Facultade de Química
Universidade de Vigo

"Computational approaches in drug discovery: expectations and reality."

Stefano Moro

Molecular Modeling Section (MMS)

Department of Pharmaceutical and Pharmacological Sciences

University of Padova

Vigo.... nice memories:

2007 2010

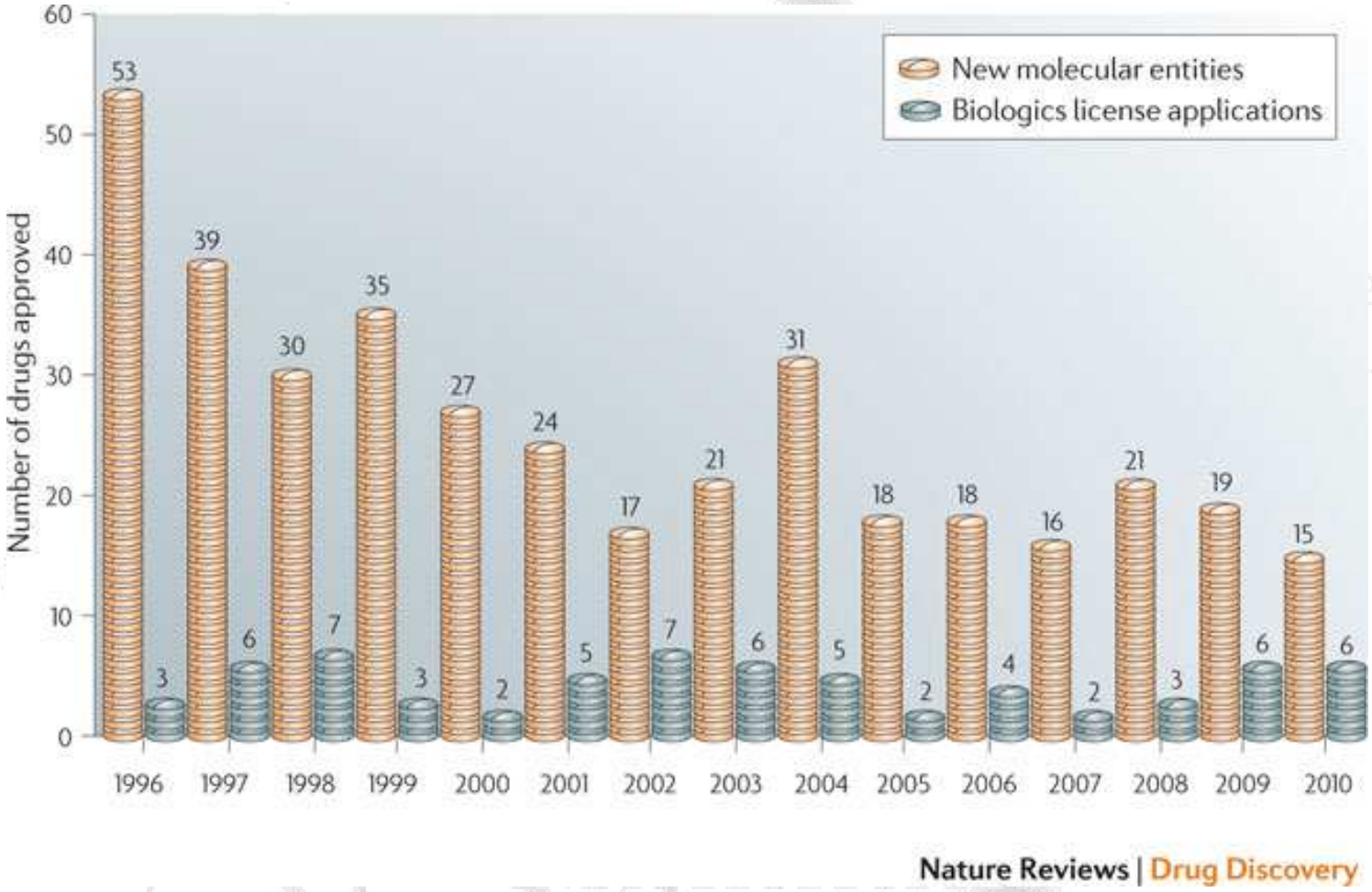
.... just to summarize the aim of our short course:

Farmaco *(drug)* ⟷ **Progettazione** *(design)*

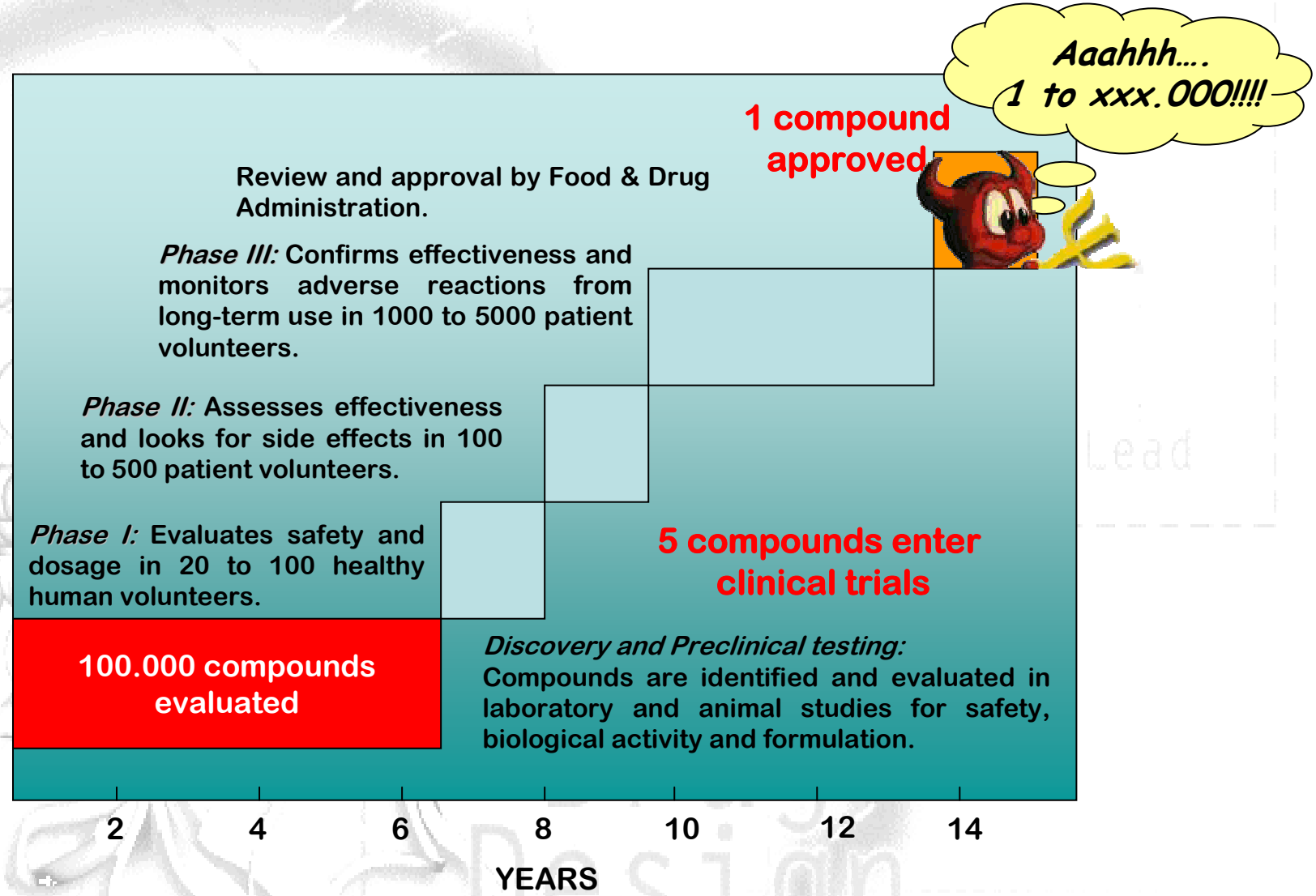
Informatica *(informatics)*

Why informatics can bridged drug-design?

I would like to start with a very pragmatic consideration:



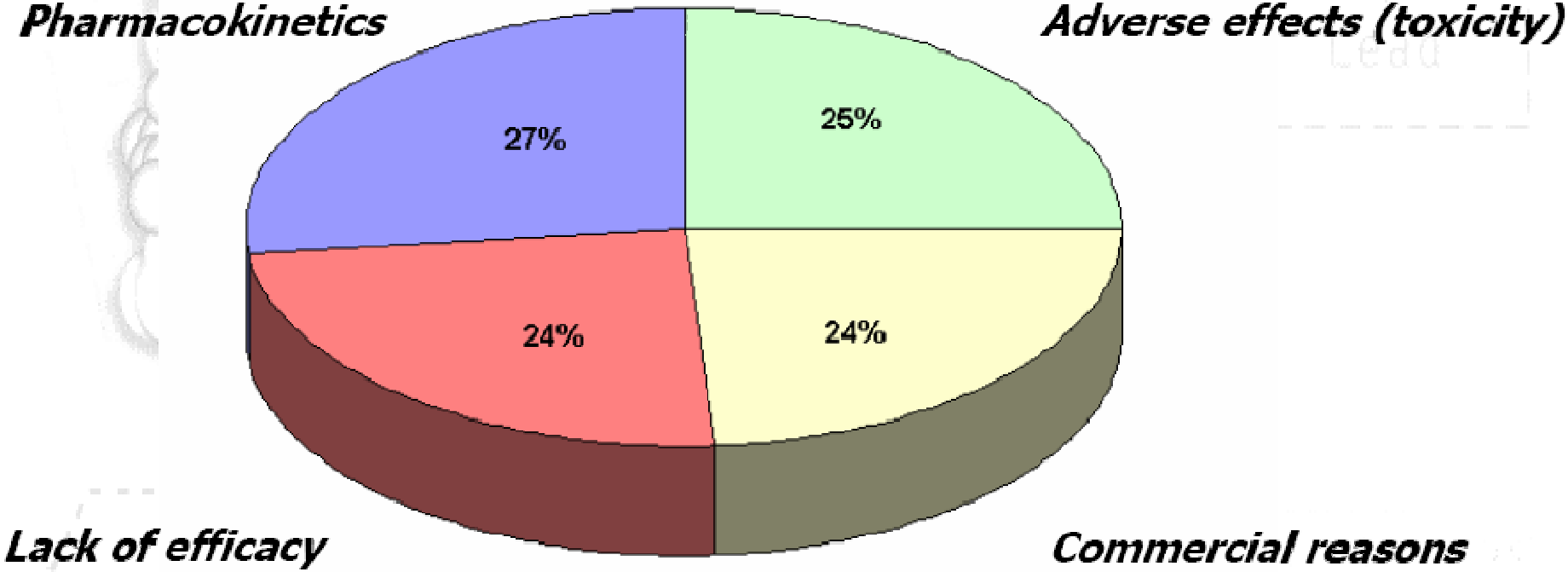
TIME is the worse enemy in drug discovery…



Bringing a new drug to market can take 8-14 years and costs between $400 and $900 million

One of the possible strategy is the analysis of all failures:

B.Abrahamsson et al., in Drug Bioavailability, H.van der Waterbeemd Ed., Vol 18, Wiley-VCH, p495 (2002)



How we can reduce the unfavorable ratio xxx.000 to 1?

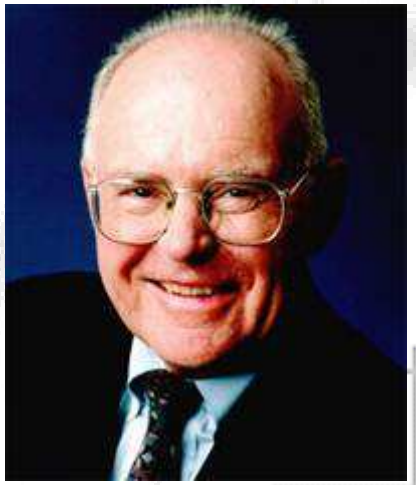
The worthwhile "pharma" goal (dream):

From Lead to Drug in Five Years

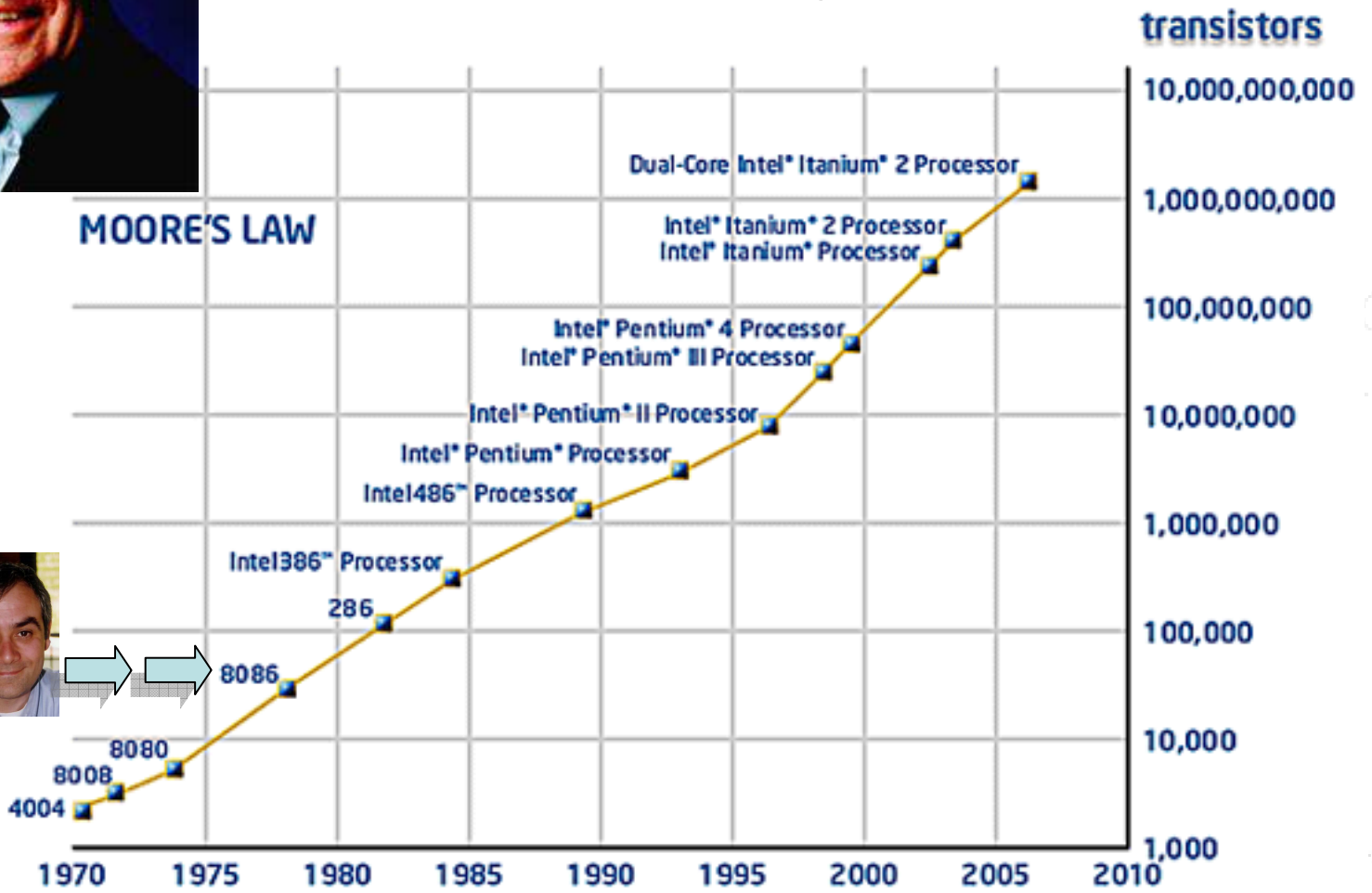
3st International Drug Discovery and Development Summit – December 2-5, 2008, Honolulu, Hi, USA

Priorities:

Lead selection using virtual screening;
In vitro **screening for drug toxicity;**
Solubility and bioavailability;
Animal models: accelerating discovery;
Accelerating clinical studies.



Informatics as synonymous of speed?



Just an example…

Informatics can help us to solve simple procedures in interactive way:

The calculation of a molecular weight is trivial, but the calculations of 4,5 millions of MW is not!

or solve extremely complex procedures:

Such as the calculation of symmetries and energies of molecular orbitals of a chemical structure or ligand-receptor binding energy (ΔG_{bind}, kcal x mol)!

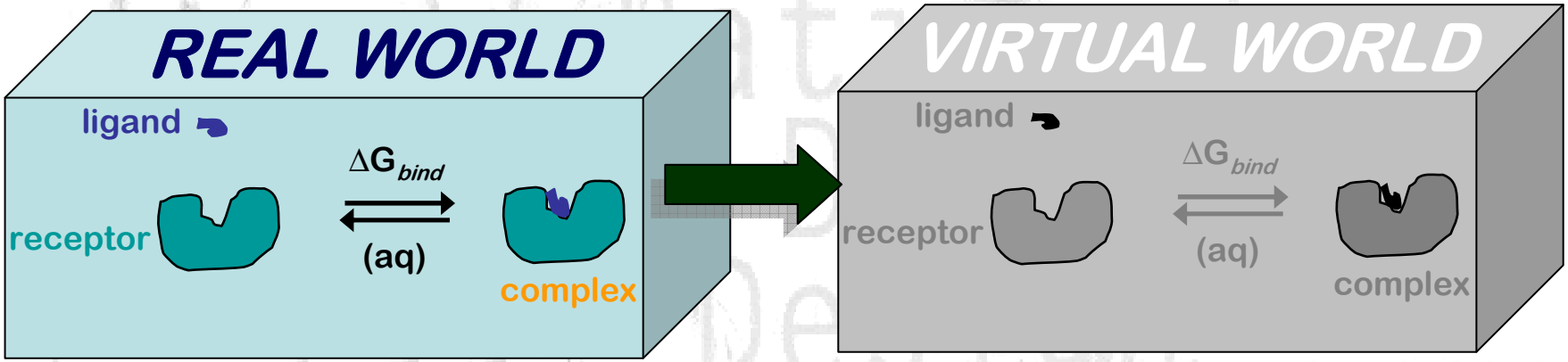
Some costs details:

Experiment Typical Cost per Compound (€)

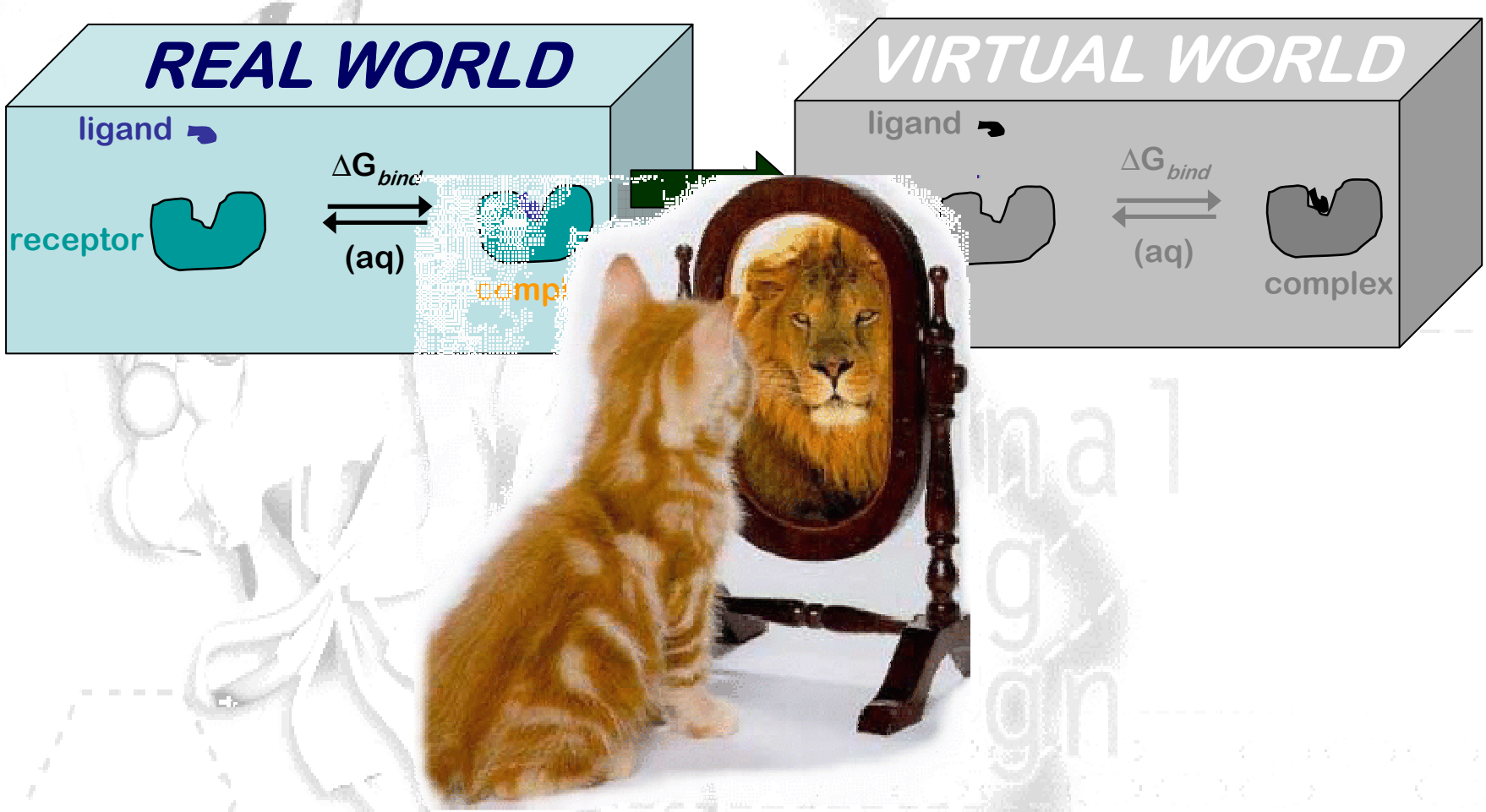
Computer modeling	**7**
Biochemical assay	**270**
Cell culture assay	**2.700**
Rat acute toxicity	**8.100**
Protein crystal structure	**68.000**
Animal efficacy trial	**200.000**
Rat 2-year chronic oral toxicity	**550.000**
Human clinical trial	**3.500.000**

Informatics is the basic science of any virtualization process:

From an informatics point of view, any compu tool is a *virtualization process*: the creation of a virtual version of the real process.

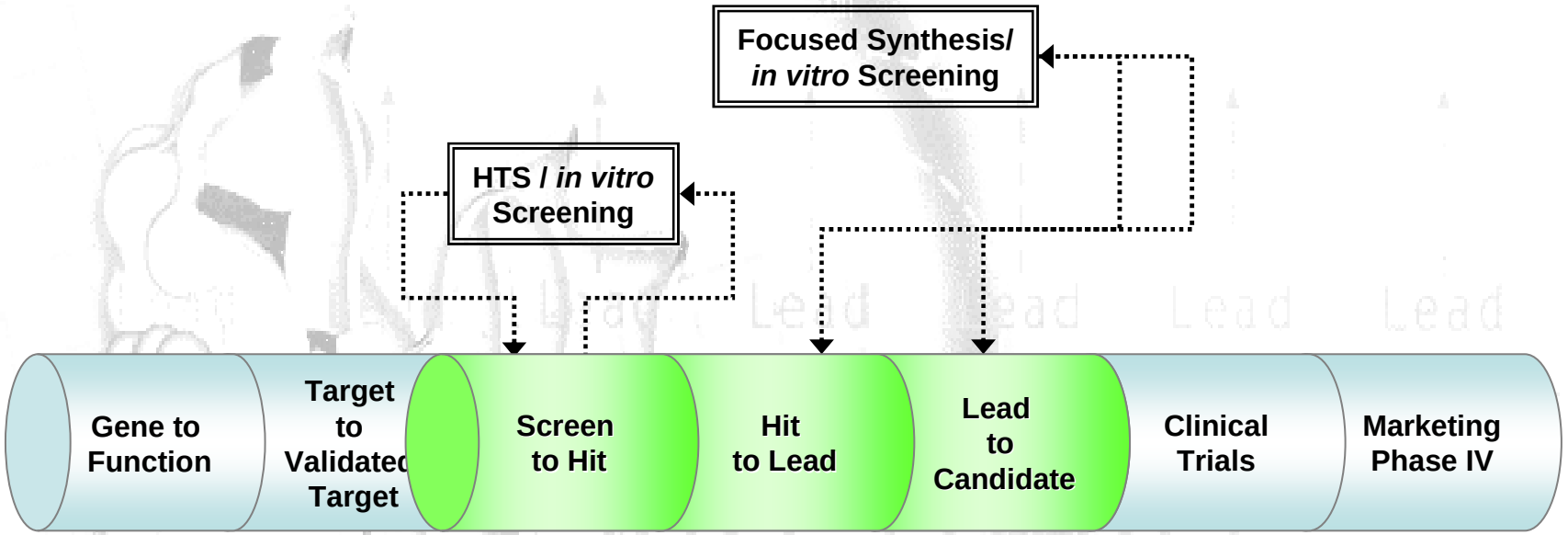


The *accuracy* of this virtualization process is crucial:



But what we need to virtualize in drug discovery process?

Flowcharts in Drug Discovery:



A easy way to define a screening process:

Biological Space

Profiling

Chemical Space

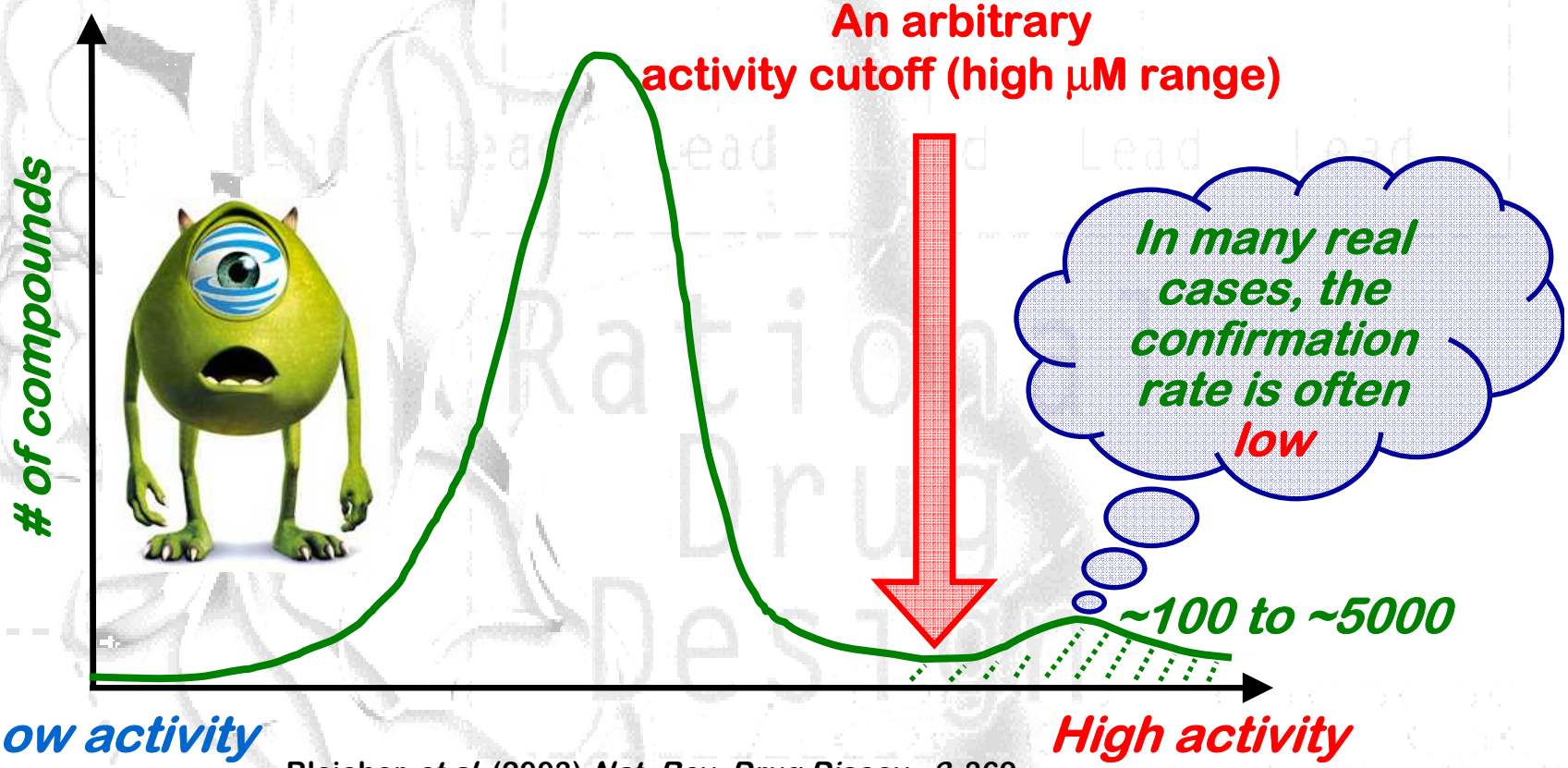
	TargetA	TargetB	TargetC	TargetD	TargetN
Comp.1	$K_1(A)$	$K_1(B)$	$K_1(C)$	$K_1(D)$	$K_1(N)$
Comp.2	$K_2(A)$	$K_2(B)$	$K_2(C)$	$K_2(D)$	$K_2(N)$
Comp.3	$K_3(A)$	$K_3(B)$	$K_3(C)$	$K_3(D)$	$K_3(N)$
Comp.4	$K_4(A)$	$K_4(B)$	$K_4(C)$	$K_4(D)$	$K_4(N)$
Comp.n	$K_n(A)$	$K_n(B)$	$K_n(C)$	$K_n(D)$	$K_n(N)$

Screening

Chemical Proteomics

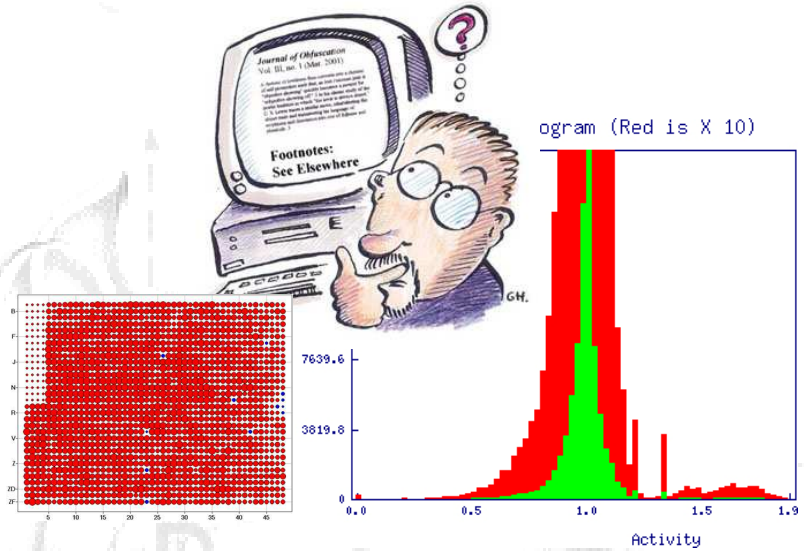
What *hit is…*

HIT: a compound that acts on an assay system.



Bleicher *et al.* (2003) *Nat. Rev. Drug Discov., 2*, 369

the "*hit-to-lead* paradigm": clear the *xxx*.000:1 ratio?



- >1,000,000
- 1,000,000
- 1,000
- 100

- *Initial HTS campaign*
- *Quality control*
- *Primary hit selection*
- *Hit validation*

Bleicher *et al.* (2003) *Nat. Rev. Drug Discov., 2*, 369

Hit definition …

- **hits**
  - active in assay
  - defined and confirmed structures
  - drug-like potential

HTS hits from this database typically show micromolar activity with a median "pPotency" of 6. The median molecular mass and lipophilicity (logP) was 359 Da and 3.8, respectively.

What about costs?

Experiment Typical Cost per Compound (€)

Computer modeling	7
Biochemical assay	270
Cell culture assay	2.700
Rat acute toxicity	8.100
Protein crystal structure	68.000
Animal efficacy trial	200.000
Rat 2-year chronic oral toxicity	550.000
Human clinical trial	3.500.000

.... do you understand how crucial could be the prediction of possible new hits?

Hit to lead …

- **leads**
  - potency established
  - selectivity/specificity
  - Mechanism of action (MOA) established
  - *in vivo* efficacy
  - ADME/Tox
  - pharmaceutically acceptable

Well… now it's trivial

Experiment Typical Cost per Compound (€)

Computer modeling	7
Biochemical assay	270
Cell culture assay	2.700
Rat acute toxicity	8.100
Protein crystal structure	68.000
Animal efficacy trial	200.000
Rat 2-year chronic oral toxicity	550.000
Human clinical trial	3.500.000

…. do you understand how magic could be the prediction of possible new leads?

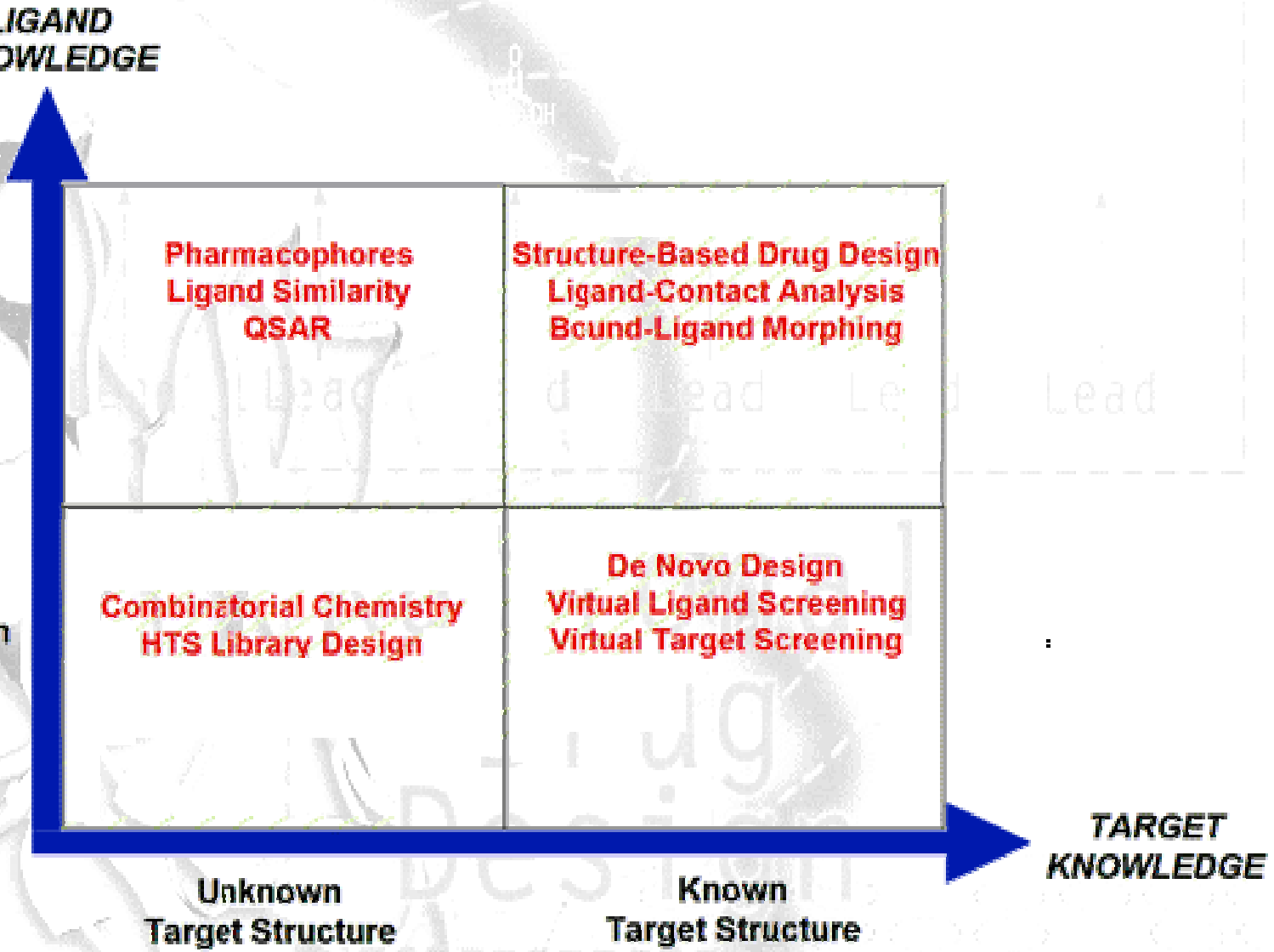
We surely need informatics but:

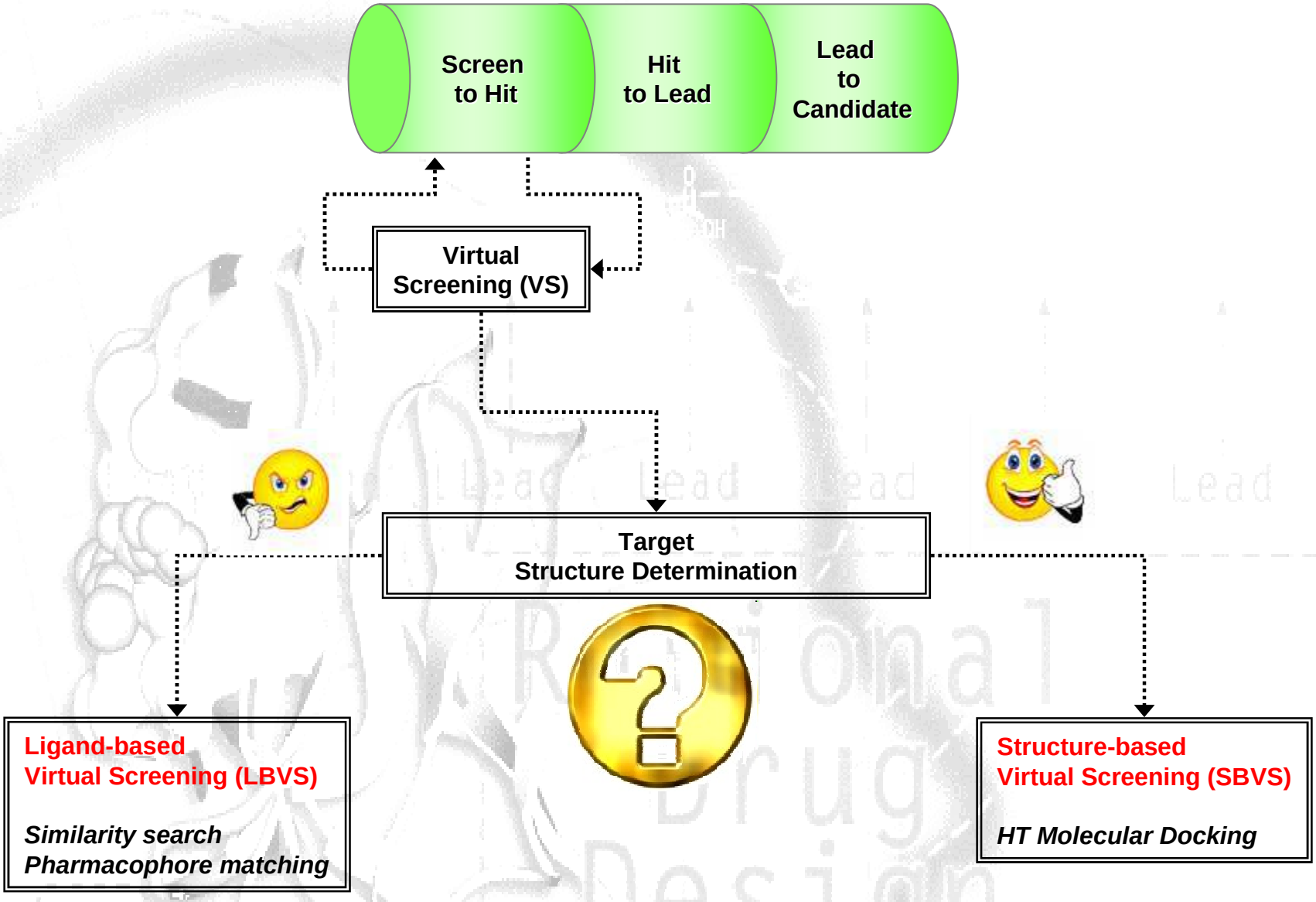
- ✓ Ligand-based Drug Design (LBDD)
- Structure-based Drug Design (SBDD)

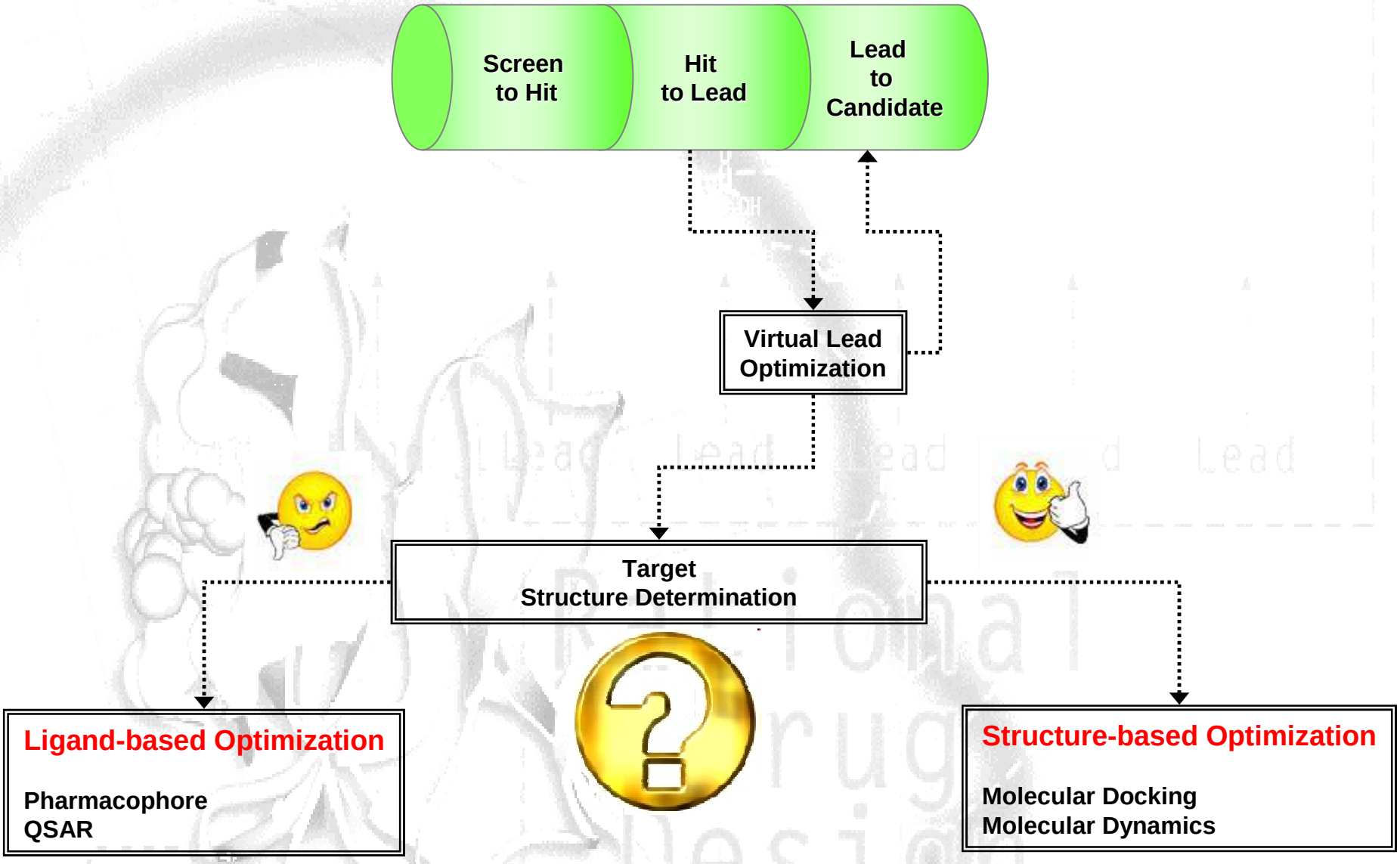
xxx.000 to 1?

Is this ratio really acceptable for a pharma company?

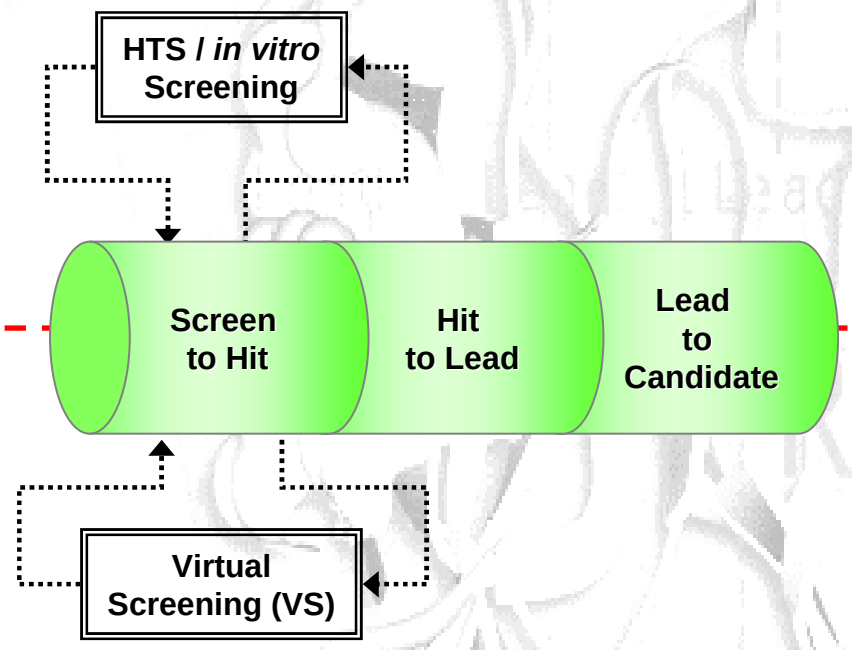
We surely need informatics but:




Screen
to Hit
Hit
to Lead
Lead
to
Candidate
Virtual
Screening (VS)
Target
Structure Determination
Ligand-based
Virtual Screening (LBVS)
Similarity search
Pharmacophore matching
Structure-based
Virtual Screening (SBVS)
HT Molecular Docking


Screen to Hit
Hit to Lead
Lead to Candidate
Virtual Lead Optimization
Target Structure Determination
Ligand-based Optimization
Pharmacophore
QSAR
Structure-based Optimization
Molecular Docking
Molecular Dynamics

Before starting: lesson 0

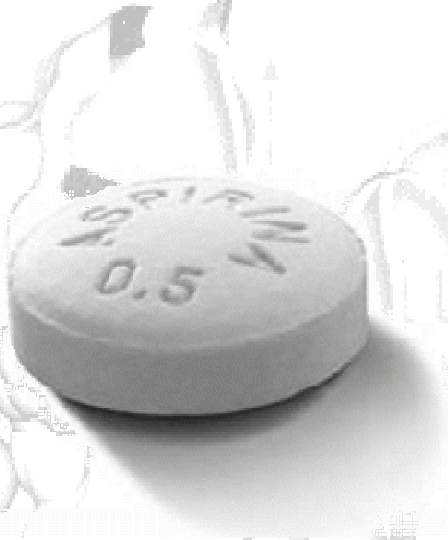


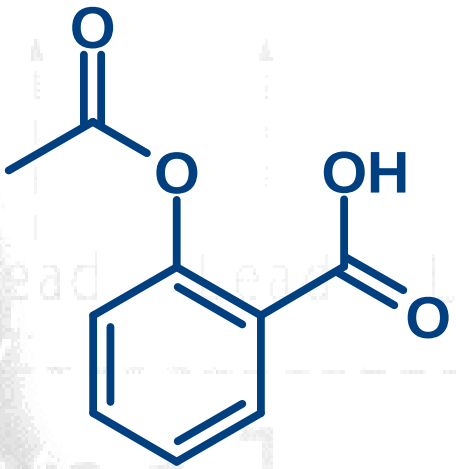
- **hits**
  - active in assay
  - defined and confirmed structures
  - drug-like potential

REAL

VIRTUAL

We do NEVER forget:





This is a chemical

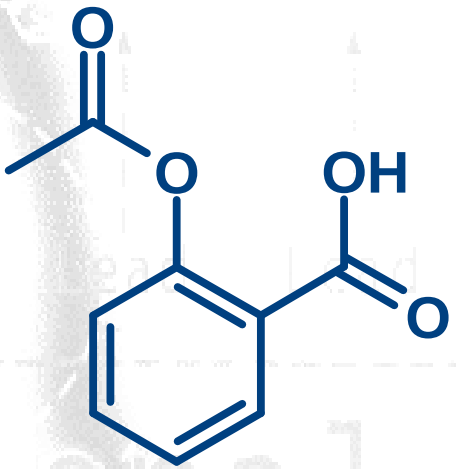
This is one of its possible chemical representations

.... follow me in this logic comparison:

Real world	*Virtual world*
Chemical Compound (CC)	
⇓	
Chemical Structure (CS)	**Numerical representazions of CS**
⇓	⇓
Chemical Properties (CP)	**Molecular Descriptors (MD)**

With how many chemical representations we can deal?

C9H8O4



…again the salicylic acid?

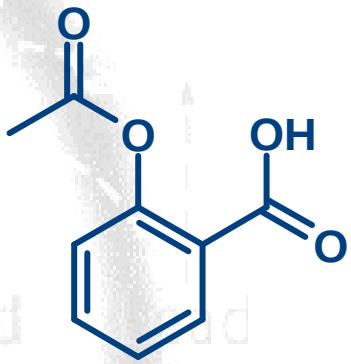
The crucial informatics differences:

C9H8O4

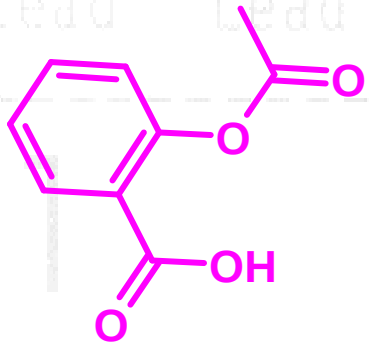
… this is a simple *string* (sequence) of alphanumeric characters and it is very easy to manage… informatically!!!

Just a simple example: are these two representations identical?

C9H8O4



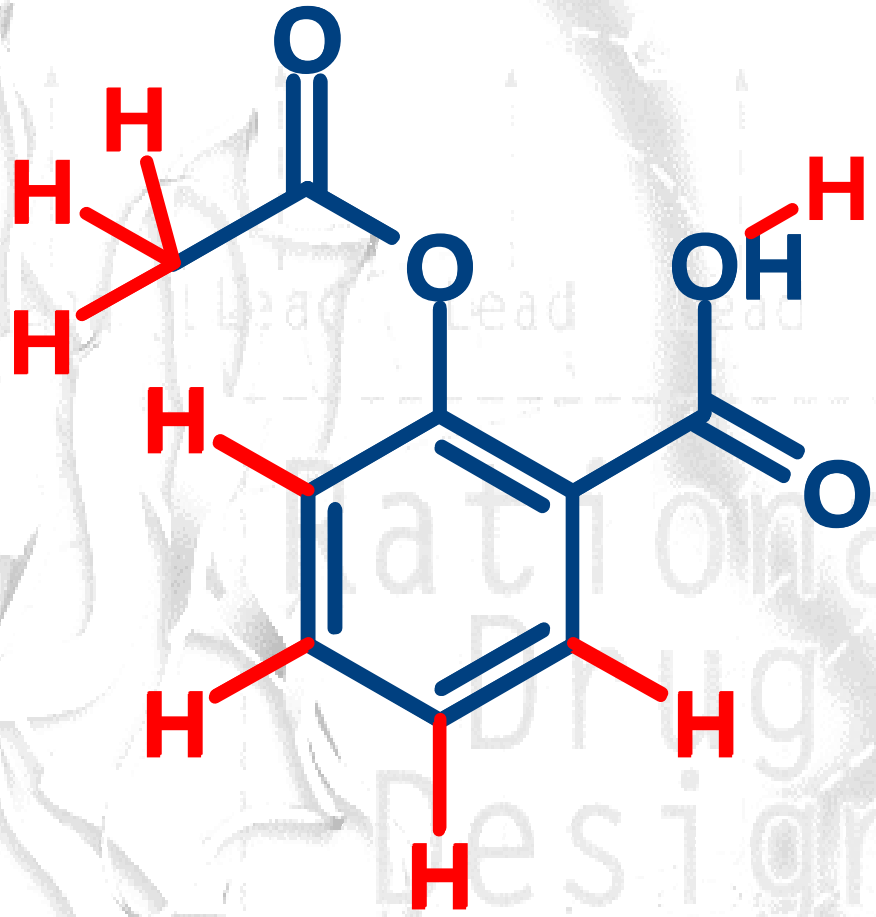
C9H8O4



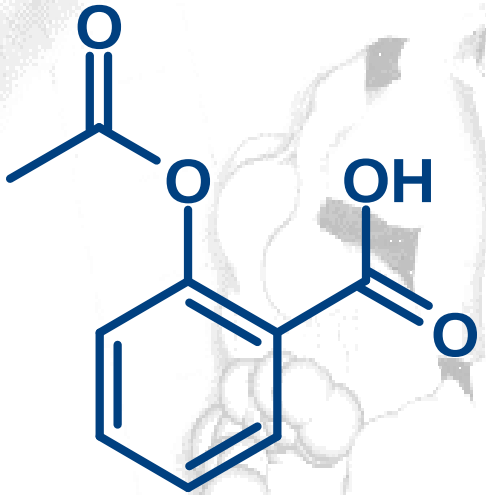
Time of answer (sec):

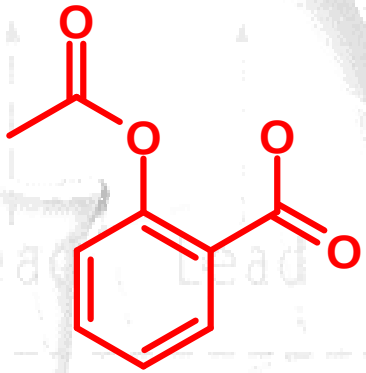
Time of answer (sec):

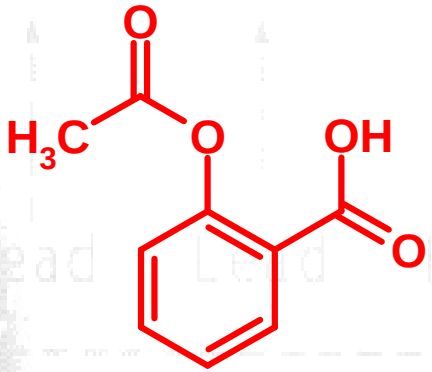
Be careful to the chemical *slang*....

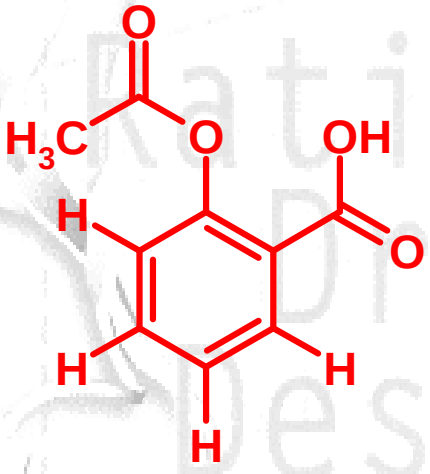


Remember, all of these are not identical… informatically specking!



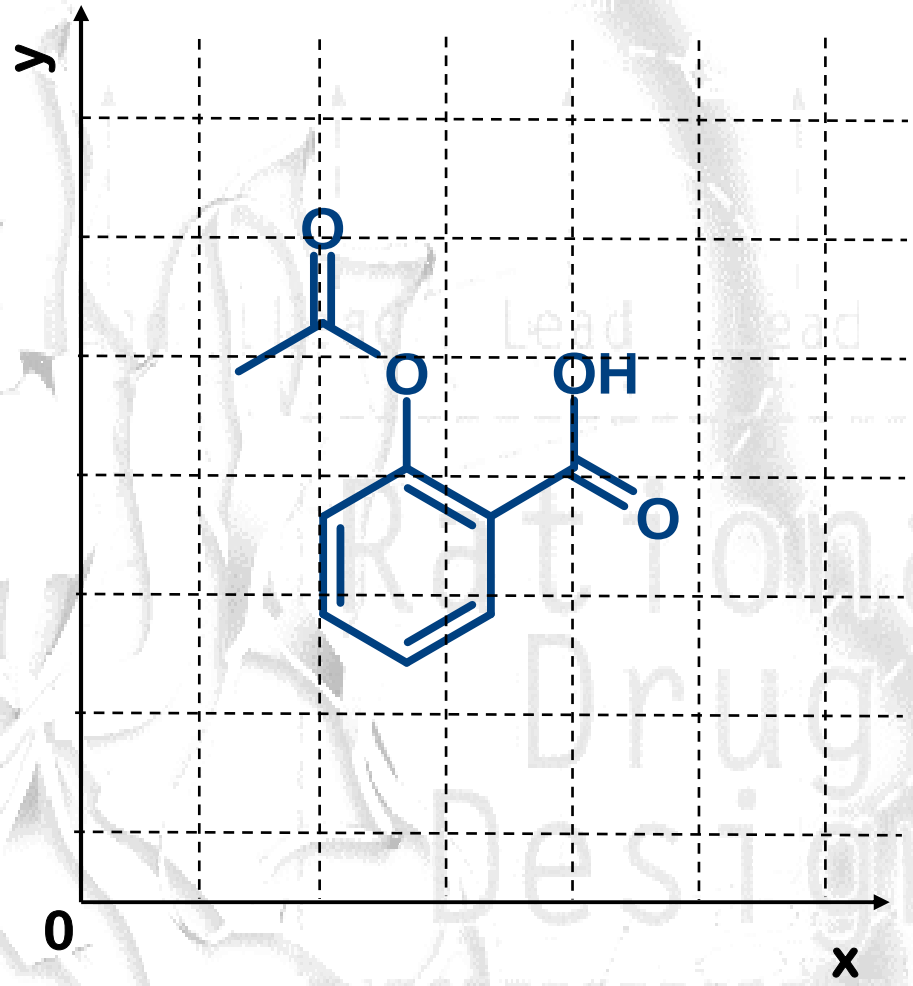






…

Here is the informatics anatomy of this representation:



Numero di atomi

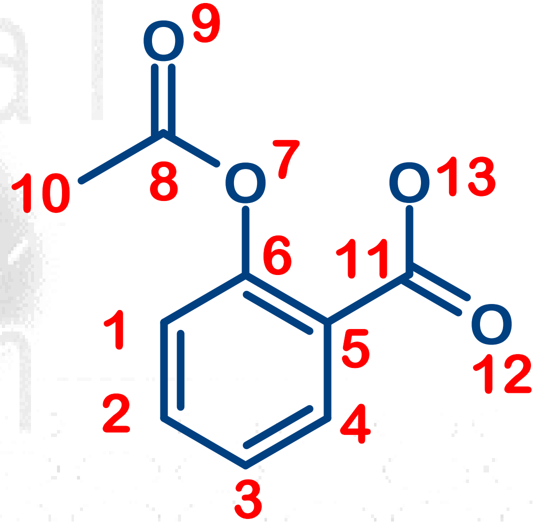
Numero di legami

tipo di atomo

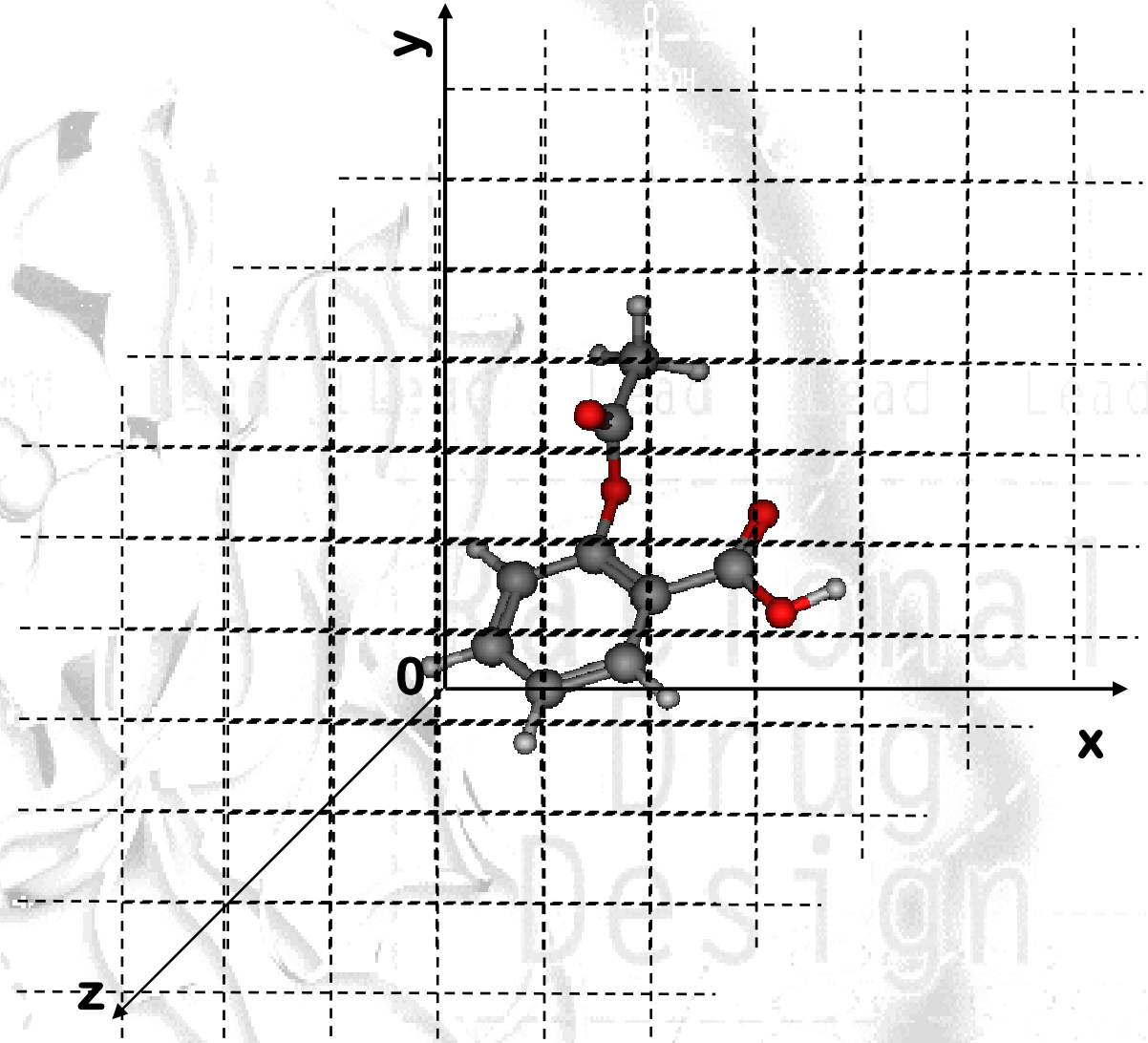
```
 ISIS

13 13
   1.0986  -13.6500    0.0000 C
   1.0974  -14.4773    0.0000 C
   1.8122  -14.8902    0.0000 C
   2.5287  -14.4769    0.0000 C
   2.5258  -13.6463    0.0000 C
   1.8104  -13.2372    0.0000 C
   1.8080  -12.4122    0.0000 O
   1.0923  -12.0018    0.0000 C
   1.0898  -11.1768    0.0000 O
   0.3790  -12.4165    0.0000 C
   3.2387  -13.2311    0.0000 C
   3.9547  -13.6409    0.0000 O
   3.2356  -12.4061    0.0000 O
  1  2  2
  6  7  1
  3  4  2
  7  8  1
  8  9  2
  4  5  1
  8 10  1
  2  3  1
  5 11  1
  5  6  2
 11 12  2
  6  1  1
 11 13  1
M  END
```

Coordinate: x y z

Il primo legame è tra l'atomo 1 e l'atomo 2 ed è di ordine 2 (doppio legame).



Now, this is perfect understandable!



```
  MOE2007     3D
                                                  
 21 21
    1.1522    0.5113   -0.0537 C
    0.2183    1.2057   -0.8346 C
    0.5665    2.3445   -1.5679 O
   -1.0985    0.7524   -0.9667 C
   -1.8050    1.2020   -1.5911 H
   -1.5038   -0.3936   -0.2868 C
   -2.5276   -0.7473   -0.3775 H
   -0.5938   -1.0870    0.5099 C
   -0.9101   -1.9816    1.0411 H
    0.7269   -0.6400    0.6244 C
    1.4145   -1.2086    1.2469 H
    0.5056    3.5095   -0.8113 C
    0.9559    4.6611   -1.6596 C
    0.9409    5.5777   -1.0629 H
    0.2773    4.7842   -2.5073 H
    1.9783    4.4900   -2.0055 H
    0.1372    3.6103    0.3517 O
    2.5419    0.9700    0.0441 C
    3.0443    1.9108   -0.5504 O
    3.2755    0.2254    0.8972 O
    4.1677    0.6310    0.8800 H
  1 10  2
  1  2  1
  1 18  1
  ...
  20 21  1
M  END
```

Numero di atomi

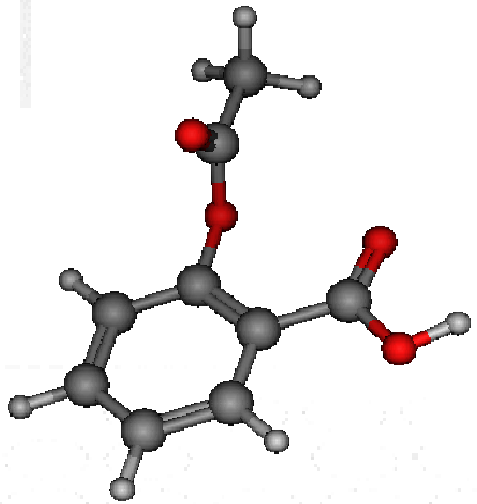
Numero di legami

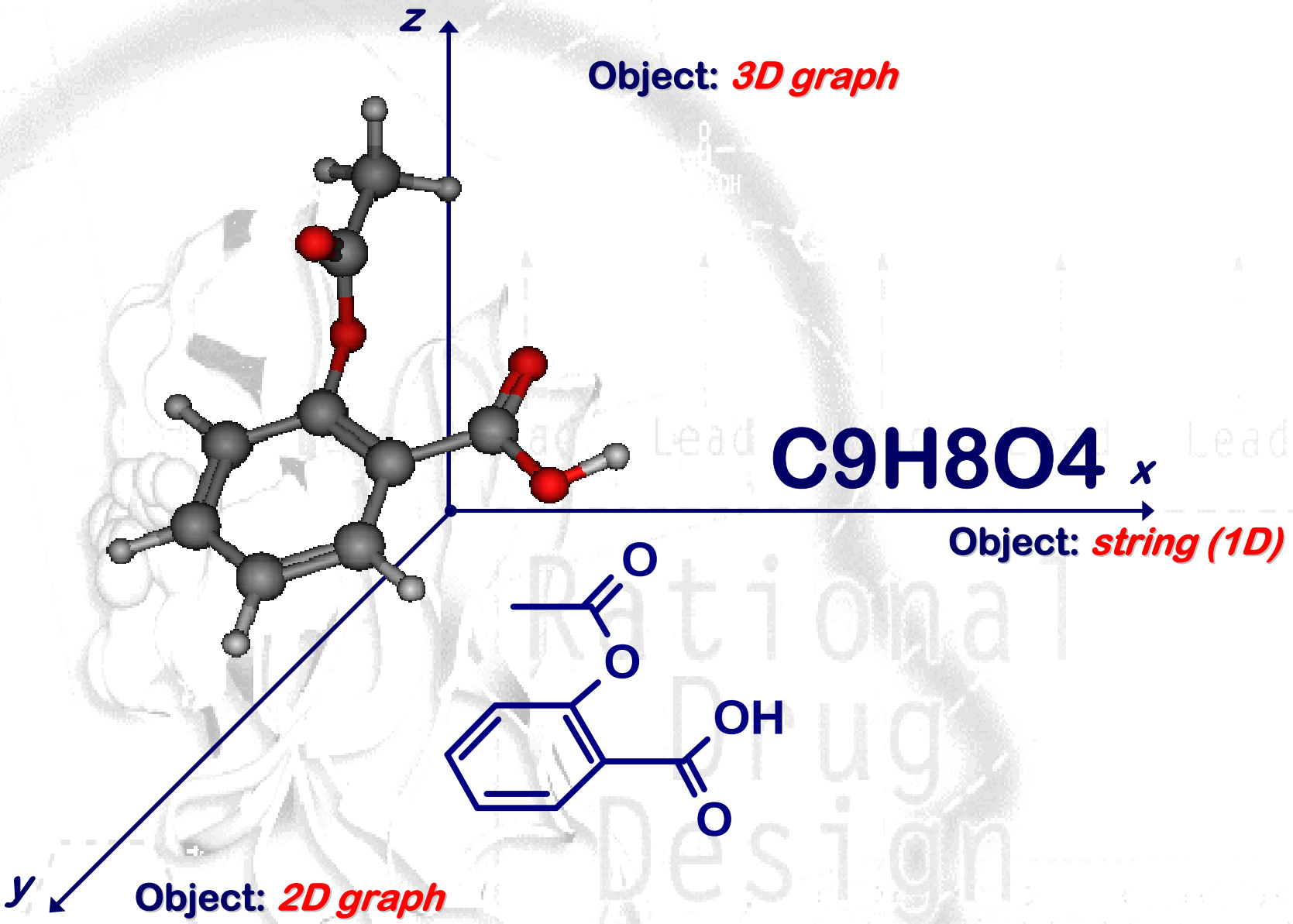
tipo di atomo

Coordinate:

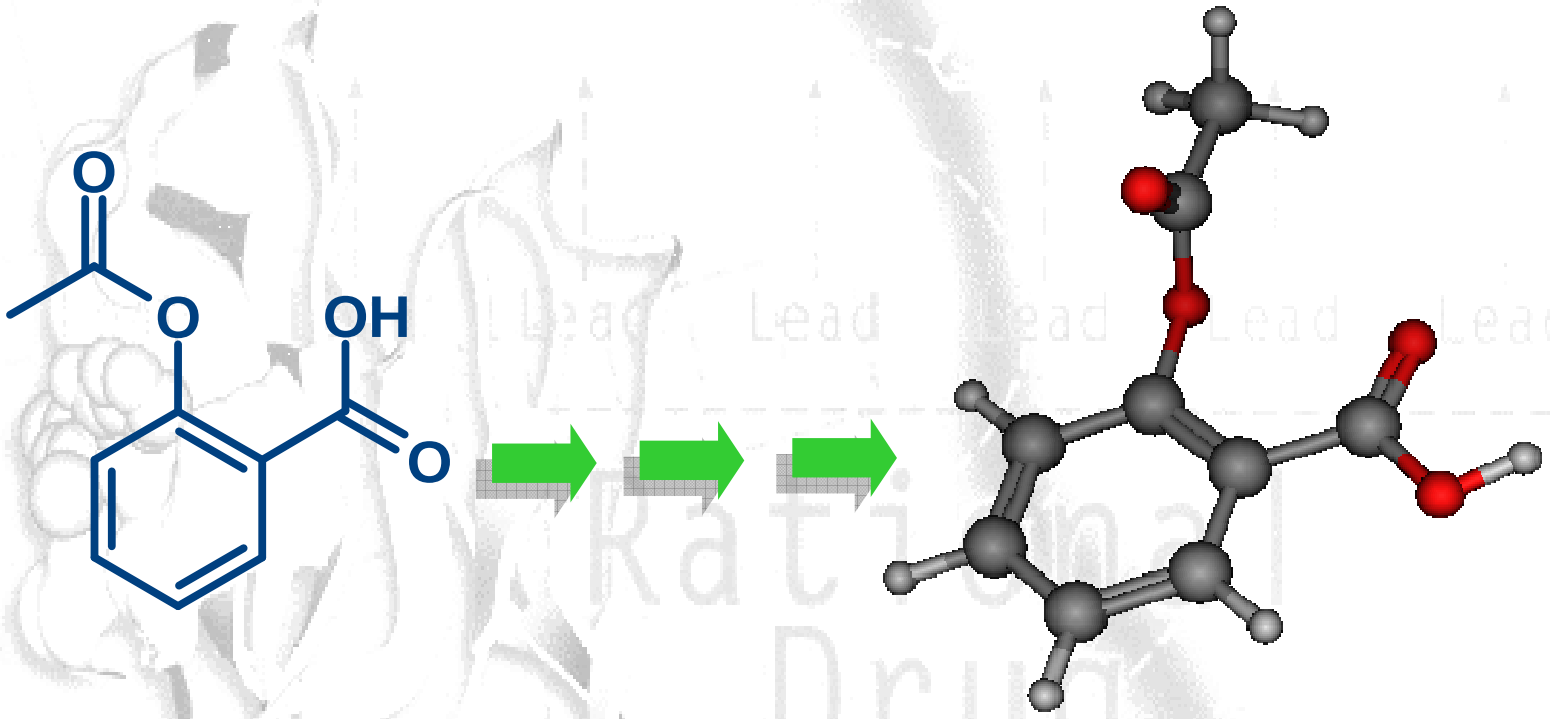
x **y** **z**

Il primo legame è tra l'atomo 1 e l'atomo 10 ed è di ordine 2 (doppio legame).

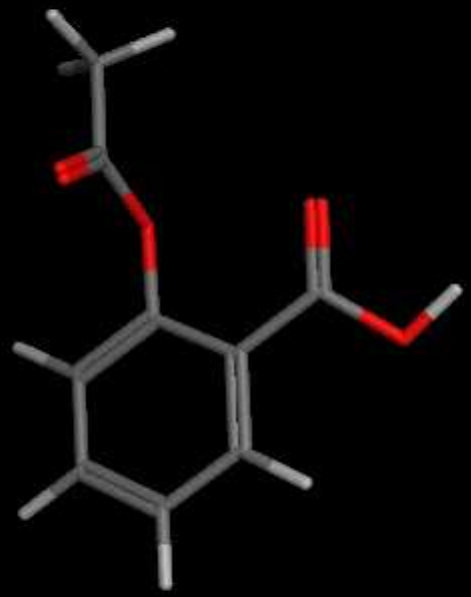



z
Object: 3D graph
C9H8O4 x
Object: string (1D)
O
O
OH
O
y
Object: 2D graph

Are we fully satisfied?



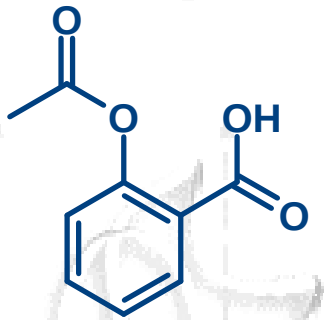
What we are really missing...

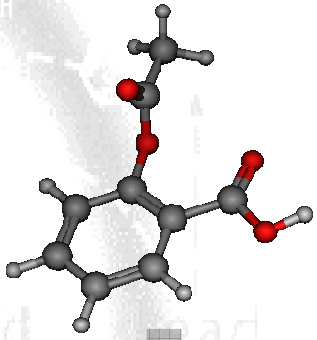


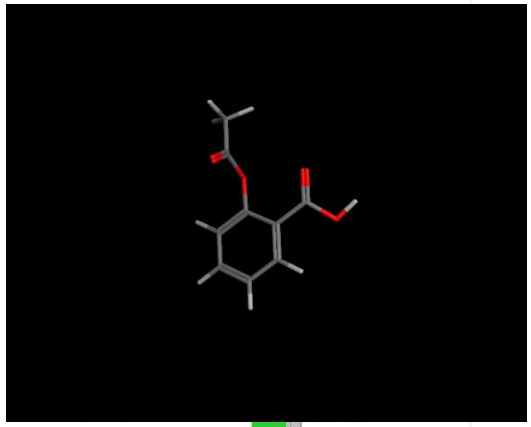
… the illusion of time!

Another important informatics difference!

C9H8O4







6 byte

1.299 byte

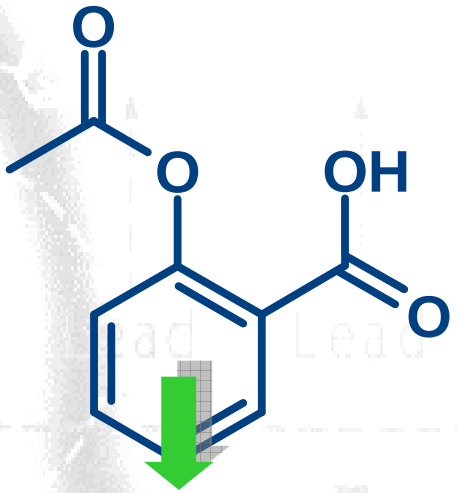
2.051 byte

73.728 byte

1 byte = 8 bit (1 bit = 0 o 1, true o false)

Combining business with pleasure ?

C9H8O4

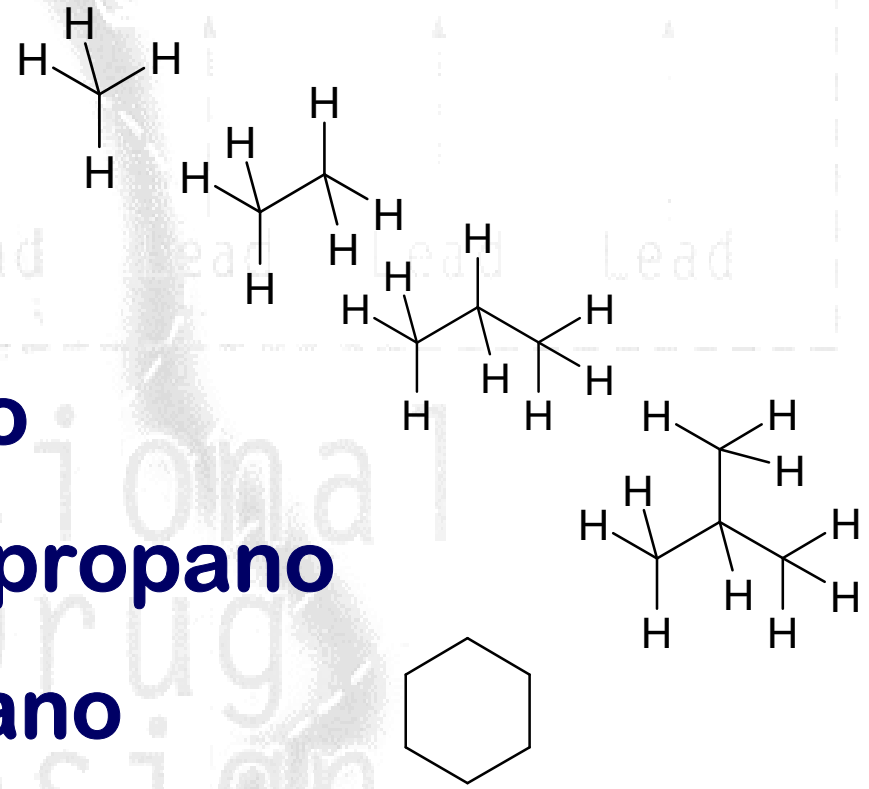


SMILES (Simplified Molecular Input Line Entry Specification)

SMILES (Simplified Molecular Input Line Entry Specification)

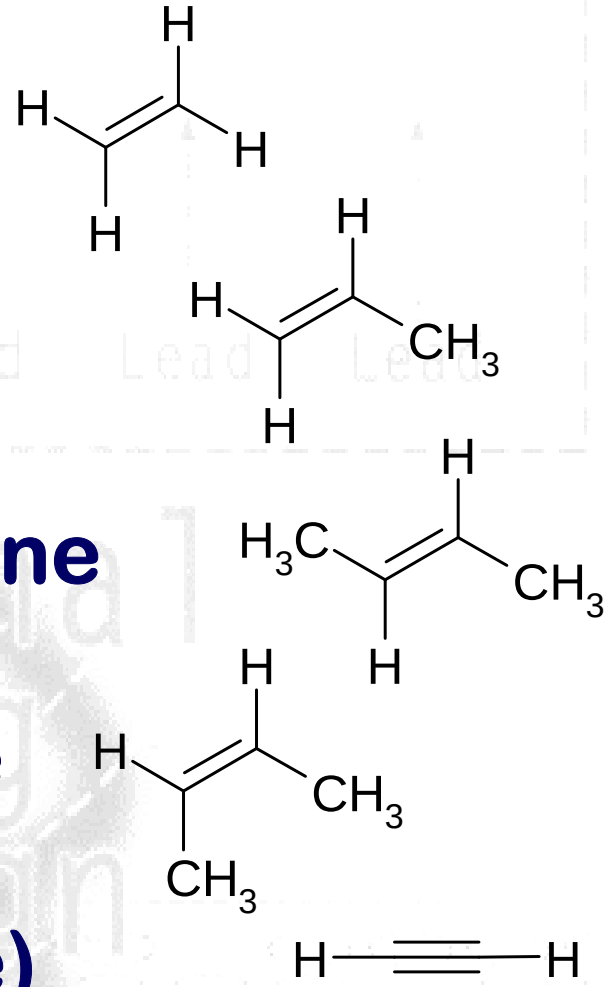
Some SMILES rules:

C	metano
CC	etano
CCC	propano
CC(C)C	2-metil-propano
C1CCCCC1	cicloesano



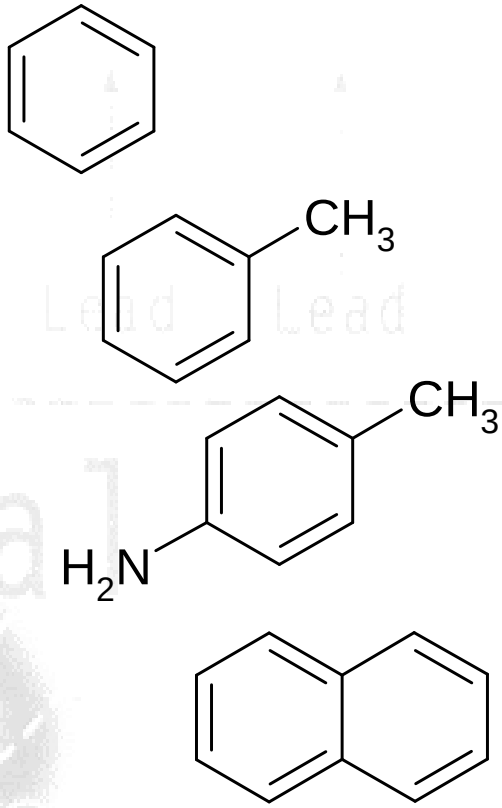
SMILES (Simplified Molecular Input Line Entry Specification)

C=C	etilene
C=CC	propene
C/C=C/C	*trans* (E)-2-butene
C/C=C\C	*cis* (Z)-2-butene
C#C	etino (acetilene)



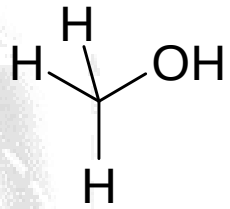
SMILES (Simplified Molecular Input Line Entry Specification)

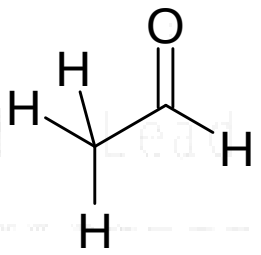
c1ccccc1	benzene
c1ccccc1C	toluene
c1ccc(N)cc1C	4-metil-anilina
c12c(cccc1)cccc2	naftalene

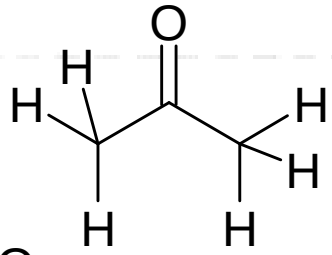


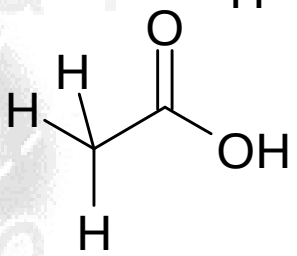
SMILES (Simplified Molecular Input Line Entry Specification)

CCO	etanolo
CC=O	etanale
CC(=O)C	acetone
CC(=O)O	acido acetico





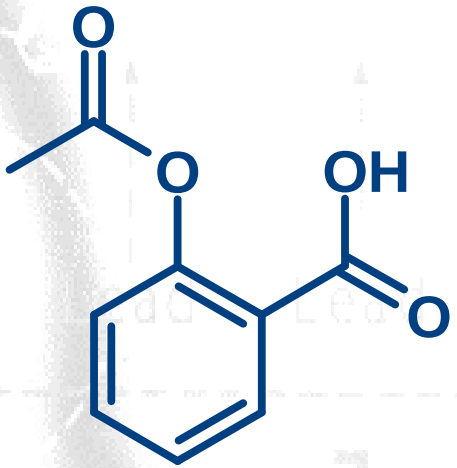




SMILES (Simplified Molecular Input Line Entry Specification)

reassuming:

C9H8O4

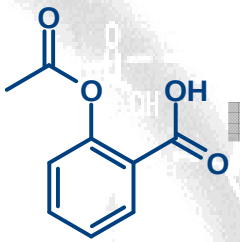


OC(=O)c1ccccc1(OC(=O)C)

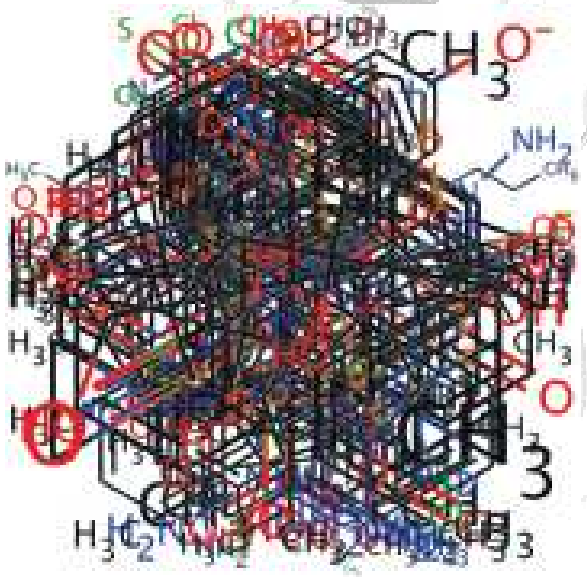
Three faces of the same medal!!!

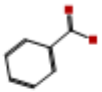
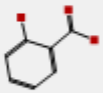
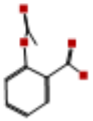
Searching in a database is now very easy!

Query

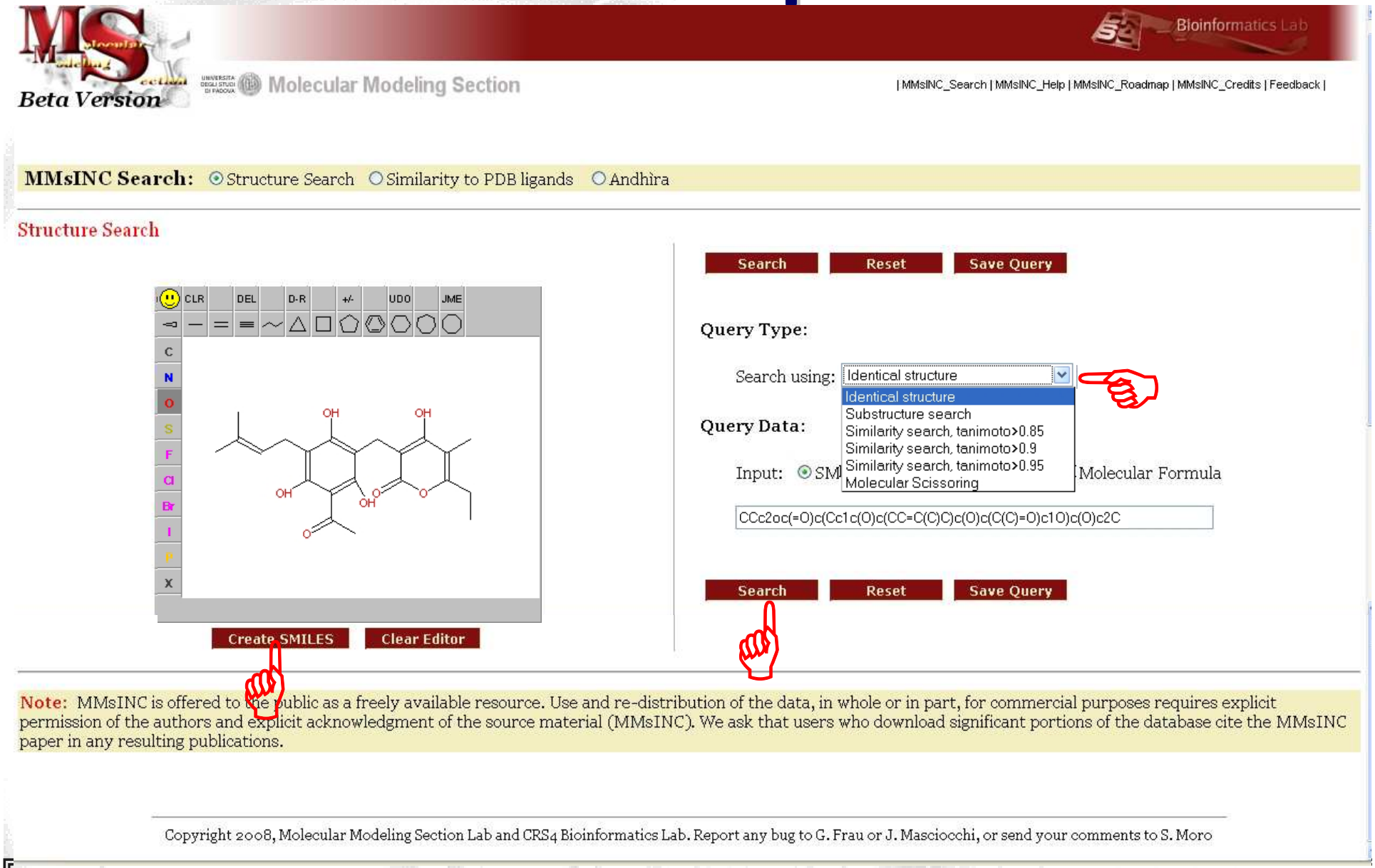


OC(=O)c1ccccc1(OC(=O)C)



	F.BRUTA	mol	SMILES	
1	C7H6O2		OC(=O)c1ccccc1	☹
2	C7H6O3		OC(=O)c1ccccc1(O)	☹
3	C9H8O4		OC(=O)c1ccccc1(OC(=O)C)	☺

MMsINC web-platform:



Arzanol

MMsInc code: MMs03967057

Type: Neutral
Formula: $C_{22}H_{26}O_7$
SMILES: O1C(=O)C(Cc2c(O)c(C(=O)C)c(O)c(CC=C(C)C)c2O)=C(O)C(C)=C1CC
InChI: InChI=1/C22H26O7/c1-6-16-11(4)18(24)15(22(28)29-16)9-14-19(25)13(8-7-10(2)3)20(26)17(12(5)23)21(14)27/h7,24-27H,6,8-9H2,1-5H3

Sdf File: MMs03967057.sdf

Similarity to PDB ligands

search links for this molecule: PubChem Compound Andhira

Ions/Tautometers related molecules: no related molecules available.

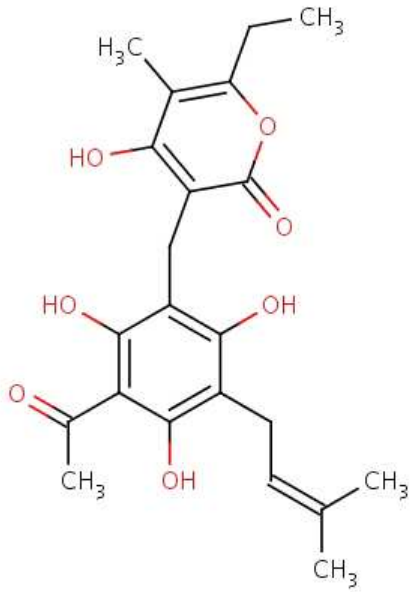
Arzanol

MMsInc code: MMs03967057

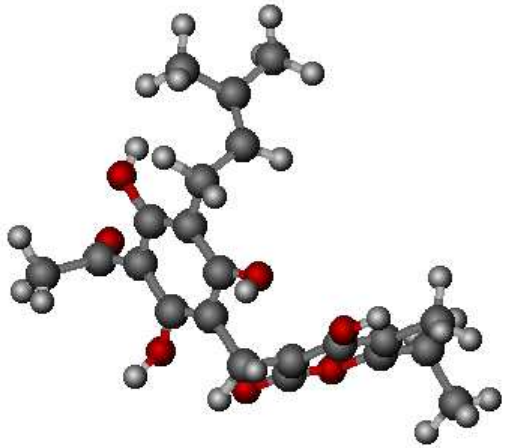
Type: Neutral
Formula: $C_{22}H_{26}O_7$
SMILES: O1C(=O)C(Cc2c(O)c(C(=O)C)c(O)c(CC=C(C)C)c2O)=C(O)C(C)=C1CC
InChI: InChI=1/C22H26O7/c1-6-16-11(4)18(24)15(22(28)29-16)9-14-19(25)13(8-7-10(2)3)20(26)17(12(5)23)21(14)27/h7,24-27H,6,8-9H2,1-5H3

Sdf File: MMs03967057.sdf

Similarity to PDB ligands



download 2D Mol File



download 3D Mol File

Potential Energy (e)
Epot(MMFF94)=107.71 kcal/mol

Beta Version

Molecular Modeling Section

Bioinformatics Lab

| MMsINC_Search | MMsINC_Help | MMsINC_Roadmap | MMsINC_Credits | Feedback |

RCSB PDB PROTEIN DATA BANK

A MEMBER OF THE PDB

An Information Portal to Biological Macromolecular Structures

As of Tuesday Jun 03, 2008 there are 51155 Structures | PDB Statistics

CONTACT US | HELP | PRINT PAGE PDB ID or keyword Author Site Search | Advanced Search

Home | Search | Structure

- 1GCZ
- Download Files
- FASTA Sequence
- Download Original Files
- Display Files
- Display Molecule
- Structural Reports
- External Links
- Structure Analysis
- Help

Quick Tips:

When exploring a structure, select *Structure Analysis* and then *Geometry* from the left menu to view a **Ramachandran Plot**.

Are you missing data updates? The PDB archive has moved to ftp://ftp.wwpdb.org. For more information click here.

Help | Structure Summary | Biology & Chemistry | Materials & Methods | Sequence Details | Geometry

1gcz

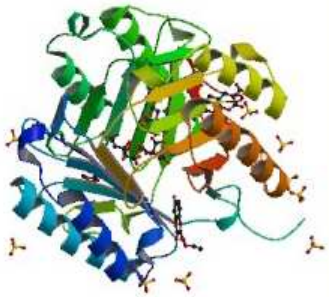
DOI 10.2210/pdb1gcz/pdb

Red - Derived Information

Title	MACROPHAGE MIGRATION INHIBITORY FACTOR (MIF) COMPLEXED WITH INHIBITOR.
Authors	Katayama, N., Kurihara, H.
Primary Citation	Orita, M., Yamamoto, S., Katayama, N., Aoki, M., Takayama, K., Yamagiwa, Y., Seki, N., Suzuki, H., Kurihara, H., Sakashita, H., Takeuchi, M., Fujita, S., Yamada, T., Tanaka, A. (2001) Coumarin and chromen-4-one analogues as tautomerase inhibitors of macrophage migration inhibitory factor: discovery and X-ray crystallography. *J.Med.Chem.* ***44:*** *540-547* [Abstract]
History	Deposition 2000-08-24 Release 2001-02-21
Experimental Method	Type X-RAY DIFFRACTION Data [EDS]

Images and Visualization

<< Biological Molecule >>



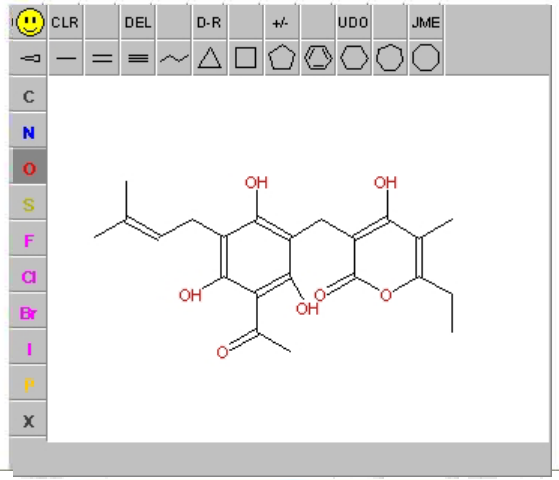
Display Options

KiNG

Jmol

MMsINC Search: ◉ Structure Search ○ Similarity to PDB ligands ○ Andhìra

Structure Search



Create SMILES | Clear Editor

Search | Reset | Save Query

Query Type:

Search using: Identical structure

- Identical structure
- Substructure search
- Similarity search, tanimoto>0.85
- Similarity search, tanimoto>0.9
- Similarity search, tanimoto>0.95
- Molecular Scissoring

Query Data:

Input: ◉ SMI Molecular Formula

CCc2oc(=O)c(Cc1c(O)c(CC=C(C)C)c(O)c(C(C)=O)c1O)c(O)c2C

Molecular Descriptors:

⊟ Physical Properties

Molecular Weight (MW) from		to	
logS from		to	
SlogP from		to	
Reactive groups	Select a value		

+ Topological Properties
+ Surface and Volume Properties
+ Pharmacophoric Properties
+ Drug- and Lead-like Properties

Search | Reset | Save Query

Molecule found with a Tanimoto Coefficent superior or equal to 0.85

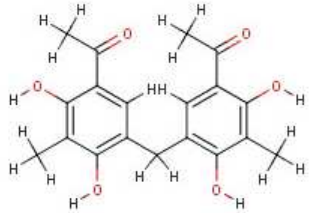
Items found 1 - 20 of 200

1 of 10 Go to Page

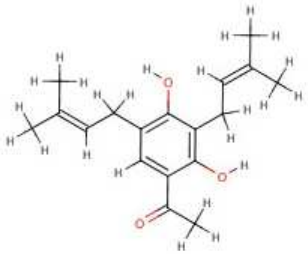
Next >>

Select All | Deselect All

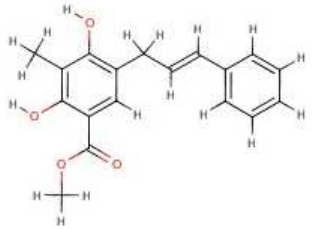
add to cart [cart empty]



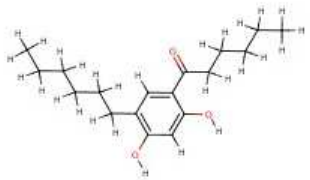
☐ MMs02176936
tanimoto score: 0.900826



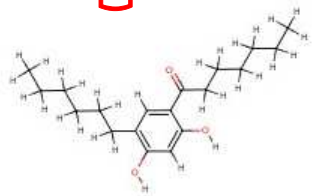
☐ MMs02690319
tanimoto score: 0.89916



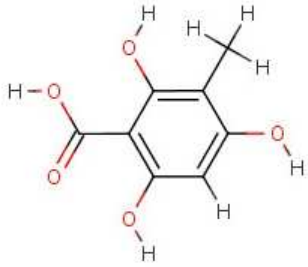
☐ MMs02670131
tanimoto score: 0.896825



☐ MMs03262991
tanimoto score: 0.891667



☐ MMs03262992
tanimoto score: 0.891667



☐ MMs02254042
tanimoto score: 0.890756

☐ MMs03378492
tanimoto score: 0.887931

MDPI-ZINC03843495 **MMsInc code: MMs02176936**

Type: Neutral
Formula: $C_{19}H_{20}O_6$
SMILES: Oc1c(C)c(O)c(cc1Cc1cc(C(=O)C)c(O)c(C)c1O)C(=O)C
InChI: InChI=1/C19H20O6/c1-8-16(22)12(6-14(10(3)20)18(8)24)5-13-7-11(4)21)19(25)9(2)17(13)23/h6-7,22-25H,5H2,1-4H3

Sdf File: MMs02176936.sdf

NCBI PubChem Compound

PubChem Compound GO

PubMed | Entrez | Structure | PubChem | Help

PubChem » Compound Summary

ZINC03843495 - Compound Summary (CID: 2748081)

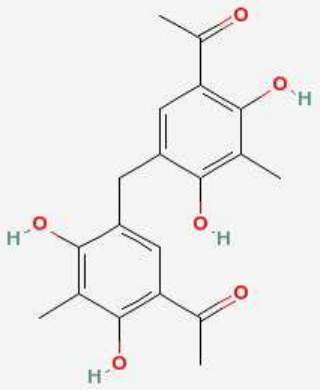
Table of Contents
- Synonyms
- Properties
- Descriptors
- Compound Info
- Substance Info
  - Substance Category
- Exports

Depositor-Supplied Synonyms:

ZINC03843495

Properties Computed from Structure:

Molecular Weight	344.3585 [g/mol]
Molecular Formula	$C_{19}H_{20}O_6$
XLogP	1.5
H-Bond Donor	4



Compound ID	2748081
Molecular Weight	344.3585 [g/mol]
Molecular Formula	$C_{19}H_{20}O_6$
XLogP	1.5
H-Bond Donor	4
H-Bond Acceptor	6

Beta Version

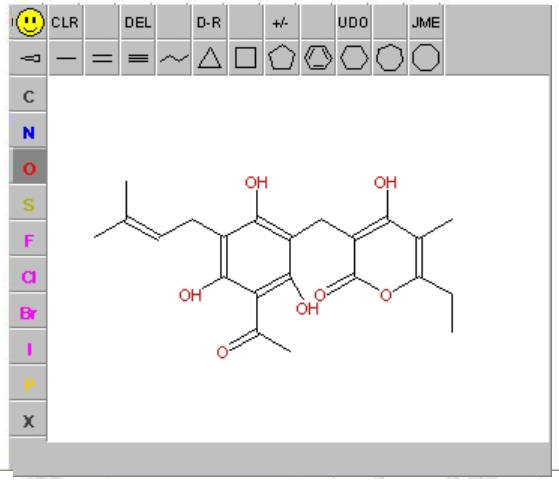
Molecular Modeling Section

Bioinformatics Lab

| MMsINC_Search | MMsINC_Help | MMsINC_Roadmap | MMsINC_Credits | Feedback |

MMsINC Search: ◉ Structure Search ○ Similarity to PDB ligands ○ Andhìra

Structure Search



Create SMILES | Clear Editor

Search | Reset | Save Query

Query Type:

Search using: Identical structure

- Identical structure
- Substructure search
- Similarity search, tanimoto>0.85
- Similarity search, tanimoto>0.9
- Similarity search, tanimoto>0.95
- Molecular Scissoring

Query Data:

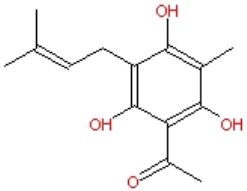
Input: ◉ SMI Molecular Formula

CCc2oc(=O)c(Cc1c(O)c(CC=C(C)C)c(O)c(C(C)=O)c1O)c(O)c2C

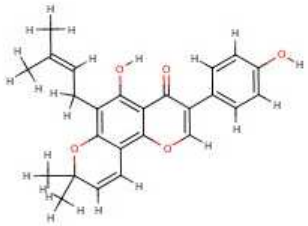
Search | Reset | Save Query

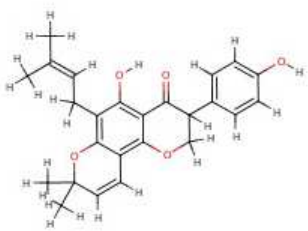
Molecule result, page: 1 Items found 1 - 20 of 75

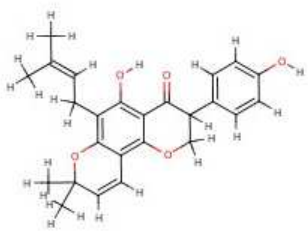
1 of 4 Go to Page
Next >>

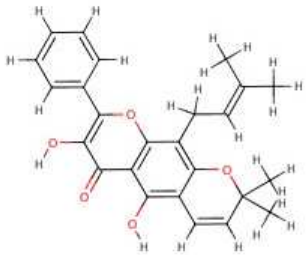


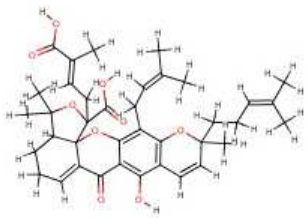
Select All Deselect All add to cart [cart empty]

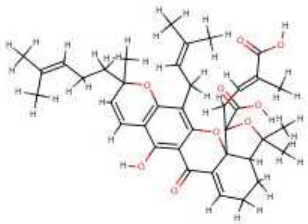

MMs00138587

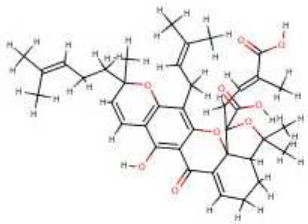

MMs01790323

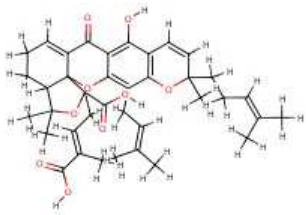

MMs01790322


MMs02188021









Beta Version

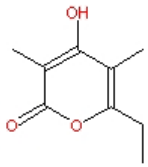
Molecular Modeling Section

Bioinformatics Lab

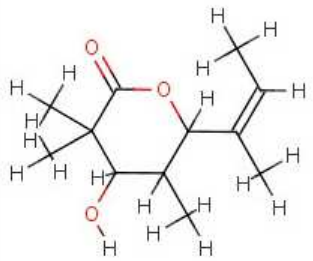
| MMsINC_Search | MMsINC_Help | MMsINC_Roadmap | MMsINC_Credits | Feedback |

Molecule result, page: 1 Items found 3

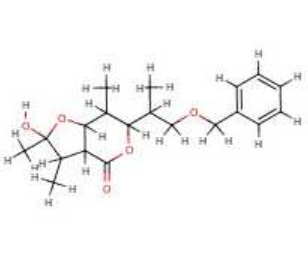
1 of 1 Go to Page



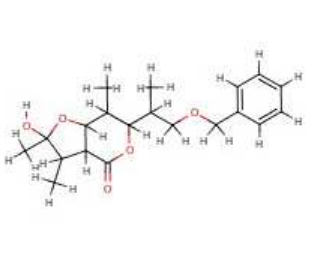
Select All Deselect All add to cart [cart empty]



MMs03207547



MMs03520027



MMs03520028