



A vivid green aurora is reflected in the Jokulsrion Glacier lagoon at Vatnajokull National Park in Iceland

British photographer James Woodend won the title of Astronomy Photographer of the Year.
James Woodend, Courtesy of Royal Museums Greenwich

Diseño de Compuestos Activos Basado en el Ligando

Contenido

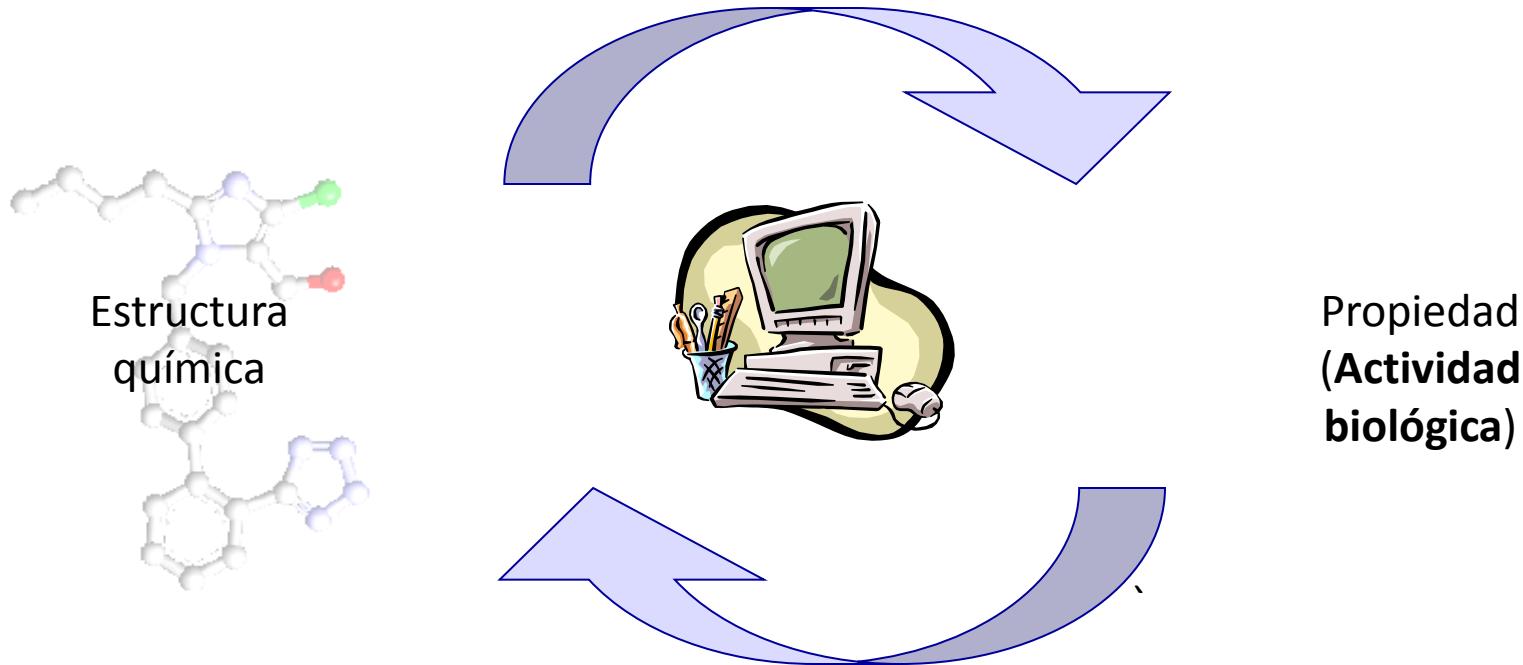
- Relaciones (cuantitativas) estructura-actividad:
(Quantitative) Structure-Activity Relationships, (Q)SAR
 - Relaciones estructura múltiple-actividad (*SmAR*)
- Panoramas de actividad (*activity landscape*) y acantilados de actividad (*activity cliffs*)
- Similitud molecular
- Búsqueda por farmacóforo
- Cribado virtual basado en el ligando
- Casos de estudio (presentaciones de alumnos)



RELACIONES (CUANTITATIVAS) ESTRUCTURA-ACTIVIDAD (QUANTITATIVE) STRUCTURE-ACTIVITY RELATIONSHIPS, (Q)SAR

Relaciones Cuantitativas Estructura-Propiedad

Quantitative structure-property relationships (QSPR)



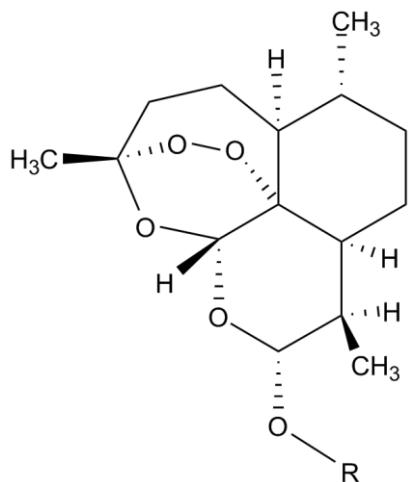
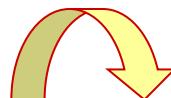
Relaciones cuantitativas estructura-actividad

*Quantitative structure-**activity** relationships (QSAR)*

Hopfinger, A. J. *J. Med. Chem.* 1985, 28, 1133

Relación Cuantitativa Estructura- Actividad (QSAR)

¿Relación matemática?



Objetivo: predecir la actividad de nuevos compuestos

SD₉₀	R	Descripción (Descriptor)
3.65	H	D ₁
1.16	CH ₃	D ₂
1.20	COCH ₃	D ₃
0.66	COC ₂ H ₃	D ₄
3.42	COC ₆ H ₅	D ₅
???	Y	

Historia del QSAR: Kubini, H. *Quant. Struct.-Act. Relat.* 2002, 21, 348

(Q)SPR / (Q)SAR en Química

- Práctica básica
 - Análisis retrospectivo: descripción
 - Aplicaciones prospectivas – diseño de compuestos: predicción
- Estrategias
 - Cualitativo: SPR/SAR
 - **Cuantitativo: QSPR/QSAR**
 - No necesariamente predictivo



Ejemplos de Métodos Cuantitativos

- QSAR
 - Actividad en escala continua (K_i , IC_{50} , EC_{50} ...)
- Rule-based
 - Actividad en escala categórica (activo/inactivo; alto/medio/bajo): decision trees, rough set theory
- Machine learning
- Activity landscape: panoramas de actividad
 - Muchos métodos están enfocados en describir el SAR de colecciones grandes de compuestos
 - Entender el SAR antes de predecir actividad



Requerimientos de la Representación Molecular

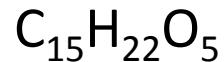
- Preservación de la información o permisión controlada de perdida de información
- Características de selección
 - Por conocimiento de interacciones (e.g. unión con el receptor, cualquier información del mecanismo de acción)
 - Por verificación de la consideración de «vecinos cercanos»
 - Por métodos de selección de características
 - Ejemplos: PCA, agrupamiento, etc.
 - Los compuestos deben agruparse compactamente dentro de una clase y estar alejados entre diferentes clases
- Combinación de diferentes medidas (consensus)



¿Cómo Describir a una Molécula Cuantitativamente?

Representación

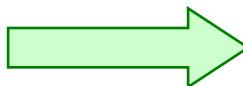
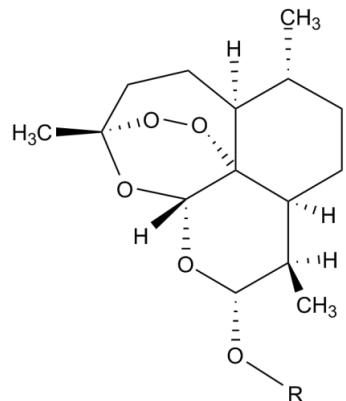
1D



Descriptores

peso molecular,
número de átomos, etc.

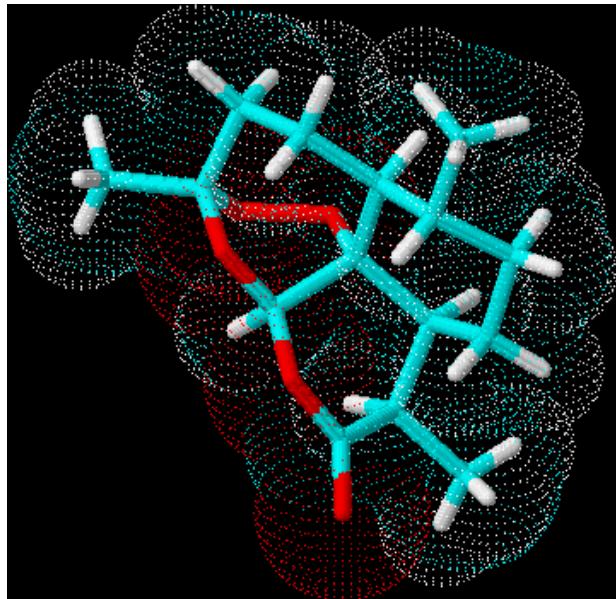
2D



conectividad; número
de anillos; coeficiente
de partición (logP), etc.



¿Cómo Describir a una Molécula Cuantitativamente?



3D



Descriptores

volumen; superficie
accesible al disolvente;
energía; HOMO; LUMO;
etc.

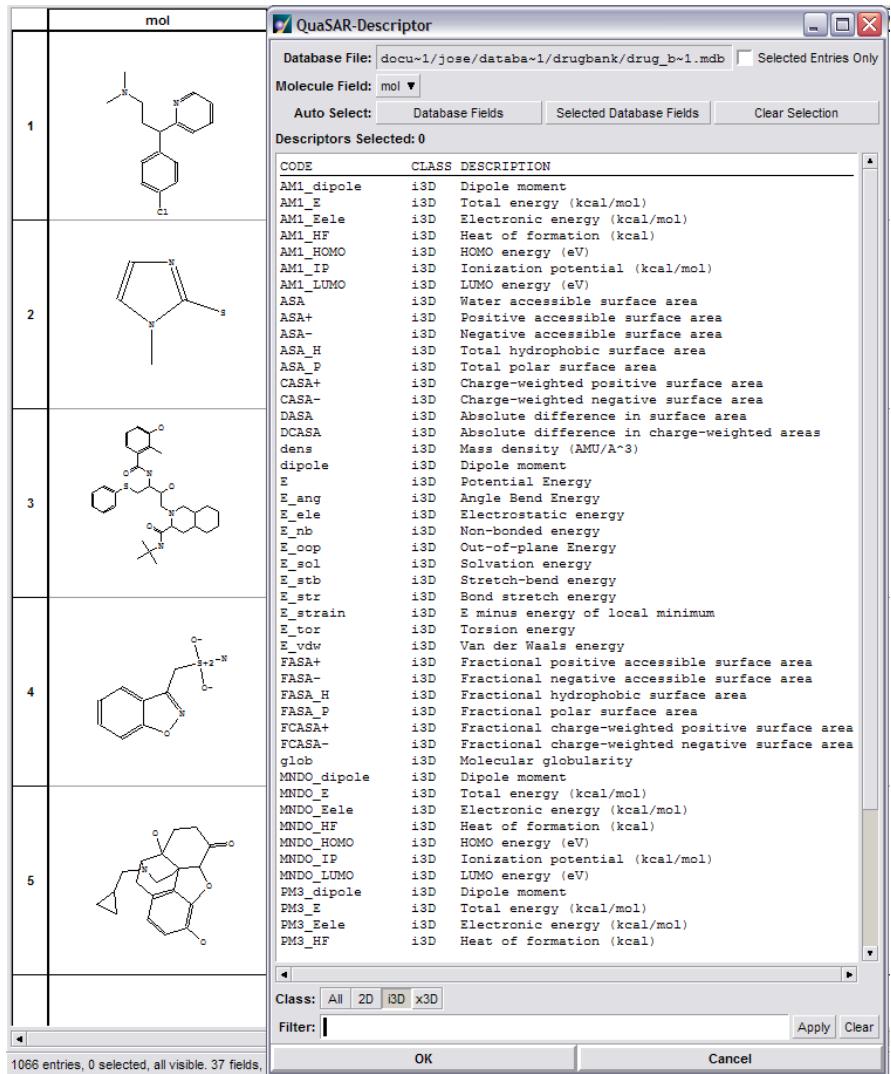


Cálculo de Descriptores

Programas de cómputo

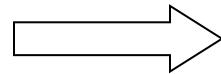
- MOE
- Hyperchem
- ACD Labs
- Alchemy
- Jaguar
- Gaussian
- Spartan
- Sybyl
- Macromodel, etc.

MOE



Etapas de un Análisis QSAR

I. Estructuras (ca. > 10) y su actividad en escala continua*



II. Cálculo de descriptores

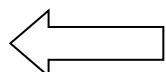
SD ₉₀	R
3.65	H
1.16	CH ₃
1.20	COCH ₃
0.66	COC ₂ H ₃

R	PM	V	HOMO	...	n _z
H					
CH ₃					
COCH ₃					
COC ₂ H ₃					

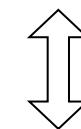
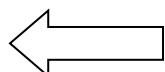


III. Relación actividad-descriptores
(análisis estadístico)

IV. Evaluación del modelo



Modelo
QSAR



Selección de descriptores

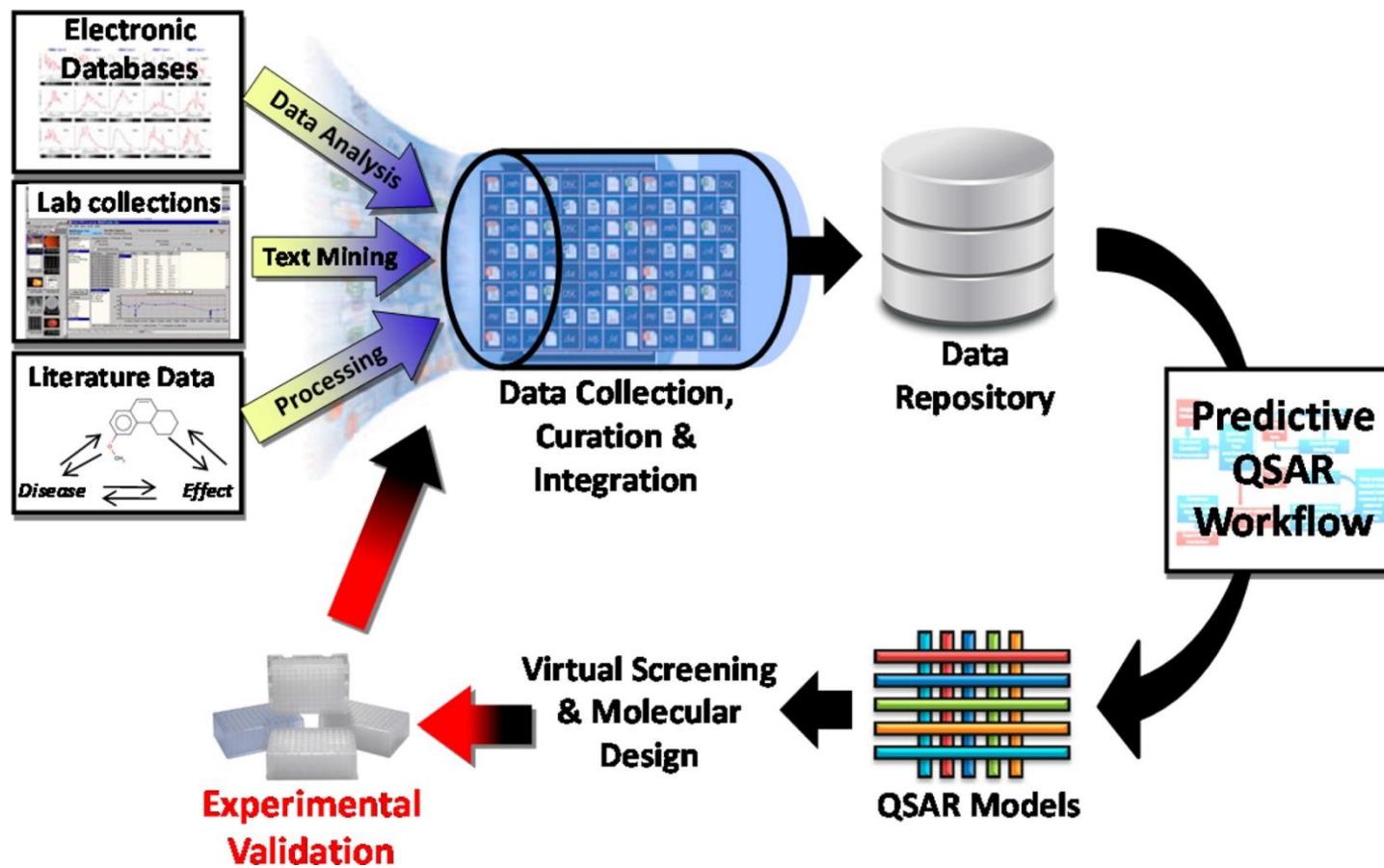
*Actividades discretas (activo/inactivo) pueden analizarse con otros métodos

Etapas de un Análisis QSAR

1. Recolección y preparación de datos
2. Identificación de *activity cliffs*
3. Cálculo de descriptores
4. Construcción de modelos
5. Validación externa
6. Interpretación de modelos
7. Uso y aplicación



Etapas de un Análisis QSAR



Cherkasov A, et al. J. Med. Chem. 2014, 57, 4977

QSAR 2D: Análisis de Hansch

La actividad biológica / propiedad (variable dependiente) es una combinación de los descriptores (variables independientes)

$$\log 1/C = a D_1 + b D_2 + \dots + z D_n + \text{constante}$$

C: Concentración molar (dosis) que produce una respuesta biológica (ED_{50} , IC_{50} , LD_{50})

D₁, D₂, D_n: Descriptores

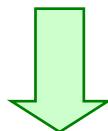
a, b, z: Coeficientes obtenidos por análisis de regresión

Hasch C, Fujita T. *J. Am. Chem. Soc.* 1964, 86, 116

Análisis de Hansch

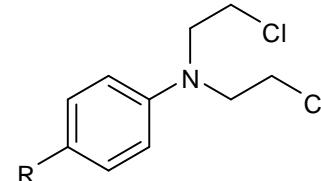
R	Actividad	Descriptores	
	$\log 1/(B + 100)$	σ	π
Cl	1.02	0.23	0.71
NH ₂	1.40	-0.66	-1.23
OCH ₃	1.33	-0.27	-0.02
CH ₃	1.77	-0.17	0.56
n-Bu	1.77	-0.16	2.05
OPh	2.13	-0.03	2.08
CN	0.57	0.66	-0.57

* $\log / (B+100)$: medida de mutagenicidad



R
SO ₂ NH ₂
CH=C(CN) ₂
NHCONH ₂
CH=CHC ₆ H ₅

Nuevas moléculas
¿Cuáles serán más mutagénicas?



$$\log / (B + 100) = 0.31 \pi - 0.82 \sigma + 1.22$$

$n = 7, r = 0.93, s = 0.24$

La mutagenicidad aumenta:

- Con sustituyentes R hidrófobos (π alto)
- Con grupos R electrodonadores ($\sigma < 0$)



Análisis de Hansch

1. Cálculo de descriptores

R	σ	π
SO_2NH_2	0.57	-1.82
$\text{CH}=\text{C}(\text{CN})_2$	0.7	0.05
NHCONH_2	-0.24	-1.3
$\text{CH}=\text{CHC}_6\text{H}_5$	-0.07	2.68

Modelo QSAR

2. Cálculo de actividad

R	$\log 1/(B + 100)$
SO_2NH_2	0.19
$\text{CH}=\text{C}(\text{CN})_2$	0.66
NHCONH_2	1.01
$\text{CH}=\text{CHC}_6\text{H}_5$	2.11

Mayor prioridad
Mutagenicidad calculada



Análisis de Hansch

1. Cálculo de descriptores

R	σ	π
SO_2NH_2	0.57	-1.82
$\text{CH}=\text{C}(\text{CN})_2$	0.7	0.05
NHCONH_2	-0.24	-1.3
$\text{CH}=\text{CHC}_6\text{H}_5$	-0.07	2.68

Modelo QSAR
→

2. Cálculo de actividad

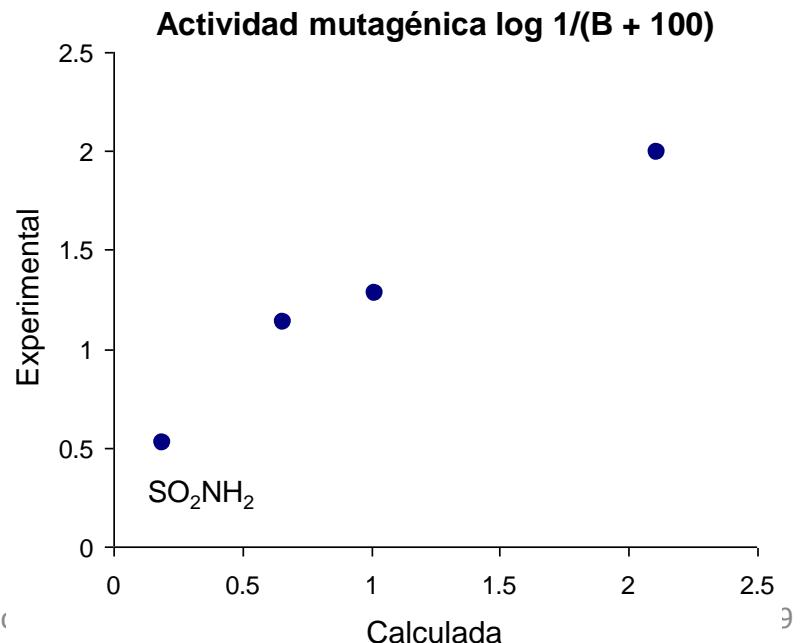
R	$\log 1/(B + 100)$
SO_2NH_2	0.19
$\text{CH}=\text{C}(\text{CN})_2$	0.66
NHCONH_2	1.01
$\text{CH}=\text{CHC}_6\text{H}_5$	2.11

Mayor prioridad
↓
Mutagenicidad calculada

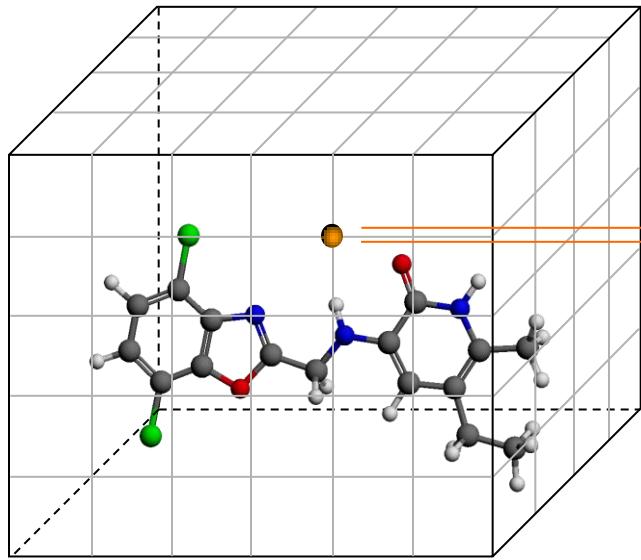
3. Comparación con la *tendencia* experimental

R	Log $1/(B + 100)$	
	Calculada	Experimental
SO_2NH_2	0.19	0.52
$\text{CH}=\text{C}(\text{CN})_2$	0.66	1.13
NHCONH_2	1.01	1.28
$\text{CH}=\text{CHC}_6\text{H}_5$	2.11	1.99

La tendencia es la misma



QSAR 3D Comparative Molecular Field Analysis



Molécula **1** con n átomos i

● Átomo de prueba j en el punto x de la rejilla

Modelo CoMFA

$$B = a + c_{S_1} S_1 + \dots + c_{S_x} S_x + c_{E_1} E_1 + \dots + c_{E_x} E_x$$

PLS

Química Farmacé

Energía de interacción electrostática (E_x)

$$E_C = \sum_{i=1}^n \frac{q_i q_j}{Dr_{ij}}$$

Energía de interacción estérica (S_x)

$$E_{vdW} = \sum_{i=1}^n (A_{ij} r_{ij}^{-12} - C_{ij} r_{ij}^{-6})$$

Tabla QSAR

	B	S₁	...	S_x	E₁	...	E_x
1	A ₁	S _{1,1}	...	S _{x,1}	E _{1,1}	...	E _{y,1}
...
m	A _n	S _{1,n}	...	S _{x,n}	E _{1,n}	...	E _{y,n}



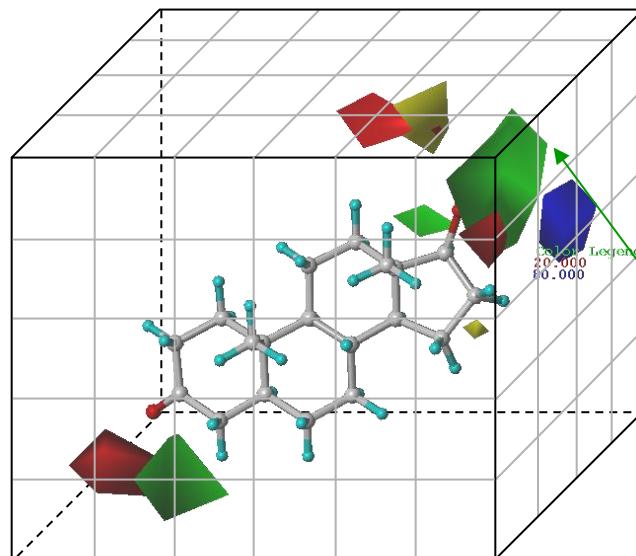
Visualización e Interpretación del Modelo CoMFA

$$B = a + c_{S_1}S_1 + \dots + c_{S_x}S_x + c_{E_1}E_1 + \dots + c_{Ex}E_x$$

Se grafica para cada punto ($stdev * c_{Sx}$): el coeficiente multiplicado por la raíz cuadrada de la **varianza de un campo** en ese punto. Se despliegan los gráficos (contornos) que tienen un valor por encima y por debajo de un *threshold*.

Contornos rojos:
Interacciones con $q > 0$
desfavorecen actividad
(nota: frecuentemente se usa un probe atom con carga 1+)

Contornos azules:
Interacciones con $q > 0$
favorecen actividad
(nota: frecuentemente se usa un probe atom con carga 1+)



Mapas de contorno

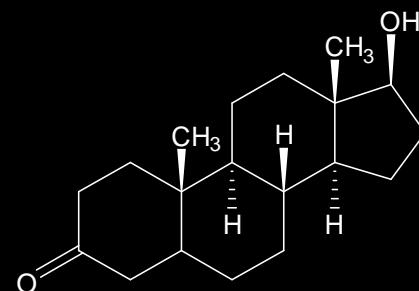
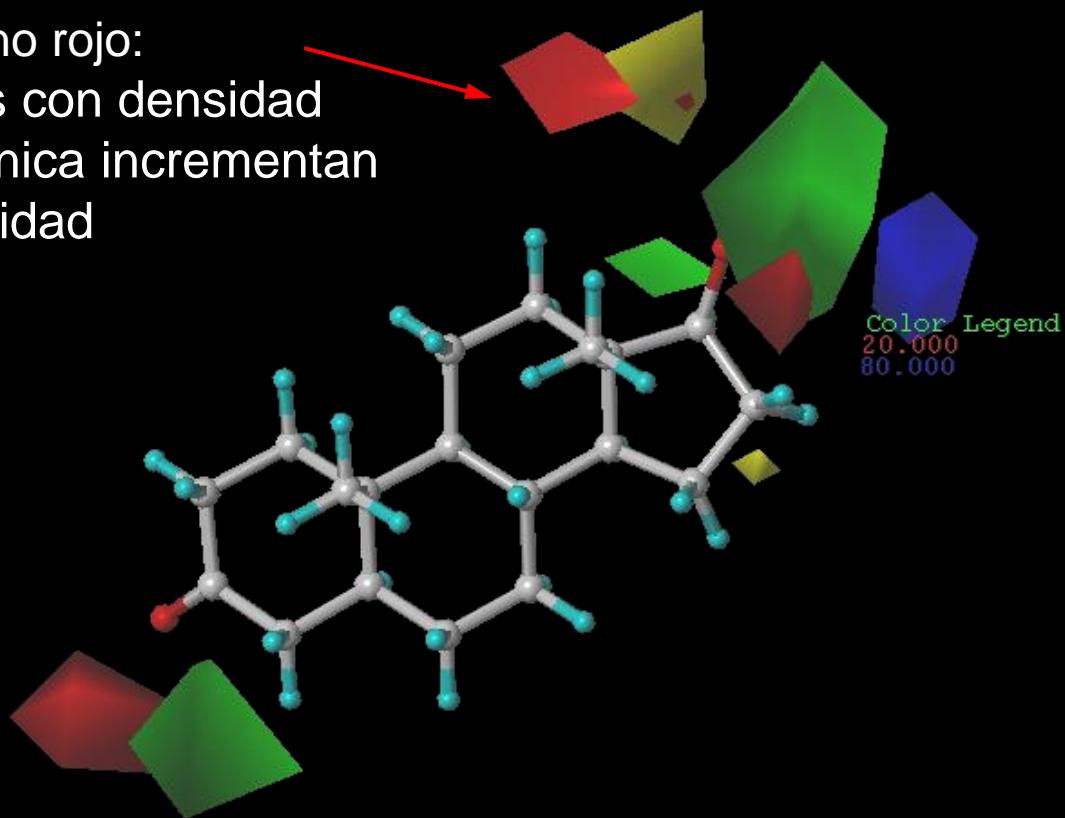
Contornos amarillos:
Interacciones estéricas
desfavorecen actividad

Contornos verdes:
Interacciones estéricas
favorecen actividad

QSAR-3D: Análisis Comparativo de Campos Moleculares (CoMFA)

Contorno rojo:

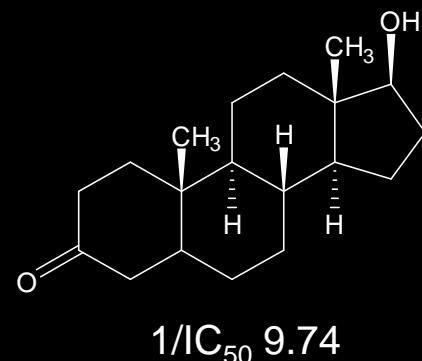
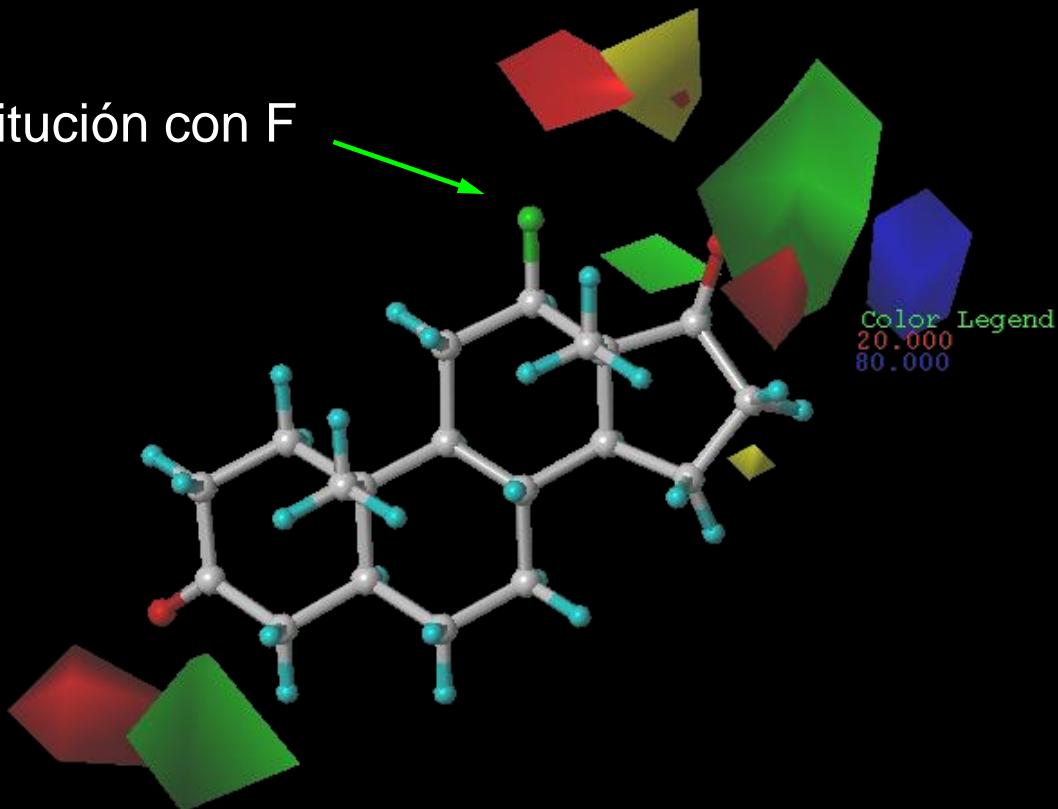
Grupos con densidad electrónica incrementan la actividad



Dehidrotestosterona
 $1/\text{IC}_{50} \ 9.74$

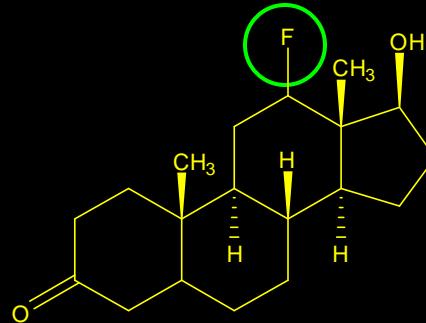
Optimización de la actividad con CoMFA

Sustitución con F

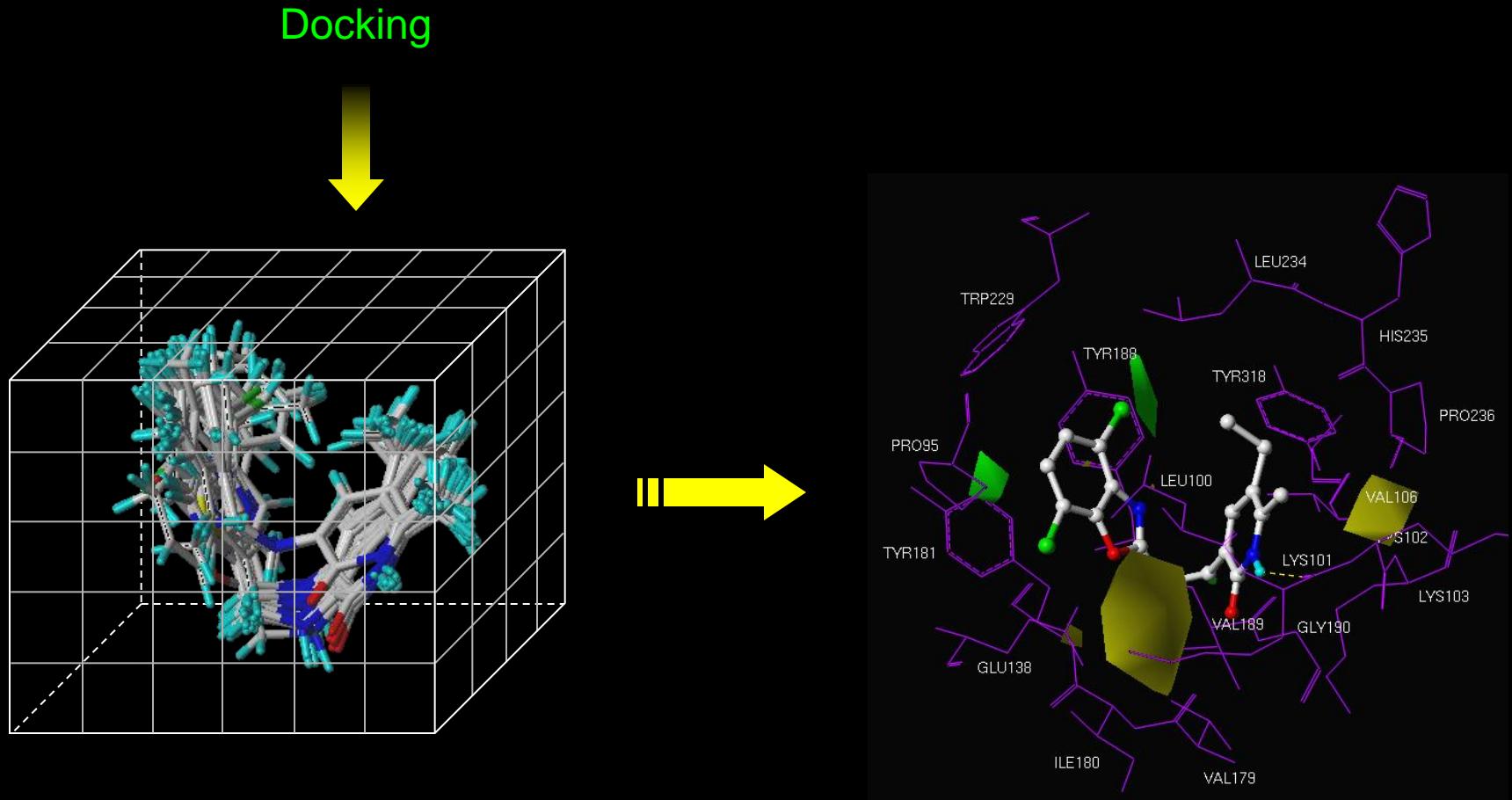


Incremento en la actividad

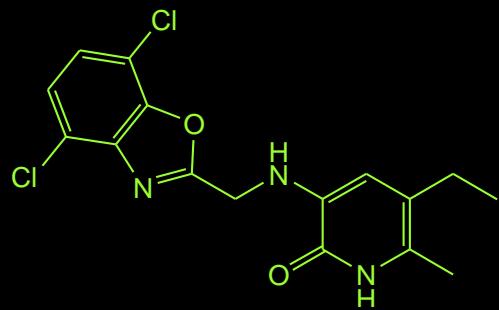
$1/\text{IC}_{50} 9.98$



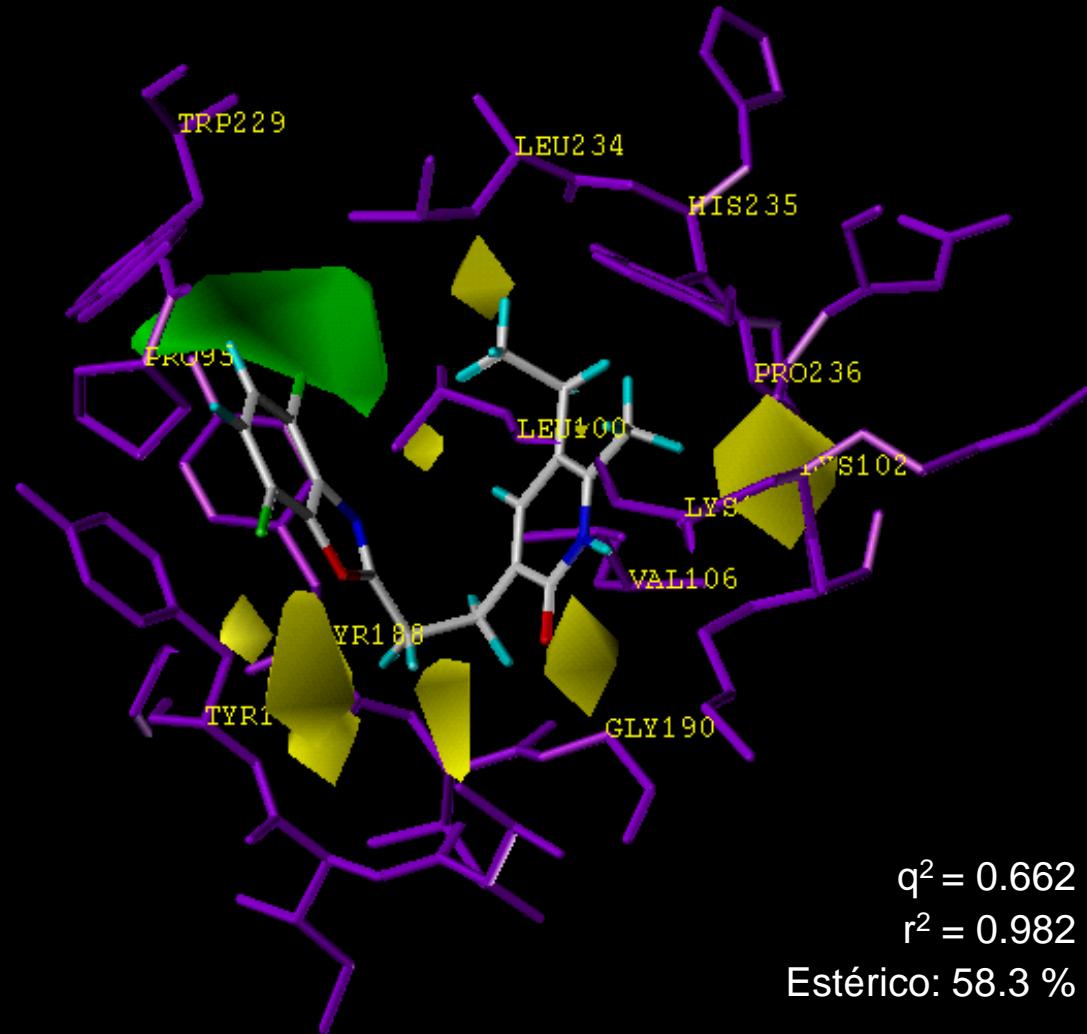
Combinación *docking* con CoMFA



QSAR 3D (CoMFA)



Inhibidores de la transcriptasa
reversa de VIH-1

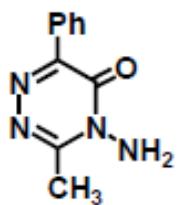


$$q^2 = 0.662$$

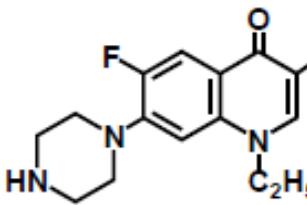
$$r^2 = 0.982$$

Estérico: 58.3 %

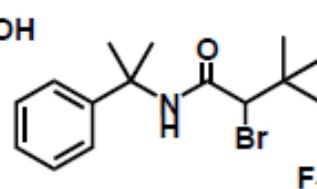
Commercialized Drugs developed with the Aid of Classical QSAR



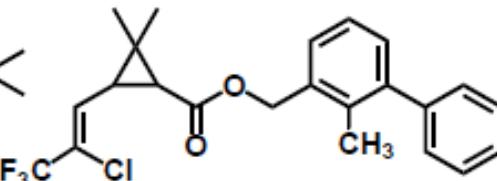
Metamitron
(Sugar-beet
Herbicide)
Bayer 1975



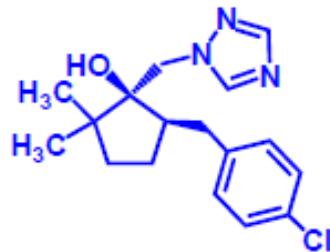
Norfloxacin
(Antibacterial)
Kyorin 1983



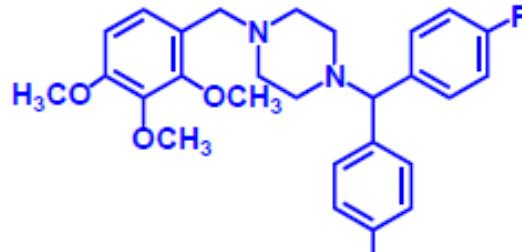
Bromobutide
(Paddy Field
Herbicide)
Sumitomo 1984



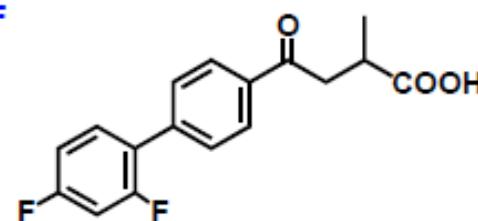
Bifenthrin
(Foliar Insecticide)
FMC 1984



Metconazole
(Wheat Fungicide)
Kureha 1994



Lomerizine
(Antimigraine, Antiglaucoma)
Organon Japan-Upjohn 1999



Flobufen
(Long-acting Antiinflammatory)
Kuchar et al.-Virbac 2000

Toshio Fujita; Kyoto University, Kyoto, Japan



PANORAMAS DE ACTIVIDAD Y ACANTILADOS DE ACTIVIDAD *ACTIVITY LANDSCAPE & ACTIVITY CLIFFS*

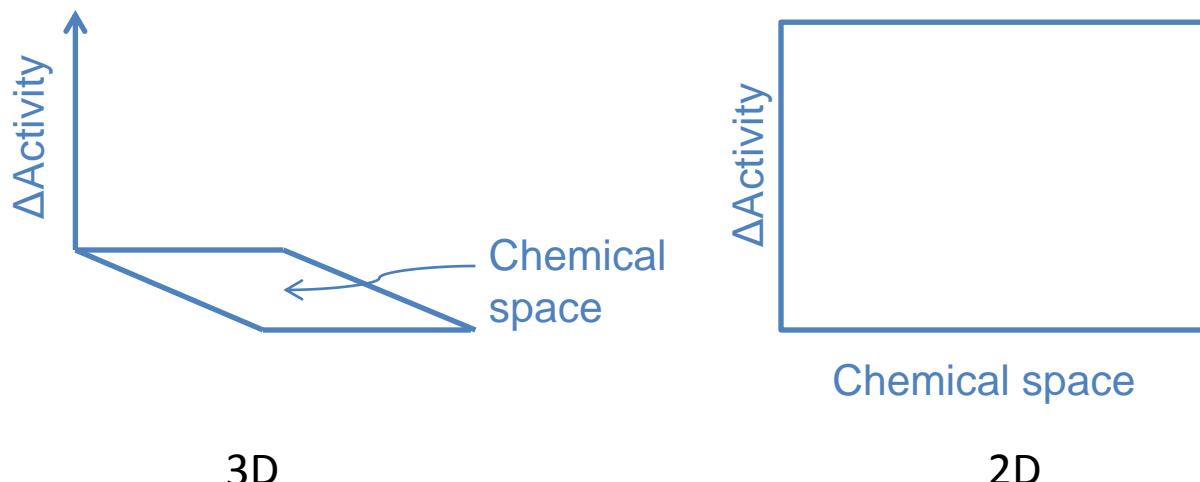


Perfil SAR

Activity Landscape Modeling

“Any representation that integrates the analysis of the **structure similarity** and **potency differences** between compounds sharing the same biological activity.”

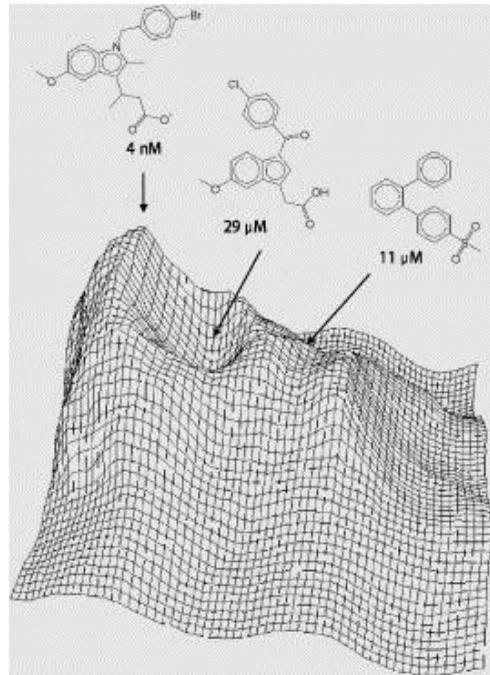
Wassermann, AM; Wawer M; Bajorath, J. *J.Med. Chem.*, 53, 8209, 2010



Perfil SAR y *Activity Cliffs*

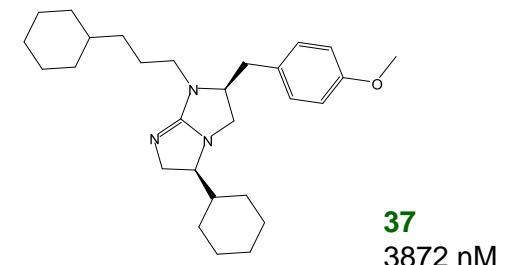
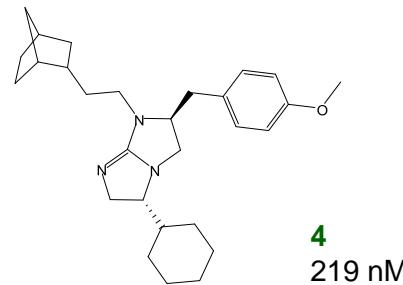
Principio de similitud

Moléculas semejantes tienen propiedades semejantes



Activity cliff

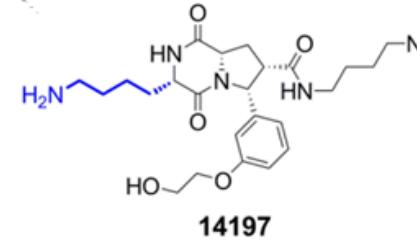
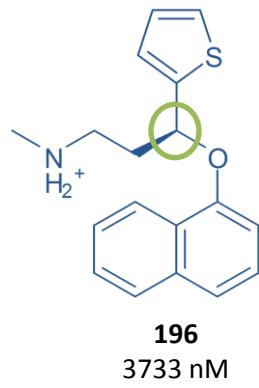
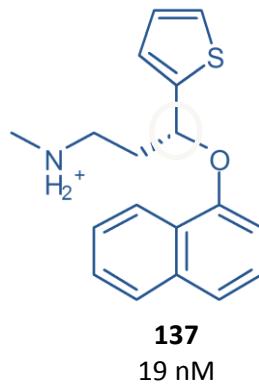
Moléculas **semejantes** tienen propiedades **muy diferentes**



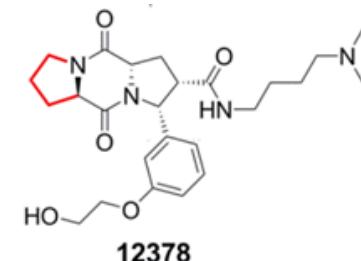
Bajorath J. et al. Navigating structure-activity landscapes *Drug Discovery Today* 2009, 14, 698

Acantilados de Actividad

Activity cliffs



Binds 76 proteins



Binds 1 protein

Activity cliff

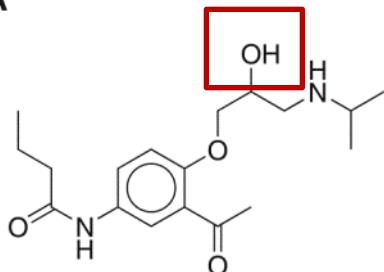
“Compounds, often series of analogues, where small structural modifications lead to significant changes in potency”

Medina-Franco, J. L. J. Chem. Inf. Model. 2012, 52, 2485

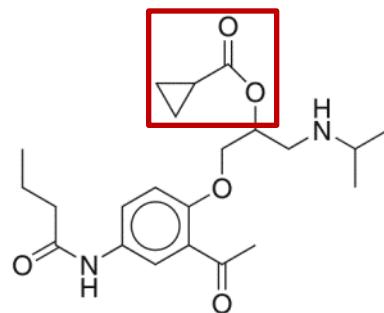
Acantilados de Propiedad

Property cliffs

A

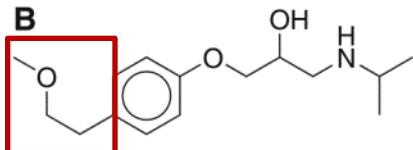


Acebutolol
 $\log P_{\text{eff}} = -5.83$

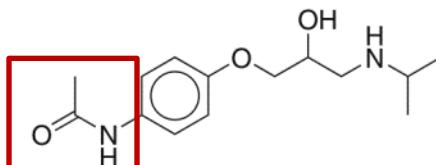


Acebutolol ester
 $\log P_{\text{eff}} = -4.61$

B



Metroprolol
 $\log P_{\text{eff}} = -4.59$



Practolol
 $\log P_{\text{eff}} = -6.05$

Acantilados de propiedad
Par de compuestos con
estructura similar pero
permeabilidad Caco-2 muy
diferente (>1 unidad logarítmica)

Rojas-Aguirre, Y.; Medina-Franco, J. L. Mol. Diversity, 2014, 18, 599

Identificación de *Activity Cliffs*; para qué?

- Alto contenido de información de SAR:
 - Elementos estructurales que rigen la actividad biológica
- Anticipar problemas de modelos QSAR
 - Los *activity cliffs* reducen la capacidad predictiva de modelos QSAR



Identificación Sistemática de *Activity Cliffs*

- Análisis visual
 - Series pequeñas de moléculas (<20 compuestos)
- Panoramas de actividad
 - Series medianas o grandes de compuestos



Métodos para Detectar Activity Cliffs

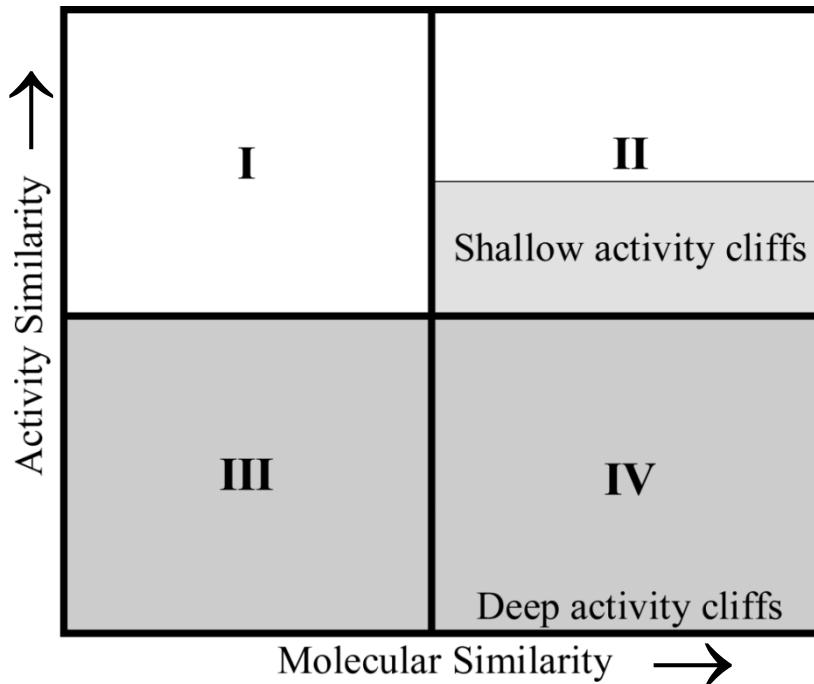
Approach	Basis	Applications	Refs.
Structure–activity relationship index (SARI)	Combine normalized continuous and discontinuous (for compounds with structure similarity and potency difference above a predefined cut-off) scores in a set of compounds or discontinuity of a compound	Global and local SAR characteristics and detect activity cliffs	[29, 30]
Structure–activity landscape index (SALI)	Potency-guided connection of nodes (individual compounds) if they form an activity cliff greater than a predefined SALI score	Detect activity cliffs	[13, 14]
Structure–activity similarity (SAS) maps	Depict pairwise structure and activity similarities. Activity similarities are represented as normalized potency difference. Data points denote compound pairs	Detect activity cliffs, continuous and discontinuous local SARs. Consensus activity cliffs for multiple structural representations	[23, 27, 31, 32]
Network-like similarity graphs (NSG)	Similarity–potency relationships. Connect nodes (individual compounds) if their similarity values are higher than a predefined cut-off. Color gradient reflects potency. Nodes are sized based on contributions of the corresponding compounds to the local SAR discontinuity	Explore global and local SAR patterns. Detect activity and selectivity cliffs.	[30, 33]
Multidimensional scaling (MDS)	Project chemical reference space onto a 2D surface. Color gradient reflects potency levels	Detect activity cliffs, continuous and discontinuous local SARs	[37]
Dual Activity Difference (DAD) maps	Pairwise comparison of the activity differences of a set of compounds across two targets. Each data point represents a pair of compounds, colored according to defined zones in the plots or structural similarity	Investigate activity, selectivity profile of a set of compounds against two targets, including selectivity cliffs	[27]
Combinatorial analog graphs (CAGs)	Utilize local similarity based on R-group substitution patterns to organize analog series in a hierarchical tree. Nodes represent compound subsets with substitutions at specified sites, while edges link subsets with one or more substitutions sites in common. Nodes are also annotated with normalized discontinuity scores	Study local SARs of closely related analogs of a single compound in advanced stages of drug discovery	[34]
Similarity-potency trees (SPT)	Hierarchical representation of compounds in a data set with increasing dissimilarity from a single reference compound down to a similarity threshold. Color gradient reflects potency but without node scaling	Describe local SARs	[38, 39]
3D activity landscape	Compound potency included as a third dimension to a 2D projection of chemical space representation	Detect activity cliffs continuous and discontinuous local SARs	[36]

Medina-Franco JL, Yonge AB, López-Vallejo F. En *Statistical Modeling of Molecular Descriptors in QSAR/QSPR*, Dehmer, M; Varmuza, K. Danail Bonchev (Eds.), Wiley-VCH, 2012, pp. 309

Structure-Activity Similarity (SAS) maps

$$AS_{a,b}(T) = 1 - \frac{|A_a - A_b|}{\max - \min}$$

$$\Delta pKi_{a,b}(T) = |pKi_a(T) - pKi_b(T)|$$

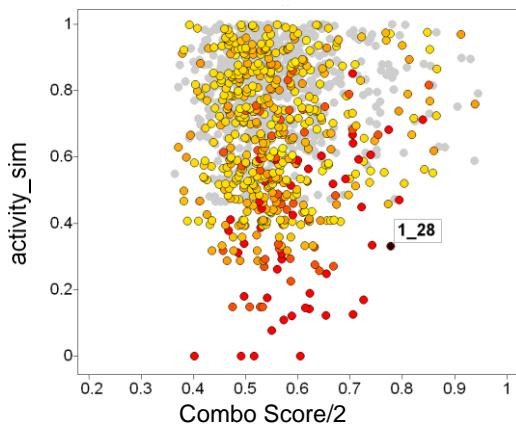
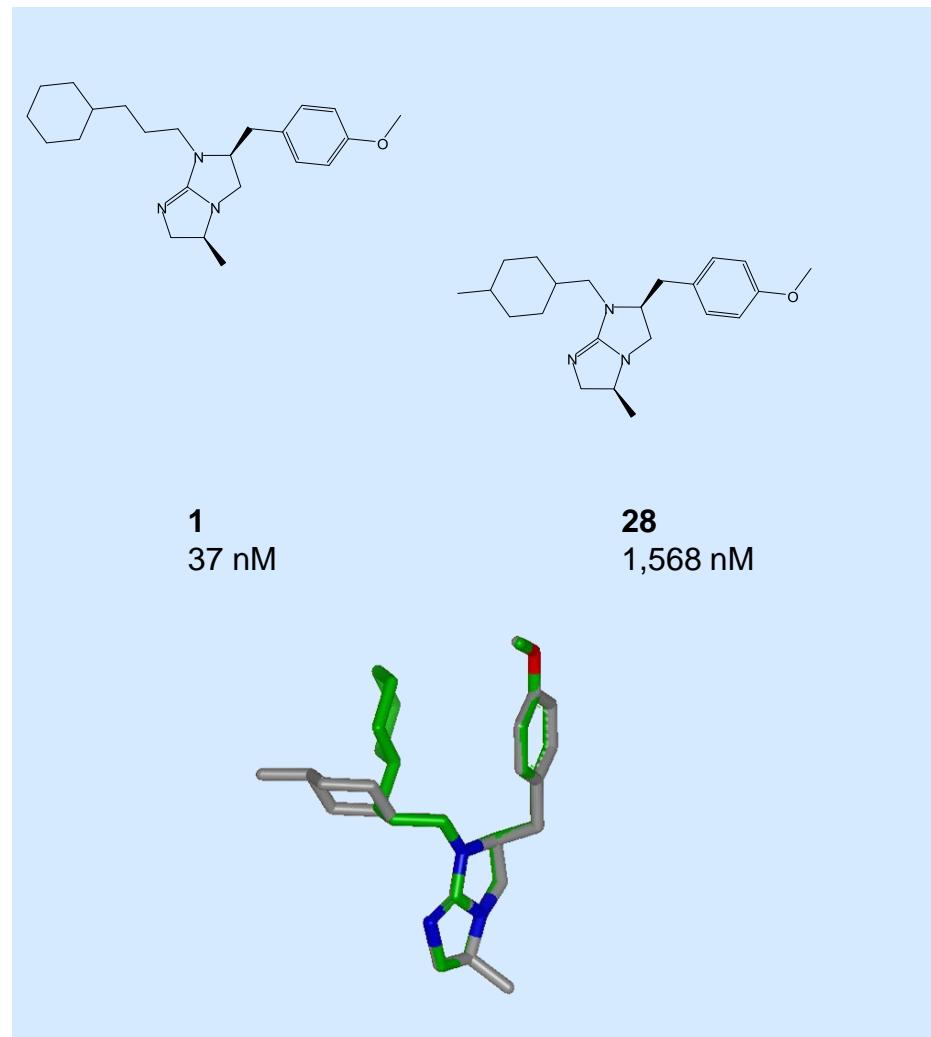
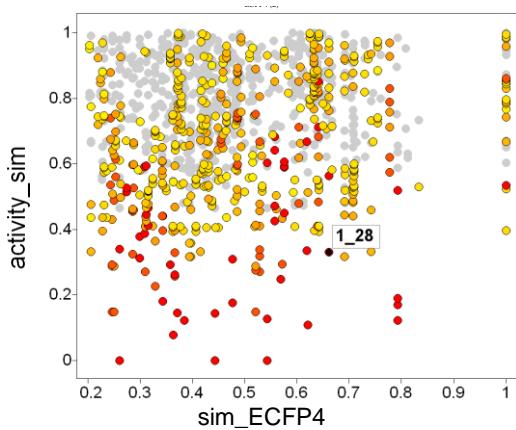


$$Tn_{a,b} = \frac{m}{i + j - m}$$

$$ED_{a,b} = \left\{ \sum_{k=1}^K (f_{ka} - f_{kb})^2 \right\}^{\frac{1}{2}}$$

Shanmugasundaram V. and Maggiora, GM. 222nd ACS National Meeting, Chicago, IL, USA, 2001

Activity cliff en SAS Map



2D similarity

MACCS: 0.98

GpiDAPH3: 0.84

ECFC4: 0.75

ECFP4: 0.66

Activity similarity: 0.33

(Delta pIC_{50} : 1.63)

3D similarity

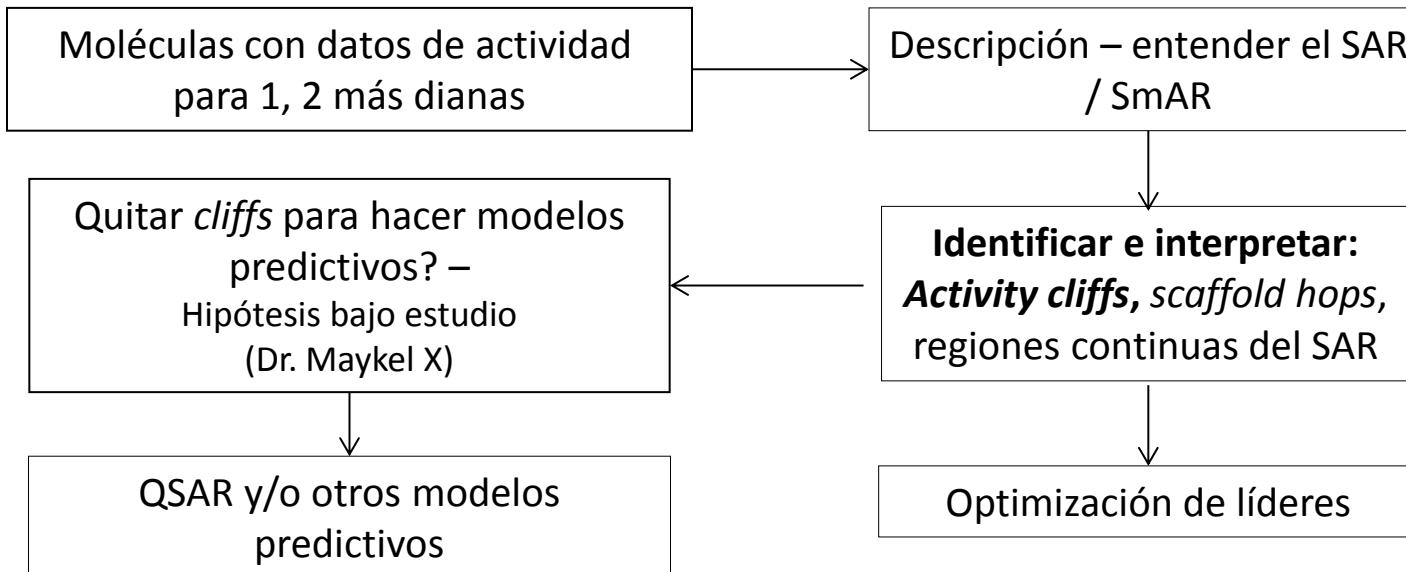
Comboscore/2: 0.78

piDAPH4: 0.69

Medina-Franco, J. L.; Martínez-Mayorga, K. et al. *J. Chem. Inf. Model.* 2009, 49, 477

Diagrama de Flujo para Analizar SAR de Compuestos

Análisis *antes de hacer estudios QSAR*



Cruz-Monteagudo, M.; Medina-Franco, J.L. et al. *Drug Discovery Today* 2014, 19, 1069





Foto:
Cortesía de CP. María Teresa Medina

SIMILITUD MOLECULAR



Bases del Concepto de *Similitud*

- Principio de similitud molecular: “Moléculas semejantes, tienen propiedades semejantes”
- Similitud es opuesto a distancia
- Similitud (*similarity*) es un concepto subjetivo
 - Depende del método para codificar la información química relevante
 - Depende del espacio químico
- Similitud estructural o molecular
 - Generalmente codifican información de la estructura o sub-estructura
- Similitud química
 - Generalmente codifica información de propiedades químicas macroscópicas (solubilidad, logP) o también energías calculadas (HOMO, LUMO, etc).

Johnson MA. and Maggiora GM., eds (1990) *Concepts and Applications of Molecular Similarity*, John Wiley

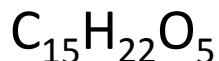
Bases del Concepto de *Similitud*

- La similitud química/estructural/molecular es una relación de pares
- Valores de similitud están dados por medidas de similitud (*similarity measures*). Componentes esenciales:
 - Representación estructural o molecular
 - Peso individual a cada elemento estructural o químico relevante
 - Función de similitud o distancia: transforma (o mapea) la información de las estructuras en una escala de valores reales en el intervalo [0,1]
- El cálculo de similitud se hace en un espacio multi-dimensional
- La visualización de la similitud se hace (en forma *aproximada*) en 2D/3D

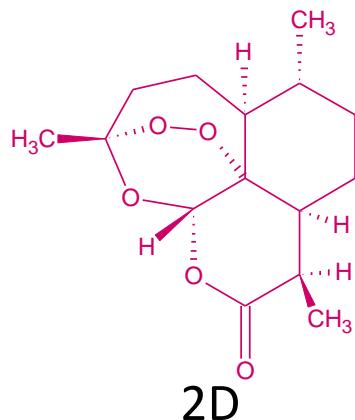


Descripción Cuantitativa de Moléculas

Artemisina



1D



2D

Descriptores

peso molecular

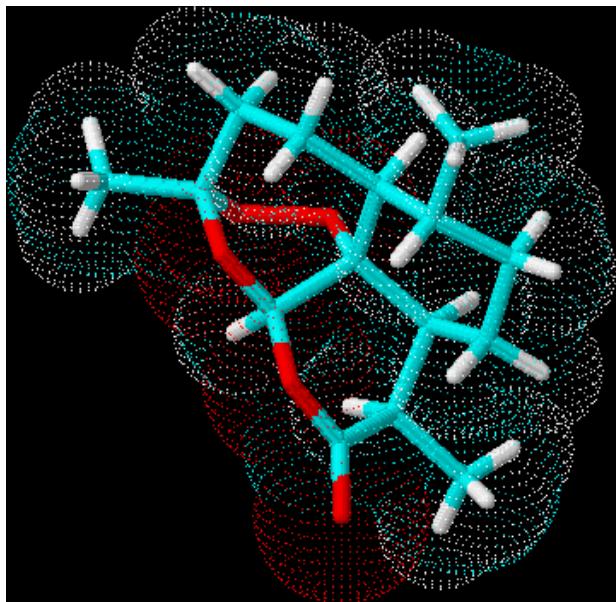
conectividad; número
de anillos; coeficiente
de partición (logP), huellas
moleculares digitales, etc.



Sabias
que ...

*Las huellas digitales moleculares (*molecular fingerprints*) son de las representaciones más empleadas en el cálculo de similitud molecular de bases de datos moleculares para el desarrollo de fármacos

Descripción Cuantitativa de Moléculas



3D

Descriptores

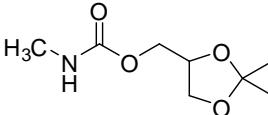
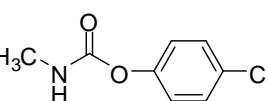
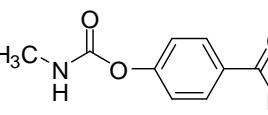
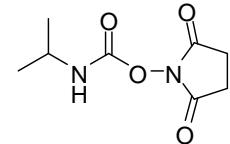
volumen; superficie
accesible al disolvente;
energía; HOMO; LUMO;
etc.



Representación de Moléculas con Huellas Digitales Bidimensionales

Número *p* de fragmentos estructurales bidimensionales (2D)

El número *p*
depende del
método

Molécula	NC(O)O	O heterociclo	CH ₃ -Q	N-O	CN
1 	1	1	1	0	0
2 	1	0	1	0	1
3 	1	0	1	0	0
...	...				
<i>n</i> 	1	0	0	1	0

Huella digital
Vector binario
Bit vector

Las estructuras se representan en 2D

- Cálculos muy rápidos
- Método apropiado para bibliotecas grandes



Tipos de “Huellas Digitales Moleculares”

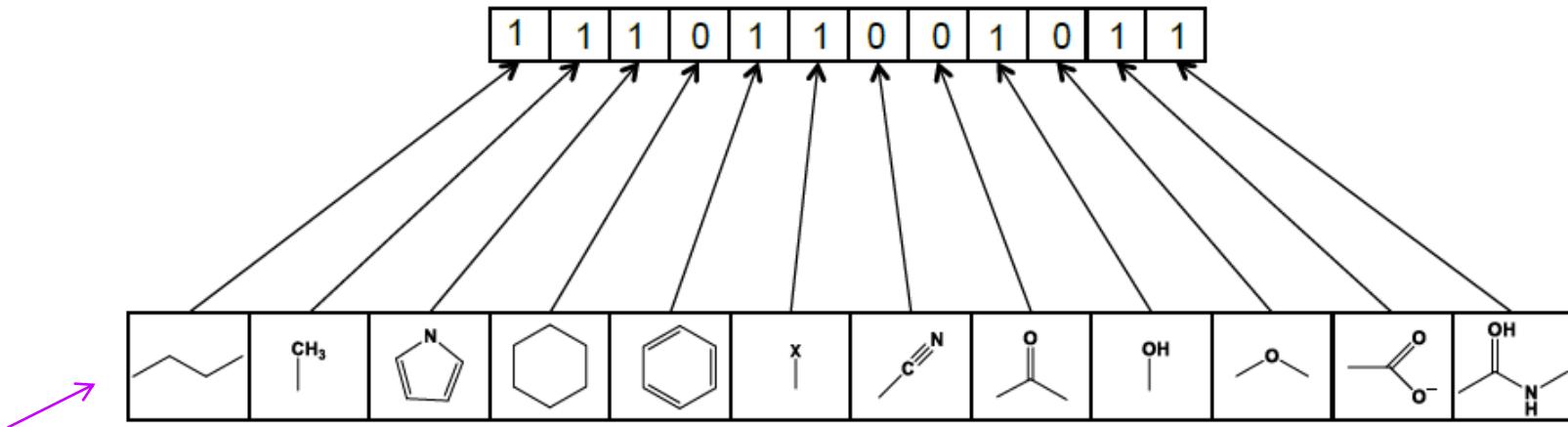
Molecular fingerprints

- 3D
 - 2D, Tipos principales
 - Basados en diccionario
 - » *substructure or dictionary based*
 - Topológicos
 - » *topological or path-based*
 - Circulares
 - » *circular*
- Número *p* de elementos en el vector
- Independientes de la molécula
- Dependientes de la molécula



Basados en Diccionario

Substructure or dictionary based



- Diccionario de características preestablecida
- Tamaño de vector fijo

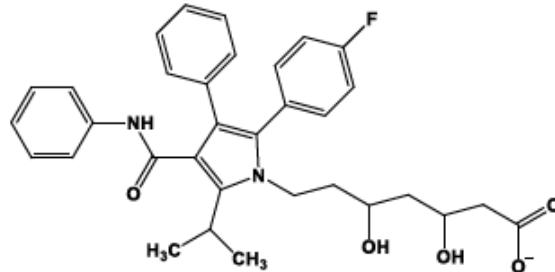


Figure 1. An example, based on the drug Lipitor, of a simplified *molecule-independent* directory-based binary structural fingerprint with its corresponding set of descriptors. The symbol 'X' corresponds to any of the halogen atoms (F, Cl, Br, I)



Basados en Diccionario

Substructure or dictionary based

- **MACCS keys:** one with 960 and the other with 166 structural keys based on SMARTS patterns
- **PubChem fingerprint:** 881 structural keys
- **BCI fingerprints:** BCI fingerprints can be generated using different numbers of bits and can be modified by the user in several ways, but the standard substructure dictionary includes 1052 keys
- **TGD and TGT fingerprints:** These are two-point and Three-point pharmacophoric fingerprints calculated from a 2D molecular graph, consisting, respectively of 735 and 13,824 bits. TGD encodes atom-pair descriptors using seven-atom features and distances up to 15 bonds. TGT encodes triplets of four-atom features using three graph distances divided into six distance ranges. They are both available in MOE software package.



Basados en Diccionario

Substructure or dictionary based

- Ventaja
 - Se puede asociar un bit con un fragmento estructural
- Desventaja
 - Moléculas diferentes pueden tener el mismo vector binario (representación idéntica)
 - Incluir número de fragmentos disminuye esta probabilidad (incluir contadores de fragmentos)
 - Se puede perder mucha información (dependiendo del diccionario)



Circulares, Radiales

- Extended-connectivity FPs (ECFPs) developed by Rogers and Hahn
- Represent circular atom neighborhoods and produce fingerprints of variable length. They are most commonly used with a diameter of 4 (ECFP4). ECFP6 are also commonly used
- The local “circular” environments surrounding each non-hydrogen atom constitute the substructural features of a given molecule

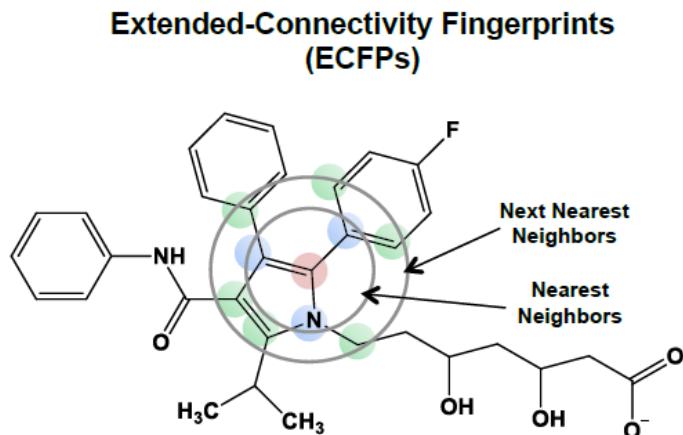


Figure 3. Examples of *molecule-dependent* ECFPs descriptors depicted with respect to the drug Lipitor.

Programas para Cálculo de *Fingerprints*

Ejemplos

- OEChem TK: This OpenEye toolkit [21] is able to produce 166-bit MACCS, LINGO, Circular, Path (Daylight-like) and Tree (Daylight-like with non-linear, “tree” fragments) fingerprints. It has interfaces to C++, Java, Python, and C#.
- JChem from ChemAxon [30]: This is a java library that provides access to several hashed fingerprints, ECFP fingerprints with all their variants (ECFC, FCFP, FCFC), and pharmacophoric fingerprints. ChemAxon also provides packages for .NET and is usable in Python through cinfo [40].
- Open Babel [24,41]: This is a free and open-source cheminformatics toolkit, which implements MOLPRINT2D, 166-bit MACCS, a Daylight-like fingerprint (FP2), and 2 structural key fingerprints with 55 (FP3) and 307 bits. It can be used from C++, Java, Python, C#, and Perl.
- RDKit [31]: This is also a free and open-source cheminformatics toolkit that provides access to several fingerprints: 166-bit MACCS, “Topological” (Daylight-like), “Atom pairs” (based on the atomic environments and shortest path separations of every atom pair in the molecule [42]), “Morgan” (ECFP and its variations), “Torsion” (based on the topological torsion descriptor [43]), and “Layered” (an experimental topological fingerprint intended to make fingerprinting queries more straightforward). It is usable from C++, Python, Java, and C#.

Cereto-Massagué A, Ojeda MJ, Valls C, Mulero M, Garcia-Vallvé S, et al. (2015) *Methods* 71:58

Función de Similitud

Coeficiente de Tanimoto

Medida del número de fragmentos comunes entre dos estructuras

$$S_{AB} = \frac{c}{a + b - c}$$

S_{AB} = similitud entre estructura A y B

c = fragmentos comunes entre A y B (“1” bits comunes)

a = fragmentos en A (“1” bits en A)

b = fragmentos en B (“1” bits en B)

Ejemplo:

f₁ *f₂* *f₃* *f₄* *f₅*



2 fragmentos en común

A		1	1	1	0	0
B		1	0	1	0	1

↓

$$S_{AB} = \frac{2}{3 + 3 - 2} = 0.5$$

Valores posibles de coeficiente de Tanimoto: $0 \leq S_{AB} \leq 1$



Funciones de Similitud

Ejemplos

Table 1

Some similarity coefficients and distances used with fingerprints.

Measure	Expression	Range
Tanimoto/Jaccard coefficient	$\frac{c}{a+b-c}$	0 to 1
Euclidean distance	$\sqrt{a + b - 2c}$	0 to N
City-block/Manhattan/Hamming distance	$a + b - 2c$	0 to N
Dice coefficient	$\frac{2c}{a+b}$	0 to 1
Cosine similarity	$\frac{c}{\sqrt{ab}}$	0 to 1
Russell-RAO coefficient	$\frac{c}{m}$	0 to 1
Forbes coefficient	$\frac{cm}{ab}$	0 to 1
Soergel distance	$\frac{a+b-2c}{a+b-c}$	0 to 1

Where, given the fingerprints of two compounds, A and B, m equals the total amount of bits present in the fingerprints, a equals the amount of bit set to 1 in A, b equals the amount of bits set to 1 in B and c equals the amount of bits set to 1 in both A and B.

FARMACÓFORO

BÚSQUEDA Y DISEÑO



Definición de Farmacóforo

Farmacóforo: “... **ensamble molecular** que contiene (*phoros*) las **características esenciales** para la actividad biológica de un fármaco (*pharmacon*).”

Erlich, P. *Dtsch. Chem. Ges.* 1909, 42, 17.

The ensemble of steric and electronic features that is necessary to ensure optimal interactions with a specific biological target structure and to trigger (or to block) its biological response



Búsqueda del Modelo de Farmacóforo

Etapas

1. Determine identity of a “lead compound”:

- Screen natural and synthetic banks of compounds for activity
- Folk medicine
- Natural ligand
- Drug already known
- Computer-aided drug design
- Computerized search of structural databases

2. Data collection: Publications; patents; biological activity; NMR and X-ray data; physiochemical properties

- Determine the effects of structural changes on activity of drug: structure-activity relationships (SARs)

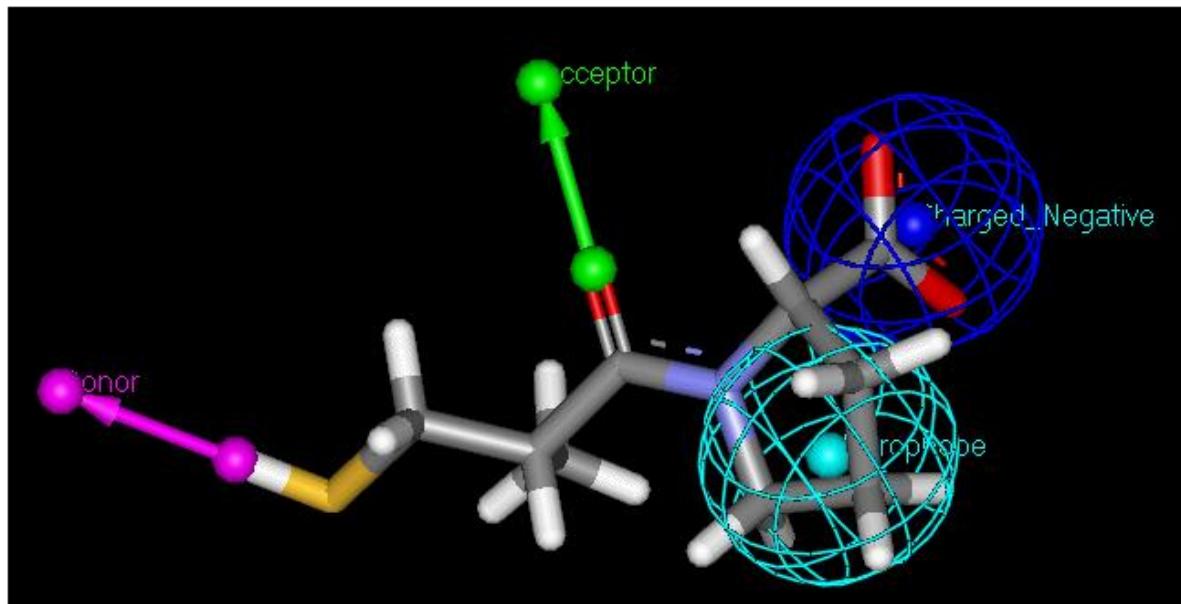


Búsqueda del Modelo de Farmacóforo

Etapas

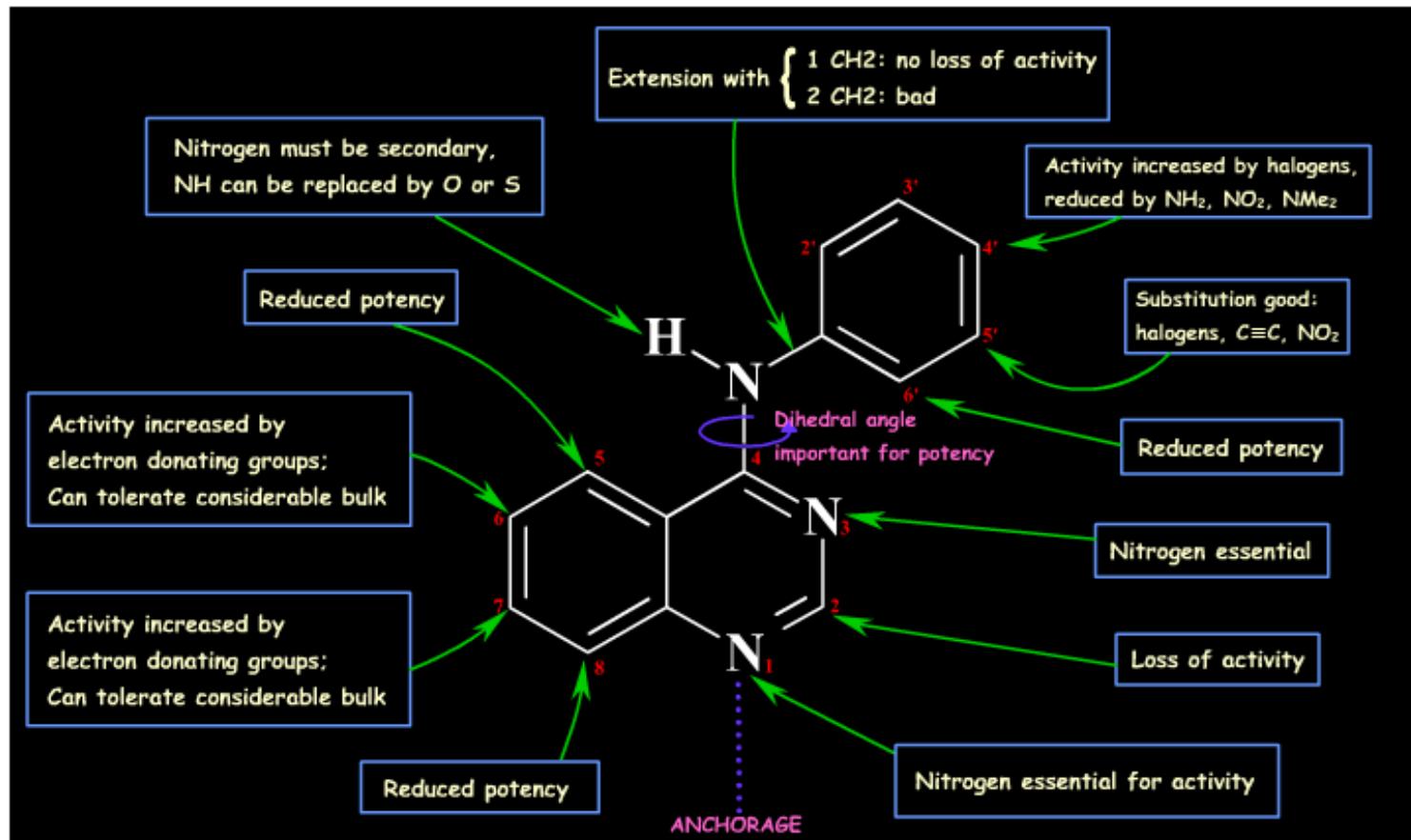
3. Analysis: integrate information about drug (and target) to generate hypothesis about activity

- This information will result in the identification of a pharmacophore...
- What functional groups are important to biological activity?



Análisis de Datos: Relaciones Estructura-Actividad (SAR)

Ejemplo

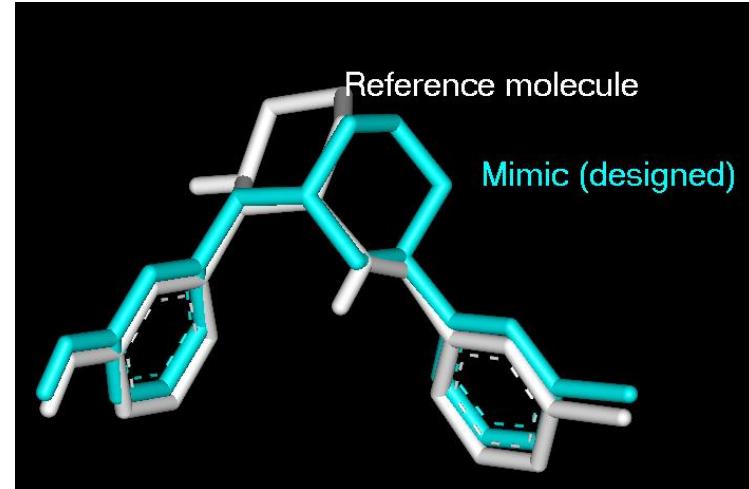


Búsqueda del Modelo de Farmacóforo

Etapas

- *Design new structures:* If you know the pharmacophore for your target, you can create new lead compounds based on the pharmacophore!
- Why make new lead compounds?
 - Increase activity (make binding stronger)
 - Decrease side effects (increase selectivity)
 - Improve ease and efficiency of administration to patient
 - Potentially find a better synthetic route

Approach: Molecular mimicry



Uso del Modelo de Farmacóforo

- *Design: use analyzed data to design new compounds - hopefully with better properties*
- *Four methods used to design better drugs:*
 1. *Chemical modification*
 2. *Database searching*
 3. *De novo*
 4. *Manual*
- *These approaches generate more data, which yet again can be used to generate new hypotheses and structures, etc.*

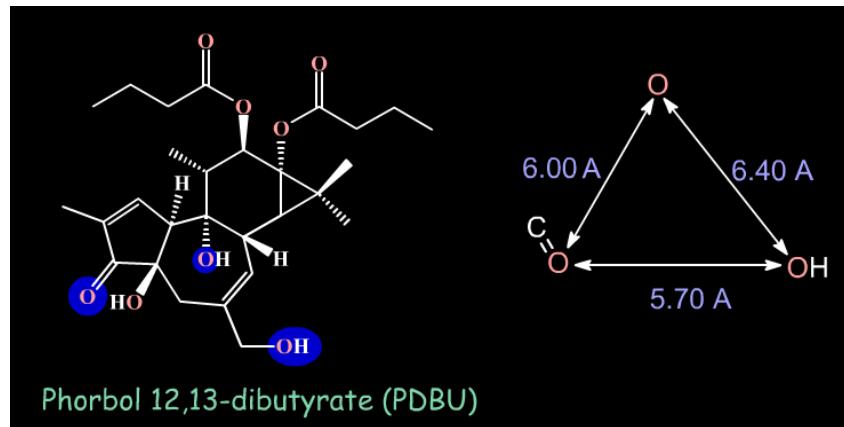


Búsqueda en Bases de Datos

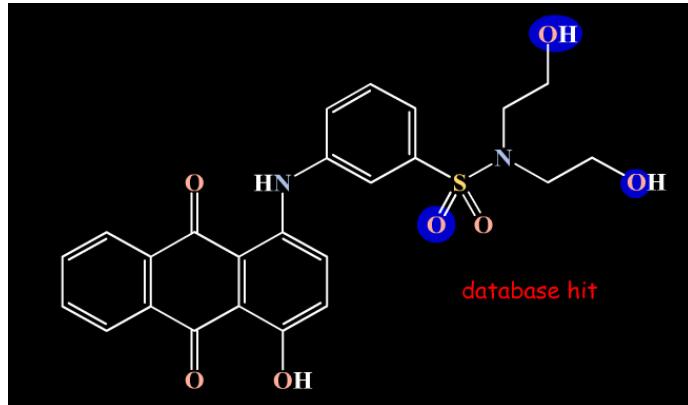
- Use databases of known compounds – no new synthesis!
- Be careful of multiple conformations
- Content of database is crucial

3D Search for a 3D pharmacophore

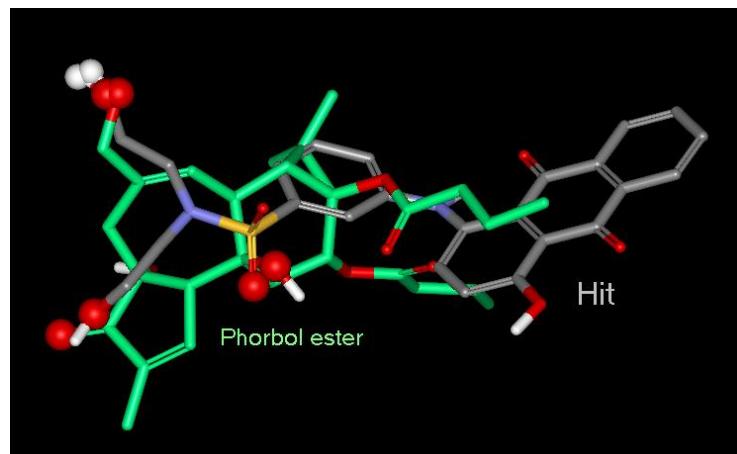
Example: Protein kinase C enzymes are targets for chemotherapeutic intervention against cancer. The pharmacophore was deduced from active phorbol esters like PDBU



Búsqueda en Bases de Datos



The 3D database search led to the discovery of a new potent protein kinase C inhibitor that is chemically very different from the original reference phorbol esters. Alignment of the two:



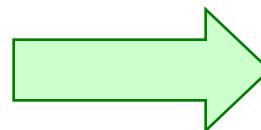
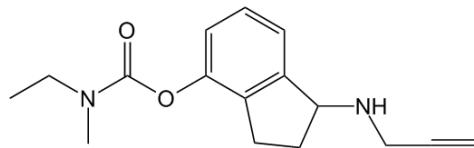
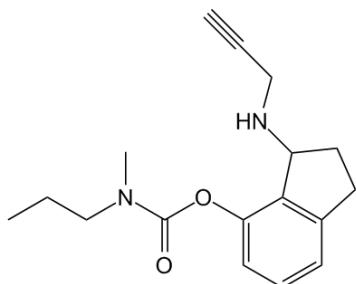
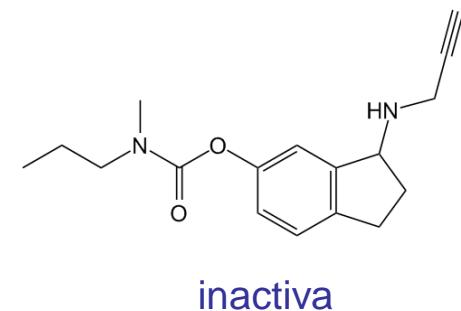
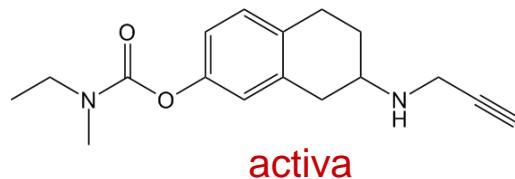
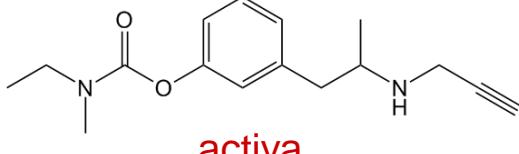
Start over with this “hit” as a new lead; chemical modification, etc...



Detección por Métodos Computacionales

Modelado del Farmacóforo

Inhibidores de la MAO

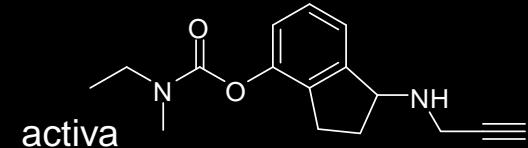
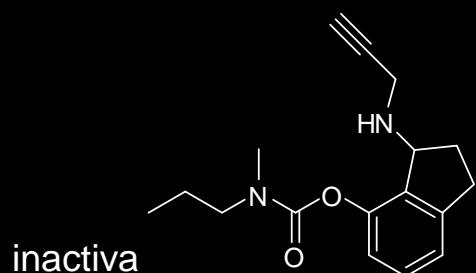
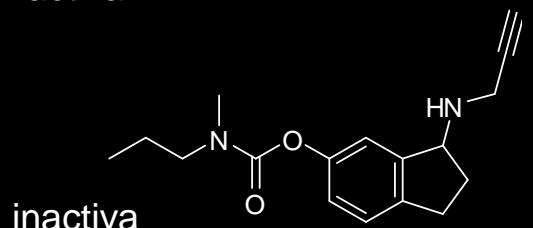
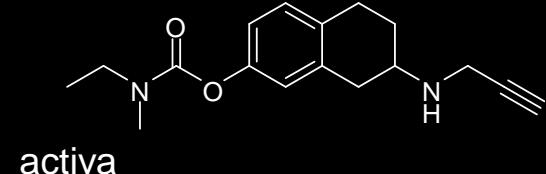
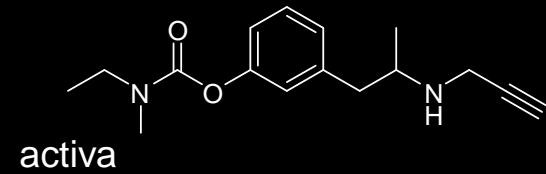


¿Cuál es el
farmacóforo?

Sterling J, Herzig Y, Goren T, Finkelstein, N. et. al., *J. Med. Chem.* 2002, 45, 5260



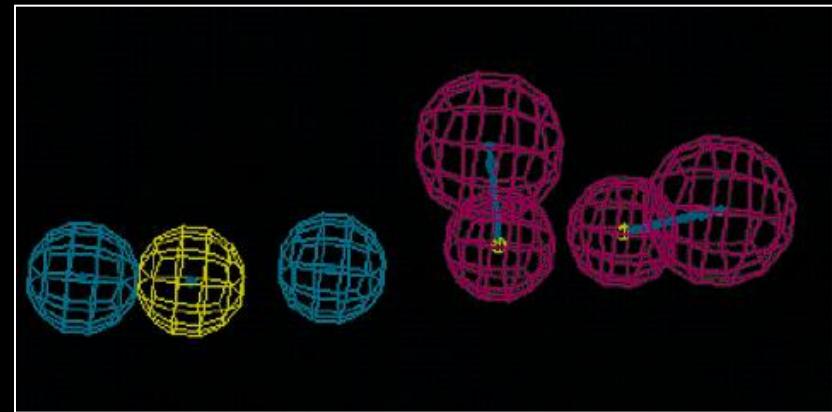
Modelo del Farmacóforo



Hidrófobo

Carbamato

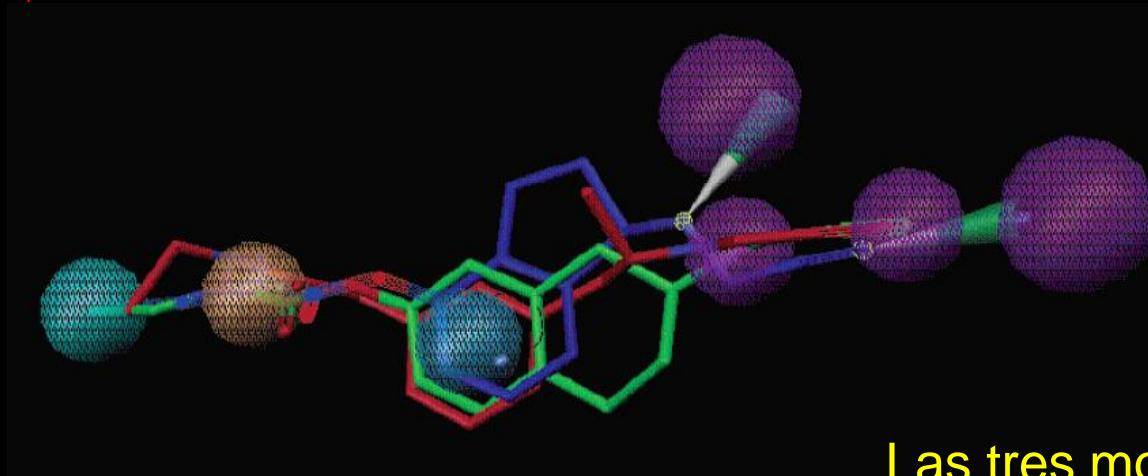
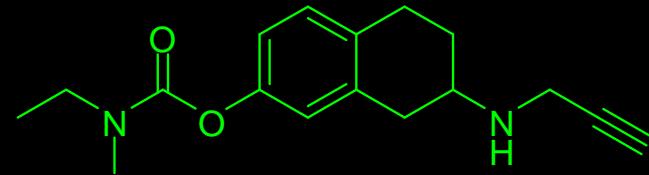
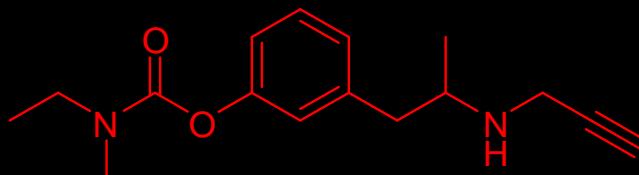
Puente de hidrógeno
(donador o aceptor)



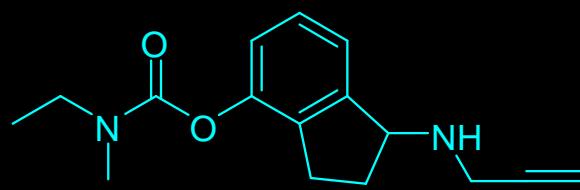
Modelo del farmacóforo
Ensamble molecular de
características esenciales

Sterling J. et. al., *J. Med. Chem.* 2002, 45, 5260

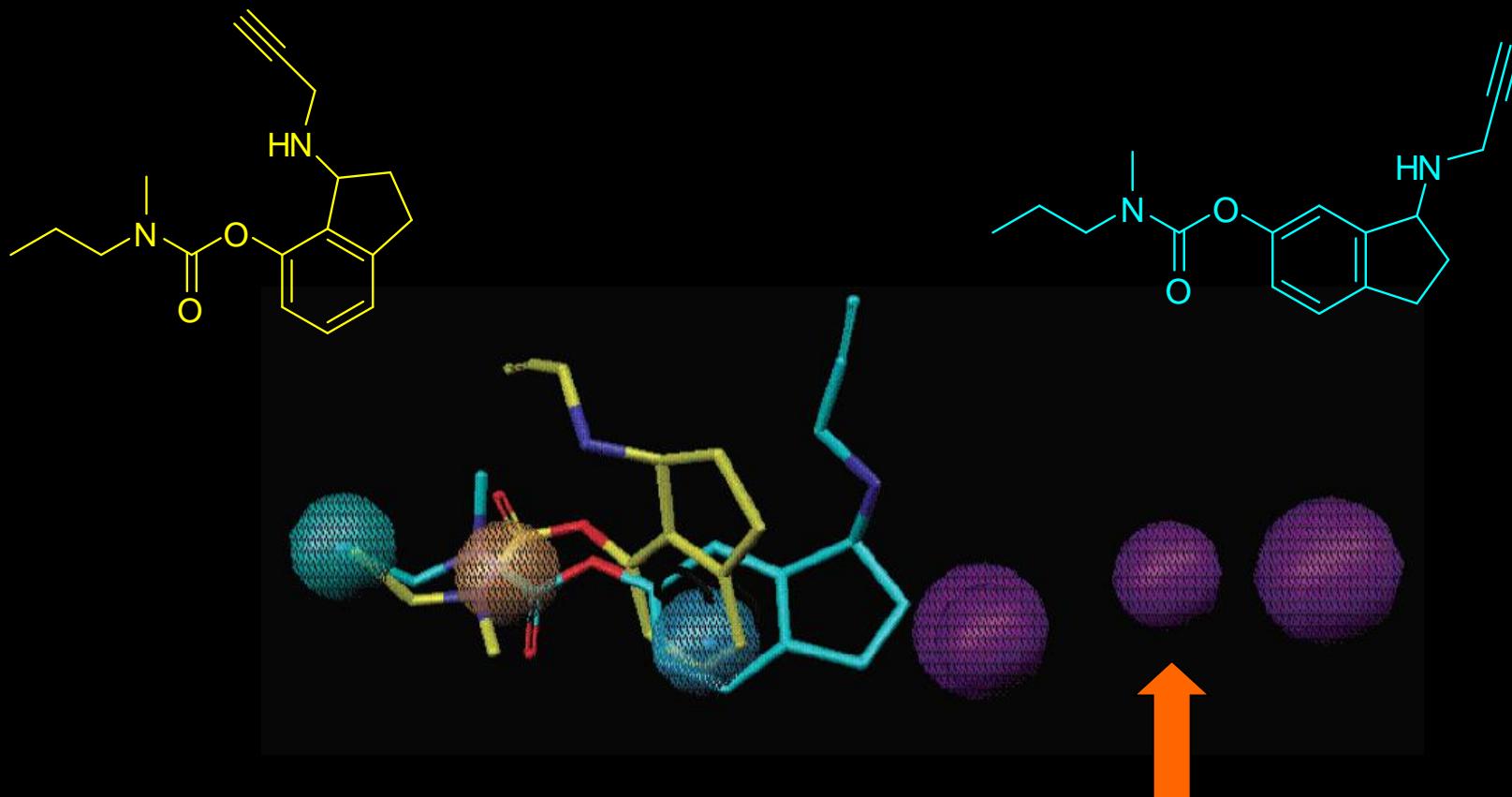
Compuestos Activos



Las tres moléculas coinciden
con el modelo del farmacóforo



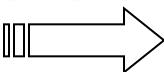
Compuestos Inactivos



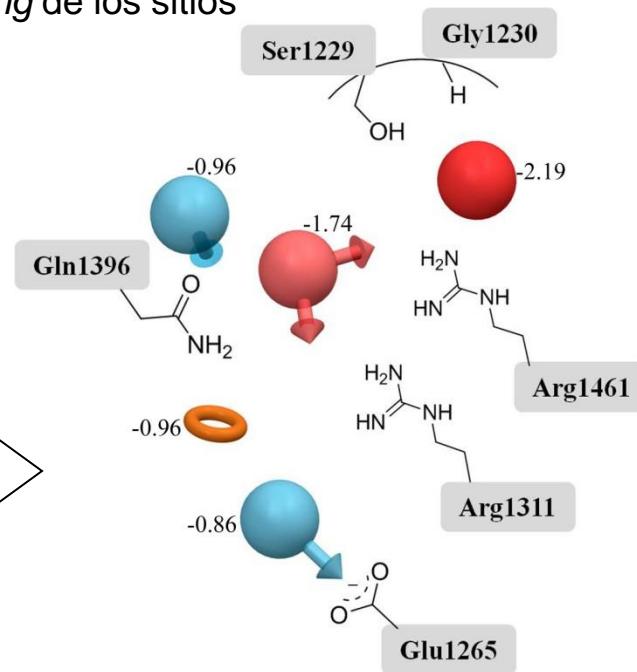
No coinciden con el farmacóforo

Modelo del Farmacóforo de Inhibidores de DNA Metiltransferasas

Modelos de unión de 14 inhibidores

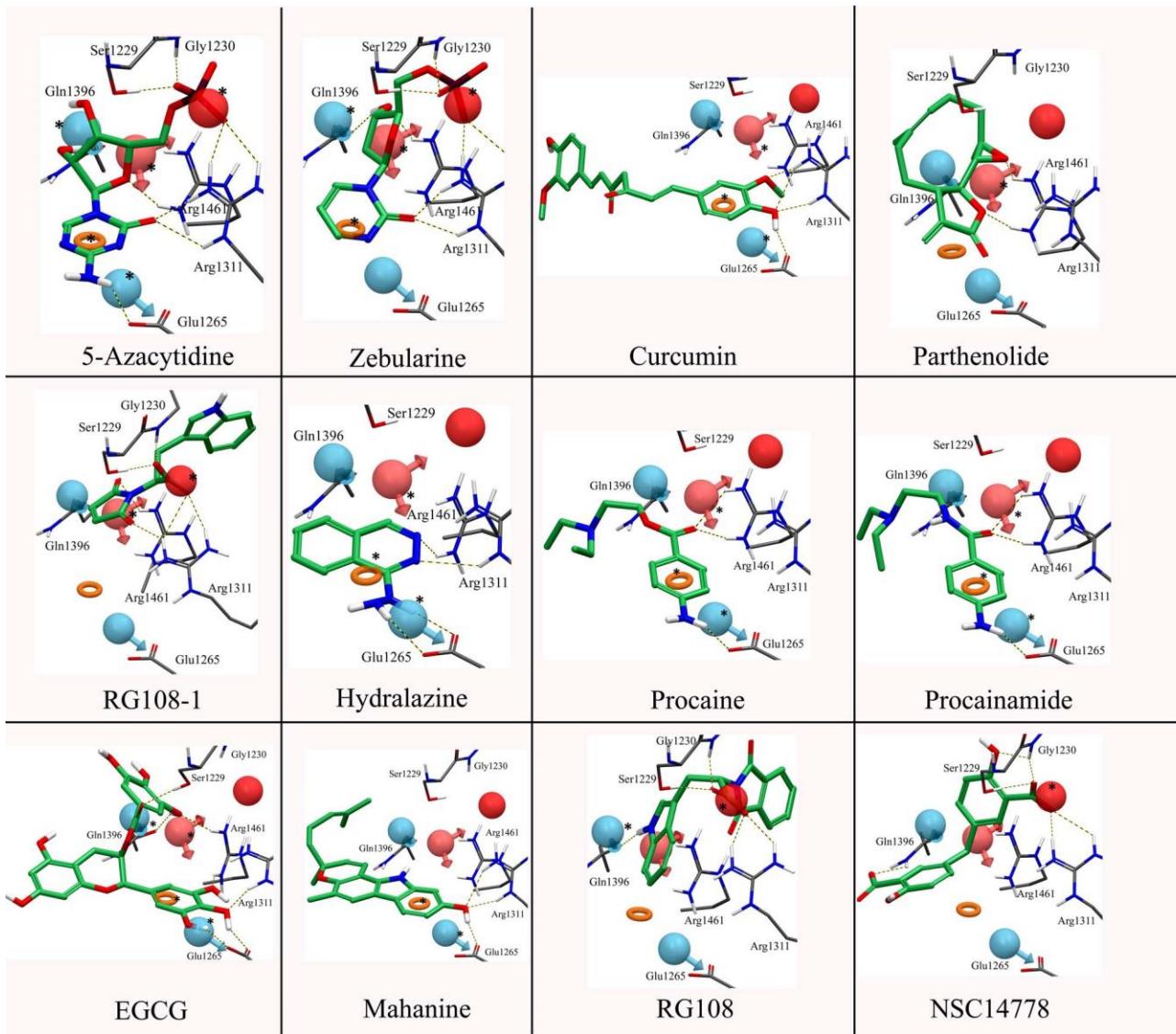


1. Mapeo de interacciones de *docking* en los átomos
2. Generar sitios farmacofóricos
3. Sumar las energías de los átomos correspondientes a los sitios farmacofóricos
4. *Ranking* de los sitios



Yoo J, Medina-Franco JL. *J. Comp.-Aided Mol. Des.*
2011, 25, 555

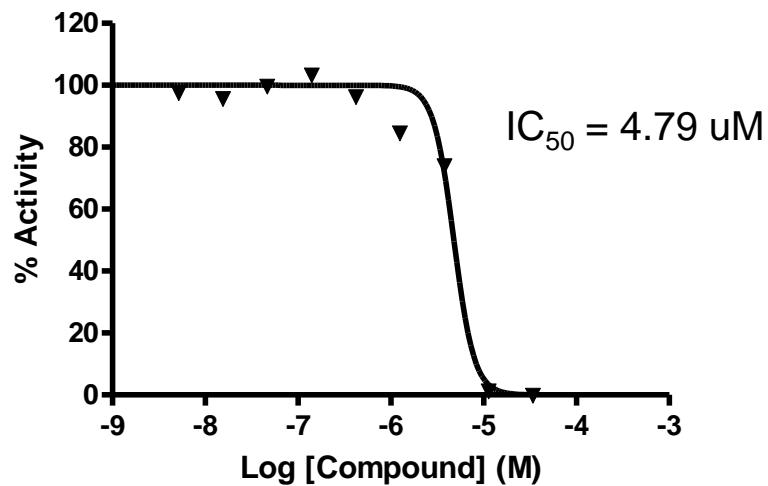
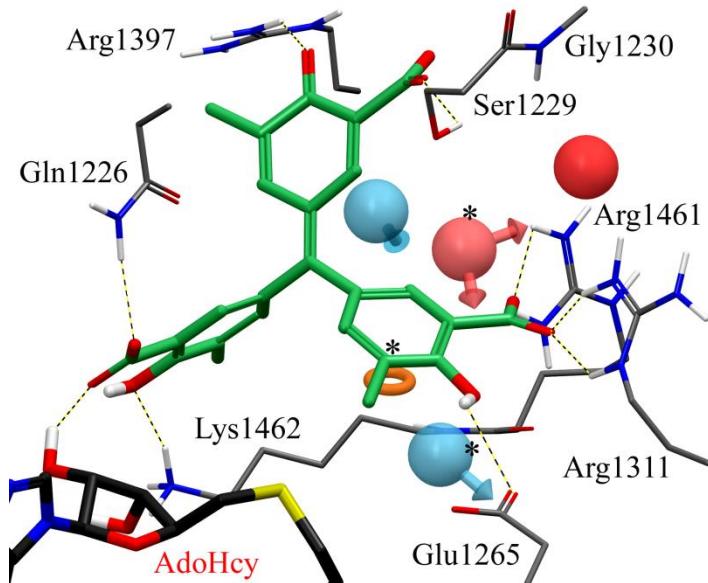




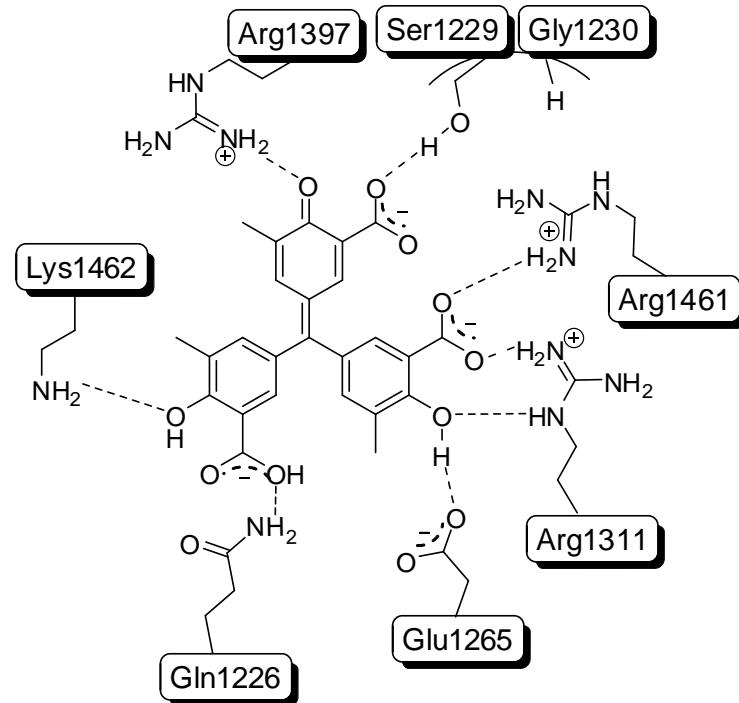
La mayoría de los inhibidores coinciden con 3 o más sitios farmacofóricos



Modelo del Farmacóforo y MeATA

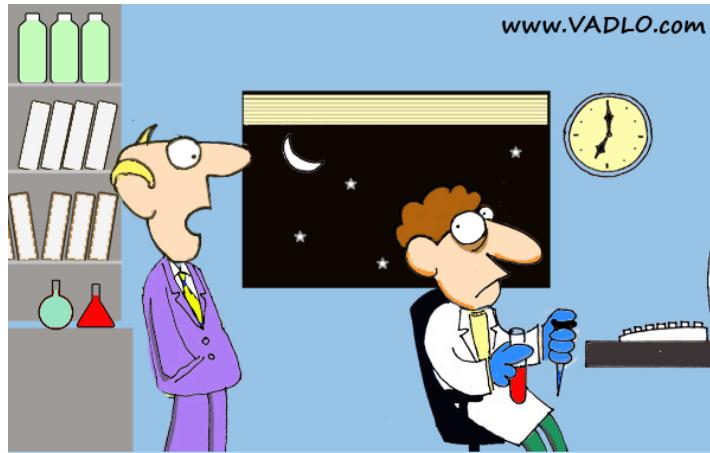


El ácido trimetil aurintricarboxílico es inhibidor de DNMT1



Yoo J, Medina-Franco JL. *J. Mol. Model.* 2012, 18, 1583-1589





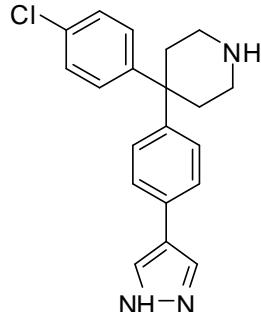
"Just work till midnight, you need to relax too"

CRIBADO VIRTUAL BASADO EN EL LIGANDO

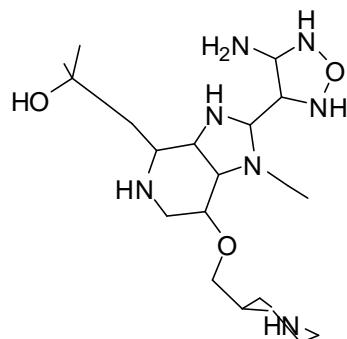


Screening Virtual de Inhibidores de AKT2

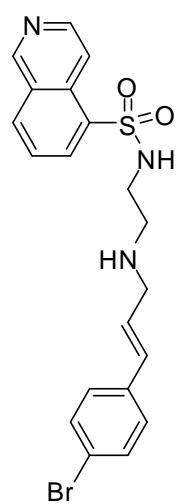
- Estructuras de rayos X
 - Inhibidores conocidos
- } → Screening basado en el ligando y la estructura



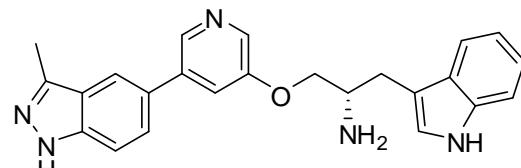
GVP (18 nM)



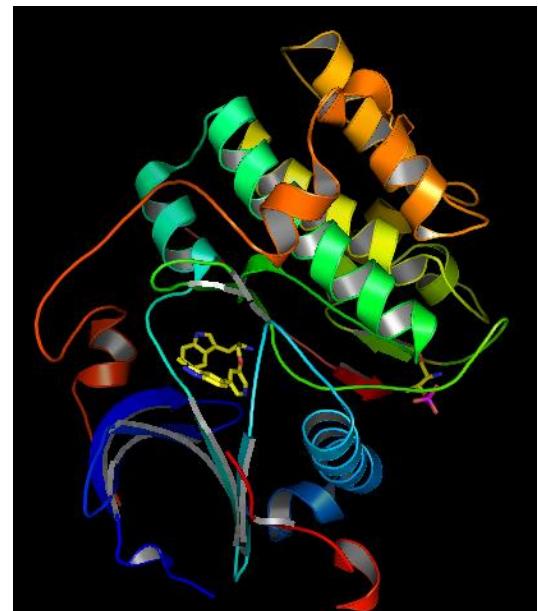
GSK690693 (13 nM)



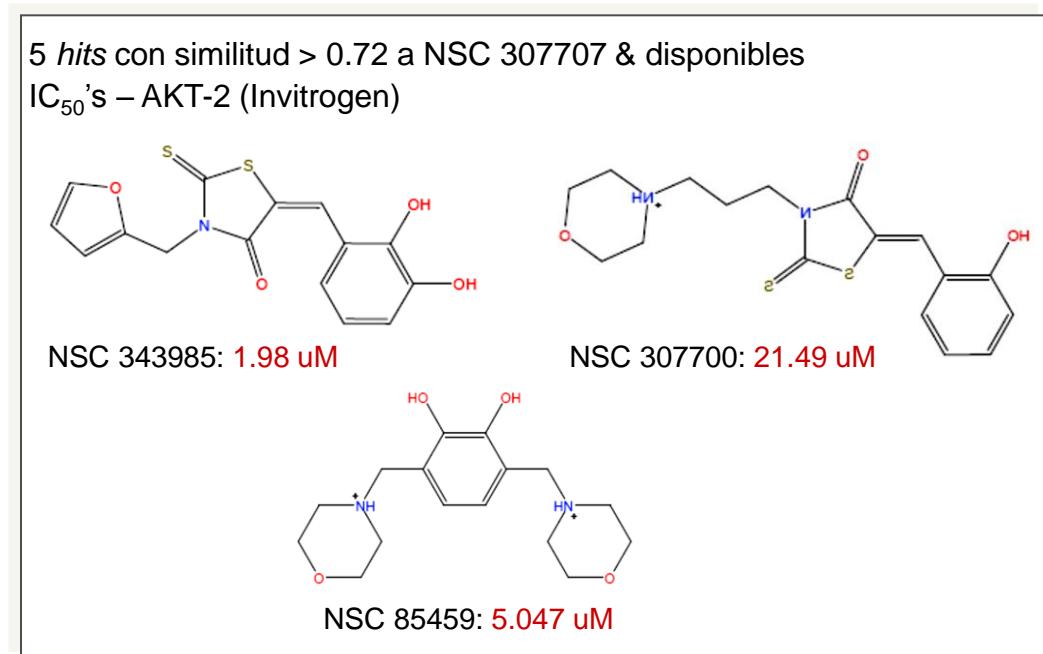
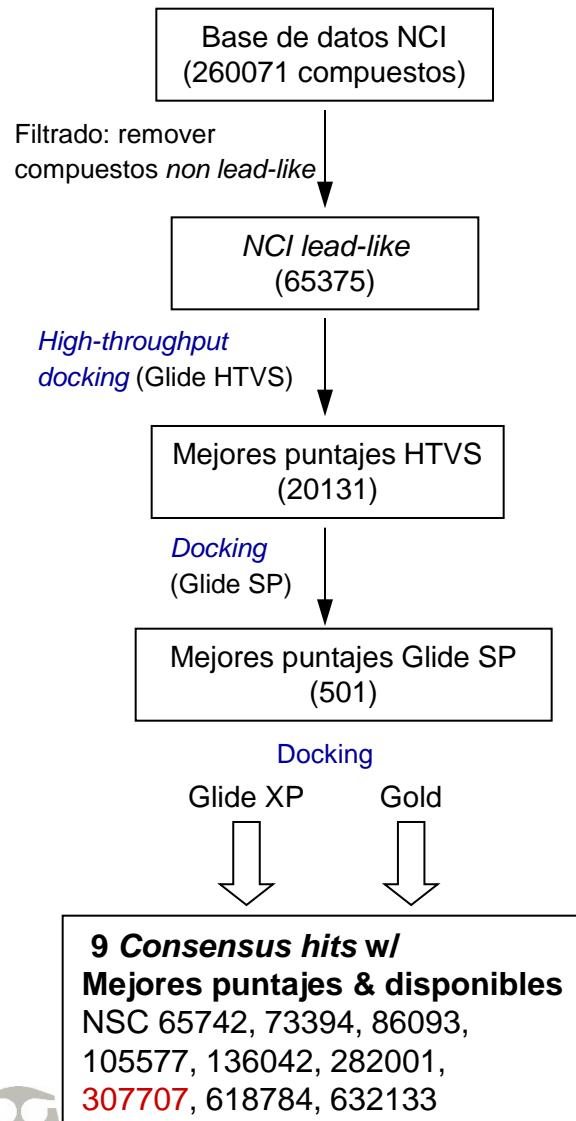
H-89 (0.3 μM)



A-443654 (0.5 nM)

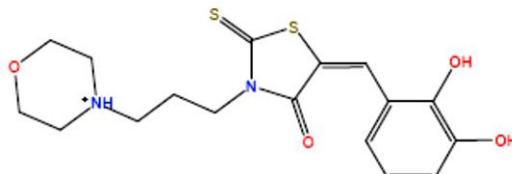


Screening de la Base de Datos NCI



Búsqueda en NCI lead-like
MACCS keys/Tmt > 0.72

Prueba experimental:
a 25 y 50 uM
AKT-2 (Invitrogen)



NSC 307707
IC₅₀ = 0.476 uM

Diseño de Compuestos Activos Basado en el Ligando

RESUMEN

