Bull. Chem. Soc. Jpn., 78, 281–283 (2005)

The Causes of Different Acid Strengths in Water: A Multivariate Statistical Approach

Roberto Aruga

Department of Analytical Chemistry, University of Turin,
via Giuria 5, I 10125 Turin, Italy

Received February 5, 2004; E-mail: roberto.aruga@unito.it

Multivariate thermodynamic data of proton dissociation of 84 aromatic acids in aqueous medium have been processed with factor analysis. It is demonstrated that the concept of latent factor allows quantitative and rigorous determinations of the number and of the importance of the causes of different acid strengths.

The traditional idea, according to which differences in acid strengths are mainly attributable to molecular and electronic properties of acids (e.g., polarization of the N–H or O–H bond, etc.) began to be criticized when, as well as pK_a data, accurate values of enthalpy and entropy of deprotonation in aqueous solution were supplied by direct calorimetric techniques. On the basis of calorimetric results, the importance of environmental factors (i.e., solute–solvent and solvent–solvent interactions) began to be pointed out, although, initially, from a merely qualitative point of view.

Later, as a consequence of these findings, the studies in this field were mainly carried on in the direction of a splitting up of conventional thermodynamic quantities into partial components, each of them being related to molecular or environmental effects. The studies of Hepler and co-workers[1] and those of Nancollas[2] should be cited first in this field.

As a consequence of the number of thermodynamic variables which are available today for acid dissociation in water, the study of the causes of different acid strengths could be made in a more rigorous and quantitative way by the use of the multivariate techniques of chemometrics. These techniques, indeed, have been proposed in order to obtain maximum information from data consisting of many variables.

Taking into account the above considerations, we have processed multivariate data of proton dissociation in water of 84 aromatic acids with factor analysis. The acids belong to the following classes: protonated anilines (symbol: An; 25 species investigated); protonated pyridines (symbol: Py; 20 species); benzoic acids (symbol: B; 19 species); phenols (symbol: Ph; 20 species).

The variables considered in this study are, first, the usual three fundamental quantities of proton dissociation: ΔG°, ΔH°, ΔS°. The values of these parameters were taken from the literature.[3–6] Electrostatic and nonelectrostatic contributions of Gibbs function and enthalpy have also been considered ($\Delta G_{\rm el}$, $\Delta G_{\rm non}$, $\Delta H_{\rm el}$, $\Delta H_{\rm non}$, respectively).[2] The electrostatic contribution represents long-range electrostatic interactions, which are affected by the dielectric nature of the solvent and by the temperature. Therefore, this part is determined by solute–solute electrostatic bonds and by long-range solute–solvent interactions. The nonelectrostatic contribution is determined by all the short-range factors, independent of dielectric constant and temperature. Of consequence, this contribution is mainly determined by solute–solute covalent bonds and by solute–solvent contact interactions (i.e., solvation interactions).[7] Moreover, taking into account that the difference between $\Delta G_{\rm non}$ and $\Delta H_{\rm non}$ is constant, as it is equal to the cratic part,[8] we have considered the following three partial quantities in this study: $\Delta G_{\rm el}$, $\Delta G_{\rm non}$, and $\Delta H_{\rm el}$. They were calculated by means of the equations:[9]

$$\Delta G_{\rm el} = RC\ {\rm e}^{T/\theta} \tag{1}$$

$$\Delta G_{\rm non} = nRT \ln M_{\rm s} + RCa \tag{2}$$

$$\Delta H_{\rm el} = RC(1 - T/\theta){\rm e}^{T/\theta} \tag{3}$$

where R is the gas constant, n is the number of moles of the reagents minus the moles of the products for the reaction considered, T is the temperature, θ is a constant characteristic of the solvent, and C and a are quantities dependent on the dielectric constant of the solvent. Fuller details on the calculations, together with applications and critical evaluations of the above partial thermodynamic quantities are available in the literature.[7–9]

Following the treatment of Hepler,[1] one can also consider the enthalpy of proton dissociation in aqueous solution in terms of "internal" and "external" (or environmental) contributions. The former are related to electronic effects, which act inside the molecule of the acid. The latter are related to interactions between solute particles and the surrounding medium and to other external effects. In any case the external part, owing to its generic and rather undefined characterization, will not be considered in this study. The internal part of the proton dissociation enthalpy has been calculated by using the equations repeatedly reported by Hepler and co-workers.[1,10,11]

The data referring to the seven thermodynamic variables used in this study are reported in Table 1. The different values for the acids have been expressed, on the whole, as the means and standard deviations of the variables for each class. The Hammett σ constants have also been considered, the values of which are available in the literature.[12] The complete set of the data is available on request.

As factor analysis is coming into widespread use, only a brief account will be given, together with references to the literature[13,14] for fuller details. Factor analysis was carried out by a previous evaluation of principal components and their oblique rotation with varimax,[14] in order to find the directions of variance corresponding to significant physical factors (or latent factors). The number of latent factors was determined with the average variance method. A corresponding group of experimental variables is associated with each factor. The variables of each group are correlated (or anticorrelated) with the latent factor representing them, and, of course, they are correlated with each other. The correlation of a certain factor with each experimental variable is expressed by the loading. Therefore the matrix of loadings allows one to determine the group of

Published on the web February 11, 2005; DOI 10.1246/bcsj.78.281

Table 1. Thermodynamic Quantities of Proton Dissociation in Aqueous Medium at 25 °C and Zero Ionic Strength for Various Classes of Acids[a)]

Variable	An	Py	B	Ph
ΔG°	21.6 (6.4)	24.6 (12.8)	22.8 (2.7)	51.1 (6.4)
ΔH°	27.9 (4.8)	18.0 (11.4)	0.5 (3.7)	21.6 (2.5)
ΔS°	21.2 (6.1)	−22.0 (7.4)	−75.0 (12.0)	−99.4 (14.7)
ΔG_{el}	2.6 (1.3)	11.8 (1.6)	23.0 (2.5)	28.1 (3.1)
ΔG_{non}	19.0 (5.2)	12.8 (11.8)	−0.14 (3.0)	23.0 (3.5)
$-\Delta H_{el}$	1.0 (0.5)	4.8 (0.6)	9.3 (1.0)	11.4 (1.2)
$-\Delta H_{int}$	6.3 (8.2)	44.2 (37.6)	1.1 (2.6)	5.5 (6.2)

a) Mean values for each class (standard deviation in parentheses). Gibbs functions and enthalpies in kJ mol^{-1}, entropies in J K^{-1} mol^{-1}. An: protonated anilines; Py: protonated pyridines; B: benzoic acids; Ph: phenols.

Table 2. Loadings of Eight Variables on Three Latent Factors (F1–F3) after Rotation by Varimax, for 84 Aromatic Acids in Aqueous Medium

Variable	F1	F2	F3
ΔG°	0.72	0.68	0.12
ΔH°	−0.35	0.93	0.09
ΔS°	0.99	−0.09	0.04
ΔG_{el}	0.99	−0.09	0.04
ΔG_{non}	0.00	0.99	0.12
ΔH_{el}	−0.99	0.09	−0.04
ΔH_{int}	0.29	0.32	0.62
σ[a)]	0.01	0.12	0.90

a) Hammett σ constant.

experimental variables best correlated with each single factor and, in consequence, to point out the nature of the factor itself. The importance of each factor is expressed in terms of the corresponding fraction of the total variance. The data were always autoscaled[13] before factor analysis.

A factor analysis was then carried out on the matrix referring to 84 acids and 8 variables. It pointed out three significant factors (F), with the following values of variance: F1, 3.7 (46.7% of the total variance); F2, 2.7 (33.5%); F3, 1.0 (12.0%). The presence of two main factors and of a third one, less important, is pointed out in this way. In order to study the nature of the factors, the loadings of the eight variables on F1–F3, after varimax rotation, have been collected in Table 2.

On the basis of these values, one sees that factor F1 appears strictly connected with ΔG_{el} and ΔH_{el}, and also with ΔS°, which is a long-range variable, as it mainly reflects the solvent structure variations.[2,8] F2, on the contrary, is correlated with ΔG_{non} and ΔH°. A long-range, electrostatic and a short-range, nonelectrostatic nature can then be attributed to F1 and F2, respectively. It must be noted that the association of a group of variables with a certain factor is indicated by the absolute values of the corresponding loadings, while positive or negative signs only indicate if this association is based on a positive or negative correlation. For example, the opposite signs of the loadings of ΔS° and ΔH_{el} on F1 are the consequence of the fact that the aqueous solvent structuring process caused by an ionogenic proton dissociation (ΔS° unfavourable) gives rise to an electrostatic bond formation between the solvent molecules (ΔH_{el} favourable). It must also be noted that the different values of quantities such as ΔG_{el} and ΔH_{el} in Eqs. 1 and 3 have no connection with the equal absolute values of the loadings of Table 2, as the latter express correlations.

The third factor (F3), on the other hand, is correlated with variables ΔH_{int} and σ. Taking into consideration the nature of these two variables, we conclude that F3 may represent the causes related to "internal" molecular effects, which depend on different electronic structures between an acid and its conjugate base and on other internal electronic effects. It can also be said that ΔH_{int} proves to be in agreement with the nature previously attributed to it.[1] Considering now that the nonelectrostatic contribution proposed by Nancollas includes all effects the nature of which is not electrostatic or long-range, we argue that the internal contribution proposed by Hepler is likely to represent a part of Nancollas's nonelectrostatic contribution. In other words, the two treatments appear to be complementary, in the sense that Hepler's concept of the internal part of enthalpy, being very specific, allows one to study the corresponding factor that otherwise could remain undefined in the more general concept of nonelectrostatic contribution of Nancollas.

The angles between the factors, after rotation, are the following: 89.5° between F1 and F2 (i.e., F_{el} and F_{non}); 70° between F2 and F3 (i.e., F_{non} and F_{int}). Taking into account that a closeness to orthogonality indicates a lack of correlation between two factors, and vice versa, the conclusion can be drawn that F_{el} and F_{non} represent two mutually independent causes, while, in accordance with what has been said above, a certain degree of interdependence is shown by F_{non} and F_{int}.

In conclusion, factor analysis points out three causes of different acid strength. Among these causes, that of internal and molecular nature, which is connected with F_{int} and is independent of the solvent, is by far the least important (12.0% of the total variance). It is then confirmed, in a quantitative way, that the cause traditionally indicated as the most important (i.e., polarization of the O–H or N–H bond, etc.) is actually of secondary importance. In other words, being obvious that the origin of different acid strengths lies in different molecular structures, the conclusion can be drawn that this primary cause does not act *directly*, but acts mainly *through the solvent*.

The effectiveness shown here by the concept of latent factor, in comparison with the treatments based on conventional thermodynamic quantities[3,4] can be explained by taking into account that conventional thermodynamic quantities (ΔG° and ΔH° above all, see the loadings in Table 2) can be considered as "mixed" quantities, in the sense that they are influenced by more than one factor simultaneously, while latent factors considered by factor analysis are the expression of a "pure" (i.e., single and homogeneous) cause. It must also be pointed out that Nancollas, although he did not use multivariate techniques, proposed partial quantities (ΔG_{el} and ΔG_{non} above all) which are "homogeneous" from a thermodynamic point of view. Of consequence, these quantities can be considered conceptually similar to the latent factors of factor analysis and of the same potential utility in the thermodynamic study of solution equilibria.

References

1 J. W. Larson and L. G. Hepler, "Solute–Solvent Interactions," ed by J. F. Coetzee and C. D. Ritchie, Dekker, New York (1969), Vol. 1.

2 G. H. Nancollas, "Interactions in Electrolyte Solutions," Elsevier, Amsterdam (1966).

3 C. L. Liotta, E. M. Perdue, and H. P. Hopkins, Jr., *J. Am. Chem. Soc.*, **95**, 2439 (1973).

4 C. L. Liotta, E. M. Perdue, and H. P. Hopkins, Jr., *J. Am. Chem. Soc.*, **96**, 7308 (1974).

5 J. J. Christensen, L. D. Hansen, and R. M. Izatt, "Handbook of Proton Ionization Heats and Related Thermodynamic Quantities," Wiley, New York (1976).

6 R. Aruga, *Bull. Soc. Chim. Fr.*, **1984**, I293.

7 G. H. Nancollas, *Q. Rev., Chem. Soc.*, **14**, 402 (1960).

8 R. Aruga, *Can. J. Chem.*, **64**, 780 (1986).

9 S. Murakami and T. Yoshino, *J. Inorg. Nucl. Chem.*, **43**, 2065 (1981).

10 L. G. Hepler and W. F. O'Hara, *J. Phys. Chem.*, **65**, 811 (1961).

11 L. G. Hepler, *J. Am. Chem. Soc.*, **85**, 3089 (1963).

12 C. Hansch and A. Leo, "Substituents Constants for Correlation Analysis in Chemistry and Biology," Wiley, New York (1979).

13 D. L. Massart, B. G. M. Vandeginste, L. M. C. Buydens, S. De Jong, P. J. Lewi, and J. Smeyers-Verbeke, "Handbook of Chemometrics and Qualimetrics," Elsevier, Amsterdam (1997).

14 E. R. Malinowski, "Factor Analysis in Chemistry," Wiley, New York (1991).