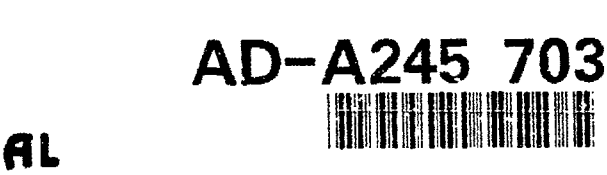


CHEMICAL
RESEARCH,
DEVELOPMENT &
ENGINEERING
CENTER

CRDEC-SP-009

MEETING ON SOLUTE/SOLVENT INTERACTIONS

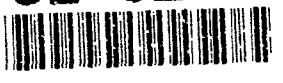
George R. Famini

RESEARCH DIRECTORATE

January 1992

U.S. ARMY
ARMAMENT
MUNITIONS
CHEMICAL COMMAND

Aberdeen Proving Ground, Maryland 21010-5423



Disclaimer

The findings in this report are not to be construed as an official Department of the Army position unless so designated by other authorizing documents.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	1992 January	Final, 91 May - 91 May

4. TITLE AND SUBTITLE
Meeting on Solute/Solvent Interactions

5. FUNDING NUMBERS
PR-1C16266A553L

6. AUTHOR(S)
Famini, George R.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
CDR, CRDEC, ATTN: SMCCR-RSP-C,
APG, MD 21010-5423

8. PERFORMING ORGANIZATION REPORT NUMBER
CRDEC-SP-009

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words)

The First U.S. Army Chemical Research, Development and Engineering Center (CRDEC) meeting on Solute/Solvent Interactions was held at CRDEC, 29-30 May 1991. The meeting brought together researchers from government and academia who have been addressing problems associated with developing models for solvent effects. This report is a compilation of the papers presented at this meeting.

14. SUBJECT TERMS
Solute/solvent interactions
Linear free energy relationships

15. NUMBER OF PAGES
160

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UL

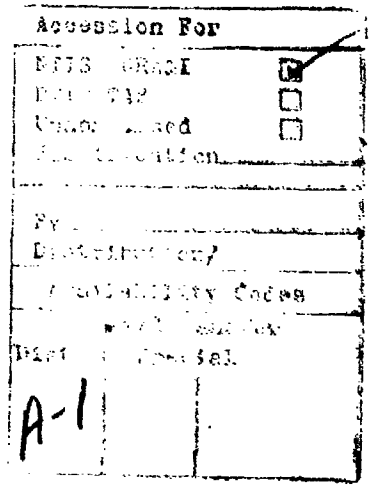
Blank

PREFACE

The work described in this report was authorized under Project No. 1C162622A553L, CB Defense Assessment Technology. This work was started and completed in May 1991.

The use of trade names or manufacturers' names in this report does not constitute an official endorsement of any commercial products. This report may not be cited for purposes of advertisement.

This report has been approved for release to the public.



BLANK

CONTENTS

BLANK

INTRODUCTION

The *Meeting on Solute/Solvent Interactions* represents the first such endeavor at the U.S. Army Chemical Research, Development and Engineering Center (CRDEC). All told, there were 16 presenters from CRDEC, Naval Research Laboratory, Environmental Protection Agency, and academic institutions. In addition, four international scientists were in attendance.

This meeting was modeled around the "Laguna Beach" meetings organized by Mortimer J. Kamlet in the mid to late 1980s. In those meetings, as well as these, there formal presentations by each of the invited attendees. In addition, there were numerous periods for informal discussions. This report contains the written version of most of the presentations.

Over the two day period, the presentations were divided into 4 basic session areas: Equilibrium, Theoretical, New Approaches, and Polymers. Many topic areas applicable to solute/solvent interactions were covered. Most prevelant were applications invloving the generalized linear solvation energy relationship (LSER), and new methods to measure the necessary parameters. In addition, theoretical approaches were presented that apply both *ab initio* and semi-empirical quantum chemical methodologies to the LSER. Two new computer programs were also presented dealing with the prediction of solute/solvent properties. The first, *QSAR*, is a compilation and thesaurus of quantitative structure activity relationships (QSAR) compiled over the years. The second, *SPARC*, is an attempt to provide relevant, accurate predictions of properties for EPA related projects.

Finally, I would like to thank all of the attendees to donating their time to share their research. Without their contributions, this meeting would not have been possible.

George R. Famini
Meeting Chairman

BLANK

Meeting on Solute/Solvent Interactions
Wednesday, 29 May, 1991

Inherent Structural and Solvent Effects on Proton Transfer and Hydrogen-Bond Acidities

R.W. Taft, Department of Chemistry, University of California, Irvine

Recent ICR studies of gas phase acidities in our laboratory have been extended to include (neutral organic) proton transfer acids of very great acidity. In particular, the trifilate substituent, SO_2CF_3, and its derivatives have been investigated. The gas phase proton transfer acidities can be compared to those in dimethyl sulfoxide. The introductory overhead illustrates the great differences that are found between $pK_a = \frac{\Delta G^\circ}{1.364}$ ($HA \rightleftarrows H^+ + A^-$; ΔG°) in the gas, dimethyl-sulfoxide (hereafter given by s or Me_2SO) and aqueous phases for familiar acids. The very large differences of pK_a values in solution phases compared to gas phase result from the reaction of H^+ with solvent and the ion solvation. However, very significant ΔpK_a differences for pair wise comparisons of acids (the corresponding proton transfer equilibria) are also to be noted between the three different phases. These ΔpK_a differences in solution compared to the corresponding gas phase ΔpK_a reveal the differentials between the free energies of solvation of anions and neutral acids. For different families of acids the ΔpK_a values frequently are widely different, whereas, at least qualitatively, it is less common to find different acidity orders within a given family.

These generalities are illustrated in the second overhead for several series of trifilated and related acids by the gas and dimethylsulfoxide phase pK_a values. Also illustrated is the fact that successive substitutions into methane can product much stronger acids than the corresponding single substitution in the water molecule. For example, $(CF_3SO_2)_3CH$ is 4.6 powers of ten stronger than trifilic acid (CF_3SO_2OH) in the gas phase. An interesting inversion from the intrinsic gas phase order is shown for the acidities of $(CF_3SO_2)_2CH_2$ (the stronger) and $(CF_3SO_2)_2NH$ (the weaker) in Me_2SO. A probable explanation is given in the third overhead. Such inversions are common for general comparisons of pK_a(gas) vs pK_a(s). In order to help single out the differentials in the solvation energies of a given HA/A^- pair, Taft and Bordwell (reference given in the first overhead) have used the proton transfer equilibria shown in overhead 4, in which 9-C_6H_5 fluorene and its conjugate anion represent a series of very poorly non-specifically solvated acid/base pairs. These reference unsaturated hydrocarbons have no

heteroatoms and have highly charged delocalized carbanions, a structural feature that leads to no effect of Me_2SO on their relative acidities (relative acidities are the same for gas and Me_2SO phase). Plots of $\Delta pK_a(g)$ (or $\Delta\Delta G°(g)$) vs $\Delta pK_a(s)$ (or $\Delta\Delta G°(s)$) for the 9-Ph fluorene reference then reveal HA/A^- specific solvation effects as deviations from the line of unit slope (overhead 4)). When A^- solvation by dipolar (but non-HBD) Me_2SO exceeds dipolar and $AH\cdots OSMe_2$ H-bonding solvation of its conjugate acid, large deviations to the right of the line of unit slope are observed (as for H_2O or CH_3OH, for example). The points for acids that lie to the left of the line of unit slope (eg CF_3SO_2OH), involve relatively strong H-bonding solvation of HA and decreased anionic stabilization by dipolar substituents that occurs in dipolar solvents as compared to the gas phase.

The equilibrium in a gas phase proton transfer reaction is determined only by intrinsic structural effects (there are no counter ion or solvation effects). The corresponding solution phase reaction has superimposed upon the main proton transfer equilibrium a set of solvation equilibria. Hydrogen-bonding equilibria of the type $AH + OSMe_2 \rightleftarrows AH\cdots OSMe_2$ are important contributors to $pK_a(s)$ values and to the direction and magnitude of the kinds of deviations illustrated in overhead 4. This follows from the fact that structural effects on proton transfer acidity and corresponding hydrogen-bond donor acidities in general are simply unrelated. This is shown in the highly disordered plot of overhead 5 which compares currently available α_2^H values with corresponding available gas phase proton transfer $\frac{\Delta G°_{HA} - \Delta G°_{HI}}{1.364}$ values. Even within families of acids, deviations from a linear relation between a corresponding series of $\frac{\Delta G°_{HA} - \Delta G°_{HI}}{1.364}$ and α_2^H, values are common. Clearly the structural features that particularly stabilize anions are nearly totally missing for corresponding $AH\cdots B$ hydrogen-bonded complexes. We must exclude hydrogen-bonded ion pairs from this conclusion since their structural and solvent stabilizations are affected in a totally different manner than for the corresponding HB complexes.

One of the important kinds of anion stabilization that occurs in both HB-D and in certain non-HB-D solvents (such as Me_2SO) has been called substituent solvation assisted resonance (SSAR) effects. Overhead 6 illustrates these effects for the proton transfer acidities of a series of m- and p-substituted phenols relative to phenol. Pi electron delocalization from the anionic center (here phenoxide) to certain (SSAR) para pi electron acceptor substituents create new solvation sites in these "ambident" anions.

Substituent site solvation by this electronic mechanism requires significant localization of pi electronic charge at a solvent accessible electronegative atom of the substituent. Sharing of the delocalized charge among many such atoms reduces the substituent solvation to the point that it may be negated by the entropy loss contribution to solvation free energies. Accordingly the SSAR substituents in decreasing order of importance are: p-NO, p-NO_2, p-SO_2CF_3, p-CH_3CO p-CO_2CH_3, CN. Substituents such as p-CF_3, p-SCF_3, p-SC_6H_5, p-$SOCH_3$, p-SO_2CH_3 and all meta substituents have been recognized as non-SSAR.

Very recently completed work with Yagupol'skii's superacceptor substituent in the p-position of aniline is shown in overhead 7. Koppel's group has determined the corresponding pK_a(s) for this highly unusual aniline. In a plot of ΔpK_a(g) vs ΔpK_a(s) for m- and p-substituted anilines (not shown) SSAR effects are significantly larger than those shown for the corresponding phenols in overhead 6. The superacceptor substituent $SO(CF_3)NSO_2CF_3$ has a very large σ_F value (1.09) and a quite significant σ_R^- value (.31). In view of the number of resonance forms that may be written in which pi electronic charge is delocalized from NH^- of the anilide ion to electronegative atoms of this substituent, it is very reasonable to expect and it is found in the ΔpK_a(g) vs ΔpK_a(s) plot that this para superacceptor gives the same magnitude of SSAR effect as does the para SO_2CF_3. For the trifilate substituent the σ_F and σ_R^- values (.84 and .21, resp.) are both smaller.

The structures that lead to gas phase superacidities (that is, ΔG_{acid} values of less than 300 kcal/mole or pK_a less than 220) have many potentially useful properties. One example comes from the work of Profs. D.D. Des Marteau and E. Yeager. The substructure shown in overhead 8 has been incorporated into polymers and these are found to have unusually good electrical conductivities.

BLANK

New Developments in Solute Hydrogen Bonding Basicity Scales

Michel BERTHELOT

Laboratoire de spectrochimie moléculaire
Université de NANTES
2 , rue de la Houssinière
44072 NANTES CEDEX 03
FRANCE

Hydrogen bonding basicity is one of the major intrinsic properties of a solvent , together with its polarity and its hydrogen bonding acidity . Corresponding β , π^* and α scales have been carefully elaborated[1] and the multilinear solvatochromic equation :

$$XYZ = (XYZ)_0 + s\pi^* + a\alpha + b\beta$$

has proved useful in rationalizing solvent effects in many physico-chemical properties and in organic as well as bio-organic reactivities . However , with practice , the Linear Solvation Energy Relationships have brought to light some limitations of the available data in the field of hydrogen bonding and have necessitated statistical co-additions of various measurements involving several probes , different temperatures and many solvents.

Since 1981 , we have been interested in overcoming one aspect of these limitations : the lack of a general H.B. basicity scale of solute molecules . Two different thermodynamic (pK_{HB}) and spectroscopic ($\Delta\nu$(OH)) H.B. scales of solute bases have been simutaneously re-examined and developed by means of FTIR spectroscopy . The definition of these scales is outlined as :

p-fluorophenol + Base ⟷ Complex (CCl_4 25°C)

$K = (Complex) / (p\text{-fluorophenol})(Base)$

$pK_{HB} = \log K$

MeOH + Base ⟷ MeOH···Base (CCl_4 20°C)

$\Delta\nu(OH) = \nu(OH) - \nu(OH\cdots)$

THE SPECTROSCOPIC SCALE

Many authors have already suggested that the frequency shift of a reference RXH (fig 3) could provide a convenient evaluation of the H.B. strength . However , the selection of the probe is of primary importance because :

- the relationships between the different acids may be family dependent (fig 4) . Therefore , special care must be exercised when averaging data from several probes .

- phenols must not be used as reference acids , since it is impossible to get the maximum of their complexed absorption with all amines and with most pyridines (fig 5) .

We have selected methanol which gives accurate shifts from 20 cm^{-1} for halogen and π bases to 480 cm^{-1} for the strongest nitrogen bases .

The $\Delta\nu$ (OH) scale is found to be very sensitive to the covalent part of the hydrogen bonding interaction (fig 6) and practically insensitive to steric effects as shown in fig 7 where ortho , meta , and para substituted pyridines fall on the same line and where the only significant deviations are due to water solvation on some specific substituents . Another interesting feature of this scale is its ability to detect the original behavior of some oxygen acceptors which may present two stereoisomeric complexes (fig 8) .

THE THERMODYNAMIC SCALE

Our purpose was to extend the scale defined by Taft and co-workers in 1969 (fig 9) . We have now examined the characteristics of more than 400 bases , the pK_{HB} of which ranges (fig 10) from −0.85 to +2.6 .

The structural effects on this thermodynamic scale may be analysed in terms of the nature of the accepting atom and of the substituents linked to this group . A strong basicity requires two characteristics of the molecule :

- the accepting atom must have a high p character in order to favor the charge transfer interaction ,

- the molecule must have polar bonds in order to get a strong electrostatic stabilisation .

The predominance of the latter effect explains the order of basicity presented in fig 11 for nitrogen acceptors .

Inside the different families of bases , the substituent effects - steric , resonance , field/inductive , proximity , cyclisation , alkyl effects - play their usual roles (fig 12 - fig 15) . The optimum blend of some of these effects is found in a cyclic amidine (fig 15) which is the strongest molecular base known .

THE pK_{HB} vs $\Delta\nu$(OH) PLOT

Whenever the nature of the accepting atom or its state of hybridization is changed a new family is defined provided that no steric effect occurs from the substituents (fig 16) . Very important new families are presented such as :

- monomeric water and alcohols (fig 17) ,
- nitro and nitroso derivatives (fig 18) including some solvatochromic indicators and super basic nitro compounds ,
- aromatic , ethylenic and acetylenic families (fig 19) ,
- halogen bases (fig 20) ,
- anions (figs 21-22) ; in opposition to neutral bases these anions form a homogeneous family whatever the nature of the accepting group .

For a selected set of 22 bases , the $pK_{HB}/\Delta\nu$ (OH) plot and the graph between F2 and Fi vectors as taken from the study of Gal and co-workers are found to be equivalent (figs 23 and 24) .

THE POLYFUNCTIONALITY

Most molecules of biological interest are potentially polyfunctional toward H.B. donors because they bear many insaturations or heteroatoms (fig 25) . The polyfunctionality may be underscored on the ν (OH) spectrum of the acid (fig 26) and in some favorable cases in the spectrum of the base (fig 27) . When the constant is obtained from the intensity of the free absorption of the acid , it is the sum of the microscopic constants related to the different sites of the molecule . These microscopic constants may be obtained with a good accuracy from our $pK_{HB}/\Delta\nu$ (OH) plot and their sums are in good agreement with the experimental ones (figs 28 and 29) . With the help of this diagram , we have been able to draw general rules concerning the polyfunctionality :

- the first rule is that the proton sharing process is less selective than the proton transfer . All insaturations and heteroatoms usually behave as acceptor sites (fig 30) .

- the second rule concerns push-pull molecules . When direct conjugation occurs between two functional groups , the electron donating one is inactive toward H.B. donors , whatever the nature of the transmitting group (fig 30) . These rules may be applied on many molecules of biological interest (fig 31) .

1

HYDROGEN BONDING BASICITY OF SOLUTE MOLECULES IN TETRACHLOROMETHANE

2

F.T.I.R. SPECTROSCOPY

FREQUENCY SHIFT OF A REFERENCE ACID RX–H

$\Delta\nu(XH) = \nu(XH) - \nu(XH...)$

SPECTROSCOPIC SCALE

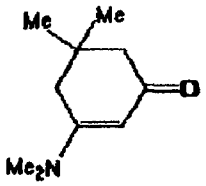
EQUILIBRIUM CONSTANT OF THE H.B. COMPLEX FORMATION

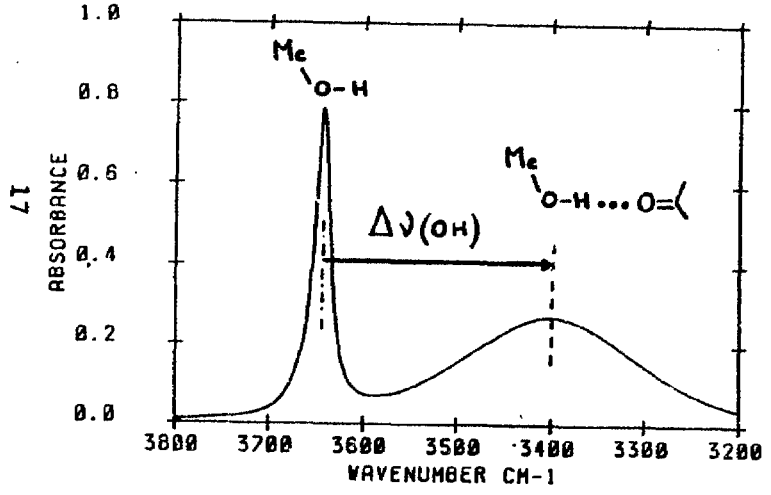
$$K_{HB} = \frac{(Complex)}{(Acid)(Base)}$$

THERMODYNAMIC SCALE

SOLVENT : CCl_4

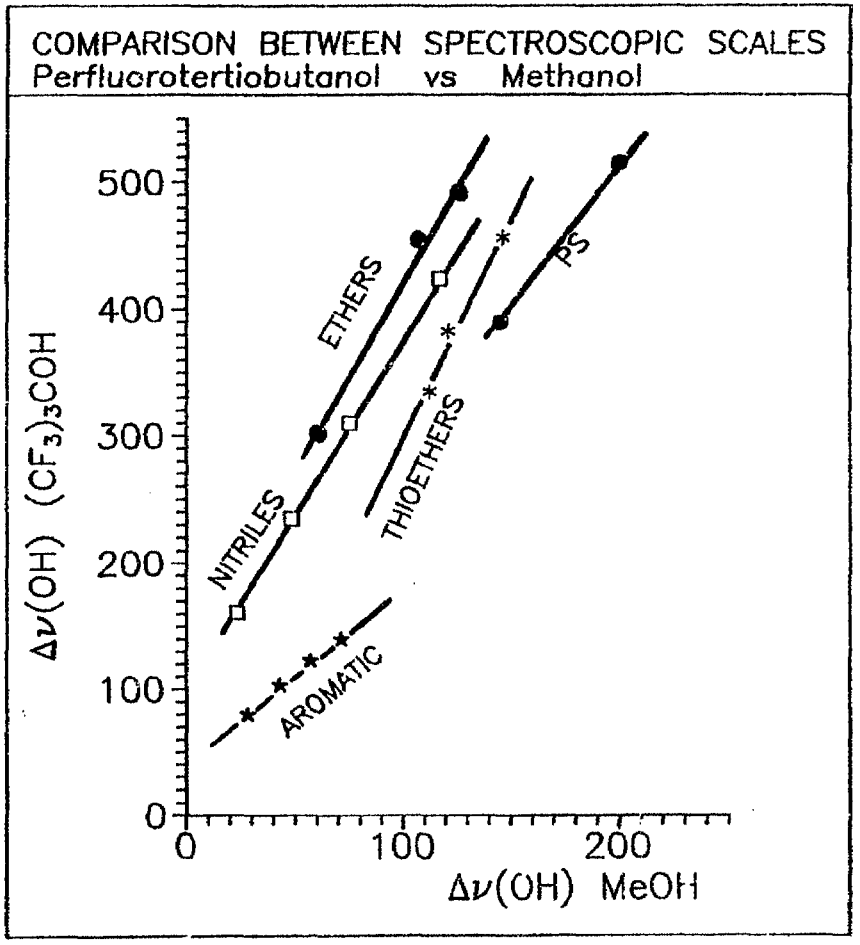
3


Me
Me
O
Me2N

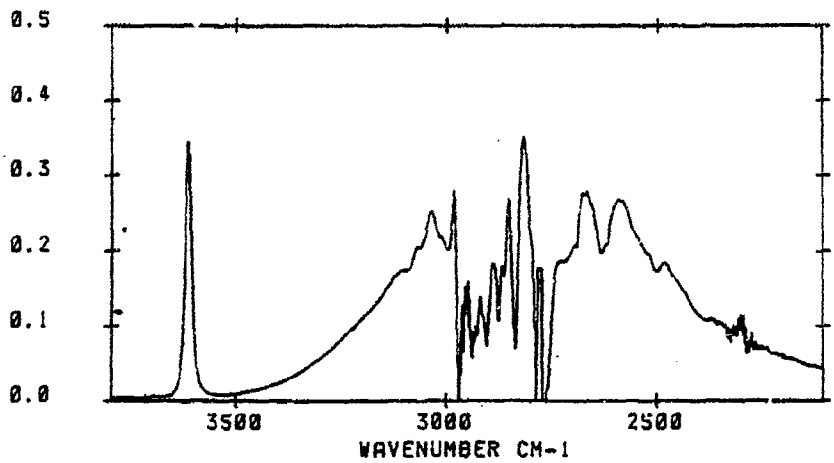

1.0
0.8
0.6
0.4
0.2
0.0
ABSORBANCE
Me
O-H
Δν(OH)
Me
O-H ••• O=(
3800
3700
3600
3500
3400
3300
3200
WAVENUMBER CM-1

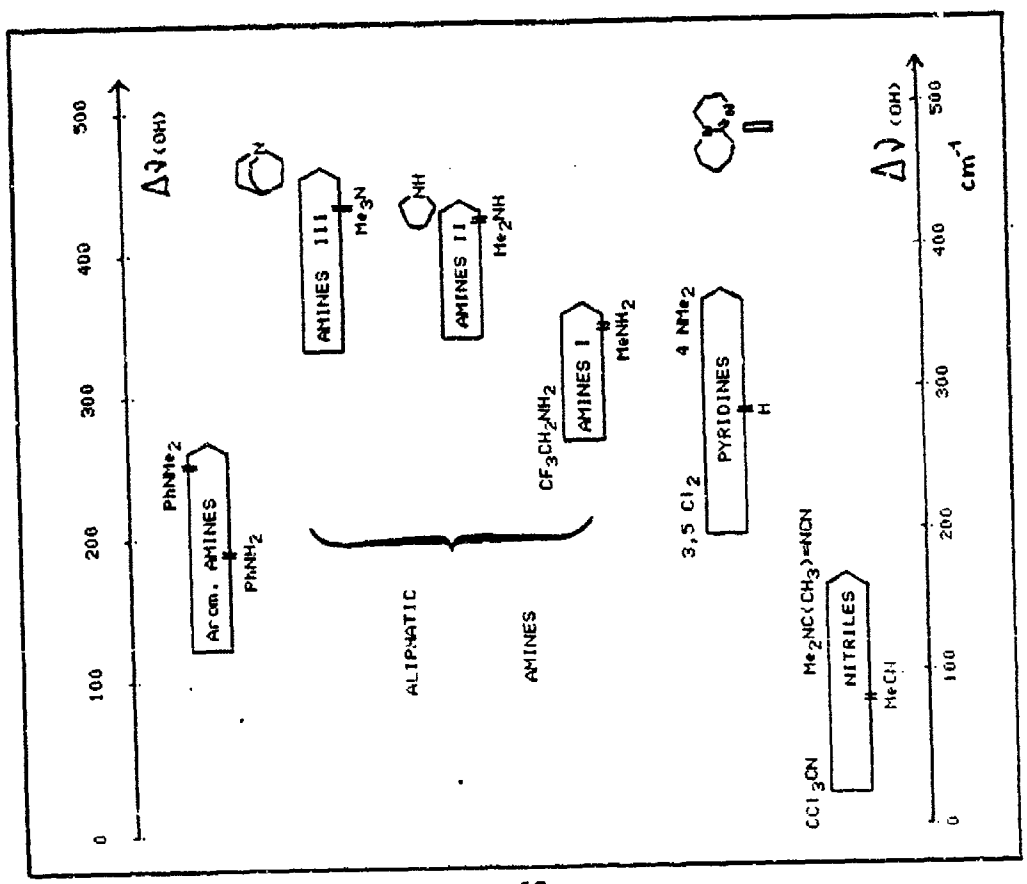
17

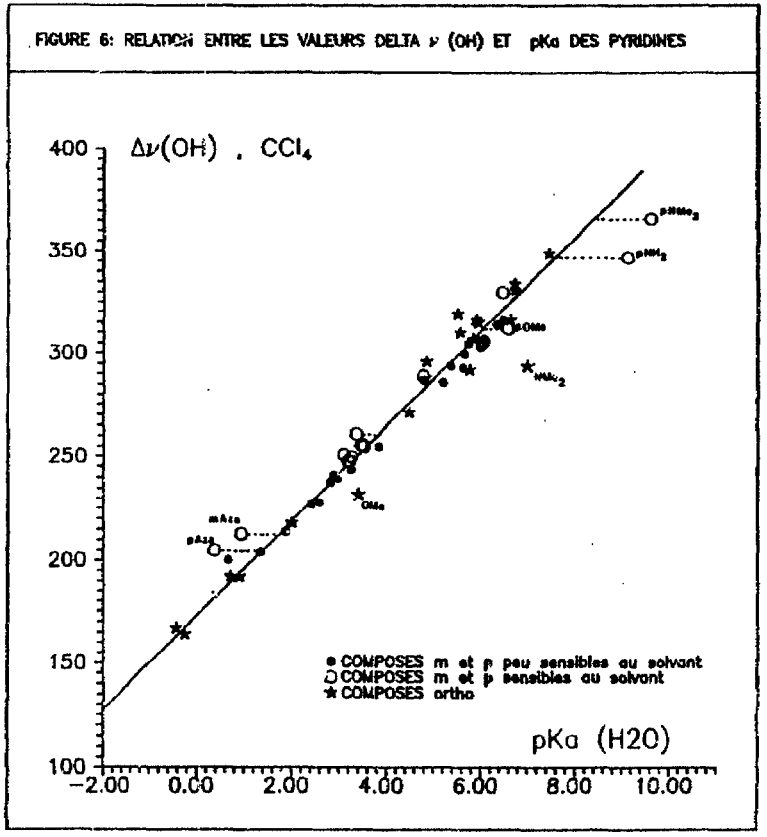
4


COMPARISON BETWEEN SPECTROSCOPIC SCALES
Perfluorotertiobutanol vs Methanol
Δν(OH) (CF3)3COH
500
400
300
200
100
0
ETHERS
PS
NITRILES
THIOETHERS
AROMATIC
0
100
200
Δν(OH) MeOH

p-fluorophenol / N-methyl pyrrolidine







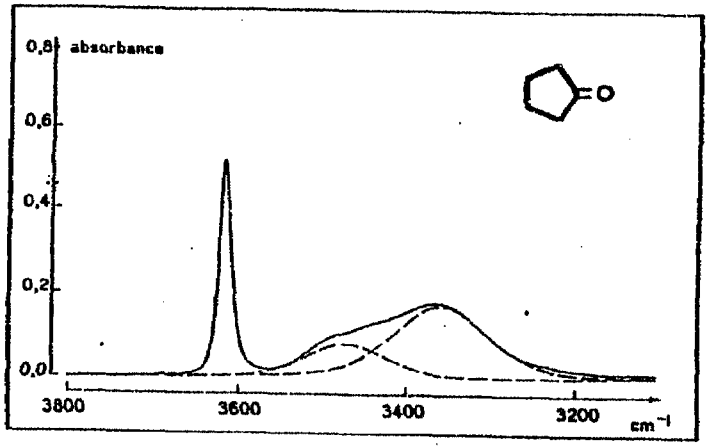


Figure 15. Spectre IR du p-fluorophénol (0,0045 M) en présence de cyclopentanone (0,0230 M).

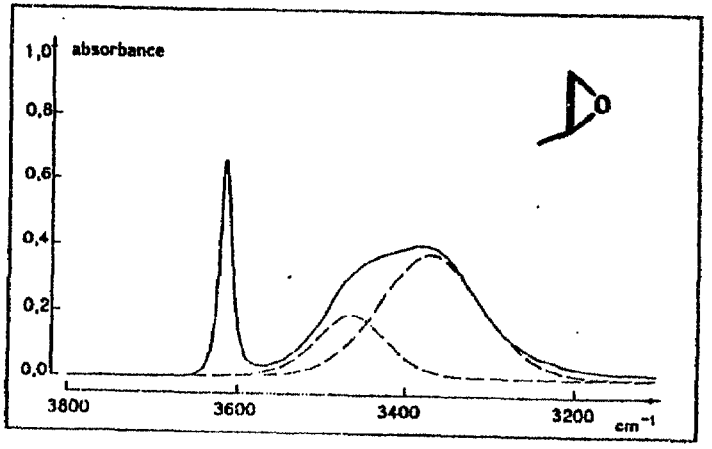
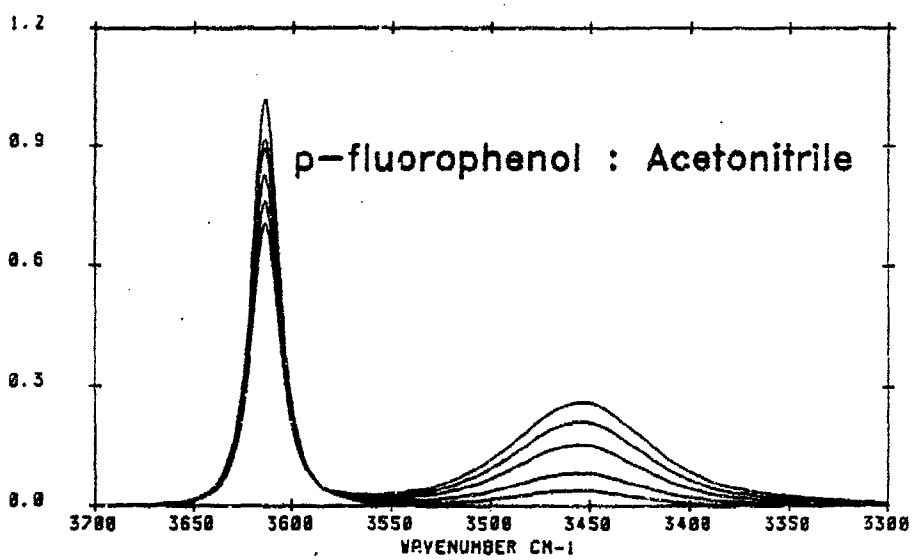


Figure 16. Spectre IR du p-fluorophénol (0,0045 M) en présence d'oxyde de propylène (0,1360 M).

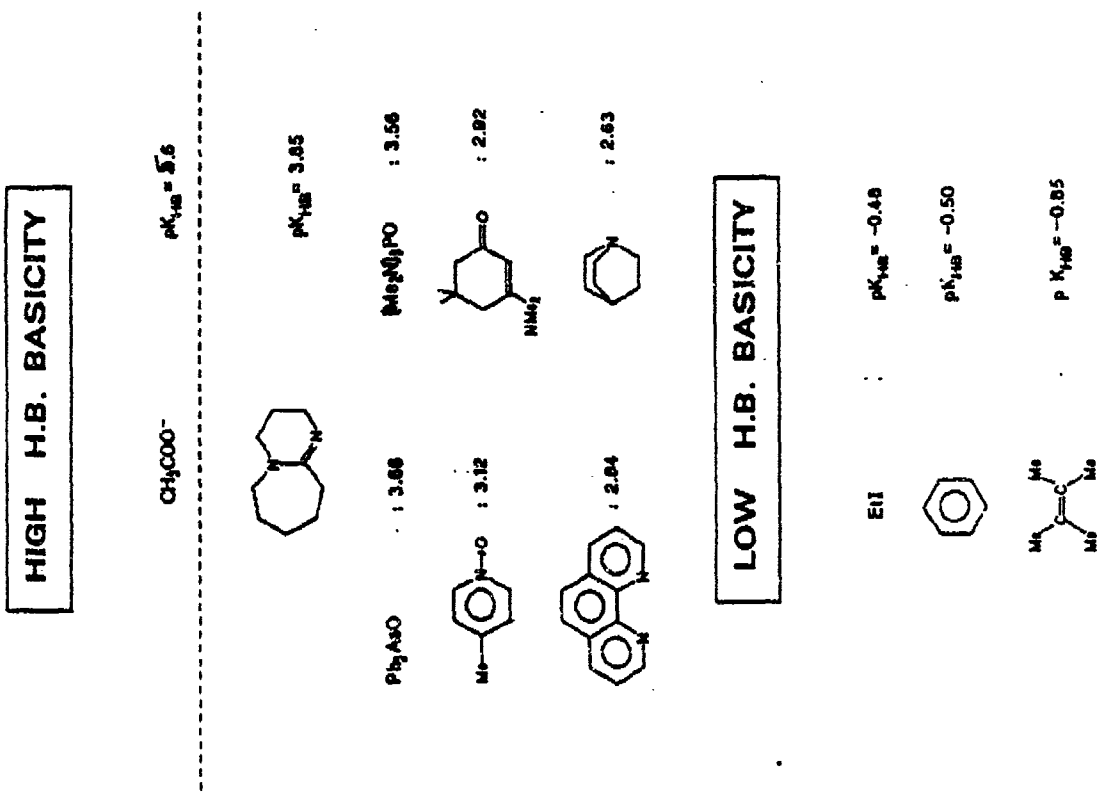
$$AH + B \underset{}{\overset{CCl_4,\ 298\ K}{\rightleftharpoons}} AH\cdots B$$

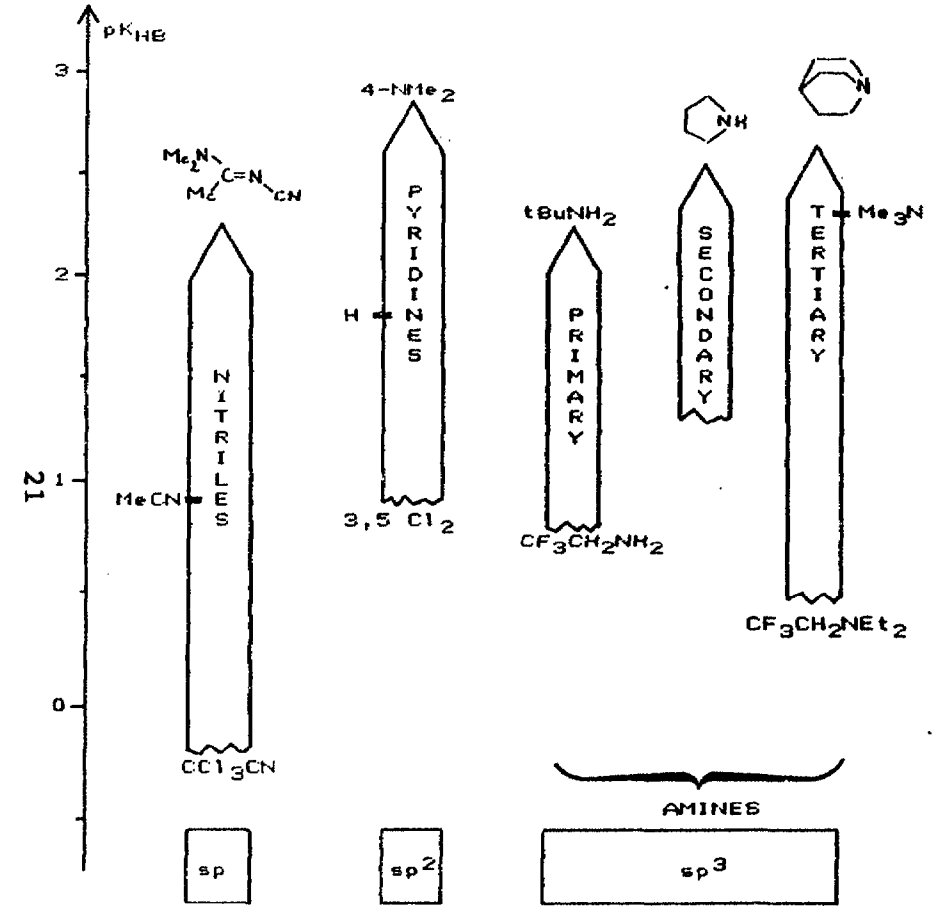
$$K = \frac{(AH\ldots B)}{(AH)(B)} \qquad pK_{HB} = \log K \quad \text{(p-fluorophenol)}$$

$$\beta_2^{H} = (pK_{HB} + 1.1) / 4.636$$



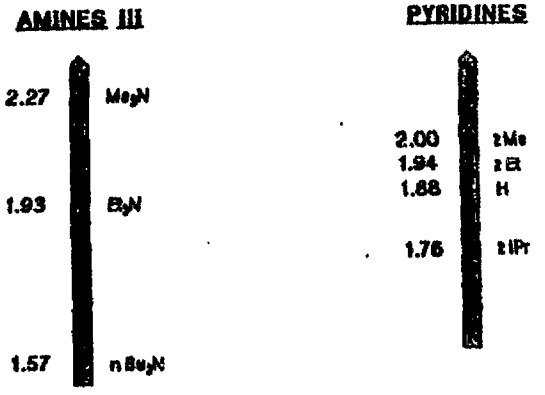
10



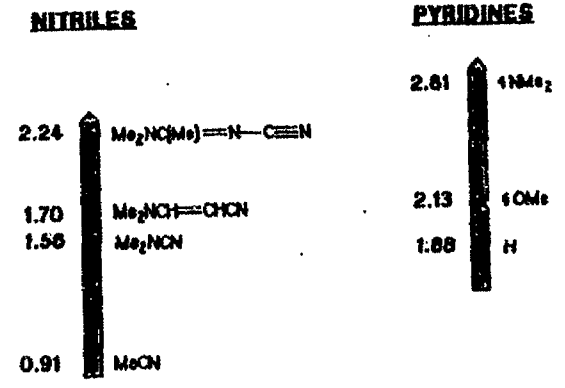


LONE PAIR WITH HIGH P CHARACTER	⟶	CHARGE TRANSFER INTERACTION
HIGHLY POLAR C-N BOND	⟶	ELECTROSTATIC STABILISATION

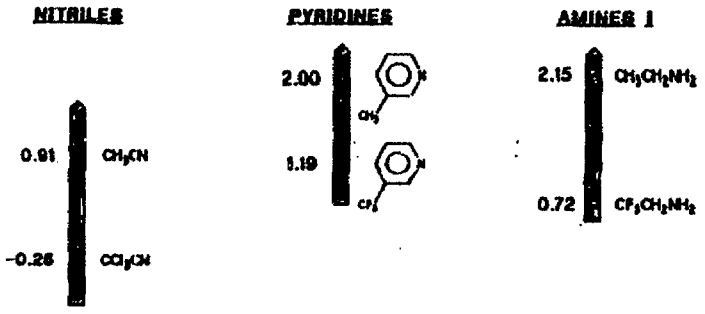
STERIC EFFECT



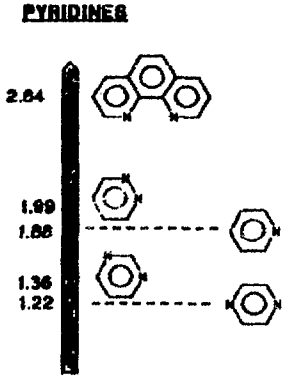
RESONANCE EFFECT



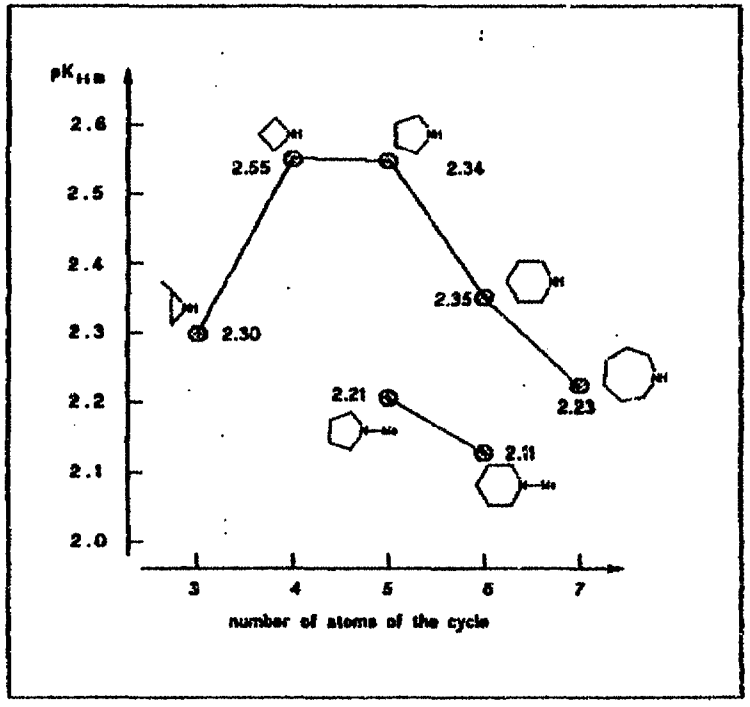
FIELD/ INDUCTIVE EFFECT



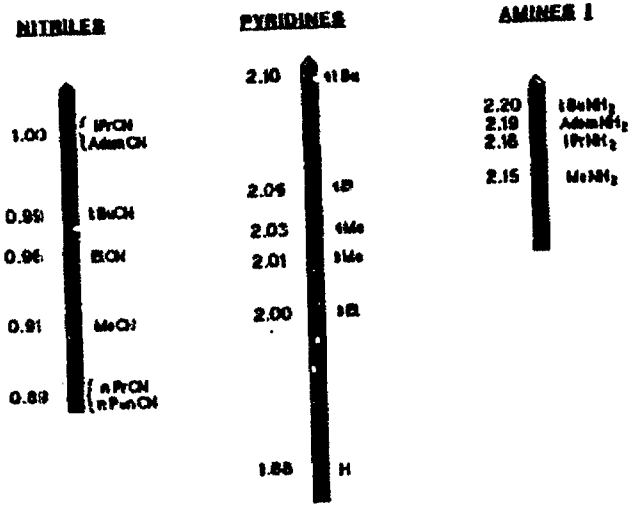
PROXIMITY EFFECT



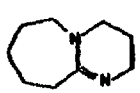
CYCLISATION



ALKYL SUBSTITUENTS

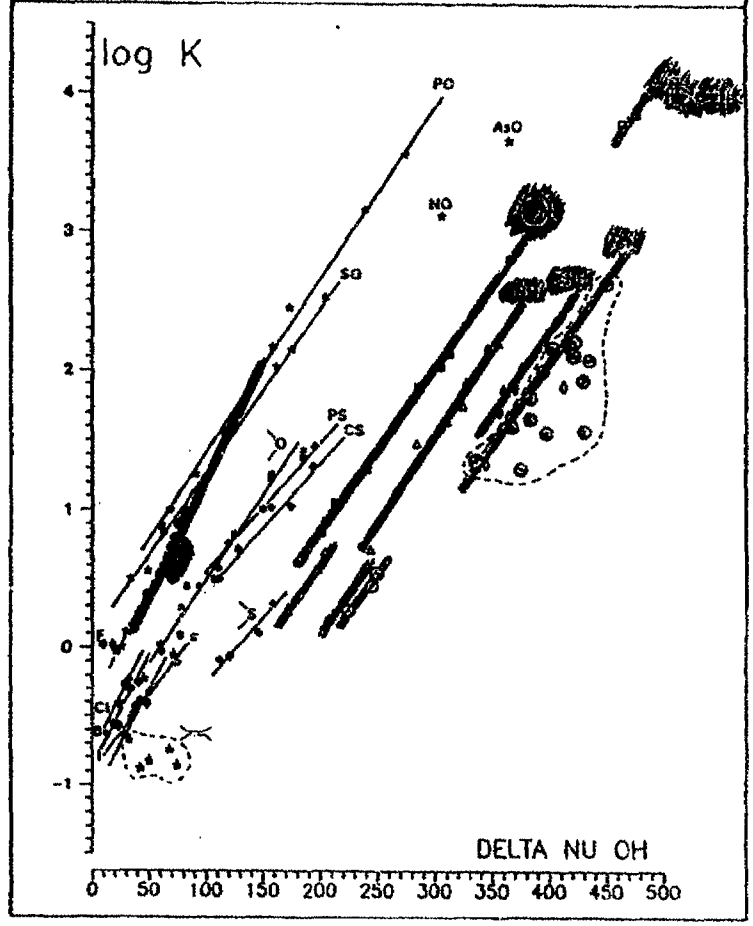


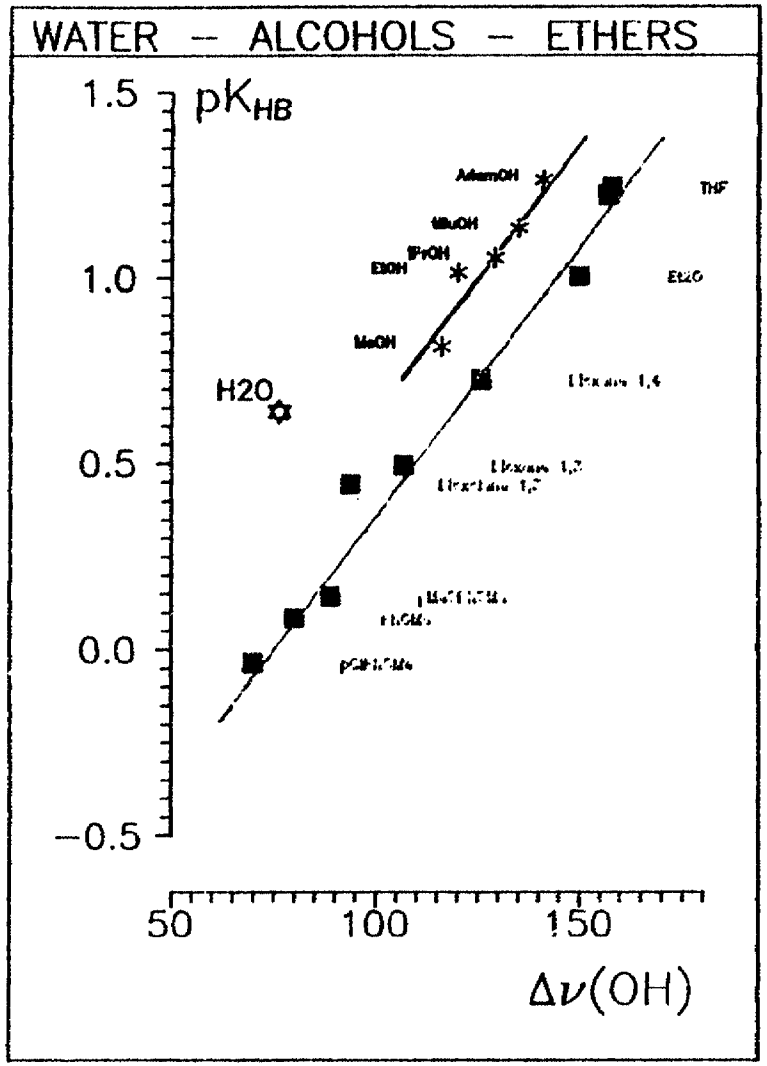
1,8 DIAZABICYCLO[5.4.0]UNDEC-7-ENE

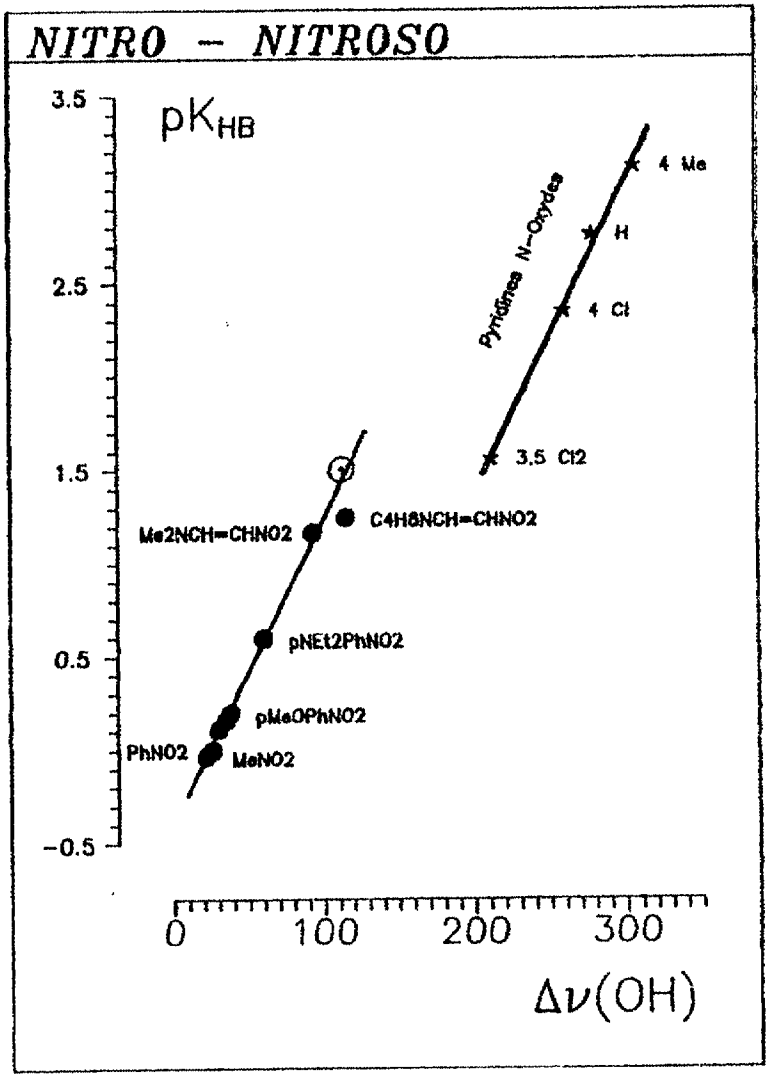


pK_{MeOH} = 3.85

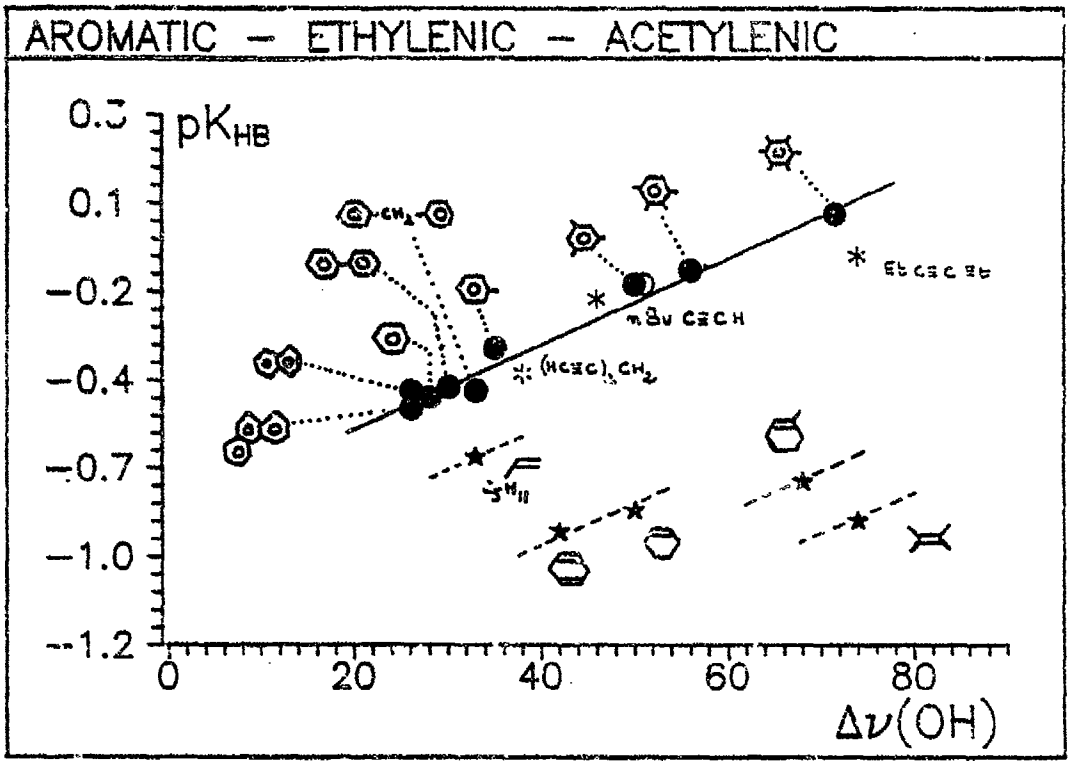
Figure 11 : Relations linéaires par familles entre pK_{HB} et Delta nu(OH) MeOH



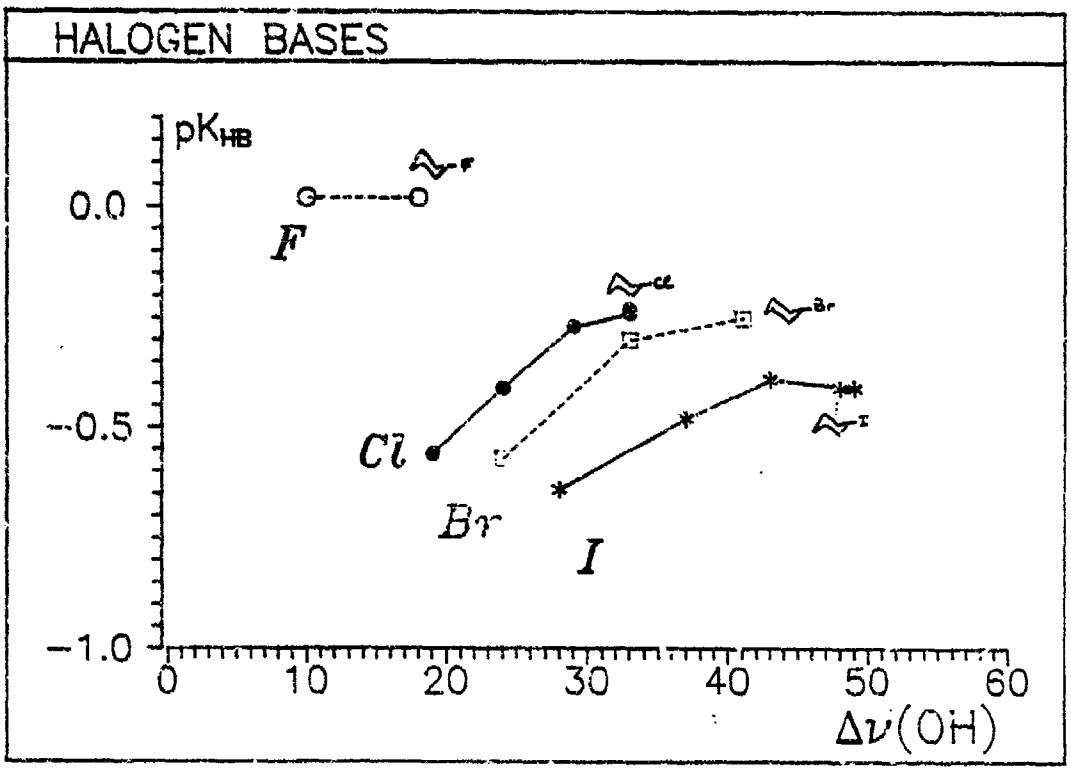

WATER – ALCOHOLS – ETHERS
pK$_{HB}$
1.5
1.0
0.5
0.0
−0.5
H2O
MeOH
EtOH
iPrOH
tBuOH
AdamOH
THF
Et2O
50
100
150
Δν(OH)

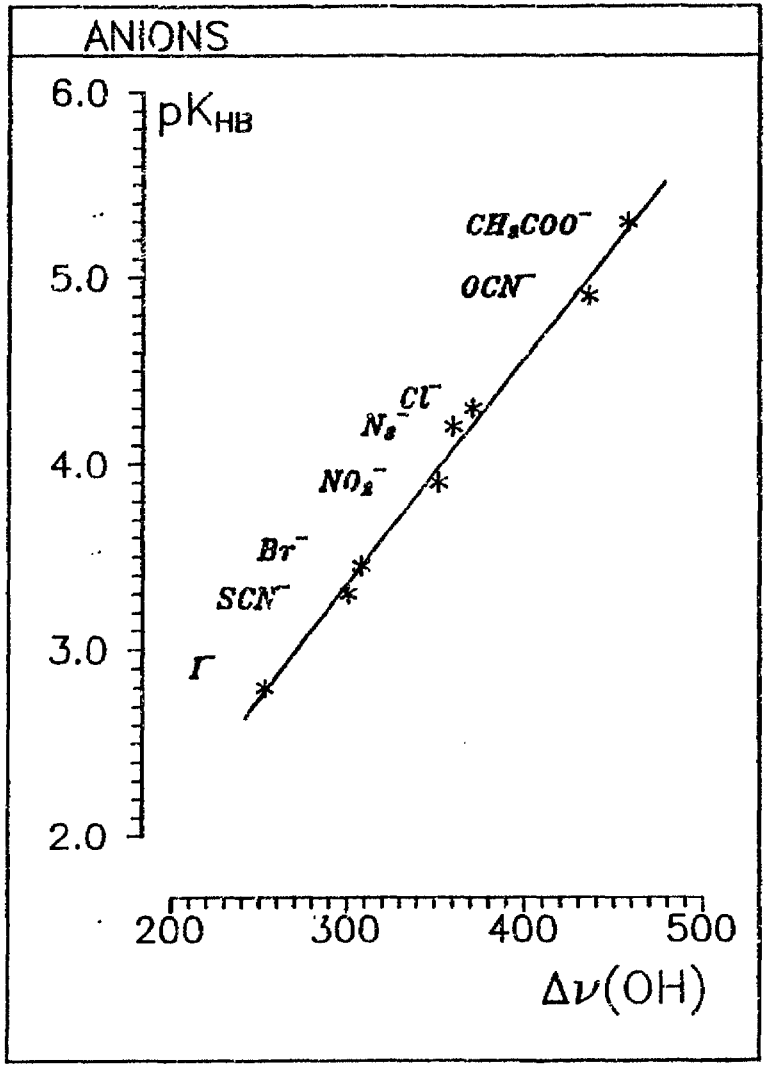

NITRO – NITROSO
pK$_{HB}$
3.5
2.5
1.5
0.5
−0.5
Pyridines N-Oxydes
4 Me
H
4 Cl
3.5 Cl2
Me2NCH=CHNO2
C4H8NCH=CHNO2
pNEt2PhNO2
pMeOPhNO2
PhNO2
MeNO2
0
100
200
300
Δν(OH)

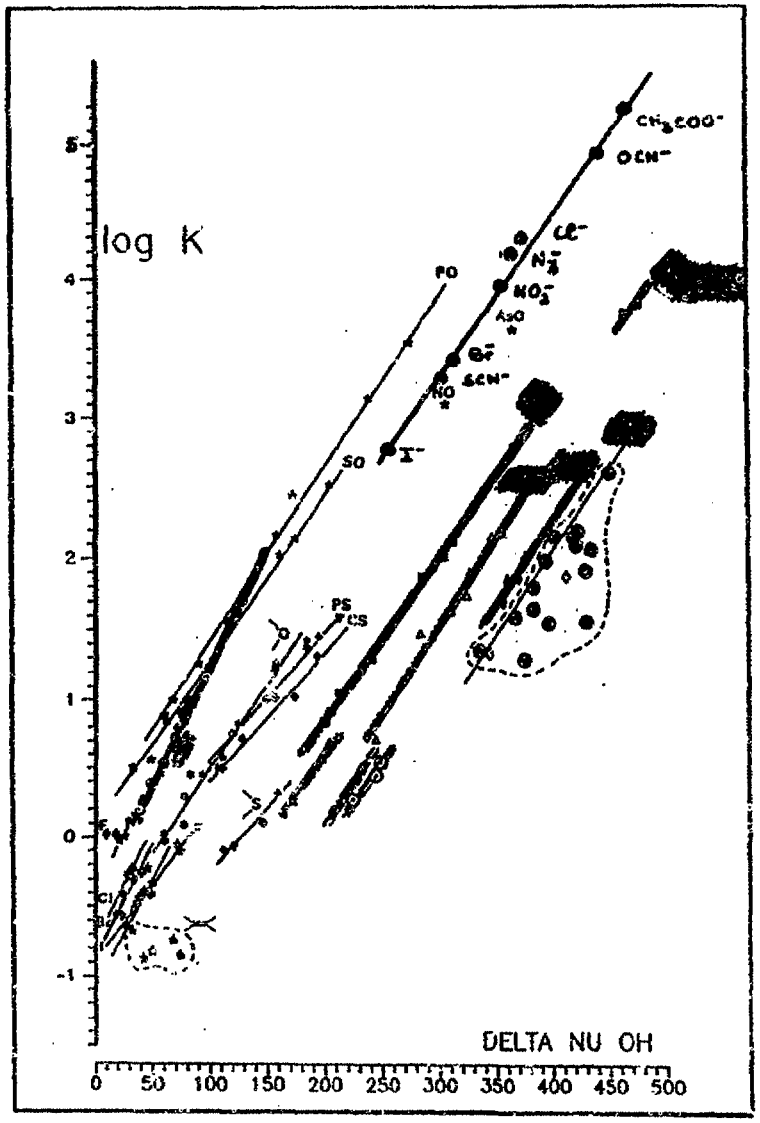
19

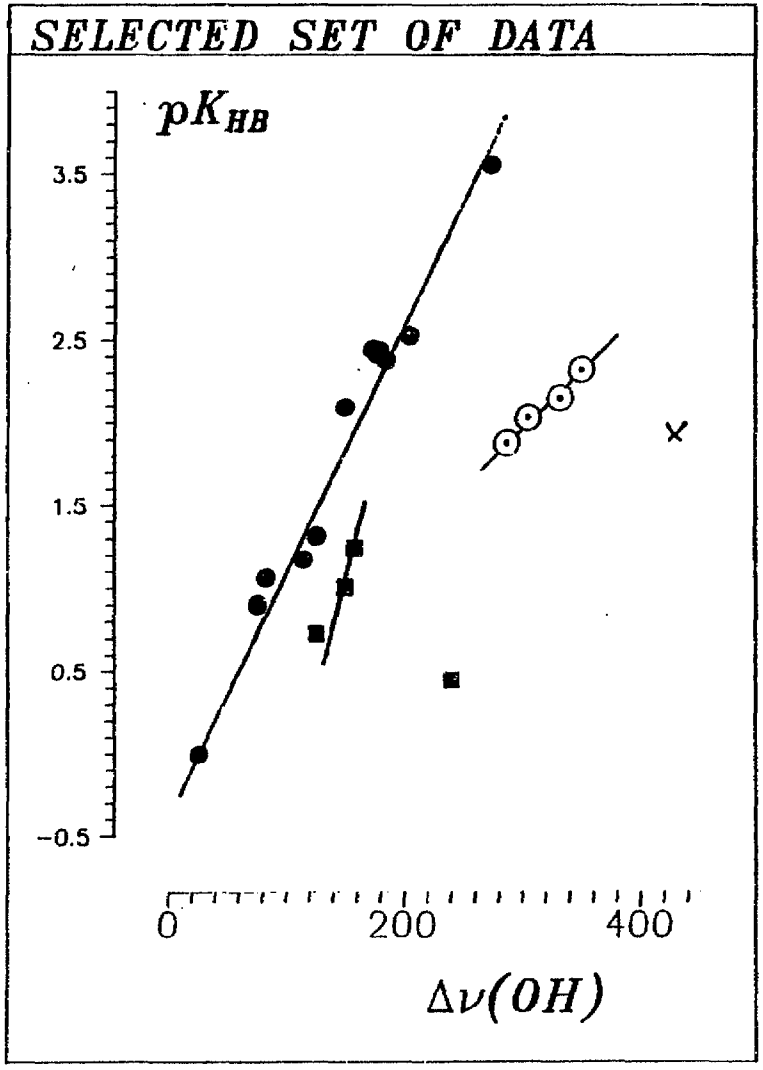

AROMATIC — ETHYLENIC — ACETYLENIC
pK_HB
0.3
0.1
−0.2
−0.4
−0.7
−1.0
−1.2
0
20
40
60
80
Δν(OH)

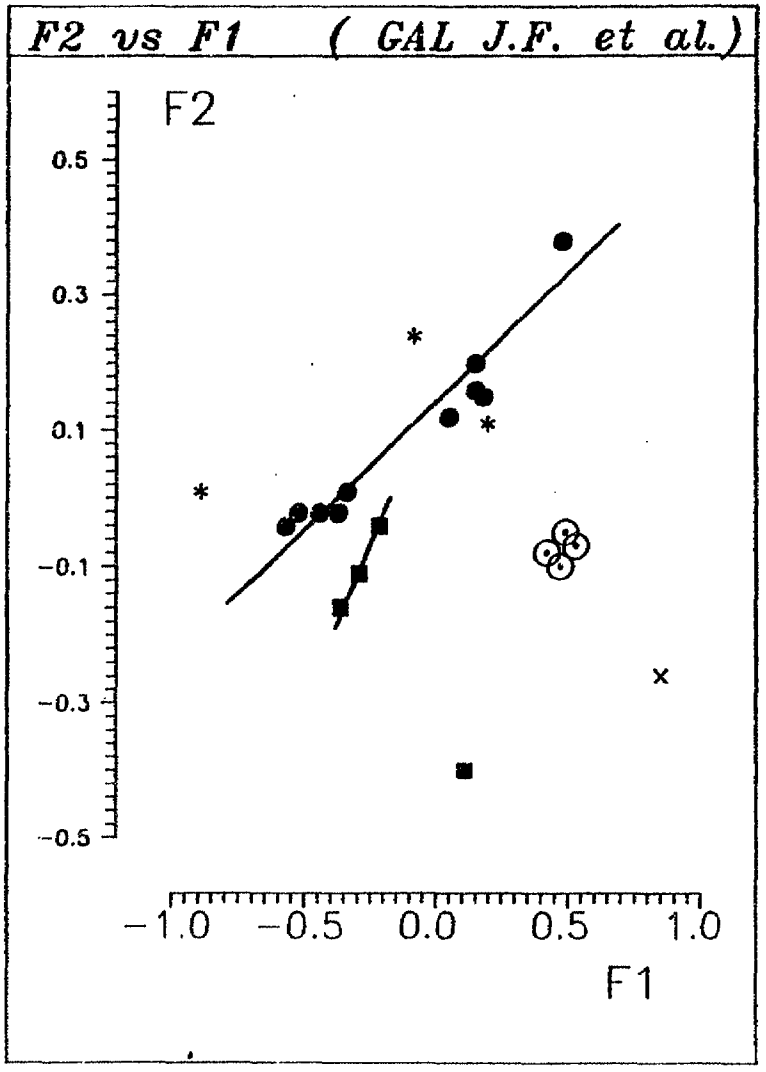
20

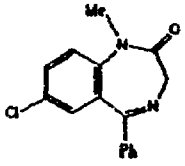

HALOGEN BASES
pK_HB
0.0
−0.5
−1.0
F
Cl
Br
I
0
10
20
30
40
50
60
Δν(OH)


ANIONS
6.0
5.0
4.0
3.0
2.0
pK_HB
CH3COO⁻
OCN⁻
Cl⁻
N3⁻
NO2⁻
Br⁻
SCN⁻
I⁻
200
300
400
500
Δν(OH)

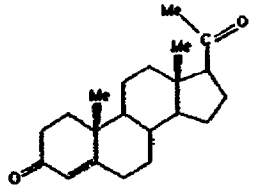

log K
DELTA NU OH
0
50
100
150
200
250
300
350
400
450
500


SELECTED SET OF DATA
pK_{HB}
3.5
2.5
1.5
0.5
−0.5
0
200
400
$\Delta\nu(OH)$

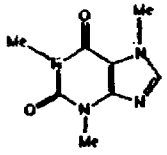

F2 vs F1 (GAL J.F. et al.)
F2
0.5
0.3
0.1
−0.1
−0.3
−0.5
−1.0
−0.5
0.0
0.5
1.0
F1



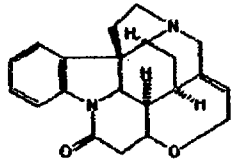
valium



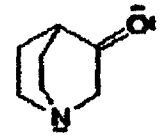
progesterone

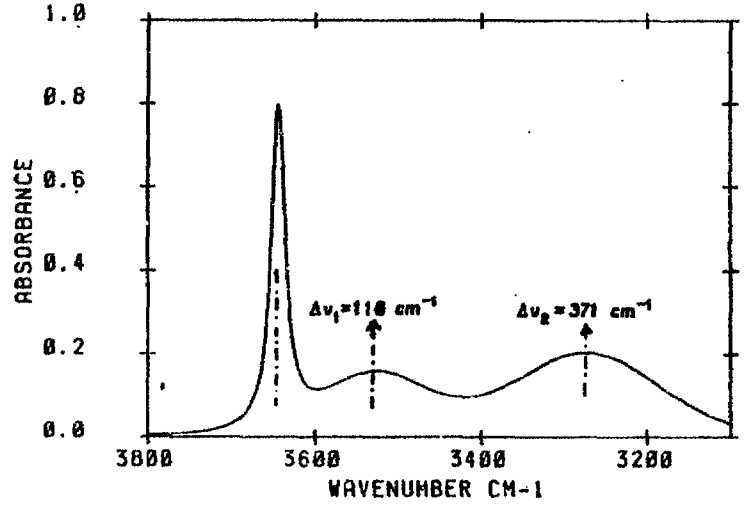


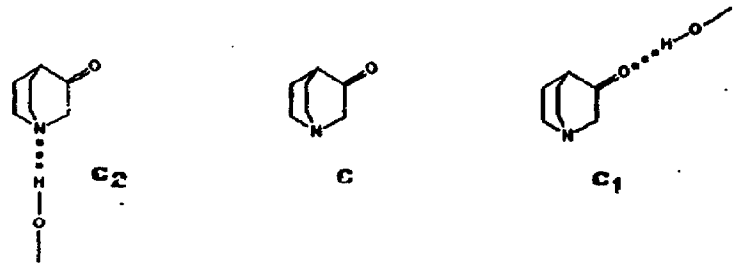
caffeine

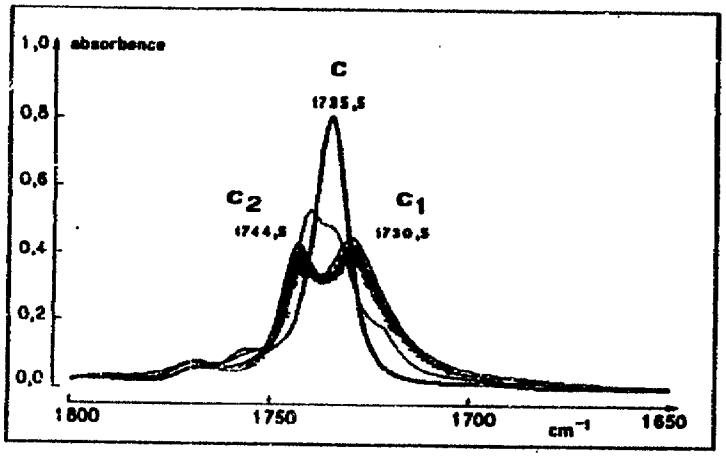


strychnine



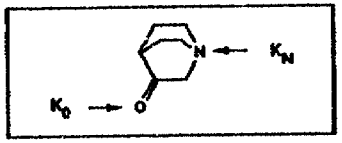


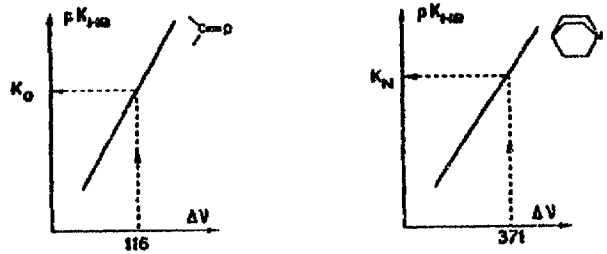




$$AH + B \rightleftharpoons C_1 \qquad K_1$$
$$AH + B \rightleftharpoons C_2 \qquad K_2$$
$$\cdots\cdots\cdots\cdots$$
$$AH + B \rightleftharpoons C_n \qquad K_n$$

$$K_{exp} = K_T = \frac{(C_1 + C_2 + \cdots + C_n)}{(AH)\,(B)} = K_1 + K_2 + \cdots + K_n$$

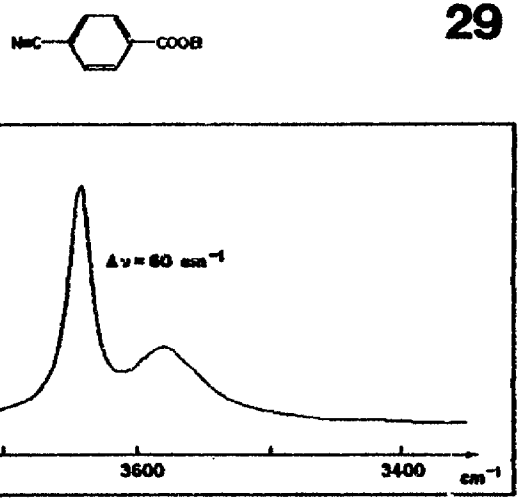


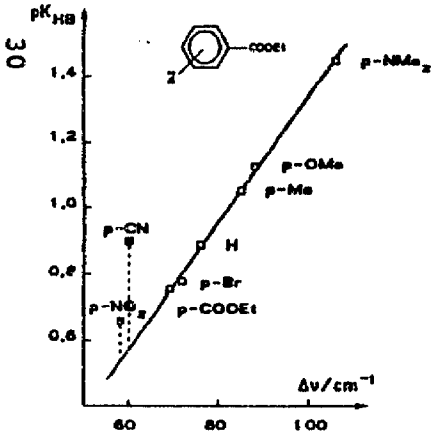


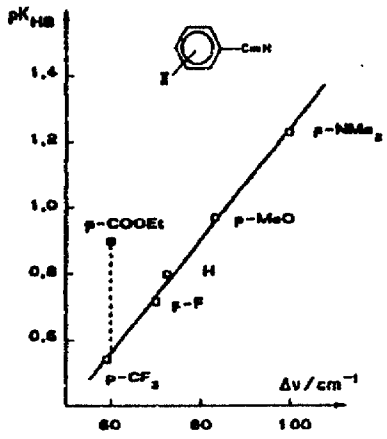
$K_O = 15.3$ $\qquad$ $K_N = 49.2$

$K_T = 64.5$

$K_{exp} = 63.1$



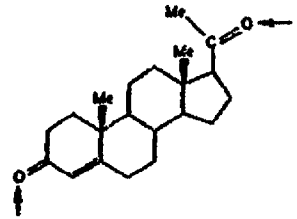


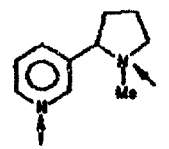


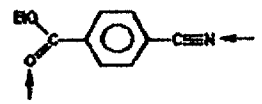
K_1 = 3,7 ℓ.mol^{-1} K_2 = 3,8 ℓ.mol^{-1}

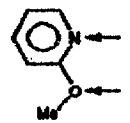
$K_T = K_1 + K_2$ = 7,5 ℓ.mol^{-1}

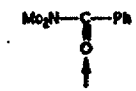
[redacted]

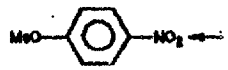


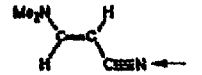


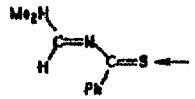




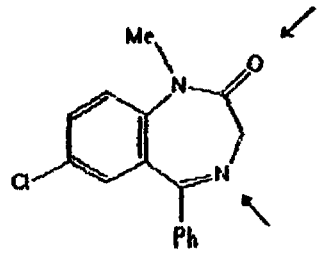




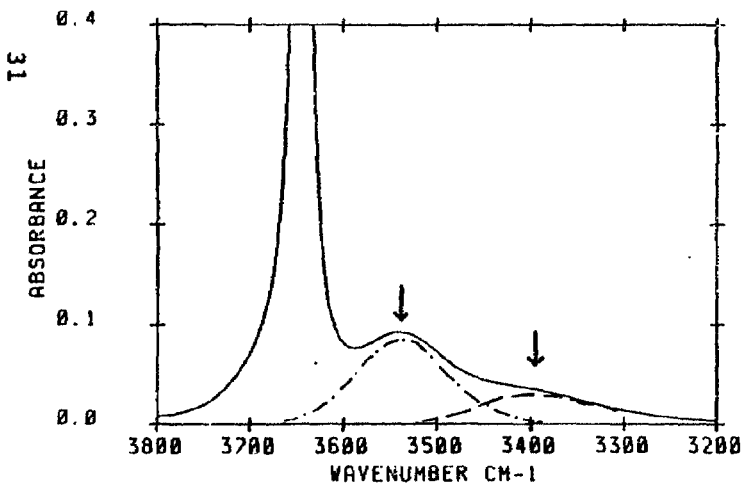




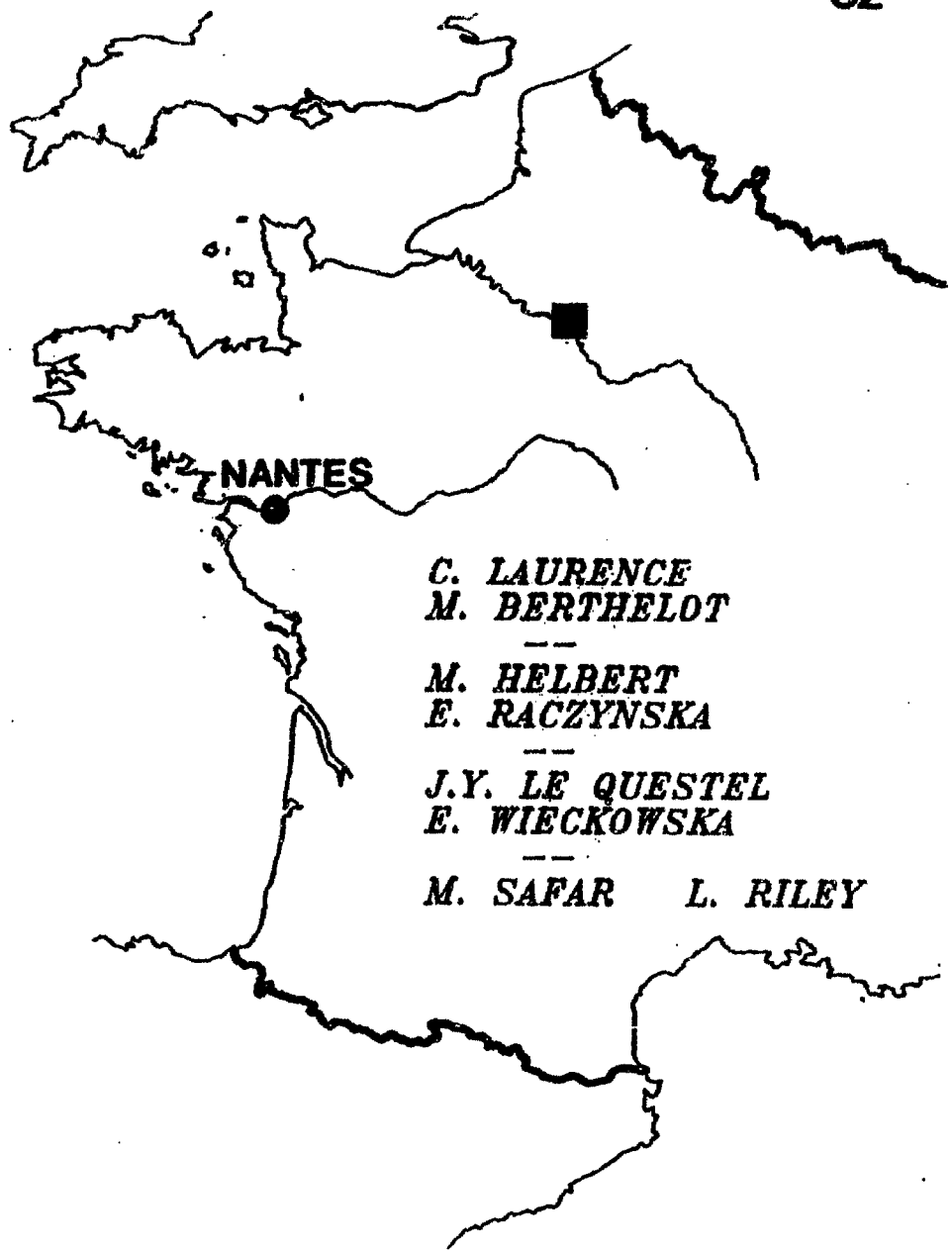
31



$K_{exp} = 43\ \ell.mol^{-1}$
$K_o = 22\ \ell.mol^{-1}$



32



BLANK

Presentation by José-Luis M. Abboud to the Aberdeen Meeting on Solvent-Solute Interactions

Two main topics were discussed:

I) The comparison between (i) Principal Component Analysis (PCA)[1] and (ii) Empirical Methods[2] for the study of medium effects on solute properties.

II) The influence of solvent "polarizability" on the properties of solutes.

I) Consider a property P of a solute in a series of solvents (S_0, S_1,.....). Within the PCA methodology, $\vec{P}_s$ can be expressed as:

$$\vec{P}_s = \vec{P}_0 + \sum_I a_I \vec{\theta}_I \qquad (1)$$

Empirically, one has:

$$\vec{P}_s = \vec{P}_0' + \sum_j C_j \vec{\varphi}_j \qquad (2)$$

$\vec{P}_0$ and $\vec{P}'_0$ are the values taken by P in reference solvents.

The fundamental mathematical difference between these two methods lies in the fact that the vectors $\vec{\theta}_i$ are mutually orthogonal while, in general, the vectors $\vec{\varphi}_j$ are not.

In order to compare both methods it was decided to study medium effects on the NMR, IR and UV-Visible spectra of a variety of a variety of solutes, using the following "rules of the game":

a) Given the variety of empirical scales of "solvent effects" currently available, the Kamlet-Taft descriptors of dipolarity-polarizability (π^*), hydrogen-bonding acidity (α) and hydrogen-bonding basicity (β) were chosen as the empirical φ_js.

b) The treatment would be "naïve", that is, in all cases, the π^* parameters would be used at face-value, no use being made of the $d\delta$ correction term, generally appropriate for aromatic and polyhalogenated solvents. We knew that this would generally lower the quality of the correlations through equation (2) but it is a fact that this is what many "non-specialist" users do.

c) The solvents used in this study were selected according to the following criteria:

- They had to span the widest possible range of π^*, α and β values. The only limitations being dictated by solubility and/or spectral interferences.
- For any given number of solvents their information contents should be maximal.
- The corrsponding sets $\{\pi^*\}$, $\{\alpha\}$ and $\{\beta\}$ were to be <u>mutually orthogonal</u> (in practice, $r^2 < 0.06$).

A set of 17 solvents, meeting these requirements was constructed by means of the NEMROD package[3]. They are listed in Table I.

We notice that spectral properties were chosen, among other reasons, because they avoid the complications inherent to cavity effects.

Also for comparison purposes, we have used Reichardt's E_T^N[4] and Swain's "acity" (A) and "basity" (B)[5].

The spectral properties studied were as follows:

1) <u>NMR Spectra</u>

<u>^{13}C Chemical Shifts of:</u> Methyl phenylsulfoxide (A), Tropone (B), Pyridine (C), 4-Methylpyridine (D), 4-Dimethylaminopyridine (E), 4-Acetylpyridine (F), 4-Cyanopyridine (G), Chlorobenzene (H), α,α,α-Trifluoromethylbenzene (I), Tropolone (J), DMSO (K).
Carbon atoms in these compounds are numbered in Chart I.

All chemical shifts in a solvent S are referred to that of the meta carbon of (I) and are defined through equation (3).

$$(\delta \int_{C_m}^{C_i})_S = (\delta C_i)_S - (\delta C_m)_S \qquad (3)$$

<u>^{31}P Chemical shifts</u>. Two cases were examined:
Triphenylphosphine oxide (L)/referred to triphenylphosphite.

Trimethyl phosphate (M)/referred to triisopropyl phophite
2) <u>IR spectra</u>. Frequencies ν of the following stretching bands were examined:
CO [in CF_3COOMe (N), CCl_3COOMe (O), CH_3COCH_3 (P) and $HCONMe_2$ (R)], CN in n-C_3H_7-CN (Q) and SO in DMSO (K).

3) <u>UV-Visible Spectra</u>. Frequencies of the longest wavelength absorption maximum for the following compounds were determined:
4-Nitroaniline (S), N-Methyl-4-nitroaniline (T), N,N-Dimethyl-4-nitroaniline (U), N,N,-Diethyl-4-nitroaniline (V), 4-Nitrophenol (W), 4-Nitroanisole (X), 3,4-Dinitrophenol (Y) and β-carotene (Z).

We are aware of the fact that, complicating factors are present in some of the IR data, particularly: (i) Fermi resonance and (ii) features related to the various extents of hydrogen-bonding association to the solvent. In the case of the UV-Visible Spectra, the existance of residual, variable vibrational structure introduces uncertainties in the determination of ν_{max}.

These facts were not taken into account, being, as they are, representative of the actual situations prevailing in many studies of "Solvent effects".

These sets of data were examined by means of equations (1) and (2). More precisely:

- Empirical treatments include correlations with (π^*, α, β),

E_T^N and (A,B).

- PCA were carried out with up to six componentes.

The results of these analyses are summarized in Tables II-IV.

Relevant conclusions derived therefrom are as follows:

1) The Kamlet-Taft formalism, even in its simplest form, generally performs better than the other empirical methods.

2) For properties displaying a significant sensitivity to the solvent, π^*, α and β account in general for at least 95% of the variance. A case-by-case study of the various correlations thusly generated shows that they "make good physical sense".

3) For the same properties, variables θ_1 to θ_6 explain the following average percentages of the variance: θ_1, 57.6%; θ_1 and θ_2, 90%; θ_1 to θ_3, 95.02%; θ_1 to θ_4, 96.22%; θ_1 to θ_5, 97.15% and θ_1 to θ_6, 98.02%. It is clear that the contributions from θ_4, θ_5 and θ_6 are rather marginal.

4) Most important, under the same conditions of orthogonality, the quality of the correlations obtained with (π^*, α, β) is quite comparable to and often better than that obtained with θ_1, θ_2 and θ_3. This is remarkable because, indirectly, the θ's account up to some extent for the "specificity" of aromatic and polyhalogenated solvents.

5) It is of interest that, while an analysis in terms of π^*, α and β can be visualized in terms of physical interactions, this is not the case for the θ's. The latter can in fact be expressed as "blends" of the Kamlet-Taft parameters. For example:

$\theta_1 = (-0.603 \pm 0.034) - (0.519 \pm 0.047)\pi^* + (0.421 \pm 0.041)\beta + (1.691 \pm 0.023)\alpha$

$r^2 = 0.991$; sd = 0.11

6) From a practical standpoint, it would seem that, for predictive purposes, if one wishes to predict the value taken by some solute property P in a given solvent, once (i) several cognate properties have been studied in a variety of media, including the solvent of interest and, (ii) P has already been determined in several of these solvents, a PCA treatment seems quite convenient.

On the other hand, if one is interested in predicting P in solvents not covered by these preliminary studies and/or if an insight into the mechanism of solvent-solute interaction is being sought, use of the Kamlet-Taft formalism seems warranted.

7) Last we notice that neither PCA nor the empirical scales should be applied outside their range of physical validity. This is the case for ν_{max} in the UV-Visible spectrum of β-carotene. Here, correlations either with (π^*, α, β) or (θ_1, θ_2 and θ_3) are poor.

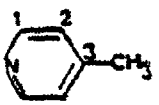
This is so because, as shown by Abe [6], medium effects on this property are essentially determined by n, the refractive index of the solvent. The properties examined in this work are much less sensitive to changes in n, and so are θ_1, θ_2 and θ_3. The physical contributions to π^*, α and β are well known and immediately explain their limited usefulness in this case.

CHART I

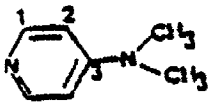
A

B

C

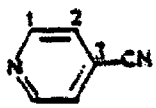


D

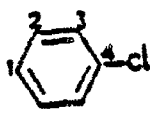


E

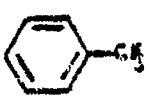
F



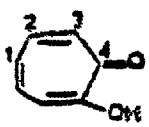
G



H



I



J

TABLE I

Solvents used in the comparison PCA/EM

Cyclohexane
Triethylamine
Tributylamine
Tetrachloromethane
Diethyl carbonate
Toluene
Acetonitrile
2, 6-Dimethylpyridine
Dichloromethane
Dimethylsulfoxide
Nitrobenzene
Tert-butanol
Butanol
Methanol
Ethylene glycol
2, 2, 2-Trifluoroethanol
1, 1, 1, 3, 3, 3-Hexafluoro-2-propanol

Regarding Tables II and III

1. These Tables are taken from the original text of a paper to appear in French in the Canadian Journal of Chemistry. Thus, commas stand for decimal points.

2. Values of A and B are not available for all the solvents studied in this work. Values in parentheses are for correlations limited to solvents for which these values are available.

3. ETE stands for the standard deviation of the correlation.

Table II. Correlations using NMR Data

Solute	P	π^*, α, β		E_T^N		A, B		Θ_1, Θ_2, Θ_3	
		r^2	ETE	r^2	ETE	r^2	ETE	r^2	ETE
A	δC_1	0,987(0,981)	0,117(0,068)	0,782	0,424	0,898	0,150	0,987	,1 16
	δC_2	0,960(0,871)	0,083(0,077)	0,611	0,242	0,745	0,102	0,954	0,088
	δC_3	0,920(0,846)	0,109(0,116)	0,277	0,308	0,752	0,140	0,929	0,103
	δC_4	0,982(0,955)	0,340(0,386)	0,930	0,530	0,962	0,334	0,993	0,213
	δC_5	0,900(0,802)	0,411(0,388)	0,914	0,351	0,877	0,290	0,953	0,279
B	δC_1	0,984(0,961)	0,196(0,187)	0,902	0,448	0,959	0,183	0,993	0,125
	δC_2	0,985(0,976)	0,284(0,269)	0,957	0,451	0,979	0,235	0,980	0,327
	δC_3	0,734(0,834)	0,355(0,312)	0,446	0,504	0,831	0,299	0,774	0,346
	δC_4	0,936(0,975)	0,219(0,201)	0,901	0,549	0,930	0,321	0,963	0,362
C	δC_1	0,950(0,912)	0,228(0,221)	0,934	0,245	0,913	0,208	0,983	0,134
	δC_2	0,978(0,956)	0,096(0,098)	0,861	0,224	0,913	0,131	0,970	0,112
	δC_3	0,992(0,984)	0,122(0,122)	0,918	0,372	0,971	0,158	0,994	0,107
D	δC_1	0,919(0,932)	0,380(0,210)	0,871	0,447	0,938	0,190	0,965	0,251
	δC_2	0,974(0,958)	0,110(0,086)	0,797	0,286	0,888	0,134	0,977	0,104
	δC_3	0,973(0,987)	0,400(0,156)	0,857	0,862	0,982	0,175	0,988	0,265
E	δC_1	0,933(0,933)	0,376(0,283)	0,873	0,478	0,936	0,261	0,961	0,288
	δC_2	0,685(0,58 1)	0,115(0,110)	0,057	0,184	0,619	0,101	0,736	0,105
	δC_3	0,938(0,87 7)	0,313(0,188)	0,627	0,718	0,830	0,211	0,962	0,247
F	δC_1	0,960(0,924)	0,128(0,128)	0,949	0,134	0,948	0,101	0,990	0,063
	δC_2	0,981(0,947)	0,093(0,108)	0,741	0,316	0,870	0,162	0,979	0,096
	δC_3	0,970(0,959)	0,123(0,112)	0,592	0,438	0,941	0,127	0,955	0,156
G	δC_1	0,909(0,809)	0,165(0,154)	0,882	0,174	0,886	0,112	0,961	0,107
	δC_2	0,976(0,943)	0,103(0,100)	0,819	0,262	0,899	0,126	0,984	0,082
	δC_3	0,431(0,575)	0,580(0,458)	0,004	0,715	0,651	0,394	0,496	0,546
H	δC_1	0,880(0,921)	0,042(0,026)	0,534	0,077	0,804	0,039	0,930	0,032
	δC_2	0,781(0,820)	0,151(0,149)	0,220	0,265	0,867	0,122	0,872	0,115
	δC_3	0,826(0,883)	0,034(0,021)	0,376	0,061	0,717	0,032	0,926	0,022
	δC_4	0,843(0,855)	0,279(0,302)	0,200	0,587	0,888	0,252	0,907	0,215
I	$\delta(C_o\text{-}C_o)$	0,943(0,956)	0,086(0,081)	0,206	0,292	0,902	0,115	0,933	0,091
	$\delta(C_p\text{-}C_o)$	0,943(0,962)	0,119(0,109)	0,212	0,410	0,927	0,144	0,949	0,111
	$\delta(C_p\text{-}C_o)$	0,749(0,826)	0,081(0,080)	0,166	0,138	0,802	0,081	0,799	0,073
J	δC_1	0,965(0,904)	0,292(0,312)	0,831	0,591	0,852	0,366	0,972	0,261
	δC_2	0,960(0,872)	0,240(0,269)	0,789	0,505	0,820	0,301	0,965	0,223
	δC_3	0,967(0,936)	0,203(0,205)	0,839	0,412	0,830	0,315	0,971	0,189
	δC_4	0,197(0,303)	0,895(0,957)	0,011	0,913	0,072	1,041	0,195	0,895
K	δC	0,919(0,858)	0,324(0,329)	0,955	0,225	0,922	0,232	0,942	0,275
L	δP	0,944(0,904)	0,370(0,323)	0,785	0,652	0,824	0,416	0,940	0,371
M	δP	0,987(0,977)	0,582(0,663)	0,926	1,300	0,977	0,613	0,989	0,532

Table III. Correlations using IR and UV-Visible Data

Solute P	π^*, α, β		E_T^N		A, B		θ_1, θ_2, θ_3	
	r^2	ETE	r^2	ETE	r^2	ETE	r^2	ETE
N $\sqrt{_{co}}$	0,946(0,994)	1,222(0,243)	0,604	3,071	0,661	1,731	0,986	0,621
O $\sqrt{_{co}}$	0,930(0,862)	1,575(1,307)	0,609	3,279	0,630	2,027	0,960	1,126
P $\sqrt{_{co}}$	0,959(0,960)	1,242(1,031)	0,925	1,546	0,971	0,835	0,971	1,038
R $\sqrt{_{co}}$	0,972(0,950)	2,362(2,503	0,921	3,621	0,966	1,947	0,976	2,172
K $\sqrt{_{so}}$	0,935(0,954)	7,682(6,355)	0,866	10,166	0,915	8,192	0,965	5,639
Q $\sqrt{_{co}}$	0,994(0,991)	0,449(0,381)	0,538	4,054	0,923	1,078	0,992	0,592
S $\sqrt{_{aaa}}$	0,984(0,990)	0,218(0,202)	0,449	1,186	0,732	0,989	0,974	0,275
T $\sqrt{_{aaa}}$	0,963(0,984)	0,297(0,192)	0,752	0,713	0,865	0,530	0,971	0,265
U $\sqrt{_{aaa}}$	0,954(0,976)	0,328(0,202)	0,767	0,684	0,988	0,133	0,984	0,191
V $\sqrt{_{aaa}}$	0,959(0,990)	0,303(0,124)	0,779	0,655	0,986	0,139	0,986	0,178
W $\sqrt{_{aaa}}$	0,895(0,969)	0,369(0,241)	0,271	0,899	0,559	0,856	0,812	0,493
X $\sqrt{_{aaa}}$	0,950(0,969)	0,214(0,151)	0,673	0,505	0,973	0,133	0,976	0,148
Y $\sqrt{_{aaa}}$	0,963(0,969)	0,380(0,396)	0,407	1,352	0,757	1,031	0,927	0,530
Z $\sqrt{_{aaa}}$	0,745(0,691)	0,192(0,186)	0,198	0,315	0,641	0,189	0,693	0,211
E_T	0,953(0,907)	2,500(3,233)			0,974	1,613	0,968	0,214

II) "Some comments on "polarizability effects"

The starting point of this discussion is an important study by Laurance, Nicolet and Reichardt[1], dealing with the comparison of medium effects on E_T and π^*. Examined therein were a highly lipophilic betaine (A) and 4-nitroanisole (b) in a variety of saturated cyclic and acyclic hydrocarbons, perfluorinated solvents and in the gas-phase [the range of media was narrower for (A)].

An extremely important finding was that, while π^* shows a strong dependence and varies linearly with $f(n) = (n^2-1)/(2n^2+1)$ over the entire range of values for n, E_T is essentially independent of f(n).

On the other hand, it is known[2] that over the range c-C_6H_{12} - DMSO, E_T and ($\pi^* - d\delta$) for non hydrogen-bond donor solvents are linearly related. It seems, therefore, that if some property is well correlated by π^* [or ($\pi^* - d\delta$)] over the range c-C_6H_{12} - DMSO ($0 \leq \pi^* \leq 1.0$) one cannot safely conclude that such a correlation will still hold for negative π^* values, (i.e., in the range gas-phase -c-C_6H_{12}).

Use of the simplified form of McRae's equation [equation (1)] shows why it is so:

$$\pi^* = \text{constant} + M[(n^2-1)/(2n^2+1)] + N\{[(\epsilon-1)/(2\epsilon+1)] - [(n^2-1)/(2n^2+1)]\}$$

In general, in the range c-C_6H_{12} - DMSO, π^* values largely reflect the importance of the last term in eq. (1). Over the range gas-phase -c-C_6H_{12}, this term is generally small or nil and changes in π^* essentially reflect changes in the refractive index of the medium. Now, for medium effects or any property P to be lenearly related to π^* over the entire range gas-phase - DMSO, the ratio N/M has to be the same for P as for π^*.

"A priori", there is no reason for this behavior to be general.

I presented two examples of properties showing a very different response to changes in the refractive index of the medium. These properties are known to satisfactorily correlate with (π^*-dδ) for $\pi^*>0$.

Differential free-energies of activation for the quaternization of Et_3N with MeI are seen to be extremely sensitive to changes in n, as well as to more specific "aromatic" effects[3], Fig. 1. It is likely that some contributions from cavity effects are also embodied in this linear relationship.

Standard free-energy changes ΔG° for the enolization of acetylacetone, equation (2):

$$CH_3COCH_2COCH_3 \rightleftharpoons CH_3COHCHCOCH_3 \qquad (2)$$

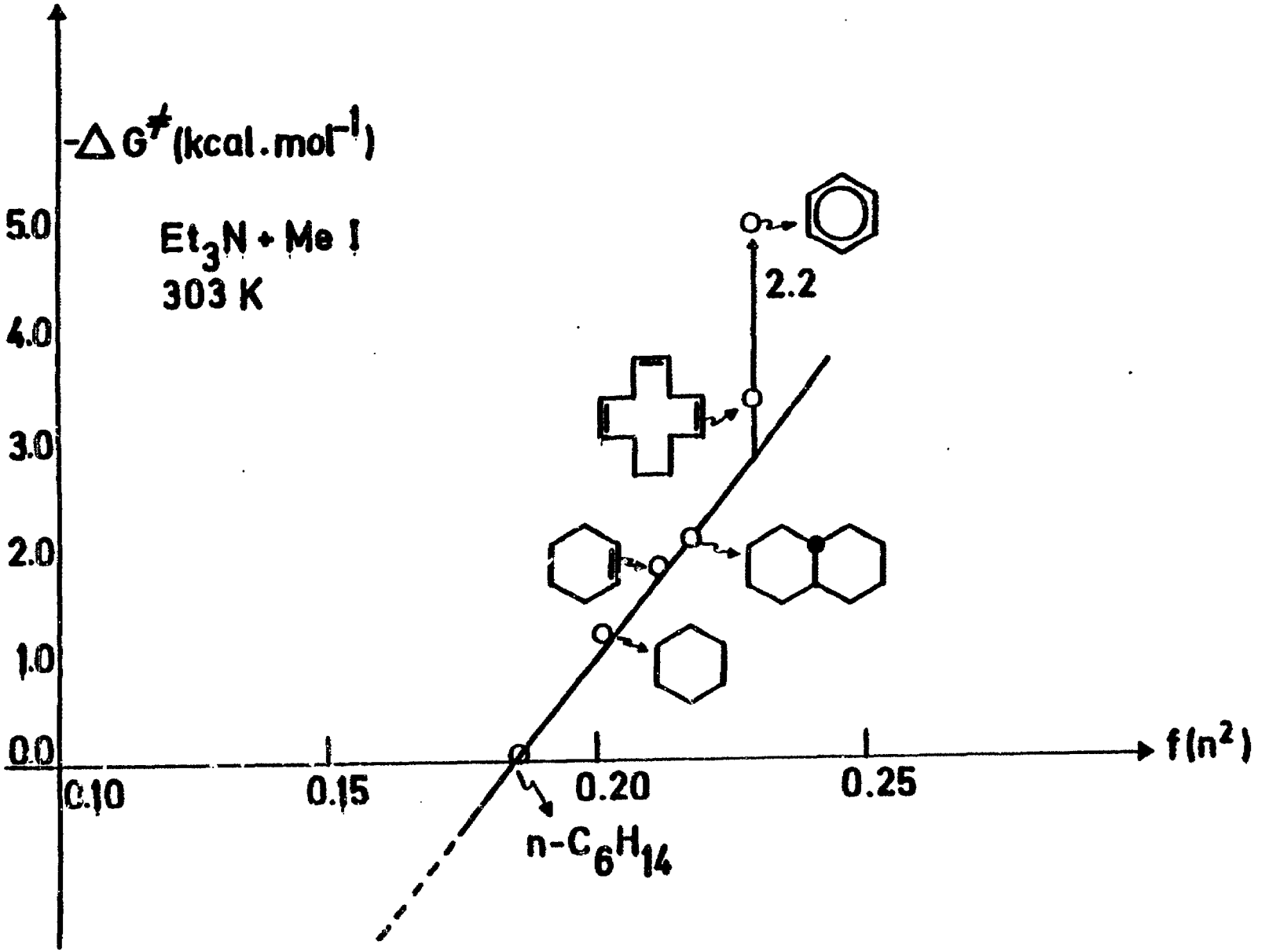
have been determined in the gas-phase[4] and in a variety of solvents[5].

We were stricken by the fact that the value for ΔG° in the gas-phase was nearly the same (within the limits of experimental error) than that in c-C_6H_{12} solution.

We are presently exploring this reaction in solvents of low or nil dipolarity having widely different refractive indexes. I presented preliminary results in perfluorodecalin, n-heptane, cis.decalin, quadricyclane and trans, trans, cis- 1, 5, 9-cyclododecatriene. It was clear that contributions other than dipolar or multipolar effects (case of benzene) play a very minor role (if any) on the position of equilibrium (2).

It thus seems that good correlations with π^* or(π^*-dδ) for $\pi^*\geq 0$ do not necessarily imply any such correlation for $\pi^*<0$.

Figure 1


-ΔG* (kcal.mol⁻¹)
Et3N + Me I
303 K
5.0
4.0
3.0
2.0
1.0
0.0
2.2
0.10
0.15
0.20
0.25
f(n²)
n-C6H14

REFERENCES

PART I

(1) (a) M. Sjöström and S. Wold, J. Chem. Soc. Perkin Trans. 2, 1981, 104 and Acta Chem. Scand., 1981, B35, 537. (b) S. Wold and M. Sjöström, Acta Chem. Scand., 1986, B40, 270.

(2) (a) M. J. Kamlet and R. W. Taft, Acta Chem. Scand., 1985, B39, 611. (b) M. J. Kamlet, R. M. Doherty, G. R. Famini and R. W. Taft, Acta Chem. Scand., 1987, B41, 589.

(3) D. Mathieu and R. Phan Tan Luu, University of Marseille.

(4) "Reduced" values of Reichardt's E_T as given in the updated data bases: (a) C. Reichardt, E. Harbusch-Görnert, Liebigs Ann. Chem., 1983, 721. (b) C. Laurence, P. Nicolet and C. Reichardt, Bull. Soc. Chim. Fr. 1987, 125. (c) C. Reichardt, "Solvents and Solvent Effects in Organic Chemistry", Second Edition, VCH, Weinheim, Germany, 1988.

(5) C. G. Swain, M. S. Swain, A. L. Powell and S. Alunni, J. Am. Chem. Soc., 1983, 105, 502.

(6) T. Abe, Bull. Chem. Son. Jpn., 1990, 63, 2328.

PART II

(1) See Part I, ref. 4b.

(2) M. J. Kamlet, J.-L. M. Abboud and R. W. Taft, Prog. Phys. Org. Chem., 1983, 13, 485.

(3) J.-L. M. Abboud, R. Notario, M. Solá and J. Bertrán, Prog. Phys. Org. Chem., in press.

(4) M. M. Folkendt, B. E. Weiss-Lopez, J. Paul Chauvel, Jr., and N. S. True, J. Phys. Chem., 1985, 89, 3347.

(5) J. Emsley and N. J. Freeman, J. Mol. Struct., 1987, 161, 193.

BLANK

FORMATION CONSTANTS IN C-H HYDROGEN BONDING

John P. Lorand
Malcolm H. Filson Laboratories, Department of Chemistry
Central Michigan University
Mt. Pleasant, MI 48859

Formation constants for Hydrogen Bond (H-bond) complexes of a wide variety of C-H donors with HMPA in CCl_4 at 35° have been measured by a pmr method.[1,2] The donors fall into classes in which the C atom hybridization is sp^3, sp^2 (ethylenes as well as substituted benzenes), and sp (terminal alkynes). In several cases where constants were too small to measure in CCl_4, cyclohexane as solvent gave measurable K values.

Proton chemical shifts for series of solutions were analyzed via the Higuchi equation,[3] which is presented in viewgraph 1, together with a description of its use. This equation is derived with no approximations for 1:1 complexes. No analogous equation has been derived for 2:1 complexes, $(HA)_2 \cdot B$.

As the examples in viewgraph 2 show, plots are invariably linear when the donor contains only one "acidic" proton. When two or more equivalent "acidic" protons are present, as shown in viewgraphs 3 and 4, the Higuchi plot may be concave downward, particularly at higher base concentrations. This curvature generally disappears at lower [HMPA], the range depending on the strength of the proton donor. Examples are fumaronitrile, malononitrile, and the 1,2,3,5- and 1,2,4,5-tetrahalobenzenes, and 1,3,5-trihalobenzenes. K values are here reported only for low concentrations, and have been statistically corrected by dividing by the number of equivalent H's, giving "K/H", when two or more are present.

Data for $C(sp^3)$-H donors appear in viewgraphs 5 and 6. Several observations emerge:

(1) The heavy halogens are about equally effective at facilitating H-bonding when present on the α-Carbon.

(2) F is less effective than the other halogens when on the α-Carbon, but more effective when on the β-Carbon.

(3) The cyano group is c. 8 times as effective as a heavy halogen, but the nitro group only c. 5 times as effective, in the dihaloacetonitriles. When, however, two of each group are present, cyano and nitro are equally effective. The largest K value measured thus far is 205 M^{-1}, for bromomalononitrile[1] (not shown in viewgraphs).

(4) Carbonyl groups are less effective than halogens, and methoxy is entirely ineffective.

(5) Viewgraph 7 shows a tendency for K to increase with the sum of the inductive/field substituent constants, but with considerable scatter. The scatter can be explained only in part by steric effects. The order of K's is somewhat scrambled as compared to that for substituent constants.

(6) K is measurable for diphenylacetonitrile, though it contains only one significantly polar functional group.

(7) Shifts of the C-H stretching band in the IR, $-\Delta\nu$ (cm^{-1}), measured in the presence of DMSO-d_6,[14] are almost entirely uncorrelated with K values. Viewgraph 8 shows this plainly. Difluoroacetonitrile and phenyldinitromethane show no shift at all; however, their C-H stretching bands both increase in intensity.

(8) Substituent effects observed for Br and CN can be used to predict K's for methyl bromide and for acetonitrile. K/H for the latter is significant and should be measurable. Both values can in turn be extrapolated to essentially the same value, 0.007, for K/H for methane. This might suggest that methane undergoes H-bonding with HMPA. However, the K/H value is an order of magnitude below the minimum found by Abraham[5] for a wide variety of donors and acceptors. Therefore, any interaction that may occur between methane and HMPA is probably not an H-bond.

K's for several alkenes with HMPA appear in viewgraph 9. Again, the cyano group is much more effective than halogen; cyano and nitro are about equally effective; β-cyano as at least as effective as α-cyano; and IR frequency shifts are nearly independent of K.

K's for several terminal alkynes with HMPA are shown in viewgraph 10. The three smallest values were measured in cyclohexane, while the three largest, measured in CCl_4, are calculated using the fact that in general $k(\text{cyclohexane})/k(CCl_4) = 4$. The K's cover a wide range, cyano again giving the largest value. The order of K's follows the order of substituent constants except for ethoxy. Most likely the inductive/field electron withdrawing effect of ethoxy is counteracted by n-electron donation to the triple bond. All the IR frequency shifts are large, and they show more correlation with K's than do those for alkenes.

K/H values for series of polyhalobenzenes appear in viewgraph 11. Many of these were measured in cyclohexane because of their small magnitudes in CCl_4. However, most values for polyfluorobenzenes had to be measured in CCl_4 because the fluorines split their signals sufficiently to render them invisible by the FT technique in the presence of the very strong cyclohexane signal.

For polyfluorobenzenes, the more F atoms, the larger is K/H. Comparison of tetrafluorobenzenes with pentafluorobenzene shows that F has the largest effect at the meta position, the smallest at the ortho position. Polychlorobenzenes behave similarly; however, K/H is larger for 1,2,3,4-tetrachlorobenzene than for pentachlorobenzene, rather than smaller, and K/H for H-5 of 1,2,3-trichlorobenzene is larger still. K/H values tend to fall in the order F > Cl > Br, but values for 1,3,5-trihalobenzenes are indistinguishable, and so are those for 1,4-dihalobenzenes. A value of K for 2,4,6-tribromo-m-xylene could not be measured, at least in CCl_4, because HMPA did not increase the chemical shift.

These observations imply that the sizes of halogen atoms are important in determining K/H. The principal effect is "buttressing", the deformation of C-C-X and even C-C-C bond angles by two or more larger halogens in consecutive

positions on the ring. Thus, it is known from crystallographic data that not only hexachloro-[6] but also pentachlorobenzene[7] are non-planar, both the ring carbons and the chlorine atoms lying either above or below an "average" plane. Hexafluorobenzene[8] is, however, accurately planar. Replacing Cl-1 in pentachlorobenzene by H to form 1,2,3,4-tetrachlorobenzene both relieves part of this deformation and makes each H more accessible to an HMPA molecule. The 1,3,5-trihalobenzenes have no buttressing, so their K/H values reflect the intrinsic substituent effects of the halogens. We may surmise that these are indistinguishable--as is consistent with their similar inductive/field constants. Buttressing in 2,4,6-tribromo-m-xylene must be severe, and K may well be smaller than that of pentabromobenzene.

The effects of replacing halogen atoms in particular locations with H atoms may be calculated and used to predict a value of K/H for benzene (cf. viewgraph 12). The polyfluorobenzenes give a value of c. 0.05 M^{-1}, while the polychlorobenzenes give c. 0.25 M^{-1}, both in cyclohexane. An attempt to measure K/H for benzene in cyclohexane gave a fairly good Higuchi plot and a K/H value of c. 0.05 M^{-1}, in good agreement with prediction for the polyfluoro series. Presumably the substituent effects in the polychloro series, determined for several compounds experiencing buttressing, are different from those which apply in the absence of buttressing. It is noteworthy that our rough value for benzene falls near the Abraham borderline, and it is desirable to seek independent evidence as to whether benzene indeed does or does not represent a second unsubstituted hydrocarbon, after terminal alkynes, which engages in H-bonding.

In conclusion, this work has shown that K's for H-bonding by a variety of compounds with HMPA in either CCl_4 or cyclohexane can be measured by pmr; that several polar substituents are slightly to very effective at facilitating H-bond formation; that K's overall correlate very poorly with $-\Delta\nu$ from IR spectra; that K's correlate fairly well with inductive/field substituent constants, with a few exceptions; and that benzene--and less likely, methane--may engage in H-bonding with strong acceptors. It had already been shown by IR, mainly by Allerhand and Schleyer,[4] that H-bonding by C-H bonds is more common than had been supposed. Given the proper tools, it should be possible to show that the C-H bonds of acetonitrile, nitromethane, and many other compounds can also form H-bonds.

References

1. F. M. Slasinski, J. M. Tustin, F. J. Sweeney, A. M. Armstrong, Q.A. Ahmed, & J. P. Lorand, **J. Org. Chem.**, **41**, 2693 (1976).
2. J. P. Lorand, J. P. Nelson, R. D. Gilman, K. L. Staley, J. R. Chambers, H. D. Kirk, K. J. Moeggenborg, & D. L Farlow, **J. Phys. Org. Chem.**, **3**, 659 (1990).
3. M. Nakano, N. I. Nakano, & T. Higuchi, **J. Phys. Chem.**, **71**, 3954 (1967).
4. A. Allerhand & P. v. R. Schleyer, **J. Amer. Chem. Soc.**, **85**, 1715 (1963).
5. M. H. Abraham, P. P. Duce, P. L. Grellier, D. V. Prior, J. J. Morris, & P. J. Taylor, **Tetrahedron Letters**, **29**, 1587 (1988).
6. G. M. Brown & O. A. W. Strydom, **Acta Crystallogr., Sect. B 30**, 801 (1974).
7. P. Marsh & D. E. Williams, **Acta Crystallogr., Sect. B 37**, 705 (1981).
8. N. Boden, P. P. Davis, C. H. Stam, & G. A. Wesselink, **Mol. Phys.**, **25**, 81 (1973).

Acknowledgements

The author thanks Prof. Robert W. Taft and Dr. George Famini for their kind invitation to attend the 1991 Workshop on Solvent-Solute Interactions. Belated acknowledgements are also due to the late Prof. Harold Kwart for pointing out the probable operation of C-H H-bonding in solutions of substituted acetonitriles in dipolar aprotic solvents, and to Profs. Arnett, Schleyer, and Taft for pointing out HMPA as a powerful proton acceptor. Support from the U. S. Army Research Office-Durham early on and the Faculty Research and Creative Endeavors Committee of Central Michigan University is gratefully acknowledged, as is the labor of the many student coworkers listed in references 1 and 2.

MEASUREMENT of K by PMR

THE HIGUCHI EQUATION

Derived for: A(cid) + B(ase) = C(omplex),

assuming 1:1 complexes **only**:

$$\frac{C_b}{\delta_{obs} - \delta_a} = \frac{C_a + C_b - C_c}{\delta_c - \delta_a} + \frac{1}{K(\delta_c - \delta_a)}$$

C's = total concentrations (M)
δ's = chemical shifts

Procedure:

1) Plot Left side vs. $C_a + C_b$: calculate slope

2) Calculate $C_c = C_a \frac{\delta_{obs} - \delta_a}{\delta_c - \delta_a}$; replot vs. $C_a + C_b - C_c$ & recalc. slope

3) Repeat (2) until slope converges

4) Reject deviant points & repeat

5) K = slope/intercept

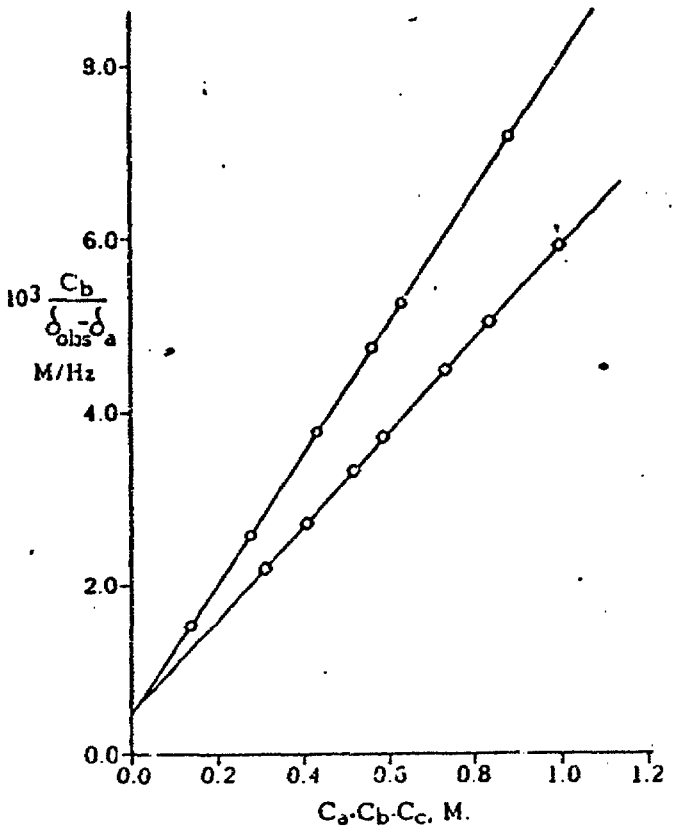


Figure 1. Higuchi plots for "strong" donors with HMPA in CCl_4, 35.0 °C: ○, Br_2CH-CN, c 0.09 M; ◇, H-C≡C-CN, c 0.08 M. The coincidence of the intercepts is not significant.

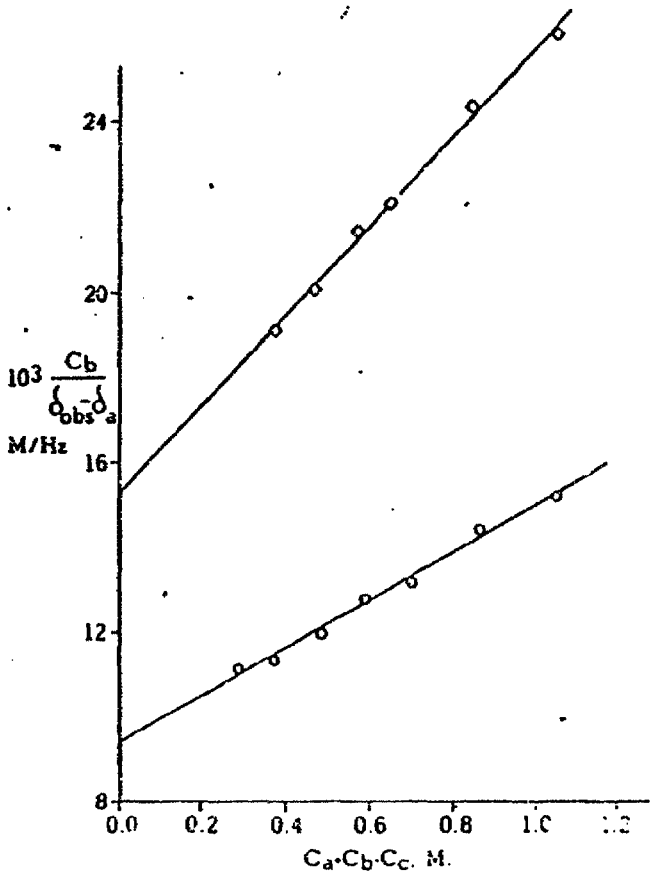


Figure 2. Higuchi plots for "weak" donors with HMPA in CCl_4, 35.0 °C: ○, pentachlorocyclopropane, c 0.10 M; ◇, $PhCHBr_2$, 0.11 M.

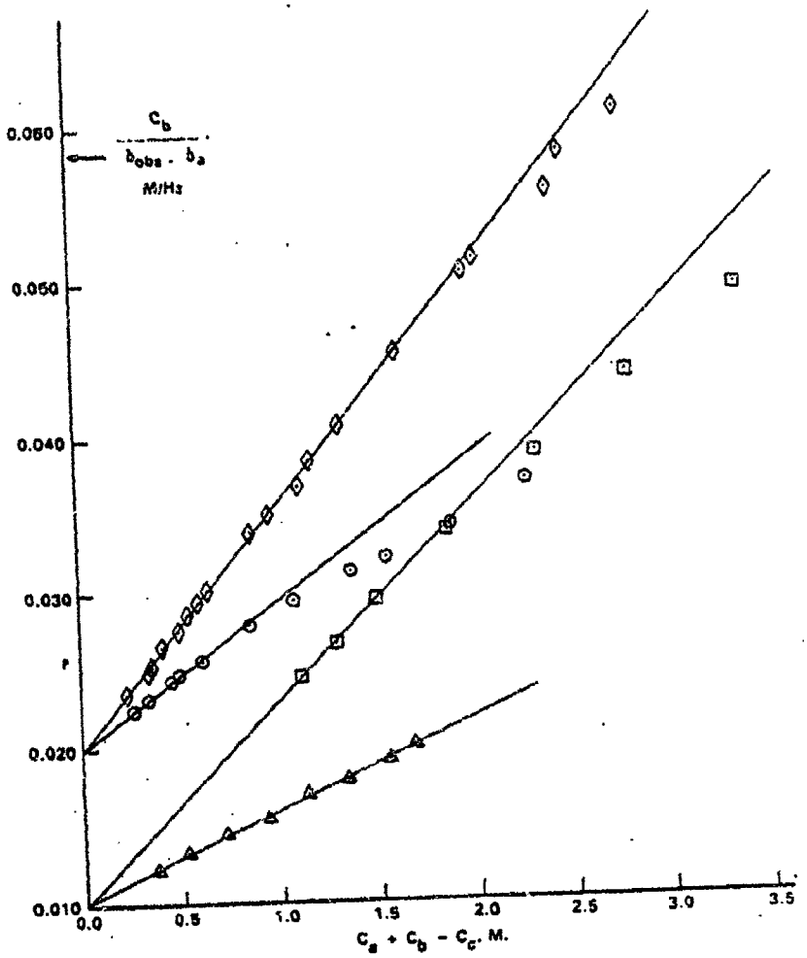


Figure 2. Higuchi plots for association of polyfluorobenzenes with HMPA. △, C_6HF_5 in CCl_4, $K = 0{\cdot}56$ l mol^{-1}, $\delta_c - \delta_a = 2{\cdot}20$ ppm; □, 1,2,3,5-$C_6H_2F_4$ in cyclohexane, $K = 1.11$ l mol^{-1}, $\delta_c - \delta_a = 1{\cdot}37$ ppm (from low concentration points); ○, 1,2,3,5-~~1,2,3,4-~~$C_6H_2F_4$ in CCl_4, $K = 0.64$ l mol^{-1}, $\delta_c - \delta_a = 1{\cdot}01$ ppm (from low concentration points); ◇, 1,2,3,4-$C_6H_2F_4$ in CCl_4, $K = 0.71$ l mol^{-1}, $\delta_c - \delta_a = 0.85$ ppm

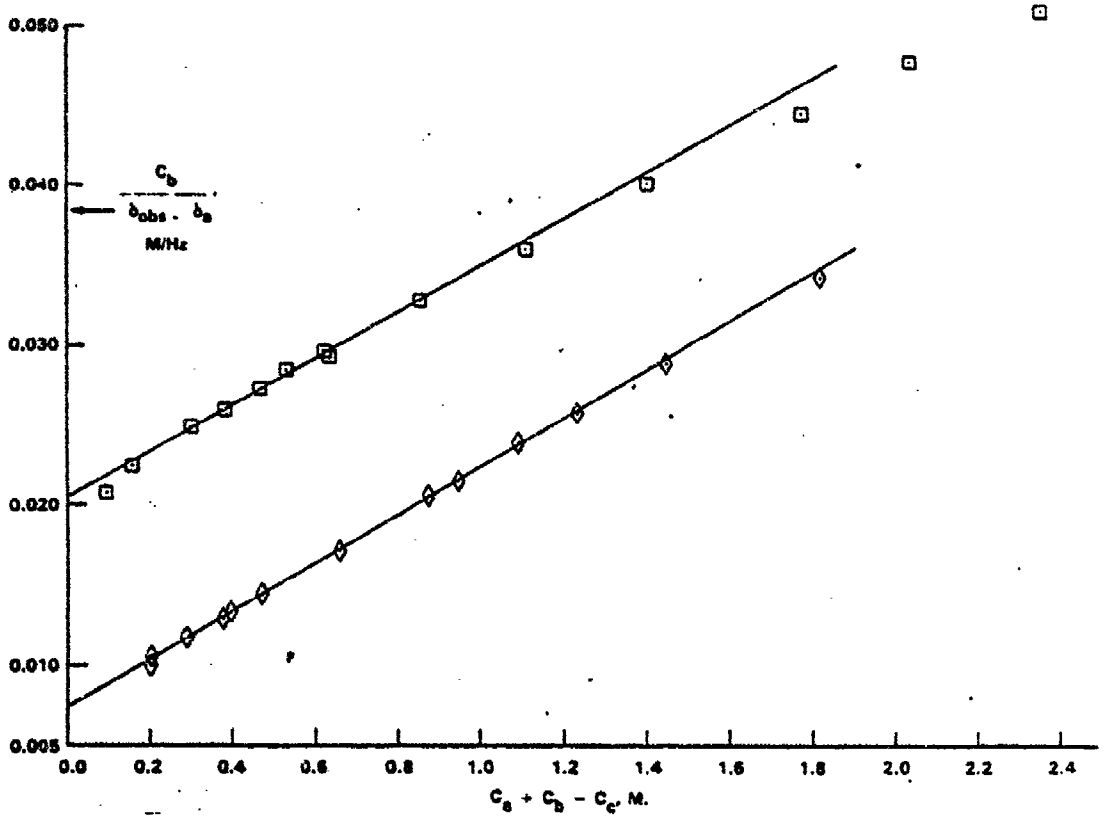


Figure 1. Higuchi plots for association of tetrachlorobenzenes with HMPA in cyclohexane. ○, 1,2,3,4-Isomer, $K = 2{\cdot}06$ l mol^{-1}, $\delta_c - \delta_a = 0{\cdot}84$ ppm; □, 1,2,3,5-isomer, $K = 0{\cdot}76$ l mol^{-1}, $\delta_c - \delta_a = 0{\cdot}83$ ppm (from low concentration points)

C(sp^3)-H DONORS with HMPA in CCl_4

DONOR	K (M^{-1})	$-\Delta\nu$ (cm^{-1})
$HCCl_3$	2.3	29
$HCBr_3$	2.0	50
HCI_3	2.8	
HCF_2CN	10.7	0
$HCCl_2CN$	20.6	66
$HCBr_2CN$	16.5	80
$HCBr_2NO_2$	10.3	80
$HCCl_2CONMe_2$	1.44	
$HCCl_2COMe$	0.95	
$HCCl_2CO_2Me$	0.68	26
$HCCl_2OMe$	c 0.3	
$HCBr_2CBr_3$	1.1	58
$HCBrClCF_3$	2.5	

K: $CN > NO_2 > Cl, Br, I, CF_3 > F, COX, CBr_3 > OMe$

σ: $NO_2 > CN > F > Cl, CF_3, Br > I > COX, OMe$

MOSTLY DISUBSTITUTED C(sp^3)-H DONORS w. HMPA in CCl_4

DONOR	K, M^{-1}	K/H, M^{-1}
CH_2Br_2	0.68	0.34
$PhCHBr_2$	0.68	0.68
p-O_2N-C_6H_4-$CHBr_2$	5.5	5.5
$ClCH_2CN$	4.6	2.3
$BrCH_2CN$	7.1	3.5
ICH_2CN	6.5	3.2
Ph_2CHCN	1.0	1.0
$CH_2(SCN)_2$	6.5	3.3
$CH_2(CN)_2$	60.	30.
With Acetone:		
$CH_2(CN)_2$	3.39	1.7
Ph-$CH(CN)_2$	1.70	1.7
Ph-$CH(NO_2)_2$	1.69	1.7
With HMPA, Estimated:		
CH_3Br		0.05
CH_3CN		0.43
CH_4		0.007

K: CN > NO_2 > SCN ~ p-O_2N-C_6H_4 > F, Cl, Br > $CONMe_2$> COMe > CO_2Me, Ph > OMe

σ: NO_2>CN>F>Cl, CF_3, Br>I>CO_2Me, COMe, OMe > $CONH_2$ > Ph

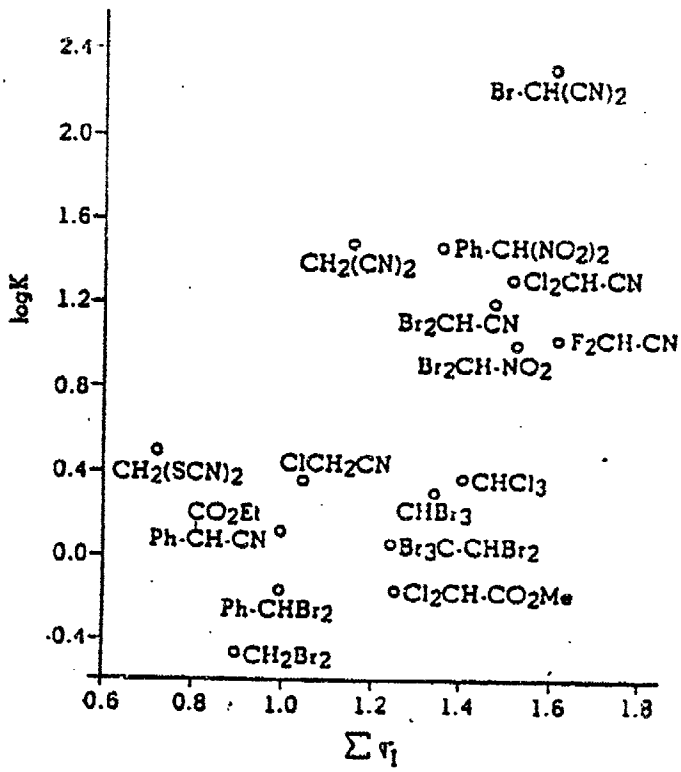


Figure 4. Relation between K's and sum of Taft σ_I values (cf. ref 18).

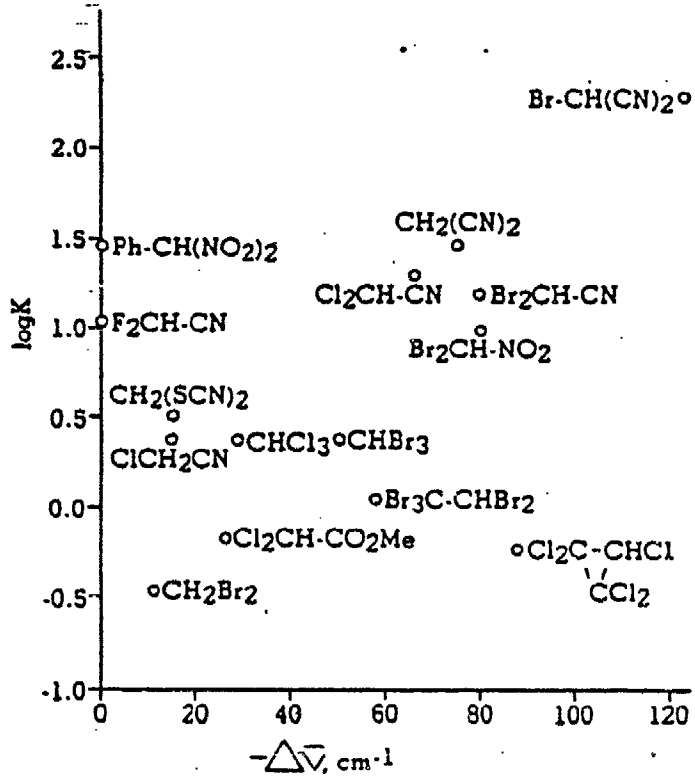


Figure 3. Relation between donor–HMPA formation constants (K's) and infrared C–H stretching wavenumber shifts.

$C(sp^2)$-H DONORS w. HMPA in CCl_4
ALKENES

DONOR	K/H (M^{-1})	$-\Delta\nu$ (cm^{-1})
$Cl_2C=CH-Cl$	0.27	41
$Br_2C=CH-Br$	0.23	40
$Me_2C=CH-CN$	0.26	
E-PhCH=CH-CN	1.24	
E-PhCH=CH-NO_2	1.06	
E-NC-CH=CH-CN	7.2	45
Z-NC-CH=CH-CN	6.2	
$(NC)_2C=CH-OEt$	18.	55

$$CN \sim NO_2 >> X$$

$$CN \sim X_3$$

C(sp)-H DONORS w. HMPA in CYCLOHEXANE: ALKYNES

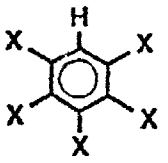
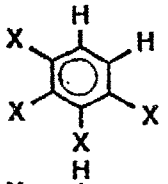
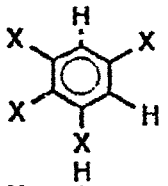
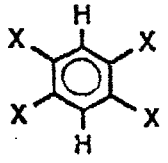
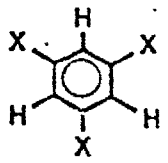
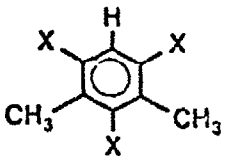
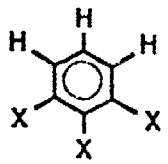
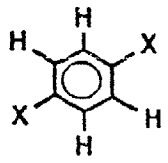
DONOR	K (M^{-1})	$-\Delta\bar{\nu}$ (cm^{-1})
$^tBu-C\equiv C-H$	0.44	82
$EtO-C\equiv C-H$	0.60	
$Ph-C\equiv C-H$	1.00	105
$BrCH_2-C\equiv C-H$	2.12	102
$EtO_2C-C\equiv C-H$	6.6	
$NC-C\equiv C-H$	48.	155

K: $NC^- > EtO_2C^- > BrCH_2^- > Ph^- > EtO^- > R^-$

σ: $NC^- > EtO_2C^- > EtO^- > BrCH_2^- > Ph^- > R^-$

K for EtO- is out of place.

C(sp^2)-H DONORS + HMPA in CYCLOHEXANE: AROMATICS

DONOR		K/H (M^{-1}) X = F	X = Cl	X = Br
		2.0	0.77	~0.1
		1.5	1.05	
		0.55	0.38	
		1.3	0.53	0.57
		0.22	0.21	0.23
				too small to measure
	H-4,6		0.60	
	H-5		1.5	
		0.18	0.23	

C(sp^2)-H DONORS w. HMPA in CYCLOHEXANE: AROMATICS, Cont'd.

	K/H (M^{-1})	
DONOR	**X = F**	**X = Cl**
Benzene–measured	**~0.05**	**~0.05**
Benzene–extrapolated	**0.05**	**0.25**

Steric crowding is important:

1) F > Cl > Br, **except** for 1,3,5-Tri- & 1,4-Di-;

2) C_6HCl_5 < 1,2,3,4-$C_6H_2Cl_4$;

3) 2,4,6-Tri-Br-m-xylene too small to measure;

4) K/H for Benzene extrapolated from poly-Cl-benzenes disagrees w. measured value.

BLANK

Using Theoretical Descriptors in Quantitative Structure Activity Relationships

George R. Famini and Leland Y. Wilson, U.S. Army Chemical Research, Development and Engineering Center, Aberdeen Proving Ground, MD, 21010, and Loma Linda University, Riverside, CA, 92515

1. Introduction

Quantitative Structure Activity Relationships (QSAR) have been used extensively in correlating structural features to physical, biological and toxicological properties. The basic tenet of QSAR is that there is a connection between the microscopic (molecular structure) and the macroscopic (empirical) properties. Further, this connection can be used to predict empirical properties directly from the molecular structure. This relationship was first quantized by Hammett, who developed the *Linear Free Energy Relationship* (LFER).

1.1 Linear Solvation Energy Relationships

Based on the concept of Hammett, Kamlet and Taft developed a methodology for developing Free Energy Relationships based on solute/solvent interactions.[1-4] This relationship, defined as the *General Linear Solvation Energy Relationship*, is shown in equation 1.

$$LOG\ Property = Steric + Polarizability + Hydrogen\ Bonding \quad (1)$$

In this way, a given property can be described as a linear expression consisting of contributions due to steric factors, polarization and polarizability factors, and hydrogen bonding factors. In the multiple solute - single solvent system, the specific terms are: Steric- Molar Volume (V_m), Polarizability- spectroscopically determined polarizability (π^*), Hydrogen Bonding- spectroscopically determined acidity and basicity terms (α and β). In practice, not all four terms are required for every relationship. The coefficients and associated statistical t-scores can be used to gauge the importance of each descriptor for every property correlation. In this way the LSER can be used to infer insight to solute/solvent interactions.

One major difficulty of this approach has been the nature of the descriptors. All are empirically determined, therefore reducing the usefulness of this approach for *a priori* predictions. Some attempts have been made correlating more fundamental structural and electronic descriptors with the Kamlet-Taft solvatochromic parameters with moderate degrees of success.[5]

1.2 Applications of Theoretical Chemistry

Theoretical chemistry has been used in the past to supply structural and electronic descriptors for QSAR and QSAR-like equations. In this way, empirically derived descriptors can be replaced in relationships such as the LSER with descriptors that will make the relationships more generally applicable. The Theoretical Linear Solvation Energy Relationship (TLSER) is such a derivation, using the LSER philosophy and general structure, but replacing the empirically derived descriptors with computationally derived descriptors.

2. Calculational Procedures

All geometries were optimized using the MNDO algorithm within MOPAC.[6,7] The molecular volumes were generated from the optimized geometries using the method of Hopfinger, as incorporated in the U.S. Army developed molecular modeling package **MMADS.**[8,9] All experimental data and LSER parameters were taken from works of the original authors.

3. The TLSER Descriptors

The TLSER descriptors have been developed with two main goals in mind. First, the TLSER descriptors should correlate optimally with the LSER parameters. Second, the property correlation equations with the TLSER should yield correlation coefficients and standard deviations as accurate as the LSER. Further, the TLSER descriptors should be as generally applicable to any solute/solvent interaction as are the LSER descriptors.

The TLSER descriptors that have been developed follow that of the LSER, and fit into one of the categories listed in equation 1. The steric term for the TLSER is the molecular Van der Waal's volume (V_{mc}). Volume calculations of this type are standard in most molecular modeling packages available today. As would be expected, correlation between V_m and V_{mc} is high, with a correlation coefficient of 0.979 and a standard error of the estimate of 5 ml/mol.

The polarizability term used in the TLSER is determined from the method of Stewart and Dewar incorporated in MOPAC[10]. Dividing the resulting polarizability by the V_{mc} results in a size independent Polarizability Index (π_I). This term defines the eases in which the electron cloud can be moved or polarized. Aromatics would rank high on the scale, and alkanes low. In addition, the π_I is inversely proportional to the electronegativity.

Like the LSER, the hydrogen bonding term is separated into acid (acceptor) and and base (donor) terms. Furthermore, because the hydrogen bond (or the Lewis bond or in essence any bond) can be separated into covalent and electrostatic parts, individual descriptors are needed for the TLSER to describe this. The energies of the highest occupied molecular orbital (representing the basicity, (ϵ_b) and the lowest

unoccupied molecular orbital (representing the acidity, ϵ_a) are used to describe the covalent interactions. Similarly, the electrostatic portions of the basicity and acidity are represented by selected atomic formal charges in the molecule, either the most negative (for the basicity, q_-) or the most positive hydrogen (for the acidity, q_+). The basicity terms (ϵ_I and q_-) correlate highly with the LSER β term, with a correlation coefficient of 0.9518 and a standard error of the estimate of 0.09.

The final generalized TLSER equation, is shown below:

$$\text{LOG P} = mV_{mc} + p\pi_I + a\epsilon_a + a'qH_+ + b\epsilon_b + b'q_-$$

The relationship is similar in appearance to the LSER, and contains all of the key points that have made the LSER successful.

4. Representative Correlations

Similar to the LSER, all six TLSER descriptors are seldom required or are statistically significant for a given solute/solvent property. Generally, the equations reduce to a three or four descriptor correlation. Table 1 shows a representative set of correlations using the TLSER descriptors.[11-16] In each case, the descriptors that were significant at the 95% confidence level were included, and all others dropped from the correlations. An in depth treatment of this technique is given in a recent journal article, and several technical reports.[17,18] The next paper will describe some correlations completed in depth pertaining to selected physical properties.

Table 1. TLSER Correlations

Property	V_{mc}	π_I	ϵ_b	ϵ_a	q_-	qH_+	C	N	R	SD
Kow	2.995	-0.847	1.730	n/a	-5.415	n/a	-3.960	64	0.957	0.357
LC_{50}	-0.928	-10.557	-1.442	n/a	0.443	n/a	18.082	32	0.943	0.574
EC_{50}	-4.067	-4.147	3.902	n/a	-2.777	n/a	11.356	25	0.982	0.332
EA	n/a	-2444	n/a	3540	1430	-383	20545	23	0.951	308
k_{75}	1.185	-0.610	n/a	n/a	-2.290	-1.077	-0.083	19	0.986	0.073
k_{OH}	-97.7	874	1256	n/a	n/a	n/a	-2184	10	0.986	26.9
C(Tadpole)	-2.13	-4.891	-4.980	n/a	5.00	n/a	11.99	39	0.979	0.237
C(Gold Orfe)	2-2.64	94.885	n/a	n/a	4.15	-1.77	7.55	30	0.975	0.237

5. Conclusions

The diversity of the properties successfully correlated with the TLSER show the general applicability of both the LSER and the application of theoretical techniques to the LSER. In addition, the TLSER descriptors, based solely on theoretically derived and determined parameters, result in as good a correlation as the LSER descriptors. Using these equations, then, solute/solvent interactions can be related to fundamental descriptions (structural and electronic) of the molecule. Furthermore, using the TLSER descriptors, *a priori* prediction of solute/solvent properties can be made using these equations.

6. References

1. Kamlet, M.J., Taft, R.W., Abboud, J-L.M., *J. Am. Chem. Soc.*, **91**, 8325, (1977).
2. Kamlet, M.J., Taft, R.W., Abboud, J-L.M., Abraham, M.J., *J. Prog. Org. Chem.*, **48**, 2877, (1983).
3. Kamlet, M.J. and Taft, R.W., *Acta Chem. Scand.*, **B39**, 616, (1985).
4. Kamlet, M.J., Taft, R.W., Famini, G.R., Doherty, R.M., *Acta Chem. Scand*, **B41**, 589, (1987).
5. Lewis, D.F.V., *J Comp Chem*, **8**(8), 1084, 1988.
6. Thiel, W. and Dewar, M.J.S., *J. Am. Chem. Soc.*, **99**, 4899, (1977).
7. Stewart, J.J.P, "MOPAC: A General Molecular Orbital Package", FJSRL-TR-86-0003, Frank J. Seiler Research Laboratory, U.S. Air Force Academy, Colorado Springs, CO, June 1988.
8. Hopfinger, A.J., *J. Am. Chem. Soc.*, **102**, 7126, (1980).
9. Leonard, J.M. and Famini, G.R., "A User's Guide to the Molecular Modeling Analysis and Display System", CRDEC-TR-030, U. S. Army Chemical Research, Development and Engineering Center, Aberdeen Proving Ground, MD., January 1989.
10. Stewart, J.J.P. and Dewar, M.J.S., *Chem. Phys. Lett.*, **111**, 416, (1984).
11. Kamlet, M.J., Taft, R.W., Abboud, J-L.M., Famini, G.R., Doherty, R.M., *J. Pharm. Sci.*, **74**(8), 807, (1985).
12. De Zwart, D., Slooff, W., *Aquat. Toxicol.*, **4**, 129, (1983).
13. Hafkensheid, T.L and Tomlinson, E., *I. J. Pharmeceut.*, **17**, 1, (1983).
14. Larsson, L., *Acta Chem. Scand.*, **11**, 1131, (1953).
15. Dorohio, D. and Iancu, D., *Anal. Stiintiche ale Univer.*, **20**, 59, (1981).
16. Giusti, D.M., Conway, R.J., Lawson, C.T., *J. Water Poll. Control Fed.*, M. Randic, *J Am Chem Soc*, *97*, 6609(1975).
17. L.Y. Wilson and G.R. Famini, *J Med Chem*, **34**,1668(1991).
18. For a more complete treatment of the TLSER descriptors, see either reference 9, or the following U.S. Army Technical Reports, available through NTIS: G.R. Famini, *Using Theoretical Descriptors in Structure Activity Relationships I. Molecular Volume*, CRDEC-TR-88031, U.S. Army Chemical Research, Development and Engineering Center, Aberdeen Proving Ground, MD,

January 1988, UNCLASSIFIED Report, NTIS# ADA191522; b) G.R. Famini, *Using Theoretical Descriptors in Structure Activity Relationships II. Polarizability Index*, CRDEC-TR-88137, U.S. Army Chemical Research, Development and Engineering Center, Aberdeen Proving Ground, MD, September 1988, UNCLASSIFIED Report, NTIS# ADA199594; c) L.Y. Wilson and G.R. Famini, *Using Theoretical Descriptors in Structure Activity Relationships III. Electronic Descriptors*, CRDEC-CR-88083, U.S. Army Chemical Research, Development and Engineering Center, Aberdeen Proving Ground, MD, August 1988, UNCLASSIFIED Report, NTIS# ADA200482; d) G.R. Famini, *Using Theoretical Descriptors in Structure Activity Relationships IV. Molecular Orbital Basicity and Electrostatic Basicity*, CRDEC-TR-013, U.S. Army Chemical Research, Development and Engineering Center, Aberdeen Proving Ground, MD, November 1988, UNCLASSIFIED Report, NTIS# ADA202132; e) G.R. Famini, *Using Theoretical Descriptors in Structure Activity Relationships V. A Review of the Theoretical Parameters* , CRDEC-TR-085, U.S. Army Chemical Research, Development and Engineering Center, Aberdeen Proving Ground, MD, July 1989, UNCLASSIFIED Report, NTIS# ADA213580;

USING THEORETICAL DESCRIPTORS IN QUANTITATIVE STRUCTURE ACTIVITY RELATIONSHIPS: SOME PHYSICOCHEMICAL PROPERTIES

George R. Famini, Carl A. Penski and Leland Y. Wilson*

U.S. Army Chemical Research, Development and Engineering Center,Aberdeen Proving Ground, MD 21010 and *Department of Chemistry, La Sierra University Riverside, Riverside, CA 92515

Introduction:

Quantitative Structure Activity Relationships (QSAR) have been used extensively to correlate particular structural features of compounds with their known biological properties (activities). For a selected property, an equation is obtained that permits the prediction of biological activities of compounds analogous to the compounds in the data set[1]. Often the chosen structural parameters (descriptors) have been limited in their ability to account for anything but local data sets. QSAR, as developed by Hansch, has been greatly modified by several researchers to increase its ability to correlate properties[2,3]. Kamlet and Taft developed an empirical parameter set, called the solvatochromic parameters, for use in Linear Solvation Energy Relationships (LSER). These LSER parameters have been very successful in correlating a wide range of biological activities; moreover, a strong point of these descriptors is that they also correlate a wide range of chemical and physical properties involving solute/solvent interactions [4].

Theoretical Linear Solvation Energy Relationships:

Based on the LSER concept, a new set of parameters for correlating a wide variety of properties has been developed [5,6]. Termed the Theoretical Linear Solvation Energy Relationship (TLSER) descriptors, these parameters are determined solely from computational methods thus permitting *a priori* prediction of properties. As indicated by George (Famini) in the previous presentation there are six TLSER descriptors: molecular volume, V_{mc}; polarizability index, π_I; covalent basicity, ϵ_b; electrostatic basicity, q_-; covalent acidity, ϵ_a; and electrostatic acidity, q_{H+}. The last four descriptors describe the covalent and electrostatic contributions to the hydrogen bonding basicity and acidity. Increasing ϵ_a and ϵ_bb values indicate decreasing acidity and basicity, respectively. The parameters are scaled by powers of ten for convenience in presentation and comparison of coefficients. The generalized equation is

$$\log P = aV_{mc} + b\pi_I + c\epsilon_b + dq_- + e\epsilon_a + fq_{H+} + g$$

where P represents the property. For a given property, the coefficients, a through g, are determined by fitting data for a set of compounds. Often not all terms are needed.

Table 2 shows the results of the application of the TLSER parameters to some biological activities from a previous paper [6]. Table 1 give the results for the corresponding LSER correlations. Both give good correlations and have physically reasonable interpretations. The descriptors have very small cross correlation in these data sets.

Properties in this presentation include pKa[7,8], charcoal adsorption[9], HPLC retention index[10], electronic absorption of a pyridinium ylide[11], Kow (octanol-water partition coefficient)[12], and the rate constant for the hydrolysis of organophosphonothioates[13].

Calculational Procedure:

Molecular geometries were optimized and TLSER descriptors were calculated using the MNDO algorithm contained in MOPAC[14,15]. The molecular volume for the optimized geometry was determined using the algorithm of Hopfinger[16]. The in house developed molecular modeling package, MMADS, was used to construct and view all molecular structures[17]. Multilinear regression analysis was used to obtain the coefficients in the correlation equation. Terms significant at the 0.95 level were retained and compounds with residuals greater than three standard deviations were considered outliers.

The properties and compounds were chosen to coincide with sets of data that describe a variety of solvent/solute interactions.

Results:

The TLSER equation obtained for the pKa of a series of organic compounds (including mono- and dicarboxylic acid, alcohols, thiols, ketones, phosphorous and nitrogen containing species) is

$$pKa = -86.7\,\epsilon_b + 228\epsilon_a - 62.3q_{H+}$$

(-6.44) (13.1) (-8.19)

n=43 R=0.971 s.d.=2.78 F=220

The quantities in parentheses are the t-statistic values for the coefficients. There is a strong correlation between ϵ_b and ϵ_a indicating that one should be eliminated from the equation. The resulting equation is

$$pKa = 127\,\epsilon_a - 48.7\,q_{H+}$$

(11.7) (-6.94)

n=43 R=0.940 s.d.=2.79 F=156

These equations lead to the following observations. (1) Only the electronic descriptors are significant (more specifically the hydrogen bonding terms). (2) The pKa increases with decreasing acidity. (3) The pKa increases with increasing basicity. These inferences are chemically reasonable.

For charcoal adsorption from aqueous solution for a set of compounds primarily made up of alkyl alcohols, ketones, and aldehydes the relation is

$$\log A = 2.46(V_{mc}/100) + 5.78\pi_I - 4.66q_- - 6.68.$$

n=33 R=0.955 s.d.=0.24 F=99.8

This equation leads to the following observations. (1) The molecular volume, V_{mc}, , increases absorption on charcoal; larger molecules would tend to be excluded from the aqueous phase. (2) The polarizability index, π_I, increases absorption on charcoal; greater polarizability would permit greater interaction with the pi electrons on the charcoal. (3) The electrostatic basicity decreases absorption on charcoal; greater charge would tend to increase the interaction with the (polar) water molecules.

For a set of alkyl and aryl compounds the equation for the HPLC retention index in a 50% methanol-water mixture is

$$\log k50 = 1.83(V_{mc}/100) - 3.05q_{-} - 1.52q_{H+} - 0.459.$$

(11.8) (-10.8) (-2.96) (-2.71)

n=21 R=0.980 s.d.=0.124 F=141

The equation leads to the following observations. (1) The volume term increases the retention on the column; the larger molecules would tend to be excluded from the polar solvent. (2) The electrostatic bonding descriptors decrease the retention; there is greater interaction with the (polar) water molecules. We might expect these if the HPLC column is not polar. The relation for a 75% methanol-water solution has the same terms and signs and has R=0.981.

For the electronic absorption for the three pyridinium ylides (in a set of alkyl and aryl solvents) examined here the best fit was for a compound with two ethoxy groups. The relation for the electronic absorption (cm^{-1}) of this compound is

$$EA = -45734\pi_I + 8733q_{H+} + 27629.$$

(-7.40) (9.02) (39.4)

n=22 R=0.960 s.d.=279 F=112

This equation leads to the following observations. (1) The absorption peak position depends on electronic descriptors; this is reasonable since it involves transitions in electronic states. (2) The polarizability index decreases the position of the absorption peak (a red shift); this implies an interaction of the pi electrons on the ylide ring with the electrons of the solvent. (3) The electrostatic acidity increases the position of the absorption peak (a blue shift); it suggests an interaction of an acidic site on the solvent molecule with a basic site on the pyridinium ylide (maybe the negatively charged carbon next to the nitrogen on the ring). The relations for two other ylides have the same terms and signs as in the above equation; their R values are 0.904 and 0.930.

The rate constant for the hydrolysis of a set of organophosphonothioates is described by this equation.

$$\log kOH = -13.1\,\pi_I - 6.11\,\epsilon_a + 7.09.$$

(-2.98) (-40.2) (13.6)

n=35 R=0.990 s.d.=0.117 F=817

This relation makes the following points. (1)The rate constant depends only on the electronic descriptors; the bulk/steric term of the substituents in this set of compounds was not significant. It could imply that the mechanism involves attack near the P-S bond away from the R groups. (2) The rate decreases with increasing polarizability; greater ease in distortion of the electron "cloud" interferes with the reaction site. (3) The rate increases with increasing hydrogen bonding acidity; this is in keeping with the involvement of a hydroxide ion.

A selection of aliphatic and aromatic compounds gave the following relationship for the octanol-water partition coefficient.

$$\log Kow = 3.00\,(V_{mc}/100) - 8.47\,\pi_I + 17.3\,\epsilon_b - 5.42\,q_{-} - 3.96.$$

(11.7) (-0.37) (2.76) (-11.6)

N=64 R=0.957 s.d.=0.36

Observations from this relation follow. (1) The bulk/steric (V_{mc}) term is significant as it is in the previous distribution equilibria: charcoal absorption and HPLC retention index. The larger the volume the more readily the molecule tends to be excluded from the polar solvent. (2) The hydrogen bonding basicity enhances the solubility in the octanol over that in water; an interpretation is that an interaction with the lowest unoccupied molecular orbital (LUMO) in the octanol tends to be favored over the corresponding one in water. (3) The more negative the formal charge the less the tendency to dissolve in the octanol; this is in keeping with a charge-dipole interaction between the negative charge and the polar water molecule. (4) The polarizability index is not significant but was included because the corresponding LSER relation contained an analogous term; however the magnitudes and signs were very similar.

Conclusions:

The application of the TLSER descriptors to the six physico-chemical properties in this paper as well as the five toxicological properties studied earlier point out these features: (1) The descriptors give good to very good correlations ; (2) They give physically reasonable interpretations ; (3) They apply to a wide range of properties ; (4) They apply to a wide range of compounds.

An attractive feature is that they are obtained from computation and not experiment. Consequently they permit almost *a priori* prediction of properties. Furthermore, the physical interpretation can suggest modes of interaction between solute and solvent molecules; in that sense they provide a probe.

References:

1. Gupta, S. *Chem. Rev.* **1987**, *87*, 1183.
2. Hansch, C. *Accts. Chem. Res.* **1969**, *B2*, 232.
3. Kamlet, M.J., Taft, R.W., Abboud, J-L.M., *J. Am. Chem. Soc.*, **1977**, *91*, 8325.
4. Kamlet, M.J., Taft, R.W., Famini, G.R., Doherty, R.M. *Act. Chem. Scand.* **1987**, *B41*, 589.
5. Famini, G.R, *Using Theoretical Descriptors in Structure Activity Relationships V*; CRDEC-TR-085, U.S. Army Chemical Research, Development and Engineering Center, Aberdeen Proving Ground, MD, July 1989.
6. Wilson, L.Y., Famini, G.R. *J. Med. Chem.* **1991**, *34*, 1668.
7. Körtum, G., Vogel, W., Andrewsson, K. *Pure and Appl. Chem.* **1960**, *1*, 190.
8. Serjent, E., Dempsey, B. *Ionic Constants of Organic Acids: Aqueous Solutions* Pergamon Press, New York, 1979.
9. Abe, H., Hayashi, K., Kitigawa, M., Uruhata,T. *Bull.Chem.Soc. Japan* **1980**,*53*,1199.
10. Hafkensheid, T.L., Tomlinson, E. *I. J. of Pharmaceutics* **1983**,*17*, 1.
11. Dorohoi, D., Iancu, D. *Anale Stiintiche ale Univeritatii* **1981**,*20*, 59.
12. Kamlet, M. J., Taft, R. W., Abboud, J-L.M., Famini, G.R., Doherty, R.M., Abraham, M.J. *J. Pharm. Sci.* **1985**, *4*, 807.
13. Kabachnik, M.I, et. al. *Russian Chemical Reviews* **1970**, *39*, 485.

14. Dewar, M.J.K., Thiel, W. *J. Am. Chem. Soc.* **1977**,*99*, 4899.

15. Stewart,J.J.P. *Mopac Manual*; FJSRL-TR-88-0007; Frank J. Seiler Research Laboratory, U.S. Air Force Academy; Colorado Springs, CO, December 1988.

16. Hopfinger, A.J. *J. Am. Chem. Soc.* **1980**, *102*, 7126.

17. Leonard, J.M., Famini, G.R. *A User's Guide to the Molecular Modeling Analysis and Display System*; CRDEC-TR-030; U.S. Army Chemical Research, Development and Engineering Center; Aberdeen Proving Ground, MD, January 1989.

BLANK

AB INITIO QUANTUM CHEMICAL INVESTIGATIONS OF SOLVATOCHROMIC AND SUBSTITUENT PARAMETERS

Peter Politzer, Jane S. Murray, Tore Brinck and M. Edward Grice

Department of Chemistry
University of New Orleans
New Orleans, Louisiana 70148

Introduction

We have found several correlations between computed properties [the electrostatic potential $V(\mathbf{r})$ and the average local ionization energy $\bar{I}(\mathbf{r})$] and (a) the solvatochromic hydrogen-bond-donor and -acceptor parameters α and β, (b) the Hammett constants σ_p and σ_m, and (c) the inductive parameters σ_I [1-7]. These relationships will be discussed following a brief synopsis of our computational methodology.

Computed Properties

1. *Electrostatic Potential V(r):*

The electrostatic potential $V(\mathbf{r})$ that the nuclei and electrons of a molecule create in the surrounding space is expressed rigorously by eq. (1):

$$V(\mathbf{r}) = \sum_A \frac{Z_A}{|\mathbf{R}_A - \mathbf{r}|} - \int \frac{\rho(\mathbf{r}')\, d\mathbf{r}'}{|\mathbf{r}' - \mathbf{r}|} \qquad (1)$$

Z_A is the charge on nucleus A, located at $\mathbf{R}_A$; $\rho(\mathbf{r})$ is the electronic density function of the molecule. $V(\mathbf{r})$ expresses the net electrical effect of the nuclei and electrons of a molecule at any point $\mathbf{r}$ in the surrounding space, and is well established as an effective tool in interpreting and predicting the reactive behavior of molecules [3,8-11]. The sign of $V(\mathbf{r})$ in any particular region depends upon whether the effects of the nuclei or the electrons are dominant there. Thus an approaching electrophile will initially be attracted to those regions in which $V(\mathbf{r})$ is negative, and in particular to the points where $V(\mathbf{r})$ has its most negative values (the local minima, or V_{min}).

While sites susceptible to electrophilic attack can be identified and ranked quite readily by means of the local $V(\mathbf{r})$ minima [3,8-11], the situation is not as straightforward for nucleophilic interactions, due to the fact that $V(\mathbf{r})$ maxima are found only at the positions of the nuclei [12]. These reflect the magnitudes of the nuclear charges and therefore cannot be assumed to indicate relative reactivities toward nucleophiles. There is accordingly no criterion for nucleophilic interactions that corresponds directly to V_{min} for electrophilic ones. However we have developed

techniques that do allow the electrostatic potential to be used to interpret nucleophilic processes [5,13-18]. For example, investigation of V(r) on two- or three-dimensional surfaces significantly removed from the nuclei has revealed buildups of positive potential that do reflect relative affinities for nucleophiles [5,14-18]. One particularly promising approach involves plotting V(r) on a molecular surface defined by the 0.002 or 0.001 electron/bohr^3 contour of the electronic density function $\rho(r)$ [19]. These contours have been shown to contain at least 95% of the total electronic charge, and to provide physically reasonable molecular dimensions [20,21].

2. *Average Local Ionization Energy $\bar{I}(r)$:*

We have recently introduced a property which we interpret as the average local ionization energy, $\bar{I}(r)$ [2]. It is rigorously defined within the framework of self-consistent-field molecular orbital (SCF-MO) theory by eq. 2.

$$\bar{I}(r) = \sum_i \frac{\rho_i(r)\,|\varepsilon_i|}{\rho(r)} \qquad (2)$$

$\rho_i(r)$ is the electronic density of the i^{th} molecular orbital at the point r, ε_i is the orbital energy and $\rho(r)$ is the total electronic density function of the molecule. We have shown that $\bar{I}(r)$ plotted on well-defined molecular surfaces is a quantitative indicator of reactivity toward electrophiles in a variety of chemical systems [2,7,22-26]. The positions where the surface $\bar{I}(r)$ has its lowest values ($\bar{I}_{S,min}$) are the sites with the least tightly bound electrons, and thus most susceptible toward electrophilic attack.

Hydrogen-Bond-Donor and -Acceptor Correlations

Linear solvation energy relationships based upon the use of "solvatochromic parameters" have resulted from extensive efforts to quantify solvent effects on various experimentally-observed quantities (e.g. rate constants, equilibrium constants and IR, NMR, ESR and UV/vis absorption maxima and intensities) [27-31]. Two of the solvatochromic parameters, designated as α and β, have been interpreted as providing measures of a solvent's ability to donate or accept a proton, respectively, in solute-solvent hydrogen bonding [31].

We have recently found that good relationships exist between β and the most negative electrostatic potentials associated with the hydrogen-bond-accepting heteroatoms in four families of compounds, taken separately: ten azines, four primary amines, four alkyl ethers and fifteen molecules containing double-bonded oxygens [4]. The correlation coefficients were found to be 0.96, 0.98, 0.94 and 0.95, respectively. We computed the V_{min} at the STO-5G level, using STO-3G optimized geometries.

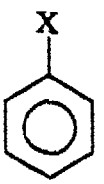
In view of the good correlations between β and V_{min}, it was suggested that there might also be one between the hydrogen-bond-donating parameter α and positive regions of V(r) [32]. Since this corresponds to interaction with a nucleophile, we computed V(r) on the surfaces of the molecules; these were defined by the 0.001 electron/bohr^3 contour of the molecular electronic density. We did indeed find good relationships between α and the most positive surface electrostatic potential, $V_{S,max}$, for two groups of hydrogen bond donors taken separately [5,6]; the correlation coefficients for both are 0.97. In all instances the $V_{S,max}$ are associated with hydrogens that could be donated in a hydrogen bond. The electrostatic potentials were computed at the STO-5G level (using STO-3G optimized geometries).

These relationships between α and $V_{S,max}$ and β and V_{min} confirm that the electrostatic potential in the space around a gas phase molecule is a key (but not the sole) factor in determining its ability to donate or accept a proton in a hydrogen bond, and they provide a predictive capability for obtaining unknown α and β values.

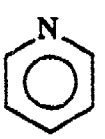
Hammett Constant Correlation

We have shown for a series of monosubstituted benzenes (I) [2] and some azines and azine N-oxides [e.g. pyridine (II) and pyridine N-oxide (III)] [23,24] that $\bar{I}_{S,min}$ values are a quantitative measure of the activating/deactivating and directing tendencies of the various substituents in regard to electrophilic aromatic substitution. The ring carbon $\bar{I}_{S,min}$ of the systems

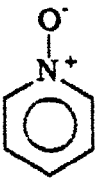
X = NH_2, CH_3, H, F, CHO, NO_2



I



II



III

bearing *ortho*- and *para*- directing substituents (or atoms) (NH_2, CH_3, F, N^+-O^-) are located above the *ortho* and *para* ring positions, while those for species with *meta*-directing substituents or atoms (NO_2, CHO, N) are positioned above the *meta* [2,23,24]. For the former, the $\bar{I}_{S,min}$ are smaller in magnitude than for benzene, showing that the ring has been activated, while for the latter the reverse is observed. $\bar{I}_{S,min}$ are also found in the vicinities of heteroatoms containing lone pairs, e.g. O, N, F and S.

The relationship between the ring carbon $\bar{I}_{S,min}$ and aromatic reactivity can also be put on a more quantitative basis, as evidenced by the excellent correlation found between the ring carbon $\bar{I}_{S,min}$ of the monosubstituted benzenes and the corresponding Hammett constants σ_p and/or σ_m [2]. (Linear correlation coefficients of 0.99 are obtained with $\bar{I}(r)$ computed at both the 6-31G*

and STO-5G levels, using STO-3G geometries [2,7].) The Hammett constants are well-established measures of the overall electron-withdrawing and -donating tendencies of substituent groups on benzene rings [33-35]. This relationship allows us to make predictions of Hammett constants for substituents for which these are not known. For example, we have predicted σ_p for NHF to be 0.06, and σ_p and σ_m for NF_2 to be 0.49 and 0.54, respectively [2].

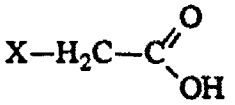
Correlations with σ_I

We have demonstrated that the amine nitrogen V_{min} of NH_2-X molecules provide a quantitative measure of the total electron-attracting tendencies of the substituents X [1]. We computed V_{min} at the STO-5G level using STO-3G geometries for twenty-four NH_2-X molecules and found correlations between V_{min} and σ_I, and V_{min} and $\sigma_I + \sigma_R$ when $\sigma_R > 0$ (electron-withdrawing); the correlation coefficients were 0.90 and 0.92, respectively [1]. For a subgroup of these molecules, we have more recently observed that both V_{min} and $\bar{I}_{S,min}$ correlate well with σ_I or $\sigma_I + \sigma_R$, when $\sigma_R > 0$; correlation coefficients range from 0.97 to 0.99 [6].

Correlations with pK_a

Since $\bar{I}_{S,min}$ is evidently a good index of reactivity toward electrophiles, and H^+ is certainly an electrophile, we have investigated the possibility that aqueous acidities (pK_a's) may correlate with the $\bar{I}_{S,min}$ of the conjugate bases. We found, for a wide variety of carbon, oxygen and nitrogen acids, that this is indeed the case [23,25,26]. In this manner we are able to predict pK_a values even for very strong acids for which aqueous acidities cannot be measured because of the leveling effect of water. It is important to mention that although both V_{min} and $\bar{I}_{S,min}$ can be viewed conceptually as indicative of reactivity toward electrophiles, $\bar{I}_{S,min}$ appears to be better suited to be a general measure of aqueous acidity [6,26].

As an extension of our $\bar{I}_{S,min}$ versus pK_a studies, we have investigated the surface $\bar{I}(r)$ of the conjugate bases of a series of substituted acetic acids (IV) [7]. The pK_a's of this series of acids



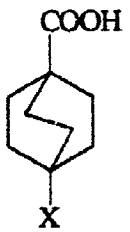
IV

X = H, CH_3, CH_2OH, NH_2, OH, F, Cl, CHO, CF_3, CN, NO_2

have been used as a basis for defining inductive substituent constants, σ_I [34,35]. In light of our recent findings, it seemed reasonable therefore to anticipate that $\bar{I}_{S,min}$ of the conjugate bases of substituted acetic acids might correlate with σ_I values of the corresponding substituent groups. The relationship between 6-31G*/3-21G $\bar{I}_{S,min}$ and σ_I for the molecules listed above, excluding

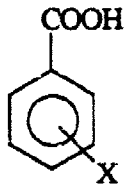
CH_2F-COOH, was found to be quite good; the linear correlation coefficient is 0.97. The point for fluoroacetic acid is outlying, as we anticipated from earlier studies of V_{min} and $\bar{I}_{S,min}$ of NH_2-X systems [1,6] and other experimental and theoretical studies [25,36-39] supporting the view that fluorine has a limited capacity for accepting additional electronic charge, despite an initial strong attraction for it.

Our relationship between $\bar{I}_{S,min}$ and σ_I presents a promising approach for theoretically determining substituent constants. In principle, the same procedure could be carried out for other reaction series used as means for determining substituent constants (σ_I, σ_m, σ_p, etc.), e.g. bicyclooctane carboxylic acids (V), protonated quinuclidines (VI) and benzoic acids (VII) [35].



V

VI



VII

Future Directions

As part of our continuing studies of molecular reactive behavior, we plan to:

1) investigate the relationship between $\bar{I}(\mathbf{r})$ and polarizability, possibly leading to the determination of local polarizabilities;
2) examine possible correlations between $V(\mathbf{r})$ and the *solute* parameters α_2^H and β_2^H, as was suggested by Dr. R. W. Taft and Dr. M. H. Abraham;
3) explore the use of the *para* -substituted aniline framework in measuring substituent effects as suggested by Dr. C. Hansch; and
4) compute surface $\bar{I}(\mathbf{r})$ for the conjugate bases of a variety of strong acids with the aim of predicting pK_a's.

Acknowledgement

We greatly appreciate the support of this work by the Office of Naval Research (contract #N00014-85-K-0217) and by the Army Research Office (contract #DAAL03-90-G-0205).

References

1. J. S. Murray and P. Politzer, Chem. Phys. Letters, 152, 364 (1988).
2. P. Sjoberg, J. S. Murray, T. Brinck, and P. Politzer, Can. J. Chem., 68, 1440 (1990).
3. P. Politzer and J. S. Murray, in Reviews of Computational Chemistry, K. B. Lipkowitz and D. B. Boyd, eds., VCH Publishers, New York, in press.
4. J. S. Murray, S. Ranganathan and P. Politzer, J. Org. Chem., in press.
5. J. S. Murray and P. Politzer, J. Org. Chem., submitted.
6. J. S. Murray, T. Brinck, M. E. Grice and P. Politzer, J. Mol. Struct. (THEOCHEM), submitted.
7. J. S. Murray, T. Brinck, and P. Politzer, J. Mol. Struct. (THEOCHEM), submitted.
8. E. Scrocco and J. Tomasi, Adv. Quantum Chem., 11, 115 (1978).
9. P. Politzer and K. C. Daiker, in The Force Concept in Chemistry, D. M. Deb, editor, Van Nostrand Reinhold Co., 1981, ch. 6.
10. P. Politzer and D.G. Truhlar, eds., Chemical Applications of Atomic and Molecular Electrostatic Potentials, Plenum Press, New York, 1981.
11. P. Politzer, P. R. Laurence and K. Jayasuriya, in Structure-Activity Correlation in Mechanism Studies and Predictive Toxicology, J. McKinney, ed., Special Issue of Environmental Health Perspectives, Vol. 61, 191 (1985).
12. R. K. Pathak and S. R. Gadre, J. Chem. Phys. 93 (1990) 1770.
13. P. Politzer, S. J. Landry and T. Warnheim, J. Phys. Chem. 86 (1982) 4767.
14. P. Politzer, L. Abrahmsen and P. Sjoberg, J. Amer. Chem. Soc. 106 (1984) 855.
15. P. Politzer, P. R. Laurence, L. Abrahmsen, B. A. Zilles and P. Sjoberg, Chem. Phys. Letters 111 (1984) 75.
16. P. Sjoberg and P. Politzer, J. Phys. Chem. 94 (1990) 3959.
17. J. S. Murray, P. Lane and P. Politzer, J. Mol. Struct. (THEOCHEM) 209 (1990) 163.
18. J. S. Murray, P. Lane, T. Brinck, P. Politzer and P. Sjoberg, J. Phys. Chem. 95 (1991) 844.
19. The original programs for computing and plotting properties on molecular surfaces were written by Dr. Per Sjoberg and Mr. Tore Brinck.
20. R. F. W. Bader, W. H. Henneker and P. E. Cade, J. Chem. Phys. 46 (1967) 3341.
21. R. F. W. Bader and H. J. T. Preston, Theor. Chim. Acta 17 (1970) 384.
22. J. S. Murray, J. M. Seminario, P. Politzer and P. Sjoberg, Internat. J. Quantum Chem.: Quantum Chemistry Symposium 24, 645 (1990).
23. T. Brinck, J. S. Murray, P. Politzer and R. E. Carter, J. Org. Chem., 56, 2934 (1991).
24. P. Lane, J. S. Murray and P. Politzer, J. Mol. Struct. (THEOCHEM), in press.
25. T. Brinck, J.S . Murray and P. Politzer, J. Org. Chem., in press.

26. J. S. Murray, T. Brinck, and P. Politzer, Int. J. Quant. Chem., Quantum Biology. Symp. 18, in press.
27. M.J. Kamlet and R. W. Taft, J. Am. Chem. Soc. 98 (1976) 377.
28. M. J. Kamlet, M. E. Jones, R. W. Taft and J.-L. M. Abboud, J. Chem. Soc. Perkin Trans. 2 (1979) 342.
29. M. J. Kamlet, A. Solomonovici and R. W. Taft, J. Am. Chem. Soc. 101 (1979) 3734.
30. M. J. Kamlet, J.-L. M. Abboud and R. W. Taft, Prog. Phys. Org. Chem. 13 (1981) 485.
31. M. J. Kamlet, J.-L M. Abboud, M. H. Abraham and R. W. Taft, J. Org. Chem. 48 (1983) 2877.
32. G. R. Famini, private communication.
33. L. P. Hammett, Physical Organic Chemistry, McGraw-Hill, New York, 1940.
34. O. Exner, Correlation Analysis of Chemical Data, Plenum Press, Czechoslavakia, 1988.
35. C. Hansch, A. Leo and R. W. Taft, Chem. Rev. 91 (1991) 165.
36. J. Hine and N. W. Burske, J. Am. Chem. Soc. 78 (1956) 3337.
37. K. R. Brower, B. Gay and T. L. Konkol, J. Am. Chem. Soc. 88 (1966) 1681.
38. P. Politzer, J. Amer. Chem. Soc. 91 (1969) 6235.
39. P. Politzer and J. W. Timberlake, J. Org. Chem. 37 (1972) 3557.

BLANK

AB INITIO QUANTUM CHEMICAL INVESTIGATIONS OF SOLVATOCHROMIC AND SUBSTITUENT PARAMETERS

Peter Politzer

Department of Chemistry
University of New Orleans
New Orleans, LA 70148

May 1991

AB INITIO QUANTUM CHEMICAL INVESTIGATIONS OF SOLVATOCHROMIC AND SUBSTITUENT PARAMETERS

We have found several correlations to exist between computed properties [the electrostatic potential $V(\mathbf{r})$ and the average local ionization energy $\bar{I}(\mathbf{r})$] and solvatochromic and substituent parameters. These relationships are listed below.

Nature of Correlation:

1. Hydrogen-bond-acceptor parameter β vs. $V(\mathbf{r})$ minima
2. Hydrogen-bond-donor parameter α vs. surface $V(\mathbf{r})$ maxima
3. Hammett constants σ_m and σ_p vs. surface $\bar{I}(\mathbf{r})$ minima
4. Inductive substituent constants σ_I vs.
   (a) V(r) minima
   (b) surface $\bar{I}(\mathbf{r})$ minima

ELECTROSTATIC POTENTIAL

$$V(\mathbf{r}) = \sum_{A} \frac{Z_A}{|\mathbf{R}_A - \mathbf{r}|} - \int \frac{\rho(\mathbf{r}')\, d\mathbf{r}'}{|\mathbf{r}' - \mathbf{r}|}$$

Z_A is the charge on nucleus A, located at $\mathbf{R}_A$, and $\rho(\mathbf{r})$ is the electronic density function of the molecule.

The electrostatic potential is a real physical property which expresses the net electrical effect of the nuclei and electrons of a molecule, and has emerged as an effective tool for studying the reactive behavior of molecules, including both electrophilic and nucleophilic processes, and also intermolecular hydrogen bonding and recognition interactions.

NEGATIVE REGIONS OF ELECTROSTATIC POTENTIAL

The electrostatic potential around a free neutral atom is positive everywhere. In response to electronic rearrangements that occur when atoms interact to form a molecule, regions of negative potential normally develop. Most of these can be classified into three general categories:

1) lone pair regions of heteroatoms:

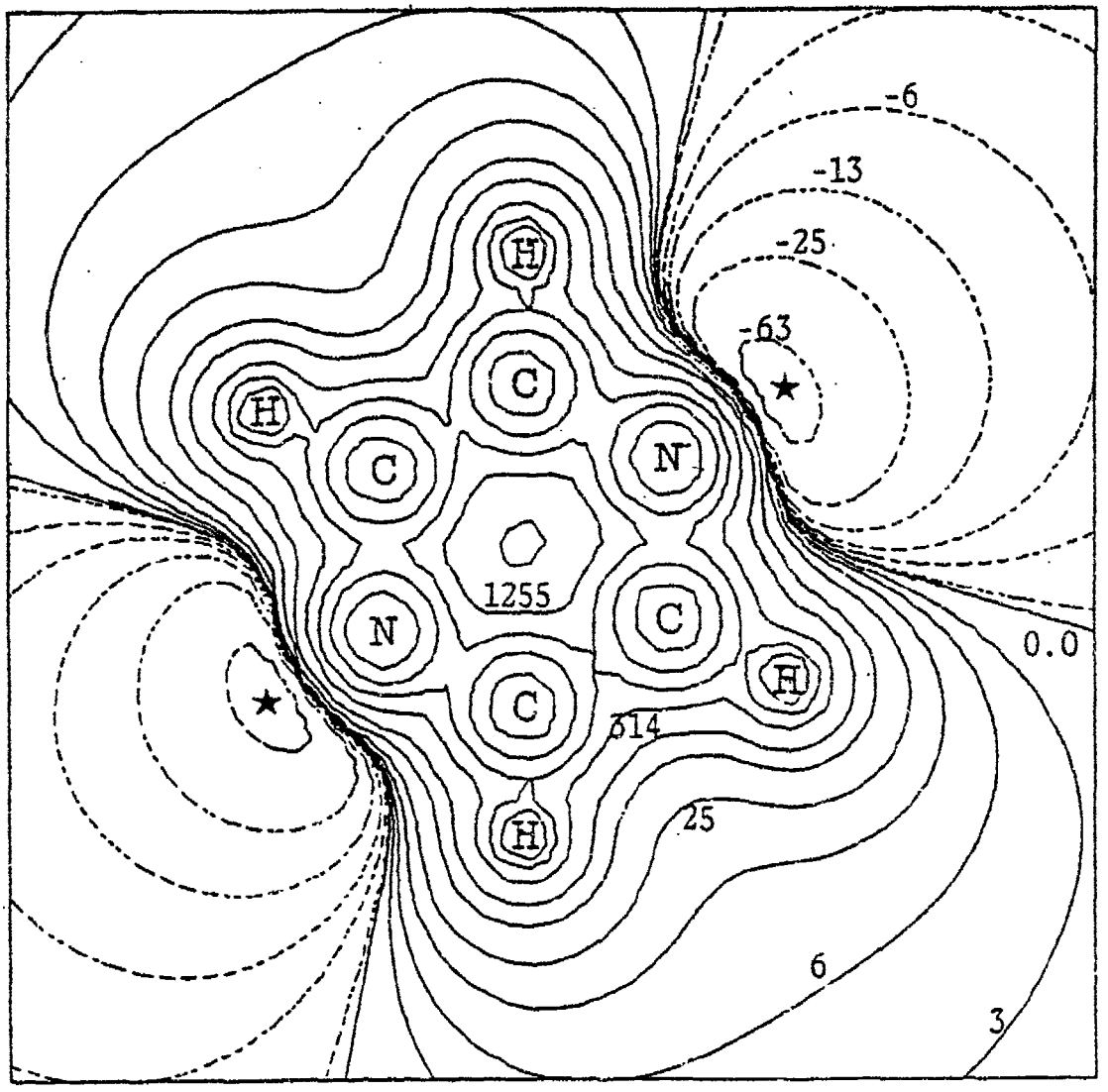
e. g. N, O, F, Cl, S, Br

2) π regions of unsaturated molecules:

e.g.  , $H_2C{=}CH_2$

3) C-C bond regions of strained hyrocarbons:

e.g. 



CALCULATED ELECTROSTATIC POTENTIAL IN THE MOLECULAR PLANE OF PYRAZINE

V_{min} = -82.4 kcal/mole (★)

RELATIONSHIPS BETWEEN HYDROGEN-BOND-ACCEPTOR PARAMETER β AND ELECTROSTATIC POTENTIAL MINIMA

The parameter β has been interpreted as providing a measure of a solvent's ability to accept a proton in a solute-to-solvent hydrogen bond. We have found correlations between β and the most negative electrostatic potentials (V_{min}) associated with the hydrogen-bond-accepting heteroatoms in four families of molecules, taken separately:

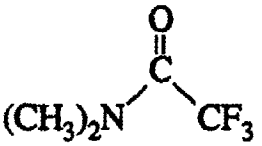
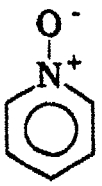
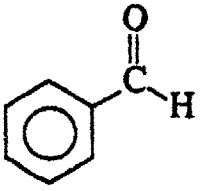
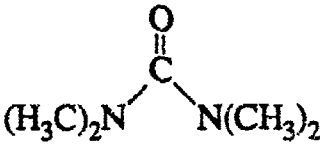
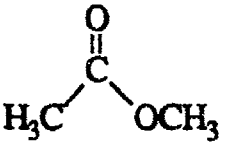
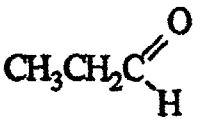
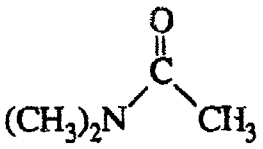
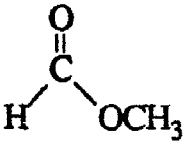
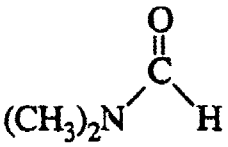
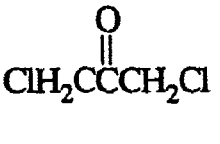
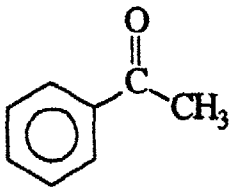
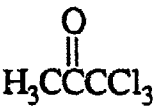
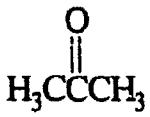
a) azines

b) primary amines

c) alkyl ethers

d) molecules containing double-bonded oxygens

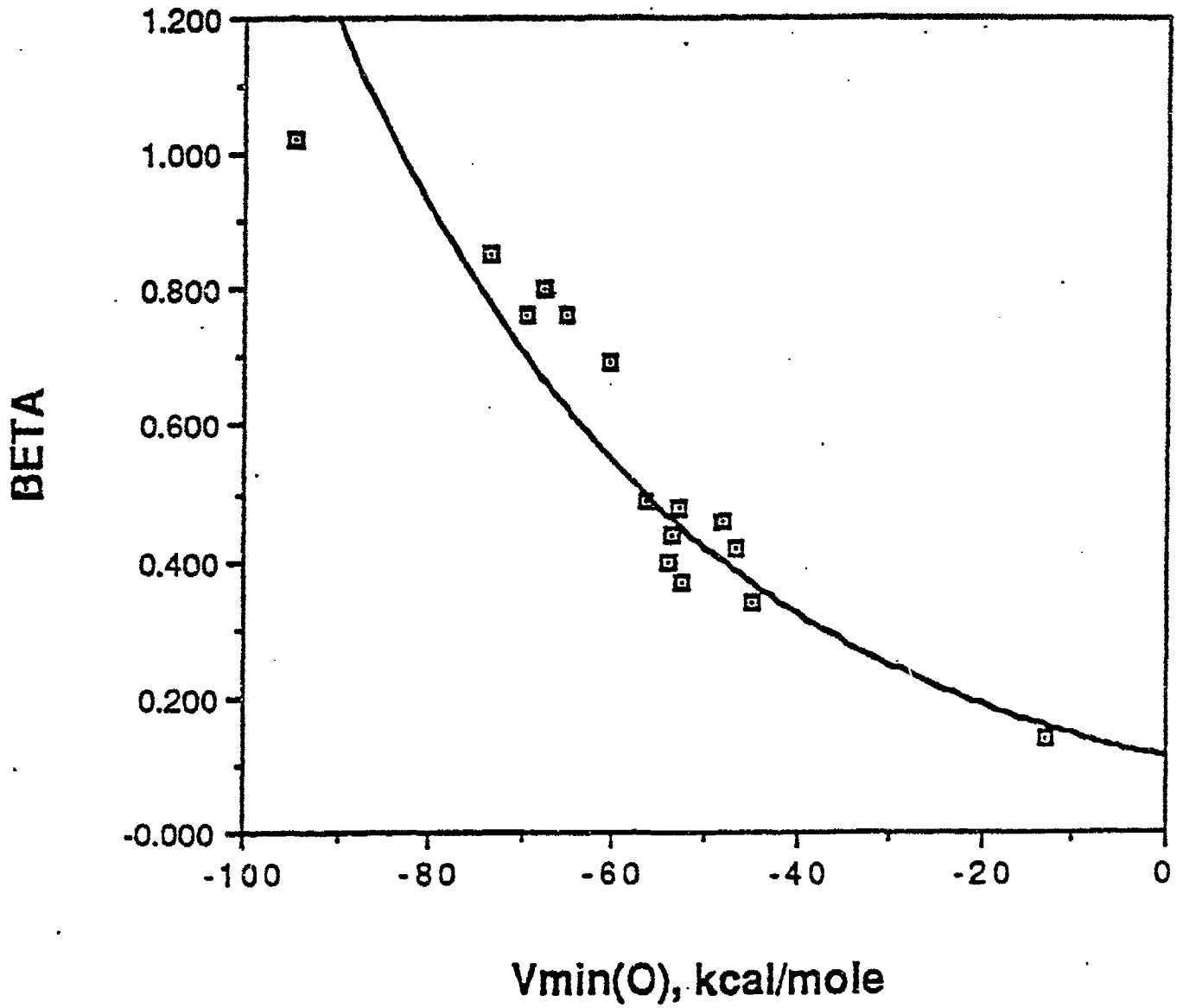
MOLECULES CONTAINING DOUBLE-BONDED OXYGENS:

Molecule	β	V_{min} (kcal/mole)	Molecule	β	V_{min} (kcal/mole)
$O{=}P(CH_3)_3$	1.02	-94.5		0.46	-48.2
	0.85	-73.4		0.44	-53.5
	0.80	-68.0		0.42	-46.6
$O{=}S(CH_3)_2$	0.76	-69.5		0.40	-54.0
	0.76	-65.2		0.37	-52.5
	0.69	-60.5		0.34	-45.0
	0.49	-56.4		0.14*	-13.0
	0.48	-52.9			

*Dr. R. W. Taft has pointed out that this value may be unreliable.

Correlation Coefficient = 0.95

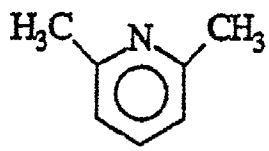
MOLECULES CONTAINING DOUBLE-BONDED OXYGENS:

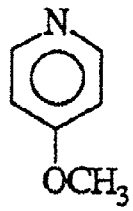


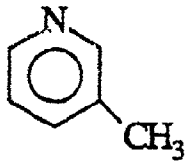
Correlation Coefficient = 0.95

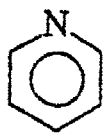
β versus V_{min}

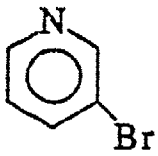
AZINES: Correlation Coefficient = 0.96

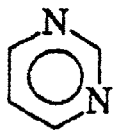


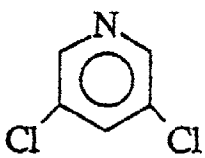






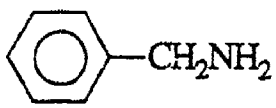


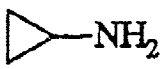




PRIMARY AMINES: Correlation Coefficient = 0.98

$CH_3(CH_2)_3NH_2$





$F_3CCH_2NH_2$

AKYL ETHERS: Correlation Coefficient = 0.94

$O[CH(CH_3)_2]_2$

$O(CH_2CH_3)_2$

$O(CH_2CH_2CH_3)_2$

$H_3COCH_2CH_2OCH_3$

POSITIVE REGIONS OF ELECTROSTATIC POTENTIAL

Maxima in $V(\mathbf{r})$ occur only at the positions of the nuclei of molecules. These reflect the magnitudes of the nuclear charges and do not indicate relative affinities for nucleophiles. Thus, there is no criterion for predicting nucleophilic attack that corresponds directly to V_{min} for electrophilic attack.

It is nevertheless possible to use $V(\mathbf{r})$ to interpret and predict nucleophilic processes. One useful approach involves examining $V(\mathbf{r})$ on a three-dimensional surface encompassing the molecule.

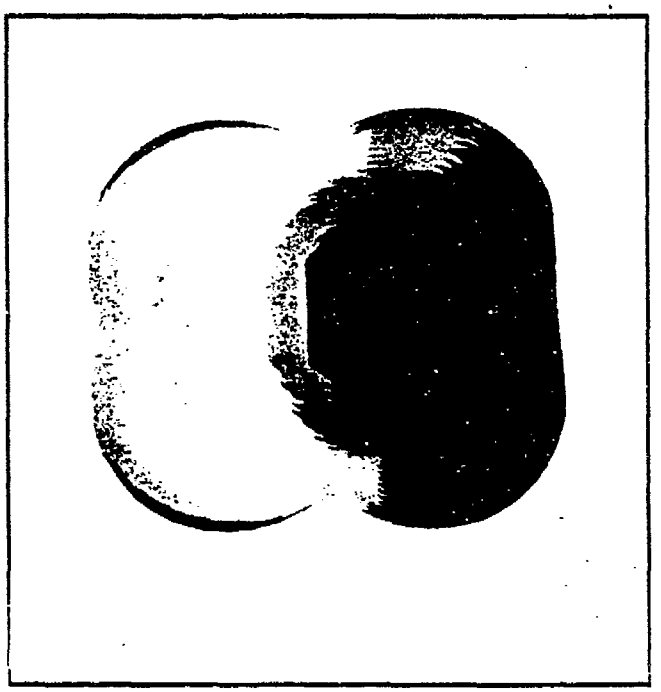
RELATIONSHIPS BETWEEN HYDROGEN-BOND-DONOR PARAMETER α AND SURFACE ELECTROSTATIC POTENTIAL MAXIMA

We have shown for two groups of hydrogen bond donors that surface V(**r**) maxima ($V_{S,max}$) correlate with the parameter α, which is viewed as a measure of a solvent's ability to donate a proton in a solvent-to-solute hydrogen bond.

V(**r**) Ranges (kcal/mole):
BLACK: Greater than 27.
DK.GRAY: Between 10 and 27.
LT.GRAY: Between 0 and 10.
WHITE: Negative.

$V_{S,max}$ = +28.6

[0.001 contour of electron density; hydroxyl group faces viewer.]



Surface V(**r**) of Ethanol

CORRELATIONS BETWEEN α AND $V_{S,max}$

1. Molecules containing OH groups:

$F_3CCH(OH)CF_3$

F_3CCH_2OH

H_2O

CH_3COOH

CH_3OH

$HOCH_2CH_2OH$

CH_3CH_2OH

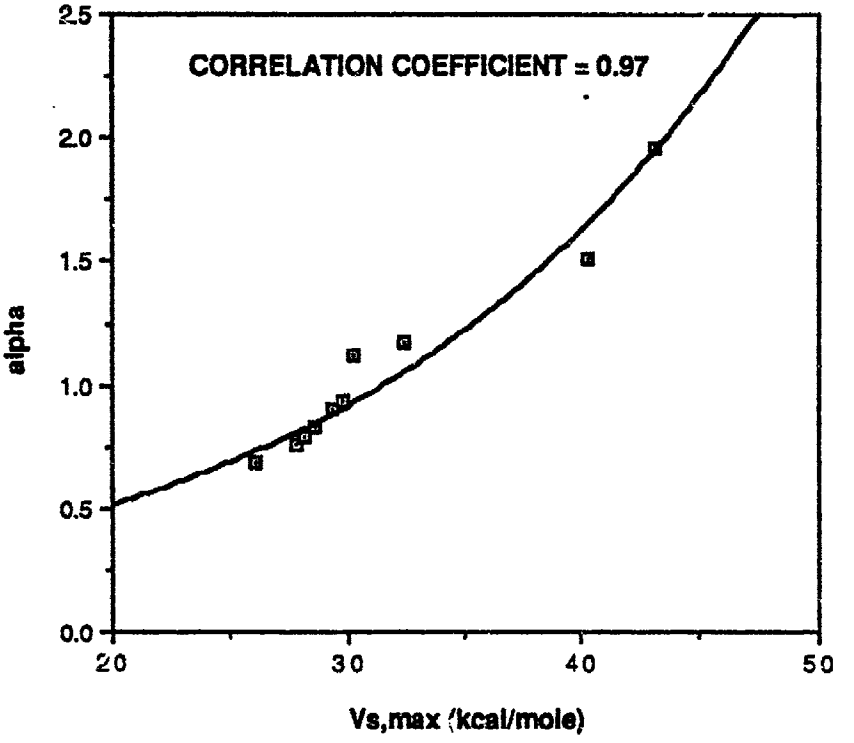
$CH_3CH_2CH_2CH_2OH$

$CH_3CH_2CH_2OH$

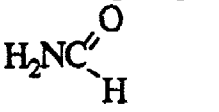
$CH_3CH(OH)CH_3$

$(CH_3)_3COH$

increasing α

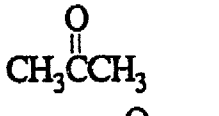


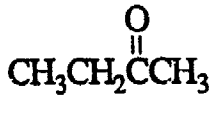
2. Molecules without OH groups:



CH_3NO_2

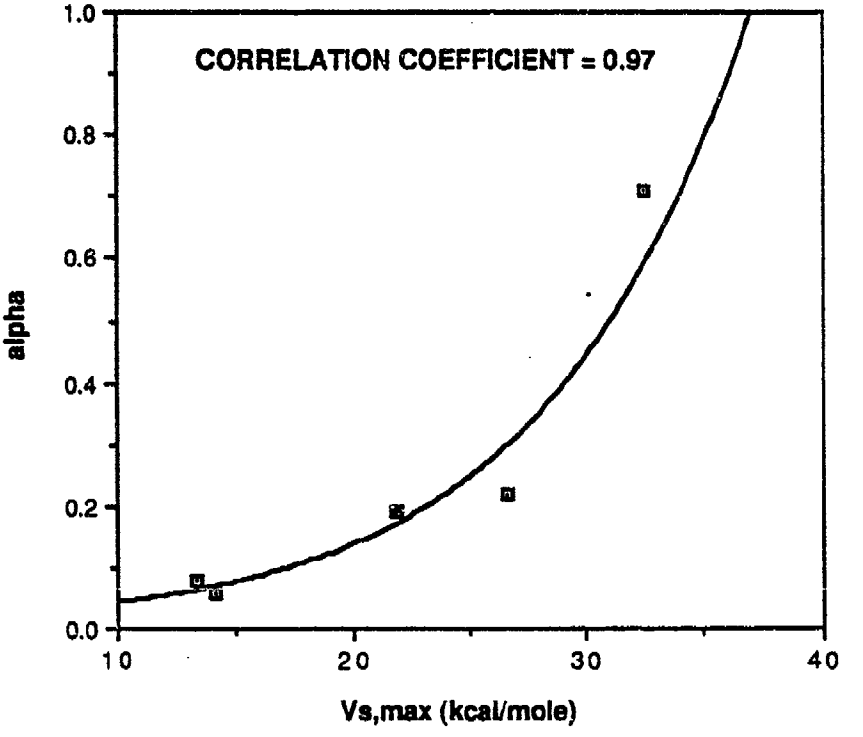
CH_3CN





increasing α

(0.001 contours of electron density used to define surfaces)



AMINE NITROGEN V_{min} OF NH_2-X MOLECULES AS MEASURES OF THE TOTAL ELECTRON-ATTRACTING TENDENCIES OF SUBSTITUENTS

We have found correlations between the amine nitrogen V_{min} of NH_2-X molecules and σ_I or $\sigma_I + \sigma_R$ when $\sigma_R > 0$ (electron-withdrawing), making it possible to estimate either of these quantities.

Examples:

NH_2-X	V_{min} (kcal/mole)	σ_I
X = H	-109	0.00
X=NCH_3F	-91	0.28*
X=NHF	-83	0.43*
X=NF_2	-77	0.53*
X=CN	-68	0.57

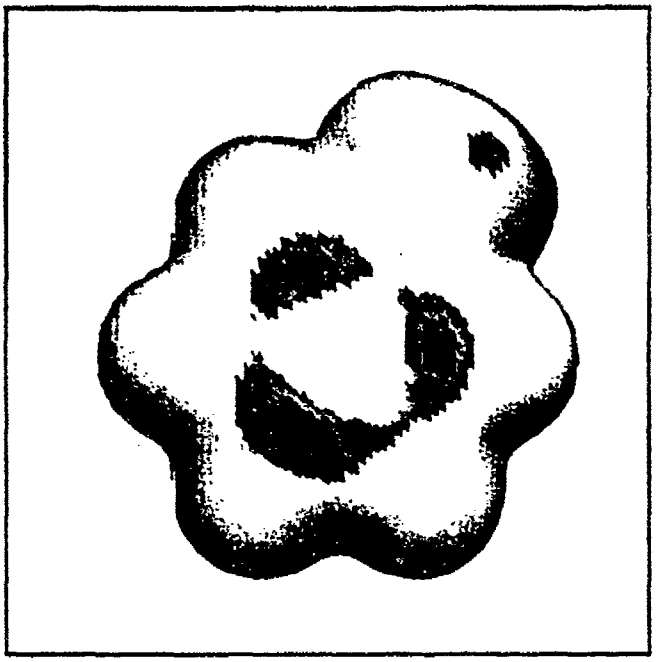
*Estimated from our V_{min} vs. ($\sigma_I + \sigma_R$) when $\sigma_R > 0$ correlation.

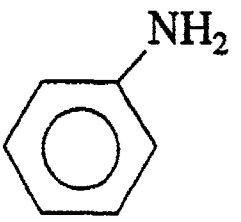
AVERAGE LOCAL IONIZATION ENERGY

$$\bar{I}(\vec{r}) = \sum_i \frac{\rho_i(\vec{r})|\varepsilon_i|}{\rho(\vec{r})}$$

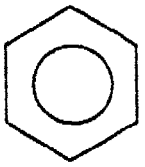
$\rho_i(\vec{r})$ is the electronic density of the i^{th} molecular orbital at the point $\vec{r}$, ε_i is the orbital energy of the i^{th} molecular orbital and $\rho(\vec{r})$ is the total electronic density.

EXAMPLES OF SURFACE Ī(r) PLOTS:

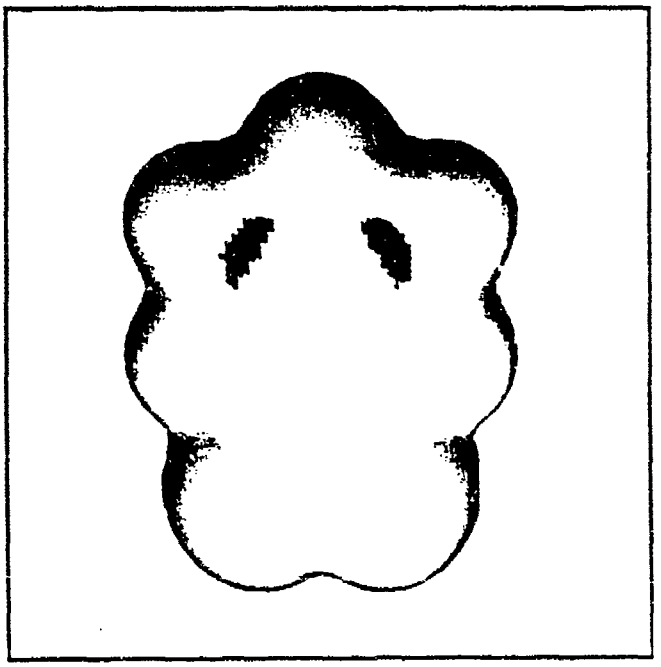


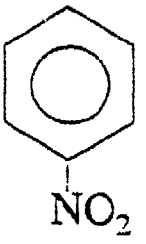


BLACK: less than 11.8 eV.
GRAY: between 11.8 and 12.1 eV.
WHITE: greater than 12.1 eV.
MINIMA: 11.5, 11.7 eV.

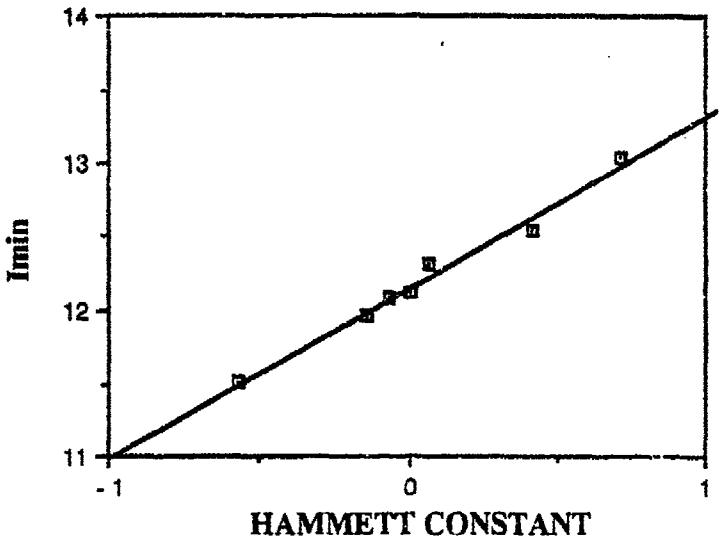


MINIMA: 12.1 eV (ABOVE CARBONS)

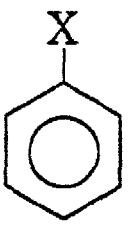




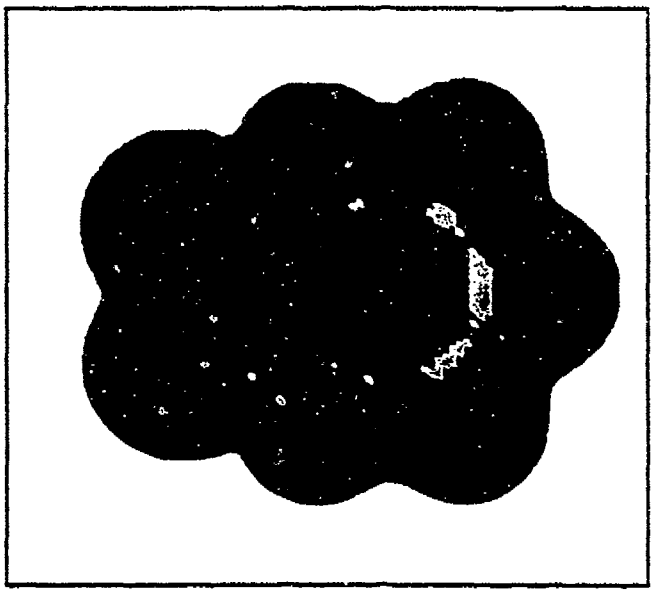
BLACK: less than 13.2 eV.
WHITE: greater than 13.2 eV.
MINIMA: 13.0 eV.

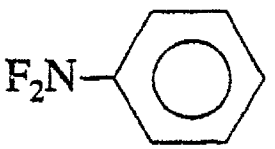


CORRELATION COEFFICIENT = 0.99



X = NH_2, CH_3 H, F, CHO, NO_2





LT.GRAY: Less than 12.8 eV.

DK.GRAY: Greater than 12.8 eV.

MINIMA: 12.7, 12.8 eV.

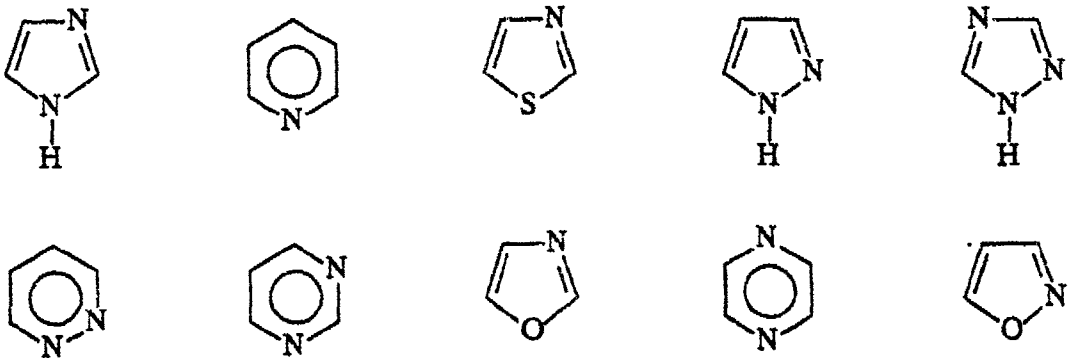
PREDICTIONS

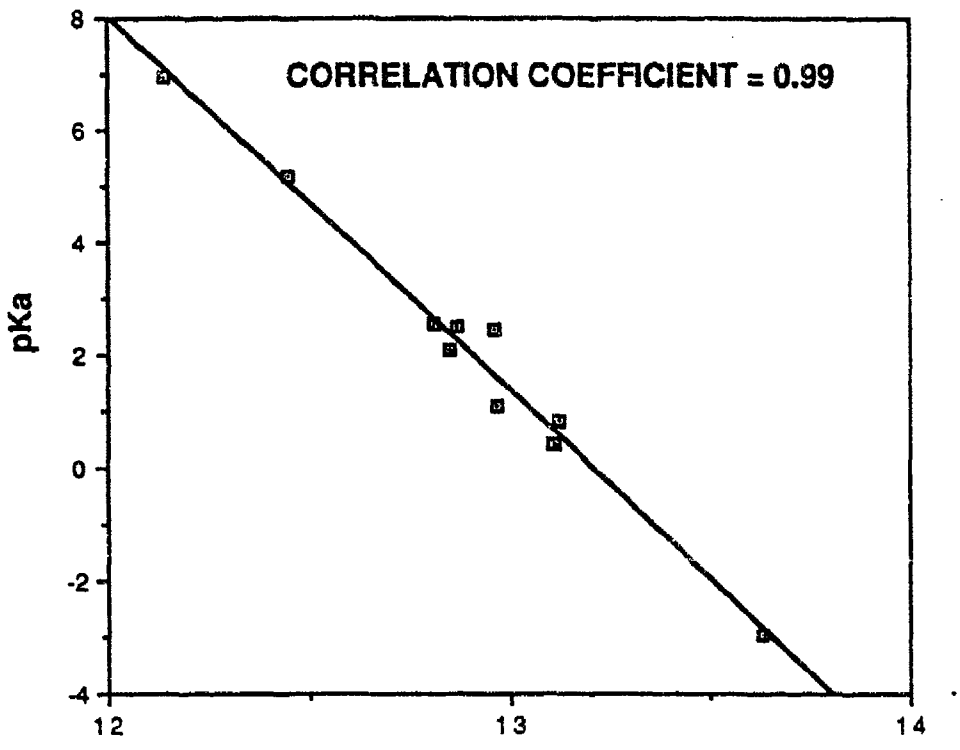
	Hammett Constants		Taft constants, predicted from earlier work and $\sigma_p \approx \sigma_I + \sigma_R$	
	σ_p	σ_m	σ_I	σ_R
NF_2	0.49	0.54	0.53	-0.04
NHF	0.06	-----	0.43	-0.37

CORRELATIONS BETWEEN pK_a AND CONJUGATE BASE SURFACE $\bar{I}(r)$ MINIMA

Our finding that $\bar{I}(r)$ on molecular surfaces provides quantitative predictions of reactivity toward electrophiles led us to investigate whether $\bar{I}(r)$ might be related to acidity, since the latter reflects the tendency to interact with the electrophile H^+.

1. Azines and Azoles:

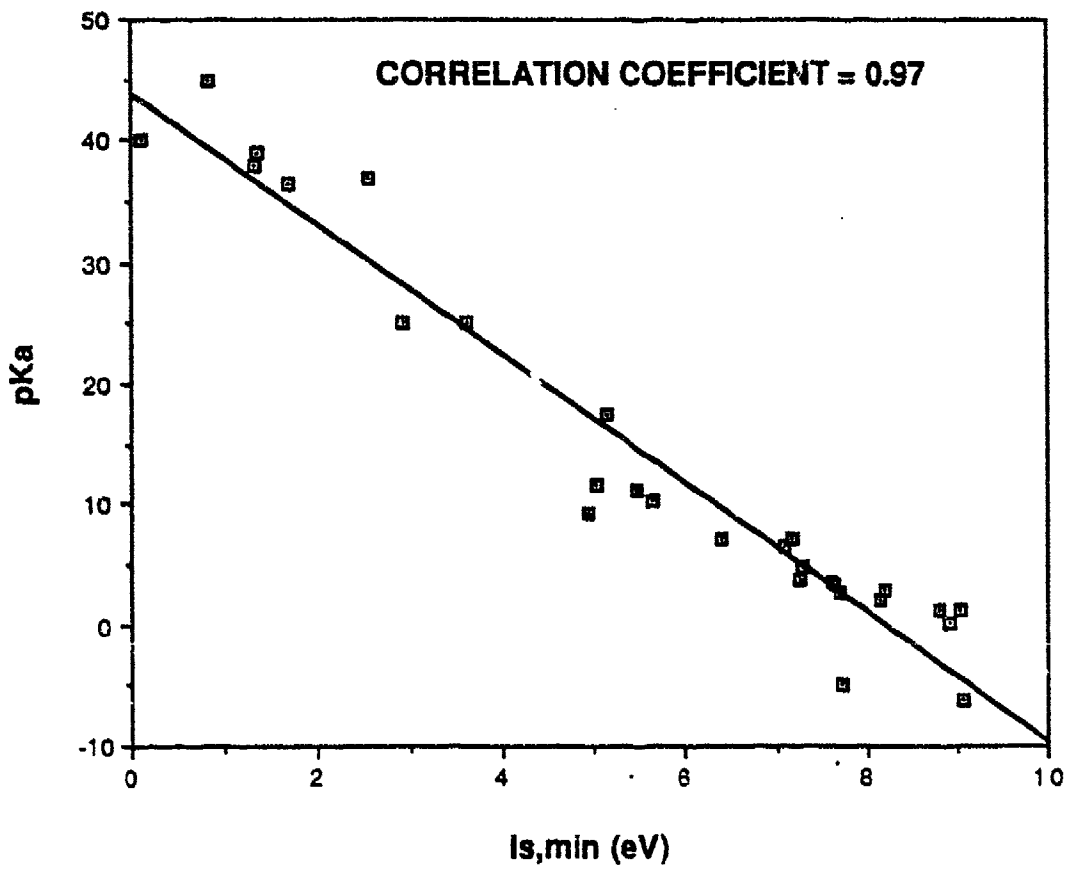




CORRELATIONS BETWEEN pK_a AND CONJUGATE BASE SURFACE $\bar{I}(r)$ MINIMA

2. Carbon, Nitrogen and Oxygen Acids:

	CH_3CN	ClOH	CH_2FCOOH
CH_4	NH_2CONH_2	NH_2NO_2	$HClO_2$
	HOOH	CH_3COOH	$CHCl_2COOH$
NH_3	$CH_2(CN)_2$	HCOOH	HOOCCOOH
	CH_3NO_3	$CH_2(NO_2)_2$	$CH(NO_3)_3$
H_2CCH_2	HCN	HNO_2	$CH(CN)_3$
HCCH	CH_2ClNO_2	$CH_2ClCOOH$	$CH(NO_2)_2CN$

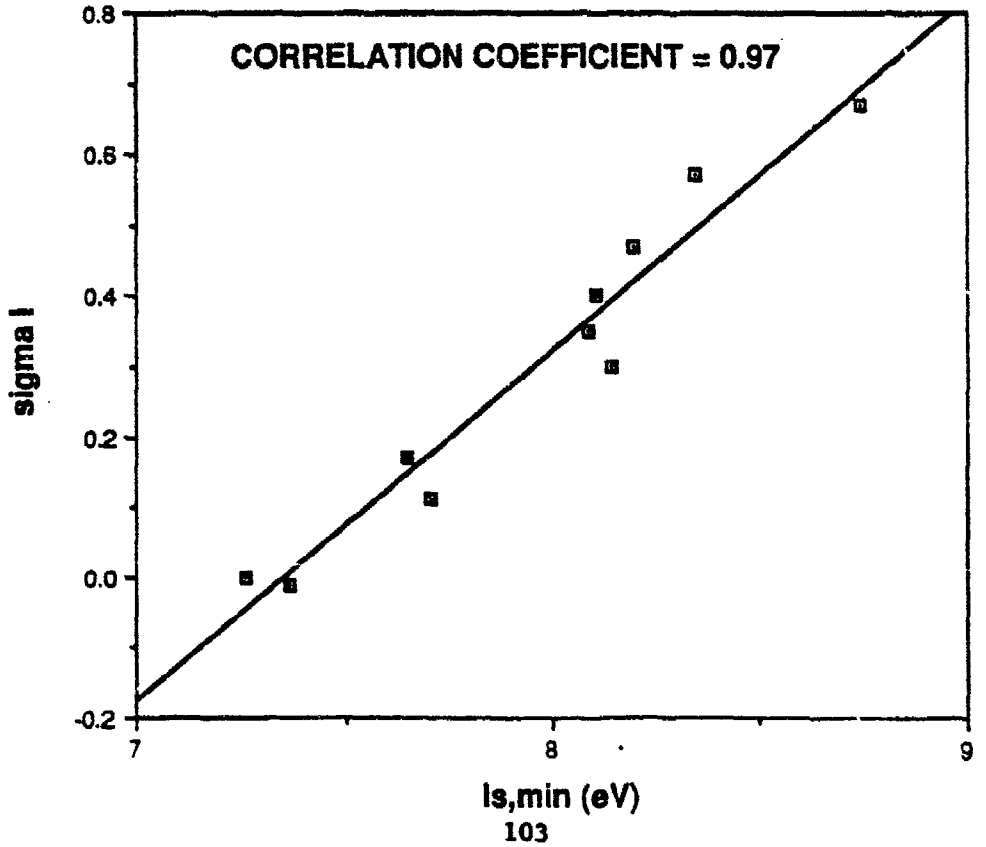


RELATIONSHIP BETWEEN σ_I AND SURFACE $\bar{I}(\mathbf{r})$ MINIMA

The pK_a's of substituted acetic acids have provided a basis for defining the inductive substituent parameters σ_I. As an extension of our $\bar{I}_{S,min}$ vs. pK_a studies, we have investigated the surface $\bar{I}(\mathbf{r})$ of the conjugate bases of a series of substituted acetic acids in relation to the σ_I values of the substituents.

Series:

$X\text{–}CH_2\text{–}COO^-$ where X = H, CH_3, CH_2OH, NH_2, OH, CHO, CF_3, Cl, CN, NO_2



PERSONS INVOLVED IN THIS WORK:

Mr. T. Brinck
Ms. A. H. Buckel
Dr. M. E. Grice
Dr. J. S. Murray
Dr. S. Ranganathan
Dr. P. Sjoberg

FINANCIAL SUPPORT:

We appreciate the support of this work from the Office of Naval Research and the Army Research Office.

A Comprehensive Approach to Structure-Activity Relationship

Corwin Hansch, David Hoekman and Peng Li

Department of Chemistry, Pomona College, Claremont, CA 91711

One of the largest of the Big-Science projects, although it is not usually thought of in this sense, is the study of how organic chemicals interact with and affect living organisms (or parts thereof such as enzymes, organelles, etc.). Many billions of dollars are spent each year by the drug companies of the world in this effort to say nothing of all of the work in pesticide science. The environmental toxicology of both industrial and natural products is also of increasingly intense interest. Thousands of academic scientists are studying the basic biochemistry, molecular biology and physiology of xenobiotics and we have no idea of the effort being expended in the area of chemical warfare. What is needed in all of these areas is better means for predicting the biological activity of untested chemicals from a knowledge of their chemical structure. The dimensions of this problem are enormous, in the first place because there are almost an infinite number of possible organic compounds. Secondly there are an unknown number of species of life: animals, plants, birds, insects, bacteria, viruses, etc. containing a huge number of biochemical systems. As yet there has been no systematic approach to this subject. Scientists from many fields, quantum chemistry, physical organic chemistry, biochemistry, medicinal chemistry, pharmacology, computer technology and applied mathematics are, in their own ways, studying what might be called

selective or comparative toxicology. One of the best ways to organize this fast growing multidisciplinary effort is around the subject of quantitative structure-activity relationships (QSAR).

The QSAR paradigm was first outlined in 1964, and although great progress has been made since then, there is enormous confusion at present because of the multitude of approaches to the subject and the lack of organized data. QSAR are now being published in an astonishing variety of journals, monographs and special publications. Only a handful of the world's libraries would contain all of these sources. Besides collecting such data in computerized form there is an urgent need for some kind of common language to discuss QSAR. We believe that it is now time to attempt this even though we by no means have the knowledge for an enduring system. Even a tentative system and language can do much to aid discussion.

The Med-Chem project at Pomona College has been working at building a database of published and unpublished QSAR for many years. It is only recently that the software has evolved to something approaching a finished product which is user friendly. The task now is to greatly increase the bank of QSAR and clean up and standardize many of the early data entries.

The database has been organized into two major forms: 1) physical organic chemistry and 2) biomedicinal QSAR. Each of these divisions is split into two parts: 1) the raw data (biological activity, chemical structures in terms of SMILES and parameters used to derive the QSAR) and 2) the QSAR equations and their statistics.

These can be searched concurrently or independently in a multitude of ways. The data bank is also tied to a model-building regression

program into which any particular data set can be loaded and studied with new or additional parameters and if an improved QSAR is found it can be returned to the bank to displace the old one.

There are three operational modes in which to study the database: **browse**, **search** and **show**. Using the Browse Menu (Table II) one simply enters a menu number to get a listing of all the items of this class now in the database. Entering 1 lists all of the various types of common names which have been used to characterize system (eg. cat, E. coli, urease, red cell, protoplast, worm, etc.). One uses the **search** mode to find the items of interest and then goes to the **show** mode to display the hits found in the search.

Organization of QSAR database

In the bio database each set is assigned nine key title elements

1. System
2. Class
3. Type of compound
4. Action
5. Reference
6. Source
7. Check
8. Notes
9. Set number
10. Date

System is assigned a common name such as frog nerve, mouse, E. coli chloroplast, etc. to characterize how the chemicals were tested. Similarly a common name is supplied to compounds studied: benzodiazepines, benzoic acids, miscellaneous, etc. Under action the type of end point observed is

noted: LD_{50}, MIC, increase in blood pressure, $1/K_m$ or k_{cat}, etc. Item 5 is simply a reference to the article from which the data were taken.

The QSAR are assigned to a particular class (Table III and IV) so that all of the data for a particular class can be studied as a unit. For example one might wish to study all of the QSAR on hydrolases which could be retrieved by means of the label B2C. There are so many different common names that it becomes important to have the systems grouped by class. For instances under animals (B6A) multiple names such as bovine, cow, steer or mouse, mice, murine might be used. To avoid missing data sets one can retrieve everything under B6A and scan the results using search.

Under source the name of the person entering the data is listed and under notes the definition of unusual variables or procedures can be given. Each data set is assigned a set number.

It is planned to have each data set checked by a second person. Searching under check we now find that 615 data sets have been independently checked for errors. However, this does not mean that the current equation is up to the present methods or is based on this latest parameter values. Most of the QSAR were entered before the advent of the bilinear model and many can now be improved by its use.

For each compound a SMILES[1,2] notation is entered so that one can easily search the database for any structure.

The physical database is organized somewhat differently as shown in Table IV. There is nothing comparable to the biological system so that here system refers to the solvent used, or the gas phase, for example. The individual classifications need pruning and reorganization, but as they stand they do provide a useful means for studying different areas of physical organic chemistry. Thus one can compare oxidation-reduction

QSAR from simple organic reactions in homogeneous solution with biological oxidation reactions by, say, oxidoreductase or microsomes.

Common Nomenclature

The need for a common language to discuss structure-activity relationships is obvious and urgently needed. Many different parameters and mathematical models are now being published so that the field is in a state of bewilderment. While any parameters or regression models can be used in our current system, it is largely based on three major physicochemical properties of substituents or molecules.

1. Electronic $S^{\bullet}$, S, S^{+}, S^{-}, F, R, SI, SR, S^{*}
2. Hydrophobic log P, Pi
3. Steric ES, ESC, MR, MV

In the above S = σ. F, SI and S^{*} are field/inductive parameters while R and SR are resonance constants for substituents. A few QSAR based on molecular orbital indices are beginning to be entered so that terms such as EHOMO, ELUMO, q_2, etc. are entering the database.

Unless otherwise noted P and π (Pi) are from the octanol-water system and refer to the unionized forms of the chemicals. If not, P′ and Pi′ are used to note the difference and these must be defined in the notes.

The Taft steric parameter (Es) and Hancock's variation of it (E_S^C) refer to substituents while molar refractivity (MR) or molar volume (MV) may

also refer to the size of the whole molecular. The sterimol parameter are labeled L, B1, B2, etc. refer only to substituents.

Of course, any parameters can be used such as the number of α-H (AH), hydrogen bonding parameters, etc.

In entering the dependent variable for biological activity, where ever possible, log 1/C is used. C is the molar concentration of chemical producing a standard biological response in a fixed time interval. In some instances responses are given relative to a standard substance so that the term RBR for relative biological response is used. Rate constants (especially for enzymic QSAR) are sometimes available and these must be defined.

There are two modes for studying the data: search and display. Table V shows 19 ways in which the search can be structured. It also notes that in the Bio database as of July 1, 1991 there were 2447 data sets for which 2869 equations had been derived and stored. Sometimes more than one equation is stored for a particular data set and, in fact, any number can be stored. At present there are 3116 data sets in the physical database. We are now cooperating with several laboratories to greatly expand the number of QSAR especially for biological data.

Search Procedure For Bio Database

With the search menu (Table V) one can formulate various kinds of searches. Entering 1 cat will find all of the QSAR on cats. Entering **2 B4B** will yield all equations on bacteria. Entering **3 benzoic acid** will find data sets on benzoic acid. Entering **4 LD_{50}** yields sets with this type of action. Searching on reference (5) one can look for all references by the

name of any author (eg. **5 Jones**) or one can search by year or by specific article, say, **5 J. Med. Chem. Soc. 76, 4525 (1954)**. The numbers 1-5 refer to Table V.

More complex studies can be organized. A. Entering **12 c1ccccc10** (SMILES notation) will uncover all QSAR data sets containing phenol as one of the compounds studied. At present there are 185 such data sets. By entering **15 " S "** one searches this set of 185 for all QSAR on phenols which have a term in the Hammett parameter σ. This turns out to be only 15.

B. Searching the whole database with **15 " S "** yields 298 hits, searching on **15 " S^+ "** finds 44 QSAR having a term in σ^+ and using **" S^- "** yields 69 examples with terms in σ^-. The hit list of 298 could be approached in another way by entering **16 2 < " S " < 3.4** which finds all equations having a coefficient (ρ) with σ between 2 and 3.4. This produces 24 hits.

C. A search can be made for equations of a certain quality by entering, for instance, **18 r > .95** which finds all equations with a correlation coefficient greater than 0.95. At present this yields 1238 hits. However, this does not mean all of these equations are, indeed, of high quality since there are many small sets which have a low ratio of data points/variable. To consider larger sets we can search under the command **18 n > 50** which shows that there are now 171 data sets having 50 or more data points each.

D. To study steric effects in the reactions of hydrolases we can enter the commands: **2 B2C** and **15 " Es "** which uncovers 37 such QSAR.

E. To ascertain the importance of the hydrophobic character of insecticides we first isolate all insect QSAR by searching with **B6B** and find

80 such equations. The " or " search **15** " **Pi** " " **P** " uncovers the fact that all equations contain a term in log P or π. As of the present, the entry of hydrophobic terms has not been as tightly standardized as we plan to make it. Using the browse mode and entering 10 we can inspect all of the various labels used for the hydrophobic properties of chemicals and with these we can discover just how important the hydrophobic terms are for the whole database. Entering **15** " **Pi**" " **P**" " **P'** " " **log P** " we obtain 2447 hits on the 2854 equations or 86% of the QSAR contain a hydrophobic term. This estimate is actually a bit low since there are a number of other ways in which hydrophobicity has been termed. We estimate that at least 90% of the QSAR contain a hydrophobic term. A similar search for QSAR with electronic terms uncovers only 651 examples.

Another illustration of a complex search is to first isolate all QSAR on animals (**2 B6**) to obtain 286 sets. Then using the " and " search **15** " **P** ", " **S** ", " **Es** " we find a total of eleven equations for animals containing all three terms.

Displaying the information obtained in a search is done by entering: **show**. The show menu (Table VI) is essentially the same as that for search (Table V).

To obtain all information on each data set the entry **/and 1,20** is made. To display only the references for each QSAR one enters **/and 5**.

There is also an important sort option. For instance, under B above we found 69 equations containing a term in σ^-. These equations could be sorted and printed according to the ρ associated with each σ^- term: **/and /sort-16 1,5 15,18**. The program then asks for the item to sort on and we enter " **S⁻** ". It is then found that the lowest value of ρ is -2.42 with all QSAR ordered according to increasing size of ρ.

Any data set uncovered by the above procedures can by means of its set # be loaded into the regression program for further study.

Motivation For Building a QSAR Data Bank

There are obvious reasons of convenience for constructing the above described system. But the most important reason, which we have not mentioned, is that of establishing the validity of a QSAR. This can be illustrated by the following example.

Mutagenicity of $X\text{-}C_6H_4\text{-}CH_2\overset{CH_3}{\overset{|}{N}}\text{-}N\text{=}O$ in the Ames test

$$\log 1/C = 3.55\,\sigma - 3.88\,\sigma^2 + 1.62\,{}^{3}\chi_p^{v} - 5.11 \qquad (1$$

$n = 13,\ r = 0.873$

$$\log 1/C = 0.92\,\pi + 2.08\,\sigma - 3.26 \qquad (2$$

$n = 12,\ r = 0.891$

In these QSAR C is the concentration of nitrosoamine producing a standard set of mutations and ${}^{3}\chi_p^{v}$ is a complex connectivity parameter developed by Kier and Hall[3] and used in this instance by Singer[4] to form QSAR 1. We have developed QSAR 2 from the same data. Which equation provides the most useful information and in which should we place our confidence? Statistics are of no help since the values of r are close to each other. One data point was dropped to obtain eq. 2, but eq 1 uses an

additional parameter. We believe that eq. 2 is the one to select because it contains terms which tie it into the basic approach from physical organic chemistry and in particular because it contains a hydrophobic term. As discussed below this is a hallmark of QSAR for the Ames test.

"Proof" of a concept in science generally arises as the number of facts which can be rationalized by it in a self consistent way increases. There is no fundamental set of laws or even a single such law which can be used as a starting point to relate chemical structure with the biological activity of a set of complex compounds acting in a mouse or even with an enzyme. Except for very small molecules, quantum chemistry cannot precisely define the electron distribution in a compound or how it might interact with a reagent or a receptor. As shown above the operationally defined hydrophobic parameters are of great importance, but we do not know what all is contained in log P or π.

Steric effects in QSAR are the most difficult of all to deal with. Statistical tests such as correlation coefficients or F tests are no firm guide to the value of parameters or mathematical models in a general sense. We believe that only via lateral correlation of a given QSAR with many others can one develop confidence in the validity of any approach to the correlation of structure with biological activity. In so far as possible the biological QSAR must be related to the far better understood reactions of simple organic compounds in homogeneous solution or to enzymic or receptor QSAR. To do this requires a language based on the concepts and nomenclature of physical organic chemistry and a very large database of QSAR with which comparisons can be made. The following are a very few examples which illustrate the utility of the present QSAR database and the current nomenclature. Obviously, this will evolve with time.

Lateral Correlation in Terms of Hydrophobic Parameters

There are simple model systems which can be used to compare with the effect of hydrophobic compounds on the more complex biological systems.[5]

ROH penetration into phosphatidyl choline monolayers

$$\log 1/C = 0.87 \log P + 0.62 \quad (3$$

$n = 4, r = 0.994, s = 0.082$

ROH disagregation of silanized glass beads

$$\log 1/C = 0.92 \log P - 0.71 \quad (4$$

$n = 4, r = 0.999, s = 0.035$

Change in electrical resistance of black lipid membrane caused by ROH

$$\log 1/C = 1.08 \log P - 0.40 \quad (5$$

$n = 7, r = 0.983, s = 0.280$

ROH inhibition of the aggregation of bovine blood platelets

$$\log 1/C = 1.00 \log P + 0.18 \quad (6$$

$n = 4, r = 0.997, s = 0.064$

The above QSAR can be compared to examples found in the present data bank by looking at QSAR on nerves.

ROH causing 10 mV decrease in lobster axon

$$\log 1/C = 0.87 \log P - 0.24 \quad (7$$

$n = 5, r = 0.993, s = 0.100$

Salicylate analogs causing a 20 mV increase in mollusc buccal ganglion

$$\log 1/C = 0.84 \log P + 3.31 \quad (8$$

$n = 30, r = 0.979, s = 0.177$

Inhibition of contraction of frog sartorius muscle by miscellaneous compounds

$$\log 1/C = 0.88 \log P + 0.63 \quad (9$$

$n = 25, r = 0.955, s = 0.297$

The dependence of nerve (membrane?) perturbation on log P parallels that seen in the artificial model systems. While there are relatively few data points supporting each model QSAR so that no single equation would give one the confidence to place much weight on the comparisons, when all are taken together the commonality is striking and deserves further study.

Studies on mutagenesis by various organic compounds in the Ames test yields the following QSAR.

Mutagenicity of X-C_6H_4N=$NN(CH_3)$ in TA92

$$\log 1/C = 1.04 \log P - 1.63 \sigma^+ + 3.06 \quad (10$$

$n = 17, r = 0.974, s = 0.315$

Mutagenicity of X-quinolines in TA100

$$\log TA100 = 1.08 \log P - 56.3 q_2 - 5.91 \quad (11$$

$n = 21, r = 0.827, s = 0.645$

Mutagenicity of arylamines in TA98

$$\log TA98 = 1.08 \log P + 1.28 \varepsilon_{LUMO} + 1.46 I_L + 7.20 \quad (12$$

$n = 88, r = 0.898, s = 0.860$

Mutagenicity of arylamines in TA100

$$\log TA100 = 1.13 \log P + 1.66 \varepsilon_{HOMO} - 0.63 \varepsilon_{LUMO} + 1.09 \quad (13$$

$n = 67, r = 0.881, s = 0.714$

Mutagenicity of aryl nitro compounds in TA100

$$\log TA100 = 1.10 \log P - 3.37 \log (\beta \bullet 10^{\log P} + 1) + 1.84 \varepsilon_{LUMO}$$
$$- 3.35 I_a + 5.64 \quad (14$$

$n = 117, r = 0.858, s = 0.920$

Mutagenicity of aryl nitro compounds in TA98

$$\log TA98 = 0.65 \log P - 2.90 \log (\beta \bullet 10^{\log P} + 1) - 1.38\ \varepsilon_{LUMO} + 1.88\ I_L - 2.89\ I_a - 4.15 \quad (15$$

$n = 188, r = 0.900, s = 0.886$

In eq. 10 C is the molar concentration producing 30 mutations above background in 10^8 bacteria. In the other equations log TA refers to revertants/nmol. In eq. 11, q_2 refers to the change on position 2 of the quinoline ring calculated using the AM1 procedure. The variable I_L in eq. 12 and 15 is assigned the value of 1 for all compounds containing more than 2 fused rings (eg. anthracene, carbazol, acridines, etc.) and 0 for those containing less (eg. benzene, naphthalene, quinoline, etc.). It is of interest that this term occurs with the TA98 organisms in eq. 12 and 13 but not with the TA100 organisms. Thus one has some evidence for its validity beyond statistics. The indicator variable I_a refers to 5 complex hydrocarbons which have much lower than expected activity. Except for eq. 10, the other QSAR for mutagenesis are not sharp correlations, however, it is found that the log P terms in all of the equations are the most important variables and moreover except in one example the rising slope has a coefficient near 1. Thus eqs. 10 - 15 with lateral correlation to eq. 2 help one to choose between eq. 1 and 2. Not only does eq. 2 have a term in log P but its coefficient is near that seen for the other QSAR. It is only by building up matrices of such QSAR that validity for any particular approach can be established.

Another set of QSAR which again shows the significance of the octanol-water log P parameter comes from the studies of the induction of P450.

50% Induction of chick embryo P450 by ROH

$$\log 1/C = 0.78 \log P + 1.46 \quad (16$$

$n = 7, r = 0.988, s = 0.095$

50% Induction of chick embryo P450 by pyrazoles

$$\log 1/C = 0.85 \log P + 1.93 \quad (17$$

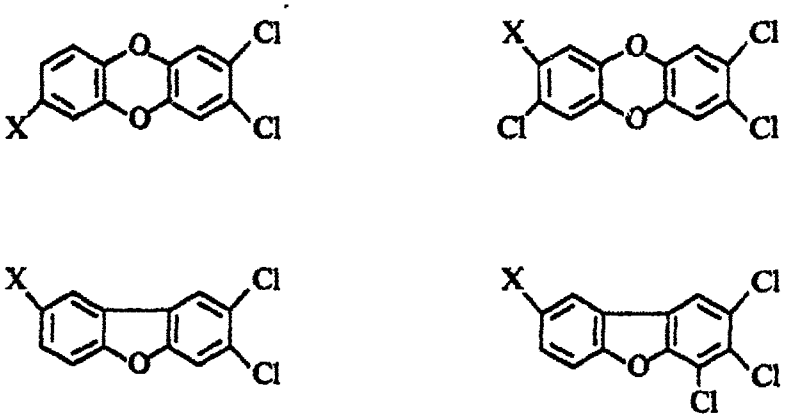
$n = 8, r = 0.970, s = 0.305$

50% Induction of chick embryo P450 by barbiturates

$$\log 1/C = 1.02 \log P + 2.75 \quad (18$$

$n = 9, r = 0.984, s = 0.186$

Induction of rat P450 by benzofurans and dioxins



$$\log 1/C = 1.14 \log P + 0.64 B_5 - 1.24 E_s + 0.58 \quad (19$$

$n = 43, r = 0.908, s = 0.670$

Equations 16 - 19 show the great importance of hydrophobicity in the induction of P450; this coupled with the mutagenicity of lipophilic compounds increases concern about lipophilic compounds in the environment. In the case of the arylamines, triazenes and quinolines P450 is necessary for activity. This is also true of aromatic hydrocarbons. Carcinogenicity has also been shown to be highly dependent on log P as seen in eq. 20. The lingua franca of the octanol-water log P enables us to see far reaching consequences of lipophilic compounds in environmental toxicology.

$$\log A = 0.55 \log P - 1.15 \log (\beta \bullet 10^{\log P} + 1) + 0.38 LK + 0.47 \varepsilon_{HOMO} + 1.98 \quad (20$$

$n = 161, r = 0.844, s = 0.347$

In this expression, A is skin carcinogenic potency as defined by the Iball index and LK is an indicator variable which takes the value of 1 for instances where a substituent is placed on the L or K region of the aromatic hydrocarbon or heterocycle. These compounds are about twice as potent as expected, presumably because substitution at these points prevents nonproductive side reactions (oxidation).

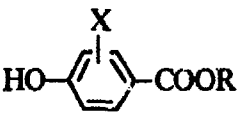
Analysis of a set of literature data aroused our interest in the toxic action of benzimidazoles.

I_{75} Influenza B virus by benzimidazoles

$$\log 1/C = 0.57 \log P + 1.58 \qquad (21$$

$n = 13, r = 0.931, s = 0.166$

The question that arises from the QSAR is, is this a specific toxic effect or nonspecific activity as the low value of the coefficient suggests? Entering the search **16 .5 < " P " < .7, 1.3 < const < 1.7** yields QSAR for comparison having similar slopes and intercepts.

Inhibition of B. subtilis by 

$$\log 1/C = 0.55 \log P + 1.61 \qquad (22$$

$n = 26, r = 0.920, s = 0.220$

Inhibition of B. subtilis by $X\text{-}C_6H_4OCC(CH_3)=CH_2$ (with O above C)

$\log 1/C = 0.62 \log P + 1.53$ (23

$n = 10, r = 0.976, s = 0.214$

Inhibition of S. typhosa by X-2-naphthols

$\log 1/C = 0.63 \log P + 1.32$ (24

$n = 22, r = 0.898$

Inhibition of hydroxyindole O-methytransferase by N-acyltryptamines

$\log 1/C = 0.60 \log P + 1.49$ (25

$n = 21, r = 0.948, s = 0.170$

Inhibition of rat mitochondral respiration by aromatic amines and phenols

$\log 1/C = 0.60 \log P + 1.52$ (26

$n = 19, r = 0.925, s = 0.289$

Division of sea urchin egg by $ROCONH_2$

$\log 1/C = 0.65 \log P + 1.33$ (27

$n = 10, r = 0.975, s = 0.147$

Equations 22 - 27 show that this is a typical "nonspecific" kind of activity shown by a variety of chemicals in a variety of systems.

Lateral Correlations in Terms of Electronic Effects

The following are a few examples which show how electronic properties of bioactive compounds can be related back to a simple organic reaction.

Inhibition (50%) of E. coli by $X\text{-}C_6H_4N{=}C{=}S$

$$\log 1/C = 2.27\,\sigma + 4.31 \qquad (28$$

$n = 9,\ r = 0.963,\ s = 0.161$

Equation 28 is unusual in that no hydrophobic effect is found. The aromatic isothiocyantes are very reactive compounds and may inhibit growth by reaction with nucleophiles near the membrane surface.

QSAR 28 can be compared with 29 and 30 for reaction with two types of nucleophiles.

$$X\text{-}C_6H_4N{=}C{=}S + CH_3CH_2OH \longrightarrow X\text{-}C_6H_4NHC(=S)OC_2H_5$$

$$\log k = 2.16\,\sigma - 4.80 \qquad (29$$

$n = 8,\ r = 0.975,\ s = 0.181$

$$X\text{-}C_6H_4N\text{-}C\text{-}S + C_6H_5NH_2 \longrightarrow X\text{-}C_6H_4NH\overset{\overset{S}{\|}}{C}NHC_6H_5$$

$$\log k = 2.14\,\sigma - 3.13 \qquad (30$$

$$n = 4,\ r = 0.994,\ s = 0.060$$

The value of ρ for the three QSAR are in very good agreement suggesting a mechanism for the toxic action of the isothiocyanates.

LD_{50} of house flies by $X\text{-}C_6H_4O\overset{\overset{O}{\|}}{P}(OC_2H_5)_2$

$$\log 1/C = 2.44\,\sigma^- + 0.26\,\pi - 0.61 \qquad (31$$

$$n = 8,\ r = 0.988,\ s = 0.219$$

Inhibition of fly head acetylcholinesterase by $X\text{-}C_6H_4O\overset{\overset{O}{\|}}{P}(OC_2H_5)_2$

$$\log 1/C = 2.45\,\sigma^- - 0.56\,E_s3 + 4.82 \qquad (32$$

$$n = 13,\ r = 0.962,\ s = 0.408$$

In the case of eq. 31 only one meta substituted congener was examined. Several were studied for eq. 32 and it was found that for enzyme inhibition a small steric effect is present for meta substituents which is accounted for by E_s3. Again we see good agreement between the two values of ρ, and σ^- is Hammett constant of choice as one might expect. A small hydrophobic effect is seen in the whole fly.

Sulfonamides inhibiting carbonic anhydrase yield QSAR 33:

$$\log 1/ki = 0.95\,\sigma + 0.54\,\pi - 0.35\,B_{5,3} + 6.29 \qquad (33$$

$n = 31,\ r = .968,\ s = .168$

The sterimol parameter B_5 applies only to substituents in the meta (3) position. It is thought that the ionized form of the sulfonamide binds to a positively charged zinc in the enzyme.

QSAR 34 shows the effect of substituents on the ionization of the sulfonamides.

$$\Delta \log K = 0.86\,\sigma + 0.08 \qquad (34$$

$n = 16,\ r = 0.962,\ s = 0.146$

After corrections for the hydrophobic and steric effects of substituents in QSAR 33 the ρ value is in good agreement with that of eq. 34 indicating that inhibitory effects follow propensity for ionization.

The data in Table I represents a different approach to enzymic QSAR. Instead of testing many compounds on a single enzyme two substrates have been tested with a variety of enzymes. The value of ρ comes from equations of the type: $\log 1/K_m = \rho\sigma + h\pi + aMR + b$ where K_m is the Michaelis constant. Since the structure of the active site of each enzyme is different, terms in π or MR or both may or may not appear in the final QSAR, nevertheless, we see that there is little variation in ρ. The mechanism of action by the various enzymes in forming the ES complex is much the same for the two different substrates in terms of the electronic effects of X.

One example from the physical-organic database will now be considered. Over the years there has been discussion about the value of ρ in the Hammett equation. That is, can it be expected that for similar reactions under similar conditions similar values of ρ are found? No systematic study has been made of the problem and few people have much interest beyond their own narrow compartment. However, in building a generalized approach to QSAR one wants to wring as much information as possible from each QSAR.

Searching the physical data base under the command **2 P2** we obtain 307 hits for QSAR on hydrolysis. Then searching on **2 < " S " < 2.6** we obtain hits of which eq. 35 to 43 are relevant.

Alkaline aqeous hydrolysis $X\text{-}C_6H_4O\overset{O}{\overset{\|}{C}}\text{-}NHCH_3$ at 37°

$\log k = 2.48\,\sigma + 3.03$ (35

$n = 6,\ r = 0.977,\ s = 0.156$

Alkaline hydrolysis in 70% dioxane $X\text{-}C_6H_4COOC_2H_5$ at 25°

$\log k = 2.26\,\sigma - 2.47$ (36

$n = 4,\ r = 0.989,\ s = .068$

Alkaline hydrolysis $X\text{-}C_6H_4COOC_2H_5$ in 87% ethanol at 30°

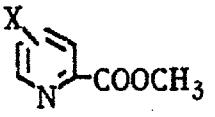
$\log k = 2.51\,\sigma - 1.28$ (37

$n = 18,\ r = 0.993,\ s = 0.105$

Alkaline hydrolysis of $X\text{-}C_6H_4\text{-}CN$ in 60% ethanol at 82°

$$\log k = 2.13\,\sigma - 1.0 \qquad (38$$

$n = 5,\ r = 0.981,\ s = 0.184$

Alkaline hydrolysis  in 85% methanol at 25°

$$\log k = 2.03\,\sigma - 2.09 \qquad (39$$

$n = 10,\ r = 0.997,\ s = 0.083$

Alkaline hydrolysis of $X\text{-}C_6H_4COOCH_3$ in 85% methanol at 25°

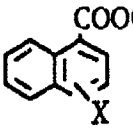
$$\log k = 2.25\,\sigma - 3.72 \qquad (40$$

$n = 10,\ r = 0.996,\ s = 0.076$

Aqueous hydrolysis 25° of $X\text{-}C_6H_4OCOCH_2CH_2COO^-$

$$\log k = 2.49\,\sigma - 2.55 \qquad (41$$

$n = 6,\ r = 0.969,\ s = 0.214$

Alkaline hydrolysis  in 85% ethanol at 50°

$$\log k = 2.13\,\sigma - 2.58 \qquad (42$$

$n = 9,\ r = 0.998,\ s = 0.048$

Alkaline hydrolysis X-$C_6H_4SO_2OC_6H_5$ in 30% dioxane at 50°

$\log k = 2.30\,\sigma - 3.47$ (43

$n = 6,\ r = 0.998,\ s = 0.077$

There is considerable variation in the basic structures and reaction conditions from which eq. 35 - 43 have been derived, yet the electronic effects of substituents show a strong central tendency toward a ρ of 2.2. This rough agreement can be helpful in bringing more general order to SAR research. It is of interest that both carbamates (eq. 35) and nitriles (eq. 38) behave in a similar fashion in terms of ρ. The sulfonates (eq. 43) have a ρ close to that of the carboxylates.

These results do not mean that all hydrolysis reactions of esters of amides will always have a ρ of about 2.2. For example, making the same kind of search in the range of 1.5 - 2.1 for ρ yields the following relevant hits.

Hydrolysis of X-C_6H_4COCl in 95% acetone at 25°

$\log k = 1.78\,\sigma - 4.20$ (44

$n = 5,\ r = 0.992,\ s = 0.112$

Neutral Hydrolysis of X-C_6H_4OCOH in 10% acetonitrile at 25°

$\log k = 1.59\,\sigma + 0.49$ (45

$n = 8,\ r = 0.988,\ s = 0.104$

Hydrolysis of $(X\text{-}C_6H_4CO)_2O$ in 75% dioxane at 25°

$\log k = 1.60\,\sigma - 5.18$ (46

$n = 9,\ r = 0.986,\ s = 0.220$

Alkaline hydrolysis of $X\text{-}C_6H_4COOC_6H_5$ in 60% acetone at 25°

$\log k = 1.92\,\sigma - 1.40$ (47

$n = 11,\ r = 0.983,\ s = 0.100$

Alkaline hydrolysis of $X\text{-}C_6H_4C{\equiv}CCOOC_2H_5$ in 88% ethanol at 20°

$\log k = 1.91\,\sigma - 0.15$ (48

$n = 4,\ r = 0.957,\ s = 0.231$

Hydrolysis of $X\text{-}C_6H_4OCOCl$ in 65% acetone at 25°

$\log k = 1.54\,\sigma - 2.71$ (49

$n = 11,\ r = 0.988,\ s = 0.107$

Alkaline hydrolysis of $X\text{-}C_6H_4OCOCH_3$ in 56% acetone at 1°

$\log k = 1.87\,\sigma - 1.06$ (50

$n = 15,\ r = 0.974,\ s = 0.124$

Equations 47, 48 and 50 are just slightly outside of our first sweep. The anhydrides and acid chlorides of eq. 44, 46 and 49 are so reactive that one would not expect substituent effects to be as important as with normal esters and amides. It is interesting to see the hydrolysis of formate esters (eq. 45) under neutral conditions has lower ρ value.

Even though there is not enough uniform data in hand to evaluate just how similar ρ would be for similar reactions this assumption is clearly of some value just as the coefficients with log P are.

Work on the two data bases was started about 1970 and largely discontinued about 1977. In the last two years we have resumed the addition of more data, especially to the Bio database. We now need to check all of the data, install a more standardized nomenclature and redo many of the QSAR. Only a few of the older QSAR have been developed using the bilinear model, also improved values of log P are now available. In a number of instances data sets which are too small in terms of data points/variable need to be deleted. Even with these shortcomings the present bank is a useful tool to those heavily involved in structure-activity analysis. However, if we can enter essentially all of the published QSAR and keep up with those currently being published we shall have an indispensable tool for all areas of structure activity analysis. This would entail building a bank about two to three times that of our current two banks which have in all about 5500 data sets.

Acknowledgements: We are grateful to R.J. Reynolds, Burroughs Welcome, Abbott Laboratories and du Pont for the support of this project. The work was first initiated by an NIH grant about 1970.

References

1. D. Weininger and J.L. Weininger in Comprehensive Medicinal Chemistry Vol. 4, C.A. Ramsden Ed., Pergamon Press, 1990, p. 59
2. D. Weininger, J. Chem. Inf. Comput. Sci. 28, 31 (1988).
3. L.B. Kier and L.H. Hall, Molecular connectivity in structure-activity analysis, Research Studies Press, Letchworth, England, 1986.
4. G.M. Singer, A.W. Andrews and S.-M. Guo, J. Med. Chem. 29, 40 (1986).
5. C. Hansch, D. Kim, A.J. Leo, E. Novellino, C. Silipo and A. Vittoria, CRC Critical Rev. Toxicol. 19, 185 (1989).

Table I

Values for ρ in the correlation of hydrolysis of
$X\text{-}C_6H_4OCOCH_2NHCOC_6H_5$ (I) and $X\text{-}C_6H_4OCOCH_2NHSO_2CH_3$ (II)

Enzyme	Substrate	ρ	pH	Class of Hydrolase
Papain	I	0.57	6	Cysteine
Papain	II	0.55	6	Cysteine
Ficin	I	0.57	6	Cysteine
Ficin	II	0.62	6	Cysteine
Actinidin	I	0.74	6	Cysteine
Bromelain B	I	0.70	6	Cysteine
Bromelain B	II	0.68	6	Cysteine
Bromelain D	I	0.63	6	Cysteine
Subtilisin	I	0.49	7	Serine
Chymotrypsin	I	0.42	6.9	Serine
Trypsin	I	0.71	7	Serine

Mean = 0.61

Table II

Browse Menu

1. System
2. Class
3. Compound
4. Action
5. Reference
6. Souce
7. Check
8. Note
10. Parameters
11. Substituents
12. Smiles

Table III

Biological Classifications

B0	Uknown
B1	Nonenzymatic Macromolecules DNA, Fibrin, Hemoglobin, Albumin
B2	Enzymes
B2A	Oxidoreductases
B2B	Transferases
B2C	Hydrolases
B2D	Lyases
B2E	Isomerases
B2F	Ligase, Thiokinase
B2G	Receptors
B3	Organelles
B3A	Mitochondria
B3B	Microsome
B3C	Chloroplasts
B3M	Membranes
B3S	Synaptosomes
B4	Single-Celled Organisms
B4A	Algae
B4B	Bacteria
B4C	Cells in culture
B4E	Erythrocytes
B4F	Fungi, Molds
B4P	Protozoa
B4V	Viruses
B4Y	Yeasts
B5	Organs/Tissues
B5G	Gastro-intestinal tract
B5H	Heart
B5I	Internal/soft organs
B5N	Nerves, Brain, Muscles
B5S	Skin
B6	Multi-cellular organisms
B6A	Animal (vertebrates)
B6B	Insect (bugs)
B6C	Cancer
B6H	Human
B6I	Invertebrates
B6P	Plant

Table IV

Physical Classifications

P0	Unknown	P7	Addition
		P7D	Dimerization
P1	Ionization	P7E	Electrophilic addition
P1P	Ionization potential	P7N	Nucleophilic addition
P1X	Proton exchange	P7P	Polymerization
P2	Hydrolysis	P8	Elimination
		P9	Rearrangement
P3	Solvolysis	P10	Oxidation
		P12	Radical reactions
P4	Spectra	P13	Complex formation
P4I	Ionization spectra		
P4E	ESR spectra	P14	Partitioning
P4M	Mass spectra	P14C	Chromatography
P4N	NMR spectra		
P4R	IR spectra	P15	Pyrolysis
P4U	UV spectra	P16	H-bonding
		P17	Electrochemical
P5R	Miscellaneous reactions	P18	Bronsted
		P19	Esterification
P6	Substitution	P20	Photochemical
P6E	Electrophilic substit.	P21	Hydrogenation
P6N	Nucleophilic substit.	P22	Isokinetic

Table V

Search Menu

1. System
2. Class
3. Compound
4. Action
5. Reference
6. Source
7. Check
8. Note
9. Date
10. Parameters
11. Substituents
12. SMILES
14. Prm max/min
15. Terms in eqn.
16. Coefs. in eqn.
17. Ideal/log B
18. Statistics
19. Residuals
20. Predicted

Table VI

Show Menu

0. ID numbers only	11. Substituents
1. System	12. SMILES
2. Class	13. Parameter value
3. Compound	14. Prm max/min
4. Action	15. Terms in eqn.
5. Reference	16. Coefs. in eqn.
6. Source	17. Ideal/log B
7. Check	18. Statistics
8. Note	19. Residuals
9. Date	20. Predicted
10. Parameters	

BLANK

Recent Progress and Problems in Calculating Log P_{oct}

by Albert J. Leo

Pomona College Medicinal Chemistry Project

The early algorithms used to calculate log P_{oct} from structural 'fragments' yielded values for very lipophilic solutes (>6.0) which were often much higher than those measured. For example, DDT was calculated as 6.91, but in 1973 Metcalf and Lu, using a radiolabeled solute, measured it at 3.98. Seeing no reason for the calculations to be so greatly in error, we resisted the temptation to insert in the algorithm some 'tailing-off factor' if more than a certain number of hydrophobic fragments were added together. Waiting for the analytical methods to improve paid off, for, as Fig. 1 shows, the most reliable technique (slow-stir) matches that from the CLOGP algorithm for DDT (6.91) and is very close for DDE (6.96 vs. 6.94).

Tetrahydrocannabinol (Δ-9-THC) is another solute that seemed to require an algorithm change to bring calculated value into agreement with measurment. Early measurements put its Log P_{oct} at slightly below 4.0, while it is calculated by CLOGP as 7.24. Martin and co-workers found they could reliably measure the more hydrophilic analogs of THC by normal shake-flask technique, but they had great difficulty with those as hydrophobic as the parent. Using reversed phase HPLC with a C-8 column, they used known standards to establish a relationship between K' and log P_{oct}. As seen in Fig 2, they found a good relationship between CLOGP and Log P_{HPLC} values with hydrophobic analogs up to log 11.5. It appears, then, that as a model for this type of hydrophobic behavior (partial de-solvation as on a hydrocarbon-coated HPLC support), CLOGP values can be used to an upper limit of at least 12.0.

The more recent versions of CLOGP contain an ortho correction (-0.12) for adjacent chlorines on an aromatic ring. As seen in Figs. 3 and 3a. this becomes important with multiple substitution. For polychlorobiphenyls (Figs. 4. & 4a.) a further negative correction factor (-0.25) is required for chlorines in the ortho position to the biphenyl bond. These empirically-derived correction factors agree quite well with the lower-than-normal solvent accessible surface area (SASA) as calculated by the SAVOL program[1]. However, it is also known that in hexachlorobenzene the

chlorine atoms are forced slightly out of plane[2], and the negative correction may in part arise from less delocalization of the C-Cl bonds. In any case, the empirical corrections allow CLOGP to calculate 20 PCBs with an average deviation of only 0.07.

In Fig. 4a it can be seen that for the last three analogs (Cpds. 21, 22, & 23) CLOGP overpredicts by successively greater amounts. It seems very likey (as proposed by DeBruijn & Hermans who provided the slow-stir data[3]) that the octanol present in the saturated water phase is acting as a "surfactant" which solubilizes very lipophilic solutes by some pre-micellar mechanism. If the behavior being modelled may parallel this action, then CLOGP values above log 7.5 should be reduced in a manner that places an asymtotic upper level of about 9.0.

Halogens attached to the same or adjacent aliphatic carbon atoms are more lipophilic than those isolated from each other. CLOGP uses a topological measure of this separation, and it has been known for some time that information of actual spatial separation, as obtained from a molecular mechanics program such as CONCORD[1] would be useful. The empirical correction factor devised for CLOGP works well for many multi-halogenated aliphatics, such as the four hexachlorocyclohexanes as seen in Fig. 5. However, in aldrin two of the chlorines crowded on the bridged rings have a *topological* separation of three carbons, but they are actually as close as those separated by only two. If CLOGP was using data based on spacial distance (instead of Cl-C-C-Cl), it would yield a value of 6.0 which is better but still considerably low. Note that crowding of chlorines on an aliphatic framework reduces the effective SASA for each atom, but the effect is to make the solute *more* hydrophobic than expected, whicl is opposite from that of the halo-aromatics.

The negative correction employed by CLOGP for the branching of alkyl chains was always thought to be based on a sound theoretical basis: namely, that branching reduced solute size. Using four types of size measurements as provided by SAVOL ("bare" area or volume; or the area or volume as covered with a 1.5 A° layer), Fig. 6 shows that the correlations with the normal alkanes is almost perfect ($r^2 > 0.995$). Adding branched and cyclic alkanes leaves only cyclopropane as an outlier, which is not unexpected in view of the polarity of the strained carbon bonds.

This work impelled us to remeasure 2,3-dimethylbutane, as it appeared as a mild outlier. Using this new more reliable value, it appears that Eq. C4a is preferred to relate measured log P of alkanes to their size, namely the surface area with a 1.5 A^o 'sweater'.

The attempt to relate 'polar group branching' to solute size runs into difficulty, as seen in Fig. 7 with the four isomeric butanols. After overall size is accounted for by using Eq. C2 from Fig. 6, then most of the reduction of 0.53 in log P (0.88 to 0.35) still must come from the changed polarity of the hydroxyl group; i.e, it must account for a lowering of -2.36 to -2.76 in log P. An increase in hydrogen bond basicity, β, could account for this, but the data now available show only a slight change (0.45 to 0.49) in this parameter. Hydrogen bond acidity, α, is not thought to change for these isomers. Acccunting for just the size difference of the hydrocarbon portion of these butanols leaves even a larger apparent increase in hydroxyl polarity to be explained.

The early data for polar group branching on alkanes was provided largely by the alcohols and amines. As more solute types were measured to determine the relationship between log P and normal vs branched SASA, it was discovered that the negative branching correction for nitriles, carboxylic acids, and di-branched ketones was either very small or completely negligible. See Fig. 8. Reasons for this behavior are still being sought.

References

1. Generoulsy supplied by Dr. Robert Pearlman, University of Texas, Austin Tex.

2. Lorand,J. et. al., J.Phys.Org.Chem., 3, 659 (1990)

3. De Bruijn,J., Busser,F., Seinen,W. & Hermens,J., Envirn.Tox.Chem., 8, 499 (1989).

DDT & ANALOGS

DDT

CLOGP **6.91**

MEAS. 3.98(Lu & Metcalf,'73) Radiolabel
5.76(Fujita) Std. Shake-Flask
6.36(Karickhoff,'79) Re-extraction
6.60(Briggs)
6.91(de Bruijn & Hermens) Slow-Stir

DDE

CLOGP 6.94

MEAS. 6.96

FIGURE 1

0022-3565/90/2552-0824$03.00/0
THE JOURNAL OF PHARMACOLOGY AND EXPERIMENTAL THERAPEUTICS
Copyright © 1990 by The American Society for Pharmacology and Experimental Therapeutics
Vol. 255, No. 2
Printed in U.S.A.

Characterization of the Lipophilicity of Natural and Synthetic Analogs of Δ^9-Tetrahydrocannabinol and Its Relationship to Pharmacological Potency[1]

BRIAN F. THOMAS, DAVID R. COMPTON and BILLY R. MARTIN
Department of Pharmacology and Toxicology, Medical College of Virginia/Virginia Commonwealth University, Richmond, Virginia
Accepted for publication August 7, 1990

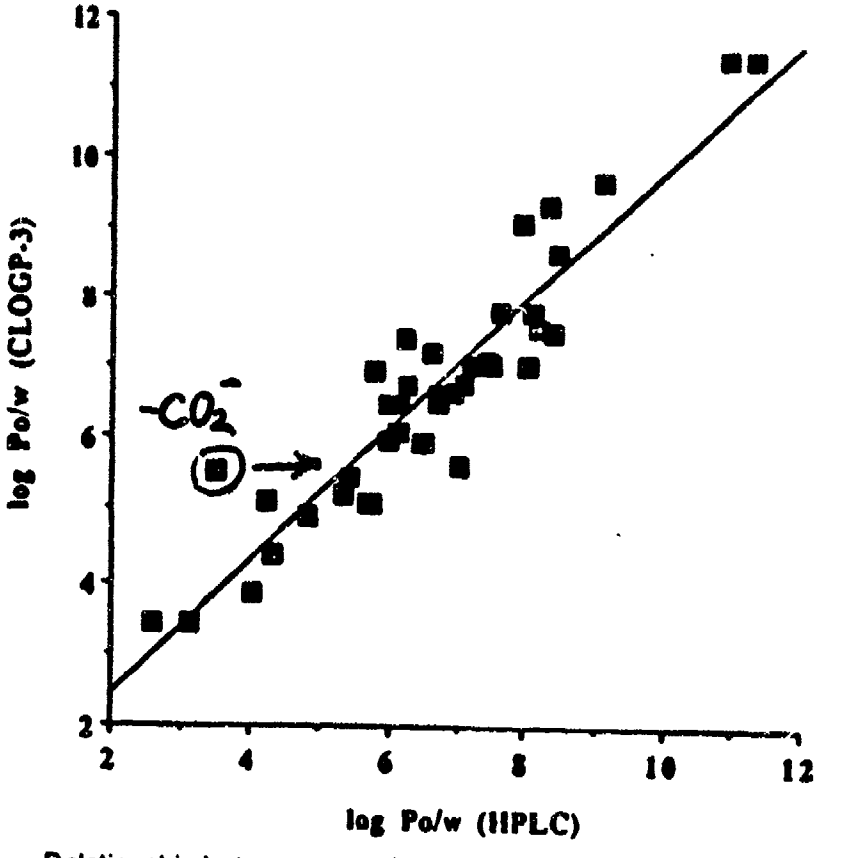
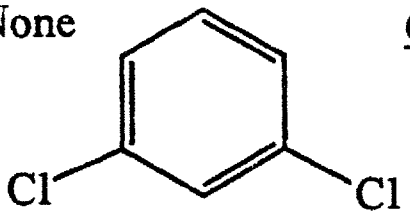
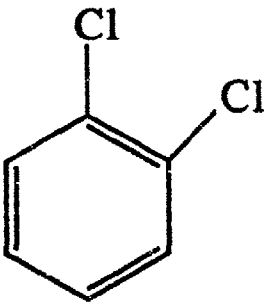
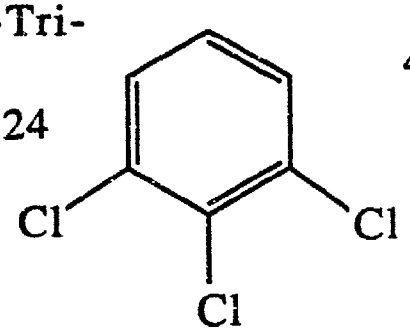


Fig. Relationship between experimentally determined log Po/w values by HPLC and computer-calculated log Po/w values for cannabinoid analogs. Correlation coefficient = 0.93.
FIGURE 2

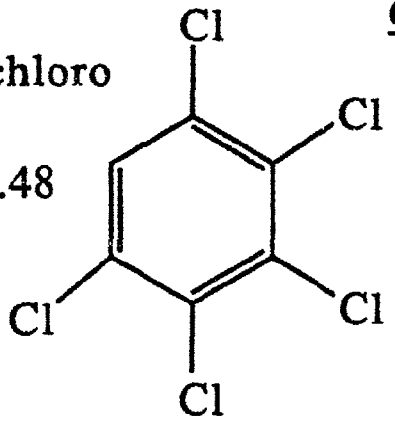
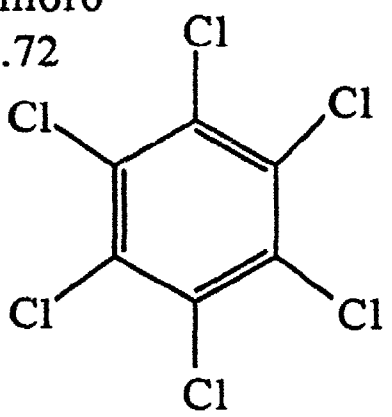
Proximity Corrections: Chlorobenzenes

		CLOGP	Meas.	Diff
1. Meta: None		3.57	3.53[1]	-0.04
2. 1,2,Di- F = -0.12		3.45	3.43[1] 3.31[2]	-0.02
3. 1,2,3-Tri- F = -0.24		4.04	4.14[1] 4.04[3]	0.10
4. 1,2,3,4-Tetra F = -0.36		4.63	4.64[1] 4.55[3]	0.01
5. 1,2,4,5-Tetra F = -0.24		4.75	4.60[1] 4.70[4]	-0.15

(1)DeBruijn, slow-stir. (2) Shake-flask, Pom.MedChem.
(3)Generator Column. (4) Chiou, G.

FIGURE 3-1

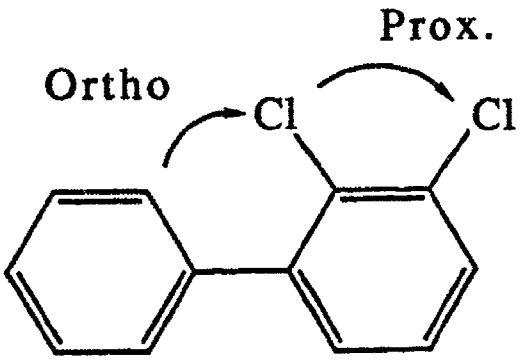
Proximity Corrections: Chlorobenzenes (cont)

		CLOGP	Meas.	Diff.
6. Pentachloro F = -0.48		5.23	5.18[1] 5.20[4]	-0.05
7. Hexachloro F = -0.72		5.70	5.73[1] 5.47[3] 5.2-7.0[5]	0.03

(1) DeBruijn, slow-stir. (2) Shake-Flask, PomMedChem. (3) Generator Column. (4) Chiou, G. (5) Platford, octanol as bulk phase vs. surface film

FIGURE 3-2

Chlorobiphenyls: Proximity & Ortho Corrections



Ortho = -0.25 each
Prox. = -0.12 each

Substituents	Meas.	Calc.	Diff.	Ortho	Prox.
1. 2-Cl	4.53	4.49	+0.04	-0.25	0.00
2. 2,4-Cl_2	5.16	5.21	-0.05	-0.25	0.00
3. 4-Cl	4.61	4.74	-0.13	0.00	0.00
4. 3,4-Cl_2	5.29	5.34	-0.05	0.00	-0.12
5. 2,6-Cl_2	4.98	4.96	+0.02	-0.50	0.00
6. 2,3,4-Cl_3	5.86	5.68	+0.16	-0.25	-0.24
7. 2,4,6-Cl_3	5.71	5.67	+0.04	-0.50	0.00
8. 2,4,5-Cl_3	5.90	5.80	+0.10	-0.25	-0.12
9. 2,3,4,5-Cl_4	6.41(1	6.27	+0.14	-0.25	-0.36
10. 2,3,4,5,6-Cl_5	6.74	6.62	+0.12	-0.50	-0.48
11. 2,2'-Cl_2	4.97	4.96	+0.01	-0.50	0.00
12. 2,3,2'-Cl_3	5.31(2	5.55	-0.24	-0.50	-0.12
13. 2,3,4,2'-Cl_4	6.11	6.14	-0.03	-0.50	-0.24

- - - -

(1) Not by slow-stir; 6.18 by generator column.

(2) By HPLC

FIGURE 4-1

Chlorobiphenyls: CLOGP Calc. Continued.

Substituents	Meas.	Calc.	Diff.	Ortho	Prox.
14. **2,3,2',3'-Cl_4**	6.18	6.14	+0.04	-0.50	-0.24
15. **2,6,2',6'-Cl_4**	5.94	5.88	+0.06	-1.00	0.00
16. **2,4,2',4'-Cl_4**	6.29(2	6.38	-0.07	-0.50	0.00
17. **2,2',4,5'-Cl_4**	6.36	6.38	-0.02	-0.50	0.00
18. **Bis-2,3,6-Cl_6**	7.12	7.07	+0.05	-1.00	-0.24
19. **Bis-2,4,6-Cl_6**	7.29	7.31	+0.02	-1.00	0.00
20. **Bis-2,3,4-Cl_6**	7.32	7.33	-0.01	-0.50	-0.48
***					
21. **Bis-3,4,5-Cl_6**	7.41	7.83	-0.42	0.00	-0.48
22. **Bis2,3,5,6Cl_8**	7.73	8.25	-0.52	-1.00	-0.48
23. **Decachloro**	8.27	9.20	-0.93	-1.00	-0.96

---

(2) Not by slow-stir.

FIGURE 4-2

HEXACHLOROCYCLOHEXANES

CLOGP = 3.75

MEASURED:

1. Alpha	3.80
2. Beta	3.78
3. Gamma	3.72
4. Delta	4.14

ALDRIN

MEAS.	6.50
CLOGP	5.41

ΔSASA/CL(in Aldrin)=15.17

ΔSASA/CL(in MeCl) = 17.39

FIGURE 5

Relating SAVOL to Log P

Volumes with and without 1.5A° surface: $V^{1,5}$ and V^0

Surface Areas (ditto) $A^{1.5}$ and A^0

HYDROCARBONS:

A. Straight Chain: Methane to Hexane:

1. Log P = 0.033 V^0 + 0.256; n=6; r^2=0.996; s=0.076
2. Log P = 0.010 $V^{1.5}$-0.497; n=6; r^2=0.998; s=0.055
3. Log P = 0.026 A^0 - 0.043; n=6; r^2=0.996; s=0.071
4. Log P = 0.017 $A^{1.5}$-1.452 n=6; r^2=0.999; s=0.036

B. Straight & Branched Chains: Above + i-Butane, 2,3-DiMe-Butane & Neopentane

1. Log P = 0.032 V^0 + 0.286; n=9; r^2=0.988; s=0.110
2. Log P = 0.010 $V^{1.5}$-0.537; n=9; r^2=0.992; s=0.091
3. Log P = 0.026 A^0 - 0.057; n=9; r^2=0.986; s=0.115
4. Log P = 0.018 $A^{1.5}$-1.583; n=9; r^2=0.987; s=0.114

C. Incl. Cyclic Hydrocarbons: Cy-propane (1.72) an Outlier

1. Log P = 0.032 V^0 + 0.299; n=11; r^2=0.987; s=0.101
   Pred.LogP L3TJ = 2.02; dev.= -0.305
2. Log P = 0.010 $V^{1.5}$-0.538; n=11: r^2=0.992; s=0.081
   2.04; dev.=-0.317
3. Log P = 0.026 A^0 - 0.058; n=11; r^2=0.985; s=.109
   1.99; dev.=-0.270
4. Log P = 0.018 $A^{1.5}$-1.598; n=11; r^2=0.987; s=0.103
   2.07; dev.=-0.351

4a LogP(iY4Y) 3.95 → 3.42

LogP = 0.017 $A^{1.5}$ - 1.44; n=11; r^2 = 0.995; s = 0.63

Pred. L3TJ 2.07

FIGURE 6

Relating SAVOL to Log P

Volumes with and without 1.5A^{0} surface: $V^{1.5}$ and V^{0}

Surface Areas (ditto) $A^{1.5}$ and A^{0}

Butanols

A. Using Eq.C-2 for Total Solute Volume:

$$\text{Log } P = 0.010\ V^{1.5} - 0.538$$

		Effect OH	β
1.Q4 (M=0.88)	3.775-0.538 =	3.24;	-2.36; .45
2.Q1Y (M=0.76)	3.690-0.538 =	3.15;	-2.39; .45
3.QY2(M=0.61)	3.700-0.538 =	3.16;	-2.55; .47
4.QX (M=0.35)	3.650-0.538 =	3.11;	-2.76; .49

B. Using Eq.C-4 for Area of Hydrocarbon Only

$$\text{Log } P = 0.018\ A^{1.5} - 1.60$$

1.Q4 (M=0.88)	(210.2)0.18 -1.6 =	2.18;	-1.30; .45
2.Q1Y (M=0.76)	(208.7)0.18 -1.6 =	2.16;	-1.40; .45
3.QY2 (M=0.61)	(218.8)0.18 -1.6 =	2.34;	-1.73; .47
4.QX (M=0.35)	(212.1)0.18 -1.6 =	2.22;	-1.87; .49

FIGURE 7

Relating SAVOL to Log P

Volumes with and without 1.5A^{o} surface: $V^{1,5}$ and V^{0}

Surface Areas (ditto) $A^{1.5}$ and A^{0}

Using Eq. C-2 with Total Solute Volume:

Pentanoic Acids:				**Effect CO2H**
1. QV4 (M=1.39)	4.355	-0.538	= 3.82;	-2.43
2. QV1Y(M=1.16)	4.272	-0.538	= 3.73;	-2.57
3. QVY2(M=1.18)	4.292	-0.538	= 3.75;	-2.57
4. QVX (M=1.47)	4.196	-0.538	= 3.66;	-2.19

For All: α = **0.54;** β = **0.41**

Nonanones:				**Effect =O;**	β
1. 4V4 (M=2.97)	6.315	-0.538	= 5.78;	-2.71	.48
2. t-Bu (M=3.00)	5.540	-0.538	= 5.00;	-2.00	.46

FIGURE 8

BLANK

APPENDIX A

AUTHORS OF PAPERS IN THESE PROCEEDINGS

BLANK

APPENDIX B

ORGANIZATIONS OF AUTHORS IN THESE PROCEEDINGS

Central Michigan University	49
Instituto de Quimica Fisica	33
La Sierra University, Riverside	71
Pomona College	105,139
Université de NANTES	13
University of California, Irvine	9
University of New Orleans	77,85
U.S. Army Chemical Research, Development and Engineering Center	7,65,71

BLANK

APPENDIX C

Meeting on Solute/Solvent Interactions

Agenda

Wednesday, 29 May Seminar Center

0800-0815	Welcome/Introduction

Equilibrium

0815-0845	R. Taft
0845-0925	I. Koppel
0925-1000	M. Berthelot
1000-1030	J. Abboud
1030-1100	J. Lorand

Theoretical

1100-1130	P. Politzer
1130-1200	G. Famini
1200-1230	L. Wilson

1400-1700	Demonstrations, E3160

Thursday, 30 May Seminar Center

New Approaches

0900-930	C. Hansch
0930-1000	A. Leo
1000-1030	S. Karickhoff
1030-1100	L. Carreira

Wednesday Afternoon

Polymers

1300-1350	M. Abraham
1350-1420	M. Harris
1420-1450	J. Grate
1450-1520	W. Shuely
1520-1550	A. McGill

Blank

APPENDIX D

ATTENDEES LIST

Meeting on Solute/Solvent Interactions
29-30 May 1991
Aberdeen Proving Gd, MD

Jose-Luis Abbud
Instituto de Quimica Fisica
"Rocasolano"
Consejo Superior de
Investigaciones Cientificas
C/ Serrano 119
28006 Madrid
Spain

Micheal Abraham
Dept of Chemistry
University College London
20 Gordon St
London WC1HOAJ
United Kingdom

William Ashman
SMCCR-RSP-C
U.S. Army Chemical RD&E Center
APG, MD 21010

Phillip Bartram
SMCCR-PPD
U.S. Army Chemical RD&E Center
APG, MD 21010

Michel Berthelot
Laboratorie de Spectrochimie
Moleculaire
Universite de NANTES
2, Rue de la Houssiniere
44072 NANTES
CEDEX 03
France

Lionel Carreira
Dept of Chemistry
University of Georgia
Athens, GA 30602

Christopher Cramer
SMCCR-RSP-C
U.S. Army Chemical RD&E Center
APG, MD 21010

George Famini
SMCCR-RSP-C
U.S. Army Chemical RD&E Center
APG, MD 21010

Jay Grate
Naval Research Laboratory
Code 6170
Washington, DC 20375

Alice Harper
SMCCR-RSP-C
U.S. Army Chemical RD&E Center
APG, MD 21010

J. Milton Harris
Dept of Chemistry
University of Alabama
Huntsville, AL 35899

Corwin Hansch
Department of Chemistry
Pomona College
Claremont, CA 91711

Samuel Karrickhoff
Environmental Research Laboratory
Chemical Division
U.S. Environmental Protection Agency
Athens, GA 30613

I. A. Koppel
Department of CHemistry
Tartu University

Tartu Estonia
USSR

Albert Leo
Department of Chemistry
Pomona College
Claremont, CA 91711

John Lorand
Dept of Chemistry
Central Michigan University
Mt. Pleasant, MI 48859

Andrew McGill
Naval Research Laboratory
Code 6170
Washington, DC 20375

Peter Politzer
Dept of Chemistry
University of New Orleans
New Orleans, LA 70148

Wendel Shuely
SMCCR-RSC-P
U.S. Army Chemical RD&E Center
APG, MD 21010

Robert Taft
Dept of Chemistry
University of California-Irvine
Irvine, CA 92717

Joseph Urban
SMCCR-RSP-C
U.S. Army Chemical RD&E Center
APG, MD 21010

Leland Wilson
Dept of Chemistry
La Sierra University
Riverside, CA 92515