Drug Discovery informatics and modeling

BT 305

Chemo informatics

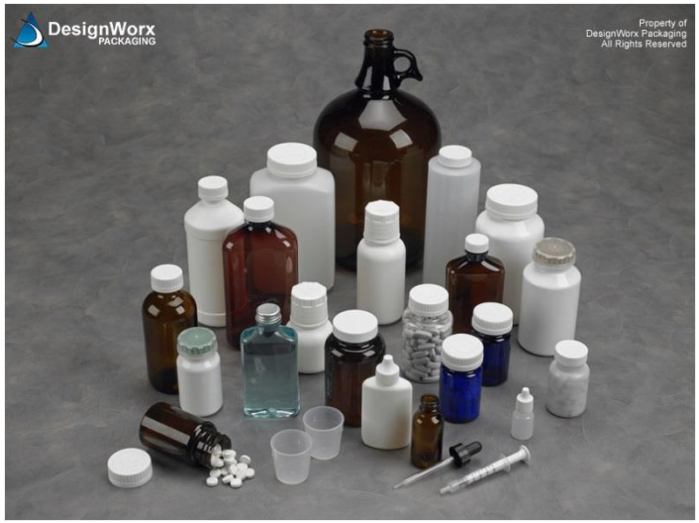
What *Chemo informatics* is ……

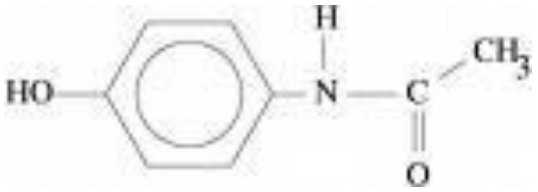
Cheminformatics is about collecting, storing, and analyzing [usually large amounts of] chemical data.

- Pharmaceutical research
- Materials design
- Computational/Automated techniques for analysis
- Virtual high-throughput screening (VHTS)

Use of Information Technology and resources in Chemistry

Our Interest – Drug Space





Chemicals that interact macromolecular targets in the body to produce a pharmacological effect

What is a Drug ?

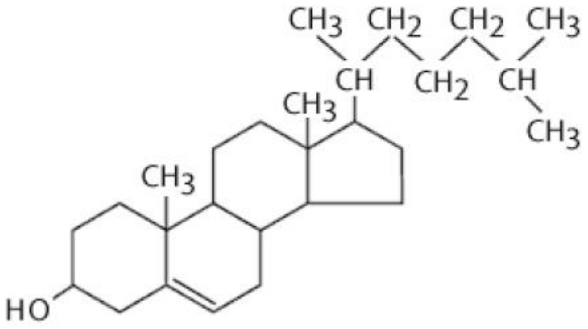
Drug (FDA Definition): An active ingredient that is intended to furnish pharmacological activity or other direct effect in the diagnosis, cure, treatment or prevention of a disease or to affect the structure of any function of the human body, but does not include intermediates used in the synthesis of such ingredient

FDA: Food & Drug Administration United States

Economics of Drug Discovery:

Average cost for discovery of a new drug: USD 500 million

Global Pharma Sales (2020, in USD) **Total: 1.2 trillion Approx.**	
North America	51 %
EU	22 %
Asia Pacific/Africa	8 – 12 %



Cholesterol:

Fatlike substance present in blood

LDL Carries cholesterol from liver to cells: attaches to receptors on the cell surface and taken into the cell

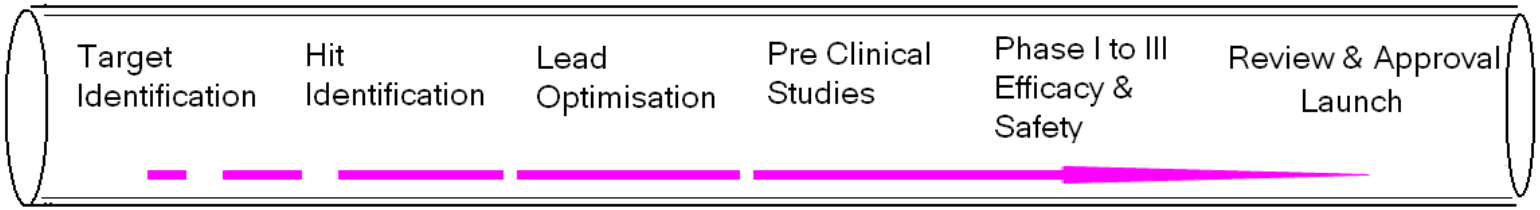
When there is excessive cholesterol inside the cell, reduce the synthesis of LDL receptors.

HDL transports cholesterol from other parts of the body to liver, where its degraded to bile acids

Traditional Drug Discovery .. Seven Steps

- Disease Selection
- Target hypothesis
- Lead compound identification
- Pre-clinical trial
- Clinical Trial
- Pharmacogenic Optimization

Drug discovery pipeline



New Chemical Entities (NCE)

Average Cost of creating a NCE in a major pharmaceutical company was estimated at around USD 7500 per compound.

Biologists can now test thousands of individual compounds can be screened per drug.

But Can Chemists make thousands of compounds a day????

Some (inevitable) jargons ……

Target

A **biological target** is a biopolymer such as a protein or nucleic acid whose activity can be modified by an external stimulus.

Drug discovery:

- Finding out the target that causes the disease
- Chemical or Biological compounds are screened and tested against these targets or assays
- Leading drug candidates for development

The most common drug targets of currently marketed drugs

1. G protein-coupled receptors (target of 50% of drugs)
2. enzymes (especially protein kinases) 28%
3. ion channels 5%
4. DNA
5. Hormones and others 11%

Receptor, Lead …..

Receptor is a protein on the cell membrane or within the cytoplasm or cell nucleus that binds to a specific molecule (a ligand), such as a neurotransmitter, hormone, or other substance, and initiates the cellular response to the ligand.

Ligand-induced changes in the behavior of receptor proteins result in physiological changes that constitute the biological actions of the ligands

A **lead compound** in drug discovery is a chemical compound that has pharmacological or biological activity and whose chemical structure is used as a starting point for chemical modifications in order to improve potency, selectivity, or pharmacokinetic parameters.

What Kind of molecules become drugs ????

Drug Like Criteria – Lipinski Rule of Five

The rule states that potential drug candidates are likely to have poor absorption and permeability if they have:

- **>5 Hydrogen bond donors (sum of –OH and NH_2 groups)**
- **Molecular weight > 500**
- **Log P > 5 (Partition coefficient which indicates lipophilicity)**
- **> 10 hydrogen bond acceptors**

Lead Optimisation

Adsorbtion

Distribution

Metabolism

Excretion

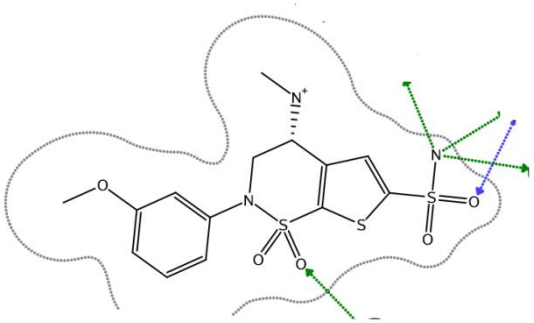
Toxicity

ADME-Tox

The Drug must reach the site of action in a timely manner in sufficient concentration to produce the desired effect

Some more (inevitable) jargons ……

Pharmocophore



A pharmocophore element is defined as an atom or a group of atoms (functional groups) common for active compounds with respect to a receptor and essential for the activity of the compounds.

A collection of pharmocophore elements and its spatial arrangements constitute a pharmocophore.

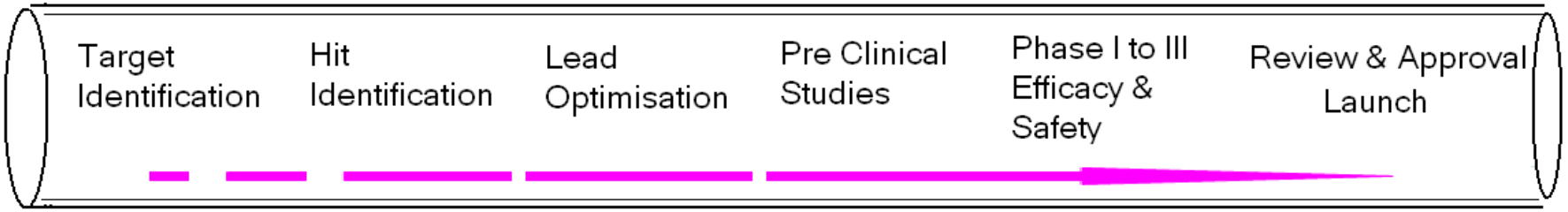
Biophore

Key Interacting features and additional attributes of a receptor – molecule interaction like shape , electrostatics, hydrogen bonding etc. together forms a biophore.

QSAR

Quantitative structure-activity relationships (QSAR) represent an attempt to correlate structural or property descriptors of compounds with activities. These physicochemical descriptors, which include parameters to account for hydrophobicity, topology, electronic properties, and steric effects, are determined empirically or, more recently, by computational methods

Drug discovery pipeline

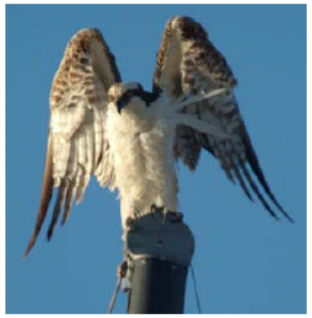


Where do a computer Scientist fit in …….

……. In quantifying the property from structure ….

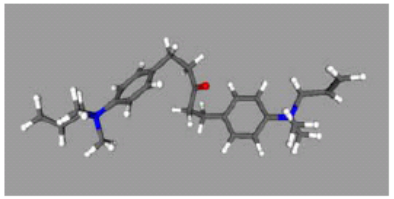
--- generate and maintain such a database ------

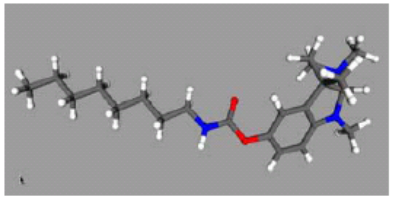
QSAR (of birds):



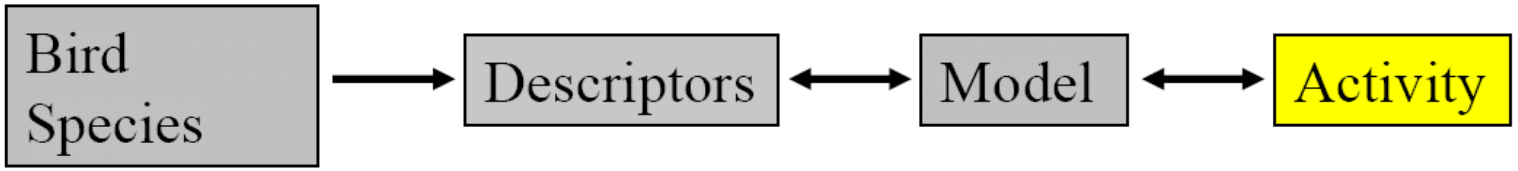
QBSPR: What is flight performance? Performance = muscle efficiency * flight time as

(P = Efficiency*Time)

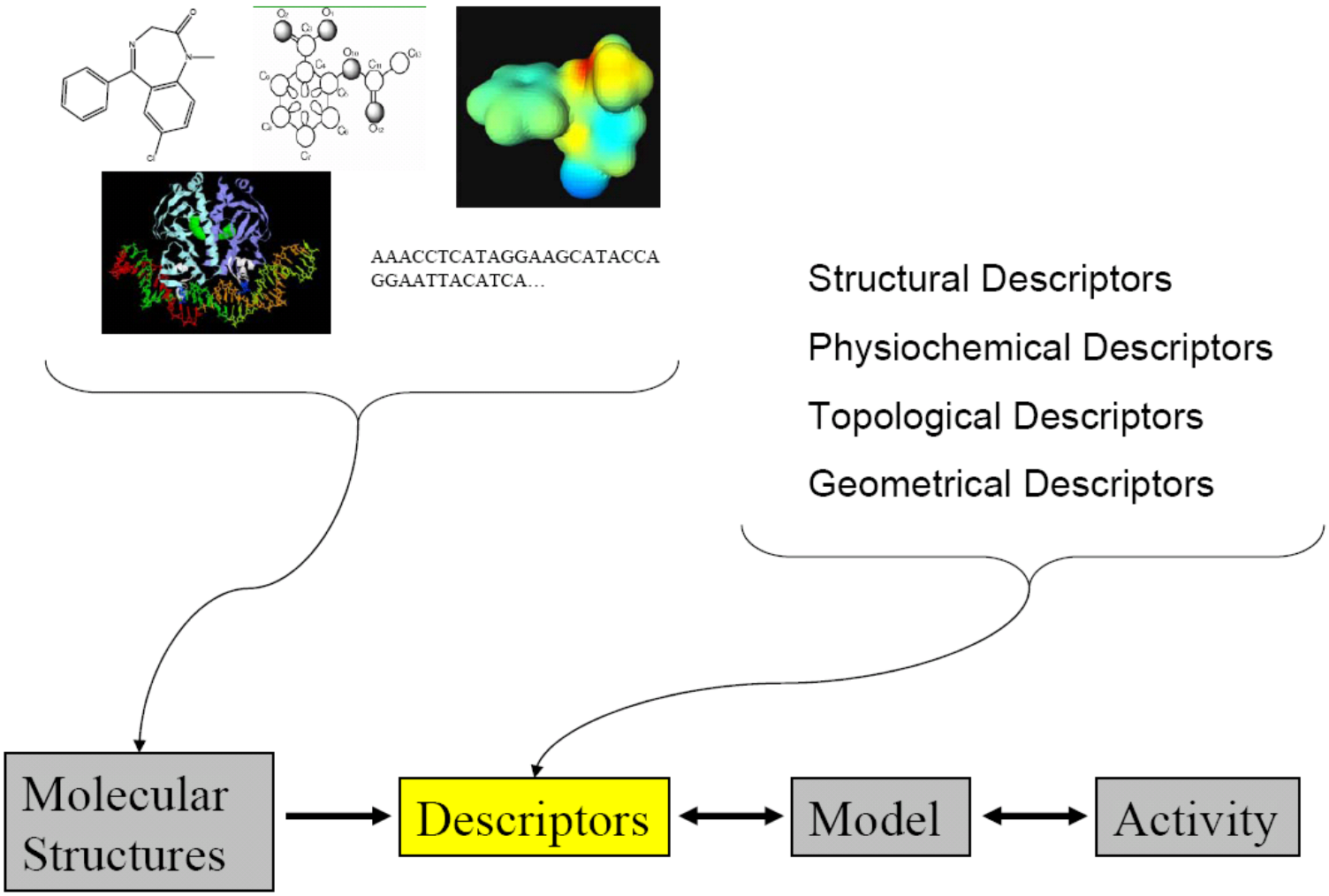




QSAR: What is AChE inhibition? IC50 is a measure of the concentration required for 50% inhibition.



Courtesy: Matt Sundling, RECCR Troy NY.



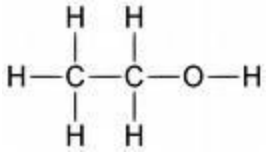
Courtesy: Matt Sundling, RECCR Troy NY.

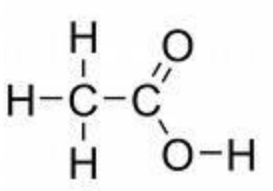
Canonicalization of chemical data……

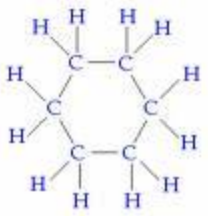
SMILES

Simplified Molecular Input Line Entry System

Ethanol	CCO
Acetic acid	CC(=O)O
Cyclohexane	C1CCCCC1
Pyridine	c1cnccc1
Trans-2-butene	C/C=C/C
L-alanine	N[C@@H](C)C(=O)O
Sodium chloride	[Na+].[Cl-]
Displacement reaction	C=CCBr>>C=CCI







concept of a graph with nodes as atoms and edges as bonds to represent a molecule.

Parentheses are used to indicate branching points and numeric labels designate ring connection points.

David Weininger 1980

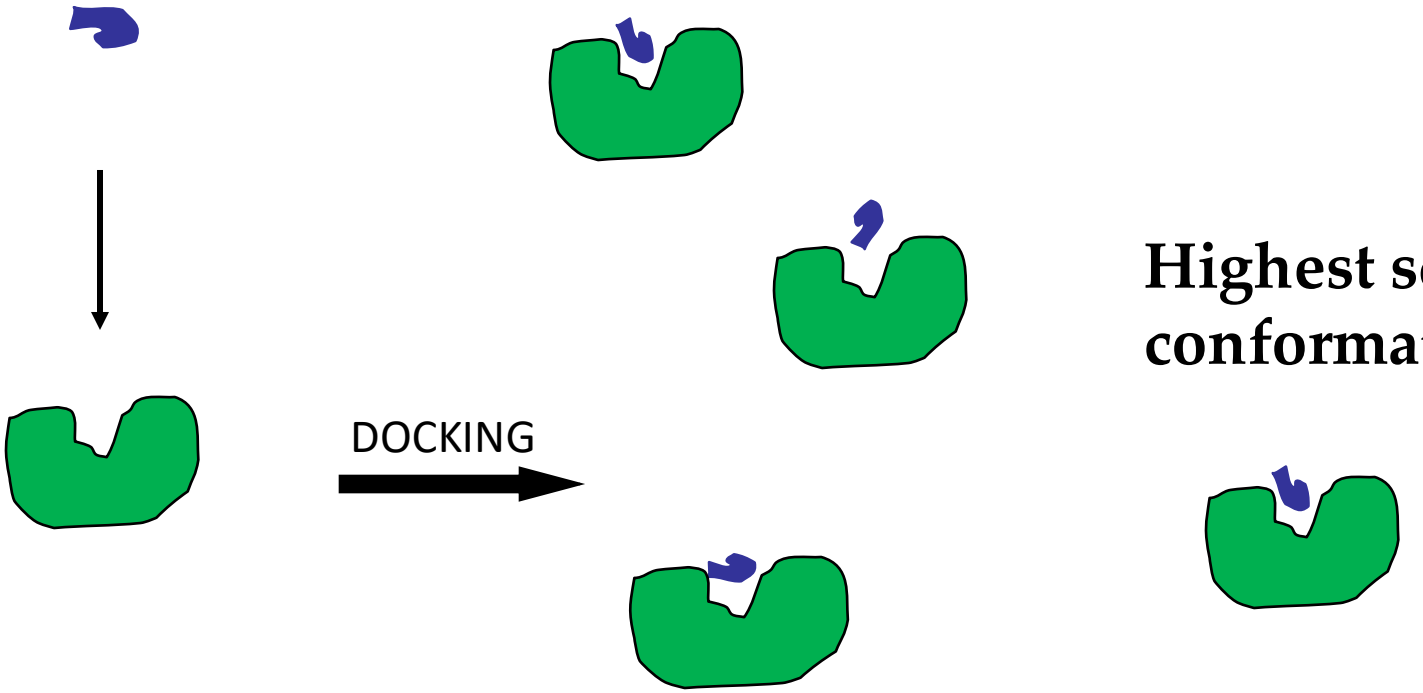
Docking

What is docking?

- Molecular modeling method
- Predicts the preferred orientation of one molecule to a second when to each other to form a stable complexes.
- Rational of design of drugs

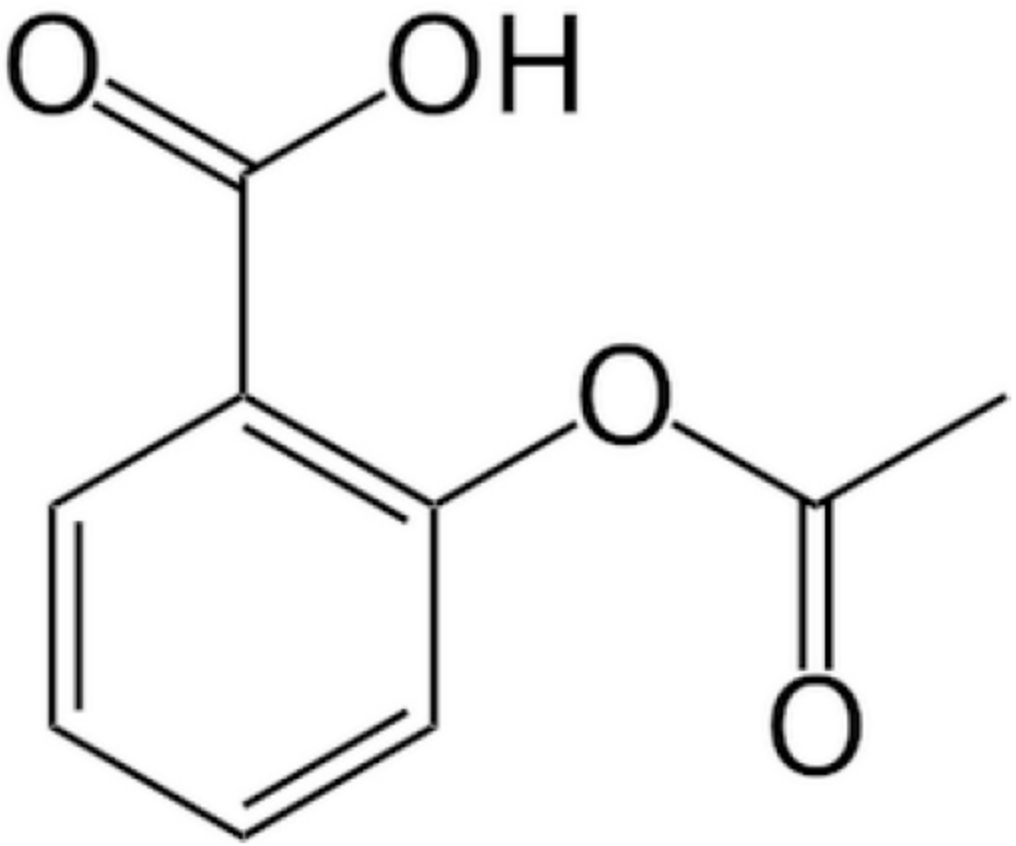
Initially,molecular docking was used to predict and reproduce_protein-ligand complexes.

Best Fit ??????



- *Focus of molecular docking is to achieve an optimized conformation for both the protein/DNA/RNA/etc and ligand and relative orientation between target and ligand*

Topological Molecular Graph



Molecular Descriptor

TABLE 1.2. Different types of molecular descriptors

Descriptor category	Examples
Physical properties	Molecular weight
	logP(o/w)
Atom and bond counts	Number of nitrogen atoms
	Number of aromatic atoms
	Number of rotatable bonds
Pharmacophore features	Number of hydrogen bond acceptors
	Sum of van der Waal surface areas of basic atoms
Charge descriptors	Total positive partial charge
	Dipole moment from partial charges
Connectivity and shape descriptors	Kier and Hall molecular shape indices
Surface area and volume	Solvent-accessible surface area

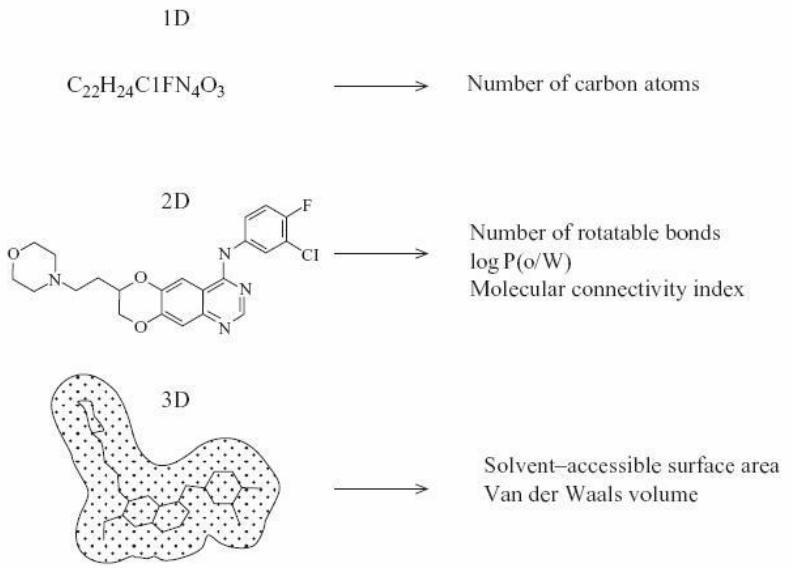


Figure 1.3. Examples of descriptors classified according to dimensionality (adapted from Bajorath 2002)

No generally preferred descriptor spaces - Context Dependant

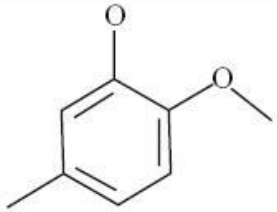
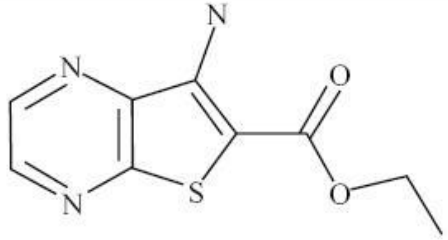
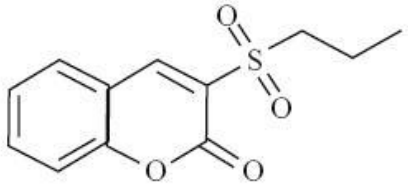
Similarity Searching –Structural queries and graphs

- Contemporary substructure search methods are mostly based on dictionaries of predefined molecular fragments.

- Queries can be transformed into an machine-readable format such as Simplified Molecular Input Line Entry Specification (SMILES) code.

- SMILES encodes 2D representation of molecules as linear strings of alpha-numeric characters.

1D String Representation … SMILES

Structures	Strings
	c1ccccc1
	Oc1cc(C)ccc1OC
	s1c2[nH0]cc[nH0]c2c(N)c1C(=0)OCC
	[S+2]([O–])([O–])(CCC)C1=Cc2ccccc2OC1=O

Scaffolds, Linkers and Sidechains (Functional Groups)

Rings Systems: Cycles within the molecular graphs or rings sharing an edge or vertex in the molecular graph.

Linkers: Edges (bonds) that connect two ring systems

Sidechains: Those atoms that are neither rings or linkers

Frameworks: Ring Systems connected by linkers

Similarity Searching –Structural queries and graphs

- Detection of structural fragments or substructures is a simple but popular form of similarity searching.

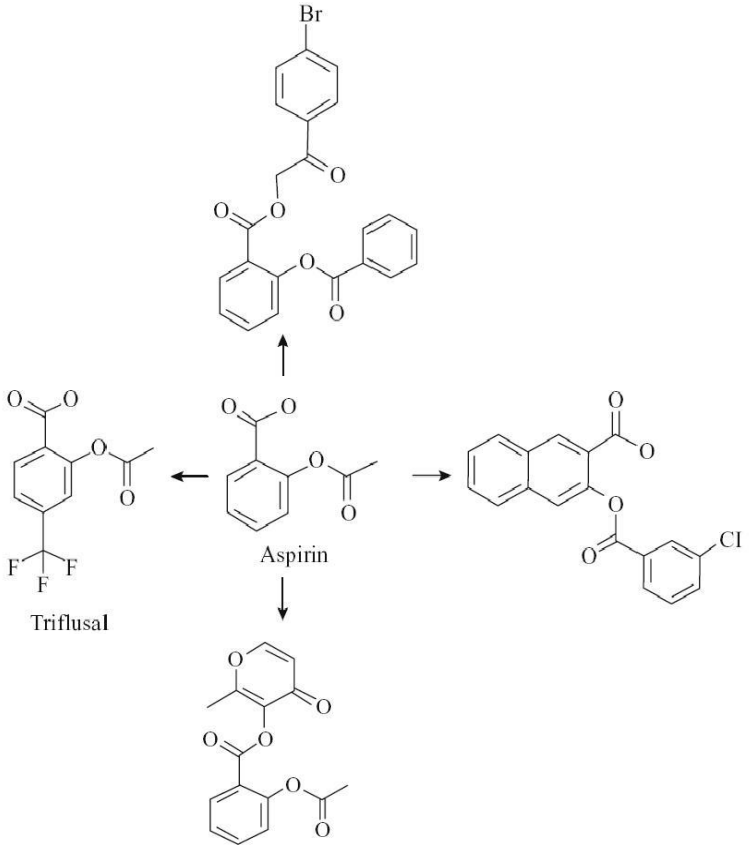


Figure 1.10. Example of compounds containing Aspirin as a substructure that can be used as a query for database searching

Similarity Searching –Structural queries and graphs (Reduced graph)

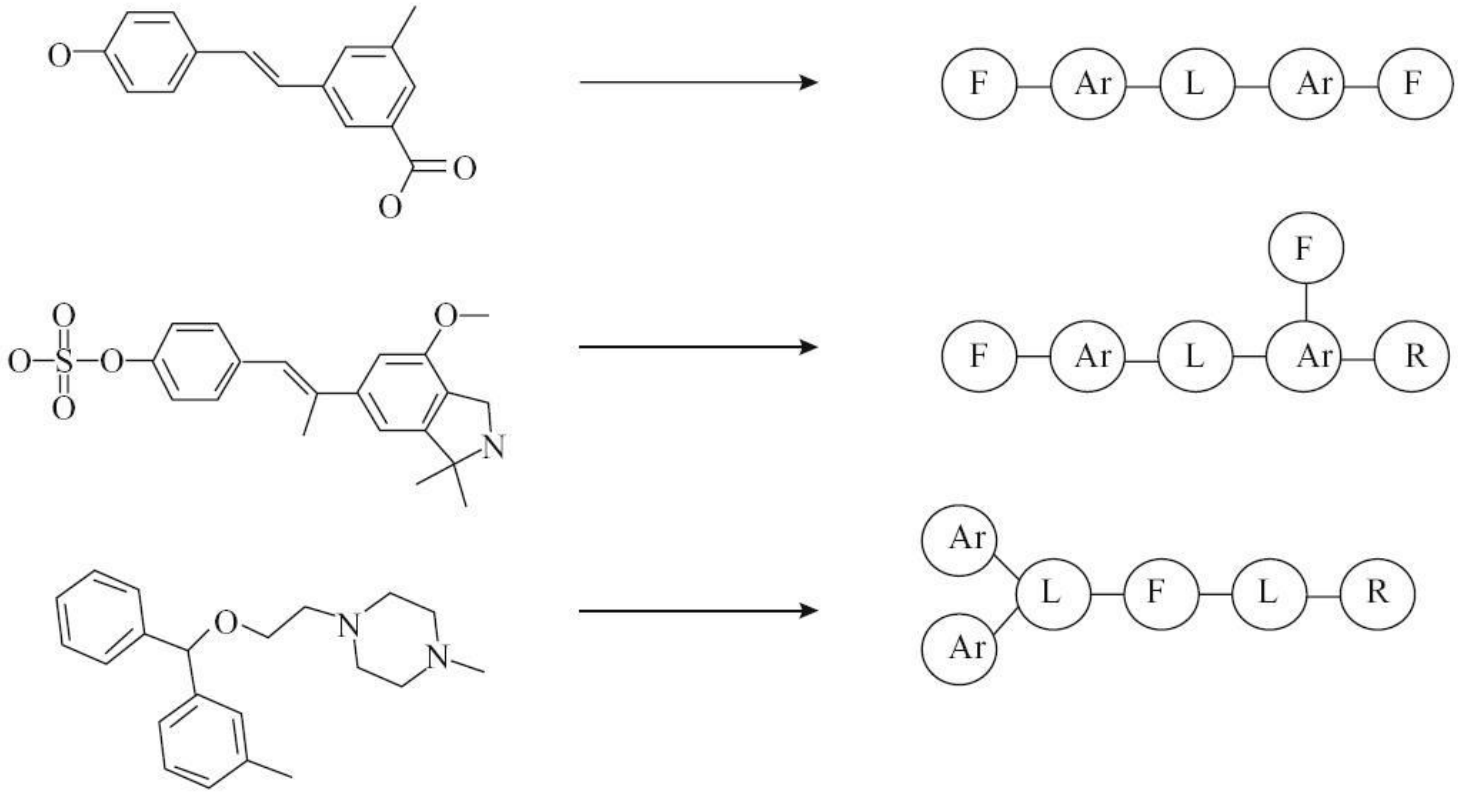


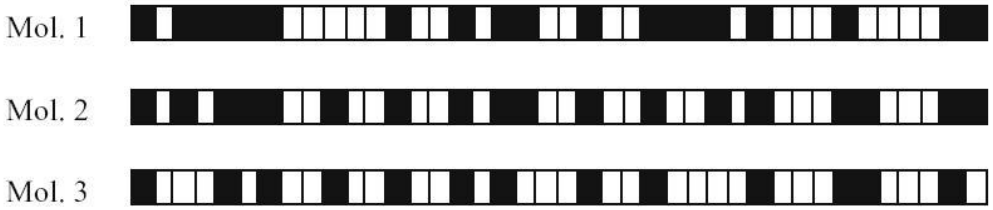
Figure 1.12. Examples of reduced graphs. Nodes corresponding to aromatic rings (Ar), aliphatic rings (R), functional groups (F) and linking groups (L) are shown (adapted from Gillet *et al.* 2003)

Similarity Searching – Pharmacophore

- A molecular framework that carries the essential features responsible for drug's biological activity
- Spatial arrangements of atoms or groups that are responsible for biological activity
- Often used as 3D queries for database searching

Similarity Searching –Fingerprints

- Fingerprints：
  - widely used similarity search tools.
  - consist of various descriptors that are encoded as bit strings
  - Bit strings of query and database compared using similarity metric such as Tanimoto coefficient



$$Tc(Mol1, Mol2) = \frac{16}{(19+18)-16} = 0.76 \qquad Tc(Mol1, Mol3) = \frac{11}{(19+13)-11} = 0.52$$

Figure 1.14. Model fingerprints and Tc comparisons

Scaffold Hopping

The Concept of scaffold-hopping aims at finding molecules that possess ***different scaffolds*** but exhibit identical or very ***similar pharmocological*** activity

Drug Likeness

TABLE 1.5. Drug-like versus lead like compound characteristics

Drug-like	Lead-like
MW < 500	MW < 350
ClogP < 5	ClogP < 3.0
Hydrogen bond donors < 5	Chemically stable
Hydrogen bond acceptors < 10	
Number of rotatable bonds ≤ 10	
PSA $\leq 140 Å^2$	
Peptides not suitable	Non-substrate peptides suitable
Eliminate reactive functional groups, promiscuous inhibitors, and metabolically unstable compounds	

BIOACTIVE CONFORMATION

- Configuration and conformation are different.
- Receptor bound conformation is the bio-active conformation.
- Increase exponentially with the number of rotatable bonds.
- In general receptor bound conformations are almost impossible to predict from the ensemble of possible conformers.