QSAR - Hansch and Free Wilson Analyses

Hugo Kubinyi

Germany

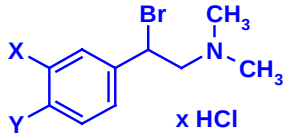
E-Mail kubinyi@t-online.de
HomePage www.kubinyi.de

Robin Ganellin

Professor of Medicinal Chemistry University College, London, UK

S. L. Carney (DDT 9, 158-160 (2004)):
Has there been a single development that, in your opinion, has moved the field of medicinal chemistry ahead more than any other?
Robert Ganellin:
I would go back to the 1960s to the work of Corwin Hansch on the importance of lipophilicity. ... I think that changed the way of thinking in medicinal chemistry. .... I think that the application of physical organic chemical approaches to structure–activity analysis have been very important.

Antiadrenergic Activities of *meta*-, *para*-, and *meta*,*para*-Disubstituted *N*,*N*-Dimethyl-α-bromophenethylamines

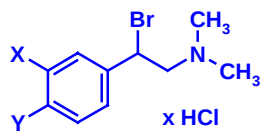


meta	para	log 1/C	meta	para	log 1/C
H	H	7.46	Cl	F	8.19
H	F	8.16	Br	F	8.57
H	Cl	8.68	Me	F	8.82
H	Br	8.89	Cl	Cl	8.89
H	I	9.25	Br	Cl	8.92
H	Me	9.30	Me	Cl	8.96
F	H	7.52	Cl	Br	9.00
Cl	H	8.16	Br	Br	9.35
Br	H	8.30	Me	Br	9.22
I	H	8.40	Me	Me	9.30
Me	H	8.46	Br	Me	9.52

Hugo Kubinyi, www.kubinyi.de

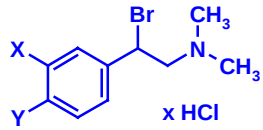
meta (X)	*para* (Y)	log 1/C obsd.	π	σ^+	E_s^{meta}	log 1/C calc.	log 1/C calc.
H	H	7.46	0.00	0.00	1.24	7.82	7.88
H	F	8.16	0.15	-0.07	1.24	8.09	8.17
H	Cl	8.68	0.70	0.11	1.24	8.46	8.60
H	Br	8.89	1.02	0.15	1.24	8.77	8.94
H	I	9.25	1.26	0.14	1.24	9.06	9.26
H	Me	9.30	0.52	-0.31	1.24	8.87	8.98
F	H	7.52	0.13	0.35	0.78	7.45	7.43
Cl	H	8.16	0.76	0.40	0.27	8.11	8.05
Br	H	8.30	0.94	0.41	0.08	8.30	8.22
I	H	8.40	1.15	0.36	-0.16	8.61	8.51
Me	H	8.46	0.51	-0.07	0.00	8.51	8.36
Cl	F	8.19	0.91	0.33	0.27	8.38	8.34
Br	F	8.57	1.09	0.34	0.08	8.57	8.51
Me	F	8.82	0.66	-0.14	0.00	8.78	8.65
Cl	Cl	8.89	1.46	0.51	0.27	8.75	8.77
Br	Cl	8.92	1.64	0.52	0.08	8.94	8.94
Me	Cl	8.96	1.21	0.04	0.00	9.15	9.08
Cl	Br	9.00	1.78	0.55	0.27	9.06	9.11
Br	Br	9.35	1.96	0.56	0.08	9.25	9.29
Me	Br	9.22	1.53	0.08	0.00	9.46	9.43
Me	Me	9.30	1.03	-0.38	0.00	9.56	9.47
Br	Me	9.52	1.46	0.10	0.08	9.35	9.33

Matrix for Hansch Analysis



meta (X)	*para* (Y)	log 1/C obs.	*meta-* F	Cl	Br	I	Me	*para-* F	Cl	Br	I	Me	log 1/C calc.
H	H	7.46											7.82
H	F	8.16						1					8.16
H	Cl	8.68							1				8.59
H	Br	8.89								1			8.84
H	I	9.25									1		9.25
H	Me	9.30										1	9.08
F	H	7.52	1										7.52
Cl	H	8.16		1									8.03
Br	H	8.30			1								8.26
I	H	8.40				1							8.40
Me	H	8.46					1						8.28
Cl	F	8.19		1				1					8.37
Br	F	8.57			1			1					8.60
Me	F	8.82					1	1					8.62
Cl	Cl	8.89		1					1				8.80
Br	Cl	8.92			1				1				9.02
Me	Cl	8.96					1		1				9.04
Cl	Br	9.00		1						1			9.05
Br	Br	9.35			1					1			9.28
Me	Br	9.22					1			1			9.30
Me	Me	9.30					1					1	9.53
Br	Me	9.52			1							1	9.51

Matrix for Free Wilson Analysis



Hugo Kubinyi, www.kubinyi.de

Free Wilson Analysis, Results: $\mu = 7.82$

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

(n = 22; r = 0.97; s = 0.19)

Hansch Analyses, Results:

C. Hansch and E. J. Lien, Biochem. Pharmacol. 17, 709 (1968)

$\log 1/C = 1.221\ \pi - 1.587\ \sigma + 7.888$

(n = 22; r = 0.918; s = 0.238)

A. Cammarata, J. Med. Chem. 15, 573 (1972)

$\log 1/C = 0.747\ (\pm 0.12)\ \pi_m - 0.911\ (\pm 0.25)\ \sigma_m + 1.666\ (\pm 0.12)\ r_v^{para} + 5.769$

(n = 22; r = 0.961; s = 0.164)

Hansch Equations

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)
log 1/C = 1.259 (±0.19) π - 1.460 (±0.34) σ^+
+ 0.208 (±0.17) E_s^{meta} + 7.619 (±0.24)
(n = 22; r = 0.959; s = 0.173; F = 69.24)

Free Wilson Equation

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

Hugo Kubinyi, www.kubinyi.de

Regression Analysis

Mathematically exact procedure
for the treatment of data with experimental errors
(cf. mean value, standard devation).
Minimization of the **sum of squared errors** (= squared deviations between y_i and y_{calc}) produces the **best fit** of the observed values to a certain model.

Regression analysis describes the relationship between:
- **independent variables x_i** (definition: can be determined without experimental error), and
- **dependent variables y_i** (contain experimental error).

Hypothesis: there is a significant relationship (95% level) between x_i and y_i values: yes / no
F test for overall significance
t tests for individual significances in multiple regression.

Formulas and Meaning of Statistical Parameters

Correlation coefficient r (relative quality of fit)

$$r^2 = 1 - \Sigma\Delta^2/S_{yy}$$

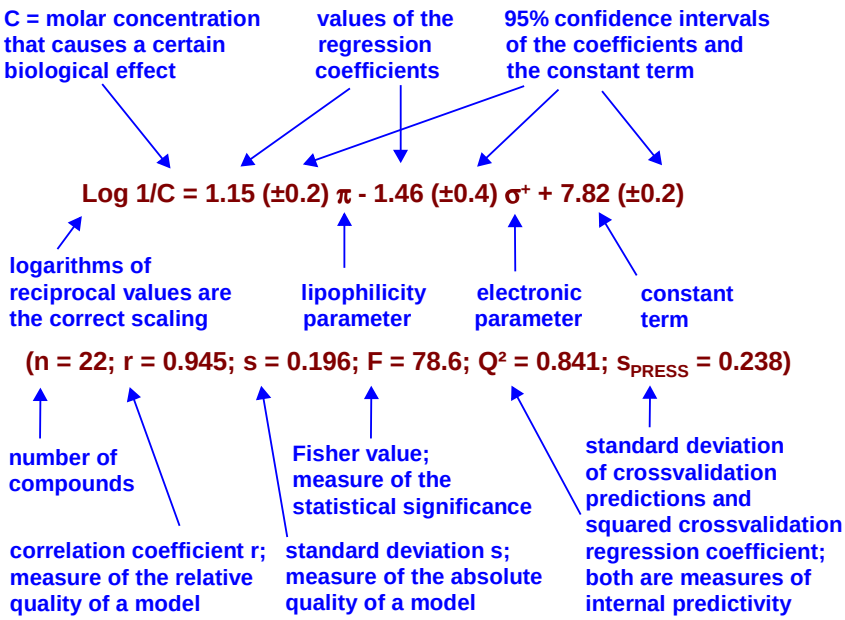
Standard deviation s (absolute quality of fit)

$$s^2 = \Sigma\Delta^2/(n - k - 1)$$

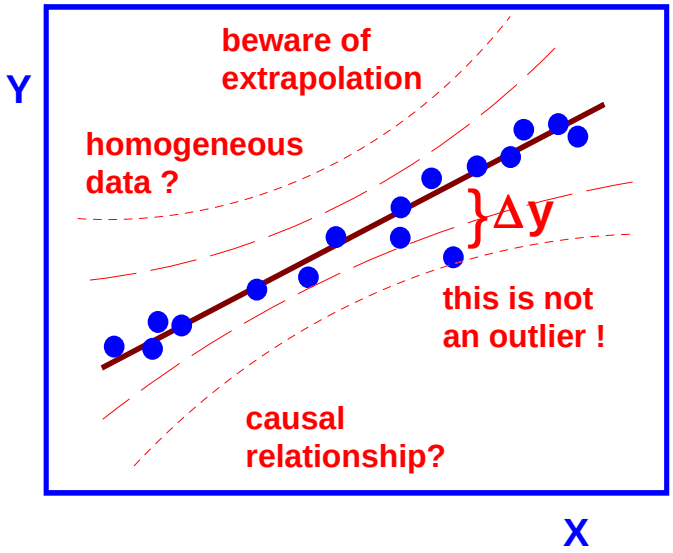
F test (Fisher value; level of statistical significance)

$$F = r^2.(n - k - 1)/(k.(1 - r^2))$$

Confidence intervals of k_i $\pm s.t.\sqrt{c_{ii}}$



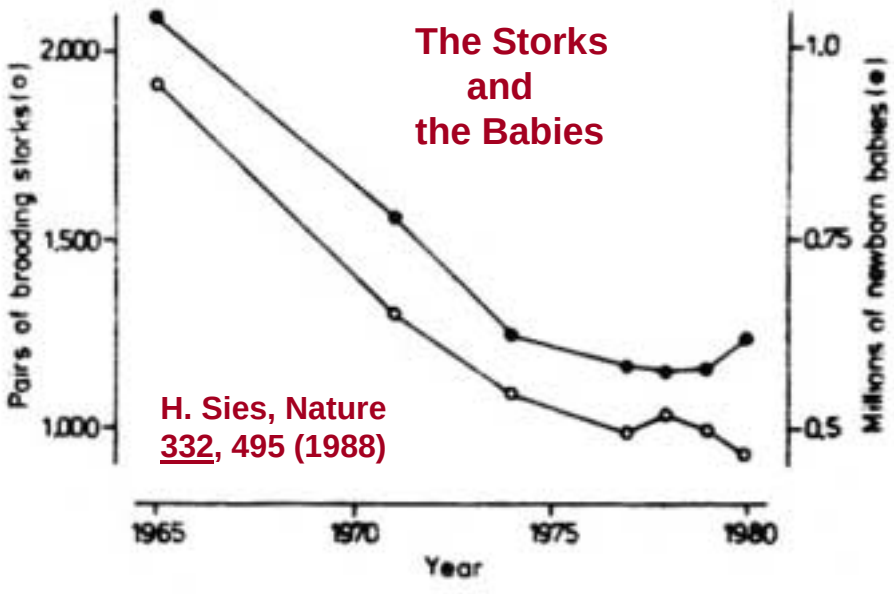
Regression Analysis - A Common QSAR Tool



regression line

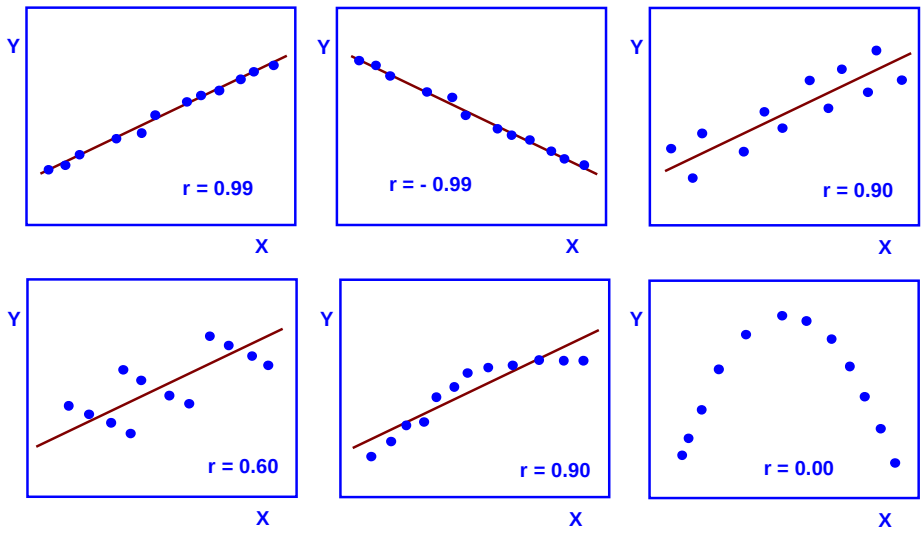
95% confidence intervals of the regression line

95% confidende intervals for a new observation



H. Sies, Nature 332, 495 (1988)

Most Important is a Graphical Analysis of the Data !



A Diagram Tells You More Than Thousand Equations

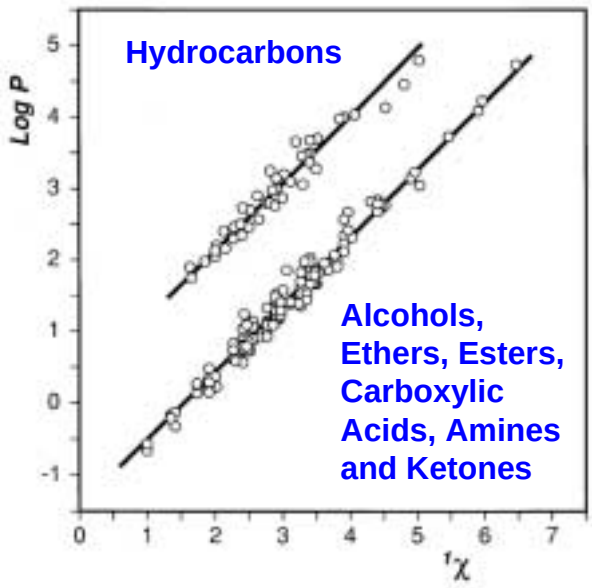
Anil K. Saxena, Quant. Struct.-Act. Relat. 14, 142-150 (1995)

183 Hydrocarbons, Alcohols, Ethers, Esters, Carboxylic Acids, Amines and Ketones

MR *vs.* $^{1}\chi$	r = 0.908; s = 0.380; F = 855.26
MR *vs.* $^{2}\chi^{v}$	r = 0.826; s = 0.419; F = 389.58
log P *vs.* $^{1}\chi$	r = 0.719; s = 0.632; F = 193.36
log P *vs.* $^{2}\chi^{v}$	r = 0.635; s = 0.574; F = 122.33

$$\log P = 0.941\ (\pm 0.02)\ {}^{1}\chi - 1.693\ (\pm 0.05)\ I + 0.244\ (\pm 0.08)$$

(n = 183; r = 0.990; s = 0.150; F = 4,633)



Log P vs. $^1\chi$ and I

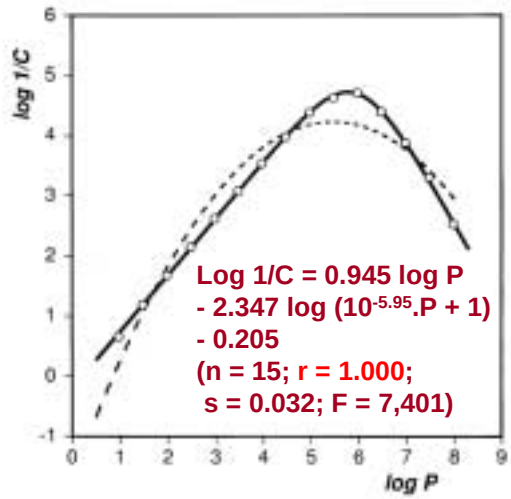
n = 183
r = 0.990
s = 0.150
F = 4,633

Antibacterial Activity of Homologous Aliphatic Amines vs. *Rhinocladium beurmanni*

(E. J. Lien and P. H. Wang, *J. Pharm. Sci.* 69, 648-650 (1980))

Parabolic Model, log P
r = 0.967;
s = 0.354;
F = 85.61

Parabolic Model and log MW term
r = 0.995;
s = 0.148;
F = 345.1



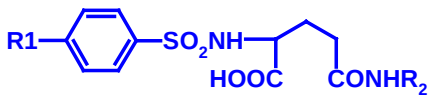
Statistical unicorns, „beasts that exist on paper but not in reality“ (C. Hansch, 1973):

sinus functions
hyperbolic functions
tangens functions

(for references cf. H. Kubinyi, QSAR: Hansch Analysis and Related Approaches, VCH, Weinheim, 1993)

Antitumor Activity of Glutamic Acid Amides

T. Jha et al., Ind. J. Chem. 25B, 169-174 (1986)



$$\log \%\,ITW = 1.4111\,(\log P)^2 - 0.5971 \log P - 0.1714\,\pi^2_{Ali} - 3.2293\,\sigma_{Ali} + 0.9595\,E_{s\,Ali} - 6.6199\,\sigma_i + 1.3249$$

(n = 8; r = 0.9912; s = 0.0310)

http://www.epa.gov/oppt/newchems/chemcat.htm

log fish 96-h LC50 (moles/L) = -0.449 log Kow - 3.314
n=54, R^2=0.53 CLOGP<6, MW<1000

log fish 96-h LC50 (millimoles/L) = -1.176 - 0.371 log Kow (CLOGP)
n=2, R^2=1.0 CLOGP<6, MW<1000

log fish 96-h LC50 (millimoles/L) = -1.021 - 0.396 log Kow (SRC)
n=2, R^2=1.0 SRC<6, MW<1000

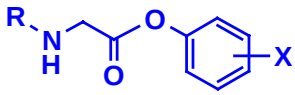
log daphnid ChV (millimoles/L) = -1.798 - 0.488 log Kow (CLOGP)
n=2, R^2=1.0 CLOGP<8, MW<1000

log algal ChV (millimoles/L) = -2.313 - 0.348 log Kow (CLOGP)
n=2, R^2=1.0 CLOGP<8, MW<1000

log fish 96-h LC50 (mmoles/L) = -3.502 + 0.017 log Kow (CLOGP)
n=3, R^2=0.25 CLOGP<6, MW<1000

Binding of Glycine Ester Amides to Papain

C. Hansch and D. F. Calef, J. Org. Chem. 41, 1240 (1976)
C. Hansch et al., Arch. Biochem. Biophys. 183, 383 (1977)



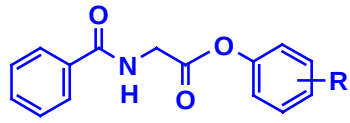
X	R	log 1/K_m	π	MR	σ	I
4-NH_2	-COC_6H_5	3.58	-1.23	0.54	-0.66	0
4-Me	-COC_6H_5	4.02	0.56	0.56	-0.17	0
H	-COC_6H_5	3.77	0.00	0.10	0.00	0
4-Cl	-COC_6H_5	4.00	0.71	0.60	0.23	0
4-F	-COC_6H_5	3.69	0.14	0.09	0.06	0
3-NO_2	-COC_6H_5	4.74	-0.28	0.74	0.71	0
4-NO_2	-COC_6H_5	4.85	-0.28	0.74	0.78	0

Hugo Kubinyi, www.kubinyi.de

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

(MR values scaled by a factor 0.1)

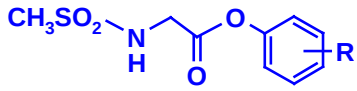
Benzamides (n = 7)



π (n.s.) 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 (n.s.), MR (n.s.) r = 0.726; s = 0.426; F = 2.23
π (n.s.), 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 (n.s.) 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

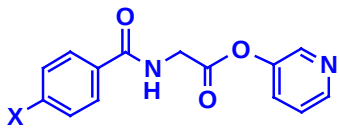
$\log 1/K_m = 0.529\ (\pm 0.23)\ MR + 0.370\ (\pm 0.20)\ \sigma + 1.877\ (\pm 0.13)$
(n = 13; r = 0.935; s = 0.105; F = 34.51)

Hugo Kubinyi, www.kubinyi.de

All Compounds (n = 20)

MR (n.s.) r = 0.154; s = 0.992; F = 0.44
σ (n.s.) r = 0.250; s = 0.972; F = 1.20
I r = 0.931; s = 0.366; F = 117.79
MR (n.s.), σ (n.s.) 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

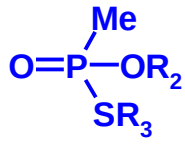
$\log 1/K_m = 0.569\ (\pm 0.26)\ MR + 0.561\ (\pm 0.19)\ \sigma - 1.922\ (\pm 0.15)\ I + 3.743\ (\pm 0.17)$
(n = 20; r = 0.990; s = 0.148; F = 272.04)



$\log 1/K_m = 1.01\ (\pm 0.11)\ \pi + 1.46$
(n = 16; r = 0.981; s = 0.165)

Inhibition of Chymotrypsin by Thiophosphonates

C. Silipo et al., Arch. Biochem. Biophys. 194, 552 (1979)

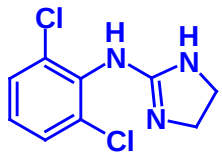


$\log K_i = 1.60\ (\pm 0.22)\ MR_2$
$- 3.85\ (\pm 1.17) \log (\beta \cdot 10^{MR2} + 1)$
$- 4.76\ (\pm 0.51)$
$\log \beta = - 3.86$ optimum $MR_2 = 3.72$
$(n = 19;\ r = 0.978;\ s = 0.258)$

$\log K_i = 1.47\ (\pm 0.10)\ MR_2$
$- 3.43\ (\pm 0.74) \log (\beta \cdot 10^{MR2} + 1)$
$+ 0.34\ (\pm 0.09)\ MR_3 + 1.25\ (\pm 0.19)\ \sigma^*_3$
$- 1.06\ (\pm 0.31)\ I - 5.26\ (\pm 0.38)$
$\log \beta = - 3.85$ optimum $MR_2 = 3.71$
$(n = 53;\ r = 0.985;\ s = 0.243)$

Central and Peripheral Activities of Clonidine Analogs

P. B. M. W. M. Timmermans et al., J. Med. Chem. 24, 502-507 (1981); J. Med. Chem. 27, 495 - 503 (1984)



$\log P_{app}$ *n*-octanol / buffer pH 7.4

C_{25} central antihypertensive activity (molar dose that causes a 25% blood pressure decrease; anesthetized rat, i.v. application)

C_{60} peripheral hypertensive activity (molar dose that causes a 60 mm blood pressure increase; pithed rat, i.v. application)

$IC_{50}\alpha_1$ binding affinity to α_1 receptors (replacement of prazosin)

$IC_{50}\alpha_2$ binding affinity to α_2 receptors (replacement of clonidine)

Peripheral Activity: Blood Pressure Increase

$pC_{60} = 1.16\ (\pm 0.21)\ pIC_{50}\alpha_2 - 0.96$

$(n = 21;\ r = 0.936;\ s = 0.317)$

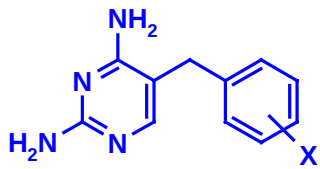
Central Activity: Blood Pressure Decrease

$pC_{25} = 1.07\ (\pm 0.20)\ pIC_{50}\alpha_2 + 0.81\ (\pm 0.22)\ \log P$
$- 3.37\ (\pm 1.02)\ \log (\beta P + 1) - 1.16$
$\log \beta = -1.99 \qquad \log P_{opt} = 1.48$
$(n = 21;\ r = 0.971;\ s = 0.284)$

Peripheral vs. Central Action on Blood Pressure

$pC_{25} = 0.83\ (\pm 0.20)\ pC_{60} + 0.78\ (\pm 0.26)\ \log P$
$- 3.69\ (\pm 1.39)\ \log (\beta P + 1) - 0.19$
$\log \beta = -2.08 \qquad \log P_{opt} = 1.51$
$(n = 21;\ r = 0.954;\ s = 0.354)$

Binding of Benzylpyrimidines to DHFR



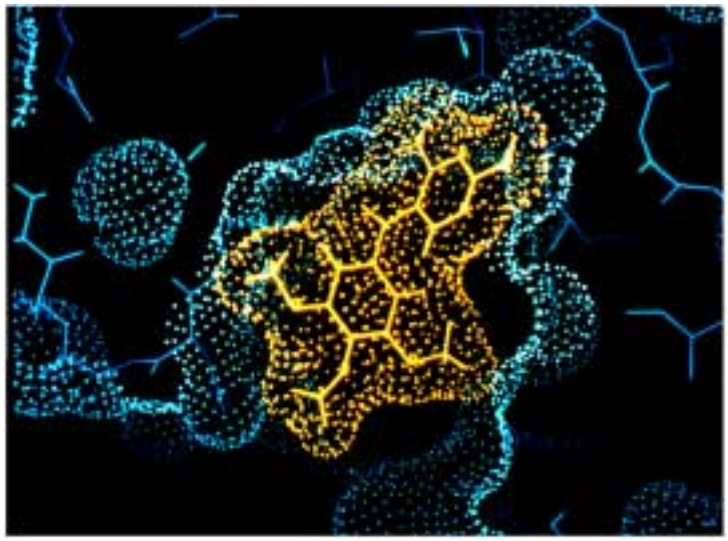
Escherichia coli DHFR

$\log 1/K_{i\ app} = 0.75\ (\pm 0.26)\ \pi_{3,4,5}$
$- 1.07\ (\pm 0.34)\ \log (\beta.10^{\pi\ 3,4,5} + 1)$
$+ 1.36\ (\pm 0.24)\ MR'_{3,5} + 0.88\ (\pm 0.29)\ MR'_4 + 6.20$
$\log \beta = 0.12 \qquad \text{optimum } \pi = 0.25$
$(n = 43;\ r = 0.903;\ s = 0.290)$

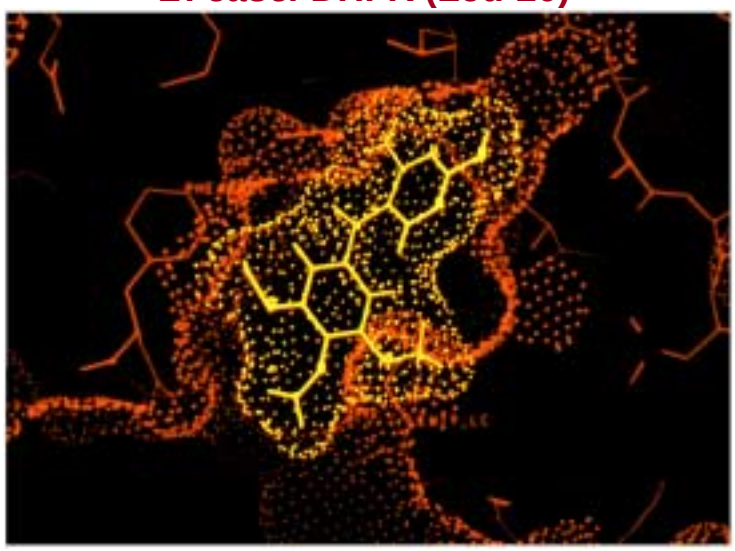
Lactobacillus casei DHFR

$\log 1/K_{i\ app} = 0.31\ (\pm 0.11)\ \pi_{3,4} - 0.88\ (\pm 0.24)\ \log (\beta.10^{\pi 3.4} + 1)$
$+ 0.95\ (\pm 0.21)\ MR'_{3,4} + 5.32$
$\log \beta = -1.33 \qquad \text{optimum } \pi = 1.05$
$(n = 42;\ r = 0.876;\ s = 0.222)$

Binding Mode of Trimethoprim to *E. coli* DHFR (Met-20)

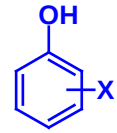


Binding Mode of Trimethoprim to *L. casei* DHFR (Leu-20)



Free Wilson Analyses

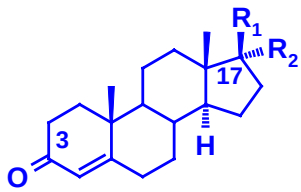
Antibacterial activity of phenols vs. *Staphylococcus aureus*

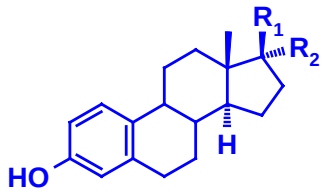


X = H, Cl

log 1/C = 0.503 (±0.13) [Cl] + 2.578
(n = 9; r = 0.960; s = 0.256; F = 83.06)

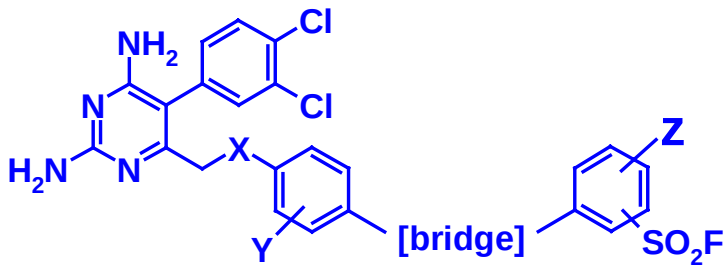
Corticosteroid-binding globulin affinities of steroids





log 1/CBG = 2.022 (±0.52) [4.5 >C=C<] + 5.186 (±0.36)
(n = 21; r = 0.882; s = 0.568; F = 66.41;
Q^2 = 0.726; s_{PRESS} = 0.630)

DHFR Inhibition by 2,4-Diaminopyrimidines



log 1/C = 0.365 (±0.12) I-1 + 1.013 (±0.12) I-8
- 0.784 (±0.19) I-9 + 0.419 (±0.20) I-13
- 0.220 (±0.09) I-15 + 0.513 (±0.18) I-20
+ 0.674 (±0.23) I-4·I-8 + 7.174 (±0.07)
(n = 105; r = 0.903; s = 0.229)

Quaternary Acetylcholine Receptor Ligands

X in X-CH_2CH_2-N^+($R^1R^2R^3$)	a_i
CH_3CH_2O-	-2.479
$CH_3CH_2CH_2$-	-2.175
$C_6H_5CH_2COO$-	-1.228
$C_6H_5CH_2CH_2O$-	-1.177
$C_6H_5CH_2CH_2CH_2$-	-0.909
$C_6H_{11}CH_2COO$-	-1.035
$C_6H_{11}CH_2CH_2O$-	-0.819
$C_6H_{11}CH_2CH_2CH_2$-	-0.683
$(C_6H_5)_2CHCOO$-	0.872
$(C_6H_5)_2CHCH_2O$-	-0.070
$(C6H5)_2CHCH_2CH_2$-	0.374
$C_6H_5(C_6H_{11})CHCOO$-	2.035
$(C_6H_5)_2C(OH)COO$-	2.047
$(C_6H_{11})_2CHCOO$-	1.467
$(C_6H_{11})_2CHCH_2O$-	0.806
$C_6H_5(C_6H_{11})C(OH)COO$-	2.975

$\mu = 6.499$

Hugo Kubinyi, www.kubinyi.de

Nonadditivity of Receptor Binding Affinities

Quaternary ammonium compounds, postganglionic acetylcholine receptor (R^1, R^2, R^3 = alkyl, cycloalkyl)

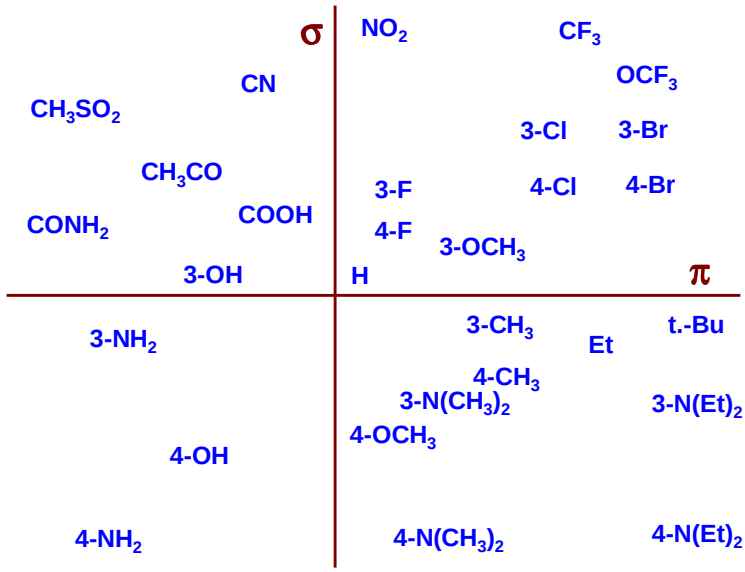
$\mu = 6.499$ (n = 128; r = 0.991; s = 0.231)

Differences in receptor affinity (logarithmic scale), by changing			
R in group X from:	H to C_6H_5	H to C_6H_{11}	C_6H_5 to C_6H_{11}
R-$CH_2CH_2CH_2$-	1.266	1.492	0.226
R-CH_2CH_2O-	1.302	1.660	0.358
$C_6H_5CH(R)COO$-	2.100	3.263	1.163
$C_6H_{11}CH(R)COO$-	3.070	2.502	-0.568

$$\log K = 1.258\ (\pm 0.09)\ [PHE] + 1.545\ (\pm 0.11)\ [c\text{-}HEX] + 1.069\ (\pm 0.19)\ I_{OH} + 0.755\ (\pm 0.17)\ I_{COO} + 0.769\ (\pm 0.19)\ [INT] + 4.142\ (\pm 0.12)$$

(n = 128; r = 0.983; s = 0.290; F = 711.14)

Craig Diagram (P. Craig, J. Med. Chem. 14, 680 (1971))



TOPLISS Scheme

J. Topliss, J. Med. Chem. 15, 1007 (1972)

Start with substituents H and 4-Cl

a) if 4-Cl > H then 3,4-Cl_2

- if 3,4-Cl_2 > 4-Cl then 3,4-Br_2, I, CF_3
- if 3,4-Cl_2 < 4-Cl then Br, I, 4-NO_2

b) if 4-Cl < H then 4-OMe

- if 4-OMe > H then 4-NMe_2
  - if 4-NMe_2 > 4-OMe then 4-NEt_2
  - if 4-NMe_2 < 4-OMe then NH_2, OH
- if 4-OMe < H then 3-Cl

c) if 4-Cl ≈ H then 4-Me

- if 4-Me > H, 4-Cl then Et, Pr, *i*-Pr, Bu, *t*-Bu
- if 4-Me < H, 4-Cl then 3-Cl
  - if 3-Cl > 4-Me then 3,5-Cl_2, 3-Br, 3-I
  - if 3-Cl < 4-Me then 3-CH_3, 3-NMe_2, 4-F, 4-NO_2, 4-CN, 4-$CONH_2$