Rational Drug Design lecture 10

Łukasz Berlicki

Od eksperymentu do teorii

- **1964, C. Hansch i T. Fujita:** QSAR
- **1984, P. Andrews:** functional group contributions to interactions with molecular target
- **1985, P. Goodford:** GRID (binding points on protein surface)
- **1988, R. Cramer:** 3D QSAR
- **1992, H.-J. Böhm:** LUDI, interactions, docking, scoring functions
- **1997, C. Lipinski:** bioavailability, rule of five
- **1998, Ajay, W. P. Walters and M. A. Murcko; J. Sadowski i H. Kubinyi:** drug-like compounds

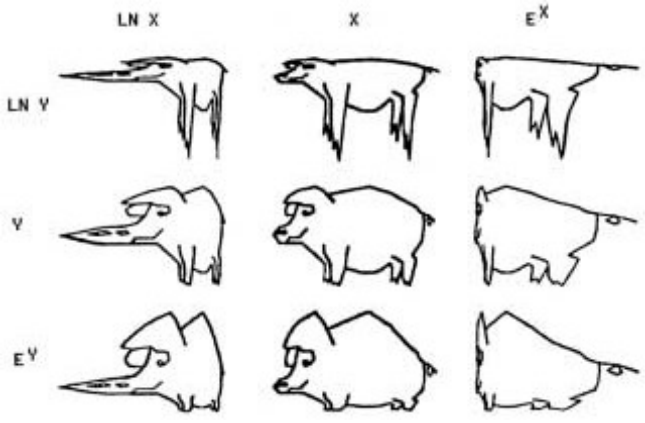
Statistics

- **Lies, damned lies, and statistics (Benjamin Disraeli)**

Benjamin Disraeli
Prime minister of Great Britain
XIX c.

- **Statistical data analysis** (**Bailar,** Clin. Pharmacol. Therapeutics, 1979)
  - There are only right answers.
  - Statistics are not the only way to wisdom.
  - Rare cases happen constantly.
  - No sample is big enough (and what?).
  - No analysis is perfect (and what?).
  - Something is always wrong with the data.
- **All models are bad, but some are useful**
- **Scaling variables changes the result**
- **The graph says more than equations**
- **Validation - a very important but difficult problem.**

Scaling variables



QSAR

- **QSAR** – *quantitative structure-activity relationships* - **statistical methods** by means of which the dependence of biological activity on structural elements, physicochemical properties or three-dimensional structures can be obtained.

Activity =

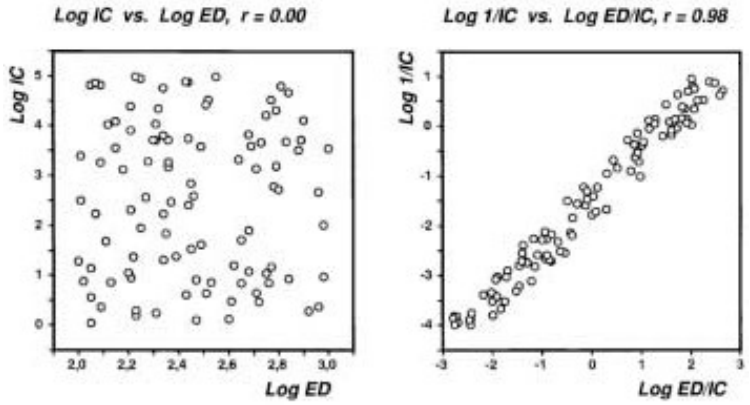
f(physicochemical and/or structural properties)

QSAR Models

- **Data selection and validation of results in QSAR models**
  - Careful evaluation of independent variables;
  - Meaning of variables;
  - Occam's razor (one should strive for a model based on the smallest number of variables);
  - The right number of compounds per variable;
  - The qualitative meaning of the model (biophysical).

Dependent / independent variables

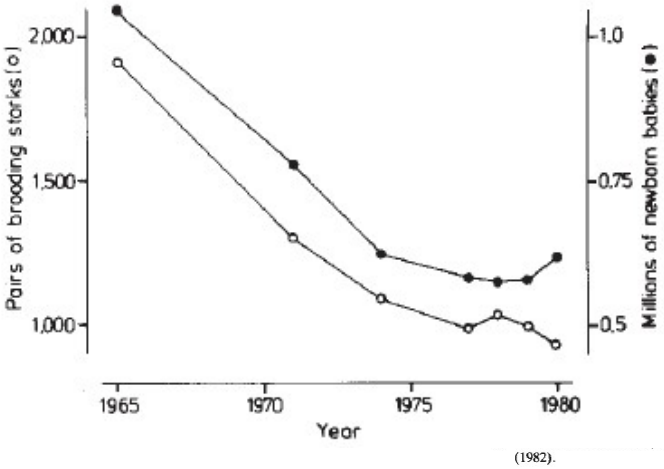
- Variables IC and ED do NOT correlate
- Variables 1/IC and ED/IC correlate but are not independent

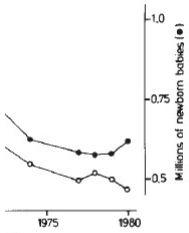


Starks!

A new parameter for sex education

SIR—There is concern in West Germany over the falling birth rate. The accompanying graph[1,2] might suggest a solution that kes sense.





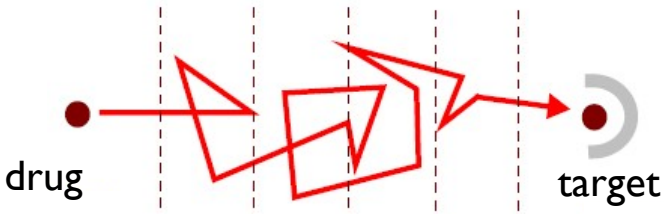
HELMUT SIES
sche Chemie I,
f,
000 Düsseldorf,

Bevölkerung (Statistisches
uttgart, 1984).
Die Vogelwarte **31**, 183–191
(1982).

Starks!

The route of the drug to the target

- Drug activity is the result of drug transport and drug interaction with the molecular target.



water and lipophilic barriers

Action of the drug

- The action of the drug depends on:
  - Lipophilicity and ionic character - transport and distribution in biological systems
  - Geometric fit and complementarity to the target surface - interaction with the binding site.
- **Which conformation is biologically active?**
  - in a vacuum
  - in the crystal
  - in a water solution
  - at the binding site

QSAR studies

- **Basic requirements in QSAR**
  - All compounds belong to one class
  - All compounds have the same mechanism of action
  - All compounds bind in a similar way
  - The effects of isosteric replacement are predictable
  - Binding is correlated with the energy of interaction
  - Biological activity correlates with binding

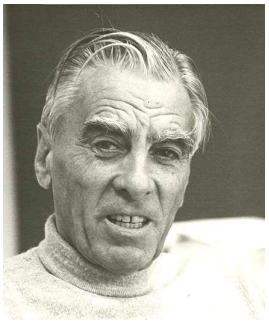
Molecular properties

Molecular properties	Oddziaływania
lipophilicity	Hydrophobic interactions
polarizability	Van der Waals interactions
Electron density	Ionic bonds, dipole-dipole interactions, hydrogen bonds, charge transfer interactions
topology	Geometric fit, steric hindrance

Hansch model

- The Hansch model was the first model of the QSAR type
- Dependence of biological properties on the properties of compounds
- Linear model

Log 1/C = a log P + b σ + c MR + ... + k



[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

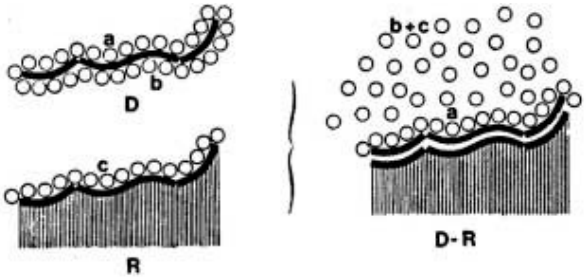
By Corwin Hansch and Toshio Fujita[1]

Received August 19, 1963

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 houseflies, thyroxine derivatives on rodents, diethylaminoethyl benzoates on guinea pigs, and carcinogenic compounds on mice.

lipophilicity

- Hydrophobic interactions are an important element of drug / receptor binding
- Lipophilicity describes the partition coefficient between water and organic (n-octanol)
- $P = c_{org} / c_{aq}$



lipophilicity

- Examples - carbamates

$R\text{-}OCONH_2$	P	Log P	Δ logP
methyl	0.22	-0.66	
ethyl	0.70	-0.15	0.51
propyl	2.30	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

lipophilicity

- The n-octanol / water system:
  - similar to membrane structures
  - available hydrogen bond donors and acceptors
  - n-octanol practically insoluble in water
  - low vapor pressure
  - transparent in the UV range
  - large database of logP values
- Additivity of π values (Hansch 1964):
  - $\pi_x = \log P_{R\text{-}X} - \log P_{R\text{-}H}$

lipophilicity

- Substituents of benzene ring

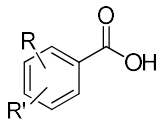
R	π_{meta}	π_{para}
H	0.0	0.0
CH_3	0.51	0.52
Br	0.76	0.70
Cl	0.94	1.02
OH	-0.49	-0.61
OCH_3	0.12	-0.04
NO_2	0.11	0.24

lipophilicity

- **Calculation of logP values for *m*-chlorotoluene :**
  - **log P = $\log P_{Benzene}$ + πCl + πMe = 2.13 + 0.71 + 0.56 = 3.40**
  - **log P = $\log P_{Toluene}$ + πmeta-Cl = 2.69 + 0.76 = 3.45**
  - **log P = $\log P_{Chlorobenzene}$ + πmeta-Me = 2.84 + 0.51 = 3.35**
- **Experimental value : $\log P_{exp}$ = 3.28**

lipophilicity

- Toxicity of substituted benzoic acids for mosquito larvae.
- -log C = 0.519 π + 1.54
- R^2 = 0.955

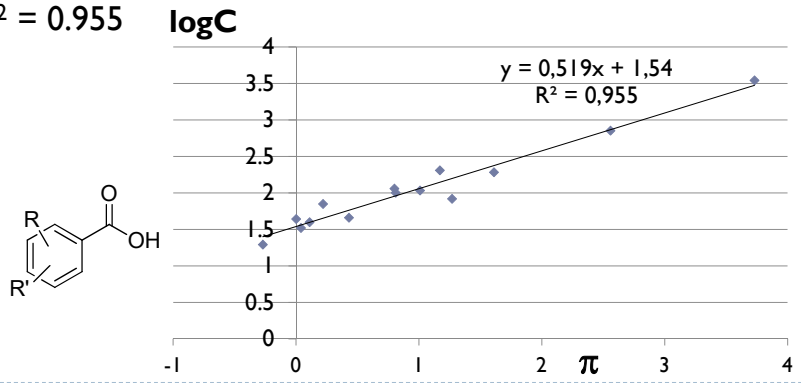


TOXICITY OF BENZOIC ACIDS TO MOSQUITO LARVAE

Functions	Σσ	Σπ[a]	−log C[b] Calcd.	−log C[b] Obsd.	Δ log C
3,4,5-Tri-I	0.98	3.73	3.476	3.540	0.064
3,5-Di-I	0.70	2.56	2.869	2.850	0.019
4-I	0.28	1.17	2.147	2.310	0.163
3,4-Di-Cl	0.60	1.61	2.376	2.280	0.096
4-Cl	0.23	0.80	1.955	2.060	0.105
4-Br	0.23	1.01	2.064	2.030	0.034
3-Cl	0.37	0.81	1.960	2.000	0.040
3,4-$(CH)_4$	0.17	1.27[c]	2.199	1.920	0.279
4-F	0.06	0.22	1.654	1.850	0.196
4-CH_3	−0.17	0.43	1.763	1.660	0.103
H	0.00	0.00	1.540	1.640	0.100
4-OCH_3	−0.27	0.11	1.597	1.600	0.003
4-NO_2	0.78	0.04	1.561	1.520	0.041
4-OH	−0.36	−0.27	1.400	1.290	0.110

lipophilicity

- Toxicity of substituted benzoic acids for mosquito larvae.
- -log C = 0.519 π + 1.54
- R^2 = 0.955



polarizability

- Parameters defining the polarizability of a molecule :
  - **Molecular volume (MV):**

$$V_m = \frac{V}{n} = \frac{M}{\rho}$$

M – molecular mass, ρ – density

  - **Molar refraction (MR):**

$$R = V_m \frac{n^2 - 1}{n^2 + 2}$$

n – refractive index

polarizability

- **Binding of neutral compounds to BSA**
  - **Log 1/C = 0.751 (±0.07) log P + 2.300**
    - **(n = 42; r = 0.960)**
  - **Log 1/C = 0.024 (±0.02) MR + 2.901**
    - **(n = 42; r = 0.307)**
- **Binding of ß-D-glicosides to Concanavalin A**
  - **Log M_{50} = 0.971 (±0.56) π + 2.37**
    - **(n = 19; r = 0.664)**
  - **Log M_{50} = 0.019 (±0.003) MR + 2.23**
    - **(n = 19; r = 0.954)**

polarizability

- **Inhibition of maleate dehydrogenase by 4-hydroxychinolinecarboxylic acids**
  - **pI_{50} = 0.70 (±0.17) MR + 2.29**

    (n = 13; r = 0.939)
- **Inhibition of cell growth by 4-hydroxychinoline carboxylic acids**
  - **pI_{50} = 0.46 (±0.11) π + 3.22**

    (n = 14; r = 0.933)
- **Binding of amino acid derivatives $R^1CONH\text{-}CH(R^2)\text{-}COOMe$ to chymotrypsin**
  - **Log 1/Km = 0.082 (±0.02) MR1 + 1.382 (±0.87) π2 – 3.876**

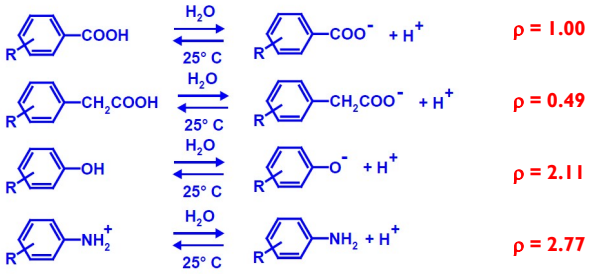
    (n = 21; r = 0.934)

Hammett's equation

- Hammett's equation describes the effect of substituents on the equilibrium of the reaction

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

- **ρ** – describes reaction (independent of the substituent)
- **σ** – describes substituent (independent of the reaction)



Coefficients of substituents - σ

- Coefficient of substituent, σ describes the electron properties of functional groups
- The influence of inductive and mesomeric effects on the reaction

$$\log K_{RX} = \text{Log } K_{RH} + \rho\sigma$$

Substituent	σ_{para}	σ_{meta}
$-N(CH_3)_2$	-0.83	-0.211
$-NH_2$	-0.66	-0.161
$-OCH_3$	-0.268	+0.115
$-OC_2H_5$	-0.25	+0.015
$-CH_3$	-0.170	-0.069
-H	0.000	0.000
-F	+0.062	+0.337
-Cl	+0.227	+0.373
-Br	+0.232	+0.393
-I	+0.276	+0.353
-CN	+0.66	+0.56
$-NO_2$	+0.778	+0.710

Hammett's equation

- It can be used to calculate constant chemical reactions for compounds with various substituents
- **3,5-Dinitro-4-methylbenzoic acid (pK_a = 2.97)**
  - **$pK_{a\ benzoic\ acid}$= 4.20**
  - **$pK_{a\ calc}$= 4.20 - (0.71 - 0.17 + 0.71) = 2.95**
- ***m*-Hydroxybenzoic acid (pK_a = 4.06 i 9.92)**
  - **$pK_{a\ benzoic\ acid}$ = 4.20, $pK_{a\ phenol}$= 9.92**
  - **ρ_{phenol} = 2.23; $\sigma_{meta\ OH}$ = 0.12;**
  - **$\sigma_{meta\ COOH}$ = 0.37; $\sigma_{meta\ COO^-}$ = -0.10)**
  - **$pK_{a\ COOH}$ = 4.20 - 0.12 = 4.08**
  - **$pK_{a\ OH}$ = 9.92 - 2.23 (-0.10) = 10.14**

Other descriptors

- The number of hydrogen bond donors
- The number of hydrogen bond acceptors
- The number of halogens
- The number of aromatic rings
- The number of ionizable groups
- Number of bonds
- Molecular weight
- Surface
- Charge
- ……

Dragon 6 program
Calculates 4885 descriptors

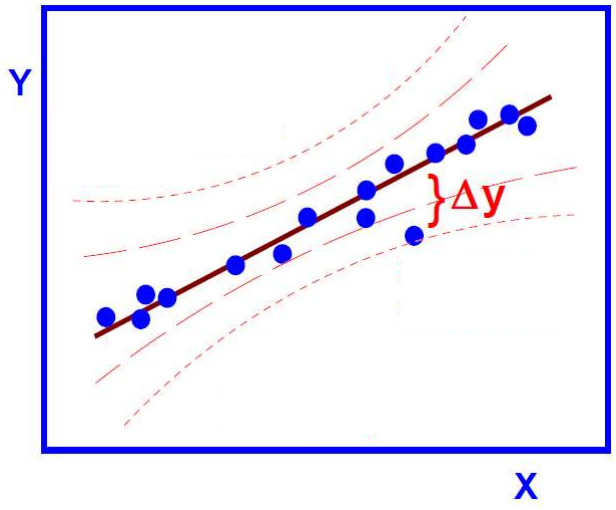
- You can calculate several thousand different descriptors

Descriptors from quantum mechanics

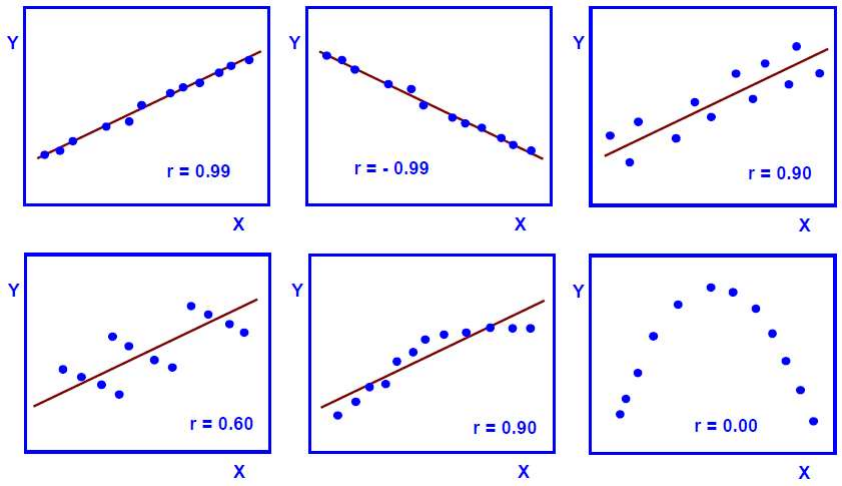
- Atomic partial charges
- The dipole moment
- HOMO/LUMO
  - **HOMO = energy of highest occupied molecular orbital, "nuclephilicity"**
  - **LUMO = energy of lowest unoccupied molecular orbital, "electrophilicity"**
- Others ….

Regression

- Regression allows you to determine the relationship between variables
- Minimization of the sum of squared deviations



Linear Regression



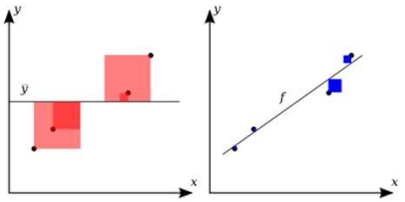
Parameters for statistical analysis

- Coefficient of determination :

$$R^2 = 1 - \frac{SS_{\text{res}}}{SS_{\text{tot}}}$$

$$SS_{\text{res}} = \sum_i (y_i - f_i)^2$$

$$SS_{\text{tot}} = \sum_i (y_i - \bar{y})^2,$$

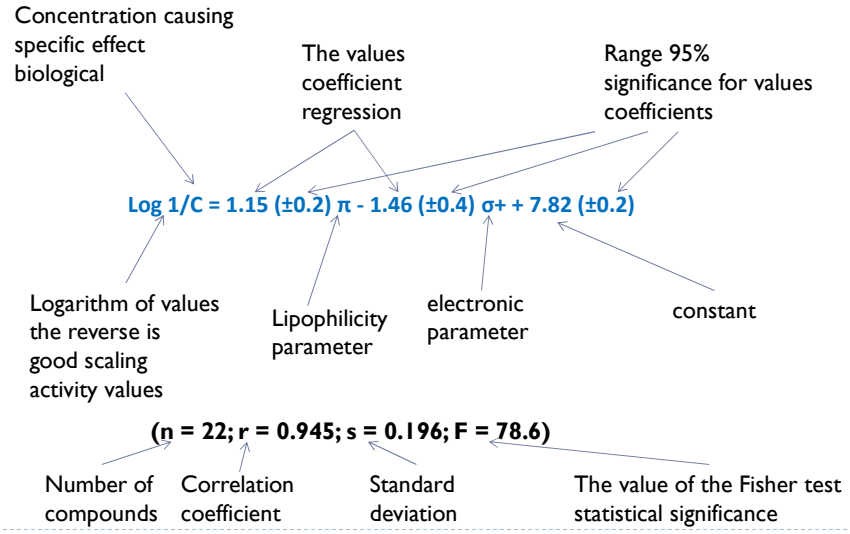


- Standard deviation :

$$s = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \overline{x})^2}{n-1}} = \sqrt{\frac{n}{n-1}(\overline{x^2} - (\overline{x})^2)}$$

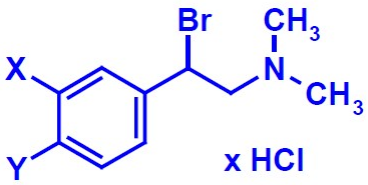
- Fischer test

QSAR equation



Example

- Anti-adrenergic activity of dimethyl-phenylethylamine



meta (X)	*para* (Y)	log 1/C obsd.	π	σ+
H	H	7.46	0.00	0.00
H	F	8.16	0.15	-0.07
H	Cl	8.68	0.70	0.11
H	Br	8.89	1.02	0.15
H	I	9.25	1.26	0.14
H	Me	9.30	0.52	-0.31
F	H	7.52	0.13	0.35
Cl	H	8.16	0.76	0.40
Br	H	8.30	0.94	0.41
I	H	8.40	1.15	0.36
Me	H	8.46	0.51	-0.07
Cl	F	8.19	0.91	0.33
Br	F	8.57	1.09	0.34
Me	F	8.82	0.66	-0.14
Cl	Cl	8.89	1.46	0.51
Br	Cl	8.92	1.64	0.52
Me	Cl	8.96	1.21	0.04
Cl	Br	9.00	1.78	0.55
Br	Br	9.35	1.96	0.56
Me	Br	9.22	1.53	0.08
Me	Me	9.30	1.03	-0.38
Br	Me	9.52	1.46	0.10

Example

- Hansch model

log 1/C = 1.151 (±0.19) π - 1.464 (±0.38) σ+ + 7.817 (±0.19)

(n = 22; r = 0.945; s = 0.196; F = 78.63)

Wilson model

- **The Wilson model is the first model of structure-activity relationship**
- **log 1/C = Σ a_i + μ**
  - **a_i = contributions of substituents**
  - **μ = contribution of the reference compound activity**
- **Mixed model - Hansch/Wilson**
  - **log 1/C = a (log P)^2 + b log P + c σ + ... + Σ a_i + k**
  - **log 1/C = a log P - b log (ßP + 1) + c σ + ... + Σ a_i + k**

Example

- Wilson model
- Data showing substituent positioning.

meta (X)	*para* (Y)	log 1/C obs.	*meta*- F	Cl	Br	I	Me	*para*- F	Cl	Br	I	Me
H	H	7.46										
H	F	8.16						1				
H	Cl	8.68							1			
H	Br	8.89								1		
H	I	9.25									1	
H	Me	9.30										1
F	H	7.52	1									
Cl	H	8.16		1								
Br	H	8.30			1							
I	H	8.40				1						
Me	H	8.46					1					
Cl	F	8.19		1				1				
Br	F	8.57			1			1				
Me	F	8.82					1	1				
Cl	Cl	8.89		1					1			
Br	Cl	8.92			1				1			
Me	Cl	8.96					1		1			
Cl	Br	9.00		1						1		
Br	Br	9.35			1					1		
Me	Br	9.22					1			1		
Me	Me	9.30					1					1
Br	Me	9.52			1							1

Example

- Wilson model
- **log 1/C = - 0.301 (±0.50) [m-F] + 0.207 (±0.29) [m-Cl]**
  + 0.434 (±0.27) [m-Br] + 0.579 (±0.50) [m-I]
  + 0.454 (±0.27) [m-Me] + 0.340 (±0.30) [p-F]
  + 0.768 (±0.30) [p-Cl] + 1.020 (±0.30) [p-Br]
  + 1.429 (±0.50) [p-I] + 1.256 (±0.33) [p-Me]
  + 7.821 (±0.27)

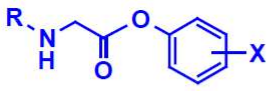
Position	H	F	Cl	Br	I	Me
meta	0.00	-0.30	0.21	0.43	0.58	0.45
para	0.00	0.34	0.77	1.02	1.43	1.26

μ = 7.82

(n = 22; r = 0.969; s = 0.194; F = 16.99)

Example 2

- Inhibition of papain by glycine esters



X	R	log 1/K_m	π	MR	σ	I
4-NH_2	-COC_6H_5	3.58	-1.23	0.54	-0.66	0
4-Me	-COC_6H_5	4.02	0.56	0.56	-0.17	0
H	-COC_6H_5	3.77	0.00	0.10	0.00	0
4-Cl	-COC_6H_5	4.00	0.71	0.60	0.23	0
4-F	-COC_6H_5	3.69	0.14	0.09	0.06	0
3-NO_2	-COC_6H_5	4.74	-0.28	0.74	0.71	0
4-NO_2	-COC_6H_5	4.85	-0.28	0.74	0.78	0

Example 2

X	R	log 1/K_m	π	MR	σ	I
4-OH	-SO_2Me	2.05	-0.67	0.28	-0.37	1
4-OMe	-SO_2Me	2.13	-0.02	0.79	-0.27	1
4-Me	-SO_2Me	2.08	0.56	0.56	-0.17	1
3-Me	-SO_2Me	2.23	0.56	0.56	-0.07	1
H	-SO_2Me	1.79	0.00	0.10	0.00	1
4-F	-SO_2Me	1.95	0.14	0.09	0.06	1
3-OMe	-SO_2Me	2.29	-0.02	0.79	0.12	1
4-CHO	-SO_2Me	2.33	-0.65	0.69	0.42	1
4-Cl	-SO_2Me	2.38	0.71	0.60	0.23	1
3-F	-SO_2Me	1.98	0.14	0.09	0.34	1
4-COMe	-SO_2Me	2.57	-0.55	1.12	0.50	1
3-NO_2	-SO_2Me	2.53	-0.28	0.74	0.71	1
4-NO_2	-SO_2Me	2.71	-0.28	0.74	0.78	1

Example 2

- **Benzamides (n = 7)**
  - π — r = 0.038; s = 0.554; F = 0.01
  - MR — r = 0.714; s = 0.388; F = 5.19
  - σ — r = 0.892; s = 0.251; F = 19.41
  - π, MR — r = 0.726; s = 0.426; F = 2.23
  - π, σ — r = 0.916; s = 0.248; F = 10.46
  - MR, σ — r = 0.971; s = 0.148; F = 32.85

    $\log 1/K_m = 0.771\ (\pm 0.67)\ MR + 0.728\ (\pm 0.37)\ \sigma + 3.623\ (\pm 0.34)$

    (n = 7; r = 0.971; s = 0.148; F = 32.85)
- **Mesylamides (n = 13)**
  - π — r = 0.271; s = 0.271; F = 0.87
  - MR — r = 0.813; s = 0.164; F = 21.45
  - σ — r = 0.730; s = 0.192; F = 12.52
  - π (n.s.), MR — r = 0.817; s = 0.170; F = 10.02
  - π (n.s.), σ — r = 0.732; s = 0.201; F = 5.79
  - MR, σ — r = 0.935; s = 0.105; F = 34.51

    $\log 1/K_m = 0.529\ (\pm 0.23)\ MR + 0.370\ (\pm 0.20)\ \sigma + 1.877\ (\pm 0.13)$

    (n = 13; r = 0.935; s = 0.105; F = 34.51)

Example 2

- **All compounds (n = 20)**
  - MR — r = 0.154; s = 0.992; F = 0.44
  - σ — r = 0.250; s = 0.972; F = 1.20
  - I — r = 0.931; s = 0.366; F = 117.79
  - MR, σ — r = 0.259; s = 0.998; F = 0.61
  - MR, I — r = 0.967; s = 0.263; F = 122.35
  - σ, I — r = 0.977; s = 0.220; F = 179.70
  - MR, σ, I — r = 0.990; s = 0.148; F = 272.04

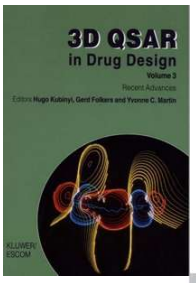
$\log 1/K_m = 0.569\ (\pm 0.26)\ MR + 0.561\ (\pm 0.19)\ \sigma$

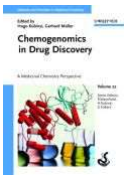
$- 1.922(\pm 0.15)\ I + 3.743\ (\pm 0.17)$

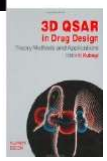
(n = 20; r = 0.990; s = 0.148; F = 272.04)

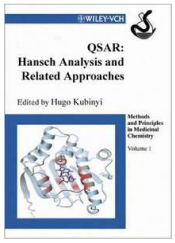
Hugo Kubinyi

- www.kubinyi.de









Summary

- QSAR models are the simplest quantitative estimation of the biological activity of compounds
- QSAR models have many limitations.