Russian Journal of General Chemistry, Vol. 74, No. 6, 2004, pp. 828–833. Translated from Zhurnal Obshchei Khimii, Vol. 74, No. 6, 2004, pp. 900–905.

Quantitative Structure–Property Relationship: XVII.[1] Properties of Branched Hydrocarbon Molecules

I. B. Golovanov and S. M. Zhenodarova

Institute of Theoretical and Experimental Biophysics, Russian Academy of Sciences, Pushchino, Moscow oblast, Russia

Received June 13, 2002

Abstract — A simple approach to estimating properties of branched molecules is suggested: A property of any nonlinear molecule is considered as a sum of the property of the corresponding linear molecule and correction "for branching," determined by the interaction of atoms of the substituting group with atoms of the main chain. The potential of this approach is demonstrated by the calculation of the melting points, heat capacities, entropies, and enthalpies and free energies of formation for 117 saturated hydrocarbon molecules, including all the linear C_1–C_{20} molecules and branched C_4–C_{10} molecules with methyl substituents; also the heats of vaporization are calculated for 72 molecules including all linear C_1–C_{20} molecules and branched C_4–C_9 molecules with methyl substituents. The accuracy of all the estimates is high. When the linear contribution is taken into account more accurately, with correction for nonlinear variation of properties of linear molecules, it becomes possible to highly accurately in estimate various properties of both linear and branched molecules, using the molecular connectivity indices.

In [2] we suggested a general expression for the quantitative structure–property relationship, having the following form for linear molecules of A_n type:

$$P_{(n)} = n\alpha + (n - 1)\beta + (n - 2)\gamma + \ldots, \quad (1)$$

where α is the one-center contribution to a property of fragment A_i; β is the two-center contribution corresponding to interaction of fragments i and $i + 1$; γ is the two-center contribution corresponding to interaction of fragments i and $i + 2$, etc. In nonlinear molecules, the number of contributions α is always equal to n, and the number of the other contributions depends on the extent of branching and can be determined from the structural formula of the molecule.

Here we apply the suggested approach to estimation of various properties of a large number of isomers of saturated hydrocarbons and compare the results with those furnished by well-known alternative approaches.

In [3, 4] it was suggested to use the molecular connectivity indices as descriptors determining a property:

$$^h\chi = \Sigma(\delta_i, \delta_{i+1}, \delta_{i+h})^{-0.5}, \quad (2)$$

where δ_i are valences of the molecular graph vertices, $h \geq 0$, and summation is over all paths of length h, with h varying from 0 to $n - 1$ (n is the number of graph vertices). For $^0\chi$, summation is over all graph vertices (path length 0), and the number of contributions $\delta_i^{-0.5}$ in this case is equal to n. For $^1\chi$, summation is over all pathways of length 1 (bonds), and the number of contributions $(\delta_i, \delta_{i+1})^{-0.5}$ is $(n - 1)$, etc. Since the contributions $(\delta_i, \delta_{i+1}, \ldots, \delta_{i+h})^{-0.5}$ are close, it is evident that $^0\chi \sim n$, $^1\chi \sim (n - 1)$, etc., i.e., the values of $^i\chi$ determining a property according to (3) should correlate with the values of n_i determining the numbers of different contributions in (1):

$$P = k + \Sigma k_i^i\chi. \quad (3)$$

Table 1 illustrates an example of the determination of properties for a series of simplest hydrocarbon molecules by Eqs. (1) and (2). The values of n_i and $^i\chi$ show a strong mutual correlation; the correlation coefficients exceed 0.96.

From the formal mathematical viewpoint, expressions (1) and (3) are equivalent; however, the following distinction should be noted: Relationship (1) for individual molecules follows from the molecular orbital theory, the numbers n_i of contributions P_i are preset by the molecular structure, and, since matrix elements P_{ij} depend on the distance between the fragments, it is possible to construct schemes in which P_{1n} is determined via P_{13} or P_{14}, as it was done in [2]. This allows approximate consideration of all remote

[1] For communication XVI, see [1].

1070-3632/04/7406-0828

Table 1. Determination of properties of some saturated hydrocarbon molecules

Molecule[a]	Property	
	by (1)	by (3)
2	$2\alpha + \beta$	$2k_0 + k_1$
3	$3\alpha + 2\beta + \gamma$	$2.70k_0 + 1.41k_1 + 0.71k_2$
4	$4\alpha + 3\beta + 2\gamma + \delta_{14}$	$3.31k_0 + 1.91k_1 + 1.0k_2 + 0.5k_3$
5	$5\alpha + 4\beta + 3\gamma + 2\delta_{14} + \delta_{15}$	$4.12k_0 + 2.41k_1 + 1.35k_2 + 0.71k_3 + 0.75k_4$
2-Me3	$4\alpha + 3\beta + 3\gamma$	$3.57k_0 + 1.73k_1 + 1.73k_2$
2-Me4	$5\alpha + 4\beta + 4\gamma + 2\delta_{14}$	$4.28k_0 + 2.27k_1 + 1.80k_2 + 0.80k_3$
2,2-$(Me)_2$3	$5\alpha + 4\beta + 6\gamma$	$4.50k_0 + 2.0k_1 + 3.0k_2$

[a] Here and in Tables 2–4, the hydrocarbon formulas are given in a simplified form [4]: The figure denotes the number of carbon atoms in the linear chain, and for branched molecules, the number and position of substituents is indicated, e.g., 2,2-dimethylropane is denoted as 2,2-$(Me)_2$3.

contributions with a small number of parameters, whereas with relationship (3) consideration of each contribution requires determination of its weight k_i. In [2], in accordance with (1), we estimated the boiling points of linear saturated hydrocarbons C_1–C_{20} using five parameters and reached the estimation quality characterized by r 1.0000 and s 0.5971, whereas with relationship (3) such a high estimation quality cannot be attained even with a considerably greater number of parameters.

It was intended to apply relationship (3) involving topological indices $^h\chi$ to estimation of properties of branched molecules [3]. Since the properties of linear molecules cannot be determined with a high accuracy with this approach, the problem can be subdivided in two steps: development of a scheme for calculating properties of linear molecules and consideration of branching in terms of the perturbation theory.

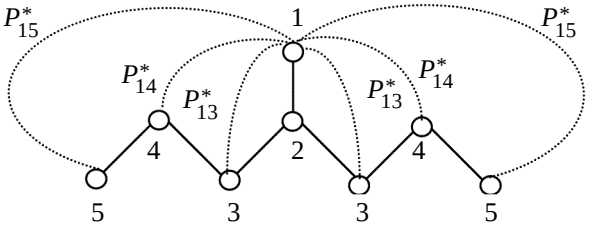
In [1] we suggested a simple method for estimating properties of linear molecules. The property of any nonlinear molecule can be represented as

$$P_{\text{nonlin}} \cong P_{\text{lin}} + \Delta P. \quad (4)$$

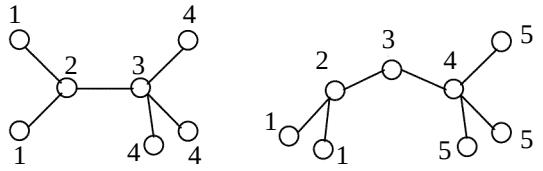
Calculation of ΔP can be based on the previously described approaches [2]. Within the framework of (4), it is natural to assume that

$$\Delta P \cong k_1 + n_{13}P_{13} + n_{14}P_{14} + n_{15}P_{15} + n_{14}^{23}P_{14}^{23} + n_{15}^{23}P_{15}^{23} + n_{14}^{33}P_{14}^{33} + n_{15}^{33}P_{15}^{33}, \quad (5)$$

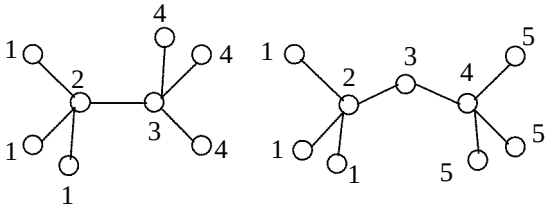
where P_{ik}^* is the contribution corresponding to a $1\cdots k$ interaction of an atom of a branch-forming substituent with atoms of the main chain, as shown in the scheme below.



P_{14}^{23} and P_{15}^{23} are the contributions corresponding, respectively, to the following Mayer graphs:



P_{14}^{33} and P_{15}^{33} are the contributions corresponding to the following graphs:



and n_{ij} is the number of the corresponding contributions.

To explain the choice of the relationship determining ΔP, it is useful to compare the differences in ΔT_b between linear and branched isomers with those determined by relationship (4). Some examples are given in Table 2. It follows from the table that $P_{13}^* \approx 5.5$, $P_{14}^* \approx -3.0$, $P_{15}^* \approx 1.0$, and $P_{16}^* \approx 0.1$. Thus, to estimate

Table 2. Corrections ΔP and differences in the boiling points of isomeric hydrocarbons

Isomer		ΔP	ΔT_b	Parameter value
linear	branched			
4	2-Me3	$2P^*_{13}$	11.2	P^*_{13} 5.6
5	2-Me4	$2P^*_{13} + P^*_{14}$	8.2	P^*_{14} –3.0
6	3-Me5	$2P^*_{13} + 2P^*_{14}$	5.5	P^*_{14} –2.7
7	3-Me6	$2P^*_{13} + 2P^*_{14} + P^*_{15}$	6.6	P^*_{15} 1.1
8	4-Me7	$2P^*_{13} + 2P^*_{14} + 2P^*_{15}$	8.1	P^*_{15} 1.5
6	2-Me5	$2P^*_{13} + P^*_{14} + P^*_{15}$	8.5	
7	2-Me6	$2P^*_{13} + P^*_{14} + P^*_{15} + P^*_{16}$	8.4	P^*_{16} –0.1
9	4-Me8	$2P^*_{13} + 2P^*_{14} + 2P^*_{15} + P^*_{16}$	8.3	P^*_{16} 0.2
5	2,2-$(Me)_2$3	$5P^*_{13}$	26.6	P^*_{13} 5.3
7	3,3-$(Me)_2$5	$5P^*_{13} + 4P^*_{14}$	12.4	P^*_{14} –3.7
9	4,4-$(Me)_2$6	$5P^*_{13} + 4P^*_{14} + 4P^*_{15}$	15.6	P^*_{15} 0.8

Table 3. Corrections $\Delta^i\chi$ and differences in the boiling points of some isomeric hydrocarbons

Isomer		$\Delta^0\chi$	$\Delta^1\chi$	$\Delta^2\chi$	$\Delta^3\chi$	$\Delta^4\chi$	$\Delta^4\chi_?$[a]	ΔT_b
linear	branched							
5	2,2-$(Me)_2$3	–0.379	0.414	–1.646	–0.707	0.754	0.50	26.6
7	3,3-$(Me)_2$5	–0.368	0.293	–0.810	–0.707	0.527	0.25	12.4
9	4,4-$(Me)_2$7	–0.326	0.293	–0.896	–0.146	–0.449	0.25	15.6

[a] $\Delta^4\chi_c$ is the connectivity index for the star graph  (cluster) [4].

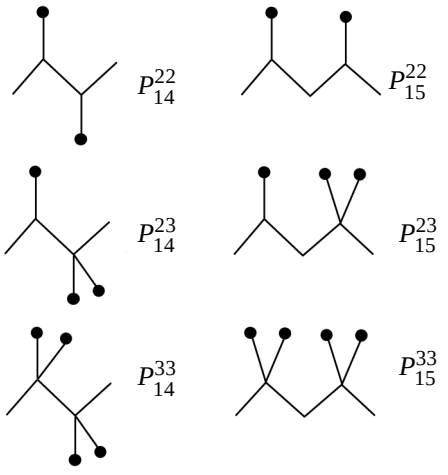
the boiling points with a ~1 K accuracy, contributions more remote than P^*_{15} can be neglected. This conclusion is not so evident when ΔP is expressed via $\Delta^h\chi$. As example, we give in Table 3 certain values of $\Delta^h\chi$ determining ΔP and the corresponding values of ΔT_b. Even for isomers strongly differing in the boiling points, variations parallel with ΔT_b are exhibited only by $\Delta^2\chi$. At the same time, it is quite plausible that a combination of different contributions $\Delta^h\chi$ can give accurate estimations of ΔT_b.

The contributions P^*_{13}, P^*_{14}, and P^*_{15} in Eq. (4) coincide in physicochemical sense with the contributions $P_{13} = \gamma$, $P_{14} = \delta_{14}$, and $P_{15} = \delta_{15}$ in the expression for determination of a property of linear molecules, but differ from them in value. This is due to the fact that bond lengths and bond angles, as well as mutual arrangement of more remote fragments, are different in linear and branched hydrocarbon molecules [5]. For example, the *n*-butane molecule can exist in the form of one *trans* and two *gauche* conformers having close weights in the equilibrium mixture, i.e., $P_{14} \sim 1/3(P^{trans}_{14} + 2P^{gauche}_{14})$. The 2-methylbutane molecule has also three conformers, but the weight of one of them will be considerably lower; therefore, both P_{14} and P^*_{14} will differ from those characteristic of linear molecules. This is also true for the contributions corresponding to more remote interactions.

A branched molecule can be simulated by a linear molecule with various quasiatoms. For example, if the *n*-butane molecule is simulated by a C–C–C–C molecule, the 2-methylbutane molecule will be simulated by C–C*–C–C, and the 2,2-dimethylbutane molecule, by C–C**–C–C. Interactions between quasiatoms C* and C** will be different, and, to take them into account, it is necessary to introduce new parameters taking into account the "interaction of branchings" and correspodning to the following Mayer graphs.

Table 4. Boiling points and heats of vaporization of nonane isomers containing methyl groups

Molecule	T_b, K			ΔH_{vap}, kJ mol^{-1}		
	experiment [6, 7]	calculation by		experiment [6, 7]	calculation by	
		Eqs. (4), (5)	Eq. (7)		Eqs. (4), (5)	Eq. (7)
9	423.9	422.5	423.4	37.74	37.45	37.49
2-Me8	416.4	415.2	414.7	36.62	36.41	36.41
3-Me8	417.4	418.1	417.2	36.74	36.70	36.62
4-Me8	415.6	416.6	416.3	36.58	36.62	36.53
2,2-$(Me)_2$7	405.8	403.5	404.2	34.74	34.78	34.78
2,3-$(Me)_2$7	413.6	417.7	413.0	36.07	35.95	35.99
2,4-$(Me)_2$7	406.0	407.8	407.9	35.32	35.53	35.45
2,5-$(Me)_2$7	409.1	410.8	408.5	35.57	35.66	35.57
2,6-$(Me)_2$7	408.4	407.9	405.9	35.49	35.36	35.32
3,3-$(Me)_2$7	410.2	409.4	410.5	35.28	35.32	35.36
3,4-$(Me)_2$7	413.7	415.2	414.7	36.32	36.18	36.03
3,5-$(Me)_2$7	409.1	412.2	411.1	35.61	35.86	35.74
4,4-$(Me)_2$7	408.3	406.4	409.7	35.32	35.20	35.20
2,2,3-$(Me)_3$6	406.8	407.4	407.8	34.74	34.74	34.78
2,2,4-$(Me)_3$6	399.7	400.4	400.4	33.98	34.07	34.07
2,2,5-$(Me)_3$6	397.2	396.2	396.3	33.73	33.73	33.86
2,33-$(Me)_3$6	410.8	410.3	411.3	34.94	35.03	35.07
2,3,4-$(Me)_3$6	412.2	410.8	411.7	35.66	35.36	35.45
2,3,5-$(Me)_3$6	404.5	404.9	404.2	34.78	34.86	34.86
2,4,4-$(Me)_3$6	403.8	403.4	403.3	34.28	34.32	34.28
3,3,4-$(Me)_3$6	413.6	413.2	414.6	35.11	35.28	35.24
2,2,3,3-$(Me)_4$5	413.4	412.8	411.9	35.24	34.90	34.65
2,2,3,4-$(Me)_4$5	406.2	405.8	406.9	34.23	34.11	34.28
2,2,4,4-$(Me)_4$5	395.4	396.0	392.7	32.81	32.81	32.56
2,3,3,4-$(Me)_4$5	414.7	412.8	414.8	34.90	34.78	35.03
r		0.9997	0.9999		0.9995	0.9995
s		1.6187	1.1405		0.0470	0.0446



Since we take into account the interactions P_{13}^*, P_{14}^*, and P_{15}^* arising from branchings, with these parameters the contributions P_{14}^{22} and P_{15}^{22} can be approximately taken into account, whereas for the remaining contributions the estimates will be rougher, since in such strained structures the interactions corresponding to these contributions will be different because of different weights of conformers in the equilibrium mixture. Therefore, we will approximately take into account the contributions P_{14}^{22} and P_{15}^{22} using a combination of the contributions P_{13}^*, P_{14}^*, and P_{15}^*, whereas the other four contributions should be taken into account explicitly. Since there are no sufficient grounds to choose the relationship between these contributions, no less than seven parameters are required to estimate ΔP, as it is written in (5).

With topological indices ${}^n\chi$, we can act similarly. Let us assume that the properties of linear hydrocarbons will be estimated within a specific scheme and the corrections "for branching," with relationship (6):

Table 5. Parameters used for calculating boiling points and ΔH_{vap}

Parameter	Relationships (4) and (5)		Parameter	Relationship (7)	
	T_b	ΔH_{vap}		T_b	ΔH_{vap}
k_1	–41.8808 ± 1.6653	–4.5718 ± 0.3427	$k_1(\sqrt{n}$	154.7842 ± 3.6118	7.5078 ± 2.4163
k_2	158.8016 ± 0.8020	12.7415 ± 0.2619	$k_2(\exp\sqrt{n})$	–0.6431 ± 0.0775	1.9497 ± 2.1439
k_3	–0.5992 ± 0.0311	0.1904 ± 0.0255	$k_3({}^0\chi)$	–42.3315 ± 3.8106	–0.4100 ± 0.1197
P^*_{13}	–4.3747 ± 0.1246	–0.6213 ± 0.0234	$k_4({}^1\chi)$	52.6317 ± 2.0881	6.1844 ± 1.1811
P^*_{14}	2.9269 ± 0.1552	0.2740 ± 0.0272	$k_5({}^2\chi)$	10.2585 ± 1.7994	2.1342 ± 0.6150
P^*_{15}	–1.4868 ± 0.1822	–0.0594 ± 0.0376	$k_6({}^3\chi_p)$[a]	5.7018 ± 1.2203	–0.7657 ± 0.2402
P^{23}_{14}	2.4243 ± 0.5488	0.1837 ± 0.1050	$k_7({}^4\chi_p)$[a]	–4.1527 ± 1.3448	–3.2577 ± 0.5958
P^{23}_{15}	4.2934 ± 0.6515	0.1632 ± 0.1381	$k_8({}^5\chi_c)$	12.9881 ± 1.4632	4.4978 ± 0.8619
P^{33}_{14}	7.7382 ± 1.1799	1.0744 ± 0.2305	$k_9({}^6\chi_c)$	–11.8411 ± 2.0048	0.5004 ± 0.1079
P^{33}_{15}	17.4288 ± 1.3064	0.9632 ± 0.2870	$k_{10}({}^5\chi_{pc})$[a]	5.2549 ± 0.5517	0.4050 ± 0.1134

[a] ${}^n\chi_p$ is the connectivity index for a linear graph of the  type (path); ${}^n\chi_{pc}$ is the connectivity index for a graph of the  type (path–cluster) [4].

Table 6. Parameters used in calculations of C_p, S^0, ΔH^0_f, and ΔG^0_f and statistical criteria of their estimation

Parameter	C_p	S^0	ΔH^0_f	ΔG^0_f
k_1	6.4726 ± 1.2113	157.5498 ± 1.4874	–43.1299 ± 0.7552	–51.1214 ± 0.9828
k_2	22.8999 ± 0.1075	38.8765 ± 0.1318	–20.6794 ± 0.0669	8.3584 ± 0.0870
k_3	14.0428 ± 8.0964	–30.7545 ± 9.9403	–25.4111 ± 5.0484	–16.4394 ± 6.5697
P^*_{13}	0.6611 ± 0.2155	–8.3542 ± 0.2644	–4.8488 ± 0.1343	–2.3665 ± 0.1749
P^*_{14}	–1.3016 ± 0.2653	3.0957 ± 0.3255	2.6769 ± 0.1653	1.7527 ± 0.2150
P^*_{15}	–0.4824 ± 0.3109	0.2527 ± 0.3816	0.6669 ± 0.1937	0.5987 ± 0.2523
P^{23}_{14}	2.9849 ± 0.9389	–3.9581 ± 1.1527	4.0786 ± 0.5853	5.3053 ± 0.7619
P^{23}_{15}	–0.5690 ± 1.1008	–1.7184 ± 1.3518	–6.5350 ± 0.6856	5.9174 ± 0.8933
P^{33}_{14}	8.2488 ± 2.0263	–13.0269 ± 2.4878	11.2934 ± 1.2636	15.2582 ± 1.6443
P^{33}_{15}	1.7313 ± 2.2071	–4.4965 ± 2.7096	20.0844 ± 1.3761	21.4284 ± 1.7908
r	0.9991	0.9996	0.9996	0.9962
s	0.6623	0.8131	0.4130	0.5374

$$\Delta P = k + \Sigma k_i \Delta^h\chi, \quad (6)$$

where $\Delta^h\chi$ - ${}^h\chi_{branch}$ – ${}^h\chi_{lin}$.

Since there are no clear physicochemical principles for assessing the significance of various contributions, it is necessary first to use a wide set of $\Delta^h\chi$ and then to discard insignificant contributions.

Let us estimate P_{lin} as it was suggested in [1]. For the boiling points and ΔH_{vap}, $P_{lin} = k_1 + k_2\sqrt{n} + k_3 e^{\sqrt{n}}$, and for C_p, S^0, ΔH^0_f, and ΔG^0_f, $P_{lin} = k_1 + k_2 n + k_3 e^{-n}$.

The quantities ΔP for any properties are determined by relationships (5) and (6). In the final expression for a property P_{nonlin}, the free terms from the expressions for P_{lin} and ΔP are combined in one term, and within approach (4) it appears necessary to determine ten parameters: k_1, k_2, and k_3 for P_{lin} and seven parameters P^*_{ij} and P^{mn}_{kl} entering into (5).

Using this approach, we calculated such properties as the boiling point, heat capacity, entropy, and enthalpy and free energy of formation of 117 saturated hydrocarbons for which the corresponding systematic data are available. This group included all the linear C_1–C_{20} molecules and C_4–C_{10} isomers with methyl substituents. Also, we calculated the heats of vaporization of 72 saturated hydrocarbons including linear C_1–C_{20} molecules and C_4–C_9 isomers with methyl substituents. Selected data on the boiling points and ΔH_{vap} are given in Tables 4 and 5, and the coefficients

and statistical criteria for estimating C_p, S^0, ΔH_f^0, and ΔG_f^0, in Table 6. In all the cases, we obtained very accurate estimates of the propertuies of various saturated hydrocarbon isomers. Virtually the same results are obtained in calculations with (6). When approach (3) is used directly, the accuracy is poor. For example, for the above-mentioned 72 saturated hydrocarbons, we obtained for boiling points r 0.9912, but s was as large as 12.5553, even though we used 15 parameters; the largest deviations (>34.0) were observed with linear molecules. When the linear contribution was taken into account more accurately using (7), we obtained highly accurate estimates of properties with only ten parameters.

$$P = k_1 + k_2\sqrt{n} + k_3 e^{\sqrt{n}} + \Sigma k_i {}^{i}\chi. \qquad (7)$$

The corresponding results for the boiling points and ΔH_{vap} are listed in Tables 4 and 5. Thus, by introducing a correction for nonlinear variation of properties of linear molecules as in Eq. (7), it is possible to obtain highly accurate estimates for various properties using the molecular connectivity indices ${}^{h}\chi$ also.

We used relationships (4) and (5) to estimate the same properties of 57 monohydric alcohols of various structures. The accuracy of the results obtained was close to that attained with hydrocarbons, and there is no doubt that these relationships will describe with a high accuracy the properties of other compounds R–X also.

REFERENCES

1. Golovanov, I.B. and Zhenodarova, S.M., *Zh. Obshch. Khim.*, 2004, vol. 74, no. 5, p. 765.
2. Golovanov, I.B. and Zhenodarova, S.M., *Zh. Obshch. Khim.*, 2003, vol. 73, no. 1, p. 90.
3. Randič, M., *J. Am. Chem. Soc.*, 1975, vol. 97, no. 23, p. 6609.
4. Kier, L.B. and Hall, L.H., *Molecular Connectivity in Chemistry and Drug Research*, New York: Academic, 1976.
5. Vilkov, L.V., Mastryukov, V.S., and Sadova, N.I., *Opredelenie geometricheskogo stroeniya svobodnykh molekul* (Determination of the Geometric Structure of Free Molecules), Leningrad: Khimiya, 1978.
6. *CRC Handbook of Chemistry and Physics*, Lide, D.R., Ed., Boca Raton: CRC, 1993–1994, sect. 3, 5, 6.
7. Stull, D. R., Westrum, E. F., and Sinke, G. C., *The Chemical Thermodynamics of Organic Compounds*, New York: Wiley, 1969.