主成分分析法による有機化合物の引火点と化学構造との相関[†]

鈴木孝弘・太田口和久・小出耕造

東京工業大学工学部 化学工学科^{††}

化学物質の引火点と化学構造との定量的な関係は従来ほとんど検討されていない。本研究では引火点に関係している構造因子の次元数を見いだすために、50種の化合物の引火点およびこれと何等かの関係がある他の10種の物性のデータセットについて主成分分析法による解析を行った。その結果、引火点を記述するためには2つの構造因子だけで十分であることがわかった。第1の因子は分子のバルクすなわち大きさや形と関係していると考えられた。そしてこの因子は分子結合性とよい相関があることがわかった。第2の因子は寄与が小さいが、分子中の官能基の特有な極性能と関係していると考えられ、藤田の無機性値と相関づけることができた。この解析結果に基づき、引火点を化学構造の情報のみから予測する推算式を提案した。さらに、本推算式の適用性を検討するために164種のさまざまなタイプの化合物についてそれらの引火点の予測を行いその有用性を確認した。

緒 言

化学物質の引火点は、化学防災上重要な基礎的物性値である。現在までに約1,000万の化合物が合成されてきており、新規化合物が出現する速度は年々増加している。また、産業界で使用される化合物の種類も急速にその範囲を広げている。しかし、引火点の実測は容易ではなく、実測値が整備されている化合物はきわめて少ない。したがって、実測値が得られない場合、既存化合物の使用および製造段階ばかりでなく、新規化合物の合成あるいは設計段階において、何らかの方法でその引火点を予測することが望まれている。

引火点の推算法は、すでに種々提案されている。原理的には、物質の飽和蒸気圧曲線と爆発下限界濃度（または引火下限界濃度）とから求められる^{8,12)}。また、炭化水素やアルコール等の同族列の化合物群については、引火点と標準沸点との間の相関関係を利用して引火点が予測できる^{1,2,14,16)}。しかし、これらの他の物性値から引火点を予測する方法は、適用範囲が構造の単純な同族列の化合物群に限られる。

一方、化学構造から引火点を予測する方法は、実測値を必要とせず便利である。Shimiyによって提案されている方法¹⁹⁾は、発火点とラジカルの水素原子数をパラメータとして引火点を予測するものであり、発火点はあらか

じめ分子中の炭素原子数と分岐上の炭素原子数とから推定される。しかし、この方法は中間パラメータとして発火点を求めなければならない、アルカン、オレフィン、ベンゼン誘導体、アルコールに適用範囲が限られる。

したがって、多様な構造の化合物に対して適用できる汎用性のある引火点の予測法ははまだ提案されていない。この問題に対する一つの解決策は、単に引火点と標準沸点等との単相関を議論するのではなく、引火点と何らかの関係のある他の特性値との関係を多変量の問題として捉えることであろう。化学データの規則性を見出すため提案されている種々の数学的手法の内、データセット中の線形関係を最もよく明確にするものの一つが主成分分析法である。この手法は、薬物の定量的構造活性相関(QSAR)の分野において、回帰式において有意になる独立変数の数等を決定する上で必要な洞察を提供し、その有用性が確認されている¹⁰⁾。

本研究では化学構造と引火点の関連性を検討するために主成分分析法を適用し、化学構造のみからその化合物の概略の引火点を予測する方法を提案することを目的とする。

1. 解析方法

1.1 主成分分析法の適用

主成分分析法(Principal Component Analysis)とは、ある問題に対して互いに相関のある多種類の特性値の持つ情報をそれら特性値の線形結合で表される合成変量で要約させる手法である¹⁰⁾。

[†] 1990年2月8日受理；化学工学会第55年会（名古屋，1990年4月）にて発表

^{††} 〒152 目黒区大岡山2-12-1

いま、 n 個の特性値 $V_1, V_2, \dots, V_n$ があると仮定すると、次の 2 つの条件を満足する m 個 ($\leq n$) の合成変数 $Z_1, Z_2, \dots, Z_m$ を求めることが主成分分析を行うことであり、 Z_1 を第 1 主成分、 Z_2 を第 2 主成分、 $\dots$ 、 Z_m を第 m 主成分とよぶ。

$$Z_1 = a_{11} V_1 + a_{12} V_2 + \dots + a_{1n} V_n \quad (1)$$

$$Z_2 = a_{21} V_1 + a_{22} V_2 + \dots + a_{2n} V_n \quad (2)$$

.....

$$Z_m = a_{m1} V_1 + a_{m2} V_2 + \dots + a_{mn} V_n \quad (3)$$

条件

- (1) 個々の主成分とそれ以外の他の全ての主成分との相関は 0 である。
- (2) Z_1 の分散は全ての特性値の線形結合の中で最大である。 Z_2 の分散は Z_1 以外の残りの全ての線形結合式の中で最大であり、以下、順次 $Z_3, \dots, Z_m$ が分散の大きさに基づき求められる。

本研究では、種々の化合物の引火点とこれに関連のある沸点等の諸物性値のデータセットについて主成分分析法を適用し、引火点を記述するために必要な構造因子の数を検討する。次にそれらの構造因子の意味を考察し、化学構造式から求められる構造記述子と引火点との相関関係を検討する。

1.2 解析に用いた化合物-物性データ

主成分分析に用いた 50 化合物とそれらの引火点等の 11 種の物性値を Appendix の Table A-1 に示した。化合物は分子量では 46.1 の Ethanol から 280.3 の Ethyl (o-(ethoxy-carbonyl) benzoxyloxyacetate) まで、極性では炭化水素から多官能性化合物まで多様な構造を有するものを選択した。物性としては、引火点 (T_f)、標準沸点 (T_b) の他、標準沸点との間に比例関係が認められる臨界温度 (T_c) と蒸発潜熱 (ΔH_v)、標準沸点の推算式¹³⁾でパラメータとして用いられている分子屈折 (MR) とパラコール (PA)、臨界温度の推算式⁶⁾で用いられるファンデルワールス式の定数 a と b 、その他分子の大きさを示す分子量 (MW) と分子容 (MV)、極性と関係した分子磁化率 (XM) の 11 種類を対象とした。以上の物性値は文献記載^{3,4)}の値を用いたが、実測値のない場合には推算値で代用した。

なお、引火点の値は測定装置に依存し、一般に密閉式引火点測定装置による値は、解放式によるものよりも 3 ~ 6°C 低い値を示す。また、測定誤差は同一の測定装置を用いた場合、1 ~ 3°C 程度である¹¹⁾。本研究では、密閉式引火点測定装置による引火点の実測値を用いた。

Table 1 Results of the principal component analysis

Component	Eigenvalue	Cumulative Proportion
1	8.93530	0.8123
2	1.37555	0.9373
3	0.24157	0.9593
4	0.16565	0.9744
5	0.12437	0.9857
6	0.08425	0.9933
7	0.03158	0.9962
8	0.02122	0.9981
9	0.01049	0.9991
10	0.00634	0.9997
11	0.00370	1.0000

2. 結果および考察

2.1 主成分分析の結果

Table A-1 のデータセットについて主成分分析を行った結果を Table 1 に示す。なお、各物性値の単位が異なっているため、平均を 0、分散を 1 とした標準化したデータについて分散共分散行列による主成分分析を行った。

累積寄与率は、第 1 主成分で 81.2%、第 2 主成分までで 93.7% に達し、第 1, 2 主成分の 2 変数のみでデータ集合の分散の 93.7% までを説明できることがわかった。第 3 主成分以下は寄与が小さいので以後の議論では無視する。この第 1 主成分 (Z_1) と第 2 主成分 (Z_2) を Table 2 にその関係を Fig. 1 に示す。これより Z_1 と Z_2 が直交しており、引火点は 2 つの独立変数のみによって記述できることがわかった。

T_f と Z_1 および Z_2 との関係は Eq. (4) で表され、 T_b のみを独立変数とした場合の相関 ($r=0.942$) よりも優れた相関式が得られた。

$$T_f = 13.721 Z_1 - 20.770 Z_2 + 38.91 \quad (4)$$

($n=50, r=0.969, s=12.41$)

Eq. (4) に基づく T_f の計算値と実測値との関係を Fig. 2 に示す。両者は良好に一致しており、目的としている化合物の Z_1, Z_2 値が得られれば、Eq. (4) に基づき T_f 値を予測することができる。しかし、 Z_1, Z_2 を直接に構造式と関係づけることは難しいため、化学構造から求められる構造記述子との相関を検討した。

2.2 主成分と構造記述子との相関

第 1, 第 2 主成分と用いた物性値との間の線形関係を仮定した場合の相関を Table 3 に示した。第 1 主成分は分子屈折、パラコール、ファンデルワールス式の定数 a 等とよい相関がある。また、 Z_1 の値は、Table 2 よりデータ集合中で分子量が最大の Ethyl(o-(ethoxycarbonyl) benzoxyloxyacetate) が 11.373、中程度の Cyclohexanol が 0.679、2-Pentanone が -1.077、Pentanoic acid が

Table 2 First two principal components, molecular connectivities and inorganicity values for training set compounds

Compound	Z_1	Z_2	${}^1\chi$	ΣI
1	7.171	-1.185	6.933	105
2	0.500	0.731	3.932	15
3	0.684	0.636	3.788	15
4	2.438	-0.963	4.966	60
5	-0.136	1.493	3.394	10
6	-3.521	-1.001	1.414	100
7	3.429	0.885	4.270	100
8	-2.461	-0.259	1.732	100
9	1.528	-1.536	3.932	115
10	0.679	-0.703	3.394	110
11	2.566	-1.345	4.326	115
12	-0.024	-1.745	3.394	115
13	-2.730	1.816	2.414	20
14	2.628	2.356	4.414	20
15	-5.268	1.256	1.414	20
16	-3.789	0.387	1.524	65
17	-2.619	0.616	2.414	65
18	-2.315	0.464	2.270	65
19	-1.131	-0.599	2.894	75
20	-1.077	0.607	2.770	65
21	2.251	-2.286	4.304	165
22	-0.678	-1.011	2.643	150
23	-3.170	-1.054	1.732	150
24	0.780	-1.136	3.270	150
25	2.448	-0.376	4.843	75
26	0.451	1.281	3.626	60
27	2.923	1.556	4.664	60
28	-0.649	0.836	3.308	60
29	-1.925	1.019	2.270	70
30	2.187	0.087	4.304	85
31	1.585	-1.156	3.804	85
32	2.924	2.015	4.414	70
33	2.700	-1.720	4.966	85
34	-2.986	-0.472	1.914	70
35	-1.870	-0.484	2.414	70
36	1.075	-1.603	4.304	85
37	-3.079	-1.380	1.732	70
38	-1.917	0.143	2.500	35
39	-3.656	0.484	1.414	20
40	-2.867	1.363	1.732	10
41	-1.393	1.055	2.414	10
42	-0.063	-0.057	3.394	25
43	-1.917	0.049	1.914	20
44	0.108	-1.898	2.770	200
45	1.824	-0.184	4.164	125
46	-3.126	0.229	1.732	75
47	-1.736	-0.207	3.000	40
48	3.130	-0.113	5.464	35
49	-1.275	0.947	2.914	40
50	11.373	2.165	9.596	195

0.780, 最小の Dimethyl ether が -5.268 であった。これより, 第 1 主成分は, 分子のバルクの特徴を示す大きさおよび分子の形状と関係していると考えられる。

分子のバルクの特徴を記述できる構造記述子は種々あ

Fig. 1 Plot of the first principal component against the second principal component

Fig. 2 Comparison of measured and recalculated T_f values for 50 compounds listed in Table A-1

るが²⁰⁾, ここでは分子結合性 (Molecular Connectivity Index)⁹⁾ との相関を検討した。

分子結合性として種々のパラメータが定義されているが, 最も簡単な 0 次のインデックスは Eq. (5) で与えられる。

$${}^0\chi = \sum \delta_i^{-1/2} \quad (5)$$

$${}^1\chi = \sum (\delta_i \delta_j)^{-1/2} \quad (6)$$

ここに, δ_i は i 番目の原子から出ている結合の数から水素原子との結合の数を引いた数であり, Σ は分子を構成するすべての原子についての総和を示す。Eq. (6) は, 1 次のインデックスであり, 互いに隣接している原子の δ_i と δ_j について積をとり, 各結合についての総和を示している。

Table 3 Correlation coefficients between principal components and original properties

Property	Z_1	Z_2
T_f	0.833	-0.495
T_b	0.882	-0.436
T_c	0.794	-0.557
MW	0.907	0.107
MR	0.967	0.143
MV	0.886	0.421
ΔH_v	0.856	-0.381
PA	0.950	0.285
XM	0.948	0.240
a	0.960	0.158
b	0.914	0.335

同様に、2次、3次、...、 n 次のインデックスが定義されるが、ここでは0次と1次のインデックスについて検討した。 δ_i については、ヘテロ原子の種類に依存した値⁹⁾ (価電子の数からその原子に結合している水素原子の数を差し引いた数値として定義され、たとえばエーテル酸素原子では6となる)をEqs. (5), (6)に代入した場合、 ${}^0\chi^v$, ${}^1\chi^v$ と表記して区別する。

Z_1 と ${}^0\chi$, ${}^1\chi$, ${}^0\chi^v$, ${}^1\chi^v$ および ${}^0\chi$ と ${}^1\chi$ については関数型を種々変えた場合(${}^0\chi$ の1/2乗, 1/3乗の ${}^0\chi^{1/2}$, ${}^0\chi^{1/3}$ 等)との相関関係を比較検討した結果をTable 4に示す。いずれの場合についてもきわめてよい相関が確認できたが、 ${}^1\chi$ を使用した場合が最も優れていた。 Z_1 と ${}^1\chi$ との関係をFig. 3およびEq. (7)に示す。

$$Z_1 = 1.915 {}^1\chi - 6.399 \quad (7)$$

$$(n=50, r=0.978, s=0.633)$$

なお、データセットの化合物に対する ${}^1\chi$ の計算値はTable 2中に掲げてある。

第2主成分については、Table 3に示したように用いた物性値と統計的に有意な相関がなく、また寄与が小さいため解釈が難しいが、 $-\text{COOH}$, $-\text{OH}$ 等の特定の原子団に対する依存性が認められた。 Z_1 が分子中のヘテロ原子の寄与を全く考慮していない ${}^1\chi$ と相関づけられた観点からも、原子団の凝集能やある種の極性能と関係があることが予想される。

そこで分子構造から求められる凝集エネルギー、分子引力係数、藤田の無機性値⁷⁾を用いた相関式を検討した。No. 50の化合物は偏差が何れの場合にも大きかったため解析から外し、49種の化合物について Z_2 と最も相関がよかったものは藤田の無機性値であった($r=0.710$)。

Z_2 には Z_1 と比べてかなり不確定な誤差等が入っている可能性を考慮してデータセット中の化合物から Z_2 との相関性の悪いものを順次除いていき、藤田の無機性値 ΣI との相関をさらに検討した結果、Fig. 4に示すように

Table 4 Correlation for various single-parameter linear regressions based on variations of the connectivity index versus Z_1

descriptor	r	s
${}^0\chi$	0.963	0.819
${}^0\chi^{1/2}$	0.967	0.764
${}^0\chi^{1/3}$	0.965	0.796
${}^0\chi^{1/4}$	0.963	0.819
${}^0\chi^{1/5}$	0.961	0.835
${}^0\chi^{2/3}$	0.968	0.756
${}^0\chi^2$	0.892	1.367
${}^1\chi$	0.978	0.633
${}^1\chi^{1/2}$	0.972	0.709
${}^1\chi^{1/3}$	0.965	0.790
${}^1\chi^{1/4}$	0.960	0.836
${}^1\chi^{1/5}$	0.958	0.865
${}^1\chi^{2/3}$	0.976	0.649
${}^1\chi^2$	0.916	1.210
${}^0\chi^v$	0.953	0.917
${}^1\chi^v$	0.937	1.056

Fig. 3 Relationship between Z_1 and ${}^1\chi$

30種の化合物について統計的に意味のある相関を確認できた。その相関式をEq. (8)に示す。

$$Z_2 = -0.0171 \Sigma I + 1.17483 \quad (8)$$

$$(n=30, r=0.898, s=0.438)$$

ここに、 ΣI は化合物中の構成原子団についてTable A-2の I 値の総和をとることによって求められる。データセットの化合物に対する ΣI の計算値はTable 2に示した。

なお、分子の幾何学的形状および極性度の双方に関する分子の複雑さを表すパラメータとして偏心因子¹⁵⁾が広く用いられているが、 Z_1 , Z_2 との相関は相関係数がそれぞれ0.511, 0.094と統計的に意味のある関係は見い出されなかった。

2.3 構造記述子からの T_f の予測

前節の解析によって Z_1 , Z_2 をそれぞれ ${}^1\chi$, ΣI で置き

Fig. 4 Relationship between Z_2 and ΣI . The number indicates the compounds in Table A-1

換えられることがわかった。そこで、データセットの化合物の T_f について $^1\chi$ および ΣI を独立変数として重回帰分析を行い次式を得た。

$$T_f = 19.938 ^1\chi + 0.5398 \Sigma I - 68.104 \quad (9)$$

$$\text{相関範囲: } -45.0^\circ\text{C} \leq T_f \leq 185.0^\circ\text{C}$$

$$(n=50, r=0.928, s=18.69)$$

この式は Eq. (4) よりも相関性がやや低く、 T_f に対する平均絶対誤差は 15.0°C であるが、化学構造のみから引火点が予測できる。

2.4 引火点推算式の適用性の検討

データセットに含まれない多様な 164 化合物をテストセットとして Eq. (9) に基づく予測値と実測値^{3,4)}との比較を行った結果を Fig. 5 に示す。 r は 0.949、平均絶対誤差は、 14.5°C であった。

Table A-3 にテストセットとして用いた化合物の一覧を示す。従来、引火点の予測が難しかったハロゲンを含む化合物や多官能性化合物についても炭化水素や単官能性化合物と同程度の精度で予測できることがわかった。さらに、推算精度を高めるためには Z_2 の意味をより明らかにしていく必要があるが、Eq. (9) は広範囲の化合物を対象としている点を考慮すれば、概略の引火点を予測するために有用であると考えられる。

Fig. 5 Comparison of measured and predicted T_f values for 164 compounds listed in Table A -3

結 言

化学物質の引火点と化学構造との関係を見出すために、50種類の多様な構造の有機化合物の引火点、標準沸点、パラコール等 11 種類の特性値のデータセットについて主成分分析を行った。その結果、引火点は近似的に 2 つの構造因子の線形結合によって記述できることがわかった。第 1 の因子は、分子のトポロジカルな特性を表す分子結合性、第 2 の因子は、分子の極性と関係した藤田の無機性値とそれぞれ関係づけられた。

さらに、化学構造から求められる分子結合性と無機性値に基づき引火点を予測する推算式を提案し、その有用性を確認した。

Appendix

Table A-1 Data table on which principal component analysis were performed

Compound	T_f	T_b	T_c	MW	MR^a	MV	ΔH_v	PA	XM	a	b
1. Anthracene	121.0	340.0	883.0	178.2	56.98 ^c	142.58	13.50	418.2 ^f	130.00	65.63 ^h	0.27 ^h
2. Ethylbenzene	15.0	136.2	617.1	106.2	35.78	122.46	8.50	290.1	77.20	28.60 ^h	0.17 ^h
3. <i>m</i> -Xylene	27.0	139.1	617.0	106.2	35.80	122.32	8.69	289.2	76.56	30.36	0.18
4. Naphthalene	79.0	218.0	748.4	128.2	36.82	110.30	10.34	282.2	91.90	39.74	0.19
5. Methylcyclohexane	-4.0	101.1	572.1	98.2	32.50	127.62	7.44	288.5	78.91	27.02 ^h	0.17 ^h
6. Ethanol	13.0	78.3	516.2	46.1	12.93	58.36	9.26	129.2	33.60	12.02	0.08
7. 2-Octanol	87.8	180.0	637.0	130.2	40.17	158.68	10.60	364.9	102.25 ^g	44.42 ^h	0.25 ^h
8. 2-Propanol	12.0	82.4	508.3	60.1	17.60	76.51	9.51	167.4	45.79	16.69 ^h	0.13 ^h
9. Benzyl alcohol	93.0	205.4	677.0	108.1	32.43	103.84	12.07	258.2	71.83	32.02 ^h	0.16 ^h
10. Cyclohexanol	68.0	161.1	625.0	100.2	29.29	106.37	10.88	259.3	73.40	28.61 ^h	0.16 ^h
11. <i>p</i> -Ethylphenol	103.9	219.0	716.4	122.2	36.97	120.84	12.10	300.0 ^f	84.15 ^g	34.48 ^h	0.18 ^h
12. Phenol	79.0	181.8	694.2	94.1	28.00	88.98	9.73	228.6	60.21	21.73 ^h	0.12 ^h
13. Diethyl ether	-45.0	34.5	466.7	74.1	22.51	103.90	6.38	216.7	55.10	17.38	0.13
14. Dibutyl ether	25.0	142.0	580.0	130.2	40.98	169.36	8.83	377.9	102.10 ^g	41.39 ^h	0.26 ^h
15. Dimethyl ether	-41.0	-24.9	400.0	46.1	13.08 ^c	69.70	5.14	137.1	26.30	8.07	0.07
16. Propionaldehyde	-30.0	47.9	496.0	58.1	16.06	71.96	6.77	159.0 ^f	34.32	15.10 ^h	0.09 ^h
17. Butyraldehyde	-6.7	74.8	524.0	72.1	20.79	89.96	6.34	204.4	46.08	19.14 ^h	0.11 ^h
18. Ethyl methyl ketone	-9.0	79.5	535.5	72.1	20.68 ^c	89.75 ^d	7.46	199.0 ^f	46.85 ^g	19.64 ^h	0.13 ^h
19. Cyclopentanone	26.0	130.7	626.0	84.1	23.15	88.46	8.69	216.0	51.63	22.61 ^h	0.13 ^h
20. 2-Pentanone	7.2	101.9	564.0	86.1	25.23	106.40	7.98	237.7	58.20 ^g	24.44 ^h	0.16 ^h
21. Benzoic acid	121.0	250.0	752.0	122.1	32.46 ^c	113.08	12.10	269.4 ^f	70.28	32.56 ^h	0.15 ^h
22. Isobutyric acid	55.6	154.5	609.0	88.1	22.11	92.75	9.83	211.2	56.06	22.83 ^h	0.13 ^h
23. Acetic acid	39.0	117.8	594.4	60.7	13.14	57.82	5.66	134.8	31.54	17.59	0.11
24. Pentanoic acid	88.0	185.5	651.0	102.1	26.83	108.76	10.53	252.2	67.55 ^g	28.33 ^h	0.16 ^h
25. Methyl benzoate	83.0	199.5	692.0	136.2	37.89	124.53	10.30	313.3	81.59	38.15 ^h	0.21 ^h
26. Isobutyl acetate	18.0	116.0	561.0	116.2	31.50	132.83	8.57	298.8	78.52	28.50	0.18
27. Isobutyl butyrate	50.0	157.2	611.0	144.2	40.68 ^c	148.78 ^d	9.61 ^e	374.8 ^f	100.55 ^g	43.05 ^h	0.25 ^h
28. Methyl butyrate	14.0	102.5	554.4	102.1	26.83	113.73	8.08	260.4	66.40	23.94	0.16
29. Isobutylamine	-9.0	68.0	516.0	73.1	24.33	101.02	7.32	224.7	59.80	20.37 ^h	0.14 ^h
30. <i>N,N</i> -Dimethylaniline	63.0	193.5	687.0	121.2	40.89	126.77	9.17 ^e	315.0	89.66	37.49	0.20
31. <i>o</i> -Toluidine	85.0	199.7	694.0	107.2	35.34	107.33	10.84	274.6	76.00	32.72 ^h	0.17 ^h
32. Dibutylamine	33.0	159.6	596.0	129.8	42.99	170.70	9.50	387.3	103.70	40.17 ^h	0.24 ^h
33. Quinoline	99.0	237.1	794.4	129.2	41.82	117.95	10.99 ^e	305.9	86.00	34.69 ^h	0.14 ^h
34. Propionitrile	2.0	97.4	564.4	55.1	15.77	70.45	7.35	164.4	38.50	16.44	0.11
35. 3-Butenenitrile	21.0	119.0	585.0	67.1	19.75	80.43	8.19	190.4	37.95 ^g	23.58 ^h	0.15 ^h
36. Nitrobenzene	88.0	211.0	732.0	123.1	32.77	102.17	9.74	266.6	61.80	21.72 ⁱ	0.15 ⁱ
37. Nitromethane	35.0	101.3	588.0	75.1	15.34	66.30	8.12	167.5	21.10	6.35 ⁱ	0.08 ⁱ
38. Thiophene	-1.1	84.0	579.4	84.1	24.41	79.58	7.52	192.2	57.38	20.72	0.13
39. Ethanethiol	-17.0	35.0	498.6	62.1	18.91	74.72	6.40	167.2	43.85 ^g	11.24	0.08
40. Isopropyl chloride	-32.0	34.8	485.0	78.5	21.13	91.72	6.34	197.2	56.50 ^g	16.65 ^h	0.12 ^h
41. Butyl chloride	-9.0	78.5	542.0	92.6	25.43	104.43	7.38	235.8	67.10	22.01 ^h	0.15 ^h
42. Chlorobenzene	29.0	132.0	632.4	112.6	31.17	101.74	8.73	249.2	69.97	25.43	0.15
43. 1, 2-Dichloroethane	13.0	83.4	561.0	99.0	21.01	78.97	7.65	192.7	60.00 ^g	19.52 ^h	0.12 ^h
44. <i>N</i> -Ethylacetamide	110.0	205.0	702.0 ^b	87.1	24.29 ^c	97.05 ^d	8.73 ^e	231.9 ^f	54.35 ^g	22.28 ^h	0.14 ^h
45. Ethyl acetate	57.0	180.4	673.0	130.2	31.93	125.53	10.60 ^e	304.1	71.67	36.56 ^h	0.21 ^h
46. Acetyl chloride	4.0	50.9	508.0	78.5	16.80	71.11	6.85	161.6	38.90	18.53 ^h	0.11 ^h
47. 1, 4-Dioxane	12.0	101.4	587.0	88.1	21.69	85.29	8.16	209.3	52.16	19.19 ^h	0.12 ^h
48. Bromopentafluorobenzene	87.0	137.0	670.0	247.0	34.02	126.85	12.29 ^e	309.5 ^f	102.50 ^g	35.04 ^h	0.17 ^h
49. 1, 2-Dimethoxyethane	1.0	85.2	536.0	90.1	24.17	104.44	7.75	274.9	61.70 ^g	20.49 ^h	0.15 ^h
50. Ethyl (<i>o</i> -(ethoxycarbonyl) benzoyloxyacetate)	185.0	312.0	740.0 ^b	280.3	67.86 ^c	230.95 ^d	13.88 ^e	597.4 ^f	155.85 ^g	94.03 ^h	0.44 ^h

a) Computed using the Lorentz-Lorenz equation, b) Estimated based on Lydersen's method [Ref. 17],
c) Estimated based on Eisenlohr's method [Ref. 13], d) Estimated based on Exner's method [Ref. 5],
e) Estimated based on Sastri's method [Ref. 18], f) Estimated based on Sugden's method [Ref. 13],
g) Estimated based on Van Krevelen's value [Ref. 21], h) Estimated based on Forman and Thodos's method [Ref. 6], i) Estimated based on Kier and Hall's equation [Ref. 9].

Table A-2 Fujita's inorganicity value⁷⁾

Structure	I
-SO ₂ -NH-CO-	260
-N=N-NH ₂	260
=N-OH	250
-SO ₃ H	250
-NH-SO ₂ -NH-	250
-CO-NH-CO-NH-CO-	250
-S-OH	240
-CO-NH-CO-NH-	240
-SO ₂ -NH-	240
-CS-NH-	230
-CO-NH-CO-	230
=N-OH	220
NH-CO-NH-	220
=N-NH-	210
-CO-NH-NH ₂	210
-CO-NH-	200
-COOH	150
-SO ₂ -	170
-SO-	140
Lactone ring	120
-CO-O-CO-	110
Anthracene ring	105
Phenanthrene ring	105
-OH	100
-NH-NH-	80
-O-CO-O-	80
-CSSH	80
-SCN	80
-CSOH	80
-COSH	80
-NCS	75
-NO ₂	70
-CN	70
-N<, -NH ₂ , -NH-	70
>CO, -CHO	65
-COO (R)	60
Naphthalene ring	60
>C=NH	50
-CSSR	50
-CSOS	50
-COSR	50
-NO	50
-O-NO ₂	40
-NC	40
-NCO	30
-N=N-	30
-O-	20
-O-NO	20
-SH, -S-	20
Benzene ring	15
Aliphatic monocyclic ring	10
-I	10
-Br	10
-Cl	10
-F	5
Triple Bond	3
Double Bond	2

Table A-3 Compounds in the testing set for predicting flash points based on Eq. (9)

Compound	$T_{f, \text{obsd.}}^{3, 4)}$	$T_{f, \text{estd.}}$
Hydrocarbons		
1. Ethylcyclohexane	18	16
2. Octane	13	10
3. Cumene	36	26
4. 2, 3-Dimethylbutane	-29	-15
5. <i>p</i> -Cymen	47	34
6. Styrene	32	20
7. Benzene	-11	-0.2
Oxygen containing compounds		
8. Allyl alcohol	21	25
9. Isobutyl alcohol	28	31
10. 2-Ethyl-1-hexanol	73	73
11. 2-Phenylethanol	102	82
12. 1-Pentanol	33	44
13. 2-Methyl-1-pentanol	54	52
14. 2-Butanol	26	31
15. 1-Decanol	82	94
16. 1-Dodecanol	127	114
17. <i>t</i> -Butyl alcohol	11	26
18. <i>p</i> -Cresol	86	70
19. 2-Naphthol	153	125
20. <i>p</i> -Nonyl phenol	140.6	150
21. Biphenyl-2-ol	123.9	129
22. <i>o</i> - <i>t</i> -Pentyl phenol	103.9	105
23. <i>o</i> -Cyclohexyl phenol	133.9	127
24. Isopropyl vinyl ether	-32	-1
25. Ethyl methyl ether	-37.2	-19
26. Diphenyl ether	115	88
27. Dibenzyl ether	135	107
28. Tetrahydrofuran	-20	-9
29. Phenetole	62.8	39
30. Butyl ethyl ether	4.4	11
31. Methyl propyl ether	-20.0	-9
32. Dipropyl ether	21.1	11
33. Diethyl ether	77	71
34. Methoxy benzene	51	29
35. Acrylaldehyde	-18	6
36. Isobutylaldehyde	-18	12
37. Octanal	52	55
38. Crotonaldehyde	13	16
39. Benzaldehyde	63	54
40. 2-Ethylhexanal	44	54
41. Acetophenone	77	61
42. Acetone	-20	2
43. Cyclohexanone	44	23
44. 2, 6-Dimethyl-1, 4-heptanone	49	57
45. 2-Tridecanone	107.2	102
46. 2-Nonanone	60	62
47. 3-Buten-2-one	-7	13
48. 2-Hexanone	25	32
49. 2-Heptanone	39	42
50. 4-Heptanone	49	43
51. 4-Methyl-2-pentanone	18	29
52. 2-Octanone	71.1	52
53. 2-Heptadecanone	120	142
54. Oleic acid	189	209
55. Formic acid	68.9	41
56. 2-Methylvaleric acid	107.2	89

Table A-3 (Continued)

Compound	$T_{f, \text{obsd.}}^{3,4)}$	$T_{f, \text{estd.}}$	Compound	$T_{f, \text{obsd.}}^{3,4)}$	$T_{f, \text{estd.}}$
57. Butyric acid	72	68	113. Butyl bromide	18	-15
58. Propionic acid	52	58	114. Trichloroethylene	32	-6
59. Stearic acid	196	208	115. Fluorobenzene	-12	10
60. Acrylic acid	54	59	116. 2-Chlorobutane	-15	-17
61. Isopropyl benzoate	98.9	86	117. Allyl bromide	-1	-23
62. Ethyl benzoate	88	79	Multifunctional compounds		
63. Benzyl benzoate	148	137	118. Adipic acid	196	186
64. Butyl isovalerate	53	67	119. Acetylacetone	34	64
65. Isopropyl formate	-6	20	120. Ethyl acetoxyacetate	82	90
66. Butyl formate	18	32	121. Acetone cyanohydrin	74	64
67. Methyl formate	-19	2	122. 1-Amino-2-propanol	77	69
68. Ethyl crotonate	2.2	30	123. Anthraquinone	185	195
69. Ethyl acetate	-4	20	124. Dibutyl isophthalate	161	175
70. Phenyl acetate	80	68	125. Ethylene glycol	111	78
71. Butyl acetate	37	39	126. 2-Ethoxyethanol	43	55
72. Propyl acetate	13	30	127. 3-Ethoxypropionaldehyde	38	46
73. Benzyl acetate	90	78	128. 3-Ethoxypropionic acid	107.2	99
74. Methyl acetate	-10	-2	129. Octanoyl chloride	82	68
75. Methyl stearate	152.8	170	130. Benzoyl chloride	72	66
76. Pentyl laurate	148.9	150	131. 2-Chloroethanol	60	29
77. Ethyl acrylate	15	31	132. Ethyl chloroformate	16	25
Nitrogen containing compounds			133. 2-Chloro-1-propanol	51.7	37
78. Aniline	70	45	134. 2-Chloropropionic acid	107.2	71
79. Octylamine	60	58	135. <i>o</i> -Chlorobenzylidene trifluoride	59	54
80. <i>N,N</i> -Diethylaniline	85	85	136. Salicylaldehyde	77.8	116
81. <i>N,N</i> -Dibutylaniline	110	125	137. Methyl salicylate	96	131
82. Decylamine	99	78	138. 1, 3-diamino-2-propanol	132.2	117
83. Piperidine	16	35	139. Diethylene glycol	124	119
84. Pyrrol	39	27	140. 1, 2-Diethoxy ethane	35	32
85. <i>s</i> -Butylamine	-9	20	141. Dichloroacetyl chloride	66.1	36
86. 2, 3-Xylidine	97	62	142. 2, 4-Dichlorophenol	113.9	89
87. Aziridine	-11.1	5	143. <i>N,N</i> -Dimethyl formamide	58	50
88. Dipropylamine	17	38	144. 1, 2-Dimethoxy ethane	-2	11
89. Pyridine	20	38	145. Ethyl nitrate	10	9
90. Acetonitrile	5	-2	146. Propyl nitrate	20	19
91. Acrylonitrile	0	9	147. Butyl nitrate	36.1	29
92. Benzonitrile	71	56	148. Diethyl carbonate	25	51
93. Butyronitrile	16	18	149. 1, 4-Thioxane	42.2	19
94. Allyl isothiocyanate	46	32	150. Triethylenetetramine	135	181
95. Butyl isothiocyanate	65.6	41	151. Ethylene diacetate	88	89
96. Nitroethane	28	15	152. Phenylhydrazine	88	94
97. <i>o</i> -Nitrotoluene	85	72	153. 2-Phenoxyethanol	121.1	103
98. 2-Nitropropane	24	22	154. <i>N</i> -(2-Phenoxyethyl) aniline	170	145
99. 2-Nitrobiphenyl	143	131	155. Diphenyl phthalate	223.9	258
100. 1-Nitronaphthalene	164	127	156. Dimethyl phthalate	146	138
101. Acetoanilide	173	143	157. 1, 3-Butanediol	121	95
102. <i>N</i> -Butylacetamide	116	115	158. 2-Butoxyethanol	62	75
Sulfur containing compounds			159. 2-Propanediol	99	85
103. 1-Butanethiol	1.7	-9	160. Ethyl bromoacetate	47.8	36
104. Diethyl sulfide	-9	-9	161. 2, 5-Hexanedione	79	74
105. Dimethyl sulfide	-36	-29	162. 2-(Methylamino) ethanol	73.9	72
Halogenated hydrocarbons			163. 4-Methylmorpholine	23.9	53
106. Allyl chloride	-32	-25	164. 2-Methoxyethanol	39	45
107. Ethyl chloride	-43	-35			
108. Chlorocyclohexane	32	10			
109. <i>o</i> -Chlorotoluene	47	21			
110. Chloroprene	-20	-15			
111. 1, 1-Dichloroethane	-6	-23			
112. <i>o</i> -Dichlorobenzene	65	27			

Nomenclature

a	= van der Waal's constant	$[l^2 \cdot \text{atom} \cdot \text{mol}^{-2}]$
b	= van der Waal's constant	$[l \cdot \text{mol}^{-1}]$
ΔH_v	= enthalpy of vaporization at normal boiling point	$[\text{kcal} \cdot \text{mol}^{-1}]$
I	= inorganicity	$[-]$
MR	= molar refraction	$[\text{cm}^3 \cdot \text{mol}^{-1}]$
MV	= liquid-state molar volume	$[\text{cm}^3 \cdot \text{mol}^{-1}]$
MW	= molecular weight	$[\text{g} \cdot \text{mol}^{-1}]$
n	= number of data points in set	$[-]$
PA	= parachor	$[(\text{cm}^3 \cdot \text{mol}^{-1})(\text{dyn} \cdot \text{cm}^{-1})^{1/4}]$
r	= multiple correlation coefficient	$[-]$
s	= standard deviation	$[^\circ\text{C} \text{ or } -]$
T_b	= normal boiling point	$[^\circ\text{C}]$
T_c	= critical temperature	$[\text{K}]$
T_f	= flash point	$[^\circ\text{C}]$
XM	= molar magnetic susceptibility	$[10^{-6} \text{cm}^3 \cdot \text{mol}^{-1}]$
Z_i	= principal component	$[-]$
δ	= number assigned to each atom	$[-]$
$^0\chi$	= the zero-th order molecular connectivity index	$[-]$
$^1\chi$	= the first-order molecular connectivity index	$[-]$
$^0\chi^v$	= the zero-th order valence molecular connectivity index	$[-]$
$^1\chi^v$	= the first-order valence molecular connectivity index	$[-]$

Literature cited

- 1) Affens, W.A.: *J. Chem. Eng. Data*, **11**, 197 (1966)
- 2) Butler, R.M., G.M. Cooke, G.G. Lukk and B.G. Jameson: *Ind. Eng. Chem.*, **48**, 808 (1956)
- 3) Dean, J.A. (ed.): "Lange's Handbook of Chemistry", 13th Ed, McGraw-Hill, New York (1985)
- 4) Dean, J.A.: "Handbook of Organic Chemistry", McGraw-Hill, New York (1987)
- 5) Exner, O.: *Collect. Czec. Chem. Comm.*, **32**, 1 (1967)
- 6) Forman, J.C. and G. Thodos: *AIChE J.*, **6**, 206 (1960)
- 7) Fujita, A.: *Pharm. Bull. (Tokyo)*, **2**, 163 (1954)
- 8) Gooding, C.H.: *Chem. Eng.*, Dec. 12, 88 (1983)
- 9) Kier, L.B. and L.W. Hall: "Molecular Connectivity in Structure-Activity Analysis", Research Studies Press (1986)
- 10) Kowalski, B.R.: *Anal. Chem.*, **52**, 112R (1980)
- 11) Lance, R.C., A.J. Barnard and J.E. Hooyman: *J. Hazard. Mater.*, **3**, 107 (1979)
- 12) Lyman, W.J., W.F. Reehl and D.H. Rosenblatt: "Handbook of Chemical Property Estimation Methods", Chap. 18, McGraw-Hill (1982)
- 13) Meissner, H.P.: *Chem. Eng. Prog.*, **45**, 149 (1949)
- 14) Patil, G.S.: *Fire & Materials*, **12**, 127 (1988)
- 15) Pitzer, K.S.: *J. Am. Chem. Soc.*, **77**, 3427 (1955)
- 16) Prugh, R.W.: *J. Chem. Educ.*, **50**, A85 (1973)
- 17) Reid, R.C., J.M. Prausnitz and T.K. Sherwood: "The Properties of Gases and Liquids", 3rd ed., McGraw-Hill, New York (1977)
- 18) Sastri, S.R.S., M.V. Ramana Rao, K.A. Reddy and L.K. Doraiswamy: *Br. Chem. Eng.*, **14**, 959 (1969)
- 19) Shimy, A.A.: *Fire Technol.*, **6**, 135 (1970)
- 20) Suzuki, T., K. Takeuchi and M. Ishida: *Kagaku Kougaku*, **53**, 823 (1989)
- 21) Van Krevelen, D.W. and P.J. Hoftyzer: "Properties of Polymers. Their Estimation and Correlation with Chemical Structure", Elsevier, Amsterdam (1976)

Correlation Between Flash Points and Chemical Structures of Organic Compounds Using Principal Component Analysis

Takahiro Suzuki, Kazuhisa Ohtaguchi and Kozo Koide

Dept. of Chem. Eng., Tokyo Inst. of Tech., Tokyo 152

Key Words : Flash Point, Principal Component Analysis, Molecular Connectivity Index

The quantitative relationship between chemical structure and the flash point is not well known. In this paper, a principal component analysis of a set of data including flash points and ten other physicochemical properties of 50 organic compounds is performed to determine the number of significant structural factors affecting the flash point. Only two structural factors were found to be sufficient to reproduce the flash point data. The first factor is identified as the bulk or size parameter of a molecule and is related to its connectivity index. The second factor, while making a relatively small contribution to the flash point, is related to the specific polar characteristics of functional groups in a molecule. A significant correlation is found between the second factor and Fujita's inorganicity value. Analyses of this work lead to a predictive equation for the flash point using only the chemical structure of a molecule of interest. The robustness and utility of this method were examined by predicting the flash points of 164 compounds not included in the original data set.