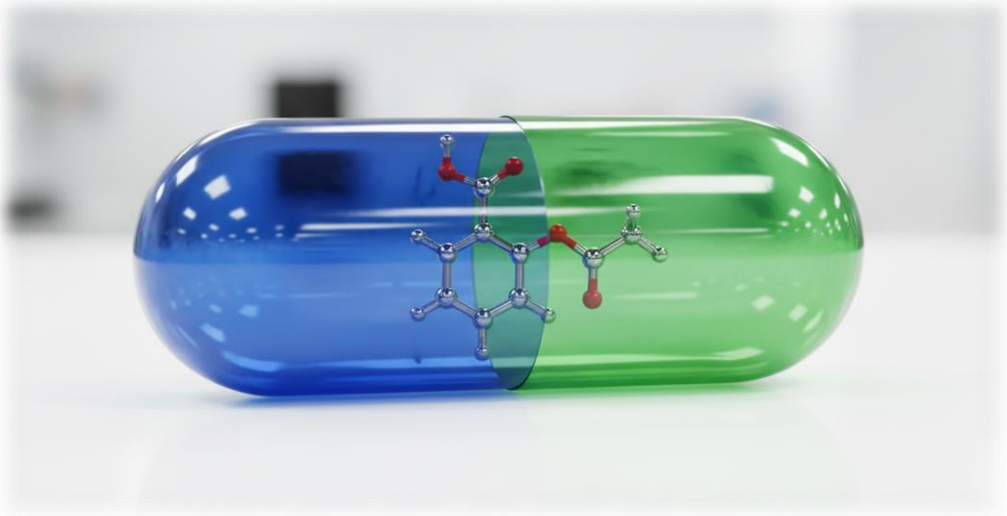
CADD – Computer aided drug design



Course structure

- 5 Theoretical classes (2h each)
- 5 Practical classes (2h each)
- Topics:
  - Introduction to Computational Drug Design
  - Molecular representations, formats and descriptors
  - Molecular similarity, fingerprinting and scaffolds
  - Programmatic access to Chemical Databases (ChEMBL, PubChem)
  - Molecular mechanics, energy minimization and molecular dynamics
  - Molecular docking and virtual screening
  - Machine Learning in Drug Design

- Practical classes use *Colab Notebooks*, offering fre and online computational capacity

Introduction to Drug Design and Computational Approaches

Drug design and CADD

- **What is Drug Design?**
  Systematic process of identifying molecules with therapeutic potential, through a combination of biological, chemical and computational techniques

- **What is Computer-Aided Drug Design (CADD)?**
  Drug design process that uses computational techniques, tools and models to facilitate the discovery of new drugs.

Drug discover *versus* development

- **Drug discovery** - all the experimentation and studies designed to move a program from the initial identification of a biological target and associated disease state to the identification of single compound with the potential to be clinically relevant.

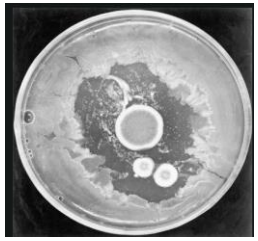
- **Drug Development** - typically begins once a single compound has been identified, which is then progressed through various studies designed to support its approval for sale by the appropriate regulatory bodies

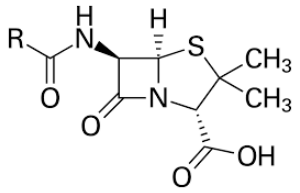
Rational drug design

- **Rational drug design** – discovery of new drugs based on the knowledge of structure and mechanism rather than trial and error

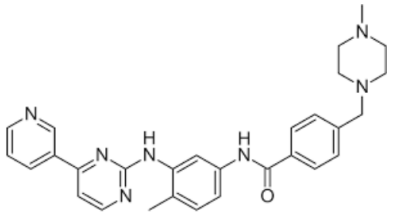
- **Empirical or "irrational" drug discovery** – random screening of natural compounds or fortuitous observation of new effect (*serendipity)*

- Examples:
  - Penicilin (fortuitous) :





  - Imatinib (rational drug design)

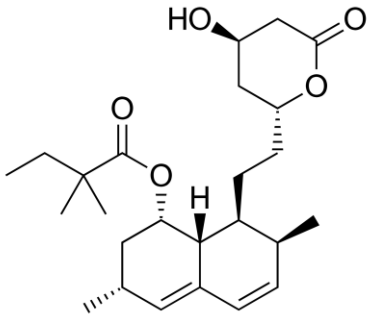


Disease -> Mechanism –> Target -> Drug

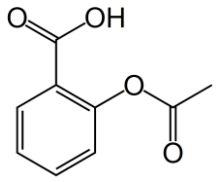
How are new drugs found ?

- Natural products (e.g. Aspirin)
- Screening assays
- Synthetic chemistry
- Combinatorial chemistry
- Similarity with know drugs (“Me too” drugs)
- Re-purposing (searching known drugs for a new effect)
- Serendipity:
  - Drugs found by chance (e.g. Penicilin)
  - Unforeseen side-effect of a drug or candidate

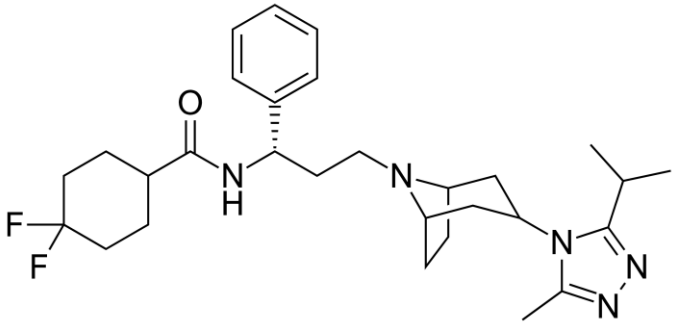
Drugs found by different methods



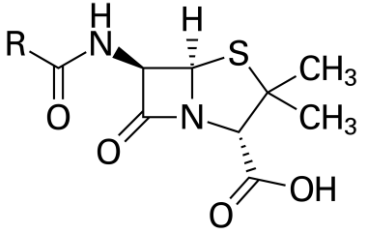
Simvastatin
(“me too”)



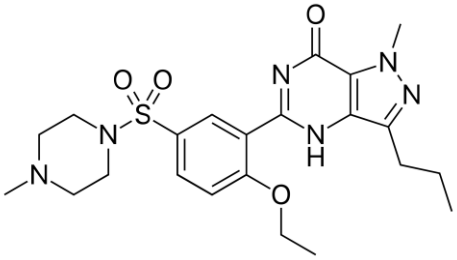
Aspirin
(natural product)



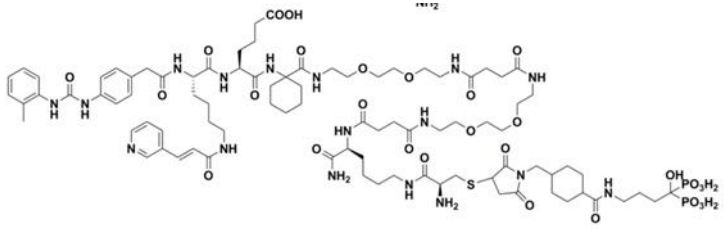
Maraviroc
(HTS assay)



Penicillin
(serendipity)



Sildenafil
(repurposing)

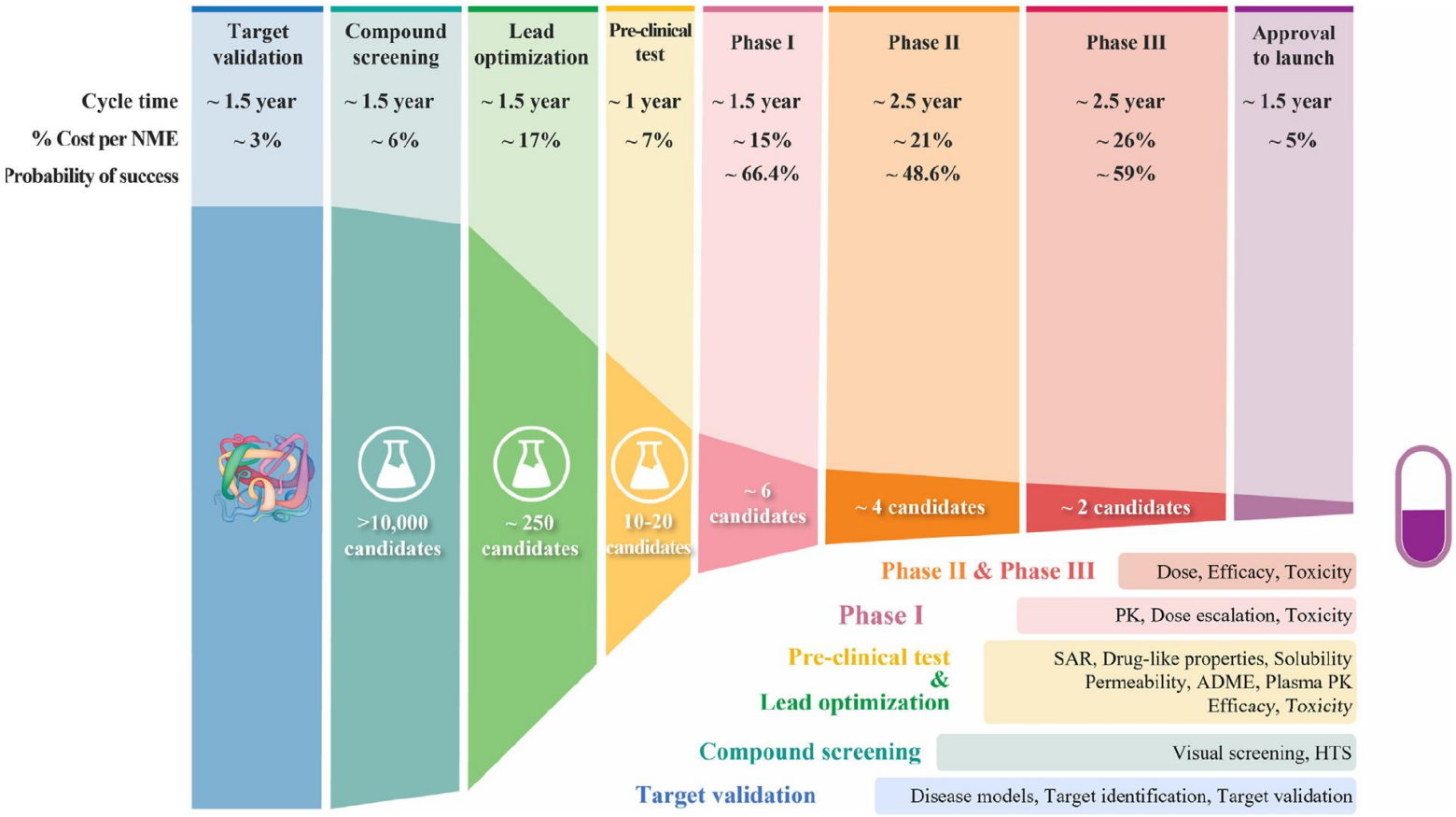


LLP2A-Ale
(combinatorial chemistry)

The challenge of drug discovery and design

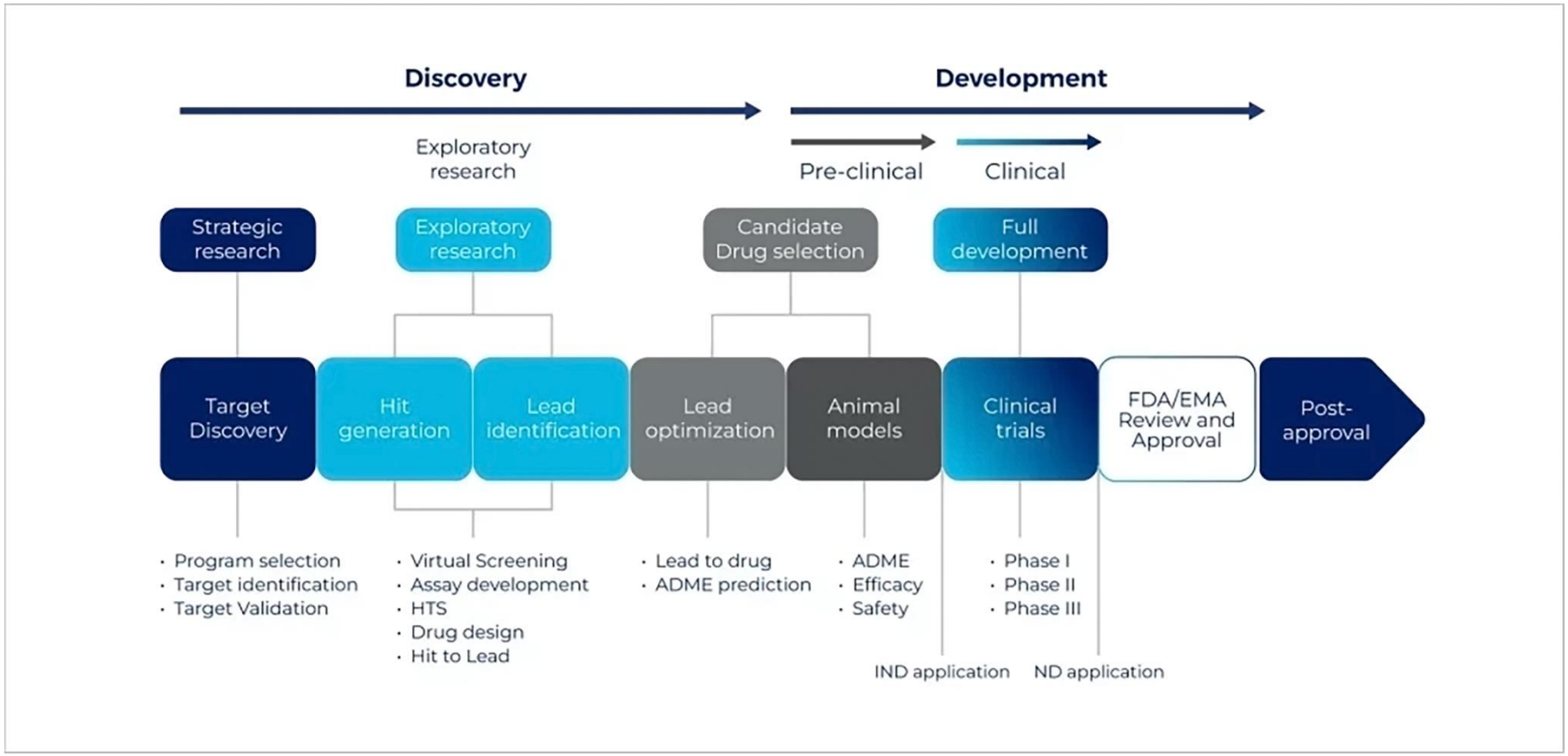
- The task of discovering new drugs is hard, expensive, lengthy and dependent on a very large number of scientific disciplines, techniques and expertise.
- Millions of compounds may have to be screened in activity tests to select but a few candidates (hits), of which only a few show promise as drug candidate (leads).
- Lengthy and thorough clinical testing in both animals and humans is required, without guarantee of approval by the regulatory entities.
- Millions (or billions) of dollars and ~5-15 years are required for the whole process.
- A large share of the profit generate by the pharmaceutical industries comes from only a few drugs.
- Patent expiry narrows the profitability range of drugs and pushes the “me too” drug concept

The drug discovery pipeline/ funnel



Sun, Duxin, et al. *Acta Pharmaceutica Sinica B* 12.7 (2022): 3049-3062.

Drug Discovery & Development



Drugs are expensive

JAMA | Original Investigation

Estimated Research and Development Investment Needed to Bring a New Medicine to Market, 2009-2018

Olivier J. Wouters, PhD; Martin McKee, MD, DSc; Jeroen Luyten, PhD

Mean Cost:
$1.3 billion

Table 4. Mean And Median Expected Research and Development Expenditure on New Therapeutic Agents Approved by the US Food and Drug Administration (2009-2018) by Therapeutic Area

Therapeutic Area[a]	Sample Size	Expenditure in US$, Millions (95% CI)[b]	
		Median	Mean
Antineoplastic and immunomodulating agents	20	2771.6 (2051.8-5366.2)	4461.2 (3114.0-6001.3)
Alimentary tract and metabolism	15	1217.6 (613.9-1792.4)	1430.3 (920.8-2078.7)
Nervous system	8	765.9 (323.0-1473.5)	1076.9 (508.7-1847.1)
Antiinfectives for systemic use	5	1259.9 (265.9-2128.3)	1297.2 (672.5-1858.5)
Dermatologicals	4	747.4	1998.3
Cardiovascular system	3	339.4	1152.4
Musculoskeletal system	3	1052.6	937.3
Blood and blood-forming organs	2	793.0	793.0
Sensory organs	2	1302.8	1302.8
Other[c]	1	1121.0	1121.0

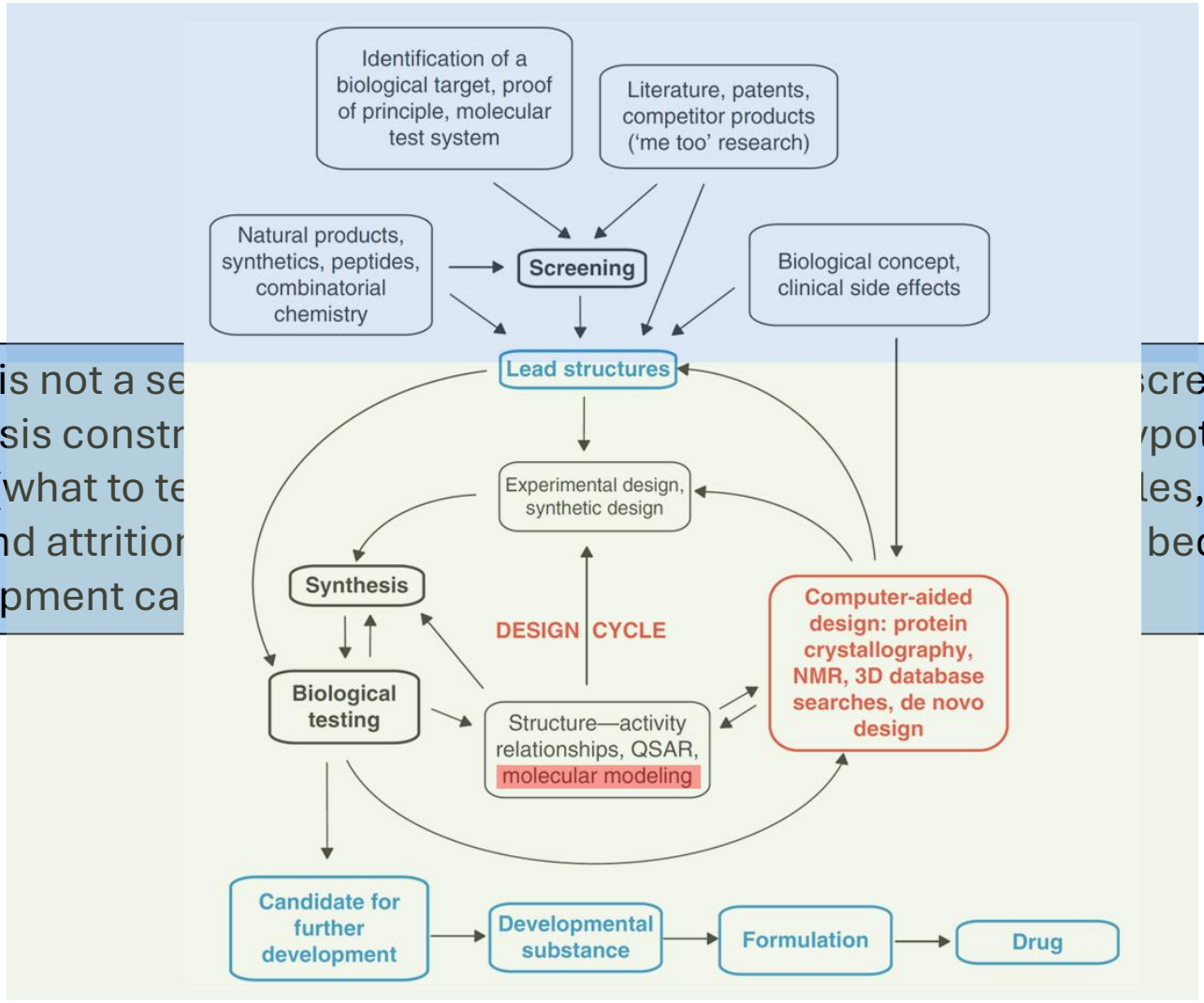
Wouters(2020) *J.Am.Medical Assoc.*, 323:884

Where does CADD fit ?

- Target modeling → homology modeling, molecular dynamics
- Hit discovery → virtual screening, de novo design, similarity search
- Lead optimization → QSAR, free energy calculations, QM calculations
- ADME/Tox → in silico property prediction, predictors, machine learning / AI

CADD does not replace the experiments – *it guides them !*

The Drug Design cycle and CADD



CADD is not a se creening data, synthesis constr potheses faster (what to te les, lowers cost and attrition becomes a development ca

Structure-based *versus* ligand-based DD

- **Structure-based Drug Design (SBDD):** relies knowledge of the 3D *structure* of the *target* and ligand to predict and optimize activity (docking, virtual screening, molecular dynamics

- **Ligand-based Drug Design (LBDD):** relies on *similarity* to known actives to find and expand new actives (chemical similarity, descriptors, pharmacophores, QSAR)

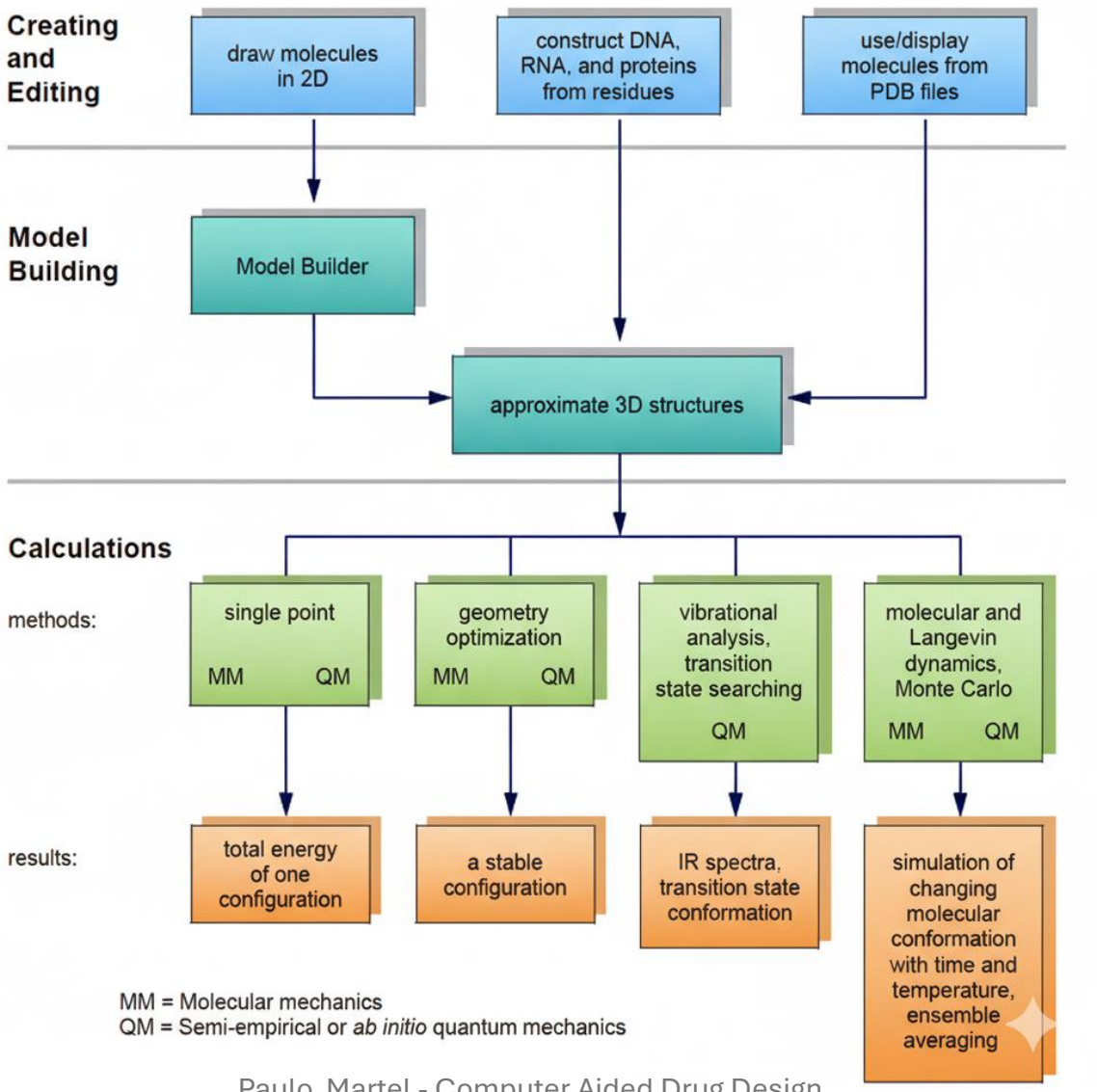
Examples:

- **SBDD:** designing kinase inhibitors from crystal structures
- **LDBD:** predicting analogues for GPCR ligands when no receptor structure is known

Core computational techniques for DD

- **Molecular modelling** – building and visualizing molecules
- **Molecular Mechanics** – molecular energetics (approximate) and conformation search
- **Quantum Chemistry** – electrostatics, reactivity, bond energy, rigorous molecular energetics
- **Molecular dynamics** – target flexibility, induced fit
- **Docking & Virtual screening** – binding prediction and energetics
- **Machine Leaning and AI** – predict properties, automated discovery, generate new molecules

Modelling flow



Key databases in CADD

- **PDB** – protein structures
- **Uniprot** – *annotated* protein sequences
- **PubChem** – small molecules + targets
- **ChEMBL** – small molecules + bioactivity data
- **DrugBank** – drugs + targets + pharmacology
- **ZINC** – virtual screening libraries (very large collections)
- **PDBind** – protein-ligand complexes and Ki / Kd
- **CCDB** – small molecule (crystal structures)

HIV protease inhibitors

1989

X-ray structure determination of HIVE protease

Visualizing the active site and figuring the mechanism

Candidate Binders

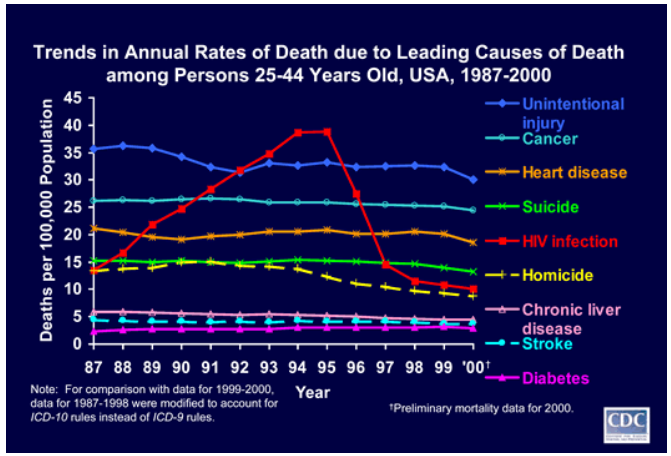
Computational docking

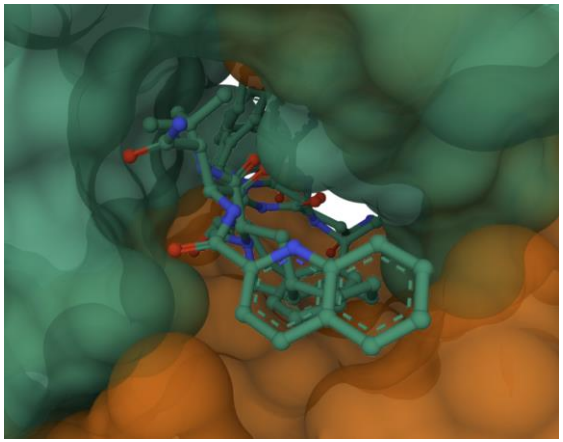
Synthesis

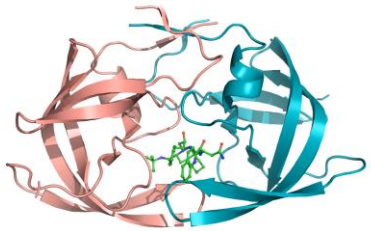
Testing

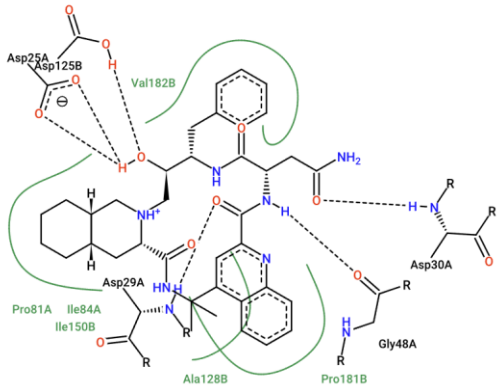
Saquinavir

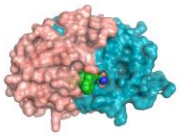
1995







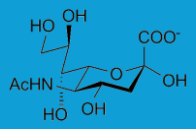




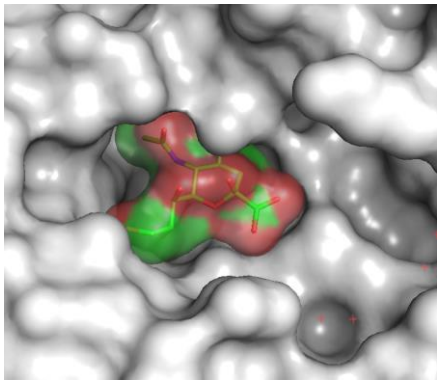
HIV protease inhibitors – CADD contribute

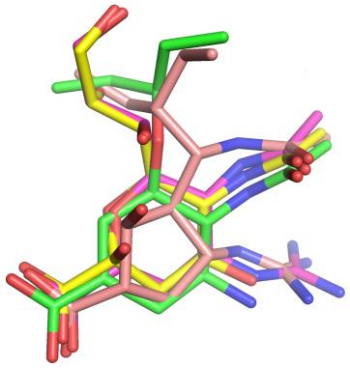
- **Accelerated the design process** - Rather than random screening, researchers could rationally design molecules
- **Provided visualization** - Scientists could "see" how molecules might interact with the target
- **Enabled prediction** - They could predict binding modes and relative affinities before synthesis
- **Guided optimization** - Structural data from crystallography combined with modelling guided improvements

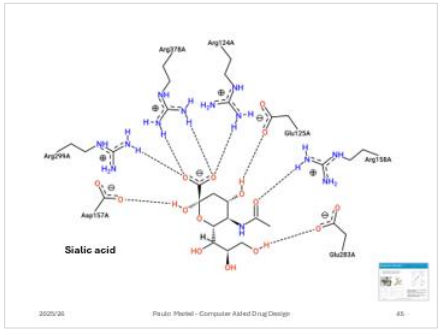
Oseltamivir (Tamiflu)

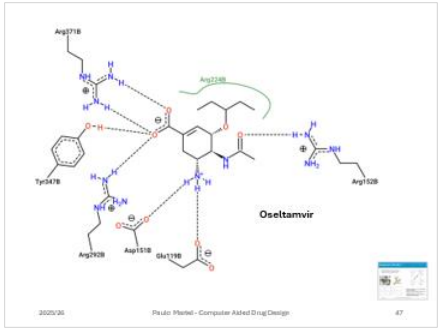


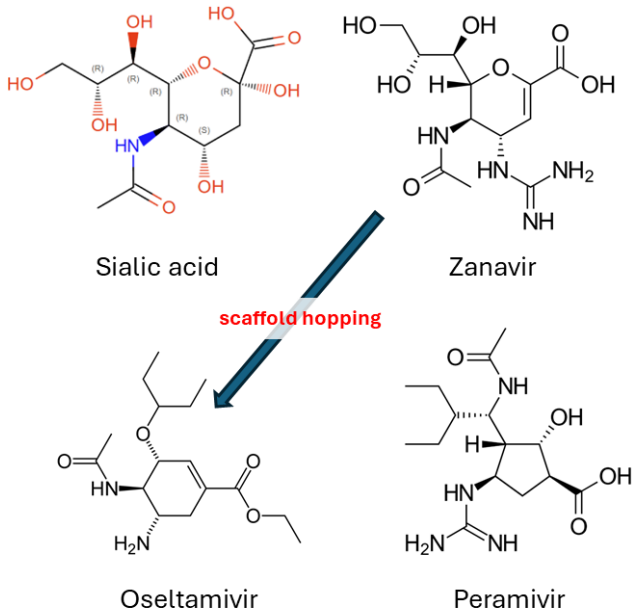
- Neuraminidase cleaves sialic acid from cell surface to allow flu virus particles to leave the host cell









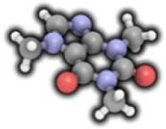
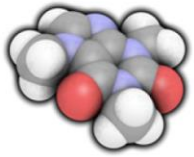


- Zanavir is too polar – poor oral araciality, taken by inhalation
- Oseltamvir can be taken via the oral route
- Both are *transition state analogues*
- **Scaffold hopping** to explore related, but chemically distinct entities.

Oseltamivir (Tamiflu) – CADD contribute

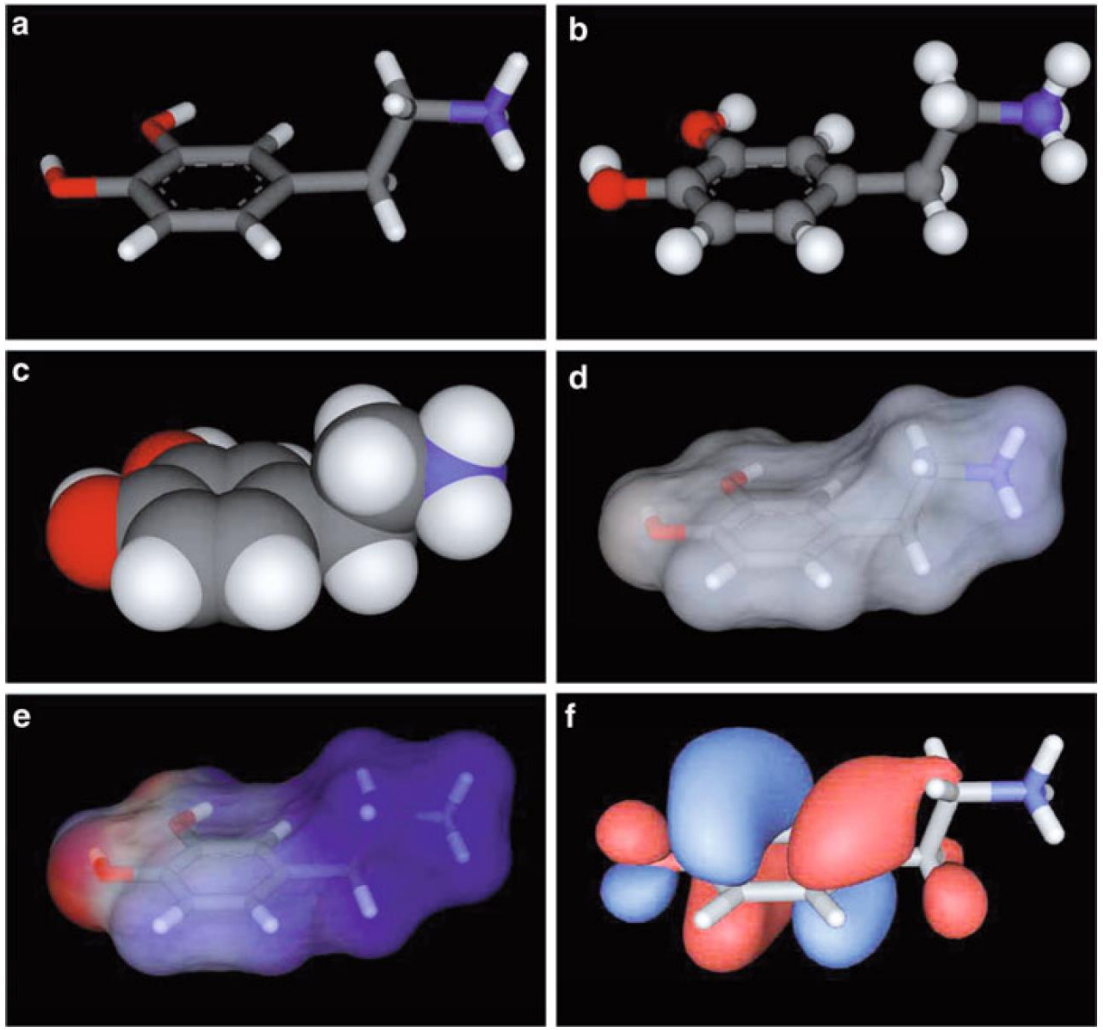
- Structure-based design using X-ray crystallography
- Molecular modeling and visualization
- Iterative design cycles
- More emphasis on **scaffold replacement** rather than substrate mimicry
- Significant focus on **predicting pharmacokinetic properties** (absorption, distribution)
- Computational chemistry helped identify which molecular features were essential vs. dispensable
- Greater emphasis on optimizing **drug-like properties** computationally

Representing Chemical Structures

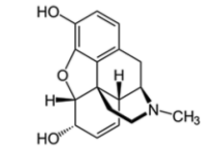
Representation Name	Representation of Caffeine
Common Name	Caffeine
Synonyms	Guaranine
	Methyltheobromine
	1,3,7-Trimethylxanthine
	Theine
Empirical Formula	$C_8H_{10}N_4O_2$
IUPAC Name	1,3,7-trimethylpurine-2,6-dione
CAS Registry Number	58-08-2
ChEMBL ID	CHEMBL113
Wiswesser Line Notation (WLN)	T56 BN DN FNVNVJ B F H
SMILES	CN1C=NC2=C1C(=O)N(C(=O)N2C)C
Aromatic SMILES	CN1C(=O)N(C)c2ncn(C)22C1=O
InChI	1S/C8H10N4O2/c1-10-4-9-6-5(10)7(13)12(3)8(14)11(6)2/h4H,1-3H3
InChIKey	RYYVLZVUVIJVGH-UHFFFAOYSA-N
Topography	
Surface	

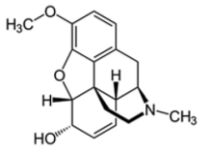
Visualizing Chemical Structures

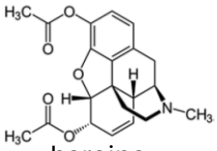
a – dreiding model
b – ball-and-stick
c – vdW (CPK)
d – molecular surface
e – surface potential
f – HOMO orbitals



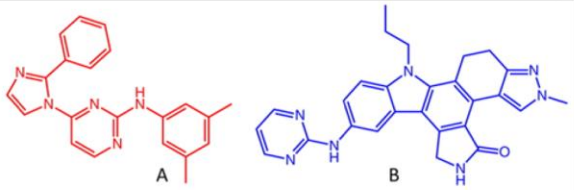
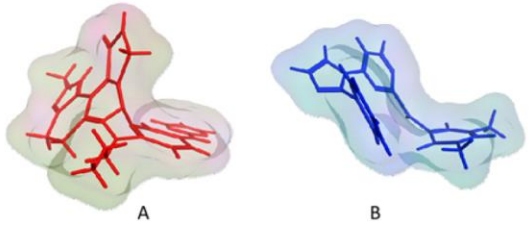
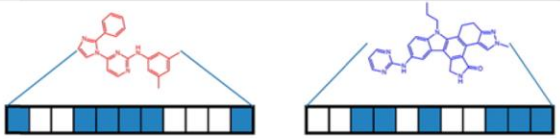
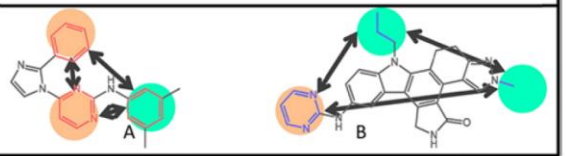
The importance of molecular similarity


morphine


codeine


heroine

Similar structures
similar functions

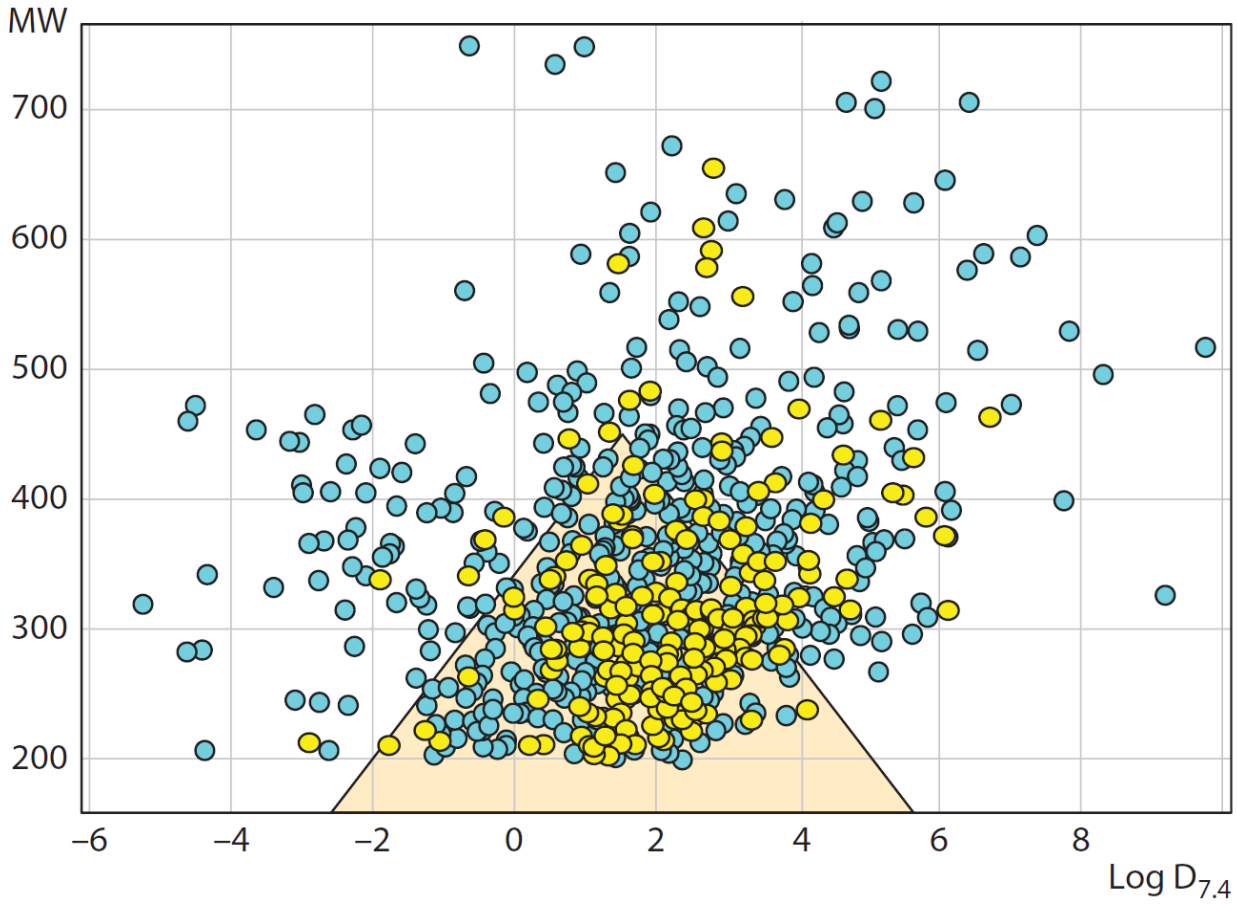
Chemical similarity		Mol. weight	LogP	Rotatable bonds	Aromatic rings	Heavy atoms
	A	341.4	5.23	4	4	26
	B	463.5	4.43	4	5	35
Molecular similarity / 2D similarity						
3D similarity						
Biological similarity		Vascular endothelial growth factor receptor 2	Tyrosine-protein kinase TIE-2			
	A	active	inactive			
	B	active	active			
Global similarity						
Local similarity						

Descriptors in chemical space

Finding the essential chemical descriptors (dimensionality reduction), classifying, filtering, selecting.
Machine learning-methods

Lipinski's rule of 5

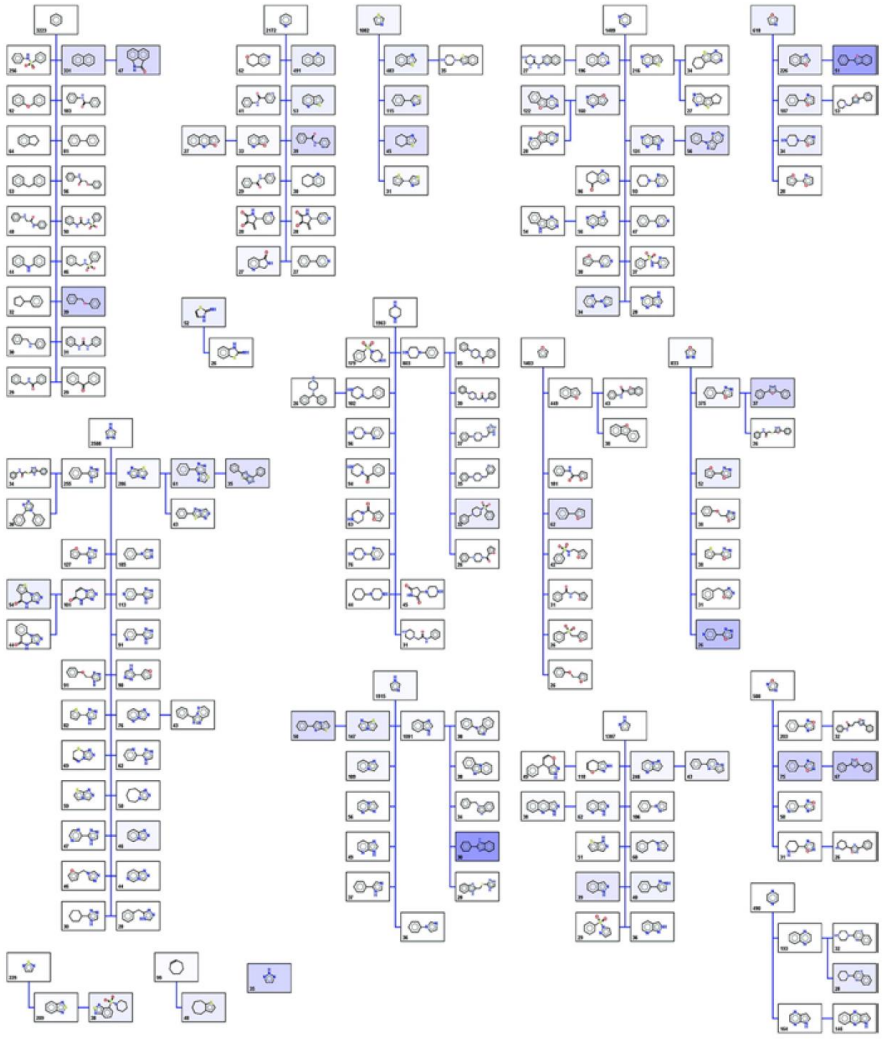
Peripheral drugs
84% Ro5 compliant 53% inside the Golden Triangle 70% have CNS MPO score > 4
CNS drugs
92% Ro5 compliant 77% inside the Golden Triangle 70% have CNS MPO score > 4



Searching through chemical space

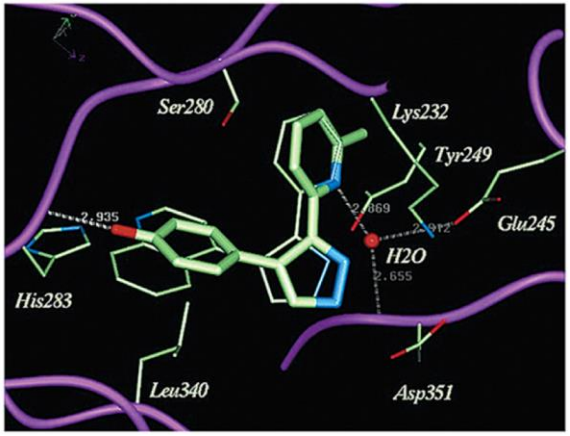
Scaffold - core structure or framework of a molecule that forms the central backbone to which various functional groups or substituents are attache

Scaffold hopping - Guided search through chemical space.

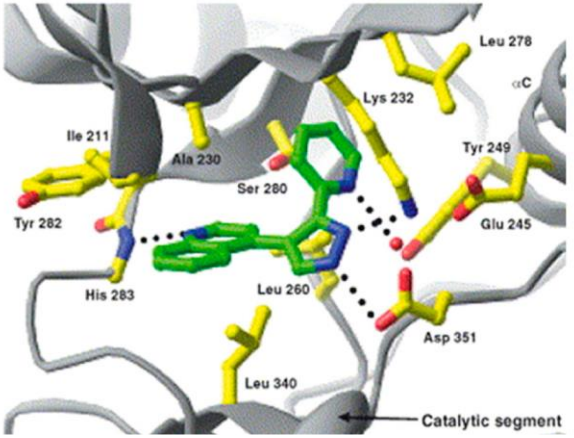


Molecular docking

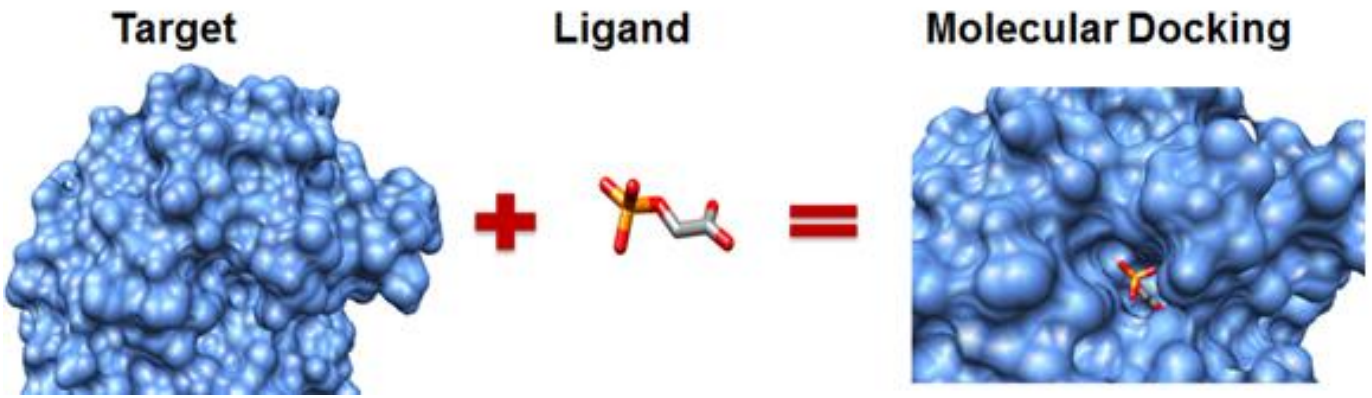
- Molecular docking can produce results similar to hight-throughput screening (HTS)
- Requires knowledge of ligand and target strcutures



HTS

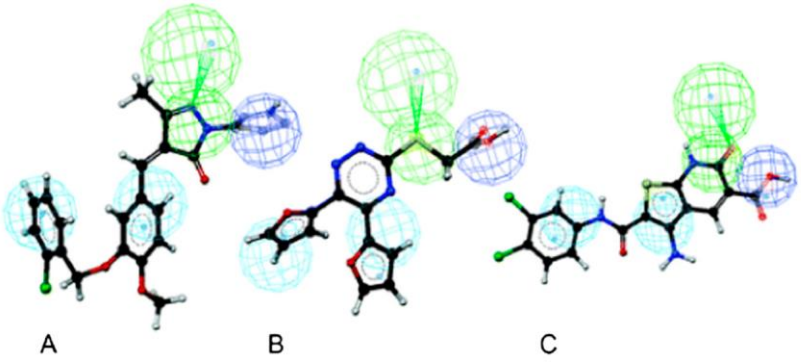


Docking



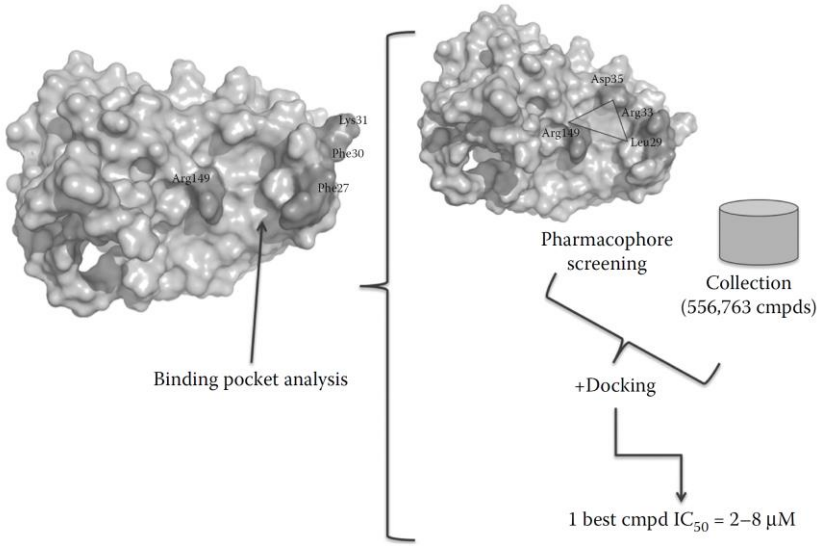
Computational pharmacophore screening

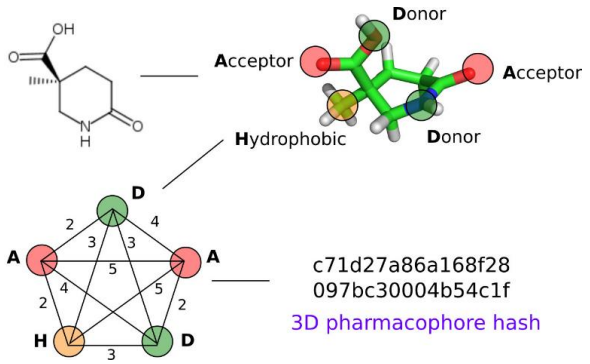
- **SBS** – pharmacophore derived from the structure of the biding site
- **LBS** – pharmacophore is derived from the structure of a known ligand



Ligand-based screening (LBS)

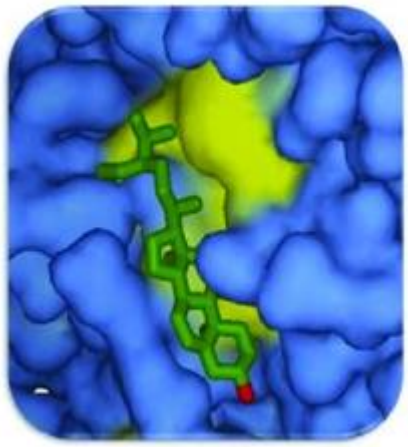
Structure-based screening (SBS)





Molecular conformation is important

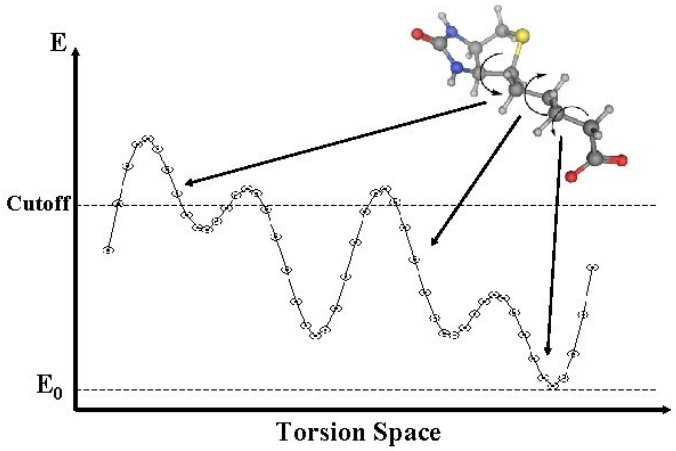
- Single conformations obtained from chemical databases don't tell the whole story!
- Molecular simulation methods can be used to explore the space of possible conformations and find **low-energy** conformations

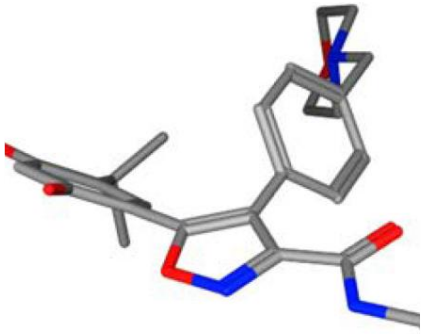


Molecule docks in the "right" conformation

Systematic Conformational Search

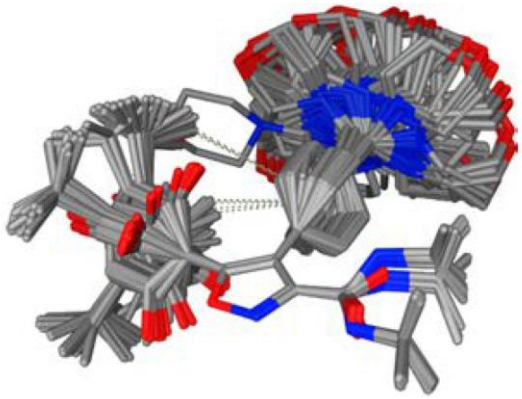
- Exhaustive incremental dihedral rotation search





Single conformation

Scanning Conformational Space



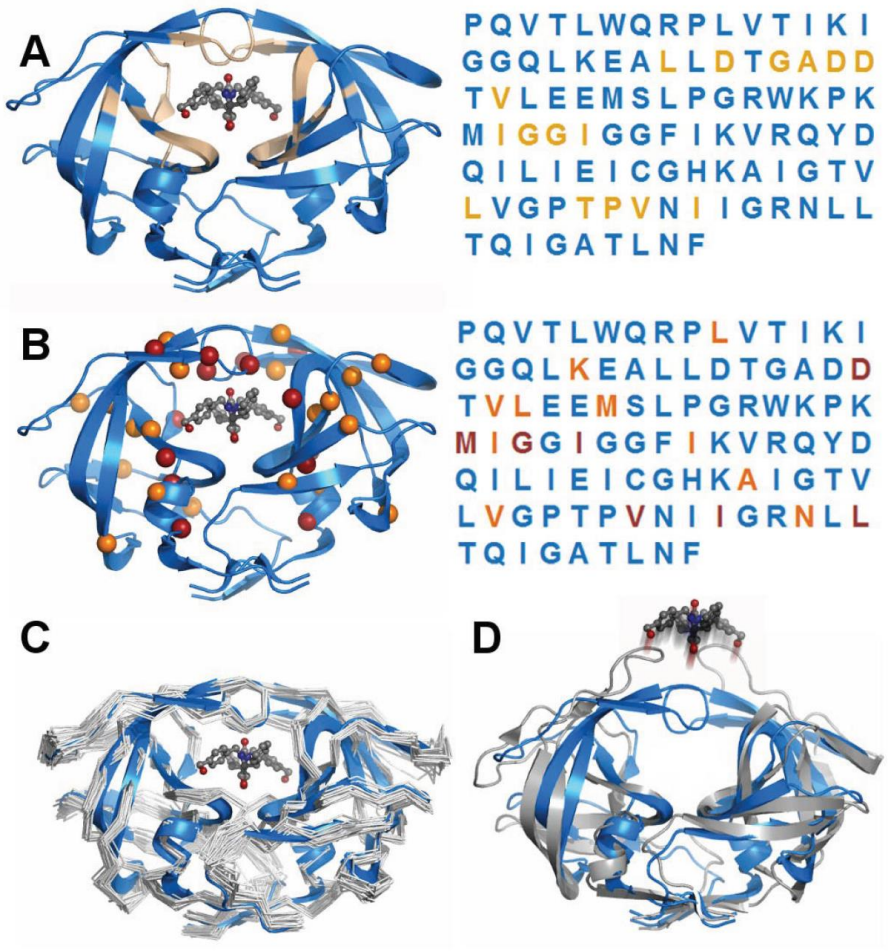
Multiple conformations

Sequence/structure analysis of targets

- Sequence analysis can reveal patterns characteristic of ligand binding and mutations affect function or stability
- Molecular dynamics (MD) studies of the HIV protease reveal the dynamics of the ligand-free protein and effect of mutations on flexibility

HIV protease

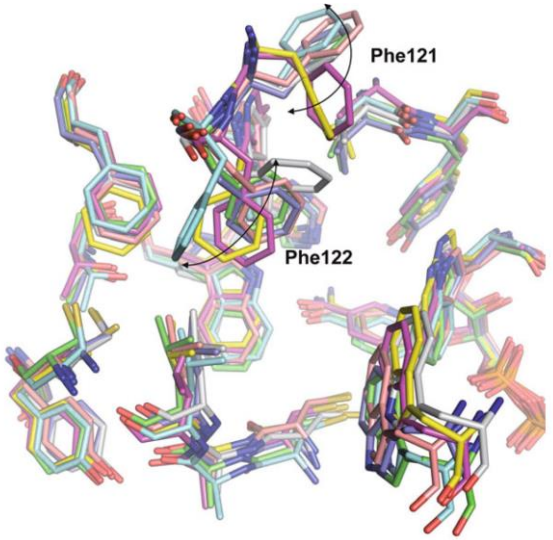
A – Residues near bound inhibitor
B – Mutations leading to resistance
C – Mutations can affect flexibility
D – Dynamics of ligand free protein (studied by *MD simualtions*)

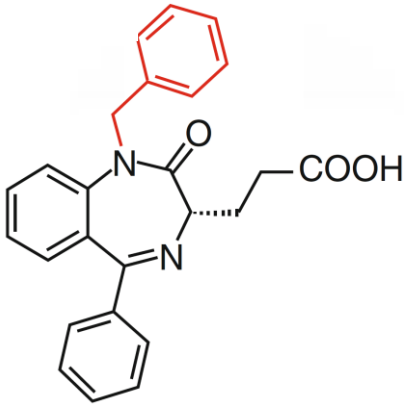


Simulating the dynamics of molecules

- Capturing the dynamic aspect of ligand and target structure is often crucial to predict binding
- Molecular dynamics (MD) simulations predict the evolution of conformation with time.

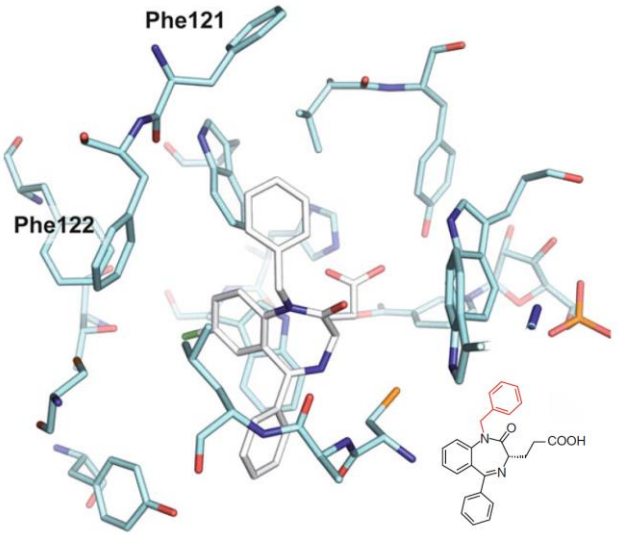
MD simulation shows wide movement of Phe121 residue, enlarging the binding pocket of the receptor





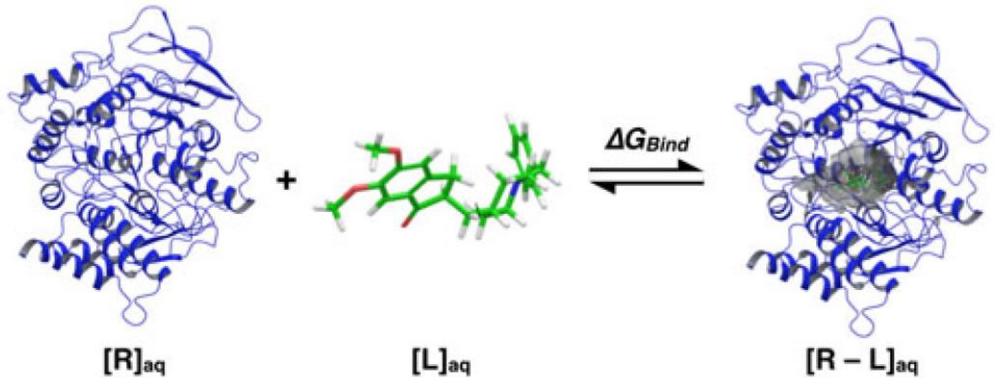
Benzodiazepine-like inhibitor

The open conformation can accommodate ligands with extended functional groups, like the red group of the benzodiazepine-like inhibitor,



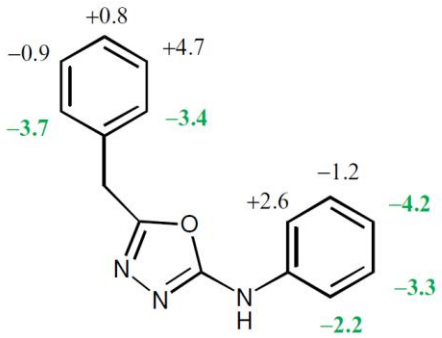
Bind free energy by MD simluations

- Estimation of the binding potency of a ligand is a most important computational task
- Molecular dynamics (MD) simulations can estimate both **relative** and **absolute** biding free energies (**ΔG** or **Δ Δ G** of binging).

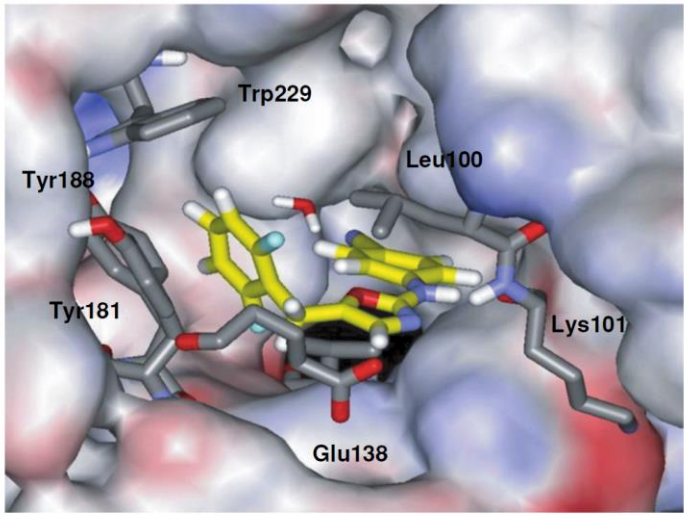


$$K_d = \frac{[R][L]}{[RL]}$$

$$\Delta G = RT \ln K_d$$



a)

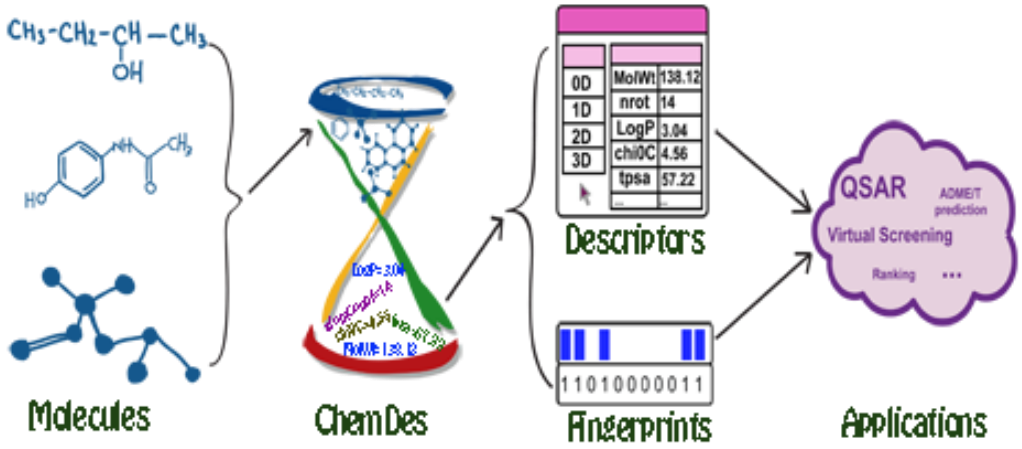


b)

Drug Descriptors and QSAR

- Ligand molecules can be represented by **molecule descriptors** and **fingerprints**
- These molecule-dependent parameters can be used to fit QSAR models or to *train* Machine Learning models

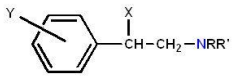
ChemDes system – generation of fingerprints and descriptors



Train a machine learning model on experimental IC50 data

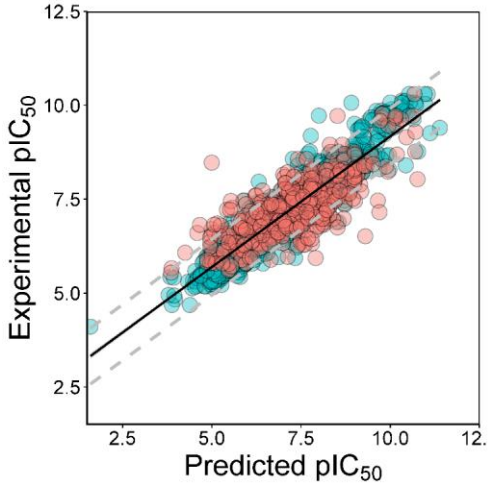
Hansch Equation

Example: Adrenergic blocking activity of β-halo-β-arylamines



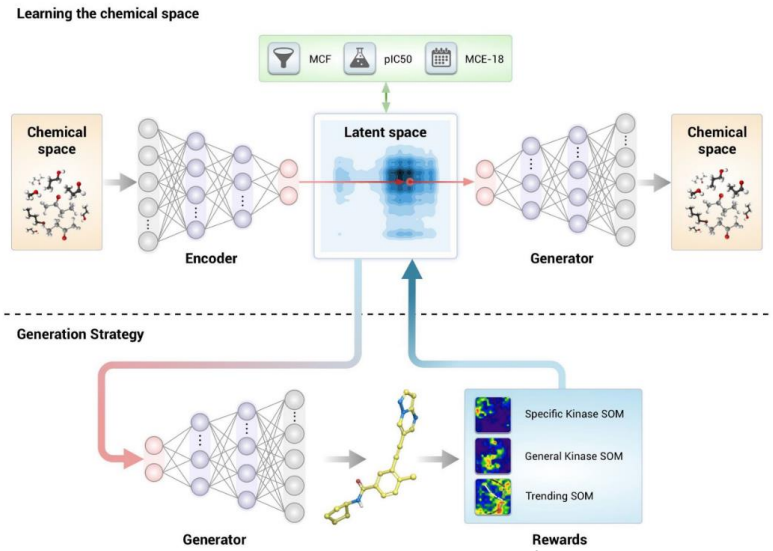
$$\text{Log}\left(1/C\right) = 1.22\,\pi - 1.59\,\sigma + 7.89$$

Conclusions:
- Activity increases if π is +ve (i.e. hydrophobic substituents)
- Activity increases if σ is negative (i.e. e-donating substituents)

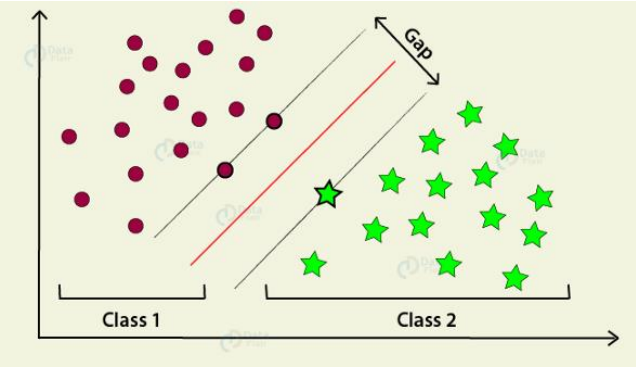


Machine learning and AI

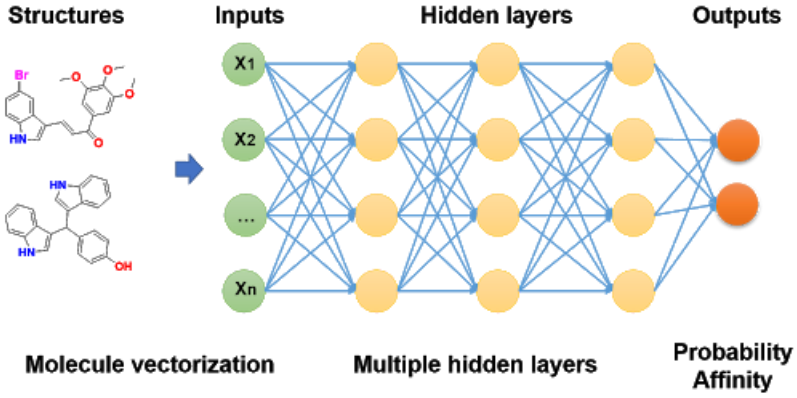
- Machine learning (ML) models are **trained** on experimental data (activity, toxicity, IC50, logP, etc)
- After learning, ML models can be used to:
  - Classification (eg:. Toxic / non-Toxic)
  - Property prediction (eg: IC50, logP)
  - Generation of new compounds (eg: new structures or scaffolds with affinity to specific targets
  - Reactivity prediction (eg: CYP metabolization)
  - Binding poses (hard)
  - Protein structure prediction(AlphaFold)
  - ...much more....



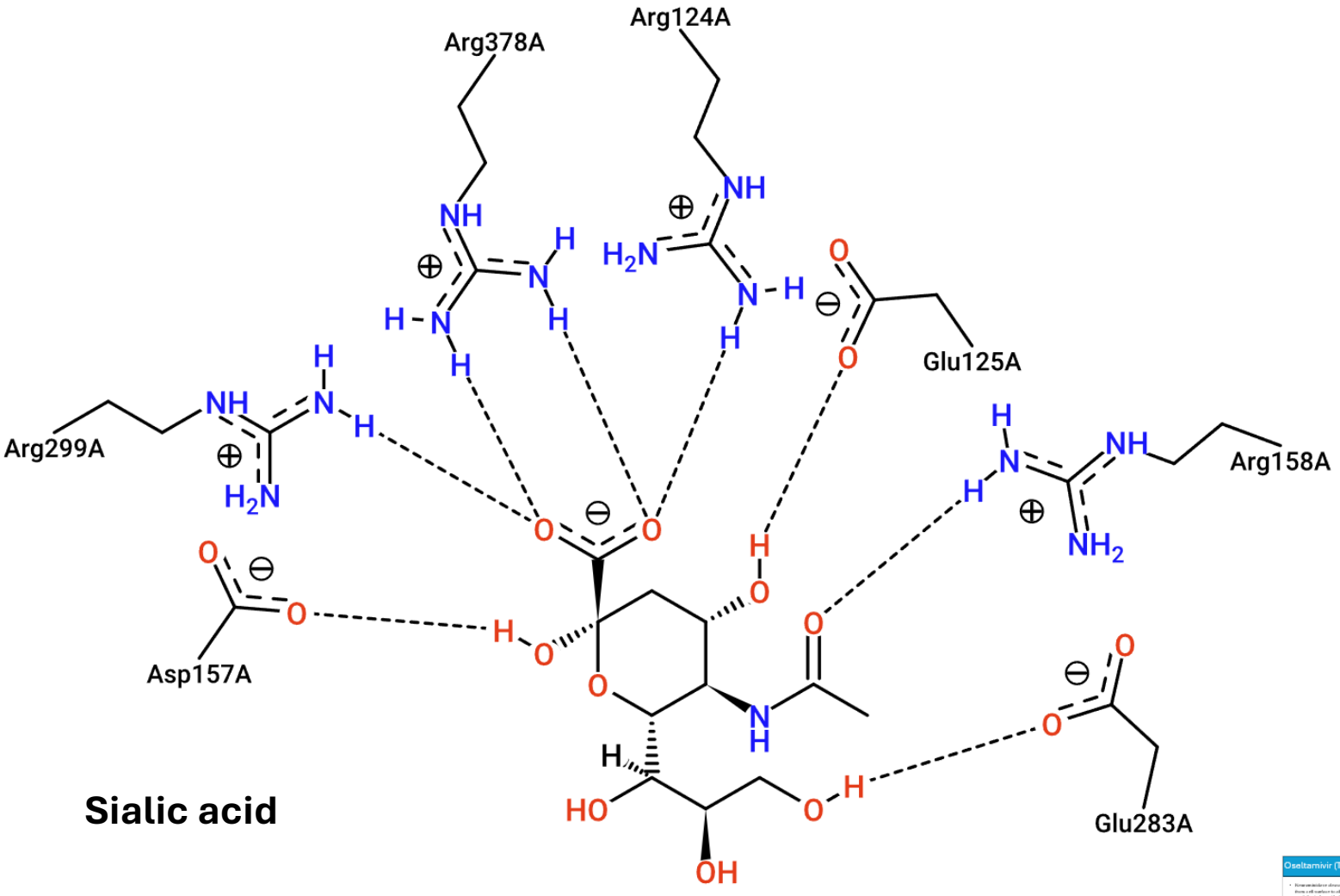
New compound generation

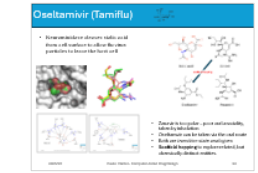


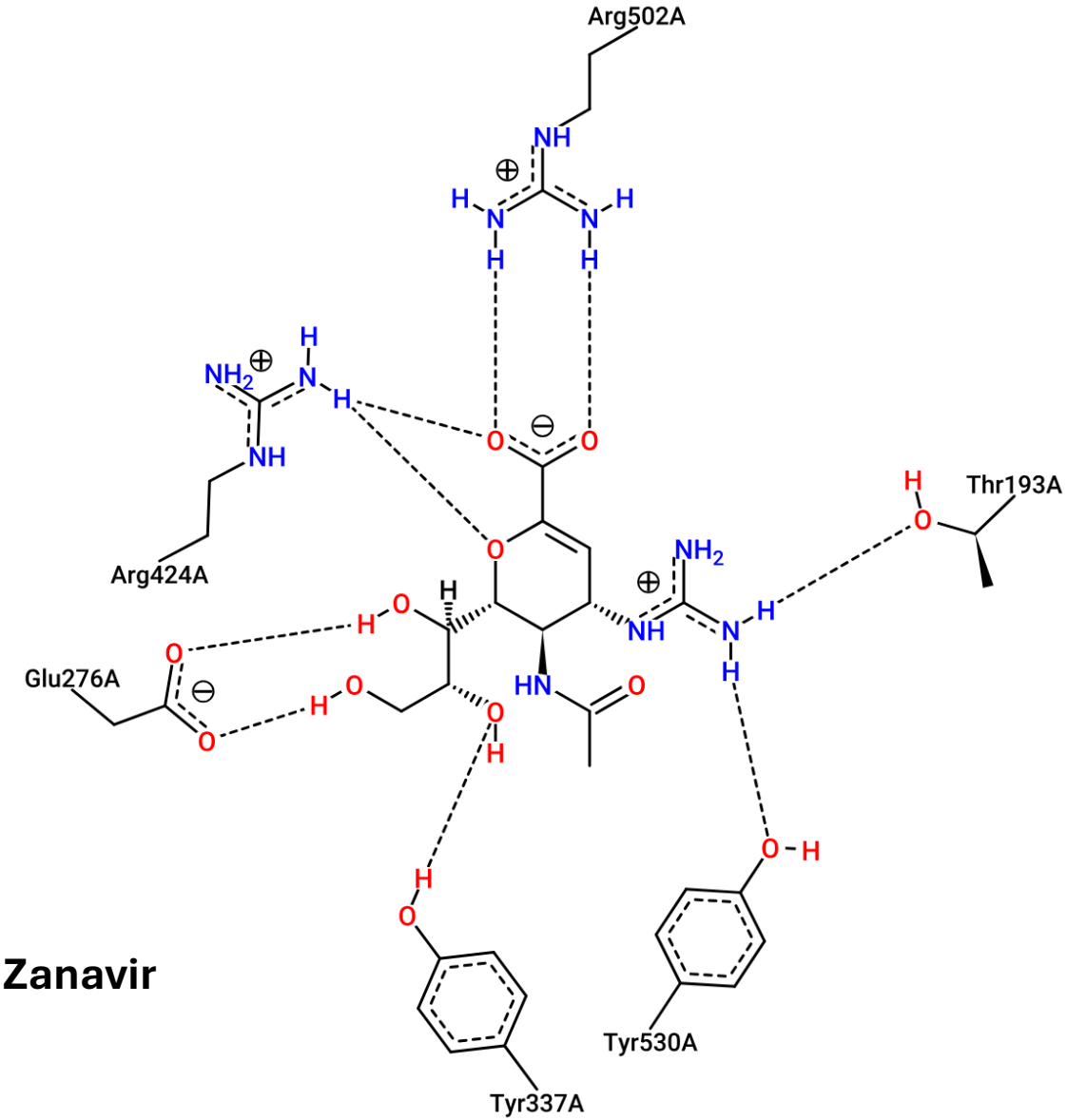
Automatic classification



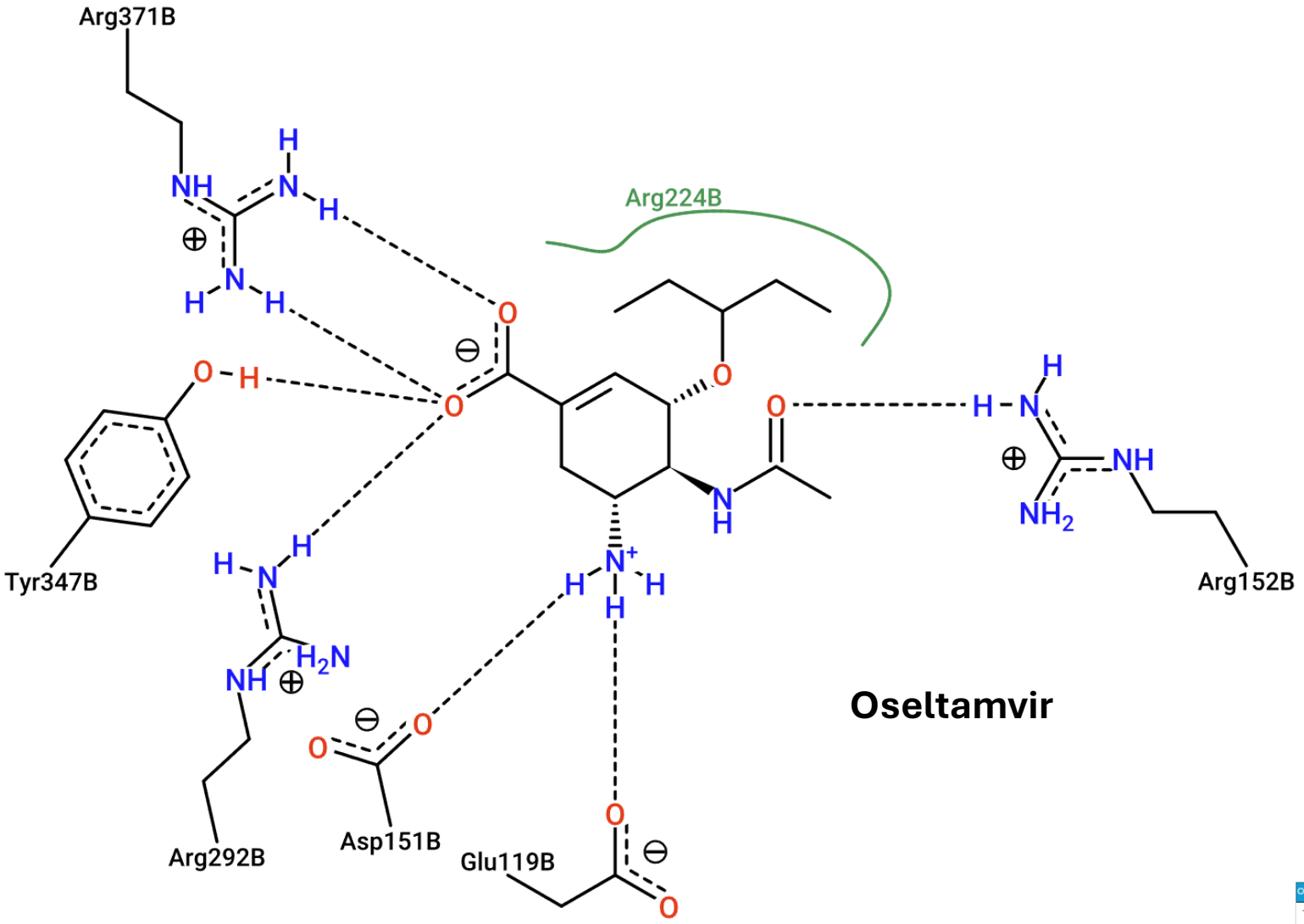
Neural Network for Property prediction

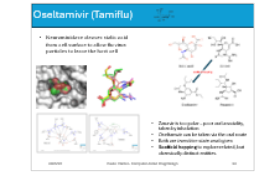

Arg378A
Arg124A
Glu125A
Arg299A
Arg158A
Asp157A
Glu283A
Sialic acid


Oseltamivir (Tamiflu)



Zanavir


Arg371B
Arg224B
Tyr347B
Arg292B
Asp151B
Glu119B
Arg152B
Oseltamvir


Oseltamivir (Tamiflu)