Scientific
Research
Publishing

Biological Activity Relationship of Cyclic and Noncyclic Alkanes Using Quantum Molecular Descriptors

Ahmad Nazib Alias*, Zubainun Mohamed Zabidi, Nurul Aimi Zakaria, Zaidatul Salwa Mahmud, Rosliza Ali

Faculty of Applied Sciences, Universiti Teknologi MARA Cawangan Perak, Tapah Road, Malaysia
Email: *ahmadnazib111@uitm.edu.my

Abstract

A 3-Dimension-Quantitative Structure-Activity Relationship (3D-QSAR[1]) approach is applied for the prediction of accurate chemical products made from biological activity and toxicity. Quantum chemical technique allows the construction of the molecular descriptors. The molecular quantum descriptors are classified into five principal component factors. Various linear regression equations are obtained using the statistical technique. In this study, the researchers propose the three best regression equations based on quantum molecular descriptors discussed earlier in this study. The observed EC50 vs calculated EC50 is plotted using the best fitting with the quantum descriptors.

Keywords

3D-QSAR, Quantum Molecular Descriptor, Biological Activity, Toxicology

1. Introduction

Industrial advancement has resulted in the use of large amounts of chemicals entering the human life cycle. According to the American Chemistry Council 2020, the business of chemistry is a large user of natural gas and petroleum products [1]. The petroleum products include the saturated hydrocarbons or alkanes. The saturated hydrocarbon falls into a large n-linear, branch, and cyclic alkanes. The physical phase alkane series consists of gases (methane, ethane, propane, and butanes), liquid phase from pentanes to hexadecane, and longer chain solids.

---

[1]QSAR is a mathematical modelling that using the machine learning as a tool in the modelling development.

These saturated hydrocarbons have been the subject of extensive attention because they encompass one of the foundations of petrochemistry. In petrochemistry, alkanes are the major constituents of natural gas and crude oil [2]. For example, pentane is mostly used in the formulation of gasoline. On the other hand, hexane is well known to be used as a solvent in glues, varnishes, cements, and other product. Meanwhile, 2-methylpropane is used in principal feedstock of alkylation units of refineries [3]. 2-methylpropane, 2-methylbutane, n-Butane, and propane are widely used in the cosmetic industry as aerosol propellants [4]. The day-to-day life demand for alkane products may impact the environment, especially in terms of toxicity to aquatic life and subsequently to biodiversity [5]. Furthermore, when inhaled, this hydrocarbon can cause central nervous system depression [6]. Liquid hydrocarbon such as pentane, hexane, and octane is also solvents for fats. When these liquid hydrocarbons are in contact with the skin, they are capable of removing fat, and resulting in dryness, scaling, and skin inflammations [7]. Surprisingly, the exposure to n-hexane or solvent containing a high concentration of n-hexane shows polyneuropathy syndrome [8].

Upon realizing this issue, a method for biological activity and toxicity assessment needs to be investigated on alkane-based compounds on the molecular structure. According to Phillips and his co-workers, the prediction of the toxicant using QSAR method must use chemicals with the same group compound; otherwise, the different toxicant actions involved will limit the prediction of toxicity [9]. Getting toxicants with the same group compounds is very challenging in QSAR, due to the limitation of retrieving experimental data. The mechanisms involved in biological action are also too complex, and it is inadequate to describe it using a model alone. Therefore, this study investigates the biological activity and toxicity of alkane groups. In this study, the relationship of molecular structure and biological activity and toxicity are investigated based on the molecular descriptors. The molecular descriptors used in this study include chemical hardness, chemical potential, dipole moment, non-linear polarizability, and graph energy index. These molecular descriptors are known as non-linear molecular properties. The molecular has been calculated using semi-empirical calculation due to fast comparison with another *ab initio* method and low cost in terms of computer resource [10]. The major limitations of semi-empirical are some parameters that may not be available especially for transition metal [11]. However, in this study, we only employ the alkanes which are in the range of the calculation. The quantum molecular descriptor relationships are employed to predict the unmeasured values of the considered properties of compounds. Furthermore, the relationships can be extended to design non-existent structures possessing some desirable properties. The volatile component of the alkane from the biological experiment using chromatography is also investigated. To best our knowledge, there are non-articles on 3-D QSARs examining the biological effects and toxicology of alkanes using the semi-empirical method as a molecular descriptor, especially on non-linear molecular properties.

2. Methods and Calculation

2.1. Calculation of the Quantum Molecular Descriptor

Semi-empirical quantum chemical method is the computational method to calculate the electronic and molecular orbital properties using self-consistent field molecular orbital theory. The integration method employs parametric method 6 (PM6). This quantum molecular descriptor was calculated using semi-empirical MOPAC2016, Version: 21.002 James J. P. Stewart software [12]. The input of the structure and geometry optimization was generated using Avogadro version 1.2.0. The geometry optimization used force field method MMFF94s with step per update 4. The molecular structure was generated using 3-dimension Avogadro interface, and the molecular descriptor involved three coordinates. Therefore, this method employed the 3D-QSAR approach. The calculation was done using PM6 parameters.

2.2. Definitions of the Quantum Molecular Descriptors Used

α_{ij}: the dynamic linear polarizability constant at $\omega = 0.25$ eV. The value of i and j represented tensor vector in x, y or z.

β_i: the average value of the hyperpolarizability at $\omega = 0.25$ eV. The quantity of interest is defined as $\beta_i = [\beta_{iii} + \beta_{ijj} + \beta_{ikk}]$ [13].

γ_{ijkl}: the second order hyperpolarizability at $\omega = 0.25$ eV. The value i, j, k and l represented tensor vector in x, y and z.

$\varepsilon_{\text{HOMO}}$: the energy of the highest occupied molecular orbital.

$\varepsilon_{\text{LUMO}}$: the energy of the lowest unoccupied molecular orbital.

d_p: the molecular dipole moment contributed from the net charge density [14].

d_{hyb}: the molecular dipole moment contributed from a hybridization of molecular orbital [14].

COS: the Conductor like screening model (COSMO) surface area.

μ: the chemical potential. The chemical potential can defined as, $\mu = (\varepsilon_{\text{HOMO}} - \varepsilon_{\text{LUMO}})/2$.

η: the molecular chemical hardness. The η can be defined as, $\eta = (\varepsilon_{\text{HOMO}} + \varepsilon_{\text{LUMO}})/2$.

W: the electrophilicity index. The electrophilicity is defined by Parr *et al.* that is $W = \mu^2/2\eta$ [15].

W^-: the electrodonating power. This index explain chemical response to the donation of electron which given by $(3\varepsilon_{\text{HOMO}} - \varepsilon_{\text{LUMO}})^2/16(\varepsilon_{\text{HOMO}} + \varepsilon_{\text{LUMO}})$ [16].

GE^1: the graph energy index. The graph energy index is the sum of eigen value. The index is defined as $GE^1 = \sum \left| -\lambda_i \right|$ where λ is the negative eigen value.

GE^2: the graph energy index with positive eigen value $GE^2 = \sum \left| +\lambda_i \right|$ where λ is the positive eigen value.

Est: the Estrada index. The Estrada energy index is defined as $Est = \sum \exp(\lambda_i)$ [17].

2.3. Data Preparation and Analysis

The chemical data was collected using the Reaxys database. Reaxys is a well-known curated database for finding and comparing relevant chemical information from different sources that attempt to bring together chemical, biological activity and toxicogenomic into their center location [18] [19]. In this study, all biological activity and toxicology values were extracted from a single reference. This was done to ensure a compatible relationship among them. The data used in this study are presented in the appendix section [20]-[26]. Inferential statistical analyses were performed. Multiple linear regression was applied to generate the prediction models. The multiple linear regression is a relatively simple approach pursuit to model the linear relationship between both independent and explanatory variables by fitting the linear equation. Variable selection in a block was entered and calculated using a single step. The researchers performed regression using SPSS statistical software. The chosen regression was based on a good fitting with the regression equation's standard of deviation not less than 0.0001. The observed plot vs calculated toxicity with a 95% degree of confidence was plotted using Minitab. The descriptors had very small standard deviations, being inadequate to represent the changes in electronic structure from molecule to molecule [27]. Principal Component Analysis (PCA) was used to determine the inherent dimensionality of groups of molecular descriptor properties. The orthogonal PROMAX rotation was used during the PCA analysis.

3. Results and Discussion

3.1. Quantum Molecular Descriptor

Alkanes are hydrocarbons derived from petroleum processing streams. These alkanes contain only carbon and hydrogen atoms. It contains carbon numbers ranging from approximately C5 - C20 with three-type constituents: normal paraffins, isoparaffins, and cycloparaffins. These constituents result in different physical and chemical properties [28]. **Tables 1-3** show the electronic and molecular orbital for alkanes generated from the semi-empirical calculation. To understand the factor contribution of each descriptor, the researchers plotted the component analysis based on ProMax rotation as shown in **Figure 1**. Five principal component factors are classified. The first component is the dynamic linear polarizability constant (a_{xx}, a_{yy} and a_{zz}), conductor-like screening model (COSMO), Ge^1 Ge^2 second order hyperpolarizability of γ_{xxxx} and γ_{xxyy}. The second component is the average value of the hyperpolarizability of β_z, hybrid and the net charge dipole moment. The third component is the second order hyperpolarizability of γ_{yyyy} and γ_{zzzz}. The fourth component is the average value of hyperpolarizability of β_x and β_y. The last component is the electrophilicity index, second order hyperpolarizability of γ_{yyyy} and γ_{yyzz}.

Inside each cluster, the molecular properties differed for their dimension, polarizability, and isomerization. The occurrence of various descriptor components was clustered using a different classification of molecular electronic property

Table 1. The value of quantum descriptor of alkanes.

Chemical Structure	COSMO Area A^2	η	ω	μ_{pc}	μ_h
n-Decane	234.87	7.443	0.793	0.047	0.120
2-Methylnonane	232.49	7.418	0.783	0.082	0.183
3-Methylnonane	229.96	7.385	0.773	0.062	0.126
2,6-Dimethyloctane	226.65	7.383	0.757	0.129	0.253
3,3-Dimethyloctane	220.98	5.354	2.638	0.006	0.006
4-n-Propylheptane	225.51	7.313	0.723	0.126	0.228
2,3,6-Trimethylheptane	212.64	7.296	0.718	0.098	0.184
3,4,5-Trimethylheptane	206.55	7.282	0.692	0.139	0.224
n-Undecane	257.24	7.399	0.801	0.034	0.107
4-Methyldecane	250.07	7.346	0.768	0.059	0.130
2,3-Dimethylnonane	244.82	7.309	0.758	0.051	0.085
n-Dodecane	277.63	7.374	0.798	0.002	0.001
n-Tridecane	297.49	7.356	0.796	0.035	0.108
2,5-Dimethylundecane	286.32	7.302	0.753	0.042	0.074
4,7-Dimethylundecane	283.64	7.281	0.746	0.036	0.102
3,7-Dimethylundecane	283.49	7.287	0.746	0.027	0.013
4,6-Dimethylundecane	281.24	7.272	0.741	0.142	0.281
3,5-Dimethylundecane	282.4	7.275	0.744	0.108	0.212
5,7-Dimethylundecane	281.14	7.271	0.740	0.102	0.217
n-Tetradecane	317.71	7.341	0.794	0.002	0.001
6-Methyltridecane	310.69	7.295	0.767	0.067	0.195
Cyclopentane	120.54	7.488	0.825	0.004	0.006
1,2-Dimethylcyclopentane	154.9	7.331	0.813	0.080	0.055
Cyclohexane	133.05	7.507	0.712	0.000	0.000
Methylcyclohexane	151.34	7.334	0.737	0.065	0.157
Methylcycloheptane	138.31	7.417	0.839	0.060	0.135
1,4-Dimethylcyclohexane	169.72	7.185	0.731	0.083	0.208

Table 2. The value of quantum descriptor of alkanes.

Chemical Structure	α_{xx} (0.25 eV)	α_{yy}	α_{zz}	γ_{xxxx}	γ_{yyyy}	γ_{zzzz}
n-Decane	137.227	123.988	123.286	10,921.916	370.927	472.400
2-Methylnonane	138.181	122.200	123.837	13,689.352	382.056	299.579
3-Methylnonane	138.146	123.903	122.015	13,384.662	566.525	435.113
2,6-Dimethyloctane	136.550	124.163	122.271	10,934.942	1517.305	367.384
3,3-Dimethyloctane	136.465	122.664	123.542	11,264.970	849.936	490.718
4-n-Propylheptane	131.754	129.602	121.610	5420.640	4747.696	652.743

Continued

2,3,6-Trimethylheptane	133.270	126.719	121.217	5779.869	2742.074	676.217
3,4,5-Trimethylheptane	132.223	125.111	123.388	4582.494	1868.465	1008.981
n-Undecane	153.880	135.314	133.121	18,831.790	185.117	104.268
4-Methyldecane	152.032	135.764	133.303	15,795.449	836.608	469.570
2,3-Dimethylnonane	150.861	137.008	132.783	13,918.112	1954.583	706.052
n-Dodecane	167.867	146.983	144.531	21,494.591	243.207	119.222
n-Tridecane	182.013	158.564	155.896	24,390.277	200.777	91.155
2,5-Dimethylundecane	178.548	159.852	155.880	18,772.366	1755.394	836.012
4,7-Dimethylundecane	179.218	159.458	155.725	18,924.502	1684.924	820.365
3,7-Dimethylundecane	178.704	158.700	156.803	18,638.242	540.801	621.957
4,6-Dimethylundecane	177.501	160.140	156.321	15,596.389	1965.595	669.194
3,5-Dimethylundecane	178.233	158.961	156.764	17,062.351	1683.910	1006.581
5,7-Dimethylundecane	177.690	159.673	156.595	15,720.519	1827.475	849.678
n-Tetradecane	196.108	170.201	167.284	27,241.110	222.666	100.396
6-Methyltridecane	194.377	170.589	167.568	23,891.478	900.386	481.141
Cyclopentane	63.057	63.067	60.330	793.376	793.704	127.117
1,2-Dimethylcyclopentane	88.172	88.324	82.775	2076.560	2255.927	297.892
Cyclohexane	76.319	76.378	72.404	2043.471	2059.324	519.776
Methylcyclohexane	89.488	88.489	83.451	3062.855	2310.723	517.874
Methylcycloheptane	76.408	74.982	71.604	1902.674	1039.159	224.067
1,4-Dimethylcyclohexane	103.724	98.981	94.418	4451.934	1393.516	401.794

Table 3. The value of quantum descriptor of alkanes.

Chemical Structure	γ_{xxyy}	γ_{xxzz}	γ_{yyzz}	Ge^-	Ge^+	E_i
n-Decane	1237.583	565.863	661.277	511.293	180.572	14,598.726
2-Methylnonane	−223.237	1069.741	715.335	511.122	181.083	14,968.295
3-Methylnonane	1089.874	−131.352	504.704	510.940	181.194	15,016.888
2,6-Dimethyloctane	465.877	714.942	351.140	510.416	182.034	15,574.127
3,3-Dimethyloctane	−103.587	1076.307	655.080	509.976	182.612	15,995.395
4-n-Propylheptane	1372.817	341.779	42.212	510.276	181.603	15,271.279
2,3,6-Trimethylheptane	1009.459	486.336	176.131	509.049	183.481	16,402.403
3,4,5-Trimethylheptane	1131.160	808.387	392.264	508.514	183.541	16,448.644
n-Undecane	1103.336	−387.544	664.719	561.245	197.427	15,742.815
4-Methyldecane	1258.864	−114.797	455.156	560.517	198.382	16,321.782
2,3-Dimethylnonane	1288.221	−132.765	167.095	559.885	199.135	16,864.032
n-Dodecane	1245.634	−424.550	713.448	610.819	214.603	17,047.743
n-Tridecane	1340.749	−471.636	797.054	660.391	231.772	18,341.952
2,5-Dimethylundecane	1622.389	−68.747	262.380	659.055	233.620	19,490.018

Continued

4,7-Dimethylundecane	1617.831	−70.734	247.355	658.835	233.684	19,523.875
3,7-Dimethylundecane	1565.988	15.874	891.825	658.698	233.680	19,464.326
4,6-Dimethylundecane	1867.661	709.949	243.295	658.640	233.825	19,594.046
3,5-Dimethylundecane	1360.275	685.151	114.204	658.659	233.769	19,573.046
5,7-Dimethylundecane	1699.145	868.613	156.785	658.593	233.833	19,600.918
n-Tetradecane	1448.013	−509.036	849.137	709.959	248.929	19,637.477
6-Methyltridecane	1685.807	−114.738	620.085	709.176	249.915	20,230.303
Cyclopentane	309.504	69.021	70.600	247.719	86.083	6575.445
1,2-Dimethylcyclopentane	580.484	142.377	66.542	346.347	121.590	9887.844
Cyclohexane	680.196	−29.347	−41.162	296.21047	104.207	8154.244
Methylcyclohexane	974.572	−47.446	29.353	345.64672	121.866	9792.494
Methylcycloheptane	491.749	103.323	62.678	297.11836	103.776	8215.064
1,4-Dimethylcyclohexane	1167.114	−33.403	250.275	394.91231	139.494	11,454.047
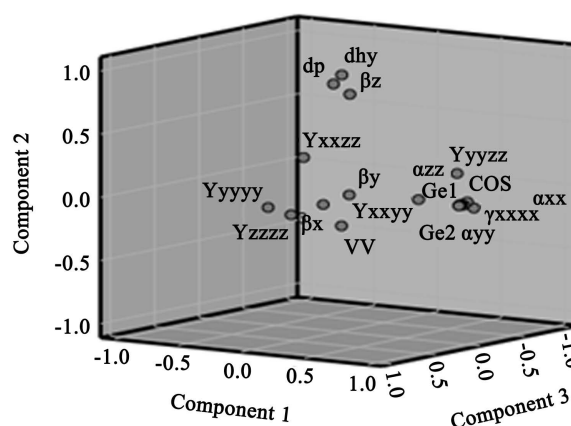
Figure 1. Component plot for quantum molecular descriptor of alkanes.

descriptors, based on the calculation of the molecular spherical coordinate system, the localization of electron, oscillating electrical field, and electron-electron interaction. The cumulative total variance of all five components is 85.136%, by which most of the contributions came from component 1 (40.434%). This contribution might be classified according to the position of carbon atoms to the molecule in the spherical coordinate system based on self-consistency molecular orbital [29]. The second principal contribution is 19.363% which is plausibly related to the orientation of dipole moment in the molecule [30]. The third contribution for the cumulative total variance is 13.115%. The fourth and fifth contributions are 6.242% and 5.982% respectively.

3.2. Quantitative Structure Retention Relationship (QSRR) Analysis

The biological activity is normally related with the molecular interaction that

goes through the compound during the transport through biological membranes, or in the reaction with the active site. This interaction is strongly related to the molecular chemical structure [20]. A change in the structure can result in a change in biological response. One of the techniques to measure the biological response is by manifesting a change in the chromatographic retention data. The prediction of chromatography retention indexes can also be explained using quantum chemical descriptors or parameters. The volatile component of Scorzonera hispanica L has been demonstrated by a wide variety of molecules [21]. The statistical analysis yields the following results in which n is the number of samples, r and r^2 are the correlation coefficients, and s is the standard deviation. In this study, the researchers only discussed three best regression Equations (with the highest value of r and r^2) that are given in Equations (1a)-(1c). The chromatography retention indexes for alkanes in this series were computed. The three best statistical analyses yield the following results:

$$\log Rt1 = 0.006333COS\left(\pm0.000455\right) + 0.630636d_p\left(\pm4.127\right)$$
$$-0.477d_h\left(\pm2.15171\right) \tag{1a}$$
$$n = 25, r = 0.9742, r^2 = 0.948, s = 0.311$$

$$\log Rt1 = 0.000307\alpha_{xx}\left(\pm0.017201\right) - 0.001759\alpha_{yy}\left(\pm0.041598\right)$$
$$+0.014028\alpha_{zz}\left(\pm0.052832\right) \tag{1b}$$
$$n = 25, r = 0.970, r^2 = 0.942, s = 0.329$$

$$\log Rt1 = 0.000044\gamma_{xxxx}\left(\pm0.000014\right) + 0.000856\gamma_{zzzz}\left(\pm0.00034\right)$$
$$+0.000515\gamma_{xxyy}\left(\pm0.000210\right) - 0.000083\gamma_{xxzz}\left(\pm0.000392\right) \tag{1c}$$
$$n = 25, r = 0.9996, r^2 = 0.99916, s = 0.047$$

Linear alkanes were also used as a reference to calculate the retention index for the volatile components from marjoram oil sample [22]. The computed isotherm retention time for alkanes in this series used statistical analysis, yielding the following results:

$$\log Rt2 = 0.00375COS\left(\pm0.00002\right) + 189.6037d_p\left(\pm3.33786\right)$$
$$-58.47219d_h\left(\pm1.01753\right) \tag{2a}$$
$$n = 5, r = 0.9999, r^2 = 0.9999, s = 0.0047$$

$$\log Rt2 = 0.072612\alpha_{xx}\left(\pm0.00806\right) - 0.075058\alpha_{zz}\left(\pm0.009266\right)$$
$$n = 5, r = 0.9998, r^2 = 0.9996, s = 0.027780 \tag{2b}$$

$$\log Rt2 = 0.000041\gamma_{xxxx}\left(\pm0.00001\right) + 0.000137\gamma_{zzzz}\left(\pm0.0001\right)$$
$$+0.000356\gamma_{xxyy}\left(\pm0.000064\right) \tag{2c}$$
$$n = 5, r = 0.9996, r^2 = 0.99916, s = 0.047$$

The researchers computed the series of correlation with chromatography retention index to produce branched alkanes by Cynobaterium *Microcoleus vigantus*. [20] The correlation with quantum molecular descriptor was found, yielding the following results:

$$\log Rt3 = 0.173W\left(\pm 0.051\right) + 0.168\eta\left(\pm 0.006\right)$$
$$n = 24, r = 0.997, r^2 = 0.994, s = 0.106$$

(3a)

$$\log Rt3 = -0.03275\alpha_{xx}\left(\pm 0.00517\right) + 0.00518\alpha_{yy}\left(\pm 0.01265\right)$$
$$+ 0.04186\alpha_{zz}\left(\pm 0.01503\right)$$
$$n = 24, r = 0.997, r^2 = 0.994, s = 0.108$$

(3b)

$$\log Rt3 = 0.000082Est + \left(\pm 0.000002\right)$$
$$n = 24, r = 0.994, r^2 = 0.989, s = 0.149$$

(3c)

The obtained results show that the best regression r^2 descriptors are different between experiments. This might be due to the number of samples being different from each other. To determine the experiment modeling with the quantum chemical descriptor, the molecular structure plays an important role. Equations (1a) and (2a), and the retention index can fit with COSMO and dipole moment descriptor. They also showed the best r and r^2 values in both results. In chromatography, the retention data are proportional to the free-energy change of solute-stationary phase interactions changed by the mobile phase. This shows that the molecular interaction with the molecules affect the regression model. The molecules such as butane, 3,4-Dimethylhexane and hexane (the total dipole moment is zero) tend to induce London dispersion forces. While the molecules with the total dipole moment is not zero, influenced by van der Waals forces. The linear polarizability constant (α) and second hyperpolarizability (γ) are the physical quantities derived from dipole moment interaction with static and fluctuation of electrical field. Both quantities show a good correlation with the retention index. Therefore, the effects of charge screening in molecule and dipole moment are correlated with the retention index.

The analysis of electrophilicity index and the molecular chemical hardness (in Equation (3a)) is related to molecular electron density distribution on HOMO and LUMO that are localized in the atoms. Chemical hardness is a property that measures the stability and reactivity of a molecule [15]. Therefore, the change of molecular structure, sightly changes the electron cloud deformation which affects the retention constant. The Estrada index also gives a good regression relation in the quantitative retention structure analysis. The Estrada index is based on the topological molecular descriptor based on eigenvalues of an adjacency matrix. The Estrada index is also suitable for molecular descriptor which is comparable with the molecular topology index using the graph theory approach [31]. This index is able to represent a mathematical numeric to character the molecules, especially the isomers structures which fit the retention constant.

3.3. Toxicology Analysis

Hydrocarbons include a vast number of existing chemicals in the environment and consumerism, and there is a need to study the toxicology impact on both environment and human. Hence, an alternative approach is needed to seek the generalities of "structure-activity relationships" from the presence of toxicological data. This can be used to predict another chemical isomer with similar ef-

fects. The in vitro oxidation by reconstituted enzyme system using alkane as substrate is demonstrated [24]. This report can provide additional information on the toxicity of substrate to the whole cells. The best three statistical analyses yield the following results:

$$\log Ea = 7.537W\left(\pm 3.670\right) - 0.578\eta\left(\pm 0.381\right)$$
$$n = 24, r = 0.997, r^2 = 0.994, s = 0.106 \tag{4a}$$

$$\log Ea = -0.063886\alpha_{xx}\left(\pm 0.030311\right) + 0.086243\alpha_{zz}\left(\pm 0.033889\right)$$
$$n = 26, r = 0.951, r^2 = 0.905, s = 0.897 \tag{4b}$$

$$\log Ea = 0.00716COS\left(\pm 0.00089\right) - 1.9520d_p\left(\pm 10.1960\right)$$
$$+ 2.0527d_h\left(\pm 4.9964\right) \tag{4c}$$
$$n = 26, r = 0.948, r^2 = 0.899, s = 0.542203$$

The acute toxicology of aqueous solutions of hydrocarbon to *Daphnia magna* is reported [23]. The researchers found the correlation with quantum molecular descriptors, yielding the following results. **Figure 2** shows the correlation between the experimental and predicted LC50 values obtained from the dynamic linear polarizability descriptor with our Equation (5a):

$$\log LC50 = 0.037333\alpha_{xx}\left(\pm 0.028125\right) - 0.002384\alpha_{yy}\left(\pm 0.025083\right)$$
$$- 0.088385\alpha_{zz}\left(\pm 0.054480\right) + 5.380546\left(\pm 0.186554\right) \tag{5a}$$
$$n = 7, r = 0.9989, r^2 = 0.9977, s = 0.0694$$

$$\log LC50 = -0.006069Ge1\left(\pm 0.00286\right) - 0.01367Ge^2\left(\pm 0.00825\right)$$
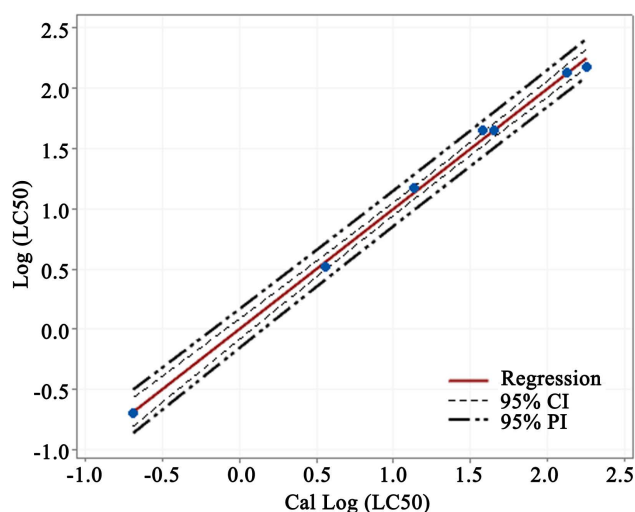$$+ 4.926273\left(\pm 0.13140\right) \tag{5b}$$
$$n = 7, r = 0.9978, r^2 = 0.9955, s = 0.083796$$

$$\log LC50 = -0.023260COS\left(\pm 0.002623\right) - 39.969094d_p\left(\pm 24.824302\right)$$
$$+ 14.012292d_h\left(\pm 9.573416\right) + 5.053425\left(\pm 0.419972\right) \tag{5c}$$
$$n = 7, r = 0.9852, r^2 = 0.9706, s = 0.2485$$

Figure 2. Plot of observed LC50 vs calculated LC50 using dynamic linear polarizability descriptor.

A number of aliphatic hydrocarbons after intravenous injections of emulsion formulations into Mice were examined [25]. The correlation with quantum molecular descriptor was found, yielding the following results. **Figure 3** shows the correlation between the experimental and predicted LD100 values obtained from the second order hyperpolarizability descriptor with our Equation (6a).

$$\log LD100 = -0.000079\gamma_{xxxx}\left(\pm 0.000047\right) + 0.00457\gamma_{yyyy}\left(\pm 0.007558\right)$$
$$+ 0.001743\gamma_{zzzz}\left(\pm 0.004965\right) - 0.004269\gamma_{xxyy}\left(\pm 0.003920\right)$$
$$- 0.002150\gamma_{xxzz}\left(\pm 0.001030\right) + 0.008519\gamma_{yyzz}\left(\pm 0.006301\right) \tag{6a}$$

$$n = 10, r = 0.9822, r^2 = 0.9647, s = 0.3100$$

$$\log LD100 = -0.012545Ge^1\left(\pm 0.009357\right) + 0.041093Ge^2\left(\pm 0.026479\right) \tag{6b}$$
$$n = 10, r = 0.9664, r^2 = 0.9340, s = 0.3000$$

$$\log LD100 = 0.00427COS\left(\pm 0.00065\right) - 18.51654d_p\left(\pm 23.85805\right)$$
$$+ 6.64021d_h\left(\pm 8.22676\right) \tag{6c}$$
$$n = 10, r = 0.9659, r^2 = 0.9329, s = 0.32297$$

The aquatic toxicity of hydrocarbon to aquatic organism was reported [26]. The correlation with quantum molecular descriptor was found, yielding the following results. **Figure 4** shows the correlation between the experimental and predicted EC50 values obtained from the dynamic linear polarizability descriptor with our Equation (5a).

$$\log EC50 = -0.000191\gamma_{xxxx}\left(\pm 0.000037\right) - 0.000503\gamma_{yyyy}\left(\pm 0.000373\right)$$
$$+ 0.000763\gamma_{zzzz}\left(\pm 0.001163\right) - 0.000364\gamma_{xxyy}\left(\pm 0.000527\right)$$
$$- 0.000278\gamma_{xxzz}\left(\pm 0.000903\right) - 0.000955\gamma_{yyzz}\left(\pm 0.001035\right) \tag{7a}$$
$$+ 2.0700\left(\pm 0.310773\right)$$

$$n = 12, r = 0.98034, r^2 = 0.961065, s = 0.233098$$
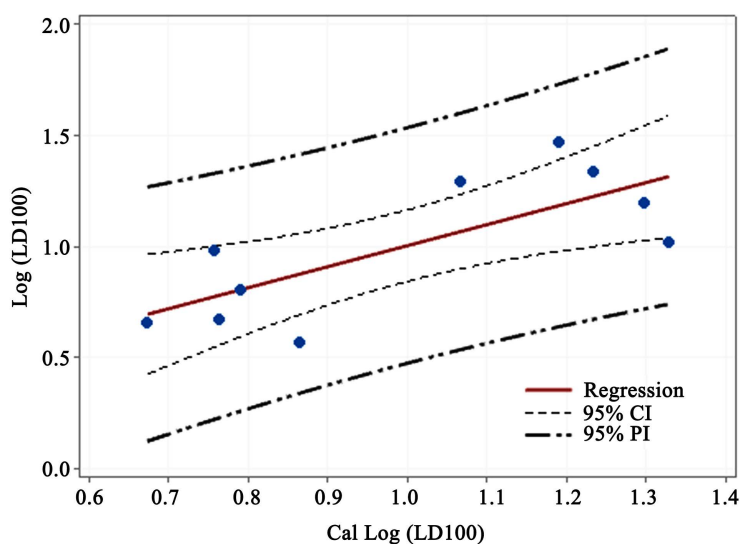
Figure 3. Plot of observed LD100 vs calculated LD100 using second order hyperpolarizability.
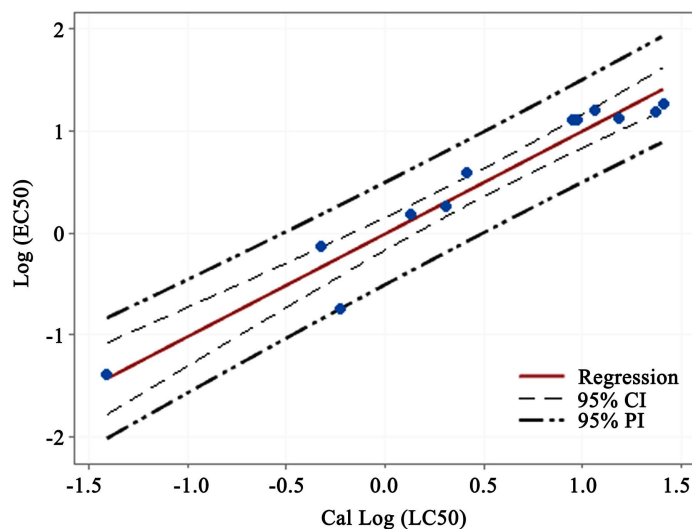
Figure 4. Plot of observed EC50 vs calculated EC50 using second order hyperpolarizability.

$$\log EC50 = -0.034559\alpha_{xx}\left(\pm0.058434\right)+0.099846\alpha_{yy}\left(\pm0.084750\right)$$
$$-0.094943\alpha_{zz}\left(\pm0.117111\right)+3.086897\left(\pm0.516546\right) \qquad (7b)$$
$$n = 12, r = 0.9566, r^2 = 0.9151, s = 0.2980$$

$$\log EC50 = -0.019657Ge^1\left(\pm0.010376\right)+0.031923Ge^2\left(\pm0.029880\right)$$
$$+3.251170\left(\pm0.336243\right) \qquad (7c)$$
$$n = 12, r = 0.9574, r^2 = 0.9167, s = 0.2784$$

The process of toxicology analysis was done through computational techniques, and the relation of the molecular structure with the interaction subject can be represented by this relation:

$$\text{Activity} = f\left(\text{physiochemistry properties}\right) \qquad (8)$$

The physiochemistry includes chemical properties, electronic properties, molecular topology, thermodynamic and optical properties [32]. The modification in molecular structure will change the biological activity. In linear form, the activity can be described by:

$$\text{Activity} = a_0 + a_1 x_1 + a_2 x_2 + a_3 x_3 + \cdots + a_n x_n \qquad (9)$$

where x_n is the molecular descriptor and a_n is the constant. The investigation of quantitative relationships between chemical structures characterized by electronic properties is one of the most important tools in biological activity such as toxicology analysis. The regression Equations (4)-(7) show that the electronic properties such as the electrophilicity index, chemical hardness, dynamic linear polarizability, second order hyperpolarizability, graph energy, dipole moment and conductor like screening model show a good relationship with the biological activity.

The electrophilicity index and chemical hardness show the best regression for *in vitro* oxidation by the reconstituted enzyme system. The electrophilicity index

shows the stabilization energy when the molecules receive the charges from the environment. The electrophilicity is also related to the chemical potential, which can be related to the binding of environment in the biological system. The chemical hardness describes the transfer of charge in the system with high charge density. These two descriptors are useful indicators to estimate enzyme activities on alkane substrates. The regression result shows that both descriptors are best fitting to relate with enzyme activity system. This is also agreeable with the result by Grillo and his coworkers [33].

The GE index is a new molecular descriptor related to the characteristic value of eigen function. The eigen function is assigned to Linear Combination Atomic Orbital Self-Consistent Filed (LCOASCF) molecular orbital. This index corresponds to orbital energy of the molecular structure. GE index is capable of describing the binding of energy interaction, the total electron exchange energy, electrostatic that exists in the molecule and the resonance energy [34] [35]. Therefore, GE index can explain the chemical interpretation based on the electronic properties of molecular structure involved in the biochemical interaction. Concerning the GE index as molecular quantum description in the biological activity and toxicity, it is a good fitting with the regression equation.

The non-linear polarizability of the molecule is an important physical property that draws our attention in chemical-biological interaction. The non-linear polarizability is a measurement of distortion of molecules in an electric field. This property measures the strength of molecular interactions such as long-range intermolecular induction, dispersion forces, scattering and electrons interaction [30]. The linear polarizability (a) shows a good regression fitting for most biological activities and toxicities. Hansch and Kurup also reported that the linear polarizability shows a good regression result with their toxicity study [36]. The hyperpolarizability descriptor is related with non-linear polarizability constant. This descriptor shows a moderate regression fitting (with the value of r between 0.901 - 0.788). The second order hyperpolarizability shows the best regression fitting in most biological activities and toxicities. This might be due to the induced non-linear dipole moment interaction with the applied field, giving a fine electronic property in the molecular structure. The electron-correlation effect in the microscopic polarizability calculation produced a good correlation in predicting the qualitative trends for structure-property relationships [37].

The conductor like screening model (COSMO) area with molecular dipole moment is surprisingly a good fitting for enzyme activity and toxicity. COSMO area is related with the surface charge densities on the neighboring segments. The COSMO area is an effective area of the screening surface. The screening surface is related to the perturbation Coulomb interaction in the molecule. The screening depends on the localization of charge, and the molecular polarizability. The molecular polarization is contributed by electronic, vibration and rotation. This molecular polarization contribution is also related with molecular dipole moment. In the researchers' work, the COSMO area and molecular dipole moment are important molecular properties in relating the biological and toxi-

cology activities.

4. Conclusions

The researchers have demonstrated that the molecular quantum descriptors calculated from semi-empirical calculation are applicable for the development of quantitative structure activity/retention relationship. These molecular quantum descriptors include the electrophilicity index, chemical hardness, dynamic linear polarizability, conductor-like screening model, graph energy index, second-order hyperpolarizability, Estrada index, and molecular dipole moment. The molecular quantum descriptors are classified into five principal component factors. The molecular quantum descriptors generated from semi-empirical calculations give a good correlation with the retention index, biological activity, and toxicity. This shows that the molecular quantum descriptors have good reliability to become a new approach in QSAR and QSRR.

In future studies, a comparison calculation of molecular descriptor can be calculated using density functional theory (DFT). DFT includes electron correlation in the *ab initio* calculation. The non-linear molecular properties also can be calculated using B3LYP, Moller-Plesset perturbation, configuration interaction and coupled-cluster theory [38]. This method can be used to compare QSAR accuracy with semi-empirical method.

Acknowledgements

The authors would like to thank Dr. James J. P. Stewart from MOPAC Inc. for his permission to use the MOPAC software, and UiTM's Department of Infostructure for the SPSS and MiniTab software usage permission.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] American Chemistry Council (2020) 2020 Guide to the Business of Chemistry. American Chemistry Council, Washington DC.

[2] Wang, Y., Hu, P., Yang, J., Zhu, Y.-A. and Chen, D. (2021) C-H Bond Activation in Light Alkanes: A Theoretical Perspective. *Chemical Society Reviews*, **50**, 4299-4358. https://doi.org/10.1039/D0CS01262A

[3] Vora, B.V., Kocal, J.A., Barger, P.T., Schmidt, R.J. and Johnson, J.A. (2003) Alkylation. In: Othmer, K., Ed., *Kirk-Othmer Encyclopedia of Chemical Technology*, John Wiley & Sons, Inc., Hoboken, 169-203. https://doi.org/10.1002/0471238961.0112112508011313.a01.pub2

[4] Moore, A.F. (1982) Final Report of the Safety Assessment of Isobutane, Isopentane, N-Butane, and Propane. *Journal of the American College of Toxicology*, **1**, 127-142. https://doi.org/10.3109%2F10915818209021266

[5] Christian, W.C., Butler, T.M., Ghannam, R.B., Webb, P.N. and Techtmann, S.M.

(2020) Phylogeny and Diversity of Alkane-Degrading Enzyme Gene Variants in the Laurentian Great Lakes and Western Atlantic. *FEMS Microbiology Letters*, **367**, Article No. fnaa182. https://doi.org/10.1093/femsle/fnaa182

[6] Tormoehlen, L.M., Tekulve, K.J. and Nañagas, K.A. (2014) Hydrocarbon Toxicity: A Review. *Clinical Toxicology*, **52**, 479-489. https://doi.org/10.3109/15563650.2014.923904

[7] Carter, D.E. and Fernando, Q. (1979) Chemical Toxicology. Part I. Organic Compounds. *Journal of Chemical Education*, **56**, 284. https://doi.org/10.1021/ed056p284

[8] MacFarland, H.N. (1986) Toxicology of Solvents. *American Industrial Hygiene Association Journal*, **47**, 704-707. https://doi.org/10.1080/15298668691390511

[9] Phillips, J.C., Gibson, W.B., Yam, J., Alden, C.L. and Hard, G.C. (1990) Survey of the QSAR and *in Vitro* Approaches for Developing Non-Animal Methods to Supersede the *in Vivo* LD_{50} Test. *Food and Chemical Toxicology*, **28**, 375-394. https://doi.org/10.1016/0278-6915(90)90112-Z

[10] Zerner, M.C. (1991) Semiempirical Molecular Orbital Methods. In: Lipkowitz, K.B. and Boyd, D.B., Eds., *Reviews in Computational Chemistry*, Vol. 2, Wiley-VCH, Inc., New York, 313-365. https://doi.org/10.1002/9780470125793.ch8

[11] Lewars, E.G. (2016) Computational Chemistry. 3rd Edition, Springer International Publishing, Cham, 728. https://doi.org/10.1007/978-3-319-30916-3

[12] Stewart, J.J.P. (2016) MOPAC 2016. Colorado Springs.

[13] Kurtz, H.A., Stewart, J.J.P. and Dieter, K.M. (1990) Calculation of the Nonlinear Optical Properties of Molecules. *Journal of Computational Chemistry*, **11**, 82-87. https://doi.org/10.1002/jcc.540110110

[14] Pople, J.A., Santry, D.P. and Segal, G.A. (1965) Approximate Self-Consistent Molecular Orbital Theory. I. Invariant Procedures. *The Journal of Chemical Physics*, **43**, S129-S135. https://doi.org/10.1063/1.1701475

[15] Parr, R.G., Szentpály, L.V. and Liu, S. (1999) Electrophilicity Index. *Journal of the American Chemical Society*, **121**, 1922-1924. https://doi.org/10.1021/ja983494x

[16] Gázquez, J.L., Cedillo, A. and Vela, A. (2007) Electrodonating and Electroaccepting Powers. *The Journal of Physical Chemistry A*, **111**, 1966-1970. https://doi.org/10.1021/jp065459f

[17] Estrada, E. (2000) Characterization of 3D Molecular Structure. *Chemical Physics Letters*, **319**, 713-718. https://doi.org/10.1016/S0009-2614(00)00158-5

[18] Akhondi, S.A., Rey, H., Schwörer, M., Maier, M., Toomey, J., Nau, H., *et al.* (2019) Automatic Identification of Relevant Chemical Compounds from Patents. *Database*, **2019**, Article No. baz001. https://doi.org/10.1093/database/baz001

[19] Komp, E. and Valleau, S. (2020) Machine Learning Quantum Reaction Rate Constants. *The Journal of Physical Chemistry A*, **124**, 8607-8613. https://doi.org/10.1021/acs.jpca.0c05992

[20] Dembitsky, V.M., Dor, I., Shkrob, I. and Aki, M. (2001) Branched Alkanes and Other Apolar Compounds Produced by the Cyanobacterium Microcoleus Vaginatusfrom the Negev Desert. *Russian Journal of Bioorganic Chemistry*, **27**, 110-119. https://doi.org/10.1023/A:1011385220331

[21] MacLeod, G. and Ames, J.M. (1991) Gas Chromatography-Mass Spectrometry of the Volatile Components of Cooked Scorzonera. *Phytochemistry*, **30**, 883-888. https://doi.org/10.1016/0031-9422(91)85272-2

[22] Harangi, J. (2003) Retention Index Calculation without *N*-Alkanes—The Virtual Carbon Number. *Journal of Chromatography A*, **993**, 187-195.

https://doi.org/10.1016/S0021-9673(03)00320-0

[23] Bobra, A.M., Shiu, W.Y. and Mackay, D. (1983) A Predictive Correlation for the Acute Toxicity of Hydrocarbons and Chlorinated Hydrocarbons to the Water Flea (Daphniamagna). *Chemosphere*, **12**, 1121-1129. https://doi.org/10.1016/0045-6535(83)90118-2

[24] van Beilen, J.B., Kingma, J. and Witholt, B. (1994) Substrate Specificity of the Alkane Hydroxylase System of *Pseudomonas oleovorans* GPo1. *Enzyme and Microbial Technology*, **16**, 904-911. https://doi.org/10.1016/0141-0229(94)90066-3

[25] Jeppsson, R. (1975) Parabolic Relationship between Lipophilicity and Biological Activity of Aliphatic Hydrocarbons, Ethers and Ketones after Intravenous Injections of Emulsion Formulations into Mice. *Acta Pharmacologica et Toxicologica*, **37**, 56-64. https://doi.org/10.1111/j.1600-0773.1975.tb00822.x

[26] Peterson, D.R. (1994) Calculating the Aquatic Toxicity of Hydrocarbon Mixtures. *Chemosphere*, **29**, 2493-2506. https://doi.org/10.1016/0045-6535(94)90052-3

[27] Kikuchi, O. (1987) Systematic QSAR Procedures with Quantum Chemical Descriptors. *Quantitative Structure-Activity Relationships*, **6**, 179-184. https://doi.org/10.1002/qsar.19870060406

[28] Ashraf, C., Joshi, N., Beck, D.A.C. and Pfaendtner, J. (2021) Data Science in Chemical Engineering: Applications to Molecular Science. *Annual Review of Chemical and Biomolecular Engineering*, **12**, 15-37. https://doi.org/10.1146/annurev-chembioeng-101220-102232

[29] Brown, R.E. and Simas, A.M. (1982) On the Applicability of CNDO Indices for the Prediction of Chemical Reactivity. *Theoretica Chimica Acta*, **62**, 1-16. https://doi.org/10.1007/BF00551049

[30] Khan, M.F., Nahar, N., Rashid, R.B., Chowdhury, A. and Rashid, M.A. (2018) Computational Investigations of Physicochemical, Pharmacokinetic, Toxicological Properties and Molecular Docking of Betulinic Acid, a Constituent of *Corypha taliera* (Roxb.) with Phospholipase A2 (PLA2). *BMC Complementary and Alternative Medicine*, **18**, Article No. 48. https://doi.org/10.1186/s12906-018-2116-x

[31] Gutman, I., Furtula, B., Glišić, B., Marković, V. and Vesel, A. (2007) Estrada Index of Acyclic Molecules. *Indian Journal of Chemistry A*, **46**, 723-728.

[32] Jana, G., Pal, R., Sural, S. and Chattaraj, P.K. (2020) Quantitative Structure-Toxicity Relationship Models Based on Hydrophobicity and Electrophilicity. In: Roy, K., Ed., *Ecotoxicological QSARs*, Humana, New York, 661-679. https://doi.org/10.1007/978-1-0716-0150-1_27

[33] Grillo, I.B., Urquiza-Carvalho, G.A., Fernando Ruggiero Bachega, J. and Bruno Rocha, G. (2020) Elucidating Enzymatic Catalysis Using Fast Quantum Chemical Descriptors. *Journal of Chemical Information and Modeling*, **60**, 578-591. https://doi.org/10.1021/acs.jcim.9b00860

[34] Rakshit, S., Banerjee, M. and Hazra, B. (1988) Molecular Branching Topology & Quantum Mechanical Quantities. *Indian Journal of Chemistry A*, **27**, 183-187.

[35] Brankov, V., Stevanović, D.P. and Gutman, I. (2004) Equienergetic Chemical Trees. *Journal of the Serbian Chemical Society*, **69**, 549-554. https://doi.org/10.2298/JSC0407549B

[36] Hansch, C. and Kurup, A. (2003) QSAR of Chemical Polarizability and Nerve Toxicity. 2. *Journal of Chemical Information and Computer Sciences*, **43**, 1647-1651. https://doi.org/10.1021/ci030289e

[37] Perrin, E. and Prasad, P.N. (1989) *Ab Initio* Calculations of Polarizability and

Second Hyperpolarizability in Benzene Including Electron Correlation Treated by Mo/ller-Plesset Theory. *The Journal of Chemical Physics*, **91**, 4728-4732. https://doi.org/10.1063/1.456761

[38] Barca, G.M.J., Bertoni, C., Carrington, L., Datta, D., De Silva, N., Emiliano Deustua, J., *et al.* (2020) Recent Developments in the General Atomic and Molecular Electronic Structure System. *The Journal of Chemical Physics*, **152**, Article ID: 154102. https://doi.org/10.1063/5.0005188

Appendix

Table A1. Table of the retention index (Log Rt1, Log Rt2 and Log Rt3), biological activity (Log (Ea)) and toxicity (Log LC50, Log LD50, Log LD100 and Log EC50) used in this study.

	Log $Rt1$	Log $Rt2$	Log $Rt3$	Log (Ea)	Log $LC50$	Log $LD50$	Log $LD100$	Log $EC50$
n-Butane						0.7924		1.1858
2-Methylpropane								1.2622
n-Pentane		0.6803		1.6902	2.1303		1.2867	1.1153
2-methylbutane				0.3010				1.1970
2,2-Dimethylpropane								
n-Hexane		0.8762		1.9138	1.6532		0.9841	1.1086
2-Methylpentane				1.2041				1.0966
3-Methylpentane				1.4624				
2,2-Dimethylbutane		0.7356						
2,3-Dimethylbutane		0.7945						
n-Heptane		1.1206		1.9494		0.3464	0.6599	0.2577
2-Methylhexane				1.3617				0.5922
3-Methylhexane		1.0592		1.3222				
3-Ethylpentane	1.1271							
2,2-Dimethylpentane		0.9274						
n-Octane	1.2496	1.3448	0.67797	1.9685	0.5185		0.5740	−0.1427
2-Methylheptane	1.2045			1.9445				
3-Methylheptane		1.2916		1.6435				
4-Methylheptane	1.1987	1.2778		1.7782				
2,5-Dimethylhexane		1.2739		0.9542				
3,4-Dimethylhexane		1.2828						
2,2,4-Trimethylpentane								0.1732
n-nonane		1.5004	0.80236	2.0000		0.2330	0.6721	
2-Methyloctane				1.9345				−0.7447
3-Methyloctane				1.8129				
4-Methyloctane				1.7709				
3-ethylheptane	1.3237							
2,3-dimethylheptane	1.3130	1.4382						
2,6-dimethylheptane		1.3943						
2,2,4-trimethylhexane	1.3520							
2,2-dimethyl-3-ethylpentane	1.3429							
2,2,3,3-tetramethylpentane	1.3734							
2,2,3,4-tetramethylpentane	1.3485							
n-decane		1.6084	0.96205	2.0000	−0.6990		0.8069	

Continued

2-methylnonane				1.8513				−1.3979
3-methylnonane	1.3680							
2,6-dimethyloctane				1.6128				
3,3-dimethyloctane	1.3559							
4-n-propylheptane	1.3734							
2,3,6-trimethylheptane		0.1761						
3,4,5-Trimethylheptane	1.3782							
n-Undecane		1.6974	1.15152	1.9777		0.5198	1.0199	
4-Methyldecane	1.3837							
2,3-Dimethylnonane	1.3889							
n-Dodecane		1.7637	1.36306	2.0170			1.1953	
n-Tridecane		1.8176				0.8000	1.3322	
2,5-Dimethylundecane	1.4059							
4,7-Dimethylundecane	1.4210							
3,7-Dimethylundecane	1.4252							
4,6-Dimethylundecane	1.4408							
3,5-Dimethylundecane	1.4476							
5,7-Dimethylundecane	1.4747							
n-Tetradecane	1.4960	1.8631		0.8451			1.4658	
6-Methyltridecane	1.4978							
cyclopentane				1.2041	2.1761			
cyclohexane		1.0330		1.1139	1.6532			
methylcyclohexane		1.1824		1.7404	1.1761			
Methylcycloheptane		1.3981						
1,4-dimethylcyclohexane				0.3010				