

# Follow-up of offline activities

## 1. Questions about Tsai et al.

- **How many** compounds were screened? ([20,000](#)) What information is available about their **properties**? ([kinase inhibition, molecular weight between 150 and 350 daltons](#))
- **How** were the compounds **screened**? ([single-dose 200 uM, crystallography with structurally divergent kinases](#))
- What was the **initial chemical structure** that was found to bind to the ATP-binding site? ([7-azaindole](#))
- By overlapping structures, the team aimed to optimizing what **two properties of the compounds**? ([potency and selectivity](#))
- What types of compounds were tested in the **subsequent screening**? ([mono- and di-substituted analogs built around the 7-azaindole core](#))
- What properties does the PLX4720 compound have that make it **particularly attractive** as a drug? ([affinity, selectivity, and a good safety profile](#))

## 2. Questions from the anonymous survey:

- **Difference between divide-and-conquer and dynamic programming:** they are indeed different strategies (thanks David Sommer!). [These discussions](#) on StackOverflow may help you recognize the commonalities and differences
- **How were papers selected?** Based on four considerations: (1) topic relevant for drug discovery, (2) reasonably well written, (3) balancing recent and classic literature, and (4) widely-used information resources. They remain however a limited and biased selection.

# Exercises of lecture 3 and 4

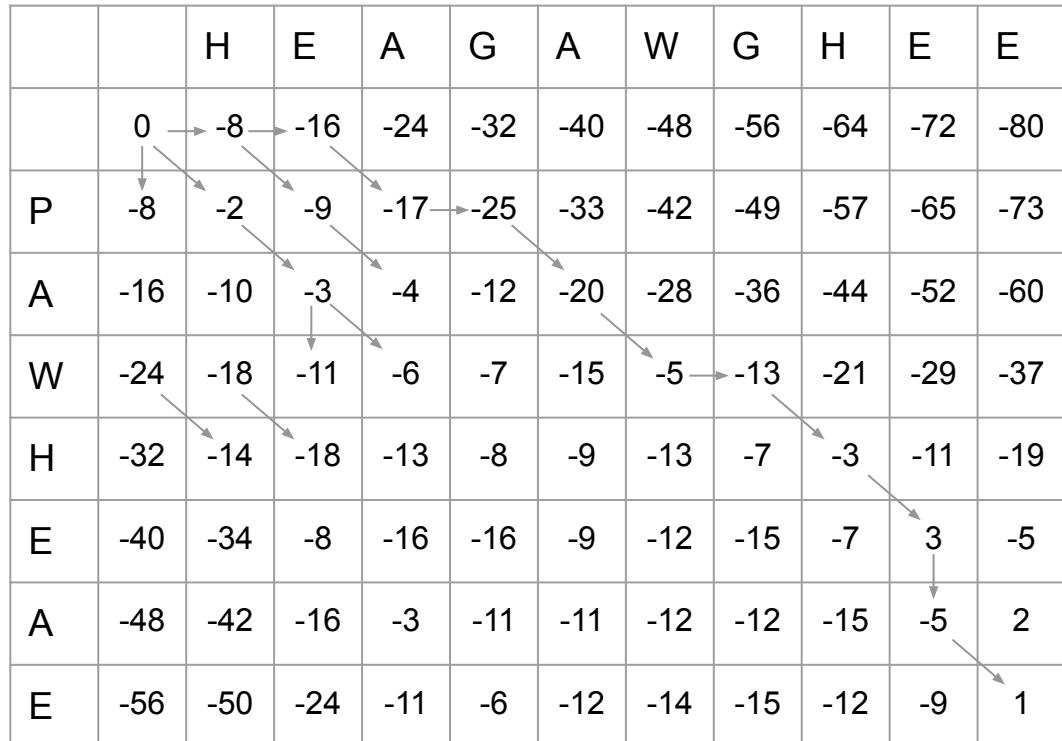
	H	E	A	G	A	W	G	H	E	E
P	-2	-1	-1	-2	-1	-4	-2	-2	-1	-1
A	-2	-1	5	0	5	-3	0	-2	-1	-1
W	-3	-3	-3	-3	-3	15	-3	-3	-3	-3
H	10	0	-2	-2	-2	-3	-3	10	0	0
E	0	6	-1	-3	-1	-3	0	0	6	6
A	2	-1	5	0	5	-3	0	2	-1	-1
E	0	6	1	3	1	-3	-3	0	6	6

Adapted from *Biological Sequence Analysis* (R. Durbin, S. Eddy, A. Krogh, G. Mitchison), Figure 2.3. We assume that a gap cost per unaligned residue of  $d=-8$ . Try to use the information to perform global alignment between the two amino-acid sequences:

1. HEAGAWGHEE
2. PAWHEAE

## What does Fomivirsen target?

It is possible to search for local sequence matches in large databases of nucleotides, for instance using the BLAST algorithm. An implementation is freely available at National Institute of Health (NIH, US): <https://blast.ncbi.nlm.nih.gov/Blast.cgi>. Try to search for the RNA/protein targeted by fomivirsen, given its sequence 5'-GCG TTT GCT CTT CTT CTT GCG-3'.



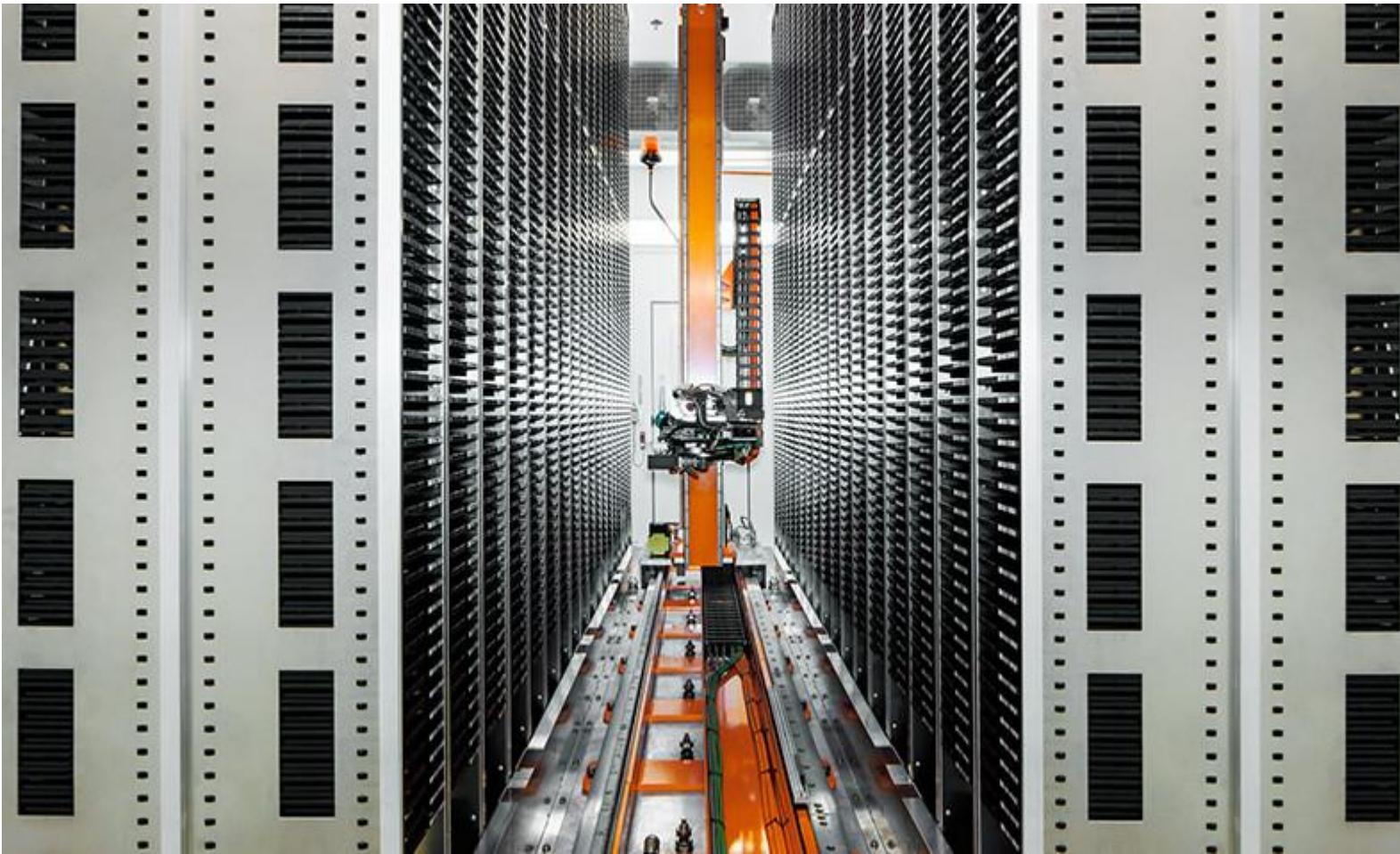
<input checked="" type="checkbox"/> Human betaherpesvirus 5 strain SYD-SCT1, complete genome	42.1	42.1	100%	0.14	100.00%	<a href="#">MT044485.1</a>
<input checked="" type="checkbox"/> Human betaherpesvirus 5 strain HAN-SOT4, complete genome	42.1	42.1	100%	0.14	100.00%	<a href="#">MT044484.1</a>
<input checked="" type="checkbox"/> Human betaherpesvirus 5 strain HAN-SOT3, partial genome	42.1	42.1	100%	0.14	100.00%	<a href="#">MT044483.1</a>
<input checked="" type="checkbox"/> Human betaherpesvirus 5 strain GLA-SOT3, complete genome	42.1	42.1	100%	0.14	100.00%	<a href="#">MT044482.1</a>
<input checked="" type="checkbox"/> Human betaherpesvirus 5 strain GLA-SOT2, complete genome	42.1	42.1	100%	0.14	100.00%	<a href="#">MT044481.1</a>
<input checked="" type="checkbox"/> Human betaherpesvirus 5 strain SYD-SCT2, complete genome	42.1	42.1	100%	0.14	100.00%	<a href="#">MT044480.1</a>
<input checked="" type="checkbox"/> Human betaherpesvirus 5 strain HAN-SOT5, complete genome	42.1	42.1	100%	0.14	100.00%	<a href="#">MT044479.1</a>
<input checked="" type="checkbox"/> Human betaherpesvirus 5 strain HAN-SOT1, complete genome	42.1	42.1	100%	0.14	100.00%	<a href="#">MT044478.1</a>
<input checked="" type="checkbox"/> Human betaherpesvirus 5 strain GLA-SOT4, complete genome	42.1	42.1	100%	0.14	100.00%	<a href="#">MT044477.1</a>

# An example of HMM

```
library(HMM)
hmmModel <- initHMM(States=c("A", "D"), # A=Angel, D=Devil
                      Symbols=c("B", "R"), # B=Blessing, R=Rant
                      startProbs=c(0.5, 0.5),
                      transProbs = matrix(c(0.8, 0.2, 0.2, 0.8), nrow=2),
                      emissionProbs = matrix(c(0.9, 0.1, 0.1, 0.9), nrow=2))
simHmm <- simHMM(hmmModel, 100)
simStates <- paste(simHmm$states, collapse="")
simSymbols <- paste(simHmm$observation, collapse="")
```

ADDADDADDADDADDADDAAAAA  
BBRRBRRBRRBRRRRRB BBBB  
AAAAADAADDADDAAADAAA  
BBBBBRBRBRRRBRRBB BBBB  
AAADDADDADDADDADDAAA  
BBBRRRRRBRRRBRRRB  
AAAADDAAAAAAADDADD  
BBBRRRBBBBBBRRRRRRR  
DDDDDDDDDDDDAADAADD  
RRRRRBRRRBRRBBRRBRR

# AMIDD Lecture 5: Molecular Screening and Modelling



*The chemical library at Novartis headquarters in Basel currently contains roughly 3 million molecules. We aim to expand that number radically within the next few years.*

Jay Bradner, President of NIBR, in [an interview](#) in 2017

**Dr. Jitao David Zhang, Computational Biologist**

**<sup>1</sup> Pharmaceutical Sciences, Pharma Research and Early Development, Roche Innovation Center Basel, F. Hoffmann-La Roche**

**<sup>2</sup> Department of Mathematics and Informatics, University of Basel**

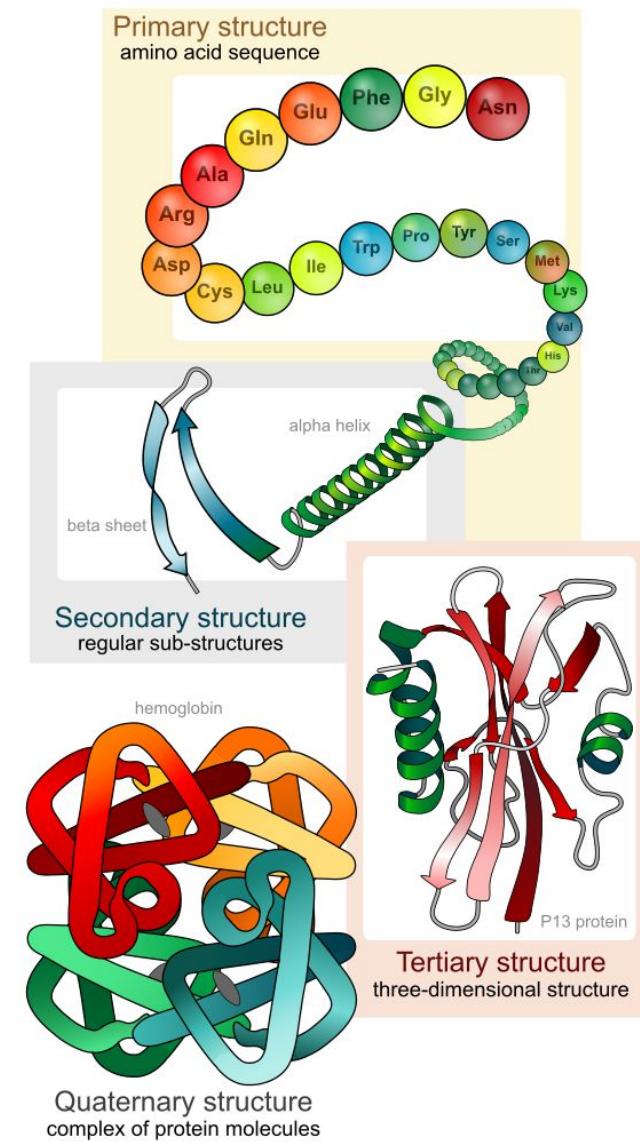
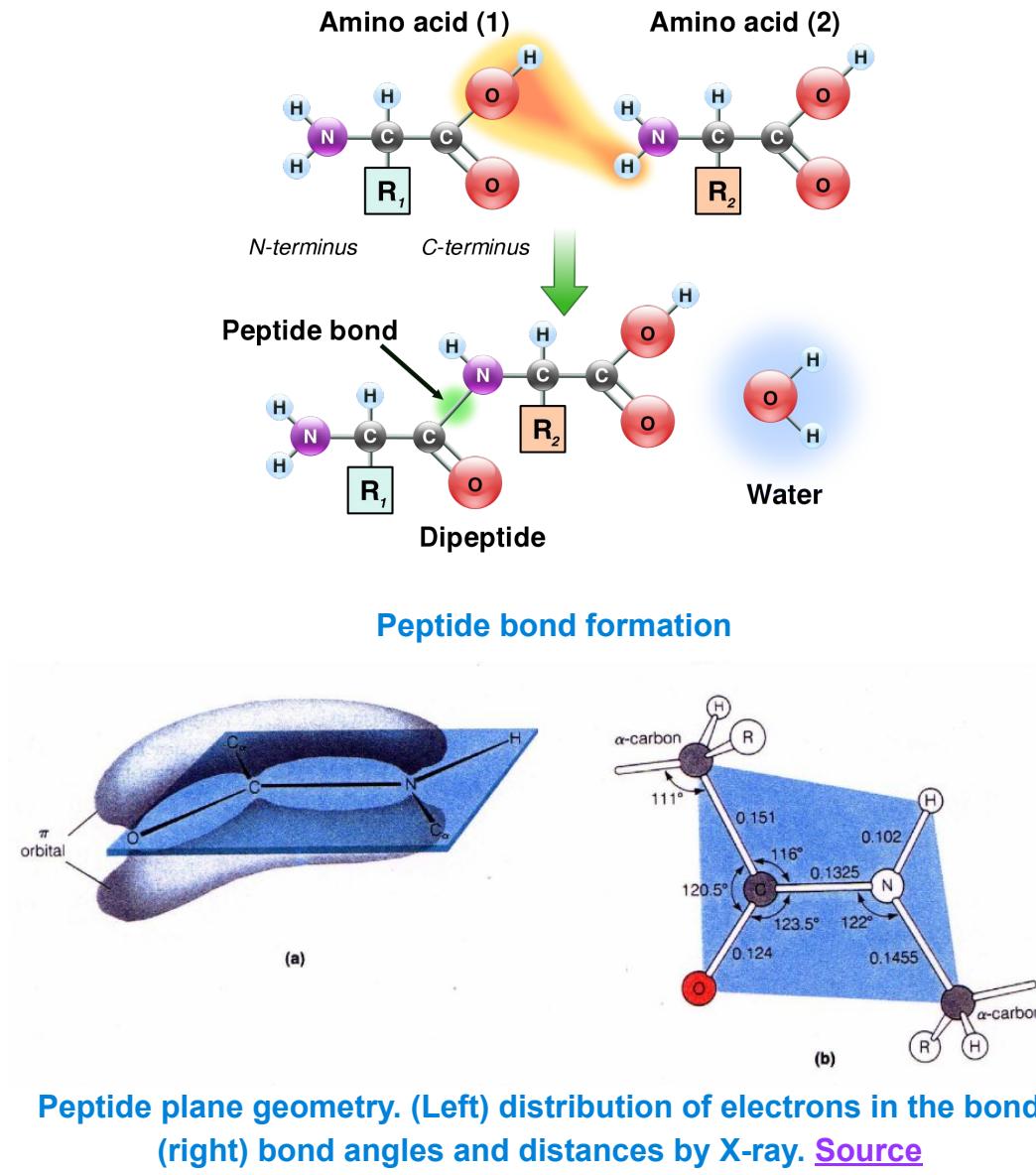
# Today's goals

- Protein biology and structure determination
- Representation and molecular descriptors of small molecules
- Two views of ligand-target binding

# What properties must a drug satisfy?

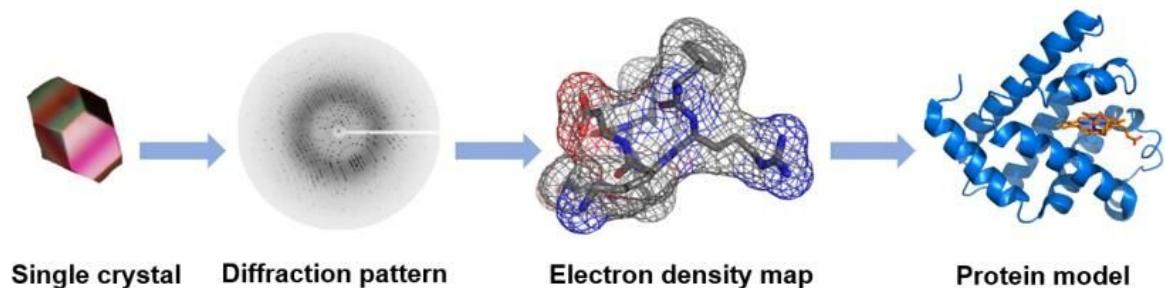
# From amino acids to proteins

- Translation of mRNA means that two consecutive amino acids specified by 3-nucleotide codons form **peptide bonds** (top left panel). The peptide bonds concatenate amino acids together into *peptides* or *proteins*.
- The peptide plane geometry, determined by X-ray crystallography, is used to model structures and proteins. (bottom left panel).
- Protein structures can be thought of as hierarchical: primary amino-acid sequences form secondary structures (alpha helices and beta sheets), which form 3D structures of proteins, which can further form complexes (right panel).



Four levels of protein structures

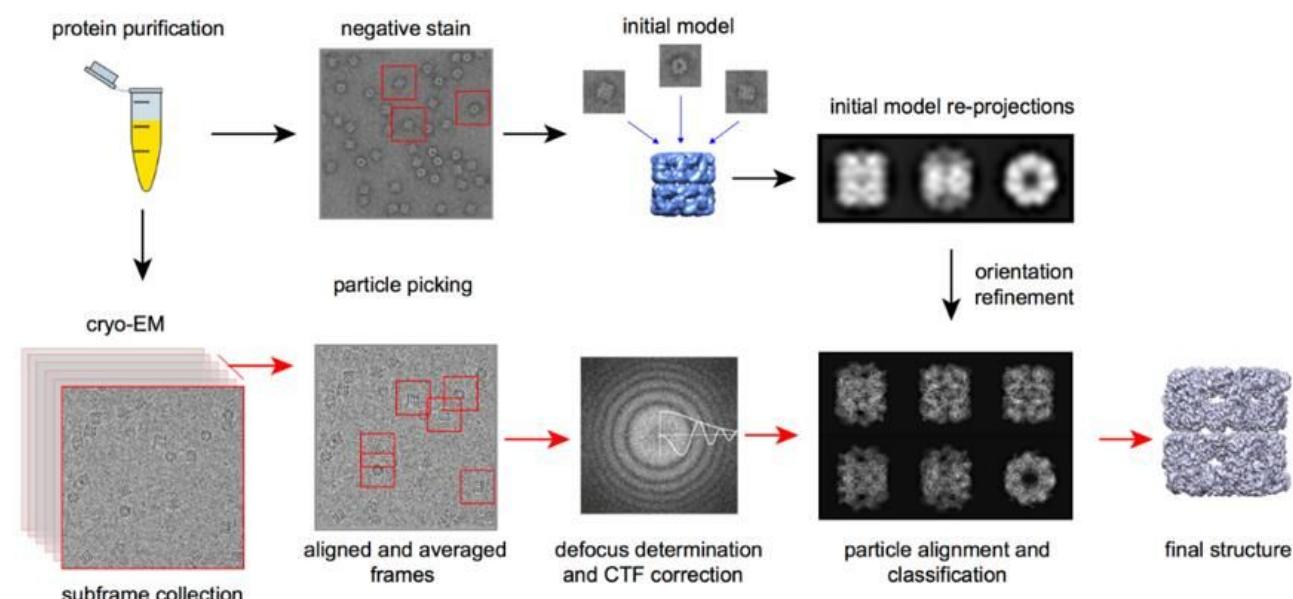
# Three major experimental approaches to determining protein structures



**X-ray crystallography**



**Nuclear Magnetic Resonance (NMR)**



**Cryo-electron microscopy (CryoEM)**

Figure sources:

[https://www.creative-biostructure.com/comparison-of-crystallography-nmr-and-em\\_6.htm](https://www.creative-biostructure.com/comparison-of-crystallography-nmr-and-em_6.htm)

# Three major experimental approaches to determining protein structures

Method	Underlying physical properties	Main mathematical technique used	Advantages	Limitations
X-ray crystallography	The crystalline structure of a molecule causes a beam of incident X-rays to diffract into many specific directions.	Fourier series and Fourier transform	<ul style="list-style-type: none"> <li>Established</li> <li>Broad molecular weight range</li> <li>High resolution</li> </ul>	<ul style="list-style-type: none"> <li>Crystallization</li> <li>Static model</li> </ul>
Nuclear Magnetic Resonance (NMR)	Nuclei with odd number of protons and/or neutrons in a strong constant magnetic field, when perturbed by a weak oscillating magnetic field, produce an electromagnetic signal with a frequency characteristic of the magnetic field at the nucleus.	Distance geometry (the study of matrices of distances between pairs of atoms) or discrete differential geometry of curves	<ul style="list-style-type: none"> <li>3D structure in solution</li> <li>Dynamic study possible</li> </ul>	<ul style="list-style-type: none"> <li>High sample purity needed</li> <li>Molecular weight limit (~&lt;40-50 kDa)</li> <li>Sample preparation and computational simulation</li> </ul>
Cryo-electron microscopy	An electron microscope using a beam of accelerated electrons (instead of protons) as a source of illumination. Samples are cooled to cryogenic temperatures and embedded in an environment of vitreous water (amorphous ice).	An inverse problem of reconstruction - the estimation of randomly rotated molecule structure from a projection with noise; Fourier transform; iterative refinement	<ul style="list-style-type: none"> <li>Easy sample preparation</li> <li>Native-state structure</li> <li>Small sample size</li> </ul>	<ul style="list-style-type: none"> <li>Costly EM equipment</li> <li>Challenging for small proteins</li> </ul>

# *In silico* presentation of protein structures: PDB

**RCsb** **PDB**  
PROTEIN DATA BANK

157145 Biological  
Macromolecular Structures  
Enabling Breakthroughs in  
Research and Education

Sea

Adva

PDB-101

WORLDWIDE  
**PDB**  
PROTEIN DATA BANK

EMDataResource  
Unified Data Resource for IEDB

ndb  
NUCLEIC ACID  
DATABASE

Worldwide  
Protein Data Bank  
Foundation

Structure Summary

3D View

Annotations

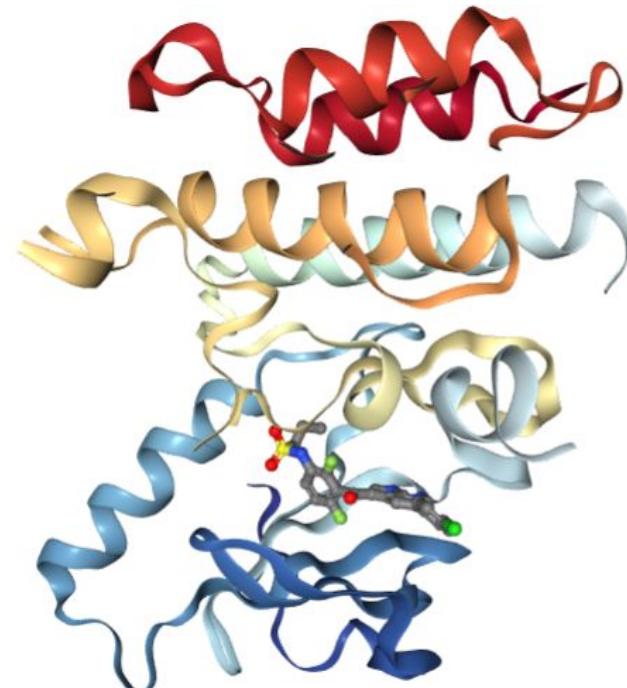
Sequence

Sequence

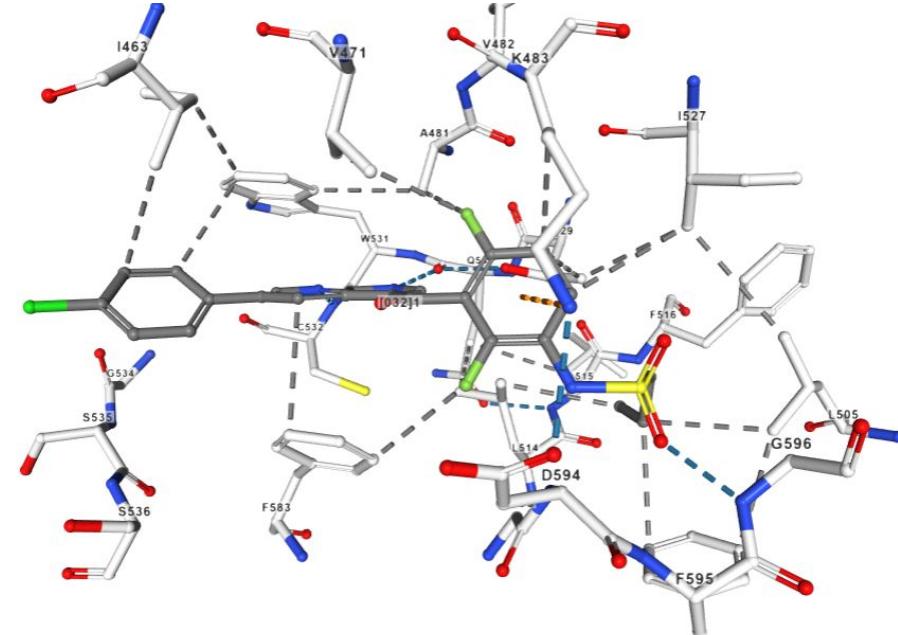
# 3OG7

B-Raf Kinase V600E oncogenic mutant in complex with PLX4032

<http://www.rcsb.org/3d-view/3OG7>



## Structural view



## Ligand view

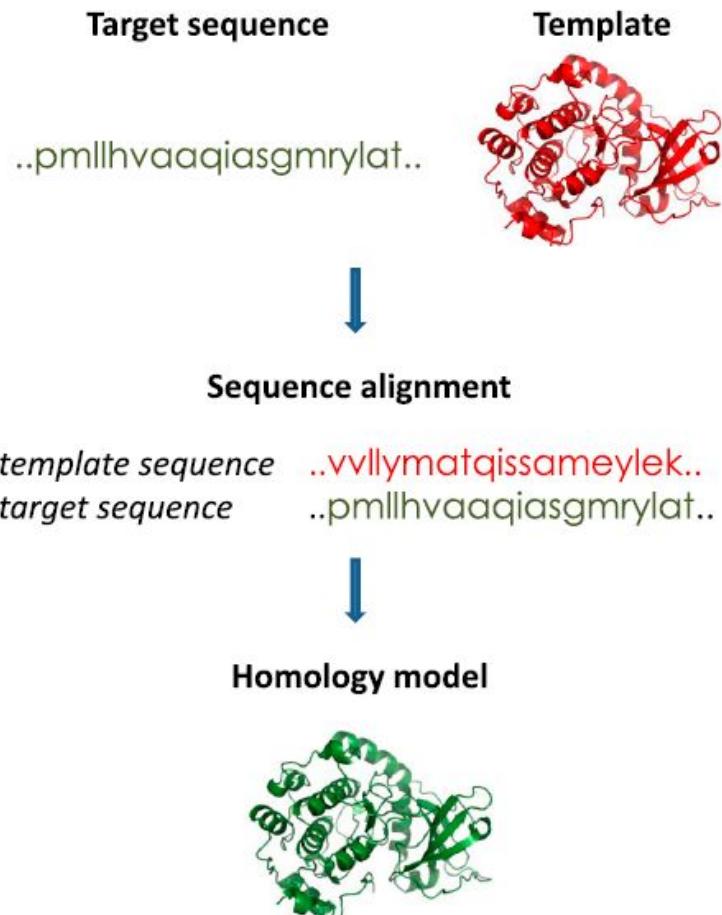
Balls and sticks: protein V600E and ligand (PLX4032)

**Blue dashes**: hydrogen bonds (<3.5 Angstrom)

**Gray dashes:** hydrophobic interactions (<4 Angstrom)

## Working with PDB files with **PyMol** from the command-line

# If no structure is available, homology model building and *in silico* prediction may help



Sliwski, Gregory, Sandeepkumar Kothiwale, Jens Meiler, und Edward W. Lowe. „Computational Methods in Drug Discovery“. *Pharmacological Reviews* 66, Nr. 1 (1. Januar 2014): 334–95. <https://doi.org/10.1124/pr.112.007336>.

W296–W303 *Nucleic Acids Research*, 2018, Vol. 46, Web Server issue  
 doi: 10.1093/nar/gky427

Published online 21 May 2018

## SWISS-MODEL: homology modelling of protein structures and complexes

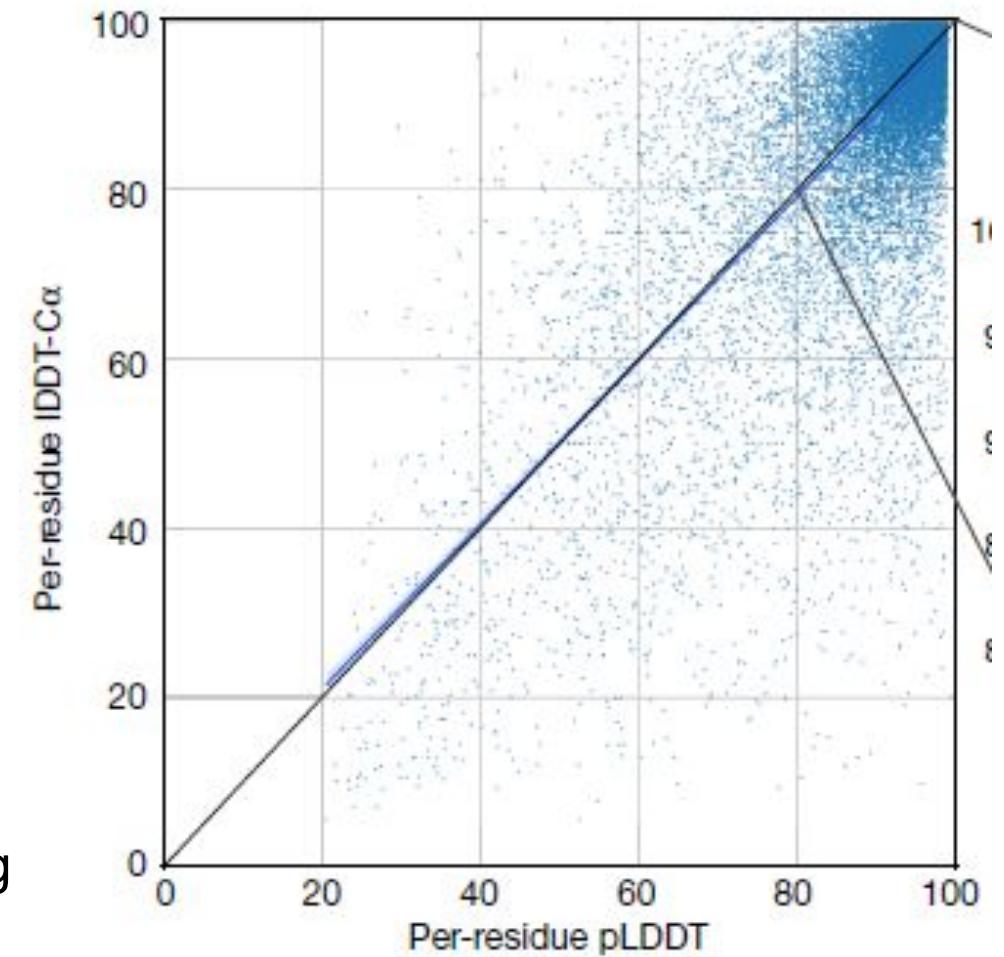
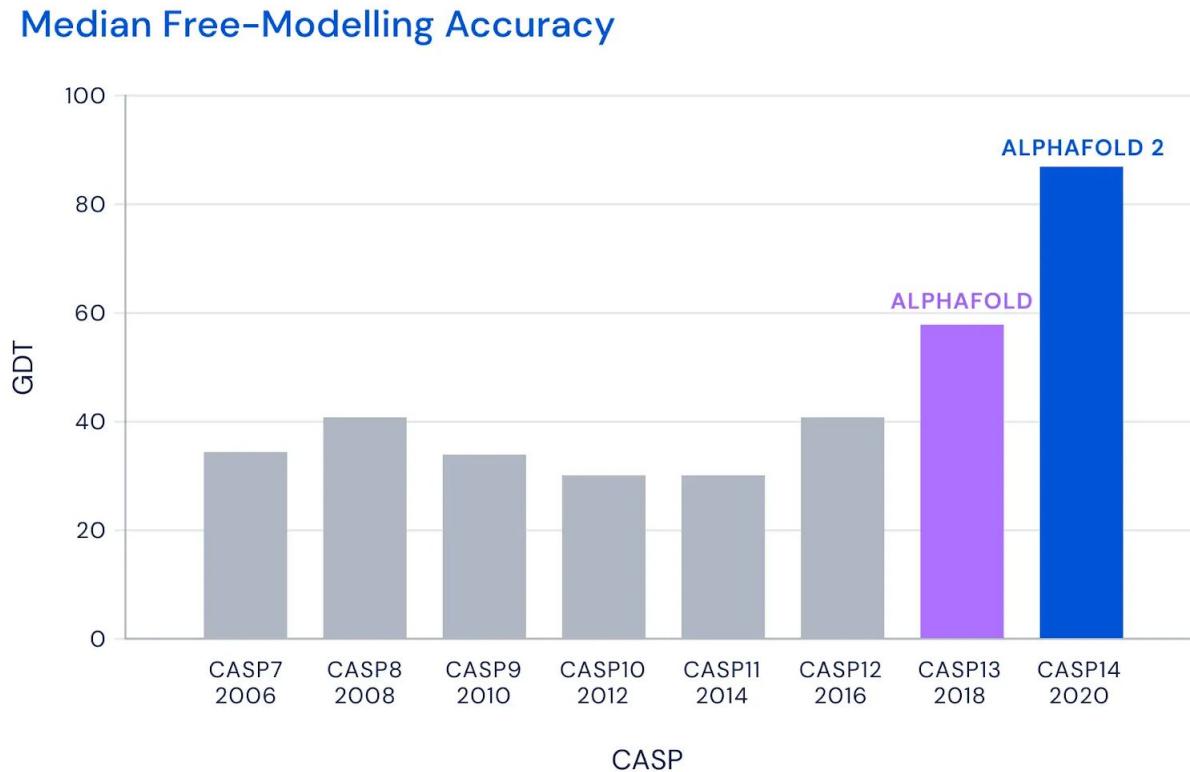
Andrew Waterhouse<sup>1,2,†</sup>, Martino Bertoni<sup>1,2,†</sup>, Stefan Bienert<sup>1,2,†</sup>, Gabriel Studer<sup>1,2,†</sup>, Gerardo Tauriello<sup>1,2,†</sup>, Rafal Gumienny<sup>1,2</sup>, Florian T. Heer<sup>1,2</sup>, Tjaart A. P. de Beer<sup>1,2</sup>, Christine Rempfer<sup>1,2</sup>, Lorenza Bordoli<sup>1,2</sup>, Rosalba Lepore<sup>1,2</sup> and Torsten Schwede<sup>1,2,\*</sup>

<sup>1</sup>Biozentrum, University of Basel, Klingelbergstrasse 50–70, CH-4056 Basel, Switzerland and <sup>2</sup>SIB Swiss Institute of Bioinformatics, Biozentrum, University of Basel, Klingelbergstrasse 50–70, CH-4056 Basel, Switzerland

Received February 09, 2018; Revised May 01, 2018; Editorial Decision May 02, 2018; Accepted May 07, 2018

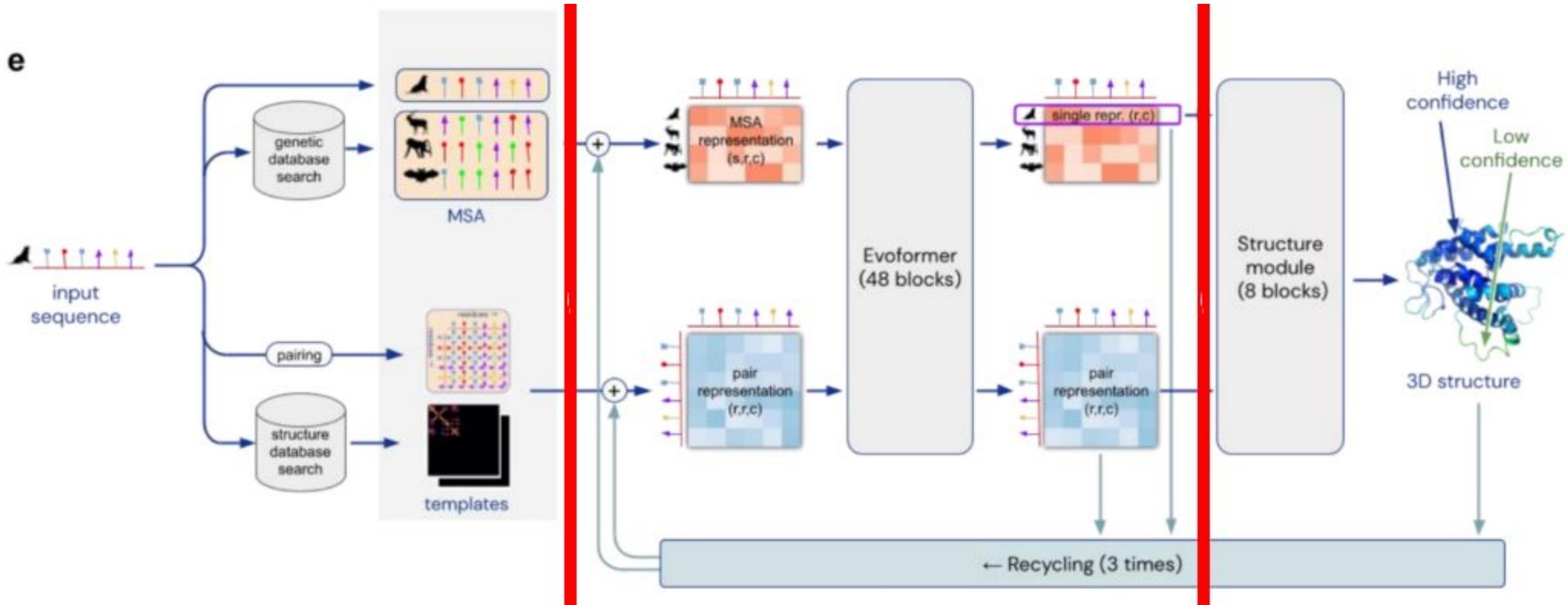
- **Levinthal's paradox:** *It would take a protein the present age of the universe to explore all possible configurations and find the minimum energy configuration. Yet proteins fold in microseconds.*
- **CASP: Critical Assessment of Techniques for Protein Structure Prediction**
- A thought-provoking blog from Mohammed AlQuraishi: [AlphaFold @ CASP13: “What just happened?”](#), with an informal but good overview of history of protein structure prediction, and his indictment (criminal accusations) of both academia and pharma.
- By 2021 AlphaFold2 and RoseTTAFold reach experiment-level accuracy in some predictions of protein static structure

# AlphaFold2 reaches prediction accuracy comparable to experimental approaches



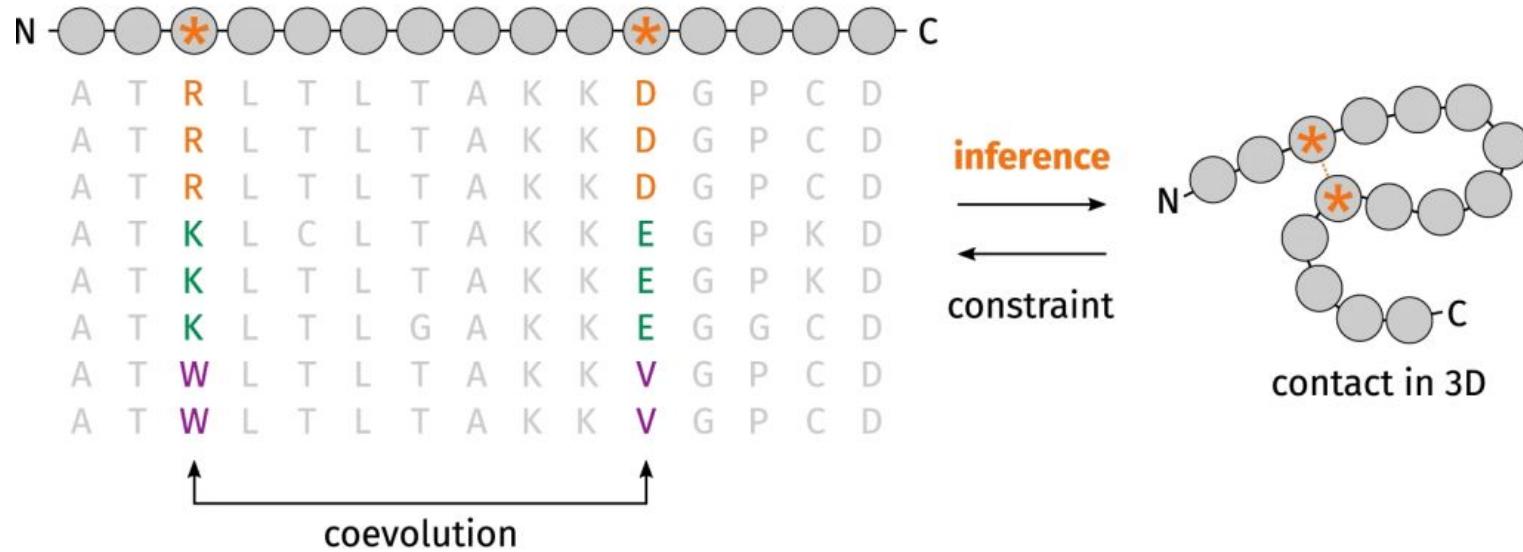
pLDDT: Predicted local distance difference test, estimating how the prediction differs from the experimental structure based on the local distance difference test (C-alpha, IDDT)

# AlphaFold2 uses co-evolution of residues, determined structures, and neural networks to achieve the high performance



- Jumpe et al. “Highly Accurate Protein Structure Prediction with AlphaFold.” *Nature* 596, no. 7873 (August 2021): 583–89. <https://doi.org/10.1038/s41586-021-03819-2>.
- A blog post that explains how AlphaFold2 works: [blogpig.com](http://blogpig.com)

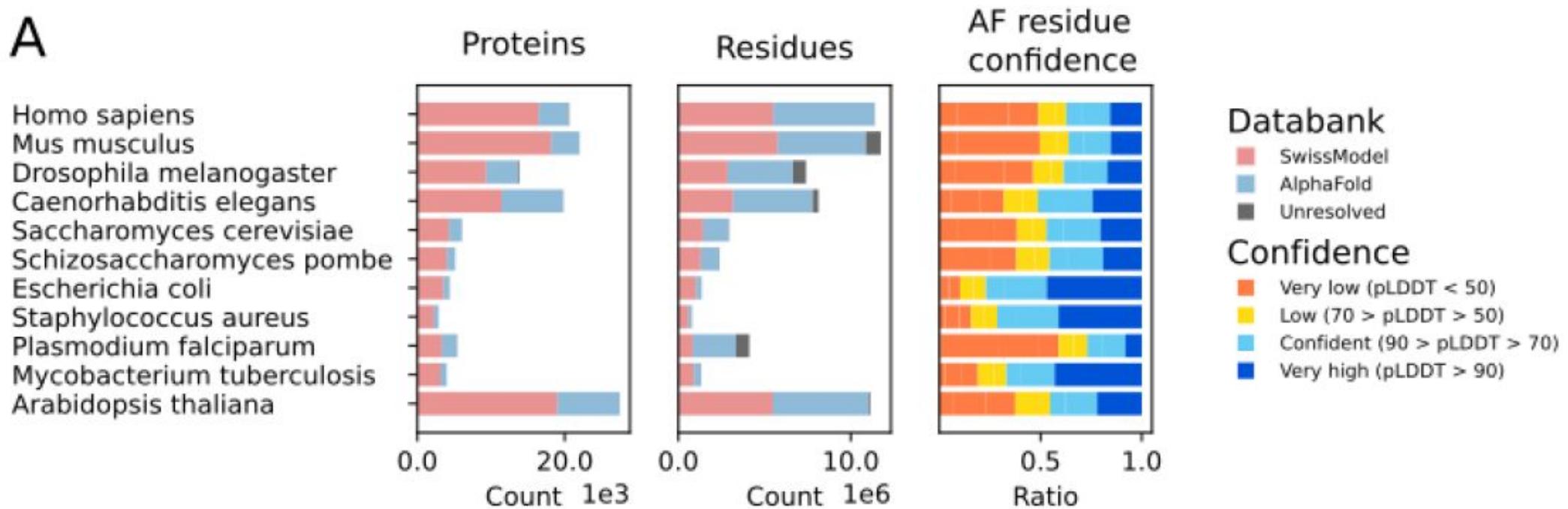
**The key idea (beyond using 2D and 3D structure mapping):  
learning from evolutionary constraints**



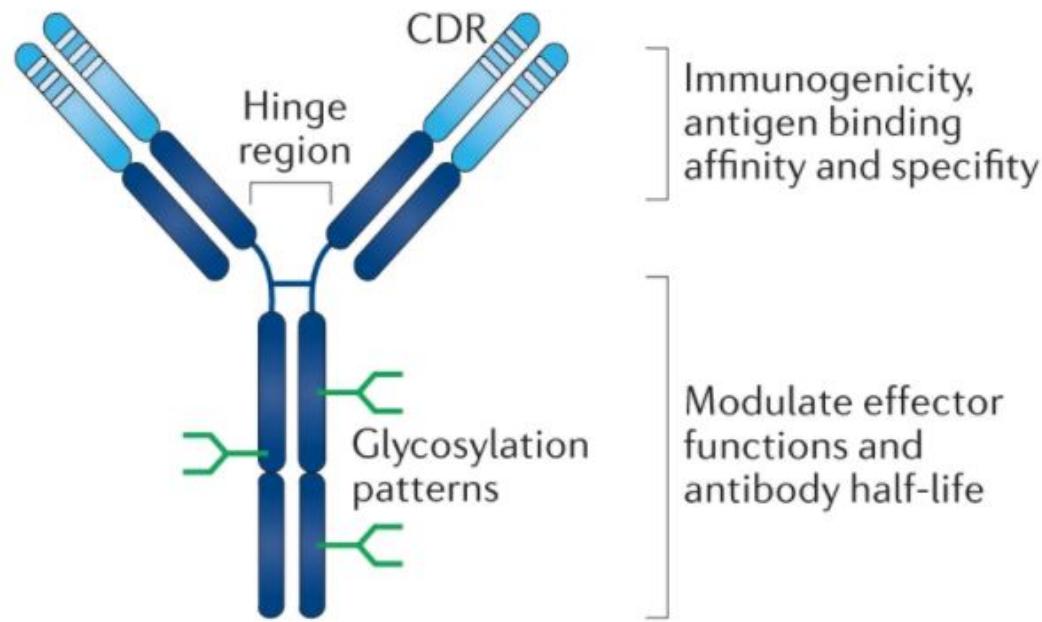
Marks, Debora S., Lucy J. Colwell, Robert Sheridan, Thomas A. Hopf, Andrea Pagnani, Riccardo Zecchina, and Chris Sander. "Protein 3D Structure Computed from Evolutionary Sequence Variation." *PLOS ONE* 6, no. 12 (December 7, 2011): e28766.  
<https://doi.org/10.1371/journal.pone.0028766>.

# AlphaFold2 & RoseTTAfold extend our understanding of protein biology, while their impact on drug discovery remains to be seen

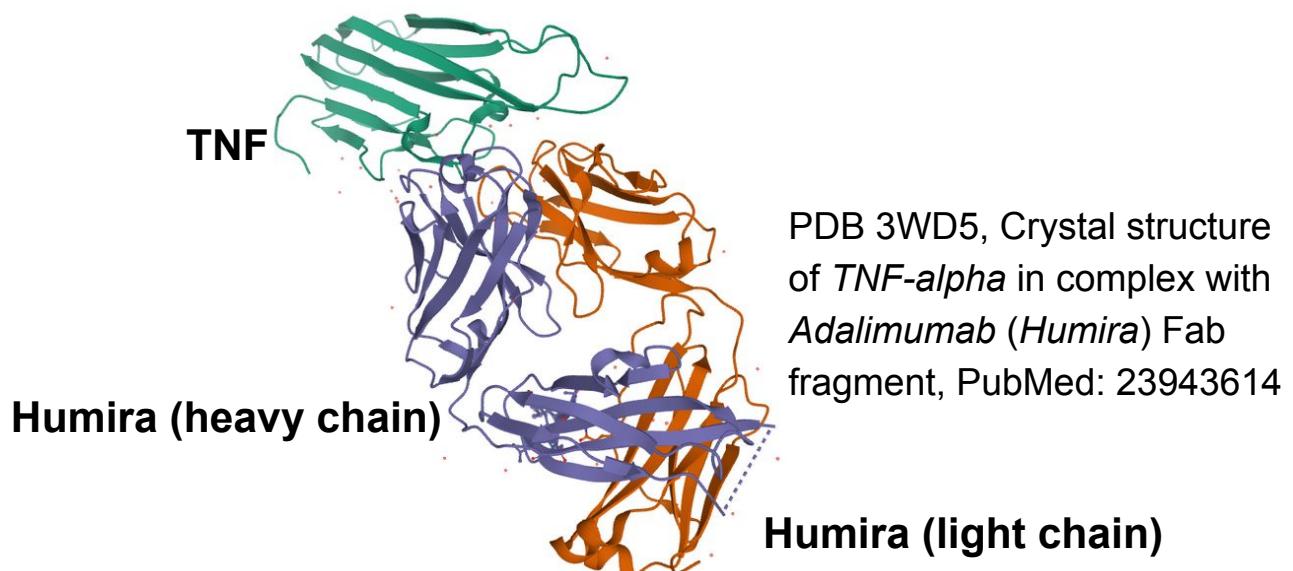
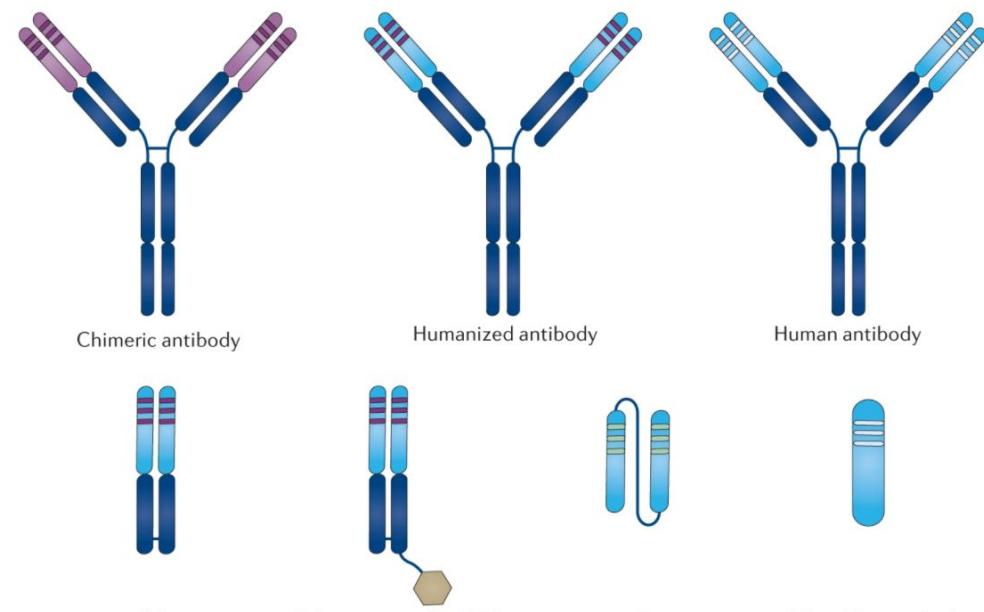
**A**



# Antibodies are also proteins



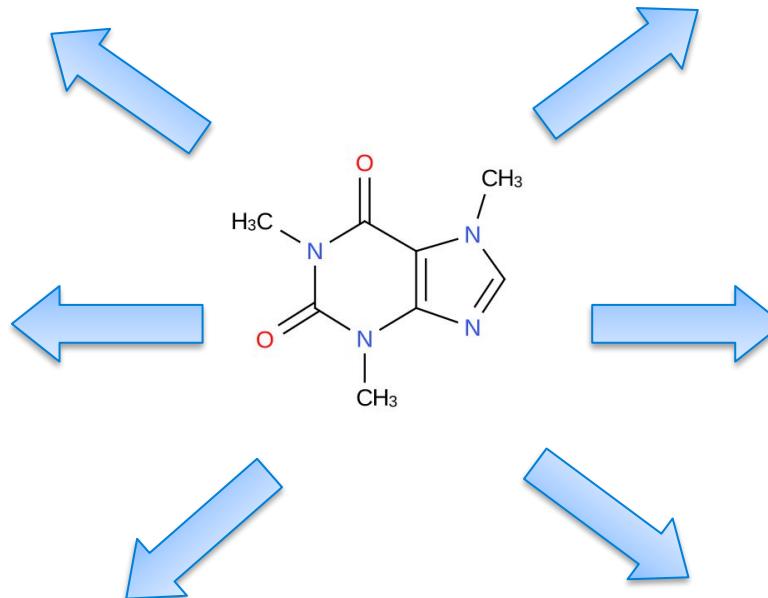
Attwood, Misty M., Jörgen Jonsson, Mathias Rask-Andersen, and Helgi B. Schiöth. 2020. "Soluble Ligands as Drug Targets." *Nature Reviews Drug Discovery* 19 (10): 695–710.  
<https://doi.org/10.1038/s41573-020-0078-4>.



# ChEMBL as information source of small molecules

## Nomenclature

*caffeine*  
*1,3,7-trimethylxanthine*  
*methyltheobromine*



## Bioactivity

*Affinity to human proteins and drug targets*

## Chemical data

*Formula: C<sub>8</sub>H<sub>10</sub>N<sub>4</sub>O<sub>2</sub>*  
*Charge: 0*  
*Mass: 194.19*

## Database Xrefs

*PubChem: CID2519*  
*BindingDB: 1849*

## Chemical Informatics

*InChI=1/C8H10N4O2/c1-10-4-9-6-5(10)7(13)  
 12(3)8(14)11(6)2/h4H,1-3H3*

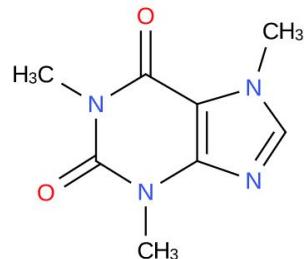
*SMILES: CN1C(=O)N(C)c2ncn(C)c2C1=O*

## Visualisation



A subset of available information from EBI ChEBI/ChEMBL,  
 inspired by EBI's roadshow *Small Molecules in Bioinformatics*

# Representation of small molecules



<b>Molfile:</b>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>
<b>Canonical SMILES:</b>	<chem>CN1C(=O)N(C)c2ncn(C)c2C1=O</chem>			
<b>Standard InChI:</b>	<chem>InChI=1S/C8H10N4O2/c1-10-4-9-6-5(10)7(13)12(3)8(14)11(6)2/h4H,1-3H3</chem>			
<b>Standard InChI Key:</b>	RYYVLZUVIJVGH-UHFFFAOYSA-N			

- Simplified Molecular-Input Line-Entry System (SMILES)
- IUPAC International Chemical Identifier (InChI)
- InChiKey: a 27-character, hash version of InChI
- Molfile: a type of [chemical table files](#)

CHEMBL113

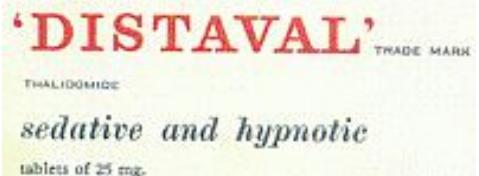
SciTegic12231509382D

```
14 15 0 0 0 0 999 V2000
-1.1875 -9.6542 0.0000 C 0 0
-1.1875 -8.9625 0.0000 C 0 0
-1.8125 -10.0292 0.0000 N 0 0
-2.4167 -8.9625 0.0000 N 0 0
-2.4167 -9.6542 0.0000 C 0 0
-1.8125 -8.6000 0.0000 C 0 0
-0.5000 -9.8917 0.0000 N 0 0
-0.5000 -8.7625 0.0000 N 0 0
-0.1125 -9.3042 0.0000 C 0 0
-3.0250 -10.0375 0.0000 O 0 0
-1.8125 -7.8917 0.0000 O 0 0
-1.8125 -10.7417 0.0000 C 0 0
-3.0250 -8.6000 0.0000 C 0 0
-0.2917 -8.0750 0.0000 C 0 0
2 1 2 0
3 1 1 0
4 5 1 0
5 3 1 0
6 2 1 0
7 1 1 0
8 2 1 0
9 7 2 0
10 5 2 0
11 6 2 0
12 3 1 0
13 4 1 0
```

# The tragedy of thalidomide and the importance of representation



A complete sedative and hypnotic range – in a single preparation. That is 'Distaval' . . . the safe day-time sedative which is equally safe in hypnotic doses by night. 'Distaval' is especially suitable for infants, the aged, and patients under severe emotional stress.



(1957)

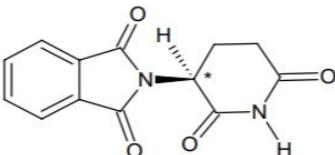
I thank Manuela Jacklin for her help preparing this slide.



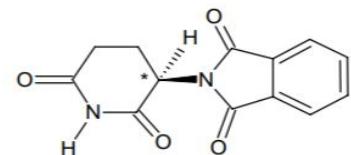
Frances Oldham Kelsey received the President's Award for Distinguished Federal Civilian Service from President John F. Kennedy, 1962

Canonic SMILES of thalidomide

C1CC(=O)NC(=O)C1N2C(=O)C3=CC=CC=C3C2=O



(-)(S)-thalidomide



(+)(R)-thalidomide

Isomeric SMILES of (-)(S)-thalidomide

C1CC(=O)NC(=O)[C@H]1N2C(=O)C3=CC=CC=C3C2=O

Isomeric SMILES of (+)(R)-thalidomide

C1CC(=O)NC(=O)[C@@H]1N2C(=O)C3=CC=CC=C3C2=O

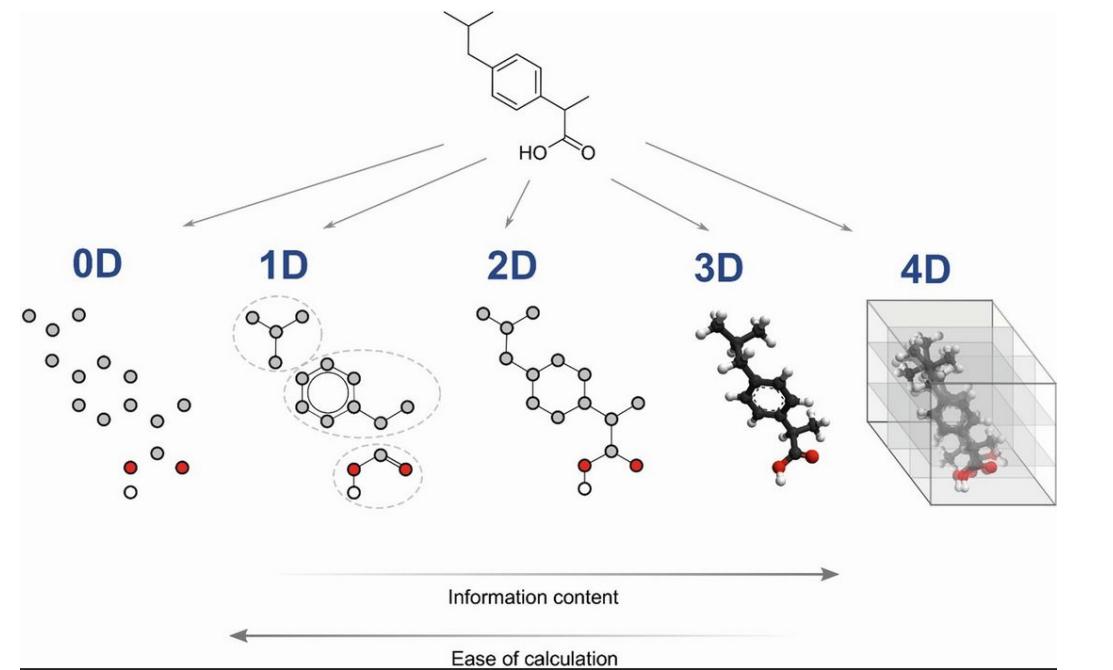
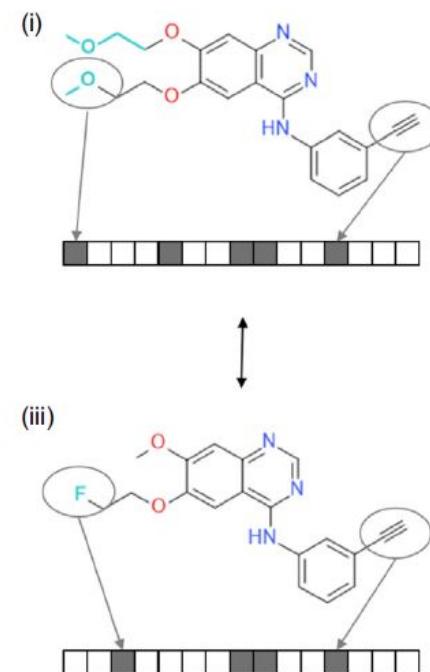
# Molecular descriptors: numeric values that describe chemical molecules.

In contrast to symbolic representations, molecular descriptors enable **quantification of molecular properties**. It allows mathematical operations and statistical analysis that associate biophysical/biochemical properties with molecule structures.



$$\log P_{\text{oct/wat}} = \log \left( \frac{[\text{solute}]^{\text{un-ionized}}_{\text{octanol}}}{[\text{solute}]^{\text{un-ionized}}_{\text{water}}} \right)$$

**logP** is an experimental molecular descriptor. Calculated version (cLogP) exists as well.



- |  |  |  |  |  |
|--|--|--|--|--|
| <ul style="list-style-type: none"> <li>- Atom count</li> <li>- Molecular weight</li> <li>- Sum of atomic properties</li> </ul> | <ul style="list-style-type: none"> <li>- Fragment counts, e.g. # of -OH</li> <li>- Fingerprints</li> </ul> | <ul style="list-style-type: none"> <li>- Topological descriptors, e.g. the Wiener Index, sum of lengths of the shortest paths between all non-H atoms</li> </ul> | <ul style="list-style-type: none"> <li>- Geometrical</li> <li>- Atomic coordinates</li> <li>- Energy grid</li> </ul> | <ul style="list-style-type: none"> <li>- Combination of atomic coordinates and sampling of possible conformations</li> </ul> |
|--|--|--|--|--|

# Lipinski's Rule of Five of small-molecule drugs

- **HBD $\leq$ 5:** No more than **5 hydrogen-bond donors**, e.g. the total number of nitrogen–hydrogen and oxygen–hydrogen bonds.
- **HBA $\leq$ 10:** No more than **10 hydrogen-bond acceptors**, e.g. all nitrogen or oxygen atoms
- **MW $<$ 500:** A **molecular weight** less than **500 Daltons, or 500 g/mol.**  
Reference: ATP has a molecular mass of  $\sim$ 507.
- **logP $\leq$ 5:** An **octanol-water partition coefficient (log P)** that does not exceed **5.** (10-based)

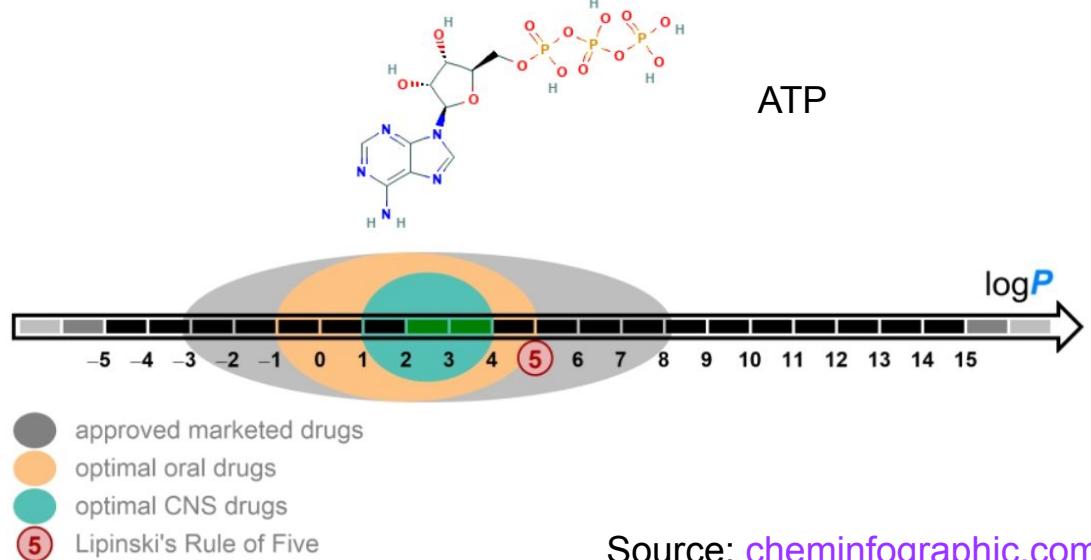


Table 1. New FDA Approvals (2014 to Present)<sup>a</sup> of Oral bRo5 Drugs

drug	year approved	therapeutic area	MW	cLogP	HBD	N+O
velpatasvir	2016	HCV	883.02	2.5	4	16
venetoclax	2016	oncology	868.44	10.4	3	14
elbasvir	2016	HCV	882.0	2.6	4	16
grazoprevir	2016	HCV	766.90	-2.0	3	15
cobimetinib	2015	oncology	531.31	5.2	3	5
daclatasvir	2015	HCV	738.88	1.3	4	14
edoxaban	2015	cardiovascular	548.06	-0.9	3	11
ombitasvir	2014	HCV	894.13	1.3	4	15
paritaprevir	2014	HCV	765.89	1.1	3	14
netupitant	2014	nausea from chemotherapy	578.59	6.8	0	5
ledipasvir	2014	HCV	889.00	0.9	4	14
ceritinib	2014	oncology	558.14	6.5	3	8

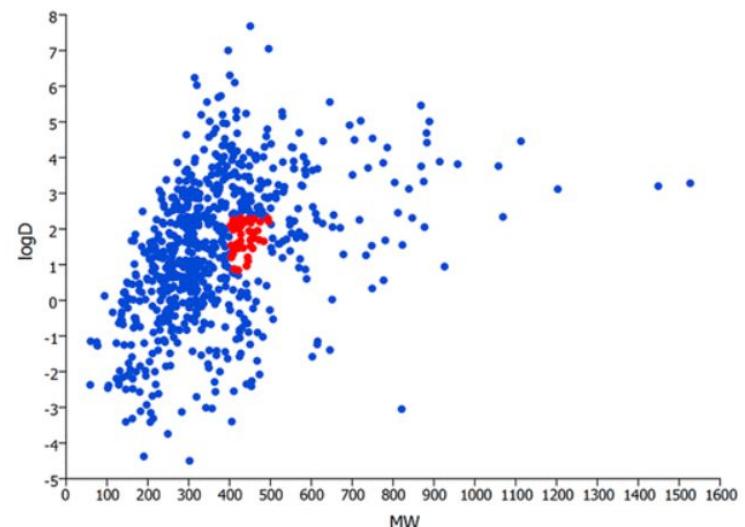
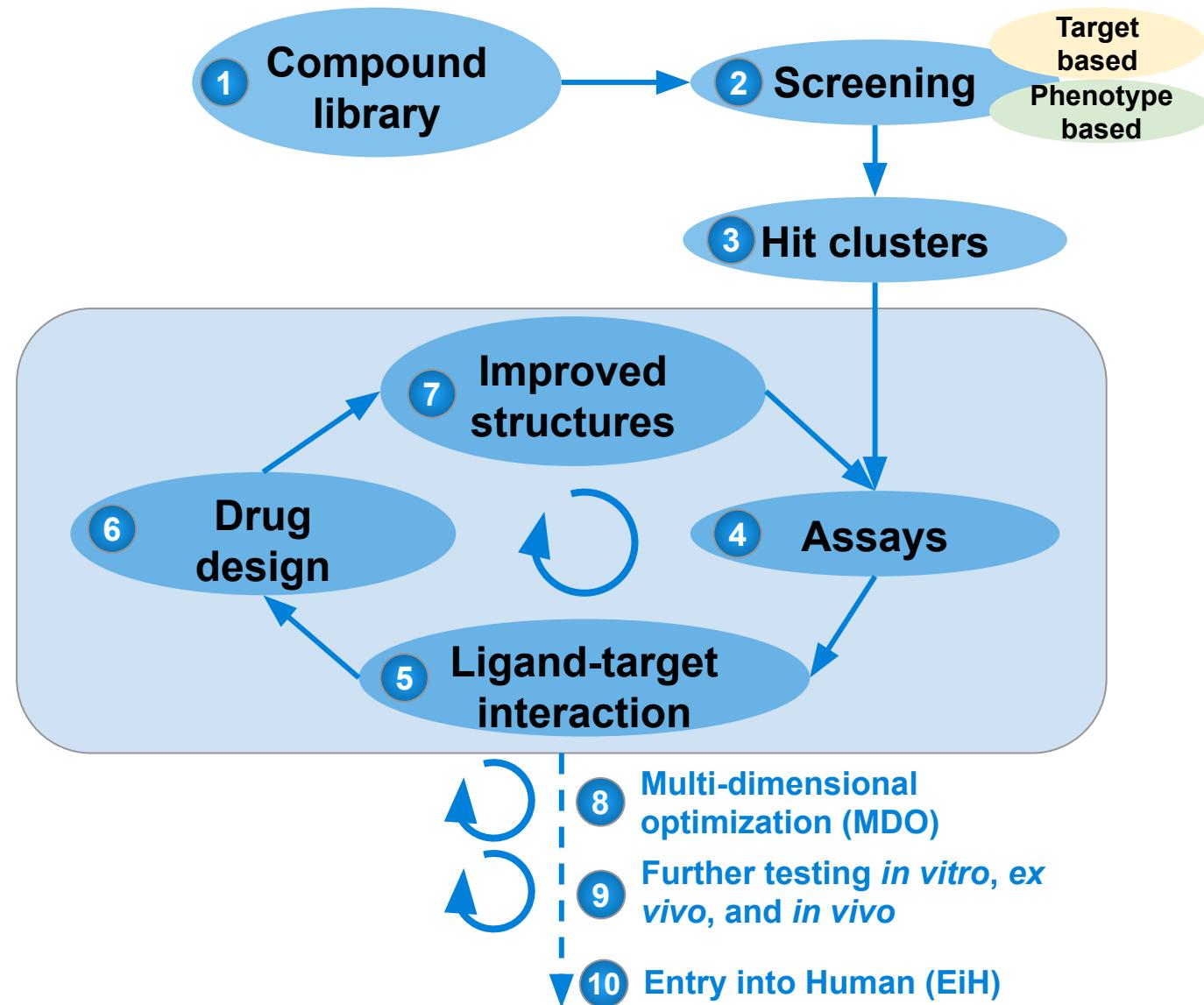


Figure 7: Plot of MW vs cLogD of FDA approved oral drugs. Red points: ‘high probability area’ supposed by (questionable) data analysis. Shultz, Michael D. 2019. “[Two Decades under the Influence of the Rule of Five and the Changing Properties of Approved Oral Drugs.](#)” Journal of Medicinal Chemistry 62 (4): 1701–14.

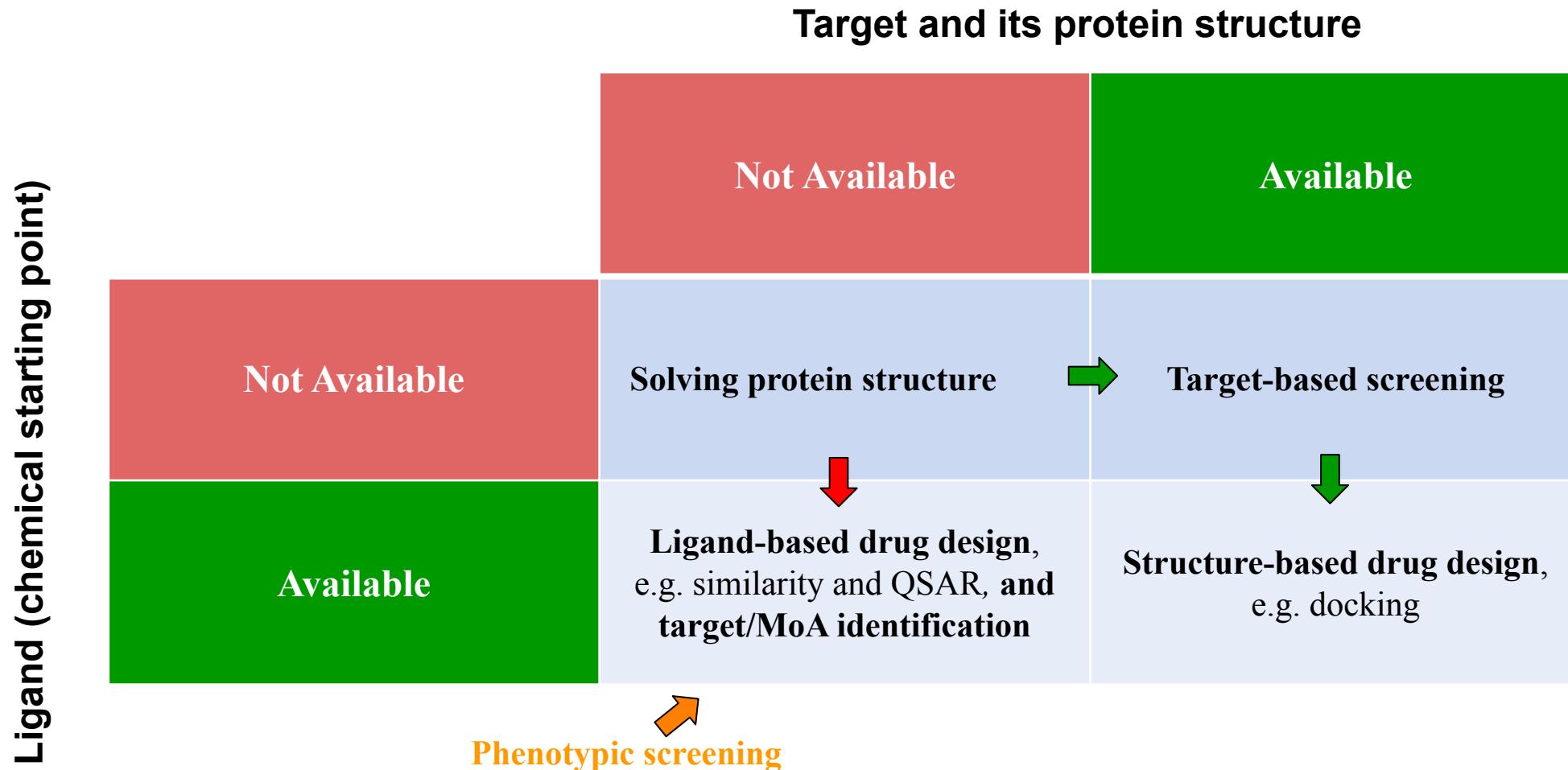
DeGoey, et al.. 2018. “[Beyond the Rule of Five: Lessons Learned from AbbVie’s Drugs and Compound Collection.](#)” Journal of Medicinal Chemistry 61 (7): 2636–51.

# Workflow in a typical drug-discovery program

1. Compound library construction;
2. Screening compounds with ***bioassays***, or **assays**, which determine potency of a chemical by its effect on biological entities: proteins, cells, etc;
3. Hit identification and clustering;
4. More assays, complementary to the assays used in the screening, maybe of lower throughput but more biologically relevant;
5. Analysis of ligand-target interactions, for instance by getting the co-structure of both protein (primary target, and off-targets if necessary) and the hit;
6. *Drug design*, namely to modify the structure of the drug candidate;
7. Analog synthesis and testing (back to step 4);
8. Multidimensional Optimization (MDO), with the goal to optimize potency, selectivity, safety, bioavailability, etc;
9. Further *in vitro*, *ex vivo*, and *in vivo* testing, and preclinical development;
10. Entry into human (Phase 0 or phase 1 clinical trial).



# Ligand-based and structure-based drug design



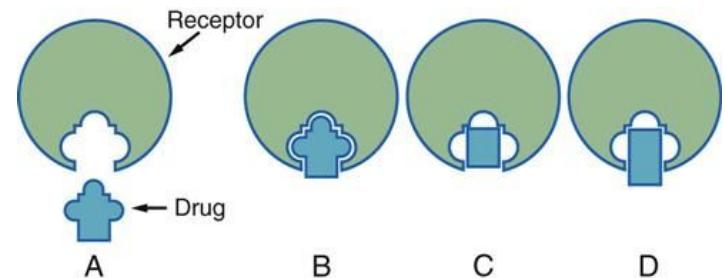
QSAR= quantitative structure activity relationship; MoA= mechanism of action, or mode of action

# Selected mathematical concepts

- **Affinity**
  - The (bio)physical view
  - The (bio)chemical view
- The **Michaelis-Menten model** and enzymatic kinetics
- Example of structure-based drug design: **molecular docking**
- Example of ligand-based drug design: **similarity and quantitative structure-activity relationship (QSAR)**

# The biophysical and biochemical views of ligand-target binding

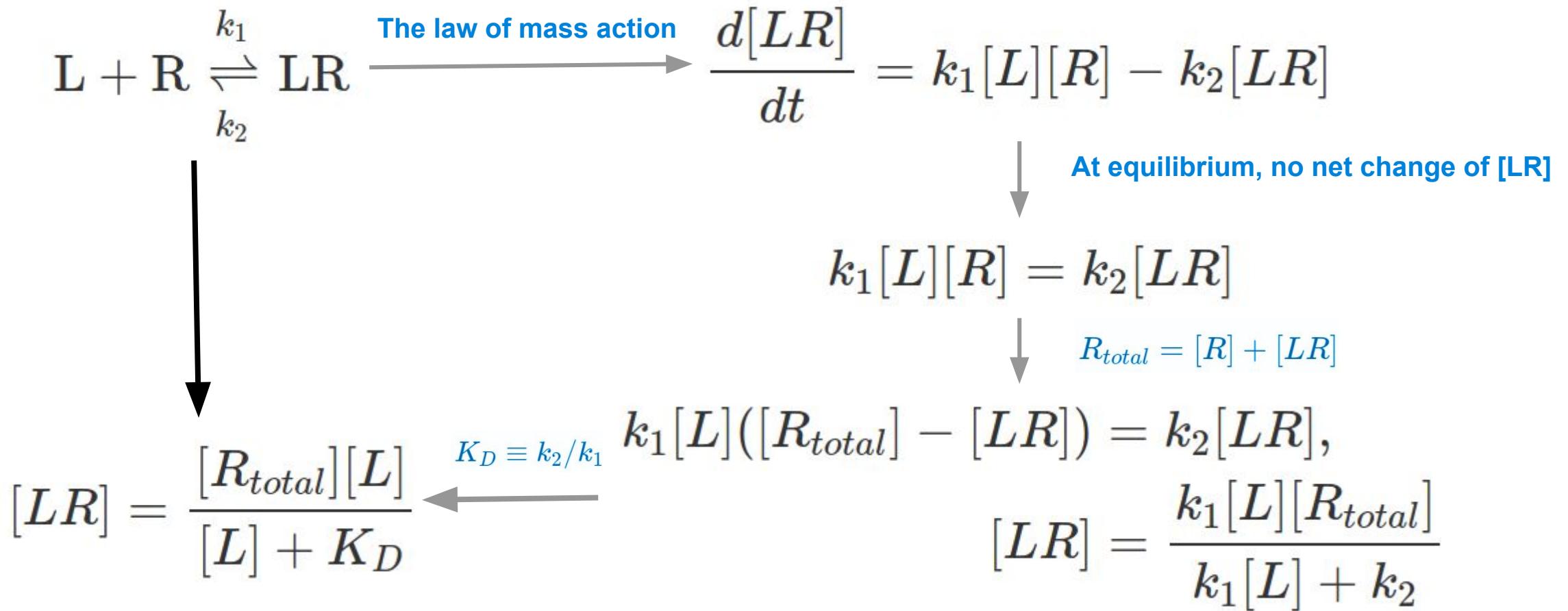
- A **ligand** is a substance that forms a complex with a biomolecule to serve a biological purpose. For instance, a drug can produce a signal by binding to a site on a target protein.
  - A ligand that binds to and alters the function of the receptor that triggers a physiological response is called a receptor **agonist**.
  - A ligand that binds to a receptor but fail to activate the physiological response is a receptor **antagonist**.
- **The biophysical view of binding:** Binding occurs in favourable steric, *i.e.* spatial, configurations (**the ‘lock-and-the-key’ model**) and is mediated by intermolecular forces, such as electrostatic interactions (ionic bonds, hydrogen bonds), Van der Waals forces (dipole interactions),  $\pi$ -effects (interactions of  $\pi$ -orbitals of a molecular system), and hydrophobic effect. Both enthalpy and entropy contribute to the binding energy.
- **The biochemical view of binding:** The *rate* of binding is called affinity, often expressed in  $K_d$  or, for inhibitors,  $K_i$ . A closely related, and often confusing, concept is  $IC_{50}$ . We will talk about them in the next lecture when we talk about the Michaelis-Menten model, the dose-response curve, and the Hill function.
- **Binding affinity data alone does not determine the overall potency of a drug.** Potency depends on binding affinity, the ligand efficacy, and many other factors.



Competitive	Uncompetitive
$E + S \xrightleftharpoons[K_m]{+} ES \xrightarrow{k_{cat}} E + P$ $K_i \parallel$ EI	$E + S \xrightleftharpoons[K_m]{+} ES \xrightarrow{k_{cat}} E + P$ $K_i \parallel$ ESI
Non-competitive	Mixed
$E + S \xrightleftharpoons[K_m]{+} ES \xrightarrow{k_{cat}} E + P$ $K_i \parallel$ $I \parallel$ $E + S \xrightleftharpoons[K_m]{+} ESI$	$E + S \xrightleftharpoons[K_m]{+} ES \xrightarrow{k_{cat}} E + P$ $K_i \parallel$ $I \parallel$ $E + S \xrightleftharpoons[\alpha K_m]{+} ESI$

Four basic types of kinetic mechanism of inhibition, source: [sciencesnail.com](http://sciencesnail.com)

# From the law of mass action to ligand-target interaction



# Four classical classes of mathematical models

## Compartment models

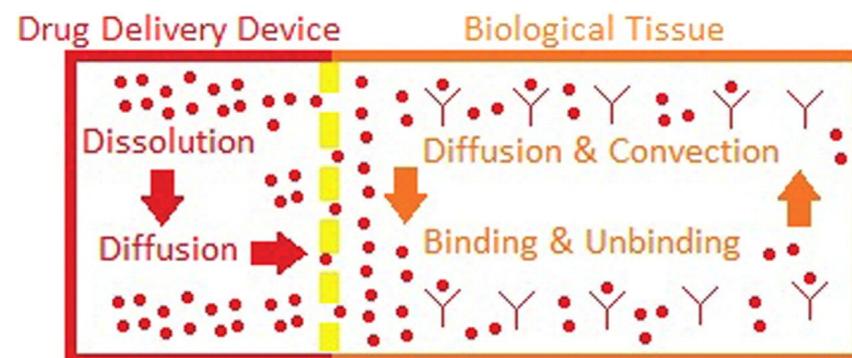
$$\begin{aligned} \frac{dx}{dt} &= \alpha x - \beta xy, \\ \frac{d[LR]}{dt} &= k_1[L][R] - k_2[LR] \\ \frac{dy}{dt} &= -\gamma y + \delta xy, \end{aligned}$$

Kinetics of ligand-target interaction

$$\begin{aligned} \frac{dS}{dt} &= -\frac{\beta IS}{N}, \\ \frac{dI}{dt} &= \frac{\beta IS}{N} - \gamma I, \\ \frac{dR}{dt} &= \gamma I \end{aligned}$$

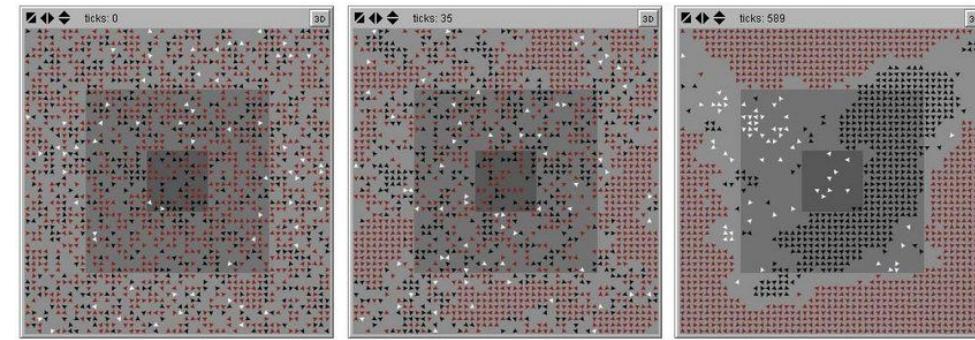
The SIR (S=susceptible, I=infectious, R=removed) model of epidemiology

## Transport models



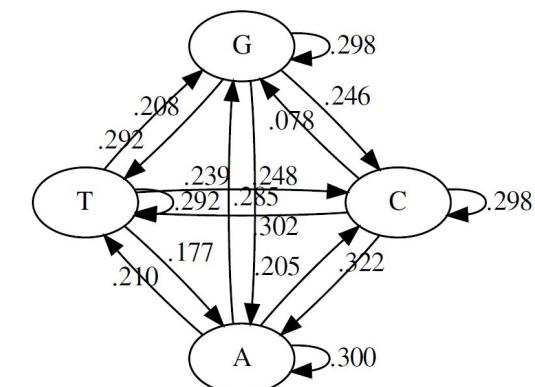
McGinty, Sean, and Giuseppe Pontrelli. 2015. "[A General Model of Coupled Drug Release and Tissue Absorption for Drug Delivery Devices](#)." *Journal of Controlled Release* 217 (November): 327–36.

## Particle models



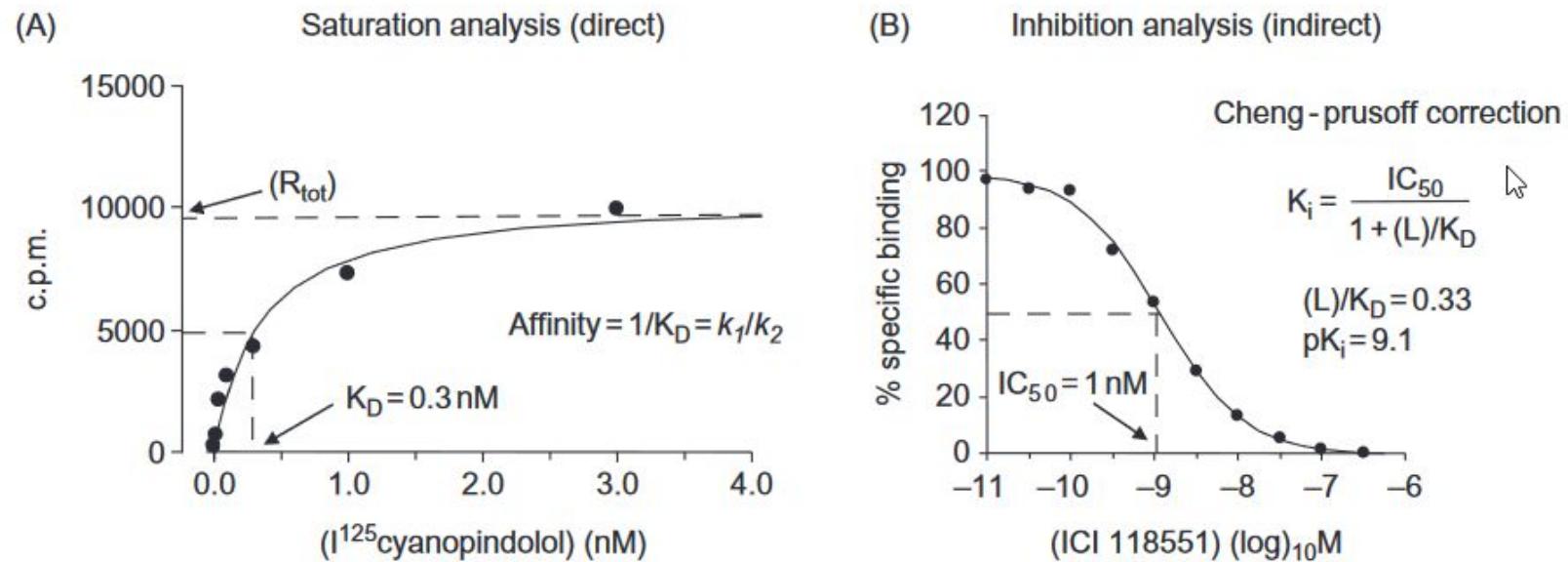
A Study on Socio-spatial Segregation Models Based on Multi-agent Systems by Quadros *et al.* (2012).  
 10.1109/BWSS.2012.14.

## Finite state models



A finite-state Markov chain modelling DNA sequences

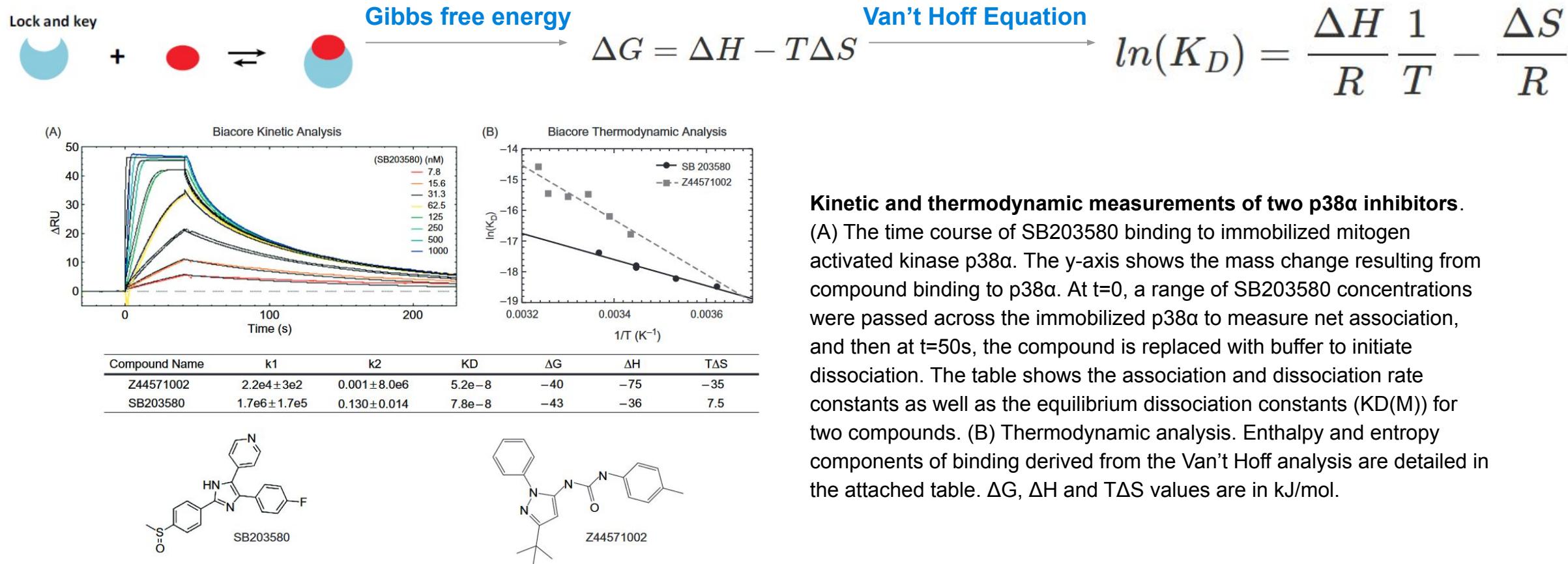
# The biochemical (kinetic) view of binding affinity: the hyperbola curve and the dissociation constant $K_D$



**Binding assays with direct and indirect measurements.** (A) A direct binding assay using  $I^{125}$  labelled cyanopindolol as a  $\beta 2$ -adrenoceptor ligand. The curve describes a rectangular hyperbola which saturates at high ligand concentration. The ligand dissociation constant ( $K_D$ ) was estimated as 0.3 nM and is a measure of the ligand affinity. (B) A typical inhibition analysis using membranes expressing the human  $\beta 2$ -adrenoceptor and employing 0.1 nM  $I^{125}$  cyanopindolol as the labeled ligand. The displacing ligand, the selective  $\beta 2$ -adrenoceptor antagonist ICI 118551, produces complete inhibition of the specific binding yielding an  $IC_{50}$  of 1 nM. From *Evaluation of the Biological Activity of Compounds: Techniques and Mechanism of Action Studies*, by Iain G. Dougall and John Unitt.

Questions: (1) how can we interpret the hyperbola curve? (2) if  $f(x)$  is a function with the form of  $Ax/(k+x)$ , what will be the form of function  $g(f(x))$  where  $g(x)=Bx/(k'+x)$ ? What implications does this have?

# The biophysical (thermodynamic) view of binding affinity: enthalpy and entropy

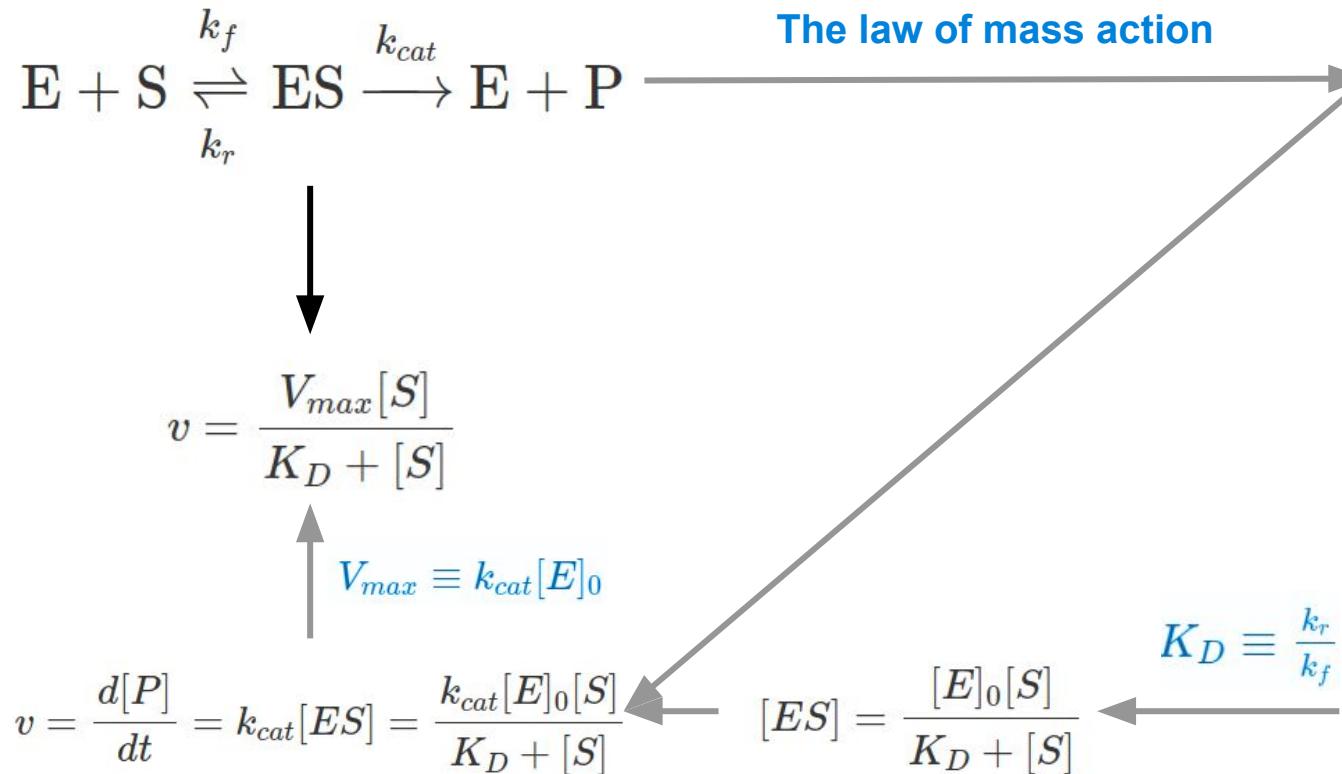


## Kinetic and thermodynamic measurements of two p38α inhibitors.

(A) The time course of SB203580 binding to immobilized mitogen activated kinase p38α. The y-axis shows the mass change resulting from compound binding to p38α. At t=0, a range of SB203580 concentrations were passed across the immobilized p38α to measure net association, and then at t=50s, the compound is replaced with buffer to initiate dissociation. The table shows the association and dissociation rate constants as well as the equilibrium dissociation constants ( $K_D(M)$ ) for two compounds. (B) Thermodynamic analysis. Enthalpy and entropy components of binding derived from the Van't Hoff analysis are detailed in the attached table.  $\Delta G$ ,  $\Delta H$  and  $T\Delta S$  values are in kJ/mol.

For a thorough discussion about enthalpic and entropic contributions to molecular interactions, see [A Medicinal Chemist's Guide to Molecular Interactions](#) (Journal of Medicinal Chemistry 53 (14): 5061–84) by Bissantz et al.

# Modelling enzyme kinetics with the Michaelis-Menten model



$$\frac{d[E]}{dt} = -k_f[E][S] + k_r[ES] + k_{cat}[ES],$$

$$\frac{d[S]}{dt} = -k_f[E][S] + k_r[ES],$$

$$\frac{d[ES]}{dt} = k_f[E][S] - k_r[ES] - k_{cat}[ES],$$

$$\frac{d[P]}{dt} = k_{cat}[ES],$$

**Assuming that  $k_f[E][S] = k_r[ES]$**

$$k_f([E]_0 - [ES])[S] = k_r[ES]$$

$$k_f[E]_0[S] - k_f[ES][S] = k_r[ES]$$

$$k_f[E]_0[S] = k_r[ES] + k_f[ES][S]$$

$$k_f[E]_0[S] = [ES](k_r + k_f[S])$$

$$[ES] = \frac{k_f[E]_0[S]}{k_r + k_f[S]}$$

$$[ES] = \frac{k_f[E]_0[S]}{k_f(\frac{k_r}{k_f} + [S])}$$

# The dose-response curve and IC<sub>50</sub>: The Hill function and *in vitro* pharmacology

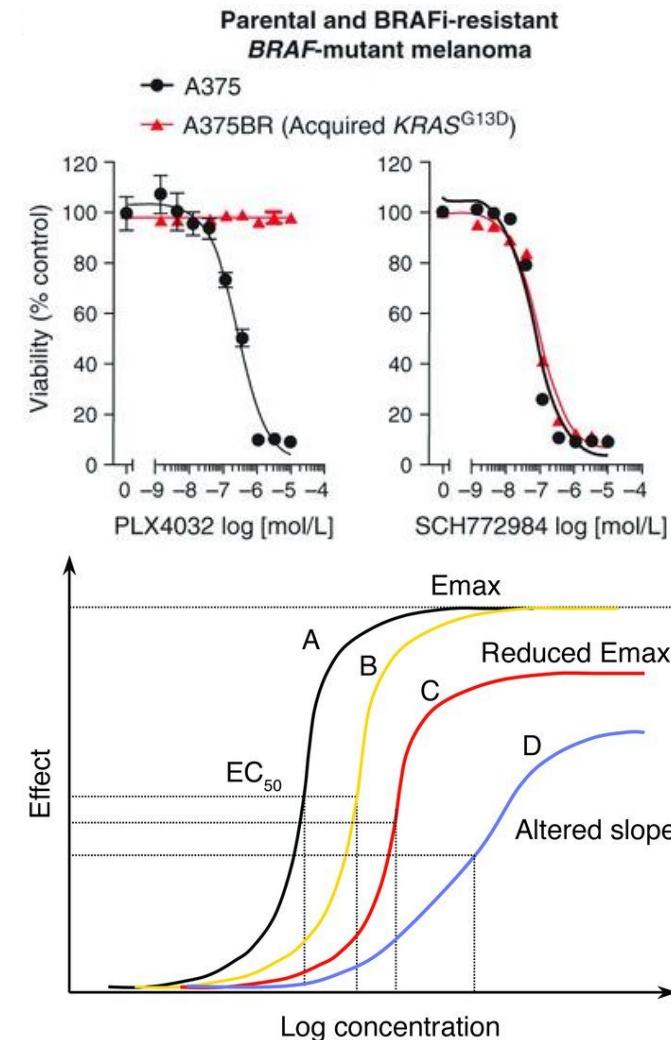
- The Hill function is one of the mostly useful non-linear functions to model biological systems.
- In its general form,  $H_{max}$  indicates the maximal value to which the function is asymptotic,  $n$  is the shape parameter (known as the Hill's coefficient), and  $k$  is the reflection point, often abbreviated as  $XC_{50}$  ( $X=I, E, C, \dots$ ), the half-saturation constant.
- The Michaelis-Menten model is a special case of the Hill function with  $n=1$ .

$$H = H_{max} \frac{x^n}{k^n + x^n}$$

**General form of the Hill function**

$$\begin{aligned} E &= E_{max} \frac{[L]^n}{EC_{50}^n + [L]^n} \\ &= E_{max} \frac{1}{1 + \left(\frac{EC_{50}}{[L]}\right)^n} \end{aligned}$$

**Modelling dose-dependent effect**



[Morris et al. Cancer Discov; 3\(7\): 742-50.](#)  
©2013 AACR.

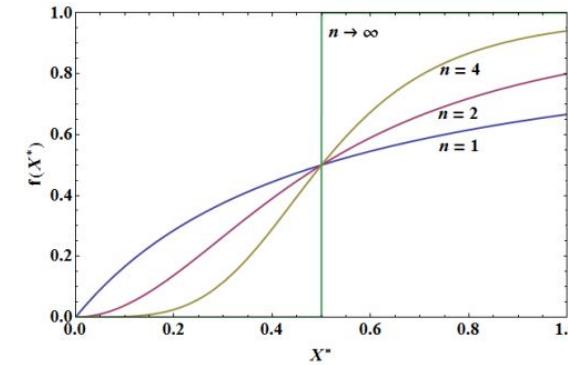
White. J Clin Invest. 2004;113(8):1084-1092.  
<https://doi.org/10.1172/JCI21682>.

Suppose it is an antiviral drug, compared with curve B, what does curve A, C, and D suggest?

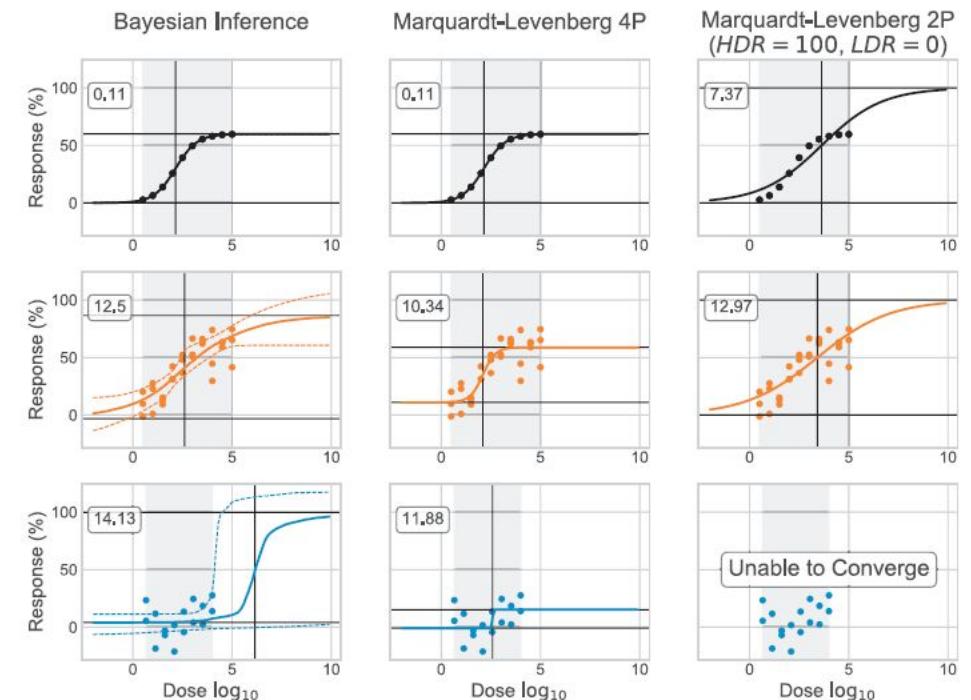
# More about the Hill function and dose-response curves

- The Hill function is often used to model either *target occupancy* or *tissue response*. In pharmacology, it is often used to model the tissue response.
- The Hill function can be approximated by a step function when  $n$  goes towards infinity (top panel). This can be seen as one of the theoretical foundations of Boolean network modelling.
- The Hill function can be deduced from statistical mechanics of binding, a particle modelling approach. See for instance [an article on Biophysics Wiki by Andreas Piehler](#) for details.
- Data needs to be fit to the model, and in reality data can look quite different from the ideal curve (bottom panel). By setting priors, it is possible to perform inference even with ill-looking data.

The Bayesian inference approach versus the Marquardt-Levenberg algorithm for non-linear regression fitting (an alternative to gradient descent and Gauss-Newton methods). 4P: four parameter model; 2P: two parameter model (IC50 and  $n$ ). Numbers in boxes are root mean square errors of fitting. Figure 2 from Labelle, Caroline, Anne Marinier, and Sébastien Lemieux. 2019. ["Enhancing the Drug Discovery Process: Bayesian Inference for the Analysis and Comparison of Dose–Response Experiments."](#) *Bioinformatics* 35 (14): i464–73.



From [the biophysics wiki article](#) by Andreas Piehler



# The principle of molecular docking, a case study of structure-based drug design

- Docking is like a discotheque: it is all about posing and scoring – Roger Sayle (NextMove Software Limited)
- Three basic methods to represent target and ligand structures *in silico*
  - **Atomic**: used in conjunction with a potential energy function, computational complexity high
  - **Surface**: often used in protein-protein docking
  - **Grid representation**:
    - Basic idea: to store information about the receptor's energetic contributions on grid points so that it only needs to be read during ligand scoring.
    - In the most basic form, grid points store two types of potentials: **electrostatic** and **van der Waals forces**.

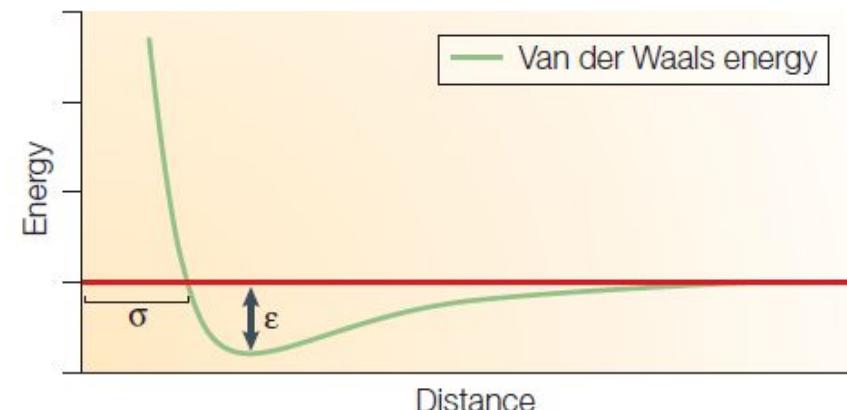
$$E_{coul}(r) = \sum_{i=1}^{N_A} \sum_{j=1}^{N_B} \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}}$$

Coulombic interactions

$$E_{vdW}(r) = \sum_{j=1}^N \sum_{i=1}^N 4\epsilon \left[ \left( \frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left( \frac{\sigma_{ij}}{r_{ij}} \right)^6 \right]$$

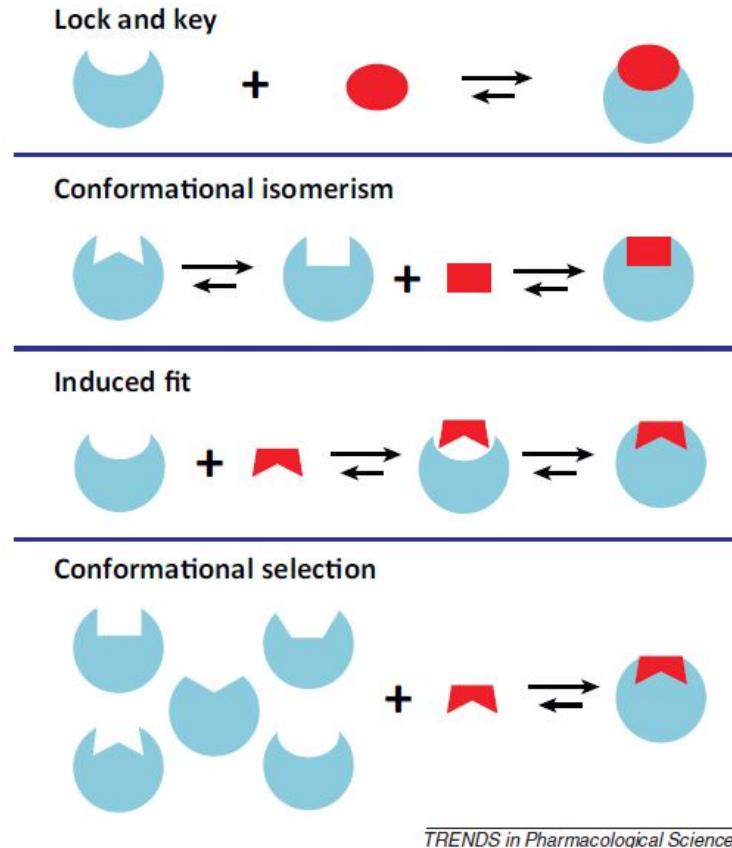
Lennard-Jones 12–6 function

- $\epsilon$  is the **well depth** of the potential
- $\sigma$  is the **collision diameter** of the respective atoms  $i$  and  $j$ .

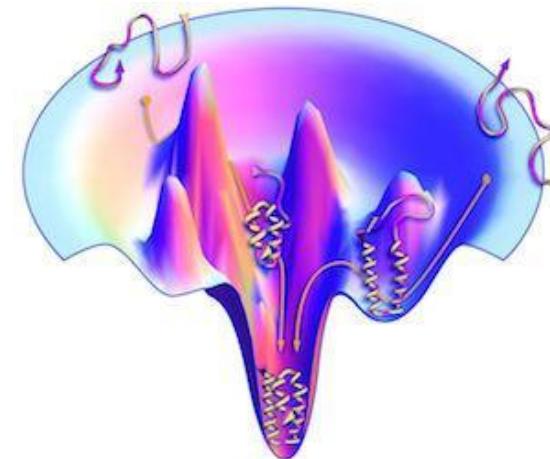
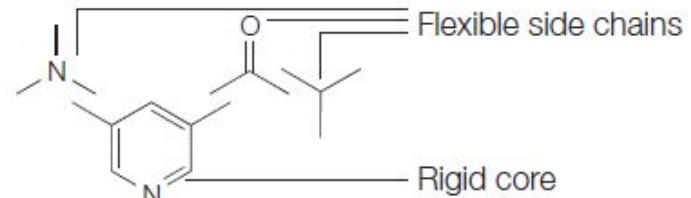


Kitchen, Douglas B., Hélène Decornez, John R. Furr, und Jürgen Bajorath. „Docking and Scoring in Virtual Screening for Drug Discovery: Methods and Applications“. *Nature Reviews Drug Discovery* 3, Nr. 11 (November 2004): 935–49. <https://doi.org/10.1038/nrd1549>.

# Posing: dealing with flexibility of ligand and of protein



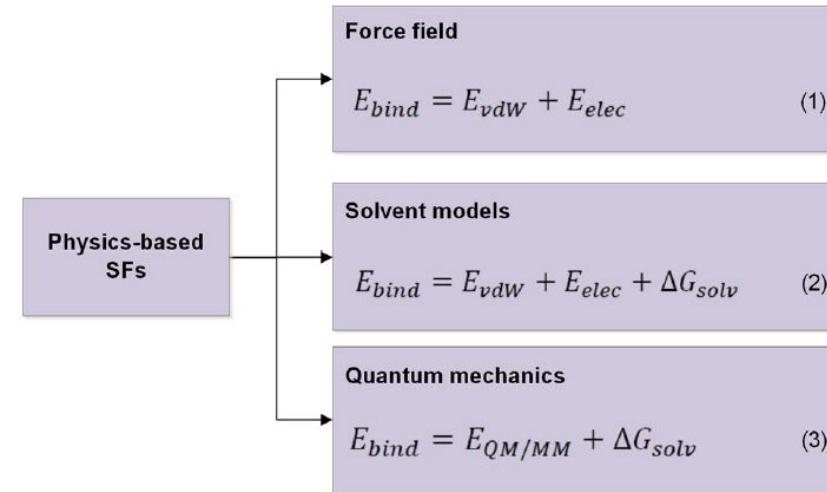
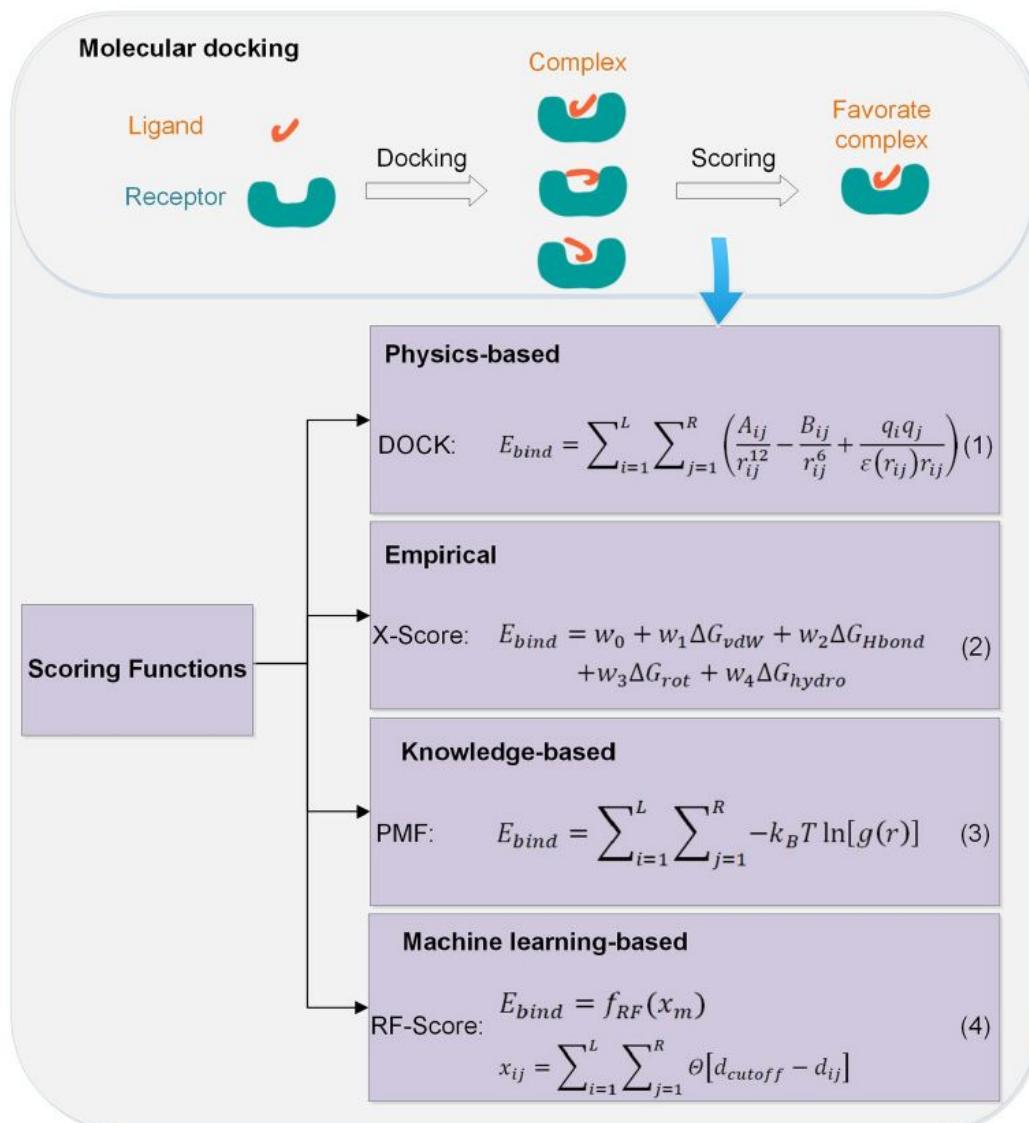
Chen, Yu-Chian. „Beware of docking!“ *Trends in Pharmacological Sciences* 36, Nr. 2 (1. Februar 2015): 78–95.  
<https://doi.org/10.1016/j.tips.2014.12.001>.



## Methods to deal with ligand and protein flexibility

- Systematic search
- Random search, such as Monte-Carlo and genetic algorithms
- Simulation methods, such as molecular dynamics

# Types of scoring functions



- Empirical scoring functions estimate the binding affinity of a complex by **summing up the important energetic factors for protein–ligand binding**, such as hydrogen bonds, hydrophobic effects, steric clashes, etc. It relies on training set and regression analysis.
- Knowledge-based scoring functions derive the desired pairwise potentials from three-dimensional structures of a large set of protein–ligand complexes based **on the inverse Boltzmann distribution**. It is assumed that the frequency of different atom pairs in different distances is related to the interaction of two atoms and converts the frequency into the distance-dependent potential of mean force.
- Machine learning-based scoring functions are usually used for rescoring to improve the initial docking.

Li, Jin, Ailing Fu, und Le Zhang. „An Overview of Scoring Functions Used for Protein–Ligand Interactions in Molecular Docking“. *Interdisciplinary Sciences: Computational Life Sciences* 11, Nr. 2 (1. Juni 2019): 320–28. <https://doi.org/10.1007/s12539-019-00327-w>.

# Interested in learning more about molecular modelling?

PROTOCOL

## Computational protein–ligand docking and virtual drug screening with the AutoDock suite

Stefano Forli, Ruth Huey, Michael E Pique, Michel F Sanner, David S Goodsell & Arthur J Olson

---

Department of Integrative Structural and Computational Biology, The Scripps Research Institute, La Jolla, California, USA. Correspondence should be addressed to A.J.O. (olson@scripps.edu).

Published online 14 April 2016; doi:10.1038/nprot.2016.051

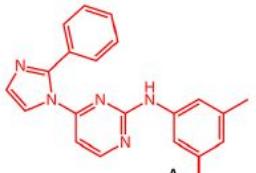
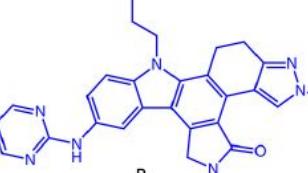
Computational docking can be used to predict bound conformations and free energies of binding for small-molecule ligands to macromolecular targets. Docking is widely used for the study of biomolecular interactions and mechanisms, and it is applied to structure-based drug design. The methods are fast enough to allow virtual screening of ligand libraries containing tens of thousands of compounds. This protocol covers the docking and virtual screening methods provided by the AutoDock suite of programs, including a basic docking of a drug molecule with an anticancer target, a virtual screen of this target with a small ligand library, docking with selective receptor flexibility, active site prediction and docking with explicit hydration. The entire protocol will require ~5 h.

- Try docking yourself by following this protocol: Forli, Stefano, Ruth Huey, Michael E. Pique, Michel F. Sanner, David S. Goodsell, und Arthur J. Olson. „Computational Protein–Ligand Docking and Virtual Drug Screening with the AutoDock Suite“. *Nature Protocols* 11, Nr. 5 (Mai 2016): 905–19. <https://doi.org/10.1038/nprot.2016.051>.
- In-depth reading: Sliwoski, Gregory, Sandeepkumar Kothiwale, Jens Meiler, und Edward W. Lowe. „Computational Methods in Drug Discovery“. *Pharmacological Reviews* 66, Nr. 1 (1. Januar 2014): 334–95. <https://doi.org/10.1124/pr.112.007336>.
- A more advanced talk by Arthur Olson can be found [here](#), Workshop on the Mathematics of Drug Design/Discovery, June 4 - 8, 2018, The Fields Institute.
- Courses available at the University of Basel and beyond.

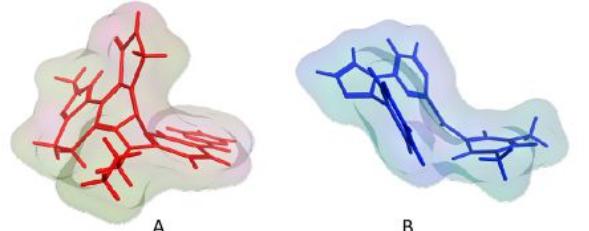
# Molecular similarity and similarity measures

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

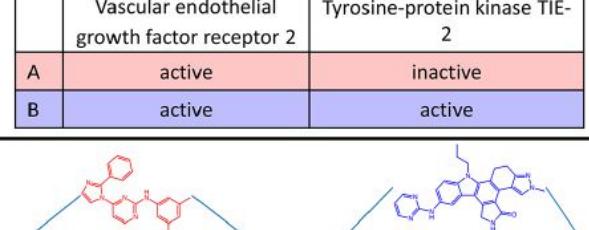
  

Molecular similarity		
----------------------	---	---

2D similarity	
---------------	---

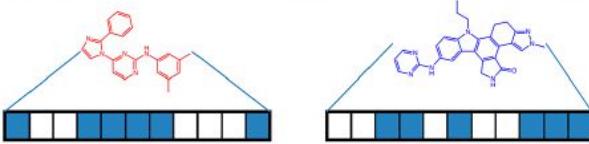
  

3D similarity	
---------------	--

Biological similarity	Vascular endothelial growth factor receptor 2	Tyrosine-protein kinase TIE-2
	A active	inactive
	B active	active

Global similarity	
-------------------	---

Local similarity	
------------------	---

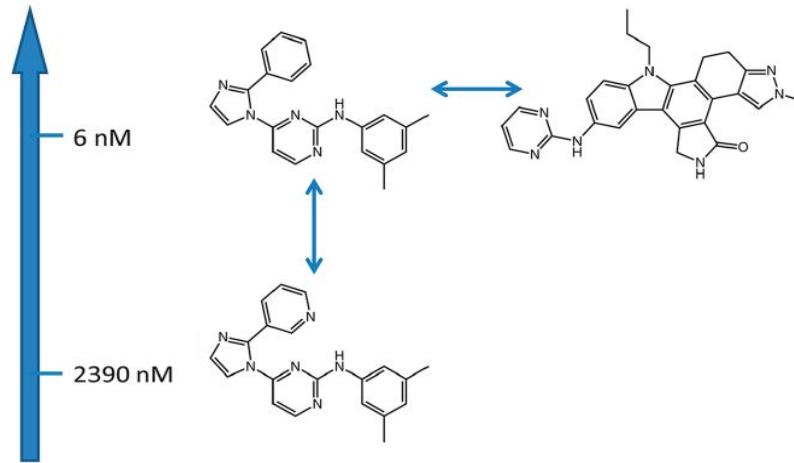
**Table 2 Formulas for the various similarity and distance metrics**

Distance metric	Formula for continuous variables <sup>a</sup>	Formula for dichotomous variables <sup>a</sup>
Manhattan distance	$D_{A,B} = \sum_{j=1}^n  x_{jA} - x_{jB} $	$D_{A,B} = a + b - 2c$
Euclidean distance	$D_{A,B} = \left[ \sum_{j=1}^n (x_{jA} - x_{jB})^2 \right]^{1/2}$	$D_{A,B} = [a + b - 2c]^{1/2}$
Cosine coefficient	$S_{A,B} = \left[ \sum_{j=1}^n x_{jA} x_{jB} \right] / \left[ \sum_{j=1}^n (x_{jA})^2 \sum_{j=1}^n (x_{jB})^2 \right]^{1/2}$	$S_{A,B} = \frac{c}{[ab]^{1/2}}$
Dice coefficient	$S_{A,B} = \left[ 2 \sum_{j=1}^n x_{jA} x_{jB} \right] / \left[ \sum_{j=1}^n (x_{jA})^2 + \sum_{j=1}^n (x_{jB})^2 \right]$	$S_{A,B} = 2c/[a + b]$
Tanimoto coefficient	$S_{A,B} = \frac{\sum_{j=1}^n x_{jA} x_{jB}}{\sum_{j=1}^n (x_{jA})^2 + \sum_{j=1}^n (x_{jB})^2 - \sum_{j=1}^n x_{jA} x_{jB}}$	$S_{A,B} = c/[a + b - c]$
Soergel distance <sup>b</sup>	$D_{A,B} = \left[ \sum_{j=1}^n  x_{jA} - x_{jB}  \right] / \left[ \sum_{j=1}^n \max(x_{jA}, x_{jB}) \right]$	$D_{A,B} = 1 - \frac{c}{[a+b-c]}$

$S$  denotes similarities, while  $D$  denotes distances. The two can be converted to each other by  $\text{similarity}=1/(1+\text{distance})$ .  $x_{jA}$  means the  $j$ -th feature of molecule A.  $a$  is the number of *on* bits in molecule A,  $b$  is number of *on* bits in molecule B, while  $c$  is the number of bits that are *on* in both molecules.

(Left) Maggiora, Gerald, Martin Vogt, Dagmar Stumpfe, und Jürgen Bajorath. „[Molecular Similarity in Medicinal Chemistry](#)“. *Journal of Medicinal Chemistry* 57, Nr. 8 (24. April 2014): 3186–3204. (Right) Bajusz, Dávid, Anita Rácz, and Károly Héberger. 2015. “[Why Is Tanimoto Index an Appropriate Choice for Fingerprint-Based Similarity Calculations?](#)” *Journal of Cheminformatics* 7 (1): 20.

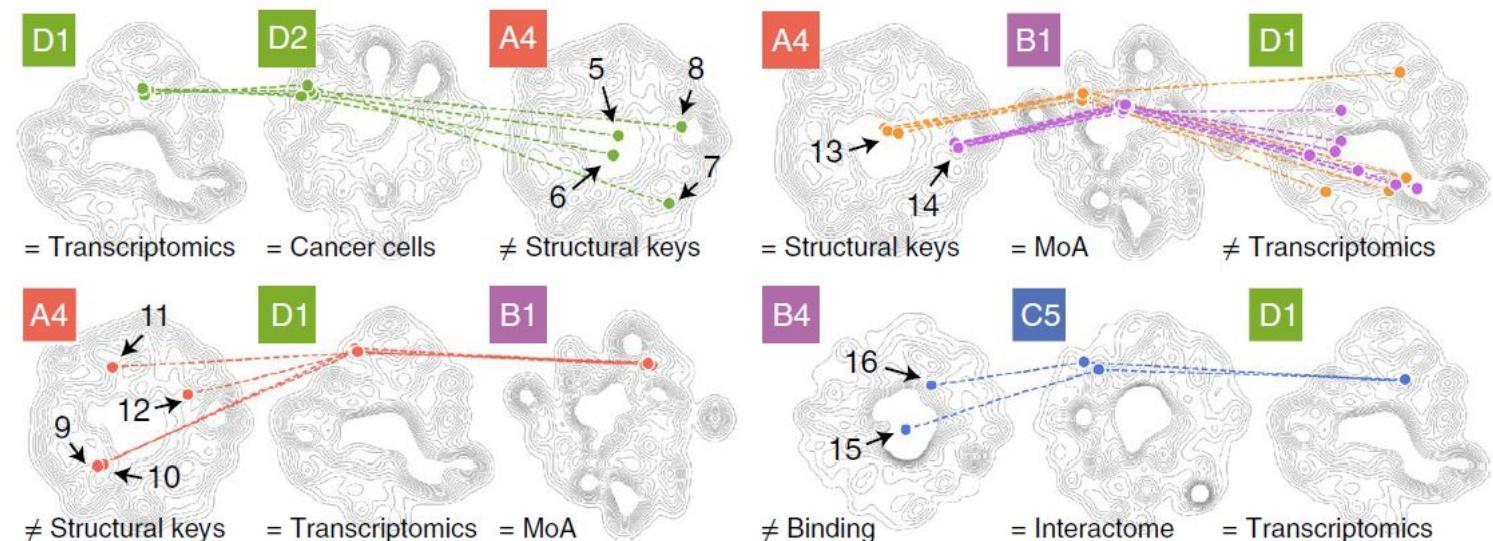
# Molecular similarity does not equal biological similarity



**Watch out biological activity cliffs!** Similarity does not imply activity. Three vascular endothelial growth factor receptor 2 (VEGFR2) ligands are shown that represent different similarity–activity relationships.

a	1	2	3	4	5	A1: 2D fingerprints	A2: 3D fingerprints	A3: Screens	A4: Structural keys	A5: Physicochemistry
A	Red	Red	Red	Red	Red	B1: Mechanisms of action	B2: Metabolic genes	B3: Crystals	B4: Binding	B5: HTS bioassays
B	Purple	Purple	Purple	Purple	Purple	C1: Small molecule roles	C2: Small molecule pathways	C3: Signaling pathways	C4: Biological processes	C5: Interactome
C	Blue	Blue	Blue	Blue	Blue	D1: Transcription	D2: Cancer cell lines	D3: Chemical genetics	D4: Morphology	D5: Cell bioassays
D	Green	Green	Green	Green	Green	E1: Therapeutic areas	E2: Indications	E3: Side effects	E4: Diseases & toxicology	E5: Drug–drug interactions
E	Orange	Orange	Orange	Orange	Orange					

- A: Chemistry**
- B: Targets**
- C: Biological network**
- D: Cells**
- E: Clinical readout**



Duran-Frigola, Miquel, Eduardo Pauls, Oriol Guitart-Pla, Martino Bertoni, Víctor Alcalde, David Amat, Teresa Juan-Blanco, and Patrick Aloy. 2020. “[Extending the Small-Molecule Similarity Principle to All Levels of Biology with the Chemical Checker](#).” Nature Biotechnology, May, 1–10.

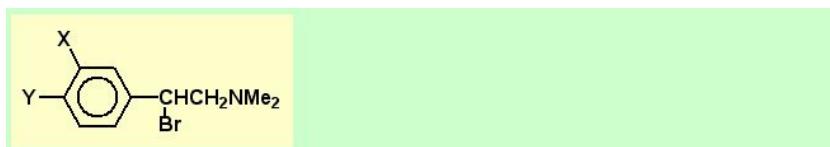
# Quantitative Structure-Activity Relationships (QSARs)

QSAR is a statistical modelling of correlation between biological activity and physicochemical properties, or  $\Delta\phi=f(\Delta S)$ , where  $\phi$  indicates a biological activity and  $S$  indicates a chemical structure (1868-1869).

		Molecular Descriptors (MD)					
	Target property	MD <sub>1</sub>	MD <sub>2</sub>	...	MD <sub>M</sub>		
C <sub>1</sub>	y <sub>1</sub>	x <sub>1,1</sub>	x <sub>1,2</sub>	...	x <sub>1,M</sub>		
C <sub>2</sub>	y <sub>2</sub>	x <sub>2,1</sub>	...	...	...		
C <sub>3</sub>	y <sub>3</sub>	...	...	...	...		
C <sub>4</sub>	y <sub>4</sub>	...	...	...	...		
...	...	...	...	...	...		
...	...	...	...	...	...		
C <sub>N</sub>	y <sub>N</sub>	x <sub>N,1</sub>	x <sub>N,2</sub>	...	x <sub>N,M</sub>		

The basic form of a QSAR model: find a function  $f$  that predicts  $y$  from  $x$ ,  $y \sim f(x)$

An example: **The Free-Wilson analysis.** The assumption: the biological activity for a set of analogues could be described by the contributions that substituents or structural elements make to the activity of a parent structure.



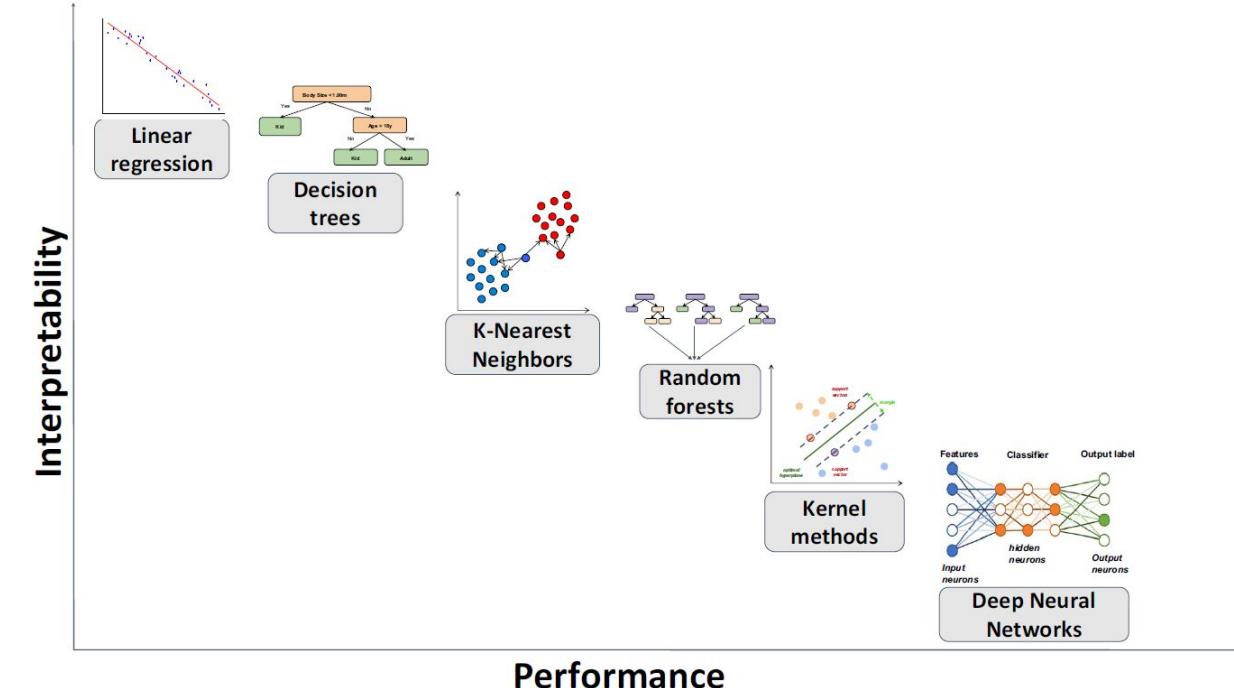
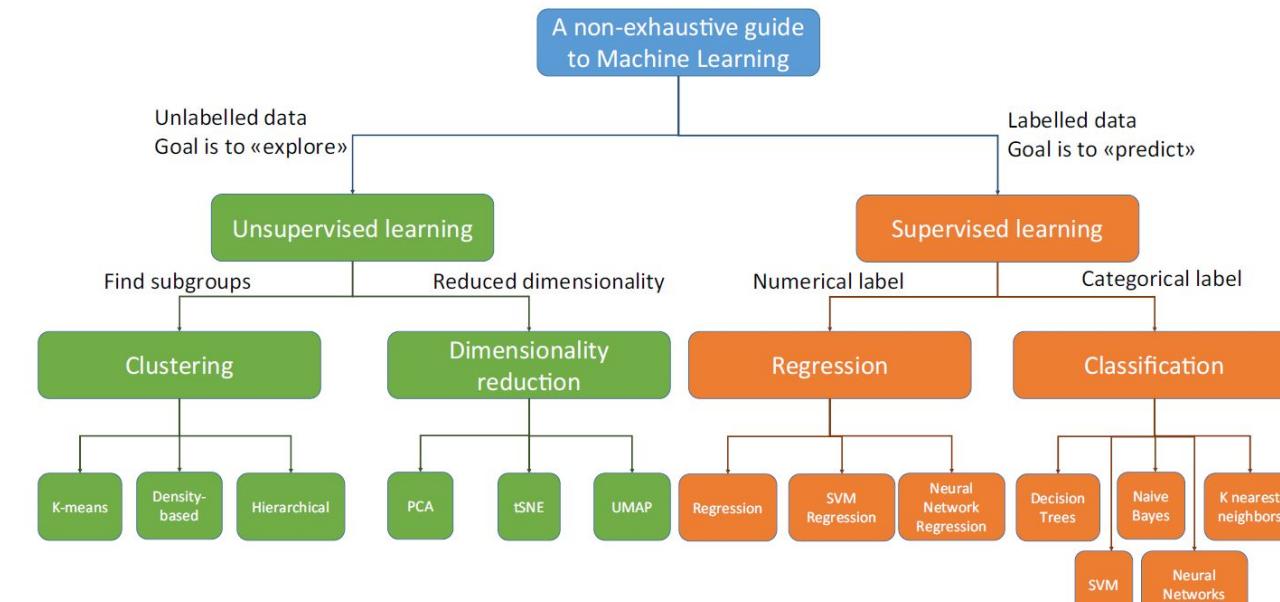
meta	para	meta-	para-					log 1/C	log 1/C					
(X)	(Y)		F	Cl	Br	I	Me	F	Cl	Br	I	Me	obsd.	calc.a)
H	H												7.46	7.82
H	F							1					8.16	8.16
H	Cl								1				8.68	8.59
H	Br									1			8.89	8.84
H	I										1		9.25	9.25
H	Me											1	9.30	9.08
F	H		1										7.52	7.52
Cl	H			1									8.16	8.03
Br	H				1								8.30	8.26
I	H					1							8.40	8.40
Me	H						1						8.46	8.28
Cl	F			1				1					8.19	8.37
Br	F				1				1				8.57	8.60
Me	F					1	1						8.82	8.62
Cl	Cl			1					1				8.89	8.80
Br	Cl				1					1			8.92	9.02
Me	Cl					1		1					8.96	9.04
Cl	Br			1						1			9.00	9.05
Br	Br				1						1		9.35	9.28
Me	Br					1					1		9.22	9.30
Me	Me						1					1	9.30	9.53
Br	Me							1					9.52	9.51

Multivariate regression analysis

$$\begin{aligned} \log(1/ED_{50}) = & -0.301[m-F] + 0.27[m-Cl] + 0.434[m-Br] + 0.579[m-I] \\ & + 0.454[m-Me] + 0.340[p-F] + 0.768[p-Cl] + 1.020[p-Br] \\ & + 1.429[p-I] + 1.256[p-Me] + 7.821 \\ n = 22, r^2 = 0.94, s = 0.194, F = 17.0 \end{aligned}$$

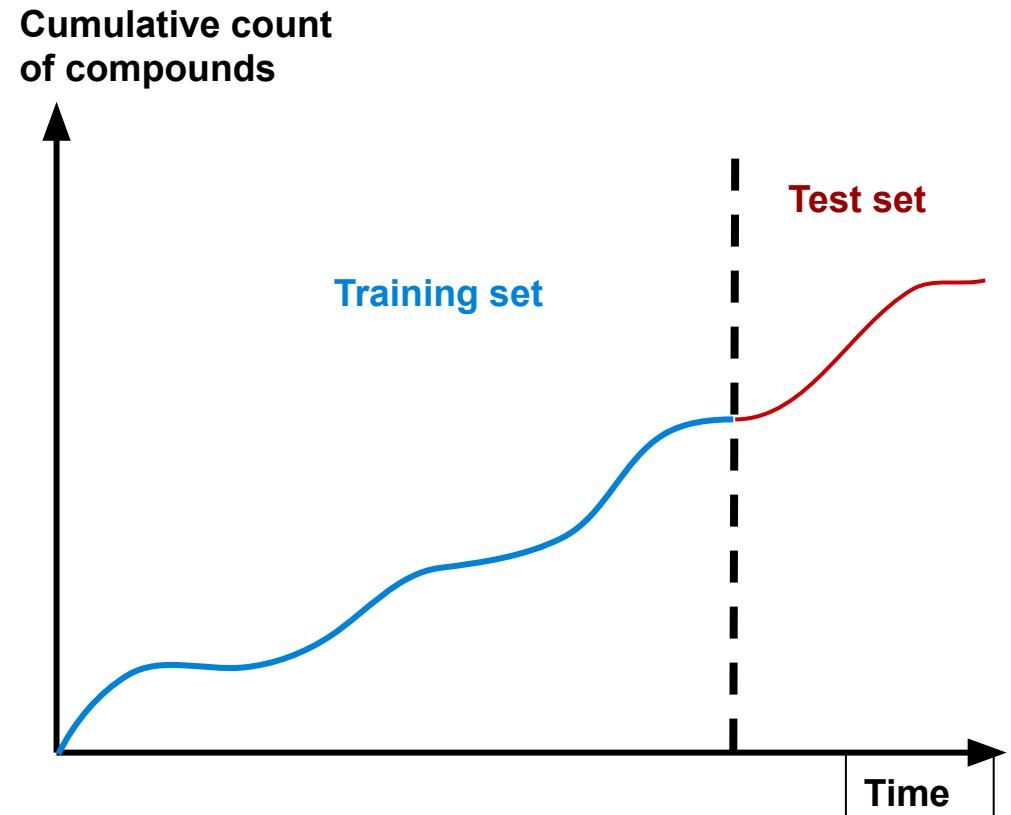
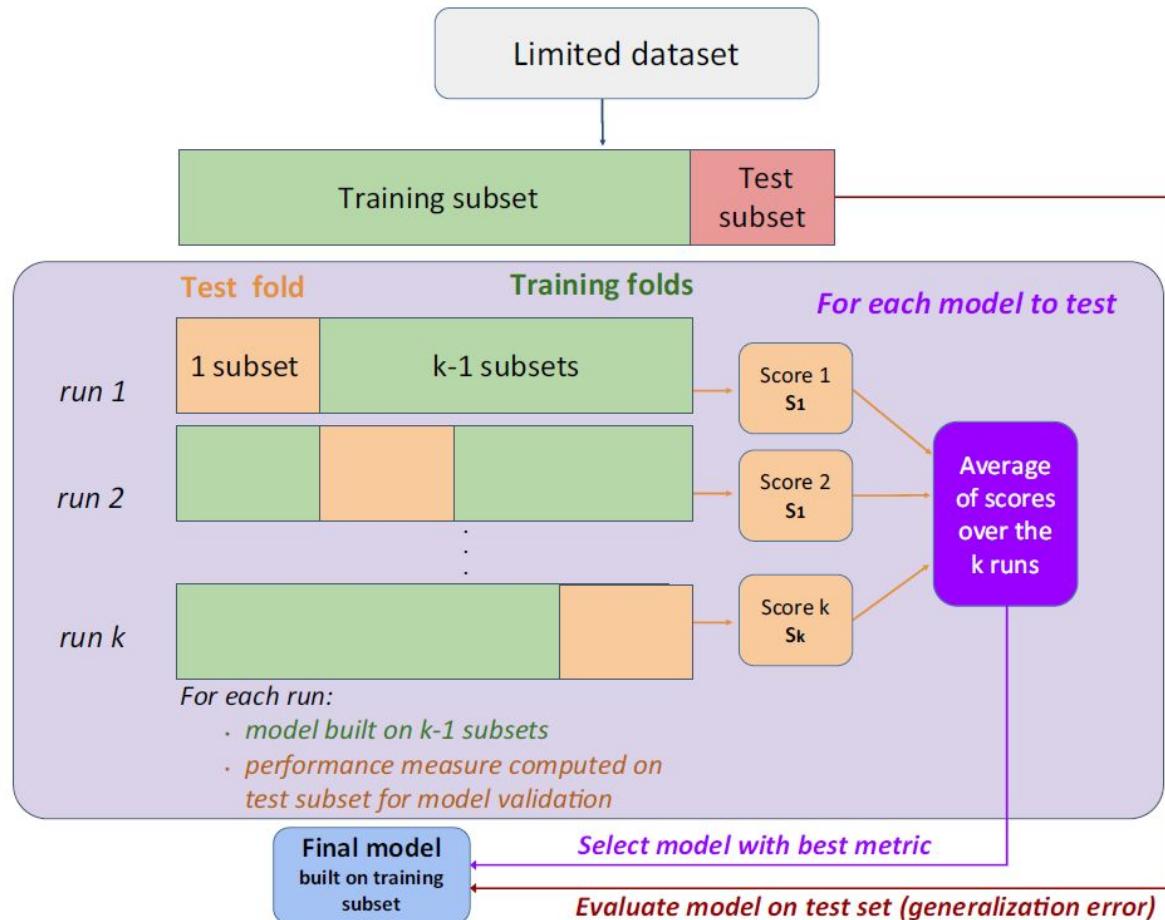
# A basic introduction to machine learning

- QSAR is among the earliest subjects that used machine learning and pattern recognition in drug discovery.
- **Advantages:** technically easy, fast, and many models are useful as filters.
- **Disadvantages:** statistical models cannot capture mechanistic aspects of biochemical interactions, limited ability to debug when a model fails to work, and findings may not be generalizable.



Badillo, Solveig, et al. 2020. “An Introduction to Machine Learning.” *Clinical Pharmacology & Therapeutics*.

# The general practice of training a supervised learning model

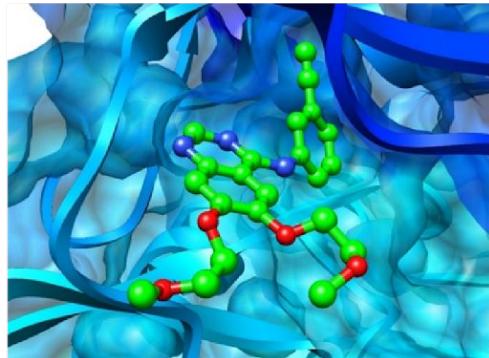


(Left) To assess the generalization ability of a supervised learning algorithm, data are separated into a training subset used for building the model and a test subset used to assess the generalization error (from Badillo *et al.*, 2020) (Right) Temporal validation is especially important for drug discovery, because chemical structures used in the training set may differ substantially from those that will be tested.

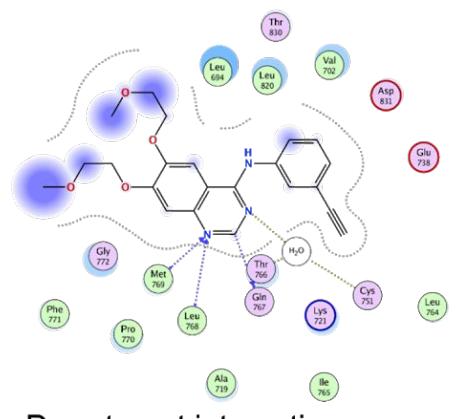
# Summary and Q&A

**A**

## 3D protein structure-based approaches



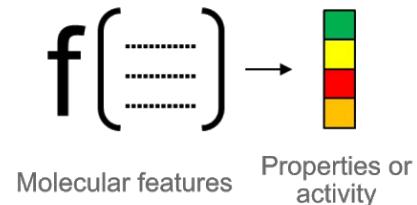
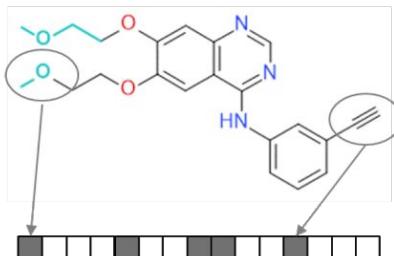
3D model of drug-target complex



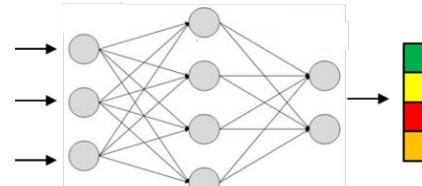
Drug-target interaction map

**B**

## Ligand-based approaches



QSAR



Artificial neural network

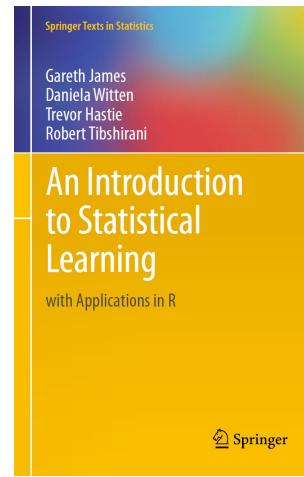
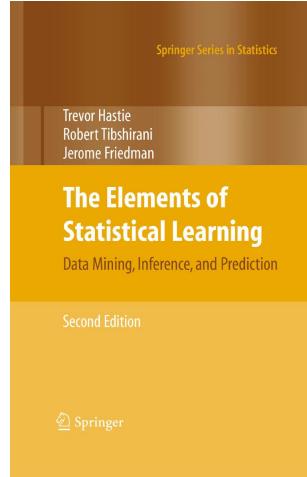
Matched molecular pairs and whole-molecule similarity

Overview of **non-sequence-based, molecular-level modelling techniques**: (A) 3D protein structure-based approaches (B) Ligand-based approaches.

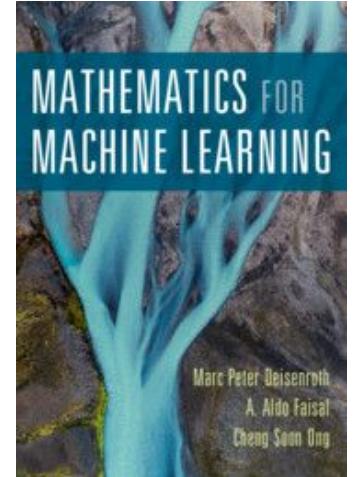
From: Zhang, Jitao David, Lisa Sach-Peltason, Christian Kramer, Ken Wang, and Martin Ebeling. 2020. "[Multiscale Modelling of Drug Mechanism and Safety](#)." *Drug Discovery Today* 25 (3): 519–34.

# Resources for learning about machine learning

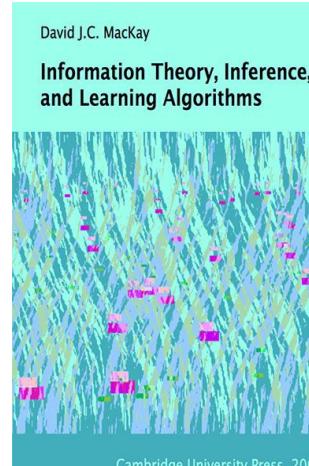
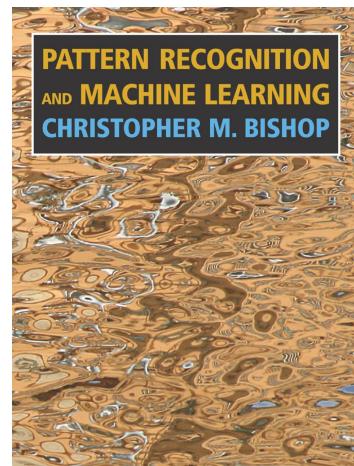
ESL and ISL: From a frequentist view (almost)



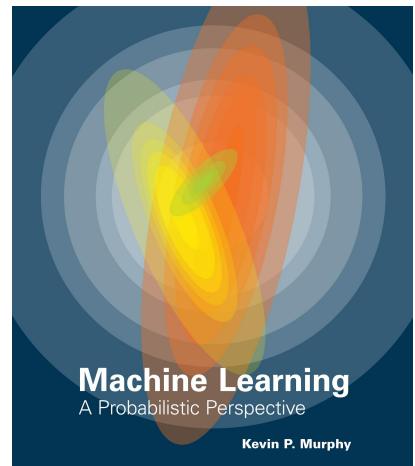
Mathematical foundations



PRML and ITILA: From a Bayesian view



MLaPP: Application oriented, more accessible, and balanced views



# Offline activities

- Read selected pages of *Computational Methods in Drug Discovery* by Sliwoski *et al.* Please submit your results to the Google Form, the link of which will be sent via a separate email.
- Optional and recommended:
  - Fill the anonymous survey #5 (link will be sent via a separate email).
  - Recommended readings:
    - Badillo *et al.* 2020. “[An Introduction to Machine Learning](#).” Clinical Pharmacology & Therapeutics.
    - Jiménez-Luna, José, Francesca Grisoni, and Gisbert Schneider. 2020. “[Drug Discovery with Explainable Artificial Intelligence](#).” Nature Machine Intelligence 2 (10): 573–84..

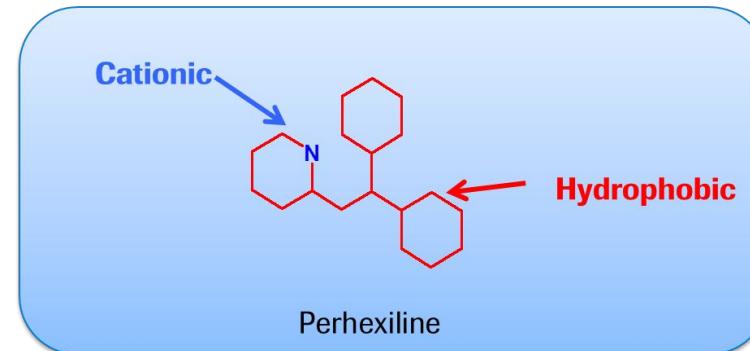
# More about the Free-Wilson analysis

- [A Mathematical Contribution to Structure-Activity Studies](#) by Spencer M. Free and James W. Wilson, Journal of Medicinal Chemistry, 1964, and reviewed by [Kubinyi](#), 1988.
- A Python implementation on [GitHub](#), and a [blog post](#) going through examples, is shared by Pat Walters.
- Free-Wilson nonadditivity is a research topic, for instance see [Cramer et al., 2015](#)
- Source of the example shown in the lecture: QSAR of the [ACCVIP](#) project (The Australian Computational Chemistry via the Internet Project)

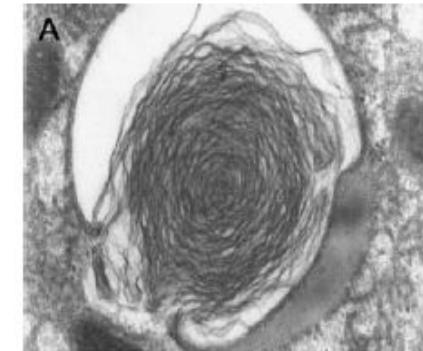
# Drug-induced phospholipidosis is correlated with amphiphilicity

- Phospholipidosis is a lysosomal storage disorder characterized by the excess accumulation of phospholipids in tissues.
- Drug-induced phospholipidosis is caused by cationic amphiphilic drugs and some cationic hydrophilic drugs.
- Clinical pharmacokinetic characteristics of drug-induced phospholipidosis include (1) very long terminal half lives, (2) high volume of distribution, (3) tissue accumulation upon frequent dosing, and (4) deficit in drug metabolism.

Fischer et al. (Chimia 2000) discovered that it is possible to predict the amphiphilicity property of druglike molecules by calculating the amphiphilic moment using a simple equation.



Lüllmann et al., Drug Induced Phospholipidosis, *Crit. Rev. Toxicol.* 4, 185, 1975



Anderson and Borlak, Drug-Induced Phospholipidosis, *FEBS Letters* 580, Nr. 23 (2006): 5533–40.

$$\vec{A} = \sum_i d \cdot \vec{\alpha}_i$$

$\vec{A}$ : Calculated amphiphilic moment

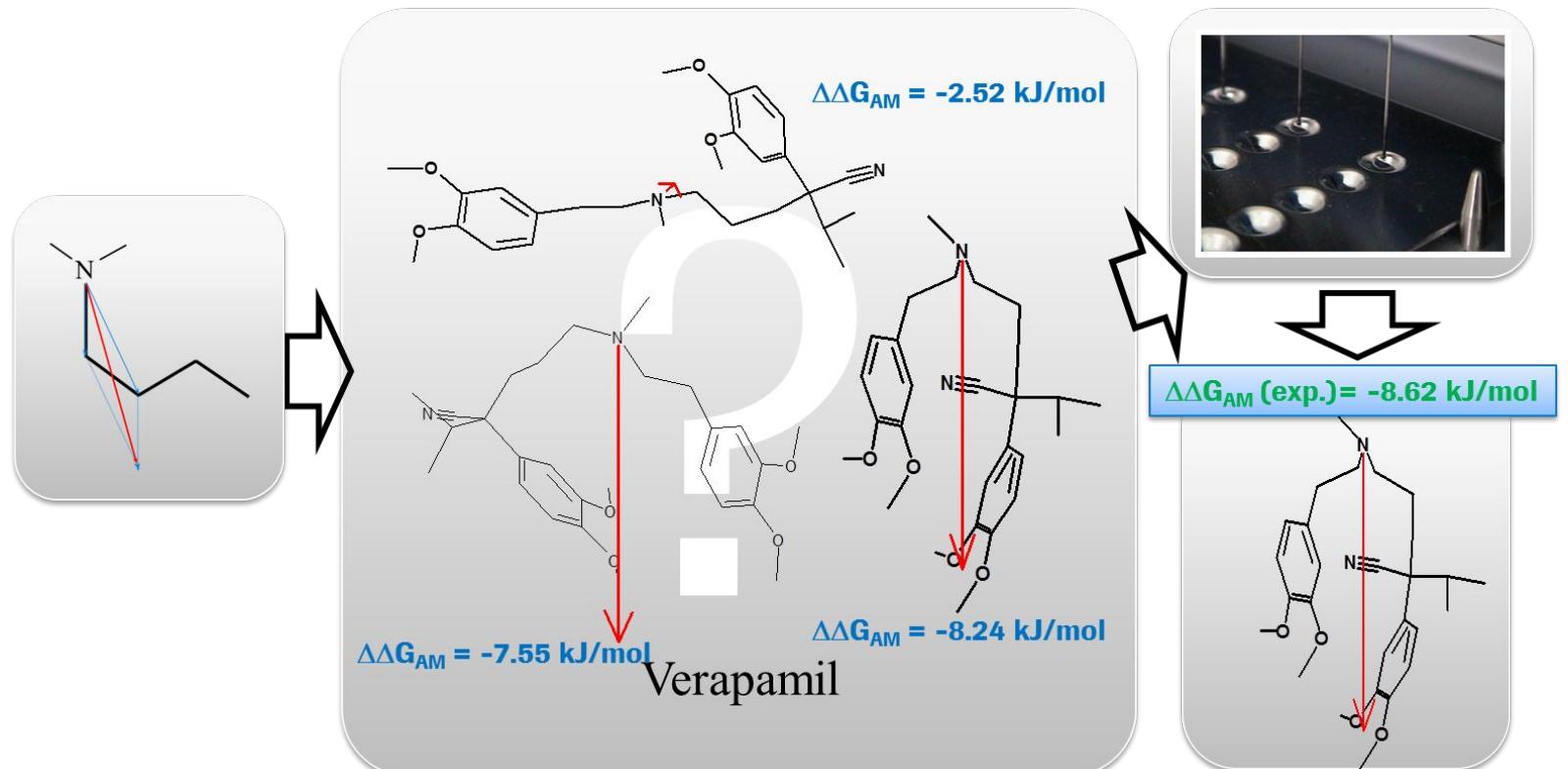
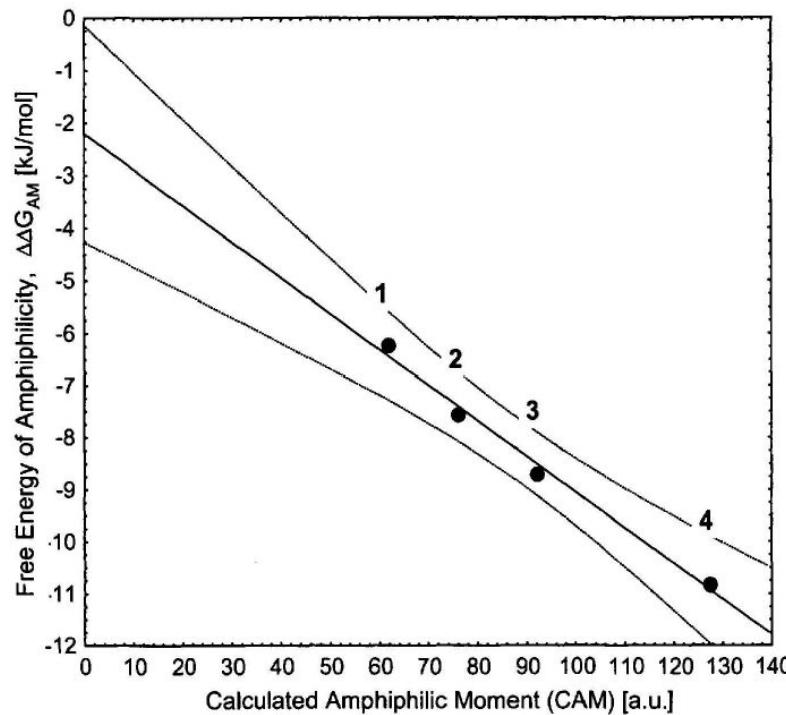
$d$ : distance between the center of gravity of the charged part of a molecule and the hydrophobic/hydrophilic remnant of the molecule

$\vec{\alpha}_i$ : the hydrophobic/hydrophilic contribution of atom/fragment  $i$

**In silico calculation of amphiphilicity property may be used to predict phospholipidosis induction potential**

# In silico prediction of amphiphilicity

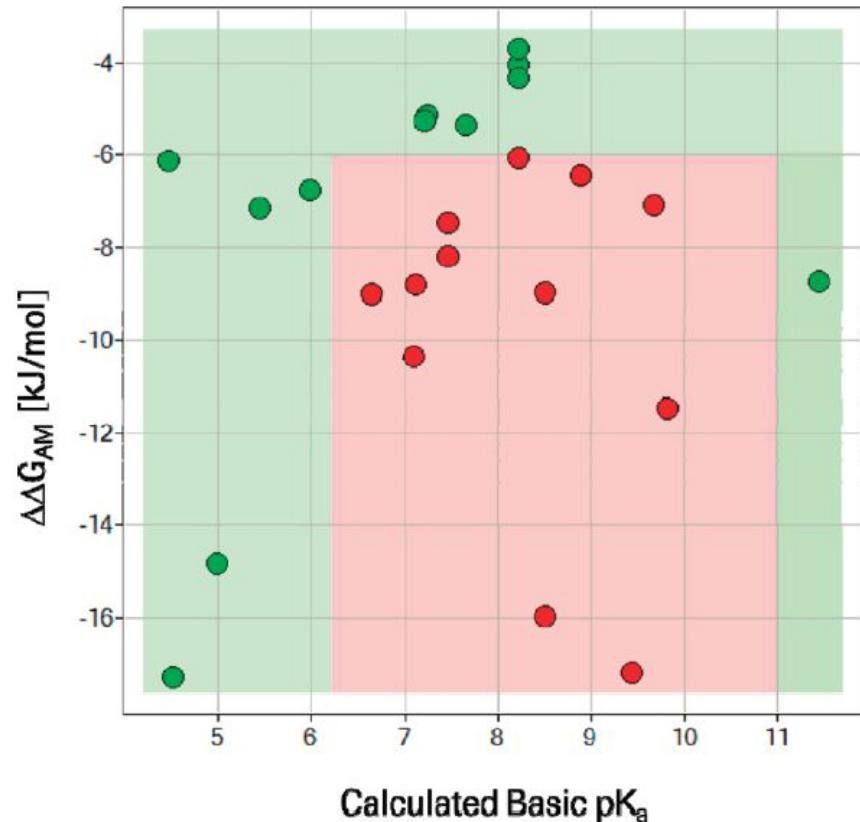
*Development of CAFCA (CAlculated Free energy of amphiphilicity of small Charged Amphiphiles)*



Iterative model building, experimentation, and model refining led to the predictive tool CAFCA

# Validation of in silico phospholipidosis prediction

*Model Validation from 1999-2004*



Plot of amphiphilicity ( $\Delta\Delta G_{AM}$ ) versus calculated basic  $pK_a$  for the training set of 24 compounds. The red area defines the region where a positive PLD response is expected, and the green area defines where a negative response is expected according to the tool.

in vitro/ in vivo	in silico/ in vivo	Exp. PC/ in vivo	In silico/ in vitro	n=36
94%	81%	89%	89%	

in vitro/in silico			n=422
Accuracy [(TP+TN)/(P+N)]	Sensitivity [True Positive Rate]	Specificity [True Negative Rate]	Precision [TP/(TP+FP)]
86%	80%	90%	84%

Fischer et al., J. Med. Chem, 55 (1), 2012

We gained mechanistic insights of phospholipidosis induction by cationic amphiphilic drugs with the model

# Phospholipidosis: lessons learned

- Cationic amphiphilic properties of a molecule is an early marker for safety in drug discovery and early development.
  - Phospholipidosis in dose range finding studies
  - Cardiac ion channel interactions (hERG, sodium channel, ...)
  - Receptor binding promiscuity
  - P-gp inhibition
  - Mitochondrial toxicity in case of safety relevant findings, e.g. in dose range finding studies
- Extreme basic amphiphilic properties should be avoided because of a higher risk of PLD, QT-prolongation, mitochondrial toxicity. However, basic compounds with moderate amphiphilic properties are still a preferred scaffold for many therapeutic areas (especially CNS).
- Generally, some safety liabilities, despite complex underlying biological and chemical mechanisms, can be predicted by molecular modelling well, sometimes with surprisingly elegant models!**

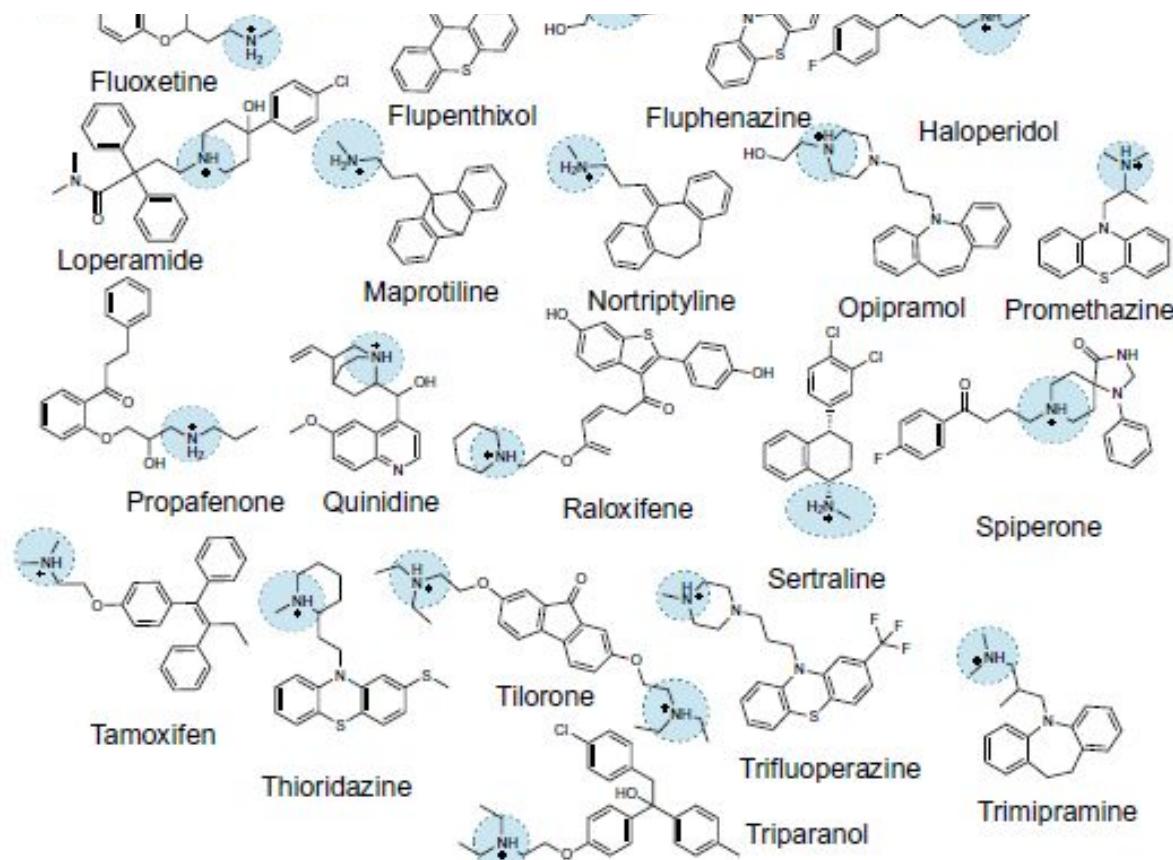


Fig. 1. Representative examples of CADs that are identified in SARS-CoV-2 drug repurposing screens.

Tummino, Tia A., Veronica V. Rezelj, Benoit Fischer, Audrey Fischer, Matthew J. O'Meara, Blandine Monel, Thomas Vallet, et al. "Drug-Induced Phospholipidosis Confounds Drug Repurposing for SARS-CoV-2." *Science* 373, no. 6554 (July 30, 2021): 541–47. <https://doi.org/10.1126/science.abi4708>.

# Summary and Q&A

- **Protein biology and structure determination:** X-ray, NMR, and CryoEM. In case no structure is available, homology modelling can be used.
- **Representation and molecular descriptors of small molecules:** symbolic representations and cheminformatic resources in ChEMBL, molecular descriptors, and Lipinski's Rule of Five.
- **Two views of ligand-target binding:** foundation of ligand-based and structure-based drug design.

# Offline activities

- **Anonymous post-lecture survey of Lecture #5:** <https://forms.gle/BVdgcSbyJYmG8SSA>
- **Required reading:** selected pages of *Evaluation of the Biological Activity of Compounds: Techniques and Mechanism of Action Studies* by Dougall and Unitt and answer questions (see the next slide). Please submit your results to [the Google Form](#).
- **Optional reading** based on your interests:
  - [Machine learning and drug discovery] Mullard, Asher. “What Does AlphaFold Mean for Drug Discovery?” *Nature Reviews Drug Discovery* 20, no. 10 (September 14, 2021): 725–27.  
<https://doi.org/10.1038/d41573-021-00161-0>.
  - [Mathematics and structural biology] Mathematical techniques used in biophysics by J. R. Quine.

# Questions about *Evaluation of the Biological Activity of Compounds: Techniques and Mechanism of Action Studies*

Q1. An important chemical and mathematical concept was not described in the book chapter: what does *the Law of Mass Action* mean?

Q2: Which quantity measures binding affinity directly: dissociation constant ( $K_D$ ) or the concentration of the test compound that produces 50 percent inhibition ( $IC_{50}$ )?

Q3: In Figure 2.3, what do x- and y-axis represent in panel (A) and panel (B), respectively?

Q4: What is a sigmoidal curve?

Q5: Do  $IC_{50}$  values indicate a particular mechanism of action (MoA)?

Q6: In a certain enzymatic assay,, two compounds have the following pIC50 values: 7.2 (Compound A), 9.3 (Compound B). If all other conditions are held constant, what is the relationship between binding affinities of the two compounds with regard to the target?

Q7: Why is DMSO often used in bioassays?

Q8: Can you use your own language to describe what is the Hill function?

Q9: What statistical measure is used to measure the signal-noise ratio in screening? Can you use your own language explaining it?

Q10: Why logarithm (usually base 10) transformation is often preferred to represent quantities such as  $IC_{50}$  and  $K_i$ ?

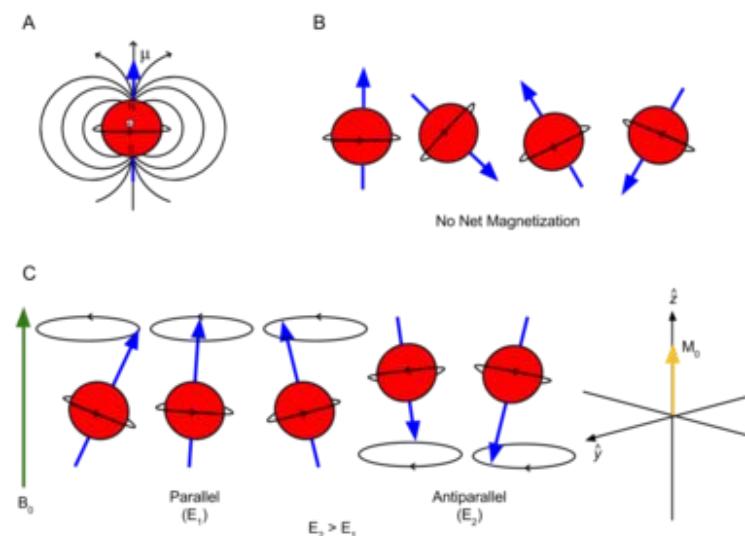
# Resources

# Resources about mathematics behind approaches to determine molecular structure

- **Mathematical and physical foundations**
  - Recommended reading: [Mathematical techniques used in biophysics](#)
  - [Background on imaging physics](#) at xrayphysics.com
- **X-ray diffraction by electrons**
  - An [AMS Feature Column](#) by Tony Phillips
  - Stanford open course [Fourier transform and its applications](#)
- **Nuclear Magnetic Resonance (NMR)**
  - [A beautiful video tutorial](#) about the principles of magnetic resonance imaging (MRI), which is a variant of NMR
- **Cryo-electron microscopy (CryoEM)**
  - [A three-minute introduction to CryoEM](#)
  - [Nobel Prize Talk by Joachim Frank](#)
  - [Talk on Mathematics of CryoEM](#), by Prof Amit Singer, with a manuscript available at arXiv: <https://arxiv.org/abs/1803.06714>



Swiss Light Source, the synchrotron at the Paul Scherrer Institute (PSI), copyright of PSI



Adapted from Bushberg JT, [The Essential Physics of Medical Imaging](#): Lippincott Williams & Wilkins; 2002

# Physics for life-science students

University of Maryland as part of the National Experiment in Undergraduate Science Education (NEXUS): <https://www.compadre.org/nexusph/>

NΦ NEXUS/Physics  
Physics for life-science students