Iraqi Journal of Science, 2023, Vol. 64, No. 6, pp: 2648-2668
DOI: 10.24996/ijs.2023.64.6.1

ISSN: 0067-2904

Topology Molecular Indices Relationship of Electronic Properties of N-Alkanes and Branched Alkanes

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

Faculty of Applied Sciences, Universiti Teknologi MARA Perak Branch, Tapah Campus, 35400 Tapah Road, Perak, Malaysia

Received: 1/9/2021 Accepted: 21/7/2022 Published: 30/6/2023

Abstract

Eight electronic properties; HUMO, LUMO, HOMO-LUMO energy gap, dipole moment point-charge, dipole moment hybrid, molecular weight, heat of formation and zero-point energy of 60 normal and branched alkanes were examined using topology molecular indices. All the electronic properties were calculated using semi-empirical self-consistent molecular orbital theory. The relationship of electronic calculation properties with seven models of topology indices based on degree and/or distance were obtained in terms of their correlation, regression and principal component analysis. Most of the properties were well-modelled ($r^2 > 0.82$) by topology molecular indices except the dipole moment point-charge and hybrid. The PCA resulted in 7 properties and 60 structures of alkanes that produced two principal components with eigenvalues of greater than 1. The first principal component explained 60.388%, while the second principal component explained 26.457%, bringing a cumulative value of 86.845% to the data variation.

Keywords: topology indices, molecular electronic properties, QSPR, semi-empirical calculation.

1. Introduction

A topology index is a mathematical number representing or describing some aspects of molecular structure. Topology indices using the graph theory approach are especially used to explain the isomerism rationalised by the chemical structure theory. The constitutional isomers of the members of certain homologous series, such as alkanes can be represented using analytical forms called trees[1]. The number of alkane constitutional isomers (trees) for the number of carbon atoms 1, 2 and 3 is 1. The numbers of a constitutional isomer of alkanes for the number of atoms 4, 5 and 6 are 2, 3 and 5, respectively. However, the number of isomers increases abruptly with the number of carbons of more than 7[2]. Therefore, there is a need for a numerical quantity to represent the indices based on various invariants or characteristics of molecular isomers. Topology indices are also widely used in establishing the correlations between the structure of a molecular compound and its physicochemical properties or biological activities[3].

A graph is a discrete mathematical concept to visualise a structure that has a collection of points (vertexes) or lines connected to points (edges)[4]. The vertex degree is the number of connected edges in a particular vertex. In chemical graph theory, the vertex represents atoms, and the edge represents bonds. Weiner carried out a pioneer topolog,y indices which can be

*Email: ahmadnazib@uitm.edu.my

easily visualised based on the distance path from hydrogen-suppressed graph[5]. Topological indices are divided into three types: degree-based indices, distance-based indices and spectrum-based indices[6]. The well-known distance-based indices and degree-based indices that have been widely used are Weiner, Randic, Zagreb, Balaban, Schultz and Xu. Since 2000, over 400 topological indices have been used in structure versus property or biological activity [7]. However, topology indices have a major drawback of degeneration, such as two or three chemical constitutions giving rise to the same topology values [8]. To date, several researches have been carried out to establish the novel indices that meet these criteria [9, 10]: novel indices should have a direct structural interpretation (so that it can meaningfully contribute to building molecular models); novel indices should involve structural features that existing indices have failed to accommodate adequately, and novel indices should dominate a correlation with some molecular properties. Many indices provided have always been used to calculate the prediction of physicochemical properties or biological activities for n-alkanes and branched alkanes [11].

The electronic structure of n-alkanes and branched alkanes have been the subject of extensive attention because they comprise one of the foundations of chemistry and petrochemistry. In petrochemistry, n-alkanes and branched alkanes are the major constituents of natural gas, crude oil and raw material for processing or manufacturing chemical industry[12]. Since the applicability of n-alkanes and branched alkanes in process or manufacturing, many groups have developed and investigated the electronic properties using experimental or computational quantum chemistry. In experimental approach, the electronic properties of alkanes or branched alkanes have been investigated using vacuum ultraviolet spectroscopy or gas chromatography [13]. Whilst, many researchers have investigated the electronic properties of n-alkanes and branched alkanes using computational quantum chemistry such as time-dependent density functional theory (TD-DFT), symmetry-adapted cluster configuration interaction (SAC-CI), Møller-Plesset and semi-empirical calculation [14-16]. The main problem in quantum computational chemistry is to find the wavefunction solution within a reasonable time and using available computational resources [17]. The drawback of this method can be overcome using a quantitative structure-property relationship approach. However, choosing a suitable descriptor is still being investigated [18, 19]. Therefore, this study aimed to investigate the impact of molecular descriptors on the electronic calculation properties such as HUMO, LUMO, HOMO-LUMO energy gap, dipole moment point-charge, dipole moment hybrid, molecular weight, heat of formation and zero-point energy of the alkanes. We also evaluated the performance of molecular descriptor based on the topology indices, which are suitable for electronic properties purpose. For the To aid this effort, it was utilised the correlation analysis, principal component, and regression analysis for molecular descriptors (topology indices).

2. Methods

2.1 Calculation of Topology Indices

Topology indices are one of the molecular descriptors that apply graph theory. The atoms in the molecular structure are represented by vertexes. The chemical bonding is described by the edge [20, 21]. To describe the method of calculation of topology indices, it is illustrated with a specific example, molecules of 2,2-dimethyl butane. This molecule can be represented by graph G(V, E) with vertex (V) and edge set (E). Figure 1(a) shows there are 6 vertexes (V) and 5 numbers of E. Therefore, set V can be represented by {1, 2, 3, 4, 5, and 6}[22]. While set E can be represented by {(1, 2), (2, 3), (3, 4), (2, 5), (2, 6)}. Figure 1(b) shows the degree of the vertex for 2,2-dimethyl butane. This paper described the vertex degree between edge i and j as u_i and $u_{j,}$ respectively. The topology indices are defined in equations (1-7). The example of calculation for 2, 2-dimethylbutane using all topology indices can be found in the supplementary.

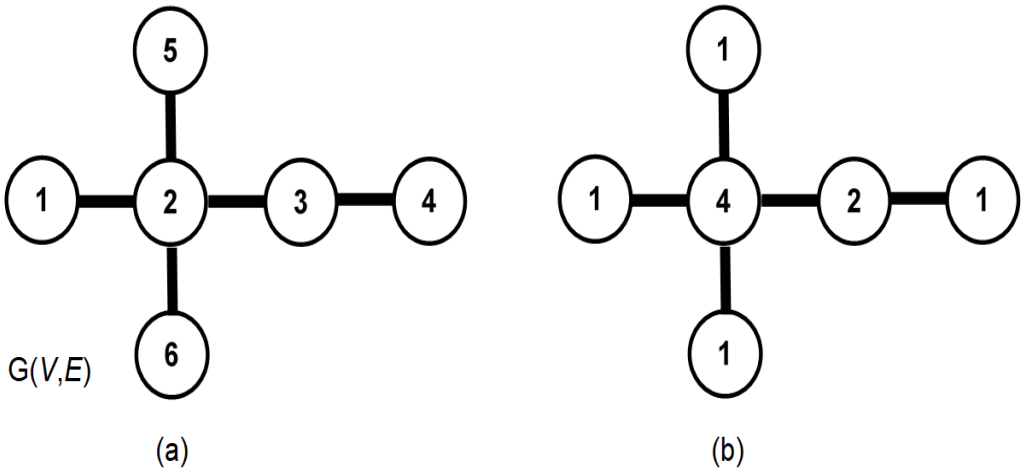

Figure 1 (a): The hydrogen-supressed molecular graph of 2,2-dimethyl butane **(b)-** vertex degree for 2,2-dimethylbutane

Topology I: Randic Indices (The connectivity indices). The Randic indices of a graph G by χ^h (G) a defined as[23]:

$$\chi^h(G) = \sum_{u_i u_j \ldots u_{h+1}} \frac{1}{\sqrt{u_i \ldots u_{h+1}}} \tag{1a}$$

The summation is calculated over all possible paths of the degree of vertex *u*. Therefore, χ^0 is the sum of reciprocal square of the degree of vertex *u*. The sum connectivity is defined as[24]:

$$\chi^+(G) = \sum_{uv \in G} \frac{1}{\sqrt{u_u + u_v}} \tag{1b}$$

Topology II: Modified Zagreb Indices. The hyper-Zagreb indices are defined as below[25]:

$$HM(G) = \sum_{uv \in G} \left(u_u + u_v\right)^2 \tag{2a}$$

In this paper, we define the inverse hyper-Zagreb indices as

$$HM^*(G) = \sum_{uv \in G} 1/\left(u_u + u_v\right)^2 \tag{2b}$$

$$AM(G) = \sum_{uv \in G} \left(\frac{u_u u_v}{u_u + u_v - 2}\right)^3 \tag{2c}$$

Topology III: Weiner indices. The Weiner indices are calculated from the distance matrix [26]. The Weiner indices, *w*, is given by equation (3a). Randic also improved the Weiner indices calledhyper-Wiener indices[27]. The hyper-Wiener indices, *ww* is given by equation (3b)[28]:

$$w = \frac{1}{2}\sum_{ij} d_{ij} \tag{3a}$$

$$ww(G) = \sum_{ij}\left(d_{ij}^2 + d_{ij}\right) \tag{3b}$$

Where d_{ij} is the shortest distance between vertices i and j. Weiner also defines the structural variable, known as the polarity number, *p*. The polarity number is half the number of the pairs

of carbon atoms separated by three carbon-carbon bonds or the number of diagonal elements of *w* with a distance of 3. Another parameter in Weiner topology is Δw which is given as:

$$\Delta w = w - w_0 \tag{3c}$$

where w_o is the Weiner indices of the straight-chain member of the group of isomers. While w in equation (3c) is the Weiner indices of the respective isomers.

Topology IV: Harary indices, $h = h\,[G\,(V, E)]$ of a graph $G\,(V, E)$ is related to reciprocal distance indices that is given by equation (4a)[29].

$$h = \frac{1}{2}\sum_{ij}\frac{1}{d_{ij}} \tag{4a}$$

The diagonal element for the Harary distance matrix equals zero the for carbon atom. While the second-order Harary, indices, *hh* can be denoted as

$$hh = \frac{1}{2}\sum_{ij}\frac{1}{d_{ij}^2} \tag{4b}$$

Topology V: Balaban index, $J = J\,[G\,(V, E)]$ is calculated using the average-distance sum connectivity and defined as

$$J = q\sum_{ij}\frac{1}{\left(D_i D_j\right)^{1/2}} \tag{5}$$

where q is the number q of vertex adjacencies and D_i is the distance sum of $G\,(V, E)$[30, 31].

Topology VI: Molecular Topological Index (MTI Index). MTI index has been introduced using adjacency (A), degree (*v*) and distance matrix (D)[5]. The adjacency matrix A of G (V, E) element is defined as 1 if u_i neighbouring with u_j and otherwise is zero. While the matrix *v* is the sum of each column element in matrix A. Matrix *v* also represents the degree of a vertex in a matrix arrangement. MTI index usealgebraic matrix operation. The index is simplified using the following mathematical equation:

$$MTI = \sum v\left(D + A\right) \tag{6}$$

vi. Topology VII: Xu index is defined as [32]:

$$Xu = n\log\left(\sum_{n}^{i} v_i s_i^2 \,/\, \sum_{n}^{i} v_i s_i\right) \tag{7}$$

where s_i is the distance sum of $G\,(V, E)$ and v_i is the sum vertex-degree matrix of $G\,(V, E)$. All the calculated topology indices are shown in Table 1.

2.2 Calculation of electronic properties

The physical properties examined in this study (the observed value) HUMO, LUMO, HOMO-LUMO energy gap, dipole moment point-charge, dipole moment hybrid, molecular weight, heat of formation and zero-point energy were calculated using semi-empirical self-consistent molecular orbital theory. These properties were computed using MOPAC2016, Version: 21.002 W, James J. P. Stewart software. The calculation was done using PM6 parameters. The input of the structure and geometry optimization were generated using 3-dimension Avogadro version 1.2.0. The observation of the physical properties of the molecular structure is shown in Table 2.

2.3 Statistical Analysis

2.3.1 Correlation

Correlation is primarily to determine a relationship or a connection between the variables. Correlation uses covariance which measures the linear association between the variables. The degree of correlation is quantitively measured by correlation coefficient $r_{xy.}$ This coefficient is also known as product-moment correlation or Pearson correlation. The value of r_{xy} can range between -1 and +1. If the slope has a negative value, the linear interrelationship also has a negative value. The relationship has a strong value with the condition that reaches one.

2.3.2 Linear Regression

Linear Regression is the predictive model to determine the best fit linear line between the dependent and independent variable. The linear mathematic equation will be established based on the independent variable via the least square method. The simple Linear Regression equation is stated as in equation 8:

$$y = b_o + b_1 x_1 + b_2 x_2 + ... + b_n x_n \quad (8)$$

where b_o is the intercept; b_1, b_2 … b_n are coefficient of n independent variable; x_1 , x_2 , … are the n independent variable, and y is the predictive (calculated) values. We utilized the standard entry method in this calculation, in which the independent variables were entered into the equation at the same time. Regression equations and other statistical measures were obtained using options in the SPSS software package. The final equations were selected based on their standard errors (SE) and F-value. SE is the expected distribution of an estimated regression coefficient based on the coefficient across multiple data. While the F-value was calculated by dividing the mean regression sum of squares by the mean error sum of squares.

It was used a single-cross validation procedure for the validation of regression [33, 34]. The effectiveness of the regression model was estimated from one data sample that would be applied to other selected data samples. The number of samples must be larger than the number of predictors. The data were split into two subsamples of unequal size using random splitting. The correlation coefficient (r_2) sample size of approximately 75% of all cases was computed. Another subgroup correlation coefficient (r_1) was also computed. The square differences ($ds^2 = r_1^2 - r_1^2$) between these two values were estimated[34]. The estimated ds^2 explains how well the regression model would predict a future sample and can be very useful.

3.2.3 Principal Component Analysis

Principal component analysis (PCA) was used to determine the inherent dimensionality of the groups of properties. This is a data compression method based on the correlation among variables. It aims to group those correlated variables, replacing the original descriptors with a new set called principal components, PCs, onto which the data are projected. These PCs are completely uncorrelated and are built as a simple linear combination of original variables. It is important to point out here that the PCs contain most of the variability in the data set, albeit in a much lower-dimensional space. The first principal component, PC1, is defined in the direction of maximum variance of the whole data set. PC2 is the direction that describes the maximum variance in the orthogonal subspace to PC1. The subsequent components are taken orthogonal to those previously chosen and describe the maximum of the remaining variance. Once the redundancy is removed, only the first few principal components are required to describe most of the information contained in the original data set. The orthogonal PROMAX rotation was used during the calculation.

3. Results

Table 1: Calculated topology indices for 60 n-alkanes and branched alkanes

Structure	0χ	1χ	2χ	3χ	4χ	$^{+}$χ	HM	*HM	AM	w	ww	Δw	p	h	hh	J	MT
n-Butane	3.414	1.914	1.000	1.000	0.000	1.655	34	0.285	24.000	10	30	0	1	4.333	3.611	1.975	38
Methylpropane	3.577	1.732	1.732	0.000	0.000	1.500	48	0.188	24.000	9	22	1	0	4.500	3.750	2.324	36
n-Pentane	4.121	5.414	1.354	0.707	0.354	2.157	50	0.347	32.000	20	70	0	2	6.417	5.035	2.191	74
-Methylbutane	4.284	2.270	1.802	0.816	0.000	2.115	66	0.276	22.750	18	54	18	2	6.917	5.597	2.661	67
2,2-imethylpropane	4.500	2.000	3.000	0.000	0.000	1.789	100	0.160	9.481	16	44	19	0	7.000	5.500	3.024	64
n-Hexane	4.828	2.914	1.707	0.957	0.500	2.655	66	0.410	40.000	35	140	0	3	8.700	6.498	2.339	128
Methylpentane	4.992	2.770	2.183	0.866	0.577	2.525	82	0.339	30.675	32	116	3	3	9.000	6.708	2.627	118
Methylpentane	4.992	2.808	1.922	1.394	0.255	2.549	84	0.365	35.375	31	108	4	4	9.083	6.757	2.737	114
2,2-imethylbutane	5.207	2.561	2.914	1.061	0.000	2.327	120	0.259	23.111	28	88	7	3	9.500	7.083	3.168	106
2,3-imethylbutane	5.155	2.643	1.911	1.333	0.000	2.408	100	0.278	24.891	29	94	6	4	9.333	6.944	2.878	108
n-Heptane	5.536	3.414	2.061	1.207	0.677	3.155	82	0.472	48.000	56	252	0	4	11.150	7.990	2.447	204
-Methylhexane	5.699	3.750	2.536	1.135	0.612	3.025	98	0.401	34.125	52	216	4	4	11.483	8.212	2.678	190
-Methylhexane	5.699	3.308	2.302	1.478	0.697	3.049	100	0.427	43.375	50	198	2	5	11.617	8.283	2.832	316
-Ethylpentane	5.699	3.346	2.091	1.732	0.574	2.091	102	0.453	48.000	48	180	0	6	11.750	8.354	2.992	174
2,2-imethylpentane	5.914	3.061	3.311	1.000	0.750	2.827	136	0.321	34.125	46	168	2	4	12.083	8.632	3.154	170
2,3-imethylpentane	5.862	3.181	2.630	1.782	0.471	2.933	118	0.366	37.516	46	166	2	6	12.000	8.542	3.144	168
2,4-imethylpentane	5.862	3.126	3.023	0.943	0.943	2.894	114	0.330	29.500	47	178	1	3	12.333	9.194	2.983	176
3,3-imethylpentane	5.914	3.121	2.871	1.664	0.250	2.866	140	0.378	36.741	44	152	4	6	12.250	8.729	3.36	162
2,2,3-rimethylbutane	6.077	2.943	3.521	1.732	0.000	2.720	156	0.265	27.685	42	138	6	6	12.500	8.917	3.541	156
n-Octane	6.243	3.914	2.414	1.457	0.854	3.655	98	0.535	56.000	84	420	9	5	13.743	9.502	2.53	306
Methylheptane	6.406	3.770	2.890	1.385	0.803	3.525	114	0.464	46.750	79	370	4	5	14.100	9.731	2.716	288
Methylheptane	6.406	3.808	2.656	1.747	0.757	3.549	116	0.490	51.375	75	336	0	6	14.767	10.564	2.895	276
Methylheptane	6.406	3.808	2.683	1.359	1.130	3.549	116	0.490	51.375	75	330	0	6	14.317	9.837	2.92	272
3-Ethylhexane	6.406	3.846	2.471	1.852	1.105	3.574	118	0.516	56.000	72	300	3	6	14.483	9.920	3.074	242
2,2-imethylhexane	6.621	3.561	3.664	1.280	0.707	3.327	152	0.384	39.111	71	298	4	6	14.767	10.176	3.112	249
2,3-imethylhexane	6.569	3.681	2.677	2.677	0.789	3.433	134	0.429	45.516	70	286	5	7	14.733	10.108	3.171	254
2,4-imethylhexane	6.569	3.664	3.263	1.571	0.971	3.419	132	0.419	42.125	71	294	4	6	14.650	10.059	3.099	258
2,5-imethylhexane	6.569	3.626	2.788	1.322	0.667	3.626	130	0.393	37.500	74	322	1	5	14.467	9.966	2.928	270
3,3-imethylhexane	6.621	3.621	3.268	1.634	0.854	3.366	156	0.420	44.740	67	262	8	7	15.033	10.318	3.424	229
3,4-imethylhexane	6.569	3.719	2.771	2.259	0.805	3.457	136	0.455	50.141	68	268	7	8	14.867	10.179	3.292	246
3-Ethyl-2-Methylhexane	6.569	3.719	2.532	1.992	3.474	1.231	136	0.455	53.516	67	258	8	7	14.917	10.201	3.355	242
3-Ethyl-3-Methylhexane	6.621	3.682	2.621	2.561	0.750	3.404	160	0.457	50.370	64	236	11	9	15.250	10.438	3.583	232
2,2,3-imethylpentane	6.784	3.481	3.675	2.091	0.612	3.244	174	0.354	40.301	63	230	12	8	15.417	10.576	3.623	230
2,2,4-imethylpentane	6.784	3.547	3.666	1.021	1.225	3.274	168	0.313	29.861	66	254	9	5	15.167	10.431	3.389	243

2,3,3-imethylpentane	6.784	3.504	3.497	2.520	0.408	3.258	176	0.364	41.315	62	222	13	9	15.500	10.625	3.708	226
2,3,4-imethylpentane	6.732	3.553	3.347	2.103	0.770	3.316	152	0.368	39.656	65	244	10	8	15.167	10.389	3.464	220
2,2,3,3-etrametylbutane	7.000	3.325	4.500	2.250	0.000	3.037	214	0.256	33.185	68	194	7	9	16.000	11.000	4.02	214
n-Nonane	6.950	4.414	2.768	1.707	1.030	4.155	114	0.597	64.000	120	660	12	6	16.461	11.029	2.595	438
-Methyloctane	7.113	4.270	3.243	1.635	0.979	4.025	130	0.526	54.750	114	594	6	6	16.836	11.263	2.747	421
-Methyloctane	7.113	4.385	3.009	1.709	0.947	4.049	132	0.552	59.375	110	550	2	7	17.026	11.354	2.877	400
-Methyloctane	7.113	4.308	3.036	1.832	1.255	4.049	132	0.552	59.375	108	528	0	7	17.110	11.389	2.955	392
-Ethylheptane	7.113	4.346	2.825	2.121	1.070	4.074	134	0.578	56.000	104	484	4	8	17.300	11.479	3.092	376
-Ethylheptane	7.113	4.346	2.852	1.971	1.513	4.074	134	0.578	64.000	102	462	6	8	17.383	11.514	3.175	368
2,2-imethylheptane	7.328	4.347	4.018	1.530	0.905	4.025	168	0.446	47.111	104	494	4	6	17.550	11.723	2.305	688
2,3-imethylheptane	7.276	4.181	2.622	2.151	0.859	3.933	150	0.491	53.516	102	470	6	8	17.550	11.667	3.155	370
2,4-Dimetylheptane	7.276	4.164	3.523	1.655	1.415	4.164	141	0.530	54.750	102	468	6	7	17.517	11.641	2.757	370
2,5-imethylheptane	7.276	4.164	3.485	1.934	0.822	4.164	148	0.481	50.125	104	488	4	7	17.417	11.596	2.678	378
2,6-imethylheptane	7.276	4.126	3.719	1.563	0.934	3.894	146	0.455	45.500	108	530	0	6	17.217	11.500	2.186	394
3,3-imethylheptane	7.328	4.121	3.828	2.164	0.832	3.866	172	0.483	52.741	98	434	10	8	17.883	11.889	3.33	356
3,4-imethylheptane	7.276	4.219	3.151	2.359	1.142	3.957	152	0.406	58.141	98	430	10	8	17.767	11.773	3.325	354
3,5-imethylheptane	7.276	4.202	3.263	2.199	1.020	3.944	150	0.507	54.750	100	448	8	8	17.633	11.702	3.223	362
4,4-imethylheptane	7.328	4.121	3.664	1.854	1.479	3.866	172	0.483	52.741	96	414	12	8	17.983	11.934	3.431	348
3-Ethyl,2-Methylhexane	7.276	4.219	3.201	2.127	1.380	3.957	141	0.530	54.750	96	408	12	9	17.850	11.808	3.836	346
4-Ethyl,2-Methylhexane	7.276	4.202	3.312	1.959	1.289	3.944	150	0.507	54.750	98	426	10	8	17.717	11.736	3.307	354
3-Ethyl,3-Methylhexane	7.328	4.182	3.268	2.561	1.207	3.904	176	0.519	58.370	92	376	16	10	18.233	12.066	4.522	332
3-Ethyl-4-Methylhexane	7.276	4.257	2.962	2.497	1.427	3.982	154	0.544	62.766	94	390	14	10	17.983	11.879	4.374	338
2,2,3-rimethylhexane	7.492	3.981	4.056	2.200	0.866	3.744	190	0.417	48.310	92	380	16	9	18.350	12.183	4.037	315
2,2,4-rimethylhexane	7.492	3.955	4.278	1.658	1.190	3.722	186	0.401	42.486	94	396	14	7	18.183	12.085	3.034	342
2,2,5-rimethylhexane	7.492	3.917	4.493	1.472	0.722	3.697	184	0.375	37.861	98	434	10	6	17.950	11.969	2.461	358
2,3,3-rimethylhexane	7.492	4.004	3.893	2.457	0.933	3.758	192	0.427	48.315	90	362	18	10	18.483	12.254	4.628	309

Table 2: Observation of electronic properties of 60 n-alkanes and branched alkanes

Structure	HOMO	LUMO	ΔE	μ_Q	μ_{hyb}	MW	HF	ZPE
Butane	-11.130	4.218	15.348	0.000	0.000	58.120	-26.039	75.449
2-Methylpropane	-11.215	4.412	15.627	0.083	0.186	58.120	-27.506	75.044
n-Pentane	-11.044	4.150	15.194	0.030	0.106	72.150	-31.017	91.973
2-Methylbutane	-11.048	4.214	15.262	0.064	0.134	86.180	-36.376	107.541
2,2-Dimethylpropane	-11.339	4.427	15.766	0.000	0.000	72.150	-35.021	90.641

n-Hexane	-11.004	4.082	15.086	0.001	0.000	86.180	-36.034	108.650
2-Methylpentane	-11.000	4.168	15.168	0.080	0.184	86.180	-36.666	108.186
3-Methylpentane	-11.024	4.162	15.186	0.063	0.124	86.180	-35.678	108.136
2,2-Dimethylbutane	-10.923	4.213	15.136	0.034	0.063	86.180	-38.003	107.540
2,3-Dimethylbutane	-10.778	4.100	14.878	0.001	0.000	86.180	-36.376	107.541
n-Heptane	-10.841	4.096	14.937	0.048	0.111	100.200	-39.984	125.174
2-Methylhexane	-10.946	4.098	15.044	0.060	0.136	100.200	-41.662	124.733
3-Methylhexane	-10.901	4.086	14.987	0.070	0.123	100.200	-40.714	124.625
3-Ethylpentane	-10.789	4.115	14.904	0.053	0.083	100.200	-39.081	124.680
2,2-Dimethylpentane	-10.921	4.183	15.104	0.023	0.056	100.200	-43.122	123.827
2,3-Dimethylpentane	-10.677	4.166	14.843	0.078	0.147	100.200	-39.626	124.257
2,4-Dimethylpentane	-10.922	4.199	15.121	0.147	0.282	100.200	-41.681	124.086
3,3-Dimethylpentane	-10.817	4.176	14.993	0.043	0.115	100.200	-40.766	123.485
2,2,3-Trimethylbutane	-10.647	4.119	14.766	0.060	0.128	100.200	-41.921	123.421
n-Octane	-10.981	4.011	14.992	0.000	0.000	114.230	-46.040	141.749
2-Methylheptane	-10.907	4.066	14.973	0.072	0.186	114.230	-46.649	141.289
3-Methylheptane	-10.735	4.084	14.819	0.059	0.103	114.230	-44.652	140.928
4-Methylheptane	-10.735	4.081	14.816	0.109	0.226	114.230	-44.577	140.896
3-Ethylhexane	-10.744	4.071	14.815	0.065	0.079	114.230	-44.179	141.216
2,2-Dimethylhexane	-10.707	4.116	14.823	0.011	0.049	114.230	-47.057	140.485
2,3-Dimethylhexane	-10.699	4.099	14.798	0.108	0.215	114.230	-45.388	140.670
2,4-Dimethylhexane	-10.765	4.108	14.873	0.139	0.245	114.230	-45.653	140.785
2,5-Dimethylhexane	-10.776	4.193	14.969	0.148	0.272	114.230	-46.557	140.735
3,3-Dimethylhexane	-10.756	4.081	14.837	0.040	0.105	114.230	-46.083	140.067
3,4-Dimethylhexane	-10.674	4.082	14.756	0.106	0.184	114.230	-44.517	140.666
3-Ethyl-2-Methylhexane	-10.655	4.073	14.728	0.073	0.171	114.230	-43.633	140.743
3-Ethyl-3-Methylhexane	-10.487	4.066	14.553	0.065	0.112	114.230	-42.720	140.355
2,2,3-Trimethylpentane	-10.628	4.099	14.727	0.044	0.077	114.230	-46.048	139.767
2,2,4-Trimethylpentane	-10.462	4.183	14.645	0.154	0.175	114.230	-44.918	139.956
2,3,3-Trimethylpentane	-10.628	4.099	14.727	0.042	0.078	114.230	-46.047	139.484
2,3,4-Trimethylpentane	-10.605	4.050	14.655	0.085	0.174	114.230	-42.777	140.512
2,2,3,3-Tetrametylbutane	-10.511	4.141	14.652	0.001	0.001	114.230	-46.589	138.508
n-Nonane	-10.934	3.989	14.923	0.035	0.104	128.260	-51.044	158.310
2-Methyloctane	-10.867	4.031	14.898	0.059	0.137	128.260	-51.662	157.844
3-Methyloctane	-10.803	4.023	14.826	0.067	0.121	128.260	-50.745	157.730
4-Methyloctane	-10.694	4.032	14.726	0.065	0.126	128.260	-49.721	157.736
3-Ethylheptane	-10.727	4.096	14.823	0.024	0.069	128.260	-48.654	157.809
4-Ethylheptane	-10.736	4.037	14.773	0.081	0.123	128.260	-49.360	157.847
2,2-Dimethylheptane	-10.788	4.087	14.875	0.025	0.06	128.260	-53.116	157.119
2,3-Dimethylheptane	-10.617	4.177	14.794	0.119	0.205	128.260	-49.564	157.210
2,4-Dimetylheptane	-10.759	4.171	14.93	0.141	0.244	128.260	-50.170	157.330
2,5-Dimethylheptane	-10.750	4.077	14.827	0.029	0.084	128.260	-51.437	157.238
2,6-Dimethylheptane	-10.715	4.118	14.833	0.114	0.189	128.260	-51.631	157.328

3,3-Dimethylheptane	-10.735	4.059	14.794	0.022	0.081	128.260	-51.088	156.618
3,4-Dimethylheptane	-10.669	4.081	14.75	0.092	0.155	128.260	-49.958	157.232
3,5-Dimethylheptane	-10.743	4.071	14.814	0.077	0.111	128.260	-50.787	157.162
4,4-Dimethylheptane	-10.738	4.068	14.806	0.036	0.063	128.260	-51.184	156.561
3-Ethyl,2-Methylhexane	-10.533	4.111	14.644	0.115	0.183	128.260	-50.020	157.173
4-Ethyl,2-Methylhexane	-10.662	4.062	14.724	0.088	0.152	128.260	-49.152	157.350
3-Ethyl,3-Methylhexane	-10.516	4.084	14.6	0.075	0.094	128.260	-47.020	157.195
3-Ethyl-4-Methylhexane	-10.589	4.051	14.64	0.054	0.106	128.260	-48.050	157.179
2,2,3-Trimethylhexane	-10.555	4.047	14.602	0.059	0.133	128.260	-51.193	156.104
2,2,4-Trimethylhexane	-10.464	4.092	14.556	0.116	0.122	128.260	-49.100	156.362
2,2,5-Trimethylhexane	-10.724	4.115	14.839	0.067	0.147	128.260	-53.742	156.229
2,3,3-Trimethylhexane	-10.435	4.013	14.448	0.044	0.095	128.260	-48.612	156.354

Table 3: Correlations among the physical properties examined.

	HOMO	***LUMO***	**ΔE**	**μ_Q**	**μ_{hyb}**	**MW**	**HF**	**ZPE**
HOMO	1							
LUMO	-0.55147	1						
ΔE	-0.9673	0.745016	1					
μ_{hyb}	0.256304	0.015267	-0.20031	1				
μ_{hyb}	0.134337	0.018456	-0.10181	0.92574	1			
MW	0.757216	-0.69924	-0.81809	0.236943	0.169372	1		
HF	-0.63308	0.611621	0.692191	-0.1734	-0.12269	-0.96469	1	
ZPE	0.744851	-0.70752	-0.81072	0.24154	0.175042	0.999517	-0.96355	1

3.1 Correlation analysis

The correlation analyses examined the electronic properties of HUMO, LUMO, HOMO-LUMO energy gap, dipole moment point-charge, dipole moment hybrid, molecular weight, heat of formation and zero-point energy of the alkanes, as shown in Table 3. The results showed that most of the properties were moderately correlated. HOMO and LUMO had a moderate correlation with each other (0.55147). HOMO-LUMO energy gap (ΔE) had a strong correlation with HOMO. ΔE also had a correlation of more than 0.8 for MW and ZPE. As seen in Table 3, the thermodynamic property of molecular weight (MW) had a strong correlation with the heat of formation (HF) and zero-point energy (ZPE). The correlation for hybrid and point-charge dipole moment was high at 0.9257. These indicate that both are correlated to each other. However, hybrid and point-charge dipole moments have low correlation with different properties.

3.2 Regression analysis

In regression analysis, electronic information such as HUMO, LUMO, ΔE, dipole moment point- charge, dipole moment hybrid, MW, HF and ZPE contain the information on the shape of a molecule, represented by molecular topology. The formulation of the multiple linear regression (MLR) model for the property of a molecule is given by:

$$\text{property} = a + \sum_{i=1} [b_i \ \text{index}_i] \quad (9)$$

where a and b_i are constant contributions to molecular topology. Linear regression was formulated for Xu and MTI indices. Regression was analysed by value or determination (R^2), and SE and F-values are shown in Tables 4 – 11.

3.2.1 Frontier molecular orbital energies

The relations of molecular topology indices with the energies of the highest occupied and lowest unoccupied molecular orbitals (HOMO and LUMO) were calculated by quantum chemistry methods via PM6, as given in Tables 4 and 5. In contrast, the relation of topology indices with ΔE is given in Table 7. Harary indices showed the highest value of R^2 of HOMO, LUMO and ΔE, which was 0.991. It was followed by Randic and Zegrab indices which showed similar patterns for all values for frontier molecular orbital energies. The R^2 values of HOMO energy for Balaban, Xu and Weiner indices were 0.965, 0.950 and 0.914, respectively. For Balaban, Xu, and Weiner indices, the R^2 values of LUMO energy were 0.968, 0.948, and 0.916, respectively. According to Table 7, for Balaban, Xu, and Weiner indices, the R^2 values of ΔE were 0.966 (for Balaban index), 0.949 (for Xu index), and 0.914 for Weiner index. All the topology indices resulted in R^2 values for frontier molecular orbital energies of more than 0.9 except for the MTI index. MTI showed that the values of R^2 were 0.826, 0.824 and 0.826 for HOMO, LUMO and Δ*E*, respectively. Harary indices had the highest regression R^2 value of 0.991.

Alkane is a saturated aliphatic hydrocarbon thatentirely composed of sigma(σ) bonds. HOMO and LUMO energy depend on the molecular structure of the molecular orbital band[35]. The electronic transition occurred at the range of less than 200 nm. Harary indices had a high value of R^2. The high value of R^2 for Harary indices might be because these indices use inverse distance matrix in the calculation, which plays a role in frontier molecular orbital energies. Nasiri reported that the Δ*E* was inversely proportional to the length of the *n*-alkane chain[36]. The decrease in HOMO and LUMO energy gap happens as the number of atoms increases, as proposed by Morisawa and co-workers [15]. They proposed that adding atoms in the molecular structure will cause the number of electrons that occupy the orbitals to increase. The increase of electrons in the orbital will induce a high HOMO energy level known as orbital "destabilization". These changes will cause the LUMO energy to either decrease or remain the same to observe the decrease in the HOMO-LUMO gap. An increase in the 'branch' in the alkane structure will cause a change to the Rydberg state, which changes with the distance of the nuclei[37]. The inverse distance matrix (Harary indices) has the highest regression constant. The indices can be represented by the structure of the molecule and the number of carbon atoms in the molecules, which are comparable to PM6 molecular calculation. Randic connectivity indices also showed good regression values. The calculation of these indices was based on the inverse degree of a vertex of carbon atom in the structure, showing that these indices can represent the branch of molecules and the number of carbon atoms. The high order of Randic indices indicates the path tree, in which the analogy is the contribution of the C–C bond that influences the next C–C neighbour by its second neighbour, third neighbour and so on. The lowest value of R^2 for MTI indices is plausibly inexhaustive to represent the molecular branch and the reduction of HOMO and LUMO energies as the carbon atom increases.

Table 4: Multiple regression equations for the HOMO

Properties	Indices	Equation
HOMO	Randic χ^o, χ^1, χ^2, χ^3, χ^4, χ^+	HUMO =-4.510(±0.436) χ^o-0.683(±0.368)χ^1+2.141(±0.384)χ^2+2.239(±0.371)χ^3 +2.652(±0.343)χ^4+2.541(±0.413)χ^+ N = 60, S.E = 1.116710, F = 921.343, R^2 = 0.990
	Zegrab – HM, HM^+,AM	HUMO = -0.035(±0.006) HM-46.436(±7.020) HM^++0.310(±0.310)AM N = 60, S.E = 1.956596, F = 587.110, R^2 = 0.969

	Weiner - w, ww, Δw, p	HUMO = -0.770(±0.135)w+0.124(±0.023)ww+0.151(±0.087)Δw+1.009(±0.456)p N = 60, S.E = 3.274366, F = 148.316, R^2 = 0.914
	Harary – h, hh	HUMO = 4.100(±0.240)h-7.053(±0.352)hh N = 60, S.E = 1.030365, F = 3249.405, R^2 = 0.991
	Balaban	HUMO = -3.373(±0.083)J N = 60, S.E = 2.021941, F = 1643.699, R^2 = 0.965
	MTI	HUMO = -0.034(±0.002)MTI N = 60, S.E = 4.526690, F = 280.714, R^2 = 0.826
	Xu	HUMO = -3.063(±0.092)Xu N = 60, S.E = 2.435627, F = 1114.420, R^2 = 0.950

Table 5: Multiple regression equations for the LUMO

Properties	Indices	Equation
LUMO	Randic – χ^o, χ^1, χ^2, χ^3, χ^4, χ^+	LUMO =1.747(±0.166)χ^o+0.224(±0.140)χ^1-0.776(±0.146)χ^2-0.842(±0.141)χ^3-1.024(±0.131)χ^4-1.018(±0.157)χ^+ N = 60, S.E = 0.425569, F = 925.396, R^2 = 0.990
	Zegrab – HM, HM^+,AM	LUMO = 0.014(±0.002)HM+17.494(±2.716)HM^+-0.119(±0.027)AM N = 60, S.E = 0.757165, F = 571.365, R^2 = 0.968
	Weiner - w, ww, Δw, p	LUMO = 0.300(±0.051)w - 0.049(±0.009)ww -0.053(±0.033)Δw – 0.400(±0.172)p N = 60, S.E = 1.232618, F = 153.073, R^2 = 0.916
	Harary – h, hh	LUMO = 2.695(±0.132)h – 1.567(±0.090)hh N = 60, S.E = 0.385876, F = 3380.552, R^2 = 0.991
	Balaban	LUMO = 1.290(±0.031)J N = 60, S.E = 0.745087, F = 1769.977, R^2 = 0.968
	MTI	LUMO = 0.013(±0.001)MTI N = 60, S.E = 1.742206, F = 275.521, R^2 = 0.824
	Xu	LUMO = 1.169(±0.035)Xu N = 60, S.E = 0.941520, F = 1086.417, R^2 = 0.948

Table 6: Multiple regression equations for the HOMO-LUMO Energy Gap (ΔE)

Properties	Indices	Equation
HOMO-LUMO Energy Gap (ΔE)	Randic – χ^o, χ^1, χ^2, χ^3, χ^4, χ^+	ΔE = 6.257(±0.601)χ^o+0.907(±0.508)χ^1-2.907(±0.508)χ^2-3.081(±0.511)χ^3-3.676(±0.473)χ^4-3.559(±0.569)χ^+ N = 60, S.E = 1.53911 , F = 926.242 , R^2 = 0.990
	Zegrab – HM, HM^+,AM	ΔE = 0.049(±0.009)HM + 63.930(±9.731)HM^+- 0.429(±0.095)AM N = 60, S.E = 2.712371, F = 583.277 , R^2 = 0.968
	Weiner - w, ww, Δw, p	ΔE = 1.070(±0.186)w – 0.173(±0.032)ww -0.205(±0.120)Δw – 1.410(±0.627)p N = 60, S.E = 4.50621 , F = 149.656 , R^2 = 0.914
	Harary – h, hh	ΔE = -5.667(±0.329)h –+9.747(±0.482)hh N = 60, S.E = 1.411423, F = 3307.347, R^2 = 0.991
	Balaban	ΔE = 4.663(±0.114)J N = 60, S.E = 2.764905 , F = 1679.777, R^2 = 0.966
	MTI	ΔE = 0.047(±0.003)MTI N = 60, S.E = 6.26786, F = 279.358, R^2 = 0.826
	Xu	ΔE = 4.233(±0.127)Xu N = 60, S.E = 3.37519, F = 1107.859, R^2 = 0.949

3.2.2 Point charge dipole moment

In semi-empirical calculation, the total molecular moment is given by[38]

$$\mu_{total} = \mu_Q + \mu_{hyb} \quad (10)$$

where μ_Q is the dipole moment due to the contribution of net point-charge located at the nuclear position. While μ_{hyb} is the contribution dipole moment from the effect of atomic polarization result from hybrid atomic orbital. Table 7 shows the multiple linear regression equation for μ_Q. All the topology indices showed the values of R^2 of less than 0.8. The highest R^2 was Randic indices at 0.778 followed by Weiner, Harary, Zegrab Balaban and Xu indices, as shown in Table 7. MTI indices showed the lowest value of $R^2 = 0.663$, which is lower than 0.7. The point charge dipole moment was related to the net charge distribution and the atoms' position vector. It also considered the electrostatic interaction between the charge interaction in the molecules[38]. The low value of R^2 in the contribution of dipole moment due to the calculation of topology indices was not based on a direct graph[39]. The topology indices did not imply the charge distribution and atomic hybrid orbital. Even though Weiner indices had introduced the polarity index *p* graph for trees, the value of R^2 was still lower. The graph tree approach for the number of unordered pairs of carbon atoms separated by three carbon-carbon bonds contributed to the molecule dipole moments.

3.2.3 Hybrid dipole moment

The contribution of hybrid dipole moment also showed that the R^2 value was much higher than the point charge of dipole moment. Randic indices showed the highest value of R^2, which was 0.804. It was followed by Harary, Xu, Zegrab, Weiner and Balaban indices, as shown in Table 8. The high value of R^2 in hybrid dipole moment compared to the point charge of dipole moment was plausibly due to the effect of σ-character of alkanes molecules due to the orbital hybridization in the molecular structure. The lowest value of R^2 was the MTI index, with the value of R^2 as 0.684. We believe that the low value of MTI in both hybrid and point charges might be due to non-isomorphic trees of the same size that may have the same value of MTI index[40].

Table 7: Multiple regression equations for the point charge dipole moment

Properties	Indices	Equation
POINT-CHARGE, μ_Q	Randic – χ^o, χ^1, χ^2, χ^3, χ^4, χ^+	μ_Q = 0.025(±0.015)χ^o-0.017(±0.012)χ^1-0.015(±0.013)χ^2-0.021(±0.012)χ^3 +0.020(±0.012)χ^4+0.010(±0.014)χ^+ N = 60, S.E = 0.037610 , F = 31.486 , R^2 = 0.778
	Zegrab – HM, HM^+,AM	μ_Q = 0.000210 (± 0.000126)HM+ 0.196590 (± 0.14414) HM^+- - 0.00106(± 0.00141)AM N = 60, S.E = 0.040178, F = 51.955, R^2 = 0.732
	Weiner - w, ww, Δw, p	μ_Q = 0.006(±0.002)w – 0.001(±0.000)ww – 0.001(±0.001)Δw – 0.012(±0.005)p N = 60, S.E = 0.03812, F = 42.734 , R^2 = 0.753
	Harary – h, hh	μ_Q = -0.008 (±0.009)h + 0.018(±0.013)hh N = 60, S.E = 0.038671, F = 85.886 , R^2 = 0.748
	Balaban	μ_Q = 0.020(±0.002)J N = 60, S.E = 0.040868, F = 146.738, R^2 = 0.713
	MTI	μ_Q = 0.000214(±0.000020)MTI N = 60, S.E = 0.044303, F = 116.069, R^2 = 0.663
	Xu	μ_Q = 0.019(±0.001)Xu N = 60, S.E = 0.039020, F = 166.684 , R^2 = 0.739

Table 8: Multiple regression equations for the hybrid dipole moment

Properties	Indices	Equation
HYBRID,	Randic χ^o, χ^1, χ^2, χ^3, χ^4, χ^+	μ_{hyb} = 0.050(±0.025)χ^o-0.014(±0.021)χ^1-0.032(±0.022)χ^2-0.042(±0.022)χ^3

μ_{hyb}		+0.023(±0.020)χ^4-0.001(±0.024)χ^+ N = 60, S.E = 0.065000, F = 36.883, R^2 = 0.804
	Zegrab – HM, HM^+,AM	μ_{hyb} = 0.00001(±0.00001)HM+0.471(±0.247)HM^+-0.003(±0.002)AM N = 60, S.E = 0.068785, F = 62.945, R^2 = 0.768
	Weiner - w, ww, Δw, p	μ_{hyb} = 0.011(±0.003)w – 0.002(±0.000)ww – 0.003(±0.002)Δw – 0.019(±0.010)p N = 60, S.E = 0.069892 , F = 45.526 , R^2 = 0.765
	Harary – h, hh	μ_{hyb} = -0.028(±0.015)h + 0.054(±0.023)hh N = 60, S.E = 0.065971, F = 104.624, R^2 = 0.783
	Balaban	μ_{hyb} = 0.038(±0.003)J N = 60, S.E = 0.070875, F = 172.549 , R^2 = 0.745
	MTI	μ_{hyb} = 0.0004(±0.000035)MTI N = 60, S.E = 0.078929, F = 127.702 , R^2 = 0.684
	Xu	μ_{hyb} = 0.036(±0.003)Xu N = 60, S.E = 0.067231, F = 198.327 , R^2 = 0.771

3.2.4 Molecular Weight

According to the results of regression analyses, all the topology indices showed a good relationship with the molecular weight. All the molecular topology indices had the values of R^2 of more than 0.91 for molecular weight, as shown in Table 9. This indicates that the topology indices are versatile to represent the molecular structure. Gumus and Turker also reported that their TG index correlated with the molecular weight of alkanes and alkenes[41]. The main objective of topology indices is to represent the molecular structure with a numerical value for isomer molecules. Most organic chemical structures have the same atomic composition or molecular weight but a different line or stereochemical formula which have other physical or chemical properties. The representation of molecular weight is not enough for further application in quantitative structure-activity or properties relationship (QSAR/QSPR) due to isomerization in molecular structure. The high value of R^2 shows that all the topology indices are capable of representing the molecular structure as a molecular descriptor for QSAR/QSPR application.

Table 9: Multiple regression equations for the molecular weight

Properties	Indices	Equation
MOLECULAR WEIGHT, MW	Randic χ^o, χ^1, χ^2, χ^3, χ^4, χ^+	MW =15.190(±1.164)χ^o+1.116(±0.983)χ^1-1.461(±1.026)χ^2-0.496(±0.989)χ^3 +1.806(±0.916)χ^4+4.308(±1.102)χ^+ N = 60, S.E = 2.98023 , F = 14352.389, R^2 = 0.999
	Zegrab – HM, HM^+,AM	MW = 0.389(±0.018)HM+209.006(±20.640)HM^+-0.647(±0.202)AM N = 60, S.E = 5.75313, F = 7688.565, R^2 = 0.998
	Weiner - w, ww, Δw, p	MW = 5.320(±0.742)w -0.788(±0.128)ww -0.800(±0.478)Δw – 4.409(±2.504)p N = 60, S.E = 17.98795 , F = 577.321 , R^2 = 0.976
	Harary – h, hh	MW = -10.027(±1.069)h + 25.923(±1.568)hh N = 60, S.E = 4.58836, F = 18147.172, R^2 = 0.998
	Balaban	MW = 35.405(±0.844)J N = 60, S.E = 20.52483 F = 1757.718, R^2 = 0.968
	MTI	MW = 0.376(±0.014)MTI N = 60, S.E = 32.35563, F = 672.050 , R^2 =0.918
	Xu	MW = 32.882(±0.295)Xu N = 60, S.E = 7.82947, F = 12425.810, R^2 = 0.995

3.2.5 Heat of formation

The heat of formation is the energy required to form one mole of an alkane compound in the gas phase from its elements in their natural state at standard temperature and pressure. The heat of formation of alkanes was calculated using the semi-empirical (PM6) method, showing a

strong correlation with molecular structure. The result (Table 10) showed that the molecular topology indices strongly correlated with the value of R^2 of more than 0.98. Bond energy plays an important role in predicting the heat of formation and thermochemistry properties. Furthermore, the bond order also contributes to the value of heat of formation. Randic index is based on the connectivity of carbon atoms in the molecular structure. Zagreb indices are based on the degree of edges. While Weiner, Harary and Balaban indices are based on the distance of vertex and the position of atoms. Meanwhile, MTI and Xu calculation indices are based on the distance of the vertex and the degree of edges. Randic indices showed a high value of R^2 = 0.999. The atom-bond connectivity index also agreed with the heat of formation calculated from *ab initio* and DFT (MP2, B3LYP) quantum chemical calculation[42].

3.2.6 Zero-Point Energy

All the molecular topology indices had the values of R^2 of more than 0.91 for zero-point energy, as shown in Table 11. These topology indices reflect the effects of molecular structure branches that have a high quality structure-property relationship. The zero-point energies (ZPE) of molecules can be related to the frequencies of the normal vibration modes in the molecular structure. Schulman and Disch reported that the zero-point energy of the hydrocarbons with molecular stoichiometry C_nH_m could be related to the empirical relation[43]:

$$ZPE = 3.8n + 7.12m - 6.19 \quad (11)$$

where *n* and *m* are the number of carbon and hydrogen in the molecules. While Rahal et al. reported that ZPE empirical relationship must have adjusted due to the number of atoms and branching[22]. The high values of R^2 in this calculation show that this approach may provide another alternative for empirical relation based on the branches' locations in the molecular structure. Thus, ZPE exhibits the structural characteristics of the bonding, shapes, and the number of atom increments related to molecular weight.

Table 10: Multiple regression equations for the heat of formation

Properties	Indices	Equation
HEAT OF FORMATION	Randic χ^o, χ^1, χ^2, χ^3, χ^4, χ^+	HF =-7.661(±0.688)χ^o-0.881(±0.581)χ^1+0.543(±0.606)χ^2+2.987(±0.584)χ^3 +1.582(±0.542)χ^4-0.110(±0.651)χ^+ N = 60, S.E =1.761018647, F =6539.024, R^2 = 0.999
	Zegrab – HM, HM^+,AM	HF = -0.167(±0.011)HM-110.155(±12.484)HM^++0.548(±0.122)AM N = 60, S.E = 3.479740849, F = 3335.087, R^2 = 0.994
	Weiner - w, ww, Δw, p	HF = -2.471(±0.343)w +0.371(±0.059)ww +0.353(±0.221)Δw +2.969(±1.160)p N = 60, S.E = 8.330070642, F =424.967, R^2 = 0.968
	Harary – h, hh	HF = 6.381(±0.573)h –13.820(±0.841)hh N = 60, S.E = 2.459899924 , F = 10038.567, R^2 = 0.997
	Balaban	HF = -14.127(±0.337)J N = 60, S.E = 8.184607974, F =1759.836, R^2 = 0.968
	MTI	HF = -0.149(±0.006)MTI N = 60, S.E = 13.66258291, F =593.715, R^2 = 0.910
	Xu	HF = -13.083(±0.174)Xu N = 60, S.E = 4.618178425, F =5653.786 , R^2 = 0.990

Table 11: Multiple regression equation for zero-point energy

Properties	Indices	Equation
ZERO-POINT ENERGY	Randic χ^o, χ^1, χ^2, χ^3, χ^4, χ^+	ZPE =20.695(±1.524)χ^o+1.843(±1.287)χ^1-3.831(±1.342)χ^2-2.284(±1.294)χ^3 +0.781(±1.200)χ^4+3.991(±1.442)χ^+

		N = 60, S.E = 3.900613, F =12701.546, R^2 = 0.999
	Zegrab – HM, HM^+,AM	ZPE = 0.465(±0.024)HM+276.284(±27.182)HM^+-0.935(±0.266)AM N = 60, S.E = 7.576597, F =6718.699, R^2 = 0.997
	Weiner - w, ww, Δw, p	ZPE = 6.676(±0.958)w – 0.993(±0.165)ww -1.069(±0.618)Δw – 5.705(±3.236)p N = 60, S.E = 23.242735, F = 522.966, R^2 = 0.974
	Harary – h, hh	ZPE = -14.237(±1.467)h +34.682(±2.151)hh N = 60, S.E = 6.295510, F =14609.258, R^2 = 0.998
	Balaban	ZPE = 43.606(±1.031)J N = 60, S.E = 25.062599, F = 1788.264, R^2 = 0.968
	MTI	ZPE = 0.462(±0.018)MTI N = 60, S.E = 40.487126, F = 648.860, R^2 = 0.917
	Xu	ZPE = 40.477(±0.383)Xu N = 60, S.E = 10.169703, F = 11160.274, R^2 = 0.9953

3.3 Validation

Cross-validation is a statistical method to investigate the predictive validity of a linear regression equation. Table 12 shows the value of ds^2 for all electronic properties. In this calculation method, the stable value ds^2 must be closer to the value of zero and positive[34, 44]. All the topology descriptors showed a positive value and were closer to zero for HOMO energy and molecular weight. The energy of LUMO, Randic, Balaban, and Xu indices showed a negative value of ds^2. For the HOMO-LUMO energy gap (ΔE), MTI and Xu showed a negative value. All the molecular topology descriptors showed a positive value for point-charge of dipole moment except for MTI. Randic, Zagreb, and Harary showed a relatively large value of ds^2. Meanwhile, the hybrid dipole moment exhibited a positive value of all descriptors. Two descriptors had a rather significant value of ds^2, which were Randic and MTI. Three descriptors; Zagreb, MTI, and Xu showed the negative value of ds^2. Finally, all descriptors showed positive values of zero-point energy except for Harary indices.

Table 12: The value of ds^2 validation

	HOMO	LUMO	ΔE	μ_Q	μ_{hyb}	MW	HF	ZPE
Randic	0.00261	-0.00020	0.00933	0.07481	0.06801	0.01806	0.00097	0.00090
Zagreb	0.02822	0.00299	0.02333	0.09765	0.03269	0.00109	-0.00712	0.00240
Wiener	0.04721	0.01569	0.02587	0.01415	0.00085	0.01027	0.00207	0.01094
Harary	0.00151	0.00296	0.00519	0.04820	0.01546	0.00926	0.00060	-0.00053
Balaban	0.01233	-0.00015	0.00788	0.02604	0.01655	0.00895	0.00203	0.00856
MTI	0.02531	0.00168	-0.03744	-0.03035	0.14541	0.02527	-0.04554	0.03113
Xu	0.00440	-0.00956	-0.01838	0.07176	0.01819	0.00220	-0.00232	0.00203

3.4 Principal Component Analysis

Correlation analyses were performed between the electronic properties of HOMO, LUMO, HOMO-LUMO energy gap (ΔE), the molecular dipole moment of point-charge and hybrid, molecular weight, the heat of formation and zero-point energy of the alkanes. The correlations of the examined electronic properties are shown in Table 3. The PCA results are tabulated in Figure 2. In the analysis for 7 properties and 60 stuctures of alkanes, as indicated in Table 13, two principal components with eigenvalues of greater than 1 were produced. The first principal component explained 60.388%, while the second principal component explained 26.457%, bringing a cumulative of 86.845% to the data variation. From Table 13, the classification of the properties was divided into three main areas. The first one was thermodynamic properties,

as Molecular Weight (MW) had a positive correlation with Heat of Formation (HF) but had a negative correlation with Zero-Point Energy (ZPE). The other classification areas were indicated to the molecular electronic properties, which indicate the correlations with HOMO and LUMO. From Table 14, it is shown that HOMO also has a negative correlation with LUMO. This could be due to HOMO and LUMO being known as frontier orbitals. HOMO can be found by locating the outermost orbital containing an electron and LUMO, which is the first orbital that does not have electron. The last classification of the properties is related to the atomic electronic properties – by which this study focused on point-charge dipole moment and hybrid dipole moment. These two properties showed strong correlations with each other. HOMO-LUMO energy gap (ΔE) showed a negative correlation with HOMO and a strong correlation with LUMO. However, PCA analysis could not be conducted for the HOMO-LUMO energy gap (ΔE) as it would give no positive definite (NPD) to the matrix. The Matrices can be NPD because ΔE has a linear dependency with HOMO and LUMO variables. Table 13 shows the component matrix of the correlations between the seven properties and the components. The results indicate that all the properties except dipole moment point-charge and dipole moment hybrid strongly affected PCA1. However, only dipole moment point-charge and hybrid dipole are significant on PCA2.

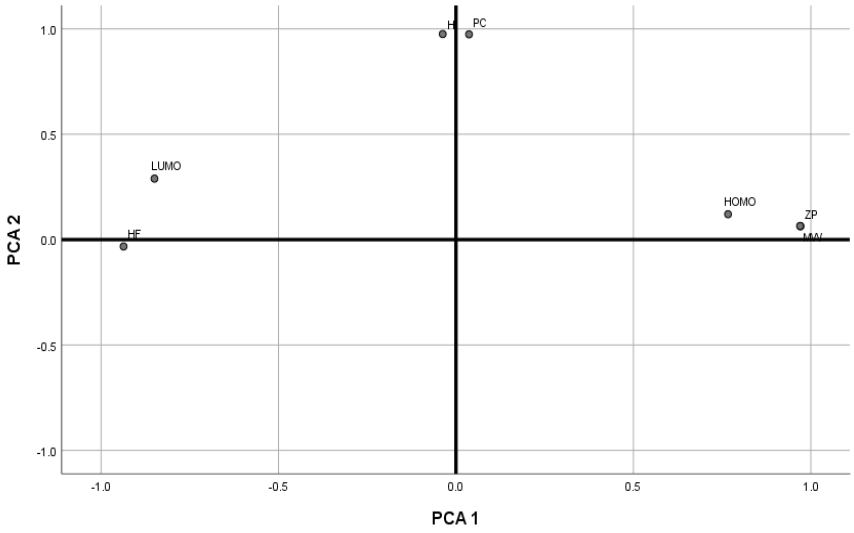

Figure 2: Principal components PC1 and PC2 for seven physical properties

Table 13: Results of principal component analysis for the physical properties: eigenvalues % variance and cumulative %

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	4.227	60.388	60.388	4.227	60.388	60.388
2	1.852	26.457	86.845	1.852	26.457	86.845
3	0.453	6.472	93.316			
4	0.389	5.563	98.879			
5	0.062	0.880	99.760			
6	0.017	0.239	99.999			
7	9.004E-05	0.001	100.000			

Table 14: PCA factor loadings for the physical properties

	Component	
	PCA1	PCA2
HOMO	0.800	-0.020
LUMO	-0.730	0.415
PC, μ_Q	0.397	0.898
H	0.325	0.912
MW	0.980	-0.108
HF	-0.935	0.131
ZPE	0.979	-0.107

4. Conclusions

The relationships of electronic calculation properties with seven degree-based indices and distance-based indices were obtained in correlation, regression, and principal component analysis. The results indicated that Harary indices obtained the highest regression R^2 value of 0.991, with the energies of the highest occupied and lowest unoccupied molecular orbitals (HOMO and LUMO). The inverse distance matrix (Harary indices) had the highest regression constant. The indices can be represented by the structure of the molecule and the number of carbon atoms in the molecules – comparable to PM6 molecular calculation. All the topology indices showed the value of R^2 of less than 0.8 for a molecular moment. The highest R^2 was Randic indices (0.778), and the lowest indices were MTI (R^2 = 0.663). The topology indices do not imply the charge distribution and atomic hybrid orbital. All the molecular topology indices had the values of R^2 of more than 0.91 for molecular weight, heat formation, and zero-point energy. These topology indices reflect the effects of branches of the molecular structures with high-quality structure-property relationships. The PCA results for seven properties and 60 stuctures of alkanes produced two principal components with eigenvalues of greater than 1. The first principal component explains 60.388%, while the second principal component explains 26.457%, bringing a cumulative of 86.845% to the data variation.

Acknowledgement

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. Additionally, the authors wish to express gratitude to Universiti Teknologi MARA for financial assistance.

References

[1] S. Noureen, A. A. Bhatti, and A. Ali, "Towards the solution of an extremal problem concerning the Wiener polarity index of alkanes," *Chaos, Solitons & Fractals,* vol. 144, pp. 110633, 2021.

[2] M. Dehmer, Z. Chen, F. Emmert-Streib, A. Mowshowitz, Y. Shi, S. Tripathi, and Y. Zhang, "Towards detecting structural branching and cyclicity in graphs: A polynomial-based approach," *Information Sciences,* vol. 471, pp. 19-28, 2019.

[3] S. A. K. Kirmani, P. Ali, and F. Azam, "Topological indices and QSPR/QSAR analysis of some antiviral drugs being investigated for the treatment of COVID-19 patients," *International Journal of Quantum Chemistry,* vol. 121, no. 9, pp. e26594, 2021.

[4] Z. Shao, A. Jahanbani, and S. M. Sheikholeslami, "Multiplicative Topological Indices of Molecular Structure in Anticancer Drugs," *Polycyclic Aromatic Compounds,* pp. 1-14, 2020.

[5] I. Gutman, "On degree-and-distance-based topological indices," *Revue Roumaine de Chimie,* vol. 66, no. 2, pp. 119-123, 2021.

[6] S. Hayat, and S. Khan, "Quality testing of spectrum-based valency descriptors for polycyclic aromatic hydrocarbons with applications," *Journal of Molecular Structure,* vol. 1228, pp. 129789, 2021.

[7] M. Randić, and J. Zupan, "On Interpretation of Well-Known Topological Indices," *Journal of Chemical Information and Computer Sciences,* vol. 41, no. 3, pp. 550-560, 2001.

[8] A. T. Balaban, "On pyrylium cations, molecular graphs, topological indices for QSAR, and various other structural problems," *Structural Chemistry,* pp. 1-11, 2019.

[9] M. M. Dehmer, Barbarini, N.N., Varmuza, K.K. and Graber, A.A., "Novel topological descriptors for analyzing biological networks," *BMC structural biology,* vol. 10, no. 1, pp. 1-17, 2010.

[10] S. Mondal, A. Dey, N. De, and A. Pal, "QSPR analysis of some novel neighbourhood degree-based topological descriptors," *Complex & Intelligent Systems,* vol. 7, no. 2, pp. 977-996, 2021.

[11] F. Safa, and M. Jafari Ghadimi, "Graph Theoretical Atom-Type-Based Descriptors for Structural Characterization and Retention Prediction of Acyclic Alkanes," *Moscow University Chemistry Bulletin,* vol. 76, no. 2, pp. 157-168, 2021.

[12] Y. Wang, Hu, P., Yang, J., Zhu, Y.A. and Chen, D., "C–H bond activation in light alkanes: a theoretical perspective," *Chemical Society Reviews,* 2021.

[13] J. X. Mao, P. Kroll, and K. A. Schug, "Vacuum ultraviolet absorbance of alkanes: an experimental and theoretical investigation," *Structural Chemistry,* vol. 30, no. 6, pp. 2217-2224, 2019.

[14] O. Rahaman, and A. Gagliardi, "Deep Learning Total Energies and Orbital Energies of Large Organic Molecules Using Hybridization of Molecular Fingerprints," *Journal of Chemical Information and Modeling,* vol. 60, no. 12, pp. 5971-5983, 2020.

[15] Y. Morisawa, S. Tachibana, M. Ehara, and Y. Ozaki, "Elucidating Electronic Transitions from σ Orbitals of Liquid n- and Branched Alkanes by Far-Ultraviolet Spectroscopy and Quantum Chemical Calculations," *The Journal of Physical Chemistry A,* vol. 116, no. 48, pp. 11957-11964, 2012.

[16] M. Ehara, and Y. Morisawa, "Theoretical and Experimental Molecular Spectroscopy of the Far-Ultraviolet Region," *Molecular Spectroscopy,* pp. 119-145, 2019.

[17] P. O. Dral, "Quantum Chemistry in the Age of Machine Learning," *The Journal of Physical Chemistry Letters,* vol. 11, no. 6, pp. 2336-2347, 2020.

[18] B. Huang, and O. A. von Lilienfeld, "Ab Initio Machine Learning in Chemical Compound Space," *Chemical Reviews,* vol. 121, no. 16, pp. 10001-10036, 2021.

[19] Y. Han, I. Ali, Z. Wang, J. Cai, S. Wu, J. Tang, L. Zhang, J. Ren, R. Xiao, Q. Lu, L. Hang, H. Luo, and J. Li, "Machine learning accelerates quantum mechanics predictions of molecular crystals," *Physics Reports,* vol. 934, pp. 1-71, 2021.

[20] J. Cao, U. Ali, M. Javaid, and C. Huang, "Zagreb Connection Indices of Molecular Graphs Based on Operations," *Complexity,* vol. 2020, pp. 7385682, 2020.

[21] R. Mustafa, A. M. Ali, and A. M. Khidhir, "M_n – Polynomials of Some Special Graphs," *Iraqi Journal of Science,* vol. 62, no. 6, pp. 1986-1993, 2021.

[22] M. Rahal, I. Bouimadaghene, R. Drissi El Bouzaidi, I. Bouabdallah, F. Malek, and A. El Hajbi, "Accessible approaches for vibrational zero point energy calculation of organoboron compounds," *Vibrational Spectroscopy,* vol. 110, pp. 103131, 2020.

[23] Z. Du, A. Jahanbai, and S. M. Sheikholeslami, "Relationships between Randic index and other topological indices," *Communications in Combinatorics and Optimization,* vol. 6, no. 1, pp. 137-154, 2021.
[24] W. Carballosa, D. Pestana, J. M. Sigarreta, and E. Tourís, "Relations between the general sum connectivity index and the line graph," *Journal of Mathematical Chemistry,* vol. 58, no. 10, pp. 2273-2290, 2020.
[25] G. H. Shirdel, H. Rezapour, and R. Nasiri, "Lower and Upper Bounds for Hyper-Zagreb Index of Graphs," *Iraqi Journal of Science,* vol. 61, no. 6, pp. 1401-1406, 2020.
[26] H. Wiener, "Structural determination of paraffin boiling points," *Journal of the American chemical society,* vol. 69, no. 1, pp. 17-20, 1947.
[27] M. Randić, "Novel molecular descriptor for structure—property studies," *Chemical Physics Letters,* vol. 211, no. 4, pp. 478-483, 1993.
[28] Z.-B. Peng, A. R. Nizami, Z. Iqbal, M. M. Munir, H. M. Waqar Ahmed, and J.-B. Liu, "Wiener and Hyper-Wiener Indices of Polygonal Cylinder and Torus," *Complexity,* vol. 2021, pp. 8882646, 2021.
[29] J. Cai, P. Wang, and L. Zhang, "Harary index of Eulerian graphs," *Journal of Mathematical Chemistry,* vol. 59, no. 5, pp. 1378-1394, 2021.
[30] A. T. Balaban, "Highly discriminating distance-based topological index," *Chemical Physics Letters,* vol. 89, no. 5, pp. 399-404, 1982.
[31] K. C. Das, "On the Balaban Index of Chain Graphs," *Bulletin of the Malaysian Mathematical Sciences Society,* vol. 44, no. 4, pp. 2123-2138, 2021.
[32] B. Osaghi, and F. Safa, "QSPR study on the boiling points of aliphatic esters using the atom-type-based AI topological indices," *Rev. Roum. Chim.,* vol. 64, no. 2, pp. 183-189, 2019.
[33] C. I. Mosier, "I. Problems and designs of cross-validation 1," *Educational Psychological Measurement,* vol. 11, no. 1, pp. 5-11, 1951.
[34] K. Kelley, and J. H. H. Bolin, "Multiple Regression," *Handbook of Quantitative Methods for Educational Research,* T. Teo, ed., pp. 71-101, Rotterdam: SensePublishers, 2013.
[35] U. W. Pohl, "Electronic Properties of Organic Semiconductors," *Epitaxy of Semiconductors,* pp. 177-205: Springer, 2020.
[36] R. Nasiri, and V. M. Gun'ko, "Quantum Mechanical effects in n-alkane droplets," in ILASS 2013 : 25th European Conference Liquid Atomization & Spray Systems, Chania, Crete, Greece, 2013.
[37] J. X. Mao, Kroll, P. and Schug, K.A., "Vacuum ultraviolet absorbance of alkanes: an experimental and theoretical investigation," *Structural Chemistry,* vol. 30, no. 6, pp. 2217-2224, 2019.
[38] J. A. Pople, and G. A. Segal, "Approximate Self-Consistent Molecular Orbital Theory. II. Calculations with Complete Neglect of Differential Overlap," *The Journal of Chemical Physics,* vol. 43, no. 10, pp. S136-S151, 1965.
[39] J. Monsalve, and J. Rada, "Vertex-degree based topological indices of digraphs," *Discrete Applied Mathematics,* vol. 295, pp. 13-24, 2021.
[40] W. R. Mueller, K. Szymanski, J. V. Knop, and N. Trinajstic, "Molecular topological index," *Journal of Chemical Information and Computer Sciences,* vol. 30, no. 2, pp. 160-163, 1990.
[41] S. Gumus, and L. Turker, "TG index and its application on alkanes and alkenes," *Chem. J,* vol. 1, pp. 103, 2015.
[42] A. Mahdieh, and I. Ali, "Atom-Bond Connectivity Index," *New Frontiers in Nanochemistry,* Mihai V. Putz, ed., Burlington, Canada: Apple Academic Press, 2020.
[43] J. M. Schulman, and R. L. Disch, "A simple formula for the zero-point energies of hydrocarbons," *Chemical Physics Letters,* vol. 113, no. 3, pp. 291-293, 1985.
[44] L. Fish, "The Importance of Invariance Procedures as against Tests of Statistical Significance," *ERIC,* 1986.