

Projektowanie związków biologicznie czynnych **wykład 10**

Łukasz Berlicki

Od eksperymentu do teorii

- ▶ **1964, C. Hansch i T. Fujita:** QSAR
 - ▶ **1984, P. Andrews:** wkład grup funkcyjnych do oddziaływania z celem molekularnym
 - ▶ **1985, P. Goodford:** GRID (punkty wiążące na powierzchni białka)
 - ▶ **1988, R. Cramer:** 3D QSAR
 - ▶ **1992, H.-J. Böhm:** LUDI miejsca oddziaływanie, dokowanie, funkcje oceniające
 - ▶ **1997, C. Lipinski:** biodostępność, zasada pięciu
 - ▶ **1998, Ajay, W. P. Walters and M. A. Murcko; J. Sadowski i H. Kubinyi:** związki drug-like
-



Statystyka

▶ Lies, damned lies, and statistics (Benjamin Disraeli)

▶ Statystyczna analiza danych (Bailar, Clin. Pharmacol. Therapeutics, 1979)

- ▶ Nie ma jedynie słusznych odpowiedzi.
- ▶ Statystyka nie jest jedyną drogą do mądrości.
- ▶ Rzadkie przypadki zdarzają się ciągle.
- ▶ Żadna próba nie jest wystarczająco duża (i co z tego?).
- ▶ Żadna analiza nie jest doskonała (i co z tego?).
- ▶ Coś jest zawsze źle z danymi.

▶ Wszystkie modele są złe, ale niektóre są użyteczne

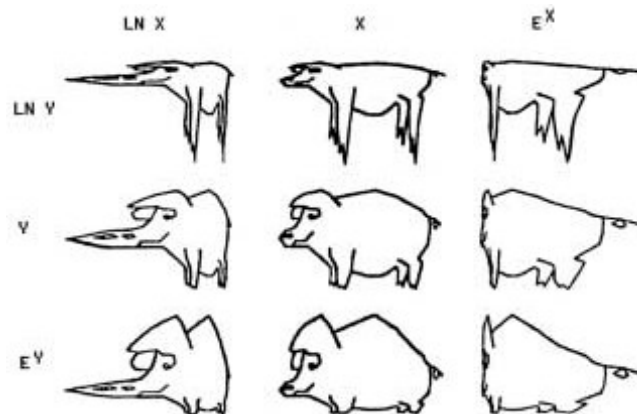
- ▶ Skalowanie zmiennych zmienia wynik
- ▶ Wykres mówi więcej niż równania
- ▶ Walidacja – bardzo ważny ale trudny problem.



Benjamin Disraeli
Premier Wielkiej Brytanii
XIX w.



Skalowanie zmiennych



QSAR

- ▶ **QSAR** – (*quantitative structure-activity relationships*) **metody statystyczne** za pomocą których można otrzymać zależność aktywności biologicznej od elementów strukturalnych, właściwości fizykochemicznych lub struktur trójwymiarowych.

Aktywność =

f(właściwości fizykochemiczne i/lub strukturalne)



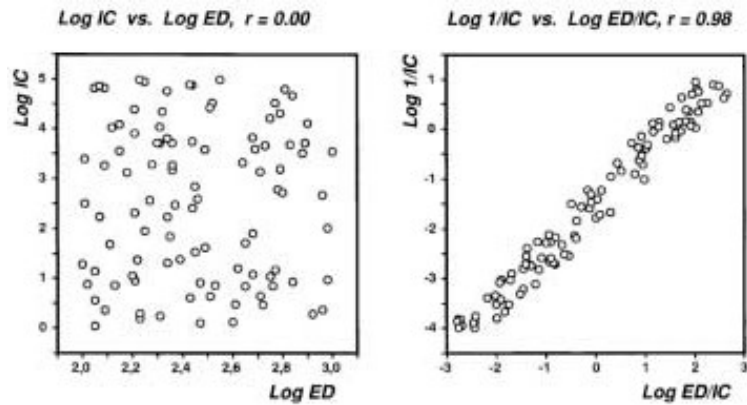
Modele QSAR

- ▶ **Selekcja danych i walidacja wyników w modelach QSAR**
 - ▶ Uważna ocena zmiennych niezależnych;
 - ▶ Znaczenie zmiennych;
 - ▶ Brzytwa Ockhama (należy dążyć do modelu opierającego się na jak najmniejszej liczbie zmiennych);
 - ▶ Odpowiednia liczba związków na zmienną;
 - ▶ Jakościowe znaczenie modelu (biofizyczne).
-



Zmienne zależne/niezależne

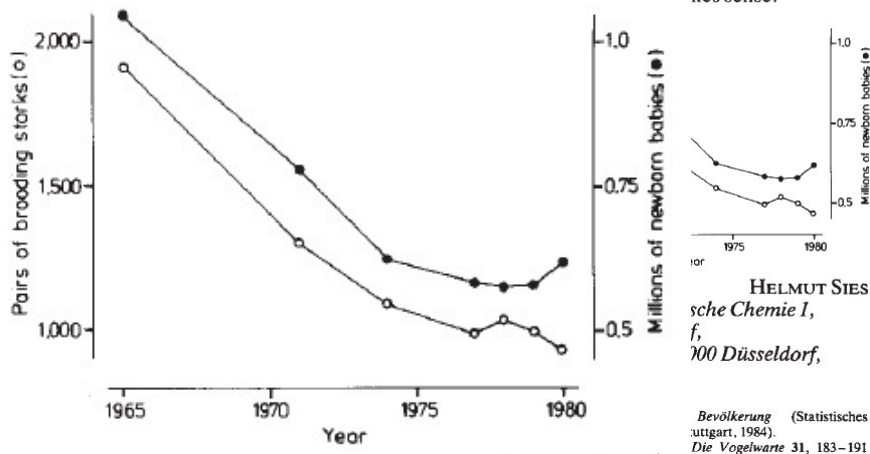
- ▶ Zmienne IC i ED nie korelują
- ▶ Zmienne I/IC i ED/IC korelują ale nie są niezależne.



Bociany!

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 makes sense.



HELMUT SIES
sche Chemie I,
f,
100 Düsseldorf,

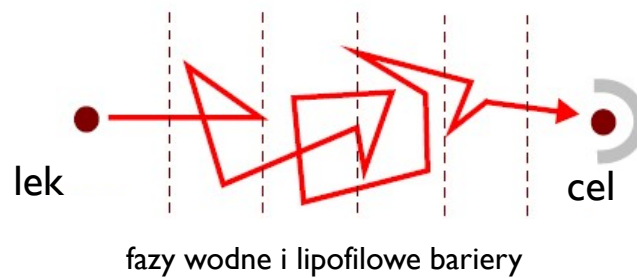
Bevölkerung (Statistisches
utgart, 1984).
Die Vogelwarte 31, 183–191

(1982).



Droga leku do celu

- ▶ Aktywność leku jest wynikiem transport leku i oddziaływanie leku z celem molekularnym.



Działanie leku

- ▶ Działanie leku zależy od:
 - ▶ Lipofilowości i charakteru jonowego – transport i dystrybucja w układach biologicznych
 - ▶ Geometryczne dopasowanie i komplementarność do powierzchni celu – oddziaływanie z miejscem wiązania.
- ▶ **Która konformacja jest aktywna biologicznie?**
 - ▶ w próżni
 - ▶ w kryształach
 - ▶ w roztworze wodnym
 - ▶ **w miejscu wiązania**

Badania QSAR

▶ **Podstawowe wymagania w QSAR**

- ▶ Wszystkie związki należą do jednej klasy
- ▶ Wszystkie związki mają taki sam mechanizm działania
- ▶ Wszystkie związki wiążą się w podobny sposób
- ▶ Efekty zamian izosterycznych są przewidywalne
- ▶ Wiązanie jest skorelowane z energią oddziaływania
- ▶ Aktywność biologiczna jest skorelowana z wiązaniem



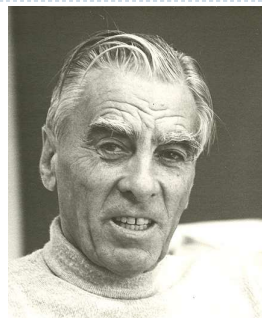
Właściwości cząsteczkowe

Właściwości cząsteczkowe	Oddziaływania
Lipofilowość	Oddziaływania hydrofobowe
Polaryzowalność	Oddziaływania Van der Waalsa
Gęstość elektronowa	Wiązania jonowe, oddziaływania dipol-dipol, wiązania wodorowe, oddziaływania przeniesienia ładunku
Topologia	Dopasowanie geometryczne, zawady steryczne



Model Hanscha

- ▶ Model Hanscha był pierwszym modelem typu QSAR
- ▶ Zależność właściwości biologicznych od właściwości związków
- ▶ Model liniowy



$$\text{Log I/C} = a \text{ log P} + b \sigma + c \text{ MR} + \dots + k$$

[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 FUJITA¹

RECEIVED AUGUST 19, 1963

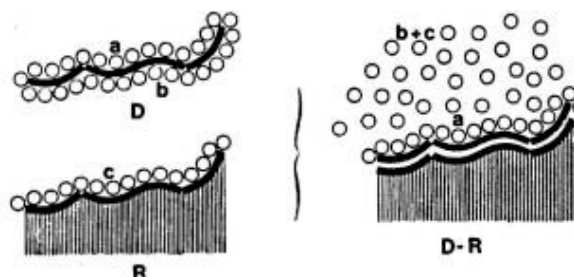
Using the substituent constant, σ , and a substituent constant, π , defined as $\pi = \log P_X - \log P_H$ (P_H is the partition coefficient of a parent compound and P_X that of a derivative), regression analyses have been made of the effect of substituents on the biological activity of benzoic acids 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.



Lipofilowość

- ▶ Oddziaływania hydrofobowe są ważnym elementem wiązania leku do receptora/białka
- ▶ Lipofilowość opisuje współczynnik podziału pomiędzy fazę wodną a organiczną (*n*-oktanol)

$$P = c_{\text{org}} / c_{\text{aq}}$$



Lipofilowość

▶ Przykład - karbaminiany

R- <chem>CONH2</chem>	P	Log P	$\Delta \log P$
metyl	0.22	-0.66	
etyl	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
heksyl	70.8	1.85	0.50
heptyl	230	2.36	0.51
oktyl	700	2.85	0.49



Lipofilowość

▶ System *n*-oktanol / woda:

- ▶ podobny do struktur membran
- ▶ dostępne donory i akceptory wiązań wodorowych
- ▶ *n*-oktanol praktycznie nierozpuszczalny w wodzie
- ▶ niska prężność par
- ▶ przezroczysty w zakresie UV
- ▶ duża baza danych wartości logP

▶ Addytywność wartości π (Hansch 1964):

- ▶ $\pi_x = \log P_{R-X} - \log P_{R-H}$



Lipofilowość

▶ Podstawniki pierścienia benzenowego

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



Lipofilowość

▶ Obliczenie wartości logP dla m-chlorotoluenu:

$$\text{▶ } \log P = \log P_{\text{Benzen}} + \pi_{\text{Cl}} + \pi_{\text{Me}} = 2.13 + 0.71 + 0.56 = \mathbf{3.40}$$

$$\text{▶ } \log P = \log P_{\text{Toluen}} + \pi_{\text{meta-Cl}} = 2.69 + 0.76 = \mathbf{3.45}$$

$$\text{▶ } \log P = \log P_{\text{Chlorobenzen}} + \pi_{\text{meta-Me}} = 2.84 + 0.51 = \mathbf{3.35}$$

▶ **Wartość eksperymentalna: $\log P_{\text{exp}} = 3.28$**

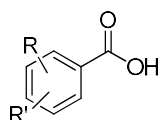


Lipofilowość

- ▶ Toksyczność podstawionych kwasów benzoesowych dla larw komarów.

- ▶ $-\log C = 0.519 \pi + 1.54$

- ▶ $R^2 = 0.955$



TOXICITY OF BENZOIC ACIDS TO MOSQUITO LARVAE

Functions	$\Sigma\sigma$	$\Sigma\pi^o$	$-\log C^o$		$\Delta \log C$
			Calcd.	Obsd.	
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 ₃) ₂	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

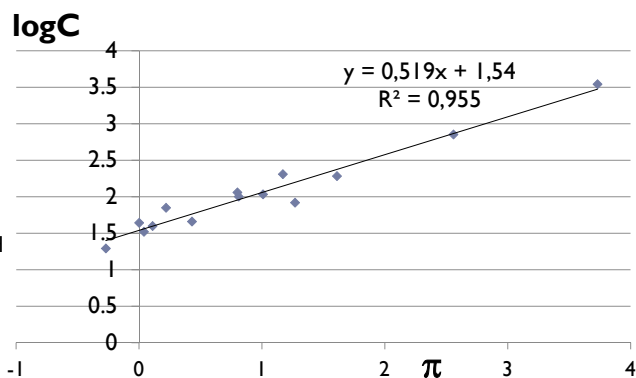
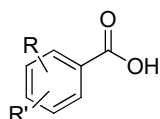


Lipofilowość

- ▶ Toksyczność podstawionych kwasów benzoesowych dla larw komarów.

- ▶ $-\log C = 0.519 \pi + 1.54$

- ▶ $R^2 = 0.955$



Polaryzowalność

- ▶ Parametry określające polaryzowalność cząsteczki:

- ▶ **Objętość molowa (MV):**

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

M – masa molowa, ρ – gęstość

- ▶ **Refrakcja molowa (MR):**

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

n – współczynnik załamania światła



Polaryzowalność

- ▶ **Wiązanie neutralnych związków do BSA**

- ▶ **Log I/C = 0.751 (± 0.07) log P + 2.300**

- ▶ (n = 42; r = 0.960)

- ▶ **Log I/C = 0.024 (± 0.02) MR + 2.901**

- ▶ (n = 42; r = 0.307)

- ▶ **Wiązanie β -D-glikozydów do Konkanawaliny A**

- ▶ **Log M₅₀ = 0.971 (± 0.56) π + 2.37**

- ▶ (n = 19; r = 0.664)

- ▶ **Log M₅₀ = 0.019 (± 0.003) MR + 2.23**

- ▶ (n = 19; r = 0.954)



Polaryzowalność

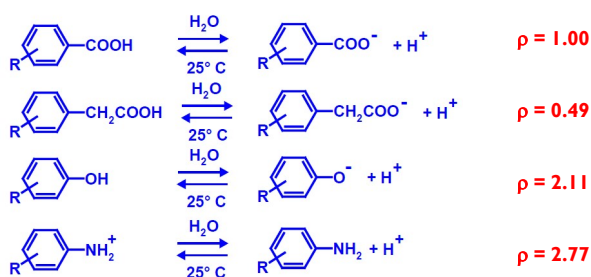
- ▶ **Inhibicja dehydrogenazy maleinowej przez kwasy 4-hydroksychinolinokarboksylowe**
 - ▶ $pI_{50} = 0.70 (\pm 0.17) MR + 2.29$
($n = 13; r = 0.939$)
- ▶ **Inhibicja wzrostu komórek przez kwasy 4-hydroksychinolinokarboksylowe**
 - ▶ $pI_{50} = 0.46 (\pm 0.11) \pi + 3.22$
($n = 14; r = 0.933$)
- ▶ **Wiązanie pochodnych aminokwasowych $R^1CONH-CH(R^2)-COOMe$ do chymotrypsyny**
 - ▶ $\text{Log } I/K_m = 0.082 (\pm 0.02) MRI + 1.382 (\pm 0.87) \pi_2 - 3.876$
($n = 21; r = 0.934$)

Równanie Hammetta

- ▶ Równanie Hammetta opisuje wpływ podstawników na równowagę reakcji

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

- ▶ ρ – opisuje reakcję (niezależny od podstawnika)
- ▶ σ – opisuje podstawnik (niezależny od reakcji)



Współczynniki podstawników σ

- ▶ Współczynnik podstawnika, σ opisuje właściwości elektronowe grup funkcyjnych
- ▶ Wpływ efektów indukcyjnych i mezomerycznych na przebieg reakcji

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

Podstawnik	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 ₂ H 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



Równanie Hammetta

- ▶ Może służyć do obliczania stałych reakcji chemicznych dla związków z różnymi podstawnikami
- ▶ **Kwas 3,5-dinitro-4-metylobenzoesowy (pK_a = 2.97)**
 - ▶ pK_a kwas benzoesowy = 4.20
 - ▶ pK_a calc = 4.20 - (0.71 - 0.17 + 0.71) = 2.95
- ▶ **Kwas m-hydroksybenzoesowy (pK_a = 4.06 i 9.92)**
 - ▶ pK_a kwas benzoesowy = 4.20, pK_a fenol = 9.92
 - ▶ ρ_{fenole} = 2.23; σ_{meta OH} = 0.12;
 - ▶ σ_{meta COOH} = 0.37; σ_{meta COO-} = -0.10)
 - ▶ pK_a COOH = 4.20 - 0.12 = 4.08
 - ▶ pK_a OH = 9.92 - 2.23 (-0.10) = 10.14



Inne deskryptory

- ▶ Liczba donorów wiązania wodorowego
- ▶ Liczba akceptorów wiązania wodorowego
- ▶ Liczba halogenów
- ▶ Liczba pierścieni aromatycznych
- ▶ Liczba grup jonizowalnych
- ▶ Liczba wiązań
- ▶ Masa cząsteczkowa
- ▶ Powierzchnia
- ▶ Ładunek
- ▶



Program Dragon 6
oblicza 4885 deskryptorów

- ▶ Można obliczyć kilka tysięcy różnych deskryptorów



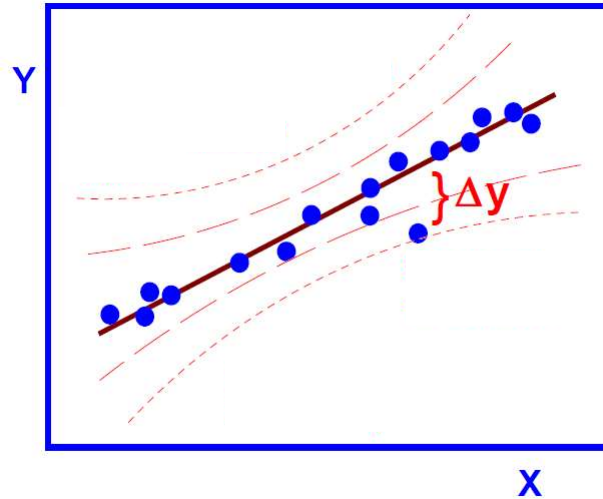
Deskryptory z mechaniki kwantowej

- ▶ Atomowe ładunki cząstkowe
- ▶ Moment dipolowy
- ▶ HOMO/LUMO
 - ▶ **HOMO** = energy of highest occupied molecular orbital, "nukleofilowość"
 - ▶ **LUMO** = energy of lowest unoccupied molecular orbital, "elektrofilowość"
- ▶ Inne

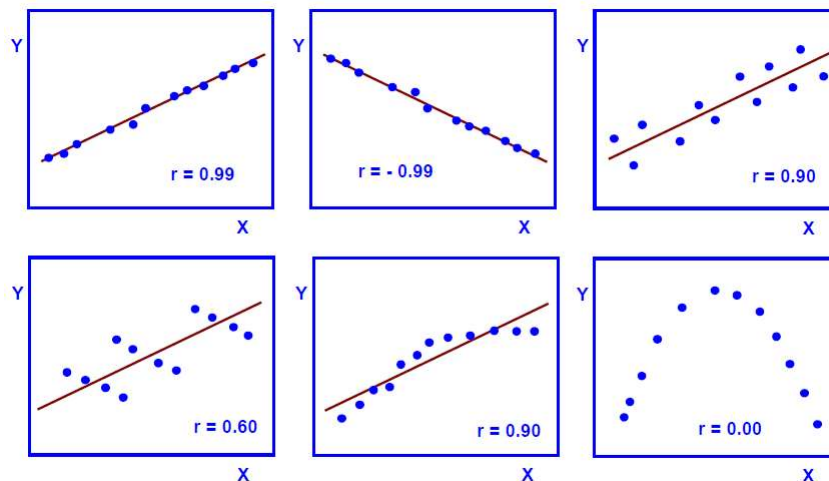


Regresja

- ▶ Regresja pozwala na wyznaczenie zależności pomiędzy zmiennymi
- ▶ Minimalizacja sumy kwadratów odchyleń



Regresja liniowa



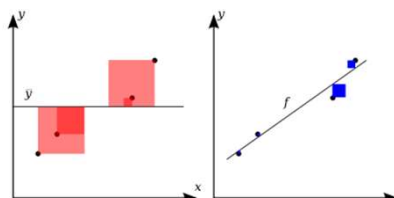
Parametry oceny analizy statystycznej

- ▶ Współczynnik determinacji:

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

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

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



- ▶ Odchylenie standardowe:

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

- ▶ Test Fischera



Równanie QSAR

Stężenie powodujące określony efekt biologiczny

Wartości współczynników regresji

Zakres 95% istotności dla wartości współczynników

$$\text{Log } 1/C = 1.15 (\pm 0.2) \pi - 1.46 (\pm 0.4) \sigma + 7.82 (\pm 0.2)$$

Logarytm z wartości odwrotnej jest dobrym skalowaniem wartości aktywności

Parametr lipofilowości

Parametr elektronowy

Czynnik stały

$$(n = 22; r = 0.945; s = 0.196; F = 78.6)$$

Liczba związków

Współczynnik korelacji

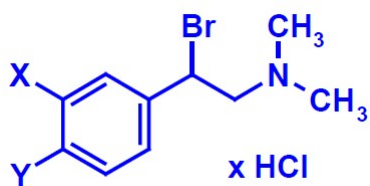
Odchylenie standardowe

Wartość testu Fischera istotność statystyczna



Przykład

- ▶ Aktywność
antyadrenergiczna
dimetylo-feniloetyloamin



<i>meta</i> (X)	<i>para</i> (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

Przykład

- ▶ Model Hanscha

$$\log 1/C = 1.151 (\pm 0.19) \pi - 1.464 (\pm 0.38) \sigma^+ + 7.817 (\pm 0.19)$$

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

Model Wilsona

▶ **Model Wilsona jest pierwszym modelem zależności struktura-aktywność**

- ▶ $\log I/C = \sum a_i + \mu$
- ▶ a_i = wkłady podstawników
- ▶ μ = wkład aktywności związku referencyjnego

▶ **Model mieszany Hansch/Wilson**

- ▶ $\log I/C = a (\log P)^2 + b \log P + c \sigma + \dots + \sum a_i + k$
- ▶ $\log I/C = a \log P - b \log (\beta P + I) + c \sigma + \dots + \sum a_i + k$



Przykład

- ▶ Model Wilsona
- ▶ Dane dla poprzedniego przykładu wskazujące położenie podstawników

meta (X)	para (Y)	log 1/C obs.	meta-					para-						
			F	Cl	Br	I	Me	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



Przykład

► Model Wilsona

$$\begin{aligned} \log I/C = & -0.301 (\pm 0.50) [m-F] + 0.207 (\pm 0.29) [m-Cl] \\ & + 0.434 (\pm 0.27) [m-Br] + 0.579 (\pm 0.50) [m-I] \\ & + 0.454 (\pm 0.27) [m-Me] + 0.340 (\pm 0.30) [p-F] \\ & + 0.768 (\pm 0.30) [p-Cl] + 1.020 (\pm 0.30) [p-Br] \\ & + 1.429 (\pm 0.50) [p-I] + 1.256 (\pm 0.33) [p-Me] \\ & + 7.821 (\pm 0.27) \end{aligned}$$

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

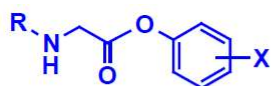
$\mu = 7.82$

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



Przykład 2

► Inhibicja papainy przez estry glicyny



X	R	log 1/K _m	π	MR	σ	I
4-NH ₂	-COC ₆ H ₅	3.58	-1.23	0.54	-0.66	0
4-Me	-COC ₆ H ₅	4.02	0.56	0.56	-0.17	0
H	-COC ₆ H ₅	3.77	0.00	0.10	0.00	0
4-Cl	-COC ₆ H ₅	4.00	0.71	0.60	0.23	0
4-F	-COC ₆ H ₅	3.69	0.14	0.09	0.06	0
3-NO ₂	-COC ₆ H ₅	4.74	-0.28	0.74	0.71	0
4-NO ₂	-COC ₆ H ₅	4.85	-0.28	0.74	0.78	0



Przykład 2

X	R	log 1/K _m	π	MR	σ	I
4-OH	-SO ₂ Me	2.05	-0.67	0.28	-0.37	1
4-OMe	-SO ₂ Me	2.13	-0.02	0.79	-0.27	1
4-Me	-SO ₂ Me	2.08	0.56	0.56	-0.17	1
3-Me	-SO ₂ Me	2.23	0.56	0.56	-0.07	1
H	-SO ₂ Me	1.79	0.00	0.10	0.00	1
4-F	-SO ₂ Me	1.95	0.14	0.09	0.06	1
3-OMe	-SO ₂ Me	2.29	-0.02	0.79	0.12	1
4-CHO	-SO ₂ Me	2.33	-0.65	0.69	0.42	1
4-Cl	-SO ₂ Me	2.38	0.71	0.60	0.23	1
3-F	-SO ₂ Me	1.98	0.14	0.09	0.34	1
4-COMe	-SO ₂ Me	2.57	-0.55	1.12	0.50	1
3-NO ₂	-SO ₂ Me	2.53	-0.28	0.74	0.71	1
4-NO ₂	-SO ₂ Me	2.71	-0.28	0.74	0.78	1

Przykład 2

► Benzamidy (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)

► Mezyloamidy (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)

Przykład 2

► Wszystkie związki (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 $r = 0.990; s = 0.148; F = 272.04$

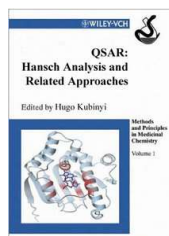
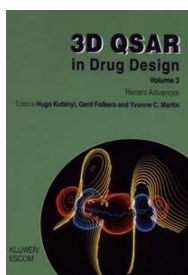
$$\log I/K_m = 0.569 (\pm 0.26) \text{ MR} + 0.561 (\pm 0.19) \sigma - 1.922 (\pm 0.15) I + 3.743 (\pm 0.17)$$

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



Hugo Kubinyi

► www.kubinyi.de



Podsumowanie

- ▶ Modele QSAR są najprostszym ilościowym oszacowaniem aktywności biologicznej związków
- ▶ Modele QSAR mają wiele ograniczeń.

