

HERRAMIENTAS PARA EL DISEÑO DE COMPUESTOS BIOACTIVOS Y ANÁLISIS CUANTITATIVO DE RELACIÓN ESTRUCTURA QUÍMICA-BIORESPUESTA (Q.S.A.R.)

CURSO DE PROFUNDIZACIÓN PEDECIBA-QUÍMICA

NOVIEMBRE 2022

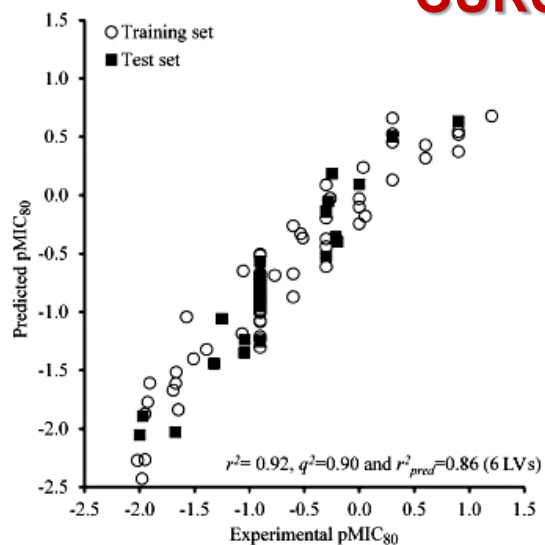
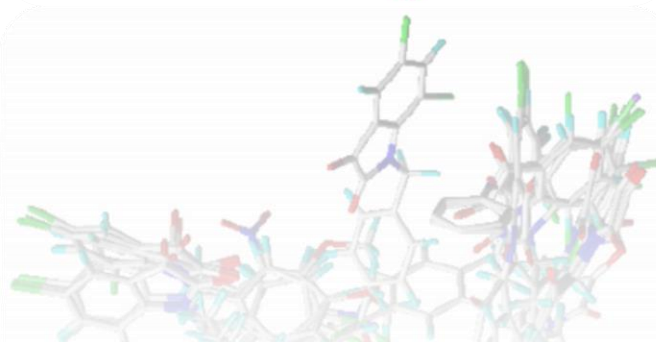
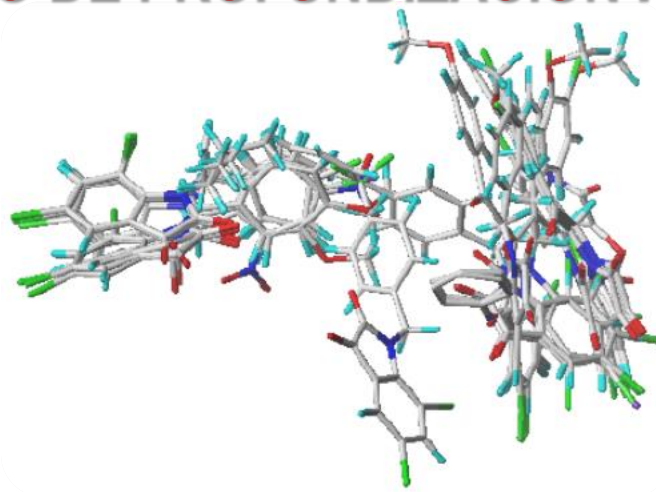
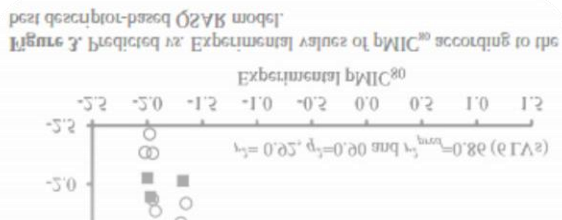


Figure 3. Predicted vs. Experimental values of pMIC₈₀ according to the best descriptor-based QSAR model.



Dr. Hugo Cerecetto
Dr. Marcos Couto

HERRAMIENTAS PARA EL DISEÑO DE COMPUESTOS BIOACTIVOS Y ANÁLISIS CUANTITATIVO DE RELACIÓN ESTRUCTURA QUÍMICA-BIORESPUESTA (Q.S.A.R.)

PROGRAMA

1. HERRAMIENTAS PARA EL DISEÑO DE COMPUESTOS BIOACTIVOS. Generalidades.

1.1. Propiedades tipo-fármaco.

1.2. Estrategias para integrar propiedades tipo-fármaco en el proceso de descubrimiento de agentes bioactivos.

1.3. Optimización química del líder: i) ¿Cómo mejorar la biodisponibilidad oral? ii) ¿Cómo controlar la estabilidad metabólica? iii) ¿Cómo controlar la estabilidad plasmática? iv) ¿Cómo mejorar la penetración de la barrera hematoencefálica? v) ¿Cómo evitar la inhibición/inducción de las CYPs? vi) ¿Cómo evitar la interacción con hERG? vii) ¿Cómo prevenir la toxicidad?

2. DISEÑO DE SERIES.

2.1. Métodos de optimización directa.

2.2. Métodos de exploración completa del campo estructural.

3. ANÁLISIS CUANTITATIVO DE RELACIÓN ESTRUCTURA QUÍMICA-BIORESPUESTA.

Antecedentes S.A.R. y Q.S.A.R. Trabajos de Ferguson, Hansch y Kubinyi.

3.1. Parámetros Biológicos. Respuesta biológica a dosis fija y a dosis variable.

3.2. Descriptores Físicoquímicos.

3.2.1. Descriptores hidrofóbicos. Coeficiente de reparto, constante de Hansch (π).

Determinación experimental y teórica de la hidrofobicidad.

HERRAMIENTAS PARA EL DISEÑO DE COMPUESTOS BIOACTIVOS Y ANÁLISIS CUANTITATIVO DE RELACIÓN ESTRUCTURA QUÍMICA-BIORESPUESTA (Q.S.A.R.)

PROGRAMA

3. ANÁLISIS CUANTITATIVO DE RELACIÓN ESTRUCTURA QUÍMICA-BIORESPUESTA.

3.2.2. Descriptores electrónicos. Constantes de Hammett (σ), constantes de Taft (σ^* , σ^0), constantes de Charton, constantes de Swain-Lupton (F y R). Otros descriptores: momento dipolar, pKa, enlaces de hidrógeno, desplazamiento químico (RMN), frecuencias (IR).

3.2.3. Descriptores estéricos. Constante de Taft (E_s), refractividad molar (MR), parámetros STERIMOL.

3.2.4. Variables indicadoras.

3.2.5. Otras variables independientes. Constantes de equilibrio o cinéticas, energía de HOMO, energía de LUMO, dureza, índices de conectividad, peso molecular, área, propiedades de la superficie molecular, log D.

3.3. Parámetros Estadísticos. Parámetros en las ecuaciones de regresión: coeficiente de correlación, coeficientes de las variables, desviación estandar, errores estándar de los coeficientes, valores F.

3.4. Modelos Q.S.A.R. en Uso. 5.1. Aproximación extratermodinámica. 5.2.

Aproximación *de novo*. 5.3. Elección de la mejor ecuación de correlación. 5.4.

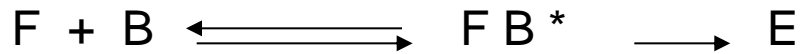
Q.S.A.R. tridimensional: análisis comparativo de campos moleculares.

3.5. Otras Correlaciones. Relaciones estructura química-biodisponibilidad, estructura química-toxicidad.

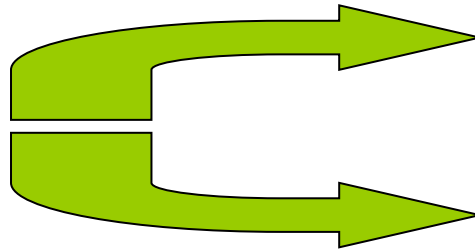
4. EJEMPLOS Y EJERCICIOS.

3.1. PARÁMETROS BIOLÓGICOS

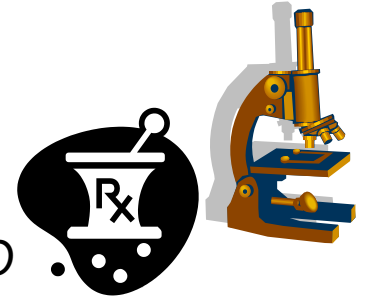
$$\text{Log (1/C)} = k_1 \sigma + k_2 (\log P) - k_3 (\log P)^2 + k_4 E_s + k_5$$



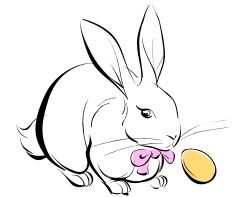
MEDIDA DE EFECTO



IN VITRO



IN VIVO



MEDIDA DEL EFECTO

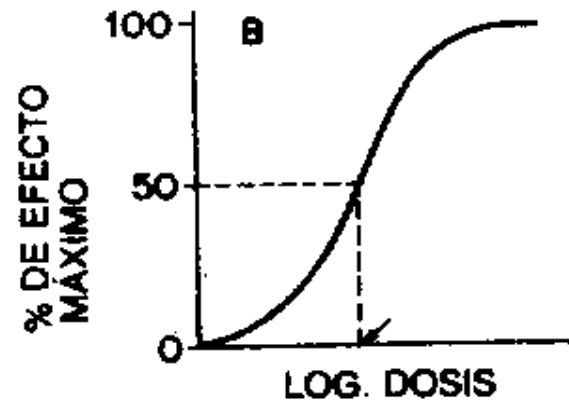
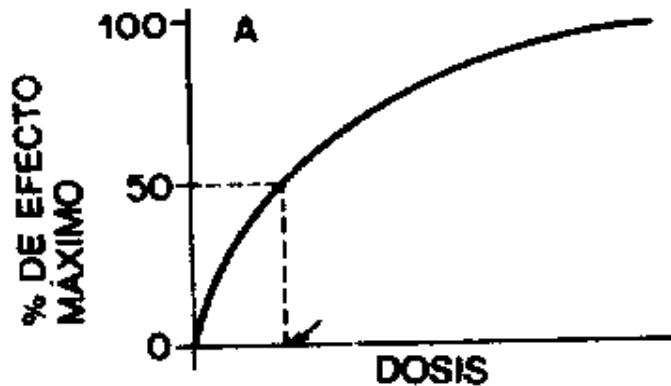
Actividad relativa o actividad absoluta	Porcentaje de inhibición de actividad enzimática
	Porcentaje de inhibición de crecimiento
	Medida de generación de productos
	Cambio en peso, tamaño y color (por ej. hojas, tallos y raíces)
	Medida de parámetros fisiológicos

Formas de determinar la actividad biológica:

- 1) A concentración o dosis fija
- 2) A concentración o dosis variable

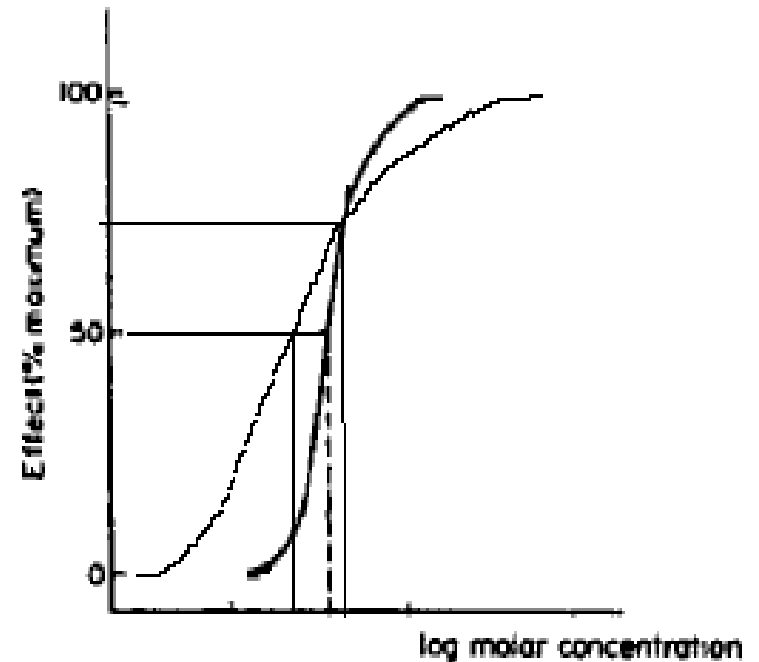
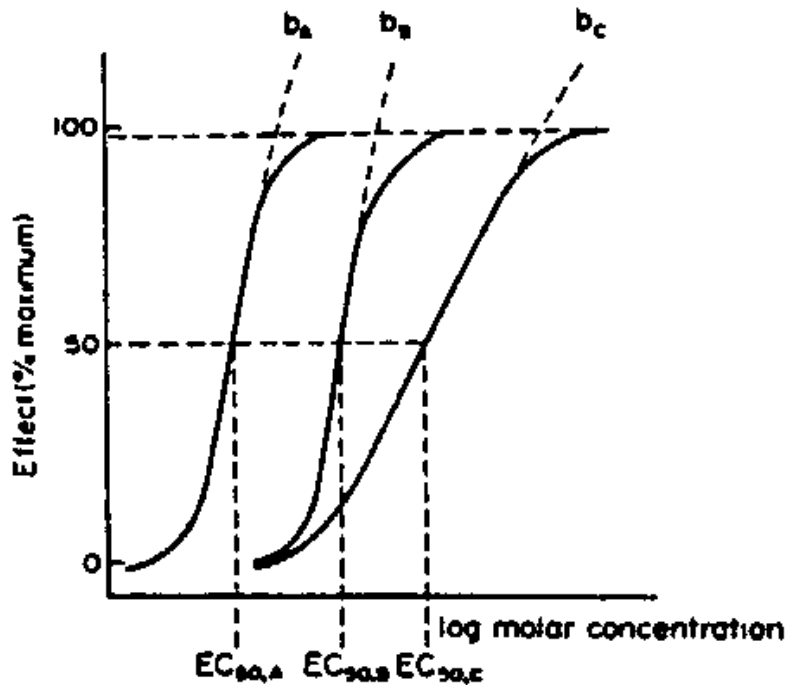
} Variable tiempo

A CONCENTRACIÓN VARIABLE



MEDIDA A CONCENTRACIÓN VARIABLE

- Es preferida dado que:
- 1) permite ver semejanza en mecanismos de acción,
 - 2) “minimiza errores” de dosis fija,
 - 3) “normaliza”



Parámetros biológicos usados en QSAR: ID_{50} , IC_{50} , ED_{50} , LD_{50} , etc.

1. Isolated receptors

Rate constants

Log k

Michaelis-Menten

Log $1/K_m$

Constants

Inhibition constants

Log $1/K_i$

Affinity data

pA_2 ; pA_1

2. Cellular systems

Inhibition constants

Log $1/IC_{50}$

Cross resistance

Log CR

In vitro biological data

Log $1/C$

Mutagenicity states

Log TA_{98}

3. In vivo systems

Biocentration factor

Log BCF

In vivo reaction rates

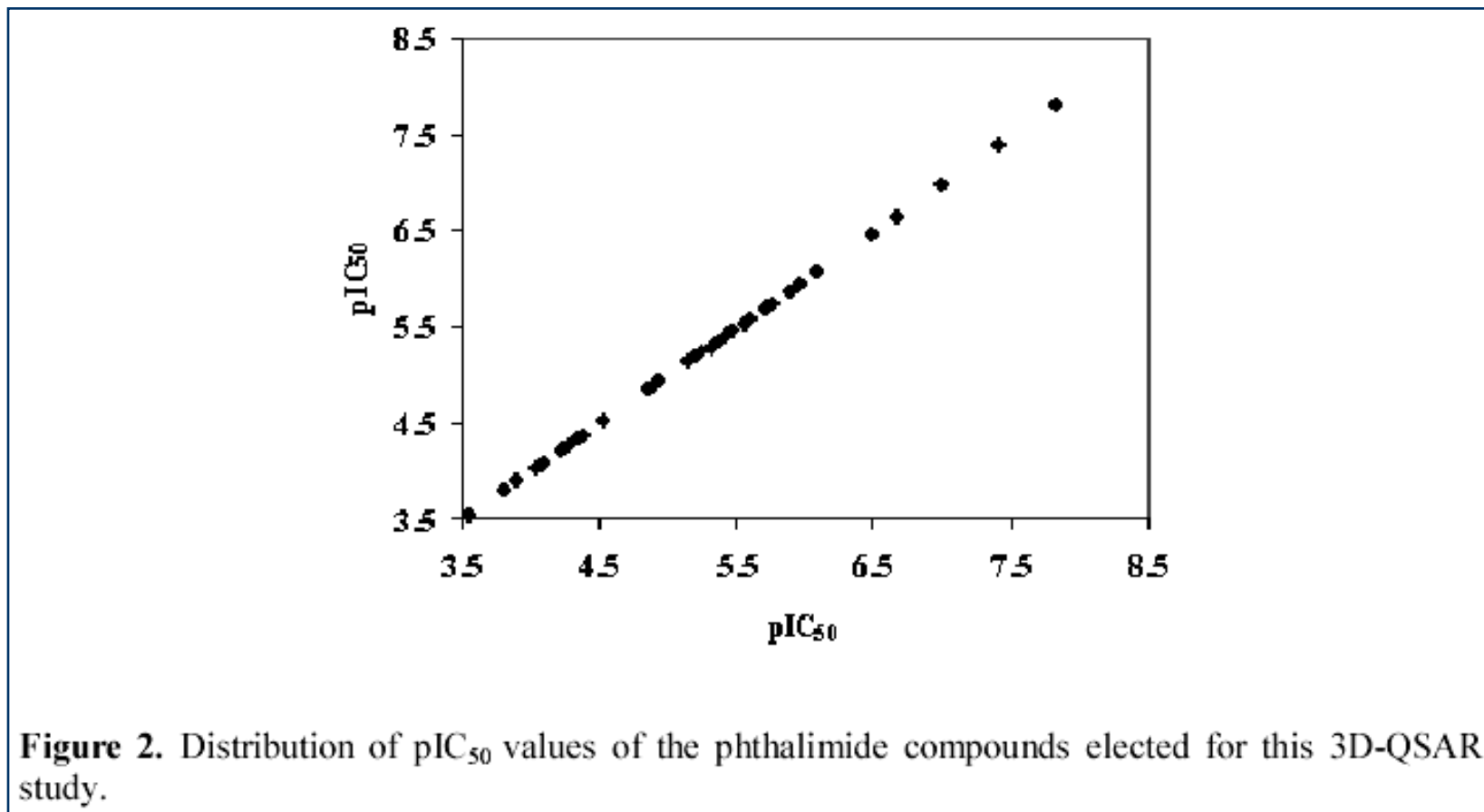
Log I (Induction)

Pharmacodynamic

Log T (total clearance)

Rates

Los parámetros biológicos usados en un QSAR deben cubrir el mayor rango posible de actividades estudiadas



3.2.4. VARIABLES INDICADORAS

$$\text{Log } (1/C) = k_1 (\log P) - k_2 (\log P)^2 + k_3 \sigma + k_4 E_s + k_5$$

Los términos propuestos en la aproximación extratermodinámica algunas veces pueden resultar insuficientes, o pueden no describir completamente a toda la familia en estudio



Hidrólisis de ésteres por papaina

X	R	log 1/K _m	MR	σ
4-OH	SO ₂ Me	2.05	0.28	-0.37
4-OMe	SO ₂ Me	2.13	0.79	-0.27
4-Me	SO ₂ Me	2.08	0.56	-0.17
3-Me	SO ₂ Me	2.23	0.56	-0.07
H	SO ₂ Me	1.79	0.10	0.00
4-F	SO ₂ Me	1.95	0.09	0.06
		2.29	0.79	0.12
		2.33	0.69	0.42
		2.38	0.60	0.23
		1.98	0.09	0.34
		2.57	1.12	0.50
		2.53	0.74	0.71
		2.71	0.74	0.78
4-NH ₂	COC ₆ H ₅	3.58	0.54	-0.66
4-Me	COC ₆ H ₅	4.02	0.56	-0.17
H	COC ₆ H ₅	3.77	0.10	0.00
4-Cl	COC ₆ H ₅	4.00	0.60	0.23
4-F	COC ₆ H ₅	3.69	0.09	0.06
3-NO ₂	COC ₆ H ₅	4.74	0.74	0.71
4-NO ₂	COC ₆ H ₅	4.85	0.74	0.78

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

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

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

$$(n = 7; r = 0.971; s = 0.148; F = 32.85)$$

3.2.4. VARIABLES INDICADORAS

$$\text{Log } (1/C) = k_1 (\log P) - k_2 (\log P)^2 + k_3 \sigma + k_4 E_s + k_5$$

Otras veces pueden existir una propiedad característica en el sistema en estudio para la que no existe un descriptor fisicoquímico (de los ya descritos)

enlace de hidrógeno intramolecular
dador o aceptor de EDH
efecto *orto*
isomerismo *cis/trans*
diferentes patrones estructurales, etc

EN ESTOS CASOS SE UTILIZARÁN EN EL ANÁLISIS LO QUE LLAMAMOS
VARIABLES INDICADORAS

ESTAS VARIABLES QUE INDICAN
PRESENCIA/AUSENCIA, POSICIONES, TIPO Y NÚMERO DE SUSTITUYENTES,
ETC.

ADOPTAN VALORES NUMÉRICOS:

1 / 0

2 / 1 / 0,5 / 0

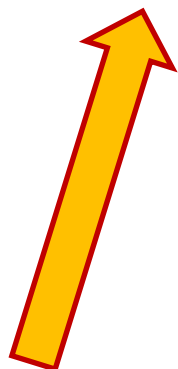
3.2.4. VARIABLES INDICADORAS

$$\log 1/K_m = 0.569(\pm 0.26) MR + 0.561(\pm 0.19) \sigma - 1.922(\pm 0.15) I + 3.743(\pm 0.17)$$

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



Hidrólisis de ésteres por papaina



X	R	log 1/K _m	MR	σ
4-OH	SO ₂ Me	2.05	0.28	-0.37
4-OMe	SO ₂ Me	2.13	0.79	-0.27
4-Me	SO ₂ Me	2.08	0.56	-0.17
3-Me	SO ₂ Me	2.23	0.56	-0.07
H	SO ₂ Me	1.79	0.10	0.00
4-F	SO ₂ Me	1.95	0.09	0.06
		2.29	0.79	0.12
		2.33	0.69	0.42
		2.38	0.60	0.23
		1.98	0.09	0.34
		2.57	1.12	0.50
		2.53	0.74	0.71
		2.71	0.74	0.78
4-NH ₂	COC ₆ H ₅	3.58	0.54	-0.66
4-Me	COC ₆ H ₅	4.02	0.56	-0.17
H	COC ₆ H ₅	3.77	0.10	0.00
4-Cl	COC ₆ H ₅	4.00	0.60	0.23
4-F	COC ₆ H ₅	3.69	0.09	0.06
3-NO ₂	COC ₆ H ₅	4.74	0.74	0.71
4-NO ₂	COC ₆ H ₅	4.85	0.74	0.78

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

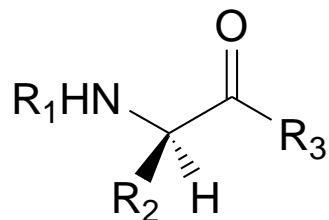
$$(n = 13; r = 0.935; s = 0.105; F = 34.51)$$

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

$$(n = 7; r = 0.971; s = 0.148; F = 32.85)$$

EJEMPLOS:

Hidrólisis de ésteres de aminoácidos por *Quimiotripsina*



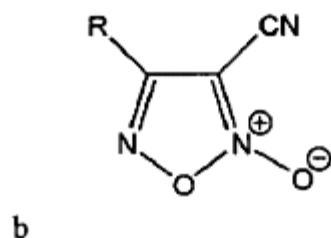
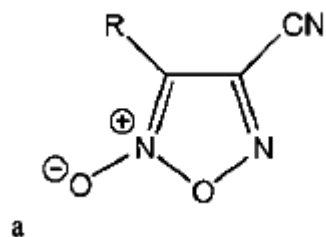
$$\log_{10} (1/K_m) = 0.47 (\pm 0.31) MR_2 + 1.38 (\pm 0.59) MR_1 + 1.38 (\pm 0.72) I + 2.76 (\pm 1.9)$$

$$n = 15, r^2 = 0.986, s = 0.267$$

I = toma el valor 1 en el caso que $\text{R}_3 = \text{OC}_6\text{H}_4\text{-4}'\text{-NO}_2$

toma el valor 0 en caso contrario

EJEMPLOS:



No	R	obsd. log % NO ₂
1	C ₆ H ₅	0.780
2	C ₆ H ₅	1.600
3	CH ₃	0.000
4	CH ₃	1.560
5	n-C ₃ H ₇	0.300
6	n-C ₃ H ₇	1.400
7	CH ₂ N(CH ₃) ₂	1.750
8	CH ₂ SCH ₂ CH ₂ N(CH ₃) ₂	1.580
9	CH ₂ SCH ₂ CH ₂ N(CH ₃) ₂	1.70
10	O C ₂ H ₅	1.790
11	OC ₆ H ₅	1.970

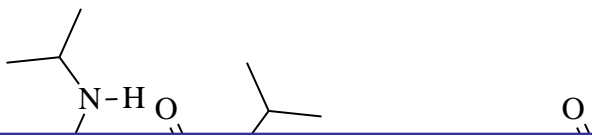
$$\log \% \text{NO}_2^- = 1.127 (0.408) I_B + 1.54 (1.428) F_R + 0.370 (0.296)$$

$$n = 8, r = 0.973, q^2 = 0.863, s = 0.199$$

Gasco AM, Boschi D, Di Stilo A, Medana C, Gasco A, Martorana PA, Schönafinger K. Characterisation of furoxan carbonitriles as a new class of vasodilators. *Arzneimittel. forschung/Drug Res* 1998;48:212 p.

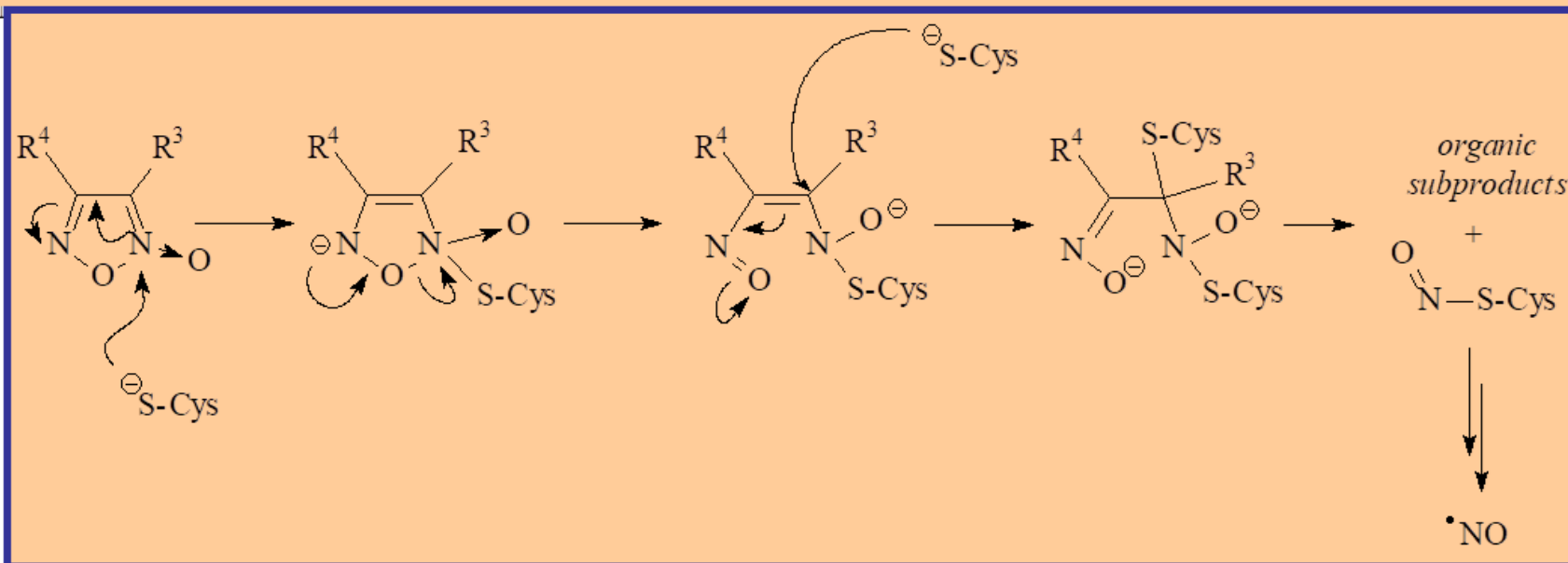
EJEMPLOS:

Capacidad de furoxanos en producir $\cdot\text{NO}$ en diferentes condiciones



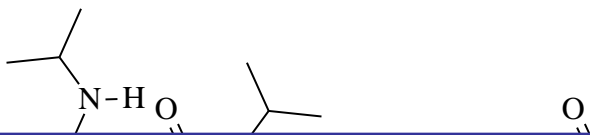
$$\text{Log}_{10}[\% \text{NO}_2^-]_{\text{Cys}, 1\text{h}} = -6.54(\pm 1.78) - 1.12(\pm 0.34) I_{\text{GEW}} + 3.29(\pm 0.80) q_{\text{N}_2} - 1.09(\pm 0.48) q_{\text{N}_5}$$
$$r = 0.818, r^2_{\text{adj.}} = 0.569, n = 14, s = 0.464, p = 0.0092$$

I_{GEW} : variable indicadora que indica la presencia de un buen grupo electrón atrayente en el sustituyente de la posición 3- o 4- del furoxano, tomando el valor 1 cuando es bueno ó 0 cuando no lo es



EJEMPLOS:

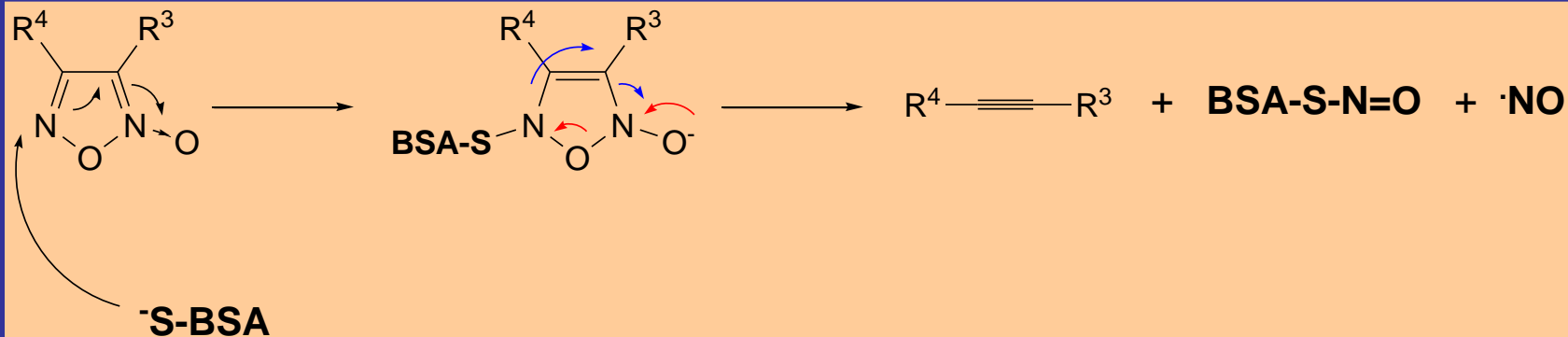
Capacidad de furoxanos en producir $\cdot\text{NO}$ en diferentes condiciones



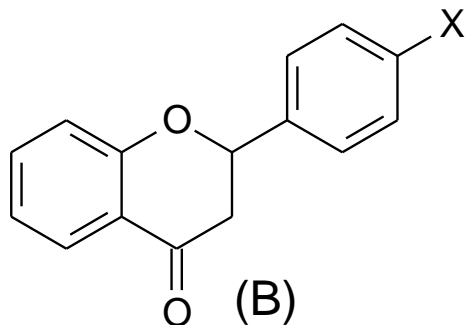
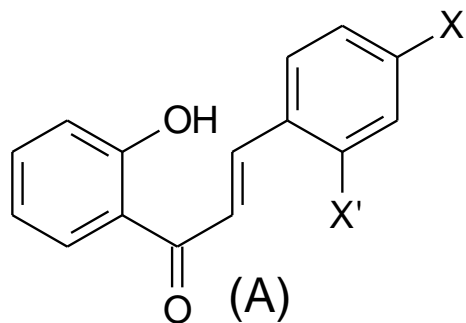
$$\text{Log}_{10}[\% \text{NO}_2^-]_{\text{Cys}, 1\text{h}} = -6.54(\pm 1.78) - 1.12(\pm 0.34) I_{\text{GEW}} + 3.29(\pm 0.80) q_{\text{N}_2} - 1.09(\pm 0.48) q_{\text{N}_5}$$
$$r = 0.818, r^2_{\text{adj.}} = 0.569, n = 14, s = 0.464, p = 0.0092$$

Ph
N

$$\text{Log}_{10}[\% \text{NO}_2^-]_{\text{BSA}, 24\text{h}} = -3.56(\pm 1.14) - 1.65(\pm 0.55) q_{\text{N}_5} + 1.78(\pm 0.43) \text{cLogP}$$
$$r = 0.847, r^2_{\text{adj.}} = 0.636, n = 10, s = 0.452, p = 0.0120$$



EJEMPLOS:



Tipo (A)		% Supervivencia células TK10		Tipo (B)	
X	X'			X	X'
H	H	0.1	79.4	H	H
OMe	H	4.0	35.5	OMe	H
NMe ₂	H	76.2	-	-	-
OH	H	4.7	-	-	-
H	Br	3.4	54.2	Br	H
NO ₂	H	34.0	-	-	-
OMe	OMe	1.9	68.6	OMe	OMe

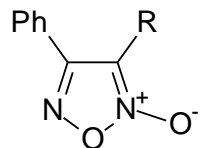
$$\log_{10} (\% \text{ Sup.}_{\text{TK10}}) = 0.04 + 2.8 \sigma_p^2 + 1.6 \text{ PC} \quad r = 0.904$$

PC = Presencia de ciclo (1)

En caso contrario 0

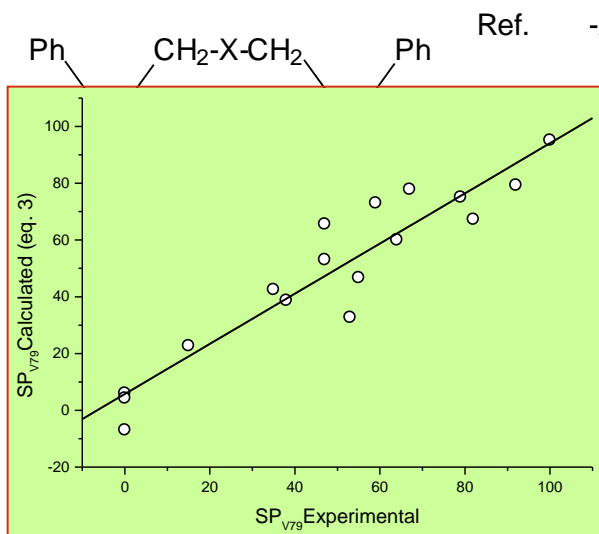
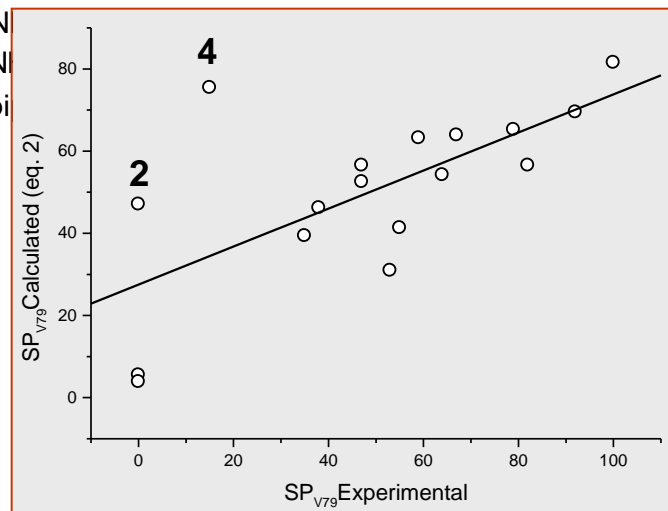
EJEMPLOS:

Citotoxicidad de furoxanos en oxia



Ref.	-R
1	-CHO
2	-CH ₂ Cl
3	-CN
4	-CH ₂ SPh
5	-CH=NOH
6	-CH ₂ -(tiomorfolina-4-yl)
7	-CH=NNH(CO)NH-β-naphtyl

Ref.	-R
8	-CH=NNH(CO)N
9	-CH=NNH(CS)N
10	-CH ₂ -(1-phenylpi)
11	-CH ₂ OH
12	-CH ₂ SCH ₃
13	-CH ₂ SOCH ₃
14	-CH ₂ SO ₂ CH ₃



$$SP_{V79} = -377 (\pm 323) q_{N2} + 45 (\pm 20) E_{LUMO} + 8 (\pm 4) cLogP$$

$$r^2 = 0.458, r^2_{adj} = 0.333, n = 17, p < 0.041$$

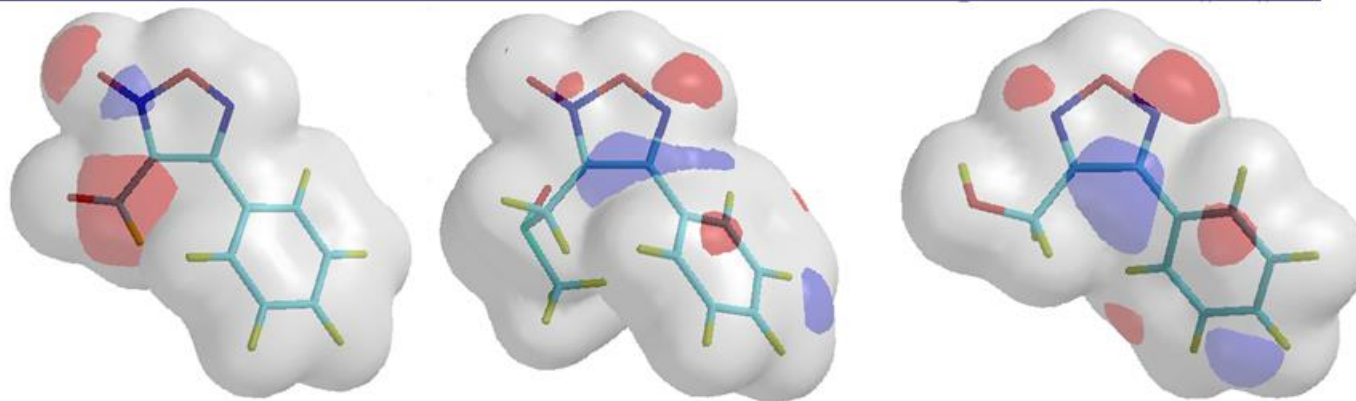
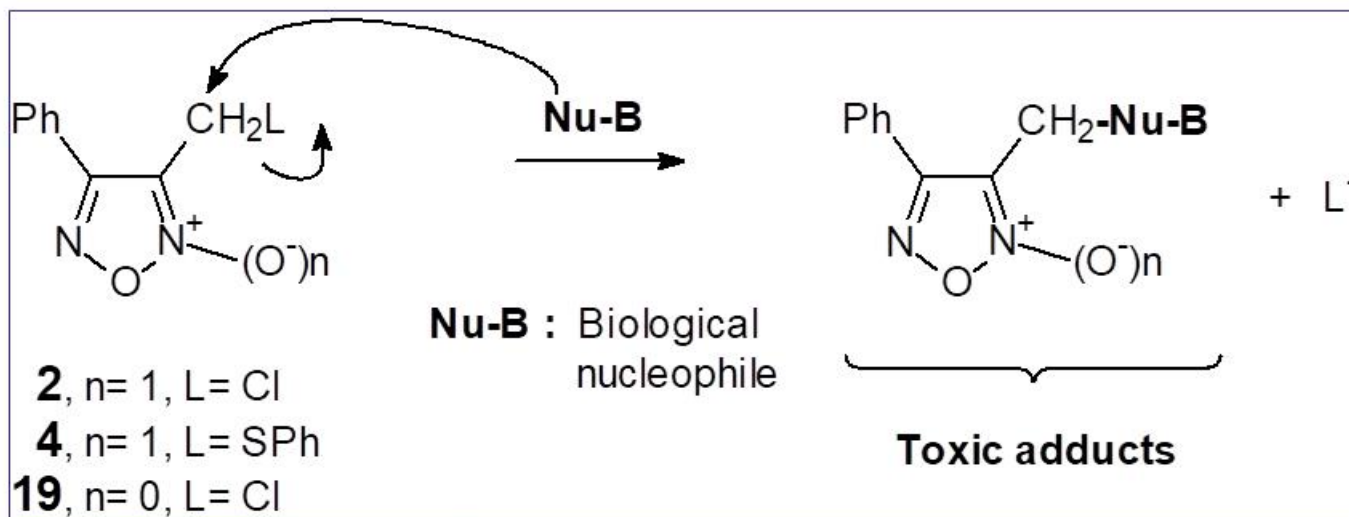


$$SP_{V79} = -35 (\pm 29) - 256 (\pm 156) q_{N2} + 60 (\pm 10) E_{LUMO} + 9 (\pm 2) cLogP - 64 (\pm 10) I_{GLG}$$

$$r^2 = 0.885, r^2_{adj} = 0.846, n = 17, p < 0.0001$$

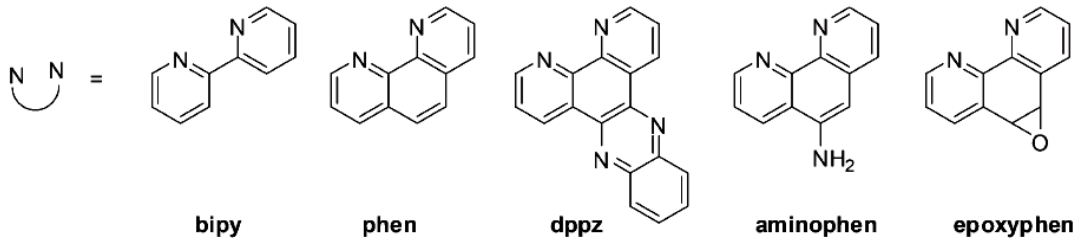
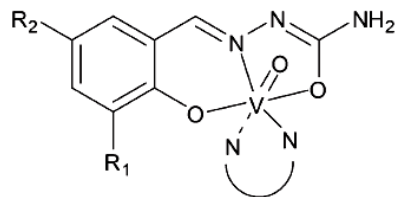
EJEMPLOS:

Citotoxicidad de furoxanos en oxia

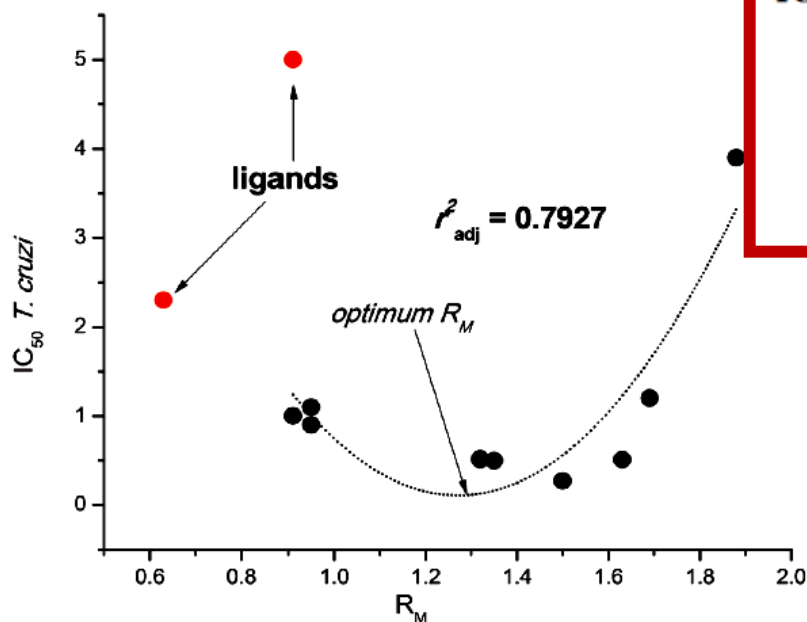


$$SP_{V79} = -35 (\pm 29) - 256 (\pm 156) q_{N2} + 60 (\pm 10) E_{LUMO} + 9 (\pm 2) cLogP - 64 (\pm 10) I_{GLG}$$
$$r^2 = 0.885, r^2_{adj} = 0.846, n = 17, p < 0.0001$$

EJEMPLOS:



I_{col} **1** **2** **2** **3** **4**



$$IC_{50, T. cruzi} = (506 \pm 97) - (592 \pm 137)R_M + (205 \pm 51)R_M^2 - (31 \pm 4) I_{col}$$

$$r_{adj}^2 = 0.7847, n = 22, F = 26.52, p = 8 \times 10^{-7}$$

RSC Publishing

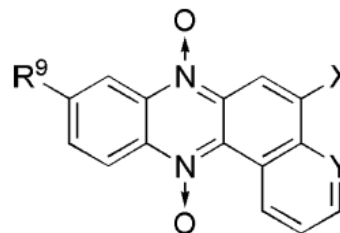
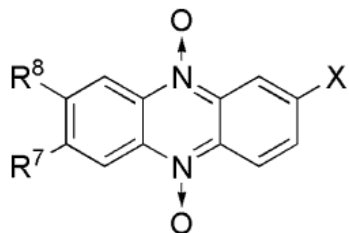
[View Article Online](#)
[View Journal](#)

A new series of heteroleptic oxidovanadium(IV) compounds with phenanthroline-derived co-ligands: selective *Trypanosoma cruzi* growth inhibitors

Mariana Fernández,^{†a} Javier Varela,^{†b} Isabel Correia,^c Estefanía Birriel,^b Jorge Castiglioni,^d Virtudes Moreno,^e Joao Costa Pessoa,^c Hugo Cerecetto,^b Mercedes González^b and Dinorah Gambino^{*a}

Fig. 5 IC_{50} values (in μM units) for the anti-*T. cruzi* activity of the compounds vs. R_M values as a measure of the lipophilicity of the compounds.

EJEMPLOS:



$$\text{SFhyp} = -179(\pm 91) + 75(\pm 6)IV_{\text{OH}} + 54(\pm 74)$$

$$- 727(\pm 203)q_{N_5/N_{12}} - 0.35(\pm 0.45) (\text{cLogP})^2$$

$$\text{Outliers : 9 and 14} \quad n = 28, r = 0.8745, r_{\text{adj}}^2 = 0.8527,$$

$$F = 40.1, p < 1 \times 10^{-9}$$

Chem Biol Drug Des 2011; **78**: 960–968

© 2011 John Wiley & Sons A/S
doi: 10.1111/j.1747-0285.2011.01237.x

Research Article

2D- and 3D-Quantitative Structure-Activity Relationship Studies for a Series of Phenazine *N,N'*-Dioxide as Antitumour Agents

Jonathan Da Cunha¹, María Laura Lavaggi¹, María Inés Abasolo^{2,*}, Hugo Cerecetto^{1,*} and Mercedes González^{1,*}

EJEMPLOS:

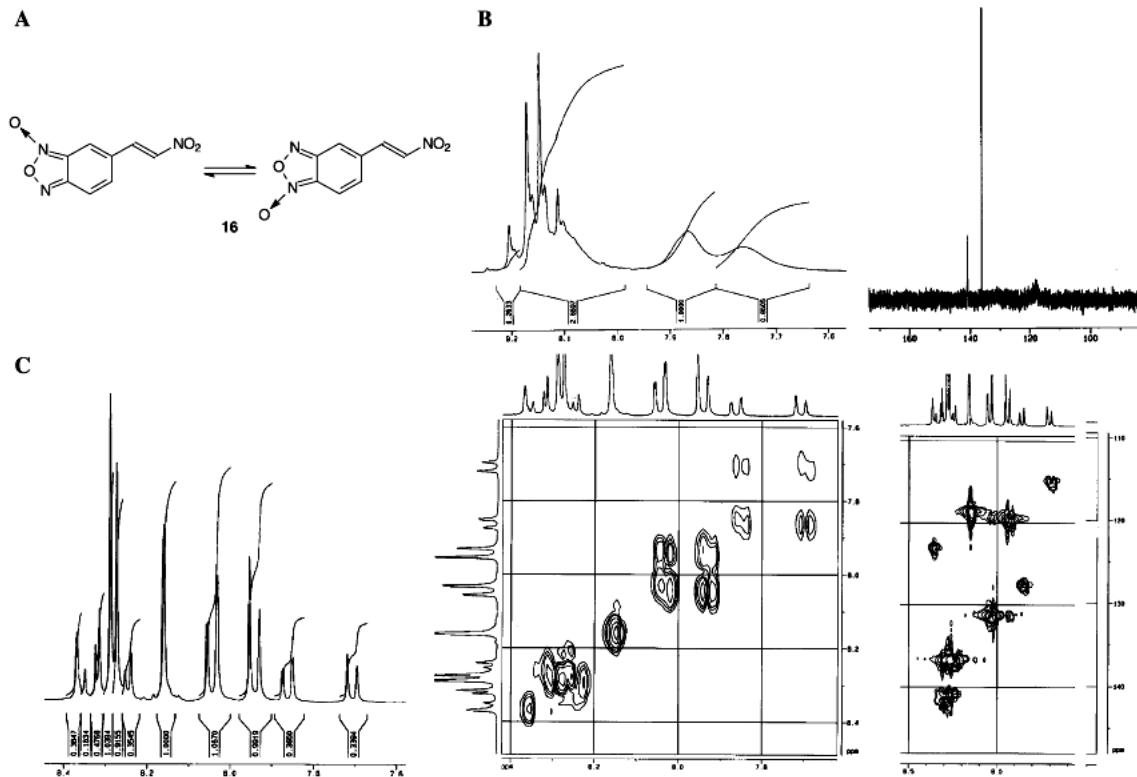
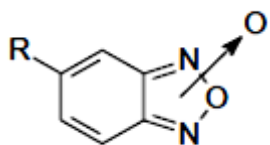
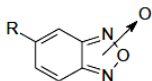


Figure 1. (A) Tautomeric equilibrium of derivative 16. (B) ^1H and ^{13}C NMR spectra of 16 in acetone- d_6 at 303 K. (C) ^1H NMR, COSY, and HMQC spectra of 16 in acetone- d_6 at 230 K.

Table 1. ^1H NMR data at low temperature



Compound	-R	Temperature (K) ^a /solvent	6-Tautomer/5-tautomer ratio
16	-E-CH=CHNO ₂	230/CD ₃ COCD ₃	2.90
17	-NO ₂	233/CDCl ₃	2.70
10	-CHO	233/CDCl ₃	1.90
18	-CH=NOH	233/CD ₃ COCD ₃	1.60
3	-E-CH=CHPh	233/CDCl ₃	1.38
19	-CH ₂ I	233/CDCl ₃	1.33
20	-CH ₂ Cl	233/CDCl ₃	1.02
21	-CH ₂ OH	263/CD ₃ COCD ₃ ^b	1.00 ^b
22	-Ph	243/CDCl ₃	0.95 ^c
23	-CH ₃	227/CDCl ₃ ^d	0.92 ^d
24	-OCH ₃	248/CDCl ₃	0.56

^a Conditions used in NMR experiments.

^b Data taken from Ref. 9a.

^c Ratio was estimated since signals were not resolved.

^d Data taken from Ref. 8c.



VIER

Available online at www.sciencedirect.com

SCIENCE @ DIRECT®

Bioorganic & Medicinal Chemistry 13 (2005) 6336–6346

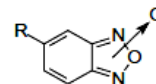
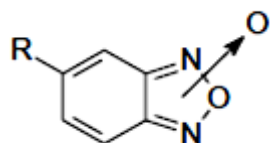
Bioorganic & Medicinal Chemistry

New potent 5-substituted benzofuroxans as inhibitors of *Trypanosoma cruzi* growth: Quantitative structure–activity relationship studies

Gabriela Aguirre,^a Lucía Boiani,^a Mariana Boiani,^a Hugo Cerecetto,^{a,*} Rossanna Di Maio,^a Mercedes González,^{a,*} Williams Porcal,^a Ana Denicola,^b Oscar E. Piro,^c Eduardo E. Castellano,^d Carlos Mauricio R. Sant'Anna^e and Eliezer J. Barreiro^f

Table 3. Benzofuroxan derivatives and substituent descriptors used to derive Eq. (1) of the 2D-QSAR model

EJEMPLOS:



Compound	-R	log ₁₀ (% GI) ^a	I ₆ ^b	MR ^c	F	I _{HBD}
1	-H	1.54	0	0.10	0.00	0
2	-N(CH ₃) ₂	1.41	0	1.56	0.15	0
3	-E-CH=CH-Ph	1.99	0.5	3.42	0.10	0
5	-CO ₂ H	0.70	0.5	0.69	0.34	1
8	-NHCO ₂ C(CH ₃) ₃	1.26	0	3.05 ^f	0.13	1
10	-CHO	1.83	1	0.69	0.33	0
16	-E-CH=CH-NO ₂	1.96	1	1.64	0.35	0
17	-NO ₂	1.95	1	0.74	0.65	0
18	-CH=NOH	1.58	1	1.03	0.28	1
19	-CH ₂ I	1.97	0.5	1.86	0.12	0
20	-CH ₂ Cl	1.99	0.5	1.05	0.13	0
21	-CH ₂ OH	1.32	0.5	0.72	0.03	1
22	-Ph	1.49	0	2.54	0.12	0
23	-CH ₃	1.34	0	0.56	0.01	0
24	-OCH ₃	1.00	0	0.79	0.29	0
25	-CH=C(CN) ₂	1.82	1	1.97	0.57	0
26	-CN	0.95	1	0.63	0.51	0
27	-CH=NNHC(S)NH ₂	1.23	0.5	2.96	0.46	1
28	-SO ₂ NH ₂	0.85	0.5	1.23	0.49	1
29	-Cl	0.70	0	0.60	0.42	0
30	-Br	0.70	0	0.89	0.45	0

^a % GI: percentage of *T. cruzi* growth inhibition at 25 μM.

^b I₆: indicator variable that takes value 1 when the substituent is preferentially 6-positioned with respect to *N*-oxide moiety, value 0 when it is preferentially 5-positioned and value 0.5 when in equilibrium there exists an equal quantity of 5- and 6-positioned tautomers at low temperature.

^c To autoscale the independent variables, the MR reported is MR_{lit}/10, being MR_{lit} the value taken from Ref. 16b.



SCIENCE @ DIRECT®

Bioorganic &
Medicinal
Chemistry

Journal of Medicinal Chemistry 13 (2005) 6336–6346

$$\log_{10}(\% \text{ GI}) = 1.31(\pm 0.10) + 1.0(\pm 0.1)I_6 + 0.15(\pm 0.04)MR - 1.31(\pm 0.24)F - 0.51(\pm 0.09)I_{\text{HBD}} \quad (1)$$

$$n = 20, r = 0.939, r_{\text{adj}}^2 = 0.849, s = 0.184, \text{ and } p < 0.0001$$

6-Substituted benzofuroxans as inhibitors of *T. cruzi* growth: Quantitative structure-activity relationship studies

Gaibriela Aguirre,^a Lucía Boiani,^a Mariana Boiani,^a Hugo Cerecetto,^{a,*} Rossanna Di Maio,^a Mercedes González,^{a,*} Williams Porcal,^a Ana Denicola,^b Oscar E. Piro,^c Eduardo E. Castellano,^d Carlos Mauricio R. Sant'Anna^e and Eliezer J. Barreiro^f

3.2.5. OTRAS VARIABLES INDEPENDIENTES

$$\text{Log } (1/C) = k_1 \sigma + k_2 (\log P) - k_3 (\log P)^2 + k_4 E_s + k_5$$

- pK_a
- Propiedades electroquímicas
- E_{HOMO} , contribución atómica a los orbitales moleculares frontera
- E_{LUMO} , contribución atómica a los orbitales moleculares frontera
- Dureza ($E_L - E_H / 2$), GAP ($E_L - E_H$)
- Cargas atómicas, densidad de carga, módulo del momento dipolar, etc.
- Índices de conectividad (topología)
- Solubilidades
- Datos espectroscópicos (IR, RMN, UV-visible)

EJEMPLOS DE PARÁMETROS FRAGMENTARIOS:

Símbolo	Descripción	Tipo
π	Constante hidrofóbica de sustituyente	h
σ	Constante de Hammett	e
F, R	Constantes de efecto inductivo y resonante	e
E_s	Parámetro estérico de Taft	s
MR	Refractividad molar del sustituyente	s

EJEMPLOS DE PARÁMETROS MOLECULARES:

Símbolo	Descripción	Tipo
$\log P$	Logaritmo del coeficiente de reparto octanol-agua	h
MR	Refractividad molar molecular	s
pK_a	Constante de ionización	e
μ	Momento dipolar	e
δ	Desplazamiento químico (RMN)	e
ν	Frecuencias (IR)	e
K	Constantes de equilibrio o cinéticas	-

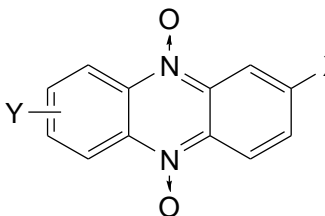
OTROS PARÁMETROS:

Símbolo	Descripción	Tipo
E_{HOMO}	Energía del orbital molecular ocupado de mayor energía	e
E_{LUMO}	Energía del orbital molecular desocupado de menor energía	e
Cargas efectivas sobre los átomos, densidad electrónica		e
Potencial de óxido-reducción		e
PM	Peso molecular	s
A	Área	s
Índice de conectividad de Kier		s

3.2.5. OTRAS VARIABLES INDEPENDIENTES

Ej.

-Y -X = -NH, -O



$$SF_{\text{hypoxia}} = 936(\pm 258) + 239(\pm 62) E_{pc} - 331(\pm 119) c\text{LogP} + 40(\pm 16) c\text{LogP}^2$$

-H , 8 , 9

-CH₃ , 10, 11

-OCH₃, 12, 13

-Br , 14, 15

-Cl , 16, 17

<i>r</i>	<i>r</i> ² <i>adj</i>	<i>s</i>	<i>p</i>	<i>F</i> value	<i>n</i>
0.875	0.665	26.74	0.013	7.63	11

Used compounds: 7 – 17

H. Cerecetto, M. González, M. L. Lavaggi, M. A. Aravena, C. Rigol, C. Olea-Azar, A. Azqueta, A. López de Cerain, A. Monge, A. M. Bruno. Phenazine 5,10-dioxide derivatives as hypoxic selective cytotoxins: Part II. Structure-activity relationship studies. *Med. Chem.*, 2006, 2, 511-521.

- *pKa*
- Propiedades electroquímicas
- E_{HOMO} , contribución atómica a los orbitales moleculares frontera
- E_{LUMO} , contribución atómica a los orbitales moleculares frontera
- Dureza ($E_{\text{L}} - E_{\text{H}} / 2$), GAP ($E_{\text{L}} - E_{\text{H}}$)
- Cargas atómicas, densidad de carga, módulo del momento dipolar, etc.
- Índices de conectividad (topología)
- Solubilidades
- Datos espectroscópicos (IR, RMN, UV-visible)

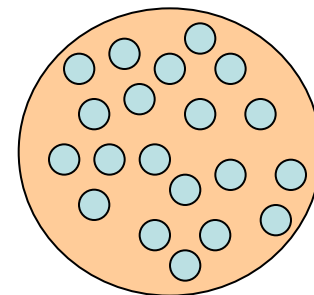
3.3. PARÁMETROS ESTADÍSTICOS

$$\text{Log } (1/C) = k_1 \sigma + k_2 (\log P) - k_3 (\log P)^2 + k_4 E_s + k_5$$

¿CÓMO SABEMOS QUÉ PARÁMETRO(S) DE LA ECUACIÓN DE HANSCH TIENE(N) REALMENTE RELEVANCIA?

• Cada parámetro debe ser **estadísticamente validado (estadísticamente significativo)** para poder ser utilizado con seguridad en la ecuación de correlación entre la actividad y las propiedades fisicoquímicas de las moléculas

• **CONSIDERACIÓN:** en **todos** los experimentos biológicos (aunque estén muy bien diseñados) existe una gran variabilidad en la respuesta por lo que se debe recurrir siempre al **TRATAMIENTO ESTADÍSTICO DE DATOS**



PARÁMETROS ESTADÍSTICOS MÁS USUALES PARA EL MANEJO DE ACTIVIDADES BIOLÓGICAS:

- Consideraciones:
- Medida sobre varios sujetos (6)
 - Grupos estadísticamente similares (animales con igual peso, células en igual día de crecimiento)
 - Iguales condiciones de medidas

- 1) Como la **Actividad Biológica** varía mucho se usa:
Medida de Actividad = **Media** aritmética

$$\bar{X} = (\sum x)/n$$

donde x = valores de respuesta y n = nº de valores de respuesta

- 2) Como medida de variabilidad entre grupos se usa:
varianza: permite comparar dos grupos o más,

$$s^2 = [\sum (x - \bar{X})^2] / (n - 1)$$

Para valorar la dispersión se define la **desviación estándar**,

$$s = \sqrt{s^2}$$

También se emplea el **error estándar de la Media**,

$$s.e. = s / \sqrt{n}$$

PARÁMETROS ESTADÍSTICOS MÁS USUALES PARA EL MANEJO DE ACTIVIDADES BIOLÓGICAS:

- 3) Test paramétricos: - **test-t** (para comparar medias)
- **test-F** y **ANOVA** (para comparar varianzas).

SE ASUME QUE LA MUESTRA PRESENTA UNA DISTRIBUCIÓN GAUSIANA NORMAL.

O SEA QUE LA RESPUESTA BIOLÓGICA ES UNA VARIABLE CONTINUA Y QUE LA DISPERSIÓN DE LOS DATOS ES SIMÉTRICA POR ENCIMA Y DEBAJO DE LA MEDIA

- 4) Coeficiente de correlación: Se utiliza frecuentemente el análisis de correlación, surgiendo el **coeficiente de correlación r** , para variables continuas.

Para normalizar muchas veces se transforman las **medidas de actividad** en **Log**, en $\sqrt{\quad}$ u **otras funciones**

PARÁMETROS EN LA ECUACIÓN DE REGRESIÓN:

EL ANÁLISIS DE CORRELACIÓN MÁS COMÚN EN VARIABLES QUE SIGUEN UNA DISTRIBUCIÓN NORMAL ES LA **REGRESIÓN LINEAL MÚLTIPLE** O **REGRESIÓN MULTIVARIABLE**

La consideración de la relación entre la actividad biológica y las propiedades fisicoquímicas de los sustituyentes sobre el “cabeza de serie” según el modelo de Hansch puede resolverse aplicando **REGRESIÓN LINEAL MÚLTIPLE**. (programas de análisis estadísticos: ORIGIN®, STATISTIC®)

$$\text{Log (1/C)} = k_1 \log P - k_2 \log P^2 + k_3 \sigma + k_4 E_s + k_5$$

PARÁMETROS EN LA ECUACIÓN DE REGRESIÓN:

EJEMPLO DE ANÁLISIS (MODELO EXTRATERMODINÁMICO DE HANSCH)

$$\text{Log (1/C)} = k_1 \log P - k_2 \log P^2 + k_3 \sigma + k_4 E_s + k_5$$

r mínimo **ACEPTADO** en Q.S.A.R. **0.7** (explica al menos el 50 % de la variación estudiada)

coeficiente de correlación que explica (en este caso) el 6.05 % ($0.246^2 \times 100$) de la variación estudiada con los datos experimentales de actividad biológica

error estándar de los coeficientes de regresión (también puede ser el 95 % del límite de confianza)

Ec. 5

$$\text{Log (1/C)} = 0.015 (\pm 0.021) E_s + 5.972$$

$n=17$, $r=0.246$, $s=0.280$

PARÁMETROS EN LA ECUACIÓN DE REGRESIÓN:

EJEMPLO DE ANÁLISIS (MODELO EXTRATERMODINÁMICO DE HANSCH)

$$\text{Log (1/C)} = k_1 \log P - k_2 \log P^2 + k_3 \sigma + k_4 E_s + k_5$$

el coeficiente de correlación, para esta ecuación que considera sólo la variable electrónica, sigue siendo inadecuado

coeficientes con errores estándar muy altos

por lo que los coeficientes son menos precisos o sea menos contribuyente es la variable

Ec. 4

$$\text{Log (1/C)} = 0.071 (\pm 0.070) \sigma + 2.002$$

Ec. 5

$$\text{Log (1/C)} = 0.015 (\pm 0.021) E_s + 5.972$$

$n=17, r=0.246, s=0.280$

$n=17, r=0.246, s=0.280$

PARÁMETROS EN LA ECUACIÓN DE REGRESIÓN:

EJEMPLO DE ANÁLISIS (MODELO EXTRATERMODINÁMICO DE HANSCH)

$$\text{Log (1/C)} = k_1 \log P - k_2 \log P^2 + k_3 \sigma + k_4 E_s + k_5$$

el coeficiente de correlación mejoró muchísimo

la desviación estándar sube un poco
(la misma debe ser pequeña para
indicar una buena predicción)

Ec. 3

$$\text{Log (1/C)} = 4.981 (\pm 0.994) \log P + 2.416$$

$n=17$, $r=0.895$, $s=0.314$

Ec. 4

$$\text{Log (1/C)} = 0.071 (\pm 0.070) \sigma + 2.002$$

$n=17$, $r=0.246$, $s=0.280$

Ec. 5

$$\text{Log (1/C)} = 0.015 (\pm 0.021) E_s + 5.972$$

$n=17$, $r=0.246$, $s=0.280$

PARÁMETROS EN LA ECUACIÓN DE REGRESIÓN:

EJEMPLO DE ANÁLISIS (MODELO EXTRATERMODINÁMICO DE HANSCH)

$$\text{Log (1/C)} = k_1 \log P - k_2 \log P^2 + k_3 \sigma + k_4 E_s + k_5$$

Ec. 2

$$\text{Log (1/C)} = 3.416 (\pm 0.096) \log P - 0.942 (\pm 0.011) \log P^2 + 6.303$$

$n= 17, r= 0.938, s= 0.141$

Ec. 3

$$\text{Log (1/C)} = 4.981 (\pm 0.994) \log P + 2.416$$

$n= 17, r= 0.895, s= 0.314$

Ec. 4

$$\text{Log (1/C)} = 0.071 (\pm 0.070) \sigma + 2.002$$

$n= 17, r= 0.246, s= 0.280$

Ec. 5

$$\text{Log (1/C)} = 0.015 (\pm 0.021) E_s + 5.972$$

$n= 17, r= 0.246, s= 0.280$

PARÁMETROS EN LA ECUACIÓN DE REGRESIÓN:

EJEMPLO DE ANÁLISIS (MODELO EXTRATERMODINÁMICO DE HANSCH)

$$\text{Log (1/C)} = k_1 \log P - k_2 \log P^2 + k_3 \sigma + k_4 E_s + k_5$$

Ec. 1

$$\text{Log (1/C)} = 2.342 (\pm 0.105) \log P - 0.731 (\pm 0.041) \log P^2 + 0.036 (\pm 0.019) \sigma + 0.195 (\pm 0.176) E_s + 5.816$$

$$n=17, r=0.957, s=0.126$$

Ec. 2

$$\text{Log (1/C)} = 3.416 (\pm 0.096) \log P - 0.942 (\pm 0.011) \log P^2 + 6.303$$

$$n=17, r=0.938, s=0.141$$

Ec. 3

$$\text{Log (1/C)} = 4.981 (\pm 0.994) \log P + 2.416$$

$$n=17, r=0.895, s=0.314$$

Ec. 4

$$\text{Log (1/C)} = 0.071 (\pm 0.070) \sigma + 2.002$$

$$n=17, r=0.246, s=0.280$$

Ec. 5

$$\text{Log (1/C)} = 0.015 (\pm 0.021) E_s + 5.972$$

$$n=17, r=0.246, s=0.280$$

PARÁMETROS EN LA ECUACIÓN DE REGRESIÓN:

EJEMPLO DE ANÁLISIS (MODELO EXTRATERMODINÁMICO DE HANSCH)

$$\text{Log (1/C)} = k_1 \log P - k_2 \log P^2 + k_3 \sigma + k_4 E_s + k_5$$

Ec. 1

$$\text{Log (1/C)} = 2.342 (\pm 0.105) \log P - 0.731 (\pm 0.041) \log P^2 + 0.036 (\pm 0.019) \sigma + 0.195 (\pm 0.176) E_s + 5.816$$

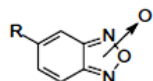
$$n = 17, r = 0.957, s = 0.126$$

los coeficientes de la regresión indican la contribución de la variable a la actividad biológica



no olvidar el orden de magnitud del descriptor usado!!! antes de sacar cualquier conclusión

Table 3. Benzofuroxan derivatives and substituent descriptors used to derive Eq. (1) of the 2D-QSAR model



Compound	-R	log ₁₀ (% GI) ^a	I ₆ ^b	MR ^c	\mathcal{F}	I _{HBD}
1	-H	1.54	0	0.10	0.00	0
21	-CH ₂ OH	1.32	0.5	0.72	0.03	1
22	-Ph	1.49	0	2.54	0.12	0
23	-CH ₃	1.34	0	0.56	0.01	0
24	-OCH ₃	1.00	0	0.79	0.29	0
25	-CH=C(CN) ₂	1.82	1	1.97	0.57	0
26	-CN	0.95	1	0.63	0.51	0
27	-CH=NNHC(S)NH ₂	1.23	0.5	2.96	0.46	1
28	-SO ₂ NH ₂	0.85	0.5	1.23	0.49	1
29	-Cl	0.70	0	0.60	0.42	0
30	-Br	0.70	0	0.89	0.45	0

Vol (Å ³)	logP
430	1.1
790	-0.8
810	0.3
390	2.1
560	3.7

^a % GI: percentage of *T. cruzi* growth inhibition at 25 μM.

^b I₆: indicator variable that takes value 1 when the substituent is preferentially 6-positioned with respect to *N*-oxide moiety, value 0 when it is preferentially 5-positioned and value 0.5 when in equilibrium there exists an equal quantity of 5- and 6-positioned tautomers at low temperature.

^c To autoscale the independent variables, the MR reported is MR_{lit}/10, being MR_{lit} the value taken from Ref. 16b.

PARÁMETROS EN LA ECUACIÓN DE REGRESIÓN:

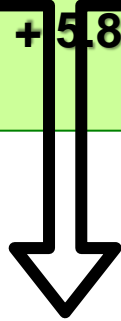
EJEMPLO DE ANÁLISIS (MODELO EXTRATERMODINÁMICO DE HANSCH)

$$\text{Log (1/C)} = k_1 \log P - k_2 \log P^2 + k_3 \sigma + k_4 E_s + k_5$$

Ec. 1

$$\text{Log (1/C)} = 2.342 (\pm 0.105) \log P - 0.731 (\pm 0.041) \log P^2 + 0.036 (\pm 0.019) \sigma + 0.195 (\pm 0.176) E_s + 5.816$$

$$n = 17, r = 0.957, s = 0.126$$



Algo más sobre los errores estándar de los coeficientes de la regresión

Tienen que ser PEQUEÑOS

TAMBIÉN se puede mirar el valor p (p value), que en QSAR se exige un $p < 0.05$, es decir que exista un máximo de 5 % de probabilidad al azar

PARÁMETROS EN LA ECUACIÓN DE REGRESIÓN:

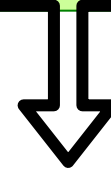
EJEMPLO DE ANÁLISIS (MODELO EXTRATERMODINÁMICO DE HANSCH)

$$\text{Log (1/C)} = k_1 \log P - k_2 \log P^2 + k_3 \sigma + k_4 E_s + k_5$$

Ec. 1

$$\text{Log (1/C)} = 2.342 (\pm 0.105) \log P - 0.731 (\pm 0.041) \log P^2 + 0.036 (\pm 0.019) \sigma + 0.195 (\pm 0.176) E_s + 5.816$$

$$n = 17, r = 0.957, s = 0.126, F = 10.9$$



El análisis de la significancia en la introducción de una nueva variable en la ecuación de regresión puede utilizarse

El valor F mide la capacidad de la ecuación de regresión en rechazar la hipótesis nula

coeficiente de correlación ajustado " r^2_{adj} "

Cuanto **MÁS GRANDE** mejor

este coeficiente no mejora cuando se introduce una nueva variable en una ecuación, salvo que sea significativa estadísticamente



Cuando no hay otros elementos para decidir entre dos ecuaciones se elegirá la ecuación con mayor F

CRITERIOS PARA ABORDAR UN Q.S.A.R.:

1) SELECCIÓN DE VARIABLES INDEPENDIENTES: examinar el mayor número posible de variables ortogonales ($r^2 < 0.5$). Examinar los análisis de regresión con todas las combinaciones de variables independientes.

2) JUSTIFICACIÓN DE LA ELECCIÓN DE VARIABLES INDEPENDIENTES: desviación estándar (s) baja y todos los términos estadísticamente significativo por lo menos al nivel del 95 % ($p < 0.05$).

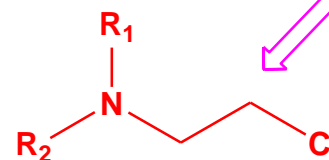
3) PRINCIPIO DE PARSIMONIA: cuando todo los parámetros estadísticos son iguales, se acepta el modelo más simple.

4) NÚMERO DE TÉRMINOS POR VARIABLE INDEPENDIENTE: 5 ó 6 datos experimentales (5 ó 6 compuestos con sus actividades) por variable independiente que se incluya en la ecuación.¹

5) SIGNIFICADO FÍSICO-QUÍMICO ORGÁNICO: no conformarse con obtener una correlación estadísticamente correcta, sino que cada término debería poder ser explicado en términos de actividad biológica.

¹ J. Topliss, R. Costello. *J. Med. Chem.*, 1972, 15, 1068.

5) SIGNIFICADO FISICO-QUÍMICO ORGÁNICO



$$\text{Log (1/C)} = 4.2 (\pm 0.2) V_{R1,R2} - 1.7 (\pm 0.4) (V_{R1,R2})^2 + 3.6 (\pm 0.2) \text{Nucl}_{\text{CH}_2\text{Cl}} + 0.1$$

$n = 150, r = 0.999$

Un término relacionado con la nucleofilia de ese carbono
NO TIENE SIGNIFICANCIA BIOMÉDICA

..... **o sí**



Se debe buscar la explicación

5) SIGNIFICADO FÍSICO-QUÍMICO ORGÁNICO

Ej.



Available online at www.sciencedirect.com

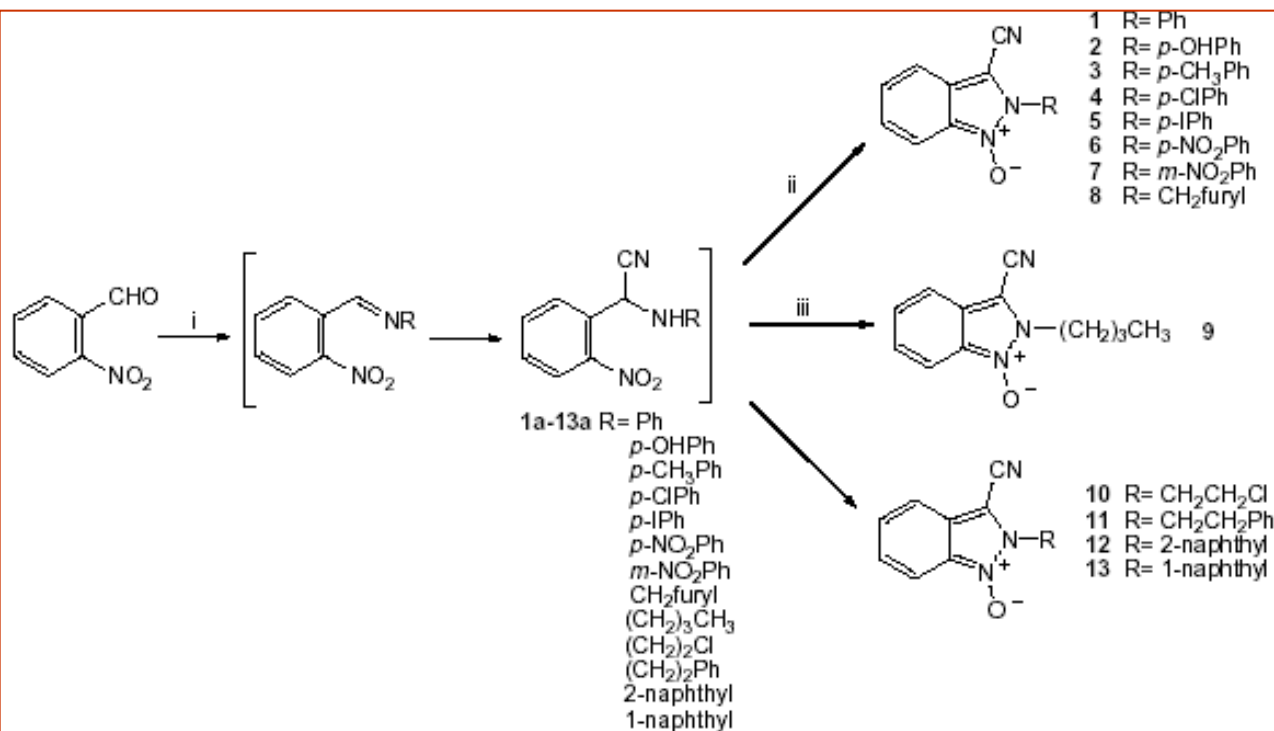
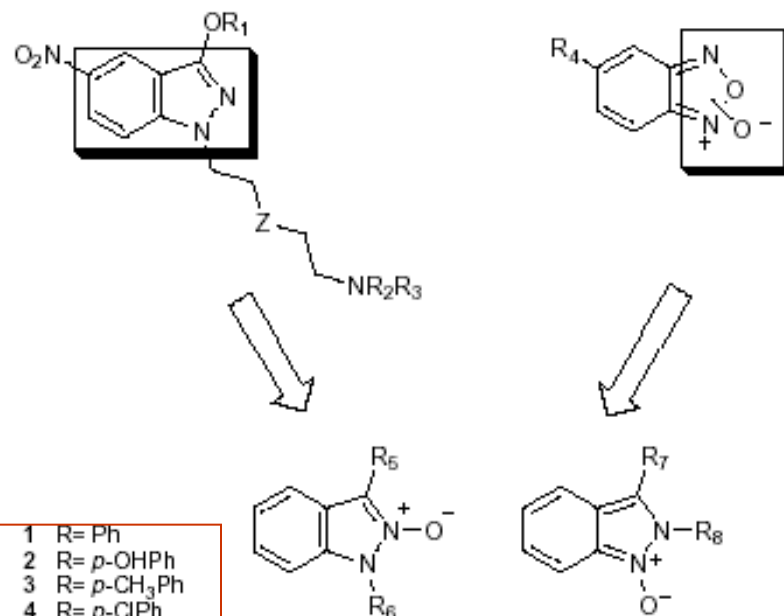
SCIENCE @ DIRECT®

Bioorganic & Medicinal Chemistry 14 (2006) 3467–3480

Bioorganic &
Medicinal
Chemistry

Indazole *N*-oxide derivatives as antiprotozoal agents: Synthesis, biological evaluation and mechanism of action studies

Alejandra Gerpe,^a Gabriela Aguirre,^a Lucía Boiani,^a Hugo Cerecetto,^{a,*}
Mercedes González,^{a,*} Claudio Olea-Azar,^b Carolina Rigol,^b Juan D. Maya,^c
Antonio Morello,^c Oscar E. Piro,^d Vicente J. Arán,^e Amaia Azqueta,^f
Adela López de Ceráin,^f Antonio Monge,^f María Antonieta Rojas^g and Gloria Yaluff^f



5) SIGNIFICADO FÍSICO-QUÍMICO ORGÁNICO

Ej.

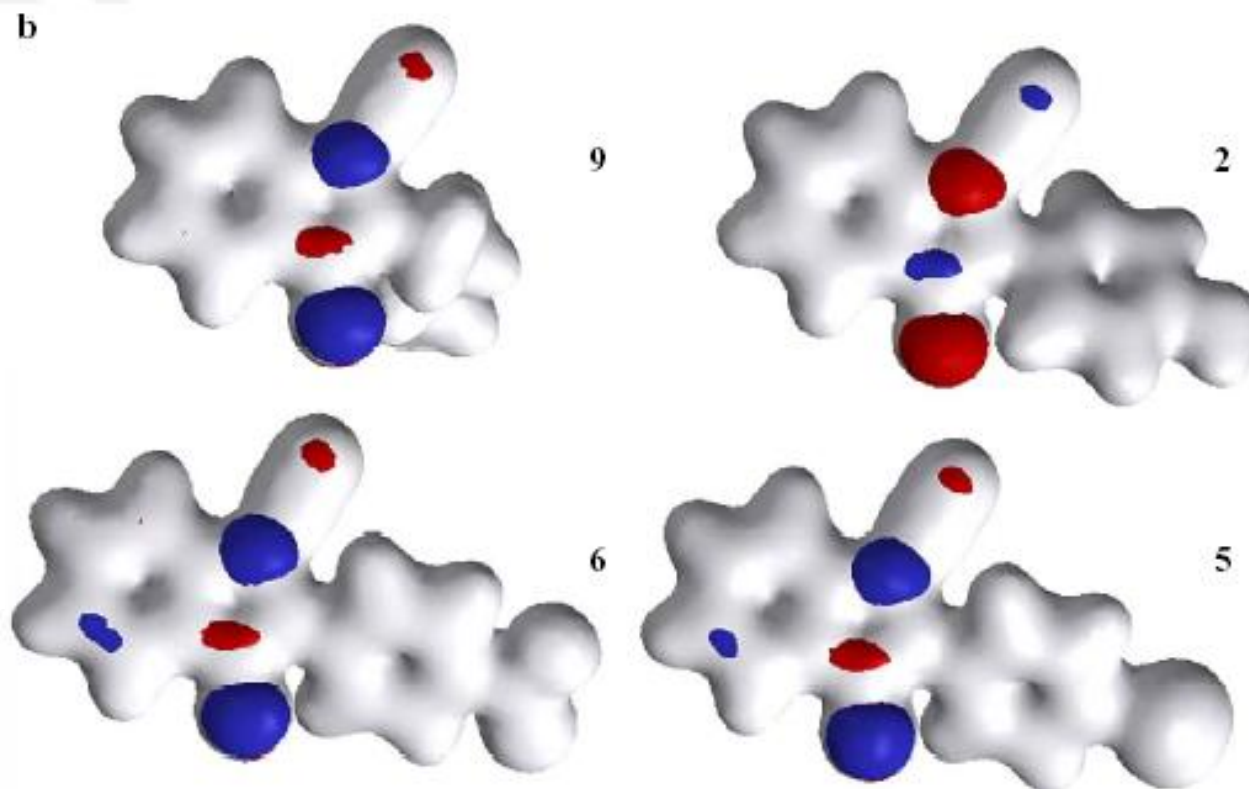


Table 7. Summary of QSAR results

Equation	Statistic parameters						Expressions
	r^2_{adj}	r^2	s	p	F value	n	
1	0.780	0.820	0.294	0.0004	20.49	12	$\log_{10}(GI_{BRENER}) = -3.63(\pm 0.84) + 0.33(\pm 0.06) \mu + 0.73(\pm 0.14) \text{ clog} P$
2	0.734	0.782	0.323	0.0010	16.18	12	$\log_{10}(GI_{BRENER}) = -19(\pm 4) - 3.04(\pm 0.67) E_{HOMO} + 0.71(\pm 0.15) \text{ clog} P$
3	0.573	0.679	0.356	0.0130	6.36	13	$\log_{10}(GI_{TULAHUEN}) = -19(\pm 10) + 0.14(\pm 0.08) \mu + 0.15(\pm 0.08) \text{ vol} - 0.0003(\pm 0.0001) \text{ vol}^2$
4	0.533	0.650	0.372	0.0190	5.56	13	$\log_{10}(GI_{TULAHUEN}) = -27(\pm 10) - 1.09(\pm 0.76) E_{HOMO} + 0.17(\pm 0.08) \text{ vol} - 0.0003(\pm 0.0001) \text{ vol}^2$

MAL EJEMPLO DE LA BIBLIOGRAFÍA



$$\log \% \text{ITW} = 1.4111(\log P)^2 - 0.5971 \log P - 0.1714\pi_{\text{Ali}}^2 - \\ - 3.2293\sigma_{\text{Ali}} + 0.9595E_{\text{sAli}} - 6.6199\sigma_{\text{I}} + 1.3249$$

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

Jha, T., Debnath, A. K., Mazumdar, A., Sengupta, C., and De, A. U., *Ind. J. Chem.* **25 B**, 169–174 (1986)

- Seis (6) variables para ocho (8) compuestos!!
- La concavidad de la cuadrática de logP es la incorrecta!!
- El coeficiente de σ es extremadamente grande dado que logITW varía sólo 0.24 unidades en todo el rango
- No se indican las desviaciones estándar o los intervalos de confianza de los coeficientes de la regresión

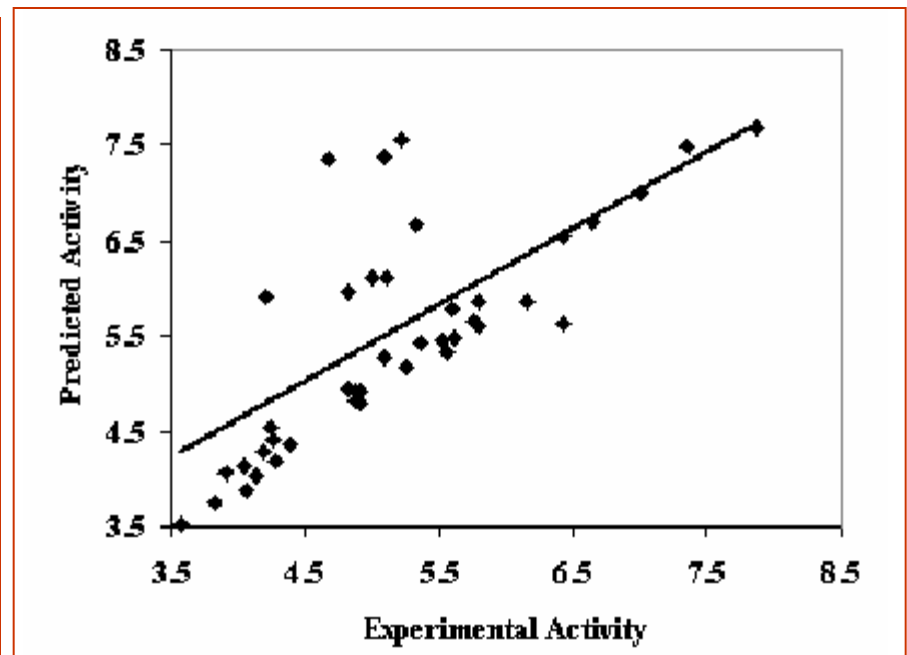
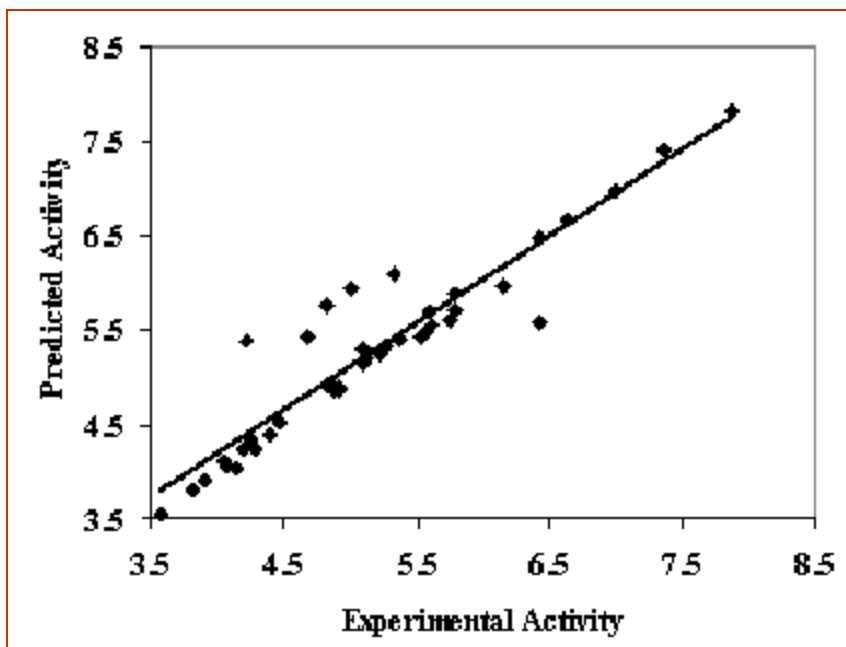
ADEMÁS DE ADECUADOS PARÁMETROS ESTADÍSTICOS

Ec. 1

$$\text{Log (1/C)} = 2.342 (\pm 0.105) \log P - 0.731 (\pm 0.041) \log P^2 + 0.036$$
$$(\pm 0.019) \sigma + 0.195 (\pm 0.176) E_s + 5.816$$

$$n = 17, r = 0.957, s = 0.126, F = 10.9$$

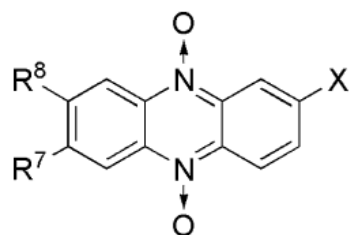
..... CON LA ECUACIÓN OBTENIDA SIEMPRE SE DEBE GRAFICAR ACTIVIDADES PREDICHAS vs ACTIVIDADES OBSERVADAS



OTROS ASPECTOS A TENER EN CUENTA

- **Cuando las variables no cumplen la distribución normal, NO pueden aplicarse los test paramétricos**
- **En estos casos se deben usar test no paramétricos (por ej. método ALS)**
- **El método ALS es aplicable para respuestas biológicas no continuas:**
 - presencia o ausencia de respuesta**
 - categorización +++, ++, +, 0, -, --, etc**

**VEAMOS EL
SIGUIENTE
EJEMPLO:**



$$\begin{aligned} SF_{\text{fox}} = & -812(\pm 499) - 491(\pm 431)q_{N_5/N_{12}} - 12(\pm 4)\mu \\ & + 252(\pm 137)c\text{LogP} - 19(\pm 10)(c\text{LogP})^2 \end{aligned} \quad (1)$$

$$n = 30, r^2 = 0.4336, r_{\text{adj}}^2 = 0.3430, F = 4.8, p = 0.0053$$

$$\begin{aligned} SF_{\text{hyp}} = & -393(\pm 128) + 63(\pm 9)IV_{\text{OH}} + 234(\pm 103)c\text{LogP} \\ & - 939(\pm 306)q_{N_5/N_{12}} - 1.4(\pm 0.6)(c\text{LogP})^2 \end{aligned} \quad (2)$$

$$n = 30, r^2 = 0.7012, r_{\text{adj}}^2 = 0.6534, F = 14.7, p < 0.00001$$

$$\begin{aligned} SF_{\text{hyp}} = & -322(\pm 114) + 68(\pm 8)IV_{\text{OH}} + 175(\pm 92)c\text{LogP} \\ & - 870(\pm 268)q_{N_5/N_{12}} - 1.1(\pm 0.6)(c\text{LogP})^2 \end{aligned} \quad (3)$$

$$\text{Outlier : 9} \quad n = 29, r^2 = 0.7740, r_{\text{adj}}^2 = 0.7363, F = 20.5, \\ p < 0.000001$$

$$\begin{aligned} SF_{\text{hyp}} = & -289(\pm 121) + 68(\pm 9)IV_{\text{OH}} + 146(\pm 98)c\text{LogP} \\ & - 836(\pm 276)q_{N_5/N_{12}} - 0.9(\pm 0.6)(c\text{LogP})^2 \end{aligned} \quad (4)$$

$$\text{Outlier : 14} \quad n = 29, r = 0.7628, r_{\text{adj}}^2 = 0.7232, F = 19.3, \\ p < 0.000001$$

$$\begin{aligned} SF_{\text{hyp}} = & -179(\pm 91) + 75(\pm 6)IV_{\text{OH}} + 54(\pm 74)c\text{LogP} \\ & - 727(\pm 203)q_{N_5/N_{12}} - 0.35(\pm 0.45)(c\text{LogP})^2 \end{aligned} \quad (5)$$

$$\text{Outliers : 9 and 14} \quad n = 28, r = 0.8745, r_{\text{adj}}^2 = 0.8527, \\ F = 40.1, p < 1 \times 10^{-9}$$

“OUTLIERS”:

Son términos del Q.S.A.R. (compuestos, sustituyentes) que al quitarlos de una correlación, la misma se robustece estadísticamente

Son datos que NO PUEDEN ser explicados por el modelo

¿Es bueno que existan outliers?

Muchas veces nos dan tanta o más información que los compuestos o sustituyentes que se ajustan adecuadamente al modelo desarrollado.

Tipos de outliers:

- 1. Aquellos que no pertenecen a la familia en estudio.**
- 2. Aquellos que poseen un mecanismo de acción diferente.^(*) Un caso interesante son aquellos que pueden presentar una química o bioquímica diferencial.**
- 3. Aquellos cuyo descriptor está sobre- o sub-evaluado.**
- 4. Falta de fiabilidad en los resultados biológicos.**

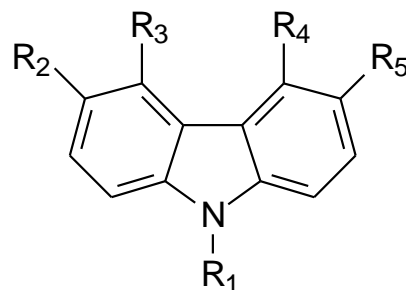
^(*)H. Gao, W.A. Denny, R. Garg, C. Hansch. *Chem-Biol. Inter.*, 1998, 116, 157.

3.4. MODELOS Q.S.A.R. EN USO

3.4.2. Aproximación *de novo*

- HAY AUTORES QUE DICEN QUE ES EL ÚNICO MODELO AUTÉNTICO DE RELACIÓN ESTRUCTURA-ACTIVIDAD, YA QUE SE CALCULA LA CONTRIBUCIÓN DE CADA FRAGMENTO CONSTITUTIVO.
- PARA PODER APLICARLO SE DEBE POSEER UNA ADECUADA POBLACIÓN DE ESTUDIO.

- ÚTIL PARA CUANDO SE TIENEN POCOS SUSTITUYENTES DISTINTOS Y OCUPAN DIFERENTES POSICIONES EN LA MOLÉCULA.



R₁-R₅= Me, Et, Prop

- ASUME QUE LAS CONTRIBUCIONES SON CONSTANTES Y ADITIVAS E INDEPENDIENTES DE LOS OTROS POSIBLES CAMBIOS QUE PUEDAN HACERSE EN LA ESTRUCTURA DE LA MOLÉCULA

$$\text{Log } A = \sum b_{jk} X_{jk} + \mu$$

- S. Free, J. Wilson. *J. Med. Chem.*, 1964, 7, 395.
- T. Fujita, T. Ban. *J. Med. Chem.*, 1971, 14, 148.

3.4. MODELOS Q.S.A.R. EN USO

3.4.2. Aproximación *de novo*

$$\text{Log A} = \sum b_{jk} X_{jk} + \mu$$

- μ es la actividad determinada para un compuesto de referencia
- X_{jk} es una variable indicadora que toma el valor 1 cuando el sustituyente está presente en una posición dada y el valor 0 en caso contrario
- b_{jk} es la contribución a la actividad del sustituyente X_{jk} en una determinada posición (es el término que se calcula en la regresión múltiple)

Según Martín(*) el número mínimo de análogos a incluir en el estudio estará dado por:

$$N_{\min} = \sum_j (n_i - 1) + 1$$

Siendo j el número de posiciones e i el número de grupos a variar

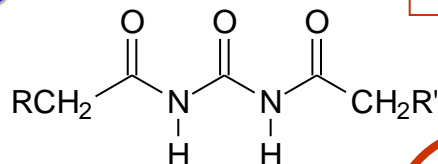
- S. Free, J. Wilson. *J. Med. Chem.*, 1964, 7, 395. (*) Y.C. Martin. *Quantitative Drug Design. A Critical Introduction*. Dekker, New York, 1978.
- T. Fujita, T. Ban. *J. Med. Chem.*, 1971, 14, 148.

3.4. MODELOS Q.S.A.R. EN USO

3.4.2. Aproximación *de novo*

Se obtiene multiplicando el número de carbonos de cada resto acilo

Ej. Diacilureas con actividad hipnótica:

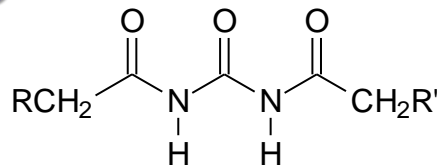


Comp\R	H	Me	Et	Pro	But	Pent	N ₁ ·N ₂	Log(1/C)
<i>X_{jk}</i>	<i>X₁</i>	<i>X₂</i>	<i>X₃</i>	<i>X₄</i>	<i>X₅</i>	<i>X₆</i>		
1	1	1	0	0	0	0	6	1.84
2	0	2	0	0	0	0	9	2.06
3	1	0	1	0	0	0	8	2.16
4	0	1	1	0	0	0	12	2.23
5	1	0	0	1	0	0	10	2.27
6	0	0	2	0	0	0	16	2.40
7	0	1	0	1	0	0	15	2.35
8	1	0	0	0	1	0	12	2.46
9	0	0	1	1	0	0	20	2.38
10	0	1	0	0	1	0	18	2.25
11	1	0	0	0	0	1	14	2.55
12	0	0	0	2	0	0	25	2.32
13	0	0	1	0	1	0	24	2.28

3.4. MODELOS Q.S.A.R. EN USO

3.4.2. Aproximación *de novo*

Ej. Diacilureas con actividad hipnótica:



$$\text{Log}(1/C) = 1.930 (\pm 0.13) + 0.430 (\pm 0.17) X_2 + 0.910 (\pm 0.28) X_3 + 1.249 (\pm 0.39) X_4 + 1.487 (\pm 0.45) X_5 + 1.813 (\pm 0.49) X_6^* - 0.085 (\pm 0.03) N_1 \cdot N_2$$

R-Square(COD)

Root-MSE(SD)

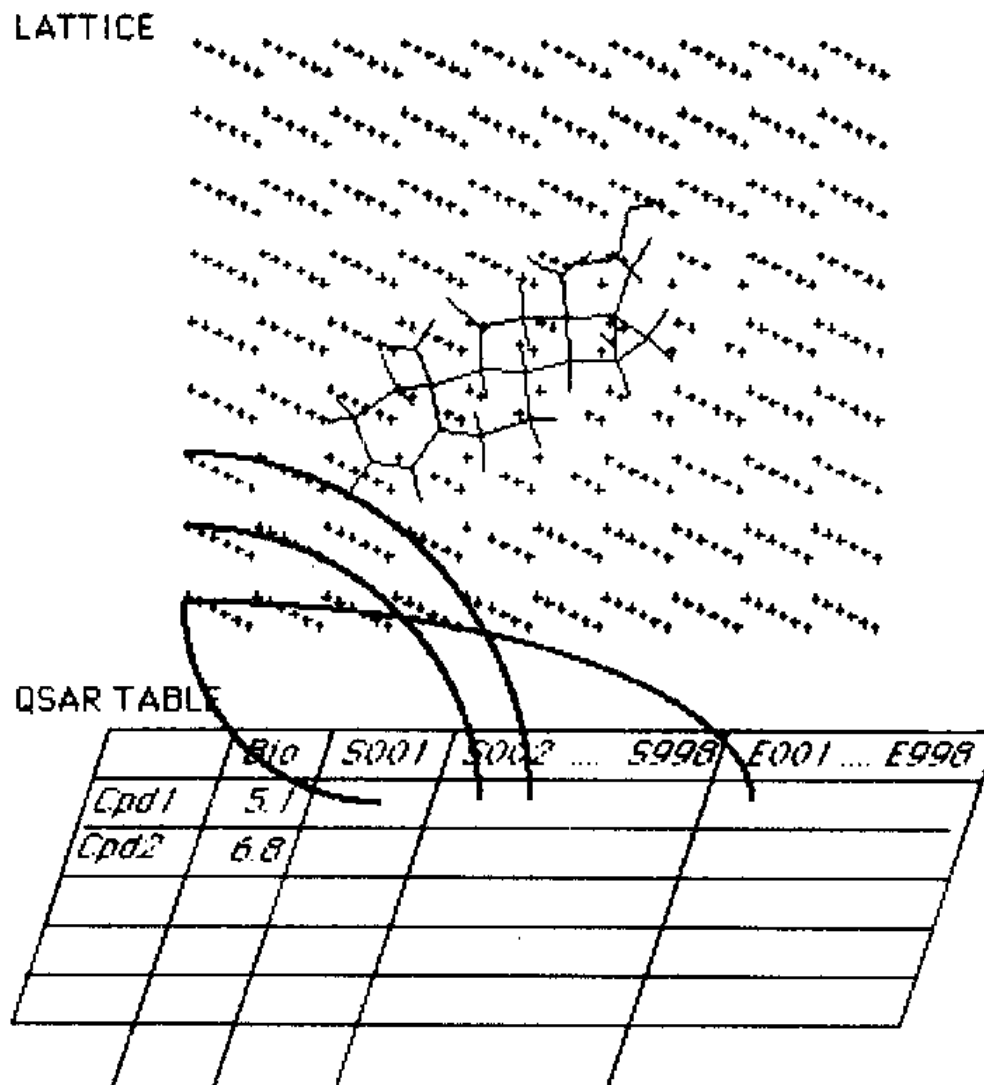
0.96375

0.04888

3.4. MODELOS Q.S.A.R. EN USO

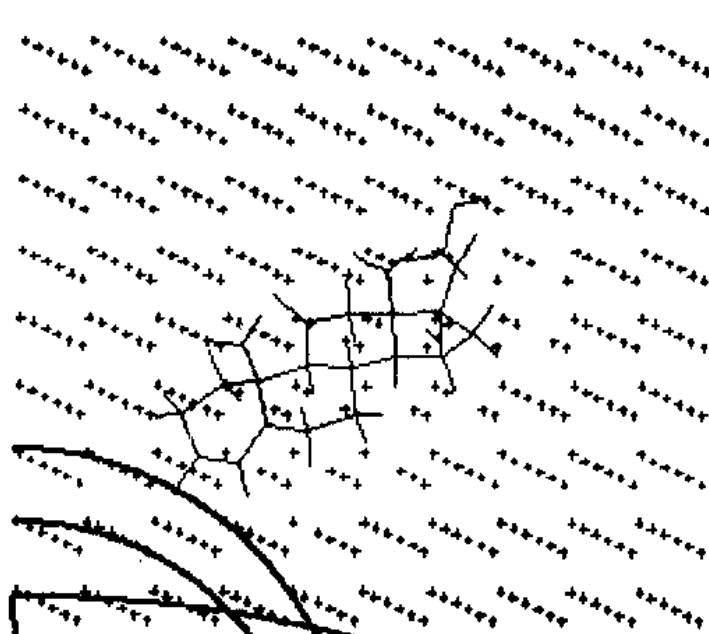
3.4.4. QSAR tridimensional

Método CoMFA (Comparative Molecular Field Analysis) implementado en SYBYL



Método CoMFA

LATTICE



QSAR TABLE

	Bio	S001	S002	...	S998	E001	...	E998
Cpd1	5.1							
Cpd2	6.8							

PLS

EQUATION

$$Bio = y + a \times S001 + b \times S002 + \dots + m \times S998 + n \times E001 + \dots + z \times E998$$

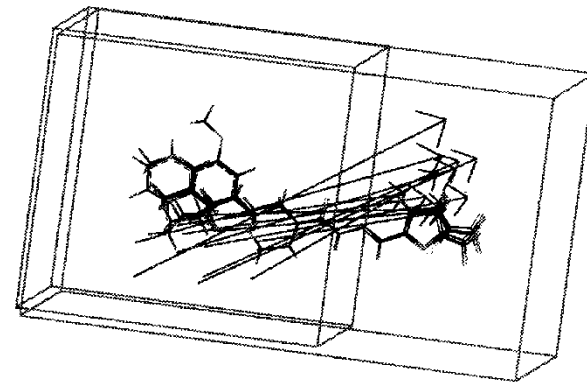
Método CoMFA

- **DEFINICIÓN DEL ENREJILLADO (GRID):** Para los dos tipos de campos se utiliza, por defecto, un átomo de prueba: un C_{sp3} con carga +1. Generando un espacio tridimensional con “descriptores” de interacciones estéricas y electrostáticas.

El átomo de prueba se espacia regularmente, dentro de la caja, cada 2 Å por el procedimiento automático del módulo SYBYL-CoMFA.

- Las nuevas versiones de SYBYL-CoMFA traen otros átomos de prueba, por ej. – OH, capaces de establecer interacciones tipo enlace de hidrógeno.

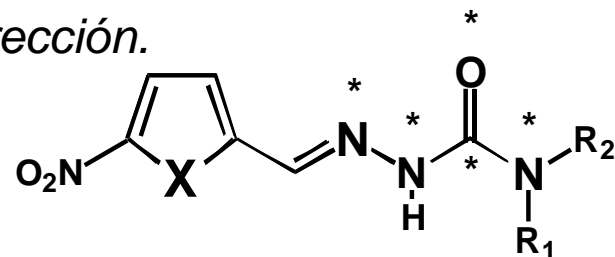
- **CAJAS:** Por defecto el modulo SYBYL-CoMFA genera una caja que involucra a todas las moléculas y se extiende por lo menos 4 Å a lo largo de los tres ejes.



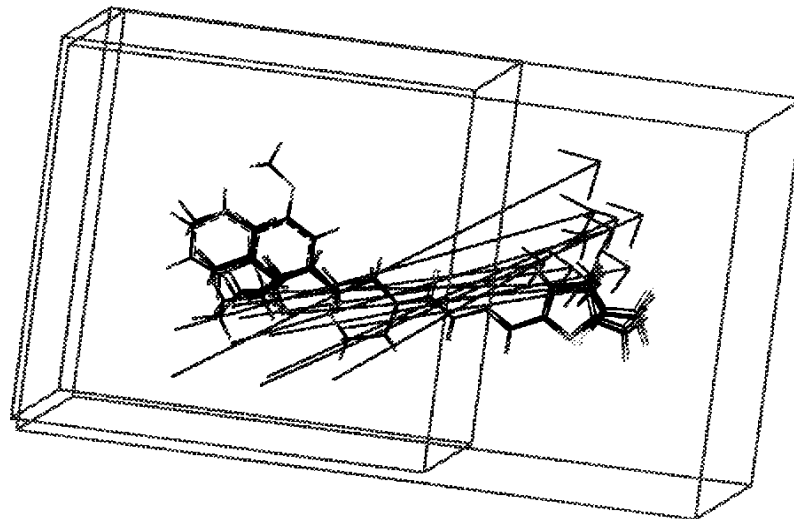
Ej. Nitrofuranos y nitrotiofenos con actividad antichagásica *in vitro* e *in vivo*

- **CONSTRUCCIÓN DE LOS COMPUESTOS Y CRITERIOS DE ALINEACIÓN:** Se construyeron usando el módulo *standard*, distancias y ángulos de enlace, de molecular modelling de SYBYL. Para el sistema $-N=N=C$ se introdujeron datos cristalográficos de otro 5-nitrofurfuriliden derivado (Caracelli I, Stamato FM, Mester B, Paulino M, Cerecetto H (1996) *Acta Crystallogr C* 52: 1281-1282). La energía de cada compuesto fue minimizada por el método de mecánica molecular implementado en SYBYL, MAXIMIN2, usando los campos de fuerzas de mecánica molecular de TRIPOS y como criterio de convergencia un gradiente de $0.01 \text{ kcal}/(\text{mol}\cdot\text{Å})$. Las cargas atómicas parciales se calcularon por el método semiempírico MNDO implementado dentro del paquete MOPAC 6.0 de SYBYL. La carga neta de los compuestos 1 – 7 y 11 – 16 se asignó como cero y para los derivados 8 – 10, 17, y 18 como +1.

- Se ensayaron varias alineaciones y finalmente la alineación que se hizo por los átomos (*) dio los mejores resultados estadísticos. En este alineamiento todas las moléculas muestran un momento dipolar en la misma dirección.

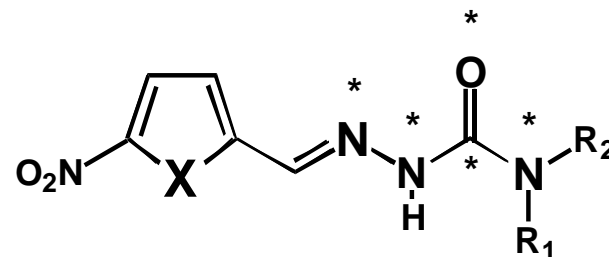


Ej. Nitrofuranos y nitrotiofenos con actividad antichagásica *in vitro* e *in vivo*



- **CAJAS:** En este ejemplo, además de la caja por defecto del modulo SYBYL-CoMFA se generó una caja que dejase fuera el sistema heterocíclico (ya que era la menor de las modificaciones dentro de la familia).

- **MÉTODO ESTADÍSTICO:** SIMCA. Este método estadístico se utiliza cuando las propiedades son variables categóricas más que variables continuas.

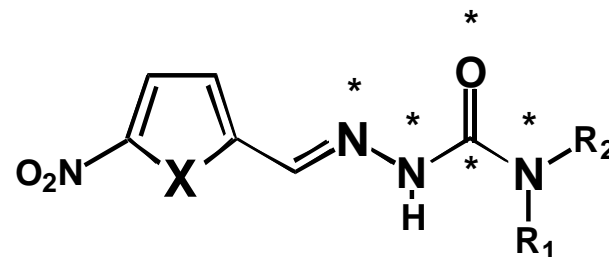


Ej. Nitrofuranos y nitrotiofenos con actividad antichagásica *in vitro* e *in vivo*

- **RESULTADO ESTADÍSTICO:** Para que el CoMFA tenga capacidad predictiva se eligen los modelos generados cuyo valor de q^2 , coeficiente de correlación de cross-validación (leave-one-out), sean mayores que 0.5. Por supuesto que existen muchos otros criterios, este es el inicial para elegir un modelo.

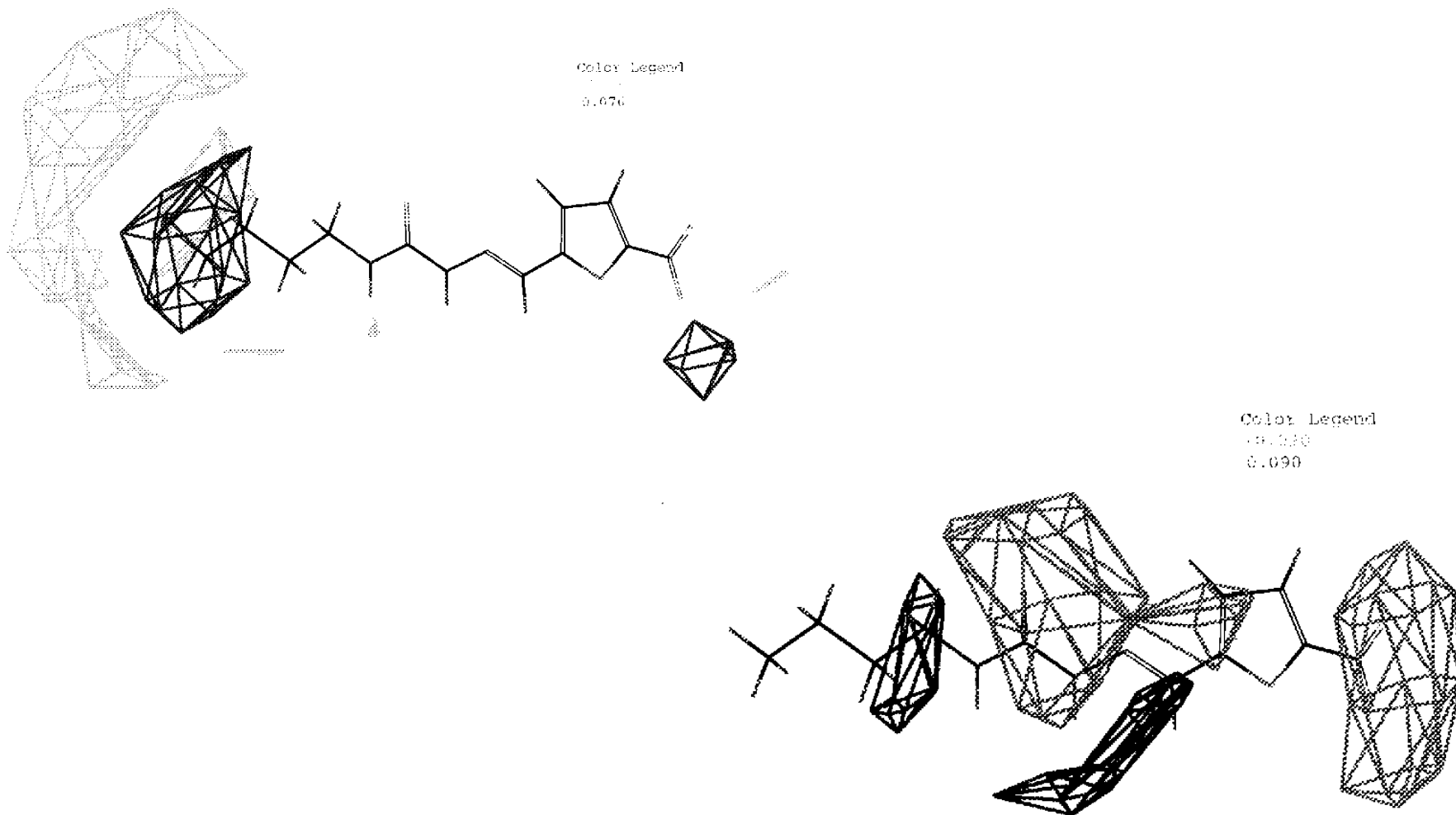
- **GRUPO DE PRUEBA:** Se deja una serie de compuestos, grupo de prueba, cuyas actividades serán predichas con los modelos generados. Estas predicciones servirán para elegir entre más de un modelo.

- **GENERACIÓN DE VISUALIZACIONES:** Con el(los) mejor(es) modelo(s) se genera un mapa tridimensional de exigencias estéricas y electrónicas de acuerdo al modelo QSAR generado.



Ej. Nitrofuranos y nitrotiofenos con actividad antichagásica *in vitro* e *in vivo*

- GENERACIÓN DE VISUALIZACIONES:



Ej. Nitrofuranos y nitrotiofenos con actividad antichagásica *in vitro* e *in vivo*

- GENERACIÓN DE VISUALIZACIONES:

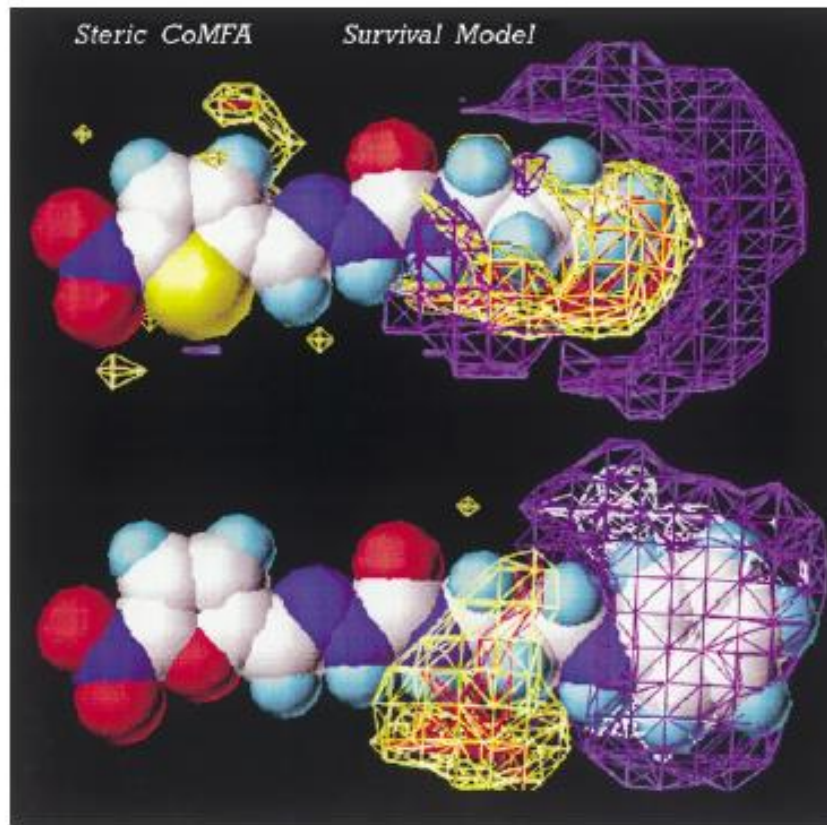


Figure 1. Contour maps of the a_{SURV} steric CoMFA field loadings for the first component of the active (upper) or inactive (lower) class in SIMCA model 1. Contour levels are -0.06 white, -0.03 purple, 0.03 yellow and 0.06 red. Reference molecules 11 (upper) and 7 (lower) are displayed in space-filling format with the atoms represented by the usual colours.



Figure 2. CoMFA contour maps of the a_{SURV} steric CoMFA field loadings for the first component of the active (upper) or inactive (lower) class in SIMCA model 1. Reference molecules 11 (upper) and 7 (lower) are displayed in space-filling format with the atoms represented by the usual colours.

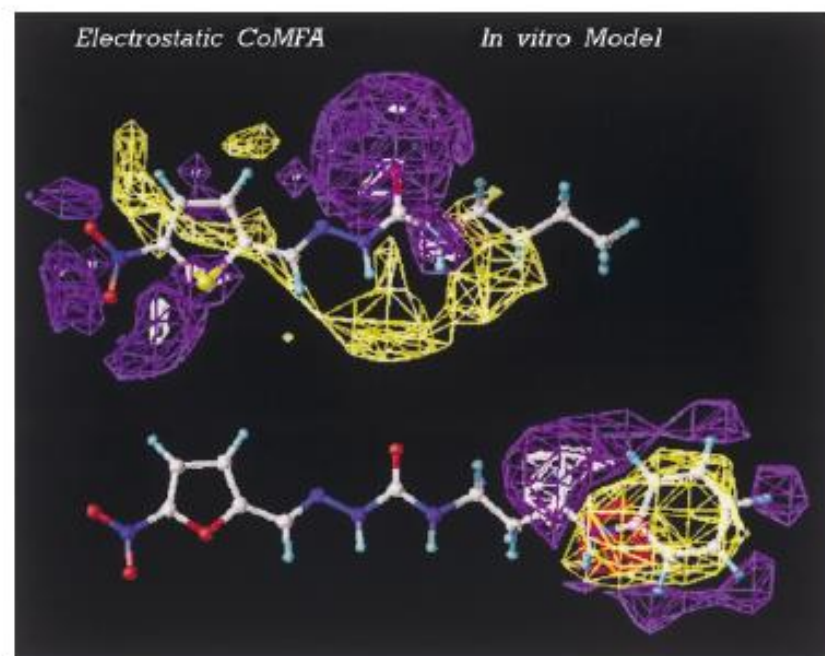


Figure 3. Contour maps of the a_{INVIT} electrostatic CoMFA field loadings for the first component of the active (upper) or inactive (lower) class in SIMCA model 2. Contour levels are as in Figure 1. Reference molecules 11 (upper) and 7 (lower) are displayed in ball and stick format with the atoms represented by the usual colours.

Ej. Nitrofuranos y nitrotiofenos con actividad antichagásica *in vitro* e *in vivo*

- Nueva generación de nitrofuranos activos

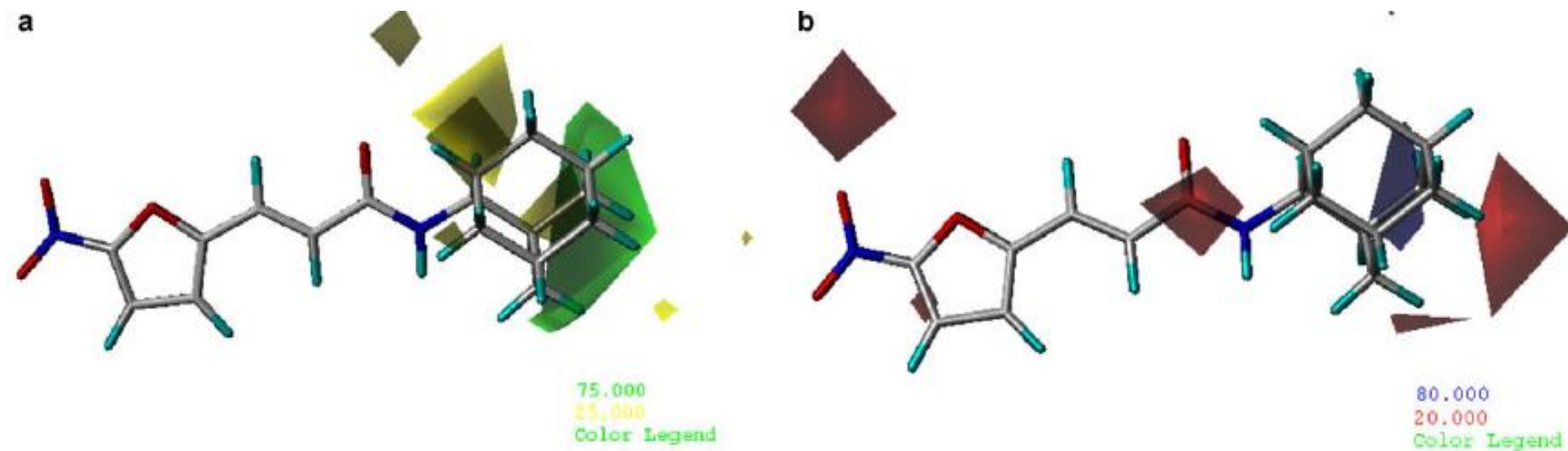
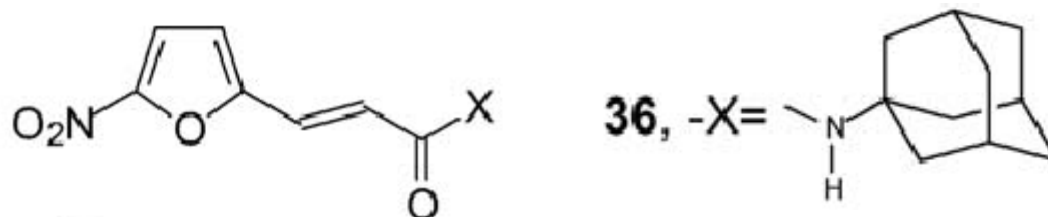


Fig. 3. **a)** Views of the CoMFA steric field contour maps for the model VI, the steric field was contoured at 0.075 and -0.075 levels. Compound 36 is superimposed in the map. **b)** Views of the CoMFA electrostatic field contour maps for the model VI, the electrostatic field was contoured at 0.075 and -0.075 levels. Compound 36 is superimposed in the maps.



OTROS TIPOS DE ANÁLISIS MOLECULARES COMPARATIVOS:

Método CoMSIA (Comparative Molecular Similarity Indices Analysis) implementado en SYBYL

Table 4. Statistical results for the best CoMSIA models obtained for phthalimide-containing TNF α modulators.

	Steric/Electrostatic	Steric	Electrostatic	Donor/Acceptor	Donor	Acceptor	Hydrophobic
q^{2a}	0.868	0.783	0.808	0.665	0.629	0.545	0.826
N^b	6	6	6	6	6	6	6
S_{cv}^c	0.417	0.533	0.519	0.663	0.687	0.934	0.487
r^{2d}	0.983	0.937	0.970	0.841	0.800	0.822	0.976
SEE^e	0.157	0.304	0.210	0.485	0.544	0.513	0.187
F^f	234.426	59.596	130.013	21.105	16.634	18.424	165.431

^aCross-validation correlation coefficient. ^bNumber of components. ^cStandard error of prediction. ^dCorrelation coefficient. ^eStandard error of estimative. ^fF-ratio.

Table 4. Statistical results for the best CoMSIA models obtained for phthalimide-containing TNF α modulators.

	Steric/Electrostatic	Steric	Electrostatic	Donor/Acceptor	Donor	Acceptor	Hydrophobic
q^{2a}	0.868	0.783	0.808	0.665	0.629	0.545	0.826
N^b	6	6	6	6	6	6	6
S_{cv}^c	0.417	0.533	0.519	0.663	0.687	0.934	0.487
r^{2d}	0.983	0.937	0.970	0.841	0.800	0.822	0.976
SEE^e	0.157	0.304	0.210	0.485	0.544	0.513	0.187
F^f	224.426	50.506	120.012	21.105	16.624	18.424	165.421

^aCross-validated coefficient of determination

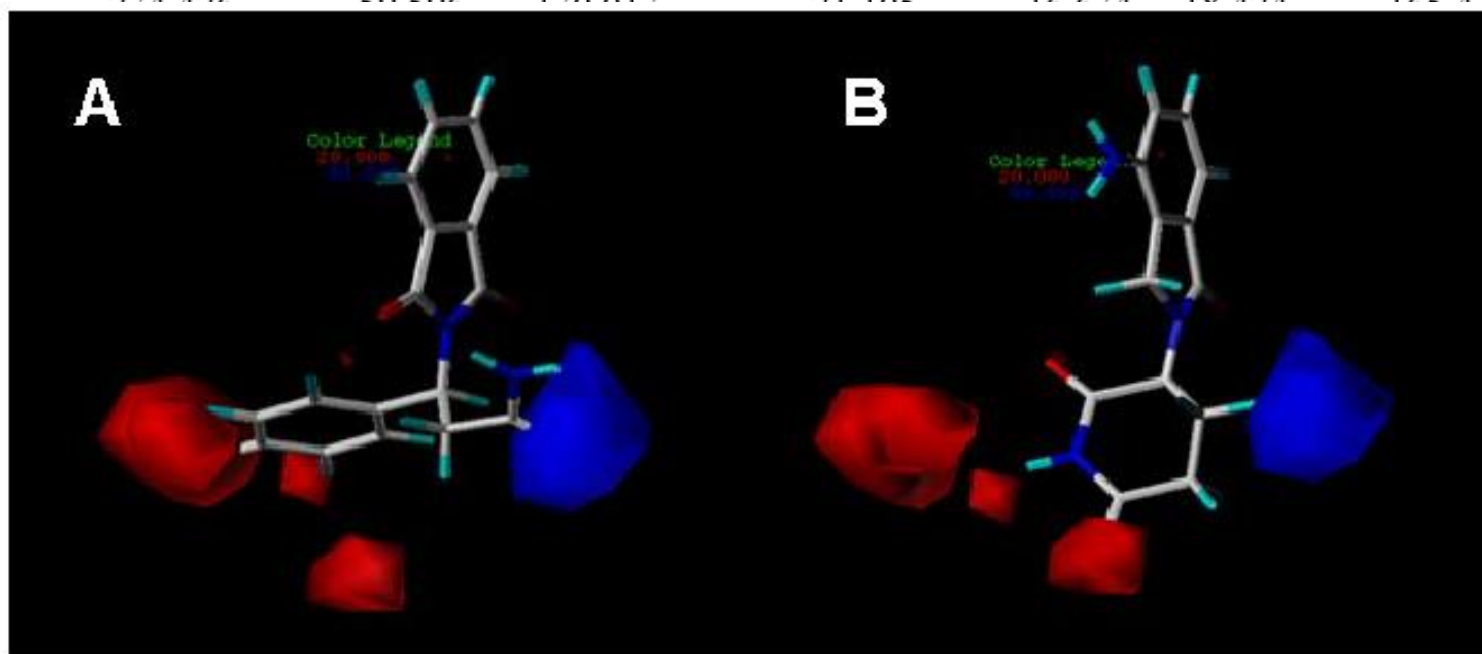


Figure 6. Electrostatic contour map from the best CoMSIA model. Compound **3** (A) and lenalidomide (**2**) (B), the least and the most active derivatives, respectively, are shown inside fields. Blue (80% contribution) and red (20% contribution) contours encompass areas where an increase in positive or negative charge, respectively, are favorable for binding properties.

Table 4. Statistical results for the best CoMSIA models obtained for phthalimide-containing TNF α modulators.

	Steric/Electrostatic	Steric	Electrostatic	Donor/Acceptor	Donor	Acceptor	Hydrophobic
q^{2a}	0.868	0.783	0.808	0.665	0.629	0.545	0.826
N^b	6	6	6	6	6	6	6
S_{cv}^c	0.417	0.533	0.519	0.663	0.687	0.934	0.487
r^{2d}	0.983	0.937	0.970	0.841	0.800	0.822	0.976
SEE^e	-----	-----	-----	-----	-----	-----	-----
F^f	-----	-----	-----	-----	-----	-----	-----

^aCross-validated coefficient of determination

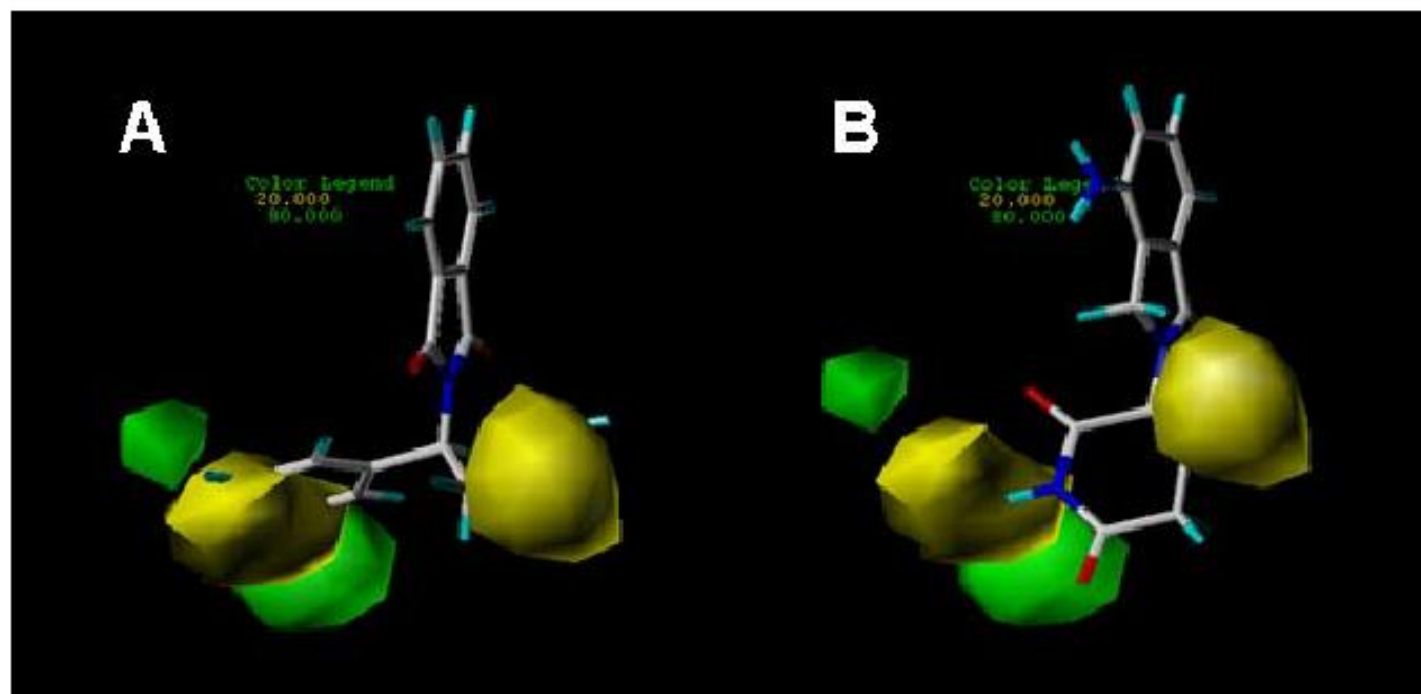


Figure 7. Contour plots of steric fields from the best CoMSIA model. Compound **3** (A) and lenalidomide (**2**) (B), the least and one of the most active derivatives, respectively, are shown inside fields. Regions where increasing the molecular volume increases bioactivity are in green (80% contribution) and regions where increasing the molecular volume decreases the activity are in yellow (20% contribution).

Table 4. Statistical results for the best CoMSIA models obtained for phthalimide-containing TNF α modulators.

	Steric/Electrostatic	Steric	Electrostatic	Donor/Acceptor	Donor	Acceptor	Hydrophobic
q^{2a}	0.868	0.783	0.808	0.665	0.629	0.545	0.826
N^b	6	6	6	6	6	6	6
S_{cv}^c	0.417	0.533	0.519	0.663	0.687	0.934	0.487
r^{2d}	0.983	0.937	0.970	0.841	0.800	0.822	0.976

SEE

F^f

^aCross
coeffic

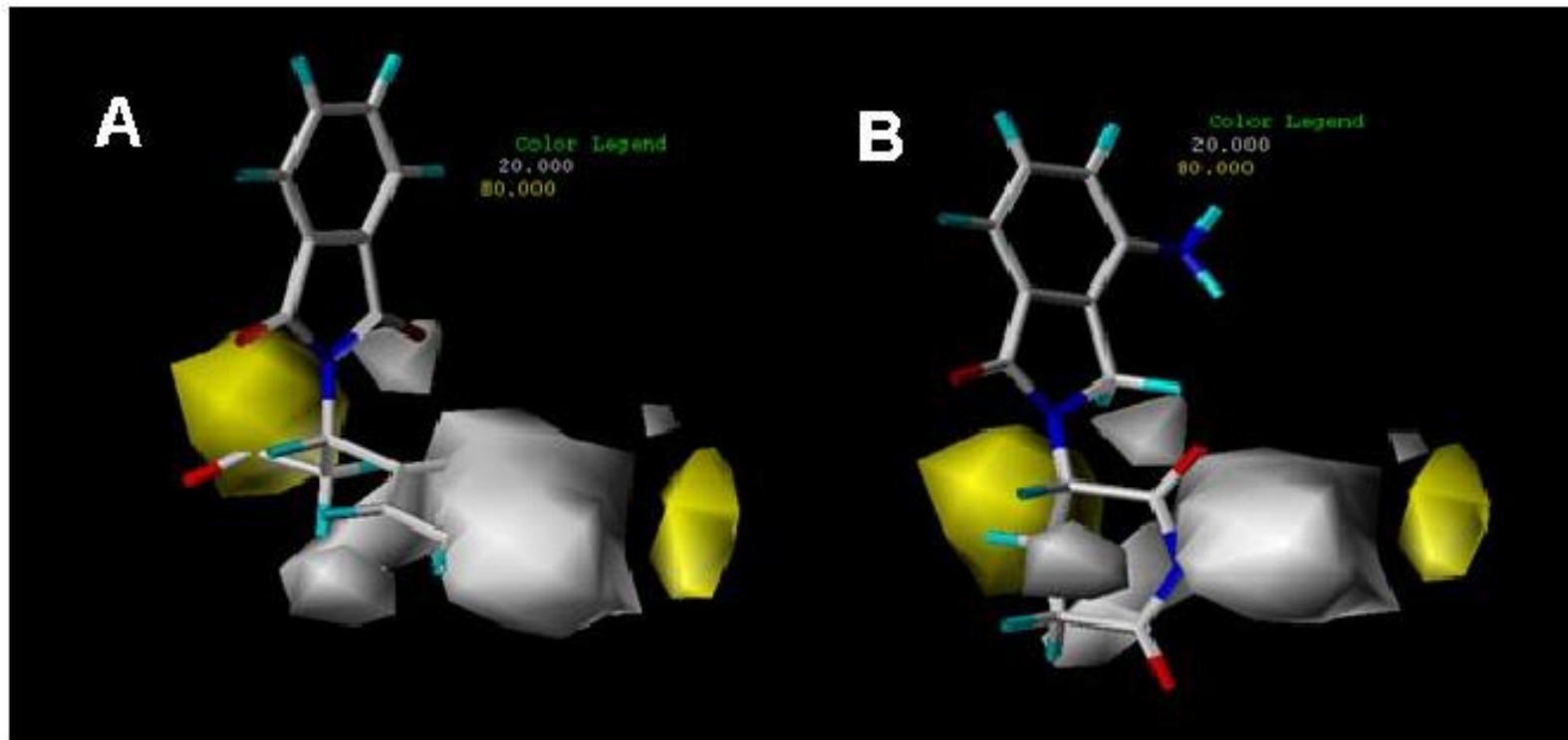
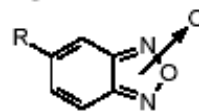


Figure 8. CoMSIA hydrophobic contour maps. Compound **3** (A) and lenalidomide (**2**) (B) are shown inside fields. Yellow regions (80% contribution) indicate areas where hydrophobic groups increase activity and white regions (20% contribution) indicate areas where hydrophobic groups decrease activity.

Table 3. Benzofuroxan derivatives and substituent descriptors used to derived **equation 1** of the 2D-QSAR model

Ref.	-R	Log ₁₀ (%Inh) ^a	I ₆ ^b	MR ^c	\mathcal{F}	I _{HBD}
1	-H	1.54	0	0.10	0.00	0
2	-N(CH ₃) ₂	1.41	0	1.56	0.15	0
3	- <i>E</i> -CH=CH-Ph	1.99	0.5	3.42	0.10	0
5	-CO ₂ H	0.70	0.5	0.69	0.34	1
8	-NHCO ₂ C(CH ₃) ₃	1.26	0	3.05 ^f	0.13	1
10	-CHO	1.83	1	0.69	0.33	0
16	- <i>E</i> -CH=CH-NO ₂	1.96	1	1.64	0.35	0
17	-NO ₂	1.95	1	0.74	0.65	0
18	-CH=NOH	1.58	1	1.03	0.28	1
19	-CH ₂ I	1.97	0.5	1.86	0.12	0
20	-CH ₂ Cl	1.99	0.5	1.05	0.13	0
21	-CH ₂ OH	1.32	0.5	0.72	0.03	1
22	-Ph	1.49	0	2.54	0.12	0
23	-CH ₃	1.34	0	0.56	0.01	0
24	-OCH ₃	1.00	0	0.79	0.29	0
25	-CH=C(CN) ₂	1.82	1	1.97	0.57	0
26	-CN	0.95	1	0.63	0.51	0
27	-CH=NNHC(S)NH ₂	1.23	0.5	2.96	0.46	1
28	-SO ₂ NH ₂	0.85	0.5	1.23	0.49	1
29	-Cl	0.70	0	0.60	0.42	0
30	-Br	0.70	0	0.89	0.45	0

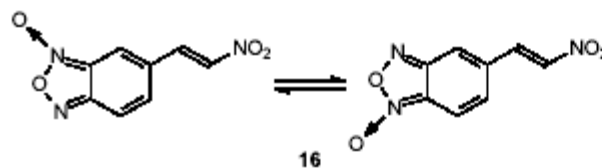
^a % Inh: percentage of *T. cruzi* growth inhibition at 25 μM. ^b I₆: indicator variable that takes value 1 when the substituent is preferentially 6-positioned respect to *N*-oxide moiety, value 0 when is preferentially 5-positioned and value 0.5 when in the equilibrium exist equal quantity of 5- and 6-positioned tautomers at low temperature. ^c To auto scale these independent variables the MR reported are MR_{lit} / 10, being MR_{lit} the values taken from lit. [16b].

Ej. QSAR-2D vs QSAR-3D

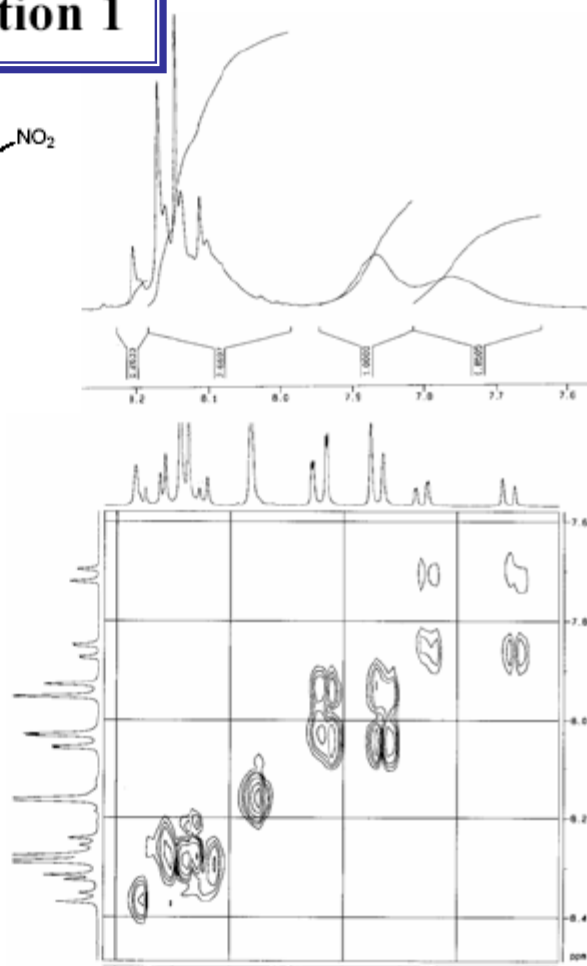
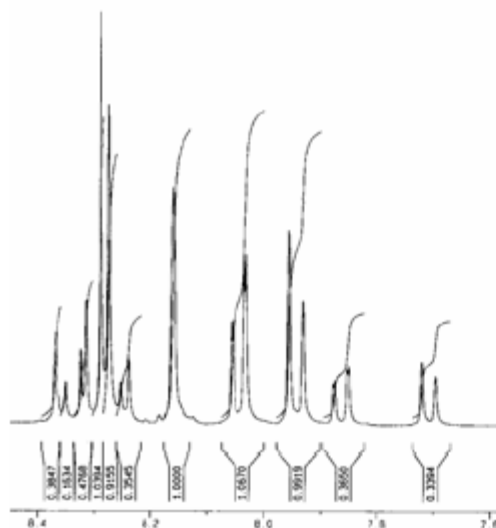
$$\text{Log}_{10}(\% \text{ Inh}) = 1.31(\pm 0.10) + 1.0(\pm 0.1) I_6 + 0.15(\pm 0.04) \text{MR} - 1.31(\pm 0.24) F - 0.51(\pm 0.09) I_{\text{HBD}}$$

$n = 20, r = 0.939, r^2_{\text{adj}} = 0.849, s = 0.184, p < 0.0001$

equation 1



c)



Ej. QSAR-2D vs QSAR-3D

$$\text{Log}_{10}(\% \text{ Inh}) = 1.31(\pm 0.10) + 1.0(\pm 0.1) I_6 + 0.15(\pm 0.04) \text{MR} - 1.31(\pm 0.24) F - 0.51(\pm 0.09) I_{\text{HBD}}$$

$n = 20, r = 0.939, r^2_{\text{adj}} = 0.849, s = 0.184, p < 0.0001$

equation 1

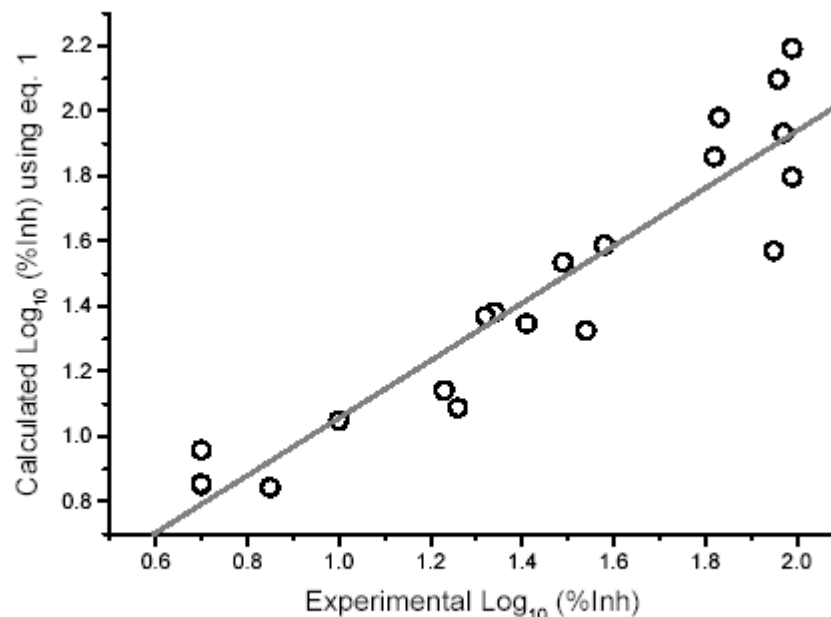


Figure 3. Actual versus calculated activity of 2D-QSAR model (eq. 1).

Ej. QSAR-2D vs QSAR-3D

$$\text{Log}_{10}(\% \text{ Inh}) = 1.31(\pm 0.10) + 1.0(\pm 0.1) I_6 + 0.15(\pm 0.04) \text{MR} - 1.31(\pm 0.24) F - 0.51(\pm 0.09) I_{\text{HBD}}$$

$n = 20, r = 0.939, r^2_{\text{adj}} = 0.849, s = 0.184, p < 0.0001$

equation 1

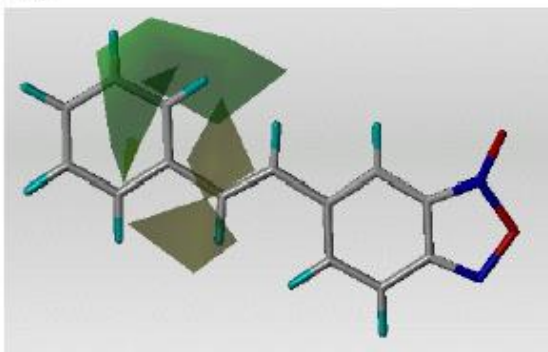
Table 4. Summary of CoMFA results.

CoMFA Model	PLS Analysis						Contribution				<i>n</i>
	<i>q</i> ²	Components	SEP	<i>r</i> ²	SEE	<i>F</i> value	Steric ^a	Electrostatic ^a	Other variables I ₆ cLogP		
I	0.559	6	0.392	0.974	0.096	69.44	---	65 ^b	26	9	18
II	0.620	6	0.329	0.989	0.055	169.34	26 ^c	57 ^c	17	---	18
III	0.667	6	0.310	0.991	0.055	215.79	25 ^d	63 ^d	12	---	19 ^e
IV	0.680	5	0.286	0.981	0.070	164.92	33 ^f	57 ^f	10	---	22 ^g
V	0.712	6	0.289	0.992	0.049	279.18	31 ^d	57 ^d	12	---	21 ^h
VI	0.753	6	0.271	0.994	0.040	343.21	32 ^d	58 ^d	10	---	20 ⁱ
VII	0.802	8	0.266	0.997	0.031	492.45	30 ^d	59 ^d	11	---	20 ^j

^a Contribution truncated to: ^b ±5 kcal/mol, ^c ±10 kcal/mol, ^d ±30 kcal/mol, ^e ±20 kcal/mol. ^e Compounds used in model II (see text) and derivative 4. ^g Compounds used in model III and derivatives 11-13. ^h Compounds used in model IV without derivative 8. ⁱ Compounds used in model V without derivative 10. ^j Compounds used in model VI without derivative 13 and with derivative 14.

Ej. QSAR-2D vs QSAR-3D

A)



B)

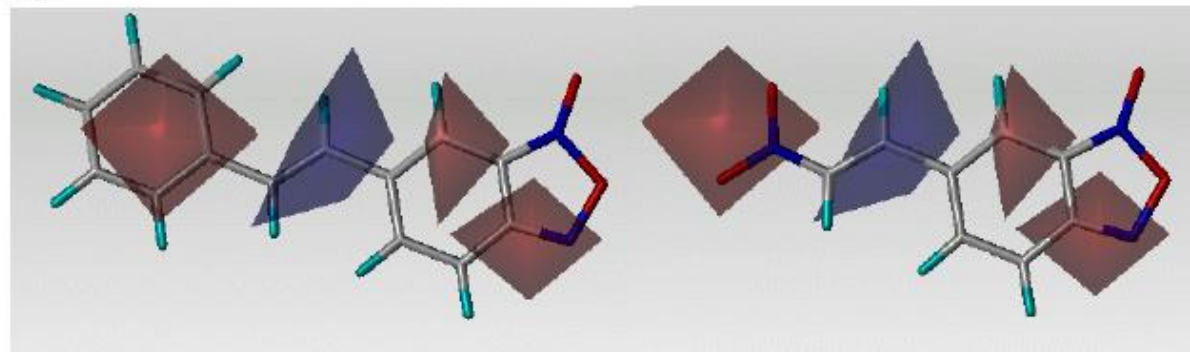


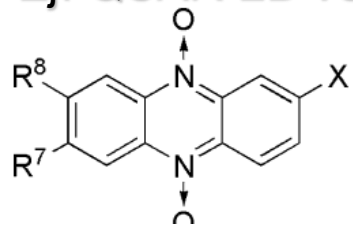
Figure 5. A) Views of the CoMFA steric field contour maps for model VII, the steric field was contoured at 0.075 and -0.075 levels. Compound **3** is superimposed in the map. B) Views of the CoMFA electrostatic field contour maps for model VII, the electrostatic field was contoured at 0.075 and -0.075 levels. Compounds **3** and **16** are superimposed in the maps.

Table 4. Summary of CoMFA results.

CoMFA Model	PLS Analysis						Contribution				
	q^2	Components	SEP	r^2	SEE	F value	Steric ^a	Electrostatic ^a	Other variables I_6	cLogP	n
I	0.559	6	0.392	0.974	0.096	69.44	---	65 ^b	26	9	18
II	0.620	6	0.329	0.989	0.055	169.34	26 ^c	57 ^c	17	---	18
III	0.667	6	0.310	0.991	0.055	215.79	25 ^d	63 ^d	12	---	19 ^e
IV	0.680	5	0.286	0.981	0.070	164.92	33 ^f	57 ^f	10	---	22 ^g
V	0.712	6	0.289	0.992	0.049	279.18	31 ^d	57 ^d	12	---	21 ^h
VI	0.753	6	0.271	0.994	0.040	343.21	32 ^d	58 ^d	10	---	20 ⁱ
VII	0.802	8	0.266	0.997	0.031	492.45	30 ^d	59 ^d	11	---	20 ^j

^a Contribution truncated to: ^b ± 5 kcal/mol, ^c ± 10 kcal/mol, ^d ± 30 kcal/mol, ^e ± 20 kcal/mol. ^e Compounds used in model II (see text) and derivative **4**. ^g Compounds used in model III and derivatives **11-13**. ^h Compounds used in model IV without derivative **8**. ⁱ Compounds used in model V without derivative **10**. ^j Compounds used in model VI without derivative **13** and with derivative **14**.

Ej. QSAR-2D vs QSAR-3D



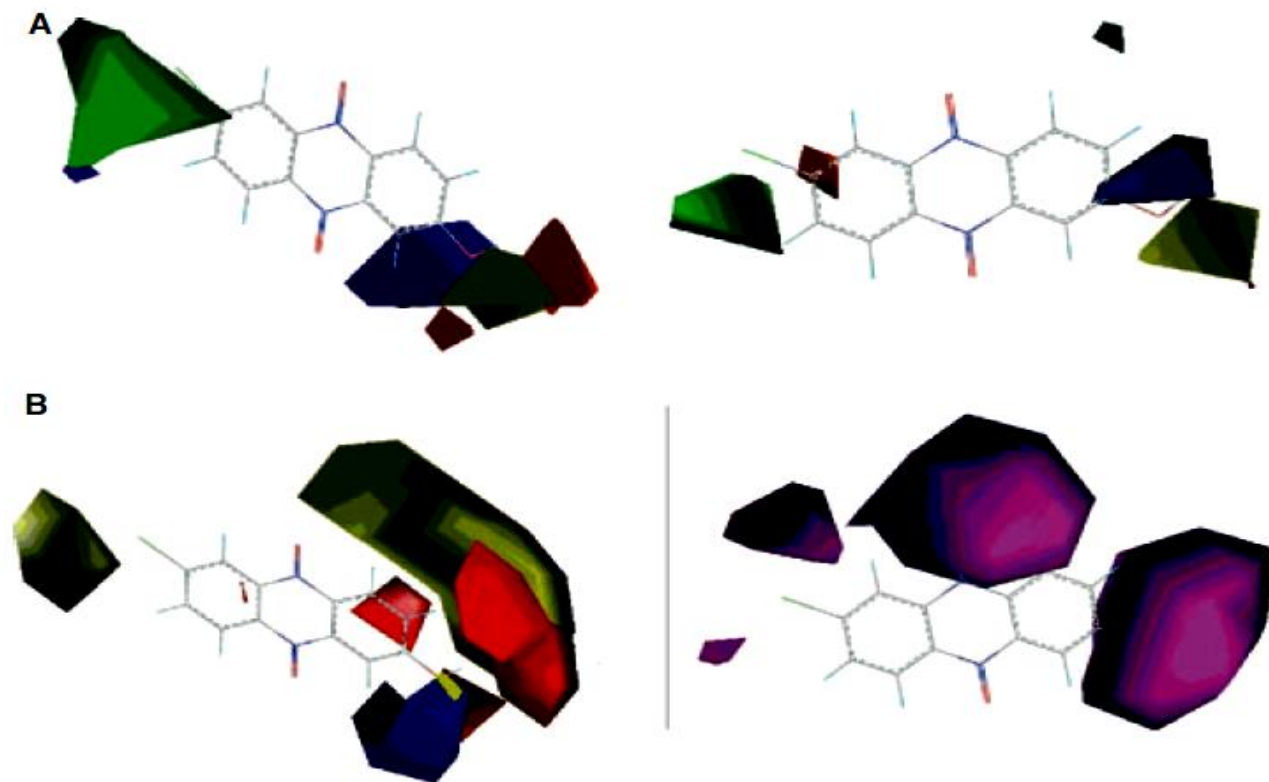
$$SF_{hyp} = -179(\pm 91) + 75(\pm 6)IV_{OH} + 54(\pm 74) cLogP - 727(\pm 203)q_{N_5/N_{12}} - 0.35(\pm 0.45) (cLogP)^2 \quad (5)$$

Outliers : 9 and 14 $n = 28, r = 0.8745, r_{adj}^2 = 0.8527,$
 $F = 40.1, p < 1 \times 10^{-9}$

Dependent variables	Analysis	Model	Other independent variables	r^2	q^2	n	F
SFhyp	CoMSIA (A)	14 ^a	$IV_{OH} + cLogP + (cLogP)^2$	0.858	0.737	5	21.79

^aModels with outliers.

r^2 , non-cross-validated correlation coefficient; q^2 , leave-one-out cross-validated correlation coefficient; n , optimum number of components; F , F -test; S , steric field; E , electrostatic field; A , acceptor hydrogen bond field; CoMFA, comparative molecular field analysis; CoMSIA, comparative molecular similarity indices analysis; SFox, survival fraction in oxia; SFhyp, survival fraction in hypoxia.



RESUMIENDO:

INFORMACIÓN QUE SE PUEDE EXTRAER DEL Q.S.A.R.

Ec. 1

$$\text{Log (1/C)} = 2.342 (\pm 0.105) \log P - 0.731 (\pm 0.041) \log P^2 + 0.036 (\pm 0.019) \sigma + 0.195 (\pm 0.176) E_s + 5.816$$

$$n = 17, r = 0.957, s = 0.126, F = 10.9$$

- 1) **Extrapolar moléculas para aumentar la actividad.**
- 2) **Calcular el valor óptimo de variables independientes.**
- 3) **Plantear nueva serie de compuestos y rediseñar el modelo.**

3.5. OTRAS CORRELACIONES .

3.5.1. RELACIONES ESTRUCTURA QUÍMICA-BIODISPONIBILIDAD

•ESTUDIOS DE UNIÓN CON SEROALBÚMINA de COMPUESTOS NEUTROS MISCELANEOS

$$\log 1/C = 0.75 (\pm 0.07)\log P + 2.30 (\pm 0.15)$$
$$n = 42, r^2 = 0.992, s = 0.159$$

C = conc. Molar que produce complejo ligando: macromolécula 1:1

Temp. = 4°C

$$\log 1/C = 0.67 (\pm 0.10)\log P + 2.60 (\pm 0.22)$$
$$n = 25, r^2 = 0.893, s = 0.242$$

C = conc. Molar que produce complejo ligando: macromolécula 1:1

Temp. = 37 °C

Consideraciones:

-LA AFINIDAD POR LA ALBÚMINA DEPENDE SÓLO DE HIDROFILIA

-NO IMPORTA LA FORMA DEL LIGANDO (hidroxiadamantano, neopentanol, indol, alcánforquinona, naftaleno, acetofenona).

-NO IMPORTA INTERACCIÓN POR EDH.

F. Helmer, K. Kiehs, C. Hansch, *Biochemistry*, 1968, 7, 2858.

J.M., Vandenbelt, C. Hansch, C. Church, *J. Med. Chem.*, 1972, 15, 787.

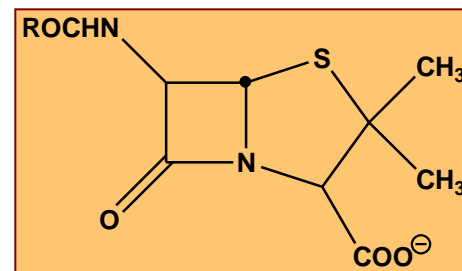
QSAR ANTIBIÓTICOS: Penicilinas actividad frente a *S. aureus*

Actividad *in vivo*

$$\text{Log } 1/\text{CD}_{50} = -0.45\pi + 5.67$$
$$n = 22, r^2 = 0.826, s = 0.191$$



Más potentes



Más hidrófilas

Actividad *in vitro*

No correlación entre actividad y π

Actividad *in vitro* con agregado de seroalbúmina

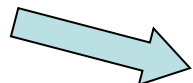
$$\text{Log } 1/\text{MIC} = -0.47\pi + 6.44$$
$$n = 21, r^2 = 0.734, s = 0.267$$



Idénticas dependencias *in vivo* e *in vitro*

UNIÓN A ALBÚMINA HUMANA

$$\text{Log } (B/F) = 0.49\pi - 0.63$$
$$n = 79, r^2 = 0.854, s = 0.134$$



B = conc. Penicilina unida a albúmina
 F = conc. Penicilina libre

Signo opuesto

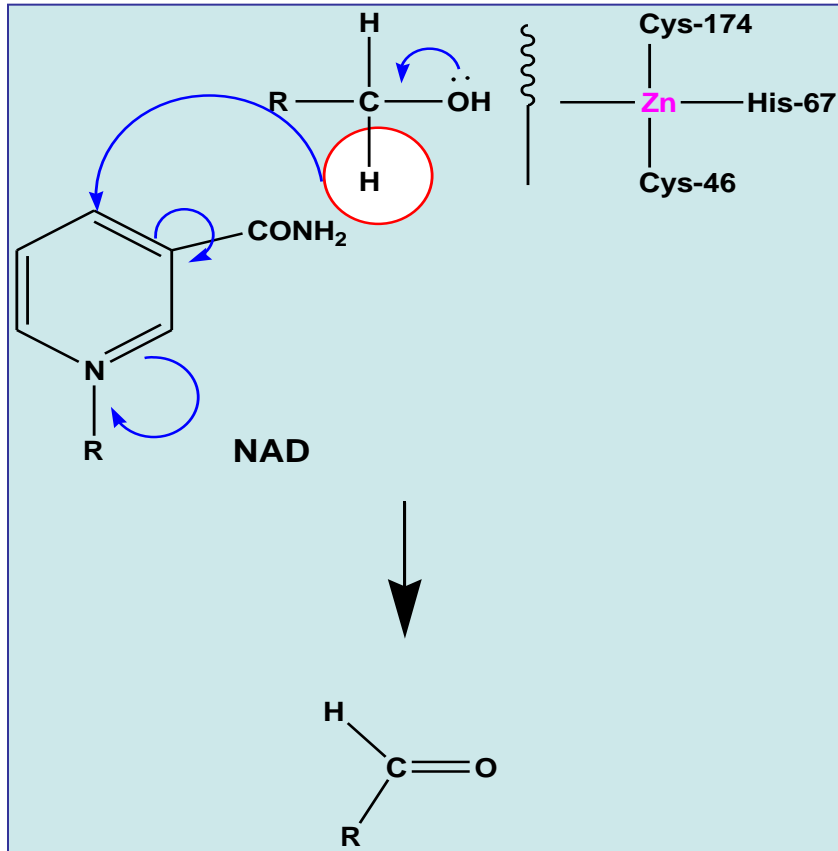


Más hidrofóbico
Mayor unión

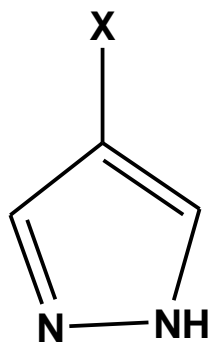
3.5.2. RELACIONES ESTRUCTURA QUÍMICA-ACTIVIDAD ENZIMÁTICA

ALCOHOL DESHIDROGENASA (ADH)

-Blanco para antídoto para intoxicación con MetOH
-Alcoholismo



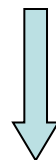
QSAR INHIBIDORES DE ADH DE HÍGADO DE RATA: 4-X-Pirazoles



COMP.	X	COMP.	X	COMP.	X
1	H	6	OCH ₃	11	CH ₃
2	CN	7	OC ₂ H ₅	12	C ₃ H ₇
3	NO ₂	8	OC ₃ H ₇	13	C ₅ H ₁₁
4	NH ₂	9	O-i-C ₃ H ₇	14	C ₆ H ₇
5	NHCOCH ₃	10	I		

$$\text{Log } 1/K_i = 1.22 (\pm 0.33) \log P - 0.20 (\pm 0.78) \sigma_m + 4.87 (\pm 0.28)$$

$$n = 14, r^2 = 0.970, s = 0.316$$



X= grupos electrón donores (σ negativo) aumentan el término σ , aumento de la inhibición

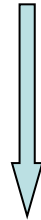
Se explica por aumento de densidad carga negativa sobre N y mayor unión a Zn

Coef. de $\log P \sim 1$ puede indicar una completa desolvatación ocurriendo en la unión

QSAR INHIBICIÓN DE LA OXIDACIÓN DE ALCOHOL EN HEPATOCITOS DE RATA: 4-X-Pirazoles

$$\text{Log } 1/K_i = 1.27 (\pm 0.33) \log P - 0.20 (\pm 0.12)(\log P)^2 - 1.80 (\pm 0.87) \sigma_m + 4.75 (\pm 0.29)$$

$n = 14, r^2 = 0.941, s = 0.320, \log P_0 = 3.2$



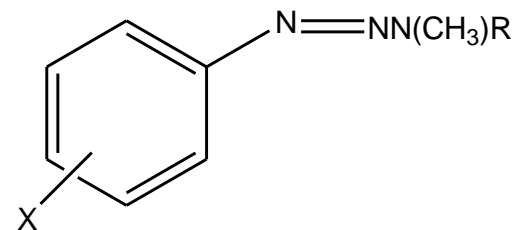
- **Coef. de log P** ~ al QSAR de enzima aislada demuestra que es el proceso *in vivo* que se está considerando.
- **(Log P)²** indica que congéneres muy lipófilos ($\log P > 3.2$) tienen menor acceso a la enzima dentro de la célula.

3.5.3. RELACIONES ESTRUCTURA QUÍMICA-TOXICIDAD

QSAR MUTAGENICIDAD: TEST DE AMES

□ **TRIAZENOS** (usados para el tratamiento de melanoma)

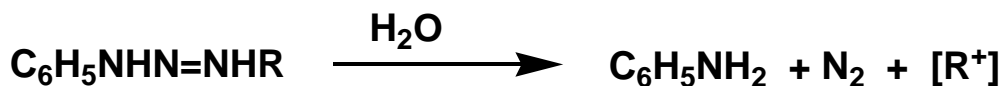
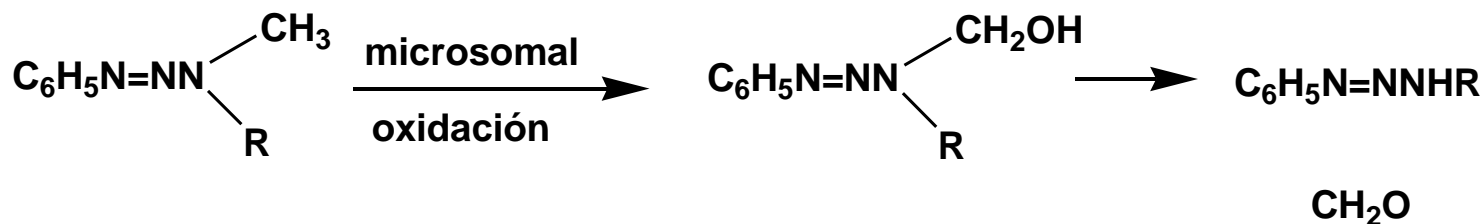
- Ceba *Salmonella typhimurium* TA92 con S9
- C = conc. que produce 30 revertantes por encima del control en 10^8 bacterias



$$\text{Log } 1/C = 1.04 (\pm 0.17) \text{ LogP} - 1.63 (\pm 0.34) \sigma^+ + 3.06 (\pm 0.43)$$

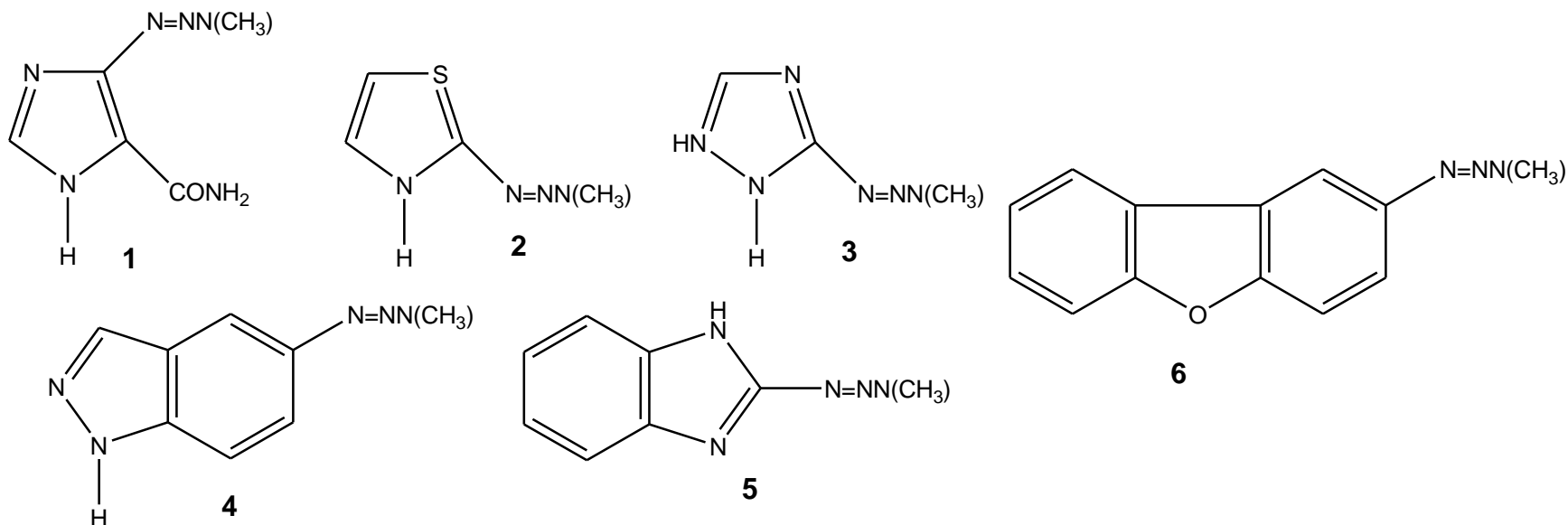
$$n = 17, r^2 = 0.949, s = 0.315$$

Consideraciones: $-\text{Log P} \longrightarrow 58\%$ varianza de $\text{Log}1/C$
 $-\sigma^+ \longrightarrow 37\%$ varianza de $\text{Log}1/C$



Grupos electrón dadores promueven MUTAGENICIDAD

Inclusión de HETEROTRIAZENOS



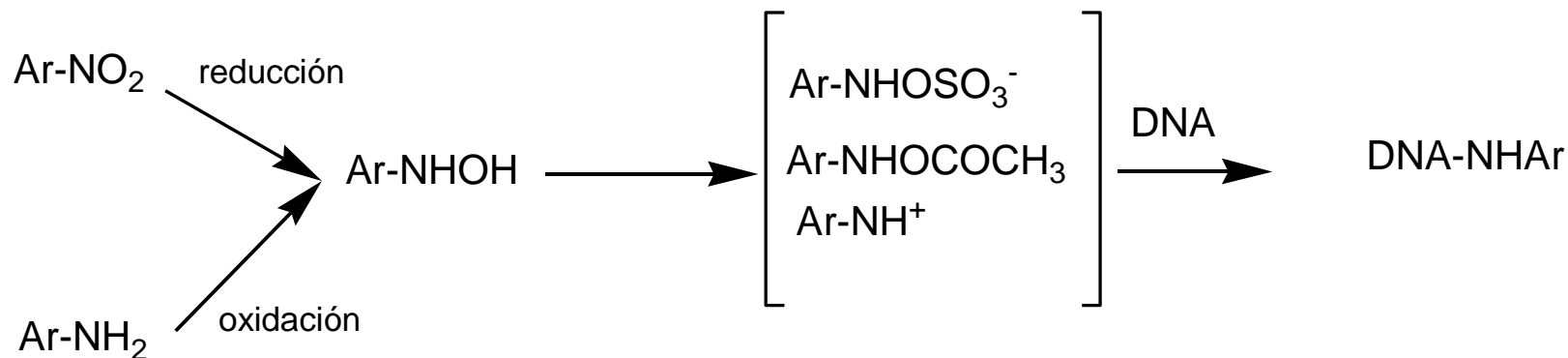
$$\text{Log } 1/C = 0.95 (\pm 0.25) \text{ LogP} + 2.22 (\pm 0.88) E_{\text{HOMO}} + 22.69$$
$$n = 21, r^2 = 0.845, s = 0.631$$

Correlación no tan buena (r^2 y s) pero igual permite explicar en el mismo sentido la mutagenicidad observada

Derivado 6 más mutagénico por ser muy lipófilo y tener el O como grupo dador de electrones

□ NITROAROMÁTICOS Y HETERONITROAROMÁTICOS (PESTICIDAS, FÁRMACOS, RESIDUOS DE PETRÓLEO, NITROPIRENOS EN POLLOS ASADOS)

Sin S9



$$\begin{aligned} \text{Log TA98} = & 0.65(\pm 0.16) \text{ LogP} - 2.90(\pm 0.59) \text{ Log}(\beta 10^{\text{LogP}} + 1) - \\ & 1.38(\pm 0.25) E_{\text{LUMO}} + 1.88 (\pm 0.39) I_1 - 2.89 (\pm 0.81) I_a - 4.15 (\pm 0.58) \\ & n = 188, r^2 = 0.810, s = 0.886, \text{LogP}_o = 4.93 (\pm 0.35) \end{aligned}$$

TA98 = revertantes/ nmol de comp.

I₁ = congéneros con más de 2 anillos fusionados (antracenos, acridinas, etc.)

I_a = grupo acentrileno

$$\text{Log TA98} = 0.65 (\pm 0.16) \text{ LogP} - 2.90 (\pm 0.59) \text{ Log} (\beta 10^{\text{LogP}} + 1) -$$

$$1.38 (\pm 0.25) E_{\text{LUMO}} + 1.88 (\pm 0.39) I_1 - 2.89 (\pm 0.81) I_a - 4.15 (\pm 0.58)$$

$$n = 188, r^2 = 0.810, s = 0.886, \text{LogP}_o = 4.93 (\pm 0.35)$$

Consideraciones: -menor E_{LUMO} más rápida reducción por tanto MÁS MUTAGÉNO

La vel. de reducción depende del E_{LUMO} queda demostrado por:

$$\text{Log K} = -1.53 (\pm 0.36) E_{\text{LUMO}} - 0.06 (\pm 0.36)$$

$$n = 21, r^2 = 0.805, s = 0.242,$$

$$\text{Log K} = 1.09 (\pm 0.20) \sigma^- + 1.76 (\pm 0.11)$$

$$n = 21, r^2 = 0.876, s = 0.192,$$

Log K = vel. de reducción nitrobenzenos por xantina oxidasa

Coef. E_{LUMO} es similar entre las tres ecuaciones similar mecanismo reducción.

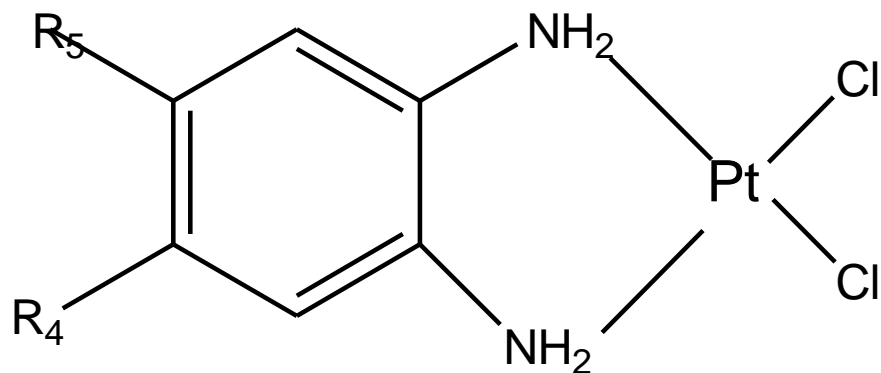
$$\begin{aligned} \text{Log TA98} = & 0.65 (\pm 0.16) \text{LogP} - 2.90 (\pm 0.59) \text{Log} (\beta 10^{\text{LogP}} + 1) - \\ & 1.38 (\pm 0.25) E_{\text{LUMO}} + 1.88 (\pm 0.39) I_1 - 2.89 (\pm 0.81) I_a - 4.15 (\pm 0.58) \\ & n = 188, r^2 = 0.810, s = 0.886, \text{LogP}_o = 4.93 (\pm 0.35) \end{aligned}$$

Consideraciones: -menor E_{LUMO} más rápida reducción por tanto **MÁS MUTAGÉNO**
-relación bilinear con hidrofobicidad indica que la reducción es endocelular
- I_1 y I_a estaría marcando la importancia de factores estéricos

MUTAGENICIDAD DIRECTA

Complejos de Pt y aminas

Mutagenicidad medida en TA92

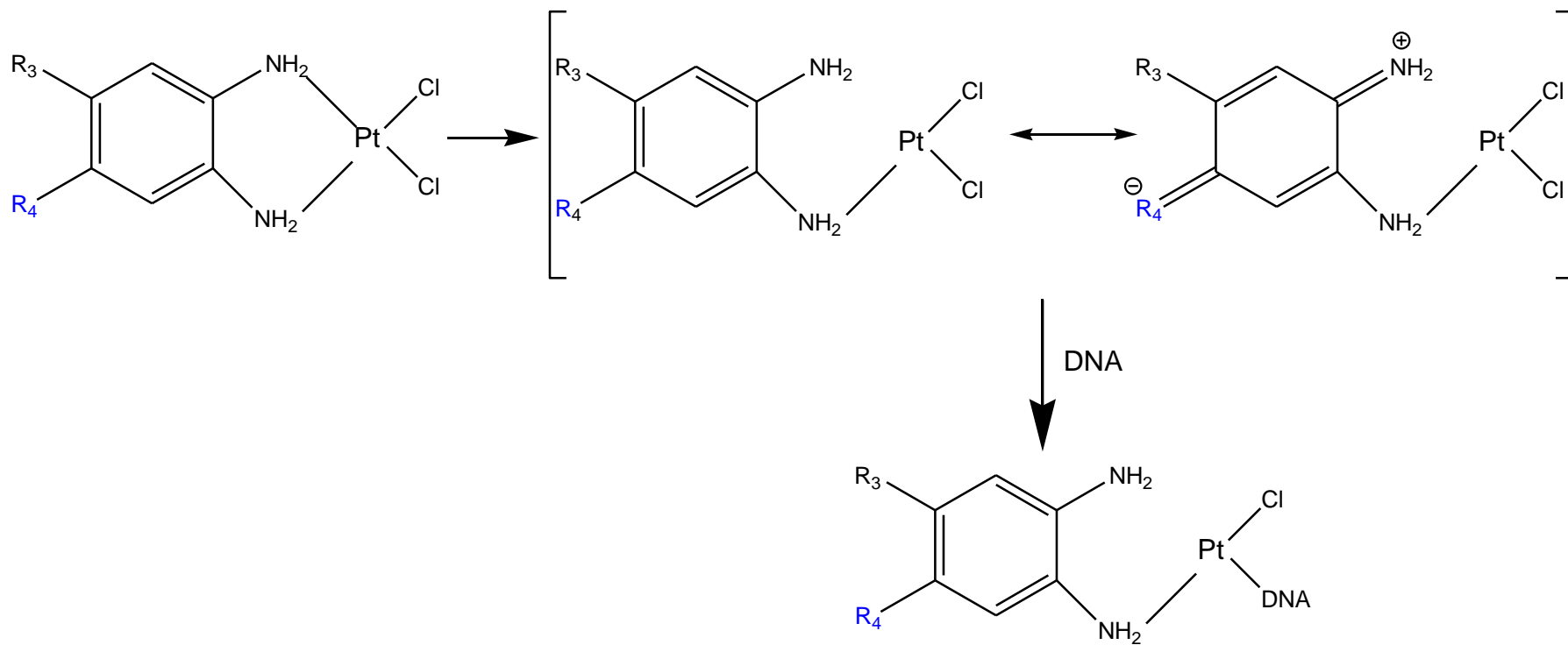


$$\text{Log } 1/C = 2.23 (\pm 0.32) \sigma_p^- + 5.78 (\pm 0.18)$$

$$n = 13, r^2 = 0.956, s = 0.260$$

4-NO₂ 10.000 VECES MÁS MUTAGÉNICO QUE EL MENOS 4,5-DI-OCH₃

Mecanismo propuesto de acuerdo QSAR



EJEMPLOS FUERA DEL TEMA DE FÁRMACOS

QSAR Models for the Acute Toxicity of 1,2,4-Triazole Fungicides to Zebrafish (*Danio rerio*) Embryos

<https://doi.org/10.1016/j.envpol.2020.114837>

Kun Qiao^{1,2}, Wenjie Fu³, Yao Jiang¹, Lili Chen¹, Shuying Li¹, Qingfu Ye², Wenjun Gui^{1*}



$$pLC_{50} = -7.24 - 0.30 X^V PC_4 + 0.76 \log D - 26.15 Q_{N1} - 0.08 \mu$$

descriptor topológico

descriptor hidrofobicidad

carga (descriptor electrónico)

momento dipolar (descriptor electrónico)



ELSEVIER

Contents lists available at ScienceDirect

Data in brief

journal homepage: www.elsevier.com/locate/dib

Data Article

Acute toxicity dataset for QSAR modeling and predicting missing data of six pesticides

Ya-Qian Xu ^{a, c}, Shu-Shen Liu ^{a, b, c, *}, Bing-Qing Lu ^c,
Ze-Jun Wang ^b

2.2. QSAR prediction

Various molecular descriptors including two-dimensional (2D) topological indices, constitutional indices, ring descriptors, walk and path counts, CATS 2D, and 2D atom-pair descriptors were computed by DRAGON (version 7) [4,5]. Further, variable selection and optimization were performed using the VIPLS [6] and VSMP programs [7]. For more details, refer to the original article, “*Conlecs: A novel procedure for deriving the concentration limits of chemicals outside the criteria of human drinking water using existing criteria and species sensitivity distribution based on quantitative structure-activity relationship prediction*” [1]. Ten robust QSAR models were developed by the above method and the acute toxicities of six pesticides for some species were predicted by the ten developed QSAR models. The values of the optimal descriptors in the QSAR models are listed in Table 1, and the values of the acute toxicities predicted by the models are provided in Table 2.



Efficacy coefficient method assisted quadruple-activities 3D-QSAR pharmacophore model for application in environmentally friendly PAE molecular modification

Meijin Du^{1,2} · Youli Qiu³ · Qing Li^{1,2} · Yu Li^{1,2}

Received: 23 January 2020 / Accepted: 1 April 2020
 © Springer-Verlag GmbH Germany, part of Springer Nature 2020

Abstract

Phthalate acid esters (PAEs) are among the most widely used plasticizers in plastic products. They are easily diffused from plastic during use and seriously affect the environment and human health. Therefore, designing environmentally friendly PAE derivatives has important practical applications. In this paper, the environmentally friendly molecular modification of PAEs was carried out according to a comprehensive structural evaluation based on a three-dimensional quantitative structure–activity relationship (3D-QSAR) pharmacophore model of four activity modes. First, the efficacy

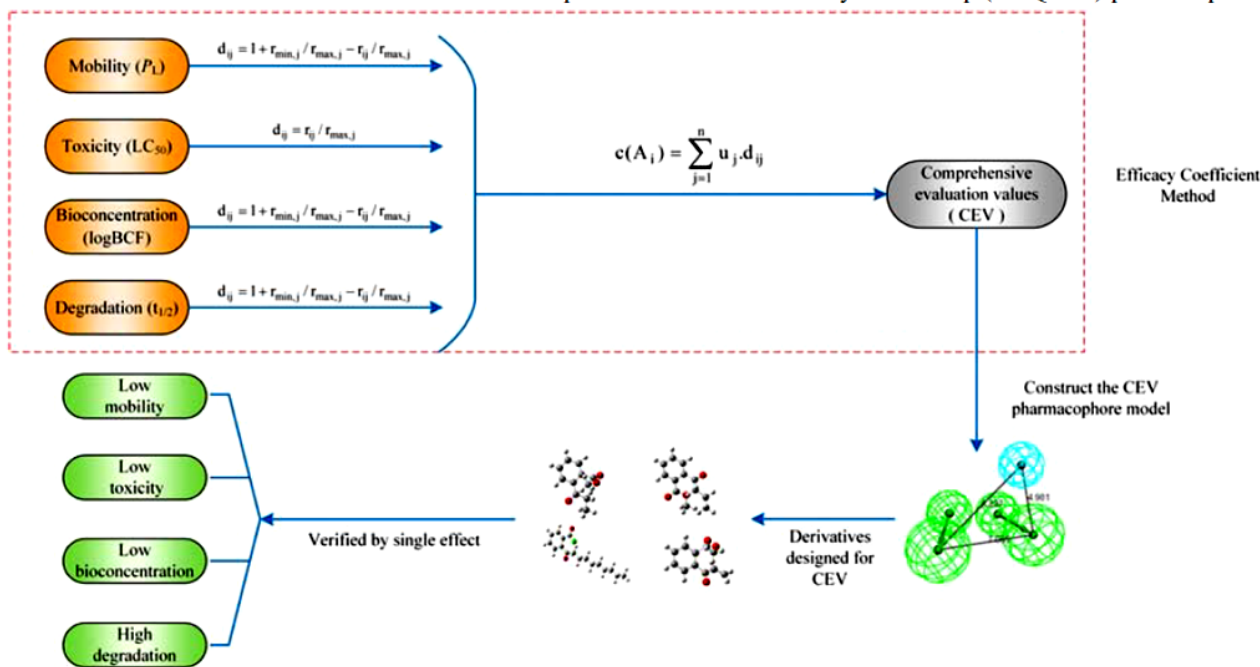


Fig. 1 The process of calculating the comprehensive evaluation values by using the efficacy coefficient method and the design of the novel molecules developed using the pharmacophore model



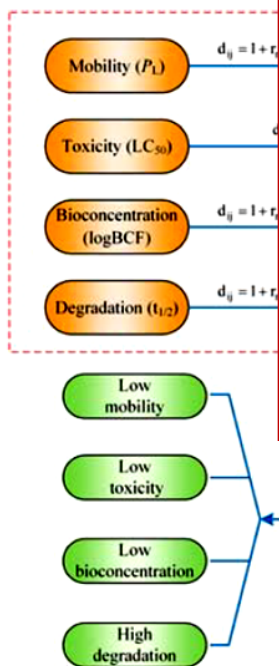
Efficacy coefficient method assisted quadruple-activities 3D-QSAR pharmacophore model for application in environmentally friendly PAE molecular modification

$$LC_{50} = 377.717 - 2.565\mu - 322.271q^+ - 20.611EG + 0.05TE - 54.939qH^+ \quad (6)$$

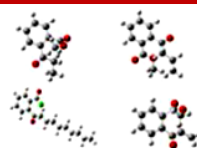
$$t_{1/2} = 7187.145 - 279.119\mu - 2728.18q^+ - 130.215EG + 3.083TE + 795.723qH^+ \quad (7)$$

$$P_L = 0.116 - 0.03\mu - 0.381q^+ + 0.125EG - 0.704qH^+ \quad (8)$$

$$\log BCF = -4.002 - 0.047\mu + 9.199q^+ - 0.658EG - 0.004TE + 3.519qH^+ \quad (9)$$



Verified by single effect



Derivatives designed for CEV

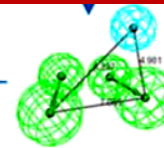


Fig. 1 The process of calculating the comprehensive evaluation values by using the efficacy coefficient method and the design of the novel molecules developed using the pharmacophore model