(3) Balaban, A. T. *MATCH* **1976**, *2*, 51.
(4) Balaban, A. T. *Rev. Roum. Chim.* **1981**, *26*, 407.
(5) Balasubramanian, K.; Kaufman, J. J.; Koski, W. S.; Balaban, A. T. *J. Comput. Chem.* **1980**, *1*, 149.
(6) Knop, J. V.; Szymanski, K.; Jeričević, Ž.; Trinajstić, N. *J. Comput. Chem.* **1983**, *4*, 23.
(7) Trinajstić, N.; Jeričević, Ž.; Knop, J. V.; Müller, W. R.; Szymanski, K. *Pure Appl. Chem.* **1983**, *55*, 379.
(8) Knop, J. V.; Szymanski, K.; Jeričević, Ž.; Trinajstić, N. *MATCH* **1984**, *16*, 119.
(9) Knop, J. V.; Müller, W. R.; Szymanski, K.; Trinajstić, N. *Computer Generation of Certain Classes of Molecules*; Association of Chemists and Technologists of Croatia: Zagreb, 1985.
(10) He, W.; He, W. *Theor. Chim. Acta* **1985**, *68*, 301.
(11) Brunvoll, J.; Cyvin, S. J.; Cyvin, B. N. *J. Comput. Chem.*, in press.
(12) Doroslovački, R.; Tošić, R., private communication, Novi Sad, Yugoslavia.
(13) Polansky, O. E.; Rouvray, D. H. *MATCH* **1976**, *2*, 63.
(14) Polansky, O. E.; Rouvray, D. H. *MATCH* **1977**, *3*, 97.
(15) Vögtle, F.; Staab, H. A. *Chem. Ber.* **1968**, *101*, 2709.
(16) Jenny, W.; Baumgarten, P.; Paioni, R. *Proceedings of the Symposium on the Nonbenzenoid Aromatic Compounds*; Sendai, Japan, 1970; p 183.
(17) Diederich, F.; Staab, H. A. *Angew. Chem.* **1978**, *90*, 383.
(18) Staab, H. A.; Diederich, F. *Chem. Ber.* **1983**, *116*, 3487.
(19) Vogler, H. *THEOCHEM* **1985**, *122*, 333.
(20) Jenny, H.; Peter, R. *Angew. Chem.* **1965**, *77*, 1027.
(21) A large selection of names has been suggested for this concept: hexagonal animal, hexanimal, hexagonal polyomino, polyhex, PAH-6 (PAH = polycyclic aromatic hydrocarbon), fusene (catafusene, perifusene). We are using the term benzenoid, but no standard terminology seems to prevail at present.
(22) Professor N. Trinajstić has informed us privately that the wrong number (48) is due to a typing error in ref 8. It was unfortunately repeated in ref 9. The error has been noticed by several researchers.
(23) Gutman, I. *Bull. Soc. Chim., Beograd* **1982**, *47*, 453.
(24) Smith, F. T. *J. Chem. Phys.* **1961**, *34*, 793.
(25) Gordon, M.; Davison, W. H. T. *J. Chem. Phys.* **1952**, *20*, 428.
(26) Hall, G. G. *Proc. R. Soc. London, A* **1955**, *229*, 251.
(27) Gutman, I.; Cyvin, S. J. *THEOCHEM* **1986**, *138*, 325.
(28) Coulson, C. A.; Longuet-Higgins, H. C. *Proc. R. Soc. London, A* **1947**, *129*, 16.
(29) Longuet-Higgins, H. C. *J. Chem. Phys.* **1950**, *18*, 265.
(30) Balaban, A. T.; Tomescu, I. *MATCH* **1983**, *14*, 155.
(31) Cyvin, S. J.; Gutman, I. *J. Serb. Chem. Soc.* **1985**, *50*, 443.
(32) Gutman, I. *Croat. Chem. Acta* **1974**, *46*, 209.
(33) Dias, J. R. *J. Chem. Inf. Comput. Sci.* **1984**, *24*, 124.
(34) Randić, M. *J. Chem. Soc., Faraday Trans. 2* **1976**, *72*, 232.
(35) Cyvin, S. J.; Bergan, J. L.; Cyvin, B. N. *Acta Chim. Hung.*, in press.
(36) Brown, R. L. *J. Comput. Chem.* **1983**, *4*, 556.

Atomic Physicochemical Parameters for Three-Dimensional-Structure-Directed Quantitative Structure–Activity Relationships. 2. Modeling Dispersive and Hydrophobic Interactions

ARUP K. GHOSE* and GORDON M. CRIPPEN*

College of Pharmacy, University of Michigan, Ann Arbor, Michigan 48109

Received July 22, 1986

In an earlier paper (Ghose A. K.; Crippen, G. M. *J. Comput. Chem.* **1986**, *7*, 565) the need of atomic physicochemical properties for three-dimensional-structure-directed quantitative structure–activity relationships was demonstrated, and it was shown how atomic parameters can be developed to successfully evaluate the molecular water–1-octanol partition coefficient, which is a measure of hydrophobicity. In the present work the atomic values of molar refractivity are reported. Carbon, hydrogen, oxygen, nitrogen, sulfur, and halogens are divided into 110 atom types of which 93 atomic values are evaluated from 504 molecules by using a constrained least-squares technique. These values gave a standard deviation of 1.269 and a correlation coefficient of 0.994. The parameters were used to predict the molar refractivities of 78 compounds. The predicted values have a standard deviation of 1.614 and a correlation coefficient of 0.994. The degree of closeness of the linear relationship between the atomic water–1-octanol partition coefficients and molar refractivities has been checked by the correlation coefficient of 89 atom types used for both the properties. The correlation coefficient has been found to be 0.322. The low value suggests that both parameters can be used to model the intermolecular interaction. The origin of these physicochemical properties and the types of interaction that can be modeled by these properties have been critically analyzed.

INTRODUCTION

In the process of drug design, medicinal chemists evaluate the binding energy of some closely related ligands with a biological receptor. The explicit structure of the receptor in most cases is unknown. The ultimate objective of any quantitative structure–activity relationship (QSAR) is to portray the receptor by the structural, physicochemical, and biological properties of the ligand. Not only is the task difficult but the inherent weakness of the approach ought to make the portrait misty. Explanation of the simplest biological data, namely, the binding energy of the ligand on the purified receptor, involves (1) the three-dimensional structure of the biological receptor[1] and its conformational flexibility,[2] (2) knowledge of the active site,[1] (3) the conformational behavior of the ligand,[3,4] (4) the interaction of the biophase[5] with the ligand/receptor, and, most important, (5) the interaction of the ligand with the receptor. Each process has its energetic (enthalpic) and entropic contribution. The energetic contribution often is easier to model than the entropic part. Entropy is related to the flexibility of the ligand and the receptor as well as the structural randomness of the biophase around the ligand and the receptor before and after binding. The complexity of these processes leads to very slow development along this line and urges some method that can allow us a rough estimate of the active site.

Most QSAR approaches therefore correlate the binding energy of the ligand with different physicochemical properties for different parts of the ligand. If these physicochemical properties represent the different types of molecular forces, one can guess the nature of interaction at different regions. The first problem is therefore to identify the possible types of forces in the biomolecular interaction and next to identify the physicochemical properties that can model these forces. Unlike the intermolecular interaction between simple molecules, the biochemical interaction of a drug involves a macromolecule on one side. The macromolecule is assumed to have low flexibility under physiological conditions, and hence the steric fit of the ligand structure at the active site often con-

0095-2338/87/1627-0021$01.50/0

stitutes a major factor. The flexibility in turn is a complex function of the intramolecular forces within the biomolecule and in the biophase. The interaction of the biophase with the ligand constitutes another important factor in the biochemical process. If the ligand is highly solvated and needs desolvation for the binding process, such binding ought to be weak unless it is compensated by strong interaction with the receptor. The interaction of the biophase with the ligand or the receptor is often governed by entropy rather than by enthalpy.[5,6] The inert gases and simple hydrocarbons are only slightly soluble in water, although they have a favorable (negative) enthalpy of solution. The negative enthalpy comes from two sources, the dispersive force between the solute and the solvent and structuring of the water around the solute. It is the latter factor that gives unfavorable (negative) entropy. Both enthalpic and energetic factors are responsible for the hydrophobic interaction. The term *hydrophobic interaction* refers to the force or the corresponding energy that operates between two or more nonpolar solutes in liquid water. Although the theoretical work on hydrophobic interactions led to a clear understanding of the molecular structure of aqueous solution, it has hardly begun to build a satisfactory theoretical description of the process that has a wide range of practical applicability. In such a situation, medicinal chemists try to model this interaction using a physicochemical property that closely parallels the hydrophobicity. They use the partition coefficient of the ligand molecules between water and a nonpolar solvent (usually 1-octanol) as a measure of hydrophobicity. This property, in fact, represents nonspecific dispersive and electrostatic forces and the consequent entropic factor. However, biological interaction has some regiospecific dispersive and electrostatic forces and thus urges the use of some physicochemical properties that can handle these forces. The formal charge density[4] on the atoms or the electrostatic potential near the van der Waals surface is a good measure of the electrostatic forces. Since the primary objective of this paper is to develop parameters that can be used to model the dispersive interaction of the ligand at the receptor site, we shall consider this interaction in greater detail in what follows.

THEORY OF DISPERSIVE FORCE AND ATOMIC REFRACTIVITY

London first showed that the attractive force between nonpolar molecules is due to correlation of the electron motion. It is therefore known as London forces or dispersive forces.[7,8] An accurate quantum chemical treatment of the process is very difficult.[9] Since the polarizability is closely related to the dispersive force, all approximate formulas for the latter are obtained by replacing unevaluated terms by it, if they approximately represent polarizability. Thus, according to London, the dispersive interaction between two spherically symmetrical systems A and B is

$$E_L = \frac{-3\alpha_A\alpha_B}{2R^6}\frac{U_AU_B}{U_A + U_B} \tag{1}$$

where α is the polarizability and U is the approximate ionization energy. On the other hand, according to Slater–Kirkwood

$$E_L = \frac{-3\alpha_A\alpha_B}{2R^6[(\alpha_A/N_A)^{1/2} + (\alpha_B/N_B)^{1/2}]} \tag{2}$$

where N is an empirical parameter known as the effective number of electrons. Equations 1 and 2 are strictly applicable for spherically symmetrical systems and are not suitable for most molecular systems. However, Pitzer[7] first used this idea to calculate the intramolecular dispersion interaction. The dispersion energies were summed for all pairs of nonbonded atoms. The approximation of atom-pair dissection of the dispersive force ultimately led to the development of the molecular mechanics method for conformational analysis.[3] This method is found to be successful for evaluating intermolecular interactions.[10] Theoretical estimation of the ligand–receptor binding energy from the properties of the ligand is based on the idea that the properties of the ligand and the receptor can be separated

$$E_L = Kf(A)f(B) \tag{3}$$

where $f(A)$ and $f(B)$ represent functions characteristic of the ligand and the receptor, respectively. If the receptor is relatively rigid and the different ligands bind in the same region of the receptor, the distance R in eq 1 or 2 may be assumed to be constant. The other part (containing the ionization energy) can be separated in the form of eq 3 if in eq 1

$$\text{(a) } U_A \gg U_B \quad \text{or} \quad \text{(b) } U_A \ll U_B \quad \text{or} \quad \text{(c) } U_A \approx U_B \tag{4}$$

If atom-pair dissection of the dispersive interaction is accepted, then the interaction of a particular ligand atom with the receptor leads to different expressions under different conditions: under condition 4a

$$E_L = \frac{-3\alpha_A}{2}\left(\sum_i \frac{\alpha_{B_i}U_{B_i}}{R_i^6}\right) \tag{5}$$

Under the summation is over all the receptor atoms. The quantity within brackets in eq 5 for a particular receptor will be constant for a small specified region due to the distance factor. In other words, if the receptor is rigid, the proportionality constant of dispersive force with the polarizability of the ligand α_A will be different in different regions. This is why in our three-dimensional-structure-directed quantitative structure–activity relationships[11] the hypothetical site cavity is divided into small pockets of different types.
under condition 4b

$$E_L = \frac{-3\alpha_AU_A}{2}\left(\sum_i \frac{1}{R_i^6}\right) \tag{6}$$

Clearly here polarizability of the ligand alone cannot be used for modeling the dispersive interaction. Here the appropriate quantity is α_AU_A.

Under condition 4c the corresponding expression for dispersive interaction becomes half of (5) or (6).

On the other hand, the Slater–Kirkwood equation (eq 2) can be separated in the form of eq 3 if

$$\text{(a) } \frac{\alpha_A}{N_A} \gg \frac{\alpha_B}{N_B} \quad \text{or} \quad \text{(b) } \frac{\alpha_A}{N_A} \ll \frac{\alpha_B}{N_B} \quad \text{or} \quad \text{(c) } \frac{\alpha_A}{N_A} \approx \frac{\alpha_B}{N_B} \tag{7}$$

Under conditions 7a and 7b the expression for the total dispersive interaction of a ligand atom with the receptor becomes eq 8 and 9, respectively:

$$E_L = \frac{-3(\alpha_AN_A)^{1/2}}{2}\left(\sum_i \frac{(\alpha_BN_B)^{1/2}}{R_i^2}\right) \tag{8}$$

$$E_L = \frac{-3\alpha_A}{2}\left(\sum_i \frac{(\alpha_BN_B)^{1/2}}{R_i^2}\right) \tag{9}$$

Here also we reach similar conclusions that under certain conditions the dispersive force is a linear function of polarizability, and under some other condition it is a linear function of $(\alpha_AN_A)^{1/2}$.

Although assuming only one of the ionization energy conditions (4a–c) or one of the polarizability conditions (7a–c) for all receptor atoms may seem very crude, in practice it is not that bad, since in a particular region only those atoms of

the receptor which are close to the ligand atom will make major contributions to the dispersive interaction. However, it may be a good idea to make the dispersive interaction a linear function of both α_A and $(\alpha_A N_A)^{1/2}$.

The polarizability α of a substance is directly proportional to its molar refractivity, MR,[12] as

$$MR = 4\pi N\alpha/3 \quad (10)$$

where N is Avogadro's number. It is therefore obvious that for a small region in the hypothetical receptor the dispersive interaction may be modeled as a linear function of the molar refractivity. The proportionality constant characterized by the receptor and the position should be adjusted so that it can represent the observed binding energies of the ligand with the receptor.

It can be deduced from electrostatics[12] that for a spherical molecule

$$\alpha = r^3 \quad (11)$$

where r is the radius of the molecule. Inserting eq 11 in eq 10, we see that molar refractivity is equal to the actual volume of the molecules in 1 mol. If this interpretation holds in general, then the atomic contribution to molar refractivity is the volume of the atom in the molecule. Such volume should be different from the isolated atomic volume due to (1) the effect of polarity of the bonds on the atomic volume and (2) the overlap of the electron clouds of the bonded atoms.

METHOD OF CALCULATION

Classification of the Atoms. In an earlier work,[13] we evaluated atomic hydrophobic parameters from water–octanol partition coefficients. That involved representing commonly occurring atomic states of carbon, hydrogen, oxygen, nitrogen, halogens, and sulfur in organic molecules by 110 atom types. Since the factors considered in classifying the atoms also affect the molar refractivity and the identical classification allows checking the correlation between the two properties, the atom classification was kept unaltered in this work also (Table I). This classification partly differentiates (1) the polarizing effect of the heteroatoms and (2) the effect of overlapping with non-hydrogen atoms. The classification, however, may be weak in differentiating the conjugation effects. The atoms thus classified cover most of the common neutral organic molecules containing the above-mentioned atoms. The classification may not completely cover all organic molecules and we are not overly concerned, since addition of atom types is always feasible. Since the constitutive factor of the property has been included (at least partly) by giving them different types, the evaluation of the individual atomic value is based on the idea that the sum of the atomic values (a_i) is the molecular value:

$$MR_{calcd} = \sum n_i a_i \quad (12)$$

Preparation of Data. The preparation of data involves two distinct steps: (1) collection of the molar refractivities of various compounds and (2) classification of the atoms according to their environment in the structure. Since in the atom classification a large number of atom types are used, it is necessary to have an even larger number of molecules in the data set to get a statistically significant result. However, classification of the atoms from a long list of atom types is extremely error prone. In order to keep the data accurate, the molecular structure (topology and bond type) was generated by a computer program CHEMSTRUC[13] using simple commands comparable to CAS ONLINE substructure generation. The correctness of the structure is checked by graphics, and the program has some other logical checks that assure the correctness of the structure even when visual aids fail to detect structure errors. Even then we feel that the best way to prepare absolutely error-free input data is to have the structures generated by more than one person and accept them if they are identical. However, in the present work such error checking was not done due to lack of resources. The structural information is kept in the Cambridge Crystallographic Data File format with minor modifications. Another program, CLASIF, uses this information to classify the atom types according to Table I.

Mathematics of Evaluation. Although the least-squares technique is the most standard procedure for fitting the data in an equation like eq 12, it cannot be used here. The physical concept of molar refractivity is the volume of the molecule or atom, which cannot have a negative value. In simple least-squares method such a condition cannot be maintained. Constrained least-squares fitting, however, is a special case of quadratic programming,[14,15] which has been used here. Another advantage of this method is that with some modification, quadratic programming can be used to confine the solution to any desired region of the solution space. This feature is sometimes helpful in confining the solution to a physically realistic region. For the present study the quadratic programming problem can be defined as follows:

minimize

$$F = \sum[MR_{calcd} - MR_{obsd}]^2 \quad (13)$$

where MR_{calcd} is given by eq 12,

subject to the constraints

$$a_i \geq l_i \qquad i = 1, 2, ..., n \quad (14)$$

where the a_i's are the atomic refractivities and the l_i's are the corresponding desired lower limits of the solution. It is important to note that this formulation of the problem becomes identical with least squares if the lower limits of the variables, as given by eq 14, are kept sufficiently low.

RESULTS AND DISCUSSION

The compounds used to evaluate the atomic refractivity are shown in Table II. The molar refractivity values were either obtained from the compilation of Vogel[16] or evaluated from the molecular weight, density, and refractive index values.[17] Some of the parameters were evaluated from a limited number of compounds due to the unavailability of molecules having that atom type. Getting a stable solution is a difficult problem when a large number of parameters are used in a fitting study. When the number of compounds was much lower, the solution for the different carbons was very unstable in the sense that adding more molecules resulted in substantially different fitted values. A relatively stable solution was obtained when the number of compounds was nearly 400. One hundred more compounds were added after this stage for even greater stability. The Lemke algorithm for quadratic programming[15] was used for the initial evaluation of the parameters; the resultant values were finally refined by using the pattern search technique.[18]

In order to explain the classification of the atoms, 10 selected molecules are presented with their skeletal structures and complete atom classification in Table III and Figure 1.

During this study we found some inconsistencies in the values of molar refractivities calculated from the data of the *CRC Handbook*.[17] Some of these compounds should be mentioned. For 2-chloroacetophenone (24 369) the refractive index (n_D) has been given as 1.685, which led to the molar refractivity of 48.90. This value was far from the calculated one, approximately 40.5, but the original reference of Beilstein ($B7^3$, 963) showed the refractive index to be 1.5404, which gave the molar refractivity to be 40.39. For 3,4-benzoisoxazole (25 151) the density and refractive index were given to be 1.8127 and 1.5845, respectively, which suggested the molar refractivity value to be 22.008, whereas the fitted value in most

Table I. Classification of Atoms and Their Contributions to Molar Refractivity and Hydrophobicity

type	description[a]	atomic refrac[b] I	atomic refrac[b] III	no. of compd	freq of use	partition coeff[c]
	C in					
1	:CH_3R, CH_4	1.0330	2.3000	251	399	−0.6232
2	:CH_2R_2	1.4336	2.3071	165	372	−0.3957
3	:CHR_3	2.0068	2.4926	32	34	−0.2821
4	:CR_4	1.8489	2.3000	11	11	0.2112
5	:CH_3X	2.4666	3.4006	67	97	−1.1423
6	:CH_2RX	2.6338	3.2624	173	252	−0.9557
7	:CH_2X_2	3.1274	3.6770	15	16	0.2041
8	:CHR_2X	2.7332	3.0137	40	45	−0.9679
9	:$CHRX_2$	2.7885	3.225	15	19	0.5335
10	:CHX_3	3.0075	3.2401	5	5	0.6684
11	:CR_3X	2.5823	2.6140	15	16	−1.1165
12	:CR_2X_2	2.7286	3.1488	16	25	1.0525
13	:CRX_3	2.1784	2.3010	13	16	0.5390
14	:CX_4	3.1677	3.3559	4	4	1.1390
15	:=CH_2	2.8557	3.5071	39	45	−0.2386
16	:=CHR	4.1009	4.4814	45	57	−0.0363
17	:=CR_2	3.7162	3.7781	7	7	0.3324
18	:=CHX	3.6247	3.6211	11	11	−0.3295
19	:=CRX	4.4024	4.4310	14	16	−0.4739
20	:=CX_2	1.9708	3.2000	8	12	−0.2407
21	:≡CH	3.1472	3.4161	12	13	0.1922
22	:≡CR, R=C=R	4.2943	4.3043	18	23	0.1517
23	:≡CX	3.2593	3.4905	7	7	
24	:R--CH--R	3.0745	3.4127	184	834	−0.0447
25	:R--CR--R	4.3404	4.3725	105	132	0.3301
26	:R--CX--R	3.7428	3.8182	98	144	−0.1244
27	:R--CH--X	1.3896	2.5001	14	16	0.0571
28	:R--CR--X	1.3044	2.5000	15	16	0.2218
29	:R--CX--X	1.6607	2.7967	4	4	−0.1456
30	:X--CH--X	1.0000	2.5000	1	1	−0.1179
31	:X--CR--X			0	0	0.6739
32	:X--CX--X	1.0001	2.5000	1	3	0.0740
33	:R--CH···X	3.1193	3.4372	19	22	0.0117
34	:R--CR···X	3.7755	3.4494	14	15	0.1381
35	:R--CX···X	2.7215	3.1048	4	5	−0.2710
36	:Al-CH=X	3.3750	3.8251	15	15	0.1671
37	:Ar-CH=X	4.0211	4.5401	15	15	−0.0909
38	:Al—C(=X)—Al	3.3265	3.7529	15	16	−0.6828
39	:Ar—C(=X)—R	3.8103	4.1288	8	8	−0.4963
40	:R—C(=X)—X					
	R—C≡X, X=C=X	2.3957	2.7938	112	138	−0.4672
41	X—C(=X)—X	1.8242	2.4165	13	13	−0.2992
42	:X--CH···X	2.0234	3.0606	6	6	−0.6503
43	:X--CR···X	1.0004	2.5001	5	5	−0.2786
44	:X--CX···X	1.0006	2.5001	2	2	−0.2992
45	unused					
	H attached to[d]					
46	:C_{sp3}^0	1.5903	1.1461	270	1695	0.4283
47	:C_{sp3}^1, C_{sp2}^0	1.1616	0.8000	389	1773	0.3607
48	:C_{sp3}^2, C_{sp2}^1 C_{sp}^0	1.0309	0.8006	69	95	−0.0069
49	:C_{sp3}^3, C_{sp2}^2 C_{sp2}^3, C_{sp}^3	1.0001	0.8001	55	61	−0.2539
50	:heteroatom	1.0001	0.8000	104	145	−0.4001
51	:α-C	1.4647	1.0026	113	330	0.2271
52–55	unused					
	O in					
56	:alcohol	1.1231	1.4430	21	23	0.0101
57	:phenol, enol carboxyl OH	1.2847	1.4090	30	35	0.5624
58	:=O	1.8485	1.6506	169	202	0.1230
59	:Al—O—Al	1.0000	1.2000	30	44	0.1017
60	:Al—O—Ar, Ar_2O R···O···R, R—O—C=X	1.8787	1.8434	115	145	0.3914
61[e]	:--O	1.0005	1.6001	21	45	1.8239
62–65	unused					
	N in					
66	:Al—NH_2	1.8396	2.5001	10	11	0.3562
67	:Al_2NH	1.5242	2.5001	9	9	0.3712
68	:Al_3N	2.1094	2.5377	8	8	0.4886
69	:Ar—NH_2, X—NH_2	3.1442	3.6195	9	12	0.5128
70	:Ar—NH—Al	2.6774	2.9832	8	8	1.1484
71	:Ar—NAl_2	3.8031	3.9733	10	11	0.7387
72	:RCO—N<, >N—X=X	2.8495	3.0059	16	17	0.2160

Table I (Continued)

type	description[a]	atomic refrac[b] I	atomic refrac[b] III	no. of compd	freq of use	partition coeff[c]
73	:Ar_2NH, Ar_3N					
	Ar_2N—Al, R···N···R[f]	2.4082	2.6295	7	7	0.4777
74	:R≡N, R=N–	3.3952	3.1464	27	29	0.1989
75	:R--N--R,[g] R--N--X	6.2666	4.5123	24	27	0.1605
76	:Ar—NO_2, R--N(--R)--O[h]					
	RO—NO_2	5.9990	4.7725	15	17	−3.1845
77	:Al—NO_2	3.9660	3.0389	6	6	−3.3406
78	:Ar—N=X, X—N=X	3.4136	3.6838	11	14	−0.1367
79–80	unused					
	F attached to					
81	:C_{sp3}^1	1.0001	0.8060	8	8	0.4929
82	:C_{sp3}^2	1.0001	0.8000	7	32	−0.1394
83	:C_{sp3}^3	1.4160	1.3484	7	28	0.1457
84	:C_{sp2}^1	1.0001	0.8000	21	34	0.6128
85	:C_{sp2}^{2-4}, C_{sp}^1					
	C_{sp}^4, X	2.2548	1.6440	6	12	0.4989
	Cl attached to					
86	:C_{sp3}^1	5.2233	5.3647	22	27	1.1021
87	:C_{sp3}^2	5.7784	5.6484	16	28	0.3333
88	:C_{sp3}^3	5.7328	5.6858	11	32	0.4402
89	:C_{sp2}^1	4.6108	5.0000	26	29	1.0372
90	:C_{sp2}^{2-4}, C_{sp}^1					
	C_{sp}^4, X	6.4057	5.9312	28	37	0.7220
	Br attached to					
91	:C_{sp3}^1	8.2314	8.3379	21	25	1.1263
92	:C_{sp3}^2	8.6483	8.5393	10	21	0.4640
93	:C_{sp3}^3	8.9016	8.8635	3	9	
94	:C_{sp2}^1	8.0271	8.0866	14	14	1.3343
95	:C_{sp2}^{2-4}, C_{sp}^1					
	C_{sp}^4, X	9.2260	9.0569	9	9	1.0137
	I attached to					
96	:C_{sp3}^1	13.5880	13.7535	7	8	1.4608
97	:C_{sp3}^2	13.6990	13.6306	4	7	
98	:C_{sp3}^3	13.4388	13.4586	3	3	
99	:C_{sp2}^1	12.8225	12.8876	5	5	1.8362
100	:C_{sp2}^{2-4}, C_{sp}^1					
	C_{sp}^4, X	13.6716	13.5530	1	1	1.0859
101–105	unused halogens					
	S in					
106	:R—SH	7.4314	7.7751	9	10	1.1181
107	:R_2S, RS—SR	7.5003	7.3151	18	20	1.0769
108	:R=S	9.4004	9.2916	6	7	0.3726
109	:R—SO—R	4.6036	5.3957	7	7	−0.5594
110	:R—SO_2—R	4.4935	5.4662	8	8	−0.6864

[a] R represents any group linked through carbon; X represents any heteroatom (O, N, S, and halogens); Al and Ar represent the aliphatic and aromatic groups, respectively; -- represents aromatic bonds as in benzene or delocalized bonds as the N–O bond in nitro group; ··· represents aromatic single bond as the C–N bond in pyrrole. [b] Atomic refractivity of only one atom. [c] A different data set was used to evaluate these values. The data set here is similar to the one reported earlier (ref 11) with two additional compounds, 2-methylbenzoimidazole and phenylacetaldehyde. [d] The subscript represents hybridization and the superscript its formal oxidation number. [e] As in nitro, =*N*-oxides. [f] Pyrrole type structure. [g] Pyridine type structure. [h] Pyridine-*N*-oxide type.

studies was approximately 34. Beilstein (B27[2], 17) showed the density and refractive index to be 1.1866 and 1.5789, respectively. These values suggested a molar refractivity of 33.36. For benzylidene dibromide (27 296) the density and refractive index are 1.51 and 1.6147, respectively, suggesting a molar refractivity of 57.743. The fitted value was much lower, near 47.5. In the original reference of Beilstein (B5[4], 836) these values are given as 1.8365 and 1.6106, respectively, leading to a value of 47.222. For thiazole (37 802) the density is given as 1.998, giving a molar refractivity of 14.515, while the fitted value was around 21. Another source[19] gave the density to be 1.1998, giving the molar refractivity of 24.192. Since in most other cases the agreement between the experimental and the fitted values was good enough, we did not try to check those values from Beilstein. Also, the *Handbook* reference of Beilstein did not always give the density and refractive index values, but in turn cited some other reference. There are two compounds for which we did not find any discrepancy in the reported density and refractive index values but still may be incorrect: *p*-chloro-*N*-methylaniline (24 937) and trichloro-(3-chlorophenyl)methane (27 078). When these values were corrected, the calculated values showed a standard deviation of 1.269, a correlation coefficient of 0.994, and an explained variance of 0.984. These parameters were used to predict the molar refractivity of 78 molecules listed in Table IV. The calculated values showed a standard deviation of 1.614 and a correlation coefficient of 0.994.

If we look at the atomic values of the various carbons (study I), we see that the saturated carbons have values around 2.5, lower than the roughly 3.5 for the ethylenic or acetylenic carbons. The effect of carbon substitution on these carbons usually goes through a maximum, as is indicated by the value of the subsets: carbon replacing hydrogen in a saturated carbon when no heteroatom is present, 1-1.0330, 2-1.4336, 3-2.0068, 4-1.8489 (here the first number indicates the atom type and the second one its refractivity; see Table I for the definition of the atom types); when one heteroatom is present, 5-2.4666, 6-2.6338, 8-2.7332, 11-2.5823; when two heteroatoms are present, 7-3.1274, 9-2.7885, 12-2.7286 (is one side of the peak missing here?; in the earlier subsets the value started declining at the fourth place); when three heteroatoms are present, 10-3.0075, 14-3.1677; in ethylenic carbon, 15-

Table II. Compounds Used to Evaluate the Atomic Refractivity

no.	ID[a]	compd	obsd	calcd from study		
				I	II	III
1	1001	methyl malonate	28.62	28.51	28.81	28.49
2	1002	methyl succinate	33.01	32.87	33.46	32.80
3	1004	methyl adipate	42.18	42.10	42.75	42.00
4	1008	ethyl malonate	37.89	38.13	38.10	38.09
5	1009	ethyl succinate	42.35	42.49	42.75	42.40
6	1011	ethyl adipate	51.51	51.72	52.04	51.60
7	1018	methyl dimethylmalonate	37.73	37.61	38.10	37.95
8	1022	methyl dipropylmalonate	56.07	56.06	56.69	56.35
9	1026	1,1-bis(methoxycarbonyl)cyclohexane	49.16	49.07	50.21	49.47
10	3001	cyclopentanone	23.31	23.13	23.47	23.23
11	3002	3-methylcyclopentanone	27.97	27.92	28.12	28.00
12	3003	cyclohexanone	27.87	27.74	28.12	27.83
13	3004	2-methylcyclohexanone	32.51	32.66	32.76	32.75
14	3005	3-methylcyclohexanone	32.65	32.53	32.76	32.60
15	3009	methylenecyclopentane	27.29	27.35	28.07	27.28
16	3010	methylenecyclohexane	32.15	31.97	32.72	31.88
17	3011	3-methylmethylenecyclohexane	37.19	36.75	37.37	36.66
18	3016	cyclopentene	22.40	24.37	23.43	24.36
19	3018	3-methylcyclopentanol	29.37	29.26	29.52	29.23
20	3021	cyclohexanol	29.16	29.09	29.52	29.05
21	3032	cycloheptanol	34.00	33.70	34.16	33.65
22	5001	acetone	16.11	16.03	16.01	16.02
23	5002	2-butanone	20.67	20.77	20.66	20.76
24	5007	2-hexanone	30.04	30.00	29.95	29.96
25	5008	4-methyl-2-pentanone	30.15	30.17	29.95	30.14
26	10117	toluene	31.10	31.32	31.19	31.17
27	10119	*n*-propylbenzene	40.42	40.55	40.49	40.37
28	10120	isopropylbenzene	40.39	40.73	40.49	40.55
29	11138	acetophenone	36.27	36.61	36.08	36.52
30	11139	propiophenone	40.83	41.35	40.72	41.27
31	12141	diethyl ether	22.51	22.52	22.05	22.40
32	12142	dipropyl ether	31.68	31.75	31.35	31.60
33	12143	diisopropyl ether	31.71	32.01	31.35	31.78
34	12150	methyl *n*-butyl ether	27.02	26.94	26.70	26.80
35	12156	2,2′-dichlorodiethyl ether	31.94	31.27	31.44	31.38
36	12159	phenyl methyl ether (anisole)	32.88	32.75	32.83	32.53
37	12161	*n*-propyl phenyl ether	42.28	42.18	42.12	41.93
38	12162	isopropyl phenyl ether	42.39	42.30	42.12	42.02
39	12166	allyl phenyl ether	41.73	42.20	42.31	41.98
40	12167	dimethoxymethane	19.20	19.09	19.04	19.28
41	12168	diethoxymethane	28.53	28.71	28.33	28.88
42	12177	1,1-dipropoxyethane	42.37	42.37	42.28	42.56
43	13180	ethyl formate	17.71	17.88	17.64	17.69
44	13181	*n*-propyl formate	22.41	22.50	22.29	22.29
45	13190	*n*-propyl acetate	26.95	26.93	26.94	26.80
46	13191	isopropyl acetate	26.96	27.05	26.94	26.89
47	13199	methyl propionate	22.14	22.24	22.29	22.14
48	13236	diethyl oxalate	33.56	33.77	33.46	33.78
49	13242	dimethyl succinate	32.99	32.87	33.46	32.80
50	13250	dimethyl adipate	42.20	42.10	42.75	42.00
51	13263	dimethyl methylmalonate	33.18	33.42	33.46	33.41
52	13325	chlorobenzene	31.14	29.53	31.24	29.88
53	14279	1,2-dichloroethane	21.00	20.36	20.51	20.45
54	14280	1,2-dichloropropane	25.69	25.10	25.16	25.14
55	14281	benzyl chloride	36.03	35.70	35.89	35.66
56	14282	1,3-dichloropropane	25.50	24.98	25.16	25.05
57	14283	methyl chloroacetate	22.34	22.86	22.34	22.72
58	14285	*n*-propyl chloroacetate	31.72	32.28	31.63	32.12
59	14287	1,2-dibromoethane	26.96	26.38	26.65	26.40
60	14288	1,2-dibromopropane	31.77	31.12	31.30	31.09
61	14289	1,3-dibromopropane	31.13	30.99	31.30	31.00
62	14290	*n*-propyl bromoacetate	34.57	35.29	34.70	35.09
63	14292	ethyl α-bromopropionate	34.35	35.12	34.70	34.98
64	14295	1-bromo-2-phenylethane	43.81	43.32	43.60	43.24
65	14296	ethyl 2-bromoethyl ether	29.41	29.91	29.82	29.86
66	14298	1,3-diiodopropane	41.51	41.70	41.69	41.83
67	14299	1-iodo-2-phenylethane	48.78	48.68	48.80	48.65
68	14300	propyl iodoacetate	39.72	40.65	39.90	40.51
69	14302	1-fluoropentane	24.99	25.60	25.21	25.20
70	14306	fluorobenzene	25.98	25.92	26.69	25.68
71	14307	4-fluorotoluene	30.74	31.83	31.34	31.58
72	14308	α-fluoronaphthalene	43.73	43.08	42.56	42.85
73	14309	4-chlorotoluene	35.99	35.44	35.89	35.78
74	14310	*m*-dichlorobenzene	36.16	33.65	35.94	34.49
75	14311	benzenesulfonyl fluoride	34.87	35.37	35.62	35.29

Table II (Continued)

no.	ID[a]	compd	obsd	calcd from study		
				I	II	III
76	14312	benzenesulfonyl chloride	41.03	39.52	40.16	39.58
77	15313	methyl benzoate	37.81	37.60	37.71	37.52
78	15315	*n*-propyl benzoate	47.22	47.02	47.01	46.92
79	15317	methyl phenylacetate	41.84	41.96	42.36	41.84
80	15318	ethyl phenylacetate	46.55	46.77	47.01	46.64
81	15325	bromobenzene	33.99	32.95	34.31	32.97
82	15327	iodobenzene	39.15	37.75	39.51	37.77
83	15328	α-methylnaphthalene	48.65	48.48	47.06	48.35
84	16331	vinylacetic acid	21.73	21.33	22.48	21.35
85	16332	methyl vinylacetate	26.30	26.88	27.13	26.79
86	16334	*n*-propyl vinyl acetate	35.65	36.30	36.42	36.19
87	16349	ethyl allylmalonate	51.27	52.30	52.24	52.26
88	16352	allyl acetate	26.39	26.95	27.13	26.85
89	16355	allyl succinate	50.85	51.77	52.43	51.70
90	16356	allyl chloride	20.42	20.62	20.66	20.62
91	16359	methyl maleate	33.18	34.67	33.65	34.74
92	16366	ethyl fumarate	43.20	44.29	42.94	44.34
93	17390	methyl but-3-yne-1-carboxylate	29.32	29.52	30.09	29.52
94	17391	ethyl but-3-yne-1-carboxylate	34.05	34.33	34.73	34.32
95	17398	dimethyl acetylenedicarboxylate	32.72	32.74	31.96	32.79
96	17399	diethyl acetylenedicarboxylate	42.22	42.36	41.25	42.39
97	17408	methyl cyanide	11.09	11.22	11.85	11.25
98	17409	ethyl cyanide	15.75	15.96	16.50	15.99
99	17415	allyl cyanide	19.67	20.60	21.34	20.64
100	17417	phenyl cyanide	31.58	31.31	31.92	31.38
101	17418	benzyl cyanide	35.22	35.67	36.57	35.69
102	18419	methyl cyclopropyl ketone	23.91	23.30	23.47	23.41
103	18420	cyclopropanecarboxylic acid	20.77	19.23	20.46	19.35
104	18421	methyl cyclopropanecarboxylate	25.34	24.77	25.11	24.78
105	18426	diethyl cyclopropane-1,1-dicarboxylate	45.60	44.85	45.57	45.28
106	18428	dimethyl cyclobutane-1,1-dicarboxylate	40.70	39.84	40.92	40.27
107	18432	cyclobutanecarboxylic acid	25.14	23.84	25.11	23.95
108	18433	methyl cyclobutanecarboxylate	29.71	29.39	29.75	29.38
109	19438	methyl cyclopentyl ether	29.42	29.30	29.52	29.21
110	19440	cyclopentyl formate	29.53	29.47	29.75	29.30
111	19441	cyclopentyl acetate	34.07	33.90	34.40	33.81
112	19442	cyclopentyl chloride	27.96	27.58	27.93	27.58
113	19444	cyclopentyl iodide	36.38	35.94	36.19	35.96
114	19445	dicyclohexyl	53.22	53.34	53.93	53.27
115	19446	methyl cyclohexyl ether	34.02	33.92	34.16	33.81
116	19450	cyclohexyl chloride	32.99	32.19	32.58	32.17
117	20453	methyl alcohol	8.22	8.07	8.11	8.04
118	20454	ethyl alcohol	12.90	12.88	12.76	12.84
119	20467	allyl alcohol	16.98	17.52	17.60	17.49
120	20468	2-methoxyethanol	19.18	18.99	19.04	18.97
121	20474	acetic acid	12.99	11.96	13.00	11.96
122	20475	propanoic acid	17.51	16.70	17.64	16.70
123	21483	ethanethiol	19.02	19.19	18.44	19.18
124	21484	propanethiol	23.71	23.81	23.09	23.78
125	21494	thiophenol	34.52	33.35	33.87	33.46
126	21495	methyl phenyl thioether	39.42	38.38	38.51	38.00
127	21496	ethyl phenyl thioether	44.19	43.18	43.16	42.80
128	22503	propylamine	19.45	19.22	19.70	19.30
129	22505	isobutylamine	23.98	24.00	24.34	24.08
130	22513	ethylenediamine	18.23	17.59	18.98	17.93
131	22514	aniline	30.56	30.07	30.47	30.10
132	22515	benzylamine	34.45	34.32	35.12	34.40
133	22517	diethylamine	24.30	24.05	24.34	24.50
134	22518	di-*n*-propylamine	33.51	33.27	33.64	33.70
135	22525	dicyclohexylamine	56.91	56.46	57.86	56.92
136	22526	ethyl *N*-methylcarbamate	25.73	26.11	26.22	26.12
137	22528	*N*-nitroso-*N*-methylaniline	39.97	38.99	39.81	39.02
138	22529	*N*-methylaniline	35.67	34.55	35.12	34.47
139	22535	tripropylamine	47.68	48.24	47.58	48.14
140	22541	*N*,*N*-dimethylaniline	40.81	40.63	39.77	40.46
141	23546	ethyl dichloroacetate	32.16	32.69	31.68	32.41
142	23549	methyl trichloroacetate	32.47	31.45	31.73	31.45
143	23553	dichloromethane	16.38	16.75	15.87	16.57
144	23554	dibromomethane	21.90	22.49	22.00	22.36
145	23555	diiodomethane	32.54	32.59	32.40	32.54
146	23557	1,1,2,2-tetrachloroethane	30.60	30.75	29.90	30.64
147	23558	chloroform	21.37	21.21	20.56	21.10
148	23559	methylchloroform	26.20	25.18	25.21	25.10
149	23560	carbon tetrachloride	26.45	26.10	25.26	26.10
150	23562	1,1,2,2-tetrabromoethane	41.97	42.23	42.18	42.21

Table II (Continued)

no.	ID[a]	compd	obsd	calcd from study		
				I	II	III
151	23 563	bromoform	29.86	30.71	29.77	30.63
152	23 565	ethyl orthoformate	39.30	39.29	39.26	39.44
153	23 566	propyl orthoformate	53.28	53.13	53.20	53.24
154	23 568	thionyl chloride	22.12	19.26	19.09	18.91
155	23 569	sulfuryl chloride	21.43	21.00	20.14	20.63
156	23 573	dimethyl-*N*-nitrosoamine	19.27	20.01	19.74	19.94
157	23 574	diethyl-*N*-nitrosoamine	28.43	29.63	29.03	29.54
158	23 577	nitromethane	12.36	12.83	13.35	12.65
159	23 578	nitroethane	17.02	17.33	18.00	17.24
160	23 584	nitrobenzene	32.38	32.92	33.42	32.85
161	23 585	*n*-butyl nitrite	26.87	27.13	26.74	26.98
162	23 588	ethyl nitrate	19.28	20.64	19.63	20.42
163	23 591	dimethyl carbonate	18.97	19.33	19.28	19.36
164	23 601	propyl xanthate	52.72	51.35	50.92	51.27
165	24 009	5-bromoacenaphthene	59.54	59.54	57.64	59.66
166	24 010	5-chloroacenaphthene	56.07	56.12	54.57	56.57
167	24 011	5-iodoacenaphthene	64.03	64.34	62.83	64.46
168	24 021	acetaldehyde	11.52	11.65	11.36	11.58
169	24 023	aminoacetaldehyde diethyl acetal	36.57	36.14	36.91	36.59
170	24 026	bromoacetaldehyde dimethyl acetal	29.89	30.91	31.45	31.23
171	24 036	acetaldehyde diethyl mercaptal	45.74	46.15	44.35	45.60
172	24 040	diphenylacetaldehyde	60.04	60.74	60.79	60.64
173	24 041	ethoxyacetaldehyde	22.46	23.55	22.29	23.34
174	24 042	hydroxyacetaldehyde	12.43	13.91	13.00	13.79
175	24 046	acetaldoxime	15.66	15.48	15.32	15.29
176	24 047	phenylacetaldehyde	35.88	36.11	36.08	36.02
177	24 051	tribromobenzaldehyde	35.74	35.11	34.65	35.17
178	24 057	trimethylacetaldehyde	25.13	25.48	25.30	25.79
179	24 058	acetamide	15.21	14.52	15.29	14.36
180	24 060	diacetylethylamine	34.46	32.95	34.11	33.11
181	24 067	*N*-acetyl-*N*-butylaniline	58.11	57.43	58.59	57.44
182	24 073	diethylacetamide	33.08	34.04	33.87	33.96
183	24 176	allyl acetate	26.45	26.95	27.13	26.85
184	24 177	acetic anhydride	22.37	21.22	22.53	21.35
185	24 178	trifluoroacetic anhydride	23.83	23.22	23.41	23.42
186	24 183	bromomethyl acetate	25.24	25.39	25.41	25.41
187	24 187	*sec*-butyl acetate	31.28	31.67	31.58	31.49
188	24 190	*tert*-butyl acetate	31.45	31.54	31.58	31.42
189	24 193	2-chloro-2-propyl acetate	32.19	31.67	31.63	31.87
190	24 306	acetone	16.18	16.03	16.01	16.02
191	24 307	acetone azine	36.17	35.15	34.83	35.03
192	24 309	bromoacetone	23.38	24.40	23.77	24.32
193	24 314	1,3-dichloroacetone	25.70	26.75	25.40	26.67
194	24 351	acetophenone	36.51	36.61	36.08	36.52
195	24 369	2-chloroacetophenone	40.39	40.72	40.77	41.13
196	24 370	3-chloroacetophenone	40.57	40.72	40.77	41.13
197	24 451	1-phenyl-1-propyne	40.05	39.91	38.99	39.78
198	24 480	acraldehyde	16.22	16.67	16.20	16.66
199	24 481	2-chloroacraldehyde	20.79	20.42	20.90	20.81
200	24 484	2-methylacraldehyde	20.94	20.92	20.85	20.90
201	24 486	acrylic acid	17.44	16.97	17.84	17.04
202	24 519	acrylyl chloride	21.18	21.09	20.90	20.76
203	24 770	2-bromoaniline	37.86	37.60	38.23	37.79
204	24 777	3-bromoaniline	38.56	37.60	38.23	37.79
205	24 791	*N*-butylaniline	49.26	48.59	49.06	48.46
206	24 797	2-*tert*-butylaniline	49.01	49.43	49.06	49.78
207	24 801	4-*tert*-butylaniline	49.01	49.43	49.06	49.78
208	24 834	*N*,*N*-dibutylaniline	68.92	68.71	67.65	68.45
209	24 844	*N*,*N*-diethylaniline	50.15	50.25	49.06	50.06
210	24 872	*N*,*N*-dimethylaniline	40.89	40.63	39.77	40.46
211	24 876	*N*,*N*-dimethyl-2-bromoaniline	47.97	48.16	47.53	48.15
212	24 879	2-chloro-*N*,*N*-dimethylaniline	45.32	44.75	44.46	45.06
213	24 883	2-nitro-*N*,*N*-dimethylaniline	48.86	48.14	46.64	48.04
214	24 887	2,3-dimethylaniline	39.94	41.88	39.77	41.90
215	24 937	4-chloro-*N*-methylaniline	29.34	38.67	39.81	39.07
216	24 941	*N*-methyl-*N*-nitrosoaniline	40.14	38.99	39.81	39.02
217	24 971	*N*-propylaniline	45.12	43.98	44.41	43.87
218	24 972	*N*-isobutylaniline	49.26	48.76	49.06	48.64
219	24 999	3-methoxybenzaldehyde	38.87	39.73	37.71	39.68
220	25 151	3,4-benzisooxazole	33.36	34.88	33.54	34.32
221	25 154	*tert*-butyl nitrite	26.77	27.14	26.74	27.01
222	25 326	antimalarine	91.20	88.40	87.42	88.66
223	25 327	antipyrine	57.44	58.19	56.23	58.52
224	25 338	2,5-dimethoxysaffrole	68.27	63.21	59.33	63.10

Table II (Continued)

no.	ID[a]	compd	obsd	calcd from study		
				I	II	III
225	25 453	nonanedioic acid	39.08	44.86	47.40	44.93
226	25 465	azobenzene	53.66	56.67	58.88	57.13
227	25 510	3,3′-dimethyldiazobenzene	72.51	68.49	68.17	68.93
228	25 584	azomethane	19.75	20.55	18.74	20.18
229	25 590	azoxybenzene	60.72	60.26	60.72	59.82
230	25 654	benzaldehyde	32.28	32.39	31.43	32.43
231	25 679	2-chlorobenzaldehyde	36.74	36.51	36.13	37.03
232	25 684	3-chlorobenzaldehyde	36.90	36.51	36.13	37.03
233	25 687	4-chlorobenzaldehyde	37.74	36.51	36.13	37.03
234	25 734	3-ethoxybenzaldehyde	43.81	44.54	42.36	44.48
235	25 742	salicylaldehyde	34.52	34.18	33.07	34.24
236	25 758	4-hydroxybenzaldehyde	35.52	34.18	33.07	34.24
237	25 767	*N*-ethylbenzaldehyde imine	44.45	44.70	43.05	44.52
238	25 776	3-methoxybenzaldehyde	37.78	39.73	37.71	39.68
239	25 793	benzoxime	36.83	36.22	35.39	36.13
240	25 991	*tert*-butylbenzene	44.99	44.78	45.13	44.95
241	25 998	4-methyl-*tert*-butylbenzene	49.92	50.69	49.78	50.85
242	26 014	2,3-dinitrochlorobenzene	45.36	44.55	44.98	45.04
243	26 042	2-chloro-2-phenylpropane	40.01	44.93	45.18	44.89
244	26 049	pentafluorochlorobenzene	33.06	32.07	31.98	31.91
245	26 062	*m*-phenylenediamine	36.15	34.72	34.40	34.93
246	26 093	2,4-dichloronitrobenzene	41.43	41.16	42.81	42.07
247	26 103	catechol	32.95	29.00	29.82	28.91
248	26 108	2,4-difluoronitrobenzene	32.92	33.94	33.71	33.67
249	26 149	4-nitroethylbenzene	42.74	43.45	42.71	43.35
250	26 153	fluorobenzene	26.15	25.92	26.69	25.68
251	26 155	4-iodofluorobenzene	34.96	38.25	39.65	38.18
252	26 156	*o*-nitrofluorobenzene	33.78	33.43	33.57	33.26
253	26 157	*m*-nitrofluorobenzene	32.69	33.43	33.57	33.26
254	26 159	2,4,6-trimethylfluorobenzene	40.35	43.65	40.63	43.38
255	26 163	hexafluorobenzene	26.49	28.46	27.43	27.71
256	26 274	pyrogallol	28.11	30.79	31.45	30.72
257	26 356	benzenesulfinyl chloride	25.46	37.78	39.11	37.86
258	26 416	ethyl benzenesulfonate	45.63	45.75	46.40	49.09
259	26 429	propyl benzenesulfonate	50.19	50.37	51.05	50.69
260	26 450	benzenesulfonyl fluoride	35.05	35.37	35.62	35.29
261	26 498	1-methylbenzimidazole	40.25	39.67	40.48	39.60
262	26 512	phenyldichlorofluoromethane	41.29	40.58	40.73	40.46
263	26 531	benzoic acid	33.64	32.05	33.07	32.09
264	26 692	3-ethylbenzoic acid	44.84	42.57	42.36	42.59
265	26 710	salicylic acid	31.18	33.84	34.70	33.90
266	26 718	propyl 4-hydroxybenzoate	50.28	48.81	48.64	48.74
267	26 848	benzonitrile	31.48	31.31	31.92	31.38
268	26 866	4-fluorobenzonitrile	31.77	31.82	32.07	31.78
269	26 869	2-hydroxybenzonitrile	33.67	33.10	33.56	33.19
270	26 876	3-methylbenzonitrile	34.81	37.22	36.57	37.27
271	27 070	trichlorophenylmethane	45.92	44.90	45.28	44.79
272	27 072	trichloro-(3-chlorophenyl)methane	41.02	49.02	49.97	49.40
273	27 078	benzothiazole	38.99	38.81	39.23	38.49
274	27 088	2-chlorobenzothiazole	44.22	43.20	43.92	43.06
275	27 103	2-methylbenzothiazole	43.94	42.22	43.87	42.44
276	27 113	5-methylbenzothiophene	46.56	46.86	45.53	46.34
277	27 119	2-chlorobenzoxazole	37.33	38.81	42.79	40.39
278	27 128	2-methylbenzoxazole	35.01	37.83	42.74	39.77
279	27 136	benzoyl bromide	39.60	38.99	39.19	38.94
280	27 137	benzoyl chloride	37.15	36.17	36.13	35.81
281	27 160	benzyl alcohol	32.55	32.60	32.83	32.54
282	27 174	3,4-dimethoxybenzyl alcohol	45.79	47.27	45.39	47.04
283	27 175	2-phenyl-2-propanol	44.03	41.83	42.12	41.77
284	27 176	1-phenylpropanol	41.73	41.96	42.12	41.83
285	27 196	benzylamine	34.27	34.32	35.12	34.40
286	27 200	4-(methylbenzylamino)-1-butyne	56.22	56.58	56.85	56.62
287	27 207	benzyldimethylamine	43.54	44.49	44.41	44.44
288	27 211	benzylethylamine	43.37	43.76	44.41	44.20
289	27 212	benzylethylaniline	69.25	69.97	69.13	69.75
290	27 221	benzylaniline	61.84	59.08	59.83	58.96
291	27 222	*N*-benzyl-2-methylaniline	65.29	64.99	64.48	64.86
292	27 254	benzyl chloromethyl ether	41.89	42.45	42.17	42.43
293	27 259	benzyl fluoride	31.09	31.48	31.34	31.10
294	27 261	benzyl iodide	44.94	44.07	44.15	44.05
295	27 263	benzyl isothiocyanate	45.69	45.67	45.74	45.53
296	27 264	phenylmethanethiol	38.80	38.91	38.51	38.87
297	27 296	benzylidene dibromide	47.22	46.64	46.72	46.54
298	27 311	benzylideneethylamine	44.36	44.70	43.05	44.52
299	27 312	benzylidenemethylamine	39.40	39.89	38.40	39.72

Table II (Continued)

no.	ID[a]	compd	obsd	calcd from study		
				I	II	III
300	27313	benzylidene difluoride	30.76	31.34	31.49	31.06
301	27383	butane-2,3-diol	23.61	23.64	23.69	23.59
302	27405	2-chloro-6-phenylphenol	58.25	56.95	57.59	57.29
303	27432	3,3′-difluorobiphenyl	52.18	52.06	51.56	51.68
304	27483	2-iodobiphenyl	64.63	63.37	64.22	63.37
305	27602	bromoacetic acid	20.43	20.32	20.76	20.26
306	27610	bromoacetyl bromide	27.54	27.26	26.89	27.11
307	27611	chlorobromoacetic acid	25.76	25.21	25.45	25.07
308	27631	1,2-butadiene	20.27	20.54	19.86	20.43
309	27638	1,3-butadiene	22.46	20.88	20.80	20.78
310	27639	2-bromo-1,3-butadiene	27.94	28.05	28.57	28.01
311	27640	1-chloro-1,3-butadiene	25.77	24.97	25.50	25.09
312	27646	1,1-dichloro-1,3-butadiene	30.69	30.49	30.19	30.73
313	27647	1,2-dichloro-1,3-butadiene	29.96	28.72	30.19	29.24
314	27655	2-fluoro-1,3-butadiene	20.75	21.02	20.95	20.73
315	27657	hexafluoro-1,3-butadiene	24.06	23.77	21.69	23.44
316	27658	2-iodo-1,3-butadiene	33.76	32.85	33.76	32.81
317	27668	butane-1,3-diyne	17.16	16.94	17.42	17.04
318	27673	*n*-butylamine	24.08	23.83	24.34	23.90
319	27679	2-methyl-2-aminobutane	28.61	28.45	28.99	28.53
320	27696	2-methyl-2-bromobutane	33.37	32.84	32.83	32.77
321	27701	1-chloro-4-fluorobutane	25.33	25.37	25.26	25.09
322	27738	1,2,3,4-diepoxybutane	20.18	19.70	20.03	19.75
323	27747	2,3-epoxy-2,3-dimethylbutane	29.67	29.38	29.52	29.38
324	27756	1,4-butanedithiol	35.50	36.01	35.06	36.07
325	27761	butyl fluoride	20.46	20.99	20.57	20.61
326	27770	isopentyl iodide	38.13	38.36	38.02	38.33
327	27772	*sec*-butyl iodide	33.83	33.71	33.38	33.64
328	27781	1-nitrobutane	27.45	26.56	27.29	26.44
329	27782	2-nitrobutane	25.61	26.39	27.29	26.33
330	27787	1,1,2,2-tetrabromobutane	51.08	51.46	51.47	51.41
331	27788	1,2,2,3-tetrabromobutane	51.42	51.14	51.47	51.32
332	27798	1,2,2-trimethylbutane	44.10	43.63	43.71	43.77
333	27801	2,2,3-tribromobutane	43.90	43.76	43.71	43.86
334	27816	2-methyl-2-nitropropane	26.40	25.96	27.29	26.07
335	27829	butane-1,3-diol	23.71	23.52	23.69	23.50
336	27832	butane-1,3-diol sulfite	30.31	29.48	29.73	29.75
337	27851	butane-1-thiol	28.74	28.42	27.74	28.37
338	27854	butane-2-thiol	28.29	28.55	27.74	28.46
339	27861	1-hydroxy-2-aminobutane	25.38	25.23	25.98	25.36
340	27876	2,2,3,3,4,4,4-heptafluorobutane	22.95	22.96	23.09	22.95
341	27943	*cis*-1-bromo-1-butene	27.54	28.36	28.37	28.13
342	27944	*trans*-1-bromo-1-butene	27.61	28.36	28.37	28.13
343	27945	2-bromo-1-butene	27.61	28.03	28.37	27.96
344	27946	2-bromo-3-methyl-1-butene	32.51	32.81	33.02	32.74
345	27947	2-bromo-4-phenyl-1-butene	51.71	52.36	53.09	52.26
346	27949	*cis*-1-chloro-1-butene	25.00	24.95	25.31	25.04
347	27950	*trans*-1-chloro-1-butene	25.01	24.95	25.31	25.04
348	27951	1-chloro-2-methyl-1-butene	28.50	29.20	29.95	29.28
349	27953	2-chloro-1-butene	24.98	24.61	25.31	24.88
350	28017	crotonic acid	22.46	22.86	22.48	22.95
351	28029	ethyl 4-bromocrotonate	39.98	40.60	39.54	40.65
352	28057	methyl vinyl ketone	20.03	21.04	20.85	21.10
353	28065	but-1-en-3-yne	18.42	18.91	19.11	18.91
354	28066	1-chlorobut-1-yn-3-ene	23.89	24.40	23.81	24.11
355	28067	1-methoxybut-1-en-3-yne	25.83	26.22	25.39	25.87
356	28081	1-(*N*,*N*-dimethylamino)butane	33.82	34.00	33.64	33.94
357	28085	2-aminobutane	21.40	23.96	24.34	23.99
358	28088	ethyl-*sec*-butylamine	33.48	33.40	33.64	33.79
359	28098	*tert*-butyl bromide	28.86	28.23	28.18	28.17
360	28101	*sec*-butyl chloride	26.48	25.34	25.11	25.25
361	28102	*tert*-butyl chloride	25.81	25.22	25.11	25.19
362	28107	1-chloro-2-methyl-1-propene	25.06	24.95	25.31	25.04
363	28109	1,1-dichloro-2-methylpropane	29.79	30.11	30.00	30.32
364	28123	ethyl *tert*-butyl ether	31.43	31.76	31.35	31.63
365	28137	isobutylisothiocyanide	35.19	35.35	34.96	35.21
366	28140	2-methylpropanethiol	28.43	28.59	27.74	28.55
367	28144	1,1-dimethylethanethiol	28.71	28.43	27.74	28.40
368	28151	butyl nitrate	28.32	29.87	28.92	29.62
369	28152	*sec*-butyl nitrate	28.23	30.00	28.92	29.71
370	28153	isobutyl nitrite	26.91	27.30	26.74	27.16
371	28173	butyl sulfite	50.51	50.19	50.15	50.33
372	28174	isobutyl sulfite	50.56	50.53	50.15	50.69
373	28178	butyl sulfoxide	54.14	47.64	46.88	47.45

Table II (Continued)

no.	ID[a]	compd	obsd	calcd from study		
				I	II	III
374	28 179	butyl thiocyanate	31.50	33.28	33.11	33.05
375	28 192	1-chloro-2-methyl-1-propene	25.06	24.59	25.31	24.68
376	28 207	2-butynedinitrile	21.66	20.17	20.37	20.49
377	28 233	3-methylbutanal oxime	29.64	29.62	29.26	29.41
378	28 234	2,2,3-trichlorobutyraldehyde	35.42	35.43	34.74	35.64
379	28 240	butyramide	24.32	23.88	24.58	23.70
380	28 246	*N,N*-dimethylbutyramide	33.43	33.78	33.87	33.70
381	28 263	butyric acid	22.21	21.31	22.29	21.30
382	28 282	2-bromobutyric acid	28.53	29.38	30.05	29.34
383	28 421	butyronitrile	20.37	20.57	21.15	20.59
384	28 426	2-methylbutyronitrile	25.09	25.48	25.79	25.51
385	28 434	2-bromoisobutyronitrile	28.11	28.21	28.91	28.37
386	28 435	2-hydroxyisobutyronitrile	22.12	22.10	22.78	22.27
387	28 443	isobutyroyl bromide	29.14	28.55	28.42	28.47
388	28 445	butyroyl chloride	25.80	25.43	25.35	25.03
389	28 460	isobutyroyl chloride	25.83	25.73	25.35	25.35
390	28 659	*N,N*-diethylcarbamic acid	32.00	30.33	30.86	30.48
391	28 661	ethyl carbamate	22.60	21.16	21.57	21.12
392	28 666	methyl *N*-nitro-*N*-ethylcarbamate	32.18	33.11	33.09	33.29
393	28 717	carbon disulfide	21.50	21.20	21.42	21.38
394	28 812	monobutyl catechol ether	48.56	48.58	48.40	48.34
395	28 921	chloroacetic acid	17.56	17.32	17.69	17.29
396	28 930	ethyl hydroxychloroacetate	28.58	28.95	28.62	28.89
397	28 938	chloroacetone cyanohydrin	26.90	26.48	27.48	26.76
398	28 939	chloroacetonitrile	16.02	16.58	16.55	16.57
399	28 949	bis(1-chloroethyl) ether	32.63	31.80	31.44	32.02
400	28 952	methyl 1-chloroethyl ether	23.17	22.35	22.10	22.41
401	28 961	2-chloroethyl chloroformate	27.66	27.09	27.03	26.93
402	28 962	chloromethyl chloroformate	22.60	22.92	22.39	22.77
403	28 975	trichloromethyl chloroformate	32.58	32.32	31.78	32.26
404	28 976	bis(chloromethyl) ether	22.59	22.94	22.15	23.05
405	28 988	2-chloro-1,3-butadiene	25.23	24.63	25.50	24.93
406	28 995	ethyl chlorosulfinate	27.33	25.50	25.33	25.42
407	28 996	chlorosulfonic acid	26.87	27.24	26.38	27.14
408	28 997	methyl chlorosulfonate	22.02	22.43	21.73	22.34
409	29 070	cinnamaldehyde	44.20	42.27	40.92	42.28
410	29 071	β-bromocinnamaldehyde	50.76	49.44	48.68	49.51
411	29 159	cinnamonitrile	42.96	41.84	41.41	41.94
412	29 162	cinnamoyl chloride	49.99	46.70	45.61	46.37
413	29 216	3-allylpiperidine	39.38	40.32	41.29	40.85
414	29 226	2-propylpiperidine	40.60	40.25	41.10	40.71
415	29 365	2-bromo-4-methylphenol	40.08	40.65	40.59	40.68
416	29 373	2-nitro-4-methylphenol	40.76	40.62	39.70	40.57
417	29 456	perfluorocyclobutene	18.80	20.26	19.67	19.96
418	29 457	phenyl cyclobutyl ketone	48.60	48.49	48.19	48.51
419	29 466	azacycloheptane	31.61	30.90	31.81	31.42
420	29 528	cyclohexane epoxide	27.40	27.25	27.68	27.22
421	29 533	fluorocyclohexane	27.54	27.97	28.03	27.62
422	30 163	*N,N*-dimethyl-2-methylpropane	33.85	34.17	33.64	34.12
423	30 171	*N,N*-dimethylpentane	38.28	38.62	38.28	38.54
424	30 195	ethyl 3,5-dinitrobenzoate	59.97	57.42	56.10	57.48
425	30 207	1,3-dioxane	21.41	21.72	21.86	22.00
426	30 215	1,4-dioxane	21.68	21.83	21.86	21.85
427	30 220	glycol methylene ether	16.84	17.10	17.21	17.40
428	30 221	glycerolethylidene ether	27.76	27.56	28.14	27.95
429	30 224	1,2-ethylenediol carbonate	16.72	17.34	17.45	17.48
430	30 225	1,2-propanediol carbonate	21.36	22.09	22.09	22.17
431	30 361	trimethylene 1,3-disulfide	28.76	30.10	28.58	29.63
432	30 829	ethyl 2-propyn-1-yl ether	24.71	25.19	25.20	25.18
433	30 847	1-chloro-1,2,2-trifluoroethene	17.52	17.11	16.46	17.26
434	30 863	1,1-dichloroethene	20.35	19.96	20.71	20.17
435	30 864	1,1-dichloro-2-fluoroethene	20.43	20.44	20.86	20.28
436	30 867	1,2-dichloro-1,2-difluoroethene	20.48	21.26	21.00	21.55
437	30 871	1,2,2-trichloro-1-fluoroethene	25.36	25.41	25.55	25.84
438	30 922	methoxyacetylene	16.28	15.27	15.91	15.35
439	30 923	phenylacetylene	33.43	33.99	34.34	33.96
440	30 924	propoxyacetylene	24.88	24.69	25.20	24.75
441	31 116	furan	18.16	18.65	19.33	18.74
442	31 117	2-acetylfuran	29.58	29.36	28.86	29.04
443	31 119	2-bromofuran	26.11	26.45	27.10	26.67
444	31 121	2-*tert*-butylfuran	37.46	37.54	37.92	37.47
445	31 122	2-chlorofuran	23.41	23.63	24.03	23.54
446	31 170	furfural	25.44	25.15	24.22	24.95
447	31 177	5-methylfurfural	30.53	30.20	28.86	29.47
448	31 562	1-fluoroheptane	34.39	34.83	34.51	34.40

Table II (Continued)

no.	ID[a]	compd	obsd	calcd from study I	calcd from study II	calcd from study III
449	31 563	perfluoroheptane	36.90	36.50	36.72	36.44
450	31 766	1-bromo-6-fluorohexane	37.57	37.60	37.62	37.27
451	31 790	2,2-dichlorohexane	39.89	39.74	39.10	39.72
452	31 813	1-fluorohexane	29.74	30.22	29.86	29.80
453	31 824	perfluorohexane	31.58	31.77	31.78	31.69
454	31 828	1,1,2,2-tetrachlorohexane	49.27	49.31	48.49	49.30
455	32 325	imidazole	18.77	19.24	19.96	19.34
456	32 328	1-methylimidazole	23.27	24.19	24.61	24.34
457	32 329	4-methylimidazole	23.33	23.58	24.61	23.85
458	32 947	2,4,6-triamino-1,3,5-triazine	36.48	37.23	33.35	36.70
459	33 044	chloroiodomethane	24.31	24.67	24.13	24.56
460	33 065	dichloroiodomethane	29.50	28.91	28.83	28.87
461	33 070	diiodomethane	32.57	32.59	32.40	32.54
462	33 083	trichloroiodomethane	34.92	33.80	33.52	33.87
463	33 084	trifluoroiodomethane	19.18	20.85	19.88	20.86
464	33 379	α-fluoronaphthalene	43.80	43.08	42.56	42.85
465	34 202	2,4-dimethyloxazole	26.09	25.45	26.97	26.21
466	34 203	2,5-dimethyloxazole	25.63	26.17	26.97	26.22
467	34 619	1-bromopentyne	31.31	31.81	31.33	31.79
468	34 626	1-iodopentyne	36.26	36.26	36.52	36.28
469	34 756	2-fluorophenyl ethyl ether	37.47	38.07	37.62	37.73
470	34 757	3-fluorophenyl ethyl ether	37.47	38.07	37.62	37.73
471	34 758	4-fluorophenyl ethyl ether	37.33	38.07	37.62	37.73
472	35 196	phenylacetylene	34.98	33.99	34.34	33.96
473	35 622	4-benzylpiperidine	54.61	55.40	56.52	55.90
474	35 623	*N*-butylpiperidine	45.75	45.86	45.75	45.86
475	35 790	2-chloro-2-bromopropane	29.12	28.76	28.23	28.81
476	35 796	1-chloro-2,2-difluoropropane	20.64	20.71	20.76	20.71
477	35 807	1-chloro-1-nitropropane	26.36	26.08	27.34	26.51
478	35 843	2,2-difluoropropane	15.79	16.34	16.07	16.23
479	35 846	2,2-diiodopropane	41.96	41.73	41.69	41.89
480	36 303	propoxyacetylene	24.88	24.69	25.20	24.75
481	36 337	1,3-dibromo-1-propyne	29.61	29.97	29.80	30.05
482	36 441	pyridine	24.07	23.75	24.89	23.75
483	36 460	2-butoxy-5-aminopyridine	50.08	49.54	49.04	49.72
484	36 463	2-benzylpyridine	50.67	52.55	54.25	52.70
485	36 466	2-bromopyridine	31.44	32.25	32.65	32.30
486	36 467	3-bromopyridine	31.49	31.29	32.65	31.44
487	36 471	2-chloropyridine	29.20	29.43	29.58	29.18
488	36 488	2,3-dimethylpyridine	34.14	34.00	34.18	34.16
489	36 490	2,6-dimethyl-4-ethylpyridine	43.49	42.96	43.48	43.26
490	36 492	2-(dimethylamino)pyridine	39.25	38.73	38.11	38.82
491	36 599	4-methylpyrimidine	26.85	27.89	27.88	27.65
492	36 623	pyrrole	20.65	20.18	21.62	20.33
493	36 626	1-methyl-2-acetylpyrrole	37.01	35.84	35.80	35.63
494	36 630	2,4-dimethylpyrrole	30.55	31.14	30.92	30.75
495	36 861	thiophene	24.36	24.27	25.02	24.22
496	37 802	thiazole	24.19	23.33	23.36	23.23
497	37 808	2,4-dimethylthiazole	32.00	31.08	32.65	31.68
498	37 871	2-bromothiophene	32.53	32.07	32.78	32.14
499	37 872	2-bromo-5-chlorothiophene	37.09	37.05	37.47	36.94
500	37 974	allylthiourea	32.33	35.32	36.07	35.37
501	38 420	ethyl tribromoacetate	45.97	45.77	45.58	45.78
502	38 430	*N,N*-dimethyltrichloroacetamide	40.42	38.37	38.67	38.41
503	38 499	*n*-propyl trifluoroacetate	27.70	27.92	27.38	27.83
504	38 953	ethyl xanthate	43.25	42.13	41.63	42.07

[a] The compound ID is given for easy reference. All molecules having ID numbers less than 24 000 were taken from ref 14. For these compounds the right three digits represent the compound number, and the remaining digits beyond that represent the paper sequel number, e.g., compound 14 287 was taken from paper 14 and its number was 287. Since in the first few papers the molecules were not numbered by the authors, we used arbitrary numbers. Molecules having ID numbers greater than 24 000 were taken from ref 15. Simply subtract 24 000 to get the compound number of the *CRC Handbook*; e.g., the compound 24 484 is compound 484 in the handbook.

2.8557, 16-4.1009, 17-3.7162. The heteroatom substitution for hydrogen is even more confusing: heteroatom replacing hydrogen on saturated carbon when there is no carbon substitution, 5-2.4666, 7-3.1274, 10-3.0075, 14-3.1677; when there is one carbon substitution, 6-2.6338, 9-2.7885, 13-2.1784; in ethylenic carbon, 15-2.8557, 18-3.6247, 20-1.9708. There may be several factors involved in the changes. The substituting atoms may have a direct effect on the volume of the atom concerned, e.g., more electronegative atoms lead to volume contraction due to electron withdrawal. The volume loss due to greater overlapping may also affect the atomic refractivities. The nature of the bonds also plays an important role in its value.

Table I (study I) shows that the hydrogens have a relatively small span of values ranging from 1.0 to 1.5. These values are decreased by electron-attracting atoms. Double-bonded oxygens, like the multiple-bonded carbons, have higher values compared to their single-bonded counterpart. The aryl ether or ester oxygens also have high values. Unexpectedly, the oxygens with a delocalized bond, as in the nitro group, have low values. The nitrogen has a higher value in arylamines than in aliphatic amines. The nitrogens in aromatic heterocyclic

Table III. Classification of Atoms in Selected Molecules

molecule ID	structure[a]	atom type (atom list)
14 300	I	1 (7), 2 (6), 6 (2, 5), 40 (3), 46 (13–17), 47 (11, 12), 51 (9, 10), 58 (8), 60 (4), 96 (1)
23 573	II	5 (1, 5), 47 (6–11), 58 (4), 72 (2), 78 (3)
24 484	III	1 (5), 15 (1), 17 (2), 36 (3), 46 (9–11), 47 (6, 7), 49 (8), 58 (4)
25 151	IV	24 (1, 4–6), 25 (2), 28 (3), 33 (7), 47 (10–13), 48 (14), 60 (8), 75 (9)
26 108	V	24 (3, 5, 6), 26 (1, 2, 4), 47 (12–14), 61 (8, 9), 76 (7), 84 (10, 11)
27 088	VI	24 (6–9), 28 (4), 34 (5), 44 (2), 47 (11–14), 75 (3), 90 (10), 107 (1)
27 263	VII	6 (7), 24 (2–6), 25 (1), 40 (9), 47 (11–17), 74 (8), 108 (10)
27 658	VIII	15 (1, 4), 16 (3), 19 (2), 47 (6–10), 99 (5)
30 922	IX	5 (4), 21 (1), 23 (2), 47 (6–8), 48 (5), 60 (3)
32 329	X	1 (6), 28 (4), 33 (5), 42 (2), 48 (9), 49 (8), 50 (7), 51 (10–12), 73 (1), 75 (3)

[a] See Figure 1 for the chemical structure of the molecules and their atom numbering.

compounds and aromatic nitro compounds have unexpectedly high values. Each individual halogen has little variation in its values, although fluorine, chlorine, and bromine attached to unsaturated oxidized carbon showed some high values.

Since a very small number of parameters are known to express the molar refractivities of many organic molecules[17] and the present calculation showed discrepancies in a few parameters, the data set was allowed to fit in terms of a very small number of parameters by converting all saturated carbons (1–14) to the same type, all ethylenic carbons (15–18) to the same type, and so on, as in Table V. Such a simplified classification (study II) used only 22 atom types, yet the fit of the data set was remarkably good, having a standard deviation of 1.527, a correlation coefficient of 0.991, and an explained variance of 0.981. When these parameters were allowed to predict the molar refractivity of the 78 molecules, the calculated values showed a standard deviation of 1.618 and a correlation coefficient of 0.995. Since here the fitting was done by using simple least-squares technique, the statistical goodness of fit of each parameter is also given by their *t*-test values.

Although the statistical fit with such few parameters gives very good *t*-test values, they cannot represent the subtle changes that may occur due to the change in the nature of the substituents. An intermediate step (study III) was taken to get a solution that would keep the atom classification of study I but would not show unexpected variation from this average value and at the same time reflect these changes. We used quadratic programming subject to the constraints that the solution will not deviate beyond 20% of its base value as obtained in study II. The calculated values of this study gave a standard deviation of 1.2897, a correlation coefficient of 0.993, and an explained variance of 0.984. These parameters predicted the values of the 78 molecules with a standard deviation of 1.5817 and a correlation coefficient of 0.995. The statistics of fit and the predictive power of the various studies are presented in Tables VI and VII. The standard deviations of studies I and III are somewhat better than that of study II. However, the correlation coefficients and the explained variances are almost identical. The standard deviation of the predicted values is slightly better for study III, while for studies I and II it is almost identical. The comparison of the parameters obtained from studies I and III shows that in general the parameters having low values in study I have a tendency toward lower values within the allowed limits in study III. Similarly, the high values in I tended to be high in III. It should be remembered that although the number of parameters used in studies III and I is the same, the number of degrees

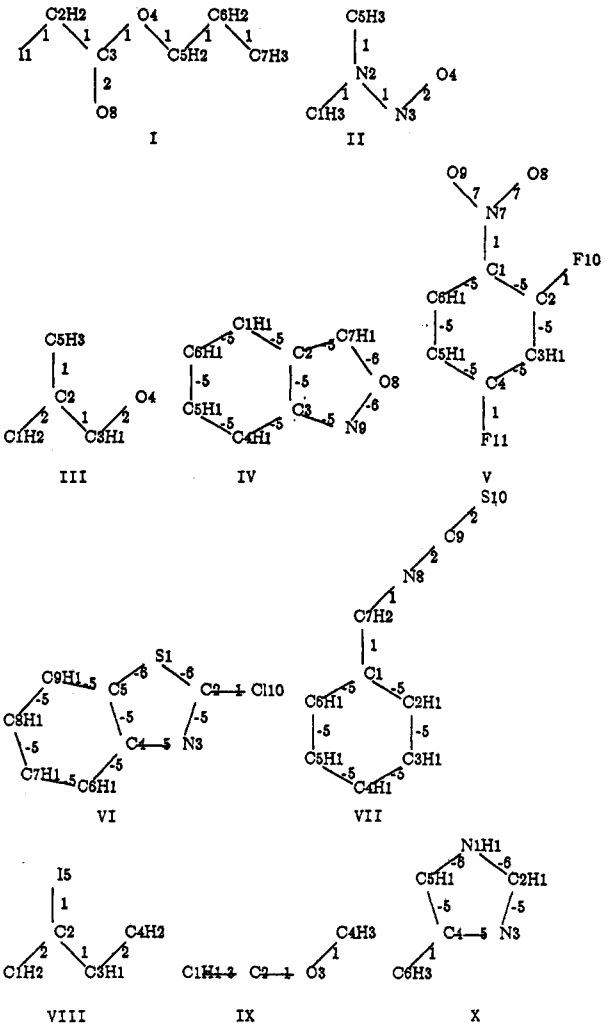

Figure 1. Schematic representation of the structures of the molecules used to illustrate the atom classification. The number after non-hydrogen atoms indicates the atom label, while the number after hydrogen indicates the quantity. The atom label for hydrogen can be easily obtained from the label of their point of attachment. The numbering starts from the lowest non-hydrogen atom and proceeds toward the higher numbered atoms. The number in between bonded atoms indicates the bond type. The structural information was kept according to the Cambridge Crystallographic Data File, with minor modification. The aromatic bonds in pyrrole type structure, for example, were represented by two types of bonds, –5 and –6.

of freedom for regression is much lower in III due to the boundaries formed by the constraints.

Molar Refractivity and Hydrophobicity. The hydrophobicities on the scale of water–octanol partition coefficients are presented in Table I. Except for a few cases, these values are very close to those reported earlier.[13] Since in the present study we used quadratic programming to evaluate the atomic refractivities, we wanted to evaluate the partition coefficient values also using this program. In theory, if the lower limit on the solution of the quadratic programming is lower than the value evaluated by the least-squares technique and if there are no other constraints on the solution, it should lead to the same values of the parameters. Except for 12 parameters, exactly the same values were obtained by this method. The discrepancy in the 12 parameters was found to be due to the singularity or near singularity in the least-squares matrix. The singularity was removed by setting prameters 41 and 44 equal, since they are chemically very similar. Under such a condition both methods gave exactly the same solution. The present solution was obtained by introducing two more molecules, 2-methylbenzimidazole and phenylacetaldehyde. This allowed us to evaluate parameters 36 and 43 from more than one

Table IV. Compounds Used to Check the Predictive Power of the Parameters

no.	ID[a]	compd	obsd	calcd from study		
				I	II	III
1	24079	*N*-methylacetamide	19.73	19.47	19.93	19.36
2	24099	*N*-butyl-*N*-phenylacetamide	58.15	57.43	58.59	57.44
3	24194	chloromethyl acetate	22.47	22.52	22.34	22.52
4	24261	1-methyl vinylacetate	26.90	26.94	27.13	26.87
5	24325	2-methyl phenylacetate	17.66	18.29	17.64	18.22
6	24331	acetone oxime	20.11	19.86	19.96	19.72
7	24342	acetonitrile	11.07	11.22	11.85	11.25
8	24343	cyclohexylidieneacetonitrile	36.51	37.84	38.09	38.00
9	24345	dichloroacetonitrile	21.13	21.60	21.24	21.46
10	24376	dichloroacetophenone	46.19	46.99	45.47	46.74
11	24394	4-fluoroacetophenone	36.19	37.11	36.23	36.93
12	24399	3-hydroxyacetophenone	38.55	38.40	37.71	38.34
13	24410	2-iodoacetophenone	49.37	48.94	49.04	49.02
14	24434	acetyl bromide	20.12	18.90	19.12	18.81
15	24440	acetyl iodide	26.15	23.34	24.32	23.31
16	24441	acetyl isothiocyanide	26.81	24.86	25.90	24.98
17	24490	ethyl 1-bromoacrylate	34.00	34.49	34.89	34.51
18	24525	1,6-hexanedial	29.70	30.40	30.19	30.37
19	24538	monoethyl adipate	46.04	41.37	42.75	41.37
20	24572	3-(methylamino)propionitrile	24.27	23.59	25.07	24.22
21	24597	2-bromoallyl alcohol	24.85	24.69	25.36	24.73
22	24604	allylamine	18.98	19.24	19.89	19.35
23	24624	allyl vinyl ether	25.68	27.11	27.09	26.62
24	24626	diallyl thioether	37.00	38.30	37.42	37.82
25	24627	diallyl sulfoxide	38.04	38.46	37.97	38.36
26	24658	ethyl glycinate	25.62	26.29	26.22	26.26
27	24678	(*N*,*N*-dimethylamino)methyl cyanide	24.07	25.37	25.07	25.35
28	24769	benzalaniline	59.73	58.86	58.47	58.80
29	24803	*o*-chloroaniline	35.48	34.19	35.17	34.71
30	24811	*p*-chloroaniline	28.64	34.19	35.17	34.71
31	24923	*m*-fluoroaniline	30.32	30.57	30.62	30.51
32	24924	*p*-fluoroaniline	28.79	30.57	30.62	30.51
33	24949	4-(methylthio)aniline	44.05	43.03	42.44	42.82
34	25065	*o*-nitroanisole	36.89	40.26	39.70	40.10
35	25067	4-nitroanisole	37.38	40.26	39.70	40.10
36	25375	aramite	87.47	86.64	85.26	86.92
37	25379	arecoline	42.42	43.64	43.17	43.81
38	25499	2,2′-dimethylazobenzene	72.12	68.49	68.18	68.93
39	25597	3,3′-dimethylazoxybenzene	78.03	72.08	70.01	71.62
40	25774	*o*-methoxybenzaldehyde	38.87	39.73	37.71	39.68
41	25781	3-methylbenzaldehyde	37.08	38.30	36.08	38.32
42	25796	*p*-isopropylbenzaldehyde	46.94	47.70	45.37	47.70
43	25948	(α-bromoethyl)benzene	44.06	43.45	43.60	43.33
44	25949	(β-bromoethyl)benzene	43.86	43.32	43.60	43.24
45	25962	*m*-nitrobromobenzene	40.45	40.46	41.18	40.55
46	26179	phenyl isocyanate	33.94	32.56	33.72	32.47
47	26195	α-nitroisopropylbenzene	45.62	45.68	47.36	45.77
48	26952	benzophenone imine	58.62	59.25	58.47	58.95
49	26960	benzophenone 4-(*N*-methylimine)	63.74	65.16	63.11	64.85
50	27015	2-methyl-7,8-benzoquinoline	63.25	61.21	61.27	61.64
51	27071	trichloromethyl-2-chlorobenzene	50.64	49.02	49.97	49.40
52	27074	trifluoromethylbenzene	30.76	31.95	31.64	31.78
53	27110	benzothiophene	41.97	40.95	40.88	40.44
54	27298	(dichloromethyl)benzene	40.87	40.90	40.58	40.76
55	27662	perfluoroisoprene	25.97	28.51	26.63	28.33
56	27665	1,2,3,4-tetrachlorobutadiene	40.04	36.56	39.58	37.71
57	27698	*n*-butyl chloride	25.44	25.21	25.11	25.16
58	27742	1,4-difluorooctachlorobutane	58.84	58.69	58.27	58.93
59	27744	1,4-diiodobutane	46.28	46.32	46.34	46.43
60	27849	butane-1,4-dithiol	35.50	36.01	35.06	36.07
61	27938	2-chlorocrotonaldehyde	25.95	26.30	25.54	26.73
62	27941	3-methylcrotonaldehyde	26.06	26.81	25.50	26.81
63	28020	3-chlorocrotonic acid	28.06	26.61	27.18	27.10
64	28124	butoxyacetylene	29.15	29.31	29.85	29.35
65	28187	*n*-butyl nitrite	26.83	27.13	26.74	26.98
66	28195	1-butyne	19.17	18.89	18.92	18.86
67	28201	2-butyn-1-al	19.59	20.62	19.16	20.62
68	28663	diethyl carbamate	30.31	30.92	30.86	30.92
69	28685	ethyl thiocarbamate	29.91	28.71	28.94	28.76
70	28811	(2-phenylmethoxy)phenol	58.61	59.07	59.18	58.84
71	29546	methylcyclohexylamine	35.33	35.44	36.45	35.91
72	30174	dimethyl sulfone	20.47	21.91	20.05	21.58
73	30175	dimethyl sulfoxide	20.04	20.17	19.00	19.86
74	30695	ethanesulfonic acid	21.48	21.84	21.68	21.98
75	30696	ethanesulfonyl chloride	25.61	25.96	24.74	25.70
76	30697	2-bromoethanesulfonyl chloride	33.06	33.35	32.51	33.17
77	30780	ethyl chlorosulfinate	27.33	25.50	25.33	25.42
78	30781	ethyl chlorosulfonate	26.87	27.24	26.38	27.14

[a] See the footnote of Table II.

Table V. Atomic Refractivities As Obtained in Study II

type	atomic refrac	no. of compd	freq of use	*t* test
1–14	2.8158	821	1311	100
15–20	3.8278	124	148	100
21–23	3.8974	37	43	100
24–35, 42–44	3.5090	468	1205	100
36–41	3.0887	178	205	100
46–51	0.9155	999	4099	100
56, 57, 59, 60	1.6351	195	247	100
58	1.7956	169	202	100
61	2.1407	21	45	100
66–73	3.0100	77	83	100
74	3.2009	27	29	100
75	2.7662	24	27	100
76, 77	3.5054	21	23	100
78	3.8095	11	14	100
81–85	1.0632	49	114	100
86–90	5.6105	103	153	100
91–95	8.6782	57	78	100
96–100	13.8741	20	24	100
106, 107	7.3190	27	30	100
108	9.1680	6	7	100
109	6.0762	7	7	100
110	5.3321	8	8	100

Table VI. Statistics of the Various Studies

study	no. of compd	no. of parameters	std dev	correl coeff	explained variance
I	504	93	1.2685	0.994	0.984
II	504	22	1.5265	0.991	0.981
III	504	93	1.2897	0.993	0.984

Table VII. Statistics of Predictive Power of the Various Studies

study	no. of compd	no. of parameters	std dev	correl coeff
I	78	93	1.6135	0.994
II	78	22	1.6184	0.995
III	78	93	1.5817	0.995

datum. When the atomic partition coefficient values are correlated with the atomic refractivities of the various studies, study I showed a correlation coefficient of 0.322, study II 0.358, and study III 0.340. The low coefficient suggests a poor linear correlation between the two parameters, thereby suggesting the use of both parameters in correlation studies. However, it should be remembered that the correlation coefficient evaluated here is based on the complete atom type set and assumes equal weighting. In a particular QSAR study, such a condition may not hold. So, one should be careful when using both parameters to evaluate the correlation for the particular data set.

Modeling Repulsive Nonbonded Interaction. Although molar refractivity is suitable for modeling the dispersive force or van der Waals attractive interaction, often an important factor for a strongly bound ligand is its steric fit with the receptor cavity. This is the consequence of repulsive nonbonded interaction. In the Lennard-Jones formulation,[20] this interaction is represented by (a/r_{ij}^{12}), where r_{ij} is the distance between two atoms. Unfortunately, in most cases of interest to medicinal chemists the explicit structure of the receptor is not known, making it extremely difficult to model the repulsive interaction. This property is largely dependent on the flexibility of the ligand. An artificial way to model the situation is to measure the volume of the molecule beyond a selected region of the hypothetical receptor cavity and model the interaction in terms of this volume. A study along this line is in progress and will be communicated in the future.

CONCLUSION

The objective of the paper is to make the partially additive, partially constitutive, properties of the ligands, which are related to molecular interaction, into additive ones by hiding the constitutive part in the atom classification. Since the constitutional factors cannot be discretized as we did, it should be considered as an approximate empirical technique. The advantage of this approach is comparable to the advantage of molecular mechanics calculations over quantum mechanical calculations. Our approach gives great flexibility in a correlation study since the local value of the necessary property can be easily calculated in any region of three-dimensional space. An added advantage is that the approximate value of these properties for any molecule can be evaluated by this approach. Although a better approach is to give the atomic values on the basis of some more fundamental properties, such as molecular orbital indices using some physical model, such a method will suffer from the burden of doing such calculations, and the various inaccuracies in those calculations may easily be transmitted to the evaluated atomic property.

ACKNOWLEDGMENT

This work was supported by grants from the National Science Foundation (PCM-8314998) and the National Institutes of Health (5-R01-GM37123-02). We thank Prof. P. K. Ponnuswamy for his comments which helped to improve the manuscript.

REFERENCES AND NOTES

(1) Volz, K.; Matthews, D. A.; Alden, R. A.; Freer, S. T.; Hansch, C.; Kaufmann, B. T.; Kraut, J. *J. Biol. Chem.* **1982**, *257*, 2528.
(2) Scheraga, H. A. In *Advances in Physical Organic Chemistry*; Gold, V., Ed.; Academic: London, 1968.
(3) Berkert, U.; Allinger, N. L. In *Molecular Mechanics*; American Chemical Society: Washington, DC, 1982.
(4) Csizmadia, I. G. In *Theory and Practice of MO Calculations on Organic Molecules*; Elsevier: Amsterdam, 1976.
(5) Nemethy, G. *Angew. Chem., Int. Ed. Engl.* **1967**, *6*, 195.
(6) Pratt, L. R. *Annu. Rev. Phys. Chem.* **1985**, *36*, 433.
(7) Pitzer, K. S. In *Advances in Chemical Physics*; Prigogine, I., Ed.; Interscience: New York, 1959; Vol. II, p 59.
(8) Claverie, P. In *Intermolecular Interactions from Diatomics to Biopolymers*; Pullman, B., Ed.; Wiley: New York, 1978; p 69.
(9) Kochanski, E. In *Intermolecular Forces*; Pullman, B., Ed.; Reidel: Dordrecht, Holland, 1981; p 15.
(10) Blaney, J. M.; Weiner, P. K.; Dearing, A.; Kollman, P. A.; Jorgensen, E. C.; Oatley, S. J.; Burridge, J. M.; Blake, C. C. F. *J. Am. Chem. Soc.* **1982**, *104*, 6424.
(11) Ghose, A. K.; Crippen, G. M. *J. Med. Chem.* **1985**, *28*, 333.
(12) Glasstone, S. In *Textbook of Physical Chemistry*; Macmillan: London, 1948; p 543.
(13) Ghose, A. K.; Crippen, G. M. *J. Comput. Chem.* **1986**, *7*, 565.
(14) Ravindran, A. *Commun. ACM* **1972**, *15*, 818.
(15) Ravindran, A. *Commun. ACM* **1974**, *17*, 157.
(16) Vogel, A. I. *J. Chem. Soc.* **1948**, 1833 and previous articles in the series.
(17) *CRC Handbook of Chemistry and Physics*, 65th ed.; Weast, R. C., Ed.; CRC: Boca Raton, FL, 1984. The molar refractivity, MR, was evaluated from refractive index, n_D, molecular weight, M, and density, d, by using the expression $MR = [(n_D^2 - 1)/(n_D^2 + 2)](M/d)$.
(18) Hooke, R.; Jeeves, T. A. *J. Assoc. Comput. Mach.* **1961**, *8*, 212.
(19) *Dictionary of Organic Compounds*; Oxford University: New York, 1965; Vol. 5, p 3044.
(20) Momany, F. A.; Carruthers, L. M.; McGuire, R. F.; Scheraga, H. A. *J. Phys. Chem.* **1974**, *78*, 1595.