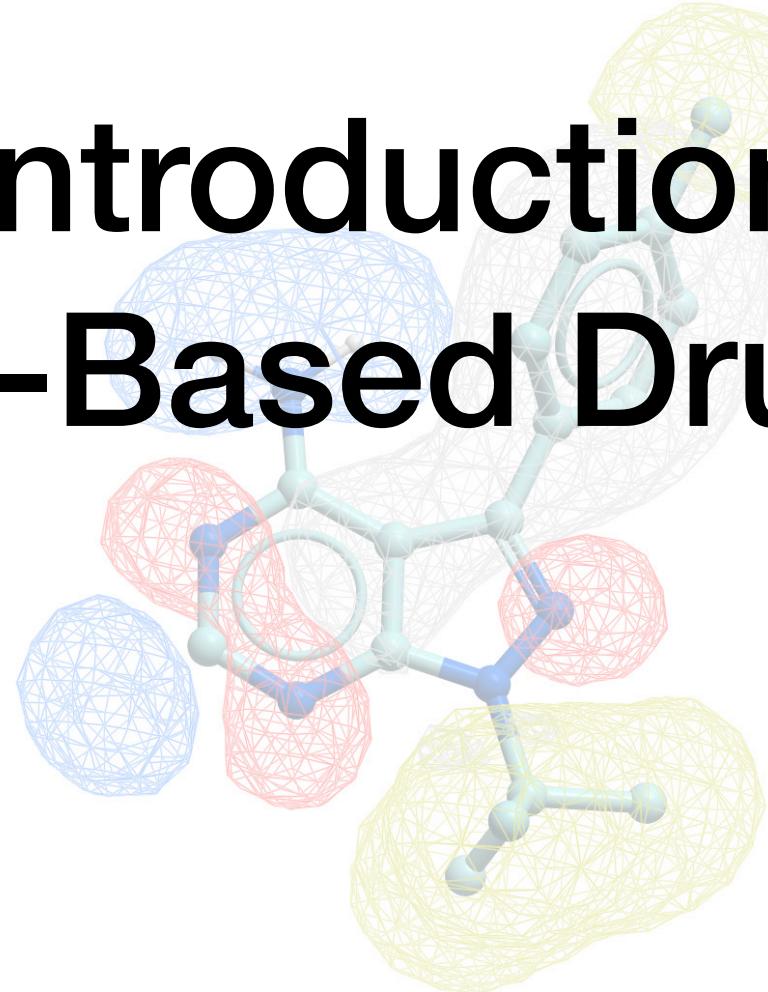


Introduction to Ligand-Based Drug Design



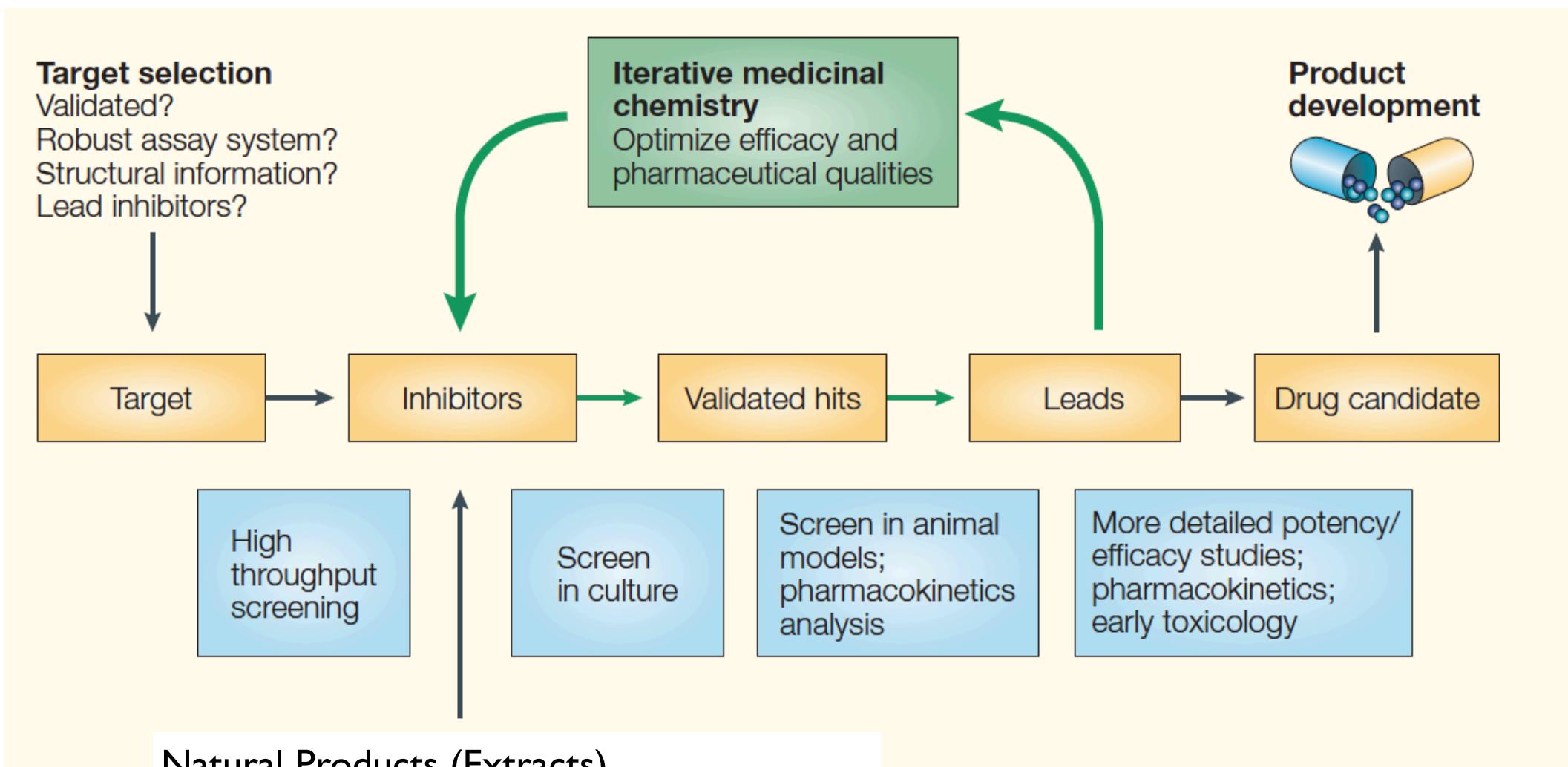
Dr. Fabien Plisson

Chemoinformatics in Drug Discovery

LANGEBIO, UGA CINVESTAV

October 15-18, 2019 - Irapuato, Mexico

Drug Discovery and Design - long and iterative process



Natural Products (Extracts),
Synthetic Compounds
Peptides, Peptidomimetics, Antibodies...

Clinical trials can cost

up to
\$4 billion dollars

and take
10 to 15 years.

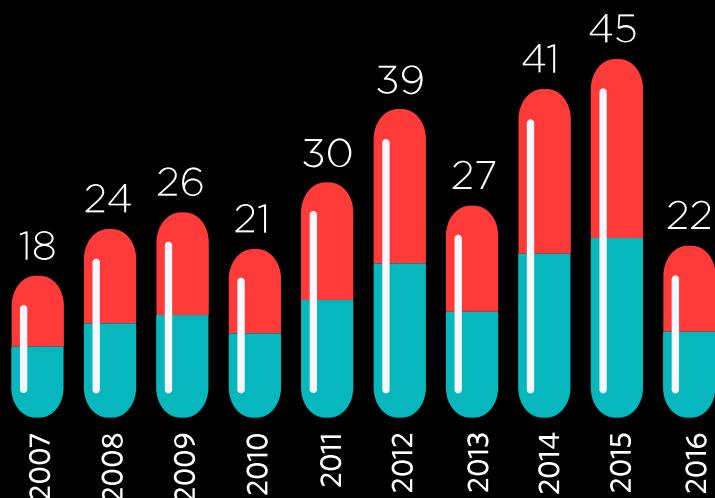
Fewer than
10 percent

make it to market.

THE COST OF DRUG DEVELOPMENT

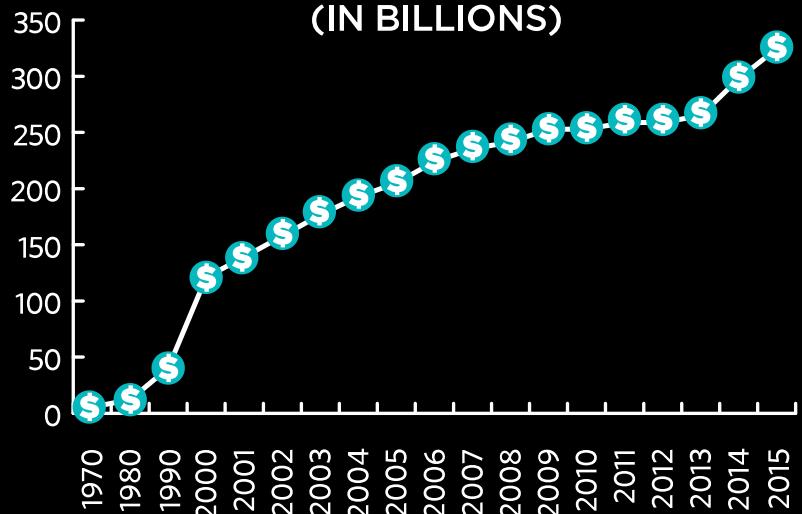
EXPENSIVE CLINICAL TRIALS AND FEW DRUG APPROVALS CAN DRIVE UP DRUG PRICES FOR CONSUMERS.

NEW DRUG APPROVALS



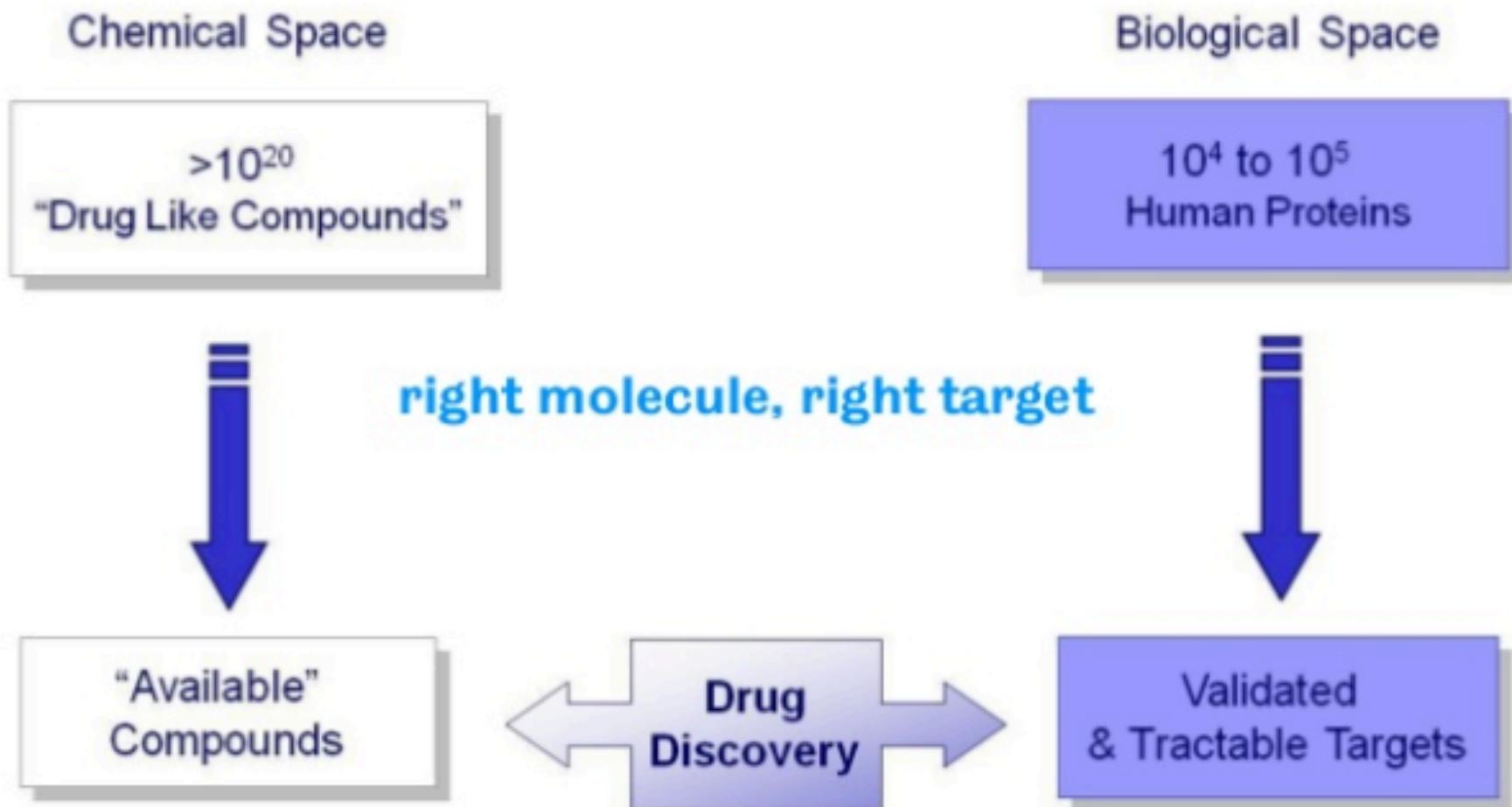
Only 22 novel drugs were approved last year—a 57 percent drop from approvals in 2015.

MONEY SPENT ON PRESCRIPTION DRUGS (IN BILLIONS)

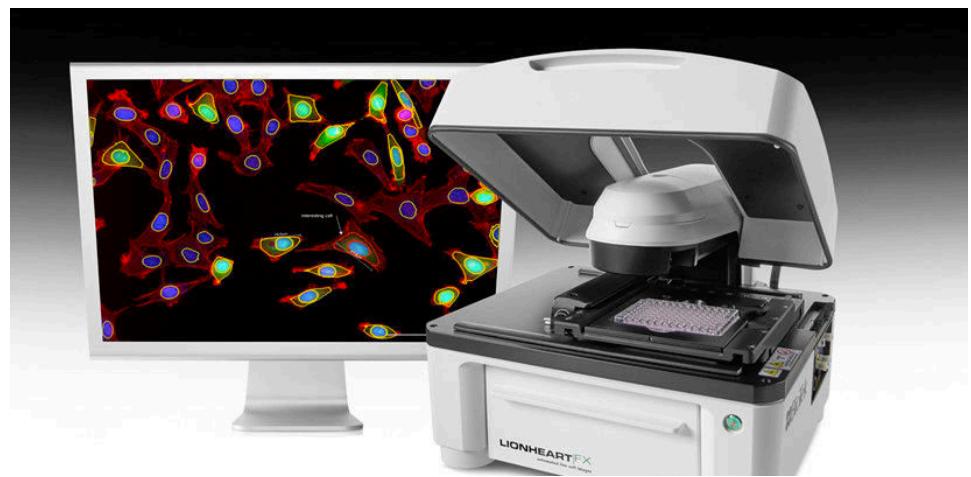


Americans spent \$324.6 billion on prescription drugs in 2015. This amount represents almost 20 percent of US health care costs per capita.

Drug Discovery Challenge



Automating Drug Discovery and Design



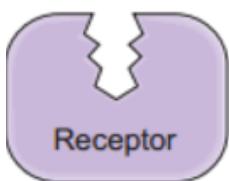
Computer-Aided Drug Design (CADD) since 1980s



“Drug companies know they simply cannot be without these computer techniques. They make drug design more rational. How? By helping scientists learn what is necessary, on the molecular level, to cure the body, then enabling them to tailor-make a drug to do the job.”

Fortune magazine October 05, 1981

Computer-Aided Drug Design



Protein
structure

Known

Unknown

Known

Unknown



Ligand

Structure-based drug design (SBDD)

Receptor-ligand docking

De novo design

Virtual screening

Fragment-based drug design (FBDD)

Ligand-based drug design (LBDD)

1 or more ligands

Similarity searching

Several ligands

Pharmacophore searching

Many ligands (20+)

Quantitative Structure-Activity Relationships (QSAR)

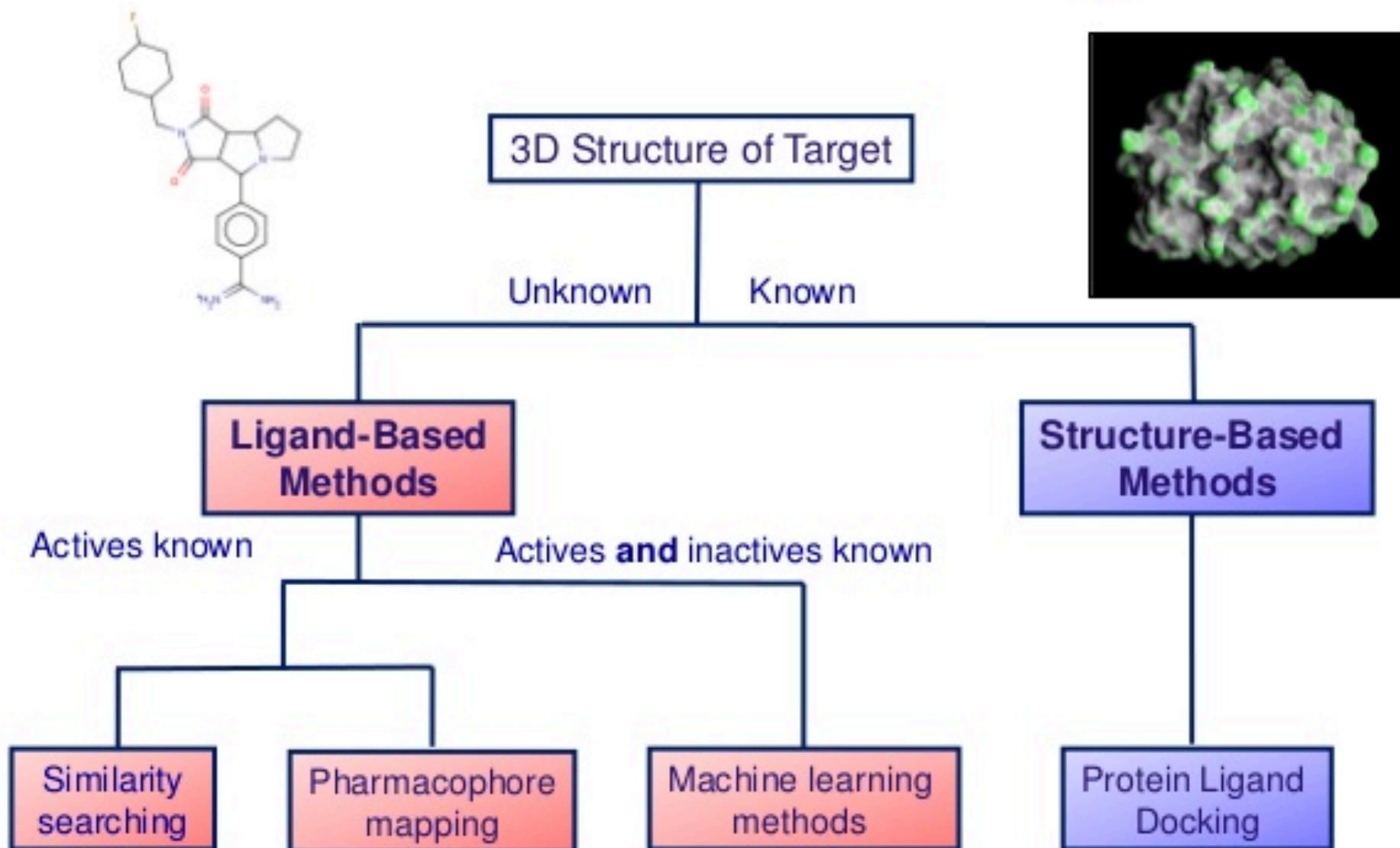
CADD of no use

Need experimental data of some sort

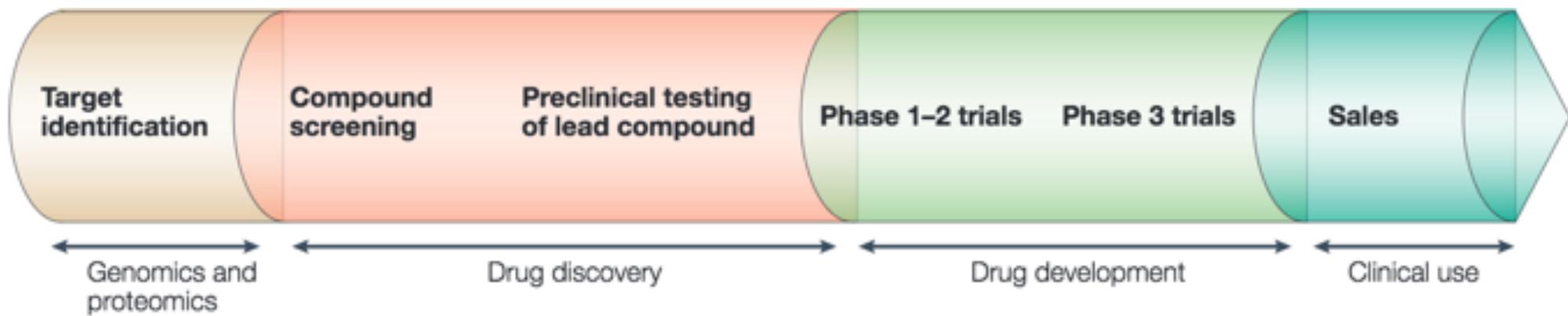
Can apply filters

e.g. ADMET, Lipinski Ro5, PAINS

Computer-Aided Drug Design



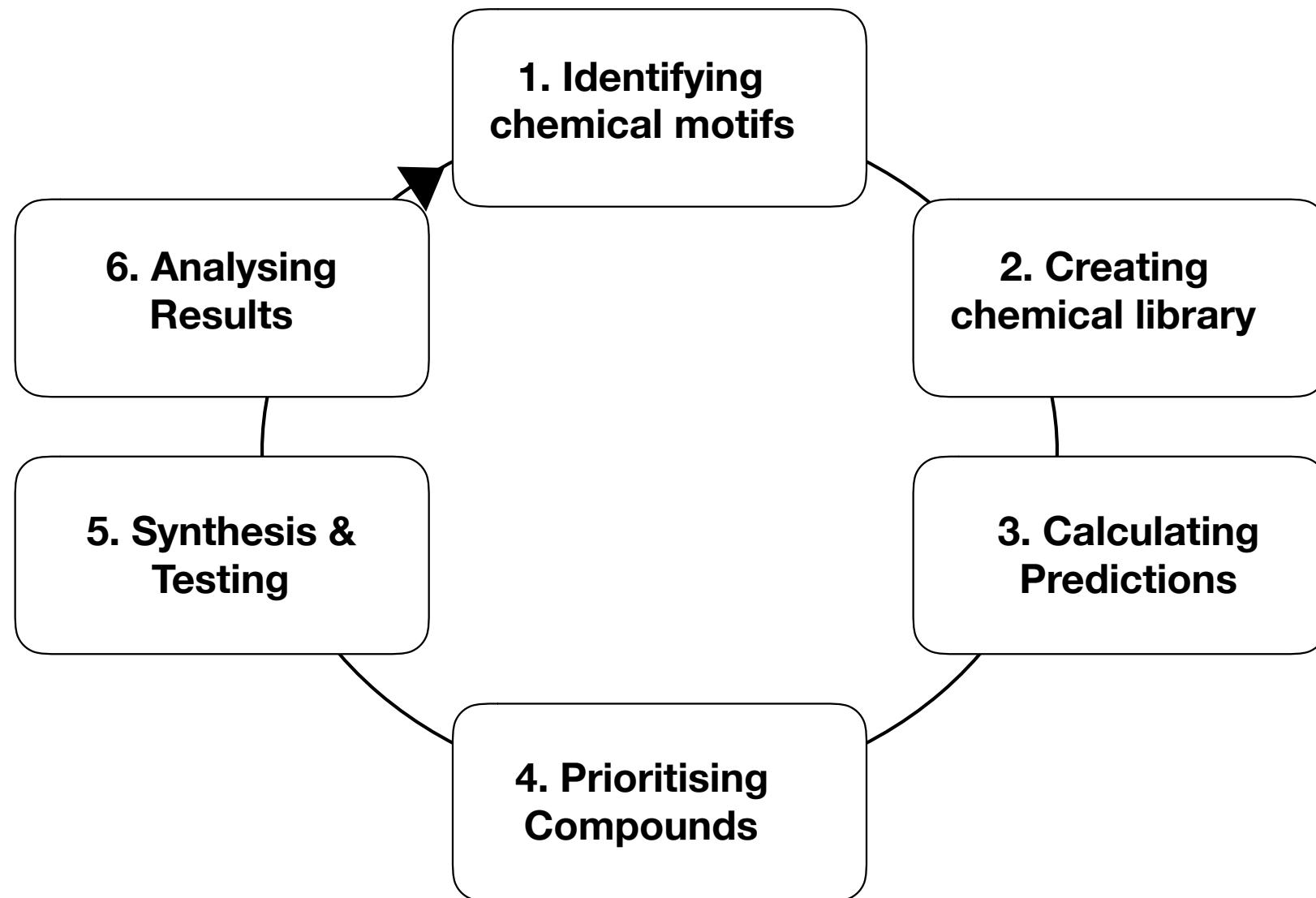
Computer-Aided Drug Design in Drug Discovery Pipeline



- Bioinformatics
 - Library design
 - QSAR
 - ADMET prediction
 - Reverse docking
 - Virtual screening
 - Pharmacophore
 - Pharmacokinetic simulations
 - Homology modeling
 - Docking / Scoring
 - Structure-based opt.
 - Target druggability
 - *De novo* design
 - Similarity search

Ligand-based Drug Design

Computational methods to design, select and prioritise chemical compounds that contribute positively to a medicinal chemistry project.



Program

Molecular Similarity

Pharmacophore Modelling

Quantitive Structure-Activity Relationship

Program

Molecular Similarity

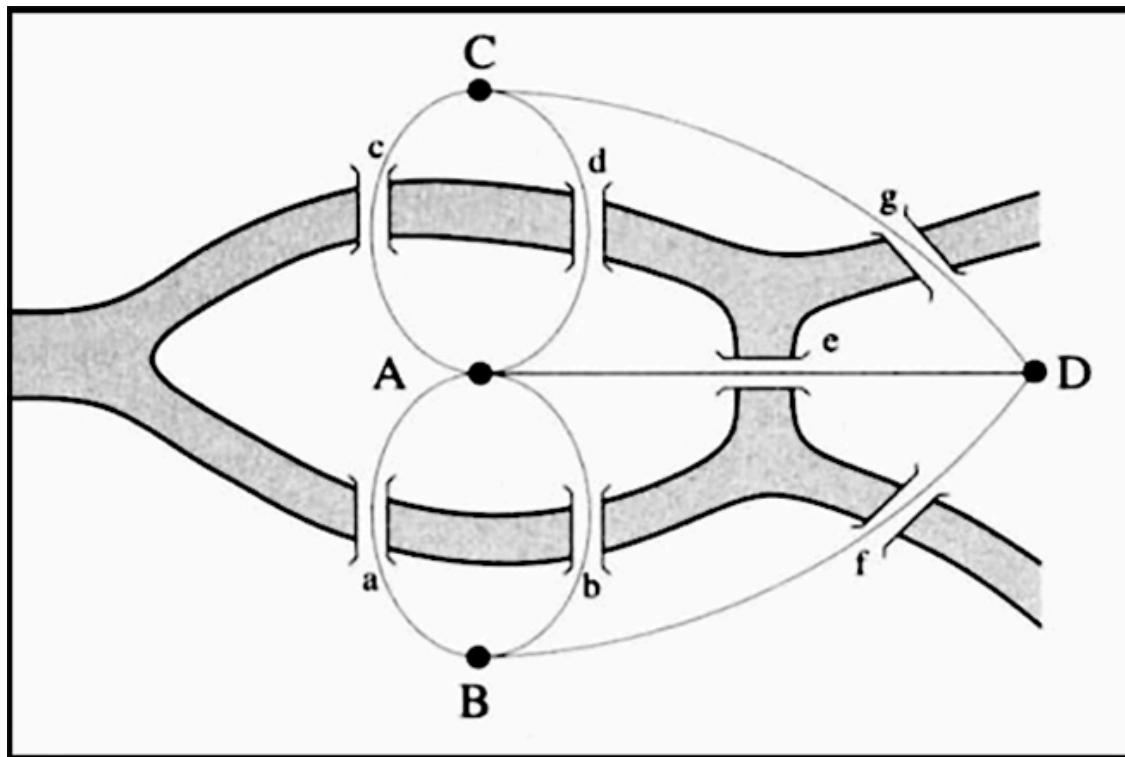
Pharmacophore Modelling

Quantitive Structure-Activity Relationship

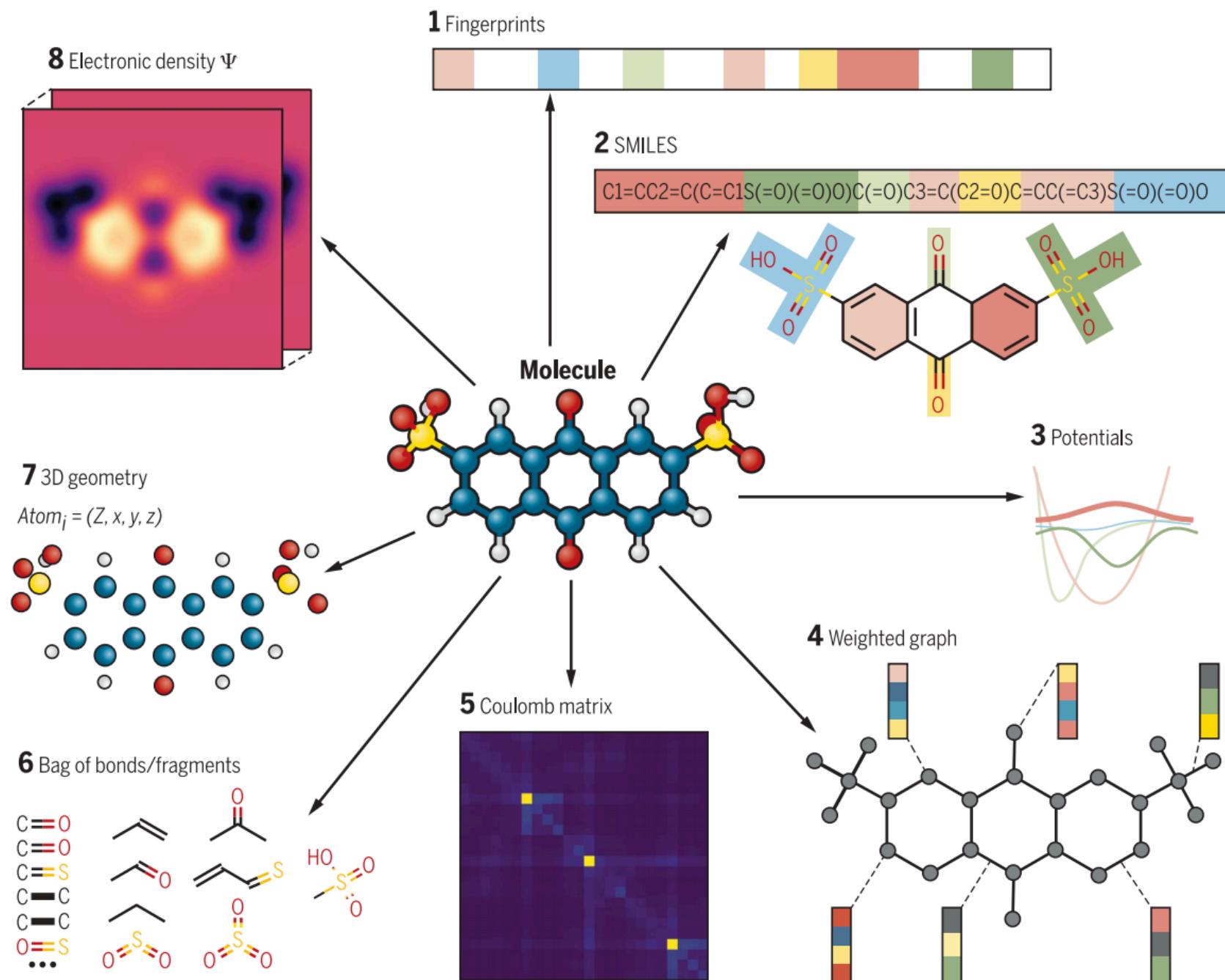
Molecular Representation takes its origin in Graph Theory

Leonhard Euler and the **Seven Bridges of Königsberg** problem (1736).

Given the graph in the image, is it possible to construct a path (or a cycle, i.e. a path starting and ending on the same vertex) that visits each edge exactly once?



Encoding Molecules



Rationale for Similarity Searching

The **similar property principle** states that structurally similar molecules tend to have similar properties (cf. neighbourhood principle)

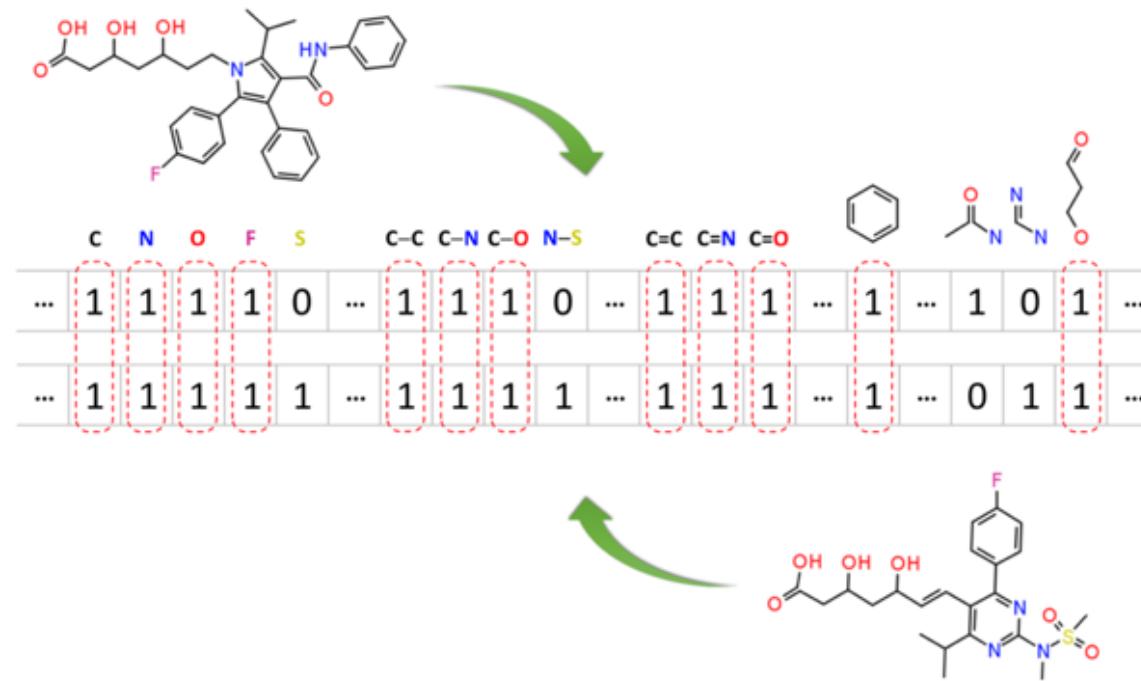


Basis of medicinal chemistry efforts and of all ligand-based virtual screening methods, despite the existence of “activity cliffs”.

Three components of a similarity measure

1. **Molecular descriptors** - numerical values assigned to structures
 - Physicochemical properties: e.g. MW, logP, MR, PSA....
 - 2D properties: fingerprints, topological indices, maximum common substructures
 - 3D: fingerprints, molecular fields
2. **Similarity coefficients** - a quantitative measure of similarity between two sets of molecular descriptors
3. Can also use a **weighting function** to ensure equal (or non-equal) contributions from all parts of the measure.

2D fingerprints: molecules represented as binary vectors

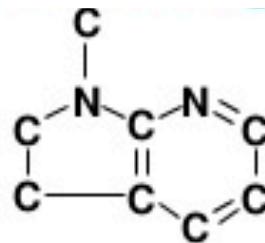


Each bit in the bit string (binary vector) represents one molecular fragment. Typical length is ~1000 bits.

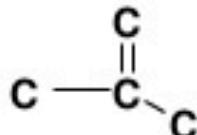
The bit string for a molecule records the presence ("1") or absence ("0") of each fragment in the molecule.

Originally developed for speeding up substructure search.

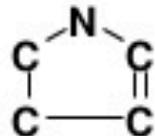
Example fragments



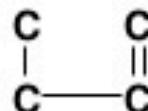
a. Augmented Atom
C rs C rd C rs C



d. Ring Composition
N rs C rd C rs C rs C rs



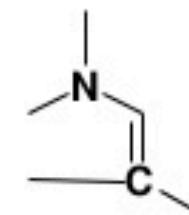
b. Atom Sequence
C rs C rs C rd C



e. Ring Fusion
XX3 XX3 XX3 XX2 XX2



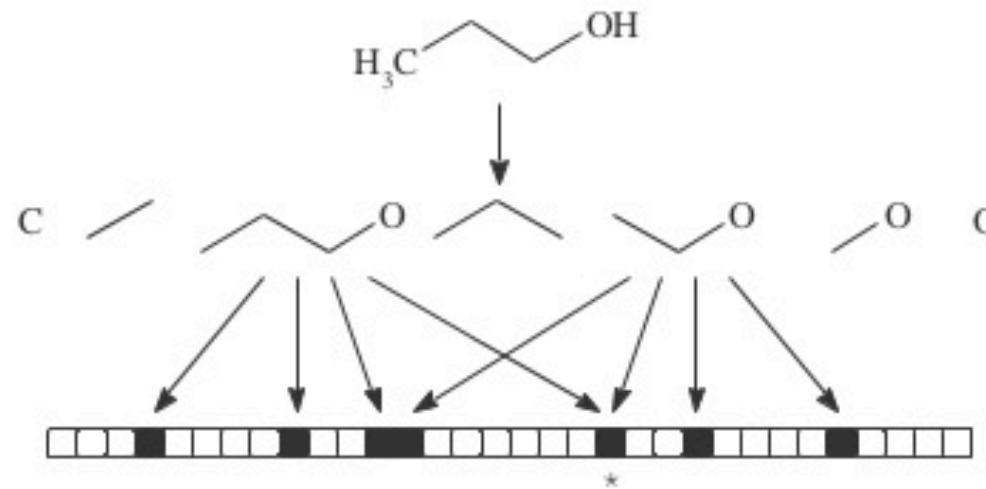
c. Bond Sequence
AA rs AA rs AA rd AA



f. Atom Pair
N 0;3 - C 0;3

Dictionary-based fingerprints: pre-defined fragments each of which maps to a single bit. Examples include MACCS Keys, BCI fps

Hashed fingerprints



Fragments are generated algorithmically without the need for a dictionary - e.g. all paths up to 7 non-hydrogen atoms.

Each fragment is processed using several different hashing functions, each of which sets a single bit in the fingerprint.

There is a one-to-many mapping between a fragment and bits in the bit string and a given bit may be set by different fragments.

Examples: *Daylight, UNITY fingerprints*

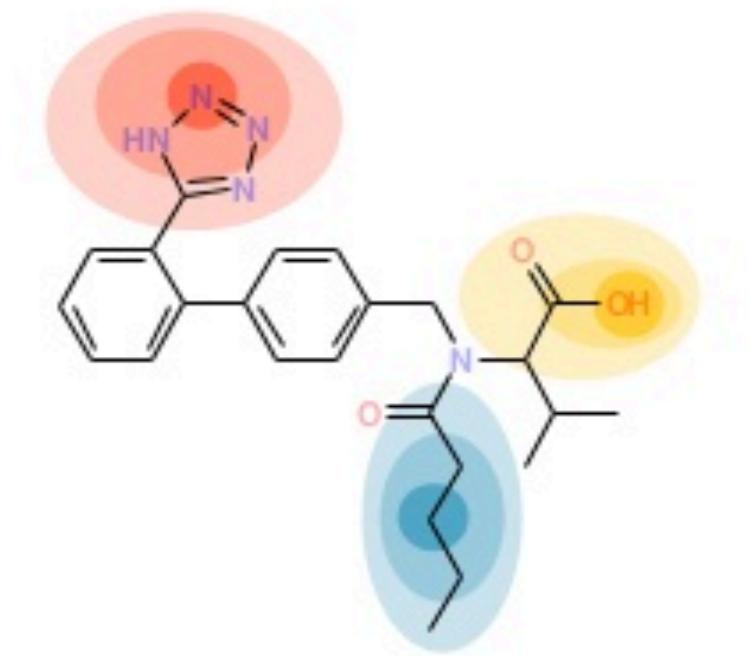
Circular substructures

Each atom is represented by a string of integers obtained by an adaptation of the Morgan algorithm.

Pipeline Pilot (Accelrys) descriptors, e.g.
ECFP2, ECFP4, ECFP6, FCFP2,...

ECFP fragments encode atomic type, charge and mass. 2/4/6 denote the diameter (in bonds) of the circular substructure.

RDKit variant Morgan, FeatMorgan



Similarity coefficients

Similarity = determine the number of bits common to 2 structures.

Tanimoto coefficient for binary bit strings

$$SIM_{RD} = \frac{C}{R + D - C}$$

- C bits set in common in the reference and database structure
- R bits set in reference structure
- D bits set in database structure

More complex form for use with non-binary data

e.g. *physicochemical property vectors*

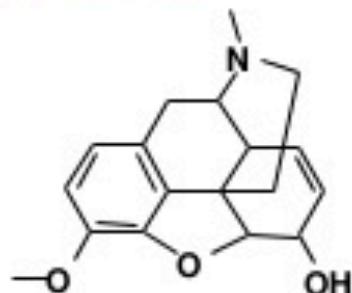
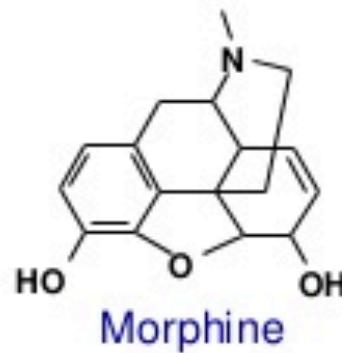
Many other types of similarity coefficient exist

e.g. *cosine coefficient, Euclidean distance, Tversky index*

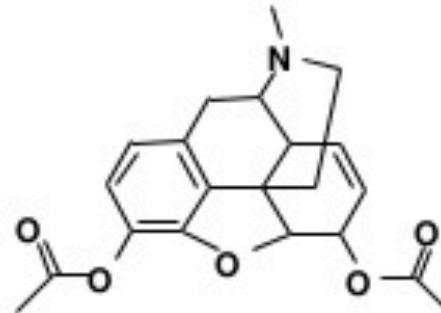
Limitations of 2D descriptors

2D fingerprints are very good at identifying close analogues.

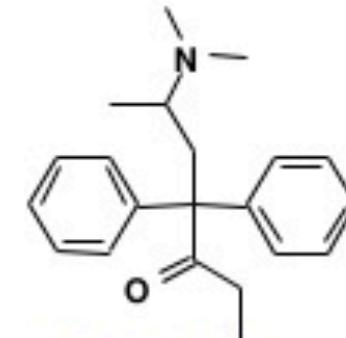
Daylight fingerprints;
Tanimoto similarities



0.99 similar
Codeine



0.95 similar
Heroin



0.20 similar
Methadone

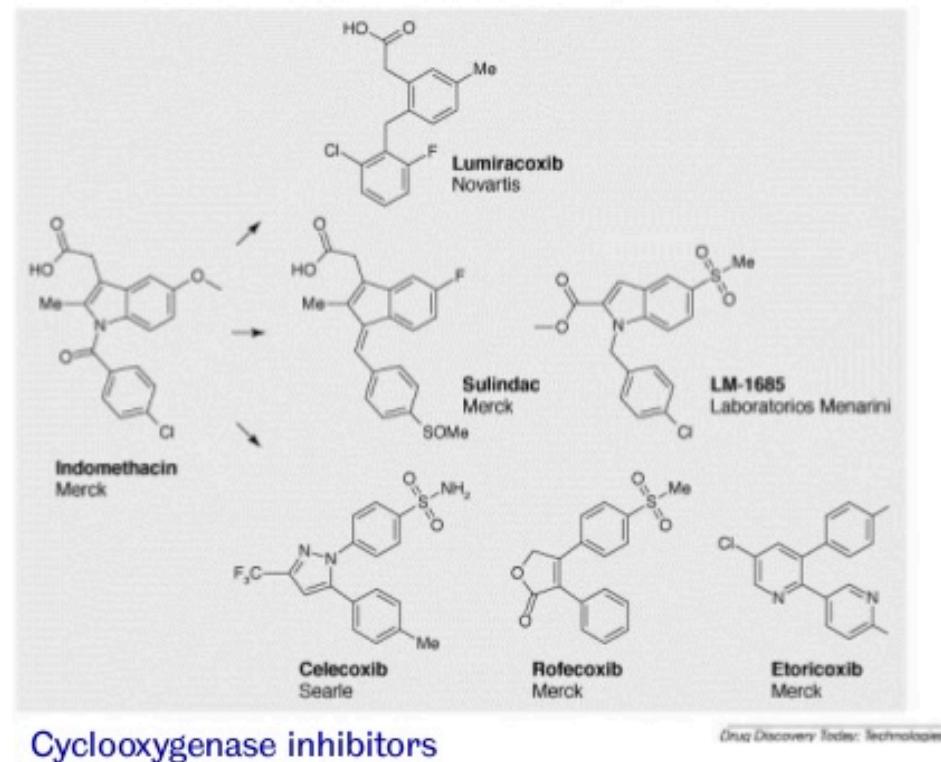
Scaffold Hopping

Scaffold Hopping: Identification of structurally novel compounds by modifying the central core structure of the molecule.

- Patent reasons: move away from competitor compounds
- Provide alternate lead series if problems arise due to difficult chemistry or poor ADME properties

Descriptors for scaffold hopping

- Topological pharmacophore keys
- Reduced graphs
- 3D descriptors



Cyclooxygenase inhibitors

Drug Discovery Today: Technologies

Bohm et al. Drug Discovery Today: Technologies (2004), 1, 217-224.
Langdon et al. Molecular Informatics (2010), 29, 366-385.

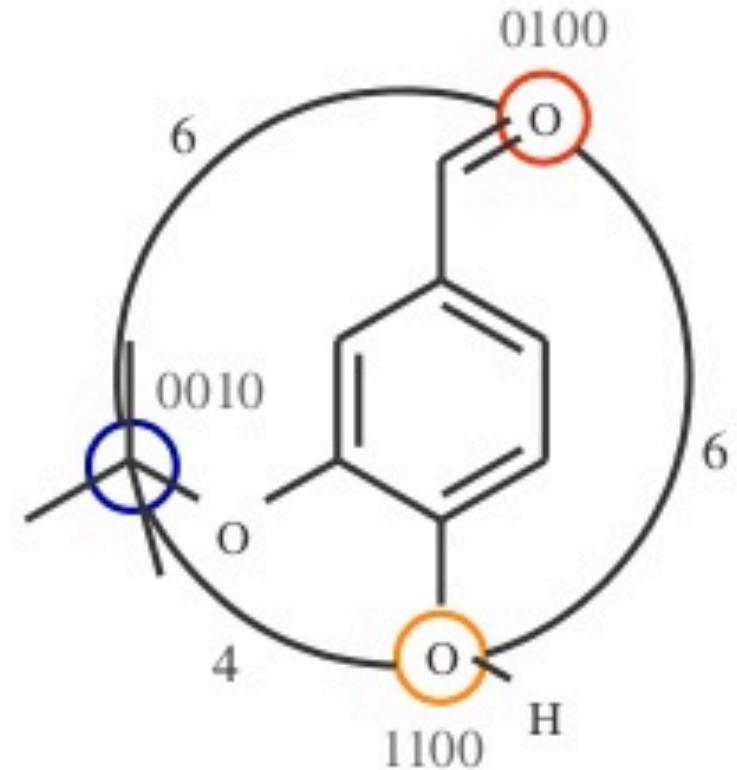
Pharmacophore vectors: Similogs

Similogs keys

Atom typing scheme based on four properties: HBD, HBA, bulkiness and electropositivity.

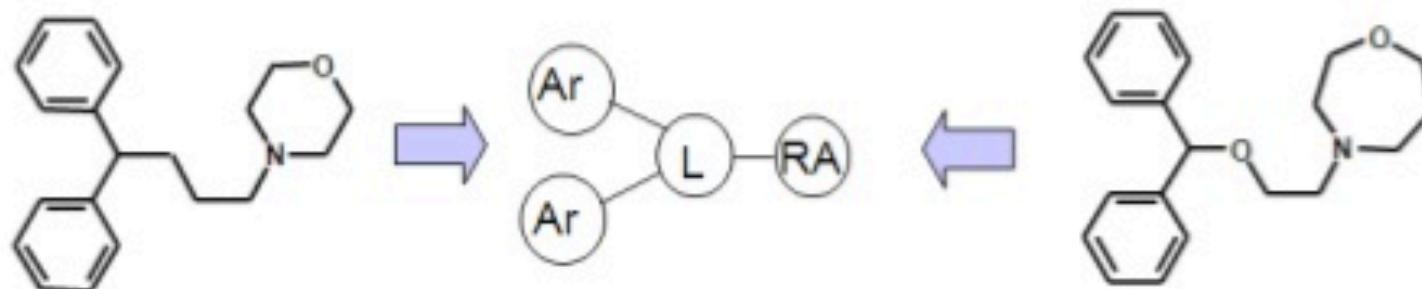
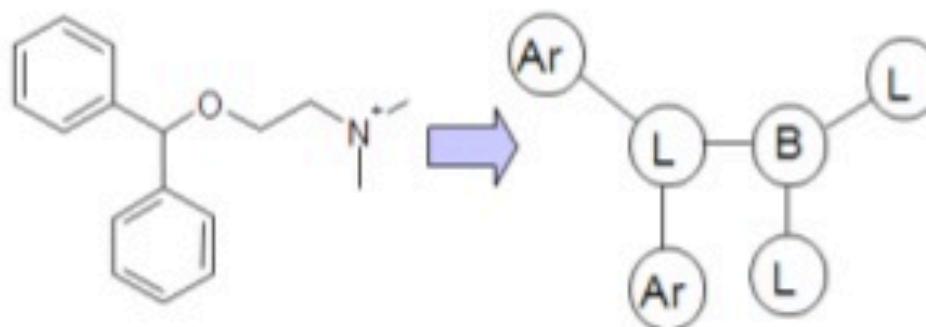
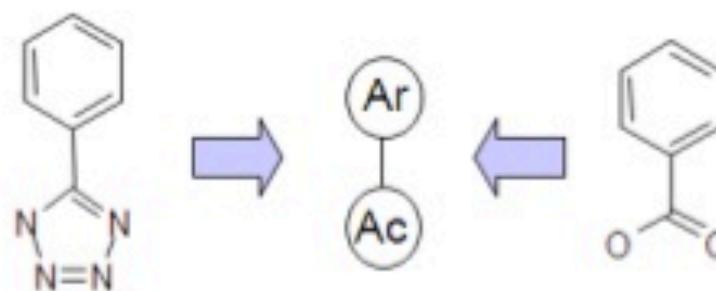
Atom triplets of strings encoding absence and presence of properties, plus distance encoding form, a DABE key.

Vector contains a count for each of 8031 possible DABE keys.



0010-4-1100-6-0100-6-

Reduced Graphs



3D Similarity Searching

Computationally more expensive than 2D methods.

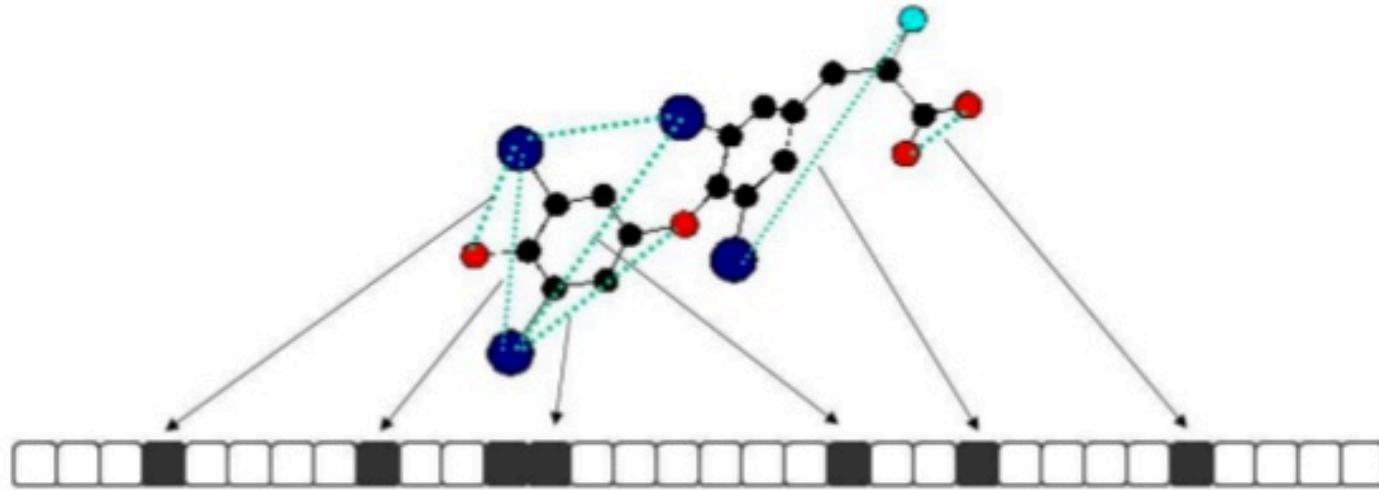
Requires consideration of conformational flexibility:

- Rigid Search (a single conformer)
- Flexible Search (ensemble of conformers at search time)

Methods that require aligning molecules are most costly than vector-based calculations.

- **Alignment independent** with fingerprint approaches
- **Alignment-based** (field-based, surface-based methods)

3D Fingerprints



Presence or absence of geometric features

- Pairs of atoms at given distance range
- Triplets of atoms and associated distance
- Pharmacophore pairs and triplets (donors, acceptors, aromatic centres...)
- Valence angles
- Torsion angles

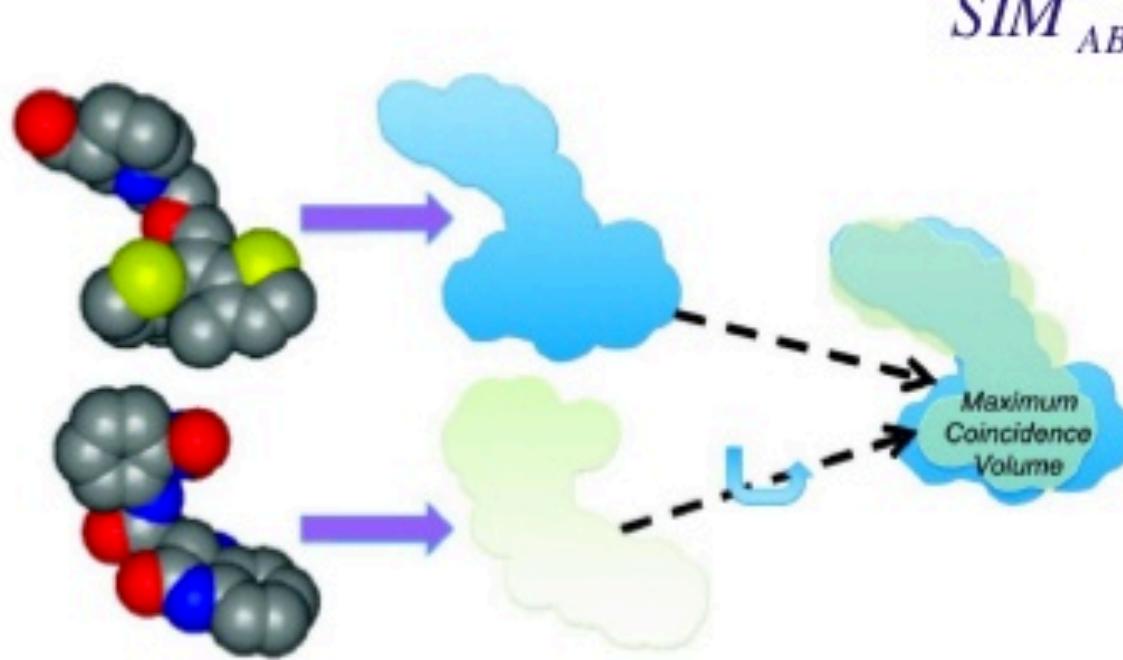
Shape-based alignment 3D similarity

ROCS (Rapid Overlay of Chemical Structures)

Molecules are aligned in 3D

Similarity score is based on common volume

$$SIM_{AB} = \frac{V_C}{V_A + V_B - V_C}$$



Program

Molecular representations

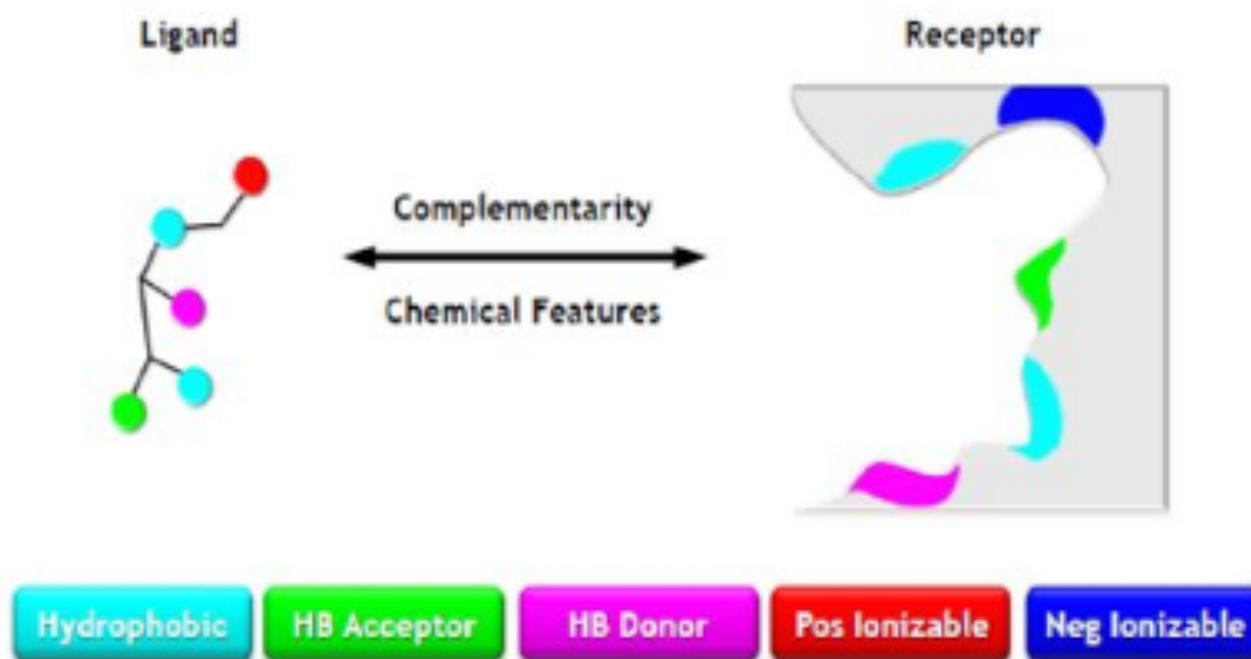
Molecular Similarity

Pharmacophore Modelling

Quantitive Structure-Activity Relationship

Pharmacophore mapping

A **Pharmacophore** is an ensemble of steric and electronic features that is necessary to ensure the optimal supramolecular interactions with a specific biological target structure and to trigger (or to block) its biological response.

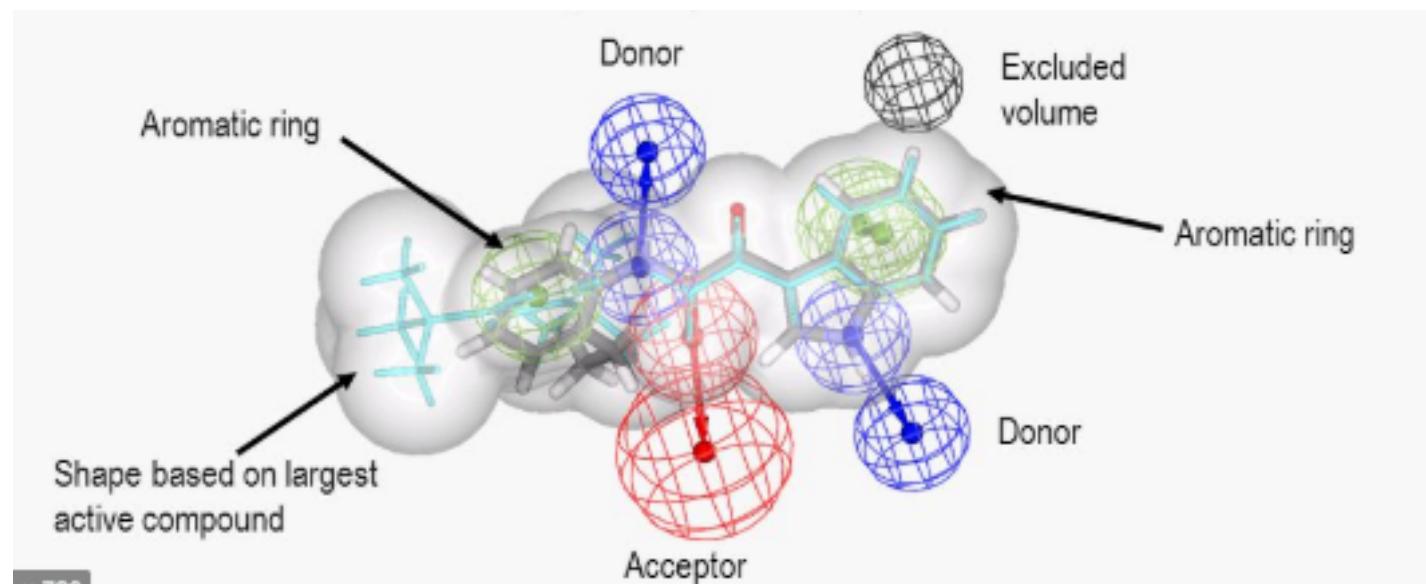


Glossary of terms in Medicinal Chemistry (IUPAC Recommendations 1998)
Pure & App. Chem. (1998), 70 (5), 1129-1143

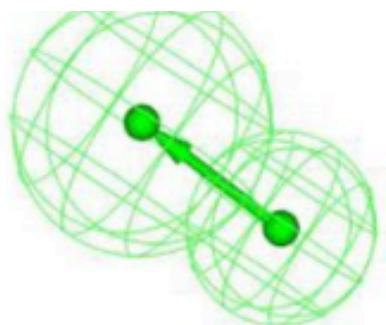
Pharmacophore mapping

Typical features:

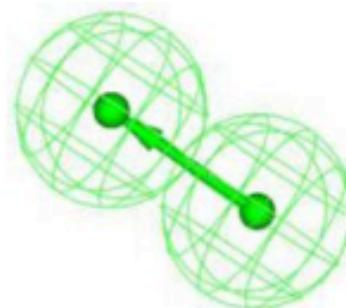
- Hydrophobicity
- Aromaticity
- H-bond acceptor
- H-bond donor
- Net charge
- Moieties



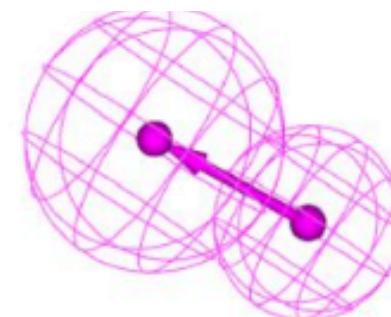
Pharmacophoric Features



H bond Acceptor



H bond Acceptor-Lipophilic



H bond Donor



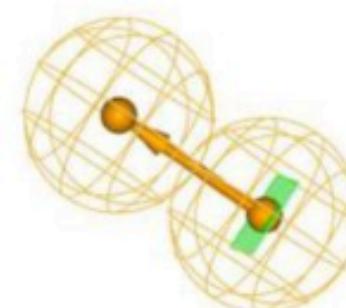
Hydrophobic



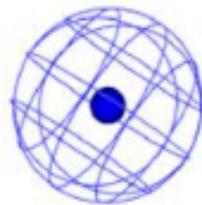
Hydrophobic-Aliphatic



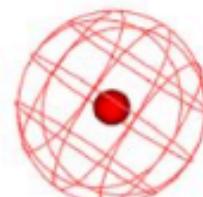
Hydrophobic-Aromatic



Ring aromatic



Negatively Ionizable



Positively Ionizable

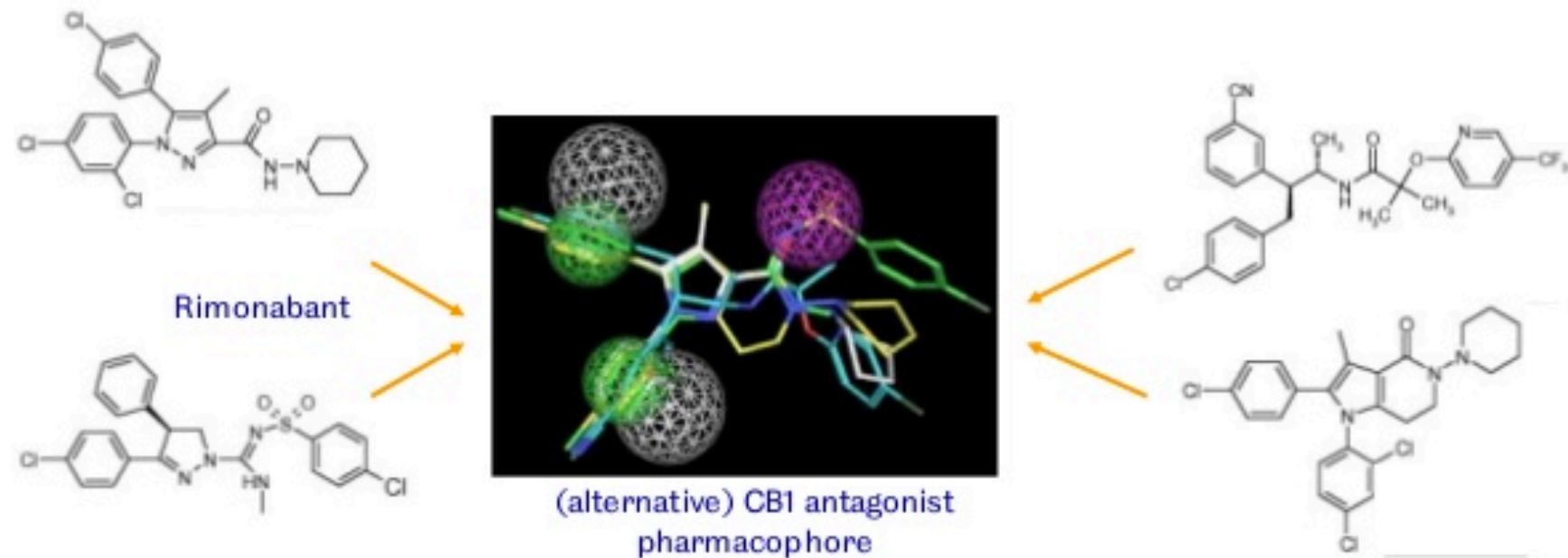


Negative Charge



Positive Charge

Generating pharmacophore models



Trying to predict how the ligands will bind to the receptor without knowing the structure of the receptor.

Pharmacophore generation methods

Pharmacophoric features in each ligand identified:

- Donors, acceptors, hydrophobic groups...
- Often SMARTs-based to allow user-definitions

Ligands aligned such that corresponding features overlaid

Conformational space explored

- On-the-fly using e.g. genetic algorithm
- Generating ensemble of conformations with each conformer in turn.

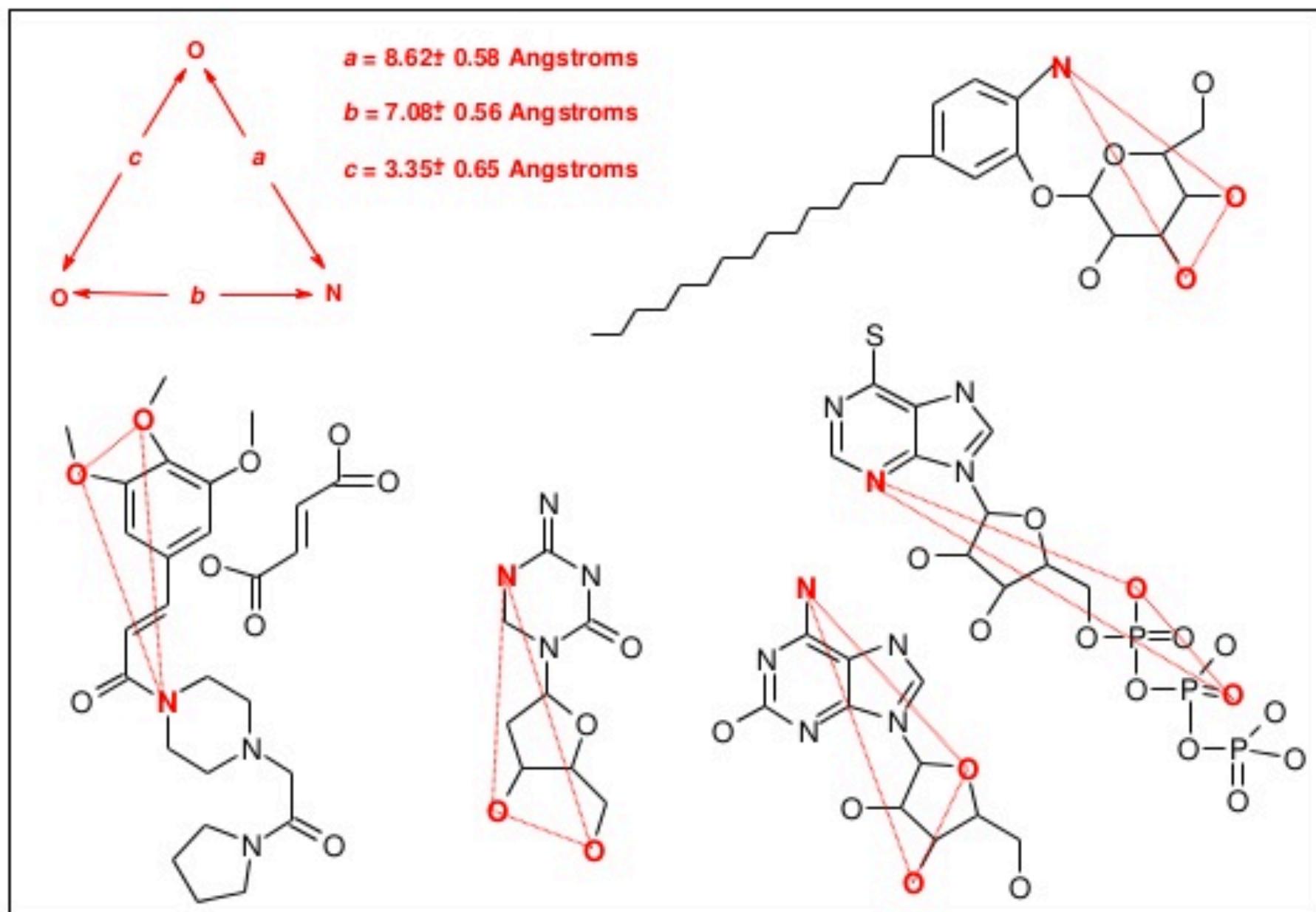
Pharmacophore hypotheses are scored.

e.g. number of features, goodness of fit to features, conformational energy, volume of the overlay...

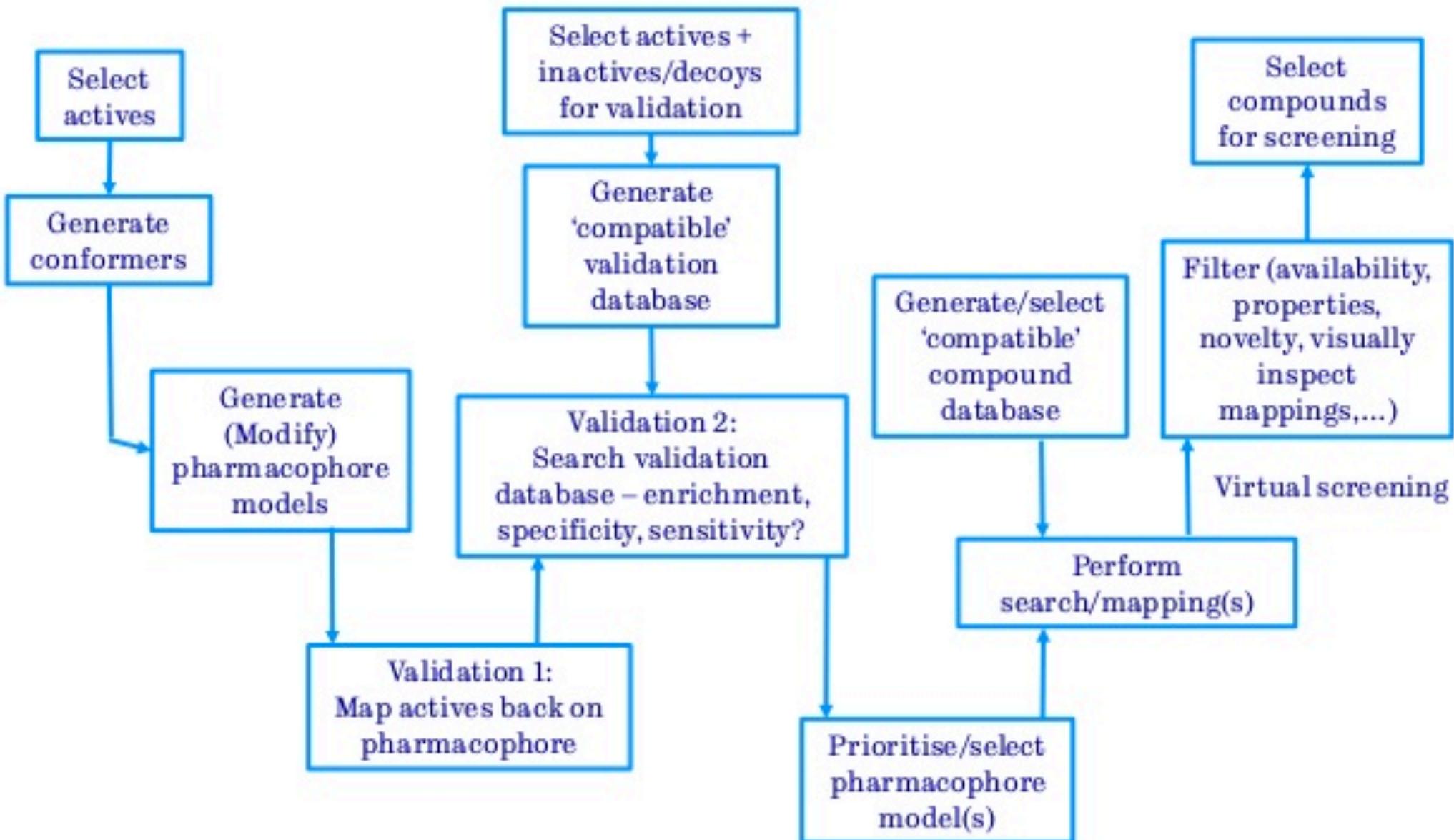
Practical aspects

1. Select ‘a representative’ set of actives
 - Most methods assume similar binding modes
 - One or more rigid molecules are preferred
 - Ligands should be diverse
2. Molecule preparation (tautomeric form, protonation state, 3D structure, generation of conformations)
3. Use pharmacophore tool to generate pharmacophores (biased or unbiased?)
4. Select preferred pharmacophore model(s) and validate it/them
 - Visual inspection
 - Do the “actives” fit the pharmacophore model?
 - Can the model separate actives from decoys?

Pharmacophore Searching



Pharmacophore Searching Workflow



Software Examples for Pharmacophore Searching

Software	Source	Recent published use cases
Catalyst (Discovery Studio)	Accelrys	http://dx.doi.org/10.1007/s00894-011-1105-5 http://dx.doi.org/10.1016/j.bmcl.2010.12.131
GASP	Tripos	http://dx.doi.org/10.1016/j.jmgm.2010.02.004
GALAHAD	Tripos	http://dx.doi.org/10.1016/j.bmc.2011.09.016 http://dx.doi.org/10.1016/j.ejmech.2010.09.012
Ligandscout	Inte:ligand	http://dx.doi.org/10.1016/j.epilepsyres.2011.08.016
MOE	Chemical Computing Group	http://dx.doi.org/10.1007/s10822-011-9442-0 http://dx.doi.org/10.1016/j.ejmech.2010.07.020
Phase	Schrödinger	http://10.1111/j.1747-0285.2011.01130.x http://cs-test.ias.ac.in/cs/Volumes/100/12/1847.pdf

Program

Molecular representations

Molecular Similarity

Pharmacophore Modelling

Quantitive Structure-Activity Relationship

Quantitative Structure-Activity Relationship (QSAR)

QSAR modelling gives the mathematical relation between structural attributes and target response for a set of chemicals. Use knowledge of known active and inactive compounds to build a predictive model.

Established QSAR models (Hansch-Fujita, Free-Wilson) involve the correlation of various dimensional features (e.g. 2D, 3D) with biological activity. Generally restricted to small, homogeneous datasets such as lead optimisation.

In general, qualitative approaches more than quantitative ones.

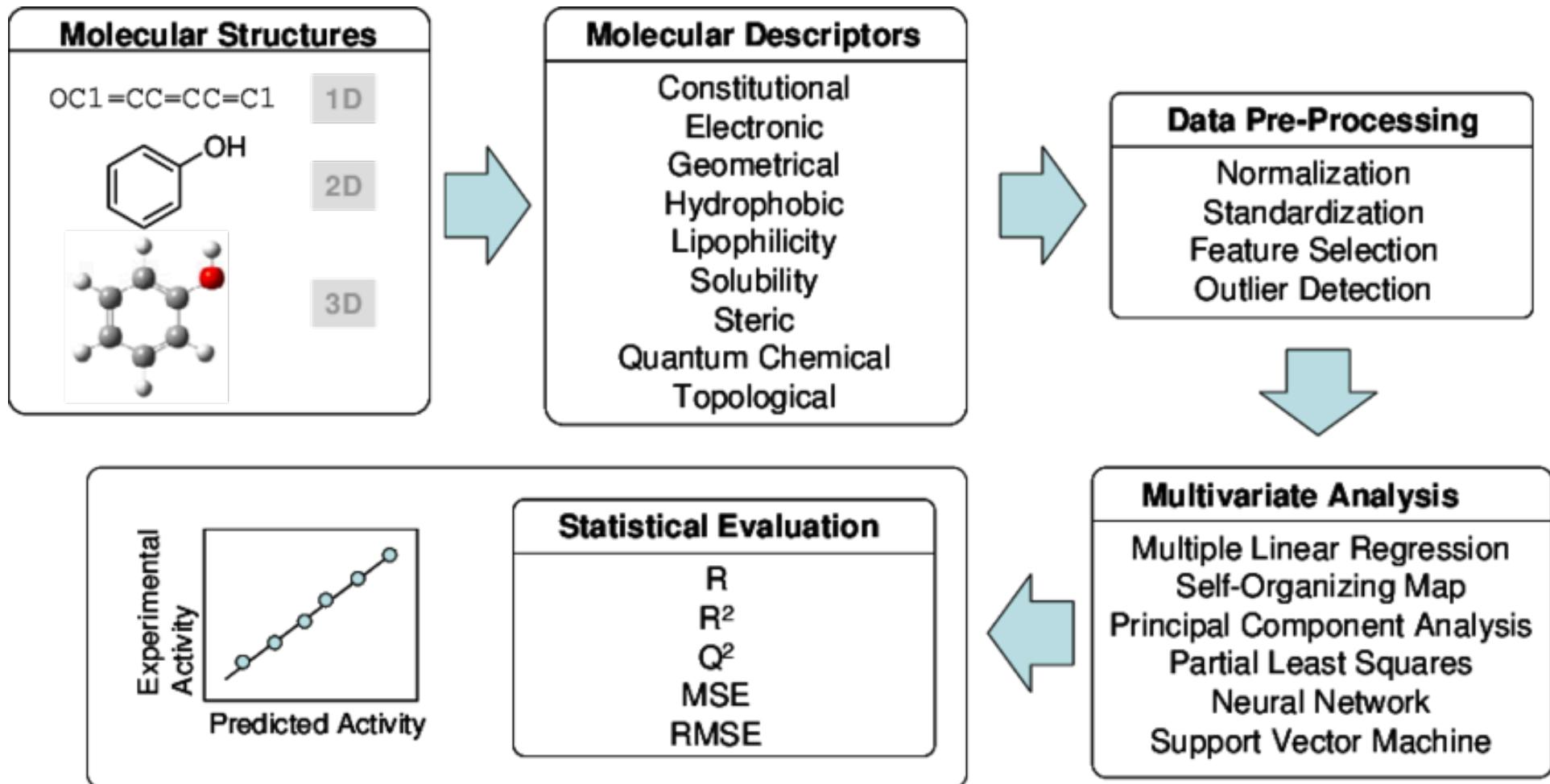
Can be used with data consisting of diverse structural classes and multiple binding modes.

Resulting models used to prioritise compound for lead finding.

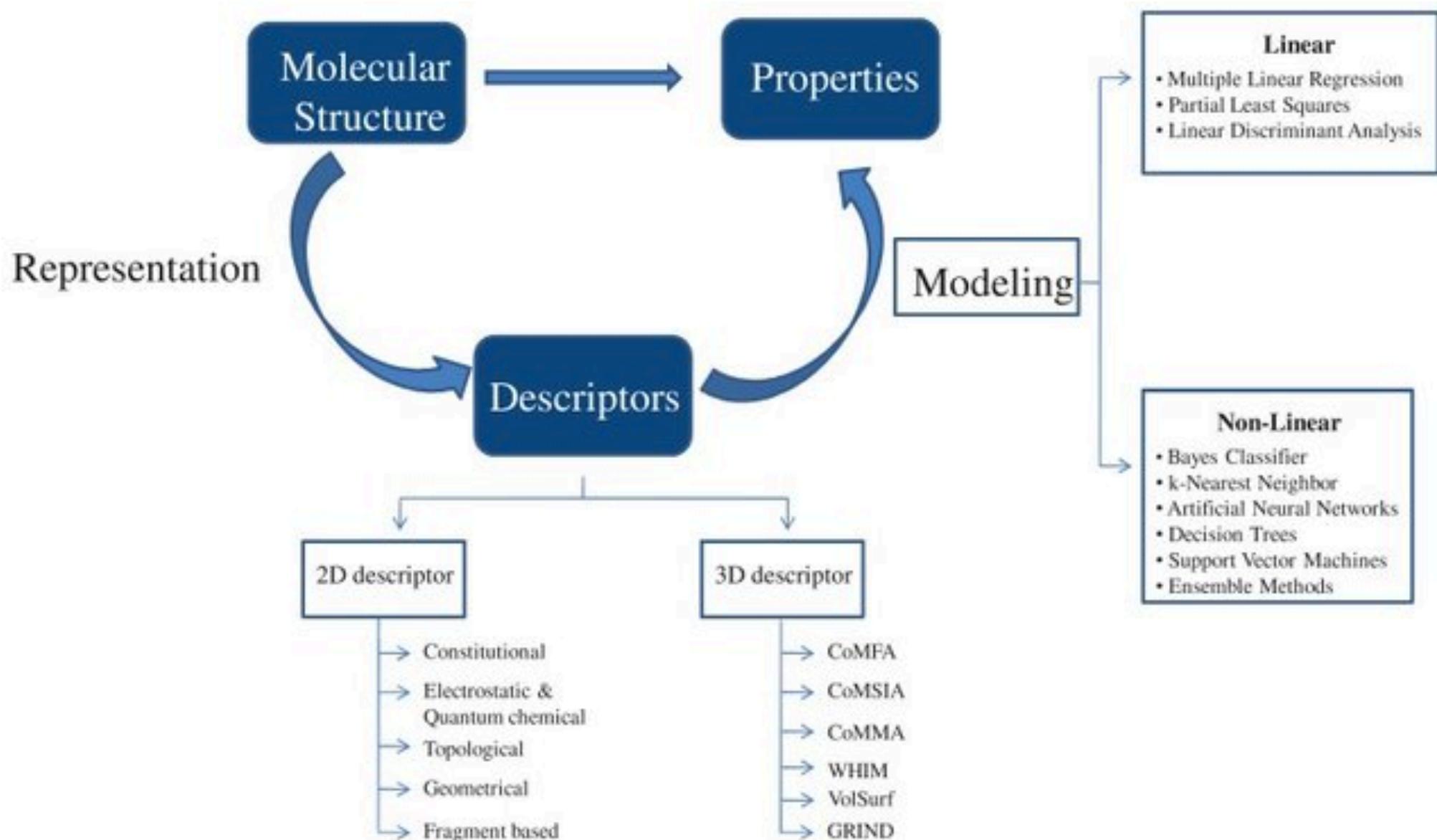
Typical QSAR Workflow

1. **Database Search:** list active compounds binding the desired drug target and their activities.
2. **Data Preparation:** collect physicochemical features that may be affecting the biological activity.
3. **Model Building:** apply machine learning algorithms between physicochemical features and biological activity.
4. **Validation** of the QSAR predictive model(s).
5. **Use QSAR models** to optimise known active compounds and maximise the biological activity.
6. **Experimental validation.**

Typical QSAR Workflow

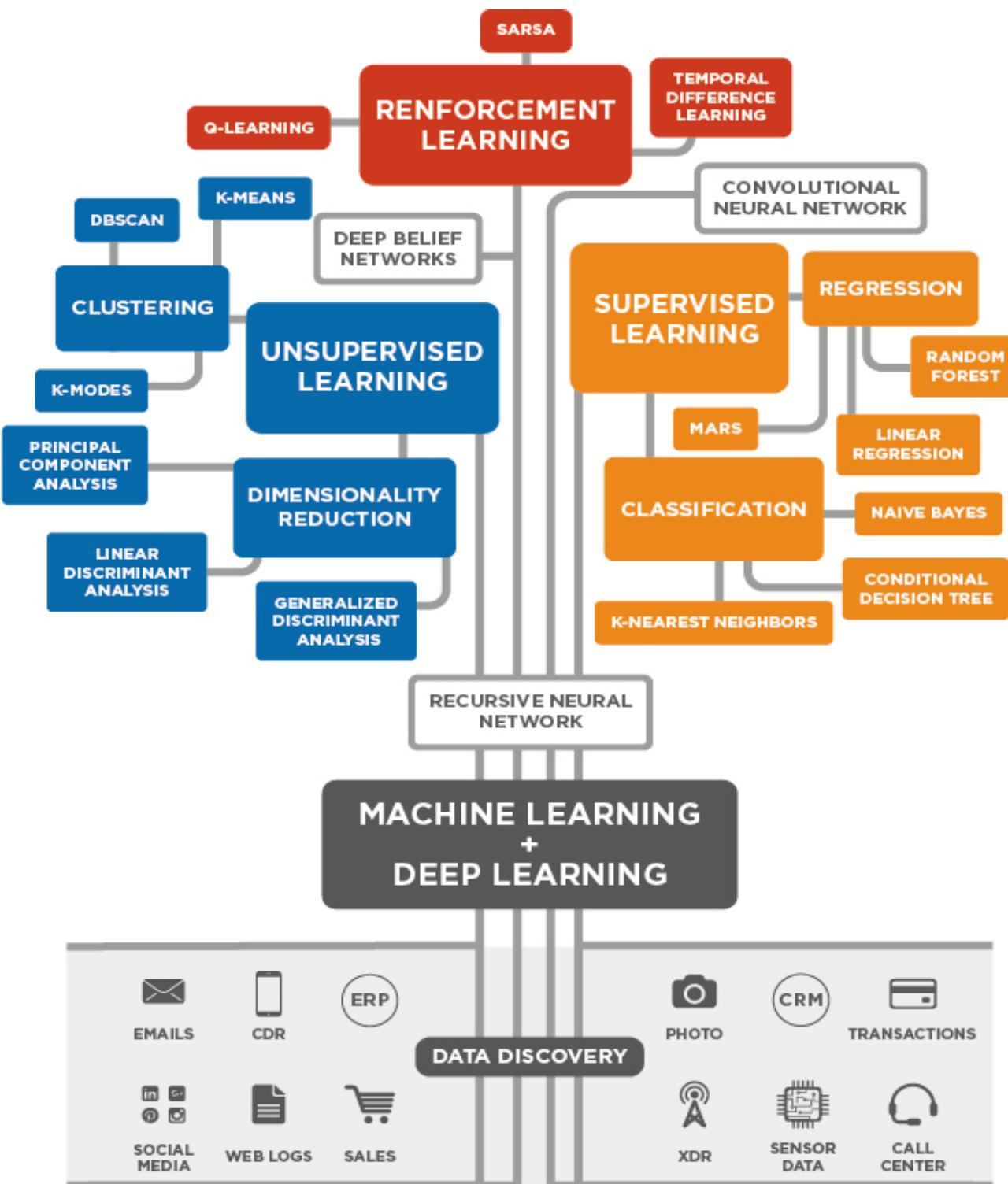


3D-QSAR methodologies

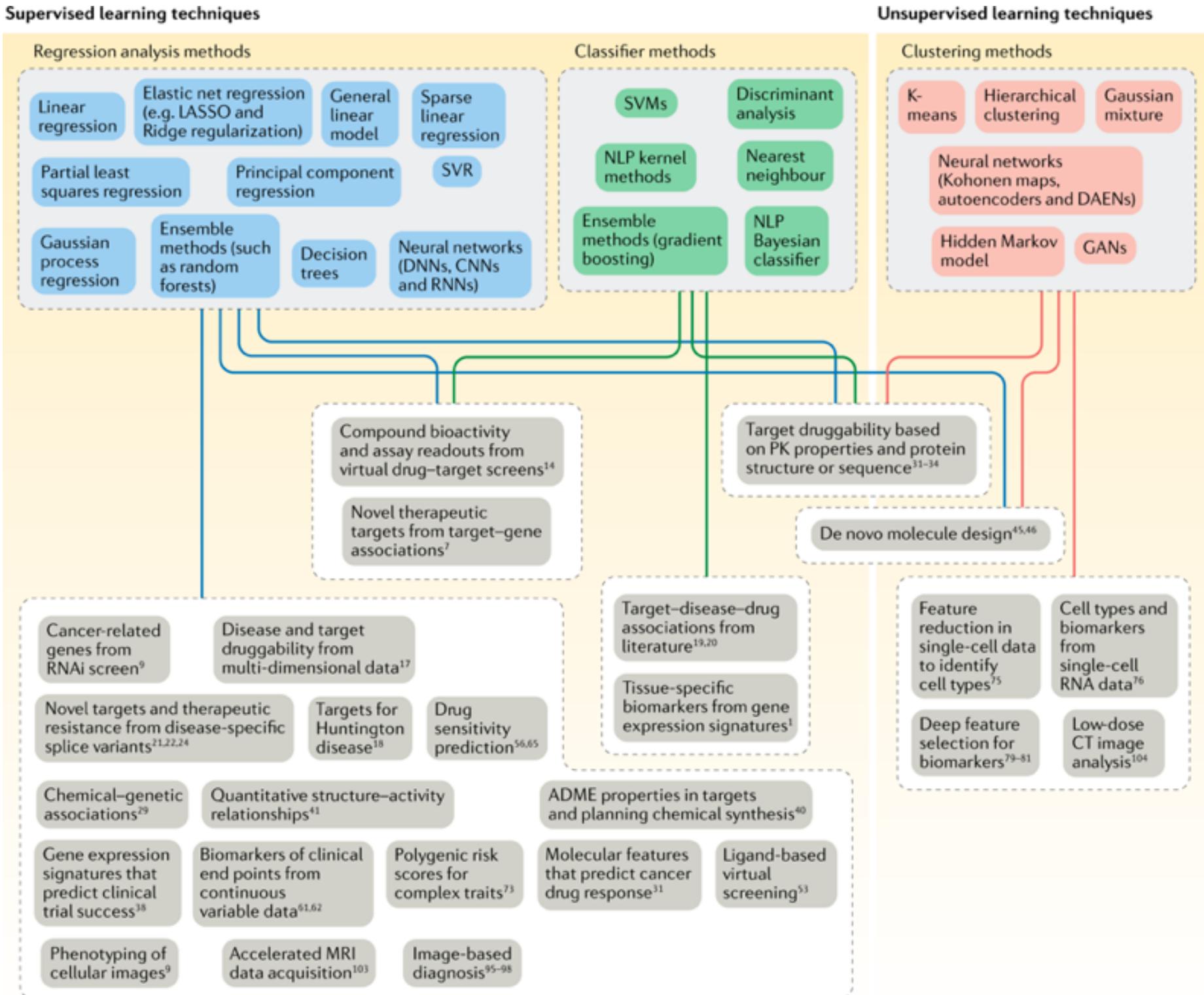


Molecular Descriptors

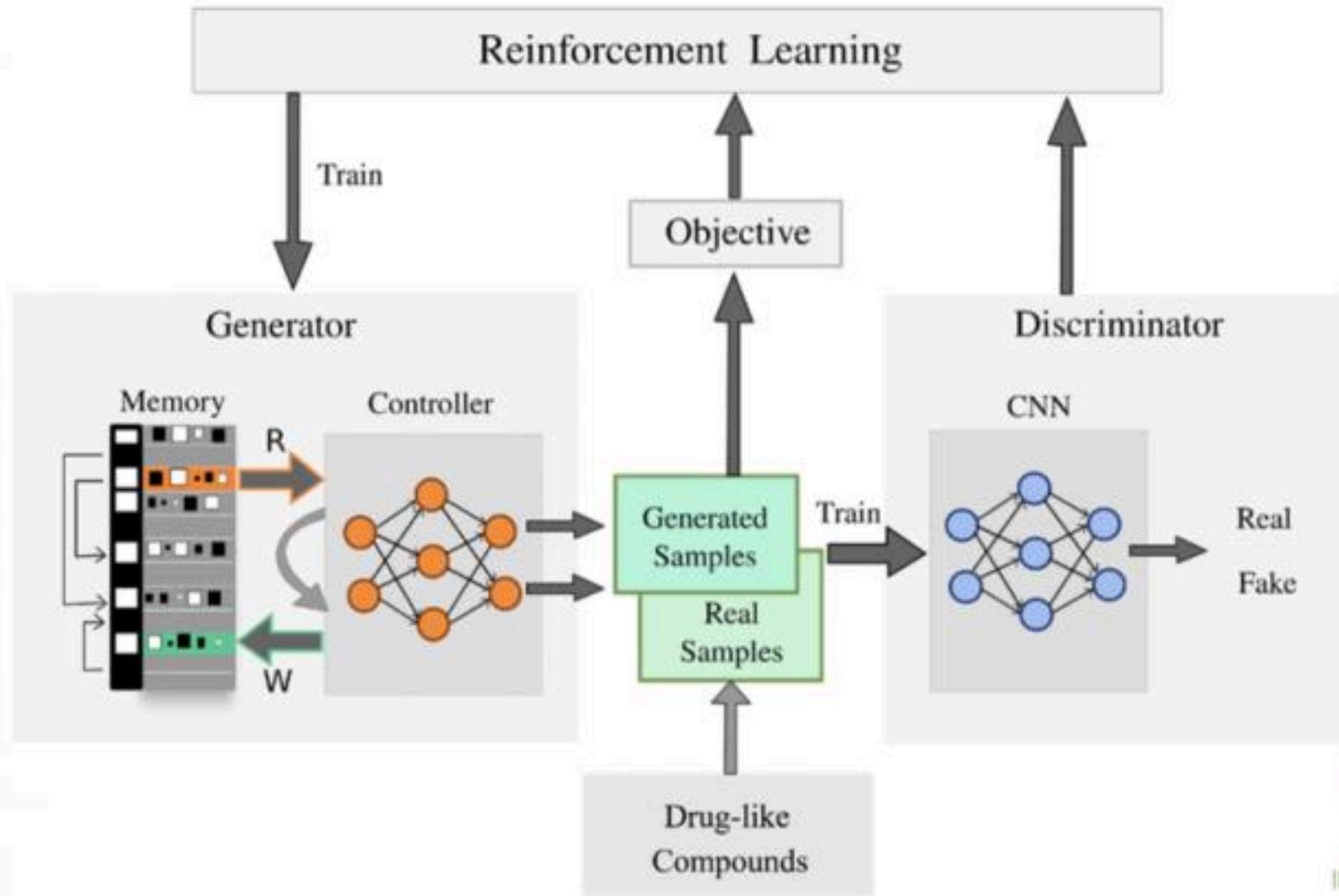




Machine Learning in Drug Discovery

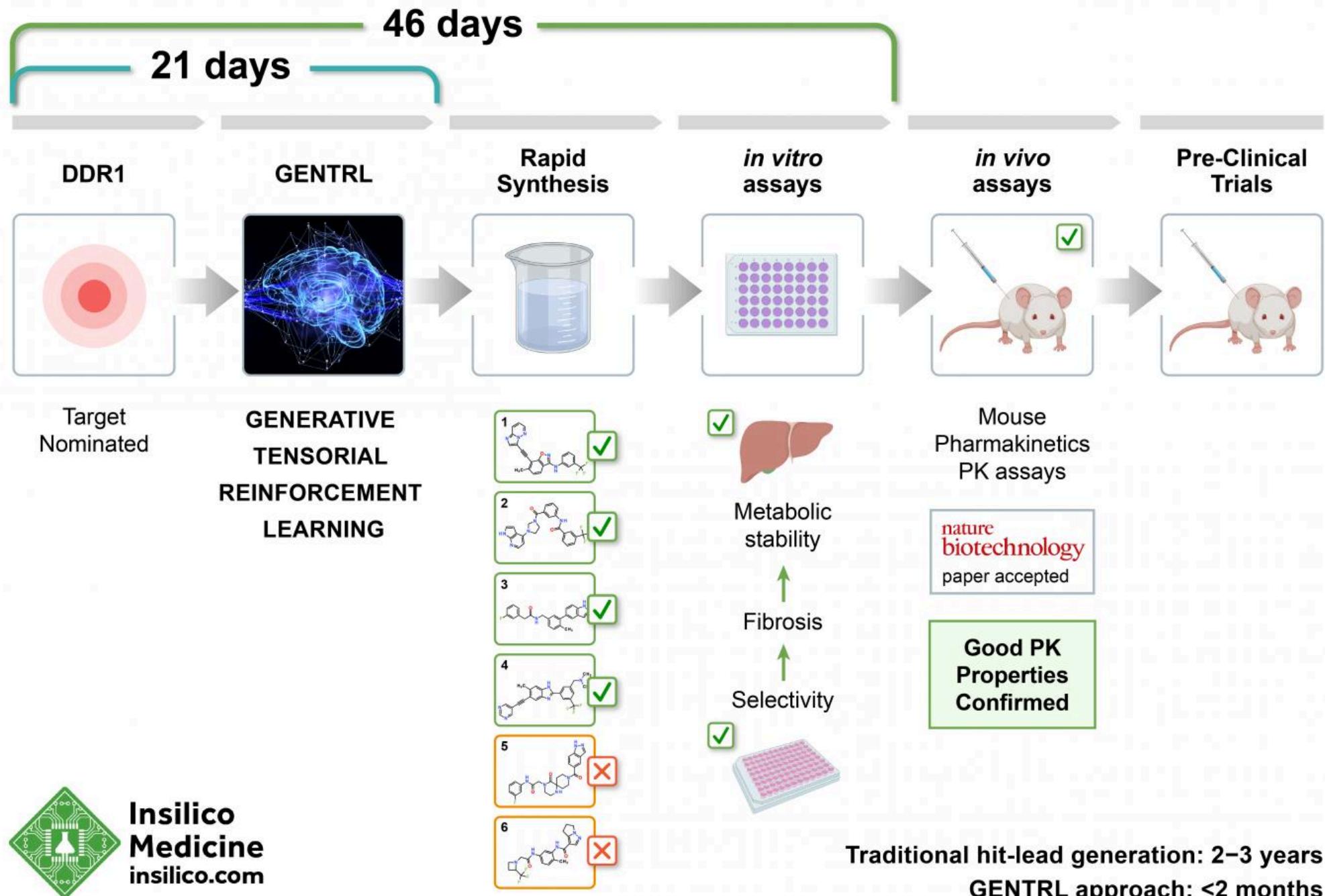


Machine Learning in Drug Discovery



INSILICO MEDICINE
insilico.com

DEEP LEARNING ENABLES RAPID IDENTIFICATION OF POTENT DDR1 KINASE INHIBITORS



**Insilico
Medicine**
insilico.com

References

1. Nathan Brown, - Henry Stewart Talks
2. Leach et al. J. Med. Chem. (2010), 53, 539-558.
3. Seidel et al. Drug Discovery Today: Technologies (2010), 7, e221-e228.
4. Hessler et al. Drug Discovery Today: Technologies (2010), 7, e263-269.
5. Hein et al. Drug Discovery Today: Technologies (2010), 7, e2229-2236.
6. Caporuscio F, Tafi A. Curr. Med. Chem. (2011), 18, 2543-2553.
7. Wallach I. Drug Dev. Res. (2011), 72, 17-25.
8. Val Gillet - University of Sheffield (2015), Ligand Based and Structure-Based Virtual Screening (on SlideShare)