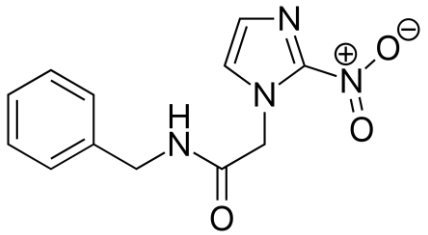
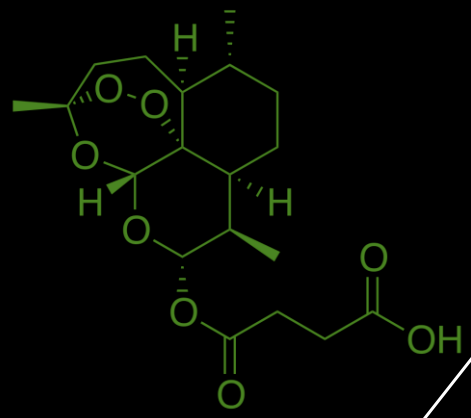


INTRODUCCIÓN A LA QUIMIOINFORMÁTICA

Fernán Agüero
Instituto de Investigaciones Biotecnológicas, UNSAM

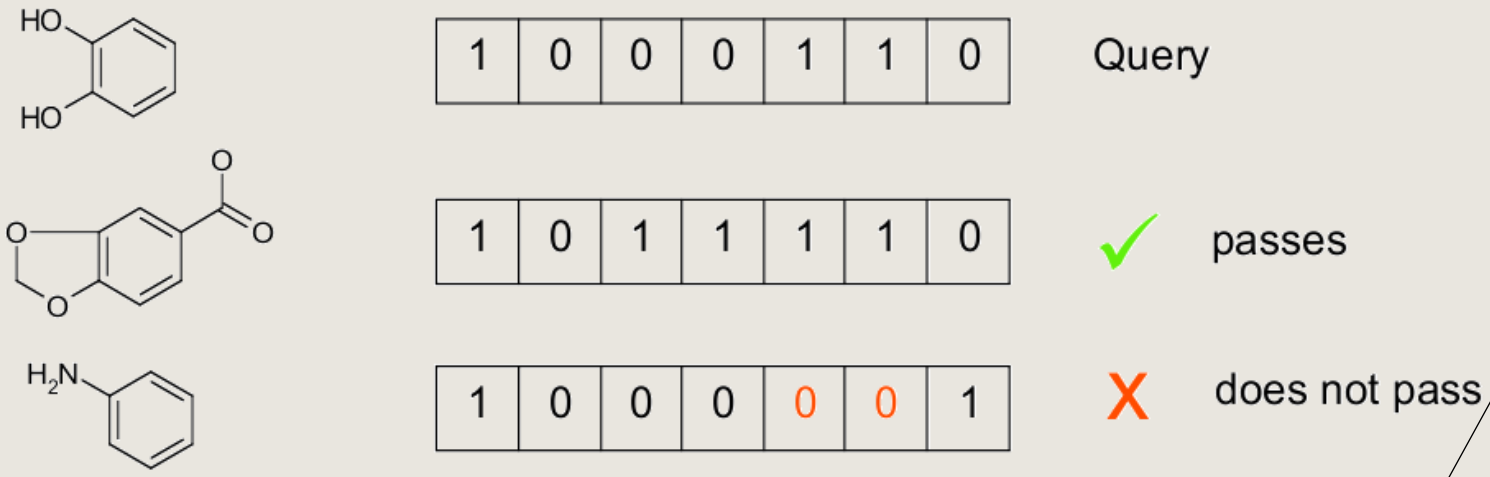
CHEMICAL FINGERPRINTS



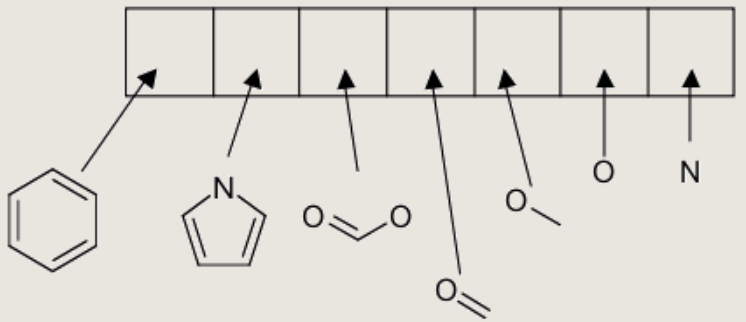
BÚSQUEDA DE SUBESTRUCTURAS: FINGERPRINTS

Fingerprint: representación abstracta de características o propiedades de una molécula (features)

- Presencia/ausencia de cada elemento
- Configuraciones electrónicas inusuales (carbono sp3, nitrógeno unido con un triple enlace)
- Anillos y sistemas de anillos (naftaleno, piridina, cyclohexano)
- Grupos funcionales (alcoholes, aminas, carboxilos, etc.)
- Se suelen utilizar tanto para búsquedas de subestructuras como para detectar similitud



Un fingerprint



BÚSQUEDA DE SUBESTRUCTURAS/SIMILITUD: FINGERPRINTS

Cuestiones a tener en cuenta

El fingerprint debe ser definido de antemano: bits, folding, count vectors...

Distintas aplicaciones pueden generar distintos tipos de fingerprints

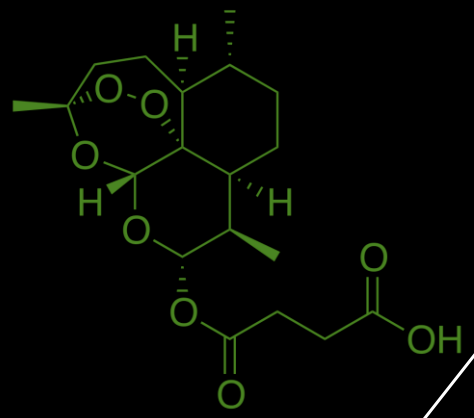
OpenBabel: https://openbabel.org/docs/dev/Features/Fingerprints.html

```
$ obabel -L fingerprints
ECFP0    Extended-Connectivity Fingerprints (ECFPs)
ECFP10    Extended-Connectivity Fingerprints (ECFPs)
ECFP2    Extended-Connectivity Fingerprints (ECFPs)
ECFP4    Extended-Connectivity Fingerprints (ECFPs)
ECFP6    Extended-Connectivity Fingerprints (ECFPs)
ECFP8    Extended-Connectivity Fingerprints (ECFPs)
FP2    Indexes linear fragments up to 7 atoms.
FP3    SMARTS patterns specified in the file patterns.txt
FP4    SMARTS patterns specified in the file SMARTS_InteLigand.txt
MACCS    SMARTS patterns specified in the file MACCS.txt
```

Daylight: https://www.daylight.com/dayhtml/doc/theory/theory.finger.html

RDKit: https://www.rdkit.org/docs/GettingStartedInPython.html#fingerprinting-and-molecular-similarity

O'Boyle NM, Banck M, James CA, Morley C, Vandermeersch T, Hutchison GR. Open Babel: An open chemical toolbox. J Cheminform. 2011 Oct 7;3:33. doi: 10.1186/1758-2946-3-33. PMID: 21982300; PMCID: PMC3198950.



DETOUR

Mirar patrones SMARTS en Github

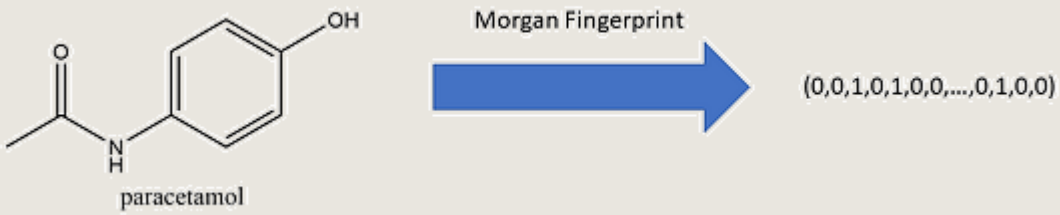
FP3
https://github.com/openbabel/openbabel/blob/master/data/patterns.txt
FP4
https://github.com/openbabel/openbabel/blob/master/data/SMARTS_InteLigand.txt
MACCS
https://github.com/openbabel/openbabel/blob/master/data/MACCS.txt

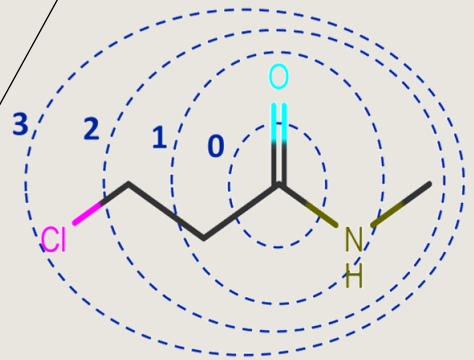
EXTENDED CONNECTIVITY FINGERPRINTS

Circular fingerprints

Concepto similar al de "***extended connectivity***" de Morgan

1. Assign each atom with an identifier
2. Update each atom's identifiers based on its neighbors
3. Remove duplicates
4. Fold list of identifiers into a 2048-bit vector (a Morgan fingerprint)





Extended Connectivity Circular Fingerprints

ECFP6 (radius = 3)

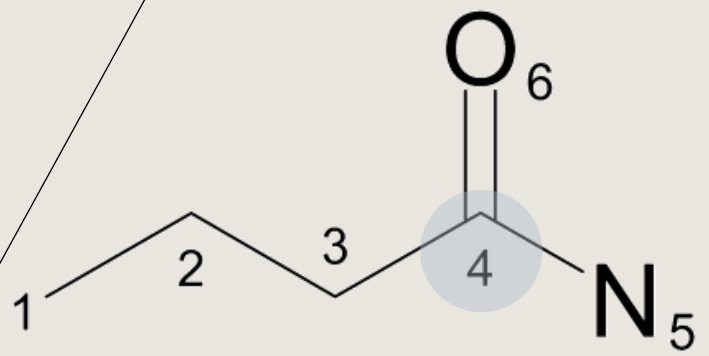
1024 or 2048 bits

EXTENDED CONNECTIVITY FINGERPRINTS

1. Assign each atom with an identifier

We choose an atom in the molecule (e.g. #4) and take note of:

- number of nearest-neighbour non-hydrogen atoms: **3**
- number of bonds attached to the atom (not including bonds to hydrogens): **4**
- atomic number: **6**
- atomic mass: **12**
- number of hydrogens connected to the atom: **0**
- is the atom in a ring (1) or not (0)?: **0**
- **Resulting list of numbers is (3,4,6,12,0,0)**
- **Hash this list of numbers into an integer (identifier)**
  - In Python: `hash((3, 4, 6, 12, 0, 0, 0))` ➔ `-5700861834356229464`



A beginner's guide for understanding Extended-Connectivity Fingerprints(ECFPs). Manish Kumar (2021). https://chemicbook.com/2021/03/25/a-beginners-guide-for-understanding-extended-connectivity-fingerprints.html

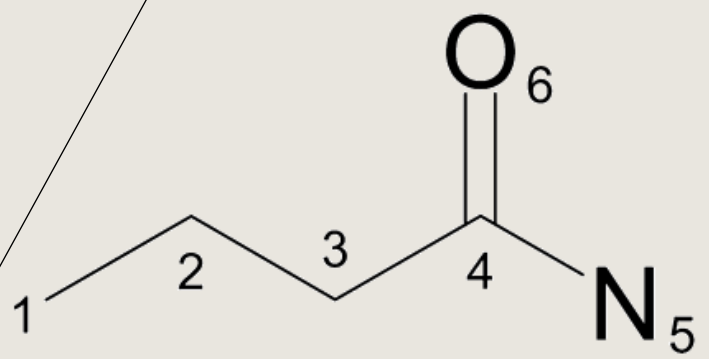
EXTENDED CONNECTIVITY FINGERPRINTS

```
# identificadores para cada atomo
atomo1 = hash((1, 1, 6, 12, 0, 3, 0)) # -CH3
atomo2 = hash((2, 2, 6, 12, 0, 2, 0)) # -CH2
atomo3 = hash((2, 2, 6, 12, 0, 2, 0)) # -CH2
atomo4 = hash((3, 4, 6, 12, 0, 0, 0)) # -C
atomo5 = hash((1, 2, 7, 14, 0, 0, 0)) # -NH2
atomo6 = hash((1, 2, 8, 16, 0, 0, 0)) # =O
```

```
atomo 1 4940186308562569707
atomo 2 -78159851478978265 76
atomo 3 -78159851478978265 76
atomo 4 -57008618343562294 64
atomo 5 -62963877442778008 66
atomo 6 86184117556823738 92
```

https://andrewbrookins.com/technology/pythons-default-hash-algorithm/

List of features (6)



A beginner's guide for understanding Extended-Connectivity Fingerprints(ECFPs). Manish Kumar (2021). https://chemicbook.com/2021/03/25/a-beginners-guide-for-understanding-extended-connectivity-fingerprints.html

EXTENDED CONNECTIVITY FINGERPRINTS

Update each atom's identifiers based on its neighbors

Each atom collects its identifier and the identifiers of its immediately neighboring atoms, into an array (list)

And we hash this list again into a new identifier.

Paso anterior

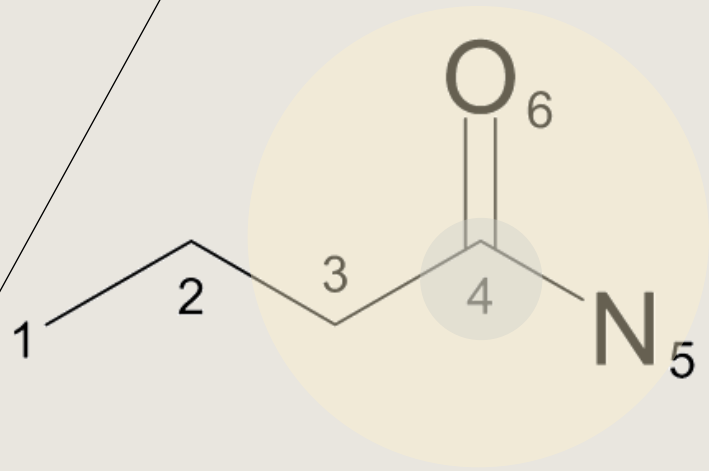
```
atomo 1 4940186308562569707
atomo 2 -7815985147897826576
atomo 3 -7815985147897826576
atomo 4 -5700861834356229464
atomo 5 -6296387744277800866
atomo 6 8618411755682373892
```

```
atomo4_updated = hash((
  1, -5700861834356229464,
  1, -7815985147897826576,
  1, -6296387744277800866,
  2, 8618411755682373892
))

-6784272694619664722
```

repetimos para los 6 átomos

List of features (12)

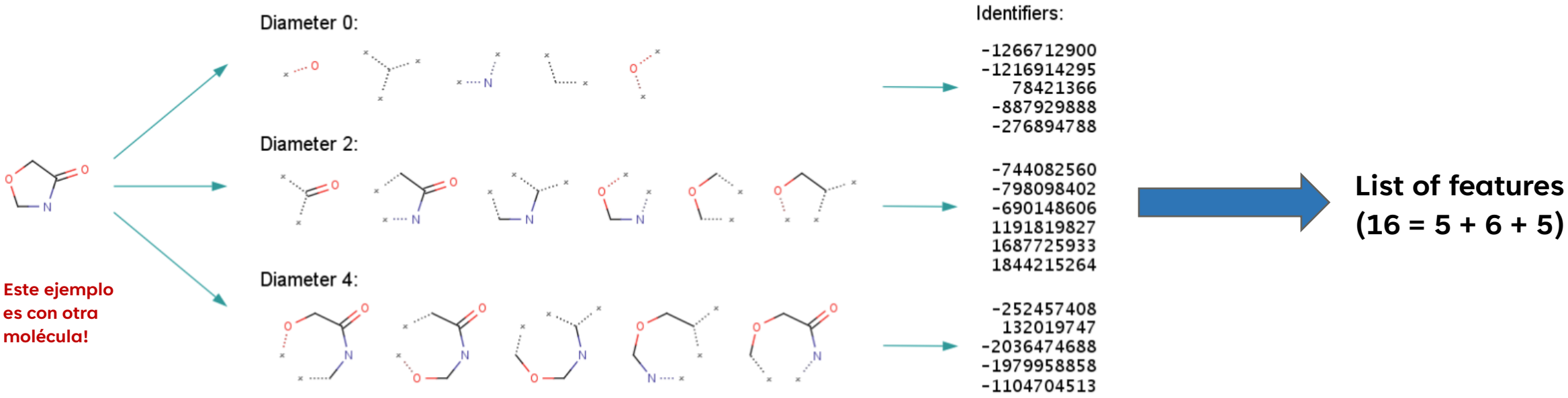


A beginner's guide for understanding Extended-Connectivity Fingerprints(ECFPs). Manish Kumar (2021). https://chemicbook.com/2021/03/25/a-beginners-guide-for-understanding-extended-connectivity-fingerprints.html

EXTENDED CONNECTIVITY FINGERPRINTS

- After that, several iterations are performed to combine the initial atom identifiers with identifiers of neighboring atoms ***until a specified diameter is reached***. Each iteration captures larger and larger circular neighborhoods around each atom

- ECFP4 = Extended Circular Fingerprint with **diameter = 4** ***(radius = 2)***
- ECFP6 = Extended Circular Fingerprint with **diameter = 6** ***(radius = 3)***



FINGERPRINTS: FOLDING AND BIT COLLISIONS

Para acomodar estos ***features*** en un fingerprint de 1024 bits

- Inicializar el fingerprint con **todos los bits en 0 (OFF)**
- Dividir cada identificador por 1024, y anotar el **resto de la división**
  - **En Python: operador módulo (%)**
- **Ese es el número de bit ➔ que se pone en 1 (ON)**

Ejemplos:

```
132019747  % 1024 = 547
1687725933 % 1024 = 877
-798098402 % 1024 = 30
```

Bit Collision:

```
-14439656419269748   % 1024 = 908
-4080868480043360372 % 1024 = 908
```

Solution: increase fingerprint size

```
-14439656419269748   % 2048 = 908
-4080868480043360372 % 2048 = 1932
```

Folding

Fixed-length binary representation

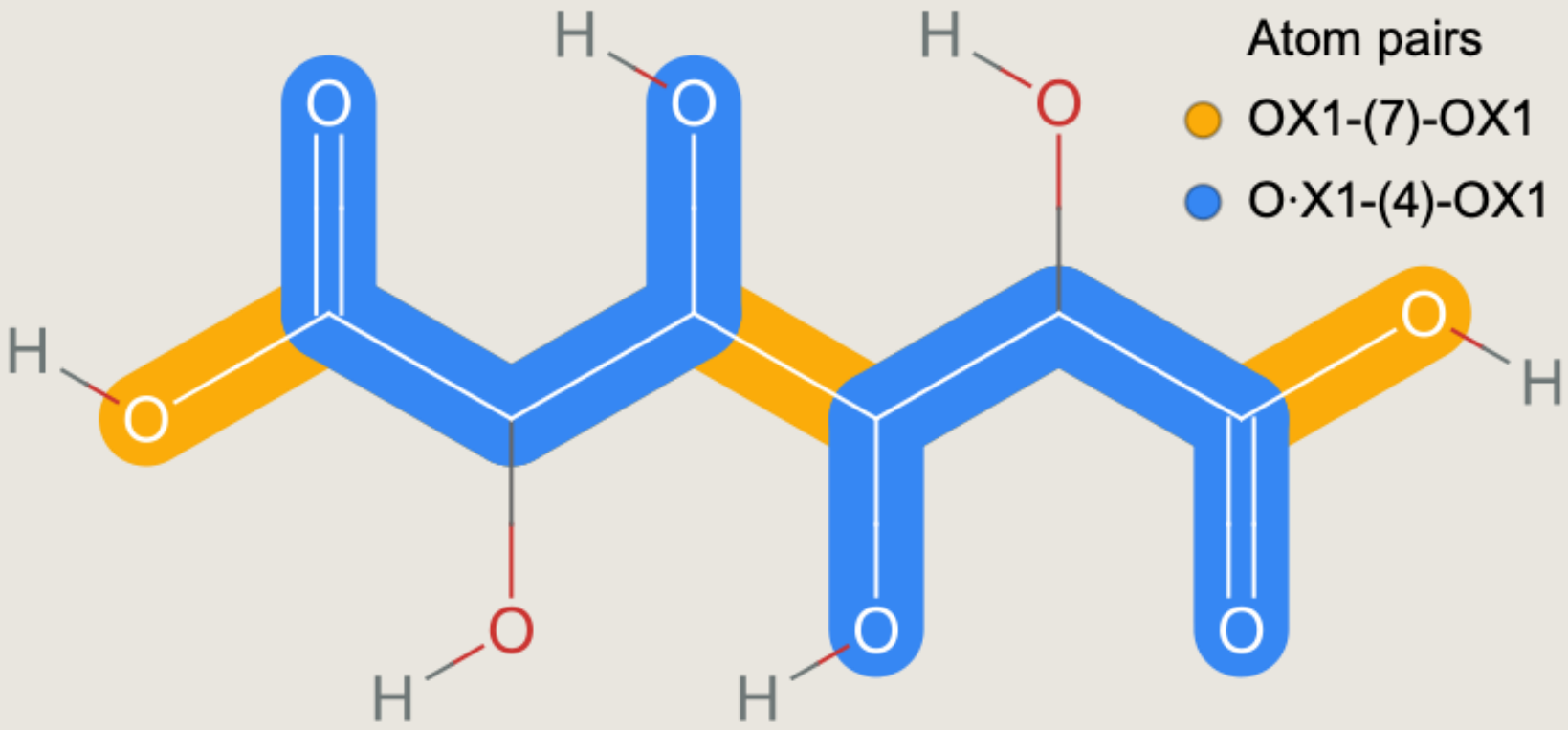
0001000000001000001000000110010000100010000000000000000100000[...]0000100000000000010

TYPES OF FINGERPRINTS

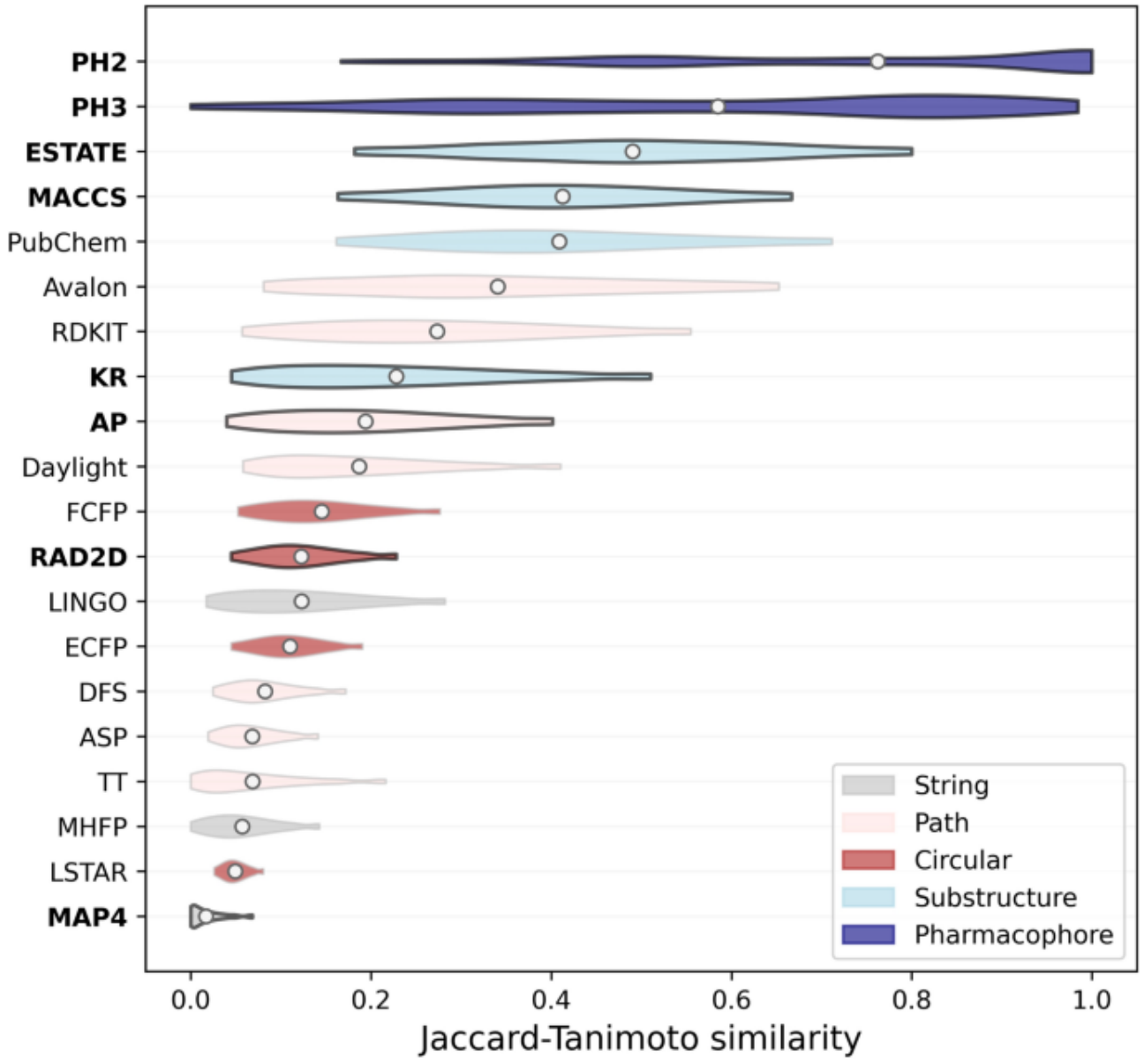
Table 1. Molecular Fingerprints[a]

ID	name	description	features	reference(s)
FP1	AP2D	topological atom pairs	1211	(44)
FP2	ASP	all-shortest paths	26,194	(45)
FP3	AT2D	topological atom triplets	56,963	(44)
FP4	DFS	all-paths (depth-first search)	48,448	(46)
FP5	ECFP	extended connectivity fingerprints	42,672	(47)
FP6	LSTAR	local path environments	85,232	(48)
FP7	MACCS	MDL public keys (166 keys)	155	(49)
FP8	PHAP2POINT2D	topological pharmacophore pairs	17	(50)
FP9	PHAP3POINT2D	topological pharmacophore triplets	302	(50)
FP10	RAD2D	topological molprint-like fingerprints	92,191	(48)
FP11	RDKit	topological daylight-like fingerprints	65,183	(43,51)

TOPOLOGICAL ATOM PAIRS



https://www.wolfram.com/language/12/molecular-structure-and-computation/topological-similarity-searching.html.en



ARE ALL FINGERPRINTS EQUAL?

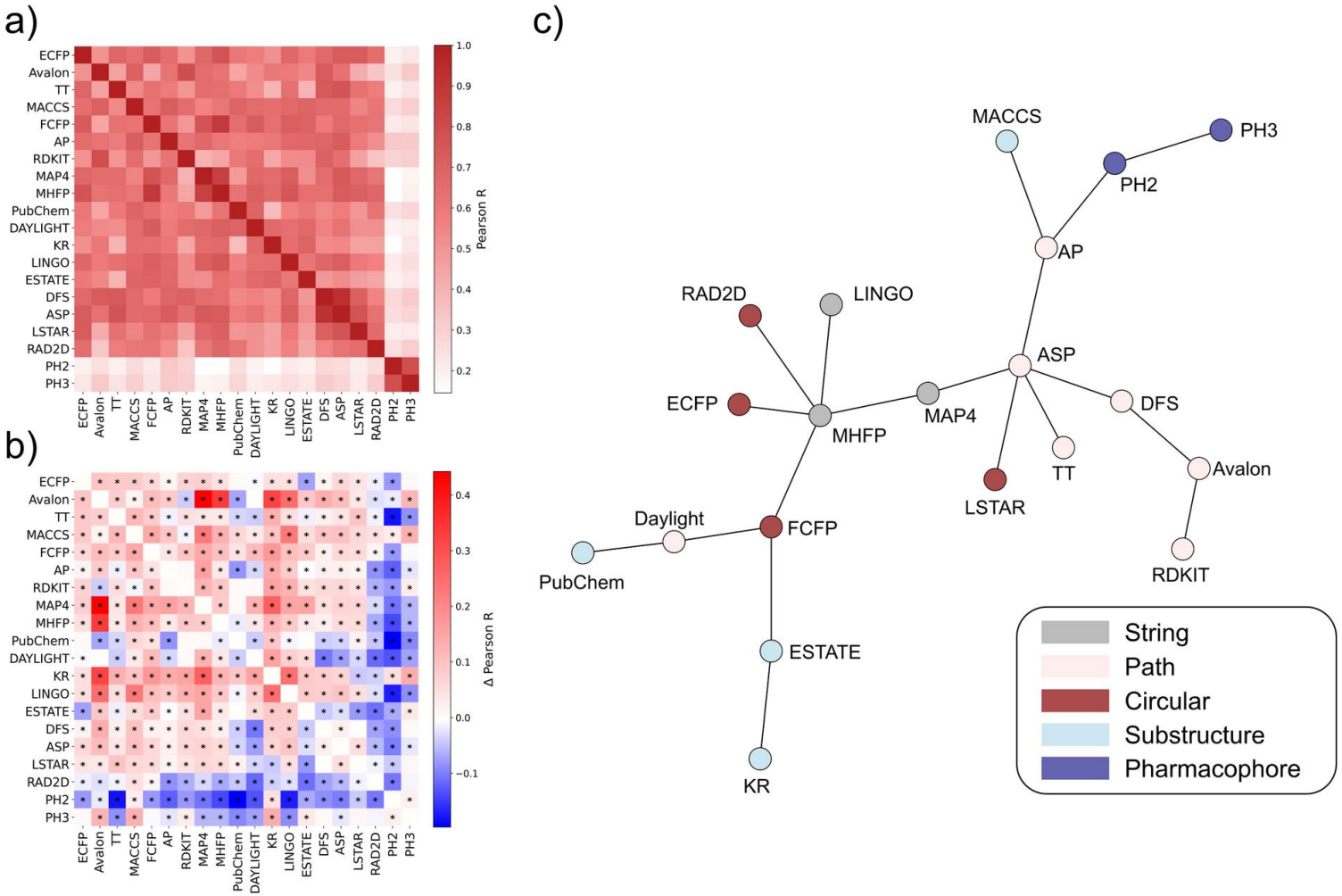
Special case: ***Natural compounds***

Different structural motifs in comparison with typical drug-like compounds, e.g.

- a wider range of molecular weight,
- multiple stereocenters
- higher fraction of sp^3-hybridized carbons

Boldini D, Ballabio D, Consonni V, Todeschini R, Grisoni F, Sieber SA. Effectiveness of molecular fingerprints for exploring the chemical space of natural products. J Cheminform. 2024 Mar 25;16(1):35. doi: 10.1186/s13321-024-00830-3. PMID: 38528548; PMCID: PMC10964529.

CORRELATION SIMILARITY ACROSS FINGERPRINT TYPES



Jaccard-Tanimoto similarity correlation analysis for all fingerprints. **a** Correlation matrix for all fingerprints evaluated in this study on the COCONUT dataset. **b** Difference between the correlation matrix obtained for the COCONUT dataset and for the Drug Repurposing Hub. Positive values indicate higher fingerprint correlation in the NP space, while negative values denote higher correlation in the drug-like space. Asterisks denote statistical significance according to one-sample Mann Whitney tests with Benjamini–Hochberg correction ($\alpha = 0.05$). **c** MST constructed from the fingerprint correlation matrix obtained for the NP chemical space. Each encoding is colored on the basis of its category

Boldini D, Ballabio D, Consonni V, Todeschini R, Grisoni F, Sieber SA. Effectiveness of molecular fingerprints for exploring the chemical space of natural products. J Cheminform. 2024 16(1):35. doi: 10.1186/s13321-024-00830-3. PMID: 38528548; PMCID: PMC10964529.

TYPES OF FINGERPRINTS

Path-based

Pharmacophore

Substructure-based

Circular

String-based

Boldini D, Ballabio D, Consonni V, Todeschini R, Grisoni F, Sieber SA. Effectiveness of molecular fingerprints for exploring the chemical space of natural products. J Cheminform. 2024 16(1):35. doi: 10.1186/s13321-024-00830-3. PMID: 38528548; PMCID: PMC10964529.

PHARMACOPHORIC FINGERPRINTS

The **pharmacophore** concept was first introduced by Paul Ehrlich in the late 19th century

A **pharmacophore** is the set of **steric** and **electronic** characteristics required to ensure better interactions with a particular biological target and to generate its biological response.

“a pharmacophore is the arrangement of a molecule’s characteristics that are accountable for a biological effect”

REPRESENTATION OF PHARMACOPHORIC HYPOTHESES

POSP FORMAT

Pharmacophore types + 3D coordinates (X, Y, Z)

EDGEP FORMAT

Pharmacophore types + shortest-path distances between points

Types:

- AROM: aromatic ring
- POSC: cation
- HACC: hydrogen bond acceptor
- HDON: hydrogen bond donor
- HYBL: hydrophobic group (ring)
- LHYBL: hydrophobic group (non-ring)

POSP format

```
AROM 5.1847 6.7106 22.5996
HACC 7.5917 7.7800 23.0800
HDON 3.8247 4.1464 21.0520
HDON 5.4160 8.6459 23.8762
```

EDGEP format

```
4
1 AROM
2 HACC
3 HDON
4 HDON
1 2 2.41
1 3 3.05
1 4 2.04
2 3 5.5
2 4 2.19
3 4 5.45
```

Number of atoms

Pharmacophore types

Points & Distances

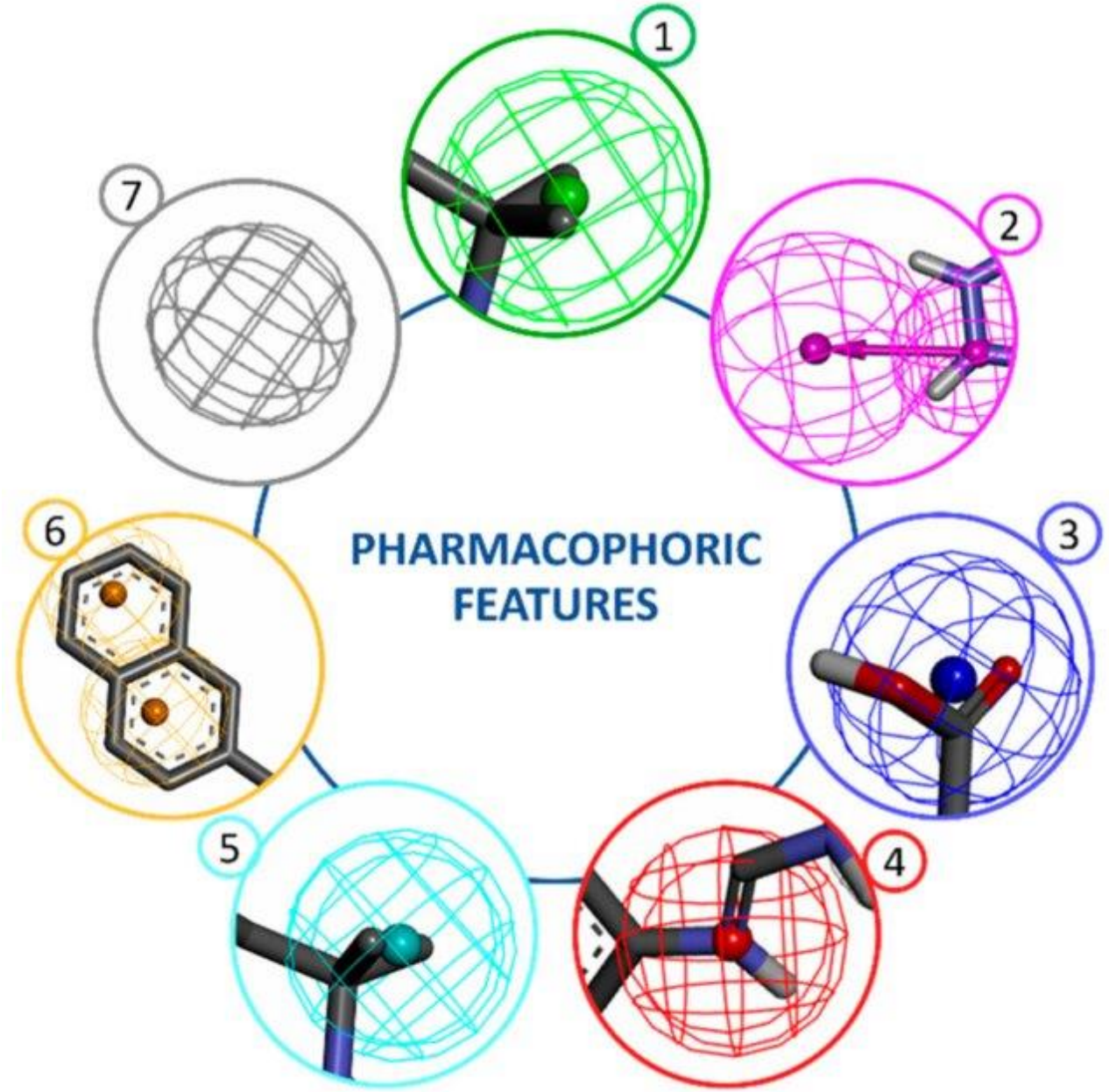
Zhu H, Zhou R, Cao D, Tang J, Li M. A pharmacophore-guided deep learning approach for bioactive molecular generation. Nat Commun. 2025 Oct 6;14(1):6234. doi: 10.1038/s41467-023-41454-9. PMID: 37803000; PMCID: PMC10558534.

Pharmacophoric types.
Hydrogen Acceptor (HA)
Hydrogen Donor (HD)
Negative ionizable (NI)
Positive ionizable (PI)
Hydrophobic (HB)
Aromatic (A)

Aggarwal R, R Koes D. PharmRL: pharmacophore elucidation with deep geometric reinforcement learning. BMC Biol. 2024 22(1):301. doi: 10.1186/s12915-024-02096-5. PMID: 39736736; PMCID: PMC11687028.

Pharmacophoric types.
1—hydrogen bond acceptor (HBA)
2—hydrogen bond donor (HBD)
3—negative ionizable (NI)
4—positive ionizable (PI)
5—hydrophobic (H)
6—aromatic (AR)
7—exclusion volume (XVOL).

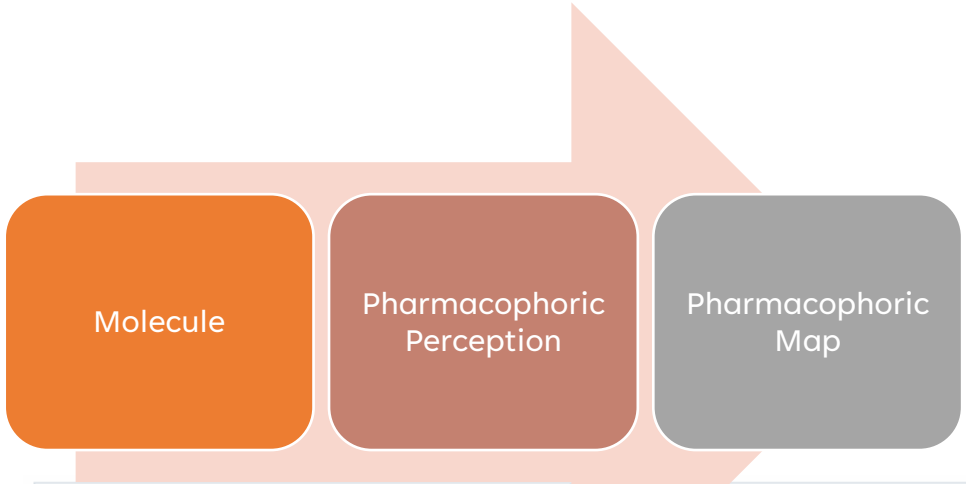
Giordano D, Biancaniello C, Argenio MA, Facchiano A. Drug Design by Pharmacophore and Virtual Screening Approach. Pharmaceuticals (Basel). 2022 15(5):646. doi: 10.3390/ph15050646. PMID: 35631472; PMCID: PMC9145410.



PHARMACOPHORIC FINGERPRINTS

Ahora que entendemos farmacóforos ...

PERCEPTION OF PHARMACOPHORIC FEATURES

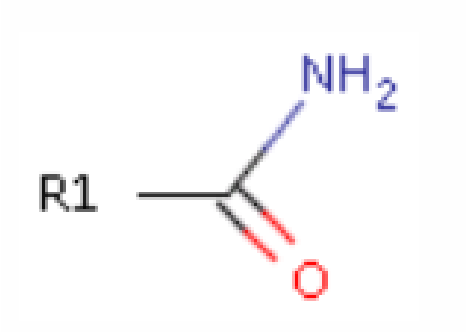


- **Atoms** are **pharmacophore points**
- Perceive **pharmacophore type** of each point
  - Using calculation (pKa, electronic potential) – Quantitative
  - Use knowledge, rules – Qualitative
    - Lots of exceptions and special cases!!!

SMILES	name	pharmacophore map
CCC=O	1-propanal	h;h;h;a
CCCO	1-propanol	h;h;h;a/d
C1=CC=CC=C1C(C)Cl	1-chloro-ethyl-benzene	r;r;r;r;r;r;h;h;h
SC(CN)C(O)=O	cystein	d;h;h;+/d;-;-/a;-/a

https://docs.chemaxon.com/display/docs/fingerprints_pharmacophore-fingerprint.md

PHARMACOPHORIC PERCEPTIONS: RULES AND EXCEPTIONS



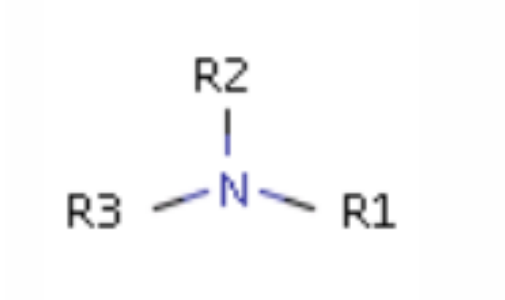
Primary amides

N = hydrogen bond donor

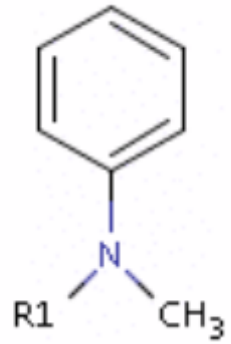
O = hydrogen bond acceptor

*nitrogen in tertiary amines is **hydrogen bond donor**, **except** when attached to an sp2 atom or, for instance, when part of N-cyano-methyl piperidine.*

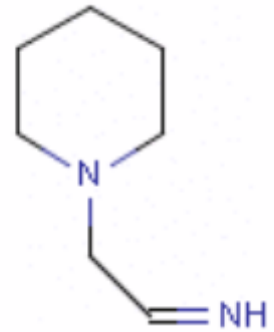
EXAMPLE EXCEPTIONS



Tertiary amine, N is protonated at pH 7.0 and ***it is a H bond donor.***

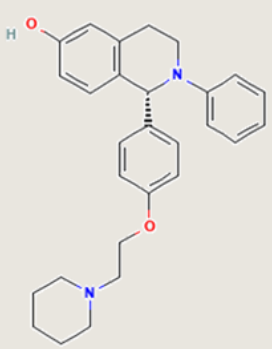


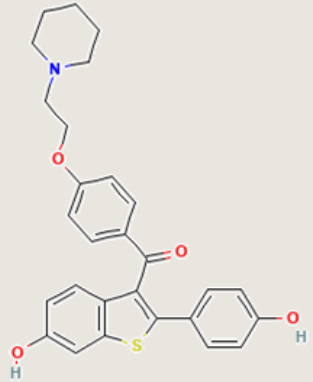
Tertiary amine where the N is next to an *sp2* atom, its lone electron pair delocalizes, thus ***it is neither a donor nor an acceptor***



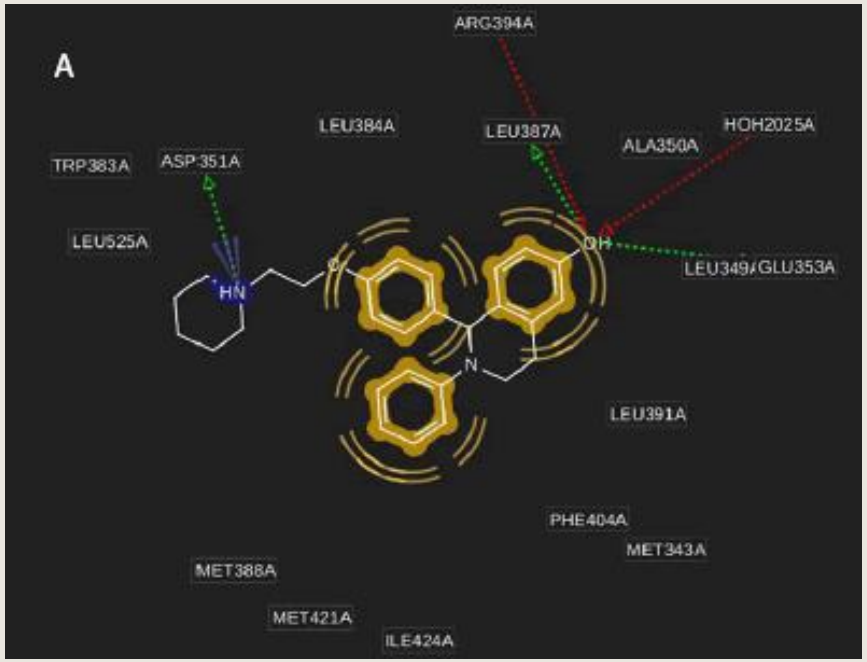
N-cyano-methyl piperidine: the N in the piperidine ring is deprotonated at pH 7.0, so ***it is neither a donor nor an acceptor.***

PHARMACOPHORIC MODELS BASED ON PROTEIN STRUCTURES

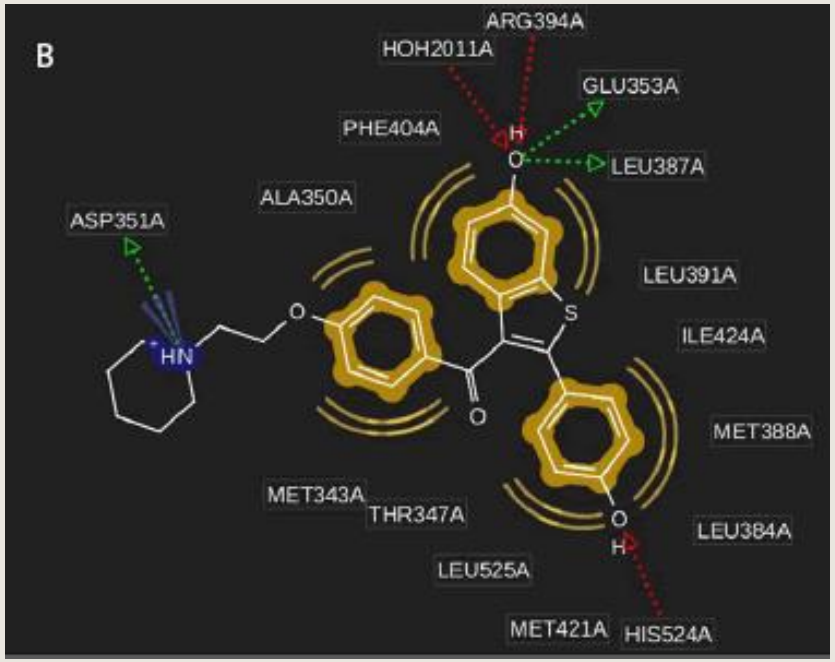




Model of 1UOM + CID448915

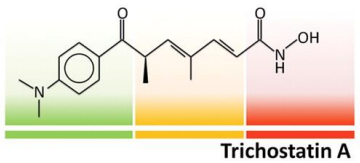


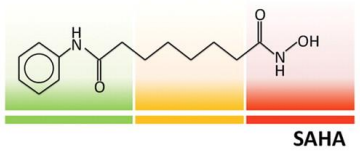
Model of 2JFA + raloxifene

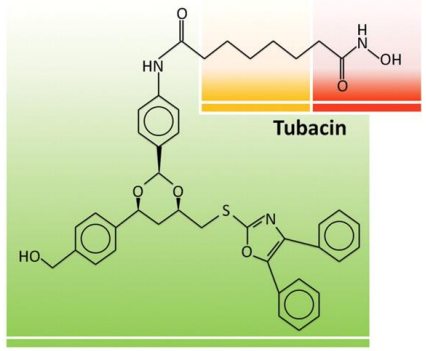


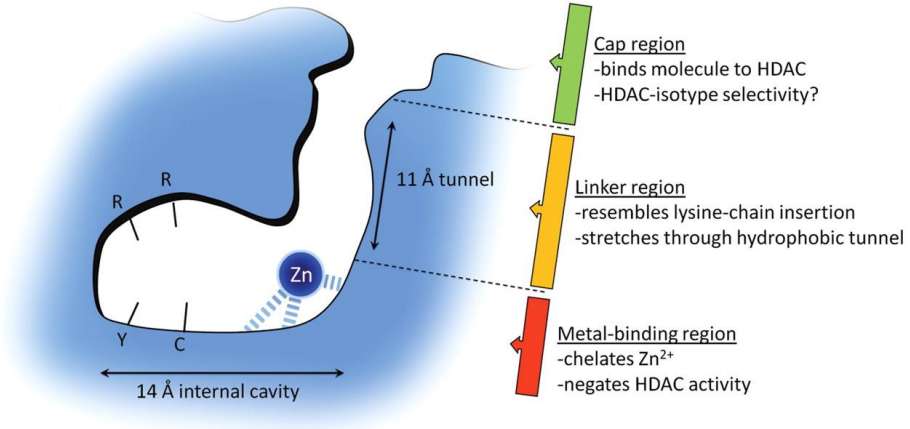
https://www.rcsb.org/structure/1UOM
https://www.rcsb.org/structure/2JFA

PHARMACOPHORIC MODELS FOR HISTONE DEACETYLASES

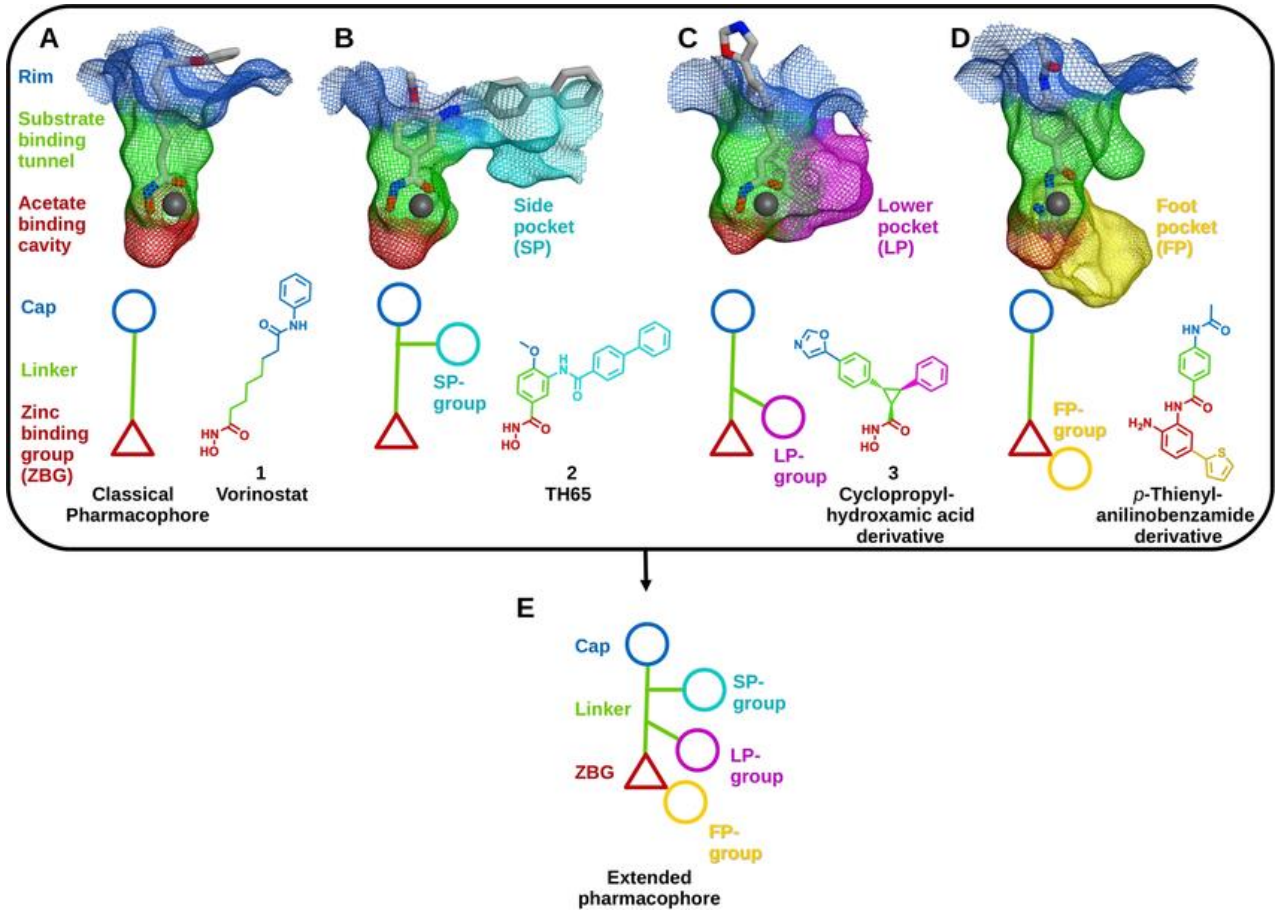








Hancock WW, Akimova T, Beier UH, Liu Y, Wang L. HDAC inhibitor therapy in autoimmunity and transplantation. Ann Rheum Dis. 2012 71 Suppl 2:i46-54. doi: 10.1136/annrheumdis-2011-200593. PMID: 22460138.



Melesina J, Simoben CV, Praetorius L, Bülbül EF, Robaa D, Sippl W. Strategies To Design Selective Histone Deacetylase Inhibitors. ChemMedChem. 2021 16(9):1336-1359. doi: 10.1002/cmdc.202000934. Epub 2021 Feb 19. PMID: 33428327.

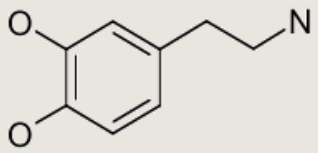
BÚSQUEDA DE SUBESTRUCTURAS

Screenings

Simple:

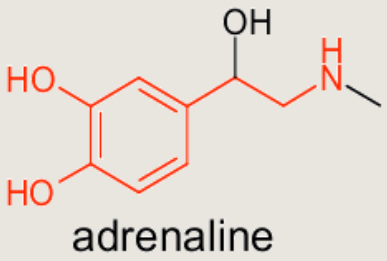
- Usa la fórmula molecular
  - La fórmula de todos los compuestos está almacenada en la base de datos
  - La fórmula de la molécula *query* se calcula al inicio de la búsqueda
  - Se descartan moléculas a las que les faltan átomos requeridos

Query:

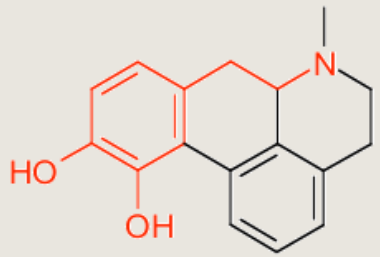


MF: C8 O2 N (H implícito)

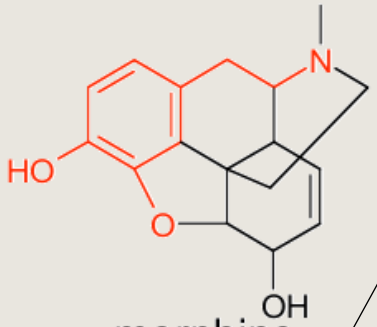
Hits:



adrenaline



apomorphine

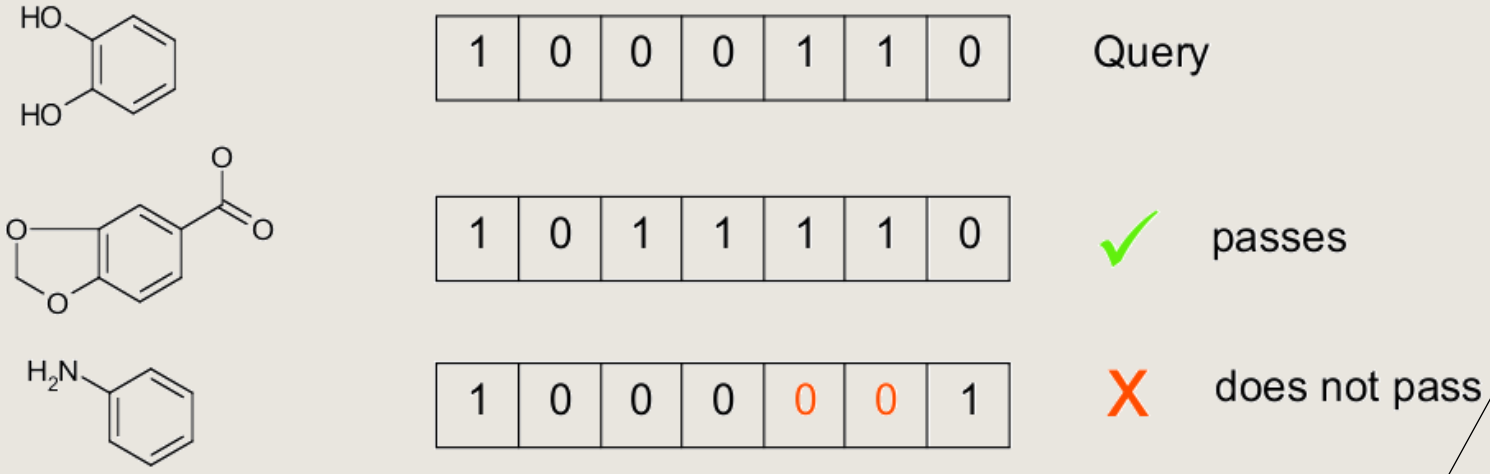


morphine

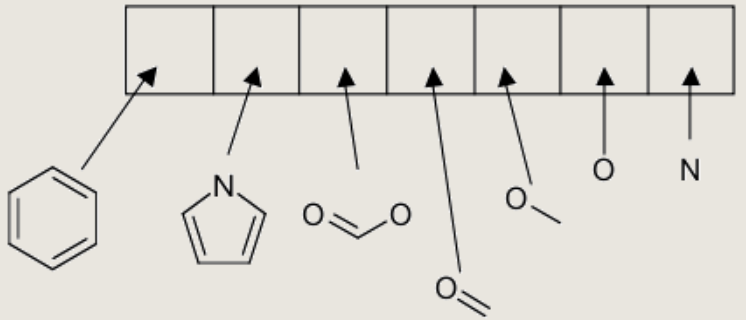
BÚSQUEDA DE SUBESTRUCTURAS: FINGERPRINTS

Fingerprint: representación abstracta de características o propiedades de una molécula (features)

- Presencia/ausencia de cada elemento
- Configuraciones electrónicas inusuales (carbono sp3, nitrógeno unido con un triple enlace)
- Anillos y sistemas de anillos (naftaleno, piridina, cyclohexano)
- Grupos funcionales (alcoholes, aminas, carboxilos, etc.)
- Se suelen utilizar tanto para búsquedas de subestructuras como para detectar similitud



Un fingerprint



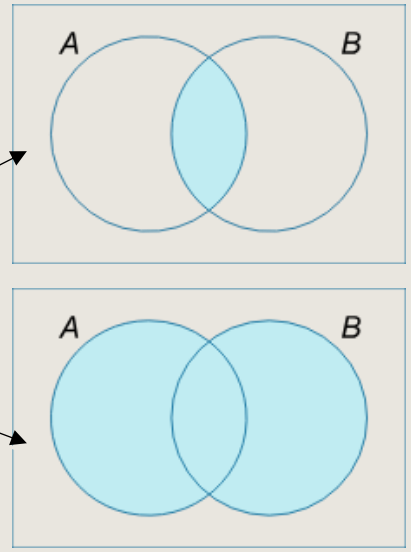
BÚSQUEDA DE SUBESTRUCTURAS Y SIMILITUD: FINGERPRINTS

Ventajas: screening extremadamente rápido

Se evalúa equivalencia entre conjuntos de bits usando el operador AND binario

Se pueden calcular distancias de similitud a partir de los bits significativos

```
A          10001101
B          01010111
-----------------
A AND B    00000101
-----------------
A OR  B    11011111
```



X = 2

Y = 4

Z = 5

W = 8

x is the number of bits set in ***both fingerprints***

y is the number of bits set in the ***first fingerprint***

z is the number of bits set in the ***second fingerprint***

w is the total number of bits in the bit string

DISTANCE METRICS: SIMILARITY, DISIMILARITY

Jaccard index (J) = Jaccard similarity coefficient = Tanimoto Index = Tanimoto similarity coefficient

(tambien llamado "Intersection Over Union")

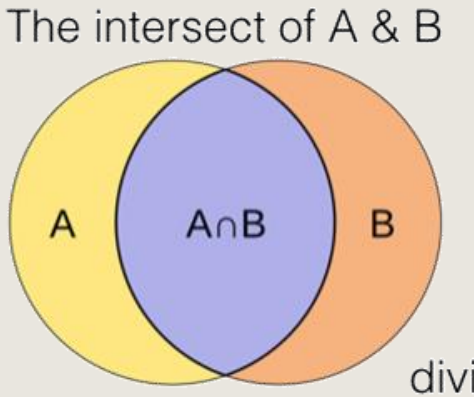
Compara similitudes entre conjuntos de datos finitos

Jaccard distance (d_J)

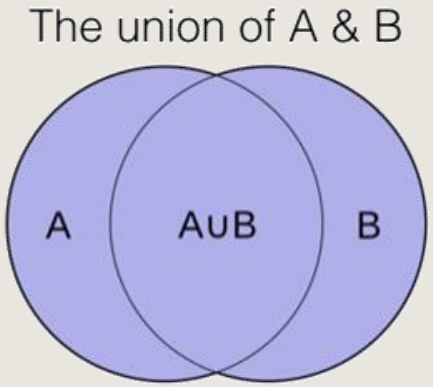
$$J(A,B) = \frac{|A \cap B|}{|A \cup B|} = \frac{|A \cap B|}{|A| + |B| - |A \cap B|}$$

$$d_J(A,B) = 1 - J(A,B) = \frac{|A \cup B| - |A \cap B|}{|A \cup B|}$$

Cociente entre el tamaño de la intersección y el tamaño de la unión de los conjuntos de datos



division



https://en.wikipedia.org/wiki/Jaccard_index

SIMILARITY METRICS

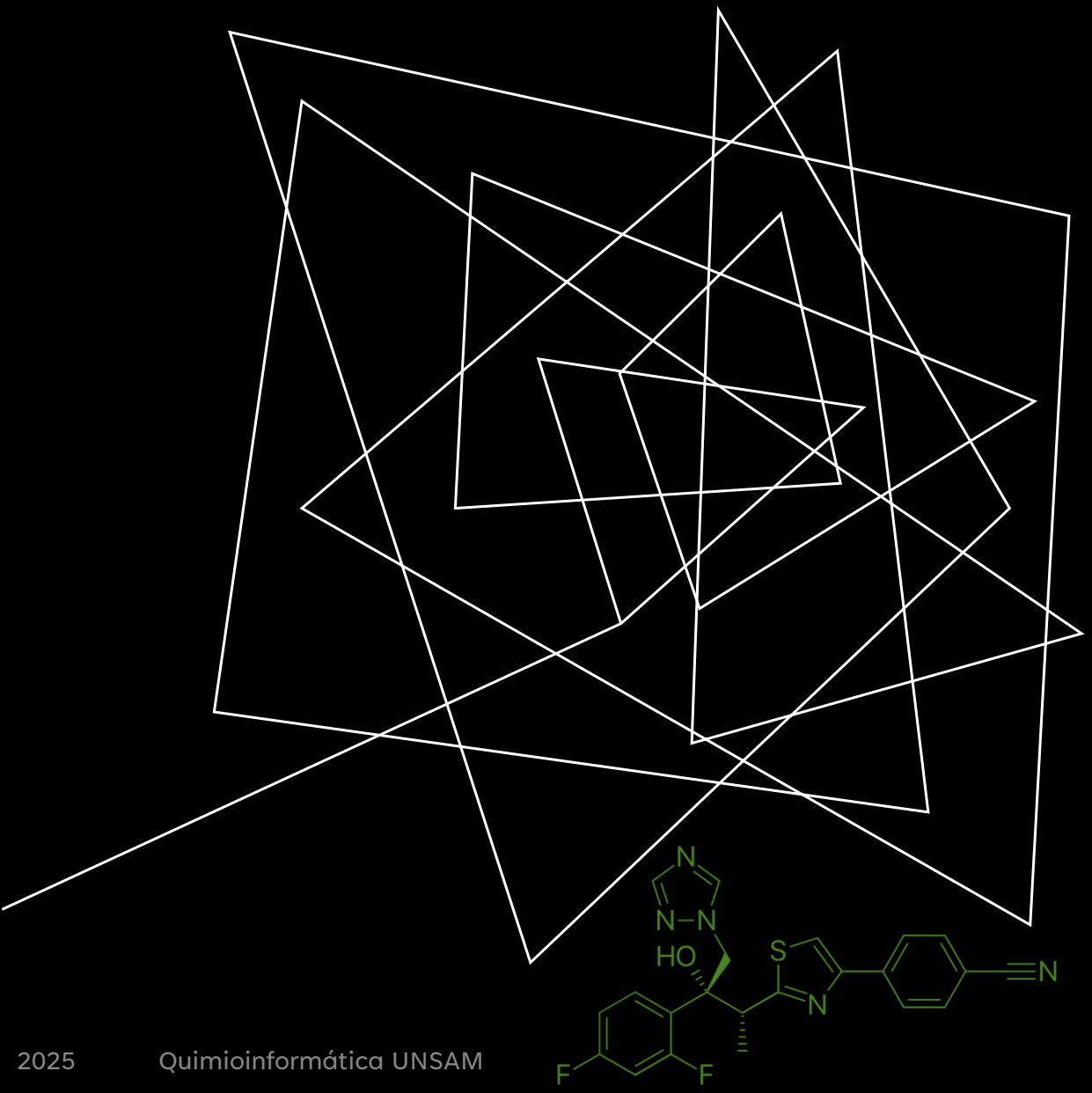
Safizadeh H, Simpkins SW, Nelson J, Li SC, Piotrowski JS, Yoshimura M, Yashiroda Y, Hirano H, Osada H, Yoshida M, Boone C, Myers CL. Improving Measures of Chemical Structural Similarity Using Machine Learning on Chemical-Genetic Interactions. J Chem Inf Model. 2021 61(9):4156-4172. doi: 10.1021/acs.jcim.0c00993.

Table 2. Similarity Coefficients[a]

name	measurement	range
Braun-Blanquet	$x/\max(y,z)$	0 to 1
Cosine	$\frac{x}{\sqrt{yz}}$	0 to 1
Dice	$\frac{2x}{y+z}$	0 to 1
Dot-product	x	0 to ∞
Euclidean	$\frac{1}{1+\sqrt{y+z-2x}}$	0 to 1
Kulczynski	$\frac{x(y+z)}{2yz}$	0 to 1
McConnaughey	$\frac{x(y+z)-yz}{yz}$	−1 to 1
Russel/Rao	x/w	0 to 1
Simpson	$x/\min(y,z)$	0 to 1
Sokal/Sneath	$\frac{x}{2y+2z-3x}$	0 to 1
Tanimoto	$\frac{x}{y+z-x}$	0 to 1
Tullos	XYZ	0 to 1
Tversky	$\frac{x}{\alpha(y-x)+(1-\alpha)(z-x)+x}$ $\alpha \in [0,1]$	0 to 1

INTERVALO

15 minutos

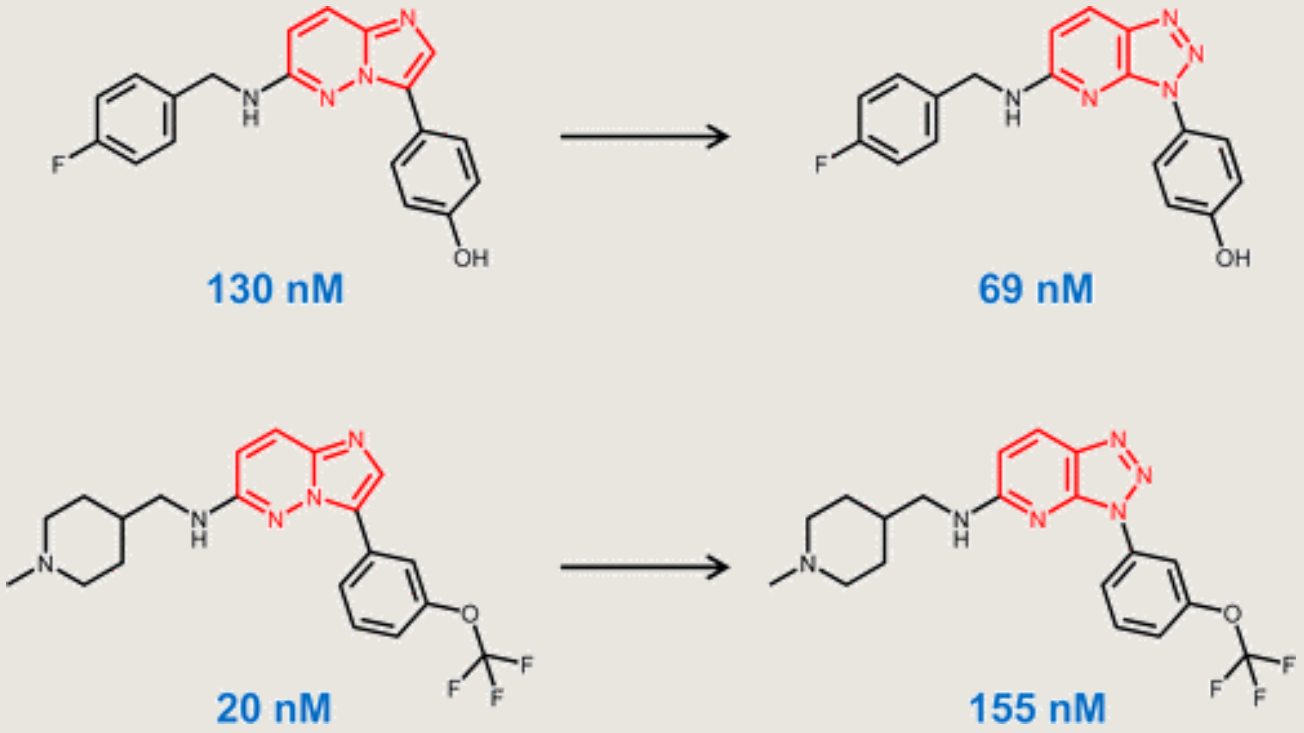


EJEMPLOS

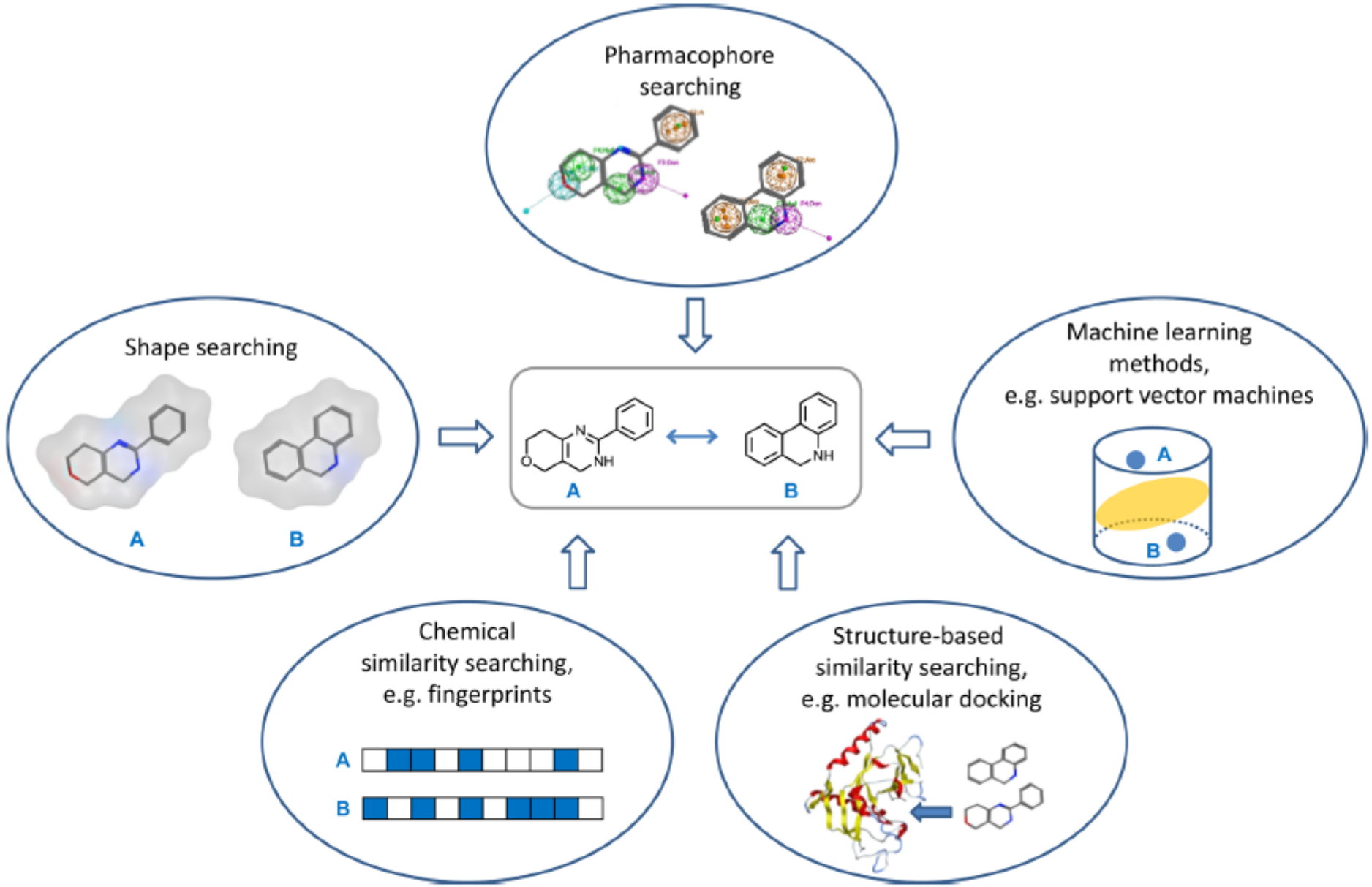
fingerprinting-with-rdkit.ipynb

SCAFFOLD HOPING

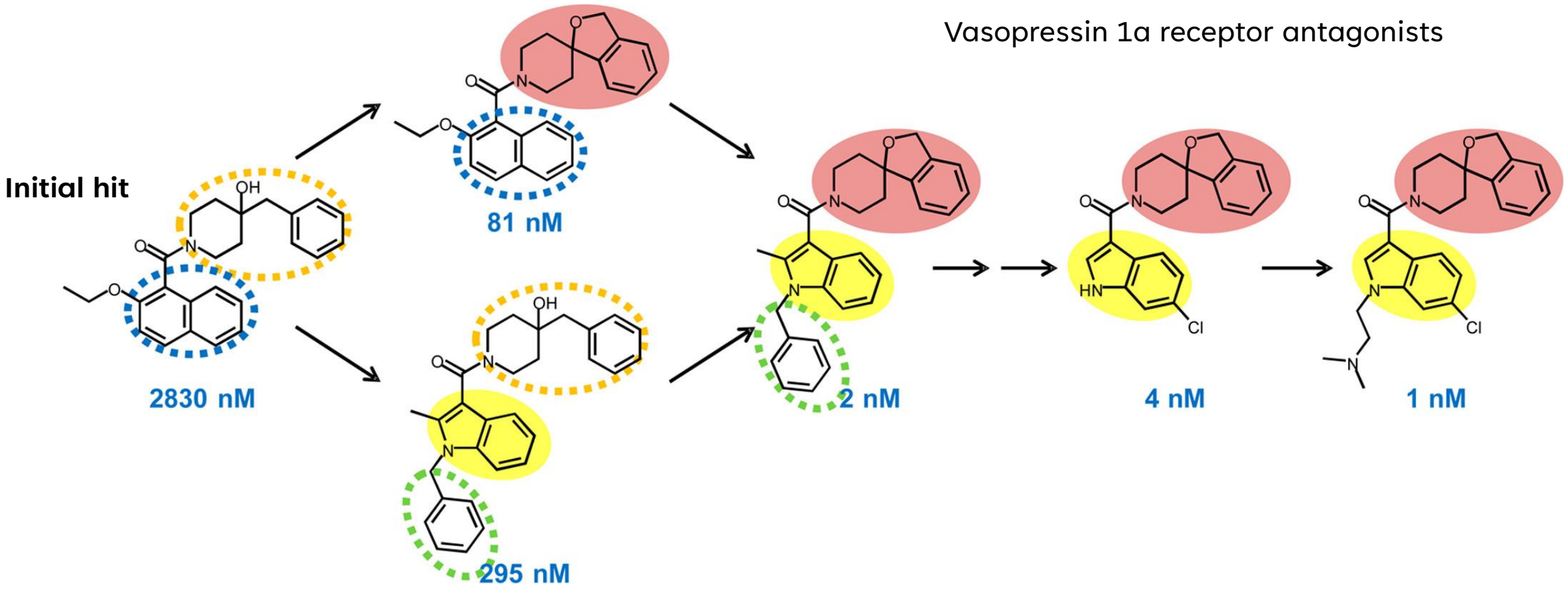
computer-aided search for active compounds containing different core structures



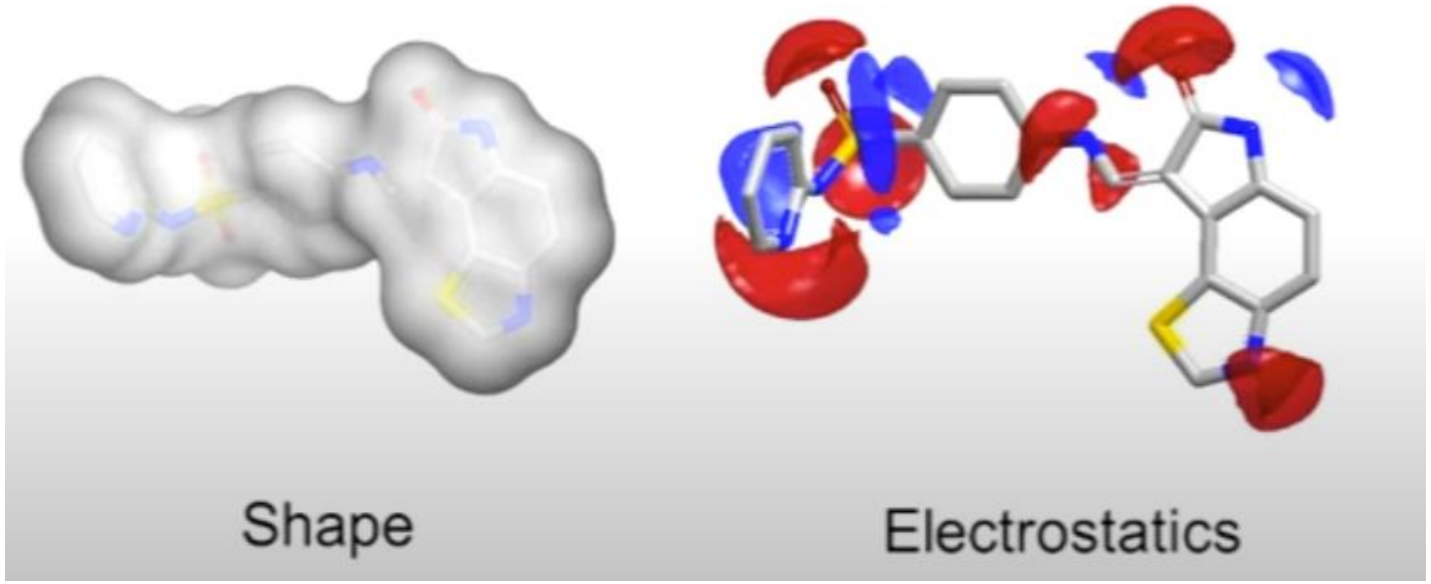
Hu Y, Stumpfe D, Bajorath J. Recent Advances in Scaffold Hopping. J Med Chem. 2017 Feb 23;60(4):1238-1246. doi: 10.1021/acs.jmedchem.6b01437. Epub 2016 Dec 21. PMID: 28001064.


Pharmacophore searching
Shape searching
A
B
A
B
Machine learning methods, e.g. support vector machines
A
B
Chemical similarity searching, e.g. fingerprints
A
B
Structure-based similarity searching, e.g. molecular docking

SCAFFOLD HOPING EXAMPLE

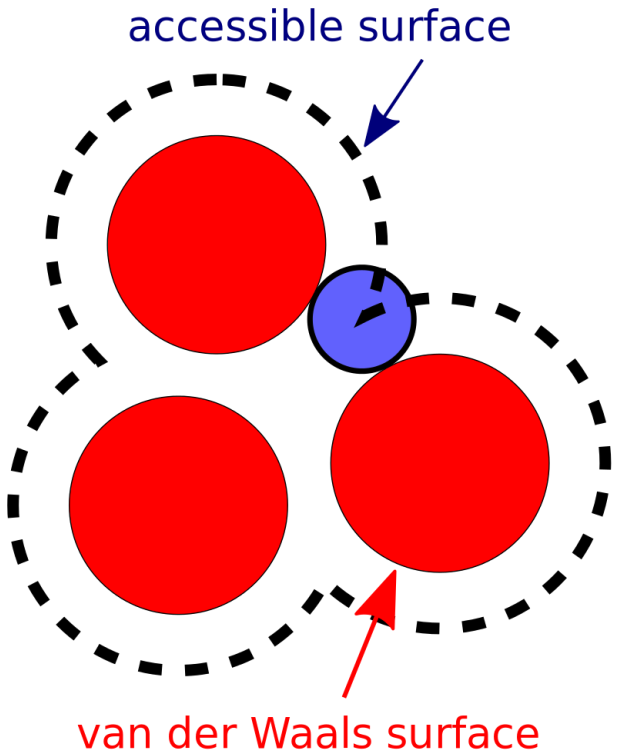


OTRAS REPRESENTACIONES DE MOLÉCULAS



SOLVENT ACCESSIBLE SURFACE AREA CALCULATION

- VSA = van der Waals Surface Area
- AS = Accessible Surface Area



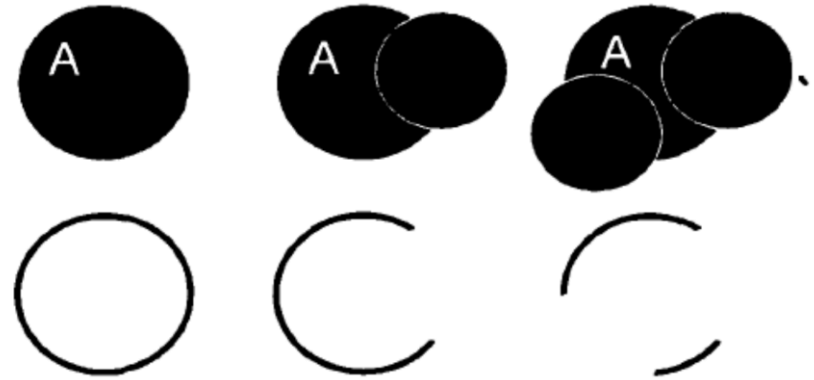
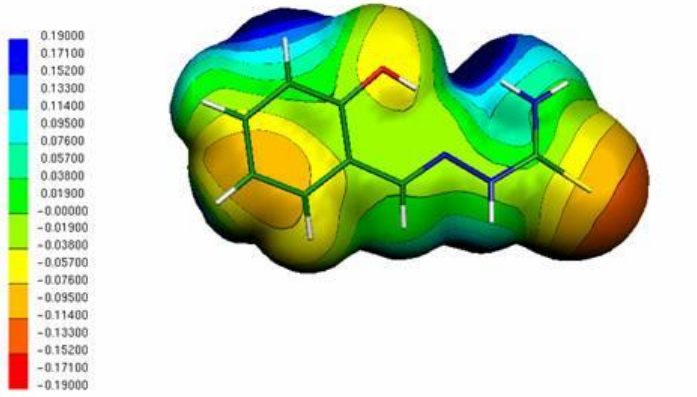


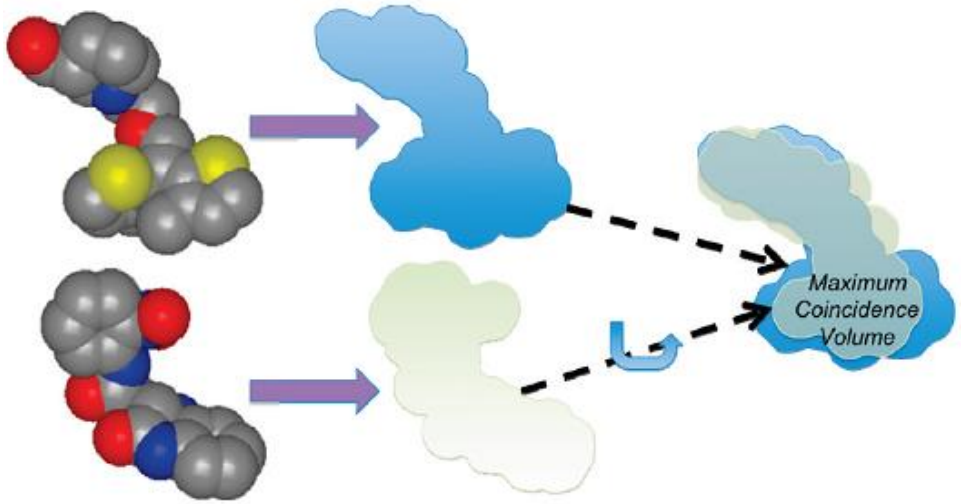
Figure 1. Assuming spherical atoms, the surface area of atom A is the amount of surface area not contained in other atoms.

Mitternacht S. FreeSASA: An open source C library for solvent accessible surface area calculations. F1000Res. 2016 Feb 18;5:189. doi: 10.12688/f1000research.7931.1. PMID: 26973785; PMCID: PMC4776673.

REPRESENTACIÓN DE MOLECULAS: 3D

- Una representación tridimensional de la molécula requiere no sólo especificar coordenadas espaciales de átomos
  - También hay que especificar
    - ***Volumen***
      - **Fused spheres**
      - **Atom-centered Gaussians**
    - ***Superficie***
    - ***Forma***
      - **Coincidencia de volumenes**



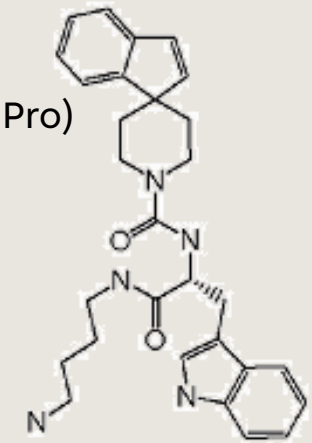


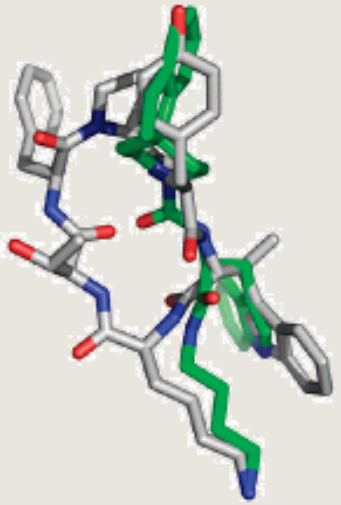
Molecular shape and medicinal chemistry: a perspective. 2010. A Nicholls *et al*. J Med Chem 53: 3862

REPRESENTACIÓN DE FORMA (SHAPE)

Aplicaciones posibles

- Búsqueda de moléculas similares
  - En este caso la similitud es a nivel de forma
  - Se pueden agregar adicionalmente limitaciones
  - Varias implementaciones en la industria farmacéutica
    - Virtual screening
      - Merck, primer aplicación publicada del método
        - Identificación de análogos no-peptídicos de:
          - antagonista endógeno del receptor de fibrinógeno (Arg-Gly-Pro)
          - Somatotrophin release inhibitor factor





REPRESENTACIÓN DE FORMA (SHAPE)

KIN: Bristol-Myers Squibb

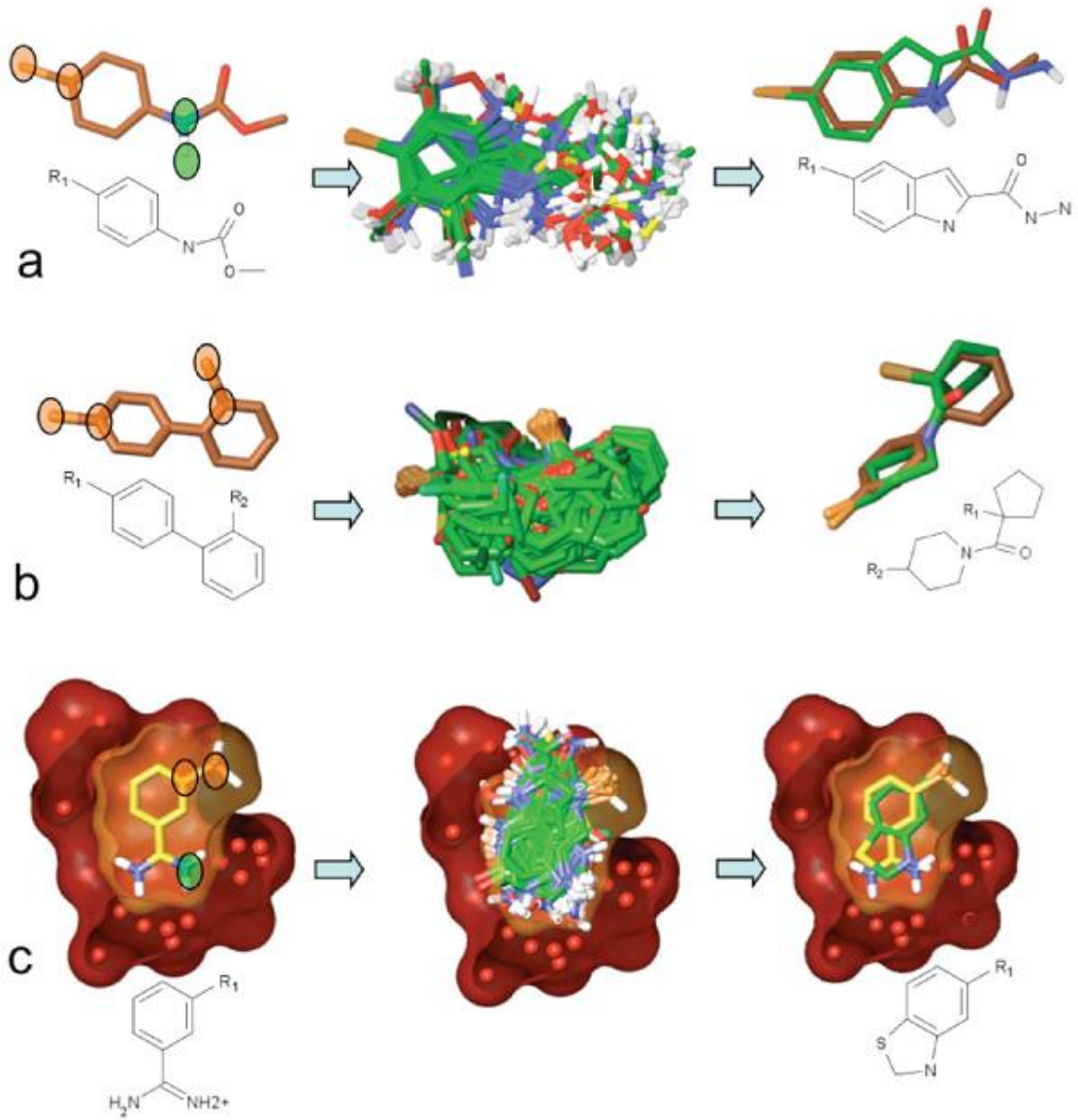
Varias aplicaciones posibles:

Lead optimization

Uno cuenta con una molécula activa que quiere optimizar

Scaffold Hoping

Facilmente explorable utilizando metodos computacionales



Molecular shape and medicinal chemistry: a perspective. 2010. A Nicholls *et al.* J Med Chem 53: 3862

CALCULO DE PROPIEDADES

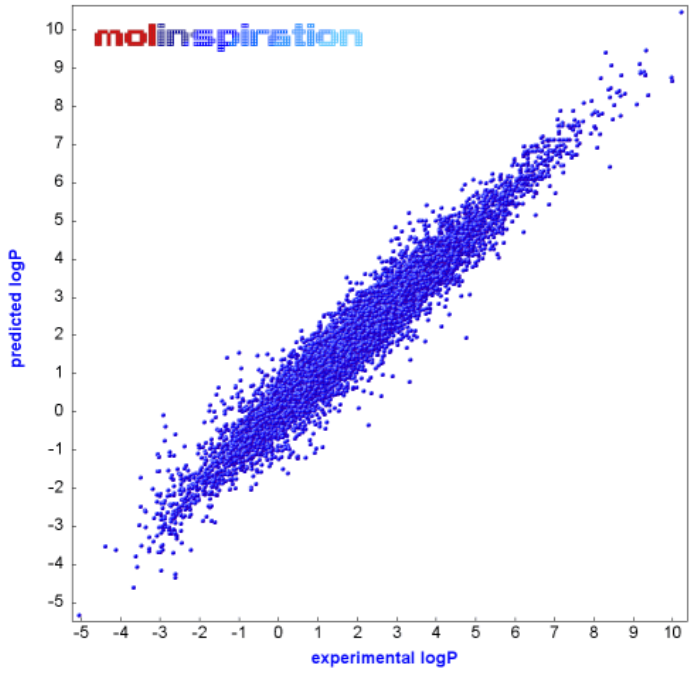
Enlaces rotables

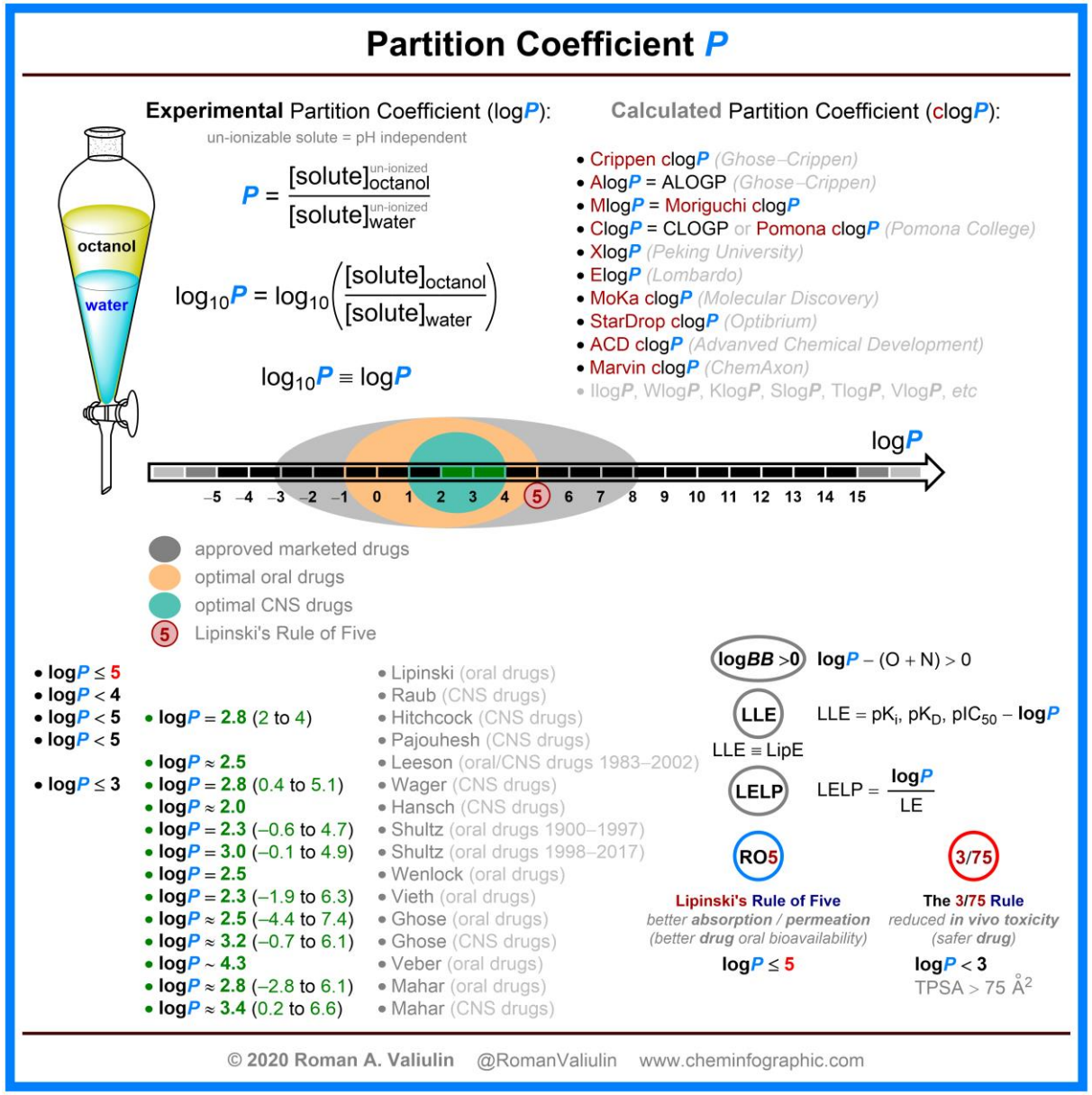
Dadores / Aceptores de puentes de hidrógeno

cLogP (coeficiente de partición octanol / agua)

PSA (polar surface area) / TPSA (topological surface area)

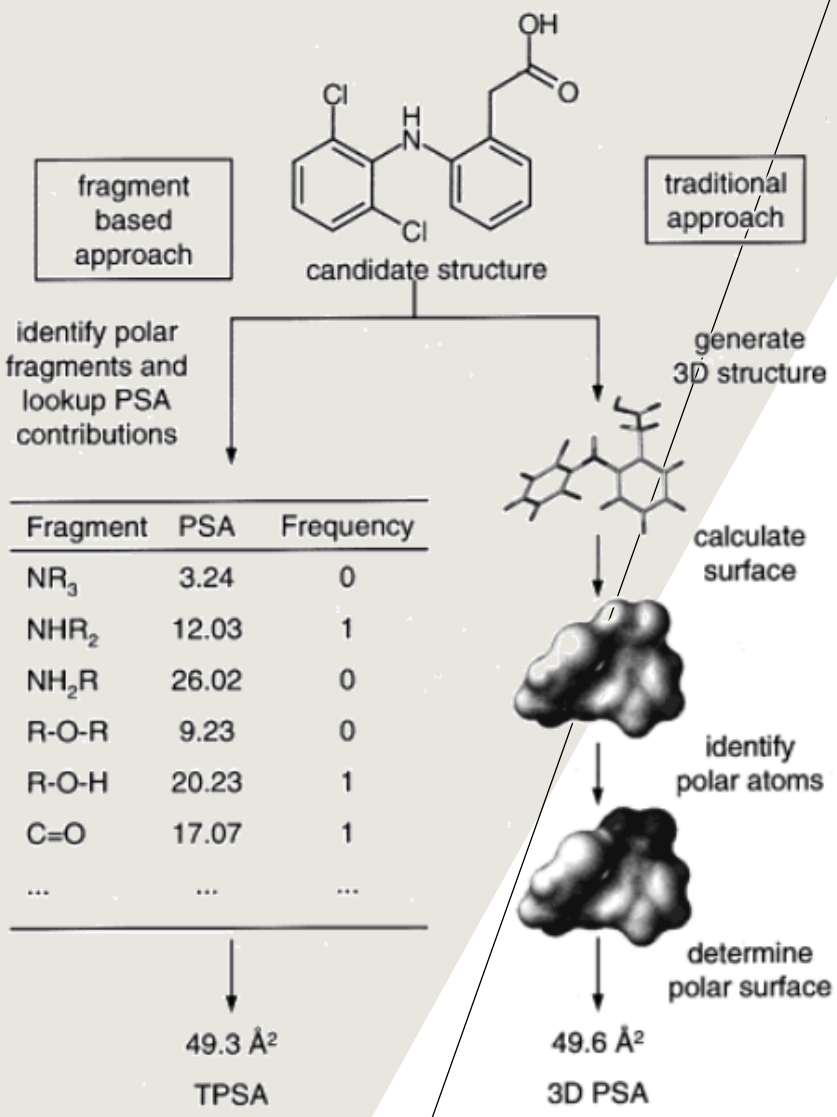
LOGP PARTITION COEFFICIENT



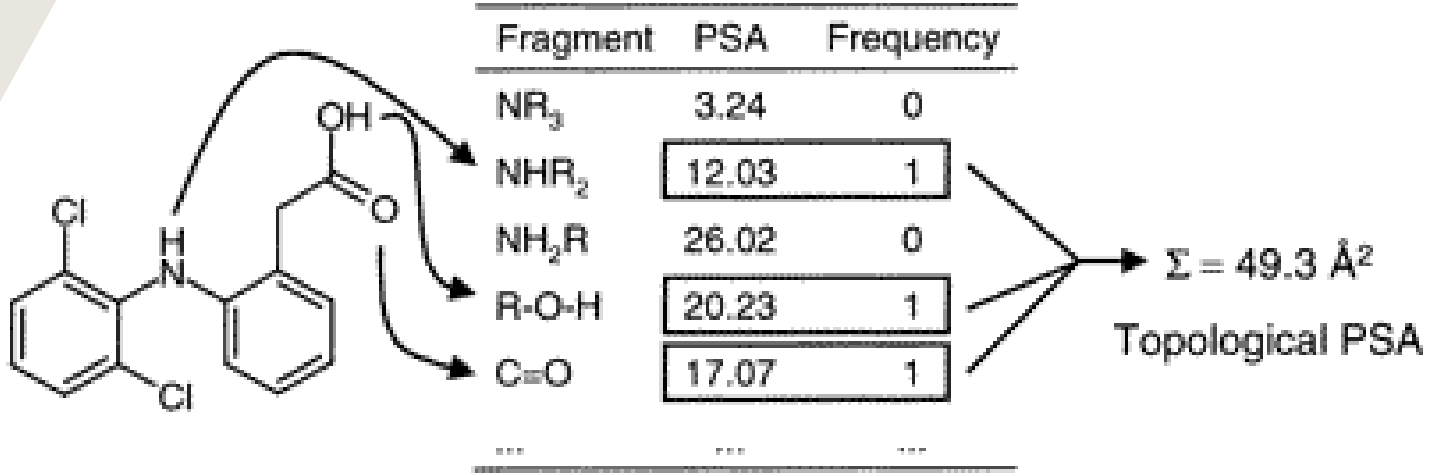


https://cheminfographic.wordpress.com/2020/05/01/partition-coefficient-p-logp/

PSA / TPSA



Fragment	PSA	Frequency
NR_3	3.24	0
NHR_2	12.03	1
NH_2R	26.02	0
R-O-R	9.23	0
R-O-H	20.23	1
C=O	17.07	1
...	...	...



Polar Surface Area (PSA, costoso)

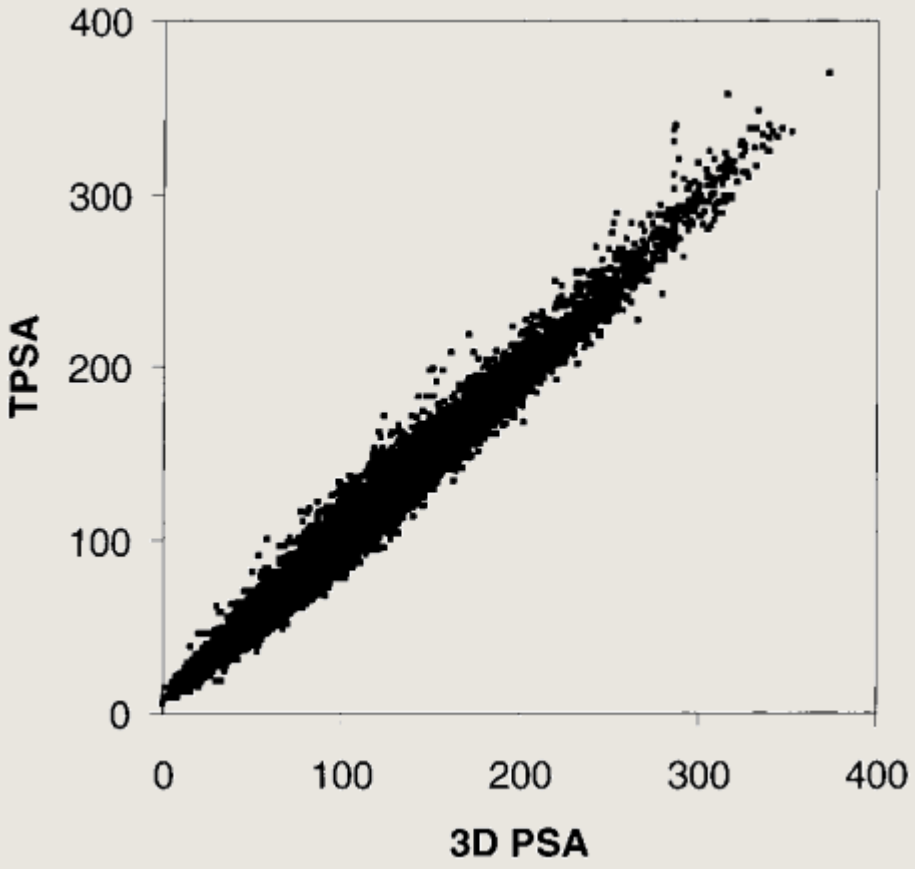
- Requiere generar conformeros 3D para calcular SA (Surface Area)

Topological Polar Surface Area (TPSA)

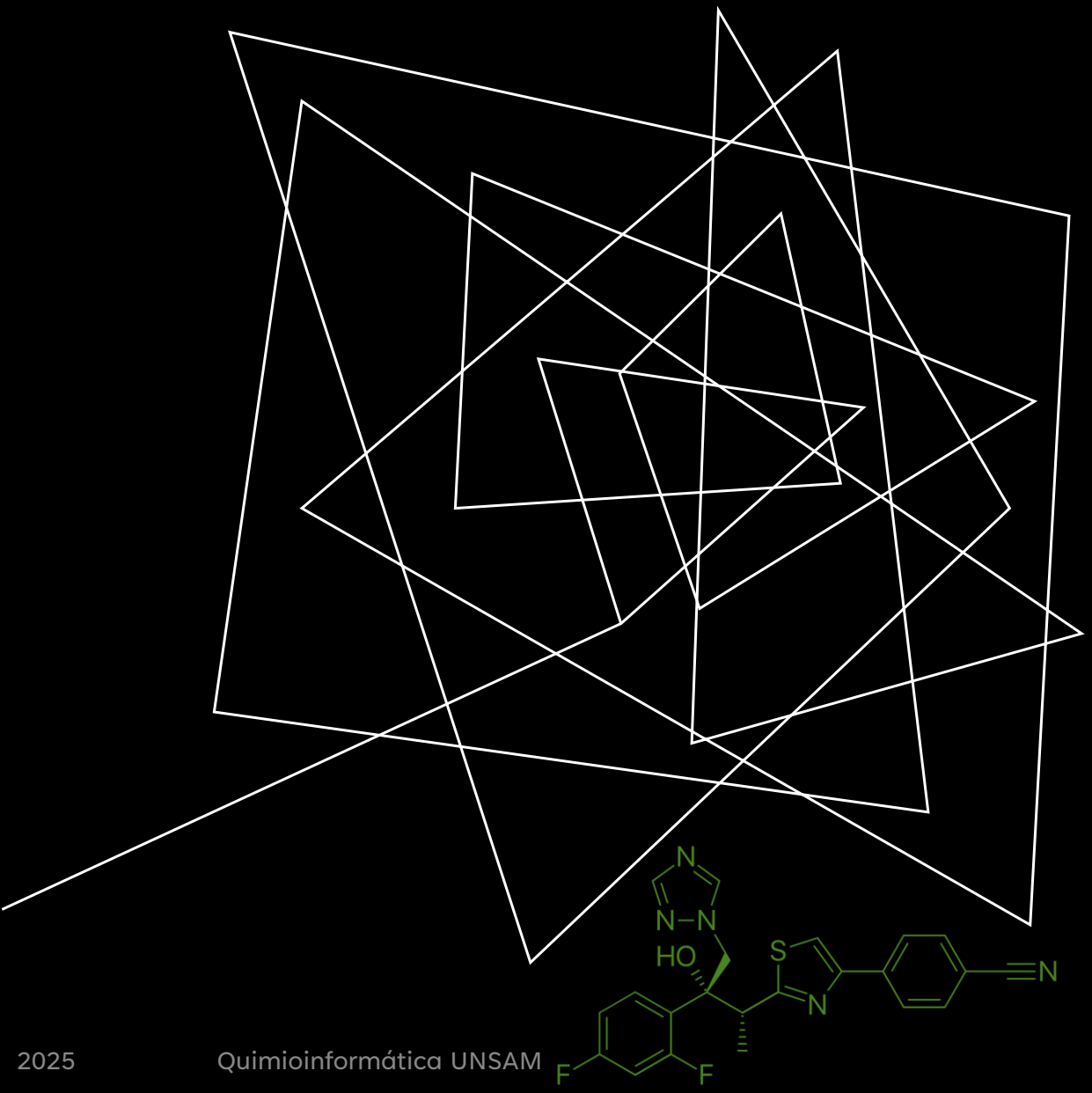
- Sumatoria de contribuciones tabuladas de fragmentos polares

Ertl P, Rohde B, Selzer P. Fast calculation of molecular polar surface area as a sum of fragment-based contributions and its application to the prediction of drug transport properties. J Med Chem. 2000 Oct 5;43(20):3714-7. doi: 10.1021/jm000942e. PMID: 11020286.

TPSA VS PSA (3D)



Ertl P, Rohde B, Selzer P. Fast calculation of molecular polar surface area as a sum of fragment-based contributions and its application to the prediction of drug transport properties. J Med Chem. 2000 Oct 5;43(20):3714-7. doi: 10.1021/jm000942e. PMID: 11020286.



ONTOLOGIAS

Organizando datos químicos

QUE ES UNA ONTOLOGIA?

Es una formalización de un área del conocimiento mediante reglas

QUÉ SIGNIFICA ONTOLOGÍA?

Webster's Revised Unabridged Dictionary

Ontology

- the things which exist
- *The department of the science of metaphysics which investigates and explains the nature and essential properties and relations of all beings, ...*

The Free On-Line Dictionary of Computing

Ontology

- Phylosophy: a systematic account of experience
- Artificial Inteligence: an explicit formal specification of how to represent the objects, concepts and other entities that are assumed to exist in some area of interest and the relationships that hold among them. [...]
- Information Science: the hierarchical structuring of knowledge about things by subcategorizing them according to their essential (or at least relevant and/or cognitive) qualities.

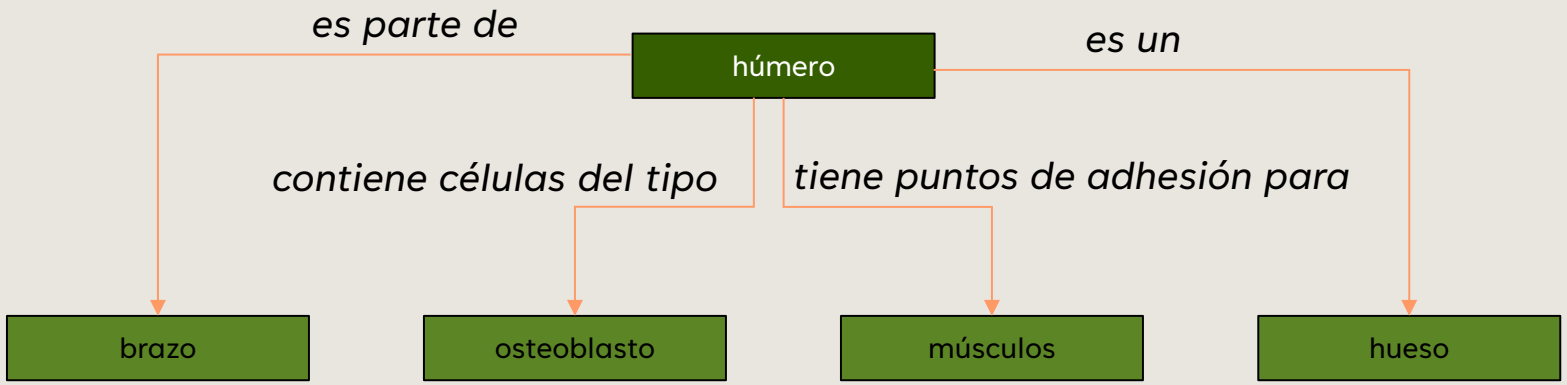
EJEMPLO: UNA ONTOLOGIA ANATOMICA

Una ontología es un área del conocimiento que ha sido formalizada:

- **Términos** (conceptos) individuales
- **Afirmaciones** (reglas) que conectan términos entre sí

Ejemplo: una ontología anatómica:

- Términos: húmero, brazo, osteoblasto, músculo, hueso
- Conexiones: es parte de, contiene células del tipo, tiene puntos de adhesión para, es un



ONTOLOGIAS QUIMICAS

ChEBI – Chemical Entities of Biological Interest

https://www.ebi.ac.uk/chebi/

Son varias ontologías agrupadas:

- **Molecular Structure**
  - **Términos:** hydrocarbons, carboxylic acids, tertiary amines
  - **Afirmaciones:** (reglas): is a (es un), has part (contiene)

caffeine (CHEBI:27732) **is a** purine alkaloid (CHEBI:26385)
caffeine (CHEBI:27732) **is a** trimethylxanthine (CHEBI:27134)

caffeine monohydrate (CHEBI:31332) **has part** caffeine (CHEBI:27732)
sodium caffeine benzoate (CHEBI:32140) **has part** caffeine (CHEBI:27732)

ONTOLOGIAS QUIMICAS

ChEBI – Chemical Entities of Biological Interest

https://www.ebi.ac.uk/chebi/

Son varias ontologías agrupadas:

- **Role**
  - **Chemical Role**
    - ej ligand, inhibitor, surfactant
  - **Biological Role**
    - antibiotic, antiviral agent, coenzyme, hormone
  - **Application**
    - pesticide, antirheumatic drug, fuel
  - **Términos:** por ej alguna entidad molecular (o una parte)
  - **Afirmaciones:** (reglas): has role (tiene rol de)

caffeine (CHEBI:27732) **has role** environmental contaminant (CHEBI:78298)
caffeine (CHEBI:27732) **has role** adenosine A2A receptor antagonist (CHEBI:53121)
caffeine (CHEBI:27732) **has role** food additive (CHEBI:64047)

CHEBI: AFIRMACIONES (REGLAS)

CHEBI: AFIRMACIONES (REGLAS)

CHEBI:16482 naphthalene

CHEBI:50715 methylnaphthalene

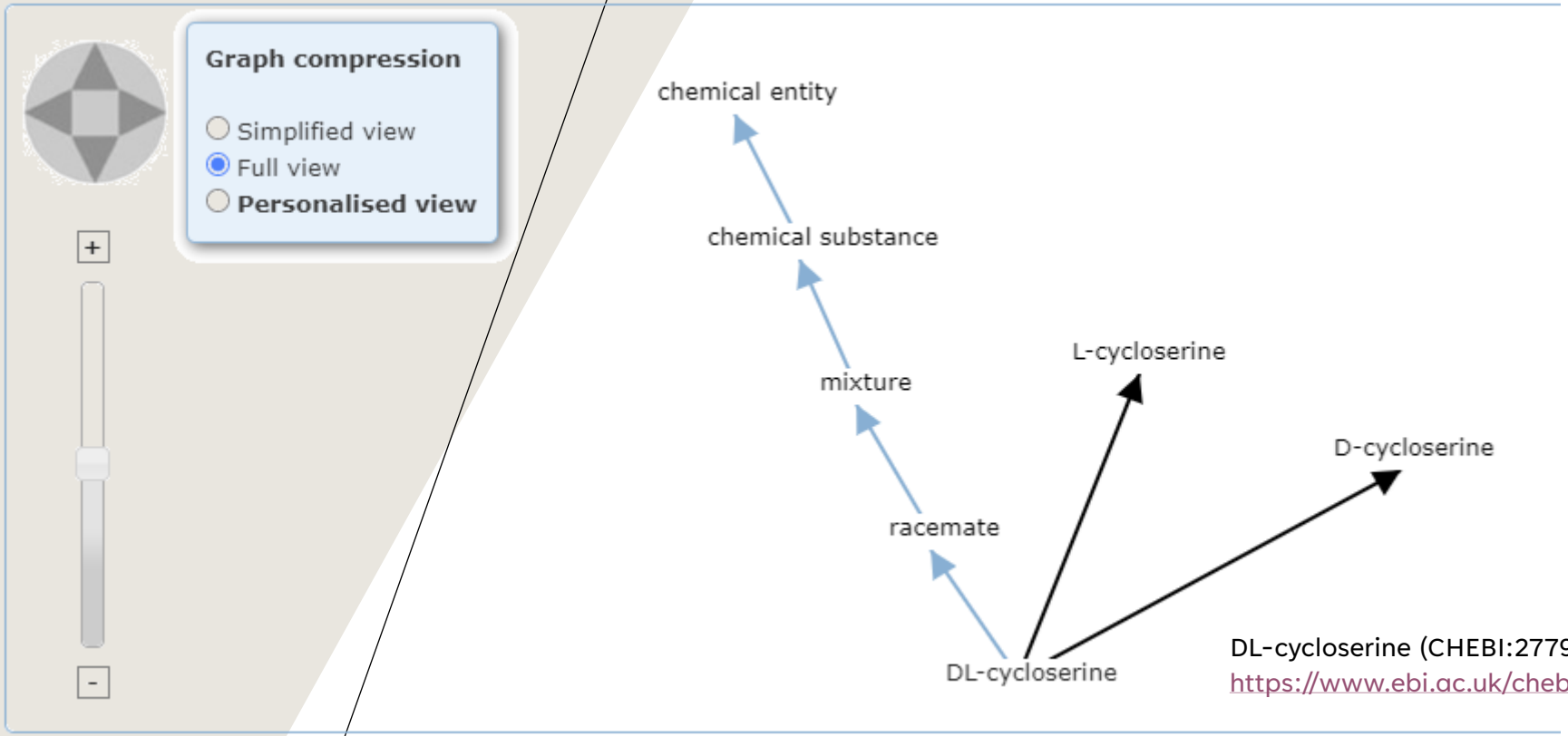
CHEBI:16414 L-valine

CHEBI:32853 L-valyl group

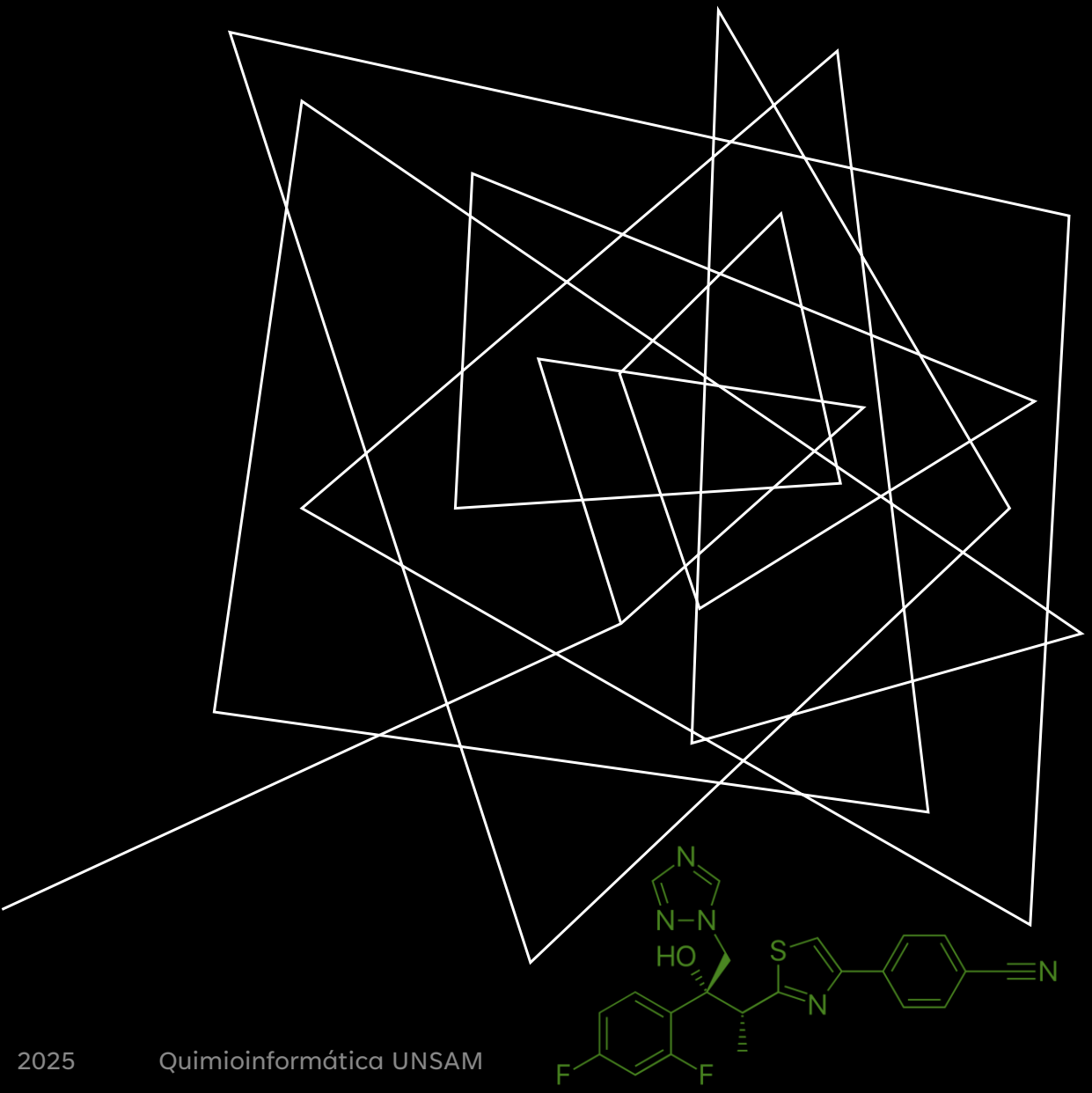
CHEBI:35482 opioid analgesic

CHEBI:17303 morphine

NAVEGACIÓN GRAFICA DE LA ONTOLOGIA



DL-cycloserine (CHEBI:27792)
https://www.ebi.ac.uk/chebi/chebiOntology.do?chebiId=CHEBI:27792



EJEMPLOS

https://www.ebi.ac.uk/chebi/searchId.do?chebiId=CHEBI:27732

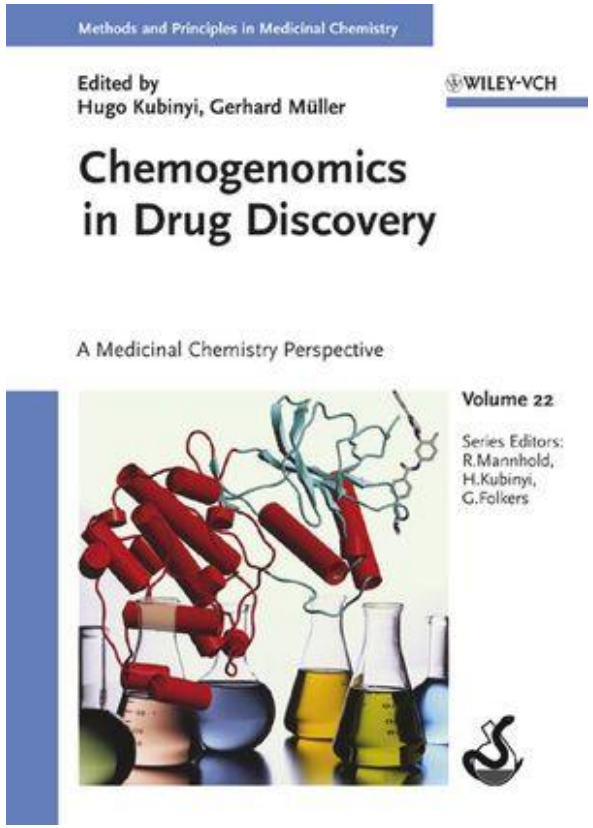
https://www.ebi.ac.uk/chebi/searchId.do?chebiId=CHEBI:78298

...

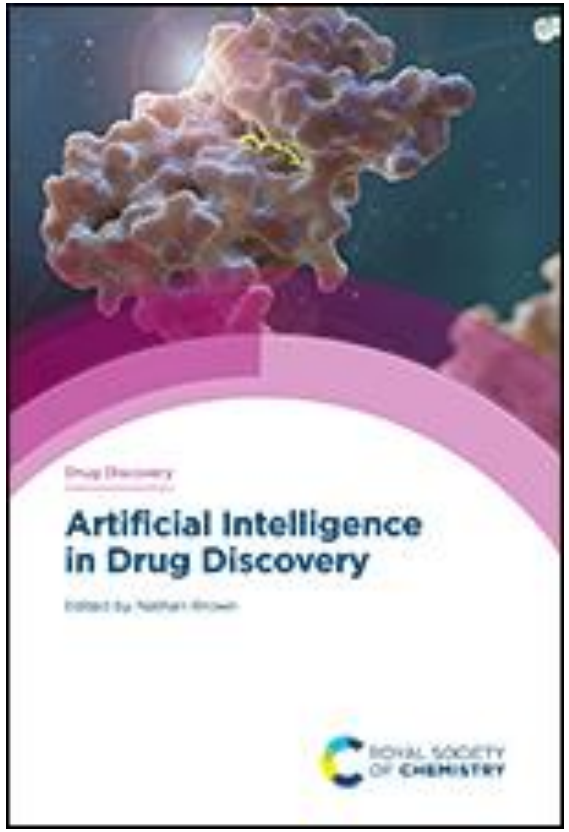
BIBLIOGRAFÍA | MATERIAL DE LECTURA

ChEBI
USER MANUAL

https://www.ebi.ac.uk/chebi/aboutChebiForward.do



Chemogenomics in Drug Discovery: A Medicinal Chemistry Perspective (2006). Edited by Hugo Kubinyi & Gerhard Müller, Wiley-VCH. https://www.wiley.com/en-us/Chemogenomics+in+Drug+Discovery%3A+A+Medicinal+Chemistry+Perspective-p-9783527604029



Artificial Intelligence in Drug Discovery (2020). Edited by Nathan Brown. Royal Society of Chemistry. https://doi.org/10.1039/9781788016841

RDKit
Open-Source Cheminformatics and Machine Learning

The RDKit Book (2025). https://www.rdkit.org/docs/RDKit_Book.html