

Multivariate Analysis of Solvolysis Kinetic Data; an Empirical Classification paralleling Charge Delocalization in the Transition State

By **Christer Albano** and **Svante Wold**,* Research Group for Chemometrics, Institute of Chemistry, Umeå University, S-901 87 Umeå, Sweden

A multivariate data analysis of solvolysis rate constants measured for a range of solvents of different ionization power provides a pattern parallel to the degree of charge delocalization in the transition state. In this way direct evidence is obtained for a difference in kinetic behaviour between *exo*- and *endo*-norbornyl solvolyses.

THE degree of charge delocalization in the transition state of solvolysis reactions has been under continuous debate following the proposal by Winstein and his co-workers^{1,2} on non-classical carbonium ions being intermediates in certain solvolysis reactions.

In a nonclassical ion the positive charge would be delocalized by a single bond or by a double or triple bond which is not in the allylic position. The 'classical' ion has the positive charge delocalized by resonance (involving an unshared pair of electrons or a double or triple bond in the allylic position).

The researchers in the field have divided into two schools. One which presently has the stronger support, assumes the existence of nonclassical carbonium ions as intermediates in some solvolyses. The second school assumes the intermediate to be a rapid equilibrium between several 'classical' ions.³

The evidence for or against the nonclassical concept has been collected by studying the solvolytic reactions of certain organic molecules (the norbornyl system is popular) adopting techniques such as n.m.r. spectroscopy in cooled superacid media, deuterium isotopic effects, variation of solvolytic conditions, variation of the substitution pattern on the parent molecule, *etc.* The empirical data have mostly been used in attempts to provide more or less detailed information on the transition states of the reactions. This can be called the mechanistic approach to this type of problem. The approach is elucidated in many textbooks,^{4,5} review articles,⁶⁻⁸ and in special publications.^{3,9}

This paper is an attempt at a more direct approach to the study of charge delocalization in the transition state. This approach is based on the multivariate analysis of parts of the huge amount of solvolysis kinetic data which have been collected over the years. The advantage with this approach is that it is not based on a specific theory relating the observed data to the structure of the transition state. Rather, the approach is based on a general theory of the behaviour of data observed for similar processes. This theory, which is an extension of the familiar concept of linear free energy relationships, makes it possible to arrive at reliable conclusions provided that the data have been selected in a way relevant to the actual problem.

Design of the Study.—The study is based on the property of certain mathematical models of adequately approximating multivariate data observed for similar

systems or processes. By selecting processes of known types and multivariate data observed for these processes it is possible to use these mathematical models to (a) find the 'pattern' in the data specific for each type of process and (b) thereafter classify new processes as belonging to this or that type of process.

In the present case we studied processes in the form of solvolysis reactions. Two types of processes are of interest, those with a delocalized charge in the transition state and those lacking such charge delocalization. One can, in principle, have a continuum in the degree of charge delocalization from a completely localized charge on one atom in, say, the solvolysis of methyl tosylate to the delocalized charge of, say, benzyl tosylate. Therefore, in the present case we have analysed all processes in one single group and instead tried to relate the position within the group to the degree of charge delocalization in the transition state. In order to minimize the bias of missing data estimates, however, values of these estimates have been calculated separately for primary and secondary substrates.

The variables used in an empirical similarity study must be such that they are related to the classification problem in question. When trying to distinguish between different degrees of charge delocalization, we have chosen kinetic data at two temperatures and from seven solvents or solvent mixtures. In that way each reaction is characterized by 14 variables. We assume that part of the variation of the kinetic data in different solvents is related to the different charge stabilizing ability of the solvents. Thus it might be possible for a series of reactions in different solvents to be separated into classes depending on their degree of charge delocalization in the transition state. We expect reactions where the transition state charge is localized to show greater differences (in velocity) between solvents of different polarity, than reactions where the charge is delocalized. Such differences would be interpretable provided that the same fractional charge is developed in each reaction in the series.

The requirement to construct a data matrix with contains a number of solvolysis reactions run in several solvents limited us to certain common solvent systems and to the use of sulphonate esters as substrates. In spite of this limitation the data matrix was not complete. Out of 364 possible rate constants, 100 are missing. The handling of the problem of the missing

data is discussed later. Substrates (R-X) with the leaving groups benzenesulphonate (X = OBz), toluene-*p*-sulphonate (X = OTs), and *p*-bromobenzenesulphonate (X = OBs) have been chosen among the sulphonate esters because they have been frequently studied in kinetic investigations and also in order to obtain a data set representative of different degrees of charge delocalization in the transition state. The following division of substrates used in the study has been carried out: (a) R groups known for resonance delocalization, benzyl [substrate (22)] and allyl (24); (b) R groups expected to have nonclassical delocalization, *exo*-2-norbornyl (20) and (25) and cyclopropylmethyl (23); (c) R groups not expected to have delocalization (classical), *endo*-2-norbornyl (21) and (26); (d) R groups considered as normal classical ions, primary, methyl (1) and (4), ethyl (2), (5), and (7), isobutyl (10), neopentyl (16), and phenylethyl (17), secondary, isopropyl (3), (6), and (8), *s*-butyl (9), 2-adamantyl (14), cyclopentyl (11), cyclohexyl (12), and 1-phenyl-1-methylethyl (18); (e) R groups which are interesting but difficult to classify, 1-adamantyl (19) and cyclo-octyl (13) and (15).

Most solvolytic data used in this study were obtained from the collection published in 1972 by the Laboratory of Chemical Kinetics and Catalysis at Tartu State University which contains all rates for first-order solvolytic substitution and elimination reactions at non-aromatic reaction centres. These data were completed with more recently published data (see Table 1). As some of the data are old and perhaps not so reliable and as many log *k* values are missing we would appreciate if the readers of this paper could help us make the data matrix more complete.

Mathematical Similarity Models.—By means of a general mathematical derivation¹⁰⁻¹³ it can be shown that the data Y_{ik} observed for an ensemble of similar processes (process *k*, variable *i*) can be adequately approximated by the principal components model (1).

$$Y_{ik} = \alpha_i + \sum_{a=1}^A \beta_{ia} \theta_{ak} + \epsilon_{ik} \quad (1)$$

Here α and β are parameters specific to the variable *i*, in the present case the logarithmic reaction rate for the solvent-temperature combination *i*. The parameter θ is specific to the process *k*, in the present case substrate *k*. The number of *A* terms is smaller the closer the similarity between the processes in the ensemble. The residual ϵ_{ik} represent the part of *y* which shows random behaviour.

Model (1) is seen to have the form of a generalized linear free energy relationship (l.f.e.r.). It should be valid provided that the following assumptions are fulfilled. (1) The processes (index *k*) have some degree of similarity. (2) The variables (index *i*) are related to this similarity and have continuity properties such that the difference in one variable between two processes is smaller the closer the similarity between the two processes. These two assumptions are presumably well fulfilled in the present study. All studied processes

(reactions) are of the same type (solvolyse) and thus similar. The variables characterizing the processes are directly related to this similarity: closely similar substrates show a greater similarity in rate constants than less similar ones. For a given data matrix, in the present case of dimension 14×26 (14 solvent-temperature combinations and 26 substrates), the method of estimation of the parameters *A*, α , β , and θ which best approximate the data matrix is called principal components analysis.¹⁴ Methods of estimating the number of *A* terms, have been developed only for complete matrices¹⁵ while, for a given *A*, the other parameters can be estimated for incomplete data matrices. We have therefore analysed the present data matrix as follows.

(a) *Estimation of the Number of Terms A from a Complete Submatrix.*—Substrates (4)—(7), (11), (12), and (19)—(22) define a complete data matrix for the variables (1)—(4), (7), (8), (11), and (12) (see Table 1). A cross-validation analysis¹⁴ shows that a two-component model [equation (1) with *A* = 2] adequately describes this matrix. Hence, we have used *A* = 2 in the following analysis of the incomplete data matrices. This corresponds to the model of two phenomenological factors varying between the 26 different solvolyses.

(b) *Analysis of Incomplete Matrices.*—The data were then analysed in two steps. In the first step the data for primary substrates [(4), (5), (7), (10), (16), (17)] were analysed using a one-component model. This gave estimates of the missing observations for the primary substrates. Thereafter the secondary substrates [(6), (8), (9), (11)—(15), (18)] were separately analysed in the same way using the missing data estimation previously described.^{13,15} This method is equivalent to the DOVE method of Swain *et al.*¹⁶ The missing observations of the included substrates were filled in and then in a final step the whole data matrix was analysed using the same method. The analysis thus gives for each substrate *k* the two-component values θ_{1k} and θ_{2k} . These express the position of the solvolysis of the substrate in relation to the other substrates. Hence, the plot of θ_{1k} against θ_{2k} shown in Figure 1 for all substrates displays similarities and dissimilarities between the corresponding solvolyses.

Interpretation of Figure 1.—Figure 1 reveals a most interesting pattern. In the bottom of the Figure the primary compounds [(1), (2), (4), (5), (7), (10), (16), (17)] are arranged in a linear structure. That confirms other evidence that phenylethyl tosylate (17) behaves like an ordinary primary compound.¹⁷ In the middle of the Figure another linear structure containing the secondary substrates [(3), (6), (8), (9), (12), (14)] parallels the primary compounds. The two *endo*-2-norbornyl compounds (21) and (26) and the secondary compound (18) very nicely fall into the linear structure of the ordinary secondary substrates. Allyl and benzyl are shifted up above the 'primary region'. It must be pointed out, however, that of the 14 variables for allyl tosylate 8 are missing. That means that the place of allyl tosylate in the Figure is somewhat uncertain. Also

the locations of the cyclo-octyl and cyclopentyl points are somewhat imprecise because of missing values for the extreme variables 1, 2, 13, and 14.

Compounds (20), (23), and (25) group together above

TABLE 2

Parameters from the principal component analyses of data matrix given in Table 1

Variable i^*	α_i	β_{i1}	β_{i2}
1	6.395	0.206	-0.302
2	5.010	0.188	-0.260
3	6.897	0.174	-0.403
4	5.546	0.174	-0.343
5	5.744	0.141	-0.150
6	4.495	0.140	-0.159
7	5.891	0.101	-0.131
8	4.580	0.108	-0.076
9	5.336	0.393	0.029
10	4.044	0.373	0.049
11	7.673	0.328	-0.167
12	6.165	0.313	-0.122
13	5.604	0.384	0.505
14	4.227	0.393	0.440

* Variables 1 and 2 refer to methanol, variables 3 and 4 to ethanol, and so forth. The order of variables is as in Table 1.

the 'normal' secondary class at the top of the Figure. This cluster is probably due to the similarity in the delocalization of the charge in the transition state. This confirms the non-localized charge in the transition state of *exo*-norbonyl compounds.

DISCUSSION

By selecting multivariate kinetic data that should relate to the degree of charge delocalization in the

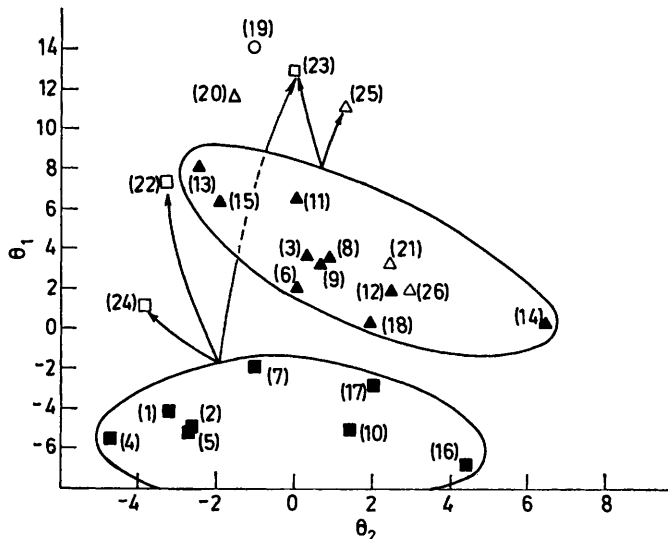


FIGURE 1 θ_1 plotted against θ_2 . Values from Table 3. Filled squares refer to primary compounds and filled triangles to secondary compounds [substrates (1)–(18)]. The empty squares and triangles refer to substrates (19)–(26) except for 1-AdoTs (19) which is marked with a circle

transition state of solvolysis reactions and then subjecting these data to a principal component analysis, direct evidence has been found for differences between *exo*- and *endo*-norbonyl substrates. The deviations between *exo*-norbonyl compounds and the 'normal secondary'

class parallels those between allyl, benzyl, and the 'normal primary' class. This indicates that the difference between *exo*- and *endo*-compounds is largely due to charge delocalization in the transition state for the *exo*-compounds.

The data analysis and the pattern revealed indicates

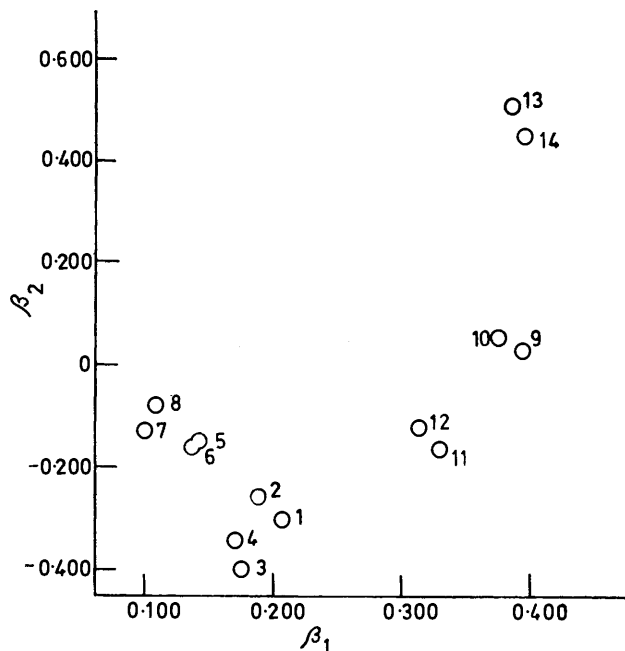


FIGURE 2 β_2 plotted against β_1 . The numbers refer to variables 1–14

the presence of two phenomenological factors affecting the solvolyses.

One, the degree of charge delocalization, is easily seen and interpreted. Solvents of high ionizing power, *i.e.* formic acid and trifluoroacetic acid stabilize a localized

TABLE 3

Values of θ_1 and θ_2

Substrate a^*	θ_1	θ_2
(1)	-4.19	-3.22
(2)	-4.45	-2.62
(3)	3.65	0.24
(4)	-5.63	-4.66
(5)	-5.22	-2.70
(6)	2.07	0.01
(7)	-1.89	-1.09
(8)	3.52	-0.88
(9)	3.16	0.66
(10)	-5.06	1.43
(11)	6.58	0.03
(12)	1.93	2.45
(13)	8.09	-2.50
(14)	0.35	6.44
(15)	6.37	-1.93
(16)	-6.77	4.42
(17)	-2.84	2.02
(18)	0.35	1.90
(19)	14.20	-1.12
(20)	11.70	-1.63
(21)	3.30	2.43
(22)	7.33	-3.33
(23)	13.00	-0.02
(24)	1.10	-3.83
(25)	11.10	1.23
(26)	1.96	2.92

charge better than solvents of lower ionizing power, *i.e.* methanol. Hence, a reaction with localized charge in the transition state should be favoured more by the ionizing solvents than a reaction with delocalized charge. This interpretation is confirmed by the plot of the parameters β_{2i} against β_{1i} for the 14 variables (Figure 2). The plot shows a continuous change over the seven solvents in accordance with their ionizing power.

The second phenomenological factor seems to be related to the size of the alkyl group in the substrate. Possible effects related to this size factor include the inductive effect and the polarizability but not the hyperconjugative effect which is related to charge delocalization.

The clear similarity between, on the one hand, *endo*-norbornyl substrates and the ordinary cyclic secondary substrates such as cyclohexyl and cyclopentyl and, on the other, between *exo*-norbornyl, methylcyclopropyl, and substrates with delocalized charge in the transition state such as benzyl is direct kinetic evidence for the presence of nonclassical charge delocalization in *exo*-norbornyl and methylcyclopropyl solvolyses. This direct evidence relies only on empirical similarities and dissimilarities in model compounds of known behaviour and is therefore, in our view, more straightforward than earlier evidence relying on theories of the detailed behaviour of solvolytic transition states. Though the analytical method is probably unfamiliar to most organic chemists, it is simple and straightforward and based on standard methods of multivariate data analysis widely available at computer centres.

Support from the Swedish Natural Science Research Council is gratefully acknowledged. We are grateful to Professor V. A. Palm for copies of the compilation of rate data.

[9/1106 Received, 16th July, 1979]

REFERENCES

- ¹ S. Winstein and D. S. Trifan, *J. Amer. Chem. Soc.*, **1949**, **71**, 2953.
- ² S. Winstein, *Bull. Soc. chim. France*, **1951**, C55.
- ³ H. C. Brown (with comments by P. v. R. Schleyer), 'The Non-classical Ion Problem,' Plenum Press, New York, **1977**.
- ⁴ J. March, 'Advanced Organic Chemistry: Reactions, Mechanisms and Structure,' McGraw-Hill, New York, **1978**, 2nd edn.
- ⁵ J. B. Hendrichson, D. J. Cram, and G. S. Hammond, 'Organic Chemistry,' McGraw-Hill, New York, **1970**, 3rd edn.
- ⁶ I. D. R. Stevens in 'Organic Reaction Mechanisms,' Wiley, New York, **1973**, pp. 295—338.
- ⁷ D. J. Raber, M. J. Harris, and P. v. R. Schleyer in 'Ions and Ion Pairs in Organic Reactions,' Wiley, New York, **1974**, vol. **2**, pp. 247—374.
- ⁸ T. W. Bentley and P. v. R. Schleyer, *Adv. Phys. Org. Chem.*, **1977**, **14**, 1.
- ⁹ P. D. Barlett, 'Nonclassical Ions (Reprints and Commentary),' Benjamin, New York, **1965**.
- ¹⁰ J. E. Leffler and E. Grunwald, 'Rates and Equilibria of Organic Reactions,' Wiley, New York, **1963**.
- ¹¹ V. A. Palm, 'Grundlagen der Quantitativen Theorie Organischer Reaktionen,' Akademie-Verlag, Berlin, **1971**.
- ¹² S. Wold, *Chemica Scripta*, **1974**, **5**, 97.
- ¹³ S. Wold and M. Sjöström in 'Correlation Analysis in Chemistry: Recent Advances,' eds. J. Shorter and N. B. Chapman, Plenum Press, London, **1978**.
- ¹⁴ S. Wold, *Technometrics*, **1978**, **20**, 379.
- ¹⁵ A. Christoffersson, Thesis, Uppsala University, **1970**.
- ¹⁶ P. F. Strong, C. G. Swain, and M. S. Swain, *J. Chem. Information Computing Sci.*, **1979**, **19**, 13.
- ¹⁷ G. A. Olah and D. Porter, *J. Amer. Chem. Soc.*, **1971**, **93**, 6877.