Estimation of Pure Compound Properties Using Group-Interaction Contributions

Jorge Marrero-Morejón and Eladio Pardillo-Fontdevila
Centro de Química Farmacéutica, Ave. 200 y 21 Atabey, POB 16042, 11600-Habana, Cuba

A family of models proposed here estimate the critical constants and normal boiling points of pure organic compounds from chemical structure by using a new structural approach called group-interaction contributions, which considers the contributions of interactions between bonding groups in the molecule instead of the contributions of simple groups. Compared to the conventional group-contribution techniques, the proposed models demonstrate significant improvements in accuracy, as well as the ability to distinguish among isomers.

Introduction

The basis for the design of chemical processing units is the set of physical and thermodynamic properties of the compounds that undergo transformation. However, it is not always possible to find experimental values of properties for the compounds of interest in the literature. Since it is not practical either to measure them as the need arises, estimation methods are profusely employed.

For the estimation of properties of pure compounds, group-contribution methods (Joback and Reid, 1987; Lyderson, 1955; Ambrose, 1978; Klincewicz and Reid, 1984; Lyman et al., 1990) have been widely used. In these methods, the property of a compound is a function of structurally dependent parameters (Σ), which are determined by summing the number frequency of each group occurring in the molecule multiplied by its contribution. These methods have the advantage of supplying quick estimates without requiring substantial computational resources. However, many of them are of questionable accuracy and are unable to distinguish among isomers due to the oversimplification of the molecular structure, which the simple group-contribution approach can.

To overcome this limitation, several attempts have been reported in the literature. Constantinou et al. (1993, 1994) have proposed a quite complex estimation technique, which is based on conjugate forms (alternative formal arrangements of valence electrons). This technique provides accurate estimations of several properties of pure compounds and allows the capture of the differences among isomers. However, the generation of conjugate forms is a nontrivial issue and requires a symbolic computing environment (Prickett et al., 1993).

A less complex method has been proposed by Constantinou and Gani (1994). Estimation is performed at two levels: the basic level uses the contribution from simple first-order groups, while the second level uses a small set of second-order groups that have the first-order ones as building blocks. The role of the second-order groups is to consider the proximity effects and to distinguish among isomers. For a proper determination of which of them occur in a given molecule, however, it is necessary to apply supplementary and nonreported rules (Abildskov, personal communication, 1995). Otherwise, the same molecule may be described in different ways.

Pardillo-Fontdevila and González-Rubio (1997) have recently proposed a new structural approach [called group-interaction contribution (GIC)], that considers the contributions of interactions between bonding groups instead of the contributions of simple groups. In this work, we propose new models, based on this new approach, for the estimation of normal boiling points and critical constants of pure organic compounds.

Development of the Models

In this work, the properties of a compound are considered to be functions of structurally dependent parameters (Σ), which are thereby determined by summing the number fre-

---

Correspondence concerning this article should be addressed to J. Marrero at his current address: Dept. of Chemical Engineering, CAPEC, Technical Univ. of Denmark, Building 228, DK-2800 Lyngby, Denmark.

Table 1. Estimation Equations Used in this Work*

Property	Symbol	Unit	Equation	
Normal boiling point**	Tb	K	$Tb = M^a\Sigma + b$	(1)
Normal boiling point	Tb	K	$Tb = a + \Sigma$	(2)
Critical temperature	Tc	K	$Tc = Tb/[a + b\Sigma - \Sigma^2]$	(3)
Critical pressure†	Pc	bar	$Pc = [a + bn_A - \Sigma]^{-2}$	(4)
Critical volume	Vc	cm^3/mol	$Vc = a + \Sigma$	(5)

*Σ = sum of the products each simple group (or interaction between groups) appears in the molecule and their contributions
**M = molecular weight
†n_A = total number of atoms in the molecule

quency of each group-interaction occurring in the molecule times its contribution. We selected 39 simple groups to generate a consistent set of group-interactions that allows one to treat a wide variety of organic compounds. These groups are the same as those used by Joback and Reid (1987), and earlier by Lydersen (1955), with the omission of $=NH$ and $=N-$ (nonring). We compiled property values of 507 pure compounds, obtained exclusively from Reid et al. (1987), for use in the development of the models.

With the object of comparing the group-contribution approach to the GIC approach, a least-square analysis was carried out to determine the contributions of both simple groups and group interactions. The estimation equations for each property are shown in Table 1. The fitted parameters of these equations and the statistical summary of regression analysis are presented in Tables 2 and 3, respectively. The equations are the same as those used by Joback and Reid (1987). In addition, we propose a new alternative nonlinear equation for estimating the normal boiling point, which makes a significant improvement in accuracy as shown in the statistical results.

The contributions of simple groups and group-interactions are presented in Tables 4 and 5, respectively. Notice that there are group interactions that have not been calculated because of the lack of property values of the compounds involving these interactions. We used the singular-value decomposition procedure (Forsythe et al., 1977) as the optimization algorithm for linear regression. For nonlinear regression, we employed the well-known Levenberg−Marquardt procedure. In both cases, the objective function was to minimize the sum of squares of the differences between experimental and estimated values.

Table 2. Fitted Parameters of Estimation Equations

	Group-Contribution Approach		Group-Interaction Contribution Approach	
Eq.	*a*	*b*	*a*	*b*
(1)	−0.366	149.84	−0.404	156.00
(2)	218.13	—	204.66	—
(3)	0.5881	−0.9305	0.5851	−0.9286
(4)	0.1218	0.4609	0.1285	−0.0059
(5)	26.3	—	25.1	

Table 3. Statistical Summary of Regression Analysis

			Simple Group-Contribution			Group-Interaction Contribution		
Eq.	Property	Data Points	AAE*	STD**	APE†	AAE	STD	APE
(1)	Tb††	507	10.11	13.94	2.86	4.99	7.35	1.34
(2)	Tb	507	12.66	17.01	3.58	6.48	9.53	1.73
(3)	Tc	491	4.71	6.32	0.82	2.79	4.39	0.48
(4)	Pc	445	2.03	3.09	5.30	1.06	1.70	2.92
(5)	Vc	289	7.32	10.59	2.33	4.56	6.68	1.45

*AAE = average absolute error
**STD = standard deviation of regression
†APE = average percent error
††Equation 1 is the one proposed for estimating Tb.
Note: AAE and STD have units of K for Tb and Tc, bar for Pc, and cm^3/mol for Vc.

Table 4. Contributions of Simple Groups

	Tb	Tb*	Tc	Pc	Vc
Nonring groups					
$-CH_3$	11.41	98.48	−0.0144	1.8345	58.6
$-CH_2-$	22.90	198.22	−0.0185	1.3731	56.6
$>CH-$	34.09	264.13	−0.0161	0.9124	48.3
$>C<$	48.08	328.01	−0.0102	0.4525	42.2
$=CH_2$	2.82	67.90	−0.0076	1.3750	50.5
$=CH-$	27.76	194.17	−0.0136	0.9144	47.8
$=C<$	31.63	258.41	−0.0119	0.4527	38.0
$=C=$	28.10	223.18	−0.0079	0.4604	35.5
$\equiv CH$	−10.06	61.28	−0.0115	0.9183	42.1
$\equiv C-$	34.38	221.38	0.0032	0.4602	37.7
Ring groups					
$-CH_2-$	22.50	174.48	−0.0092	1.3768	47.3
$>CH-$	30.48	247.01	−0.0141	0.9125	42.3
$>C<$	46.39	324.31	−0.0046	0.4567	33.0
$=CH-$	23.24	179.86	−0.0088	0.9180	40.7
$=C<$	38.42	290.21	−0.0103	0.4566	34.0
Halogen groups					
$-F$	−14.79	−45.91	−0.0108	0.4560	19.9
$-Cl$	25.63	267.42	−0.0084	0.4577	51.0
$-Br$	43.53	508.17	−0.0074	0.4562	73.5
$-I$	89.39	950.06	−0.0017	0.4605	87.4
Oxygen groups					
$-OH$ (alcohol)	84.51	474.98	−0.0678	0.9309	20.5
$-OH$ (phenol)	59.95	350.99	−0.0273	0.9352	−7.6
$-O-$ (nonring)	25.83	203.21	−0.0174	0.4577	16.1
$-O-$ (ring)	25.78	181.91	−0.0092	0.4637	11.1
$>CO$ (nonring)	72.21	522.16	−0.0327	0.9165	66.5
$>CO$ (ring)	96.98	605.02	−0.0183	0.9203	52.4
$-CHO$ (aldehyde)	68.99	451.84	−0.0306	1.3837	76.6
$-COOH$ (acid)	146.99	948.08	−0.0843	1.8407	85.4
$-COO-$ (ester)	80.61	610.80	−0.0469	1.3716	88.3
$=O$ (other)	−17.05	31.79	−0.0092	0.4669	32.7
Nitrogen groups					
$-NH_2$	58.67	325.69	−0.0210	1.3870	31.1
$>NH$ (nonring)	56.91	362.92	−0.0291	0.9204	44.4
$>NH$ (ring)	64.96	369.36	−0.0154	0.9305	28.6
$>N-$ (nonring)	49.72	347.77	−0.0158	0.4565	34.6
$=N-$ (ring)	44.74	240.89	−0.0108	0.4692	−24.1
$-CN$	111.45	639.92	−0.0525	0.9107	86.3
$-NO_2$	144.77	912.64	−0.0409	1.3878	88.3
Sulfur groups					
$-SH$	52.67	428.90	0.0048	0.9313	62.8
$-S-$ (nonring)	72.12	519.86	−0.0100	0.4618	59.5
$-S-$ (ring)	66.09	444.31	0.0008	0.4652	30.1

*Contributions to structurally dependent parameter (Σ) of estimation equation (Eq. 1).

Table 5. Contributions of Group Interactions

	Tb	*Tb**	*Tc*	*Pc*	*Vc*		*Tb*	*Tb**	*Tc*	*Pc*	*Vc*
Interactions with CH_3- (via single bond)											
CH_3- and CH_3-	−20.82	61.28	−0.0213	−0.0576	123.2	CH_3- and $-OH$ (a)	133.04	736.93	−0.0876	−0.0180	92.9
CH_3- and $-CH_2-$	33.19	194.25	−0.0227	−0.0430	88.6	CH_3- and $-O-$	31.94	228.01	−0.0205	−0.0321	66.0
CH_3- and $>CH-$	26.94	194.27	−0.0223	−0.0376	78.4	CH_3- and $>CO$	64.46	445.61	−0.0362	−0.0363	88.9
CH_3- and $>C<$	22.71	186.41	−0.0189	−0.0354	69.8	CH_3- and $-CHO$	89.34	636.49	−0.0606	−0.0466	128.9
CH_3- and $=CH-$	18.17	137.18	0.8526	0.0654	81.5	CH_3- and $-COOH$	186.44	1228.84	−0.0890	−0.0499	145.9
CH_3- and $=C<$	23.91	182.20	0.1792	0.0851	57.7	CH_3- and $-COO-$ (o)	58.87	456.92	0.0267	0.1462	93.3
CH_3- and $\equiv C-$	23.04	194.40	0.3818	−0.2320	65.8	CH_3- and $-COO-$ (c)	65.95	510.65	−0.0974	−0.2290	108.2
CH_3- and $>CH-$ (r)	25.68	176.16	−0.0214	−0.0396	58.3	CH_3- and NH_2-	62.14	443.76	−0.0397	—	—
CH_3- and $>C<$ (r)	20.25	180.60	0.1117	−0.0597	49.0	CH_3- and $>NH$	41.60	293.86	−0.0313	−0.0317	—
CH_3- and $=C<$ (r)	19.61	145.56	0.0987	−0.0746	71.7	CH_3- and $>N-$	23.78	207.75	−0.0199	−0.0348	76.3
CH_3- and $-F$	−9.96	160.83	−0.0370	−0.0345	88.1	CH_3- and $-CN$	150.14	891.15	−0.0766	−0.0507	147.9
CH_3- and $-Cl$	44.44	453.70	−0.9141	−0.0231	113.8	CH_3- and $-NO_2$	169.64	1148.58	−0.0591	−0.0385	148.1
CH_3- and $-Br$	71.94	758.44	−0.9166	−0.0239	—	CH_3- and $-SH$	74.44	588.31	−0.9192	−0.0244	119.7
CH_3- and $-I$	111.04	1181.44	−0.9146	−0.0241	—	CH_3- and $-S-$	53.24	409.85	−0.0181	−0.0305	87.9
Interactions with nonring $-CH_2-$ (via single bond)											
$-CH_2-$ and $-CH_2-$	22.00	244.88	−0.0206	−0.0272	56.6	$-CH_2-$ and $-OH$ (a)	108.85	673.24	−0.0786	−0.0119	45.2
$-CH_2-$ and $>CH-$	19.78	244.14	−0.0134	−0.0219	40.2	$-CH_2-$ and $-O-$	25.03	243.37	−0.0205	−0.0184	34.5
$-CH_2-$ and $>C<$	22.17	273.26	−0.0098	−0.0162	32.0	$-CH_2-$ and $>CO$	50.77	451.27	−0.0256	−0.0204	62.3
$-CH_2-$ and $=CH-$	25.30	201.80	0.8636	0.0818	50.7	$-CH_2-$ and $-CHO$	88.49	648.70	−0.0267	−0.0210	106.1
$-CH_2-$ and $=C<$	27.34	242.47	0.1874	0.1010	24.0	$-CH_2-$ and $-COOH$	156.34	1180.39	−0.0932	−0.0253	114.0
$-CH_2-$ and $\equiv C-$	26.20	207.49	0.4160	−0.2199	33.9	$-CH_2-$ and $-COO-$ (o)	49.04	475.65	0.0276	0.1561	69.9
$-CH_2-$ and $>CH-$ (r)	14.60	238.81	−0.0149	−0.0265	31.9	$-CH_2-$ and $-COO-$ (c)	53.64	541.29	−0.0993	−0.2150	79.1
$-CH_2-$ and $>C<$ (r)	19.06	260.00	0.1193	−0.0423	52.1	$-CH_2-$ and $-NH_2-$	70.84	452.30	−0.0301	−0.0214	63.3
$-CH_2-$ and $=C<$ (r)	4.10	167.85	0.1012	−0.0626	—	$-CH_2-$ and $>NH-$	35.62	314.71	−0.0248	−0.0203	49.4
$-CH_2-$ and $-F$	6.27	166.59	−0.0255	−0.0161	49.3	$-CH_2-$ and $>N-$	18.11	240.08	−0.0161	−0.0170	32.7
$-CH_2-$ and $-Cl$	62.72	517.62	−0.0162	−0.0150	80.8	$-CH_2-$ and $-CN$	130.85	869.18	−0.0654	−0.0329	113.5
$-CH_2-$ and $-Br$	84.49	875.85	−0.0205	−0.0140	101.3	$-CH_2-$ and $-SH$	70.35	612.31	−0.0137	−0.0163	93.3
$-CH_2-$ and $-I$	107.75	1262.80	−0.0210	−0.0214	—	$-CH_2-$ and $-S-$	47.45	451.03	−0.0192	−0.0173	57.9
Interactions with nonring $>CH-$ (via single bond)											
$>CH-$ and $>CH-$	21.94	291.41	−0.0039	−0.0137	18.3	$>CH-$ and $-OH$ (a)	84.70	585.99	−0.0757	−0.0093	45.6
$>CH-$ and $>C<$	31.03	344.06	0.0025	−0.0085	8.6	$>CH-$ and $-O-$	14.40	215.94	−0.0162	−0.0155	23.7
$>CH-$ and $=CH-$	14.44	179.96	0.8547	0.0816	48.9	$>CH-$ and $>CO$	45.66	434.45	−0.0194	−0.0112	39.3
$>CH-$ and $=C<$	33.24	249.10	0.1969	0.1080	4.3	$>CH-$ and $-CHO$	78.46	630.07	−0.0406	−0.0280	92.2
$>CH-$ and $>CH-$ (r)	21.15	295.33	0.0025	−0.0168	—	$>CH-$ and $-COOH$	170.37	1270.16	−0.1054	−0.0358	110.2
$>CH-$ and $=C<$ (r)	−5.51	132.66	0.1187	−0.0556	—	$>CH-$ and $-COO-$ (c)	44.23	497.23	−0.0918	−0.2098	72.3
$>CH-$ and $-F$	−2.06	68.80	−0.0200	−0.0147	37.7	$>CH-$ and $-NH_2$	47.06	388.44	−0.0286	−0.0212	39.2
$>CH-$ and $-Cl$	47.08	438.47	−0.0142	−0.0131	68.6	$>CH-$ and $>NH$	22.34	260.32	−0.0158	−0.0162	—
Interactions with nonring $>C<$ (via single bond)											
$>C<$ and $>C<$	46.38	411.56	0.0084	0.0002	22.7	$>C<$ and $-Cl$	33.83	360.79	−0.0133	−0.0129	63.7
$>C<$ and $=CH-$	23.36	286.30	0.8767	0.0953	23.4	$>C<$ and $-Br$	50.42	610.26	−0.0084	−0.0121	85.7
$>C<$ and $=C<$	41.20	286.42	0.2061	0.1109	8.8	$>C<$ and $-OH$ (a)	76.39	540.38	−0.0780	−0.0094	40.6
$>C<$ and $>CH-$ (r)	25.89	340.00	0.0049	−0.0111	—	$>C<$ and $-O-$	23.46	267.26	−0.0156	−0.0103	40.8
$>C<$ and $=C<$ (r)	2.61	188.99	0.1249	−0.0510	—	$>C<$ and $>CO$	38.63	373.71	−0.0114	−0.0085	62.1
$>C<$ and $-F$	−7.70	−16.64	−0.0176	−0.0161	30.0	$>C<$ and $-COOH$	164.43	1336.54	−0.1008	−0.0455	89.0
Interactions with $=CH_2$ (via double bond)											
$=CH_2$ and $=CH_2$	−35.36	51.13	−0.9129	−0.0476	105.3	$=CH_2$ and $=C<$	13.11	215.27	−0.4158	−0.2709	99.2
$=CH_2$ and $=CH-$	28.66	205.73	−0.8933	−0.1378	77.4	$=CH_2$ and $=C=$	17.02	183.55	−0.0123	−0.0239	68.4
Interactions with nonring $=CH-$ (via double bond)											
$=CH-$ and $=CH-$	35.33	334.64	−1.7660	−0.2291	47.8	$=CH-$ and $-Cl$	42.28	370.60	0.8613	0.0919	68.7
$=CH-$ and $=C<$	36.03	354.41	−1.2909	−0.3613	73.6	$=CH-$ and $-O-$	14.95	204.81	0.8565	0.0947	36.4
$=CH-$ and $=C=$	40.99	316.46	−0.8945	−0.1202	43.6	$=CH-$ and $-CHO$	92.68	658.53	0.8246	0.0801	—
$=CH-$ and $=CH-$	20.63	174.18	1.7377	0.1944	42.1	$=CH-$ and $-COOH$	180.68	1245.86	0.7862	0.0806	107.4
$=CH-$ and $=C<$	36.87	228.38	1.0731	0.2146	16.6	$=CH-$ and $-COO-$ (o)	44.27	423.86	0.8818	0.2743	55.2
$=CH-$ and $\equiv C-$	22.58	174.39	1.2865	−0.1087	26.3	$=CH-$ and $-COO-$ (c)	59.38	525.35	0.7780	−0.1007	64.1
$=CH-$ and $=C<$ (r)	18.08	184.20	0.9929	0.0533	—	$=CH-$ and $-CN$	117.18	761.36	0.8122	0.0771	107.4
$=CH-$ and $-F$	−32.32	5.57	0.8623	0.0929	41.4						
Interactions with nonring $=C<$ (via double bond)											
$=C<$ and $=C<$	45.70	399.58	−0.8155	−0.4920	93.7	$=C<$ and $=C=$	44.51	321.02	−0.4009	−0.2502	58.1

*Contributions to structurally dependent parameter (Σ) of estimation equation (Eq.1).

(Table continued)

(a) −OH in a nonaromatic compound.
(p) −OH in an aromatic compound.
(c) Interaction with ester group via carbon atom.
(o) Interaction with ester group via oxygen atom.
(r) Ring group.
(rr) Ring group interacting with a group occurring in a different ring (heterocyclic compounds).
(—) Contribution that has not been calculated.

Table 5. Contributions of Group Interactions (Continued)

	Tb	*Tb**	*Tc*	*Pc*	*Vc*		*Tb*	*Tb**	*Tc*	*Pc*	*Vc*
Interactions with nonring =C< (via simple bond)											
=C< and =C< (r)	29.92	220.88	0.3043	0.0705	—	=C< and −Cl	36.54	367.05	0.1886	0.1102	43.3
=C< and −F	−13.78	−37.99	0.1868	0.1064	14.6						
Interactions with nonring =C= (via triple bond)											
=C= and =O	10.32	160.42	−0.0159	−0.0010	51.4						
Interactions with nonring ≡CH (via triple bond)											
≡CH and ≡CH	−16.26	120.85	−0.0288	−0.0226	87.6	≡CH and ≡C−	22.20	222.40	−0.4222	0.1860	73.1
Interactions with nonring ≡C− (via triple bond)											
≡C− and ≡C−	49.36	333.26	−0.7958	0.3933	64.3						
Interactions with ring $-CH_2-$ (via single bond)											
$-CH_2-$ (r) and $-CH_2-$ (r)	25.62	201.89	−0.0098	−0.0221	47.2	$-CH_2-$ (r) and −O− (r)	29.60	225.52	−0.0092	−0.0119	29.2
$-CH_2-$ (r) and >CH− (r)	21.77	209.40	−0.0093	−0.0181	47.5	$-CH_2-$ (r) and >CO (r)	61.01	451.74	−0.0148	−0.0177	50.7
$-CH_2-$ (r) and >C< (r)	20.34	182.74	−0.1386	0.0081	49.9	$-CH_2-$ (r) and >NH (r)	39.47	283.55	−0.0139	−0.0127	38.8
$-CH_2-$ (r) and =CH− (r)	31.27	218.07	0.0976	−0.1034	42.5	$-CH_2-$ (r) and −S− (r)	56.34	424.13	−0.0071	—	—
$-CH_2-$ (r) and =C< (r)	9.91	106.21	0.1089	−0.0527	—						
Interactions with ring >CH− (via single bond)						*Interactions with ring >C< (via single bond)*					
>CH− (r) and >CH− (r)	19.23	210.66	−0.0055	−0.0088	33.9	>C< (r) and >C< (r)	20.52	348.23	0.0329	0.0057	36.2
>CH− (r) and >C< (r)	22.71	220.24	−0.1341	0.0162	—	>C< (r) and =C< (r)	−3.11	−25.81	−0.0038	0.0072	—
>CH− (r) and −O− (r)	22.17	169.17	−0.0218	−0.0091	19.2	>C< (r) and >C< (rr)	33.67	550.72	0.0662	−0.0509	—
>CH− (r) and >CH− (rr)	20.17	242.01	−0.0059	−0.0071	—	>C< (r) and =C< (rr)	16.16	408.64	0.1615	0.1542	53.9
>CH− (r) and −OH (a)	81.38	597.82	−0.0737	−0.0220	—	>C< (r) and −F	−0.85	−41.35	−0.0314	−0.0119	18.4
Interactions with ring =CH− (via double bond)											
=CH− (r) and =CH− (r)	8.41	112.0	−0.2246	0.1490	36.5	=CH− (r) and =N− (r)	30.44	221.55	0.3913	0.1356	8.3
=CH− (r) and =C< (r)	36.01	291.15	−0.3586	−0.1822	34.4						
Interactions with ring =CH− (via single bond)											
=CH− (r) and =CH− (r)	44.57	285.07	0.2089	−0.1324	39.3	=CH− (r) and >NH (r)	68.48	420.54	0.0947	—	—
=CH− (r) and =C< (r)	24.95	237.22	0.2190	—	29.8	=CH− (r) and =N− (r)	49.83	321.44	−0.4067	−0.1491	65.9
=CH− (r) and −O− (r)	22.15	171.59	0.0990	−0.0935	40.3	=CH− (r) and −S− (r)	45.58	348.00	0.1027	−0.0916	40.8
Interactions with ring =C< (via double bond)											
=C< (r) and =C< (r)	66.09	477.77	−0.4848	0.1432	37.8	=C< (r) and =N− (r)	43.35	334.09	0.2541	—	—
Interactions with ring =C< (via single bond)											
=C< (r) and =C< (r)	1.99	180.07	0.2318	−0.0809	20.6	=C< (r) and −O−	3.66	199.70	0.0997	−0.0523	—
=C< (r) and −O− (r)	14.56	134.23	0.1104	−0.0374	−0.3	=C< (r) and >CO	38.88	437.51	0.1112	−0.0528	46.3
=C< (r) and =N− (r)	16.03	174.31	−0.3972	−0.0971	−93.2	=C< (r) and −CHO	92.60	700.06	0.0919	−0.0597	—
=C< (r) and =C< (rr)	−32.07	153.05	0.2424	−0.0792	51.7	=C< (r) and −COOH	151.44	1232.55	0.0313	−0.0684	100.2
=C< (r) and −F	−8.96	−48.79	0.1069	−0.0504	23.7	=C< (r) and −COO− (c)	23.85	437.78	0.0241	−0.2573	55.2
=C< (r) and −Cl	30.76	347.33	0.1028	−0.0512	60.3	=C< (r) and $-NH_2$	77.47	517.75	0.0830	−0.0579	33.2
=C< (r) and −Br	51.77	716.23	0.1060	−0.0548	83.2	=C< (r) and >NH	40.53	411.29	0.0978	−0.0471	—
=C< (r) and −I	90.04	1294.98	0.1075	−0.0514	110.2	=C< (r) and >N−	48.18	422.51	0.0938	−0.0462	—
=C< (r) and −OH (p)	64.74	456.25	0.0931	−0.0388	8.5	=C< (r) and −CN	92.74	682.19	0.0768	−0.0625	—
Interactions with −Cl (via single bond)											
−Cl and >CO	54.79	532.24	−0.0191	−0.0125	84.0						
Interactions with −O− (via single bond)											
−O− and >CO	42.16	367.83	0.1987	0.3209	—	−O− (r) and =N− (r)	57.78	382.25	—	—	—
Interactions with nonring >CO (via single bond)											
>CO and >CO	83.64	734.86	−0.4957	−0.6898	—						
Interactions with −H (forming formaldehyde, formic acid, ...)											
−H and −CHO	49.34	387.17	−0.0422	−0.0123	—	−H and −COO− (c)	44.47	298.12	−0.0781	−0.1878	51.2
−H and −COOH	169.14	1022.45	−0.0690	—	—						
Interactions with $-NH_2$ (via single bond)						*Interactions with nonring −S− (via single bond)*					
$-NH_2$ and >NH	115.75	673.59	−0.0301	—	—	−S− and −S−	61.17	597.59	−0.0124	—	—

*Contributions to structurally dependent parameter (Σ) of estimation equation (Eq.1).
(a) −OH in a nonaromatic compound.
(p) −OH in an aromatic compound.
(c) Interaction with ester group via carbon atom.
(o) Interaction with ester group via oxygen atom.
(r) Ring group.
(rr) Ring group interacting with a group occurring in a different ring (heterocyclic compounds).
(—) Contribution that has not been calculated.

Table 6. Absolute Average Errors: Training Set vs. Testing Set

		Training Set			Testing Set		
Eq.	Property	Data Points	AAE* (GC**)	AAE* (GIC†)	Data Points	AAE* (GC**)	AAE* (GIC†)
(1)	TB††	407	10.07	4.87	98	11.01	5.22
(2)	Tb	407	12.43	6.36	99	13.12	7.01
(3)	Tc	391	4.54	2.61	90	5.02	2.73
(4)	Pc	345	1.97	0.98	87	2.12	1.00
(5)	Vc	189	7.13	4.33	83	7.88	4.69

*AAE = average absolute error.
**The notation GC stands for group contribution.
†The notation GIC stands for group-interaction contribution.
††Equation 1 is the one proposed for estimating *Tb*.
Note: AAE has units of K for *Tb* and *Tc*, bar for *Pc*, and cm^3/mol for *Vc*.

Reliability of the Models

In order to test the reliability of the models, we generated two disjointed subsets by randomly excluding 100 compounds from the entire data set of 507 compounds. The largest subset was used as a training set for fitting the equations presented in Table 1 and determining the contributions of both simple groups and group interactions. The excluded compounds were used as a testing set by estimating their property values through the contributions and equations previously obtained using the training-set compounds.

The estimation errors for the testing-set compounds are comparable with those for the training-set compounds, as shown in Table 6, confirming the reliability of the proposed models. Notice that some compounds were removed from the testing set because they involved interactions whose contributions could not be obtained from the regression in which the randomly selected training set was used. Since the models presented perform well for compounds not included in the training set, it is more suitable to employ the contributions calculated by using the entire data set because of its larger size. These contributions are reported in Table 5.

Examples and Discussion

To illustrate our proposed application of the GIC approach, we have estimated the values of the normal boiling point and critical constants using 1,3,5-trichlorotrifluorobenzene, ethyl acrylate, isoquinoline, and *m*-terphenyl as example compounds. In Tables 7a, 7b, 7c, and 7d, we show the GIC summations for each property using the values from Table 5. In Tables 8a, 8b, 8c, and 8d, we summarize the results and compare them with experimental values.

Table 7a. Summations of Group-Interaction Contributions for 1,3,5-Trichlorotrifluorobenzene

Interactions		No.	*Tb*	*Tb**	*Tc*	*Pc*	*Vc*
−Cl and −F and ring =C<							
=C<	−Cl	3	30.76	347.33	0.1028	−0.0512	60.3
=C<	−F	3	−8.96	−48.79	0.1069	−0.0504	23.7
Interactions within the ring							
via double bond							
=C<	=C<	3	66.09	477.77	−0.4848	0.1432	37.8
via single bond							
=C<	=C<	3	1.99	180.07	0.2318	−0.0809	20.6
Structurally dependent parameter Σ							
			269.64	2869.14	−0.1299	−0.1179	427.2

*Contributions to structurally dependent parameter (Σ) of estimation equation (Eq. 1).

Table 7b. Summations of Group-Interaction Contributions for Ethyl Acrylate

Interactions		No.	*Tb*	*Tb**	*Tc*	*Pc*	*Vc*
Interaction with $=CH_2$ *(via double bond)*							
$=CH_2$	=CH−	1	28.66	205.73	−0.8933	−0.1378	77.4
Interaction with $-CH_2-$							
$-CH_3$	$-CH_2-$	1	33.19	194.25	−0.0227	−0.0430	88.6
Interactions with ester group							
via oxygen atom							
$-CH_2-$	−COO−	1	49.04	475.65	0.0276	0.1561	69.9
via carbon atom							
=CH−	−COO−	1	59.38	525.35	0.7780	−0.1007	64.1
Structurally dependent parameter Σ							
			170.27	1400.98	−0.1104	−0.1254	300.0

*Contributions to structurally dependent parameter (Σ) of estimation equation (Eq. 1).

Table 7c. Summations of Group-Interaction Contributions for Isoquinoline

Interactions		No.	*Tb*	*Tb**	*Tc*	*Pc*	*Vc*
Interactions within the rings (via double bonds)							
=CH−	=CH−	3	8.41	112.00	−0.2246	0.1490	36.5
=C<	=C<	1	66.09	477.77	−0.4848	0.1432	37.8
=CH−	=N−	1	30.44	221.55	0.3913	0.1356	8.3
Interactions within the ring (via single bond)							
=CH−	=N−	1	49.83	321.44	−0.4067	−0.1491	65.9
=CH−	=CH−	1	44.57	285.07	0.2089	−0.1324	39.3
=CH−	=C<	4	24.95	237.22	0.2190	n.d.	29.8
Structurally dependent parameter Σ							
			315.96	2590.71	−0.0891	0.4443	380.0

*Contributions to structurally dependent parameter (Σ) of estimation equation (Eq. 1).
n.d. = interaction that could not be determined because of lack of data.

Table 7d. Summations of Group-Interaction Contributions for *m*-Terphenyl

Interactions		No.	*Tb*	*Tb**	*Tc*	*Pc*	*Vc*
Interactions within the rings (via double bonds)							
=CH−	=CH−	5	8.41	112.00	−0.2246	0.1490	36.5
=CH−	=C<	4	36.01	291.15	−0.3586	−0.1822	34.4
Interactions within the ring (via single bonds)							
=C<	=C<	2	−32.07	153.05	0.2424	−0.0792	51.7
=CH−	=CH−	5	44.57	285.07	0.2089	−0.1324	39.3
=CH−	=C<	4	24.95	237.22	0.2190	n.d.	29.8
Structurally dependent parameter Σ							
			444.60	4404.93	−0.1521	−0.8042	739.20

*Contributions to structurally dependent parameter (Σ) of estimation equation (Eq. 1).
n.d. = interaction that could not be determined because of lack of data.

Table 8a. Results for 1,3,5-Trichlorotrifluorobenzene

Property	Eq. Used	Est. Value (GIC)*	Est. Value (GC)**	Exp. Value
Tb (K)	(1)†	471.9	475.8	471.5
Tb (K)	(2)	474.3	481.2	471.5
Tc (K)	(3)††	684.5	672.6	684.9
Pc (bar)	(4)‡	32.4	33.8	32.7
Vc (cm^3/mol)	(5)	452.3	443.0	448.0

*The notation GIC stands for group-interaction contribution.

**The notation GC stands for group contribution.

†The molecular weight (M) was 235.419 g/mol.

††The experimental value of *Tb* was employed in Eq. 4. If the value of *Tb* (471.9 K) estimated from Eq. 1 and group-interaction contributions had been used, a value of 685.0 K for *Tc* would have been obtained.

‡n_A, total number of atoms, is 12.

Note: Summation of contributions from Table 7a. Experimental values were obtained from Reid et al. (1987).

Table 8c. Results for Isoquinoline

Property	Eq. Used	Est. Value (GIC)*	Est. Value (GC)**	Exp. Value
Tb (K)	(1)†	591.5	501.0	516.4
Tb (K)	(2)	520.6	502.4	516.4
Tc (K)	(3)††	782.5	775.4	803.0
Pc (bar)	(4)‡	n.e.	45.2	n.a.
Vc (cm^3/mol)	(5)	405.1	355.1	n.a.

*The notation GIC stands for group-interaction contribution.

**The notation GC stands for group contribution.

†The molecular weight (M) was 129.162 g/mol.

††The experimental value of *Tb* was employed in Eq. 4. If the value of *Tb* (519.5 K) estimated from Eq. 1 and group-interaction contributions had been used, a value of 787.2 K for *Tc* would have been obtained.

‡n_A, total number of atoms, is 17.

Note: Summation of contributions from Table 7c. n.e. = value that cannot be estimated because of lack of contribution. n.a. = experimental data not available. Experimental values were obtained from Reid et al. (1987).

We have used a common set of experimental data to obtain the contribution values of both simple groups and group interactions. It is worth mentioning that it would not have been a fair comparison between both structural techniques if we had not used a common set of data. As shown in Table 3, the results achieved by using group interactions are significantly better than those obtained with the typical group-contribution technique.

Another important feature of the GIC-based models presented is the ability to capture the fine differences among isomers. Consider, for example, the estimation of the normal boiling points of octane isomers. Table 9 shows the estimations through the GIC and group-contribution approaches, using Eq. 1 as the estimation equation, and the group-contribution technique proposed by Constantinou and Gani (1994). Clearly, by using the GIC approach, a higher number of isomers is distinguished and the estimates are significantly more accurate.

The proposed GIC-based models were found to be quite satisfactory for most of the chemical classes (alkanes, alkenes, cycloalkanes, cycloalkenes, aromatic hydrocarbons, ethers, aldehydes, ketones, acids, esters, as well as nitrogen, sulfur, and halogenated compounds), and relatively poor for alcohols, phenols, and large heterocyclic compounds. All the models yield large errors for polyhydroxy alcohols. It is remarkable that the GIC-based models perform better than the group-contribution ones for all the classes.

Conclusion

A family of models for the estimation of normal boiling points and critical constants of pure organic compound has been developed by using the group-interaction contribution (GIC) approach. Compared to the classic group-contribution techniques, the proposed models exhibit a drastically superior accuracy and are better able to distinguish among isomers. In addition, a nonlinear equation for estimating the normal boiling point, making a significant improvement in accuracy, was reported.

Acknowledgment

Financial support of the Centro de Química Farmacéutica (CQF), Havana, is gratefully acknowledged.

Table 8b. Results for Ethyl Acrylate

Property	Eq. Used	Est. Value (GIC)*	Est. Value (GC)**	Exp. Value
Tb (K)	(1)†	373.9	366.5	373.0
Tb (K)	(2)	374.9	363.6	373.0
Tc (K)	(3)††	552.2	555.2	552.0
Pc (bar)	(4)‡	36.6	36.0	37.4
Vc (cm^3/mol)	(5)	325.1	328.1	320.0

*The notation GIC stands for group-interaction contribution.

**The notation GC stands for group contribution.

†The molecular weight (M) was 100.119 g/mol.

††The experimental value of *Tb* was employed in Eq. 4. If the value of *Tb* (373.9 K) estimated from Eq. 1 and group-interaction contributions had been used, a value of 553.6 K for *Tc* would have been obtained.

‡n_A, total number of atoms, is 15.

Note: Summation of contributions from Table 7b. Experimental values were obtained from Reid et al. (1987).

Table 8d. Results for *m*-Terphenyl

Property	Eq. Used	Est. Value (GIC)*	Est. Value (GC)**	Exp. Value
Tb (K)	(1)†	645.3	652.3	638.0
Tb (K)	(2)	649.3	697.2	638.0
Tc (K)	(3)††	907.3	893.5	924.9
Pc (bar)	(4)‡	n.e.	27.1	35.1
Vc (cm^3/mol)	(5)	764.3	732.1	768.0

*The notation GIC stands for group-interaction contribution.

**The notation GC stands for group contribution.

†The molecular weight (M) was 230.31 g/mol.

††The experimental value of *Tb* was employed in Eq. 4. If the value of *Tb* (645.3 K) estimated from Eq. 1 and group-interaction contributions had been used, a value of 917.7 K for *Tc* would have been obtained.

‡n_A, total number of atoms, is 32.

Note: Summation of contributions from Table 7d. n.e. = value that cannot be estimated because of lack of contribution. Experimental values were obtained from Reid et al. (1987).

Table 9. Normal Boiling Point Estimates of Octane Isomers

Isomeric Compounds	Exp. Values	GIC* Estimates Using Eq. 1		GC** Estimates Using Eq. 1		GC** Estimates (Constantinou and Gani, 1994)	
n-Octane	398.8	394.0	(4.8)	394.5	(4.3)	406.6	(7.8)
2-Methylheptane	390.8	386.4	(4.4)	388.5††	(2.3)	393.1	(2.3)
3-Methylheptane	392.1	386.3†	(5.8)	388.5††	(3.6)	396.5‡	(4.4)
4-Methylheptane	390.9	386.3†	(4.6)	388.5††	(2.4)	396.5‡	(5.6)
2,2-Dimethylhexane	380.0	379.8	(0.2)	382.2††	(2.2)	384.5	(4.5)
2,3-Dimethylhexane	388.8	385.8	(3.0)	382.5††	(6.3)	391.4‡	(2.6)
2,4-Dimethylhexane	382.6	378.7	(3.9)	382.5††	(0.1)	382.3‡	(0.3)
2,5-Dimethylhexane	382.3	378.9	(3.4)	382.5††	(0.2)	378.6	(3.7)
3,3-Dimethylhexane	385.1	385.1	(0.0)	382.2††	(2.9)	386.0‡	(0.9)
3,4-Dimethylhexane	390.9	385.7†	(5.2)	382.5††	(8.4)	391.4‡	(0.5)
3-Ethylhexane	391.7	386.2	(5.5)	388.5††	(3.2)	396.5‡	(4.8)
2,2,3-Trimethylpentane	383.0	382.6	(0.4)	376.2‡‡	(6.8)	383.5	(0.5)
2,2,4-Trimethylpentane	372.4	372.2	(0.2)	376.2††	(3.8)	369.5	(3.0)
2,3,3-Trimethylpentane	387.9	388.1	(0.2)	376.2††	(11.7)	385.0	(2.9)
2,3,4-Trimethylpentane	386.6	385.3	(1.3)	376.6	(10.0)	386.1	(0.5)
2-Methyl-3-ethylpentane	388.8	385.7†	(3.1)	382.5††	(6.3)	382.3‡	(6.5)
3-Methyl-3-ethylpentane	391.4	390.5	(0.9)	382.2††	(9.2)	386.0‡	(5.4)
2,2,3,3-Tetramethylbutane	379.6	381.8	(2.2)	369.9	(9.7)	383.8	(4.2)
		AAE‡‡	2.7	AAE‡‡	5.2	AAE‡‡	3.3

*GIC = Group-interaction contribution.

**GC = Group contribution.

†Isomeric compounds that cannot be distinguished by the GIC approach.

††Isomeric compounds that cannot be distinguished by the GC approach.

‡Isomeric compounds undistinguished by the technique proposed by Constantinou and Gani.

‡‡AAE = Absolute average error.

Note: Experimental values, estimates, and errors have units of K. Experimental values were obtained from Reid et al. (1987).

Literature Cited

Ambrose, D., "Correlation and Estimation of Vapor-Liquid Critical Properties: I. Critical Temperatures of Organic Compounds," NPL Rep. Chem., Nat. Physical Lab., Teddington, UK (1978).

Constantinou, L., S. E. Prickett, and M. L. Mavrovouniotis, "Estimation of Thermodynamic and Physical Properties of Acyclic Hydrocarbons Using the ABC Approach and Conjugation Operators," *Ind. Eng. Chem. Res.*, **32**, 1734 (1993).

Constantinou, L., S. E. Prickett, and M. L. Mavrovouniotis, "Estimation of Properties of Acyclic Organic Compounds Using Conjugation Operators," *Ind. Eng. Chem. Res.*, **39**, 395 (1994).

Constantinou, L., and R. Gani, "New Group Contribution Method for Estimating Properties of Pure Compounds," *AIChE J.*, **40**, 1697 (1994).

Forsythe, G. E., M. A. Malcolm, and C. B. Moler, *Computer Methods for Mathematical Computations*, Prentice Hall, Englewood Cliffs, NJ (1977).

Joback, K. G., and R. C. Reid, "Estimation of Pure-Component Properties from Group-Contributions," *Chem. Eng. Commun.*, **57**, 233 (1987).

Klincewicz, K. M., and R. C. Reid, "Estimation of Critical Properties with Group-Contribution Methods," *AIChE J.*, **30**, 137 (1984).

Lydersen, A. L., "Estimation of Critical Properties of Organic Compounds," Rep., Univ. Cool. Exp. Stn., Madison, WI (1955).

Lyman, W. J., W. F. Reehl, and D. H. Rosenblatt, *Handbook of Chemical Property Estimation Methods*, American Chemical Society, Washington, DC (1990).

Pardillo-Fontdevila, E., and R. González-Rubio, "A Group-Interaction Contribution Approach. A New Strategy for the Estimation of Physico-Chemical Properties of Branched Isomers," *Chem. Eng. Commun.*, **163**, 245 (1997).

Prickett, S. E., L. Constantinou, and M. L. Mavrovouniotis, "Computational Identification of Conjugate Paths for Estimation of Properties of Organic Compounds," *Mol. Simulation*, **11**, 205 (1993).

Reid, R. C., J. M. Prausnitz, and B. E. Poling, *The Properties of Gases and Liquids*, 4th ed., McGraw-Hill, New York (1987).

Manuscript received May 15, 1998, and revision received Nov. 17, 1998.