QSAR Parameters

Hugo Kubinyi

Germany

E-Mail kubinyi@t-online.de
HomePage www.kubinyi.de

What is QSAR ?

QSAR (quantitative structure-activity relationships) includes all statistical methods, by which biological activities (most often expressed by logarithms of equipotent molar activities) are related with structural elements (Free Wilson analysis), physicochemical properties (Hansch analysis), or fields (3D QSAR).

Classical QSAR analyses (Hansch- and Free Wilson analyses) consider only 2D structures. Their main field of application is in substituent variation of a common scaffold.

3D-QSAR analysis (CoMFA) has a much broader scope. It starts from 3D structures and correlates biological activities with 3D-property fields.

Drug Action: From Experience to Theory to Rules

- 1900, H. H. Meyer and C. E. Overton: lipoid theory of narcosis
- 1930's, L. Hammett: electronic sigma constants
- 1964, C. Hansch and T. Fujita: QSAR
- 1984, P. Andrews: affinity contributions of functional groups
- 1985, P. Goodford: GRID (hot spots at protein surface)
- 1988, R. Cramer: 3D QSAR
- 1992, H.-J. Böhm: LUDI interaction sites, docking, scoring
- 1997, C. Lipinski: bioavailability rule of five
- 1998, Ajay, W. P. Walters and M. A. Murcko; J. Sadowski and H. Kubinyi: drug-like character

[Contribution from the Department of Chemistry, Pomona College, Claremont, California]

ρ–σ–π Analysis. A Method for the Correlation of Biological Activity and Chemical Structure

By Corwin Hansch and Toshio Fujita[1]

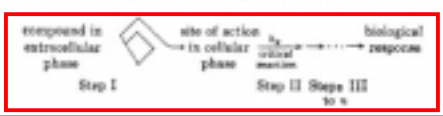
Received August 19, 1963

JACS 86, 1616 (1964)

Using the substituent constant, σ, and a substituent constant, π, defined as $\pi = \log P_X - \log P_H$ (P_H is the partition coefficient of a parent compound and P_X that of a derivative), regression analyses have been made of the effect of substituents on the biological activity of benzoic acids on mosquito larvae, phenols on gram-positive and gram-negative bacteria, phenyl ethyl phosphate insecticides on housefllies, thyroxine derivatives on rodents, diethylaminoethyl benzoates on guinea pigs, and carcinogenic compounds on mice.

Recently[2,3] we have shown the advantage of using partition coefficients in connection with the Hammett equation to rationalize the substituent effect on the growth-promoting activity of the phenoxyacetic acids and the bactericidal action of chloromycetin derivatives on various bacteria. In particular, it was found that a substituent constant, π, patterned after the Hammett σ-constant was useful in evaluating the lipo-hydrophilic character of a molecule upon which biological activity is highly dependent. π is defined as $\pi = \log (P_X/P_H)$ where P_H is the partition coefficient of a parent compound and P_X is the value for a derivative. The reference system is octanol-water, and all of the work reported in this paper refers to this pair of solvents. The purpose of this report is to show that our previously employed expression appears to have general applicability.

We have assumed that the rate-limiting conditions for many biological responses to chemicals can be defined in the simplest and most general way as follows.

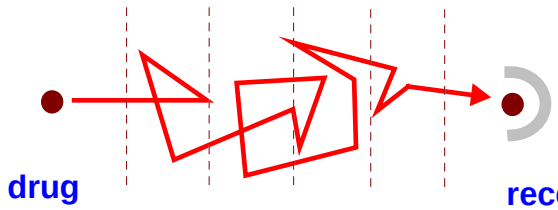


question makes its way from a very dilute solution outside the cell to a particular site in the cell which may be within an organelle. This is visualized as being a relatively slow process, the rate of which is highly dependent on the molecular structure of the compound in question. It is assumed, as a first approximation, for many types of biologically active molecules there will be one key rate-controlling reaction at the active sites. This could be formulated as in eq. 1. A is the probability of a molecule reaching a site of action in a given time interval and C is the extracellular molar concentration of the compound being tested. The product AC represents the "effective" concentration at the sites of action. The constant k_X might be either an equilibrium or rate constant. It is assumed that a relatively large number of reaction sites are available so that these remain essentially constant during the test interval. In certain instances, it may be that new sites are being constantly generated. It also is considered that the many reactions which may occur subsequently to the one critical one (Steps III to x) before the visible response is elicited can be neglected for a first approximation.

$$\text{rate of biological response} = \frac{d(\text{response})}{dt} = ACk_X \quad (1)$$

Drug Transport and Drug Receptor Interaction

The "random walk" process



drug

aqueous phases and lipophilic barriers

receptor (binding site)

biological activity = f (transport + binding) =

$$- k_1 (lipo)^2 + k_2 (lipo) + k_3 (pol) + k_4 (elec) + k_5 (ster) + k_6$$

Specifics of Drug Action

Lipophilicity and dissociation / ionization are responsible for transport and distribution of drugs in biological systems.
The geometric fit and the complementarity of the surface 3D properties of a ligand are responsible for its affinity to a binding site.

- **Which conformation is the biologically active conformation ?**
  - **conformation *in vacuo***
  - **conformation in the crystal**
  - **conformation in aqueous solution**
  - **conformation at the binding site**

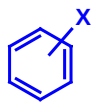
Basic Requirements in QSAR Studies

- all analogs belong to a congeneric series
- all analogs exert the same mechanism of action
- all analogs bind in a comparable manner
- the effects of isosteric replacement can be predicted
- binding affinity is correlated to interaction energies
- biological activities are correlated to binding affinity

Molecular Properties and Their Parameters

Molecular Property	Corresponding Interaction	Parameters
Lipophilicity	hydrophobic interactions	log P, π, f, R_M, χ
Polarizability	van-der-Waals interactions	MR, parachor, MV
Electron density	ionic bonds, dipol-dipol interactions, hydrogen bonds, charge transfer interactions	σ, *R*, *F*, κ, quantum chemical indices
Topology	steric hindrance geometric fit	E_S, r_V, L, B, distances, volumes

Hammett equation



$$\rho\sigma = \log k_{RX} - \log k_{RH}$$

QSAR Models - Hansch model (property-property relationship)

Definition of the lipophilicity parameter π

$\pi_X = \log P_{RX} - \log P_{RH}$

Linear Hansch model

$\text{Log } 1/C = a \log P + b\,\sigma + c\,MR + \ldots + k$

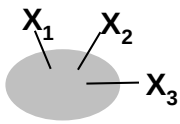
Nonlinear Hansch models

$\log 1/C = a (\log P)^2 + b \log P + c\,\sigma + \ldots + k$

$\log 1/C = a\,\pi^2 + b\,\pi + c\,\sigma + \ldots + k$

$\log 1/C = a \log P - b \log (\beta P + 1) + c\,\sigma + \ldots + k$

Free-Wilson model (structure-property relationship)



$\log 1/C = \Sigma\, a_i + \mu$

a_i = substituent group contributions

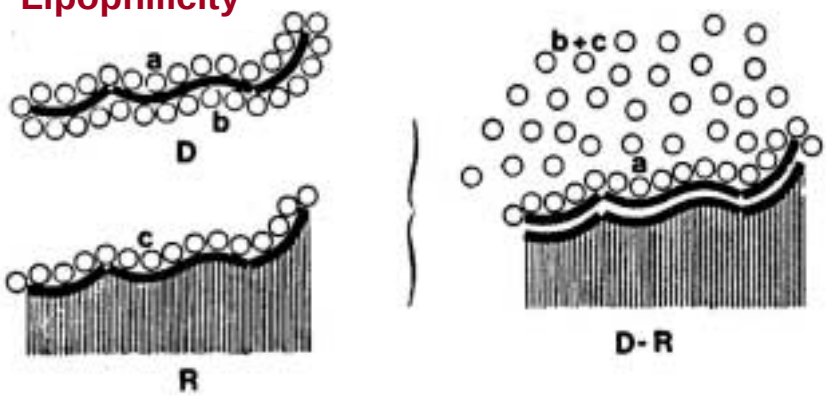
μ = activity contribution of reference compound

Mixed Hansch/Free-Wilson model

$\log 1/C = a (\log P)^2 + b \log P + c\,\sigma + \ldots + \Sigma\, a_i + k$

$\log 1/C = a \log P - b \log (\beta P + 1) + c\,\sigma + \ldots + \Sigma\, a_i + k$

Lipophilicity



Hydrophobic interaction between a drug and a binding site at a receptor

Definition of Partition Coefficients

$P = c_{org}/c_{aq}$ (n-octanol/water system)

Octanol/Water Partition Coefficients of Carbamates

J. B. Houston et al., J. Pharmacol. Exp. Ther. **189**, 244 (1974)

R-OCONH$_2$	P	log P	Δ log P = π_{CH_2}
Methyl	0.22	-0.66	
Ethyl	0.70	-0.15	} 0.51
Propyl	2.3	0.36	} 0.51
Butyl	7.1	0.85	} 0.49
Pentyl	22.5	1.35	} 0.50
Hexyl	70.8	1.85	} 0.50
Heptyl	230	2.36	} 0.51
Octyl	700	2.85	} 0.49
sec-Butyl	4.5	0.65	-0.20 *)
tert-Butyl	3.0	0.48	-0.37 *)

*) relative to *n*-butyl carbamate

n-Octanol/Water as a Standard System

- membrane analogous structure
- hydrogen bond donor and acceptor
- practically insoluble in water
- no desolvation on transfer into organic phase
- very low vapor pressure
- transparent in the UV region
- large data base of log P values

Additivity Principle of π Values (C. Hansch, 1964)

$$\pi_X = \log P_{R\text{-}X} - \log P_{R\text{-}H}$$

The lipophilicity parameter π is an additive, constitutive molecular parameter; compare the Hammett Equation:

$$\rho\sigma_X = \log K_{R\text{-}X} - \log K_{R\text{-}H}$$

Hugo Kubinyi, www.kubinyi.de

π Values of Aromatic Substituents

Substituent	π_{meta}	π_{para}	$\pi_{Benzene}$
H	0.00	0.00	0.00
CH_3	0.51	0.52	0.56
Cl	0.76	0.70	0.71
Br	0.94	1.02	0.86
OH	-0.49	-0.61	-0.67
OCH_3	0.12	-0.04	-0.02
NO_2	0.11	0.24	-0.28

→ cf.: Lipophilicity contribution of an aliphatic CH_2 group = 0.50

Toxicity of Substituted Benzoic Acids in Mosquito Larvae

all compounds, σ

$\log 1/C = 1.069\,\sigma + 1.780$

(n = 14; r = 0.711; s = 0.427)

without the 4-nitro analog, σ (Hansen, 1962)

$\log 1/C = 1.454\,\sigma + 1.787$

(n = 13; r = 0.918; s = 0.243)

all compounds, π (Hansch and Fujita, 1964)

$\log 1/C = 0.535\,\pi + 1.602$

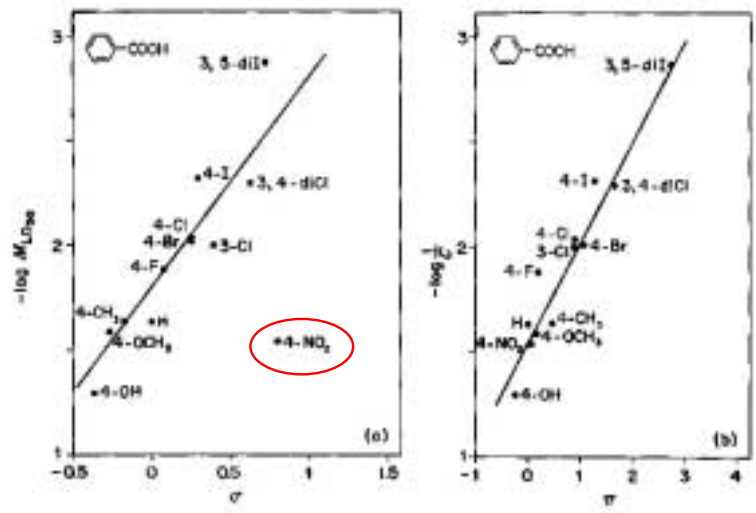
(n = 14; r = 0.969; s = 0.151)

intercorrelation π vs. σ (without 4-nitro analog)

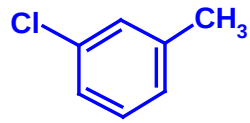
r = 0.91

Hugo Kubinyi, www.kubinyi.de

Toxicity of Substituted Benzoic Acids in Mosquito Larvae



Calculation of the Log P Value of m-Chloro-toluene



$\log P = \log P_{Benzene} + \pi_{Cl} + \pi_{Me}$

$= 2.13 + 0.71 + 0.56 = 3.40$

$\log P = \log P_{Toluene} + \pi_{meta\text{-}Cl}$

$= 2.69 + 0.76 = 3.45$

$\log P = \log P_{Chlorobenzene} + \pi_{meta\text{-}Me}$

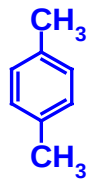
$= 2.84 + 0.51 = 3.35$

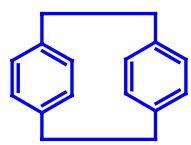
$\log P_{exp} = 3.28$

Hugo Kubinyi, www.kubinyi.de

Nonadditivity of π Values

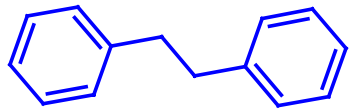
- intramolecular hydrogen bonds
- ortho effects (phenols)
- polysubstituted aromatic compounds
- conjugation (push pull effect)
- heterocycles
- cyclophanes





surface area	150.3	196.6
exptl. Log P	3.15	2.33

Wrong Application („Failure“) of the π Concept



$\pi_X = \log P_{R-X} - \log P_{R-H}$ (correct)

$\log P = \Sigma\pi_i$ (wrong)

$\log P_{C_6H_5CH_3} = \log P_{C_6H_6} + \pi_{CH_3}$ (correct)

$\log P_{C_6H_5CH_2CH_2C_6H_5} = 2 \log P_{C_6H_6} + 2\,\pi_{CH_3}$

$= 4.26 + 1.00 = 5.26$ (wrong)

1,2-diphenylethane, exp. log P = 4.79

Hydrophobic Fragmental Constants f

R. Rekker, The Hydrophobic Fragmental Constant, Elsevier, Amsterdam 1977; R. Rekker. Eur. J. Med. Chem. 14, 479 (1979)

$\log P = \Sigma\, a_i f_i$ (R. Rekker, 1973)

Fragment	Rekker constant f	Leo constant f	$\pi_{benzene}$
H	0.175	0.23	0.00
CH_3	0.702	0.89	0.56
$>CH_2$	0.530	0.66	
>CH-	0.235	0.43	
C_6H_5	1.886	1.90	1.96
C_6H_4	1.688		
C_6H_3	1.431		

$\log P_{C_6H_5CH_2CH_2C_6H_5} = 2\, f_{C_6H_5} + 2\, f_{CH_2}$

$= 3.77 + 1.06 = 4.83$ (correct)

Fragment	Rekker constant f		Leo constant f		$\pi_{benzene}$
	Aliph.	Arom.	Aliph.	Arom.	
F	-0.462	0.399	-0.38	0.37	0.14
Cl	0.061	0.922	0.06	0.94	0.71
Br	0.270	1.131	0.20	1.09	0.86
I	0.587	1.448	0.60	1.35	1.12
OH	-1.491	-0.343	-1.64	-0.40	-0.67
-O-	-1.581	-0.433	-1.81	-0.57	
COOH	-0.954	-0.093	-1.09	-0.03	-0.32
NH_2	-1.428	-0.854	-1.54	-1.00	-1.23
NH	-1.825	-0.964	-2.11	-1.03	
NO_2	-0.939	-0.078	-1.26	-0.02	-0.28
$CONH_2$	-1.970	-1.109	-2.18	-1.26	-1.49
>C=O	-1.703	-0.842	-1.90	-0.32	
CF_3	0.757	1.331			0.88
C≡N	-1.066	-0.205	-1.28	-0.34	-0.57
SH	0.000	0.620			0.39
-S-	-0.510	0.110	-0.79	0.03	

Experimental Determination of Log P Values

- Shake flask method
- Reversed phase thin layer chromatography
- High performance liquid chromatography (HPLC)

High Performance Liquid Chromatography (HPLC)

$\log P = a \log k' + const.$ $\quad k' = (t_r - t_o)/t_o$

different standard compounds, n-octanol-saturated, persilylated CORASIL C-18®, pH 7 phosphate buffer

$\log P = 1{,}025\ (\pm 0.06) \log k' + 0.797$
$(n = 33;\ r = 0.987;\ s = 0.127)$

S. Unger et al., J. Pharm. Sci. 67, 1364 (1978)

Lipophilicity Estimation Software

CLogP (Daylight Chemical Information Systems)
http://www.daylight.com/daycgi/clogp

KowWin (Syracuse Research Corp., SRC; W. M. Meylan and P. H. Howard, J. Pharm. Sci. 84, 83-92 (1995))
http://esc.syrres.com/interkow/kowdemo.htm

Interactive Analysis's LogP Program (Mark Parham; includes Lipinski numbers)
http://www.logp.com/main.html

Neuro-Heuristic Program ALogPS (Igor Tetko, Lausanne)
http://www.vcclab.org

Experimental LogP values (SRC, n = 13250)
http://esc.syrres.com/interkow/PhysProp.htm

Hugo Kubinyi, www.kubinyi.de

Polarizability Parameters

Molar volume, Molar Refractivity, Parachor

$$MV = \frac{MW}{d} \qquad MR = \frac{n^2 - 1}{n^2 + 2} \cdot \frac{MW}{d}$$

$$PA = \gamma^{1/4} \cdot \frac{MW}{d}$$

d = density; n = refraction index;
γ = surface tension

(MR is most often scaled by a factor of 0.1)

Binding of Neutral Compounds to BSA

Log 1/C = 0.751 (±0.07) log P + 2.300
(n = 42; r = 0.960; s = 0.159)

Log 1/C = 0.024 (±0.02) MR + 2.901
(n = 42; r = 0.307; s = 0.536)

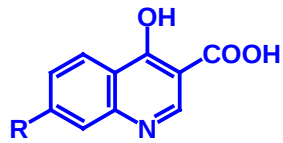
Binding of Phenyl ß-D-Glucosides to Concanavalin A

Log M_{50} = 0.971 (±0.56) π + 2.37
(n = 19; r = 0.664; s = 0.095)

Log M_{50} = 0.019 (±0.003) MR + 2.23
(n = 19; r = 0.954; s = 0.038)

4-Hydroxyquinoline Carboxylic Acids, Inhibition of Malate Dehydrogenase

pI_{50} = 0.70 (±0.17) MR + 2.29
(n = 13; r = 0.939; s = 0.315)



4-Hydroxyquinoline Carboxylic Acids, Inhibition of Ascites Cell Respiration

pI_{50} = 0.46 (±0.11) π + 3.22
(n = 14; r = 0.933; s = 0.280)

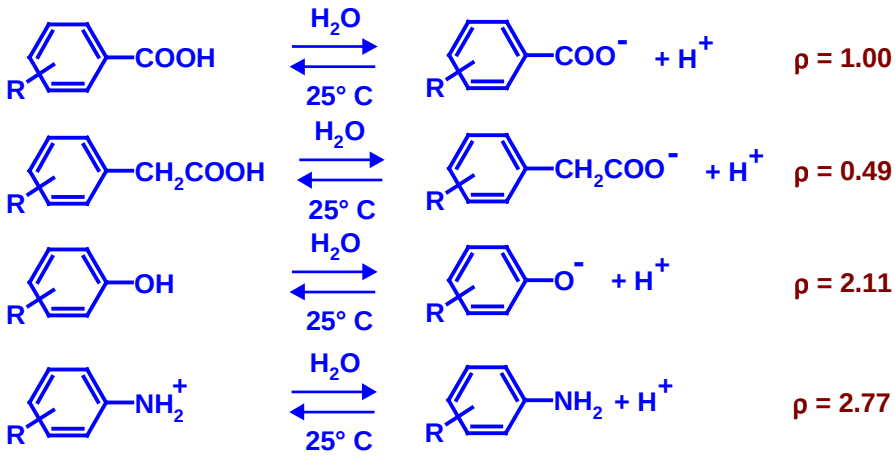
Binding of N-Acyl L-Amino Acid Methyl Esters $R_1CONH\text{-}CH(R_2)\text{-}COOMe$ to Chymotrypsin

Log $1/K_m$ = 0.082 (±0.02) MR_1 + 1.382 (±0.87) π_2 – 3.876
(n = 21; r = 0.934; s = 0.331)

Electronic Parameters

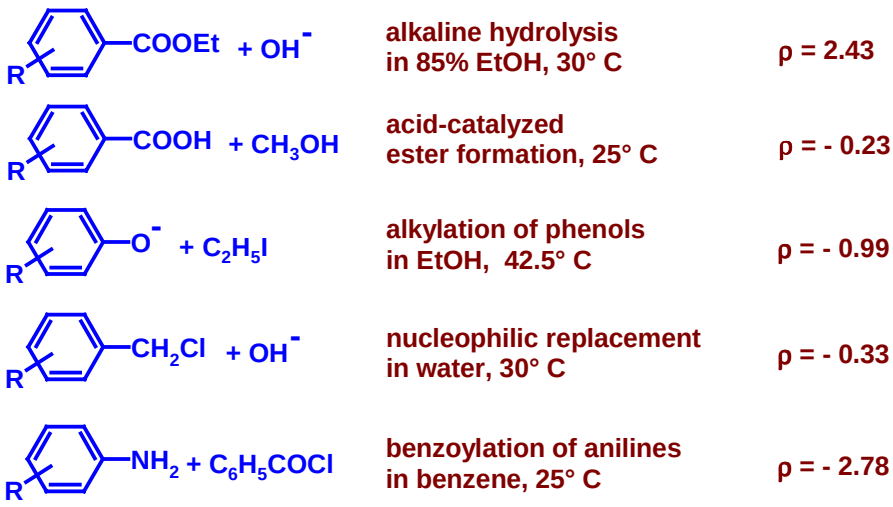
Hammett Equation

$$\rho\sigma = \log K_{RX} - \log K_{RH}$$



Hammett Equation

$$\rho\sigma = \log k_{RX} - \log k_{RH}$$



Calculation of pK_a Values $pK_{a\ R\text{-}X} = pK_{a\ R\text{-}H} - \rho\sigma$

pK_a value of 3,5-dinitro-4-methyl-benzoic acid

(pK_a benzoic acid = 4.20)
experimental value = 2.97
calculated value = 4.20 - (0.71 - 0.17 + 0.71) = 2.95

pK_a value of m-hydroxybenzoic acid

(pK_a benzoic acid = 4.20, phenol = 9.92;
$\rho_{phenols}$ = 2.23; -OH: σ_{meta} = 0.12; -COOH: σ_{meta} = 0.37;
$-COO^-$: σ_{meta} = -0.10)
experimental values = 4.06 and 9.92
calculated for the carboxylate group
= 4.20 - 0.12 = 4.08
calculated for the phenolic OH group
= 9.92 - 2.23·(-0.10) = 10.14

Hugo Kubinyi, www.kubinyi.de

pK_a value of 3-methoxy-5-nitro-aniline

(pK_a aniline = 4.57; $\rho_{aniline}$ = 2.81)
experimental value = 2.13
calculated value = 4.57 - 2.81 (0.12 + 0.71) = 2.24

pK_a values of chloro-, trichloro- and cyanoacetic acid

(pK_a RCOOH = 4.66; ρ_{RCOOH} = 1.62; σ^* for $-CH_2Cl$, $-CCl_3$ and $-CH_2C{\equiv}N$ = 1.05, 2.65 and 1.30; σ^* = aliphatic σ value)

chloroacetic acid, experimental value = 2.85
calculated value = 4.66 - 1.62·(1.05) = 2.96

trichloroacetic acid, experimental value = 0.70
calculated value = 4.66 - 1.62·(2.65) = 0.37

cyanoacetic acid, experimental value = 2.45
calculated value = 4.66 - 1.62·(1.30) = 2.55

Quantum Mechanical Descriptors

Atom partial charges:

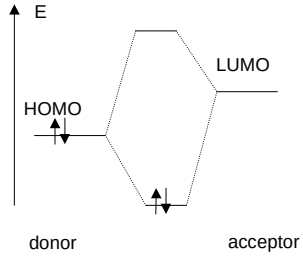
Mulliken population analysis (orbital population)
ESP charges (mapping EP to atom locations)

Dipole moment:

strength and orientation behavior of a molecule in an electrostatic field

HOMO / LUMO ("frontier orbital theory"):

HOMO = energy of highest occupied molecular orbital, "nucleophilicity"
LUMO = energy of lowest unoccupied molecular orbital, "electrophilicity"



Superdelocalizability:

estimate for the reactivity of positions in aromatic hydrocarbons

Steric Parameters

E_s values (acid-catalyzed hydrolysis of RCOOR´, based on CH_3COOR´)

$$E_s = \log (k/k_o)_A$$

Van-der-Waals radii of symmetric substituents ($r_{v\,(av)}$ = mean radius)

$$E_s = -1.839\ r_{v\,(av)} + 3.484$$

(n = 6; r = 0.996; s = 0.132)

Charton's steric substituent parameter (r_v = minimal van-der-Waals radius)

$$\upsilon_X = r_{vX} - r_{vH} = r_{vX} - 1.20$$

$$E_s = -2.062\ (\pm 0.86)\ \upsilon - 0.194\ (\pm 0.10)$$

(n = 104; r = 0.978; s = 0.250)

Hancock E_s^c values (corrected for the number of α-hydrogen atoms, n_H)

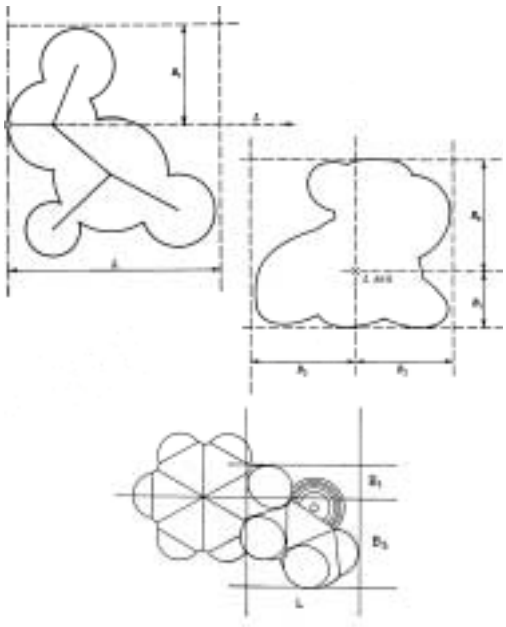
$E_s^c = E_s + 0.306\ (n_H - 3)$

Fujita E_s^c values for unsymmetrical substituents

$CR^1R^2R^3$: weighted sum of the E_s^c values of R^1, R^2 and R^3

Verloop's STERIMOL parameters

L, B1, B2, B3, and B4
revised parameters L, B1, and B5



Sterimol Parameters (Arië Verloop)

A graphical definition of second generation STERIMOL parameters (Verloop, 1987), using the substituent $CH_2CH_2COCH_3$ as an example; L is the length, B_1 the minimal width and B_5 a newly defined maximal width

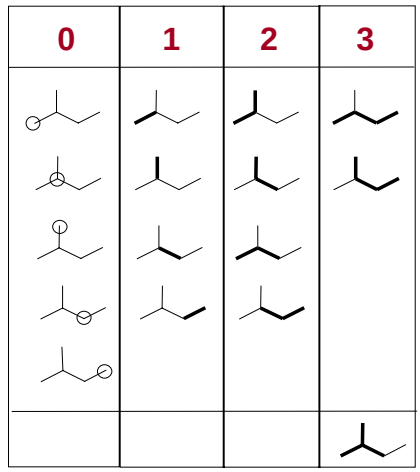
Kier-Hall Molecular Connectivity Indices $^{i}\chi$

Bond-path indices for isopentane = electron-weighted subgraph counts

$^{0}\chi = \Sigma$ (# σ-electrons of i)$^{-0.5}$

$^{1}\chi = \Sigma$ ($^{0}\chi$(i).$^{0}\chi$(j))$^{-0.5}$ (over all bonds ij)

... etc.



$^{0}\chi$ $^{1}\chi$ $^{2}\chi$ $^{3}\chi_{P}$, $^{3}\chi_{C}$

Parameter Tables and Log P Data Bases

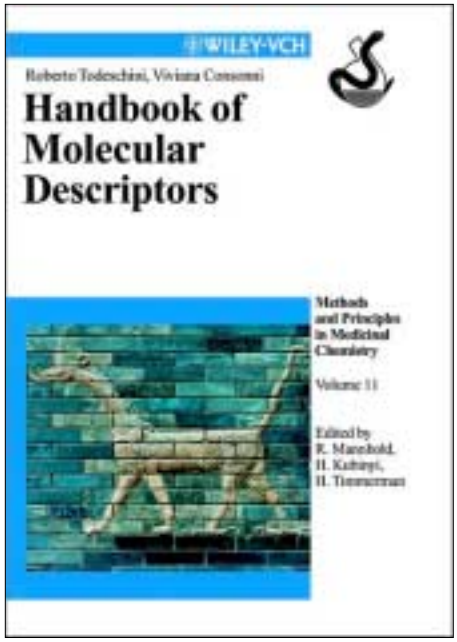
C. Hansch et al., J. Med. Chem. 16, 1207 (1973); J. Med. Chem. 20, 304 (1977); 284 aromatic substituents, π, σ_{meta}, σ_{para}, *F*, *R* and MR

C. Hansch and A. Leo, Substituent Constants for Correlation in Chemistry and Biology, Wiley, New York, 1979. All known substituent parameters; about 15,000 log P values

C. Hansch, A. Leo, and D. Hoekman, Exploring QSAR. Hydrophobic, Electronic, and Steric Constants, American Chemical Society, Washington, 1995.

R. Todeschini and V. Consonni, Handbook of Molecular Descriptors, Methods and Principles in Medicinal Chemistry, Volume 11, R. Mannhold, H. Kubinyi and H. Timmerman, Eds., Wiley-VCH, Weinheim, 2000.

MedChem and BioByte Databases, Program DRAGON



Program DRAGON

Roberto Todeschini

free download of the DRAGON program from www.disat.unimib.it/chm/Dragon.htm

calculates 1,481 molecular descriptors, distributed in 18 blocks, from MOL, SDF or Hyperchem files.

Hugo Kubinyi, www.kubinyi.de

The Software DRAGON

R. Todeschini, © Talete srl
calculates 1,481 molecular descriptors, distributed in 18 blocks.

General features of the version 2.1:
Each molecule has to be represented by 3D geometrical coordinates, hydrogen included.
The maximum number of atoms (hydrogen atoms included) is 150.
Calculations are quite fast; however, several condensed or adjacent rings can slow down the calculations.
The total number of molecules was extended to 1,000.

Features of the commercial version:
up to 2,500 molecules (instead of 1,500),
maximum number of 300 atoms, including hydrogen (instead of 150)
maximum number of 100 y responses (instead of 20)
number of molecular descriptors: 1,486 (instead of 1,481;
new models for the calculation of experimental properties)
loading of SMILES format (not possible in the Web version)

Hugo Kubinyi, www.kubinyi.de