

Molecular Connectivity in the Determination of Physical Properties

R.P. Semwal* & J.P. Balodi
Department of Chemistry, SGRR(PG) College,
Dehradun 248 001 INDIA

ABSTRACT

We have used molecular connectivity to develop polynomials of the form $y = a_0 + a_1x + a_2x^2 + \dots + a_nx^n$ for predicting b.p., m.p., density and r.i. of normal hydrocarbons. These properties of the systems determined by the polynomials are very close to their experimental values. For the correlations corresponding to each physical property the statistical parameters, r , s , F and ESS have been determined. Except for m.p. of small molecules, in most cases the error is less than one per cent.

1. Introduction

The quest for a suitable method for the determination of several physical, chemical and biological properties of molecules that takes the structural features into consideration has led to the concept of molecular connectivity (MC)¹. It has been extensively studied²⁻⁴ and applied in many areas⁵⁻⁷, nevertheless, in this paper we have developed polynomial equations using MC indices to predict boiling points (b.p.), melting points (m.p.), densities and refractive indices (r.i.) of normal hydrocarbons. It is gratifying to note that the properties of the systems determined by these polynomials are very close to their experimental values.

In this section we introduce MC indices and their selection in developing polynomials. The next section describes the polynomials, followed by the results and discussion. The simplest MC index ${}^m\chi$ is the first-order ($m=1$) one

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

where the sum is over all connections or edges in the hydrogen suppressed graph and an atom connectivity δ_i is the number of (non-hydrogen) atoms directly bonded to atom i . The nature of the atoms is not considered in the calculation.

To account for the nature and unsaturation of the bonds in χ , Kier and Hall⁴ proposed the valence MC χ^v . The valence term of first order, ${}^1\chi^v$, is expressed by;

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

Where δ_i^v is the valence connectivity, representing non-hydrogenic valence of atoms depicted by that vertex, it is defined as

$$\delta_i^v = Z_i^v - h_i \quad (3)$$

in which Z_i^v is the number of valence electrons of atom i , and h_i is the number of hydrogen atoms attached to the atom.

The general equation for the calculation of valence MC index of order m is

$${}^m\chi^v = \sum_{j=1}^{n_m} \prod_{i=1}^{m+1} (\delta_{ij}^v)^{-1/2} \quad (4)$$

where n_m is the number of m th order subgraphs (the number of edges in a subgraph is the order of the subgraph) of the graph.

All the valence MC indices up to third order have been calculated for each molecule. The program for regression is applied to indices. First each index was individually treated to obtain an idea about its statistical importance. The indices whose results have been so found to be encouraging were taken up to third degree and their promising interaction with other indices were also determined. In case results corresponding to a linear-expansion in the single indices were not good, further trial was made by taking their square, cube or interaction with other indices (but of the same molecule) to see if results could still be improved. The programme is so developed that after feeding the input indices of different order and degree and their interaction, we obtain optimal combinations of indices so that there is minimum standard error(s) and sum of square of residue, maximum correlation coefficient(r) and F value.

2. Polynomial equations

The first trend for the correlation came from a plot of b.p. of the normal alkanes against $^{\circ}\chi$, the curve obtained gave an indication that it would better fit a converging polynomial series. Plots of m.p., density and r.i. of these normal alkanes against $^{\circ}\chi$ also showed the same trend. Therefore, an nth degree polynomial equation

$$Y = a_0 + a_1x^1 + a_2x^2 + \dots + a_nx^n \quad (5)$$

was adopted for the best fit using step-wise techniques of regression, where x represents the best indices, y represents the physical property and $a_0, a_1, a_2, \dots, a_n$ are coefficients. It has been observed that the predictive ability of such polynomial equations improves on increasing the value of n to a certain extent. The optimal choice of n was based on the criteria

(i) With increasing value of n, the values of s and ESS should decrease implying the n versus s and n versus ESS curves should approach an asymptotic value, and the sequence of n values should be terminated when an opposite trend starts.

(ii) To delete the insignificant terms in the polynomial equation, the t-values corresponding to each regression coefficient at a given degree of freedom were used. This is based on Efforyson's technique for step wise regression analysis.

If the trend observed for normal alkanes is a fundamental one, it should also be applicable to structurally similar but different classes of compounds like normal unsaturated hydrocarbons. Where instead of sp^3 there is sp^2 or sp hybridization between carbon atoms having double or triple bond respectively. The new polynomials developed for these systems were found to be equally valid in determining their mentioned physical properties.

In Table 1, a polynomial equation of the form of Eq. (5) for each of the four properties is given for the normal alkanes and the unsaturated hydrocarbons. The statistical inferences, namely r , s , ESS and F values for each of these polynomial equations, along with the optimal value of degree of polynomial n and the size of sample N taken have been given in Table 2. Except in the case of m.p. (for reason see the next section) the value of n remains constant with the change in properties. This indicates that this study is not just curve fitting but has scientific significance. It has also been found that for all the properties, further increase in n from the optimal value deteriorates the achieved fit in terms of the statistical parameters.

3. Results and discussion

We have taken 34 normal alkanes from C_1 to C_{34} , and 148 normal unsaturated hydrocarbons ranging from 2 to 32 carbon atoms. Estimated the b.p. of normal alkanes by Eq. (5). The estimated values are in excellent agreement with those of the experimental values⁸. The correlation coefficient(r), F value and standard error are 0.9999, 161433 and 0.18 $^{\circ}C$ respectively. The maximum deviation is 4.5 $^{\circ}C$ for nonane. For the normal unsaturated hydrocarbons too a good agreement between the observed and the estimated values of b.p. can be seen in Table-2. The maximum deviation is 5.8 $^{\circ}C$ for Eicosane. The r , s , ESS and F values are 0.9998, 0.48, 6.03 and 98327 respectively.

The m.p. of the normal alkanes were calculated by Eq.(5). The maximum deviation was 4.2 $^{\circ}C$ for heptane. The corresponding r , s , ESS and F values are 0.9988, 2.83, 216.82 and 2347 respectively. In the case of unsaturated hydrocarbons (Eq.5), the maximum deviation r , F and the standard deviation for m.p. are 5.7 $^{\circ}C$ (for 1, 6-heptadiene-3-yne), 0.9887, 3105 and 3.25 respectively. These results for m.p. show more error in our values, particularly for small molecules. This occurs, because, in normal hydrocarbons with change in carbon numbers their is alternating nature of m.p.. This nature decreases with increase in the chain length.

For density and $n.i.$ of the alkanes the values of standard error and ESS are low and the correlation coefficient and F value are fairly high, which prove the validity of Eq.(5) in predicting the properties. For the unsaturated systems also, an excellent match between the estimated and experimental values is observed for density and refractive index. The maximum difference between the observed and calculated values for density and $n.i.$ are 0.074 and 0.12 respectively for δ , 8-Tetradecadiyne and 1, 3-Pentadiyne respectively.

For comparison, some calculated values are given in Table 3, along with the experimental values. Except for m.p. of small molecules, in most cases the error is less than one percent. Hence Eq. (5) is a universal equation and any physical property of systems of a series can be determined by this equation provided we know the coefficients a_0, a_1, \dots, a_n and the degree corresponding to that physical property. This type of study can be extended to branched molecules by including cluster connectivity terms as well as to other physical properties. The results can still be improved to very small extent by taking higher than third order valence MC terms.

References

1. Randic M, J Am Chem Soc, 97 (1975) 6609.
2. Kier L B, Hall L H, Murray W J & Randic M, J Pharm Sci, 64 (1975) 1971.
3. Murray W J, Hall L H & Kier L B, J Pharm Sci, 64 (1975) 1978.
4. Kier L B & Hall L H, J Pharm Sci, 65 (1976) 1806.
5. Kier L B & Hall L H, Molecular Connectivity in Chemistry and Drug Research (Academic Press, New York), 1976.
6. Kier L B & Hall L H, Molecular Connectivity in Structure Activity Analysis (Research Studies Press, Letchworth), 1986.
7. Kier L B, in: Physical Chemical Properties of Drugs, eds. Yalkowsky S H, Sinkula A A & Valvani S C (Dekker, New York), 1980.
8. Weast R C, Astle M J & Bayer W H, CRC Handbook of Chemistry and Physics, 68th Edition (CRC Press Inc., Florida), 1988.

Table 1. Coefficients in the polynomial $y = a_0 + a_1x + a_2x^2 + \dots + a_nx^n$ corresponding to the best regression equation.

y	a_0	a_1	a_2	a_3	a_4	a_5	a_6
NORMAL ALKANES							
B.P.	-0.28468E+3	0.99301E+2	-0.16320E+2	0.17014E+1	-0.11235E+0	0.43043E-2	-0.11204E-3
M.P.	0.78655E-1	-0.51413E-1	0.64191E-2	-0.48957E+1	0.12304E+3		
Den.	0.52947E-1	0.26940E+0	-0.50767E-1	0.50171E-2	-0.29707E-3	0.90036E-3	-0.19433E-6
R.I.	0.10261 E+1	0.21164E+0	-0.43303E-1	0.4972E-2	0.34219E-3	0.13402E-4	-0.30284E-6
NORMAL UNSATURATED HYDROCARBONS							
B.P.	-0.18926E+3	0.82686E+2	-0.5767E+1	0.26975E+0	-0.70121E-2	0.73016E-4	0.18719E-3
M.P.	0.55638E-1	-0.90712E-1	0.61131E-1	-0.71072E+1	-0.81292E-2	0.44706E-1	
Den.	0.10459E+0	0.28017E+0	-0.53188E-1	0.60530E-2	-0.37012E-3	0.13422E-4	-0.26291E-6
R.I.	0.14283E+1	-0.39715E-1	0.97722E+0	-0.12686E-2	0.66288E-4	0.21511E-5	0.20101E-7

Table 2. Statistics of the best regression equation

Property	n	N	r	F	s	ESS
NORMAL ALKANES						
B.P.	6	32	0.9999	161433	0.18	3.98
M.P.	4	33	0.9988	2347	2.83	216.82
Den.	6	29	0.9996	85815	0.00	8.02
R.I.	6	28	0.9994	55892	0.62	6.47
NORMAL UNSATURATED HYDROCARBONS						
B.P.	6	110	0.9998	98327	0.48	6.03
M.P.	5	93	0.9887	3105	3.25	307.69
Den.	6	112	0.9997	21722	0.81	13.41
R.I.	6	105	0.9992	4216	2.37	24.01

n = Degree of polynomial equation.
 N = Sample size
 r = Correlation coefficient
 F = F value
 s = Standard error
 ESS = sum of squares of residues.

Table 3. MC calculated values of b.p., m.p., den. and r.i. of some compounds compared with experimental values.

Compound	B.P. ^b		M.P. ^b		Density, ^c		R.I. ^d	
	Cal.	Exp. ^a	Cal.	Exp. ^a	Cal.	Exp. ^a	Cal.	Exp. ^a
Pentane	35.0	36.1	-128.7	-130.0	0.5847	0.6262	1.4582	1.4568
Octane	126.2	125.7	-53.1	-56.8	0.6724	0.7025	1.4056	1.3974
Pentadecane	269.1	270.6	9.5	10.0	0.8052	0.7685	1.4502	1.4315
Eicosane	339.6	343.0	37.3	36.8	0.8139	0.7886	1.4403	1.4425
Triacotane	448.1	446.4	64.9	65.8	0.8102	0.8097	1.4535	1.4536
Pentatriacontane	488.9	490.0	74.3	75.0	0.8311	0.8157	1.4529	1.4568
Propylene	-32.8	-34.5	-137.1	-136.0	0.8154	0.7870	1.3958	1.4168
1-Butene	-5.3	-6.3	-186.1	-185.3	0.6310	0.5951	1.4056	1.3962
2-Pentyne	54.1	56.0	-99.3	-101.0	0.6873	0.7107	1.4206	1.4039
1,5-Hexadiene	60.4	59.5	-142.0	-141.0	0.7154	0.6880	1.3991	1.4042
1-Heptyne	100.3	99.7	-82.3	-81.0	0.6980	0.7328	1.4292	1.4087
1-Nonyne	149.4	150.8	-51.5	-50.0	0.8629	0.7568	1.3912	1.4217
1-Decene	168.2	170.5	-67.6	-66.3	0.7618	0.7408	1.3990	1.4215
5-undecyne	199.4	198.0	-75.3	-74.1	0.8013	0.7753	1.4503	1.4369
1-Dodecyn	213.6	215.0	-18.5	-19.0	0.8152	0.7788	1.4433	1.4340
1-Tridecene	229.5	232.8	-12.6	-13.0	0.7551	0.7658	1.4433	1.4340
1-Hexadecyne	286.2	284.0	15.4	15.0	0.7706	0.7965	1.4391	1.4440
1-Octadecyne	309.2	313.0	23.2	22.5	0.8352	0.8025	1.4458	1.4774
1-Nonadecyne	230.7	227.0	38.1	37.0	0.8154	0.8054	1.4559	1.4488
1-Eicosyne	345.8	340.0	34.7	36.0	0.7853	0.8073	1.4468	1.4501

a) Ref. 8

b) Boiling and melting points in °C at atmospheric pressure (760mm)

c) Densities in g./cm³ at 20°C

d) Refractive indices for the D line of sodium spectrum (n) at 20°C.