JOURNAL
OF THE AMERICAN CHEMICAL SOCIETY

VOLUME 95, NUMBER 18 SEPTEMBER 5, 1973

The Estimation of Relative Gas Phase Acidities of Substituted Benzoic Acids from Solution Measurements by Factor Analysis[1a]

Paul H. Weiner[1b]

Contribution from the Chemistry Department, The University of Mississippi, University, Mississippi 38677. *Received February 7, 1973*

Abstract: Data on the relative acidities of substituted benzoic acids in a variety of solvents were analyzed using the mathematical technique of factor analysis. It was found that four factors were required to account for the variance of the data in ethylene glycol solvent while three factors were sufficient for all other solvents studied. Using the rotation part of the factor analysis scheme, successful attempts were made to identify the abstract factors with physically significant parameters such as relative gas phase acidities, electrostatic interactions, and van der Waals' effects. Values of the relative gas phase acidities were estimated.

Recently, the study of gas phase acidities has received much attention[2,3] due to the development of methods for determining these quantities and also due to their importance in aiding our understanding of reaction rates in solution. One of the main experimental methods for obtaining gaseous acidity data consists of using ion cyclotron resonance spectroscopy. However, due to problems of sample decomposition and low volatility, this technique has been limited to small molecular weight compounds. An alternate approach of obtaining gas phase data, which is not limited by the above constraints, is by using the technique of factor analysis[4] to mathematically isolate the appropriate gas phase acidity data from solution measurements.

Factor analysis can be used whenever the measured quantity of the system, Q, is expressable as a linear sum of terms in product function form.

$$Q(i,\alpha) = \sum_{j=1}^{n} U(i,j)V(j,\alpha) \tag{1}$$

In a solute–solvent interaction problem, $Q(i,\alpha)$ would correspond to the measured property of the system for solute i in solvent α; $U(i,j)$ would correspond to the jth solute property of solute i; $V(j,\alpha)$ would correspond to the jth solvent property of solvent α; and the sum would be taken over the n important solute–solvent interactions. At present, factor analysis has proved valuable in helping to understand several multidimensional problems: (1) linear free energy relationships such as the Hammett equation;[4a,b] (2) physical properties of the isomeric paraffin hydrocarbons;[4a] (3) solvent effects in nmr;[4c–f] (4) activity coefficients determined by gas–liquid chromatography;[4g] (5) solute structure–gas chromatographic retention index relationships;[4h–l] (6) drug structure–biological activity relationships;[4m] (7) analysis of gas chromatographic mixtures of perfumes and tree tars;[4n] (8) prognosis of cirrhosis of the liver;[4o] and (9) solvent effects in polarography.[4p] In the present paper, factor analysis will be applied to the experimental data of Elliott and Kilpatrick[5] who reported the acid

---

(1) (a) Page charges assisted by the Chemistry Department, University of Mississippi. (b) Address correspondence to author at Department of Chemistry, Northeastern University, Boston, Mass. 02115.

(2) (a) D. T. Bohme, E. Lee-Ruff, and L. B. Young, *J. Amer. Chem. Soc.*, **94**, 5153 (1972); (b) D. K. Bohme and L. B. Young, *ibid.*, **92**, 3301 (1970); (c) *ibid.*, **93**, 4608 (1971).

(3) (a) J. I. Brauman and L. K. Blair, *ibid.*, **92**, 5986 (1970); (b) *ibid.*, **90**, 5636 (1968); (c) *ibid.*, **93**, 4315 (1971); (d) *ibid.*, **91**, 2126 (1969); (e) P. Kriemler and S. E. Buttrill, Jr., *ibid.*, **92**, 1123 (1970).

(4) (a) E. R. Malinowski, Ph.D. Thesis, Stevens Institute of Technology, Hoboken, N. J., 1961; (b) S. Wold and M. Sjöström, *Chem. Scripta*, **2**, 49 (1972); (c) P. H. Weiner, Ph.D. Thesis, Stevens Institute of Technology, Hoboken, N. J., 1971; (d) P. H. Weiner, E. R. Malinowski, and A. R. Levinstone, *J. Phys. Chem.*, **74**, 4537 (1970); (e) P. H. Weiner and E. R. Malinowski, *ibid.*, **75**, 1207 (1971); (f) *ibid.*, **75**, 3160 (1971); (g) P. T. Funke, E. R. Malinowski, D. E. Martire, and L. Z. Pollara, *Separ. Sci.*, **1**, 661 (1967); (h) P. H. Weiner and D. G. Howery, *Can. J. Chem.*, **50**, 448 (1972); (i) P. H. Weiner and D. G. Howery, *Anal. Chem.*, **44**, 1189 (1972); (j) P. H. Weiner, C. Dack, and D. G. Howery, *J. Chromatogr.*, **69**, 249 (1972); (k) P. H. Weiner and J. F. Parcher, *J. Chromatogr. Sci.*, **10**, 612 (1972); (l) P. H. Weiner and J. F. Parcher, *Anal. Chem.*, **45**, 302 (1973); (m) M. L. Weiner and P. H. Weiner, *J. Med. Chem.*, **16**, 655 (1973); (n) S. C. Elliot, N. A. Hartman, and S. J. Hawkes *Anal. Chem.*, **43**, 1938 (1971); (o) A. Gauthier, J. Zurli, R. C. Cros, and H. Sarles, *Rev. Eur. Etud. Clin. Biol.*, **17**, 574 (1972); (p) D. G. Howery, *Bull. Chem. Soc. Jap.*, **45**, 2643 (1972).

(5) (a) J. H. Elliott and M. Kilpatrick, *J. Phys. Chem.*, **45**, 454 (1941); (b) *ibid.*, **45**, 466 (1941); (c) *ibid.*, **45**, 472 (1941); (d) *ibid.*, **45**, 485 (1941); (e) J. H. Elliott, *ibid.*, **46**, 221 (1942); (f) M. Kilpatrick, *ibid.*, **46**, 159 (1942).

strengths of substituted benzoic acids in a series of solvents. An attempt will be made to isolate and thus determine the gas phase acidities from these reported solution measurements.

Development of Model

Previously, Malinowski applied the technique of factor analysis to the study of the effect of substituent on the acidity of benzoic acids, phenols, and anilines.[4a] He reported that an equation of the following form could be fitted to the experimental data.

$$\log K = \log K_0 + \rho\sigma + k'r \qquad (2)$$

Here, K is the equilibrium constant for the substituted benzoic acid, phenol, or aniline; K_0 is the equilibrium constant for the corresponding unsubstituted parent compound; ρ is a constant dependent on the reaction; σ is a constant dependent on the substituent group; r is a constant proportional to the covalent radii of substituents in the ortho position; and k' is a constant relating to the ortho substituent effect.

In an independent study of the same area, Wold and Sjöström[4b] also concluded that component factor analysis could be used to great advantage in the study of substituent effects on the Hammett equation. Equations of similar form to eq 1 have also been proposed by several workers in studies of acid–base equilibrium data. Swain and Scott[6] have proposed that a four parameter equation might be used to correlate all types of polar displacement reactions. Edwards[7] proposed a double scale equation to correlate free energy changes. Drago and Wayland[8] and Jolly, Illige, and Mendelsohn[9] proposed that the same type of equation will hold for enthalpy of adduct formations.

It is therefore also reasonable that the effect of solvent on the equilibrium constant for proton transfer in benzoic acids can be expressed by an equation in the form of eq 1.

$$\ln \frac{K(i,\alpha)}{K_0} = \sum_{j=i}^{n} U(i,j)V(j,\alpha) \qquad (3)$$

Here, $K(i,\alpha)$ is the measured equilibrium constant for the substituted benzoic acid with substituent i in solvent α; K_0 is the measured equilibrium constant for benzoic acid in solvent α; $U(i,j)$ would correspond to the difference of the jth solute property between the substituted and unsubstituted benzoic acid, for solute i; and $V(i,\alpha)$ would correspond to the jth solvent property of the αth solvent; the sum j would be taken over the n important solute–solvent interactions present.

In a manner exactly analogous to that used to study solvent effects in proton magnetic resonance,[4c–f] one of the terms of the right-hand side of eq 3 will be associated with the relative acidities in the gas phase, while the other terms will be used to account for the presence of the solvent. Therefore, eq 3 can be rewritten as

$$\ln \frac{K(i,\alpha)}{K_0} = \ln \frac{K(i,\text{gas})}{K_0}\cdot 1 + U(i,2)V(2,\alpha) + \ldots + U(i,n)V(n,\alpha) \qquad (4)$$

where each term is the correct product function form. By applying factor analysis to the data of Elliott and Kilpatrick, it should be possible to: (1) test if a mathematical solution exists in the form of eq 4; (2) find out how many terms are required in the summation to account for the data; and (3) test various models for the different interaction terms. If it is possible to identify all important interaction terms, then it should be possible to extract relative gas phase acidities from the final solution to the problem.

Application of Factor Analysis to the Proton Transfer Data of Elliott and Kilpatrick

Elliott and Kilpatrick's data (shown in Table V), which were measured by a concentration cell potentiometric method, were subjected to factor analysis.[4d] Four factors were required to reproduce the data within ±0.01 ln unit, which is their reported experimental accuracy.[5a–e] This indicates that there are four terms in eq 4. Before attempting to interpret these results, it was considered wise to perform a "uniqueness test"[4c,h,i] to ascertain whether any of the solutes or solvents in the data set contain a unique factor not present in the other data. If a high uniqueness is found for any row of the data matrix, which is not accountable on chemical grounds, then it might indicate a possible experimental error in the data of that row. The result of the uniqueness test for the solutes is shown in Table I, while that for the solvents is shown in Table II. Each row of the tables corresponds to the results of a separate uniqueness

Table I. Unique Test of Solutes, All Solvents[a]

No.	Substituent	Uniqueness[b]	Additional values
1	*o*-NO_2	0.53	(2) 0.14, (3) 0.13, (4) 0.22, (12) −0.19, (15) 0.23, (17) −0.13
2	*o*-I	0.26	(1) 0.14, (3) 0.26, (4) 0.22, (5) 0.13
3	*o*-Br	0.29	(1) 0.13, (2) 0.26, (4) 0.24, (12) 0.13, (15) −0.12
4	*o*-Cl	0.22	(1 and 2) 0.22, (3) 0.24, (5) 0.11
5	*o*-CH_3	0.12	(2) 0.13, (3) 0.10, (4) 0.13, (6) 0.11, (7) −0.19
6	*o*-OCH_3	0.45	(5) 0.11, (7) −0.25, (8) 0.15, (9) 0.21, (10) 0.12, (11) 0.11, (14) −0.23
7	*o*-OH	0.56	(5) −0.19, (6) 0.25, (8) 0.12, (10) 0.12, (12) 0.24, (14) 0.13, (15) 0.15
8	*m*-NO_2	0.22	(6) 0.15, (7) 0.12, (9) 0.12, (11) 0.10, (14) −0.12, (15) 0.22, (19) −0.12
9	*m*-I	0.16	(6) 0.21, (8) 0.12, (10) 0.14, (11) 0.11, (12) 0.10
10	*m*-Br	0.16	(1) −0.11, (6) 0.12, (7) 0.12, (9) 0.14, (11) 0.11, (12) 0.17
11	*m*-Cl	0.08	(6) 0.11, (8) 0.10, (9) 0.11, (10) 0.11
12	*m*-F	0.27	(1) −0.19, (3) 0.13, (7) 0.24, (9) 0.11, (10) 0.17, (15) −0.10
13	*m*-CH_3	0.03	
14	*m*-OH	0.14	(6) −0.23, (7) 0.12, (8) −0.11
15	*p*-NO_2	0.33	(1) 0.22, (3) −0.11, (7) 0.15, (8) 0.22, (12) −0.10, (19) −0.14
16	*p*-F	0.02	
17	*p*-CH_3	0.05	(1) −0.13
18	*p*-OCH_3	0.02	
19	*p*-OH	0.07	(8) −0.12, (15) −0.14

[a] Four factors in the rotation matrix. [b] Each row is the result of a separate uniqueness test vector.

---

(6) C. G. Swain and C. B. Scott, *J. Amer. Chem. Soc.*, **75**, 141 (1953).

(7) J. O. Edwards, *ibid.*, **76**, 1541 (1954).

(8) (a) R. S. Drago and B. B. Wayland, *ibid.*, **87**, 3571 (1965); (b) R. S. Drago, G. C. Vogel, and T. E. Neldham, *ibid.*, **93**, 6014 (1971).

(9) W. L. Jolly, J. D. Illige, and M. H. Mendelsohn, *Inorg. Chem.*, **11**, 869 (1972).

Table II. Uniqueness Test of Solvents[a]

A	Methanol	0.33	(B) 0.19, (E) 0.23, (F) 0.27, (H) −0.20
B	Ethanol	0.33	(A) 0.19, (C) −0.10, (D) 0.26
C	Ethylene glycol	0.93	(B) −0.10, (D) 0.19
D	Butanol	0.33	(B) 0.26, (C) 0.19, (E) 0.24, (F) −0.18, (H) 0.14
E	Propanol	0.32	(A) 0.23, (B) 0.32, (D) 0.24
F	Dioxane–water ($\epsilon = 55$)[b]	0.58	(A) 0.27, (D) −0.18, (G) 0.35
G	Dioxane–water ($\epsilon = 40$)	0.39	(F) 0.35, (H) 0.31
H	Dioxane–water ($\epsilon = 13$)	0.79	(A) −0.20, (D) 0.14, (G) 0.31

[a] Four factors in the rotation matrix. [b] Dielectric constant of mixture.

test in which a test vector was created containing a unity for the solute or solvent in question and zeros for all other values on the vector. Also shown for each test vector are the other solvents or solutes which yielded relatively "high" predicted values even though they were assigned a zero value on the test vector. These additional high values indicate that the solute or solvent in question share a common factor with the case tested. In the present paper, our main interest with these tables is in ascertaining if any of the solutes or solvents is behaving in a unique fashion which could introduce an additional dimension or factor into the factor space. In other words, one of the four factors found necessary to reproduce the data might be localized to a given solute or solvent in the scheme. This would indicate that it is not a general type of interaction. In Table I, containing tests for uniqueness of the solutes, there is no one solute which has a high uniqueness value that does not also have several other solutes sharing this same uniqueness. Therefore, there does not seem to be any one solute which is behaving in a unique manner. In Table II for the solvents, ethylene glycol has quite a high predicted uniqueness value, which is not shared by any other solvents, indicating that there is some unique factor associable only with this solvent. Therefore, for the acidity data in all solvents except ethylene glycol, only three terms are required in eq 4, while four terms are required for the data in ethylene glycol. Since we have already postulated that one of the terms in eq 4 is related to the gas phase acidity ratio, it is only necessary to account for two dominant solute–solvent interaction terms for all solvents except ethylene glycol. For ethylene glycol, three interaction terms must be introduced. Furthermore, to fit our model, each interaction term must be expressable in product form.[4d] The gas phase acidity term is already in the correct product function form with the solute part corresponding to the ln of the relative gas phase acidities and the solvent part being unity.

The two important solute–solvent interaction terms which will be considered are the electrostatic and van der Waals' terms. The theoretical expression for the former has been worked out by Kirkwood and Westheimer.[10] The expression that they arrived at is shown below.

$$\left(\ln \frac{K(i,\alpha)}{K_0}\right)_E = \left[\frac{e\mu_i \cos\theta}{2.303RTr^2}\right]_u \left[\frac{1}{\epsilon_\alpha}\right]_v \tag{5}$$

In this expression $(\ln(K(i,\alpha)/K_0))_E$ is defined by us as the electrostatic part of the measured solution acidity ratio, while Kirkwood and Westheimer defined e as the charge of an electron; μ_i is the bond moment of the substituted group in the benzoic acid; R is the gas constant, T the solution temperature, r the distance between the dipole of the substituent and the acid group, θ the angle between the dipole axis and the hydrogen of the acid, and ϵ_α the solvent dielectric constant. This expression is already in the correct product form to test by factor analysis, with the terms in the second bracket referring to solvent properties.

For the van der Waals' or dispersion term, we have used the model of Grunwald and Price.[11] These authors used the dispersion model of Linder,[12] in which the solute and solvent are treated as though they contain oscillating dipoles. Grunwald and Price proposed that the van der Waals' interaction is caused by localized interactions between solute and solvent oscillating dipoles. In their model, one is faced with having detailed microscopic information concerning the structure of the solvent cage surrounding the various solute structures. This information is not always readily available. An alternate approach is to treat the solvent as a dielectric continuum.[12] Then the solute, rather than interacting with individual solvent molecules, can be treated as interacting with a dielectric continuum having the bulk properties of the solvent. Using this latter model, Weiner and Malinowski[4f] studied the van der Waals' interactions on solvent effect observed in nmr. They redeveloped this model and showed that it could account for the observed gas to solution chemical shifts of nonpolar solutes. This shift is usually thought to arise solely from van der Waals' interactions. It is proposed that this same continuum model be used to account for the observed dispersion interactions in the present problem. An equation to account for the dispersion interaction can be derived as an extension of Linder's theory.[12b]

Linder[12] stated that the partial molal dispersion free energy, F, of a solute, 2, in a solvent, 1, is given by

$$F_2 = -{}^1/_4\langle m_2{}^2\rangle g[\nu_1/(\nu_1 + \nu_2)] \tag{6}$$

where $\langle m_2{}^2\rangle$ is the average square of the solute oscillating dipole moment, ν refers to a mean absorption frequency of the molecule, and g is given by

$$g = \frac{2n_1{}^2 - 2}{2n_1{}^2 + 1}\frac{1}{a_2{}^2} \tag{7}$$

where n is the index of refraction and a_2 is the Onsager radius of the solute.

Now according to Linder

$$\langle m^2\rangle = {}^3/_2 h\nu\alpha \tag{8}$$

where h is Planck's constant and α is the polarizability. Also, $\langle m^2\rangle$ can be related to the radius of the electric cloud of the molecule by the following equation

$$\langle m^2\rangle = e^2\sum\langle r_j{}^2\rangle \tag{9}$$

where e is the charge on the electron. Substituting eq 7 and 8 into 6, we can write

$$F_2 = \frac{-3h}{8}\frac{2n_1{}^2 - 2}{2n_1{}^2 + 1}\frac{\alpha_2}{a_2{}^3}\frac{\nu_1\nu_2}{\nu_1 + \nu_2} \tag{10}$$

---

(10) J. G. Kirkwood and F. H. Westheimer, *J. Chem. Phys.*, **6**, 506, 513 (1938).

(11) E. Grunwald and E. Price, *J. Amer. Chem. Soc.*, **86**, 4517 (1964).

(12) (a) B. Linder, *J. Chem. Phys.*, **33**, 668 (1960); (b) B. B. Howard, B. Linder, and M. T. Emerson, *ibid.*, **36**, 485 (1962).

Table III. Values for Various Parameters Used in Eq 14[a]

	α[b]	χ_d[c]	ϵ[d]
Methanol	3.2	23.8	31.5
Ethanol	5.19	36.0	24.2
Propanol	7.01	48.0	20.2
Butanol	8.8	59.7	17.5
Ethylene glycol	5.76	43.7	37.6
Dioxane–H_2O (ϵ = 55)[f]	1.98	15.1	55
Dioxane–H_2O (ϵ = 40)[f]	2.48	19.1	40
Dioxane–H_2O (ϵ = 15)[f]	4.13	30.8	15
Dioxane	8.8	64[g]	2.2[e]
Water	1.48	12	78.5[e]

[a] All experimental values of α and χ_d were taken from the book by Dorfman, ref 14. [b] Polarizability × 10^{-24} cm^3/molecule. [c] Diamagnetic susceptibility × 10^{-6} cm^3/mol. [d] Values of dielectric constants reported in papers of Elliott and Kilpatrick, see ref 5. [e] R. C. Weast, Ed., "Handbook of Chemistry and Physics," 51 ed. Chemial Rubber Co., Cleveland, Ohio 44128. [f] Values of α and χ_d were calculated as mole fraction averages over the pure constituent values. [g] Estimated from group contributions.

Table IV. Test of Solvent Factors

	Unity		ϵ^{-1} [a]		van der Waals' effect[b]		Uniqueness for ethylene glycol	
	Exptl	Pred	Exptl	Pred	Exptl	Pred	Exptl	Pred
Methanol	1	1.02	3.17	2.89	7.44	7.41	0	0.00
Ethanol	1	1.00	4.14	4.98	6.93	6.78	0	−0.10
Ethylene glycol	1	0.96	2.66	2.95	7.59	7.59	1	0.93
Butanol	1	0.97	5.73	5.31	6.78	6.72	0	0.19
Propanol	1	1.03	4.98	4.66	6.85	7.08	0	0.00
Dioxane–water (ϵ = 55)	1	1.00	1.82	1.29	7.84	7.84	0	0.00
Dioxane–water (ϵ = 40)	1	1.00	2.50	3.36	7.72	7.68	0	0.00
Dioxane–water (ϵ = 15)	1	1.00	6.68	6.23	7.46	7.49	0	0.00

[a] Dielectric constants reported by Elliot and Kilpatrick in their papers, ref 5, ×10^{-2}. [b] ×10^{-9}, values needed for the solvent part of eq 13 were obtained from the book by Dorfman, ref 14. See also ref 4f for further discussion of this calculation.

Now if we let $V_2 = {}^4/_3\pi N a_2{}^3$, where V_2 is the molar volume of the solute, eq 10 can be rewritten as

$$F_2 = \frac{-h\pi N\alpha_2}{2V_2}\left[\frac{2n_1{}^2 - 2}{2n_1{}^2 + 1}\right]\frac{\nu_1\nu_2}{\nu_1 + \nu_2} \qquad (11)$$

With the help of eq 8 and 9, this can be rearranged to yield

$$F_2 = \frac{-1}{3}\left[\frac{e^2\pi N\alpha_2}{V_2}\right]\left[\frac{2n_1{}^2 - 2}{2n_1{}^2 + 1}\cdot\frac{\Sigma\langle r^2(1)\rangle}{\alpha_1}\cdot\frac{\nu_2}{\nu_1 + \nu_2}\right] \qquad (12)$$

Now if we make the assumption that maximum absorption occurs when $\nu_1 \approx \nu_2$, which was also made by Linder, eq 12 finally simplifies to

$$F_2 = \frac{-1}{6}e^2\pi N^2\frac{\alpha_2}{V_2}\left(\frac{2n_1{}^2 - 2}{2n_1{}^2 + 1}\frac{\Sigma\langle r^2(1)\rangle}{\alpha_1}\right) \qquad (13)$$

This formula gives the free energy of dispersion interaction of an oscillating dipole with a continuum. The solvent term, $\Sigma\langle r^2(1)\rangle$, can be evaluated from the diamagnetic susceptibility.[4f] What we are interested in here is the difference in free energy between two different solutes in a given solvent; therefore, eq 12 can be written for each solute. The difference between these two equations corresponds to the difference in free energy for dispersion interactions of the two solutes with the solvent α. This equation is developed for the interaction of an oscillating dipole with a solvent. If there is more than one solute oscillating dipole which is important, or we are dealing with a oscillating dipole of an ionic species, then the solute part of eq 13 may change. However, we will assume that the solvent dependence should remain the same.

For ethylene glycol, the identity of the fourth factor is not known. It can be accounted for in our analysis, however, by using the ethylene glycol uniqueness test vector itself as a measure of this factor. As a further test of the validity of this substitution, ethylene glycol can also be dropped as a solvent from the scheme. If there is truly a unique factor associable with the ethylene glycol, then the reduced data set should only need three factors to span the space, rather than four. Furthermore, the same final results and conclusions (*i.e.*, predicted relative gas phase acidities) should be reached in terms of the three remaining factors. In the following sections, the results of both approaches will be presented.

One can apply factor analysis to a data set to study properties associated with either the solutes or the solvents in the scheme, depending on whether one originally is considering the data matrix or its transpose.[4a,d] In the present problem, several of the solute properties such as the gas phase acidity ratio and the solute part for the unique ethylene glycol factor are unknown. The four solvent factors can be evaluated; therefore, the data matrix will be set up to test solvent properties. The equation being tested is

$$\ln\frac{K(i,\alpha)}{K_0} = \ln\frac{K(i,\text{gas})}{K_0}\cdot 1 + S_E(i)S_E(\alpha) + S_D(i)S_D(\alpha) + S_{EG}(i)S_{EG}(\text{ethylene glycol}) \qquad (14)$$

or

$$\ln\frac{K(i,\alpha)}{K_0} = \ln\frac{K(i,\text{gas})}{K_0}\cdot 1 + S_E(i)\frac{1}{\epsilon_\alpha} + S_D(i)\frac{\Sigma\langle r^2\rangle}{\alpha} + S_{EG}(i)S_{EG}(\text{ethylene glycol})$$

The numerical values of the various parameters used to calculate the terms in eq 14 are shown in Table III. In this expression, each term is in the correct product

Table V. Test of Solvent Factors, without Ethylene Glycol[a]

	Unity		ϵ^{-1}[b]		VW[c]	
	Exptl	Pred	Exptl	Pred	Exptl	Pred
Methanol	1	1.02	3.17	2.97	7.44	7.41
Ethanol	1	0.99	4.17	5.04	6.93	6.86
Butanol	1	0.96	5.73	5.28	6.78	6.59
Propanol	1	1.03	4.98	4.73	6.84	7.14
Dioxane–water (ϵ = 55)	1	0.98	1.82	1.36	7.84	7.78
Dioxane–water (ϵ = 40)	1	1.00	2.50	3.42	7.72	7.74
Dioxane–water (ϵ = 15)	1	1.01	6.68	6.20	7.46	7.47

[a] A three dimensional rotation was used. [b] See Table IV, footnote *a*. [c] See Table IV, footnote *b*.

Table VI. Comparison between Original and Recalculated ln $(K(i,\alpha)/K_0)$ Using Three and Four Factors

	Solvents							
Solute	Methanol	Ethanol	Propanol	Butanol	Ethylene glycol[g]	Dioxane–water[a]	Dioxane–water[b]	Dioxane–water[c]
o-NO$_2$	1.83[d]	1.77	1.80	1.78	1.74	1.78	1.66	1.58
	1.77,[e] 1.77[f]	1.83, 1.84	1.79, 1.80	1.76, 1.77	1.76	1.75, 1.76	1.74, 1.74	1.53, 1.52
o-I	1.19	1.08	1.10	1.04	1.10	1.00	0.87	0.75
	1.04, 1.04	1.16, 1.16	1.11, 1.11	1.06, 1.06	1.13	1.00, 1.01	0.99, 0.99	0.66, 0.66
o-Br	1.27	1.16	1.16	1.09	1.20	1.13	1.00	0.89
	1.14, 1.14	1.22, 1.22	1.17, 1.17	1.12, 1.12	1.21	1.13, 1.14	1.10, 1.11	0.80, 0.80
o-Cl	1.21	1.12	1.09	1.08	1.14	1.09	0.97	0.84
	1.09, 1.09	1.17, 1.17	1.12, 1.12	1.08, 1.08	1.16	1.08, 1.09	1.06, 1.06	0.77, 0.77
o-CH$_3$	0.09	0.02	0.04	0.00	0.05	0.04	−0.08	−0.21
	0.02, 0.02	0.08, 0.08	0.04, 0.04	0.00, 0.00	0.06	0.02, 0.02	0.00, 0.00	−0.27, 0.27
o-OCH$_3$	0.17	0.24	0.26	0.28	0.18	−0.06	−0.10	−0.07
	0.07, 0.08	0.28, 0.29	0.28, 0.28	0.28, 0.28	0.20	−0.06, −0.05	−0.03, −0.03	−0.12, −0.12
o-OH	1.49	1.63	1.57	1.50	1.50	1.49	1.60	1.78
	1.54, 1.54	1.52, 1.51	1.56, 1.56	1.60, 1.60	1.45	1.52, 1.51	1.55, 1.54	1.80, 1.81
m-NO$_2$	1.05	1.17	1.15	1.10	0.93	0.89	0.97	1.06
	1.02, 1.02	1.13, 1.12	1.15, 1.15	1.17, 1.17	0.91	0.93, 0.92	0.96, 0.95	1.05, 1.06
m-I	0.55	0.62	0.62	0.57	0.49	0.37	0.40	0.46
	0.49, 0.49	0.61, 0.61	0.62, 0.62	0.62, 0.62	0.48	0.40, 0.40	0.42, 0.42	0.42, 0.43
m-Br	0.60	0.65	0.63	0.58	0.54	0.43	0.47	0.53
	0.54, 0.53	0.63, 0.62	0.64, 0.64	0.64, 0.64	0.53	0.47, 0.46	0.48, 0.48	0.50, 0.50
m-Cl	0.59	0.63	0.61	0.59	0.52	0.44	0.48	0.53
	0.53, 0.53	0.62, 0.61	0.62, 0.62	0.63, 0.63	0.51	0.47, 0.47	0.49, 0.49	0.51, 0.51
m-F	0.51	0.53	0.52	0.42	0.45	0.41	0.41	0.49
	0.46, 0.46	0.50, 0.50	0.51, 0.50	0.51, 0.51	0.43	0.44, 0.43	0.44, 0.44	0.45, 0.46
m-CH$_3$	−0.09	−0.06	−0.10	−0.10	−0.09	−0.12	−0.14	−0.14
	−0.11, −0.11	−0.07, −0.07	−0.08, −0.08	−0.08, −0.08	−0.09	−0.12, −0.12	−0.12, −0.12	−0.16, −0.16
m-OH	−0.11	−0.16	−0.16	−0.16	−0.03	0.01	0.00	−0.02
	−0.07, −0.07	−0.17, −0.17	−0.18, −0.18	−0.18, −0.18	−0.03	0.00, 0.00	−0.02, −0.02	0.00, 0.00
p-NO$_2$	1.02	1.17	1.14	1.14	0.97	0.97	1.10	1.18
	1.06, 1.06	1.11, 1.11	1.14, 1.14	1.17, 1.18	0.94	1.00, 1.00	1.04, 1.03	1.22, 1.22
p-F	0.19	0.23	0.21	0.22	0.17	0.15	0.22	0.25
	0.19, 0.19	0.20, 0.20	0.22, 0.22	0.23, 0.23	0.16	0.17, 0.17	0.19, 0.18	0.27, 0.27
p-CH$_3$	−0.18	−0.18	−0.21	−0.19	−0.17	−0.23	−0.22	−0.21
	−0.20, −0.20	−0.19, −0.19	−0.19, −0.19	−0.19, −0.19	−0.17	−0.22, −0.22	−0.21, −0.22	−0.21, −0.22
p-OCH$_3$	−0.36	−0.32	−0.36	−0.36	−0.32	−0.33	−0.32	−0.36
	−0.34, −0.34	−0.35, −0.35	−0.35, −0.35	−0.36, −0.36	−0.33	−0.33, −0.33	−0.34, −0.34	−0.36, −0.35
p-OH	−0.53	−0.55	−0.57	−0.56	−0.45	−0.42	−0.47	−0.53
	−0.50, −0.50	−0.55, −0.55	−0.57, −0.57	−0.58, −0.58	−0.44	−0.45, −0.45	−0.47, −0.46	−0.52, −0.53

[a] Dioxane–water mixture, dielectric constant equals 55. [b] Dioxane–water mixture, dielectric constant equals 40. [c] Dioxane–water mixture, dielectric constant equals 15. [d] Experimental data of Kilpatrick and Elliott. [e] Predicted data using four factors. [f] Predicted data using three factors. [g] Three factors are not applicable for ethylene glycol.

form of solute and solvent properties, with $S_E(i)$ corresponding to the difference in the solute part of electrostatic free energy for the substituted and unsubstituted benzoic acid; $S_E(\alpha)$ being the solvent part of this expression (see eq 5); $S_D(i)$ being the difference in the solute dispersion interaction formula and $S_D(\alpha)$ being the solvent part (see eq 13); and $S_{EG}(i)$ being the solute part of the unique factor associable with ethylene glycol and S_{EG}(ethylene glycol) being the uniqueness test factor for ethylene glycol. Here we will be testing the solvent part of each term. In eq 14, we have lumped the index of refraction term of the van der Waals' effect into the solute coefficient since its variation is almost negligible among the solvents.

The tests of the four solvent factors are shown in Table IV, while the results with three factors without the relative solution acidity data in ethylene glycol included in the data set are shown in Table V. The fit of each suspected test factor will be discussed separately below.

(a) Solvent Electrostatic Term, $S_E(\alpha)$. The solvent part of the electrostatic interaction term will be approximated by the inverse of the bulk solvent dielectric constants as reported by Kilpatrick and Elliot.[5] As pointed out by many authors[13] this is only an approximate relationship; therefore, we cannot expect more than to predict the right trend. Both in Tables IV and V, the agreement between experimental and predicted values for this test factor is satisfactory.

(b) Solvent Part of Gas Phase Term. The solvent part of the gas phase term is unity, since this term is

---

(13) See, for example, S. Ehrenson, "Progress in Physical Organic Chemistry," Vol. 2, S. G. Cohen, A. Streitwieser, Jr., and R. W. Taft, Ed., Interscience, New York, N. Y., 1964.

supposedly independent of solvent. Therefore, the test of a constant unity factor is shown in Tables IV and V. Again the agreement between experimental and predicted values is quite good, indicating that the gas phase term can be separated in this manner.

(c) Solvent Part of the Dispersion Term, $S_D(\alpha)$. The data needed to calculate the solvent part of the dispersion term was taken from the book by Dorfman[14] who reported the diamagnetic susceptibilities for the solvents employed here. These can be used to calculate the $\Sigma\langle r^2\rangle$ term in eq 13. For the mixtures, it was assumed that the quantities needed for eq 13 could be calculated as a simple mole fraction average of the pure solvent properties. Again the fit in Tables IV and V is good.

(d) Uniqueness for Ethylene Glycol Solvent, $S_{EG}(\alpha)$. Shown in Table IV is the fit of the uniqueness test on ethylene glycol. Again the fit is good.

The real test, needed to ascertain if all four important factors have been accounted for by our choice of test factors, is to attempt to reproduce the original data set in terms of these four solvent parameters. The original data set will only be reproduced if the four test factors truly account for the four important abstract factors of the space. Each factor can separately test as a factor, but until one attempts to replace simultaneously all abstract factors with their corresponding physically significant parameters, it is not possible to judge whether all important factors have been accounted for. The results of this test using the four calculated solvent factors in Table IV and the three calculated solvent test factors in Table V are shown in Table VI. Also shown in Table VI are the raw relative solution acidity data of Kilpatrick and Elliot. The mean error in the four factors case is ±*0.03* ln unit and in the three factors case is also ±*0.03* ln unit. If one compares the fit on a point by point basis in Table VI, then it is also observed that the fit is the same for both cases.

The fit is the best for the para-substituted acids and is progressively less satisfactory for the meta and ortho substituents. Possibly, some factor such as steric effects are not being totally accounted for in the present formulation of the four test factors. Among the solvents, the best overall fit occurs in the case of propanol and butanol, with the poorest occurring in the case of methanol. At present, no explanation is apparent for these results.

Isolation and Prediction of Relative Gas Phase Acidities

Table VII. Comparison of Solute Coefficients Using Three and Four Factors

	Solute coefficients of solvent term			
Solute	Unity	ϵ^{-1}	VW	Uniq EG[c]
o-NO_2	(3.949)[a]	(−0.069)	(−0.263)	
	3.907[b]	−0.067	−0.259	−0.000
o-I	(4.677)	(−0.107)	(−0.443)	
	4.637	−0.105	−0.439	0.010
o-Br	(4.001)	(−0.096)	(−0.343)	
	3.973	−0.094	−0.341	0.075
o-Cl	(3.848)	(−0.092)	(−0.331)	
	3.808	−0.089	−0.327	0.071
o-CH_3	(2.343)	(−0.081)	(−0.227)	
	2.306	−0.079	−0.273	0.046
o-OCH_3	(4.097)	(−0.055)	(−0.517)	
	4.064	−0.053	−0.514	0.169
o-OH	(−0.179)	(0.076)	(0.198)	
	−0.095	0.072	0.190	−0.082
m-NO_2	(2.385)	(0.013)	(−0.189)	
	2.433	0.010	−0.194	−0.078
m-I	(2.572)	(−0.016)	(−0.274)	
	2.594	−0.017	−0.275	0.023
m-Br	(2.045)	(−0.008)	(−0.199)	
	2.075	−0.010	−0.202	0.018
m-Cl	(1.787)	(−0.040)	(−0.167)	
	1.809	−0.050	−0.169	−0.003
m-F	(1.047)	(−0.001)	(−0.078)	
	1.091	−0.002	−0.082	−0.029
m-CH_3	(0.622)	(−0.014)	(−0.092)	
	0.622	−0.014	−0.092	0.024
m-OH	(−1.862)	(0.018)	(0.233)	
	−1.863	0.018	0.233	0.012
p-NO_2	(0.890)	(0.046)	(0.003)	
	0.939	0.043	−0.002	−0.102
p-F	(−0.069)	(0.023)	(0.025)	
	−0.045	0.022	0.023	−0.030
p-CH_3	(0.101)	(−0.027)	(−0.040)	
	0.104	−0.003	−0.040	0.036
p-OCH_3	(−0.412)	(−0.003)	(−0.010)	
	−0.403	−0.003	−0.010	0.009
p-OH	(−1.176)	(−0.009)	(0.095)	
	−1.189	−0.008	0.096	0.037

[a] Coefficients determined with three factors. [b] Coefficients determined with four factors. [c] Solute part of uniqueness of ethylene glycol solvent.

Since there is good agreement between experimental and predicted solution acidity ratios for both the three and four factor cases, then one can state that the present chosen physical factors do vary in similar manner to the true factors of the solute–solvent interaction space and that all important interactions have correctly been accounted for. One could not get such close agreement between experimental and predicted values unless there were some validity in the formulism of the test factors (*i.e.*, each test factor could separately test well, but the original data do not have to be accurately reproduced by an arbitrary combination of these test factors). One can therefore make a simultaneous rotation into the four (three) solvent factors and determine the appropriate solute factor terms or loadings of each solute on each solvent factor. These results are shown in Table VII.

Referring to eq 14, one can see that the solute coefficients of the solvent unity test factor term should correspond to the relative gas phase acidities values that we were originally seeking. For clarity, these values will also be reported in Table VIII as the average of the two values determined with and without ethylene glycol as

Table VIII. Calculated Values of Gas Phase Acidity Relative to That for Benzoic Acid[a]

Solute	$\ln \frac{K(i,\text{gas})}{K_0}$	Solute	$\ln \frac{K(i,\text{gas})}{K_0}$
o-NO_2	3.93	m-Cl	1.80
o-I	4.66	m-F	1.06
o-Br	3.98	m-CH_3	0.62
o-Cl	3.83	m-OH	−1.86
o-CH_3	2.32	p-NO_2	0.91
o-OCH_3	4.08	p-F	−0.05
o-OH	−0.14	p-CH_3	0.10
m-NO_2	2.40	p-OCH_3	−0.40
m-I	2.58	p-OH	−1.18
m-Br	2.06		

[a] Solute coefficients of solvent unity test factor; see eq 14, also see column 1, Table VI.

---

(14) Ya. G. Dorfman, "Diamagnetism on the Chemical Bond," Elsevier, New York, N. Y., 1965.

a solvent in the data space. Unfortunately, there are no experimental values available for comparison purposes. However, a paper has recently appeared by Bohme, *et al.*,[2a] where many experimental relative gas phase acidities are reported. The order of magnitude of our results for the relative gas phase acidity rates reported in Table VIII is quite reasonable in the light of their experimental data. Our data indicate that an ortho substituent has a greater effect on gas phase acidities than on the corresponding solution acidities. This effect decreases as the substituent is moved to the metal and para positions. However, a more detailed critique of our predicted values must wait until the appropriate experimental data are available.

Referring back to Table VII, it is interesting to further note that the solute coefficients of the uniqueness for ethylene glycol solvent have relatively high values only for o-OCH_3 and p-NO_2. This might indicate that either a true unique type of solute–solvent interaction is occurring for these two solutes in ethylene glycol, or possibly that there is some error in the reported values of these cases.

In conclusion, an attempt was made to show that the mathematical technique of factor analysis can be quite useful in studying the effects of solvent on reaction rates. In the present case, it was shown that for substituted benzoic acids in various solvents only two main solvent effects must be accounted for, namely electrostatic and van der Waals' interactions, and that the latter interaction can be calculated in terms of the continuum model of Linder.[12a] Further, the linear free energy model, as expressed by eq 4, can be used to isolate and predict relative gas phase acidities from solution data. Listings of the factor analysis computer program are available on request.

Acknowledgments. The author would like to thank Dr. Andrew P. Stefani, Dr. Robert B. Scott, Dr. Robert M. Metzger, Dr. Norman E. Heimer, and Dr. John I. Brauman for their helpful comments and discussions during the preparation of this manuscript.

Gas-Phase Reaction of *tert*-Butyl Ions with Arenes. Remarkable Selectivity of a Gaseous, Charged Electrophile

Fulvio Cacace* and Pierluigi Giacomello

Contribution from the Laboratorio di Chimica Nucleare del C.N.R.,
Istituto di Chimica Farmaceutica, University of Rome, 00100 Rome, Italy.
Received March 6, 1973

Abstract: *tert*-Butyl ions from the γ radiolysis of neopentane were allowed to react in the gas phase with benzene, toluene, and their mixtures, contained at low concentrations in the gas. The reaction was studied at neopentane pressures ranging from 8.7 to 720 Torr, in the presence of a thermal radical scavenger (O_2). The effects of a gaseous base (EtOH) and of an inert gas (Ar) on the reaction were also investigated. The *tert*-butyl ions react with toluene in the gas phase, yielding *p*- and *m-tert*-butyltoluenes, whose ratio markedly increases with the pressure of neopentane and the concentration of the base. At the higher neopentane and base pressures, the isomeric composition of the products appears to be kinetically controlled and corresponds to a para : $^1/_2$ meta ratio in excess of 35. This value is fully comparable to the highest ratios measured in the study of the liquid-phase *tert*-butylation and reflects the remarkable *positional* selectivity of the gaseous, unsolvated *tert*-butyl cation. Furthermore, the gaseous reagent displays a significant *substrate* selectivity, as indicated by a k_T/k_B ratio in excess of 55 measured in the competition experiments. The mechanism of the gas-phase alkylation is discussed and compared with those of different gaseous electrophiles, characterized by a considerably lower selectivity. It is concluded that a low *positional* and *substrate* selectivity does not necessarily represent a typical mechanistic feature of the gas-phase aromatic substitutions by charged reagents and that the low selectivity observed in previous studies is likely to arise from the extreme electrophilic character of the reagents so far investigated.

New experimental techniques that allow the determination of the neutral reaction products and their isomeric composition have been recently applied to the study of gas-phase aromatic substitutions involving charged electrophiles, whose positional selectivity could be therefore measured for the first time.

The results indicate that the reagents employed, including He^3H^+ from the β decay of $^3H_2$,[1-3] $^{80}Br^+$ from the isomeric transition of $CH_3{}^{80m}Br$,[4] and radiolytically formed $^2H_2{}^3H^+$[5] and i-$C_3H_7^+$,[6,7] display indeed a significant *positional* selectivity.

On the other hand, the selectivity measured in the gas phase was always substantially lower than in most aromatic substitutions occurring in solution.

From the available data it could hardly be decided whether the low positional selectivity represented a general mechanistic feature of the gas-phase attack or should rather be traced to the specific reactivity of the

---

(1) F. Cacace and S. Caronna, *J. Amer. Chem. Soc.*, **89**, 6848 (1967).
(2) F. Cacace and G. Perez, *J. Chem. Soc. B*, 2086 (1971).
(3) F. Cacace, R. Cipollini, and G. Ciranni, *J. Chem. Soc. B*, 2089 (1971).
(4) F. Cacace and G. Stöcklin, *J. Amer. Chem. Soc.*, **94**, 2518 (1972).
(5) F. Cacace, R. Cipollini, and G. Occhiucci, *J. Chem. Soc., Perkin Trans. 2*, 84 (1972).
(6) S. Takamuku, K. Iseda, and H. Sakurai, *J. Amer. Chem. Soc.*, **93**, 2420 (1971).
(7) F. Cacace and E. Possagno, *J. Amer. Chem. Soc.*, **95**, 3397 (1973).