Approach to a General Classification of Solvents Using a Multivariate Statistical Treatment of Quantitative Solvent Parameters

M. Chastrette,*[†] M. Rajzmann,[‡] M. Chanon,*[‡] and Keith F. Purcell[§]

Contribution from the Laboratoire de Chimie Organique et Physique, Université C. Bernard Lyon I 43, Bd. du 11 Novembre 1918, 69622, Villeurbanne Cedex, France, the Laboratoire de Chimie Inorganique Moleculaire, Unité Associeé à CNRS, Numero 126, Université de Droit, d'Economie et de Sciences, Aix-Marseille III, Centre de St. Jérôme, Rue H. Poincaré 13397, Marseille Cedex 13, France, and the Department of Chemistry, Kansas State University, Manhattan, Kansas 66506. Received April 27, 1984

Abstract: A general classification procedure for organic solvents is proposed by treating a basis set of eight solvent physiochemical variables (the Kirkwood function (K), molecular refraction (MR), molecular dipole moment (μ), the δ parameter of Hildebrand, index of refraction (n), boiling point (bp), and the energies of HOMO and LUMO) by the Fage nonhierarchical multivariate statistical method. This classification is based on the representation of 83 solvents as points in an eight-dimensional space, solvent similarity being measured from the distance between two points within this space. The original eight-dimensional space may be reduced to a three-dimensional subspace with only an 18% loss of information. This subspace is defined by the principal components F_1 (strongly correlated with molecular refraction, refractive index, and HOMO energy), F_2 (strongly correlated with the Kirkwood function, dipole moment, and boiling point), and F_3 (strongly correlated with LUMO energy). The finding that the HOMO and LUMO energies are important descriptors of solvent properties is an important result. The 83 solvents may be grouped into 9 classes by their clustering of component values: aprotic dipolar (AD), aprotic highly dipolar (AHD), aprotic highly dipolar and highly polarizable (AHDP), aromatic apolar (ARA), aromatic relatively polar (ARP), electron pair donor (EPD), hydrogen bonding (HB), hydrogen bonding strongly associated (HBSA), and miscellaneous (MISC). Among the possible applications of a quantitative classification, the design of new solvents and the semiquantitative evaluation of physicochemical properties (here solubility in water) are specifically considered. Consideration of the regiospecificities of frontier orbitals provides the basis for the ambiclassification of some solvents.

The choice of the right solvent for a particular application is an everyday decision for the chemist: which solvent should be the best to dissolve certain products, and what solvent should lead to increased reaction yields and/or rates for a reaction? This choice is traditionally based on solvent classification according to the parameters polarity,[1a,2b] basicity, and acidity.[1b,3] The most-used organic classification, due to Parker, recognizes mainly three classes of solvents: protic, dipolar aprotic, apolar aprotic.[4] Chemical experience, however, suggests that more than three solvent classes may be necessary to classify solute/solvent interactions if a useful range of solvents is to be covered. The general aim of this report is, therefore, to quantify chemical intuition in the classification of solvents. The statistical method[5] used here does not lead to an all encompassing, rigid classification of solvents; rather, the exploration of a quantitative, logical approach on which to base a classification provides the means to identify the origin of an ambiguity in the classification of a particular solvent. Another advantage of the procedure is the quantification of the

loss of information incurred by the suppression of any basis parameter(s). Redundancy in the choice of the basis set members is, therefore, avoided, as is over-simplification (see Appendix section).

In preceding papers, one of us[2] described this approach for a set of 22 solvents. The present paper treats 83 solvents and extends the foregoing papers in three directions: (a) establishment of a

(1) Reichardt, C. "Solvent Effects in Organic Chemistry"; Verlag Chemie: Weinheim, New York, 1979; (a) p 41; (b) p 43; (c) pp 11, 270; (d) p 7; (e) p 348; (f) p 256.

(2) (a) Chastrette, M. *Tetrahedron* **1979**, 1441. (b) Chastrette, M.; Caretto, J. *Ibid.* **1982**, *38*, 1615.

(3) (a) Brönsted, J. N. *Ber. Dtsch. Chem. Ges.* **1928**, *61*, 2049. (b) Drago, R. S.; Purcell, K. F. *Progr. Inorg. Chem.* **1964**, *6*, 217. (c) Gutmann, V. "Coordination Chemistry in Non Aqueous Solvents"; Springer: Wien, New York, 1968. (d) Purcell, K. F.; Kotz, J. C. "Inorganic Chemistry"; W. B. Saunders: Philadelphia, 1977; p 235.

(4) (a) Parker, A. J. *Chem. Rev.* **1969**, *69*, 1. (b) Parker, A. J. *Adv. Phys. Org. Chem.* **1967**, *5*, 173.

(5) (a) Dagnellie, P. "Analyse Statistique à Plusieurs Variables"; Presses Agronomiques: Gembloux, 1977; p 185. (b) Fages, R., Seminaire IRIA Lyon, 1974. (c) Lebart, L.; Morineay, A.; Fenelon, J. P. "Traitement des Données Statistiques"; Dunod: Paris, 1979. (d) David, J. G. "Statistics and Data Analysis in Geology"; Wiley: New York, 1973.

[†] Laboratoire de Chimie Organique et Physique, Université C. Bernard Lyon I 43.
[‡] Laboratoire de Chimie Inorganique Moleculaire, Université de Droit.
[§] Department of Chemistry, Kansas State University.

representation for solvents in the three- or four-dimensional space best suited for establishing their similarities, (b) separation of a large number of solvents into classes using a well-defined, quantitative procedure, and (c) use of the multidimensional analysis to explore the possibility that frontier orbital concepts of molecules[6-9] may bring new insights to solvent classification.

Data Base and Analysis Method

Table I lists the 83 solvents selected to provide a range of each of 8 basis variables (see next paragraph). The traditional solvent description[4] would separate 54 of them into sets of apolar and dipolar aprotic solvents (not involved in hydrogen bonding interactions; among them 20 aromatics and 5 chlorinated aliphatics) and 29 protic solvents (capable of entering into hydrogen bonding interactions; 20 alcohols, water, 4 amines, 2 amides, and 2 acids).

Choice of Basis Variables. Five basis variables (the Kirkwood function (K), molecular refraction (MR), the Hildebrand parameter (δ), index of refraction (n), and boiling point (bp)) are properties of bulk solvents while three (dipole moment (μ) and HOMO and LUMO energies (ϵ_H, ϵ_L)) are molecular properties. An important question is whether bulk solvent physiochemical data, characterizing the interaction of an aprotic solvent with itself, are appropriate for a classification which must take into account the possibilities for interactions of those solvents with protic solutes. That δ and bp roughly correlate with H-bonding abilities[10] partly resolves this question; the inclusion of HOMO and LUMO energies as basis variables is further intended to compensate for a possible inadequacy of this type.

The "polarity" of solvents is linked both to the dipole moment, μ, and to the dielectric constant, ϵ. Taft thoroughly discussed different theoretical definitions of solvent polarity in terms of ϵ[11-17] (see the footnotes to Table II for five such definitions). Before retaining one of the definitions, we checked their degree of similarity for $2 \leq \epsilon < 80$. Table II shows that the correlations of all definitions with the Kirkwood function are excellent and, in practice, all the functions are proportional for $\epsilon > 6$. Only the Bloch–Walker definition, which tries to take into account the saturation effect[18a] by treating ϵ as changing when going from the cavity containing the solute to the bulk of the solvent, shows consistently poorer correlations[18b] (number 4 in Table II). The set of correlations described in Table II convinced us to retain

only the Kirkwood function $((\epsilon - 1)/(2\epsilon + 1))$[7] as a characteristic solvent parameter.

Indexes of refraction[1c] were included to describe the polarizabilities of the solvents. The molecular refraction, MR (molecular volume $\times (n^2 - 1)/(n^2 - 2))$, also serves to describe the solvent polarizability. At the suggestion of a referee we investigated the effect of replacing MR by molar volume (MV). The two parameters are fairly well correlated $(r = 0.922)$ for the 83 solvents. The solvents whose deviations are the most important are highly polarizable, such as Ph$_2$O and IPh. Deviations of opposite sign are observed for solvents having high MV and low polarizabilities, such as $(n\text{-Bu})_2$O and $(i\text{-Pr})_2$O. Replacing MR by MV in the principal component analysis described below gives some changes, generally quite small, but for the extremes just cited. The replacement of MR by MV does not change the substance of our conclusions, however,

The cohesion of a solvent (as a measure of the energy required to create a cavity in the solvent) was approximated by using the δ parameter of Hildebrand;[10] more indirectly, one may use the boiling point, bp. These variables implicitly include a term measuring the H-bonding ability of the solvent.

To take into account solute/solvent acid–base interactions, the HOMO and LUMO energies of the solvents were calculated by the EHT method.[19] We use this model of solvent electronic structure because the large number of solvents studied requires use of a not too expensive method for consistently treating both saturated and unsaturated molecules. Even though the EHT method does not give accurate absolute values for the energy levels (the CNDO/S method[20] is more successful in individual cases), all that is required is that energy trends are correctly described. We verified that the trends are correct by comparing EHT and CNDO/S results for a restricted set of solvents. Furthermore, Paolini et al.[21] have used the SCFMO ab initio (STO-3G) formalism to show that for a sample of nine solvents, a relation exists between AN and DN parameters[3c] and the energies of LUMO and HOMO, respectively. Their HOMO and LUMO energies correlate with those from EHT as shown in Figure 1; the correlation is not bad for HOMO, but less satisfactory for LUMO.

In Figure 2 we show histograms for the values of the eight physiochemical basis variables for the full set of 83 solvents. An interesting result of this graphical representation of the data base variables is that it shows their distribution characteristics (even or peaked) and the peculiar absence of molecules with ϵ_L in the -3 eV to -7 eV range.

Methods of Analysis. The eight basis variables $(K, \text{MR}, \delta, \mu, n_D, \text{bp}, \epsilon_H, \text{and } \epsilon_L)$ define an eight-dimensional space in which every solvent may be represented as a point. The eight variables are correlated as shown in Table III. The most noticeable correlations are between K and μ, MR and n, MR and ϵ_H, and n and ϵ_H. In contrast δ and LUMO display a relative independency with regard to the other variables. The dimensionality needed to describe the space solvent classification should, because of the existence of correlations between variables, be lower than 8.

Principal components analysis uses linear combinations of basis parameters to replace the starting set of n basis variables (here $n = 8$) by another set of p orthogonal principal components $(p \leq n)$. Every principal component is a linear combination of the starting variables, and the new set is built in such a way that each principal component bears in turn the maximum of variance. A geometrical analogy may clarify this process: the first principal component is directed along the principal axis of the ellipsoid defined by the cloud of points (in the eight-dimensional space initially defined). The other principal components are then built from the first one by applying the orthogonality condition within the n-dimensional space. This process of orthogonalization aims to replace the starting set of more or less correlated variables by a new set of fully independent variables. Furthermore, there is

(6) (a) Fujimoto, H.; Fukui, K. In "Chemistry Reactivity and Reaction Paths"; Klopman, G., Ed.; Wiley: New York, 1974; Chapter 3. (b) Woodward, R. B.; Hoffman, R. *Angew. Chem., Int. Ed. Engl.* **1969**, *8*, 781. (c) Pearson, R. G. *Top. Curr. Chem.* **1973**, *41*, 75. (d) Flemming, I. "Frontier Orbitals and Organic Chemical Reactions"; Wiley: New York, 1976. (e) Fukui, K. *Angew. Chem., Int. Ed. Engl.* **1982**, *21*, 801.

(7) Kirkwood, J. G. *J. Chem. Phys.* **1934**, *2*, 351.

(8) Hildebrand, J. H.; Prausnitz, J. M.; Scott, R. L. "Regular and Related Solutions"; Van Nostrand-Reinhold: Princeton, 1970.

(9) (a) Griffiths, T. R.; Pugh, D. C. *Coord. Chem. Rev.* **1979**, *29*, 129. (b) Gutmann, V. *Ibid.* **1976**, *18*, 225. (c) Reichardt, C. "Solvent Effects in Organic Chemistry"; Verlag Chemie: New York, 1979. (d) Dimroth, K.; Reichardt, C.; Siepmann, T.; Bohlmann, F. *Liebigs Ann. Chem.* **1963**, *661*, 1. (e) Kamlet, M. J.; Taft, R. W. *J. Am. Chem. Soc.* **1976**, *98*, 377, 2886. (f) Mayer, U.; Gutmann, V.; Gerger, W. *Montsh. Chem.* **1975**, *106*, 1235.

(10) Jensen, W. B. "The Lewis Acid Base Concept"; Wiley: New York, 1980; (a) p 118; (b) p 127; (c) p 153.

(11) Abboud, J. L. M.; Kamlet, M. J.; Taft, R. W. *Prog. Phys. Org. Chem.* **1981**, *13*, 485.

(12) (a) Onsager, L. *J. Am. Chem. Soc.* **1936**, *58*, 1486. (b) Laidler, K. J.; Kandskroener, P. A. *Trans. Faraday Soc.* **1956**, *52*, 200. (c) Abraham, R. J.; Cooper, M. A. *J. Chem. Soc. B* **1967**, 202.

(13) Kosower, E. M. *J. Am. Chem. Soc.* **1958**, *80*, 3523.

(14) (a) Brooker, L. G. S.; Keyes, G. H.; Heseltine, D. W. *J. Am. Chem. Soc.* **1951**, *73*, 5350. (b) Brooker, L. G. S.; Craig, A. C.; Heseltine, D. W.; Jenkins, P. W.; Lincoln, L. L. *J. Am. Chem. Soc.* **1965**, *87*, 2443.

(15) Dubois, J. E.; Goetz, E.; Bienvenue, A. *Spectrochim. Acta* **1964**, *20*, 1815.

(16) Gutmann, V.; Wychera, E. *Inorg. Nucl. Chem. Lett.* **1966**, *2*, 257.

(17) Drago, R. S. *Struct. Bond.* **1973**, *15*, 73.

(18) (a) Bloch, M.; Walker, S. M. *Chem. Phys. Lett.* **1973**, *19*, 363. (b) Motivated by a reviewer comment, we can add that for the 83 solvents in this work, the linear correlation between the Bloch–Walker and Kirkwood functions is rather good $(r = 0.958, s = 0.001)$. Replacing the latter by the former in the principal component analysis described below gave only small changes in the representation of solvents as points in a three-dimensional space. Changes were noticeable only for a few solvents having large ϵ values.

(19) Hoffmann, R. *J. Chem. Phys.* **1963**, *39*, 1397. The program actually used was an updated version from K. F. Purcell's laboratory.

(20) Del Bene, J.; Jaffe, H. H. *J. Chem. Phys.* **1968**, *48*, 1807.

(21) Sabatino, A.; La Manna, G.; Paolini, L. *J. Phys. Chem.* **1980**, *84*, 2641.
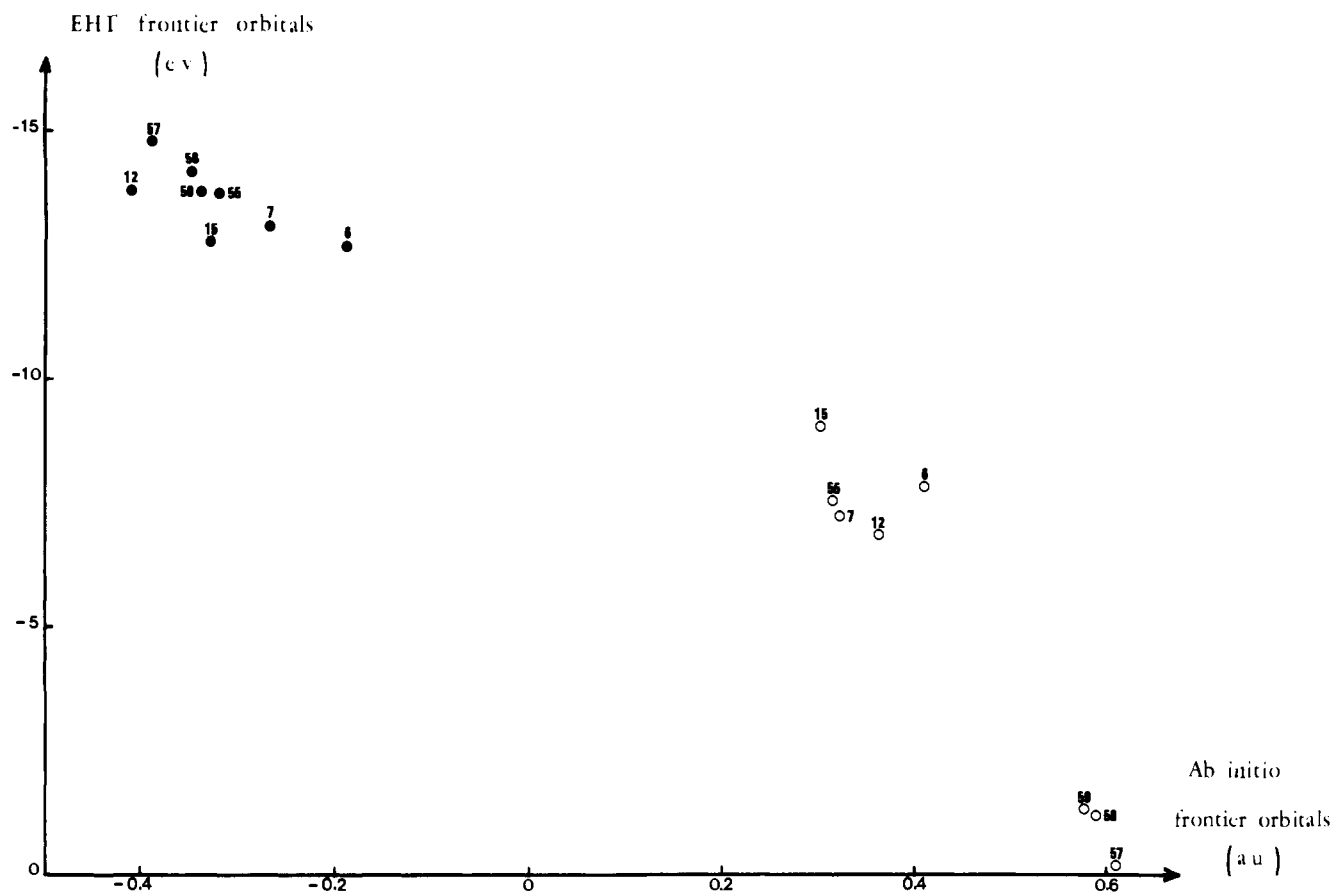
EHT frontier orbitals

Figure 1. EHT frontier orbital energies compared to ab initio frontier orbitals energies: filled points: HOMO; open points: LUMO.

a quantified hierarchy in the information content of the variable space: the first principal component plays a determining role in the description of the whole population of individuals; the last principal component plays the least significant role; suppression of the least significant principal component may cause no important distortion in the description of the population. The distortion incurred by component supression may be expressed as a percentage of information lost. As a consequence, it is possible to pass from an n-dimensional space to a simpler p space by the successive removal of $n - p$ components, and at each removal the loss of information is quantitatively assessed. In this work the suppression of the four least significant principle components entrains a loss of only 8% of the total information. The suppression of five principle components, providing an easily visualized three-dimensional space, results in an 18% loss of information.

Results

Figure 3 is a presentation of the 83 solvent points in the space of F_1, F_2, and F_3. Values (Table IV-S) of the first principal component (F_1), associated with 39.7% of the total variance, distinguishes the set of hydroxylated solvents (water, methanol, and glycol) and dipolar aprotic solvents ($MeNO_2$, CH_3CN) from a group of aromatic solvents such as iodobenzene and diphenyl ether, which are characterized by high polarizability. The second component values (F_2) distinguish hexane, cyclohexane, and carbon tetrachloride from a set of dipolar solvents (PDC, sulfolane, formamide, ...). The third component values (F_3) discriminate between chlorinated solvents such as CH_2Cl_2, $CHCl_3$, and CCl_4 and hydroxylated solvents (glycol, diiethylene glycol, water, and benzylic alcohol).

Although the principal components are mathematical constructs and, as such, do not *necessarily* embody a direct physical significance, one purpose of seeking principal components for a data set is that an otherwise hidden physical interpretation will emerge. Table IV contains the information needed for the detection of approximate physical components. The first principal component (F_1) is strongly correlated with molecular refraction, refractive

index, and the energy of HOMO. F_1 is, therefore, interpreted as an index of the *polarizability of the solvent*. The second principal component is well correlated with the Kirkwood function, dipole moment, and boiling point. It is, therefore, representative of the *polarity of the solvent*. The third component is correlated with the energies of LUMO levels, which are linked to the *acceptor aspect* of donor/acceptor interactions and with electron affinities.[22] For example, McIver recently showed that the relative electron affinities of a series of substituted nitrobenzenes are excellently correlated with the relative gas-phase acidities of substituted anilines and phenols.[23-33] The Hildebrand parameter δ accounts for approximately 25% of the variance in the four first principal

(22) (a) Hoijtink, G. J.; de Boer, E.; van der Meij, P. H.; Weyland, W. P. *Recl. Trav. Chim. Pays-Bas* **1956**, *75*, 487. (b) Matsen, F. A. *J. Chem. Phys.* **1956**, *24*, 602. (c) Dewar, M. J. S.; Hashmall, J. A.; Trinajstic, N. *J. Am. Chem. Soc.* **1970**, *92*, 5555. (d) Kampars, V.; Neilands, O. *Russian Chem. Rev.* **1977**, *46*, 503. (e) Modelli, A.; Jones, D.; Colonna, F. P.; Distegano, G. *Chem. Phys.* **1983**, *77*, 153.

(23) Fukuda, E. K.; McIver, R. T. *J. Phys. Chem.* **1983**, *87*, 2993.

(24) Dack, M. J. R. "Solutions and Solubilities"; Weissberger, A., Ed.; Wiley: New York, 1976; Vol. 8 Part 2.

(25) Bermann, H. A.; Stengle, T. R. *J. Phys. Chem.* **1975**, *79*, 1001.

(26) Parker, A. J. *Quart. Rev.* **1962**, *16*, 63.

(27) Davies, M. M. In "Chemistry of Non Aqueous Solvents"; Lagowski, J. J., Ed.; Academic Press: New York, 1970, Vol. 3, p 1.

(28) Kolthoff, I. M. *Anal. Chem.* **1974**, *46*, 1982.

(29) Snyder, L. R. *J. Chromatogr.* **1974**, *92*, 223.

(30) Ewell, R. H.; Harrison, J. M.; Berg, L. *Ind. Eng. Chem.* **1944**, *36*, 871.

(31) Elguero, J.; Fruchier, A. *An. Quim.* **1983**, *79*, C 72.

(32) (a) Koppel, I. A.; Palm, V. A. In "Advances in Linear Free Energy Relationships"; Chapman, N. B., Shorter, J., Ed.; Plenum Press: New York, 1972; Chapter 5. (b) Fowler, F. W.; Katritzky, A. R.; Rutherford, J. D. *J. Chem. Soc. B* **1971**, 460. (c) Chapman, N. B.; Dack, M. J. R.; Newman, D. J.; Shorter, J.; Wilkinson, J. *J. Chem. Soc., Perkin Trans. 2* **1974**, 962. (d) Abboud, J. L.; Kamlet, M. J.; Taft, R. W. *J. Am. Chem. Soc.* **1977**, *99*, 8325. (e) Schanze, K. S.; Mattox, T. F.; Whitten, D. G. *J. Org. Chem.* **1983**, *48*, 2808. (f) Svoboda, P.; Pytela, O.; Vecera, M. *Collect. Czech. Chem. Commun.* **1983**, *48*, 3287.

(33) Symons, M. C. R. *Chem. Soc. Rev.* **1983**, *12*, 1.

Table I. Variables Associated with the Population of Solvents

no.	solvent	K^a	MR^b	μ^c	δ^d	N^e	bp^f	$\epsilon_H{}^g$	$\epsilon_L{}^h$
1	hexane	0.185	29.9	0	7.27	1.375	69	−12.81	0.94
2	cyclohexane	0.202	27.7	0	8.19	1.476	81	−13.12	0.09
3	triethylamine	0.243	33.1	2.9	7.42	1.401	90	−12.37	−0.66
4	carbon disulfide	0.261	21.3	0	10	1.626	46	−14	−10.7
5	sulfonate	0.483	27.2	16.05	12.5	1.482	287	−10.8	−9.22
6	dimethyl sulfoxide	0.484	20.1	13	13	1.478	189	−12.7	−8.9
7	dimethylformamide	0.48	19.9	12.88	11.79	1.431	152	−13.04	−8.3
8	dimethylacetamide	0.48	24.2	12.41	10.8	1.438	166	−13	−8.1
9	hexamethylphosphotriamide	0.475	47.7	18.48	8.9	1.458	235	−12.7	−6.2
10	N-methylpyrrolidone	0.478	27	13.64	11.3	1.47	204	−13	−8.1
11	nitromethane	0.479	12.5	11.8	12.9	1.381	101	−13.7	−10.4
12	acetonitrile	0.48	11.1	11.48	12.11	1.344	82	−13.8	−7.9
13	butyronitrile	0.464	21.2	13.58	10.17	1.384	118	−13.31	−7.87
14	propionitrile	0.473	15.8	11.91	10.73	1.366	97	−13.56	−7.80
15	acetone	0.465	16.2	9.54	9.62	1.359	56	−12.93	−9.1
16	butanone	0.461	20.7	9.21	9.45	1.379	80	−12.85	−9.11
17	3-pentanone	0.457	25.2	9.41	9.06	1.392	102	−12.73	−9.09
18	cyclohexanone	0.46	27.9	10.04	10.42	1.451	156	−12.7	−9.04
19	diethyl ether	0.345	22.1	4.34	7.53	1.352	35	−13.7	−2
20	dibutyl ester	0.291	40.8	3.94	7.76	1.399	142	−13.09	−1.64
21	diisopropyl ether	0.329	31.7	4.2	7.06	1.368	68	−12.92	−1.77
22	tetrahydrofuran	0.407	19.9	5.84	9.32	1.407	66	−13.4	−2.7
23	dioxane	0.223	21.6	1.5	10.13	1.422	101	−12.9	−2.5
24	dimethoxy-1,2-ethane	0.403	24.1	5.7	8.3	1.38	85	−13.5	−2.6
25	acetic anhydride	0.465	22.4	9.41	10.65	1.39	140	−13.48	−9.12
26	methyl acetate	0.396	17.6	5.37	9.46	1.361	56	−13.56	−8.46
27	ethyl acetate	0.385	22.3	6.27	8.1	1.372	77	−13.37	−8.46
28	diethyl carbonate	0.274	28.4	3.63	8.8	1.385	127	−13.8	−8.03
29	propanediol 1,2-carbonate	0.489	21.6	16.7	13.5	1.421	242	−12.9	−7.5
30	benzene	0.23	26.2	0	9.16	1.501	80	−12.8	−8.3
31	toluene	0.24	31.1	1.43	8.93	1.497	111	−12.63	−8.27
32	o-xylene	0.258	35.8	2.07	9.06	1.505	144	−12.52	−8.21
33	p-xylene	0.232	36	0	8.83	1.496	138	−12.45	−8.28
34	mesitylene	0.23	40.8	0	8.88	1.499	165	−12.53	−8.12
35	styrene	0.244	36.4	0	9.35	1.547	146	−12.36	−9.03
36	fluorobenzene	0.373	26.1	4.9	9.11	1.jl8	85	−12.63	−8.27
37	chlorobenzene	0.377	31.2	5.14	9.67	1.525	132	−12.6	−9.75
38	bromobenzene	0.373	33.7	5.17	9.87	1.56	156	−12.13	−9.03
39	iodobenzene	0.354	39.2	4.64	10.13	1.62	188	−11.78	−8.57
40	o-dichlorobenzene	0.428	35.9	7.57	10.04	1.552	181	−12.43	−9.84
41	m-dichlorobenzene	0.364	36	5.74	9.8	1.546	172	−12.46	−9.77
42	pyridine	0.442	24.1	7.91	10.62	1.51	115	−12.8	−9.3
43	nitrobenzene	0.479	32.7	13.44	10.4	1.553	211	−12.8	−10.3
44	benzonitrile	0.471	31.6	13.51	10.7	1.528	191	−12.8	−9.3
45	acetophenone	0.458	36.3	9.87	10.58	1.534	202	−12.43	−9.79
46	diphenyl ether	0.321	52.8	3.87	10.1	1.589	258	−12.1	−8.29
47	anisole	0.345	33	4.17	9	1.52	154	−12.4	−8.27
48	ethylbenzoate	0.385	42.5	6.67	9.75	1.501	213	−12.81	−9.41
49	carbon tetrachloride	0.226	25.8	0	8.55	1.46	77	−14.35	−10.61
50	chloroform	0.359	21	3.84	9.16	1.445	61	−14	−10.3
51	dichloromethane	0.42	16	5.17	9.88	1.424	40	−13.9	−10.06
52	sym-dichloroethane	0.431	21	6.2	9.86	1.445	84	−13.5	−9.8
53	trichloroethylene	0.309	25.5	2.7	9.16	1.478	87	−12.24	−10.28
54	dichloro-1,1-ethane	0.429	21.2	6.87	8.92	1.417	57	−13.47	−10.04
55	formamide	0.493	10.6	11.24	19.2	1.448	211	−13.8	−8.24
56	N-methylformamide	0.496	15.3	12.88	16.1	1.432	180	−13.3	−8.32
57	water	0.49	3.7	6.07	23.53	1.333	100	−14.8	−0.2
58	methanol	0.477	8.2	5.67	14.5	1.328	65	−14.2	−1.2
59	ethanol	0.47	12.8	5.77	12.78	1.361	78	−13.8	−1.3
60	butanol	0.458	22.1	5.84	11.6	1.399	118	−13.26	−2.23
61	tert-butyl alcohol	0.442	22.2	5.54	10.5	1.388	82	−13.17	−2.53
62	glycol	0.481	14.4	7.61	17.05	1.432	198	−13.64	−2.29
63	isopropyl alcohol	0.463	17.5	5.54	11.44	1.377	82	−13.39	−2.53
64	s-butyl alcohol	0.456	22.1	5.54	11.08	1.397	100	−13.11	−2.27
65	diethylene glycol	0.477	28.4	7.71	14.24	1.448	245	−13.34	−2.98
66	benzyl alcohol	0.445	32.3	5.54	12.05	1.54	26	−12.02	−8.27
67	1-propanol	0.464	17.4	5.54	12.18	1.386	97	−13.57	−2.23
68	isobutyl alcohol	0.459	22.2	5.97	11.24	1.396	108	−13.41	−2.22
69	isoamyl alcohol	0.451	26.7	6.07	11.09	1.407	131	−13.17	−2.23
70	2-pentanol	0.448	26.8	5.54	10.77	1.46	119	−12.85	−2.27
71	2-pentanol	0.445	26.8	5.47	10.16	1.41	115	−12.06	−2.34
72	1-pentanol	0.448	26.8	5.94	10.9	1.41	138	−13.1	−2.22
73	tert-pentyl alcohol	0.381	26.8	5.7	10.5	1.405	102	−13.2	−2.38
74	1-octanol	0.431	40.7	5.81	10.3	1.429	195	−12.6	−2.22
75	cyclohexanol	0.452	28.7	6.2	11.4	1.464	161	−12.9	−2.4
76	trifluoroethanol	0.472	12.4	8.41	12.4	1.291	77	−14.1	−0.96
77	2-methoxyethanol	0.457	19.2	6.81	11.4	1.402	125	−13.5	−2.59

Table I (Continued)

no.	solvent	K^a	MR^b	μ^c	δ^d	N^e	bp^f	$\epsilon_H{}^g$	$\epsilon_L{}^h$
78	acetic acid	0.387	12.9	5.6	13.01	1.372	118	−13.79	−8.50
79	trifluoroacetic acid	0.417	13.7	7.54	10.6	1.285	72	−14	−8.06
80	piperidine	0.381	26.7	3.97	9.45	1.453	106	−12.79	−0.51
81	aniline	0.399	30.6	5.04	10.3	1.586	184	−12.1	−8.27
82	propylamine	0.371	19.4	3.9	8.87	1.387	49	−13.1	0.21
83	diethylamine	0.317	24.3	3.07	0.04	1.386	56	−12.58	0

^{a} Kirkwood function.[7] ^{b} Molecular refraction. ^{c} Dipole moment. ^{d} Hildebrandt function.[8] ^{e} Refractive index. ^{f} Boiling point. ^{g} Energy of the highest occupied molecular orbital calculated by EHT, given in electronvolts. ^{h} Energy of the lowest unoccupied molecular orbital by EHT, given in electronvolts. Except for g and h, which were calculated for this work, the foregoing physicochemical values were extracted from reviews (ref 9a–f).

Table II. Correlations between Different Functions of ϵ and the Kirkwood Function $(\epsilon - 1)/(2\epsilon + 1)^a$

functions^b	$n = 2$	$n = 4$	$n = 6$	$n = 8$
1	0.997	0.999	0.999	0.999
2	0.981	0.994	0.997	0.998
3	0.997	0.999	0.999	0.999
4	0.891	0.927	0.943	0.948
5	0.995	0.998	0.995	0.999

^{a} For 20 values of ϵ between n and 80. ^{b} Function 1 = $(\epsilon - 1)/(\epsilon + 1)$ (ref 12c); function 2 = $(\epsilon - 1)/(\epsilon + 2)$; function 3 = $(3\epsilon - 1)/(8\epsilon + 1)$ (ref 12b); function 4 = $(3\epsilon \ln \epsilon)/(\epsilon \ln \epsilon - \epsilon + 1) - 6/(\ln \epsilon) - 2$ (ref 21, 15–20); function 5 = $1/\epsilon$. For other definitions of polarity, see ref 9a, 12, and 13–17.

Table III. Correlation between the Basis Physicochemical Variables Used to Build the Present Classification^a

K	MR	μ	δ	n	bp	ϵ_{Homo}	Δ_{Lumo}
	−0.42	*0.80*^{a}	0.58	−0.26	0.27	−0.19	−0.03
		−0.19	−0.54	*0.65*	0.51	*0.71*	−0.15
			0.38	−0.12	0.47	0.02	−0.26
				−0.15	0.30	−0.36	0.08
					0.54	*0.62*	−0.51
						0.53	−0.26
							−0.20

^{a} Italic coefficients exceed 0.6.

Table IV. Principal Components in Eight-Dimensional Space

principal component	dimensionality							
	1	2	3	4	5	6	7	8
% variable	39.7	31.0	11.6	9.3	3.8	2.7	1.3	0.5
cumulated variance		70.7	82.4	91.7	95.5	98.2	99.5	100

Correlation Coefficients^a

variable				
K	−0.53	*0.74*	0.07	−0.21
MR	*0.91*	0.02	0.24	−0.10
μ	−0.27	*0.85*	−0.02	−0.40
δ	−0.55	0.56	0.12	0.58
n	*0.82*	0.27	−0.19	0.34
bp	0.46	*0.78*	0.28	0.18
ϵ_{Homo}	*0.81*	0.23	0.25	−0.19
ϵ_{Lumo}	−0.37	−0.40	*0.82*	0.03

^{a} For a four-dimensional space.

components, without being significantly correlated to any of them. The first two principal components are balanced in that they account for similar proportions of the total information (respectively 39.7% and 31.1%); this indicates that polarity and polarizability have comparable importance to the relative positions of solvents.

Table IV may be used further to approximately evaluate in which direction a point should move when a basis variable is changed. For example, if we want to evaluate the effect of substituting a lower lying orbital than HOMO for benzylic alcohol (aiming by this alternative choice to select a molecular orbital of oxygen lone pair character in place of the π HOMO used in this study), then we note that the energy of HOMO is positively correlated with F_1. Therefore, increasing the absolute value of ϵ_H should displace this solvent away from aromatics and nearer to the alcohols, in accordance with chemical intuition. For our

Table V. Characteristics and Basis of Some Previous Classifications of Solvents^a

starting variables	dimensionality of space	no. of classes	ref
ϵ, $\epsilon x \mu$	2′	4	24
ϵ, DN	2	2	25
acidity, basicity	2	4	3a
ϵ, proticity	2	3	26
ϵ, E_T, proticity	3	3	1
ϵ, acidity, basicity	3	8	27
proticity, rel acidities, rel basicities	3	6	28
μ, H-bond acceptors H-bond donors	3	8	29
ϵ, donor atoms, networks, H-bonding	4	5	30
polarity, polarizability electrophilicity, nucleophilicity	4	12	31

^{a} This table does not include multiparametric correlation studies (see ref 1f, 32), which do not aim to classify solvents. A recent review (9a) treats in detail the monodimensional approaches, as characterized by the choice of only one parameter (E_T, Z, etc.) determined from physicochemical or spectroscopic measurements. A recent contribution shows the advantages of combining several spectroscopic methods to gain deeper insights into solute–solvent interactions (ref 33).

basis of eight variables, all but the HOMO and LUMO energies are experimentally unique and do not present the opportunity for this type of intramolecular variable selection.

This last interpretation of the correlation coefficients of the F_i's and the basis variables also has the important feature of complementing the solvent classification results to be discussed later. For points which fall at a frontier region between the parts of space defining different solvent classes, the correlation coefficients suggest what should be done to displace a solvent situated near a "border" from one class to another.

Classification of Solvents. Reichardt[1a] and Griffiths[9a] have reviewed and discussed the several classifications of solvents which have been proposed and in which the number of classes considered varied from one to four. The summary presented in Table V illustrates the scope of these approaches. In most cases the number of classes and the number of classification parameters were arbitrarily chosen. In the present work we started with eight parameters, but the number of classes follows from the nonhierarchical taxonomy that we used.

The Appendix section describes the principle of Fage's method[5b] for treatment of our data. This method allows one, starting from an unique population of 83 solvents, to progressively sort solvents into classes. Every decision to create a new class is based on a mathematical analysis measuring the dispersion of points in the n-dimensional space. The advantage of Fage's approach is that being nonhierarchical, it distributes optimally all the individuals at each new partition decision, in place of only partitioning the most dispersed class. We have illustrated this point in Figure 4 by labeling two specific solvents (⊙ and ⊖) in order that one can follow their assignments through successive classes as the process of classification progresses.

A basis for deciding the number of classes (nine) to be retained follows from the fact that the percentage of information retained in the process of classification increases with the number of classes but slowly when this number exceeds eight (72% of the total variance is accounted for with eight classes, 74% with nine classes, 76% with ten classes). There are, therefore, no compelling reasons
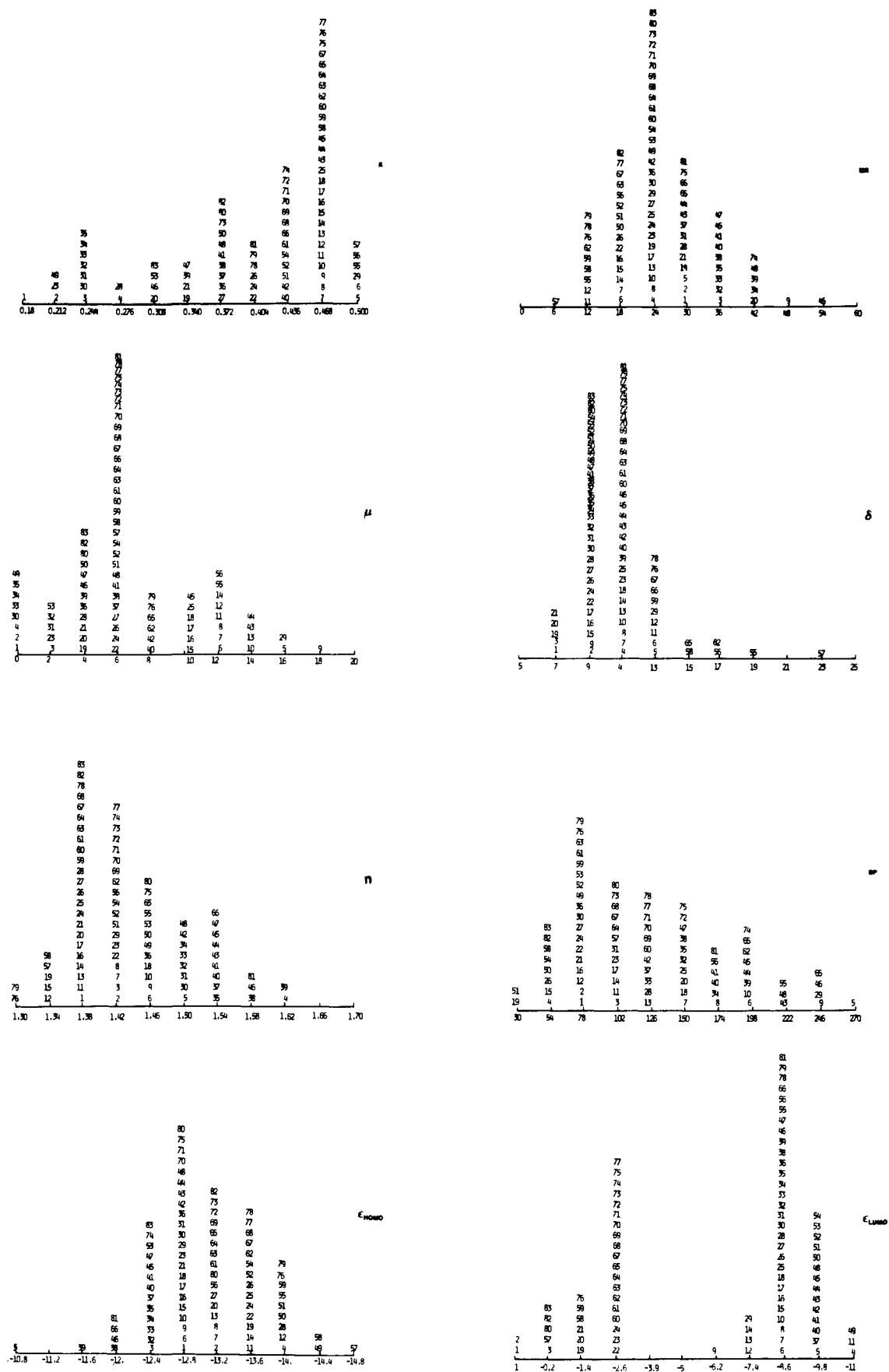
Figure 2. Histograms showing the distribution of each of the eight basis variables used in this work.

to distinguish between eight or more classes. The partitions nine and ten are interesting, however, because they point out the special behavior of two well-known dipolar aprotic solvents, HMPA and sulfolane (which constitute a class when we allow for nine classes)

and of water (a class in itself when we allow for ten classes).

An interesting result of the taxonomy is that were we wishing to define three classes, only then would these three *not* be the ones defined by Parker.[4] It is only at the level of six classes that dipolar

General Classification of Solvents

J. Am. Chem. Soc., Vol. 107, No. 1, 1985 7
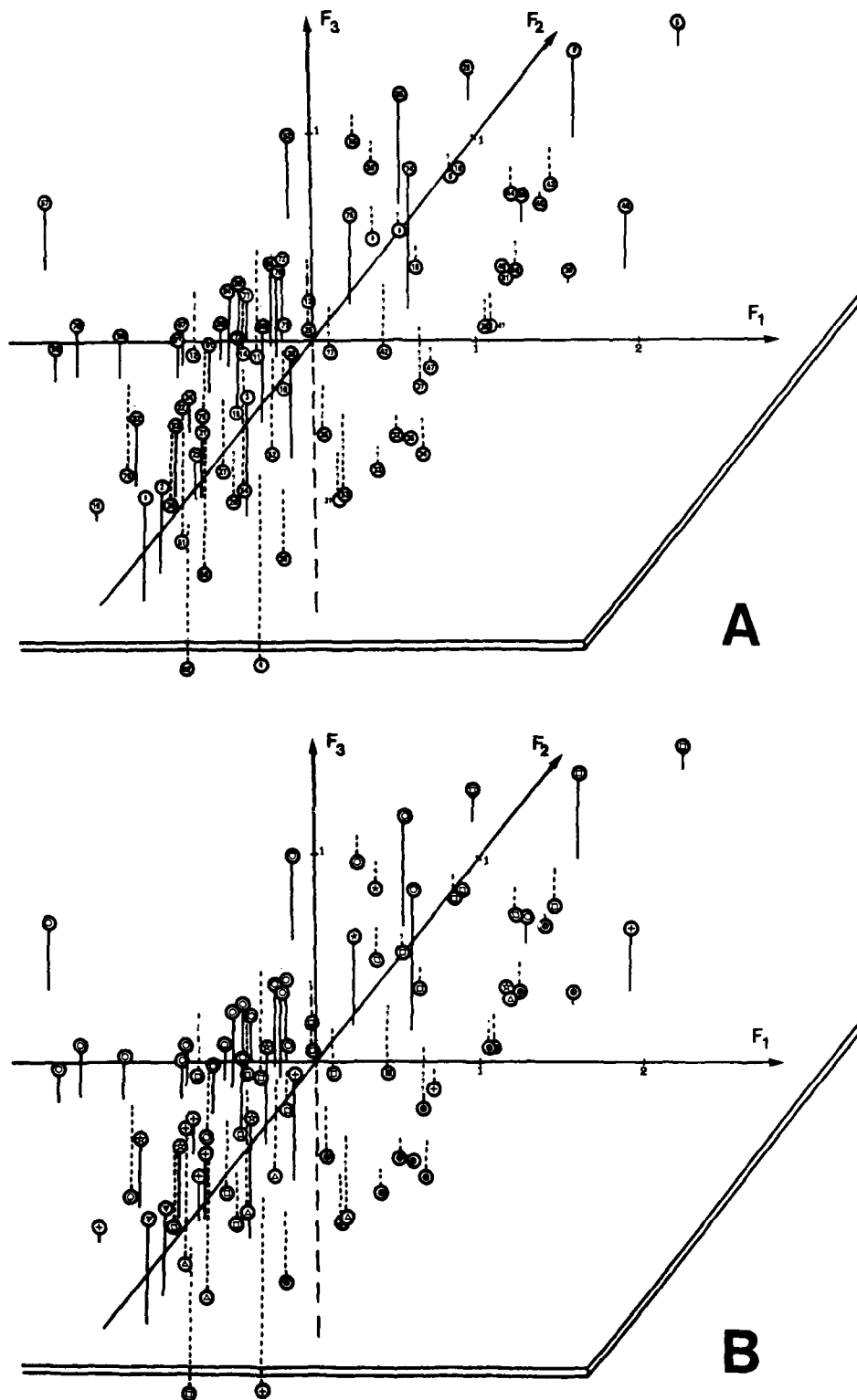
Figure 3. (a) Solvent points by solvent number (Table I) and (b) solvent points by classes according to the code: AD, AHD, AHDP (□); HB, HBSA, H₂O (O); ARA, ARP (●); DPE (+); and MISC (△).

aprotic solvents are separated from other solvents and they fall in not one but two classes.

In the subsequent discussion we focus on description of the class members at the nine-class level. In the first class, named AD (for aprotic dipolar), are found the usual aprotic solvents having a relatively low polarity. In the second class, named AHD (aprotic highly dipolar), are found the more polar aprotic solvents. A third class, differing from the second by the addition of high polarizability, is defined by only two members (HMPA and sulfolane); this class is named AHDP. To summarize, we list by number the solvents that comprise each of these three classes ((*) marks solvents not expected to be classed as found): AD, 11, 12, 13,

14, 15, 16, 17, 25, 26, 27, 51, 52, 54, 79*; AHD, 6, 7, 8, 10, 18, 29, 42, 43, 44; and AHDP, 5, 9.

The apolar solvents are divided into three classes. Aromatic solvents give two of these classes, one covering the aromatic apolar solvents (ARA), the other covering the aromatic polar solvents (ARP): ARA, 30, 31, 32, 33, 34, 35, 36*, 53*; and ARP, 37, 38, 39, 40, 41, 45, 46, 47, 48, 49*, 66*, 74*.

Solvents which typically act as electron pair donors constitute a sixth class. This class, labeled EPD (electron pair donor), contains amines and ethers: EPD, 1*, 2*, 3, 19, 20, 21, 23, 80, 82, 83.

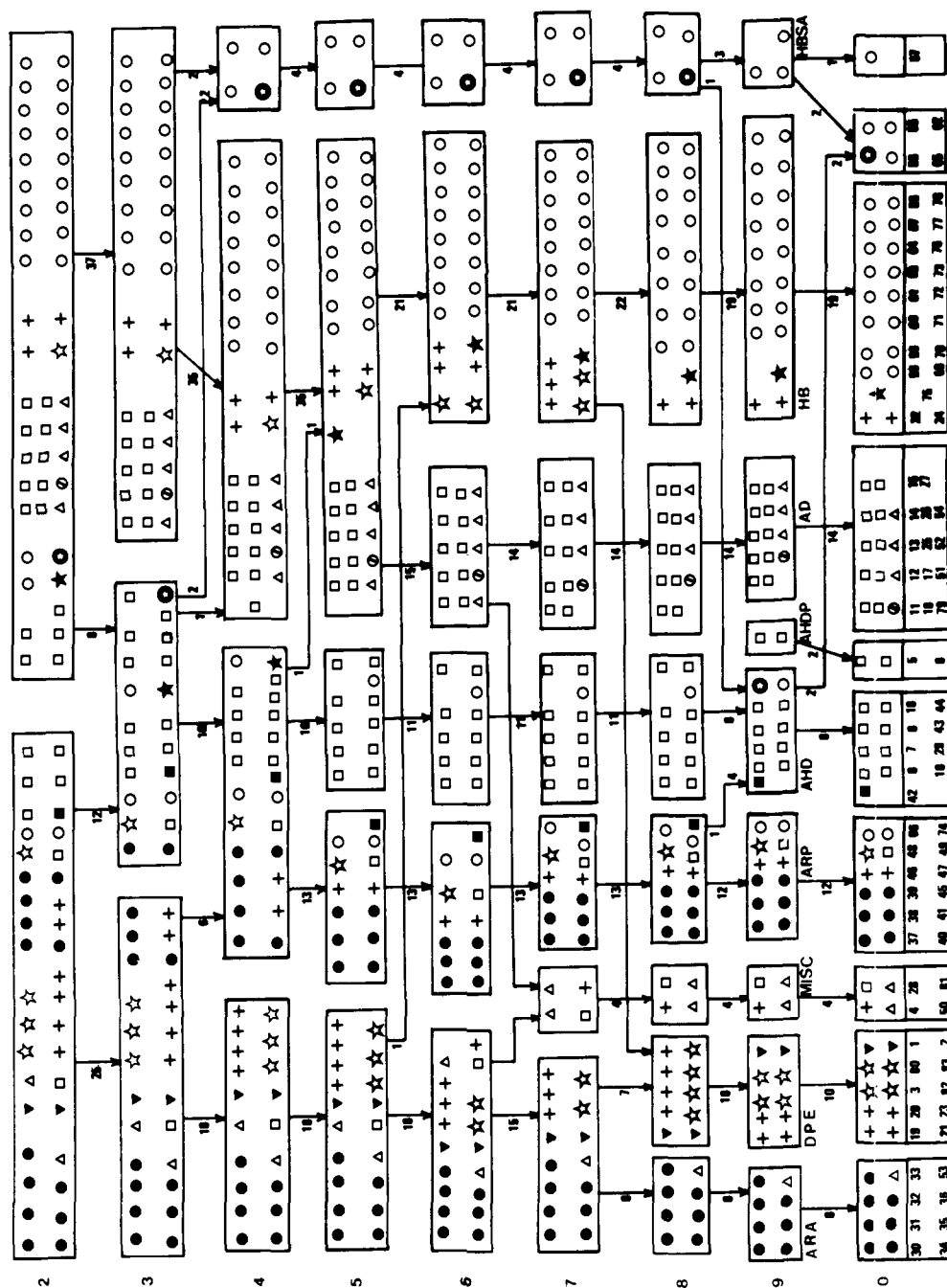The hydrogen bonding solvents are clearly separated into two

Figure 4. Flow diagram of successive partitions in the population of solvents. The two solvents labeled with (⊙) and (⊖) show the nonhierarchical nature of the assignment process; a new class may incorporate individuals from more than a single preceding class as the taxonomy progresses.

classes. Most of the alcohols are in a class named HB (hydrogen bonding), while strongly associated solvents constitute a class of their own named HBSA (hydrogen bonding strongly associated): HB, 22*, 24*, 58, 59, 60, 61, 63, 64, 67, 68, 69, 70, 71, 72, 73, 75, 76, 77, 78; and HBSA, 55, 56, 57, 62, 65. A subclass of HBSA is constituted by water when ten classes are allowed.

The final class is constituted by four solvents which share the characteristic of high polarizability but nothing else. This class is accordingly named MISC (for miscellaneous): MISC, 4, 28, 50, 81.

The overall classification reached by this quantitative treatment overlaps quite strongly with the chemist's intuition. Limitations are apparent, however. OH-bearing solvents provide two especially interesting examples: trifluoroacetic acid appears with the class AD, and benzylic alcohol and 1-octanol are classified with ARP. These classifications will be discussed from a general point of view in the Conclusions section.

The quantification of the successive decisions taken to obtain classification allows a clear identification of which factors intervene in the creation of a new class. The weight given a variable in the decision process increases dramatically at those stages that its discriminating power (see Appendix section) is effective. This idea is graphically presented in Figure 5 and quantitatively in Table VI. For the first partition the discriminating variables are molecular refraction, refractive index, the energy of the HOMO (all of them associated with solvent polarizability), and the Kirkwood function. The second partition decision utilizes the Kirkwood, dipole moment, and boiling point variables (all of them represented in the second F principal component). For the subsequent decisions important variables are successively δ (Hildebrand), again μ, and the energy of LUMO. The decision to increase the number of classes from five to six is therefore a consequence of the energy of LUMOs. Beyond eight classes, all the variables play a role of comparable importance.

Application to the Solubilities of Liquids in Water. Hildebrand[8] expressed the old rule "similia similibus solvuntur" by noting that the solubility of two nonpolar solvents is the greatest when their δ values are close. This illustrates an application of the similarity

General Classification of Solvents

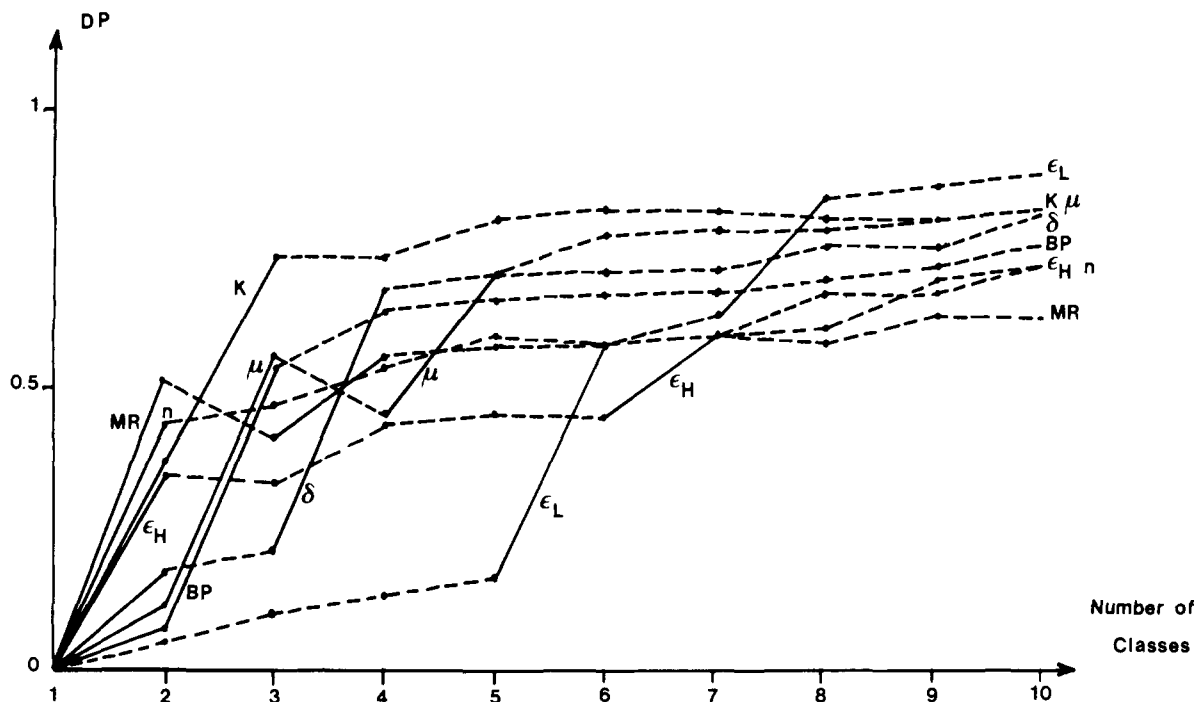*J. Am. Chem. Soc., Vol. 107, No. 1, 1985* 9

Figure 5. Discriminating power of the solvent properties during the successive partitions. A variable plays an important role in the creation of a new partition when its discriminating power strongly increases. The dotted lines indicate an unimportant weight for the considered variable and therefore an unimportance in the descision to create a new class.

Table VI. Discriminating Power of Variables in the Successive Decisions of Partition[a]

no. of classes	variable							
	K	MR	μ	δ	n	bp	ϵ_H	ϵ_L
2	*0.39*	*0.52*	0.13	0.18	*0.43*	0.08	*0.36*	0.06
3	*0.74*	0.41	*0.56*	0.21	0.47	*0.55*	0.34	0.11
4	0.73	0.55	0.45	*0.68*	0.54	0.64	0.44	0.13
5	0.80	0.58	*0.71*	0.70	0.59	0.65	0.45	0.17
6	0.82	0.58	0.76	0.71	0.58	0.66	0.45	*0.58*
7	0.81	0.59	0.77	0.71	0.59	0.66	*0.60*	0.63
8	0.80	0.58	0.77	0.76	0.68	0.69	0.61	*0.84*
9	0.80	0.63	0.80	0.75	0.68	0.71	0.69	0.85
10	0.80	*0.80*	0.81	0.81	0.70	0.75	0.72	0.88

[a] Underlined discriminating powers indicate an important role for the particular variable in the partition.

rule in a "one-dimensional" space. Other authors have tackled the problem of miscibilities by preliminarily resolving δ into components linked to orientation/dispersion interactions and to hydrogen bonding.[34] Considering that the limited representation space was obtained with a minimum of distortion, the distance between two points gives a reasonable[35] measure of the liquids' similarity and, therefore, of their miscibility.

We have examined the relation between solvent solubility (S) in water and the hyperspace distance (D_n) of the 82 solvents from water. The hyperspace distance is defined in the Euclidian sense as

$$D_n = (\sum_{i=1}^{n} [(F_i)_{solv} - (F_i)_{water}]^2)^{1/2} = [\sum_{i=1}^{n} d_i^2]^{1/2}$$

where i = index of the ith principal component, n = dimensional of the hyperspace space, and F_i = projection of the solvent point on the ith principal component axis. The values of D_4 for the 4D hyperspace range from 1.21 (acetonitrile) to 3.61 (diphenyl ether) and differ from those of the full 8D space with a standard error of only 0.01. A study of the S/D correlations shows the correlation to be 0.4 for less than four F components but reaches $r = 0.77$

(34) (a) Hansen, C. M. *Ind. Eng. Chem. Prod. Res. Dev.* **1969**, *8*, 2. (b) Hansen, C. M. *J. Paint Technol.* **1967**, *39*, 104, 505.

(35) Reichardt (ref 1d) gives good reasons for not expecting a close adherence to this proposition.

($s = 0.59$) for four components. The 4D correlation is presented in Figure 6 for the 50 solvents for which $0.0001 < S < 5$ (M).

One should expect that all the solvents for which the representative solvent hyperspace point lies near that of water are highly soluble in it and that there be at least a rough correlation of solubility with hyperspace separation. Of the 51 solvents soluble ($0.2 < S < 5$ (M)) or completely miscible in water ($S > 5$ M; not shown in Figure 6), 47 have hyperdistances of less than 2.75 units. Sulfolane (3.30) and HMPA (3.27), although completely miscible, stand apart, as do aniline (3.00) and benzyl alcohol (2.85) (from the soluble group). In the subclass of 28 solvents of low solubility in water ($S < 0.2$ M), 24 are distant from water by a value greater than 2.75. The four which do not adhere to the ($S < 0.2$ M)/($D > 2.75$) discrimination are chloroform (2.47), 1,2-dichloroethane (2.40), fluorobenzene (2.74), and diethyl carbonate (2.57), which is insoluble.

A second way to analyze the solvent/water point separation is to consider the individual component distances (d_i) of the various solvents from water. Attempted correlations between log S and d_i in 6D space show that the distances on the i-1,2 axes suffice to obtain as reasonable a correlation as obtained from 4D space ($r = 0.794$, $s = 0.55$). This suggests that the variables used in the construction of the first two principal components (mainly polarizability and polarity) account approximately for solvent solubility in water. The solvents which stand outside of this correlation are hexane, cyclohexane, diphenyl ether, and aniline. Discarding these four solvents increases r to 0.896 ($s = 0.26$). The fact that alcohols adhere to this correlation (except octanol, which behaves like hexane) suggests that the analysis correctly takes into account the hydrogen bonding interactions.

Conclusions

Because the nature of a specific chemical problem may depend on a certain solvent parameter to a degree not reflected in the present basis variable set, the method used here to achieve solvent classification does not necessarily provide *the* most chemically useful classification; however, it has the advantage of clearly delineating the successive decisions leading to the classification. This is expected to be of value when disagreements between different classification schemes arise, and the origins of the discrepancies may be traced out and resolved (if indeed they are resolvable). This advantage derives from the fact that *this*
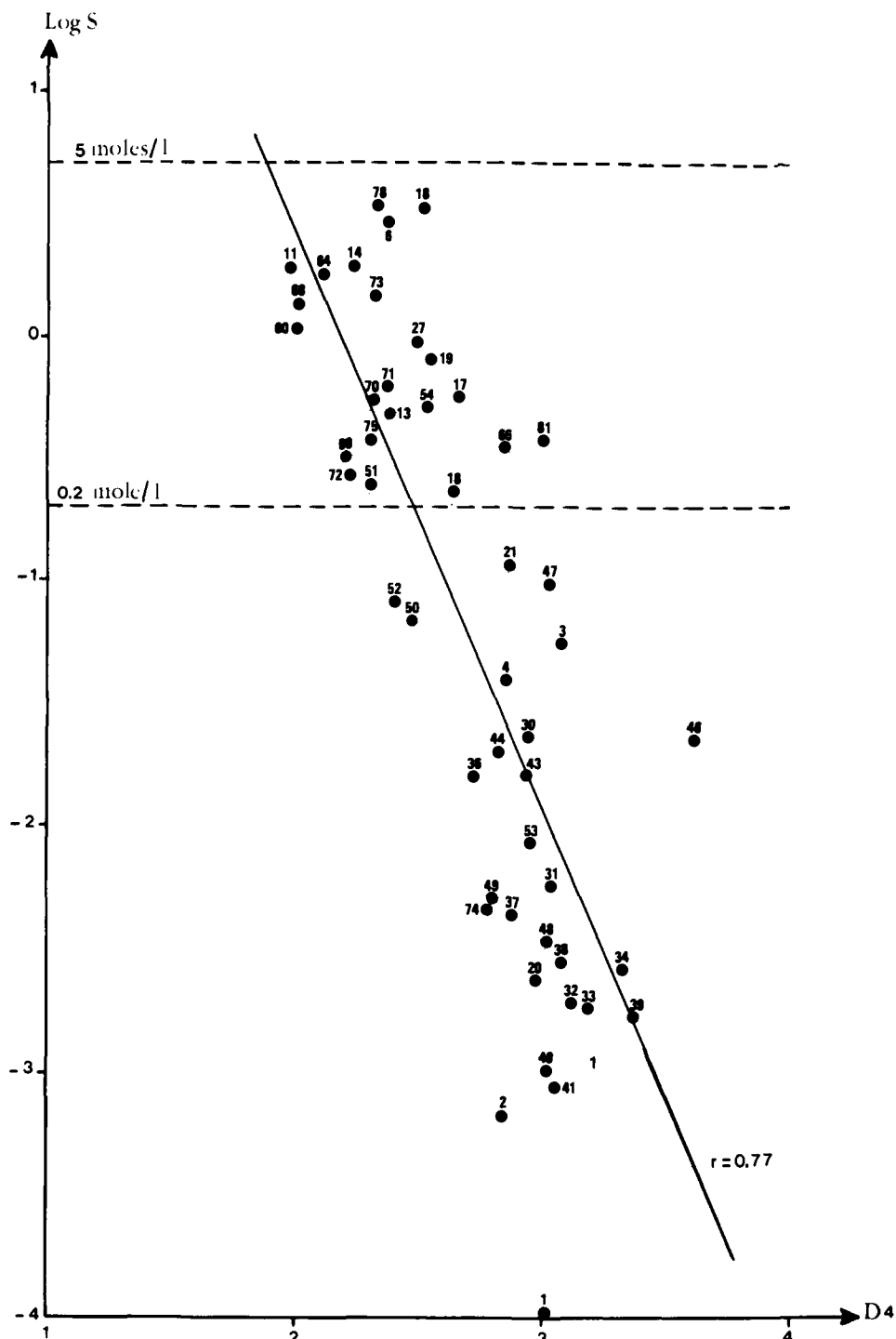
Figure 6. Correlation between solubility in water (*S*) and intersolvent distances (*D_4*) in 4D principle component space.

classification is quantitative and mathematically rigorously defined.

As an example of the identification of the source of a discrepancy, we use the substituted benzenes. In the discussion of the classification results above, we noted the "unexpected" inclusion of benzyl alcohol in the ARP group. Diphenyl ether and aniline are structurally related to benzyl alcohol (all are heteroatom-substituted benzenes), and they appear as "unusual" in failing to fit the S/D_2 correlation. A possible explanation for the unusual behavior of such solvents may lie in the fact that the HOMO energies for all three are those for mainly aromatic π orbitals. Each also has a high-lying heteroatom lone pair MO which could conceivably be a more appropriate HOMO choice in applications where formation of a hydrogen bond is important; solubility in water appears to be such a case. One of us has shown in earlier experimental work[36] that the HOMO may play a secondary role

in intermolecular interactions when a deeper orbital provides better overlap with the LUMO of the acid. Strict adherence to the definition of HOMO for benzyl alcohol derives from recognition of the energy-match criterion for adduct formation, ignores the overlap criterion, and implicitly biases the treatment of benzyl alcohol as an aromatic solvent, rather than as an alcohol. This treatment would probably be justifiable if we were seeking to correlate solvent solubility in benzene with hyperspace distance from benzene.

We, therefore, reach the conclusion *that there could be a degeneracy of representations associated with the choice of HOMO and LUMO of multifunctional solvents.* By the term degeneracy we mean that a solvent may have two points in hy-

(36) Arbelot, M.; Metzger, J.; Chanon, M.; Guimon, C.; Pfister-Guillouzo, G. *J. Am. Chem. Soc.* **1974**, *96*, 6217.

perspace associated with it; for different chemical applications one point is more relevant than the other as dictated by the choice of a chemically relevant HOMO (or LUMO). This is somewhat reminiscent of the situation already met when acids and bases are classified in a generalized approach.[37] It must be stressed that duality in parameter choice does not reduce interest in the classification; rather, our results show the fallability of searching for a unique, global solvent classification and also verify the utility of the method for identifying degenerate solvent representations.

As further examples of the limitations of this work's final classification, we note that the eight-member basis variable set has no representative related to solvent H atom donor ability, which may play a determining role in the selectivity of some radical reactions. As a consequence, our classification space is very probably not so well adapted for treating solvents in radical reactions. As another example, because it is the *overall* size of the solvent which is reflected in its molecular refraction, the operation of steric hindrance around a *specific* solvating site within the solvent is not taken into account. Similarly, the current classification will be totally useless, as are all other main classifications (Table VI), when solvent chirality is the chemically relevant property.

In addition to its value for the quantity of information that it summarizes in a logical way, the utility of the present classification scheme also derives from its role as a basis from which to derive new ideas about solvent applications and interpretations of solvent properties. In a personal communication, Carlson[38] has drawn our attention to the value of such a quantitative approach for experiment planning[32] to systematically improve the yield of a given reaction. Data organized as in Figures 2 and 3 pave the way for applications in which solvent similarity can be judged from the proximity of points in hyperspace and also in which "empty" regions of this space are clearly evident. For example, the use of HMPA, which is becoming more and more proscribed because of possible carcinogenic properties, may be avoided by substituting for it a solvent whose position in the hyperspace (for example diethylacetamide or tetramethylurea) is very close to that of HMPA. The predictive utility of the approach used in this work derives from the introduction of theoretical parameters (frontier orbital energies) in the construction of a space for solvent classification. This could lead to "a priori design" of new, interesting solvents intended for a specific purpose. Moreover, that approximate values of refractive index and dipole moment are now accessible from theoretical techniques raises hope for estimating the position of a not yet synthesized solvent in hyperspace. Such knowledge would provide guidance and impetus in the search for new solvents with properties not like those examined here.

(37) Purcell, K. F.; Kotz, J. C.; "Inorganic Chemistry"; W. B. Saunders: Philadelphia, 1977; p 224.
(38) Carlson, R. *Acta Chem. Scand.*, submitted.

Appendix

The partition method used in this work rests on a minimization of each solvent class dispersion D_{I_k}

$$D_{I_k} = \sum_{i \in I_k} \sum_{j=1}^{p} (x_{ij} - x_{kj})^2$$

where k is the identification index of each solvent class, I_k is the set of indexes identifying each solvent included in the kth class, j counts over the p basis variables, and x_{ij} is the jth basis variable for the ith member of the kth solvent class. x_{kj} is defined as

$$x_{kj} = [\sum_{i \in I_k} \chi_{ij}] / n_k$$

where n_k is the number of solvents in the kth class.

The dispersion D (total active variance) at a given partition level is the sum of the dispersions of each class.

$$D = \sum_k D_{I_k}$$

According to the Ward criterion, the partition is optimum when D is minimized.

The taxonomy method used in this work is nondescendant and nonhierarchical. The decision to further partition from n to $n + 1$ classes follows from the identification of the class having the highest dispersion. The full impact of the partition decision is, however, achieved by transferring, as necessary, some elements from one previous class to another.

The number of classes is chosen after examination of the curve giving the total active variance, D, as a function of the number of classes (Figure 5). Alternately, the correct number of classes is considered reached when the tree of decisions becomes hierarchical: at this stage new decisions do not transfer elements from one class to another class.

The discriminating power of each variable at each partition decision measures the weight given that variable on the subsequent partition; it is derived from the sum of the variances associated with that variable over the p classes. The discriminating power DP of a variable for a given number of classes is given by

$$\mathrm{DP} = 1 - \frac{\sum_k n_k \sigma_{jk}^2}{N \sigma_j^2}$$

where k is the number of classes, n_k is the number of individuals in a given class, N is the total number of individuals, σ_{jk}^2 is variance of the variable j inside the class k, and σ_j^2 is variance of the variable j in the whole population. When the number of classes increases, the discriminating power tends toward 1. Its variations are generally of interest only for the first partitions.