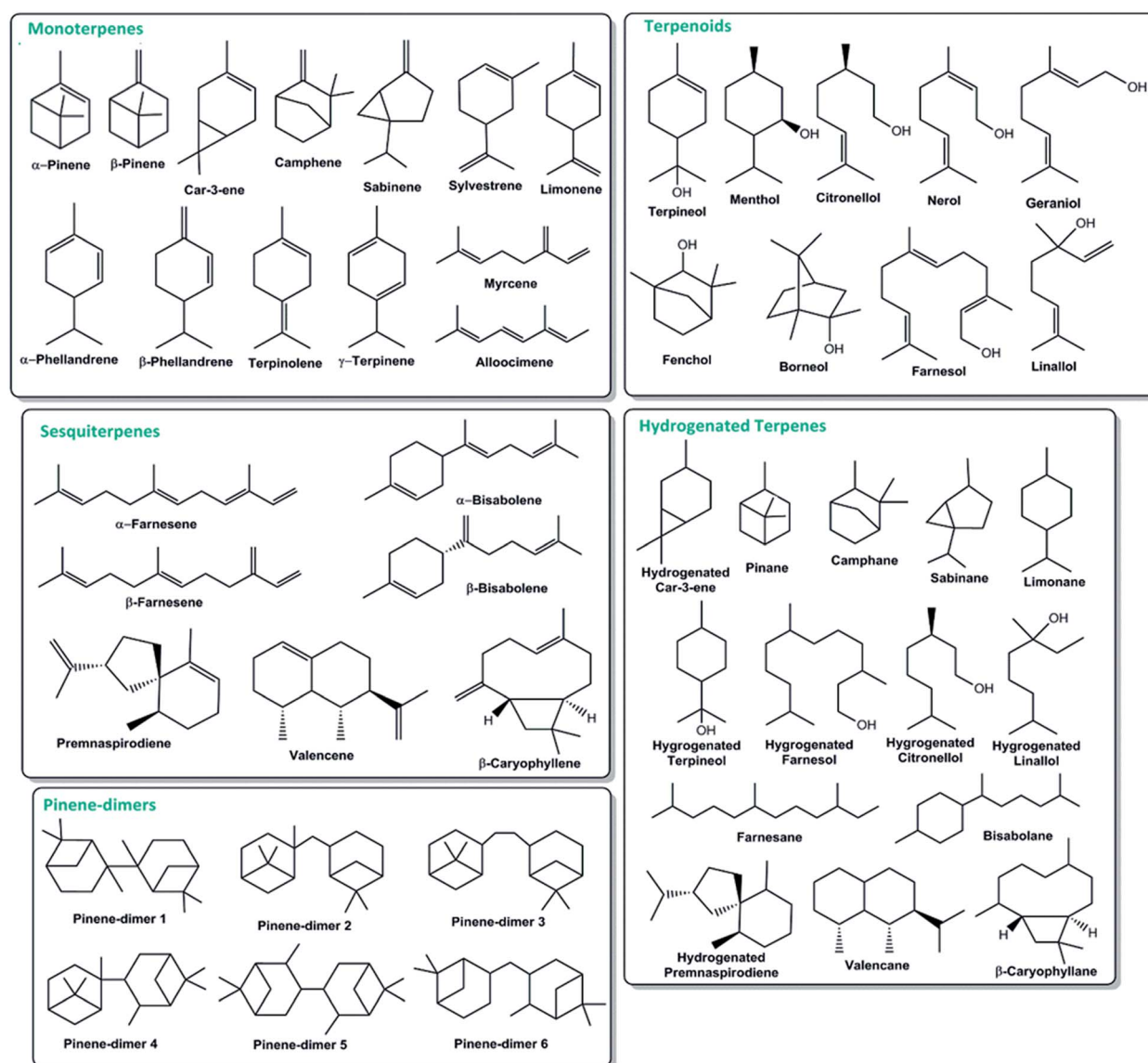


Chart 1 Terpenes and terpenoids evaluated in this study.

biofuels more suitable for cold-weather applications.^{3,40} It might also lower cetane ratings, but they can be improved by complete or partial reduction of a terpene molecule's double bonds.⁶ Branching is overall favorable, as it naturally provides terpene molecules with tertiary carbon centers, with an inherent ability to stabilize radical species formed under high-pressure thermal conditions. Moreover, their dense hydrocarbon architecture results in an abundance of dispersion interactions, which also stabilize radicals. Such properties reduce unwanted knocking and increase the compressibility of the fuel.^{7,10,19,32,33}

In the current work, we present an *in silico* strategy to compute several properties of terpenes and terpenoids relevant to biofuels (this work will discuss pure terpene-based biofuels and not blends). We use density functional theory (DFT) and *ab initio* quantum chemistry methods with the Gaussian G09 program⁴¹ in order to compute the following thermodynamic and physical properties:

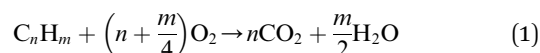
- Enthalpy (heat) of combustion (ΔH_{comb}).
- Enthalpy (heat) of vaporization (ΔH_{vap}).
- Enthalpy (heat) of formation (ΔH_f).
- Cetane rating/number (CN).
- Boiling point.
- Vapor pressure.

In Cht 1, we present the terpenes and terpenoids that were studied as potential biofuel alternatives.

Theoretical methods

Computation of enthalpy of combustion

The enthalpy of combustion is the heat released in the form of energy during a combustion reaction. The general combustion reaction is described by chemical eqn (1):



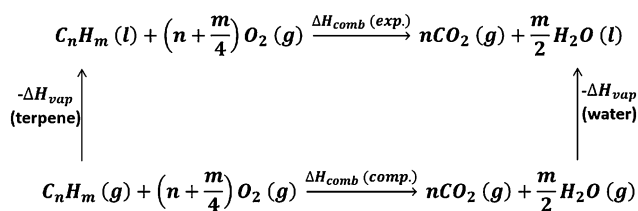
The calculations are performed using eqn (2):⁴²

$$\begin{aligned} \Delta H_r^\circ(298\text{ K}) &= \Delta H_{\text{combustion}}^\circ(298\text{ K}) \\ &= \sum_{i, \text{Products}} v_i (\epsilon_{0,i} + H_i^{\text{corr}}) - \sum_{j, \text{Reactants}} v_j (\epsilon_{0,j} + H_j^{\text{corr}}) \end{aligned} \quad (2)$$

where v_i is the stoichiometric constant of product i , $\epsilon_{0,i}$ is the electronic energy of product i and H_i^{corr} is the enthalpy correction to the electronic energy of product i , while j is the same for reactants (for further explanation, see the ESI†).

This property is calculated using a hybrid DFT method.⁴³ However, two major corrections need to be added to the calculations in order to achieve accurate results.

First, an addition of enthalpy of vaporization is required for the computed enthalpies of the terpenes and water. Quantum calculations of individual molecules in the absence of external fields correspond to the gas phase, while reaction enthalpies are measured experimentally when all reactants and products are at 298 K and 1 atm, *i.e.* the liquid phase for water and many terpenes. A schematic way to demonstrate how this correction was added is presented in Scheme 1 and eqn (3).



Scheme 1

$$\begin{aligned} \Delta H_{\text{combustion}} &= n(\epsilon_0 + H^{\text{corr}})_{\text{CO}_2(\text{g})}^{298\text{ K}} \\ &+ \frac{m}{2} \left[(\epsilon_0 + H^{\text{corr}})_{\text{H}_2\text{O}(\text{g})}^{298\text{ K}} - \Delta H_{\text{vap}, \text{H}_2\text{O}} \right] \\ &- \left[(\epsilon_0 + H^{\text{corr}})_{\text{C}_n\text{H}_m(\text{g})}^{298\text{ K}} - \Delta H_{\text{vap}, \text{C}_n\text{H}_m} \right] \\ &+ \left(n + \frac{m}{4} \right) (\epsilon_0 + H^{\text{corr}})_{\text{O}_2(\text{g})}^{298\text{ K}} \end{aligned} \quad (3)$$

A second correction was added in order to account for the tendency of many DFT methods to overbind the electrons in molecular O_2 , resulting in the enthalpy of O_2 being too negative,^{44–46} due to electron self-interactions⁴⁷ (additional explanation can be found in the ESI†).

Computation of enthalpy of vaporization

The enthalpies of vaporization, defined as the energy required to transfer a single molecule from the liquid phase to the gas phase, were calculated using the SMD solvation model.⁴⁸ SMD is an implicit solvation model that computes the free energy of solvation, ΔG_{sol} , defined as the free energy required to transfer a molecule from the gas phase to the liquid phase. Practically, by using multiple linear regression (MLR), the computed free energy of solvation is converted into the enthalpy of vaporization, by summing with a key factor accounting for the entropic effect of solvation – the solvent accessible surface area (SASA, eqn (4)). The use of the MLR allows for the conversion of ΔG_{sol} into ΔH_{vap} without explicitly calculating the entropy of solvation.

$$-\Delta G_{\text{sol}} + T \cdot \text{SASA} \propto \Delta H_{\text{vap}} \quad (4)$$

This property is then added to the ΔH_{comb} calculations, as described above. See the ESI† for further details.

Computation of enthalpy of formation

The enthalpy of formation is defined as the difference in enthalpy between a molecule and the atoms that make up that molecule in their naturally occurring elemental form. In other words, it indicates the amount of heat released or absorbed when a molecule is formed or how much enthalpy is required to completely break down the molecule.

This property is calculated using the *ab initio* G4MP2 method.^{49–51} This level of theory is required in order to get the high accuracy needed for computing the atomization energies of the constituent elements of the reactants in their reference standard states (graphite for C, H_2 and O_2 at 1 atm for atomic H and O).

Estimation of the cetane number

The cetane number describes the ignition quality of fuels suitable for a compression-ignition engine, which is of great importance for diesel fuel formulations. Low cetane numbers indicate poor fuel ignition, which results in difficulty starting and running the engine.

Specifically, the cetane number quantifies the ignition delay of a fuel, and is influenced mainly not only by the fuel's chemical structure (causing "chemical delay" in ignition), but also by physical properties such as its density, viscosity, enthalpy of vaporization, heat capacity and more (causing "physical delay").^{4,52,53}

The finite rate of the radical-forming oxidation reaction cascade occurring under harsh conditions in the engine's cylinders is responsible for the "chemical delay".⁵² We estimate this property by calculating the rate of what is presumed to be the rate determining step in the combustion reaction, *i.e.* the initiation step, where an oxygen molecule extracts a hydrogen atom to form a terpene-radical (Scheme 2).⁵⁴ The energy barrier for this step was calculated by performing a transition state (TS) search using hybrid DFT.

This property was combined in a MLR calculation with properties causing the physical delay, *e.g.* the enthalpy of vaporization and heat capacity (both influence the time required for a drop of fuel to heat and vaporize when injected into the cylinder). The R^2 correlation was checked by the F distribution test to estimate the likelihood of the correlation occurring by chance.

Computation of the boiling point and vapor pressure

The boiling point and vapor pressure were calculated using the ADF COSMO-RS (Conductor like Screening Model for Realistic Solvents) program.^{55–59} In the COSMO-RS model, the molecules are regarded as a collection of surface segments and the chemical potential of each segment is self-consistently determined from a statistical mechanical relationship.⁵⁹ By calculating the difference in the chemical potential of the molecules in the liquid phase and the gas phase, the vapor pressure can be estimated.⁵⁷ By estimating the vapor pressure at different temperatures in an iterative manner, the boiling point can be calculated as well. The boiling points and vapor pressures computed by COSMO-RS deviate from the experimental values by approx. 4.6% and 29.3%, respectively. Thus, linear regression was applied in order to calibrate the computed values with the experimental ones (see the ESI† for further information).

Computational details

The enthalpy of combustion was calculated using the M06-2X functional⁴³ in conjunction with the 6-31+G(p,d) basis set, as this approach gave the most accurate results compared to the other DFT methods tested in this study (see ESI Tables S1–S7

and Fig. S1a–f†). We have also shown that this method performs well for terpene chemistry in enzyme systems.^{60–63} The enthalpies of vaporization were calculated using the SMD solvation model,⁴⁸ with M06-2X/6-31+G(d,p). The enthalpies of formation were calculated using the *ab initio* Gaussian-4 method with second-order Møller–Plesset (MP2) perturbation theory (*i.e.* G4MP2).⁴⁹ All the thermodynamic properties described were calculated from the basic physical properties of the molecule, as obtained from the electronic energy, and translational, rotational and vibrational motion of the molecule within the harmonic approximation. Boltzmann averages over conformers for flexible molecules were taken into consideration as well (explanation and details are available in the ESI†). The cetane number was estimated using transition state calculations (QST3 keyword), the computed heat capacity at a constant volume (Freq keyword), and the enthalpies of vaporization (SMD model), all calculated at the M06-2X/6-31+G(d,p) level of theory. These different properties were combined as a measure of cetane number using MLR. All quantum mechanical calculations were carried out using the Gaussian 09 program (version B.01).⁴¹ All optimized geometries were confirmed as local minima by having all real vibrational frequencies. All TS structures were confirmed as having one imaginary vibration, and in order to verify that the TS connects the reactant and product wells we performed intrinsic reaction coordinate (IRC)^{64,65} calculations. The boiling point and vapor pressure were calculated using the ADF COSMO-RS program.^{55,58} The geometry optimization required for calculating the COSMO-RS properties was performed with ADF, with a small core TZP basis set, the Becke–Perdew functional (BP86), the relativistic scalar ZORA method, and good numerical integration quality (default settings). We note that all values were computed at 298 K, with the exception of the transition state calculations, which were performed at 1000 K (explained in the following section).

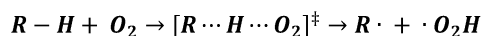
Results

In the current paper, we propose a computational protocol for calculating the inherent thermodynamic properties of several promising terpene molecules, which show high compatibility with petroleum fuel criteria. The computational protocol is calibrated and validated using known experimental data.

We employ DFT and *ab initio* computational chemistry methods to generate a statistical correlation model, which is used to calculate thermodynamic properties, including the enthalpy of combustion, enthalpy of formation, enthalpy of vaporization, cetane number, boiling point and vapor pressure.

Enthalpy of vaporization

In order to calculate the ΔH_{comb} , ΔH_{vap} calculations are performed using the SMD model combined with MLR, where x_1 is the computed negative free energy of solvation and x_2 is the temperature (298.15 K) multiplied by the computed solvent accessible surface area (SASA, computed as part of the SMD calculations). We applied cross-validation to estimate the reliability of the predicted values.



Scheme 2

Fig. 1 presents the correlation between the predicted and the experimental ΔH_{vap} . The values and deviations from the experimental ΔH_{vap} , the cross-validation procedure, and further details on the calculation procedure and the SMD model are presented in Fig. S2a and b, and Table S8 (see the ESI†).

The results show that the strategy adopted here, combining quantum chemistry and simple statistical modeling, can provide highly accurate vaporization enthalpies for terpenes, with maximal errors of approx. 1 kcal mol⁻¹ for the training set. The correlation coefficient values (R^2) obtained from the MLR calculations after the cross-validation procedure are as follows: 0.97 for the mono- and sesquiterpenes (Fig. 1a) and 0.69 for the terpenoids (Fig. 1b). The Q^2 values, obtained from the correlation between the experimental enthalpies of vaporization and the predicted values obtained from the cross-validation, are 0.92 for terpenes and 0.46 for terpenoids (see Fig. S2a and b†). The results for the terpenoids are less accurate than those for the terpenes, likely because of the lack of hydrogen-bonds in the implicit SMD solvation model. However, the predicted enthalpies of vaporization still have a deviation of only approx. 1 kcal mol⁻¹. The biggest error is obtained for linalool, with an error of -0.97 kcal mol⁻¹.

Since the coefficient of the x_2 value is small ($\sim 10^{-5}$), a linear regression calculation was performed using the computed ΔG_{sol}

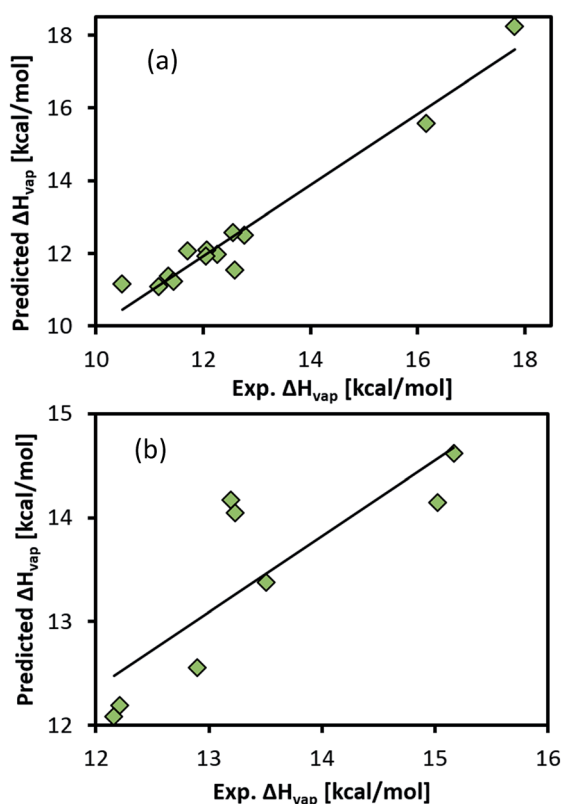


Fig. 1 Calculated enthalpies of vaporization vs. experimental values. The predicted values were obtained using the SMD solvation model with M06-2X/6-31+G(d,p) and MLR with cross-validation. (a) Mono- and sesquiterpenes, regression equation: $y = 1.006643x_1 + 0.000097x_2 + 0.228989$, $R^2 = 0.97$ and (b) terpenoids, regression equation: $y = -0.153031x_1 + 0.000247x_2 - 1.115814$, $R^2 = 0.69$. y is ΔH_{vap} , x_1 is $-\Delta G_{\text{sol}}$, and x_2 is T·SASA.

and the experimental ΔH_{vap} , to judge whether the T·SASA variable is negligible. The results showed lower R^2 values and higher standard errors for the linear regression, indicating that the T·SASA is important. More likely the small value of the coefficient of x_2 balances the large absolute value of x_2 (see Table S9† for more statistical data).

Enthalpy of combustion

As explained above, after adding ΔH_{vap} to the ΔH_{comb} calculations, linear regression was applied (correlating the experimental ΔH_{comb} values with the computed ones) in order to account for the deviation in the computed values. The deviation in the computed ΔH_{comb} is possibly caused by O₂ overbinding, which is due to DFT's tendency to overbind the electrons in the O₂ molecule (see the ESI† for further details).^{44–46} The correlation plot is presented in Fig. 2 (also see ESI Fig. S3a and b, and Table S10†). The difference between the mono- and sesquiterpene's values is causing a bias towards the high R^2 value (0.99) obtained in this regression analysis. Zooming in on the mono- and sesquiterpene regions in Fig. 2 suggests that the calculated values indeed possess a good correlation with the experimental values. The LR calculations used 12 terpenes as a training set and 7 terpenes as a test set (all included in Fig. 2). Two additional similar correlation plots were constructed, but with different terpenes as training and test sets, to ensure that the correlation with the experimental data persists for randomly chosen training and test sets. For further details and for the predicted ΔH_{comb} values obtained from the linear regression analysis, see Fig. S3a and b, and Table S10.†

The enthalpy of combustion is predicted rather accurately, with an RMS error of 5.5 kcal mol⁻¹, which corresponds to an error of approx. 0.3%.

By dividing the enthalpy of combustion (the heat released in the combustion reaction) by the molecular mass, we obtain the specific energy (Fig. 3 and Table S11†). In Fig. 3, we present the

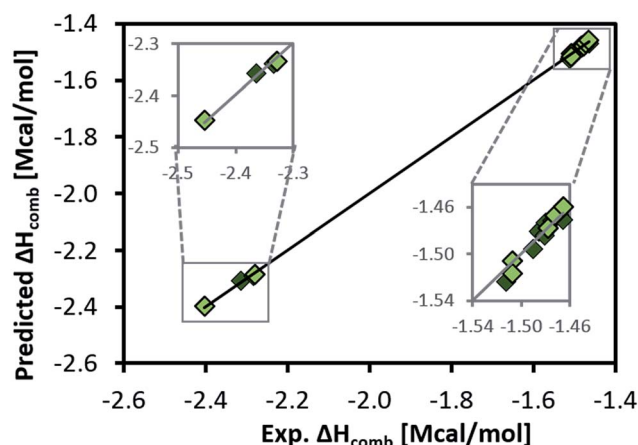


Fig. 2 Correlation between experimental and computed combustion enthalpy (using M06-2X/6-31+G(d,p)). The abscissa values are the calculated ΔH_{comb} after ΔH_{vap} correction and before O₂ overbinding correction, while the ordinate values are the experimentally known ΔH_{comb} . Regression equation: $y = 1.070922x + 24.998307$, $R^2 = 0.99$. Dark green symbols: training set and light green symbols: test set.

specific energy for terpenes and their saturated forms. Inspection of the figure shows an increase in the heat released during combustion of the hydrogenated terpenes.

Enthalpy of formation

The computed enthalpies of formation can be used as a guide in determining the viability of the biofuel molecules, as higher enthalpies of formation are favorable. The ΔH_f values were calculated using the *ab initio* G4-MP2 (ref. 49–51) method (Table S12[†]), with a deviation of less than 2 kcal mol^{−1} from the known experimental values. We note that not much experimental data were available for this property. Fig. 4 presents the ΔH_f values for terpenes alongside the experimental data. For the complete list of molecules, see Table S12.[†]

The computed enthalpies of formation can also serve as an input for EXPLO5 software.^{66,67} The EXPLO5 computer program is based on the chemical equilibrium steady-state model of detonation that predicts the performance of ideal high explosives, propellants and pyrotechnic mixtures. Further explanation about this program and these calculations can be found in the ESI (Table S13[†]). This might indicate whether a terpene can serve as a potential explosive, in addition to serving as a biofuel.

Cetane number

The cetane number values were calculated using a MLR with the following three variables: the logarithmic expression of the reaction rate constant (causing the “chemical delay”, calculated using the energy barrier of the radical-forming, rate determining step in Scheme 2), the calculated enthalpies of vaporization, and the computed heat capacity at a constant volume (causing the “physical delay”), given by the Gaussian electronic structure calculations. The rate constant was calculated at 1000 K to mimic the engine temperature.⁵⁴

It is worth noting that experimental cetane data are not always reliable, as uniform literature values often do not exist and the purity of some of the tested compounds is unknown. Also, peroxides found in many compounds are known to affect the measured cetane number.^{52,68}

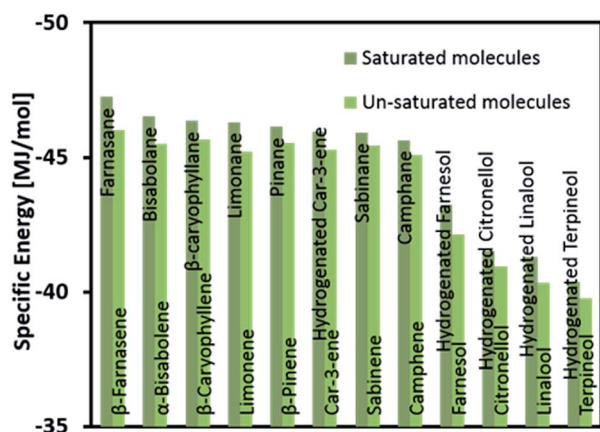


Fig. 3 Computed specific energy for unsaturated and saturated terpenes.

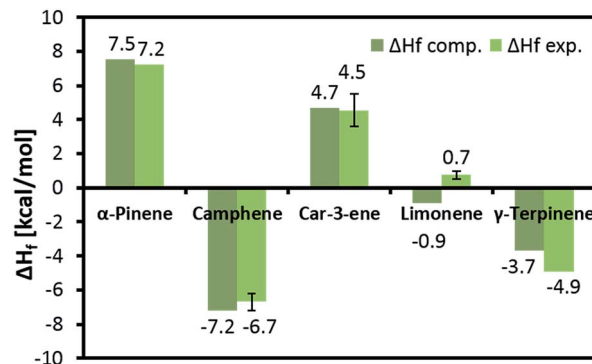


Fig. 4 Experimental and computed enthalpy of formation using the *ab initio* G4MP2 method.

Table 2 presents the values used to build the MLR, and the calculated data are compared with available experimental data. The full list of the predicted cetane number values is available in Table S14a.[†]

To estimate the relative importance of chemical and physical delay, we compute the average $\sum_{i=1}^3 c_i x_i$ for the training set in

Table 2, where c_i and x_i are the regression coefficients and variables, respectively. The average correlation values are 54, −19, and −68 for heat capacity, enthalpies of vaporization, and global log k , respectively. Based on these regression data, we conclude that the contribution of chemical delay to the cetane number is greater than that of physical delay.

Another MLR calculation was performed to check whether additional models can be useful for predicting the cetane number. In this case, we employed two variables: the log k (chemical delay) and computed enthalpy of vaporization (physical delay).

Slightly lower R^2 values (0.86) and higher standard error values were attained, but with slightly higher F distribution values. This suggests that the correlation in both models (two and three variables) is good and both can predict cetane numbers, as both contain the main factors influencing CN, namely chemical and physical delay. All statistic information can be found in Table S14b.[†]

Boiling point and vapor pressure

By correlating the computed boiling point and vapor pressure values with the experimental ones, the following correlation constants were obtained, followed by their RMSD: boiling point – $R^2 = 0.74$, RMSD = 22.3 K, which corresponds to 4.6% error (note that LR was applied to predict the boiling point values); vapor pressure – $R^2 = 0.91$, RMSD = 0.63 mbar, which corresponds to 29.3% error. The linear regression figures are presented in Fig. 5a and b. For the complete list of predicted values, see Tables S16a and b in the ESI.[†]

Discussion

In this work, we presented a combined quantum chemistry and statistical model to calculate thermodynamic and physical

Table 2 Predicted and experimental cetane numbers and associated regression data. The regression equation: $y = 1.033869x_1 - 1.348807x_2 - 6.603378x_3 + 60.143412$, $R^2 = 0.91$. y is the cetane number, x_1 is the heat capacity, x_2 is ΔH_{vap} , and x_3 is the global log k

Terpenes	Global log k (electronic energy) x_3	Calculated ΔH_{vap} [kcal mol ⁻¹] x_2	Heat capacity, C_v [cal (mol K) ⁻¹] x_1	Predicted cetane number y	Experimental cetane number	Absolute deviation
Training set						
α -Pinene	10.13	10.88	40.03	19.9	17.1 (ref. 68)	2.8
β -Pinene	10.08	10.73	39.45	19.9	19 (ref. 68)	0.9
3-Carene	9.99	11.61	41.00	20.9	27 (ref. 52)	6.1
Limonene	9.85	11.48	40.87	21.9	17.1 (ref. 68)	4.8
γ -Terpinene	11.01	12.02	41.69	14.3	18.7 (ref. 68)	4.4
β -Bisabolene	10.08	16.98	64.43	37.3	32.6 (ref. 68)	4.7
β -Caryophyllene	11.20	15.07	62.40	30.4	29 (ref. 52)	1.4
α -Farnesene	11.42	18.02	68.29	31.0	32 (ref. 68)	1.0
Farnesane	8.87	17.61	74.36	54.7	58 (ref. 52)	3.3
Test set						
Pinane	9.78	10.67	41.20	23.8	23 (ref. 69)	0.8
Limonane	9.24	11.33	44.48	29.9	29.1 (ref. 69)	0.8
β -Farnesene	10.36	17.73	67.79	37.9	32 (ref. 68)	5.9
β -Caryophyllane	9.01	14.99	65.84	48.5		
Sabinane	9.59	11.43	42.33	25.2		
Average absolute deviation						3.1
Standard deviation						2.1
RMS						3.6
Maximum absolute deviation						6.1

properties for a series of terpenes and terpenoids that are promising biofuel alternatives. Based on the current results, it seems that using a combined DFT, *ab initio*, and simple statistical approach, we can predict the thermodynamic properties of terpene molecules with reasonable accuracy. The presented predictive approach can be performed entirely *in silico*, prior to experimental work.

The computed thermodynamic and physical properties were the enthalpy of vaporization, enthalpy of combustion, enthalpy of formation, cetane number, boiling point and vapor pressure. For several of these properties, we employed DFT in conjunction with regression analysis, while for enthalpy of formation, we employed *ab initio* quantum chemistry. The agreement with the experimental data for the computed enthalpies of vaporization and combustion using DFT and continuum solvation models is good, with errors less than approx. 1 kcal mol⁻¹ and 8.9 kcal mol⁻¹, respectively. This corresponds to 8.4% and 0.3% errors for the computed enthalpies of vaporization and combustion, respectively. Predictions of the enthalpies of formation were performed with the *ab initio* G4MP2 method (without regression analysis), and the agreement with the experimental data is within approx. 1–2 kcal mol⁻¹. This level of accuracy is presumed to be sufficient to be useful for selecting promising biofuel candidates.

The cetane number results are given with a maximum deviation of 6.1. We believe that this deviation is acceptable, as the experimental values of the cetane number may vary by up to 15 units.⁵² A correlation was found between the computed specific energy values and the experimental cetane number, suggesting that for terpenes, better ignition and higher energy density go together (Table S15†).

The boiling point and vapor pressure results show a rather good correlation with the experimental data, although the maximum deviations for the boiling point and vapor pressure are 54 K and 1.4 mbar, respectively, corresponding to 10.1% and 54% errors.

We now turn our attention to the energetic considerations for terpenes as biofuels. Clearly, the stability of each molecule, as reflected in the enthalpy of formation, contributes to its enthalpy of combustion and as seen from the correlation in Table S15,† to its cetane number as well, in addition to other factors (*e.g.* non-covalent intermolecular interactions). During combustion, the compound releases its internal energy in the form of heat, and the less stable the molecule (*i.e.* higher enthalpy of formation), the more heat is released.⁷⁰

It is well established that the more branched a molecule is, the more stable it is. This is the result of the following three reasons: first, the difference in zero-point energy^{70,71} (*i.e.*, the sum of $h\nu_i/2$ for each individual vibrational mode, i), second, dispersion interactions (*i.e.* London forces) between non-bonded atoms,^{70,71} and finally hyper-conjugation.^{33,72–74} However, these effects might compete with ring strain considerations for some of the terpenes.

Monoterpenes with bi-cyclic structures, where one ring consists of less than 6-carbons (*e.g.* pinenes, sabinene and car-3-ene), have more negative combustion enthalpies than monocyclic monoterpenes consisting of only a 6-membered ring (Table S10†). This reflects a balance between branching and ring-strain, as the monocyclic ones are more stable, having less angular strain. Camphene is an example of a bi-cyclic 6-carbon ring and is the most stable monoterpene (with the least negative ΔH_{comb} and most negative ΔH_f). The stability of camphene is

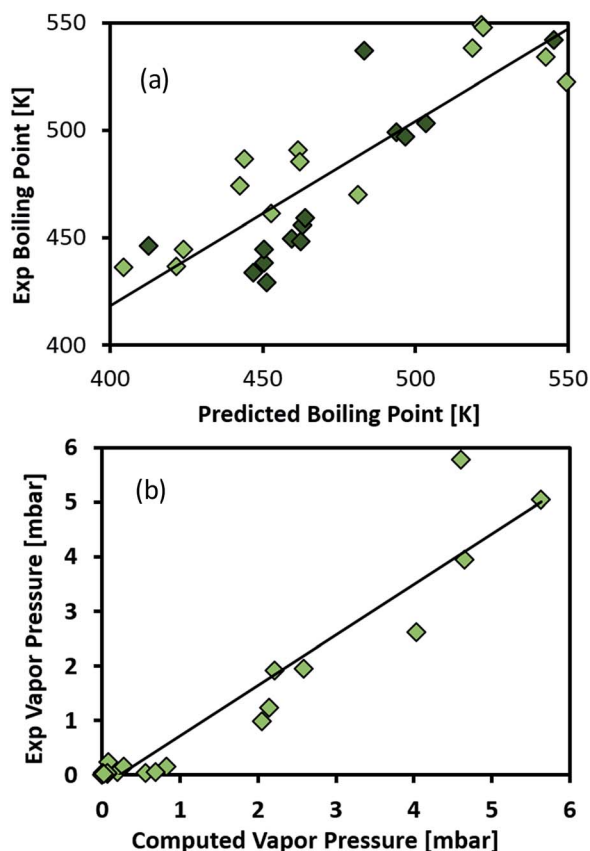


Fig. 5 (a) Correlation between the experimental and predicted data for the boiling point, $R^2 = 0.74$ and training set molecules are shown in dark green. For the training set figure, see Fig. S4a.† (b) Correlation between the experimental and computed vapor pressure, $R^2 = 0.91$ (no correlation was made for the vapor pressure data). All values were computed using COSMO-RS. Dark green symbols: training set and light green symbols: test set.

the result of a highly branched carbon skeleton without much angular and steric strain, in contrast to some other mono-terpenes that have four-membered rings. The most negative combustion enthalpies are found for the linear monoterpenes (alocimene and myrcene), as these are less branched.

For the sesquiterpenes, similarly farnesene and farnesane have the most negative enthalpy of combustion as these compounds are less branched and thus less stable. β -Caryophyllene has a 4-carbon ring with high angular strain, but is

also branched, and therefore its enthalpy of combustion is less negative.

The terpenoids overall will produce less heat during combustion (in comparison with the other terpenes), suggesting that they may not be the best choice for alternative fuels. Also, their enthalpy of formation is very negative, implying that terpenoids are more stable. This difference may be explained by bond strength considerations. For example, terpineol and limonene resemble each other, except for one C=C double bond ($146 \text{ kcal mol}^{-1}$) that is saturated to a single C-C bond (83 kcal mol^{-1}) with the addition of a C-O bond ($85.5 \text{ kcal mol}^{-1}$), O-H bond ($111 \text{ kcal mol}^{-1}$) and C-H bond (99 kcal mol^{-1}).⁷⁶ Indeed, limonene's enthalpy of formation is higher than that of terpineol ($-0.88 \text{ kcal mol}^{-1}$ vs. $-49.21 \text{ kcal mol}^{-1}$, respectively). Another possible way to explain this difference in the ΔH_{comb} values is that terpenoids already have some of the desirable C-O or O-H bonds that are formed during the combustion reaction, producing CO_2 and water. This brings the reactants closer to their product state along the reaction coordinate, reducing the reaction energy.

We also observe that saturated molecules have higher (more negative) specific energies, or energy density, than unsaturated ones. As shown in Fig. 3, after hydrogenating all terpenes, all specific energies went up by $1\text{--}2 \text{ MJ kg}^{-1}$. This means that for terpene-based biofuels, we can relate the increase in the H/C ratio⁷⁷ and increase in the fuel's molar mass to the higher fuel energy content. However, many studies suggest using terpenes in their natural, unsaturated form as a biofuel option as well.^{1,22,34,36}

It turned out that pinene dimers^{34,78,79} release less heat than the smaller mono-terpenes. Judging by the specific energy, the monoterpenes are packed with more energy and hence seem to be better suited as biofuels based on the energy content criterion.

Conclusions

Terpenes are abundant, promising biofuel materials that have many of the needed characteristics of fuels used in combustion engines. In the current work, we presented a combined theoretical and statistical model for calculating inherent fuel-related physical properties for several promising terpenes. We used statistical modeling combined with density functional theory and *ab initio* quantum chemistry methods to compute the

Table 3 Comparison of the experimental/predicted data for diesel and biodiesel fuels with the experimental/predicted values for most energy dense terpenes

Fuel property	Diesel ¹	Biodiesel, no. 1-B grade ¹	Farnesane	Bisabolane	β -Caryophyllene
Density g cm^{-3} @15 °C	0.851	0.875	0.77 ^a	0.82 (ref. 40)	0.9 ^b
Kinetic viscosity $\text{mm}^2 \text{s}^{-1}$ @40 °C	1.3–4.1	4.0–6.0	3.53 (@20 °C) ^a	2.91 (ref. 40)	—
Boiling point °C	180–340	315–350	249.1 ^b	265.3 ^b	264 (ref. 75)
Flash point °C	60–80	100–170	103.1 ^b	111 (ref. 40)	104.9 ^b
H/C ratio	1.79	1.86	2.13	2.00	1.87
Cetane number	40–55	48–65	58 (ref. 52)	41.9 (ref. 40)	29 (ref. 52)

^a <http://echa.europa.eu>. ^b <http://www.chemspider.com> (predicted using ACD/Labs software).

enthalpy of combustion, enthalpy of vaporization, enthalpy of formation, cetane number, boiling point and vapor pressure for a range of terpenes, with good accuracy. The current *in silico* strategy presents a promising strategy for finding suitable petroleum substitutes, while avoiding costly experimental trial and error approaches.

Overall, the terpenes that show the highest specific energy are the most linear and the saturated ones, while the terpenoids have overall lower energy density than the terpenes. Specifically, the most energy dense terpene is farnesane^{34,40,80,81} (hydrogenated α/β -farnesene, H/C ratio of 2.13) with -47.3 MJ kg^{-1} , followed by bisabolane^{34,35,40} (hydrogenated α/β -bisabolene, H/C ratio of 2.00) with -46.5 MJ kg^{-1} and hydrogenated β -caryophyllene^{34,36} (H/C ratio of 1.87) with -46.4 MJ kg^{-1} . These are $\text{C}_{15}\text{H}_{32}$, $\text{C}_{15}\text{H}_{30}$, and $\text{C}_{15}\text{H}_{28}$ -branched sesquiterpenes. In Table 3, we compare the experimentally determined properties for diesel and commercial biodiesel fuels, and the most promising terpenes. Importantly, the flash point, boiling point, and density of the terpenes match those of commercial biodiesel and diesel fuels. These terpene molecules are being investigated as biofuel candidates, both in their saturated and unsaturated forms.^{34,36,82}

Conflicts of interest

There are no conflicts to declare.

Author contributions

Dan T. Major planned the research. Efrat Pahima conducted the research. Dan T. Major and Efrat Pahima wrote the manuscript. Shmaryahu Hoz assisted in discussions of the results. Moshe Ben-Tzion conducted EXPLO5 calculations. All authors have given approval to the final version of the manuscript.

Acknowledgements

This work has been supported by the Israeli Ministry of Science and the Israel Science Foundation (Grant # 2146/15). The authors thank Prof. Hanoch Senderowitz for helpful discussions.

References

- 1 T. L. Alleman and R. L. McCormick, *Biodiesel Handling and Use Guide*, 5th edn, National Renewable Energy Laboratory, 2016.
- 2 A. J. Ragauskas, C. K. Williams, B. H. Davison, G. Britovsek, J. Cairney, C. A. Eckert, W. J. Frederick Jr, J. P. Hallett, D. J. Leak, C. L. Liotta, J. R. Mielenz, R. Murphy, R. Templer and T. Tschaplinski, *Science*, 2006, **311**, 484–489.
- 3 H. R. Beller, T. S. Lee and L. Katz, *Nat. Prod. Rep.*, 2015, **32**, 1508–1526.
- 4 J. Yanowitz, M. A. Ratcliff, R. L. McCormick, J. D. Taylor and M. J. Murphy, *National Renewable Energy Laboratory, Compendium of Experimental Cetane Numbers*, 2017.
- 5 D. Tilman, R. Socolow, J. A. Foley, J. Hill, E. Larson, L. Lynd, S. Pacala, J. Reilly, T. Searchinger, C. Somerville and R. Williams, *Science*, 2009, **325**, 270–271.
- 6 S. K. Lee, H. Chou, T. S. Ham, T. S. Lee and J. D. Keasling, *Curr. Opin. Biotechnol.*, 2008, **19**, 556–563.
- 7 J. L. Fortman, S. Chhabra, A. Mukhopadhyay, H. Chou, T. S. Lee, E. Steen and J. D. Keasling, *Trends Biotechnol.*, 2008, **26**, 375–381.
- 8 M. Hannon, J. Gimpel, M. Tran, B. Rasala and S. Mayfield, *Biofuels*, 2010, **1**, 763–784.
- 9 M. E. Himmel, S.-Y. Ding, D. K. Johnson, W. S. Adney, M. R. Nimlos, J. W. Brady and T. D. Foust, *Science*, 2007, **315**, 804–807.
- 10 P. P. Peralta-Yahya, F. Zhang, S. B. del Cardayre and J. D. Keasling, *Nature*, 2012, **488**, 320–328.
- 11 J. Gershenzon and N. Dudareva, *Nat. Chem. Biol.*, 2007, **3**, 408–414.
- 12 E. J. Steen, Y. Kang, G. Bokinsky, Z. Hu, A. Schirmer, A. McClure, S. B. Del Cardayre and J. D. Keasling, *Nature*, 2010, **463**, 559–562.
- 13 M. A. Rude, T. S. Baron, S. Brubaker, M. Alibhai, S. B. Del Cardayre and A. Schirmer, *Appl. Environ. Microbiol.*, 2011, **77**, 1718–1727.
- 14 F. Zhang, S. Rodriguez and J. D. Keasling, *Curr. Opin. Biotechnol.*, 2011, **22**, 775–783.
- 15 J. E. Martínez, V. Raghavan, F. González-Andrés and X. Gómez, *Int. J. Mol. Sci.*, 2015, **16**(5), 9385–9405.
- 16 J. L. Fortman, S. Chhabra, A. Mukhopadhyay, H. Chou, T. S. Lee, E. Steen and J. D. Keasling, *Trends Biotechnol.*, 2008, **26**, 375–381.
- 17 F. Zhang, J. M. Carothers and J. D. Keasling, *Nat. Biotechnol.*, 2012, **30**, 354–359.
- 18 E.-B. Goh, E. E. K. Baidoo, H. Burd, T. S. Lee, J. D. Keasling and H. R. Beller, *Metab. Eng.*, 2014, **26**, 67–76.
- 19 P. Hellier, L. Al-Haj, M. Talibi, S. Purton and N. Ladommatos, *Fuel*, 2013, **111**, 670–688.
- 20 R. Mewalal, D. K. Rai, D. Kainer, F. Chen, C. Külheim, G. F. Peter and G. A. Tuskan, *Trends Biotechnol.*, 2016, **35**, 227–240.
- 21 R. Firn, *Nature's chemicals: the natural products that shaped our world*, 2010.
- 22 R. F. Sawyer, *Environ. Health Perspect.*, 1993, **101**, 5–12.
- 23 J. Hill, E. Nelson, D. Tilman, S. Polasky and D. Tiffany, *Proc. Natl. Acad. Sci. U. S. A.*, 2006, **103**, 11206–11210.
- 24 C. B. Granda, L. Zhu and M. T. Holtzapple, *Environ. Prog.*, 2007, **26**, 233–250.
- 25 A. Hull, I. Golubkov, B. Kronberg, T. Marandzheva and J. van Stam, An Alternative Fuel for Spark Ignition Engines, 2006, 7(3), 203–214.
- 26 G. Knothe, *Fuel Process. Technol.*, 2005, **86**, 1059–1070.
- 27 A. W. Schirmer, M. A. Rude and S. A. Brubake, *US Pat.*, 8,846,371, 2014.
- 28 Z. Hu and V. Arlogadda, *US Pat.*, 9,068,201, 2015.
- 29 D. L. Greeneld, A. W. Schirmer, E. J. Clarke, E. S. Groban, B. M. DaCosta and Z. Hu, *US Pat.*, 200150064782, 2015.
- 30 A. Schirmer, M. A. Rude, X. Li, E. Popova and S. B. del Cardayre, *Science*, 2010, **329**, 559–562.
- 31 M. O. Park, K. Heguri, K. Hirata and K. Miyamoto, *J. Appl. Microbiol.*, 2005, **98**, 324–331.

- 32 R. M. Phelan, O. N. Sekurova, J. D. Keasling and S. B. Zotchev, *ACS Synth. Biol.*, 2015, **4**, 393–399.
- 33 C. R. Kemnitz, J. L. Mackey, M. J. Loewen, J. L. Hargrove, J. L. Lewis, W. E. Hawkins and A. F. Nielsen, *Chem.–Eur. J.*, 2010, **16**, 6942–6949.
- 34 J. Lane, *9 Advanced Molecules that Could Revolutionize Jet and Missile Fuel*, 2014.
- 35 T. S. Lee, P. P. Peralta-Yahya and J. D. Keasling, *US Pat.*, 9,109,175, 2013.
- 36 J. Yang, Z. Li, L. Guo, J. Du and H. J. Bae, *Renewable Energy*, 2016, **99**, 216–223.
- 37 K. Tiefenbacher and Q. Zhang, *Nat. Chem.*, 2015, **7**, 197–202.
- 38 Q. Zhang, L. Catti, J. Pleiss and K. Tiefenbacher, *J. Am. Chem. Soc.*, 2017, **139**, 11482–11492.
- 39 S. Sarria, B. Wong, H. G. Martín, J. D. Keasling and P. P. Peralta-Yahya, *ACS Synth. Biol.*, 2014, **3**, 466–475.
- 40 P. P. Peralta-Yahya, M. Ouellet, R. Chan, A. Mukhopadhyay, J. D. Keasling and T. S. Lee, *Nat. Commun.*, 2011, **2**, 1–8.
- 41 M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery Jr, J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, T. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, O. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski and D. J. Fox, *Gaussian 09*, Revision E.01, 2009.
- 42 J. W. Ochterski, *Thermochemistry in Gaussian*, 2000.
- 43 Y. Zhao and D. G. Truhlar, *Acc. Chem. Res.*, 2008, **41**, 157–167.
- 44 O. Gunnarsson and R. O. Jones, *Phys. Rev. B*, 1985, **31**, 7588–7602.
- 45 L. Wang, T. Maxisch and G. Ceder, *Phys. Rev. B*, 2006, **73**, 1–6.
- 46 V. L. Chevrier, S. P. Ong, R. Armiento, M. K. Y. Chan and G. Ceder, *Phys. Rev. B*, 2010, **82**, 1–11.
- 47 P. Mori-sánchez, A. J. Cohen and W. Yang, *J. Chem. Phys.*, 2006, **125**, 1–4.
- 48 A. V. Marenich, C. J. Cramer and D. G. Truhlar, *J. Phys. Chem. B*, 2009, **113**, 6378–6396.
- 49 L. A. Curtiss, P. C. Redfern and K. Raghavachari, *J. Chem. Phys.*, 2007, **127**, 1–8.
- 50 P. C. Redfern, P. Zapol, L. A. Curtiss and K. Raghavachari, *J. Chem. Phys.*, 2000, **104**, 5850–5854.
- 51 D. Doron, D. T. Major, A. Kohen, W. Thiel and X. Wu, *J. Chem. Theory Comput.*, 2011, **7**, 3420–3437.
- 52 J. Yanowitz, M. A. Ratcliff, R. L. McCormick, J. D. Taylor and M. J. Murphy, *NREL – Compendium of Experimental Cetane Numbers Compendium of Experimental Cetane Numbers*, 2017.
- 53 W. Chen, S. Shuai and J. Wang, *Energy Fuels*, 2010, **24**, 856–862.
- 54 H. Abou-Rachida, K. El Marrounia and S. Kaliaguinea, *J. Mol. Struct.*, 2003, **631**, 241–250.
- 55 C. C. Pye, T. Ziegler, E. van Lenthe and J. N. Louwen, *Can. J. Chem.*, 2009, **87**, 790–797.
- 56 A. Klamt, V. Jonas, T. Bürger and J. C. W. Lohrenz, *J. Phys. Chem. A*, 1998, **102**(26), 5074–5085.
- 57 A. Klamt, *J. Phys. Chem.*, 1995, **99**, 2224–2235.
- 58 C. C. Pye and T. Ziegler, *Theor. Chem. Acc.*, 1999, **101**, 396–408.
- 59 S. T. Lin, J. Chang, S. Wang, W. A. Goddard and S. I. Sandler, *J. Phys. Chem. A*, 2004, **108**, 7429–7439.
- 60 D. T. Major, *ACS Catal.*, 2017, **7**, 5461–5465.
- 61 M. Dixit, M. Weitman, J. Gao and D. T. Major, *ACS Catal.*, 2017, **7**, 812–818.
- 62 M. Weitman and D. T. Major, *J. Am. Chem. Soc.*, 2010, **132**, 6349–6360.
- 63 D. T. Major, Y. Freud and M. Weitman, *Curr. Opin. Chem. Biol.*, 2014, **21**, 25–33.
- 64 C. Gonzalez and H. B. Schlegel, *J. Phys. Chem.*, 1990, **94**, 5523–5527.
- 65 K. Fukui, *Acc. Chem. Res.*, 1981, **14**, 363–368.
- 66 M. Sućeska, in *Proc. of 30th Int. Annual Conference of ICT*, Karlsruhe, 1999, pp. 50/1–50/14.
- 67 M. Sućeska, in *Proc. of 32nd Int. Annual Conference of ICT*, Karlsruhe, 2001, pp. 110/1–110/13.
- 68 M. Dahmen and W. Marquardt, *Energy Fuels*, 2015, **29**, 5781–5801.
- 69 B. G. Harvey, W. W. Merriman and T. A. Koontz, *Energy Fuels*, 2015, **29**, 2431–2436.
- 70 E. V. Anslyn and D. A. Dougherty, *Modern Physical Organic Chemistry*, 2006, vol. 43.
- 71 K. B. Wiberg, *Angew. Chem., Int. Ed. Engl.*, 1986, **25**, 312–322.
- 72 D. H. Ess, S. Liu and F. De Proft, *J. Phys. Chem. A*, 2010, **114**, 12952–12957.
- 73 J. K. Badenhoop and F. Weinhold, *J. Chem. Phys.*, 1997, **107**, 5406.
- 74 J. K. Badenhoop and F. Weinhold, *Int. J. Quantum Chem.*, 1999, **72**, 269–280.
- 75 M. Hoskovec, D. Grygarová, J. Cvačka, L. Streinz, J. Zima, S. P. Verevkin and B. Koutek, *J. Chromatogr. A*, 2005, **1083**, 161–172.
- 76 S. J. Blanksby and G. B. Ellison, *Acc. Chem. Res.*, 2003, **36**, 255–263.
- 77 L. Yue, G. Li, G. He, Y. Guo, L. Xu and W. Fang, *Chem. Eng. J.*, 2016, **283**, 1216–1223.
- 78 B. G. Harvey, M. E. Wright and R. L. Quintana, *Energy Fuels*, 2010, **24**, 267–273.
- 79 B. G. Harvey, M. E. Wright and R. L. Quintana, *US Pat.* 8,227,651, 2012.
- 80 J. A. Ryder, *US Pat.* 7,589,243, 2009.
- 81 N. S. Renninger and D. J. McPhee, *US Pat.* 7,399,323, 2008.
- 82 S. H. Won, S. Dooley, P. S. Veloo, H. Wang, M. A. Oehlschlaeger, F. L. Dryer and Y. Ju, *Combust. Flame*, 2014, **161**, 826–834.