Prediction of autoignition temperatures (AITs) for hydrocarbons and compounds containing heteroatoms by the quantitative structure–property relationship †

Yeong Suk Kim,[a] Sung Kwang Lee,[c] Jae Hyun Kim,*[b] Jung Sup Kim[c] and Kyoung Tai No*[c]

[a] *Department of Chemistry, Kongju National University, Kongju 314-700, Korea*
[b] *Department of Chemical Education, Kongju National University, Kongju 314-700, Korea*
[c] *Computer Aided Molecular Design Research Center and Department of Chemistry, Soong Sil University, Seoul 156-743, Korea*

Received (in Cambridge, UK) 23rd July 2002, Accepted 27th September 2002
First published as an Advance Article on the web 4th November 2002

Regression models that are useful for the explanation and prediction of autoignition temperatures of diverse compounds were provided by a quantitative structure–property relationships study (QSPR).

Genetic functional approximation was used to find the best multiple linear regression within 72 molecular descriptors. After validation by correlation of the prediction set, nine descriptor models were evaluated in the best model. The nine descriptors were I_{al}, I_{ke}, radius of gyration, $^1\chi^v$, SC-2, the Balaban index JX, density, Kappa-3-AM and Jurs-FNSA-2, and information of structure features and their interactions was provide.

The result of the best regression model showed that the square of the correlation coefficient (R^2) for the autoignition temperature of the 157-member training set was 0.920, and the root mean square error (RMSE) was 25.876. The R^2 of AIT for a 43-member prediction set was 0.910, and the RMSE was 28.968.

Introduction

Autoignition temperature (AIT) is the lowest temperature at which a material generates heat automatically without an external ignition source, reaching ignition and combustion by accumulated heat. Since autoignition occurs in air without the presence of an ignition source, it is an important fire safety parameter of combustible materials. Autoignition is also related to the phenomenon of engine knock.[1,2] The octane number of a fuel is a measure of how readily knocking occurs and depends on the chemical nature of the fuel.

The origins of the automatic generation of heat can be the decomposition, oxidation, absorption or polymerization of a material. Usually, the aforementioned chemical processes occur in oxidative materials or unsaturated compounds. AIT is generally accelerated by high temperatures and humidity in the air. The autoignition mechanism proceeds by a free radical reaction and the stability of the free radical intermediates determines the ease of oxidation. For aliphatic hydrocarbons, this stability follows the pattern, tertiary > secondary > primary > methyl radical, *i.e.* the more branches there are, the more stable molecule.[3] In addition, the position of the branch is important. As the number of the methylene of the branch increases, so does the AIT. The structural features that affect the temperature of autoignition are the chain length, addition of methyl groups, the degree of unsaturation, the degree of branching, aromaticity and the functional groups of a compound.

Experimentally determined values of AIT depend on the methods (such as size and shape of containers and method of heating)[4–8] and different ignition temperatures of the same substance have been reported by different laboratories.

Several studies have been performed to describe the relationship between AIT and structure,[9–13] since autoignition is a physicochemical phenomenon that is closely related to molecular structure and physical properties. Several studies applied the use of physical properties such as the critical pressure of the gas and parachor (defined by the equation $P_A = \Gamma^{1/4} M/(D - d)$, where P_A is parachor, M is molecular weight, D is density of liquid state, d is vapour density and Γ is the surface tension).[9–11]

Using a quantitative structure–property relationship(QSPR) study with calculated molecular descriptors has advantages that it decreases experimental costs and eliminates any danger involved in the experiment. Several kinds of QSPR models for AIT prediction were developed using multiple linear regressions and computational neural networks with calculated molecular descriptors.[12,13]

The purpose of this study is to obtain informative descriptors that can describe AIT for all hydrocarbons and compounds containing heteroatoms, and to develop a QSPR model which calculates the AIT using descriptors without grouping according to compound.

Computation

A Database construction

In this study, 200 compounds summarized in Tables 3and 4, were introduced for the QSPR study of molecular properties and AIT. 92 molecules are hydrocarbons; 26 alkanes, 16 alkenes, 20 cycloalkanes, 3 cycloalkenes and 27 aromatic hydrocarbons.

A subset of 157 of the original 200 structures was selected at random as a training set, and was composed of 70 hydrocarbon compounds and 87 heteroatom containing compounds (Table 3). The remaining 43 prediction set contains 22 hydrocarbon compounds and 21 heteroatom containing compounds (Table 4).

The AIT of the hydrocarbons ranged from 201 to 540 °C. The remaining 108 molecules were heteroatom containing compounds; 33 alcohols/phenols, 13 organic acids, 33 esters,

† Electronic supplementary information (ESI) available: descriptors involved in the best models derived for AIT. See http://www.rsc.org/suppdata/p2/b2/b207203c/

Table 1 The best QSPR equation using 4–12 descriptors and their regression statistics in the training and prediction set[a]

No. of descriptor	Descriptors	Training set		Prediction set	
		R^2	RMSE	R^2	RMSE
4	I_{al}, RoG, $^1\chi^v$, SC-2	0.871	32.867	0.908 (0.918)	26.837 (25.557)
5	I_{al}, RoG, $^1\chi^v$, SC-2, Jurs-FPSA-3	0.888	30.577	0.904 (0.916)	28.270 (26.699)
6	I_{al}, I_{ke}, RoG, $^1\chi^v$, SC-2, density	0.898	29.448	0.903 (0.919)	28.556 (26.362)
7	I_{al}, I_{ke}, RoG, $^1\chi^v$, SC-2, density, JX	0.908	27.793	0.898 (0.917)	30.224 (27.315)
8	I_{al}, I_{ke}, RoG, $^1\chi^v$, SC-2, density, JX, Jurs-FNSA-2	0.916	26.578	0.906 (0.928)	29.264 (25.843)
9	I_{al}, I_{ke}, RoG, $^1\chi^v$, SC-2, density, JX, Jurs-FNSA-2, Kappa-3-AM	0.920	25.876	0.910 (0.932)	28.968 (25.212)
10	I_{al}, I_{ke}, RoG, $^1\chi^v$, SC-2, density, JX, Jurs-RPCS, MolRef, Jurs-DPSA-3	0.922	25.462	0.899 (0.923)	30.711 (27.004)
11	I_{al}, I_{ke}, RoG, $^1\chi^v$, SC-2, density, JX, Jurs-RPCS, MolRef, Jurs-DPSA-3, Kappa-3-AM	0.925	25.002	0.903 (0928)	30.254 (26.189)
12	I_{al}, I_{ke}, RoG, $^0\chi^v$, SC-2, Kappa-2, JX, Jurs-FNSA-2, Kappa-2-AM, Vm, Rotlbonds, Area	0.928	24.388	0.896 (0.904)	28.911 (28.060)

[a] R^2: squared coefficient of correlation; RMSE: root mean square error. Values in brackets are statistic values of the prediction set without butane.

8 amines, 6 aldehydes, 5 ethers, 7 ketones and 3 halogenated, and the AIT of these compounds lie between 187 and 590 °C. The conformations of the molecules having energetically stable potential energy were determined by energy minimization with Merck Molecular Force Field (MMFF).[14]

B Calculation of molecular descriptors

For computation of all 72 molecular descriptors, we used the Cerius[2] v.4.6 molecular modeling software from Accelrys Inc. (San Diego, CA). These descriptors were grouped into six categories: topological, electronic, spatial, electrostatic, thermodynamic, and structural descriptors. Firstly, the topological descriptors, such as the Wiener index,[15] Balaban index,[16] Randic indices[17] and Kier–Hall indices,[18] were introduced to describe the degree of branching of a molecule.

Secondly, the sum of atomic polarizability, dipole moment, highest occupied molecular orbital HOMO energy, and lowest unoccupied molecular orbital LUMO energy of a molecule, obtained by semi-empirical AM1 single point calculation, were used for electronic descriptors.

Thirdly, the radius of gyration, density and volume descriptors were used as the spatial descriptors.

Fourthly, charged partial surface area (CPSA) descriptors were proposed for electrostatic descriptors as in Jurs *et al.*[19] The set of 30 CPSA descriptors was calculated as a combination of the contributions of atomic partial charges and the total molecular solvent-accessible surface area.

Fifthly, thermodynamic descriptors (*i.e.*, AlogP,[20] hydration free energy and 1-octanol solvation free energy) were introduced to describe the degree of interaction with the environment.

Finally, the structural descriptors such as molecular weight, number of rotational bonds, functional group indicator descriptors, number of hydrogen bond acceptors and hydrogen bond donors, were included to illustrate the influence of the shape and structural features of a molecule on AIT.

C Determination of the optimum descriptor set and description of the AIT by multiple linear regression

Genetic functional approximation (GFA), based on evolutionary principles, offers a means of exploring descriptor combinations nonexhaustively but with a good chance of finding effective combinations. GFA works with a population of individuals, each of whom is a candidate solution to the problem. These individuals mate with each other, mutate, crossover, reproduce and evolve through successive generations towards an optimum solution. Using the GFA software in the Cerius[2] and beginning with a population of 5000 initial equations, 300 000 generation of equations were sufficient to achieve convergence. We controlled the number of descriptors in initial equation to a set between 4–12.

The most commonly quoted statistic used to describe the fitness of data for a regression model is the square of the correlation coefficient (R^2) and the root mean square error (RMSE).

$$R^2 = 1 - \frac{\sum_{i=1}^{n}\left(\mathrm{AIT}_i^{\mathrm{obs}} - \mathrm{AIT}_i^{\mathrm{pre}}\right)^2}{\sum_{i=1}^{n}\left(\mathrm{AIT}_i^{\mathrm{obs}} - \overline{\mathrm{AIT}}\right)^2} \quad (1)$$

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}\left(\mathrm{AIT}_i^{\mathrm{obs}} - \mathrm{AIT}_i^{\mathrm{pre}}\right)^2}{n}} \quad (2)$$

where i represents ith molecule, n is the number of compounds, and $\mathrm{AIT}_i^{\mathrm{obs}}$, $\mathrm{AIT}_i^{\mathrm{pre}}$ and $\overline{\mathrm{AIT}}$ symbolize observed, predicted and average value of the AIT, respectively.

All statistical analyses of the best equations obtained by GFA was performed using the SAS JMP package (V.3.2.5, SAS Institute Inc, Cary, NC, USA).

Results and discussion

Genetic functional approximation (GFA) was used to select the descriptors for the best model and the quality thereof was determined by examining root mean square error (RMSE) and the square of the correlation coefficient (R^2) for the training and prediction. The results of the best QSPR models using 4–12 descriptors are given in Table 1 and the descriptors in these models are summarized in Table 2.

The most common descriptors of these models using 4–12 descriptors are I_{al}, radius of gyration (RoG), $^1\chi^v$ and SC-2, indicating that these descriptors are the most significant descriptors in AIT prediction.

Among these models, four-descriptor and nine-descriptor models show high stability for the regression equation owing to high R^2 values of the prediction set.

$$\mathrm{AIT} = 410.326 - 132.94(I_{al}) + 62.348(\mathrm{RoG}) - 148.211(^1\chi^v) + 32.3665(\mathrm{SC\text{-}2}) \quad (3)$$

where $n = 157$, $R^2_{\mathrm{training}} = 0.871$, $F = 256.75$, RMSE = 32.867, $R^2_{\mathrm{cv}} = 0.865$, $n = 43$, $R^2_{\mathrm{prediction}} = 0.908$, RMSE = 26.837.

$$\mathrm{AIT} = 147.606 - 134.152(I_{al}) + 57.0323(I_{ke}) + 86.2026(\mathrm{RoG}) - 140.845(^1\chi^v) + 27.8304(\mathrm{SC\text{-}2}) + 200.084(\mathrm{density}) + 28.3369(\mathrm{JX}) - 4.55002(\mathrm{Kappa\text{-}3\text{-}AM}) + 66.811(\mathrm{Jurs\text{-}FNSA\text{-}2}) \quad (4)$$

Table 2 Descriptors involved in the best models derived for AIT

Label	Type	Definition
SC-2	Topological	Number of pairs with connected edges
Kappa-1		First order Kier's shape index
Kappa-2		Second order Kier's shape index
Kappa-3-AM		Third order Kier's alpha-modified shape index
JX		Balaban index
$^0\chi^v$		Zero order valence connectivity index
$^1\chi^v$		First order valence connectivity index
Density	Spatial	Density
RoG		Radius of gyration
Area		Molecular surface area
V_m		Molecular volume
Jurs-RPCS	Electrostatic	Relative positive charged surface area
Jurs-FNSA-2		Fractional total charge weighted partial negative surface area
Jurs-FPSA-3		Fractional atomic charge weighted partial positive surface area
Jurs-DPSA-3		Difference in atomic charge weighted partial surface area
I_{al}	Structural	Aldehyde group indicator
I_{ke}		Ketone group indicator
Rotbonds		Number of rotatable bonds

where $n = 157$, $R^2_{training} = 0.920$, $F = 187.88$, RMSE = 25.876, $R^2_{cv} = 0.909$, $n = 43$, $R^2_{prediction} = 0.910$, RMSE = 28.968.

Butane was a common outlier in these models, having a large difference between calculated and experimental AIT for the prediction set (Fig. 1). This compound may not have been

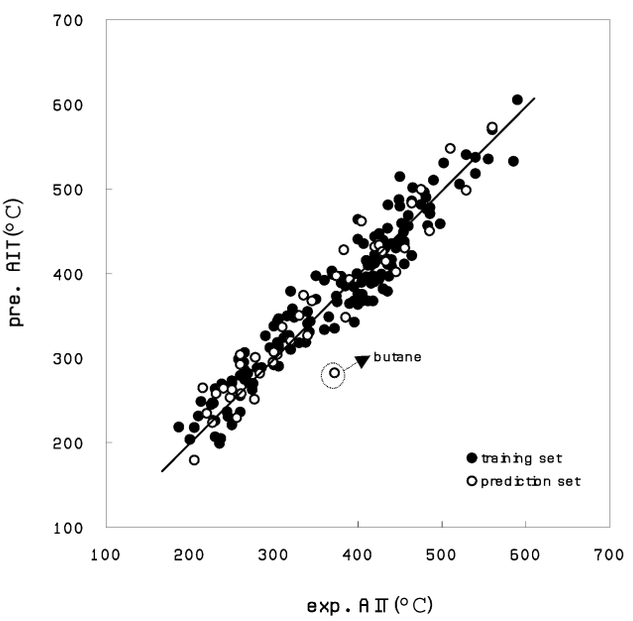

Fig. 1 Experimental *versus* predicted AIT values of a MLR selected nine-descriptor model.

modeled well as it has the lowest molecular weight in the entire data set. Excluding butane, the remaining prediction set of 42 compounds has an R^2 of 0.932 and RMSE (root mean square error) of 25.212 by the nine-descriptor model. The experimental error in AIT is estimated to be ±30 °C. For a good model, training and prediction errors should be similar to each other and close to the experimental error.[9,10]

Therefore, we concluded that a nine-descriptor model (eqn. (4)) could be considered a stable final model. Table 3 and 4 show the calculated and the experimental AIT values for the 157 training set and the 43 prediction set using the nine-descriptor model and these data are plotted in Fig. 1. The correlation matrix (Table 5) demonstrates that the descriptors involved in the nine-descriptor model are not mutually interrelated by correlation, so this model is of relevance.

The statistic results of the nine-descriptor model for the entire data set of 200 compounds are given in Table 6. The important descriptor parameters are indicated by the sign and magnitude of the *t*-test statistics in Table 6.

The most significant descriptors in the nine-descriptor model are topological indices such as $^1\chi^v$ and SC-2. $^1\chi^v$ is a direct representation of the molecular structure that encodes the various degrees of skeletal branching. The number of second order subgraphs (SC-2) in a molecular graph is the number of pairs with connected edges. These descriptors as quantifiers of molecular branching within a short ranges of 1–2 order also show the importance of the molecular shape in determining AIT.

Some topological descriptors correlate with AIT in the nine-descriptor model. The Balaban index,[17] JX, is based on a Randic formula,[18] replacing vertex degrees by average distance sums. JX reflects the relative connectivity and effective size of the carbon chain to which multiple methyl groups are attached. The more the number of methyl groups increases, the higher the AIT value. Therefore, the AIT value has a linear relationship with the Balaban index JX. The octane number of alkanes depends mainly on branching. The Kappa-3-AM descriptor[19] (*i.e.*, Kier alpha modified shape index based on the three-bond paths in the molecules) is an index that encodes information about the centrality of branching.

The descriptors I_{al} and I_{ke}, indicating the aldehyde and ketone groups, were assigned values of 1 for molecules with these functional groups, while other compounds were assigned the value 0. The negative regression coefficient for I_{al} and positive regression coefficient for I_{ke} in Table 6 reflect the fact that relative stable ketone compounds have higher AITs and unstable aldehydes have lower AITs.[12] Jurs *et al.*[11] developed predictive models of AIT and classified them into several groups, *e.g.* hydrocarbon, halohydrocarbons, and compounds containing oxygen, sulfur and nitrogen, based on multiple linear regressions and computational neural networks. However, we tried to discriminate between the aldehydes, ketones and the other groups using I_{al} and I_{ke}, and developed a model for the entire data set without dividing the data set.

The radius of gyration is defined as the largest radius that the primary axis can sweep out upon rotation. The positive regression coefficient of this descriptor in these models reflects the fact that structures with extended conformation have higher AITs than those with wrapped conformation in the case of isomeric compounds having the same number of carbons.

Density is defined as the ratio of molecular weight to molecular volume. It reflects the types of atoms and how tightly they are packed in a molecule.

When the number of carbons are the same, the AIT increases as unsaturated bonds or the number of aromatic rings increases. In other words, the compounds containing

Table 3 Comparison of AIT values between the experimental values of 70 hydrocarbon compounds and 87 compounds containing heteroatom in a training set and the predicted AIT values obtained using 9 descriptors

No.	Structure name	AIT_{exp}/°C	AIT_{pre}/°C	No.	Structure name	AIT_{exp}/°C	AIT_{pre}/°C
Hydrocarbon				Heteroatom			
1	Pentane	265	274.8	71	Ethanol	400	363.1
2	Heptane	213	248.6	72	1-Propanol	372	335.1
3	Nonane	205	217.8	73	2-Butanol	390	364.3
4	Decane	201	203.8	74	*tert*-Butanol	460	468.6
5	2-Methylpentane	264	294.8	75	Cyclohexanol	300	337.8
6	2-Methyloctane	227	246.8	76	Benzylalcohol	436	481.0
7	3-Methyloctane	228	246.3	77	1-Hexanol	285	288.7
8	4-Methyloctane	232	244.7	78	1-Octanol	260	255.2
9	2-Methylnonane	214	231.6	79	Allyl alcohol	370	398.9
10	4-Ethyloctane	237	205.0	80	3,5-Dimethylphenol	555	535.2
11	2,2-Dimethylbutane	405	375.2	81	3-Methyl-1-butanol	340	340.8
12	2,3-Dimethylpentane	338	318.6	82	3-Pentanol	365	348.6
13	2,3-Dimethylbutane	396	342.5	83	4-Heptanol	295	312.3
14	3,3-Dimethylheptane	330	317.9	84	1-Nonanol	260	236.0
15	2,2,3-Trimethylbutane	412	409.8	85	2,4-Dimethyl-3-pentanol	395	384.2
16	2,3,3-Trimethylpentane	430	382.2	86	2-Methyl-1-propanol	405	367.3
17	2,2,4-Trimethylpentane	418	367.2	87	2-Methyl-2-butanol	435	416.0
18	2,3,3,4-Tetramethylpentane	437	396.7	88	2-Octanol	265	285.5
19	2,2,3,3-Tetramethylpentane	430	439.6	89	2-Pentanol	343	343.4
20	1-Pentene	290	326.4	90	2-Propanol	399	399.7
21	1-Hexene	265	306.6	91	4-Methyl-2-pentanol	340	354.7
22	1-Octene	250	270.5	92	2-Ethyl-1,3-hexanediol	360	333.5
23	1-Decene	244	236.5	93	2-Ethyl-1-hexanol	270	271.2
24	2-Methyl-1-pentene	306	346.4	94	2,2-Dimethyl-1,3-propanediol	400	463.7
25	2,3-Dimethyl-2-butene	407	435.2	95	1-Decanol	250	220.7
26	2,4,4-Trimethyl-1-pentene	420	416.7	96	Ethylene glycol	410	396.1
27	2,3-Dimethyl-1-butene	369	402.7	97	1-Heptanol	275	270.2
28	2-Ethyl-1-butene	324	348.3	98	1,2-Propanediol	421	415.9
29	3-Methyl-1-butene	374	373.2	99	Glycerol	400	440.6
30	4-Methyl-1-pentene	304	343.3	100	Acetic acid	464	485.8
31	1,3-Hexadiene	320	378.8	101	Pentanoic acid	400	375.8
32	2-Methylpropene	464	421.3	102	Tetradecanoic acid	235	199.2
33	Cyclohexane	320	279.1	103	Propionic acid	440	435.0
34	Cyclopentane	260	310.0	104	2,2-Diethylpropionoic acid	450	514.7
35	Methylcyclohexane	258	298.7	105	Isopentanoic acid	416	409.8
36	Ethylcyclohexane	238	269.5	106	Isobutyric acid	460	455.8
37	1,3-Dimethylcyclohexane	306	314.3	107	Dodecanoic acid	230	225.7
38	1,4-Dimethylcyclohexane	304	318.5	108	Decanoic acid	230	263.8
39	Isobutylcyclohexane	274	262.5	109	*o*-Phthalic acid	590	605.3
40	*tert*-Butylcyclohexane	342	331.1	110	Ethyl formate	435	410.9
41	4-Isopropyl-1-methylcyclohexane	306	290.3	111	Isobutyl formate	425	421.9
42	n-Hexylcyclopentane	228	226.0	112	Ethyl acetate	425	441.0
43	Butylcyclohexane	246	231.0	113	Propyl acetate	435	416.7
44	Propylcyclopentane	269	282.0	114	Butyl acetate	380	389.0
45	1,3,5-Trimethylcyclohexane	314	325.1	115	Isobutyl acetate	420	393.6
46	Decalin	268	281.1	116	Methyl propionate	455	458.2
47	Cyclopentene	395	367.6	117	Ethyl propionate	445	430.1
48	1,3-Cyclohexadiene	360	392.0	118	Methyl butyrate	455	433.8
49	Benzene	498	458.6	119	Ethyl benzoate	490	510.2
50	Toluene	482	475.7	120	Butyl benzoate	435	453.3
51	Ethylbenzene	432	429.2	121	Pentyl acetate	375	366.3
52	n-Butylbenzene	412	367.5	122	Ethyl butyrate	440	408.5
53	1,4-Diethylbenzene	430	419.6	123	Phenyl benzoate	560	569.7
54	Biphenyl	540	517.9	124	Isopropyl acetate	425	447.1
55	Naphthalene	540	537.4	125	Isobutyl propionate	435	415.7
56	1-methylnaphtalene	529	540.7	126	Isopropyl butyrate	435	415.6
57	1,2,3-Trimethylbenzene	479	496.1	127	Propyl propionate	430	381.5
58	1,2,4-Trimethylbenzene	521	505.6	128	n-Butyl formate	322	415.4
59	1,2-Diethylbenzene	404	389.2	129	*sec*-Butyl acetate	410	357.9
60	1,3-Diethylbenzene	455	411.0	130	Butyl butyrate	350	369.2
61	*tert*-Butylbenzene	450	479.6	131	Butyl propionate	385	385.1
62	1-Ethylnaphthalene	481	489.7	132	Methyl formate	450	439.0
63	1-Methyl-2-ethylbenzene	448	456.3	133	Isopentyl acetate	380	396.6
64	1-Methyl-4-ethylbenzene	483	435.2	134	Propyl formate	435	379.1
65	Isobutylbenzene	428	399.0	135	Propyl butyrate	420	390.5
66	Diphenylmethane	486	470.7	136	Methyl acetate	475	481.6
67	2-Ethylbiphenyl	449	530.4	137	Isopropyl formate	440	416.4
68	2-Propylbiphenyl	452	458.9	138	Phenyl acetate	585	532.6
69	*sec*-Butylbenzene	418	397.1	139	1,2-Propanediamine	416	388.4
70	2-Methylbiphenyl	502	487.1	140	Diisopropylamine	316	349.9

Table 3 (*Contd.*)

No.	Structure name	AIT_{exp}/°C	AIT_{pre}/°C
Heteroatom			
141	Diethylamine	312	324.4
142	Dimethylamine	400	363.7
143	Propylamine	318	326.9
144	2-Propenal	300	292.0
145	Crotonaldehyde	280	288.2
146	Butyraldehyde	230	206.9
147	2-Ethylcrotonaldehyde	250	272.9
148	Dihexyl ether	187	218.3
149	Dimethyl ether	350	396.9
150	Propylene oxide	455	437.7
151	Cyclohexanone	420	412.3
152	2-Pentanone	452	435.0
153	3-Pentanone	454	449.0
154	Acetone	465	501.2
155	2-Hexanone	423	416.5
156	1,1,1-Trichloroethane	486	478.1
157	Trichloroethylene	420	443.3

Table 4 Comparison of AIT values between the experimental values of 43 compounds used as the prediction set and the predicted AIT values obtained using 9 descriptors

No.	Structure name	AIT_{exp}/°C	AIT_{pre}/°C
Hydrocarbons			
158	Butane	372	282.3
159	Octane	220	234.5
160	2,3-Dimethyloctane	231	258.1
161	Hexane	240	264.0
162	Dodecane	203	179.2
163	3-Methylpentane	278	300.9
164	2,2-Dimethylpropane	456	429.8
165	1,5-Hexadiene	345	367.4
166	1-Heptene	263	292.4
167	2,3,3-Trimethyl-1-pentene	383	427.9
168	Ethylcyclopentane	262	303.7
169	*sec*-Butylcyclohexane	277	251.0
170	n-Propylcyclohexane	248	253.1
171	Bicyclohexyl	256	229.4
172	*trans*-1,2-Dimethylcyclohexane	304	304.1
173	Isopropylcyclohexane	283	282.1
174	Cyclohexene	310	336.4
175	Isopropylbenzene	424	436.3
176	1,2-Dimethylbenzene	464	483.3
177	1-Methyl-3-ethylbenzene	485	450.5
178	2-Butylbiphenyl	433	414.5
179	1,4-Dimethylbenzene	529	498.4
Compounds containing heteroatoms			
180	2,2-Dimethyl-1-propanol	420	431.3
181	2-Methyl-1-butanol	385	348.1
182	1-Butanol	340	327.0
183	1-Pentanol	300	307.1
184	Hexanoic acid	330	350.2
185	2-Ethylbutyric acid	390	392.8
186	Butyric acid	445	401.6
187	Isopentyl butyrate	335	374.1
188	Methyl benzoate	510	548.0
189	Isopropyl propionate	425	433.6
190	n-Decyl acetate	215	264.7
191	Dipropylamine	299	294.9
192	Diisopropanolamine	374	397.1
193	2-Diethylaminoethanol	320	320.7
194	Isobutyraldehyde	261	257.9
195	Propionaldehyde	227	223.5
196	Ethylene oxide	429	425.8
197	Methoxybenzene	475	499.3
198	2-Butanone	404	461.7
199	Acetophenone	560	572.9
200	1-Chlorobutane	250	262.6

unsaturation and aromaticity have enhanced radical stability and autoignition stability. For example, the AIT values of 1-hexene (AIT = 265 °C, density = 0.7837) and cyclohexene (AIT = 310 °C, density = 0.8595) are higher than those of hexane (AIT = 240 °C, density = 0.7694) and cyclohexane (AIT = 260 °C, density = 0.8256). Similarly, the AIT of naphthalene (AIT = 540 °C, density = 1.009) is higher than that of benzene (AIT = 498 °C, density = 0.9326). The straight-chain compounds have lower AITs than cyclohydrocarbons, and the AITs of cyclohydrocarbons are lower than those of aromatic compounds. Linear compounds have lower density than the cyclohydrocarbon compounds, and are also smaller than the aromatic compounds. Density is related to factors such as unsaturation, aromaticity and linearity of compounds. Accordingly, the value of density shows a positive coefficient in these equations.

The fractional total atomic charge weighted partial negative surface area, Jurs-FNSA-2,[16] is obtained as the ratio of total charge weighted partial negative charged surface area (PNSA-2) to total molecular surface area (TMSA). PNSA-2 is defined as the product of the summation of the negatively charged solvent-accessible atomic surface area and that of the individual atomic partial negative charge. Owing to a negative charge value, the value of Jurs-FNSA-2 is negative for the entire data set. Jurs-FNSA-2 is related to the relative distribution of negative charged surface area in the molecules. The absolute values of Jurs-FNSA-2 values for compounds containing aromatic groups, halogens, acids or esters are larger than those for the remaining molecules because these contain negatively charged carbons of aromatic group, halogen and oxygen.

In order to test the predictive power of the model, the whole data set was divided into hydrocarbons- and heteroatoms-containing compounds in both the training and prediction set, and the results of each subset are given in Table 7. The predicted AITs of the hydrocarbons in the prediction set were less successful (R^2 = 0.8890, MAE = 26.081 °C), but the results not including butane, improved to an R^2 of 0.9320 and mean absolute errors (MAE) of 23.052. These results demonstrate that the nine-descriptor model shows good correlation between the predicted and experimental AITs, whereas the MAE doesn't deviate by more than twenty five degrees.

Conclusion

Quantitative structure–property relationships were used to predict the autoignition temperatures for a diverse set of 200 compounds. With the genetic functional approximation, the best model developed was a nine-descriptor model containing I_{al}, I_{ke}, radius of gyration, $^1\chi^v$, SC-2, the Balaban index JX, Kappa-3-AM and Jurs-FNSA-2 descriptors.

Unlike previously reported models, this model can be applied to hypothetical compounds without dividing them into subsets of hydrocarbons and compounds containing heteroatoms, or other unverifiable structures such as acid halides or amides. Not only does this model provide accurate AIT value for unknown compounds, but also the descriptors in this model provide information about the structure's features and their interactions.

The factors determining the AIT values of molecules are chain length, addition of methyl groups, unsaturation, branching, aromaticity, conformation, stability of functional groups and polar interactions.

The autoignition mechanism proceeds by a free radical reaction, so more branches which can help to stabilize the molecules, result in an increase of the AIT. In addition, the AIT of ring compounds is higher than that of chains, and the AIT of the aromatic ring observed indeed is higher than the AIT of the aliphatic ring tested. Four topological descriptors ($^1\chi^v$, SC-2, the Balaban index JX and Kappa-3-AM) were found to

Table 5 Correlation matrix of the descriptors involved in the nine-descriptor model[a]

	1	2	3	4	5	6	7	8	9
1	1.000	0.056	−0.017	−0.083	0.078	0.030	0.013	0.055	0.016
2	0.056	1.000	−0.014	−0.034	0.079	−0.016	0.012	0.028	0.073
3	−0.017	−0.014	1.000	0.117	−0.730	−0.655	−0.142	−0.555	−0.231
4	−0.083	−0.034	0.117	1.000	−0.284	0.106	0.220	0.120	−0.235
5	0.078	0.079	−0.730	−0.284	1.000	0.048	0.037	0.457	0.658
6	0.030	−0.016	−0.655	0.106	0.048	1.000	0.243	0.407	−0.450
7	0.013	0.012	−0.142	0.220	0.037	0.243	1.000	0.711	0.111
8	0.055	0.028	−0.555	0.120	0.457	0.407	0.711	1.000	0.287
9	0.016	0.073	−0.231	−0.235	0.658	−0.450	0.111	0.287	1.000

[a] 1: I_{al}, 2: I_{ke}, 3: $^1\chi^v$, 4: JX, 5: SC-2, 6: RoG, 7: density, 8: Jurs-FNSA-2, 9: Kappa-3-AM.

Table 6 The QSPR nine-descriptor model for autoignition temperature of a 200 entire data set

	Coefficient	Std. error	*t*-Test	*F*-Test	Name of the descriptor
0	166.612	31.908	5.22		Intercept
1	−132.442	11.527	−11.49	132.018	I_{al}
2	47.741	10.496	4.55	20.690	I_{ke}
3	−140.212	6.080	−23.06	531.866	$^1\chi^v$
4	28.026	1.410	19.88	395.035	SC-2
5	83.044	9.010	9.22	84.959	RoG
6	194.170	30.087	6.45	41.648	Density
7	24.789	4.764	5.2	27.074	JX
8	67.861	20.363	3.33	11.106	Jurs-FNSA-2
9	−4.506	1.533	−2.94	8.637	Kappa-3-AM

$R^2 = 0.9189$, F = 239.32, significant level(p)<0.001, RMSE = 26.95, N = 200.

Table 7 Validation of correlation for the nine-descriptor model

Set	Group	N^a	$R^{2\,b}$	RMSE[c]	MAE(°C) [d]
Training set	Hydrocarbon compounds	70	0.9303	24.270	21.857
	Heteroatoms compounds	87	0.9082	25.225	20.516
Prediction set	Hydrocarbon compounds	22	0.8890 (0.9320)	31.399 (25.039)	26.081 (23.052)
	Heteroatoms compounds	21	0.9317	25.569	19.776
Whole set	Hydrocarbon compounds	92	0.9218 (0.9318)	25.835 (24.175)	22.867 (22.132)
	Heteroatoms compounds	108	0.9114	25.428	20.372

[a] *N*: number of compounds. [b] R^2: squared coefficient of correlation. [c] RMSE: root mean square error. [d] MAE: mean absolute error. Values in parentheses are statistic values of the prediction set without butane.

be of a significant utility in describing branching differences in the molecules. Moreover, density was shown to be related to unsaturation and aromaticity of the chemical structures, and the radius of gyration related to conformation. Jurs-FNSA-2 was useful in noting the polar interaction by distribution of negative charge. I_{al} and I_{ke} encoded ketones or aldehydes from the data set. The model used here employs only descriptors calculated from chemical structure, and this approach is applicable in principle to organic compounds. But to understand more clearly how the radical stability depends on the molecule structure, we need supplementary descriptors based molecular orbital theory which will be considered in further studies.

Acknowledgements

Y. S. Kim thanks research workers in the Computer Aided Molecular Design Research Center at Soong Sil University for their technical advice.

References

1 C. Morley, *Combust. Sci. Technol.*, 1987, **55**, 115.
2 L. J. Kirsch and C. P. Quiun, *J. Chim. Phys.*, 1985, **82**, 459.
3 R. T. Morison and R. N. Boyd, *Organic Chemistry*, Allyn and Bacon, Boston, 4th edn., 1983, ch. 3.
4 C. J. Hilado and S. W. Clark, *Chem. Eng. (N. Y.)*, 1972, **79**, 75.
5 D. E. Swarts and M. Orchin, *Ind. Eng. Chem.*, 1957, **49**, 432.
6 C. E. Frank and A. U. Blackham, *Ind. Eng. Chem.*, 1952, **44**, 862.
7 *Fire Protection Guide on Hazardous Materials*, National Fire Protection Association, Quincy, MA, 10th edn., 1991.
8 *Fire Protection Handbook*, Boston, USA, 14th edn., 1976.
9 T. Suzuki, K. Ohtaguchi and K. Koide, *J. Chem. Eng. Jpn.*, 1992, **25**, 606.
10 J. Tetteh, E. Metcalfe and S. L. Howells, *Chemom. Intell. Lab. Syst.*, 1996, **32**, 177.
11 J. Tetteh, S. L. Howells, E. Metcalfe and T. Suzuki, *Chemom. Intell. Lab. Syst.*, 1998, **41**, 17.
12 B. H. Michell and P. C. Jurs, *J. Chem. Inf. Comput. Sci.*, 1997, **37**, 538.
13 L. M. Egolf and P. C. Jurs, *Ind. Eng. Chem. Res.*, 1992, **37**, 1798.
14 K. B. Lipkowitz, D. B. Boyd, *Reviews in Computational Chemistry 9*, VCH Publishers, New York, 1996, p. 167.
15 H. Wiener, *J. Am. Chem. Soc.*, 1947, **69**, 17.
16 A. T. Balaban, *Chem. Phys. Lett.*, 1982, **89**, 399.
17 M. Randic, *J. Am. Chem. Soc*, 1975, **97**, 6609.
18 L. B. Kier, L. H. Hall, *Molecular Connectivity on Chemistry and Drug Research*, Academic Press, New York, 1976.
19 D. T. Stanton and P. C. Jurs, *Anal. Chem.*, 1990, **62**, 2323.
20 A. K. Ghose, V. N. Viswanadhan and J. J. Wendoloski, *J. Phys. Chem. A*, 1998, **102**, 3762.