(19) C. Hansch, J. E. Quinlan, and G. L. Lawrence, *J. Org. Chem.*, **33**, 347 (1968); A. Leo, C. Hansch, and C. Church, *J. Med. Chem.*, **12**, 766 (1969); A. Cammarata, S. J. Yan, and K. S. Rogers, *ibid.*, **14**, 1211 (1971).

(20) A. Leo, C. Hansch, and P. Y. C. Yow, *J. Med. Chem.*, **19**, 611 (1976). See also ref 28.

(21) E. R. Malinowski in ref 6a, Chapter 3, p 53.

(22) S. Wold, *Technometrics*, in press; S. Wold and M. Sjostrom in ref 6a, Chapter 12, p 243.

(23) P. H. Weiner, *Chemtech*, **7**, 321 (1977). Studies were carried out at the University of Pennsylvania Medical School Computer Facility.

(24) Inspection of Figure 1 will show that, unless data are standardized, the use of widely different measurement scales for different properties will tend to introduce spurious structure. For example, if boiling points were recorded in hundreds of degrees rather than degrees, variance along the boiling-point axis would virtually disappear. See discussion of "autoscaling" in ref 6c.

(25) R. A. Pierotti, *Chem. Rev.*, **76**, 717 (1976).

(26) D. G. Howery in ref 6a, Chapter 4, p 73. See also ref 6-12. By the criteria in ref 7c, our procedure of factor analysis would be characterized by the following quote: "the best way to obtain correct parameters is to . . . narrow the scope of the study to a full subset having no missing data and then use principal components analysis followed by a valid transformation." According to these authors' findings, popular methods of transformation can give physically absurd results. This seems to justify our decision to perform no transformation whatsoever, beyond the principal components analysis described.

(27) F. M. Richards, *Annu. Rev. Biophys. Bioeng.*, **6**, 151 (1977), gives an instructive discussion of the difficulties in defining "molecular volume", one plausible major component of **BCDEF** space.

(28) An example of this phenomenon is cited by Harman (ref 7a). Given a set of data on the falling times of various balls through various media, the factor analyst presumably would discover that two variables correlate the observations. These two variables would not be identical with weight and volume, however, because the weights and volumes of balls are partially correlated. Instead one variable would probably be weight, but the second would be "volume corrected for weight".

(29) The earliest reference is to M. Randic, *J. Am. Chem. Soc.*, **97**, 6609 (1975), and a recent one to T. DiPaolo, L. B. Kier, and L. H. Hall, *J. Pharm. Sci.*, **68**, 39 (1979). A review is L. B. Kier and L. H. Hall, "Molecular Connectivity in Chemistry and Drug Research", Academic Press, New York, 1976.

(30) C. Tanford, "The Hydrophobic Effect", Wiley, New York, 1973, and references cited therein.

(31) P. Mukerjee, *Adv. Colloid Interface Sci.*, **1**, 241 (1967); O. W. Howarth, *J. Chem. Soc., Faraday Trans. 1*, **71**, 2303 (1975); R. A. Wolfenden and C. A. Lewis, *J. Theor. Biol.*, **59**, 231 (1976); R. D. Cramer, III, *J. Am. Chem. Soc.*, **99**, 5408 (1977); K. Shinoda, *J. Phys. Chem.*, **81**, 1300 (1977); J. H. Hildebrand, *Proc. Natl. Acad. Sci. U.S.A.*, **76**, 194 (1979). M. H. Abraham, *J. Am. Chem. Soc.*, **101**, 5477 (1979), has very recently responded to the Cramer and Wolfenden criticisms. In brief reply, the central issue should perhaps be "Are there any experimental data which *require* a 'hydrophobic effect'?" instead of "Can the experimental data be manipulated so as to *allow* postulation of a 'hydrophobic effect'?" More specifically, Abraham asserts "hydrophobicity" to be an attribute of hydrocarbon but *not* of the completely apolar rare gases. Of what value can such a construct be?

BC(DEF) Parameters. 2. An Empirical Structure-Based Scheme for the Prediction of Some Physical Properties[†]

Richard D. Cramer, III

Contribution from the Department of Chemistry, Research and Development, Smith Kline and French Laboratories, Philadelphia, Pennsylvania 19101.

Received June 11, 1979

Abstract: Based on either a hierarchically organized additive-constitutive model or a subset of four physical properties, for calculation of intermediate **BC(DEF)** values where **BCDEF** are the principal components of a matrix of six physical properties of 114 compounds, all experimental values of 18 common physical properties for 139 additional compounds of diverse structure have been "predicted". The rms difference between the 1142 predicted and experimental values is 22% of the variance in the experimental values, corresponding to a "correlation coefficient" or "r" of 0.88. For the 118 compounds and 10 properties to which application of the **BC(DEF)** model is clearly warranted, the rms difference between the 749 predicted and actual values is 6% of the overall variance; that is, the "r" is 0.97. Predictions using the **BC(DEF)** model are at least as accurate as those of existing additive-constitutive models for individual properties. There is no significant difference in predictive accuracy between **BCDEF** values derived from the additive-constitutive model and **BCDEF** values derived from the property subset. The five-factor **BCDEF** model is more accurate than the two-factor **BC** model for compounds having reasonable structural similarity to any of the 114 used to derive the **BCDEF** scale, but the two-factor model is the less likely to give completely misleading results for very different structures.

In the preceding paper,[1] analysis of a collection of physical-property data for a variety of pure liquid compounds showed that more than 95% of the variance in most of the properties can be explained in terms of a two-, three-, or five-component "**BC(DEF)**" model, where the components are derived by factorization of a matrix constructed from the values of activity coefficient, partition coefficient, boiling point, molar volume, refractivity, and heat of vaporization for 114 compounds. In this paper, the generality and utility of this model will be investigated by "predicting" the experimentally known properties of 139 compounds not among the 114 used for derivation of the model.

Prediction of a property using the **BC(DEF)** scheme has two steps: (1) calculation of the **BC(DEF)** values for the compound, either from previously known properties or from its structure alone; (2) calculation of the property, from the **BC(DEF)** values and the appropriate previously derived "property equation" (Table IV[1]).

Although structurally based schemes have been proposed for calculating some of the physical properties encompassed by the **BC(DEF)** models,[2] little attention has been given to scope and limitations. One notable exception is Exner's discussions of the significance of the long-known additive-constitutive behaviors of molar volume and parachor.[3] Types of information which add to the utility of any predictive scheme include answers to the following questions: (1) What kinds of molecules (and properties) can the scheme confidently be applied to? (2) What must be known about a molecule in order to calculate an unknown property? (3) How accurate are the results? These questions provide an outline for the following description of our data and methods.

Scope of the BC(DEF) Model. In choosing the 139 compounds whose properties were to be predicted, the major objectives were a large number of examples of values for the rarer properties and a structurally diverse data set. The completed

---

[†] Presented in part at the 177th National Meeting of the American Chemical Society, Honolulu, Hawaii, 1979.

0002-7863/80/1502-1849$01.00/0

Table I. Compounds and Properties Whose Values Have Been Predicted Using the **BC(DEF)** Scheme[a]

CLASS I COMPOUNDS

	B	C	D	E	F	ID #	1 AC	2 PC	3 MR	4 BP	5 MV	6 HVP	7 X	8 CT	9 VDW A	10 VDW B	11 E	12 SLP	13 CP	14 STN	15 TCD	16 VIS	17 CMP	18 ET	19 U	20 MP	21 MW
2-METHYLBUTANE	-0.001	-0.211	0.021	-0.022	0.002	1	+		=	--	=	=	=	-	=	=	=		=	=		+			=	=	+
HEXANE	0.082	-0.248	0.018	-0.029	0.012	2	=		=	--	=	=	=	--	-	-	=	-	=	-	-	++	--	=		=	=
2-METHYLPENTANE	0.075	-0.249	0.032	-0.026	0.005	3	=		=	--	+	--	=	-					=							=	=
2,3-DIMETHYLBUTANE	0.068	-0.250	0.047	-0.023	-0.002	4		--	=	--	++	-	=	--					=							=	=
METHYLCYCLOPENTANE	0.049	-0.186	-0.016	-0.017	0.000	5	=		=	--	=	--	=	-					=						=	+	+
ETHYLCYCLOPENTANE	0.125	-0.224	-0.006	-0.020	0.003	6			=	--	=	--		-					=							+	=
METHYLCYCLOHEXANE	0.125	-0.224	-0.006	-0.020	0.003	7	=		=	--	=	-	=	-				=	=		=					=	=
CIS-2-BUTENE	-0.076	-0.138	0.001	-0.004	0.001	8			=	--	+	--	+	-					=							=	+
1-PENTENE	0.001	-0.176	0.012	-0.008	0.005	9	=		++	--	++	-		-			=	=	=							=	+
1-HEXENE	0.077	-0.215	0.023	-0.011	0.008	10	=		++	--	+	-	+	-				-	=							=	=
CYCLOHEXENE	0.050	-0.152	-0.025	-0.002	0.003	11	--	++	++	--	++		+	-			=								=	=	+
1,4-PENTADIENE	-0.005	-0.143	0.018	0.010	0.001	12	=	++	++	-	+	-													=	=	=
BIALLYL	0.071	-0.181	0.029	0.006	0.005	13		++	++	--	++		++	-					=							=	=
1-PENTYNE	-0.041	-0.078	0.039	0.004	0.008	14	+	-	=	=	+														=	=	=
1,2,3-TRIMETHYLBENZENE	0.255	-0.159	0.001	0.008	0.003	15			=	--	++	-		=					=							=	=
MESITYLENE	0.255	-0.159	0.001	0.008	0.003	16		++	=	=	=	=	=		=	=	=	=			=					=	=
1,2,3,5-TETRAMETHYLBENZENE	0.331	-0.197	0.012	0.005	0.007	17			=	--	++	--		=					=							=	=
DURENE	0.331	-0.197	0.012	0.005	0.007	18		++	=	--	--	--	=		-	--									=	--	=
P-CYMENE	0.324	-0.198	0.026	0.008	-0.000	19			=	+	+	-	=	=	=	--		=	=		=					=	=
2-BUTYLBENZENE	0.324	-0.198	0.026	0.008	-0.000	20			-	++	+	-	=		+	=										=	=
STYRENE	0.173	-0.088	-0.005	0.030	-0.004	21		=	=	--	++	=	+	=			=	=	=							=	=
1-METHYLNAPHTHALENE	0.401	-0.057	-0.036	0.032	-0.015	22		--	-	++	=		=	++					+							++	=
CHLOROCYCLOHEXANE	0.155	-0.116	-0.027	-0.013	0.004	23			=	--	-															=	=
FLUOROBENZENE	0.029	-0.075	-0.029	0.008	-0.002	24		++	+	--	=	=	=	=	=	=			=			=			=	=	=
IODOBENZENE	0.253	-0.059	-0.075	0.016	-0.013	25		+	=	++	=														=	+	-
METHYL ETHYL ETHER	-0.103	0.021	0.046	0.002	-0.020	26			=	+	+	=		=	=	=			=						=		=
METHYL PROPYL ETHER	-0.027	-0.017	0.057	-0.002	-0.017	27	--		=	=	+	--															=
DIISOPROPYL ETHER	0.111	-0.096	0.108	-0.004	-0.023	28	++		=	++	=	=	-	=			=		=		++					=	=
ETHOXYBENZENE	0.213	-0.066	0.036	0.020	-0.016	29		+	=	--	++	--	=	=	=	-			=	-	=					=	=
3-METHYLBUTANOL	0.135	0.095	0.036	-0.018	0.023	30		+	=	+	+	--		=											=		=
1-HEPTANOL	0.295	0.020	0.044	-0.028	0.036	31	=		-	++	+	--		=					=		=	=	=			=	=
M-CRESOL	0.271	0.163	-0.058	-0.008	0.003	32		--	=	++	--	++	=	++	+	=	+		++		++	--				+	=
2-NAPHTHOL	0.452	0.188	-0.047	0.028	0.009	33		--		++	++	++	=													=	=
2-PENTANONE	0.065	0.074	0.057	0.002	-0.016	34	=	=	=	++	=	--	=					=							-	=	=
3-PENTANONE	0.065	0.074	0.057	0.002	-0.016	35			=	++	=	--	=	=			-		=			+				=	=
PROPIONIC ACID	0.037	0.226	-0.011	-0.014	0.014	36	-	--	++	--	=	--	=	=	-	=	++	++	=	=		=			=	=	=
PENTANOIC ACID	0.233	0.149	0.003	-0.032	0.013	37			++	++	++	+	=	++			++		+							=	=
HEXANOIC ACID	0.309	0.111	0.014	-0.036	0.017	38			++	++	++	--	=													=	=
METHYL FORMATE	-0.145	0.144	0.026	0.007	-0.022	39	++		=	-	=	=	=	=	=	=	=	=	=	=					=	=	=
ETHYL FORMATE	-0.069	0.106	0.037	0.003	-0.018	40	++		=	=	=	=	=	=	=	=	=	=	=	=		=			=	=	=
ISOBUTYL FORMATE	0.076	0.029	0.073	-0.001	-0.018	41	++		=	=	=	=	-		=	=										=	=
N-PENTYL FORMATE	0.160	-0.008	0.070	-0.008	-0.008	42			=	=	=			=	=	=			=							=	=
BUTYL ACETATE	0.160	-0.008	0.070	-0.008	-0.008	43	+		=	+	=		=	=			=	=	=		+	+				=	=
ISOBUTYL ACETATE	0.153	-0.009	0.084	-0.005	-0.015	44	++		=	++	=	+	-		=	-										=	=
METHYL PROPIONATE	0.008	0.068	0.048	-0.001	-0.015	45	=		=	=	=	=	=	=	=	=	=	=	=							=	=
METHYL BUTYRATE	0.084	0.030	0.059	-0.004	-0.011	46	=		=	+	=	+	=	=	=	=	=		=							=	=
ETHYL BUTYRATE	0.160	-0.008	0.070	-0.008	-0.008	47	+		-	++	=	++	-	+	=	--	=		=							=	=
2-AMINOPROPANE	-0.053	0.116	0.063	0.032	-0.001	48		-	+	=	-			-												=	=
BUTYLAMINE	0.031	0.079	0.060	0.025	0.009	49	--	=	++	--	=		=	--											=	=	=
PENTYLAMINE	0.107	0.041	0.071	0.021	0.013	50	-	-	+	--	=														=	=	=
DIMETHYL AMINE	-0.141	0.155	0.068	0.042	-0.012	51	=		++	--	+	-		=	=	=	=		-	=					=	=	=
DIPROPYLAMINE	0.164	0.003	0.112	0.027	0.002	52	=	--	+	=	=			=	=	=			-						=	=	=
PYRROLIDINE	-0.015	0.142	0.042	0.043	-0.010	53	--	--	++	--	--		-												=		=
3-METHYLPYRIDINE	0.113	0.101	0.014	0.021	-0.004	54	-	-	=	=	=		=	=											=	=	=
4-METHYLPYRIDINE	0.113	0.101	0.014	0.021	-0.004	55	-	-	=	=	=		=	=											=	=	=
2-METHYLPYRIDINE	0.113	0.101	0.014	0.021	-0.004	56	=	=	=	++	=		=	=											=	+	=
2-ETHYLPYRIDINE	0.190	0.063	0.025	0.018	-0.000	57	=	=	=	++	+															+	=
3,4-DIMETHYLPYRIDINE	0.190	0.063	0.025	0.018	-0.000	58	--		--	+	=		=	=													=
2,6-DIMETHYLPYRIDINE	0.190	0.063	0.025	0.018	-0.000	59	=	=	-	++	=		=	++										+		=	=
BUTYRONITRILE	0.029	0.117	0.005	-0.022	-0.021	60	=		=	++	=	=	=		--	--		=							-	+	=
TRIFLUOROMETHANE	-0.291	-0.021	0.014	-0.024	-0.004	61	=	=		--		=		-					=						=	=	-
CHFCL2	-0.118	-0.022	-0.015	-0.011	-0.006	62		-	=	=	-	++	=	=			=	=	=	+	=	=			=	=	=
METHYLENE BROMIDE	-0.013	0.040	-0.050	0.007	-0.007	63	=		=	--	+	=	--				=					-	+		=	=	--
BROMOFORM	0.125	0.060	-0.056	0.014	-0.007	64	++		-	++	=	++	--				+			-		-			+	--	--
CF3CH3	-0.215	-0.059	0.025	-0.028	-0.000	65				--				-											-	-	-
CF3CF3	-0.214	-0.199	0.070	-0.056	-0.003	66			=	--	++	--		--					=		=				=	--	--
CF3CF2CL	-0.128	-0.199	0.055	-0.049	-0.005	67	--		=	--	--	=		-					=	=		=			=	--	--
CF3CCL3	0.045	-0.200	0.026	-0.036	-0.008	68	-		+	+	++	++		=			=		=	=	++	=			=	--	--
CFCL2CFCL2	0.132	-0.201	0.011	-0.030	-0.009	69	--		+	++	++	++		=				=	=	=	+	=			=	--	--
ALLYL CHLORIDE	-0.053	-0.032	-0.006	0.006	-0.005	70	=		+	--	+	-	=	-					=			=			-	+	=
1,3-DICHLOROPROPANE	0.052	0.004	-0.008	-0.005	-0.004	71	=	--	=	--	=														=	+	=
CHCL2CHCL2	0.149	-0.052	-0.018	-0.007	-0.012	72	--		+	+	++	=	--							-	++	-	+		=	=	-
VINYL CHLORIDE	-0.134	-0.043	-0.024	0.004	-0.005	73	+		++	++	=	-						=							=	+	=
CHCL=CHCL	-0.040	-0.024	-0.028	0.005	-0.005	74	-		++	+	+	=	=				-						=		=	=	=
CCL2=CCL2	0.129	-0.101	-0.038	-0.003	-0.000	75	++		++	++	++	+	-					=		=	-		=		+	=	-

[a] For every compound are given (1) its **BCDEF** values computed by the additive-constitutive scheme of Table IV and (2) the deviation between the predicted and actual values of those of its properties which have been experimentally determined. = indicates that the difference between predicted and experimental value is less than *s*; + indicates a predicted value exceeding the experimental value by more than 1*s* but less than 2*s*; ++ indicates a predicted value exceeding the experimental value by more than 2*s* (or the 95% confidence interval of a prediction); – indicates

list was divided into three classes, of decreasing similarity to the original 114 and consequently of expected decreasing accuracy in property prediction. These classes are (I) compounds which are either isomers or one- or two-carbon homologues[4] of the original 114; (II) compounds not in class I which contain no more than one structural fragment or grouping of structural fragments not found among the original 114; (III) all other compounds. The first column of Table I lists the 139 compounds by class.

Even the class I compounds represent enough structural dissimilarity from the compounds of Table I in the preceding paper to make their property prediction a significant challenge for the **BC(DEF)** model. Less than half of these are isomers, representing a true interpolation of the model. On the other hand, the properties of six alkylpyridines must be inferred from pyridine itself, and all of the polyhaloalkanes with their notoriously irregular properties fall into class I.

At this writing there were 21 properties for which **BC(DEF)** equations had been derived (Table IV[1]). However, the last three of the 21, dipole moment, melting point, and molecular weight, do not depend upon nonspecific intermolecular interactions in the liquid state, so a low accuracy in predicting these properties was expected and found.

Obtaining BC(DEF) Values from Property Data. Inasmuch as the original 114 **BC(DEF)** values[1] are simply the results of factorization of a somewhat arbitrarily constituted matrix, the definition of **BC(DEF)** values for other compounds is not a trivial problem. In general the only test of the "accuracy" of a **BC(DEF)** value will be its ability to reproduce measured physical properties. This situation is not unique in chemistry; for example, the accuracy of a wave function also can be assessed only by its ability to reproduce derived observations. A

Table I (*Continued*)

CLASS II COMPOUNDS

	B	C	D	E	F	ID #	1 AC	2 PC	3 MR	4 BP	5 MV	6 HVP	7 X	8 CT	9 VDW A	10 VDW B	11 E	12 SLP	13 CP	14 STN	15 TCD	16 VIS	17 CMP	18 ET	19 U	20 MP	21 MW
HEPTANE	0.159	-0.286	0.029	-0.032	0.015	1			=	--	+	=	=	--	--	--		-	=		-	++	-			=	=
OCTANE	0.235	-0.324	0.040	-0.036	0.019	2	+		=	--	++	=	=	--			=	--	-	--	-	++	=			-	=
2,2,3,3-TETRAMETHYLBUTANE	0.195	-0.327	0.091	-0.017	0.006	3			+	--		--		--				--	--							--	=
DECANE	0.388	-0.400	0.062	-0.044	0.026	4			=	--	++		=	-	-	--	-		-		=		=			-	=
HEXADECANE	0.846	-0.628	0.127	-0.066	0.046	5			-	++	++	=	-									++	-			-	=
CYCLOPROPANE	-0.173	-0.071	-0.064	-0.008	-0.004	6		+	=	++	--	-	=	+											+	=	+
DECALIN	0.320	-0.276	-0.007	-0.019	0.002	7			=	--	=	++	=	=				-	=							=	=
1-HEPTENE	0.153	-0.253	0.034	-0.015	0.012	8			++	--	++			-			=									=	=
ALLENE	-0.158	-0.067	-0.004	0.017	-0.006	9				+		=		=					=						=	=	+
METHYLALLENE	-0.082	-0.105	0.007	0.014	-0.002	10			++	--	++	--		-					=						=	=	+
CYCLOPENTADIENE	0.009	-0.041	-0.067	0.002	-0.025	11			++	++	--														+	+	++
ANTHRACENE	0.624	0.007	-0.062	0.052	-0.029	12	++	--		++		=	=												++	-	+
PHENANTHRENE	0.624	0.007	-0.062	0.052	-0.029	13	++	--		++		++	=					++								+	+
DIPHENYL	0.428	-0.071	-0.016	0.045	-0.006	14		--	+	++	--	++	=	++	=	--	=		+			=			+	=	=
DIPHENYLMETHANE	0.504	-0.109	-0.005	0.042	-0.002	15		+	=	++	--	++	=		++	++										+	=
DIPHENYL ETHER	0.453	-0.115	0.015	0.042	-0.015	16		--	=	++	++	++	=											--		=	=
ETHYLENE OXIDE	-0.206	0.122	-0.014	0.014	-0.029	17		+	=	-	--	--	=	=			=		=	=		=			=	=	=
PROPYLENE OXIDE	-0.137	0.083	0.012	0.014	-0.032	18			=	-	--	--	-	=				=	=						=		=
THYMOL	0.451	0.008	0.026	0.000	0.027	19		-	--	++	--	++		++					=							=	=
EUGENOL	0.528	0.137	0.027	0.005	-0.006	20			=	++	++	++	=									=				++	=
ACETALDEHYDE	-0.146	0.168	0.000	-0.005	-0.015	21			=	++	=	=	+	+			=		=	+	=	+			=	=	=
HEXANAL	0.159	0.016	0.044	-0.020	-0.001	22	+		-	++	=															=	=
2-HEXANONE	0.141	0.036	0.068	-0.002	-0.013	23		=	=	++	=	--														=	=
ISOPENTYL ACETATE	0.229	-0.047	0.095	-0.009	-0.011	24	++		=	--	=	=		=					=		++	+			=	=	=
HEXYLAMINE	0.184	0.003	0.082	0.018	0.016	25	-	-	+	--	+			-											=	-	=
DIBUTYLAMINE	0.317	-0.073	0.134	0.019	0.009	26	=	-	+	=	++		-	=											=	=	=
TRIETHYLAMINE	0.142	-0.040	0.115	0.022	0.007	27	-	++	=	--	=		-	-	=	=			-				++		=	=	=
TRIPROPYL AMINE	0.371	-0.154	0.148	0.011	0.018	28		++	-	=	+			=												=	-
ANILINE	0.118	0.106	0.017	0.054	0.003	29		++	+	--	++	--	=	--	=	+	=	=	=	--	--	--	++	--	=	+	=
P-TOLUIDINE	0.194	0.068	0.028	0.051	0.007	30		++	=	--	=		=				=			-		-				-	=
N-METHYLANILINE	0.175	0.068	0.058	0.060	-0.008	31			+	--	++	--	-	--			=		-	--	--	--				+	=
N-ETHYLANILINE	0.251	0.030	0.069	0.056	-0.004	32		--	=	--	++	--	-	-						++		--				+	=
N,N-DIMETHYLANILINE	0.229	-0.013	0.072	0.051	0.002	33		--	-	--	++	--	-	--	-	=		-	-	--		-				=	=
DIPHENYL AMINE	0.491	-0.019	0.048	0.078	-0.003	34		-		--	++	--	=	-			=					+				=	=
QUINOLINE	0.336	0.165	-0.012	0.041	-0.018	35		--	--	++	--	++	=	++			+	++		=				+	=	++	=
2-METHYLPYRAZINE	0.115	0.228	0.048	0.022	-0.001	36	++	--	=	++	++																=
NITROMETHANE	-0.127	0.165	-0.022	-0.019	-0.024	37		++	--	--	=	--	+	--				--	=	--	--	=		=	-	-	=
METHANETHIOL	-0.134	0.079	-0.036	-0.002	-0.001	38	++			++	=	++		=					-						=	=	=
THIOMETHYLBENZENE	0.229	-0.053	-0.008	0.037	-0.017	39	-	=	=	--	++		=												=		=
ACROLEIN	-0.080	0.113	0.009	0.003	-0.011	40			=	=	++	=													-	=	=
CHLOROACETIC ACID	0.055	0.216	-0.022	-0.013	0.004	41			++	--	++	--	=				+			-						-	=
DICHLOROACETIC ACID	0.143	0.198	-0.021	-0.014	0.003	42			++	=	++	-	=				++	+		=						=	-
PYRUVIC ACID	0.089	0.364	0.034	0.001	-0.015	43			++	++	++	++														=	=

CLASS III COMPOUNDS

	B	C	D	E	F																						
EPICHLORDHYDRIN	-0.033	0.094	-0.001	0.012	-0.030	1			=	--	--	--		=												=	=
GLYCEROL	0.251	0.666	0.040	-0.012	0.055	2	++		=	++	++	++	-	--			++	+		--	-	--	-			+	-
PARALDEHYDE	0.135	0.104	0.193	0.034	-0.073	3			=	++	++	--	--	++			-			=					+	-	-
CARBON DISULFIDE	-0.299	0.071	0.128	0.125	0.031	4			--	--	=	--	--		--	=	--	--		--		--	++	--	-	-	--
ALLYL ISOTHIOCYANATE	-0.003	0.090	0.129	0.076	0.020	5			--	--	++	--					--			--						=	--
XENON	-0.211	-0.017	-0.081	0.038	0.010	6		++		++	=			++	++	+			=						=	=	--
GERMANIUM TETRACHLORIDE	-0.024	0.007	-0.137	0.042	-0.002	7			--	++	--			++	--	--	+		++						+	+	--
AMMONIA	-0.363	0.414	0.197	0.144	0.040	8	++		++	--	++			--			--	--	--		+	--			-	--	--
ACETIC ANHYDRIDE	0.042	0.153	0.039	0.001	-0.013	9			=	--	--	--			=	=	=	=		--	--	=				=	=
OLEIC ACID	1.177	-0.311	0.160	-0.052	0.062	10			++		++	++	-							-	=	++				++	--
NITROGYLCERIN	0.293	0.166	0.159	-0.019	0.015	11			--	--	++															-	--
HYDRAZINE	-0.214	0.442	0.117	0.103	0.021	12				--	++			--			--			--		--			=	--	-
CYANOGEN	-0.042	0.206	-0.029	-0.040	-0.053	13			++	++	++	++	++	++	++	++									++	=	+
FURAN	-0.031	0.011	-0.003	0.020	-0.030	14		=	++	++	++		+	++			+		=						+	=	+
THIOPHENE	0.061	-0.014	-0.036	0.033	-0.028	15		++	++	++	++	-	+	++	+	+	+		=						+	=	+
THIAZOLE	0.015	0.128	-0.000	0.040	-0.020	16		+	=	--	++		=												=		=
ETHYLENE DIAMINE	-0.028	0.229	0.096	0.064	0.002	17	--		++	--	++	--	=												=	-	=
NITROUS OXIDE	-0.376	-0.072	-0.015	0.019	0.013	18	--	++	--	--		--	=	--	--	=			--						=	--	=
N-METHYLACETAMIDE	-0.006	0.401	0.129	0.065	-0.033	19	=	--	+	--	+														=	-	-
DIMETHYLACETAMIDE	0.048	0.320	0.143	0.056	-0.024	20	+	--	=	--	++														-	=	-
ETHYLENE GLYCOL	0.043	0.340	-0.003	-0.020	0.035	21	++	++	++	--	++	--	=				=	-		--	--	--	++	++		=	=

a predicted value between $1s$ and $2s$ lower than the experimental value; - - indicates a predicted value more than $2s$ lower than the experimental value; where s is the standard error of estimate for the corresponding property equation in Table IV of the preceding paper[1] (which see for s values).

second problem is that the following methods of obtaining **BC(DEF)** values for additional compounds implicitly assume negligible error in the 6 × 114 matrix's experimental values. However, some justification for this assumption is the previously described[1] collinearity of the 6 × 114 matrix **BC(DEF)** eigenvectors with the vectors obtained from factorization of two other matrices, suggesting that existing **BC(DEF)** values will not be perturbed significantly by the addition of new data to the defining matrix.

Because the original **BC(DEF)** values[1] were defined in terms of a compound's properties, rather than its structure, we at first believed that a compound's unknown properties would be most accurately predicted when the intermediate **BC(DEF)** values are defined by a set of its known properties. One approach would be a regression study of the original data (Table I[1]), which would yield linear equations which predict **B, C, D, E,** or **F** values as functions of various properties from among activity and partition coefficients, boiling point, and molar refractivity, volume, and heat of vaporization.

Table II shows some properties of those of the 126 possible sets of linear equations which involve four or five of the six properties. The s value, the amount of variance in **B, C, D, E,** or **F** left unexplained by an equation, is shown for all such property combinations, along with an average of s over the six **BC(DEF)** equations which is weighted to reflect the far greater importance of **B** and **C** in predicting properties. The most useful or informative grouping of properties will have a relatively low weighted mean s. Comparison of the weighted mean s values for the six five-property equations suggests that the single property most important for **BC(DEF)** estimation is molar volume. The five-property block of Table II also gives some insight into the significance of the data-set-dependent **E** and **F** eigenvectors. The correlation coefficients of the **E** equations are depressed by exclusion of refractivity or molar volume,

Table II. The Fit between the **BC(DEF)** Values and Property Data Found in Table I in the Preceding Paper,[1] When Various of the Six Defining Properties Are Missing

	Missing One Property											
fit	1 (AC)		2 (PC)		3 (MR)		4 (bp)		5 (MV)		6 (ΔH_{vap})	
to	r^2	s (×10^2)	r^2	s (×10^2)	r^2	s (×10^2)	r^2	s (×10^2)	r^2	s (×10^2)	r^2	s (×10^2)
B	0.999	0.5	0.999	0.6	0.997	1.0	0.997	0.9	0.993	1.6	0.998	0.8
C	0.991	1.3	0.991	1.3	0.999	0.4	0.999	0.4	0.994	1.0	0.998	0.5
D	0.984	0.5	0.927	1.1	0.989	0.4	0.958	0.8	0.479	2.9	0.979	0.4
E	0.905	0.6	0.996	0.1	0.403	1.6	0.952	0.4	0.633	1.2	0.870	0.7
F	0.823	0.7	0.726	0.8	0.976	0.3	0.229	1.4	0.998	0.0	0.747	0.8
weighted mean[a]		0.7		0.8		0.8		0.8		1.5		0.7

	Missing Two Properties (s × 10^2) Values														
	1,2	1,3	1,4	1,5	1,6	2,3	2,4	2,5	2,6	3,4	3,5	3,6	4,5	4,6	5,6
B	0.6	1.3	1.1	1.6	1.4	1.8	1.2	1.7	1.3	1.6	3.1	1.2	1.8	2.1	1.9
C	4.9	1.2	1.4	1.7	1.9	1.9	1.3	1.6	1.3	0.5	1.7	0.7	1.1	1.1	1.1
D	2.5	0.5	0.8	2.8	0.6	1.2	1.3	3.1	1.3	0.8	3.8	0.6	3.0	1.4	2.8
E	1.3	2.0	0.7	1.5	0.8	1.9	0.5	1.2	0.8	1.6	1.6	1.8	1.3	1.4	1.5
F	0.9	0.7	1.5	0.7	1.6	0.9	1.6	0.9	1.4	1.6	0.3	0.9	1.4	1.4	0.8
weighted mean[a]	2.0	1.2	1.2	1.7	1.5	1.8	1.2	1.7	1.3	1.2	2.6	1.0	1.6	1.8	1.7

[a] Weighted by **B** = 0.64, **C** = 0.31, **D** = 0.03, **E** = 0.01, **F** = 0.005 (eigenvalues of original factorization[1]).

Table III. Regression Equations Expressing the Original **B, C, D, E,** and **F** Values[1] as Functions of a Compound's Partition Coefficient, Molar Refractivity, Boiling Point, and Molar Volume[a]

	r	s
$\mathbf{B} = -0.3123 - 0.002(\pm 0.004)\mathrm{PC} + 0.00552(\pm 0.00109)\mathrm{MR} + 0.000968(\pm 0.000060)\mathrm{bp} + 0.00138(\pm 0.00022)\mathrm{MV}$	0.997	0.014
$\mathbf{C} = 0.1688 - 0.094(\pm 0.006)\mathrm{PC} + 0.00056(\pm 0.00151)\mathrm{MR} + 0.000825(\pm 0.000084)\mathrm{bp} - 0.00095(\pm 0.00031)\mathrm{MV}$	0.989	0.019
$\mathbf{D} = -0.1370 - 0.042(\pm 0.002)\mathrm{PC} + 0.00090(\pm 0.00045)\mathrm{MR} - 0.000416(\pm 0.000025)\mathrm{bp} + 0.00236(\pm 0.00009)\mathrm{MV}$	0.979	0.006
$\mathbf{E} = -0.0227 - 0.019(\pm 0.002)\mathrm{PC} + 0.00768(\pm 0.00068)\mathrm{MR} - 0.000340(\pm 0.000034)\mathrm{bp} - 0.00113(\pm 0.00012)\mathrm{MV}$	0.925	0.008
$\mathbf{F} = -0.0049 - 0.0002(\pm 0.005)\mathrm{PC} + 0.00043(\pm 0.00124)\mathrm{MR} - 0.000049(\pm 0.000069)\mathrm{bp} - 0.000015(\pm 0.00011)\mathrm{MV}$	0.179	0.016

[a] Values in parentheses are 95% confidence intervals. All equations are based on n = 114.

suggesting that the **E** eigenvector of the 6 × 114 matrix serves primarily to bring any disparate values of molar volume and molar refractivity into line with one another. Since molar refractivity is defined as molar volume times a fraction which is roughly proportional to the square of the velocity of light passing through the substance, $\mathbf{E}_6$ might be regarded as a "measure" of the relative velocity of light, that is, the mean electronic density. By a similar argument, the $\mathbf{F}_6$ parameter seems to serve only to align boiling point with the trends in all the other variables, and thus embodies the relatively structure-specific aspects of boiling point.

For comparison with predictions based on "additive-constitutive" **BCDEF**'s, a single set of equations for "property-derived" **BCDEF**'s was chosen from Table II. The set, labeled "1,6", comprises the four properties partition coefficient, molar refractivity, boiling point, and molar volume, a combination which has been measured for a relatively large number of the compounds in Table I. The coefficients needed to estimate **BC(DEF)** values from these properties appear in Table III.

Obtaining BCDEF Values from Structure Alone. The Additive-Constitutive BC(DEF) Models. Certainly structure is the most generally useful basis for estimation of a molecular property. The simplest such structural basis is the summation of contributions of individual fragments; for example, molecular weight is computed by summing atomic nuclear contributions. However, when other properties are being estimated, it has usually been found that such a purely "additive" model can be readily improved by the inclusion of "constitutive" contributions. Thus in most additive-constitutive models an ester group is treated differently either from "two oxygens and a carbon" or from "a carbonyl and an ether". It is also recognized[5] that the selection of descriptors is fundamentally an arbitrary, ad hoc procedure. Objectives in selecting **BC(DEF)** additive-constitutive descriptors were the desire to say at least something about any molecule, regardless of how novel its fragments might be; unambiguous fragment definitions which would readily be applicable to a computer perception process[6] but also convenient for human perception; and the minimum number of fragment definitions possible, to maximize the degrees of freedom in the calculation of fragment contributions. These objectives led to purely "hierarchical" fragment definitions, such that fragments at a lower hierarchy can be thought of as "corrections" to an approximation obtained higher in the hierarchy. The more usual "linear" type of additive-constitutive model can say nothing about a molecule which cannot be unambiguously decomposed into previously encountered fragments.

A hierarchical additive-constitutive model which was fitted by regression to the original 114 **BCDEF** values is presented in Table IV. Starting at its top, evaluation of molecular weight and enumeration of oxygen, nitrogen, and sulfur atoms, and of the various types of carbon-carbon, carbon-hydrogen, and carbon-halogen bonds, all present no perceptual problems. The redundancy in description which characterizes a hierarchical model becomes most evident among the oxygen and nitrogen functionalities next encountered. For example, an ester group possesses a molecular weight, two oxygen atoms, half of a C-C bond, a carbonyl moiety, and an ether moiety, as well as the ester moiety, and all of these effects must be considered in using Table IV to calculate the ester **BC(DEF)** parameter contributions. This point is illustrated by the left hand of the two sample "**B**" calculations in Table VI.

The last six features in Table IV require comment. Tertiary and quaternary carbons are defined as carbon atoms having four attachments, either three or four of which are not hydrogen or halogen atoms. The "X-C-X tracing" is a novel treatment of the well-known observation that successive replacement of hydrogens by halogen or other electronegative attachments seldom produces exactly additive effects on molecular properties. The various polarizations introduced by the

Table IV. Hierarchical Additive-Constitutive Models Found for the Original **BC(DEF)** Values[1],[f]

	B × 10^2	**C** × 10^2	**D** × 10^2	**E** × 10^2	**F** × 10^2
intercept	−50.647	−5.605	.663	3.12	2.80
coefficients					
mol wt	0.2251(±0.010)	0.0294(±0.012)	−0.0669(±0.011)	−0.0051(±0.0083)	−0.014(±0.0063)
no. of O atoms	0.82(±1.35)	−0.32(±1.56)	−0.31(±1.41)	−0.39(±1.11)	−0.28(±0.84)
no. of N atoms	1.17(±1.35)	−1.31(±1.57)	.55(±1.41)	−0.53(±1.11)	−0.30(±0.85)
no. of S atoms	2.31(±2.84)	8.10(±3.30)	8.67(±2.97)	4.73(±2.34)	0.58(±1.78)
no. of C—C bonds	−4.09(±0.99)	−3.91(±1.15)	3.69(±1.03)	1.67(±0.81)	2.11(±0.62)
no. of aromatic C—C bonds	−0.08(±0.61)	−1.71(±0.70)	2.77(±0.63)	1.27(±0.50)	1.08(±3.80)
no. of C═C bonds	4.36(±0.96)	−0.85(±2.15)	2.47(±1.00)	1.32(±0.79)	0.16(±0.60)
no. of C≡C bonds	9.26(±1.85)	8.80(±2.15)	0.34(±1.93)	0.34(±1.53)	−1.14(±1.16)
no. of C—H bonds	4.28(±0.41)	−0.15(±0.48)	−0.83(±0.43)	−1.05(±0.39)	−0.78(±0.26)
no. of C—F bonds	1.48(±0.90)	6.24(±1.05)	2.20(±0.94)	−1.17(±0.74)	−0.66(±0.56)
no. of C—Cl bonds	6.44(±0.73)	5.70(±0.85)	1.84(±0.76)	−0.59(±0.60)	−0.57(±0.46)
no. of C—**Br** bonds	1.30(±1.52)	5.23(±1.76)	3.89(±1.59)	−0.37(±1.25)	0.15(±0.95)
no. of C—I bonds	−0.38(±1.18)	4.64(±1.37)	4.78(±1.23)	−0.84(±0.97)	−0.24(±0.74)
no. of —OH groups	13.52(±1.47)	30.30(±1.71)	2.05(±1.54)	−0.35(±1.21)	1.84(±0.92)
no. of —O— groups	−4.14(±2.32)	11.39(±2.69)	11.20(±2.42)	3.89(±1.91)	0.44(±1.45)
no. of —(C═O)— groups	10.43(±1.89)	27.99(±2.19)	4.59(±1.97)	0.88(±1.55)	−3.56(±1.18)
no. of —CH═O groups	1.85(±2.78)	−2.04(±3.24)	−2.37(+2.91)	−1.80(±2.30)	1.23(±1.75)
no. of —COOH groups	−1.11(±2.36)	−22.99(±2.74)	−6.02(±2.47)	−2.71(±1.95)	1.61(±1.48)
no. of —COO— groups	−2.43(±2.02)	−19.86(±2.35)	−5.91(±2.11)	−2.53(±1.67)	2.67(±1.27)
no. of —SH groups	11.28(±3.65)	4.44(±4.24)	−7.22(±3.81)	−5.15(±3.01)	−0.41(±2.29)
no. of —$\mathrm{NH_2}$ groups	9.84(±2.17)	25.76(±2.52)	6.05(±2.27)	4.02(±1.79)	0.17(±1.36)
no. of —NH— groups	−0.46(±2.81)	22.02(±3.27)	13.60(±2.94)	7.68(±2.32)	1.62(±1.77)
no. of —N—— groups	−11.06(±4.25)	13.95(±4.95)	18.39(±4.45)	9.90(±3.51)	5.08(±2.67)
no. of ═N— groups	3.82(±3.09)	16.12(±3.59)	7.76(±3.23)	2.55(±2.55)	1.72(±1.94)
no. of —CN groups	19.11(±2.10)	29.38(±2.44)	−1.38(±2.20)	−2.72(±1.73)	−4.50(±1.32)
no. of —$\mathrm{CONH_2}$ groups	14.22(±3.30)	−1.59(±3.85)	−11.70(±3.46)	−8.91(±2.73)	2.34(±2.08)
no. of —$\mathrm{NO_2}$ groups	8.55(±3.43)	22.73(±3.99)	3.82(±3.59)	−8.35(±2.83)	−1.18(±2.15)
no. of cycles (rings)	10.45(±2.28)	9.96(±2.65)	−10.34(2.38)	−2.85(±1.88)	−4.20(±1.43)
no. of tertiary carbons[a]	−0.74(±0.95)	−0.12(±1.10)	1.44(±.99)	0.30(±0.78)	−0.67(±0.60)
no. of quaternary carbons[b]	−1.99(±1.16)	−0.16(±1.35)	2.57(±1.21)	0.95(±0.96)	−0.61(±0.73)
no. of X—C—X tracings[c]	−0.99(±0.52)	−5.70(±0.60)	−0.11(±0.54)	−0.44(±0.43)	0.19(±0.32)
no. of $\mathrm{X_3C}$—$\mathrm{CX_3}$ bonds[d]	−0.77(±3.52)	−17.62(±4.09)	−0.63(±3.68)	−1.42(±2.90)	−0.51(±2.21)
no. of C═C—X tracings[e]	−0.47(±0.50)	−4.96(±0.58)	−0.74(±0.52)	−0.58(±0.41)	0.31(±0.31)
no. of aromatic fusions	−0.31(±2.82)	−6.57(±3.28)	1.81(±2.95)	−1.40(±2.33)	2.41(±1.77)
r^2	0.9970	0.9921	0.9284	0.8321	0.8480
s	0.0120	0.0139	0.0125	0.0099	0.0075
SD of residuals (includes $\mathrm{CH_4}$ and $\mathrm{H_2O}$)	0.0126	0.0117	0.0121	0.0089	0.0065

[a] Count one for each sp^3 carbon having three nonhydrogen, nonhalogen attachments. [b] Count one for each sp^3 carbon having four nonhydrogen, nonhalogen attachments. [c] X = not hydrogen or carbon. Count one for *each* distinct path (i.e., for CF4, X-C-X = 3 + 2 + 1 = 6). [d] X = not hydrogen or carbon. Count one for each such C-C bond, and fractions for any bond having electronegative attachments (see text). [e] X = not hydrogen or carbon, unless carbon is C═X. One for *each* path (i.e., one for CH═CHCl but two for PhCl).

first electronegative atom seem to be opposed by the polarizations from subsequent groups. We propose that this type of attenuation might be proportional to the total number of attenuating interactions possible. Thus there is only one attenuating interaction in $\mathrm{CH_2Cl_2}$, but three attenuating interactions in $\mathrm{CHCl_3}$ (between the first and second Cl's, between the first and third Cl's, and between the second and third Cl's), and six in $\mathrm{CCl_4}$. Such an "X-C-X" type of correction for attenuation yields an excellent fit to the **BCDEF** values and consumes far fewer degrees of freedom than the alternative approach of defining many individual fragments for evaluation. A similar rationale underlies the "$\mathrm{X_3C}$-$\mathrm{CX_3}$" correction, which applies to bonds between carbons each bearing at least one atom not hydrogen or carbon. Every such bond makes an "$\mathrm{X_3C}$-$\mathrm{CX_3}$" contribution equal to 1 − $\frac{1}{6}$ (number of carbons and hydrogens attached to the two end atoms). The "C═C—X" correction allows a generalized distinction between aromatic and aliphatic functionalities. This "C═C—X" correction is applied once for each electronegative moiety attached to an alkene carbon and twice for each moiety attached to an aryl ring. The final feature in Table IV, the benzo fusion, is self-explanatory.

The excellent overall fit of the additive-constitutive model to the original **BCDEF** parameters is shown by the r^2 and s values at the bottom of Table IV. The 112 observations require only 35 fragment definitions, a most satisfactory result considering the structural diversity of the compounds. (Methane and water, which have both anomalous **B** or **C** values and unique fragments, were excluded from the model.) The individual fragment assignments for the compounds in Tables I of both this and the preceding paper appear in the supplementary material.

For hand calculations a linear model is more convenient than the hierarchical model. Consequently a linear additive-constitutive model for evaluating **BCDEF** parameters, derived from the equations of Table IV, is shown in Table V. However, note that the complex corrections, that is, the last seven features in Table IV, will often still be necessary when using the linear model.

The group contributions of Table V also facilitate recapitulation of the mechanistic rationale for the **BC(DEF)** parameters.[1] The surprises are found among the **B** values. One is the relatively high effective "bulk", or **B** contribution, of polar groups. A hydroxyl appears in Table V to be much "bigger" than an amino or methyl group, and a primary amide to be little "smaller" than a phenyl group. The relatively low or even negative bulk ascribed by **B** to atoms bearing few hydrogens can hardly be literally true. However, the **C** and **D** group contributions conform with their "cohesiveness" and

Table V. **BC(DEF)** Values of Individual Fragments: a Linear Additive-Constitutive Model Derived from Table IV

	B	C	D	E	F
-H	0.066	0.018	−0.027	−0.019	−0.019
—CH$_3$	0.142	−0.020	−0.016	−0.023	−0.015
—CH$_2$—	0.076	−0.038	0.011	−0.004	0.003
>CH—[c]	0.003	−0.058	0.053	0.018	0.015
>C<[c]	−0.075	−0.076	0.091	0.043	0.034
—CH=CH—	0.147	−0.043	0.028	0.010	0.003
—CH=CH$_2$	0.212	−0.025	0.000	−0.009	−0.015
>C=CH$_2$	0.147	−0.043	0.028	0.010	0.003
—C≡CH	0.171	0.074	0.027	0.002	−0.012
—C$_6$H$_5$	0.467	−0.007	0.012	0.007	−0.017
=CH— (aromatic)	0.088	0.002	−0.007	0.001	−0.003
-naphthyl	0.766	0.018	−0.026	0.024	−0.028
-cyclohexyl	0.489	−0.148	0.004	−0.029	−0.009
—F[a,b]	0.078	0.088	0.009	−0.019	−0.020
—Cl[a,b]	0.165	0.087	−0.024	−0.012	−0.021
—Br[a,b]	0.213	0.095	−0.033	−0.008	−0.020
—I[a,b]	0.302	0.103	−0.056	−0.010	−0.031
—CF$_3$[b]	0.150	0.017	0.035	−0.037	−0.013
—CCl$_3$[b]	0.410	0.015	−0.009	−0.017	−0.017
—OH[a]	0.202	0.324	−0.012	−0.015	0.003
—O—[a]	0.044	0.155	0.061	0.019	−0.022
—C=O—[a]	0.135	0.246	0.061	0.023	−0.021
—CH=O[a]	0.219	0.244	0.010	−0.014	−0.027
—COO—[a]	0.167	0.170	0.062	0.015	−0.027
—COOH[a]	0.323	0.342	−0.011	−0.017	0.008
—NH$_2$[a]	0.167	0.269	0.037	0.027	−0.014
—NH—[a]	0.082	0.251	0.095	0.056	−0.010
—N—[a]	−0.006	0.189	0.125	0.069	0.014
—CN[a]	0.241	0.269	−0.007	−0.023	−0.041
—N=[a] (pyridine)	0.102	0.183	0.031	−0.011	−0.020
—NO$_2$[a]	0.238	0.241	−0.012	−0.027	−0.037
—CONH$_2$[a]	0.444	0.499	−0.019	−0.039	−0.012
—S—[a]	0.136	0.130	0.028	0.032	−0.020
—SH[a]	0.231	0.155	−0.026	−0.011	−0.013
molecule	−0.5065	−0.056	0.007	0.031	0.028

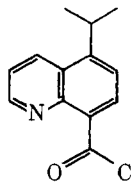
[a] Value when attached to aliphatic system. Note the correction for "C=C—X" in Table IV, to be applied once for this group when attached to an alkenyl carbon and twice when this group is attached to aromatic carbon. [b] Note correction for "X-C-X" in Table IV, which must be applied when more than one halogen or other nonhydrogen, noncarbon atom is attached to the same carbon atom. The —CF$_3$ and —CCl$_3$ values already reflect the "X—C—X" correction. [c] Includes the "tertiary" or "quaternary" correction of Table IV.

"dispersion" rationalizations.

Evaluation of a Collection of Predictions. As mentioned above, evaluation of a predictive scheme has not often been attempted in the chemical literature. Therefore a brief general discussion of the issues involved will precede the presentation of our results.

"Prediction" might be defined as an attempt to reduce the uncertainty surrounding the value of an unknown property. In the worst instance, with no predictive scheme available, the most reasonable guess for an unknown property value might be the mean of a representative set of known values of the property. The "probable error" involved in such a guess would be the standard deviation of the set of known values. Therefore a predictive scheme should at the least yield values whose deviations from a property mean are greater, in general, than their deviations from the corresponding actual or experimentally determined values. A perfect predictive scheme would of course yield values that are indistinguishable from the experimental value. These two extreme cases establish a scale of "predictive accuracy", on which the perfect predictive result is rated as 1.0 and the result which is no better a predictor than the property mean is rated as 0.0. Intermediate prediction results can be placed onto this scale in a natural way, by using the expression

Table VI. Sample Calculation of a **B** Value Using the Hierarchical Scheme of Table IV or the Linear Scheme of Table V



hierarchical scheme		linear scheme	
intercept	−0.506	intercept	−0.506
+ mol wt (=267.25)	0.602	+ 2 —CH$_3$'s	0.284
+ 1 O atom and 1 N atom	0.020	+ 1 >CH—	0.003
+ 5 C—C bonds	−0.205	+ 1 C$_{10}$H$_7$-	0.766
+ 9 aromatic C···C; bonds	−0.007	− 1 —H	−0.066
+ 12 C—H bonds	0.514	−1 =CH—	−0.088
+ 3 C—F bonds	0.044	+ 1 =N—	0.102
+ 1 —C=O— group	0.104	+1 —C=O—	0.135
+ 1 =N— group	0.038	+ 1 —CF$_3$	0.150
+ 2 cycles (rings)	0.209		
+ 1 tertiary carbon	−0.007		
+ 3 X—C—X tracings	−0.030		
+ 0.833 X$_3$C—CX$_3$ bond (between C=O and CF$_3$)	−0.007	+ 0.833 X$_3$C—CX$_3$ bond	−0.007
+ 2 C=C—X tracings (C—X is the aromatic to C=O bond)	−0.009	+ 2 C=C—X tracings	−0.009
+ 1 aromatic fusion	−0.003		
B=	0.757	**B=**	0.764

"prediction r^2" or "accuracy of prediction"

$$= 1 - \frac{1}{ns^2}\sum_{n} (\text{predicted value} - \text{exptl value})^2$$

where n is the number of predictions and s is the standard deviation of a representative and comparable, or "typical", set of property values. It is evident that this expression yields a maximum value of 1.0 when all predicted and experimental values are equal and falls below 0.0 whenever the average root mean square of (predicted − experimental)^2 becomes greater than s^2. This scale for prediction accuracy is almost identical in definition and interpretation with the familiar r^2 criterion for the fit between a regression equation and the data from which the equation is derived. However, it must be appreciated that a high "prediction r^2" is more difficult to achieve than a high "regression r^2" because in the prediction case there are no adjustable coefficients.

A judgmental problem arising with this criterion for prediction accuracy is the choice of the "typical" compound set for computation of s^2. The higher the s^2 used for comparison (greater spread in values), the higher will be the "prediction r^2" of any particular set of predictions. In our case, any s^2 estimate based on a set of actual values is probably an underestimate of s^2 for the theoretical objective, predicting values of a property for all possible compounds. We therefore compute "prediction r^2" based either on the s^2 calculated from the experimental values being predicted (Table I of this paper) or on the s^2 from values of that property represented among the original 114 compounds, whichever value is higher.

The Prediction Studies. Results and Discussion. "Predictions" were carried out for every known value of the 21 properties of the 139 compounds in Table I, applying the "property equations" (Table IV[1]) both to a set of **BC(DEF)** values derived from structure and, where possible, to a set obtained from the four properties indicated above. An illustration of how a "prediction" is obtained using a property equation and a set of **BCDEF** values was given previously (Table III[1]).

The results of these 1142 predictions are tabulated in three different ways. First, Table I gives the **BCDEF** values for the 139 compounds, based on the additive-constitutive model of Table IV, and indicates the closeness of each structure-based **BCDEF** prediction to the experimental value. (The actual experimental values and three predictions apiece (**BC, BCD,** and **BCDEF**) appear in the supplementary material.) Second, Table VII assesses the accumulated prediction results by property type and by compound class, and also compares the predictions based on structure-derived **BC(DEF)** values with those based on property-derived **BC(DEF)** values. Finally, Table VIII summarizes the assessments of Table VII, giving averaged "prediction r^2" and a count of the property r^2 values within Table VII that exceed various cutoffs. Both Table VII and Table VIII allow comparisons of the two-parameter **BC** predictions with the five-parameter **BCDEF** predictions. (Predictions using the three-parameter **BCD** model seem in all respects to be intermediate between the two- and five-parameter model and so these results are relegated to the supplementary material.)

Of these, Table VII is the most important. The accuracies of prediction are given by compound class for each of the 21 properties listed. For example, the first row of Table VII indicates that among the class I compounds of Table I there are 47 known activity coefficient values. Their standard deviation is taken as 2.00. Ordinarily this value is the standard deviation of the known values themselves, but, as shown by the footnote, in this instance the standard deviation of the original 114 activity coefficients has the larger value.

The next four items summarize the accuracy of predicting these activity coefficients using structure-derived **BC(DEF)** values. The **BC** equation yielded predictions whose root mean square deviation from the experimental values was 0.38, a "prediction r^2" of 0.964 according to the previously given formula. The **BCDEF** equation gave better predictions, a root mean square deviation of 0.30 yielding a prediction r^2 of 0.978. The last five items compare activity coefficient predictions using **BC(DEF)** values derived by applying Table III to four predicting properties (partition coefficient, boiling point, molar volume, and refractivity) with those from the additive-constitutive model. The first of these items indicates that the comparison includes only the 15 class I compounds which have known values of both activity coefficient and the four predicting properties. The next two items compare the root mean square deviations, 0.70 in prediction using "four-property"-derived **BC** parameters with 0.43 using "additive-constitutive" **BC** parameters. The additive constitutive model is significantly more accurate. Finally, the last two items show that additive-constitutive predictions are superior to the four property predictions also when the five-parameter **BCDEF** model is used.

These 1142 prediction experiments can be summarized as follows.

(1) The accuracy in predicting the 1142 property values from structure alone, averaged over all properties of all compounds, is about 78% (lower right hand corner of part A of Table VIII). This corresponds to a "prediction r" of $\sqrt{0.78}$ or 0.88.

(2) As would be expected, the accuracy of prediction improves for lower numbered compound classes (those compounds which are most similar to those from which the **BC(DEF)** model was derived) and for lower numbered properties (those properties whose factorization most cleanly produced the **BC(DEF)** vectors). The average accuracy of prediction for the "reliable subset", the 749 values of properties 1-10 for compounds in classes I and II, using the five-parameter model, is about 94%, a "prediction r" of 0.97.

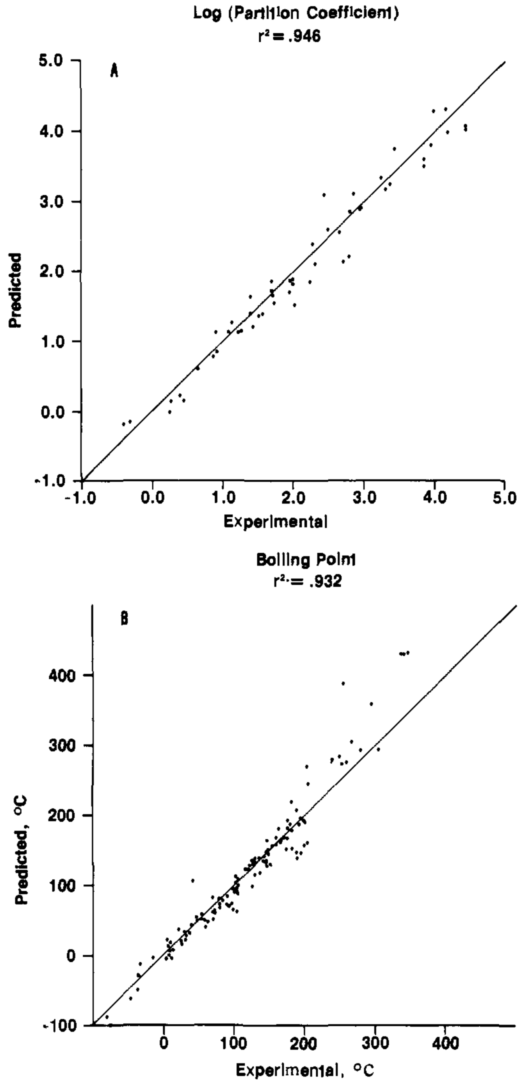


Figure 1. Plots of predicted vs. experimental values of (A) partition coefficients and (B) boiling points of all class I and class II compounds.

(3) On the other hand, only 10 of the 18 properties of class I compounds are predicted with an average accuracy of >90%, and only 4 with an average accuracy >95%. The overall average accuracy is skewed upwards because those properties whose known values are most numerous proved to be easiest to predict.

(4) The **BCDEF** and **BC** models perform almost identically in predicting all properties for all compounds. However, when interest is limited to the "reliable subset", the five-parameter **BCDEF** model reduces the error of prediction by about one-third, a highly significant amount statistically.

(5) On the other hand, the two-parameter **BC** model is much the more "robust" when the properties of class III compounds, those having the least resemblance to those on which the model is based, are being predicted. ("Robust" connotes a model which can be extrapolated without giving increasingly divergent predictions. The more robust a model, the more likely it is to represent, and not merely reflect, physical reality.)

(6) Perhaps surprisingly, use of the four-property model instead of the additive-constitutive model to calculate **BC(DEF)** values did not usually improve the accuracy of predicting other compound properties, according to the right-hand side of Table VII. This comparison has to be limited to those compounds whose partition coefficient, boiling point, and molar refraction and volume are known.

To allow visualization of the average accuracy of predictions, two representative sets of data having "prediction r^2" of 0.94,

Table VII. Averaged Accuracies of "Predicting" Known Values of 21 Properties of 139 Compounds (Table I) Having Various Degrees of Structural Similarity to the Original Structures[1,j]

			additive-constitutive **BCDEF**[i]				property **BCDEF**[i]				
class	n	s	**BC—**[h] rms	**BC—** "r^2"	**BCDEF**[h] rms	**BCDEF** "r^2"	4-prop n	**BC**[h] (4p) rms	**BC**[h] (ac) rms	**BCDEF**[h] (4p) rms	**BCDEF**[h] (ac) rms
1. Activity Coefficient											
I	47	2.00[a]	0.38	0.964[d]	0.30	0.978[d]	15	0.70	0.43[e]	0.65	0.23[e]
II	11	2.00[a]	0.81	0.835[d]	0.74	0.863[d]	5	0.63	0.69	0.50	0.55
III[c]	7	2.62	2.55	0.053	2.85	<0.000	3	1.30	1.74	1.24	1.83
				0.787[d]		0.760[d]					
2. Partition Coefficient											
I	30	1.22[a]	0.39	0.899[d]	0.31	0.935[d]					
II	22	1.46	0.48	0.891[d]	0.28	0.962[d]	used to calculate **BCDEF**				
III	8	1.33	0.45	0.885[d]	0.75	0.682					
				0.895[d]		0.912[d]					
3. Molar Refractivity											
I	72	10.32[a]	1.95[g]	0.964[d]	0.85	0.994[d]	used to calculate **BCDEF**				
II	38	14.44	2.82	0.962[d]	1.05	0.994[d]					
III	19	17.34	6.51	0.859[d]	3.55	0.958[d]					
				0.941[d]		0.986[b,d]					
4. Boiling Point											
I	75	104.4[a]	21.08[g]	0.958[d]	17.12	0.972[d]	used to calculate **BCDEF**				
II	43	104.4[a]	39.94	0.854[d]	38.70	0.863[d]					
III	20	102.0	62.80	0.769[d]	99.98	0.039					
				0.880[d]		0.808[d]					
5. Molar Volume											
I	73	30.1[a]	11.0[g]	0.867[d]	3.51	0.986[d]	used to calculate **BCDEF**				
II	39	50.5	13.0	0.933[d]	7.15	0.989[d]					
III	20	61.8	19.9	0.897[d]	17.40	0.922[d]					
				0.904[d]		0.964[b,d]					
6. Heat of Vaporization											
I	54	3.07[a]	0.81	0.931[d]	0.77	0.937[d]	12	1.08	0.83	1.13	0.97
II	32	3.08	1.41	0.790[d]	1.38	0.799[d]	12	0.92	1.28	0.78	1.27
III	12	4.50	1.85	0.832[d]	2.46	0.701[d]	2	0.58	0.18	0.62	0.11
				0.867[d]		0.843[d]					
7. Magnetic Susceptibility											
I	49	20.1[a]	6.41[g]	0.899[d]	6.45[g]	0.897[d]	19	5.42	6.23	5.45	6.29
II	28	36.9	5.59	0.976[d]	5.70[g]	0.976[d]	14	6.21	5.94	6.36	6.05
III	11	52.2	12.3	0.945[d]	12.2	0.945[d]	4	2.59	5.27	2.57	5.71
				0.929[d]		0.927[d]					
8. Critical Temperature											
I	51	134.1[a]	42.9[g]	0.897[d]	35.45	0.929[d]	18	31.8	44.5	18.4[f]	45.9
II	27	122.4	71.7	0.714[d]	48.45	0.869[d]	13	82.1	97.2	19.6[f]	57.1
III	11	176.0	85.2	0.766[d]	168.0	0.089	2	40.6	15.2	11.5[f]	69.5
				0.824[d]		0.806[d]					
9. (van der Waals A)$^{1/2}$											
I	22	0.99	0.23[g]	0.947[d]	0.20[g]	0.959[d]	7	0.11	0.25	0.12	0.21
II	7	0.96[a]	0.65	0.542	0.65	0.542	5	0.66	0.76	0.26[f]	0.72
III	7	1.20	0.79	0.567	0.92	0.412	1	0.08	0.01	0.10	0.13
				0.771[d]		0.772[d]					
10. van der Waals B											
I	22	0.04	0.02	0.832[d]	0.02	0.887[d]	7	0.02	0.02	0.01	0.01
II	7	0.05	0.03	0.632	0.03	0.684	5	0.02	0.02	0.01	0.02
III	7	0.04	0.02	0.714	0.02	0.717	1	0.00	0.00	0.00	0.00
				0.794[d]		0.839[d]					
11. Log (Dielectric Constant)											
I	23	0.35[a]	0.31[g]	0.215	0.31	0.194	6	0.34	0.38	0.39	0.39
II	13	0.34	0.34	0.027	0.24	0.498	5	0.34	0.35	0.18	0.17
III	11	0.52	0.39	0.437	0.59	<0.000	3	0.33	0.40	0.35	0.33
				0.304		0.070					
12. Solubility Parameter											
I	18	1.84[a]	0.79[g]	0.704[d]	0.73[g]	0.843[d]	5	0.53	1.12	0.98	1.15
II	11	1.84[a]	1.48[g]	0.353	1.78	0.064	4	1.13	1.55	0.69	1.40
III	5	3.18	2.30	0.477	4.90	<0.000	1	0.00	0.43	0.30	0.38
				0.557[d]		<0.000					
13. Critical Pressure											
I	41	12.7[a]	8.25[g]	0.576[d]	6.87[g]	0.706[d]	10	7.27	7.60	7.34	11.12
II	19	15.9	9.48[g]	0.643[d]	10.8[g]	0.533[d]	7	10.31	11.4	10.4	10.25
III	6	25.7	22.7	0.222	47.30	<0.000	2	7.56	8.3	0.64	2.15
				0.563[d]		<0.000					

Table VII (*Continued*)

			additive-constitutive **BCDEF**[i]				property **BCDEF**[i]				
class	n	s	**BC—**[h] rms	**BC—** "r^2"	**BCDEF**[h] rms	**BCDEF** "r^2"	4-prop n	**BC**[h] (4p) rms	**BC**[h] (ac) rms	**BCDEF**[h] (4p) rms	**BCDEF**[h] (ac) rms
14. Surface Tension											
I	14	9.58	5.99	0.608[d]	2.42[g]	0.935[d]	3	1.72	1.83	2.31	2.22
II	12	8.66	8.31	0.100	6.98	0.368	7	8.98	9.80	7.12	10.25
III	8	22.2	30.8	<0.000	32.5	<0.000	1	201	234	149	348
				0.099		0.283					
15. Thermal Conductivity											
I	15	0.73[a]	0.52[g]	0.493	0.70	0.081	4	0.27	0.51	0.29	0.59
II	8	0.73[a]	0.86	<0.000	0.81	<0.000	2	0.73	1.06	0.65	0.97
III	5	2.26	1.66	0.461	0.95	0.823	1	0.81	0.45	0.34	1.15
				0.372		0.548[d]					
16. Log (Viscosity)											
I	17	0.55	0.29[g]	0.733[d]	0.23	0.824[d]	4	0.49	0.32	0.52	0.29
II	15	0.47	0.32	0.564	0.51	<0.000	7	0.27	0.25	0.29	0.43
III	7	1.36	0.73	0.712[d]	1.16	0.272	1	0.34	0.21	0.70	0.26
				0.664[d]		0.335					
17. Isothermal Compressibility											
I	6	3.82	2.68[g]	0.508	1.39[g]	0.867[d]	0				
II	6	4.25	4.11	0.067	5.33	<0.000	2	3.16	3.81	6.67	8.33
III	3	3.36[a]	3.90	<0.000	10.60	<0.000	1	5.40	3.02	2.50	2.54
				0.173		<0.000					
18. E_T (Solvent Effect on Electronic Transition)											
I	2	8.69[a]	4.88[g]	0.686	2.12[g]	0.941[d]	1	10.56	19.3	0.03	8.57
II	4	8.69[a]	3.85[g]	0.805	5.28	0.630	4	3.75	3.85	3.89	5.28
III	2	16.8	7.3	0.812	11.20	0.555	1	34.5	68	5.18	10.6
				0.776[d]		0.920[d]					
19. Dipole Moment											
I	38	1.09[a]	0.68[g]	0.529[d]	0.57[g]	0.724[d]	18	0.57	0.57	0.56	0.39
II	19	1.08	0.92[g]	0.266	0.84	0.389	10	0.73	0.83	0.59	0.60
III	14	1.30	1.06	0.280	1.33	<0.000	5	0.66	0.65	0.64	0.91
				0.420		0.445[d]					
20. Melting Point											
I	70	73.6[a]	41.63[g]	0.681[d]	41.2[g]	0.686[d]	26	33.8	34.2	34.4	37.9
II	40	78.3	52.59[g]	0.549[d]	55.9	0.491[d]	17	47.1	48.1	40.5	40.0
III	20	73.6[a]	61.8	0.295	54.2	0.458	5	28.2	29.1	17.9	30.0
				0.581[d]		0.591[d]					
21. Molecular Weight											
I	75	40.3[a]	33.37	0.315[d]	31.6	0.386[d]	28	33.8	34.2	17.2	17.2
II	43	44.1	17.67[g]	0.839[d]	19.8[g]	0.799[d]	19	8.96	8.91	14.2	16.66
III	21	66.2	43.56	0.567[d]	57.1	0.256	6	5.20	13.4	11.6	29.8
				0.516[d]		0.494[d]					

[a] The standard deviation of all the original values for this property,[1] rather than the (smaller) standard deviation of the n experimental values being predicted. [b] Significantly more accurate in prediction than the corresponding **BC** equation ($P < 0.05$). [c] These predictions would be much better if the intramolecular interaction between vicinal —OH groups and —NH_2 groups was included (for ethylene glycol, glycerol, and ethylenediamine). [d] Significantly superior to "predicting" simply that all compounds have the mean property value (a "predictive r^2" of 0) ($P < 0.05$). [e] Significantly superior to the predictions using **BC(DEF)** values derived from the additive-constitutive model of Table IV ($P < 0.05$). [f] Significantly superior to the predictions using **BC(DEF)** values derived from the property-based equations of Table III ($P < 0.05$). [g] Not significantly inferior to the original fit, between the property equation (Table VI, preceding paper) and the data from which the property equation was derived ($P < 0.05$). [h] The units of measurement are indicated in Table IV of the previous paper. [i] Explanation of table headings: n is the number of predictions attempted, i.e., the number of compounds in Table I for which this property is known. s is a "typical" standard deviation of this property for this class of compounds. Except where noted, it is the standard deviation of the n experimental values. See discussion in text. rms is $\sum_n$ (predicted − actual)$^2/n$. "r^2" is the fraction of the variance in experimental values accounted for by the predictions, as defined in the text. [j] The intermediate **BCDEF** values are calculated either from an additive-constitutive model (Table IV) or from four physical properties (Table III), as indicated by column headings.

the boiling points and partition coefficients of class I and class II compounds, have been plotted in Figure 1. The outliers on these plots represent either shortcomings of the **BCDEF** scheme or, possibly, experimental error.

In deciding whether the scheme is accurate enough to be used for prediction of a particular property, attention should be focused on the rms columns for that property and compound class in Table VII. About two-thirds of the time, a predicted value will deviate from the actual values by less than this rms value. The accuracies of each individual prediction are presented symbolically in the right-hand half of Table I. Absence of any symbol in the appropriate space indicates an unknown experimental value. An "=" implies that the experimental value and value predicted from the additive-constitutive **BCDEF** model differ by less than the standard deviation or average fit between the **BCDEF** property equation and its derivation data (Table IV[1]). A "+" or "−" entry implies that the calculated value is greater or less, respectively, than the experimental value by more than 1.0, but less than 2.0, standard deviations, while a "++" or "−−" value denotes a cal-

Table VIII. Summarization of the Average Prediction Accuracies for the 18 Liquid State Properties Presented in Table VII

A. Mean "Prediction r^2" Values[a]

compd class	properties 1-6			properties 7-10			properties 11-18			properties 1-10			all 18 properties		
	n	BC-r^2	BCDEF-r^2	n	BC-r^2	BCDEF-r^2	n	BC-r^2	BCDEF-r^2	n	BC-r^2	BCDEF-r^2	n	BC-r^2	BCDEF-r^2
I	351	0.931	0.972	144	0.895	0.909	136	0.544	0.617	495	0.921	0.954	631	0.840	0.881
II	185	0.885	0.915	69	0.795	0.860	88	0.338	0.330	254	0.861	0.900	342	0.726	0.740
III	86	0.780	0.596	36	0.772	0.535	47	0.371	0.152	122	0.778	0.578	169	0.664	0.460
all compounds	622	0.896	0.903	249	0.850	0.841	271	0.447	0.443	871	0.883	0.886	1142	0.780	0.776

B. Number of Properties (1-18 in Table VII) Whose Mean "Prediction r^2" Exceeds Various Given Values

compd class	$r^2 > 0.00$		$r^2 > 0.00$, statistically significant		$r^2 > 0.50$		$r^2 > 0.75$		$r^2 > 0.90$		$r^2 > 0.95$	
	BC	BCDEF	BC	BCDEF	BC	BCDEF	BC	BCDEF	BC	BCDEF	BC	BCDEF
I	18	18	14	15	15	16	10	15	6	10	3	4
II	17	15	9	9	13	11	<8	8	3	4	2	4
III	16	12	8	4	11	8	8	4	1	3	0	1
all compounds	18	15	14	13	14	12	11	10	6	5	0	2

[a] As discussed in the text, "prediction r^2" = $1/ns^2\sum_n$ (predicted value − experimental value)^2 where s is the standard deviation of a representative set of property values. In this case, where r^2 is being averaged over several properties, "prediction r^2" = $\sum n_k(r^2)k/\sum_k n_k$, where (r^2) and n_k are the values given in Table VII for the prediction accuracy and the number of predictions, respectively, of the kth property for the indicated class of compound.

Table IX. Comparison of Predictions Using Additive-Constitutive Schemes from the Literature[a] for Individual Properties with the Predictions Using the **BC(DEF)** Values

property/compound class	no. of predictions	rms of prediction[c] using lit. scheme	rms of prediction[c] using BCDEF, same compds	rms[b] ratio²
		1. Activity Coefficient (H_2O)		
class I	47	0.55	0.30	3.20**
class II	11	1.39	0.74	3.54**
class III	6	3.30	2.85	1.34
	64	1.57	0.92	2.89**
		2. Partition Coefficient		
class I	30	0.37	0.31	1.45
class II	22	0.64	0.28	5.14**
class III	5	0.95	0.75	1.61
	57	0.55	0.36	2.35**
		3. Molar Refractivity		
class I	72	0.83	0.85	0.94
class II	38	0.97	1.05	0.85
class III	17	2.58	3.24	0.63
	127	1.25	1.46	0.73*
		4. Boiling Point		
class I	47	17.25	13.62	1.60*
class II	25	17.71	23.46	0.57
class III	12	75.37	92.52	0.66
	84	32.73	38.61	0.72
		7. Magnetic Susceptibility		
class I	49	5.83	6.45	0.82
class II	27	4.39	5.80	0.57
class III	8	6.00	12.00	0.25
	84	5.43	6.99	0.60*

[a] See ref 2 for the literature individual property additive-constitutive schemes. [b] Compares the individual property scheme accuracy of prediction with the **BCDEF** model accuracy of prediction. A ratio greater than 1.0 implies relative superiority of the **BCDEF** model and less than 1.0 implies relative inferiority of the **BCDEF** model. [c] Units are given in Table IV of the preceding paper. ** This difference in predictive accuracy between the two models has a less than 1% probability of being a chance occurrence, according to an F-test of this variance ratio. * This difference in predictive accuracy has a less than 5% probability of being a chance occurrence.

culated value which is more than 2.0 standard deviations higher or lower than the experimental value. Statistical theory teaches that a "well-behaved" model of normally distributed data should fit about two-thirds of its data points within one standard deviation and 95% of its data points within two standard deviations. Predictions of properties 7-21 for class I compounds fulfill this expectation very well, 64% of predictions in Table I being within one standard deviation and 90% being within two standard deviations. In other words, the predictions for compounds resembling those on which the model is based are as good as the original fit by the model. That the predictions of properties 1-6 are not as "well behaved" is understandable, because in these instances the original standard deviations are made artifactually small by the manner of definition of the original **BCDEF** values.

As a final assessment of the **BC(DEF)**-based predictions, existing additive-constitutive schemes were used to "predict" the values of activity coefficient, partition coefficient, molar refractivity, boiling point, and magnetic susceptibility for all applicable compounds in Table I. Their average errors of prediction are compared in Table IX with the **BCDEF** structure-based predictions for the same compounds. These comparisons must be assessed cautiously, because (1) all of the literature additive-constitutive schemes were forced into a fragment pattern much like that of Table IV, in particular by requiring a general "C═C—X" correction instead of separate aromatic and aliphatic parameters for every function; (2) some of the compounds in Table I were probably used to *define* fragment values in previously existing schemes, and thus "prediction" of those properties is not a prediction at all. However, the rms ratios in Table IX do suggest that the general additive-constitutive **BCDEF** model, if anything, tends to be more accurate than these individual property additive-constitutive models. Furthermore, the existing models often cannot be applied to as wide a variety of structures. Thus it seems appropriate to seek a **BC(DEF)** model rather than a new additive-constitutive model when one hopes to predict a physical property which depends upon nonspecific intermolecular interactions.

Conclusions

Prediction of any property which depends mostly upon nonspecific and noncovalent intermolecular interactions, from structure alone, appears from the foregoing to be technically feasible within established error limits. On an absolute basis, the ability to forecast 94% of the variance in many properties of the most commonly encountered compounds would seem to have areas of practical applicability. On a relative basis, the **BC(DEF)** model is as accurate as previously existing prediction schemes, and might be considered preferable to these on the grounds of demonstrated compatibility with the widest variety of experimental observations.

Further studies now underway are intended to extend the range of **BC(DEF)** predictions, both to compound types not yet considered, such as multiply functionalized substances usually existing as solids, and to other properties, such as boiling points or partition coefficients under different conditions, or those aspects of biological behavior which are not structurally specific.

Supplementary Material Available: Experimental and calculated values for the indicated properties of the compounds in Table I of this paper and the structural fragments and resulting calculated **BCDEF** values for all compounds in Tables I of this and the preceding paper (40 pages). Ordering information is given on any current masthead page.

References and Notes

(1) R. D. Cramer, III, *J. Am. Chem. Soc.*, preceding paper in this issue.

(2) (a) The activity coefficient scheme is from J. Hine and P. R. Mookerjee, *J. Org. Chem.*, **40**, 292 (1975). (b) The partition coefficient scheme is that of A. Leo, P. Y. C. Jow, C. Silipo, and C. Hansch, *J. Med. Chem.*, **18**, 865 (1975). (c) The molar refractivity scheme is from A. I. Vogel, *J. Chem. Soc.*, 1833 (1948), as adapted by C. Hansch, A. Leo, S. H. Unger, K. H. Kim, D. Nikaitani, and E. J. Lien, *J. Med. Chem.*, **16**, 1207 (1973). (d) The boiling-point scheme is that of C. R. Kinney, *J. Am. Chem. Soc.*, **60**, 3032 (1938), as cited by (e) H. F. Herbrandson and F. C. Nachod in "Determination of Organic Structures by Physical Methods", E. A. Braude and F. C. Nachod, Eds., Academic Press, New York, 1955, p 16. (f) The magnetic susceptibility scheme is that of Pascal, *C. R. Acad. Sci.*, **156**, 323 (1913), as cited by A. Hutchison in ref 2e, p 300.

(3) O. Exner, *Collect. Czech. Chem. Commun.*, **32**, 1, 24 (1967).

(4) One-carbon compounds such as CH_3NO_2 and CH_3SH are known to behave anomalously in this type of model (ref 3) and have been placed in class II.

(5) S. W. Benson and J. H. Buss, *J. Chem. Phys.*, **29**, 546 (1958), and ref 3.

(6) M. F. Lynch, J. M. Harrison, W. G. Town, and J. E. Ash, "Computer Handling of Structural Information", American Elsevier, New York, 1972; E. J. Corey, W. T. Wipke, R. D. Cramer III, and W. J. Howe, *J. Am. Chem. Soc.*, **94**, 431 (1972).

Kinetics and Mechanism of Hydrolysis of a Silicate Triester, Tris(2-methoxyethoxy)phenylsilane

K. J. McNeil, J. A. DiCaprio, D. A. Walsh, and R. F. Pratt*

Contribution from the Department of Chemistry, Wesleyan University, Middletown, Connecticut 06457. *Received July 23, 1979*

Abstract: The kinetics of hydrolysis in dilute aqueous solution of tris(2-methoxyethoxy)phenylsilane to phenylsilanetriol have been studied. The hydrolysis exhibits specific acid and general base catalysis, the latter with a Brønsted β value of 0.7. The specific acid catalysis mechanism is probably A-2 ($k_{H_3O^+/D_3O^+} = 1.24$, $\Delta S^{\ddagger} = -39$ cal deg^{-1} mol^{-1}). At high pH (>10) the rate of appearance of the triol is limited by the rate of hydrolysis of one of the intermediates in the hydrolysis sequence, bis(2-methoxyethoxy)phenylsilanol, which, under these conditions, forms an inert anion. At lower pH the hydrolysis of bis(2-methoxyethoxy)phenylsilanol is several times faster than that of tris(2-methoxyethoxy)phenylsilane while that of the second intermediate, 2-methoxyethoxyphenylsilanediol, is probably faster than the above two hydrolyses at all pHs. It is argued that the form of general base catalysis observed suggests that the base-catalyzed reactions involve either an S_N2^{**}-Si or S_N2^{*}-Si mechanism with formation of a pentacoordinate intermediate. Generalization of the argument used here is explored.

Many processes of importance to both geology[1] and, as is being increasingly appreciated, biology[2] involve the reactions in aqueous solution of simple silicates, i.e., of the $Si(OX)_n$ functional group, where X = H, C, or Si and $n > 1$. Although many qualitative and semiquantitative studies have been made of the hydrolyses of such compounds,[3] practically no systematic kinetic/mechanistic or thermodynamic studies have been made. Quantitative interpretation of much of the early work is complicated by the presence of polymerization phenomena and by the use of a variety of mixed aqueous-organic solvent systems.

We present here what appears to be the first example of a kinetic study of the hydrolysis in a homogeneous, purely aqueous medium of a silicate triester, tris(2-methoxyethoxy)-

0002-7863/80/1502-1859$01.00/0