SUBSTITUENT INTERACTION EFFECTS
AND
MATHEMATICAL-STATISTICAL DESCRIPTION
OF
RETENTION IN LIQUID CHROMATOGRAPHY

SPANJER, M.C.

wegen Vreneli
à Sophi

VOORWOORD
DANKWOORD

Promoveren lijkt in menig opzicht op de eerstbeklimming van een berg[1],
peinzend begin je onderaan de wand, je lijkt zo nietig als een dwerg.
Je zult dan ook zelden geheel en al in je eentje op de top geraken[2],
vandaar dus mijn dank aan diegenen, die dit van nabij bleven meemaken:

Hooggeleerde De Ligny, beste Leo, met jou als gids op de top gekomen,
kunnen we nu, vermoeid, al terugblikkend, over de gevolgde route bomen.
Als te doen gebruikelijk bij 't openen van een volslagen nieuwe route,
weet je vooraf nooit geheel precies welke richting je op zou moeten.
Is er dan opeens een moment waarop het in de wand niet verder gaat,
dan wijk je uit naar de zijkant, wat ons deed belanden op de topgraat.
Sneeuwstormen, wächten en nevel doen ons nog op de voortop bivakkeren,
toch kunnen verijzingen, gendarmes en de messcherpe kam ons niet deren.
Want niet één van deze hindernissen hield uiteindelijk voor ons stand,
en nu dan, er eenmaal bovenop gekomen, schudden we elkaar dus de hand[3].

De copromotor droeg bij tot forse uitbreiding van mijn rekenfantasie,
die tot dan toe praktisch niet verder ging dan multipele regressie.
Enkel denken aan gecompliceerder gereken maakte me bijkans hoogteziek,
en dan het vatten van die duizelingwekkende mathematische statistiek.
Gelukkig werd ik bijgestaan door een doctor met kennis van die dingen,
mijn dank aan de snel werkende en zeergeleerde Hans van Houwelingen.
Niet te vergeten aangaande het (uit)voeren van rekenwerk en discussie,
zijn, eveneens in wiskundig verband, de heren A.de Meijer en J.Weesie.

Zeergeleerde Hammers, beste Wim, in grote mate was het aan jouw hand,
dat ik na een aanloop[4] destijds aan de voet van die wand ben beland.
Nu vermeld ik je voor je bijdragen aan de wetenschappelijke discussie,
en hulp bij de nog wel eens 'Highly Problemfull Liquide Chromatografie'.
In dat geval wist je er uiteindelijk dan wel zo'n draai aan te geven,
dat de juiste betekenis 'Hoogwaardig Presterend' voor mij ging herleven.

Mejuffrouw Roozen, lieve Annemarie, voor jou 'the ultimate compliment',
en dat voor de afdelingsanaliste zonder het doen van één experiment.
Want moge dat dan noodzakelijk zijn, dat doen van soms vele metingen,
belangrijker zijn uiteraard wel wat verhevener en hoogstaander dingen.
In eerste instantie zorgde je voor het relevantste extractieprodukt[5],
een waardig koffiekoker gelijk, ben je naar een hogere rang opgerukt[6].
We klommen de Cima d'Innominata (EX), gewaardeerd met uiterst extreem[7],
hoewel gewend aan de Torre d'Avorio (D sup.), voor jou geen probleem.
Het hoger niveau van het op die top door ons waargenomen vergezicht,
is zoveel meerder dan het voor de feitelijke inhoud vereiste inzicht.
Bij winnen van letterlijke hoogtemeters bibberde je,en niet van de kou,
want stijgen boven 4000-meter-peil gebeurde met mevrouw aan mijn touw.

Hulp van buiten de afdeling verplaatst mij mede naar de glasblazerij,
want ook werkplaats, electronica en bibliotheek horen thuis in die rij.
Van het personeel in de chemicaliën-, glas- en instrumentenmagazijnen,
vermeld ik hier het ondoorgrondelijke stel Jan Ebskamp en Rob Heinen.
Want die twee hebben me door hun eigengereide stijl van klaverjassen,
een toch wel fiks aantal malen, zij het figuurlijk, de oren gewassen.

Bianco Mollema verrichtte een scheikundige prestatie van topformaat,
ze verdeelde mijn grijsgeschreven wirwar in wat nu zwart op wit staat.
En dat in een, vergeleken met mijn aanslagenniveau, flitsend typetempo,
immer er naar strevend dat de datum niet nog eens in 1986 viel, of zo.

De heren E.Berhitu en J.den Boesterd betrachtten geduld in vele uren,
om mijn tekeningen om te toveren tot de elders gepresenteerde figuren.
Een sleutelpositie werd zeker ingenomen door de dames van de kantine,
van levensbelang was 't vullen van de door mij genuttigde soepterrinen.

Buiten 't ACL stonden familie, vrienden en bekenden op de achtergrond,
verdroegen de wisselende stemmingen en kalmeerden als ik me opwond.
Mijn eettafelgenoten hadden het jarenlang vaak 't zwaarst te verduren,
ondanks mijn soms volmondig gesputter, staan ze bij tot d'laatste uren.

In het bijzonder denk ik aan de wil tot kosten en moeiten investeren,
van mijn ouders, omwille van zoonlief die dan zo nodig moest studeren.
'n Speciale plaats wordt altijd ingenomen door mijn beide lieve zussen,
tezamen met zwagers, neefje en nichtje, wel 'n stel hele dikke plussen.

Al die steun echter voor een tekst, waarvoor geldt, ieder zegt het na,
zeg, is dàt nu onze dank:'Machen Sie es mir nicht so kompliziert,ja'![8]
Daarom poog ik in deze bewoordingen en enkele regels, àlles te vatten,
om hiermee het zo immense belang tòch naar juiste waarde te schatten.

Biwak[9] am Dom, 3.49 m, eerste, laatste en definitieve versie:14-3-1984.
M.C. Skyhook.

Ten bate van hen die onvoldoende algemene ontwikkeling hebben genoten,
volgen hieronder tenslotte nog enige,wellicht verhelderende,voetnoten:

1. H.Harrer, 'Die Weisse Spinne', Ullstein Verlag, Wien, 1961, p.75-120.
2. R.Messner, 'Alleingang Nanga Parbat', Müller Verlag, Rüschlikon,1979.
3. H.Maeder, 'Lockende Berge', Silva Verlag, Zürich, 1971, p.24.
4. W.E.Hammers, M.C.Spanjer and C.L.deLigny, Journal of Chromatography, Uitgeverij Elsevier, Amsterdam, aflevering 174, 1979, p.291-305.
5. A.Franquin, 'Die Reuze Flater', Guust album nummer 10, Uitgeverij Dupuis, Sittard, 1977, p.19.
6. A.Faasse,'Zee en Eiland',Drukkerij De Waal,Middelharnis,1965, p.55.
7. M.Vaucher,'Les Alpes Valaisannes',Editions Denoël,Paris,1979, p.33.
8. W.Fux (Wirt von der Bordierhütte, Mischabel-Kette, Walliser Alpen, Schweiz), Persönliche Mitteilung, Sommer 1980.
9. E.Pertl und R.Messner, 'Zurück in die Berge', Verlagsanstalt Athesia, Bozen, 4. Auflage, 1977, p.46.

This dissertation was written under the never failing,
ever stimulating influence of $\mathit{An}_2\mathit{At}_2\mathit{Lt}_2\mathit{JuHrHa}$*,*
after which it was printed by the Domstedelijke Stichting
Pressa Trajectina

SUBSTITUENT INTERACTION EFFECTS AND MATHEMATICAL-STATISTICAL DESCRIPTION OF RETENTION IN LIQUID CHROMATOGRAPHY

substituent interactie effecten en een
mathematisch-statistische beschrijving
van de retentie in de vloeistofchromatografie
(met een samenvatting in het Nederlands)

PROEFSCHRIFT

ter verkrijging van de graad van doctor
in de Wiskunde en Natuurwetenschappen aan de
Rijksuniversiteit te Utrecht, op gezag van
de Rector Magnificus Prof. Dr. O.J. de Jong,
volgens besluit van het College van Decanen
in het openbaar te verdedigen op maandag
18 juni 1984 des namiddags te 2.30 uur

door

MARTINUS CASPER SPANJER

geboren op 5 oktober 1956 te Veenendaal

PROMOTOR: PROF. DR. C.L. DE LIGNY
COPROMOTOR: DR. J.C. VAN HOUWELINGEN

CONTENTS PAGE

INTRODUCTION

Substituent interaction effects in reversed-phase HPLC

The influence of a substituent on the properties of a parent molecule is a well-known feature in phase-equilibrium constants, *e.g.* in octanol-water partition coefficients (1). In most cases it has been accounted for by the additivity concept. If a molecule consists of a skeleton ϕ and m(eventually different)substituents i, such a model considers a chemical or physical property χ to be composed of group contributions x :

$$\chi^{\text{molecule}} = x^{\phi} + \sum_{i=1}^{m} x^{i} \qquad (1)$$

In this study we investigated the mutual *interaction* between the substituents i and k in disubstituted benzenes, denoted by $i\phi k$. The existence of this *interaction* effect has first been recognized by Hine (3). Fujita *et al.* studied it in octanol-water partition coefficients and defined:

$$\log P^{i\phi k} = \log P^{i\phi H} + \log P^{k\phi H} - \log P^{H\phi H} + \log P^{i \leftrightarrow k} \qquad (2)$$

where the last term accounts for the mutual interaction between both substituents.

Since partition coefficients are directly related to retention behaviour in partition chromatography, Nieuwdorp *et al.* investigated the presence of a substituent interaction effect in gas- liquid chromatography. Their study (5) proved the existence of such an effect in substituted phenols, anilines and pyridines in helium/apiezon and helium/carbowax systems at 206°C. They described it with an extension of the well-known Hammett and Taft linear substituent free energy relationships (6). Earlier, Bark and Clarke (7) and Cook and Raushel (8) had ignored the interaction effect in their gaschromatographic experiments on alkylphenols and benzene derivatives, respectively. This leads to incorrect predictions of retention volumes, and to an incorrect choice of optimal separation conditions.

The earliest observations on substituent effects in the field of liquid chromatography are those of Green and Marcinkiewicz (9). They measured numerous retardation factors of multisubstituted benzenes in a reversed-phase paper chromatographic system. The values were compared with those calculated on the basis of additivity of the contributions of singular substituents and of the contribution of the benzene nucleus. Similar approaches in this area are still made, *e.g.* by Kuchar *et al.* (10), who studied phenylacetic, cinnamic and α-methylcinnamic acid derivatives. Their work led Chen and Horvath (11) towards the evaluation of substituent contributions for chromatographic retention of catecholamine derivatives in reversed-phase HPLC systems. Riley *et al.* (12) investigated the functional group behaviour in ion-pair reversed-phase HPLC systems, also with an additivity model.

As the presence of a substituent interaction effect has been proved in GLC systems it seems rather odd to neglect it in HPLC experiments. We therefore set out to demonstrate its presence in some commonly used reversed-phase HPLC systems. Retention volumes of monosubstituted benzoic acids, phenols, anilines and pyridines were measured in buffered methanol-water, acetonitrile-water and tetrahydrofuran-water eluents on an octadecylsilylsilica adsorbent. Chapters I and II describe the results of this work.

Linear substituent free energy relationships (LSFER's)

Description of substituent interaction effects on chemical or phase-equilibrium constants or reaction rate constants inevitably leads to consideration of the several possible models. The first proposal was made by Hammett (13), as follows:

$$\log K_X / K_H = \rho\sigma_{m,p} \tag{3}$$

The symbol K denotes an equilibrium or reaction rate constant, the subscripts X and H refer to the substituted and the unsubstituted compound, respectively, ρ is a reaction-dependent parameter, and σ is a variable that depends on the nature of the substituent, and on its position: different σ values are required for *m*- and *p*- substituents.

This equation has contributed tremendously towards systematizing data on chemical reactions, but on the other hand it soon became clear that so simple an equation has only a limited range of application. To widen the scope of applications, modifications of equation (3) were proposed, at first by Taft (14). His equation reads as follows:

$$\log K_X / K_H = \rho_I\sigma_I + \rho_R\sigma_R \tag{4}$$

The subscripts I and R refer to the inductive and the resonance effect of a substituent, respectively. For a particular substituent only one σ_I value is given. This value holds for both *m*- and *p*-substituents in reactions of various types. However, the value of σ_R for a particular substituent depends on the reaction type. In a summarizing paper (15) four types of reactions, and thus four sets of σ_R values, are proposed:

- σ_R^- type reactions, where direct (through, exalted) resonance is possible between an electron-withdrawing substituent and an electron-donating reaction centre,
- σ_R^+ type reactions, where direct resonance is possible between an electron-donating substituent and an electron-withdrawing reaction centre,
- $\sigma_R(BA)$ type reactions, of substituted benzoic acids and related compounds,
- σ_R^o reactions, where direct resonance is considered to be impossible.

However, in a number of cases this equation fails to explain the variance of the data satisfactorily (15). Nieuwdorp *et al.* (16) therefore proposed the equation

$$\log K_X / K_H = \sum_j \rho_j \sigma_j \tag{5}$$

It was established that equation (5) with only three terms, *i.e.*, in the form

$$\log K_X / K_H = \rho_I \sigma_I + \rho_R \sigma_R + \rho_E \sigma_E \tag{6}$$

is able to describe data on σ_R^o , $\sigma_R(BA)$, σ_R^- and σ_R^+ reactions and on reactions of rigid saturated compounds. In equation (6) the last term accounts for the exalted resonance, and σ_R is equivalent with Taft's σ_R^o . The values of the substituent constants σ were estimated by applying an advanced statistical procedure (factor analysis with missing data) to a large body of data, so that the experimental errors in the data are averaged out.

Recently, two other models were thoroughly revised by the original authors. Those are:

- the model of Yukawa and Tsuno (17), reformulated as the *Linear Substituent Free Energy Relationship*:

$$\log K_X / K_H = \rho_I \sigma_I + \rho_\pi^+ \sigma_\pi^+ + \sigma_\pi^- \sigma_\pi^- \qquad (7)$$

where σ_I is the inductive substituent constant, σ_π^+ the resonance substituent constant for electron-donating substituents (σ_π^+ is zero for electron-withdrawing substituents) and σ_π^- is the resonance substituent constant for electron-withdrawing substituents (σ_π^- is zero for electron-donating substituents).

- the model of Swain and Lupton (18):

$$\log K_X / K_H = fF + rR + h \qquad (8)$$

In equation (8) F is a nonresonance or field constant and R a resonance constant. The symbols f, r and h represent the free parameters. Swain and Lupton's original proposal (19) of F and R values turned out to be unreliable. In a reinvestigation they used a larger reference set of series, and a newly developed mathematical procedure to obtain strictly orthogonal F and R values (20).

In Chapter III we compared the abilities of the latter three linear substituent free energy relationships (6,7 and 8) to describe data on substituent effects by regression analysis. As a reference, the

Taft relationship was used, and 211 series of data were selected from literature that cannot be described satisfactorily by this equation. To explore the scope of applicability of the most successful relationship, data were selected on the effects of ortho-, meta- and para substituents on chemical and physical properties, in rigid saturated systems, small π-systems and extended π-systems. The medium ranged from water to apolar organic solvents and even to the gas phase.

Three-mode factor analysis

In all mentioned LSFER's, but equation (6), sets of σ values figure that are obtained from precisely defined reference reaction series. This procedure yields σ values that are used as variables, in regression analysis with equations (3) - (8). Nieuwdorp *et al.* (16), on the other hand, considered reaction type and substituent as equivalent, which results in the factor analysis model:

$$y_{r,s} = \sum_{j=1}^{u} R_{r,j} \cdot S_{s,j} \qquad (9)$$

where $y = \log K_X / K_H$, r and s denote the reaction and substituent, respectively, and R and S are adjustable parameters, dependent on the reaction and substituent respectively. (Equation (9) is an alternative formulation of equation (5)).

However, the simultaneous description of the influence of reaction type and substituent on equilibrium or reaction rate constants is not the only possibility. The simultaneous description of the influence of solvent and substituent is also possible. Similar to equation (9) we then obtain:

$$y_{p,s} = \sum_{j=1}^{u} P_{p,j} \, S_{s,j} \qquad (10)$$

where p denotes the solvent and P_j are the adjustable parameters, characterizing the solvent. An example of this model is the work of Weiner (21), who described data on the ionization of substituted benzoic acids in a variety of solvents with equation (10).

The combined influence of solvent and reaction type has also been described by two-mode factor analysis. Bacon and Maciel (22) used the equation:

$$y_{p,r} = \sum_{j=1}^{u} P_{p,j} \, R_{r,j} \qquad (11)$$

to model data on the solvent dependence of the NMR chemical shift of ^{1}H, ^{13}C and ^{29}Si nuclei in 36 solvents.

Thus, it is a logical step to try to describe simultaneously the influence of solvent, reaction type and substituent on equilibrium constants by means of three-mode factor analysis. This mathematical-statistical model describes data that are classified with respect to three modes by an equation of the form:

$$y_{p,r,s} = \sum_{j=1}^{u} \sum_{k=1}^{v} \sum_{l=1}^{w} c_{j,k,l} \, P_{p,j} \, R_{r,k} \, S_{s,l} \qquad (12)$$

wherein c represents the elements of the three-mode core matrix of scaling constants. Thus far, equation (12) has only been applied in the field of the social sciences and for the case where data exist for each combination of p, r and s (23). After the recent development of a method to estimate $P_{p,j}$, $R_{r,k}$ and $S_{s,l}$ for the case of missing data (24), three-mode factor analysis is now also suited for applications in the field of the natural sciences where missing data are likely to occur due to the difficulty or cost, involved in obtaining complete large matrices of measurements.

In Chapter IV we describe applications of three-mode factor analysis to data on ionization constants for 15 series of substituted compounds in 3 solvents and to data on phase equilibrium constants of 6 series of substituted compounds in 9 two-phase systems.
The prediction of missing values, another important feature of this mathematical-statistical model is also illustrated in this chapter.

In Chapter V we describe retention in normal-phase HPLC on chemically bonded phases, by means of three-mode factor analysis. In this case the three modes consist of solute, adsorbent and eluent. The main interest is the possibility to predict missing data more accurately than can be done by the commonly applied physical model of Snyder (25).

Finally, in an Appendix the mathematical aspects of three-mode factor analysis are elucidated.

References

1. A. Leo, *J.Chem.Soc., Perkin Trans. II,* (1983)825.
2. J. Shorter, 'Correlation Analysis of Organic Reactivity', J. Wiley, New York, 1982.
3. J. Hine, 'Physical Organic Chemistry', Mc Graw-Hill, New York, 1962, 2nd edition, p. 82.
4. T. Fujita, J. Iwasa and C. Hansch, *J.Am.Chem.Soc.,* 86(1964)5175.
5. G.H.E. Nieuwdorp, C.L. de Ligny and N.G. van der Veen, *J.Chromatogr.,* 154(1978)133.
6. G.H.E. Nieuwdorp, C.L. de Ligny and J.C. van Houwelingen, *J.Chem. Soc., Perkin Trans. II,* (1979)537.

7. L.S. Bark and K.F. Clarke, *J.Chromatogr.*, 48(1970)418.
8. L.E. Cook and F.M. Raushel, *J.Chromatogr.*, 65(1972)556.
9. J. Green and S. Marcinkiewicz, *Chromatogr.Rev.*, 5(1963)64.
10. M. Kuchar, V. Rejholec, M. Jelinkova and O. Nemecek, *J.Chromatogr.*, 150(1978)419.
11. B.K. Chen and C. Horvath, *J. Chromatogr.*, 171(1979)15.
12. C.M. Riley, E. Tomlinson and T.M. Jefferies, *J.Chromatogr.*, 185(1979)197.
13. L.P. Hammett and H.L. Pfluger, *J.Am.Chem.Soc.*, 55(1933)4079.
14. R.W. Taft and I.C. Lewis, *J.Am.Chem.Soc.*, 80(1958)2436.
15. S. Ehrenson, R.T.C. Brownlee and R.W. Taft, *Progr.Phys.Org.Chem.*, 9(1958)287.
16. C.L. de Ligny, G.H.E. Nieuwdorp, W.K. Brederode, W.E. Hammers and J.C. van Houwelingen, *Technometrics*, 23(1981)91.
17. M. Sawada, M. Ichihara, Y. Yukawa, T. Nakachi and Y. Tsuno, *Bull. Chem.Soc. Jpn.*, 53(1980)2055.
18. C.G. Swain, S.H. Unger, N.R. Rosenquist and M.S. Swain, *J.Am.Chem. Soc.*, 105(1983)492.
19. C.G. Swain and E.C. Lupton, *J.Am.Chem.Soc.*, 90(1968)4328.
20. P.F. Strong, C.G. Swain and M.S. Swain, *J.Chem.Inf.Comput.Sci.*, 19(1979)13.
21. P.H. Weiner, *J.Am.Chem.Soc.*, 95(1973)5845.
22. M.R. Bacon and G.E. Maciel, *J.Am.Chem.Soc.*, 95(1973)2413.
23. P.M. Kroonenberg and J. de Leeuw, *Psychometrica*, 45(1980)69.
24. H.M. Weesie and J.C. van Houwelingen, GEPCAM User's Manual, Institute of Mathematical Statistics, Utrecht, 1983 (available on request).
25. L.R. Snyder, 'Principles of Adsorption Chromatography', Marcel Dekker, New York, 1968.

Journal of Chromatography, 253 (1982) 23–40
Elsevier Scientific Publishing Company, Amsterdam — Printed in The Netherlands

CHROM. 15,254

SUBSTITUENT INTERACTION EFFECTS IN AROMATIC MOLECULES IN REVERSED-PHASE LIQUID CHROMATOGRAPHY

M. C. SPANJER and C. L. DE LIGNY*
Laboratory for Analytical Chemistry, State University, Croesestraat 77A, 3522 AD Utrecht (The Netherlands)

(Received June 21st, 1982)

SUMMARY

Retention volumes of monosubstituted benzenes, chlorobenzenes, benzoic acids, phenols, anilines and pyridines have been measured in the reversed-phase mode of a high-performance liquid chromatographic system. Buffered methanol–water and acetonitrile–water eluents were used with an octadecylsilylsilica adsorbent. From the net retention volumes a substituent interaction effect was derived, and described with the Taft equation. The resultant values of the ϱ-parameters were compared with values holding for gas–liquid chromatography and for normal-phase liquid chromatography, derived in previous investigations, and with values holding for batch partition coefficients. They were interpreted in terms of hydrogen bonding between the solutes and the eluent.

Batch partition experiments with monosubstituted benzenes, phenols and pyridines were carried out with an *n*-hexadecane/methanol–water system, in order to obtain more information concerning the retention mechanism. A general discussion is given on retention mechanisms in reversed-phase chromatography.

INTRODUCTION

The influence of substituents on the properties of the parent molecule has since long been recognized. An attempt to quantify these effects has been made by Hammett[1] who proposed a set of substituent parameters based on the dissociation constants of benzoic acids. These σ-parameters turned out to be useful for interpreting the influence of substituents on other chemical equilibrium constants and even on reaction rate constants.

Further developments in describing the influence of substituents were made in the field of adsorption and partition. Examples can be found in the well known compilation of Leo *et al.*[2] of partition data in octanol–water systems. Most of the literature[3–5] describes only the individual substituent effect of a single group. Mutual interactions in multisubstituted molecules were mostly neglected. However, simple addition of individual group contributions is incorrect for an accurate prediction of retention. Nieuwdorp *et al.*[6] demonstrated the presence of a substituent interaction effect with gas–liquid chromatography (GLC). These workers investigated substituted phenols, anilines and pyridines and developed an extension of the Hammett and Taft[7] equations for a better description of those systems.

In view of the explosive growth of liquid chromatography an analogous approach in this field is a logical continuation. Hammett-like formula have been tested in normal-phase systems by Snyder[8] for silica and alumina as adsorbents and Hammers and co-workers[9–11] have investigated non-polar and polar bonded phases in normal-phase adsorption chromatography. Nowadays reversed-phase liquid chromatography is a popular method[12] and we have therefore set out to investigate substituent interaction effects in this chromatographic system, and to describe them with the equations developed by Hammett, Taft or Nieuwdorp. In the course of this work we became interested in the retention mechanism and so performed batch partition experiments, and compared the results with those from corresponding high-performance liquid chromatographic (HPLC) systems.

THEORETICAL

The series of solutes used in this study can be described schematically as $i\phi k$, a general formula for disubstituted benzenes. The variable substituent is denoted by i, while within a certain series k is the fixed polar group. The substituents i were fluoro, chloro, bromo, iodo, methyl, nitro, cyano, methoxy, acetyl and carboxymethyl ester. Phenols, anilines, pyridines and benzoic acids formed series of test solutes.

In this article $i\phi$OH, $i\phi NH_2$, $i\phi$N and $i\phi$COOH will be used as notations for the investigated series. (For the sake of simplicity pyridines are also denoted by $i\phi k$ and $i\phi N$, although the nitrogen atom is a part of the ring.)

For chemical equilibrium constants K the Hammett equation has the following form:

$$\log K^{i\phi k} - \log K^{H\phi k} = \varrho_k \, \sigma_i \tag{1}$$

The reaction constant ϱ_k reflects the sensitivity of the fixed group k towards electronic shifts in the benzene nucleus caused by substituent i. The substituent parameter σ_i indicates the influence of a varying substituent i on the fixed functional group. Hammett developed a σ_i scale by defining $\varrho_k \equiv 1$ for dissociation constants of benzoic acids in water at 25°C. This set of σ values can be used to determine ϱ values for other systems.

The σ values depend on the position of the substituent, which results in σ_m values for *meta* and σ_p values for *para* substituents. The reaction constant ϱ has the same value for *meta*- and *para*-substituted solutes. To increase the somewhat limited applicability of the Hammett equation, an extension was formulated by Taft and Lewis[7] which reads as follows:

$$\log K^{i\phi k} - \log K^{H\phi k} = \varrho_I \, \sigma_I + \varrho_R \cdot \sigma_R \tag{2}$$

Taft and Lewis use σ_I and σ_R parameters, and so distinguish an inductive (I) and a resonance (R) effect for each substituent, which are equal for *meta* and *para* positions. *Meta*- and *para*-substituted solutes must be considered separately in this model, which results in different ϱ_I and ϱ_R values for *meta* and *para* series.

In *ortho*-substituted compounds steric and other short-range intramolecular interactions interfere with a description of substituent effects in terms of σ parameters only[8]. For this reason we have not carried out measurements on *ortho*-substituted solutes.

Because in this work retention volumes are measured we have to convert from chemical equilibrium constants to retention volumes. The basic retention formula for a solute in liquid chromatography[13] is:

$$V_N = V_R - V_M = K\Phi_s \qquad (3)$$

where V_N is the net retention volume, V_M is the mobile phase volume and the instrumental dead volume, Φ_s is either the volume of the stationary phase in partition chromatography or the adsorbent surface area in adsorption chromatography, and K is the distribution coefficient of an eluted compound.

An important detail of this equation is the fact that the symbol K is not restricted to either an adsorption or a partition process taking place in the column. This is important because there exists a lack of full understanding of the retention mechanism in reversed-phase HPLC[14–18]. For the members of a series with a fixed centre k we can write

$$\log K^{i\phi k} = \log V_N^{i\phi k} - \log \Phi_s \qquad (4)$$

and

$$\log K^{H\phi k} = \log V_N^{H\phi k} - \log \Phi_s \qquad (5)$$

from which it follows that

$$\log \frac{K^{i\phi k}}{K^{H\phi k}} = \log \frac{V_N^{i\phi k}}{V_N^{H\phi k}} \qquad (6)$$

Before eqns. 1 (or 2) and 6 can be combined, a correction must be made for the contribution of group i to $\log K$. This contribution, the "primary substituent effect", can be calculated from data on substituted benzenes:

$$\log \frac{K^{i\phi H}}{K^{H\phi H}} = \log \frac{V_N^{i\phi H}}{V_N^{H\phi H}} \qquad (7)$$

Combination of eqns. 2, 6 and 7 results in the final formula:

$$\log \frac{V_N^{i\phi k}}{V_N^{H\phi k}} - \log \frac{V_N^{i\phi H}}{V_N^{H\phi H}} = \varrho_I\, \sigma_I + \varrho_R \cdot \sigma_R \qquad (8)$$

In this equation the left hand side represents the substituent interaction effect. Eqn. 8 will further be abbreviated to

$$\Delta \log V_N^{i\phi k} - \Delta \log V_N^{i\phi H} = \Delta\Delta \log V_N^{i\phi k} = \varrho_I\, \sigma_I + \varrho_R \cdot \sigma_R \qquad (9)$$

In the batch partition experiments the two phases consisted of *n*-hexadecane and the eluent used in the HPLC experiments, mutually saturated. Partition coefficients, P, can be calculated from the UV absorbances of the eluent by the formula[19]

$$P = \frac{C_{\text{hexadecane}}}{C_{\text{eluent}}} = \frac{A_{\text{before equilibration}} - A_{\text{after equilibration}}}{A_{\text{after equilibration}}} \cdot \frac{V_{\text{eluent}}}{V_{\text{hexadecane}}} \tag{10}$$

where C represents the concentration of the solute, A its absorbance in the eluent and V the volume of the equilibrated quantities of both phases.

The Taft equation for partition coefficients is analogous to eqn. 8:

$$\log \frac{P^{i\phi k}}{P^{H\phi k}} - \log \frac{P^{i\phi H}}{P^{H\phi H}} = \varrho_I \sigma_I + \varrho_R \cdot \sigma_R \tag{11}$$

In further dealing with this equation it will be abbreviated in the same way as eqn. 8.

EXPERIMENTAL

Chemicals

The solutes were from Fluka (Buchs, Switzerland) (grade purissimus), Baker (Deventer, The Netherlands) (analysed-reagent grade), Aldrich Europe (Beerse, Belgium) (laboratory-use quality) and ICN Pharmaceuticals (Plainview, NY, U.S.A.) (qualified by "for investigational use").

The eluent components methanol, acetonitrile and phosphoric acid were from Baker (analysed-reagent grade). Disodium hydrogen phosphate and sodium dihydrogen phosphate were from E. Merck (Darmstadt, G.F.R.) (pro analysi). Water was distilled twice from an alkaline potassium permangate solution.

LiChrosorb 10 RP-18 (Merck) was used as adsorbent. It has a reported[20] carbon content of 0.286 g per g bare LiChrosorb Si-100. This corresponds with a surface concentration of 4.4 μmole octadecyl groups per square meter.

The *n*-hexadecane was a product of Aldrich Europe, with a purity of 99%.

Apparatus

The liquid chromatograph was a Packard-Becker Model 8200 (Delft, The Netherlands), equipped with a UV detector with a fixed wavelength of 254 nm. The eluent flow-rate was continuously monitored with a siphon counter (Waters Assoc., Milford, MA, U.S.A.) which had been calibrated with a type ABU 12 autoburette (Radiometer, Copenhagen, Denmark). The column and eluent vessel were kept at (25.0 ± 0.1)°C with a Haake circulating-water thermostat Model E52 (Karlsruhe, Germany). The column (precision-bore stainless steel, 25 cm × 2.1 mm I.D.) was packed by forcing an ultrasonically degassed and homogenized slurry of the adsorbent in carbon tetrachloride (*ca.* 10%, w/v) into it with *n*-hexane at 350 atm. The packing was settled by flushing 300 ml of *n*-hexane and 300 ml of methanol through the column[21]. The column was weighed when filled with *n*-hexane and with methanol, respectively, to obtain the void volume of the column[22]. This volume, measured twice, was (622 ± 10) μl. A similar measurement with *n*-hexane and tetrachloromethane gave a slightly higher value. The weight of the packing material was 0.63 g. Dead volumes of connecting capillaries were measured by coupling column inlet and outlet directly together with a volumeless device and injecting a solute. Samples of 3 μl were injected on-stream with a high-pressure resistant syringe (SGE, Melbourne, Australia).

Procedure

Chromatographic experiments. Eluent modifier concentrations were chosen to obtain capacity ratios (k') between 1 and 10. The eluent flow was *ca.* 1 ml min^{-1}, which required pressure drops of 100 and 150 atm with acetonitrile–water and methanol–water, respectively. Triplicate measurements were made with reproducibilities of *ca.* 20 μl or 4 % for the most strongly retained solutes.

To take into account any possible slight differences in eluent compositions, the retention of the nitro-containing compounds in each series of solutes was measured daily, and if necessary, minor corrections were made. These measurements served also as a control for column performance.

Special attention had to be paid to measurements in eluents with pH 2, since literature reports different opinions on adsorbent stability at very low pH values[23,24]. When measurements at pH 2 had been made, the column was filled with eluent of pH 3.5 at the end of the day, so as to avoid adsorbent deterioration overnight. A buffer concentration of 25 mM sufficed because of the very small sample concentration. This factor also enlarged column lifetime[24]. In the course of this study no sign of column deterioration has been observed.

Suppression of ionization of all compounds was accomplished by choosing appropriate pH values.

The pH of the aqueous phosphate buffer was adjusted while using a Methrohm E 350 B pH meter (Herisau, Switzerland). Methanol or acetonitrile was then added. Upon addition of an organic solvent to water, the pK of the phosphoric acid in the buffer solution increases, and hence its pH. In 50 % (v/v) methanol we estimate the increase at *ca.* 0.5 unit. However, the pK values of the acidic solutes increase by about the same amount, whereas the pK values of the basic solutes decrease by about 0.5 unit[25]. This implies that all solutes are in the uncharged form when the pH before addition of the organic solvent is adjusted to 2.0 for benzoic acids, 3.5 for phenols and 7.1 for anilines and pyridines[26]. Special care was taken in degassing the eluents. In contrast with recent reports[27] we achieved excellent performance by ultrasonically degassing after gently heating the mixtures to 40°C.

Batch experiments. Before performing the actual partition experiments both the eluent and the *n*-hexadecane were saturated with each other. Both phases were equilibrated during a week by permanent shaking, the eluent being refreshed four times during this period. Constancy of composition was checked by gas chromatographic analysis. Afterwards the layers were roughly separated by a separatory funnel and thoroughly centrifuged. Solutes were dissolved in the eluent up to concentrations giving absorbances between 0.4 and 0.6. Since phenols and pyridines have log ε values of 3 and higher[28] solute concentrations were low enough to be sure of the absence of dimerization in the *n*-hexadecane layer[29,30]. Aliquots (each 1.5 ml) of the sample solutions were equilibrated with 7.5 ml of *n*-hexadecane for 30 min by means of a testtube rotator. To ensure the absence of emulsions the tubes were centrifuged at 1300 g for 30 min. After these manipulations a constant blank was measured spectrophotometrically.

Before drawing a sample of the aqueous bottom layer, the upper layer and the upper part of the bottom layer were carefully removed by means of a water suction pump. Absorbances of 1-ml aliquots of the eluent phase only (not only the organic phase[31], nor both phases[32]), were measured spectrophotometrically at 254 nm in the quartz cell of a Vitatron MPS type 940-320 spectrophotometer (Dieren, The Netherlands). Solutes were used as received because the prior HPLC experiments had not shown any spurs due to impurities.

RESULTS

The results of the measurements in methanol–water are shown in Tables I and II, while those in acetonitrile–water are in Table III. The precision of the log V_N values can be estimated by comparing data for the series of substituted benzenes at pH 2.0, 3.5 and 7.1 and of substituted chlorobenzenes at pH 3.5 and 7.1. Doing so (for substituents 0–10) we find a standard deviation of 0.007 in log V_N, corresponding with an experimental uncertainty of 1.6% in the means of the triplicate V_N determinations. This precision can be expected to hold for all measurements, except those on bromo- and iodobenzene. The mean V_N values of these solutes have relative standard deviations of *ca.* 3% due to tailing. Before testing the two-parameter eqn. 8, graphs of $\Delta\Delta$ log V_N *versus* $-\Delta$pK (where K is the ionization constant of the disubstituted

TABLE I

LOG V_N VALUES FOR MONOSUBSTITUTED BENZENES AND MONOSUBSTITUTED CHLOROBENZENES AT SEVERAL pH VALUES IN METHANOL–WATER (50:50, v/v)

Substituent, i	*Substituent No.*	*Log V_N*						
		Series (pH)						
		iϕH (2.0)	*iϕH (3.5)*	*iϕH (7.1)*	*iϕCl (3.5)*		*iϕCl 7.1*	
					m	*p*	*m*	*p*
H	0	0.64	0.64	0.64	–	–	–	–
F	1	0.62	0.71	0.73	1.06	0.98	1.06	0.98
Cl	2	0.96	0.99	1.00	1.41	1.33	1.41	1.34
Br	3	1.11	1.14	1.10	1.51	1.43	1.53	1.45
I	4	1.27	1.26	1.30	–	–	–	–
CH_3	5	0.97	1.02	1.01	1.38	1.38	1.38	1.37
NO_2	6	0.40	0.50	0.49	0.85	0.74	0.86	0.75
CN	7	0.23	0.25	0.23	0.58	0.52	0.59	0.52
OCH_3	8	0.59	0.63	0.68	1.11	1.07	1.13	1.08
$COCH_3$	9	0.30	0.34	0.34	0.71	0.71	0.73	0.72
$COOCH_3$	10	–	0.59	0.60	–	–	–	–
OH	11	−0.07	0.02	0.01	0.53	0.50	–	–
NH_2	12	0.12	0.17	−0.12	–	–	0.36	0.32
=N–	13	–	–	−0.07	–	–	–	–
COOH	14	0.20	–	–	–	–	–	–

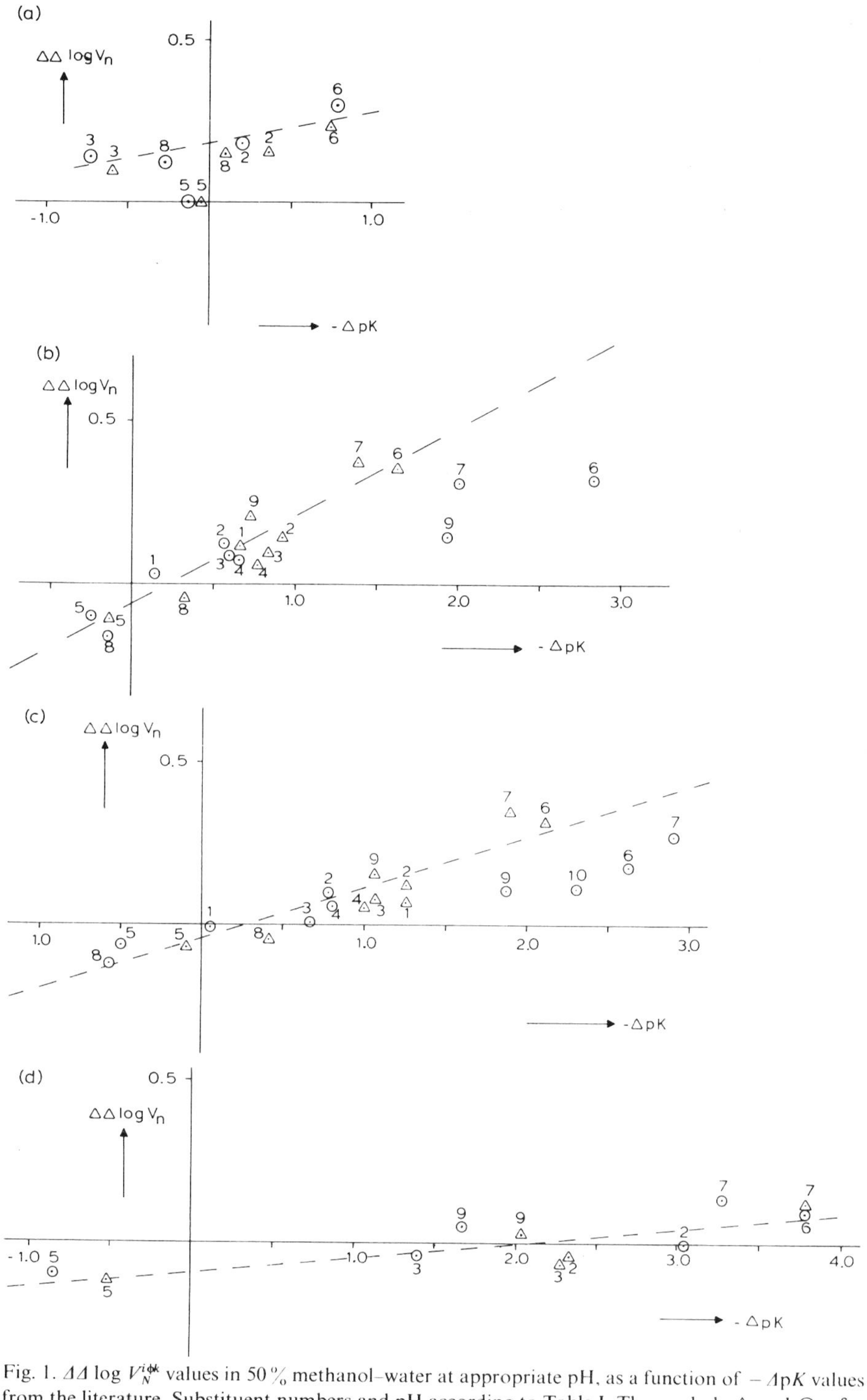


Fig. 1. ΔΔ log $V_N^{i\phi k}$ values in 50% methanol–water at appropriate pH, as a function of −ΔpK values taken from the literature. Substituent numbers and pH according to Table I. The symbols △ and ⊙ refer to the *meta* and *para* positions, respectively. (a), k = COOH, pK from refs. 28 and 33; (b), k = OH, pK from refs. 33–40; (c), k = NH$_2$, pK from refs. 34 and 35; (d), k = =N–, pK from refs. 35 and 41.

TABLE II

LOG V_N VALUES IN METHANOL–WATER (50:50, v/v)

i	*Log* V_N			
	Series (pH)			
	$i\phi COOH$ (2.0)	$i\phi OH$ (3.5)	$i\phi NH_2$ (7.1)	$i\phi N$ (7.1)
m-F	0.34	0.21	0.04	–
p-F	0.32	0.12	−0.03	–
m-Cl	0.68	0.53	0.36	0.26
p-Cl	0.70	0.50	0.32	0.28
m-Br	0.77	0.62	0.43	0.34
p-Br	0.80	0.61	0.42	0.36
m-I	0.86	0.72	0.59	0.41
p-I	0.98	0.72	0.60	–
m-CH_3	0.51	0.29	0.18	0.19
p-CH_3	0.50	0.30	0.19	0.20
m-NO_2	0.19	0.25	0.05	–
p-NO_2	0.26	0.20	−0.10	−0.13
m-CN	0.05	0.02	−0.18	−0.36
p-CN	0.01	−0.05	−0.25	−0.34
m-OCH_3	0.30	−0.03	−0.13	–
p-OCH_3	0.27	−0.15	−0.21	–
m-$COCH_3$	−0.04	−0.06	−0.25	−0.34
p-$COCH_3$	0.01	−0.13	−0.31	−0.31
m-$COOCH_3$	–	0.15	–	–
p-$COOCH_3$	–	0.16	−0.05	–

benzene) were drawn to visualize the presence of a substituent interaction effect. Since acid–base equilibration can be described with σ values, a correlation of $\Delta\Delta$ log V_N with ΔpK is an indication for a correlation of $\Delta\Delta$ log V_N with σ constants. These graphs are shown in Fig. 1 for all solute series and confirm the existence of a substituent interaction effect that is related to σ constants. Any missing point in these figures is due to a lack of available pK values. Deviations from straight lines in these plots can be regarded as limitations of the two-dimensional presentation. An effect that must be described with a two-parameter equation should be graphically illustrated with a three-dimensional figure. Testing of the two-parameter equation by regression analysis seemed to be justified by these plots.

The $\Delta\Delta$ log V_N values for all series were correlated with the substituent constants σ_I and σ_R given by Nieuwdorp *et al.*[42]. (Nieuwdorp's values of σ_I and σ_R have been estimated by an advanced statistical procedure, taking into account the whole body of relevant data). This treatment generates regression coefficients identical to the ϱ_I and ϱ_R parameters in eqn. 8. The results are given in Table IV for both methanol–water and acetonitrile–water eluents. The s_y values in this table correspond with the standard deviation of an individual $\Delta\Delta$ log V_N value from the calculated value. An extraordinary group of substances consists of solutes with both *i* and *k* being OH, NH_2, =N– or COOH. The log V_N values of these compounds are collected in Table V. The results of the batch partition experiments are given in Table VI and

TABLE III

LOG V_N VALUES IN ACETONITRILE–WATER (30:70, v/v)

i	*Log V_N*				
	Series (pH)				
	iφH (2.0)	*iφH* (3.5)		*iφCOOH* (2.0)	*iφOH* (3.5)
H	0.77	0.77		–	–
F	0.88	0.88	*m*	0.41	0.35
			p	0.43	0.23
Cl	1.22	1.20	*m*	0.75	0.57
			p	0.78	0.55
Br	1.32	1.29	*m*	0.85	0.70
			p	0.88	0.68
I	1.49	1.47	*m*	1.01	0.84
			p	1.03	0.83
CH_3	1.19	1.17	*m*	0.59	0.39
			p	0.59	0.40
NO_2	0.73	0.71	*m*	0.31	0.33
			p	0.37	0.28
CN	0.50	0.52	*m*	0.06	0.18
			p	0.11	0.10
OCH_3	0.84	0.82	*m*	0.32	0.15
			p	0.33	0.05
$COCH_3$	0.49	0.46	*m*	−0.03	−0.01
			p	−0.01	−0.11
$COOCH_3$		0.78	*m*	–	0.24
			p	–	0.20
OH	–	0.06			
COOH	0.17	–			

the calculated ϱ_I, ϱ_R and s_y values are collected in Table IV. The precision of the log *P* values can, at first instance, be estimated by comparing data for the series of substituted benzenes at pH 3.5 and 7.1, respectively. Doing so we find a standard deviation of 0.024 in log *P*, corresponding with an experimental uncertainty of 5.5% in the means of the duplicate *P* determinations.

DISCUSSIONS

The primary substituent effect

Before exploring the substituent interaction effect, which is a secondary substituent effect, we must justify the correction for the primary substituent effect in eqn. 8. Therefore we examined following equations:

$$\log V_N^{i\phi H} - \log V_N^{H\phi H} = \varrho_I\,\sigma_I + \varrho_R\cdot\sigma_R \tag{12}$$

$$\log V_N^{i\phi OH} - \log V_N^{H\phi OH} = \varrho_I\,\sigma_I + \varrho_R\cdot\sigma_R \tag{13}$$

$$\log V_N^{i\phi OH} - \log V_N^{H\phi OH} = \varrho_I\,\sigma_I + \varrho_R\cdot\sigma_R + q(\log V_N^{i\phi H} - \log V_N^{H\phi H}) \tag{14}$$

TABLE IV

CALCULATED ϱ_I AND ϱ_R PARAMETERS FOR $\Delta\Delta$ LOG V_N DATA IN METHANOL–WATER (50:50, v/v) (A) AND ACETONITRILE–WATER (30:70, v/v) (B), AND FOR $\Delta\Delta$ LOG P DATA IN n-HEXADECANE/METHANOL–WATER (50:50, v/v) (C)

	i	k		ϱ_I	ϱ_R	s_y
A	1–9	COOH	*m*	0.34 ± 0.04	0.03 ± 0.08	0.05
	1–9		*p*	0.41 ± 0.02	0.08 ± 0.04	0.03
	1–10	OH	*m*	0.49 ± 0.03	0.42 ± 0.05	0.04
	1–10		*p*	0.40 ± 0.03	0.50 ± 0.05	0.03
	1–9	NH_2	*m*	0.42 ± 0.03	0.41 ± 0.05	0.03
	1–10		*p*	0.28 ± 0.03	0.37 ± 0.06	0.04
	2–5,7,9	=N–	3	0.07 ± 0.06	0.38 ± 0.15	0.06
	2,3,5,6,7,9		4	0.10 ± 0.04	0.34 ± 0.11	0.04
B	1–9	COOH	*m*	0.28 ± 0.01	0.01 ± 0.01	0.01
	1–9		*p*	0.36 ± 0.01	0.03 ± 0.01	0.01
	1–10	OH	*m*	0.46 ± 0.04	0.33 ± 0.07	0.05
	1–10		*p*	0.35 ± 0.02	0.37 ± 0.03	0.02
C	1–10	OH	*m*	1.99 ± 0.03	0.59 ± 0.06	0.10
	1–10		*p*	1.65 ± 0.04	0.70 ± 0.07	0.08
	2,3,5–8	=N–,	3	0.20 ± 0.09	1.00 ± 0.08	0.07
	2,3,5–8		4	0.26 ± 0.08	0.89 ± 0.07	0.08

Calculated values of the regression coefficients ϱ_I, ϱ_R and q are shown in Table VII. If we survey the column of s_y values in this table we see that the primary substituent effect can by no means be described by a linear combination of σ_I and σ_R constants, neither with substituted benzenes nor with substituted phenols. On the other hand, with eqn. 14 we obtained a standard error of fit which is similar to the experimental

TABLE V

LOG V_N VALUES FOR $i\phi k$ COMPOUNDS IN WHICH BOTH i AND k ARE POLAR SUBSTITUENTS

In the case of the pyridines, 3 and 4 have to be read instead of *m* and *p*, respectively.

i	*Log V_N*					
	Eluent					
	Methanol–water (50:50, v/v)				*Acetonitrile–water (30:70, v/v)*	
	k(pH)					
	COOH (2.0)	*OH (3.5)*	*NH_2 (7.1)*	*=N– (7.1)*	*COOH (2.0)*	*OH (3.5)*
m-OH	−0.30	−0.59	−0.66	−0.50	−0.28	−0.46
p-OH	−0.39	−0.80	−0.77	−0.94	−0.42	−0.62
m-NH_2	−0.54	−0.48	−0.74	−0.49	−0.72	−0.53
p-NH_2	−0.63	−0.53	−0.90	−0.55	−0.63	−0.71
m-COOH	−0.48	–	–	–	−0.68	–
p-COOH	−0.65	–	–	–	−0.74	–

TABLE VI

LOG *P* VALUES IN *n*-HEXADECANE/METHANOL–WATER (50:50, v/v)

i	*Log P*					
	Series (pH)					
	iϕH *(3.5)*	*iϕH* *(7.1)*		*iϕOH* *(3.5)*		*iϕN* *(7.1)*
H	0.02	0.02		–		–
F	0.04	0.05	*m*	−0.39	3	–
			p	−0.60	4	–
Cl	0.37	0.39	*m*	−0.02	3	−1.10
			p	−0.17	4	−1.07
Br	0.51	0.47	*m*	0.14	3	−1.00
			p	−0.04	4	−1.01
I	0.62	0.66	*m*	0.20	3	−0.80
			p	0.05	4	–
CH_3	0.34	0.36	*m*	−0.90	3	−1.12
			p	−0.92	4	−1.13
NO_2	−0.44	−0.41	*m*	−0.27	3	–
			p	−0.50	4	−1.43
CN	−0.53	−0.52	*m*	−0.48	3	−1.64
			p	−0.65	4	−1.66
OCH_3	−0.11	−0.09	*m*	−1.02	3	–
			p	−1.17	4	–
$COCH_3$	−0.62	−0.58	*m*	−1.13	3	−1.72
			p	−1.19	4	−1.74
$COOCH_3$	−0.72	−0.69	*m*	−1.20	3	–
			p	−1.28	4	–
OH	−1.15	–				
=N–	–	−1.34				

error (see below, in *Statistics* section). In eqn. 14 the primary substituent effect is accounted for in the last term. Furthermore, the least-squares estimates of q are about unity. This points towards the correctness of the use of eqn. 8 for further calculations on the substituent interaction effect as eqn. 8 is identical with eqn. 14 for $q = 1$.

TABLE VII

CALCULATED ϱ_I, ϱ_R AND q VALUES FOR LOG V_N DATA OF SUBSTITUTED BENZENES AND PHENOLS IN METHANOL–WATER (50:50, v/v) AT pH 3.5 WITH SUBSTITUENTS 1–10

Equation		ϱ_I	ϱ_R	q	s_y
12		−0.01 ± 0.24	−0.87 ± 0.44	–	0.31
13	*m*	0.49 ± 0.22	−0.45 ± 0.40	–	0.28
	p	0.40 ± 0.25	−0.37 ± 0.46	–	0.32
14	*m*	0.49 ± 0.02	0.33 ± 0.04	0.90 ± 0.03	0.02
	p	0.40 ± 0.03	0.53 ± 0.06	1.03 ± 0.04	0.03

Substituent interaction effects

The main goal of this work is to determine whether eqn. 8 is a valuable tool for describing solute retention. A survey of Table IV for the standard deviations s_y of the regressions on chromatographic experiments shows that they are not much larger than the experimental error in $\Delta\Delta \log V_N$ (0.014). Therefore eqn. 8 can be concluded to hold well (see *Statistics* section).

As can be expected, s_y values from calculations on batch partition experiments, shown in the same table, are worse. This is mainly due to the rather large number of manipulations involved in measuring a single P value, each one of which introduces a small error[2]. As already mentioned in the Results section, the experimental error for these experiments is more than three times greater than the corresponding error for chromatographic experiments.

We now turn to the evaluation of the calculated ϱ parameters. Starting with values calculated from chromatography (Table IV), the positive sign of the ϱ values of the acidic carboxylic acids and phenols as well as the basic anilines and pyridines is striking. This contrasts with ϱ values calculated from the GLC measurements mentioned previously[6] (Table VIII). The GLC results have been explained by the occurrence of hydrogen bonding, with the hydroxyl or amino group acting as proton donor and the stationary phase polyethylene glycol (PEG) as proton acceptor. The opposite sign of the ϱ values for the pyridines was explained in terms of a dipole–dipole interaction between the pyridines and the stationary phase (hydrogen bonding cannot occur in this case).

TABLE VIII

CALCULATED ϱ_I AND ϱ_R PARAMETERS FOR DATA GIVEN BY NIEUWDORP *et al.*[6] WITH SUBSTITUENTS 1–8

Stationary phase	k		ϱ_I	ϱ_R	s_y
	OH	*m*	0.43 ± 0.02	0.22 ± 0.03	0.02
		p	0.62 ± 0.02	0.56 ± 0.04	0.02
PEG	NH_2	*m*	0.38 ± 0.01	0.14 ± 0.03	0.02
		p	0.75 ± 0.05	0.80 ± 0.10	0.06
	=N–	3	−0.51 ± 0.03	−0.19 ± 0.08	0.03
		4	−0.40 ± 0.01	0.02 ± 0.03	0.01
	OH	*m*	0.16 ± 0.02	0.04 ± 0.03	0.02
		p	0.27 ± 0.03	0.23 ± 0.05	0.03
Apiezon M	NH_2	*m*	0.12 ± 0.01	0.10 ± 0.01	0.01
		p	0.30 ± 0.03	0.36 ± 0.06	0.03
	=N–	3	−0.16 ± 0.02	−0.10 ± 0.04	0.01
		4	−0.25 ± 0.01	−0.17 ± 0.02	0.01

Another difference is found when our results are compared with those on retention behaviour in normal-phase systems as investigated by Hammers and coworkers[9–11]. In this mode liquid–solid adsorption is the retention mechanism. The reported ϱ values (Table IX) were explained in terms of hydrogen bonding between the solutes and the adsorbent. The positive signs of the ϱ values for the phenols indicate the proton-donating properties of these solutes.

TABLE IX

ϱ PARAMETERS IN NORMAL-PHASE SYSTEMS AS CALCULATED BY HAMMERS *et al.*[9–11] ON OCTADECYLSILYLSILICA (ODS), N-2-CYANOETHYL-N-METHYLAMINOSILICA (CNA) AND AMINOBUTYLSILYLSILICA (ABS)

Bonded phase	*k*	*i*	ϱ	s_y
ODS	OH	1–3,5–8	1.04 ± 0.15	0.19
	NH_2	1–3,5–8	-2.60 ± 0.16	0.16
	=N–	2,3,5,7	-3.39 ± 0.21	0.12
CNA	OH	1–3,5–10	0.82 ± 0.16	0.21
	NH_2	1–3,5–9	-1.76 ± 0.16	0.16
	=N–	2,3,5,7	-3.03 ± 0.35	0.19
ABS	OH	1–3,5–10	2.56 ± 0.20	0.26
	NH_2	1–3,5–9	-0.46 ± 0.26	0.27
	=N–	2,3,5,7	-2.42 ± 0.45	0.28

Anilines and pyridines act as proton acceptors towards the silanol sites of the adsorbent, and have negative ϱ parameters. These results are analogous to earlier findings by Snyder[43] for phenols and pyridines on both bare silica and alumina.

The signs of the ϱ values in our reversed-phase system agree with those calculated from batch partition experiments (Table IV), and with the signs of ϱ values calculated from literature data of the octanol–water system (Table X). Thus the positive signs exhibited by all investigated series, whether acidic or basic, are not a peculiar feature of the complicated reversed-phase system. We believe that all solutes examined act as proton acceptors towards the aqueous phase. This is a necessity for the pyridines, and rather obvious for the anilines, but, at first sight, seems odd for the phenols and benzoic acids. However it must be realized that in the case of the phenols the aqueous phase was acidified to pH 3.5 and in the case of the benzoic acids to pH 2.0, in order to suppress the basic properties of the aqueous phase towards these substances. It is plausible that the so-generated H_3O^+ and $CH_3OH_2^+$ ions act as proton donors towards phenols and benzoic acids.

Interaction between two polar substituents

A molecule such as aminophenol has *two* ways of hydrogen bonding with the

TABLE X

CALCULATED ϱ_I AND ϱ_R PARAMETERS FOR THE OCTANOL–WATER SYSTEM[2]

i	*k*		ϱ_I	ϱ_R	s_y
1–8	COOH	*m*	0.29 ± 0.05	-0.08 ± 0.09	0.06
1–8		*p*	0.37 ± 0.07	0.24 ± 0.13	0.08
1–10	OH	*m*	1.15 ± 0.06	0.62 ± 0.11	0.07
1–10		*p*	1.01 ± 0.04	0.89 ± 0.08	0.05
1,2,3,5,6,8	NH_2	*m*	0.96 ± 0.06	0.58 ± 0.10	0.06
1,2,3,5,6,8		*p*	0.97 ± 0.16	0.66 ± 0.27	0.15

eluent molecules: one by its hydroxyl and the other by its amino group. This means that the influence of both the amino group on the electron density in the hydroxyl group and of the hydroxyl group on the electron density in the amino group has to be taken into account. In other words, for the case of two polar substituents, eqn. 8 becomes

$$\Delta\Delta \log V_N^{i\phi k} = \{(\varrho^i\sigma^k)_I + (\varrho^i\sigma^k)_R\} + \{(\varrho^k\sigma^i)_I + (\varrho^k\sigma^i)_R\} \tag{15}$$

Calculations using this formula can be carried out since ϱ parameters can be taken from Table IV. Calculations cannot be performed for benzoic acids and pyridines because no σ values for the carboxyl and pyridyl groups are available. This limits discussion to the aminophenols, dihydroxybenzenes and diaminobenzenes in methanol–water (50:50, v/v). The results (Table XI) show poor agreement between measured and calculated values. The aminophenols may have been partly ionized, at both pH 3.5 and 7.1, but this should make the experimental $\Delta\Delta$ log V_N values *smaller* than the calculated ones.

TABLE XI

VALUES OF $\Delta\Delta$ LOG $V_N^{i\phi k}$ IN METHANOL–WATER (50:50, v/v) OF COMPOUNDS IN WHICH *i* AND *k* ARE HYDROXYL AND AMINO GROUPS

k		$\Delta\Delta \log V_N^{i\phi k}$					
		i = OH (pH 3.5)			*i = NH₂ (pH 7.1)*		
		Meas.	*Calc.*	*Diff.*	*Meas.*	*Calc.*	*Diff.*
OH	*m*	0.01	−0.13	0.14	0.09	−0.23	0.32
	p	−0.20	−0.25	0.05	−0.02	−0.31	0.29
NH_2	*m*	−0.03	−0.23	0.20	0.13	−0.32	0.19
	p	−0.08	−0.31	0.23	−0.02	−0.32	0.30

Retention mechanisms in reversed-phase chromatography with alkyl-silica sorbents

A discussion of this subject is appropriate for two reasons. First, it is important to clarify the retention mechanism in the chromatographic systems that we have investigated. Second, our data from the batch partition experiments can shed some light on this mechanism.

Several possible retention mechanisms can be envisaged: (1) partition between the eluent and the bound organic layer, which may have taken up preferentially one of the components from the eluent[14]; (2) partition between the bulk eluent and a thick layer of adsorbed eluent, the composition of which is different from that of the bulk[15]; (3) interaction with residual silanol groups[16]; (4) adsorption on a monomolecular layer of adsorbed eluent[17]; (5) adsorption on the bound organic layer[18].

Mechanism 1. If retention is governed by partition between the eluent and the bound organic layer, the chromatographic distribution coefficient, K, is equal to the partition coefficient, P, determined by batch experiments. In other words,

$$\log V_N = \log K\Phi_s = \log P + \log \Phi_s \tag{16}$$

Our data on log *P* (Table VI) enable us to investigate the validity of this relationship for the methanol–water (50:50, v/v)/octadecyl silica system, and a series of solutes ranging in polarity from benzene to substituted phenols. Fig. 2 shows data on log V_N as a function of log *P*. Inspection of this figure reveals that eqn. 16 must be rejected on the following arguments: (i) the data points are severely scattered; (ii) they do not lie on a straight line with a slope equal to unity; (iii) although the data points for the apolar solutes do lie on such a line, its intercept is much too large. From the intercept it would follow that V_s is equal to 5.06 ml, whereas the actual volume of the bound alkyl layer in the column is *ca.* 260 μl. Analogous conclusions were drawn by Hammers *et al.*[44] for the water/octadecyl silica system with a series of 27 solutes comprising methylbenzenes, fused arenes, halogenated benzenes, chloroanilines, chlorophenols and polar monosubstituted benzenes.

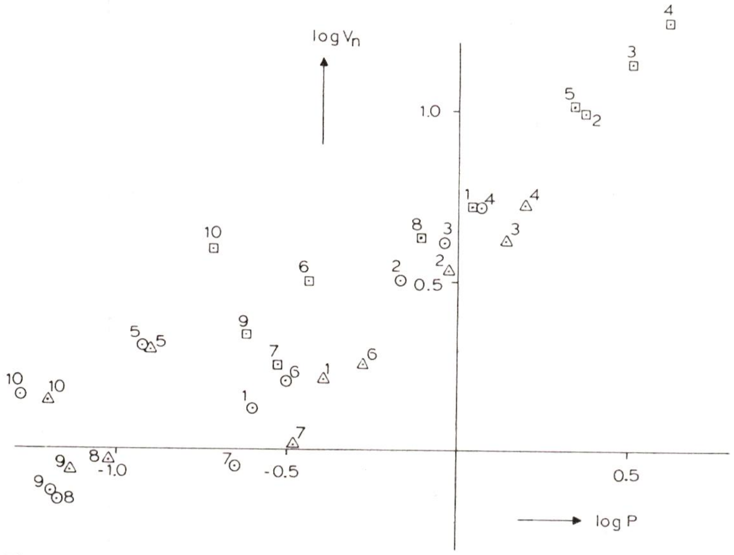


Fig. 2. Log V_N values of substituted benzenes (⊡) and *meta*- (△) and *para*- (⊙) substituted phenols, measured in methanol–water (50:50, v/v) at pH 3.5, as a function of the corresponding log *P* values for batch partition in the *n*-hexadecane/methanol–water (50:50, v/v) system.

Mechanism 2. With methanol–water mixtures any contribution to retention from this mechanism can be ruled out, as adsorption of methanol to the alkyl silica does not proceed beyond a monolayer[44]. However, alkyl silica does adsorb large amounts of less polar organic co-solvents from their mixtures with water, and in these cases a partition mechanism may contribute to retention.

According to Slaats *et al.*[45] the volume of the solvent layer that is adsorbed on to alkyl silica is, in acetonitrile–water mixtures, about four times as large as in methanol–water mixtures. Organic co-solvents that are less polar than acetonitrile are even more strongly adsorbed. For instance, the adsorbed amount of tetrahydrofuran is about twice as large as that of acetonitrile[46]. The behaviour of a more exotic co-solvent, 1-pentanol, is enlightening in this respect[47]. This compound is only slightly soluble in water with a maximum concentration of 2.30 % (v/v). When the concentration is 30–70 % of this value, a monomolecular layer is adsorbed onto the surface of octyl silica. When the concentration of 1-pentanol in the solution is increased from 70 to 100 % saturation, the adsorbed amount increases by a factor of 5. At the saturation point almost the complete pore volume is filled with 1-pentanol. The authors estimate that the contribution of partition to the retention volume varies from about 30 % at 85 % saturation to 95 % at 100 % saturation.

Mechanism 3. Interaction of the solutes with residual silanol groups is not likely to occur[48]. These groups are shielded by the octadecyl chains[49] and, besides, the eluent offers a vast number of hydroxyl groups[50]. Furthermore, it appears that log V_N data for acidic, neutral and basic solutes correlate well with the corresponding log P data for the octanol–water system[20,44].

Mechanism 4. In our opinion, adsorption on a monomolecular layer of adsorbed eluent rarely occurs. There are no arguments[46,51] for the assumption that the exchange of adsorbed eluent molecules such as methanol or acetonitrile for sample molecules does not occur. (It is, formally, the mechanism prevailing in ion-pair chromatography, if the counter ion that is added to the eluent is strongly adsorbed by the alkyl silica).

Mechanism 5. Adsorption on the bound organic layer is probably the retention mechanism in methanol–water eluents. Retention data are in good agreement with Locke's model for competitive adsorption in a monolayer[44,52]. Retention data in 1-pentanol–water eluents can also be described by assuming competitive adsorption in a monolayer, up to 70 % of the saturation concentration of 1-pentanol.

Summarizing, we conclude that in our experiments with 50 % (v/v) methanol and probably also with 30 % (v/v) acetonitrile, retention is governed by adsorption on the bound alkyl layer.

Statistics

As has been mentioned above, the standard deviation of the $\Delta\Delta$ log V_N values is 0.014. This is much better than the standard errors of fit of the two-parameter regression analyses. These latter values vary from 0.01 to 0.06, as can be seen from Table IV. This can be understood if the exact expression for $\Delta\Delta$ log V_N is considered. If we overlook, for the sake of simplicity, the splitting into inductive and resonance parts we can write the exact eqn. 8 as

$$\Delta\Delta \log V_N^{i\phi k} = \varrho_i\sigma_k + \varrho_k\sigma_i \qquad (17)$$

Within a certain solute series the term $\varrho_i\sigma_k$ varies because the value of ϱ_i for the variable substituent i varies. The $\varrho_k\sigma_i$ term is the part we have been dealing with when calculating the ϱ parameters found in Table IV. The $\varrho_i\sigma_k$ term accounts for the influence of the fixed group k on the variable group i. It is small compared to the $\varrho_k\sigma_i$ term, but not negligible; its presence implies that the s_y values of the regressions are composed of the experimental error (0.014) and the neglected $\varrho_i\sigma_k$ terms:

$$(s_y)^2 = (0.014)^2 + (s')^2 \qquad (18)$$

where s' accounts for the neglected $\varrho_i\sigma_k$ terms. It turns out that s' varies from zero to 0.06. That this hypothesis is correct can be investigated as follows. Eqn. 17 should also hold for the substituted chlorobenzenes. However, in this case the fixed substituent k (Cl) and the variable substituents i are of comparable polarity. Hence, the terms $\varrho_i\sigma_k$ and $\varrho_k\sigma_i$ will be of comparable magnitude. This means that the variance within the $\Delta\Delta$ log V_N values of the substituted chlorobenzenes is equal to

$$s^2 = (0.014)^2 + 2(s')^2 \tag{19}$$

With eqn. 19 we find, from the data on *meta*- and *para*-substituted chlorobenzenes at pH 3.5 and 7.1, an estimate of 0.04 for s'. This is in good agreement with the estimate of 0.06 given above. Therefore we can conclude that the proposed two-parameter equation is valid as a description of the substituent interaction effect in all the solute series. Addition of a $\varrho_E\sigma_E$ term[6] does not give any improvement in the standard error of fit for our regressions. The single-term Hammett equation, on the other hand, does not hold at all.

CONCLUSIONS

The influence of substituents on the retention behaviour of benzoic acids, phenols, anilines and pyridines in reversed-phase liquid chromatography can be described using the Taft equation. For molecules with more than one strongly polar substituent the description fails. The signs of the values of the ϱ parameters in the Taft equation correspond with the signs of the ϱ parameters holding for batch partition coefficients, but not with the signs of the ϱ parameters holding for normal-phase liquid chromatography or for GLC. The signs can be explained by assuming that all the solutes (pyridines, anilines, phenols and benzoic acids) act as proton acceptors in hydrogen bonding with the (acidified) mixed aqueous–organic eluents.

Adsorption on the bound layer is the prevailing retention mechanism in our experiments.

REFERENCES

1 L. P. Hammett, *Physical Organic Chemistry*, McGraw–Hill, New York, 1940, p. 184.
2 A. Leo, C. Hansch and D. Elkins, *Chem. Rev.*, 71 (1971) 525.
3 S. R. Bakalyar, R. McIlwrick and E. Roggendorf, *J. Chromatogr.*, 142 (1977) 353.
4 N. Tanaka, H. Goodell and B. L. Karger, *J. Chromatogr.*, 158 (1978) 233.
5 T. Hanai, C. Tran and J. Hubert, *J. High Resolut. Chromatogr. Chromatogr. Commun.*, 4 (1981) 454.
6 G. H. E. Nieuwdorp, C. L. de Ligny and N. G. van der Veen, *J. Chromatogr.*, 154 (1978) 133.
7 R. W. Taft and J. C. Lewis, *J. Amer. Chem. Soc.*, 80 (1958) 2436.
8 L. R. Snyder, *Principles of Adsorption Chromatography*, Marcel Dekker, New York, 1968, p. 64.
9 W. E. Hammers, R. H. A. M. Janssen, A. G. Baars and C. L. de Ligny, *J. Chromatogr.*, 167 (1978) 273.
10 W. E. Hammers, C. H. Kos, W. K. Brederode and C. L. de Ligny, *J. Chromatogr.*, 168 (1979) 9.
11 W. E. Hammers, M. C. Spanjer and C. L. de Ligny, *J. Chromatogr.*, 174 (1979) 291.
12 N. H. C. Cooke and K. Olsen, *J. Chromatogr. Sci.*, 18 (1980) 512.
13 E. L. Johnson and R. Stevenson, *Basic Liquid Chromatography*, Varian Assoc., Palo Alto, CA, 1978, p. 36.
14 H. Colin and G. Guiochon, *J. Chromatogr.*, 158 (1978) 183.
15 C. H. Löchmuller and D. R. Wilder, *J. Chromatogr. Sci.*, 17 (1979) 574.
16 A. Nahum and Cs. Horváth, *J. Chromatogr.*, 203 (1981) 53.
17 R. P. W. Scott and P. Kucera, *J. Chromatogr.*, 142 (1977) 213.
18 Cs. Horváth, W. Melander and I. Molnár, *J. Chromatogr.*, 125 (1976) 129.
19 D. O. Rauls and J. K. Baker, *J. Med. Chem.*, 22 (1979) 85.
20 H. Könemann, R. Zelle, F. Busser and W. E. Hammers, *J. Chromatogr.*, 178 (1979) 559.

21 R. E. Majors, *Anal. Chem.*, 44 (1972) 1722.
22 G. E. Berendsen, P. J. Schoenmakers, L. de Galan, G. Vigh, Z. Varga-Puchony and J. Inczédy, *J. Liquid Chromatogr.*, 3 (1980) 1669.
23 K. Karch, I. Sebastian and I. Halász, *J. Chromatogr.*, 122 (1976) 3.
24 P. E. Barker, B. W. Hatt and S. R. Holding, *J. Chromatogr.*, 206 (1981) 27.
25 C. L. de Ligny, *Rec. Trav. Chim.*, 79 (1960) 731.
26 A. Albert and E. P. Serjeant, *Ionization Constants of Acids and Bases*, Wiley, New York, 1962.
27 S. R. Bakalyar, M. P. T. Bradley and R. Honganen, *J. Chromatogr.*, 158 (1978) 277.
28 *Handbook of Chemistry and Physics*, Chemical Rubber Co., Cleveland, OH, 56th edn., 1975–1976.
29 R. Mecke, *Disc. Farad. Soc.*, 9 (1950) 161.
30 G. Aknes and T. Gramstad, *Acta Chim. Scand.*, 14 (1960) 1485.
31 J. W. Mathison and R. Tidwell, *J. Med. Chem.*, 18 (1975) 1227.
32 A. Nahum and Cs. Horváth, *J. Chromatogr.*, 192 (1980) 315.
33 L. G. Sillén and A. E. Martell, *Stability Constants of Metal–Ion Complexes*, Special Publications nos. 17 and 25, Chemical Society, London, 1964.
34 H. C. Brown and D. H. McDaniel, *J. Amer. Chem. Soc.*, 77 (1955) 3756.
35 D. D. Perrin, *Dissociation Constants of Organic Bases in Aqueous Solution*, Butterworths, London, 1965.
36 M. M. Fickling, A. Fischer, B. R. Mann, J. Packer and J. Vaughan, *J. Amer. Chem. Soc.*, 81 (1959) 4226.
37 P. D. Bolton, F. M. Hall and I. H. Reece, *J. Chem. Soc., B*, (1967) 709.
38 C. van Hooidonk and L. Ginjaar, *Rec. Trav. Chim.*, 86 (1967) 449.
39 M. Bartusek and L. Sommer, *J. Inorg. Nucl. Chem.*, 27 (1965) 2397.
40 R. R. Rao and P. K. Bhattacharya, *Current Sci.*, 36 (1967) 71.
41 A. Bryson, *J. Amer. Chem. Soc.*, 82 (1960) 4871.
42 G. H. E. Nieuwdorp, C. L. de Ligny and H. C. van Houwelingen, *J. Chem. Soc., Perkin Trans. II*, 75 (1979) 537.
43 L. R. Snyder, *Principles of Adsorption Chromatography*, Marcel Dekker, New York, 1968, p. 304.
44 W. E. Hammers, G. J. Meurs and C. L. de Ligny, *J. Chromatogr.*, 246 (1982) 169.
45 E. H. Slaats, W. Markovski, J. Fekete and H. Poppe, *J. Chromatogr.*, 207 (1981) 299.
46 R. M. McCormick and B. L. Karger, *Anal. Chem.*, 52 (1980) 2249.
47 K. G. Wahlund and J. Beijersten, *Anal. Chem.*, 54 (1982) 128.
48 S. Marcinkiewicz and J. Green, *J. Chromatogr.*, 10 (1963) 372.
49 P. Roumeliotis and K. K. Unger, *J. Chromatogr.*, 149 (1978) 211.
50 J. L. M. v.d. Venne, J. P. M. Rindt, G. J. M. M. Coenen and C. A. M. G. Cramers, *Chromatographia*, 13 (1980) 11.
51 R. M. McCormick and B. L. Karger, *J. Chromatogr.*, 199 (1980) 259.
52 H. J. Möckel and T. Freyholdt, *Z. Anal. Chem.*, 368 (1981) 401.

CHAPTER II

SUBSTITUENT INTERACTION EFFECTS IN AROMATIC MOLECULES IN RP-HPLC WITH ACETONITRILE-WATER AND TETRAHYDROFURAN-WATER ELUENTS

ABSTRACT

Retention volumes of monosubstituted benzenes, benzoic acids, phenols and anilines have been measured in reversed-phase liquid chromatography. Buffered acetonitrile-water and tetrahydrofuran-water eluents were used with an octadecylsilylsilica adsorbent. From the net retention volumes a substituent interaction effect was calculated and described with the linear free energy relationship developed by Nieuwdorp *et al.* The positive sign of the values of the ρ-parameters, figuring in this relationship, was interpreted in terms of hydrogen bonding between the solutes and the eluent.

Further, batch partition experiments were carried out in *n*-hexadecane/eluent systems, with monosubstituted benzenes and phenols. The results were compared with those of the chromatographic measurements, and it was concluded that retention cannot be described in terms of partition between the eluent and the stationary phase, saturated with this eluent.

INTRODUCTION

The effect of a substituent in a parent molecule on its properties can be described by the Hammett (1) or the Taft (2) equation. These equations can be used for the description of the substituent (interaction) effect on chemical eqiulibria, reaction rates, physical properties and partition equilibria (3). Nieuwdorp *et al.* (4) were the first to investigate the substituent interaction effect in gas-liquid chromatography. These workers investigated the retention of substituted phenols, anilines and pyridines (and, as a reference, the retention of substituted benzenes) on two stationary phases. Interaction between substituents is an important phenomenon in chromatography, because its occurrence implies that simple addition of group contributions (5-8) does not lead to an accurate prediction of retention. We focus our attention on the substituent interaction effect in reversed-phase liquid chromatography.

In a previous publication (9) methanol-water was used as the eluent, and a limited amount of data was presented for acetonitrile-water also. In this paper, the substituent interaction effect in acetonitrile-water and tetrahydrofuran-water eluents is investigated. These three eluents are used in the majority of applications of reversed-phase chromatography. It is our purpose to give a quantitative description of the substituent interaction effect with the linear substituent free energy relationship proposed by Taft (2) or Nieuwdorp (4).

THEORETICAL

The solute series investigated in this study consisted of substituted benzoic acids ($i\phi$COOH), phenols ($i\phi$OH) and anilines ($i\phi NH_2$), with the substituents fluoro, chloro, bromo, iodo, methyl, nitro, cyano, methoxy, acetyl and carboxymethyl ester. They can be represented schematically by $i\phi k$, a general expression for a benzene nucleus with a variable substituent i and a fixed polar group k. (Although in pyridines the nitrogen atom is a part of the ring, they are also denoted by $i\phi k$, for the sake of simplicity).

Hammett described the influence of substituents on a property X of a compound $i\phi k$ by

$$\log X^{i\phi k} - \log X^{H\phi k} = \rho_k \sigma_i \qquad (1)$$

Herein ρ_k expresses the sensitivity of the fixed group k towards electronic shifts in the benzene nucleus caused by substituent i, and the substituent constant σ_i indicates the influence of a varying substituent i on the fixed functional group k. Depending on the position of the groups, σ_m constants are used for *meta* and σ_p values are used for *para* substituents, whereas the ρ parameter has the same value for *meta*- and *para*-substituted compounds. Taft *et al.* distinguished between an inductive (I) and a resonance (R) effect, thus extending equation (1) to:

$$\log X^{i\phi k} - \log X^{H\phi k} = \rho_I \sigma_I + \rho_R \sigma_R \qquad (2)$$

Meta- and *para*-substituted compounds have the same σ_I and σ_R values, but they are treated separately, which results in ρ_I and ρ_R values that differ for *meta* and *para* series. A disadvantage of this model is the necessity to make a choice between four different σ_R sets, σ_R^{o}, σ_R^+, σ_R^- and σ_R (BA) (10).

Often it is not obvious, a priori, which set should be used, so that the choice has to be made afterwards, on the basis of the residual standard deviations of the four regressions.

Nieuwdorp improved on this point in his equation:

$$\log X^{i\phi k} - \log X^{H\phi k} = \rho_I\sigma_I + \rho_R\sigma_R + \rho_E\sigma_E \tag{3}$$

The σ_I and σ_R values in this formula are comparable with Taft's σ_I and σ_R^{o} values, respectively. The σ_E values represent a mixture of the inductive effect, the resonance effect and the effect of direct resonance.

The left-hand side of equations (1)-(3) represents a substituent *interaction* effect. However, when X is a chromatographic net retention volume or a batch partition coefficient (rather than a chemical equilibrium constant or a reaction rate constant), $X^{i\phi k}$ contains also a contribution of the substituent i proper, which does not cancel in the difference $\log X^{i\phi k} - \log^{H\phi k}$. Therefore, for our purpose (*i.e.*, the application to phase eqilibria), the left-hand side of equations (1)-(3) must be modified to:

$$\text{l.h.s.} = (\log X^{i\phi k} - \log X^{H\phi k}) - (\log X^{i\phi H} - \log X^{H\phi H}) \equiv \Delta\Delta \log X^{i\phi k} \tag{4}$$

This can be proved as follows.

If $\log X^{i\phi k}$ is composed of group contributions and first-order interactions between the moieties i, ϕ and k, *i.e.*:

$$\log X^{i\phi k} = a^{i} + a^{\phi} + a^{k} + a^{i\phi} + a^{\phi k} + a^{ik} \qquad (5)$$

it follows that:

$$\Delta\Delta\log X^{i\phi k} = (a^{ik} - a^{Hk}) - (a^{iH} - a^{HH}) \qquad (6)$$

Both terms within parentheses in the right-hand side of equation (6) represent a first-order substituent-interaction effect, for which equation (3) holds:

$$\begin{aligned}\Delta\Delta\log X^{i\phi k} &= (\rho_I\sigma_I + \rho_R\sigma_R + \rho_E\sigma_E) - (\rho'_I\sigma_I + \rho'_R\sigma_R + \rho'_E\sigma_E)\\ &= \rho''_I\sigma_I + \rho''_R\sigma_R + \rho''_E\sigma_E \qquad (7)\end{aligned}$$

where $\rho'' = \rho - \rho'$.

In this study X is the net retention volume V_N or the batch partition coefficient P. Retention volumes were measured in the chromatographic systems octadecylsilylsilica/eluent and partition coefficients were measured by means of batch experiments in n-hexadecane/eluent systems. The latter data were calculated from UV absorbances in the eluent phase by:

$$P = \frac{C_{\text{hexadecane}}}{C_{\text{eluent}}} = \frac{A_{\text{before equilibration}} - A_{\text{after equilibration}}}{A_{\text{after eqilibration}}} \cdot \frac{V_{\text{eluent}}}{V_{\text{hexadecane}}} \qquad (8)$$

where C represents the concentration of the solute, A its absorbance in the eluent and V the volume of the equilibrated quantities of both phases (11).

Measurements on *ortho*-substituted compounds were not carried out as steric effects and intramolecular hydrogen bonding disturb a description of substituent influences in terms of σ constants (12). No measurements on solutes wherein $i=k$ (*i.e.*, $C_6H_4(COOH)_2$, $C_6H_4(OH)_2$ and $C_6H_4(NH_2)_2$) were made, because previous work (9) had already demonstrated the failure of the description of substituent interaction effects in these molecules by eqns. (1)-(3).

EXPERIMENTAL

All experimental procedures have been described previously (9), so in this paper we mention only the modifications.

Chemicals

The eluent modifier tetrahydrofuran was purchased from Aldrich Europe (Beerse, Belgium). For chromatographic experiments 99+% quality and for batch partition experiments gold label 99.5+% quality was used.

Procedure

The void volume of the newly packed column was determined by the static weighting method, as before (9). The volume was measured in triplicate, with *n*-heptane/tetrachloromethane, *n*-heptane/methanol and methanol/tetrachloromethane as solvent combinations, giving (620 ± 10) µl. Within experimental error the weigth of the column packing was 0.63 g, as before. The tetrahydrofuran percentage was 32.5 v/v %. This eluent composition gives retention volumes of a reasonable magnitude for substituted benzenes as well as for substituted benzoic acids, phenols etc.

Ultrasonic degassing of this eluent after gently heating up to 40°C (as done previously with methanol-water and acetonitrile-water eluents) was not sufficient. We therefore degassed the required daily portion by nitrogen sparging overnight, this being the best method according to Bakalyar *et al.* (13). Afterwards ultrasonic degassing was performed to expel the nitrogen. Eventual differences in eluent composition, caused by nitrogen sparging, were corrected for by measuring the retention volumes of the nitro-substituted compounds in each series every day.

As in MeOH-water mixtures, the pK_a of the phosphoric acid in the buffer used can be expected to increase upon addition of the organic modifier and the same applies to the pH of the buffered eluent. Our earlier estimate of an increase of *ca.* 0.5 unit in 50 v/v% MeOH-water has been confirmed by Leitold and Vigh (14). A similar increase can be expected in 30 v/v % ACN-water or 32.5 v/v % THF-water mixtures. Further the pK_a values of all solutes will change.

In MeOH-water mixtures they increase for carboxylic acids and decrease for anilines (15). A similar increase or decrease can be expected in ACN-water and THF-water mixtures. Hanai *et al.* (16) derived a relation for benzoic acids in ACN-water mixtures from which we calculated a pK_a increase of 0.66. Summarizing, upon addition of the organic modifier to the aqueous phosphate buffer, both the pH of the eluent and the pK_a of the investigated acids increase about 0.5. The pK_a of the investigated bases decreases. Therefore we believe all solutes to be unionized when we buffer both ACN-water and THF-water eluents at pH = 2.0 for benzoic acids, 3.5 for phenols and 7.1 for anilines.

RESULTS

The results of the chromatographic experiments are shown in Table I for the acetonitrile-water and in Table II for the tetrahydrofuran-water eluent. The presented values are the means of duplicate measurements and have a standard deviation of 0.01, corresponding with a precision in $\bar{V}_N$ of 2.1%. The precision is slightly worse for bromo- and iodobenzene, where tailing deteriorates the standard deviation in $\bar{V}_N$ to ca. 3.5%.

Table I - Log V_N values in acetonitrile-water (30:70, v/v), at pH = 7.1 (V_N is measured in ml.)

Substituent		Series		
		Benzenes	Anilines	
i	No.		m	p
H	0	0.85	0.01	0.01
F	1	0.89	0.30	0.14
Cl	2	1.21	0.62	0.56
Br	3	1.30	0.71	0.67
I	4	1.49	0.87	0.86
CH_3	5	1.17	0.34	0.33
NO_2	6	0.73	0.32	0.22
CN	7	0.52	0.09	0.02
OCH_3	8	0.82	0.03	-0.16
$COCH_3$	9	0.46	-0.21	-0.36
$COOCH_3$	10	0.78	-	0.13

Table II - Log V_N values in tetrahydrofuran-water (32.5:67.5, v/v) at several pH values (V_N is measured in ml.)

Substituent	Series (pH)						
	Benzenes	Benzoic acids (2.0)		Phenols(3.5)		Anilines(7.1)	
i		*m*	*p*	*m*	*p*	*m*	*p*
H	0.88	0.26	0.26	0.30	0.30	0.01	0.01
F	0.91	0.52	0.50	0.59	0.46	0.36	0.50
Cl	1.12	0.72	0.76	0.80	0.75	0.59	0.48
Br	1.19	0.80	0.84	0.86	0.82	0.67	0.56
I	1.30	0.86	0.93	0.96	0.93	0.77	0.69
CH_3	1.11	0.49	-	0.50	0.49	0.30	0.26
NO_2	0.69	0.54	0.66	0.65	0.60	0.48	0.28
CN	0.40	0.24	0.34	0.35	0.23	0.14	-0.02
OCH_3	0.74	0.27	0.18	0.22	0.03	-0.10	0.31
$COCH_3$	0.24	-0.08	0.02	-0.12	-0.28	-0.51	-0.21
$COOCH_3$	0.58	-	-	0.26	0.23	-	-0.03

The results of the batch partition experiments are collected in Table III for both solvent systems. The precision of the log P values is worse than of the log V_N values. Here we find a standard deviation of 0.05 in log P, corresponding with an uncertainty of 11% in the means of duplicate P determinations. It follows that the standard deviations of $\Delta\Delta$log V_N and $\Delta\Delta$log P are equal to 0.02 and 0.10, resp. (*i.e.*, twice those in log V_N and log P). To demonstrate the presence of the substituent interaction effect in acetonitrile-water and tetrahydrofuran-water eluents, graphs of the $\Delta\Delta$ log V_N values *versus* the corresponding values in methanol-water were drawn. They are presented in figure 1

Table III - Log P values in n-hexadecane/acetonitrile-water (30:70, v/v) and n-hexadecane/tetrahydrofuran-water (32.5:67.5, v/v) at pH 3.5

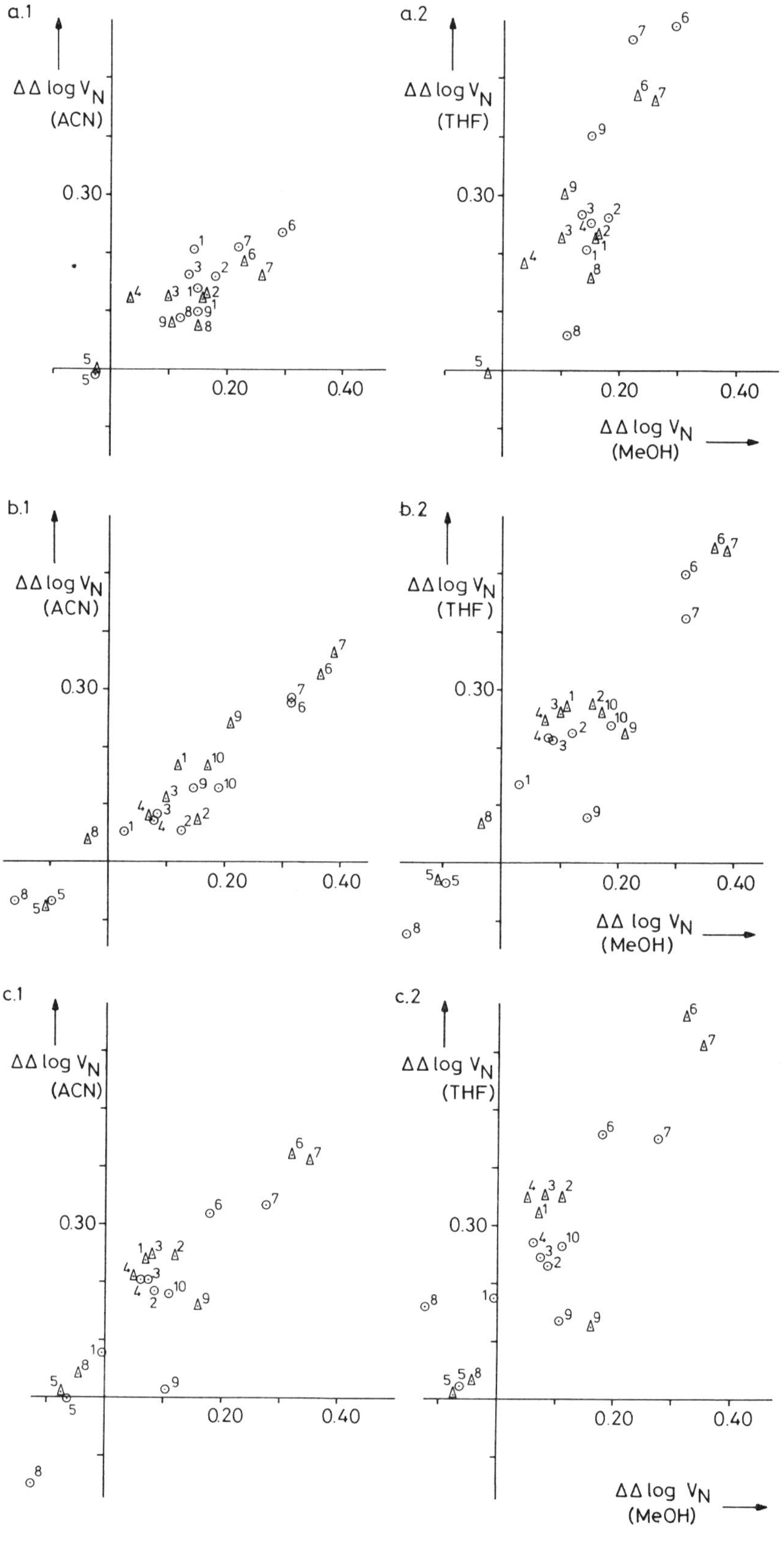
Substituent	Series (pH)					
	Acetonitrile-water			Tetrahydrofuran-water		
	Benzenes	Phenols (3.5)		Benzenes	Phenols (3.5)	
i		m	p		m	p
H	0.11	-0.47	-0.47	0.35	-0.22	-0.22
F	0.20	-0.25	-0.80	0.45	0.88	0.73
Cl	0.35	-0.02	0.03	0.48	0.81	0.85
Br	0.53	-0.10	0.10	0.59	0.87	1.08
I	0.65	0.58	0.12	0.69	0.76	0.79
CH_3	0.40	-0.30	-0.32	0.48	-0.16	-0.19
NO_2	-0.26	0.03	-0.12	0.13	0.49	0.68
CN	-0.12	-0.05	-0.20	0.10	0.54	0.62
OCH_3	0.18	-0.76	-0.73	0.23	0.06	0.06
$COCH_3$	-0.37	-0.37	-0.43	0.05	0.20	0.32
$COOCH_3$	-0.17	-0.73	-0.74	-0.03	0.21	0.28

Figure 1 - Graphs of ΔΔ log V_N values of *meta*- (Δ) and *para*- (o) substituted benzoic acids measured in acetonitrile-water (30:70,v/v) (a.1.) and in tetrahydrofuran-water (32.5:67.5, v/v) (a.2.) as a function of the corresponding ΔΔ log V_N values measured in methanol-water (50:50, v/v).

b.1 and 2 : idem, for phenols

c.1 and 2 : idem, for anilines

Data in methanol-water, and on the benzoic acids and phenols in acetonitrile-water are from Ref. (9).


a.1
ΔΔ log V_N (ACN)
0.30
0.20
0.40
a.2
ΔΔ log V_N (THF)
0.30
0.20
0.40
ΔΔ log V_N (MeOH)
b.1
ΔΔ log V_N (ACN)
0.30
0.20
0.40
b.2
ΔΔ log V_N (THF)
0.30
0.20
0.40
ΔΔ log V_N (MeOH)
c.1
ΔΔ log V_N (ACN)
0.30
0.20
0.40
c.2
ΔΔ log V_N (THF)
0.30
0.20
0.40
ΔΔ log V_N (MeOH)

for all series. In figure 2 the magnitude of the subsituent interaction effect in chromatographic experiments is compared with that in batch partition experiments. For that purpose the $\Delta\Delta\log V_N$ values for phenols were plotted *versus* the corresponding $\Delta\Delta\log P$ values in all three eluents.

The $\Delta\Delta\log V_N$ and $\Delta\Delta\log P$ values were analysed by the Taft equation (2), using the values of the substituent constants σ_I and σ_R given by Nieuwdorp *et al.* (17). This analysis yields the regression coefficients ρ_I and ρ_R in eqn. (2). The results are given in Table IV for both acetonitrile-water and tetrahydrofuran-water eluents. The s_y values in this table correspond with the standard deviation of an individual $\Delta\Delta\log V_N$ or $\Delta\Delta\log P$ value from the calculated value.

DISCUSSION

Substituent interaction effects

Figure 1 visualizes the presence of the substituent interaction effect in both acetonitrile-water and tetrahydrofuran-water eluents. It can be seen that its magnitude is comparable with that in the methanol-water mixture. From figure 2 we learn that for all eluents the effect is larger in batch partition systems. These observations are reflected in the values of the ρ parameters in Table IV. All ρ parameters are positive, as they are in methanol-water. The values of the ρ parameters for batch partition systems are larger than those for chromatographic measurements.

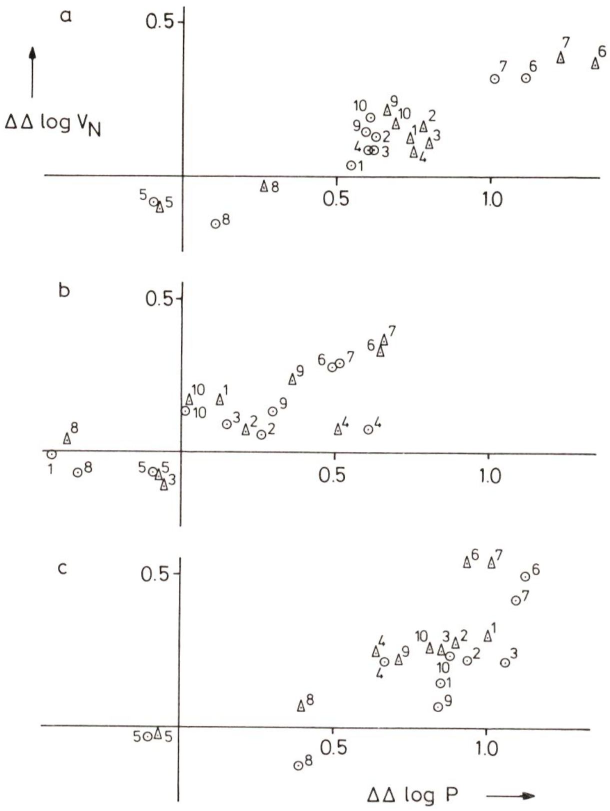


Figure 2 - Graphs of ΔΔ log V_N values of *meta*- (Δ) and *para*- (o) substituted phenols measured in methanol-water (50:50, v/v) (a), acetonitrile-water (30:70, v/v) (b) and tetrahydrofuran-water (32.5:67.5, v/v) (c) as a function of the corresponding ΔΔ log P values for the batch partition experiments in the *n*-hexadecane/eluent systems.

Data in methanol-water, and on ΔΔ log V_N in acetonitrile-water are from Ref. (9).

Table IV - Values of the ρ_I and ρ_R parameters in eqn. (2), calculated from $\Delta\Delta$ log V_N data in acetonitrile-water and tetrahydro furan-water, and from $\Delta\Delta$ log P data in n-hexadecane/ acetonitrile-water and n-hexadecane/tetrahydrofuran-water

		Acetonitrile-water			Tetrahydrofuran-water		
Chromatography		ρ_I	ρ_R	s_y	ρ_I	ρ_R	s_y
Benzoic acids	*m*	0.29 ± 0.01	0.02 ± 0.01	0.01	0.70 ± 0.05	0.30 ± 0.10	0.06
	p	0.37 ± 0.01	0.04 ± 0.01	0.01	0.86 ± 0.04	0.60 ± 0.09	0.05
Phenols	*m*	0.47 ± 0.05	0.37 ± 0.10	0.06	0.78 ± 0.03	0.38 ± 0.05	0.03
	p	0.35 ± 0.02	0.43 ± 0.04	0.03	0.62 ± 0.06	0.52 ± 0.12	0.07
Anilines	*m*	0.63 ± 0.02	0.25 ± 0.05	0.03	0.92 ± 0.07	0.39 ± 0.15	0.09
	p	0.45 ± 0.07	0.39 ± 0.14	0.08	0.67 ± 0.04	0.22 ± 0.09	0.05
Batch Partition							
Phenols	*m*	0.79 ± 0.16	1.05 ± 0.33	0.20	1.87 ± 0.13	0.15 ± 0.27	0.17
	p	0.50 ± 0.15	1.17 ± 0.30	0.18	2.07 ± 0.13	0.42 ± 0.28	0.17

In some cases the residual standard deviations s_y in Table IV are larger than experimental error. We tried to improve the fit of the regressions in two ways. The first way was to perform calculations with the appropriate σ_R values of Taft for the investigated series. Then, for the benzoic acids the σ_R *(BA)* scale and for the phenols and anilines the σ_R^- scale has to be used. The second way was to carry out regression analyses according to the three-parameter equation (3). A survey of all s_y values that are obtained is given in Table V. From this compilation we can conclude

Table V - Standard deviations s_y of regression analyses of $\Delta\Delta$ log V_N and $\Delta\Delta$ log P data according to equations (2) and (3).

Eluent		Acetonitrile-water			Tetrahydrofuran-water		
Equation		2		3	2		3
		σ_R^N	σ_R^T		σ_R^N	σ_R^T	
Chromatography							
Bezoic	*m*	0.01	0.01	0.01	0.06	0.06	0.03
acids	*p*	0.01	0.01	0.01	0.05	0.06	0.03
Phenols	*m*	0.06	0.05	0.06	0.03	0.04	0.03
	p	0.03	0.04	0.03	0.07	0.09	0.05
Anilines	*m*	0.03	0.04	0.02	0.09	0.10	0.08
	p	0.08	0.10	0.05	0.05	0.06	0.06
Batch Partition							
Phenols	*m*	0.20	0.22	0.20	0.17	0.16	0.18
	p	0.18	0.19	0.20	0.17	0.17	0.18

σ_R^N = using σ_R values of Nieuwdorp *et al.* (17)

σ_R^T = using σ_R *(BA)* values for benzoic acids and σ_R^- values for phenols and anilines, both given by Taft (10).

that the description of the data with equation (2), with σ_I and σ_R values according to Nieuwdorp, can hardly be improved by adding the term $\rho_E\sigma_E$. Use of Taft's σ_R (BA) and σ_R^- values gives even worse results.

For a correct interpretation of our results on s_y it is necessary to consider the exact expression for $\Delta\Delta$ log V_N (9). If we use, for the sake of simplicity, the formalism of the Hammett equation (1) we can write the exact eqn. (2) as

$$\Delta\Delta\log V_N^{i\phi k} = \rho_i\sigma_k + \rho_k\sigma_i \tag{9}$$

The $\rho_k\sigma_i$ term is the part we have been dealing with when calculating the ρ parameters given in Table IV. The $\rho_i\sigma_k$ term accounts for the influence of the fixed group k on the variable group i. Within a given solute series the term $\rho_i\sigma_k$ varies because the value of ρ_i for the variable substituent i varies. This term is small compared to the $\rho_k\sigma_i$ term, but not negligible. Its presence implies that the s_y values of the regressions are composed of 1) the experimental error, 2) a contribution s' associated with the neglected $\rho_i\sigma_k$ terms and 3) a contribution s'', associated with the error in the Taft or Nieuwdorp model.

$$(s_y)^2 = (0.02)^2 + (s')^2 + (s'')^2 \tag{10}$$

The magnitude of s' has been estimated to be 0.04 from measurements on substituted chlorobenzenes in methanol-water (9). As it is a reasonable assumption that this value also holds for acetonitrile-water and tetrahydrofuran-water eluents we can calculate that experimental error and the neglect of the term $\rho_i\sigma_k$ in eqn. (9) give rise to a contribution of 0.04 to s_y. The pooled values of the standard deviations given in Table V are 0.05 and 0.06 for eqn. (2), using Nieuwdorp's σ_R constants and Taft's σ_R (BA) and σ_R^- constants, respectively, and 0.04 for eqn. (3). It follows that the Nieuwdorp equation describes our data perfectly, whereas the Taft equation is beset with a model error of about 0.04. The results of the regression analyses on the $\Delta\Delta\log P$ data are bad in all cases. This may be due to the large experimental error in the $\log P$ values.

We now turn to the interpretation of the sign of the ρ parameters. For all solutes and in all eluents, the sign is positive. We presume that the substituent interaction effect on retention and on the partition coefficient reflects mainly the influence of substituents on hydrogen bonding of the solutes with the aqueous phase. When hydrogen bonding is stronger, the retention or the partition coefficient is smaller. So, the positive ρ values in Table IV correspond with a type of hydrogen bonding of the solutes with the aqueous phase that is characterized by negative ρ values. Seidel *et al.* (18) have demonstrated that for the dimerization of phenols by hydrogen bonding, the influence of substituents in the proton donor molecule is described by a positive ρ value, while the influence of substituents in the proton acceptor molecule is characterized by a negative ρ value (*i.e.*, electron withdrawing substituents, with positive σ constants, increase the proton donating ability and decrease the proton accepting ability of the parent molecule). This phenomenon holds probably generally. It follows that, in the hydrogen bonding of the solutes with the aqueous phase, observed in this study, the solutes act as proton acceptors towards the aqueous phase. This is rather obvious for the anilines, and understandable for the phenols and benzoic acids, when it is kept in mind that the aqueous phase is buffered to suppress its basic properties towards the last mentioned compounds, *i.e.* to suppress their ionization (9).

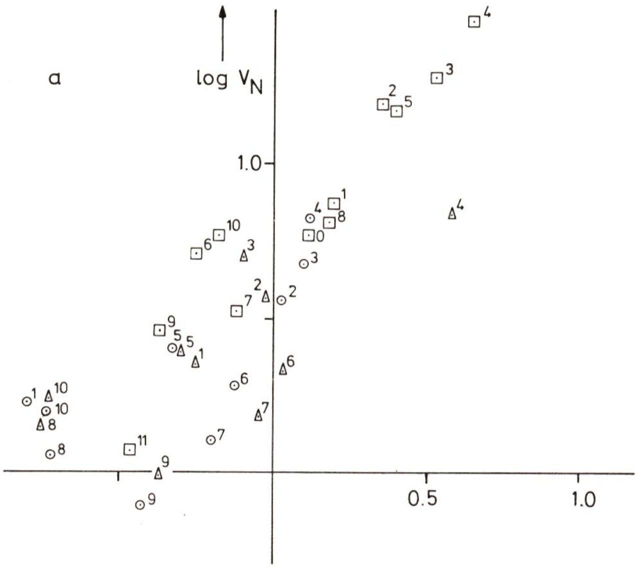
Comparison of chromatographic retention with batch partition

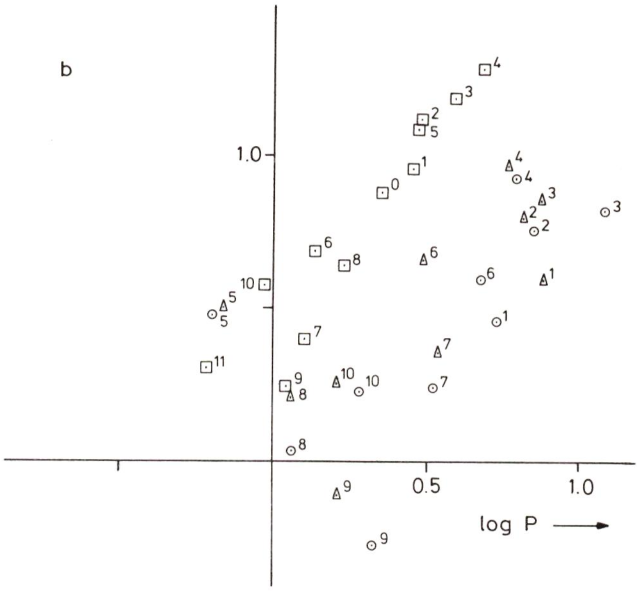
Our data permit to judge whether the sorption process in the chromatographic column resembles partition between the eluent and a long-chain hydrocarbon phase, saturated with the eluent. If this is the case, it should hold that:

$$\log V_N = \log P + \log V_s \tag{11}$$

where V_s is the volume (in ml) of the stationary phase in the column. In Fig. 3^a log V_N is plotted *vs.* log P for monosubstituted benzenes and phenols in acetonitrile-water, and in Fig. 3^b in tetrahydrofuran-water. It follows from these figures that equation (11) does not hold,

Figure 3 - Log V_N values of substituted benzenes (□) and *meta*- (Δ) and *para*- (o) substituted phenols, measured in acetonitrile-water (30:70, v/v) (a) and in tetrahydrofuran-water (32.5:67.5, v/v) (b) as a function of the corresponding log P values for the batch partition in the *n*-hexadecane/acetonitrile-water (30:70, v/v) and the *n*-hexadecane/tetrahydrofuran-water (32.5:67.5, v/v) systems respectively, all at pH 3.5.
Number 11 represents phenol. Data on log V_N in acetonitrile-water are from Ref. (9).


a
log V_N
1.0
0.5
1.0


b
1.0
0.5
1.0
log P

in neither solvent. This argument is confirmed by the plots of ΔΔ log V_N *vs.* ΔΔ log P, given in Fig. 2. These plots are severely scattered and their slopes are clearly smaller than the expected value 1.

In tetrahydrofuran-water the correlation between log V_N and log P is particularly bad (see also Fig. 2 of Ref. 9). This is remarkable in view of a report of Martire and Boehm (19) who state that the RP-18 layer is more extended in tetrahydrofuran-water, compared to its collapsed structure in methanol-water and acetonitrile-water eluents. For an extended layer one might expect partition to contribute to retention. However, this is clearly not the case.

For acetonitrile-water and methanol-water mixtures similar findings are reported by Colin *et al.* (20). They plot log k' *vs.* log K_{LLE} (the liquid-liquid extraction constant, *i.e.*, the partition coefficient for the *n*-hexadecane/eluent system) and obtain a collection of lines, each of which holds for a certain class of solutes.

The volume of the bound alkyl layer in the column is *ca.* 260 μl, if it has the density of *n*-hexadecane. Thus, log V_s in equation (11) is equal to *ca.* - 0.6. It follows from Fig. 3^a that retention in acetonitrile-water is an order of magnitude larger than corresponds with equation (11).

The same conclusion can be drawn from Fig. 3^b for retention of substituted benzenes in tetrahydrofuran-water.

CONCLUSIONS

The influence of interaction between substituents on the retention behaviour of benzoic acids, phenols and anilines in reversed-phase liquid chromatography can be described by the Taft equation, and slightly better by the Nieuwdorp equation. The signs of the ρ parameters obtained from chromatographic experiments and from corresponding batch partition experiments are all positive. This can be explained by assuming that the solutes act as proton acceptors in hydrogen bonding with the buffered aqueous-organic eluents.

The data on log V_N are not correlated with the data on log P for partition between *n*-hexadecane and the aqueous-organic eluent. The retention in acetonitrile-water eluents, and the retention of substituted benzenes in tetrahydrofuran-water eluents is an order of magnitude larger than corresponds with a partition mechanism.

REFERENCES

1. L.P. Hammett, *Physical Organic Chemistry*, McGraw-Hill, New York, 1940, p. 184.
2. R.W. Taft and J.C. Lewis, *J. Amer. Chem. Soc.*, 80:2436-43(1958).
3. A. Leo, C. Hansch and D. Elkins, *Chem. Rev.*, 71:525-616(1971).
4. G.H.E. Nieuwdorp, C.L. de Ligny and N.G. van der Veen, *J. Chromatogr.*, 154:133-59(1978).
5. J.Green and S. Marcinkiewicz, *Chromatogr. Rev.* 5:58-160(1963).
6. L.S. Bark and K.F. Clarke, *J. Chromatogr.* 48:418-429 (1970).
7. L.E. Cook and F.M. Raushel, *J. Chromatogr.* 65:556-9(1972).

8. R.F. Rekker, *The Hydrophobic Fragmental Constant*, W.T. Nauta and R.F. Rekker, Elsevier Scientific Publishing Company, Amsterdam, 1977, p. 25.

9. M.C. Spanjer and C.L. de Ligny, *J. Chromatogr.*, 253:23-40(1982).

10. S. Ehrenson, R.T.C. Brownlee and R.W. Taft, *Prog. Phys. Org. Chem.*, 10:1-80(1973).

11. D.O. Rauls and J.K. Baker, *J. Med. Chem.*, 22:81-6(1979).

12. T. Fujita and T. Nishioka, *Prog. Phys. Org. Chem.*, 12:49-89(1976).

13. S.R. Bakalyar, M.P.T. Bradley and R. Honganen, *J. Chromatogr.*, 158:277-93(1978).

14. A. Leitold and G. Vigh, *J. Chromatogr.*, 257:384-6(1983).

15. A.L. Bacarella, E. Grunwald, H.P. Marshall and E.L. Purlee, *J. Org. Chem.*, 20:747-62(1955).

16. T. Hanai, K.C. Tran and J. Hubert, *J. Chromatogr.*, 239:385-95(1982).

17. G.H.E. Nieuwdorp, C.L. de Ligny and H.C. van Houwelingen, *J. Chem. Soc.*, Perkin Trans. II, 75:537-544(1979).

18. H. Seidel, C. Ritter, J. Fruwert and G. Geiseler, *Spectrochim. Acta*, 32A:705-708(1976).

19. D.E. Martire and R.E. Boehm, *J. Phys. Chem.*, 87:1045-62(1983).

20. H. Colin, A. Krstulovic, G. Guiochon and Z. Yun, *J. Chromatogr.*, 255:295-309(1983).

CHAPTER III

A COMPARISON OF SOME LINEAR SUBSTITUENT FREE ENERGY RELATIONSHIPS

ABSTRACT

The ability of three recently proposed linear substituent free energy relationships to describe data on substituent effects by regression analysis has been compared. As a reference, the Taft relationship was used, and 211 series of data were selected from literature that cannot be described satisfactorily by this equation To explore the scope of applicability of the most successful relationship, data were selected on the effects of ortho-, meta- and para substituents on chemical and physical properties, in rigid saturated systems, small π-systems and extended π-systems. The medium ranged from water to apolar organic solvents and even to the gas phase.

INTRODUCTION

A frequently used way of description of the influence of substituents on the properties of a parent molecule is the dual parameter equation introduced by Taft (1). However, in a number of cases this equation fails to explain the variance of data satisfactorily (1). Nieuwdorp et al. (2) therefore proposed an equation with three ρ parameters, which described substituent effects in a set of 76 series of data more precisely. This equation was originally applied to chemical reactivity and ionization equilibria of meta- and para monosubstituted benzene- derivatives (2). Nieuwdorp et al. (3) also proved its usefulness in the field of gas-chromatography. We used this equation to describe the influence of meta- and para substituents on the retention of substituted benzoic acids, phenols and anilines in liquid chromatography (4). In both cases (3,4) the Taft equation was slightly worse in explaining the variance of the data. This has prompted us to make a more elaborate investigation of the merits of Nieuwdorp's equation for those cases where the Taft equation does not give good results. We try to explore its scope of applicability by investigating, *inter alia*, data on the substituent effect in more extended π-systems and data on the substituent effect on physical properties. Further, we compare the merits of Nieuwdorp's equation with those of the equations of Swain and Lupton and of Yukawa and Tsuno. These latter models were recently revised by the original authors (5,6) and were claimed to give good results for a wide range of data on substituent effects.

THEORY

In order to describe the effect of substituents on chemical equilibrium constants and reaction rate constants, Hammett (7) has proposed the equation

$$\log K_X / K_H = \rho\sigma_{m,p} \qquad (1)$$

The symbol K denotes an equilibrium or reaction rate constant, the subscripts X and H refer to the substituted and the unsubstituted compound, respectively, ρ is a reaction-dependent parameter, and σ is a variable that depends on the nature of the substituent, and on its position: different σ values are required for *m*- and *p*- substituents.

This equation has contributed tremendously towards systematizing data on chemical reactions, but on the other hand it soon became clear that so simple an equation has only a limited range of application. To widen the scope of applications modifications of equation (1) were proposed, at first by Taft (1,8,9). His equation reads as follows:

$$\log K_X / K_H = \rho_I\sigma_I + \rho_R\sigma_R \qquad (2)$$

The subscripts I and R refer to the inductive and the resonance effect of a substituent, respectively. For a particular substituent only one σ_I value is given. This value holds for both *m*- and *p*-substituents in reactions of various types. However, the value of σ_R for a particular substituent depends on the reaction type. In a summarizing paper (1) four types of reactions, and thus four sets of σ_R values, are proposed:

- σ_R^- type reactions, where direct (through, exalted) resonance is possible between an electron-withdrawing substituent and an electron-donating reaction centre,
- σ_R^+ type reactions, where direct resonance is possible between an electron donating substituent and an electron - withdrawing reaction centre,
- σ_R *(BA)* type reactions, of substituted benzoic acids and related compounds,
- σ_R° reactions, where direct resonance is considered to be impossible.

Swain and Lupton (10) proposed an equation that contains two substituent constants and three parameters:

$$\log K_X / K_H = fF + rR + h \qquad (3)$$

In equation (3) F is a nonresonance or field constant and R a resonance constant. The symbols f, r and h represent the free parameters. Swain and Lupton's original proposal (10) of F and R values turned out to be unreliable. In a reinvestigation (5) they used a larger reference set of series, and a newly developed mathematical procedure to obtain strictly orthogonal F and R values (11).

We now turn to equations with three substituent constants and three parameters. The first equation of this kind is the *Linear Aromatic Substituent Reactivity Relationship* of Yukawa and Tsuno (12-14):

$$\log K_X / K_H = \rho\ (\sigma_{m,p} + r^+ \Delta \sigma_R^+ + r^- \Delta \sigma_R^-) \qquad (4)$$

where the last two terms account for direct resonance in σ_R^+ and σ_R^- reactions, respectively. Thus the Yukawa-Tsuno equation is an extension of the Hammett equation in which two more substituent constants ($\Delta\sigma_R^+$ and $\Delta\sigma_R^-$) are introduced, in addition to the constant $\sigma_{m,p}$ figuring in the Hammett equation. The Yukawa-Tsuno equation holds for all four reaction types mentioned above. However, it is not able to describe data on rigid saturated compounds, where substituent effects are proportional to σ_I. In later work (6) Yukawa and Tsuno reformulated equation (4) as the *Linear Substituent Free Energy Relationship*:

$$\log K_X / K_H = \rho_I \sigma_I + \rho_\pi^+ \sigma_\pi^+ + \rho_\pi^- \sigma_\pi^- \qquad (5)$$

where σ_I is the inductive substituent constant, σ_π^+ the resonance substituent constant for electron donating substituents (σ_π^+ is zero for electron - withdrawing substituents) and σ_π^- is the resonance substituent constant for electron - withdrawing substituents (σ_π^- is zero for electron-donating substituents).

Nieuwdorp et al. (2) investigated the merits of the equation:

$$\log K_X / K_H = \sum_i \rho_i \sigma_i \qquad (6)$$

in which no constraints are posed on the values of σ_i. It was established that equation (6) with only three terms, *i.e.*, in the form

$$\log K_X / K_H = \rho_I \sigma_I + \rho_R \sigma_R + \rho_E \sigma_E \qquad (7)$$

is able to describe data on $\sigma_R°$, $\sigma_R(BA)$, σ_R^- and σ_R^+ reactions and on reactions of rigid saturated compounds. In equation (7) the last term accounts for the exalted resonance, and σ_R is equivalent with Taft's $\sigma_R°$. The values of the substituent constants σ were estimated by applying an advanced statistical procedure (factor analysis with missing data (2,15)) to a large body of data, on chemical reactions of *m*- and *p*- substituted compounds (76 series of data). In this way the experimental errors in the data are averaged out.

As mentioned in the Introduction it is the purpose of this paper to compare the merits of equations (2), (3), (5) and (7).

SELECTION OF LITERATURE DATA AND RESULTS OF THE REGRESSION ANALYSES

An important source of the data to be considered is the investigation of the applicability of the Taft equation by Ehrenson, Brownlee and Taft (1). From their compilation we selected the data series that can not be described well by the Taft equation, according to the following criteria: the residual standard deviation SD of the regression is more than ten times the last significant digit, and at the same time more than 10% of the total standard deviation RMS of the data. In using the last criterion we follow Ehrenson *c.s.* who state that: " experience has shown that correlations of good precision are those for which SD/RMS $\equiv f \leqslant 0.1$, where SD is the root mean square of the deviations and RMS is the root mean square of the data. Evidence has been presented that this simple f measure of statistical precision is more trustworthy in measuring the precision of structure-reactivity correlations than is the more conventional correlation coefficient". Another criterion that we used for data selection is that a series must contain data on at least five substituents of three chemical types. Other important data sources are the investigation of the applicability of the Hammett equation by Exner (16) and the recent paper by Swain *et al.* (5). From these sources 211 series of data on chemical reactions and physical properties were selected that could not be described well by the Taft equation, according to the above-mentioned criteria. Of these series, 48 could be described satisfactorily by equation (7), proposed by Nieuwdorp et al. These 48 series are presented in Table I, and the corresponding $\log K_X / K_H$ data are given in Table II. The values of the substituent constants for the various models are shown in Table III. A compilation of f values calculated with equations (2), (3) (5), and (7) is given in Table IV. The detailed results of the regression analysis of the data by equation (7) are shown in Table V. In Table VI a classification is given of the 211 investigated series and of the 48 series that can be described satisfactorily by equation (7).

Table I Listing of the reactions and physical properties of the 48 series, for which equation (7) holds good

Series	Reaction or Physical Property	Ref.
1	Ionization of ortho-substituted benzoic acids in 43,5 w/v % dioxane-water at 25°C.	17
2	Ionization of 2-substituted pyridines in water at 25°C.	18
3	Ionization of 3-substituted pyridines in water at 25°C.	19
4	Ionization of 4-substituted pyridines in water at 25°C.	19
5	Ionization of 3-substituted quinolines in water at 25°C.	19
6	Ionization of 7-substituted isoquinolines in water at 25°C.	19
7	Ionization of *m*-substituted benzylidenemeldrumacids in 99 v/v % methanol-water at 25°C.	19
8	Ionization of *m*-substituted benzylidene indandiones in 99 v/v % methanol-water at 25°C.	20
9	Ionization of *m*-substituted naphthylammonium cations in water at 25°C.	21
10	Ionization of 4-substituted quinuclidinium ions in water at 25°C.	22
11-17	Ionization of 4β, 5β, 7β, 3α, 4α, 6α and 7α-substituted naphthoic acids in 50 v/v % ethanol-water at 25°C.	23
18	Ionization of 4-substituted diphenylamines in DMSO at 25°C.	24
19	Radical polymerization of *p*-substituted styrenes with methylmethacrylate at 25°C.	25
20	Halogenation of substituted benzenes in acetic acid at 25°C.	26
21	Menschutkin reaction of 4-substituted pyridines with ethyl iodide in nitrobenzene at 25°C.	27
22	Reaction of *p*-substituted benzaldehydes with semicarbazide in 25 v/v % ethanol at 25°C.	28
23	Reaction of piperidine with *p*-substituted, 2-nitrochlorobenzene in benzene at 45°C.	29
24	Solvolysis of triphenylcarbonchloride in 60 v/v % ethylether in ethanol at 60°C.	30
25,26	NMR shifts of the amino protons in *o*- and *p*-substituted anilines respectively, in DMSO at 25°C.	31
27,28	NMR shifts of the hydroxyl protons in *m*- and *p*-substituted phenols in DMSO at 25°C.	32

Series	Reaction or Physical Property	Ref.
29	NMR shifts of the nuclear proton at the para position in *o*-substituted duroles in *n*-hexane at 25°C.	33
30	NMR shifts of the nuclear proton at the para position in substituted benzenes in *n*-hexane at 25°C.	34
31	NMR shifts of the hydroxyl protons in *p*-substituted phenols in DMSO at 40°C.	35
32-34	NMR shifts of the hydroxyl protons in 4-substituted 2-cresols, 2,6 xylenols and 3-cresols respectively in DMSO at 25°C.	36
35	NMR shifts of fluorine in 6β-substituted 2-fluoronaphthalenes in DMF at 25°C.	37
36	NMR shifts of the carboxyl carbons in *p*-substituted benzoic acid anions in 80 v/v % $DMSO\text{-}d_6$ at pH > 13.0 at 25°C.	38
37	IR shifts of the carbonyl stretching frequency in *o*-substituted benzoic acids in chloroform at 29°C.	39
38-40	IR shifts of the hydroxyl stretching frequency in *o*- and *m*-substituted and of the carbonyl stretching frequency in *m*-substituted benzoic acids respectively, in carbon tetrachloride at 29°C.	40
41,42	IR shifts of the hydroxyl stretching frequency in Schiff bases, *m*- and *p*-substituted in ring B, in carbon tetrachloride at 25°C.	41
43,44	Shifts in the wavelengths in the visible region of the spectra of 2- and 1-substituted azulenes in cylcohexane at 25°C.	42
45	Ionization potentials of *p*-substituted benzyl radicals in the gas phase.	43
46	Ionization potentials of *p*-substituted acetophenones in the gas phase.	44
47,48	Changes in the Kovats retention indices of *m*- and *p*-substituted chlorobenzenes in gas-liquid chromatography.	45

Table II $\log K_X / K_H$ values for the

					Substituent			
Series	CH_3	C_2H_5	i-C_3H_7	t-C_4H_9	C_6H_5	F	Cl	Br
1	-0.08						0.97	1.00
2	0.05	0.05						
3	-0.46				0.41	2.24	2.40	2.36
4	-0.82				-0.14		1.38	1.46
5	-0.24					2.17	2.27	2.17
6	-0.44					0.65	1.05	1.03
7	-0.10						0.90	1.00
8	-0.10						1.04	1.10
9	-0.13				0.44	1.01	1.09	1.07
10	0.11	0.03	-0.08		0.94		2.51	2.65
11	-0.13						0.40	0.38
12								0.27
13	-0.07					0.22	0.27	0.29
14	-0.08						0.46	0.52
15	-0.21					0.08	0.39	0.45
16	-0.07						0.25	0.28
17	-0.11							0.11
18	-0.51						1.11	
19	0.056						0.046	0.067
20	3.30			2.80	3.40	0.80	-0.40	-0.60
21	0.32		0.28		0.04			
22	-0.31						0.18	
23	-0.82				0.32		0.79	0.98
24	0.61					-0.12	-0.49	-0.55
25	-10						24	24
26	-6				21		21	21
27	-0.14					0.52	0.52	
28	-0.28				0.12	0.04	0.30	0.34
29						14	4	1
30						22	13	6
31	-0.20	-0.23	-0.23	-0.20	0.32	0.10	0.37	0.39
32	-0.263						0.352	0.387
33	-0.25							0.346
34	-0.257						0.326	0.355
35								-1.23
36	-0.10					-1.40	-1.40	
37						- 88	- 83	- 83
38	-2					- 11	- 13	- 12
39	0					- 4	- 5	- 5
40	-2					4	4	4
41	-1.3					9.6	11.8	
42	-3.5					5.2	6.5	8.5
43	-18						- 29	- 23
44	28							25
45	-0.30		-0.34			0.02	0.19	
46							9.47	
47							156.5	
48							162.0	

48 series listed in Table I

I	OCH$_3$	COCH$_3$	COOCH$_3$	CN	NO$_2$	OH	NH$_2$	CF$_3$
0.87	-0.09				1.66			
	-0.19		-0.38				0.11	
1.92	0.40	1.99	2.12	3.86	4.03	0.42	-0.83	
1.20	-1.37	1.70	1.72	3.35	3.82	-2.12	-3.91	
					3.85	0.65	-0.01	
					2.52	-0.58	-1.71	1.72
	0.10				1.35	0.00		
	0.10				1.55	0.05		
0.99	0.37	1.03	1.04	1.86	2.14	0.41	-0.04	1.10
2.34	1.81		1.70	3.04	3.48	1.68		
0.34	-0.02			0.86	0.92	-0.01	-0.13	
	-0.01			0.56	0.61	-0.04		
0.27	-0.01			0.54	0.56	-0.14	-0.23	
				0.90	0.93	0.09		
	-0.55			1.20	1.25	-0.79	-1.09	
	-0.09			0.50	0.62	-0.12		
	-0.12			0.47	0.55	-0.15		
	-0.78				6.77		-1.76	
0.107	0.230			0.321				
	9.80					11.80		
	0.42	-0.63		-1.47			1.21	
	-0.66			1.08	-0.80			
1.08	-1.60		2.96	3.77	4.61		-3.00	2.72
-0.47	1.95				-1.96			
	- 12				152			
	- 18				110			
					1.02	-0.30		
0.34	-0.56				1.66	-0.82		
2					- 16	29	42	
3					- 33	43	54	
0.37	-0.43	1.05	1.01	1.32	1.70	-0.68		1.00
	-0.468	1.047	1.044	1.327	1.779	-0.746	-1.050	
	-0.455	0.975	0.986	1.281	1.710	-0.715		
	-0.447	0.908	0.923	1.276	1.575	-0.710	-0.942	
	3.72			-5.45	-6.54	4.87	6.96	
	-0.60				-2.60			
- 84					- 77		-123	
- 10	- 10				- 20			
	- 1				- 10			
	- 3				8			
	1.7				18.6	-1.2		
	-7.5				22.5	-8.7		
- 20	-57			55				
	104		- 30		- 24	- 48		
	-0.94			0.60				
	8.62				10.07	8.70	8.17	
				140.3	143.8	181.2	213.3	
				152.1	143.2	181.4	217.3	

Table III Values of the substituent constants in equations (2), (3), (5) and (7)

Equation	Symbol	Substituent CH_3	C_2H_5	i-C_3H_7	t-C_4H_9	C_6H_5	F	Cl	Br	I
2	σ_I	-0.04				0.10	0.50	0.46	0.44	0.39
	σ_R	-0.11				-0.11	-0.34	-0.23	-0.19	-0.16
	σ_R^+	-0.25				-0.30	-0.57	-0.36	-0.30	-0.25
	σ_R^-	-0.11				0.04	-0.45	-0.23	-0.19	-0.11
	$\sigma_R(BA)$	-0.11				-0.11	-0.45	-0.23	-0.19	-0.16
3	F	-0.01	-0.02		-0.11	0.25	0.74	0.72	0.72	0.65
	R	-0.41	-0.44		-0.29	-0.37	-0.60	-0.24	-0.18	-0.12
5	σ_I	-0.045	-0.045	-0.053	-0.060	0.11	0.363	0.348	0.337	0.289
	σ_π^+	-0.078	-0.069	-0.051	-0.034	-0.09	-0.118	-0.070	-0.061	-0.068
	σ_π^-	0.00	0.00	0.00	0.00	0.12	0.00	0.00	0.00	0.00
7	σ_I	0.0004	0	-0.0091	-0.0193	0.1527	0.4627	0.4498	0.4495	0.4136
	σ_R	-0.1133	-0.1070	-0.0929	-0.1002	-0.0861	-0.2862	-0.1772	-0.1615	-0.1322
	σ_E	-0.0395	-0.0500	-0.0666	0.0115	-0.1754	0.3837	0.2434	0.2251	0.1573

Table III continued

Equation	Symbol	Substituent OCH_3	$COCH_3$	$COOCH_3$	CN	NO_2	OH	NH_2	CF_3
2	σ_I	0.27	0.28	0.30	0.56	0.65		0.12	0.45
	σ_R	-0.45	0.16	0.14	0.13	0.15		-0.48	0.08
	σ_R^+	-1.02	0.16	0.14	0.13	0.15		-1.61	0.08
	σ_R^-	-0.45	0.47	0.34	0.33	0.46		-0.48	0.17
	$\sigma_R(BA)$	-0.61	0.16	0.14	0.13	0.15		-0.82	0.08
3	F	0.54	0.50	0.70	0.90	1.00	0.46	0.38	0.64
	R	-1.68	0.90	-0.04	0.71	1.00	-1.89	-2.52	0.76
5	σ_I	0.185	0.214	0.23	0.42	0.46	0.19	0.10	0.34
	σ_π^+	-0.281	0.00	0.00	0.00	0.00	-0.34	-0.46	0.00
	σ_π^-	0.00	0.266	0.22	0.226	0.332	0.00	0.00	0.143
7	σ_I	0.2793	0.2714	0.2824	0.5656	0.6126	0.2743	0.1884	0.4740
	σ_R	-0.4116	0.1390	0.1342	0.0847	0.1520	-0.4746	-0.5799	0.0025
	σ_E	-0.0326	-0.4750	-0.2260	-0.0948	-0.3020	-0.0949	-0.4680	0.2695

Table IV Values of f for the 48 series listed in Table I, calculated with equations (2), (3), (5) and (7)

Series No.	Equation (2)	(3)	(5)	(7)
1	0.13	0.13	0.16	0.09
2	0.13	0.24	0.34	0.04
3	*	0.10	0.07	0.07
4	*	0.16	0.09	0.06
5	*	0.03	0.05	0.03
6	*	0.10	0.10	0.08
7	*	0.17	0.14	0.07
8	*	0.16	0.13	0.06
9	*	0.11	0.10	0.08
10	*	*	0.09	0.04
11	*	0.17	0.10	0.09
12	*	0.09	0.07	0.05
13	*	0.08	0.13	0.09
14	*	0.09	0.10	0.07
15	*	0.10	0.15	0.10
16	*	0.05	0.08	0.06
17	*	0.19	0.09	0.09
18	0.11	0.36	0.04	0.08
19	0.69	0.66	0.16	0.05
20	*	0.15	0.08	0.03
21	0.11	*	0.13	0.07
22	1.17	1.20	0.03	0.04
23	0.11	0.23	0.10	0.06
24	0.14	0.19	0.12	0.08
25	0.16	0.40	0.02	0.04
26	0.12	0.36	0.09	0.09
27	*	0.16	0.13	0.07
28	*	0.21	0.13	0.06
29	*	0.11	0.16	0.10
30	*	0.23	0.16	0.07
31	*	*	0.08	0.10
32	*	0.19	0.07	0.05
33	*	0.18	0.07	0.06
34	*	0.18	0.08	0.04
35	*	0.05	0.05	0.02
36	0.12	0.11	0.14	0.10
37	0.13	0.08	0.10	0.08
38	0.19	0.15	0.13	0.09
39	0.11	0.10	0.15	0.08
40	0.17	0.24	0.09	0.08
41	*	0.11	0.15	0.09
42	*	0.08	0.10	0.08
43	0.13	0.54	0.11	0.08
44	*	1.14	0.09	0.10
45	0.25	*	0.12	0.07
46	*	0.05	0.05	0.02
47	*	0.04	0.07	0.05
48	*	0.06	0.09	0.07

* The value of f cannot be calculated because of the lacking substituent constants

Table V Values of the ρ_I, ρ_R, and ρ_E parameters in equation (7) for the 48 series listed in Table I, and the standard deviation of the regressions

Series	ρ_I	ρ_R	ρ_E	SD
1	2.53±0.10	1.78±0.21	0.55±0.22	0.09
2	- 1.33±0.02	- 0.42±0.01	- 0.24±0.02	0.01
3	6.18±0.12	3.04±0.17	0.41±0.16	0.15
4	5.02±0.12	7.03±0.16	1.50±0.17	0.14
5	5.77±0.07	1.93±0.09	- 0.03±0.09	0.06
6	3.42±0.11	3.53±0.15	0.41±0.15	0.11
7	2.22±0.06	1.16±0.08	0.61±0.11	0.05
8	2.52±0.07	1.28±0.09	0.62±0.13	0.06
9	2.95±0.07	1.10±0.10	- 0.29±0.09	0.09
10	5.63±0.07	- 0.38±0.12	- 0.21±0.15	0.08
11	1.26±0.04	0.83±0.05	- 0.21±0.07	0.04
12	0.85±0.02	0.60±0.03	- 0.04±0.06	0.02
13	0.82±0.02	0.64±0.03	0.07±0.04	0.03
14	1.39±0.04	0.64±0.08	- 0.06±0.10	0.04
15	1.65±0.07	2.40±0.09	0.14±0.10	0.08
16	0.82±0.02	0.73±0.03	0.05±0.05	0.02
17	0.64±0.03	0.75±0.04	- 0.20±0.08	0.03
18	7.51±0.35	7.75±0.40	- 3.11±0.50	0.27
19	0.47±0.01	- 0.19±0.02	- 0.78±0.03	0.01
20	- 5.29±0.47	-26.10±0.43	-10.91±0.47	0.15
21	- 2.28±0.09	- 2.40±0.08	- 0.63±0.10	0.05
22	1.26±0.04	2.41±0.04	0.18±0.07	0.03
23	5.47±0.12	7.68±0.20	- 0.77±0.20	0.15
24	- 2.53±0.09	- 6.18±0.21	- 1.80±0.19	0.09
25	161 ± 3	145 ± 6	- 103 ± 7	3
26	121 ± 5	123 ±10	- 58 ±10	4
27	2.01±0.05	2.70±0.06	- 0.32±0.06	0.05
28	1.83±0.04	2.46±0.04	- 0.24±0.05	0.04
29	- 13 ± 2	- 70 ± 3	- 7 ± 3	2
30	- 24 ± 2	- 106 ± 3	7 ± 3	2
31	2.01±0.06	2.50±0.11	- 0.34±0.09	0.08
32	2.01±0.04	2.70±0.05	- 0.32±0.06	0.05
33	1.90±0.06	2.57±0.09	- 0.32±0.11	0.06
34	1.83±0.03	2.46±0.04	- 0.24±0.05	0.04
35	- 7.49±0.12	-14.18±0.14	- 0.48±0.22	0.11
36	- 3.79±0.19	- 0.98±0.38	0.35±0.35	0.16
37	- 148± 8	137 ±11	33 ±10	7
38	- 29± 1	5 ± 3	9 ± 2	1
39	- 14± 0	- 7 ± 1	1 ± 1	0
40	12± 0	15 ± 1	6 ± 1	0
41	28.6±1.1	16.6 ±1.5	4.5 ±1.8	0.9
42	29.2±1.0	35.6 ±1.5	3.2 ±1.6	0.9
43	54± 4	183 ± 7	- 89 ± 9	3
44	- 19± 6	- 259 ±11	- 23 ±10	5
45	0.84±0.05	2.73±0.08	1.11±0.09	0.04
46	-18.24±0.25	- 7.98±0.26	- 0.31±0.36	0.19
47	273± 8	- 232 ±11	- 47 ±15	8
48	283±14	- 235 ±18	- 44 ±24	13

Table VI Classification of the investigated series, and of the series that can be described satisfactorily by equation (7)

Classification of the series of data	Number of investigated series	Percentage that can be described satisfactorily by equation (7)
Effects of *m*- and *p*-substituents on chemical properties in rigid saturated systems and small π-systems	67	13 %
Effects of *m*- and *p*-substituents on chemical properties in extended π-systems	32	34 %
Effects of *m*- and *p*-substituents on NMR chemical shifts, on IR and visible absorption frequencies, and on ionization potentials	48	44 %
Effects of *o*-substituents	59	14 %
Miscellaneous	5	40 %

DISCUSSION

1. *General observations*

In Table III some substituent constants are missing for Taft's model. This implies that f can not be calculated with Taft's model for series that contain data on the effects of these substituents, unless these data are omitted. However, omission of data in the application of Taft's model only would give a favorable bias to this model, in a comparison with the other models. On the other hand, omission of data for *all* models, would not do justice to the ability of the Swain-Lupton, Yukawa-Tsuno and Nieuwdorp models to cope with the effects of these substituents. Therefore, we decided not to omit any data. Consequently, for the Taft equation many f values are missing in Table IV.

For Swain-Lupton's model the constants for the isopropyl substituent are not available. This substituent occurs in four series in Table II. For these four series, the f value is thus missing in Table IV.

As can be seen from Table IV the f values for 48 of the investigated 211 series are lower than or equal to 0.10, when equation (7) is used. The decreases of f, compared with the values for the Taft equation vary from rather drastic ones like the decrease from 1.17 to 0.04 in series 22, to slight decreases like that from 0.11 to 0.08 in series 18. With the other models a smaller number of series fulfills the criterion $f \leqslant 0.10$, *viz.* 15 with Swain-Lupton's and 33 with Yukawa-Tsuno's model. The remaining 163 series could not be described satisfactorily by any of the models discussed in this paper.

2. *The Taft equation*

It must be taken in mind that the relative success of the Taft equation is due to the introduction of four sets of ρ_R constants. An advantage of the use of four σ_R scales might be that an outstanding result of regression analysis with one of the σ_R scales might offer an opportunity to distinguish reactions in four distinct classes, since each scale is related with a certain reaction type. However, this is not always possible. For example, it is clear that series 1 belongs to the σ_R *(BA)* type, and series 23 to the σ_R^+ type. Yet, the best results are obtained with the σ_R^+ and σ_R° scales respectively. The same holds for physical properties. Series 37 - 40 are of the σ_R *(BA)* type but best results are obtained with the σ_R^+, σ_R^+, σ_R^-, and σ_R^+ scale, respectively.

3. The Swain-Lupton equation

In this equation the effects of the substituents are also characterized by two substituent constants, *viz.* F and R, a rather complete set of which is available. In addition to the corresponding regression coefficients f and r, the equation contains an adjustable constant h. Nevertheless the ability of the Swain-Lupton equation to describe data on substituent effects is not better than that of the Taft equation. In 9 cases the f value for the Taft equation is the smaller one; in 5 cases the Swain-Lupton equation is better. However, it also describes well 13 out of the 31 series for which the Taft equation cannot be used for lack of data on substituent constants. The only advantage of the Swain-Lupton equation over the Taft equation thus is its ability to describe the effect of the hydroxyl substituent, and of the ethyl and *t*-butyl substituents.

4. The Yukawa-Tsuno equation

In this equation the effects of the substituents are characterized by three substituent constants, *viz.*, σ_I, σ_π^+ and σ_π^-. Due to the presence of three corresponding ρ parameters its ability to describe data on substituent effects is clearly better than that of the Swain-Lupton equation: The Yukawa-Tsuno equation describes 33 series with $f \leq 0.10$, the Swain-Lupton equation only 15. It is also better than the Taft equation: in 12 cases the f value for the Yukawa-Tsuno equation is the smaller one, in 5 cases the f value for the Taft equation is lower.

5. The equation, proposed by Nieuwdorp et al.

This equation is of the same mathematical form as the Yukawa-Tsuno equation. The effects of the substituents are characterized by three substituent constants, σ_I, σ_R and σ_E, in combination with three ρ parameters. Yet, its ability to describe data on substituent effects is much better than that of the Yukawa-Tsuno equation, as for all series in Table IV it holds that $f \leqslant 0.10$. This has two causes: (i) the values of the substituent constants in Nieuwdorp's equation are not subject to any constraint, in contrast with the substituent constants in the Yukawa-Tsuno equation, and (ii) the values of the substituent constants are based on 76 series of accurate data, rather than on a few reference series. In this way the experimental error is averaged out. The scope of applicability of Nieuwdorp's equation can be judged from the last column of Table VI. It appears to hold as well for substituent effects in extended π-systems as for substituent effects in rigid saturated systems and small π-systems, and as well for substituent effects on physical properties as for effects on chemical properties. However, it seems to be less suited for the description of the effects of ortho-substituents. Table I shows that Nieuwdorp's equation holds in solvents that range in polarity from water to hexane or carbon tetrachloride and that it even holds in the gas phase. In previous work it has been shown that this equation holds not only for substituent effects on chemical equilibria, but also for substituent effects on phase equilibria as occur in gas chromatography (3), reversed phase liquid chromatography and liquid-liquid partition (4).

CONCLUSIONS

The ability of some recently proposed linear substituent free energy relationships to describe data on substituent effects increases in the following order: Taft ≈ Swain and Lupton < Yukawa and Tsuno < Nieuwdorp et al..The scope of applicability of the latter relationship comprises chemical properties, physical properties (like proton, fluorine and carbon-13 NMR shifts, absorption frequencies in the visible and infrared region of the spectrum and ionization potentials), and phase equilibria. The medium can vary from water to apolar organic solvents or even to the gas phase.

Its applicability ranges from rigid saturated systems and small π-systems to more extended π-systems. However, effects of ortho-substituents are not well described.

REFERENCES

1. S. Ehrenson, R.T.C. Brownlee and R.W. Taft, *Progr. Phys. Org. Chem.*, 1958, 9, 287.
2. G.H.E. Nieuwdorp, C.L. de Ligny and H.C. van Houwelingen, *J. Chem. Soc., Perkin Trans.* 2, 1979 ,537.
3. G.H.E. Nieuwdorp, C.L. de Ligny and N.G. van der Veen, *J. Chromatography*, 1978, 154, 133.
4. M.C. Spanjer and C.L. de Ligny, *J. Chromatogr.Sci.*, submitted for publication.
5. C.G. Swain, S.H. Unger, N.R. Rosenquist and M.S. Swain, *J. Am. Chem. Soc.*, 1983, 105, 492.
6. M. Sawada, M. Ichihara, Y. Yukawa, T. Nakachi and Y. Tsuno, *Bull. Chem. Soc. Jpn.*, 1980, 53, 2055.
7. L.P. Hammett, *Physical Organic Chemistry*, McGraw-Hill, New York, 1940, p. 184.
8. R.W. Taft and I.C. Lewis, *J. Am. Chem. Soc.*, 1958, 80, 2436.
9. R.W. Taft, C.N. Deno and P.S. Skell, *Ann. Rev. Phys. Chem.*, 1958, 9, 287.
10. C.G. Swain and E.C. Lupton, Jr., *J. Am. Chem. Soc.* , 1968, 90, 4328.
11. P.F. Strong, C.G. Swain and M.S. Swain, *J. Chem. Inf. Comput. Sci.*, 1979, 19, 13.
12. Y. Yukawa and Y. Tsuno, *Bull. Chem. Soc. Jpn.*, 1959, 32, 971.
13. Y. Yukawa, Y. Tsuno and M. Sawada, *Bull. Chem. Soc. Jpn.*, 1966, 39, 2274.
14. Idem, *ibidem*, 1972, 45, 1198.
15. C.L. de Ligny, G.H.E. Nieuwdorp, W.K. Brederode, W.E. Hammers and J.C. van Houwelingen, *Technometrics*, 1981, 23, 91.
16. O. Exner, Advances in Linear Free Energy Relationships, eds. N.B. Chapman and B. Shorter, Plenum, New York, 1972, p.1.
17. J.H. Elliot and M. Kilpatrick, *J. Phys. Chem.*, 1941, 45, 485.
18. M. Charton, *J. Am. Chem. Soc.*, 1964, 86, 2033.
19. D.D. Perrin , Dissociation constants of organic bases in aqueous solution, Butterworth, London, 1965.
20. P. Schuster, O.E. Polansky and F. Wessely, *Tetrahedron Letters*, Suppl.8, Part II, 1966, 463.
21. A. Bryson, *J. Am. Chem. Soc.*, 1960, 82, 4862 .
22. C.A. Grob and M.E. Schlageter, *Helv. Chim. Acta*, 1976, 59, 264.

23. P.R. Wells and W. Acdock, *Austr. J. Chem.*, 1965, 18, 1365.

24. D. Dolman and R. Stewart, *Can. J. Chem.*, 1967, 45, 911.

25. C. Walling, E.R. Briggs, K.B. Wolfstirn and F.R. Mayo, *J. Am. Chem. Soc.*, 1948, 70, 1537.

26. P.B.D. de la Mare, *J. Chem. Soc.*, 1954, 4450.

27. A. Fischer, W.J. Galloway and J. Vaughan, *J. Chem. Soc.*, 1964, 3596.

28. R. Wolfender and W.P. Jencks, *J. Am. Chem. Soc.*, 1961, 83, 2763.

29. W. Greizerstein, R.A. Bonelli and J.A. Brieux, *J. Am. Chem. Soc.*, 1962, 84, 1026.

30. A.C. Nixon and G.E.K. Branch, *J. Am. Chem. Soc.*, 1936, 58, 492.

31. B.M. Lynch, B.C. MacDonald and J.G.K. Webb, *Tetrahedron Letters*, 1968, 24, 3595.

32. J.G. Traynham and G.A. Knesel, *J. Org. Chem.*, 1966, 31, 3350.

33. P. Diehl and G. Svegliado, *Helv. Chim. Acta*, 1963, 46, 461.

34. P. Diehl, *Helv. Chim. Acta*, 1961, 44, 829.

35. M.T. Tribble and J.G. Traynham, *J. Am. Chem. Soc.*, 1969, 91, 379.

36. M. Fujio, M. Mishima, Y. Tsuno, Y. Yukawa and Y. Takai, *Bull. Chem. Soc. Jpn.*, 1975, 48, 2127.

37. W. Adcock and M.J.S. Dewar, *J. Am. Chem. Soc.*, 1967, 89, 379.

38. Y. Kosugi and Y. Furuya, *Tetrahedron Letters*, 1980, 36, 2741.

39. D.G. O'Sullivan and P.W. Sadler, *J. Chem. Soc.*, 1957, 2839.

40. C.J.W. Brooks, G. Eglinton and J.F. Morman, *J. Chem. Soc.*, 1961, 106.

41. A.W. Baker and A.T. Schulgin, *J. Am. Chem. Soc.* 1959, 81, 1523.

42. R.N. McDonald, J.M. Richmond, J.R. Curtis, H.E. Petty and T.L. Hoskins, *J. Org. Chem.*, 1976, 41, 1811.

43. A.G. Harrison, P. Kebarle and F.P. Lossing, *J. Am. Chem. Soc.*, 1961, 83, 777.

44. A. Foffani, S. Pignataro, B. Cantone and F. Grasso, *Z. Phys. Chem.*, 1964, 42, 228.

45. S.D. West and R.C. Hall, *J. Chromatogr. Sci.*, 1975, 13, 5.

CHAPTER IV

SIMULTANEOUS DESCRIPTION OF THE INFLUENCE OF SOLVENT, REACTION TYPE AND SUBSTITUENT ON EQUILIBRIUM CONSTANTS BY MEANS OF THREE-MODE FACTOR ANALYSIS

ABSTRACT

The equations, proposed by Hammett, Taft and Nieuwdorp et al., respectively, for the simultaneous description of the influence of reaction type and substituent on equilibrium and reaction rate constants are discussed. The latter equation represents an example of factor analysis. This mathematical-statistical technique has also been applied to describe simultaneously the influence of solvent and substituent and the influence of solvent and reaction type. It is thus a logical step to classify equilibrium and reaction rate constants with respect to three modes: solvent, reaction type and substituent and to try to describe the influence of these three variables by three-mode factor analysis. Two examples of the application of this technique to literature data are given. The first one concerns data on ionization constants for 15 series of substituted compounds in 3 solvents. The second example concerns data on phase equilibrium constants of 6 series of substituted compounds in 9 two-phase systems. The two-phase systems comprise gas-liquid as well as liquid-liquid systems. The precision of the fit of the observations and the precision of the prediction of the missing data are discussed. For the first example the predictions of the 237 missing data are tabulated. Among them are 90 data that cannot be predicted at all by the Hammett, Taft or Nieuwdorp equations (*viz.*, for reactions on which no measurements at all are available in a particular solvent). The standard deviation of the prediction of the latter data by three-mode factor analysis ranges from 0.1 to 0.2. Further, it is shown that the number of parameters that is required to fit the observations by three-mode factor analysis is less that the number of parameters in the corresponding regression analysis model, *viz.*, the Taft model.

INTRODUCTION AND THEORY

The description of the influence of the solvent on equilibrium and reaction rate constants is an important problem in chemistry, especially in physical organic chemistry. A satisfactory solution of this problem has not yet been obtained. In this paper we present a new approach, and illustrate its merits by applying it to data on chemical equilibrium constants of 15 series of substituted aromatic compounds in three solvents, and to data on phase equilibrium constants of six series of substituted aromatic compounds in nine widely different two-phase systems.

Our approach is an extension of the work by Nieuwdorp, De Ligny and Van Houwelingen[1] on the simultaneous description of the influence of the reaction type and substituents on equilibrium and reaction rate constants. It leads to the simultaneous description of the influence of solvent, reaction type and substituents. Nieuwdorp *et al.*'s study, on its turn, is an extension of the well-known Hammett[2] and Taft[3] approach.

The Hammett relationship is given in equation (1).

$$\log(K_X/K_H) = \rho\sigma_{m,p} \tag{1}$$

The subscripts X and H refer to the substituted and the unsubstituted compounds, respectively, ρ is a parameter that depends on the reaction type and σ is a variable that depends on the substituent and its position (*i.e.*, whether it is *m* or *p*). It was soon recognized that the Hammett relation could describe substituent effects in only a relatively small area of reaction types. To widen the scope of applications Taft proposed equation (2).

$$\log(K_X/K_H) = \rho_I\sigma_I + \rho_R\sigma_R \tag{2}$$

The subscripts *I* and *R* refer to the inductive effect and the resonance effect of a substituent, respectively.

For each substituent only one σ_I value is given(independent of its position), but the σ_R value depends on the type of reaction. In a review by Ehrenson *et al.*[4] four types of reactions, and thus four sets of σ_R values, were proposed: the well-known σ_R^0 , σ_R(benzoic acids), σ_R^- and σ_R^+ sets.

Nieuwdorp *et al.* showed that data on the four reaction types that correspond with the above mentioned sets of σ_R values can be described with universal σ values when a third term is added to the right hand side of equation (2).

$$\log(K_X/K_H) = \rho_I\sigma_I + \rho_R\sigma_R + \rho_E\sigma_E \tag{3}$$

In this equation each substituent has only one, universal, σ_I , σ_R and

σ_E value, that is independent of its position. The subscript E refers to the direct (through, exalted) resonance effect of a substituent on the reaction centre. As in the Taft equation, the ρ parameters have different values for *m*- and *p*-substituted series. The wide scope of applicability of equation (3) has been demonstrated recently[5].

Further, Nieuwdorp *et al.* stressed the point that, in principle, in equation (1) - (3) both σ and ρ are adjustable parameters. Previously, in the application of equations (1) and (2) reference reaction series were chosen to define a scale of σ values. *E.g.*, the σ values in the Hammett equation (1) are defined by the ionization constants of benzoic acids in water at 25°C, for which, by definition, $\rho \equiv 1$. Thus, the σ values in equation(1), and the σ_I and σ_R values in equation(2), are considered to be known *a priori*, *i.e.*, σ is considered as a variable instead as a parameter, and equations (1) and (2) are applied in regression analysis. Nieuwdorp *et al.* on the other hand, considering both σ and ρ as adjustable parameters, determined their optimal values by factor analysis of 570 data on 76 reactions and 17 substituents. To do so, a procedure had to be developed for the application of factor analysis on data sets in which a large fraction of the data is missing[7]. (Of course, once optimal values of σ_I , σ_R and σ_E have been calculated, they can be used in conjunction with equation (3) for regression analysis of other data).

In the language of factor analysis, equation (3) can be reformulated as (4).

$$y_{r',s} = \sum_{l=1}^{w} R_{r',l}\, S_{s,l} \qquad (4)$$

The symbol y stands for $\log(K_X/K_H)$, r' and s denote the reaction and

the substituent, respectively, and R and S are adjustable parameters, dependent on the reaction and the substituent, respectively.
It is the purpose of factor analysis to give a precise description of the data y with a small number w of terms. For Nieuwdorp *et al.*'s data, three terms were sufficient to reduce the standard deviation to 0.03.

In equation (4) reactions of different type (*e.g.*, the ionization of phenols and of benzoic acids in water) and reactions of the same type in different solvents (*e.g.*, the ionization of phenols in water and in 50% ethanol) are all considered as "different" reactions. In other words, r' in equation (4) is a combination mode of reaction type and solvent. If we denote the reaction type by r and the solvent (or, for a phase equilibrium constant, the two-phase system) by p, equation (4) can be reformulated as (5). Then, it is a logical

$$y_{(p,r),s} = \sum_{l=1}^{w} R_{(p,r),l}\, S_{s,l} \tag{5}$$

step to classifiy the data y in three modes (with respect to solvent, reaction type and substituent, respectively) and to try to describe them by three-mode factor analysis. Extension of (5) to three modes is formulated in (6). The symbol P represents a

$$y_{p,r,s} = \sum_{j=1}^{u} \sum_{k=1}^{v} \sum_{l=1}^{w} c_{j,k,l}\, P_{p,j}\, R_{r,k}\, S_{s,l} \tag{6}$$

solvent - dependent paremeter and c represents the elements of the so-called three-mode core matrix of scaling constants. As far as we know, equation (6) has only been applied in the field of the social sciences, and for the case where data exist for each combination of p, r and s[8]. Two of the present authors have recently devised a method to estimate $P_{p,j}$, $R_{r,k}$ and $S_{s,l}$ for the case of missing data[9], and the present investigation is one of the first applications of this method in the field of the natural sciences.

We believe that it is obvious to try the merits of equation (6) for the simultaneous description of the influence of solvent, reaction type and substituent on equilibria and reaction rates, as 1) the influence of reaction type and substituent on $y_{r,s}$ can be described simultaneously by two-mode factor analysis, *i.e.* by equation (5), with three factors[1], and 2) the influence of solvent and substituent on $y_{p,s}$ can also be described simultaneously by two-mode factor analysis, *i.e.*, by equation(7). This was demonstrated by

$$y_{p,s} = \sum_{j=1}^{u} P_{p,j} \; S_{s,j} \tag{7}$$

Weiner[10] for data on the ionization of substituted benzoic acids in a variety of solvents; four factors were required to reproduce the data within the experimental precision of 0.01 and 3) The combined influence of solvent and reaction type on $y_{p,r}$ has also been described by two-mode factor analysis, *i.e.*, by equation (8), for data on the

$$y_{p,r} = \sum_{j=1}^{u} P_{p,j} \; R_{r,j} \tag{8}$$

solvent dependence of the NMR chemical shift of ^{1}H, ^{13}C and ^{29}Si nuclei in 36 solvents[11]. Two factors were sufficient to reproduce the data within experimental error. Thus, it is logical to combine equations (5), (7) and (8) to equation (6). It may be expected that not more than four factors for each mode are required to describe adequately the influence of solvent , reaction type and substituent.

It should be realized that the same data set (*i.e.*, data on substituent effects on a number of reactions in a number of solvents) may be either described by three-mode factor analysis, equation (6), or by two-mode factor analysis, equation(5), or, if optimal values of the parameters $S_{s,l}$ have been estimated from other data, by regression analysis. In regression analysis, equation(5) is also used, but now w is fixed and $S_{s,l}$ is a variable. Roughly speaking, the difference between equation (5), whether used in factor analysis or in regression analysis, and equation (6) is that the parameters $R_{(p,r),l}$ in equation (5) are freely adjustable, whereas in equation (6) the constraint (9) is posed on their values.

$$R_{(p,r),l} = \sum_{j=1}^{u} \sum_{k=1}^{v} c_{j,k,l} \; P_{p,j} \; R_{r,k} \quad , \quad \text{for } l = 1 - w \tag{9}$$

Thus it can be expected that the residual standard deviation with equation (5) is smaller than with equation (6). What then is the advantage of equation (6)? It is twofold. 1) More important than the standard deviation $\hat{\sigma}$ is the length of the prediction interval $\hat{I}$. This length is equal to $2t\hat{\sigma}$, where the factor t increases with decreasing number of degrees of freedom DF. With equation (5), used in regression analysis, DF is equal to the number of observations minus the number of estimated parameters for a particular reaction in a particular solvent. Thus, DF may be quite small, and accordingly t may be rather large. With equation (6), DF is equal to the number of observations minus the number of estimated parameters in the whole data set, and this is likely to be a large number. Thus, even if $\hat{\sigma}$ is smaller with equation (5), $\hat{I}$ may be smaller with equation(6). 2) If no observations are available in a particular solvent, they cannot be predicted with equation (5), whether used in factor analysis or in regression analysis. On the other hand, equation (6) offers the possibility to predict these missing data as well. An example will be worked out in the Discussion.

DATA

We shall apply equation (6) to two cases.

1) Data on chemical equilibria, *i.e.*, ionization constants of 15 series of substituted aromatic compounds in three (admittedly rather closely related) solvents. These data are given in Tables 1^a - 1^c.

Table 1^a Investigated solvents

Code	Solvent
1	water ,25°C
2	50 v/v ethanol-water, 20 - 25°C
3	95 v/v % ethanol-water, 20 - 25°C

Table 1^b Investigated reactions (ionization of the tabulated compounds)

Code	Compounds	References
1	trans-4-X-cyclohexanecarboxylic acids	17
2	m-X-phenylacetic acids	18 - 21
3	p-X-phenylacetic acids	18 - 21
4	p-X-β-phenylpropionic acids	22,23
5	m-X-benzoic acids	18-20, 24-28
6	p-X-benzoic acids	18-20, 24-28
7	p-X-benzenephosphonic acids, first step	29
8	p-X-benzenephosphonic acids, second step	29
9	m-X-phenols	30
10	p-X-phenols	30
11	m-X-thiophenols	30,31
12	p-X-thiophenols	31
13	2-X-thiophene-5-carboxylic acids	32
14	m-X-anilines	33,34
15	p-X-anilines	33,35

Table 1^c Data on log (K_X/K_H), where K is an ionization constant

Solvent	Reaction	Substituent									
		F	Cl	Br	I	CH_3	NO_2	CN	OCH_3	$COCH_3$	$COOCH_3$
1	1		0.32					0.42	0.24		0.24
	2		0.172		0.153	-0.051	0.345				
	3	0.066	0.122	0.124	0.134	-0.058	0.461		-0.048		
	4		0.053			-0.024	0.187		-0.029		
	5	0.338	0.373	0.391	0.352	-0.069	0.710		0.115	0.376	
	6	0.062	0.226	0.232		-0.17	0.778		-0.268	0.502	
	7		0.31	0.34		-0.17			-0.27		
	8		0.17			-0.15	0.59				
	9		0.32	0.24		-0.17	0.84				
	10		0.76	0.83		-0.20	1.50		0.24		
	11		0.54	0.62	0.68		2.90	2.04	-0.23	1.88	
	12		0.835			-0.045	1.374		0.230		
	13		0.480	0.595		-0.205	1.900		-0.160	1.285	
	14	1.09	1.13			-0.11	2.12	1.83	0.39	1.03	1.04
	15	0.02	0.68	0.77	0.86	-0.45	3.67	2.92	-0.73		2.14
2	1		0.50					0.66	0.38		0.36
	2	0.18	0.19	0.19	0.16	-0.09	0.49	0.45	0.04		
	3	0.11	0.19	0.19	0.22	-0.09	0.61	0.52	-0.06		
	4		0.096			-0.053	0.323		-0.034		
	5		0.53	0.58		-0.08	1.14	0.95		0.59	0.64
	6		0.44	0.45		-0.20	1.27	1.10	-0.32	0.70	0.51
	7		0.44	0.43		-0.22			-0.39	0.84	
	8		0.22	0.27		-0.17	0.85				
	9		0.27	0.25		-0.12	0.84				
	10		1.11	1.07	1.08	-0.15			0.29	0.85	0.78
	11		0.68	0.78	0.87	-0.33	3.60		-0.24	2.07	1.70
	12		0.93	0.88	0.93	-0.21			0.24	0.85	0.80
	13		0.72	0.78	0.79	-0.29	2.79		-0.30	1.85	1.61
	14	1.08	1.17	1.20	1.22	-0.11					
	15	0.02	0.64	0.76	0.92	-0.49	3.68				
3	5		0.58			-0.20	1.35				
	6		0.43			-0.20	1.45				
	10		1.22	1.30	1.33	-0.17			0.17	0.99	0.87
	11		0.88	0.98	1.06	-0.33	3.04		-0.27	2.23	1.88
	12		1.15	1.10	1.24	-0.24			0.12	0.97	0.92
	13		0.87	0.95	1.00	-0.38	2.90		-0.44	2.04	1.82

2) Data on phase equilibria of 6 series of substituted aromatic compounds, *i.e.*, net retention volumes in reversed-phase HPLC and in gas-liquid chromatography and partition coefficients in several widely different two-phase systems. These data are given in Tables 2^a-2^c. It has been proved previously[12] that data on the influence of substituents on phase equilibria are, in principle, equivalent with data on chemical equilibria, when a correction is applied for the influence of the substituent itself on the phase equilibrium constant. In other words, the left hand side of equations (1)-(3) reads for phase equilibria as in (10).

Table 2^a Investigated two-phase systems

Code	Two-phase system	References
4	50 v/v % methanol-water/octadecylsilylsilica at 25 °C	16
5	30 v/v % acetonitrile-water/octadecylsilylsilica at 25 °C	16,36
6	32.5 v/v % tetrahydrofuran-water/octadecylsilylsilylsilica at 25 °C	36
7	helium/apiezon m at 206 °C	14
8	helium/polyethylene glycol 20,000 at 206 °C	14
9	*n*-octanol/water	37
10	50 v/v % methanol-water/*n*-hexadecane at 25 °C	16
11	30 v/v % acetonitrile-water/*n*-hexadecane at 25 °C	16,36
12	32,5 v/v % tetrahydrofuran-water/*n*-hexadecane at 25 °C	36

2^b Investigated reactions (partition of the tabulated compounds)

Code	Compounds	Code	Compounds	Code	Compounds
16	*m*-benzoic acids	18	*m*-phenols	20	*m*-anilines
17	*p*-benzoic acids	19	*p*-phenols	21	*p*-anilines

Table 2[c] Data on $\Delta\Delta \log V_N$, where V_N is a chromatographic net retention volume, and $\Delta\Delta \log P$, where P is a partition coefficient

Solvent	Reaction	Substituent									
		F	Cl	Br	I	CH_3	NO_2	CN	OCH_3	$COCH_3$	$COOCH_3$
4	16	0.158	0.162	0.100	0.034	-0.024	0.229	0.258	0.150	0.104	
	17	0.144	0.181	0.134	0.148	-0.026	0.295	0.223	0.118	0.151	
	18	0.120	0.154	0.099	0.073	-0.107	0.363	0.368	-0.035	0.210	0.170
	19	0.028	0.126	0.086	0.080	-0.096	0.314	0.314	-0.158	0.146	0.186
	20	0.071	0.117	0.081	0.051	-0.074	0.317	0.351	-0.045	0.161	
	21	-0.005	0.085	0.076	0.060	-0.065	0.176	0.274	-0.125	0.105	0.110
5	16	0.126	0.129	0.125	0.120	-0.003	0.184	0.161	0.076	0.079	
	17	0.154	0.161	0.163	0.142	-0.004	0.237	0.211	0.087	0.102	
	18	0.166	0.072	0.108	0.076	-0.078	0.324	0.362	0.036	0.236	0.165
	19	0.052	0.054	0.089	0.073	-0.068	0.276	0.284	-0.072	0.128	0.129
	20	0.239	0.244	0.244	0.211	0.007	0.417	0.407	0.040	0.161	
	21	0.077	0.185	0.204	0.205	-0.002	0.317	0.332	-0.152	0.014	0.177
6	16	0.223	0.228	0.224	0.182	-0.006	0.471	0.463	0.154	0.302	
	17	0.206	0.261	0.264	0.247		0.590	0.564	0.059	0.401	
	18	0.269	0.274	0.257	0.247	-0.030	0.544	0.538	0.066	0.221	0.259
	19	0.136	0.219	0.213	0.216	-0.035	0.497	0.421	-0.123	0.066	0.237
	20	0.321	0.347	0.350	0.347	0.057	0.667	0.611	0.028	0.123	
	21	0.175	0.232	0.243	0.270	0.018	0.460	0.451	0.159	0.136	0.263
7	18	0.056	0.065	0.051	0.062	0.030	0.104	0.108	0.023		
	19	0.054	0.042	0.042	0.049	0.000	0.232	0.213	0.015		
	20	0.030	0.029	0.029	0.028	-0.015	0.093	0.092	-0.001		
	21	0.016	0.024	0.033	0.051	-0.025	0.297	0.206	-0.034		
8	18	0.130	0.150	0.166	0.158	-0.045	0.299	0.258	-0.007		
	19	0.096	0.147	0.157	0.157	-0.047	0.513	0.425	-0.056		
	20	0.135	0.146	0.158	0.146	-0.028	0.265	0.213	0.016		
	21	0.051	0.134	0.146	0.151	-0.056	0.688	0.493	-0.085		
9	16	0.140	0.100	0.140	0.140	-0.060	0.240	0.180	0.170		-0.030
	17	0.070	0.070	0.130	0.030	-0.160	0.300	0.260	0.110		
	18	0.330	0.330	0.310	0.350	-0.060	0.820	0.810	0.140	0.480	
	19	0.170	0.220	0.270	0.330	-0.080	0.730	0.710	-0.100	0.440	0.510
	20	0.260	0.230	0.340		-0.060	0.750		0.050		
	21	0.110	0.160	0.500		-0.070	0.770		0.070		
10	18	0.741	0.782	0.802	0.749	-0.074	1.344	1.221	0.264	0.662	0.688
	19	0.533	0.626	0.621	0.603	-0.091	1.109	1.053	0.108	0.597	0.613
11	18	0.126	0.204	-0.054	0.511	-0.119	0.643	0.654	-0.369	0.360	0.024
	19	-0.422	0.258	0.151	0.054	-0.133	0.488	0.508	-0.334	0.300	0.012
12	18	0.998	0.893	0.850	0.637	-0.070	0.930	1.010	0.394	0.711	0.810
	19	0.847	0.937	1.058	0.665	-0.102	1.120	1.090	0.391	0.840	0.876

$$\Delta\Delta \log K \equiv \log (K_{X\phi Y}/K_{H\phi Y}) - \log (K_{X\phi H}/K_{H\phi H}) \qquad (10)$$

where X is a variable substituent, ϕ a common hydrocarbon skeleton, and Y a common reaction centre. It has been demonstrated also that data on liquid-liquid partition equilibria[13] and retention data in gas chromatography[14] and in normal phase[15] and reversed phase[16] liquid chromatography can be fitted excellently by equations (1), (2) or (3).

RESULTS AND DISCUSSION, *Case 1*

The results of three-mode factor analysis with missing data[9], applied to the data in Table 1[c], are presented in Table 3. The results of an analysis of variance, in which only the main effects of the solvents, reaction types and substituents (but not their interactions) were taken into account, are also presented. An analysis of variance with firstorder interactions is impossible because the number of parameters in the model (198) approaches the number of observations (213).

It follows from the one-but-last column of Table 3 that the present data can be fitted better by the factor analysis model (even with only one factor for each mode) than by an (additive) analysis of variance model. With two factors for each mode the residual standard deviation is 0.10 and this figure can be improved to 0.07 by introduction of a third factor for the reaction types and the subsituents.

However, the ability of a statistical model to fit observations is not its most useful property. Far more important is its ability to predict missing data.

The average standard deviation of the calculated values of the missing data can be computed by extension of the procedure, developed for the two-mode factor analysis case[38], to the three-mode situation. Using a linearization of the model around the true parameters, the σ^2_{pred} = {variance of (new observation - prediction)} can be computed for all missing data. The square root of its average is given in

---

Table 3 Summary of the results of an analysis of variance and of three-mode factor analysis of the data, presented in Table 1[c]

Model	NFP	NFR	NFS	NPAR	DF	$\hat{\sigma}$	$\hat{\sigma}_{pred.}$
Analysis of variance				26	187	0.68	
Three-mode factor analysis	1	1	1	26	187	0.24	0.27
" " "	2	2	2	52	161	0.10	0.13
" " "	2	2	3	61	152	0.10	0.17
" " "	2	3	2	66	147	0.08	0.16
" " "	2	3	3	77	136	0.07	0.36

Number of observations: 213

Number of missing values: 237

NFP = number of factors for the solvents

NFR = number of factors for the reactions

NFS = number of factors for the substituents

NPAR = number of estimated parameters

DF = degrees of freedom

$\hat{\sigma}$ = standard deviation of the model

$\hat{\sigma}_{pred.}$ = average standard deviation of (new observation-prediction) for the missing data the last column of Table 3. It follows that the factor analysis model (2,2,2) is the best one for the prediction of missing data, and that model (2,3,3) is of no use because it gives imprecise predictions. Detailed

Table 4 Fit of the three-mode factor analysis model (2,2,2) for the individual solvents, reactions and substituents

Solvents			Reactions			Substituents		
Code	NO	RMRSS	Code	NO	RMRSS	Symbol	NO	RMRSS
1	84	0.09	1	8	0.09	F	9	0.08
2	93	0.08	2	12	0.03	Cl	36	0.06
3	36	0.10	3	15	0.00	Br	26	0.05
			4	8	0.03	I	17	0.06
			5	18	0.05	CH_3	33	0.05
			6	18	0.12	NO_2	27	0.16
			7	9	0.10	CN	9	0.07
			8	7	0.05	OCH_3	26	0.11
			9	8	0.06	$COCH_3$	16	0.08
			10	19	0.07	$COOCH_3$	14	0.11
			11	23	0.15			
			12	18	0.06			
			13	22	0.12			
			14	13	0.12			
			15	15	0.08			

NO = number of observations

RMRSS = $(RSS/NO)^{\frac{1}{2}}$ = root of the mean residual sum of squares.

Table 5 Predictions of the missing data in Table 1^c by the three-mode

						Substit
Solvent	Reaction	F	Cl	Br	I	CH_3
1	1	0.46 ± 0.13		0.37 ± 0.12	0.36 ± 0.12	-0.02 ± 0.11
	2	0.14 ± 0.12		0.15 ± 0.11		
	3					
	4	0.05 ± 0.13		0.08 ± 0.11	0.08 ± 0.11	
	5					
	6				0.33 ± 0.11	
	7	0.13 ± 0.12			0.35 ± 0.11	
	8	0.13 ± 0.13		0.20 ± 0.11	0.22 ± 0.11	
	9	0.12 ± 0.13			0.24 ± 0.11	
	10	0.91 ± 0.15			0.81 ± 0.12	
	11	0.32 ± 0.13				-0.31 ± 0.11
	12	0.80 ± 0.14		0.72 ± 0.12	0.73 ± 0.12	
	13	0.36 ± 0.12			0.71 ± 0.11	
	14			0.97 ± 0.12	1.00 ± 0.12	
	15					
2	1	0.57 ± 0.14		0.50 ± 0.12	0.51 ± 0.12	-0.05 ± 0.11
	2					
	3					
	4	0.06 ± 0.14		0.09 ± 0.12	0.10 ± 0.12	
	5	0.48 ± 0.12			0.52 ± 0.11	
	6	0.24 ± 0.12			0.41 ± 0.11	
	7	0.16 ± 0.13			0.42 ± 0.12	
	8	0.16 ± 0.15			0.27 ± 0.12	
	9	0.15 ± 0.14			0.30 ± 0.12	
	10	1.14 ± 0.16				
	11	0.39 ± 0.13				
	12	0.10 ± 0.15				
	13	0.44 ± 0.13				
	14					
	15					
3	1	0.62 ± 0.15	0.58 ± 0.13	0.57 ± 0.13	0.58 ± 0.13	-0.07 ± 0.11
	2	0.19 ± 0.13	0.22 ± 0.12	0.22 ± 0.12	0.23 ± 0.11	-0.05 ± 0.11
	3	0.12 ± 0.12	0.18 ± 0.11	0.20 ± 0.11	0.21 ± 0.11	-0.07 ± 0.11
	4	0.07 ± 0.14	0.10 ± 0.13	0.10 ± 0.12	0.11 ± 0.12	-0.03 ± 0.11
	5	0.52 ± 0.13		0.56 ± 0.12	0.58 ± 0.12	
	6	0.26 ± 0.12		0.41 ± 0.11	0.44 ± 0.11	
	7	0.17 ± 0.14	0.37 ± 0.13	0.40 ± 0.12	0.44 ± 0.13	-0.18 ± 0.11
	8	0.17 ± 0.15	0.26 ± 0.13	0.27 ± 0.13	0.29 ± 0.13	-0.10 ± 0.11
	9	0.17 ± 0.15	0.28 ± 0.13	0.30 ± 0.12	0.32 ± 0.12	-0.11 ± 0.11
	10	1.24 ± 0.16				
	11	0.42 ± 0.14				
	12	1.08 ± 0.15				
	13	0.47 ± 0.13				
	14	1.27 ± 0.15	1.40 ± 0.13	1.41 ± 0.13	1.46 ± 0.13	-0.31 ± 0.11
	15	-0.01 ± 0.15	0.67 ± 0.14	0.80 ± 0.14	0.92 ± 0.14	-0.51 ± 0.12

factor analysis model (2,2,2) and their standard deviations

uent				
NO_2	CN	OCH_3	$COCH_3$	$COOCH_3$
0.52 ± 0.13			0.07 ± 0.14	
	0.30 ± 0.19	0.02 ± 0.11	0.18 ± 0.13	0.16 ± 0.12
	0.41 ± 0.12		0.31 ± 0.12	0.27 ± 0.12
	0.19 ± 0.12		0.14 ± 0.13	0.12 ± 0.12
	0.68 ± 0.12			0.34 ± 0.12
	0.81 ± 0.11			0.53 ± 0.11
1.22 ± 0.16	1.00 ± 0.14		0.84 ± 0.14	0.71 ± 0.13
	0.54 ± 0.13	-0.05 ± 0.12	0.41 ± 0.15	0.35 ± 0.13
	0.63 ± 0.13	-0.08 ± 0.12	0.50 ± 0.14	0.43 ± 0.13
	1.15 ± 0.16		0.42 ± 0.15	0.46 ± 0.13
				1.52 ± 0.12
	1.10 ± 0.15		0.45 ± 0.14	0.48 ± 0.13
	1.85 ± 0.12			1.25 ± 0.12
			2.65 ± 0.17	
0.88 ± 0.15			0.26 ± 0.13	
			0.25 ± 0.12	0.23 ± 0.12
			0.38 ± 0.12	0.33 ± 0.12
	0.24 ± 0.12		0.18 ± 0.13	0.15 ± 0.12
		0.07 ± 0.11		
1.44 ± 0.17	1.18 ± 0.15	0.83 ± 0.13		
	0.66 ± 0.12	-0.06 ± 0.12	0.50 ± 0.14	0.43 ± 0.12
	0.76 ± 0.12	-0.09 ± 0.12	0.60 ± 0.14	0.51 ± 0.12
2.16 ± 0.17	1.77 ± 0.17			
	2.58 ± 0.14			
2.03 ± 0.16	1.66 ± 0.16			
	2.23 ± 0.13			
3.02 ± 0.19	2.47 ± 0.17	0.14 ± 0.12	1.46 ± 0.14	1.36 ± 0.13
	3.17 ± 0.17	-0.79 ± 0.14	2.94 ± 0.17	2.42 ± 0.14
1.07 ± 0.17	0.87 ± 0.14	0.19 ± 0.12	0.37 ± 0.13	0.38 ± 0.12
0.56 ± 0.13	0.48 ± 0.12	0.01 ± 0.11	0.29 ± 0.12	0.26 ± 0.12
0.65 ± 0.13	0.53 ± 0.12	-0.06 ± 0.11	0.41 ± 0.12	0.36 ± 0.12
0.31 ± 0.14	0.26 ± 0.12	-0.02 ± 0.12	0.19 ± 0.13	0.17 ± 0.12
	1.10 ± 0.12	0.06 ± 0.11	0.65 ± 0.12	0.60 ± 0.11
	1.08 ± 0.12	-0.10 ± 0.11	0.82 ± 0.12	0.71 ± 0.11
1.52 ± 0.18	1.25 ± 0.16	-0.20 ± 0.12	1.03 ± 0.14	0.88 ± 0.14
0.88 ± 0.14	0.72 ± 0.13	-0.07 ± 0.12	0.54 ± 0.13	0.47 ± 0.12
1.00 ± 0.14	0.82 ± 0.13	-0.10 ± 0.12	0.64 ± 0.13	0.55 ± 0.12
2.54 ± 0.18	2.08 ± 0.18			
	2.75 ± 0.14			
2.36 ± 0.17	1.93 ± 0.17			
	2.40 ± 0.14			
3.43 ± 0.21	2.81 ± 0.19	0.12 ± 0.13	1.70 ± 0.15	1.57 ± 0.14
3.99 ± 0.19	3.27 ± 0.20	0.82 ± 0.15	3.03 ± 0.19	2.50 ± 0.16

information on the fit of the observations by this model is given in Table 4. It appears that there are no large discrepancies in the fit for the individual solvents, reactions and substituents. The rather large values of the root of the mean residual sum of squares for reaction 11 and for the NO_2 substituent correspond with the large substituent effects observed for this reaction and for this substituent. The values, predicted by the factor analysis model (2,2,2) for the 237 missing data are given in Table 5, together with their standard deviation. This table gives a striking illustration of the merits of the proposed method. For, 90 of the missing data are from reactions on which no observations at all are available in solvent 3 (95 v/v % ethanol). The traditional way to give a mathematical-statistical description of the data is by the Hammett (1), Taft (2) or Nieuwdorp (3) equation, *i.e.*, for each reaction and each solvent separately. Of course, this approach does not offer the possibility to predict data on a reaction in solvents where no measurements at all are available. Further, for 53 missing data that can be predicted by the Nieuwdorp equation, the 0.90 probability interval is larger than for the predictions by three-mode factor analysis, for reasons set forth in the theoretical part. In only 73 cases the 0.90 probability interval for predictions by the Nieuwdorp equation is smaller. For 3 series only 3 data points are available, which makes it impossible to compare the intervals for the remaining 21 series.

We conclude from Tables 3-5 that the combined influence of solvent, reaction type and substituent on chemical equilibrium constants (and presumably also on reaction rate constants) can be described satisfactorily by equation (6).

Case 2. It must be admitted that the range of solvents in case 1 is rather small: water, 50 and 95 v/v % ethanol. Therefore we applied equation (6) to a set of phase equilibrium constants where the range of two-phase systems is very large (see Table 2^a). The results of three-mode factor analysis with missing data, applied to these data (listed in Table 2^c) are given in Table 6. Also given are the results of an analysis of variance. These results are analogous to the results shown in Table 3. Again the three-mode factor analysis model is better than the analysis of variance model. With the factor analysis model, and one factor for the reactions and two factors for the two-phase systems and for the substituents, the residual standard deviation of the observations is equal to 0.09, and the standard deviation of the

Table 6 Summary of the results of an analysis of variance and of three-mode factor analysis of the data, presented in Table 2^c

Model	NFP	NFR	NFS	NPAR	DF	$\hat{\sigma}$	$\hat{\sigma}_{pred.}$
Analysis of variance				23	319	0.37	
Three-mode factor analysis	1	1	1	23	319	0.11	0.12
" " "	2	1	2	39	303	0.09	0.10
" " "	3	1	3	53	289	0.09	0.10
" " "	2	2	2	46	296	0.07	1.21
" " "	3	3	3	75	267	0.06	0.77

Number of observations 342

Number of missing data 198

predictions is equal to 0.10. Thus, even in this case where the phases range from a gas, via apolar liquids, to water, three-mode factor analysis gives a good description of the observations and an accurate prediction of the 37% missing data. The values of $\hat{\sigma}$ and $\hat{\sigma}_{pred.}$ for the (2,1,2) model are even better than those for the (2,2,2) model in Table 3.

Regression analyses of the data in Table 2^c with the Taft equation (2) yield a pooled value for the residual standard deviation that is

Table 7 Fit of the three-mode factor analysis model (2,1,2) for the individual solvents, reactions and substituents

Solvents			Reactions			Substituents		
Code	NO	RMRSS	Code	NO	RMRSS	Symbol	NO	RMRSS
1	57	0.06	1	36	0.09	F	38	0.09
2	57	0.05	2	34	0.10	Cl	38	0.07
3	56	0.09	3	85	0.08	Br	38	0.06
4	48	0.11	4	86	0.08	I	36	0.08
5	32	0.08	5	49	0.06	CH_3	37	0.04
6	32	0.04	6	52	0.09	NO_2	38	0.11
7	20	0.06				CN	36	0.09
8	20	0.14				OCH_3	38	0.09
9	20	0.09				$COCH_3$	26	0.10
						$COOCH_3$	17	0.08

NO = number of observations

RMRSS = $(RSS/NO)^{\frac{1}{2}}$ = root of the mean residual sum of squares equal to that, obtained with the factor analysis model (2,1,2). However, the number of parameters in the factor analysis model is much less than that in the regression analysis model, *viz.*, 39 *vs.* 78. Moreover, the ability to predict missing data is better for the factor analysis model (see the introductory section).

Detailed information on the fit of the observations by the factor analysis model (2,1,2) is given in Table 7. It appears from Tables 6 and 7 that the combined influence of two-phase system, reaction type and substituent on phase equilibrium constants can be described adequately by equation (6), even when the two-phase systems differ widely. We believe that an analogous conclusion holds for chemical equilibrium constants and reaction rate constants, in widely different solvents.

CONCLUSIONS

Three-mode factor analysis is a good mathematical-statistical model for the simultaneous description of the influence of solvent, reaction type and substituents on equilibrium constants, and presumably also on reaction rate constants. The residual standard deviation of the observations is of the same magnitude as is obtained by a regression analysis model, *viz.*, by applying the Taft equation to each reaction in each solvent separately. However, the number of parameters in the factor analysis model is far less than that in the regression analysis model. The advantage of the factor analysis model is, that it is better suited for the prediction of missing data. It can even predict data that cannot be predicted at all with the Taft equation, *viz.*, for reactions on which no data at all are available in a particular solvent.

REFERENCES

1. G.H.E. Nieuwdorp, C.L. de Ligny, and J.C. van Houwelingen, *J.Chem.Soc., Perkin II,* 1979, 537.
2. L.P. Hammett,'Physical Organic Chemistry', McGraw-Hill, New York, 1940, p. 184.
3. R.W. Taft and I.C. Lewis, *J.Am.Chem.Soc.*, 1958, 80, 2436.
4. S. Ehrenson, R.T.C. Brownlee, and R.W. Taft, *Progr. Phys. Org. Chem.*, 1973, 10, 1.
5. M.C. Spanjer and C.L. de Ligny, in course of publication.
6. H.H. Harman, 'Modern Factor Analysis', Univ. Chicago Press, Chicago, 1970.
7. C.L. de Ligny, G.H.E. Nieuwdorp, W.K. Brederode, W.E. Hammers, and J.C. van Houwelingen, *Technometrics,* 1981, 23, 91.
8. P.M. Kroonenberg and J. de Leeuw, *Psychometrica,* 1980, 45,69.
9. H.M. Weesie and J.C. van Houwelingen, GEPCAM User's Manual, Institute of Mathematical Statistics, Utrecht, 1983, (available on request).
10. P.H. Weiner, *J.Am.Chem.Soc.,* 1973, 95, 5845.
11. M.R. Bacon and G.E. Maciel, *J.Am.Chem.Soc.,* 1973, 95, 2413.
12. C.L. de Ligny, N.J. Koole, H.D. Nelson and G.H.E. Nieuwdorp, *J. Chromatogr.* 1975, 114, 63.
13. T. Fujita, J. Iwasa, and C. Hansch, *J.Am.Chem.Soc.,* 1964, 86, 5175.
14. G.H.E. Nieuwdorp, C.L. de Ligny, and N.G. van der Veen, *J.Chromatogr.,* 1978, 154, 133.
15. L.R. Snyder, 'Principles of Adsorption Chromatography', Marcel Dekker, New York, 1968.
16. M.C. Spanjer and C.L. de Ligny, *J.Chromatogr.,* 1982, 253,23.
17. S. Siegel and J.M. Komarmy, *J.Am.Chem.Soc.,*1960, 82, 2547.
18. J.F.J. Dippy and F.R. Williams, *J.Chem.Soc.,*1934, 1888.

19. J.W. Baker, J.F.J. Dippy and J.E. Page, *J.Chem.Soc.*, 1937, 1774.

20. J.F.J. Dippy, S.R.C. Hughes and B.C. Kitchiner, *J.Chem.Soc.*, 1964, 1275.

21. A.J. Hoefnagel, J.C. Monshouwer, E.C.G. Snorn and B.M. Wepster, *J.Am.Chem.Soc.*, 1973, 95, 5350.

22. J.F.J. Dippy and J.E. Page, *J.Chem.Soc.*, 1938, 357.

23. R. Fuchs, C.A. Kaplan, J.J. Bloomfield and L.F. Hatch, *J.Org.Chem.*, 1962, 27, 733.

24. J.F.J. Dippy, F.R. Williams and R.H. Lewis, *J.Chem.Soc.*, 343, 1935.

25. J.F.J. Dippy and R.H. Lewis, *J.Chem.Soc.*, 644, 1936.

26. L.G. Bray, J.F.J. Dippy and R.S.C. Hughes, *J.Chem.Soc.*, 1957, 265.

27. D.H. McDaniel and H.C. Brown, *J.Org.Chem.*, 1958, 23, 420.

28. W.L. Bright and H.T. Briscoe, *J.Phys.Chem.*, 1933, 37, 787.

29. H.H. Jaffé, L.D. Freedman and G.O. Doak, *J.Am.Chem.Soc.*, 1953, 75,2209.

30. G. Schwarzenbach and E. Rudin, *Helv.Chim.Acta*, 1939, 22, 360.

31. P. de Maria, A. Fini and F.M. Hall, *J.Chem.Soc. Perkin II*, 1973,1969.

32. A.R. Butler, *J.Chem.Soc.* (B), 1970, 867.

33. A. Bryson, *J.Am.Chem.Soc.* 1960, 82, 4858.

34. G. Baddeley, J. Chadwick and H.T. Taylor, *J.Chem.Soc.*, 1954, 2405.

35. A. de Courville, *Compt.Rend.Acad.Sci.*, Paris, 1966, 262, 1196.

36. M.C. Spanjer and C.L. de Ligny, *J.Chromatogr.Sci.*, submitted for publication.

37. A. Leo, C. Hansch and D. Elkins, *Chem.Rev.*, 1971, 71, 525.

38. J.C. van Houwelingen, Principal components of large matrices with missing elements, Proceedings of the 3rd Prague Symposium on Asymptotic Statistics, North Holland Inc., to appear in 1984. (Momentally available as preprint nr. 310.).

CHAPTER V

THREE-MODE FACTOR ANALYSIS OF DATA ON THE RETENTION IN NORMAL-PHASE HPLC

SUMMARY

It is shown that the Snyder equation, which has proven its value for bare silica and alumina adsorbents, is not quite satisfactory for fitting retention data in normal-phase HPLC on chemically bonded phases.

This equation is in fact a special case of the mathematical-statistical three-mode factor analysis model. This model, in its general form, has been used to fit two sets of literature data on the retention in normal-phase HPLC, of 19 solutes on 6 adsorbents with 2 eluents, and of 39 solutes on 3 adsorbents with two eluents, respectively.

As far as we are aware, this study represents the first application of three-mode factor analysis with missing data,and also the first application of three-mode factor analysis in the field of the natural sciences.

The accuracy of the fit of the observations and of the prediction of the missing data, for various numbers of extracted factors, is discussed.

INTRODUCTION

For the correlation of data that can be classified in two modes, factor analysis[1] is often a good mathematical model. In factor analysis, the data y are correlated by the equation:

$$y_{a,s} = \sum_{j=1}^{n} c_j \, A_{a,j} \, S_{s,j} \qquad (1)$$

where a and s are the two modes in which the data can be classified, the parameters (or "factors") A depend only on a and the factors S depend only on s and c_j is a scaling constant. The objective is to describe the data with a small number n of factors.

An example of data that can be classified in two modes are gas-chromatographic data on the retention of a number of solutes s on a number of absorbents a. Several authors, e.g.[2-6], have applied eqn. (1) to data of this kind. In the case of missing data the straightforward solution of eqn. (1) is precluded, but the authors have developed an iterative procedure[7] for the estimation of $A_{a,j}$ and $S_{s,j}$. Once this has been performed, eqn.(1) can be used to estimate the missing data.

Of course, many data sets must be classified in three or even more modes. An example is a data set on the liquid-chromatographic retention of a number of solutes s on a number of adsorbents a, obtained with a number of eluents e.

The extension of eqn. (1) to three modes is[8]:

$$y_{a,e,s} = \sum_{j=1}^{p} \sum_{k=1}^{q} \sum_{l=1}^{r} c_{j,k,l} \, A_{a,j} \, E_{e,k} S_{s,l} \qquad (2)$$

where c denotes the three-mode core matrix of scaling constants.

For the cases where equation (2) is applied to j, k and l are usually larger than one.

As far as we know, eqn. (2) has only been applied in the field of the social sciences[9], and only for cases where data exist for each combination of a, e and s. Two of the present authors have recently devised a method to estimate $A_{a,j}$, $E_{e,k}$ and $S_{s,l}$ for the case of missing data[10], and we will apply this method here to a case from the field of the natural sciences, *viz.*, data on the retention in normal-phase HPLC.

It is quite obvious to investigate the merits of eqn. 2 for the correlation of data of this kind. For normal-phase liquid chromatography on bare silica and alumina a good physical model exists, *viz.* that of Snyder[11], and his resulting equation is a special case of eqn. (2). For monosubstituted benzene solutes it can be written as:

$$y_{a,e,s} = A_{a,1} E_{e,1} S_{s,1} c_{1,1,1} + A_{a,2} E_{e,1} S_{s,2} c_{2,1,2} + A_{a,2} E_{e,2} S_{s,3} c_{2,2,3} + A_{a,3} E_{e,2} S_{s,4} c_{3,2,4} \qquad (3)$$

whereas for polycyclic aromatic hydrocarbon solutes an additional term:

$$A_{a,4} E_{e,1} S_{s,5} c_{4,1,5} \qquad (4)$$

is required. The symbols in the Snyder equation that are equivalent with y and with the parameters A, E and S in eqns. (3) and (4) are given in Table I.

However, while the Snyder equation holds good for bare adsorbents, it is less suited for the correlation of data, obtained on chemically bonded phases. Hammers et al. [12-15] have demonstrated this repeatedly. So, as a matter of fact these has grown a need for a model that can correlate observations and predict missing data more accurately. It will be shown that three-mode factor analysis fulfils these requirements.

Table I

Symbols in the Snyder equation that are equivalent with y and with the parameters A, E and S in eqn. (3)

j,k,l	1	2	3	4	5
$y = \log V_N/W$					
A_j	$\log V_a$	α	$\alpha\gamma$	$\alpha\zeta$	
E_k	1	ε^o			
S_l	1	S_o	Σa_i(calc.)	$\Sigma\Delta a_i(\mathrm{SiO_2})$	$n-6$

V_N = chromatographic net retention volume, $\mathrm{cm^3}$

W = weight of the adsorbent, g

V_a measure for the specific surface area of the adsorbent

α measure for the strength of the adsorbent

γ,ζ measures for the heterogeneity of the adsorbent

ε^o measure for the strength of the eluent

S_o measure for the Lewis acid or base strength of the solute

Σa_i(calc.) = surface area of the solute molecule

$\Sigma\Delta a_i(\mathrm{SiO_2})$ empirical increment of the surface area of the solute molecule for silica adsorbents

n = number of aromatic carbon atoms in the solute molecule

DATA

The data that we shall analyse are taken from recent investigations in our laboratory[12-15]. The first set contains data on simple solutes, *viz.* , monosubstituted benzenes and polycyclic aromatic hydrocarbons. The adsorbents are described in Table II and the solutes in Table III. The eluents were *n*-hexane(1) and 35 v/v % methylene chloride in *n*-hexane (2), and the temperature was 25°C. The data are given in Table IV.

A rough estimate of the precision of the fit of these data by the Snyder equation can be obtained as follows. For the adsorbents 4-6, the eluent *n*-hexane and monosubstituted benzene solutes the standard deviation of the fit was found to be 0.14, 0.12 and 0.10, respectively, when three strongly deviating data were excluded from the regression analyses by the Snyder equation[15]. For polycyclic aromatic hydrocarbon solutes the standard deviation of the fit was found to be 0.08, 0.08 and 0.12, respectively, when two strongly deviating data were excluded from the regression analyses[15]. For the adsorbents 1-3 the fits of the regressions are better, but here too several data that deviate rather strongly (0.2-0.3) from the values, predicted by the Snyder equation, were noted[12-14]. From these considerations we estimate that the overall precision of the fit of the data in Table IV by the Snyder equation is not better than 0.15.

The second set contains data on more complicated solutes, *viz.*, monosubstituted phenols,anilines and pyridines. Retention data for these solutes have been measured on adsorbents 1-3 with eluents 2 and 3 (methylene chloride). They are given in Table V. The standard deviations of the fit of these data by the Snyder equation range from 0.10 to 0.28 [12-14].

Table II

Investigated adsorbents

Code	Nature of the adsorbent	Reference
1	Octadecyl-silica	12
2	N-cyanoethyl-N-methylamino-silica	13
3	Aminobutyl-silica	14
4	2,4-dinitroanilino-silica	15
5	Bis-(3-nitrophenyl)-sulfone-silica	15
6	2,4,7-trinitrofluorenimine-silica	15

Table III

Investigated solutes in Table IV

Code	Monosubstituted benzenes	Code	Polycyclic aromatic hydrocarbons
1	Anisole	7	Naphthalene
2	Thioanisole	8	Acenaphthene
3	Nitrobenzene	9	Fluorene
4	Benzonitrile	10	Bibenzyl
5	Acetophenone	11	Anthracene
6	Methyl benzoate	12	Phenanthrene
		13	Pyrene
		14	Fluoranthene
		15	Chrysene
		16	3,4 Benzpyrene
		17	Perylene
		18	Triphenylene
		19	Coronene

Table IV

Data on log V_N/W in normal-phase HPLC, for the adsorbents and the solutes listed in Tables II and III, respectively

Adsorbents	1		2		3		4		5		6	
Eluents	1	2	1	2	1	2	1	2	1	2	1	2
Solutes												
1	0.69	-0.28	0.54	-0.33	0.45	-	0.19	-	0.26	-	0.35	-
2	0.35	-0.76	0.24	-0.55	0.32	-	0.14	-	0.17	-	0.27	-
3	1.05	-0.26	0.85	-0.17	0.82	-0.33	0.88	-0.07	0.92	-0.04	0.86	-0.20
4	1.45	0.16	1.34	0.14	1.12	-0.17	1.28	0.15	1.35	0.21	1.15	-0.01
5	-	0.72	1.79	0.55	1.38	0.07	1.49	0.35	1.64	0.46	1.46	0.27
6	1.66	0.33	1.25	0.19	0.97	-0.20	1.05	-0.03	1.18	0.10	1.00	-0.08
7	0.09	-	-0.06	-0.60	0.21	-	0.15	-	0.13	-	0.23	-
8	0.18	-0.95	0.00	-	0.31	-	0.29	-	0.28	-	0.34	-
9	0.37	-	0.18	-0.50	0.52	-0.73	0.45	-	0.46	-	0.56	-
10	0.47	-	0.23	-0.66	0.38	-	-0.01	-	0.00	-	0.23	-
11	0.37	-0.91	0.22	-0.58	0.66	-0.68	0.78	-0.19	0.79	-0.16	0.85	-0.14
12	0.39	-0.91	0.23	-0.48	0.73	-0.59	0.79	-0.16	0.80	-0.16	0.86	-0.12
13	0.45	-0.92	0.29	-0.50	-	-	1.15	0.21	1.14	0.19	1.20	0.20
14	0.51	-0.91	0.32	-	0.94	-0.43	1.13	0.13	1.14	0.11	1.21	0.14
15	0.70	-0.76	0.51	-0.47	1.17	-0.29	1.46	0.30	1.47	0.28	1.56	0.42
16	0.78	-0.68	0.59	-0.45	1.31	-0.24	1.85	0.68	1.85	0.63	1.98	0.86
17	0.82	-0.61	0.64	-0.35	1.41	-0.16	1.92	0.78	1.92	0.73	2.02	0.93
18	0.70	-0.67	-	-	1.18	-0.31	-	-	-	-	-	-
19	0.96	-0.46	0.77	-0.31	1.63	0.01	-	-	-	-	-	-

RESULTS

The results of three-mode factor analysis with missing data[10], applied to the data in Table IV and V are presented in Table VI and IX,respectively. The results of an analysis of variance in which only the main effects of the adsorbents, eluents and solutes (but not their interactions) were taken into account, are also presented in these tables. The computer program, called GEPCAM (Generalized Principal Components Analysis with Missing values), as well as a mathematical treatment of its underlying theory[16], are available on request at the second pair of authors.

DISCUSSION

It follows from Table VI that the present data can be fitted better by the three-mode factor analysis model (even with only one factor for each mode) than by the (additive) analysis of variance model. An analysis of variance with first order interactions is hardly feasible because it involves too many free parameters, namely 138, and it is less suited for making predictions.

Table VI shows further that the factor analysis model, eqn.(2), with only 3 factors for the solutes, and 2 factors for the adsorbents and for the eluents gives already a better fit of the data in Table IV than the Snyder equation. Introduction of a third factor for the adsorbents reduces the standard deviation of the fit even further, to a value that is only 1/3 of our estimate for the overall precision of the Snyder equation. With 3 factors for the solutes and the adsorbents, and 2 factors for the eluents, eqn.(2) explains 99.7% of

Table V

Data on log V_N/W in normal-phase HPLC, for monosubstituted phenols, anilines and pyridines

	Adsorbents		1		2		3	
	Eluents		2	3	2	3	2	3
	Solutes							
Code	Series	Substituent						
20	Phenols	*m*F	-	0.79	1.00	0.53	1.95	1.18
21		*p*F	-	0.42	1.01	0.67	1.81	1.03
22		*m*Cl	0.78	0.48	0.98	0.71	2.01	1.22
23		*p*Cl	0.83	0.46	1.00	0.68	1.93	1.14
24		*m*Br	0.78	0.49	1.00	0.76	2.00	1.17
25		*p*Br	0.84	0.48	1.02	0.71	1.98	1.09
26		$m\mathrm{CH_3}$	0.75	0.37	0.96	0.55	1.57	0.73
27		$p\mathrm{CH_3}$	0.78	0.41	0.98	0.57	1.57	0.74
28		$m\mathrm{OCH_3}$	1.22	0.72	1.34	0.87	1.92	0.94
29		$p\mathrm{OCH_3}$	1.32	0.43	1.41	0.84	1.84	0.87
30		$m\mathrm{NO_2}$	1.45	1.08	1.64	1.27	-	1.57
31		$p\mathrm{NO_2}$	1.63	1.38	1.88	1.36	-	2.09
32		*m*CN	-	1.12	1.94	1.23	-	1.51
33		*p*CN	1.77	1.30	2.04	1.46	-	1.82
34		$m\mathrm{COOCH_3}$	-	1.22	1.84	1.23	-	1.26

35		*p*COOCH$_3$	-	1.37	1.94	1.37	-	1.54
36		*m*COCH$_3$	-	1.61	2.16	1.48	-	1.52
37		*p*COCH$_3$	-	1.77	2.37	1.78	-	1.82
38	Anilines	*m*F	0.80	0.09	0.72	0.12	0.63	-0.25
39		*p*F	1.22	0.55	1.03	0.47	0.85	0.00
40		*m*Cl	0.78	0.06	0.68	0.10	0.63	-0.24
41		*p*Cl	1.01	0.27	0.85	0.24	0.75	-0.12
42		*m*Br	0.78	0.05	0.67	0.08	0.66	-0.24
43		*p*Br	0.97	0.21	0.83	0.21	0.75	-0.14
44		*m*CH$_3$	1.25	0.61	0.94	0.46	0.73	-0.09
45		*p*CH$_3$	1.29	0.80	1.13	0.45	0.83	0.03
46		*m*OCH$_3$	1.66	0.85	1.33	0.64	1.01	0.03
47		*p*OCH$_3$	-	1.28	1.67	0.99	1.24	0.29
48		*m*NO$_2$	1.15	0.09	1.01	0.19	1.00	-0.13
49		*p*NO$_2$	1.20	-0.02	1.19	0.22	1.26	0.08
50		*m*CN	1.56	0.49	1.34	0.47	1.14	-0.03
51		*p*CN	1.48	0.28	1.33	0.38	1.21	0.02
52		*m*COCH$_3$	-	1.33	1.93	1.02	1.45	0.26
53		*p*COCH$_3$	-	1.19	2.00	1.04	1.58	0.33
54	Pyridines	3Cl	-	0.96	1.10	0.84	0.57	-0.07
55		3Br	-	0.89	1.17	0.81	0.57	-0.08
56		4CH$_3$	2.07	2.01	2.13	1.70	1.30	0.74
57		3CN	-	1.18	1.61	1.04	0.94	-0.06
58		4CN	1.74	1.15	1.65	1.11	1.01	0.06

Table VI

Summary of the results of analysis of variance and of three-mode factor analysis of the data, presented in Table IV

Model	NFA	NFE	NFS	NPAR	DF	$\hat{\sigma}$	$\sqrt{S_p}$	$\hat{\sigma}_{pred.}$
Analysis of Variance				25	158	0.54		
Factor analysis	1	1	1	25	158	0.38	0.41	0.41
" "	2	2	2	50	133	0.23	0.27	0.26
" "	3	2	2	55	128	0.22	0.26	0.25
" "	2	2	3	68	115	0.11	0.14	0.20
" "	3	2	3	75	108	0.05	0.07	0.10
" "	3	2	4	93	90	0.05	0.07	32.15

Number of observations : 183

Number of missing values: 45

NFA = number of factors for the adsorbents

NFE = number of factors for the eluents

NFS = number of factors for the solutes

NPAR = number of estimated parameters

DF = degrees of freedom

$\hat{\sigma}$= standard deviation of the model

$\sqrt{S_p}$ = standard deviation of (new observation - prediction) when the new observations are generated by a random process

$\hat{\sigma}_{pred.}$= average standard deviation of (new observation - prediction) for the missing values in the data set

---

the variance of the data in Table IV. It can also be concluded that observations that strongly deviate from predictions by the Snyder equation can be fittted well by the three-mode factor analysis model.

Detailed information on this model is given in Table VII, which shows that the fit for each adsorbent is about equally good. The same conclusion applies to the eluents and to the solutes.

Table VII

Fit of the three-mode factor analysis model (3,2,3) for the individual adsorbents, eluents and solutes in Table IV

Adsorbents			Eluents			Solutes		
Code	NO	RMRSS	Code	NO	RMRSS	Code	NO	RMRSS
1	34	0.04	1	105	0.03	1	8	0.07
2	34	0.04	2	78	0.04	2	8	0.03
3	31	0.03				3	12	0.03
4	28	0.03				4	12	0.06
5	28	0.03				5	11	0.03
6	28	0.05				6	12	0.04
						7	7	0.03
						8	7	0.03
						9	8	0.03
						10	7	0.08
						11	12	0.03
						12	12	0.03
						13	10	0.04
						14	11	0.03
						15	12	0.03
						16	12	0.03
						17	12	0.03
						18	12	0.00
						19	6	0.03

NO = number of observations

RMRSS = root of the mean residual sum of squares

It must be realized that a comparison of the fit of the data by the Snyder model and the factor analysis model alone does not do justice to the merits of the former. The Snyder model contains far less parameters than the factor analysis model (3,2,3), *i.e.*, 24 *vs.* 75, as the values for the constants, characterizing the eluents and the solutes, can be taken from literature. Moreover, the Snyder model gives a good deal of physical insight, whereas the factor analysis model is not a physical, but a mathematical-statistical model.

The ability of a statistical model to fit observations is not its most useful property. Far more important is its ability to predict accurate values for missing data. (This statement holds for a science like chemistry, where data are usually very precise, but their measurement is often costly or time-consuming. In such a situation there is little need for data smoothing but a clear need for the prediction of missing data. So, the choice for a particular statistical model should be made on the basis of its ability to predict missing data. In the social sciences the reverse situation exists, and here the choice of a statistical model should be based on its fit to the obervations). The ability of a model to predict missing data can be measured by the so called S_p-criterion, defined by $S_p = \hat{\sigma}^2(1 + \frac{p}{n - p - 1})$, where p = number of parameters (noted as NPAR in Table VI), n = number of observations and $\hat{\sigma}^2$ = variance of the model. The quantity S_p is an estimate of the variance of $(y_{\text{new}} - y_{\text{pred.}})$, where y_{new} is a new observation and $y_{\text{pred.}}$ its prediction, under the assumption that the observations are generated by a kind of random mechanism[17]. Using this criterion we learn from the one-but-last column of Table VI that (3,2,3) is the best model and that it is no use to introduce more factors.

The assumption about the way observations are generated is not very plausible in our situation. A better idea about the appropriateness of a model can be obtained by computing $\sigma^2_{\text{pred.}}$ = variance of $(y_{\text{new}} - y_{\text{pred.}})$ averaged over the missing data. The value of $\sigma^2_{\text{pred.}}$ can be computed by extension of the procedure, developed for the two-mode factor analysis case[10], to the three-mode situation. Using a linearization of the model around the true parameters, the variance of $(y_{\text{new}} - y_{\text{pred.}})$ can be computed for all missing data. From the values of $\sigma_{\text{pred.}}$, presented in the last column of table VI, we learn again that (3,2,3) is the model to be preferred and that (4,2,3) is of no use, because it gives nonsensical predictions. Moreover, this technique enables us to give 0.95-prediction intervals for future observations of a missing datum. These prediction intervals, calculated with model (3,2,3), are given in Table VIII together with the calculated values for the missing data.

Values for missing data can also be calculated by the Snyder equation and these results are collected in the last column of Table VIII, On comparing the last two columns of this table one notices several values, predicted with the Snyder equation, that deviate so strongly that they do not fall within the 0.95-prediction interval calculated with the (3,2,3) model. (These cases are marked with an asterisk in the last column of Table VIII).

The correctness of the three-mode factor analysis model for these relatively simple compounds raised our interest in further investigations on more complicated molecules. Therefore a data set for substituted phenols, anilines and pyridines, measured in eluents 2 and 3 (pure methylene chloride) on adsorbents 1,2 and 3 was compiled from references 12-14 (Table V). The results of an analysis of variance,

Table VIII

Predictions of the missing data in Table IV by the three-mode factor analysis model (3,2,3) and their 0.95-intervals. Also given, for the sake of comparison, are predictions by Snyder's model

Missing datum			Predictions according to	
j	k	l	model (3,2,3)	Snyder
3	2	1	- 0.58 ± 0.16	- 0.69
4	2	1	- 0.52 ± 0.17	- 0.42
5	2	1	- 0.50 ± 0.17	- 0.39*
6	2	1	- 0.58 ± 0.18	- 0.36
3	2	2	- 0.78 ± 0.17	- 0.74
4	2	2	- 0.64 ± 0.18	- 0.65
5	2	2	- 0.65 ± 0.18	- 0.63
6	2	2	- 0.67 ± 0.19	- 0.51
1	1	5	2.22 ± 0.18	2.07
1	2	7	- 0.83 ± 0.18	- 0.86
3	2	7	- 0.69 ± 0.20	- 0.67
4	2	7	- 0.49 ± 0.19	- 0.48
5	2	7	- 0.52 ± 0.19	- 0.48*
6	2	7	- 0.49 ± 0.20	- 0.23*
2	2	8	- 0.66 ± 0.13	- 1.09
3	2	8	- 0.78 ± 0.17	- 0.86
4	2	8	- 0.51 ± 0.18	- 0.64
5	2	8	- 0.54 ± 0.18	- 0.65
6	2	8	- 0.51 ± 0.19	- 0.37
1	2	9	- 0.81 ± 0.15	- 0.90
4	2	9	- 0.39 ± 0.15	- 0.28
5	2	9	- 0.41 ± 0.15	- 0.29*
6	2	9	- 0.40 ± 0.16	- 0.02*
1	2	10	- 0.96 ± 0.19	- 1.18
3	2	10	- 1.05 ± 0.23	- 0.94*
4	2	10	- 0.94 ± 0.23	- 0.54*
5	2	10	- 0.96 ± 0.23	- 0.56*
6	2	10	- 1.00 ± 0.24	- 0.24*
3	1	13	0.91 ± 0.12	0.53*
3	2	13	- 0.45 ± 0.12	- 0.30*
2	2	14	- 0.56 ± 0.12	- 0.81*
2	1	18	0.54 ± 0.13	0.31*
2	2	18	- 0.42 ± 0.13	- 0.86*
4	1	18	1.53 ± 0.21	1.92
4	2	18	0.46 ± 0.20	0.62*
5	1	18	1.53 ± 0.20	1.94
5	2	18	0.44 ± 0.20	0.61*
6	1	18	1.64 ± 0.22	2.00*
6	2	18	0.50 ± 0.22	0.89*
4	1	19	2.25 ± 0.23	3.07*
4	2	19	1.08 ± 0.23	1.62*
5	1	19	2.26 ± 0.23	3.09*
5	2	19	1.07 ± 0.23	1.60*
6	1	19	2.41 ± 0.25	3.12
6	2	19	1.63 ± 0.25	1.88

and of three-mode factor analysis of this data set, are summarized in Table IX. Again it appears that the data can be fitted better by the three-mode factor analysis model (even with only one factor for each mode) than by the analysis of variance model. With 3 factors for the solutes and 2 factors for the adsorbents and the eluents the standard deviation of the fit by the factor analysis model ($\hat{\sigma} = 0.08$) is already better than the standard deviation of the fit by regression analyses according to the Snyder equation (0.10 - 0.28). With the (3,2,4) model, $\hat{\sigma}$ is as low as 0.05, an impressive figure for these complicated solutes. From the data on $\hat{\sigma}_{pred.}$ in the last column of Table IX it follows that the (3,2,4) model is the best

Table IX

Summary of the results of analysis of variance and of three-mode factor analysis of the data, presented in Table V

Model	NFA	NFE	NFS	NPAR	DF	$\hat{\sigma}$	$\sqrt{Sp}$	$\hat{\sigma}_{pred.}$
Analysis of variance				42	171	0.58		
Factor analysis	1	1	1	42	171	0.38	0.42	0.47
" "	2	2	2	84	129	0.21	0.27	0.32
" "	2	2	3	122	91	0.08	0.12	0.14
" "	3	2	3	126	87	0.07	0.11	0.13
" "	3	2	4	164	49	0.05	0.11	0.11

Number of observations	:	213
Number of missing values	:	21
For other symbols	:	see Table VI

one to make predictions for the missing data in Table V*.

Detailed information on this model is given in Table X. Again it appears that the fit for each adsorbent, each eluent and each solute is about equally good.

This model has been used to calculate values for missing data, which are presented in Table XI, together with the 0.95-prediction intervals.

For the sake of comparison we made also predictions according to the Snyder equation (which is for these solutes more complicated than eqn. (3)). The predicted values are given in the last column of Table XI. Now we notice that in almost all cases predictions by the Snyder equation deviate significantly from the values, calculated by the (3,2,4) model (These cases are marked with an asterisk in Table XI).

---

[*] In the (3,2,3) and the (3,2,4) model the maximum number of factors for the adsorbents (3) and the eluents (2) is extracted. The consequence is that these two modes can be combined to a single adsorbent/eluent mode, so that the data can be classified in two modes. Accordingly, they can be analysed with two-mode factor analysis with identical results.

Table X

Fit of the three-mode factor analysis model (3,2,4) for the individual adsorbents, eluents and solutes in Table V

Adsorbents			Eluents			Solutes		
Code	NO	RMRSS	Code	NO	RMRSS	Code	NO	RMRSS
1	65	0.00	2	96	0.03	20	5	0.00
2	78	0.03	3	117	0.03	21	5	0.00
3	70	0.01				22	6	0.00
						23	6	0.00
						24	6	0.03
						25	6	0.03
						26	6	0.03
						27	6	0.03
						28	6	0.00
						29	6	0.05
						30	5	0.00
						31	5	0.00
						32	4	0.00
						33	5	0.00
						34	4	0.00
						35	4	0.00
						36	4	0.00
						37	4	0.00
						38	6	0.03
						39	6	0.00
						40	6	0.03
						41	6	0.00
						42	6	0.03
						43	6	0.00
						44	6	0.04
						45	6	0.00
						46	6	0.03
						47	5	0.00
						48	6	0.00
						49	6	0.04
						50	6	0.00
						51	6	0.03
						52	5	0.00
						53	5	0.03
						54	5	0.04
						55	5	0.03
						56	6	0.04
						57	5	0.03
						58	6	0.03

NO = number of observations

RMRSS = root of the mean residual sum of squares

Table XI

Predictions of the missing data in Table V by the three-mode factor analysis model (3,2,4) and their 0.95-intervals. Also given, for the sake of comparison, are predicions by Snyder's model

Missing datum			Predictions according to	
i	*k*	*l*	model (3,2,4)	Snyder
1	1	20	0.96 ± 0.22	1.10
1	1	21	0.81 ± 0.17	1.09*
3	1	30	2.58 ± 0.24	1.26*[a]
3	1	31	3.31 ± 0.28	1.62*[a]
1	1	32	1.78 ± 0.20	1.97
3	1	32	2.91 ± 0.27	1.11*[a]
3	1	33	3.12 ± 0.25	1.34*[a]
1	1	34	1.74 ± 0.19	1.95*
3	1	34	2.44 ± 0.26	1.16*[a]
1	1	35	1.79 ± 0.20	2.00*
3	1	35	2.73 ± 0.27	1.37*[a]
1	1	36	2.08 ± 0.21	2.44*
3	1	36	2.83 ± 0.28	1.28*[a]
1	1	37	2.17 ± 0.20	2.52*
3	1	37	3.14 ± 0.28	1.61*[a]
1	1	47	1.85 ± 0.18	2.35*
1	1	52	2.13 ± 0.18	2.59*
1	1	53	2.12 ± 0.17	2.48*
1	1	54	1.29 ± 0.18	*b*
1	1	55	1.32 ± 0.18	*b*
1	1	57	1.84 ± 0.18	*b*

a - calculated with unpublished β, ρ and δ values of the authors of reference 14

b - impossible to predict because of lacking A_s values for pyridine[11]

CONCLUSIONS

The Snyder equation, which has proven its value for bare silica and alumina adsorbents, is not quite satisfactory for fitting retention data in normal-phase HPLC on chemically bonded phases.

This equation is in fact a special case of the mathematical-statistical three-mode factor analysis model. This model, in its general form, appears to be able to fit data on the retention of 19 simple solutes (*viz.*, monosubstituted benzenes and unsubstituted polycyclic aromates) on 6 adsorbents with the eluents *n*-hexane and 35 v/v% methylene chloride in *n*-hexane very satisfactory. With only 3 parameters or "factors" for the solutes and the adsorbents, and 2 factors for the eluents, the standard deviation of the fit of log V_N/W data is only 0.05. This model is also able to fit a set of data on 39 more complicated solutes (*viz.*, substituted phenols, anilines and pyridines) on 3 adsorbents with the eluents 35 v/v % methylene chloride in *n*-hexane and pure methylene chloride. With 4 factors for the solutes, 3 factors for the adsorbents and 2 factors for the eluents, the standard deviation of the fit of the observations is only 0.05. The most important feature of a mathematical-statistical model is its ability to predict missing data. In the analysed sets of data, about 20% and 10% of the data are missing, respectively. The missing data can be predicted by three-mode factor analysis with a standard deviation of 0.10 and 0.11 respectively.

REFERENCES

1. H.H. Harman, Modern Factor Analysis, University of Chicago Press, Chicago (1970).

2. P. T. Funke, E.R. Malinovski, D.E. Martire and L.Z. Pollara, Separ. Sci., 1(1966)661.
3. P.H. Weiner and J.R. Parcher, J. Chromatogr. Sci., 10(1972)612.
4. S. Wold and K. Andersson, J. Chromatogr., 80(1973)43.
5. R.B. Selzer and D.G. Howery, J. Chromatogr., 115(1975)139.
6. D.H. Mc Closkey and S.J. Hawkes, J. Chromatogr. Sci., 13(1975)1.
7. C.L. de Ligny, G.H.E. Nieuwdorp, W.K. Brederode, W.E. Hammers and J.C. van Houwelingen, Technometrics, 23(1981)91.
8. L.R. Tucker, in: C.W. Harris(Ed.), Problems in Measuring Change, University of Winconsin Press, Madison (1963).
9. P.M. Kroonenberg and J. de Leeuw, Physchometrica 45(1980)69.
10. H.M. Weesie and J.C. van Houwelingen, GEPCAM Users' Manual, Institute of Mathematical Statistics, Utrecht, 1983.
11. L.R. Snyder, Principles of Adsorption Chromatography, Dekker, New York (1968).
12. W.E. Hammers, R.H.A.M. Janssen, A.G. Baars and C.L. de Ligny. J. Chromatogr. 167(1978)273.
13. W.E. Hammers, C.H. Kos, W.K. Brederode and C.L. de Ligny, J. Chromatogr., 168(1979)9.
14. W.E. Hammers, M.C. Spanjer and C.L. de Ligny, J. Chromatogr., 174(1979)291.
15. W.E. Hammers, A.G.M. Theeuwes, W.K. Brederode and C.L. de Ligny J. Chromatogr., 234(1982)321.
16. J.C. van Houwelingen, Principal components of large matrices with missing elements, Proceedings of the 3rd Prague Symposium on Asymptotic Statistics, North-Holland Inc., to appear in 1984. (momentally available as preprint nr. 310).
17. L. Breiman and D. Freedman, J. Amer. Stat. Ass., 78(1983)131.

APPENDIX

Mathematical aspects of three-mode principal components analysis

A.0. Introduction

In this appendix we describe a method which models a three-dimensional data matrix. Such a data set can be considered as consisting of an additional mode called 'layer' to the rows and columns of a two-dimensional matrix. We describe firstly the mathematical background of the model, and secondly the estimation of parameters for a complete data set by the three-mode principal components analysis.

The program GEPCAM (GEneralized Principal Components Analysis with Missing values) (1) computes these estimates also for very incomplete cases. An additional program is able to calculate also prediction intervals for missing values. Further we describe a criterion with which one can choose the optimum model.

A.1. Presentation of the model

Consider a three-dimensional data matrix with elements

$$X_{ijk} = M_{ijk} + \varepsilon_{ijk}\ , \text{ where } \begin{array}{l} i = 1, \ldots, I, \\ j = 1, \ldots, J \text{ and} \\ k = 1, \ldots, K. \end{array}$$

The random error ε_{ijk} fulfills the requirements $E\varepsilon_{ijk} = 0$, the variance of $\varepsilon_{ijk} = \sigma^2$ and ε_{ijk} values are mutually independent.

With three-mode principal components analysis we model the data by

$M_{ijk} = \sum_{p=1}^{P} \sum_{q=1}^{Q} \sum_{r=1}^{R} L_{pqr} A_{pi} B_{qj} C_{rk}$, where L is the core matrix and A, B and C are component matrices. The elements in the core matrix are a measure for the importance of the combination of the three factors. The rows of the component matrices $A_{P\times I}$ $(P \leqq I)$, $B_{Q\times J}$ $(Q \leqq J)$ and $C_{R\times K}$ $(R \leqq K)$ are the principal components or factors.

The representation of M_{ijk} is not unique. M_{ijk} does not change if A is replaced by $\tilde{A}_{pi} = \sum_{p'} T_{pp'} A_{p'i}$ and L by $\tilde{L}_{pqr} = \sum_{p'} T^{-1}_{p'p} L_{p'qr}$, where T is an arbitrary P×P matrix and $TT^{-1} = I_p$. This transformation can be made in such a way that the rows of A are orthonormal. Similarly, the rows of B and C can be made orthonormal.

Then for the first mode we can write:

$$V_{ii'}^{(1)} = \sum_{jk} M_{ijk} M'_{i'jk} = \sum_{pp'} A_{pi} A_{p'i'} W_{pp'} ,$$

where $W_{pp'} = \sum_{qr} L_{pqr} L_{p'qr}$.

Hence the rows of A span the same space as the rows of $V^{(1)}$. Therefore we can choose A in such a way that the rows are the first P orthonormal eigenvectors of $V^{(1)}$, and $W_{pp'} = 0$ for $p \neq p'$. Similarly B and C can be chosen such that they contain eigenvectors of $V^{(2)}$ and $V^{(3)}$, respectively:

$$V_{jj'}^{(2)} = \sum_{ik} M_{ijk} M'_{ij'k} \text{ and } V_{kk'}^{(3)} = \sum_{ij} M_{ijk} M'_{ijk'} .$$

With this choice for A, B and C it can be shown that

$L_{pqr} = \sum_{ijk} M_{ijk} A_{pi} B_{qj} C_{rk}$, since $\sum_i A_{pi} A_{p'i} = 1$ if $p = p'$ or zero if $p \neq p'$. It also holds that $\sum_{qr} L_{pqr} L_{p'qr} = W_{pp'} = 0$ if $p \neq p'$.

Similar relations can be derived for the second and third mode.

The relation above can be read as a description of P orthogonal vectors in a Q×R dimensional space. If $P > Q \times R$, this is only possible when a number of vectors are zero. The latter ones are omitted, so that only a number of $Q \times R$ non-zero vectors remain. In the special case of $R = 1$, we have $P \leqq Q$, $Q \leqq P$, hence $P = Q$ and $L_{pq1} = 0$ if $p \neq q$.

A.2. Estimates of L, A, B and C in the complete case

For a complete three-mode data matrix X, the model parameters are estimated by the least squares method, solving the minimization of

$\sum_{ijk} (X_{ijk} - \sum_{pqr} L_{pqr} A_{pi} B_{qj} C_{rk})^2 = RSS$, where RSS is the residual sum of squares.

As initialisation values we choose the eigenvectors

of $\sum_{jk} X_{ijk} X_{i'jk}$, $\sum_{ik} X_{ijk} X_{ij'k}$, $\sum_{ij} X_{ijk} X_{ijk'}$ and $L_{pqr} = \sum_{ijk} X_{ijk} A_{pi} B_{qj} C_{rk}$. Then we improve A_{pi} , with $i = 1, \ldots\ldots, i_o, \ldots\ldots, I$ if we fix the index column by column. So, for a certain column i_o we can write:

$$\sum_{jk} [X_{i_o jk} - \sum_{p=1}^{P} A_{pi_o} \sum_{qr} L_{pqr} B_{qj} C_{rk}]^2$$

If one reads $X_{i_o jk}$ as y, A_{pi_o} as b and $\sum_{qr} L_{pqr} B_{qj} C_{rk}$ as x, this clearly resembles a least squares problem. The term $\sum_{qr} L_{pqr} B_{qj} C_{rk}$ depends on p, j and k and can be expressed as U_{pjk}. This reveals the following minimization:

$$\left| \begin{pmatrix} X_{i_o 11} \\ X_{i_o 12} \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ X_{i_o JK} \end{pmatrix} - \begin{pmatrix} U_{111} & U_{211} & \cdots\cdots & U_{P11} \\ U_{112} & U_{212} & \cdots\cdots & U_{P12} \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ U_{1JK} & U_{2JK} & \cdots\cdots & U_{PJK} \end{pmatrix} \begin{pmatrix} A_{1i_o} \\ A_{2i_o} \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ A_{Pi_o} \end{pmatrix} \right|^2$$

In this way the component matrix A is columnwise improved. The same procedure is adapted to B and C. Thereafter the core matrix L is calculated by minimization of:

$$\left| \begin{pmatrix} X_{111} \\ X_{112} \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ X_{IJK} \end{pmatrix} - \begin{pmatrix} A_{11} B_{11} C_{11} & A_{11} B_{11} C_{21} & \cdots & A_{P1} B_{Q1} C_{R1} \\ A_{11} B_{11} C_{12} & A_{11} B_{11} C_{22} & \cdots & A_{P1} B_{Q1} C_{R2} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ A_{1I} B_{1J} C_{1K} & A_{1I} B_{1J} C_{2K} & \cdots & A_{PJ} B_{QJ} C_{RK} \end{pmatrix} \begin{pmatrix} L_{111} \\ L_{112} \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ L_{PQR} \end{pmatrix} \right|^2$$

After each iteration step RSS is further reduced, until a, hopefully global, minimum is reached.

The order of the iteration cycle used in GEPCAM is : A, L, B, L, C, L, which empirically turned out to be the best.

A.3. Estimates of L, A, B and C in the incomplete case

In case of missing data the procedure remains in principle the same. The summation is then over the available observations, with the restriction that at each regression step the number of observations is not less than the number of parameters to be computed. In this case we start with randomly generated L, A, B and C matrices.

The procedure can be improved by computing least squares estimates for the missing data by means of the more simple analysis of variance model, which accounts only for the main effects. A missing value X_{ijk} is estimated by $\hat{M}_{ijk} = \hat{\mu} + \hat{\alpha}_i + \hat{\beta}_j + \hat{\gamma}_k$, where $\sum_i \hat{\alpha}_i = \sum_j \hat{\beta}_j = \sum_k \hat{\gamma}_k = 0$, and $\hat{\mu}$, $\hat{\alpha}$, $\hat{\beta}$, and $\hat{\gamma}$ are obtained by minimization of

$$\sum_{ijk} (X_{ijk} - \mu - \alpha_i - \beta_j - \gamma_k)^2 .$$

Afterwards, the completed matrix can be used to obtain initial values for L, A, B and C, as described in A.2.

A.4. Estimation of the variance of the prediction

The number of estimated free parameters NPAR can be computed by:

$\text{NPAR} = \text{NFR} \times I + \text{NFC} \times J + \text{NFL} \times K + P \times Q \times R - P^2 - Q^2 - R^2$, where

NFR = number of row factors, here denoted by P,

NFC = number of column factors, here denoted by Q, and

NFL = number of layer factors, here denoted by R.

The common variance of the error component for the three-mode matrix is estimated by $\hat{\sigma}^2 = \text{RSS}_{\text{conv.}}/\text{DF}$, where $\text{RSS}_{\text{conv.}}$ = RSS after convergence has occurred, and DF=degrees of freedom = number of observations - NPAR. The estimate of $\hat{\sigma}$ permits us to compare the fit of the models with different numbers of factors.

A.5. Model selection

Achieving the smallest possible standard deviation of the model is not the most important feature of the goodness of fit. Predicting values for missing data most accurately is a more valuable point. This can be judged by the so called S_p criterion (2), defined as

$S_p = \hat{\sigma}^2 (1 + \frac{p}{n - p - 1})$, where p = the number of estimated parameters (here denoted as NPAR) and n = the number of observations (here given by (I × J × K) - (number of missing data)).

The quantity S_p is an estimate of the variance of (new observation - prediction).

A more accurate method is to compute the variance of (new observation - prediction) for each missing observation, and to use the average value (averaged over all missing data) as a criterion to select the best model.

A.6. The calculation of prediction intervals

The method by which prediction intervals for $\hat{M}_{ijk}$ are calculated is based on the extension of the procedure that is applied to two-mode cases (3).

Considering the estimates $\hat{M}_{ijk}$ as a linearization around the parameters, we start with computing the derivatives of $\hat{M}_{ijk}$ with respect to the parameters.

Filling a matrix Z with all $\frac{\partial \hat{M}_{ijk}}{\partial PAR}$ values for all observed M_{ijk} , we calculate the variance of $\hat{M}_{ijk}$ by $\sigma^2 D'(Z'Z)^{-}D$ as introduced by Rao (4), where ' indicates a transpose, and $^{-}$ indicates a generalized inverse matrix. D denotes the matrix of derivatives of $\hat{M}_{ijk}$ with respect to the parameters for all missing values.

We will illustrate this by the following example: Let $P = Q = R = 1$, $I = J = K = 2$ and X_{ijk} be given for $i = j = k = 1,2$ with exception of X_{222} . We then calculate

$$\begin{array}{c} Z = \\ \\ D' = \end{array}\left(\begin{array}{ccccccc} A_1B_1C_1 & LB_1C_1 & 0 & LA_1C_1 & 0 & LA_1B_1 & 0 \\ A_1B_1C_2 & LB_1C_2 & 0 & LA_1C_2 & 0 & 0 & LA_1B_1 \\ A_1B_2C_1 & LB_2C_1 & 0 & 0 & LA_1C_1 & LA_1B_2 & 0 \\ A_1B_2C_2 & LB_2C_2 & 0 & 0 & LA_2C_2 & 0 & LA_1B_2 \\ A_2B_1C_1 & 0 & LB_1C_1 & LA_2C_1 & 0 & LA_2B_1 & 0 \\ A_2B_1C_2 & 0 & LB_1C_2 & LA_2C_2 & 0 & 0 & LA_2B_1 \\ A_2B_2C_1 & 0 & LB_2C_1 & 0 & LA_2C_1 & LA_2B_2 & 0 \\ \hdashline A_2B_2C_2 & 0 & LB_2C_2 & 0 & LA_2C_2 & 0 & LA_2B_2 \end{array}\right)$$

The columns of Z represent derivatives of M_{ijk} with respect to L , A_1 , A_2 , B_1 , B_2 , C_1 , and C_2 , respectively.

The part under the dotted line represents the derivatives of M_{222} with respect to the parameters, to be calculated as D.

We can now calculate the value of Z'Z and combine it to

$\sigma^2_{\text{pred.}} = \hat{\sigma}^2[1 + D' \ (Z'Z)^- \ D]$, which will be illustrated by the following numerical example: Let $A_1 = B_1 = C_1 = 0.600$, $A_2 = B_2 = C_2 = 0.800$ and $L = 2.000$, for which we obtain :

$$Z = \begin{pmatrix} 0.216 & 0.72 & 0 & 0.72 & 0 & 0.72 & 0 \\ 0.288 & 0.96 & 0 & 0.96 & 0 & 0 & 0.72 \\ 0.288 & 0.96 & 0 & 0 & 0.72 & 0.96 & 0 \\ 0.384 & 1.28 & 0 & 0 & 0.96 & 0 & 0.96 \\ 0.288 & 0 & 0.72 & 0.96 & 0 & 0.96 & 0 \\ 0.384 & 0 & 0.96 & 1.28 & 0 & 0 & 0.96 \\ 0.384 & 0 & 0.96 & 0 & 0.96 & 1.28 & 0 \end{pmatrix} \quad \text{and } D = \begin{pmatrix} 0.512 \\ 0 \\ 1.28 \\ 0 \\ 1.28 \\ 0 \\ 1.28 \end{pmatrix}$$

Then $D' \; (Z'Z)^{-} \; D = 2.384$ and the variance of $(X_{222} - \hat{X}_{222})$ $= \hat{\sigma}^2(1 + 2.384)$, so $\sigma_{\text{pred.}}$ for $X_{222} = 1.840 \; \hat{\sigma}$.

A.7. References

1. H.M. Weesie and J.C. van Houwelingen, GEPCAM User's Manual, Institute of Mathematical Statistics, Utrecht, 1983, (available on request).
2. L. Breiman and D. Freedman, J. Am. Statist. Ass., 78(1983)131.
3. J.C. van Houwelingen, Principal components of large matrices with missing elements, Proceedings of the 3rd Prague Symposium on Asymptotic Statistics, North Holland Inc., to appear in 1984, (momentally available as preprint nr. 310).
4. C.R. Rao, Linear Statistical Inference and its applications, J. Wiley, New York, 2nd edition, 1973, p. 224.

SUMMARY

The main subjects of this dissertation are the investigation of a substituent interaction effect in disubstituted benzenes in reversed-phase high-performance liquid chromatography (RP-HPLC) and a mathematical-statistical description of retention in normal-phase liquid chromatography.

In chapter I and II the influence of the interaction between substituents on the retention behaviour of benzoic acids, phenols, anilines and pyridines in RP-HPLC is investigated. Buffered methanol-water, acetonitrile-water and tetrahydrofuran-water eluents were used with an octadecylsilylsilica adsorbent. Net retention volumes of monosubstituted benzoic acids, phenols, anilines, pyridines and, as a reference series, benzenes were measured. For a similar purpose we carried out batch partition experiments with monosubstituted benzenes, phenols and pyridines, with the corresponding n-hexadecane/eluent systems. These experiments also gave us more information concerning the retention mechanism in RP-HPLC. From these net retention volumes and batch partition coefficients a substituent interaction effect was derived. It was described with the linear substituent free energy relationships of Taft or Nieuwdorp. The values of the ρ parameters for the chromatographic measurements and for the batch partition experiments were positive. This holds as well for the n-hexadecane/eluent systems as for literature data on octanol-water systems. The signs do not correspond with the signs of ρ parameters obtained from literature data on normal-phase liquid chromatography and gas-liquid chromatography. The positive signs are explained by assuming that all solutes act as proton acceptors in hydrogen bonding with the, acidified, mixed aqueous-organic eluents. The data on chromatographic

retention are not well correlated with the data on partition between n-hexadecane and the eluent. Therefore, partition between the eluent and the bound alkyl layer, saturated with the eluent, is not an important contribution to the retention mechanism. In methanol-water we assume adsorption on the bound alkyl layer to be the prevailing retention mechanism. The retention in acetonitrile-water eluents, and the retention of substituted benzenes in tetrahydrofuran-water eluents is an order of magnitude larger than corresponds with a partition mechanism.

In chapter III the ability of three recently proposed linear substituent free energy relationships (LSFER's) to describe data on substituent effects by regression analysis has been compared. The Taft equation was used as a reference LSFER. A number of 211 series of data were selected from literature that could not be described satisfactorily by this equation. To explore the scope of applicability of the most successful relationship, data were selected on the effects of ortho-, meta- and para substituents on chemical and physical properties (like proton, fluorine and carbon-13 NMR shifts, absorption frequencies in the visible and infrared region of the spectrum, and ionization potentials), and phase equilibria, in rigid saturated systems and small as well as extended π-systems. The medium ranged from water to apolar organic solvents and even to the gas phase. The standard deviation of the fit of these data by the various LSFER's decreases in the following order: Taft ≈ Swain and Lupton > Yukawa and Tsuno > Nieuwdorp et al.. It can thus be concluded that the latter equation can be applied most universally.

In chapter IV we classified equilibrium and reaction rate constants with respect to three modes: solvent, reaction type and substituent and described simultaneously the influence of these three variables by the mathematical-statistical three-mode factor analysis model. Applications of this technique to literature data and to our own measurements in chapters I en II are given. The first example concerns literature data on ionization constants for 15 series of substituted compounds in 3 solvents. The second example concerns data on phase equilibrium constants of 6 series of substituted compounds in 9 two-phase systems, comprising both gas-liquid as well as liquid-liquid systems. The residual standard deviation of the observations is about equal to that obtained by a regression analysis model, viz., by applying the Taft equation to each reaction in each solvent separately. However, the number of parameters in the factor analysis model is far less than that in the regression analysis model. The advantage of the factor analysis model is, that it is better suited for the prediction of missing data (in the first example 53% of the data are missing). It can even predict data that cannot be predicted at all with the Taft equation, viz., for reactions on which no data are available in a particular solvent, which is the case for 20% of the data in the first example.

In chapter V we applied three-mode factor analysis to data on the retention in normal-phase HPLC. It was shown that the Snyder equation, which has proven its value for bare silica and alumina adsorbents, is less satisfactory for fitting retention data in normal-phase HPLC on chemically bonded phases. This equation is in fact a special case of the mathematical-statistical three-mode factor analysis model. The latter model, in its general form, appeared to be able to fit data on the retention of 19 simple solutes (viz., monosubstituted benzenes and unsubstituted polycyclic aromates) on 6 adsorbents with the eluents n-hexane and 35 v/v% methylene chloride in n-hexane very satisfactorily. With only 3 parameters or "factors" for the solutes and the adsorbents, and 2 factors for the eluents, the standard deviation of the fit of $\log V_N/W$ data is only 0.05. This model also fitted a set of data on 39 more complicated solutes (viz., substituted phenols, anilines and pyridines) on 3 adsorbents with the eluents 35 v/v % methylene chloride in n-hexane and pure methylene chloride. With 4 factors for the solutes, 3 factors for the adsorbents and 2 factors for the eluents, the standard deviation of the fit of the observations is only 0.05. The most important feature of the mathematical-statistical model is its ability to predict missing data. In the analysed sets of data, about 20% and 10% of the data are missing, respectively. The missing data can be predicted by three-mode factor analysis with a standard deviation of 0.10 and 0.11 respectively.

Finally, an appendix elucidates the mathematical aspects of the three-mode factor analysis model.

SAMENVATTING

De voornaamste onderwerpen van dit proefschrift zijn een onderzoek van het substituent interactie effect in digesubstitueerde benzenen in de hoge druk vloeistofchromatografie met apolaire adsorbentia (RPHPLC) en een mathematisch-statistische beschrijving van de retentie in vloeistofchromatografie met polaire adsorbentia (NPHPLC).

In de eerste twee hoofdstukken wordt de invloed van de interactie tussen substituenten op het retentiegedrag van benzoëzuren, phenolen, anilines en pyridines in RPHPLC onderzocht. Gebufferde methanol-water, acetonitril-water en tetrahydrofuraan-water mengsels werden gebruikt als eluentia, met octadecylsilylsilica als adsorbens. De netto retentie volumina van monogesubstitueerde benzoënzuren, phenolen, anilines, pyridines en als referentie serie, benzenen, werden gemeten. Met hetzelfde doel (onderzoek van interacties tussen substituenten) werden verdelingsexperimenten uitgevoerd met monogesubstitueerde benzenen, phenolen en pyridines, in de overeenkomstige n-hexadecaan/eluens systemen. Deze experimenten verschaften ons meer inzicht in het retentiemechanisme bij RP-HPLC. Het substituent interactie effect dat uit deze netto retentievolumina en verdelingscoëfficiënten werd afgeleid, werd beschreven met de lineaire vrije energie relaties voor substituenten van Taft of Nieuwdorp. De waarden van de ρ parameters voor de chromatografische metingen en voor de verdelingsexperimenten waren positief. Dit geldt zowel voor de n-hexadecaan/eluens systemen als voor literatuurwaarden voor octanol-water systemen. Het teken van de ρ parameters komt niet overeen met de tekens die worden gevonden voor literatuurwaarden in NP-HPLC en gas-vloeistof chromatografie. Het positieve teken van de ρ waarden kan worden verklaard met de hypothese dat alle stoffen optreden als proton acceptoren bij de waterstofbrugvorming met de aangezuurde, uit mengsels van water en organisch oplosmiddel bestaande, eluentia. Er bestaat geen duidelijk verband tussen de grootte van de chromatografische retentie en de waarde van de verdelingscoëfficiënt voor de verdeling tussen n-hexadecaan en het eluens. Blijkbaar levert de verdeling tussen het eluens en de gebonden alkyllaag, verzadigd met het eluens, geen belangrijke bijdrage aan het retentie mechanisme. Adsorptie op de gebonden alkyllaag is waarschijnlijk het overheersende retentie mechanisme in methanol-water. De retentie in acetonitril-water en die van gesubstitueerde benzenen in tetrahydrofuraan-water eluentia is een orde groter dan overeen zou komen met een verdelingsmechanisme.

In hoofdstuk III worden de mérites van drie onlangs voorgestelde lineaire vrije energierelaties voor substituenten (LSFER's) vergeleken. Als referentie LSFER werd de Taft vergelijking gebruikt. Uit de literatuur werden 211 getallenseries geselecteerd die niet bevredigend met deze vergelijking konden worden beschreven. Om het toepassingsgebied van de best bruikbare vergelijking af te bakenen werden gegevens verzamelt betreffende effecten van ortho-, meta- en para substituenten op chemische en fysische eigenschappen (zoals proton, fluor en koolstof-13 NMR verschuivingen, absorptie frequenties in het zichtbare en infrarode gebied van het spectrum en ionisatiepotentialen) en fase evenwichten, zowel in kleine, starre verzadigde systemen, als in kleine en uitgebreidere π-systemen. Het medium varieerde van water, via apolaire organische oplosmiddelen, tot de gas fase. De volgorde waarin de standaardafwijking van de beschrijving van deze gegevens met de verschillende LSFER's afnam, was: Taft ≈ Swain en Lupton > Yukawa en Tsuno > Nieuwdorp e.a.. Hieruit kan worden geconcludeerd dat laatstgenoemde vergelijking het meest universeel toepasbaar is.

In hoofdstuk IV werden evenwichts- en reactiesnelheidsconstanten in de drie categorieën oplosmiddel, reactie type en substituent ingedeeld en de invloed van deze drie variabelen werd gelijktijdig beschreven met het mathematisch-statistische factoranalyse model voor drie-indelingen. Toepassingen van deze methode op literatuurgegevens en op eigen metingen zoals weergegeven in de eerste twee hoofdstukken worden gegeven. Het eerste voorbeeld betreft literatuurwaarden van ionisatieconstanten van 15 series gesubstitueerde verbindingen in 3 oplosmiddelen. Het tweede voorbeeld betreft getalwaarden van verdelingscoëfficiënten van 6 series gesubstitueerde verbindingen in 9 twee-fase systemen, die zowel gas-vloeistof als vloeistof-vloeistof systemen omvatten. De residuele standaardafwijking van de waarnemingen is ongeveer gelijk aan de waarde die verkregen wordt bij toepassing van regressieanalyse, b.v. door toepassing van de Taft vergelijking op elke reactie in elk oplosmiddel afzonderlijk. Het aantal parameters in het factoranalyse model is echter veel kleiner dan in het regressieanalyse model. De grotere geschiktheid voor het voorspellen van ontbrekende gegevens (in het eerste voorbeeld: 53% van de gevallen) is het grote voordeel van het factoranalyse model. Het kan zelfs gegevens voorspellen die absoluut niet voorspeld kunnen worden met de Taft vergelijking, b.v. voor reacties waarvoor geen enkel gegeven voor een bepaald oplosmiddel beschikbaar is, wat het geval is in 20% van de gevallen in het eerste voorbeeld.

In hoofdstuk V werd factoranalyse met drie-indelingen toegepast op getalwaarden voor de retentie in de NP-HPLC. Er werd aangetoond dat de vergelijking van Snyder, die zijn waarde bewezen heeft voor silica en aluminium adsorbentia, minder geschikt is om retentiegegevens in de NP-HPLC met chemisch gebonden fasen te beschrijven. In feite is deze vergelijking een bijzonder geval van het mathematisch-statistische factoranalyse model voor drie-indelingen. Laatstgenoemd model bleek, in zijn algemene formulering, retentiegegevens van 19 eenvoudige stoffen (zoals monogesubstitueerde benzenen en ongesubstitueerde polyaromaten) op 6 adsorbentia, in de eluentia n-hexaan en 35 v/v% methyleenchloride in n-hexaan, goed te beschrijven. De residuele standaardafwijking van de beschrijving van log V_N/W gegevens bedraagt slechts 0.05, bij gebruik van slechts 3 parameters of "factoren" voor de stoffen en de adsorbentia, en 2 factoren voor de eluentia. Ook gegevens voor 39 ingewikkelder stoffen (zoals gesubstitueerde phenolen, anilines en pyridines) op 3 adsorbentia in de eluentia 35 v/v% methyleenchloride in n-hexaan en zuivere methyleenchloride konden met dit model goed worden beschreven. De residuele standaardafwijking bedraagt slechts 0.05 bij gebruik van 4 factoren voor de stoffen, 3 factoren voor de adsorbentia en 2 factoren voor de eluentia. Het vermogen om ontbrekende getalwaarden te voorspellen is het belangrijkste kenmerk van het mathematisch-statistische model. In de onderzochte getalseries ontbraken ongeveer 20%, resp. 10% van de getalwaarden. De ontbrekende gegevens kunnen door het factoranalyse model met drie-indelingen voorspeld worden met standaardafwijkingen van 0.10 en 0.11, resp.

Een appendix heldert tenslotte de wiskundige punten van het factoranalytische model met drie-indelingen op.

CURRICULUM VITAE

De schrijver van dit proefschrift werd op 5 oktober 1956 geboren te Veenendaal. De basis van de nu voltooide opleiding werd gelegd op de lagere school, waaraan zijn vader als hoofd verbonden is. Het voorbereidend wetenschappelijk onderwijs werd van september 1968 tot juni 1974 gevolgd aan het Christelijk Streeklyceum te Ede. Het diploma Atheneum-B, in de vakken Scheikunde, Wiskunde 1 en 2, Natuurkunde, Biologie, Engels en Nederlands, werd behaald op 14 juni 1974. Per september 1974 werd aan de Rijksuniversiteit te Utrecht begonnen met de scheikundestudie. Het kandidaatsexamen S2 werd afgelegd op 7 november 1977. Het doctoraalexamen chemie met hoofdvak analytische scheikunde (dr. W.E. Hammers en prof. dr. C.L. de Ligny), bijvak toxicologie (dr. R.H. Drost en prof. dr. R.A.A. Maes) en de aantekening onderwijsbevoegdheid (dr. R.A.W. Docters van Leeuwen en prof. dr. R.S. Mossel) werd afgelegd op 10 maart 1980. Het genoegen waarmee laatstgenoemd diploma werd uitgereikt kwam tot uitdrukking in het daaropvolgend aangeboden wetenschappelijk assistentschap. Het in dat kader uitgevoerde promotiewerk werd op het Analytisch Chemisch Laboratorium verricht van 15 maart 1980 tot 15 maart 1984. Voor wat betreft de wiskundige aspecten van de rekenkundige werkzaamheden werd samengewerkt met het Instituut voor Mathematische Statistiek. Na de promotie zal hij zijn militaire dienstplicht gaan vervullen.