

QSAR = QUANTITATIVE STRUCTURE – ACTIVITY  
RELATIONSHIPS

KVANTITATIVNÍ VZTAHY MEZI (CHEMICKOU) STRUKTUROU A  
(BIOLOGICKOU) AKTIVITOU

Hledáme vztah, v němž je biologická aktivita funkcí struktury, resp. parametrů, které se strukturou souvisejí  
**A = f (struktura)**

2 základní přístupy QSAR

- **regresní analýza** – hledá matematický popis funkce, většinou s použitím lineární apod. regrese
- **empirické metody** – hledají pouze extrémy (maxima a minima) dané funkce

### Regresní analýza

hledá rovnici ve tvaru  $A = a_0 + a_1x_1 + a_2x_2 + \dots + a_nx_n$ ,

kde  $A$  je biologická aktivita,  $x$  jsou parametry vyplývající ze struktury sloučeniny,  $a$ ,  $b$  jsou regresní koeficienty ( $a_0$  ... absolutní člen), získané výpočtem. V případě tzv. **Hanschovy metody** jsou  $x$  fyzikálně-chemické parametry vyplývající ze struktury, v případě tzv. **Free-Wilsonova** přístupu parametry  $x$  vyjadřují přítomnost nebo nepřítomnost určitého substituentu nebo strukturního fragmentu v molekule.

## Hanschova metoda regresní analýzy

$$A = a_0 + a_1 x_1 + a_2 x_2 + \dots + a_n x_n$$

A ... biologická aktivita, často kvůli linearizaci vztahu v převrácené hodnotě nebo logaritmu (1/MIC ... převrácená hodnota minimální inhibiční koncentrace u antimikrobních látek, log LD<sub>50</sub>, logaritmy konstant vyjadřujících afinitu látky k receptoru, také ale některé farmakokinetické parametry, např. log BB, vyjadřující schopnost látky pronikat hematoencefalickou bariérou aj. a<sub>1</sub> ... a<sub>n</sub> ... regresní koeficienty, tj. koeficienty získané výpočtem s použitím např. lineární regrese

“Klasické” parametry x<sub>1</sub> ... x<sub>n</sub>

- hydrofobní
- elektronové
- sterické

a) **Hydrofobní parametry** - v rovnici často v kvadratickém tvaru – vyjadřují poměr rozpustnosti látky ve vodě a v lipidech; mají pro účinek často zásadní význam, zejména pro průnik bariérovými systémy organismu; např. **log P(oktanol/voda)**, **log P(cyklohexan/voda)** aj., parametr **R<sub>m</sub>** z rozdělovací TLC na tzv. reverzní fázi (stacionární fáze lipofilní, mobilní fáze hydrofilní):

$$R_m = \log\left(\frac{1}{R_f} - 1\right),$$

dále logaritmus tzv. **kapacitního faktoru log k'** z rozdělovací chromatografie plynové (GC) nebo vysokoúčinné kapalinové (HPLC)

$$\log k' = \log\left(\frac{t_r - t_0}{t_0}\right),$$

kde t<sub>r</sub> je retenční čas dané látky a t<sub>0</sub> je tzv. mrtvý retenční čas, tj. retenční čas látky, která se na koloně nezadržuje (např. u HPLC na reverzní fázi s oktadecylovaným silikagelem se pro stanovení t<sub>0</sub> používá dusitan sodný).

(Hanschův) **lipofilní parametr**  $\pi$  – pro řady (série) látek, obsahující na stejném strukturním fragmentu (nejčastěji benzenovém jádře) různé substituenty

$$\pi = \log \frac{P_X}{P_H} = \log P_X - \log P_H$$

kde  $P_x$  je rozdělovací koeficient substituované látky a  $P_H$  rozděl. koeficient látky nesubstituované.

### Vypočtené hydrofobní parametry

Kromě experimentálně stanovených hydrofobních parametrů se dnes stále častěji používají odhady těchto parametrů získaných výpočtem podle různých algoritmů, z nichž nejjednodušší jsou patrně výpočetní postupy pro odhad  $\log P$  (oktanol/voda) součtem  $\log P$  inkrementů podle **Rekera** a **Nyssa**

$$\log P = \sum_i a_i f_i,$$

kde  $f_i$  je  $\log P$  příslušného fragmentu a  $a_i$  je počet, kolikrát se tento fragment vyskytuje, nebo podle

**Hansche** (a Lea) vzorcem

$$\log P = \sum_i a_i f_i + \sum_j b_j f_j,$$

kde  $f_i$  je fragmentová konstanta,  $f_j$  je korekční faktor a  $a_i$  a  $b_j$  je odpovídající počet výskytů daného parametru. Takto vypočtený parametr se běžně označuje **ClogP**. Často se používají i složitější postupy vyžadující počítače a vhodný software, který často umožňuje i optimalizaci struktury metodami molekulové mechaniky a výpočty dalších parametrů pro výpočty QSAR (pro PC např. Molgen, HyperChem). Shoda vypočteného odhadu  $\log P$ , např. Clog P, s experimentálně stanovenou hodnotou, bývá pro různé výpočetní postupy velmi odlišná, často však pro výpočty postačí, existuje-li lineární závislost mezi  $\log P$  a Clog P.

### Mlog P (Moriguchi et al.)

- odvozen statistickou analýzou log P and strukturních dat 1230 sloučenin
- definován vzorcem

$$\begin{aligned} \log P = & 1.244(CX)^{0.6} - 1.017(NO)^{0.9} + 0.406PRX \\ & (t=60.5) \quad (t=58.5) \quad (t=33.8) \\ & - 0.145(UB)^{0.8} + 0.511HB + 0.268POL - 2.215AMP \\ & (t=9.5) \quad (t=5.9) \quad (t=19.6) \quad (t=19.5) \\ & + 0.912ALK - 0.392RNG - 3.684QN + 0.474NO2 \\ & (t=9.5) \quad (t=13.1) \quad (t=22.1) \quad (t=10.8) \\ & + 1.582NCS + 0.773BLM - 1.041 \\ & (t=16.4) \quad (t=5.0) \end{aligned}$$

$$n = 1230, \quad r = 0.952, \quad s = 0.411, \quad F_0(13, 1216) = 900.4$$

# Mlog P (pokračování)

kde

Parameter	Type <sup>a)</sup>	Description
<i>CX</i>	N	Summation of numbers of carbon and halogen atoms weighted by C: 1.0, F: 0.5, Cl: 1.0, Br: 1.5, and I: 2.0
<i>NO</i>	N	Total number of N and O atoms
<i>PRX</i>	N	Proximity effect of N/O; X-Y: 2.0, X-A-Y: 1.0 (X, Y: N/O, A: C, S, or P) with a correction (-1) for carboxamide/sulfonamide
<i>UB</i>	N	Total number of unsaturated bonds except those in NO <sub>2</sub>
<i>HB</i>	D	Dummy variable for the presence of intramolecular hydrogen bond as <i>ortho</i> -OH and -CO-R, -OH and -NH <sub>2</sub> , -NH <sub>2</sub> and -COOH, or 8-OH/NH <sub>2</sub> in quinolines, 5 or 8-OH/NH <sub>2</sub> in quinoxalines, <i>etc.</i>
<i>POL</i>	N	Number of aromatic polar substituents (aromatic substituents excluding Ar-CX <sub>2</sub> - and Ar-CX=C<, X: C or H)
<i>AMP</i>	N	Amphoteric property; α-aminoacid: 1.0, aminobenzoic acid: 0.5, pyridinecarboxylic acid: 0.5
<i>ALK</i>	D	Dummy variable for alkane, alkene, cycloalkane, or cycloalkene (hydrocarbons with 0 or 1 double bond)
<i>RNG</i>	D	Dummy variable for the presence of ring structures except benzene and its condensed rings (aromatic, heteroaromatic, and hydrocarbon rings)
<i>QN</i>	N	Quaternary nitrogen: >N <sup>+</sup> <, 1.0; N oxide, 0.5
<i>NO2</i>	N	Number of nitro groups
<i>NCS</i>	N	Isothiocyanato (-N=C=S), 1.0; thiocyanato (-S-C≡N), 0.5
<i>BLM</i>	D	Dummy variable for the presence of β-lactam

a) N, numerical variable; D, dummy variable.

## Variace na Hansche a Lea

### VG

Viswanadhan, V. N.; Ghose, A. K.; Revankar, G. R.; Robins, R. K., J. Chem. Inf. Comput. Sci., 1989, 29, 163-172

- separace log P na „atomové hydrofobicity“, lišící se podle toho, v jakém seskupení je atom umístěn
- celkem 120 „druhů atomů“
- odvozeny analýzou souboru 893 látek o známých lipofilitách

### KLOP

Klopman, G.; Li, Ju-Yun.; Wang, S.; Dimayuga, M.: J.Chem.Inf.Comput.Sci., 1994, 34, 752

- multivariační regresní analýza souboru log P 1663 organických látek vzniklo 98 seskupení s parciálními log P a 98 korekčních faktorů

Příklad QSAR závislosti pouze s hydrofobním parametrem  
 Účinnost fenolů jako induktorů apoptozy nádorových buněk  
*Hansch, C. et al.: Bioorg. Med. Chem. 11, 617 (2003)*

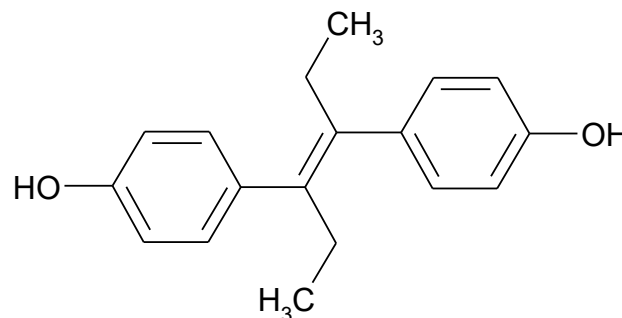
Table 1. Data for QSAR 3

Compd	Log 1/C	Pred log 1/C	Dev	Clog P
1 Estradiol	2.79	2.88	-0.09	3.78
2 4-MeO-phenol	1.27	1.41	-0.14	1.57
3 4-C <sub>6</sub> H <sub>5</sub> O-phenol	2.67	2.74	-0.07	3.57
4 4-CH <sub>2</sub> COO-phenol <sup>a</sup>	3.01	1.33	1.68	1.46
5 Bisphenol A	2.84	2.81	0.03	3.67
6 4-(Me) <sub>3</sub> C-phenol	2.65	2.56	0.09	3.30
7 4-CN-phenol	1.44	1.43	0.01	1.60
8 Diethylstilbestrol <sup>a</sup>	2.89	3.66	-0.77	4.96
9 4-I-phenol	2.08	2.29	-0.21	2.90
10 Phenol <sup>a</sup>	3.10	1.35	1.75	1.48
11 4-MeS-phenol <sup>a</sup>	2.60	1.72	0.88	2.03
12 4-C <sub>2</sub> H <sub>5</sub> O-phenol	2.51	2.12	0.39	2.63

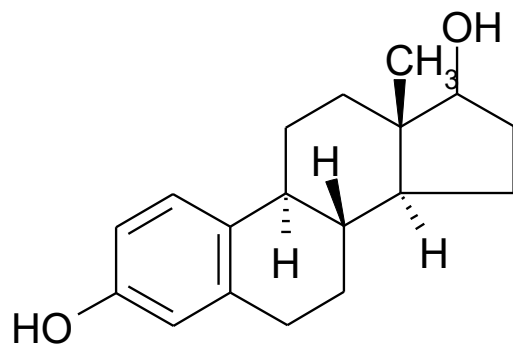
<sup>a</sup>Data points not used in deriving QSAR 3.

$$\log 1/C = 0,67(\pm 0,21)\text{ClogP} + 0.37(\pm 0.63)$$

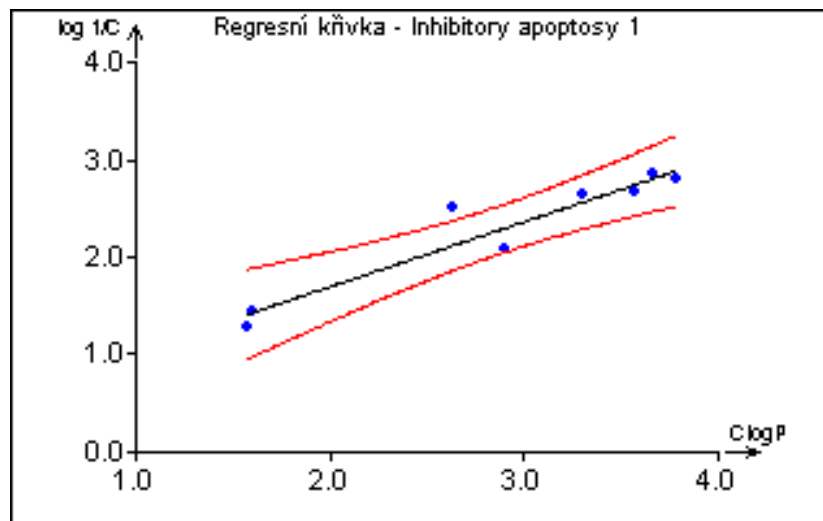
$$n = 8, r^2 = 0,910, s = 0,201, q^2 = 0,863$$



diethylstilbestrol (8)



estradiol (1)





## b) Elektronové parametry

-souvisejí s elektronovým obalem molekuly

• **Hammettova konstanta**  $\sigma$  - pro *m*- a *p*-substituované deriváty benzenu; vyjadřuje elektrondonorové (+M, +I) nebo elektronakceptorové (-M, -I) vlastnosti substituentu, resp. jeho vliv na elektronovou hustotu substituentu v pol. 1; definována na základě Hammetovy rovnice, odvozené pomocí rychlostních konstant hydrolýzy ethylesterů substituovaných benzoových kyselin uváděné obvykle ve tvaru

$$\log k = \log k_H + \rho \cdot \sigma,$$

kde *k* je rychlostní konstanta reakce,  $\rho$  tzv. reakční konstanta, charakterizující podmínky reakce (zejména rozpouštědlo a teplota) a  $k_H$  rychlostní konstanta hydrolýzy nesubstituované látky, tj. ethylbenzoátu. (Pro *o*-deriváty tato lineární závislost neplatí, uplatňují se zde ve zvýšené míře sterické vlivy). Pro Hammetovu konstantu  $\sigma$  tedy platí

$$\sigma = \frac{\log k - \log k_H}{\rho}$$

$\sigma$  vyjadřuje míru elektrondonorového nebo elektronakceptorového působení substituentu, charakterizuje tedy souhrnně jeho mezomerní a indukční efekt. Pro elektrondonorové substituenty nabývá záporných hodnot ( $\sigma < 0$ ), pro elektronakceptorové kladných ( $\sigma > 0$ ). Běžně se rozlišují následující typy Hammetovy konstanty:

$\sigma_m$  konstanta pro substituent v *m*-poloze

$\sigma_p$  konstanta pro substituent v *p*-poloze

Tyto konstanty (a lineární tvar Hammetovy rovnice) platí obecně pro reakce na **druhém nebo dalším atomu** od benzenového jádra a pro reakce na první atomu, který není silně konjugován, pro reakční centra na **silně konjugovaném atomu** (dusík anilinu, kyslík fenolu) byly pro substituenty v poloze *para* zavedeny tzv. duální konstanty, pro akceptorové substituenty  $\sigma_p^-$ , použitelné např. pro disociace na dusíku anilinu nebo kyslíku fenolu a pro donorové  $\sigma_p^+$ , používané u reakcí, kde na  $\alpha$ -atomu vůči jádru vzniká aspoň částečný kladný náboj.

- odvozené konstanty:  $\sigma_m$ ,  $\sigma_i$ ,  $\sigma^*$ , podobné Swain-Luptonovy konstanty  $\mathcal{F}$ ,  $\mathcal{R}$

- parametry ze **spekter** a jiných fyzikálních měření – chem. posuny  $\delta$  z NMR, vlnová délka maxima absorpce  $\lambda_{\max}$  z UV-VIS spekter, vlnočet  $\nu$  výrazného absorpčního pásu v IČ spektrech, půlvlnný potenciál  $E_{1/2}$  z polarografie

- vypočtené elektronové parametry: polarita, polarizovatelnost, parciální náboj na určitém atomu

### c) Sterické parametry

- vyjadřují “celkovou objemnost” molekuly

• **van der Waalovy poloměry  $v_F$**

**Taftova sterická konstanta  $E_s$**  odvozená pomocí rychlostních konstant hydrolýzy esterů alkanových kyselin

$$E_s = \log \frac{k_x}{k_h},$$

kde  $k_x$  je rychlostní konstanta hydrolýzy esteru příslušné alkanové kyseliny  $\text{RCOOR}'$  a  $k_h$  obdobná konstanta pro ester kyseliny octové  $\text{CH}_3\text{COOR}'$  - standard.  $E_s$  není čistě sterickým parametrem, zahrnuje částečně i elektronové vlivy (+I).

$E_s(\text{CH}_3) = 0$ , objemnější substituenty  $E_s < 0$ , méně objemné  $E_s > 0$

**Sterická konstanta  $v$  odvozená z van der Waalových poloměrů [pm]**

$$v[\text{pm}] = r[\text{pm}] - 120\text{pm}$$

120 pm je poloměr vodíku

• pro symetrické substituenty získány z modelů, pro nesymetrické rovněž z kyselé hydrolýzy esterů

Hodnoty sterických konstant  $E_s$  a  $v$  pro některé substituenty

Substituent	$E_s$	$v$
-H	+1,24	0
-CH <sub>3</sub>	0	0,52
-C <sub>2</sub> H <sub>5</sub>	-0,07	0,56
-C <sub>3</sub> H <sub>7</sub>	-0,36	0,68
-CH(CH <sub>3</sub> ) <sub>2</sub>	-0,47	0,76
-C <sub>4</sub> H <sub>9</sub>	-0,39	0,68
-CH <sub>2</sub> CH(CH <sub>3</sub> ) <sub>2</sub>	-0,93	0,98
-CH(CH <sub>3</sub> )C <sub>2</sub> H <sub>5</sub>	-1,13	1,02
-C <sub>5</sub> H <sub>11</sub>	-0,40	0,68
cyklo-C <sub>5</sub> H <sub>9</sub>	-0,51	-
cyklo-C <sub>6</sub> H <sub>11</sub>	-0,79	1,25

## Další parametry užívané v QSAR

- >většinou vypočtené
- >zahrnují často 2 i 3 typy vlivů (hydrof.+elektron.+ster.)

“Klasické”

### •parachor

$$P_r = \frac{M}{d} \gamma^{(1/4)}$$

, kde  $\gamma$  je povrchové napětí, M molární hmotnost a d hustota.

•**molární refrakce** (= molekulární refraktivita) **MR** (též CMR); def. vztah označován Lorentzova-Lorenzova rovnice

$$MR = n^2 - \frac{1}{n^2} + \frac{2M}{d} = \frac{(n^2 - 1) M}{(n^2 + 2) d},$$

kde n je index lomu.

“Neklasické”

•**solvatační energie** – je-li pro vodu, pak **hydratační energie**  $\Delta G_o^w$

•**povrchy molekuly** různého typu – polární van der Waalsův, nepolární, přístupný vodě, dynamický polární (DPSA), topologický polární (TPSA) aj.

•**objemy molekuly** – polární, přístupný vodě aj.

## Free- Wilsonova metoda regresní analýzy

- hledá závislost biologické aktivity na přítomnosti nebo nepřítomnosti určitých substituentů nebo strukturních fragmentů v molekule. Jde o vlastně o **statistickou separaci aktivity** na příspěvky určitých částí molekul, tj. předpokládá se aditivita vlivu substituentů nebo jiných částí molekuly. Metoda vede k řešení soustav rovnic o větším počtu neznámých, které se v jednoduchých případech dají řešit maticovým počtem, jinak pomocí statistických programů umožňujících multilineární regresi (MLR).
- obě metody lze též kombinovat. Část nezávislých proměnných pak vyjadřuje fyzikálně-chemické vlastnosti látek, další, nazývané „**indikátorové proměnné**“ (symbol I), udávají přítomnost nebo nepřítomnost určitých molekulových fragmentů. Indikátorových proměnných bývá obvykle malý počet, často pouze jediná.

## Free- Wilsonova metoda

- aby poskytovala prognózy, jsou nutná nejméně 2 obměňovaná místa
- každá látka popsána rovnicí
- indexy značíme jednak substituenty, jednak obměňovaná místa molekuly  $\Rightarrow$  každé nezávislé proměnné a každému regresnímu koeficientu přiřazeny 2 indexy, z nichž první označuje příslušnost k obměňované poloze (např. indexy 1,2 ... až j) a druhý určitý substituent či jiný obměňovaný fragment (např. indexy 1,2 ... až m u obměň. na prvním obměň. místě; 1,2 ... až n na druhém obměň. místě; 1,2 ... až p na třetím obměň. místě atd. až např. po z)
- biologické aktivity y jednotlivých látek rovněž označíme indexy



$$y_1 = a_0 + a_{1,1} x_{1,1} + a_{1,2} x_{1,2} + \dots + a_{1,m} x_{1,m} + a_{2,1} x_{2,1} + a_{2,2} x_{2,2} + \dots + a_{2,n} x_{2,n} + a_{3,1} x_{3,1} + a_{3,2} x_{3,2} + \dots + a_{3,p} x_{3,p} + \dots +$$

$$a_{j,z} x_{j,z}$$

$$y_2 = a_0 + a_{1,1} x_{1,1} + a_{1,2} x_{1,2} + \dots + a_{1,m} x_{1,m} + a_{2,1} x_{2,1} + a_{2,2} x_{2,2} + \dots + a_{2,n} x_{2,n} + a_{3,1} x_{3,1} + a_{3,2} x_{3,2} + \dots + a_{3,p} x_{3,p} + \dots +$$

$$a_{j,z} x_{j,z}$$

$$y_3 = a_0 + a_{1,1} x_{1,1} + a_{1,2} x_{1,2} + \dots + a_{1,m} x_{1,m} + a_{2,1} x_{2,1} + a_{2,2} x_{2,2} + \dots + a_{2,n} x_{2,n} + a_{3,1} x_{3,1} + a_{3,2} x_{3,2} + \dots + a_{3,p} x_{3,p} + \dots +$$

$$a_{j,z} x_{j,z}$$

·  
·  
·  
·

$$y_i = a_0 + a_{1,1} x_{1,1} + a_{1,2} x_{1,2} + \dots + a_{1,m} x_{1,m} + a_{2,1} x_{2,1} + a_{2,2} x_{2,2} + \dots + a_{2,n} x_{2,n} + a_{3,1} x_{3,1} + a_{3,2} x_{3,2} + \dots + a_{3,p} x_{3,p} + \dots +$$

$$a_{j,z} x_{j,z}$$

- skutečná soustava rovnic jednodušší, řada členů rovna 0
- přesto v daném stavu neřešitelná; proměnné x nejsou nezávislé  $\Rightarrow$  nutnost zjednodušujícího předpokladu

### Řešení dle Freeho a Wilsona

- předpoklad: součet příspěvků vlivů jednotlivých substituentů na biol. aktivitu ve sledované skupině pro každé obměňované místo je nulový  $\Rightarrow$  **symetrické rovnice**
- b ... počet výskytů daného fragmentu; indexy mají stejný význam jako v předchozí soustavě rovnic

$$b_{1,1} a_{1,1} + b_{1,2} a_{1,2} + \dots + b_{1,m} a_{1,m} = 0$$

$$b_{2,1} a_{2,1} + b_{2,2} a_{2,2} + \dots + b_{2,m} a_{2,m} = 0$$

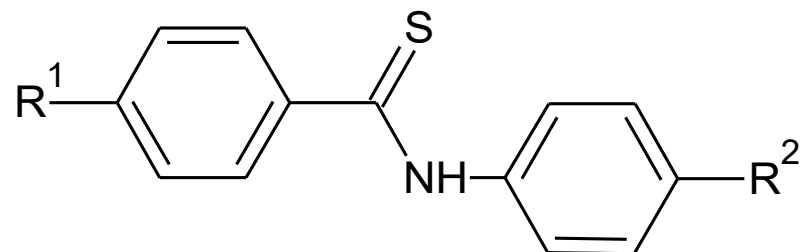
·  
·  
·

$$b_{i,1} a_{i,1} + b_{i,2} a_{i,2} + \dots + b_{i,j} a_{i,j} = 0$$

- pomocí symetrických rovnic se na každém obměňovaném místě jeden substituent nahradí (vyjádří) kombinací ostatních a soustava se stává řešitelnou



## Příklad – biologicky aktivní thiobenzanilidy



	$R^1$	$R^2$
1. thiobenzanilid	H	H
2. 4'-chlor-4-methylthiobenzanilid	$CH_3$	Cl
3. 4-chlor-4'-methylthiobenzanilid	Cl	$CH_3$
4. 4,4'-dichlorthiobenzanilid	Cl	Cl
5. 4'-methylthiobenzanilid	H	$CH_3$

•v souboru sloučenin 2 obměňovaná místa

### Matrice struktury

Název	$x_{1,1}$	$x_{1,2}$	$x_{1,3}$	$x_{2,1}$	$x_{2,2}$	$x_{2,3}$
	(H)	( $CH_3$ )	(Cl)	(H)	( $CH_3$ )	(Cl)
thiobenzanilid	1	0	0	1	0	0
4'-brom-4-methylthiobenzanilid	0	1	0	0	0	1
4-chlor-4'-methylthiobenzanilid	0	0	1	0	1	0
4,4'-dichlorthiobenzanilid	0	0	1	0	0	1
4'-methylthiobenzanilid	<u>1</u>	<u>0</u>	<u>0</u>	<u>0</u>	<u>1</u>	<u>0</u>
	<b>2</b>	<b>1</b>	<b>2</b>	<b>1</b>	<b>2</b>	<b>2</b>
	<b>b</b> <sub>1,1</sub>	<b>b</b> <sub>1,2</sub>	<b>b</b> <sub>1,3</sub>	<b>b</b> <sub>2,1</sub>	<b>b</b> <sub>2,2</sub>	<b>b</b> <sub>2,3</sub>

- první obměňovanou polohou pol. 4; v matici první 3 sloupce
- druhou obměňovanou polohou pol. 4'; v matici další 3 sloupce
- pod tabulkou uvedeno, kolikrát se daný substituent vyskytuje (b)
- pro každé substituované místo lze napsat symetrickou rovnici:

$$2 a_{1,1} + a_{1,2} + 2 a_{1,3} = 0$$

$$a_{2,1} + 2 a_{2,2} + 2 a_{2,3} = 0$$

- ze symetrických rovnic lze libovolný regresní koeficient v každé obměňované poloze vyjádřit pomocí ostatních, např.

$$a_{1,2} = -2 a_{1,1} - 2 a_{1,3}$$

$$a_{2,1} = -2 a_{2,2} - 2 a_{2,3}$$

- vztahů se využije k úpravě (zjednodušení) matice proměnných x, které již lze považovat za nezávislé

Název	$x_{1,1}$	$x_{1,3}$	$x_{2,2}$	$x_{2,3}$
	(H)	(Cl)	(CH <sub>3</sub> )	(Cl)
thiobenzanilid	1	0	-2	0
4'-chlor-4-methylthiobenzanilid	-2	-2	0	1
4-chlor-4'-methylthiobenzanilid	0	1	1	0
4,4'-dichlorthiobenzanilid	0	1	0	1
4'-methylthiobenzanilid	1	0	1	0

- v matici na každém místě o jeden substituent méně

•přepíšeme-li do soustavy rovnic, dostáváme

$$y_1 = a_0 + a_{1,1} - 2 a_{2,2} - 2 a_{2,3}$$

$$y_2 = a_0 - 2 a_{1,1} - 2 a_{1,3} + a_{2,3}$$

$$y_3 = a_0 + a_{1,3} + a_{2,2}$$

$$y_4 = a_0 + a_{1,3} + a_{2,3}$$

$$y_5 = a_0 + a_{1,1} + a_{2,1}$$

•soustava 5 rovnic o 5 neznámých je řešitelná (metoda dosazovací, sčítací, maticový počet, multilineární regrese na počítači)

•za y dosadíme příslušné hodnoty biologické aktivity jednotlivých látek

•získáme hodnoty regresních koeficientů  $a_0$ ,  $a_{1,1}$ ,  $a_{1,3}$ ,  $a_{2,2}$ ,  $a_{2,3}$

•hodnoty  $a_{1,2}$  a  $a_{2,2}$  se získají dosazením do předchozích rovnic, ve kterých byly vyjádřeny pomocí ostatních regresních koeficientů

• $a_0$  zde vyjadřuje průměrnou hodnotu studované aktivity souboru, ostatní hodnoty a s příslušnými indexy udávají, jakým způsobem obměňované substituenty v jednotlivých polohách tuto aktivitu ovlivňují (tj. jsou to kladné nebo záporné odchylky od průměrné aktivity)

• pomocí získaných hodnot regresních koeficientů lze výpočtem odhadnout aktivity látek s kombinacemi použitých substituentů, které ve výchozí matici nebyly zahrnuty (tj. 4-chlorthiobenzanilid, 4'-chlorthiobenzanilid a 4,4'-dimethylthiobenzanilid)

- Aktivity: antituberkulotické, vyjádřené jako log MIC proti *Mycobacterium kansasii*

## Řešení dle Fujity a Bana

- jednodušší: místo symetrických rovnic se vynechá jeden ze sloupců pro každé ze substitucí obměňovaných míst
- vhodné vynechat sloupce odpovídající substituci dané polohy vodíkem  $\Rightarrow$  zavedení nulové hodnoty pro každé substituční místo; hodnota odpovídá substituentu ve vynechaném sloupci
- příspěvky odpovídající vlivům substituentů na biol. aktivitu jsou pak relativní vůči této hodnotě  $\Rightarrow$  výhodné mít v souboru nesubst. látku

### Zjednodušená matice struktury

Název	$x_{1,2}$	$x_{1,3}$	$x_{2,2}$	$x_{2,3}$
	(CH <sub>3</sub> )	(Cl)	(CH <sub>3</sub> )	(Cl)
thiobenzanilid	0	0	0	0
4'-brom-4-methylthiobenzanilid	1	0	0	1
4-chlor-4'-methylthiobenzanilid	0	1	1	0
4,4'-dichlorthiobenzanilid	0	1	0	1
4'-methylthiobenzanilid	0	0	1	0

- soustava rovnic k řešení pak vypadá následovně:

$$y_1 = a_0$$

$$y_2 = a_0 + a_{1,2} + a_{2,3}$$

$$y_3 = a_0 + a_{1,3} + a_{2,2}$$

$$y_4 = a_0 + a_{1,3} + a_{2,3}$$

$$y_5 = a_0 + a_{2,2}$$

- na výstupu hodnota  $a_0$  odpovídá aktivitě nesubst. látky (zde thiobenzanilid)
- ostatní hodnoty  $a$  odpovídají příspěvkům jednotlivých substituentů k aktivitě

### **Empirické metody QSAR**

- používány s výhodou tam, kde nelze jednoduše zjistit matem. popis funkce  $A = f(\text{struktura})$
- hledají pouze extrémy (maxima a/nebo minima) dané funkce; matem. popis přitom zůstává “černou skříňkou”
- při jejich aplikaci syntetik usměrňuje volbu látky k syntéze na základě biol. hodnocení látky předchozí

## Fibonacci – Leonardo Pisánský (cca 1175 - 1250)



- přinesl do Evropy arabský číselný systém
- Fibonacciho posloupnost: nekonečná posloupnost přirozených čísel, začínající 0, 1, 1, 2, 3, 5, 8, 13, 21, ... (čísla nacházející se ve Fibonacciho posloupnosti jsou někdy nazývána Fibonacciho čísla), kde každé číslo je součtem dvou předchozích.
- rekurzivní zápis:

$$F(n) = \begin{cases} 0, & \text{pro } n = 0; \\ 1, & \text{pro } n = 1; \\ F(n-1) + F(n-2) & \text{jinak.} \end{cases}$$

- explicitní vyjádření – vztah pro n-tý člen:

$$F(n) = \frac{\phi^n}{\sqrt{5}} - \frac{(1-\phi)^n}{\sqrt{5}}; \quad \text{kde } \phi = (1 + \sqrt{5})/2 \approx 1,62$$



## **Optimalizace podle jednoho strukturního parametru – Fibonacciho optimalizace**

Látky se seřadí podle rostoucí hodnoty strukturního parametru, u něhož se předpokládá, že aktivitu významně ovlivňuje. Počet látek musí odpovídat počtu bodů v některém z Fibonacciho intervalů (viz tab. 1), není-li tomu tak, některá z krajních látek, u nichž je malá pravděpodobnost, že by byly nejúčinnější, se vypustí nebo se naopak přidá fiktivní krajní látka. K syntéze se vyberou látky, které mají v intervalu pořadí uvedené ve druhém sloupci tabulky. Provede se biologické hodnocení těchto látek a podle jeho výsledků se část intervalu od jednoho z krajních bodů směrem k méně aktivní sloučenině vyloučí. Výsledný soubor je dalším Fibonacciho intervalem. Výběr se opakuje tak dlouho, dokud se nedospěje k nejúčinnější sloučenině. Metoda umožňuje podstatně snížit počet syntetizovaných a testovaných látek, např. namísto 589 látek, které by bylo nutno připravit a otestovat, abychom mezi nimi našli nejúčinnější, stačí připravit a stanovit biol. aktivitu u pouhých 13 sloučenin (viz sloupec C tab. 1).

Tab. 1. Fibonacciho optimalizace

**Legenda:**     **A ... počet látek příslušného Fibonacciho intervalu**  
                  **B ... pořadí látek určených pro syntézu a testování v rámci intervalu**  
                  **C ... počet látek potřebný k dosažení optimalizace**

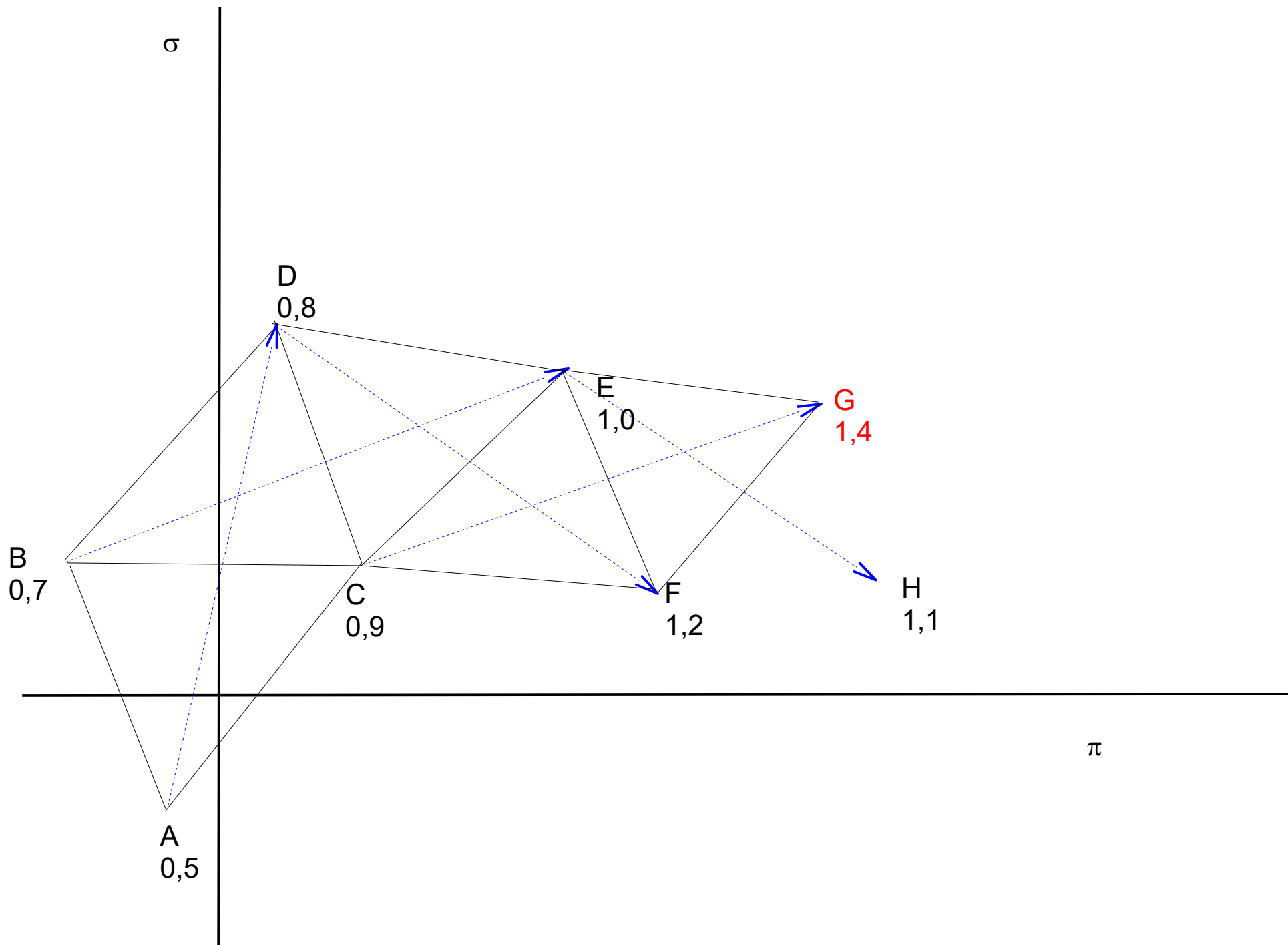
A	B	C	A	B	C	A	B	C
2	1 a 2	2	20	8 a 13	6	143	55 a 89	10
4	2 a 3	3	33	13 a 21	7	222	89 a 144	11
7	3 a 5	4	54	21 a 34	8	366	144 a 233	12
12	5 a 8	5	88	34 a 55	9	589	233 a 377	13

## Optimalizace podle více strukturních parametrů

### Simplexová metoda

Každou látku lze charakterizovat jako **bod v n-rozměrném prostoru**, ve kterém **biologická aktivita** je jednou souřadnicí a zbývajících  $n - 1$  souřadnic představují **fyzikální a fyzikálně-chemické vlastnosti** sloučeniny, o nichž se předpokládá, že aktivitu ovlivňují. Pokud pracujeme v (klasickém) trojrozměrném prostoru, tj. optimalizujeme-li pouze dva parametry, můžeme optimalizaci provést i graficky na milimetrovém papíře. Pracujeme vlastně v průmětu do roviny vlastností. K (syntéze a) hodnocení se vyberou 3 látky, které v rovině souřadnic vlastností nejsou od sebe příliš vzdálené s jejichž souřadnice v této rovině tvoří trojúhelník, nejlépe rovnostranný. Porovnáme aktivity těchto 3 sloučenin a z bodu, patřícího látce o nejnižší aktivitě, vedeme polopřímku středem spojnice dvou bodů s vyššími aktivitami (alternativně bodem, který vznikne rozdělením této spojnice v převráceném poměru aktivit) a na ní nalezneme bod stejně vzdálený, ale opačně orientovaný, než bod s nejnižší aktivitou. Pokud tento bod neodpovídá žádné sloučenině, použijeme k hodnocení látku nejbližší. Z tohoto bodu a dvou předchozích vytvoříme další trojúhelník, který podrobíme stejnému optimalizačnímu pochodu. Postup opakujeme, dokud nacházíme látky o rostoucí aktivitě. Jakmile aktivita začne klesat, můžeme látku s dosaženou nejvyšší aktivitou považovat za optimální.

# Simplexová metoda

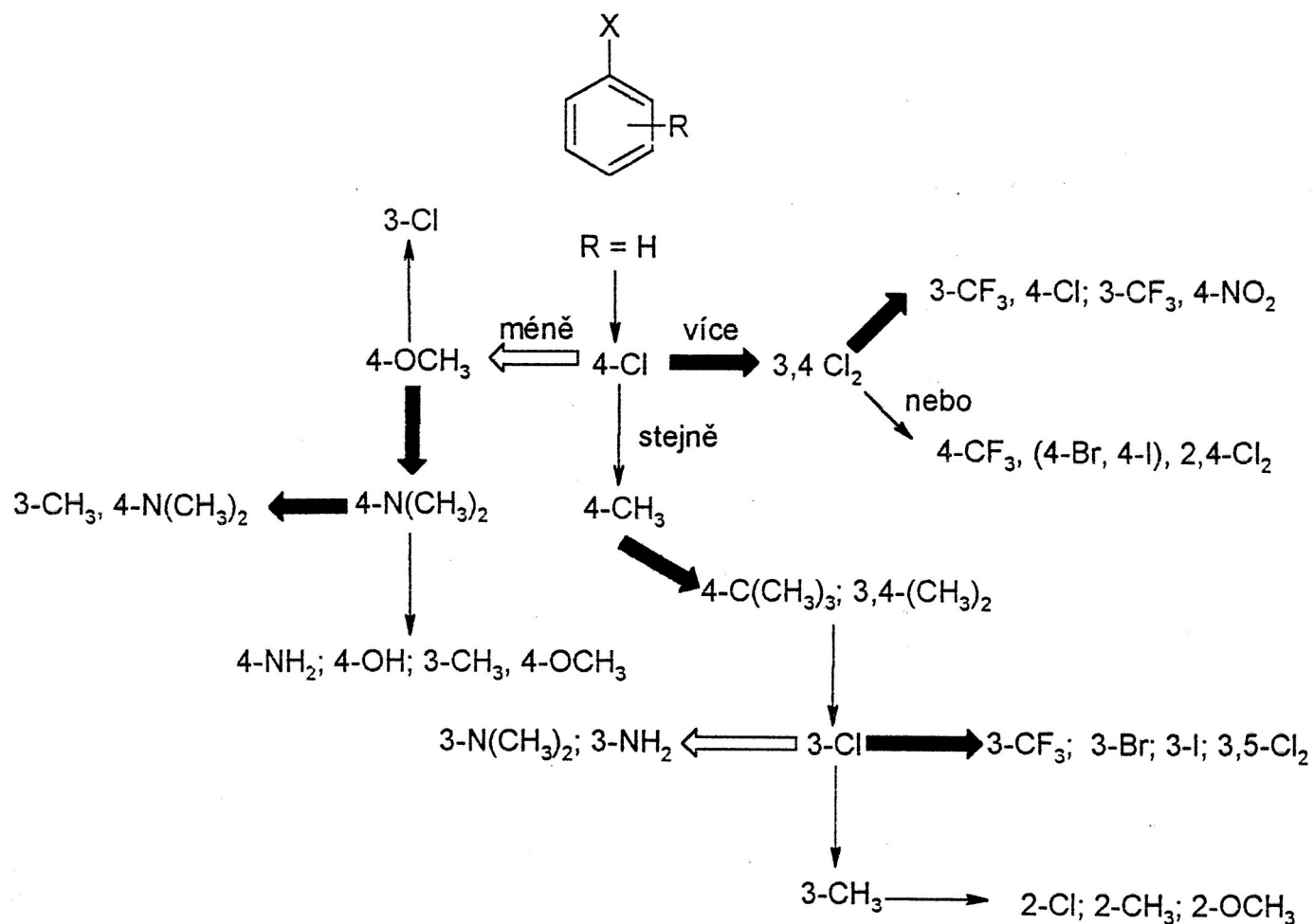


## Optimalizace podle více strukturních parametrů

### Optimalizační schémata

- postup racionálních myšlenkových pochodů farm. chemika
- zohledňují hydrofobní, elektronové a sterické parametry
- nejsou univerzální; pro určitý typ obměn na určité struktury bývá nutné sestavit nové

Schéma obměn substituentů na fenylu (Topliss 1972)



## Komentář ke schématu obměn substituentů na fenylu

Při optimalizaci podle uvedeného schématu postupujeme tak, že nejprve připravíme **nesubstituovanou sloučeninu** a její **4-chlorderivát**. Substituce chlorem sníží elektronovou hustotu v poloze 1 a zároveň zvýší lipofilitu (4-Cl:  $\sigma = 0,23$ ;  $\pi = 0,71$ ); je-li 4-chlorderivát aktivnější, lze lipofilitu a elektronakceptorové vlastnosti zvýšit další substitucí chlorem. Je-li 4-chlorderivát méně aktivní, můžeme předpokládat, že pokles elektronové hustoty ovlivnil aktivitu negativně, a připravíme **4-methoxyderivát**, který má lipofilitu prakticky stejnou jako nesubst. látka, el. hustota v pol. 1 je však vyšší (4-OCH<sub>3</sub>:  $\sigma = -0,27$ ,  $\pi = 0,02$ ). Není-li zřetelný rozdíl mezi aktivitou nesubst. látky a 4-chlorderivátu; můžeme předpokládat, že vliv el. hustoty a lipofility působí proti sobě a zkusíme připravit **4-methylderivát** (4-CH<sub>3</sub>:  $\sigma = -0,17$ ,  $\pi = 0,56$ ). Je-li u všech látek substituovaných v pol. 4 aktivita nižší než u nesubstituované látky, je zřejmé, že substituce v poloze 4 je stericky nevýhodná, a budeme připravovat látky substituované v polohách 2 a 3. V jednotlivých větvích tohoto schématu je možné postupovat dále, dokud nedojdeme ke sloučenině s maximální aktivitou.